UAB

Universitat Autònoma de Barcelona

Escola d'Enginyeria

Departament d'Enginyeria Química, Biològica i Ambiental

Developing strategies for systems metabolic engineering in *Pichia pastoris*

Memòria per obtenir el Grau de Doctor
per la universitat Autònoma de Barcelona
dins del Programa de Doctorat en Biotecnologia
sota la direcció dels doctors
Joan Albiol Sala i Pau Ferrer Alegre

Màrius Tomàs Gamisans

Bellaterra, Juny 2017

El Dr. Joan Albiol i Sala i el Dr. Pau Ferrer i 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 Màrius Tomàs Gamisans ha dut a terme sota la nostra direcció el treball que, amb títol "**Developing strategies for systems metabolic engineering of *Pichia pastoris*"**, es presenta en aquesta memòria i constitueix la seva Tesi per optar al Grau de Doctor per la Universitat Autònoma de Barcelona en el programa de Doctorat en Biotecnologia.

I per tal que se'n prengui coneixement i consti als efectes oportuns, signem la present a Bellaterra, 26 de juny de 2017

Dr. Joan Albiol i Sala   Dr. Pau Ferrer i Alegre   Màrius Tomàs i Gamisans

(Co-director)   (Co-director)   (Autor)

Title   Developing strategies for systems metabolic engineering of *Pichia pastoris*

Author   Màrius Tomàs Gamisans

Supervisors   Dr Joan Albiol Sala, Dr Pau Ferrer Alegre

Keywords   *Pichia pastoris,* systems metabolic engineering, genome-scale metabolic model, systems biology, recombinant protein production

The way of progress is
neither swift nor easy

Marie Curie

Abstract

Pichia pastoris has become one of the most extensively used platform cell factories for recombinant protein and high-value added metabolite production. In the past recent years, important breakthroughs in the systems-level quantitative analysis of its physiology have been achieved. This wealth of information has allowed the development of genome-scale metabolic models, which make new approaches possible for host cell and bioprocess engineering. Previous to this work, three different genome-scale metabolic models were available for *P. pastoris*. Nevertheless, these models showed some inconsistencies regarding certain pathways, including the terminology for both metabolites and reactions and annotations. Furthermore, some *P. pastoris* specific metabolic traits were misrepresented. Therefore, in this study, a consensus genome-scale metabolic model has been developed, thereby integrating the prior models. In addition, a comprehensive revision of metabolic pathways was performed and several pathways were curated and updated according to the currently available literature. As a result, the new model, iMT1026, is able to more accurately reproduce experimental growth parameters using glucose as carbon source and different oxygen availability conditions. In order to expand the capabilities of the consensus model, new physiological datasets of cells growing on two of the most relevant substrates for this cell factory were generated. Specifically, a series of chemostat cultivations were performed to characterise the physiologic profile and macromolecular biomass composition of *P. pastoris* growing on glycerol and methanol as sole carbon sources. Also, macromolecular biomass composition was analysed, allowing us to incorporate new carbon-source specific stoichiometric biomass equations into the model, as well as to estimate the associated energetic parameters. Overall, a new version of the model (iMT1026 v3.0) was validated for these growing conditions.

In addition to the validation of iMT1026 v3.0 for a wider range of carbon sources and growth conditions, we have further tested its performance in two different applications, namely, the generation of reduced metabolic models suitable for ^{13}C-based metabolic flux analysis and, assisting the interpretation of physiological growth parameters of redox-cofactor engineered strains. In particular, the genome-scale metabolic model has been reduced into a core model and used for ^{13}C-based metabolic flux analysis of cells growing on glycerol at different growth rates. To our knowledge, this is the first study ever reported of 13C-MFA using glycerol as sole carbon source. Notably, flux analyses are highly consistent with pioneering ^{13}C-based metabolic profiling studies of *P. pastoris* growing on glycerol. iMT1026 v3.0 was also employed for assisting the interpretation of the physiological profiles obtained for redox-cofactor engineered strains. A recombinant strain producing an antibody fragment was engineered to overexpress a heterologous NADH kinase, aiming at increased NADPH regeneration rates. Notably, the redox-engineered strains showed an increase in recombinant protein production and altered macroscopic growing profiles. *In silico* analysis of the impact of NADH kinase overexpression using the iMT1026 model predicted possible metabolic changes associated to the redox cofactor imbalance that were in agreement with the observed physiological phenotypes.

Overall, a refined tool for systems metabolic engineering is provided in the present study. Moreover, such tool has been validated for a wide range of environmental conditions and employed in two different applications, confirming its reliability.

Ressenya

Pichia pastoris s'ha convertit en una de les plataformes cel·lulars més utilitzades per a la producció de proteïnes recombinants i metabòlits d'alt valor afegit. En els darrers anys s'han aconseguit fites importants en l'anàlisi quantitativa a nivell de sistemes de la seva fisiologia. Aquesta gran quantitat d'informació ha permès desenvolupar models metabòlics a escala genòmica, que permeten el desenvolupament de noves estratègies per l'enginyeria de soques i de bioprocés. Amb anterioritat a aquest estudi s'havien publicat tres models metabòlics a escala genòmica per a *P. pastoris*. No obstant això, aquests models presentaven algunes inconsistències en algunes vies metabòliques, en la nomenclatura de metabòlits i reaccions, així com les anotacions associades a certes vies. A més, algunes de les rutes metabòliques o característiques específiques de *P. pastoris* eren representades de forma errònia o incompleta. És per això que en aquest estudi es desenvolupa un model metabòlic a escala genòmica consens, que integra els models anteriors. A més, també es fa una revisió exhaustiva de diverses rutes metabòliques i nombroses vies es corregeixen i actualitzen d'acord amb les publicacions disponibles. Com a resultat, el nou model, iMT1026, pot reproduir amb més precisió els paràmetres de creixement experimentals de cèl·lules creixent en glucosa i diferents nivells de disponibilitat d'oxigen. Amb la voluntat d'expandir les capacitats del model, es generen noves dades fisiològiques fent servir dos dels substrats més importants per aquesta factoria cel·lular. Es realitzen unes series de cultius en continu per a la caracterització del perfil fisiològic de *P. pastoris* creixent en glicerol i metanol com a fonts úniques de substrat. A més, també es caracteritza la composició macromolecular de la biomassa. Posteriorment, s'incorporen en el model noves equacions de biomassa específiques per a cada font de carboni.

Aquestes noves dades experimentals han permès estimar els paràmetres energètics associats a les fonts de carboni i validar el model (iMT1026 v3.0) per aquestes condicions de creixement.

Tot i la validació de iMT1026 v3.0 en un rang més ampli de condicions, en aquest treball es prova en dues aplicacions diferents: en l'anàlisi de fluxos metabòlics basat en ^{13}C i com a eina de suport per a la interpretació de resultats en soques amb modificades en el metabolisme redox. Tot i que hi ha un únic estudi on s'analitza la relació entre fluxos metabòlics en cèl·lules creixent en glicerol, no es té constància de cap estudi d'anàlisi de fluxos metabòlics en aquesta font de carboni. Així doncs, es redueix el model metabòlic a escala genòmica a un model del metabolisme central i es fa servir per a l'anàlisi de fluxos metabòlics basats en ^{13}C en cèl·lules creixent amb glicerol a diferents velocitats de creixement. Els resultats obtinguts són molt consistents amb els cultius previs en glicerol. També s'utilitza iMT1026v3.0 com a suport per a la interpretació del perfil fisiològic obtingut en soques amb el metabolisme redox modificat. Una soca que expressa un fragment d'anticòs es modifica genèticament mitjançant l'expressió d'una NADH quinasa, de manera que el balanç de cofactors redox queda pertorbat. Les soques generades mostren una producció de proteïna recombinant més elevada i una alteració en el perfil macroscòpic de creixement. Mitjançant l'anàlisi *in silico* dels perfils fisiològics resultants, es prediuen possibles canvis metabòlics associats a l'alteració del balanç de cofactors que estan d'acord amb el perfil macroscòpic observat. Així doncs, en línies generals, en aquest treball es desenvolupa una eina precisa per a l'enginyeria de sistemes metabòlics. A més, és validada en condicions vàries i s'utilitza en dues aplicacions diferents que demostren la seva fiabilitat.

TABLE OF CONTENTS

List of abbreviations

CI	Confidence interval
D	Dilution rate (h-1)
DCW	Dry cell weight
EMU	Elementary metabolite units
ER	Endoplasmatic reticulum
ETC	Electron transport chain
GAME	Growth-associated maintenance energy
GC-MS	Gas chromatography coupled with mass spectrometry
GRH	Growth rate hypothesis
GSH	Glutathione, reduced form
GSMM	Genome-scale metabolic model
GSSG	Gluthathione oxidised, form
HPLC	High-performance liquid chromatography
MDV	Mass distribution vector
MFA	Metabolic flux analysis
MID	Mass isotope distribution
NADH/NAD$^+$	Nicotinamide adenine dinucleotide reduced / oxidised form
NADPH/NADP$^+$	Nicotinamide adenine dinucleotide phosphate reduced / oxidised form
NGAME	Non-growth associated maintenance energy
PDI	Protein disulphide isomerase
PPP	Pentose phosphate pathway
PTM	Post-translational modifications
q$_{CO2}$	Specific CO_2 production rate
q$_{O2}$	Specific oxygen consumption rate
qP	Specific product generation rate
qS	Specific substrate consumption rate
RD	Reduction Degree
RNA	Ribonucleic acid
RQ	Respiratory coefficient
TCA	Tricarboxilic acid
UPR	Unfolded protein response
Yxs	Biomass yield
μ	Growth rate (h^{-1})
τ	Residence times

1

General introduction

Table of contents

1. General introduction

The United Nations, in 1992 defined biotechnology in the Convention on Biological Diversity as any technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use.

Despite the definition was established 25 years ago, ancient human civilisations already performed the first biotechnological processes: microorganisms were used for beer, bread, cheese and yogurt production. These processes were at first rudimentary and discovered by chance. Initially, and for centuries, biotechnological products were just basic consumer goods. Nevertheless, more recently, advances in research and scientific discoveries allowed for the exploitation of microorganisms in a broad range of applications. For example, from the development of the acetone-butanol-ethanol process, citric acid production in *Aspergillus niger*, the antibiotic production in the first half of the 20th century, to the current production of drugs, vaccines and a wide variety of enzymes and valuable chemicals. Nowadays, biotechnology rests on the foundations stablished during the past century, with the discovery of the double helix of DNA, the capacity of genome sequencing, the development of DNA-editing technologies that enabled the development of genetic engineering and the improved computational capacity and methods that facilitates processing large volumes of information.

The application of these new knowledge and technologies boosted a rise in the field of biotechnology. In fact, this field is in an exponential growth with the implementation of new tools and processes that can replace conventional industrial processes. Generally, the concern on the environment protection made industries to overhaul the actual manufacturing processes, turning them into more environmentally-friendly, using renewable sources and evolving towards a bio-based industry. New discoveries in biological sciences and its implementation together with the reassessment of the industrial technologies, point to the key role of biotechnology on facing the new challenges in the present century that has been considered the biotechnology century.

1.1. YEAST BIOTECHNOLOGY: *P. PASTORIS* AS CELL FACTORY

Yeast have been used for thousands of years for alcoholic fermentation of beverages and processing a variety of foods. The term yeast includes a wide diversity of species, nevertheless it is commonly associated to *Saccharomyces cerevisiae* that is the most common and extensively used yeast system.

In industrial biotechnology, the productivity (space-time yield) is one of the major determinants for process profitability together with process simplicity and cheapness. When using organisms, the productivity is dependent on the specific growth and high turnover rates. Thus, organisms with higher growth rates would be a preferable choice for industrial processes [1]. Bacteria are the simplest organisms with higher growth and turnover rates and therefore the first option to choose in industrial processes. Nevertheless, their expression system has some limitations such as low stress tolerance and the inability of performing posttranslational modifications and consequently to properly express complex proteins [2]. Yeasts share some of the advantages with bacterial systems: they are unicellular organisms, easy to grow at high cell densities in cheap media and easy to genetically manipulate [3]. Moreover, unlike bacteria, yeasts are eukaryotic organisms and it confers them extended highly valuable capabilities for industrial applications. Together with mammalian cells, yeasts are able to perform posttranslational modifications (PTM) and secreting proteins, thus able of expressing complex proteins. Overall, yeasts are in between bacterial and mammalian platforms, thereupon share some advantages and disadvantages with these two other systems (**Table 1-1**). Despite there is not a universal platform of choice for biotechnological processes because it depends specifically on each particular case, yeasts are very versatile and well stablished platforms for multiple industrial applications such as the expression of recombinant proteins or the production of ethanol and other value-added chemicals.

S. cerevisiae is known for growing in high sugar concentration environments and for its high capacity for producing ethanol. However, among the wide variety of yeasts species, there are a few of them valuable for other industrial applications with a reduced capacity on fermentative pathways [3,4]. These species are also called 'non-conventional' yeast and are able to grow in a variety of substrates with alternative metabolic routes, different regulatory

patterns and the ability to generate alternative products [5]. *Pichia pastoris*, one of the most prominent non-conventional yeast is studied in the present thesis.

Table 1-1. Main characteristics of different industrial biological systems.

	Advantages	Disadvantages
Bacteria	High growth rates High product yields Cultivation at high cell densities Cheap cultivation media Easy to manipulate genetically Scalable processes	Low stress tolerance Lack of posttranslational modifications No cellular compartments
Yeast	High growth rates High product yields Cultivation to high cell densities Cheap cultivation media Easy to manipulate genetically Scalable processes Posttranslational modifications Protein folding and secretion Compartmentalisation of the cell	Glycosylation patterns different than mammalian
Mammalian cells	Posttranslational modifications Protein folding and secretion Compartmentalisation of the cell	Expensive cultivation media Difficult to scale up processes

    P. pastoris was isolated in 1970s by Phillips Petroleum Company looking for yeasts capable to metabolise methanol. A decade later, *P. pastoris* was developed as heterologous protein production platform [6] taking advantage of the strong methanol inducible *AOX* promoter. In fact, it has become one of the most commonly used hosts for recombinant protein production [3,7] including biopharmaceuticals [8,9]. In the mid 1990s, *Pichia* was reassigned to the genus *Komagataella* [10]. Later, industrial *P. pastoris* strains were analysed and reclassified as *Komagataella spp.*, that include two different strains (*K. phaffi* and *K. pastoris*) [11]. Since 1995, the number of heterologously expressed genes in this methylotrophic yeast has steadily increased [12]. *P. pastoris* has many engaging properties for recombinant protein production. One of the main advantages is the availability of well-established protocols and

techniques for its genetic manipulation [13–15] that enables strain engineering. As eukaryote has the capability of performing PTMs that makes it suitable for the expression and secretion of recombinant proteins, even those with intricate structure that require PTMs [16]. Moreover, there are a number of promoters available which are suitable for regulation with different carbon sources [17,18]. As most of the yeasts, *P. pastoris* is able to grow up to high cell densities and achieving high protein production yields [19,20]. Due to the increasing interest in using this advantageous methylotrophic yeast, several efforts have provided tools to better understand its operation, from physiological characterization to metabolic modelling. Furthermore, progress in synthetic biology together with its extensive use in recombinant protein production have also opened the door towards utilising *P. pastoris* as whole cell biocatalyst for non-native value-added metabolite production [21–23].

1.1.1. Recombinant protein production

Nowadays *P. pastoris* has become the most frequently used eukaryotic system for recombinant protein production [12]. The fact that secretes low amounts of endogenous proteins extremely facilitates de downstream process for protein purification, conferring to this production platform an enormous advantage among other alternative hosts [24]. Moreover, recombinant product titters are reported to be higher than other yeast species [1,7] due to its particular secretory system. Furthermore, protein glycosylation patterns in *P. pastoris* have been extensively studied and reported to be much shorter than the hypermanosylated chains in *S. cerevisiae* [25–27]. The glycosylation pattern in recombinant proteins is important in order to ensure the proper protein activity and avoid antigenicity [28]. In addition, *P. pastoris* strains have been modified for producing human-like glycosylation patterns [29] allowing for the expression of biopharmaceuticals [30], including antibodies and antibody fragments both for therapeutic and immunodetection purposes [31,32].

Bottlenecks on recombinant protein production

Despite the advances and extensive use of *P. pastoris* on recombinant protein production, there are some bottlenecks that hamper the enhancement of recombinant protein production and the increase in production yields. During heterologous protein

overexpression, two main limitations have been identified: the protein folding and secretion pathway and the metabolic burden.

Protein folding and secretion

In yeast, folding and secretion processes are major limitations when secreting recombinant proteins (**Fig. 1-1**). In the folding and secretory pathway of all eukaryotic systems, the protein is translocated to the endoplasmic reticulum (ER), where takes place the major part of folding and PTMs. Protein folding is facilitated by ER specific proteins such as ATP-dependent chaperones and PDI (protein disulfide isomerases) that help disulfide bond formation [33]. Despite high expression levels of the target protein, in those cases overexpressing complex proteins that require disuphide bond formation and post-
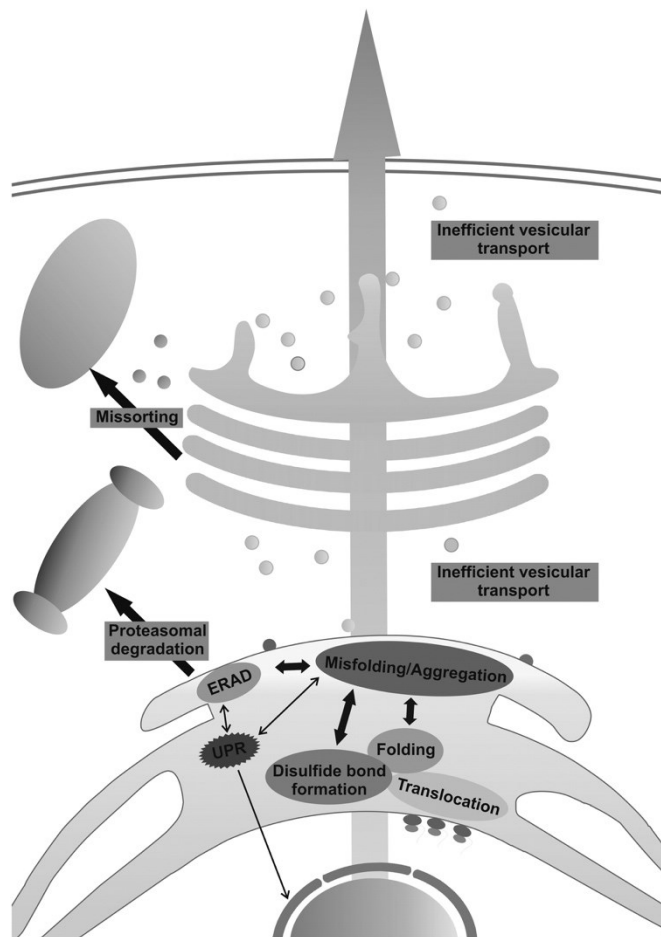
Fig. 1-1. Bottlenecks encountered by recombinant proteins on their way through the secretory pathway in eukaryotic hosts. Reproduced from [35] with permission of Mary Ann Liebert, Inc.

translational modifications, the folding and secretion machinery is not able to properly fold and secrete the recombinant protein, and a fraction is retained intracellularly [34]. Consequently, misfolded or unfolded proteins accumulate leading to ER stress and the activation of the unfolded protein response (UPR) signalling pathway [35] that triggers misfolded proteins refolding or elimination via ER-associated degradation (ERAD). ATP-dependent chaperones aid in protein folding; thus, several attempts of protein refolding are ATP expensive for the cell. In addition, complex proteins require a correct pairing of sulfhydryl groups for correct folding. The oxidized environment in ER favours disulphide bound formation, in addition, PDI oxidises the sulfhydryl groups for sulphide bound formation. ERO1 subsequently oxidises PDI, so it can be reused to further catalyse the disulphide bounds formation. At the same time, ERO1 is reoxidised transferring the electrons to O_2 leading to the reactive oxygen species (ROS) production (**Fig. 1-2**). ROS detoxification is mediated by the oxidation of glutathione (GSH) to GSSG (oxidised gluthathione). Thus, glutathione reductase converts oxidised GSSG to reduced GSH at expenses of NADPH in order maintain the glutathione redox potential [36,37].

In addition, wrong sulfhydryl groups pairing leads to misfolded proteins. Therefore, reduced glutathione (GSH) assists in reducing the non-native disulphide bonds and PDI can oxidise anew the sulfhydryl group pairs. Thus, when misfolded proteins accumulate, a futile cycle in disulphide bond formation (with PDI) and breakage (with GSH) may increase the NADPH demand for GSH reoxidation. In addition, it implies the formation of additional ROS that will also lead in an extra demand of NADPH for its detoxification [38]. Therefore, when producing a complex protein, an additional supply of both ATP and NADPH is required for folding and secretion processes, contributing to the global metabolic burden.
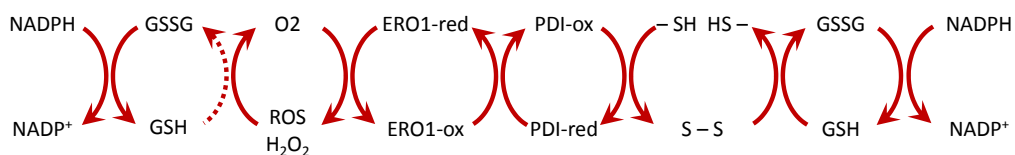
Fig. 1-2. Oxidative disulphide bond formation in ER. Sulphydryl groups (*–SH*) are oxidised to form disulphide bounds (S–S) by protein disulphide isomerase (PDI). PDI is subsequently reoxidised by ERO1 that uses O_2 for its reoxidation and generates reactive oxygen species (ROS) and H_2O_2. ROS are eliminated by glutathione (GSH) that is oxidised and requires NADPH for being regenerated. Disulphide bonds can be undone by GSH for further PDI folding attempts.

Once the protein is properly folded, it is subsequently glycosylated and passes the quality control in ER to further being transported through Golgi apparatus. After additional glycosylation steps, the protein can be secreted or transported to other organelles. Protein glycosylation is an additional resource-consuming step that could drain cell resources affecting its fitness. The last step of protein sorting and transport can also be a limitation in efficient protein secretion [35,39]. Failure in the process of sorting to the vacuoles leads to protein degradation and recycling [40]. In order to overcome this limitations, some strategies focusing at different levels of the secretory pathway can be applied (extensively reviewed in [41–43]). Thus, by overexpressing certain transcription factors, chaperones, foldases and other secretion enzymes, the secretion process may be debottlenecked and the recombinant protein production can be improved.

Limitations in the energy and building block supply chain

Since the early beginnings of the establishment of recombinant proteins expression processes, the existence of a metabolic burden consequence of the overexpression of recombinant proteins was reported: the expression of foreign DNA adds an additional load to the cell due to the drain of cell resources into the protein production instead of for its own cell metabolism [44]. When cell is not able to cope with this overload, not only several physiological changes may occur, but also a limitation in protein production would impair high production yields [45]. Indeed, heterologous protein overexpression has a major impact on cells and can alter its composition, flux distribution and even reduce its maximal growth rate [46]. Both in *Schizosaccharomyces pombe* and *P. pastoris* an increase in lipid content is observed in recombinant protein expressing strains that could be correlated with an increased demand of membranes for vesicles in the secretory pathway [47–49].

In order to elucidate the derived physiological changes of protein overexpression, several studies focused on the metabolic flux redistribution analysis in recombinant strains [47,49–51]. Overall, cell metabolism responds similarly, with an increase in the pentose phosphate pathway (PPP), that is the main source of NADPH in cytosol. In addition, flux to tricarboxylic acid (TCA) cycle also seems to be affected, but in diverse manner among different yeast species [52]. Therefore, energy supply would be also conditioned by the expression of the recombinant protein. In fact, the metabolic burden is caused both by the

increased demand in precursors for amino acids biosynthesis and the additional energetic requirements in protein synthesis, folding and secretion. Yeasts are usually cultivated in minimal media, consequently, amino acids have to be synthesized de novo. Therefore, there is an additional demand on ATP, NADPH and precursors in central carbon metabolism that have to be redirected to amino acid biosynthesis, leading to a deep impact in central carbon metabolism [52,53]. Moreover, protein biosynthesis is a very expensive process that requires GTP and ATP consumption for aminoacyl-tRNA charging [44]. Furthermore, as detailed above, protein folding and secretion are energy intensive processes, consuming ATP for chaperone binding, and NADPH to assist disulphide bond formation and alleviating ER oxidative stress. Hence, the metabolic burden is associated to the supply of the additional resources (energetic and building blocks) drained in the increased anabolic processes as well as in protein biosynthesis.

Strategies for overcoming limitations on recombinant protein production

In production processes, the detection of bottlenecks is a key point for facing the limitations in the employed system and basic for choosing the best strategies to improve productivities. There is not a universal strategy that could improve global yields and therefore it has to be case-specific. Nevertheless, one of the first steps to check is whether the production system is performing in stable and non-stressing conditions. Subsequently, several advanced strategies can be considered. Generally, strategies for improving protein production can be classified into process and genetic engineering strategies.

Process engineering strategies

This group of actions is mainly related to the modification of cultivation conditions. Choosing the most appropriate operational mode and feeding profile can significantly improve cell performance [19,54–56]. In addition, several strategies focus on alleviating the impact of the metabolic burden caused by the expression of the recombinant protein. Detailed studies on operational settings of cultivation, showed a strong influence of environmental conditions on the productivity in yeast cells [57,58]. The reduction of growth temperature has generally showed an improvement in recombinant protein production [58]. Indeed, in *P. pastoris*, the reduction of the growth temperature from 30ºC to 20ºC resulted in a 3-fold increase in antibody fragment production [59]. Authors suggest that lower temperature might

reduce ER folding stress due to a higher protein stability and a reduction of protein denaturation, aggregation and subsequent refolding. Surprisingly, a reduction in oxygen availability is likewise a successful strategy to improve recombinant protein production in yeast [60–63]. Furthermore, another of these strategies, and specially implemented in *P. pastoris*, is the utilization of mixed substrates for improving the energy supply [64,65]. With co-substrate utilisation, one of the substrates is commonly used for yielding biomass, while the other can generate the energy and cofactor surplus to balance the metabolic burden. Several examples can be found in yeast literature about using substrate mixtures such as *S. pombe* grown on glycerol and acetate [66] and *P. pastoris,* on mixtures of sorbitol/methanol [46], glucose/methanol [49], or glycerol/methanol [67]. Similarly, supplementation of cultivation media with amino acids relieves consuming cell resources and energy in de novo amino acid biosynthesis and allows for an improved performance [52,53]. Besides, media can include limiting precursors required for the desired product: for example, hemin supplementation is able to improve the recombinant production of an heme peroxidase in *P. pastoris* [68].

A success case: hypoxic cultivation enhances recombinant protein production

Among environmental strategies, cultivation in hypoxic conditions has centered the attention of *Pichia* community and comprehensive studies have focused on the physiological response of *P. pastoris* in reduced oxygen supply cultivations [69]. *P. pastoris* is a Crabtree-negative yeast, nevertheless, in hypoxic conditions it turns its metabolism to respirofermentative. Thus, ethanol and other by-products such as arabitol are detected in oxygen limited conditions [69]. Similarly, *S. cerevisiae* produces ethanol and glycerol in hypoxic conditions for redox cofactor regeneration. Surprisingly, strains expressing an antibody fragment (Fab) grown in hypoxic conditions showed a 2-fold increase in productivity [62].

Moreover, several physiological changes are reported in such conditions that may explain the increase in productivity. Biomass composition was clearly affected by different oxygen levels [48], particularly, important differences were appreciated in lipid profile [70]. These differences may influence membrane fluidity, and facilitate Fab secretion [71]. In addition, it has also been reported a strong impact on transcriptional profile and metabolic carbon flux redistribution [69]. Hypoxia induces stress to the cell and the activation of the

UPR. With the overexpression of chaperones, and the protein folding machinery, the secretion of the target protein is enhanced. Furthermore, the flux through PPP is reduced and redirected to the glycolytic pathway. Variations in flux distribution are in correlation with the same pattern in transcriptional analysis, where an increase of transcriptional levels of glycolytic genes is reported [69]. Thus, as Fab is expressed under the control of the glycolytic constitutive GAP promoter, it seems plausible that the increased productivity would be attributable to the increased transcriptional levels of glycolytic genes.

Genetic engineering strategies

The other alternative for enhancing productivities is strain engineering to improve cell capabilities allowing for better performance. Rational strain design has evolved since the first strategies based on current knowledge of cell physiology and biochemical pathways. In *P. pastoris* as well as in other organisms, rational strategies for improving recombinant protein secretion or metabolite production have been developed together with analytical and computational methods. In a first generation of engineered strains, the biological information of cell physiology was used for rationally designing new strains. Thus, an example of first generation of strain engineering would be the amplification of the gene copy number [7,72]. Alternatively, the overexpression of chaperones, PDI, or the transcription factor *HAC1* can also increase the productivity by overcoming limitations in protein folding and secretion [39,73,74].

The second generation of rational strain engineering strategies include those targets inferred from –omics data (e.g. transcriptomics, proteomics, metabolomics, fluxomics) [75]. Basically, strains are further engineered for mimicking the observed physiological data in overexpressing strains [70,71]. Finally, third generation of rational strain engineering is a systems biology-based approach (Systems metabolic engineering or systems biotechnology as later discussed in section 1.2). Briefly, this strategy encompasses all the available information of an organism, integrating different layers of '-omics' information to in silico predict gene targets that optimizes the selected objective (e.g. recombinant protein production, metabolite secretion, etc.) [75,76].

1.1.2. *P. pastoris* as platform factory for biorefining

Citric acid, penicillin and ethanol were the first chemicals produced using microorganisms particularly its respective native producers (*A. niger*, *Penicillium chrysogenum*, *S. cerevisiae*). Originally such processes were optimized by random mutations using UV light or chemical mutant agents and the subsequent screening of mutants [77]. Later, progress in synthetic biology as well as wide experience in strain engineering have also opened the door towards using yeast and other organisms as whole cell biocatalyst and for both non-native value-added metabolite production. Finally, in the last decades, the development of bioinformatics and application of mathematical modelling strategies boosted the production of chemicals improving the traditional processes and enabling the production of new chemicals in novel cell factories [78]. Thus, with current technologies, a certain chemical can be produced by non-native hosts platforms and achieve greater performances.

When developing a production process, choosing the production platform depends on various factors: whether the molecule of interest can be produced with any existing metabolic pathway and organism (**Fig. 1-3**), its potential toxicity for cell, the available substrate, and the metabolic networks of alternative platforms. Then, if the production of the target metabolite in the native organism is scalable, it will be a good starting point for testing the production. Otherwise, alternative platforms would be the best choice [79]. Hence, synthetic biology takes an important role for metabolic engineering of cell factories [80]. In this context, the knowledge in the physiology and metabolic capabilities of the host organism is also basic to determine the feasibility and expected yields of the production process. Therefore, studies on metabolic flux analysis give valuable information for designing and further improving the novel bioprocesses [81]. Yeast arose as excellent cell factories. Unlike bacteria, the compartmentalization of cellular metabolism allows for specifically directing reactions into compartments, avoiding unspecific reactions or undesired products [82]. In addition, the wide variety of yeasts allows for using a huge range of substrates, including inexpensive and renewable carbon feedstocks [83,84].

Generally, production processes are in continuous reassessment in a design-build-test cycle. The producing strains are further improved by cell engineering methods. Traditionally,

adaptive evolution methods were used for overcoming product toxicity, growing cells in alternative conditions or improving the substrate uptake capacity of the producing strains [85]. More recently, the development of new promoters, and genetic engineering tools allowed for fine-tuning the overexpression of each enzyme in a pathway, taking into account reaction kinetics and pathway bottlenecks. As a result, more efficient and productive strains and processes are achieved [78] (**Fig. 1-3**).
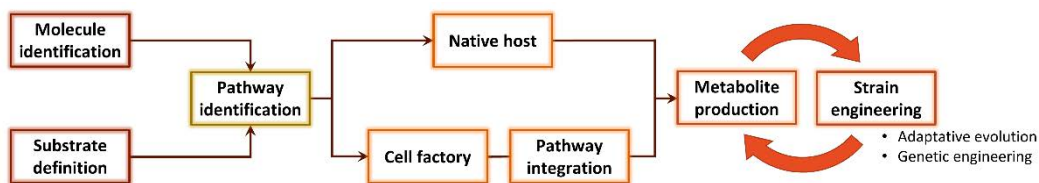
Fig. 1-3. Workflow for developing new processes for producing chemicals

Efficient growth on waste feedstocks: valorisation of glycerol

At an industrial scale, reduced cost of raw materials are as important as high production yields for cost-effective processes [86,87]. In addition, in order to optimize the metabolite biosynthesis process to obtain high yields, it is also important to select the most appropriate substrate [88]. In this context, glycerol is a by-product in the conventional biodiesel production process and therefore represents an attractive opportunity for revalorisation of an industrial waste stream, that is, for the development of a glycerol-based integrated biorefinery concept [89]. Indeed, glycerol is becoming an attractive carbon source in fermentation processes to produce highly added value compounds [90–92]. Furthermore, the reduction degree of glycerol (4.67) is different from that of glucose (4.0), and therefore higher yields of certain metabolites can be obtained from this compound [93]. Nonetheless, crude glycerol is far from being pure and contains several other compounds such as methanol [94]. Methanol is usually toxic for microbes with the exception of methylotrophic microorganisms. Thus, subsequent purification and refinement steps should be applied to the raw glycerol if it has to be used by non-methylotrophic organisms. On the other hand, methanol is also an increasingly interesting C1 compound as building block for value-added compound biosynthesis [95–97]. In this context, *P. pastoris* is able to efficiently use glycerol and/or methanol as energy and carbon sources [67,98,99]. In addition, the most extensively used promoters for heterologous gene expression in *P. pastoris* (namely, P_{GAP}, constitutive, and

P$_{AOX}$, inducible) are directly associated with glycerol and methanol metabolism [7,19]. Therefore, *P. pastoris* appears as an organism of high potential for the development of the glycerol biorefinery concept.

Glycerol uptake through specific transporters in *S. cerevisie* has a very limited efficiency (low affinity) [100], and therefore it cannot reach high growth rates. Indeed, due to the high interest in using glycerol, several efforts have recently been done for enhancing the glycerol uptake rate in *S. cerevisiae* [101,102]. Unlike *S. cerevisiae*, *P. pastoris* is able to grow in glycerol at high growth rates due to a more efficient glycerol transport [101]. Therefore, this confers a huge advantage when producing compounds easily derived from glycerol. Additionally, *P. pastoris* can tolerate and metabolize the residual methanol, can be easily genetically engineered and therefore would be an excellent platform to develop the biorefinery concept by the revalorisation of the waste glycerol generated in biodiesel production.

Nevertheless, despite *P. pastoris* can efficiently grow in glycerol there are very few studies characterising the growth of *P. pastoris* growing on glycerol as single carbon source. Therefore, the analysis of the main physiological and growth parameters will be the main subject of chapters 4 and 5 of the present thesis. In addition, the expression of a recombinant protein growing while using glycerol as single carbon source will also be tested in chapter 6.

1.2. SYSTEMS METABOLIC ENGINEERING

Metabolic engineering was proposed as new field for metabolite production in 1991 [103,104] and fully stablished in 1998 with the publication of *Metabolic Engineering: Principles and Methodologies* where it was defined as *"the directed improvement of product formation or cellular properties through the modification of specific biochemical reaction(s) or the introduction of new one(s) with the use of recombinant DNA technology"* [105]. Thus, metabolic engineering is a close discipline to genetic engineering for strain manipulation, however, the main difference lies on the conceptual design of such manipulation: while metabolic engineering aims at strain engineering to alter pathway distribution and rates, genetic engineering mainly focuses on the individual modification of genes or their derived proteins and enzymes [80]. Hence, an alternative definition that emphasises its main focus would be *"the directed modulation of metabolic pathways using methods of recombinant technology for the purpose of overproducing fuels, chemicals and pharmaceutical*

products" [80]. It is considered the seed of the systemic view of metabolic pathways and therefore preceding the application of systems biology approaches [81]. More recently, with the empowerment of computational tools and methods driven by bioinformatics and the development of high throughput techniques, systems biology emerged as a key field for integrating the multi-layer complexity of biological systems [106]. Meanwhile, genome sequencing also allowed for global systems analysis that counterweighted the classical molecular biology focused at single gene level [107]. Thereafter, advances in high-throughput techniques and bioinformatics have resulted in the development of systems-level analysis of the different biological layers also known as '–omics' (**Fig. 1-4**): genes, transcripts, proteins, metabolites, fluxes, etc [75]. The availability of these comprehensive data allowed for the improved understanding of cell physiology and the design of the 2nd generation of engineered strains for the production of proteins or chemicals.

Systems biology is also applied for systematically cataloguing all components of the different biological levels and characterising the interactions among them that result into cell function [108]. The comprehensive information and interactions of cell components can be integrated into mathematical models to discover biological mechanisms and test hypothesis on the cell physiology and behaviour. Furthermore, the application of systems biology (systems biotechnology) through large-scale integrative models, allows the prediction of genetic targets for improved strains leading into the 3rd generation of rational strain design [109]. Hence, systems metabolic engineering is the field that integrates different systems biology layers aiming at rational strategies for genetic manipulating or modulating metabolic pathways thereby enhancing the productivities of the biological process of interest [110].

1.2.1. Systems biology of *P. pastoris*

Due to the growing interest in *P. pastoris* as cell factory, several efforts have been made to provide comprehensive information on its physiology. As in other organisms, several characterisation analyses have been performed including the different systems biology layers and provided an extensive body of knowledge that enabled further strain engineering and improving biotechnological processes.

Genomics. Currently, it is available the genome sequence of three *P. pastoris* strains. *P. pastoris* has been reclassified into the genus *Komagataella*, and the most commonly used industrial strains were splited in two species: *K. phaffi* and *K. pastoris* [111]. The genome of DSMZ 7382 was the first published and corresponds to a *K. pastoris* strain [24]. Simultaneously, the genome sequence of GS115 strain (*K.* phaffi) was released [112]. Two years later, in 2011, a third strain, CBS7435 (*K. phaffi*, parental strain of GS115) was sequenced [113] and its sequence has been recently refined [114].

Transcriptomics. Several studies in *P. pastoris* provide information of gene expression in different conditions. For instance, an interesting study describe the effects of oxygen availability at different levels, including transcriptomics and compares the response to low oxygen availability with *S. cerevisiae* [69,115]. Alternatively, other studies focus on the variation in gene expression profile in different carbon sources [17,116], stress conditions, such as osmotic stress [57], or the heterologous protein expression [72,117].

Proteomics. Related to transcriptomics, proteomics also provides information on the regulation of the cell machinery in biological systems. Specifically, proteomics is used for describing the complete set of proteins expressed under certain conditions. Thus, a few number of studies have provided insight into the effect of temperature [59], the overexpression of recombinant proteins [118] and different carbon sources [116,119] on the proteome.

Metabolomics. Metabolome indicates the intracellular amounts of metabolites and valuable information on enzyme kinetics. In addition, metabolic fingerprints (endo-metabolome) and footprints (exo-metabolome) give information about the metabolic response in different conditions [120]. Moreover, metabolomic studies aid in finding out thermodynamic feasibility of reactions, bottlenecks in pathways, redox cofactor balance, the energetic state of the cells [121] and additional information that enable and constrain flux calculations [122]. Methodologies for quenching, metabolite extraction and ^{13}C-based metabolomics in *P. pastoris* have been developed [123–125] and applied in studies growing in glucose [126] and mixtures of glucose and methanol [127,128].

Lipidomics. As the name indicates, lipidomics is the study of lipid profile and could be included as a specific branch of metabolomics. It provides useful information on biomass composition and lipid metabolism. As an example, the group of Prof. Daum performed several studies on the lipid profile of different cellular organelles and membranes of cells growing in different carbon sources [129–133]. Recently, the effect of hypoxia on lipidome in an strain expressing an antibody fragment was reported [70].

Fluxomics. This omics is the closest to metabolic phenotypes (**Fig. 1-4**), and, supported by stoichiometric metabolic networks, aims to calculate metabolite turnover rates as well as flux distribution through metabolic pathways [134]. Fluxomics, unlike some of the other omics, is strictly quantitative. In addition, is considered an integrative omics due to the fact that metabolic fluxes are the result of gene expression, protein concentration, enzyme kinetics and metabolite concentrations [135]. Therefore, with the application of different approaches, fluxomics has the goal of metabolic flux analysis (MFA). Fluxomics has become widely used to study cell physiology, and thus plenty of examples can be found for *Pichia pastoris* [51,67,98,116,136]. Due to the highly informative data derived from fluxomics, a large amount of quantification and estimation techniques have been developed in order to determine the behaviour of the metabolic network. Mainly, ^{13}C-MFA and flux balance analysis (FBA) techniques are used for elucidating metabolic fluxes. Hence, in the following section an overview of these approaches is given.

The complexity of biological systems cannot be described at a single –omic level. For instance, genome sequence only represents the capabilities of an organism, but does not provide information on the specific phenotype, the response to environmental and genetic perturbations, or the regulation mechanisms [76]. Therefore, the integration of multiple –omic layers would provide more accurate snapshots of the complexity of biological systems and would allow for more reliable modelling. In addition, by integrating transcriptional and metabolic regulation information more extensive models can be generated and will be able to more precisely predict phenotypes [137]. Towards this direction, comprehensive studies have already analysed at multi-layer scale the behaviour of *P. pastoris* in different oxygen availability conditions [69], suggested a new stoichiometry for methanol uptake metabolism [116] or described the effect of genetic perturbations on metabolism and recombinant protein

production [138]. Other microorganisms have likewise been studied using multiple –omic integrative approaches: in *S. cerevisiae*, the cellular mechanisms for galactose and arabinose fermentation were identified [110] and key metabolic properties were identified in *Corynebacterium glutamicum* for the overproduction of lysine [139]. A step forward in integration of multiple layers of information, is the incorporation of multi –omic levels in genome-scale metabolic models (GSMM). Although GSMM are already –omic integrative platforms (e.g. genomics, proteomics, fluxomics), they can be further extended with additional information such as transcriptomic or proteomic data as regulatory constrains [140] resulting in improved predictions of metabolic behaviour [141]. An example of how extensive can be the multiple omics integration can be found in [142], where a whole-cell computational model of the entire life cycle of *Mycoplasma genitalium* is presented.
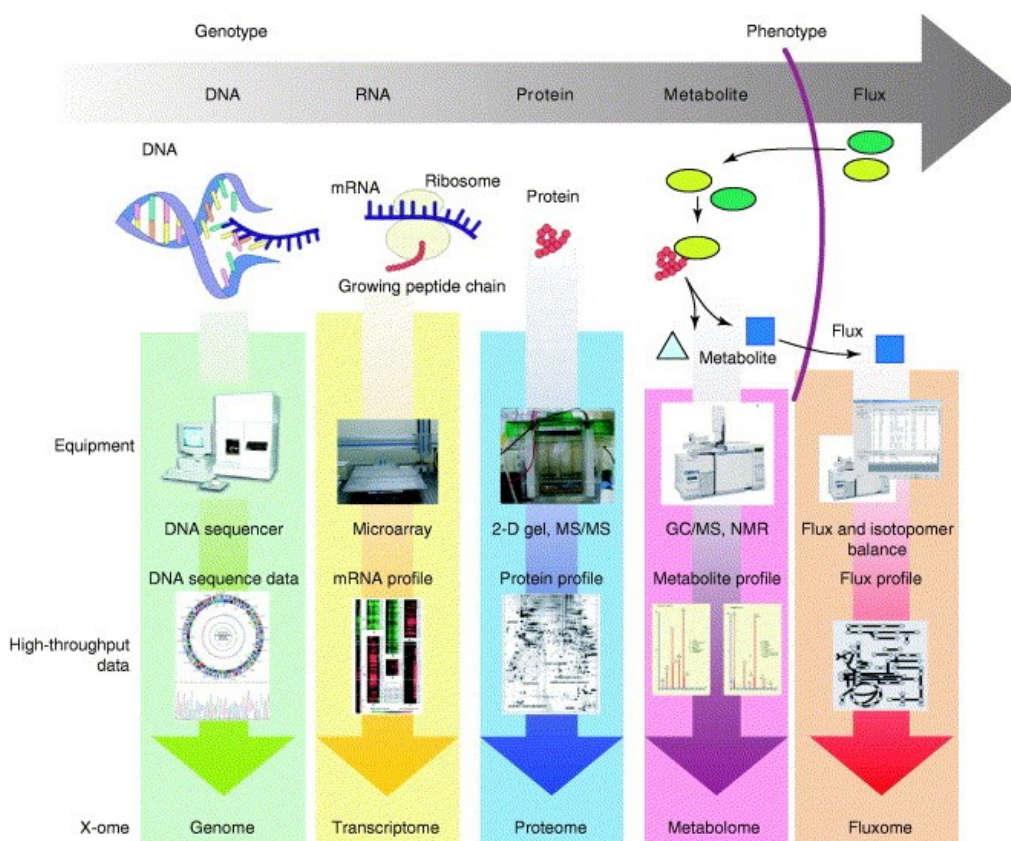
Fig. 1-4. Multiple layers in systems biology: high throughput of omics research. Figure taken from [75] (reproduced with permission of Elsevier)

The whole knowledge generated using the different –omic layers together with data integration and modelling, allows a better understanding of biological functions and phenotypes and enhances precision of the respective models. Therefore, provides solid foundations for both predicting and selecting targets for strain improvement, and together with synthetic biology, generating new and model-based 3rd generation of engineered strains.

1.2.2. ^{13}C-Metabolic flux analysis

Whereas most–omics provide direct information of each biological level (e.g. genes, transcriptional levels, protein abundance, metabolite concentration), there is no available technique for directly measuring metabolic fluxes. Therefore fluxes have to be inferred from measurable quantities [143]. The developed protocols and strategies to this purpose are generally based on stable-isotopes, typically using ^{13}C-labelled substrates (CLE, for carbon labelled experiments) and performing the experiments until reaching metabolic steady-state, at which both extracellular and intracellular fluxes are constant (i.e. chemostat cultivations). Nevertheless, due to both the length of time to reach isotopic steady state and the cost of the labelling, it is often the case that labelling experiments are discontinued before reaching steady state. Hence, correction methods for extrapolating the labelling pattern to that one in the steady state are commonly applied with good precision [144]. Alternatively, methods based on isotopically non-stationary CLE shorten labelling times and are able to resolve fluxes that are only resoluble it time-dependent information is available (e.g. fluxes related to reversible reactions or 1C substrates such as methanol) [144]. Nevertheless, the non-stationary ^{13}C-MFA requires specific infrastructures, laborious experimental setups as well as more complex computational calculations [145]. In isotopic stationary ^{13}C-MFA, two common strategies have been described for quantifying label distribution. The analysis of proteinogenic amino acids by means of NMR has been traditionally used, despite its large measurement errors, it is more informative [146]. However, nowadays the most commonly used strategy is the analysis of either the free intracellular metabolite enrichment or the proteinogenic aminoacids in by GC-MS or LC-MS/MS [143]. In this case measurement errors can be decreased down to < 1% [146]. Approaches to use both types of measurements (NMR and GC-MS) at the same time have also been performed [147].

In addition to the different options in the experimental setup and analysis, there are two major approaches for the numerical calculation of metabolic fluxes from the experimental enrichment patterns that can be classified as ^{13}C-FBA and ^{13}C-fluxomics. The first one ^{13}C-FBA, is based on the calculation of flux ratios in converging pathways [148]. The labelling patterns of proteinogenic amino acid allows for extracting the information of the labelling patterns of the respective precursors in the central carbon metabolism [149] and therefore stablishing up to 16 flux ratios when using ^{13}C glucose [150]. Subsequently, these ratios are used as constrains for the FBA and absolute net-fluxes can be calculated [143]. Metabolic flux ratio (METAFoR) analysis was the first method of this kind applied to *P. pastoris* [67,151]. Nevertheless, this method has some limitations. On the one side, the number or flux ratios available limits the degrees of freedom of the model that can be solved. On the other side, when using 1C substrates (e.g. methanol) or substrates with a reduced number of carbon bonds, the obtained information available to solve de model is very limited [148]. The other approach for flux calculation is ^{13}C-fluxomics. ^{13}C-fluxomics relies on the simultaneous iterative fitting of metabolic fluxes and carbon labelling distribution to the measured data, that is extracellular fluxes and mass isotopomer – isotope isomers – distribution (MID), minimising differences between simulated and measured data [148]. One advantage of this method is that it enables either the estimation of absolute fluxes, as well as the degree of reversibility on certain reversible reactions or the flux distribution parallel pathways [135]. A stoichiometric model of the metabolic network including the carbon atom mapping transitions for all reactions is required for the calculations in the corresponding isotopomer balancing software [152]. Nevertheless, this approach has also some limitations regarding the large number of isotopomer balancing equations needed to be solved [153] (for each metabolite the number of isotopomer balances is 2^n were n is the number of carbons). Hence, several approaches have been developed to reduce the calculation workload. One of those approaches introduces the term 'cumomer', representing the 'cumulative isotopomer'. Briefly, instead of using each of the isotopomer balance equation, cumomer balance equations are used. Then isotopomer fractions are linearly transformed into 'cumomer fractions' meaning a certain sum of isotopomer fraction of a metabolite. The solution is found by iteratively solving a sequence of linear systems formed by the cumomer balances of increasing weight (for details see [154]). This approach notably simplified computation and enabled de

resolution of additional fluxes [154]. More recently, Antoniewicz and collaborators developed a novel framework based on elementary metabolite units (EMU) that further reduces de size of the computational problem without losing information. This algorithm is based on identifying the minimum amount of labelling information required for the isotopic simulations [153].

Aside from the experimental setup, the analytical methods and flux estimation algorithms, the metabolic networks used also play a key role in flux determination. Simplified models of the central carbon metabolism have been widely employed for ¹³C-MFA. These models are based on the current biochemical knowledge of metabolic pathways at the time the calculations were done, usually before the publication of genome-scale metabolic models. In ¹³C-MFA, carbon atom transition maps are crucial for a reliable estimation of metabolic fluxes. Therefore, atom transitions are well known in central carbon metabolism [155]. However, extending the level of detail of a metabolic network including additional reactions towards a large-scale model, adds a potential source of error as an incorrectly assigned atom mapping would be critical for flux estimation. Ravikirthi and co-workers [156] reported an outstanding effort in mapping the whole atom transition genome-scale in an *E. coli* genome-scale metabolic model. More recently, this model was used for ¹³C-MFA at genome scale [157] and authors reported similar flux distributions when comparing genome-scale and simplified models. Nevertheless, the major limitation in using ¹³C-MFA at genome-scale model is the increase in computation time associated with the higher number of variables. In addition, the robustness of the metabolic network with multiple parallel and alternative pathways dramatically impairs the procedure of flux identification. Thus, several authors suggest to simplify the genome-scale metabolic model by reducing its size to achieve core-scale models [157]. Until computational processing power is not increased, allowing a significant reduction in metabolic flux calculation time and improved analytical techniques are developed enabling to determine the complete labelling information available in all involved metabolites, ¹³C-MFA at genome-scale level is still not a real option [158].

Due to the high interest in *P. pastoris,* cultivation and analysis methods [124,148] and an extensive body of knowledge, incuding fluxomics, is available describing the behavior of its metabolic network using different carbon sources and growth conditions. An extensive

review of ^{13}C-MFA in *P. pastoris* with examples of alternative labelling strategies, analysis methods and calculation approaches can be found in [159] and references therein.
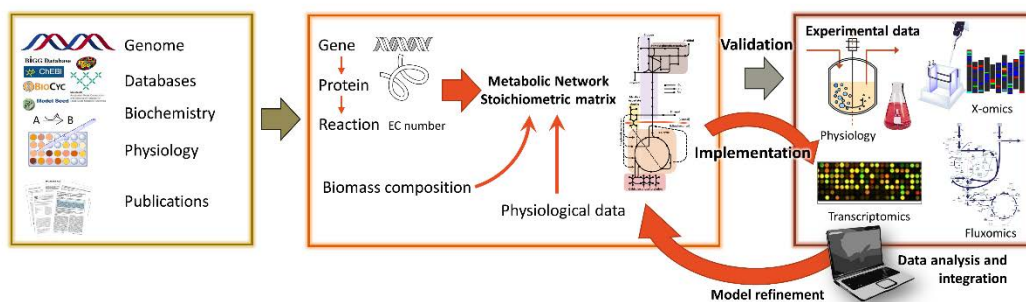
1.2.3. Genome-scale metabolic models

Genome-Scale metabolic models (GSMM) have become one of the most useful and widely employed tools in systems biology in the last fifteen years [160]. GSMM are extensive descriptions of all known metabolic reactions available in an organism inferred from the genome sequence [161]. This information is converted mathematically into a stoichiometric matrix correlating reactions and metabolites. The stoichiometry of the matrix, together with additional constrains such as lower or upper boundaries for each reaction flux, allow for performing flux balance analysis (FBA) and thus linking genotype to phenotype by calculating the (metabolic) behaviour under certain environmental conditions [75,162,163]. From a systems biology perspective, GSMM are a multi –omics tool. For their derivation, genome information is processed to find out transcribed genes encoding for metabolic proteins. Further, these proteins are associated to the reaction with the corresponding metabolites. By means of FBA or other mathematical algorithms, GSMM finally provide information on metabolic fluxes [164]. Hence, genome information is connected to transcriptome and proteome, leading to the description of a metabolic state of the cell with the corresponding metabolome and fluxome (**Fig. 1-4***).

The initial core metabolic networks were based on the available knowledge of biochemical pathways [165–167]. Later, with the improved ability to sequence and annotate genomes, larger models (genome-scale) were developed relying on the identification of substrates and products for each enzyme encoded and annotated in the genome [168]. Advances in bioinformatics and biochemical databases enabled the automatization of metabolic network generation procedures [169]. Consequently, metabolic network reconstructions are attainable for any organism with an annotated genome [161]. Since the first genome-based metabolic model was presented [170], a huge number of models have been developed for a broad variety of species, from archaea and bacteria, to higher eukaryotes [171]. Despite the biological diversity, the same steps are required for GSMM development [172]. Briefly, the first step consists in extracting and processing the information of the genome to obtain the gene-protein-reaction relationship (GPR) and assemble the network into a draft

reconstruction. GPR associations can be performed automatically by computational algorithms [169], nevertheless the resulting associations should afterwards be revised and manually curated. Subsequently, the metabolic network is complemented with bibliographic information and experimental data such as biomass composition and physiological constrains. Finally, the GSMM has to be calibrated with experimental data and subsequently validated with new datasets (**Fig. 1-5**). Validation of a GSMM for different conditions requires the availability of extensive cultivation data information describing its physiology. In addition, a wide range of information on biomass composition enables building specific biomass equations to accurately describe cell growth in alternative conditions, such as different substrates [173]. Although several software platforms facilitate the network reconstruction process by automating some of the abovementioned steps, it is usually necessary and advisable to manually review and curate the resulting model [174].

Although GSMMs have been widely used for predicting metabolic engineering targets to improve growth and production of chemicals or recombinant proteins [175–178], they have many other applications both for descriptive and predictive purposes. Nevertheless, the guidance for metabolic engineering is the most commonly used application. Since GSMMs allow to predict the phenotype of a microorganism in a range of conditions, including those derived from genetic modification [160,177], it makes GSMMs a powerful tool for the rational design of metabolic engineering strategies to enhance productivities or implementing new pathways [87,179]. Different algorithms have been developed for simulating gene knockout or under/overexpression in order to identify metabolic engineering targets [180]. Most of them showed successful results in the resulting improved strains for metabolite or recombinant protein production [138].

Fig. 1-5. Main genome-scale metabolic model development workflow.

GSMMs are also comprehensive databases of genes, proteins, metabolites and biochemical pathways that give information of the whole capabilities of the organism. Therefore, GSMMs can be used to assist in the interpretation of the experimental data. In addition, are suitable platforms to contextualise high-throughput data. Instead of analysing the variation in the expression of a single or a couple of genes, the increase in a certain flux, or the production of a metabolite, GSMMs enable to integrate high-throughput data and therefore, analyse a big pool of genes up/downregulated, correlations between fluxes or activation of different metabolic states that leads to the production of certain metabolites. This contextualisation enables a better biological interpretation of experimental data interconnecting pieces of information, even at different levels, to form a global relate [181]. Furthermore, it facilitates the discovery of network properties such as pathway redundancy, metabolites connectivity, the existence of loops, or novel alternative pathways and capabilities that remained 'hidden' in the genome [177]. Moreover, in silico methods allowed for the integration of multi –omics, including transcriptional regulation, and refined prediction accuracies [182]. The incorporation of multiple layers of information requires at the same time the development of metabolic modelling techniques able to integrate this multi-level data leading to GSMMs with a wider scope, depth and increased accuracy in precise and more realistic predictions of cellular metabolism [76,168]. Finally, GSMMs present structured and systematised information of metabolic pathways, therefore are extremely useful for comparing pathways and capabilities between organisms. When contrasting metabolic networks from different species, the conserved or missing pathways between species may allow to elucidate evolutionary relationship between organisms [178].

During model implementation, GSMM predictions are constantly evaluated. Thus, novel experimental data, including multi –omics information, can be used for refining models and to improve predictability [121]. Physiological data can be used for extending or constraining accordingly the capabilities of the model. In addition, novel literature discovering moonlight proteins [183], identifying gene functions, characterising enzyme activities and organism-specific pathways can be used for further extend a GSMM [171]. Hence, GSMMs can be expanded in number of genes, reactions and metabolites resulting in an increase in network complexity. Moreover, another important measure for model refinement is the curation of large-scale derived bugs mainly associated to the automatic missannotation of

genes [184], wrong metabolite assignments [185] or protein functions [186]. In fact, a recent example showed how after the revision of redox cofactor specificity of the enzymes and thus in the reactions, and manually curating this cofactor assignment, GSMM can improve the simulated flux distribution [187]. In order to further extend models, biomass composition would be an additional target. It has been shown that the alteration of biomass composition has a deep impact on predictions, thus adopting condition-specific compositions would allow for more realistic flux distribution determination [173]. In addition, as explained before, the integration using multi –omics information, as well as the application of thermodynamic constrains to the network expands the model and enables improved reliability [188].

The main goal for GSMM developers is to achieve models able to very closely predict the in vivo flux distribution. Therefore, GSMMs are in continuous evolution, by the curation of erroneously included information, the incorporation of novel capabilities to make them more descriptive or providing improved predictions. In *S. cerevisiae*, new versions and updated models integrate previous versions and incorporate new features and information from novel literature. In fact, since the first GSMM published for *S. cerevisiae* [163] several alternative models were published and were eventually joined into a consensus metabolic model [189]. This consensus model was further revised, upgraded and expanded up to the 7th version [190–193]. Similarly, in other organisms, available models were integrated and successive releases continuously upgrade preexisting models [194,195].

In *P. pastoris*, the first metabolic models were limited to the central carbon metabolism and were mainly used for ^{13}C-MFA [67,69,136,151]. In 2009, the genome of two different strains, DSMZ 70382 [24] and GS115 [112] was sequenced. More recently, a third strain (CBS7435) was also sequenced and annotated [113] showing some discrepancies with the previously reported sequences. Once the genomic data was available, two GSMMs were published simultaneously: iPP668 [196] corresponding to GS115 strain, and PpaMBEL1254 [197] based on the DSMZ 70382 genome. Two years later, a third model (iLC915) became available, also based on the GS115 genome [198]. Moreover, other GSMMs derived by automatic application of reconstruction algorithms are also available [169,199]. *P. pastoris'* GSMMs have been reported as potentially useful platforms for bioprocess design and optimization, as well as for strain metabolic engineering [121,200]. In fact, Nocon and co-

workers [138] have already used PpaMBEL1254 to predict the effect of gene overexpression and deletion in mutants to enhance recombinant protein production.

1.3. REFERENCES

1   Porro, D. *et al.* (2011) Production of recombinant proteins and metabolites in yeasts: when are these systems better than bacterial production systems? *Appl. Microbiol. Biotechnol.* 89, 939–48

2   Liu, L. *et al.* (2013) Frontiers of yeast metabolic engineering: diversifying beyond ethanol and *Saccharomyces*. *Curr. Opin. Biotechnol.* 24, 1023–1030

3   Corchero, J.L. *et al.* (2013) Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics. *Biotechnol. Adv.* 31, 140–153

4   Van Dijken, J.P. *et al.* (1986) Alcoholic fermentation by "non-fermentative" yeasts. *Yeast* 2, 123–127

5   Mattanovich, D. *et al.* (2014) Yeast biotechnology: teaching the old dog new tricks. *Microb. Cell Fact.* 13, 34

6   Cregg, J.M. *et al.* (1985) *Pichia pastoris* as a Host System for Transformations. *Mol. Cell. Biol.* 5, 3376–3385

7   Gasser, B. *et al.* (2013) *Pichia pastoris*: protein production host and model organism for biomedical research. *Future Microbiol.* 8, 191–208

8   Martínez, J.L. *et al.* (2012) Pharmaceutical protein production by yeast: Towards production of human blood proteins by microbial fermentation. *Curr. Opin. Biotechnol.* 23, 965–971

9   Walsh, G. (2014) Biopharmaceutical benchmarks 2014. *Nat. Biotechnol.* 32, 992–1000

10 Yamada, Y. *et al.* (1995) The Phylogenetic Relationships of Methanol-assimilating Yeasts Based on the Partial Sequences of 18S and 26S Ribosomal RNAs: The Proposal of *Komagataella* Gen. Nov. (Saccharomycetaceae). *Biosci. Biotechnol. Biochem.* 59, 439–444

11 Kurtzman, C.P. (2009) Biotechnological strains of *Komagataella* (*Pichia*) *pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis. *J. Ind. Microbiol. Biotechnol.* 36, 1435–1438

12 Bill, R.M. (2014) Playing catch-up with *Escherichia coli*: using yeast to increase success rates in recombinant protein production experiments. *Front. Microbiol.* 5, 1–5

13 Vogl, T. *et al.* (2013) New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*. *Curr. Opin. Biotechnol.* 24, 1094–1101

14 Näätsaari, L. *et al.* (2012) Deletion of the *Pichia pastoris KU70* Homologue Facilitates Platform Strain Generation for Gene Expression and Synthetic Biology. *PLoS One* 7, e39720

15 Weninger, A. *et al.* (2015) Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. *J. Biotechnol.* 235, 139–149

16 Hamilton, S.R. *et al.* (2013) Production of sialylated *O*-linked glycans in *Pichia pastoris*. *Glycobiology* 23, 1192–203

17 Prielhofer, R. *et al.* (2013) Induction without methanol: novel regulated promoters enable high-level expression in *Pichia pastoris*. *Microb. Cell Fact.* 12, 5

18 Weinhandl, K. *et al.* (2014) Carbon source dependent promoters in yeasts. *Microb. Cell Fact.* 13, 5

19 Cos, O. *et al.* (2006) Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: a review. *Microb. Cell Fact.* 5, 17

20 Potvin, G. *et al.* (2012) Bioprocess engineering aspects of heterologous protein production in *Pichia pastoris*: A review. *Biochem. Eng. J.* 64, 91–105

21 Heyland, J. *et al.* (2010) Simple enzymatic procedure for L-carnosine synthesis: whole-cell biocatalysis and efficient biocatalyst recycling. *Microb. Biotechnol.* 3, 74–83

22 Cheng, H. *et al.* (2014) Genetically engineered *Pichia pastoris* yeast for conversion of glucose to xylitol by a single-fermentation process. *Appl. Microbiol. Biotechnol.* 98, 3539–3552

23 Geier, M. *et al.* (2015) Compact multi-enzyme pathways in *P. pastoris*. *Chem. Commun.* 51, 1643–1646

24 Mattanovich, D. *et al.* (2009) Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*. *Microb. Cell Fact.* 8, 29

25 Grinna, L.S. and Tschopp, J.F. (1989) Size distribution and general structural features of *N*-linked oligosaccharides from the methylotrophic yeast, *Pichia pastoris*. *Yeast* 5, 107–15

26 Scorer, C.A. *et al.* (1993) The intracellular production and secretion of HIV-1 envelope protein in the methylotrophic yeast *Pichia pastoris*. *Gene* 136, 111–119

27 Montesino, R. *et al.* (1998) Variation in N-Linked Oligosaccharide Structures on Heterologous Proteins Secreted by the Methylotrophic Yeast *Pichia pastoris*. *Protein Expr. Purif.* 14, 197–207

28 Bretthauer, R.K. and Castellino, F.J. (1999) Glycosylation of *Pichia pastoris*-derived proteins. *Biotechnol. Appl. Biochem.* 30, 193–200

29 Jacobs, P.P. *et al.* (2009) Engineering complex-type N-glycosylation in *Pichia pastoris* using GlycoSwitch technology. *Nat. Protoc.* 4, 58–70

30 Meehl, M.A. and Stadheim, T.A. (2014) Biopharmaceutical discovery and production in yeast. *Curr. Opin. Biotechnol.* 30, 120–127

31 Bretthauer, R.K. (2003) Genetic engineering of *Pichia pastoris* to humanize *N*-glycosylation of proteins. *Trends Biotechnol.* 21, 459–462

32 Spadiut, O. *et al.* (2014) Microbials for the production of monoclonal antibodies and antibody fragments. *Trends Biotechnol.* 32, 54–60

33 Delic, M. *et al.* (2013) The secretory pathway: exploring yeast diversity. *FEMS Microbiol. Rev.* 37, 872–914

34 Garcia-Ortega, X. *et al.* (2016) A step forward to improve recombinant protein production in *Pichia pastoris*: From specific growth rate effect on protein secretion to carbon-starving conditions as advanced strategy. *Process Biochem.* 51, 681–691

35 Delic, M. *et al.* (2014) Engineering of protein folding and secretion-strategies to overcome bottlenecks for efficient production of recombinant proteins. *Antioxid. Redox Signal.* 21, 414–37

36 Delic, M. *et al.* (2012) Oxidative protein folding and unfolded protein response elicit differing redox regulation in endoplasmic reticulum and cytosol of yeast. *Free Radic. Biol. Med.* 52, 2000–12

37 Tyo, K.E.J. *et al.* (2012) Imbalance of heterologous protein folding and disulfide bond formation rates yields runaway oxidative stress. *BMC Biol.* 10, 16

38 Malhotra, J.D. and Kaufman, R.J. (2007) Endoplasmic Reticulum Stress and Oxidative Stress: A Vicious Cycle or a Double-Edged Sword? *Antioxid. Redox Signal.* 9, 2277–2294

39 Hou, J. *et al.* (2012) Metabolic engineering of recombinant protein secretion by *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 12, 491–510

40 Marsalek, L. *et al.* (2017) Disruption of genes involved in CORVET complex leads to enhanced secretion of heterologous carboxylesterase only in protease deficient *Pichia pastoris*. *Biotechnol. J.* 12, 1600584

41 Graf, A. *et al.* (2008) Novel insights into the unfolded protein response using *Pichia pastoris* specific DNA microarrays. *BMC Genomics* 9, 390

42 Gasser, B. *et al.* (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microb. Cell Fact.* 7, 11

43 Puxbaum, V. *et al.* (2015) Quo vadis? The challenges of recombinant protein folding and secretion in *Pichia pastoris*. *Appl.*

Microbiol. Biotechnol. 99, 2925–2938

44 Glick, B.R. (1995) Metabolic load and heterologous gene expression. *Biotechnol. Adv.* 13, 247–261

45 Wu, G. *et al.* (2016) Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol.* 34, 652–664

46 Ramón, R. *et al.* (2007) Sorbitol co-feeding reduces metabolic burden caused by the overexpression of a *Rhizopus oryzae* lipase in *Pichia pastoris. J. Biotechnol.* 130, 39–46

47 Klein, T. *et al.* (2014) Overcoming the metabolic burden of protein secretion in *Schizosaccharomyces pombe* - A quantitative approach using 13C-based metabolic flux analysis. *Metab. Eng.* 21, 34–45

48 Carnicer, M. *et al.* (2009) Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels. *Microb. Cell Fact.* 8, 65

49 Jordà, J. *et al.* (2012) Metabolic flux profiling of recombinant protein secreting *Pichia pastoris* growing on glucose:methanol mixtures. *Microb. Cell Fact.* 11, 57

50 Driouch, H. *et al.* (2012) Integration of in vivo and in silico metabolic fluxes for improvement of recombinant protein production. *Metab. Eng.* 14, 47–58

51 Celik, E. *et al.* (2010) Metabolic flux analysis for recombinant protein production by *Pichia pastoris* using dual carbon sources: Effects of methanol feeding rate. *Biotechnol. Bioeng.* 105, 317–29

52 Klein, T. (2015) Engineering the supply chain for protein production/secretion in yeasts and mammalian cells. *J. Ind. Microbiol. Biotechnol.* 42, 453–464

53 Heyland, J. *et al.* (2011) Carbon metabolism limits recombinant protein production in *Pichia pastoris. Biotechnol. Bioeng.* 108, 1942–53

54 Barrigón, J.M. *et al.* (2013) Searching the best operational strategies for *Rhizopus oryzae* lipase production in *Pichia pastoris* Mut+ phenotype: Methanol limited or methanol non-limited fed-batch cultures? *Biochem. Eng. J.* 75, 47–54

55 Surribas, A. *et al.* (2007) Production of a *Rhizopus oryzae* lipase from *Pichia pastoris* using alternative operational strategies. *J. Biotechnol.* 130, 291–9

56 García-Ortega, X. *et al.* (2013) Fed-batch operational strategies for recombinant Fab production with *Pichia pastoris* using the constitutive GAP promoter. *Biochem. Eng. J.* 79, 172–181

57 Dragosits, M. *et al.* (2010) The response to unfolded protein is involved in osmotolerance of *Pichia pastoris. BMC Genomics* 11, 207

58 Dragosits, M. *et al.* (2011) Influence of growth temperature on the production of antibody Fab fragments in different microbes: A host comparative analysis. *Biotechnol. Prog.* 27, 38–46

59 Dragosits, M. *et al.* (2009) The effect of temperature on the proteome of recombinant *Pichia pastoris. J. Proteome Res.* 8, 1380–1392

60 Liu, Z. *et al.* (2013) Anaerobic α-amylase production and secretion with fumarate as the final electron acceptor in *Saccharomyces cerevisiae. Appl. Environ. Microbiol.* 79, 2962–2967

61 Fredlund, E. *et al.* (2004) Oxygen- and Glucose-Dependent Regulation of Central Carbon Metabolism in *Pichia anomala. Appl. Environ. Microbiol.* 70, 59'5-5911

62 Baumann, K. *et al.* (2008) Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins. *Biotechnol. Bioeng.* 100, 177–183

63 Siso, M.I.G. *et al.* (2012) The yeast hypoxic responses, resources for new biotechnological opportunities. *Biotechnol. Lett.* 34, 2161–2173

64 Babel, W. and Muller, R.H. (1985) Mixed Substrate Utilization in Micro-organisms: Biochemical Aspects and Energetics. *Microbiology* 131, 39–45

65 Ferrer, P. and Valero, F. (2016) Coping with Physiological Stress During Recombinant Protein Production by Bioreactor Design and Operation. In *Bioreactors: Design, Operation and Novel Applications* (Mandenius, C.-F., ed), pp. 227–260, Wiley-VCH Verlag GmbH & Co. KGaA

66 Klein, T. *et al.* (2013) Metabolic fluxes in *Schizosaccharomyces pombe* grown on glucose and mixtures of glycerol and acetate. *Appl.*

Microbiol. Biotechnol. 97, 5013–5026

67 Solà, A. *et al.* (2007) Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. *Microbiology* 153, 281–90

68 Krainer, F.W. *et al.* (2015) Optimizing cofactor availability for the production of recombinant heme peroxidase in *Pichia pastoris. Microb. Cell Fact.* 14, 4

69 Baumann, K. *et al.* (2010) A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions. *BMC Syst. Biol.* 4, 141

70 Adelantado, N. *et al.* (2017) The effect of hypoxia on the lipidome of recombinant *Pichia pastoris. Microb. Cell Fact.* 16, 86

71 Baumann, K. *et al.* (2011) Protein trafficking, ergosterol biosynthesis and membrane physics impact recombinant protein secretion in *Pichia pastoris. Microb. Cell Fact.* 10, 93

72 Cámara, E. *et al.* (2017) Increased dosage of *AOX1* promoter-regulated expression cassettes leads to transcription attenuation of the methanol metabolism in *Pichia pastoris. Sci. Rep.* 7, 44302

73 Guerfal, M. *et al.* (2010) 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.* 9, 49

74 Gasser, B. *et al.* (2006) Engineering of *Pichia pastoris* for improved production of antibody fragments. *Biotechnol. Bioeng.* 94, 353–361

75 Lee, S.Y. *et al.* (2005) Systems biotechnology for strain improvement. *Trends Biotechnol.* 23, 349–58

76 Fondi, M. and Liò, P. (2015) Multi -omics and metabolic modelling pipelines: Challenges and tools for systems microbiology. *Microbiol. Res.* 171, 52–64

77 Jullesson, D. *et al.* (2015) Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnol. Adv.* 33, 1395–1402

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

79 Nielsen, J. (2015) Yeast cell factories on the horizon. *Science.* 349, 1050–1051

80 Stephanopoulos, G. (2012) Synthetic Biology and Metabolic Engineering. *ACS Synth. Biol.* 1, 514–525

81 Woolston, B.M. *et al.* (2013) Metabolic Engineering: Past and Future. *Annu. Rev. Chem. Biomol. Eng.* 4, 259–288

82 Avalos, J.L. *et al.* (2013) Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nat. Biotechnol.* 31, 335–341

83 Becker, J. and Wittmann, C. (2015) Advanced biotechnology: Metabolically engineered cells for the bio-based production of chemicals and fuels, materials, and health-care products. *Angew. Chemie - Int. Ed.* 54, 3328–3350

84 Wendisch, V.F. *et al.* (2016) The flexible feedstock concept in Industrial Biotechnology: Metabolic engineering of *Escherichia coli, Corynebacterium glutamicum, Pseudomonas, Bacillus* and yeast strains for access to alternative carbon sources. *J. Biotechnol.* 234, 139–157

85 Dragosits, M. and Mattanovich, D. (2013) Adaptive laboratory evolution – principles and applications for biotechnology. *Microb. Cell Fact.* 12, 64

86 Kroll, J. *et al.* (2010) Plasmid addiction systems: perspectives and applications in biotechnology. *Microb. Biotechnol.* 3, 634–657

87 Gustavsson, M. and Lee, S.Y. (2016) Prospects of microbial cell factories developed through systems metabolic engineering. *Microb. Biotechnol.* 9, 610–617

88 Goldman, S. (2010) Genetic chemistry: production of non-native compounds in yeast. *Curr. Opin. Chem. Biol.* 14, 390–395

89 Kiss, A.A. *et al.* (2015) A systems engineering perspective on process integration in industrial biotechnology. *J. Chem. Technol. Biotechnol.* 90, 349–355

90 Valerio, O. *et al.* (2015) Improved utilization of crude glycerol from biodiesel industries: Synthesis and characterization of sustainable biobased polyesters. *Ind. Crops Prod.* 78, 141–147

91 Johnson, D.T. and Taconi, K.A. (2007) The glycerin glut: Options for the value-added conversion of crude glycerol resulting from biodiesel production. *Environ. Prog.* 26, 338–348

92 Yang, F. *et al.* (2012) Value-added uses for crude glycerol--a byproduct of biodiesel production. *Biotechnol. Biofuels* 5, 13

93   da Silva, G.P. *et al.* (2009) Glycerol: A promising and abundant carbon source for industrial microbiology. *Biotechnol. Adv.* 27, 30–39

94   Posada, J. a *et al.* (2012) Design and analysis of biorefineries based on raw glycerol: addressing the glycerol problem. *Bioresour. Technol.* 111, 282–93

95   Schrader, J. *et al.* (2009) Methanol-based industrial biotechnology: current status and future perspectives of methylotrophic bacteria. *Trends Biotechnol.* 27, 107–115

96   Khosravi-Darani, K. *et al.* (2013) Microbial production of poly(hydroxybutyrate) from C1 carbon sources. *Appl. Microbiol. Biotechnol.* 97, 1407–1424

97   Nguyen, A.D. *et al.* (2016) Reconstruction of methanol and formate metabolic pathway in non-native host for biosynthesis of chemicals and biofuels. *Biotechnol. Bioprocess Eng.* 21, 477–482

98   Jordà, J. *et al.* (2014) Metabolic flux analysis of recombinant *Pichia pastoris* growing on different glycerol/methanol mixtures by iterative fitting of NMR-derived 13C-labelling data from proteinogenic amino acids. *N. Biotechnol.* 31, 120–132

99   Çelik, E. *et al.* (2008) Use of Biodiesel Byproduct Crude Glycerol as the Carbon Source for Fermentation Processes by Recombinant *Pichia pastoris*. *Ind. Eng. Chem. Res.* 47, 2985–2990

100   Ho, P.-W. *et al.* (2017) The sole introduction of two single-point mutations establishes glycerol utilization in *Saccharomyces cerevisiae* CEN.PK derivatives. *Biotechnol. Biofuels* 10, 10

101   Klein, M. *et al.* (2017) Glycerol metabolism and transport in yeast and fungi: established knowledge and ambiguities. *Environ. Microbiol.* 19, 878–893

102   Klein, M. *et al.* (2016) Towards the exploitation of glycerol's high reducing power in *Saccharomyces cerevisiae* -based bioprocesses. *Metab. Eng.* 38, 464–472

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

104   Stephanopoulos, G. and Vallino, J. (1991) Network rigidity and metabolic engineering in metabolite overproduction. *Science* 252, 252, 1675–1681

105   Stephanopoulos, G.N. *et al.* (1998) The Essence of Metabolic Engineering. In *Metabolic Engineering* pp. 1–20, Elsevier

106   Kitano, H. (2002) Systems Biology: A Brief Overview. *Science.* 295, 1662–1664

107   Westerhoff, H. V and Palsson, B.O. (2004) The evolution of molecular biology into systems biology. *Nat. Biotechnol.* 22, 1249–1252

108   Barabási, A.-L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113

109   Bruggeman, F.J. and Westerhoff, H. V (2007) The nature of systems biology. *Trends Microbiol.* 15, 45–50

110   Yoshikawa, K. *et al.* (2012) Design of Superior Cell Factories Based on Systems Wide Omics Analysis. In *Systems Metabolic Engineering* pp. 57–81, Springer Netherlands

111   Kurtzman, C.P. (2005) Description of *Komagataella phaffii* sp. nov. and the transfer of *Pichia pseudopastoris* to the methylotrophic yeast genus *Komagataella*. *Int. J. Syst. Evol. Microbiol.* 55, 973–976

112   De Schutter, K. *et al.* (2009) Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nat. Biotechnol.* 27, 561–6

113   Küberl, A. *et al.* (2011) High-quality genome sequence of *Pichia pastoris* CBS7435. *J. Biotechnol.* 154, 312–20

114   Sturmberger, L. *et al.* (2016) Refined *Pichia pastoris* reference genome sequence. *J. Biotechnol.* 235, 121–131

115   Baumann, K. *et al.* (2011) The impact of oxygen on the transcriptome of recombinant *S. cerevisiae* and *P. pastoris* - a comparative analysis. *BMC Genomics* 12, 218

116   Rußmayer, H. *et al.* (2015) Systems-level organization of yeast methylotrophic lifestyle. *BMC Biol.* 13, 80

117   Zhu, T. *et al.* (2011) Understanding the effect of foreign gene dosage on the physiology of *Pichia pastoris* by transcriptional analysis of key genes. *Appl. Microbiol. Biotechnol.* 89, 1127–35

118   Vanz, A.L. *et al.* (2012) Physiological response of *Pichia pastoris* GS115 to methanol-induced high level production of the Hepatitis B surface antigen: catabolic adaptation, stress responses, and autophagic processes. *Microb. Cell Fact.* 11, 103

119   Huang, C.-J. *et al.* (2011) A proteomic analysis of the *Pichia pastoris* secretome in methanol-induced cultures. *Appl. Microbiol. Biotechnol.* 90, 235–47

120   Noack, S. and Wiechert, W. (2014) Quantitative metabolomics: a phantom? *Trends Biotechnol.* 32, 238–244

121   Caspeta, L. and Nielsen, J. (2013) Toward systems metabolic engineering of *Aspergillus* and *Pichia* species for the production of chemicals and biofuels. *Biotechnol. J.* 8, 534–44

122   Toya, Y. and Shimizu, H. (2013) Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. *Biotechnol. Adv.* DOI: 10.1016/j.biotechadv.2013.05.002

123   Russmayer, H. *et al.* (2015) Metabolomics sampling of *Pichia pastoris* revisited: rapid filtration prevents metabolite loss during quenching. *FEMS Yeast Res.* 15, fov049

124   Carnicer, M. *et al.* (2012) Development of quantitative metabolomics for *Pichia pastoris*. *Metabolomics* 8, 284–298

125   Mairinger, T. *et al.* (2015) Gas Chromatography-Quadrupole Time-of-Flight Mass Spectrometry-Based Determination of Isotopologue and Tandem Mass Isotopomer Fractions of Primary Metabolites for 13C-Metabolic Flux Analysis. *Anal. Chem.* 87, 11792–11802

126   Carnicer, M. *et al.* (2012) Quantitative metabolomics analysis of amino acid metabolism in recombinant *Pichia pastoris* under different oxygen availability conditions. *Microb. Cell Fact.* 11, 83

127   Jordà, J. *et al.* (2013) Glucose-methanol co-utilization in *Pichia pastoris* studied by metabolomics and instationary 13C flux analysis. *BMC Syst. Biol.* 7, 17

128   Jordà, J. *et al.* (2014) Quantitative Metabolomics and Instationary 13C-Metabolic Flux Analysis Reveals Impact of Recombinant Protein Production on Trehalose and Energy Metabolism in *Pichia pastoris*. *Metabolites* 4, 281–299

129   Klug, L. *et al.* (2014) The lipidome and proteome of microsomes from the methylotrophic yeast *Pichia pastoris*. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* 1841, 215–226

130   Grillitsch, K. *et al.* (2014) Isolation and characterization of the plasma membrane from the yeast *Pichia pastoris*. *Biochim. Biophys. Acta - Biomembr.* 1838, 1889–1897

131   Ivashov, V. a *et al.* (2013) Lipidome and proteome of lipid droplets from the methylotrophic yeast *Pichia pastoris*. *Biochim. Biophys. Acta* 1831, 282–90

132   Wriessnegger, T. *et al.* (2009) Lipid analysis of mitochondrial membranes from the yeast *Pichia pastoris*. *Biochim. Biophys. Acta* 1791, 166–72

133   Wriessnegger, T. *et al.* (2007) Lipid composition of peroxisomes from the yeast *Pichia pastoris* grown on different carbon sources. *Biochim. Biophys. Acta* 1771, 455–61

134   Roldão, A. *et al.* (2012) Bridging Omics Technologies with Synthetic Biology in Yeast Industrial Biotechnology. In *Systems Metabolic Engineering* pp. 271–327, Springer Netherlands

135   Winter, G. and Krömer, J.O. (2013) Fluxomics - connecting 'omics analysis and phenotypes. *Environ. Microbiol.* 15, 1901–1916

136   Heyland, J. *et al.* (2010) Quantitative physiology of *Pichia pastoris* during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. *Biotechnol. Bioeng.* 107, 357–68

137   Ritchie, M.D. *et al.* (2015) Methods of integrating data to uncover genotype–phenotype interactions. *Nat. Rev. Genet.* 16, 85–97

138   Nocon, J. *et al.* (2014) Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. *Metab. Eng.* 24, 129–138

139   Kromer, J.O. *et al.* (2004) In-Depth Profiling of Lysine-Producing *Corynebacterium glutamicum* by Combined Analysis of the Transcriptome, Metabolome, and Fluxome. *J. Bacteriol.* 186, 1769–1784

140   Saha, R. *et al.* (2014) Recent advances in the reconstruction of metabolic models and integration of omics data. *Curr. Opin. Biotechnol.* 29, 39–45

141   Åkesson, M. *et al.* (2004) Integration of gene expression data into genome-scale metabolic models. *Metab. Eng.* 6, 285–293

142   Karr, J.R. *et al.* (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* 150, 389–401

143   Zamboni, N. *et al.* (2009) 13C-based metabolic flux analysis. *Nat. Protoc.* 4, 878–892

144   Nöh, K. and Wiechert, W. (2006) Experimental design principles

for isotopically instationary ¹³C labeling experiments. *Biotechnol. Bioeng.* 94, 234–251

145 Nöh, K. and Wiechert, W. (2011) The benefits of being transient: Isotope-based metabolic flux analysis at the short time scale. *Appl. Microbiol. Biotechnol.* 91, 1247–1265

146 Quek, L.-E. and Nielsen, L.K. (2014) Customization of ¹³C-MFA Strategy According to Cell Culture System. (Krömer, J. O. et al., eds), pp. 81–90, Springer New York

147 Kleijn, R.J. *et al.* (2007) Metabolic flux analysis of a glycerol-overproducing *Saccharomyces cerevisiae* strain based on GC-MS, LC-MS and NMR-derived ¹³C-labelling data. 7, 216–231

148 Ferrer, P. and Albiol, J. (2014) ¹³C-Based Metabolic Flux Analysis in Yeast: The *Pichia pastoris* Case. In *Yeast Metabolic Engineering* 1152 (Mapelli, V., ed), pp. 209–232, Springer New York

149 Szyperski, T. (1995) Biosynthetically directed fractional 13C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism. *Eur. J. Biochem.* 232, 433–448

150 Niedenführ, S. *et al.* (2015) How to measure metabolic fluxes: a taxonomic guide for ¹³C fluxomics. *Curr. Opin. Biotechnol.* 34, 82–90

151 Solà, A. *et al.* (2004) Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. *Eur. J. Biochem.* 271, 2462–70

152 Nöh, K. *et al.* (2015) Visual workflows for ¹³C-metabolic flux analysis. *Bioinformatics* 31, 346–354

153 Antoniewicz, M.R. *et al.* (2007) Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions. *Metab. Eng.* 9, 68–86

154 Wiechert, W. *et al.* (1999) Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnol. Bioeng.* 66, 69–85

155 Crown, S.B. and Antoniewicz, M.R. (2013) Publishing ¹³C metabolic flux analysis studies: A review and future perspectives. *Metab. Eng.* 20, 42–48

156 Ravikirthi, P. *et al.* (2011) Construction of an *E. coli* genome-scale atom mapping model for MFA calculations. *Biotechnol. Bioeng.* 108, 1372–1382

157 Gopalakrishnan, S. and Maranas, C.D. (2015) 13C metabolic flux analysis at a genome-scale. *Metab. Eng.* 32, 12–22

158 Gopalakrishnan, S. and Maranas, C. (2015) Achieving Metabolic Flux Analysis for *S. cerevisiae* at a Genome-Scale: Challenges, Requirements, and Considerations. *Metabolites* 5, 521–535

159 Ferrer, P. and Albiol, J. (2014) ¹³C-Based Metabolic Flux Analysis of Recombinant Pichia pastoris. In *Metabolic Flux Analysis* 1191 (Krömer, J. O. et al., eds), pp. 291–313, Springer New York

160 Kim, T.Y. *et al.* (2012) Recent advances in reconstruction and applications of genome-scale metabolic models. *Curr. Opin. Biotechnol.* 23, 617–623

161 Kerkhoven, E.J. *et al.* (2014) Applications of computational modeling in metabolic engineering of yeast. *FEMS Yeast Res.* DOI: 10.1111/1567-1364.12199

162 Famili, I. *et al.* (2003) *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc. Natl. Acad. Sci. U. S. A.* 100, 13134–9

163 Förster, J. *et al.* (2003) Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Res.* 13, 244–53

164 Simeonidis, E. and Price, N.D. (2015) Genome-scale modeling for metabolic engineering. *J. Ind. Microbiol. Biotechnol.* 42, 327–338

165 Vallino, J.J. and Stephanopoulos, G. (1994) Carbon Flux Distributions at the Glucose 6-Phosphate Branch Point in *Corynebacterium glutamicum* during Lysine Overproduction. *Biotechnol. Prog.* 10, 327–334

166 van Gulik, W.M. and Heijnen, J.J. (1995) A metabolic network stoichiometry analysis of microbial growth and product formation. *Biotechnol. Bioeng.* 48, 681–698

167 Varma, A. *et al.* (1993) Biochemical production capabilities of *Escherichia coli*. *Biotechnol. Bioeng.* 42, 59–73

168 Feist, A.M. *et al.* (2009) Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* 7, 129–43

169 Pitkänen, E. *et al.* (2014) Comparative Genome-Scale Reconstruction of Gapless Metabolic Networks for Present and Ancestral Species. *PLoS Comput. Biol.* 10, e1003465

170 Edwards, J.S. and Palsson, B.O. (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5528–5533

171 Monk, J. *et al.* (2014) Optimizing genome-scale network reconstructions. *Nat. Biotechnol.* 32, 447–52

172 Thiele, I. and Palsson, B.Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121

173 Dikicioglu, D. *et al.* (2015) Biomass composition: the "elephant in the room" of metabolic modelling. *Metabolomics* 11, 1690–1701

174 Hamilton, J.J. and Reed, J.L. (2014) Software platforms to facilitate reconstructing genome-scale metabolic networks. *Environ. Microbiol.* 16, 49–59

175 Nett, J.H. *et al.* (2013) Characterization of the *Pichia pastoris* protein-O-mannosyltransferase gene family. *PLoS One* 8,

176 Liu, Y. *et al.* (2014) Toward metabolic engineering in the context of system biology and synthetic biology: advances and prospects. *Appl. Microbiol. Biotechnol.* 99, 1109–1118

177 Oberhardt, M.A. *et al.* (2009) Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* 5, 320

178 Österlund, T. *et al.* (2012) Fifteen years of large scale metabolic modeling of yeast: Developments and impacts. *Biotechnol. Adv.* 30, 979–988

179 Cvijovic, M. *et al.* (2011) Mathematical models of cell factories: moving towards the core of industrial biotechnology. *Microb. Biotechnol.* 4, 572–584

180 Boghigian, B.A. *et al.* (2010) Metabolic flux analysis and pharmaceutical production. *Metab. Eng.* 12, 81–95

181 Lewis, N.E. *et al.* (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.* 10, 291–305

182 Machado, D. and Herrgård, M. (2014) Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism. *PLoS Comput. Biol.* 10,

183 Gancedo, C. and Flores, C.-L. (2008) Moonlighting proteins in yeasts. *Microbiol. Mol. Biol. Rev.* 72, 197–210, table of contents

184 Dikicioglu, D. *et al.* (2014) Improving functional annotation for industrial microbes: a case study with *Pichia pastoris*. *Trends Biotechnol.* 32, 396–9

185 Bernard, T. *et al.* (2014) Reconciliation of metabolites and biochemical reactions for metabolic networks. *Brief. Bioinform.* 15, 123–35

186 Schnoes, A.M. *et al.* (2009) Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLoS Comput. Biol.* 5, e1000605

187 Pereira, R. *et al.* (2016) Improving the flux distributions simulated with genome-scale metabolic models of *Saccharomyces cerevisiae*. *Metab. Eng. Commun.* 3, 153–163

188 Hamilton, J.J. *et al.* (2013) Quantitative assessment of thermodynamic constraints on the solution space of genome-scale metabolic models. *Biophys. J.* 105, 512–522

189 Herrgård, M.J. *et al.* (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat. Biotechnol.* 26, 1155–60

190 Dobson, P.D. *et al.* (2010) Further developments towards a genome-scale metabolic model of yeast. *BMC Syst. Biol.* 4, 145

191 Heavner, B.D. *et al.* (2012) Yeast 5 - an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network. *BMC Syst. Biol.* 6, 55

192 Heavner, B.D. *et al.* (2013) Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database (Oxford).* 2013, bat059

193 Aung, H.W. *et al.* (2013) Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism. *Ind. Biotechnol.* 9, 215–228

194 Swainston, N. *et al.* (2016) Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics* 12,

195 Hefzi, H. *et al.* (2016) A Consensus Genome-scale

Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Syst.* 3, 434–443.e8

196 Chung, B.K.S. *et al.* (2010) Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. *Microb. Cell Fact.* 9, 1–15

197 Sohn, S.B. *et al.* (2010) Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for *in silico* analysis of heterologous protein production. *Biotechnol. J.* 5, 705–15

198 Caspeta, L. *et al.* (2012) Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and in silico evaluation of their potentials. *BMC Syst. Biol.* 6, 24

199 Büchel, F. *et al.* (2013) Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst. Biol.* 7, 116

200 Chung, B.K.-S. *et al.* (2013) Metabolic reconstruction and flux analysis of industrial *Pichia* yeasts. *Appl. Microbiol. Biotechnol.* 97, 1865–73

2

Background, aims and outline

Table of contents

2. Background, aims and outline

2.1. Background

This study has been carried out within the research group of Bioprocess Engineering and Applied Biocatalysis of the UAB. The group has a long-standing research programme in integrative industrial biotechnology, using the methylotrophic yeast *Pichia pastoris* as a cell factory for the production of recombinant proteins. The group has performed several quantitative physiology studies to expand the knowledge base of this cell factory as a basis for the rational design of metabolic engineering and process engineering strategies.

Solà *et al.* [1] carried out the first series of physiological studies using ^{13}C-isotopic tracer labelling experiments aimed at the metabolic profiling (metabolic flux ratio analysis) of central carbon metabolism of recombinant *P. pastoris* growing under different cultivation conditions, basically different carbon sources (glycerol, glucose and mixtures of glycerol and methanol) and growth rates. Later on, the availability of the *P. pastoris* genome sequence and subsequent development of omics techniques for *P. pastoris* allowed the research grout to perform several multilevel studies in a series of PhD thesis projects. Thus, following the initial studies by Aina Solà [2,3], the research group took part of the European project GENOPHYS, in which the effect of different environmental parameters such as oxygen availability, temperature and osmolarity on recombinant protein expression was systematically studied in cell factories, including *P. pastoris*. In this context, Kristin Baumann's thesis focused on the transcriptional and proteomic analysis of *P. pastoris* growing on glucose as carbon source at different levels of oxygen availability [4,5], whereas parallel studies on the same series of experiments by Marc Carnicer applied ^{13}C-based MFA and quantitative metabolomics [6,7]. Notably, he established new methodologies and protocols for metabolomics and combined analysis of transcriptomics, fluxomics and metabolomics datasets [6,8,9]. The combination of these two studies paved the path towards the systems-level analysis of *P. pastoris* and future implementation of systems metabolic engineering strategies for this cell factory [7]. Indeed, M. Carnicer showed a first attempt of integrating the multiple layers of physiological

information by incorporating different regulation levels (transcriptomic, metabolic and thermodynamic) for energetic parameter estimation in a precise and systematic manner [9].

In addition, Joel Jordà and Elena Cámara focused their studies on the adaptation of ^{13}C-based metabolic flux analysis and metabolomics and the use of transcriptomics to the case of growth on methanol as a carbon source, respectively. In particular, J. Jordà described the effects of the recombinant protein production on the central carbon metabolism by integrating multi-level datasets. J. Jordà characterised *P. pastoris* growing on mixtures of substrates: glycerol-methanol and glucose-methanol. He mainly used metabolomic and fluxomic data and applied thermodynamic constrains for metabolic fluxes calculation with special emphasis on the energetic and redox cofactor metabolism. E. Cármara performed transcriptomics analysis of a series of strains expressing different number of copies of the recombinant gene.

Simplified models for central carbon metabolism were used in all those studies, representing a potential limitation, as important reactions enabling better calculation of fluxes could be missing. Meanwhile, the first genome-scale metabolic models for *P. pastoris* were published and thus a potential tool for integrating large-scale data became available. Nevertheless, initial studies using these genome-scale metabolic models showed clear limitations.

Finally, these systems-level studies have provided a basis for the design of novel cell engineering strategies in *P. pastoris*. For instance, Núria Adelantado targeted the membrane lipid metabolism based on transcriptomic data from K. Baumann to engineer new strains showing enhanced capabilities to secrete a recombinant antibody fragment. In addition, integration of physiological information of *P. pastoris* under methanol growing conditions enabled E. Cámara to design novel strain and process engineering strategies. Nonetheless these strain improvement strategies have been designed on the basis of direct interpretation of –omics data. Therefore, in order to fully extract information from our increasing body of physiological datasets, subsequent steps for strain improvement should rely in both data integration and modelling, leading to a new generation of rational strain engineering, i.e. model-based strain engineering.

2.2. Aims and objectives

The aim of the present study was the development of strategies for rational strain engineering based on systems metabolic engineering of *P. pastoris*. Genome-scale metabolic models are essential tools for this purpose. Thus, a major objective of this work was the development of a consensus GSMM for *P. pastoris*. In addition, we aimed to extend and validate the consensus GSMM for a reliable application in a broader range of growth conditions, such as alternative carbon sources. A second major objective was the demonstration of the model performance and applicability for systems metabolic engineering of *P. pastoris*.

2.3. Outline

The aim of the thesis was the implementation of tools and strategies for systems metabolic engineering or *P. pastoris* based on genome-scale metabolic models. Although there three independent GSMMs were previously available for *P. pastoris*, initial steps towards their use manifested some weakness and limitations. Moreover, these models presented some discrepancies in certain pathways. Therefore, a consensus GSMM has been developed, as described in **Chapter 3**, thereby integrating the three previous versions and upgrading certain pathways. This allowed for a more detailed and accurate representation of the metabolism of *P. pastoris*. The consensus GSMM, iMT1026, was initially validated for a recombinant *P. pastoris* growing on glucose as sole carbon source under different oxygen availability conditions, as well as on mixtures of glycerol and methanol.

In **Chapter 4**, iMT1026 capabilities were further expanded to accurately describe growth on glycerol and methanol as carbon sources. These carbon sources are of special interest due to its potential utilisation in revalorisation of waste feedstocks from biodiesel industry and in the context of glycerol-based biorefineries. For this purpose, a series of chemostat cultivations at different dilution rates were performed using glycerol and methanol as sole carbon sources. In addition, macromolecular biomass composition was analysed and used to define new specific biomass equations for the iMT1026 model. As a result, a new version of iMT1026 (v3.0) was developed and validated for glycerol- and methanol-grown cells.

In order to provide a better insight on the metabolic phenotype of glycerol-grown cells, in **Chapter 5**, a series of ^{13}C-labelling experiments were performed in chemostat cultures operated at different dilution rates. Moreover, iMT1026 v3.0 was used for generating a reduced model of the core metabolism of *P. pastoris* enabling ^{13}C-based metabolic flux analysis with robust and consistent results.

A second application of the GSMM is shown in **Chapter 6**. Based on previous studies of the group, process and metabolic engineering strategies were combined to enhance recombinant protein production. Thus, a heterologous NADH kinase encoding gene was

overexpressed in a *P. pastoris* strain expressing an antibody fragment. The redox-engineered strain was cultivated on glycerol and glucose as carbon sources. Moreover, cultivations on glucose were performed under normoxic and reduced (hypoxic) oxygen availabilities. The iMT1026 v3.0 model was used as a tool for interpretation of the physiological response to both the environmental and genetic perturbations. The integration of experimental and simulated data allowed a global (systems-level) interpretation of the metabolic impact of the genetic and environmental perturbations induced to the system.

Finally, **Chapter 7** summarises the major outcomes and conclusions of this work, namely the development and validation of the genome-scale metabolic model of *P. pastoris* and its exploitation in two systems metabolic engineering applications.

2.4. References

1 Solà, A. *et al.* (2004) Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. *Eur. J. Biochem.* 271, 2462–70

2 Solà, A. *et al.* (2007) Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. *Microbiology* 153, 281–90

3 Baumann, K. *et al.* (2011) The impact of oxygen on the transcriptome of recombinant *S. cerevisiae* and *P. pastoris* - a comparative analysis. *BMC Genomics* 12, 218

4 Baumann, K. *et al.* (2011) Protein trafficking, ergosterol biosynthesis and membrane physics impact recombinant protein secretion in *Pichia pastoris*. *Microb. Cell Fact.* 10, 93

5 Carnicer, M. *et al.* (2012) Development of quantitative metabolomics for *Pichia pastoris*. *Metabolomics* 8, 284–298

6 Baumann, K. *et al.* (2010) A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions. *BMC Syst. Biol.* 4, 141

7 Carnicer, M. Systematic metabolic analysis of recombinant *Pichia pastoris* under different oxygen conditions A Metabolome and Fluxome Based Study. . (2012) , Universitat Autònoma de Barcelona

3

Generation of a consensus genome-scale metabolic model for *P. pastoris*

Table of contents

3. Generation of a consensus genome-scale metabolic model for *P. pastoris*

3.1. INTRODUCTION

The first metabolic network describing the central carbon metabolism of *P. pastoris* was used for ^{13}C-based metabolic flux ratio (METAFoR) analyses. Essentially, the network included described the core metabolism (including methanol metabolism, glycolysis, pentose phosphate pathway, TCA cycle, and anaplerotic reactions such as the glyoxylate shunt) as well as amino amino biosynthetic pathways, linking proteinogenic amino acids to their precursors in central carbon metabolism [1,2]. This network was based in the well characterized *S. cerevisiae* metabolic network [3], incorporating the methanol assimilating reactions specific to methylotrophic yeast. Notably, the initial ^{13}C-isotopic tracer experiments allowed validating the amino acid biosynthetic pathways of *P. pastoris*, being essentially equivalent to those described in *S. cerevisiae* [1]. Later studies used this metabolic network as a basis for formulating the central carbon metabolism model of *P. pastoris* and using it for ^{13}C-MFA [4–7]. More recently, when the *P. pastoris* genome sequence became available [8–10], two GSMMs were developed and published in 2010: iPP668 [11] and PpaMBEL1254 [12]. Two years later, a third GSMM was released: iLC915 [13]. In addition to these three curated models, other GSMMs were published derived automatically from the genome sequence using automatic reconstruction algorithms [14,15]. Nevertheless, these automatically-generated models are not curated and thus will not be discussed in this chapter.

Each one of the three currently available models is fully compartmentalised, but they differ in the number of reactions and metabolites. The first two models, iPP668 and PpaMBEL1254 have a comparable number of associated genes, reactions and metabolites, as well as similar metabolite and reaction identifications and nomenclature. On the other hand, iLC915 incorporates more *P. pastoris'* specific gene-protein-reaction associations and hence, a larger number of genes; however, several extracellular and nuclear reactions are missing. In

general terms, these models cover the same metabolic processes, but iLC915 is more detailed. Nevertheless, there still exist some critical issues in these models, such as missing and divergent information or reactions that require manual revision and curation. These divergences may be explained due to the automatization of some of the reconstruction steps. Protein function is derived from the annotation in public databases, therefore, misannotations in reference databases may propagate errors leading to a wrong gene-protein-reaction relation [16]. In addition, some discrepancies in metabolite assignations and reaction stoichiometry are also found. Commonly, such differences depend on the reference database used for reaction stoichiometry definition [17]. Unfortunately, *P. pastoris* GSMMs apparently used different reference databases as template, thus, alternative stoichiometries hamper model comparison. Generally, gene annotations and metabolic reconstructions of non-conventional organisms is based on reference well stablished and extensively physiological characterised model organisms, such as *S. cerevisiae*. Therefore, particular characteristics of the non-conventional organism are misrepresented in both annotations and derived metabolic reconstructions [18].

Similar to other organisms, such as *S. cerevisiae*, new versions and updated models integrate previous versions and incorporate new features and information from newly published literature. In the case of *S. cerevisiae*, despite the existence of other versions, a consensus metabolic model was developed [19] and it was further upgraded, being expanded and revised up to the 7th version [20–23].

In this chapter, we compare the models of *P. pastoris* and provide an upgraded consensus version. As mentioned above two different strains were used to obtain these models. Nonetheless, there is a high degree of identity at the amino acid coding sequences level (93.7%) and functional annotation between the two genome sequences [12]. In addition, no differences were observed in reactions involved in metabolising the different carbon sources [24]. Therefore, a major objective of our study was to obtain a general model that can be applicable to both strains. Furthermore, a comprehensive analysis was performed on several pathways, comparing the three models and updating them with the newly published literature. Recent findings on *P. pastoris* physiology and metabolism enabled to complete sphingolipid biosynthesis metabolism and glycosylation pathways, as well as the oxidative phosphorylation electron transport chain. Furthermore, we included different biomass

compositions specific for each of the alternative carbon sources used. Finally, the model accuracy was tested in a variety of physiological conditions.

3.2. MATERIALS AND METHODS

3.2.1. Model merging

For the model comparison (**Fig. 3-1**), an initial step of metabolite nomenclature unification was required. In PpaMBEL1254 only the identifier (ID) was available in the SBML file, i.e. neither the complete name of metabolites nor any association to a reference database was included. These metabolite IDs, were mostly the standard IDs most commonly used and therefore also included in both in BiGG [25] and The SEED [26] reference databases. After this first metabolite parsing and renaming step was done, MetaNetX [17] database was used in order crosslink information from KEGG [27], ChEBI [28] and MetaCyc [29]. This step not only allowed unifying metabolite names but also to include its molecular formula and charge at pH 7.2.

Once all metabolite names were unified, PpaMBEL1254 and iPP668 were compared using ModelBorgifier [30], thereby obtaining a first pre-merged model. In a second step, this merged model was compared to iLC915, generating a first draft of the consensus model. Due to important differences in model structure between iLC915 and the other models, a manual comparison was necessary. This was performed by analyzing the structure of each of the remaining pathways or subsystems. Differences were resolved according to the available literature, comparing the reactions with those included in two latest versions of the consensus *S. cerevisiae* GSMMs [22,23] or in another recently published *S. cerevisiae* GSMM [31]. Divergences in gene assignments were resolved using *P. pastoris* or *S. cerevisiae* literature. The *P. pastoris* high quality sequence annotation [10] and the automatic reconstructions for *P. pastoris* [14,15] were also used to verify annotations and gene-reaction assignments from the previous models. Finally, pathway revamping and addition of new reactions was performed based on available yeast literature and metabolic pathways/reaction databases [25,29,32,33].

Eventually, the network was loaded into a convenient environment for debugging [34]. Thus, both COBRA [35] and RAVEN [36] toolboxes were used in order to ensure pathway

connectivity and biomass formation. Duplicated reactions in the final model were deleted and blocked reactions were connected (gap filling) to the network when few steps were required. In addition, the elemental mass balance of each reaction was checked and corrected when unbalanced. The final model (**File S3-1** and **S3-5**) can also be obtained in SBML format from BIOMODELS database with accession number: MODEL1508040001 [37]. The SBML model was generated with the RAVEN toolbox [36] and validated with SBMLeditor [38].

3.2.2. Biomass and recombinant protein composition

The biomass reaction is defined by the sum of biomass components, grouped in macromolecules (carbohydrates, proteins, lipids, DNA, RNA), essential cofactors and ATP consumption associated to growth. This equation was adapted depending on culture conditions or carbon source used in accordance to the available literature experimental data [39,40]. In addition, composition of each macromolecule type, such as lipid and carbohydrate, was updated and extensively detailed due to the recently published detailed information of the specific composition [41,42]. See **File S3-2** for details in composition and calculations.

The model was also tested for the expression of two different recombinant proteins under different growth conditions: i) the antibody fragment 2F5 (FAB), expressed constitutively under the *GAP* promoter [43] and, ii) a *Rhizopus oryzae* lipase (ROL), regulated by the methanol inducible AOX promoter [44]. The dataset from the FAB-producing strain was used for simulations in oxygen limiting growth conditions [4,5], whereas the dataset from the ROL producing strain was used in simulations for glycerol-methanol co-feeding experimental conditions [2,40]. Reactions for heterologous protein production were included considering different levels: DNA sequence, transcription and mRNA formation, as well as translation and protein formation. Similarly to PpaMBEL1254 and iLC915, a ratio of $1:100:10^5$ between recombinant DNA (gene copies), mRNA and heterologous protein was assumed, as described in [12]. These equations also include energetic requirements for polymer formation [45]. Details of DNA, RNA and amino acid composition of each protein, as well as equations for the biosynthesis of their components can be found in **File S3-3**.

3.2.3. Energy requirements

Before the validation step, classic energetic parameters were estimated using experimental data. These parameters are the growth-associated maintenance energy (GAME) and non-growth-associated maintenance energy (NGAME). Both are represented as ATP consumption in the model. For the NGAME calculation, a classical approach was used [46]. For the glucose-limited cultivations, the glucose uptake rate (mmol glucose·gDCW^{-1}·h^{-1}) was represented against the specific growth rate using available data from [1,39,47,48]. The *y*-intercept of the linear regression line to this data corresponds to the amount of glucose needed for maintenance by non-growing cells. Using this value and the model, the NGAME can be calculated by maximization of the ATP turnover per mmol of glucose for the case of no biomass growth. Using the obtained value as fixed value for the non-growth associated maintenance, GAME is determined by adjusting the ATP consumption coefficient in biomass equation to fit biomass-substrate yields using experimental data (including CO_2 and O_2 constrains) from Carnicer et al. [39].

However, for the case of glycerol:methanol growth conditions NGAME was directly taken from the calculated values reported by Jordà et al. [40]. This was necessary due to the range of cultivation conditions considered and the insufficient experimental data available. Using these values, and similarly to the glucose-only growth condition, the GAME values for glycerol:methanol conditions were calculated by fitting the predicted values to the range of experimental biomass-substrate yields previously reported [40].

3.2.4. Model analysis and validation

Model analysis and validation were performed using both RAVEN [36] and COBRA [35] toolboxes as described below.

Carbon assimilation capabilities were determined maximizing growth rate and arbitrarily constraining the carbon source influx to 10 mmol·gDCW^{-1}·h^{-1} except were otherwise stated.

Reaction essentiality was determined performing an additional set of simulations. The procedure consisted on sequentially deleting each reaction of the model, maximizing biomass production and calculating the ratio of the resulting growth rate over the wild type result (GR$_{KO}$= growth$_{KO}$/growth$_{WT}$). The ratios obtained allowed to classify each reaction into three

categories: i) essential (GR_{KO} = 0), ii) partially-essential ($0 < GR_{KO} < 1$) iii) non-essential (GR_{KO} = 1).

Evaluation of the effect of oxygen limiting conditions on glucose cultures was performed constraining glucose and oxygen uptake rates to the measured values [5]. For the glycerol:methanol experimental conditions, only glycerol and methanol uptake rates were constrained to the experimental values [1] while O_2 uptake rate was left unconstrained. In all these cases biomass, CO_2 and by-products were left as unconstrained positive values and therefore appeared as calculated products were necessary. In all these cases biomass production was the maximized objective function.

3.3. RESULTS AND DISCUSSION

3.3.1. Model merging

As described in Materials and Methods section and summarized in **Fig. 3-1** the generation of the new model consists of several steps. In the first step of reconciliation, PpaMBEL1254 and iPP668 were automatically compared using modelBorgifier [30]. After the initial pairing step, 75% of the complete set of reactions was identified as identical (exact coincidence) reactions. This pre-merged model was compared with iLC915 resulting in a low number of identical or equivalent reactions (36% of the complete set). Nevertheless, a larger number of reactions were comparable. Those mainly differ in having different stoichiometric coefficients, being assigned to different compartments, decomposed in multi-step reactions, or using alternative names for metabolites (different synonyms, generic names or corresponding to enantiomer compounds).

The iPP668 and PpaMBEL1254 models were the first models to be published, and they are more similar to each other than to iLC915. Approximately 83% of the reactions from PpaMBEL1254 are present in iPP668 and 89% of the reactions in iPP668 are shared with PpaMBEL1254. Overall, models iPP668 and PpaMBEL1254 have a 75% of reaction identity. Furthermore, similar nomenclature and abbreviations are used in these two models. In addition, model structure and detail are similar to iND750 [49], from *S. cerevisiae*, and those models in the BiGG database [25].

The third published model, iLC915, has many differences with the previous two models. Its nomenclature and structure is KEGG-based [27]. Therefore, its metabolites are fully protonated and many pathways include the same number of steps described in KEGG. That is, many condensed or simplified metabolic branches in the other two models appear decomposed as multi-step reactions in iLC915. Such reaction differences among the models are one of the major reasons for the low pairing of iLC915 with the other two models and seem to be the result of the main database or model scaffold used as basis for model reconstruction.
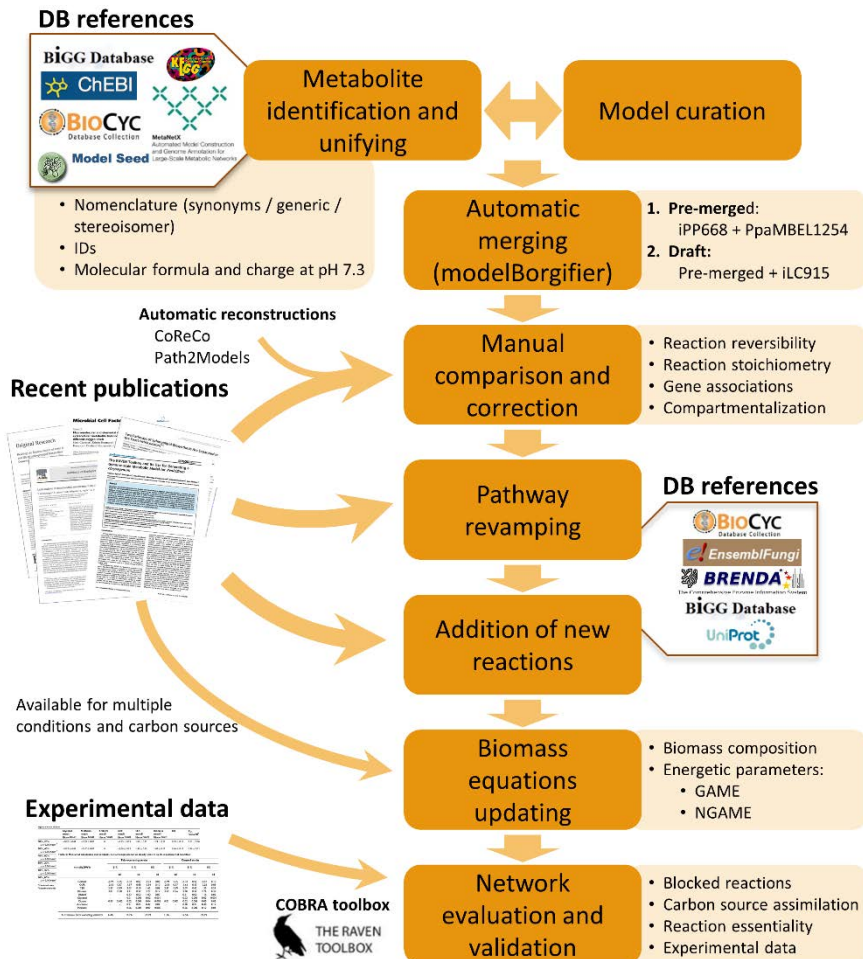
Fig. 3-1. Schematic overview of the major steps involved in the construction of P. pastoris GSMM iMT1026. The process of GSMMs integration started with the metabolite identification, unifying nomenclature and curation steps of iPP668, PpaMBEL1254 and iLC915. The continuation steps were performed on the resulting pre-merged model and subsequent drafts. Experimental data for model validation was taken from [2,5,39,40].

3.3.2. Updated pathways

As a result of the model comparison and merging step, some divergences in several pathways of the three existing models were evident. In addition, some of them were incomplete or misrepresented in all three models. Therefore, we engaged in a curation step according to the recently published literature and database information. Nevertheless, some difficulties in gene and pathway verification were found. The fact that new genomes are usually automatically annotated, at least in an initial step, and that enzyme activities or functions are inferred by homology, propagates errors from already annotated sequences to the new ones [16]. This issue arises when there is limited biological knowledge of the organism. Furthermore, not only genome annotations are based in other organisms and sequences, but GSMMs are also commonly developed from previous existing models. As a result, annotation errors or misrepresented pathways are also spread to the subsequent new/updated models. Moreover, as new GSMMs are mostly based in pre-existing reconstructions, few new metabolic reactions are often incorporated in the novel GSMMs versions. Consequently, the metabolic potential and biological diversity is often not fully reflected in the GSMMs and the total number of enzymatic activities present in the existing models remains far below from the complete catalogue of enzymatic steps described in the literature for each organism [50].

In the case of *P. pastoris*, the current annotation of its genome is rather limited [18], with most of all annotations being inferred by homology, mainly from *S. cerevisiae*. According to the best annotation available *P. pastoris* has 5040 protein-coding annotated genes of which only 3532 has been assigned an Ontology term and all but 21 annotations are automatically inferred [18]. Despite to the fact that *P. pastoris* and *S. cerevisiae* belong to the same family (*Saccharomycetaceae*), they present significant differences in their metabolic capabilities. Thus, besides *P. pastoris* well known additional pathways such as the methanol incorporation steps, other significant metabolic differences exist. More specifically, in this work, pathways such as sphingolipid biosynthesis, oxidative phosphorylation and glycosylation pathway, were adapted and redefined, as described below.

Fatty acid biosynthesis

Due to the limited information on specific fatty acid (FA) metabolic pathways in *P. pastoris*, it was assumed that most of *S. cerevisiae* fatty acid pathways were identical in the *P. pastoris* case. According to Hiltunen and co-workers [51], the latest version of yeast consensus model [23] and iTO977 model of *S. cerevisiae* [31], fatty acid biosynthesis takes place both in mitochondria and cytosol by fatty acid synthase (FAS) type II and I, respectively [52]. FAS type II has individual enzymes for each reaction in fatty acid *de novo* biosynthesis and elongation. Despite it is well known that mitochondrial FAS type II synthetizes at least up to C_8 fatty acid, some evidences suggest that this system can synthesize longer fatty acids [51,53]. While in *S. cerevisiae* 7th version of the consensus model [23] mitochondrial biosynthesis is up to C_8, this model also include reactions for up to C_{18} biosynthesis.

On the other hand, cytosolic FAS is a complex formed by Fas1p and Fas2p within which the successive elongation reactions take place and only the final acyl-CoA is released [54]. The final products of this cytosolic complex are considered to be C_{14} to C_{18} acyl-CoAs, mainly because they are the main fatty acids found in *P. pastoris* [41]. The biosynthesis takes place inside the complex in a number of four step cyclic reactions for each acetyl-CoA added. Different number of cycles results in a range FA (C_{14}, C_{16} and C_{18} acyl-CoAs). In addition to *de novo* biosynthesis, *P. pastoris* also has fatty acid elongation enzymes, which are able to extend C_{12-14} fatty acids and generate very long chain fatty acids (up to C_{26}).

Activation of free fatty acids (FFA) was considered to take place in cytosol only for C_{14}, C_{16} and C_{18} FFA, as well as their respective acyl-CoA hydrolysis. Finally, only acyl-CoA desaturations were included (that is not acyl-ACP or FFA) according to the pathway defined in *S. cerevisiae* [23].

Fatty acid oxidation

Two different transport mechanisms are commonly described depending on the FA chain length [55–57], both being closely coupled to its activation to acyl-CoA [58–62]: on the one hand, a simple diffusion and further activation of medium-chain fatty acids (up to C_{12} chain length) and, on the other hand, long and very long chain fatty acids are translocated as acyl-CoA concomitant with the corresponding ATP hydrolysis [63]. For the active transport

mechanism, the ATP has a cytosolic origin in ILC915, while in PpaMBEL1254 and iPP668 the required ATP is peroxisomal. According to [64,65], peroxisomal ATP is only required for medium chain fatty acid activation, therefore long and very long chain fatty acids transport should be dependent on cytosolic ATP.

Each cycle of β-oxidation is represented by 4 reactions. Nevertheless, for unsaturated fatty acid degradation, and due to its highly complex degradation steps which depend on the position of its double bounds [66,67], lumped reactions up to the generation of acetyl-CoA were taken from iPP668. However, the desaturation reaction of $C_{18:3}$ to $C_{18:2}$ was taken from iLC915.

Sphingolipid metabolism

General sphingolipid biosynthetic pathways in yeast are partially homologous to *S. cerevisiae* and they are extensively described in the literature [68,69]. Nevertheless, unlike *S. cerevisiae*, some yeasts such as *P. pastoris*, are able to produce glucosylceramides (GlcCer) from sphingolipid bases [70–73].

None of the three models include GlcCer biosynthesis. Ternes and co-workers [74,75] identified the gene role in GlcCer pathway and described the fatty acid specific composition in GlcCer and other sphingolipids, as well as the main chain sphingoid bases in *P. pastoris*.

Sphingoid bases can be derived from palmitoyl-CoA and stearoyl-CoA. However, only a 5% of the detected species correspond to the last one. In fact, Ternes and co-workers [75] characterized sphingolipid composition assuming all the species were formed with a palmitoyl-CoA derived sphingoid base. This sphingolipid composition is in agreement with other literature sources [41,76,77]. As palmitoyl-CoA bases represents around 95% of sphingoid bases, only palmitoyl-CoA derived ones are taken into account in this model.

Glycosylation pathways

Protein glycosylation pathways are not accurately described in previous models of *P. pastoris*. Only iLC915 partially included *N*-glycosylation, *O*-glycosylation and glycosylphosphatidylinositol-anchor (GPI-anchor) biosynthesis pathways. However, compartmentalization of several reactions of this pathway also required revision.

The first part of the *N*-glycosylation process is highly conserved among eukaryotes [78,79]. It takes place in the cytosol up to the addition of 5 mannose residues (Man) forming (Man)5(GlcNAc)2(PP-Dol)1 oligosaccharide. At this point, the oligosaccharide is transferred to the endoplasmic reticulum (ER), where up to 9 Man and 3 glucose residues (Glc) are further added [80]. In the second and less conserved part of the pathway, the dolichol diphosphate attachment to the protein is represented by a pseudo-reaction forming the compound (Glc)3(Man)9(GlcNAc)2(Asn)1 in which Asn represents an asparagine residue from the targeted protein. Once the oligosaccharide is attached to the Asn residue of the target protein, it is further modified by the removal of one Man. The resulting glycoprotein is transported to the Golgi Apparatus [81,82]. There, an heterogeneous pattern of glycosylation has been observed corresponding to the different heterologous proteins expressed in *P. pastoris* [83–85]. As an example, differences in Man residues range from 6 to 18 [86–88] and even may include hypermanosylation [89]. Due to its complexity and variability, in this model an average glycan is assumed to consist of (Man)9(GlcNAc)2(Asn). The resulting oligosaccharide contributes to the formation of a mannan (Man polymer represented by 1 mannose residue polymer) and a chitin (*N*-Acetylglucosamine polymer). Both contribute to the biomass formation as a specific component of the carbohydrate fraction.

Similarly to *N*-glycosylation in mannan formation, *O*-glycosylation is included assuming an average of 3 Man oligosaccharides [90–93]. *O*-glycosylation is also represented by a pseudo-reaction forming the compound (Man)1(Ser/Thr)1 in which Ser/Thr represents a serine or threonine residue of a protein.

Finally, GPI-anchor biosynthesis was also reviewed and compartments reassigned according to Orlean and Menon [94].

Oxidative phosphorylation

There are *P. pastoris* specific traits in respiratory chain that should be included in the GSMM. As an example, while *complex I* is not present in *S. cerevisiae* [95], it is described in *P. pastoris* [96,97]. None of the previous *P. pastoris* models include respiratory *complex I* and so, an important proton translocation step was missing. In general, two main traits of oxidative

phosphorylation were deeply analyzed: the proton stoichiometry and the complexes integrating the electron transport chain.

In this model the mitochondrial intermembrane space has been included. Hence proton translocation is assumed to occur from the mitochondrial matrix to the intermembrane space. Thus, protons pumped out from the mitochondria do not merge with the high amount of protons from the cytosolic space. Regarding the electron transfer chain reactions, each of the previous *P. pastoris* models shows a different stoichiometry for proton translocation. After a review of the stoichiometry and relevant literature it was decided to apply a stoichiometry that satisfies the H$^+$ balance of the metabolites' charged formula and including *complex I* stoichiometry considerations proposed by Wikström and Hummer [98]. This includes the *complex I* translocating 4 H$^+$ to the intermembrane space [98,99].

In the present model, reactions for non H$^+$ translocating outer and inner mitochondrial membrane NAD(P)H dehydrogenases (cytoplasmic side and matrix side) were also included. While inner NADH dehydrogenase appears in all three models, outer NADH dehydrogenase was only present in iPP668 and PpaMBEL1254, despite both dehydrogenases have been previously reported [96,100,101]. We also included PAS_chr1-4_0299 putative NADPH dehydrogenase homologue to *Kluyveromyces lactis* [102,103] and also included in its metabolic reconstruction [104].

On the other hand, *complex III* and *IV* are included in all three models, but several discrepancies exist on the details of the H$^+$ balance due to the consideration of alternative metabolite's molecular formulas or the result of using different criteria when considering chemical and translocated protons [105]. An additional trait for *complex III* equations is the complexity on the stoichiometric representation of Q-cycle [106,107]. Therefore, in this model stoichiometric coefficients of equations for *complex III* and *IV* reactions were chosen with special attention to the proton balance. The selected stoichiometry was of 2H$^+$/2e$^-$ for *complex III* and 4H$^+$/2e$^-$ for *complex IV* according to recent literature [99,106,107].

One additional characteristic of *P. pastoris* is the presence of an alternative oxidase that could bypass *complex III* and *IV* in the respiratory chain [95,108] which seems to be active only in certain growth conditions [95,100,109]. Although this oxidase was only included iLC915,

our consensus model also incorporates this reaction in the electron transport chain module. Finally, the stoichiometry for ATP synthase was maintained as in PpaMBEL1254 and iPP668 ($4H^+$/ATP, resulting in a final maximum theoretical stoichiometry of $2.5e^-$/ATP) as its H^+ balance is in agreement with the available literature [110].

Other reviewed pathways

Metabolization of some sugars was also updated. Some sugars such as starch, maltose or cellobiose were able to be assimilated in the previous models. However, Kurtzman [24] and Naumov and co-workers [111] characterized substrate assimilation in *P. pastoris* and reported no growth for these carbon sources. Consequently, reasons for their metabolization were revised and discarded reactions are detailed in **File S3-4**. L-rhamnose assimilation was only possible in PpaMBEL1254. However, the included metabolic steps for its metabolization were not typical of yeast species. Therefore reactions and genes associated to the metabolization of L-rhamnose were added as suggested in [112], similarly as also described for *P. stipitis* [113,114].

A full list of modifications from the original models, also including some additional pathways and subsystems not discussed above such as phosphatidylinositol synthesis or transport reactions, is provided in **File S3-4**.

3.3.3. General characteristics of the model

After following all the steps described above, an extended model is obtained that integrates previous *P. pastoris'* GSMMs: iMT1026, which is provided in **File S3-1**, **File S3-5** and also available at BIOMODELS database (MODEL1508040001) [37]. The general characteristics of the resulting model are described in the following.

In the first place, this model includes an increased number of gene-protein-reaction relationships as can be seen in .

Our final model incorporated 185 new reactions that didn't appear in previous models and has 614 common reactions in all three models (**Fig. 3-2**). Reactions appearing in **Fig. 3-2** as common to two or all three different models include those reactions that have been taken directly from the previous model. Therefore, they do not include those reactions that are

either not the same but equivalent, have been decomposed in several reactions or are a combination of several other reactions. Thus, taking into account these multi-step, lumped or decomposed reactions, in the final model there are up to 721 equivalent reactions in common in all three models, 504 in two of the three models and 638 reactions in only one of the models, without any clear equivalence in any of the other two models.

Table 3-1. Comparison of the main features of iMT1026 and previous P. pastoris GSMMs

	iPP668	PpaMBEL1254	iLC915	iMT1026
Genes	668	540	915	1026
Metabolites [a]	1177 (684)	1058 (696)	1302 (899)	1689 (1018)
Reactions	1354	1254	1423	2035
Cytosolic	623	604	790	1059
Mitochondrial	163	155	205	268
Peroxisomal	66	66	64	102
Extracellular	12	11	0	16
Endoplasmic reticulum	15	7	34	41
Golgi apparatus	4	8	4	13
Vacuolar	3	6	12	9
Nuclear	16	17	0	17
Intercompartmental/Transport	452	328	314	510

[a] Total number of metabolites, with compartment, and unique metabolites (in brackets).
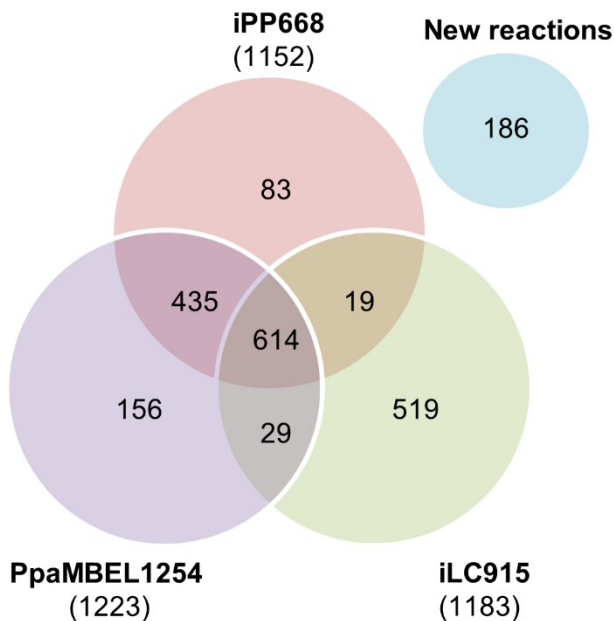
Fig. 3-2. Reactions from PpaMBEL1254, iPP668 and iLC915 included in iMT1026 model

3.3.4. Maintenance and growth-associated ATP calculations

As described in Materials and Methods, data from [1,39,47,48] was used to calculate the NGAME. As an initial step, a value of 32 mol ATP/ mol glucose, obtained by maximizing the ATP turnover using 1 mmol/(gDCW·h) of glucose. The *y*-intercept from a representation of glucose uptake rate versus growth rate was 0.0878 mmol glucose/(gDCW·h) which corresponds to 2.81 mmol ATP/(gDCW·h) using the glucose-ATP conversion factor obtained in the initial step. This calculated glucose NGAME value is similar to the 2.3 mmol ATP/(gDCW·h), previously proposed by [11] for glucose, and close to the NGAME estimated for *Pichia* (*Scheffersomyces*) *stipitis* also growing in glucose [115].

On the other hand, GAME for glucose was estimated by fitting the calculated biomass-substrate yields to experimental data. This way a value of 72 mmol ATP/gDCW was obtained. This amount of ATP associated to cell growth is also similar to 70.5 mmol ATP/gDCW calculated previously for *P. pastoris* [13] and close to the experimentally calculated values for *S. cerevisiae* of 62-71 mmol ATP/gDCW [116,117], and to the 69.2 mmol ATP/gDCW, *in silico* estimated [118].

For the case of the glycerol and methanol co-feeding cultivations, ATP maintenance values calculated by Jordà and co-workers [40] were used as NGAME as available data was insufficient for a new determination. These values range from 4.5 to 5.7 mmol ATP/gDCW and are similar to 6 mmol ATP/gDCW, proposed in iLC915 [13]. For the different conditions tested, specific GAMEs were calculated by fitting the simulations to experimental data from [40]. These experimental data show that the ratio of the glycerol or methanol uptake rates with the growth rate is different for each pair of glycerol:methanol feeding conditions (80:20, 60:40 and 40:60, % w/w at 0.05 and 0.16 h^{-1}growth rates), therefore a specific GAME was calculated for each case. The obtained values ranged within the 69.8 and 125.6 mmol ATP/gDCW interval. These GAME values increase with the fraction of methanol in the mixed feeding and are in agreement with those calculated by Caspeta and co-workers [13], who estimated a maximum GAME for methanol as sole carbon source of 150 mmol ATP/gDCW.

3.3.5. Carbon source assimilation

The model agreement with *P. pastoris* utilization of different carbon sources was tested and compared to experimental data [24,111]. A total of 47 carbon sources were evaluated (**Table 3-2**) using an *in silico* minimal medium, with ammonium, phosphate and biotin. The model successfully predicts carbon assimilation for all sources tested.

Table 3-2. Evaluation of the substrate assimilation capabilities in P. pastoris.

Substrate	Experimental [a]	In silico	Substrate	Experimental [a]	In silico
D-glucose	+	+	Erythrol	-	-
DL-lactate	+	+	Galactitol	-	-
Mannitol	+	+	Galactose	-	-
Ethanol	+	+	Hexadecane	-	-
Glycerol	+	+	*myo*-Inositol	-	-
L-Rhamnose	+	+	Inulin	-	-
Methanol	+	+	Lactose	-	-
Succinate	+	+	L-Arabitol	-	-
Trehalose	+	+	L-Arabinose	-	-
Sorbitol	+	+	L-Sorbose	-	-
Citrate	v	+	L-Tryptophan	-	-
D-Xylose	- / + [b]	+	Maltose	-	-
Xylitol	- / + [c]	+	Meleziose	-	-
5-keto-D-gluconate	-	-	Melibiose	-	-
Arbutin	-	-	Methyl-α -D-glucoside	-	-
Cellobiose	-	-	N-acetylglucosamine	-	-
D-arabinose	-	-	Raffinose	-	-
D-Galacturonate	-	-	Ribitol	-	-
D-gluconate	-	-	Saccharate	-	-
D-glucono-1,5-lactone	-	-	Salicin	-	-
D-Glucosamine	-	-	Soluble starch	-	-
D-Glucuronate	-	-	Sucrose	-	-
D-Ribose	-	-			

+, carbon assimilation and growth; -, no carbon assimilation; v, variable growth/non-growth experiments observed.
[a] Experimental data from [24,119].
[b] Described in [120].
[c] Described in [121].

3.3.6. Reaction essentiality

An interesting trait of the obtained GSMMs is the identification of those reactions critical or essential for biomass growth (essential reactions). Simulations to determine reaction essentiality were performed for glucose in normoxia, limited oxygen and hypoxia conditions and also for mixtures of glycerol and methanol at different growth rates. No significant difference in reaction essentiality was observed in all the conditions tested. Thus, similar patterns of distribution of essential reactions in each pathway are observed for all the cases. The results, grouped into major metabolic pathways, are summarized in **Fig. 3-3** (and **Fig.**

S3-1). From a global point of view, essential reactions represent a 15-16% of the total reactions, while 76-79% of the reactions are classified as non-essential. The remaining 6-9% are partially-essential and its deletion causes a decrease in growth rate.
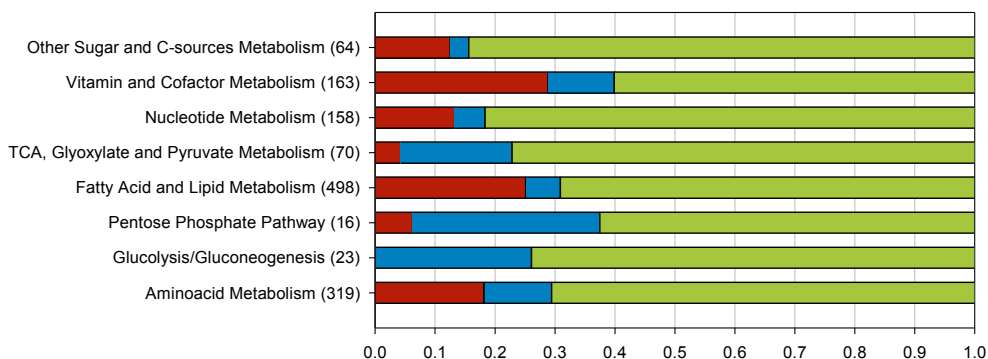
Fig. 3-3. Summary of reaction essentiality results for glucose simulations grouped into major pathways. FBA was performed by optimizing biomass production and sequentially constraining to 0 each reaction in the corresponding simulations. The resulting growth rate was compared with the wild type one. Metabolic reactions were classified in three categories according to the relative growth rate obtained: Essential (E), partially-essential (PE) and non-essential (NE). X axis represent the fraction of each type of reactions in each category of E (in red), PE (in blue) and NE (in green). Reactions are distributed in 8 major subsystems (Y axis). Numbers between brackets indicate number of reactions in each group. Equivalent figures for oxygen limiting conditions and glycerol:methanol simulations can be found in **Fig. S3-1**.

The results show that there are 312 condition independent essential reactions, which are essential in all the performed simulations. These essential reactions can be grouped in three main groups: in the first group, most of them are associated with lipid metabolism (40%), in the second group most reactions belong to the amino acid metabolism (18.6%) while the third group mostly includes cofactor related essential reactions as a 14.1% of all common essential reactions. Similarly to other GSMM models, such as *S. cerevisiae*, extending the model and including more detailed biomass composition results in an increase of essential reactions directly linked to the biomass related metabolites. Nevertheless, this essentiality could be overestimated *in silico*, due to the fact that *in vivo* systems are able to replace de missing species with other similar biomass components.

3.3.7. Model validation

An additional step of model validation was performed by comparing the model predicted values with an additional set of experimental data including diverse combinations of glucose, glycerol and methanol chemostats [5,40].

For the glucose chemostats, simulations with different oxygen availabilities, growth rates and CO_2 production were successfully predicted with errors lower than 6% for both FAB expressing and non-expressing strains (**Fig. 3-4 A-C**). According to the available experimental data [5,39], when the oxygen availability decreases ethanol and other metabolites are secreted. The present model is also able to predict by-product formation when oxygen-limited conditions are simulated. However, the model predicts a slightly higher production of ethanol and none of the other by-products, such as arabitol or pyruvate. However, when ethanol secretion flux is constrained to the experimentally measured value, arabitol secretion is also predicted (**Fig. 3-4** C). In addition to arabitol, pyruvate is also secreted in the *in silico* predictions when both ethanol and arabitol are constrained to the experimental values. This discrepancy of the model to directly predict arabitol or pyruvate secretion if no additional constraint is imposed points to additional regulatory constraints other than those strictly stoichiometric. In addition, different cofactor utilization by combinations of isoenzymes [122,123], and their impact on $NAD(P)^+/NAD(P)H$ regeneration, could be one of the key factors for the production of those alternative products [124,125]. Constraining the model with additional ^{13}C-labelling data [126] as well as studying the impact of cofactor perturbation and analyzing these cofactor demands for cell growth, as done in other organisms [127] would be interesting approaches to consider in future studies.

For the second dataset, (glycerol:methanol mixtures), specific growth rate, together with specific O_2 consumption and CO_2 production rates were predicted within a 11% of deviation, as shown **Fig. 3-4 E-G**. Similarly to the above described glucose tests, arabitol was only produced when ethanol was constrained to the experimental values; otherwise, ethanol is the preferred product of the stoichiometric model. As in the previous glucose case these results point to another possible level of regulation for arabitol production not included in the model, as without it ethanol production appears as the most efficient way to regenerate NAD^+ for maximum biomass production.

In order to compare the accuracy of our model with the previous existing models a set of simulations were performed for glucose and glycerol:methanol cultivations (**APPENDIX I-B**). The same constrains were set to all the models and growth rate, CO_2 production and O_2 consumption (only in the glycerol:methanol simulations) were compared to the experimental
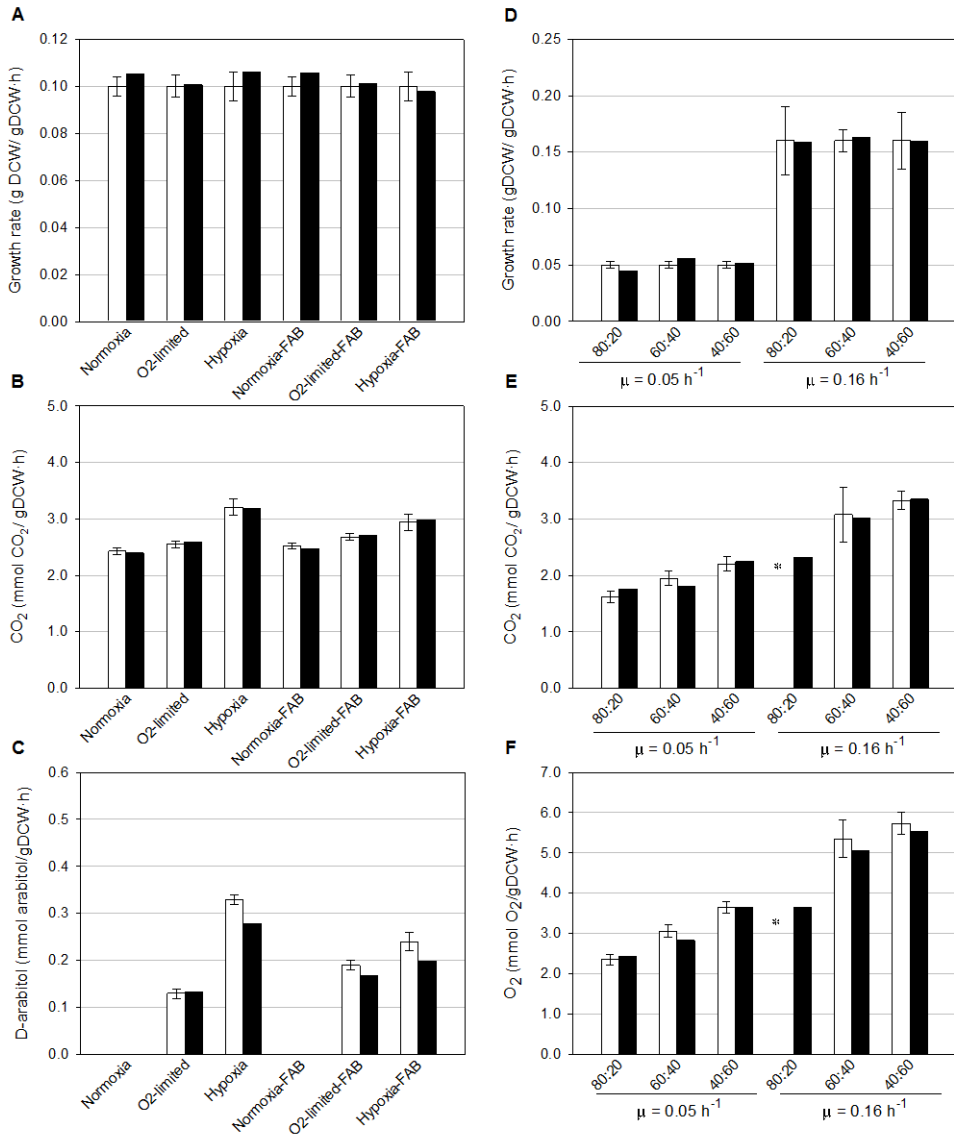
Fig. 3-4. Results of the model validation. Graphs with (A) growth rate (B) CO_2 and (C) D-arabitol production predictions simulating glucose chemostats at different oxygen conditions with and without recombinant protein production [5,39] with glucose, O_2 and ethanol fluxes constrained to the experimental values. (D) Growth rate (E) CO_2 production and (F) O_2 consumption predictions simulating different glycerol:methanol chemostats [2,40]. White and black bars correspond to experimental and predicted data respectively.* Not determined in [2].

values [5,40]. As shown in **APPENDIX I-B**, iMT1026 can predict the evaluated macroscopic cultivation parameters more accurately, i.e. with smaller deviations from experimental data **Table S3-1**. Moreover, our model is also able to describe byproduct secretion under respirofermentative conditions.

3.4. CONCLUSIONS

In summary, a consensus GSMM of the yeast *P. pastoris* integrating the three preexisting models has been obtained. Importantly, the new GSMM, iMT1026, is more complete and includes a comprehensive revision and upgrading of several metabolic processes (e.g. fatty acid and sphingolipid metabolism, protein glycosylation and energy metabolism) based on new information emerged from recent literature. Furthermore, the new GSMM has been validated using different sets of experimental data corresponding to a wider range of physiological states than previous GSMMs. From our point of view this GSMM improves the capabilities in terms of accuracy of predictions/simulations in relation to previous models. Overall, we provide an improved tool to the *P. pastoris* community for the physiological analysis and understanding of this yeast. It is expected that on-going efforts in the functional (re)annotation of the *P. pastoris* genome will allow for further improvements of its GSMMs by all the *P. pastoris* community. In a more wide perspective, it also has to be pointed out the importance of curating and manually revising new GSMMs of non-model organisms that are based on GSMM scaffolds from related model organisms. Despite the comprehensiveness of these scaffolds, an exhaustive analysis of specific metabolic traits of the non-model organism is still essential to construct a GSMM describes/predicts its metabolic phenotype accurately.

3.5. REFERENCES

1 Solà, A. *et al.* (2004) Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. *Eur. J. Biochem.* 271, 2462–70

2 Solà, A. *et al.* (2007) Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. *Microbiology* 153, 281–90

3 Maaheimo, H. *et al.* (2001) Central carbon metabolism of Saccharomyces cerevisiae explored by biosynthetic fractional 13C labeling of common amino acids. *Eur. J. Biochem.* 268, 2464–2479

4 Baumann, K. *et al.* (2010) A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions. *BMC Syst. Biol.* 4, 141

5 Carnicer, M. *et al.* (2012) Quantitative metabolomics analysis of amino acid metabolism in recombinant *Pichia pastoris* under different oxygen availability conditions. *Microb. Cell Fact.* 11, 83

6 Jordà, J. *et al.* (2012) Metabolic flux profiling of recombinant protein secreting *Pichia pastoris* growing on glucose:methanol mixtures. *Microb. Cell Fact.* 11, 57

7 Jordà, J. *et al.* (2013) Glucose-methanol co-utilization in *Pichia pastoris* studied by metabolomics and instationary 13C flux analysis. *BMC Syst. Biol.* 7, 17

8 Mattanovich, D. *et al.* (2009) Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*. *Microb. Cell Fact.* 8, 29

9 De Schutter, K. *et al.* (2009) Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nat. Biotechnol.* 27, 561–6

10 Küberl, A. *et al.* (2011) High-quality genome sequence of *Pichia pastoris* CBS7435. *J. Biotechnol.* 154, 312–20

11 Chung, B.K.S. *et al.* (2010) Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. *Microb. Cell Fact.* 9, 1–15

12 Sohn, S.B. *et al.* (2010) Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for *in silico* analysis of heterologous protein production. *Biotechnol. J.* 5, 705–15

13 Caspeta, L. *et al.* (2012) Genome-scale metabolic reconstructions of *Pichia stipitis* and *Pichia pastoris* and in silico evaluation of their potentials. *BMC Syst. Biol.* 6, 24

14 Pitkänen, E. *et al.* (2014) Comparative Genome-Scale Reconstruction of Gapless Metabolic Networks for Present and Ancestral Species. *PLoS Comput. Biol.* 10, e1003465

15 Büchel, F. *et al.* (2013) Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst. Biol.* 7, 116

16 Schnoes, A.M. *et al.* (2009) Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLoS Comput. Biol.* 5, e1000605

17 Bernard, T. *et al.* (2014) Reconciliation of metabolites and biochemical reactions for metabolic networks. *Brief. Bioinform.* 15, 123–35

18 Dikicioglu, D. *et al.* (2014) Improving functional annotation for industrial microbes: a case study with *Pichia pastoris*. *Trends Biotechnol.* 32, 396–9

19 Herrgård, M.J. *et al.* (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat. Biotechnol.* 26, 1155–60

20 Dobson, P.D. *et al.* (2010) Further developments towards a genome-scale metabolic model of yeast. *BMC Syst. Biol.* 4, 145

21 Heavner, B.D. *et al.* (2012) Yeast 5 - an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network. *BMC Syst. Biol.* 6, 55

22 Heavner, B.D. *et al.* (2013) Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database (Oxford).* 2013, bat059

23 Aung, H.W. *et al.* (2013) Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism. *Ind. Biotechnol.* 9, 215–228

24 Kurtzman, C.P. (2005) Description of *Komagataella phaffii* sp. nov. and the transfer of *Pichia pseudopastoris* to the methylotrophic yeast genus *Komagataella*. *Int. J. Syst. Evol. Microbiol.* 55, 973–976

25 Schellenberger, J. *et al.* (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* 11, 213

26 Henry, C.S. *et al.* (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* 28, 977–82

27 Kanehisa, M. *et al.* (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205

28 Degtyarenko, K. *et al.* (2007) ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res.* 36, D344–D350

29 Caspi, R. *et al.* (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 40, D742-53

30 Sauls, J.T. and Buescher, J.M. (2014) Assimilating genome-scale metabolic reconstructions with modelBorgifier. *Bioinformatics* 30, 1036–1038

31 Österlund, T. *et al.* (2013) Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst. Biol.* 7, 36

32 Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30

33 Schomburg, I. *et al.* (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res.* 41, D764-72

34 Thiele, I. and Palsson, B.Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5, 93–121

35 Schellenberger, J. *et al.* (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* 6, 1290–307

36 Agren, R. *et al.* (2013) The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum. *PLoS Comput. Biol.* 9, e1002980

37 Chelliah, V. *et al.* (2015) BioModels: ten-year anniversary. *Nucleic Acids Res.* 43, D542–D548

38 Rodriguez, N. *et al.* (2007) SBMLeditor: effective creation of models in the Systems Biology Markup Language (SBML). *BMC Bioinformatics* 8, 79

39 Carnicer, M. *et al.* (2009) Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels. *Microb. Cell Fact.* 8, 65

40 Jordà, J. *et al.* (2014) Metabolic flux analysis of recombinant *Pichia pastoris* growing on different glycerol/methanol mixtures by iterative fitting of NMR-derived 13C-labelling data from proteinogenic amino acids. *N. Biotechnol.* 31, 120–132

41 Grillitsch, K. *et al.* (2014) Isolation and characterization of the plasma membrane from the yeast *Pichia pastoris*. *Biochim. Biophys. Acta - Biomembr.* 1838, 1889–1897

42 Roca, C. *et al.* (2012) Production of yeast chitin–glucan complex from biodiesel industry byproduct. *Process Biochem.* 47, 1670–1675

43 Gasser, B. *et al.* (2006) Engineering of *Pichia pastoris* for improved production of antibody fragments. *Biotechnol. Bioeng.* 94, 353–361

44 Minning, S. *et al.* (1998) Functional expression of *Rhizopus oryzae* lipase in *Pichia pastoris*: high-level production and some properties. *J. Biotechnol.* 66, 147–156

45 Stephanopoulos, G. *et al.* (1998) Review of Cellular Metabolism. In *Metabolic Engineering* 16pp. 21–79, Elsevier

46 Pirt, S.J. (1982) Maintenance energy: a general model for energy-limited and energy-sufficient growth. *Arch. Microbiol.* 133, 300–302

47 Chung, B.K.-S. (2012) , Yeast Systems Biotechnology for

Production of Value-Added Biochemicals. , National University of Singapore

48 Rebnegger, C. *et al.* (2014) In *Pichia pastoris*, growth rate regulates protein synthesis and secretion, mating and stress response. *Biotechnol. J.* 9, 511–25

49 Duarte, N.C. *et al.* (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* 14, 1298–309

50 Monk, J. *et al.* (2014) Optimizing genome-scale network reconstructions. *Nat. Biotechnol.* 32, 447–52

51 Hiltunen, J.K. *et al.* (2009) Mitochondrial fatty acid synthesis type II: more than just fatty acids. *J. Biol. Chem.* 284, 9011–5

52 Tehlivets, O. *et al.* (2007) Fatty acid synthesis and elongation in yeast. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* 1771, 255–270

53 Kursu, V.A.S. *et al.* (2013) Defects in mitochondrial fatty acid synthesis result in failure of multiple aspects of mitochondrial biogenesis in Saccharomyces cerevisiae. *Mol. Microbiol.* 90, 824–840

54 Schweizer, E. and Hofmann, J. (2004) Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiol. Mol. Biol. Rev.* 68, 501–517

55 van Roermund, C.W.T. *et al.* (2001) Identification of a peroxisomal ATP carrier required for medium-chain fatty acid beta-oxidation and normal peroxisome proliferation in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 21, 4321–9

56 Theodoulou, F.L. *et al.* (2006) Peroxisomal ABC transporters. *FEBS Lett.* 580, 1139–55

57 Visser, W.F. *et al.* (2007) Metabolite transport across the peroxisomal membrane. *Biochem. J.* 401, 365–75

58 Hiltunen, J.K. *et al.* (2003) The biochemistry of peroxisomal beta-oxidation in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* 27, 35–64

59 Watkins, P.A. *et al.* (1998) Disruption of the *Saccharomyces cerevisiae* FAT1 Gene Decreases Very Long-chain Fatty Acyl-CoA Synthetase Activity and Elevates Intracellular Very Long-chain Fatty Acid Concentrations. *J. Biol. Chem.* 273, 18210–18219

60 Watkins, P.A. and Ellis, J.M. (2012) Peroxisomal acyl-CoA synthetases. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1822, 1411–1420

61 Zou, Z. *et al.* (2003) Vectorial acylation in *Saccharomyces cerevisiae*. Fat1p and fatty acyl-CoA synthetase are interacting components of a fatty acid import complex. *J. Biol. Chem.* 278, 16414–22

62 Obermeyer, T. *et al.* (2007) Topology of the yeast fatty acid transport protein Fat1p: mechanistic implications for functional domains on the cytosolic surface of the plasma membrane. *J. Lipid Res.* 48, 2354–64

63 van Roermund, C.W.T. *et al.* (2012) Peroxisomal fatty acid uptake mechanism in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 287, 20144–53

64 Nakagawa, T. *et al.* (2000) Peroxisomal Membrane Protein Pmp47 Is Essential in the Metabolism of Middle-chain Fatty Acid in Yeast Peroxisomes and Is Associated with Peroxisome Proliferation. *J. Biol. Chem.* 275, 3455–3461

65 Rottensteiner, H. and Theodoulou, F.L. (2006) The ins and outs of peroxisomes: Co-ordination of membrane transport and peroxisomal metabolism. *Biochim. Biophys. Acta - Mol. Cell Res.* 1763, 1527–1540

66 Goepfert, S. and Poirier, Y. (2007) Beta-oxidation in fatty acid degradation and beyond. *Curr. Opin. Plant Biol.* 10, 245–51

67 Gurvitz, A. *et al.* (2001) Peroxisomal degradation of trans-unsaturated fatty acids in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* 276, 895–903

68 Cowart, L.A. and Obeid, L.M. (2007) Yeast sphingolipids: Recent developments in understanding biosynthesis, regulation, and function. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* 1771, 421–431

69 Dickson, R.C. (2008) Thematic Review Series: Sphingolipids. New insights into sphingolipid metabolism and function in budding yeast. *J. Lipid Res.* 49, 909–921

70 Aguilera-Romero, A. *et al.* (2014) Sphingolipid homeostasis in the web of metabolic routes. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* 1841, 647–656

71 Börgel, D. *et al.* (2012) Metabolic engineering of the non-conventional yeast *Pichia ciferrii* for production of rare sphingoid

bases. *Metab. Eng.* 14, 412–26

72 Cheon, S.A. *et al.* (2012) Distinct roles of two ceramide synthases, CaLag1p and CaLac1p, in the morphogenesis of *Candida albicans*. *Mol. Microbiol.* 83, 728–45

73 Takakuwa, N. *et al.* (2008) Significance of the *KlLAC1* gene in glucosylceramide production by *Kluyveromyces lactis*. *FEMS Yeast Res.* 8, 839–45

74 Ternes, P. *et al.* (2006) Identification of fungal sphingolipid C9-methyltransferases by phylogenetic profiling. *J. Biol. Chem.* 281, 5582–92

75 Ternes, P. *et al.* (2011) Two pathways of sphingolipid biosynthesis are separated in the yeast *Pichia pastoris*. *J. Biol. Chem.* 286, 11401–14

76 Ejsing, C.S. *et al.* (2006) Collision-induced dissociation pathways of yeast sphingolipids and their molecular profiling in total lipid extracts: a study by quadrupole TOF and linear ion trap-orbitrap mass spectrometry. *J. Mass Spectrom.* 41, 372–89

77 Klug, L. *et al.* (2014) The lipidome and proteome of microsomes from the methylotrophic yeast *Pichia pastoris*. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* 1841, 215–226

78 Lehle, L. *et al.* (2006) Protein Glycosylation, Conserved from Yeast to Man: A Model Organism Helps Elucidate Congenital Human Diseases. *Angew. Chemie Int. Ed.* 45, 6802–6818

79 Delic, M. *et al.* (2013) The secretory pathway: exploring yeast diversity. *FEMS Microbiol. Rev.* 37, 872–914

80 Burda, P. and Aebi, M. (1999) The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta - Gen. Subj.* 1426, 239–257

81 Bretthauer, R.K. (2003) Genetic engineering of *Pichia pastoris* to humanize N-glycosylation of proteins. *Trends Biotechnol.* 21, 459–462

82 Rich, J.R. and Withers, S.G. (2009) Emerging methods for the production of homogeneous human glycoproteins. *Nat. Chem. Biol.* 5, 206–15

83 Cereghino, J.L. and Cregg, J.M. (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol. Rev.* 24, 45–66

84 Daly, R. and Hearn, M.T.W. (2005) Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J. Mol. Recognit.* 18, 119–38

85 Vervecken, W. *et al.* (2004) In Vivo Synthesis of Mammalian-Like, Hybrid-Type N-Glycans in *Pichia pastoris*. *Appl. Environ. Microbiol.* 70, 2639–2646

86 Grinna, L.S. and Tschopp, J.F. (1989) Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast, *Pichia pastoris*. *Yeast* 5, 107–15

87 Montesino, R. *et al.* (1998) Variation in N-Linked Oligosaccharide Structures on Heterologous Proteins Secreted by the Methylotrophic Yeast *Pichia pastoris*. *Protein Expr. Purif.* 14, 197–207

88 Hirose, M. *et al.* (2002) Characterization of N-linked oligosaccharides attached to recombinant human antithrombin expressed in the yeast *Pichia pastoris*. *Yeast* 19, 1191–202

89 Scorer, C.A. *et al.* (1993) The intracellular production and secretion of HIV-1 envelope protein in the methylotrophic yeast *Pichia pastoris*. *Gene* 136, 111–119

90 Lommel, M. and Strahl, S. (2009) Protein O-mannosylation: conserved from bacteria to humans. *Glycobiology* 19, 816–28

91 Nett, J.H. *et al.* (2013) Characterization of the *Pichia pastoris* protein-O-mannosyltransferase gene family. *PLoS One* 8,

92 Govindappa, N. *et al.* (2013) PMT1 gene plays a major role in O-mannosylation of insulin precursor in *Pichia pastoris*. *Protein Expr. Purif.* 88, 164–71

93 Kannan, V. *et al.* (2009) A tandem mass spectrometric approach to the identification of O-glycosylated glargine glycoforms in active pharmaceutical ingredient expressed in *Pichia pastoris*. *Rapid Commun. Mass Spectrom.* 23, 1035–42

94 Orlean, P. and Menon, A.K. (2007) Thematic review series: lipid posttranslational modifications. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycophospholipids. *J. Lipid Res.* 48, 993–1011

95 Veiga, A. *et al.* (2003) Cyanide-resistant respiration, a very frequent metabolic pathway in yeasts. *FEMS Yeast Res.* 3, 239–245

96 Bridges, H.R. *et al.* (2009) The respiratory complexes I from the mitochondria of two *Pichia* species. *Biochem. J.* 422, 151–9

97 Bridges, H.R. *et al.* (2010) The subunit composition of mitochondrial NADH:ubiquinone oxidoreductase (complex I) from *Pichia pastoris*. *Mol. Cell. Proteomics* 9, 2318–26

98 Wikström, M. and Hummer, G. (2012) Stoichiometry of proton translocation by respiratory complex I and its mechanistic implications. *Proc. Natl. Acad. Sci. U. S. A.* 109, 4431–6

99 Mazat, J.-P. *et al.* (2013) Mitochondrial energetic metabolism-some general principles. *IUBMB Life* 65, 171–9

100 González-Barroso, M.M. *et al.* (2006) Isolation and bioenergetic characterization of mitochondria from *Pichia pastoris*. *Yeast* 23, 307–13

101 Veiga, A. *et al.* (2003) Energy conversion coupled to cyanide-resistant respiration in the yeasts *Pichia membranifaciens* and *Debaryomyces hansenii*. *FEMS Yeast Res.* 3, 141–148

102 Tarrío, N. *et al.* (2005) The nuclear genes encoding the internal (*KlNDI1*) and external (*KlNDE1*) alternative NAD(P)H:ubiquinone oxidoreductases of mitochondria from *Kluyveromyces lactis*. *Biochim. Biophys. Acta - Bioenerg.* 1707, 199–210

103 Tarrío, N. *et al.* (2006) Characterization of the second external alternative dehydrogenase from mitochondria of the respiratory yeast *Kluyveromyces lactis*. *Biochim. Biophys. Acta - Bioenerg.* 1757, 1476–1484

104 Dias, O. *et al.* (2014) iOD907, the first genome-scale metabolic model for the milk yeast *Kluyveromyces lactis*. *Biotechnol. J.* 9, 776–790

105 Guerra, G. *et al.* (2002) On the $H^+/2e^-$ stoichiometry of the respiratory chain. *Biochem. Mol. Biol. Educ.* 30, 363–367

106 Crofts, A.R. (2004) The Cytochrome bc_1 Complex: Function in the Context of Structure. *Annu. Rev. Physiol.* 66, 689–733

107 Nicholls, D.G. and Ferguson, S.J. (2003) Respiratory chains. In *Bioenergetics* (Nicholls, D. G. and Ferguson, S. J., eds), pp. 89–XIII, Academic Press

108 Veiga, A. *et al.* (2000) Cyanide-resistant respiration is frequent, but confined to yeasts incapable of aerobic fermentation. *FEMS Microbiol. Lett.* 190, 93–97

109 Kern, A. *et al.* (2007) *Pichia pastoris* "just in time" alternative respiration. *Microbiology* 153, 1250–60

110 Petersen, J. *et al.* (2012) Comparison of the H^+/ATP ratios of the H^+-ATP synthases from yeast and from chloroplast. *Proc. Natl. Acad. Sci.* 109, 11150–11155

111 Naumov, G.I. *et al.* (2013) *Komagataella kurtzmanii* sp. nov., a new sibling species of *Komagataella* (*Pichia*) *pastoris* based on multigene sequence analysis. *Antonie Van Leeuwenhoek* 104, 339–347

112 Prielhofer, R. *et al.* (2015) *Pichia pastoris* regulates its gene-specific response to different carbon sources at the transcriptional, rather than the translational, level. *BMC Genomics* 16, 167

113 Watanabe, S. *et al.* (2008) Eukaryotic and Bacterial Gene Clusters Related to an Alternative Pathway of Nonphosphorylated L-Rhamnose Metabolism. *J. Biol. Chem.* 283, 20372–20382

114 Koivistoinen, O.M. *et al.* (2012) Characterisation of the gene cluster for L-rhamnose catabolism in the yeast *Scheffersomyces* (*Pichia*) *stipitis*. *Gene* 492, 177–185

115 Balagurunathan, B. *et al.* (2012) Reconstruction and analysis of a genome-scale metabolic model for *Scheffersomyces stipitis*. *Microb. Cell Fact.* 11, 27

116 Verduyn, C. (1991) Physiology of yeasts in relation to biomass yields. *Antonie Van Leeuwenhoek* 60, 325–353

117 Verduyn, C. *et al.* (1991) A theoretical evaluation of growth yields of yeasts. *Antonie Van Leeuwenhoek* 59, 49–63

118 Förster, J. *et al.* (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 13, 244–53

119 Kurtzman, C.P. (2009) Biotechnological strains of *Komagataella* (*Pichia*) *pastoris* are *Komagataella phaffii* as determined from multigene sequence analysis. *J. Ind. Microbiol. Biotechnol.* 36, 1435–1438

120 Li, P. *et al.* (2015) Construction of efficient xylose utilizing *Pichia pastoris* for industrial enzyme production. *Microb. Cell Fact.* 14, 1–10

121 Cheng, H. *et al.* (2014) Genetically engineered *Pichia pastoris* yeast for conversion of glucose to xylitol by a single-fermentation process. *Appl. Microbiol. Biotechnol.* 98, 3539–3552

122 González-Siso, M.I. *et al.* (2009) Sugar metabolism, redox balance and oxidative stress response in the respiratory yeast *Kluyveromyces lactis*. *Microb. Cell Fact.* 8, 46

123 Tarrío, N. *et al.* (2006) Reoxidation of cytosolic NADPH in *Kluyveromyces lactis*. *FEMS Yeast Res.* 6, 371–380

124 Fonseca, C. *et al.* (2008) Use of in vivo ${}^{13}C$ nuclear magnetic resonance spectroscopy to elucidate L-arabinose metabolism in yeasts. *Appl. Environ. Microbiol.* 74, 1845–1855

125 Wong, B. *et al.* (1993) D-Arabitol metabolism in *Candida albicans*: Studies of the biosynthetic pathway and the gene that encodes NAD-dependent D-arabitol dehydrogenase. *J. Bacteriol.* 175, 6314–6320

126 Gopalakrishnan, S. and Maranas, C.D. (2015) 13C metabolic flux analysis at a genome-scale. *Metab. Eng.* 32, 12–22

127 Cheung, C.Y.M. *et al.* (2013) A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. *Plant J.* 75, 1050–1061

APPENDIX I-A

Additional information discussed in the main text can be found attached as Supporting Information Files in the published article corresponding to this chapter:

Tomàs-Gamisans, M., Ferrer, P., Albiol, J., 2016. **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 11, e0148031.

- **File S3-1. iMT1026 model.** *Pichia pastoris* GSMM model in xlsx format for use in the RAVEN toolbox.

- **File S3-2. Biomass composition.** Details on biomass composition for cultivations on glucose and glycerol:methanol mixtures.

- **File S3-3. Recombinant protein composition.** DNA, RNA and amino acid compositions for the FAB and ROL recombinant proteins expressed in *P. pastoris* and tested in the model.

- **File S3-4. Reaction changes.** List of modified, added and excluded reactions in iMT1026.

- **File S3-5. iMT1026 model in SBML format**. *Pichia pastoris* GSMM model in SBML format generated with the RAVEN toolbox.

This Appendix I-A contains the reaction essentiality analysis in different cultivation conditions (**Fig. S3-1**) in addition to **Fig. 3-3** that summarises the glucose-case under normoxic conditions.

A

B
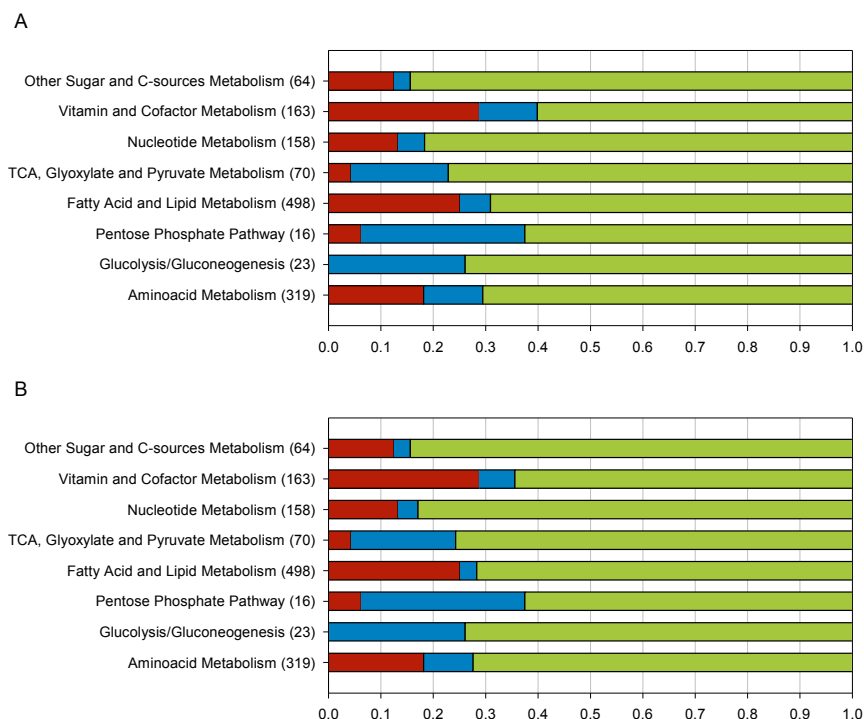
Fig. S3-1. Reaction essentiality analysis in different cultivation conditions. FBA was performed optimizing biomass production and sequentially constraining to 0 each reaction in the corresponding simulations. The resulting growth rate was compared with the wild type growth and reactions were classified in three categories according to the relative growth rate obtained: essential (E), partially essential (PE) and non-essential (NE). Reactions are grouped in 8 major pathways or global subsystems (Y axis). In brackets, the number of reactions included in each subsystem. X axis represent the fraction of each type of reactions in each category of E (in red), PE (in blue) and NE (in green). Reaction essentiality was evaluated with glucose chemostats with different oxygen conditions and glycerol:methanol mixtures chemostats: (A) Glucose and normoxia; (B) Glucose and limited oxygen; (C) Glucose and hypoxia; (D) Glycerol:methanol (80:20 w/w) at μ=0.05 h^{-1}; (E) Glycerol:methanol (60:40 w/w) at μ=0.05 h^{-1}; (F) Glycerol:methanol (40:60 w/w) at μ=0.05 h^{-1}; (G) Glycerol:methanol (80:20 w/w) at μ=0.16 h^{-1}; (H) Glycerol:methanol (60:40 w/w) at μ=0.05 h^{-1}; (I) Glycerol:methanol (40:60 w/w) at μ=0.05 h^{-1}.

C

D

E

F
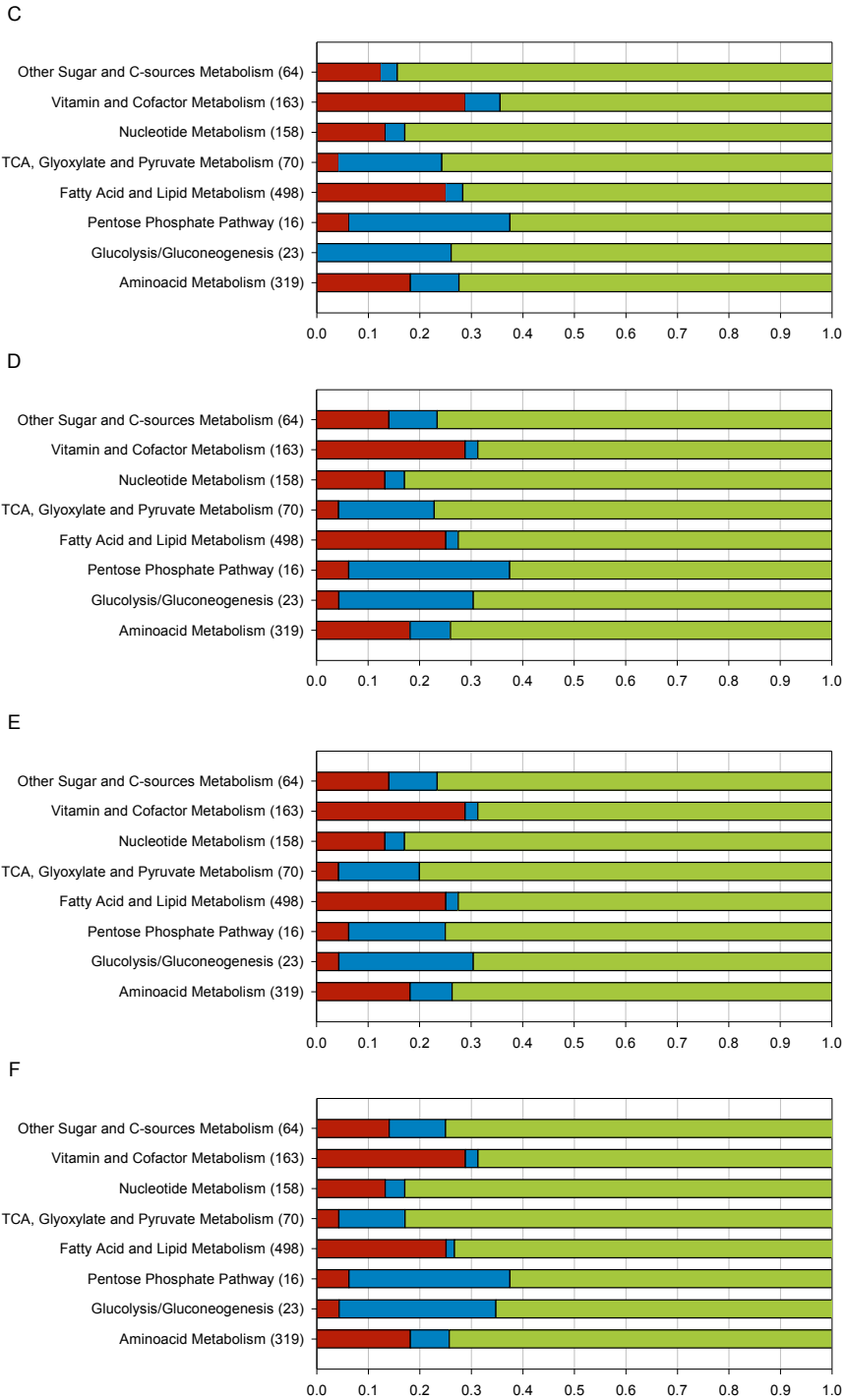
Fig. S3-1. Continued

Models were compared by flux balance analysis performing a maximization of biomass production (x) under the constraints of the experimental variables as described below.

Simulation parameters

Glucose – O_2 levels

Glucose and oxygen uptake rates were constrained according to Carnicer et al. [1]. Ethanol production was also constrained for O_2-limited and hypoxic cases. Only the wild type (wt) strain was evaluated. Biomass growth and CO_2 production were calculated by the model.

Glycerol:methanol mixtures

Only glycerol and methanol uptake rates were constrained for the glycerol and methanol cases according to Jordà et al. [2]. When arabitol is experimentally produced, its measured production rate is set as lower bound for arabitol secretion only for models iMT1026 and iLC915. The other two models, iPP668 and PpaMBEL1254, are not able to generate arabitol.

Simulation results

Glucose – O_2 levels

By setting the normoxic conditions in iLC915, some flux loops occur which were avoided using the following strategy. Reactions involving bidirectional formation of DNA[m], DNA[c], RNA[c] and RNA[m] by a single nucleotide were set to 0. Otherwise reactions such as Diphosphate[c] + DNA[c] <=> dATP[c] could occur without energetic costs. Ethanol is also produced in these conditions. First constrain set of constrains for iLC915 that enable proper growth with glucose (instructions for COBRA Toolbox [3]):

```
iLC915b=changeRxnBounds(iLC915E,{'r66','r910','r1104','r239','r111','r106','r490','r791','r243','r252','r253','r307','r308','r404','r405','r1320','r639','r640','r641','r642','r649','r650','r651','r652','r645','r646','r643','r644','r653','r654','r655','r656','r534'},0,'b').
```

When simulating oxygen restricted conditions, none of the previous models predict arabitol secretion when ethanol production is constrained, and iMT1026 is the only one that predicts this by-product secretion.
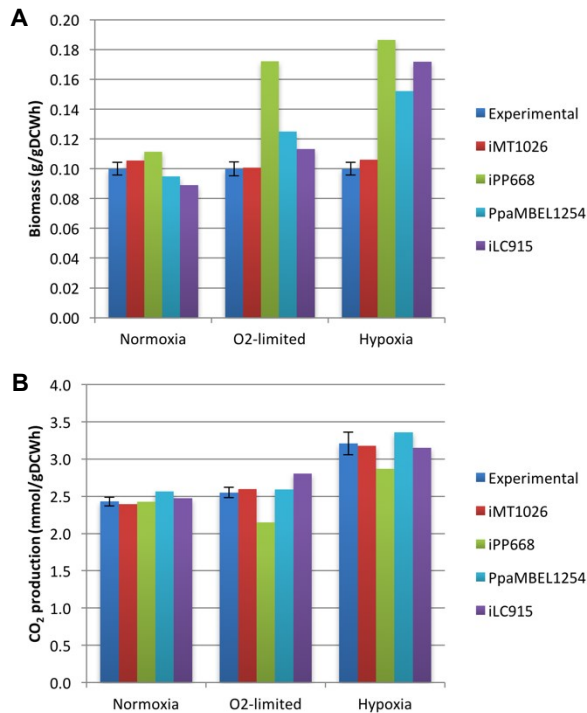
Fig. S3-2. Comparison of predicted values for each model with experimental data on glucose. Graphs with (A) growth rate (B) CO_2 production predictions simulating glucose chemostats at different oxygen conditions with and without recombinant protein production with glucose, O_2 and ethanol fluxes constrained to the experimental values.

Glycerol:methanol mixtures

In order to compare the performance of the models using mixtures of glycerol:methanol as carbon sources, the experimental data corresponding to the strain *P. pastoris* X-33 pGAPZαA_ROL expressesing a *Rhizopus oryzae* lipase were used. Due to the fact that none of the other models incorporate ROL production equations, the simulation using iMT1026 with no ROL production was also performed. Comparing iMT1026 expressing ROL and iMT1026 without the ROL expression, similar values were observed in the evaluated fluxes. Thus, to compare the accuracy of the model using these experimental data, iMT1026 was used setting the expression of ROL to zero. iLC915 has also to be constrained, for the above explained reasons, for the simulation of glycerol:methanol feeding conditions (instructions for COBRA Toolbox [3]):

iLC915=changeRxnBounds(iLC915E,{'r639','r640','r641','r642','r649','r650','r651','r65 2','r645','r646','r643','r644','r653','r654','r655','r656'},0,'b')

PpaMBEL1254 has also to be modified. Reaction for O_2 diffusion to peroxisomes has to be added in order to enable growth and assimilation of methanol. Thus, results showed represent the modified PpaMBEL1254, otherwise no growth was possible.
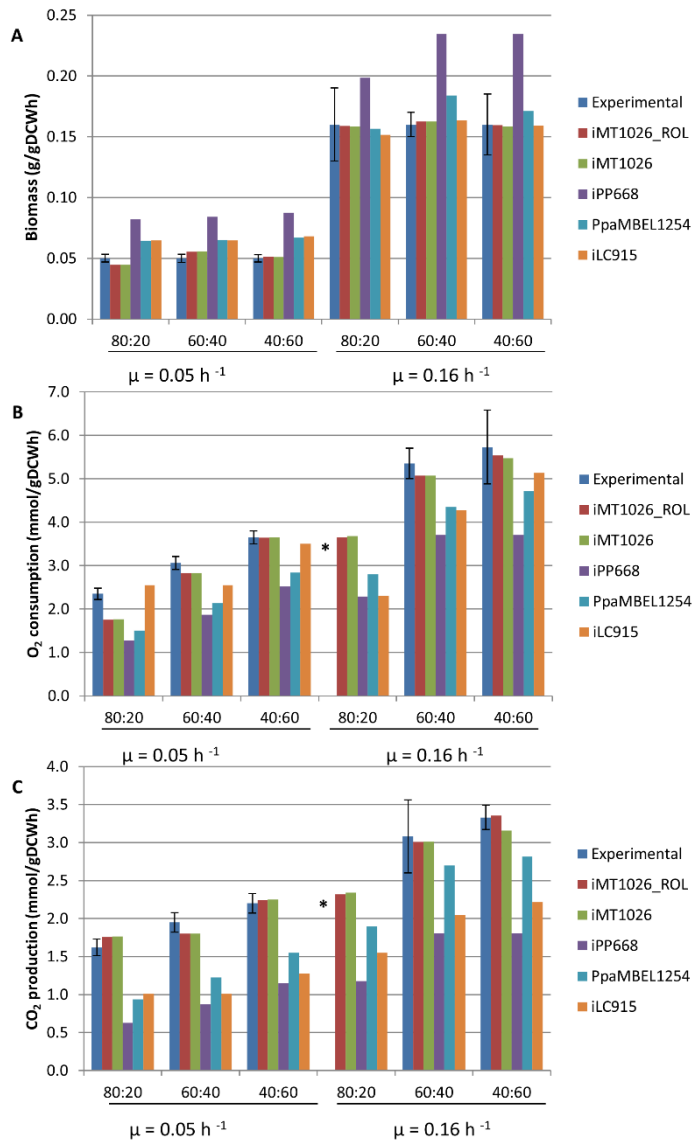
Fig. S3-3. Comparison of predicted values for each model with experimental data on mixtures of glycerol and methanol. Graphs with (A) Growth rate (B) O_2 consumption and (C) CO_2 production predictions simulating different glycerol:methanol chemostats. * Not determined experimentally.

Model evaluation

For evaluating and comparing all the models, a statistical parameter (*eq.* 1) was calculated for all the predicted variables: biomass formation and CO_2 production in the glucose and different O_2 levels case; and biomass formation, O_2 consumption and CO_2 production in the glycerol:methanol simulations.

$$\frac{1}{n} \cdot \sum_{i=1}^{n} \frac{\sqrt{\left(v_{s_i} - v_{e_i}\right)^2}}{v_{e_i}} \%$$

(*eq* 1)

Where v_{s_i} is the resulting flux from the simulation for the variable i and v_{e_i} is the experimental flux for this variable. The total number of predicted and compared fluxes with the experimental values is n and equals to 21.

Table S3-1. Comparison of deviations from experimental values. Calculations were performed according to (eq. 1).

	iMT1026_ROL	iMT1026	iPP668	PpaMBEL1254	iLC915
Glucose		2.71%	32.73%	15.66%	18.30%
Glycerol:methanol	6.90%	7.27%	48.93%	26.40%	25.14%
OVERALL		**6.08%**	**44.70%**	**23.60%**	**23.36%**

References

1. Carnicer M, ten Pierick A, van Dam J, Heijnen JJ, Albiol J, van Gulik W, et al. Quantitative metabolomics analysis of amino acid metabolism in recombinant Pichia pastoris under different oxygen availability conditions. Microb Cell Fact. 2012;11: 83.

2. Jordà J, De Jesus SS, Peltier S, Ferrer P, Albiol J. Metabolic flux analysis of recombinant Pichia pastoris growing on different glycerol/methanol mixtures by iterative fitting of NMR-derived 13C-labelling data from proteinogenic amino acids. N Biotechnol. 2014;31: 120–132.

3. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc. 2011;6: 1290–307.

4

Benchmarking iMT1026 for growth on methanol or glycerol as sole carbon sources

TABLE OF CONTENTS

4. Benchmarking iMT1026 for growth on methanol or glycerol as sole carbon sources

4.1. INTRODUCTION

In the first chapter of the thesis, the development of iMT1026, a consensus genome-scale metabolic model for P. pastoris, is described. This model is an integrative and upgraded version of the previously existing models, validated for P. pastoris growing on glucose under different levels of oxygen availability and expressing an antibody fragment using data previously published by our research group [1,2]. Furthermore, iMT1026 was also validated for growth on mixtures of glycerol and methanol at low and high growth rate [3]. As described in the 'General introduction', industrial bioprocesses are tending to be eco-friendlier, reusing waste streams for novel applications. In this context, glycerol is one of the main by-products in biodiesel synthesis and can be used as carbon source for certain organisms. Thus, developing a glycerol-based biorefinery would allow for glycerol revalorisation, for the production of value-added chemicals, and to improve the viability of biodiesel as alternative biofuel [4]. Commonly, glycerol can be transformed into 1,3-propanediol, succinate, dihydroxyacetone, cell lipids and other value-added products [5,6]. Moreover, due to the residual content of methanol in the crude glycerol [7,8], P. pastoris, as methylotrophic yeast, is an ideal choice for bioprocesses using this carbon source. In fact, glycerol from biodiesel production has already been used as carbon source in P. pastoris [9,10].

Production of chemicals is typically associated to (systems) metabolic engineering strategies for strain engineering in order to enhance production yields. In this context, metabolic models are usually adapted and validated for the specific substrates used in each particular application. Despite the potential of glycerol as carbon source, only a few physiological studies of P. pastoris growing in glycerol as single carbon source can be found [11,12]. Validation of GSMMs for different conditions requires the availability of extensive cultivation data information describing its physiology. In addition, a wide range of information

on biomass composition enables building specific biomass equations to accurately describe cell growth in each case [13]. Given the impact of biomass composition on the model predictive accuracy in a context-dependent manner [13], iMT1026 model was still not suitable for describing growth on glycerol or methanol as single carbon sources. This is because biomass composition equations take a major role on prediction reliability, and small changes in that composition, or using an inadequate biomass equation, may rend model calculations inaccurate [13]. Hence, GSMMs are in continuous evolution (e.g. for Saccharomyces cerevisiae [14]) usually involving error-fixing steps related to poor or wrong gene annotation [15], or extending GSMM capabilities for a broader range of cultivation conditions.

In this work, we expand the iMT1026 model capabilities by implementing the capacity to accurately describe P. pastoris growth phenotype when using glycerol or methanol as sole carbon sources.

A series of chemostat cultures were performed at a wide range of growth rates using glycerol or methanol as sole carbon sources in order to provide comprehensive physiological datasets needed to upgrade the model. This included analysis of the elemental and macromolecular biomass composition for each tested growth condition. This allowed to introduce new biomass reaction equations to the metabolic model specific for growth on glycerol or methanol. Furthermore, the new version of the model (v3.0) was validated for growth on these two substrates within the tested growth rate range. This way, a tool for further developing the glycerol-based biorefinery concept in P. pastoris is provided.

.

4.2. MATERIALS AND METHODS

4.2.1. Strain and cultivation conditions

Pichia pastoris wildtype X-33 (Invitrogen) was cultivated in carbon-source limiting chemostat cultures at a range of dilution rates. Continuous cultures were performed at a working volume of 1 L in a 2-L benchtop bioreactor Biostat B (Sartorius AG, Göttingen, Germany) for glycerol cultures, and in a Biostat B+ (Sartorius AG, Göttingen, Germany) for methanol cultivations. Two independent chemostat series were performed for each carbon source of increasing dilution rates (D) of 0.035, 0.050, 0.065, 0.100, 0.130 and 0.160 h^{-1} for glycerol and 0.035, 0.050, 0.065, 0.080, 0.100 and 0.130 h^{-1} for methanol. For pre-culture, 150 mL of YPG media (2% (w/v) peptone, 1% (w/ v) yeast extract, and 2% (w/v) glycerol) in 1-L shake flasks were inoculated with a cryostock at an initial OD_{600} of $0.15 - 0.30$ and incubated at 150 rpm and 25°C (Infors HT Multitron, Bottmingen, Switzerland) for approximately 24 h. Cells were centrifuged and resuspended in sterile demineralised water and used to inoculate the bioreactor for the batch phase. Once the batch phase was concluded, chemostat phase was initiated at the specific growth rate by appropriately setting the corresponding inlet flow and enabling outlet flow to keep the reactor volume constant to 1 L. Both for batch and chemostat culture, stirring was set to 700 rpm, aeration rate to 1 vvm, temperature was maintained at 25°C and pH 5.0 automatically controlled with 15% ammonia. The off-gases were cooled dawn in a condenser at 4°C and further desiccated in two silica gel columns. For the glycerol cultures, off-gas CO_2 and O_2 fractions were analysed through BCP-CO_2 and BCP-O_2 Sensors (BlueSens gas sensor GmbH, Herten, Germany). On the other hand, methanol off-gas composition was analysed by means of a mass spectrometer Omnistar™ 300 02 (Balzers Instruments, Balzers, Liechtenstein). Each dilution rate was kept for at least 5 residence times, and 3 culture samples were taken along the last volume change.

Batch medium composition was previously described by Baumann et al. [16]. Chemostat medium composition was also taken from [16], except that glucose was replaced by glycerol or methanol as carbon source. Thus, briefly chemostat medium contained per liter: 50 g carbon source (glycerol or methanol), 0.9 g citric acid, 4.35 g (NH4)$_2$HPO$_4$, 0.01 g CaCl$_2$ · 2H$_2$O, 1.7 g KCl, 0.65 g MgSO$_4$ · 7H$_2$O, 1 mL Biotin (0.2 g L^{-1}), and 1.6mL PTM1 trace

salts stock solution (prepared as described previously [16]). pH was adjusted to 5.0 with 25% HCl.

4.2.2. Analytical methods

Extracellular metabolite quantification

Glycerol, methanol, arabitol, succinate, acetate and ethanol were analysed by HPLC. Triplicate samples (2 mL each) were centrifuged at 12,000 rpm for 2 min (Minispin, Eppendorf, Hamburg, Germany). The supernatant was collected and filtered through 0.45 μm nitrocellulose membrane filters (Millipore). Duplicate samples were analysed by HPLC (HP 1050 liquid chromatograph, Dionex Corporation, Sunnyvale, CA, USA) using an ICSep ICE COREGEL 87H3 column (Transgenomic Inc., Omaha, NE, USA). The mobile phase was 8 mM sulphuric acid. Injection volume was 20 μL. Data was quantified by Chromeleon 6.80 Software (Dionex Corporation, Sunnyvale, CA, USA). Average relative standard deviation (RSD) of the analysis was about 1%.

Biomass quantification

Biomass in culture broth was monitored during cultivation by measuring the optical density at 600 nm. Dry cell weight (DCW) was quantified accordingly to the method described in Jordà *et al.* (2012). Biomass concentration was determined in triplicate. Biomass concentration average RSD was about 2%.

Biomass composition analysis

Both for the glycerol and methanol cultivations, biomass composition was analysed at the following growth rates: 0.035, 0.065, 0.100 h^{-1}. Additionally, for the glycerol cultivations, biomass analyses were also carried out at 0.160 h^{-1}.

Elemental analysis.

C, H, N, S were analysed by combustion at 1200°C and subsequent gas chromatography in a Flash 2000 Elemental Analyzer (ThermoFisher Scientific, USA). Oxygen was determined through an oxygen-specific pyrolysis at 1060°C. Ash content was determined by subtraction of the C, H, N, O, S fractions as remaining component.

Amino acid analysis.

15 mg of lyophilised biomass were hydrolysed with 6M HCl for 24 h at 105°C. Then, deionised water (MiliQ) was added up to complete 50 mL. Filtered aliquots were vacuum dried and finally resuspended in water. Samples were then derivatised wit 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate according to AccQ-Tag method (Waters, Milford, USA). Derivatised amino acids were analysed with a Waters 2487 (Waters, Milford, USA) UV detector at 254 nm in a gradient system HPLC Waters 600 (Waters, Milford, USA).

Biomass samples for the determination of total protein, carbohydrates, DNA and RNA content were prepared and analysed as described in Carnicer et al. [2].

4.2.3. Statistical analysis

Standard reconciliation procedures [18,19] were applied to elemental composition and major macromolecular components (proteins, carbohydrates, DNA and RNA). The resulting biomass elemental composition was subsequently used to check chemostat cultivation data reconciliation and consistency [20]. Both for biomass macromolecular and elemental composition and for chemostat data, statistical consistency test was passed with a confidence level of 95%. Consequently, there was no evidence for gross measurement errors.

Global macromolecular, amino acid and lipid composition data was analysed with statistical tests available in Microsoft Excel. 2-tailed Student's *t*-Test was used to determine statistically significant differences in biomass composition between carbon source and growth rates.

4.2.4. Modelling

Pichia pastoris iMT1026 v2.0 [21] updated at BioModels database [22] ID: MODEL1508040001 (**File S4-1**) was used as starting model for further updating. The model was edited incorporating new biomass equations using standard scripts from COBRA Toolbox v2.0.6 [23]. **File S4-3** includes the COBRA commands necessary to add these new equations into the existing model. All simulations were carried out with the COBRA Toolbox v2.0.6 under Matlab 2014 (Mathworks, USA) with SBML toolbox v4.1.0 [24] and libSBML library v5.12.0 [25]. Flux balance analysis (FBA) with linear optimization was used to predict

metabolic phenotypes by setting the appropriate flux constraints. To test model accuracy, biomass production was constrained to growth rate and the absolute value of substrate uptake rate was minimised. iMT1026 v3.0 was validated in SBML format and submitted to BioModels database with the ID: **MODEL1612130000**. This model is also available in **File S4-2**.

4.2.5. Energetic parameters calculation

ATP requirement for cellular maintenance was determined by the following energetic parameters estimation procedure. These requirements were divided into growth-associated maintenance-energy (GAME) and non-growth associated maintenance-energy (NGAME). For NGAME calculation, the substrate uptake rate was represented against the growth rate (μ) according to Pirt's equation [26]. In the *y*-intercept of this linear regression, ATP turnover was maximised ($\mu = 0$). These ATP values (for glycerol and methanol) are set as lower bounds in 'ATPM' reaction, representing NGAME.

Using the obtained values for NGAME, GAME was determined by adjusting ATP stoichiometric coefficient in the corresponding biomass equation to fit biomass-substrate yields according to the experimental data. These simulations were carried out by maximising the biomass production in a FBA, constraining the substrate uptake rate according to the experimental data.

4.3. RESULTS AND DISCUSSION

4.3.1. Physiological macroscopic parameters

P. pastoris X-33 strain was cultivated in carbon-limited chemostat cultures at different dilution rates to characterise its physiology using different carbon sources. This information was used to estimate the energetic parameters and to calibrate the model for such carbon sources. Glycerol cultivations were carried out at different dilution rates (D): 0.035, 0.050, 0.065, 0.100, 0.130 and 0.160 h^{-1}. At 0.160 h^{-1} the inflowing gas was supplied with an air:O_2 mixture (92.5:7.5) due to the higher O_2 demand and cell concentration. Due to this operational limitation, no higher dilution rates were tested, despite *P. pastoris* has been reported to grow at higher growth rates [27]. Methanol limiting chemostats were run at 0.035, 0.050, 0.065,

0.080, 0.100 and 0.130 h^{-1}. At 0.130 h^{-1} bioreactor washed out. Biomass concentration, CO_2 production and O_2 consumption continuously decreased, and methanol accumulated. According to a chemostat washout kinetics [28], maximum growth rate on methanol was estimated to be between 0.11 and 0.12 h^{-1}, which is in agreement with previously reported values [29].

Based on the chemostat data, specific productivities and yields were calculated for each condition (**Table 4-1**). In both glycerol and methanol cultivation series, main growth parameters show a linear correlation with growth rate (μ). Regarding biomass yields (Y_{XS}), there is a slight decrease at lower growth rates on both carbon sources, similarly as reported by Van Dijken et al. [30] and Rebnegger et al. [31]. These points at a carbon redirection towards biomass production occurring at higher growth rates at expenses to of CO_2 production, as reflected in the respiratory quotient (RQ). As shown in **Table 4-1**, there is a modest decrease (< 10 %) of RQ at higher growth rates, especially in the glycerol case in relation to the lower growth rates. Despite this apparent correlation, there are no statistically significant differences within the tested range, and average Y_{XS} and RQ can be calculated for the abovementioned range of growth rates. Average Y_{XS} in methanol is 0.40 $g_X \cdot g_S^{-1}$ and is in accordance with yields previously reported for *P. pastoris* and other yeast [32]. This value is considerably lower than 0.67 $g_X \cdot g_S^{-1}$, the average Y_{xs} for glycerol. The Y_{XS} for glycerol ranged between 0.62 – 0.71 $g_X \cdot g_S^{-1}$, similar to yields on this substrate reported for different *Pichia* species and other yeasts [33].

4.3.2. Macromolecular and elemental biomass composition

Growth rate dependent stoichiometry

To investigate the potential impact of growth rate on biomass composition, samples of the cultures were taken for analysis of the biomass elemental and macromolecular composition at different dilution rates. In particular, we analysed the biomass composition at four different growth rates for glycerol (μ = 0.035, 0.065, 0.100 and 0.160 h^{-1}) and three for methanol (μ = 0.035, 0.065 and 0.100 h^{-1}). The experimental datasets and the calculated (reconciled) biomass composition are summarised in **Fig. 4-1** (and **Table S4-2** for full details).

Table 4-1. Macroscopic growth parameters after the reconciliation procedure for glycerol and methanol cultivations at different growth rates.

C-source	μ_{SP}	μ_{exp}	q_S	q_{O2}	q_{CO2}	q_X	Y_{XS}	RQ
Glycerol	0.035	0.035 ± 0.001	-0.58 ± 0.05	-0.82 ± 0.13	0.53 ± 0.11	1.22 ± 0.04	0.65 ± 0.03	0.64 ± 0.03
	0.050	0.049 ± 0.002	-0.85 ± 0.06	-1.26 ± 0.10	0.84 ± 0.08	1.70 ± 0.14	0.62 ± 0.05	0.67 ± 0.07
	0.065	0.064 ± 0.000	-1.07 ± 0.01	-1.52 ± 0.02	1.00 ± 0.02	2.22 ± 0.02	0.64 ± 0.00	0.65 ± 0.00
	0.100	0.094 ± 0.004	-1.52 ± 0.08	-2.04 ± 0.11	1.28 ± 0.08	3.27 ± 0.15	0.71 ± 0.04	0.63 ± 0.01
	0.130	0.124 ± 0.001	-1.92 ± 0.08	-2.36 ± 0.15	1.41 ± 0.13	4.36 ± 0.24	0.71 ± 0.04	0.60 ± 0.07
	0.160	0.154 ± 0.002	-2.41 ± 0.03	-2.92 ± 0.01	1.74 ± 0.00	5.47 ± 0.08	0.70 ± 0.00	0.60 ± 0.00
	Average						0.67 ± 0.04	0.63 ± 0.03
Methanol	0.035	0.035 ± 0.001	-2.81 ± 0.16	-2.98 ± 0.22	1.59 ± 0.14	1.22 ± 0.02	0.38 ± 0.01	0.53 ± 0.01
	0.050	0.049 ± 0.000	-3.88 ± 0.10	-4.07 ± 0.15	2.15 ± 0.10	1.73 ± 0.01	0.39 ± 0.01	0.53 ± 0.01
	0.065	0.065 ± 0.001	-4.87 ± 0.22	-4.97 ± 0.29	2.55 ± 0.18	2.33 ± 0.04	0.41 ± 0.01	0.51 ± 0.01
	0.080	0.084 ± 0.001	-6.23 ± 0.16	-6.36 ± 0.18	3.27 ± 0.12	2.96 ± 0.13	0.42 ± 0.02	0.51 ± 0.02
	0.100	0.099 ± 0.001	-7.82 ± 0.28	-8.22 ± 0.37	4.34 ± 0.24	3.47 ± 0.04	0.40 ± 0.01	0.53 ± 0.00
	Average						0.40 ± 0.01	0.52 ± 0.01

μ_{SP} (h^{-1}) corresponds to the set point growth rate and μ_{exp} (h^{-1}), the measured experimental μ. q_S, q_{O2} and q_{CO2} are expressed in mmol · g_{DCW}^{-1}· h^{-1}. Units for q_X are Cmmol · g_{DCW}^{-1}· h^{-1}. Y_{XS} represents biomass yield (g_X · g_S^{-1}) RQ is respiratory quotient.

Notably, the protein and RNA fractions positively correlate with growth rate in both glycerol- and methanol-fed cultivation series. This increment on protein and RNA with increasing growth rates is at expenses of carbohydrate content. This trade-off between RNA-protein and carbohydrate content has been widely reported in yeast species [34,35], including in *P. pastoris* [3]. The increase in protein fraction is consistent with the measured changes in the elemental composition: the nitrogen content is also higher at high growth rates (**Table S4-1**). Nonetheless, only the correlation of RNA and growth rate is statistically significant. This stoichiometric dependence of biomass components on growth can be described on the basis of the growth rate hypothesis (GRH). Essentially, GRH attributes this shift to the tight control of the expensive protein synthesis system [36]. At higher growth rates, cells need a higher ribosomal content to maintain the enzymatic machinery. The ribosomes are reported to consist of 53% RNA and 47% protein in *Aspergillus niger* [37]. Thus, the increase in ribosome levels has a deep impact in overall cell protein and RNA content.

Carbon source effects on biomass composition

Besides the impact of the specific growth rate on biomass composition described above, other factors like the carbon source are also known to have a significant influence [3]. In our case, the effect of the carbon source can be appreciated in **Fig. 4-1**: cells grown on methanol show a significantly higher protein fraction than those grown on glycerol. Jordà et al. [3] cultivated *P. pastoris* in chemostats using different ratios of glycerol:methanol mixtures as carbon source. The corresponding biomass composition analyses showed that protein content increased at higher methanol/glycerol ratios. Similarly, *P. pastoris* cells growing on a glucose:methanol mix in a chemostat show higher protein content than when growing on glucose as a sole carbon source under analogous conditions [17]. Consequently, the increase in cell protein content seems to be directly related to methanol-utilisation, and more specifically, to the amount of enzymes needed for methanol assimilation [38]. In fact, genes encoding for the methanol utilisation pathway such as the alcohol oxidase (AOX) and dihydroxyacetone synthase (DHAS), two major enzymes involved in the initial steps of methanol metabolism, are highly induced in the presence of methanol [38]. They are reported to account for up to 10-20% of total protein in methylotrophic yeasts [30,39].
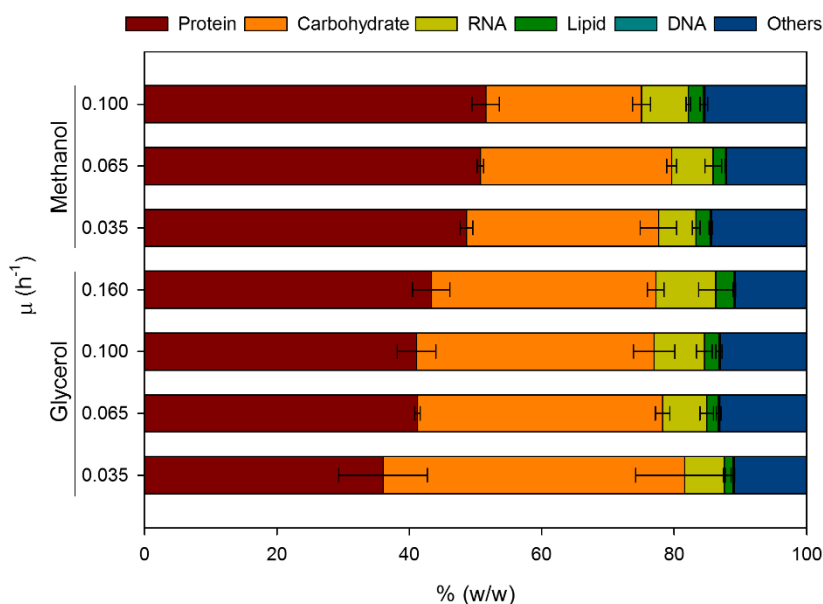
Fig. 4-1. Comparison of the reconciled macromolecular composition of glycerol and methanol cultures at different growth rates.

On the other hand, amino acid composition analysis of the cell proteome showed no significant differences when comparing growth rates of cells grown each substrate **Table S4-2**. Nevertheless, the amino acid composition of biomass differed significantly for some amino acids when comparing glycerol versus methanol grown cells (**Table S4-2**). In addition, the subset of amino acids showing significant differences of relative abundances in methanol-grown cells (compared to the glycerol condition) was compared to the amino acid composition of enzymes related to the methanol metabolisation pathway (**Fig. 4-2**). This analysis clearly reveals how the amino acid composition of the methanol metabolisation enzymes affects the overall cell amino acid composition with respect to glycerol. Therefore, the higher protein fraction in biomass composition in methanol appears to be related to the increased content of methanol-assimilating enzymes.

In terms of cell total lipid fraction, no statistically significant differences are found across the series of methanol biomass samples collected at different growth rates. Conversely, the cell lipid fraction shows a statistically significant positive correlation (*P*-value < 0.05) with the growth rate of glycerol-grown cells. In addition, the lipid content in glycerol-grown cells is on average 1.25-fold higher than in methanol-grown cells. Regarding the lipid composition

profile, there are significant differences depending on the carbon source (**Fig. 4-3**, and additional details in **Table S4-3**). Specifically, these differences are found in triacylglycerols (TAG), free fatty acids (FFA) and phosphatidic acid (PA). There is a higher content of TAG and PA at expenses of FFA in glycerol-grown cells, whereas in methanol-grown cells, FFA is the major lipid fraction and TAG and PA are present only in trace amounts. Both TAG and PA are lipid molecules with a glycerol backbone. Therefore, these differences seem to reflect the lower synthesis cost of TAG and PA from glycerol.
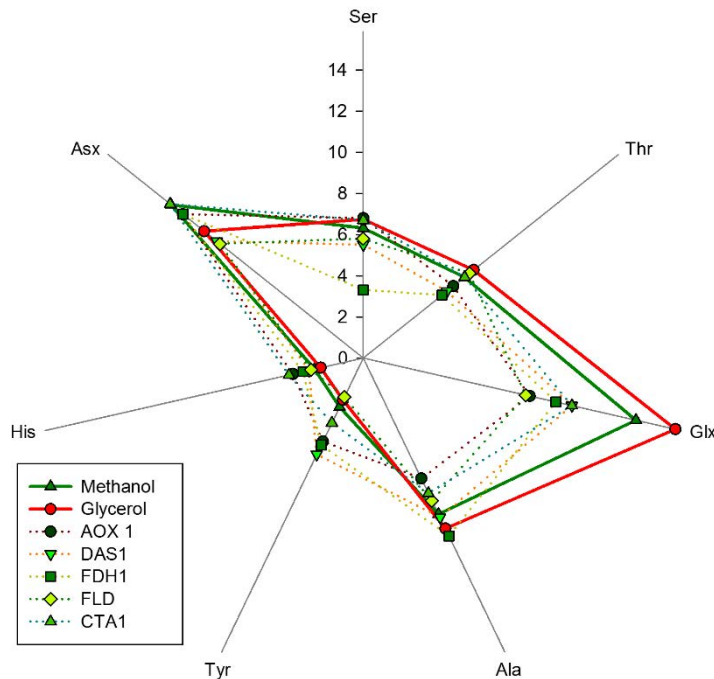
Fig. 4-2 Comparison of average amino acid profiles from glycerol and methanol cultures in relation to amino acid abundance in the most abundant proteins in methanol metabolisation. Amino acid abundance is presented as mol/mol %. Glycerol and methanol represent the average amino acid composition of glycerol and methanol cultivations respectively. Other variables correspond to the most abundant proteins in the presence of methanol: alcohol oxidase (AOX1), dihydroxyacetone synthase (DAS1), formate dehydrogenase (FDH1), formaldehyde dehydrogenase (FLD), catalase (CTA1). Glx and Asx represent the pair of Asp/Asn and Glu/Gln respectively.

When formulating a biomass equation for glycerol and methanol growth conditions, despite that certain biomass components appear to be correlated with biomass specific growth rate, statistical analyses do not show significant differences associated with growth rate. In contrast, statistically significant differences are found when comparing average glycerol and methanol biomass compositions. Consequently, new biomass equations have been formulated

for growth on glycerol and methanol incorporating protein, lipid and macromolecular composition equations specific to each carbon source growth condition.
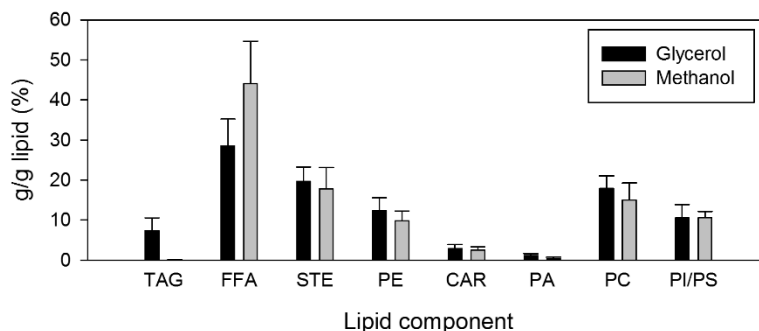
Fig. 4-3. Average lipid profile for biomass grown on glycerol (black) and methanol (gray). Triacylglycerols (TAG), free fatty acids (FFA), sterols (STE), cardiolipin (CAR), phosphatidic acid (PA), phosphatidylcholine (PC) and phosphatidylinositol / phosphatidylserine (PI/PS). Detailed composition of lipids can be found in **Table S4-1**

4.3.3. Energetic parameters estimation

Prior to model validation, energetic parameters have to be estimated in order to assure accurate predictions of cell performance. These parameters are the growth associated and the non-growth associated maintenance energy (GAME and NGAME, respectively). NGAME, values differed significantly for glycerol and methanol growth conditions. On the one hand, growth on glycerol showed a NGAME of 2.51 mmol ATP $\cdot$ g_{DCW}^{-1} $\cdot$ h^{-1}, which is similar to the corresponding value previously calculated for glucose growth conditions, 2.81 mmol ATP $\cdot$ g_{DCW}^{-1} $\cdot$ h^{-1} [21,40]. In contrast, the NGAME calculated for methanol growth is 0.44 mmol ATP $\cdot$ g_{DCW}^{-1} $\cdot$ h^{-1}, i.e. much lower compared to the corresponding values calculated for the other carbon sources.

For GAME estimation for growth on glycerol, physiological parameters corresponding to the $\mu = 0.035$ h^{-1} condition were not considered, as a metabolic shift seems to change the phenotypic profile at this (and lower) growth rates [31]. This can be directly inferred from the specific CO_2 production rate (q_{CO2}) and specific O_2 consumption rate (q_{O2}) observed at this growth rate, which do not follow the same linear trend as in the rest of measured range (**Fig. 4-4**). Hence, taking into account this consideration, GAME for glycerol was estimated to be 70.66 mmol ATP $\cdot$ g_{DCW}^{-1}, that is, 2.4-fold lower than for methanol (166.77 mmol ATP $\cdot g_{DCW}^{-1}$). As mentioned above, there is an important change in protein composition in methanol-

grown cells compared to glycerol growth due to the high levels of enzymes associated to methanol metabolisation. The metabolic overload resulting from the maintenance of this cell machinery could be one of the reasons for the higher GAME besides the fact that growth on highly reduced substrates like methanol (reduction degree (RD) of 6) is known to be usually less efficient (higher energy dissipation and lower biomass yields) compared to glycerol (RD 4.67) or glucose (RD 4) [41]. When compared to glucose culture conditions, GAME for glycerol growth is very similar to the 72 mmol ATP $\cdot$ g_{DCW}^{-1} calculated for glucose growth in the previous chapter [21].
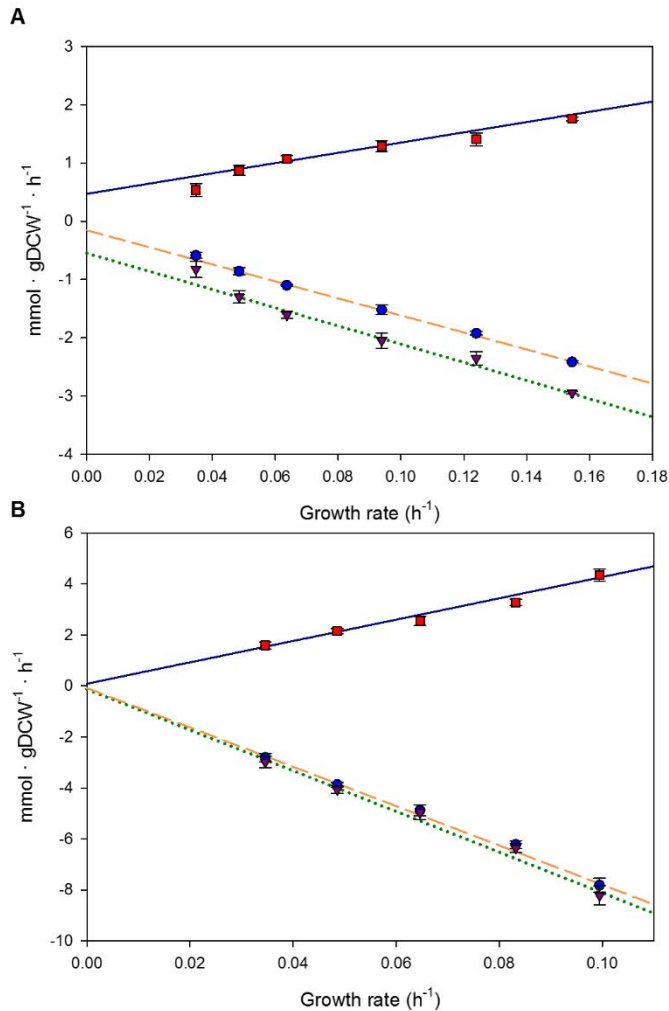
Fig. 4-4 Evaluation of simulated and experimental macroscopic variables for the growth in glycerol and methanol. For each carbon source, the growth rate was constrained and the absolute value of substrate uptake rate was minimised. (A) Chemostats on glycerol; (B) Chemostats on methanol. $q_{S,glycerol/methanol}$ ($\bullet$), q_{CO_2} ($\blacksquare$), q_{O_2} ($\blacktriangledown$); predicted $q_{S,glycerol/methanol}$ (dashed line), predicted q_{CO_2} (continuous line), predicted q_{O_2} (dotted line).

4.3.4. Model validation

The updated model, iMT1026 v3.0 (**File S4-2**) and available at BioModels Database with model ID **MODEL1612130000**), integrating the new specific biomass equations for growth on glycerol and methanol as sole carbon sources, was used to estimate the main macroscopic growth parameters as described in the Experimental procedures section. The version 3.0 of iMT1026 accurately predicts macroscopic growth parameters within the range of tested growth rates for both carbon sources (**Fig. 4-4**). In comparison to iMT1026 v2.0, this new version improves the accuracy in the main macroscopic parameters prediction for glycerol- or methanol-grown biomass (**Fig. S4-1**).

Despite the great overall performance, model deviates from the experimental data by overestimating q_{O_2} and q_{CO_2} in the case of glycerol growth at 0.035 h^{-1} (**Fig. 4-4**). *P. pastoris* has been reported to reduce maintenance energy requirements at very low growth rates associated with metabolic adaptations and changes in gene expression [31,40]. In order to take into account this lower maintenance energy requirement, a series of additional simulations were carried out by constraining the NGAME at values lower than 2.90 mmol ATP $g_{DCW}^{-1} \cdot$ h^{-1} (i.e. the default value set for glycerol-grown cells) and maximising growth at a given substrate uptake rate. As a result, iMT1026 v3.0 is also able to accurately predict the main macroscopic growth parameters for glycerol growth at 0.035 h^{-1} when NGAME is lowered (**Fig. S4-2**). In particular, values between 1 and 1.5 mmol ATP $g_{DCW}^{-1} \cdot$ h^{-1} allow the best accuracy in predicting the experimental data at 0.035h^{-1}, as shown in Figure S2 According to these calculations, there is between a 2 to 3-fold reduction of NGAME at the lower growth rate range. Notably, these results are in agreement with Rebnegger et al. [31], who reported 3-fold reduction of the maintenance requirements at low growth rates.

4.4. CONCLUSIONS

In this study, we analysed the performance of *P. pastoris* growing in chemostat cultures using glycerol or methanol as single carbon source over a wide range of growth rates. The observed biomass composition changes in terms of protein and RNA content as a function of growth rate further support the growth-rate effect hypothesis on biomass composition, i.e. for both carbon sources, higher content of protein and RNA was observed at higher growth rates. Moreover, biomass composition also showed a strong dependence on carbon source, as protein content in biomass was higher in methanol-grown cells. In addition, the carbon source has a significant impact on lipid and amino acid profiles.

Overall, the information gathered on biomass composition at different growth rates and carbon sources allowed to calculate average biomass compositions for glycerol- and methanol-grown biomass. This allowed us to extend the iMT1026 model with new biomass equations for growth on glycerol or methanol as sole carbon sources. Energetic maintenance requirements were estimated for the first time in *P. pastoris* in both carbon sources. Furthermore, the model was validated for the range of growth rates tested and it accurately described the experimental physiological data. Minor discrepancies between experimental data and simulations were found for glycerol at lower growth rates, where a non-linear behaviour of growth parameteres has been reported due to a metabolic shift on metabolism that enables *P. pastoris* to reduce its maintenance energy requirements. Such discrepancies can be easily taken into account by decreasing the value of maintenance energy requirements included in the model.

In summary, the third version of iMT1026, v3.0, consensus model for *P. pastoris* provides to the scientific community an improved metabolic engineering and analysis tool with expanded capabilities for predicting the metabolic phenotype in a broader range of conditions as well as an improved tool for future design of model-based metabolic engineering of the *P. pastoris* cell factory.

4.5. REFERENCES

1   Carnicer, M. *et al.* (2012) Quantitative metabolomics analysis of amino acid metabolism in recombinant *Pichia pastoris* under different oxygen availability conditions. *Microb. Cell Fact.* 11, 83

2   Carnicer, M. *et al.* (2009) Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels. *Microb. Cell Fact.* 8, 65

3   Jordà, J. *et al.* (2014) Metabolic flux analysis of recombinant *Pichia pastoris* growing on different glycerol/methanol mixtures by iterative fitting of NMR-derived 13C-labelling data from proteinogenic amino acids. *N. Biotechnol.* 31, 120–132

4   Garlapati, V.K. *et al.* (2016) Bioconversion technologies of crude glycerol to value added industrial products. *Biotechnol. Reports* 9, 9–14

5   da Silva, G.P. *et al.* (2009) Glycerol: A promising and abundant carbon source for industrial microbiology. *Biotechnol. Adv.* 27, 30–39

6   Litsanov, B. *et al.* (2013) Glycerol as a substrate for aerobic succinate production in minimal medium with Corynebacterium glutamicum. *Microb. Biotechnol.* 6, 189–195

7   Valerio, O. *et al.* (2015) Improved utilization of crude glycerol from biodiesel industries: Synthesis and characterization of sustainable biobased polyesters. *Ind. Crops Prod.* 78, 141–147

8   Posada, J. a *et al.* (2012) Design and analysis of biorefineries based on raw glycerol: addressing the glycerol problem. *Bioresour. Technol.* 111, 282–93

9   Çelik, E. *et al.* (2008) Use of Biodiesel Byproduct Crude Glycerol as the Carbon Source for Fermentation Processes by Recombinant *Pichia pastoris*. *Ind. Eng. Chem. Res.* 47, 2985–2990

10 Tang, S. *et al.* (2009) *Pichia pastoris* fermentation for phytase production using crude glycerol from biodiesel production as the sole carbon source. *Biochem. Eng. J.* 43, 157–162

11 Solà, A. *et al.* (2004) Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. *Eur. J. Biochem.* 271, 2462–70

12 Prielhofer, R. *et al.* (2015) *Pichia pastoris* regulates its gene-specific response to different carbon sources at the transcriptional, rather than the translational, level. *BMC Genomics* 16, 167

13 Dikicioglu, D. *et al.* (2015) Biomass composition: the "elephant in the room" of metabolic modelling. *Metabolomics* 11, 1690–1701

14 Aung, H.W. *et al.* (2013) Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism. *Ind. Biotechnol.* 9, 215–228

15 Dikicioglu, D. *et al.* (2014) Improving functional annotation for industrial microbes: a case study with *Pichia pastoris*. *Trends Biotechnol.* 32, 396–9

16 Baumann, K. *et al.* (2008) Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins. *Biotechnol. Bioeng.* 100, 177–183

17 Jordà, J. *et al.* (2012) Metabolic flux profiling of recombinant protein secreting *Pichia pastoris* growing on glucose:methanol mixtures. *Microb. Cell Fact.* 11, 57

18 Lange, H.C. and Heijnen, J.J. (2001) Statistical reconciliation of the elemental and molecular biomass composition of *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 75, 334–344

19 Verheijen, P. (2010) Data Reconciliation and Error Detection. In *The Metabolic Pathway Engineering Handbook* (Smolke, C. D., ed), pp. 8.1-8.13, CRC Press

20 Noorman, H.J. *et al.* (2000) Classification, error detection, and reconciliation of process information in complex biochemical systems. *Biotechnol. Bioeng.* 49, 364–376

21 Tomàs-Gamisans, M. *et al.* (2016) 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* 11, e0148031

22 Chelliah, V. *et al.* (2015) BioModels: ten-year anniversary. *Nucleic Acids Res.* 43, D542–D548

23 Schellenberger, J. *et al.* (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* 6, 1290–307

24 Keating, S.M. *et al.* (2006) SBML.Toolbox: An SBML toolbox for MATLAB users. *Bioinformatics* 22, 1275–1277

25 Bornstein, B.J. *et al.* (2008) LibSBML: An API library for SBML. *Bioinformatics* 24, 880–881

26 Pirt, S.J. (1982) Maintenance energy: a general model for energy-limited and energy-sufficient growth. *Arch. Microbiol.* 133, 300–302

27 Cos, O. *et al.* (2006) Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: a review. *Microb. Cell Fact.* 5, 17

28 Doran, P.M. (1995) Reactor Engineering. In *Bioprocess Engineering Principles* pp. 333–391, Elsevier

29 Barrigon, J.M. *et al.* (2015) A macrokinetic model-based comparative meta-analysis of recombinant protein production by *Pichia pastoris* under *AOX1* promoter. *Biotechnol. Bioeng.* 112, 1132–1145

30 Van Dijken, L.P. *et al.* (1976) Growth of *Hansenula polymorpha* in a methanol-limited chemostat. *Arch. Microbiol.* 111, 137–144

31 Rebnegger, C. *et al.* (2014) In *Pichia pastoris*, growth rate regulates protein synthesis and secretion, mating and stress response. *Biotechnol. J.* 9, 511–25

32 Hazeu, W. and Donker, R.A. (1983) A continuous culture study of methanol and formate utilization by the yeast *Pichia pastoris*. *Biotechnol. Lett.* 5, 399–404

33 Taccari, M. *et al.* (2012) Screening of yeasts for growth on crude glycerol and optimization of biomass production. *Bioresour. Technol.* 110, 488–495

34 Verduyn, C. *et al.* (1990) Energetics of *Saccharomyces cerevisiae* in Anaerobic Glucose-Limited Chemostat Cultures. *J. Gen. Microbiol.* 136, 405–412

35 Verduyn, C. (1991) Physiology of yeasts in relation to biomass yields. *Antonie Van Leeuwenhoek* 60, 325–353

36 Henriksen, C.M. *et al.* (1996) Growth energetics and metabolic fluxes in continuous cultures of *Penicillium chrysogenum*. *J. Biotechnol.* 45, 149–164

37 Hangeraaf, P.P.F. and Muller, E.B. (2001) The dynamics of the macromolecular composition of biomass. *J. Theor. Biol.* 212, 237–251

38 Rußmayer, H. *et al.* (2015) Systems-level organization of yeast methylotrophic lifestyle. *BMC Biol.* 13, 80

39 Stewart, M.Q. *et al.* (2001) Alcohol oxidase and dihydroxyacetone synthase, the abundant peroxisomal proteins of methylotrophic yeasts, assemble in different cellular compartments. *J. Cell Sci.* 114, 2863–8

40 Rebnegger, C. *et al.* (2016) *Pichia pastoris* Exhibits High Viability and a Low Maintenance Energy Requirement at Near-Zero Specific Growth Rates. *Appl. Environ. Microbiol.* 82, 4570–4583

41 Heijnen, J.J. and Van Dijken, J.P. (1992) In search of a thermodynamic description of biomass yields for the chemotrophic growth of microorganisms. *Biotechnol. Bioeng.* 39, 833–858

APPENDIX II

In this appendix, supplementary details on biomass composition are provided. **Table S4-1** to **Table S4-3**show specific information on biomass composition for all the conditions tested. These include the reconciled macromolecular and elemental composition (**Table S4-1**), amino acid composition (**Table S4-2**) and lipid profile of each condition (**Table S4-3**).

These tables also include the averaged compositions for glycerol and methanol used for building the new biomass equations in iMT1026 v3.0 and detailed in **File S4-3**.

As a result, new glycerol and methanol specific biomass equations were incorporated into the model. This allowed enhancing model accuracy when experimental data of the corresponding glycerol and methanol chemostats is simulated (**Fig. S4-1**).

The reduction of maintenance energy requirements allowed predicting macroscopic growth parameters using iMT1026 v3.0 (**Fig. S4-2**).

Table S4-1. Detailed reconciled elemental and macromolecular composition of glycerol- and methanol-grown cells at different growth rates. Values represent weight/weight % ± sd.

D (h⁻¹)	Glycerol				Average glycerol [a]	Methanol			Average methanol [a]
	0.035	0.065	0.100	0.160		0.035	0.065	0.100	
Protein	36.04 ± 6.68	41.20 ± 0.41	41.07 ± 2.91	43.31 ± 2.81	40.97 ± 1.46	48.64 ± 0.96	50.72 ± 0.48	51.54 ± 2.07	50.1 ± 0.76
Carbohydrate	45.58 ± 7.46	37.04 ± 1.08	35.92 ± 3.12	33.90 ± 1.27	35.89 ± 2.01	28.98 ± 2.79	28.89 ± 0.76	23.53 ± 1.33	27.32 ± 1.46
Lipid	1.25 ± 0.28	1.79 ± 0.35	2.19 ± 0.47	2.75 ± 0.03	2.48 ± 0.43	2.17 ± 0.24	1.87 ± 0.06	2.35 ± 0.57	1.97 ± 0.17
RNA	5.97 ± 0.14	6.68 ± 1.06	7.58 ± 1.19	9.08 ± 2.58	7.79 ± 0.55	5.69 ± 0.58	6.28 ± 1.25	7.05 ± 0.37	6.98 ± 0.59
DNA	0.19 ± 0.00	0.19 ± 0.01	0.18 ± 0.01	0.18 ± 0.01	0.19 ± 0.00	0.19 ± 0.00	0.18 ± 0.00	0.18 ± 0.02	0.18 ± 0.01
SO₄	0.28 ± 0.06	0.40 ± 0.00	0.45 ± 0.10	0.46 ± 0.11	0.46 ± 0.08	0.63 ± 0.00	0.69 ± 0.05	0.66 ± 0.06	0.63 ± 0.04
H₂O	5.76 ± 0.63	5.66 ± 0.19	6.58 ± 2.51	6.57 ± 0.19	5.62 ± 0.71	8.36 ± 0.08	6.16 ± 1.82	8.06 ± 0.70	7.16 ± 0.91
Metals	5.27 ± 1.00	7.04 ± 0.11	6.03 ± 0.94	5.76 ± 1.14	7.02 ± 0.59	5.33 ± 2.08	5.2 ± 1.95	6.62 ± 0.11	6.32 ± 0.61
C	42.31 ± 0.07	41.94 ± 0.06	41.86 ± 0.91	42.44 ± 1.00	41.98 ± 0.27	42.19 ± 0.76	43.21 ± 0.10	41.88 ± 0.21	42.28 ± 0.53
H	6.29 ± 0.03	6.24 ± 0.00	6.34 ± 0.18	6.44 ± 0.08	6.24 ± 0.06	6.57 ± 0.09	6.45 ± 0.17	6.46 ± 0.07	6.43 ± 0.05
N	7.41 ± 1.16	8.42 ± 0.11	8.73 ± 0.73	9.22 ± 0.24	8.58 ± 0.31	9.69 ± 0.08	10.16 ± 0.12	10.44 ± 0.43	10.06 ± 0.16
O	37.81 ± 2.17	35.36 ± 0.04	36.00 ± 2.57	34.98 ± 0.40	35.11 ± 0.58	35.22 ± 1.26	33.89 ± 1.61	33.43 ± 0.74	33.83 ± 0.46
S	0.25 ± 0.06	0.30 ± 0.00	0.25 ± 0.04	0.30 ± 0.04	0.30 ± 0.03	0.41 ± 0.01	0.43 ± 0.02	0.43 ± 0.03	0.41 ± 0.01
Ashes	5.93 ± 1.04	7.74 ± 0.00	6.82 ± 1.07	6.61 ± 1.28	7.79 ± 0.55	5.92 ± 2.03	5.86 ± 1.83	7.36 ± 0.14	6.98 ± 0.59

[a] Average compositions are weighted averages using 1/sd.

Table S4-2. Amino acid composition of cell protein extracts for all the growth conditions tested. Values represent % mol/mol ± sd.

D (h⁻¹)	Glycerol				Methanol			Average glycerol [c]	Average methanol [c]	p-value [d]
	0.035	0.065	0.100	0.160	0.035	0.065	0.100			
Ala	9.00 ± 0.80	8.68 ± 0.72	9.18 ± 0.32	9.92 ± 0.09	8.22 ± 0.11	8.11 ± 0.68	8.94 ± 0.82	8.97 ± 0.53	8.54 ± 0.45	0.045
Arg	5.12 ± 1.44	4.76 ± 0.51	6.14 ± 1.80	5.28 ± 0.90	5.51 ± 0.03	5.59 ± 0.18	5.64 ± 0.58	5.52 ± 0.59	5.62 ± 0.06	0.548
Asx [a]	10.3 ± 2.03	10.18 ± 0.21	9.43 ± 1.31	9.59 ± 0.00	12.03 ± 0.09	11.87 ± 0.19	11.92 ± 1.49	10.00 ± 0.43	11.92 ± 0.09	0.001
Cys	0.44 ± 0.08	0.38 ± 0.20	0.20 ± 0.01	0.28 ± 0.14	0.41 ± 0.33	0.33 ± 0.13	0.43 ± 0.21	0.36 ± 0.10	0.40 ± 0.05	0.510
Glx [a]	16.15 ± 2.59	14.35 ± 0.43	16.50 ± 2.55	15.2 ± 0.25	13.32 ± 0.84	13.52 ± 0.48	13.93 ± 0.52	16.17 ± 0.97	13.54 ± 0.31	0.012
Gly	5.89 ± 1.08	7.76 ± 0.05	7.31 ± 0.57	7.73 ± 0.03	7.39 ± 0.69	7.47 ± 0.67	6.81 ± 2.03	6.46 ± 0.88	7.06 ± 0.36	0.926
His	2.08 ± 0.10	2.14 ± 0.05	2.14 ± 0.13	2.12 ± 0.06	2.54 ± 0.11	2.49 ± 0.05	2.22 ± 0.02	2.12 ± 0.03	2.50 ± 0.17	0.005
Ile	4.5 ± 0.63	4.62 ± 0.11	4.44 ± 0.49	4.87 ± 0.20	4.74 ± 0.19	4.92 ± 0.36	4.68 ± 0.14	4.55 ± 0.19	4.82 ± 0.12	0.289
Leu	7.14 ± 1.13	7.69 ± 0.61	7.05 ± 0.12	7.39 ± 0.45	7.66 ± 0.22	7.69 ± 0.12	7.72 ± 0.12	7.34 ± 0.29	7.68 ± 0.03	0.116
Lys	6.95 ± 0.96	6.88 ± 0.01	6.61 ± 0.39	6.73 ± 0.15	6.67 ± 0.21	6.67 ± 0.14	7.11 ± 0.35	6.86 ± 0.15	6.89 ± 0.26	0.894
Met	1.00 ± 0.16	1.00 ± 0.14	0.63 ± 0.33	0.89 ± 0.23	0.97 ± 0.31	1.02 ± 0.07	1.03 ± 0.12	0.83 ± 0.17	0.99 ± 0.03	0.253
Phe	3.19 ± 0.48	3.50 ± 0.12	3.25 ± 0.25	3.27 ± 0.12	3.43 ± 0.07	3.43 ± 0.04	3.47 ± 0.11	3.26 ± 0.14	3.45 ± 0.02	0.155
Pro	4.74 ± 0.01	4.59 ± 0.20	4.45 ± 0.13	4.31 ± 0.06	4.71 ± 0.03	4.79 ± 0.03	4.29 ± 0.12	4.52 ± 0.18	4.45 ± 0.27	0.555
Ser	6.77 ± 0.08	6.95 ± 0.31	6.63 ± 0.32	6.60 ± 0.06	6.37 ± 0.12	6.38 ± 0.22	6.17 ± 0.17	6.79 ± 0.16	6.31 ± 0.12	0.002
Thr	7.31 ± 0.24	6.81 ± 0.18	6.88 ± 0.22	6.46 ± 0.24	6.42 ± 0.25	6.38 ± 0.06	6.07 ± 0.09	6.88 ± 0.35	6.33 ± 0.19	0.003
Trp [b]	1.18 ± n.d.	1.01 ± n.d.	1.12 ± n.d.	0.99 ± n.d.	0.98 ± n.d.	0.82 ± n.d.	0.91 ± n.d.	1.08 ± 0.09	0.90 ± 0.08	0.040
Tyr	2.29 ± 0.29	2.35 ± 0.11	2.10 ± 0.07	2.28 ± 0.05	2.75 ± 0.11	2.58 ± 0.13	2.58 ± 0.05	2.28 ± 0.11	2.64 ± 0.10	2.4e-4
Val	5.96 ± 0.72	6.35 ± 0.14	5.93 ± 0.43	6.08 ± 0.03	5.87 ± 0.04	5.93 ± 0.12	6.09 ± 0.05	6.01 ± 0.19	5.96 ± 0.11	0.424

[a] Asx and Glx represent the pair of Asp/Asn and Glu/Gln respectively.

[b] Trp was not measured, thus, recalculated according to values in [1].

[c] Average compositions for each carbon source are weighted averages using 1/sd of the different analysed growth rates.

[d] p-value resulting of the comparison of glycerol and methanol datasets for each amino acid by applying a 2-tailed Student's t-Test. Statistically significant differences on composition were considered when p-value < 0.05.

Table S4-3. Amino acid composition of cell protein extracts for all the growth conditions tested. Values represent % mol/mol ± sd.

D (h⁻¹)	Glycerol				Methanol			Average glycerol [a]	Average methanol [a]
	0.035	0.065	0.100	0.160	0.035	0.065	0.100		
TAG	7.50 ± 2.45	5.89 ± 1.72	9.36 ± 4.66	6.82 ± 4.99	0.10 ± 0.13	0.00 ± 0.00	0.00 ± 0.00	7.13 (6.87) ± 1.47	0.00 (0.00) ± 0.06
AG	36.06 ± 8.84	27.72 ± 1.34	28.77 ± 1.05	21.33 ± 4.54	53.74 ± 1.27	42.43 ± 2.76	36.15 ± 14.91	28.41 (27.40) ± 6.03	46.38 (44.74) ± 8.91
ERG	20.38 ± 1.63	16.62 ± 5.73	18.51 ± 0.61	22.99 ± 3.23	16.26 ± 6.16	17.69 ± 2.77	19.51 ± 9.48	19.61 (18.91) ± 2.72	16.55 (15.96) ± 1.63
PE	11.25 ± 7.21	14.04 ± 1.00	12.02 ± 3.36	11.96 ± 0.34	6.88 ± 1.71	10.69 ± 0.03	11.77 ± 1.46	12.61 (12.16) ± 1.20	10.00 (9.64) ± 2.57
CAR	2.69 ± 1.63	3.64 ± 0.51	2.80 ± 1.40	2.21 ± 0.51	1.71 ± 0.95	2.61 ± 0.52	3.18 ± 0.18	2.92 (2.82) ± 0.59	2.69 (2.60) ± 0.74
PA	0.48 ± 0.68	1.55 ± 0.64	0.38 ± 0.54	1.10 ± 0.01	0.40 ± 0.56	0.30 ± 0.42	0.30 ± 0.43	1.10 (1.06) ± 0.55	0.31 (0.30) ± 0.06
PC	15.29 ± 4.22	19.11 ± 1.57	17.29 ± 1.96	19.89 ± 4.29	10.53 ± 3.96	17.32 ± 0.93	17.09 ± 4.29	18.36 (17.71) ± 2.05	15.20 (14.66) ± 3.86
PI/PS	6.35 ± 0.82	11.44 ± 3.92	10.88 ± 1.76	13.70 ± 1.34	10.39 ± 0.12	8.96 ± 0.04	11.99 ± 0.57	9.85 (9.50) ± 3.08	8.87 (8.56) ± 1.52
SPH [a]								(3.55)	(3.55)

TAG: Triacylglyceorls; FFA: free fatty acids; STE: sterols; CAR: cardiolipin; PA: phosphatidic acid; PC: phosphatidylcholine; PI/PS: phosphatidylinositol / phosphatidylserine; SPH: sphingolipids.

[a] SPH content was not measured in lipid analysis. Thus, as well as in [2], values are taken from [3].

[b] Average compositions are weighted averages using 1/sd. In brackets, values used in iMT1026 v3 and represent the rescaled lipid content with the addition of the SPH fraction.
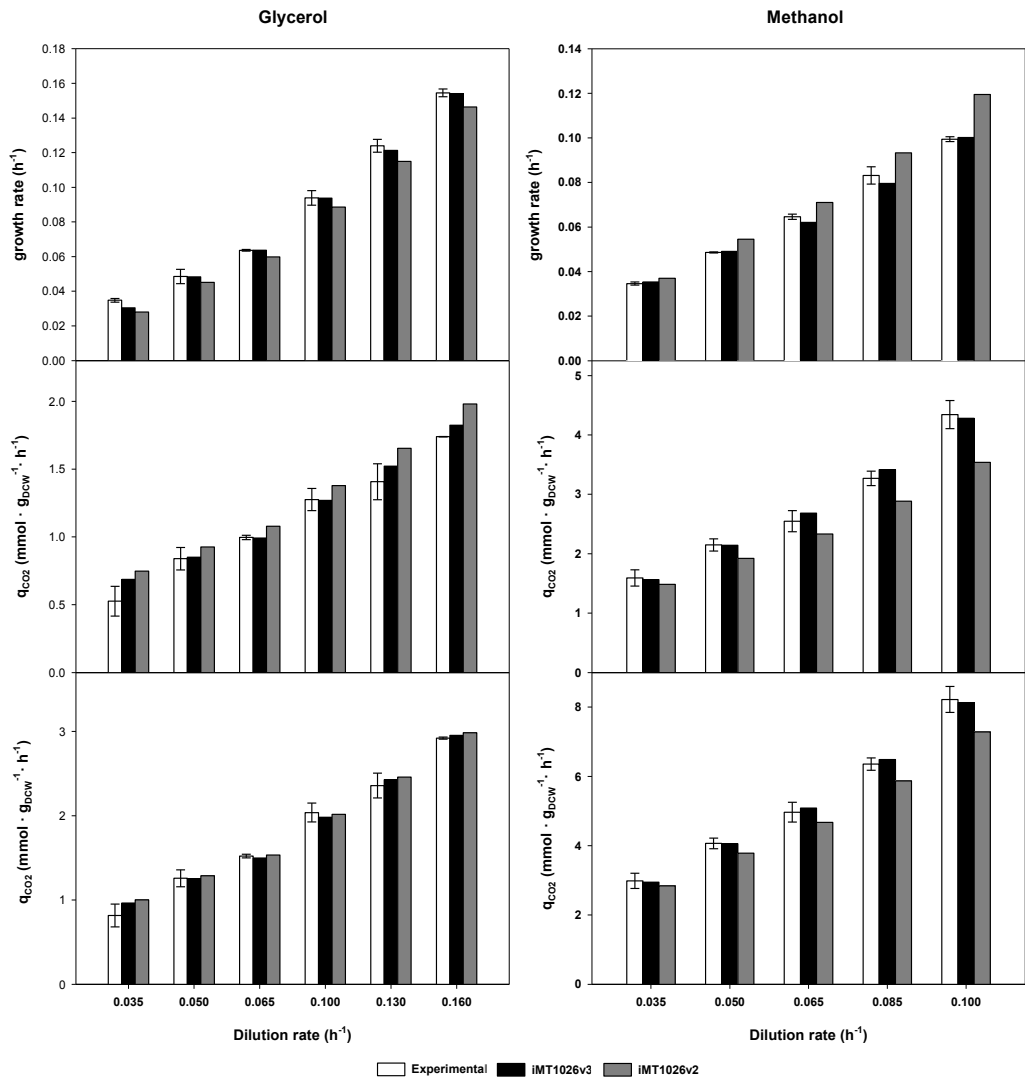
Fig. S4-1. Performance of iMT1026 v3.0 and iMT1026 v2.0 models compared to experimental data for the glycerol and methanol cultivations at different growth rates. For the simulations, the specific substrate uptake rate was set as constrain and biomass was maximised. In iMT1026 v3.0, the specific biomass equations, as well as new non-growth associated maintenance energy values for glycerol and methanol were enabled accordingly to the corresponding carbon source specific simulation.
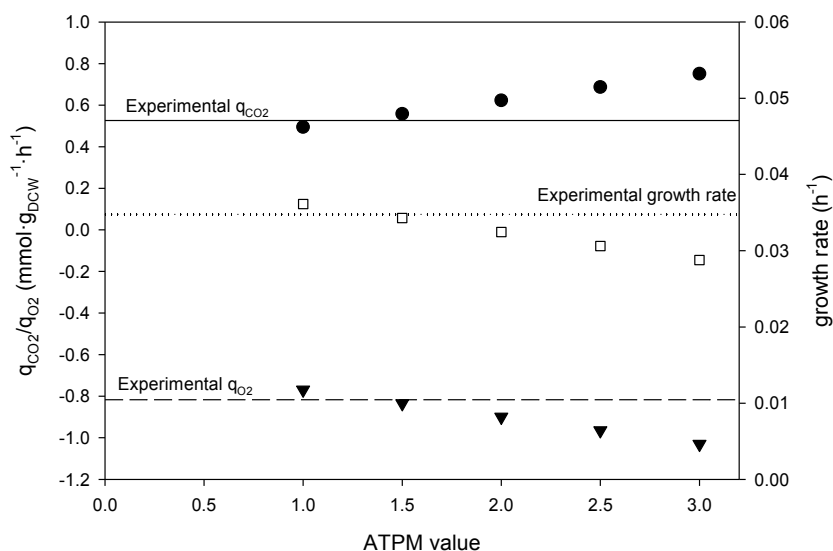
Fig. S4-2. Prediction of macroscopic growth parameters in glycerol-grown cells at 0.035 h-1 using different values for non-growth associated maintenance (ATPM). Substrate uptake rate was constrained according to the experimental data and different values for the 'ATPM' reaction were tested. Default ATPM corresponding to glycerol-grown biomass is 2.9 mmol ATP·gDCW-1·h-1. qCO2: experimental (solid line) and predicted ($\bullet$); growth rate: experimental (dotted line) and predicted ($\square$); qO2: experimental (dashed line) and predicted ($\blacktriangledown$).

References

1. Carnicer, M., Baumann, K., Töplitz, I., Sánchez-Ferrando, F., Mattanovich, D., Ferrer, P., Albiol, J., 2009. Macromolecular and elemental composition analysis and extracellular metabolite balances of Pichia pastoris growing at different oxygen levels. Microb. Cell Fact. 8, 65.

2. Tomàs-Gamisans, M., Ferrer, P., Albiol, J., 2016. 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 11, e0148031.

3. Grillitsch, K., Tarazona, P., Klug, L., Wriessnegger, T., Zellnig, G., Leitner, E., Feussner, I., Daum, G., 2014. Isolation and characterization of the plasma membrane from the yeast Pichia pastoris. Biochim. Biophys. Acta - Biomembr. 1838, 1889–1897.

5

^{13}C-based MFA of *P. pastoris* growing on glycerol

Table of contents

5. ^{13}C-based MFA of *P. pastoris* growing on glycerol

5.1. INTRODUCTION

The first metabolic networks for 13C-labelling experiments (CLE) were based on the pre-existing knowledge on biochemical pathways that included the amino acid biosynthesis pathways [1]. These central carbon metabolism networks have been extensively used with only minor modifications that improved the experimental datasets adjustment in flux estimations [2]. Early studies reported negligible fluxes in certain parts of the network, such as glyoxylate pathway or malic enzyme, in the tested conditions [3]. Subsequent studies, omitted such reactions and pathways in the respective 13C-MFA [4,5]. Despite there is no clear evidence of large fluxes in glyoxylate pathway in P. pastoris, the default omission of certain pathways may lead to inaccurate flux estimation under novel untested conditions. In addition, traditional metabolic models are simplifications of the whole metabolic network, therefore, even with the addition of particular reactions that improve flux estimation, Jordà et al. [4] reported a redox cofactor imbalance which could be due to the incomplete network in which key reactions that would fully balance redox cofactors would be missing or erroneously assigned using the incorrect cofactor.

In GSMM, the complete metabolic network is represented, thus at a glance, GSMM would be an alternative for ^{13}C-MFA avoiding missing reactions. Nevertheless, large-scale ^{13}C-MFA have significant drawbacks. One of the inconveniences is the requirement for an accurate atom transition mapping. Although there are databases including the reaction atom mapping of biological pathways, GSMM-specific reactions would require the additional effort to accurately annotate atom transitions. Another important hindrance is the huge number of variables that impacts computation complexity. Finally, a major handicap is the inability to resolve all fluxes due to the robustness and redundancy of metabolic pathways which include parallel and alternative pathways as well as compartmentalization [6]. Until more powerful computational methods are developed, GSMM are not a practical alternative to core models. Gopalakrishnan and Maranas [7] used a GSMM for ^{13}C-MFA. However, they conclude that reducing GSMM to a similar size than the currently used core metabolism models used in ^{13}C-MFA would be a feasible alternative instead of full GSMM. Reduced models would have more

complete biomass equation as well as alternative pathways that would allow better adjustments in flux distribution. In this regard, algorithms for reducing GSMM to core models have been developed [8]. Briefly, by protecting relevant reactions and applying phenotypic constrains in simulations the algorithm successively eliminates reactions until a minimal model that fulfills the phenotypic constrains is obtained. This method has been employed in *E. coli* and reported to produce equivalent predictions for flux distribution than the original GSMM [9].

First ^{13}C-MFA studies in *P. pastoris* used biosynthetically directed fractional (BDF) ^{13}C-labeling of proteinogenic amino acids with 2D-NMR that enabled the analysis of metabolic flux ratios (METAFoR analysis) [10]. This method relies on the identification of conserved C-C bounds. Therefore, the information derived of certain substrates with a low number of carbon bounds is limited. Alternatively, further studies in our research group employed non-stationary CLE [2,11] and were coherent with equivalent studies using ^{13}C-labelling data of proteinogenic amino acids and METAFoR analysis [12]. In addition, the latter method allowed the determination of flux distributions through an extended network and include flux reversibility in certain parts of the network. Between these two approaches for flux determination, is the analysis of proteinogenic amino acids with subsequent iterative fitting of fluxes to the measured MIDs and cell exchange fluxes. Despite this method does not allow the resolution achieved in non-stationary CLE, it is expected to improve the resolved fluxes in comparison to METAFoR analysis. Thus, in this chapter, the genome-scale metabolic model is reduced to a glycerol-specifc core model. This reduced model is further used for ^{13}C-MFA of *P. pastoris* growing on glycerol as carbon source at different growth rates.

5.2. MATERIAL AND METHODS

5.2.1. Strain and cultivation conditions

Pichia pastoris X-33 strain was used throughout this study. Duplicate carbon limited chemostat cultivations were performed using a Sartorius 0.5-L bioreactor (Sartorius AG, Göttingen, Germany) at dilution rates (D) of 0.05, 0.10 and 0.16 h^{-1} with a working volume of 0.3 L maintained by a gravimetrically controlled peristaltic pump. Chemostat cultivations were performed for at least 5 residence times (τ).

Batch and chemostat media were adapted from Baumann et al. [14] and diluted to achieve an approximately final biomass of 6 g/L when the steady state is reached. Thus, briefly, batch medium contained: 9.98 g/L glycerol, 0.46 g/L citric acid, 3.15 g/L $(NH4)_2HPO_4$, 0.006 g/L CaCl2 · $2H_2O$, 0.225 g/L KCl, 0.125 g/L MgSO4 · $7H_2O$, 0.5 mL Biotin (0.2 g/L), 1.15 mL PTM1 trace salts stock solution (prepared as described in Baumann et al., 2008). pH was adjusted to 5.0 with 25% HCl. Chemostat medium contained: 10 g/L glycerol, 0.818 g/L citric acid, 4.35 g/L $(NH4)_2HPO_4$, 0.01 g/L CaCl2 · $2H_2O$, 1.7 g/L KCl, 0.65 g/L MgSO4 · 7H2O, 1.0 mL Biotin (0.2 g/L), 1.6 mL PTM1 trace salts stock solution.

An Inoculum was cultivated overnight at 30ºC, 150 rpm in a 0.5 L shake flask containing 75 mL of basal medium with glycerol and supplemented with biotin (1% yeast nitrogen base, 4 · 10^{-5} % biotin, 1% glycerol). Bioreactor was inoculated an initial OD_{600} of 0.3 – 0.5. Once glycerol is completely depleted, continuous cultivations were started by initiating the feeding at the corresponding flow rate. The aeration rate was set to 1 vvm with sterile atmospheric air and the off-gas O_2 and CO_2 concentrations were measured a Prima Pro Process Mass Spectrometer (Thermo Fischer Scientific). Cultivation conditions were set to stirring rate of 500 rpm, pH 5.0 controlled by automatic addition of 15% ammonia and the temperature was maintained at 25ºC.

5.2.2. Labelling experiment and biomass harvest

After a minimum of 5τ of continuous cultivations with non-labelled glycerol, the feed was switched to the labelled medium. Labelled feed medium composition was the same than unlabeled feed medium composition, replacing glycerol for 20% [1,3-^{13}C]glycerol (CortecNet)

and 80% unlabeled glycerol. Labelled medium was feed for at least 2τ. 50 – 100 mL of biomass were collected and centrifuged 15 min at 16000g. The supernatant was discarded and pellets were frozen with liquid N_2 and stored at -80ºC for further extraction and analysis of proteinogenic amino acids.

5.2.3. Analytical methods

Cell density and dry cell weight

Cell density was monitored by optical density at 600 nm. Dry cell weight (DCW) was measured in duplicate by gravimetric methods. A known volume of sample (5 to 10 mL) was filtered throughout a dried pre-weighted 0.45 µm polyether sulfone filters (Frisenette, Knebel, Denmark) and washed with distilled water. Filters were dried in a microwave oven at 150 W for 20 min and cooled down in a desiccator for at least 2 h and finally weighted. For external metabolite analysis, samples were taken along the cultivation, filtered through a 0.22 µm syringe filters and stored at -20ºC until subsequent analysis. Glycerol, was the only and main peak detected in HPLC analyses performed as described in [15].

5.2.4. Proteinogenic amino acid MID determination

Amino acid extraction, derivatization and GC-MS analysis

Isotope distribution of the proteinogenic amino acid was determined as described by Knudsen [16]. Briefly, 5 mg of biomass pellets were hydrolyzed with 6 M hydrochloric acid at 105ºC for 16h. Once at room temperature, samples were dried for 3h under a stream of nitrogen. Samples were redissolved in water and filtered through a Strata SCX columns and washed with 50% ethanol to remove all the impurities. Samples were eluted with 1 N NaOH and additionally with the elution solution [16]. Two types of derivatives were prepared for GC-MS analysis: N-ethoxycarbonyl-amino ethyl-esters (ECF) and N-dimethyl-amino-methylene-methyl-esters (DMFDMA) following the accurate protocol described by Knudsen [16]. Derivatised amino acid samples were analyzed in a GC-MS Agilent 6890 gas chromatograph coupled to an Agilent 5973 quadruple MS accordingly to the specified settings [16]. GC/MS Translator (Agilent) was used to convert the row files into readable for Agilent MassHunter Qualitative Analysis.

MID are uncorrected for naturally labelled atoms other than carbon backbone [17,18]. OpenFlux [13] was used for correcting the MID of each amino acid according to the expected fragmentation ions obtained in GC-MS analysis [19]. In addition, biomass was harvested after $2\ \tau$ and thus, the fraction of labelled biomass ($X_{labelled}$) at the steady state was calculated according to a first-order wash-out kinetics [20]: $X_{labelled} = 1 - e^{-t/\theta}$, where t is the labelling time and θ the residence time of the chemostat.

5.2.5. Statistical analysis

Chemostat cultivation data was checked for consistency and standard reconciliation procedures were applied [21]. Biomass molecular formula was set according to the growth rate and values reported in **Chapter 4** of the present thesis. In all the cultivation sets, statistical consistency test was passed with a confidence level of 95%. Consequently, there was no evidence of gross measurement errors.

5.2.6. Core model generation

In order to obtain a complete network of the central carbon metabolism of *P. pastoris*, including all the relevant reactions a core model was generated by reducing iMT1026 v3.0 using NetworkReducer [8] with CellNetAnalyser 2016.1[22] under Matlab 2011. A detailed procedure and commands for model reduction can be found in **File S5-1**. Default glycerol growing flux constrains were set and glycerol was set as the only possible carbon source. Maximal growth rate in the reduced network is constrained to be 99.9% of the maximal growth rate for the iMT1026 v3.0 (here *PpaGS*, meaning '*P. pastoris* Genome-Scale') corresponding to the glycerol chemostat cultivations in Chapter 4. In a first step, 46 reactions of the central carbon metabolism were protected (**File S5-1**) and *PpaGS* is reduced to a pruned model (*PpaPruned*). Subsequently, *PpaPruned* is further reduced to *PpaCore* by a compressing procedure with a new set of 56 protected reactions (File S5-1). Due to numerical reasons the 'cof' metabolite was removed from biomass equation. The resulting *PpaCore* was tested for growth in glycerol and the same maximal growth rates obtained in *PpaGS* were achieved for a constrained uptake of glycerol.

5.2.7. ^{13}C-Metabolic flux analysis (^{13}C-MFA)

Flux calculations were performed with OpenFLUX [13] under Matlab 2011 using FMINCON from Matlab's optimization toolbox. Previously to perform MFA, *PpaCore* was adapted to OpenFLUX requirements in three steps: (1) reversible reactions were expressed as two paired of irreversible reactions; (2) reactions mapping label distribution in the measured proteinogenic amino acids were added according to the appropriate compartmentalization [1]; (3) carbon atom transition equations were added according to previous *P. pastoris* models [5] and databases [23]. In order to avoid biased solutions, O_2, reduction equivalents, energetic cofactors and additional non-carbon-balanceable metabolites were defined as excluded metabolites. Moreover, those reactions in *PpaCore* that uniquely contained excluded metabolites were also removed from the final model for ^{13}C-metabolic flux calculations. The resulting model (**Table S5-1**) was used for ^{13}C-MFA. MIDs showed an average deviation below 5% and the model was fitted to the experimental data by the least squares method detailed in [24] using the glycerol uptake rate and specific biomass generation rates as constrains. The parameter estimation procedure was repeated 100 times. Subsequently, the solution cluster with lower residual error was used for sensitivity analysis using the non-linear approach developed by Antoniewicz et al. [25]. Sensitivity analysis was performed in order to find the lower and upper confidence interval boundaries of calculated fluxes at a 95% confidence level [13]. Default configuration settings were used for the sensitivity analysis. Due to the impossibility to determine individually forward and reverse fluxes in certain pairs of reversible reactions, only the net fluxes were calculated and submitted to the sensitivity analysis for those forward/reverse reaction pairs. ^{13}C flux fitting to MID was performed for each experimental replicate and subsequently averaged.

5.2.8. Calculation of redox cofactor regeneration rates and energy requirements

Rates for redox cofactor regeneration and ATP synthesis were derived from the determined fluxes. Once the solution of the metabolic system was found, estimated fluxes were further used for calculating the remaining reaction fluxes that contained excluded metabolites. Redox cofactor balance was checked and any surplus of reduction equivalents (NADH both cytosolic and mitochondrial) was considered a source of electrons to be

transferred to the electron transport chain (ECT). Therefore, assuming the complete electron transfer from the reduction equivalents to ETC and taking into account the oxygen requirements for biomass synthesis (included in the full biomass equation), the oxygen uptake rate could be calculated and compared to the experimental values. In addition, since the model includes the complete ETC with proton translocation to the mitochondrial intermembrane space and the consecutive reaction for ATP synthesis [26], the theoretical maximal ATP generation rate can be estimated. This way, the total ATP generation can be taken later into account for calculating the energetic parameters. Essentially, two parameters were determined: growth and non-growth associated maintenance energy (GAME and NGAME respectively) according to Pirt's equation [27] and calculating the linear regression between ATP generation and growth rates [28]. The y-intercept would correspond to the NGAME ($\mu = 0.0$ h^{-1}) and the slope to GAME. Those values can be later compared to those obtained using the complete *PpaGS* model.

5.3. RESULTS AND DISCUSSION

5.3.1. Macroscopic growth parameter characterisation

Despite the rising interest on using glycerol for microbial cultivations, there is limited information on the physiology of *P. pastoris* growing on such carbon source. In Chapter 4, a series of cultivations were performed at different growth rates using glycerol as sole carbon source. In this initial work, a macroscopic landscape is provided by analysing the effects of growth rate on biomass composition and its growth profile which are both described in more detail. Thus, in this chapter, a series of chemostat cultivation were performed at three different dilution rates that were already tested in previous experiments [1] as well as in in Chapter 4: 0.05, 0.10 and 0.16 h^{-1}.

As in previous experiments, cultivation of *P. pastoris* using glycerol as sole carbon source showed no by-product secretion or accumulation. In addition, glycerol can be considered as completely consumed in the reactor, as its measurement in the broth samples was below the detection limit of the analytical method. All the measured external macroscopic fluxes showed a clear correlation with the growth rate ($R^2 > 0.99$), that is in agreement with results in Chapter

4 where a linear range for the macroscopic cell's variables growing on glycerol is established between 0.05 and 0.16 h^{-1}.

In this new set of cultivations at different growth rates using glycerol as carbon source, biomass yields were calculated. Y_{XS} ranges between 0.70 – 0.72 g$_X$ · g$_S^{-1}$ (**Table 5-1**). These values are in the upper range of Y_{XS} calculated for glycerol in Chapter 4 (0.62 – 0.71 g$_X$ · g$_S^{-1}$) and are in agreement with biomass yields on glycerol for different yeast species [29]. Similarly, average RQ (0.58 ± 0.2) is close to 0.63 ± 0.03, the average RQ calculated in Chapter 4 for glycerol-grown cells. Therefore, despite reducing the carbon source concentration in the feeding media, cells showed similar macroscopic profile and thus cultivations were performed in comparable conditions.

Table 5-1. Reconciled macroscopic growth parameters for glycerol cultivations at different growth rates.

D_{SP}	0.05	0.10	0.16
D_{exp}	0.046 ± 0.006	0.098 ± 0.009	0.166 ± 0.006
q_S	-0.72 ± 0.11	-1.45 ± 0.13	-2.52 ± 0.14
q_X	1.62 ± 0.21	3.43 ± 0.31	5.89 ± 0.23
q_{CO_2}	0.55 ± 0.12	0.93 ± 0.07	1.65 ± 0.18
q_{O_2}	-0.90 ± 0.18	-1.66 ± 0.14	-2.89 ± 0.25
Y_{XS}	0.70 ± 0.02	0.73 ± 2e-3	0.72 ± 0.01
RQ	0.60 ± 0.02	0.56 ± 0.01	0.57 ± 0.01

D_{SP} (h^{-1}) corresponds to the set point dilution rate and D_{exp} (h^{-1}), the measured experimental D.
q_S, q_{O_2} and q_{CO_2} are expressed in mmol · g$_{DCW}^{-1}$· h^{-1}. Units for q_X are Cmmol · g$_{DCW}^{-1}$· h^{-1}.
Y_{XS} represents biomass yield (g$_X$ · g$_S^{-1}$)
RQ is respiratory quotient.

5.3.2. Reduction of *P. pastoris* genome-scale model

iMT1026 v3.0 was used for generating a derived central carbon metabolism model in two reduction steps, generating *PpaPruned* and *PpaCore*. The main model characteristics are summarized in **Table 5-2**. In order to avoid infeasibilities due to numerical tolerance constrains, 'cof' metabolite corresponding to cofactors in biomass equation was removed. 'COF' reaction has stoichiometric coefficients of the order of 1e-6 that can be lower than the minimal calculation tolerance. By removing COF, no significant impact on flux distribution nor predictability capacity was observed. In the first step of reduction, 46 reactions were protected including the main equivalent reactions in previous *P. pastoris* central carbon

metabolism model [5] and additional transport reactions due to the compartmentalization of the model. Protected phenotypes were stablished in order to ensure the accuracy of the predictions, thus, for a given glycerol uptake rate, the 99.9% of the maximal predicted growth rate, obtained using the original model, was used as a constraint. As reported in Chapter 4, no by-product generation was detected when growing in glycerol-limited chemostats. None of the by-product formation reactions were protected and consequently none of them were present in *PpaPruned*. Initially, the genome scale metabolic model contained 2237 reactions and 1881 metabolites (175 external). As a result of the network reduction the number of reactions and metabolites was considerably reduced (495 reactions and 513 metabolites). The degrees of freedom (*dof*) are also strongly reduced (from 485 to 4). This reduction could be due to the fact that the pruned model only takes into account growth in glycerol with no by-product generation. In the subsequent step, we further reduced *PpaPruned* applying a loss-free network compression [8], where linear and transport reactions were lumped and several other reactions were also grouped. As a result, *PpaCore* was generated and included 77 reactions with 102 metabolites with no further reduction in degrees of freedom. In addition, the maximal predicted growth rate was maintained as in the original model.

Table 5-2. Main properties of P. pastoris models used and generated in this study.

	iMT1026 v3.0	*PpaPruned*	*PpaCore*	Openflux *PpaCore*
# reactions	2237	495	77	120
# internal metabolites	1706	513	102	100
# external metabolites	175	9	9	9
degrees of freedom	485	4	4	-
μ_{max}	0.0940	0.0939	0.0939	-

In addition, the amino acid biosynthetic pathways included in the reduced model were checked and compared to those present in previous models of the central carbon metabolism. In addition, such pathways were verified based on available databases [23] to ensure a proper carbon atom mapping. Hence, amino acid carbon atom transitions were adapted from previous models [4,5,30].

5.3.3. ^{13}C-Metabolic flux analysis on glycerol-grown cells

The generated model was implemented in OpenFLUX code for ^{13}C-MFA determination employing the experimental MIDs corrected both for the naturally labelled atoms other than carbon backbone and extrapolating the MID to the steady state using a first order washout kinetics (**Table S5-2**). Furthermore, the metabolic fluxes were estimated independently for each experimental replicate and subsequently averaged. Despite an optimal flux value is obtained at the end of the calculations, here results are presented as the 95% confidence interval (CI) in order to provide a more informative description of the results taking into account the true uncertainty of estimated fluxes [25]. On one hand, the expression as mean of the optimal value ± SEM, would provide an overview of changes in the estimated fluxes, but it would not indicate the real uncertainty of the measurement due to the asimetry of the interval. Thus, if similar flux distributions were predicted between replicates, it would provide poor information on the confidence of the estimated fluxes. Then, in this chapter results will be expressed as confidence interval boundaries.

To our knowledge, there is only a single study analysing the metabolic flux distribution of glycerol-grown yeasts with ^{13}C-labelling experiments [1]. Although this previous study performed only a metabolic flux ratio analysis (METFoR) and thus no specific flux estimation values were provided, comparable flux ratio results were obtained in our case when analysing flux ratios in converging pathways with the resulting estimated fluxes. One of the conclusions by Solà et al. [1] was that the activity through glyoxylate cycle was very low. This observation was based on the available carbon labelling evidence (or better the lack of contrary evidence) as well as on previous studies in *S. cerevisiae* and *P. stipitis* grown in glucose, in which the glyoxylate pathway was almost inactive [31]. In addition Solà et al. [1] measured the activity of isocitrate lyase (ICL) and showed similar activities for both in glucose and glycerol chemostats. Since glucose is known to repress the glyoxylate pathway in *S. cerevisiae* [32] and measured activities of ICL in *P. pastoris* were similar for glucose and glycerol, it was assumed that the flux through glyoxylate cycle was negligible. This consideration has been assumed in subsequent studies [5,11,30,33,34] by omitting the glyoxylate pathway from the metabolic network for MFA. As shown in **Fig. 5-1**, the activity through glyoxylate cycle found in the present results is very low, almost negligible, and thus confirms the previous assumptions by Solà et al. [1]. A second observation was that the fraction of mitochondrial pyruvate derived

from malate was also very low or negligible, thus indicating that the malic enzyme is likewise almost inactive. Our calculations are in agreement to this observation and very low fluxes are estimated for malic enzyme. However, for the lowest dilution rate tested, the relative flux through the malic enzyme reaction is predicted to be higher than at the other dilution rates and accounting for less than 20% of the contribution to the mitochondrial pyruvate node. Flux ratios between gluconeogenesis and pentose phosphate pathway cannot be assessed in glycerol nor mixtures of glycerol and methanol cultivations using biosynthetically directed fractional ^{13}C labelling of proteinogenic amino acids [1,3]. Nevertheless, when applying ^{13}C-constrained metabolic flux analysis, fluxes through the oxidative branch of the pentose phosphate pathway (PPP) can be estimated [4]. Similarly to previous cultivations on mixtures of glycerol and methanol [4], our results indicate that the flux through the oxidative branch of PPP is almost negligible (**Fig. 5-1**). Thus, the majority of NADPH generated in cytosol would be produced in other reactions, such as the glycerol uptake pathway. Flux directionality in non-oxidative branch of PPP could only be determined for $D = 0.05$ h[-1] within the 95% CI. In the other two conditions tested, estimated CI includes both reaction directions as feasible **Fig. 5-1**. This uncertainty on PPP was previously described and attributed to the operation of PPP reactions close to the thermodynamic equilibrium and thus bidirectionally feasible [2].

The resulting core model included several mitochondrial transporters that in addition act as redox shuttles [35]. Nevertheless, the malate/aspartate shuttle appears to be the major redox and C4 intermediate metabolite transporter **Fig. 5-1**, while the flux through other transporters is estimated to be very low, almost negligible.

5.3.4. Impact of dilution rate on metabolic flux distribution

Although the flux through the oxidative branch of PPP is very low, and could be considered as negligible, the upper bound of the CI of the relative flux through the oxidative part of PPP appears to increase with the growth rate. Hence, while cultivations at 0.05 are estimated to have a relative flux between 0 and 0.2, those at 0.10 and 0.16 h[-1] showed a CI upper bound of 6.7 and 13.9 respectively. Thus, it would suggest that at higher growth rates the relative flux through the oxidative branch of PPP may be higher than in lower growth rates. Consequently, the upper bounds for the non-oxidative part of the PPP are also increased

at higher growth rates. As detailed in Chapter 4, the growth rate hypothesis (GRH) describes a positive correlation between the growth rate and the RNA and protein content [36–38]. Moreover, the analysis performed of the RNA content of cells growing on glycerol at different growth rates also depicts this positive correlation. Considering that major precursors for nucleic acids biosynthesis are generated in the pentose phosphate pathway [39], an additional demand of RNA would require an increase in PPP activity. Thus, the estimated increase in relative flux in the oxidative branch of PPP at high growth rates would supply the additional demand for RNA precursors. Therefore, an additional evidence for the GRH is provided that is in agreement with results described in the previous chapter.

The split ratio between gluconeogenesis-PPP and lower glycolysis at the glyceraldehyde 3-phosphate node is then altered with the growth rate. As a result of the increase in the relative flux though gluconeogenesis and PPP at higher growth rates, there is a reduction of relative flux through the lower glycolysis part and consequently to the TCA cycle (**Fig. 5-1**). Similar results were reported by Jordà et al. [4] in chemostat cultivations using mixtures of glycerol and methanol as carbon sources at low and high dilution rates (0.05 and 0.16 h^{-1}). Among the different mixtures of glycerol and methanol tested, a comparison with the condition with lower methanol:glycerol ratio (20:80, w%) can be performed. Authors estimated invariant absolute fluxes through TCA cycle (citrate synthase reaction). However, the substrate uptake rate is much higher in the highest growth rate. Therefore, the relative flux through citrate synthase reaction to the substrate uptake rate is significantly reduced at the high growth rate which agrees with our results. In addition, the relative flux through the malate/α-ketoglutarate transporter is correlated with the growth rate. The upper bound of the CI is increasing with the dilution rate, and the relative flux for D=0.16 h^{-1} would be the highest possible with respect to the other conditions. These results are also in agreement to those described by Solà et al. [1]. There, authors described that at higher growth rates the cytosolic-mitocondrial exchange flux of oxaloacetate was largely unidirectional from the cytosol to the mitochondria. Therefore, an increase in malate/α-ketoglutarate exchange supports the observed unidirectional transport of oxaloacetate into the mitochondria, since malate is subsequently reduced to oxaloacetate in mitochondria.
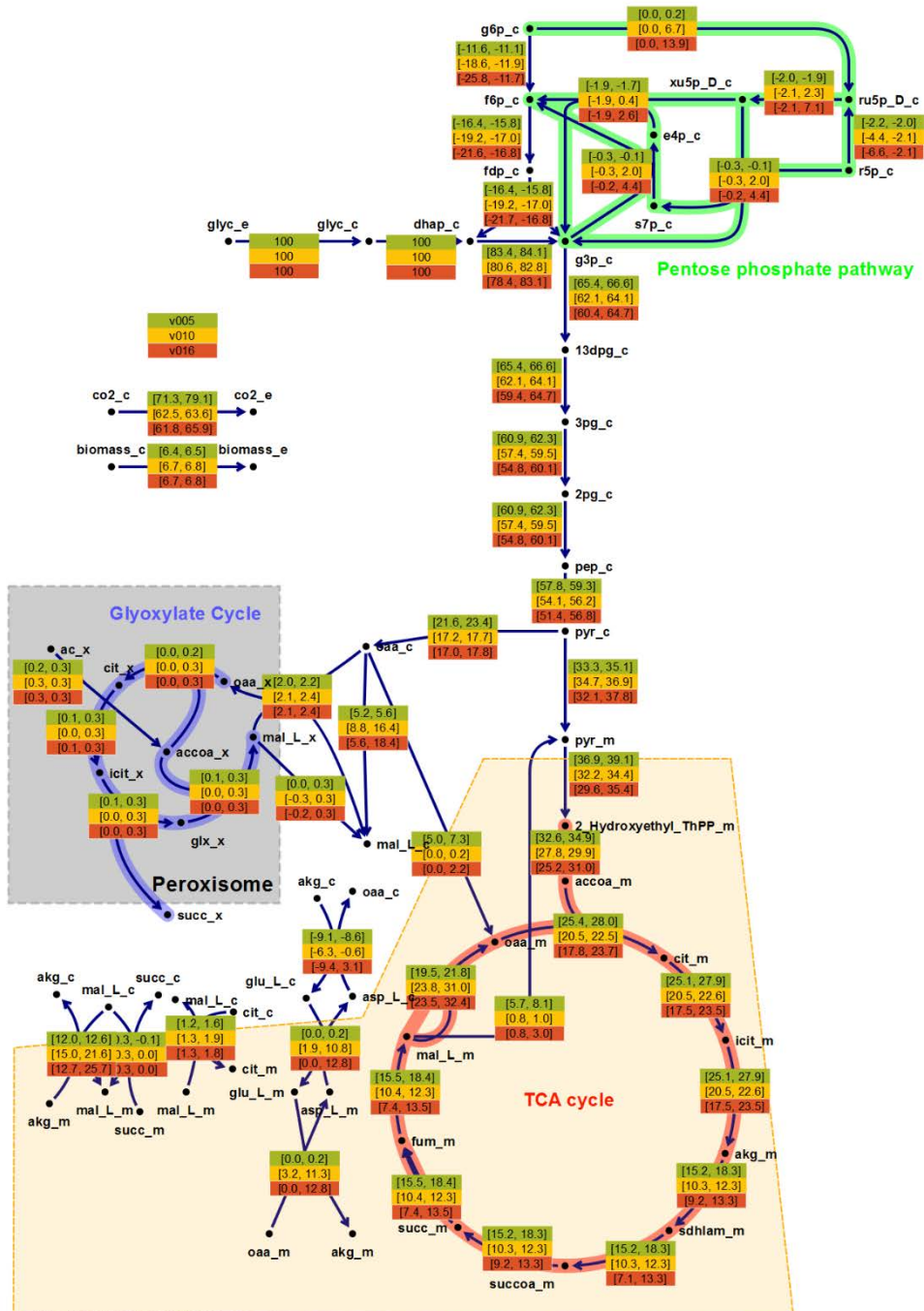
Fig. 5-1. Metabolic flux distribution estimated for *Pichia pastoris* growing on glycerol at different dilution rates: 0.05 h⁻¹ (top), 0.10 h⁻¹ (middle) and 0.16 h⁻¹ (bottom). Results are expressed as the 95% confidence interval of the estimated fluxes relative to the glycerol uptake rate in mmol glycerol · g_{DCW}^{-1} · h⁻¹. Lower and upper bounds of CI correspond to the maximum and minimum CI bounds between the replicates. Flux directionality assumption was represented by arrows; therefore, negative fluxes describe opposite direction.

5.3.5. Redox cofactor regeneration and energy metabolism

Redox cofactors were excluded from the metabolic flux calculation in order to avoid biased solutions. Consequently, estimated fluxes are derived exclusively from the adjustment of MIDs to the metabolic carbon stoichiometry with no interference of other additional metabolites. Once the metabolic fluxes were estimated, a calculation of both oxygen and cofactor regeneration was performed in order to check whether the estimated solution implicitly satisfies the electron balance. Initially, the estimated flux distribution solution predicted an excess of cytosolic NADH that could not be oxidised. Cells have redox shuttles that are able to transport NADH indirectly from cytosol to mitochondria [40]. One of these redox shuttles is the malate-aspartate shuttle (**Fig. 5-2**). In this particular case, despite being present in *PpaCore*, the exact specific activity of the mitochondrial shuttle cannot be calculated as no carbon rearrangement takes place. Consequently, only the net flux in the shuttle system can be calculated (i.e. difference between the cytosolic and mitochondrial flux). Therefore scaled fluxes in both compartments would result in identical flux distributions, but with additional amounts of NADH translocated to the mitochondria. With regards to the inability to predict scaled up fluxes in redox shuttles employing the currently used constrains, an additional flux fitting was performed by adding the cytosolic NAD(H) balance to the stoichiometric matrix for flux estimation. As a result, identical flux distributions were obtained for reactions other than the mitochondrial redox shuttle. Moreover, the increase in flux of mitochondrial redox shuttle in the new calculation corresponds exactly to the quantified excess of cytosolic NADH. Correspondingly, the theoretical oxygen consumption rates were calculated assuming that all the NADH excess is consumed in the ETC. Oxygen requirements included in the biomass equation were also taken into account and constrained in the ETC calculations. Thus, for each growth rate, the theoretically calculated oxygen requirements account over 97% of the experimentally determined oxygen consumption rates. Therefore, metabolic fluxes and calculated variables are highly consistent with the experimental data.

With the global electron balance and metabolic flux distribution, maximal ATP generation rates were also calculated. Thus, as described in the 5.2.8 section in Materials and methods, growth associated and non-associated maintenance energy could be calculated. The y-intersection in the regression of maximal ATP generation rates with the growth rate is 1.22 ± 0.48 mmol ATP $\cdot$ g_{DCW}^{-1} $\cdot$ h^{-1} and corresponds to the NGAME. Despite the difference, this

value is comparable with the 2.51 reported in the previous chapter for cultivations on glycerol. In addition, GAME was estimated to be 88.8 ± 4.1 mmol ATP $\cdot$ g_{DCW}^{-1}, that is higher but comparable to the 70.7 mmol ATP $\cdot$ g_{DCW}^{-1} calculated in the previous chapter, given that only 3 different growth rates are available.
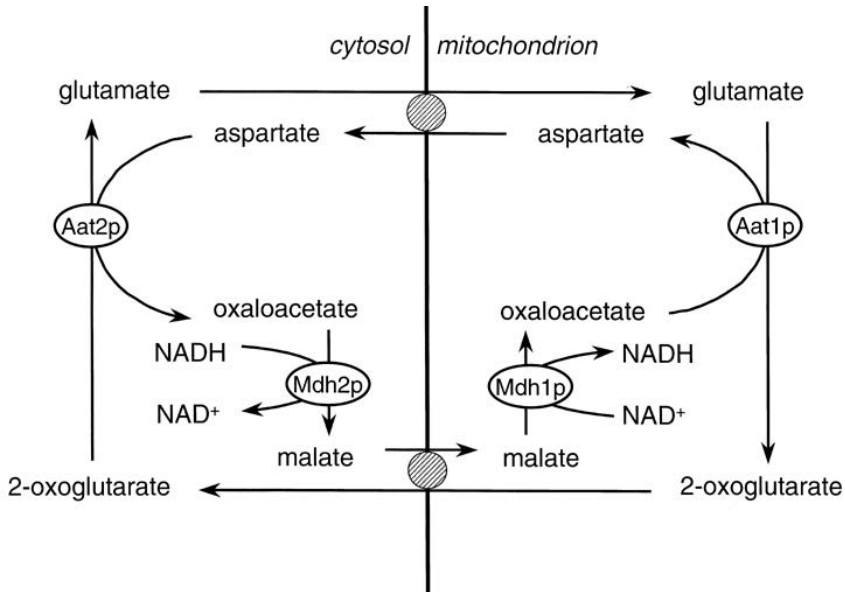
Fig. 5-2. The malate-aspartate shuttle, exchanging cytosolic NADH for mitochondrial NAD+ . Figure from [40] (reproduced with permission of Oxford University Press).

It is worth to mention that the oxidative branch of PPP is usually considered the major source of cytosolic NADPH [11,41]. Nevertheless, the predicted flux through this pathway in this case is very low, almost negligible, similarly than previous results on mixtures of glycerol and methanol [4]. Therefore, Jordà et al. already suggested that alternative reactions must supply the required cytosolic NADPH, otherwise a NADPH imbalance was observed. GSMM contains all the feasible reactions producing NADPH in cytosol. As a result of automatic model reduction, *PpaCore* contained the NADPH-dependent glycerol uptake pathway, although previous core models proposed NADH-dependent glycerol uptake. Therefore, from our point of view it is sensible to assume that in glycerol-grown cells, the glycerol uptake pathway uses de NADPH-dependent variant and this way it would be the major source for cytosolic NADPH formation. This way it will provide sufficient NADPH to fulfil the global NADPH requirements.

5.4. CONCLUSIONS

In order to get a better insight on the impact of growth rate on cell physiology, we extended the analysis performed in Chapter 4 to an additional level of detail by determining the metabolic fluxes distribution. Thus, a series of chemostat cultivations were performed and showed comparable results to those in previous chapter, with a similar growth profile and comparable results for energetic parameters.

The analysis of the labelling pattern in proteinogenic amino acids in steady state CLE allowed for the estimation of metabolic fluxes within an acceptable confidence range. Nevertheless, only net fluxes in bidirectional reactions can be determined in our case using this method, and alternative strategies such as non-stationary CLE should be used for identifying each of the directional fluxes [42]. The resulting metabolic flux distributions showed appreciable differences between the low and high growth rates. Furthermore, similarities in flux distribution found with previously obtained data [1] allowed for validating such data with alternative methods. In addition, results are comparable with ^{13}C-MFA based on cultivations under mixtures of glycerol and methanol (80:20 w%) at high and low growth rates. Furthermore, the used methodology enabled to resolve additional parts of the network in comparison to the METAFoR analysis. Moreover, the observed variations in flux distribution are in agreement to the results in Chapter 4, in which evidences for the growth rate hypothesis are provided. Thus, the described increase in RNA fraction at high growth rates would correlate with higher relative fluxes through the pentose phosphate pathway, the precursor supply for ribonucleic acids. The estimated flux distributions were checked and validated for the electron balance. Calculations based on the NADH generation rates and the electron transport chain accurately predicted the experimental measured oxygen uptake rate.

In this work, we provide an alternative central carbon metabolism derived from the genome-scale metabolic model of *P. pastoris* that successfully enables the determination of metabolic fluxes. Previously used metabolic models were developed from the available knowledge in biochemical pathways. Thus, central carbon metabolism could be underrepresented and relevant reactions could be missing or misrepresented. Therefore, the generation of core models from GSMMs ensures a proper representation of case-specific biomass equation. In addition, these core models may include additional reactions that would

be significant for metabolic flux estimation. The generated model obtained from the full genome-scale model included redox cofactor balances. Although cofactors were not used in metabolic flux adjustment, the obtained flux distribution satisfied electron and reduction equivalents balance. Finally, we demonstrate the feasibility of using core models obtained from the reduction of GSMM for ^{13}C-MFA. On the one hand, these models contain extended information of the metabolic network than 'classical' core metabolic models do. And, on the other hand, these models are more suitable to be used than whole genome-scale models. GSMM contain a huge number of variables, degrees of freedom and alternatives that require large computational times and amounts of experimental information to be solved, which are unfeasible for routine ^{13}C-MFA experiments.

5.5. REFERENCES

1   Solà, A. *et al.* (2004) Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. *Eur. J. Biochem.* 271, 2462–70

2   Jordà, J. *et al.* (2013) Glucose-methanol co-utilization in *Pichia pastoris* studied by metabolomics and instationary 13C flux analysis. *BMC Syst. Biol.* 7, 17

3   Solà, A. *et al.* (2007) Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. *Microbiology* 153, 281–90

4   Jordà, J. *et al.* (2014) Metabolic flux analysis of recombinant *Pichia pastoris* growing on different glycerol/methanol mixtures by iterative fitting of NMR-derived 13C-labelling data from proteinogenic amino acids. *N. Biotechnol.* 31, 120–132

5   Baumann, K. *et al.* (2010) A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions. *BMC Syst. Biol.* 4, 141

6   Gopalakrishnan, S. and Maranas, C. (2015) Achieving Metabolic Flux Analysis for *S. cerevisiae* at a Genome-Scale: Challenges, Requirements, and Considerations. *Metabolites* 5, 521–535

7   Gopalakrishnan, S. and Maranas, C.D. (2015) 13C metabolic flux analysis at a genome-scale. *Metab. Eng.* 32, 12–22

8   Erdrich, P. *et al.* (2015) An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. *BMC Syst. Biol.* 9, 48

9   Hädicke, O. and Klamt, S. (2017) EColiCore2: a reference network model of the central metabolism of *Escherichia coli* and relationships to its genome-scale parent model. *Sci. Rep.* 7, 39647

10   Maaheimo, H. *et al.* (2001) Central carbon metabolism of Saccharomyces cerevisiae explored by biosynthetic fractional ^{13}C labeling of common amino acids. *Eur. J. Biochem.* 268, 2464–2479

11   Jordà, J. *et al.* (2014) Quantitative Metabolomics and Instationary 13C-Metabolic Flux Analysis Reveals Impact of Recombinant Protein Production on Trehalose and Energy Metabolism in *Pichia pastoris*. *Metabolites* 4, 281–299

12   Jordà, J. *et al.* (2012) Metabolic flux profiling of recombinant protein secreting *Pichia pastoris* growing on glucose:methanol mixtures. *Microb. Cell Fact.* 11, 57

13   Quek, L.-E. *et al.* (2009) OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis. *Microb. Cell Fact.* 8, 25

14   Baumann, K. *et al.* (2008) Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins. *Biotechnol. Bioeng.* 100, 177–183

15   Liu, X. *et al.* (2013) Expression and functional studies of genes involved in transport and metabolism of glycerol in Pachysolen tannophilus. *Microb. Cell Fact.* 12, 27

16   Knudsen, P.B. (2015) , Development of scalable high throughput fermentation approaches for physiological characterisation of yeast and filamentous fungi. , Technical University of Denmark

17   Fernandez, C.A. *et al.* (1996) Correction of ^{13}C Mass Isotopomer Distributions for Natural Stable Isotope Abundance. *J. Mass Spectrom.* 31, 255–262

18   Van Winden, W.A. *et al.* (2002) Correcting mass isotopomer distributions for naturally occurring isotopes. *Biotechnol. Bioeng.* 80, 477–479

19   Christensen, B. and Nielsen, J. (1999) Isotopomer Analysis Using GC-MS. *Metab. Eng.* 1, 282–290

20   Ferrer, P. and Albiol, J. (2014) ^{13}C-Based Metabolic Flux Analysis in Yeast: The *Pichia pastoris* Case. In *Yeast Metabolic Engineering* 1152 (Mapelli, V., ed), pp. 209–232, Springer New York

21   Noorman, H.J. *et al.* (2000) Classification, error detection, and reconciliation of process information in complex biochemical systems. *Biotechnol. Bioeng.* 49, 364–376

22   Steffen, K. *et al.* (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst. Biol.* 1, 1–13

23   Caspi, R. *et al.* (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* 42, 459–471

24   Antoniewicz, M.R. *et al.* (2007) Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions. *Metab. Eng.* 9, 68–86

25   Antoniewicz, M.R. *et al.* (2006) Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metab. Eng.* 8, 324–337

26   Tomàs-Gamisans, M. *et al.* (2016) 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* 11, e0148031

27   Pirt, S.J. (1982) Maintenance energy: a general model for energy-limited and energy-sufficient growth. *Arch. Microbiol.* 133, 300–302

28   Stephanopoulos, G.N. *et al.* (1998) CHAPTER 2 - Review of Cellular Metabolism BT - Metabolic Engineering. pp. 21–79, Academic Press

29   Taccari, M. *et al.* (2012) Screening of yeasts for growth on crude glycerol and optimization of biomass production. *Bioresour. Technol.* 110, 488–495

30   Nocon, J. *et al.* (2014) Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. *Metab. Eng.* 24, 129–138

31   Fiaux, J. *et al.* (2003) Metabolic-Flux Profiling of the Yeasts *Saccharomyces cerevisiae* and *Pichia stipitis*. *Eukaryot. Cell* 2, 170–180

32   Gancedo, J.M. (1998) Yeast carbon catabolite repression. *Microbiol. Mol. Biol. Rev.* 62, 334–61

33   Heyland, J. *et al.* (2010) Quantitative physiology of *Pichia pastoris* during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. *Biotechnol. Bioeng.* 107, 357–68

34   Celik, E. *et al.* (2010) Metabolic flux analysis for recombinant protein production by *Pichia pastoris* using dual carbon sources: Effects of methanol feeding rate. *Biotechnol. Bioeng.* 105, 317–29

35   Bakker, B.M. *et al.* (2001) Stoichiometry and compartmentation of NADH metabolism in Saccharomyces cerevisiae. *FEMS Microbiol. Rev.* 25, 15–37

36   Kyle, M. *et al.* (2006) Coupling of growth rate and body stoichiometry in Daphnia: A role for maintenance processes? *Freshw. Biol.* 51, 2087–2095

37   Elser, J.J. *et al.* (2003) Growth rate-stoichiometry couplings in diverse biota. *Ecol. Lett.* 6, 936–943

38   Franklin, O. *et al.* (2011) Optimization of biomass composition explains microbial growth-stoichiometry relationships. *Am. Nat.* 177, E29--42

39   Stephanopoulos, G. *et al.* (1998) Review of Cellular Metabolism. In *Metabolic Engineering* 16pp. 21–79, Elsevier

40   Bakker, B.M. *et al.* (2001) Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* 25, 15–37

41   Blank, L.M. *et al.* (2005) Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *FEMS Yeast Res.* 5, 545–58

42   Nöh, K. and Wiechert, W. (2011) The benefits of being transient: Isotope-based metabolic flux analysis at the short time scale. *Appl. Microbiol. Biotechnol.* 91, 1247–1265

APPENDIX III

File S5-1. Contains the original (*PpaGS*), reduced (*PpaPruned*) and compressed model (*PpaCore*) in Cell Net Analyzer format [1]. In addition, the instructions followed for generating the reduced model are also included and described as in [2]. *PpaGS* model is derived from the original iMT1026 v3.0, but due to numerical tolerance constrains in model reduction, the 'cof_c' metabolite was removed from the specific biomass equation describing growth on glycerol.

Table S5-1. Reduced stoichiometric model used for ¹³C-MFA. Reactions in *PpaCore* that contained exclusively metabolites from the 'excluded metabolites' list, were omitted in the OpenFLUX model and subsequent metabolic flux adjustment. The following representation of the reactions and atom transition are in OpenFLUX [3] format. 'RxnID' corresponds to the ID of each reaction; 'rxnEq' refers to the equation of the reaction; 'rxnCTrans' represents the carbon atom transition of each reaction; 'rxnType' is the reaction classification according to OpenFLUX specifications [3].

RxnID	rxnEq	rxnCTrans	rxnType
v1	glyc_e = glyc_c	abc = abc	F
v2	glyc_c + atp_c + 1 nadp_c = dhap_c + 2 h_c + adp_c + 1 nadph_c	abc + X + X =cba + X + X + X	F
v3	g6p_c = f6p_c	abcdef = abcdef	FR
v4	f6p_c = g6p_c	abcdef = abcdef	R
v5	f6p_c + atp_c = fdp_c + h_c + adp_c	abcdef + X = abcdef + X + X	FR
v6	fdp_c + h2o_c = f6p_c + pi_c	abcdef + X = abcdef + X	R
v7	fdp_c = dhap_c + g3p_c	abcdef = cba + def	FR
v8	dhap_c + g3p_c = fdp_c	abc + def = cbadef	R
v9	dhap_c = g3p_c	abc = cba	FR
v10	g3p_c = dhap_c	abc = cba	R
v11	g3p_c + pi_c + nad_c = 13dpg_c + h_c + nadh_c	abc + X + X = abc + X + X	F
v12	13dpg_c + adp_c = 3pg_c + atp_c	abc + X = abc + X	F
v13	3pg_c = 2pg_c	abc = abc	F
v14	2pg_c = pep_c + h2o_c	abc = abc + X	F
v15	pep_c + h_c + adp_c = pyr_c + atp_c	abc + X + X= abc + X	F
v16	pyr_c + hco3_c + atp_c = oaa_c + h_c + pi_c + adp_c	abc + d + X = abcd + X + X + X	F
v17	pyr_c + 1 h_c = pyr_m + 1 h_m	abc + X = abc + X	F
v18	1 g6p_c + 1 h2o_c + nadp_c + nadp_c = co2_c + ru5p_D_c + h_c + h_c + nadph_c + nadph_c	abcdef + X + X + X = a + bcdef + X + X + X + X	F
v19	ru5p_D_c = xu5p_D_c	abcde = abcde	FR
v20	xu5p_D_c = ru5p_D_c	abcde = abcde	R
v21	r5p_c = ru5p_D_c	abcde = abcde	FR
v22	ru5p_D_c = r5p_c	abcde = abcde	R
v23	xu5p_D_c + e4p_c = f6p_c + g3p_c	abcde + fghi = abfghi + cde	FR
v24	f6p_c + g3p_c = xu5p_D_c + e4p_c	abfghi + cde = abcde + fghi	R
v25	r5p_c + xu5p_D_c = g3p_c + s7p_c	abcde + fghij = cde + abfghij	FR
v26	g3p_c + s7p_c = r5p_c + xu5p_D_c	cde + abfghij = abcde + fghij	R
v27	g3p_c + s7p_c = f6p_c + e4p_c	abc + defghij = defabc + ghij	FR
v28	f6p_c + e4p_c = g3p_c + s7p_c	defabc + ghij = abc + defghij	R
v29	pyr_m + h_m + thmpp_m = co2_m + 2_Hydroxyethyl_ThPP_m	abc + X + X = a + bc	F

v30	1 2_Hydroxyethyl_ThPP_m + coa_m + 1 lpam_m = accoa_m + dhlam_m + 1 thmpp_m	ab + X + X = ab + X + X	F
v31	accoa_m + oaa_m + h2o_m= cit_m + h_m + coa_m	ab + cdef + X= fedbac + X + X	F
v32	cit_m = icit_m	abcdef = abcdef	F
v33	icit_m + nad_m = co2_m + akg_m + nadh_m	abcdef + X = f + abcde + X	F
v34	akg_m + h_m + lpam_m = co2_m + sdhlam_m	abcde + X + X = a + bcde	F
v35	sdhlam_m + coa_m = succoa_m + dhlam_m	abcd + X = abcd + X	F
v36	succoa_m + pi_m + adp_m = succ_m + atp_m + coa_m	abcd + X + X = abcd + X + X	F
v37	succ_m + fad_m = 0.5 fum_m + 0.5 fum_m + fadh2_m	abcd + X = 0.5 abcd + 0.5 dcba + X	F
v38	fum_m + h2o_m = mal_L_m	abcd + X = abcd	F
v39	mal_L_m + nad_m= oaa_m + h_m + nadh_m	abcd + X = abcd + X + X	F
v40	mal_L_m + nadp_m = pyr_m + co2_m + nadph_m	abcd + X = abc + d + X	F
v41	accoa_x + oaa_x + h2o_x = cit_x + coa_x + h_x	ab + cdef + X = fedbac + X + X	F
v42	1 cit_x + succ_x = succ_c + 1 icit_x	abcdef + ghij = ghij + abcdef	F
v43	icit_x = glx_x + 0.5 succ_x + 0.5 succ_x	abcdef = ab + 0.5 edcf + 0.5 fcde	F
v44	accoa_x + glx_x + h2o_x = mal_L_x + h_x + coa_x	ab + cd + X = cdba + X + X	F
v45	oaa_c + h_c + nadh_c = mal_L_c + nad_c	abcd + X + X = abcd + X	F
v46	ac_x + coa_x + atp_x= accoa_x + ppi_x + amp_x	ab + X + X = ab + X + X	F
v47	oaa_c + h_c = oaa_m + h_m	abcd + X = abcd + X	F
v48	mal_L_c + succ_m = mal_L_m + succ_c	abcd + efgh = abcd + efgh	FR
v49	mal_L_m + succ_c = mal_L_c + succ_m	abcd + efgh = abcd + efgh	R
v50	mal_L_c + akg_m = akg_c + mal_L_m	abcd + efghi = efghi + abcd	F
v51	cit_c = cit_x	abcdef = abcdef	FR
v52	cit_x = cit_c	abcdef = abcdef	R
v53	mal_L_x = mal_L_c	abcd = abcd	FR
v54	mal_L_c = mal_L_x	abcd = abcd	R
v55	glu_L_c + asp_L_m = glu_L_m + asp_L_c	abcde + fghi = abcde + fghi	F
v56	glu_L_m + oaa_m = akg_m + asp_L_m	abcde + fghi = abcde + fghi	FR
v57	akg_m + asp_L_m = glu_L_m + oaa_m	abcde + fghi = abcde + fghi	R
v58	mal_L_m + cit_c = mal_L_c + cit_m	abcd + efghij = abcd + efghij	FR
v59	mal_L_c + cit_m = mal_L_m + cit_c	abcd + efghij = abcd + efghij	R
v60	oaa_c + mal_L_x = mal_L_c + oaa_x	abcd + efgh = efgh + abcd	FR
v61	mal_L_c + oaa_x = oaa_c + mal_L_x	abcd + efgh = efgh + abcd	R
v62	akg_c + asp_L_c = glu_L_c + oaa_c	abcde + fghi = abcde + fghi	FR
v63	glu_L_c + oaa_c = akg_c + asp_L_c	abcde + fghi = abcde + fghi	R
v64	co2_m = co2_c	a = a	F
v65	co2_c + h2o_c = hco3_c + h_c	a + X = a + X	FR
v66	hco3_c + h_c = co2_c + h2o_c	a + X = a + X	R
v67	co2_c = co2_e	a = a	F
v68	352.2735 h2o_c + 0.82747 h2o_m + 2.4843 pyr_m + 3.331 2_Hydroxyethyl_ThPP_m + 16.665 akg_c + 1.0393 o2_c + 400.9972 atp_c + 1.0273 atp_m + 8.3136 nad_c + 2.0324 nad_m + 2.7626 nadph_m + 40.1303 nadph_c + 1.7889 r5p_c + 1.4687 pyr_c + 2.4637 pep_c + 5.4843 accoa_m + 0.22139 ppi_x + 0.22139 amp_x + 0.58824 q6_m + 1.5463 oaa_x + 0.22139 h_x + 5.6454 hco3_c + 1.2319 cit_m + 7.0936 asp_L_c + 31.3156 nh4_e + 1.27 pi_e + 2.2397 f6p_c + 0.085146 dhap_c + 8.8503 g6p_c + 3.4521 3pg_c + 1.2319 e4p_c + 0.22139 so4_e = 328.7142 h_c + 419.6639 pi_c + 0.61675 h_m + 2.2716 pi_m + 8.929 co2_c + 0.22139 h2o_x + 3.331 thmpp_m + 4.2415 co2_m + 7.0936 glu_L_c + 400.9972 adp_c + 1.0273 adp_m + 8.3136 nadh_c + 40.1303 nadp_c + 2.7626 nadp_m + 2.0324 nadh_m + 5.4843 coa_m + 2.364 mal_L_c + 0.22139 atp_x + 0.22139 ac_x + 1.874 akg_m + 0.58824 q6h2_m + 1.5463 mal_L_x + 32.8139 h_e + 0.20062 g3p_c + 1.2319 cit_c + 5.0482 biomass_c		B
v69	biomass_c = biomass_e		B
v70	pyr_m = ALA	abc = abc	S

v71	SER + THF = GLY + MTHF		abc + X = ab + c	S
v72	pyr_m + pyr_m = KVAL + CO2		abc + def = abefc + d	S
v73	KVAL = VAL		abcde = abcde	S
v74	KVAL + accoa_m = CO2 + LEU		abcde + fg = a + fgbcde	S
v75	pyr_m + oaa_c = ILE + CO2		abc + defg = debfgc + a	S
v76	akg_m = PRO		abcde = abcde	S
v77	oaa_c = THR		abcd = abcd	S
v78	3pg_c =SER		abc = abc	S
v79	oaa_c = ASX		abcd = abcd	SF
v80	oaa_m = ASX		abcd = abcd	SF
v81	pep_c + e4p_c = SKM		abc + defg = abcdefg	S
v82	SKM + pep_c = CHOR		abcdefg + hij = abcdefghij	S
v83	CHOR = CO2 + PHE		abcdefghij = a + hijgfedcb	S
v84	akg_m + accoa_m = LYS + CO2		abcde + fg = fgbcde + a	S
v85	akg_c = GLX		abcde = abcde	SF
v86	akg_m = GLX		abcde = abcde	SF
##	**excludedMetabolites**	**##**	**simulatedMDVs**	
#	atp_c	#	ALA#011	
#	nadp_c	#	ALA#111	
#	h_c	#	GLY#01	
#	adp_c	#	GLY#11	
#	nadph_c	#	VAL#01111	
#	h2o_c	#	VAL#11000	
#	pi_c	#	LEU#011111	
#	nad_c	#	ILE#011111	
#	nadh_c	#	PRO#01111	
#	h_m	#	ASX#0100	
#	thmpp_m	#	ASX#0111	
#	coa_m	#	ASX#1111	
#	lpam_m	#	GLX#01111	
#	dhlam_m	#	GLX#11000	
#	h2o_m	#	GLX#11111	
#	nadp_m	#	PHE#011111111	
#	nadph_m	#	PHE#110000000	
#	pi_m	#	LYS#011111	
#	adp_m			
#	atp_m	**##**	**inputSubstrates**	
#	fad_m	#	glyc_e	
#	fadh2_m			
#	nad_m			
#	nadh_m			
#	h2o_x			
#	coa_x			
#	h_x			
#	atp_x			
#	ppi_x			
#	amp_x			
#	h2o_e			
#	o2_e			
#	o2_m			
#	o2_c			
#	q6h2_m			
#	q6_m			
#	h_i			
#	mlthf_c			
#	methf_c			
#	glyc_e			
#	nh4_e			

#	pi_e
#	so4_e
#	co2_e
#	h_e
#	biomass_e

Table S5-2. Corrected MIDs for each experimental replicate and corresponding analysed peaks from each corresponding derivatisation method. MID of those isotopomers analysed by both derivatised methods were averaged. ECF ion cluster and DMFDMA ion cluster columns represent the m = 0 ion analysed for each isotopomer according to [4].

	0.05A	0.05B	0.10A	0.10B	0.16A	0.16B	ECF ion cluster	DMFDMA ion cluster
ALA#011	0.753	0.745	0.731	0.751	0.748	0.753	116	99
	0.235	0.245	0.256	0.246	0.249	0.235		
	0.012	0.010	0.013	0.003	0.006	0.012		
ALA#111	0.725	0.706	0.700	0.725	0.702	0.700	-	158
	0.121	0.138	0.133	0.122	0.119	0.128		
	0.150	0.154	0.161	0.150	0.167	0.165		
	0.004	0.002	0.006	0.004	0.012	0.008		
GLY#01	0.968	0.972	0.977	0.975	0.969	0.971	-	85
	0.032	0.028	0.023	0.025	0.031	0.029		
GLY#11	0.780	0.772	0.781	0.775	0.771	0.762	175	144
	0.215	0.229	0.220	0.220	0.222	0.215		
	0.005	0.000	0.000	0.005	0.008	0.022		
VAL#01111	0.628	0.612	0.616	0.620	0.613	0.626	144	127
	0.320	0.331	0.330	0.329	0.335	0.325		
	0.050	0.051	0.051	0.048	0.047	0.047		
	0.002	0.004	0.003	0.003	0.004	0.002		
	0.000	0.002	0.000	0.000	0.001	0.000		
VAL#11000	0.793	0.774	0.779	0.775	0.771	0.772	-	143
	0.203	0.216	0.220	0.212	0.223	0.217		
	0.004	0.010	0.001	0.014	0.006	0.011		
LEU#011111	0.530	0.514	0.504	0.515	0.511	0.515	158	-
	0.369	0.380	0.385	0.381	0.376	0.380		
	0.091	0.096	0.100	0.095	0.111	0.094		
	0.009	0.010	0.010	0.009	0.002	0.010		
	0.000	0.000	0.000	0.000	0.000	0.001		
	0.000	0.000	0.000	0.000	0.000	0.000		
ILE#011111	0.575	0.559	0.563	0.562	0.573	0.570	158	-
	0.314	0.326	0.322	0.321	0.327	0.325		
	0.096	0.099	0.099	0.101	0.095	0.095		
	0.014	0.015	0.016	0.016	0.005	0.009		
	0.000	0.001	0.000	0.000	0.000	0.002		
	0.000	0.000	0.000	0.000	0.000	0.000		
PRO#01111	0.648	0.630	0.637	0.637	0.618	0.622	142	-
	0.302	0.316	0.310	0.311	0.327	0.316		
	0.048	0.051	0.050	0.049	0.053	0.056		
	0.002	0.002	0.002	0.002	0.002	0.005		
	0.000	0.000	0.000	0.000	0.000	0.000		
ASX#0100	0.894	0.884	0.886	0.876	0.888	0.898		115
	0.106	0.116	0.114	0.124	0.112	0.102		
ASX#0111	0.702	0.691	0.690	0.697	0.682	0.699	188	-
	0.227	0.236	0.235	0.230	0.244	0.231		
	0.069	0.071	0.073	0.071	0.072	0.068		
	0.002	0.003	0.003	0.002	0.002	0.002		

ASX#1111	0.631	0.616	0.630	0.599	0.626	0.638	-	216
	0.242	0.243	0.216	0.221	0.211	0.221		
	0.106	0.122	0.125	0.130	0.130	0.121		
	0.019	0.020	0.021	0.032	0.031	0.019		
	0.001	0.000	0.009	0.018	0.001	0.001		
GLX#01111	0.584	0.565	0.565	0.790	0.579	0.589	202	-
	0.355	0.351	0.367	0.179	0.357	0.350		
	0.058	0.087	0.065	0.030	0.060	0.058		
	0.003	0.004	0.003	0.003	0.004	0.003		
	0.000	0.000	0.000	0.000	0.000	0.000		
GLX#11000	0.736	0.732	0.734	0.734	0.726	0.745	-	143
	0.244	0.247	0.249	0.249	0.259	0.248		
	0.021	0.021	0.017	0.016	0.015	0.007		
GLX#11111	0.499	0.487	0.503	0.517	0.492	0.515	-	230
	0.360	0.366	0.354	0.349	0.346	0.352		
	0.120	0.123	0.122	0.113	0.128	0.113		
	0.019	0.022	0.019	0.019	0.032	0.015		
	0.002	0.002	0.002	0.002	0.002	0.002		
	0.000	0.000	0.000	0.000	0.000	0.003		
PHE#01111111	0.422	0.399	0.411	0.416	0.398	0.427	192	175
	0.305	0.322	0.314	0.312	0.319	0.308		
	0.172	0.177	0.178	0.177	0.179	0.172		
	0.077	0.095	0.078	0.076	0.078	0.076		
	0.018	0.006	0.017	0.017	0.020	0.016		
	0.005	0.001	0.002	0.002	0.004	0.002		
	0.000	0.000	0.000	0.000	0.001	0.000		
	0.000	0.000	0.000	0.000	0.000	0.000		
	0.000	0.000	0.000	0.000	0.000	0.000		
PHE#1100000	0.740	0.747	0.754	0.741	0.756	0.753	-	143
	0.225	0.232	0.231	0.229	0.228	0.230		
	0.035	0.021	0.015	0.030	0.016	0.016		
LYS#01111	0.533	0.512	0.508	0.512	0.481	0.516	156	-
	0.374	0.376	0.375	0.377	0.389	0.370		
	0.085	0.099	0.101	0.099	0.117	0.102		
	0.008	0.012	0.014	0.012	0.015	0.011		
	0.000	0.001	0.002	0.001	0.000	0.001		
	0.001	0.000	0.000	0.000	0.000	0.000		

References

1. Klamt, S. and von Kamp, A. (2011) An application programming interface for CellNetAnalyzer. BioSystems 105, 162–168
2. Erdrich, P. et al. (2015) An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. BMC Syst. Biol. 9, 48
3. Quek, L.-E. et al. (2009) OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis. Microb. Cell Fact. 8, 25
4. Christensen, B. and Nielsen, J. (1999) Isotopomer Analysis Using GC-MS. Metab. Eng. 1, 282–290

6

Redox engineering by ectopic overexpression of NADH kinase in recombinant *P. pastoris*

TABLE OF CONTENTS

6. Redox engineering by ectopic overexpression of NADH kinase in recombinant *Pichia pastoris*

6.1. INTRODUCTION

Despite *P. pastoris* has been extensively used for recombinant production, there are still some limitations to overcome to further increase productivities. High-level expression and secretion of heterologous proteins has been reported to cause a metabolic burden that could significantly impact on energy metabolism and alter the central carbon metabolism flux distribution [1,2]. Producing strains may not cope with the additional demand of ATP, NADPH and precursors for *de novo* biosynthesis of amino acids, thereby leading to a suboptimal cell fitness and reduced production yields [3]. In addition, the folding and secretion processes of complex proteins is very resource-intensive, particularly of NADPH, which is required for disulphide bound formation and alleviating ER oxidative stress [4]. Thus, overproduction of recombinant proteins would result in imbalanced redox cofactor and, specifically, a reduction in NADPH availability. Such alterations in redox cofactor balance have a strong impact on cell metabolism [5]. Therefore, strain engineering strategies targeting redox metabolism have been successfully applied to improve *E. coli* [6,7], *S. cerevisiae* [8,9] and *P. pastoris* [10] strains for a range of different applications.

NADPH availability is tightly related to biomass yields and recombinant protein production yields [11]. Driouch *et al.* [12] reported that *Aspergillus niger* strains overproducing recombinant proteins show higher fluxes through the oxidative branch of pentose phosphate pathway, which is the main cytosolic NADPH generation pathway. Also, Nocon *et al.* [13] overexpressed genes coding for enzymes of the oxidative branch of pentose phosphate pathway, obtaining higher productivities in heterologous protein secretion. Indeed, a preceding study identified several metabolic engineering targets for improving recombinant protein production using a genome-scale metabolic model [14]. Interestingly, about the 50% of the identified targets pointed towards NADPH supply reactions [15].

Based on the key metabolic role of NADPH on protein synthesis and secretion and previous studies pointing at the positive effects that its increased supply appears to have on recombinant protein, we have investigated the impact of the overexpression of a heterologous NADH kinase on a *P. pastoris* strain producing an antibody fragment (Fab). Previous studies in our research group reported an increase of Fab specific productivity under reduced oxygen availability conditions [16]. In addition, the reduced oxygen availability for electron transport chain leads to higher NADH/NAD$^+$ ratios under hypoxic conditions [17]. This is concomitant with a shift to a respiro-fermentative metabolism, as reflected in the generation of ethanol and arabitol for cofactor reoxidation [18]. In this context, we postulate that the NADH excess found under hypoxic conditions could be a potential source of electrons for NADPH production and, therefore, the effects of the NADH kinase overexpression might be enhanced under hypoxia in comparison to the reference normoxic condition. In order to test our hypothesis, redox-engineered strains were grown on glycerol and glucose under normoxic conditions as well as on glucose under hypoxic conditions. Overall, we aimed to investigate the combined effect of a process strategy (hypoxic conditions) and metabolic engineering strategy to improve recombinant protein production.

Moreover, we have used the genome-scale metabolic model to evaluate the experimental physiological datasets obtained in chemostat cultivations and assisting the metabolic interpretation of the observed macroscopic changes.

6.2. MATERIALS AND METHODS

6.2.1. Strain generation

A Pichia pastoris X-33-derived strain expressing multiple copies of the genes encoding the human antigen-binding fragment (Fab) 2F5 under the transcriptional control of the constitutive GAP promoter and with the α-mating factor secretion signal sequence from *Saccharomyces cerevisiae* [19] was used in this study.

S. cerevisiae POS5 gene, encoding for the mitochondrial NADH kinase Pos5p [20,21] was codon optimised for heterologous expression in *P. pastoris* and synthetised by Geneart (ThermoFischer Scientific), cloned into a pPUZZLE vector [22] under the control of GAP promoter, thereby generating vector pPUZZLE_mPOS5 (**Fig. 6-1**). Similarly, an analogous construction, pPUZZLE_cPOS5, was constructed expressing a 5'-truncated *POS5* excluding the first 48 bp coding for the N-terminal 16 amino acids, allowing for cytosolic Pos5p localisation. *Escherichia coli* DH5α was used for plasmid propagation.

P. pastoris X-33/2F5 strain transformation and recombinant clone isolation were performed as by described [22]. The presence of insert was confirmed by colony PCR [23] using the primer pairs described in **Table S6-1**.
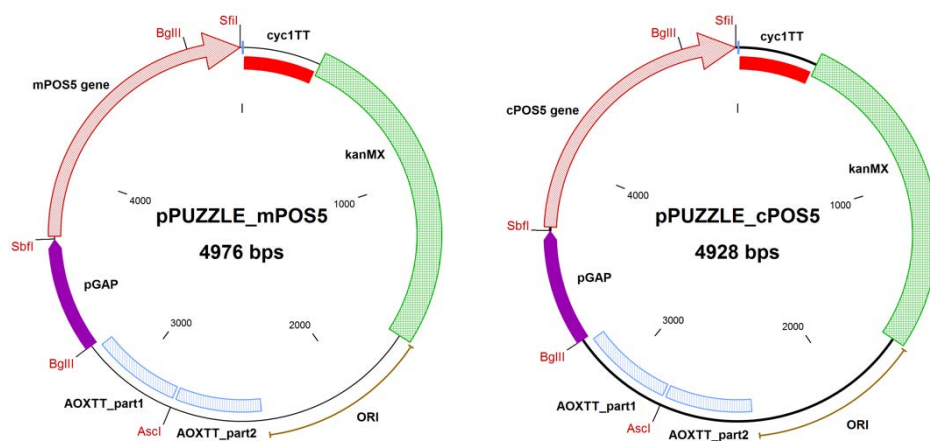
Fig. 6-1. Plasmid maps for pPUZZLE_mPOS5 and pPUZZLE_cPOS5. In red, the restriction enzymes used: SbfI and SfiI for cloning mPOS5 and cPOS5; AscI for plasmid linearization; BglII for plasmid ligation verification. pPUZZLE contains the kanMX gene encoding for kanamycin resistance (*E. coli*) and geneticin G418 resistance (*P. pastoris*).

6.2.2. Clone screening at small scale

A set of 12 recombinant clones for each strain construct were screened for growth and Fab 2F5 production in baffled shake flasks with glucose minimal medium as described by Baumann *et al.* [22].

6.2.3. Chemostat cultivations

Two independent series of carbon-limited chemostat cultivations were performed for each strain in three different growing conditions: using glycerol as carbon source, glucose under normoxic conditions (100% air in the inlet gas composition) and glucose under hypoxic conditions (25:75 of air:N_2 in the inlet gas composition). Cultivations were performed at a working volume of 1-L in a 2-L bench-top Biostat B (B. Braun Biotech International) bioreactor. Operational conditions were set to 25°C, 700 rpm, 1 vvm inlet gas flow, 0.2 bar overpressure, 0.1 h^{-1} dilution rate (D) and pH 5.0 by addition of 15% (v/v) NH_4. Samples were taken at 3rd, 4th and 5th residence times for cell density monitoring and ELISA, HPLC and DCW analysis. The off-gases were cooled dawn in a condenser at 4°C and further desiccated in two silica gel columns. The off-gas O_2 and CO_2 concentrations were measured by BCP-O_2 (zirconium dioxide) and BCP-CO_2 (infrared) BlueSens Gas Analyser (Herten, Germany), respectively.

For reactor inoculation, strains were cultivated in 1-L baffled Erlenmeyer flask containing 150 mL YPG broth (1% w/v yeast extract, 2% w/v peptone, 1% w/v glycerol) and antibiotic (100 μg·L^{-1} zeocin for control strain or 500 μg·L^{-1} geneticin G418 for NADH kinase recombinant clones) at an optical density (OD$_{600}$) between 0.15 and 0.30. Pre-cultures were incubated at 25°C under 130 rpm for 16-24 h.

Batch medium content was previously described [16]. Briefly, it contained: 40 g·L^{-1} glycerol, 1.8 g·L^{-1} citric acid, 12.6 g·L^{-1} $(NH_4)_2HPO_4$, 0.5 g·L^{-1} $MgSO_4.7H_2O$, 0.9 g·L^{-1} KCl, 0.02 g·L^{-1} $CaCl_2.2H_2O$, 4.6 mL·L^{-1} trace salts stock solution, 2 mL·L^{-1} of biotin solution (0.2 g·L^{-1}) and 250 μL·L^{-1} of Glanapon 2000 antifoam (Bussetti, Vienna). Chemostat medium was also adapted from [16]. Briefly, media contained: 50 g·L^{-1} carbon source – glycerol, glucose – , 0.92 g·L^{-1} monohydrate citric acid; 4.35 g·L^{-1} $(NH_4)_2HPO_4$; 0.65 g·L^{-1} $MgSO_4.7H_2O$; 1.7 g·L^{-1} KCl; 0,01 g·L^{-1} $CaCl_2.2H_2O$; 1.6 mL trace salt solution, 1 mL biotin solution (0,2 g·L^{-1}) and

200 µL·L^{-1} Glanapon antifoam. Trace salt solution was composed of: 6.0 g·L^{-1} CuSO$_4$.5H$_2$O; 0.08 g·L^{-1} NaI; 3.36 g·L^{-1} MnSO$_4$.H$_2$O; 0.2 g·L^{-1} Na$_2$MoO$_4$.2H$_2$O; 0.02 g·L^{-1} H$_3$BO$_3$; 0.82 g·L^{-1} CoCl$_2$.6H$_2$O; 20 g·L^{-1} ZnCl$_2$; 65 g·L^{-1} FeSO$_4$.7H$_2$O and 5.0 mL H$_2$SO$_4$ (95-98% w/w). Media pH was set to 5.0 with 6 N HCl.

6.2.4. Analytical methods

Biomass concentration

Cell density was monitored by optical density in a DR3900 spectrophotometer (Hach, Bizcaia, Spain) at 600 nm. Dry cell weight (DCW) was measured by gravimetric methods. From 2 to 10 mL of sample were filtered in Glass Fibre Prefilters (Merk Millipore), previously pre-weighted after drying at 105°C for 24 h. Each filter was washed twice with 10 mL of distillated water; died at 105°C for 24 h, cooled in desiccator and weighted.

Fermentation products analysis

Citric acid, glucose, glycerol, arabitol, succinic acid, acetic acid and ethanol were analyzed by HPLC in an UltiMate 3000 Liquid Chromatography Systems (Dionex) using an ICSep ICE-COREGEL 87H3 (Transgenomic) ion exchange column and a Waters 2410 (Waters) refraction index detector. 6 mM sulphuric acid was used as continuous phase at 0.5 mL/min flow and 20 µL sample injection volume. Data were analysed in CROMELEON software (Dionex).

Quantification of Fab

Fab 2F5 was quantified by ELISA in 96-well Immuno Plates (Nunc, Thermo Scientific) as described by Gasser and co-workers [19]. Briefly, plates were subjected to an overnight pre-coating of Fab specific Anti-Human IgG (I5260, Sigma) primary antibody in PBS buffer (1:1000). Then, plates were washed three times with PBS 1% Tween 20 and samples and Fab standard (Bethyl Inc.) were diluted in PBS buffer containing 10% (w/v) BSA (Sigma) and 0.1% (v/v) Tween 80. Plates were incubated for 2 h, washed again with PBS 1% Tween 20 three times and incubated for 1 h after addition of Anti-Human Kappa Light Chains (bound)-Alkaline Phosphatase (Sigma) secondary antibody. Plates were washed with PBS 1% Tween 20 three times, treated with pNPP phosphatase substrate (Sigma) and the absorbance was measured at 405 nm using a Multiskan™ FC Microplate reader (Thermo Scientific).

ddPCR (droplet digital PCR) analysis for recombinant *POS5* gene copy determination

Genomic DNA was purified by Wizard Genomic DNA Purification Kit (Promega), according to manufacturer instructions, and quantified in NanoDrop 2000 Spectrophotometer (Thermo Scientific). 0,5 µg DNA were digested by EcoRI and BamHI FastDigest Enzymes (Thermo Scientific) to result in DNA fragments lower than 5 kb and purified with Wizard® SV Gel and PCR Clean-Up System (Promega). Reaction conditions (1X Supermix ddPCR TaqMan, 300 nM of each primer, 200 nM of each probe and 0.02 ng·µL⁻¹ digested genomic DNA) and operational conditions were performed as suggested by Biorad and optimised to *P. pastoris* by [24]. The annealing temperature was set to 57°C, after temperature gradient determination. Primers used in for ddPCR are described in (**Table S6-1**).

NADPH/NADP⁺ ratio determination

Samples for NADPH and NADP⁺ quantification were taken and rapidly quenched with cold 60% v/v methanol [25,26]. The solutions were centrifuged and washed twice with quenching solution as described in Ortmayr et al. [25] (4000 g, -10°C, 10 min in a Centrifuge 5804 R, Eppendorf). Finally, pellets were stored at -80°C. NADPH and NADP⁺ concentrations were determined using EnzyChrom™ NADP⁺/NADPH Assay Kit (BioAssay Systems) and the optical densities were measured by means of a Multiskan™ FC Microplate reader (Thermo Scientific) at a wavelength of 595 nm.

6.2.5. Statistical analysis

Chemostat cultivation data was checked for consistency and standard reconciliation procedures were applied [27]. In all the cultivation sets, statistical consistency test was passed with a confidence level of 95%. Consequently, there was no evidence of gross measurement errors. A statistical comparison of the macroscopic growth profiles of the different strains was performed by the Microsoft Excel 2-tailed Student's *t*-Test.

6.2.6. Metabolic Modelling

iMT1026 v3.0 metabolic model of *P. pastoris* [28] (developed in Chapter 4) was used in the COBRA Toolbox v2.0.6 under Matlab 2014 (Mathworks, USA) with SBML toolbox

v4.1.0 [29] and libSBML library v5.12.0 [30]. The prediction of flux redistribution due to Fab overexpression was performed employing Flux Scanning based on Enforced Objective Function (FSEOF) [31] by maximizing the biomass production at a constrained range of Fab secretion $(0 - 0.12 \text{ mg·g}_{DCW}^{-1}·h^{-1})$. Particularly, redox cofactor turnover rates were calculated using the flux-sum analysis [32] on each resulting flux distribution. A cytosolic NADH kinase reaction was incorporated into the model (not the mitochondrial reaction, iMT1026 v3.0 already contained the endogenous NADH kinase reaction). The perturbation of the NADH kinase activity on flux distribution was calculated with Minimization of Metabolic Adjustment (MOMA) [33] performing a series of simulations enforcing a minimal flux through the NADH kinase reaction (cytosolic or mitochondrial) constraining the uptake of carbon source to the control strain experimental values in the case of normoxia (glycerol and glucose), and additionally constraining the oxygen uptake rate for the simulations in hypoxic conditions. The resulting flux distributions at different NADH kinase reaction fluxes $(0 - 2 \text{ mmol·g}_{DCW}^{-1}·h^{-1})$ were compared against the control strain (X-33/2F5).

6.3. RESULTS

6.3.1. Cytosolic and mitochondrial overexpression of *POS5* and its effect on recombinant Fab secretion in shake flask cultures

The codon optimised *POS5* gene, either containing or not its native mitochondrial signal peptide, was integrated into a *P. pastoris* X-33 strain expressing a 2F5 antibody fragment under the control of the glycolytic GAP promoter, aiming for the co-overxpression of NADH kinase in the mitochondria (mPOS5) or in the cytosol (cPOS5), respectively. Individual clones of each co-overexpressing strain were verified for integration of m*POS5* or c*POS5* prior to the screening in baffled shake flasks.

Series of twelve verified clones overexpressing one of the *POS5* forms were used in a first small scale screening in shake flasks in triplicate experiments. Product titters and biomass were measured by ELISA and dry cell weight quantification, respectively to check the effect of co-expression of m*POS5* or c*POS5* on recombinant Fab secretion. The average Fab yields were normalised to those obtained from the reference strain X-33/2F5. The results of these preliminary screening experiments are shown in **Fig. 6-2A.** The overall picture of the co-

expression of the two *POS5* gene forms demonstrated a largely unchanged protein secretion capacity. c*POS5* clones showed two distinct populations: one with the same behaviour of the control strain (FC = 1.07 ± 0.21) and another with overproducing clones (FC = 2.34 ± 0.52). A plausible explanation for the observed clonal variation would be that isolated transformants differ in the dosage of the co-expressed c*POS5* gene. In order to test this hypothesis, the relative recombinant *POS5* gene dosage was determined by droplet digital PCR (ddPCR) for a representative clone population, one giving a clear increased Fab yield and the other no significant effect on product yield compared to the reference strain.
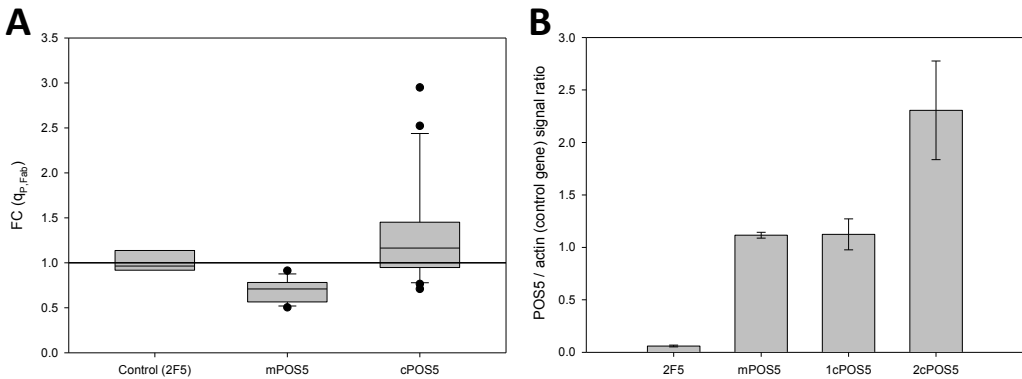
Fig. 6-2. Representation of the results of clone screening and selection. **(A)** Specific Fab production was measured and normalised to the reference strain X-33/2F5; **(B)** A representative clone of each population was analysed for determining POS5 gene copy number.

6.3.2. Physiological characterization of the co-overexpressing *POS5* strains growing in chemostat cultures

Strains X-33/FAB (reference), mPOS5 (mitochondrial expression of *POS5*), cPOS5 and 2cPOS5 (cytosolic expression with one and two copies of *POS5* respectively) were cultivated in carbon-limited chemostat cultures using glycerol or glucose as carbon source as well as two different oxygen availability conditions (normoxia and hypoxia) when using glucose to study the effects of cofactor perturbation on cell physiology and Fab production in different environmental conditions. Since Pos5p catalyses a NADPH-generating reaction, it may alter redox cofactor balance. In order to check whether *POS5* overexpression has an impact on the redox cofactor balance, the NADPH/NADP+ ratios were measured and calculated for all strains under the tested growth conditions (**Fig. 6-3**). The reference strain X-33/2F5 showed the lowest ratios in all the three conditions. Strains co-overexpressing m*POS5* and c*POS5* had comparable NADPH/NADP+ values to the reference strain when

growing on glycerol, whereas this value was 2-fold higher in the 2cPOS5 strain. Co-overexpression of *POS5* had a higher impact in cells growing on glucose, as shown in **Fig. 6-3B** and **Fig. 6-3C**. Both cPOS5 and mPOS5 strains showed a moderate increase in NADPH/NADP$^+$ ratios of about a 20% and 30% compared to the reference strain, respectively. As observed in glycerol-grown cells, the 2cPOS5 strain had a higher fold change in the NADPH/NADP$^+$ ratio during growth in glucose. When the oxygen supply was reduced, a similar response pattern was observed, but NADPH/NADP$^+$ ratios were comparatively higher for all four strains **Fig. 6-3**C. Such increase in the NADPH/NADP$^+$ ratio observed in hypoxic conditions may be a consequence of the excess or accumulation of NADH observed under these conditions [17], which is generally converted to ethanol or other by-products such as arabitol [16,34]. Thus, an increased availability of NADH would allow higher conversions to NADPH and consequently increasing NADPH/NADP$^+$ ratio.
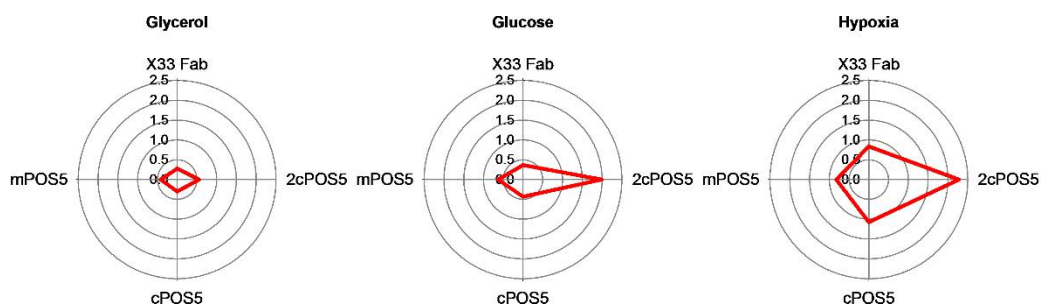
Fig. 6-3. Representation of NADPH/NADP$^+$ molar ratios in all the strains for growth on glicerol, glucose in normoxic conditions and hypoxic conditions.

These cofactor balance alterations due to *POS5* overexpression had a further impact on the physiological growth profile of the mutant strains (). Although such impact was not significant in all cases, some tendencies can be appreciated (**Fig. 6-5**). The specific oxygen uptake rate (q_{O_2}) showed a tendency to increase with the overexpression of *POS5* in all growth conditions for both cytosolic- and mitochondrial-directed expression. Notably, 2cPOS5 generally showed higher q_{O_2} than the other strains. These results would reflect an increased activity of the respiratory chain due to the additional demand of ATP that is consumed in the NADH kinase reaction, as discussed in section '6.4.3. *In silico* biological interpretation' and '6.5.3 Energy metabolism and by-product secretion'. Similarly, the specific CO_2 production rate (q_{CO_2}) increased with the NADH kinase activity. Specifically, q_{CO_2} was between 5% and 20% higher for all redox-engineered strains in normoxic conditions, showing the 2cPOS5

strain the highest q_{CO_2} in glucose. However, while cPOS5 and 2cPOS5 strains had the tendency of increasing q_{CO_2} both in normoxic and hypoxic conditions, mPOS5 showed reduced q_{CO_2} compared to the reference strain (**Fig. 6-5**). The variation in CO_2 specific production rates would be at expenses of biomass yields ($Y_{X,S}$), which show the opposite pattern: a statistically significant reduction for 2cPOS5 and an increase in mPOS5 in hypoxic conditions (**Table S6-2**).

As expected, by-product formation was detected only under respiro-fermentative conditions attained under hypoxia. Notably, redox-engineered strains had some significant alterations in the production profiles of these metabolites (**Fig. 6-5C**). In particular, while both cPOS5 and 2cPOS5 showed similar or higher arabitol and ethanol specific production rates, ethanol formation decreased in mPOS5 compared to the reference strain.
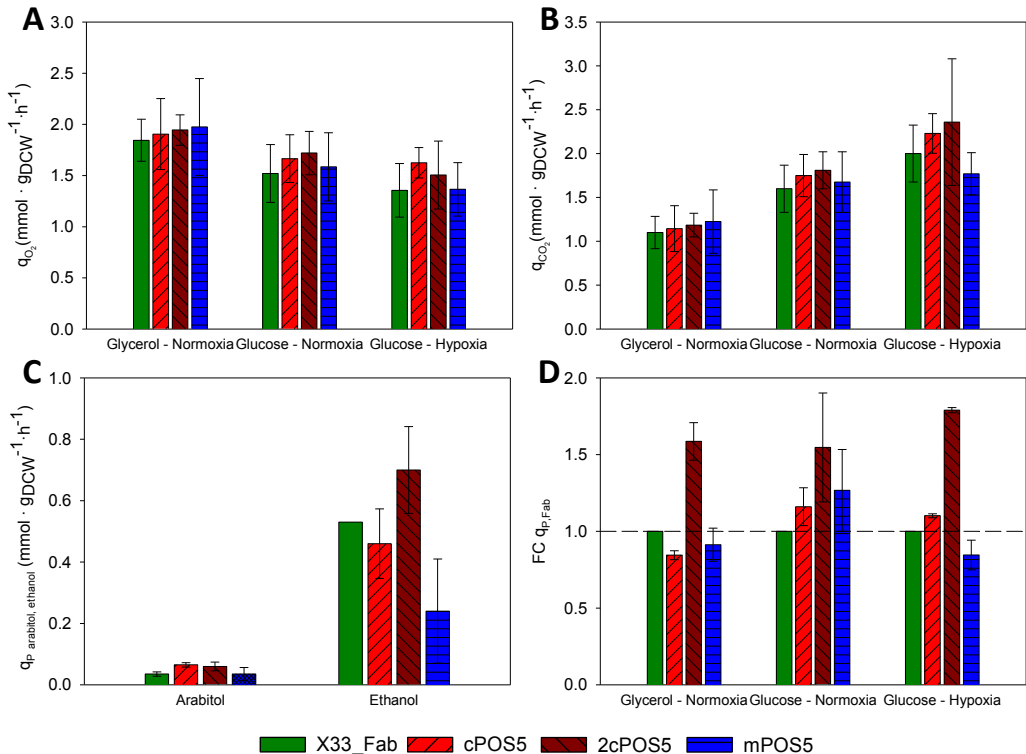
Fig. 6-5. Main macroscopic growth parameters for all strains and chemostat cultivation conditions. A: specific O_2 consumption rate; B: specific CO_2 production rate; C: specific by-product generation rate in glucose – hypoxic conditions; D: fold change of Fab productivity of the generated strains in comparison to the reference X-33/2F5 strain.

Fig. 6-4. Graphical representation of the predicted metabolic flux redistribution when overexpressing the NADH kinase in the cytosol (A-C) or mitochondria (D-F) growing in glycerol (A, D), glucose under normoxic conditions (B, E) and glucose under hypoxia (C, F). ▶
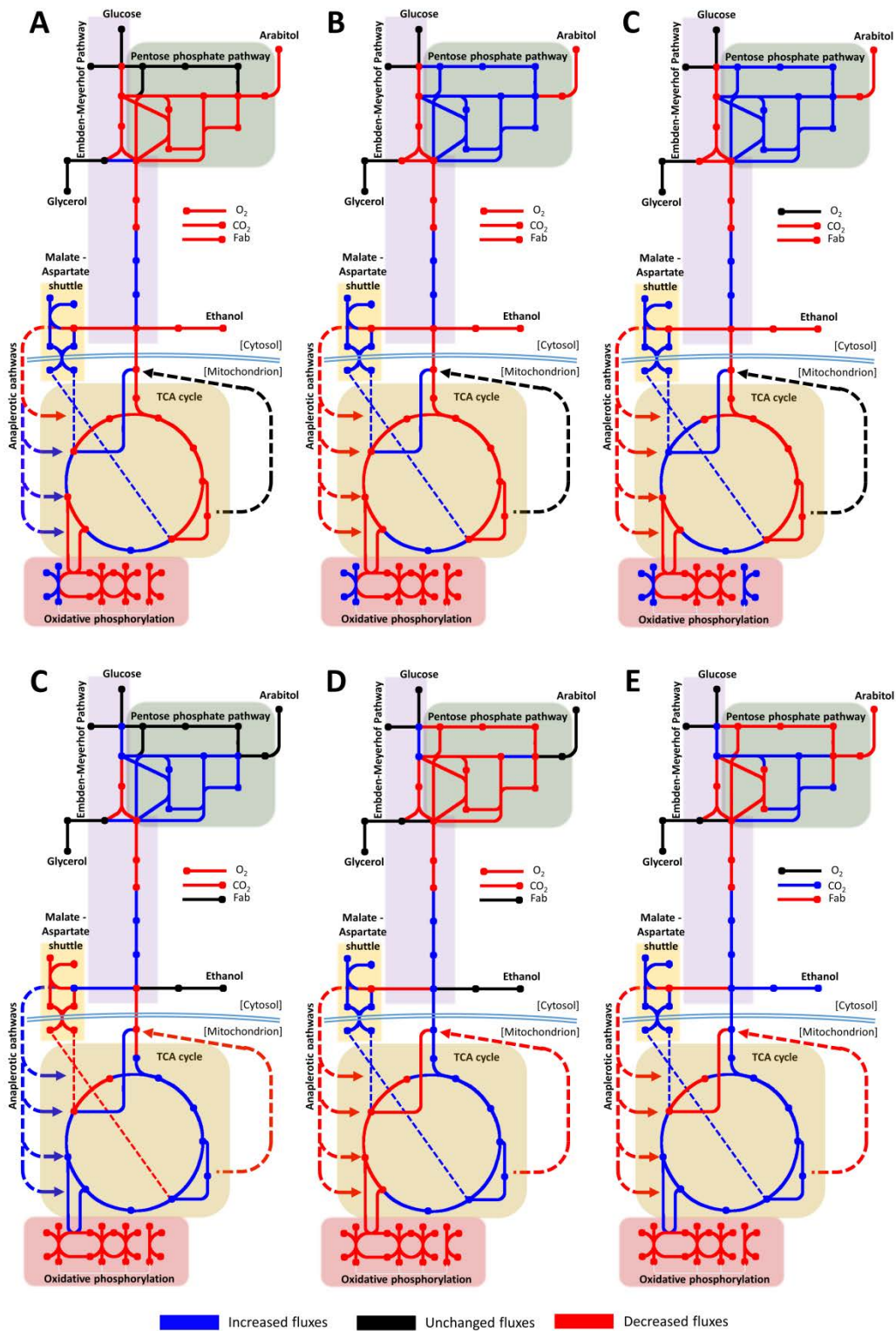
POS5 overexpression also had a clear impact on Fab production (**Fig. 6-5D**). When growing on glycerol, only 2cPOS5 improved significantly Fab productivity ($q_{P,Fab}$) with respect to the reference strain. However, all three redox-engineered strains showed significantly enhanced $q_{P,Fab}$ when growing on glucose under normoxic conditions, being 2cPOS5 the strain with the best performance (1.55-fold increase compared to the reference strain). In oxygen limiting conditions, although mPOS5 showed a decrease in productivity, the cytosolic *POS5* overexpressing strains improved up to 1.8-fold Fab production.

6.3.3. *In silico* biological interpretation

One of the applications of genome-scale metabolic models (GSMM) is to assist interpretation of biological data to reinforce or discard hypothesis. We used the iMT1026 v3.0 genome-scale metabolic model for *P. pastoris* **(Chapter 4),** that is, the third version of its consensus model [28].

The effect of Fab overproduction on flux distribution was tested by performing a series of simulations successively increasing Fab production constrains (employing FSEOF). NADPH turnover ratios were calculated for the resulting predicted fluxes. A clear correlation between Fab production and this cofactor turnover was obtained due to the increased demand of amino acid biosynthesis for Fab production (**Fig. S6-1**).

The metabolic impact of *POS5* overexpression was also modelled by constraining successively increasing fluxes through the specific NADH kinase reaction (cytosolic or mitochondrial), and using MOMA for assessing the new distribution. As a result, metabolic fluxes through some pathways are predicted to be up or downshifted concomitantly with an increased flux through NADH kinase reaction. Mitochondrial- or cytosolic-directed Pos5p overexpression impacts differently on metabolic flux distribution, as it is affecting different compartmentalised NAD(P)H pools. For example, cytosolic Pos5p overexpression supplies an important fraction of NADPH under glucose growing conditions, (**Fig. 6-4B, C**) and, consequently, the flux through the oxidative branch of the pentose phosphate pathway decreased. In contrast, mitochondrial NADPH kinase overexpression would not be able to supply enough cytosolic NADPH and therefore the flux of the NADPH generating reactions of the pentose phosphate pathway is enhanced (**Fig. 6-4D, E**). Another important predicted

difference between cytosolic and mitochondrial Pos5p overexpression is the increased activity of TCA cycle when the cytosolic NADPH kinase is overexpressed, while mitochondrial overexpression is predicted to cause a decreased flux through this pathway. However, some predicted changes are common between both c*POS5* and m*POS5* strains: in both cases, there is an increase of the activity of the oxidative phosphorylation that supplies the additional ATP demanded in the NADPH kinase reaction. The increased electron transfer to the respiratory chain is predicted to be through NADH oxidation in mPOS5 but, for cPOS5, succinate oxidation to fumarate would be the additional electron supply. Overall, an increase of oxygen consumption, CO_2 production, by-product formation as well as Fab production is predicted for all the conditions tested in simulations for the cPOS5 strain. Similarly, an increase of q_{O_2} and q_{CO_2} during growth on glycerol and glucose-normoxia is predicted for mPOS5, whereas no significant changes are predicted in terms of by-product formation or Fab production for this strain. Conversely, mitochondrial overexpression of NADH kinase leads to a reduction in CO_2, ethanol and Fab production in hypoxic conditions, coherent with the experimentally observed behaviour.

6.4. DISCUSSION

6.4.1. Increased NADPH availability enhances recombinant protein production

Several studies have addressed the impact of heterologous protein expression on metabolism and cell physiology of *P. pastoris* [1] and how environmental conditions can modulate the product yields [18,35]. Heyland et al. [11] postulated and Nocon et al. [13,14] and provided strong envidences that increasing flux through the PPP allows supplying additional NADPH, thereby compensating the extra demand caused by biosynthetic processes involved in heterologous protein production. In fact, our *in silico* simulations indicate an increase of Fab production correlates with higher NADPH turnover rates. Redox-engineered strains are able to generate additional NADPH compared to the reference X-33/2F5 strain, resulting in increased NADPH/NADP+ ratios. In addition, increased *POS5* gene dosage has a clear effect, likely yielding higher NADH kinase enzymatic activities and, subsequently, higher NADPH/NADP+ ratios in the 2cPOS5 strain. Furthermore, NADPH availability appears to be positively correlated to Fab productivity. The 2cPOS5 strain has the

highest NADPH/NADP$^+$ ratios and hence enhanced specific Fab production rates under all tested conditions. Higher NADPH/NADP$^+$ ratios may reflect more NADPH available for cellular processes, not only biosynthesis of recombinant protein, but also for its potential use in protein folding and ER oxidative stress response processes [12]. Therefore, although *POS5* overexpression leads to a drain in cell energy resources (ATP consumption), it ensures non-limiting supply NADPH, which has been demonstrated to allow for increased recombinant protein production yields [13]. Our *in silico* calculations show increased Fab secretion when NADH kinase is overexpressed in the cytosol. This may indicate that Fab biosynthesis could compensate the cofactor perturbation (boosted NADPH levels) by draining such NADPH excess, thereby restoring the redox cofactor balance.

6.4.2. Metabolic impact of *POS5* overexpression

Our results strongly suggest that Pos5p NADH kinase overexpression perturbs cofactor balance. This is further reflected in the differences observed in growth profiles of *POS5* overexpressing strains compared to the reference strain, pointing to an impact in the distribution of fluxes through the cell's metabolic network. Previous studies in other yeast and fungi have reported that the overexpression of NADH kinases has a strong effect on metabolic flux distribution [36–38], as also predicted by the simulations performed with iMT1026 v3.0. Noteworthy, the impermeability of organelle membranes to NAD(P)H leads to cofactor reoxidation in the same compartment where they are reduced [39]. Thus, different effects of the cofactor perturbation are expected in cPOS5 and mPOS5 strains. *POS5* overexpression provides a source of NADPH in addition to the oxidative branch of the pentose phosphate pathway (PPP), which is the main cytosolic NADPH-generating pathway in yeast [40]. As extra NADPH is supplied by the NADH kinase, flux through the oxidative branch of the PPP is predicted to decrease, as observed when *POS5* is cytosollicaly overexpressed in *S. cerevisiae* [37]. Indeed, in *S. cerevisiae* NADPH inhibits *ZWF1*, the first step in oxidative pentose phosphate pathway [41]. Therefore, the increased levels of NADPH in cPOS5 (and 2cPOS5), would be coherent with the predicted reduction of flux through the oxidative branch of the PPP. Conversely, mitochondrial overexpression of *POS5* would result in an increased flux through the PPP. Despite the production of additional mitochondrial NADPH, the impermeability of mitochondrial membrane to redox cofactors would force

cytosolic NADPH generation pathways to supply the required NADPH for compartment-specific biosynthetic processes. The correlation between the flux through the oxidative branch of the pentose phosphate pathway and biomass yields has been widely reported [40,42]. Accordingly, the predicted flux increase and decrease through the PPP as a consequence of m*POS5* and c*POS5* overexpression, respectively, are in agreement with the biomass yields observed in chemostat cultivations: an increase of Y_{XS} yield in mPOS5 and a reduction in cPOS5 strains (**Table S6-1**). Similarly, in other species, while mitochondrial *POS5* overexpressing strains show an increase in Y_{XS} [5,36,38], the cytosolic overexpression of the NADH kinase results in reduced biomass yields. It is noteworthy to mention that the scaled reduced costs of Fab production over biomass generation are narrow (1.01×10^{-4}) and, therefore, differences in biomass yields are mainly consequence of *POS5* overexpression and the concomitant redistribution of metabolic fluxes, rather than increased Fab yields.

6.4.3. Energy metabolism and by-product secretion

Redox cofactor balance and energy metabolism are very closely linked in the respiratory chain. Thus, a perturbation in redox cofactor levels would lead a metabolic flux redistribution to restore the energy supply capacity of the cells. In addition to the redox cofactor imbalance created by a surplus of NADPH, NADH kinase is ATP and NADH consuming. According to the performed simulations, higher fluxes in oxidative phosphorylation would compensate this ATP drain. These predictions are also in agreement with the increased oxygen consumption rates of the mutant strains during the chemostat cultivations. Nevertheless, in the cPOS5 strain, part of the cytosolic NADH generated is consumed and cannot be neither used for biosynthetic purposes nor transported by mitochondrial redox shuttles to further deliver its electrons into the respiratory chain. Despite the higher TCA cycle activity and consequent increase in mitochondrial NADH generation, it would not provide enough reducing power for the electron transport chain and a reduction in the NADH electron transfer as well as an increase in alternative electron delivery mechanisms (i.e. succinate oxidation to fumarate) are predicted for c*POS5*. Since the NADH demand is located in mitochondria in m*POS5*, the compensation of the drain would rely on flux readjustments in mitochondrially-located reactions. In this case, an activation of the malic enzyme as well as an increase in the anaplerotic feed of TCA cycle intermediates would supply the additional

NADH enabling increased electron transference to the respiratory chain. As in *S. cerevisiae* and *A. nidulans*, we do not observe a reduction in biomass yields [36,37].

When growing in hypoxic conditions, the additional supply of ATP required by Pos5p is limited due to the restricted oxygen availability that constrains the respiratory chain. We predict that when c*POS5* cells grow under hypoxia, the ATP drain caused by Pos5p activity cannot be restored at all, leading to a decrease of cell fitness and increased hypoxic effects. Accordingly, experimental data shows an increase of by-product formation when c*POS5* is overexpressed. Conversely, despite the oxygen limitations, mPOS5 would be able to compensate the drained ATP by supplying additional NADH to the respiratory chain and increasing ATP production. This strain, similarly to *A. nidulans*, is able to overcome the ATP drain and increase biomass yields, even under hypoxic conditions [36]. Although *P. pastoris* is commonly classified as Crabtree negative yeast, it can produce certain amount of ethanol and other by-products [42], particularly under hypoxic conditions [18]. By-product formation is a consequence of limitations in carbon catabolisation capacity in TCA cycle and oxidative phosphorylation that leads to an excess of reduced NAD(P)H that the cell is not able to reoxidise by the respirative pathway [42,43]. mPOS5 showed a decrease in ethanol secretion due to the reduction in available mitochondrial NADH, while arabitol production remained comparable to the reference X-33/2F5 strain. The c*POS5* strains showed increased arabitol and ethanol production both in experimental data and simulations. In hypoxic conditions, increased NADH kinase levels would convert the excess of NADH to NADPH (as reflected in the experimentally determined increased NADPH/NADP$^+$ ratio); this NADPH surplus would be subsequently reoxidised through the generation of arabitol. In addition, simulations indicate an increase in TCA cycle flux leading to enhanced NADH generation. Due to the reduced capacity of oxidative phosphorylation caused by oxygen restriction, the additional NADH generated has to be reoxidised forming ethanol, coherent with the experimental observation.

6.5. CONCLUSIONS

In this study, the *S. cerevisiae POS5* NADH kinase was overexpressed in *P. pastoris,* either directed to the cytosol or to the mitochondria. The physiological characterisation of these

strains in chemostat cultivations showed a clear effect of *POS5* overexpression on redox cofactor balance. Indeed, *POS5* overexpression increased NADPH/NADP+ ratio in all the strains and conditions tested. Furthermore, the strain containing two copies of *POS5* integrated in the cell's genome (2cPOS5) showed the highest increase in NADPH/NADP+ ratios in comparison to the reference strain. This strongly supports a positive correlation between *POS5* gene dosage and NADPH availability. Moreover, this correlation is also appreciable when comparing the strain performance: 2cPOS5 showed the greatest fold change increase in Fab productivity. These results are also in agreement with the performed simulations, in which we predict a positive correlation between NADPH turnover and Fab production as well as increased Fab productivities when flux through the cytosolic NADH kinase reaction is increased.

The differentiated behaviour of m*POS5* and c*POS5* strains indicates the complexity of cell metabolism with organelle membranes impermeable to redox cofactors and highlights the importance of directing enzymes to the appropriate compartment when designing metabolic engineering strategies.

As a result of *POS5* overexpression, metabolic fluxes through the central carbon metabolism redistribute. Notably, redox engineered strains showed higher oxygen requirements concomitant with increased oxidative phosphorylation in order to replenish the ATP pools drained in the reaction catalysed by the NADH kinase. Consequently, these strains, particularly c*POS5*, are more sensitive to O_2 (i.e. show a lower threshold for the onset of respirofermentative metabolism) and showed more extreme/drastic hypoxic effects (increased by-product formation). In fact, the effects of NADH kinase overexpression are boosted under hypoxic conditions, and redox-engineered strains show higher NADPH/NADP+ ratios and Fab productivities. Both *in silico* flux distributions and macroscopic growth parameters agreed in increased demand of oxygen, CO_2 and by-product generation profiles, as well as Fab productivities, revealing iMT1026 v3.0 as a useful tool for assessing consistently the interpretation of the cultivation results.

Overall, we demonstrated the impact of redox cofactor perturbation in cell metabolism and provided further evidence of NADPH metabolism as key cell engineering target for improved recombinant protein production.

6.6. References

1   Jordà, J. *et al.* (2012) Metabolic flux profiling of recombinant protein secreting *Pichia pastoris* growing on glucose:methanol mixtures. *Microb. Cell Fact.* 11, 57

2   Glick, B.R. (1995) Metabolic load and heterologous gene expression. *Biotechnol. Adv.* 13, 247–261

3   Wu, G. *et al.* (2016) Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol.* 34, 652–664

4   Delic, M. *et al.* (2012) Oxidative protein folding and unfolded protein response elicit differing redox regulation in endoplasmic reticulum and cytosol of yeast. *Free Radic. Biol. Med.* 52, 2000–12

5   Hou, J. *et al.* (2009) Metabolic impact of redox cofactor perturbations in *Saccharomyces cerevisiae*. *Metab. Eng.* 11, 253–61

6   Lee, H.C. *et al.* (2010) High NADPH/NADP+ ratio improves thymidine production by a metabolically engineered *Escherichia coli* strain. *J. Biotechnol.* 149, 24–32

7   Siedler, S. *et al.* (2011) Increased NADPH availability in *Escherichia coli*: Improvement of the product per glucose ratio in reductive whole-cell biotransformation. *Appl. Microbiol. Biotechnol.* 92, 929–937

8   Kim, S. and Hahn, J.S. (2015) Efficient production of 2,3-butanediol in Saccharomyces cerevisiae by eliminating ethanol and glycerol production and redox rebalancing. *Metab. Eng.* 31, 94–101

9   Geertman, J.M.A. *et al.* (2006) Physiological and genetic engineering of cytosolic redox metabolism in *Saccharomyces cerevisiae* for improved glycerol production. *Metab. Eng.* 8, 532–542

10 Krainer, F.W. *et al.* (2012) Recombinant protein expression in *Pichia pastoris* strains with an engineered methanol utilization pathway. *Microb. Cell Fact.* 11, 22

11 Heyland, J. *et al.* (2010) Quantitative physiology of *Pichia pastoris* during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. *Biotechnol. Bioeng.* 107, 357–68

12 Driouch, H. *et al.* (2012) Integration of in vivo and in silico metabolic fluxes for improvement of recombinant protein production. *Metab. Eng.* 14, 47–58

13 Nocon, J. *et al.* (2016) Increasing pentose phosphate pathway flux enhances recombinant protein production in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* 100, 5955–5963

14 Nocon, J. *et al.* (2014) Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. *Metab. Eng.* 24, 129–138

15 Mattanovich, D. *et al.* (2016) Industrial Microorganisms: *Pichia pastoris*. In *Industrial Biotechnology* pp. 687–714, Wiley-VCH Verlag GmbH & Co. KGaA

16 Baumann, K. *et al.* (2008) Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins. *Biotechnol. Bioeng.* 100, 177–183

17 Carnicer, M. (2012) , Systematic metabolic analysis of recombinant *Pichia pastoris* under different oxygen conditions. , Universitat Autònoma de Barcelona

18 Baumann, K. *et al.* (2010) A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions. *BMC Syst. Biol.* 4, 141

19 Gasser, B. *et al.* (2006) Engineering of *Pichia pastoris* for improved production of antibody fragments. *Biotechnol. Bioeng.* 94, 353–361

20 Outten, C.E. and Culotta, V.C. (2003) A novel NADH kinase is the mitochondrial source of NADPH in *Saccharomyces cerevisiae*. *EMBO J.* 22, 2015–24

21 Strand, M.K. *et al.* (2003) POS5 Gene of *Saccharomyces cerevisiae* Encodes a Mitochondrial NADH Kinase Required for Stability of Mitochondrial DNA. *Eukaryot. Cell* 2, 809–820

22 Baumann, K. *et al.* (2011) Protein trafficking, ergosterol biosynthesis and membrane physics impact recombinant protein secretion in *Pichia pastoris*. *Microb. Cell Fact.* 10, 93

23 Murray Lab, T. 01-Dec-(2012) , Colony PCR. . [Online]. Available: http://labs.mcb.harvard.edu/murray/colony_pcr.html

24 Cámara, E. *et al.* (2016) Droplet digital PCR-aided screening and characterization of Pichia pastoris multiple gene copy strains. *Biotechnol. Bioeng.* 113, 1542–1551

25 Ortmayr, K. *et al.* (2014) Sample preparation workflow for the liquid chromatography tandem mass spectrometry based analysis of nicotinamide adenine dinucleotide phosphate cofactors in yeast. *J. Sep. Sci.* 37, 2185–2191

26 Carnicer, M. *et al.* (2012) Development of quantitative metabolomics for *Pichia pastoris*. *Metabolomics* 8, 284–298

27 Noorman, H.J. *et al.* (2000) Classification, error detection, and reconciliation of process information in complex biochemical systems. *Biotechnol. Bioeng.* 49, 364–376

28 Tomàs-Gamisans, M. *et al.* (2016) 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* 11, e0148031

29 Keating, S.M. *et al.* (2006) SBMLToolbox: An SBML toolbox for MATLAB users. *Bioinformatics* 22, 1275–1277

30 Bornstein, B.J. *et al.* (2008) LibSBML: An API library for SBML. *Bioinformatics* 24, 880–881

31 Choi, H.S. *et al.* (2010) In Silico Identification of Gene Amplification Targets for Improvement of *Lycopene Production*. *Appl. Environ. Microbiol.* 76, 3097–3105

32 Chung, B. and Lee, D.-Y. (2009) Flux-sum analysis: a metabolite-centric approach for understanding the metabolic network. *BMC Syst. Biol.* 3, 117

33 Segrè, D. *et al.* (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15112–15117

34 Carnicer, M. *et al.* (2009) Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels. *Microb. Cell Fact.* 8, 65

35 Dragosits, M. *et al.* (2009) The effect of temperature on the proteome of recombinant *Pichia pastoris*. *J. Proteome Res.* 8, 1380–1392

36 Panagiotou, G. *et al.* (2009) Overexpression of a novel endogenous NADH kinase in *Aspergillus nidulans* enhances growth. *Metab. Eng.* 11, 31–39

37 Hou, J. *et al.* (2009) Impact of overexpressing NADH kinase on glucose and xylose metabolism in recombinant xylose-utilizing Saccharomyces cerevisiae. *Appl. Microbiol. Biotechnol.* 82, 909–19

38 Qiao, K. *et al.* (2017) Lipid production in Yarrowia lipolytica is maximized by engineering cytosolic redox metabolism. *Nat. Biotechnol.* 35,

39 Bakker, B.M. *et al.* (2001) Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* 25, 15–37

40 Blank, L.M. *et al.* (2005) Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *FEMS Yeast Res.* 5, 545–58

41 Llobell, A. *et al.* (1988) Glutathione reductase directly mediates the stimulation of yeast glucose-6-phosphate dehydrogenase by GSSG. *Biochem. J.* 249, 293–296

42 Heyland, J. *et al.* (2011) Carbon metabolism limits recombinant protein production in *Pichia pastoris*. *Biotechnol. Bioeng.* 108, 1942–53

43 Vemuri, G.N. *et al.* (2007) Increasing NADH oxidation reduces overflow metabolism in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 2402–7

APPENDIX IV

Table S6-1. List of primers used in this study

Primer name	Sequence (5'-3')	T_m (°C)
POS5-F [a,b]	ATCCGCGCCTGCAGGAATGTTTGTTAGAGTTAAGTTGAACAAGCCAGTTAA	67
POS5_cyt-F [a,b]	ATCCGCGCCTGCAGGAATAATGTCCACTTTGGACTCCCATTCCTTGAA	68
POS5-R [a,b]	ATGACTAGGCCGAGGCGGCCTTAGTCGTTGTCAGTCTGTCTC	68
POS5_int-R [a,b]	AGCAACACCGTCAGCAGTAG	
POS5_amp-F [c]	GGAGTGTCACTTGAAGAA	38.6
POS5_amp-R [c]	CGTCAGCAGTAGTTCTAG	36.6
POS5 probe [c]	ACTCCAACTCCTCCATCGTTACTCA (5': 6-Fam / 3': BHQ-1)	57.1

[a] Primer used for cloning *POS5* into pPUZZLE.
[b] Primer used for clone verification
c Primer used for gene copy number determination
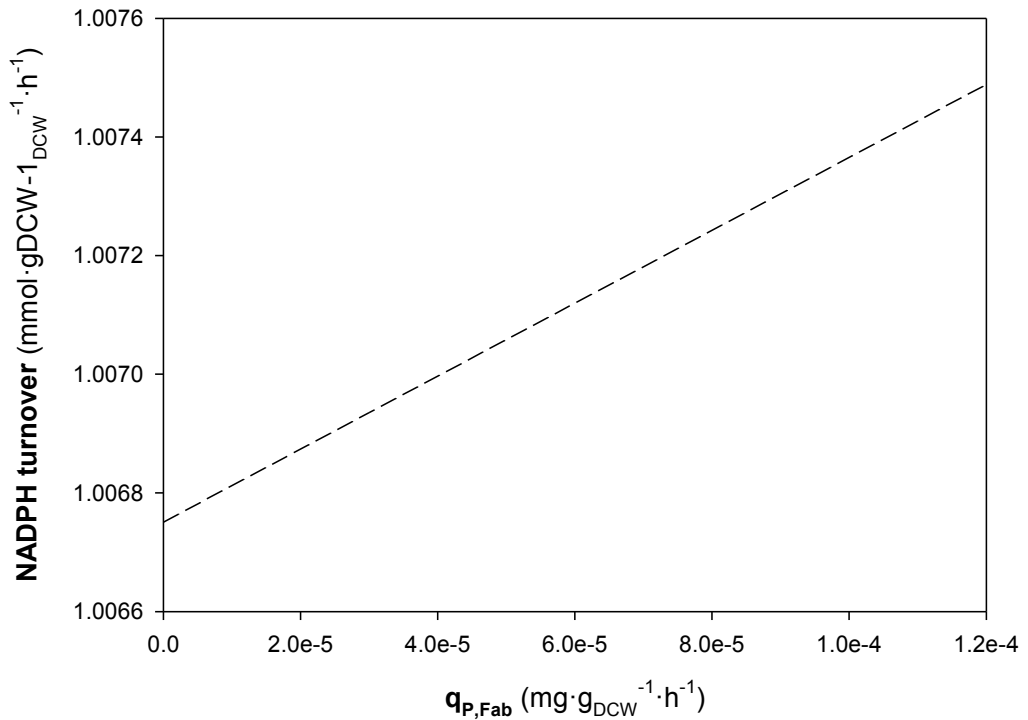
Fig. S6-1. Representation of NADPH pool turnover in relation to Fab productivity. A series of FSEOF simulations were performed by increasing the Fab productivity. A sum-flux analysis of NADPH was performed for each simulation and resulting NADPH turnovers are plotted into the graph.

Table S6-2. Macroscopic growth parameters of cultivations performed in glycerol and normoxic conditions and glucose in normoxia and hypoxia after the reconciliation procedure. Values are average ± standard deviation of two independent replicates.

	Glycerol - normoxia				Glucose - normoxia				Glucose - hypoxia			
	X-33/2F5	cPOS5	2cPOS5	mPOS5	X-33/2F5	cPOS5	2cPOS5	mPOS5	X-33/2F5	cPOS5	2cPOS5	mPOS5
Glucose	-	-	-	-	-0.80 ± 0.02	-0.93 ± 0.06	-0.93 ± 0.04	-0.93 ± 0.18	-1.20 ± 0.00	-1.26 ± 0.13	-1.37 ± 0.01	-1.03 ± 0.12
Glicerol	-1.49 ± 0.06	-1.52 ± 0.16	-1.53 ± 0.01	-1.51 ± 0.23	0.01 ± 0.01	-	-	-	-	0.01 ± 0.01	-	-
Arabitol	-	-	-	-	-	-	-	-	0.04 ± 0.01	0.07 ± 0.01	0.06 ± 0.01	0.04 ± 0.02
Ethanol	-	-	-	-	-	-	-	-	0.53 ± 0.00	0.46 ± 0.11	0.70 ± 0.14	0.24 ± 0.17
Succinic	-	-	-	-	-	-	-	-	0.01 ± 0.00	0.01 ± 0.01	0.01 ± 0.00	-
Biomass	3.36 ± 0.01	3.42 ± 0.20	3.41 ± 0.08	3.28 ± 0.34	3.15 ± 0.47	3.75 ± 0.03	3.80 ± 0.04	3.64 ± 0.37	3.54 ± 0.04	3.67 ± 0.15	3.76 ± 0.04	3.57 ± 0.21
CO2	1.10 ± 0.18	1.15 ± 0.26	1.19 ± 0.13	1.23 ± 0.36	1.60 ± 0.27	1.75 ± 0.24	1.81 ± 0.21	1.68 ± 0.35	2.00 ± 0.33	2.23 ± 0.23	2.36 ± 0.72	1.77 ± 0.24
O2	1.85 ± 0.21	1.91 ± 0.35	1.95 ± 0.15	1.98 ± 0.47	1.52 ± 0.28	1.67 ± 0.23	1.72 ± 0.21	1.59 ± 0.33	1.36 ± 0.26	1.63 ± 0.15	1.51 ± 0.33	1.37 ± 0.26
RQ [a]	0.60 ± 0.03	0.60 ± 0.03	0.61 ± 0.02	0.62 ± 0.03	1.05 ± 0.01	1.05 ± 0.01	1.05 ± 0.01	1.06 ± 0.01	1.48 ± 0.05	1.37 ± 0.02	1.55 ± 0.13	1.30 ± 0.07
Yxs [b]	0.70 ± 0.03	0.70 ± 0.03	0.69 ± 0.02	0.68 ± 0.04	0.59 ± 0.07	0.62 ± 0.02	0.60 ± 0.04	0.62 ± 0.02	0.44 ± 0.00	0.44 ± 0.07	0.41 ± 0.00	0.53 ± 0.09
Fab [c]	0.02 ± 0.01	0.04 ± 0.02	0.04 ± 0.01	0.04 ± 0.03	0.02 ± 0.01	0.04 ± 0.03	0.05 ± 0.02	0.05 ± 0.03	0.07 ± 0.03	0.07 ± 0.04	0.12 ± 0.06	0.05 ± 0.02

[a] mmol $CO_2 \cdot g_{DCW}^{-1} \cdot h^{-1}$ / mmol $O_2 \cdot g_{DCW}^{-1} \cdot h^{-1}$
[b] g biomass / g substrate
[c] mg $Fab \cdot g_{DCW}^{-1} \cdot h^{-1}$

7

General conclusions

7. General conclusions

The high interest in *P. pastoris* as a platform for recombinant protein and metabolite production has boosted the body of knowledge of its physiology over the past 10 years. Previously to this study, systematic work at different levels has been performed for a better understanding of this cell factory. With the rise of systems biology genome-scale metabolic models have become indispensable tools for rational strain engineering and contextualising biological data. Indeed, three GSMMs of *P. pastoris* were available at the onset of this study. Nevertheless, such models showed some weaknesses and discrepancies in certain metabolic pathways. Consequently, the need for developing a reliable model was of utmost relevance (or essential) for further development of systems metabolic engineering strategies for *P. pastoris*.

Therefore, a consensus GSMM has been developed, integrating and upgrading the previously available models. Following the validation of the model in a broader range of conditions than in prior models, we have demonstrated that the new model, iMT1026, shows improved capabilities and accuracy in predictions of physiological parameters compared to previous models.

Aiming at extending the applicability range of iMT1026 in biotechnologically relevant conditions, novel physiological datasets of cells growing on methanol and glycerol as sole carbon sources have been generated in a systematic and comprehensive manner. Notably, a detailed characterisation of macromolecular composition of biomass grown on such carbon sources has been performed. Several differences have been identified when comparing biomass compositions of cells grown on glycerol or methanol. One of the major spotted differences has been the higher protein content under methanolic growth conditions, probably due to the high expression levels of the methanol metabolic machinery. In addition, the lipid profile was strongly affected by the carbon source, in agreement with previous studies. Overall, these analyses have enabled the formulation of new carbon

source-specific biomass equations. Furthermore, energetic parameters have been estimated for first time and using different methods, for *P. pastoris* growth on glycerol and methanol, and compared to those obtained for glucose-grown cells. As a result, the iMT1026 v3.0 upgrading can describe cells grown in glycerol and methanol more accurately than the previous version.

Due to the potentiality of glycerol as carbon source for industrial biotechnological processes, a ^{13}C-based metabolic flux analysis (13C-MFA) of *P. pastoris* growing on this carbon source has been performed for first time. Despite central carbon metabolism models have been widely used for ^{13}C-MFA, they are typically directly derived from the generic knowledge of biochemical pathways of model organisms (e.g. *S. cerevisiae*) and, therefore, these models could misrepresent the core metabolism of our organism of interest. In this study, iMT1026 v3.0 has been systematically reduced to a 77-reaction core model and subsequently used for ^{13}C-MFA. Notably, the estimated metabolic flux distributions in glycerol-grown cells are in agreement with pioneering 13C-isotopic labelling experiments of our group in which alternative analytical techniques were employed for METAFoR analysis. Notably, flux distributions have been validated in terms of redox cofactor and electron balances, showing a great accuracy in the calculated oxygen uptake rate when compared to the experimentally measured values. Energetic parameter estimation using the calculated fluxes has also provided similar values to those estimated using the complete iMT1026 v3.0.

Moreover, the setup of ^{13}C-MFA for growth on glycerol has allowed performing a systematic study of the impact of the specific growth rate on the central carbon metabolism of cells growing on this carbon source. As a result, some significant differences have been appreciated in metabolic flux profiles. Specifically, an increase in the flux through the pentose phosphate pathway has been observed at higher growth rates. Since precursors for nucleotide biosynthesis are generated in the PPP, these results are in agreement with the growth rate hypothesis, which postulates a correlation between RNA content and growth rate Indeed, biomass compositional analyses also showed this correlation, providing further evidences for the growth rate hypothesis.

Finally, we have further proved the capabilities of this genome-scale metabolic model in another application, i.e for assisting in the interpretation of biological data. We have been

able to introduce a redox cofactor perturbation by overexpressing a recombinant NADH kinase to a Fab producing strain. As a result, increased Fab productivities have been obtained, especially when process engineering strategies have been also combined (i.e. cultivation under hypoxic conditions). These results strongly support NADPH regeneration rates as a key factor for metabolic engineering of *P. pastoris* for improved recombinant protein production. Furthermore, experimental datasets and their subsequent *in silico* biological interpretation have allowed us to postulate a metabolic interpretation for the observed macroscopic physiological changes, thereby revealing iMT1026 v3.0 as a useful tool for assessing consistently the interpretation of both genetic and environmental perturbations/modifications of our biological system.

Overall, the present study provides a consensus genome-scale metabolic model for *P. pastoris* with extended capabilities and greater accuracy than previous models, thus providing an improved tool for systems metabolic engineering of *P. pastoris*. Furthermore, its use in two different applications demonstrates its reliability, consistency and potential capabilities.

Scientific contributions

Publications

Tomàs-Gamisans, M., Ferrer, P., & Albiol, J. (2016). **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, 11(1), e0148031.

Tomàs-Gamisans, M., Ferrer, P., & Albiol, J. **Benchmarking P. pastoris genome-scale metabolic model iMT1026 for growth on methanol or glycerol as sole carbon sources.** Microbial Biotechnology (Under Review)

Chapters 5 and **6** are also expected to be submitted as articles

INDEX OF TABLES

INDEX OF FIGURES