Universitat Autònoma de Barcelona

Systems metabolic engineering for recombinant protein production in Pichia pastoris

Sergi Monforte Mercado

Universitat Autònoma de Barcelona

Escola d'Enginyeria

Departament d'Enginyeria Química, Biològica i Ambiental

Systems metabolic engineering for recombinant protein production in *Pichia pastoris*

Memòria per a optar al grau de Doctor

per la Universitat Autònoma de Barcelona,

sota la direcció del Dr. Pau Ferrer i el Dr. Joan Albiol

per

Sergi Monforte Mercado

Bellaterra, setembre 2019

El Dr. Joan Albiol Sala i el Dr. Pau Ferrer Alegre, ambdós professors agregats del Departament d'Enginyeria Química, Biològica i Ambiental, i membres del grup de recerca d'Enginyeria de Bioprocessos i Biocatàlisi Aplicada de la Universitat Autònoma de Barcelona.

Certifiquem:

Que el biotecnòleg Sergi Monforte Mercado ha dut a terme al Departament d'Enginyeria Química, Biològica i Ambiental de la Universitat Autònoma de Barcelona i amb la nostra direcció la tesi doctoral titulada **"Systems metabolic engineering for recombinant protein production in *Pichia pastoris*"**. La mateixa es presenta en aquesta memòria i constitueix el manuscrit per optar al Grau de Doctor en Biotecnologia per la Universitat Autònoma de Barcelona.

I per tal que se'n prengui coneixement i consti als efectes oportuns, signem la present declaració a Bellaterra, a 25 de setembre de 2019.

Dr. Joan Albiol Sala	Dr. Pau Ferrer Alegre	Sergi Monforte Mercado
(co-director)	(co-director)	(Autor)

Abstract

The methylotrophic yeast *Pichia pastoris* (*Komagataella sp.*) is one of the most attractive expression systems for heterologous protein production, which constitutes a continuously expanding market. The strong alcohol oxidase gene 1 promoter (P_{AOX1}), induced by methanol but repressed by glucose, glycerol or ethanol, is one of the most used for this purpose. Nevertheless, there still exist several physiological bottlenecks limiting the process.

In this context, several strategies have been proposed and tested in order to improve the heterologous production of many different types of proteins. Common approaches include increasing heterologous gene copy number, promoter engineering and modification of the folding and secretory mechanisms. The aim of this thesis has been the development of new strategies to increase recombinant protein yields, using the *Rhizopus oryzae* lipase (Rol) as model protein in a P_{AOX1}-based expression system.

Firstly, the P_{AOX1} transcription factor genes *MXR1* and *MIT1* were constitutively overexpressed aiming at improving *ROL* transcription. This was confirmed by an improved methanol assimilation capacity and an increase in relative mRNA levels of *ROL* and several genes related with methanol metabolism, i.e. reverting the titration effect caused by the transcription of multiple *ROL* expression cassettes. Despite such improvements, extracellular lipase activity levels did not increase significantly in chemostat cultures, pointing out to additional bottlenecks limiting Rol production.

Second, possible metabolic engineering targets in *P. pastoris*' cell metabolism were explored using the consensus genome-scale metabolic model (GEM) iMT1026 v3.0. This *in silico* step provided several promising knock-outs which were going to be experimentally tested using the CRISPR/Cas9 genome editing system. The simulations pointed to NADPH availability and limited supply of some amino acids (serine and cysteine) as potential Rol production limiting factors. A reduction in cell fitness affecting the viability of the obtained strains impeded to verify most of the proposed knock-outs.

Finally, since our *in silico* analyses and previously published studies identified NADPH as an important limiting cofactor in recombinant protein production, our efforts were geared towards increasing its availability through gene knock-in strategies. Specifically, we overexpressed two genes encoding redox enzymes, a NADH kinase and a NADH oxidase, with the aim to directly perturb the cell's redox balance. Further, we tested the physiological effect of these enzymes using different co-substrate/methanol mixtures as carbon source. In short, we observed an increase in recombinant protein production with different degrees of improvement depending on the carbon source(s) tested. We also performed a transcriptomic analysis and an *in silico* evaluation of our results in order to provide a better interpretation of the cell physiological state. To our knowledge, this is the first study aiming to increase NADPH generation in the P_{AOX1}-based expression system, under methanol growth conditions.

Overall, novel strain engineering strategies have been proposed and tested during the execution of this study. Furthermore, GEMs and related systems biology approaches were applied, proving to be promising powerful tools for rational engineering of industrial microorganisms.

Ressenya

El llevat metilotròfic *Pichia pastoris* (*Komagataella sp.*) és un dels sistemes d'expressió més atractius per a la producció de proteïna recombinant, mercat contínuament en expansió. El fort promotor del gen de l'alcohol oxidasa 1 (P_{AOX1}), induït per metanol però reprimit per glucosa, glicerol o etanol, és un dels més emprats per aquest propòsit. No obstant, existeixen encara diversos colls d'ampolla fisiològics que limitant el procés.

En aquest context, diferents estratègies han estat proposades i provades per tal de millorar la producció heteròloga de molts tipus diferents de proteïna. Les aproximacions més habituals inclouen increment en el nombre de còpies de gen heteròleg, enginyeria de promotors i modificació dels mecanismes de plegament i secreció. L'objectiu d'aquesta tesi ha estat el desenvolupament de noves estratègies per incrementar la producció de proteïna recombinant, emprant la lipasa de *Rhizopus oryzae* (Rol) com a proteïna model en el sistema d'expressió basat en el P_{AOX1}.

Primerament, els gens del factors de transcripció de P_{AOX1}, *MXR1* i *MIT1*, es van sobreexpressar constitutivament per tal de millorar la transcripció de *ROL*. Això es va confirmar degut a una millora en la capacitat assimilatòria de metanol i un increment en els nivells relatius de mRNA de *ROL* i varis gens relacionats amb el metabolisme del metanol, i.e. revertint l'efecte de titulació causat per la transcripció de múltiples *cassettes* d'expressió de *ROL*. Tot i aquestes millores, els nivells extracel·lulars d'activitat lipàsica no van augmentar de forma significativa en cultius en quimiòstat, apuntant a colls d'ampolla addicionals limitant la producció de Rol.

En segon lloc, es van explorar possibles dianes d'enginyeria metabòlica en el metabolisme cel·lular de *P. pastoris* emprant el model metabòlic a escola genoma (GEM) consens iMT1026 v3.0. Aquest pas *in silico* va proporcionar diversos knock-outs prometedors que serien experimentalment testats fent servir el sistema d'edició genòmica CRISPR/Cas9. Les simulacions apuntaven a la disponibilitat de NADPH i una limitada aportació de determinats aminoàcids (serina i cisteïna) com a potencials factors limitants de la producció de Rol. Una reducció en el fitnes cel·lular que

afecta a la viabilitat de les soques que es buscaven obtenir va impedir la verificació de la majoria dels knock-outs proposats.

Finalment, donat que les nostres anàlisis i estudis prèviament publicats identificaven el NADPH com un cofactor important limitant la producció de proteïna recombinant, els nostres esforços es van dirigir a incrementar la seva disponibilitat a través d'estratègies de knock-in de gens. Específicament, vam sobreexpressar dos gens que codificaven per enzims redox, una NADH quinasa i una NADH oxidasa, amb l'objectiu de pertorbar directament l'equilibri redox de la cèl·lula. A més, es va comprovar l'efecte fisiològic d'aquests enzims fent servir diferents mescles co-substrat/metanol com a font de carboni. En resum, vam observar un increment en la producció de proteïna recombinant amb diferents graus de millora depenent de la font de carboni provada. També vam realitzar anàlisis transcriptòmiques i una avaluació *in silico* dels nostres resultats per tal de presentar una interpretació millor de l'estat fisiològic de la cèl·lula. Dins del nostre coneixement, aquest és el primer estudi dirigit a incrementar la generació de NADPH en un sistema d'expressió basat en P_{AOX1}, en condicions de creixement en metanol.

A grans trets, noves estratègies d'enginyeria de soques han estat proposades i provades durant l'execució d'aquest estudi. A més a més, s'han aplicat GEMs i aproximacions relacionades amb biologia de sistemes, demostrant que són eines potents i prometedores per al disseny racional d'organismes industrials.

Table of contents

1. Introduction

1.1. Yeasts in the recombinant protein market scenario

The European Federation of Biotechnology (EFB) defines biotechnology as "the integration of natural science and organisms, cells, parts thereof, and molecular analogues for products and services", that is, obtaining goods and services using biological tools. Therefore, the origins of biotechnology back to the first attempts to bake bread and brew alcoholic beverages, but recent developments in molecular biology have led to the arisen of the term "modern biotechnology". This new designation has applications far beyond breeding or fermentation and involves genetic engineering and cell manipulation.

Since 1980s, a major focus of modern biotechnology has been the development of key technologies enabling recombinant protein production, which allows for a far more interesting alternative than protein extraction from natural sources. Natural sources of proteins are not always available in big amounts and their extraction is a costly and low profitable process. The first products obtained by means of recombinant DNA techniques were biopharmaceuticals (insulin, interferons…) and industrial enzymes (used for food treatment, detergents…). The recombinant protein market has been continuously expanding since then. A recent study valued it around 1645 million US dollars in 2017 and is expected to grow up to 2850.5 million dollars in 2022 [1].

Concerning enzyme industry, in the 20th century there was an increasing interest for enzymes over chemical catalysts due to the rise of recombinant DNA technology. Enzymes are used in the chemical industry and other industrial applications when very specific catalysts are required but as disadvantage they have a limited number of reaction cycles until their efficiency drops [2]. At present, a 60% of all the industrial enzymes are produced heterologously due to its higher production yield and the chance to modify the natural enzyme to improve its activity and stability [3], [4].

Over the years, yeast-based expression systems have become one of the most frequently employed organisms for recombinant protein production at industrial scale. Microbial eukaryotic hosts can combine the high growth rates in high-scale processes

of prokaryotic systems with the advanced protein processing machinery of mammalian and insect cell lines. Classically, the yeast *Saccharomyces cerevisiae* has been the main choice for this kind of processes but, in the last 20 years, other yeasts such as *Pichia pastoris (Komagataella spp), Hansenula polymorpha, Kluyveromcyes lactis* and *Yarrowia lipolytica* have been established as alternative yeast cell factory platforms [5].

Production of enzymes using *P. pastoris* has been employed for a wide variety of industrial applications. As feed additive, it has been proved that *P. pastoris* can produce phytases with higher specific activity than those produced by *Aspergillus niger*, *H. polymorpha* and *S. cerevisiae* [6] (other enzymes used in feed and food industry and successfully expressed in *P. pastoris* can be reviewed in [7]). *P. pastoris* has been also used in the production of laccases which can be utilized in wastewater treatment and bleaching in textile industry [8]–[10]. Hydrolytic enzymes such as cellulases and xylanases, highly used for treatment of lignocellulosic materials, have been successfully expressed in this host as well [11]. Finally, lipases, which are used for synthesis of aromatic compounds and enantioselective synthesis of fine chemicals, are one of the enzymes more extensively studied and optimized to be produced in *P. pastoris* [12]–[14].

In summary, it is expected that interest in the utilization of yeasts for recombinant protein production will further increase. To this end, knowledge about cell physiology and new genetic engineering technologies and strategies for strain improvement are steadily growing, accelerating the engineering of strains and production processes with increased yields and productivities and, ultimately, leading to a major market share of protein-based products.

1.2. Expression hosts for recombinant protein production

The selection of the most suitable expression system to produce a given protein at large scale is a critical step. This choice depends on many features of the target protein and process (productivity, bioactivity, purpose and physicochemical characteristics). All hosts have their own advantages and limitations and a compromise solution between product quantity and quality is usually unavoidable (**Table 1**).

Among the different hosts that are used for recombinant protein production, the gram-negative prokaryotic bacterium *Escherichia coli* is extensively used due to its well-known methods for genetic manipulation, high growth rates and cheap media. Also, *Bacillus subtilis* is the most known gram-positive bacterium for this purpose. However, unless we are considering very simple proteins, production in these prokaryotic cell factories could probably demand an additional *in vitro* process for the insertion of post-translational modifications. Generally, they are not the most suitable hosts when glycosylation, chemical modifications or proteolytic processing are required [15].

Table 1. Comparison of features among different industrial expression hosts.

Host	Production cost	Production time	Product quality	Scaling up	Glycosylation	Risk of pathogenicity
Bacteria	Low	Short	Low	Scalable	No	Medium
Yeast	Medium	Medium	Medium	Scalable	Yes (unusual)	Low
Mammalian cells	High	Long	Very high	Hard to scale up	Yes (human-like)	High
Insect cells	High	Long	High	Non-scalable	Yes (minor differences)	Medium

When prokaryotic hosts are not able to produce a protein with the desired biological activity or stability, eukaryotic hosts generally represent the alternative solution. In these cases, for proteins that do not require highly complex posttranslational modifications, yeasts are employed over higher eukaryotic systems due to its operational simplicity, fast growth and high expression levels. As previously stated, *Saccharomyces cerevisiae* has been the reference yeast for recombinant protein production for several decades but other yeasts such as *Pichia pastoris* are well established at present and are gaining increasing industrial relevance as powerful and cheap heterologous systems, as reflected in the large number of currently available products that are being produced using these alternative yeast platforms [16]. As a Crabtree-negative yeast, *P. pastoris* can achieve higher recombinant protein levels

than *S. cerevisiae,* since under non-limiting oxygen conditions there are not product yield losses due to ethanol production. Some of the main characteristics of *P. pastoris* as host for recombinant protein production include ease of high cell-density cultivation, high levels of productivity in defined and protein-free media, highly efficient heterologous protein secretion (although not at the same level than some animal cell lines [17]), well-established genetic manipulation tools, capacity to perform posttranslational modifications and GRAS (generally recognized as safe) status.

When the protein of interest has particularly high levels of complexity, yeasts may not be capable of correctly performing the required post-translational modifications. For example, mammalian cells are the most suitable election for properly folded and post-translationally modified glycoproteins with complex (e.g. human) glycans. This is particularly important in e.g. heterologous proteins that must replicate biological activities performed by their corresponding native counterparts in human cells, since the highest levels of quality and fidelity are minimal requisites for clinical applications [18]. However, production costs of mammalian cells systems are very high due to slow growth rates and expensive media. For this reason, it is important to define in advance a balance between protein quality and price. The most common cell lines utilized for recombinant protein production are HEK 293 (Human embryonic kidney) and CHO (Chinese hamster ovary).

Insect cells used in conjunction with the baculovirus expression vector system (BEVS) are expanding very quickly as hosts for recombinant protein production, competing with mammalian cell systems. They have some advantages over mammalian cells such as easiness of cultivation, higher resistance to changes in osmolality and by-product concentration, higher duplication rate and superior expression level when used altogether with a recombinant baculovirus. Nevertheless, their post-translational modifications are not as human-like as those of mammalian cell lines and down-stream process has some limitations due to separation of the virus-derived fraction [19]. Currently, the most widely used insect cell lines are High Five and Sf9.

1.3. Pichia pastoris as emerging cell platform for recombinant protein production

1.3.1. Historical overview

P. pastoris is a methylotrophic yeast that has become increasingly popular as a cell factory for the production of recombinant proteins over the past two decades. Its story in the biotechnological field began 40 years ago, when Phillips Petroleum used high cell density fermentations of this microorganism as animal feed additive, using methanol as carbon source. In the 80s, *P. pastoris* took its first steps as heterologous protein production host using the strong and tightly regulated *AOX1* promoter (P_{AOX1}) [20]. This allowed to launch the first industrial process to produce an industrial enzyme (hydroxynitrile lyase) in the 90s [21].

Once this expression system, initially patented by Phillips Petroleum, was made available to the research community, the characterization and know-how of this cell factory platform expanded rapidly. One of the major milestones achieved was the publication of detailed genome sequences of 3 different *P. pastoris* strains: the collection strains CBS7435 and DMSZ 70382 [22], [23], and strain GS115 (derived from CBS7435) [24]. Meanwhile, *P. pastoris* had been reclassified into the genus *Komagataella* [25]. Although we commonly refer to this host as *P. pastoris*, this name is actually used as a synonym of two different *Komagataella* species (*K. pastoris* and *K. phaffii*)) [26].

Equally important for the advent of *P. pastoris* for commercial applications was the achievement of the GRAS status by the Food and Drug Administration (FDA) and the first consent of a recombinant biopharmaceutical product using *P. pastoris*, Kalbitor® (inhibitor of the protein kallikrein used for the treatment of hereditary angioedema) [27].

1.3.2. Pichia pastoris as recombinant protein production system

P. pastoris, as a methylotrophic yeast, can assimilate methanol for growth and use it for energy production. The enzymes used for methanol metabolism are specifically compartmentalized in the peroxisomes and their expression is tightly regulated by the

presence of methanol in the medium. The genes of the enzymes catalyzing the first step of methanol metabolism, alcohol oxidase 1 and 2 (*AOX1, AOX2*), had the first promoters with this type of regulation studied (P_{AOX1}, P_{AOX2}) [28]. Due to its transcriptional efficiency and regulatory features, P_{AOX1} is the inducible promoter that has been more extensively used for recombinant protein production in this host, since it has around 10-fold higher transcription levels than its counterpart (P_{AOX2}) [29]. Later, the promoter of the gene of the glutathione-dependent formaldehyde dehygrogenase (*FLD1*), inducible by methanol or methylamine, was proposed as an attractive alternative to P_{AOX1} [30]. However, P_{AOX1} is still the most extended inducible promoter for recombinant protein production in *P. pastoris*. The main advantage of this system is the achievement of very high product yields using a carbon source characterized by its low cost and acquisition from renewable sources. Conversely, the maintenance of the cultivation operational conditions at high cell densities is quite problematic: methanol oxidation is not energetically efficient and very powerful heat dissipation devices are required. Moreover, the oxygen consumption rates using this carbon source are extremely high, and an extra supply of pure oxygen is often required [31]. These drawbacks make the use of this carbon source difficult from an operational point of view at industrial scale. In addition, logistically, the storage and delivery of an inflammable and toxic product as methanol is also problematic [32]. Following these limitations, there have been substantial efforts to create alternative expression systems for *P. pastoris*.

To this end, the first methanol-free expression system developed was based on the utilization of the glyceraldehyde-3-phosphate dehydrogenase gene (TDH3) promoter (P_{GAP}). *TDH3* is an essential gene involved in glycolysis and has a strong and constitutive expression when glucose or glycerol is used as carbon source [21], [33]. Great efforts were also made in genetic manipulation in order to obtain methanol-free P_{AOX1} strains (e.g. by removal of gene repressors and overexpression of activators). This strain allows to replace the classical glycerol/methanol shift by a glucose/glycerol shift in production processes [34]. Another interesting approach for development of methanol free systems included direct engineering of P_{AOX1}, obtaining a promoter whose activation was dependent on removal of glycerol (the repressing carbon source)

[35]. Finally, new promoters whose expression was repressed in glycerol and activated in glucose were discovered by DNA microarray analysis. Between the promoters reported, the one of the high-affinity glucose transporter gene *GTH1* (P_{GTH1}) excels due to its particularly high expression levels [36]. This promoter was engineered to obtain even higher expression levels in a recent study [37]. In summary, although there is an increasing number of alternatives for recombinant protein production in *P. pastoris* due to the new advances in synthetic biology, P_{AOX1} is still widely used as expression system.

In summary, *P. pastoris* presents some specific advantages and drawbacks compared to conventional yeasts such as *S. cerevisiae*:

- **Space-time product yields**. *P. pastoris* can achieve higher levels of correctly folded protein in unusually high-density cultures (HDC) (more than a hundred grams per liter). Nevertheless, it is important to keep in mind that HDC require an exceptionally robust monitoring and control of process parameters [38].
- **Fermentative phenotype**. Ethanol generation in yeasts has been described as a response to a change in external concentration of carbon source (Crabtree effect) or difference in the aeration levels (Pasteur effect). *S. cerevisiae* is classified as Crabtree-positive and produces ethanol (a toxic compound) at high cell-densities. Contrary, *P. pastoris* is Crabtree-negative, which means is not sensitive to variations in substrate concentration and it has a non-fermentative phenotype under aerobic conditions [39]. This increases process robustness and can reduce the harvest time in cultures such as fed-batch. Overall, the existing overflow to ethanol generation in *S. cerevisiae* derives in a less efficient carbon source usage for recombinant protein production.
- **High growth rate on alternative C-sources such as glycerol**. *S. cerevisiae* exhibits a poor growth on glycerol (carbon source of high commercial interest due to the large amounts of crude glycerol waste generated in biodiesel production) unless growth supporting supplements (amino acids, nucleobases or complex supplements) are added in culture medium [40]. Moreover, even in supplemented media, *P. pastoris* shows superior growth rates [41].

- **Genetic manipulation.** *P. pastoris* is not as genetically manageable as *S. cerevisiae*. For heterologous gene expression in *S. cerevisiae* there are both, replicative (low or high copy) or integrative plasmid vectors. Conversely, it is difficult to maintain episomal plasmid stability in *P. pastoris* [20]. For this reason, most of the expression vectors available for this host are designed to be integrative elements in its genome.

  Due to a more active non-homologous end joining (NHEJ) mechanism compared to homologous recombination (HR), specific integration events of foreign DNA in *P. pastoris* occur with less frequency than in *S. cerevisiae.* This leads to lower efficiencies of gene knock-in. In addition, in gene knock-outs, where double recombination is required for integration of the disruption cassette, the DNA reparation mechanisms of *P. pastoris* hinders even further the obtention of positive clones. For this reason, new tools with higher rate of successful genetic modification, such as the CRISPR/Cas9 system, are being adapted for *P. pastoris*.
- **Glycosylation.** Although both *S. cerevisiae* and *P. pastoris* are capable of glycosylate proteins, the glycosylation pattern is different [42]. *P. pastoris* tends to generate a less hyperglycosylated pattern (20 vs 50-150 residues length of N-linked oligosaccharides). Particularly, although both yeast species have an hyperglycosylated pattern with high content of mannose glycans, *P. pastoris* adds mannose residues to a lesser extent.
- **Protein secretion**. Both yeast expression systems very often make use of the α-factor prepro-peptide from *S. cerevisiae* for heterologous protein secretion. Since *P. pastoris* has very low levels of secreted endogenous protein [23], making easier protein purification from supernatant, it is a very attractive alternative for extracellular recombinant protein production.

1.4. Methanol metabolism in *P. pastoris*

As previously explained, *P. pastoris* is a methylotrophic yeast and can use methanol as sole carbon source. Methanol utilization requires a set of distinctive enzymes [43] which were firstly identified in *P. pastoris* by Courderc & Baratti [44]. The initial reactions occur in a specialized organelle called peroxisome. The methanol

compartmentalization in the peroxisome is crucial due to the high toxicity of this molecule. Moreover, another reason why methanol metabolism needs to be compartmentalized is the generation of reactive species, such as hydrogen peroxide (H_2O_2), in this pathway.

Firstly, methanol is introduced in the peroxisome and oxidized to formaldehyde and H_2O_2 by the enzyme alcohol oxidase. The resultant H_2O_2 (a toxic compound) is converted to oxygen and water by a catalase. The formaldehyde may follow two different destinations, which take place in different cellular compartments. Formaldehyde can leave the peroxisome and be oxidized to formate and carbon dioxide by two different cytosolic NAD^+-dependent dehydrogenases. These reactions constitute the dissimilatory pathway of methanol metabolism and are a source of energy for *P. pastoris*. The formaldehyde that remains in the peroxisome can enter the assimilatory pathway for biomass generation. Briefly, formaldehyde reacts with xylulose 5-phosphate in a reaction catalyzed by the enzyme dihydroxyacetone synthase (DAS), generating glyceraldehyde 3-phosphate and dihydroxyacetone. These products may leave peroxisome and enter the glycolytic pathway or are partially recycled for xylulose 5-phosphate formation in the peroxisome. A model proposed by Ruβmayer et al. [45] indicates that isoforms of enzymes involved in the non-oxidative route of the pentose phosphate pathway would produce the sugar recycling in the peroxisome (**Figure 1**).

Principally, the expression of methanol utilization (MUT) genes is strongly activated by methanol and repressed by glucose, ethanol and glycerol (reviewed in [46]). While this true for dihydroxyacetone synthase (*DAS*) and *AOX* expression, formaldehyde dehydrogenase (*FLD*) and formate dehydrogenase (*FDH*) genes can also be induced by methylamine or choline, even in the presence of glucose. More detailed information about P_{AOX1} regulation is provided in section 5.2. Since many reactions of the MUT pathway occur in the peroxisome, these organelles proliferate upon methanol induction.

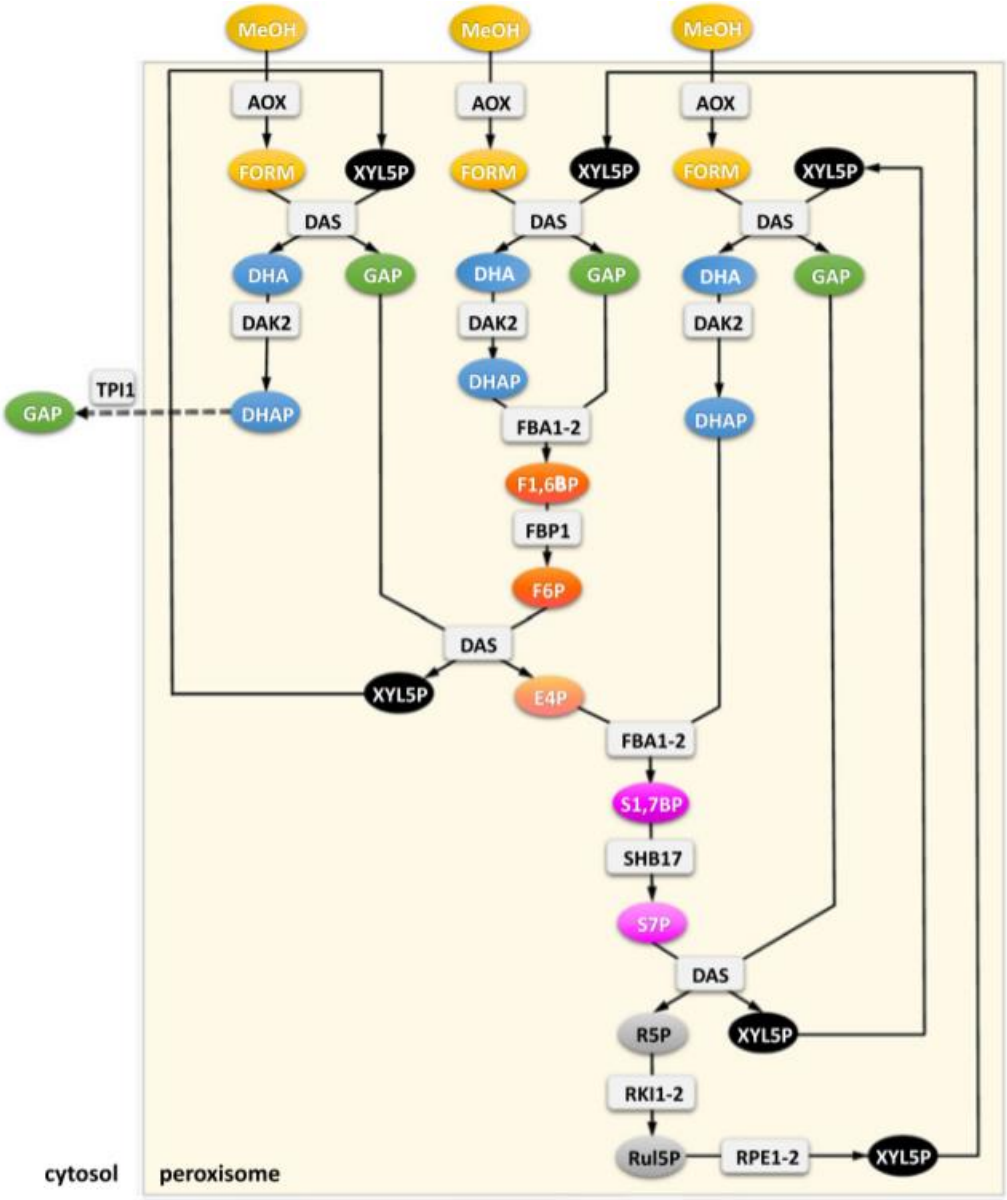


Figure 1. Model proposed by Ruβmayer and coworkers for regeneration of pentoses phosphate in the peroxisome. Taken from [45].

1.5. Using P_{AOX1}-based system for recombinant protein production

1.5.1. Methanolic phenotypes for the P_{AOX1}-based expression system

The highly expressed and tightly regulated P_{AOX1} is one of the most suitable choices for recombinant protein production. Its strong activation upon the presence of methanol and repression on glycerol or glucose is particularly useful when the protein product is known to be toxic for the cell, since growth and production can be uncoupled.

As previously indicated, despite its economical advantages, methanol utilization as carbon source carries important operational (hard temperature control and oxygen supply) and logistic (use of a highly flammable chemical) drawbacks. As a solution to minimize the amount of methanol used during fermentations, mutant strains of *P. pastoris* with genetic modifications in the *AOX* genes were created. Currently, three phenotypes of methanol metabolism have been established for the utilization of P_{AOX1} in *P. pastoris* [28]:

- Mut^+ (methanol utilization plus): both *AOX* genes are intact and active. The strain grows optimally on methanol.
- Mut^S (methanol utilization slow): only *AOX1* is knocked out. The strain has a slow growth phenotype on methanol. The reduced methanol consumption rate limits heat generation while maintaining similar heterologous protein expression levels.
- Mut^- (methanol utilization minus): both *AOX* genes are knocked out. The strain has almost negligible growth on methanol as sole carbon source. Consequently, even lower amounts of methanol are required to promote expression than Mut^S strain [47].

All three phenotypes grow equally on carbon sources other than methanol, but high-cell density cultivation strategies for each one of these strain phenotypes differ in the way methanol is employed as expression inducer. On one side, cultivation strategies for wild type Mut phenotype strains usually involve an initial growth batch phase using glycerol as carbon source followed by a switch to methanol feeding to trigger recombinant protein production. On the other side, for Mut^S and Mut^- it is a common practice to use a mix of methanol with a co-substrate for P_{AOX1} induction, since using methanol as sole carbon source would be too time-consuming (Mut^S) or even not viable (Mut^-).

Several studies have been performed to investigate the effect of these mutations on P_{AOX1}-driven heterologous gene expression levels, mainly focusing in comparative analyses between Mut^S and Mut^+ phenotypes. For intracellular protein production, it is widely stablished to employ the Mut^S variant to avoid the presence of the alcohol

oxidase 1 protein. For protein purification from extracellular medium, strain choice between Mut^+ and Mut^S is not so clear. Studies regarding yields and production level comparisons between phenotypes are not unanimous and the results seem to be strictly dependent on your protein and process conditions [48]–[50].

1.5.2. P_{AOX1} expression regulation

It has been already stated that the regulation of this expression system is completely linked to the type and concentration of the carbon source(s) of the culture medium. Empirical approaches have shown that P_{AOX1} needs two conditions to be fully induced: depletion or very low concentrations of repressing carbon sources and presence of methanol in the medium. This regulatory mechanism is therefore divided into three states: repression, derepression and induction. Expression remains suppressed as long as carbon sources such as glycerol or glucose are present in the medium. Derepression starts once this carbon sources have been almost totally consumed (expression levels in this state are still very low). Finally, feeding methanol in the medium triggers induction at its maximum capacity [51], [52], [53]. Although here and in almost all the works published in this field methanol is referred as the inducer of P_{AOX1} (what is correct from an operational point of view), it is important to remark that some studies have indicated that formate and formaldehyde, metabolites produced during methanol metabolism, are the direct inducers of this promoter [54], [55].

Operational procedures to control induction by methanol in high-cell density cultivations have been well-established since the 90s. However, the understanding of the molecular mechanisms involved in this regulatory process have been only achieved during past recent years. In order to identify the regulatory sequences of P_{AOX1}, several studies using sequential deletion analysis have been performed [53], [56]–[58]. The *cis*-acting elements identified using sequence-based approaches provided important suggestions about which *trans*-acting molecules could be involved in the regulatory mechanism. Initially, Lin-Cereghino and co-workers discovered the methanol expression regulator 1 (Mxr1), the first transcription factor known to be related with P_{AOX1} regulation [53]. Deletion studies of *MXR1* confirmed that Mxr1 is not only regulator of *AOX1*, but several other MUT (Methanol Utilization) genes. Later, Kranthi and co-workers revealed that there are six Mxr1 binding sites (MXRE, MXR response

elements) in P_{AOX1} sequence, being MXRE3 and 4 the ones with higher affinity [59]. More recently, Parua and coworkers demonstrated that 14-3-3 proteins (regulatory molecules involved with many signaling proteins in eukaryotes) are involved in the regulation of Mxr1 [60]. The interaction between these two proteins (which is triggered when ethanol or other carbon sources are present) inhibits expression of genes regulated by Mxr1, probably by avoiding recruitment of RNA polymerase II.

Subsequently, other transcription factors have been discovered. Takagi and co-workers identified a novel transcription factor named positive regulator of methanol (Prm1) [61]. *PRM1* deficient strains exhibited null growth when methanol was used as sole carbon source. In addition, transcriptional analysis revealed around a 50-fold decrease in P_{AOX1} expression levels in this mutant strain, pointing out to an activation role of this protein. At that point, the mechanism by which P_{AOX} was induced remained unclear, since Prm1p did not seem to directly interact with the promoter sequence.

Recently, Wang and co-workers described another transcription factor essential for a functional methanol metabolism in *P. pastoris.* This protein, named methanol induced transcription factor (Mit1), was found to directly interact with P_{AOX1} without competition with Mxr1 [62]. The study also proposed the first regulatory model of P_{AOX1} when *P. pastoris* is growing on methanolic conditions. In this model, methanol triggers Mxr1 transport from cytoplasm to the nucleus if there is no glucose in the medium, since it acts as repressor of this transport. Once in the nucleus, this transcription factor would act as a derepressor of P_{AOX1}. Concurrently, methanol addition starts a signaling cascade that involves Prm1p and Mit1. Prm1p induces its own and Mit1 gene transcription, while Mit1 also represses *PRM1* to avoid Prm1p accumulation.

In addition to transcriptional activators, some repressors of P_{AOX1} have been discovered. The repressor of phosphoenolpyruvate carboxykinase (ROP) is a zinc finger protein that negatively regulates expression of phosphoenolpyruvate carboxykinase gene (*PEPCK*) in conditions of methanol and biotin starvation. Kumar and co-workers discovered that ROP shares DNA sequence specificity with Mxr1 and has a role in P_{AOX1} regulation [63]. Overexpression and deletion experiments with ROP gene concluded

that it acts as a repressor in rich medium with methanol, competing with Mxr1p for DNA binding. Interestingly, ROP does not seem to act when cells are grown in minimal medium (another example of the importance of substrates in P_{AOX1} regulation).

Other P_{AOX1} transcriptional down-regulators more tightly related with catabolic repression have also been described. For instance, the role of Mit1 is not limited to methanolic activation of P_{AOX1}, but also an antagonistic function as repressor when glycerol is in the medium. In the same line, Zhang and co-workers discovered a new mechanism involved in glucose repression [64]. The hexose transporter Htx1, whose gene is transcriptionally regulated by the presence of glucose, acts as a catabolite repressor of P_{AOX1} in response to this carbon source. Briefly, by deleting this transporter, P_{AOX1} was temporarily induced without methanol. However, later pexophagy activation events led to a drop of P_{AOX1} expression. Further, another transcriptional repressor, named Nrg1, was identified and characterized by Wang and coworkers [65]. This zinc finger acts in the presence of both, glycerol and glucose. It can bind to five regions of P_{AOX1} and it has been hypothesized to be a competitor of Mxr1p. More recently, the two last P_{AOX1} repressors known so far were identified (Mig1 and Mig2). They were found by BLAST search using its homologous ScMig1 and ScMig2 from *S. cerevisiae* and by combinatorial strain engineering it was demonstrated that these regulators are implied in glycerol repression of P_{AOX1}.

All the knowledge available about P_{AOX1} regulatory mechanism has been mostly obtained during the last decade (summarized in **Figure 2**). This has opened the door to the design of re-wiring strategies of such mechanism for different cell engineering purposes. For instance, the most recent efforts have focused on genetic modifications aiming to obtain methanol-free *AOX*-based expression systems. Firstly, Wang and co-workers created a *P. pastoris* (Δ*MIG1*, Δ*MIG2*, Δ*NRG1,* P_{GAP}-*MIT1*) strain, called MF1, and proposed a novel induction strategy for recombinant protein production involving a glucose-glycerol shift instead of the standard glycerol-methanol transition [66]. Later on, Vogl and co-workers indicated that the sole overexpression of either *MXR1* or *MIT1* enables derepression of P_{AOX1} [67]. The usage of the peroxisomal catalase 1 gene promoter (P_{CAT}), which is repressed by glucose and derepressed once this is depleted, to promote *MXR1* and *MIT1* expression provided a new strategy to regulate P_{AOX1}-

driven recombinant protein production using a single carbon source. Finally, the most recent approach known so far came from Chang and co-workers, who implemented a synthetic positive feedback of Mxr1 expression by adding an extra copy of this gene under control of P_{AOX2} [68]. This allowed a smoother transition phase between glycerol and methanol. In addition, this mutant strain exhibited higher recombinant protein production levels under methanolic conditions and was able to consistently produce recombinant protein under glycerol starvation. This last discovery provides another potential alternative to induce recombinant protein production under methanol-free conditions. In short, three new ways to operationally exploit P_{AOX1} (glucose-glycerol, glucose-starvation and glycerol-starvation transitions) have been discovered in a narrow time window. Consequently, there is no reason to doubt that there is still room for discovery of new applications inside this field of research.

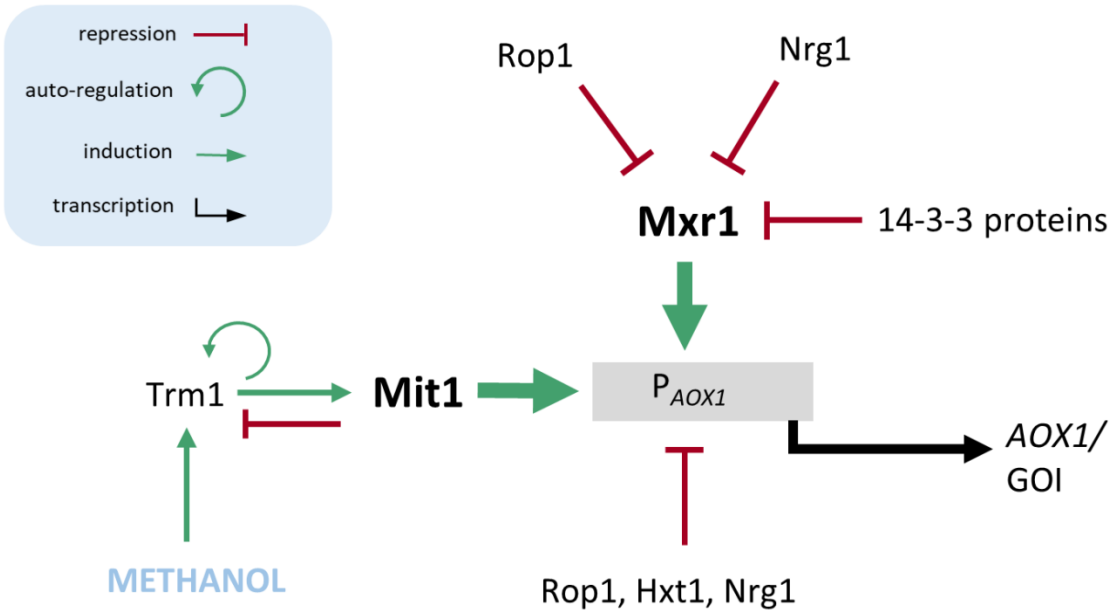


Figure 2. Overview of the interactions between regulators involved in methanol metabolism. Taken from [69].

1.6. Bottlenecks in recombinant protein production in *P. pastoris*

The utilization of strong promoters to achieve maximum product yield is usually the standard for most of the habitual expression hosts, including *P. pastoris.* In addition, the generation of multi-copy strains has been a strategy extensively used to increase the production level of heterologous proteins [70], [71]. In this context, the synthesis

of complex proteins requires to redistribute cellular resources for translation, folding, posttranslational modifications and secretion. This situation implies that any of these processes can be a potential bottleneck for recombinant protein production.

1.6.1. Protein folding and secretion

The saturation of the recombinant protein production machinery carries an important cost for cell fitness and growth rate could be reduced. For instance, if translation is not able to efficiently couple with folding and secretion, a fraction of misfolded proteins accumulates intracellularly. This situation can lead to endoplasmic reticulum (ER) stress and the activation of the unfolded protein response (UPR) and the ER-associated degradation (ERAD) (**Figure 3**) [72].

Figure 3. Major bottlenecks encountered by recombinant proteins on their way through the secretory pathway in eukaryotic hosts. Taken from [72].

The oxidative protein folding machinery in the ER requires dedicated oxidoreductases and chaperones. Firstly, disulfide bonds are formed from sulfhydryl groups of

immature proteins by the protein disulfide isomerase (PDI). Reduced PDI is then re-oxidized by FAD-dependent ER oxidoreductin (ERO1) using oxygen as electron acceptor. This reaction generates hydrogen peroxide, a reaction oxygen specie (ROS). Although ROS is a natural byproduct of metabolism, the increase of its levels due to environmental stress may result in damage to cell structures such as DNA, RNA, lipids and proteins. Consequently, the oxidative stress response is triggered to eliminate ROS from ER. ROS are detoxified by the oxidation of glutathione (GSH) to disulfide glutathione (GSSG). Finally, glutathione is recycled by converting GSSG to its reduced state, using NADPH as electron donor. Therefore, maintenance of the redox homeostasis inside the ER requires energy, and failure to do so can be harmful for the cell. Moreover, under unfolded protein stress situations, the UPR signaling pathway activates ATP-dependent chaperones to correctly refold heterologous proteins, which may involve several re-folding attempts that consume ATP.

Several strategies have been applied in order to increase the efficiency of this folding machinery. For instance, overexpression of the UPR transcriptional activator gene *HAC1* has been proved to improve secretion of many different recombinant proteins by increasing the synthesis of ER proteins related with protein folding [73]–[76]. More specifically, disulfide bond formation has been reported as rate-limiting step for a correct posttranslational modification of the heterologous protein. In this regard, by overexpression of the protein disulfide isomerase (*PDI1*), enhanced secretion of different recombinant proteins has been successfully achieved enhanced protein secretion (reviewed in [77]).

1.6.2. Energetic metabolism and availability of redox cofactors

Recombinant protein production is a highly demanding process in terms of energy and reducing power, particularly when the product is secreted (**Figure 4**). High levels of protein overproduction usually result in a phenomenon called metabolic burden. Cell tries to rearrange its metabolism to cope with this overload derived from heterologous protein synthesis, maturation, folding and secretion. Specifically, the flux redistribution over the metabolic network seem to be geared to increase supply of ATP, NADH or NADPH, while by-product formation is reduced [78]–[80].

Several ^{13}C-MFA studies have been performed in regards of the quantification of redistribution of metabolic fluxes through the central carbon metabolism of recombinant protein producing strains compared to its wild-type control. For instance, central carbon metabolism fluxes of a *P. pastoris* producing intracellularly a β-aminopeptidase under control of P_{GAP} were compared to its reference strain. A higher flux through the TCA cycle was observed in the producer strain, resulting in an increased regeneration of NADH and ATP [80], [81]. Moreover, the recombinant strain showed reduced biomass yield which correlated with the PPP activity. These results indicated that cell requires an increased demand of energy due to the metabolic burden associated to recombinant protein production.

Similar ^{13}C-MFA studies have been performed in P_{AOX1}-based expression systems. The effect of the production of a recombinant lipase from *R. oryzae* (Rol) in the metabolic network of the cell was investigated in cells grown in an 80/20 (w/w, %) glucose/methanol mix culture medium [82], [83]. Increased fluxes through glycolysis, TCA cycle and methanol dissimilatory pathway were detected this way. Remarkably, the reduction of biomass yield observed in the recombinant strain did not correlate to the oxidative PPP flux, as in previous glucose-grown ^{13}C-MFA studies. This suggests that an increased NADPH supply through PPP in Rol-producing strains would be necessary to regenerate GSH in the endoplasmic reticulum, since Rol production has been previously shown to trigger UPR in *P. pastoris* [84].

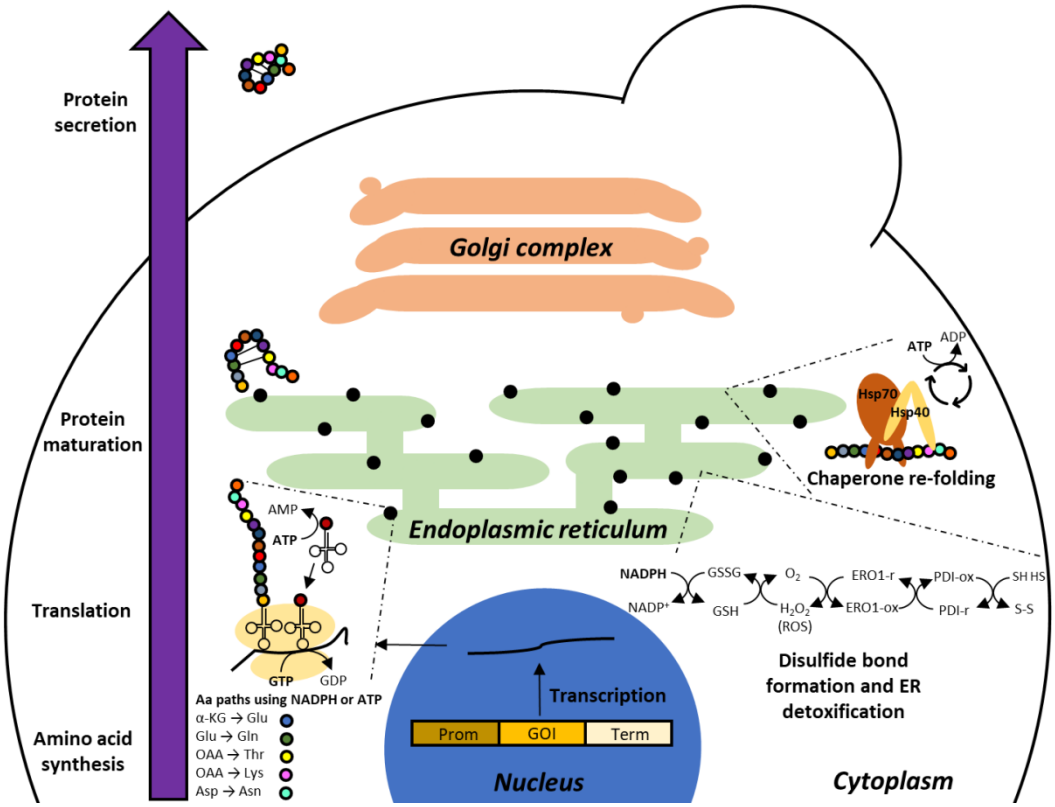


Figure 4. Main steps involved in recombinant protein production and their cofactor requirements. α-KG: α-ketoglutarate, Glu: glutamate, Gln: glutamine, OAA: oxaloacetate, Thr: threonine, Lys: lysine, Asp: aspartate, Asn: asparagine, prom: promoter, GOI: gene of interest, term: terminator, Hsp: heat-shock protein, GSSG: glutathione disulfide, GSH: reduced glutathione, ROS: reactive oxygen species, ERO1: endoplasmic reticulum oxidoreductin 1, PDI: protein disulfide isomerase.

It has been already explained that ATP and NADPH play a central role for chaperone function and disulfide bond formation. However, these molecules are also required in earlier steps of protein production. Since yeast cultures are usually performed in minimal media, cells need to synthesize their own amino acids. As main building blocks for protein translation, amino acid requirements are higher in recombinant strains, with its consequently increased cost of ATP and NADPH. Moreover, protein translation also implies an energetic cost payed in form of ATP and GTP. These molecules are used for tRNA-amino acid and aminoacyl-tRNA-ribosome binding, respectively. Overall, all processes involved in protein expression, from amino acid synthesis to protein folding and secretion, involve an energetic precursor at some point. This explains why recombinant protein production leads to a metabolic burden in cell. This way, draining part of the building and energetic blocks required for anabolic process and

endogenous protein synthesis leads to a suboptimal cellular state for the cell that results in lower production yields.

Several metabolic engineering strategies have been applied to alleviate the constraints in recombinant protein production. Initially, targeted genes were chosen by rational approaches based on existing literature. Nocon and co-workers implemented for the first time gene overexpression and knock-out strategies in *P. pastoris* based on GEM simulations [78]. In this study, nine genetic modifications related to central carbon metabolism were performed separately and five of them resulted in increased recombinant protein production levels. Notably, one of the successful approaches involved overexpressing single genes of the oxidative branch of the pentose phosphate pathway (PPP), involved in NADPH generation. In a later study, the same group applied combinatorial overexpression of several genes of the PPP aiming to achieve even a superior flux through the oxidative branch of PPP [85]. Strain performance was different depending on the genes combined, obtaining strains with reduced or improved protein production. In 2017, Jayachandran and co-workers explored a different alternative to engineer the redox balance by overexpressing a NADH oxidase from *Lactococcus lactis* (noxE) in a recombinant *P. pastoris* strain consuming this way part of the extra NADH supplied by the methanol metabolism [86]. As a result, methanol assimilation was increased due to the displacement in the thermodynamics in this pathway and the extracellular production of the *Candida antarctica* lipase B (CalB) increased. In the same study, an adelynate kinase *ADK1* from *S. cerevisiae* was also overexpressed with the aim to compensate the lower ATP supply due to the extra NADH consumption by noxE, resulting in even higher protein yields.

In addition, a line of research regarding the importance of amino acid supply has been also investigated. In this regard, strain engineering has focused on improving flux through amino acid biosynthesis. For instance, by overexpressing the activator *GCN4* (involved in several pathways of amino acid generation), the recombinant protein production of different model enzymes was significantly increased [87], [88]. The overexpression of specific enzymes involved in amino acid anabolic reactions also resulted in increased amino acid supplies and protein yields [87].

1.7. Synthetic biology of *P. pastoris*

Over the years, the creation of novel vectors using different auxotrophic and antibiotic resistance selectable markers have improved the capabilities of genetic manipulation of *P. pastoris* [89]. This supposed the first step towards increasing the genetic accessibility in this organism, which has some difficulties compared to the highly employed *S. cerevisiae.* In this regard, important advances have been achieved in both knock-in and knock-out strategies to overcome the intrinsic genetic limitations of *P. pastoris*, such as lower transformation efficiencies and random genetic integration events [90].

1.7.1. State-of-the-art in traditional gene integration techniques

Initially, genetic manipulation of *P. pastoris* was limited to gene overexpression by single homologous recombination of a linearized vector into a targeted locus. Although transformation protocols for *P. pastoris* are not as efficient as those of *S. cerevisiae,* they are usually good enough to obtain enough transformants to perform a clone screening [91]. Moreover, highly efficient methods are available to select for clones with multiple gene integrations of your gene of interest [92], [93].

Besides the lower transformation efficiency, another important drawback, which supposes an important challenge in genetic engineering of *P. pastoris*, concerns the extremely low efficiency of classic gene disruption strategies. These are based on gene replacement by double crossover homologous recombination with a disruption cassette [94]. The recombination mechanism used for DNA damage repair in *P. pastoris* is not active enough to provide a high rate of success in double recombination events. In addition, the non-homologous end joining (NHEJ) pathway is dominant over the homologous recombination (HR) pathway, which causes a high number of false positive clones due to integration of the disruption cassette in a random locus. Classically, in order to partially overcome this situation, the most common practice was to extent the homology arms of the disruption cassette to at least 1000 bp [95]. Later, more sophisticated strategies were developed. A strain of *P. pastoris* without an active NHEJ machinery was created by knock-out of the gene *KU70*. This allowed to shorten homologous sequences in the disruption cassette to less than 250 pb, but

problems regarding genetic stability of the strain were detected [90]. Another alternative method that is not dependent on a specific strain is the split-marker approach. The utilization of two overlapping DNA fragments as a disruption cassette instead of a single DNA fragment avoids the false positive issue but reduces transformation frequencies, since a triple homologous recombination is be required [96]. All these strategies have several advantages over classical methods but, as previously explained, all of them still pose some drawbacks. In this regard, the CRISPR/Cas9 toolbox for genetic editing is taking important attention during the past recent years and has the potential to eventually become a reference technique to avoid complex knock-out strategies.

1.7.2. CRISPR/Cas9 system for genome editing

CRISPR/Cas9 is naturally a defense mechanism of bacteria to protect the cell against foreign DNA sequences and it has been heterologously expressed in various organisms over the past years [97]. This technology provides and specific and efficient way for both DNA integration and gene disruption. The components of this system are a nuclease, named Cas9, and a specific gRNA (guide RNA) with 20 bp complementary to the target locus. The gRNA interacts with Cas9 and directs its activity to a specific locus, creating a DNA break. This DNA damage can be repaired by NHEJ or by HR if a homologous DNA is present. On one side, NHEJ pathway usually adds indels, displacing the reading frame of a gene. On the other side, since double strand breaks (DSBs) recruits HR machinery, the integration by double recombination of a foreign DNA cassette can be achieved with a higher success rate [98].

In *S. cerevisiae* this system was first implemented by Dicarlo and co-workers [99] and it has subsequently been improved for more complex tasks such as multiplexing and transcriptional regulation [100], [101]. The adaptation of the CRISPR technology to *P. pastoris* was firstly achieved by Weninger co-workers [102]. By implementing a combinatorial approach, 95 constructs with different codon optimized sequences of *CAS9*, gRNA sequences, promoters (some combined with ribozymes) for gRNA expression and promoters for *CAS9* expression. Only 6 from the 95 constructs showed efficient genome edition. The main features necessary to achieve success consisted in low levels of Cas9 inside the cell to avoid toxicity problems and removal of

untranslated regions (UTRs) using ribozymes for a functional gRNA structure. Recently, the same group also reported that by combining the CRISPR system with a Δ*KU70* strain they were able to achieve a near 100% efficiency in genes that had been proven problematic to eliminate in a wild type strain. In the same study, they also observed an increase in the donor cassette integration in the wild type strain by adding an autonomously replicating sequence (ARS) [103]. For a simpler application and broader diffusion of this system in *P. pastoris*, Gassler and co-workers published a detailed procedure using its designed modular vector [104]. This CRISPR/Cas9-based toolkit uses a Golden Gate-derived approach for *P. pastoris* (GoldenPiCS) [105] and provides different resistance markers and promoters for *CAS9* expression. Moreover, this procedure shows a simple way to create gRNAs by using overlapping extension PCR (OE-PCR). Finally, the most recent research regarding CRISPR studied the effect of different ARS in CRISPR/Cas9 efficiency. The ARS that showed better plasmid stability (the panARS from *Kluyveromyces lactis*) was compared with the PARS1 efficiency used by Weninger and co-workers in its CRISPR vector, showing up to a ten-fold better genome-editing efficiency [106]. Currently, CRISPR-based studies in *P. pastoris* are still rather limited compared to other expression hosts. Nevertheless, improvements of this technology are continuously appearing, so it is foreseen that its applicability will increase in the upcoming years.

1.7.3. Modular cloning in *P. pastoris*

Typically, the (re)design and construction of expression vectors with the aim of adapting them for the purpose of your study has been performed using the classic restriction-ligation approach. However, this classic methodology is time-consuming and does not allow to assemble multiple fragments simultaneously. Moreover, false-positives clones can appear due to undesired vector re-ligations. During the last 10 years, more advanced techniques such as Gibson assembly and Golden Gate assembly (GGA) have appeared. These techniques are based in the utilization of a mix of polymerase, exonuclease and ligase and type II restriction enzymes, respectively. Both are able to provide a clean (i.e. without extra base pairs) multiassembly in a single reaction step (**Figure 5**).

The breakthrough in the application of these novel techniques started with the creation of a toolkit, based on GGA, called Modular Cloning (MoClo), introduced by Ernst Weber and co-workers [107]. This system allows the assembly of up to six DNA fragments in a single reaction by using the type II restriction enzymes BsaI and BbsI (BpiI) and predefined overhangs. MoClo was adapted to DNA parts from *P. pastoris,* such as promoters (P_{AOX1}, P_{GAP}…), secretion signals and transcription terminators, and included to some existing parts from *S. cerevisiae* [108]. However, this first library was only adapted for optimization of single expression cassettes. In this regard, Prielhofer and co-workers have recently provided a modular system for overexpression of multiple genes, called GoldenPiCS, which includes 20 promoters, 10 transcription terminators, 4 genome integration targets and 4 resistance genes [105]. For Gibson assembly, a similar approach for modular vector design has been constructed by Vogl and co-workers consisting of 49 promoters and 20 terminators [109].

The main potential of these kits or toolboxes remains in their usage for metabolic engineering. For instance, classical molecular biology methods require several steps of cloning or transformation in order to introduce a set of foreign genes in yeast. By using Gibson Assembly or GGA-derived techniques it is possible to generate large vectors that encode themselves entire new pathways, thereby saving time, handwork, and resources.

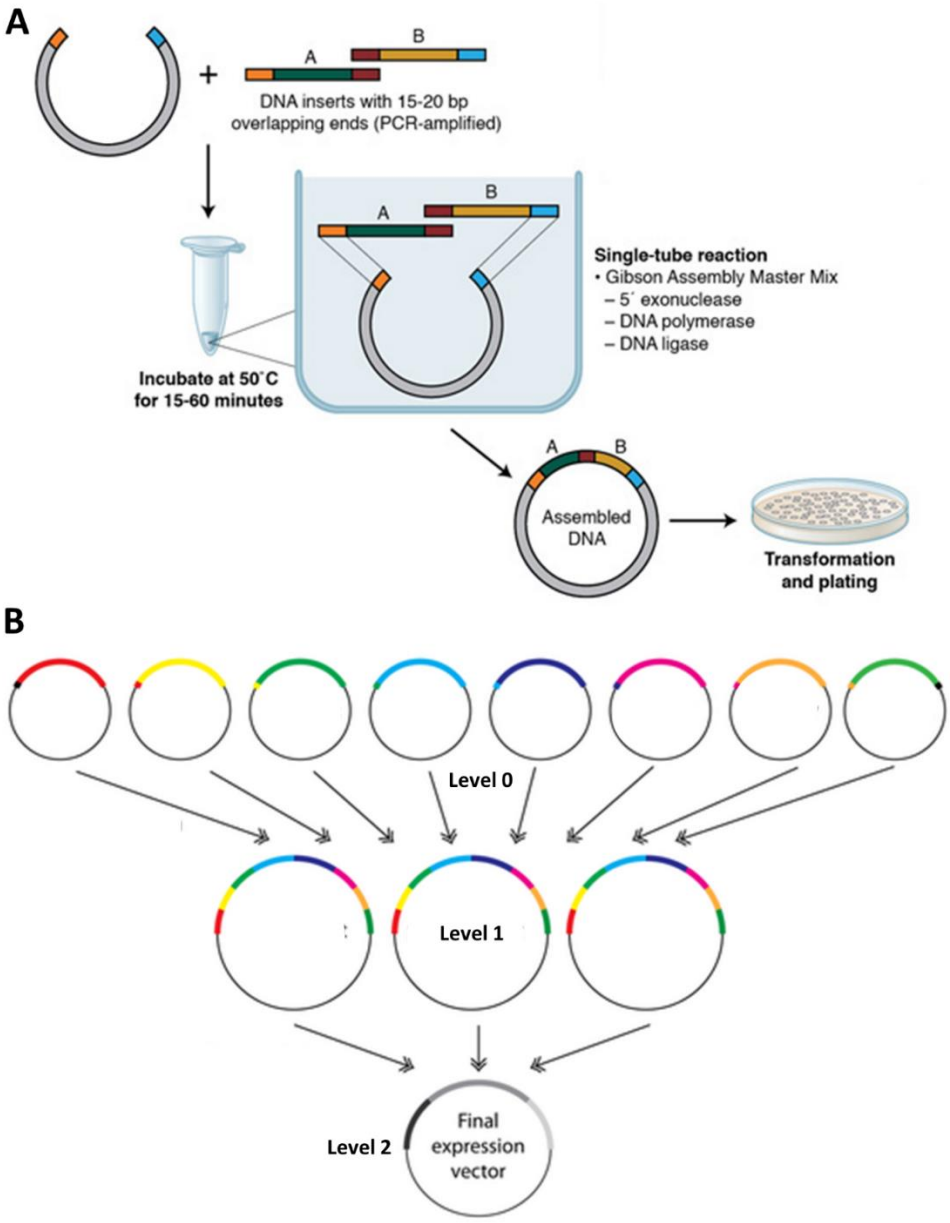


Figure 5. Overview of Gibson Assembly and Golden Gate workflow. Gibson Assembly (**A**) combines 3 different reactions in a single tube: 5' exonuclease, 5'→3' polymerase and DNA ligase activity. The exonuclease activity creates sticky ends in the overlapping sequences of the DNA fragments, the polymerase activity fills the resulting gaps produced by the exonuclease activity and the DNA ligase seals the nicks between fragments. Libraries designed for GGA modulation (**B**) have 3 vector levels. Level 0: contains the basic functional DNA parts (promoters, terminators, coding gene sequences, etc.). Level 1: contains single transcriptional units (obtained by assembling of level 0 parts). Level 2: final vector containing multiple expression cassettes (obtained by assemble of level 1 modules). Figures A and B were adapted from images from New England Biolabs and Addgene webpages, respectively. Taken from: https://www.neb.com/applications/cloning-and-synthetic-biology/dna-assembly-and-cloning/gibson-assembly
https://www.addgene.org/kits/wittbrodt-golden-gateway/

1.8. Metabolic modeling of *P. pastoris*

Genome-Scale metabolic models (GEMs) and their applications have obtained a big momentum nowadays. These models provide a mathematical approach to understand at systems level metabolic networks many different organisms. The creation of a structured platform to describe all metabolic reactions inside an organism was only possible due to the release of the first whole-genome sequences. Briefly, GEMs are created by relating metabolic genes with metabolic reactions. This way, a stoichiometric matrix that summarizes all possible biochemical reactions in a given organism can be created (**Figure 6**) [110].

A major application of these models is the prediction of the metabolic phenotype of a cell under certain growth conditions. Also, GEMs are often used for the *in silico* design of metabolic engineering strategies to increase the production levels of an existing strain or create and optimize novel non-native pathways [111], [112]. GEMs can employ a wide range of algorithms with the aim to determine targets for over/under-expression or gene removal. Moreover, models also provide a tool for -omics data integration, allowing a deeper interpretation of the metabolic behavior of your host [113], [114]. For instance, multi-level data sets can be correlated with metabolic fluxes, thereby facilitating a biological interpretation of omics data sets. Interestingly, GEMs can also be reduced to core models limited to the central carbon metabolism for a more precise ^{13}C-based metabolic flux analyses (13C-MFA) [81], [115], [116]. Finally, GEMs also provide a useful tool for comparison of metabolic networks between different species [117]. This type of *in silico* analysis can be used to detect conserved or missing pathways between species and determine evolutive connections between them.

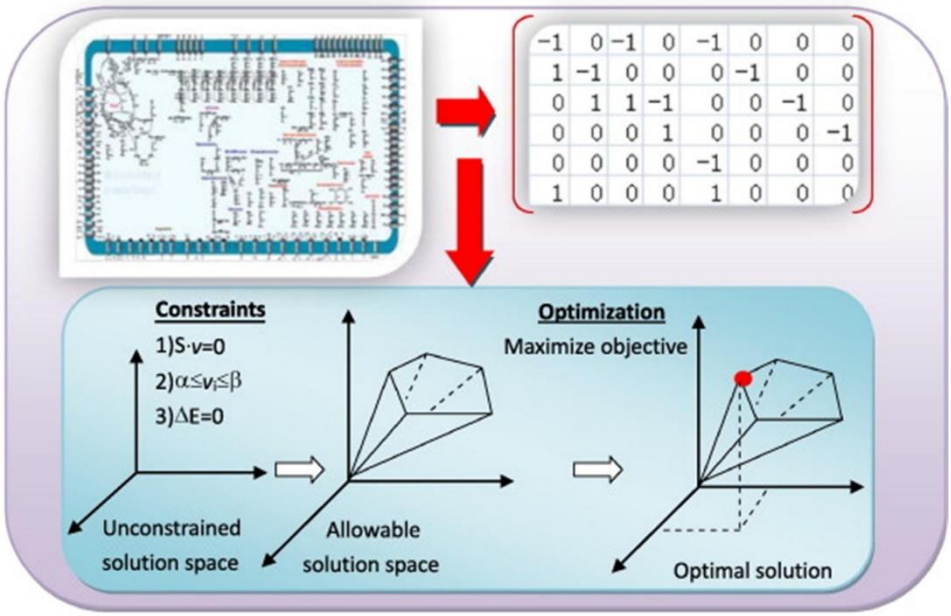


Figure 6. Conversion of the reaction list obtained from genome annotation into a constrained-based mathematical model. Taken from [110].

Although GEMs are tools that have supported the acquisition of new knowledge in several biological areas, they still have several limitations. GEMs only take into account metabolic reactions (and, in some cases, their thermodynamics, to assure the same *in silico* directionality of the reactions as *in vivo*). Metabolic networks are significantly more complex, since transcriptional regulation, enzyme kinetics and their interconnections with metabolic fluxes are not considered. Moreover, important cellular processes taking place during recombinant protein production, such as protein folding and secretion, are not included. These drawbacks suppose the major limitations for accurate whole cell predictive modelling.

The genome sequencing of *P. pastoris* allowed for the creation of the first genome-scale metabolic models (GEMs) of this yeast. More concretely, the sequencing of the strains GS115 and DSMZ 70382 allowed the simultaneous publication of the GEMs iPP668 [118] and PpaMBEL1254 [119], respectively. Afterwards, Caspeta and co-workers published another model (iLC915), based on the GS115 genome as well [119]. This model provided a 17.2% higher gene coverage compared to the previous ones. Later, a GEM based on iLC915 was created by Irani and co-workers, named

ihGlycopastoris [119]. This model included the native and humanized glycosylation processes for recombinant proteins.

Although the three first generated GEMs (iPP668, PpaMBEL1254 and iLC915) are based on the same organism, they showed differences in some reactions and nomenclatures and a manual curation was necessary. Since *P. pastoris* is a non-conventional yeast, most of the metabolic reconstructions were based in more characterized model organisms, such as *S. cerevisiae*. This led to a misinterpretation of some particular characteristics of *P. pastoris*. In this regard, Tomàs-Gamisans and co-workers created the integrative model iMT1026, based on the ones previously published [120]. iMT1026 corrected mistakes and filled gaps in the fatty acid metabolism, sphingolipid synthesis, GPI-anchor biosynthesis, N-glycosylation and oxidative phosphorylation. More recently, Tomàs-Gamisans and co-workers improved their iMT1026 model by adding accurate biomass equations of *P. pastoris* under different single carbon source conditions (glycerol and methanol as unique substrates) [121]. This way, the updated version of iMT1026 has a higher potential for phenotypic prediction. Independently, Ye and co-workers also improved the first version of iMT1026, but focusing on increasing gene annotation [122]. This upgraded version used the latest genome annotations and literature data to increase the reaction and metabolite numbers, improving the accuracy and precision of the model.

1.9. References

[1] "Protein Expression Market by Type (E. Coli, Mammalian, Yeast, Pichia, Insect, Baculovirus, Cell-free), Products (Competent Cells, Reagents, Instruments, Services), Application (Therapeutic, Research, Industrial) & End User - Global Forecast to 2022". Available online: https://www.marketsandmarkets.com/PressReleases/protein-expression.asp.

[2] S. Lobedanz *et al.*, "Enzymes in Industrial Biotechnology," in *Kirk-Othmer Encyclopedia of Chemical Technology*, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2016, pp. 1–73.

[3] J. L. Adrio and A. L. Demain, "Recombinant organisms for production of industrial products" *Bioengineered Bugs,* vol. 1, no. 2. pp. 116-131, 2010.

[4] H. Yang, J. Li, H. D. Shin, G. Du, L. Liu, and J. Chen, "Molecular engineering of industrial enzymes: Recent advances and future prospects," *Appl. Microbiol. Biotechnol.*, vol. 98, no. 1, pp. 23–29, 2014.

[5] J. L. Corchero *et al.*, "Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics," *Biotechnology Advances*, vol. 31, no. 2. pp. 140–153, Mar-2013.

[6] E. Rodriguez, E. J. Mullaney, and X. G. Lei, "Expression of the Aspergillus fumigatus phytase gene in Pichia pastoris and characterization of the recombinant enzyme," *Biochem. Biophys. Res. Commun.*, vol. 268, no. 2, pp. 373–378, 2000.

[7] S. C. Spohner, H. Müller, H. Quitmann, and P. Czermak, "Expression of enzymes for the usage in food and feed industry with Pichia pastoris," *J. Biotechnol.*, vol. 202, pp. 118–134, 2015.

[8] L. Otterbein, E. Record, S. Longhi, M. Asther, and S. Moukha, "Molecular cloning of the cDNA encoding laccase from Pycnoporus cinnabarinus I-937 and expression in Pichia pastoris," *Eur. J. Biochem.*, vol. 267, no. 6, pp. 1619–1625, 2000.

[9] D. M. Soden, J. O'Callaghan, and A. D. W. Dobson, "Molecular cloning of a laccase isozyme gene from Pleurotus sajor-caju and expression in the heterologous Pichia pastoris host," *Microbiology*, vol. 148, no. 12, pp. 4003-4014, 2002.

[10] F. Hong, N. Q. Meinander, and L. J. Jönsson, "Fermentation strategies for improved heterologous expression of laccase in Pichia pastoris," *Biotechnol. Bioeng.*, vol. 79, no. 4, pp. 438–449, 2002.

[11] J. G. Berrin, G. Williamson, A. Puigserver, J. C. Chaix, W. R. McLauchlan, and N. Juge, "High-level production of recombinant fungal endo-β-1,4-xylanase in the methylotrophic yeast Pichia pastoris," *Protein Expr. Purif.*, vol. 19, no. 1, pp. 179–187, 2000.

[12] K. K. Bhardwaj and R. Gupta, "Synthesis of chirally pure enantiomers by lipase," *J. Oleo Sci.*, vol. 66, no. 10, pp. 1073–1084, 2017.

[13] A. C. L. D. M. Carvalho *et al.*, "Recent advances in lipase-mediated preparation of pharmaceuticals and their intermediates," *Int. J. Mol. Sci.*, vol. 16, no. 12, pp. 29682–29716, 2015.

[14] F. Valero, "Recent advances in pichia pastoris as host for Heterologous Expression System for Lipases: A review," in *Methods in Molecular Biology*, vol. 1835, pp. 205–216, 2018.

[15] A. R. Gomes, S. M. Byregowda, B. M. Veeregowda, and V. Balamurugan, "An Overview of Heterologous Expression Host Systems for the Production of Recombinant Proteins," *Adv. Anim. Vet. Sci.*, vol. 4, no. 7, pp. 346–356, 2016.

[16] R. Baghban *et al.*, "Yeast Expression Systems: Overview and Recent Advances," *Mol. Biotechnol.*, vol. 61, no. 5, pp. 365–384, 2019.

[17] A. Maccani *et al.*, "Pichia pastoris secretes recombinant proteins less efficiently than Chinese hamster ovary cells but allows higher space-time yields for less complex proteins," *Biotechnol. J.*, vol. 9, no. 4, pp. 526–537, 2014.

[18] T. Lai, Y. Yang, and S. K. Ng, "Advances in mammalian cell line development

technologies for recombinant protein production," *Pharmaceuticals*, vol. 6, no. 5, pp. 579–603, 2013.

[19] J. Zitzmann, G. Sprick, T. Weidner, C. Schreiber, and P. Czermak, "Process Optimization for Recombinant Protein Expression in Insect Cells," *New Insights into Cell Cult. Technol.*, vol. 1, 2017.

[20] J. M. Cregg, K. J. Barringer, A. Y. Hessler, and K. R. Madden, "Pichia pastoris as a host system for transformations.," *Mol. Cell. Biol.*, vol. 5, no. 12, pp. 3376–3385, 2015.

[21] M. Hasslacher *et al.*, "High-level intracellular expression of hydroxynitrile lyase from the tropical rubber tree Hevea brasiliensis in microbial hosts," *Protein Expr. Purif.*, vol. 11, no. 1, pp. 61–71, 1997.

[22] A. Küberl *et al.*, "High-quality genome sequence of Pichia pastoris CBS7435," *J. Biotechnol.*, vol. 154, pp. 312–320, 2011.

[23] D. Mattanovich *et al.*, "Genome, secretome and glucose transport highlight unique features of the protein production host Pichia pastoris," *Microb. Cell Fact.*, vol. 8, no. 1, pp. 29, 2009.

[24] K. De Schutter *et al.*, "Genome sequence of the recombinant protein production host Pichia pastoris," *Nat. Biotechnol.*, vol. 27, no. 6, pp. 561–566, 2009.

[25] Y. Yamada, M. Matsuda, K. Maeda, and K. Mikata, "The Phylogenetic Relationships of Methanol-assimilating Yeasts Based on the Partial Sequences of 18S and 26S Ribosomal RNAs: The Proposal of Komagataella Gen. Nov. (Sacchar omy cet aceae)," *Biosci. Biotechnol. Biochem.*, vol. 59, no. 3, pp. 439–444, 1995.

[26] C. P. Kurtzman, "Description of Komagataella phaffii sp. nov. and the transfer of Pichia pseudopastoris to the methylotrophic yeast genus Komagataella," *Int. J. Syst. Evol. Microbiol.*, vol. 55, no. 2, pp. 973–976, Mar. 2005.

[27] C. A. Thompson, "FDA approves kallikrein inhibitor to treat hereditary angioedema," *Am. J. Heal. Pharm.*, vol. 67, no. 2, pp. 93–93, 2010.

[28] J. L. Cereghino and J. M. Cregg, "Heterologous protein expression in the

methylotrophic yeast Pichia pastoris," *FEMS Microbiol. Rev.*, vol. 24, no. 1, pp. 45–66, 2000.

[29] O. Cos, A. Serrano, J. L. Montesinos, P. Ferrer, J. M. Cregg, and F. Valero, "Combined effect of the methanol utilization (Mut) phenotype and gene dosage on recombinant protein production in Pichia pastoris fed-batch cultures," *J. Biotechnol.*, vol. 116, pp. 321–335, 2005.

[30] S. Shen, G. Sulter, T. W. Jeffries, and J. M. Cregg, "A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast Pichia pastoris,", vol. 216, no. 1., pp. 93-102, 1998.

[31] X. Garcia-Ortega, P. Ferrer, J. L. Montesinos, and F. Valero, "Fed-batch operational strategies for recombinant Fab production with Pichia pastoris using the constitutive GAP promoter," *Biochem. Eng. J.*, vol. 79, pp. 172–181, 2013.

[32] S. Macauley-Patrick, M. L. Fazenda, B. McNeil, and L. M. Harvey, "Heterologous protein production using the Pichia pastoris expression system," *Yeast*, vol. 22, no. 4. pp. 249–270, 2005.

[33] H. R. Waterham, M. E. Digan, P. J. Koutz, S. V. Lair, and J. M. Cregg, "Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter," *Gene*, vol. 186, no. 1, pp. 37–44, 1997.

[34] J. Wang *et al.*, "Methanol-Independent Protein Expression by AOX1 Promoter with trans-Acting Elements Engineering and Glucose-Glycerol-Shift Induction in Pichia pastoris," *Sci. Rep.*, vol. 7, 2017.

[35] T. Vogl *et al.*, "Methanol independent induction in Pichia pastoris by simple derepressed overexpression of single transcription factors," *Biotechnol. Bioeng.*, vol. 115, no. 4, pp. 1037–1050, 2018.

[36] R. Prielhofer *et al.*, "Induction without methanol: Novel regulated promoters enable high-level expression in Pichia pastoris," *Microb. Cell Fact.*, vol. 12, no. 1, 2013.

[37] R. Prielhofer *et al.*, "Superior protein titers in half the fermentation time:

Promoter and process engineering for the glucose-regulated GTH1 promoter of Pichia pastoris," *Biotechnol. Bioeng.*, vol. 115, no. 10, pp. 2479–2488, 2018.

[38] W. J. Holmes, R. A. J. Darby, M. D. B. Wilks, R. Smith, and R. M. Bill, "Developing a scalable model of recombinant protein yield from Pichia pastoris: The influence of culture conditions, biomass and induction regime," *Microb. Cell Fact.*, vol. 8, pp. 35, 2009.

[39] N. Bonander *et al.*, "Transcriptome analysis of a respiratory Saccharomycescerevisiae strain suggests the expression of its phenotype is glucose insensitive and predominantly controlled by Hap4, Cat8 and Mig1," *BMC Genomics*, vol. 9, pp. 365, 2008.

[40] S. Swinnen, M. Klein, M. Carrillo, J. McInnes, H. T. T. Nguyen, and E. Nevoigt, "Re-evaluation of glycerol utilization in Saccharomyces cerevisiae: Characterization of an isolate that grows on glycerol without supporting supplements," *Biotechnol. Biofuels*, vol. 6, no. 1, pp. 157, 2013.

[41] M. Klein, S. Swinnen, J. M. Thevelein, and E. Nevoigt, "Glycerol metabolism and transport in yeast and fungi: established knowledge and ambiguities," *Environmental Microbiology*, vol. 19, no. 3, pp. 878–893, 2017.

[42] R. K. Bretthauer and F. J. Castellino, "Glycosylation of Pichia pastoris -derived proteins," *Biotechnol. Appl. Biochem.*, vol. 30, no. 3, pp. 193–200, 1999.

[43] T. Egli, J. P. van Dijken, M. Veenhuis, W. Harder, and A. Fiechter, "Methanol metabolism in yeasts: Regulation of the synthesis of catabolic enzymes," *Arch. Microbiol.*, vol. 124, no. 2–3, pp. 115–121, 1980.

[44] J. Baratti and R. Couderc, "Oxidation of Methanol by the Yeast, Pichia pastoris. Purification and Properties of the Alcohol Oxidase," *Agric. Biol. Chem.*, vol. 44, no. 10, pp. 2279–2289, 1980.

[45] H. Rußmayer *et al.*, "Systems-level organization of yeast methylotrophic lifestyle," *BMC Biol.*, vol. 13, pp. 80, 2015.

[46] F. S. Hartner and A. Glieder, "Regulation of methanol utilisation pathway genes

in yeasts," *Microb. Cell Fact.*, vol. 5, pp. 39, 2006.

[47] V. Chiruvolu, J. M. Cregg, and M. M. Meagher, "Recombinant protein production in an alcohol oxidase-defective strain of Pichia pastoris in fedbatch fermentations," vol. 21, no. 4, pp. 277-283, 1997.

[48] F. W. Krainer, C. Dietzsch, T. Hajek, C. Herwig, O. Spadiut, and A. Glieder, "Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway," *Microb. Cell Fact.*, vol. 11, no. 1, pp. 22, 2012.

[49] I. A. Pla, L. M. Damasceno, T. Vannelli, G. Ritter, C. A. Batt, and M. L. Shuler, "Evaluation of Mut+ and MutS Pichia pastoris phenotypes for high level extracellular scFv expression under feedback control of the methanol concentration.," *Biotechnol. Prog.*, vol. 22, no. 3, pp. 881–8, 2006.

[50] S. J. Kim, J. A. Lee, Y. H. Kim, and B. K. Song, "Optimization of the functional expression of Coprinus cinereus peroxidase in Pichia pastoris by varying the host and promoter," *J. Microbiol. Biotechnol.*, vol. 19, no. 9, pp. 966–971, 2009.

[51] J. F. Tschopp, P. F. Brust, J. M. Cregg, C. A. Stillman, and T. R. Gingeras, "Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris," *Nucleic Acids Res.*, vol. 15, no. 9, pp. 3859–3876, 1987.

[52] M. Inan and M. M. Meagher, "The effect of ethanol and acetate on protein expression in Pichia pastoris," *J. Biosci. Bioeng.*, vol. 92, no. 4, pp. 337–341, 2001.

[53] G. P. Lin-Cereghino *et al.*, "Mxr1p, a Key Regulator of the Methanol Utilization Pathway and Peroxisomal Genes in Pichia pastoris," *Mol. Cell. Biol.*, vol. 26, no. 3, pp. 883–897, 2006.

[54] O. V. Tyurin and D. G. Kozlov, "Deletion of the FLD gene in methylotrophic yeasts Komagataella phaffii and Komagataella kurtzmanii results in enhanced induction of the AOX1 promoter in response to either methanol or formate," *Microbiology (Russian Federation)*, vol. 84, no. 3, pp. 408–411, 2015.

[55] A. Singh and A. Narang, "The Mut + strain of Komagataella phaffii (Pichia

pastoris) expresses PAOX1 5 and 10 times faster than Mut s and Mut-strains: Evidence that formaldehyde or/and formate are true inducers of AOX.," *bioRxiv*, pp. 1–34, 2019.

[56] M. Inan and M. M. Meagher, "Non-repressing carbon sources for alcohol oxidase (AOX1) promoter of Pichia pastoris," *J. Biosci. Bioeng.*, vol. 92, no. 6, pp. 585–589, 2001.

[57] B. V. Kranthi, R. Kumar, N. V. Kumar, D. N. Rao, and P. N. Rangarajan, "Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in Pichia pastoris," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1789, no. 6–8, pp. 460–468, 2009.

[58] Y. Xuan, X. Zhou, W. Zhang, X. Zhang, Z. Song, and Y. Zhang, "An upstream activation sequence controls the expression of AOX1 gene in Pichia pastoris," *FEMS Yeast Res.*, vol. 9, no. 8, pp. 1271–1282, 2009.

[59] B. V. Kranthi, H. R. V Kumar, and P. N. Rangarajan, "Identification of Mxr1p-binding sites in the promoters of genes encoding dihydroxyacetone synthase and peroxin 8 of the methylotrophic yeast Pichia pastoris," *Yeast*, vol. 27, no. 9, pp. 705–711, 2010.

[60] P. K. Parua, P. M. Ryan, K. Trang, and E. T. Young, "Pichia pastoris 14-3-3 regulates transcriptional activity of the methanol inducible transcription factor Mxr1 by direct interaction," *Mol. Microbiol.*, vol. 85, no. 2, pp. 282–298, 2012.

[61] S. Takagi, N. Tsutsumi, Y. Terui, and X. Y. Kong, "Method for methanol independent induction from methanol inducible promoters in Pichia," US patent US8236528B2, 2012.

[62] X. Wang *et al.*, "Mit1 transcription factor mediates methanol signaling and regulates the alcohol oxidase 1 (AOX1) promoter in pichia pastoris," *J. Biol. Chem.*, vol. 291, no. 12, pp. 6245–6261, 2016.

[63] N. V. Kumar and P. N. Rangarajan, "The zinc finger proteins Mxr1p and repressor of phosphoenolpyruvate carboxykinase (ROP) have the same DNA binding

specificity but regulate methanol metabolism antagonistically in Pichia pastoris," *J. Biol. Chem.*, vol. 287, no. 41, pp. 34465–34473, 2012.

[64] P. Zhang, W. Zhang, X. Zhou, P. Bai, J. M. Cregg, and Y. Zhang, "Catabolite Repression of Aox in Pichia pastoris is dependent on hexose transporter PpHxt1 and pexophagy," *Appl. Environ. Microbiol.*, vol. 76, no. 18, pp. 6108–6118, 2010.

[65] X. Wang *et al.*, "PpNrg1 is a transcriptional repressor for glucose and glycerol repression of AOX1 promoter in methylotrophic yeast Pichia pastoris," *Biotechnol. Lett.*, vol. 38, no. 2, pp. 291–298, 2016.

[66] J. Wang *et al.*, "Methanol-Independent Protein Expression by AOX1 Promoter with trans-Acting Elements Engineering and Glucose-Glycerol-Shift Induction in Pichia pastoris," *Sci. Rep.*, vol. 7, pp. 1–12, 2017.

[67] T. Vogl *et al.*, "Methanol independent induction in Pichia pastoris by simple derepressed overexpression of single transcription factors," *Biotechnol. Bioeng.*, vol. 115, no. 4, pp. 1037–1050, 2018.

[68] C. H. Chang, H. A. Hsiung, K. L. Hong, and C. T. Huang, "Enhancing the efficiency of the Pichia pastoris AOX1 promoter via the synthetic positive feedback circuit of transcription factor Mxr1," *BMC Biotechnol.*, vol. 18, no. 1, pp. 1–10, 2018.

[69] E. Cámara, S. Monforte, J. Albiol, and P. Ferrer, "Deregulation of methanol metabolism reverts transcriptional limitations of recombinant Pichia pastoris (Komagataella spp) with multiple expression cassettes under control of the AOX1 promoter," *Biotechnol. Bioeng.*, vol. 116, no. 7, pp. 1710–1720, 2019.

[70] E. Cámara, N. Landes, J. Albiol, B. Gasser, D. Mattanovich, and P. Ferrer, "Increased dosage of AOX1 promoter-regulated expression cassettes leads to transcription attenuation of the methanol metabolism in Pichia pastoris," *Sci. Rep.*, vol. 7, no. 1, 2017.

[71] M. Mansur *et al.*, "Multiple gene copy number enhances insulin precursor secretion in the yeast Pichia pastoris," *Biotechnol. Lett.*, vol. 27, no. 5, pp. 339–345, 2005.

[72] M. Delic, R. Göngrich, D. Mattanovich, and B. Gasser, "Engineering of protein folding and secretion - Strategies to overcome bottlenecks for efficient production of recombinant proteins," *Antioxidants Redox Signal.*, vol. 21, no. 3, pp. 414–437, 2014.

[73] M. Guerfal *et al.*, "The HAC1 gene from Pichia pastoris: Characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins," *Microb. Cell Fact.*, vol. 9, no. 1, pp. 49, 2010.

[74] B. Gasser, M. Maurer, J. Gach, R. Kunert, and D. Mattanovich, "Engineering of Pichia pastoris for improved production of antibody fragments," *Biotechnol. Bioeng.*, vol. 94, no. 2, pp. 353–361, 2006.

[75] M. Valkonen, M. Ward, H. Wang, M. Penttilä, and M. Saloheimo, "Improvement of Foreign-Protein Production in Aspergillus niger var. awamori by Constitutive Induction of the Unfolded-Protein Response," *Appl. Environ. Microbiol.*, vol. 69, no. 12, pp. 6979–6986, 2003.

[76] M. Valkonen, M. Penttilä, and M. Saloheimo, "Effects of inactivation and constitutive expression of the unfolded-protein response pathway on protein production in the yeast Saccharomyces cerevisiae," *Appl. Environ. Microbiol.*, vol. 69, no. 4, pp. 2065–2072, 2003.

[77] V. Puxbaum, D. Mattanovich, and B. Gasser, "Quo vadis? The challenges of recombinant protein folding and secretion in Pichia pastoris," *Applied Microbiology and Biotechnology*, vol. 99, no. 7. Springer Verlag, pp. 2925–2938, 2015.

[78] J. Nocon *et al.*, "Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production," *Metab. Eng.*, vol. 24, pp. 129–138, 2014.

[79] Y. Nie *et al.*, "Impacts of high β-galactosidase expression on central metabolism of recombinant Pichia pastoris GS115 using glucose as sole carbon source via 13 C metabolic flux analysis," *J. Biotechnol.*, vol. 187, pp. 124–134, 2014.

[80] J. Heyland, J. Fu, L. M. Blank, and A. Schmid, “Carbon metabolism limits recombinant protein production in Pichia pastoris,” *Biotechnol. Bioeng.*, vol. 108, no. 8, pp. 1942–1953, 2011.

[81] J. Heyland, J. Fu, L. M. Blank, and A. Schmid, “Quantitative physiology of Pichia pastoris during glucose-limited high-cell density fed-batch cultivation for recombinant protein production,” *Biotechnol. Bioeng.*, vol. 107, no. 2, pp. 357–368, 2010.

[82] J. Jordà, P. Jouhten, E. Cámara, H. Maaheimo, J. Albiol, and P. Ferrer, “Metabolic flux profiling of recombinant protein secreting Pichia pastoris growing on glucose:methanol mixtures,” *Microb. Cell Fact.*, vol. 11, no. 1, pp-57, 2012.

[83] J. Jordà, H. C. Rojas, M. Carnicer, A. Wahl, P. Ferrer, and J. Albiol, “Quantitative Metabolomics and Instationary 13C-Metabolic Flux Analysis Reveals Impact of Recombinant Protein Production on Trehalose and Energy Metabolism in Pichia pastoris.,” *Metabolites*, vol. 4, no. 2, pp. 281–99, May 2014.

[84] D. Resina, M. Bollók, N. K. Khatri, F. Valero, P. Neubauer, and P. Ferrer, “Transcriptional response of P. pastoris in fed-batch cultivations to Rhizopus oryzae lipase production reveals UPR induction.,” *Microb. Cell Fact.*, vol. 6, p. 21, 2007.

[85] J. Nocon *et al.*, “Increasing pentose phosphate pathway flux enhances recombinant protein production in Pichia pastoris,” *Appl. Microbiol. Biotechnol.*, vol. 100, no. 13, pp. 5955–5963, 2016.

[86] C. Jayachandran, B. P. Athiyaman, and M. Sankaranarayanan, “Cofactor engineering improved CALB production in Pichia pastoris through heterologous expression of NADH oxidase and adenylate kinase,” *PLoS One*, vol. 12, no. 7, 2017.

[87] B. Gasser, D. Mattanovich, M. Sauer, and H. Rußmayer, “Improved Protein Production," European Patent EP2952584A1, 2014.

[88] L. Gu, J. Zhang, B. Liu, G. Du, and J. Chen, “High-Level Extracellular Production of Glucose Oxidase by Recombinant Pichia Pastoris Using a Combined Strategy,”

Appl. Biochem. Biotechnol., vol. 175, no. 3, pp. 1429–1447, 2014.

[89] M. Ahmad, M. Hirz, H. Pichler, and H. Schwab, "Protein expression in Pichia pastoris: Recent achievements and perspectives for heterologous protein production," *Appl. Microbiol. Biotechnol.*, vol. 98, no. 12, pp. 5301–5317, 2014.

[90] L. Näätsaari, B. Mistlberger, C. Ruth, T. Hajek, F. S. Hartner, and A. Glieder, "Deletion of the pichia pastoris ku70 homologue facilitates platform strain generation for gene expression and synthetic biology," *PLoS One*, vol. 7, no. 6, 2012.

[91] J. M. Cregg and K. A. Russell, "Transformation," in *Pichia Protocols*, New Jersey: Humana Press, vol. 103, pp. 27–40.

[92] S. Wu and G. J. Letchworth, "High efficiency transformation by electroporation of Pichia pastoris pretreated with lithium acetate and dithiothreitol.," *Biotechniques*, vol. 36, no. 1, pp. 152–4, 2004.

[93] L. C. Piva, M. O. Bentacur, V. C. B. Reis, J. L. De Marco, L. M. P. de Moraes, and F. A. G. Torres, "Molecular strategies to increase the levels of heterologous transcripts in Komagataella phaffii for protein production," *Bioengineered*, vol. 8, no. 5. pp. 441–445, 2017.

[94] N. A. Da Silva and S. Srikrishnan, "Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae," *FEMS Yeast Research*, vol. 12, no. 2. pp. 197–214, 2012.

[95] J. H. Nett, N. Hodel, S. Rausch, and S. Wildt, "Cloning and disruption of the Pichia pastoris ARG1, ARG2, ARG3, HIS1, HIS2, HIS5, HIS6 genes and their use as auxotrophic markers," *Yeast*, vol. 22, no. 4, pp. 295–304, 2005.

[96] S. Heiss, M. Maurer, R. Hahn, D. Mattanovich, and B. Gasser, "Identification and deletion of the major secreted protein of Pichia pastoris," *Appl. Microbiol. Biotechnol.*, vol. 97, no. 3, pp. 1241–1249, 2013.

[97] A. C. Komor, A. H. Badran, and D. R. Liu, "CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes," *Cell*, vol. 168, no. 1–2. pp. 20–36, 2017.

[98] F. Storici, C. L. Durham, D. A. Gordenin, and M. A. Resnick, "Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast," *Proc. Natl. Acad. Sci.*, vol. 100, no. 25, pp. 14994–14999, 2003.

[99] J. E. Dicarlo, J. E. Norville, P. Mali, X. Rios, J. Aach, and G. M. Church, "Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems," *Nucleic Acids Res.*, vol. 41, no. 7, pp. 4336–4343, 2013.

[100] E. D. Jensen *et al.*, "Transcriptional reprogramming in yeast using dCas9 and combinatorial gRNA strategies," *Microb. Cell Fact.*, vol. 16, no. 1, pp. 46, 2017.

[101] R. Mans *et al.*, "CRISPR/Cas9: A molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae," *FEMS Yeast Res.*, vol. 15, no. 2, pp. 1–15, 2015.

[102] A. Weninger, A. M. Hatzl, C. Schmid, T. Vogl, and A. Glieder, "Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast Pichia pastoris," *Journal of Biotechnology*, vol. 235, pp. 139–149, 2016.

[103] A. Weninger, J. E. Fischer, H. Raschmanová, C. Kniely, T. Vogl, and A. Glieder, "Expanding the CRISPR/Cas9 toolkit for Pichia pastoris with efficient donor integration and alternative resistance markers," *J. Cell. Biochem.*, vol. 119, no. 4, pp. 3183–3198, 2018.

[104] T. Gassler, L. Heistinger, D. Mattanovich, B. Gasser, and R. Prielhofer, "CRISPR/Cas9-mediated homology-directed genome editing in Pichia pastoris," in *Methods in Molecular Biology*, vol. 1923, 2019, pp. 211–225.

[105] R. Prielhofer *et al.*, "GoldenPiCS: A Golden Gate-derived modular cloning system for applied synthetic biology in the yeast Pichia pastoris," *BMC Syst. Biol.*, vol. 1923, no. 1, pp. 211-225, 2017.

[106] Y. Gu *et al.*, "Construction of a series of episomal plasmids and their application in the development of an efficient CRISPR/Cas9 system in Pichia pastoris," *World J. Microbiol. Biotechnol.*, vol. 35, no. 6, p. 79, 2019.

[107] E. Weber, C. Engler, R. Gruetzner, S. Werner, and S. Marillonnet, "A modular cloning system for standardized assembly of multigene constructs," *PLoS One*, vol. 6, no. 2, 2011.

[108] U. Obst, T. K. Lu, and V. Sieber, "A Modular Toolkit for Generating Pichia pastoris Secretion Libraries," *ACS Synth. Biol.*, vol. 6, no. 6, pp. 1016–1025, 2017.

[109] T. Vogl *et al.*, "A Toolbox of Diverse Promoters Related to Methanol Utilization: Functionally Verified Parts for Heterologous Pathway Expression in Pichia pastoris," *ACS Synth. Biol.*, vol. 5, no. 2, pp. 172–186, 2016.

[110] L. Liu, R. Agren, S. Bordel, and J. Nielsen, "Use of genome-scale metabolic models for understanding microbial physiology," *FEBS Letters*, vol. 584, no. 12, pp. 2556–2564, 2010.

[111] M. Cvijovic, S. Bordel, and J. Nielsen, "Mathematical models of cell factories: Moving towards the core of industrial biotechnology," *Microbial Biotechnology*, vol. 4, no. 5. pp. 572–584, 2011.

[112] M. Gustavsson and S. Y. Lee, "Prospects of microbial cell factories developed through systems metabolic engineering," *Microb. Biotechnol.*, vol. 9, no. 5, pp. 610–617, 2016.

[113] R. Saha, A. Chowdhury, and C. D. Maranas, "Recent advances in the reconstruction of metabolic models and integration of omics data," *Current Opinion in Biotechnology*, vol. 29, no. 1. pp. 39–45, 2014.

[114] T. Y. Kim, S. B. Sohn, Y. Bin Kim, W. J. Kim, and S. Y. Lee, "Recent advances in reconstruction and applications of genome-scale metabolic models," *Current Opinion in Biotechnology*, vol. 23, no. 4. pp. 617–623, 2012.

[115] A. Solà, P. Jouhten, H. Maaheimo, F. Sánchez-Ferrando, T. Szyperski, and P. Ferrer, "Metabolic flux profiling of Pichia pastoris grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates," *Microbiology*, vol. 153, no. 1, pp. 281–290, 2007.

[116] K. Baumann *et al.*, "A multi-level study of recombinant Pichia pastoris in

different oxygen conditions," *BMC Syst. Biol.*, vol. 4, pp. 141, 2010.

[117] T. Österlund, I. Nookaew, and J. Nielsen, "Fifteen years of large scale metabolic modeling of yeast: Developments and impacts," *Biotechnology Advances*, vol. 30, no. 5. pp. 979–988, 2012.

[118] B. K. S. Chung *et al.*, "Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement," *Microb. Cell Fact.*, vol. 9, no. 1, pp-50, 2010.

[119] S. B. Sohn *et al.*, "Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production," *Biotechnol. J.*, vol. 5, no. 7, pp. 705–715, 2010.

[120] M. Tomàs-Gamisans, P. Ferrer, and J. Albiol, "Integration and validation of the genome-scale metabolic models of Pichia pastoris: A comprehensive update of protein glycosylation pathways, lipid and energy metabolism," *PLoS One*, vol. 11, no. 1, 2016.

[121] M. Tomàs-Gamisans, P. Ferrer, and J. Albiol, "Fine-tuning the P. pastoris iMT1026 genome-scale metabolic model for improved prediction of growth on methanol or glycerol as sole carbon sources," *Microb. Biotechnol.*, vol. 11, no. 1, pp. 224–237, 2018.

[122] R. Ye *et al.*, "Comprehensive reconstruction and evaluation of Pichia pastoris genome-scale metabolic model that accounts for 1243 ORFs," *Bioresour. Bioprocess.*, vol. 4, p. 22, 2017.

2. Overview of the work

In the last 15 years, great advances have been made in the understanding of how recombinant protein production can affect cell physiology and trigger stress responses. Protein expression is a complex multi-step process (gene expression → protein translation → protein folding and maturation → protein secretion) and it is therefore important to identify any potential bottlenecks in each of these steps. Essentially, the approaches used to achieve this aim can be divided in two: rational design based on the manual (re)evaluation of the existing knowledge base, or rational design of new engineering strategies assisted by novel computational techniques. The decision between these two alternatives is the starting point of the classic design-build-test-learn cycle that encompasses the choices and actions necessary to perform rational metabolic engineering experiments.

This work focuses in rational cell engineering strategies, including model-based strategies, with the objective to overcome some of these limitations and improve heterologous protein secretion in the yeast *Pichia pastoris* using the *Rhizopus oryzae* lipase (Rol) as model protein. Strain modifications have been performed using conventional knock-in techniques as well as novel protocols for gene knock-out involving the CRISPR/Cas9 technology. Finally, the suitability of the recombinant strains has been verified at both Erlenmeyer and bioreactor scales. In this way, a systematic study of the physiological changes induced by these genetic modifications was performed.

Chapter 1 focuses on overcoming transcriptional limitations that had been previously detected in *P. pastoris* strains carrying multiple copies of the *ROL* gene. These studies, performed in the framework of the PhD thesis of Elena Cámara, concluded that insertion of too many expression vectors affect negatively the transcriptional levels of the genes regulated by P_{AOX1}. This led to propose a cell engineering strategy to rewire the regulation mechanisms of this promoter. Specifically, by overexpressing the P_{AOX1} transcription factors Mxr1 and Mit1, the expression of *ROL* and genes related with methanol pathway was increased. Concurrently, this resulted in increased methanol

assimilation capacity. However, these modifications resulted in a modest or insignificant improvement of RoI production in bioreactor-scale chemostat experiments, pointing out to the existence of other physiological bottlenecks at different levels of the protein production process.

This study has been published as:

Cámara, E.[*], Monforte, S.[*], Albiol, J., Ferrer, P. (2019) Deregulation of methanol metabolism reverts transcriptional limitations of recombinant *Pichia pastoris* (*Komagataella spp*) with multiple expression cassettes under control of the *AOX1* promoter. *Biotechnol. Bioeng.* 1-11.

[*]Equal contribution to the manuscript.

Contributions:

Elena Cámara had the original conceptual idea of this study. She did the experimental design and participated in all the experiments involving construction and testing of MXR1 strains. I created the strains with extra copies of *MXR1* as well as *MIT1* strains, performed most of the cultures and related analytics. The Biotechnology graduate student Rafael Jiménez worked under my supervision, assisting in the screening and cultivation of Mit1 strains during his internship.

In **chapter 2**, in order to identify other possible targets for metabolic engineering, *in silico* simulations were done using the model iMT1026 previously created by Màrius Tomàs during his PhD project. These simulations aimed at optimizing RoI production through single gene knock-outs. As a result, 6 target genes were selected (4 related with central carbon metabolism and 2 with amino acid metabolism). Briefly, the disruption of genes from central carbon metabolism were associated with an increase in NADPH availability (important cofactor in several steps of protein synthesis and secretion), while the other three were related with an increase in the supply of specific amino acids. Once targets were selected, an attempt to create the modified strains was done using the CRISPR/Cas9 system (work performed at University of Natural Resources and Life Sciences from Vienna, under supervision of the Dr. Brigitte Gasser). Unfortunately, the most promising knock-outs according to the *in silico* predictions could not be obtained due to the high essentiality of the reactions that the corresponding encoded proteins catalyze. Two knock-outs were successfully performed but their performance at Erlenmeyer and bioreactor scale were similar to those of control strain.

Contributions:

I did the *in silico* simulations using the *P. pastoris* GEM model, designed and applied the CRISPR-related techniques for gene knock-outs and performed the cultivations of the new strains.

In **Chapter 3,** following up the outcome of the initial design-build-test (DBT) cycle of metabolic engineering performed in chapter 2, an alternative redesign strategy was proposed, based on the knowledge learned from the first DBT iteration and combining it with conclusions from previous studies of our research group performed by the Dr. Màrius Tomàs in his PhD thesis, as well as from others, pointing out NADPH supply as a key limiting cofactor in recombinant protein production. As a mean to increase cellular NADPH availability, two knock-in strategies were proposed in this study: a NADH kinase and a NADH oxidase were overexpressed separately in a Rol-producing strain. The results obtained in Erlenmeyer and bioreactor-scale chemostat cultures point out to a beneficial effect of these modifications in terms of productivity. In addition, the physiological parameters of these strains were studied under different carbon source conditions and an *in silico* interpretation of the results was also performed with aim of obtaining a deeper understanding of the behavior of the modified strains.

Monforte, S., Quesada, A., Gasser, B., Albiol, J., Ferrer, P. Engineering redox metabolism in recombinant *Pichia pastoris* for enhanced protein secretion under different methanolic growth conditions. Manuscript under preparation.

Contributions:

Ane Quesada created the strains that overexpress a NADH kinase and assisted in the cultures of the NADH kinase strains during her master thesis. I created the strains that overexpress a NADH oxidase, performed the cultures of the strains, analytical procedures and *in silico* simulations and wrote the manuscript.

3. Objectives

The main objective of this work is the application and study of the impact of different cell engineering strategies directed to increase the production of the *Rhizopus oryzae* lipase (Rol) in the methylotrophic yeast *Pichia pastoris*. In order to achieve this goal, a wide range of potential limiting steps in recombinant protein production, from transcriptional to energy and building blocks supply, have been proposed and tried to implement using different systems and synthetic biology tools and methodologies. In order to accomplish this global aim, the following partial objectives were defined:

- Rational selection of gene knock-in targets with the aim to increase Rol production.
- Utilization of genome-scale metabolic models (GEMs) as tools for detection of genes whose deletion could potentially improve Rol production.
- Construction of strains that overexpress heterologous proteins with the objective to improve expression of *ROL* (transcription factors) or modify NADPH supply for a more optimal redox balance for recombinant protein production (redox enzymes).
- Application of the novel genome-editing system CRISPR/Cas9 to perform the disruptions of the previously predicted genes and generate new Rol-producing strains.
- Strain characterization using cultures at different scale-level to study the effect of the genetic modifications performed in recombinant protein production and cell physiology.
- Transcriptomic study of the new strains for a deeper inside in variations at gene regulatory level compared to control strain.
- *In silico* interpretation using GEMs of culture results with the aim to explain the physiological changes observed in the new strains from a metabolic perspective.

4. Deregulation of methanol metabolism reverts transcriptional limitations of recombinant *Pichia pastoris* (*Komagataella* spp) with multiple expression cassettes under control of the *AOX1* promoter

The methanol-regulated alcohol oxidase promoter (P_{AOX1}) of *Pichia pastoris* (syn. *Komagataella* spp.) is one of the strongest promoters for heterologous gene expression. Although increasing the gene dosage is a common strategy to improve recombinant protein productivities, *P. pastoris* strains harbouring more than two copies of a *Rhizopus oryzae* lipase gene (*ROL*) have previously shown a decrease in cell growth, lipase production and substrate consumption, as well as a significant transcriptional downregulation of methanol metabolism. This pointed to a potential titration effect of key transcriptional factors Mxr1 and Mit1 regulating methanol metabolism caused by the insertion of multiple expression vectors.

To proof this hypothesis, a set of strains carrying one and four copies of *ROL* (1C and 4C, respectively) were engineered to co-express one or two copies of *MXR1**, coding for a Mxr1 variant insensitive to repression by 14-3-3 regulatory proteins, or one copy of *MIT1*. Small scale cultures revealed that growth, Rol productivity and methanol consumption were improved in the 4C-MXR1*and MIT1 strains growing on methanol as sole carbon source, whereas only a slight increase in productivity was observed for re-engineered 1C strains. We further verified the improved performance of these strains in glycerol/methanol-limited chemostat cultures.

Keywords: *Pichia pastoris (Komagataella spp)*, Mxr1, Mit1, *AOX1* promoter, methanol metabolism, heterologous gene dosage, recombinant protein productio

4.1. Introduction

The methylotrophic yeast *Pichia pastoris* (*Komagataella sp.*) is a widely used recombinant protein host and one of the most preferred yeast system for the production of a wide range of proteins, from biopharmaceuticals and industrial enzymes to membrane proteins and complex nanostructures [1]–[5]. Key features, established tools and methodologies for protein production in *P. pastoris* have been extensively reviewed [6]–[9]. One of the most compelling factors in this host is the presence of the strong and tightly regulated promoter of alcohol oxidase 1 (P_{AOX1}), the most frequently used option to drive the expression of foreign genes [10]. Catabolite repression of $P_{AOX1}$and other methanol utilization (Mut) genes by glucose and glycerol, among other C-sources, has been reported since long time [11]. However, the molecular mechanisms of P_{AOX1} transcriptional regulation have been unveiled only over the past 10 years. Lin-Cereghino and coworkers identified Mxr1 (*methanol expression regulator 1*), a transcription factor (TF) with a zinc finger DNA-binding domain, homologous to the TF Adr1 from *Saccharomyces cerevisiae* [12]. Mutations in this gene resulted in a strain unable to grow in methanol and oleate and induced the transcription of alcohol oxidase 1 (*AOX1*), dihydroxyacetone synthase (*DAS*), peroxin 8 (*PEX8*) and peroxin 14 (*PEX14*). Further investigations by Kranthi and coworkers revealed several Mxr1-binding sites in the promoters of *PEX8*, *AOX1* and *DAS* [13], [14]. Subsequently, Parua and coworkers characterized a 14-3-3 family protein (regulator of numerous biological processes in many eukaryotes, [15]) in *P. pastoris* [16]. Furthermore, these authors described a highly conserved yeast 14-3-3 binding motif in Mxr1, revealing that the interaction between both proteins was due to the phosphorylation of Mxr1 Ser215 in a carbon-source dependent manner, leading to the repression of Mxr1-dependent genes. Recently, Mit1 (*methanol-induced transcription factor*), Rop (*repressor of phosphoenolpyruvate carboxykinase*) and Trm1 have also been described as transcription factors responsible for methanol regulation [17]–[19]. Wang and coworkers suggested a transmission of the methanol induction signal among Mxr1, Mit1 and Trm1 through a cascade, being Mxr1 the responsible for the derepression of P_{AOX1}, whereas methanol presence would induce Trm1 and, subsequently, Mit1 expression [19]. Finally, it has been demonstrated that Rop and

Nrg1, are repressors of methanol metabolism, competing with Mxr1 for the same binding sites in P_{AOX1} [17], [20]. New insights in this complex regulatory circuit can provide innovative strategies for recombinant protein production using $P_{AOX1.}$ For instance, Vogl and coworkers have successfully overexpressed *MXR1* and *MIT1* for activation of P_{AOX1} under methanol-free carbon-limiting conditions [21].

In this context, a study by Cámara and coworkers supports the hypothesis that *P. pastoris* strains expressing more than one copy of a lipase of *Rhizopus oryzae* (Rol) encoding gene under the control of P_{AOX1} leads to a transcriptional limitation in the methanol assimilation capacity of such strains [22]. Because of *MXR1* is constitutively expressed at low levels, our observations suggested that the presence of multiple P_{AOX1} copies resulted in an insufficient number of Mxr1 molecules to fully induce the expression of methanol-regulated genes, including the recombinant *ROL* [12]. Coherently with this hypothesis, Takagi and coworkers reported an increase in recombinant protein production due to *MXR1* overexpression [23]. In addition, *MIT1*-deficient strains have marginal *AOX1* mRNA when exposed to methanol, highlighting Mit1 as an essential TF for growth on methanol [19].

In this study, we have co-expressed a deregulated *MXR1* variant, *MXR1** [16], and *MIT1* in a series of *P. pastoris* strains harbouring one and four copies of *ROL* (named 1C and 4C, respectively, [24]. Mxr1* contains a mutation in the Ser215 that completely prevents the binding of the 14-3-3 proteins to this TF, thereby ensuring a constant action of Mxr1. The effect of *MXR1** and *MIT1* co-overexpression on cell growth and Rol productivity of the re-engineered 1C and 4C strains was tested in shake flasks. Furthermore, a comparative transcriptional analysis of 8 key genes involved in different relevant cellular processes between the reference strain 4C and two strains with two additional copies of Mxr1* and one additional copy of Mit1, respectively, was performed in chemostat cultures.

4.2. Methods and materials

4.2.1. Strains and plasmids

P. pastoris strains carrying 1 and 4 copies of the lipase of *Rhizopus oryzae* (named 1C and 4C strains, respectively) were constructed previously by Cámara and coworkers

[24], and together with the non-producing strain X-33 (named 0C), were used as a reference strains. These two strains were used as starting strains for this study. pGAPHA, a pGAPZA-derived plasmid in which the Zeocin resistance marker is replaced by the Hygromycin resistance one [25], was used as *MXR1** and *MIT1* expression vector.

4.2.2. Construction of expression vector pGAPHA_MXR1* and pGAPHA_MIT1

The mutant *MXR1** gene, with the Ser215 replaced by an Ala, was amplified by PCR from the recombinant plasmid pG213 (a gift from E.T. Young, University of Washington) using MXR1*_PmlI and MXR1*_KpnI primers (listed in **Supplementary table I**) and the Q5®Hot Start High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA, USA). After amplification, PCR product was sequenced using MXR1*seq 1-5 primers (**Supplementary table I**). Afterward, *MXR1** was cloned into the pGAPHA vector between the PmlI and KpnI sites, under the control of the *GAP* promoter (P_{GAP}). The resulting plasmid, pGAPHA_MXR1*, was sequenced using the primers pGAP_fw, MXR1*_int_rev and MXR1*seq_5-6 to confirm that insertion was in-frame. Plasmid and PCR product sequencing was carried out by the Servei de Genòmica i Bioinformàtica of the UAB.

MIT1 gene was amplified by PCR from *P. pastoris* genomic DNA using MIT1F and MIT1R primers (**Supplementary table I**). pGAPHA vector was amplified using pGAPHAF and pGAPHAR primers (**Supplementary file I**). Then, pGAPHA_*MIT1* was constructed by using *Infusion cloning* (*Clontech* Laboratories, Becton Dickinson, New Jersey, USA) and sequenced using Mit1Seq1, Mit1Seq2, Mit1Seq3 and Mit1Seq4 primers (**Supplementary table I**).

4.2.3. Transformation of P. pastoris and clone selection

Competent cells were prepared following the protocol described by Cregg [26]. 100 ng of XbaI-linearized pGAPHA_MXR1* and AvrII-linearized pGAPHA_MIT1 were pulsed into *P. pastoris* competent cells by electroporation, using a Gene Pulser Xcell™ Electroporation System (Bio-Rad, Hercules, CA, USA). Instruments settings were 1500 V, 25 µF, and 200Ω. Subsequently, transformants were selected on Yeast Extract Peptone Dextrose (YPD) plates ((1% (w/v) yeast extract, 2% (w/v) peptone, 2% (w/v)

dextrose)) containing 250 µg/mL Hygromycin (Invivogen, San Diego, CA, USA). Positive clones were confirmed by PCR (prior genomic extraction (Wizard® Genomic DNA Purification Kit, Promega, Madison, WI, USA), by means of the amplification of *MXR1** and *MIT1* insert using the pGAP and MXR1*_int and pGAP and MIT1_int primers, respectively (**Supplementary table I**).

4.2.4. Growth conditions

Shake flask cultures

The *P. pastoris* strain cultures in shake flasks were performed in triplicate as shown in Cámara et al. [24]. All media were supplemented with Hygromycin (150 µg/mL).

Chemostat cultures

Chemostat cultures of selected strains were carried out at a working volume of 1 L in a 1.5-L vessel bioreactor (Biostat B Plus, Sartorius AG, Göttingen, Germany). A 1-L shake flask containing 150 mL of YPD-Zeocin or YPD-Hygromycin (for the reference and the MXR1*/MIT1 strains, respectively) medium was inoculated with 1 mL cryostock of the selected strain and incubated for approximately 24 h at 30°C and 150 rpm agitation in a Multitron II orbital shaker, and subsequently used to inoculate a volume of 1 L of batch medium.

Bioreactor conditions and monitoring of the culture were performed as reported by Cámara et al. [24], with a difference: media composition of all components for both, batch and chemostat phase, were doubled.

4.2.5. Analytical procedures

Biomass determination

Cell biomass was monitored by measuring the optical density at 600nm (OD_{600}). For cellular dry weight, the method previously reported by Jordà and coworkers was used [27]. All determinations were performed in triplicate.

Lipase activity assay

The lipolytic activity determination was performed with a colorimetric assay (Roche Diagnostics) in triplicate samples as previously described by Resina and coworkers. The

Specord 200 Plus spectrophotometer (Analytik Jena AG, Jena, Germany) was used to measure absorbance [28].

Metabolite quantification

Glycerol, methanol and other potential extracellular compounds in the shake flask and chemostat cultures were analysed as previously reported by Cámara et al. [24]. Analyses were performed in duplicate for each independent experiment.

Flow cytometry analysis

Cell counting and viability was measured using the methods previously described by Cámara et al. [24]. To determine the cell population size, the Flow Cytometry Size Calibration Kit (Thermo Fisher Scientific, Waltham, MA, USA) was used. Briefly, microspheres suspensions of 1 µm, 2 µm, 4 µm, 6 µm, 10 µm and 15 µm of diameter were sonicated for 5 s at 50W with the ultrasonic processor VC-5 (Vibracell, Sonics & Materials, Newtown, CT, USA) before being subjected to analysis. After that, different thresholds were established for each range in the analysis software to define the interval size of the processed samples. Experiments were performed using the Guava EasyCyte Mini cytometer (Millipore, Jaffrey, NH, USA), with a 488nm Argon laser. An amount of 5,000 cells were measured per analysis at a flow rate of 0.59mL/s. Viability assays were carried out in duplicates, whereas cell counting and cell size determination were performed in triplicate for each shake flask and chemostat culture.

4.2.6. Gene copy number determination by droplet digital PCR (ddPCR)

Gene dosage quantification of *ROL, MXR1* and *MIT1* was analysed by ddPCR using the method previously described by Cámara and coworkers [24], with the reference gene β-actin (*ACT1*) as endogenous control for data normalization and the set of primers for *MXR1*, *MIT1, ACT1* and *ROL* genes described in **Supplementary file I** for DNA amplification. To verify the *ROL* and *MXR1* gene dosage stability, biomass samples of the chemostat cultures were taken after five residence times, once the steady state was achieved. Reagents for ddPCR were purchased to Bio-Rad (Hercules, CA, USA), whereas primers were synthesized by Sigma Aldrich (St Louis, MO, USA).

4.2.7. Measurement of transcriptional levels by ddPCR

For transcript quantification, 5 mL of collected pellets at the end of the shake flask culture were previously treated with phenol (5% v/v) and further stored at -80°C. Total RNA extraction was performed using the Rneasy Mini Kit (Qiagen, Hilden, Germany) whereas cDNA was synthesized by using the iScript™ cDNA Synthesis Kit (Bio-Rad) following the manufacturer's instructions. RNA quality was tested by measuring the ratio of absorbance at 260 and 280 nm using the Nanodrop 1000 (Thermo Fisher Scientific). ddPCR reactions contained 10 µL of QX200™ ddPCR™ EvaGreen Supermix, 200 nM of forward primer, 100 nM of reverse primer, 0.4 ng of cDNA and the required amount of Dnase/Rnase-free water up to 20 µL of final volume. Reactions were incubated at 95°C for 10 min, followed by a denaturation (94°C, 30 s) and an annealing/extension step (60.2°C, 1 min for the *KAR2* and *ROL* primers; 56.5°C for the rest) during 40 cycles. Positive droplets of each PCR reaction were normalized in relation to *ACT1* signal. Primer list is shown in **Supplementary File I**.

4.2.8. Statistics

Data are expressed as the mean ± standard deviation (SD). Statistical analyses of the data were performed using the unpaired Student's t-test using Microsoft's Excel software. A *P*-value lower than 0.05 was considered statistically significant.

4.3. Results

4.3.1. Construction and characterization of Rol-producing strains co-expressing MXR1* and MIT1

In order to increase the Mxr1 levels in the Rol-producing strains, as well as ensuring a constant activation of Mxr1-dependent promoters under induction conditions, an expression vector harboring *MXR1** (encoding for the Mxr1 variant carrying the mutation in the position Ser215) was introduced in the Rol-producing reference strains 1C and 4C. Because previous studies had reported a lethal effect of *MXR1* overexpression under the control of inducible promoters (P_{AOX1} and P_{PEX8}, [12]), but permissive growth when using constitutive promoters [18], [23], P_{GAP} was selected as the promoter to drive *MXR1** expression. Five and eleven transformants were isolated

for 1C and 4C strains, respectively. All 1C-MXR1* and 3 out of 11 for the 4C-MXR1* clones resulted to have the MXR1* expression cassette integrated in their genome. Intriguingly, only three clones of the 1C-MXR1*strain ($1C_1$, $1C_2$ and $1C_3$) were capable of growing in liquid media. We further determined the total *MXR1* (*MXR1+MXR1**) copy number for each clone by ddPCR (**Figure 1**), revealing a population with one copy (1C_MXR1*a and 1C_MXR1*b derived from the 1C strain, and 4C_MXR1*a and 4C_MXR1*b derived from the 4C strain), or two copies of *MXR1** (1C_2MXR1* and 4C_2MXR1*). Concerning *MIT1* co-overexpression, no fitness-related problems were observed in agar plates and liquid medium growth steps during the isolation of these strains, allowing to isolate a series of 1C- and 4C Rol-producing clones containing one extra copy of *MIT1*.

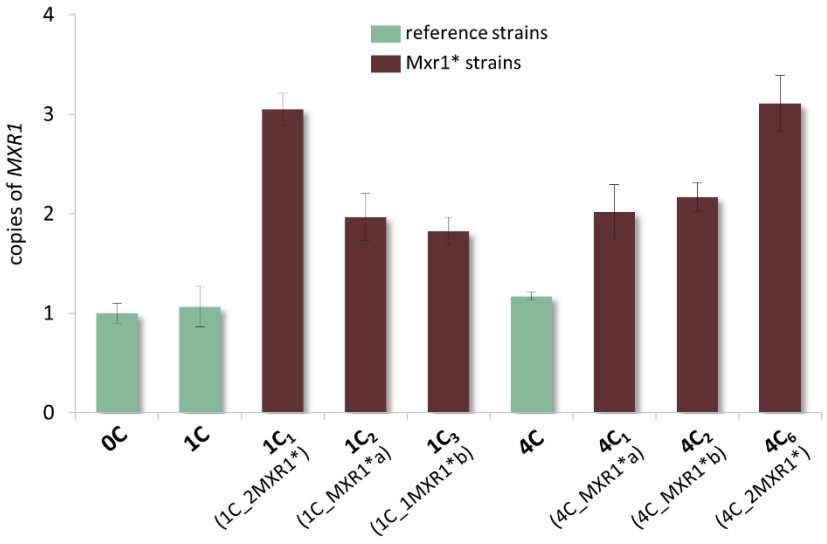


Figure 1. Characterization of *MXR1* copy number. Determination of total *MXR1* (*MXR1+MXR1**) and *MIT1* copy number by ddPCR. Because only one copy of *MXR1* is present in the genome of *P. pastoris*, clones with two or three copies of *MXR1* consequently harbor one or two copies of *MXR1**.

4.3.2. Expression of MXR1* and MIT1 increases Rol production levels and reverts detrimental effect of ROL multi-copy expression on methanol assimilation

As previously reported [22], [24], *P. pastoris* strains carrying more than one copy of a P_{AOX1}-*ROL* expression cassette exhibit a reduced growth and methanol assimilation capacity. A first series of shake flask cultivations was carried out to test the effect of *MXR1** and *MIT1* co-overexpression on cell growth, methanol consumption and Rol production after 70 h of cultivation. As expected, biomass levels of the 4C-Rol-

producing strain were 3-fold lower than those observed for its 1C counterpart (**Figure 2A**). In addition, significant levels of residual methanol were detected at the end of 4C strain culture (**Figure 2C**), while no substrate accumulation was observed in the 1C cultures. Moreover, almost all the cells in the non-producer strain (0C) culture were viable (97%), while a decrease of 20% in viability was observed for 1C and 4C cultures (**Supplementary figure I**). *ROL* overexpression further impacted cell size: whereas a slight (but significant) change in population size profile was observed between strains 0C and 1C, a remarkable increase of larger cell fractions (6µm and ≥10µm) was measured for the 4C strain, compared with the other reference strains (**Supplementary figure I**).

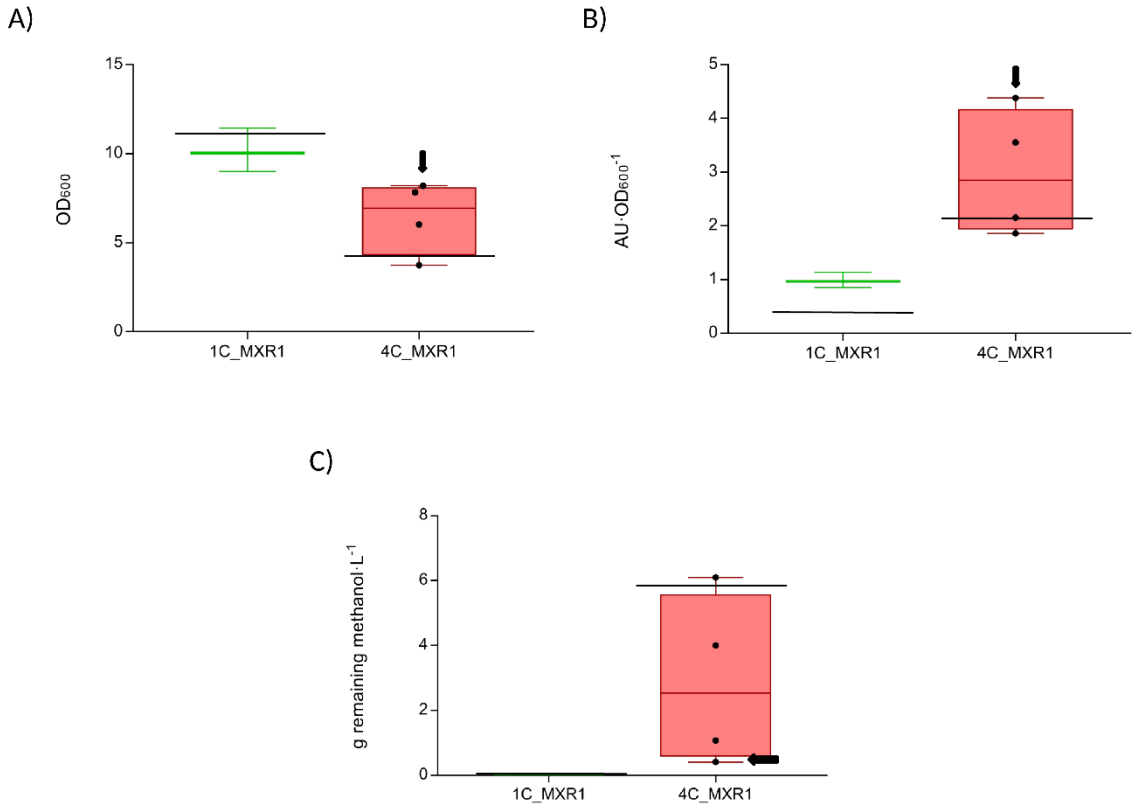


Figure 2. Screening of *MXR1 transformants in shake flask cultures. A**) Biomass levels measured by OD_{600}. **B**) Specific lipase activity in shake flask cultures. **C**) Residual methanol quantified by HPLC at the end of the cultures. Samples were taken after 70 h of culture. 1C and 4C (reference strains) were marked as horizontal lines, whereas 1C and 4C strains expressing *MXR1** were represented as green and maroon plot boxes, respectively.

Co-expression of *MXR1** and *MIT1* in 1C strains had either no or slightly negative impact in cell growth compared to the reference 1C strain, also showing no residual

methanol at the end of the culture. In contrast, the average biomass levels of each series of 4C clones presented 1.73- and 1.75-fold increase, respectively. Consistently, these clones showed also lower average residual methanol at the end of the screening (2.12- and 3.48-fold reduction, respectively). Nonetheless, MXR1* strains still showed a 10-20% of non-viable cells at the end of cultivations, and this value was even higher (around 40% non-viable cells) in the case of the 4C_MXR1*a strain.

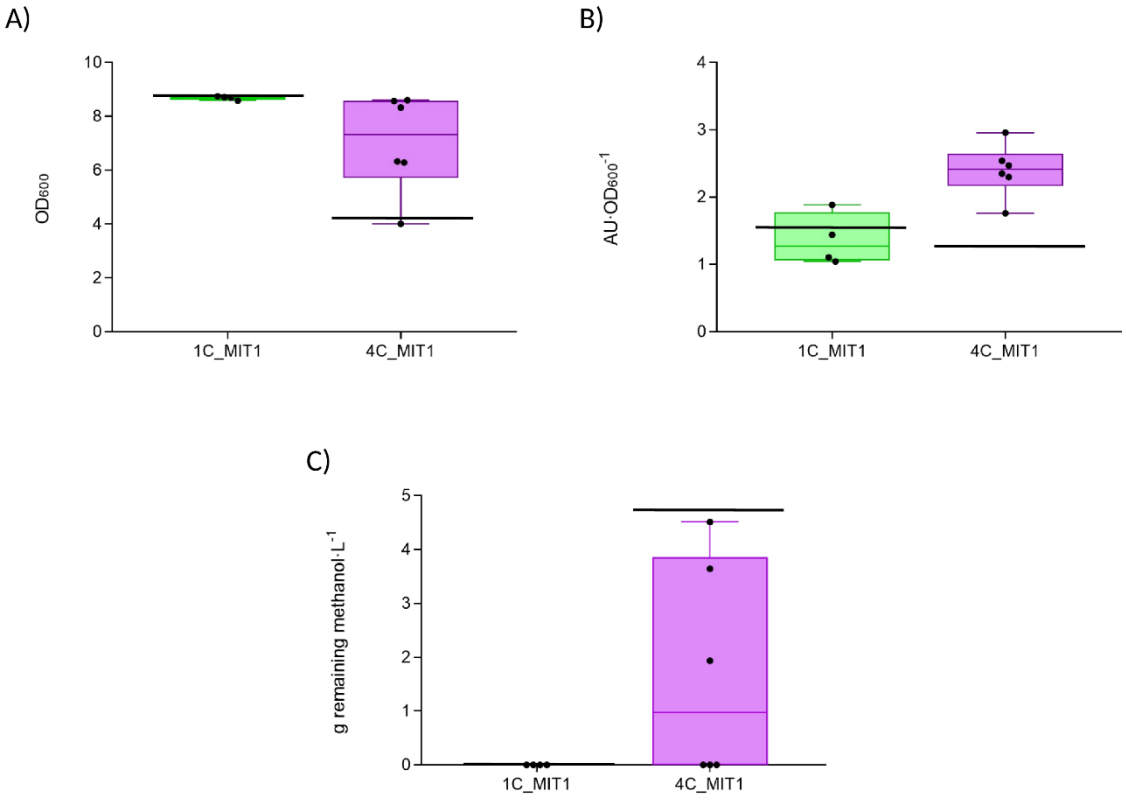


Figure 3. Screening of *MIT1* transformants in shake flask cultures. A) Biomass levels measured by OD_{600}. B) Specific lipase activity in shake flask cultures. C) Residual methanol quantified by HPLC at the end of the cultures. Samples were taken after 70 h of culture. 1C and 4C (reference strains) were marked as horizontal lines, whereas 1C and 4C strains expressing *MIT1* were represented as green and purple plot boxes, respectively.

As previously reported by Cámara et al. [22], the increase in *ROL* gene dosage led to a higher specific productivity in the 4C strain compared to the 1C strain. Interestingly, for all the 1C_MXR1* strains, Rol production was between two- and three-fold higher than the 1C control strain (**Figure 2B**). However, Rol production did not show important differences in 1C_MIT1 strains compared to control (average 20% reduction) (**Figure 3B**). Concerning the re-engineered 4C strains, 4C_MXR1 strains showed an average

1.2-fold increase and 4C_MIT1 strains a 2-fold increase. Although average RoI production improvement in MXR1* strains was not significant, a MXR1 clone exerted a 1.8-fold increase (this clone was further discovered to have 2 extra copies of *MXR1* gene instead of one). As shown in **Figure 2C and 3C**, this increase in the lipase production was concomitant with a sharp decrease in the residual methanol in the medium, reflecting a methanol consumption rate similar to the 1C strains. In fact, strains with higher methanol consumption rate (0C and 1C) presented an average population size lower than those with reduced methanol assimilation capacities (4C and 4C_MXR1*a). Notably, significant differences were also observed in terms of cell size distribution between the strains 4C_MXR1*b and 4C_2MXR1* referred to the 4C strain control (**Figure 2B**). More specifically, a 20% increase was detected for the 4µm-cell population in the 4C-MXR1*strains, while the 6µm-cell fraction was reduced around 15%, altogether resulting in a cell distribution size similar to the 1C strain control, thus indicating a possible reversion of the detrimental effects of *ROL* overexpression on the methanol assimilation capacity of the 4C strain by means of the expression of *MXR1**.

4.3.3. Chemostat cultures reveal a preference change of C-source depending on the strain

To validate the results obtained in the shake flask cultures, the 4C reference strain, a 4C strain expressing 2 copies of MRX1* and a representative clone of a 4C strain overexpressing MIT1 (4C_2MXR1* and 4C_MIT1, respectively) were further grown in chemostat cultures at a D of 0.09 h^{-1} using a mixed glycerol:methanol feed (60/40% w/w), i.e. under conditions analogous to previous studies [22], [24]. In terms of extracellular lipase activity, a 30% higher volumetric productivity was reached for the 4C_2MXR1* strain in comparison to 4C, corresponding to a slight increase of 10% in terms of specific lipase productivity (**Table 1**). Unexpectantly, a reduction of 13% and 28% in the total extracellular lipase activity and specific lipase productivity, respectively, was observed for the 4C_MIT1 strain compared to the reference 4C strain. This might be due to higher tendency to foaming observed for this strain under the tested chemostat conditions, since gas-liquid interphases can influence lipases adsorption and activity. Nonetheless, the cultivations confirmed the positive

physiological impact of both *MXR1** and *MIT1* overexpression in the 4C strain. In particular, biomass levels were about 30% and 20% higher than the 4C strain for 4C_2MXR1* and 4C_MIT1, respectively, also reflected by an increase of 15% in the biomass/substrate yield in both cases (**Table 1**).

Under the stablished culture conditions, chemostats are carbon limited when growing the reference strains 0C and 1C, but residual methanol is observed in the fermentation broth when growing multi-copy strains (i.e. 4C strain). In this experiment, a substantial change was observed regarding the specific C-source consumption rate between 4C and 4C_2MXR1*/4C_MIT1. Specific glycerol consumption decreased about a 30%, from 19.59 for 4C to 1.52 mmol g^{-1} DCW h^{-1} for the 4C_2MXR1* strain and 1.48 mmol g^{-1} DCW h^{-1} for the 4C_MIT1 strain, because despite that no glycerol accumulation was observed in the bioreactor, the increase in the biomass levels modified the assimilation rate. In contrast, the specific methanol consumption rate appeared to increase for both modified strains, from 1.44 to 1.71 and 1.70 mmol g^{-1} DCW h^{-1}, when comparing the 4C with the 4C_2MXR1* and 4C_MIT1 strain, respectively. Nevertheless, some residual methanol (3.7 g/L) was still observed in the culture media (not observed in analogous conditions for the reference strains 0C and 1C).

Table 1. Macroscopic growth parameters and lipase production of 4C and 4C_2MXR1*XR1* strains cultured in chemostat conditions. DCW, Dry Cell Weight, qs, substrate specific consumption rate; $Y_{X/S}$, biomass/substrate yield. Data are shown as means ± standard deviation based on triplicate measurements.

	DCW (g L^{-1})	**Biomass** (mmol g^{-1} DCW h^{-1})	**$Y_{x/s}$** (g C-mol^{-1})	**$q_{s,\ glycerol}$** (mmol g^{-1} DCW h^{-1})	**$q_{s,\ methanol}$** (mmol g^{-1} DCW h^{-1})	**Lipase activity** (UA mL^{-1})	**Lipase activity** (UA g^{-1} DCW)
4C	8.64 ± 0.24	3.37 ± 0.17	13.66 ± 0.38	-1.96± 0.2	-1.44± 0.1	40.13 ± 0.02	4644.33 ± 129.04
4C_2MXR1*	11.12 ± 0.11	3.37 ± 0.17	15.94 ± 0.16	-1.52± 0.2	-1.71± 0.2	57.57 ± 1.23	5177.23 ± 121.89
4C_1MIT1	10,32 ± 0.18	3.02 ± 0.12	16 ± 0.16	-1.48± 0.07	-1.70± 0.03	35.21 ± 0.97	3344.41 ± 286.65

4.3.4. Transcriptional analysis reveals an upregulation of methanol metabolism key genes in the 4C_2MXR1* and 4C_MIT1 strains

To further evaluate if the physiological changes observed in the strains 4C_2MXR1* and 4C_MIT1 compared to the reference strain 4C (i.e. higher biomass levels, increase in the methanol specific consumption rate), corresponded to a transcriptional alteration (or deregulation) of the methanol metabolic pathway, several key genes of

this route were analyzed in cells grown in chemostat cultures (**Figure 4**). Specifically, transcriptional levels of *AOX1*, the gene encoding the enzyme alcohol oxidase 1 were 2.3-fold and 2.1-fold higher for 4C_2MXR1* and 4C_MIT1 than for the reference strain 4C. Similarly, *DAS1* and *FDH1*, encoding for key enzymes of the methanol assimilatory and dissimilatory pathways were also upregulated 3 and 1.9-fold in the strain 4C_2MXR1*, and 1.75 and 1.52-fold in the 4C_MIT1 strain. In relation to *ROL* transcriptional levels, a 5-fold increase was observed with the 4C_2MXR1* strain and a 3.2-fold with 4C_MIT1 strain. However, the specific lipase productivity did not increase similarly and was only a 10% higher than the reference strain for 4C_2MXR1* and even a 30% lower for 4C_MIT1 (**Table 1**), being this last case probably due to increased foaming occurring during 4C_MIT1 chemostats. In addition, as our previous transcriptomic study revealed a high impact of the overexpression of *ROL* on the peroxisomal pentose phosphate regeneration pathway (PPP) [29] and peroxisome biogenesis in *ROL* multi-copy strains [22], the mRNA levels of fructose 1,6-bisphosphate aldolase (*FBA1-2*) and peroxine 6 (*PEX6*) were selected as markers of these pathways. In the case of *FBA1-2*, an increase of 40% and 36% was detected for 4C_2MXR1* and 4C_MIT1 strains, respectively, supporting the role of the peroxisomal PPP in methanol assimilation. Regarding *PEX6* (coding for an ATPase involved in the peroxisomal protein import and crucial for peroxisomal biogenesis, [30]), no significant regulation was detected comparing both strains.

Concerning the mRNA levels of native *MXR1* and *MIT1*, no changes were observed between both modified 4C strains and the reference 4C strain. Hence, the observed increases in total *MXR1* or *MIT1* expression levels can be assigned to the transcription factors encoding genes heterologously expressed under control of GAP promoter. Specifically, the 4C_2MXR1* and 4C_MIT1 strains showed an increase of 55% and 50% in total *MXR1* and *MIT1* transcript, respectively.

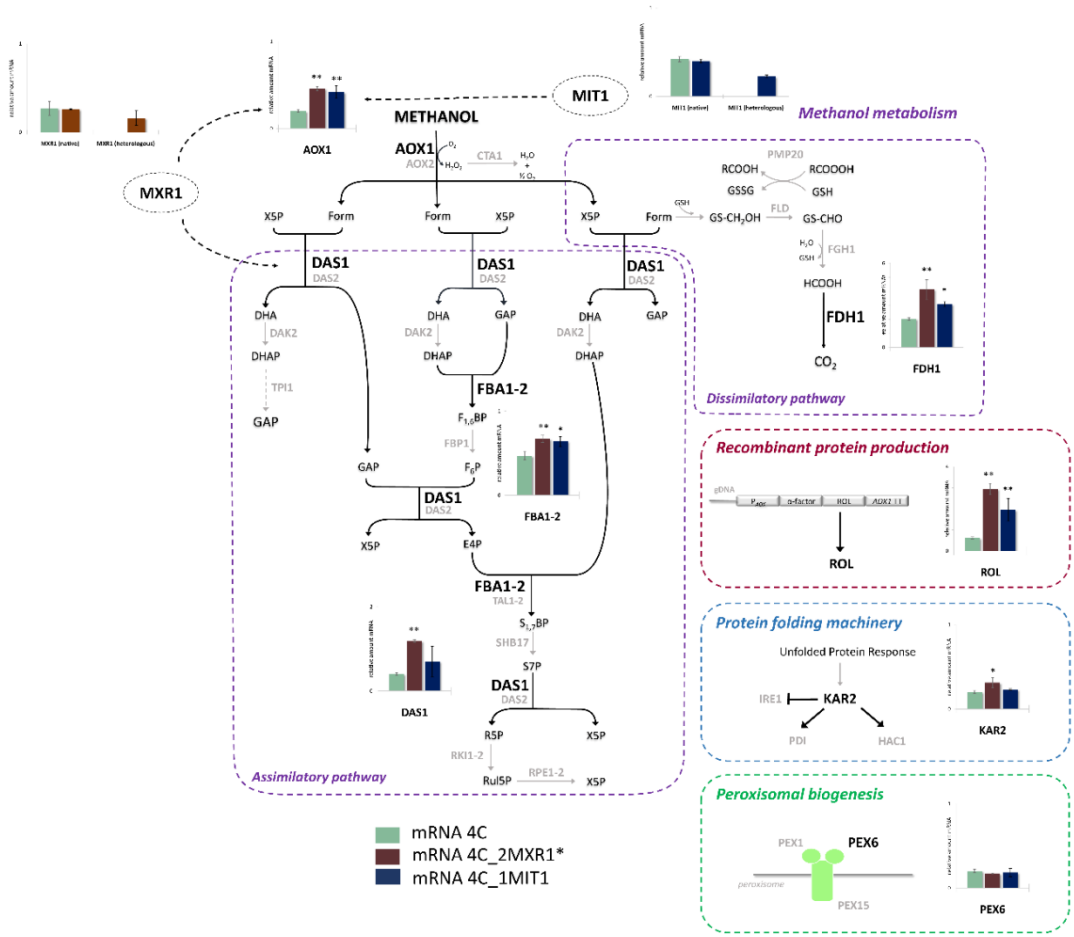


Figure 4. Transcriptional levels of selected genes. Comparative of mRNA levels of 4C (blue) and 4C_2MXR1* (maroon) strains. Relative amounts of mRNAwere calculated comparing to *ACT1* expression levels of each strain. Thicker black lines represent the routes directly controlled by measured genes. DHA, dihydroxyacetone; E4P, erythrose-4-phosphate; HCOOH, formic acid; Form, formaldehyde; F6P, fructose-6-phosphate; GAP, glyceraldehyde-3-phosphate; S1,7BP, sedoheptulose-1,7-bisphosphate; S7P, sedoheptulose-7-phosphate; GS-CH_2OH, S-(hydroxymethyl)gluthiatione, X5P, xylulose-5-phosphate; R5P, ribose-5-phosphate; Rul5P, ribulose-5-phosphate. *MXR1*, methanol expression regulator 1; *MIT1*, methanol-induced transcription factor 1; *AOX1,* alcohol oxidase 1; *DAS1*, dihydroxyacetone synthase; *FDH1,* formate dehydrogenase 1; *FBA1-2*, fructose 1,6 –bisphosphate aldolase; *ROL, Rizhopus oryzae*

Finally, as a typical Unfolded Protein Response (UPR) marker gene, *KAR2* mRNA levels were also measured. Strikingly, the 4C_2MXR1* strain showed an upregulation of about 50% compared to the reference strain, whereas for the 4C_MIT1 strain no significant changes were observed in *KAR2* mRNA levels.

4.4. Discussion

This work proposes an innovative strain engineering strategy to overcome transcriptional limitations in producing strains carrying multiple copies of P_{AOX1} cassettes, by overexpressing a mutant variation of Mxr1 and the wild-type Mit1 transcription factors. Firstly, during the strain construction process, we demonstrated the suitability of P_{GAP} as option to express *MXR1** while maintaining Mxr1 toxicity levels under control. The fact that only clones with low gene dosages of *MXR1** were capable of growth in liquid media could be due to the deleterious effect of higher gene dosages of this gene, i.e. suggesting that clones harboring more than two copies of MXR1* can only grow on agar plates.

In the shake flask cultivations, we demonstrated that the modifications proposed changed cell physiology and provided beneficial advantages for the strain. As previously reported, the strains with higher *ROL* gene dosage exhibited higher lipase specific productivity, but reduced growth and methanol assimilation capacity [22], [24]. Additionally, the fact that only 4C strains showed production and growth improvements with transcription factor deregulation supports the theory that only multicopy strains are transcriptionally limited.

The results obtained from the chemostat cultures also endorse the physiological changes observed at shake flask scale. Interestingly, the presence of some residual methanol in the culture media of 4C_2MXR1* and 4C_MIT1 chemostats may indicate the possibility of further improvement of the methanol assimilation capacity, deserving further studies. Moreover, this might also indicate a variation in the C-source preference of the strains, pointing out to a lower metabolic constraint in the methanol metabolism.

Finally, the transcriptional analysis performed allowed to evaluate the physiological changes observed. The increase in mRNA levels of the methanol metabolism genes was coherent with the higher methanol assimilation capacity of the strains observed in both shake flasks and chemostat cultures. Interestingly, the rise in the transcription of *ROL* did not results in a similar increment in the extracellular lipase production, pointing out to a translational [22] and/or posttranslational limitation [31], [32].

Moreover, increased foaming occurring during 4C_MIT1 chemostats could also partially explain the lower specific product production.

The impact of the overexpression of *ROL* on the peroxisomal PPP was analyzed with the transcriptional results of *FBA1-2*, which support the role of this pathway in methanol assimilation [29]. Regarding transcription of peroxisome biogenesis analysis (*PEX6*), further results are necessary to completely discard any regulation. It would be interesting to analyze whether there were any changes in the expression of *PEX8* and *PEX14*, also coding for peroxines implied in the protein import machinery.

By analyzing the transcriptional levels of *MXR1* and *MIT1*, we demonstrated that there was increase in total *MXR1* and *MIT1* transcripts due to the genes heterologously expressed under control of GAP promoter. These results support the hypothesis that deregulation of methanol metabolism observed in these strains is mainly related to the significant increase in both transcription factors mRNA levels.

It is remarkable that the moderate increase in the transcription levels of *MXR1* and *MIT1* could produce such a strong regulatory response in several genes. The self-regulation of transcription factors has been extensively reviewed in prokaryotes and eukaryotes [33], [34], and even described for Trm1 [19], a TF also involved in methanol metabolism regulation. Our hypothesis points out to an auto regulatory feedback loop by Mxr1 to control the pool of available molecules in these culture conditions. Despite that, the significant regulation detected in the methanol-related genes could be explained by the fact that although the transcriptional efficiency per copy (i.e. the ratio between mRNA levels and the number of copies of a certain gene) of *MXR1* decreased, the mutation introduced in a subpopulation of molecules of Mxr1 (e.g. ensuring a total derepression of P_{AOX1}) was enough to trigger an overexpression of those monitored genes. Although Parua and coworkers demonstrated that there was no 14-3-3 dependent inhibition of methanol utilization (Mut) genes in *P. pastoris* cultures with glycerol or glucose as a C-source, basal phosphorylation of Mxr1 (i.e. promoting the union of 14-3-3 proteins) were observed in the analysis published, even in total inducible conditions (only methanol as a C-source) [16]. Due to the low levels of Mxr1, this basal phosphorylation might severely affect the transcriptional levels of methanol-dependent genes in the multi-copy strains. Hence, our hypothesis points that mutation

in the Ser215, which prevents union of 14-3-3 proteins to Mxr1*, allows for the higher transcriptional levels observed in the strains expressing this *MXR1**, resulting in a better methanol assimilation capacity. Altogether could explain the decrease in methanol residual levels observed for the 4C_2MXR1* strain.

Overall, the results described above allow us to confirm the main role of Mxr1 in the recombinant protein production driven by P_{AOX1}. Due to the constitutive low expression of *MXR1* [10], [12], a limited number of Mxr1 molecules has multiple binding sites available in several genes (*AOX1*, *DAS*, *PEX8*, *ACS1*) [13], [14], [18]. In the case of the *ROL* single copy strain, this Mxr1 pool can be sufficient to accomplish an appropriate methanol assimilation concomitant with a proper recombinant protein expression (**Figure 5A**). However, in the case of recombinant *ROL* multi-copy strains (i.e. with an increased number of P_{AOX1} sequences) the number of available molecules of Mxr1 might be insufficient to deal with the regulation of all the Mxr1-dependent genes and thereby negatively affecting the corresponding transcriptional levels, including *ROL* expression (**Figure 5B**). Hence, this assumption could explain the plateau-like trend observed between gene copy number and mRNA levels of recombinant genes under P_{AOX1} control in other studies [35]–[38].

Furthermore, the underlying mechanisms of P_{AOX1} repression in glycerol culture conditions allowed for maintaining an inducible recombinant protein system, while taking advantage of the *MXR1** expression. Moreover, as Mit1 and Trm1 cooperate with Mxr1 in the regulation of methanol metabolism [19], and, considering that Trm1 is also expressed at constitutive low levels as Mxr1, the Mxr1* overexpression could be further improved by trying to enhance the Trm1 expression levels.

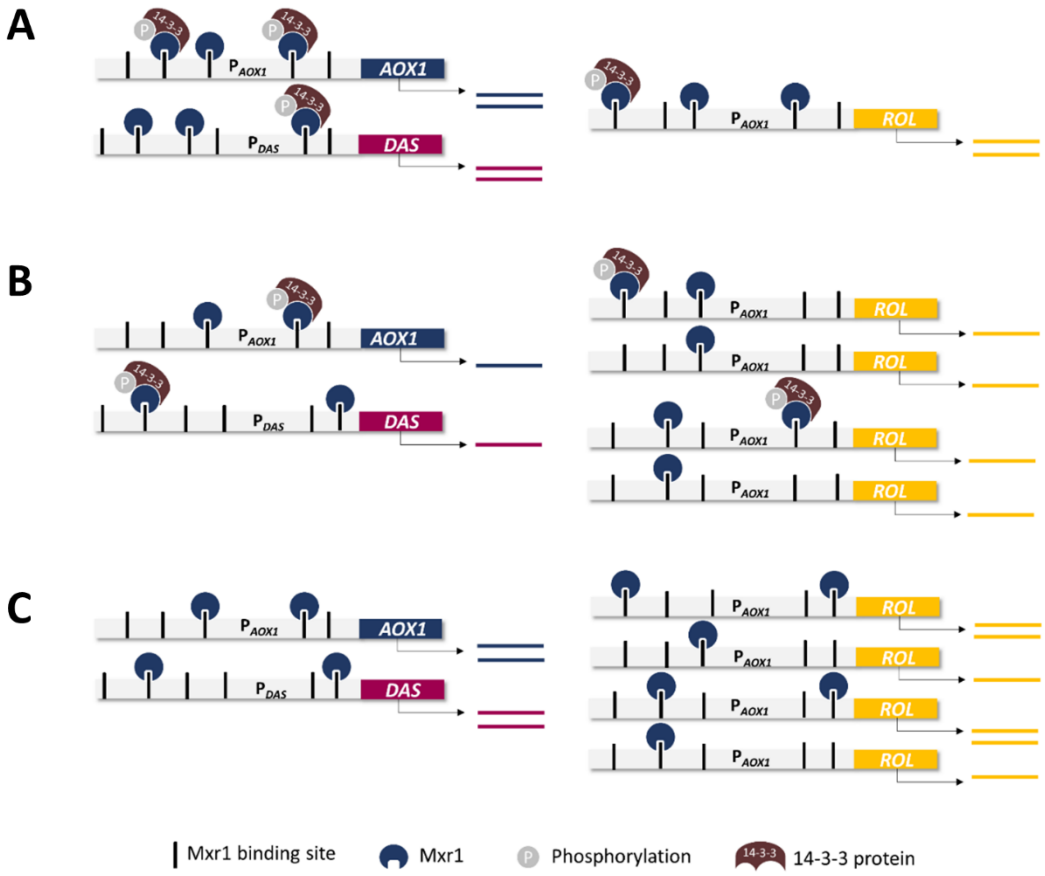


Figure 5. Regulatory model of *MXR1* in recombinant protein production driven by P_{AOX1}. Proposed regulatory model, where the presence of multiple TF binding sites contributes to a higher transcriptional levels of the gene as the number of attached TF increases, due toa synergistic stimulation of gene expression. **A**) Recombinant protein production in single copy strains. A number of limited molecules of Mxr1 are available, which are distributed among the multiple binding sites of the P_{AOX1} and P_{DAS}, among other Mxr1-regulated promoters. **B**) Recombinant protein production in a 4C copy strain. As the number of Mxr1 binding sites increases, transcription levels of endogenous genes controlled by Mxr1 decreased. **C**) Recombinant protein production in 4C strain expressing Mxr1*. Replacement in the position Ser215 by an Ala prevents its phosphorylation and the subsequent union of 14-3-3 proteins. Consequently, higher transcriptional levels of methanol related genes and *ROL* are obtained.

Finally, concerning the UPR marker gene *KAR2*, the observed upregulation in the 4C_2MXR1 strain may indicate a limiting step in folding machinery due to the protein overexpression. Further analyses beyond the scope of this study would be necessary to investigate whether this is a limiting step to couple the increased *ROL* transcription levels with the production of functional lipase.

4.5. Conclusions

Knowledge about the underlying regulation mechanisms of the *AOX1* promoter of *P. pastoris* has increased significantly over past recent years, opening the door to subsequent engineering strategies developments overcoming transcriptional limitations and enhancing the performance of P_{AOX1}-based expression systems.

Notably, in this study we show that overexpression of one of the major transcription factors strongly activating P_{AOX1} in *P. pastoris* (*MXR1* or *MIT1*) was sufficient to reverse the transcriptional attenuation of methanol metabolism caused by the insertion of multiple P_{AOX1}-based expression cassettes, further supporting our initial hypothesis. Furthermore, partial rewiring of P_{AOX1}transcriptional circuits by overproducing a deregulated form of Mxr1 can still preserve a basic output in terms of regulation pattern (derepression under glycerol-limiting conditions and induction by methanol), making the engineered strains suitable for bioreactor-scale cultivations based on mixed carbon source feeding strategies.

Finally, the proposed regulatory model of P_{AOX1} by Mxr1 in the multi-copy strains, brings light to better understand the transcriptional bottlenecks previously observed in several studies on multi-copy strains using methanol-based systems. Further quantitative physiology studies of these new strains could shed light about the global impact of these modifications in *P. pastoris* and, in particular, to novel bottlenecks that may appear downstream transcription still preventing a fully linear increase of secreted product as a function of heterologous gene dosage.

4.6. References

[1] J. L. Corchero *et al.*, "Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics," *Biotechnology Advances*, vol. 31, no. 2. pp. 140–153, 2013.

[2] R. M. Bill, "Recombinant protein subunit vaccine synthesis in microbes: a role for yeast?," *J. Pharm. Pharmacol.*, vol. 67, no. 3, pp. 319–28, 2015.

[3] B. Byrne, "Pichia pastoris as an expression host for membrane protein structural biology.," *Curr. Opin. Struct. Biol.*, vol. 32, pp. 9–17, 2015.

[4] C. Rabert, D. Weinacker, A. Pessoa, and J. G. Farías, "Recombinant proteins for industrial uses: Utilization of Pichia pastoris expression system," *Brazilian Journal of Microbiology*, vol. 44, no. 2, pp. 351-356, 2013.

[5] S. C. Spohner, H. Müller, H. Quitmann, and P. Czermak, "Expression of enzymes for the usage in food and feed industry with Pichia pastoris.," *J. Biotechnol.*, vol. 202, pp. 118–34, 2015.

[6] M. Ahmad, M. Hirz, H. Pichler, and H. Schwab, "Protein expression in Pichia pastoris: Recent achievements and perspectives for heterologous protein production," *Appl. Microbiol. Biotechnol.*, vol. 98, no. 12, pp. 5301–5317, 2014.

[7] O. Cos, R. Ramón, J. L. Montesinos, and F. Valero, "Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast Pichia pastoris under different promoters: A review," *Microbial Cell Factories*, vol. 5., pp. 17, 2006.

[8] B. Gasser *et al.*, "Pichia pastoris: protein production host and model organism for biomedical research.," *Future Microbiol.*, vol. 8, no. 2, pp. 191–208, 2013.

[9] T. Vogl, F. S. Hartner, and A. Glieder, "New opportunities by synthetic biology for biopharmaceutical production in Pichia pastoris.," *Curr. Opin. Biotechnol.*, vol. 24, no. 6, pp. 1094–101, 2013.

[10] T. Vogl and A. Glieder, "Regulation of Pichia pastoris promoters and its

consequences for protein production," *N. Biotechnol.*, vol. 30, no. 4, pp. 385–404, 2013.

[11] J. F. Tschopp, P. F. Brust, J. M. Cregg, C. A. Stillman, and T. R. Gingeras, "Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris," *Nucleic Acids Res.*, vol. 15, no. 9, pp. 3859–3876, 1987.

[12] G. P. Lin-Cereghino *et al.*, "Mxr1p, a Key Regulator of the Methanol Utilization Pathway and Peroxisomal Genes in Pichia pastoris," *Mol. Cell. Biol.*, vol. 26, no. 3, pp. 883–897, 2006.

[13] B. V. Kranthi, R. Kumar, N. V. Kumar, D. N. Rao, and P. N. Rangarajan, "Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in Pichia pastoris," *Biochim. Biophys. Acta - Gene Regul. Mech.*, vol. 1789, no. 6–8, pp. 460–468, 2009.

[14] B. V. Kranthi, H. R. V Kumar, and P. N. Rangarajan, "Identification of Mxr1p-binding sites in the promoters of genes encoding dihydroxyacetone synthase and peroxin 8 of the methylotrophic yeast Pichia pastoris," *Yeast*, vol. 27, no. 9, pp. 705–711, 2010.

[15] H. Fu, R. R. Subramanian, and S. C. Masters, "14-3-3 proteins: structure, function, and regulation.," *Annu. Rev. Pharmacol. Toxicol.*, vol. 40, pp. 617–47, 2000.

[16] P. K. Parua, P. M. Ryan, K. Trang, and E. T. Young, "Pichia pastoris 14-3-3 regulates transcriptional activity of the methanol inducible transcription factor Mxr1 by direct interaction," *Mol. Microbiol.*, vol. 85, no. 2, pp. 282–298, 2012.

[17] N. V. Kumar and P. N. Rangarajan, "The zinc finger proteins Mxr1p and repressor of phosphoenolpyruvate carboxykinase (ROP) have the same DNA binding specificity but regulate methanol metabolism antagonistically in Pichia pastoris," *J. Biol. Chem.*, vol. 287, no. 41, pp. 34465–34473, 2012.

[18] U. Sahu, K. Krishna Rao, and P. N. Rangarajan, "Trm1p, a $Zn(II)_2Cys_6$-type transcription factor, is essential for the transcriptional activation of genes of

methanol utilization pathway, in Pichia pastoris.," *Biochem. Biophys. Res. Commun.*, vol. 451, no. 1, pp. 158–64, 2014.

[19] X. Wang *et al.*, "Mit1 transcription factor mediates methanol signaling and regulates the alcohol oxidase 1 (AOX1) promoter in pichia pastoris," *J. Biol. Chem.*, vol. 291, no. 12, pp. 6245–6261, 2016.

[20] W. Shen *et al.*, "A novel methanol-free Pichia pastoris system for recombinant protein expression," *Microb. Cell Fact.*, vol. 15, no. 1, pp. 1–11, 2016.

[21] T. Vogl *et al.*, "Methanol independent induction in Pichia pastoris by simple derepressed overexpression of single transcription factors," *Biotechnol. Bioeng.*, vol. 115, no. 4, pp. 1037–1050, 2018.

[22] E. Cámara, N. Landes, J. Albiol, B. Gasser, D. Mattanovich, and P. Ferrer, "Increased dosage of AOX1 promoter-regulated expression cassettes leads to transcription attenuation of the methanol metabolism in Pichia pastoris," *Sci. Rep.*, vol. 7, 2017.

[23] S. Takagi, N. Tsutsumi, Y. Terui, and X. Y. Kong, "Method for methanol independent induction from methanol inducible promoters in Pichia," US patent US8236528B2, 2012.

[24] E. Cámara, J. Albiol, and P. Ferrer, "Droplet digital PCR-aided screening and characterization of Pichia pastoris multiple gene copy strains," *Biotechnol. Bioeng.*, vol. 113, no. 7, pp. 1542–1551, 2016.

[25] N. Adelantado *et al.*, "The effect of hypoxia on the lipidome of recombinant Pichia pastoris," *Microb. Cell Fact.*, vol. 16, no. 1, 2017.

[26] J. M. Cregg, "DNA-mediated transformation.," *Methods Mol. Biol.*, vol. 389, pp. 27–42, 2007.

[27] J. Jordà, P. Jouhten, E. Cámara, H. Maaheimo, J. Albiol, and P. Ferrer, "Metabolic flux profiling of recombinant protein secreting Pichia pastoris growing on glucose:methanol mixtures.," *Microb. Cell Fact.*, vol. 11, 2012.

[28] D. Resina, A. Serrano, F. Valero, and P. Ferrer, "Expression of a Rhizopus oryzae lipase in Pichia pastoris under control of the nitrogen source-regulated formaldehyde dehydrogenase promoter.," *J. Biotechnol.*, vol. 109, no. 1–2, pp. 103–13, 2004.

[29] H. Rußmayer *et al.*, "Systems-level organization of yeast methylotrophic lifestyle," *BMC Biol.*, vol. 13, no. 1, pp. 80, 2015.

[30] D. Saffian, I. Grimm, W. Girzalsky, and R. Erdmann, "ATP-dependent assembly of the heteromeric Pex1p-Pex6p-complex of the peroxisomal matrix protein import machinery.," *J. Struct. Biol.*, vol. 179, no. 2, pp. 126–32, 2012.

[31] D. Resina *et al.*, "Engineering of bottlenecks in Rhizopus oryzae lipase production in Pichia pastoris using the nitrogen source-regulated FLD1 promoter," *N. Biotechnol.*, vol. 25, no. 6, pp. 396–403, 2009.

[32] D. Resina, M. Bollók, N. K. Khatri, F. Valero, P. Neubauer, and P. Ferrer, "Transcriptional response of P. pastoris in fed-batch cultivations to Rhizopus oryzae lipase production reveals UPR induction.," *Microb. Cell Fact.*, vol. 6, p. 21, 2007.

[33] S. MacPherson, M. Larochelle, and B. Turcotte, "A Fungal Family of Transcriptional Regulators: the Zinc Cluster Proteins," *Microbiol. Mol. Biol. Rev.*, vol. 70, no. 3, pp. 583–604, 2006.

[34] D. J. Stekel and D. J. Jenkins, "Strong negative self regulation of Prokaryotic transcription factors increases the intrinsic noise of protein expression," *BMC Syst. Biol.*, vol. 2, 2008.

[35] H. Liu, Y. Qin, Y. Huang, Y. Chen, P. Cong, and Z. He, "Direct evaluation of the effect of gene dosage on secretion of protein from yeast Pichia pastoris by expressing EGFP.," *J. Microbiol. Biotechnol.*, vol. 24, no. 2, pp. 144–51, 2014.

[36] C. Sha, X.-W. Yu, N.-X. Lin, M. Zhang, and Y. Xu, "Enhancement of lipase r27RCL production in Pichia pastoris by regulating gene dosage and co-expression with chaperone protein disulfide isomerase.," *Enzyme Microb. Technol.*, vol. 53, no.

6–7, pp. 438–43, 2013.

[37] J. Huang *et al.*, "Improved production of a recombinant Rhizomucor miehei lipase expressed in Pichia pastoris and its application for conversion of microalgae oil to biodiesel," *Biotechnol. Biofuels*, vol. 7, no. 1, pp. 111, 2014.

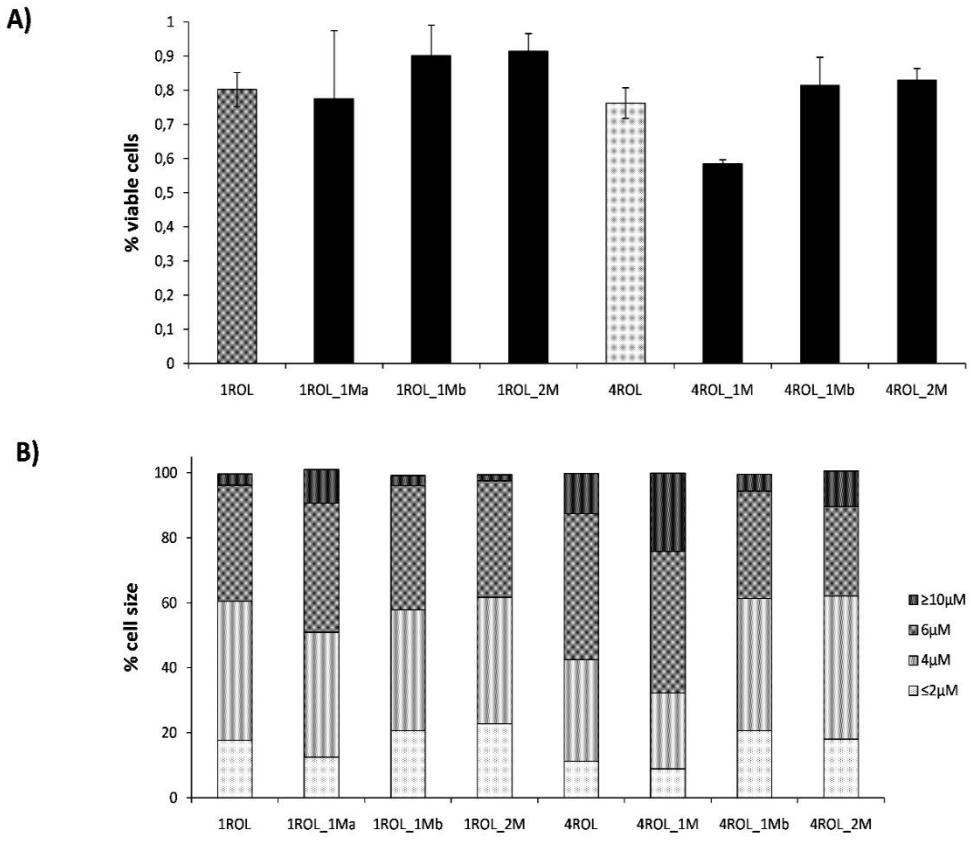
[38] H. Yang *et al.*, "High-level expression of Proteinase K from Tritirachium album Limber in Pichia pastoris using multi-copy expression strains.," *Protein Expr. Purif.*, vol. 122, pp. 38–44, 2016.

Appendix

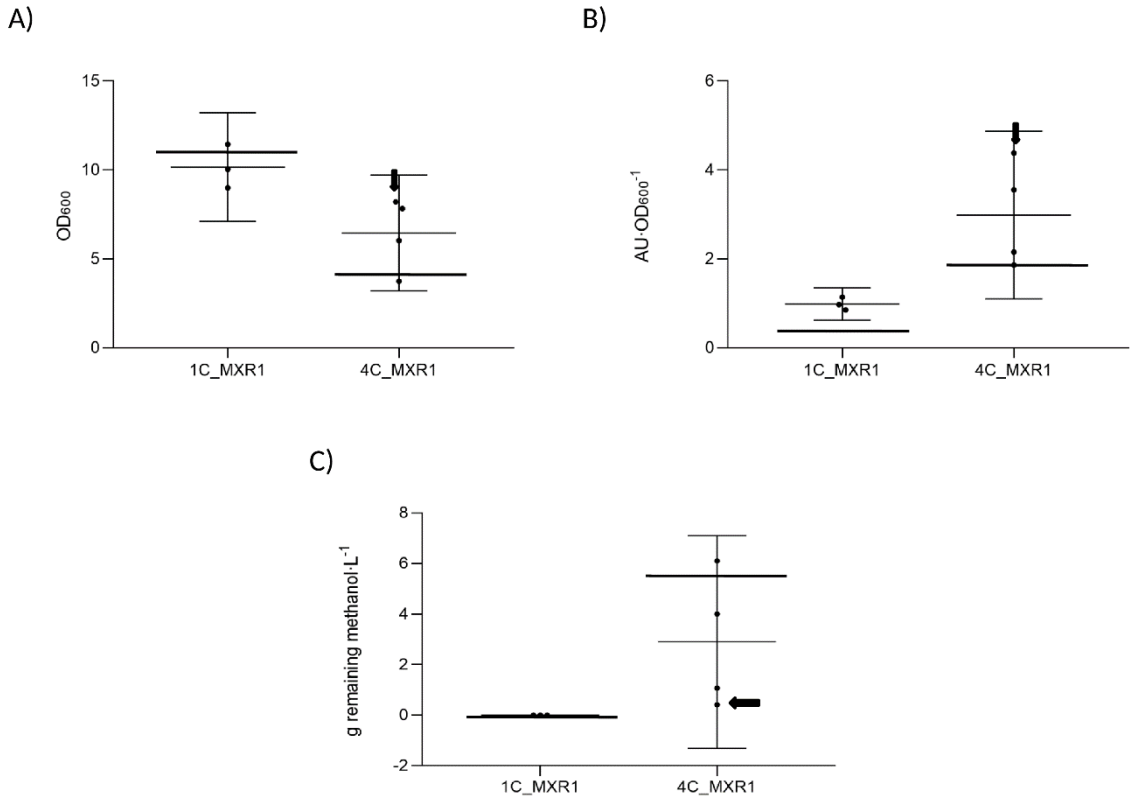
Supplementary table I. List of primers used for cloning and droplet digital PCR (ddPCR)

PRIMER	SEQUENCE (5'-3')	Tm	%GC	AMPLICON LENGHT (bp)
ROL probe	CCGGTCACTCACTCGGTGGTGCA	75	65	-
ACT probe	TCCGTATGGATCGGTGGTTCTATCCTCGCT	75	53	
ROL fwd1	CCTGTCGTCCAAGAACAAC	62	52	164
ROL rev1	GAGGACCACCAACAGTGAAG	62	53	
ACT1 fw	TGTCCGGTGGTACTACTATGTTCC	65	50	199
ACT1 rev	GATTCGTCGTACTCTTGCTTTGA	62	43	
AOX1 fw	GACATTCACGGTTTCGAAGG	61	40	78
AOX1 rev	CCTCAAGAAGTCCTGGCAAAC	63	42	
DAS1 fw	TTGAACTGGGACGGAGTG	60	44	135
DAS1 rev	CAAACCGTTGGCAATAGCAC	62	40	
FAB1-2 fw	CCCTTGGTTTGACGGAATG	60	42	97
FAB1-2 rev	TTCCTCCGACAGGTCTAAC	60	42	
FDH1 fw	GGTGCTGGAAGAATTGG	57	52	135
FDH1 rev	GACAGTGTCGACTCTTC	55	52	
KAR2 fw	GATGAAGTCGGGTCGTGTAC	61	55	110
KAR2 rev	TCTTAGCAGCATCACCCAACC	62	50	
MIT1 fw	CGTGAATCTGCAACAACAGC	57	50	151
MIT1 rev	CGGATCTGAACTGCCAGAG	58	58	
MIT1 fw(inf)	CGAGGAATTCACCGAAACGATGAGTACCGCAGCCCCAATC	71	55	2710
MIT1 rev(inf)	CATGTCTAAGGCTACAAACGATTTCTATTCTTCAACATTCCAGTAG	67	37	
MIT1 seq_1	GGGAAGTTGCATTGCTGAC	57	53	-
MIT1 seq_2	GTCAGCAATGCAACTTCCC	56	53	-
MIT1 seq_3	CTCTGACAAGTTCTCATTGAG	55	43	-
MIT1 seq_4	GGAATTGAACAAACCCTTGACC	56	45	-
MXR1 fw	TTCGCCCATTCTACTCG	57	52	101
MXR1 rev	GGGCTTGACTCATTTCAG	57	52	
pGAP_fw	GATTATTGGAAACCACCAGAATCG	61	41	450
MXR1*_int_rev	AGTCTTGTTATGGCCGCATC	62	50	
MXR1*_PmlI fw	GGCGACACGTGATGAGCAATCTACCCCCAAC	75	48	3490
MXR1*_KpnI rev	CGGCAGGTACCCTAGACACCACCATCTAGTCGG	76	60	
MXR1* seq_1	CCGGCTGTATCTGGATTTAG	59	50	-
MXR1* seq_2	CTCTCGATGTGAACAGGAAC	59	50	-
MXR1* seq_3	GTGGCTACGATAGGTGCAG	61	57	-
MXR1* seq_4	CACAGTTGGGATAGGATG	56	50	-
MXR1* seq_5	CTTTCAGCTGCGGCATTTG	62	52	-
MXR1* seq_6	GCAAAATGGCATTCTGACATCC	62	47	-
pGAPHA fw	GTTTGTAGCCTTAGACATGAC	55	43	3389
pGAPHA rev	CGTTTCGGTGAATTCCTCGTTTC	58	48	
PEX6 fw	GGTTTGGATGTGGTCAAG	57	50	148
PEX6 rev	TGGCTTTAGCAAGCAGAG	59	50	
ROL fwd	AAGTGGGACTGTGTCCAATG	61	50	158
ROL rev	GCACTTCTGAAGGAGTTG	57	50	

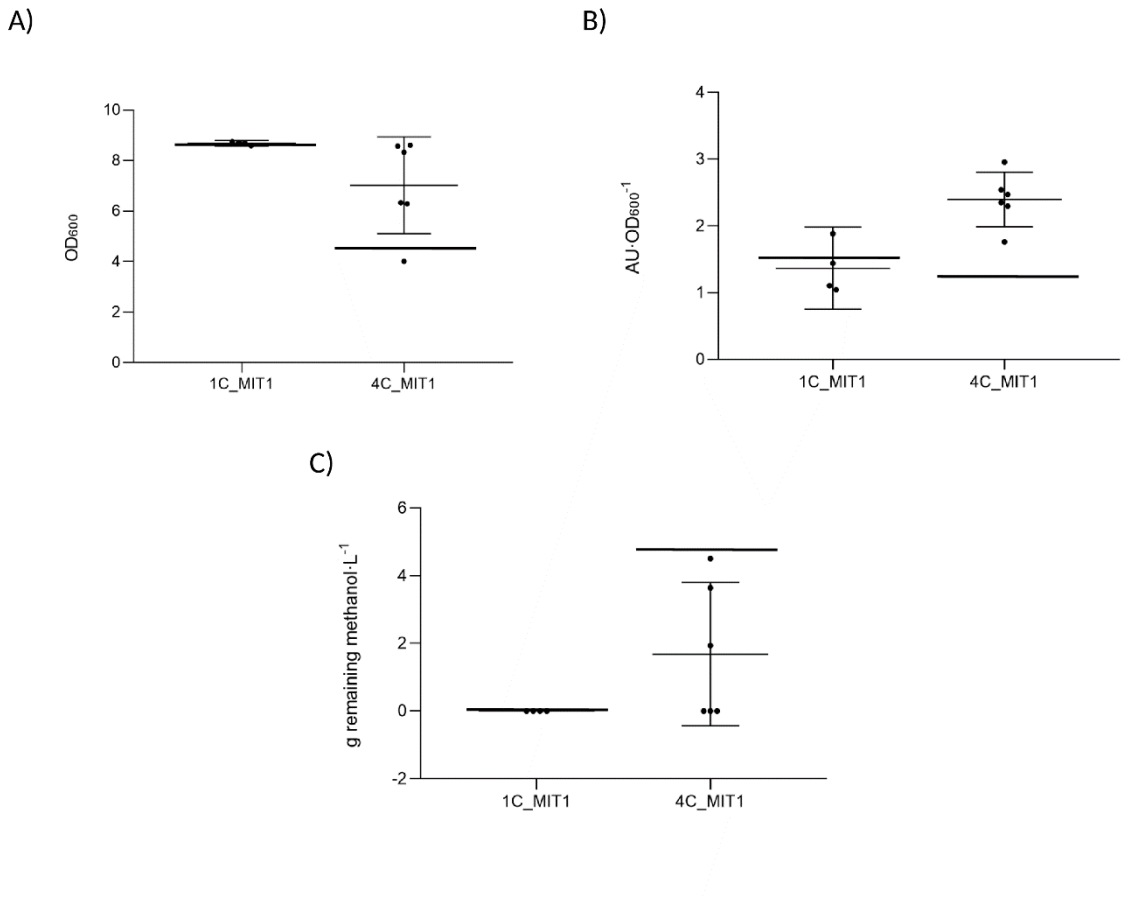
Tm, melting temperature; %GC, percentage of guanosine+cytosine.



Supplementary figure I. Flow cytometric analysis of screening samples. A) Viability of the cells at the end of the screening. B) Cell size distribution of yeast cells.



Supplementary figure II. Screening of *MXR1 transformants in shake flask cultures**. A) Biomass levels measured by OD_{600}. B) Specific lipase activity in shake flask cultures. C) Residual methanol quantified by HPLC at the end of the cultures. Samples were taken after 70 h of culture. Values for reference strains 1C and 4C are marked as horizontal black lines. Black dots indicate values of individual clones of each population. Error bars show an interval with a 95% confidence level.



Supplementary figure III. Screening of *MIT1* transformants in shake flask cultures. A) Biomass levels measured by OD_{600}. B) Specific lipase activity in shake flask cultures. C) Residual methanol quantified by HPLC at the end of the cultures. Samples were taken after 70 h of culture. Values for reference strains 1C and 4C are marked as horizontal black lines. Black dots inside plot boxes indicate values of individual clones of each population. Error bars show an interval with a 95% confidence level.

5. Application of a genome-scale metabolic model-based *in silico* strategy in tandem with the CRISPR/Cas9 system for strain engineering of *Pichia pastoris*

The development of strategies to enhance recombinant protein production of cell factories is a topic of considerable interest for the biotech industry. The multiplicity and complexity of potential factors imposing barriers to achieve high production levels of heterologous proteins, makes their identification and engineering difficult. The yeast *Pichia pastoris* has been extensively studied and engineered from different perspectives in this regard.

Here, we have used a systems-level approach for rational strain engineering based on the utilization of a genome-scale metabolic model to identify hitherto unknown metabolic engineering targets to increase heterologous protein secretion yields, using a *Rhizopus oryzae* lipase as model protein. Specifically, we directed our efforts towards the *in silico* prediction of gene knock-outs and their subsequent construction using the CRISPR/Cas9 system as a genome editing tool. Furthermore, experimental verification of the newly engineered strains has been performed in shake flask and bioreactor cultures.

From our *in silico* analyses, we came to the conclusion that there are two important metabolic processes to which engineering strategies for improved recombinant protein production should be directed, namely: energy/reducing power supply (NADPH generation) and building block supply chain (availability of specific amino acids). However, gene knock-out efficiencies using the CRISPR-Cas9 technique were lower than expected, preventing us from obtaining most of our designed strains, and therefore, our *in silico* hypothesis could not be proved. For this reason, the value of this work mainly relies in the theoretical information it provides, which will be useful for forthcoming studies, including the refinement of the model and simulation tools for improved predictive capacity.

Keywords: *Pichia pastoris (Komagataella spp)*, CRISPR/Cas9, knock-out, genome-scale metabolic model, NADPH, amino acid metabolism, central carbon metabolism

5.1. Introduction

The improvement of computational tools and high throughput techniques in recent years has allowed systems biology to become an important discipline aiming to understand biological systems at a global level, which have highly complex multi-level physiological aspects to take into consideration [1]. Due to the growing biotechnological interest in *P. pastoris*, important advances have been made in this regard and extensive knowledge is available in the different layers of its cell physiology (genomics, transcriptomics, proteomics, lipidomics and fluxomics), as reviewed in [2]–[4]. Application of genome-scale metabolic models (GEMs) can be considered an essential element of systems biology, since they are tools for integration and/or interpretation of the different –omic datasets. The final goal of these models is to be able to predict as closely as possible steady state *in vivo* flux distributions under specific environmental conditions.

The availability of several whole-genome sequences of *P. pastoris* ([5], [6]) allowed the creation of the first GEMs of this yeast [7]–[9]. Notably, the addition of equations for heterologous protein production into these models opened the door to the design of novel strain engineering strategies for improved recombinant protein production [8], [10]. *In silico* approaches provide useful information to hypothesize how cell is affected by the extra load that heterologous protein production exerts from a metabolic perspective. Although their utilization has been wide spreading over the last years, there is so far only one reported application of GEMs for optimization of *P. pastoris* strains [11]. In that study, the authors performed a series of calculations based on a GEMs to propose several gene overexpression and knock-outs with the aim to enhance recombinant protein production using the constitutive *GAP* promoter (P_{GAP})-based expression system [11][11], [12]. Other recent applications of these metabolic models have been focused in -omic data integration for a better characterization of cell physiology or model refinement [12], while their utilization as starting point for rational design of metabolic engineering strategies has been set aside.

In this study, we have used the consensus model iMT1026 v3.0. [10] to identify genes whose deletion could improve the production of our model heterologous protein (*Rhizopus oryzae* lipase, Rol) in an *AOX1* promoter (P_{AOX1})-based expression system.

Once the deletion targets have been proposed, it is necessary to verify the expected results. Thus, new strains with the specified genetic modifications need to be built. However, gene disruption requires efficient homologous of deletion cassettes and this is not always a trivial task. *P. pastoris* has a homologous recombination (HR) machinery far less effective than that of *S. cerevisiae*, even when homologous overlapping regions of several hundred bp are employed [13], [14]. Moreover, non-homologous end joining (NHEJ) machinery is particularly active in *P. pastoris,* which may lead to the random integration of your gene disruption cassette and the presence of false-positive clones [15]. For these reasons, the highly-efficient CRISPR/Cas9 system was used for this work. By co-transforming the recombination cassette with a CRISPR/Cas9 expression vector, HR efficiencies increase greatly, with different results depending on the host species [16]–[18]. The CRISPR/Cas9 system was established relatively late in *P. pastoris* compared to *S. cerevisiae*, but its implementation has been slowly spreading and refined [19]–[21].

In summary, the present work combines prediction of knock-out targets using Minimization of Metabolic Adjustment (MOMA) with the advanced genome editing technique CRISPR/Cas9 to obtain new strains with enhanced extracellular RoI production. Beyond the experimental results obtained, the interest of this study aims at providing ideas for future design of novel metabolic engineering strategies for *P. pastoris*.

5.2. Materials and methods

5.2.1. *In silico* gene targeting for gene knock-out

The iMT1026 v.3.0. GEM was loaded in the OptFlux v.3.3.3 [22] software with IBM® ILOG® CPLEX® Optimization Studio 12.7.1 as solver. As a starting point, a simulation of the reference RoI-producer strain was done by pFBA, using biomass maximization as objective function. Taking this simulation as reference, knock-out predictions were performed using Minimization of Metabolic Adjustment (MOMA), maximizing RoI as objective function. Carbon source uptake was fixed in all simulations to match the experimental chemostat data from Cámara and coworkers [23].

5.2.2. Strains and plasmids

A *P. pastoris* X-33-derived strain containing a single copy of the gene encoding the lipase of *Rhizopus oryzae* was used in this study [23]. Alternatively, a strain created from a *P. pastoris* CBS 7435 by removing the gene *KU70* (named as CBS 12694 and obtained from Central Bureau of Fungal Cultures, Netherlands) was used for construction of new Rol-producing strains [15]. The expression cassette of this gene, constructed using the pPICZαA vector (Invitrogen, California, USA), is under control of the methanol-inducible P_{AOX1} and includes the encoding region of the *S. cerevisiae* α-mating factor secretion signal (its construction has been previously described in [24]).

For deletion of target genes, the CRISPR/Cas9 system was employed as described by Gassler and co-workers [21]. The two components required for this technique (nuclease Cas9 and gRNA) were expressed using BB3_GaT_B3_026 vector [25] (plasmid map is shown in **Supplementary figure I**). The human codon-optimized Cas9 gene and the gRNA were expressed under control of *LAT* (dihydrolipoamide acetyltransferase gene) and *GAP* promoters (P_{LAT}, P_{GAP}), respectively. The vector has a CEN/ARS sequence from *S. cerevisiae* for stable plasmid amplification in yeast. To obtain the gRNA cassette, 6 overlapping primers were assembled using Q5 High-Fidelity DNA Polymerase (New England Biolabs, Massachusetts, USA). The deletion cassettes were created by overlapping extension PCR using the same polymerase. Primers used are shown in **Supplementary table I**. The assembled gRNA was cloned in GaT_B3_026 vector with Golden Gate assembly using the restriction enzymes BpiI and BsaI (New England Biolabs, Massachusetts, USA) and T4 DNA Ligase (Thermo Fischer Scientific, Waltham, Massachusetts, USA).

5.2.3. Strains construction

pPICZαA_ROL vector was linearized using the restriction enzyme PmeI (New England Biolabs, USA) and 100 ng were used for transformation in *P. pastoris* electrocompetent cells. Electroporator settings were adjusted as follows: 1550 V, 25 µF and 200 Ω. Cells were grown in YPD plates with zeocin and colonies were checked by colony PCR.

CRISPR/Cas9 system application was performed as follows: 1000 ng of disruption cassette and 3000 ng of GaT_B3_026 were introduced into *P. pastoris*

electrocompetent cells simultaneously by electroporation. The Gene PulserXcell™ Electroporation System (Bio/Rad, Hercules, CA, USA) was employed for electroporation (2000 V, 25 µF and 200 Ω). Transformants grown in YPD plates with G418 were checked by colony PCR for gene knock-out in the target locus with external primers and for gene reintegration in a different locus with internal primers. Primers for clone checking are described in **Supplementary table I**.

5.2.4. Growth conditions

Shake flask cultures

Clones obtained in the YPD plates were further tested in shake flask cultures. Each clone was cultured in triplicate, as described by Cámara and coworkers [23]. Briefly, the first screening step consisted in an overnight Erlenmeyer culture with 50 mL of Buffered Minimal Glycerol (BMG) medium. Clones were then re-inoculated at exponential phase in 25 mL of Buffered Minimal Methanol (BMM) medium (the amount of biomass was adjusted at an initial optical density of 1). Pulses of pure methanol were added (final concentration of 0.5% v/v) at 24 and 48 h post-inoculation. Samples for optical density and extracellular lipase activity were taken after 70 h of incubation.

Chemostat cultures

A representative clone of each strain was further cultured in a 2-L chemostat bioreactor (Biostat B Plus, Sartorius AG, Göttingen, Germany) with a working volume of 1 L. Cultivation conditions were set to 25°C, pH 5, 700 rpm, 1 vvm inlet gas flow, 0.2 bar overpressure and 0.1 h^{-1} dilution rate (D). The carbon source was a combination of 60% glycerol and 40% methanol (w/w) to emulate the conditions employed in the *in silico* simulations. Concentration of O_2 and CO_2 in the gas outlet was measured using a BlueinOne Gas Analyser (BlueSens, Herten, Germany). Samples were taken at the 4th and 5th residence times.

Batch medium composition: 19.95 g/L glycerol, 0.9 g/L citric acid, 6.3 g/L $(NH_4)_2HPO_4$, 0.01 g/L $CaCl_2$, 0.45 g/L KCl, 0.25 g/L $MgSO_4 \cdot 7H_2O$, 1 mL/L Biotin (0.2 g/L; Sigma,

Munich, Germany), 0.3 mL/L antifoam (Glanapon, Bussetti & Co GmbH, Wien, Austria) and 2.3 mL/L PTM1 trace salts stock solution.

Chemostat medium composition: 15 g/L glycerol, 10 g/L methanol, 0.42 g/L citric acid, 2.18 g/L $(NH_4)_2HPO_4$, 5.5 mg/L $CaCl_2 \cdot 2H_2O$, 0.85 g/L KCl, 0.32 g/L $MgSO_4 \cdot 7H_2O$, 0.5 mL Biotin (0.2 g/L), 50 µL antifoam (Glanapon) and 0.8 mL of PTM1 trace salts stock solution. The PTM1 trace salts stock solution composition was: 6.0 g/L $CuSO_4.5H_2O$; 0.08 g/LNaI; 3.36 g/L $MnSO_4.H_2O$; 0.2 g/L $Na_2MoO_4.2H_2O$; 0.02 g/L H_3BO_3; 0.82 g/L $CoCl_2.6H_2O$; 20 g/L $ZnCl_2$; 65 g/L $FeSO_4.7H_2O$ and 5.0 mL H_2SO_4 (95-98% w/w). Medium pH was set to 5 with 25% HCl.

5.2.5. Analytical methods

Biomass determination

The DR3900 spectrophotometer (Hach, Bizkaia, Spain) was used to measure the culture optical density (600 nm). The protocol employed with pre-weighed dried glass fiber filters (Millipore, Massachusetts, USA) used by Jordà and co-workers was followed for dry cell weight (DCW) quantification [26]. Both measurements were performed in triplicate.

Lipolytic activity assay

A lipase colorimetric assay was used for lipolytic activity determination in the supernatant (Roche Diagnostics). The method was the same reported by Resina and coworkers [27]. Measurements were performed in triplicate.

Quantification of extracellular metabolites

Extracellular metabolites in shake flask and bioreactor cultures were measured using a HPLC (Ultimate 3000 Liquid Chromatography Systems from Dionex combined with an aICSep ICE-COREGEL 87H3 ion exchange column from Transgenomic and a Water 2410 refraction index detector from Waters). CROMELEON software (Dionex) was used for data analysis.

5.3. Results

5.3.3. Prediction of knock-out targets and construction of knock-out strains using CRISPR/Cas9

The iMT1026 v.3.0. GEM developed by Tomàs-Gamisans and co-workers was used in order to identify metabolic targets to enhance RoI production [10], following the procedure described in material and methods section. Six potential genes whose deletion would improve RoI reaction flux were predicted using MOMA (reactions are illustrated in **Figure 1** and genes are briefly described in **Supplementary table II**). Four of these targets were related with central carbon metabolism (glycolysis and tricarboxylic acid cycle (TCA) cycle) and two of them with reactions related to amino acid metabolism (cysteine and serine biosynthesis). Briefly, removal of reactions related with central carbon metabolism have an effect in $NADP^+$ reduction fluxes. Triose-phosphate isomerase (*TPI*) and pyruvate kinase (*PYK*) gene knock-outs would enhance flux through the pentose phosphate pathway (PPP) while suppression of malic enzyme (*MAE*) and isocitrate dehydrogenase (NAD^+-dependent) (*IDH*) genes would allow for an increased flux through the isocitrate dehydrogenase ($NADP^+$-dependent) reaction, both pathways/reactions are important sources of NADPH for the cell. Regarding *MAE* deletion, it is important to indicate that previous studies have demonstrated that flux through the malic enzyme reaction in glycerol cultures is almost negligible [28], [29]. In fact, this knock-out reported the lowest improvement in the *in silico* simulations and was maintained basically as a control target.

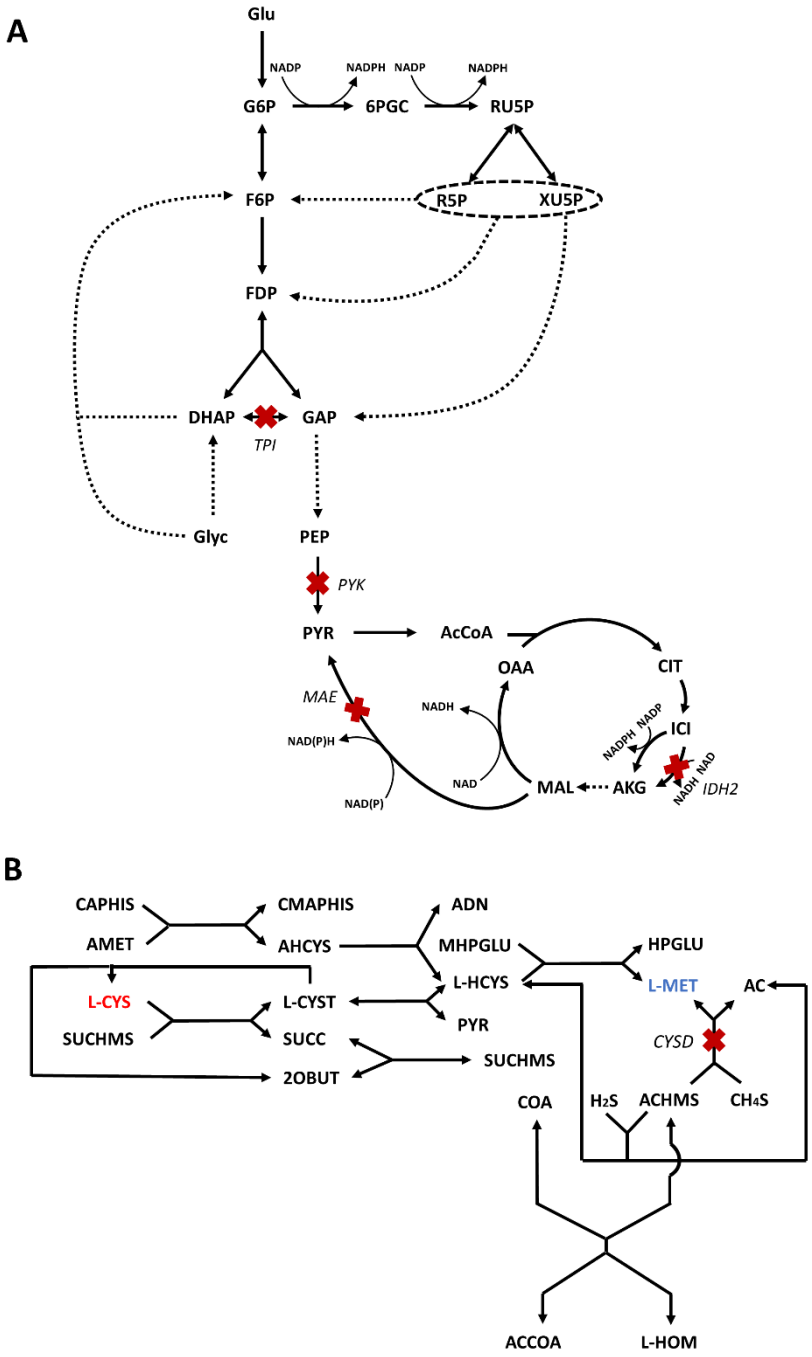

A
Glu
NADP
NADPH
NADP
NADPH
G6P
6PGC
RU5P
F6P
R5P
XU5P
FDP
DHAP
TPI
GAP
Glyc
PEP
PYK
PYR
AcCoA
OAA
CIT
MAE
NADH
NAD(P)H
NAD(P)
NAD
MAL
AKG
ICI
NADPH
NADP
NADH
NAD
IDH2
B
CAPHIS
CMAPHIS
ADN
AMET
AHCYS
MHPGLU
HPGLU
L-HCYS
L-MET
AC
L-CYS
L-CYST
PYR
CYSD
SUCHMS
SUCC
SUCHMS
2OBUT
COA
H2S
ACHMS
CH4S
ACCOA
L-HOM
C
PSER
2AMSA
PYR
Pi
NH4
THF
L-SER
L-CYS
L-CYST
L-HCYS
2OBUT
SHM
MLTHF
GLY
ACCOA
L-ALA
GLX
2AOBUT
COA

Figure 1. Knock-out targets represented in simplified metabolic maps. The different pathways show the central carbon metabolism (A), methionine and cysteine metabolism (B) and serine and glycine metabolism (C). Knock-out targets are illustrated by a red cross (the name of the gene is shown in italics). Dotted arrows are used to summarize several reactions in a single step. Red and blue metabolites represent amino acids whose input fluxes are increased or decreased in the *in silico* analysis, respectively. Glu: glucose; G6P: glucose 6-phosphate; 6PGC: 6-phophogluconate; RU5P: ribulose 5-phosphate; R5P: ribose 5-phosphate; XU5P: xylulose 5-phosphate; F6P: fructose 6-phosphate; FDP: fructose 1,6-diphosphate; DHAP: dihydroxyacetone phosphate; GAP: glyceraldehyde 3-phosphate; Glyc: glycerol; PEP: phosphoenolpyruvate; PYR: pyruvate; AcCoA: acetyl-coenzyme A; CIT: citrate; ICI: isocitrate; AKG: α-ketoglutarate; MAL: malate; OAA: oxaloacetate; CAPHIS: 2-(3-Carboxy-3-aminopropyl)-L-histidine; AMET: S-adenosyl-L-methionine; CMAPHIS: 2-[3-Carboxy-3-(methylammonio) propyl]-L-histidine; AHCYS: S-adenosyl-L-homocysteine; ADN: adenosine; MHPGLU: 5-methyltetrahydropteroyltri-L-glutamate; HPGLU: tetrahydropteroyltri-L-glutamate; L-HCYS: L-homocysteine; L-MET: L-methionine; AC: acetate; L-CYS: L-cysteine; L-CYST: L-cystathionine; SUCHMS: O-succinyl-L-homoserine; SUCC: succinate; SUCHMS: O-succinyl-L-homoserine; 2OBUT: 2-oxobutanoate; ACHMS: O-acetyl-L-homoserine; ACCOA: acetyl-CoA; L-HOM: L-homoserine; PSER: O-phospho-L-serine; 2AMSA: 2-aminomalonate semialdehyde; THF: tetrahydrofolate; L-SER: L-serine; METHF: 5,10-methenyltetrahydrofolate; GLY: glycine; 2AOBUT: L-2-amino-3-oxobutanoate; COA: coenzyme A; *TPI*: triosephosphate isomerase; *PYK*: pyruvate kinase; *IDH2*: isocitrate dehydrogenase 2; *MAE*: malic enzyme; *CYSD*: acetylhomoserine (thiol)-lyase; *SHM:* glycine hydroxymethyltransferase.

On the other side, deletion of genes codifying for the enzymes acetylhomoserine (thiol)-liase and glycine/serine hydroxymethyl transferase (*CYSD* and *SHM*, respectively) increase availability of cysteine and/or serine. Serine is an amino acid that is present in high relative proportion in Rol amino acid sequence. In addition, a previous metabolomic study from Joel and co-workers where amino acid composition of X-33 Rol-expressing strain was compared with a reference strain reported statistically significant increases in the intracellular pools of these amino acids in the Rol-producing strain [30], supporting our hypothesis. Regarding cysteine composition, despite results from a different study of Joel and co-workers did not indicate clearly an increase in cysteine relative composition between reference and Rol-producing strain [26], by direct comparison between the relative levels of cysteine in *P. pastoris* and the percentage of cysteine in the amino acid sequence of Rol, we are able to conclude that cysteine has a higher prevalence in our recombinant protein.

In order to obtain these targeted gene deletions, specific gRNAs and deletion cassettes were constructed as explained in material and methods. However, after several attempts (each involving checking by colony PCR of tens of colonies per knock-out),

only two out of six genes could be successfully removed (*MAE* and *CYSD*). In these two cases, five and four positive colonies were obtained, respectively (**Figure 2A, B**). In the other cases, all clones checked did not recombine the deletion cassette in the target locus or the clones with successful recombination experienced a reintegration event in a new locus of the target gene. This last situation was repeatedly observed for the *TPI* knock-out case (**Figure 2C**). In particular, five positive clones with target *TPI* deletion events were observed but all of them were accompanied by a reintegration of *TPI* into a new locus. Since these unspecific reintegrations were not foreseen when the disruption cassettes were designed, these were created to achieve a complete removal of the target gene sequence. Therefore, gene reintegration implies that gene expression is still possible and the knock-out cannot be considered complete.

With the aim to avoid gene reintegration events during the generation of *TPI* knock - outs, a different approach was applied. X-33 strain was replaced by a CBS 7435-derived strain with the gene *KU70* suppressed. This gene has been proved to be essential for non-homologous end joining (NHEJ) DNA repair in *P. pastoris* [15], and the modified strain has already been tested in combination with CRISPR with good results [20]. Thus, elimination of this route should avoid these undesirable integration processes while performing CRISPR. Firstly, the Δ*KU70* strain was transformed with the same vector used for X-33 Rol-producing strain generation and four clones were obtained with one and two integrated copies of the *ROL* gene. These clones were characterized for its productivity at Erlenmeyer scale (**Figure 3A**) and by recombinant gene copy number determination using droplet digital PCR (ddPCR) (**Supplementary table III**). Clones with higher *ROL* copy number have Rol production levels increased by an average of 1.67-fold in shake flask cultures, coherent with previous studies with the X-33 strain [23]. Regarding the *ROL* single copy clones, the CBS7435-Δ*KU70* strain with one *ROL* copy showed a slightly lower (16%) but non-significant specific Rol production (average of 4 independent clones) compared to our reference X-33 Rol-producing strain. Since differences in recombinant protein levels are not statistically significant, the single *ROL* copy Δ*KU70* clone with highest extracellular specific activity at the end of the screening, was selected for subsequent CRISPR-directed *TPI* gene deletion. Using this strain, we could not observe integration of the deletion cassette in any of the 30

clones tested (example shown in **Figure 3B**). This strongly suggests that integration of the gene disruption cassette results in non-viable or very slow-growing colonies. This would favour the selection of clones having the *TPI* gene reintegrated in another locus when the X-33 strain is used. Base on this observation, the same conclusion was assumed for all the other unsuccessful gene knock-outs. Consequently, the CRISPR-based construction of knock-out strains was ended at this stage due to the high demands in time and resources necessary to carry out new attempts.

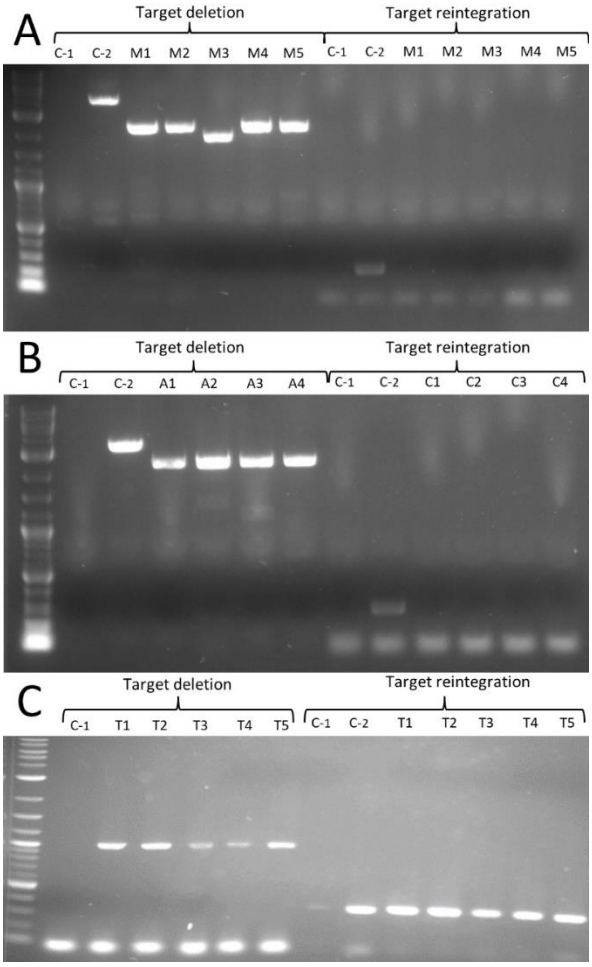


Figure 2. DNA electrophoresis gels of colony PCRs for knock-out checking. $C\text{-}_1$ and $C\text{-}_2$ refer to PCR negative controls using water and WT genomic DNA as template, respectively. **(A)** M1-M5, **(B)** C1-C4 and **(C)** T1-T5 are colonies checked for *MAE*, *CYSD* and *TPI* knock-outs, respectively. Target deletion was determined by using external primers to the integration point (positive clones have a lower band size than $C\text{-}_2$) while possible reintegration of the removed DNA sequence was performed using internal primers for that region (no band should appear if there was not reintegration). For TPI knock-outs, target deletion checking lacks $C\text{-}_2$ but band size corresponds to successful target deletion.

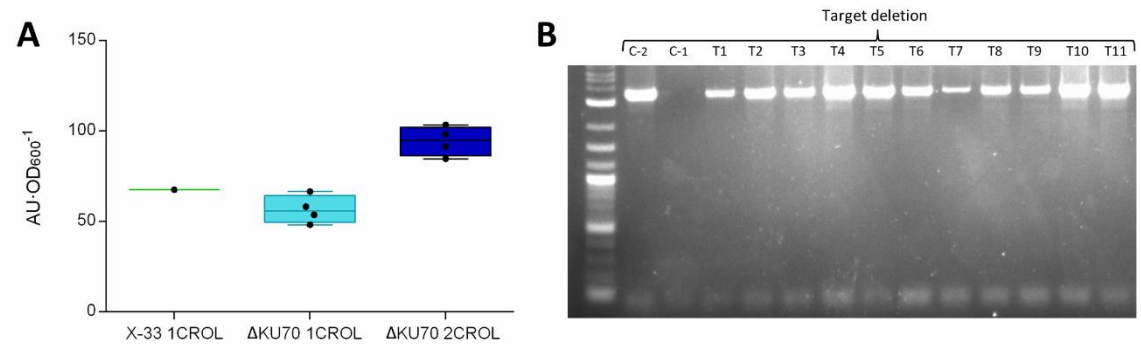


Figure 3. Characterization of *ΔKU70* strains for specific RoI production, compared to the X-33/ROL strain, and DNA electrophoresis gel for knock-out checking of the gene *TPI* in a *ΔKU70* 1CROL strain. Results of specific productivity of RoI-producing clones **(A)** are separated by gene copy number (GCN) of *ROL* gene. The clone which showed higher productivity with a single *ROL* gene was used as backbone strain for knock-out of *TPI* **(B)**. Nomenclature of the electrophoresis wells is the same as in figure 2.

5.3.4. Effect of *MAE* and *CYSD* deletion on recombinant protein production and cell growth performance

All positive clones obtained during CRISPR/Cas9 system rounds of knock-outs (five clones with *MAE* and four clones with *CYSD* deletions) were tested at Erlenmeyer scale and compared with the reference X-33_*ROL* strain (**Figure 4**). The average specific activity of the Δ*MAE* clones at the end of the screening was not significantly different to that of the reference strain. Final production levels reached by the Δ*CYSD* clones were also similar to the X-33_*ROL* strain, showing a non-significant average decrease of 7.4% in specific activity. Also, final biomass (OD_{600}) did not show significant differences between strains neither (data not shown).

In order to confirm the lack of improvement observed in terms of recombinant protein production was not a result of the particular no-steady state Erlenmeyer cultivation conditions, a representative clone of the Δ*MAE* strain was selected for a scale-up at bioreactor operated in chemostat mode. Carbon source composition was adjusted to match with the specific substrate uptake rates employed as fixed inputs in the simulations (i.e. 60% glycerol and 40% methanol, w/w) and μ was set to 0.1 h^{-1}. In this way, metabolic fluxes would be similar enough to those of the simulations and *in silico* knock-out results should be more consistent. The results summarized in **Table 1** indicate that deletion of *MAE* does not change the physiological behavior of the cell in

any observable way. Differences regarding specific lipase activity between strains were lower than 2%. Other macroscopic culture parameters such as biomass yield (Y_{xs}), specific substrate uptake rate (q_s), specific oxygen uptake rate (q_{O2}) and specific carbon dioxide uptake rate (q_{CO2}) did not show significant changes either. Therefore, it can be concluded that, at least the knock-outs that we were able to successfully achieve, the benefits predicted by our simulations were not attained.

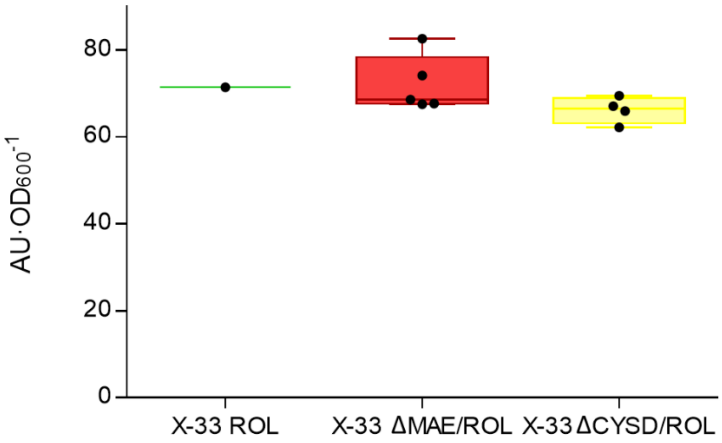


Figure 4. Characterization of knocked strains for specific Rol production. 5 clones of ΔMAE and 4 clones Δ*CYSD* were compared with the reference X-33 Rol-producing strain at Erlenmeyer scale. AU: lipase activity units; OD: optical density.

Table 1. Macroscopic parameters of X-33 ROL and X-33 ΔMAE/ROL strains growing on chemostat.

	DCW	$Y_{x/s}$	Residual methanol	q_{O2}	q_{CO2}	Lipase activity
	(g L^{-1})	(g DCW $g^{-1}s$)	(g L^{-1})	(mmol g^{-1} DCW h^{-1})	(mmol g^{-1} DCW h^{-1})	(AU g^{-1} DCW)
X-33 ROL	12.20±0.61	0.58±0.01	0	4.83±0.11	2.93±0.31	6291±362
X-33 ΔMAE/ROL	11.90±0.12	0.57±0.02	0	4.79±0.09	2.82±0.20	6410±93

DCW: dry cell weight; $Y_{x/s}$: biomass yield; q_{O2}: specific oxygen consumption rate; q_{CO2}: specific carbon dioxide production rate; AU: activity units.

5.4. Discussion

The design of optimal genetic manipulation(s) aiming to maximize the yield of the product of interest is an essential but not simple aspect of metabolic engineering, particularly when the product is a secreted protein. In this regard, constraint-based modeling provides a powerful tool for a rational re-design of cell metabolism. By using a mathematical model of the metabolic pathway of interest and a computational algorithm one can obtain expectations about gene overexpression or knock-out in order to achieve the desired goal, e.g., improve recombinant protein production. Once the gene targets have been selected, one must be able to perform the necessary genetic modifications. For complex genetic manipulations such as gene deletion, which requires a double homologous recombination (HR) event to occur, classical genetic cloning strategies may not be efficient enough in *P. pastoris*, due to its low rate of HR compared to random non-homologous end joining (NHEJ) [15]. In this study, we have implemented an experimental workflow that combines a GEM-driven modeling approach in tandem with the CRISPR/Cas9 genome editing system, with the goal to generate strains having marker-free gene deletions designed to increase RoI production.

The six targets obtained after the simulation experiments point out at two potential limiting factors affecting the recombinant protein production process (NADPH and amino acid availability). Four of the targets predicted were directed towards central carbon metabolism genes (*TPI*, *PYK*, *MAE* and *IDH2*). *PYK*, *MAE* and *IDH2* has been reported as knock-out targets for the first time whereas *TPI* as target of interest for metabolic engineering had been already reported by Nocon and coworkers [11]. As previously mentioned, *in silico* predictions for *TPI* and *PYK* removal would lead to an increase in the flux through the oxidative branch of the PPP, while *MAE* and *IDH2* deletion would imply a boost in the reaction catalyzed by the mitochondrial NADP-dependent isocitrate dehydrogenase, both sources of NADPH. PPP flux would increase NADPH levels in the cytosol whereas the isocitrate dehydrogenase would do a similar role in the mitochondrion. Several studies in different yeast species have indicated that PPP flux increases as an adaptive response to recombinant protein production [31]–[33]. Indeed, the important role of PPP in recombinant protein production has

already been predicted *in silico* and experimentally verified in *P. pastoris* through individual and combinatorial gene overexpression of PPP genes [11], [34]. However, earlier model-based cell engineering studies by Nocon and co-workers were based on the use of glucose as a sole carbon source. In this study we propose for the first time strategies to increase the oxidative branch of PPP in a P_{AOX1}-based expression system. Moreover, this is also the first time that the mitochondrial (NADP-dependent) isocitrate dehydrogenase has been identified as a potential target to metabolically engineer *P. pastoris* for improved NADPH generation.

Regarding the gene knock-outs related with amino acid metabolism (*CYSD* and *SHM*), previous studies have already foreseen changes in the cellular amino acid composition of *P. pastoris* when producing RoI under methanolic-growth conditions [26], [30]. Our target selection is consistent with the increase in the relative quantity of in amino acid composition of X-33/ROL strains that were reported there. Surprisingly, even though RoI has a higher relative amount of cysteine than average amino acid composition of *P. pastoris,* these studies did not report a clear change in cysteine pool in the producing strains.

The increase of amino acid availability with the aim to improve recombinant protein production is a concept that has been already proposed in *S. cerevisiae* and *P. pastoris* [35], [36]. However, these studies were focused on amino acid supplementation in the medium, which may not be always viable from an economical point of view. There is a study that points out to knock-out targets in amino acid metabolism as a strategy to increase the production of their model protein (Human Serum Albumin, HSA) [37]. Particularly, the main targets proposed were directed to increase tryptophan and cysteine, but they never attempted to verify their predictions. Despite the lack of success in our attempts, this work provides the first experimental approach to engineer amino acid pathways through gene disruption in *P. pastoris*.

Unfortunately, most of our gene targets seemed non-feasible or at least extremely hard to achieve even with advanced molecular biology tools such as CRISPR-Cas. Although the CRISPR/Cas9 applicability in *P. pastoris* has been maturing during the

recent years [20], [21], unforeseen factors such as gene locations with hard accessibility for the cell recombination machinery or gene functions affecting cell fitness most probably hindered our efforts. Regarding this last factor, since *TPI*, *PYK* and *IDH2* codify for enzymes involved in important metabolic networks (glycolysis and TCA cycle), it was anticipated that the obtention of knocked-out strains would be difficult, because there had been already unsuccessful attempts to knock-out *TPI* in *P. pastoris* using classical gene deletion techniques [11]. However, we tentatively assumed that it could be possible to obtain at least slow-growing clones by using a more advanced genome editing technique and increasing the number of screened clones. Nonetheless, even the application of CRISPR/Cas9 in a Δ*KU70* strain background resulted unsuccessful in our hands. Concerning the *SHM* knock-out, it is difficult to speculate on the reason(s) why CRISPR/Cas9 did not work, particularly whether or not is related with the function exerted by the coded enzyme, as very few information is available on the phenotype of the knock-out in P. pastoris.

We explored in the *S. cerevisiae* genome database (SGD) in order to correlate the gene essentiality information in there with our *in silico* analysis and observations during strain constructions (**Supplementary table II**). Briefly, *MAE* and *SHM* are reported as non-essential genes; *IDH2* deficient strains show a reduced growth rate; and *PYK, TPI* or *CYSD* defects are associated to a non-viable phenotype. It is remarkable that many selected genes have serious viability problems in *S. cerevisiae* but not in the simulations performed in *P. pastoris*. As previously indicated, we expected that *IDH2* knock-out would be compensated by using its $NADP^+$-dependent isoenzyme, but it is possible that this reaction is not able to incorporate such a high flux and TCA cycle cannot couple with the input flux from glycolysis. *PYK* deletion is most likely stopping flux through glycolysis in *S. cerevisiae*, which is in agreement with its non-viable phenotype. However, in our *in silico* analysis glycolysis can continue by producing pyruvate in a serine deamination reaction. Serine is in turn created from 3-phosphoglycerate, a previous glycolysis metabolite. Although it would be theoretically possible to bypass pyruvate generation this way, it is not probable that these reactions may restore the high fluxes needed in glycolysis for a functionally growing cell. Similarly, the growth defect produced by *TPI* deletion in *S. cerevisiae* has been studied

and explained by a highly reduced flux through glycolysis due to dihydroxyacetone phosphate (DHAP) accumulation [38], [39]. *In silico, P. pastoris* could theoretically reduce the amount of DHAP and restore glycolysis by converting DHAP to glycerol or to the byproduct methylglyoxal. On one side, glycerol would be used for glyceraldehyde generation and this metabolite would enter into glycolysis by condensation of glyceraldehyde and DHAP. Nevertheless, in *S. cerevisiae* it has been reported that this glycerol production does not occur due to a lack of NADH, which is needed in the reactions involved there [39]. On the other side, methylglyoxal would be incorporated in the central carbon metabolism through conversion to pyruvate. At the end, our results during strain construction may indicate that *P. pastoris* is not able to use *in vivo* the previously mentioned paths to compensate DHAP accumulation (two reduced metabolic maps showing the reactions involved in flux redistributions in *PYK* and *TPI* knock-out simulations are available in **Supplementary figure II**). Finally, regarding *CYSD* knock-out, although SGD associates the lack of this gene to an auxotrophic phenotype for organic sulfur sources, it has been demonstrated that this phenotype is viable in many other yeasts such as *Kluyveromyces lactis*, *Yarrowia lipolytica*, *Schizosaccharomyces pombe* and *Trichosporon cutaneum* [40]. In this study, we report that this knock-out is viable in *P. pastoris* as well.

The low success rate in the genome editing part of this study, combined with the fact that the few genetic modifications that could be performed did not show the expected phenotypic improvements, does not allow us to verify the hypothesis derived from our *in silico* analyses. In the case of *CYSD* clones it is hard to discuss the reason why we did not observe the predicted improvements, since few information is available about metabolic fluxes through amino acid biosynthesis pathways. Although the physiological impact of knocking out *MAE* might have been anticipated to some extent, our *in silico* approach for identification and selection of metabolic engineering clearly needs to be refined to improve the predictive potential. Curiously, another work attempted to increase NADPH supply in *P. pastoris* by overexpressing *MAE,* instead of removing it [41]. This discrepancy compared to our hypothesis lies in the cofactor affinity supposed for malic enzyme (while we suppose affinity for both NAD^+ and $NADP^+$, they

assume that the enzyme is $NADP^+$-dependent). At the end, the strain constructed with an extra copy of *MAE* did not demonstrate a higher recombinant protein production yield neither.

5.5. Conclusions

We applied a combination of GEM-driven interpretation of cell behavior together with advance molecular biology techniques with the objective to engineer *P. pastoris* for enhanced recombinant protein production. MOMA was employed as a tool to determine which gene knock-outs may improve the flux through the heterologous protein production while the CRISPR/Cas9 system was used for genetic manipulation due to its well-known high and selective efficiency.

From de targets obtained *in silico*, two thirds of them were directed to increase flux through NADPH-generating reactions while the rest of the targets were intended to increase the supply of certain amino acids in the cell. NADPH is a well-known cofactor involved in many biosynthetic processes including amino acid synthesis, as well as playing a key role in the oxidative protein folding in the endoplasmic reticulum. Our simulations confirm the PPP as a major pathway for NADPH generation and, therefore, being an attractive target to metabolically engineer *P. pastoris*, as previously postulated and experimentally verified by others [11], [34]. In addition, we also point at the relevance of the (NADP-dependent) isocitrate dehydrogenase reaction, a target not previously proposed in the literature. Engineering amino acid metabolism with the purpose of enhancing recombinant protein production is another approach that has not been sufficiently explored. To our knowledge, earlier attempts to customize *P. pastoris* amino acid metabolism for improved recombinant protein production, although promising, did not lead to scale up studies [41].

The high number of knock-out targets that could not be experimentally verified, implied that we were not able to provide evidence for most of the proposed targets. Nevertheless, this work presents new ideas that are potential starting points for new research lines. For instance, application of gene overexpression instead of gene deletion could be a more immediate engineering approach to achieve the same goals.

5.6. References

[1] H. Kitano, "Systems biology: A brief overview," *Science*, vol. 295, no. 5560. pp. 1662–1664, 2002.

[2] R. J. Zahrl, D. A. Peña, D. Mattanovich, and B. Gasser, "Systems biotechnology for protein production in Pichia pastoris," *FEMS Yeast Res.*, vol. 17, no. 7, pp. 1–15, 2017.

[3] P. Ferrer and J. Albiol, "^{13}C-based metabolic flux analysis of recombinant Pichia pastoris.," *Methods Mol. Biol.*, vol. 1191, pp. 291–313, 2014.

[4] D. A. Peña, B. Gasser, J. Zanghellini, M. G. Steiger, and D. Mattanovich, "Metabolic engineering of Pichia pastoris," *Metab. Eng.*, vol. 50, pp. 2–15, 2018.

[5] D. Mattanovich *et al.*, "Genome, secretome and glucose transport highlight unique features of the protein production host Pichia pastoris," *Microb. Cell Fact.*, vol. 8, pp-29, 2009.

[6] K. De Schutter *et al.*, "Genome sequence of the recombinant protein production host Pichia pastoris," *Nat. Biotechnol.*, vol. 27, no. 6, pp. 561–566, 2009.

[7] L. Caspeta, S. Shoaie, R. Agren, I. Nookaew, and J. Nielsen, "Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in silico evaluation of their potentials," *BMC Syst. Biol.*, vol. 6, pp. 24, 2012.

[8] S. B. Sohn *et al.*, "Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production," *Biotechnol. J.*, vol. 5, no. 7, pp. 705–715, 2010.

[9] B. K. S. Chung *et al.*, "Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement," *Microb. Cell Fact.*, vol. 9, pp. 50, 2010.

[10] M. Tomàs-Gamisans, P. Ferrer, and J. Albiol, "Integration and validation of the genome-scale metabolic models of Pichia pastoris: A comprehensive update of protein glycosylation pathways, lipid and energy metabolism," *PLoS One*, vol. 11,

no. 1, 2016.

[11] J. Nocon *et al.*, "Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production," *Metab. Eng.*, vol. 24, pp. 129–138, 2014.

[12] M. Tomàs-Gamisans, A. S. R. Ødum, M. Workman, P. Ferrer, and J. Albiol, "Glycerol metabolism of Pichia pastoris (Komagataella spp.) characterised by 13C-based metabolic flux analysis.," *N. Biotechnol.*, vol. 50, pp. 52–59, 2019.

[13] D. R. Higgins and J. M. Cregg, "Introduction to Pichia pastoris.," in *Methods in molecular biology (Clifton, N.J.)*, vol. 103, pp. 1–15, 1998.

[14] P. Li *et al.*, "Expression of recombinant proteins in Pichia pastoris," *Applied Biochemistry and Biotechnology*, vol. 142, no. 2. pp. 105–124, 2007.

[15] L. Näätsaari, B. Mistlberger, C. Ruth, T. Hajek, F. S. Hartner, and A. Glieder, "Deletion of the pichia pastoris ku70 homologue facilitates platform strain generation for gene expression and synthetic biology," *PLoS One*, vol. 7, no. 6, 2012.

[16] L. Enkler, D. Richer, A. L. Marchand, D. Ferrandon, and F. Jossinet, "Genome engineering in the yeast pathogen Candida glabrata using the CRISPR-Cas9 system," *Sci. Rep.*, vol. 6, 2016.

[17] A. A. Horwitz *et al.*, "Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas," *Cell Syst.*, vol. 1, no. 1, pp. 88–96, 2015.

[18] J. Z. Jacobs, K. M. Ciccaglione, V. Tournier, and M. Zaratiegui, "Implementation of the CRISPR-Cas9 system in fission yeast," *Nat. Commun.*, vol. 5, pp. 5344, 2014.

[19] A. Weninger, A. M. Hatzl, C. Schmid, T. Vogl, and A. Glieder, "Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast Pichia pastoris," *Journal of Biotechnology*, vol. 235. pp. 139–149, 2016.

[20] A. Weninger, J. E. Fischer, H. Raschmanová, C. Kniely, T. Vogl, and A. Glieder, "Expanding the CRISPR/Cas9 toolkit for Pichia pastoris with efficient donor integration and alternative resistance markers," *J. Cell. Biochem.*, vol. 119, no. 4, pp. 3183–3198, 2018.

[21] T. Gassler, L. Heistinger, D. Mattanovich, B. Gasser, and R. Prielhofer, "CRISPR/Cas9-mediated homology-directed genome editing in Pichia pastoris," in *Methods in Molecular Biology*, vol. 1923, pp. 211–225, 2019.

[22] I. Rocha *et al.*, "OptFlux: An open-source software platform for in silico metabolic engineering," *BMC Syst. Biol.*, vol. 4, Apr. 2010.

[23] E. Cámara, J. Albiol, and P. Ferrer, "Droplet digital PCR-aided screening and characterization of Pichia pastoris multiple gene copy strains," *Biotechnol. Bioeng.*, vol. 113, no. 7, pp. 1542–1551, 2016.

[24] S. Minning, A. Serrano, P. Ferrer, C. Solá, R. D. Schmid, and F. Valero, "Optimization of the high-level production of Rhizopus oryzae lipase in Pichia pastoris," *J. Biotechnol.*, vol. 86, no. 1, pp. 59–70, 2001.

[25] R. Prielhofer *et al.*, "GoldenPiCS: A Golden Gate-derived modular cloning system for applied synthetic biology in the yeast Pichia pastoris," *BMC Syst. Biol.*, vol. 11, no. 1, 2017.

[26] J. Jordà, P. Jouhten, E. Cámara, H. Maaheimo, J. Albiol, and P. Ferrer, "Metabolic flux profiling of recombinant protein secreting Pichia pastoris growing on glucose:methanol mixtures.," *Microb. Cell Fact.*, vol. 11, pp. 57, 2012.

[27] D. Resina *et al.*, "Engineering of bottlenecks in Rhizopus oryzae lipase production in Pichia pastoris using the nitrogen source-regulated FLD1 promoter," *N. Biotechnol.*, vol. 25, no. 6, pp. 396–403, 2009.

[28] A. Solà, P. Jouhten, H. Maaheimo, F. Sánchez-Ferrando, T. Szyperski, and P. Ferrer, "Metabolic flux profiling of Pichia pastoris grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates," *Microbiology*, vol. 153, no. 1, pp. 281–290, 2007.

[29] J. Jordà *et al.*, "Glucose-methanol co-utilization in Pichia pastoris studied by metabolomics and instationary 13C flux analysis," *BMC Syst. Biol.*, vol. 7, pp. 17, 2013.

[30] J. Jordà, H. C. Rojas, M. Carnicer, A. Wahl, P. Ferrer, and J. Albiol, "Quantitative Metabolomics and Instationary 13C-Metabolic Flux Analysis Reveals Impact of Recombinant Protein Production on Trehalose and Energy Metabolism in Pichia pastoris.," *Metabolites*, vol. 4, no. 2, pp. 281–99, 2014.

[31] R. Gonzalez, B. A. Andrews, J. Molitor, and J. A. Asenjo, "Metabolic analysis of the synthesis of high levels of intracellular human SOD in Saccharomyces cerevisiae rhSOD 2060 411 SGA122," *Biotechnol. Bioeng.*, vol. 82, no. 2, pp. 152–169, 2003.

[32] Pedersen, Carlsen, and Nielsen, "Identification of enzymes and quantification of metabolic fluxes in the wild type and in a recombinant aspergillus oryzae strain," *Appl. Environ. Microbiol.*, vol. 65, no. 1, pp. 11–19, 1999.

[33] H. Driouch, G. Melzer, and C. Wittmann, "Integration of in vivo and in silico metabolic fluxes for improvement of recombinant protein production," *Metab. Eng.*, vol. 14, no. 1, pp. 47–58, 2012.

[34] J. Nocon *et al.*, "Increasing pentose phosphate pathway flux enhances recombinant protein production in Pichia pastoris," *Appl. Microbiol. Biotechnol.*, vol. 100, no. 13, pp. 5955–5963, 2016.

[35] J. F. Görgens, W. H. van Zyl, J. H. Knoetze, and B. Hahn-Hägerdal, "Amino acid supplementation improves heterologous protein production by Saccharomyces cerevisiae in defined medium.," *Appl. Microbiol. Biotechnol.*, vol. 67, no. 5, pp. 684–91, 2005.

[36] J. Heyland, J. Fu, L. M. Blank, and A. Schmid, "Carbon metabolism limits recombinant protein production in Pichia pastoris," *Biotechnol. Bioeng.*, vol. 108, no. 8, pp. 1942–1953, 2011.

[37] F. Saitua, P. Torres, J. R. Pérez-Correa, and E. Agosin, "Dynamic genome-scale metabolic modeling of the yeast Pichia pastoris," *BMC Syst. Biol.*, vol. 11, no. 1,

2017.

[38] C. Compagno, L. Brambilla, D. Capitanio, F. Boschi, B. M. Ranzi, and D. Porro, "Alterations of the glucose metabolism in a triose phosphate isomerase-negative Saccharomyces cerevisiae mutant," *Yeast*, vol. 18, no. 7, pp. 663–670, 2001.

[39] K. M. Overkamp, B. M. Bakker, M. A. H. Luttik, J. P. Van Dijken, and J. T. Pronk, "Metabolic Engineering of Glycerol Production in Saccharomyces cerevisiae," *Appl. Environ. Microbiol.*, vol. 68, no. 6, pp. 2814–2821, 2002.

[40] J. Brzywczy and A. Paszewski, "Role of O-acetylhomoserine sulfhydrylase in sulfur amino acid synthesis in various yeasts," *Yeast*, vol. 9, no. 12, pp. 1335–1342, 1993.

[41] H. Rußmayer, "The Impact of Amino Acid Metabolism on Recombinant Protein Production in Pichia pastoris," *Phd Thesis*, 2015.

Appendix

Supplementary table I. Primers used for CRISPR/Cas9 and colony PCR

PRIMER	SEQUENCE (5'-3')	Comments
CR_rib_1_MAE1_Fw	GATAGGTCTCCCATGCAATCTCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly (KO MAE1)
CR_rib_2_MAE1_Fw	AAACGAGTAAGCTCGTCAGATTGCAACGCACGCTTTTGTTTTAGAGCTAGAAATAGCAAG	
CR_rib_1_ICDH(NAD)_Fw	GATAGGTCTCCCATGCGCTACCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly (KO ICDH)
CR_rib_2_ICDH(NAD)_Fw	AAACGAGTAAGCTCGTCGTAGCGTAAGATCTGGTGAAGTTTTAGAGCTAGAAATAGCAAG	
CR_rib_1_TPI_Fw	GATAGGTCTCCCATGATCCTCCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly (KO TPI)
CR_rib_2_TPI_Fw	AAACGAGTAAGCTCGTCGAGGATCAGGAAAGAGAGTGGTTTTAGAGCTAGAAATAGCAAG	
CR_rib_1_CYSD_Fw	GATAGGTCTCCCATGTTTGCACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly (KO CYSD)
CR_rib_2_CYSD_Rv	AAACGAGTAAGCTCGTCTGCAAAGTGTCGAAGTGAGAGTTTTAGAGCTAGAAATAGCAAG	
CR_rib_1_HMGS_Fw	GATAGGTCTCCCATGTCTCCACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly (KO HMGCS)
CR_rib_2_HMGS_Rv	AAACGAGTAAGCTCGTCTGGAGATTGTTAACACTTGTGTTTTAGAGCTAGAAATAGCAAG	
CR_rib_1_SHM_Fw	GATAGGTCTCCCATGATTAGACTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTC	gRNA assembly (KO SHM)
CR_rib_2_SHM_Rv	AAACGAGTAAGCTCGTCTCTAATCTGGCAATGATGAGGTTTTAGAGCTAGAAATAGCAAG	
CR_rib_1_gRNAall_Rv	CGCCATGCCGAAGCATGTTGCCCAGCCGGCGCCAGCGAGGAGGCTGGGACCATGCCGGCC	gRNA assembly (all KOs)
CR_rib_2_gRNAall_Rv	GATAGGTCTCCAAGCAGTCCAAAGCTGTCCCATTCGCCATGCCGAAGCATGTTGCCCAGCC	
CR_rib_3_gRNAall_Rv	AGGCTGGGACCATGCCGGCCAAAAGCACCGACTCGGTGCCACTTTTCAAGTTGATAACG	
CR_rib_4_gRNAall_Fw	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGT	
MAE1_KO_HR1_Fw	TGCGGGCATCGCGCATTTCTATGC	Amplify HR1 of deletion cassette
MAE1_KO_HR1_Rv	GATGGTATATCAGAGTTCCTTAGAAAGATAGCGGTATTGTTGG	
MAE1_KO_HR2_Fw	AGGATTCCGCTACTGGCATGATCCGGG	Amplify HR2 of deletion cassette
MAE1_KO_HR2_Rv	GCATAGAAATGCGCGATGCCCGCACCGAAATCCATACGCATCAGTC	
MAE1_KO_check_OHR2_Fw	GTCTCTACCAGTTTGGTCATGG	Check KO
MAE1_KO_check_OHR1_Rv	GTCTTGAAGTGGGCGATCC	
MAE1_KO_check_gene_Fw	GCTCTTCTTCCACTCTGGC	Check gene reintegration
MAE1_KO_check_gene_Rv	GCTGTTCCTGTTGATGTTATGC	
ICDH_KO_HR1_Fw	AGTTGAACGGATTGCTCTAGAATATTGTCTGAACATG	Amplify HR1 of deletion cassette
ICDH_KO_HR1_Rv	CAGTGAATAGGTTTGCTTGGTGTGGATC	
ICDH_KO_HR2_Fw	GGCTTATATGTGACCCTTATTTCAACTCGTCC	Amplify HR2 of deletion cassette
ICDH_KO_HR2_Rv	CATGTTCAGACAATATTCTAGAGCAATCCGTTCAACTCAGCTATGCCGACAAGATCGAAAACTCTG	
ICDH_KO_check_OHR2_Fw	CTGCTGCTATTGACTTTAGAACCAGAC	Check KO
ICDH_KO_check_OHR1_Rv	CACTAGTTTGACTCAACTGAAGGCTC	
ICDH_KO_check_gene_Fw	GTGATTCAACAACACCTGGAACAGAC	Check gene reintegration
ICDH_KO_check_gene_Rv	GCACCAAAGAGTTCCAATCGAGTGG	
TPI_KO_HR1_Fw	GCAGTAAGATGAGGTGATGTGGATC	Amplify HR1 of deletion cassette
TPI_KO_HR1_Rv	CTTATTTAAGAGCCCGCCAAACG	

TPI_KO_HR2_Fw	GAAGCAAAACGCAAGATCCCAGC	Amplify HR2 of deletion cassette
TPI_KO_HR2_Rv	GATCCACATCACCTCATCTTACTGCGAACATATTACTACTCTATA TTCGGG	
TPI_KO_check_OHR2_Fw	CGAGTTTCTAGTTGACAAGACTGAATGG	Check KO
TPI_KO_check_OHR1_Rv	CATAGCTGATATTGCTTATCCGTCACG	
TPI_KO_check_gene_Fw	GATCCACCGTACAAGATACGGAC	Check gene reintegration
TPI_KO_check_gene_Rv	GCTTTGAAGGACTTGGGTGTCC	
CYSD_KO_HR1_Fw	GAAGCATTACATCGAGGGGTTTCAATGTTGAC	Amplify HR1 of deletion cassette
CYSD_KO_HR1_Rv	CTCTCCTCAGTAGCTCTTATTACATCATCGGTTATTAACTG	
CYSD_KO_HR2_Fw	CCATCACAGACACAGCTTGCATCTTTTAATACGAC	Amplify HR2 of deletion cassette
CYSD_KO_HR2_Rv	GTCAACATTGAAACCCCTCGATGTAATGCTTCGAACGGTTTCGG TGCTGTATTGTCTTTTGG	
CYSD_KO_check_OHR2_Fw	GCAGGTGCTGTTGGAAATAACATTC	Check KO
CYSD_KO_check_OHR1_Rv	CATGGTTGGGCAAGTGAAGCGAAC	
CYSD_KO_check_gene_Fw	CCGTGACCATTGATCCACTTAG	Check gene reintegration
CYSD_KO_check_gene_Rv	CTTCAAGAGACTGGGAATTGAATCC	
SHM_KO_HR1_Fw	ATTTGCTAAACATTGTCTATTATGAACTATTTTGTTTCAAG	Amplify HR1 of deletion cassette
SHM_KO_HR1_Rv	CGTGCACTAACAGTCAATGGC	
SHM_KO_HR2_Fw	CTGCTTCCGGTTCTTTGGG	Amplify HR2 of deletion cassette
SHM_KO_HR2_Rv	CTTGAAACAAAATAGTTCATAATAGACAATGTTTAGCAAATGCA GGAGAATTCCCATTGTCTGTTTAG	
SHM_KO_check_OHR2_Fw	GATATTTCAACGACACCCGCTTC	Check KO
SHM_KO_check_OHR1_Rv	GCCTTCCTAGTTCAGTTACAGCAG	
SHM_KO_check_gene_Fw	GGACGGTATAGAACTGCAGTCTTCTC	Check gene reintegration
SHM_KO_check_gene_Rv	CACTTCCACTTCGGTGTTCGATG	

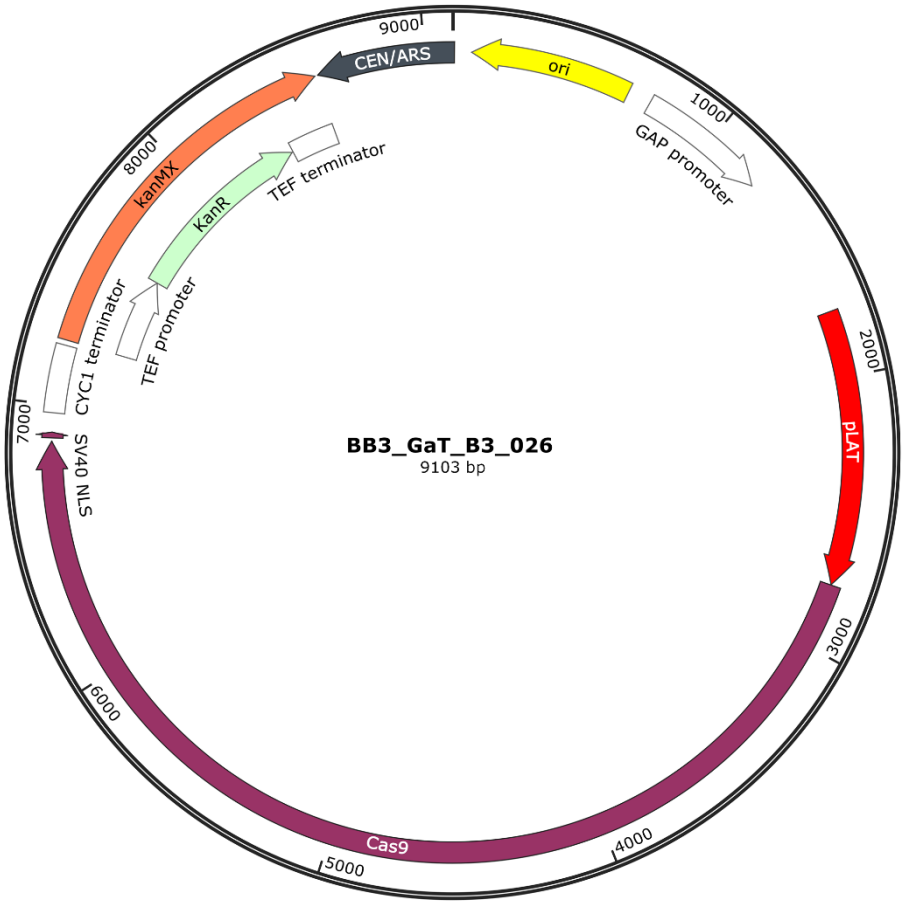
Supplementary table II. List and brief description of the candidate genes to be eliminated for RoI overproduction.

Gene name	Accession number	*S. cerevisiae* homologous gene	Function	Knock-out phenotype (according to SGD)
MAE	PAS_chr3_0181	YKL029C	Mitochondrial malic enzyme; catalyzes decarboxylation of malate to pyruvate	Viable
TPI	PAS_chr3_0951	YDR050C	Triose phosphate isomerase, catalyzes the reversible interconversion of glyceraldehyde 3-phosphate and dihydroxyacetone phosphate	Non-viable
PYK	PAS_chr2-1_0769	YAL038W	Pyruvate kinase, catalyzes conversion of phosphoenolpyruvate to pyruvate	Severe growth defect
IDH2	PAS_chr2-1_0120	YOR136W	Subunit of mitochondrial NAD(+)-dependent isocitrate dehydrogenase complex, catalyzes the oxidation of isocitrate to α-ketoglutarate in the TCA cycle	Growth defect
CYSD	PAS_chr4_0330	YLR303W	O-acetyl homoserine (thiol)-liase, catalyzes the the incorporation of sulfide into O-acetylhomoserine to form homocysteine, used in synthesis of sulfur-containing amino acids (cysteine and methionine)	Auxotroph for methionine, cysteine and homocysteine
SHM	PAS_chr4_0415	YLR058C	Serine hydroxymethyltransferase, converts serine to glycine and 5, 10 methylenetetrahydrofolate (precursor of purine, pyrimidine, amino acid and lípid biosynthesis)	Viable

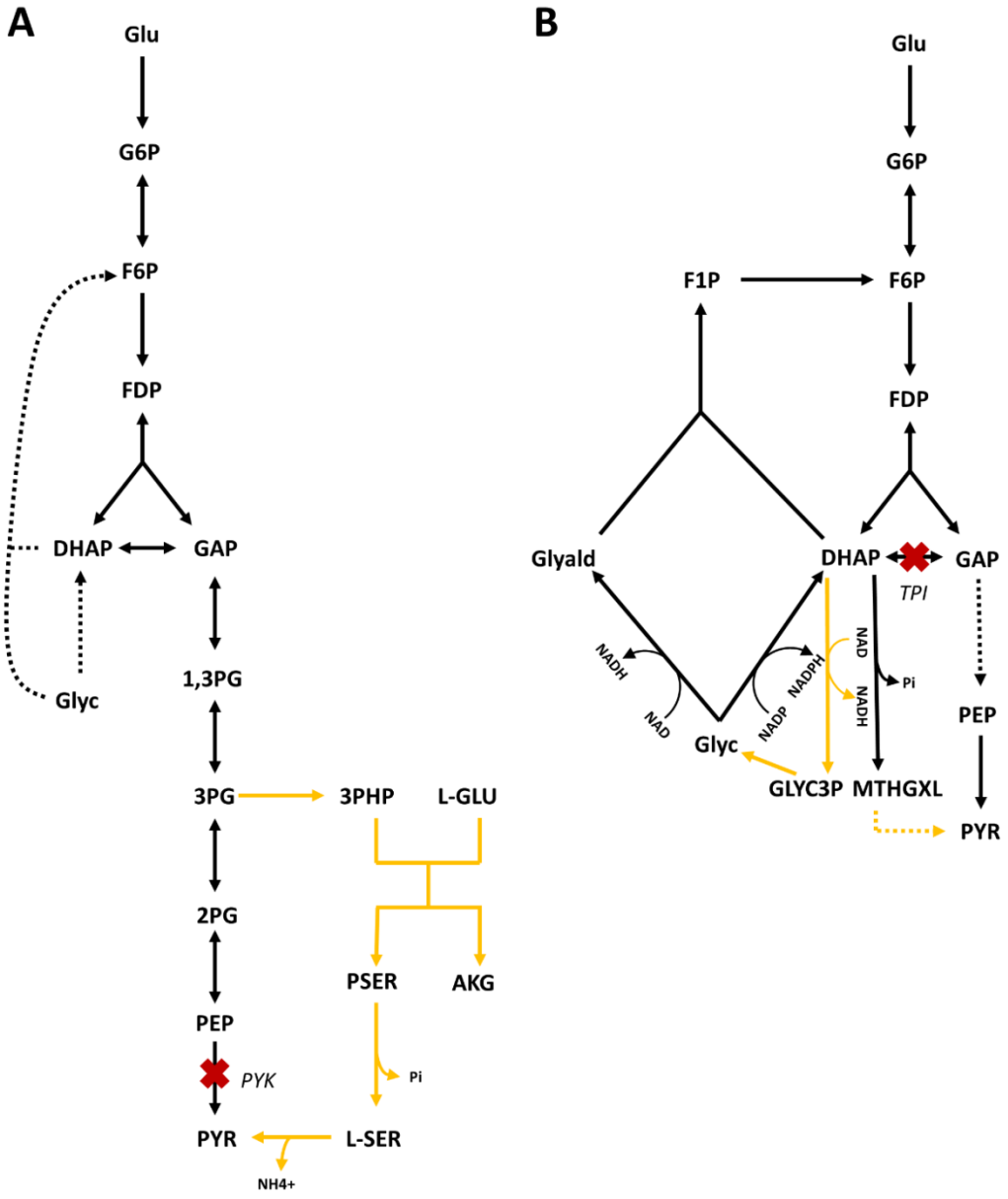
Supplementary table III. Gene copy number determination of *ROL* through droplet digital PCR (ddPCR)

		Gene copy number (Ratio *ROL* vs *ACT*)	SD	RSD (%)
Strain CBS 7435_Δ*KU70*_*ROL*	**Clone 1**	0.87	0.15	0.17
	Clone 2	0.96	0.09	0.09
	Clone 3	1.05	0.12	0.11
	Clone 4	1.21	0.1	0.08
	Clone 5	1.06	0.04	0.04
	Clone 6	2.12	0.21	0.10
	Clone 7	1.84	0.15	0.08
	Clone 8	2.09	0.32	0.15
	Clone 9	2.05	0.03	0.01
	Clone 10	2.31	0.21	0.09

ROL: Rhizopus oryzae lipase gene*; ACT:* actin gene; SD: standard deviation; RSD: relative standard deviation.



Supplementary figure I. Plasmid map of BB3_GaT_B3_026. The Cas9 expression cassette is under control of P_{LAT}. Cas9 gene has the simian virus 40 (SV40) nuclear localization signal at 3'. CEN/ARS sequence from *S. cerevisiae* allows plasmid replication in yeast. kanMX gene is kanamycin resistance marker in bacteria (*Escherichia coli*) and geneticin G418 resistance marker in yeast (*P. pastoris*).



Supplementary figure II. Simplified metabolic map showing the alternative pathways used in the *in silico* analysis of *PYK* (A) and *TPI* (B) knock-outs. Knock-out targets are shown by red crosses (the name of the gene is shown in italics), dotted arrows are used to summarize several reactions in a single step and yellow arrows indicate reactions used to redistribute fluxes. Glu: glucose; G6P: glucose 6-phosphate; F6P: fructose 6-phosphate; FDP: fructose 1,6-diphosphate; DHAP: dihydroxyacetone phosphate; GAP: glyceraldehyde 3-phosphate; Glyc: glycerol; 1,3PG: 1,3-phosphoglycerate; 3PG: 3-phosphoglycerate; 3PHP: 3-phosphohydroxypyruvate; L-GLU: L-glutamate; PSER: phosphoserine; AKG: α-ketoglutarate; L-SER: L-serine; 2PG: 2-phosphoglycerate; PEP: phosphoenolpyruvate; PYR: pyruvate; F1P: fructose 1-phosphate; Glyald: glyceraldehyde; GLYC3P: glycerol 3-phosphate; MTHGXL: methylglyoxal; *PYK*: pyruvate kinase; *TPI*: triosephosphate isomerase.

6. Engineering redox metabolism in recombinant *Pichia pastoris* for enhanced protein secretion under methanolic growth conditions

Recombinant protein production is a high-demanding process in terms of energy and redox homeostasis. Yeast such as *Pichia pastoris* seem to adapt its metabolism to some extent by increasing ATP, NADH or NADPH supply to compensate the metabolic burden exerted for such costly process. Previous studies of our group have proven that heterologous overexpression of the NADH kinase encoding *POS5* gene from *S. cerevisiae* in a recombinant *P. pastoris* has a positive effect in the production of a Fab under the constitutive GAP promoter (P_{GAP}). Moreover, the *in silico* analysis performed and explained in the previous chapter also point to NADPH availability as a limiting factor for recombinant protein production in Rol-producing strains.

In this study, we have further investigated the effect of overexpressing S. cerevisiae's *POS5* and an alternative oxidase (*AtOX*), in a *P. pastoris* strain producing the *Rhizopus oryzae* lipase (Rol) under control of the alcohol oxidase 1 (*AOX1*) promoter (P_{AOX1}), thereby assessing the impact of increased NADPH availability on recombinant protein production under methanol growth conditions. Small-scale cultures allowed us to conclude that genetic modifications increase the protein production significantly. The two strains with higher product yield (one with 2 copies of a cytosolic *POS5* and another with a single copy of *AtOX*) were selected for further physiological characterization in chemostat bioreactor cultures using different mixtures of methanol and a multicarbon source to corroborate its performance. Finally, an *in silico* interpretation of the results using a genome-scale metabolic model and a transcriptomic analysis of different metabolic marker genes was also done to gain further insights on the physiologic effect of these modifications.

Keywords: *Pichia pastoris,* NADH kinase, alternative oxidase, *ROL*, *POS5*, *AOX,* redox cofactor balance, genome-scale metabolic model

6.1. Introduction

High-level expression of heterologous genes in *Pichia pastoris* represents a significant metabolic burden to the cell due to the drain of resources for the heterologous protein production, removed from the cellular maintenance and growth [1]. The metabolic stress caused by overexpression of recombinant proteins is an important drawback that negatively impacts productivity [2], [3]. Such metabolic burden may be triggered by several factors: *de novo* synthesis of energetically costly amino acids, overloaded protein folding machinery, posttranslational protein modifications and secretion processes. These lead to energetic and redox balance alterations and a suboptimal cell physiological state to support recombinant protein production.

Several metabolic engineering strategies have been applied to overcome such limitations. Redox metabolism has been successfully engineered to improve *P. pastoris* recombinant protein production [4]–[6]. For instance, NADPH has been proven to have an important role in this regard [7]. A study by Nocon and co-workers demonstrated that the overexpression of the pentose phosphate pathway enzymes improved protein yield, in the *GAP* promoter (P_{GAP}) expression system, by increasing NADPH supply [5]. Recently, we have investigated the effect of overexpressing a heterologous NADH kinase encoded by the *POS5* gene from *Saccharomyces cerevisiae* [8] on a *P. pastoris* strain producing an antibody fragment under the control the constitutive P_{GAP} with positive results [9]. Another successful approach consisted in the overexpression of a heterologous NADH oxidase from *Lactococcus lactis* [6]. By oxidizing NADH to NAD^+, the resulting increased NAD^+ availability in the cytoplasm allowed to improve CALB production in methanol-grown *P. pastoris* cultures. In the previous chapter of this thesis we aimed to increase NADPH generation in *P. pastoris* by using gene knock-out strategies but without success, mainly due to low efficiency in the generation of new producing strains. Therefore, we focused on achieving the same goal by a much simpler knock-in strategy.

In this study, we aimed at investigating the effect of cofactor regeneration on RoI production and central carbon metabolism of *P. pastoris* growing under methanol conditions, by overexpressing a heterologous NADH kinase (*POS5*) from *Saccharomyces cerevisiae* and an alternative oxidase (*AtOX*) from *Histoplasma*

capsulatum [10], [11]. NADH kinase simply provides extra NADPH for recombinant protein production. An alternative oxidase is an enzyme located in the inner mitochondrial membrane and provides an alternative route to transfer electrons from the ubiquinol pool to oxygen, bypassing this way several proton-pumping steps. Thus, the alternative oxidase would perform a similar function in the mitochondria than that of a NADH oxidase into the cytosol.

Firstly, the redox-engineered strains were tested at shake flask scale and then further characterized in 1-L scale chemostat cultures. Secondly, to gain a deeper insight in the physiological effect of these modifications, a transcriptomic analysis from chemostat cultivation samples and an *in silico* interpretation of the effect of these genetic modifications were also performed.

6.2. Materials and methods

6.2.1. Strains and plasmids

A *P. pastoris* X-33-derived strain containing a single copy of the gene encoding the lipase of *Rhizopus oryzae* was used in this study [12]. The expression cassette of this gene, constructed using the pPICZαA vector (Invitrogen-Thermo Fisher Scientific, California, USA), is under control of the methanol-inducible *AOX1* promoter (P_{AOX1}) and includes the encoding region of the *S. cerevisiae* α-mating factor secretion signal.

The genes encoding the mitochondrial *S. cerevisiae* NADH kinase Pos5p (*POS5*) and *H. capsulatum*'s alternative oxidase (*AtOX*) were codon-optimized for *P. pastoris* and synthetized by Geneart (ThermoFischer Scientific) and cloned into a pPUZZLE vector [13] under control of GAP promoter. In addition, the first 49 bp of *POS5* ORF, encoding region of the mitochondrial signal peptide, were removed, i.e. ensuring the cytosolic location of the corresponding protein product, and cloned in the pPUZZLE vector. Overall, 3 vectors were created for this study: pPUZZLE_mPOS5, pPUZZLE_cPOS5 and pPUZZLE_AtOX (**Supplementary figure I**).

6.2.2. Strains construction

For genomic integration of the expression vector, the plasmids were linearized using AvrII (New England Biolabs, Massachusetts, USA) and transformed in *P. pastoris* using 100 ng of DNA. Transformation was performed by electroporation with a Gene PulserXcell™ Electroporation System (Bio-Rad, Hercules, CA, USA) using as parameters: 1550 V, 25 µF and 200 Ω. Positives clones were selected by growing on Yeast Peptone Dextrose (YPD) plates with 500 µg/mL of G418 (Geneticin) and checked by colony PCR (primers used are described in **Supplementary table I**).

6.2.3. Screening conditions

In order to take into account clonal variation of *P. pastoris* transformants, five colonies from each strain were selected and further tested in shake flask cultures in triplicate, as described by Cámara and coworkers [12]. Firstly, strains were grown overnight in 50 mL of Buffered Minimal Glycerol (BMG) medium. Afterwards, reinoculation from BMG to 25 mL of Buffered Minimal Methanol (BMM) medium was performed at an initial optical density of 1. One pulse of 125 µL of pure methanol was added (0.5% v/v) after 24 h and 48 h of cultivation. Samples for biomass and recombinant protein quantification were taken after 70 h of incubation.

6.2.4. Chemostat cultivation conditions

Selected strains were further characterized in duplicate chemostat cultures using three different mixed carbon sources: 60:40 glycerol:methanol, 60:40 glucose:methanol and 80:20 glucose:methanol (% w/w). Cultures were performed at a working volume of 1 L using a Biostat B Plus bioreactor (Sartorius AG, Göttingen, Germany). Cultivation conditions were set to 25°C, pH 5, 700 rpm, 1 vvm inlet gas flow, 0.2 bar overpressure and 0.1 h^{-1} dilution rate (D). A BlueinOne Gas Analyser (BlueSens, Herten, Germany) was used to measure the gas outlet O_2 and CO_2 concentrations. Samples were taken at the 4^{th} and 5^{th} residence times.

Batch medium contained: 19.95 g/L glycerol, 0.9 g/L citric acid, 6.3 g/L $(NH_4)_2HPO_4$, 0.01 g/L $CaCl_2$, 0.45 g/L KCl, 0.25 g/L $MgSO_4 \cdot 7H_2O$, 1 mL/L Biotin (0.2 g/L; Sigma, Munich, Germany), 0.3 mL antifoam (Glanapon, Bussetti & Co GmbH, Wien, Austria)

and 2.3 mL/L PTM1 trace salts stock solution. Chemostat medium contained: 25g/L total carbon source, 0.42 g/L citric acid, 2.18 g/L $(NH_4)_2HPO_4$, 5.5 mg/L $CaCl_2{\cdot}2H_2O$, 0.85 g/L KCl, 0.32 g/L $MgSO_4{\cdot}7H_2O$, 0.5 mL Biotin (0.2 g/L), 50 µL antifoam (Glanapon) and 0.8 mL PTM1 trace salts stock solution. Trace salt composition was: 6.0 g/L $CuSO_4.5H_2O$; 0.08 g/LNaI; 3.36 g/L $MnSO_4.H_2O$; 0.2 g/L$Na_2MoO_4.2H_2O$; 0.02 g/L H_3BO_3; 0.82 g/L $CoCl_2.6H_2O$; 20 g/L $ZnCl_2$; 65 g/L $FeSO_4.7H_2O$ and 5.0 mL H_2SO_4 (95-98% w/w). Media pH was set to 5 with 25% HCl.

6.2.5. Analytical methods

Biomass determination

Biomass optical density was measured at 600 nm using a DR3900 spectrophotometer (Hach, Bizkaia, Spain). Dry cell weight (DCW) was determined using pre-weighed dried glass fiber filters (Millipore, Massachusetts, USA) as reported by Jordà and coworkers [14]. Both measurements were performed in triplicate.

Lipolytic activity assay

To quantify the secretion of RoI in the supernatant, a lipase colorimetric assay was performed (Roche Diagnostics). The method used is described in Resina et al. [15].

Metabolite quantification

Residual glycerol, glucose and methanol concentrations in the cultures were measured using a HPLC. Specifically, an Ultimate 3000 Liquid Chromatography Systems (Dionex) with aICSep ICE-COREGEL 87H3 (Transgenomic) ion exchange column and a Water 2410 (Waters) refraction index detector. CROMELEON software (Dionex) was used for data analysis.

Droplet digital PCR (ddPCR) for recombinant gene copy number determination

To quantify *POS5* and *AtOX* gene dosage, ddPCR analysis was performed using the protocol described by Cámara and coworkers [12]. Primers used are described in **Supplementary table II**.

qPCR analysis of gene expression levels

The method used for RNA fixation and isolation from fermentation samples is described in the manufacturer guidelines for TRIzol™ Reagent usage (Invitrogen, California, USA). Yield of the RNA extractions was determined in a Nanodrop™1000 Spectrophotometer (Thermo Fischer Scientific, Waltham, Massachusetts, USA). 500 ng of RNA were used as template for cDNA conversion with iScript™ cDNA synthesis kit (Bio-Rad Laboratories, Hercules, California, USA). SYBR™ Select Master Mix (Thermo Fischer Scientific, Waltham, Massachusetts, USA) was the reagent used for the qPCR. qPCR components (master mix, cDNA, 10 µM primers and RNase-free water) were mixed in the plate automatically by the Epmotion 5075 (Eppendorf, Hamburg, Germany). The QuantStudio 12K Flex Real-Time PCR System from Thermo Fisher Scientific was used for qPCR reaction and amplification detection. Data was analyzed using the ThermoFisher Cloud web-based software. Actin-encoding gene *ACT1* was used as reference gene for relative quantification of gene expression. Primers used are described in **Supplementary table II**.

6.2.6. *In silico* metabolic flux analysis

The iMT1026 v.3.0 GEM was used in the COBRA v.2.0.6 [16] toolbox with Gurobi v.8.0.1 as solver. The simulations of the wild-type Rol-producer strain were done using FBA, maximizing the biomass reaction. Simulations of the redox strains were performed using Minimization of Metabolic Adjustment (MOMA), with the wild-type simulations as reference conditions. For all simulations performed, carbon source uptake rate was fixed to match the experimental data. In the simulations of *POS5* and *AtOX* strains the ectopic reactions were added at fixed flux values ranging from 0 to 2 mmol/(gDCW·h).

6.2.7. Statistical analyses

Unless otherwise stated, data are expressed as mean ± standard deviation of at least triplicate samples. To test for significance when comparing experimental data, obtained from continuous cultures and transcriptomic experiments, a Student's *t*-test was performed using the Microsoft Excel software. A p-value lower than 0.05 was considered statistically significant.

6.3. Results

6.3.1. Effect of overexpression of *POS5* and *AtOX* on *Rhizopus oryzae* lipase production

Three series of RoI-producing strains, co-overexpressing genes encoding for redox cofactor balance related enzymes (cPOS5, mPOS5 and AtOX), were tested in shake flask cultures together with a RoI-producing reference strain. Biomass and extracellular RoI activity were measured after 70 h of cultivation in methanol medium (BMM) and final specific production of recombinant lipase was calculated. As shown in **Figure 1A**, co-overexpression of these genes results in modest but consistent increase in specific extracellular lipase activity in some of the strains. Specifically, *mPOS5* showed 15% activity increase referred to the control strain. Conversely, four out of five of the *cPOS5* clones tested, did not show any significant difference respect to the control strain. However, one of the clones showed an important deviation from that behavior, with 50% higher RoI specific activity levels. Concerning AtOX strains, an average RoI specific activity increase of 30% was detected.

For a better characterization of the redox-engineered clones, gene copy number (GCN) determination of *POS5* and *AtOX* was carried out by means of ddPCR (**Figure 1B**). All clones tested for the mPOS5 and AtOX strains showed a single integration event of the corresponding heterologous gene, except for the cPOS5 clone with increased product yields, which had two copies of the *POS5* gene instead of one, referred hereafter as 2cPOS5 strain.

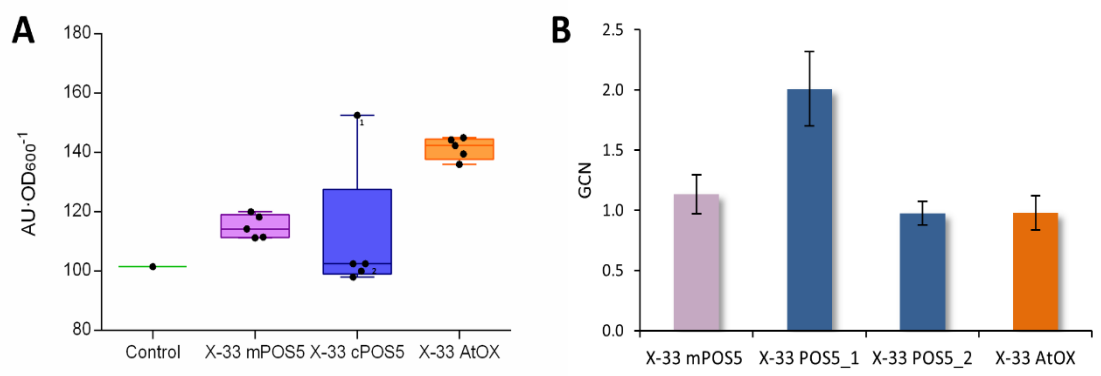


Figure 1. Screening of redox engineered strains. A) Specific RoI production in shake flask cultures. **B)** *POS5* and *AtOX* gene copy number determination. The cPOS5 clone that was an outlier was labeled as 1 and the other four clones labeled as 2 for differentiation in Figure 1B.

6.3.2. Performance of 2cPOS5 and AtOX strains in chemostat cultures

For a phenotypic characterization at bioreactor scale of the redox-engineered strains showing a significant improvement in RoI production, the 2cPOS5 clone and the AtOX clone that showed the best performance in the previous screening step were selected. These two clones and the reference strain were cultivated in multicarbon source-limited chemostats at a D of 0.1 h^{-1} using different mixtures of glycerol or glucose with methanol. Operational conditions ensured normoxic conditions, i.e. pO_2 levels remained higher than 40% during the experiments, as well as residual glucose and glycerol concentrations below detection limit, ensuring (partial) derepression of methanol metabolism [12], [14], [17]. Three carbon source compositions (60:40 glycerol:methanol, 60:40 glucose:methanol and 20:80 glucose:methanol (w/w, %)) were tested for each strain to investigate possible carbon source-related effects of these genetic modifications. Notably, methanol-only chemostat cultures of the RoI-producing strain resulted in culture wash-out even at dilution rates as low as 0.05 h^{-1} due to the burden/stress caused by RoI production. In contrast, the X-33 mock strain can be grown in methanol-fed chemostat cultures at a D up to 0.12 h^{-1} [18].

The redox balance alterations due to *cPOS5* and *AtOX* overexpression resulted in a significant physiological impact in the cells (**Figure 2**). Specifically, the 2cPOS5 strain had a modest but significant increase in terms of lipase specific productivity (q_p) in all tested conditions. The improvement showed was a 21% increase in the 60%glyc:40%meth condition. This positive effect was further enhanced when glucose was used as co-substrate, showing a 34% increase with respect to the reference strain in the 60%gluc:40%meth condition). The condition with lower amount of glucose (20%gluc:80%meth) showed the highest improvement (41%).

The AtOX strain showed a significant improvement in q_p (23% respect to the reference strain) only at the highest methanol ratio (20%gluc:80%meth), whereas in the other growth conditions the slight improvements were not statistically significant (**Figure 2A**). Despite the higher productivity improvement observed when using glucose as co-substrate, we still could observe partial repression of *ROL* expression due to residual glucose (below detection limit of the HPLC analysis) in the bioreactor medium,

resulting in product yields that are half of those observed in the chemostats that use glycerol as co-substrate (**Figure 2A**).

Although it is not statistically significant, a slight effect of *POS5* and *AtOX* overexpression on biomass yield (Y_{xs}) can be appreciated (**Figure 2B**). These strains have a reduced Y_{xs} tendency in all the conditions tested. Only in the case of AtOX strain grown in the 20%gluc:80%meth condition, the biomass yield was significantly reduced by 12.5%. The lower Y_{xs} correlates with a diminished biomass concentration in the chemostat (**Figure 2C**). Importantly, the 60%gluc:40%meth and the 20%gluc:80%meth conditions showed residual methanol accumulation in the reactor, which directly affected cell growth parameters. In particular, about a 25% and 75% of the methanol remained in the medium under these growth conditions, respectively (data shown in **Supplementary table III**).

Concerning the specific oxygen uptake rate (qO_2) values, these were not significantly affected in any condition (**Figure 2D**). Conversely, ectopic NADH kinase and NADH oxidase activity seems to generally produce a slight but not significant increase in specific carbon dioxide uptake rate (qCO_2), as shown in **Figure 2E.** This increment was only statistically significant (34%) for the AtOX strain in the 20%gluc:80%meth condition

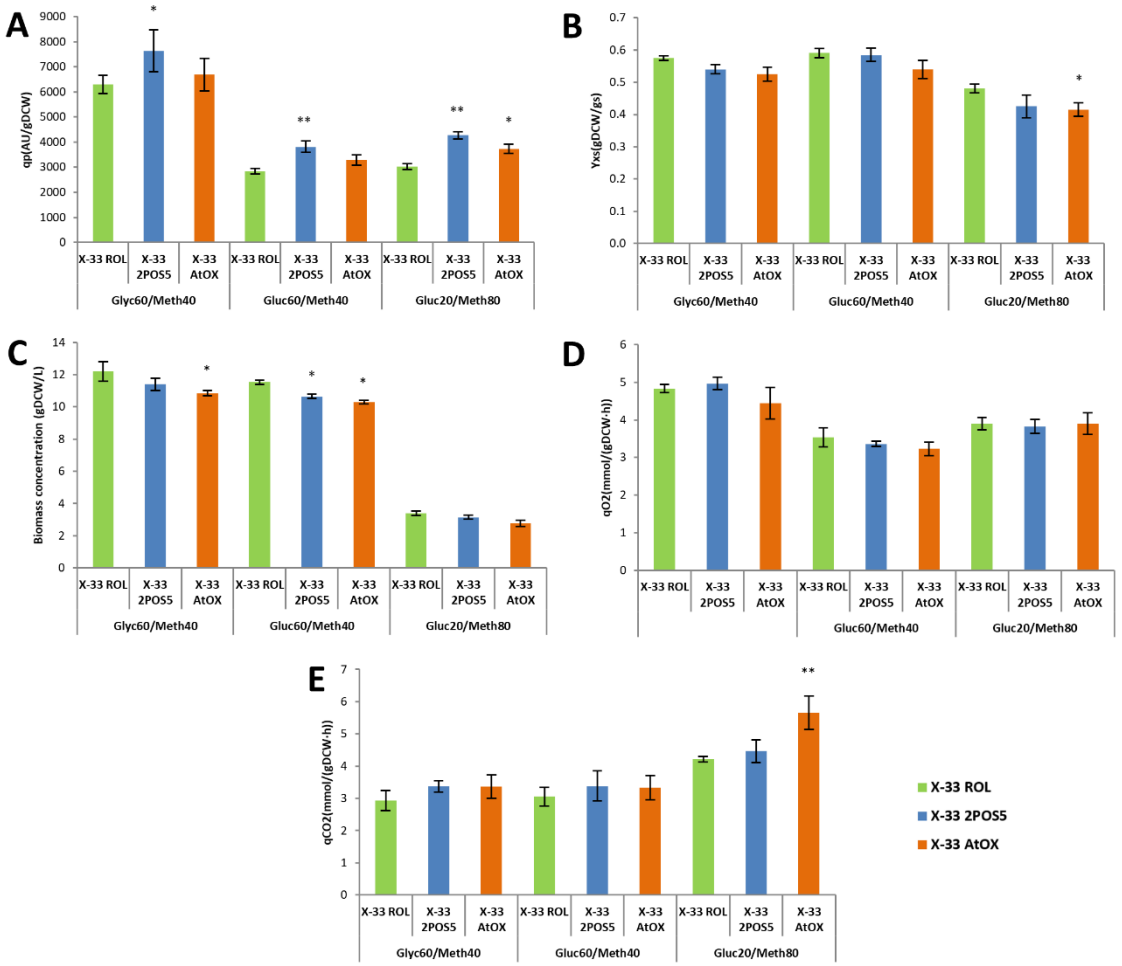


Figure 2. Macroscopic growth steady state parameters of *P. pastoris* strains under different chemostat cultivation conditions. A) Specific RoI production. **B)** Biomass yield ($Y_{x/s}$). **C)** Biomass concentration. **D)** Specific oxygen consumption rate (qO_2). **E)** Specific carbon dioxide production rate (CO_2). *p-val < 0.05, **p-val < 0.01, compared to the reference strain. C-balance error was lower than 15% in all the chemostat cultures.

6.3.3. Transcriptional effect of *POS5* and *AtOX* overexpression

To provide further insight into the mechanism(s) underlying the physiological alterations caused by *POS5* and *AtOX* overexpression, transcriptional levels of marker genes of different pathways of the central carbon metabolism (glycolysis, methanol assimilation route, Pentose Phosphate Pathway (PPP) and TCA cycle), as well as the recombinant genes (*ROL*, *POS5* and *AtOX*) in chemostat-grown cells were analyzed by qPCR.

Gene expression results are depicted in **Figure 3** (relative RNA levels normalized against the control strain for each condition are shown in **Supplementary figure IIA**).

As expected, expression levels of the genes coding for our model recombinant protein (*ROL*) and the alcohol oxidase 1 (*AOX1*), both under control of the same promoter, did not show any significant differences in any of the redox-engineered strains compared to reference strain in any condition. Coherently, the genes encoding for the other two enzymes of the methanol metabolism tested, fructose 1,6-bisphosphate aldolase (*FBA1-2*) and formate dehydrogenase 1 (*FDH1*), also showed no major changes in expression levels. In the case of glycolytic (*TDH3*) and TCA cycle (*IDH2*) marker genes, the transcriptional levels were again very similar between strains, suggesting that the redox alterations had no major impact on their transcriptional regulation and, therefore, on the transcriptional regulation of the glycolysis and TCA cycle.

Conversely, transcriptional levels of *ZWF1*, the PPP marker gene encoding for glucose-6-phosphate dehydrogenase, appeared to be highly sensitive to the redox perturbations introduced to the cells. Specifically, *POS5* overexpression had a down-regulatory effect on *ZWF1* transcriptional levels, whereas *AtOX* exerted the opposite positive effect. In particular, the POS5 strain presented a 35% reduced *ZWF1* expression levels compared to the reference strain when cells where grown on glycerol, whereas the decline in the transcription levels was only 27% and 21% when 60% and 20% of glucose was used as co-substrate, respectively. Regarding the AtOX strain, the boosting effect of the alternative oxidase reaction in *ZWF1* expression was higher when glucose was used as co-substrate than with glycerol. The fold-changes with glycerol and glucose (60:40 co-substrate:methanol ratio) were 1.41 and 2.51, respectively. Regarding the 20%gluc:80%meth condition, there was also an increase in *ZWF1* expression levels (1.46-fold higher).

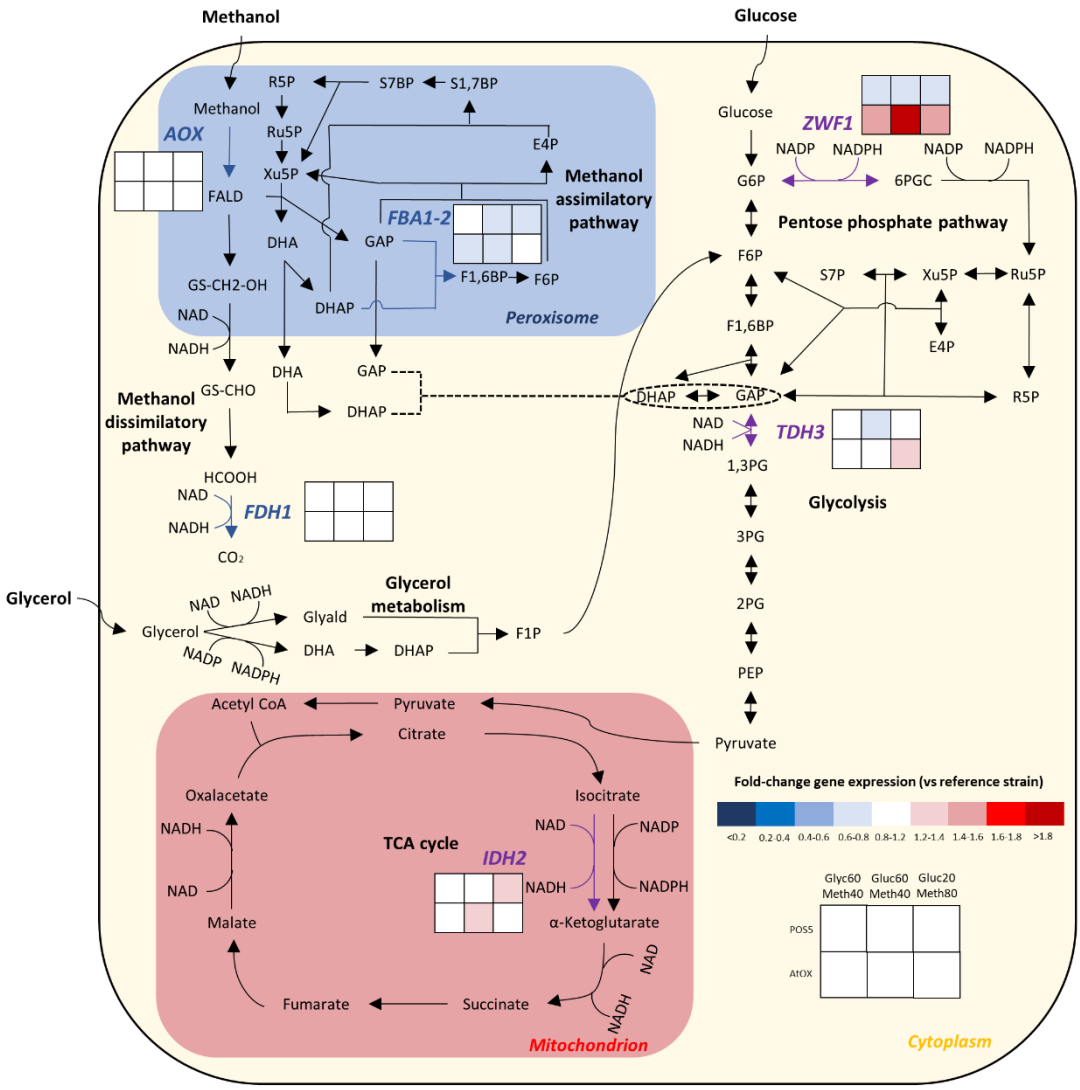


Figure 3. Metabolic pathways implied in carbon source assimilation under the different carbon source conditions tested in chemostat. Methanol and central carbon metabolism genes analyzed by qPCR are marked in blue and purple, respectively. FALD: formaldehyde; GS-CH_2-OH: hydroxymethylglutathione; GS-CHO: S-formylglutathione; HCOOH: formate; CO_2: carbon dioxide; Xu5P: xylulose 5-phosphate; DHA(P): dihydroxyacetone (phosphate); GAP: glyceraldehyde 3-phosphate; F1,6BP: fructose 1,6-bisphosphate; E4P: erythrose 4-phosphate; S1,7BP: sedoheptulose 1,7-bisphosphate; S7P: sedoheptulose 7-phosphate; R5P: ribose 5-phosphate; Ru5P: ribulose 5-phosphate; G6P: glucose 6-phosphate; 6PGC: 6-phosphogluconate; F6P: fructose 6-phosphate; 1,3PG: 1,3-bisphosphoglycerate; 3PG: 3-phosphoglycerate; 2PG: 2-phosphoglycerate; PEP: phosphoenolpyruvate; *AOX1,* alcohol oxidase 1; *FDH1,* formate dehydrogenase 1; *FBA1-2*, fructose 1,6–bisphosphate aldolase; *TDH3*, glyceraldehyde-3-phosphate dehydrogenase; *ZWF1*, glucose-6-phosphate dehydrogenase; *IDH2*, isocitrate dehydrogenase 2.

Expression levels of ectopic *POS5* and *AtOX*, under the control of the endogenous *TDH3* promoter, were normalized to those of this endogenous gene (**Figure 4**). Since these genes are expressed under control of the same promoter, one should expect similar expression levels (taking into account the GCN). However, *TDH3* transcriptional

levels appeared to be much higher than those of *POS5* and *AtOX* in all the conditions tested. For *POS5* strain, since it has twice the gene dosage of *TDH3*, it should be expected a 2-fold higher expression level. Strikingly, we observed a 50% reduction with respect to the reference *TDH3* expression levels (between 0.53-0.60 fold-change). Expected expression levels for *AtOX* (with the same gene dosage as *TDH3*) would be similar to those of *TDH3.* However, the ratio of *AtOX* against *TDH3* was even lower than that of *POS5* (0.25-0.40 fold change). These results may suggest a lower stability of the heterologous mRNAs compared to the endogenous gene. Despite these relatively low transcriptional levels, we were able to confirm that both genes are being successfully expressed.

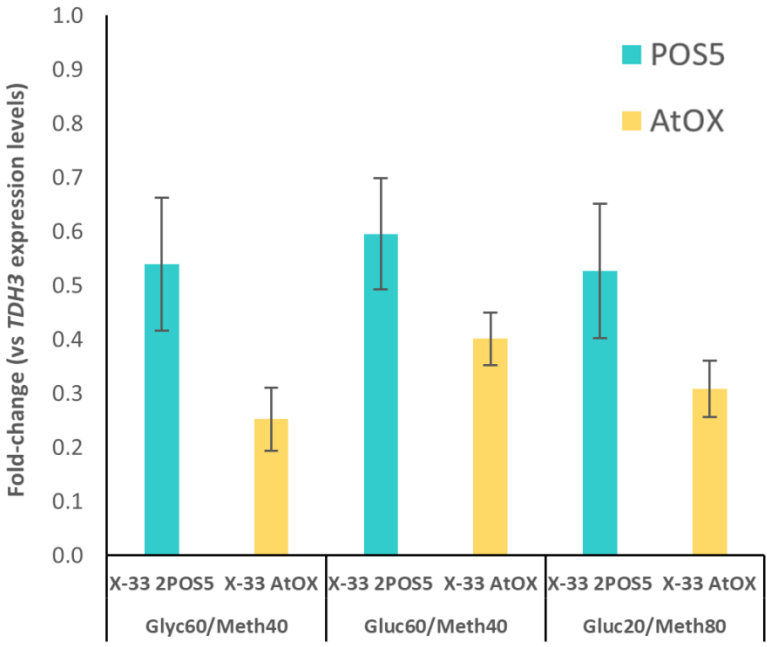


Figure 4. Transcriptional analysis of the ectopic POS5 and AtOX genes expressed under control of *GAP* promoter. Results are shown as a ratio of the heterologous genes (*POS5* and *AtOX*) against endogenous *TDH3*.

6.3.4. *In silico* analysis of the effect of POS5 and AtOX overexpression

The iMT1026 v3.0 genome-scale metabolic model of *P. pastoris* was used for interpretation of the biological data obtained in the chemostat cultures. First, simulations of the wild-type RoI producing strain were done using FBA under two different carbon source conditions (60/40% glycerol/methanol and glucose/methanol). The effect of *POS5* and *AtOX* overexpression in a RoI-producing strain was modelled by

gradually increasing the fluxes of their corresponding reactions from 0 to 2 mmol/(gDCW·h) and using MOMA with the previous wild-type simulations as reference conditions. The addition of a significant flux of a NADH kinase or alternative oxidase reaction in the model has a great impact in the cell physiological parameters (**Supplementary figure II**).

In particular, the simulations performed show important changes in macroscopic parameters such as μ, q_{CO2} and q_{O2}. Both POS5 and AtOX strains have a reduced flux in the biomass reaction of the model (**Supplementary figure II A-C**). This matches with the reduced biomass yield observed in the bioreactor cultivations. Carbon dioxide production also showed good agreement between chemostat and simulation values (**Supplementary figure II C-E**). As previously mentioned in this chapter, chemostat's off-gas analyses show that CO_2 production has a tendency to increase in redox-engineered strains compared to the reference strain. However, such variations were only statistically significant when the AtOX strain was grown under the 20%gluc:80%meth condition. Simulation results reflect a similar outcome for the AtOX strain, since increasing fluxes of alternative oxidase have a remarkable effect in the specific CO_2 production. On the contrary, the impact of the NADH kinase reaction on the CO_2 production rates of *POS5* strain is almost negligible. The *in silico* results for the AtOX strain also present an increase in the q_{O2} coupled with the previously mentioned increment in q_{CO2} (**Supplementary figure II G-I**). However, this change in q_{O2} was not reflected in the corresponding chemostat physiological parameters.

The performed simulations also reveal that both introduction of ectopic redox reactions and their subcellular location (cytosol vs mitochondria) have an important and distinctive impact on the flux distributions through the central carbon metabolism, as shown in **Figure 5**.

Simulations results indicate that methanol flux distribution through the assimilatory and dissimilatory pathways, is affected in a similar way by NADH kinase and alternative oxidase ectopic reactions, independently of the co-substrate used. All the simulations performed show that the dissimilatory pathway is favored over assimilation for biomass while the flux of these redox reactions increases.

Another important effect predicted by the simulations is a shift in the PPP when glucose is used as co-substrate. Particularly, the NADH kinase reaction reduces the flux through the oxidative branch of this route whereas the alternative oxidase has a positive effect. On the other side, when glycerol is used as co-substrate, there is no effect of NADH kinase flux in PPP (in both cases flux through this route is null) whereas alternative oxidase flux exerts a positive effect. The excess of NADPH generated by the NADH kinase tends to be compensated by reducing flux through PPP. The alternative oxidase reaction generates an extra amount of NAD^+ and PPP acts as one of the electron donors for this pool (once NAD^+ has been previously phosphorylated). Again, since the simulations do not predict any flux through this pathway when glycerol is used as second carbon source, no differences were observed for this condition. Interestingly, in this situation the cell tries to reestablish redox balance in the *POS5* strain by balancing the flux between the two possible first reactions of glycerol metabolism (which are NADH and NADPH dependent each) instead of using PPP.

Regarding the glycolytic pathway, glucose/methanol simulations indicate that the upper part of this route increases in both POS5 and AtOX strains, whereas the lower part (after glyceraldehyde 3-phosphate) is reduced in POS5 strain and increased in AtOX strains. For the glycerol/methanol simulations, increasing NADH kinase flux have a negative effect in the fluxes of the lower part of glycolysis and a positive one when alternative oxidase flux is increased.

Another important change predicted consists in a total or partial increase of the fluxes through TCA cycle in all simulations performed. Glucose/methanol results show an overflow all over the TCA cycle in AtOX strain and an almost complete increase of all the fluxes in POS5 strain. For the glycerol/methanol simulations, these showed the same tendency in the case of the AtOX strain, with a total increase of TCA cycle, whereas only the first steps of this cycle showed an increment in the POS5 strain.

Finally, in the simulations where glycerol metabolism is active, we observe that the first step in the glycerol assimilation pathway involves a competition between two NADH dehydrogenases with different cofactor affinity (NAD and NADP, dependent). Although the existence of this competition is not very clear, simulation results indicate

that the NAD-dependent enzyme flux is favored in the overexpression of both redox-related genes.

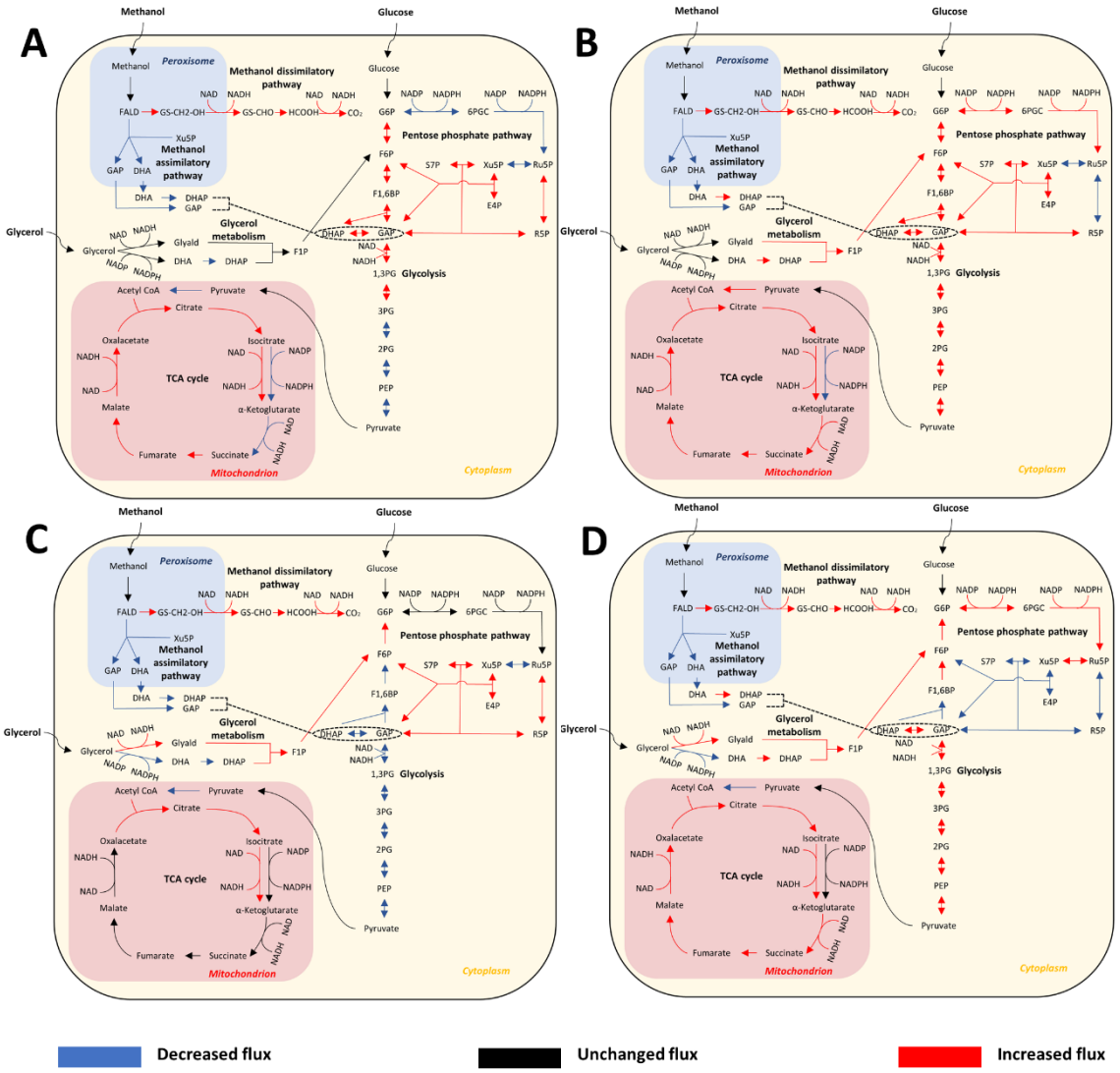


Figure 5. Graphical representation of the flux redistributions obtained *in silico*. The simulation conditions include glucose **(A, B)** and glycerol **(C, D)** as cosubstrates when *POS5* **(A, C)** and *AtOX* **(B, D)** are overexpressed. Red, blue and black arrows represent positive, negative and null changes in the fluxes compared to reference strain simulation, respectively. FALD: formaldehyde; GS-CH_2-OH: hydroxymethylglutathione; GS-CHO: S-formylglutathione; HCOOH: formate; CO_2: carbon dioxide; Xu5P: xylulose 5-phosphate; DHA(P): dihydroxyacetone (phosphate); GAP: glyceraldehyde 3-phosphate; F1,6BP: fructose 1,6-bisphosphate; E4P: erythrose 4-phosphate; S1,7BP: sedoheptulose 1,7-bisphosphate; S7P: sedoheptulose 7-phosphate; R5P: ribose 5-phosphate; Ru5P: ribulose 5-phosphate; G6P: glucose 6-phosphate; 6PGC: 6-phosphogluconate; F6P: fructose 6-phosphate; 1,3PG: 1,3-bisphosphoglycerate; 3PG: 3-phosphoglycerate; 2PG: 2-phosphoglycerate; PEP: phosphoenolpyruvate.

6.4. Discussion

Yeast systems such as *P. pastoris* are widely used platforms for recombinant protein production due to its capability to achieve high biomass and product yields. However, the production capacity of these systems is not lacking limitations. The bottlenecks are usually associated to protein folding and secretion [19], [20] and transcriptional constraints [12], [21], [22]. In this work, we focus on increasing intracellular NADPH availability, since NADPH is required for redox balancing of protein folding and amino acid biosynthesis.

Through time, various strategies have been proposed to improve cofactor regeneration [23]–[25]. In our case, we used gene knock-in of directly-related redox genes (*POS5* and *AtOX*) in order to attack the issue from a metabolic perspective. Our results with strains overexpressing *POS5* and *AtOX* point to a strong effect of these enzymes in cofactor balance and cell physiology. Previous studies have successfully achieved the overexpression of *POS5* in *P. pastoris* using P_{GAP} expression systems and an antibody fragment as a product [9]. Our work is the first attempt to use this strategy in a methanol-based system and using a secreted protein (Rol) as model of study. To our knowledge, there is no information regarding overexpression of an alternative oxidase with the aim to improve recombinant protein production neither.

From the experimental results, we can conclude that Rol production is increased in both redox modified strains. Particularly, the strain carrying 2 copies of the *POS5* gene presents the highest values in all conditions tested. As shown also by Tomàs-Gamisans and coworkers [9], our results support that glucose is a better substrate than glycerol to take advantage of these redox modifications. However, glucose repression of P_{AOX1} reduces by half the recombinant protein production. This situation makes glycerol a more suitable co-substrate for the modified strains when P_{AOX1}-based expression system is employed.

The increase in specific recombinant protein production could be directly related with the lower biomass yield observed in POS5 and AtOX strains. The altered redox balance would redirect metabolic fluxes to benefit biosynthesis of macromolecules such as amino acids. It was already postulated that synthesis of energetically costly amino

acids can be a limiting metabolic factor in recombinant protein production [26]. The increase in the intracellular amino acid pools would enhance this way protein synthesis. This redistribution of fluxes seems to also lead to an increased specific CO_2 production rate compared to control strain in all conditions tested.

Model-driven approaches have been employed extensively over the last years to guide metabolic engineering [27]–[29] and they have been applied in this work to provide insight into chemostat results *in silico* using the iMT1026 v.3.0. For instance, the reduction of the biomass yield observed experimentally is successfully predicted. Concerning CO_2 production rates, experimental results do not provide enough evidence to stablish the same pattern observed in the *in silico* results.

We also used modeling to predict how metabolic fluxes would redistribute as a result of *POS5* and *AtOX* overexpression. Since both ectopic enzymes consume NADH as substrate, most of the changes observed are directed to regenerate this cofactor. Methanol metabolism is redirected to fast NADH generation through the dissimilatory pathway and TCA cycle is also overexpressed, particularly in AtOX strain, where NADH consumption takes place in the mitochondrion. When glycerol is used as carbon source, the utilization of the NAD-dependent alcohol dehydrogenase instead of the NADP-dependent glycerol dehydrogenase is an important input to restore NADH balance. However, experimental evidence is still needed in order to assure this regulatory point in glycerol metabolism. Moreover, the results obtained indicate that PPP is an important regulatory point of redox balance, as well. This is one of the first reports about how this pathway is affected by cell's redox state [30].

A transcriptional analysis of key genes of central carbon metabolism was done with the objective to elucidate if the gene regulation could be related to the redox state inside the cell. Overall, no significant impact was observed in methanol metabolism, glycolysis or Krebs cycle. The expression levels of these genes are probably tightly connected to the carbon source presence in the medium and any possible effect related with changes in the redox balance is too low to be detected by the qPCR accuracy levels. This situation does not mean that flux changes do not exist due to overexpression of *POS5* and *AtOX*, but this flux redistribution is not coupled to the

gene regulation. Further studies using ^{13}-C flux analysis could provide more information in this regard.

In agreement with our model-driven interpretation, the PPP exhibited remarkably different expression levels in the modified strains. Overexpression of *POS5* seems to downregulate *ZWF1* whereas in the case of *AtOX* enhances its expression. Since PPP is one of the main NADPH sources in *P. pastoris*, the extra amount of this cofactor provided by the NADH kinase reaction tends to be compensated by reducing the flux through the PPP pathway. On the other side, *ZWF1* overexpression by the presence of an alternative oxidase is explained by the *P. pastoris* need to maximize the usage of the pool of NAD^+ accumulated inside cell. Since the cell is not able to restore the redox balance by using only NADH-generating pathways, such as methanol dissimilatory pathway, glycolysis or TCA cycle, it takes profit of the PPP by first converting NAD^+ to $NADP^+$. An interesting observation about these results is that the effect of *AtOX* overexpression in the gene expression levels of *ZWF1* is higher when glucose is used as co-substrate. The reason for that is the important contribution of PPP in supplying NADPH when glucose is used as carbon source compared to glycerol [25]. Glycerol metabolism includes in its metabolic route a NADP-dependent glycerol dehydrogenase (GLYCD) in the initial steps. The high flux through this reaction makes PPP a far less critical route when glycerol is used as carbon source.

6.5. Conclusions

Although *P. pastoris* has been one of the most preferred yeasts systems for recombinant protein production, some bottlenecks still constrain the attainment of higher product titers, limiting the application of this host for this purpose. Producing strains have difficulties in coping with the additional demand of NADPH required for the synthesis of amino acids, disulfide bonds formation, alleviation of endoplasmic reticulum from oxidative stress and protein secretion. Thus, overproduction of recombinant proteins can result in a redox cofactor imbalance resulting in a low NADPH availability.

We have demonstrated that recombinant protein production in *P. pastoris* can be improved by knocking-in heterologous genes coding for redox cofactor-modifying

enzymes. Bioreactor cultivations showed improvements ranging among 20-40% in Rol production when *POS5* and *AtOX* were overexpressed. In future work, it would be highly desirable to try these modifications with high-copy number strains of *ROL,* where the metabolic stress would be higher. We also showed that the extent of the improvement in recombinant protein production is tightly related to the carbon source used. Glucose has been proven to give better results in terms of fold-change increase in Rol yield, but it is not able to compensate for the repression concomitantly exerted in the P_{AOX1}. This still makes glycerol the co-substrate of choice for this expression system. The modified strains also show other altered physiological traits such as biomass yield and CO_2 generation. These results were further investigated by an *in-silico* approach using a genome-scale metabolic model, demonstrating its usefulness in understanding catabolic and anabolic processes.

Finally, transcriptomic results indicate that the gene regulation of central carbon metabolism (besides PPP) is not affected by the redox alterations achieved. This means that any flux changes in these pathways are not tied to a transcriptomic shift. To further demonstrate that our genetic modifications have a real effect on these routes, ^{13}-C flux analysis would be an interesting approach.

6.6. References

[1] B. R. Glick, "Metabolic load and heterologous gene expression.," *Biotechnol. Adv.*, vol. 13, no. 2, pp. 247–61, 1995.

[2] P. Ferrer and J. Albiol, "^{13}C-based metabolic flux analysis of recombinant Pichia pastoris.," *Methods Mol. Biol.*, vol. 1191, pp. 291–313, 2014.

[3] R. J. Zahrl, B. Gasser, D. Mattanovich, and P. Ferrer, "Detection and elimination of cellular bottlenecks in protein-producing yeasts," in *Methods in Molecular Biology*, vol. 1923, pp. 75–95, 2019.

[4] F. W. Krainer, C. Dietzsch, T. Hajek, C. Herwig, O. Spadiut, and A. Glieder, "Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway," *Microb. Cell Fact.*, vol. 11, pp. 22, 2012.

[5] J. Nocon *et al.*, "Increasing pentose phosphate pathway flux enhances recombinant protein production in Pichia pastoris," *Appl. Microbiol. Biotechnol.*, vol. 100, no. 13, pp. 5955–5963, 2016.

[6] C. Jayachandran, B. P. Athiyaman, and M. Sankaranarayanan, "Cofactor engineering improved CALB production in Pichia pastoris through heterologous expression of NADH oxidase and adenylate kinase," *PLoS One*, vol. 12, no. 7, 2017.

[7] J. Heyland, J. Fu, L. M. Blank, and A. Schmid, "Quantitative physiology of Pichia pastoris during glucose-limited high-cell density fed-batch cultivation for recombinant protein production," *Biotechnol. Bioeng.*, vol. 107, no. 2, pp. 357–368, 2010.

[8] Y. Iwahashi, A. Hitoshio, N. Tajima, and T. Nakamura, "Characterization of NADH kinase from Saccharomyces cerevisiae," *J. Biochem.*, vol. 105, no. 4, pp. 588–593, 1989.

[9] J. A. M. Tomàs-Gamisans, C. Andrade, F. Maresca, S. Monforte, P. Ferrer, "Redox engineering by ectopic overexpression of NADH kinase in recombinant Pichia pastoris (Komagataella spp): Impact on cell physiology and recombinant

production of secreted proteins.," Submitted in *Front. Bioeng. Biotechnol*.

[10] C. H. Johnson, J. T. Prigge, A. D. Warren, and J. E. McEwen, "Characterization of an alternative oxidase activity of Histoplasma capsulatum," *Yeast*, vol. 20, no. 5. pp. 381–388, 2003.

[11] A. Kern *et al.*, "Pichia pastoris 'just in time' alternative respiration," *Microbiology*, vol. 153, no. 4, pp. 1250–1260, 2007.

[12] E. Cámara, J. Albiol, and P. Ferrer, "Droplet digital PCR-aided screening and characterization of Pichia pastoris multiple gene copy strains," *Biotechnol. Bioeng.*, vol. 113, no. 7, pp. 1542–1551, 2016.

[13] G. Stadlmayr *et al.*, "Identification and characterisation of novel Pichia pastoris promoters for heterologous protein production," *J. Biotechnol.*, vol. 150, no. 4, pp. 519–529, 2010.

[14] J. Jordà, P. Jouhten, E. Cámara, H. Maaheimo, J. Albiol, and P. Ferrer, "Metabolic flux profiling of recombinant protein secreting Pichia pastoris growing on glucose:methanol mixtures.," *Microb. Cell Fact.*, vol. 11, pp. 57, 2012.

[15] D. Resina *et al.*, "Engineering of bottlenecks in Rhizopus oryzae lipase production in Pichia pastoris using the nitrogen source-regulated FLD1 promoter," *N. Biotechnol.*, vol. 25, no. 6, pp. 396–403, 2009.

[16] J. Schellenberger *et al.*, "Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0," *Nat. Protoc.*, vol. 6, no. 9, pp. 1290–1307, 2011.

[17] J. Jordà, S. S. de Jesus, S. Peltier, P. Ferrer, and J. Albiol, "Metabolic flux analysis of recombinant Pichia pastoris growing on different glycerol/methanol mixtures by iterative fitting of NMR-derived (13)C-labelling data from proteinogenic amino acids.," *N. Biotechnol.*, vol. 31, no. 1, pp. 120–32, 2014.

[18] M. Tomàs-Gamisans, P. Ferrer, and J. Albiol, "Fine-tuning the P. pastoris iMT1026 genome-scale metabolic model for improved prediction of growth on methanol or glycerol as sole carbon sources," *Microb. Biotechnol.*, vol. 11, no. 1,

pp. 224–237, 2018.

[19] A. Idiris, H. Tohda, H. Kumagai, and K. Takegawa, "Engineering of protein secretion in yeast: strategies and impact on protein production.," *Appl. Microbiol. Biotechnol.*, vol. 86, no. 2, pp. 403–17, 2010.

[20] L. M. Damasceno, C.-J. Huang, and C. A. Batt, "Protein secretion in Pichia pastoris and advances in protein production.," *Appl. Microbiol. Biotechnol.*, vol. 93, no. 1, pp. 31–9, 2012.

[21] E. Cámara, N. Landes, J. Albiol, B. Gasser, D. Mattanovich, and P. Ferrer, "Increased dosage of AOX1 promoter-regulated expression cassettes leads to transcription attenuation of the methanol metabolism in Pichia pastoris," *Sci. Rep.*, vol. 7, 2017.

[22] E. Cámara, S. Monforte, J. Albiol, and P. Ferrer, "Deregulation of methanol metabolism reverts transcriptional limitations of recombinant Pichia pastoris (Komagataella spp) with multiple expression cassettes under control of the AOX1 promoter," *Biotechnol. Bioeng.*, vol. 116, no. 7, pp. 1710–1720, 2019.

[23] W.-H. Lee, M.-D. Kim, Y.-S. Jin, and J.-H. Seo, "Engineering of NADPH regenerators in Escherichia coli for enhanced biotransformation.," *Appl. Microbiol. Biotechnol.*, vol. 97, no. 7, pp. 2761–72, 2013.

[24] J. Nocon *et al.*, "Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production," *Metab. Eng.*, vol. 24, pp. 129–138, 2014.

[25] M. Lakshmanan, K. Yu, L. Koduru, and · Dong-Yup Lee, "In silico model-driven cofactor engineering strategies for improving the overall NADP(H) turnover in microbial cell factories," *J Ind Microbiol Biotechnol*, vol. 42, pp. 1401–1414, 2015.

[26] J. Heyland, J. Fu, L. M. Blank, and A. Schmid, "Carbon metabolism limits recombinant protein production in Pichia pastoris," *Biotechnol. Bioeng.*, vol. 108, no. 8, pp. 1942–1953, 2011.

[27] S. Y. Lee, D.-Y. Lee, and T. Y. Kim, "Systems biotechnology for strain improvement.," *Trends Biotechnol.*, vol. 23, no. 7, pp. 349–58, 2005.

[28] J. W. Lee, T. Y. Kim, Y. S. Jang, S. Choi, and S. Y. Lee, "Systems metabolic engineering for chemicals and materials," *Trends in Biotechnology*, vol. 29, no. 8. pp. 370–378, 2011.

[29] M. H. Medema, R. van Raaphorst, E. Takano, and R. Breitling, "Computational tools for the synthetic design of biochemical pathways.," *Nat. Rev. Microbiol.*, vol. 10, no. 3, pp. 191–202, 2012.

[30] P. Liu *et al.*, "Combined 13C-assisted metabolomics and metabolic flux analysis reveals the impacts of glutamate on the central metabolism of high β-galactosidase-producing Pichia pastoris.," *Bioresour. Bioprocess.*, vol. 3, no. 1, p. 47, 2016.

Appendix

Supplementary table I. Primers used for colony PCR of *POS5* and *AtOX* clones

PRIMER	SEQUENCE (5'-3')	Comments
POS5_Fw	ATCCGCGCCTGCAGGAATGTTTGTTAGAGTTAAGTTGAACAAGCC AGTTA	Check *POS5* integration
POS5_Rv	ATGACTAGGCCGAGGCGGCCTTAGTCGTTGTCAGTCTGTCTC	
AtOX_Fw	CTTGAGAAGAATGAAGAGAGACAACGG	Check *AtOX* integration
AtOX_Rv	CTGAGCCAACTTCAAGAAAGACAACAAG	

Supplementary table II. List of primers used for ddPCR and qPCR.

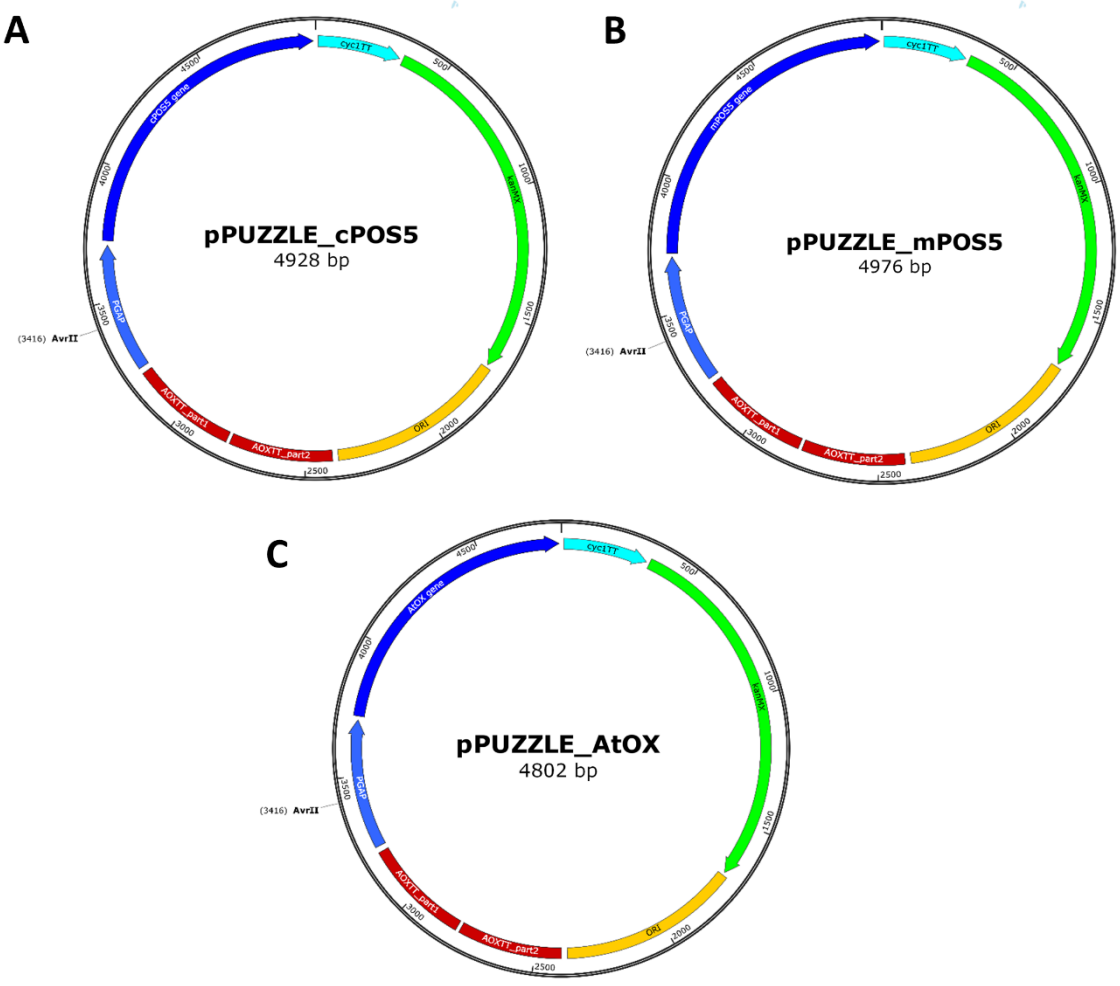
PRIMER	SEQUENCE (5'-3')	Tm	%GC	AMPLICON LENGHT (bp)
ACT1 F	TGTCCGGTGGTACTACTATGTTCC	65	50	199
ACT1 R	GATTCGTCGTACTCTTGCTTTGA	62	43	
AOX1 F	GACATTCACGGTTTCGAAGG	61	40	78
AOX1 R	CCTCAAGAAGTCCTGGCAAAC	63	42	
FBA1-2 F	CCCTTGGTTTGACGGAATG	60	42	135
FBA1-2 R	TTCCTCCGACAGGTCTAAC	60	42	
FDH1 F	GGTGCTGGAAGAATTGG	57	52	97
FDH1 R	GACAGTGTCGACTCTTC	55	52	
ICDH2 F	GCAGAGTGGTCTTTCCGTTGATG	60	52	141
ICDH2 R	CCGTGTCTTTCATAGAGGGTGAC	58	52	
ZWF1 F	CTGAGGATGGAAAGAAGCCAGC	60	55	204
ZWF1 R	CATTGAGAAATCCAGTAGACTGTTTGTATTG	58	35	
ROL F	CCTGTCGTCCAAGAACAAC	62	52	164
ROL R	GAGGACCACCAACAGTGAAG	62	53	
TDH3 F	GGTGAGGTTTCTGCCAGC	57	61	125
TDH3 R	GTGGACTCAATGACGTAGTC	56	50	
AtOX F	GGTTTTGGGTGCTCAGGGTG	61	60	88
AtOX R	GTAACCAACGAATCTGTGACAAGTTC	58	42	
POS5 F	GGAGTGTCATTGAAGAAGAA	57	40	148
POS5 R	CGTCAGCAGTAGTTCTAC	57	46	

Tm, melting temperature; %GC, percentage of guanosine+cytosine.

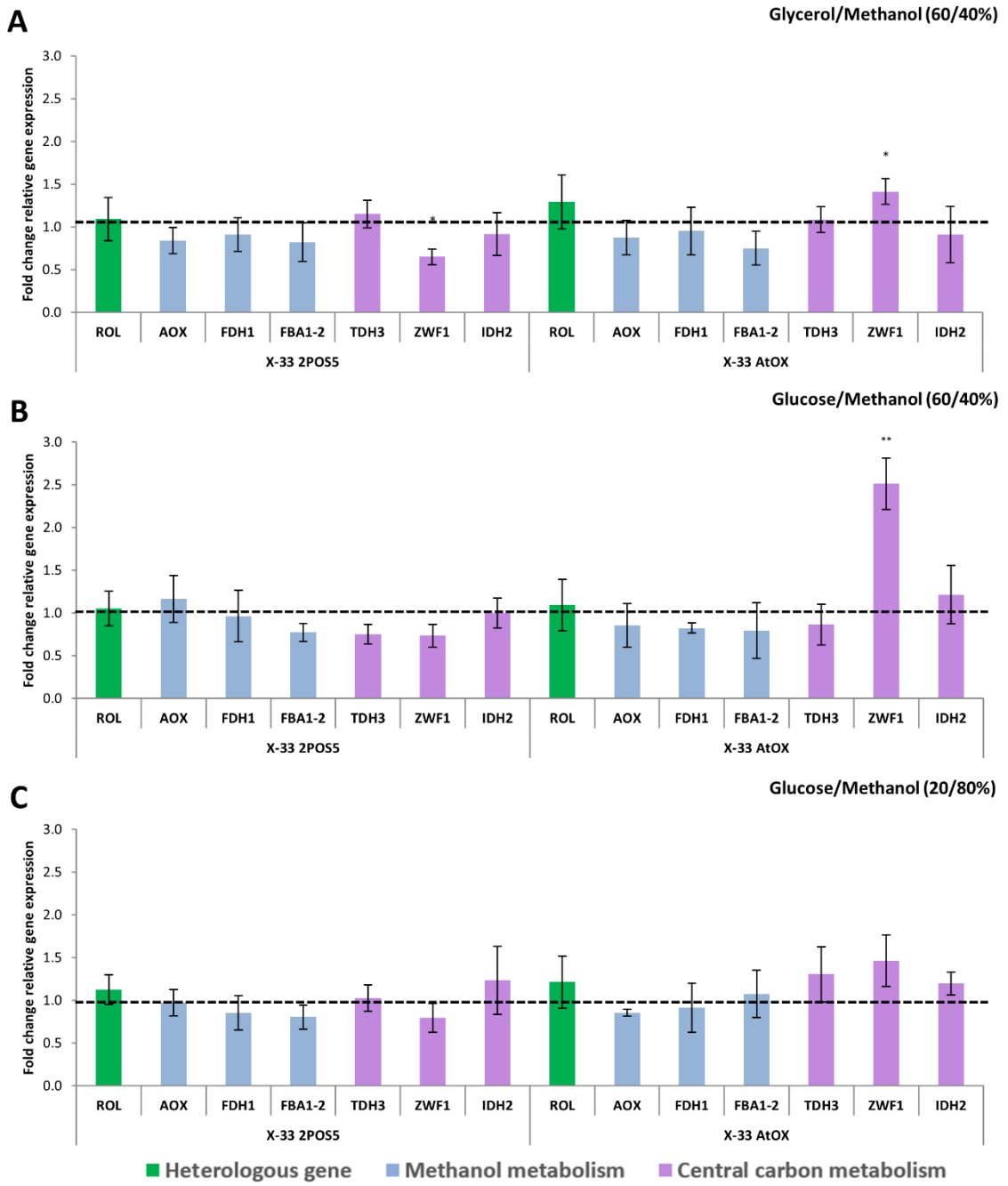
Supplementary table III. Macroscopic parameters of X-33 ROL, X-33 2POS5 and X-33 AtOX growing on different carbon sources in chemostat cultures.

		DCW ($g\ L^{-1}$)	$Y_{x/s}$ ($g\ DCW\ g^{-1}s$)	Residual methanol ($g\ L^{-1}$)	q_{O2} ($mmol\ g^{-1}\ DCW\ h^{-1}$)	q_{CO2} ($mmol\ g^{-1}\ DCW\ h^{-1}$)	RQ	Lipase activity ($AU\ g^{-1}\ DCW$)
Glyc60/Meth40	X-33 ROL	12.20±0.61	0.58±0.01	0	4.83±0.11	2.93±0.31	0.61±0.07	6291±362
	X-33 2POS5	11.39±0.39	0.54±0.01	0	4.97±0.16	3.38±0.18	0.68±0.04	7631±830
	X-33 AtOX	10.84±0.17	0.53±0.02	0	4.45±0.42	3.36±0.36	0.76±0.09	6685±647
Gluc60/Meth40	X-33 ROL	11.52±0.12	0.59±0.01	5.23±0.03	3.53±0.25	3.05±0.28	0.86±0.07	2830±111
	X-33 2POS5	10.65±0.11	0.59±0.02	4.90±0.02	3.36±0.07	3.38±0.47	1.01±0.09	3807±225
	X-33 AtOX	10.29±0.11	0.54±0.03	5.45±0.06	3.23±0.18	3.33±0.37	1.03±0.08	3283±208
Gluc20/Meth80	X-33 ROL	3.39±0.14	0.48±0.01	16.23±0.04	3.90±0.16	4.21±0.08	1.08±0.03	3014±112
	X-33 2POS5	3.15±0.13	0.43±0.04	15.39±0.05	3.82±0.18	4.46±0.35	1.17±0.06	4270±139
	X-33 AtOX	2.77±0.19	0.42±0.02	17.47±0.64	3.9±0.29	5.66±0.52	1.45±0.07	3729±189

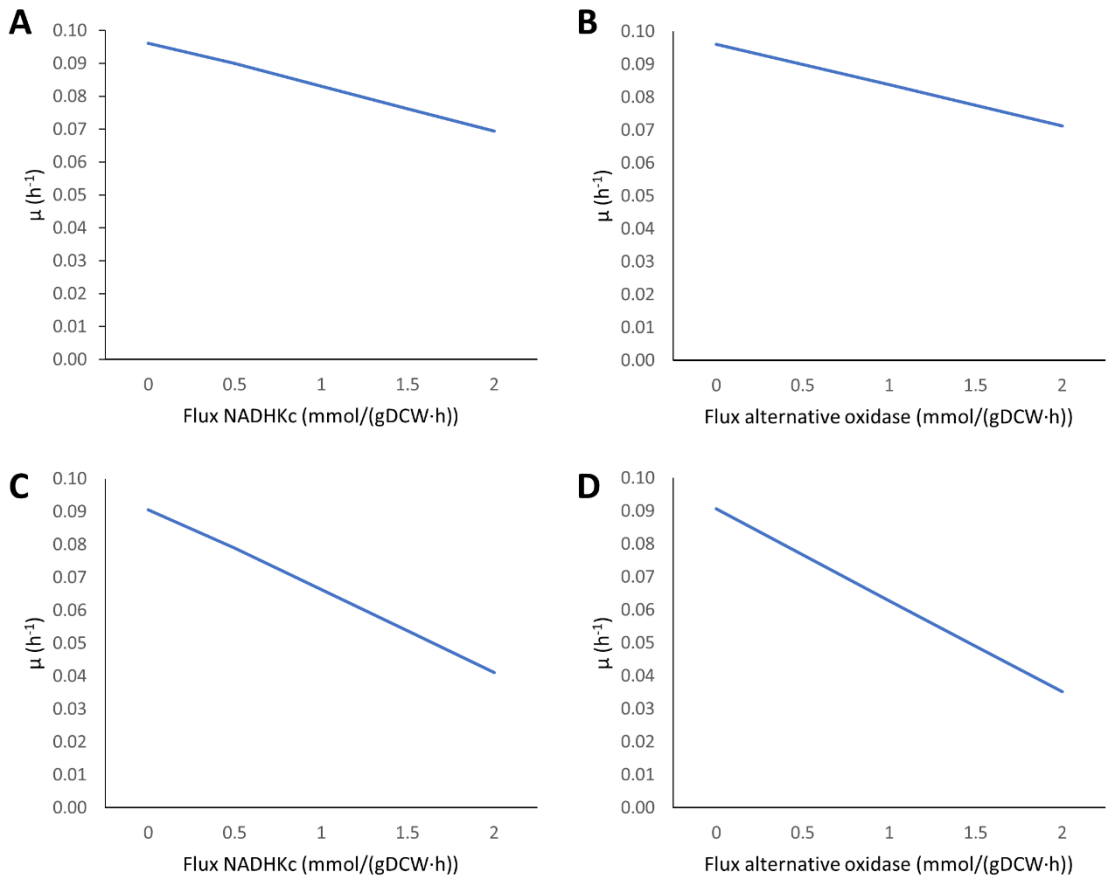
DCW: dry cell weight; $Y_{x/s}$: biomass yield; q_{O2}: specific oxygen consumption rate; q_{CO2}: specific carbon dioxide production rate; RQ: respiratory quotient; AU: activity units.



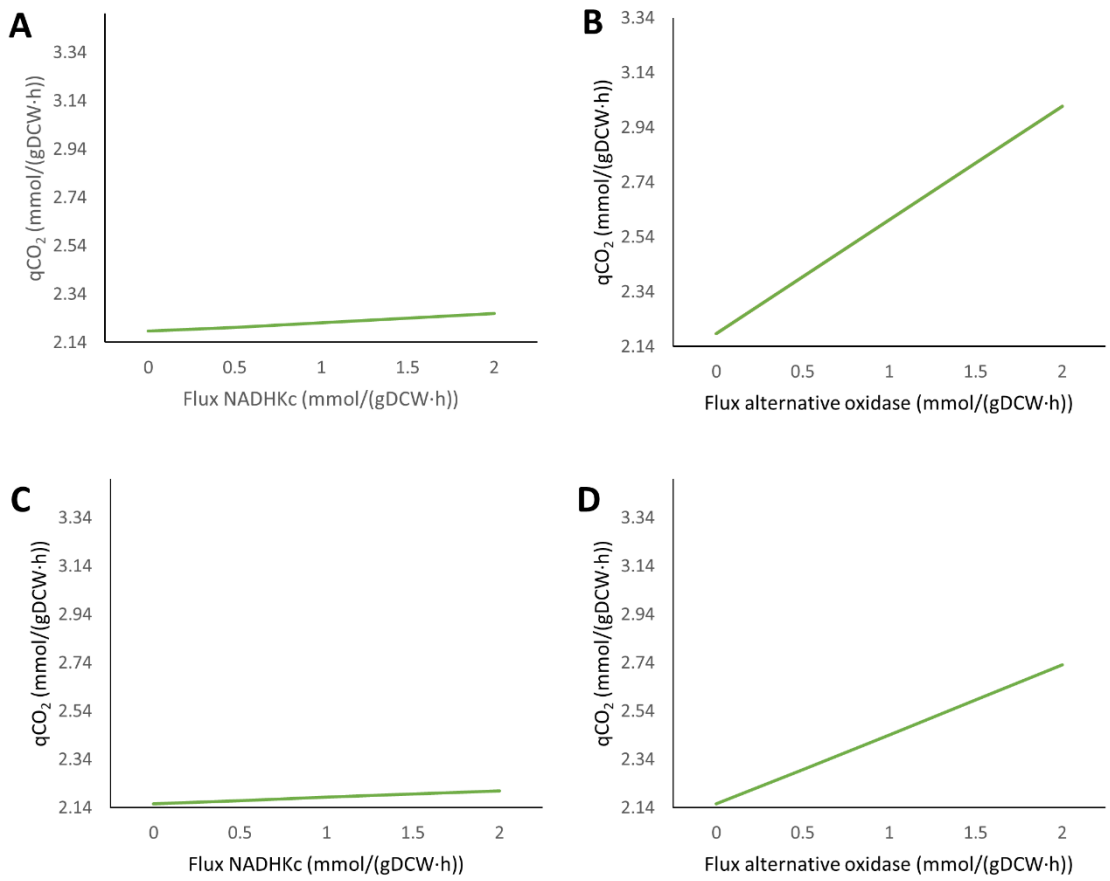
Supplementary figure I. Plasmid maps of pPUZZLE_cPOS5 (A), pPUZZLE_mPOS5 (B) and pPUZZLE_AtOX (C). The restriction enzyme used for plasmid linearization (AvrII) cuts in the middle of pGAP. pPUZZLE contains the kanMX gene for kanamycin resistance in bacteria (*Escherichia coli*) and geneticin G418 resistance in yeast (*P. pastoris*).



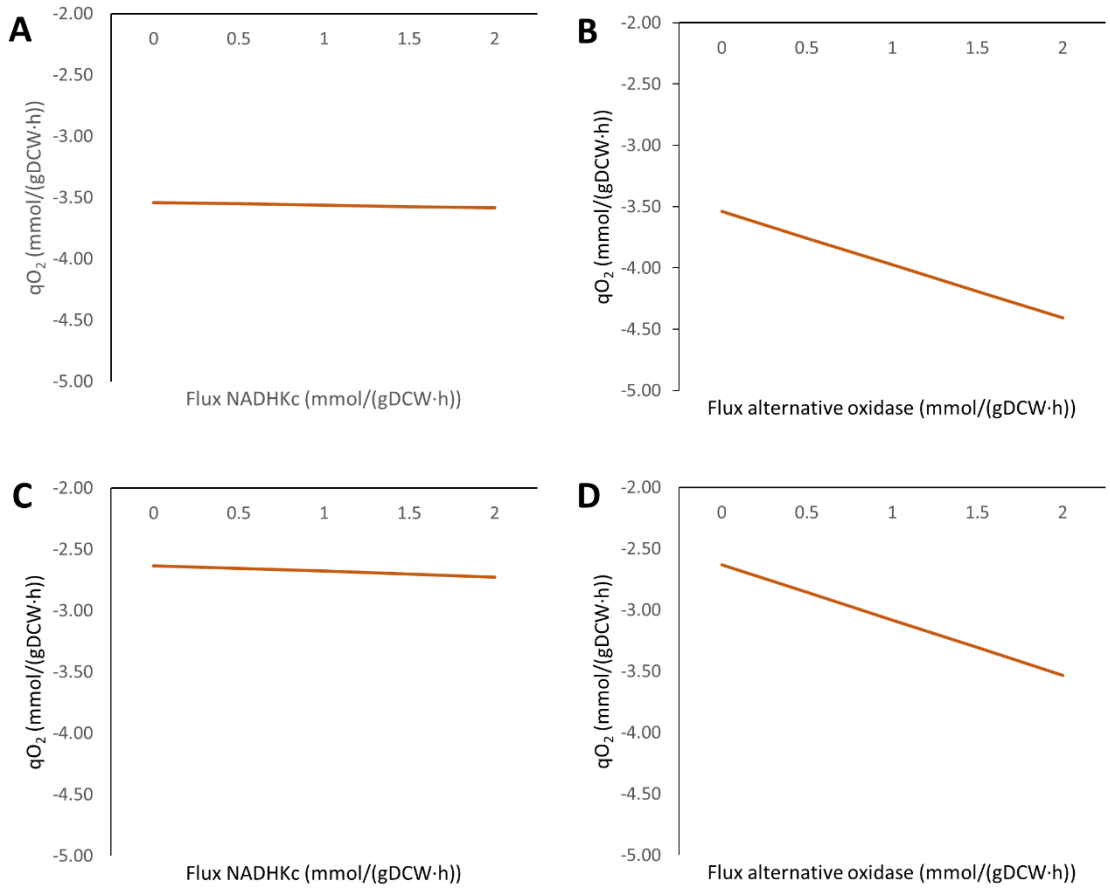
Supplementary figure II. Transcriptional analysis of chemostat cultivations. Expression values are normalized against the control strain for each condition. Each figure corresponds to a different condition (different ratio cosubstrate/methanol). **A)** Glycerol/Methanol (60/40%). **B)** Glucose/Methanol (60/40%). **C)** Glucose/Methanol (20/80%). *ROL, Rhizopus oryzae* lipase; *AOX1,* alcohol oxidase 1; *FDH1,* formate dehydrogenase 1; *FBA1-2*, fructose 1,6–bisphosphate aldolase; *TDH3*, glyceraldehyde-3-phosphate dehydrogenase; *ZWF1*, glucose-6-phosphate dehydrogenase; *IDH2*, isocitrate dehydrogenase 2. *p-val < 0.05, **p-val < 0.01, compared to control strain.



Supplementary figure III. Representation of μ in relation to the flux of NADHKc (A, C) and alternative oxidase (B, D) in simulations using glycerol (A, B) and glucose (C, D) as co-substrates.



Supplementary figure IV. Representation of q_{CO2} in relation to the flux of NADHKc (A, C) and alternative oxidase (B, D) in simulations using glycerol (A, B) and glucose (C, D) as co-substrates.



Supplementary figure IV. Representation of q_{O2} in relation to the flux of NADH kinase (A, C) and alternative oxidase (B, D) in simulations using glycerol (A, B) and glucose (C, D) as co-substrates.

7. General conclusions and future outlook

The increasing interest in the yeast *P. pastoris* as chassis platform for recombinant protein production has boosted quantitative physiological studies at different levels (genomic, transcriptomic, proteomic, metabolomic and fluxomic) over the past 10 years. Consequently, the increasing amount of datasets available has expanded the knowledge base, allowing for the development of novel strategies to improve heterologous protein production. To this end, a wealth of cell engineering studies over the past years have been focused on improving *P. pastoris'* folding and secretory machinery. In this study, we have implemented three different rational cell-engineering strategies to enhance recombinant protein production in the well-established P_{AOX1}-based *P. pastoris* expression system, aiming to target cellular processes that are less explored for cell engineering purposes.

The first strategy has been based on the manual extraction of information (i.e. *learning*) from previously obtained transcriptomic datasets, and subsequent formulation of novel hypothesis for strain engineering based on such information. In particular, earlier transcriptomic analyses pointed out to a transcriptional limitation in *ROL* multicopy strains. This led us to postulate a titration effect of essential P_{AOX1}-activating transcriptional factors caused by the presence of multiple copies of *ROL* expression cassettes. Specifically, we proposed a limitation of Mxr1 and Mit1 transcriptional factors (involved in regulation of P_{AOX1} and other promoters of methanol metabolism-related genes) as main factors involved in the reduced transcription of these genes, including *ROL*. We verified this hypothesis by constitutively overexpressing these transcription factors (TF) under control of the glyceraldehyde-3-phosphate dehydrogenase promoter (P_{GAP}) in a strain carrying 4 *ROL* copies. Notably, we also demonstrated that this approach increases methanol assimilation capacity of TF-engineered cells by de-regulating methanol metabolism. However, since the improvement in the extracellular lipase activity was not observed at bioreactor scale, we proposed that some other bottlenecks other than transcription may be limiting Rol productivity.

The emergence of systems biology, including the construction of genome-scale metabolic models (GEMs), has provided a series of powerful tools for rational strain engineering, thereby facilitating the elucidation of novel, hitherto unknown genetic targets to engineer cells for enhanced performance. Hence, the second strategy has been based on the use of the consensus genome-scale metabolic model of *P. pastoris* iMT1026 to determine gene knock-outs which could enhance Rol production. The *in silico* analyses pointed out at two potential elements limiting the production of our model protein, i.e. NADPH generation and availability of certain amino acids. Out of the six gene knock-out targets which were identified, four were genes associated with central carbon metabolism. From these, two gene deletions increased flux through the oxidative branch of the pentose phosphate pathway (PPP) and the two remaining knock-outs boosted the flux through the (NADP-dependent) isocitrate dehydrogenase (IDH) reaction. Both PPP and IDH reaction are sources of NADPH. While PPP is a well-known target in strain engineering, the potential role of IDH in recombinant protein production strategies has been proposed for the first time here. The other two *in silico*-selected gene knock-outs were related to amino acid metabolism of serine and cysteine. Serine is an amino acid that are is in high relative proportion in Rol amino acid sequence. In addition, serine doubles the amount of relative amino acid composition of *P. pastoris*. Cysteine is shown in low proportion in Rol sequence, but its relative amount is several fold higher than the average *P. pastoris'* amino acid composition. So, amino acid availability could be another important factor to take into consideration to increase heterologous protein yields. However, the lack of success in obtaining gene-knocked strains prevented us from verifying these hypotheses, leaving this work as mainly theoretical.

Third, since NADPH has been identified as an important cofactor affecting recombinant protein production, we aimed to increase its regeneration in the cell by knocking-in genes coding for ectopic enzymes directly involved in redox cofactor balance: a NADH kinase and an alternative oxidase. As a result, increased Rol activities have been obtained, further supporting the role of NADPH in maintaining high levels of recombinant protein production. Moreover, a transcriptomic analysis and an *in silico* determination of metabolic fluxes was performed to provide a metabolic

interpretation of the observed macroscopic parameters. In particular, qPCR analyses revealed a transcriptional regulation of PPP depending on the redox state of the cell, supported by the *in silico* fluxes calculated in the simulations. It is particularly intriguing that NADH kinase effect seems to have a stronger positive effect when expressed in the cytosol than in the mitochondrion. This could be explained by a strong compartmentalization of NADPH (the cofactor cannot be exchanged between cytosol and mitochondrion) and the fact that most of the NADPH-consuming reactions related with amino acid synthesis (such as proline, lysine, glutamate and aspartate metabolism) take place in the cytosol.

Overall, this work implements different tools and methodologies in order to increase recombinant protein production with different degrees of success and provides also two examples of GEM applications, i.e. for *designing* strain engineering strategies and for dataset interpretation (*extraction of knowledge* or *learning*). However, the utilization of GEMs to predict metabolic fluxes and make hypothesis to guide strain engineering strategies must be taken with caution. Most models lack of regulatory and kinetic information and are limited to the cell's biochemical reactions. In the near future, novel GEMs integrating metabolic networks and regulatory mechanisms should be developed to provide more accurate predictions of cellular phenotypes.

Another key problem in data-driven exploration of biological systems, such as the information provided by -omic data or GEMs analysis, is the extraction of knowledge from this enormous and highly complex amount of information. Machine (deep) learning algorithms are an artificial intelligence (AI)-based approach that could help to automatize this data interpretation. This field of AI focuses on improving prediction accuracy through experience, given specific processable data from which it is able to learn and evolve. Therefore, integration of GEMs and machine learning techniques is an exciting challenge of synthetic biology that could benefit metabolic engineering by finding hidden patterns in metabolic fluxes obtained during *in silico* analysis.

Concerning strain construction, further investigation should be focus on the application of multiple genetic modifications in the same strain, such as increasing *ROL* transcription and NADPH supply simultaneously. This way, it would be possible to achieve even further heterologous protein yields than the ones reported in this study.

In order to efficiently and rapidly construct these strains, it is imperative to develop multiplexed genome engineering tools, especially when performing gene knock-outs. Further optimizations of the CRISPR/Cas9 system may help in this regard.