CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

The human population is constantly growing, leading to an increasing demand for food and goods. This significantly impacts our planet. Climate change is a major challenge we need to address through scientific discoveries and engineering innovations.

One such innovation is the development of microbial cell factories - designed (engineered) microorganisms that can transform a variety of sustainable biomasses into useful products such as foods or goods.

Yarrowia lipolytica, a promising cell factory, has gained attention for its ability to produce a wide range of valuable molecules used in the food, biofuel, and pharmaceutical industries. However, to fully leverage *Y. lipolytica's* potential, more research is needed.

In this thesis I investigated the underlying biology of a *Y. lipolytica* strain whose lipid synthesis is disrupted. This strain can be used for production of non-lipid molecules. However, disrupting lipid synthesis induced stress responses, suggesting that a downregulation of these processes might be a better strategy. I then explored the use of urea as an alternative and more sustainable nitrogen source, showing that it does not alter cell physiology and can also reduce issues related to media acidification. I leveraged this information to improve a fed-batch cultivation to produce high titres of itaconic acid, a chemical that finds applications in the food, textile, and pharmaceutical industries. Additionally, I laid the foundations for single-cell transcriptomics to explore cell heterogeneity in bioreactor cultivations and developed a computational framework to minimise the variability of cell cycle genes.

Overall, this thesis explores and expands knowledge in relevant areas to develop *Y. lipolytica* as a microbial cell factory for the sustainable production of non-lipid chemicals.

CHALMERS
UNIVERSITY OF TECHNOLOGY

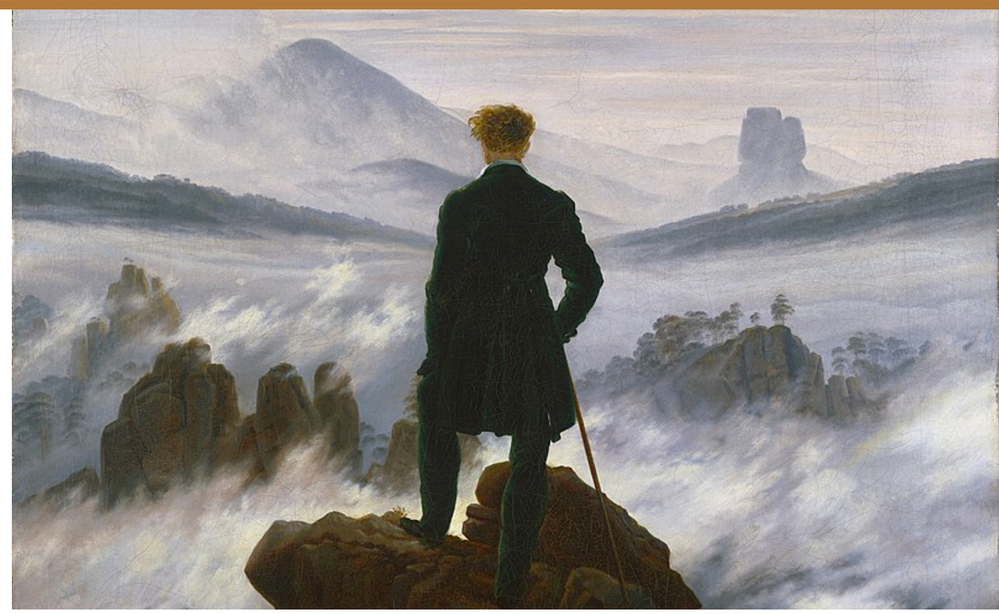
SIMONE ZAGHEN

Multi-omics approaches to unravel regulatory dynamics in yeast bioreactor cultivations

2024

CHALMERS
UNIVERSITY OF TECHNOLOGY

PHD THESIS



Multi-omics approaches to unravel regulatory dynamics in yeast bioreactor cultivations

SIMONE ZAGHEN

DEPARTMENT OF LIFE SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2024
www.chalmers.se

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Multi-omics approaches to unravel regulatory dynamics in yeast bioreactor cultivations

Simone Zaghen

Division of Systems and Synthetic Biology

Department of Life Sciences

CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2024

Multi-omics approaches to unravel regulatory dynamics in yeast bioreactor cultivations

Simone Zaghen

ISBN 978-91-8103-070-9

Acknowledgements, dedications, and similar personal statements in this thesis reflect the author's own views.

Doktorsavhandlingar vid Chalmers tekniska högskola

Ny serie nr 5528

ISSN 0346-718X

Division of Systems and Synthetic Biology

Department of Life Sciences

Chalmers University of Technology

SE-412 96 Göteborg

Sweden

Telephone +46(0)31-772 1000

Cover: *Wanderer above the Sea of Fog* (1818), Caspar David Friedrich

Printed by Chalmers Digitaltryck

Göteborg, Sweden 2024

Multi-omics approaches to unravel regulatory dynamics in yeast bioreactor cultivations

Simone Zaghen
Division of Systems and Synthetic Biology, Department of Life Sciences
Chalmers University of Technology

Abstract

Climate change is a multifaceted problem that requires multiple scientific discoveries and engineering innovations. Among the innovations that have emerged in recent years are microbial cell factories, engineered microorganisms that produce desired molecules through their metabolism.

A promising microbial cell factory is *Yarrowia lipolytica*, an oleaginous yeast that has gained significant traction since it proved a versatile host to produce lipids as well as both bulk and fine chemicals. However, further research is needed to better understand this host and to design better bioprocesses.

To improve the current understanding of *Y. lipolytica* as a microbial cell factory, I combined chemostat cultivations with transcriptomic analysis. I studied the underlying biology of a platform strain with disrupted lipid synthesis, revealing that abolishing storage lipids induces protein misfolding and stress responses. I then explored the use of urea as an alternative and more sustainable nitrogen source, demonstrating that it does not alter the cell transcriptome and can reduce media acidification. I combined this information to improve a fed-batch cultivation to produce high titres of itaconic acid.

Meanwhile, I laid the foundations for single-cell transcriptomics to explore cell heterogeneity in bioreactor cultivations. I performed a proof-of-concept analysis in the well-characterized yeast *Saccharomyces cerevisiae* to understand the potential challenges in translating single-cell transcriptomics to *Y. lipolytica*. I found that cell cycle genes are a major source of variability that needs to be minimized.

The work performed combines bioreactor cultivation with omics analyses to inform and guide future strain improvement. Overall, this thesis explores and expands knowledge in relevant areas to develop *Y. lipolytica* as a microbial cell factory for the sustainable production of non-lipid chemicals.

Keywords

Transcriptomics, RNA-sequencing, single-cell RNA-sequencing, bioreactor, fed-batch, chemostat, itaconic acid

Contents

List of publications

Papers included in the thesis

I) **Zaghen, S.***, Konzock, O.*, Fu, J., & Kerkhoven, E. J. (2023). Abolishing storage lipids induces protein misfolding and stress responses in *Yarrowia lipolytica*. *Journal of Industrial Microbiology and Biotechnology*, Volume 50, Issue 1, 2023, kuad031, https://doi.org/10.1093/jimb/kuad031

II) Konzock, O.*, **Zaghen, S.***, Fu, J., & Kerkhoven, E. J. (2022). Urea is a drop-in nitrogen source alternative to ammonium sulphate in *Yarrowia lipolytica*. *ISCIENCE*, *25*(12), 105703. https://doi.org/10.1016/j.isci.2022.105703

III) Fu, J., **Zaghen**, S., Lu, H., Konzock, O., Poorinmohammad, N., Kornberg, A., Ledesma-Amaro, R., Koseto, D., Wentzel, A., Bartolomeo, F. di, & Kerkhoven, E. J. (2024). Reprogramming *Yarrowia lipolytica* metabolism for efficient synthesis of itaconic acid from flask to semipilot scale. *Science Advances*, *10*(32). https://doi.org/10.1126/SCIADV.ADN0414

IV) **Zaghen, S.**, Jackson C.A., Kerkhoven, E.J., Gresham, D., Quantification of stochastic gene expression in *S. cerevisiae* using single cell RNA-sequencing. Manuscript

Additional papers not included in this thesis

V) Konzock, O.*, **Zaghen, S.***, & Norbeck, J. (2021). Tolerance of *Yarrowia lipolytica* to inhibitors commonly found in lignocellulosic hydrolysates. *BMC Microbiology*, *21*(1), 1–10. https://doi.org/10.1186/s12866-021-02126-0

VI) Konzock, O., Matsushita, Y., **Zaghen, S.**, Sako, A., & Norbeck, J. (2022). Altering the fatty acid profile of *Yarrowia lipolytica* to mimic cocoa butter by genetic engineering of desaturases. *Microbial Cell Factories*, *21*(1), 1–11. https://doi.org/10.1186/s12934-022-01748-x

VII) **Zaghen, S.**, Fu, J., Poorinmohammad, N., De Biaggi, J.S., Lahtvee, P.J., Kerkhoven, E.J. Proteomics analysis of *Yarrowia lipolytica* as chassis for non-lipid products. Manuscript

*both authors contributed equally to the work

Contribution summary

Paper I

Conceptualized the study; designed and performed the experiments; analysed and interpreted the results; wrote and reviewed the manuscript.

Paper II

Conceptualized the study; designed and performed the experiments; analysed and interpreted the results; wrote and reviewed the manuscript.

Paper III

Performed bioreactor cultivations; analysed and interpreted the results of those cultivations; edited and reviewed the manuscript.

Paper IV

Funding acquisition; conceptualized the study; designed and performed the experiments; analysed and interpreted the results; wrote and reviewed the manuscript.

Additional papers not included in this thesis

Paper V

Conceptualized the study; designed and performed the experiments; analysed and interpreted the results; wrote and reviewed the manuscript.

Paper VI

Performed some strain engineering and sample analysis; edited and reviewed the manuscript.

Paper VII

Analysed and interpreted the results; wrote and reviewed the manuscript.

Preface

This dissertation serves as partial fulfilment of the requirements to obtain the degree of Doctor of Philosophy at the Department of Life Sciences at Chalmers University of Technology. The PhD studies were carried out between October 2020 and October 2024 at the Division of Systems and Synthetic Biology (Sysbio) under the supervision of Eduard Kerkhoven and co-supervision of Verena Siewers. Part of the PhD studies, from October 2023 to April 2024, was carried out at the Centre for Genetics and Genomics, New York University under the supervision of David Gresham. The thesis was examined by Ivan Mijakovic. The thesis was funded by Novo Nordisk Foundation (grant NNF20CC0035580), Research Council for Environment, Agricultural Sciences, and Spatial Planning (Formas) (grant 2018-00597), Swedish Research Council (VR) (grant 2019-04624), Åforsk Foundation (grant 23-587), Adlerbertska Forskningsstiftelsen (grant C 2023-0559), and Barbro Osher Endowment (grant SC 2023-000).

Simone Zaghen

October 2024

Abbreviations

C-lim: carbon limitation

C/N ratio: carbon-to-nitrogen ratio

CAD: cis-aconitate decarboxylase

CV: coefficient of variation

FAME: fatty acid methyl esters

GO: gene ontology

GRAS: generally regarded as safe

GSA: gene set analysis

IDH: NAD+ dependent isocitrate dehydrogenase

Log_2FC: log_2 fold change

MTT: tricarboxylic acid transporter

N-lim: nitrogen limitation

NL: nitrogen limiting

NR: nitrogen replete

PCA: principal component analysis

SPE: solid phase extraction

Introduction

I remember studying biochemistry in high school and being intrigued by how complex and intertwined cell metabolism and regulatory networks are. The topic not only fascinated me, but also struck me for its complexity. It came very natural to study biotechnology, to try to understand and hopefully untangle some of this complexity. Fast forward a few years, I was learning the implications of biotechnology on society: from fighting climate change with microbial cell factories, to the ethical issues raised by genome editing. Which better combination than employing computational methods to improve microbial cell factories, and hopefully contribute to reducing our impact on earth?

Modern-day challenges

Climate change

Think about the area where you grew up and compare it to how it was when you were a child. When I think about Pianura Padana (Padan Plain) during my childhood and teen years, I recall foggy winters and humid summers with rare showers. Nowadays, foggy winters are rare, and summers are increasingly marked by extreme weather events such as hailstorms. Given that intensive agriculture, and, in the northern part of the region, wine production, are important activities, consequences on the local economy are becoming evident.

Climate change is the consequence of multiple human activities such as deforestation, farming livestock, and fossil fuel combustion[1]. These activities release significant amounts of greenhouse gases into the atmosphere, disrupting climate patterns and leading to rising global temperatures, melting ice caps, and increased frequency of extreme weather[1]. The impacts of climate change are profound, affecting not only ecosystems and weather patterns, but also societies. Vulnerable populations face disproportionate risks[2], to the extent that the concepts of climate migrant and climate refugee emerged[3].

The urgency to address climate change is clear to (almost) everyone[4]. We intuitively know that it is essential not only to stop using fossil resources, but also to abandon the linear economical model of “take, make, dispose”[5]. We know that we should transition to a circular economy based on sustainable development. We know that resource efficiency, waste reduction, and the respect of natural systems is fundamental. We know that transitioning towards circular economy requires drastic changes in attitudes and habits towards resource extraction, energy use, production and consumption of goods, urban planning and transportation, diet, waste management, and so on[6]. But why, then, hasn’t this transition happened yet?

Wicked problems

You may have guessed from the long list of problems I listed that shifting to a sustainable and circular economy is a complex and challenging task. The interconnectivity of the problems that need a solution and the multitude of actors involved make climate change a super wicked problem. Let's first define a wicked problem, and then clarify the *super* part.

The concept of wicked problem was introduced by Rittel and Webber in 1973 and it relates to problems that are inherently multifaceted and difficult to define and resolve, i.e. the characteristics of a wicked problem are intrinsic to the problem itself[7]. Rittel and Webber defined a series of characteristics typical of wicked problem, that were later summarized in six key points by Conklin[8]:

I) The problem is not understood until after the formulation of a solution: climate change becomes clearer only as we develop and implement solutions, which frequently open other questions or problems.
II) Wicked problems have no stopping rule: it is unclear when a solution is found, and there is no definitive end of when to stop addressing climate change.
III) Solutions to wicked problems are neither right nor wrong, but only trade-offs: solutions can only partially address some issues of complex systems, while either failing or ignoring to address some other issues.
IV) Wicked problems are essentially novel and unique.
V) Every solution to a wicked problem is a one-shot operation, there is no possibility for trial and error, given that every attempt has lasting consequences that alter the formulation of the problem.
VI) Wicked problems have no given alternative solutions.

Now that we defined wicked problems, let's understand why climate change is a *super* wicked problem. What makes a wicked problem *super* is the addition of extrinsic characteristics that are related to the agent trying to solve the problem. In 2012 Levin et al. defined four traits of super wicked problems[9]:

I) The problem is time sensitive i.e. time is running out: at some point the problem might be too acute, and it might be too late to stop or reverse the problem.
II) Those seeking to end the problem are also causing it: every person trying to reduce climate change has contributed to climate change, and everyday activities are major culprits.
III) There is no central authority dedicated to finding a solution, i.e. decision makers do not have control over all the choices required to alleviate the climate change problem.
IV) Policies often irrationally prioritize short-term policies over long-term benefits.

Sustainable development

Now that we defined the scale of the problems, let's talk about how to tackle them. We briefly touched upon sustainable development, but we have yet to define it. In 1987 The World Commission on Environment and Development published a report called Our Common Future, in which sustainable development is defined as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs"[10]. This report also identified three important pillars to address climate change: environmental protection, economic growth, and social equality. Throughout the years the pillars and the goals evolved, until in 2015 the United Nations formulated the 17 Sustainable Development Goals (Figure 1)[11]. Each of the 17 goals is articulated in a list of targets, and each of the targets has one to four indicators to measure progress, providing a framework to achieve sustainable development.

Climate change and sustainable development are multifaceted problems that require many solutions which need to co-exist and be implemented together. Wicked problems are intrinsically complex, and a single easy solution does not exist. Many scientific discoveries and engineering innovations were developed in recent years to tackle climate change; however, it remains uncertain which ones will prove successful. In the meantime, it is crucial to diversify research and development across multiple fields to increase the likelihood of funding and finding successful innovations. Among the many innovations that emerged in recent years we find bioprocessing and microbial cell factories.

Figure 1: The 17 Sustainable Development Goals formulated by the United Nations in 2015.

Bioprocessing

Bioprocessing is the use of biological entities or their components to produce desired products. Which, as sophisticated as it sounds, is an activity that humans have been doing for at least 8000 years[12,13]. Beer and wine production are among the many examples of bioprocesses that have been around for millennia. Cheese, tempeh, kimchi, sourdough bread, and surströmming, a Swedish delicacy, are also bioprocessing products. First employed (consciously or unconsciously) for food and beverage production, bioprocessing is now rapidly expanding to produce a wider range of products.

One of the drivers for the expansion of bioprocessing is the characterization of novel microorganisms[14]. Of the small number of known microorganisms[15], a wide variety produces secondary metabolites as evolutionary strategy to compete with other microorganisms present in the environment[16]. Over the years, scientists realized the potential of non-conventional organisms and the range of molecules produced by them: big is the interest in harnessing this biodiversity for medical, agricultural, and industrial purposes[17]. Another key driver for the expansion of bioprocessing is the development of genetic and metabolic engineering[18–20], which resulted in the development and definition of microbial cell factories.

These two approaches are not mutually exclusive. Selecting the proper host organism for a specific bioprocess is crucial for minimizing the amount of genetic engineering needed to develop a microbial cell factory, saving time and money that would otherwise be spent on engineering functions already present in other organisms. At the same time, selecting a host organism for which genetic tools are available is important to speed up the development of the microbial cell factory.

Microbial cell factories

Microbial cell factories are engineered microorganisms, designed to produce a desired product through their metabolic processes[21].

We can make an analogy between microbial cell factories and conventional factories. A traditional factory takes raw materials and energy as inputs, processes them with machinery, and produces a product, hopefully desired by consumers. Similarly, in microbial cell factories, the substrate (a carbon and energy source) act as raw material. Cell enzymes function as the machinery (native-existing pathway) that transform the input into the final product (Figure 2). The development of genetic engineering enabled scientists to modify the machinery within the factory: we can now remove or introduce new machinery, either from the same factory or from different factories; we are now getting to a stage where we can even design new machinery to produce novel products[21] (Figure 2).

The appeal of microbial cell factories for sustainable bioproduction derives not only from the range of products that are possible to produce, but also from the type of substrates that can be used. Traditional factories generally rely on fossil fuel-derived inputs, such as petroleum, which are not sustainable and harmful for the environment. On the other hand, microbial cell factories can use a wide variety of sustainable raw materials, which can be classified in three generations[22]. First generation biomasses include sugars derived from corn, soy, and sugarcane. Even though these inputs are renewable, they compete with food production and supplies. Therefore, research has focused on second generation biomasses such as agricultural residues and forestry by-products, which do not compete with food supplies. Second generation biomasses promote a circular economy since waste material gets transformed into a valuable product. Third generation biomasses involve the use of greenhouse gases such as CO_2, CO, methane, formate, and methanol as substrates for bioprocessing, capturing greenhouse gases and converting them into valuable products.

Now that we've seen why microbial cell factories are important for sustainable development, we need to dive a bit deeper into how to engineer microbial cell factories to produce desired molecules at feasible titres, rates, and yields.

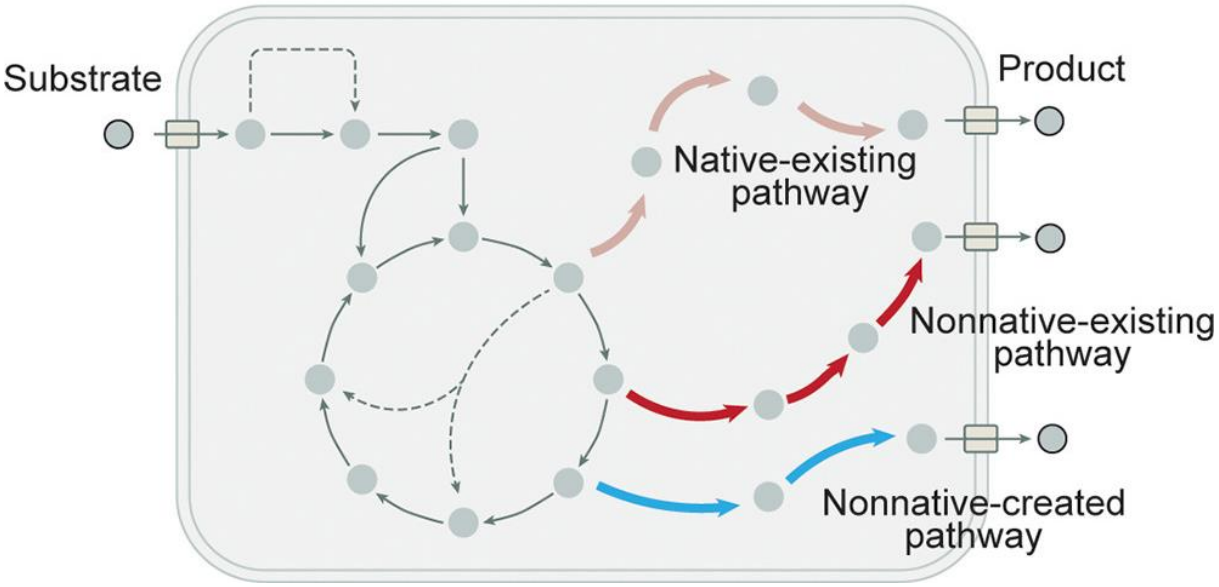


Figure 2: Schematic representation of a microbial cell factory (source[21]). Circles indicate metabolites and arrows indicate pathways.

Engineering biology

Microbes evolved to increase the likelihood of their survival, but this objective frequently does not coincide with the production of molecules we desire. Even when evolution does coincide with the production of a molecule of interest, titres are generally too low for an industrial scale production that is economically feasible. That's where metabolic engineering comes to the rescue.

Metabolic engineering involves the modification of metabolic pathways by introducing, removing, or modifying genes to fine-tune regulatory processes within cells[23,24]. The goal is generally to increase the production of a specific molecule or to introduce the production of novel molecules. Asides from production at high titres of the desired molecule, other characteristics are necessary for a microbial cell factory to achieve a viable bioprocess[25]. A microbial cell factory should:

- Grow fast and at a high cell density to ensure high product titre, rate, and yield.
- Be robust to a wide range of pH, temperatures, salt, and inhibitor concentrations to withstand different production conditions.
- Be genetically stable over prolonged cultivation times to ensure consistent production.
- Be genetically tractable to optimize metabolic pathways, enhance production, and introduce new functionalities.
- Consume multiple carbon sources to enable more sustainable bioprocesses based on second or third generation biomasses.
- Have efficient metabolic pathways to minimize by-product formation and energy loss in non-productive activities.

While this might sound straightforward on the theoretical level, it is a complicated challenge. But why is engineering biology so complicated?

Let's make another analogy. Classic engineering involves designing and assembling systems in which the function of each component is known. For example, building a radio: each component has a well-defined function, a blueprint is available, and we have knowledge on how each component contributes to the final product.

In contrast, engineering biology presents a unique set of challenges. Instead of using components with known function to build a system from scratch, we were gifted by evolution a complex "radio", i.e. microorganisms. This biological radio operates through regulatory networks that we do not fully know or understand, and for which we do not have a blueprint. To further complicate this, we are trying to engineer new features in a not-completely-known system.

Another reason why engineering biology is complicated is related to evolution: cells evolved regulatory networks to ensure homeostasis when external conditions change. Rewiring metabolism to produce a desired molecule needs to circumvent these regulatory networks that evolved for millennia, and that we do not yet completely know and understand.

Design, build, test, learn cycle

As a result of the limitations and challenges described in the previous paragraph, developing economically viable microbial cell factories can require 6–8 years and over $50 million[25]. Engineering a cell factory involves an empiric approach: several rounds of trial and error are required, and a framework based on the design-build-test-learn cycle emerged over the years[18,26,27].

At the start of the cycle, scientists formulate the design of a microbial cell factory, producing a desired molecule or consuming desired substrates. The microbial cell factory is then constructed with the help of genetic engineering tools that have been developed in recent years. Following construction we have the testing phase, in which analytical chemistry and screening techniques are employed to test the microbial cell factory and compare the actual output against the expected output. This leads to the learning phase, where detailed analysis such as pathway and omics analysis are used to understand why there are some discrepancies between expected and actual outcomes.

One of the goals of the learning phase is to clarify how the microbial cell factory functions to inform the next cycle iteration. Since we are engineering a system that we do not fully understand due to our incomplete knowledge, some modifications will produce unexpected outcomes which can be used to improve our understanding of the microbial cell factory. To accelerate and reduce the cost of the design-build-test-learn cycle, an increasing number of methods emerged over the years, including systems biology.

Systems biology

Systems biology aims to understand the emerging properties of a biological system that cannot be understood by studying their individual components in isolation. The goal is to understand how a biological system is built, which are its components, and how these components function and interact with each other. By building a holistic view of the biological system, systems biology can help guide experimental design to improve the design process and increase produduction[18,28]. This is accomplished by developing models that simulate and predict biological behaviour[29,30].

Applying environmental perturbations or genetic modifications can provide information on how various parts of cell metabolism interact with each other. One way to modify the systems is through genetic engineering, such as using CRISPR to edit the genome by inserting or removing specific genes. Another way to understand how different parts of the cellular system operate is by altering the environmental conditions in which the system is operating.

Thanks to omics technologies, combined with analytical methods, it is possible to measure the effects of modifications and perturbations, to identify how different genes, pathways, and biological processes are affected[28,30]. By understanding how each component contributes to the system's function and behaviour, we can gain insights into its functioning and optimize its performance.

Multi-omics approaches, and integrating data from genomics, transcriptomics, proteomics, and metabolomics, can provide a comprehensive view of the molecular and cellular processes, enabling the identification of key regulatory networks and interactions within the microbial cell factory. This helps identifying specific targets for further genetic modifications and adjustments in environmental conditions, significantly accelerating the design-build-test-learn cycle to construct robust and economically viable microbial cell factories.

Transcriptomics

Let’s briefly define one of the reoccuring methods in this thesis. Transcriptomics aims to sequence and quantify the transcriptome, i.e. the complete set of transcripts in a cell. The central dogma of biology states that genes are transcribed into mRNA, which is then translated into proteins. These proteins perform different tasks within the cell thanks to their catalytical activity[31]. Transcriptomics measures the quantity of mRNAs and infers that an increase in transcript quantity corresponds to an increase in protein quantity and catalytical activity.

Transcriptomics can be performed on a population of cells (bulk transcriptomics), or on a single cell (single-cell transcriptomics). Bulk transcriptomics measures the average gene expression of a large population of cells, providing a broad overview on their gene expression program, but potentially masking cell heterogeneity. Single-cell transcriptomics quantifies gene expression of individual cells, and can be used to reveal cellular heterogeneity, cell subpopulations, and dynamic regulatory processes. However, single-cell transcriptomics involves more complex data analysis, can be affected by higher noise and variability, and is significantly more expensive.

Understanding the transcriptome is essential to interpret the functional elements of the genome, to reveal the molecular processes that are taking place in a cell, and to infer which regulatory mechanisms underline specific phenotypic responses. Bulk and single-cell transcriptomics offer complementary perspectives on gene expression and cellular function. Deciding which is more suitable depends on the scientific question and experimental design.

Models and their limits

Before discussing which organism to transform into a microbial cell factory, I want to spend a few words on the role and limitations of models. I am particularly fond of the spherical cow metaphor:

> Milk production at a dairy farm was low, so the farmer wrote to the local university, asking for help from academia. A multidisciplinary team of professors was assembled, headed by a theoretical physicist, and two weeks of intensive on-site investigation took place. The scholars then returned to the university, notebooks crammed with data, where the task of writing the report was left to the team leader. Shortly thereafter the physicist returned to the farm, saying to the farmer, "I have the solution, but it works only in the case of spherical cows in a vacuum."

Intuitively, approximating the cow to a spere and increasing its size might lead to an increase in milk production. However, this approach will only work within certain limits before other factors will become significant. For example, the body and the head of the cow, initially not modelled, will eventually need to be considered. Using geometrical approximations, body and head can be approximated as spheres, while the neck can be approximated to a cylinder. When increasing the size of the cow, body and head volume will grow with the cube of the radius ($V = \frac{4}{3}\pi r^3$), while the neck will grow with the square of the radius ($V = \pi r^2 h$). At some point, the neck would not be able to sustain the head, and to use a euphemism, milk production would stop.

Models can be good approximations of reality withing certain limits, but they can lack or oversimplify crucial elements. For instance, for a microbial cell factory we might not know about certain regulatory mechanisms, like the neck in the cow metaphor. "All models are wrong, but some are useful", George Box said[32]. Since models are approximations of reality, which we do not fully understand, we need to recognize that they are only valid under specific assumptions and within certain limits, of which we need to be aware of and not forget about.

This is particularly relevant in biology, where our knowledge of the system is limited. This limitation forces us to build simplified models that work under certain assumptions and within specific boundaries. By revising models, challenging their assumptions, and pushing their boundaries, we can uncover aspects that we may have overlooked or oversimplified. A good model should aim to be simple enough, while maximizing accuracy within its underlying assumptions.

Yarrowia lipolytica

Now that we clarified the background and key principles of the thesis, let's dive into which organism we will use for bioprocessing.

Many biological entities can be used for bioprocessing thanks to the sheer biodiversity we were gifted from nature and evolution. There is currently a big rise in research on non-conventional yeasts, and scientists are trying to find microorganisms with peculiar and desirable characteristics that can be leveraged to develop bioprocesses[33,34].

During my PhD I mostly focused on the non-conventional yeast *Yarrowia lipolytica* but also worked with the well-characterised yeast *Saccharomyces cerevisiae*. In the introduction I will mostly describe *Y. lipolytica* and provide an overview on why it is such an interesting microorganism. I will spend a few words on *S. cerevisiae* in Chapter 4.

Historically classified as non-conventional yeast, *Yarrowia lipolytica's* status is now changing due to several factors: extensive research being performed, genetic engineering making it easier to manipulate, used in a growing number of industrial applications, and receiving regulatory approval (generally regarded as safe, GRAS). However, some genes and regulatory mechanisms are still unknown, as we will see later, especially compared to the conventional organism *S. cerevisiae*. All the research performed on *Y. lipolytica* is not happening by chance, but for how promising *Y. lipolytica* is.

Y. lipolytica raised interest several decades ago as host for heterologous protein production due to its ability to secrete high levels of proteins[35–38]. The initial enthusiasm rapidly evolved due to its oleaginous nature, i.e. the ability to produce high amounts of lipids. Since then, various genetic tools have been developed and optimized to allow the quick and precise genetic engineering, including CRISPR-CaS9[39,40].

Today, *Y. lipolytica* is regarded as a promising microbial cell factory, well-suited for a wide range of biotechnological applications[41,42]. Several research groups are employing its metabolic capabilities and its ample acetyl-CoA supply for production of food oils[43,44], flavonoids[45,46], commodity chemicals[47,48], pigments[49,50], pheromones[51–53], plastic degradation[54], and so on. The field is rapidly expanding, and an increasing number of start-ups are leveraging *Y. lipolytica* as chassis strain, not only for lipid derived products, but also for non-lipid products[42].

Y. lipolytica is not only promising for the wide range of molecules it can produce. This yeast is also characterized by the ability to utilize a diverse array of sugars such as glucose, fructose, mannose, and other hydrophobic carbon sources such as fatty acids, alkanes, and glycerol[55]. *Y. lipolytica* can also withstand harsh industrial conditions, such as high temperatures and osmotic pressures, enhancing its utility in various processes. This highlights its adaptability and efficiency in various fermentation processes.

Aim of the thesis

The overall goal of the thesis is to advance our understanding of *Y. lipolytica* as a microbial cell factory, contributing to efforts to address the wicked problem of climate change.

In the early stages of my journey, collaborators developed a platform strain of *Y. lipolytica* in which the lipid synthesis is disrupted to redirect carbon flux towards itaconic acid. We thought it would be insightful to study this strain to guide future strain development. In **paper I** we combined chemostat cultivations with transcriptomic analysis, revealing that abolishing storage lipids induces protein misfolding and stress responses.

To improve our *Y. lipolytica*'s bioprocesses, we explored the use of urea as an alternative and more sustainable nitrogen source, demonstrating that it does not alter the transcriptome and can reduce media acidification (**paper II**). We combined this information and by using urea as a nitrogen source we improved a fed-batch cultivation and increased itaconic acid titres (**paper III**).

Meanwhile, my interest in omics analysis evolved, and I sought a collaboration at New York University to perform single cell transcriptomics (**paper IV**). The analysis was performed on the well-characterized yeast *Saccharomyces cerevisiae* as a proof-of-concept, with the goal of translating single-cell transcriptomics to *Y. lipolytica*.

Chapter 1 – Effects of disrupting lipid synthesis

Summary

Lipid biosynthesis requires high amounts of acetyl-CoA, and it wasn't long before metabolic engineers diverted the acetyl-CoA flux from lipids towards commodity and added-value chemicals, such as flavonoids (naringenin[45], eriodictyol[56], taxifolin[56]), polyketides (triacetic acid lactone[47], resveratrol[45]), and terpenoids (lycopene[57], β-Carotene[58], limonene[59]).

Disrupting storage lipid accumulation is a justifiable strategy to enhance the production of desired chemicals in *Y. lipolytica*. However, the impact of these deletions on cell physiology and regulation has yet to be investigated.

Here we show that under nitrogen limitation disrupting lipid synthesis leads an enrichment of the unfolded protein response, and an enrichment of several biological processes related to protein refolding and degradation. Additionally, cells with disrupted lipid synthesis show an altered lipid class distribution with an abundance of potentially cytotoxic free fatty acids under.

Based on these results, we conclude that to optimize our platform strain of *Y. lipolytica*, it is preferable to downregulate the genes involved in lipid synthesis rather than delete them. This approach could ensure that these genes remain functional within the cell, maintaining homeostasis without being expressed to the extent that they consume acetyl-CoA, which can otherwise be utilized to produce other valuable products.

Introduction

Why disrupting lipid synthesis?

Cell metabolism has a bow-tie structure, where all carbon sources are converted to 12 precursor metabolites that are then used for the synthesis of all cellular building blocks and secreted metabolites[25]. A strain with a high flux through a molecule at the centre of the bow-tie can become a platform strain for synthesizing products derived from that same intermediate. For example, from acetyl-CoA, one of the 12 precursors, it is possible to synthetize lipids, polyketides, and many other bioproducts used in biochemical, biofuel, and pharmaceutical industries[60].

A platform strain with high supply of acetyl-CoA is desirable for many applications, and *Y. lipolytica* is an ideal candidate due to its high acetyl-CoA flux under nitrogen limitation[61,62]. While this flux goes towards storage lipid accumulation in wild-type strains, it can be redirected towards the production of other bioproducts at high titres thanks to metabolic engineering.

How to disrupt lipid synthesis

Lipid production can be disrupted by deleting four genes (Figure 3): *DGA1* (YALI1_E38810g), *DGA2* (YALI1_D10264g), *LRO1* (YALI1_E20049g), and *ARE1* (YALI1_F09747g)[63].

- *DGA1* and *DGA2* encode enzymes that catalyse the final step of triacylglycerol formation, using acyl-CoA to convert diacylglycerols into triacylglycerols[63,64]. *DGA2* has also been reported to affect the size and morphology of lipid droplets[63].
- *LRO1* codes for a triacylglycerol synthase that is acyl-CoA independent and uses phospholipids as acyl-donors to convert diacylglycerols into triacylglycerols[64].
- Are1p is essential for sterol esterification, and the deletion of the encoding gene (*ARE1*) abolished sterol ester synthesis[63].

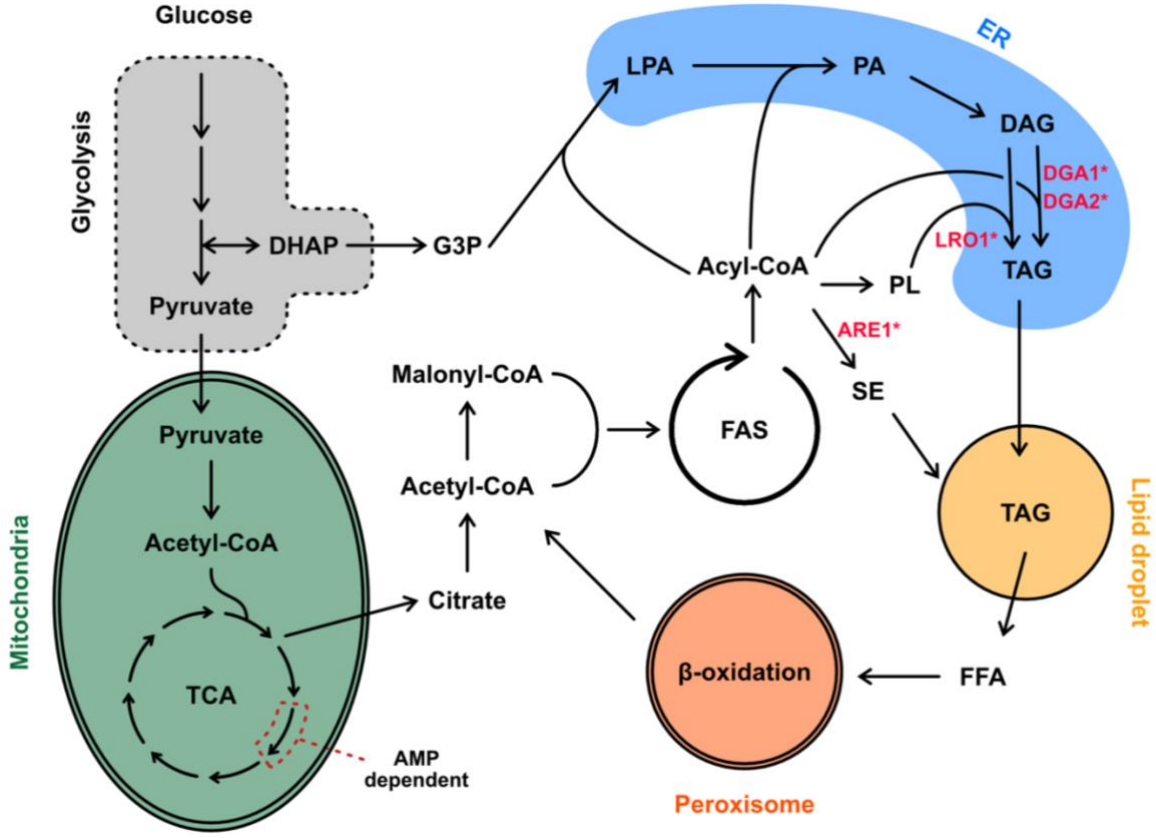


Figure 3: Lipid metabolism in *Y. lipolytica*, adapted from[43]. In red and marked with an asterisk are the genes deleted in the Q4 strain to disrupt lipid accumulation. *DAG: diacylglycerol; DHAP: dihydroxyacetone phosphate; ER: endoplasmic reticulum; FAS: Fatty acid synthase; FFA: free fatty acid; G3P: Glyceraldehyde 3-phosphate; LPA: lysophosphatidic acid; PA: Phosphatidic acid; PL: phospholipid; SE: sterol ester; TAG: triacylglycerol; TCA: tricarboxylic acid cycle.*

Decreasing lipid accumulation by deleting one or more of these genes increases production of added-value and commodity chemicals. For instance, Shi et al. increased β-farnesene titres by 56% after deleting *DGA1* and *DGA2*[65]. Similarly, itaconic acid titres were almost doubled by deleting *DGA1*, *DGA2*, *LRO1*, *ARE1*[48] (Paper III, figure 3A, strains JFYL023 vs JFYL013).

Goal of the project

A *Y. lipolytica* strain with deletions of *DGA1*, *DGA2*, *LRO1*, and *ARE1*, known as Q4 strain, was previously reported[63,66] and was used in Paper III to redirect the acetyl-CoA flux away from lipid synthesis and toward itaconic acid production.

Eliminating storage lipid accumulation is a justifiable strategy to enhance the production of desired chemicals in *Y. lipolytica*. However, the impact of these deletions on cell physiology and regulation has yet to be investigated.

Studying the effects of genetic and environmental perturbations is crucial to improve our understanding of the biological system, uncover unknown aspects of its function, and guide the next rounds of strain design.

Experimental setup

To elucidate the impact of disrupting lipid synthesis, we performed chemostat cultivations on the Q4 and on the wild-type strain (normal lipid phenotype). Since nutrient limitation impacts gene expression in *Y. lipolytica*[61,62,67], we cultivated the strains under different C/N ratios. The C/N ratio is the molar carbon-to-nitrogen ratio, which is a crucial factor that influences microbial metabolism and growth. We tested carbon limitation (C-lim, C/N ratio 3) and nitrogen limitation (N-lim, C/N ratio 116). After at least four-volume changes we measured physiological parameters, lipid abundance, composition, class distribution and we sampled for transcriptomics.

We selected a suitable C-lim C/N ratio by performing shake-flask cultivations with C/N ratios between 1.45 and 20 (Figure 4A). We tested the Q4 and a lipid overproducer strain (OKYL049[68], Δ*are1, DGA1* overexpression) since they show opposite phenotypes that might affect the threshold between carbon and nitrogen limitation. We selected a C-lim C/N ratio of 3 from this experiment. We selected a N-lim C/N ratio of 116 based on literature[66].

We chose pH-controlled chemostat cultivations to ensure a highly controlled environment that increases reproducibility. Additionally, the two strains (OKYL029 and Q4) have different growth dynamics (Figure 4B) and a chemostat culture allows to control the growth rate by setting the dilution rate. This will reduce growth-related variability, and ensure comparable results between strains with different growth dynamics.

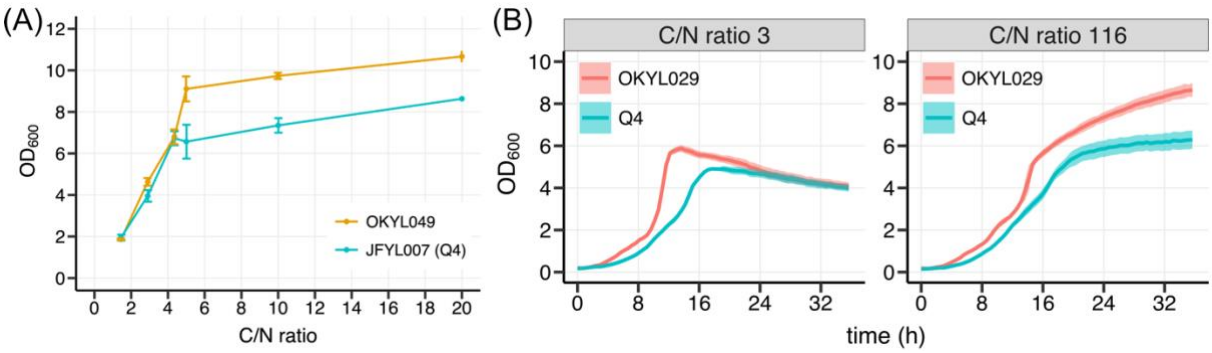


Figure 4: (A): Y. lipolytica was grown in delft media for 72 hours. The media composition was kept constant, but the glucose concentration was varied to produce C/N ratios. The OD_{600} was measured after 72h of cultivation and plotted against the C/N ratio. C/N ratios between 1.45 and 4.43 are carbon limiting for both strains. At higher C/N ratios, the nitrogen becomes limiting, and increasing the glucose concentration doesn't have a major effect on OD_{600}. Dots represent the average OD_{600} of triplicates, and error bars represent the standard deviation. (B) Strains OKYL029 and Q4 were cultivated in 96-wells plates with C/N ratio 3 (C-lim, left panels) or C/N ratio 116 (N-lim, right panels). OD_{600} was measured with the growth profiler every 30 minutes. The curves represent the average of triplicates, and the shadowed areas the standard deviation.

Results and discussion

Impact of gene deletions on cell physiology and lipid composition

The first thing we examined is the effect that the gene deletions have on cell physiology and on lipid composition. We compared the two strains under both C-lim and N-lim to determine which metabolic processes and regulatory networks might be influenced by these deletions, and to identify under which conditions these effects can be observed.

Under C-lim, we observed that cell dry weight and lipid content remained unaffected by the four deletions in the Q4 strain. Both strains had similar biomass yields, lipid yields, and specific glucose uptake rates (r-glucose) (Figure 5).

Under N-lim, the Q4 strain showed a decrease in cell dry weight, lipid content, and lipid yield. Despite these, both strains maintained similar biomass yields and identical specific glucose uptake rates (Figure 5).

We then investigated how the four deletions in the lipid pathway affect abundance and chain length of the lipid. We performed lipid extraction and converted the fatty acid chains of all lipids into fatty acid methyl esters (FAME). We then analyzed the distribution of the five most dominant fatty acids: palmitic acid (C16:0), palmitoleic acid (C16:1), stearic acid (C18:0), oleic acid (C18:1), and linoleic acid (C18:2). The deletions affected the lipid composition in both C/N ratios. The C16:0 fraction showed no statistically significant difference between strains (p-value > 0.01), regardless of the C/N ratio. However, other fatty acids (C16:1, C18:0, C18:1, C18:2) are significantly different (p-value < 0.01) between the two strains, in both C/N ratios.

The scale of the change is generally larger in N-lim, but the direction of change is the same (either more abundant in both C/N ratios or less abundant in both C/N ratios).

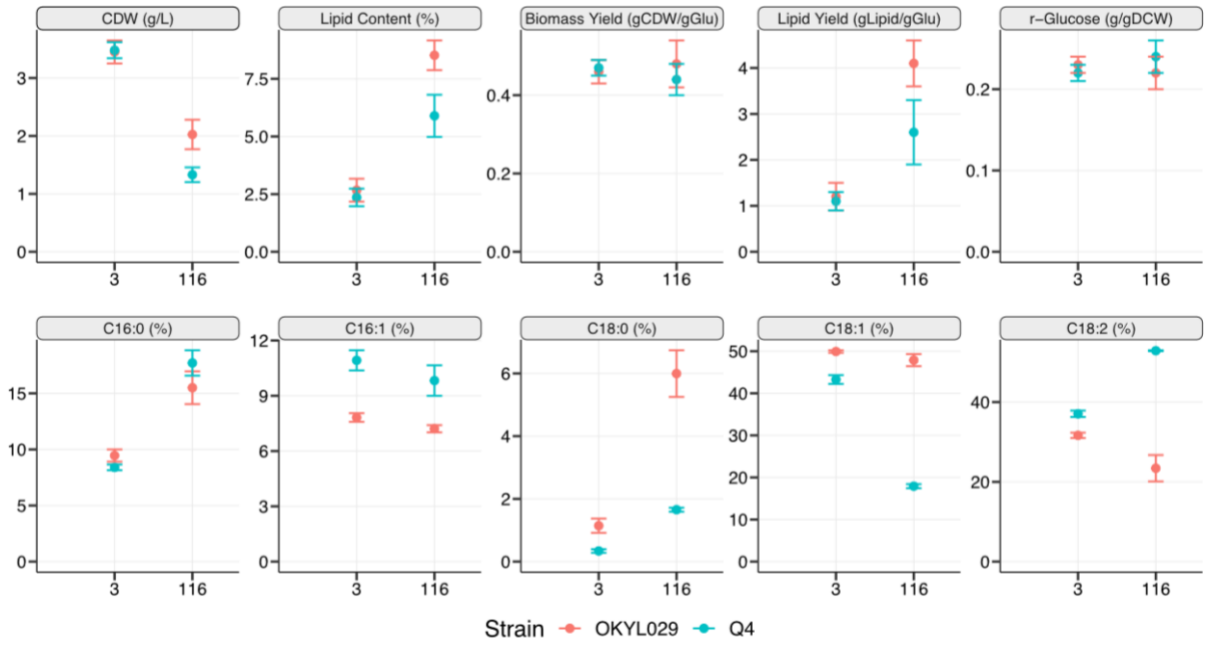


Figure 5: Physiological and lipid composition changes of the strains OKYL029 and Q4 in C-lim (C/N ratio 3) and N-lim (C/N ratio 116). Lipid content is calculated as % of the lipids on the cell dry weight, and the strains' fatty acid composition is calculated as the % of each chain length on the total amount of lipids. Displayed are the average (dot) and standard deviation (error bar) of at least three replicates.

The FAME analysis only provides insights into the changes in the overall lipid composition of the cell. However, it does not distinguish between various lipid classes, such as neutral lipids, free fatty acids, and phospholipids. Neutral lipids include diacylglycerols, triacylglycerols, and sterol esters; phospholipids are primarily found in cell membranes. Since the Q4 strain exhibited altered lipid chain abundances, we employed solid-phase extraction (SPE)[69] to separate and quantify these distinct lipid classes (Figure 6).

Under C-lim, SPE revealed no significant differences (p-value > 0.01) between the Q4 and wild-type strains (Figure 6) and both strains have similar proportions of phospholipids, nutral lipids, and free fatty acids. While the phospholipids fraction shows minor differences in chain length distribution (less than 5%), the neutral lipid fraction does not. In the free fatty acid fraction, the Q4 strain has more unsaturated lipids. However, free fatty acids constitute a small portion of the total lipid content (less than 7%).

Under N-lim, the wild-type strain predominantly contained neutral lipid, while the Q4 had reduced neutral lipid and increased phospholipids. The free fatty acid fraction in the Q4 is three times larger than in the wild-type strain, indicating that the Q4 is synthesising free fatty acids but lacks the ability to incorporate them in triacylglycerols. Regardless of the lipid fraction, the Q4 strain has higher levels of C16:1 and C18:2, while C18:1 was more abundant in the

wild-type. The saturated fatty acids (C16:0 and C18:0) only show minor changes between strains.

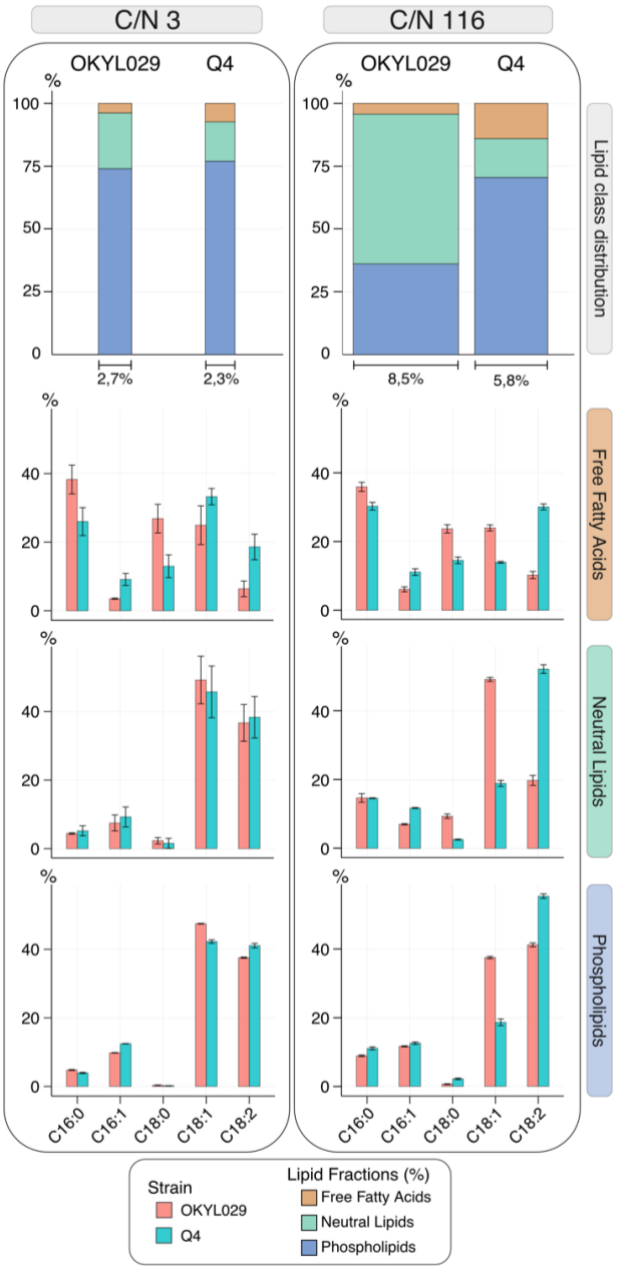


Figure 6: Solid phase extraction (SPE) of Q4 and OKYL029 in C-lim and N-lim conditions. Stacked bar charts (lipid class distribution) represent the share of each lipid class detected by SPE over the total amount of lipids present in the cell. The bar chart area is proportional to the total lipid content of the cell. The bottom three bar charts represent the fatty acid composition of each lipid fraction (free fatty acids, neutral lipids, and phospholipids), calculated as the % of each chain length on the amount of lipids in that specific lipid class. Displayed is the average and standard deviation of at least three replicates.

Lipid homeostasis is maintained by balancing neutral lipid synthesis and lipid turnover, with free fatty acids stored as biologically inert neutral lipid to avoid potential toxic and membrane-disturbing effects[70]. In *Y. lipolytica* the neutral lipid fraction mainly contains triacylglycerols, and only small amounts of sterol esters[63]. However, the Q4 strain lacks four

genes responsible for triacylglycerol and sterol ester synthesis. In C-lim lipid synthesis is not stimulated and the genotypical difference is not visible in the phenotype. Under N-lim, the high flux through the lipid accumulation pathway highlights the absence of these enzymes in the Q4 strain, preventing free fatty acids from being incorporated into triacylglycerols.

Impact of free fatty acid supplementation on growth

A storage lipid-free Q4 strain of *S. cerevisiae* (Δ*are1*, Δ*are2*, Δ*dga1*, Δ*lro1*) shows high sensibility towards free fatty acids, suggesting the important role triacylglycerols play in free fatty acid buffering and detoxification[71]. Free fatty acid could act as detergents, disrupting membrane integrity, or be incorporated into lipid species that are cytotoxic at high concentrations (ceramide, acylcarnitine, diacylglycerol)[72]. In wild-type strains excess free fatty acids are incorporated into triacylglycerols and stored into lipid droplets to prevent lipotoxicity[70,73].

The Q4 strain of *Y. lipolytica* cannot synthesize lipid droplets[74] (Figure 7), and shows higher levels of free fatty acid under N-lim. We therefore investigated *Y. lipolytica*'s sensitivity to fatty acids by testing the highest concentrations that solubility allowed in our experimental setup, supplementing cultivations with up to 8 mM of unsaturated fatty acids and up to 1 mM of saturated fatty acids.

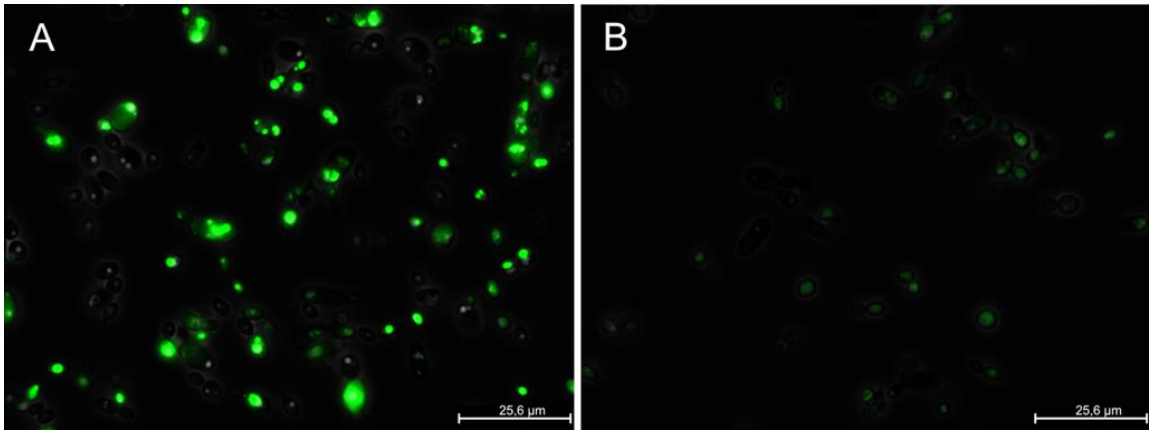


Figure 7: Microscope images of Y. lipolytica strain OKYL029 (A) and Q4 (B) grown under N-lim and stained with Bodipy® Lipid Probe. In the wild-type strain OKYL029 (A) lipid droplets are visible. No visible lipid droplets were observed in the Q4 strain.

The Q4 strain is more sensitive to high concentrations of unsaturated fatty acids, while the wild-type strain was unaffected even by high concentrations (Figure 8). Although the growth of the Q4 strain was affected by free fatty acid, it was able to grow in media supplemented with 8 mM free fatty acid. Notably, the Q4 strain of *Y. lipolytica* is less sensitive to fatty acid supplementation then the Q4 strain of *S. cerevisiae*, where concentrations of 0.5 mM delay or inhibit growth[75].

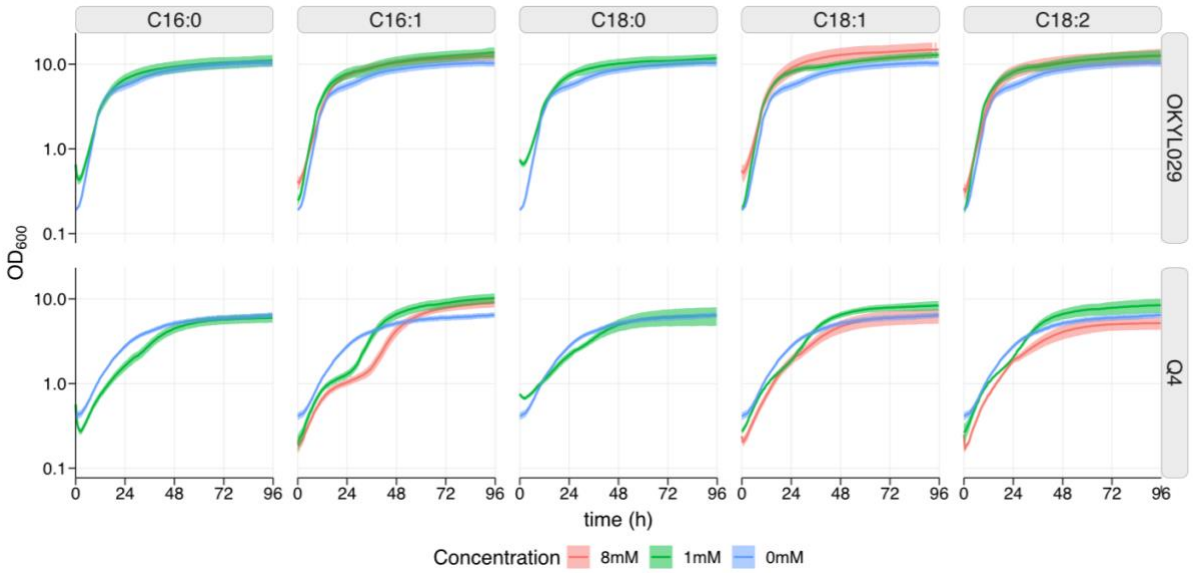


Figure 8: Growth curves of Y. lipolytica strains Q4 and OKYL029 on delft media containing 2% ethanol and 1% tween-20. The media was supplemented with different concentrations of fatty acid. Strains were cultured in 96-well plates and the OD_{600} was measured with the growth profiler every 30 minutes. The lines and shadows represent the average and standard deviation of five replicates.

Integrating a pathway that uses acetyl-CoA as a precursor into the Q4 strain could redirect the acetyl-CoA flow towards the production of other compounds, preventing free fatty acids accumulation. To test this, we performed an SPE analysis on the itaconic acid producer strain JFYL014 (built in Paper III). Our analysis (unpublished data from Manuscript VII) reveales that the free fatty acid fraction decreased from 34% of the total lipid content in JFYL007 to 25% in JFYL014, under the same cultivation conditions (C/N ratio of 116 and dilution rate of 0.1 in chemostat). However, under these cultivation conditions the itaconic acid titer is low compared to fed-batch cultivation. These results suggest that redirecting the acetyl-CoA flux towards other molecules is a feasible strategy. However, the observed reduction in free fatty acids may be more significant under batch or fed-batch cultivation conditions, when higher titres of itaconic acid are produced.

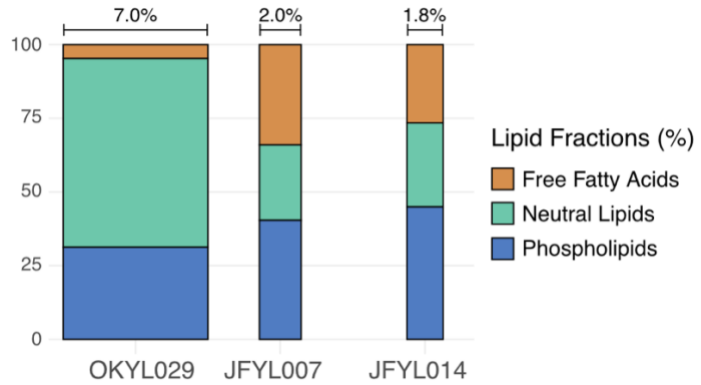


Figure 9: Solid phase extraction (SPE) of OKYL029, JFYL007, and JFYL014 (itaconic acid producer) in N-lim. Stacked bar charts (lipid class distribution) represent the share of each lipid class detected by SPE over the total amount of lipids present in the cell. The bar chart area is proportional to the total lipid content of the cell.

Impact of gene deletions on the transcriptome

To elucidate how the quadruple deletion impacts cell regulation, we performed a transcriptomic analysis (RNA-seq) on the Q4 and OKYL029 strains under carbon and nitrogen limitation.

We explored differences between samples with principal component analysis (PCA) (Figure 10A). Samples in C-lim cluster together, regardless of their genetic background, while samples are separated by genetic background in N-lim, when lipid accumulation is stimulated. These results align with phenotype and lipid measurements, where in C-lim both strains are very similar, while in N-lim the strains show major differences (Figure 5).

We then performed differential gene expression analysis and compared the Q4 strain with the OKYL029 in C-lim and N-lim.

- In C-lim we only detected 30 differentially expressed genes (absolute $log_2FC > 0.5$ and adjusted p-value < 0.05) (Figure 10B) of which only 6 are associated with a function on UniProt. Three of these genes are related to lipid metabolism ("glycerophosphocholine phosphodiesterase", "glycolipid 2-alpha-mannosyltransferase-domain-containing protein", "triacylglycerol lipase",), and might contribute to the small differences we observed in lipid composition between strains in carbon limiting conditions.
- In N-lim, when nitrogen depletion triggers lipid accumulation, we observe major difference between strains. As expected from the PCA, 953 genes are differentially expressed (absolute $log_2FC > 0.5$ and adjusted p-value < 0.05) (Figure 10). Out of the total 953 differentially expressed genes, 390 have a function annotated on UniProt.

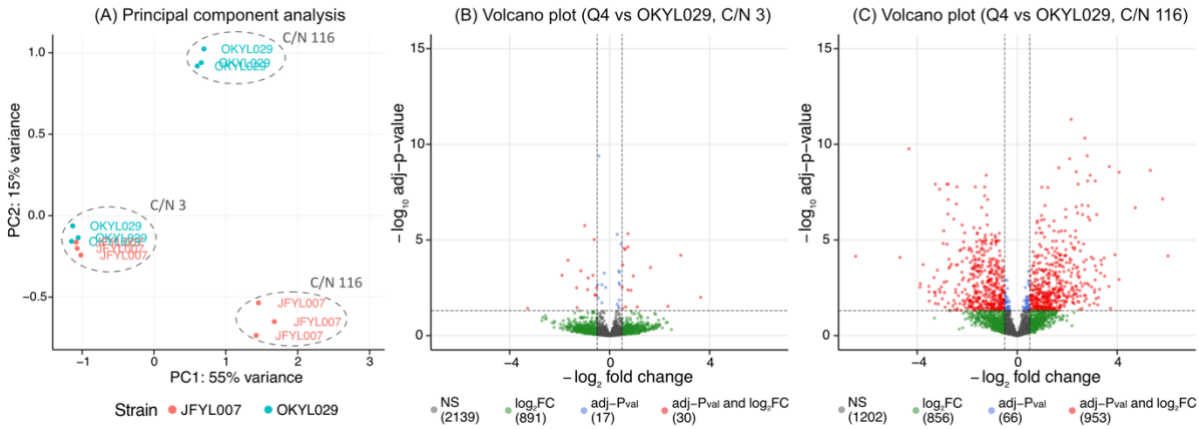


Figure 10: RNA-sequencing of Y. lipolytica strains Q4 and OKYL029 in carbon (C/N ratio 3) and nitrogen (C/N ratio 116) limitation. Panel A: principal component analysis. Volcano plots for samples in carbon (B) and nitrogen limitation (C). NS: non-significative genes. Log_2FC: genes with an absolute fold change greater than 0.5. Adjusted p-value: genes with an adjusted p-value below 0.05. Log_2FC and adjusted p-value: genes with both adjusted p-value below 0.05 and absolute log_2FC greater than 0.5.

To draw biological conclusions from the high number of differentially expressed genes in N-lim conditions, we performed a gene set analysis (GSA). A GSA leverages prior biological knowledge to determine whether a defined gene set shows significant differences between samples[76]. Gene sets can be defined using gene ontology (GO) terms[77] which are generally divided into biological process, molecular function, and cellular component[78]:

- A biological process represents a specific objective that the organism is genetically programmed to achieve and is carried out by specific gene products in a regulated manner.
- A molecular function term describes activities that occur at the molecular levels and are carried out by individual gene products or by molecular complexes composed of multiple gene products.
- Cellular component is the location occupied by a macromolecular machine when it carries out a molecular function.

For each of these levels, we performed a GSA with the R package PIANO[79] (Figure 11). The results suggest that a major alteration in lipid metabolism affects protein synthesis and functionality:

- When the Q4 strain was cultivated under N-lim, we found several GO terms contributing to the unfolded protein response and four GO terms related to chaperones and ubiquitin-dependent activities enriched. Chaperones are proteins that assist the conformational folding of proteins during or after synthesis, and after partial denaturation[80]; ubiquitin-dependent activities are responsible for targeting proteins for degradation[81,82]. The Q4 strain shows enrichment of chaperone and ubiquitin-related related GO terms, indicating that the cells are experiencing folding stress.
- This observation is further supported by the enrichment of Golgi-related GO terms: proteins are glycosylated in the Golgi apparatus before being targeted for delivery to their destination[83]. The genes in the “protein N-linked glycosylation” GO term are mainly downregulated, suggesting that newly synthesized proteins might be misfolded and targeted for degradation before being transported to the Golgi apparatus for glycosylation. The genes of the “Golgi organization” GO term are mainly upregulated, suggesting that a proper Golgi organization might be lacking.
- The Q4 strain lacks the ability to synthesize lipid droplets[74] (Figure 7) and displays alterations in the lipid quantity and distribution, activation of the unfolded protein response, and enrichment of several GO terms related to proteostasis. Cell homeostasis is linked with lipid droplet biology and functionality, as was previously shown in *S. cerevisiae*[84]. Lipid droplets not only act as lipid storage but also prevent lipotoxicity by buffering fatty acid stress[71,85] and have an active role in membrane and

organelle homeostasis[71,84,86]. Lipid droplets are important in starvation-induced autophagy[86,87], clearance of inclusion bodies[88], and, ultimately, in proteostasis[86,88].

Deleting gene involved in lipid metabolism in *Y. lipolytica* results in cells that lack lipid droplets, which are important organelles in cell homeostasis. This results in cells with altered lipid composition and proteome, and with upregulation of the unfolded protein response.

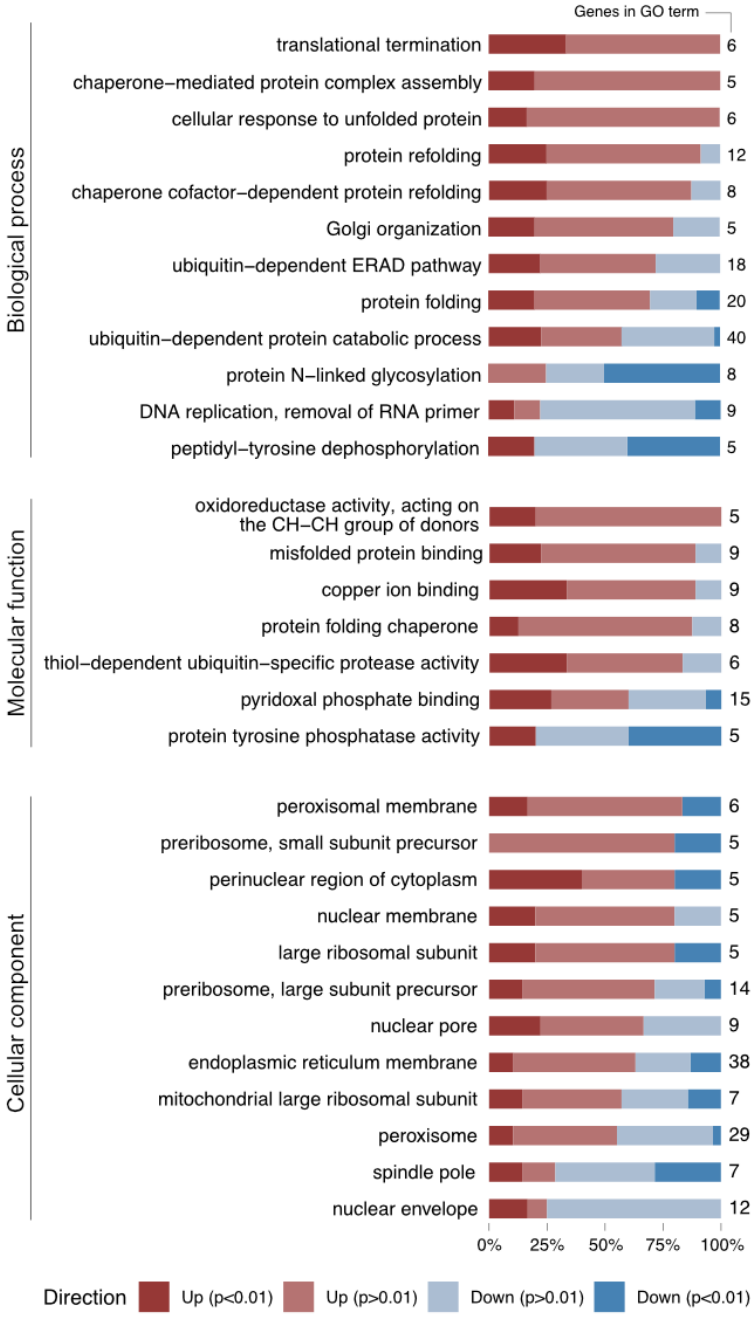


Figure 11: Gene set analysis (GSA) of Q4 vs OKYL029 in N-lim (C/N ratio 116). Gene sets are defined by GO terms (biological process, molecular function, cellular component). For each gene set that is significantly enriched, the direction of the relative changes in RNA levels (positive or negative fold change) is shown, and the genes in the gene sets are marked based on significative or non-significative adjusted p-value (cut-off 0.05). Genes are considered up or down in the Q4 strain, and the OKYL029 strain is the reference strain. The total number of genes in each gene set is reported on the right.

Conclusions and outlook

Deleting lipid genes has proven a valid strategy to boost added value and commodity chemicals production in *Y. lipolytica*.

However, under nitrogen limitation, disrupting lipid synthesis leads an enrichment of several biological processes related to protein refolding and degradation. Cells with disrupted lipid synthesis do not produce lipid droplets, which participate in many biological processes that guarantee cell proteostasis and prevent lipotoxicity. Furthermore, cells with disrupted lipid synthesis show an altered lipid class distribution with an abundance of potentially cytotoxic free fatty acids under.

Based on our findings, we conclude that to optimize our *Y. lipolytica* platform strain, it would be preferable to downregulate rather than delete the genes involved in lipid synthesis. This strategy would aim at maintaining the functional integrity of these genes within the cell, preserving metabolic homeostasis, but without the expression of these genes consuming high quantities of acetyl-CoA, which can be directed towards production of other valuable compounds. This approach would ensure a balanced allocation of cellular resources between molecule production and physiological homeostasis. Although reducing available acetyl-CoA might limit the potential yield of target molecules, it could prevent energy-consuming stress responses and disruptions to cellular physiology. This metabolic strategy could promote a more efficient resource utilization, potentially enhancing cellular robustness.

Chapter 2 – Urea as a nitrogen source

Summary

Media components, including the nitrogen source, are significant cost factors in cultivation processes[89]. While ammonium sulphate is a widely used nitrogen source for cultivating microorganisms, its production requires vast amounts of energy and releases high amounts of greenhouse gases[90]. Urea on the other hand can be a sustainable and cheap alternative if produced from municipal waste[91].

However, the nitrogen source can influence and alter cell behaviour and production[92,93]: a microbial cell factory developed and tested using ammonium sulphate may not behave and produce the same on urea.

To clarify whether to switch from ammonium sulphate to urea for our bioprocesses, we cultivated three phenotypically different strains of *Y. lipolytica*. We investigated the influence of urea as a nitrogen source compared to ammonium sulphate to study how *Y. lipolytica* might behave on urea.

We found no significant coherent changes in growth and lipid production. Transcriptomics revealed no significant coherent changes, and the genes involved in urea uptake and degradation are not up-regulated on a transcriptional level.

Our findings support urea usage, indicating that previous metabolic engineering efforts are likely translatable and can ease the way for urea as a cheap and sustainable nitrogen source in more applications, as we will also show in Chapter 3.

Introduction

Why urea as nitrogen source?

The most used nitrogen source in microbial cultivation is ammonium sulphate, produced by sulfuric acid treatment of ammonia. Ammonia is mostly produced via the energy and carbon-intense Haber-Bosch process that fixes atmospheric nitrogen with hydrogen at high temperature (400-500°C) and pressure (>100 bar)[90]. Ammonia production, combined with the energy needed to produce hydrogen and purified atmospheric nitrogen, accounts for 1% to 2% of the global energy consumption[94] and 3% to 5% of natural gas consumption[95].

Urea is currently produced through the energy intense Bazarov reaction which combines ammonia with carbon dioxide at high temperatures (170-220°C) and pressures (125-250 bar)[96]. However, urea can be an interesting alternative nitrogen source for microbial cultivation

since it can be extracted from municipal waste in an economical and environmentally friendly way[91], allowing for waste valorisation and bioprocess cost reduction. Additionally, unlike ammonium sulphate, urea consumption does not acidify the media, thus requiring less base addition during fermentation (Results and Discussion, Figure 13).

Ammonium and urea assimilation

The pathway for ammonium and urea utilization are well characterised in *S. cerevisiae*. Since *Y. lipolytica* contains homologous genes, we reconstructed the nitrogen and urea assimilation pathway in *Y. lipolytica* through homology (Figure 12).

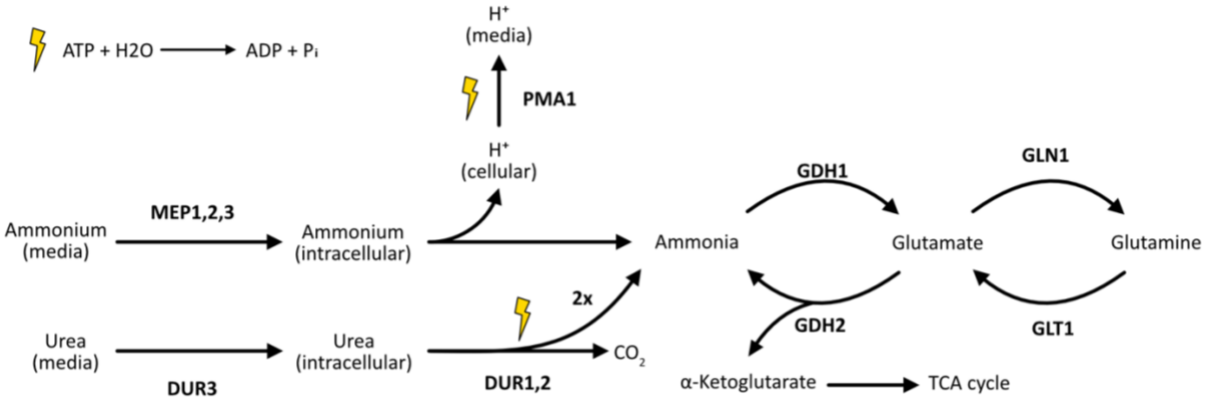


Figure 12: Schematic overview of ammonium and urea utilization in yeast. Arrows represent reactions, and gene names follow S. cerevisiae nomenclature.

Ammonium is transported into the cell by *MEP1,2,3*. Intracellular ammonium then dissociates into ammonia, releasing a proton. The proton is then transported into the media by the plasma membrane H^+-ATPase (*PMA1*), which consumes one ATP per proton and is responsible for the media acidification[97].

Urea is transported into the cell by *DUR3* and converted into two ammonia molecules by a urea amidolyase (Dur1_2). *DUR1_2* is a multifunctional enzyme with urea carboxylase and allophanate hydrolase activity. The first activity converts urea into allophanate by consuming one ATP and one bicarbonate. The second activity converts allophanate into two ammonia molecules by consuming water and releasing CO_2[98]. Urea usage is more energy efficient than ammonium usage since it consumes one ATP to yield two ammonia molecules.

The two pathways converge on ammonia, which can be incorporated into glutamate by the NADP-dependent glutamate dehydrogenase (*GDH1*) and into glutamine by the glutamine synthetase (*GLN1*). Glutamate and glutamine are both starting points for amino acid synthesis. Additionally, glutamate can be converted by the NAD-dependent glutamate dehydrogenase (*GDH2*) to ammonia and α-ketoglutarate, linking nitrogen metabolism to the tricarboxylic acid cycle[99].

Goal of the project

Urea can be a more sustainable nitrogen source than ammonium sulphate if produced from municipal waste[91].

However, changing the media composition can affect cell behaviour and impact the production performance of microorganisms[92,93,100,101]. A microbial cell factory developed and tested using ammonium sulphate may behave differently when using urea, potentially producing lower amounts of the desired product.

In *Y. lipolytica*, lipid accumulation is triggered by nitrogen limitation[62], and the fatty acid composition is crucial when producing lipid derivatives. It is important that changing nitrogen source does not alter the fatty acid profile or interfere with lipid accumulation.

Additionally, in *Y. lipolytica*, the nitrogen source was linked to dimorphic growth[102], which impacts bioreactor cultivations not only by causing line clogging, but also by altering cell physiology, leading to reduced product yield, titre, and rate.

The goal of this study is to determine whether switching from ammonium sulphate to urea would negatively impact future bioprocesses. To address this, we investigated the influence of urea compared to ammonium sulphate on the physiology and transcriptome of *Y. lipolytica*.

Experimental setup

To understand the cells' reaction to different nitrogen sources, it is not sufficient to observe single parameters such as growth rate, lipid content, and metabolite production. Instead, it is necessary to monitor the whole cell system and the interactions between the modified environment and the altered cellular response. One comprehensive approach for studying cell behaviour on a genome-wide level is transcriptomic analysis.

By comparing gene expression under different conditions, we can measure changes in cell behaviour, and gain deeper insights into how cells respond to various nitrogen sources.

In this study, we performed transcriptomic analysis of *Y. lipolytica* cultivations to investigate whether a response to the nitrogen source might differ depending on the amount of lipid accumulation, either as a function of strain genotype or nutrient limitation.

To address this, we varied three parameters:

1. *Y. lipolytica* strains, differing in their lipid accumulation ability:
    i. OKYL029: normal lipid accumulation
    ii. JFYL007 or Q4: lipid synthesis disrupted (Δ*dga1*, Δ*dga2*, Δ*lro1*, Δ*are1*)
    iii. OKYL049: lipid overproducer (*DGA1* overexpression, Δ*are1*)

2. Nutrient limitation:
    i. carbon limitation (C-lim): C/N ratio 3, as established in Chapter 1
    ii. nitrogen limitation (N-lim): C/N ratio 116, as established in Chapter 1
3. Nitrogen source:
    i. ammonium sulphate
    ii. urea

Results and discussion

Cell physiology

We performed chemostat cultivations under C-lim or N-lim with urea or ammonium sulphate as equimolar nitrogen sources. We maintained the pH at 5 by automated addition of potassium hydroxide. Urea required significantly less base addition than ammonium sulphate (Figure 13), which can reduce the costs of a bioprocess since it does not dilute the final product, as we will see in Chapter 3.

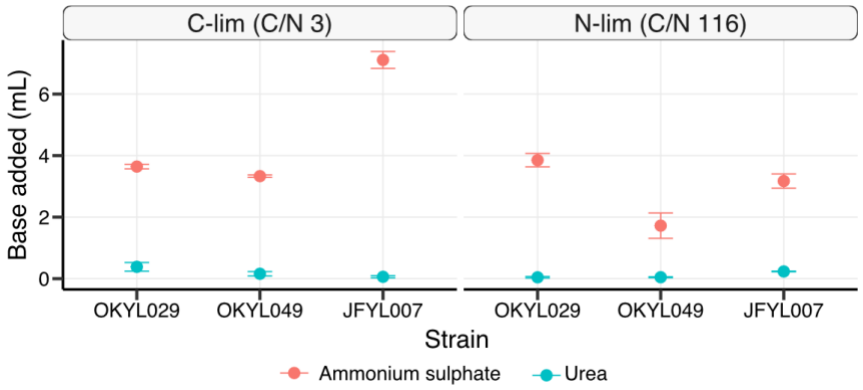


Figure 13: Addition of base to a steady-state cultivation (working volume 500 mL). Volume (mL) of base added per residence time (dilution rate 0.1, 10 hours residence time) in C/N ratio 3 and 116, normalized to cell dry weight. Dots and error bars represent the average and standard deviation of the replicates.

Since nitrogen assimilation from ammonium sulphate costs 1 ATP per ammonia, while from urea ½ ATP per ammonia (Figure 12), we hypothesized that this might impact the biomass, lipid content, their corresponding yields, or on the specific uptake rate of glucose (r-Glucose).

However, the cell physiology was largely unaffected by the nitrogen source (Table 1). Only the lipid overproducer OKYL049 showed significant changes in biomass in N-lim (p-value < 0.01), albeit in the opposite direction as anticipated. The other strains showed no statistically significant changes. We concluded that the nitrogen source does not significantly impact overall cell physiology.

*Table 1: Physiological parameters of the strains in different C/N ratios and nitrogen sources (N-source). Displayed is the mean ± standard deviation of at least three replicates. Significance was calculated between the two nitrogen sources ammonium sulphate (AS) and urea (U), with a two-tailed homoscedastic t-test. ** indicates a p-value < 0.01.*

		OKYL029		OKYL049			JFYL007	
	N-source	C/N 3	C/N 116	C/N 3	C/N 116		C/N 3	C/N 116
Biomass	AS	3.5 ± 0.2	2.0 ± 0.3	3.6 ± 0.1	2.2 ± 0.1	**	3.5 ± 0.1	1.3 ± 0.1
(g/L)	U	3.7 ± 0.3	1.7 ± 0.1	3.5 ± 0.1	1.8 ± 0.2		3.8 ± 0.0	1.4 ± 0.4
Lipid content	AS	2.7 ± 0.5	8.5 ± 0.6	2.5 ± 0.1	12.1 ± 1.6		2.4 ± 0.4	5.9 ± 0.9
(%)	U	2.6 ± 0.3	9.8 ± 1.7	2.8 ± 1.0	10.2 ± 1.3		2.2 ± 0.1	7.8 ± 1.5
Biomass yield	AS	0.46 ± 0.03	0.48 ± 0.06	0.48 ± 0.01	0.36 ± 0.01		0.47 ± 0.02	0.44 ± 0.04
(gCDW/gGlucose)	U	0.50 ± 0.03	0.42 ± 0.02	0.47 ± 0.01	0.42 ± 0.04		0.51 ± 0.01	0.36 ± 0.10
Lipid yield	AS	1.2 ± 0.3	4.1 ± 0.5	1.2 ± 0.1	4.3 ± 0.6		1.1 ± 0.2	2.6 ± 0.7
(gLipid/gGlucose)	U	1.3 ± 0.2	4.2 ± 0.8	1.3 ± 0.5	4.3 ± 0.9		1.1 ± 0.0	2.9 ± 1.1
r-Glucose	AS	0.23 ± 0.01	0.22 ± 0.02	0.22 ± 0.01	0.29 ± 0.01		0.22 ± 0.01	0.24 ± 0.02
(g/gCDW h)	U	0.21 ± 0.01	0.25 ± 0.02	0.23 ± 0.01	0.26 ± 0.02		0.20 ± 0.01	0.31 ± 0.10

Since the fatty acid composition is of interest when *Y. lipolytica* is applied to produce lipid derivatives, we performed a FAME analysis to investigate whether the nitrogen source affects the fatty acid composition (Table 2).

*Table 2: Changes in the fatty acid composition (% of total fatty acid) in different C/N ratios and nitrogen sources (N-source) in strains OKYL029, OKYL049, and JFYL007 (Q4). Displayed is the mean ± standard deviation of at least three replicates. Significance was calculated between the two nitrogen sources ammonium sulphate (AS) and urea (U), with a two-tailed homoscedastic t-test. ** indicates a p-value < 0.01.*

		OKYL029			OKYL049			JFYL007		
	N-source	C/N 3	C/N 116		C/N 3	C/N 116		C/N 3	C/N 116	
C16:0	AS	9.5 ± 0.5	15.5 ± 1.5		9.0 ± 0.1	16.7 ± 0.2	**	8.4 ± 0.3	17.7 ± 1.1	
(%)	U	8.8 ± 0.2	14.3 ± 0.5		10.3 ± 2.7	13.6 ± 0.6		8.3 ± 0.2	18.0 ± 0.4	
C16:1	AS	7.8 ± 0.2	7.2 ± 0.2	**	7.7 ± 0.7	5.8 ± 0.4		10.9 ± 0.5	9.8 ± 0.8	**
(%)	U	9.2 ± 0.2	7.0 ± 0.6		8.6 ± 0.7	5.5 ± 0.2		12.4 ± 0.2	10.5 ± 0.2	
C18:0	AS	1.1 ± 0.2	6.0 ± 0.7		3.8 ± 0.1	10.4 ± 0.7		0.3 ± 0.1	1.7 ± 0.1	**
(%)	U	0.8 ± 0.1	4.7 ± 0.3		3.4 ± 0.4	9.6 ± 0.3		0.2 ± 0.0	2.2 ± 0.1	
C18:1	AS	49.9 ± 0.3	47.9 ± 1.4	**	54.0 ± 1.9	54.1 ± 0.7		43.3 ± 1.0	17.9 ± 0.5	**
(%)	U	47.7 ± 1.2	49.6 ± 1.4		52.9 ± 2.3	54.4 ± 0.2		40.6 ± 0.3	16.1 ± 0.3	
C18:2	AS	31.7 ± 0.7	23.4 ± 3.3		25.5 ± 1.4	13.1 ± 1.2	**	37.1 ± 0.8	52.9 ± 0.1	
(%)	U	33.5 ± 1.2	24.5 ± 1.0		24.8 ± 3.8	16.9 ± 1.1		38.5 ± 0.3	53.2 ± 0.6	
Saturated/Unsaturated	AS	0.12 ± 0.01	0.27 ± 0.04		0.15 ± 0.00	0.37 ± 0.02	**	0.10 ± 0.00	0.24 ± 0.02	
	U	0.11 ± 0.00	0.23 ± 0.00		0.16 ± 0.03	0.30 ± 0.02		0.09 ± 0.00	0.25 ± 0.01	
C16/C18	AS	0.21 ± 0.01	0.29 ± 0.02		0.20 ± 0.01	0.29 ± 0.00	**	0.24 ± 0.00	0.38 ± 0.01	**
	U	0.22 ± 0.00	0.27 ± 0.02		0.23 ± 0.03	0.24 ± 0.01		0.26 ± 0.00	0.40 ± 0.01	

Storage lipid production is not triggered under C-lim, and most of the extracted fatty acids are expected to originate from phospholipids. Under C-lim, we observed a significant change between urea and ammonium sulphate in C16:1 and C18:1 for OKYL029 and Q4 (JFYL007). We did not observe any significant changes under carbon limitation in OKYL049. The minor changes that could be observed, were most likely derived from changes in membrane fatty acids, which become visible when the contribution of the storage lipids to the lipid content is low.

Under N-lim, OKYL049 showed significant changes in C16:0 and C18:2 and we observed a change towards lower saturation and longer chain length (C16/C18) in urea compared to ammonium sulphate. The Q4 strain (JFYL007) showed significant changes in the C18:0 and C18:1 but no significant changes in saturation or chain length of the fatty acids.

Overall urea does not seem to have a major impact on any of the measured parameters. However, the changes in fatty acid composition in the Q4 strain indicate that there might be some changes not captured with the measured parameters. Additionally, there might be changes in metabolites we did not measure. Therefore, we performed transcriptomics to analyse how gene expression is affected by urea.

Transcriptomics analysis

To probe whether any transcriptional changes occurred that might influence the phenotype beyond the parameters we measured, we clustered the RNA-sequencing result by samples through principal component analysis (PCA, Figure 14).

We found that the nitrogen source only resulted in minor separation across the samples (Figure 14A). Meanwhile, as the C/N ratio affected cell physiology (e.g. lipid content, Table 1), it also significantly separated the samples in the PCA (Figure 14B). Cell physiology was also affected by the strain genotype in N-lim, with the JFYL007 (Q4) strain clustering further. In C-lim, the strains showed low variance, indicating that the nitrogen source has little effect when available in copious amounts.

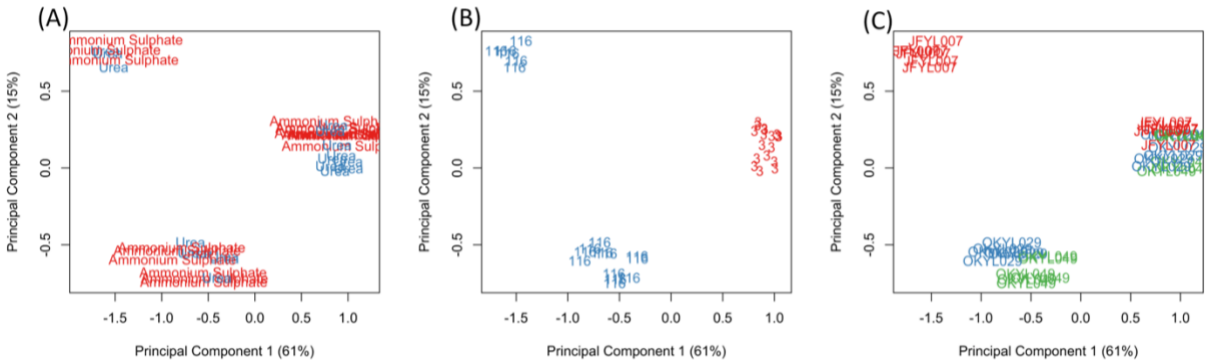


*Figure 14: Principal component analysis (PCA) plot of RNA-sequencing samples. The panels display the same PCA result, but samples are labelled based on either **(A)** nitrogen source, **(B)** C/N ratio or **(C)** strain.*

We then identified the genes that are differentially expressed as an effect of the different nitrogen sources. The expression of each gene was compared between a sample cultivated in urea versus a sample cultivated in ammonium sulphate, while strain and nutrient limitation were kept constant. The results of each comparison were filtered to retain differentially

expressed genes with an adjusted p-value below 0.5 and a $\log_2$FC above 1. We then visualized the genes in a network plot and identified the most interesting clusters (Figure 15).

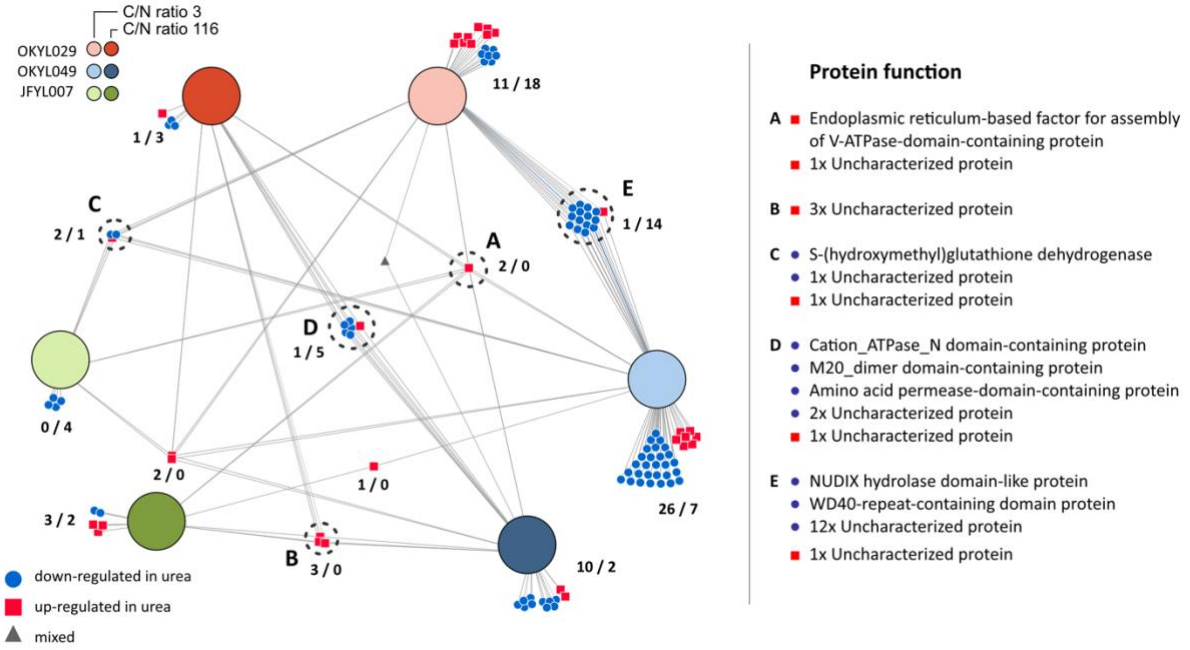


Figure 15: Overlap of differentially expressed genes in different conditions. Displayed are the differentially expressed genes (adjusted p-value < 0.05, absolute fold change > 1) of each strain and C/N ratio in urea compared to ammonium sulphate. Numbers indicate the number of down/up-regulated genes. Clusters A to E contain groups of genes discussed in the results. The UniProt protein function of the corresponding genes is listed on the right-hand side. For visualization, the DiVenn web tool was used[103].

- Cluster A. A coherent response to the nitrogen source, irrespective of lipid phenotype, is indicated by genes that are differentially expressed in all strains and under all nutrient limitations. However, only two genes are in this cluster: one encodes a protein of unknown function, the other a protein with similarity to *S. cerevisiae*'s *VPH2*. *VPH2* is essential for the vacuolar-type ATPase assembly, and its differential expression can be associated with the acidification caused by ammonium sulphate, but not urea[104].
- Clusters B and C. Since the C/N ratio had a significant impact on gene expression (Figure 14B), we checked clusters between all three strains under either C-lim or N-lim. Cluster B contained three uncharacterized proteins. Cluster C contained two uncharacterized proteins and a S-(hydroxymethyl)glutathione dehydrogenase for which we could not find a link to the nitrogen source.
- Clusters D and E. Since JFYL007 (Q4) showed a different behaviour than the other two strains (Figure 14C), we checked, for both C/N ratios, the clusters between strains whose behaviour was similar (OKYL029 and OKYL049). 16 of the 21 genes of clusters D and E were uncharacterized proteins, and for the remaining proteins we could not find a link to the nitrogen source.

The low variance identified in the PCA, the low number of differentially expressed genes, and the low overlap between differentially expressed genes indicate that the nitrogen source (urea or ammonium) has minimal effect on the overall transcriptome. Additionally, the number of uncharacterized proteins highlight one of the limitations of working with non-conventional yeast.

Urea assimilation pathway

Since the overall transcriptional changes between the two nitrogen sources is marginal, we investigated the expression of the genes of the urea and ammonium pathway (Figure 16).

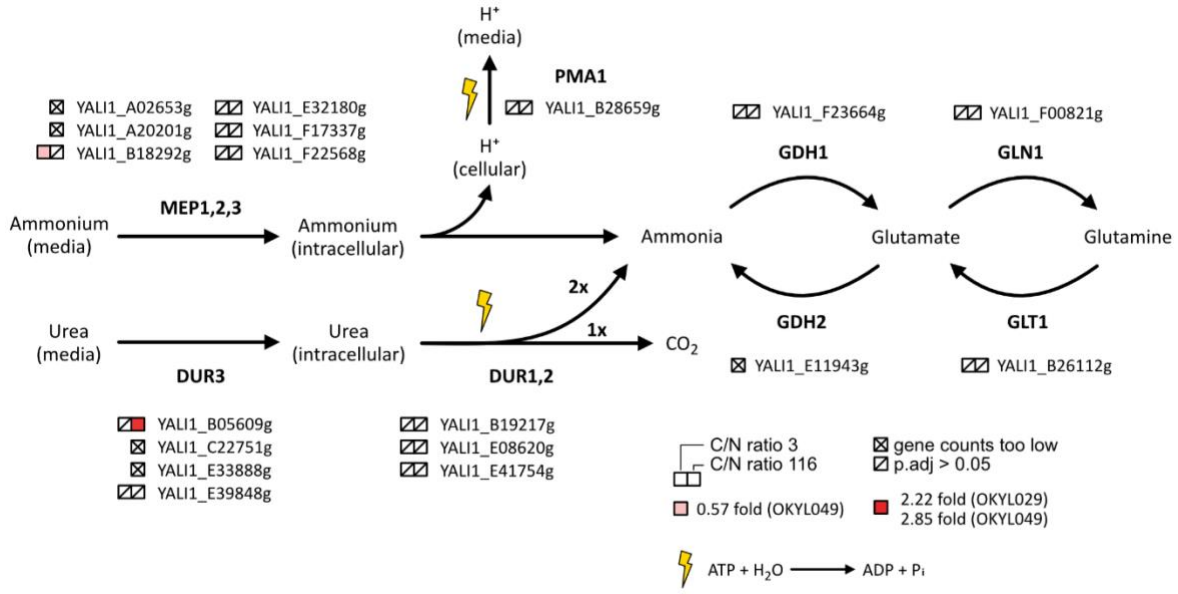


Figure 16: Potential homologous Y. lipolytica genes, and their expression changes when comparing the use of ammonium sulphate versus urea. The genes were identified from different sources, as listed in Table S1 of Paper II. Genes marked with ☒ have been removed during the filtering of the gene counts; genes marked with ⧄ did not show any significant change between the nitrogen sources (adjusted p-value < 0.05).

Five genes of the pathway were filtered out due to low reads when processing raw RNA-sequencing data (marked with ☒ in Figure 16). Eleven were not significantly different (adjusted p-value > 0.05, marked with ⧄ in Figure 16). Two genes showed significant expression changes between the two nitrogen sources, either in carbon or nitrogen limitation, for one or more strains: an ammonium transporter (YALI1_B18292g) and a urea transporter (YALI1_B05609g).

These results were unexpected since gene expression of both ammonium and urea pathways is regulated by the available nitrogen source in *S. cerevisiae* and *C. albicans*[105,106]. However, we only observed an upregulation of one of the four *DUR3* homologs. This suggests that YALI1_B05609g is the true *ScDUR3* homolog and has a similar regulation. However, this

hypothesis would require further studies for confirmation, for instance through the generation of knockout strains.

The genes downstream of ammonia were not expected to be differently expressed since both ammonium and urea metabolism end in ammonia.

Conclusions and outlook

Switching nitrogen source to urea can decrease the cost of the bioprocess and increase its sustainability if urea is extracted from waste. The goal of this study was to investigate whether switching from ammonium sulphate to urea would negatively impact our bioprocess.

In our study we found no significant coherent changes in growth or lipid production, and RNA-sequencing revealed no significant coherent changes in the transcriptome. The genes involved in urea uptake and degradation were also not up-regulated on a transcriptional level. Additionally, urea reduces media acidification and base consumption, making the process cheaper and more sustainable, especially if urea is sourced from waste materials.

Our findings support urea usage, indicating that previous metabolic engineering efforts are likely translatable to urea. Although there are some minor non-coherent changes, switching from ammonium sulphate to urea is feasible, and can bring several advantages such as reduced fermentation cost, increased bioprocess sustainability, and lower base consumption.

We will build on these advantages in Chapter 3, showing how using urea as a nitrogen source in a fed-batch cultivation decreases base consumption, resulting in lower bioreactor volume and higher titres.

Chapter 3 – Itaconic acid production

Preface

The project to produce itaconic acid in Y*. lipolytica* began before I started my PhD. Over the years, Jing Fu and collaborators developed a platform strain for non-lipid production (Q4, discussed in paper I), and subsequently engineered it for itaconic acid production.

My contribution focused on lab-scale fed-batch cultivation in 1L bioreactors. I will primarily discuss this part in the results, while briefly outlining previous work in the introduction.

Summary

Itaconic acid ranks among the top 12 building block chemicals[107] and has several applications in food, textile, and pharmaceutical industries[108].

Currently, the most promising microorganisms for itaconic acid production are *Aspergillus terreus* and *Ustilago maydis*[109,110]. However, they are pathogens and require costly bioprocess setups. *Y. lipolytica*, on the other hand, can be an economic alternative, as it is generally regarded as safe and genetic tools are available.

After previous metabolic engineering efforts, we obtained a strain with high itaconic acid production in shake flask.

This promising strain was cultivated in fed-batch bioreactors to increase production. Through trial and error, we identified the key factors to improve production: addition of yeast extract, continuous feeding, pH of 5.5, and urea as nitrogen source.

By tweaking these parameters, we reached an itaconic acid titre of 130 g/L in fed-batch cultivations, a significant leap towards establishing *Y. lipolytica* for competitive itaconic acid production.

Introduction

Why producing itaconic acid in *Y. lipolytica*?

Itaconic acid ranks among the top 12 building block chemicals[107] and has several applications in food, textile, and pharmaceutical industries[108]. Itaconic acid is a platform chemical and can be transformed into various valuable bio-based products with remarkable properties[111,112], e.g. shape memory polymers[113], polymeric hydrogels for targeted drug delivery[114,115], and anti-bacterial materials[116,117].

Currently, the primary microorganism used for itaconic acid production is the filamentous fungus *Aspergillus terreus*, capable of producing 160 g/L[109]. However, *A. terreus* is pathogenic[118], needs careful monitoring of fermentation parameters to prevent morphological switch[119], and production is inhibited by low concentrations of manganese ions[109,120]. These issues increase operational costs and the risk of failed batches[110].

Another microorganism, *Ustilago maydis*, can produce 220 g/L of itaconic acid as solid calcium salt[110]. However, it is a corn pathogen[121], its fermentation requires manual addition of calcium carbonate as suspension or powder, and in-situ precipitation of calcium itaconate to prevent product inhibition. These factors complicate scale-up efforts, making it costly.

On the other hand, *Y. lipolytica* is generally recognized as safe[122], has available genetic manipulation tools, and exhibits a high flux towards acetyl-CoA, citric acid and isocitric acid, which are itaconic acid precursors. Additionally, the formation of pseudo-hyphae in *Y. lipolytica* can be abolished by deleting *mhy1*[123], making morphological control easier than *A. terreus*.

How to produce itaconic acid in *Y. lipolytica*

The metabolic engineering work and the fermentation improvement that were performed to produce high titres of itaconic acid in *Y. lipolytica* can be divided into 4 steps (Figure 17):

1. In Step 1, cis-aconitate supply was enhanced by:
    a. removing carbon flux from lipid storage and sterol ester formation by deleting *DGA1*, *DGA2, LRO1*, and *ARE1* (Q4 or JFYL007 strain, discussed in Chapter 1)
    b. interrupting the glyoxylate cycle (deletion of *ICL1/2*) and deleting the isocitrate dehydrogenase (IDP).
2. In Step 2, the itaconic acid biosynthetic pathway was introduced and optimized.
    a. Itaconic acid production was tested in mitochondria and in the cytosol. Initially, the mitochondrial route showed higher titres. However, when a tricarboxylic acid transporter (MTT) was expressed to export cis-aconitate to the cytosol for subsequent transformation to itaconic acid, the titre surpassed that of mitochondrial production. Consequently, the strain with cytosolic production was selected for further engineering.
    b. Combining cytosolic production with a strain lacking *ICL1/2*, resulted in decreased production, while only disrupting lipid synthesis improved titre. A strain with cytosolic production, MTT transporter, and disrupted lipid synthesis was retained for further engineering.
    c. Itaconic acid production through the trans-aconitate pathway was also tested, but yields were lower than the cis-aconitate pathway.

d. Various promoters and copy numbers of MTT and CAD (cis-aconitate decarboxylase) from *A. terreus* were tested. The combination of six copies of pTef-AtCAD and one copy of pGPD-AtMTT was retained for further engineering.

3. In Step 3, carbon distribution between cell growth and itaconic acid production was optimized. After respectively 4 and 8 days of cultivation, itaconic acid titres were similar in both nitrogen replete (NR), in which nitrogen is abundant, and nitrogen limiting (NL) conditions. However, while NL resulted in higher yield, productivity was lower. In NL, the activity of adenosine monophosphate (AMP) deaminase (AMPD) increases, leading to a decrease in AMP levels. Since the isocitrate dehydrogenase IDH is inhibited by low AMP levels, low IDH activity was mimicked to enhance production:
   a. The overexpression of native AMP deaminase did not increase production.
   b. Downregulating IDH using weaker promoters increased itaconic acid titre, yield, and productivity in NR conditions. The strain JFYL122 achieved similar titres and yields in both NR and NL conditions, but in NR it only took 4 days compared to 8 days in NL, resulting in higher productivity.
4. In Step 4, we optimised fermentation parameters for fed-batch bioreactor cultivations of strain JFYL122, as we will see in the Results and discussion section.

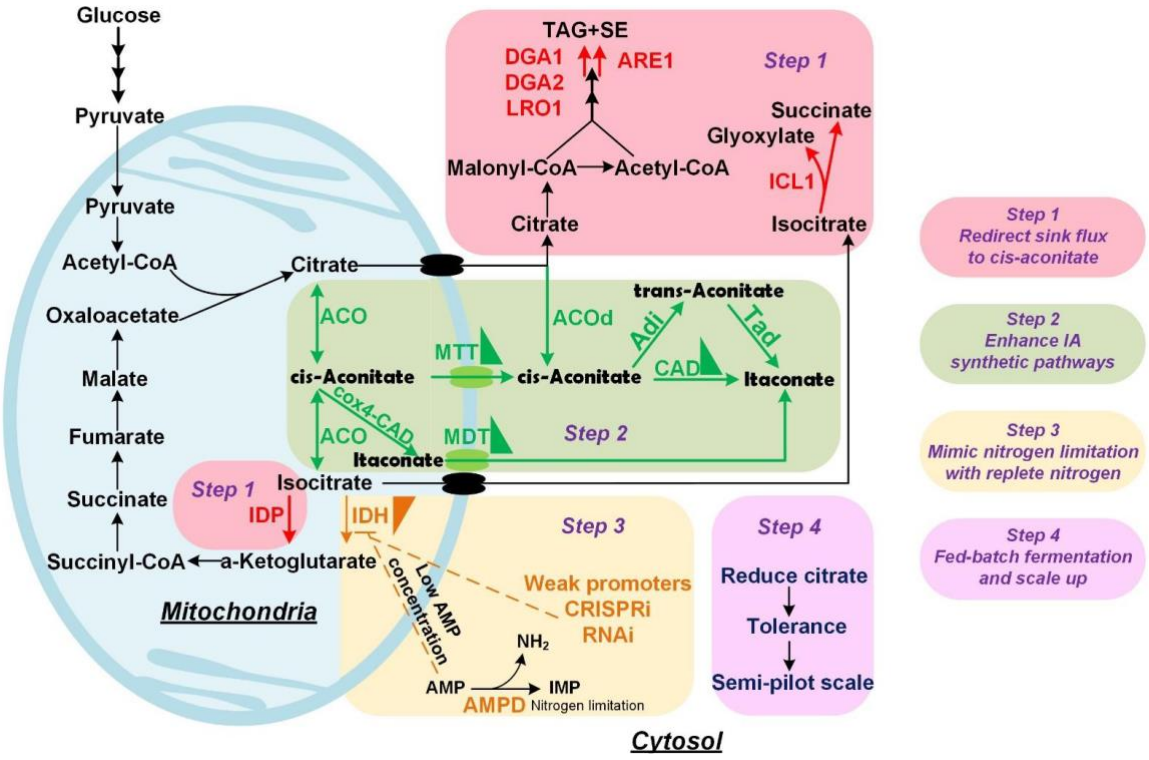


Figure 17: Overview of the metabolic engineering efforts to produce itaconic acid. ACO, aconitase; ACOd, aconitase without mitochondrial leading sequence; ADI, aconitate isomerase; AMPD, AMP deaminase; ARE1, Acyl-CoA:sterol O-acyltransferase; CAD, cis-aconitate decarboxylase; DGA1, Acyl-CoA diacylglycerol O-acyltransferase 1; DGA2, Acyl-CoA diacylglycerol O-acyltransferase 2; ICL1, isocitrate lyase; IDH, NAD+ dependent isocitrate dehydrogenase; IDP, NADP+ dependent isocitrate dehydrogenase; LRO1, phospholipid:diacylglycerol acyltransferase; MDT, mitochondrial decarboxlic transporters; MTT, mitochondrial tricarboxlic transporters; TAD, trans-aconitate decarboxylase.

Experimental setup

JFYL122 is the most promising itaconic acid producer: its lipid synthesis is disrupted, it expresses 6 copies of *AtCAD*, one copy of *AtMTT*, and its IDH expression is modulated by a weak promoter. JFYL122 produces the highest itaconic acid titre (4.3 g/L) with a yield of 0.31 mol/mol (itaconic acid/glucose) in shake flask.

We cultivated this strain in 1L fed-batch bioreactors. Fed-batch cultivations allow to control environmental conditions such as pH, nutrient concentration, dissolved oxygen, etc. Fed-batch cultivations also allow the feed of media whenever the cultivation reaches a desired stage, for example when certain nutrients become limiting. Normally the limiting nutrient is fed as a high concentration solution to keep a low flow of liquid into the reactor, resulting in a low dilution rate and a low volume increase. A fed-batch cultivation extends a culture's productive duration and enables formation of high product titers, which is important for decreasing the cost of downstream operations. However, inhibitory or toxic by-products may accumulate and compromise cell viability and productivity.

Results and discussion

The first tested strategy (NR -> NL) consists in starting the fermentation with NR condition, and having NL conditions generated whenever the nitrogen source is consumed during biomass formation. With this setup, significant amounts of the by-product citric acid started to accumulate after three days (Figure 18) and surpassed the itaconic acid titre.

Citric acid accumulation is probably due to complex regulation mechanisms between absolute nitrogen amount, C/N ratio, and reaching low glucose concentrations. Glucose was fed during the cultivation but reached low concentrations multiple times during the fermentation. Nitrogen likely became limiting at day 3, when growth halted at OD_{600} 120, and citric acid started to be secreted, indicating that it was not being converted to itaconic acid.

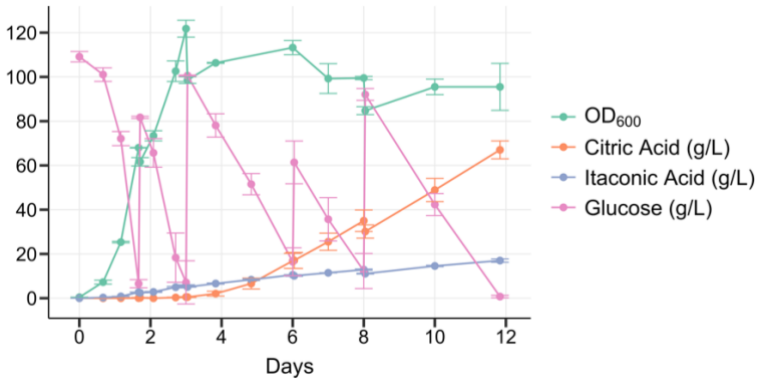


Figure 18: Fed-batch cultivation of JFYL122. Media composition: 10 g/L $(NH_4)_2SO_4$, 3 g/L KH_2PO_4, 0.5 g/L $MgSO_4 \bullet 7H_2O$, 2 mL trace metals solution stock, and 1 mL of vitamin solution stock. 100 g/L initial glucose, pH 3.5. 650 g/L glucose was fed when residual glucose is below 20 g/L.

To check if either nitrogen or other nutrient limitation was responsible for citric acid secretion, we tested phosphate limitation and sulphur limitation (Figure 19). Citric acid was not detected under phosphate limitation and sulphur limitation, but it was detected under nitrogen limitation and NL->NR conditions, indicating that nitrogen limitation should be avoided to prevent citric acid secretion.

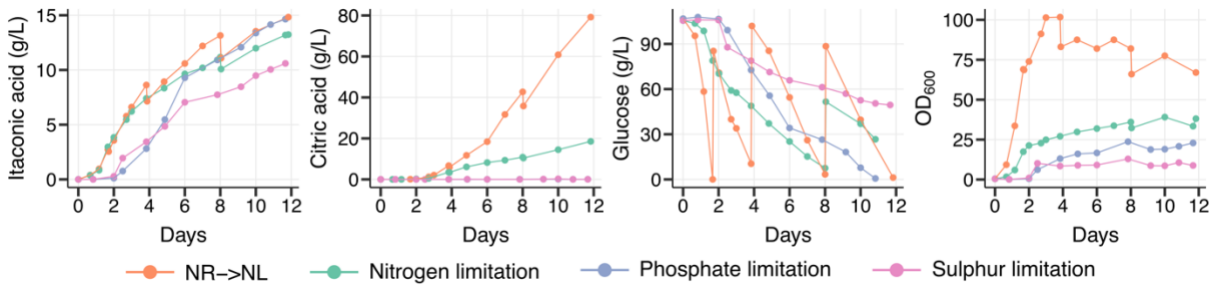


Figure 19: Testing different nutrient limitations in JFYL122. Media composition for NR -> NL: 10 g/L $(NH_4)_2SO_4$, 3 g/L KH_2PO_4, 0.5 g/L $MgSO_4 \bullet 7H_2O$, 2 mL trace metals solution stock, and 1 mL of vitamin solution stock. 100 g/L initial glucose, pH 3.5. Nitrogen limitation: same composition as NR -> NL, but 2.5 g/L of $(NH_4)_2SO_4$. Phosphate limitation: same composition as NR -> NL, but 0.2 g/L KH_2PO_4. Sulphur limitation: same composition as NR -> NL, but 0.1 g/L $MgSO_4 \bullet 7H_2O$ and 0.17 g/L $MgCl_2$. 650 g/L glucose was fed when residual glucose is below 20 g/L.

To avoid nitrogen limitation, we started feeding nitrogen together with the carbon source. Additionally, to prevent citric acid production, we tried not to reach very low glucose concentrations in the bioreactor. However, this could prove challenging since we cannot measure live glucose concentration and are unable to monitor the fermentation continuously.

To increase itaconic acid production and prevent citric acid formation, we tested the effect of feeding yeast extract as a nitrogen source together with the media feed. Yeast extract has low production cost and is frequently used to improve cell growth and productivity[124]. After starting a cultivation with the NR -> NL media, we tested different yeast extract concentrations in the media feed. Feeding with 2.5 g/L yeast extract significantly increased the itaconic acid titre to 17.3 g/L (Figure 20) within 3 days. When feeding with less yeast extract, citric acid accumulated, indicating nitrogen source depletion. Furthermore, with 2.5 g/L yeast extract the glucose concentration did not get close to zero, also preventing citric acid secretion.

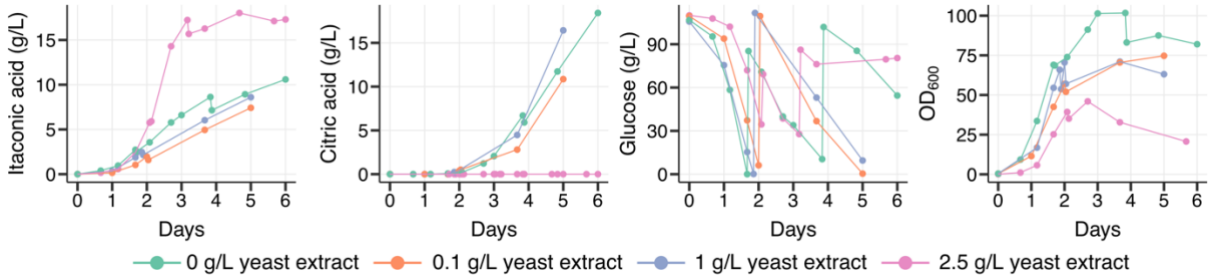


Figure 20: Testing yeast extract effect. Media composition (NR -> NL): 10 g/L $(NH_4)_2SO_4$, 3 g/L KH_2PO_4, 0.5 g/L $MgSO_4 \bullet 7H_2O$, 2 mL trace metals solution stock, and 1 mL of vitamin solution stock. 100 g/L initial glucose, pH 3.5. 650 g/L glucose was fed when residual glucose is below 20 g/L, together with 2.5, 1, 0.1, and 0 g/L yeast extract.

Since high osmotic pressure can inhibit yeast growth and decrease fermentation performance[125], we tried to change feeding mode from pulses to continuous feed. This increased itaconic acid production to 24.5 g/L on the sixth day (Figure 21). However, cells had arrested glucose consumption and itaconic acid production at that point. In high-throughput microbioreactors (BioLector) an initial pH of 3.5 showed that 20 g/L of itaconic acid interrupts cell growth, while 60 g/L are tolerated with at pH 7[126]. We then decided to increase the pH of our cultivations from 3.5 to 5.5. Growth, glucose consumption, and itaconic acid production resumed (Figure 21), reaching 29.3 g/L of itaconic acid. We therefore decided for the next round of fed-batch to keep a pH of 5.5 and a continuous feed.

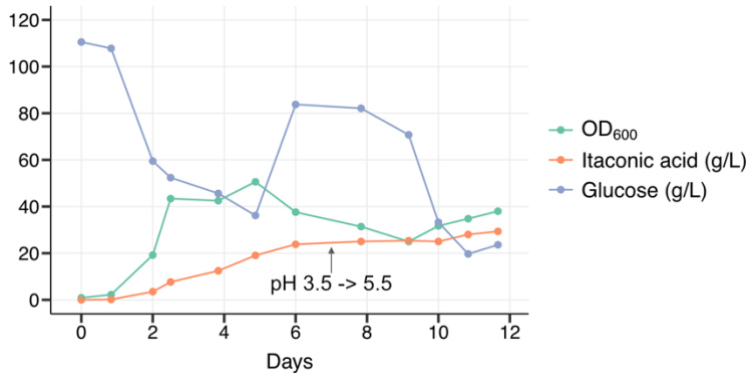


Figure 21: Media composition (NR -> NL): 10 g/L $(NH_4)_2SO_4$, 3 g/L KH_2PO_4, 0.5 g/L $MgSO_4 \bullet 7H_2O$, 2 mL trace metals solution stock, and 1 mL of vitamin solution stock. 100 g/L initial glucose. 650 g/L glucose with 30 g/L yeast extract was fed continuously. Initial pH was 3.5 and was then increased to 5.5.

The key factors that we identified are addition of yeast extract, continuous feeding, and pH of 5.5. These conditions were combined in 1 L fed-batch cultivation: the pH was kept at 5.5, and 600 g/L glucose combined with 20 g/L yeast extract were continuously fed into the reactor.

Using ammonium sulphate as nitrogen source, we produced 68.1 g/L of itaconic acid in 16.75 days (Figure 22). However, to maintain the pH at 5.5, large volumes of base were required (Figure 22). This drastically increased the cultivation volume, lowering the itaconic acid titre after 17 days, although the absolute amount of itaconic acid still increases (titre remained similar, but the volume increased: the total amount of itaconic acid increased).

In Chapter 2 we learnt that ammonium sulphate consumption reduces the pH of the media, while urea consumption does not. A cultivation with urea reduced the volume of base required to keep a pH of 5.5 – significantly increasing itaconic acid titre and yield to 130.5 g/L and 0.320 mol/mol glucose respectively (Figure 22).

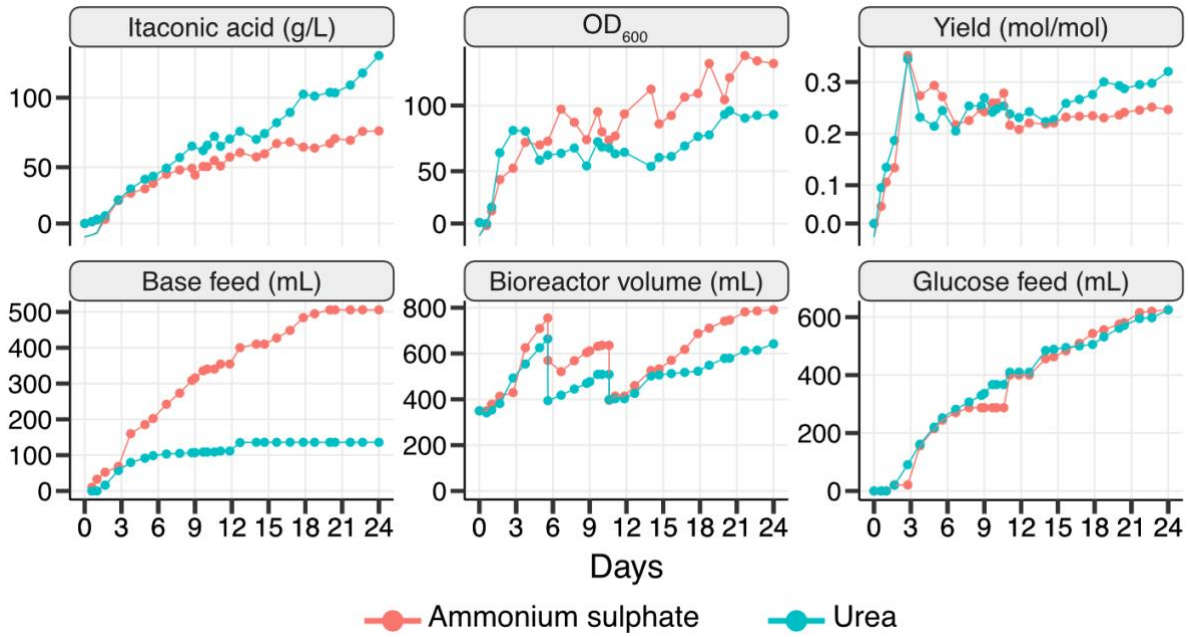


Figure 22: Fed-batch fermentation with ammonium sulphate or urea. Media composition: 10 g/L $(NH_4)_2SO_4$ or 4.5 g/L urea, 3 g/L KH_2PO_4, 0.5 g/L $MgSO_4 \cdot 7H_2O$, 2 mL trace metals solution stock, and 1 mL of vitamin solution stock. 100 g/L initial glucose, pH 5.5. 650 g/L glucose with 30 g/L yeast extract was fed continuously. Base feed: 6M KOH.

Conclusions and outlook

The goal of this study was to establish and increase itaconic production in *Y. lipolytica*. Through metabolic engineering and by optimizing fermentation conditions, we substantially increased itaconic acid titres, reaching 130 g/L in fed-batch cultivations.

The key parameters we identified to enhance itaconic acid production and minimize citric acid accumulation include the addition of yeast extract, continuous nutrient feeding, and maintaining a pH of 5.5.

Building on the findings from Chapter 2, we switched the nitrogen source from ammonium sulphate to urea. This proved effective in maintaining the desired pH without excessively increasing reactor volume, further boosting itaconic acid titres and yields. Additionally, limiting base consumptions can reduce chemical expenses, lower waste treatment cost, and facilitate downstream processing.

In Chapter 1 we learnt that disrupting lipid synthesis can induce protein misfolding and stress responses, potentially lowering cell robustness and productivity in industrial bioprocesses. It will be worth exploring the effect of downregulating *DGA1*, *DGA2*, *ARE1*, and *LRO1* to see if this increases itaconic acid production.

Future work will also need focus on (I) scaling up the fed-batch process to confirm that the identified parameters can be translated to larger scales, (II) optimizing downstream processing to recover itaconic acid from fermentation broths and lower costs, (III) investigating cheaper alternatives for yeast extract and other media components.

Chapter 4 – Single cell transcriptomics

Preface

During the years, my interest in omics analysis evolved, and I sought a collaboration at New York University to perform single cell transcriptomics.

The analysis was performed on *Saccharomyces cerevisiae* because it is a well-established model organism with extensive genomic resources and abundant existing transcriptomic data, making data interpretation more straightforward. Additionally, several genetic engineering tools are available in *S. cerevisiae*, facilitating laboratory verification of omics results.

The analysis performed on *S. cerevisiae* is a proof-of-concept and explores stochasticity in gene expression. The goal of the project, asides from investigating gene stochasticity in *S. cerevisiae*, it to identify key challenges in translating single-cell transcriptomics to *Y. lipolytica*.

In the first part I will discuss gene stochasticity, and in the last section I will outline the challenges in translating this analysis to *Y. lipolytica*.

Summary

Genetically identical populations of cells exhibit phenotypic variation due to stochastic (random) gene expression[127–130], which enables subpopulations of cells to survive adverse conditions such as antibiotic treatment[131–134].

In this study, we find that in actively dividing cells, a simplistic approach to identify genes that show the highest variance recovers genes that are differentially regulated through the mitotic cell cycle and is therefore uninformative.

To mitigate this, we computationally assigned the cell cycle phase and a discrete cell cycle time to each cell. We then divided cells into three-minute intervals to minimize cell cycle variability.

This approach allowed us to identify genes that show high variability, revealing an extensive landscape of variable gene expression that may underlie bet-hedging strategies used by cells to diversify their phenotype and increase their likelihood of surviving harsh conditions.

Introduction

Genetically identical populations of cells exhibit phenotypic variation due to stochastic (random) gene expression[127–130]. A stochastic gene is a gene whose expression is highly variable, leading to fluctuations in the gene expression level even under identical conditions.

This stochastic variability in gene expression can enable subpopulations of cells to survive adverse conditions such as antibiotic treatment[131–134]. Previous studies reported that single cell stochasticity in *S. cerevisiae* contribute to adaptation dynamics in response to nutrient shift and temperature and osmotic shock[135].

However, most studies could only focus on a handful of putative stochastic genes due to technical and technological limitations, as it was only possible to track specific mRNAs using a combination of automated microscopy, fluorescent reporter, and in-situ.

The situation changed in recent years with the development of single cell RNA sequencing, which allows to unravel not only the heterogeneity and complexity of RNA transcripts within individual cells, but also to follow how RNA transcripts abundances change over time[136].

Experimental setup

In this study we analyzed a published single cell transcriptomic dataset in *S. cerevisiae* [137] and develop a method for computationally identifying stochastic genes.

The dataset contains the sequencing of 5843 mRNAs for approx. 175000 individual cells before and after they were subjected to rapamycin treatment. To minimize the interval between individual data points, the culture was continually pumped into excess saturated ammonium sulfate and RNAlater to collect and fix cells in separate samples over sequential 10 minute intervals. The adopted sampling design captured the transcriptome of individual cells over a continuous temporal distribution, unlike a standard discrete time point sampling.

Results and discussion

Cell cycle genes are a confounding variable

We investigated if yeast has genes that are stochastically expressed under exponential growth conditions in rich media. To do this, we selected the 45000 cells from our dataset which are not treated with rapamycin, and computed the coefficient of variation (CV) for every gene in the dataset.

The CV is the ratio between standard deviation and mean, and it measures of variability in relation to the mean. The CV can be used to identify highly variable genes across cell populations[138].

We fitted a linear model between $\log_{10}$(CV) and $\log_{10}$(mean) and calculated the residual values for each gene (Figure 23A). We reason that genes with positive residuals are more variable than expected, and are candidates for highly variable, stochastic expression. However, these

highly variable genes are enriched for cell cycle-related genes (Figure 23B) when compared to an annotated list of cell cycle genes (Spellman et al. dataset[139]). We found that 51 of the 100 most variable genes are associated with the cell cycle (Figure 23B). However, cell cycle genes are not the target of our analysis, and therefore introduce a confounding factor in our analysis.

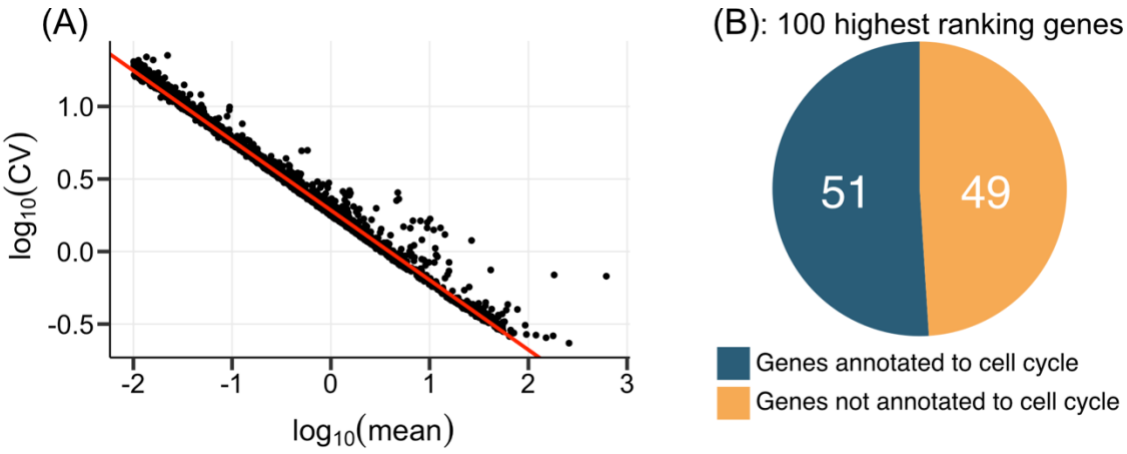


Figure 23: (A) Red line represents the linear model between log_{10}(CV) and log_{10}(mean). Dots represent individual genes, and the distance between each dot and the linear model is the residual. (B) Gene annotations for the 100 genes with highest residuals were matched against the Spellman et al. dataset[139], revealing that 51 genes are associated with the cell cycle.

Stochastic genes exhibit random fluctuations in expression due to inherent cellular noise, resulting in significant variability. In contrast, cell cycle genes have high variability due to periodic regulation which is timed and controlled by specific regulatory networks. Here, we want to extract the genes with highest residuals to reflect stochastic genes only, not genes with periodic regulation. Therefore, the variability of cell cycle introduces a confounding factor in our study, and to minimize its impact on our analysis we decided to:

I) assign cell cycle phase and time in the cell cycle to each cell,
II) divide cells into short time intervals along the cell cycle, under the assumption that if the intervals are short enough, cell cycle genes will not change in expression and will have low variability,
III) extract genes that still show higher than expected variability in each time interval,
IV) combine information across each time interval to extract the most variable and stochastic genes.

We opted for this approach, instead of removing genes annotated to the cell cycle, since certain genes might not have explicit cell cycle annotations but could still be associated and correlate with it.

Assigning time in the cell cycle to individual cells

To minimize the effect of the cell cycle as confounding variable, we assigned the cycle phase (G1, S, G2, M, M-G1) (Figure 24A) and the corresponding time within the cell cycle (1 to 90 minutes) to each cell in the dataset (Figure 24B), following an approach previously developed[137].

This approach allowed us to computationally synchronize a large population of *S. cerevisiae* cells along the 90 minutes trajectory of the cell cycle (Figure 24B). Assigning a specific time in the cell cycle to individual cells will allow us to split cells in short intervals and, assuming the interval is short enough, to minimize the variability of cell cycle genes, since their expression will not significantly change.

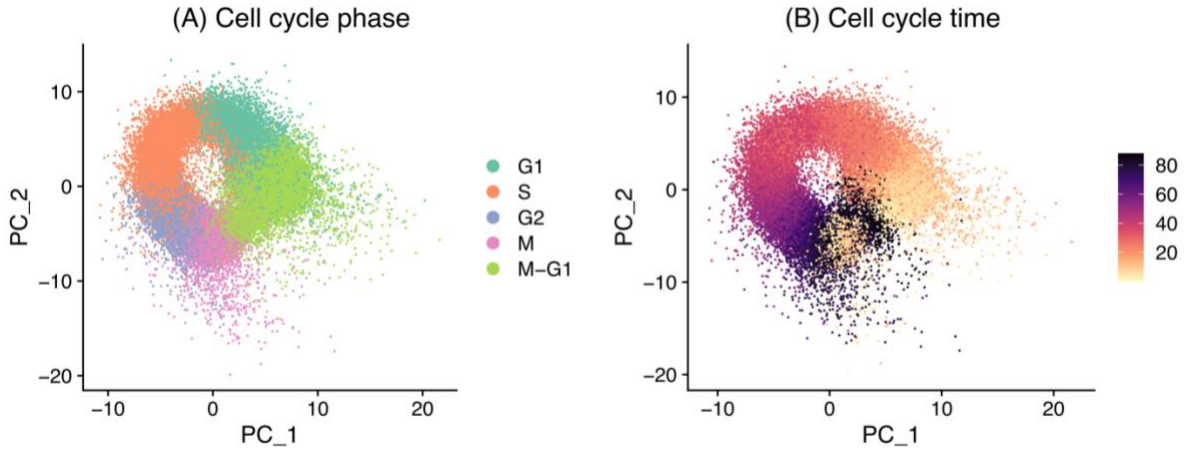


Figure 24: PCA of cells before rapamycin treatment. Cells were assigned (A) a cell cycle phase and (B) a continuous time from 1 to 90 minutes based on their gene expression program, according to markers from Spellman et al.[139]*, and using an approach previously developed*[137]*.*

To divide cells into the smallest time intervals possible along the cell cycle without losing statistical power, we calculated the minimum number of cells needed to estimate a model that accurately represents the overall population dynamics.

To do this, we employed a bootstrapping technique, subsampling the dataset with cell numbers ranging from 250 to 2000. Each subsample size was iterated 1000 times, and for every subsample size iteration we estimated a linear model and calculated the residuals of each gene. After calculating the average residual for each combination of *sample size* and *gene* we computed the Spearman's coefficient of correlation for each combination of sampling size (paper IV, table 1). Our analysis reveals that, in our dataset, 750 to 1000 cells per time window yields models that closely align with the population model, with Spearman's coefficients of 0.947 and 0.961 respectively.

To minimize the variability associated with the cell cycle we then divided cells in three-minutes intervals, which is the shortest interval that guarantees at least 750 cells in each time bin (Figure 25A).

For each time bin we computed the CV and mean for each gene across cells belonging to that time interval. After fitting a linear model and calculating the residuals for each gene, we ranked the genes according to their residuals, i.e. from genes whose variability is higher than expected to lower than expected. After repeating this process for each time bin, from 3 to 90 minutes, we combined the rank across all the time bins by summing the ranks of each gene across all time bins.

We then checked the 100 highest ranking genes (Figure 25B) and found that 21 genes out of 100 are associated with the cell cycle in the Spellman et al.[139] dataset, a major improvement compared to the previous approach, where 51 genes out of 100 were annotated to cell cycle. This approach allowed us to extract the genes with highest residuals to reflect stochastic genes only, not genes with periodic regulation.

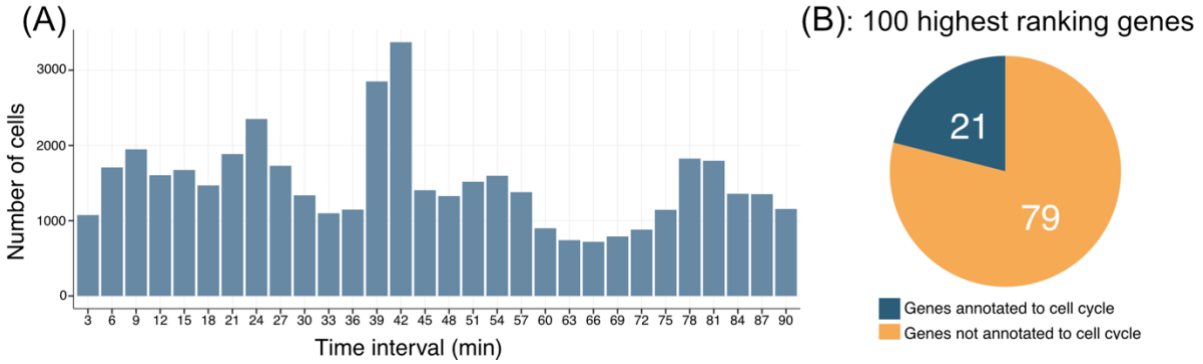


Figure 25: (A) number of cells in each three-minute time bin. (B) gene annotations for the 100 genes with highest residuals were matched against the Spellman et al. dataset[139].

Extracting biological information from stochastic genes

To extract biological information from the list of the putative stochastic genes, we performed a gene set analysis (GSA) on genes ranked based on the combined rank and with an average gene count across cells below 1 (Figure 26). This threshold was chosen because, in large systems, the addition or removal of a single molecule usually has minimal impact on system properties. However, in smaller systems, stochastic fluctuations can have more significant effects[140]. Among the enriched gene sets, we found gene sets related to:

- secondary alcohol metabolism, essential for the detoxification of harmful compounds,
- oxidoreductase activity, fundamental for cellular respiration and oxidative stress management,
- metal binding and transport, critical for maintaining cellular homeostasis as metals are vital cofactors for many enzymes and structural proteins.

The enrichment of these gene sets indicates that a significant portion of the genes with high variability are involved in biological processes typically activated in response to external perturbations, suggesting that these genes play a pivotal role in how cells adapt to changes in their environment. Cells exhibiting stochastic expression of genes involved in these gene sets might be more resistant to sudden and unexpected perturbation, since the gene(s) to respond to those perturbation are already expressed. When subject to an external perturbation, cells stochastically expressing genes involved in the response pathways can react immediately, while cells that do not express these genes must first activate the specific transcription programs. This delay in response might reduce their ability to cope with the perturbation, impairing their survival.

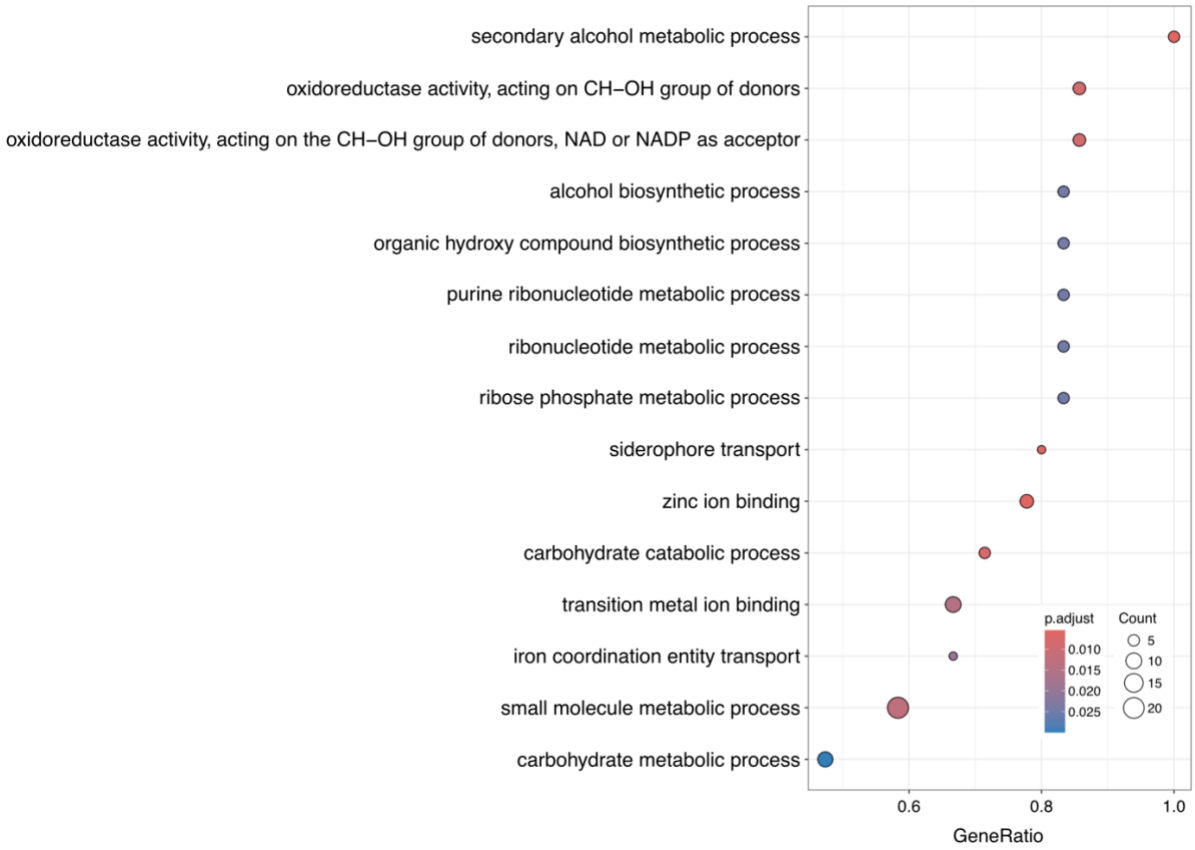


Figure 26: GSA using the package clusterProfiler[141,142]. The analysis focuses on genes ranked based on the combined rank and with counts less than 1. The dot colour represents the adjusted p-value ($-\log_{10}$(FDR)) and the dot size represents the number of DEGs in each GO term. The X-axis indicates the number of differentially expressed genes in each GO term relative to the total number of genes in the cluster.

Among the enriched gene sets, we investigated the siderophore transport gene set. Siderophore transport is defined as "*the directed movement of siderophores, low molecular weight Fe(III)-chelating substances, into, out of or within a cell, or between cells*" by the Yeast Genome Database (May 2024)[143] . Iron is an essential micronutrient because it participates as a redox cofactor in many cellular processes[144]. Iron-containing compounds need to cross the cell wall and the periplasmic space before uptake systems can shuttle the compounds through

the plasma membrane[145]. Among these, *FIT1, FIT2, and FIT3* are important for the non-reductive iron import machinery[144].

FIT2 and *FIT3* rank among the top 10 genes whose residuals are higher than expected across all timepoints. This indicates that cells are stochastically expressing different levels of these genes. Such variability in gene expression may function as a bet-hedging mechanism, a strategy that spreads risk by diversifying phenotypes within isogenic populations[146]. Some cells randomly maintain high expression levels of stochastic genes, which comes at the cost of spending energy to synthetize proteins that might not be required under current environmental conditions. However, upon a sudden environmental shift, a subpopulation synthetizing these proteins can have a selective advantage, since it might be more prepared to adapt and respond to the environmental fluctuation. In the case of *FIT2* and *FIT3* stochastic expression ensures that iron uptake can occur efficiently. This allows the population as a whole to survive and adapt to sudden changes in iron availability, preserving the gene pool of the population.

We then tested for correlated behavior of stochastically expressed genes: we examined whether the expression level of any stochastic gene correlates with the expression level of other stochastic genes. We found that most genes do not show correlated behaviour and have a coefficient of determination below 0.2 (Figure 27A). However, we found that *FIT2* and *FIT3* are exceptions. These two genes are stochastically expressed and exhibit unusually high correlated expression, with a coefficient of determination of 0.36 (Figure 27B), suggesting stochastic activity of a shared regulator.

This result requires further elucidation of the nature of this regulation. Additional experiments, such as in situ hybridization, will be necessary to achieve this. These studies will enable us to track the expression patterns of FIT2 and FIT3 under various conditions, providing deeper insights into their regulatory dynamics and potential functional interactions.

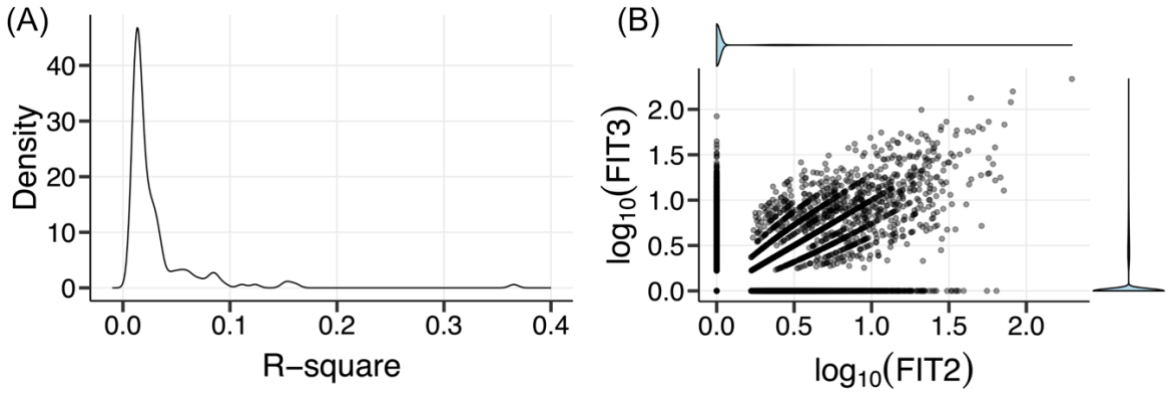


Figure 27: (A) density plot of the correlation coefficient calculated between gene counts of gene pairs, filtering for R-square above 0.01. (B) Marginal plot of FIT2 and FIT3 gene counts. Each dot represents an individual cell (45 000 cells). The violin plots on the margins illustrate the distribution of gene counts for FIT3 (right) and FIT2 (top).

Conclusions and outlook

In this study, we aimed to identify stochastic gene expression in *S. cerevisiae* by analyzing the transcriptomes of 175,000 individual cells continuously sampled over a one-hour period.

We found that in actively dividing cells, a simplistic approach to identify genes that show the highest variance recovers genes that are differentially regulated through the mitotic cell cycle. This approach is therefore uninformative for us.

To mitigate this, we computationally assigned the cell cycle phase and a discrete cell cycle time to each cell in the dataset. This allowed us to computationally synchronize a population of actively dividing cells. After grouping cells into three-minute intervals along the cell cycle we were able to identify stochastic genes whose expression is highly variable along the cell cycle.

Through gene set analysis we revealed an enrichment in gene sets related to alcohol metabolism, oxidoreductase activity, and metal transport. Their stochastic expression may serve as a bet-hedging strategy for population survival, even at the cost of decreased fitness for single cells. We also found that the genes *FIT2* and *FIT3* are stochastically expressed but show high correlated expression, suggesting stochastic activity of their regulator.

Our study establishes a method for studying stochastic gene expression in mitotic cells, revealing an extensive landscape of variable gene expression that may underlie bet-hedging strategies used by cells to diversify their phenotype and increase their likelihood of survival.

Translating single cell transcriptomics to *Y. lipolytica*

The second goal of this study was to identify the key challenges in translating single-cell transcriptomics to *Y. lipolytica*.

The 10x Genomics[147] method for single cell transcriptomics was originally developed for mammalian cells, but later adjusted by different groups to yeast cells[148,149]. The first step of the protocol consists in digesting the cell wall to obtain a spheroplast. In the group where I performed this work, we noticed that digesting the cell wall of different *S. cerevisiae* strains already required substantial changes to the protocol, and that applying the protocol to a different yeast species proved challenging on several occasions. Translating the protocol to *Y. lipolytica* will require protocol adjustment and troubleshooting.

The framework developed here relies on cell cycle markers to identify the cell cycle phase and to calculate cell trajectory in the cell cycle. However, the genome of *Y. lipolytica* is less annotated than *S. cerevisiae*, which could pose challenges during analysis. Additionally, gene annotations are very important for inferring gene function, and the lack of extensive genomic resources *Y. lipolytica* can complicate data interpretation.

Summary, conclusions and outlook

In the last chapter, I will outline the main findings of my thesis, summarizing what we learnt during this journey. In the second section I will highlight the limitations of this study and how they can be addressed. Finally, I will contextualize our study in a broader context and give an outlook on what lays ahead.

What did we learn?

The goal of this thesis was to improve our understanding of the yeast *Y. lipolytica* to deploy it as a microbial cell factory. Through a combination of transcriptomics and fed-batch cultivations we identified important factors that need to be considered for future rounds of strain design and fermentation improvement.

In Chapter 1 we studied the effect of disrupting lipid synthesis to build a platform strain with high supply of acetyl-CoA. Several molecules that have acetyl-CoA as a building block can potentially be produced through this platform strain. We found that disrupting lipid synthesis negatively affects proteostasis and leads to an enrichment of protein misfolding and degradation. Based on these findings, we concluded that to improve our *Y. lipolytica* platform strain, it would be preferable to downregulate the genes involved in lipid synthesis instead of deleting them. This approach would ensure a balanced allocation of cellular resources between molecule production and physiological homeostasis. However, gene downregulation also presents challenges such as establishing an optimal downregulated expression level, achieving efficient and specific downregulation without off-target effects, and ensuring that the cells do not activate compensatory mechanisms. Additionally, fine-tuning gene expression based on one condition might reduce the strain's adaptability to changing environments, which are for instance encountered during scale-up.

In Chapter 2 we studied the effect of using urea as a nitrogen source instead of ammonium sulphate. Urea can be a cheaper and more sustainable nitrogen source if extracted from waste. In our study we found no significant coherent changes in growth and lipid production. We found no significant coherent changes neither in the transcriptome, nor in the genes involved in urea uptake and degradation. A previous study[123] observed changes in the fatty acid profile of OKYL029 when cultivated with urea under nitrogen limitation. This is likely due to pH changes since ammonium consumption acidifies the media. Our chemostat cultures were pH-controlled, preventing any changes in pH. Our findings support urea usage, indicating that previous metabolic engineering efforts are likely translatable to urea, as we showed in Chapter 3. Additionally, first experiments using synthetic and real human urine as nitrogen sources for

cultivating *Y. lipolytica* showed promising results, with growth and biomass formations similar to ammonium sulfate[150].

In Chapter 3 we improved a fed-batch cultivation of *Y. lipolytica* to enhance itaconic acid production. Itaconic acid ranks among the top 12 building block chemicals and has several applications in food, textile, and pharmaceutical industries. The key parameters we identified to improve our fermentation outcomes include addition of yeast extract, continuous nutrient feeding, and a pH of 5.5. Guided by the findings from Chapter 2, using urea as nitrogen source proved effective in maintaining the desired pH without excessively increasing reactor volume, further boosting itaconic acid titres and yields. This study represents a significant leap forward in establishing *Y. lipolytica* for the industrial production of itaconic acid. Additionally, due to the bow-tie structure of metabolism, a strain with high acetyl-CoA supply can be adapted to produce other organic acids thought pathways with similar enzyme regulation as itaconic acid. After appropriate genetic modifications, these fermentation conditions could offer a good foundation for optimising the production of other organic acids.

In Chapter 4 we applied innovative single-cell transcriptomics to *S. cerevisiae* for studying stochastic gene expression. Until recently, most studies could only focus on the single-cell expression of a handful of genes due to technical and technological limitations. However, in recent years the development of single-cell transcriptomics allowed us to unravel the heterogeneity and complexity of RNA transcripts within individual cells. After finding that cell cycle genes are a confounding variable in identifying genes with stochastic regulation, we developed a framework to study stochastic gene expression in mitotic cells, and revealed an extensive landscape of variable gene expression in yeast that may underlie bet-hedging strategies. There are currently no reported applications of single-cell transcriptomics in yeasts other than *S. cerevisiae*[151]. However, developing this framework helped us identify the key challenges in translating single-cell transcriptomics to the non-model yeast *Y. lipolytica*.

What can we improve?

Let's see which limitations are present in this thesis, and where there is room for improvement.

Most of the work in this thesis focuses on transcriptomics, both bulk and single-cell. The central dogma of biology states that genes are transcribed into mRNA, which is then translated into proteins. These proteins perform different tasks within the cell[31]. Transcriptomics measures the quantity of mRNAs and infers that an increase in transcript quantity corresponds to an increase in protein quantity and catalytical activity.

However, this is a simplification. Multiple studies have shown that the correlation between transcriptome and proteome is insufficient in yeast[152–154]. For example, Lahtvee et al.[154] quantified absolute abundances of mRNAs and proteins under ten environmental conditions and demonstrated low correlation between mRNA and protein abundances in *S. cerevisiae*. Interestingly, the mRNA-protein correlation was higher for differentially expressed proteins. These discrepancies between mRNA and protein abundances can be attributed to various factors. For instance, due to different 3' or 5' untranslated regions, not all mRNAs are translated at the same rate or have the same stability[155,156]. Additionally, once translated, proteins can undergo post-translational modifications, such as phosphorylation, acylation, or ubiquitylation, which can affect their activity and stability[157].

By measuring mRNA abundance, we only capture the early steps in a long chain of regulatory events. Nevertheless, RNA sequencing has become an affordable tool to investigate the transcriptional changes of a cell, providing cost-effective insights into cell regulation. When necessary, transcriptomics can be integrated with other omics such as proteomics and metabolomics to increase the level of details to which a process is studied. Although multi-omics are more expensive, technological advances will likely reduce the cost in the upcoming years.

Another limitation that we should be aware of for *Y. lipolytica* and other non-conventional organisms, is the limited gene annotation available.

Strategies based on sequence homology, such as BLAST, have been extensively used over the years to infer protein function annotations[158]. These techniques rely on the assumption that, if proteins share similar sequence, they will likely have the same function. However, it was shown that proteins with sequence identity between 20-35% fall in the twilight zone, where remote homologs can be confused with random sequences[159]. Below 20% identity, in the midnight zone, homologous relationships cannot be determined with simple pairwise alignments. As sequence identity decreases, the accuracy of predictions made with sequence-homology drops rapidly when sequence identity decreases[160].

To solve this, several machine learning strategies were developed[161]. These strategies, instead of relying on sequence homology, aim to understand how protein structure and function are encoded in protein sequences. However, since specific proteins have been characterized more extensively, these models can be biased, and getting good predictions for less characterized proteins can still prove challenging.

As a result of these limitations, a large portion of genes remains un-annotated. Out of the 7894 entries for the *Y. lipolytica* W29 proteome (UP000182444) on UniProt, 3865 lack annotated functions.

Additionally, homology-based strategies can only annotate genes that are conserved across species. Since these conserved genes are not the only genes responsible for the unique phenotype traits of non-conventional yeasts, we are failing to observe genes that are differentially expressed and contributing to interesting phenotypes due to lack of annotations. For example, in Chapter 1 we saw that out of the total 953 differentially expressed genes, only 390 had a function annotated on UniProt.

The future is however promising: thanks to new technological developments and the involvement of many companies like DeepMind with AlphaFold, new tools that yield better protein structures and gene annotations are being developed. These advances will significantly improve gene annotations in the upcoming years, reducing the gap between conventional and non-conventional yeasts.

Gene set analysis, a valuable method for interpreting transcriptomics data, faces similar limitations to gene annotations. The gene sets used for gene set analysis are grouped based on prior biological knowledge. As a consequnce, some un-annotated genes are not associated with any gene set. Additionally, certain pathways and biological processes are more studied and annotated than others, introducing a bias in the analysis. These over-represented pathways are more likely to be found as enriched.

Furthermore, only few GO terms are manually annotated, while approx. 98% are computationally inferred[162]. This is done mostly through sequence similarity, structural similarity, or phylogenetic closeness: this leads to the same limitations we encountered in gene annotation.

Despite these challenges, there are many ongoing efforts in gene annotations and pathway curation. New algorithms are constantly under development, and the reliability and applicability of gene set analysis will increase in the near future.

In this thesis, we aimed to understand specific biological responses to gene deletions and environmental conditions. To do this, we analysed the transcriptome and built models that simplify the reality to be able to interpret our results. However, the conditions we tested are limited, making it hard to predict outcomes of different conditions. For instance, in Chapter 1 we studied the effect of two C/N ratios in chemostat: it is still hard to predict which changes might occur in the transcriptome over the course of a batch fermentation, when the C/N changes as the carbon and nitrogen are consumed, without further experiments. We observed, modelled, and drew conclusions based on the tested conditions and the assumptions and constraints we applied. Different assumptions and constraints might have led

to different cellular behaviours, yet the most significant changes might remain consistent. Under different conditions cell behaviour might not be the same, but the most significant changes could be consistent. Only time and additional studies will tell which part of our results are specific to our experimental design, and which ones are a part of a general behaviour.

Why is this relevant?

Climate change is one of the biggest challenges that we need to solve. We will need to find sustainable ways to produce goods and foods and transition from an economy based on fossil fuels towards a circular economy.

Of the many innovations that emerged throughout the years, microbial cell factories are a promising one. Microbial cell factories can produce goods and foods from renewable and sustainable sources, effectively enabling a green transition by reducing our dependence on finite resources and minimizing environmental impact.

An increasing number of novel organisms are continuously discovered, expanding our Swiss Army Knife of microbial cell factories. Each newly characterised organism has unique metabolic features, enhancing the diversity of applications of microbial production. Additionally, thanks to the recent development in genetic engineering and systems biology, it is also becoming faster and cheaper to characterize, engineer, and fine-tune these novel organisms.

The combination of novel organisms with peculiar characteristics and novel techniques will allow us to develop several organisms that are suitable for specific challenges, instead of relying on a few well-established organisms that might be suboptimal for a specific bioprocess. This tailored approach will ensure that the most appropriate organism is used for each application, enhancing productivity and both environmental and economical sustainability.

This will result in an increasing number of applications of non-conventional organisms, such as *Yarrowia lipolytica*. *Yarrowia lipolytica* exemplifies how a non-conventional organism, with the peculiar characteristics of secreting lipases and accumulating lipids, can gain the attention of the biotech community. This organism is now extensively studied, applied across a wide variety of applications, and leveraged by many start-ups.

The insights gained from studying *Y. lipolytica* can be translated to other non-conventional and oleaginous yeasts. The methodologies, frameworks, and findings can guide the engineering of novel yeasts with similar traits to *Y. lipolytica*. The knowledge generated expands the available toolkit for the sustainable production of goods and foods, providing valuable strategies to characterize and optimize diverse yeast species.

Acknowledgements

At the end of this journey, I'd like to thank my supervisor, Eduard Kerkhoven, and my co-supervisor, Verena Siewers, for welcoming me into Sysbio and giving me the freedom and support to follow my curiosity and explore the projects I found interesting. I would also like to thank David Gresham and his research group at NYU for welcoming me and providing me with great scientific guidance.

Doing a PhD is challenging, and it can be difficult not to let stress overshadow amazing opportunities. I am certainly guilty of this, but looking back with clearer perspective, I realize how lucky I was with the many amazing opportunities I was given. It shouldn't be taken for granted to be able to follow my own curiosity, to have the opportunity to travel around the world for conferences, and to be able to spend six months in New York doing cutting-edge research. All these experiences made me grow immensely, both as a scientist and as a person.

A heartfelt thanks goes also to all the people that made these few years in Sweden something I will always look back to and cherish. There is a quote from the last episode of The Office that really resonates with me: "*I wish there was a way to know you're in the good old days, before you've actually left them*". Thanks to everyone for making all those old days, good old days. This journey would have been incredibly dull without you. Thanks for all the moments we shared: kayaking, climbing, hiking, trips around Sweden, Bosnia, Denmark, celebrating midsummers, swimming in freezing water, saunas, after-works, parties, ski trips, cozy indoors evening sheltering from the elements, whinging about Vårdcentral, Systembolaget, and the winter darkness, making amazing pizza, drinking Negroni, barbecuing. The list goes on and on. I will miss all of this.

A huge and special thanks to Tara, Cilia, Mauri, Vero, Andre, Olha, Andrés, Luisa, Mauro, Oliver, Marta, and my family. Your support and presence have meant the world to me - thank you for everything you did and for always being there.

References

1. (IPCC), I. P. on C. C. Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. *Climate Change 2021 – The Physical Science Basis* (2023) doi:10.1017/9781009157896.
2. World Bank Group Climate Change Action Plan 2021–2025: Supporting Green, Resilient, and Inclusive Development. Preprint at https://hdl.handle.net/10986/35799 (2021).
3. The concept of 'climate refugee': Towards a possible definition | Think Tank | European Parliament. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698753.
4. Gounaridis, D. & Newell, J. P. The social anatomy of climate change denial in the United States. *Scientific Reports 2024 14:1* **14**, 1–12 (2024).
5. Transitioning to a circular economy business | Report. https://www.ellenmacarthurfoundation.org/towards-the-circular-economy-vol-1-an-economic-and-business-rationale-for-an.
6. Geissdoerfer, M., Savaget, P., Bocken, N. M. P. & Hultink, E. J. The Circular Economy – A new sustainability paradigm? *J Clean Prod* **143**, 757–768 (2017).
7. Rittel, H. W. J. & Webber, M. M. Dilemmas in a general theory of planning. *Policy Sci* **4**, 155–169 (1973).
8. Conklin, J. Social Complexity Wicked Problems. *Wicked Problems & Social Complexity in Dialogue Mapping: Building Shared Understanding of Wicked Problems* 1–20 (2005).
9. Levin, K., Cashore, B., Bernstein, S. & Auld, G. Overcoming the tragedy of super wicked problems: Constraining our future selves to ameliorate global climate change. *Policy Sci* **45**, 123–152 (2012).
10. Our Common Future: From One Earth to One World - A/42/427 Annex, Overview - UN Documents: Gathering a body of global agreements. http://www.un-documents.net/ocf-ov.htm.
11. Assembly, G. A/RES/71/313: Work of the Statistical Commission pertaining to the 2030 Agenda for Sustainable Development. (2030).

12. Mannaa, M., Han, G., Seo, Y. S. & Park, I. Evolution of Food Fermentation Processes and the Use of Multi-Omics in Deciphering the Roles of the Microbiota. *Foods* **10**, (2021).

13. McGovern, P. E. *et al.* Fermented beverages of pre- and proto-historic China. *Proc Natl Acad Sci U S A* **101**, 17593–17598 (2004).

14. Bull, A. T., Ward, A. C. & Goodfellow, M. Search and Discovery Strategies for Biotechnology: the Paradigm Shift. *Microbiology and Molecular Biology Reviews* **64**, 573 (2000).

15. Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. *Proc Natl Acad Sci U S A* **113**, 5970–5975 (2016).

16. O'Brien, J. & Wright, G. D. An ecological perspective of microbial secondary metabolism. *Curr Opin Biotechnol* **22**, 552–558 (2011).

17. Pham, J. V. *et al.* A review of the microbial production of bioactive natural products and biologics. *Front Microbiol* **10**, 449147 (2019).

18. Campbell, K., Xia, J. & Nielsen, J. The Impact of Systems Biology on Bioprocessing. *Trends Biotechnol* **35**, 1156–1168 (2017).

19. Kim, J. Y., Ahn, Y. J., Lee, J. A. & Lee, S. Y. Recent advances in the production of platform chemicals using metabolically engineered microorganisms. *Curr Opin Green Sustain Chem* **40**, 100777 (2023).

20. Kim, G. B., Choi, S. Y., Cho, I. J., Ahn, D. H. & Lee, S. Y. Metabolic engineering for sustainability and health. *Trends Biotechnol* **41**, 425–451 (2023).

21. Cho, J. S., Kim, G. B., Eun, H., Moon, C. W. & Lee, S. Y. Designing Microbial Cell Factories for the Production of Chemicals. *JACS Au* **2**, 1781–1799 (2022).

22. Lee, R. A. & Lavoie, J. M. From first- to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity. *Animal Frontiers* **3**, 6–11 (2013).

23. Bailey, J. E. Toward a science of metabolic engineering. *Science* **252**, 1668–1675 (1991).

24. Stephanopoulos, G., Aristidou, A. A. & Nielsen, J. Høiriis. Metabolic engineering : principles and methodologies. 725.

25. Nielsen, J. & Keasling, J. D. Engineering Cellular Metabolism. *Cell* **164**, 1185–1197 (2016).

26. Freemont, P. S. Synthetic biology industry: data-driven design is creating new opportunities in biotechnology. *Emerg Top Life Sci* **3**, 651–657 (2019).

27. Opgenorth, P. *et al.* Lessons from Two Design-Build-Test-Learn Cycles of Dodecanol Production in Escherichia coli Aided by Machine Learning. *ACS Synth Biol* **8**, 1337–1351 (2019).

28. Chubukov, V., Mukhopadhyay, A., Petzold, C. J., Keasling, J. D. & Martín, H. G. Synthetic and systems biology for microbial production of commodity chemicals. *NPJ Syst Biol Appl* **2**, 1–11 (2018).

29. Breitling, R. What is systems biology? *Front Physiol* **1**, (2010).

30. Pinu, F. R. *et al.* Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community. *Metabolites 2019, Vol. 9, Page 76* **9**, 76 (2019).

31. Crick, F. Central Dogma of Molecular Biology. *Nature 1970 227:5258* **227**, 561–563 (1970).

32. Box, G. E. P. Robustness in the Strategy of Scientific Model Building. *Robustness in Statistics* 201–236 (1979) doi:10.1016/B978-0-12-438150-6.50018-2.

33. Navarrete, C., Martínez, J. L., Navarrete, C. & Martínez, J. L. Non-conventional yeasts as superior production platforms for sustainable fermentation based bio-manufacturing processes. *AIMS Bioengineering 2020 4:289* **7**, 289–305 (2020).

34. Geijer, C., Ledesma-Amaro, R. & Tomas-Pejo, E. Unraveling the potential of non-conventional yeasts in biotechnology. *FEMS Yeast Res* **22**, 71 (2022).

35. Madzak, C. Yarrowia lipolytica: recent achievements in heterologous protein expression and pathway engineering. *Appl Microbiol Biotechnol* **99**, 4559–4577 (2015).

36. Madzak, C., Gaillardin, C. & Beckerich, J. M. Heterologous protein expression and secretion in the non-conventional yeast Yarrowia lipolytica: A review. *J Biotechnol* **109**, 63–81 (2004).

37. De Pourcq, K. *et al.* Engineering the yeast Yarrowia lipolytica for the production of therapeutic proteins homogeneously glycosylated with Man8GlcNAc2 and Man5GlcNAc2. *Microb Cell Fact* **11**, 1–12 (2012).

38. Nicaud, J.-M. Yarrowia lipolytica. *Yeast* **29**, 409–418 (2012).

39. Shi, T. Q., Huang, H., Kerkhoven, E. J. & Ji, X. J. Advancing metabolic engineering of Yarrowia lipolytica using the CRISPR/Cas system. *Appl Microbiol Biotechnol* **102**, 9541–9548 (2018).

40. Holkenbrink, C. *et al.* EasyCloneYALI: CRISPR/Cas9-Based Synthetic Toolbox for Engineering of the Yeast Yarrowia lipolytica. *Biotechnol J* **13**, 1–8 (2018).

41. Madzak, C. Engineering Yarrowia lipolytica for Use in Biotechnological Applications: A Review of Major Achievements and Recent Innovations. *Mol Biotechnol* **60**, 621–635 (2018).

42. Park, Y. K. & Ledesma-Amaro, R. What makes Yarrowia lipolytica well suited for industry? *Trends Biotechnol* **41**, 242–254 (2023).

43. Konzock, O., Matsushita, Y., Zaghen, S., Sako, A. & Norbeck, J. Altering the fatty acid profile of Yarrowia lipolytica to mimic cocoa butter by genetic engineering of desaturases. *Microb Cell Fact* **21**, 1–11 (2022).

44. Zeng, S. Y. *et al.* Recent Advances in Metabolic Engineering of Yarrowia lipolytica for Lipid Overproduction. *European Journal of Lipid Science and Technology* vol. 120 1700352 Preprint at https://doi.org/10.1002/ejlt.201700352 (2018).

45. Palmer, C. M., Miller, K. K., Nguyen, A. & Alper, H. S. Engineering 4-coumaroyl-CoA derived polyketide production in Yarrowia lipolytica through a β-oxidation mediated strategy. *Metab Eng* **57**, 174–181 (2020).

46. Rodriguez, A. *et al.* Metabolic engineering of yeast for fermentative production of flavonoids. *Bioresour Technol* **245**, 1645–1654 (2017).

47. Markham, K. A. *et al.* Rewiring yarrowia lipolytica toward triacetic acid lactone for materials generation. *Proc Natl Acad Sci U S A* **115**, 2096–2101 (2018).

48. Fu, J. *et al.* Reprogramming Yarrowia lipolytica metabolism for efficient synthesis of itaconic acid from flask to semipilot scale. *Sci Adv* **10**, (2024).

49. Tramontin, L. R. R., Kildegaard, K. R., Sudarsan, S. & Borodina, I. Enhancement of Astaxanthin Biosynthesis in Oleaginous Yeast Yarrowia lipolytica via Microalgal Pathway. *Microorganisms* **7**, (2019).

50. Thomsen, P. T. *et al.* Beet red food colourant can be produced more sustainably with engineered Yarrowia lipolytica. *Nature Microbiology 2023 8:12* **8**, 2290–2303 (2023).

51. Holkenbrink, C. *et al.* Production of moth sex pheromones for pest control by yeast fermentation. *Metab Eng* **62**, 312–321 (2020).

52. Petkevicius, K. *et al.* Biotechnological production of the European corn borer sex pheromone in the yeast Yarrowia lipolytica. *Biotechnol J* **16**, 2100004 (2021).

53. Hambalko, J. *et al.* Production of Long Chain Fatty Alcohols Found in Bumblebee Pheromones by Yarrowia lipolytica. *Front Bioeng Biotechnol* **8**, 593419 (2021).

54. Kosiorowska, K. E., Biniarz, P., Dobrowolski, A., Leluk, K. & Mirończuk, A. M. Metabolic engineering of Yarrowia lipolytica for poly(ethylene terephthalate) degradation. *Science of The Total Environment* **831**, 154841 (2022).

55. Ledesma-Amaro, R. & Nicaud, J. M. Metabolic Engineering for Expanding the Substrate Range of Yarrowia lipolytica. *Trends Biotechnol* **34**, 798–809 (2016).

56. Lv, Y., Marsafari, M., Koffas, M., Zhou, J. & Xu, P. Optimizing Oleaginous Yeast Cell Factories for Flavonoids and Hydroxylated Flavonoids Biosynthesis. *ACS Synth Biol* **8**, 2514–2523 (2019).

57. Matthäus, F., Ketelhot, M., Gatter, M. & Barth, G. Production of lycopene in the non-carotenoid-producing yeast Yarrowia lipolytica. *Appl Environ Microbiol* **80**, 1660–1669 (2014).

58. Gao, S. *et al.* Production of β-carotene by expressing a heterologous multifunctional carotene synthase in Yarrowia lipolytica. *Biotechnol Lett* **39**, 921–927 (2017).

59. Cheng, B. Q., Wei, L. J., Lv, Y. B., Chen, J. & Hua, Q. Elevating Limonene Production in Oleaginous Yeast Yarrowia lipolytica via Genetic Engineering of Limonene Biosynthesis Pathway and Optimization of Medium Composition. *Biotechnology and Bioprocess Engineering* **24**, 500–506 (2019).

60. Ku, J. T., Chen, A. Y. & Lan, E. I. Metabolic Engineering Design Strategies for Increasing Acetyl-CoA Flux. *Metabolites* **10**, (2020).

61. Beopoulos, A. *et al.* Control of lipid accumulation in the yeast Yarrowia lipolytica. *Appl Environ Microbiol* **74**, 7779–7789 (2008).

62. Kerkhoven, E. J., Pomraning, K. R., Baker, S. E. & Nielsen, J. Regulation of amino-acid metabolism controls flux to lipid accumulation in yarrowia lipolytica. *NPJ Syst Biol Appl* **2**, 1–7 (2016).

63. Beopoulos, A. *et al.* Identification and characterization of DGA2, an acyltransferase of the DGAT1 acyl-CoA:diacylglycerol acyltransferase family in the oleaginous yeast Yarrowia lipolytica. New insights into the storage lipid metabolism of oleaginous yeasts. *Appl Microbiol Biotechnol* **93**, 1523–1537 (2012).

64. Athenstaedt, K. YALI0E32769g (DGA1) and YALI0E16797g (LRO1) encode major triacylglycerol synthases of the oleaginous yeast Yarrowia lipolytica. *Biochim Biophys Acta Mol Cell Biol Lipids* **1811**, 587–596 (2011).

65. Shi, T. *et al.* Engineering the oleaginous yeast Yarrowia lipolytica for β-farnesene overproduction. *Biotechnol J* **16**, (2021).

66. Poorinmohammad, N., Fu, J., Wabeke, B. & Kerkhoven, E. J. Validated Growth Rate-Dependent Regulation of Lipid Metabolism in Yarrowia lipolytica. *Int J Mol Sci* **23**, (2022).

67. Hapeta, P., Kerkhoven, E. J. & Lazar, Z. Nitrogen as the major factor influencing gene expression in Yarrowia lipolytica. *Biotechnology Reports* **27**, e00521 (2020).

68. Konzock, O., Zaghen, S. & Norbeck, J. Tolerance of Yarrowia lipolytica to inhibitors commonly found in lignocellulosic hydrolysates. *BMC Microbiol* **21**, 1–10 (2021).

69. Kaluzny, M. A., Duncan, L. A., Merritt, M. V. & Epps, D. E. Rapid separation of lipid classes in high yield and purity using bonded phase columns. *J Lipid Res* **26**, 135–140 (1985).

70. Koch, B., Schmidt, C. & Daum, G. Storage lipids of yeasts: A survey of nonpolar lipid metabolism in Saccharomyces cerevisiae, Pichia pastoris, and Yarrowia lipolytica. *FEMS Microbiol Rev* **38**, 892–915 (2014).

71. Petschnigg, J. *et al.* Good fat, essential cellular requirements for triacylglycerol synthesis to maintain membrane homeostasis in yeast. *Journal of Biological Chemistry* **284**, 30981–30993 (2009).

72. Olzmann, J. A. & Carvalho, P. Dynamics and functions of lipid droplets. *Nat Rev Mol Cell Biol* **20**, 137–155 (2019).

73. Friedlander, J. *et al.* Engineering of a high lipid producing Yarrowia lipolytica strain. *Biotechnol Biofuels* **9**, 1–12 (2016).

74. Gajdoš, P., Ledesma-Amaro, R., Nicaud, J. M., Čertík, M. & Rossignol, T. Overexpression of diacylglycerol acyltransferase in Yarrowia lipolytica affects lipid body size, number and distribution. *FEMS Yeast Res* **16**, 1–8 (2016).

75. Garbarino, J. *et al.* Sterol and diacylglycerol acyltransferase deficiency triggers fatty acid-mediated cell death. *Journal of Biological Chemistry* **284**, 30994–31005 (2009).

76. Maleki, F., Ovens, K., Hogan, D. J. & Kusalik, A. J. Gene Set Analysis: Challenges, Opportunities, and Future Research. *Front Genet* **11**, 1–16 (2020).

77. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nat Genet* **25**, 25–29 (2000).

78. Thomas, P. D. The Gene Ontology Handbook. *Methods in Molecular Biology* **1446**, 1–9 (2017).

79. Väremo, L., Nielsen, J. & Nookaew, I. 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 (2013).

80. Buchner, J. Molecular chaperones and protein quality control: An introduction to the JBC Reviews thematic series. *Journal of Biological Chemistry* **294**, 2074–2075 (2019).

81. Thibault, G. & Ng, D. T. W. The endoplasmic reticulum-associated degradation pathways of budding yeast. *Cold Spring Harb Perspect Biol* **4**, 1–15 (2012).

82. Finley, D., Ulrich, H. D., Sommer, T. & Kaiser, P. The ubiquitin-proteasome system of Saccharomyces cerevisiae. *Genetics* **192**, 319–360 (2012).

83. Suda, Y. & Nakano, A. The Yeast Golgi Apparatus. *Traffic* **13**, 505–510 (2012).

84. Graef, M. Lipid droplet-mediated lipid and protein homeostasis in budding yeast. *FEBS Lett* **592**, 1291–1303 (2018).

85. Nguyen, T. B. *et al.* DGAT1-Dependent Lipid Droplet Biogenesis Protects Mitochondrial Function during Starvation-Induced Autophagy. *Dev Cell* **42**, 9-21.e5 (2017).

86. Velázquez, A. P., Tatsuta, T., Ghillebert, R., Drescher, I. & Graef, M. Lipid droplet-mediated ER homeostasis regulates autophagy and cell survival during starvation. *Journal of Cell Biology* **212**, 621–631 (2016).

87. Deretic, V. Lipid droplets and their component triglycerides and steryl esters regulate autophagosome biogenesis. *EMBO J* **34**, 2111–2113 (2015).

88. Moldavski, O. *et al.* Lipid Droplets Are Essential for Efficient Clearance of Cytosolic Inclusion Bodies. *Dev Cell* **33**, 603–610 (2015).

89. Vasconcelos, B., Teixeira, J. C., Dragone, G. & Teixeira, J. A. Oleaginous yeasts for sustainable lipid production—from biodiesel to surf boards, a wide range of "green" applications. *Appl Microbiol Biotechnol* 3651–3667 (2019) doi:10.1007/s00253-019-09742-x.

90. Wang, M. *et al.* Can sustainable ammonia synthesis pathways compete with fossil-fuel based Haber-Bosch processes? *Energy Environ Sci* **14**, 2535–2548 (2021).

91. Antonetti, E. *et al.* Waste-to-Chemicals for a Circular Economy: The Case of Urea Production (Waste-to-Urea). *ChemSusChem* **10**, 912–920 (2017).

92. Evans, C. T. & Ratledge, C. Effect of nitrogen source on lipid accumulation in oleaginous yeasts. *J Gen Microbiol* **130**, 1693–1704 (1984).

93. Albers, E., Larsson, C., Lidén, G., Niklasson, C. & Gustafsson, L. Influence of the nitrogen source on Saccharomyces cerevisiae anaerobic growth and product formation. *Appl Environ Microbiol* **62**, 3187–3195 (1996).

94. Kyriakou, V., Garagounis, I., Vourros, A., Vasileiou, E. & Stoukides, M. An Electrochemical Haber-Bosch Process. *Joule* **4**, 142–158 (2020).

95. Song, Y. *et al.* A physical catalyst for the electrolysis of nitrogen to ammonia. *Sci Adv* **4**, (2018).

96. Meessen, J. Urea synthesis. *Chemie Ingenieur Technik* **86**, 2180–2189 (2014).

97. Milne, N. *et al.* Functional expression of a heterologous nickel-dependent, ATP-independent urease in Saccharomyces cerevisiae. *Metab Eng* **30**, 130–140 (2015).

98. Zhao, J., Zhu, L., Fan, C., Wu, Y. & Xiang, S. Structure and function of urea amidolyase. *Biosci Rep* **38**, 1–12 (2018).

99. Trotter, P. J. *et al.* Glutamate dehydrogenases in the oleaginous yeast Yarrowia lipolytica. *Yeast* **37**, 103–115 (2020).

100. Fickers, P., Nicaud, J. M., Gaillardin, C., Destain, J. & Thonart, P. Carbon and nitrogen sources modulate lipase production in the yeast Yarrowia lipolytica. *J Appl Microbiol* **96**, 742–749 (2004).

101. Yang, P., Chen, Y. & Gong, A. dong. Development of a defined medium for Corynebacterium glutamicum using urea as nitrogen source. *3 Biotech* **11**, 1–10 (2021).

102. Ruiz-Herrera, J. & Sentandreu, R. Different effectors of dimorphism in Yarrowia lipolytica. *Arch Microbiol* **178**, 477–483 (2002).

103. Sun, L. *et al.* Divenn: An interactive and integrated web-based visualization tool for comparing gene lists. *Front Genet* **10**, 452359 (2019).

104. Forgac, M. Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. *Nat Rev Mol Cell Biol* **8**, 917–929 (2007).

105. Van Vuuren, H. J. J., Daugherty, J. R., Rai, R. & Cooper, T. G. Upstream induction sequence, the cis-acting element required for response to the allantoin pathway inducer and enhancement of operation of the nitrogen-regulated upstream activation sequence in Saccharomyces cerevisiae. *J Bacteriol* **173**, 7186–7195 (1991).

106. Navarathna, D. H. M. L. P., Das, A., Morschhäuser, J., Nickerson, K. W. & Roberts, D. D. Dur3 is the major urea transporter in Candida albicans and is co-regulated with the urea amidolyase Dur1,2. *Microbiology (N Y)* **157**, 270 (2011).

107. Program, B., Werpy, T. & Petersen, G. Top Value Added Chemicals from Biomass Volume I-Results of Screening for Potential Candidates from Sugars and Synthesis Gas Produced by the Staff at Pacific Northwest National Laboratory (PNNL) National Renewable Energy Laboratory (NREL) Office of Biomass Program (EERE) For the Office of the Energy Efficiency and Renewable Energy.

108. Devi, N. *et al.* Itaconic Acid and Its Applications for Textile, Pharma and Agro-Industrial Purposes. *Sustainability 2022, Vol. 14, Page 13777* **14**, 13777 (2022).

109. Krull, S., Hevekerl, A., Kuenz, A. & Prüße, U. Process development of itaconic acid production by a natural wild type strain of Aspergillus terreus to reach industrially relevant final titers. *Appl Microbiol Biotechnol* **101**, 4063–4072 (2017).

110. Hosseinpour Tehrani, H. *et al.* Integrated strain- And process design enable production of 220 g L-1 itaconic acid with Ustilago maydis. *Biotechnol Biofuels* **12**, 1–11 (2019).

111. Teleky, B. E. & Vodnar, D. C. Biomass-Derived Production of Itaconic Acid as a Building Block in Specialty Polymers. *Polymers 2019, Vol. 11, Page 1035* **11**, 1035 (2019).

112. Robert, T. & Friebel, S. Itaconic acid – a versatile building block for renewable polyesters with enhanced functionality. *Green Chemistry* **18**, 2922–2934 (2016).

113. Guo, B. *et al.* Biobased poly(propylene sebacate) as shape memory polymer with tunable switching temperature for potential biomedical applications. *Biomacromolecules* **12**, 1312–1321 (2011).

114. Raţă, D. M., Chailan, J. F., Peptu, C. A., Costuleanu, M. & Popa, M. Chitosan: poly(N-vinylpyrrolidone-alt-itaconic anhydride) nanocapsules—a promising

alternative for the lung cancer treatment. *Journal of Nanoparticle Research* **17**, 1–11 (2015).

115. Babić, M. M. *et al.* Evaluation of novel antiproliferative controlled drug delivery system based on poly(2-hydroxypropyl acrylate/itaconic acid) hydrogels and nickel complex with Oxaprozin. *Mater Lett* **163**, 214–217 (2016).

116. Boschert, D., Schneider-Chaabane, A., Himmelsbach, A., Eickenscheidt, A. & Lienkamp, K. Synthesis and Bioactivity of Polymer-Based Synthetic Mimics of Antimicrobial Peptides (SMAMPs) Made from Asymmetrically Disubstituted Itaconates. *Chemistry* **24**, 8217–8227 (2018).

117. Bajpai, S. K., Jyotishi, P. & Bajpai, M. Synthesis of nanosilver loaded chitosan/poly(acrylamide-co-itaconic acid) based inter-polyelectrolyte complex films for antimicrobial applications. *Carbohydr Polym* **154**, 223–230 (2016).

118. Walsh, T. J. *et al.* Experimental pulmonary aspergillosis due to Aspergillus terreus: pathogenesis and treatment of an emerging fungal pathogen resistant to amphotericin B. *J Infect Dis* **188**, 305–319 (2003).

119. Karaffa, L. & Kubicek, C. P. Citric acid and itaconic acid accumulation: variations of the same story? *Appl Microbiol Biotechnol* **103**, 2889–2902 (2019).

120. Karaffa, L. *et al.* A deficiency of manganese ions in the presence of high sugar concentrations is the critical parameter for achieving high yields of itaconic acid by Aspergillus terreus. *Appl Microbiol Biotechnol* **99**, 7937–7944 (2015).

121. Brefort, T. *et al.* Ustilago maydis as a Pathogen. *Annu Rev Phytopathol* **47**, 423–445 (2009).

122. Groenewald, M. *et al.* Yarrowia lipolytica: Safety assessment of an oleaginous yeast with a great industrial potential. *Crit Rev Microbiol* **40**, 187–206 (2014).

123. Konzock, O. & Norbeck, J. Deletion of MHY1 abolishes hyphae formation in Yarrowia lipolytica without negative effects on stress tolerance. *PLoS One* **15**, 1–11 (2020).

124. Tao, Z. *et al.* Yeast Extract: Characteristics, Production, Applications and Future Perspectives. *J Microbiol Biotechnol* **33**, 151–166 (2023).

125. Yang, L. B. *et al.* A novel osmotic pressure control fed-batch fermentation strategy for improvement of erythritol production by Yarrowia lipolytica from glycerol. *Bioresour Technol* **151**, 120–127 (2014).

126. Zhao, C. *et al.* Enhanced itaconic acid production in Yarrowia lipolytica via heterologous expression of a mitochondrial transporter MTT. *Appl Microbiol Biotechnol* **103**, 2181–2192 (2019).

127. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science (1979)* **297**, 1183–1186 (2002).

128. Raj, A. & van Oudenaarden, A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135**, 216–226 (2008).

129. Balázsi, G., Van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**, 910–925 (2011).

130. Hardo, G. & Bakshi, S. Challenges of analysing stochastic gene expression in bacteria using single-cell time-lapse experiments. *Essays Biochem* **65**, 67–79 (2021).

131. Farquhar, K. S. *et al.* Role of network-mediated stochasticity in mammalian drug resistance. *Nature Communications 2019 10:1* **10**, 1–14 (2019).

132. El Meouche, I., Siu, Y. & Dunlop, M. J. Stochastic expression of a multiple antibiotic resistance activator confers transient resistance in single cells. *Scientific Reports 2016 6:1* **6**, 1–9 (2016).

133. Pisco, A. O. *et al.* Non-Darwinian dynamics in therapy-induced cancer drug resistance. *Nature Communications 2013 4:1* **4**, 1–11 (2013).

134. Charlebois, D. A., Abdennur, N. & Kaern, M. Gene expression noise facilitates adaptation and drug resistance independently of mutation. *Phys Rev Lett* **107**, 218101 (2011).

135. Okabe, Y. & Sasai, M. Stable stochastic dynamics in yeast cell cycle. *Biophys J* **93**, 3451–3459 (2007).

136. Jovic, D. *et al.* Single-cell RNA sequencing technologies and applications: A brief overview. *Clin Transl Med* **12**, e694 (2022).

137. Jackson, C. A., Beheler-amass, M., Tj, A. & Gresham, D. Simultaneous estimation of gene regulatory network structure and RNA kinetics from single cell gene expression. *bioRxiv* 1–75 (2023).

138. Chen, H. I. H., Jin, Y., Huang, Y. & Chen, Y. Detection of high variability in gene expression from single-cell RNA-seq profiling. *BMC Genomics* **17**, 119–128 (2016).

139. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell* **9**, 3273–3297 (1998).

140. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A* **99**, 12795–12800 (2002).

141. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

142. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, (2021).

143. Cherry, J. M. *et al.* Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* **40**, D700 (2012).

144. Ramos-Alonso, L., Romero, A. M., Martínez-Pastor, M. T. & Puig, S. Iron Regulatory Mechanisms in Saccharomyces cerevisiae. *Front Microbiol* **11**, (2020).

145. Protchenko, O. *et al.* Three Cell Wall Mannoproteins Facilitate the Uptake of Iron in Saccharomyces cerevisiae. *Journal of Biological Chemistry* **276**, 49244–49250 (2001).

146. Morawska, L. P., Hernandez-Valdes, J. A. & Kuipers, O. P. Diversity of bet-hedging strategies in microbial communities-Recent cases and insights. *WIREs mechanisms of disease* **14**, (2022).

147. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nature Communications 2017 8:1* **8**, 1–12 (2017).

148. Jackson, C. A., Castro, D. M., Saldi, G. A., Bonneau, R. & Gresham, D. Gene regulatory network reconstruction using single-cell rna sequencing of barcoded genotypes in diverse environments. *Elife* **9**, 1–34 (2020).

149. Jariani, A. *et al.* A new protocol for single-cell RNA-seq reveals stochastic gene expression during lag phase in budding yeast. *Elife* **9**, 1–22 (2020).

150. Brabender, M., Hussain, M. S., Rodriguez, G. & Blenner, M. A. Urea and urine are a viable and cost-effective nitrogen source for Yarrowia lipolytica biomass and lipid accumulation. *Appl Microbiol Biotechnol* **102**, 2313–2322 (2018).

151. Nadal-Ribelles, M., Solé, C., de Nadal, E. & Posas, F. The rise of single-cell transcriptomics in yeast. *Yeast* **41**, 158–170 (2024).

152. Payne, S. H. The utility of protein and mRNA correlation. *Trends Biochem Sci* **40**, 1 (2015).

153. Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between Protein and mRNA Abundance in Yeast. *Mol Cell Biol* **19**, 1720 (1999).

154. Lahtvee, P. J. *et al.* Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast. *Cell Syst* **4**, 495-504.e5 (2017).

155. Mayr, C. What Are 3′ UTRs Doing? *Cold Spring Harb Perspect Biol* **11**, (2019).

156. Leppek, K., Das, R. & Barna, M. Functional 5′ UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews Molecular Cell Biology 2017 19:3* **19**, 158–174 (2017).

157. Ramazi, S. & Zahiri, J. Post-translational modifications in proteins: resources, tools and prediction methods. *Database (Oxford)* **2021**, (2021).

158. Chen, J., Guo, M., Wang, X. & Liu, B. A comprehensive review and comparison of different computational methods for protein remote homology detection. *Brief Bioinform* **19**, 231–244 (2018).

159. Rost, B. Twilight zone of protein sequence alignments. *Protein Engineering, Design and Selection* **12**, 85–94 (1999).

160. Zielezinski, A., Vinga, S., Almeida, J. & Karlowski, W. M. Alignment-free sequence comparison: benefits, applications, and tools. *Genome Biology 2017 18:1* **18**, 1–17 (2017).

161. Bordin, N. *et al.* Novel machine learning approaches revolutionize protein knowledge. *Trends Biochem Sci* **48**, 345 (2023).

162. Škunca, N., Altenhoff, A. & Dessimoz, C. Quality of Computationally Inferred Gene Ontology Annotations. *PLoS Comput Biol* **8**, 1002533 (2012).