

UNIVERSITAT AUTÒNOMA DE BARCELONA



Systematic metabolic analysis of recombinant  
*Pichia pastoris* under different oxygen conditions

*A Metabolome and Fluxome Based Study*

Memòria per obtenir el Grau de Doctor  
per la Universitat Autònoma de  
Barcelona sota la direcció de  
Pau Ferrer Alegre i Joan Albiol Sala

**Marc Carnicer Heras**

Departament d'Enginyeria Química, 2012



**Pau Ferrer Alegre** y **Joan Albiol Sala**, Professors Associats en el grup d'Enginyeria de Bioprocessos i Biocatàlisi Aplicada del Departament d'Enginyeria Química de la Universitat Autònoma de Barcelona

CERTIFIQUEN: Que el bioquímic Marc Carnicer Heras ha dut a terme sota la nostra direcció, el treball que, amb el títol "**Systematic metabolic analysis of recombinant *Pichia pastoris* under different oxygen conditions**" es presenta en aquesta memòria, la qual consisteix la seva Tesi per optar al grau de Doctor en Biotecnologia per la Universitat Autònoma de Barcelona.

I per tal que se'n prengui coneixement i consti als efectes oportuns, signem la present a Bellaterra, Abril 2012.

Dr. Pau Ferrer Alegre

Dr. Joan Albiol Sala



To My Family



*I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale.*

**Marie Curie**



## Summary

The systematic analysis of the cell physiological it is critical to gain knowledge about microorganisms. Systems biology gives the opportunity to obtain quantitatively analysis of different physiological levels which allow the *in silico* representations of the studied pathways. Overall, this field is addressed to the crucial understanding of complex biological networks behaviour.

This research project was aimed to contribute to the understandings of the *Pichia pastoris* physiology under different stress conditions. Specifically, this thesis was centred in the systematic analysis of the metabolome and fluxome omic levels, and how the different levels contribute to regulate the final metabolic flow regarding the oxygen availability and the recombinant protein production.

In **Chapter 2**, a consistent description of the biomass principal components of *P. pastoris* is obtained for different oxygenation conditions, as well as under recombinant protein producing and non-producing backgrounds obtaining the best estimation of the biomass composition. Specifically, an increased in the reduction degree was determined at lower oxygen availability. In addition, consistency of the input and output extracellular metabolic fluxes was used to identify the production of an unexpected by-product, which was subsequently identified as arabinitol. The obtained results were afterwards used in chapter 3 to perform metabolic flux analysis.

In **Chapter 3**, a biosynthetically directed fractional  $^{13}\text{C}$ -labellings were performed obtaining information about the principals metabolic flux ratios of the carbon central metabolism of *P. pastoris*. Afterwards, these values were used to perform  $^{13}\text{C}$ -constrained metabolic flux analysis under the different oxygenation conditions and producing the recombinant protein. The most prominent feature obtained, already indicated by the metabolic flux ratios, was the similarity in flux estimations between the expressing and the control strains. Nevertheless, clear differences were observed when comparing flux patterns corresponding to different oxygenations set points.

The issue of metabolome analysis was started in **Chapter 4** with the systematic evaluation of five quenching protocols applied to *P. pastoris*. Methanol concentration and temperature were the selected variables to be optimized for minimizing metabolite leakage during quenching and subsequent washing. As result, acceptable recoveries (>90%) were obtained for all quenching procedures tested. However, quenching at  $-27^{\circ}\text{C}$  in 60% v/v methanol performed slightly better in terms of leakage minimization. Thereafter, the best performing quenching protocol was used to demonstrate that five residence times under glucose limitation were enough to reach stable intracellular metabolite pools. Moreover, when comparing *P. pastoris* and *S. cerevisiae* metabolomes, under the same cultivation conditions, similar metabolite fingerprints were found in both yeasts, except for the lower glycolysis.

In Chapters 5, 6 and 7, the *P. pastoris* metabolome adaptation to recombinant protein production and to different oxygen availability is described together with a global metabolome analysis. Specifically, in

**Chapter 5**, the redox and energy stoichiometry of *P. pastoris* was validated showing increasing energy requirements as the oxygen availability became shortage. In addition, a network-embedded thermodynamic analysis of the quantitative intracellular metabolites pools was performed. From the analysis, an improvement of the non-oxidative pentose phosphate pathway fluxes was performed as well as a determination of the oxygen impact over the redox state of the cytosol.

In **Chapter 6**, the transcriptional and thermodynamic regulations over the metabolic flow between conditions were investigated for the central carbon metabolism reactions. Overall, the different omic contributions to the final metabolic flux changes were resulted to be different between *P. pastoris* and *S. cerevisiae*. Specifically, for *P. pastoris*, transcriptional regulation was found to be in most cases significant, contrarily to the baker yeast. Moreover, a possible kinetic contribution to the metabolic flux changes could be seen in the central carbon metabolism.

Finally, in **Chapter 7**, to better understand the effect and interplay of oxygen availability and foreign protein secretion on central metabolism, a first quantitative metabolomic analysis of free amino acids pools in a recombinant *P. pastoris* strain growing under different oxygen availability conditions was carried out. The obtained values indicated significant variations in the intracellular amino acid levels due to different oxygen availability conditions. Notably, even that foreign protein productivities were relatively low, recombinant protein production was found to have a limited but significant impact on the intracellular amino acid pools, which were generally decreased in the producing strain compared with the reference strain under normoxia conditions. However, observed changes in individual amino acids pools were not correlated with their corresponding relative abundance in the recombinant protein sequence, but to the overall cell protein amino acid compositional variations.

## Ressenya

L'anàlisi sistemàtic de la fisiologia cel·lular és de gran importància en la millora del coneixement sobre els microorganismes. En aquesta direcció, la biologia de sistemes ens permet l'anàlisi quantitatiu de diferents nivells fisiològics obtenint representacions *in silico* de les vies metabòliques estudiades que, en conjunt, s'adrecen a la millor comprensió del comportament de sistemes biològics complexos.

Aquest projecte de recerca va ser dissenyat per contribuir en la comprensió de la fisiologia de *Pichia pastoris* sota diferents condicions d'estrès. Específicament, aquesta tesi se centre en l'estudi sistemàtic del conjunt de metabòlits lliures dins la cèl·lula i en la velocitat que aquests es transformen. Conjuntament, també s'ha estudiat la contribució d'aquests en la regulació final dels fluxos metabòlics respecte la disponibilitat d'oxigen i la producció d'una proteïna recombinant.

En el **Capítol 2**, una descripció consistent dels principals component de la biomassa de *P. pastoris* es realitzada sota diferents condicions d'oxigenació al igual que produint i no produint la proteïna recombinant obtenint així una millor estimació de la composició de la biomassa. Específicament, un increment en el grau de reducció de la biomassa va ésser determinat en condicions de dèficit d'oxigen. Per una altra part, l'estudi de la consistència dels fluxos dels metabòlits extracel·lulars va ser utilitzat per identificar la producció d'un subproducte de fermentació inesperat posteriorment identificat com arabinitol. Aquests resultats van ser utilitzats per l'anàlisi de fluxos metabòlics del capítol 3.

En el **Capítol 3**, s'obtingué la relació entre diferents fluxos metabòlics del metabolisme central del carboni de *P. pastoris* mitjançant l'estudi del carboni marcat dels aminoàcids sintetitzats de *novo*. Posteriorment, aquest valors van ser utilitzats delimitant l'anàlisi de fluxos metabòlics sota diferents condicions d'oxigenació i producció de proteïna recombinant. Els resultat més destacat va ser la gran similitud entre els fluxos estimats per les soques productora i control en condicions de cultiu equivalents. Tot i així, diferències clares van ser observades quan es van comparar el patró de fluxos corresponents a diferents nivells d'oxigenació.

L'anàlisi dels metabòlits lliures dins la cèl·lula va començar en el **Capítol 4** amb l'avaluació sistemàtica de cinc protocols de parada ràpida del metabolisme en *P. pastoris*. El metanol i la temperatura van ser les variables seleccionades a optimitzar en la minimització de la pèrdua de metabòlits en el procés de parada i rentat de les cèl·lules. Com a resultat, recuperacions acceptables (>90%) van ser obtingudes en tots els protocols provats. Amb tot, la parada del metabolisme realitzada a -27°C amb una solució de 60% v/v de metanol a ser amb la que menys pèrdues es van observar. Consecutivament, el millor protocol de parada va ser utilitzat per demostrar que cinc temps de residència són els necessaris per arribar a un estat estacionari del metabòlits intracel·lulars en cultius amb glucosa limitant. D'altra banda, sota les mateixes condicions de treball, els perfils intracel·lulars dels principals metabòlits

centrals eren similars entre els llevats *P. pastoris* i *S. cerevisiae*, en excepció de la part baixa de la glicòlisi.

En els capítols 5, 6 i 7, l'adaptació del metabolome de *P. pastoris* a la producció de la proteïna recombinant i a les diferents condicions d'oxigen és descrita conjuntament amb un anàlisi global d'aquest. Específicament, en el **Capítol 5**, l'estequiometria redox i energètica de *P. pastoris* varen ser validades veient un augment en els requeriments energètics sota condicions deficitàries d'oxigen. A més a més, es va realitzar un anàlisi termodinàmic dels metabòlits intracel·lulars mitjançant el qual s'obtingueren millores en el patró de fluxos en la via de les pentoses fosfat conjuntament amb informació de l'estat redox del citosol.

En el **Capítol 6**, la regulació transcripcional i termodinàmica dels fluxes del metabolisme central del carboni va ser investigada entre diferents condicions de cultiu. En general, la contribució dels diferents nivells omics en el canvi final dels fluxos metabòlics van resultar ser diferents entre els llevats *P. pastoris* i *S. cerevisiae*. Específicament, per *P. pastoris*, la regulació transcripcional va ser significant en la majoria de reaccions, contràriament al que s'havia descrit per *S. cerevisiae*. D'altre banda, una possible contribució cinètica al canvi de fluxos metabòlics va ser observada en el metabolisme central del carboni en el llevat *P. pastoris*.

Finalment, en el **Capítol 7**, per un millor coneixement i comprensió de l'efecte i la interacció de la disponibilitat d'oxigen i la producció d'una proteïna forania en el metabolisme central, es va realitzar un anàlisi quantitatiu dels aminoàcids lliures intracel·lulars sota la producció d'aquesta proteïna i en diferents condicions d'oxigenació. Els valors obtinguts indicaren una variació significativa d'aquests aminoàcids intracel·lulars entre diferents disponibilitats d'oxigen. Sorprenentment, tot i que la producció de la proteïna recombinant fos relativament baixa, l'expressió d'aquesta es va veure que tenia un impacte limitat, però significant, en els pools d'aminoàcids reduint-ne la mida en condicions totalment oxidatives. Amb tot, els canvis individuals observats en els aminoàcids no presentaven una correlació amb l'abundància relativa d'aquests en la proteïna recombinant, en canvi, concordava amb la variació de la composició aminoacídica de la biomassa.

## Table of contents

SUMMARY	9
RESSENYA	11
TABLE OF CONTENTS	13
LIST OF ABBREVIATIONS	15
<b><u>CHAPTER 1: GENERAL INTRODUCTION</u></b>	<b><u>17</u></b>
<b><u>CHAPTER 2: <i>PICHIA PASTORIS</i> BIOMASS COMPOSITION</u></b>	<b><u>31</u></b>
<b><u>CHAPTER 3: <i>PICHIA PASTORIS</i> <sup>13</sup>C-BASED METABOLIC FLUX ANALYSIS</u></b>	<b><u>53</u></b>
<b><u>CHAPTER 4: <i>PICHIA PASTORIS</i> INTRACELLULAR METABOLITE QUANTIFICATION METHOD</u></b>	<b><u>71</u></b>
<b><u>CHAPTER 5: <i>PICHIA PASTORIS</i> STOICHIOMETRIC NETWORK IMPROVEMENT</u></b>	<b><u>93</u></b>
<b><u>CHAPTER 6: <i>PICHIA PASTORIS</i> CENTRAL CARBON METABOLITES ADAPTABILITY</u></b>	<b><u>123</u></b>
<b><u>CHAPTER 7: <i>PICHIA PASTORIS</i> FREE AMINO ACIDS ADAPTABILITY</u></b>	<b><u>147</u></b>
<b><u>CHAPTER 8: OUTLOOK AND FUTURE DIRECTIONS</u></b>	<b><u>161</u></b>
BIBLIOGRAPHY	166
LIST OF SCIENTIFIC CONTRIBUTIONS	178



## List of abbreviations

<b>PPP</b>	Pentose phosphate pathway	<b>αKG</b>	α-Ketoglutarate
<b>TCA</b>	Tricarboxylic acid	<b>OAA<sub>cyt</sub></b>	Cytosolic oxalacetate
<b>gDCW</b>	Grams of dry cell weight	<b>OAA<sub>mit</sub></b>	Mitochondrial oxalacetate
<b>AOX</b>	Alcohol oxidase	<b>G3P</b>	Glycerol-3-phosphate
<b>Fab</b>	Antibody fragment	<b>S7P</b>	Sedoheptulose-7-phosphate
<b>NMR</b>	Nuclear magnetic resonance	<b>Xu5P</b>	Xylulose-5-phosphate
<b>MFA</b>	Metabolic flux analysis	<b>Ribu5P</b>	Ribulose-5-phosphate
<b>HPLC</b>	High performance liquid chromatography	<b>R5P</b>	Ribose-5-phosphate
<b>LC-MS</b>	Liquid chromatography coupled with mass spectrometry	<b>E4P</b>	Eritrose-4-phosphate
<b>GC-MS</b>	Gas chromatography coupled with mass spectrometry	<b>UDP-glc</b>	Uridine-diphosphate glucose
<b>MAR</b>	Mass action ratio	<b>ATP</b>	Adenosine triphosphate
<b>G6P</b>	Glucose-6-phosphate	<b>ADP</b>	Adenosine diphosphate
<b>T6P</b>	Trehalose-6-phosphate	<b>AMP</b>	Adenosine monophosphate
<b>FBP</b>	Fructose-1,6-Bisphosphate	<b>NAD<sup>+</sup></b>	Nicotinamide adenine dinucleotide. Oxidized form
<b>6PG</b>	6-Phospho gluconate	<b>NADP<sup>+</sup></b>	Nicotinamide adenine dinucleotide phosphate. Oxidized form
<b>F6P</b>	Fructose-6-phosphate	<b>NADH</b>	Nicotinamide adenine dinucleotide. Reduced form
<b>G1P</b>	Glucose-1-phosphate	<b>NADPH</b>	Nicotinamide adenine dinucleotide phosphate. Reduced form
<b>M6P</b>	Mannose-6-phosphate	<b>PDH</b>	Pyruvate dehydrogenase
<b>GAP</b>	Glyceraldehyde-3-phosphate	<b>PGI</b>	Phosphoglucose isomerase
<b>2PG</b>	2-Phospho glycerate	<b>PGM</b>	Phosphoglucomutase
<b>3PG</b>	3-Phospho glycerate	<b>PMM</b>	Phosphomannomutase
<b>3PG<sub>cyt</sub></b>	Cytosolic 3-Phospho glycerate	<b>ENO</b>	Enolase
<b>PEP</b>	Phosphoenolpyruvate	<b>FMH</b>	Fumarase
<b>PEP<sub>cyt</sub></b>	Cytosolic Phosphoenolpyruvate	<b>Glx</b>	Glu + Gln
<b>PYR</b>	Pyruvate	<b>Asx</b>	Asp + Asn
<b>PYR<sub>cit</sub></b>	Cytosolic pyruvate	<b>RQ</b>	Respiratory coefficient
<b>PYR<sub>mit</sub></b>	Mitochondrial pyruvate	<b>P/O</b>	Efficiency of ATP generation in the oxidative phosphorylation
<b>AcCoA</b>	Acetyl Coenzyme-A	<b>C:N</b>	Carbon:Nitrogen ratio
<b>AcCoA<sub>mit</sub></b>	Mitochondrial acetyl Coenzyme-A	<b>H:O</b>	Hydrogen:Oxygen ratio
<b>AcCoA<sub>cyt</sub></b>	Cytosolic acetyl Coenzyme-A	<b>γ</b>	Degree of reduction of the biomass.
<b>Ace</b>	Acetate		



# **Chapter 1: General Introduction**



The biotechnology role is increasing in nearly every industry, including healthcare, pharmaceutical, chemical, food or agricultural industries. Biotechnological production of small-volume high-value drugs is well justified. However, production of large-volume low-value bioproducts requires the development of lower-cost and higher-yield processes [1]. Towards this goal, the heterologous protein production has succeeded in many cases for the production of pharmaceutical proteins or industrial enzymes [2, 3]. However, although high yields of active proteins had been obtained in many cases [4] there are still limitations for the production of the majority of complex proteins in large quantities and as functional entities.

The strain improvement has been carried out traditionally through random mutagenesis followed by screening processes in order to diminish or eliminate the heterologous production bottlenecks. This method has had significant success producing quite different metabolites in high-yield processes. However, rational cellular and metabolic engineering have obtained further improvements on the biotechnological production of different recombinant protein and cellular metabolites [1, 5].

Metabolic engineering is the field which has the ultimate aim to manipulate the cellular metabolism through various techniques to produce a desired end product, or alternatively, remove or breakdown an undesirable one. To accomplish this aim, understanding the metabolism of the cell is of critical importance [6].

Nutrients are transported through the cell membrane using different mechanism in order to be available to be catabolised. Once there are inside, these nutrients are the precursors to generate all the energy and reducing power to synthesize all the cell components required. The metabolic pathways ensure the proper and efficient energy and reducing power supply to obtain the desired molecules. Therefore, a solid knowledge of the cellular metabolism may allow the efficiency enhance of a particular metabolic conversion [6]. In some cases, one may wish to produce a compound in an organism which it is not present naturally like the recombinant protein or the polyhydroxyalkanoates (PHAs) among others. These compounds are valuable in a variety of industries as agriculture, pharmaceutical, food additive or bioremediation field. In these cases, a well know cell metabolism together with the characterization of the heterologous production impact may lead to the identification of the main production bottlenecks. Therefore, knowing the cell production difficulties a metabolic engineering strategy could be performed enhancing the improvement possibilities. As example, the

PHAs production in *E. coli* were improved increased over-expressing two enzymes involved in the cell NADPH availability [7].

## Systems Biology

In the metabolic engineering field, the systematic analysis of the cell physiological it is critical to gain knowledge about microorganisms. Systems biology gives the opportunity to obtain quantitatively analysis of different physiological levels which allow the *in silico* representations of the studied pathways [8]. Overall, this field is addressed to the crucial understanding of complex biological networks behaviour.

With this aim, proteomic, transcriptomic or fluxomic studies has become a tool towards strain optimization [9–12]. Besides, the metabolome study have obtained substantial improvements in genotype-phenotype relation studies on strains and mutant cells [11]. Nevertheless, not only the study of one “omic” level will lead to the complete understanding of the cellular metabolism but also the different level combination will lead to the closed picture of the microorganism *in vivo* regulation (Figure 1.1).

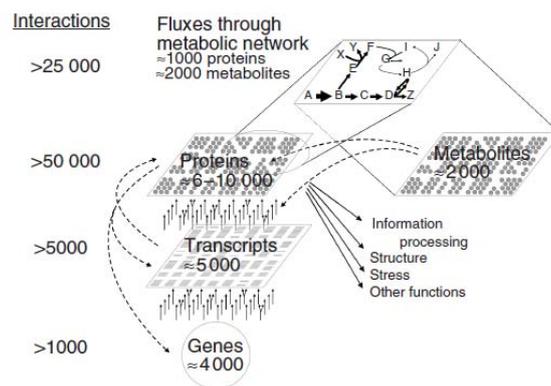


Figure 1.1: **Schematic overview of the relationship between the different “omics” levels in a complex biological network.** Picture taken from [13].

In systems biology, regulation indicates which mechanisms the cell uses to accomplish changes in, or robustness of cellular functions [14]. For the cells, these changes allow i) the internal communication between different compartments, cells or organisms, ii) the capability of living organisms to respond to changes in its environments, and iii) the maintenance of homeostasis upon external perturbations [15].

In order to study the importance of each regulatory route during a specific adaptation of an organism a quantitative approach is needed [14]. This approach was addressed by the regulation analysis theory [16] which, in its original version, enables a quantitative dissection of the *in vivo* regulation of the enzymes fluxes by gene expression and metabolite level changes as represented in Equation 1.1 [17, 18]:

$$\frac{\Delta \log J}{\Delta \log J} = \frac{\Delta \log f(e)}{\Delta \log J} + \frac{\Delta \log g(X,K)}{\Delta \log J} = \rho_h + \rho_m = 1 \quad (1.1)$$

Where  $J$  is the metabolic flux at steady state of each specific metabolic reaction,  $e$  is the enzyme concentration;  $X$  is a vector of the concentration of the different metabolites which may affect the reaction and  $K$  is a vector including all the metabolic reaction parameters.

The  $\rho_h$  coefficient in Equation 1.1 represents the hierarchical regulation which quantify to what extent the local flux through the enzyme is modulated by the change in the enzyme capacity ( $V_{max}$ ). This coefficient may be influenced by the gene expression regulation which, at the same time, could be dissected by gene transcription, mRNA processing, transport and stability, translation and post-translational modification, and protein stability regulation [14].

On the other hand, the metabolic regulation coefficient ( $\rho_m$ ) represents the intracellular interaction changes of substrates, products and allosteric effectors with the enzymes. In many publications it can be found studies related with the determination of the intracellular metabolite levels impact on the metabolic fluxes. Besides, Heijnen [19] performed a review focused on the different existing kinetic approximations used to relate the intracellular metabolite levels and the metabolic fluxes. In this review, independently of the approximation used, all the modelling approaches share two fixed kinetic assumptions which could be related with the gene expression and metabolic regulation:

1. The metabolic rate is directly proportional to the level of active enzyme. These enzyme levels together with the specific catalytic rate constant ( $k_{cat}$ ) define the enzyme capacity ( $V_{max}$ ).
2. There is a saturation type kinetic behaviour of the  $J/e$  ratio versus the concentration of the different reactant metabolites according to a Michealis-Menten mechanism.

Performing regulatory analysis, Rossell and co-workers [20] classified the possible numerical outcomes into five different categories starting from a reference steady-state:

- Purely hierarchical regulation: the metabolic changes do not make a net contribution to a change in the enzyme rate ( $\rho_m = 0$ ) being the changes in  $V_{max}$  which completely describe it ( $\rho_h = 1$ ).
- Purely metabolic regulation: oppositely to the previous category, the flux through the enzyme is only modulated by the intracellular metabolite interactions with the enzyme ( $\rho_m = 1$ ;  $\rho_h = 0$ ).
- Cooperative regulation: when there is a cooperative enzyme flux modulation for the gene expression and metabolic regulation ( $0 < \rho_h, \rho_m < 1$ ).
- Antagonistic regulation directed by  $V_{max}$ : in this category the gene expression and the metabolic regulation have antagonistic effect over the enzyme rate. Nevertheless,  $\rho_h$  is larger than  $\rho_m$  ( $|\rho_h| > |\rho_m|$ ).
- Antagonistic regulation directed by the metabolic regulation: similarly to the previous case, gene expression and the metabolic regulation have antagonistic effect over the enzyme rate. However,  $\rho_m$  in this case is the larger coefficient ( $|\rho_m| > |\rho_h|$ ).

## Fluxome

In the regulation analysis theory, the metabolic flux analysis (MFA) is taken as the closer real representation of the physiological state of the cells being it the result of the other omics levels interactions. Therefore, being the MFA outcome the combination of gene expression and metabolic regulation.

MFA has become a well-established diagnostic tool in metabolic engineering and in systems biology in the recent years [21] being it an invaluable tool for characterising different strains or physiological states of a microorganism [22]. The development of MFA started with a purely stoichiometric approach which required only measured extracellular fluxes (substrate uptake, biomass production, end product formation,...) as input data [21, 23]. Nevertheless, some major drawbacks of a purely stoichiometric approach became apparent:

1. Strong assumptions about the cellular energy metabolism are required like P/O ratios and the knowledge of all intracellular reactions steps to produce-consume ATP, NADH or NADPH [24].
2. Limitation on the detailed resolution of intracellular fluxes. In particular, the forward and backward directions of reversible reactions steps or parallel reaction steps like the various anaplerotic reactions.

These problems led to the development of the  $^{13}\text{C}$ -MFA methods which work without any energy balance and can resolve bidirectional or parallel reactions steps [21]. Specifically, a labelled carbon substrate together with unlabelled are used as carbon sources in the experiments, leading to a labelled material distribution over the central carbon metabolism. Afterwards, when the metabolic steady state of the label enrichment in each intracellular metabolite pool is achieved, NMR or MS measurements are performed to determine the label material distribution in the intracellular metabolites. These measurements become extra metabolic constraints, which allow the resolution of the  $^{13}\text{C}$ -MFA [13].

Considering the possible uses of the  $^{13}\text{C}$  derived metabolic information, two approaches have been performed in the literature. The first approach integrates simultaneously the  $^{13}\text{C}$  data, extracellular fluxes, and biosynthetic requirements in computer models. The flux distribution is then obtained by fitting the intracellular fluxes to the experimental data by minimizing the differences between the observed and simulated isotopologue spectra [21, 25]. The second method is the flux ratio method, in which the relative distribution of converging pathway to the formation of a specific metabolite is quantified from particular combination of NMR or mass pattern [13, 26–28].

Nowadays, the  $^{13}\text{C}$ -MFA has been evolved not only to measure the metabolite fluxes in steady state conditions but also in dynamic experiments [21].

**Box 1: Central Carbon Metabolism**

Metabolism is the set of chemical reactions that happen in living organisms to sustain life. The different reactions are usually divided into two categories:

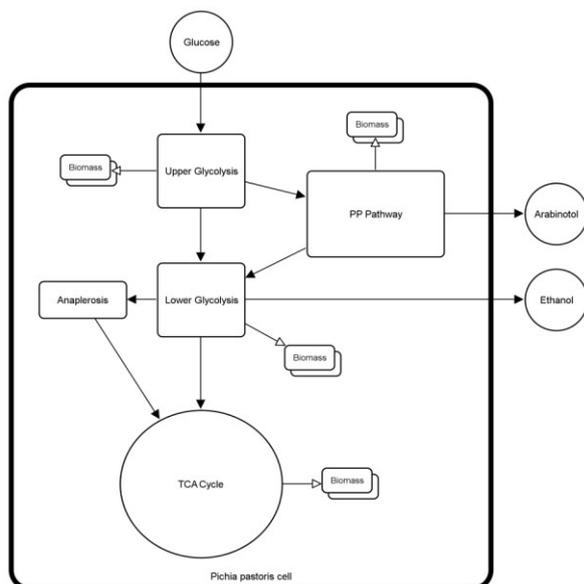


Figure 1.2. *Pichia pastoris* central carbon metabolism

- **Catabolism:** reactions which break down organic compounds in order to supply carbon precursors, reducing power and energy used for the cell growth.

- **Anabolism:** reactions which use the products from the catabolism reaction to construct the cell components such as proteins, carbohydrates, lipids and nucleic acids.

The central carbon metabolism is the set of reactions which represent the main important metabolic pathways of both metabolism categories which could change from an organism to another.

In *Pichia pastoris* case, its central carbon catabolism consists of glycolysis, pentose phosphate pathway, TCA cycle, the fermentative pathways for the ethanol and arabinol production and the anaplerotic reactions (Figure 1.2). On the other hand, the central carbon anabolism includes the biosynthetic reactions of the biomass macromolecular components.

### Mathematical framework of stoichiometric MFA

The MFA is based on the principle of mass conservation of the intracellular metabolites in a known stoichiometric network. By measuring the net conversion rates of the extracellular metabolites, assuming steady state for the intracellular metabolite concentration and neglecting the dilution effects of the growth, the mass balances of the metabolites can be written as:

$$\begin{bmatrix} S \\ R \end{bmatrix} \cdot v = \begin{bmatrix} 0 \\ r_m \end{bmatrix} \quad (1.2)$$

Where  $S$  represents the  $N$ -by- $V$  stoichiometric matrix, where  $N$  is the number of intracellular metabolite and  $V$  the total number of reactions included in the model.  $R$  is a  $M$ -by- $V$  stoichiometric matrix, where  $M$  is the number of measured net conversions rates in the rate vector  $r_m$ .

Along the thesis, all the equations systems were determined allowing a lineal computation to solve it. In case the system would have been undetermined, only a lineal combination of the fluxes could be calculated or an optimization function would have been to be applied.

## Metabolome

The metabolome of even a simple microbe exceeds 1000 different compounds making an extremely demanding task the entire metabolome analysis. The fact is that the metabolomics field has experienced a strong development over the recent past, mainly due to the improvements in MS-based analytical procedures [11, 29–31], which has greatly expanded the possibilities for quantitative analysis of intracellular metabolite pools.

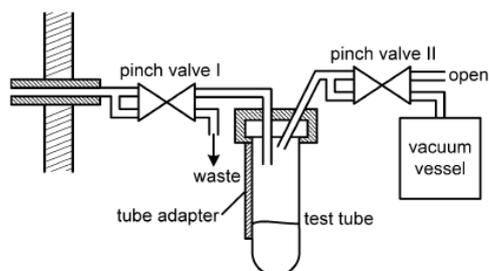


Figure 1.3: **Rapid sampling setup for metabolome fingerprint analysis.** Setup composed by two electro valves, one vacuum vessel and a tube adaptor for the quenching solution which allow sampling times lower than 1 second. Pictured derived from [32].

Although there have been many analytical measurement improvements in the field, an important step for the metabolic reaction network study is to obtain representative and accurate snapshots of the metabolome. Many metabolites, not only those related to the central carbon metabolism but e.g. also free amino acids, have turnover times in the order of seconds or less, which highlight the need for a rapid sampling technique (Figure 1.3) to quantify their actual levels, as well as a proper, leakage-free, quenching procedure to ensure absence of losses or (inter)conversion of metabolites [9, 33]. Moreover, if substantial amounts of metabolites are present in the extracellular medium, those need to be efficiently removed. Therefore, a sample treatment procedure allowing separating the intracellular and extracellular metabolite pools, as well as a degradation-free extraction method are required [34].

Although many efforts have been directed towards the development of a universal method, no consensus solution has been found because of the vast diversity in cell properties [9, 10, 35]. Recently, proper sampling, quenching, separation and extraction protocols for the intracellular metabolite quantification in *Saccharomyces cerevisiae* have been presented and successfully applied, allowing accurate, reliable and reproducible metabolite determinations [33, 34]. Unfortunately, it appears that for different microbial species, different sampling and quenching methodologies need to be developed and quantitatively evaluated and validated [10, 36, 37].

### Experimental setup and Data consistency

For systems biology studies, a well-controlled environmental condition of the experimental setup is crucial for the well physiological state characterization. For this reason, chemostat cultures were chosen as the best culture strategy due it facilitates the study of cellular responses to a particular

perturbation (e.g. oxygenation, temperature, pH,... effects) while keeping the other parameters constant. Besides, such perturbations may also include punctual stimulus (e.g. a glucose pulse) as well as other controlled parameter changes like shifts or gradients.

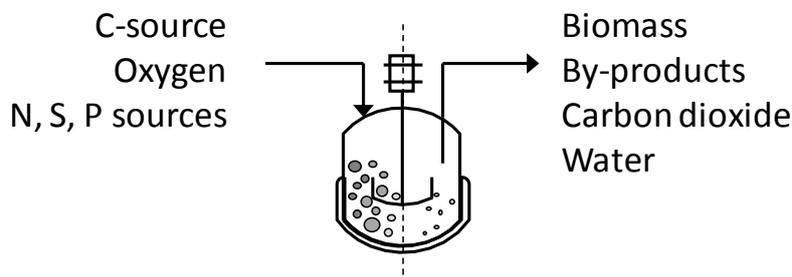


Figure 1.4: **Black box model of the chemostat experimental system**

In addition, the chemostat cultures allowed, as well as the well-controlled environmental condition, black box modelling of the experimental system due to the fully characterization of the cell metabolic steady state (Figure 1.4). Along this thesis, a black box model was performed using the elemental balances as constrains for all the chemostat cultivations realized. These models led to over-determined systems which allowed statistical consistency test [38], at 95% confidence level, proofing measurement errors.

Overall, the strict control of such cultivations provide strongly reproducible cultures and hence, little variations on data acquisition. Therefore, it is the most adequate strategy for system biology studies where high-quality data are required at the different omics levels.

## Microbial Cell Factories

One of the major requirements for development and production of recombinant proteins is the capability of a given cell factory to produce sufficient biologically active material that allows up-scaling for mass production [39]. Over the last two decades significant progress has been made in heterologous protein production, particularly due to the initiation of the genomics era. The acquisition of profound knowledge and entire genome sequences of a number of expression platforms has lead to considerable success for the production of many pharmaceutical proteins or industrial enzymes [2, 40].

Choosing an appropriate method for expressing a recombinant protein is a critical factor in obtaining the desired yields and quality of the recombinant protein in a timely fashion. Selecting a wrong expression host can result in the protein being misfolded or poorly expressed, lacking the necessary post-transcriptional modifications or containing inappropriate modifications.

In academic and industry, there are numerous expression systems currently being used although, some of these are too new and insufficiently tested to comment on their utility. In addition, some established systems for expressing recombinant proteins, such as transgenic animals, are too technically challenging, time consuming or prohibitively expensive to be a viable option for average laboratories. In

Table 1.1, there are a small characterization of four well establish expression systems which have straightforward protocols, are readily accessible and are relatively inexpensive for small-scale production.

**Table 1.1:** Summary of expression systems

Expression System	Advantages	Disadvantages
<i>Escherichia coli</i>	Rapid expression method (days)	Limited capacity for post-translational modifications
	Inexpensive bioproduction media and high density biomass	
	Simple process scale-up	Difficult to produce some proteins in a soluble, properly folded state
	Well characterized genetics	
<i>Pichia pastoris</i>	Moderately rapid expression method (weeks)	N-linked glycan structure different from mammalian forms
	Inexpensive bioproduction media and high density biomass	Enhanced safety precautions needed for large-scale bioproductions due to methanol in induction media (pAOX1 promoter)
	Most post-translational modifications and high folding capacity	
	Well characterized genetics	
Baculovirus/Insect cell	Moderately rapid expression method (weeks)	N-linked glycan structure different from mammalian forms
	Most post-translational modifications and high folding capacity	Low biomass density and expensive bioproduction media
		Difficult scale-up process
Mammalian - Transient expression	Moderately rapid expression method (weeks)	Low biomass density and expensive bioproduction media
	All post-transcriptional modifications and high folding capacity	Difficult scale-up process
Mammalian - Stable expression	All post-transcriptional modifications and high folding capacity	Lengthy expression methods (months)
		Low biomass density and expensive bioproduction media
		Difficult scale-up process

### ***Pichia pastoris* as a Host**

The methylotrophic yeast *P. pastoris* (Figure 1.5) has been developed into a highly prosperous system for the production of a variety of heterologous proteins (for a review, [4, 41]). Nevertheless, understanding the mechanisms governing efficient production of complex proteins as functional entities still remains a major challenge.

The first yeast routinely used for recombinant protein expression was *Saccharomyces cerevisiae* [42]. However, in the last 15 years, *P. pastoris* has become the yeast of choice because it typically permits higher levels of recombinant protein expression than does *S. cerevisiae* [43]. The growth of *P. pastoris* in methanol-containing medium results in the dramatic transcriptional induction of the genes for

alcohol oxidase (AOX) and dihydroxyacetone synthase [43]. After induction, these proteins comprise up to 30% of the *P. pastoris* biomass. Investigators have exploited this methanol-dependent gene induction by incorporating the strong, yet tightly regulated, promoter of the alcohol oxidase I (AOX1) gene into the majority of vectors for expressing recombinant proteins [41]. The *P. pastoris* expression vectors integrate in the genome whereas by contrast, *S. cerevisiae* vectors use the more unstable method of replicating episomally. The length of time to assess recombinant gene expression with the *P. pastoris* method is approximately 3–4 weeks which includes the transformation of yeast, screening the transformants for integration, and an expression timecourse. An appealing feature of *P. pastoris* is the extremely high cell densities achievable under appropriate culture conditions [43]. Using inexpensive medium, the *P. pastoris* culture can reach 120 g/l of dry cell weight density. An important caveat is that the induction medium requires a low methanol percentage. In large-scale cultures, the amount of methanol becomes a fire hazard requiring a new level of safety conditions. *P. pastoris* has been used to obtain both intracellular and secreted recombinant proteins. Like other eukaryotes, it efficiently generates disulfide bonds and has successfully been used to express proteins containing many disulfide bonds. To facilitate secretion, the recombinant protein must be engineered to carry a signal sequence. The most commonly used signal sequence is the pre-pro sequence from *S. cerevisiae*  $\alpha$ -mating factor [41]. Because *P. pastoris* secretes few endogenous proteins, purification of the recombinant protein from the medium is a relatively simple task. If proteolysis of the recombinant protein is a concern, expression can be completed using the pep4 protease-deficient strain of *P. pastoris*.

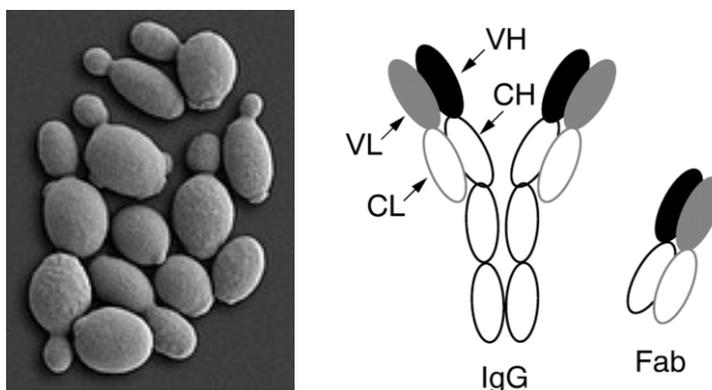


Figure 1.5: Images of *Pichia pastoris* cells (left side) and antibody fragment structure (right side).

### Antibody fragment as model for heterologous protein production

Throughout the thesis, a *P. pastoris* strains producing the antibody fragment (fab) 3H6 [39] have been used to characterize the heterologous protein production impact over the physiology (Figure 1.5). This protein has been used as model for complex protein production as it shows its typical features: it consists of multiple subunits that must bind to each other in a stoichiometric relation and its structure depends on the availability of the subunits.

Specifically, this recombinant protein is an anti-idiotypic Fab fragment directed against the human monoclonal antibody (mAb) 2F5 which is one of the most promising neutralizing antibodies against HIV-1 [39]. Besides, the crystal structure of the Fab fragment mouse antibody Ab2/3H6 in complex with mAb 2F5 Fab fragment have been achieved obtaining potential information for future design of vaccine components against human immunodeficiency virus [39].

## Oxygen availability as environmental stress

Oxygen transfer is it often described as an important issue in high cell density fermentations. While affecting growth and protein production, oxygen also influences cellular redox reactions, and these are interlinked with protein folding reactions within the cell. Moreover, protein folding related oxidative stress has been described [44–46]. Paradoxically, it has demonstrated recently that hypoxic conditions in chemostat as well as fed batch cultures significantly increased the specific productivity of recombinant *P. pastoris* [47]. However, it is not straightforward to predict whether this low oxygen availability effect would be also observed in other expression systems like *S. cerevisiae*. Yeasts do not perform in an identical way, as they can differ in productivity and with regard to their capacity to secrete, to process and to modify proteins in particular cases. As a consequence, it is important to systematically identify the complex mechanisms ruling efficient protein production, and integrated omics studies are a valuable tool for the study of biological processes.

## Aim and Outline of the Thesis

The aim of this research project was to contribute to the understandings of the *Pichia pastoris* physiology under different stress conditions. The project was in close collaboration with another thesis project realized by Kristin Baumann where complementary analyses were performed aiming to develop a multiple level analysis, from genome to fluxes, for characterizing oxygen availability and recombinant protein production physiological impact.

This thesis is centred in the systematic analysis of the metabolome and fluxome omic levels, and how the different levels contribute to regulate the final metabolic flow. In the first two chapters of the thesis the *P. pastoris* fluxome is characterized. Specifically, in **Chapter 2** a consistent description of the biomass principal components of *P. pastoris* is obtained for different oxygenation conditions, as well as under recombinant protein producing and non-producing backgrounds. In addition, consistency of the input and output extracellular metabolic fluxes is used to identify the production of an unexpected by-product, which was subsequently identified as arabinitol. Afterwards, in **Chapter 3**, a comparison of the different metabolic steady-states performed under different culture conditions was performed as well as a minimal stoichiometric by  $^{13}\text{C}$ -NMR analysis aimed to describe the metabolic flux of the different steady-states.

The issue of metabolome analysis is started in **Chapter 4** with the systematic evaluation of five quenching protocols applied to *P. pastoris*. Methanol concentration and temperature were the selected

variables to be optimized for minimizing metabolite leakage during quenching and subsequent washing. Thereafter, the best performing quenching protocol was used in two applications: 1) determination of the time needed for each intracellular metabolite to reach a metabolite steady state and 2) a metabolome comparison between *P. pastoris* and *S. cerevisiae*.

In Chapters 5, 6 and 7, the *P. pastoris* metabolome adaptation to recombinant protein production and to different oxygen availability is described together with a global metabolome analysis. Specifically, in **Chapter 5**, a network-embedded thermodynamic analysis of the quantitative intracellular metabolites pools was performed in order to obtain more information of the intracellular metabolic networks. Consecutively, in **Chapter 6**, the transcriptional and thermodynamic regulations over the metabolic flow between conditions were investigated for the central carbon metabolism reactions.

In **Chapter 7**, to better understand the effect and interplay of oxygen availability and foreign protein secretion on central metabolism, a first quantitative metabolomic analysis of free amino acids pools in a recombinant *P. pastoris* strain growing under different oxygen availability conditions was carried out.

Finally, **Chapter 8** discusses the general outcome of the studies presented in this thesis and provide the reader with possible leads for future research.



## Chapter 2: *Pichia pastoris* Biomass Composition<sup>a</sup>

---

<sup>a</sup> Published as: Carnicer M, Baumann K, Töplitz I, et al. **Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels.** *Microbial Cell Factories* 2009, **8**:65-78



## Background

Genome-scale *in silico* metabolic models are nowadays being built for a number of species to get highly informative view of the physiological cell status at each experimental condition [48, 49]. Validation and practical application of such complex models require obtaining reliable experimental data on a number of metabolic fluxes among which those related to biosynthesis precursors for cell constituents play a key role [50, 51]. Considering that cells are composed of different biopolymers and macromolecules, knowledge of their composition and quantity becomes essential for determination of the metabolic fluxes of biosynthetic precursors, as well as any other metabolic or energetic analysis. Availability of such molecular compositional data is scarce or inexistent for a particular strain or species and growth condition, particularly in non-model organisms as e.g.: *P. pastoris*.

In this chapter, a consistent description of the biomass composition of *P. pastoris* is obtained for different oxygenation conditions, as well as under recombinant protein producing and non-producing backgrounds, and consistency of the input and output extracellular metabolic fluxes is used to identify the production of an unexpected by-product, which is subsequently identified as arabinitol.

## Materials and Methods

### **Strains**

In this study two different strains were used. A *P. pastoris* X-33 (wild type phenotype, Invitrogen) transformed with pGAP $\alpha$ A (Invitrogen) as mock vector (control strain) , and a X-33-derived strain expressing the human antibody fragment (Fab) 3H6 fragment under the transcriptional control of glyceraldehyde-3-phosphate dehydrogenase (GAP) constitutive promoter, using the secretion signal sequence from the *S. cerevisiae*  $\alpha$  mating factor (Fab-expressing strain) [52].

### **Cultivation conditions**

Chemostat cultivations were performed in a 2-liter bench-top bioreactor (B. Braun Biotech International, Biostat B), as previously described [47]. Briefly, cells were grown under carbon-limited conditions at a dilution rate ( $D$ ) of  $0.1 \text{ h}^{-1}$ , with different oxygen concentrations in the bioreactor inlet air. The total inlet gas flow was controlled by mass flow meters (Bronkhorst High-Tech) at 1.5 vvm. The outlet gas flow was dried through two silica gel columns to remove the humidity and analyzed to determine its  $\text{CO}_2$  and  $\text{O}_2$  content (BCP- $\text{CO}_2$  & BCP- $\text{O}_2$  Sensors, BlueSens). In the cultures that received less than 21% of  $\text{O}_2$  in the inlet gas stream (that is, 11% and 8% of  $\text{O}_2$ ), air was mixed with the necessary volume of  $\text{N}_2$  so that the total gas flow of 1.5 vvm was maintained constant. The pressure in the culture vessel was maintained at 1.2 bars using a pressure valve (GO Inc). The pH, stirring speed and temperature were maintained at 5 (with 20%  $\text{NH}_3$ ), 700 rpm and 25 °C, respectively. During the

experiments the data acquisition and control of the different variables was done using UAB proprietary software (O. Cos, Department of Chemical Engineering, UAB).

For each independent experiment, chemostat culture conditions were maintained constant for at least 5 residence times to allow reaching a metabolic steady state, except for the hypoxic conditions (8 % O<sub>2</sub>), where a wash out after 3.5 residence times had previously been observed [47]. In this case only 3 residence times were awaited. At that point, samples were taken and analyzed for biomass composition. At least three independent samples were analyzed for each tested condition

## ***Biomass Analyses***

### Determination of biomass ash content

Culture samples of 20 ml volume were centrifuged at 5,000 rpm, 4 °C for 5 min and the pellets were washed twice in 20 ml of 20 mM Tris·HCl, pH 7.6. The resuspended cells were filtered through pre-dried glass fibre filters (Millipore) and dried overnight at 100 °C. Combustion of the pre-dried biomass was done by placing the glass fibre filters in ceramic cups (Millipore) in an oven (Hengstler) at 550 °C for 12 hours.

### Biomass lyophilisation

Most of the analyses were performed using lyophilized biomass samples. To lyophilize the biomass a sample of cultivation broth was centrifuged at 13,000 rpm for 1 min and the cell pellet was washed twice with 1 ml of 20 mM Tris·HCl, pH 7.6 to get a pellet free of culture medium. The recovered pellet was immediately frozen by immersion in an acetone-dry ice mixture followed by the lyophilization step under vacuum (Virtys Sentry).

### Biomass elemental analysis

The elemental composition of the biomass was determined taking 1mg of a lyophilized biomass, adding 1 mg of V<sub>2</sub>O<sub>5</sub> and introducing the samples in an oven at 1000 °C. The volatile compounds were measured with an Elemental Analyzer (NA2100 ThermoFisher).

### Biomass amino acid composition

Samples of 40 mg of lyophilized biomass were hydrolyzed with 3 ml HCl 6 M at 105 °C for 24 hours. A N<sub>2</sub> stream was used to remove the O<sub>2</sub>, minimizing the oxidation during hydrolysis. Following the hydrolyzation step, the volume of the samples was completed up to 50 ml with deionised water (MilliQ). A volume of 500 µl of the diluted samples and 100 µl of Nor-Leucin 2500 µM (non-proteinogenic amino acid used as internal standard), were mixed before being evaporated to dryness. The resulting pellets were dissolved with 1 ml buffer solution pH 2.2 and filtered by ultracentrifugation using 10,000 Da cut-off filter (Microcon YM10; 42408). A 50 µl volume of the filtrate was injected in an amino acid analyzer (Biochrom 30; Biochrom UK, Software: EZ Chrom) using a cation exchange

chromatography column and post derivatisation with ninhydrine, (as described by Spackman, Moore and Stein in 1958, [53]). For the amino acid identification, the eluent (lithium citrate buffer) was heated up to 135 °C and mixed with ninhydrine. Identification was done according to the known retention time of the amino acid standards in this conditions and quantification performed according to the previously performed calibration curves relating peak area with concentration (detection limit: 1 nmol, reproducibility: 1.5 % for 10 nmol).

#### Biomass total protein content

Total protein content was determined by means of the Lowry method as described in [54] from a solution of lyophilized biomass at 0.5 g/l dry weight. Protein concentration was calculated using BSA as standard.

#### Biomass total carbohydrates content

Total carbohydrates were determined by the phenol method as described in [54]. A 1 ml sample of lyophilized biomass (0.1 mg dry biomass/ml) was mixed with 1ml phenol 5 % and 5 ml 96 % sulphuric acid. After 10 min tubes were cooled (15 min 25 °C). Absorbance at 488 was measured using glucose solutions as standard. Results were corrected for the presence of nucleic acid pentoses using a relative absorbance of 0.455 and 0.264 for RNA and DNA, respectively [55].

#### Biomass glycogen content

The glycogen content was analyzed as described in the literature [23]. Glycogen was hydrolyzed adding 10 ml of 0.6 M HCl to 20 mg of freeze dried biomass and incubating in a water bath at 100 °C for 1h. After cooling, samples were filtered through 0.22 µm filter. The glucose produced was quantified using an automatic glucose analyzer (YSI 1500, Yellow Springs Instruments, OH, USA).

#### Biomass trehalose content

Determination of trehalose content of the biomass was performed according to [56]. 10 mg of lyophilized biomass was resuspended in 5 ml water and incubated for 15 minutes in a water bath. After cooling and centrifugation the cellular extract was rinsed twice with water and resuspended in 0.1 M acetate buffer pH 4.5. A sample of 800 µl of the washed cellular extract was mixed with 200 µl of a threhalase (Sigma) solution (1.37 µU/µl in acetate buffer, pH 4.5) and incubated overnight. The glucose released was determined using a glucose oxidase kit (Sigma) using the same reaction mixture without threhalase as a blank.

#### Biomass lipid content

Lipid extraction was performed according to [57]. Briefly, 150 mg of freeze dried biomass were extracted with hexane:isopropanol (3:2 v:v) overnight. After adding 0.47 M Na<sub>2</sub>SO<sub>4</sub> phases were separated with the aid of a centrifugation step. The hexane phase was recovered and evaporated to

dryness under a N<sub>2</sub> flow. The increase in dry weight of the tube was taken as total lipids. Pellet was redissolved in chloroform and further analyzed by thin layer chromatography using silica gel plates (Merck) and a mobile phase of chloroform:methanol:water:acetic acid (345:133:21:3). A standard was analyzed in parallel containing phosphatidic acid (Sigma), phosphatidyl serine (bovine, Fluka), phosphatidyl glycerol (bovine, Sigma), cardiolipine (bovine, sigma), phosphatidyl ethanolamine (bovine, sigma), phosphatidil inositol (bovine, sigma), phosphatidyl coline (bovine, sigma), ergosterol (Fluka) and triacilglycerides (Sigma) using a range of concentrations from 1 to 8 µg of each standard. After running the chromatography the plates were stained either with iodine or with Sudan Black (Sebia). The developed images were quantified using an image analysis software (Multi gauge, Fujifilm) comparing the intensity of the stains with the standards.

#### Biomass DNA content

DNA content of biomass was determined by means of the Hoechst fluorescent dye method as described in [58]. In short, lyophilized biomass samples were dissolved in TNE buffer (NaCl 1 M, EDTA 10 mM, Tris-HCl 0.1 M, pH 7.4) at a concentration of 25 mg/ml. Sample solution was mixed with 2 mL of the Hoechst dye solution (Hoechst 33258 0.5 µg/ml in TNE buffer). Fluorescence was measured using the excitation/emission wavelengths of 356/468. DNA content was calculated by interpolation in a calibration curve performed using standard DNA (DNA sodium salt from calf thymus from Sigma-Aldrich).

#### Biomass RNA content.

RNA content of biomass was determined according to Benthin [59]. Briefly 5 mg of lyophilized biomass were resuspended in 10 ml of cold HClO<sub>4</sub> 0.7 M and incubated for 5 min. After incubation biomass was centrifuged (8000 rpm, 10 min, 4 °C), washed twice and resuspended in 10 ml KOH 0.3 M. Two 5 ml aliquots of the resuspended biomass were incubated at 37 °C for 1 hour. After cooling 1ml cold HClO<sub>4</sub> 3 M was added and samples were centrifuged in the same conditions. Supernatant was collected and the pellet washed twice with 1 ml HClO<sub>4</sub> 0.5 M. The 3 supernatants collected were mixed and absorbance measured at 260 nm in a quartz cuvette. The percentage (w/w) was calculated using A<sub>260</sub>: 0.038 mgRNA/ml and taking into account the sample dilution.

#### ***Fab quantification by ELISA***

Fab amounts in soluble cell extracts and in culture broths were performed by means of a sandwich ELISA assay, as previously described [52]. For the quantification of the intracellular Fab cells were disrupted using a cell disruptor (One Shot. Constant Cell Distruption System LTD, The Netherlands). Briefly, 5 ml of 10 g/l of lyophilized biomass resuspended in PBS buffer supplemented with 2 % (w/v) SDS, complete protease inhibitor cocktail (Roche) and 0.1 % β-mercaptoethanol, were disrupted at 2 Kbar of pressure in one shot. Fab analysis was performed as for soluble celll extracts.

### ***Bioreactor culture broth concentrations***

Cell biomass was monitored by measuring the optical density at 600 nm ( $OD_{600}$ ). For cellular dry weight, a known volume of cultivation broth was filtered using pre-weighed filters; these were washed with two volumes of distilled water and dried to constant weight at 105 °C for 24 h. Samples for extracellular metabolite analyses were centrifuged at 10,000 rpm for 2 min in a microcentrifuge to remove the cells and subsequently filtered through 0.45  $\mu$ m-filters (Millipore type HAWP). Glucose, organic acids, ethanol and arabinitol were analyzed by HPLC (Series 1050, Hewlett Packard) with an ionic exchange column (Bio-Rad, Aminex HPX-87H). As mobile phase, 15 mM sulphuric acid was used. The metabolites were detected (HP 1047A, Detector IR HP, Hewlett Packard) and quantified with the Software EmpowerProfor.

### ***Identification of extracellular metabolites***

#### LC-MS analysis

Extracellular metabolites were identified by means of LC-MS. For this purpose, a volume of about 50 ml of cleared culture supernatant was filtered through a 0.22  $\mu$ m Millipore filter. The analysis of filtered supernatants were performed on a Shimadzu Prominence HPLC with a UV/VIS detector coupled to a Mass Spectrometry detector Shimadzu 2010A equipped with an ESI (Electro Spray Ionization) interface operating at a wavelength of 210 nm.

Metabolite compounds were separated on a 300 mm  $\times$  7.8 mm Aminex HPX-87H column (Biorad) using 0.15 mM formic acid in miliQ water at pH 3 as mobile phase in isocratic mode. The analyses were performed at a flow rate of 0.6 ml/min at room temperature using a 20  $\mu$ l injection volume.

#### NMR analysis

To identify the unknown extracellular metabolite by NMR, one of the replica cultivation experiments under hypoxic conditions was further used to perform a  $^{13}\text{C}$ -labelling experiment with uniformly  $^{13}\text{C}$ -labeled glucose (Cortecnet, Paris), following a previously reported procedure [60]. In brief, isotopic labelling was achieved by feeding the bioreactor (at steady state conditions) with the medium containing about 12% (w/w) of uniformly  $^{13}\text{C}$ -labeled and 88% unlabeled substrate for one volume change. After that, a volume of 50 ml of fresh sample from the chemostat culture was centrifuged at 4°C, 5000 rpm for 5 min. The supernatant was filtered through a 0.22  $\mu$ m Millipore filter, stored overnight at  $-80$  °C and subsequently lyophilized. The lyophilisate residue was resuspended in a small volume of  $\text{D}_2\text{O}$  prior to NMR analysis.

$^{13}\text{C}$  NMR spectra of samples resuspended in  $\text{D}_2\text{O}$  were analyzed using a Bruker Avance 500 MHz spectrometer, operating at 125 MHz for  $^{13}\text{C}$ , equipped with a cryoprobe and with about one hour

accumulation. Spectra were referenced against an external standard in order to preserve sample recovery.

### Numerical calculations

All numerical calculations were performed using Matlab 2007b.

## Results and Discussion

In this study, the differences in biomass composition are evaluated for two different *P. pastoris* strains, namely the X-33-derived strain expressing the human antibody fragment Fab 3H6 [61] and a control X-33 strain transformed with the corresponding mock expression vector. For these two strains, three different oxygenation conditions (normoxic, oxygen-limited and hypoxic, corresponding to O<sub>2</sub> concentrations in the bioreactor inlet gas of about 21 %, 11 % and 8 % of O<sub>2</sub>, respectively) were assayed in carbon-limited continuous cultures using glucose as carbon source. As previously described [47], *P. pastoris* shows a fully respirative metabolism under normoxic conditions, a shift towards respirofermentative metabolism can be observed under oxygen-limiting conditions, and a clearly respirofermentative metabolism is shown under hypoxic conditions. For each case, the different biomass composition analyses described in the materials and methods section were performed. At this point, the different biomass chemical analyses can be considered at two levels, resulting either in an elemental or a molecular description of the biomass constituents. The variations and similarities observed at these levels were further used to propose a consensus composition for each case.

**Table 2.1:** Amino acid composition of the whole protein extract for all the conditions and strains tested in % mol/mol.

	Expressing Strain			Control Strain		
	21%	11%	8%	21%	11%	8%
ASX	8.8	9.8	9.6	8.8	9.5	9.8
THR	5.9	6.0	5.8	5.8	5.9	6.2
SER	6.3	6.6	6.7	6.4	6.8	6.9
GLX	17.8	16.2	15.6	18.5	16.2	14.3
PRO	3.8	4.1	4.5	3.7	4.0	4.8
GLY	6.9	7.3	7.4	7.1	7.3	7.5
ALA	10.4	9.9	10.1	10.7	9.7	9.9
VAL	5.9	6.0	6.1	5.6	6.8	6.2
CYST	0.2	0.2	0.1	s.n.	0.2	0.1
MET	0.8	0.7	0.8	0.8	0.8	0.7
ILE	4.6	4.7	4.8	4.1	4.5	4.7
LEU	7.0	7.3	7.5	7.0	7.2	7.9
TYR	2.2	2.2	2.3	2.1	2.2	2.5
PHE	3.2	3.3	3.3	3.0	3.2	3.4
ORN	1.0	1.0	0.9	1.5	1.2	0.9
LYS	6.4	7.1	6.9	6.3	7.0	6.9
HIS	1.9	1.9	1.9	1.8	1.8	1.9
TRP	1.4	1.4	1.4	1.4	1.4	1.4
ARG	7.0	5.7	5.8	6.7	5.7	5.4

### **Molecular biomass composition**

At the molecular level, and for simplicity, the biomass composition is usually described in terms of major groups of biomass constituents, namely proteins, carbohydrates, lipids, DNA and RNA. In some cases, each of these groups is further analyzed for their major constituents. This procedure allows to determine, not only their mass fractions and consensus global carbon molecular formula, but also which particular precursors and corresponding amounts are required for their biosynthesis.

**Table 2.2:** Average amino acid composition of the whole protein extract between the two strains for the different experimental conditions tested in % mol/mol. \* Data from [55] for *S.cerevisiae*.

	<i>Pichia pastoris</i>			<i>Saccharomices</i>
	21%	11%	8%	<i>cerevisiae</i> *
ASX	8.7	9.5	9.5	9.3
THR	5.7	5.9	5.9	5.6
SER	6.2	6.6	6.7	5.3
GLX	17.9	16.0	14.8	15.5
PRO	3.7	4.0	4.5	4.2
GLY	6.9	7.2	7.4	9.0
ALA	10.4	9.7	9.9	9.8
VAL	5.6	6.3	6.0	7.3
CYST	0.2	0.2	0.1	0.1
MET	0.8	0.7	0.7	1.1
ILE	4.3	4.5	4.7	5.9
LEU	6.9	7.1	7.6	8.0
TYR	2.1	2.2	2.3	2.0
PHE	3.1	3.2	3.3	3.8
ORN	1.3	1.1	0.9	0.2
LYS	6.3	6.9	6.8	6.6
HIS	1.8	1.8	1.9	1.9
TRP	1.4	1.4	1.4	2.0
ARG	6.8	5.6	5.5	3.9

### **Proteins and amino acid content**

Proteins are one of the major constituents of the biomass, usually representing around 50% of the dry weight of a microorganism [62]. Consequently, detailed knowledge of their composition is important for any metabolic and energetic calculations. As there are virtually no data on the amino acid composition of *P. pastoris* available, determination of this composition, for the different strains at each of the tested oxygenation conditions, was performed as shown in Table 2.1. These analyses revealed that, for a given oxygenation condition, the differences between the control and the Fab-expressing strain were, for most of the amino acids, below 9 %. Some exceptions were found for cysteine (for which, one of the analysis resulted in amounts below detection limits), ornithine, methionine and valine. Nevertheless, these differences were apparently not correlated to changes in experimental conditions (oxygenation level or Fab production). Therefore, in a first approach, it was considered that under the same oxygenation conditions, the control and expressing strains did not show significant differences in their amino acid composition. Consequently, average amino acid composition values of *P. pastoris* (shown in Table 2.2) for each oxygenation condition were considered for further compositional

data calculations. Notably, values given in Table 2.2, suggest that some amino acids seem to follow a consistent tendency (e.g. Glx, Orn, Arg) upon a decreasing oxygen availability, while others seem to remain at the same level (e.g. Thr, Trp, His, Met). Nevertheless, considering the standard deviations of the analysis (see Materials and Methods Section), further experimental data would be required to confirm these tendencies.

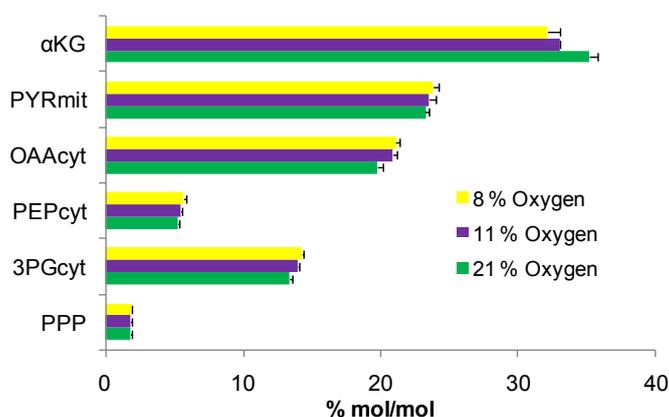


Figure 2.1: **Amount of key precursors required for protein generation at each experimental condition.** Using the average amino acid compositional data provided in Table 2.2, a calculation of the amounts of precursors consumed for protein generation was done.

From another point of view, Figure 2.1 shows the amount of precursors consumed to produce cell proteins (mols of each amino acid *per* 100 mols of amino acids) with the measured composition. For this calculation, the main *P. pastoris* biosynthetic routes for the amino acids were taken into account (revised according to its recent genome sequence [63, 64]). Amino acids were grouped according to their common precursor in the central carbon metabolism (Figure 2.1), showing that the requirements of amino acids precursors show a correlation with the variation of oxygen availability, with the exception of the precursors consumed from the pentose phosphate cycle. According to this observation, it can be expected that those variations would have an impact on the calculated metabolic fluxes and, therefore, it was decided to adopt one different biomass composition for each one of the three oxygenation conditions. In addition, data shown in Table 2.2 reveal that the measured amino acid composition for *P. pastoris* is significantly different to the previously published amino acid composition of *S. cerevisiae* [55]. Interestingly, Orn presents several fold variations in terms of relative amounts between the two yeasts, while Arg, Ile, Leu, Val, Gly, Met and Ser show relative abundance values with a more than 10% variation from the *S. cerevisiae* data. These results justify the adoption of a new amino acid composition for *P. pastoris* instead of the previously available compositional data from *S. cerevisiae*.

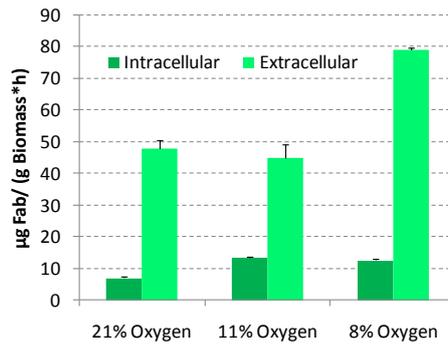


Figure 2.2: **Effect of decreasing oxygen supply on Fab heterologous protein production.** Measured Fab content in soluble cell extracts and in the extracellular medium for the different oxygenation conditions.

Total cell protein content was estimated or directly measured using two different methods. A first approximation was based on the sum of amounts for each of the amino acids extracted during the amino acid compositional analysis. Since this analytical technique involves an acid hydrolysis step of cell protein (see Materials and Methods), it was expected that, the value obtained by adding all the amino acids amounts should be on the lower margin. In the second, the protein content of the biomass was measured using the widely used Lowry method, using bovine serum albumin as standard. The different values obtained and the standard deviation of the different analytical approaches made it difficult to calculate a proper value for the protein content for each case. Therefore, total protein content values given by the two methods, together with their confidence intervals have been used in the data reconciliation step. This procedure allows to calculate the total protein content which best agrees with all the information available. The reconciled values indicate an increase in the global protein content correlated with the oxygen limitation (see reconciled biomass composition section). Besides, in the Fab-expressing strain, the amount of intracellular and extracellular heterologous protein was also measured in addition to the total protein content. Figure 2.2 shows that the amount of extracellular Fab increases when the oxygen concentration in the gas supply decreases down to 8%, in accordance with previous results [52]. At the same time, intracellular Fab content (which can also be considered as part of the total protein content) also increased significantly when reducing the oxygen availability. As the heterologous protein expression is under the control of the glycolytic GAP promoter this effect is in agreement with what can be expected. Nevertheless, further detailed studies are required in order to assess the potential impact of Fab production on total protein content and amino acid composition.

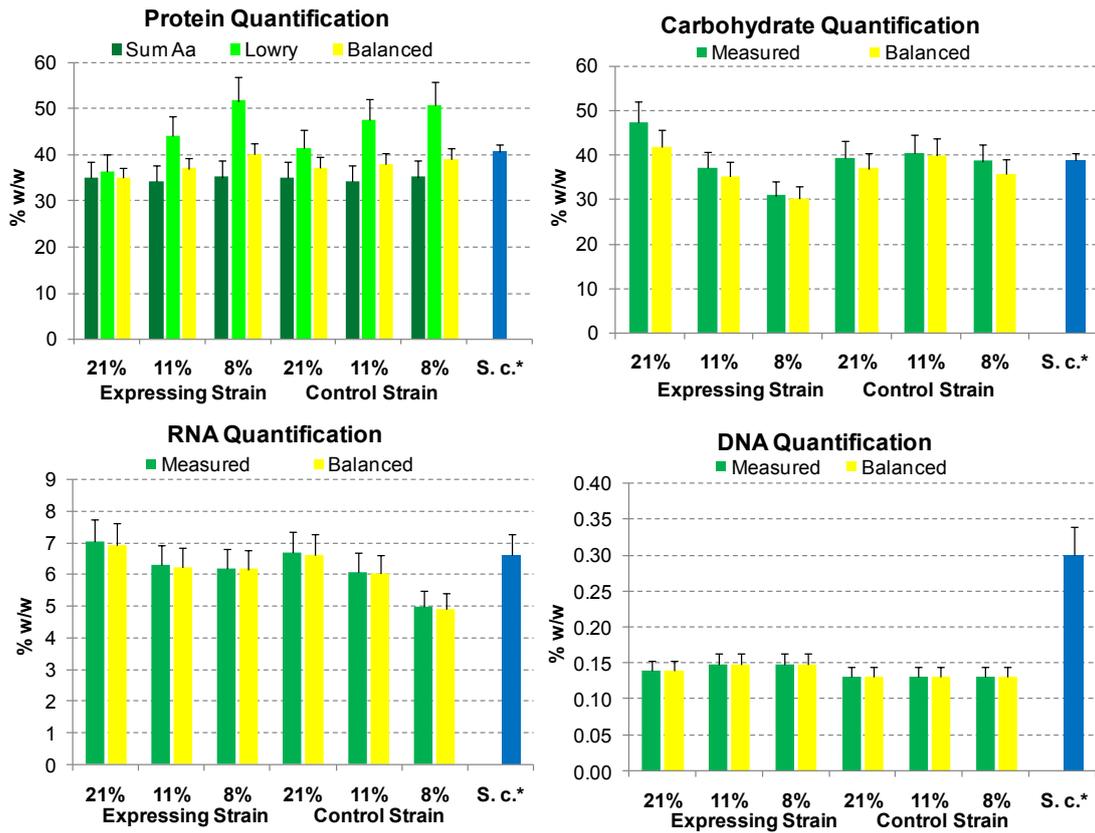


Figure 2.3: **Summary of the macromolecular biomass composition results obtained for the different experimental conditions and the final reconciled value.** Top left: Protein content in % DCW measured using the Lowry method and the sum of extracted amino acids, together with its reconciled value. Top right: Carbohydrate content % DCW together with its reconciled value. Bottom left: RNA content % DCW together with its reconciled value. Bottom right: DNA content % DCW together with its reconciled value. S.c.: *S. cerevisiae*. \* Data for *S. cerevisiae* in blue bars [55].

### RNA content

The results obtained for RNA content are shown in Figure 2.3 (bottom left). In this case, the control strain seemed to show a decrease in the RNA content as the oxygen availability was reduced, while in the Fab-expressing strain this effect is only apparent for the change from normoxic (21 % O<sub>2</sub>) to the lower oxygenation conditions and the RNA content in cells grown under oxygen limiting and hypoxic conditions remained similar.

Nevertheless, considering that RNA levels are expressed in relative terms (% weight RNA/ % dry cell weight), these variations do not appear to be relevant when compared to the corresponding variations observed for the major cell components (e.g. cell protein). On the other hand, when comparing the control strain with the Fab-expressing strain at the same growth conditions, the RNA levels appeared to be generally slightly higher in the Fab-expressing strain. Nevertheless, such differences were not statistically significant for all oxygenation condition. Notably, the relative decrease in RNA content

under reduced oxygen availability conditions seemed to be less pronounced for the Fab-expressing strain. One might speculate that the increase in heterologous protein production under hypoxic conditions could result in higher RNA levels than in the control strain. Overall, the average value of all RNA levels measured for *P. pastoris*, seems to be similar to those available from *S. cerevisiae* [55].

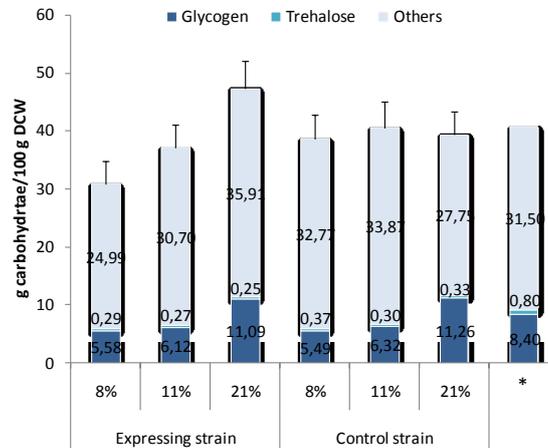


Figure 2.4: **Total carbohydrate, glycogen and trehalose amounts measured depending on the strain and oxygen conditions.** \* Data from [48] for *S.cerevisiae*.

### Carbohydrate content

Figure 2.3 (top right), shows the cell's total carbohydrate content measured for both strains grown under the three tested oxygenation conditions. The control strain seemed to present a fairly constant carbohydrate levels in all oxygenation conditions, which as an average, are comparable to the *S. cerevisiae* levels [55]. However, the Fab-expressing strain shows a significant decrease in its relative carbohydrate content as the oxygen availability decreased. This decrease is also in agreement with the increase in the relative total protein content and lower decrease in RNA content in the Fab-expressing strain than in the control strain as the oxygen availability is decreased.

Moreover, although the total carbohydrates content changes slightly, glycogen relative amounts were different depending on the oxygen conditions showing a significantly higher value in cells grown under normoxic conditions than in the oxygen-limited and hypoxic ones (Figure 2.4). These results could be explained by the need to secrete differently reduced carbon compounds to the media in limited-oxygen and hypoxic conditions to maintain the redox balance inside the cell (see Metabolite mass balance section) Interestingly, one of the glycogen synthases (*Gsy1*) from *S. cerevisiae* appears to be regulated by Rox1, a transcriptional factor that regulates responses to oxygen [65] so one could speculate that glycogen metabolism in *P. pastoris* is also regulated at the level of gene expression by an environmental stressor such as hypoxia. On the other hand, the trehalose values did not show any clear tendency when varying the oxygenation degree or the strain genetic background.

### **DNA content**

As shown in Figure 2.3 (bottom right), the measured DNA content of *P. pastoris* was lower than the amounts previously reported for *S. cerevisiae* [55]. This might partially reflect of the smaller size of the *P. pastoris* genome (haploid set of 4 chromosomes resulting in 9.43 Mbp [63]) compared to the one of *S. cerevisiae* (haploid set of 16 chromosomes resulting in 12 Mbp [66]). On the other hand, industrial strains of *S. cerevisiae* are usually found under diploid (or even polyploid) state, while *P. pastoris* is in haploid state. Combination of both factors could explain the differences among the measured DNA content of the two species.

### **Lipid content**

With respect to the lipid fraction content, an effort was done to quantify the total lipid components of the biomass as well as determining semi-quantitatively the major components of this fraction. However, the methodology employed showed low levels of total lipid recovery. Therefore, the data obtained was only used to determine the major components of the lipid family, whereas the total amount was instead estimated during the reconciliation procedure (see materials and method section). The total lipid content of the biomass calculated by the reconciliation procedure appeared to be close to the data available for *S. cerevisiae* [55].

On the other hand, the major lipid components identified in the thin layer chromatography were triacylglycerides (TAG), ergosterol (ERG), phosphatidylcholine (PC) and phosphatidylethanolamine (PE) in order of abundance. Moreover, the triacylglycerides amount seemed to increase when the oxygen availability decreased while ergosterol seemed to decrease. An average of those values was taken to calculate the lipid C-mol formula: triacylglycerides (57.2% w/w), ergosterol and ergosterol esters (26.8% w/w), phosphatidylcholine (10.0% w/w), phosphatidylethanolamine (5.2% w/w), phosphatidylinositol+phosphatidylserine (0.4% w/w), cardiolipin (0.3% w/w and phosphatidic acid (0.1% w/w).

### **Elemental biomass composition**

As explained above, besides the macromolecular composition of the biomass, its content of major chemical elements can be also considered. Usually, the elements that are taken into account are those that add up to nearly 90% of the dry weight, namely C, H, O, N, P, and S. In this study, only C, H, N, S and ash content were measured (see appendix 2.1). For the reconciliation procedure, P was included into a composite compound which also included ashes. Therefore, the reconciliation procedure calculated the oxygen content that fulfilled the global balance taking into account all the available information (see Reconciled biomass composition section).

### Consistency check and data reconciliation

As mentioned, in order to perform detailed metabolic flux analyses calculations, it is necessary to verify the consistency of the data obtained and, when different kinds of complementary or redundant data is available, to obtain the best estimation of the studied system using common data reconciliation techniques. This phase has been performed in two steps. In the first one, the different biomass compositional data available has been used to obtain a reconciled biomass composition. In the second step, the consistency check of the measured substrate consumption together with the biomass and products generated is performed.

### Reconciled biomass composition

Calculation of a reconciled biomass composition from the biomass compositional data described above was performed using the maximum likelihood method as described by Lange and Heijnen [55]. In short, the method allows the calculation of the best estimation of the biomass composition by solving a linear system which takes conservation laws of elemental and mass balances into account. The obtained experimental data is considered using a weight proportional to the confidence of the analysis.

**Table 2.3:** Biomass composition calculated after the reconciliation procedure. Biomass composition in terms of macromolecular and elemental compounds for the different experimental conditions (Normoxia: 21% O<sub>2</sub>; Oxygen-limited: 10% O<sub>2</sub> and Hypoxia: 8% O<sub>2</sub>).

	Fab Expressing strain						Control strain					
	21%		10%		8%		21%		10%		8%	
	% w/w	sd	% w/w	sd	% w/w	sd	% w/w	sd	% w/w	sd	% w/w	sd
Prot.	36.6 ± 2.0		38.4 ± 2.1		40.6 ± 2.2		38.2 ± 2.1		39.0 ± 2.2		40.1 ± 2.2	
Carbo.	41.3 ± 4.0		34.8 ± 3.3		30.0 ± 2.9		36.7 ± 3.5		39.7 ± 4.0		35.4 ± 3.5	
Lipid	5.4 ± 3.3		8.8 ± 3.1		9.8 ± 3.0		5.4 ± 3.1		6.9 ± 3.3		7.9 ± 3.2	
RNA	6.9 ± 0.7		6.2 ± 0.6		6.1 ± 0.6		6.6 ± 0.7		6.0 ± 0.6		4.9 ± 0.5	
DNA	0.14 ± 0.01		0.15 ± 0.01		0.15 ± 0.01		0.13 ± 0.01		0.13 ± 0.01		0.13 ± 0.01	
SO <sub>4</sub>	0.2 ± 0.2		0.2 ± 0.2		0.1 ± 0.2		0.3 ± 0.3		2.4 ± 3.5		0.1 ± 0.2	
H <sub>2</sub> O	3.0 ± 2.5		5.1 ± 2.4		6.7 ± 2.2		6.2 ± 2.4		0.2 ± 0.2		4.9 ± 2.4	
Metals	6.4 ± 0.4		6.4 ± 0.4		6.4 ± 0.4		6.4 ± 0.4		5.7 ± 1.2		6.5 ± 0.4	
C	44.2 ± 1.4		44.5 ± 1.4		44.4 ± 1.4		42.9 ± 1.4		45.5 ± 1.2		44.6 ± 1.4	
H	6.1 ± 0.2		6.4 ± 0.2		6.6 ± 0.2		6.3 ± 0.2		6.0 ± 0.2		6.4 ± 0.2	
N	6.9 ± 0.3		7.1 ± 0.3		7.5 ± 0.4		7.1 ± 0.3		7.2 ± 0.4		7.2 ± 0.4	
O	35.5 ± 1.4		34.7 ± 1.4		34.3 ± 1.4		36.4 ± 1.4		34.0 ± 1.3		34.6 ± 1.4	
S	0.2 ± 0.1		0.2 ± 0.1		0.1 ± 0.1		0.2 ± 0.1		0.2 ± 0.1		0.2 ± 0.1	
Ashes	7.1 ± 0.4		7.1 ± 0.4		7.1 ± 0.4		7.1 ± 0.4		7.1 ± 0.4		7.1 ± 0.4	

For the reconciliation procedure, the elemental composition of the macromolecular compounds, in terms of a C-mol formula, of each biomass component was calculated as follows: Protein elemental composition was calculated from the measured relative abundance of each amino acid in *P. pastoris*

samples. The elemental composition of carbohydrates was taken as identical as a glucose polymer. The elemental composition of lipids was calculated as an average of the major compounds. DNA elemental composition was calculated from the percentage of its nucleotide content in *P. pastoris* [63]. RNA elemental composition was calculated as proposed by Stephanopoulos [62]. Other components, such as metals, H<sub>2</sub>O and sulphate were included in an identical way as described for *S. cerevisiae* in Lange and Heijnen [55] (See appendix 2.2 for the C-mol formulas).

Table 2.3 shows the values for the biomass composition obtained after the reconciliation procedure was applied. The reconciliation procedure provides the best estimation of the biomass composition, including the lipid content or the oxygen values that were not measured. The biomass composition obtained this way is also consistent from the viewpoint that all the macromolecular elements considered add to 100% and the chemical elements composing them also add up to 100 %, while at the same time the standard deviation of the estimated values is also given. This is important for any further numerical treatment. Overall, the calculated biomass compositional values show that, as the oxygen availability is decreased, there is a slight increase in the protein content of the biomass. Also, the calculated data indicates an increase in the total lipid content, which has a higher degree of reduction than the proteins or the carbohydrates. Nevertheless, taking into consideration the standard deviation calculated for the lipids, the confidence for the calculated values is low and, therefore, more data will have to be obtained in the future to confirm the indicated tendency.

**Table 2.4:** Biomass C-molecular formula calculated after the reconciliation procedure. Biomass composition in terms of elemental compounds, expressed as the C-molecular formula, for the different experimental conditions (%O<sub>2</sub> in the inlet air: 21, 11, 8) \* Data from [55] for *S.cerevisiae*.

		<i>C-mol Biomass formula</i>				C:N Ratio	H:O Ratio	Red. Degree (γ)	
Fab Expressing strain	21 % oxygen	C	H <sub>1.665</sub>	N <sub>0.134</sub>	O <sub>0.602</sub>	S <sub>0.004</sub>	7.45	2.77	4.89
	10 % oxygen	C	H <sub>1.731</sub>	N <sub>0.137</sub>	O <sub>0.585</sub>	S <sub>0.001</sub>	7.30	2.96	4.98
	8 % oxygen	C	H <sub>1.775</sub>	N <sub>0.144</sub>	O <sub>0.579</sub>	S <sub>0.001</sub>	6.94	3.06	5.06
Control Strain	21 % oxygen	C	H <sub>1.761</sub>	N <sub>0.143</sub>	O <sub>0.636</sub>	S <sub>0.002</sub>	7.02	2.77	4.93
	10 % oxygen	C	H <sub>1.585</sub>	N <sub>0.135</sub>	O <sub>0.560</sub>	S <sub>0.002</sub>	7.39	2.83	4.88
	8 % oxygen	C	H <sub>1.726</sub>	N <sub>0.138</sub>	O <sub>0.581</sub>	S <sub>0.000</sub>	7.26	2.97	4.98
S. cerevisiae*		C	H <sub>1.748</sub>	N <sub>0.148</sub>	O <sub>0.596</sub>	S <sub>0.018</sub>	6.76	2.93	5.11

The reconciliation procedure also calculates the major chemical elements contained in the biomass, including the non measured oxygen. This allows calculating a biomass carbon elemental formula, shown in Table 2.4. It can be observed that the ratio H:O and the biomass degree of reduction increase as the oxygen availability is decreased. This is in agreement with the accumulation of more reduced biomass

components, for instance, an increase in lipid content. This effect is apparently more pronounced in the Fab-producing strain. The resulting reconciled biomass composition was used in the rest of the calculations.

### **Metabolite mass balances**

One of the key points in any fermentation study is to validate the measured consumption of substrates and generation of products. This step has been performed following the statistical procedures for the purpose [62, 67–69], which are similar to the one already followed for the reconciliation of the biomass composition data. The biomass C-molecular formula calculated in the previous step is used in this procedure. Then, the  $\chi^2$  – distributed consistency index  $h$  has been calculated for the reconciled data to test for the presence of gross errors according to a well established procedure [68, 69].

**Table 2.5:** Measured substrates and products consumed/produced at steady state in each experimental condition. Substrate consumption, biomass and metabolites production, for the different experimental conditions (%O<sub>2</sub> in the inlet air: 21, 11, 8). Bottom rows: values of the carbon balance mismatch in percentage including or not including arabinitol.

mmol/ gDW*h	Fab Expressing strain			Control strain		
	21%	11%	8%	21%	11%	8%
	sd	sd	sd	sd	sd	sd
Glucose	-0.99 ± 0.05	-1.33 ± 0.07	-1.74 ± 0.09	-0.99 ± 0.05	-1.32 ± 0.07	-1.92 ± 0.10
OUR	-2.12 ± 0.17	-1.57 ± 0.08	-0.54 ± 0.10	-2.12 ± 0.17	-1.66 ± 0.33	-0.28 ± 0.01
CER	2.37 ± 0.19	2.03 ± 0.10	1.65 ± 0.08	2.37 ± 0.19	2.09 ± 0.42	1.35 ± 0.27
Biomass	3.01 ± 0.30	3.71 ± 0.37	3.73 ± 0.19	3.59 ± 0.36	3.70 ± 0.37	3.76 ± 0.38
Ethanol	- -	0.33 ± 0.02	1.00 ± 0.05	- -	0.31 ± 0.02	1.16 ± 0.06
Glycerol	- -	0.01 ± 0.001	0.02 ± 0.001	- -	0.02 ± 0.001	0.02 ± 0.001
Citric Acid	0.03 ± 0.002	0.02 ± 0.001	0.04 ± 0.002	0.01 ± 0.001	0.02 ± 0.001	0.05 ± 0.02
Arabinitol	- -	0.21 ± 0.01	0.42 ± 0.02	- -	0.10 ± 0.01	0.40 ± 0.13
Pyruvic Acid	- -	0.06 ± 0.003	0.09 ± 0.005	- -	0.06 ± 0.002	0.12 ± 0.01
% C Balance Error w/o Arabinitol	6.4%	15.7%	23.9%	1.3%	14.5%	29.3%
% C Balance Error with Arabinitol	6.4%	2.5%	3.7%	1.3%	8.2%	11.9%

In a first approximation, it was considered that biomass and CO<sub>2</sub> generation under normoxic conditions resulted from the glucose and oxygen consumption, while reduction in the oxygen availability resulted in the generation some by-products such as pyruvate, glycerol, ethanol and citrate. As can be observed in Table 2.5, a carbon balance performed in both strains for the normoxic conditions (21 % of oxygen in the inlet air), indicated that most of the carbon was taken into account, as there was only a mismatch of 6% and 1% respectively, well in agreement with the precision of the analyses. However, in the case of cultivations under oxygen-limiting and hypoxic conditions, the carbon balance showed a significant mismatch, suggesting that a compound was most probably missing. In addition, observation of the HPLC chromatograms of cleared supernatants from these cultivations (appendix 2.3) showed an

unidentified peak which increased as oxygen limitation was more severe. The compound corresponding to this peak was isolated and further analysed by LC-MS, resulting in the identification of a compound of a molecular mass of 197 which taking into account the effect of the reacting eluent (formic acid, 46  $\text{m}\cdot\text{z}^{-1}$ ) results in a metabolite of 151  $\text{g}\cdot\text{mol}^{-1}$ .

### **Identification of the unknown extracellular metabolite**

To identify the unknown extracellular compound, a hypoxic culture under steady state conditions was fed with fresh medium containing a carbon source mixture consisting of 12% (w/w) of uniformly  $^{13}\text{C}$ -labeled glucose and 88% (w/w) of unlabeled glucose, as described in the materials and methods section. Filtered samples of output culture media were submitted to an NMR analysis in order to identify the unknown compound. The NMR results (appendix 2.4) were compared with the  $^{13}\text{C}$ -NMR Japanese database spectra SBDS (Spectral Database for Organic Compounds, [http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre\\_index.cgi?lang=eng](http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre_index.cgi?lang=eng)). By combining this information with the molecular weight determined by LC-MS and its retention time in HPLC analysis (appendix 2.3), the unknown metabolite was identified as arabinitol (MW 152).

Arabinitol, a C5 sugar alcohol linked to the pentose phosphate pathway, has been previously shown to be produced by a variety of yeast species during fermentation. For instance, glycerol and arabinitol accumulation has been previously observed in *Pichia anomala* cultures during growth in high-salt environments and on highly concentrated sugar substrates [70]. Passoth and collaborators [71] suggested that arabinitol has the same physiological role as glycerol in the protection to osmotic stress, since they also found a glycerol and arabinitol accumulation under oxygen limiting conditions in *P. anomala*.

In yeasts (*S. cerevisiae*), glycerol is produced under oxygen limitation to reoxidize the redox equivalents produced during amino acid synthesis. Arabinitol has not yet been reported to be involved in the redox metabolism during fermentative growth; however, it is possible that arabinitol formation also plays a role in the redox balance of the cells [71].

Once the missing compound was identified it was quantified by HPLC. Including the arabinitol in the mass balance allowed to significantly improve the carbon balance and the percentage of carbon recovery reached values similar to the ones obtained for the cultivations under normoxic conditions. With this additional analysis, a reconciliation step including all available data was performed resulting in improved consistency index values (h values). Figure 2.5 shows the measured values, as well as the reconciled values for the involved metabolites in terms of  $\text{mmol}/(\text{gDW}\cdot\text{h})$ . As expected, it can be observed that the oxygen consumed per unit of dry weight decreased when oxygen limitation and hypoxic conditions were applied. Also, the amount of glucose consumed per unit of dry weight is

increased. This increased specific consumption of carbon source results mainly in the production of the above mentioned metabolites as a result of the imposed oxygen limitation conditions.

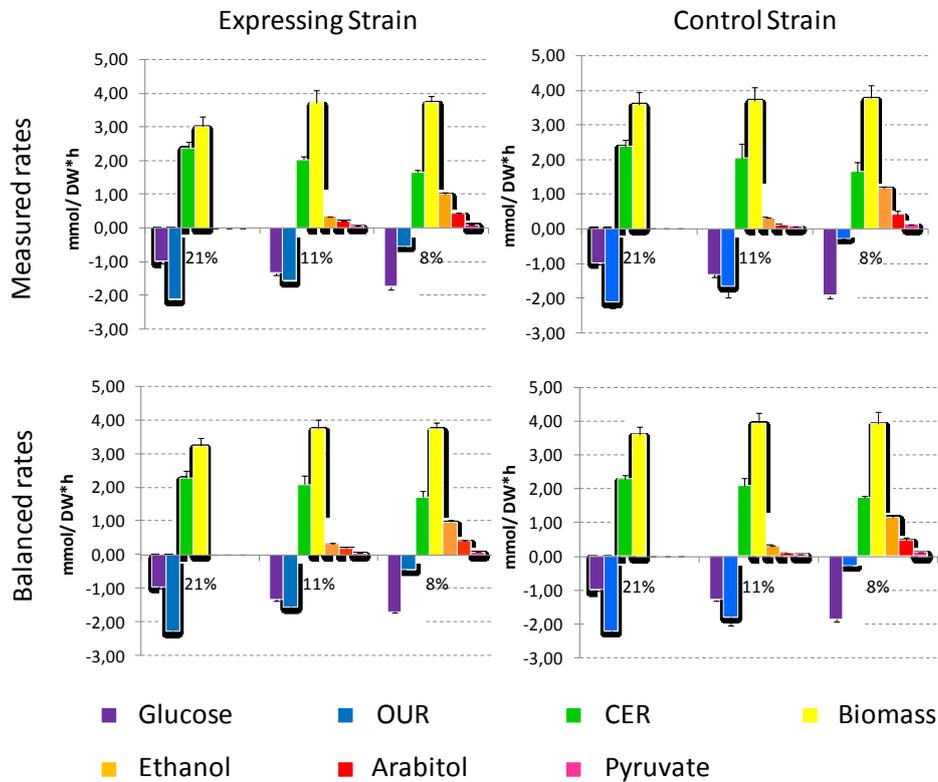


Figure 2.5: **Substrates consumed and metabolites generated for the different experimental conditions.** Top row: values measured. Left values for the Fab expressing strain. Right values for the control strain. Bottom row reconciled values. Negative values represent substrate consumed. Positive values represent products generated.

## Conclusions

In this study, different experimental methods were applied in order to obtain first hand data on the biomass composition of *P. pastoris*. The diversity of the data obtained was combined, using *ad hoc* data reconciliation techniques, and the best estimation of the biomass composition was obtained.

Application of the elemental mass balances to the input-output metabolite data allowed to detect the lacking of one major metabolite. Using complementary NMR data the missing metabolite was identified as being arabinol. Quantification of this metabolite and its incorporation into the metabolite mass balances allowed to validate the obtained data and to calculate reconciled and consistent fermentation data. The obtained results will be used in metabolic flux analyses of *P. pastoris*, which are being carried out in parallel with transcriptomics and proteomics analyses.

## Appendix

**Appendix 2.1:** Major chemical elements measured in de biomass at the different experimental conditions (%O<sub>2</sub> in the inlet air: 21, 11, 8). Data given as percentage of dry weight  $\pm$  standard deviation.

	Expressing strain			Control strain		
	21%	10%	8%	21%	10%	8%
C	43.7 $\pm$ 2.2	44.4 $\pm$ 2.2	44.3 $\pm$ 2.2	42.7 $\pm$ 2.1	43.6 $\pm$ 2.2	44.4 $\pm$ 2.2
H	6.8 $\pm$ 0.3	6.8 $\pm$ 0.3	6.8 $\pm$ 0.3	6.7 $\pm$ 0.3	6.8 $\pm$ 0.3	7.0 $\pm$ 0.3
N	7.1 $\pm$ 1.1	6.9 $\pm$ 1.0	7.5 $\pm$ 1.1	7.2 $\pm$ 1.1	7.1 $\pm$ 1.1	7.0 $\pm$ 1.0
O	35.0 $\pm$ 1.8	34.6 $\pm$ 1.7	34.2 $\pm$ 1.7	36.1 $\pm$ 1.8	35.1 $\pm$ 1.8	34.3 $\pm$ 1.7
S	0.2 $\pm$ 0.1	0.2 $\pm$ 0.1	0.1 $\pm$ 0.1	0.2 $\pm$ 0.1	0.2 $\pm$ 0.1	0.2 $\pm$ 0.1
Ashes	7.2 $\pm$ 0.4	7.2 $\pm$ 0.4	7.2 $\pm$ 0.4	7.2 $\pm$ 0.4	7.2 $\pm$ 0.4	7.2 $\pm$ 0.4

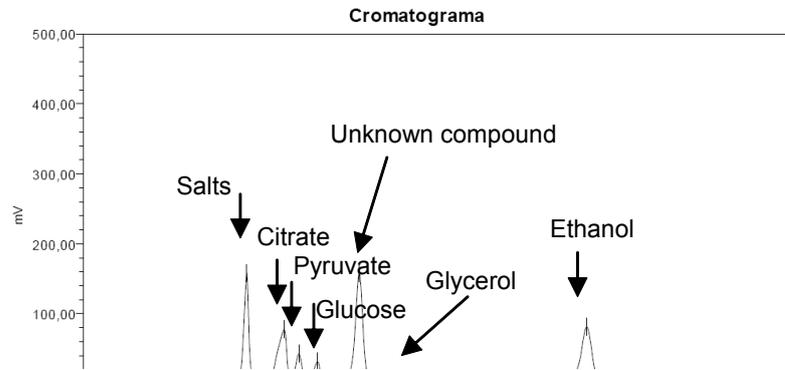
**Appendix 2.2:** Calculated C-mol formulas of the molecular biomass components at the different experimental conditions (%O<sub>2</sub> in the inlet air: 21, 11, 8). \* Data from [55] for *S.cerevisiae*.

<i>Saccharomyces cerevisiae</i> (*)							21% Oxygen Setpoint												
Prot.	C	H	1.581	N	0.275	O	0.318	S	0.003	Prot.	C	H	1.503	N	0.259	O	0.347	S	0.002
Carbo.	C	H	1.667	O	0.833					Carbo.	C	H	1.667	O	0.833				
Lipids	C	H	1.873	N	0.01	O	0.149	P	0.01	Lipids	C	H	1.697	N	0	O	0.156	P	0.004
RNA	C	H	1.232	N	0.389	O	0.737	P	0.105	RNA	C	H	1.230	N	0.4	O	0.734	P	0.105
DNA	C	H	1.255	N	0.378	O	0.312	P	0.102	DNA	C	H	1.232	N	0.39	O	0.632	P	0.105

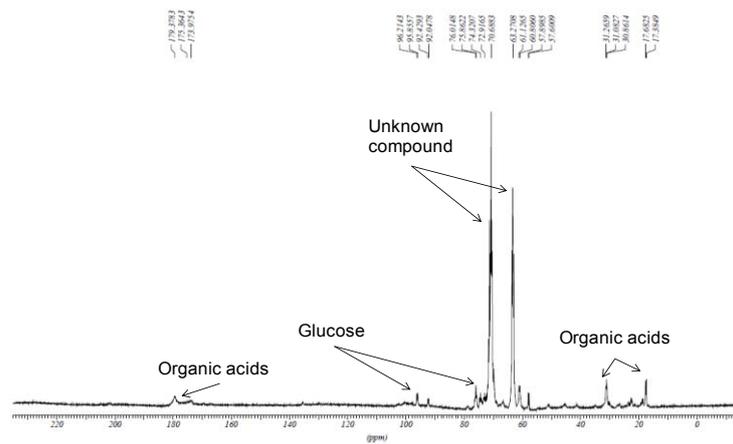
  

11% Oxygen Setpoint							8% Oxygen Setpoint												
Prot.	C	H	1.512	N	0.256	O	0.344	S	0.002	Prot.	C	H	1.511	N	0.255	O	0.34	S	0.002
Carbo.	C	H	1.667	O	0.833					Carbo.	C	H	1.667	O	0.833				
Lipids	C	H	1.746	N	0	O	0.145	P	0.003	Lipids	C	H	1.780	N	0	O	0.144	P	0.002
RNA	C	H	1.230	N	0.4	O	0.734	P	0.105	RNA	C	H	1.230	N	0.4	O	0.734	P	0.105
DNA	C	H	1.232	N	0.39	O	0.632	P	0.105	DNA	C	H	1.232	N	0.39	O	0.632	P	0.105

**Appendix 2.3:** Chromatogram with the unknown peak at retention time of 11.5 minutes in a HPLC Hewlett Packard 1050 Biorad Aminex HPX-87C ion-exchange resin column



**Appendix 2.4:**  $^{13}\text{C}$  NMR spectrum of a  $\text{D}_2\text{O}$ -resuspended lyophilised hypoxic culture broth sample in a Bruker Avance 500 MHz spectrometer.





## Chapter 3: *Pichia pastoris* <sup>13</sup>C-Based Metabolic Flux Analysis<sup>b</sup>

---

<sup>b</sup> Partly Published as: Baumann K, Carnicer M, Dragosits M, et al. **A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions.** *BMC Systems Biology* 2010, **4**:141-163.



## Background

In recombinant protein production processes, the secretion and correct folding of proteins, particularly protein complexes, may be a potential physiological bottleneck as their overexpression may result in stress conditions [46]; so far, studies have mainly concentrated on maximizing productivity, whereas the aspect of proper protein folding and secretion has stayed largely unattended. Furthermore, environmental and metabolic stresses are known to have a major impact on those processes. Thereby, since oxygen transfer rate is one of the most important parameters when we are developing a high cell density culture in recombinant protein production processes, it is important to study the different behaviors of the cells, *P. pastoris* in our case, growing under different oxygen levels.

In this chapter, *Pichia pastoris* producing a recombinant antibody fragment under different oxygenation conditions were grown with the aim to investigate the potential interrelations between energy and redox metabolism and recombinant protein production, as well as the impact of oxygen levels on such cellular processes using a Metabolic Flux Analysis (MFA) methodology based on <sup>13</sup>C-labelling data [13].

## Materials and Methods

### ***Yeast strains***

The *Pichia pastoris* strain X-33 pGAPZαA Fab3H6, secreting a the light and heavy chain chains of a human monoclonal antibody Fab fragment under the constitutive GAP promoter and the *S. cerevisiae* alpha-mating factor leader, and its empty-vector control strain.

### ***Chemostat cultivations and medium composition***

The pre-inoculation procedure was performed equal independently of the culture aim. Briefly, two litre shake flask containing 300 ml of YPD medium (10 g/l yeast extract, 20 g/l peptone, 10 g/l glucose) were inoculated with a 1.0 ml cryostock *P. pastoris* cells previously performed. The pre-inoculation cultures were grown for approximately 24 h at 30 °C with shaking at 200 rpm, and used to inoculate the different bioreactors.

The chemostat cultivations were performed in a 2-litre bench-top bioreactor (B. Braun Biotech International, BIOSTAT B) with a working volume of 1 litre. Cells were grown in a batch mode until about 24 hours, when chemostat mode was started by adding fresh medium at a dilution rate of 0.1 h<sup>-1</sup> using a peristaltic pump (Ismatec, IPC). The total gas flow inlet was controlled by Mass flow meters (Brankhorst High-Tech) at 1.5 vvm and the gas flow outlet was analysed by the CO<sub>2</sub>/O<sub>2</sub>-Sensors (BCP-CO<sub>2</sub> & BCP-O<sub>2</sub>, BlueSens) after it had been passed through a air drier (Festo LF-D-MINI) to remove all the humidity. For the cultures that received less than 21% of O<sub>2</sub> in the gas stream (oxygen-limited and hypoxic conditions), air was replaced with the equivalent volume of N<sub>2</sub> to maintain constant the 1.5

vvm. The cultures were kept at 1.2 bars of pressure using a pressure valve (GO Inc) and the pH, stirring and temperature were maintained at 5 (with 20% NH<sub>3</sub>), 700 and 25°C, respectively. During the cultures the data acquisition and the control of the different variables was done using software developed by Oriol Cos, (Department of Chemical Engineering, UAB). Samples were taken for each physiological equilibrium condition after 5 residence times, with the exception of the hypoxic set point, where a wash-out of the culture was observed after 3.5 residence times and, therefore, samples were taken just after 3 residence times.

The batch medium composition was: 1.80 g/l Citric acid, 40.00 g/l Glycerol, 12.60 g/l (NH<sub>4</sub>)<sub>2</sub>HPO<sub>4</sub>, 0.50 g/l MgSO<sub>4</sub>·7 H<sub>2</sub>O, 0.90 g/l KCl and 0.02 g/l CaCl<sub>2</sub>·2 H<sub>2</sub>O which were dissolved in deionized water to be sterilised by autoclaving together with the bioreactor. After cooled down, 4.60 ml/l trace salt stock solution and 2.00 ml/l Biotin (0.2 g/l, Sigma) were filtered through a 0.2 µm filter and putted inside the bioreactor.

The chemostat medium composition was: 0.92 g/l Citric acid, 50.0 g/l Glucose, 4.35 g/l (NH<sub>4</sub>)<sub>2</sub>HPO<sub>4</sub>, 0.65 g/l MgSO<sub>4</sub>·7 H<sub>2</sub>O, 1.70 g/l KCl, 0.010 g/l CaCl<sub>2</sub>·2 H<sub>2</sub>O, 1.60 ml/l trace salt solution and 1.00 ml/l Biotin (0.2 g/l) which were dissolved in deionized water. The pH was kept at 5.0 with 25% HCl before being filtered through a deep pore filter (Millipore Opticap 4''; 0.22 µm Hydrophilic Durapore) using a peristaltic pump. To prevent foaming, 0.1 ml/l (Glanapon 2000 Konz.) was added in the Chemostat Medium.

The Trace Salt solution composition was: 5.00 ml/l H<sub>2</sub>SO<sub>4</sub> (95-98%), 65.00 g/l FeSO<sub>4</sub>·7 H<sub>2</sub>O, 20.00 g/l ZnCl<sub>2</sub>, 6.00 g/l CuSO<sub>4</sub>·5 H<sub>2</sub>O, 3.36 g/l MnSO<sub>4</sub>·H<sub>2</sub>O, 0.82 g/l CoCl<sub>2</sub>·6 H<sub>2</sub>O, 0.20 g/l Na<sub>2</sub>MoO<sub>4</sub>·2 H<sub>2</sub>O, 0.08 g/l NaI and 0.02 g/l H<sub>3</sub>BO<sub>3</sub> which were dissolved in deionized water and sterilised by filtering through a 0.22 µm filter (Sterivex-GP, Millipore SVGPB1010).

The <sup>13</sup>C-labelled chemostat medium had exactly the same composition as the unlabelled medium except that it contained 12% (w/w) uniformly <sup>13</sup>C-labeled Glucose (Cortecnet, Paris) and 88% (w/w) unlabeled Glucose.

The macroscopic metabolite rates of two chemostat cultures performed by Kristin Baumann using the same fermentation setup [52] were taken plus the <sup>13</sup>C labelled culture performed in order to have biological replicates and perform statistical analysis.

### **Analytical procedures**

For cellular dry weight, a known volume of cultivation broth was filtered using pre-weighted filters. These were washed with two volumes of distilled water and dried to constant weight at 105 °C for 24 h.

Samples for extracellular metabolite analyses were centrifuged at 4.000 g for 2 min in a microcentrifuge to remove the cells and subsequently filtered through 0.45 µm-filters (Millipore type HAWP). Glucose,

ethanol and arabinitol were analyzed by HPLC (Series 1050, Hewlett Packard) with an ionic exchange column (Bio-Rad, Aminex HPX-87H). As mobile phase, 15 mM sulphuric acid was used. The metabolites were detected (Detector HP 1047A, Hewlett Packard) and quantified with the Software EmpowerProfor.

### ***Biosynthetically directed fractional (BDF) <sup>13</sup>C-labeling for <sup>13</sup>C-MFA***

*P. pastoris* cells were continuously fed with a non labeled chemostat medium for five residence times until reaching a metabolic steady state, as indicated by a constant cell density and constant oxygen and carbon dioxide concentrations in the bioreactor exhaust gas. Biosynthetically directed fractional <sup>13</sup>C labelling (BDF) of cells growing at steady state on a single carbon source has been described elsewhere [60, 72, 73]. After reaching the steady state, 12% (w/w) of the carbon source in the medium was replaced with uniformly <sup>13</sup>C-labeled substrate (<sup>13</sup>C-labeled glucose, isotopic enrichment of >98%, from Cortecnet, Voisins le Bretonneux, France). After one residence time, labeled cells were harvested by centrifugation at 4000 g for 10 min, resuspended in 20 mM Tris-HCl (pH 7.6) and centrifuged again. Finally, the washed cell pellets were lyophilized (Benchtop 5L Virtis Sentry). Afterwards, an amount of 100 mg of the lyophilized biomass was resuspended in 6 ml of 6 M HCl and subsequently hydrolyzed in sealed glass tubes at 110 °C for 21 h. The resulting suspensions were filtered using 0.2 µm-filters (Millex-GP, Millipore) and lyophilized. The lyophilized hydrolysates were dissolved in D<sub>2</sub>O for NMR experiments being the pH of the samples below 1 due to residual HCl.

### ***NMR spectroscopy and metabolic flux ratio (METAFor) analysis***

2D [<sup>13</sup>C,<sup>1</sup>H]-COSY spectra were acquired for both aliphatic and aromatic resonances as described [74] at 40 °C on a Varian Inova spectrometer operating at a <sup>1</sup>H resonance frequency of 600 MHz. The spectra were processed using the standard Varian spectrometer software VNMR (version 6.1, C). The program FCAL [28] was used for the integration of <sup>13</sup>C-<sup>13</sup>C scalar fine structures in 2D [<sup>13</sup>C,<sup>1</sup>H]-COSY, for the calculation of relative abundances, *f*-values (see Appendix 3.1), of intact carbon fragments arising from a single carbon source molecule [74]. Afterwards, the the flux ratios through several key pathways in central metabolism were calculated as described by Maaheimo [26] and Jouhten [27].

As described previously [26, 28, 60, 72–74], the calculation of metabolic flux ratios when using fractional <sup>13</sup>C-labeling of amino acids is based on assuming both a metabolic and an isotopomeric steady state. To establish a cost-effective protocol for a larger number of <sup>13</sup>C labelling experiments, we fed a chemostat operating in metabolic steady state for the duration of one volume change with the medium containing the <sup>13</sup>C-labeled before harvesting the biomass [72, 73]. Then, the fraction of unlabeled biomass produced prior to the start of the supply with <sup>13</sup>C-labeled medium can be calculated following simple wash-out kinetics ([75], see also [60] for additional discussion).

**Minimal metabolic model used for *P. pastoris***

In order to perform the metabolic flux analysis of *P. pastoris* a minimal metabolic model was build including the principal catabolic and anabolic pathways of cell. This model, as it represents the principal metabolic pathways of the cell, it will allow deeper understanding of the cell behavior and gaining knowledge of the recombinant protein production impact on the cell metabolism. The model was divided between the principal metabolic pathways for the better understanding of the reader. The ATP, ADP, NAD(P)H and NAD(P)<sup>+</sup> were not balanced in the calculations.

Name reactionsStoichiometry**Glycolysis**

HXT + HXK	$\text{Glc}_{\text{ext}} + 2 \text{ ATP} \Rightarrow \text{G6P} + 2 \text{ ADP}$
PGI	$\text{G6P} \Leftrightarrow \text{F6P}$
PFK + FBA	$\text{F6P} + \text{ ATP} \Rightarrow \text{GAP} + \text{ DHAP} + \text{ ADP}$
TPI	$\text{DHAP} \Leftrightarrow \text{GAP}$
GAPDH + PGK + GPM + ENO + PYK	$\text{GAP} + 2 \text{ ADP} + \text{ NAD}^+ \Rightarrow \text{Pyr} + 2 \text{ ATP} + \text{ NADH}$

**Pentose Phosphate Pathway**

G6PDH + 6PGDH	$\text{G6P} + 2 \text{ NADP}^+ \Rightarrow \text{RU5P} + 2 \text{ NADPH} + \text{ CO}_2$
RPE	$\text{RU5P} \Leftrightarrow \text{XU5P}$
RPI	$\text{RU5P} \Leftrightarrow \text{R5P}$
TK(1)	$\text{R5P} + \text{ X5P} \Leftrightarrow \text{ S7P} + \text{ GAP}$
TK(2)	$\text{E4P} + \text{ X5P} \Leftrightarrow \text{ F6P} + \text{ GAP}$
TA	$\text{S7P} + \text{ GAP} \Leftrightarrow \text{ F6P} + \text{ E4P}$

**Pyruvate Dehydrogenase**

PDH	$\text{CoA} + \text{ NAD}^+ + \text{ Pyr} \Rightarrow \text{ AcCoA}_{\text{mit}} + \text{ NADH} + \text{ CO}_2$
-----	---

**TCA cycle**

CIT Syn.	$\text{AcCoA}_{\text{mit}} + \text{ Oaa}_{\text{mit}} \Rightarrow \text{ CoA} + \text{ Cit}_{\text{mit}}$
ACO + ISODH	$\text{Cit}_{\text{mit}} + \text{ NAD}^+ \Rightarrow \text{ NADH} + \text{ CO}_2 + \text{ aKG}_{\text{mit}}$
aKGDH + SUCDH + FHM	$2 \text{ NAD}^+ + \text{ aKG}_{\text{mit}} + \text{ ADP} \Rightarrow \text{ Mal} + 2 \text{ NADH} + \text{ ATP} + \text{ CO}_2$
MALDH	$\text{Mal} + \text{ NAD}^+ \Rightarrow \text{ OAA}_{\text{mit}} + \text{ NADH}$

**Anaplerotic Pathways**

PyrCK	$\text{ATP} + \text{ CO}_2 + \text{ Pyr} \Rightarrow \text{ OAA}_{\text{cyt}} + \text{ ADP}$ (O <sub>2</sub> -limited and hypoxic <sup>13</sup> C-MFA)
-------	--

**Fermentative Pathways**

PDC	$\text{Pyr} \Rightarrow \text{AcO} + \text{CO}_2$
ADDH + ACETCoA	$\text{AcO} + \text{NADP}^+ + \text{CoA} + \text{ATP} \Rightarrow \text{NADPH} + \text{AcCoA}_{\text{cyt}} + \text{Pyrophosphate} + \text{AMP}$
ADH + ET	$\text{AcO} + \text{NADH} \Rightarrow \text{Eth}_{\text{ext}} + \text{NAD}^+$
G3PDH + GT	$\text{DHAP} + \text{NADH} + \text{ATP} \Rightarrow \text{Glycerol}_{\text{cyt}} + \text{P}_i + \text{ADP} + \text{NAD}^+$
Ara Syn.	$\text{X5P} + \text{NADH} \Rightarrow \text{ARA}_{\text{ext}} + \text{NAD}^+$

**Transport Reactions**

PyrCK_OAA_G	$\text{Pyr} + \text{ATP} + \text{CO}_2 \Rightarrow \text{Oaa}_{\text{mit}} + \text{ADP}$ (normoxia <sup>13</sup> C-MFA)
OAA_Transp_cyt-mit	$\text{Oaa}_{\text{cyt}} \Rightarrow \text{Oaa}_{\text{mit}}$ ( $\text{O}_2$ -limited and hypoxic <sup>13</sup> C-MFA)
OAA_Transp_mit-cyt	$\text{Oaa}_{\text{mit}} \Rightarrow \text{Oaa}_{\text{cyt}}$ ( $\text{O}_2$ -limited and hypoxic <sup>13</sup> C-MFA)
aKG_Transp_mit-cyt	$\text{aKG}_{\text{mit}} \Rightarrow \text{aKG}_{\text{cyt}}$
Glycerol Syn	$\text{Glycerol}_{\text{cyt}} \Rightarrow \text{Glycerol}_{\text{ext}}$
Pyr Syn	$\text{Pyr}_{\text{cyt}} \Rightarrow \text{Pyr}_{\text{mit}}$

**Biomass synthesis reactions for each oxygenation condition**

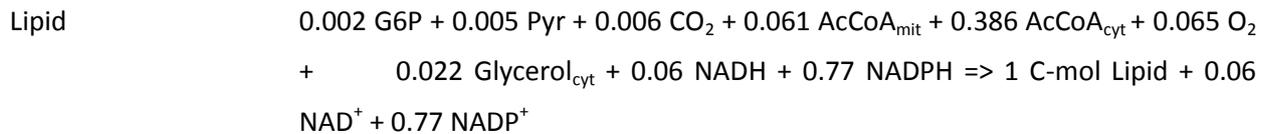
1. Protein (Composition derived from the amino acid composition of each oxygenation condition [76]; Chapter 2).

Prot_N	(Normoxic condition): $0.136 \text{ Pyr} + 0.006 \text{ R5P} + 0.013 \text{ E4P} + 0.031 \text{ Oaa}_{\text{cyt}} + 0.014 \text{ AcCoA}_{\text{mit}} + 0.009 \text{ Oaa}_{\text{mit}} + 0.075 \text{ Akg}_{\text{mit}} + 0.013 \text{ AcCoA}_{\text{cyt}} \Rightarrow 1 \text{ C-mol Protein} + 0.002 \text{ GAP} + 0.058 \text{ CO}_2$
Prot_L	(Oxygen limited conditions): $0.141 \text{ Pyr} + 0.006 \text{ R5P} + 0.013 \text{ E4P} + 0.033 \text{ Oaa}_{\text{cyt}} + 0.015 \text{ AcCoA}_{\text{mit}} + 0.009 \text{ Oaa}_{\text{mit}} + 0.070 \text{ Akg}_{\text{mit}} + 0.014 \text{ AcCoA}_{\text{cyt}} \Rightarrow 1 \text{ C-mol Protein} + 0.002 \text{ GAP} + 0.066 \text{ CO}_2$
Prot_H	(Hypoxic condition): $0.144 \text{ Pyr} + 0.006 \text{ R5P} + 0.014 \text{ E4P} + 0.033 \text{ Oaa}_{\text{cyt}} + 0.016 \text{ AcCoA}_{\text{mit}} + 0.010 \text{ Oaa}_{\text{mit}} + 0.068 \text{ Akg}_{\text{mit}} + 0.014 \text{ AcCoA}_{\text{cyt}} \Rightarrow 1 \text{ C-mol Protein} + 0.002 \text{ GAP} + 0.068 \text{ CO}_2$

2. Carbohydrate, Glycogen and Trehalose.

Carbhy.	$0.166 \text{ G6P} \Rightarrow 1 \text{ C-mol Carbohydrate}$
Glycogen	$0.166 \text{ G6P} \Rightarrow 1 \text{ C-mol Glycogen}$

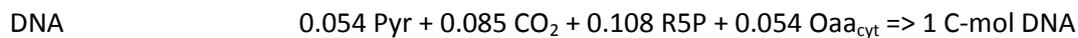
3. Lipid (Composition derived from the mean lipid composition from [76];Chapter 2).



4. RNA (Composition derived from the RNA composition proposed by Stephanopoulos and co-workers [77]).



5. DNA (Composition derived from the DNA composition published in [63]).



This model was firstly constructed based in previous studies on this microorganism using labeling carbon sources [60, 78], the stoichiometric model of central carbon metabolism formulated for *S. cerevisiae* [26, 27] and its genome-scale metabolic network [50]. However, as more specific data from *P. pastoris* was available like the specific biomass C-mol formula under different oxygen conditions [76] (Chapter 2) or from the two genome-scale metabolic network [79, 80], they were included.

### <sup>13</sup>C-Metabolic Flux Analysis

In *P. pastoris*, the metabolic pathways could be distributed in a specific cell compartment, as for glycolysis in the cytosol, or in two compartments at the same time, as for TCA present in the mitochondria and in the peroxisome. Therefore, in the model some central carbon metabolites are described to be located in different compartments at the same time (OAA, aKG and AcCoA). Moreover, the transports of OAA across the mitochondrial membrane was included but the transport of AcCoA, the final step of the cytosolic Pyruvate dehydrogenase (PDH) bypass, was omitted. The potential carbon flux through the PDH bypass was lumped into the flux through the PDH reaction, since the <sup>13</sup>C-labeling protocol used does not allow for an assessment of the split flux ratio between these two pathways. That is, given that flux through malic enzyme is essentially zero in our model (see results section), labeled Pyr being metabolized through the PDH bypass does not produce labeling patterns in mitochondrial AcCoA that are distinct from those generated when Pyr is channeled through the PDH reaction. Nevertheless, flux through the PDH bypass cannot be totally excluded, as discussed in the results section.

Moreover, the glyoxylate cycle, the PEP carboxykinase and the malic enzyme activity were omitted from the stoichiometric model since the METAFoR data showed that those pathways were either inactive or at basal levels (see [27] for details on the identification of these activities).

### <sup>13</sup>C-NMR metabolic ratios computation

The determined metabolic flux ratios were used as additional constraints for solving the metabolic network following a <sup>13</sup>C constrained flux balancing approach similarly to a previous approach [27, 75]. The flux ratios considered in the present approach were the following (equations 3.1 to 3.4):

The fraction of  $Oaa_{mit}$  originating from  $Oaa_{cyt}$ , that is,  $Oaa_{cyt}$  transport into the mitochondria (only applicable to  $O_2$ -limiting and hypoxic <sup>13</sup>C-MFA; see results section):

$$a = \frac{OAA\_transp\_cyt - mit}{OAA\_transp\_cyt - mit + MALDH} \quad (3.1)$$

The fraction of  $Oaa_{cyt}$  originating from  $Pyr_{cyt}$ , that is, the anaplerotic flux ratio (only applicable to  $O_2$ -limiting and hypoxic <sup>13</sup>C-MFA; see results section):

$$b = \frac{PyrCK}{PyrCK + OAA\_Transp\_mit - cyt} \quad (3.2)$$

Under normoxic conditions, OAA transport across the mitochondrial was no available from the same metabolic flux ratios as from the oxygen-limited and hypoxic conditions (See results section). However, a minimum OAA transport was available looking the  $Oaa_{mit}$  originated from PEP which was used for describing the anaplerotic pathway activity under normoxia. In order to use this ratio, the OAA transport under this condition was simplified pooling the  $PyrCK$  reaction with the OAA global (import-export) transport through the mitochondria ( $PyrCK\_OAA\_G$ ). Therefore, the corresponding anaplerotic flux ratio was defined as:

$$b = \frac{PyrCK\_OAA\_G}{PyrCK\_OAA\_G + MALDH} \quad (3.3)$$

As said before, this ratio is an expression which describe the minimum OAA transport under normoxic conditions. Comparing the three <sup>13</sup>C-MFA (normoxia, oxygen-limited and hypoxia),  $Pyr$  node on the normoxia was much less constrained due to the lack of <sup>13</sup>C-NMR ratios, therefore, the  $CO_2$  stoichiometry were used as further constrains in the <sup>13</sup>C-MFA for this condition.

The fraction of  $Pep$  from  $PPP$  assuming a maximal contribution of  $PPP$  could be computed differently depending on which reaction you take in to account. Equation 3.4 and 3.5 represent the ratios used for Jouhten and co-workers [27] and for Sauer and co-workers [73] respectively:

$$d \geq \frac{TK(1) + 2*(TA) + 3*(TK(2))}{2*(PFK + FBA) + TK(1) + TA} \quad (3.4)$$

$$d \geq \frac{GAPDH - 2*PGI}{GAPDH} \quad (3.5)$$

Both of them were tried giving equivalent metabolic flux fingerprint independently of the reactions used (data not shown). Finally, to perform the <sup>13</sup>C-MFA the Equation 3.4 was used. The following linear constraint equations (equations 3.6 to 3.9) were derived from the previous flux ratio equations:

$$\text{OAA\_transp\_cyt-mit} * (1-a) + \text{MALDH}*(-a) = R_a \quad (3.6)$$

$$\text{PyrCK} * (1-b) + \text{OAA\_transp\_mit-cyt} * (-b) = R_b \quad (3.7)$$

$$\text{PyrCK\_OAA\_G} * (1-c) + \text{MALDH} * (-c) = R_c \quad (3.8)$$

$$\text{TK}(1) * (1-d) + \text{TA} * (2) + \text{TK}(2) * (3-d) + (\text{PFK} + \text{FBA}) * (-2) \leq R_d \quad (3.9)$$

The complete model for <sup>13</sup>C-MFA, comprised 32 (normoxic <sup>13</sup>C-MFA) and 33 (oxygen-limited and hypoxic <sup>13</sup>C-MFA) metabolic reactions. The measured uptake and secretion metabolite rates and the rates of metabolic precursor depletion to biosynthesis were combined with a set of linearly independent equations obtained from metabolic flux ratio (METAFor) analysis to render the complete linear system solvable. The oxygen and carbon dioxide rates were not included in the input data to solve the lineal system unless it is specified.

The complete metabolic models were solved using the metabolite mass balances:

$$\begin{bmatrix} S \\ F \end{bmatrix} \cdot x = \begin{bmatrix} c \\ 0 \end{bmatrix} \equiv N \cdot x = b \quad (3.9)$$

Where S represents the stoichiometric matrix (including input/output reactions), c is a column vector with either 0 for internal reactions or the corresponding value for each one of the input/output rates and x is the vector of fluxes. Equations 3.6 to 3.9 were added to the stoichiometric model as a submatrix F, obtaining the solution of the resulting linear system using the Matlab function *lsqlin*. Irreversibility was assumed for several intracellular fluxes and for the depletion of precursors to biosynthetic reactions, that is, only positive values were allowed for these fluxes (see table 5.1).

Confidence intervals for the optimized fluxes were calculated up on the determination of their standard deviation using the Fischer Information Matrix approach (FIM)[81] as:

$$\sigma_j = \sqrt{(FIM^{-1})_{jj}} \quad (3.10)$$

Calculation of FIM was performed as:

$$FIM = \sum W^T C^{-1} W \quad (3.11)$$

Where C is the variance-covariance matrix of the measurements (assumed independent) and W is a parameter sensitivity matrix where each element of  $w_{ij}$  corresponds to:

$$w_{ij} = \frac{\partial x_i}{\partial p_j} \quad (3.12)$$

Which describes an infinitesimal change of the variable  $x_i$  (e.g. a measurement) due to an infinitesimal change in parameter  $p_j$  (a flux).

Confidence intervals for the estimated fluxes  $\hat{p}_j$  of  $p_j$  can be derived from [81].

$$\hat{p}_j - \sigma_{p_j} \cdot t_{\alpha/2}^v < p_j < \hat{p}_j + \sigma_{p_j} \cdot t_{\alpha/2}^v \quad (3.13)$$

Where  $t_{\alpha/2}^v$  corresponds to the Student t distribution, with  $v$  degrees of freedom and  $\alpha$  corresponds to the  $(1-\alpha)$  confidence interval chosen. All calculations were performed on a PC compatible computer running Matlab<sup>®</sup> 7.4 (v2007b) for Windows.

## Results and Discussion

### ***Culture parameters***

The recombinant *P. pastoris* X-33 and a wild type strains were grown under different oxygen conditions in order to gain knowledge about the impact of the oxygen availability and the recombinant protein production on the central carbon metabolism. For all chemostat cultivations, the performed the statistical consistency test, at 95% confidence level, was passed indicating that there were no proof for gross measurement errors.

Briefly, the impact of reduced oxygen supply on the core metabolism was readily observed, both in the biomass yields and the profile of secreted by-products such as ethanol and arabinitol, reflecting the adaptation from a respiratory to a respiro-fermentative metabolism. Since all cultivations were carbon-limited, the decrease in the biomass yield decreased under oxygen limiting and hypoxic conditions resulted in an increase of specific glucose uptake rates under such conditions.

For all the normoxic conditions, the respiratory quotient (RQ) were close to unit as expected from a totally respiratory metabolism indicating high reproducibility of that environmental condition. In addition, looking the other culture parameters, no significant differences were found between the control and the expressing cells entailing equal normoxic metabolite state independently of the recombinant protein production.

For the oxygen-limited conditions, the RQ ratios were around the 1.2-1.3 indicating some oxygen constriction in all the conditions although the differences between the cultures were no significant

looking at this parameter. On the other hand, the arabinitol specific rate was higher in the expressing strain than the control strain suggesting some impact of the recombinant protein production.

Table 3.1: Overview of the macroscopic growth parameters.

	Control Strain			Expressing Strain		
	Normoxic	O <sub>2</sub> -limited	Hypoxic	Normoxic	O <sub>2</sub> -limited	Hypoxic
q <sub>Glc</sub> <sup>a</sup>	-0.99 ± 0.04	-1.27 ± 0.05	-1.85 ± 0.08	-0.95 ± 0.04	-1.31 ± 0.05	-1.69 ± 0.04
q <sub>O<sub>2</sub></sub> <sup>a</sup>	-2.20 ± 0.12	-1.79 ± 0.24	-0.28 ± 0.01	-2.26 ± 0.12	-1.54 ± 0.06	-0.44 ± 0.07
q <sub>CO<sub>2</sub></sub> <sup>a</sup>	2.27 ± 0.12	2.18 ± 0.24	1.71 ± 0.07	2.29 ± 0.12	2.09 ± 0.06	1.73 ± 0.06
q <sub>X</sub> <sup>a</sup>	3.61 ± 0.24	3.94 ± 0.30	3.92 ± 0.35	3.25 ± 0.22	3.76 ± 0.27	3.75 ± 0.18
q <sub>EtOH</sub> <sup>a</sup>		0.31 ± 0.02	1.16 ± 0.06		0.33 ± 0.02	0.98 ± 0.05
q <sub>Ara</sub> <sup>a</sup>		0.10 ± 0.00	0.48 ± 0.09		0.21 ± 0.01	0.42 ± 0.02
RQ <sup>b</sup>	1.03 ± 0.08	1.22 ± 0.21	6.14 ± 0.40	1.01 ± 0.08	1.36 ± 0.07	3.90 ± 0.62

<sup>a</sup> mmol / (g Biomass · h)

<sup>b</sup> mol CO<sub>2</sub> / mol O<sub>2</sub>

Nevertheless, the hypoxic conditions resulted in a completely different oxygen limitation environment as could be observed comparing the RQ ratios. The hypoxic cultures not only presented a stronger oxygen limitation (higher RQ ratios), but also the cultures could only be run for 3.5 residence times due to a bioreactor wash out.

Table 3.2: Metabolic flux ratio (METAFor) analysis results. Bold fractions are the fractions used as <sup>13</sup>C-constrains in the metabolic flux analysis. n.d= not determined. The values are the % pool fraction ± standard deviation.

	Expressing Strain			Control Strain		
	Normoxia	O <sub>2</sub> -Limited	Hypoxia	Normoxia	O <sub>2</sub> -Limited	Hypoxia
<b>Pep from pentose phosphates, upper bound</b>	<b>50 ± 9</b>	<b>23 ± 6</b>	<b>15 ± 7</b>	<b>39 ± 9</b>	<b>32 ± 8</b>	<b>15 ± 6</b>
R5P from T3P and S7P (transketolase)	71 ± 2	78 ± 2	70 ± 2	66 ± 2	70 ± 2	62 ± 2
R5P from E4P (transaldolase)	44 ± 2	24 ± 2	23 ± 2	40 ± 2	29 ± 2	24 ± 2
Ser originating from Gly and C1-unit	61 ± 4	68 ± 4	68 ± 4	62 ± 4	69 ± 4	72 ± 4
Gly originating from CO <sub>2</sub> and C1-unit	10 ± 4	13 ± 3	13 ± 3	6 ± 4	12 ± 3	10 ± 3
PEP originating from OAA-cyt (PEPck)	0 ± 4	0 ± 8	0 ± 10	2 ± 5	0 ± 10	5 ± 11
OAA-mit originating from PEP	<b>44 ± 2</b>	32 ± 2	44 ± 2	<b>42 ± 2</b>	35 ± 2	41 ± 2
<b>OAA-mit originating from OAA-cyt</b>	nd	<b>43 ± 3</b>	<b>55 ± 3</b>	nd	<b>44 ± 3</b>	<b>51 ± 3</b>
<b>OAA-cyt originating from PEP</b>	nd	<b>63 ± 3</b>	<b>64 ± 4</b>	nd	<b>66 ± 3</b>	<b>66 ± 4</b>
OAA-cyt reversibly converted to FUM	63 ± 11	7 ± 5	10 ± 5	63 ± 11	11 ± 4	9 ± 4
Flux through malic enzyme, upper bound	1 ± 4	nd	nd	1 ± 6	nd	nd
Flux through malic enzyme, lower bound	1 ± 2	nd	nd	0 ± 3	nd	nd

### METAFor analyses for <sup>13</sup>C-MFA

Biosynthetically directed fractional (BDF) <sup>13</sup>C-labeling of proteinogenic amino acids combined with 2D-NMR enabled the analysis of metabolic flux ratios (METAFor analysis). The metabolic flux ratios were calculated using the relative abundances (*f*-values) of intact carbon fragments arising from a single source molecule of glucose (Appendix 3.1). The calculated flux ratios are shown in Table 3.2. As expected, the *f*-values obtained for this series of cultivations confirm that the proteinogenic amino

acids are primarily synthesized in *P. pastoris* according to the pathways documented for *S. cerevisiae*, as previously reported [60].

In <sup>13</sup>C-based metabolic flux analyses (<sup>13</sup>C-MFA), the metabolic flux ratios determined by METAFoR were used as additional constraints for the stoichiometric equation system to be able to solve the metabolic flux distribution (see materials and methods section). Nevertheless, some statements could already be seen looking the ratios directly. For example, in fully aerobic conditions up to 50–39% of Pep was originated from the pentose phosphate pool meanwhile the fraction of Pep from PPP assuming a maximal contribution of PPP was clearly lower under hypoxic conditions, only about 15 %.

On the other hand, the arabinitol production had a clear impact on the flux ratios involved in the PPP. The fraction of pentose phosphates showing reversible action of a transketolase reaction was generally high (> 60 % in all cultivations), with no clear trend, whereas the fraction of pentose phosphates showing the reversible action of a transaldolase clearly decreased at lower oxygen availability.

In addition, as previously observed for *P. pastoris* [60] and other yeasts (e.g. *S. cerevisiae* and *P. stipitis*, [27, 72]), cells growing aerobically in glucose-limited chemostats show a bidirectional transport of OAA and/or other TCA cycle intermediates, across the mitochondrial membrane. However, calculation of flux ratios defining the fraction of  $Oaa_{mit}$  from  $OAA_{cyt}$  and,  $OAA_{cyt}$  from Pep under normoxic conditions was not possible. The labeling patterns of cytosolic and mitochondrial OAA pools are accessible through the observation of Asp labeling patterns (shown to be synthesized from  $OAA_{cyt}$  in yeast in previous studies [26] and glutamate (synthesized from mitochondrial  $\alpha$ KG, and therefore accessing to  $OAA_{mit}$  labeling patterns). Strikingly, the fraction of intact C2-C3 bonds of OAA, which is often used to calculate these flux ratios, were approximately equal for  $OAA_{cyt}$  and  $OAA_{mit}$  (that is, the fraction of intact C $\alpha$ -C $\beta$  bonds in Asp/Thr and Glu) were equal (as revealed by the *f*-values of Asp, Thr and Glu, Appendix 3.1). This could be explained by an extremely fast exchange between TCA cycle intermediates cytosolic and mitochondrial pools (near equilibrium), resulting in identical labeling patterns in the amino acids synthesized from such pools. However, this possibility can be excluded, as the Asp-C $\beta$ , Thr-C $\beta$  and Glu-C $\alpha$  labeling patterns under normoxic conditions were not identical (Appendix 3.1). Notably, the reversibility of the inter-conversion of cytosolic OAA to other cytosolic TCA cycle intermediates (defined here as the  $OAA_{cyt}$  interconversion to fumarate ratio) was clearly higher under normoxic conditions. This results suggested that Asp, OAA, and malate might be participating in a redox shuttle (e.g. malate-Asp and/or malate-OAA shuttles) for translocation of NADH across the mitochondrial membrane, as described by Bakker and co-workers [82].

It is worth noting that, although flux through the PDH bypass was not considered in our metabolic model (see Materials and Methods section), its activity should not be totally excluded since Crabtree negative yeasts are reported to have activity on this pathway [83]. In contrast to *S. cerevisiae*, a microorganism that does not synthesize carnitine *de novo* (essential for carnitine acetyltransferase-

mediated transport of cytosolic AcCoA to the mitochondria [84]), a complete carnitine biosynthesis pathway has been characterised in *Candida albicans*, and the corresponding 4 genes have been identified [85]. Interestingly, the *P. pastoris* genome contains putative homologues to these genes [64]. Moreover, it should be mentioned that a 20-fold change in relative expression levels of the *S. cerevisiae* *ACS1* homolog encoding the mitochondrial AcCoA synthetase essential for the contribution of mitochondrial PDH bypass to the formation of mitochondrial AcCoA, was observed when comparing normoxic vs. hypoxic conditions [52]. Also, significant changes in expression levels in *CAT2* and *YAN2* homologues involved in carnitine transport to mitochondria were observed [52].

### ***P. pastoris* metabolic flux profile**

In Figure 3.1 the net fluxes for the expressing and control strains growing under different oxygenation conditions are shown. These fluxes were scale to the glucose uptake rate (Table 3.1) in each oxygen conditions to facilitate the comparison. For absolute net fluxes in mmol/(gDCW·h) see Appendix 3.2. In addition, in Figure 3.2 there are represented different flux distributions of some metabolic nodes under the different environmental conditions.

Overall, the most prominent feature, as already indicated by the METAFoR analysis (Table 3.2), was the similarity in flux estimates between corresponding Fab-expressing and control strains datasets. Nevertheless, as expected, clear differences were observed when comparing flux patterns corresponding to different oxygenation set points. In general terms, the metabolic adaptation from oxidative towards respiro-fermentative growth was accompanied by complex changes of carbon flux throughout the whole central carbon metabolism, as previously described in other yeasts (*S. cerevisiae* [27, 86], *P. anomala* [87, 88]).

Specifically, as already inferred from the macroscopic data, flux through some fermentative pathways was increased when upon adaptation from respirative to respiro-fermentative metabolism, particularly the fluxes towards the formation of ethanol and arabinitol. Production of arabinitol had a clear impact on the flux ratios and on the distribution of fluxes through the PPP as already seen in the METAFoR results. Besides, Figure 3.2 showed that under normoxic conditions there was an important net contribution of the PPP to glucose catabolism. In contrast, as oxygen availability was decreased and, particularly, when arabinitol was produced, this contribution was clearly reduced.

The carbon flux distributions at the pyruvate branching point (Figure 3.2) clearly show the shift from respiratory to respiro-fermentative metabolism: fluxes through the pyruvate dehydrogenase pathway decreased when decreasing oxygen availability, whereas flux through the pyruvate decarboxylase pathway increased, reflecting the production of ethanol. Also, anaplerotic flux through the pyruvate carboxylase pathway drastically decreased when oxygen availability was reduced. Remarkably, although variations in the carbon flux distribution around the pyruvate branch and TCA cycle activity were

observed, mitochondrial transporters such as *DIC1*, *OAC1*, *SFC1* showed no significant change in transcriptomic analysis [52]. Only *YIA6*, which is involved in NAD<sup>+</sup> transport into the mitochondria (and has a disputed role as Pyr transporter), was down regulated under hypoxic conditions. *CAT2* and *YAT2* (carnitine transporters) were also down regulated under such conditions [52].

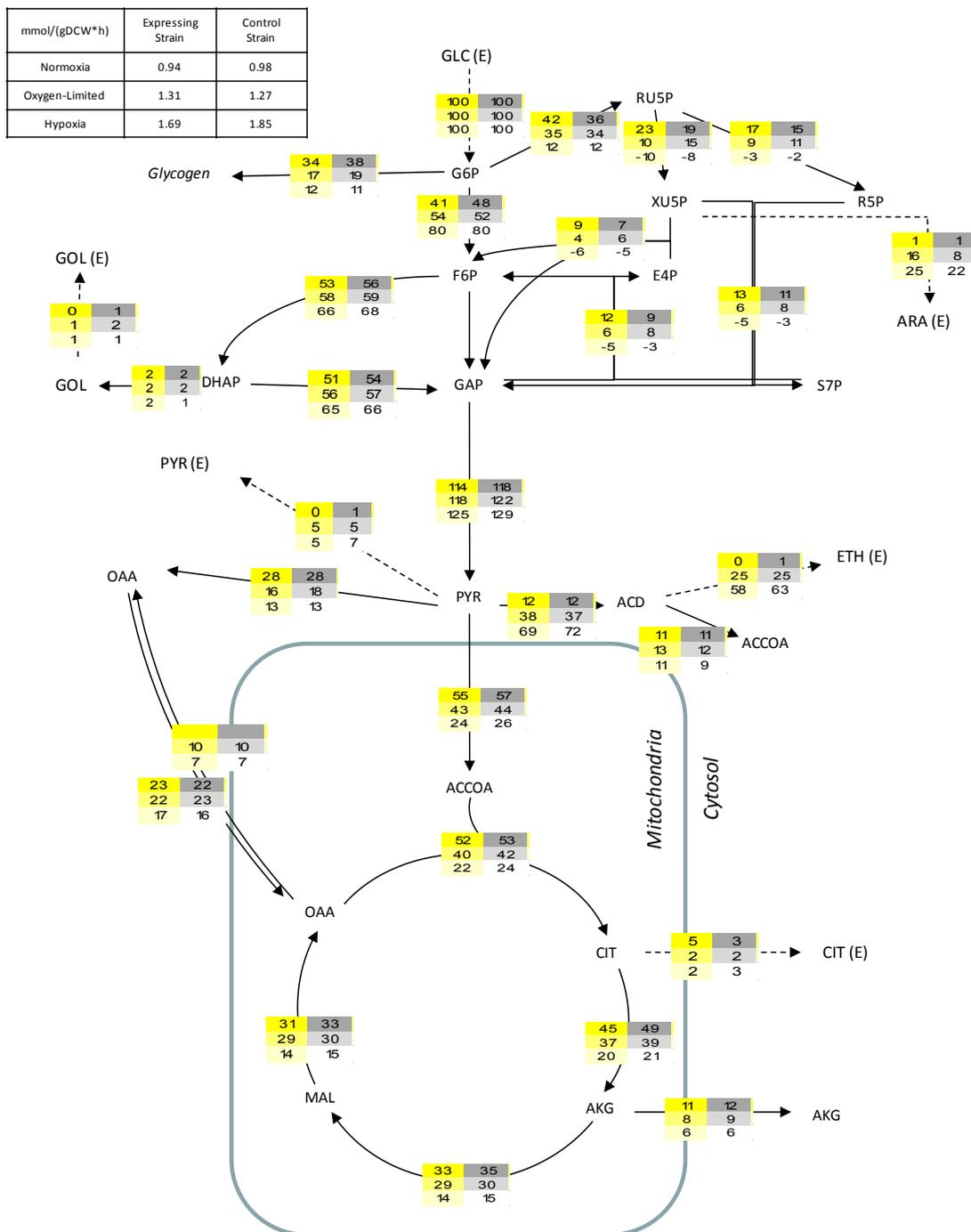


Figure 3.1: **Metabolic flux distributions in *P. pastoris* Fab-expressing and control strains under different oxygenation conditions.** Fluxes are shown as relative fluxes normalized to the specific glucose uptake rate in the corresponding experiment (Table 3.1). The fluxes for each reaction in the network corresponding to Normoxic, O<sub>2</sub>-limited and Hypoxic conditions are ordered from high to low oxygen availability. The transport of Oaa across the

mitochondrial membrane under normoxic conditions are given as a single net influx value showed in OAA\_transport\_cyt:mit reaction.

On the other hand, limitation in oxygen availability reduced the respiratory net carbon flux through the TCA cycle (that is, the net flux of aKG through the TCA cycle to OAA<sub>mit</sub>). Nevertheless, the fraction of the net carbon flux in the TCA cycle corresponding to the respirative carbon flux from aKG (or relative TCA cycle activity) remained between 50–60%, that is to say, the relative anaplerotic fluxes were between 40–50% (Figure 3.2). Similar results have been previously observed in *S. cerevisiae* under aerobic conditions [27]. No significant contribution for malic enzyme flux could be observed (Table 3.2), so the pyruvate carboxylase pathway was the only anaplerotic supply to the TCA cycle.

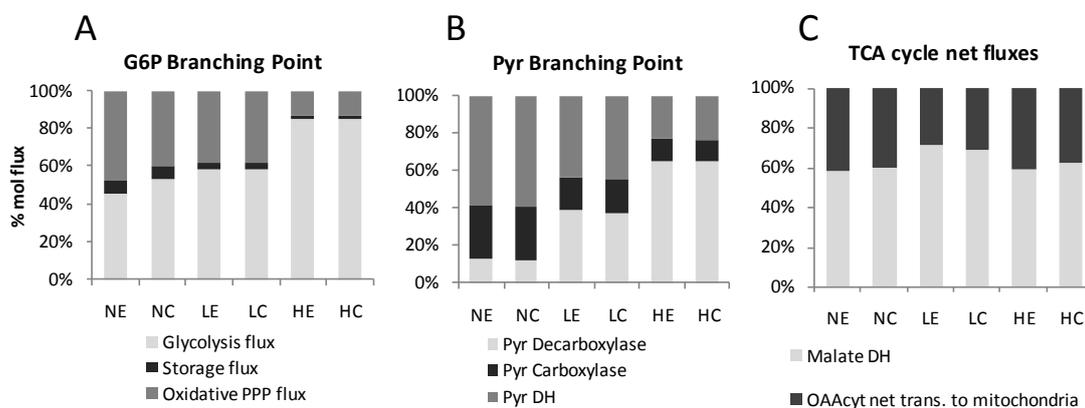


Figure 3.2: **Fractional distributions of carbon fluxes in metabolic branching points derived from MFA.** Fractional distributions of carbon fluxes from A) the G6P flux split to glycolysis and PPP, B) the pyruvate branching point and C) the TCA cycle, in *P. pastoris* Fab-expressing (E) and control (C) strains growing in glucose-limited chemostats at  $D = 0.1 \text{ h}^{-1}$ , in Normoxic (N), O<sub>2</sub>-Limited (L) and Hypoxic (H) conditions.

## Conclusions

In this chapter, new insights into the carbon metabolism of *P. pastoris* under Fab-producing and not producing conditions, at different oxygenation levels were shown. The combination of classical macroscopic physiological data with metabolic flux ratio analyses resulted in the determination of metabolic fluxes, providing new quantitative evidence on how *P. pastoris* central carbon metabolism reacts to the shift from respirative to respiro-fermentative metabolism. Nevertheless, no specific recombinant protein impact could be determined.

## Appendix

**Appendix 3.1:** Relative abundances of intact carbon fragments in proteinogenic amino acids. Relative abundances of intact C2 and C3 fragments (f-values) in proteinogenic amino acids describing the conservation of carbon chain fragments in *P. pastoris* Fab-producing and control strains growing in glucose-limited chemostats at a D = 0.1 h<sup>-1</sup> in different oxygenation conditions.

	X33 pGAP 3H6												X33 pGAP															
	21% Oxygen				11% Oxygen				8% Oxygen				21% Oxygen				11% Oxygen				8% Oxygen							
	f <sup>(1)</sup>	f <sup>(2)</sup>	f <sup>(2*)</sup>	f <sup>(3)</sup>	f <sup>(1)</sup>	f <sup>(2)</sup>	f <sup>(2*)</sup>	f <sup>(3)</sup>	f <sup>(1)</sup>	f <sup>(2)</sup>	f <sup>(2*)</sup>	f <sup>(3)</sup>	f <sup>(1)</sup>	f <sup>(2)</sup>	f <sup>(2*)</sup>	f <sup>(3)</sup>	f <sup>(1)</sup>	f <sup>(2)</sup>	f <sup>(2*)</sup>	f <sup>(3)</sup>	f <sup>(1)</sup>	f <sup>(2)</sup>	f <sup>(2*)</sup>	f <sup>(3)</sup>	f <sup>(1)</sup>	f <sup>(2)</sup>	f <sup>(2*)</sup>	f <sup>(3)</sup>
Ala α	0.06	0.11	0.03	0.80	0.01	0.07	0.00	0.92	0.01	0.05	0.00	0.94	0.07	0.14	0.00	0.80	0.00	0.07	0.00	0.93	0.01	0.07	0.00	0.93	0.01	0.07	0.00	0.93
Ala β	0.10	0.90			0.02	0.98			0.01	0.99			0.11	0.89			0.00	1.00			0.01	0.99			0.01	0.99		
Arg β	0.57	0.40	0.00	0.04	0.58	0.32	0.00	0.10	0.55	0.41	0.00	0.04	0.64	0.36	0.00	0.00	0.60	0.38	0.00	0.02	0.58	0.38	0.00	0.03	0.58	0.38	0.00	0.03
Arg δ	0.13	0.87			0.09	0.91			0.10	0.90			0.15	0.85			0.08	0.92			0.13	0.87			0.13	0.87		
Asp α	0.26	0.18	0.37	0.20	0.11	0.10	0.17	0.63	0.09	0.09	0.15	0.67	0.26	0.17	0.36	0.21	0.09	0.10	0.15	0.66	0.10	0.08	0.14	0.68	0.10	0.08	0.14	0.68
Asp β	0.25	0.29	0.37	0.09	0.12	0.70	0.15	0.03	0.09	0.72	0.14	0.05	0.27	0.28	0.36	0.10	0.08	0.73	0.15	0.04	0.12	0.71	0.13	0.04	0.12	0.71	0.13	0.04
Glu α	0.20	0.26	0.40	0.14	0.22	0.20	0.47	0.11	0.17	0.26	0.41	0.16	0.24	0.27	0.37	0.12	0.24	0.24	0.42	0.10	0.22	0.25	0.40	0.14	0.22	0.25	0.40	0.14
Glu β	0.63	0.37	0.00	0.00	0.63	0.37	0.00	0.00	0.60	0.38	0.00	0.02	0.65	0.35	0.00	0.00	0.67	0.33	0.00	0.00	0.57	0.42	0.00	0.00	0.57	0.42	0.00	0.00
Glu γ	0.07	0.00	0.93	0.00	0.02	0.00	0.98	0.00	0.00	0.01	0.99	0.00	0.05	0.00	0.95	0.00	0.00	1.00	0.00	0.00	0.03	0.00	0.97	0.00	0.03	0.00	0.97	0.00
Gly α	0.28	0.73			0.22	0.78			0.20	0.80			0.24	0.76			0.20	0.80			0.19	0.81			0.19	0.81		
His α	0.07	0.04	0.00	0.88	0.05	0.00	0.01	0.94	0.06	0.03	0.00	0.91	0.09	0.07	0.00	0.84	0.03	0.06	0.00	0.91	0.08	0.03	0.06	0.82	0.08	0.03	0.06	0.82
His β	0.16	0.55	0.00	0.29	0.12	0.65	0.00	0.22	0.10	0.59	0.00	0.31	0.15	0.51	0.00	0.34	0.08	0.62	0.00	0.30	0.11	0.51	0.02	0.37	0.11	0.51	0.02	0.37
His δ <sup>2</sup>	0.44	0.56			0.24	0.76			0.23	0.77			0.40	0.60			0.29	0.71			0.24	0.76			0.24	0.76		
Ile α	0.43	0.00	0.57	0.00	0.23	0.00	0.78	0.00	0.20	0.00	0.80	0.00	0.44	0.00	0.56	0.00	0.20	0.00	0.80	0.00	0.22	0.00	0.78	0.00	0.22	0.00	0.78	0.00
Ile γ <sup>1</sup>	0.54	0.45	0.00	0.01	0.75	0.17	0.00	0.08	0.82	0.17	0.00	0.01	0.57	0.43	0.00	0.00	0.83	0.17	0.00	0.00	0.85	0.15	0.00	0.00	0.85	0.15	0.00	0.00
Ile γ <sup>2</sup>	0.12	0.88			0.04	0.96			0.05	0.95			0.13	0.87			0.02	0.98			0.04	0.96			0.04	0.96		
Ile δ	0.55	0.45			0.84	0.16			0.82	0.18			0.57	0.43			0.82	0.18			0.91	0.09			0.91	0.09		
Leu α	0.12	0.00	0.88	0.00	0.07	0.00	0.93	0.00	0.07	0.00	0.93	0.00	0.10	0.00	0.90	0.00	0.04	0.00	0.96	0.00	0.09	0.00	0.91	0.00	0.09	0.00	0.91	0.00
Leu β	1.00	0.00	0.00	0.00	0.89	0.00	0.00	0.11	0.97	0.01	0.00	0.02	0.99	0.00	0.00	0.11	0.92	0.08	0.00	0.00	0.99	0.00	0.00	0.01	0.99	0.00	0.00	0.01
Leu δ <sup>1</sup>	0.14	0.86			0.08	0.82			0.08	0.92			0.14	0.86			0.06	0.94			0.09	0.91			0.09	0.91		
Leu δ <sup>2</sup>	1.00	0.00			0.99	0.02			1.00	0.00			1.00	0.00			1.00	0.00			1.00	0.00			1.00	0.00		
Lys α	0.11	0.00	0.89	0.00	0.03	0.07	0.75	0.15	0.09	0.00	0.90	0.01	0.12	0.00	0.88	0.00	0.06	0.00	0.94	0.00	0.11	0.00	0.89	0.00	0.11	0.00	0.89	0.00
Lys ε	0.13	0.87	0.00	0.00	0.08	0.92	0.00	0.00	0.11	0.89	0.00	0.00	0.14	0.86	0.00	0.00	0.09	0.91	0.00	0.00	0.13	0.87	0.00	0.00	0.13	0.87	0.00	0.00
Met α	0.27	0.20	0.33	0.20	0.08	0.11	0.17	0.64	0.10	0.16	0.15	0.59	0.24	0.23	0.32	0.22	0.08	0.13	0.14	0.66	0.13	0.16	0.14	0.57	0.13	0.16	0.14	0.57
Phe α	0.08	0.09	0.00	0.82	0.04	0.06	0.00	0.90	0.04	0.03	0.00	0.93	0.09	0.09	0.00	0.83	0.02	0.07	0.00	0.91	0.06	0.04	0.00	0.89	0.06	0.04	0.00	0.89
Phe β	0.06	0.94	0.00	0.00	0.01	0.99	0.00	0.00	0.03	0.97	0.00	0.00	0.09	0.91	0.00	0.00	0.03	0.97	0.00	0.00	0.05	0.95	0.00	0.00	0.05	0.95	0.00	0.00
Pro α	0.29	0.26	0.36	0.09	0.28	0.22	0.40	0.09	0.22	0.24	0.38	0.17	0.30	0.25	0.36	0.09	0.25	0.27	0.39	0.10	0.26	0.23	0.37	0.14	0.26	0.23	0.37	0.14
Pro β	0.64	0.36	0.00	0.00	0.64	0.34	0.00	0.02	0.54	0.41	0.00	0.06	0.64	0.34	0.00	0.03	0.63	0.38	0.00	0.00	0.73	0.27	0.00	0.00	0.73	0.27	0.00	0.00
Pro γ	0.11	0.86	0.00	0.03	0.06	0.92	0.00	0.02	0.05	0.90	0.00	0.05	0.14	0.86	0.00	0.00	0.05	0.91	0.00	0.04	0.10	0.88	0.00	0.02	0.10	0.88	0.00	0.02
Ser α	0.14	0.06	0.29	0.52	0.05	0.06	0.25	0.65	0.05	0.04	0.26	0.66	0.14	0.06	0.29	0.51	0.04	0.05	0.25	0.66	0.07	0.03	0.22	0.68	0.07	0.03	0.22	0.68
Ser β	0.41	0.59			0.29	0.71			0.30	0.70			0.44	0.56			0.29	0.72			0.29	0.71			0.29	0.71		
Thr α	0.26	0.19	0.35	0.20	0.12	0.10	0.17	0.62	0.10	0.08	0.14	0.68	0.27	0.17	0.36	0.20	0.11	0.09	0.13	0.67	0.13	0.09	0.13	0.65	0.13	0.09	0.13	0.65
Thr β	0.29	0.62	0.00	0.09	0.14	0.85	0.00	0.01	0.11	0.87	0.00	0.03	0.28	0.63	0.00	0.09	0.09	0.87	0.00	0.04	0.12	0.85	0.00	0.03	0.12	0.85	0.00	0.03
Thr γ <sup>2</sup>	0.51	0.49			0.83	0.17			0.86	0.14			0.51	0.49			0.84	0.16			0.86	0.14			0.86	0.14		
Tyr α	0.09	0.09	0.00	0.82	0.03	0.06	0.00	0.91	0.05	0.05	0.00	0.90	0.07	0.08	0.02	0.84	0.01	0.09	0.00	0.90	0.05	0.05	0.01	0.89	0.05	0.05	0.01	0.89
Tyr β	0.07	0.93	0.00	0.00	0.02	0.96	0.02	0.00	0.04	0.96	0.00	0.00	0.09	0.91	0.00	0.00	0.02	0.99	0.00	0.00	0.07	0.93	0.00	0.00	0.07	0.93	0.00	0.00
Tyr δ <sup>x</sup>	0.05	0.95	0.00		0.04	0.96	0.00		0.03	0.97	0.00		0.08	0.92	0.00		0.00	1.00	0.00	0.00	0.04	0.96	0.00		0.04	0.96	0.00	
Tyr ε <sup>x</sup>	0.44	0.09	0.06	0.41	0.46	0.04	0.11	0.40	0.38	0.07	0.10	0.45	0.46	0.06	0.08	0.39	0.42	0.05	0.12	0.41	0.40	0.03	0.14	0.43	0.40	0.03	0.14	0.43
Val α	0.18	0.00	0.82	0.00	0.12	0.00	0.86	0.02	0.07	0.00	0.90	0.03	0.19	0.00	0.81	0.00	0.09	0.00	0.92	0.00	0.10	0.00	0.90	0.00	0.10	0.00	0.90	0.00
Val γ <sup>1</sup>	0.08	0.92			0.05	0.95			0.03	0.97			0.12	0.88			0.02	0.98			0.06	0.94			0.06	0.94		
Val γ <sup>2</sup>	0.99	0.01			1.00	0.00			1.00	0.00			1.00	0.00			1.00	0.00			1.00	0.00			1.00	0.00		

**Appendix 3.2:** Metabolic fluxes. Metabolic fluxes in the central carbon metabolism of *P. pastoris* Fab-producing and control strain in glucose-limited chemostats at a  $D = 0.1 \text{ h}^{-1}$ , in different oxygenation conditions. The values are in  $\text{mmol}/(\text{gDCW}\cdot\text{h})$  and its standard deviations. N: Normoxia; L: O<sub>2</sub>-Limited; H: Hypoxic; C: Control Strain; E: Expressing Strain.

	Fab-expressing Strain			Control Strain		
	Normoxia	Oxygen-limited	Hypoxia	Normoxia	Oxygen-limited	Hypoxia
HXT+HXK	0.94 ± 0.03	1.31 ± 0.04	1.69 ± 0.03	0.98 ± 0.03	1.27 ± 0.03	1.85 ± 0.07
PGI	0.38 ± 0.06	0.70 ± 0.05	1.34 ± 0.06	0.47 ± 0.06	0.66 ± 0.06	1.46 ± 0.08
PFK+FBA	0.50 ± 0.03	0.76 ± 0.04	1.11 ± 0.04	0.55 ± 0.03	0.75 ± 0.03	1.24 ± 0.07
GAPDH-PYK	1.08 ± 0.06	1.55 ± 0.09	2.11 ± 0.07	1.16 ± 0.06	1.55 ± 0.06	2.38 ± 0.14
TPI	0.48 ± 0.03	0.74 ± 0.04	1.09 ± 0.04	0.53 ± 0.03	0.72 ± 0.03	1.22 ± 0.07
G6PDH-6PGDH	0.40 ± 0.07	0.46 ± 0.04	0.22 ± 0.05	0.35 ± 0.07	0.43 ± 0.07	0.23 ± 0.09
RPE	0.22 ± 0.04	0.13 ± 0.03	-0.17 ± 0.03	0.18 ± 0.04	0.18 ± 0.04	-0.14 ± 0.03
RPI	0.16 ± 0.02	0.11 ± 0.02	-0.04 ± 0.02	0.15 ± 0.02	0.14 ± 0.02	-0.03 ± 0.02
TK(1)	0.12 ± 0.02	0.08 ± 0.01	-0.07 ± 0.02	0.11 ± 0.02	0.10 ± 0.02	-0.06 ± 0.02
TA	0.09 ± 0.02	0.05 ± 0.01	-0.09 ± 0.02	0.07 ± 0.02	0.08 ± 0.02	-0.08 ± 0.02
TK(2)	0.11 ± 0.02	0.07 ± 0.01	-0.07 ± 0.02	0.09 ± 0.02	0.10 ± 0.02	-0.06 ± 0.02
PDH	0.52 ± 0.05	0.57 ± 0.09	0.42 ± 0.07	0.56 ± 0.05	0.57 ± 0.04	0.49 ± 0.12
CIT Syn.	0.49 ± 0.05	0.52 ± 0.09	0.37 ± 0.08	0.52 ± 0.05	0.53 ± 0.04	0.44 ± 0.12
ISOCITDH	0.43 ± 0.05	0.49 ± 0.09	0.33 ± 0.08	0.48 ± 0.05	0.50 ± 0.04	0.39 ± 0.11
aKGDH-FUM	0.31 ± 0.05	0.39 ± 0.09	0.23 ± 0.08	0.35 ± 0.05	0.38 ± 0.04	0.28 ± 0.11
MalDH	0.30 ± 0.05	0.38 ± 0.09	0.23 ± 0.08	0.33 ± 0.05	0.38 ± 0.04	0.28 ± 0.11
PyrCK	0.27 ± 0.01	0.22 ± 0.01	0.22 ± 0.01	0.28 ± 0.01	0.23 ± 0.01	0.23 ± 0.02
PDC	0.11 ± 0.01	0.49 ± 0.05	1.15 ± 0.06	0.12 ± 0.01	0.46 ± 0.02	1.31 ± 0.06
ADDH + ACETCoA	0.10 ± 0.01	0.16 ± 0.05	0.17 ± 0.04	0.11 ± 0.01	0.14 ± 0.01	0.15 ± 0.01
ET	0.00 ± 0.00	0.33 ± 0.02	0.98 ± 0.04	0.01 ± 0.00	0.31 ± 0.02	1.16 ± 0.06
G3PDH	0.00 ± 0.00	0.01 ± 0.00	0.02 ± 0.00	0.01 ± 0.00	0.02 ± 0.00	0.02 ± 0.00
ARA Syn.	0.01 ± 0.00	0.21 ± 0.00	0.42 ± 0.00	0.01 ± 0.00	0.10 ± 0.00	0.40 ± 0.09
PyrCK_OAA_G	0.21 ± 0.01	- ± -	- ± -	0.22 ± 0.01	- ± -	- ± -
OAA_Transp_cyt-mit	- ± -	0.28 ± 0.06	0.28 ± 0.08	- ± -	0.29 ± 0.02	0.29 ± 0.08
OAA_Transp_mit-cyt	- ± -	0.13 ± 0.06	0.12 ± 0.08	- ± -	0.12 ± 0.01	0.12 ± 0.08
aKG_Transp_mit-cyt	0.10 ± 0.01	0.10 ± 0.01	0.10 ± 0.01	0.12 ± 0.01	0.11 ± 0.01	0.10 ± 0.01
Glycerol Syn.	0.01 ± 0.00	0.02 ± 0.00	0.03 ± 0.00	0.02 ± 0.00	0.03 ± 0.00	0.03 ± 0.00
Pyruvate Syn.	0.00 ± 0.00	0.06 ± 0.00	0.09 ± 0.00	0.01 ± 0.00	0.06 ± 0.00	0.12 ± 0.01
Citrate Syn.	0.05 ± 0.00	0.02 ± 0.00	0.04 ± 0.00	0.03 ± 0.00	0.03 ± 0.00	0.06 ± 0.02
Prorein	1.14 ± 0.10	1.39 ± 0.12	1.50 ± 0.11	1.32 ± 0.09	1.49 ± 0.10	1.53 ± 0.13
Carbohydrate	1.04 ± 0.12	1.10 ± 0.13	0.93 ± 0.10	0.94 ± 0.07	1.32 ± 0.10	1.20 ± 0.11
Glycogen	0.32 ± 0.03	0.22 ± 0.02	0.20 ± 0.01	0.38 ± 0.03	0.24 ± 0.02	0.20 ± 0.02
Lipid	0.22 ± 0.02	0.37 ± 0.12	0.39 ± 0.11	0.22 ± 0.02	0.31 ± 0.02	0.34 ± 0.03
RNA	0.23 ± 0.03	0.23 ± 0.03	0.23 ± 0.03	0.24 ± 0.02	0.24 ± 0.02	0.19 ± 0.02
DNA	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00

## Chapter 4: *Pichia pastoris* intracellular metabolite quantification method<sup>c</sup>

---

<sup>c</sup> Published as: Carnicer, M., Canelas, A. B., Pierick, A., Zeng, Z., Dam, J., Albiol, J., Ferrer, P., Heijnen, J. J., and Gulik, W. ***Development of quantitative metabolomics for Pichia pastoris***. *Metabolomics* 2012, **8**:284-298.



## Background

The metabolomics field has experienced a strong development over the recent past, mainly due to the improvements in MS-based analytical procedures [11, 29–31]. In combination with the advances in analysis of microbial metabolic fluxes via  $^{13}\text{C}$  isotopic labelling [89] this has greatly expanded the possibilities for quantitative analysis of metabolic pathways. Nevertheless, an important step in the study of metabolic reaction networks is to obtain representative and accurate snapshots of the metabolome. Many metabolites, not only those related to the central carbon metabolism but e.g. also free amino acids, have turnover times in the order of seconds or less, which highlight the need for a rapid sampling technique to quantify their actual levels, as well as a proper, leakage-free, quenching procedure to ensure absence of losses or (inter)conversion of metabolites [9, 33]. Moreover, if substantial amounts of metabolites are present in the extracellular medium, those need to be efficiently removed. Therefore, a sample treatment procedure allowing to separate the intracellular and extracellular metabolite pools, as well as a degradation-free extraction method are required [34].

Although many efforts have been directed towards the development of a universal method, no consensus solution has been found because of the vast diversity in cell properties [9, 10, 35]. Recently, proper sampling, quenching, separation and extraction protocols for the intracellular metabolite quantification in *Saccharomyces cerevisiae* have been presented and successfully applied, allowing accurate, reliable and reproducible metabolite determinations [33, 34]. Unfortunately, it appears that for different microbial species, different sampling and quenching methodologies need to be developed and quantitatively evaluated and validated [10, 36, 37].

Therefore, in this chapter, a systematic evaluation of five quenching protocols applied to *P.pastoris* was carried out in order to obtain a reliable intracellular metabolite quantification method. Methanol concentration and temperature were the selected variables to be optimized for minimizing metabolite leakage during quenching and subsequent washing. Thereafter, the best performing quenching protocol was used in two applications: 1) determination of the time needed for each intracellular metabolite to reach a metabolite steady state and 2) a metabolome comparison between *P.pastoris* and *S.cerevisiae*.

## Materials and Methods

### ***Strain and cultivation conditions***

Analytical grade reagents were supplied by Sigma-Aldrich. HPLC-grade methanol and ethanol were supplied by J.T. Baker.

In this study, the *P. pastoris* X-33 (wild type phenotype, Invitrogen) transformed with pGAP $\alpha$ A (Invitrogen) as mock vector was used. Chemostat cultivation was performed in a 7-liter fermentor

(Applikon, The Netherlands) with a working volume of 4 liter. The culture media used were derived from previously described media [47]. In particular, the C-source of the chemostat medium was reduced to obtain a steady state biomass concentration of approximately 4.5 g/l. The other components were adjusted accordingly to obtain similar residual concentrations as with the original medium, thereby considering the biomass composition of *P.pastoris* [76]. The composition of the batch medium was: 8 g/l glycerol, 0.9 g/l citric acid monohydrate, 12.6 g/l  $(\text{NH}_4)_2\text{HPO}_4$ , 0.5 g/l  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 1.5 g/l  $\text{KH}_2\text{PO}_4$ , 0.02 g/l  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ , 5 ml/l trace salt solution, 2 ml/l Biotin solution (0.2 g/l). The composition of the chemostat medium was: 8.80 g/l glucose monohydrate, 0.92 g/l citric acid monohydrate, 2 g/l  $(\text{NH}_4)_2\text{HPO}_4$ , 0.3 g/l  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 1.4 g/l  $\text{KH}_2\text{PO}_4$ , 0.01 g/l  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ , 0.5 ml/l trace salt solution, 0.3 ml/l Biotin (0.2 g/l). The trace salts solution was the same as described previously [47].

One liter shake flask containing 300 ml of YPD medium (10 g/l yeast extract, 20 g/l peptone, 10 g/l glucose) was inoculated with a 1.0 ml cryostock of *P. pastoris* cells. The culture was grown for approximately 24 h at 30 °C with shaking at 200 rpm, and used to inoculate the reactor. After complete termination of the batch phase, approximately 24 h after performing the inoculation, the feed of the chemostat culture was started. During the chemostat, the cells were grown under carbon-limited conditions at a dilution rate (D) of  $0.1 \text{ h}^{-1}$ , and an aeration rate of 0.5 vvm, controlled by mass flow meters (5850 Smart Mass Flow Controller, Brooks Instrument). During the entire cultivation the  $\text{pO}_2$  was maintained above 50% ensuring fully aerobic conditions. The  $\text{O}_2$  and  $\text{CO}_2$  concentrations in the bioreactor off-gas were measured on-line using a combined paramagnetic/infrared analyzer (NGA 2000, Rosemount, USA). Pressure, pH, stirring speed and temperature were maintained at 1.2 bars, pH 5 (with 20% v/v  $\text{NH}_3$ ), 400 rpm and 25 °C, respectively.

In the leakage evaluation experiment, a chemostat culture was maintained for 5 residence times until a stable off-gas reading was obtained. This measure was used as an indicator for obtaining a steady-state condition. During steady state five duplicate samples were taken, i.e. two for each quenching protocol, for metabolite measurements in quenched and washed cells, and in quenching and washing solutions. Furthermore three filtrate samples and five whole broth samples were taken.

Once the protocols were evaluated, using the best of the tested protocols, two additional chemostat cultures were run to study the time needed for *P. pastoris* to reach a metabolic pseudo steady-state. Duplicate samples for intracellular metabolite measurement were taken approximately each 24 hours during a period of 10 residence times.

### **Sampling**

Samples for intracellular metabolite analysis were taken using a dedicated rapid-sampling setup [32]. With this setup approximately,  $0.63 \text{ g} \pm 0.01$  of broth was rapidly withdrawn and immediately injected in 5 ml of precooled quenching solution. The tubes were quickly mixed by vortexing and introduced in

the filtration unit after weighting the tube [90]. All sampling tubes were weighted before and after the sampling procedure in order to determine the exact amount of sample taken. Briefly, the cell suspensions were filtered with membrane disk filters (Pall Corporation, East Hills, NY, USA, 47mm diameter, 0.45  $\mu\text{m}$  pore size) using a vacuum pump. A washing step was performed to remove as much extracellular metabolites as possible. This washing step consisted of rapidly pouring a fresh amount of 10 ml of the previously mentioned quenching solution on the filter cake as soon as the biomass started to fall dry, thereby maintaining the vacuum on the filtration unit. The average total contact time between the cells and the quenching solution (from taking the sample until the washing solution completely passed through the filter) was 1 min +/- 7 seconds.

For accurate quantification purposes by using Isotope Dilution Mass Spectrometry [29, 91], 120  $\mu\text{l}$  of a  $^{13}\text{C}$  internal standard solution (0°C) was pipetted on top of the dry filter cake. The  $^{13}\text{C}$  internal standard solution contained all relevant metabolites as U- $^{13}\text{C}$  -labeled isotopes and was obtained from a *S. cerevisiae* fed-batch culture grown on 100 % U- $^{13}\text{C}$  -labeled glucose and ethanol. Metabolite extraction with 75 % (v/v) aqueous ethanol at 95 °C and further sample processing were carried out as described previously [90].

The quenching and washing liquids (QWS) were collected in 50 ml Falcon tubes. After thorough mixing, aliquots of 3.0 ml of the solutions were mixed with 120  $\mu\text{l}$  of a  $^{13}\text{C}$  internal standard solution and subsequently extracted in 75 % (v/v) aqueous ethanol at 95 °C, to eliminate possible enzymatic activity in these samples. The exact sample amounts taken were determined by weighting all tubes before and after sampling.

Samples from the culture filtrate (CF) and the complete culture broth (WB) were withdrawn and further processed as described earlier [33].

### ***Metabolite analysis***

Metabolite quantification was carried out with LC-ESI-MS/MS and GC-MS based isotope dilution mass spectrometry (IDMS) [34]. Each sample was analyzed in duplicate. In total, 37 metabolites, with a wide variety of chemical and physical properties, were analyzed. However, only 34 of these could actually be measured, because the concentrations of glyoxylate, mannitol-6-phosphate and fructose-2,6-bisphosphate were found to be below the detection limit in all the samples analyzed.

### ***Consistency check and data reconciliation***

The amount of each metabolite was quantified in different sample fractions, that is, in whole broth (WB), quenched/washed cells (QC), culture filtrate (CF) and quenching + washing liquid (QWS). Because the  $^{13}\text{C}$  internal standard mix was in all cases added to the different sample fractions prior the

metabolite extraction procedure, possible metabolite losses due to partial degradation could be corrected for. This implies that the following mass balance should be satisfied for each metabolite  $i$ :

$$M_i(WB) = M_i(CF) + M_i(QC) + M_i(leakage) \quad (4.1)$$

This mass balance allows calculating, from measurements in total broth, filtrate and cells, which amount was lost into the quenching and washing solutions as a result of leakage. However, because in addition also metabolite quantifications were carried out in the quenching + washing solutions, a second mass balance can be written, namely:

$$M_i(WB) = M_i(QC) + M_i(QWS) \quad (4.2)$$

This balance states that the total amount of each metabolite (i.e. the extracellular + intracellular amount) present in a whole broth sample should be equal to the amounts measured in the cell cake sample and in the quenching + washing solution. This means that the data set contains redundant information, i.e. more information than strictly necessary to quantify metabolite leakage, allowing to perform a statistical consistency check, using the calculated  $\chi^2$ -distributed consistency index  $h$ , as well as data reconciliation[69].

Briefly the data reconciliation was performed as follows: under the constraint that the mass balances (Eqs, 1 and 2) should be satisfied, the best estimates of the measured metabolite amounts in the different sample fractions as well as the calculated amount of leakage were obtained by least squares minimization of the differences between the measured and estimated amounts, weighed by their measurement errors [38].

## Results and Discussion

### ***Chemostat cultivations***

*P. pastoris* was grown in aerobic, glucose limited chemostat cultures at a dilution rate of  $0.1 \text{ h}^{-1}$ . Under these conditions biomass and carbon dioxide were the only products. Once steady state was obtained, the consumption rates of glucose and oxygen and the production rates of biomass and carbon dioxide were calculated from measurements of biomass dry weight, residual glucose and the concentrations of oxygen and carbon dioxide in the off-gas.

The experimental data consistency was verified using standard data reconciliation procedures, under the constraint that the elemental conservation relations were satisfied [38, 69]. For all chemostat cultivations performed the statistical consistency test, carried out with a confidence level of 95%, was acceptable, indicating that there was no proof for gross measurement errors. Therefore, under the applied chemostat conditions, the balanced steady state input-output rates obtained were  $-0.97 (\pm 0.01)$ ,  $3.65 (\pm 0.01)$ ,  $2.19 (\pm 0.09)$  and  $-2.11 (\pm 0.09) \text{ mmol}/(\text{gDCW}\cdot\text{h})$  for glucose uptake rate, Biomass production,  $\text{CO}_2$  evolution rate and oxygen uptake rate respectively. As expected, the rates were

equivalent to the ones obtained in previous studies under normoxia conditions, thereby making the data sets comparable [52, 76](Chapter 2 and 3).

### **Quenching optimization**

#### Effect of methanol content and quenching temperature

To study the effect of the quenching procedure on metabolite leakage, a full mass balance analysis was performed as described in [33]. Briefly, the fate of the metabolites was identified by quantification of metabolite levels in four different fractions: whole-broth fraction (WB), culture filtrate (CF), quenched/washed cells (QC) and quenching + washing methanol solution (QWS). The actual intracellular metabolite levels were estimated from the difference between the levels measured in whole broth (WB) and culture filtrate (CF)[33, 37]. However, this approach for estimation of intracellular metabolite levels has drawbacks compared to the direct measurement of quenched/washed cells. In particular, it requires a double analytical effort, as two samples have to be analyzed, and whole broth and supernatant samples may contain high amounts of salts, potentially interfering (ion suppression) with the analytical techniques [33]. On the other hand, in the quenched/washed cells, metabolism needs to be properly arrested, while avoiding leakage and degradation.

**Table 4.1:** Tested quenching protocols to investigate the effects of temperature and methanol concentration on metabolite leakage.

Protocol	Temperature (°C)	Concentration of Methanol solutions (v/v) (%)	Concentration of Methanol after sampling (v/v) (%)
A	-27	40	36
B	-27	60	54
C	-40	60	54
D	-40	80	71
E	-40	100	89

Therefore, a comparison of the determined intracellular metabolite levels between the direct measurement (QC) and the differential method (WB-CF) was performed for the 5 different variations of the cold methanol quenching protocols, to determine for which condition metabolite leakage from *P.pastoris* cells was minimal. The methanol content of the quenching solution and the quenching temperature were the parameters changed to investigate their impact on possible metabolite leakage as shown in Table 4.1.

The metabolites analyzed consisted of a wide range of different chemical compounds, such as phosphorylated intermediates, organic acids and amino acids. In some cases the fate of the metabolite could not be followed properly and therefore the mass balances could not be calculated. For example, the amounts of FBP and T6P in CF, QWS and WB were below the detection limits while for citric acid the

differential method could not be applied due to a too high extracellular level outside measuring range and therefore, QC was the only measurement which could be carried out.

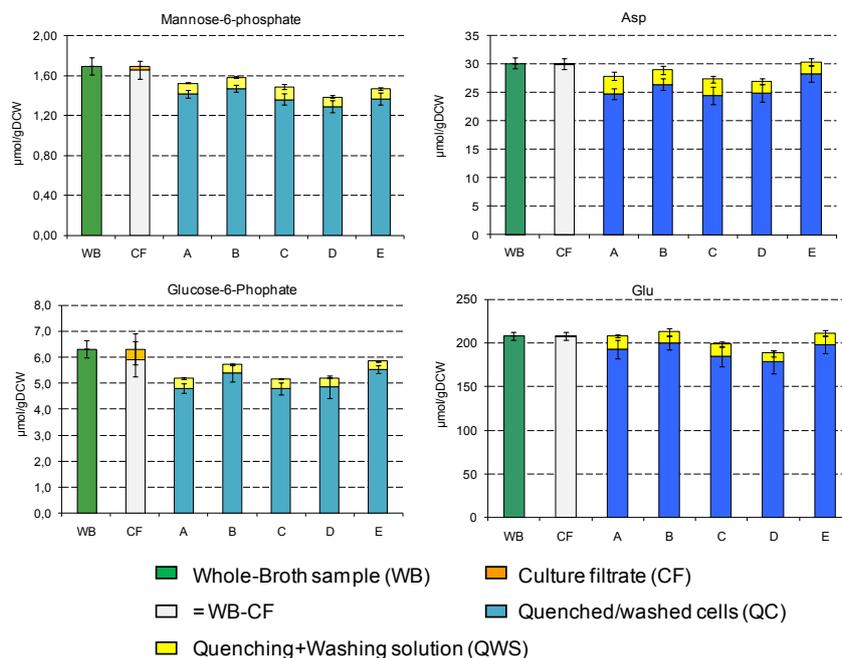


Figure 4.1: **Full mass balances of some representative metabolites.** The letters below the bars represent the five different treatments tested. **A:**  $-27^{\circ}\text{C}$ , 40% MeOH; **B:**  $-27^{\circ}\text{C}$ , 60% MeOH; **C:**  $-40^{\circ}\text{C}$ , 60% MeOH; **D:**  $-40^{\circ}\text{C}$ , 80% MeOH; **E:**  $-40^{\circ}\text{C}$ , 100% MeOH.

As a first step of the data evaluation, the average mass balance closure was calculated  $((\text{QC} + \text{QWS})/\text{WB})$  being on average 108% ( $\pm 24\%$ ). This was considered acceptable due to the analytical challenge and was in the range of values obtained in previous studies employing the mass balance approach [33]. Nevertheless, the data consistency was also checked individually for each metabolite to detect any gross errors in the measurements (See Protocol evaluation section).

Secondly, the full mass balances were analyzed metabolite by metabolite. In Figure 4.1 some examples are shown, representing different classes of metabolites. Comparison of the intracellular metabolite levels obtained with the differential method (WB-CF) and the quenched cell measurements (QC) revealed that metabolite leakage occurred to some extent in all treatments, independent of the methanol content of the quenching solution, the quenching temperature or the compound type. This could be attributed to the so called “cold shock” phenomenon, leading to a sudden release of metabolites from the cells when the broth is rapidly cooled, as reported for the bacterium *Corynebacterium glutamicum* [92, 93].

However, in a previous study performed with the yeast *S. cerevisiae* [33] the extent of metabolite leakage was found to be heavily influenced by the methanol content used in the quenching solution; whereby decreasing the methanol content resulted in increased leakage. These results highlight the

need to fine tune the quenching protocol for each microorganism. On the other hand, the much smaller influence of the methanol content observed for *P. pastoris* could also be attributed to the shorter contact time with the quenching solution (about one minute) when applying the cold filtration method. In the work of [33] the cold centrifugation method was used, whereby the cells were between 20-30 minutes in contact with the quenching solution. Detailed data of the metabolite quantification results for each protocol is provided as supplementary material (see appendix 4.2).

**Table 4.2:** Direct and differential intracellular quantification comparison. sd= standard deviation; < = Metabolite lower than detection limit; > = Metabolite higher than the calibration range.

$\mu\text{mol/gDCW}$	ICS Protocol B		WB-CF	
	Value	sd	Value	sd
<b>G6P</b>	5.4	0.3	5.9	0.7
<b>F6P</b>	1.22	0.09	1.21	0.06
<b>FBP</b>	0.70	0.07	<	<
<b>2-3PG</b>	1.93	0.22	2.14	0.12
<b>PEP</b>	0.33	0.04	0.33	0.02
<b>Citric acid</b>	7.3	0.6	>	>
<b><math>\alpha</math>KG</b>	2.09	0.15	2.25	0.18
<b>Succinate</b>	2.94	0.18	2.79	0.27
<b>Fumarate</b>	1.37	0.06	1.09	0.15
<b>Malate</b>	6.9	0.6	7.0	0.9
<b>G1P</b>	0.80	0.06	0.68	0.06
<b>6PG*</b>	0.67	0.10	0.57	0.09
<b>M6P</b>	1.47	0.03	1.66	0.09
<b>Trehal.-6P</b>	0.09	0.01	<	<
<b>G3P</b>	0.06	0.01	0.05	0.04
<b>S7P</b>	2.08	0.09	2.08	0.16
<b>UDP-glc</b>	1.18	0.31	0.95	0.17
<b>Ala</b>	23.5	0.1	23.8	2.4
<b>Val</b>	1.97	0.07	1.99	0.05
<b>Leu</b>	1.48	0.10	1.50	0.02
<b>Ile</b>	0.53	0.04	0.55	0.01
<b>Thr*</b>	3.94	0.02	4.19	0.12
<b>Pro</b>	10.8	0.4	11.7	0.2
<b>Asn</b>	7.1	0.2	7.6	0.2
<b>Asp</b>	26.4	1.0	30.1	0.9
<b>Met</b>	0.97	0.15	1.03	0.15
<b>Glu</b>	200	8	208	5
<b>Phe</b>	0.51	0.04	0.52	0.01
<b>Gln*</b>	177	8	178	5
<b>Orn</b>	56	3	54.8	0.6
<b>Lys</b>	12.8	0.7	12.9	0.4
<b>His</b>	7.4	0.5	7.3	0.1
<b>Tyr</b>	0.80	0.06	0.82	0.02
<b>Trp</b>	0.24	0.02	0.23	0.01

\* No CF quantification available. WB-CF was replaced by WB measurement.

### **Protocol evaluation**

From the evaluation of the mass balances for each metabolite, as shown in Figure 4.1, no clear distinction could be made between the different protocols. Therefore, data reconciliation was applied under the constraint that for each metabolite both mass balances which could be evaluated (Eqs. 4.1 and 4.2) were satisfied (see materials and methods section). Thus for each metabolite the data consistency was evaluated at 95% confidence level and one degree of freedom. Under these conditions the calculated,  $\chi^2$ -distributed, consistency index  $h$  should have a value of 3.84 or less. The calculated  $h$ -index values are shown in the appendix section (Appendix 4.1). When inspecting these values it can be seen that in some cases proof is obtained for gross measurements errors ( $h > 3.84$ ). Only, for two metabolites, namely Pyr and Gly, the data were inconsistent for all the treatments tested. In case of Pyr these systematic errors could be related to quantification errors in Pyr peaks due to analytical difficulties. However, for Gly no clear reason could be identified for the data inconsistency.

Interestingly, the free amino acid quantifications showed higher global consistency, having lower  $h$  values compared to the central carbon metabolites, thereby indicating more accuracy in the GC-MS analysis of amino acids.

In the subsequent calculation of the average metabolite yield ( $QC/(WB-CF)$ ) for the different quenching protocols (A to E) only the statistically consistent data ( $h < 3.84$ ) were taken into account. The recoveries and the standard error for each protocol were  $93.1 \pm 1.1$ ,  $95.4 \pm 0.7$ ,  $92.6 \pm 1.3$ ,  $94.3 \pm 1.1$  and  $93.4 \pm 1.5$  for Protocols A to E respectively (Table 4.1). Although the differences were small, application of Protocol B resulted in the highest average yield, being approximately 95%.

Consequently, based on the data consistency and average metabolite yield, protocol B was considered as the optimum quenching procedure for quantification of the intracellular metabolites in *P. pastoris*. Moreover, in order to evaluate the applicability of direct measurement using quenching protocol B, the obtained results were compared with the results obtained using with the differential method (see in Table 4.2). To test for significant differences between the results obtained with both methods a two tailed Student's T test was performed. Results show that, only in the case of Asp the obtained values were significantly different, while no significant difference was detected in any other metabolite.

Therefore it is concluded that, in spite of the fact that for one metabolite (Asp) the values were slightly deviated, direct measurement, using quenching protocol B is the preferred methodology for quantification of intracellular metabolites in *P. pastoris*. In comparison to the differential method, this procedure results, on average, in smaller measurement errors. Furthermore, it requires less analytical effort because for each measurement only one sample has to be analyzed instead of two (WB and CF) for the differential method. Besides, as for the WB and CF no medium components are eliminated, they could potentially interfere (ion suppression) with the analytical techniques [33]. For these reasons, we

applied, for the experiments described in the next section, quenching protocol B for the quantification of the *P. pastoris* metabolome.

In a recent paper of Tredwell and co-workers [94], published after our paper had been submitted, also the evaluation of procedures for sampling and cold methanol quenching of *P. pastoris* for metabolome analysis is described. Although the quenching conditions were different from ours with respect to temperature, methanol content and the addition of buffers, they also found that the different variations of the cold methanol quenching method used gave very similar results. Apparently *P. pastoris* is a relatively robust microorganism, resistant to cold methanol quenching with respect to metabolite leakage. Also Tredwell et al. present baseline metabolome data for chemostat cultured *P. pastoris* cells and present a comparison between the *P. pastoris* and *S. cerevisiae* metabolome. The conditions they applied were, however, highly different with respect to the strains of *P. pastoris* and *S. cerevisiae*, the cultivation conditions (batch vs chemostat) and the substrates (methanol and glycerol vs glucose) used. Most important difference is that the study of Tredwell at all aimed at metabolic profiling, using non-targeted analytical techniques, whereas in our work a targeted, quantitative approach was used applying isotope dilution mass spectrometry (IDMS) aimed at obtaining highly accurate metabolome data.

### **Steady-state evaluation**

Typically, the first approach in metabolomic studies is the quantification of the (pseudo) steady-state metabolite concentrations in order to obtain a better understanding of the cell behaviour during fixed environmental conditions.

For measurement of metabolite levels in steady state chemostat cultivation, it is generally assumed, but seldom verified, that the intracellular metabolites have reached their steady state levels after 5 residence times. To verify this for chemostat cultivation of our *P. pastoris* strain, two replicate chemostat cultures were carried out. Rapid sampling, combined with quenching according to protocol B and filtration, was applied for intracellular metabolite measurement at 24 hours intervals, from the start of the chemostat phase until a period of 10 residence times.

For the metabolites of which the levels evolved towards different steady state values, an exponential decay curve was fitted (Equation 4.3), whereby each measurement was weighted by its standard deviation.

$$C_i(t) = C_{i,SS} + (C_i(0) - C_{i,SS}) \cdot e^{-b \cdot t} \quad (4.3)$$

Herein  $C_i(0)$  is the average quantification pool for each metabolite at the end of the batch phase,  $C_{i,SS}$  is the steady state level and  $b$  is a time constant.

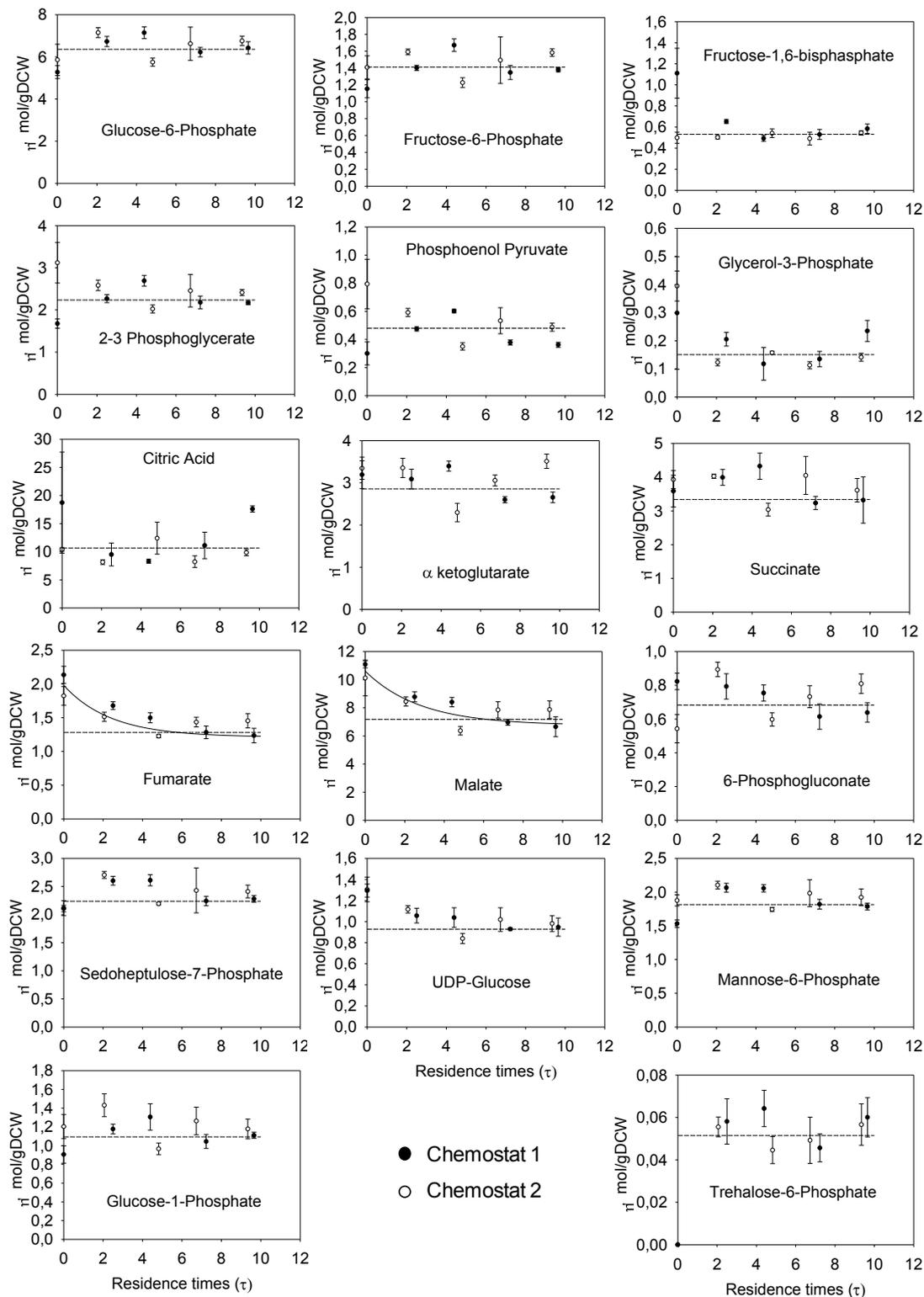


Figure 4.2: Evolution changes in intracellular pools associated with adaptation to glucose-limited conditions in two chemostats. The line represent the exponential decay curve fitted and dashed lines the weighted average of the last five residence times.

For the metabolites which were already at a stable level from the beginning, a weighted average of the 5 last residence times was taken as the steady state value. Moreover, this was also done for the metabolites which showed changing levels to compare the results from both calculations.

The results are shown in Figures 4.2 and 4.3. Globally, the small differences between the profiles obtained for the two chemostats shows the high reproducibility of the cultivations. Furthermore the relatively small errors in the individual metabolite measurements indicate a high analytical reproducibility for most of the metabolites reinforcing the choice of the Protocol B as the optimum one for the intracellular metabolite quantification. However, some biological differences were observed for few metabolites when the pool sizes of the two independent chemostat were compared.

Focusing on the profiles of metabolites from the upper glycolysis and the pentose phosphate pathway (Figure 4.2), for most metabolites no significant changes were observed during the 10 residence times of chemostat cultivation. This could be explained by the absence of pool size differences between the 2 growth conditions (batch phase and continuous phase) or a very fast adaptation after shifting from batch to continuous operation mode for that part of the metabolism. However, a slight increase of the G6P and F6P levels were observed as well as a more pronounced decrease in the pool size of G3P, probably because the carbon source was changed from glycerol to glucose when the continuous culture phase was started. In the TCA cycle, fumarate and malate could be described using an exponential decay profile as indicated by the fitted curve showing slower adaptation compared to other metabolites. Moreover, no changes were observed in the levels of other TCA cycle metabolites, such as  $\alpha$ KG and succinate, during 10 residence times. A metabolite related to the storage metabolism, such as T6P, was below the detection limit at the end of the batch phase and increased to higher steady state levels in less than 24 hours, reflecting the regulatory role of this metabolite in the glucose uptake as described elsewhere [95]. Also, the UDP-glucose level was higher during the batch phase and, after 24 hours of chemostat cultivation, reached a lower stable value showing a different regulation of storage metabolism during the two cultivation conditions.

It is remarkable that most of the measured central metabolites show no or relatively little change during the transition between batch cultivation on glycerol and chemostat cultivation on glucose as sole carbon source. One reason could be that the decrease in the growth rate during the transition is relatively small, i.e. from  $\mu = 0.17$  to  $0.1 \text{ h}^{-1}$ .

Recently, a study was performed comprising a total of 32 growth conditions of *S. cerevisiae*, covering a range of growth rates from  $0.02$  to  $0.38 \text{ h}^{-1}$  [96]. The complete set of conditions spanned a large range of metabolic fluxes with, as the median for 27 reactions investigated, a 35 fold change. The associated changes in intracellular metabolite levels for these highly different flux profiles were, however, much smaller, with, as the median, a maximum change of 3.5 fold. This shows how tightly metabolism is regulated to keep the metabolite levels between narrow regions (homeostasis).

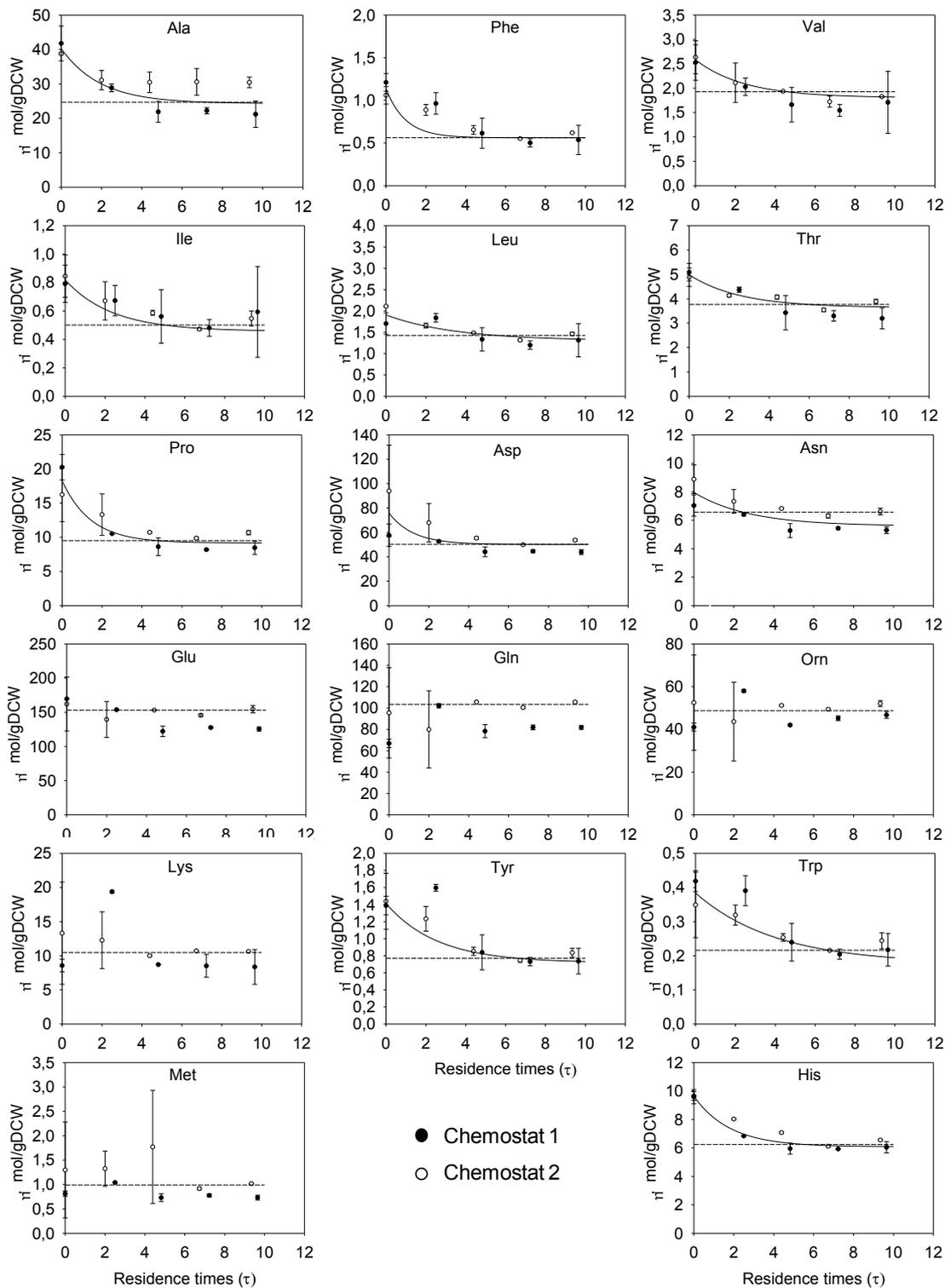


Figure 4.3: Evolution changes in intracellular pools associated with adaptation to glucose-limited conditions in two chemostats. The line represent the exponential decay curve fitted and dashed lines the weighted average of the last five residence times.

The evolution of the free amino acid pools are shown in Figure 4.3. The observation that the levels of many free amino acids decreased to lower levels could be related to the change in growth rate between

the batch phase ( $\mu = 0.17 \text{ h}^{-1}$ ) and the chemostat phase ( $\mu = 0.1 \text{ h}^{-1}$ ). This relation is strengthened when considering that usually the total protein content in the biomass increases with the growth rate, being higher at higher growth rates [97, 98] and therefore requiring higher protein synthesis rates. Moreover, a recent genomic-scale metabolic network reconstruction for *P. pastoris* showed higher biosynthetic amino acid flux requirements when the cells were grown on glycerol compared to glucose [80]. Remarkably, the amino acids belonging to the glutamate family, the largest intracellular amino acid pool, did not show any significant changes during 10 residence times of chemostat cultivation, which could also be the result of lack of variation in terms of pool sizes between the two conditions or to a fast rate of adaptation to the new growth conditions. On the other hand, for the other amino acids, the fitted curve could adequately describe the pool trends highlighting that, for most of them, 5 residence times were sufficient to achieve the steady-state values considering the experimental error. In these amino acids, the steady-state values of the fitted curves ( $C_{i,ss}$  parameter in Equation 4.3) were compared to the values from the weighted average of the last 5 residence times, showing less than 10% difference between two calculations except for Trp and Asn with a 17% and 15% of difference, respectively. These results were considered satisfactory considering the analytical challenge involved (See Appendix for detailed data).

**Table 4.3:** Physiological parameters of *P. pastoris* and *S. cerevisiae* in aerobic carbon-limited chemostats at a D of  $0.1 \text{ h}^{-1}$ .

	$Y_{sx}^a$ gDCW/C-mol glc	$q_{glc}$	$q_{o_2}$ mmol/(gDCW·h)	$q_{CO_2}$	$RQ^b$
<i>P. pastoris</i>	16.90	0.97	2.11	2.19	1.04
<i>S. cerevisiae</i> *	14.80	1.10	2.70	2.87	1.06

<sup>a</sup> Yield of biomass (gDCW/C-mol glc consumed)

<sup>b</sup> Respiratory Coefficient ( $q_{CO_2}/q_{O_2}$ )

\* Data taken from [33]

### ***Interspecies quantitative metabolome comparison***

By using the optimized quenching protocol, an accurate determination of intracellular and extracellular metabolite amounts were determined in *P. pastoris* growing in aerobic carbon-limited chemostat cultures at a dilution rate of  $0.1 \text{ h}^{-1}$ . These values were then compared with the pool sizes found in *S. cerevisiae* grown at equal culture conditions [33]. In **Table 4.3**, a summary of the rates obtained for both yeast are represented indicating higher  $q_{CO_2}$ ,  $q_{O_2}$  and  $q_{glc}$  rates in *S. cerevisiae* which lead to a lower biomass yield compared to *P. pastoris*.

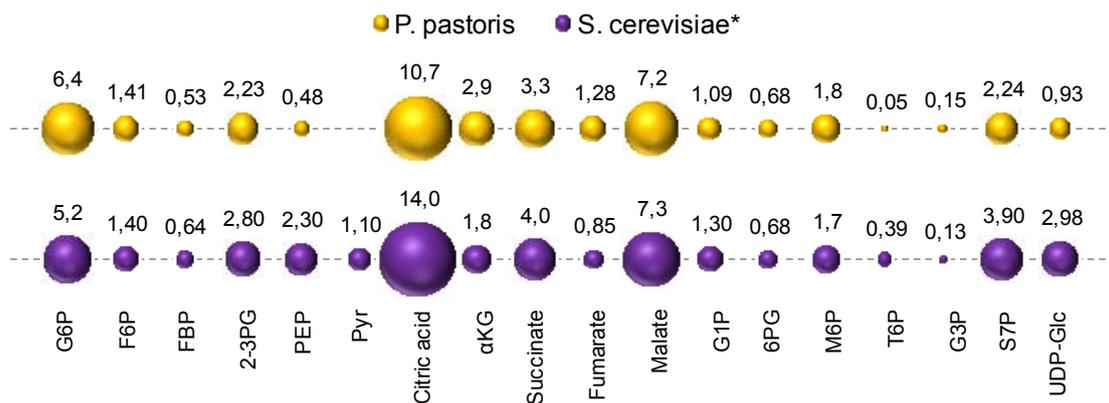


Figure 4.4: **Intracellular metabolite comparison of *P. pastoris* and *S. cerevisiae*.** The bubble areas are proportional to the pool sizes of each intracellular metabolite. The numbers indicate the exact pool size in  $\mu\text{mol/gDCW}$ . \* Data taken from [33]

#### Intracellular metabolite pools

When comparing the intracellular metabolite pools of *P. pastoris* and *S. cerevisiae* (Figure 4.4), no significant differences in the profiles could be seen in upper glycolysis. Nevertheless, in the lower part of glycolysis, the 2PG/3PG and PEP pools were lower in *P. pastoris* indicating possible differences in the thermodynamic behaviour among the yeasts. The metabolite levels of the TCA cycle were similar whereby the citric acid and malate were the most abundant in both microorganisms. Also for 6PG, G3P and M6P the levels were similar. However, the T6P pool in *S. cerevisiae* was one order of magnitude larger than in *P. pastoris*. As it is known that this metabolite has an inhibitory effect on the hexokinases [95] a lower glucose consumption rate could in principle be expected in *S. cerevisiae* if the capacity of hexokinase is similar in both organisms. However, as the maximum glucose uptake rate is higher in *S. cerevisiae*, i.e.  $20.2 \text{ mmol/gDCW}\cdot\text{h}$  [99], compared to *P. pastoris*, i.e.  $2.88 \text{ mmol/gDCW}\cdot\text{h}$  [79], it seems logical that the T6P level is higher in the former microorganism to get similar glucose consumption rates, assuming similar regulation in both yeasts.

Moreover, the mass action ratios (MAR) of enzymes from the central carbon metabolism which are expected to operate close to equilibrium (i.e. PGI, PGM, PMI, ENO and FMH) were calculated (Table 4.4) giving more information about the thermodynamic properties of these enzymes in both yeasts. Except for enolase, no significant differences could be observed in the calculated MAR's for these enzymes for both yeasts and they were all close to the equilibrium constant. However, the calculated MAR of enolase for *P. pastoris* was significantly lower than the equilibrium constant indicating a lower capacity of enolase in *P. pastoris* compared to *S. cerevisiae*. A comparable low value of the MAR of enolase in *S. cerevisiae* was measured at a much higher growth rate of  $0.33 \text{ h}^{-1}$  [96].

**Table 4.4:** Mass action ratios of some enzymes related to the central carbon metabolism of *S. cerevisiae* and *P. pastoris*. *S. cerevisiae* data were derived from [33]. The apparent *in vivo*  $k_{eq}$  were taken from [96].

Enzyme	Mass action ratios	This study	Canelas et al. 2008	Canelas et al., 2011
		<i>P. pastoris</i>	<i>S. cerevisiae</i>	<i>in vivo</i> $k_{eq}$
PGI	F6P/G6P	0.22 ± 0.11	0.26 ± 0.05	0.259 ± 0.002
PGM	G1P/G6P	0.13 ± 0.09	0.05 ± 0.02	0.063 ± 0.004
PMI	M6P/F6P	1.27 ± 0.06	1.17 ± 0.03	1.183 ± 0.013
ENO	PEP/2-3PG**	1.67 ± 0.10	3.99 ± 0.13	4.01 ± 0.09
FMH	Malate/Fumarate	5.25 ± 0.38	4.82 ± 0.12	5.15 ± 0.14

\*\* The 2PG total amounts were calculated assuming that 3PG and 2PG were in equilibrium ( $k_{eq}= 0.1$ ).

#### Estimated turnover times of central metabolites

When using the accurate determination of intracellular metabolite levels combined with the metabolic flux data from *P. pastoris* grown under analogous conditions [52] an estimation of the turnover times for these metabolites can be calculated (Figure 4.5). These turnover times are known to be an overestimation of the real turnovers inside the cell due to the usage of net fluxes instead of forward and reverse fluxes. However, even assuming that, it is interesting to see that there are already values in the order of seconds or less (FBP, PEP and 2-3PG) which highlight the importance of the rapid sampling and optimized quenching to obtain the most accurate quantification. It can be seen from Figure 4.5 that the turnover times of the central metabolite pools of *P. pastoris* and *S. cerevisiae* cultivated in glucose limited chemostat under the same conditions have the same profiles, showing smaller turnover times for the intermediates of the glycolysis pathway compared to the TCA cycle.

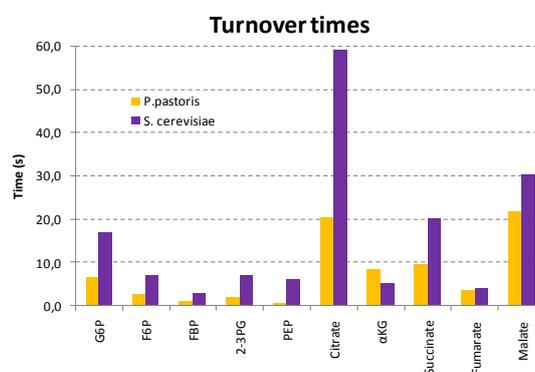


Figure 4.5: Intracellular turnover times of the central metabolism metabolites of *P. pastoris* and *S. cerevisiae*. *S. cerevisiae* data taken from [33].

#### Extracellular metabolite pools

In addition to intracellular metabolite amounts, the extracellular metabolite levels (Figure 4.6) were compared for both yeasts. In this case, the *P. pastoris* culture filtrate samples were taken from the quenching experiment under validated metabolic steady state conditions. Interestingly, in *P.pastoris* chemostat cultivations much lower extracellular metabolite levels were observed, i.e. in total 4.9

$\mu\text{mol/gDCW}$  were found in *P. pastoris*, compared to the  $41.8 \mu\text{mol/gDCW}$  for *S. cerevisiae*. In Figure 4.6, the extracellular amounts are expressed for each metabolite as percentage of the whole broth sample amount. It can be seen from this figure that for all measurements metabolites the extracellular amounts are much lower in *P. pastoris* compared to *S. cerevisiae* and that in the latter the majority of the metabolites present outside were intermediates of TCA cycle and G3P.

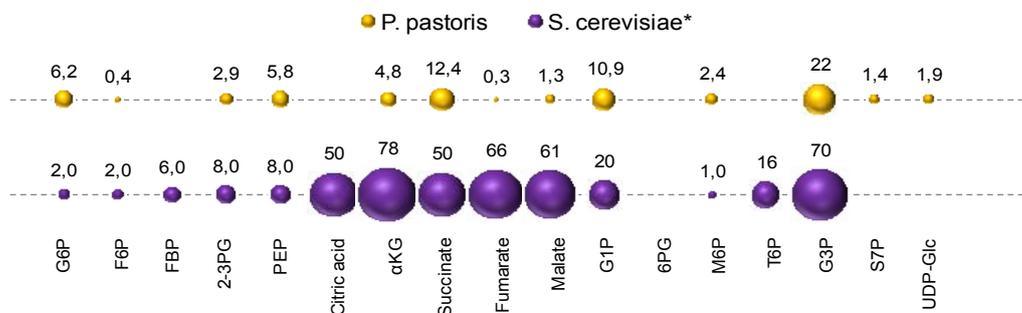


Figure 4.6: **Extracellular levels of central metabolites in glucose limited cultures of *P. pastoris* and *S. cerevisiae*.** The bubble areas represent the percentage of the total pool which is present in the culture filtrate. The numbers above each the bubbles represent the exact value. The metabolites without value could not be detected. \**S. cerevisiae* data taken from [33].

#### Intracellular amino acid pools

The free amino acid pools measured for *P. pastoris* were compared with those previously published for *S. cerevisiae* under the same conditions (Figure 4.7). Overall, amino acid pools sizes seem to follow similar trends in both microorganisms, with Glu, Ala and Asp being the major amino acids. However, in *P. pastoris* the Gln, Orn and Lys pools were larger than in *S. cerevisiae* indicating differences in the cell physiology resulting in higher accumulation levels of the amino acids derived from  $\alpha$ KG. Moreover, Val and Met amounts were one order of magnitude higher in *S. cerevisiae*, even though the protein production demand of these amino acids was similar for both yeasts (based on amino acid composition of the biomass protein content taken from for *P. pastoris* [76] and for *S. cerevisiae* [32]).

In order to have a general view of the amino acid distribution, the pool sizes of all the amino acids with the same precursor were combined (data not shown). Interestingly, all amino acid families were within similar range in both strains except for the glutamate family which was higher in *P. pastoris* as mentioned before which leads to a larger intracellular amino acid content in *P. pastoris* ( $425 \mu\text{mol/gDCW}$ ) compared to *S. cerevisiae* ( $340 \mu\text{mol/gDCW}$ ).

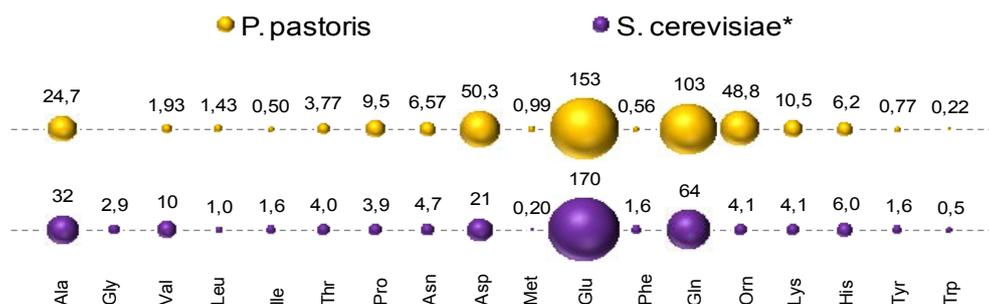


Figure 4.7: **Intracellular amino acid comparison of *P. pastoris* and *S. cerevisiae*.** The bubble area represents the proportional pool sizes of each intracellular metabolite. The number above each bubble is the exact pool size in  $\mu\text{mol/gDCW}$ . \**S. cerevisiae* data taken from [33].

### Extracellular amino acid pools

In the same way as for the extracellular levels of the central metabolites, the extracellular amino acids levels were compared for both yeasts. As was found for the central metabolites, also the total level of extracellular amino acids in *P. pastoris* chemostat cultivations was much lower (total extracellular amino acid pool of  $0.7 \mu\text{mol/gDCW}$ ) compared to *S. cerevisiae* ( $13.7 \mu\text{mol/gDCW}$ ). In Figure 4.8, the extracellular amounts for each amino acid are expressed as percentage of the whole broth sample amount. It can be concluded from the low extracellular metabolite amounts in *P. pastoris* that total broth extraction would be a valid alternative for the cold filtration method because for the majority of the metabolites measured, the extracellular amounts are too low to interfere significantly with the intracellular measurement. Removal of the extracellular medium, e.g. by cold filtration or cold centrifugation) would then only be required if constituents of the medium would interfere with the analysis method applied. Furthermore, these low extracellular metabolite levels make *P. pastoris* attractive as a cell factory because a less contaminated broth facilitates the downstream processing.

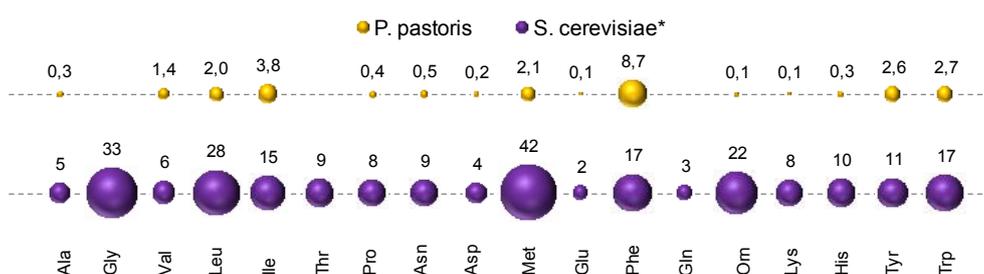


Figure 4.8: **Extracellular amino acid comparison of *P. pastoris* and *S. cerevisiae*.** The bubble areas are proportional to % Extracellular pools of each intracellular metabolite. The numbers over the bubbles are the exact % value. The metabolites without value could not be identified extracellularly. \**S. cerevisiae* data taken from [33].

## Conclusions

The aim of this study was to perform a systematic investigation of a cold methanol quenching method, combined with filtration for removal of extracellular metabolites, specifically optimised for *P. pastoris*,

thereby providing a validated technique for metabolomics studies of this yeast. Moreover, this methodology has been successfully applied to investigate the evolution of *P. pastoris* cells after a shift from batch to continuous operation mode under well-defined and controlled bioreactor cultivations. Notably, the obtained data has allowed performing a vis-a-vis comparison of *P. pastoris* and *S. cerevisiae* metabolomes, revealing similar profiles except for the lower glycolysis.

Overall, the outcome of this study will provide a starting point for the microbial metabolomics community future studies related to systems biology and systems metabolic engineering of *P. pastoris*.

## Appendix

**Appendix 4.1:** Specific metabolite consistency index for each protocol tested. Values above 3.84 are an indicator of gross measurement errors with a confidence level of 95%. Higher h values are highlighted with increasing colour intensities.

Protocol	h-index				
	A	B	C	D	E
G6P	8.2	1.5	7.7	4.1	1.6
F6P	4.4	0.8	0.1	0.0	1.8
2-3PG	3.4	0.1	5.4	0.0	0.2
PEP	1.5	5.4	1.4	2.7	3.1
Pyruvate	7.8	25.8	16.3	12.6	9.0
aKG	2.8	0.0	11.1	0.0	2.9
Succinate	0.2	4.6	0.2	0.8	1.9
Fumarate	9.7	6.4	1.9	0.6	1.0
Malate	0.1	0.1	0.1	0.1	0.2
G1P	2.8	1.9	1.5	0.8	23.0
6PG	0.4	0.6	0.0	0.0	0.2
M6P	3.2	1.4	3.9	8.7	4.7
G3P	0.4	0.3	4.2	11.2	0.0
S7P	0.0	0.2	0.8	1.7	0.0
UDP-glc	0.7	0.4	0.0	0.0	0.3
Ala	0.5	1.3	0.0	0.0	1.6
Gly	9.0	6.5	14.2	5.4	4.0
Val	2.9	4.0	0.0	0.8	2.3
Leu	2.1	1.4	0.0	5.8	2.7
Ile	3.3	2.1	0.6	3.6	3.3
Thr	0.1	0.8	0.2	0.4	0.2
Pro	0.0	0.1	1.1	2.8	0.1
Asn	0.1	0.1	0.9	2.1	0.1
Asp	2.2	0.6	2.1	3.1	0.0
Met	0.8	0.0	0.0	2.4	0.2
Glu	0.0	0.2	0.6	2.0	0.0
Phe	2.1	3.0	7.0	24.0	8.5
Gln	0.0	0.0	1.1	1.2	0.1
Orn	0.0	0.2	0.3	1.6	0.4
Lys	0.0	0.0	0.4	1.4	0.1
His	0.2	0.4	0.2	1.2	1.1
Tyr	1.3	0.3	0.0	0.1	1.7
Trp	3.7	0.1	0.1	1.5	1.4



## Chapter 5: *Pichia pastoris* stoichiometric network improvement<sup>d</sup>

---

<sup>d</sup> Manuscript in preparation as: Marc Carnicer, Angela ten Pierick, Zhen Zeng, Reza M. Seifar, Jan van Dam, Joseph J. Heijnen, Pau Ferrer, Walter van Gulik and Joan Albiol. **Oxygen impact on *Pichia pastoris* metabolome: a thermodynamic study.**



## Background

In a recent multilevel study, including transcriptome, proteome and metabolic flux analyses, information on the biological processes involved in the adaptation to low oxygen availability conditions, in *P. pastoris* have been obtained [100]. Under these conditions, the proper characterization of the pentose phosphate pathway (PPP) appeared as a key point not only due to its role as provider of cytosolic reduction equivalents for biosynthesis (NADP<sup>+</sup>/NADPH balance) but also due to the arabinitol production as a fermentative by-product.

This part of the metabolism is classically divided in two parts. The oxidative PPP produces three pentose phosphate metabolites and the non-oxidative PPP which basically interconverts PPP intermediates with F6P and GAP. The stoichiometry of the oxidative part is well established. However, the transketolase (TK) and transaldolase (TA) reactions of the non-oxidative PPP have been the subject of a number of considerations, of particular relevance when considering the precise <sup>13</sup>C-label distribution in metabolic studies [25, 101, 102].

In the present study, a wide range metabolomic analysis of different intracellular metabolites was performed in carbon limited chemostat cultivations at a fixed growth rate and different oxygenation conditions. These intracellular metabolite quantifications were analyzed using network-embedded thermodynamic (NET) calculations [103] using the public available anNET tool [104]. This allowed to validate the experimental data, obtain further metabolic fluxes constrains as well as improvements in the metabolome derived considerations. In a second step, three non-classical stoichiometric networks, that were found to agree with the thermodynamic flux constrains, were used to perform metabolic flux analysis showing different metabolic flux profiles.

## Materials and Methods

### ***Strain and cultivation conditions***

Analytical grade reagents were supplied by Sigma-Aldrich. HPLC-grade methanol and ethanol were supplied by J.T. Baker.

In this study, the *P. pastoris* X-33 (wild type phenotype, Invitrogen) transformed with pGAP $\alpha$ A (Invitrogen) as a mock vector was used. Glucose-limited chemostat cultivations at a dilution rate (D) of 0.1 h<sup>-1</sup> were performed as described in [105] using also the same culture media composition. In order to have biological replicates, two chemostat cultures for each oxygen condition were performed. Briefly, one liter shake flask containing 300 ml of YPD medium (10 g/l yeast extract, 20 g/l peptone, 10 g/l glucose) was inoculated with a 1.0 ml of cryostock *P. pastoris* cells. The culture was grown for approximately 24 h at 30 °C in a shaker at at 200 rpm, and used to inoculate a 7-liter fermentor (Applikon, The Netherlands) with a working volume of 4 liter. Once the batch phase ended,

approximately 24 h after performing the inoculation, the feed of the chemostat culture was started. During the chemostat cultivation phase, cells were grown under carbon-limited conditions at a dilution rate (D) of  $0.1 \text{ h}^{-1}$ , and an aeration rate of 0.5 vvm, controlled by two mass flow meters (5850 Smart Mass Flow Controller, Brooks Instrument). Initially the oxygen concentration in the inlet gas stream corresponding to normal air (therefore approximately 21 % v/v) leading to a totally normoxic condition (i.e.  $p\text{O}_2 > 20 \%$ , fully respiratory metabolism). Inlet gas oxygen levels were subsequently stepwise reduced by replacing different air proportions with nitrogen. Thereby creating either oxygen limited or hypoxic conditions in the bioreactor which are characterized by different ethanol and arabinitol production rates [47, 52, 76]. The  $\text{O}_2$  and  $\text{CO}_2$  concentrations in the bioreactor off-gas were measured on-line using a combined paramagnetic/infrared analyzer (NGA 2000, Rosemount, USA). Pressure, pH, stirring speed and temperature were respectively maintained constant at 1.2 bars, pH 5 (with 20%  $\text{NH}_4^+$ ), 400 rpm and 25 °C.

On the different cell cultures performed in this study, the experimental data consistency was verified applying standard procedures to elemental conservation relations as constraints [38]. For all chemostat cultivations, the statistical consistency test was passed at 95% confidence level, indicating that there was no proof for gross measurement errors.

### **Sampling**

The different chemostat experimental conditions were maintained for 5 residence times before sampling. As previously reported, this cultivation time is enough to reach a metabolic steady state in *P. pastoris* [105]. For each steady state condition, duplicate samples, for intracellular metabolite levels determination, were taken using the previously described optimized protocol for the direct measurement of *P. pastoris* metabolome [105]. Briefly, quenched/washed cells were taken using a dedicated rapid-sampling setup [32]. The sample tubes were quickly mixed by vortexing and introduced in the filtration unit after weighting the tube [90]. All sampling tubes were weighted before and after the sampling procedure in order to determine the exact amount of sample (on average  $0.51 \pm 0.04 \text{ g}$ ). The cell suspensions were filtered with membrane disk filters (Pall Corporation, East Hills, NY, USA, 47mm diameter, 0.45  $\mu\text{m}$  pore size) using a vacuum pump. A washing step was performed to remove as much extracellular metabolites as possible pouring 10 ml of fresh quenching solution (60 % v/v of Methanol/Unbuffered Water at -27 °C) on the filter cake as soon as the biomass started to dry.

For accurate quantification purposes by means of Isotope Dilution Mass Spectrometry [29, 91], 120  $\mu\text{l}$  of a  $^{13}\text{C}$  internal standard solution (0°C) was pipetted on top of the dry filter cake. The  $^{13}\text{C}$  internal standard solution contained all relevant metabolites as U- $^{13}\text{C}$  -labeled isotopes and was obtained from a *S. cerevisiae* fed-batch culture grown on 100 % U- $^{13}\text{C}$  -labeled glucose and ethanol. Metabolite extraction with 75 % (v/v) aqueous ethanol at 95 °C and further sample processing was carried out as described in [90].

Samples from the culture filtrate (CF) and the complete culture broth (WB) were withdrawn and further processed as previously described [33, 105].

### **Metabolite analysis**

The concentrations of the principal metabolites from *P. pastoris* metabolome were quantified using two different analytical platforms, LC-MS/MS and GC-MS. G6P, F6P, FBP, 2PG+3PG, PEP, citrate,  $\alpha$ KG, succinate, fumarate, malate, G1P, 6PG, M6P, T6P, G3P S7P and UDP-glc were determined by anion-exchange LC-MS/MS [30]. AMP, ADP and ATP were determined by ion-pair reverse-phase LC-MS/MS [106]. Intracellular glucose, 3PG, 2PG, DHAP, R5P, Rbu5P, Xu5P, E4P, Trehalose and isocitrate were determined by GC-MS [107]. Quantification was based on isotope dilution mass spectrometry (IDMS) [29]. In total, 33 metabolites, with a wide variety of chemical and physical properties, were analyzed. However, only 30 of these could actually be determined because the concentrations of glyoxylate, mannitol-1-phosphate and fructose-2,6-bisphosphate were found to be below the detection limit in all the analyzed samples. Moreover, it is worth to be noticed that the 2PG and 3PG were quantified together by LC-MS but also separately by GC-MS. Both data sets were included in the comparison.

### **Network-embedded thermodynamic analysis**

All reactions are constrained by the second law of thermodynamics, where the advance of the reaction is driven by a negative change of the Gibbs free energy. This reaction property has been used as a method to verify the quality of the intracellular metabolite concentration thereby improving the consistency of the metabolome data sets [103]. In addition, it was possible to predict the levels of some unmeasured metabolites. This increased the amount of information obtained from the studied system [103]. All calculations were performed using anNET software developed by Zamboni and co-workers [104].

Formally, the software uses Molar units for the intracellular metabolite data. As the experimental data was obtained in a dry weight basis, it was necessary to convert the intracellular metabolite determinations from  $\mu\text{mol/gDCW}$  to concentrations units. To this purpose, different values of cell volume were tested due to the lack of information of its real value for *P. pastoris* in the experimental conditions tested. The different cell volumes were taken from *S. cerevisiae* studies ranging from 1.4 to 2.0 ml/gDCW [96]. However, the thermodynamic derived constraints obtained were equivalent, independently of the cell volume (data not shown). Therefore, the cell volume was fixed to 1.7 ml/gDCW to perform all NET analysis. This value was also used to re-calculate the  $\mu\text{mol/gDCW}$  from the refined intracellular metabolite concentrations obtained in the analysis for comparison purposes.

Cytosol was the only cellular compartment considered in the analysis as no information on the specific volumes of the different cellular compartments was available. Thus, the metabolic pathways included in the NET analysis and the intracellular metabolites were considered to be exclusively located in this

compartment. This allowed a direct comparison between them (see Appendix for more detailed information about the metabolism pathways included).

The feasible ranges of the quantified intracellular metabolites were calculated using the respective experimental mean and standard deviation. The lower and upper limits were defined according to a Student-T distribution with a confidence interval of 0.80. In addition, the unmeasured metabolites were specifically constrained to fulfill the thermodynamic constrains [96, 103]. Specifically, for GAP and  $P_i$  it was possible to define narrower window ranges taking into account the available ranges measured in *S. cerevisiae* [96]. For the rest of metabolites a wide range of concentrations was delimited (See Appendix for detailed information about the unmeasured metabolite ranges).

For the *P. pastoris* cytosol, the assumed ionic strength was 0.15 M while the pH was assumed to be 7.2 (D. Mattanovich personal communication). This value is slightly high compared to the one of *S. cerevisiae* [108]. Nevertheless, the calculations were performed also at pH 7.0 and 7.5 globally obtaining equivalent results. Only the  $NAD^+/NADH$  minimum ratio was expected to be highly influenced by this parameter (see results and discussions section).

### **Metabolic Flux Analysis (MFA)**

Intracellular flux distributions were calculated from the measured specific production and consumption rates (see Results section) using global MFA approach described for *P. pastoris* [52] (Chapter 3). In total, in this study 4 different stoichiometric networks were used to perform MFA incorporating the thermodynamic restrictions derived from the NET analysis (Appendix). The MFA-1 was the metabolic flux profiling of *P. pastoris* already published by Baumann and co-workers included for comparison purposes [52](MFA from Chapter 3). In the MFA-2, constrains derived from the  $^{13}C$ -NMR data in previous study were replaced by two global redox balances of NADH and NADPH together with  $O_2$  and  $CO_2$  stoichiometry. In addition, a global ATP balance (taking into account maintenance coefficient. See below) was included in the stoichiometry.

The energetic biomass requirements were determined using the biomass specific macromolecular data for *P. pastoris* [76] together with the ATP specific costs for polymerization of each macromolecular [109]. Moreover, ATP requirements for maintenance were included as a constant ATP consumption reaction, not linked to growth rate, proportional to a constant coefficient. This value was taken from two *P. pastoris* genome scale metabolic network analysis and was considered to be constant independently of the culture conditions [79, 80]. The metabolite transport rates were also included (metals,  $NH_4^+$ , and  $SO_4^{2-}$  consumption) taken from the standard data reconciliation procedures of the specific consumption and production rates [38]. The ATP synthase, the main source of ATP production under normoxic conditions, does not have specific stoichiometry as its yield is linked to the proton gradient between the matrix and the mitochondrial inter-membrane space. This relationship is usually

described by a P/O ratio coefficient, which may change under different conditions. Therefore, in the MFA, the ATP synthase reaction was only constrained by the ATP balance indicating the amount of ATP which needed to be synthesized to close the balance. Once calculated, the physiological feasibility of the ATP synthase rates were validated by calculating the resulting P/O ratios of each MFA.

**Table 5.1:** Summary of non-oxidative pentose phosphate pathway reactions used in the different MFAs

		MFA-1	MFA-2	MFA-3	MFA-4	MFA-5
TK(1)	$X5P + R5P \leftrightarrow S7P + GAP$	x	x		x	x
TA	$S7P + GAP \leftrightarrow F6P + E4P$	x	x			
TK(2)	$X5P + E4P \leftrightarrow F6P + GAP$	x	x	x		
TK(1)+TA	$X5P + R5P \leftrightarrow F6P + E4P$			x	x	x
TK(3)	$E4P + S7P \leftrightarrow F6P + R5P$				x	x
TAr	$F6P + E4P \leftrightarrow S7P + GAP$					x

In the MFA-3, MFA-4 and MFA-5, the non-oxidative pentose phosphate pathway reactions included in MFA-2 were modified according to the thermodynamic flux constrains derived from the NET analysis. In Table 5.1 the specific reactions of the non-oxidative PPP considered in each calculation are shown. In the MFA-3 and MFA-4 no further constraints were needed. For the MFA-5, a flux ratio constrain derived from  $^{13}\text{C}$ -NMR analysis from equivalent metabolic steady states was included. See appendix for detailed information on the implementation of these ratios as constrains.

## Results and Discussion

### Chemostat cultivations

The experimental balanced rates of the different chemostat cultures performed in this study are shown in Table 5.2.

**Table 5.2:** Overview of the macroscopic growth parameters

	Normoxic	O <sub>2</sub> -limited	Hypoxic
$q_{\text{Glc}}^a$	$-1.00 \pm 0.02$	$-1.28 \pm 0.03$	$-1.72 \pm 0.05$
$q_{\text{O}_2}^a$	$-2.35 \pm 0.06$	$-2.01 \pm 0.07$	$-2.01 \pm 0.15$
$q_{\text{CO}_2}^a$	$2.43 \pm 0.06$	$2.55 \pm 0.06$	$3.21 \pm 0.14$
$q_X^a$	$3.57 \pm 0.15$	$3.83 \pm 0.18$	$3.77 \pm 0.23$
$q_{\text{EtOH}}^a$		$0.31 \pm 0.02$	$0.84 \pm 0.06$
$q_{\text{Ara}}^a$		$0.13 \pm 0.01$	$0.33 \pm 0.01$
$RQ^b$	$1.03 \pm 0.04$	$1.27 \pm 0.05$	$1.60 \pm 0.14$

<sup>a</sup> mmol/(g Biomass·h)

<sup>b</sup> mol CO<sub>2</sub>/mol O<sub>2</sub>

The substrate and product rates obtained at normoxic and O<sub>2</sub>-limited conditions showed no significant differences compared to the corresponding rates previously reported for *P. pastoris* in equivalent experimental conditions [52](Chapter 3), thereby reflecting equivalent metabolic states. Conversely, the

hypoxic condition appeared to be less stringent compared with the corresponding cultivation condition previously reported, as denoted by lower RQ ( $RQ > 3.0$  in [52]), ethanol and arabinitol production rates. One possible explanation for the differences in hypoxia could be that in the previous study, only 3.5 residence times could be accomplished in the hypoxic conditions [47], whereas in this study the hypoxic condition could be extended up to a minimum of 5 residence times, which is the period required to reach a true bioreactor (and therefore metabolomic) steady state [105].

### ***Thermodynamic revision of the stoichiometric network***

The intracellular metabolite quantifications are constrained by the second law of thermodynamics allowing specific directions of the metabolic fluxes. In this study, an initial NET analysis was performed using the stoichiometric network and flux directions taken from the already published calculations for *P. pastoris* under the same conditions [52]. In the initial calculations, when all the reactions were directionally constrained the system resulted to be unfeasible. In a second step, using the troubleshooting options of the software, the inconsistency between the measured metabolites and the flux directions was located in the non-oxidative part of the pentose phosphate pathway. Thereby, validating the measures of all the rest intracellular metabolites included in the NET model.

**Table 5.3:** Transformed Gibbs Energy of the non-oxidative pentose phosphate pathway reactions of *P. pastoris* under different oxygen conditions using fixed glycolysis reaction directions.

Transformed Gibbs Energy of reactions (KJ/mol)		Normoxia		Oxygen-limited		Hypoxia	
		min	max	min	max	min	max
RPI	$RU5P \leftrightarrow R5P$	1.0	2.4	-0.2	1.7	-2.2	2.9
RPE	$RU5P \leftrightarrow X5P$	-0.6	0.6	-0.5	1.5	-5.9	-2.9
TK(1)	$X5P + R5P \leftrightarrow S7P + GAP$	-12.0	-5.2	-17.9	-7.9	-10.9	-1.8
TA	$S7P + GAP \leftrightarrow F6P + E4P$	3.4	9.3	2.0	11.3	0.3	9.4
TK(2)	$X5P + E4P \leftrightarrow F6P + GAP$	-26.6	-17.9	-31.5	-18.6	-27.1	-11.7
TK(1)+TA	$X5P + R5P \leftrightarrow F6P + E4P$	-3.3	-1.1	-8.0	-4.5	-4.2	1.2
TK(3)	$E4P + S7P \leftrightarrow F6P + R5P$	-14.6	-12.7	-13.6	-10.8	-16.2	-9.9
TAr	$F6P + E4P \leftrightarrow S7P + GAP$	-9.3	-3.4	-11.3	-2.0	-9.4	-0.3

In this stoichiometry, the non-oxidative pentose phosphate pathway was implemented as classically reported in any biochemical textbook. That is including the Ribulose 5-phosphate epimerase (RPE) and the Ribulose 5-phosphate isomerase (RPI) reactions followed by two transketolases, TK(1) and TK(2), and a transaldolase (TA). In this part of the model, the performed NET analysis showed a disagreement between the intracellular metabolite pools of the pathway and the metabolic flux directions when only the glycolysis reactions were directionally constrained (Table 3). Specifically, RPI in normoxia, and TA in all the conditions, could not operate in the direction of formation R5P, F6P and E4P respectively. These were the expected directions in the previously published MFA analysis [52].

### Metabolic channeling

In the NET analysis, TA reaction inconsistency could be related to significant measurement error in its intracellular metabolites intermediates. Nevertheless, all the metabolites were analyzed from the same sample tubes (see Materials and Methods section). Thereby, reducing the possible measurement deviation of the pentose pathway intermediates although, a specific drift in one metabolite quantification could not be discarded. Besides, for the *S. cerevisiae* strain IMS0001, a thermodynamic inconsistency was also determined for the TA reaction [110] indicating that these are not isolated results. Specifically, in that study the TA showed a higher value than the equilibrium constant under glucose-limited chemostat conditions.

However, according the results obtained in this work, the metabolite channeling between TK(1) and TA reactions could be a possible explanation of the obtained intracellular metabolite flux analysis, indicating a micro-compartmentation of the reaction intermediates (Table 3). To overcome this limitation, a TK(1)+TA reaction was included in the NET analysis. The results showed that the resulting reaction was feasible towards the formation of F6P and GAP in all the cases. Nevertheless, the TK(1) and TA reverse reactions are expected to have an influence on the carbon distribution (Table 3) as indicated by the <sup>13</sup>C-label distribution found in other studies [101]. Moreover, the metabolite channeling also was determined for *S. cerevisiae* (CEN.PK 113-7D) grown on glucose showing similar Gibbs free energy values [111].

### R5P heterogeneity distribution

For RPI, the reversed flux direction in the normoxia conditions would lead to an inconsistency on the R5P balance as this reaction was the only reaction included in the MFA-1,-2 and -3 that could produce this metabolite. This disagreement was also found in other metabolome data sets such as in *E. coli* [104] or *S. cerevisiae* [110]. In both metabolome data sets, it was assumed an erroneous metabolite quantification of the R5P and/or RU5P as the most plausible explanation.

Nevertheless, considering the fact that three independent studies have reached the same inconsistency, it indicates that other possible explanatory alternatives should be assayed, besides considering the discrepancy an experimental error. Looking at the two metabolites involved, the RU5P could be present inside the cell in two isomeric forms, L-RU5P and D-RU5P, which the GC-MS analysis cannot discriminate but perhaps the enzyme can. However, taking into account the L-RU5P in the NET analysis, would not solve the RPI inconsistency as the D-RU5P pool would become lower promoting its formation instead of R5P.

On the other hand, R5P, could not be present in two isomeric forms inside the cell as RU5P does. However, R5P plays an important role as a carbon skeleton in the RNA structure being it an important component of the biomass macromolecular composition [55, 76].

In eukaryotic cells, the RNA and DNA precursors are transported from the cytosol to the nucleus through the nucleus pores where the DNA replication and repair processes and the RNA transcription and maturation ones are performed [112]. Therefore, in *P. pastoris* and *S. cerevisiae* metabolomes, the R5P role could imply a compartmentalization of this metabolite between the nucleus and cytosol allowing a heterogenic distribution throughout the cell.

In addition, besides the fact that compartmentalization is not possible in *E. coli*, a heterogenic distribution of the RNA precursors around the DNA stands could be also originated in prokaryotic cells due to the fact that DNA is concentrated forming the nucleoid allowing a R5P concentration gradient inside the cell [112].

#### TK promiscuity

On the other hand, using *S. cerevisiae* as a model, Kleijn and co-worker [101] reviewed the possible reactions in the non-oxidative PPP. As a result, they introduced another reaction catalyzed by a transketaolase, the TK(3). This reaction was verified to be thermodynamically constrained towards the R5P and F6P formation in all the conditions (Table 3). Therefore, the incorporation of the TK(3) reaction in the MFA analysis would allow the R5P formation from the non-oxidative PPP.

This evidence is important in *P. pastoris* due to the fact that, in previous studies using the same strain under equivalent conditions, it was determined that around 70% of the intracellular R5P pool had been produced by, at least, one transketolase reaction [52](Chapter 3). Nevertheless, the thermodynamic analysis of the TK(1) reaction showed a clear tendency towards the GAP and S7P formation (Table 3) which reinforce the hypothesis of R5P recirculation through the TK(3) reaction. In addition, Baumann and co-workers [52] also determined that part of the R5P also had been through a transaldolase reaction which opened the possibility of the TAr reaction.

As a consequence of the consideration of these new thermodynamic constrains on the pentose phosphate pathway of *P. pastoris*, new metabolic flux analyses were performed (see below).

### **Metabolic Flux Analysis**

#### Redox and ATP stoichiometry validation

In previous studies using equivalent environmental conditions, the metabolic flux analysis was performed using metabolic flux constrains derived from <sup>13</sup>C-NMR measures and without including the NADH or NADPH balances in the calculations named MFA-1 in this study [52].

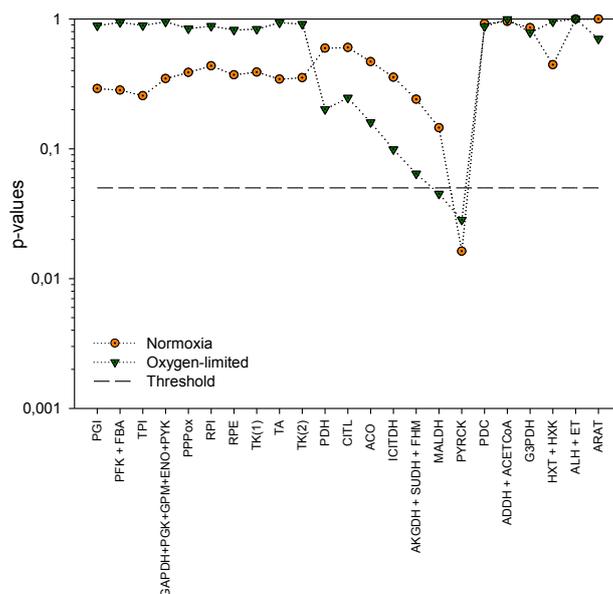


Figure 5.1: **MFA-1 and MFA-2 comparison.** Representations of the p-values according to a two tailed t-test distribution whether the different MFA calculation for each reaction was different than zero. The threshold was fixed at 0.05.

The MFA-2 was performed in order to validate the redox and the ATP stoichiometries using the MFA-1 as reference values. Figure 5.1 shows the p-values calculated while comparing the MFA-1 and MFA-2 according to a two tailed t-test distribution. These indicate whether significant differences exist in both MFA calculations for each reaction under the normoxic and oxygen-limited conditions. Notice that the biomass requirements were not compared as they were assumed to be the same under equivalent metabolic conditions. No significant difference for p-values > 0.05.

As a result, only the PYRCK presented significant differences between the two MFA being in the MFA-1 higher than the MFA-2. These results could be due to the better resolution of MFA-1 around the mitochondrial inside-out oaa transport due to the  $^{13}\text{C}$ -NMR flux ratios. Nevertheless, despite this reaction, both MFA calculations showed equivalent metabolic flux profiles. This fact in turn validates the NADH, NADPH and ATP stoichiometry used in the calculations done without  $^{13}\text{C}$ -NMR constrains. In addition, the results allowed the metabolic flux calculation of the hypoxic conditions realized in this study using the MFA-2 stoichiometry.

#### Pentose phosphate pathway stoichiometry considerations

The pentose phosphate pathway has received little attention from the MFA community with respect to model validation which is rather surprising since it plays several roles in the cell metabolism. Several previous works already mention the difficulties observed in reproducing the  $^{13}\text{C}$ -level distribution without considering a detailed enzymatic mechanism [25, 101, 102]. In the present NET analysis of the PPP, a disagreement between metabolic flux directions and the intracellular metabolite concentrations

was found (see above). Therefore, four alternative PPP stoichiometries (Table 5.1) were tested for the metabolic flux analysis in order to improve the *P. pastoris* model agreement with both requirements (Figure 5.2).

The MFA-2 represents the classical non-oxidative pentose phosphate pathway reactions which were found to be thermodynamically infeasible in *P. pastoris*. In this study, a metabolic channeling between the TK(1) and TA reactions is proposed as a possible solution leading to a equivalent metabolic flux distribution in the pentose phosphate pathway under the different oxygenation conditions (MFA-3 in Figure 5.2). However, in this alternative the S7P does not react in any non-oxidative PPP reaction indicating a too simplistic stoichiometric model.

On the other hand, in addition to the two classical TK reaction, the kinetic mechanism of this enzyme allows for another reaction stoichiometry represented as the TK(3) reaction [101] which was also thermodynamically feasible (Table 5.3). Nevertheless, only two TK reactions could be included at the same time in the MFA as they are linearly dependent ( $TK(1) + TK(3) = TK(2)$ ). In the MFA-4 alternative, the TK(2) was replaced by the TK(1), which could operate without considering any metabolite channeling (Table 5.3), and also included TK(3). The resulting flux distribution is equivalent to the previous MFA.

In the MFA-5, the previously obtained results, using  $^{13}C$ -NMR data, from the non-oxidative pentose phosphate pathway were included in the MFA analysis allowing the incorporation of the TAr reaction [52] (See appendix for more detailed information about the implementation of the flux ratio). As a result, TK(1)+TA metabolic channeling was found to be more important in the carbon flux distribution in the PP pathway than in MFA-4. For TK(3), no significant differences were found between the MFA-4 and MFA-5 flux profiles. Besides, TAr reaction was found to have an impact on the carbon distribution pattern of the PP pathway under normoxic and oxygen-limited conditions.

From a general point of view, MFA-3, MFA-4 and MFA-5 show the oxygen impact over this part of the metabolism as a result of the arabinitol secretion and the decrease in the net carbon flow through the non-oxidative PP pathway, as previously reported [52].

### ***Energetic metabolism under different oxygen conditions***

In all cases except MFA-1, the ATP global balance was incorporated in order to determine the oxygen availability impact on the energetic metabolism. In the present case, it was assumed that glucose is transported in *P. pastoris* not only by diffusion, as it is the case of *S. cerevisiae*, but also by a facilitate transport as reported to be for the Crabtree Negative yeasts (HXT reaction) [83, 113–115]. This assumption was further supported by the intracellular glucose accumulation, suggesting that the HXT could be actively regulated reaction.

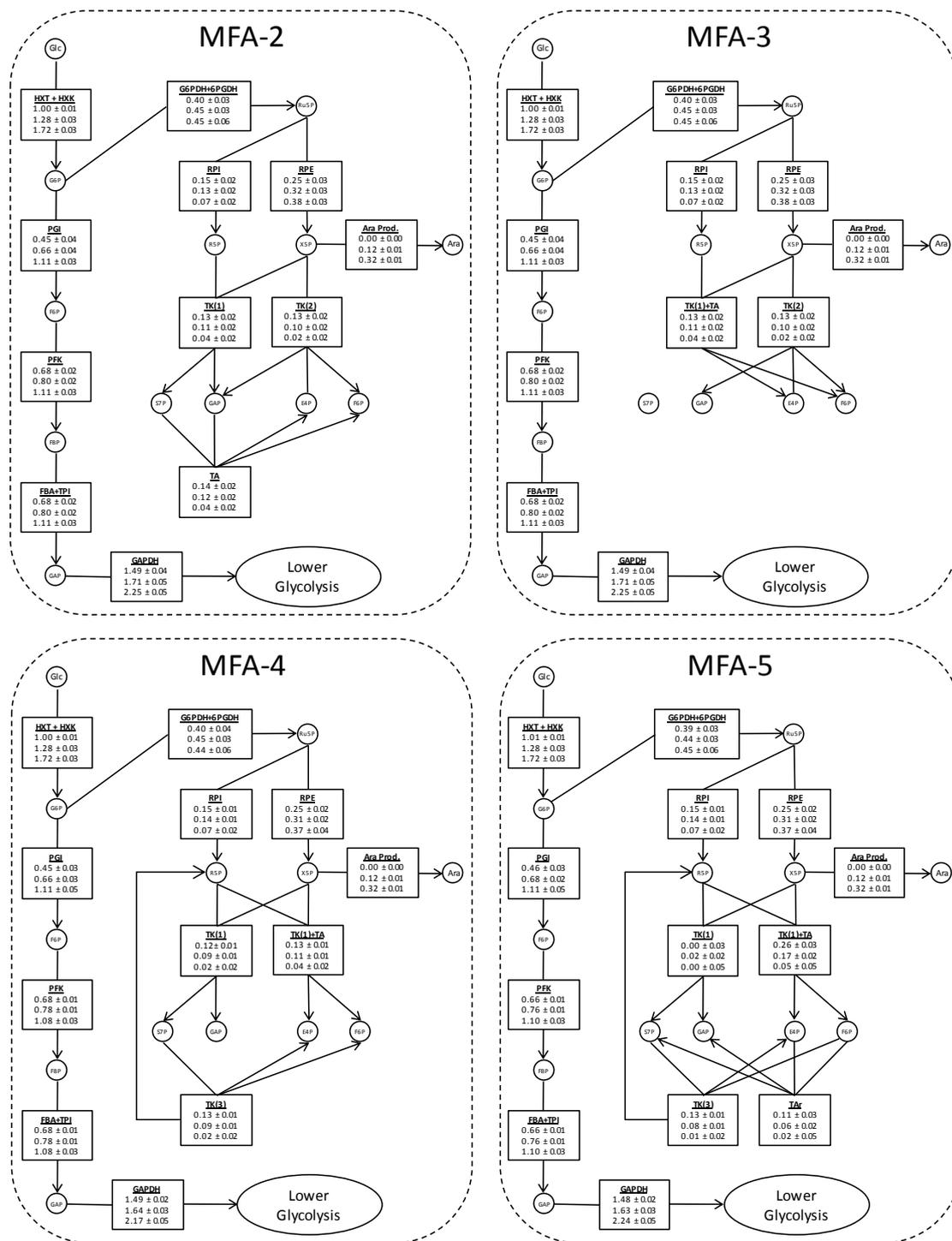


Figure 5.2: Four different *P. pastoris* pentose phosphate pathway metabolic flux analysis considering thermodynamic constrains under different oxygenation conditions. Each box shows the calculated fluxes for each of the 3 oxygen levels conditions. The conditions are ordered top-down as Normoxic, Oxygen-limited and Hypoxic.

Table 5.4 shows the energetic growth requirements for *P. pastoris* under the different oxygen conditions from the MFA-5 calculation. As explained in the materials and methods section, the maintenance requirements were fixed at 2.3 mmol ATP/(gDCW\*h) according to available published data [79, 80]. The energy needs for biomass growth ranged between 2.3-2.5 mmol ATP/(gDCW\*h) depending on the conditions although no significant differences were found related to oxygen availability. On the other hand, a clearly increased trend could be observed in the ATP requirements for glucose transport when the oxygen availability was decreased. Besides this, Ammonia, Sulphate and Metals uptake costs were equivalent in all the conditions.

As mentioned before, the ATP production at Oxidative Phosphorylation level was calculated as the value closing the ATP balance in each steady state (see materials and methods section). Notice that behind this calculation, there is the assumption that the stoichiometry used was representative of the total ATP consumption-production reactions. To verify the physiological feasibility of these ATP production rates, the corresponding P/O ratio was calculated being all the ratios inside a window range of feasible values ( $1 < \text{P/O ratio} < 1.5$ ).

**Table 5.4:** Energetic metabolism parameters calculated for *P. pastoris* under different oxygen conditions. The values are mmol ATP/(gDCW\*h) except for the two ratios.

	Normoxia	O <sub>2</sub> -limited	Hypoxia
Biomass Growth	2.36 ± 0.01	2.47 ± 0.03	2.44 ± 0.03
Biomass Maintenance	2.3 ± 0.2	2.3 ± 0.2	2.3 ± 0.2
Transport outside-in			
Glucose	1.98 ± 0.01	2.54 ± 0.07	3.43 ± 0.12
Ammonia	0.50 ± 0.02	0.52 ± 0.02	0.53 ± 0.05
Sulphate	0.007 ± 0.0004	0.006 ± 0.0005	0.005 ± 0.0005
Metals	0.16 ± 0.01	0.15 ± 0.01	0.14 ± 0.03
Oxidative Phosphorylation	5.3 ± 0.3	5.7 ± 0.3	5.8 ± 0.4
P/O Ratio	1.18 ± 0.07	1.44 ± 0.09	1.45 ± 0.12
Total ATP consumption	8.6 ± 0.3	9.4 ± 0.3	10.5 ± 0.4
Oxid Phop/Total ATP	0.62 ± 0.04	0.61 ± 0.04	0.55 ± 0.04

Interestingly, when all the ATP costs were pooled together, an increase in the cell energetic requirements was determined as the oxygen limitation became stronger. Furthermore, in all the environmental conditions 50% of ATP was produced through the respiratory chain coupled to the oxidative phosphorylation while the total ATP consumed decreases as the total biomass produced for the same amount of substrate decreases. These values were opposite to the results obtained from the Crabtree positive yeast, *S. cerevisiae*, where the oxidative phosphorylation contribution to the total cell ATP cell requirements decrease with oxygen availability [27]. These differences could be due to the fact

that *S. cerevisiae* is able to grow under anoxic condition which is not the case of *P. pastoris*. Nevertheless, further studies should be performed on this issue

### ***Metabolome data set reliability improvement***

In addition to the verification of the metabolic fluxes thermodynamic feasibility, the NET analysis also allowed the detection of any significant metabolite measurement error to constrain the metabolite concentration ranges. This improves the reliability of the experimental dataset. Figure 5.3 shows the results obtained from a NET analysis of the three oxygenation conditions using the metabolite flux directions derived from the MFA-5. Notice that the concentration units from the anNET software have been transformed back to  $\mu\text{mol/gDCW}$  using the 1.7 ml/gDCW specific volume used above.

Globally, in 14 out of the 20 measured metabolites the NET analysis could not decrease the metabolite concentration ranges, as they are found to be in agreement with the metabolic fluxes directions. This result highlights the high quality of the metabolite measurements obtained. In addition, 2PG, G1P, Ru5P and Xu5P metabolite concentrations could be further constrained. Nevertheless, for 3PG the measured concentrations were found inconsistent with the flux directions which force the utilization of the 2-3PG measurements instead of the 3PG for the NET analysis in all the cases. In Figure 3 the measured and the calculated 3PG intracellular pools are shown indicating an underestimation of this metabolite.

For the rest of the analysis, the R5P was only constrained by an upper bound due to a possibly non-homogenous distribution in the cell as explained before. For this metabolite, the NET analysis showed an overestimated quantification suggesting that not all the R5P quantified was available for the central metabolism enzymes.

In addition, some concentrations of the unmeasured metabolites could also be constrained (Figure 5.4). Specifically, Glycerol, GAP and 1,3 DPG were constrained in a narrow concentration window. Besides this, no significant differences were observed between culture conditions, for normoxic conditions the predicted ranges of metabolite levels were much narrower indicating stronger thermodynamic metabolite pool restriction. On the other hand, for Alcad, Oaa, Pi, Pyr and intracellular ethanol no further constrain could be performed.

### ***Redox state of the cell***

The redox state of the cell is of key importance to sustain the proper operation of the cell metabolism [82]. The main redox coenzymes are electron-carriers involved in hundreds of cell reactions located in a variety of cell compartments [116]. Therefore, in principle, each cell compartment could have a different redox state depending on the relative values of the different oxidation states of the redox coenzymes. Besides, not all of these coenzymes are involved in the redox state of the cell as some of

them are linked to proteins [117]. Therefore, direct quantification of the no-linked redox coenzymes becomes highly complicated.

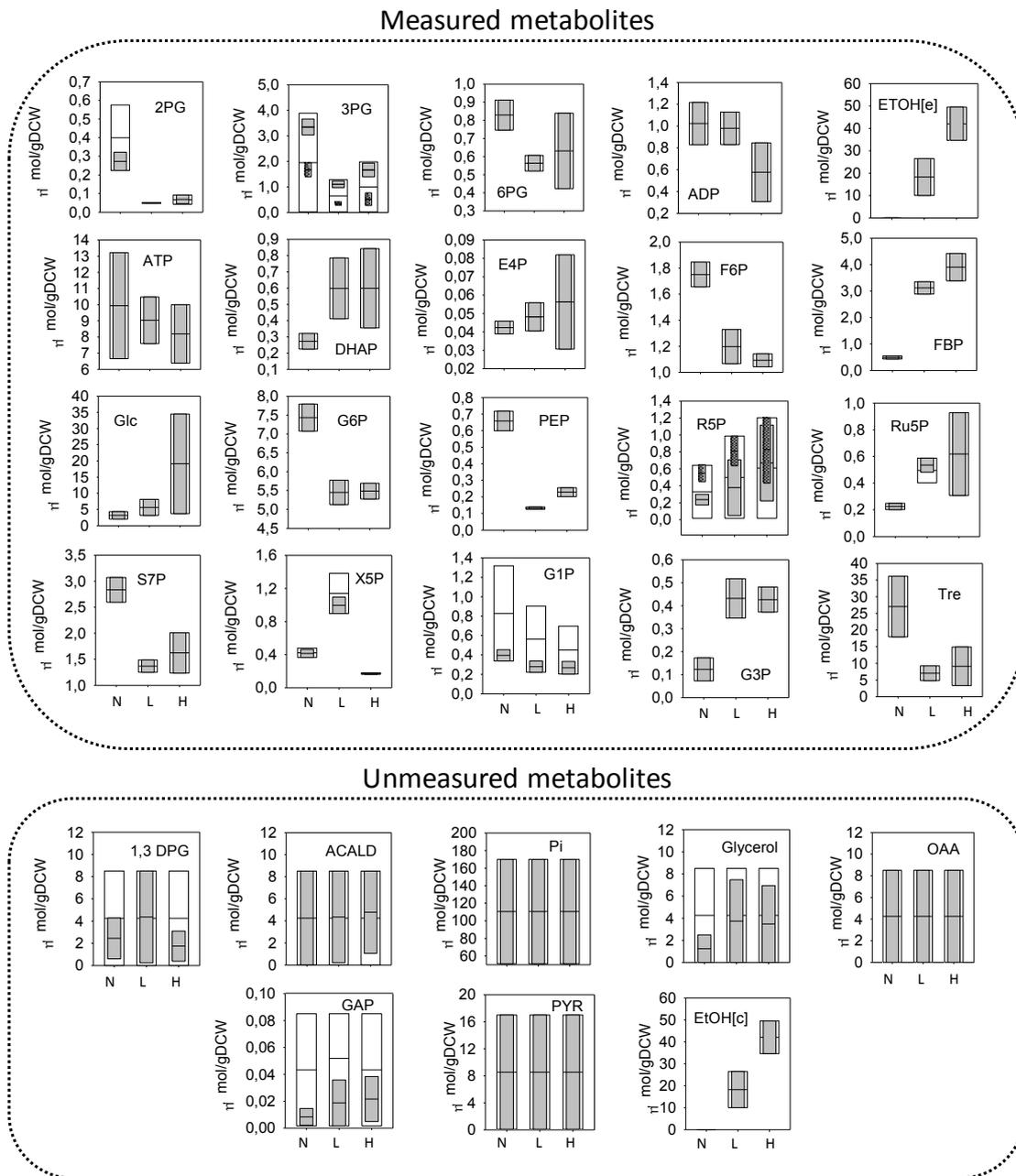


Figure 5.3: Concentration range of the measured and the most probable ranges of the non measured ones. The white bars represent *de a priori* considered metabolite ranges while the light grey bars show the corrected values after performing a network-embedded thermodynamic (NET) analysis. In case of detection of a significant metabolite quantification error, the original measurement (dark grey bars) and concentration ranges before and after the NET analysis are shown together. N: Normoxia; L: Oxygen-limited; H: Hypoxia; [c]: cytosol; [e]: extracellular.

In the current literature different approaches can be found to indirectly determine the different redox states, bypassing all the problems related to the direct quantification. These approaches are based mainly in the quantification of indicators involved in an oxido–reductive reaction and use it to obtain an approximation of the  $\text{NAD}^+/\text{NADH}$  ratio [116, 117]. Another possibility would be to use the NET analysis to obtain information about the redox state of the cell as previously proposed by Kümmel and co-workers [103].

In the present work a NET analysis was performed on the metabolome data sets available for *P. pastoris*, obtaining the minimum ratio of the two oxidation states of the coenzymes which make the system feasible in each experimental condition (Table 5.5). The oxidation state ratios, as expected, were highly influenced by the assumed cytosolic pH (in the present work 7.2. D. Mattanovich personal communication). Thereby, two alternative pHs were also tested analyse the pH sensitivity of the mentioned ratio.

The oxygen availability is known to have a huge impact over the metabolic redox state due to its role as final electron acceptor in the respiratory chain. In addition, the oxygen limitation has already described to have an impact on the biomass composition of *P. pastoris* which increases the storage of reduced compounds in direct proportion to the oxygen limitation [76](Chapter 2).

**Table 5.1:** Ratios of the oxidation states at different oxygen conditions for *P. pastoris*. Note that for *S. cerevisiae*, the reference value shown is a measured ratio while for *P. pastoris* the displayed values are the calculated minimum feasible ratios. \*Data taken from [116]

pH	$\text{NAD}^+/\text{NADH}_{\min}$		
	7.0	7.2	7.5
Normoxia	233	101	28
Oxygen Limited	6.1	2.7	0.7
Hypoxia	13.7	5.9	1.6
<i>S. cerevisiae</i> * (Normoxia)	101		

Focusing on the ratios shown on Table 5.5, it is clear that the oxygen limitation reduces the  $\text{NAD}^+/\text{NADH}_{\min}$  ratio as those are lower at oxygen limitation conditions. Thereby, promoting a more reduced cytosol as seen for the total biomass composition [76].

In the case of *S. cerevisiae*, the  $\text{NAD}^+/\text{NADH}$  ratio was indirectly measured using a heterologous redox reporter reaction genetically introduced in the cell [116]. Interestingly, comparing the results under equivalent conditions the redox states seemed to be equal.

## Conclusions

In this study, the determination of the effect of decreased levels of oxygen availability on the metabolic readjustment of the *P.pastoris* central carbon metabolism and particularly on metabolite levels was

performed. This allowed to apply a NET analysis to specific key parts of the metabolism as well as a metabolic flux analysis.

The range of performed analyses allowed to validate nearly all of the obtained metabolite concentrations as well as to identify a few measures probably suffering of significant experimental error and therefore used with caution.

After the data validation step, the analysis of the pentose phosphate data revealed an inconsistency among the measured metabolite data and any feasible metabolic flux distribution. Once discarded any possible experimental data source of error, different configurations of the pentose phosphate cycle were assayed. Results indicate that, to obtain a feasible flux distribution consistent with the thermodynamic constrains derived from the metabolite levels, an alternative configuration of the pentose phosphate cycle has to be adopted. Among the possible alternatives the inclusion of a metabolite channeling step between the TK(1) and TA reactions shows the most promising results.

Using a stoichiometric network that includes redox cofactor balances, previously validated in equivalent experimental conditions using  $^{13}\text{C}$ -labelling, a metabolic flux analysis step indicated that under oxygen limiting conditions the ATP requirements increase compared to normoxic conditions.

## Appendix

### **Appendix 5.1:** Unmeasured metabolite ranges for NET analysis

The metabolites which were not directly measured were metabolite-specific constrained to be within the thermodynamically feasible limits [96, 103]. Table S5.1 shows the metabolite ranges used for the unmeasured metabolites included in the stoichiometric model. In the cases of GAP and  $P_i$ , a narrower window range could be applied as they have been previously measured in *S. cerevisiae* giving an idea of the feasible metabolite ranges [96]. For the rest of metabolites a wide range of concentrations was used (3-4 decades).

**Table S5.1:** Unmeasured metabolite ranges used in the network-embedded thermodynamic analysis.

	Lower limit (mM)	Upper limit (mM)
1,3-DPG	0.001	5
Aald	0.005	5
Glycerol	0.005	5
GAP	0.001	0.05
$NAD^+$	0.01	100
NADH	0.01	100
$NADP^+$	0.01	100
NADPH	0.01	100
$P_i$	30	100
Oaa	0.005	5
Pyr	0.05	10

### **Appendix 5.2:** Minimal *Pichia pastoris* stoichiometric model

In this study a minimal stoichiometric model of *P. pastoris*, including only the main metabolic central pathways, is used for the metabolic flux analyses (MFA). As constrains for the MFA-2 model, the measured  $O_2$  and  $CO_2$  specific rates together with the redox constrains were used instead of the  $^{13}C$  derived constrains [52] (MFA-2). In addition, a non classical pentose phosphate pathway was implemented in order to reach a better agreement with the thermodynamic constrains (MFA-3, MFA-4 and MFA-5). Notice that for MFA-5 only the  $^{13}C$  derived constrain derived from the TA  $^{13}C$ -NMR data (see next section) was used.

The metabolic reactions used were derived from the previous work stoichiometry:

Name reactionsStoichiometry**Glycolysis**

HXT + HXK	$\text{Glc}_{\text{ext}} + 2 \text{ATP} \Rightarrow \text{G6P} + 2 \text{ADP} + \text{P}_i$
PGI	$\text{G6P} \Leftrightarrow \text{F6P}$
PFK + FBA	$\text{F6P} + \text{ATP} \Rightarrow \text{GAP} + \text{DHAP} + \text{ADP}$
TPI	$\text{DHAP} \Leftrightarrow \text{GAP}$
GAPDH + PGK + GPM + ENO + PYK	$\text{GAP} + 2\text{ADP} + \text{NAD}^+ \Rightarrow \text{Pyr} + 2\text{ATP} + \text{NADH}$

**Pentose Phosphate Pathway**

G6PDH + 6PGDH	$\text{G6P} + 2 \text{NADP}^+ \Rightarrow \text{RU5P} + 2 \text{NADPH} + \text{CO}_2$
RPE	$\text{RU5P} \Leftrightarrow \text{XU5P}$
RPI	$\text{RU5P} \Leftrightarrow \text{R5P}$
TK(1)	$\text{R5P} + \text{XU5P} \Leftrightarrow \text{S7P} + \text{GAP}$
TK(1)+TA	$\text{R5P} + \text{XU5P} \Leftrightarrow \text{F6P} + \text{E4P}$
TK(2)	$\text{E4P} + \text{XU5P} \Leftrightarrow \text{F6P} + \text{GAP}$
TK(3)	$\text{S7P} + \text{E4P} \Leftrightarrow \text{F6P} + \text{R5P}$
TA	$\text{S7P} + \text{GAP} \Leftrightarrow \text{F6P} + \text{E4P}$
TAr	$\text{F6P} + \text{E4P} \Leftrightarrow \text{S7P} + \text{GAP}$

**Pyruvate Dehydrogenase**

PDH	$\text{CoA} + \text{NAD}^+ + \text{Pyr} \Rightarrow \text{AcCoA}_{\text{mit}} + \text{NADH} + \text{CO}_2$
-----	--

**TCA cycle**

CIT Syn.	$\text{AcCoA}_{\text{mit}} + \text{Oaa}_{\text{mit}} \Rightarrow \text{CoA} + \text{Cit}_{\text{mit}}$
ACO	$\text{citrate} \Leftrightarrow \text{Isocitrate}$
ISODH	$\text{Isocitrate} + \text{NAD}^+ \Rightarrow \text{NADH} + \text{CO}_2 + \text{Akg}_{\text{mit}}$
aKGDH + SUCDH + Fum	$\text{NAD}^+ + \text{Akg}_{\text{mit}} + \text{ADP} + \text{FAD} \Rightarrow \text{Mal} + \text{NADH} + \text{FADH}_2 + \text{ATP} + \text{CO}_2$
MalDH	$\text{Mal} + \text{NAD}^+ \Rightarrow \text{OAA}_{\text{mit}} + \text{NADH}$

**Anaplerotic Pathways**

PyrCK	$\text{ATP} + \text{CO}_2 + \text{Pyr} \Rightarrow \text{Oaa}_{\text{cyt}} + \text{ADP}$
-------	--

**Fermentative Pathways**

PDC	$\text{Pyr} \Rightarrow \text{AcO} + \text{CO}_2$
ADDH + ACETCoA	$\text{AcO} + \text{NADP}^+ + \text{CoA} + \text{ATP} \Rightarrow \text{NADPH} + \text{AcCoA}_{\text{cyt}} + \text{Pyrophosphate} + \text{AMP}$
ALDH + ET	$\text{AcO} + \text{NADH} \Rightarrow \text{Eth}_{\text{ext}} + \text{NAD}^+$
G3PDH + GT	$\text{DHAP} + \text{NADH} + \text{ATP} \Rightarrow \text{Glycerol}_{\text{cyt}} + \text{Orthophosphate} + \text{ADP} + \text{NAD}^+$
Ara Syn.	$\text{XU5P} + \text{NADH} \Rightarrow \text{ARA}_{\text{ext}} + \text{NAD}^+$

**Respiratory Chain**

RC	$\text{NADH} + 0.5 \text{O}_2 \Rightarrow \text{NAD}^+ + \text{H}_2\text{O}$
Oxidative Phosphorilation	$\text{ADP} + \text{P}_i \Rightarrow \text{ATP}$

**Transport**

OAA Transp.	$Oaa_{cyt} \rightleftharpoons Oaa_{mit}$
Glycerol Transp.	$Glycerol_{cyt} \Rightarrow Glycerol_{ext}$
$NH_4^+$ -Assimilation	$NH_4^+_{ext} + ATP \Rightarrow NH_4^+_{cyt} + ADP + P_i$
$SO_4^{2-}$ -Assimilation	$SO_4^{2-}_{ext} + ATP \Rightarrow SO_4^{2-}_{cyt} + ADP + P_i$
Metal-Assimilation	$Metal_{ext} + ATP \Rightarrow Metal_{cyt} + ADP + P_i$

**Biomass synthesis reactions for each oxygenation condition**

1. Protein (Composition derived from the measured amino acid composition at each oxygenation condition [76]). The energy needed to biosynthesize 1 C-mol of protein was derived from the synthesis of each amino acid and the protein polymerization value taken from bibliographic references, (Chapter 2 [77]).

Prot\_N (Normoxic condition):  $0.136 \text{ Pyr} + 0.006 \text{ R5P} + 0.013 \text{ E4P} + 0.031 \text{ Oaa}_{cyt} + 0.014 \text{ AcCoA}_{mit} + 0.009 \text{ Oaa}_{mit} + 0.075 \text{ Akg}_{mit} + 0.013 \text{ AcCoA}_{cyt} + 0.061 \text{ NAD}^+ + 0.354 \text{ NADPH} + 1.132 \text{ ATP} \Rightarrow 1 \text{ C-mol Protein} + 0.002 \text{ GAP} + 0.058 \text{ CO}_2 + 0.061 \text{ NADH} + 0.354 \text{ NADP}^+ + 1.132 \text{ ADP} + 1.132 \text{ P}_i$

Prot\_L-H (Oxygen limited and hypoxic conditions):  $0.141 \text{ Pyr} + 0.006 \text{ R5P} + 0.013 \text{ E4P} + 0.033 \text{ Oaa}_{cyt} + 0.015 \text{ AcCoA}_{mit} + 0.009 \text{ Oaa}_{mit} + 0.070 \text{ Akg}_{mit} + 0.014 \text{ AcCoA}_{cyt} + 0.062 \text{ NAD}^+ + 0.353 \text{ NADPH} + 1.132 \text{ ATP} \Rightarrow 1 \text{ C-mol Protein} + 0.002 \text{ GAP} + 0.066 \text{ CO}_2 + 0.062 \text{ NADH} + 0.353 \text{ NADP}^+ + 1.132 \text{ ADP} + 1.132 \text{ P}_i$

2. Carbohydrate, Glycogen and Trehalose (Composition derived from [118]).

Carbhy.	$0.113 \text{ G6P} + 0.053 \text{ F6P} \Rightarrow 1 \text{ C-mol Carbohydrate}$
Glycogen	$0.166 \text{ G6P} \Rightarrow 1 \text{ C-mol Glycogen}$
Trehalose	$0.166 \text{ G6P} \Rightarrow 1 \text{ C-mol Trehalase}$

3. Lipid (Composition derived from the mean lipid composition from [76]).

Lipid  $0.002 \text{ G6P} + 0.005 \text{ Pyr} + 0.006 \text{ CO}_2 + 0.061 \text{ AcCoA}_{mit} + 0.386 \text{ AcCoA}_{cyt} + 0.065 \text{ O}_2 + 0.022 \text{ Glycerol}_{cyt} + 0.06 \text{ NADH} + 0.77 \text{ NADPH} + 0.420 \text{ ATP} \Rightarrow 1 \text{ C-mol Lipid} + 0.06 \text{ NAD}^+ + 0.77 \text{ NADP}^+ + 0.420 \text{ ADP} + 0.420 \text{ P}_i$

4. RNA (Composition derived from the RNA composition proposed by Stephanopoulos and co-workers [77]).

RNA  $0.060 \text{ Pyr} + 0.060 \text{ CO}_2 + 0.111 \text{ R5P} + 0.051 \text{ Oaa}_{cyt} + 0.181 \text{ NAD}^+ + 0.079 \text{ NADPH} + 1.170 \text{ ATP} \Rightarrow 1 \text{ C-mol RNA} + 0.181 \text{ NADH} + 0.079 \text{ NADP}^+ + 1.170 \text{ ADP} + 1.170 \text{ P}_i$

5. DNA (Composition derived from the DNA composition published in [63] and the DNA polymerization from [77]).

DNA  $0.054 \text{ Pyr} + 0.085 \text{ CO}_2 + 0.108 \text{ R5P} + 0.054 \text{ Oaa}_{cyt} + 0.161 \text{ NAD}^+ + 0.225 \text{ NADPH} + 1.208 \text{ ATP} \Rightarrow 1 \text{ C-mol DNA} + 0.161 \text{ NADH} + 0.225 \text{ NADP}^+ + 1.208 \text{ ADP} + 1.208 \text{ P}_i$

6. Biomass energetic maintenance requirements were fixed at  $2.3 \pm 0.23 \text{ mmol ATP}/(\text{gDCW} \cdot \text{h})$ . These value was derived from the two genomic-scale metabolic networks published for *P. pastoris* [79, 80].

ATP maintenance  $ATP \Rightarrow ADP + P_i$  (redox-MFA)

In the arabinitol production pathway, the enzyme L-xylulose reductase could be modulated by two redox coenzymes [119]. In *P. pastoris*, the authors were not able to find information on the redox specificity or the presence of iso-enzymes. Therefore, the MFA were performed using each one of the coenzymes to check the adjustment with the previous MFA (MFA-1). As a result, the stoichiometry which gave better agreement with the previously results, was the one using the NADH as a substrate for the L-xylulose reductase appearing in the arabinitol production reaction (Ara Syn).

The Isocitrate Dehydrogenase could present two iso-enzymes with different redox affinity in *P. pastoris*. Initially, the iso-enzyme which produces the NADPH was included in the MFA, however, it was later on discarded due to the fact that in all the calculations its resulting flux was zero.

The biomass composition was taken from the *P. pastoris* specific data [76]. For the two oxygen limited and hypoxic conditions performed in this study, the same biomass composition was used because no specific data was available.

### **Appendix 5.3:** Implementation of $^{13}\text{C}$ -NMR transaldolase ratio

The transketolase (TK) and the transaldolase (TA) enzymes are involved in the non-oxidative pentose phosphate pathway. Classically these enzymes are considered to work close to the equilibrium constant allowing reversible reactions. However, in *P. pastoris* growing under different oxygen conditions, the transformed Gibbs free energy of the reaction where these enzymes are involved disagreed with that reversibility (Table 5.3, main manuscript).

On the other hand, in Table S5.2 there are shown the  $^{13}\text{C}$ -NMR metabolic flux ratios of TK and TA enzymes from equivalent cultures performed by Baumann and co-workers [52]. These ratios were calculated looking R5P carbon skeleton cleavage (using the Histidine carbon skeleton) to determine which R5P pool fraction had been produced by, at least, one TK or TA reaction respectively [73, 74]. These results were commonly attributed to the TK and TA reversibility although a R5P re-circulation was also a possible explanation through the TK(3) reaction.

**Table S5.2:** Metabolic flux ratio analysis results from Baumann and co-workers study [52].

	Normoxia	O <sub>2</sub> -Limited	Hypoxia
Transketolase ratio	66 ± 2	70 ± 2	62 ± 2
Transaldolase ratio	40 ± 2	29 ± 2	24 ± 2

In MFA-5, the TA  $^{13}\text{C}$ -NMR ratios were implemented considering both directions of the TA stoichiometry could take place inside the cell. The classical direction was included in the metabolic channeling with the TK(1) and the inverse direction was added as an extra reaction in the model due to the fact it is thermodynamically feasible (Table 5.3, main manuscript).

In this stoichiometric network, the reaction which could produce the R5P are the RPI and the TK(3) reactions. Therefore, RPI is the source of the uncleavage carbon skeleton and the TK(3) of the cleavage one.

Focusing on the TK kinetic mechanism, the R5P of the TK(3) reactions is produced due to the S7P cleavage. Thus, looking the S7P production from the TAr reactions, the TA ratio could be defined as function of metabolic flux relation due to the TK(3) it is the only reaction taken into account which consumes S7P (Eq. S5.1)

$$TA\ ratio = a = \frac{TAr}{RPI+TK(3)} \quad (\text{Eq. S5.1})$$

The following linear constraint equation (Eq. S5.2) was derived from the flux ratio equation and implemented in the MFA-5 as described in [27, 52].

$$R_{TA} = TAr - a * (RPI) - a * (TK(3)) \quad (\text{Eq. S5.2})$$

As a result, the MFA-5 could explain much better the experimental results than the classical non-oxidative pentose phosphate pathway stoichiometry.

For the hypoxic condition performed in this study, no specific metabolic flux ratios were available and the oxygen limited values were used due to no big differences were expected from both oxygen limited conditions in this ratio.

**Appendix 5.4:** Measured intracellular metabolite pools**Table S5.3:** Measured intracellular metabolites. The values are in  $\mu\text{mol/gDCW} \pm \text{sd}$ .

	Normoxia	O <sub>2</sub> -limited	Hypoxia
G6P	7.43 $\pm$ 0.13	5.45 $\pm$ 0.12	5.49 $\pm$ 0.08
F6P	1.75 $\pm$ 0.04	1.20 $\pm$ 0.05	1.09 $\pm$ 0.02
FBP	0.49 $\pm$ 0.02	3.12 $\pm$ 0.09	3.91 $\pm$ 0.19
6PG	0.83 $\pm$ 0.03	0.56 $\pm$ 0.02	0.63 $\pm$ 0.08
S7P	2.83 $\pm$ 0.09	1.37 $\pm$ 0.04	1.62 $\pm$ 0.14
UDP-G	1.1 $\pm$ 0.04	0.9 $\pm$ 0.0	0.83 $\pm$ 0.0
M6P	2.20 $\pm$ 0.06	1.54 $\pm$ 0.06	1.34 $\pm$ 0.02
G3P	0.12 $\pm$ 0.02	0.43 $\pm$ 0.03	0.43 $\pm$ 0.02
G1P	0.83 $\pm$ 0.18	0.56 $\pm$ 0.13	0.45 $\pm$ 0.09
PEP	0.66 $\pm$ 0.02	0.13 $\pm$ 0.00	0.23 $\pm$ 0.01
Cit	7.56 $\pm$ 0.31	13.78 $\pm$ 0.83	16.96 $\pm$ 1.02
aKG	3.67 $\pm$ 0.12	6.50 $\pm$ 0.32	7.04 $\pm$ 0.44
Suc	5.86 $\pm$ 0.98	5.75 $\pm$ 0.32	8.19 $\pm$ 0.81
Fum	2.00 $\pm$ 0.32	2.92 $\pm$ 0.16	3.01 $\pm$ 0.33
Mal	8.6 $\pm$ 0.16	10.4 $\pm$ 0.4	8.52 $\pm$ 0.3
Glc int	3 $\pm$ 0.60	6 $\pm$ 1	19.10 $\pm$ 8
2PG	0.40 $\pm$ 0.09	0.05 $\pm$ 0.00	0.07 $\pm$ 0.01
IsoCit	0.083 $\pm$ 0.007	0.079 $\pm$ 0.001	0.074 $\pm$ 0.010
DHAP	0.27 $\pm$ 0.03	0.60 $\pm$ 0.10	0.60 $\pm$ 0.13
3PG	1.67 $\pm$ 0.14	0.34 $\pm$ 0.03	0.52 $\pm$ 0.12
E4P	0.04 $\pm$ 0.00	0.05 $\pm$ 0.00	0.06 $\pm$ 0.01
R5P	0.55 $\pm$ 0.05	0.81 $\pm$ 0.09	0.83 $\pm$ 0.20
Rub5P	0.22 $\pm$ 0.01	0.50 $\pm$ 0.05	0.62 $\pm$ 0.16
X5P	0.42 $\pm$ 0.03	1.14 $\pm$ 0.13	1.40 $\pm$ 0.38
Trehalose	27.1 $\pm$ 4.8	7.0 $\pm$ 1.2	9.1 $\pm$ 3.1
AMP	0.24 $\pm$ 0.02	0.23 $\pm$ 0.00	0.17 $\pm$ 0.01
ADP	1.02 $\pm$ 0.07	0.98 $\pm$ 0.06	0.61 $\pm$ 0.09
ATP	9.9 $\pm$ 1.2	9.03 $\pm$ 0.54	8.11 $\pm$ 0.64

**Appendix 5.5:** Stoichiometric network for NET analysis

<u>Abbreviation</u>	<u>Reactions</u>
HXK	glc-D + atp > g6p + adp
PGI	g6p <> f6p
PFK	atp + f6p > adp + fdp + h
FBA	fdp <> dhap + g3p
TPI	dhap <> g3p
GAPDH	g3p + nad + pi <> h + nadh + 13dpg
PGK	13dpg + adp <> 3pg + atp
GPM	3pg <> 2pg
ENO	2pg <> h2o + pep
PYK	adp + h + pep > atp + pyr
G6PDH	g6p + nadp > 6pgl + h + nadph
6PGDH	6pgl + nadp + h2o > ru5p-D + nadph + h + co2tot
RPI	ru5p-D <> r5p
RPE	ru5p-D <> xu5p-D
TK(1)+TA	r5p + xu5p-D <> f6p + e4p
TK(1)	r5p + xu5p-D <> g3p + s7p
TK(3)	s7p + e4p <> f6p + r5p
TA rev	f6p + e4p <> g3p + s7p
G3PDH	dhap + nadh <> glyc3p + nad
PYRCK	pyr + atp + co2tot <> oaa + adp + pi + h2o
PGM	g6p <> g1p
TreP	tre + h2o > (2) glc-D
PDC	pyr <> acald + co2tot
ADH	acald[c] + nadh[c] <> etoh[c] + nad[c]
ET	etoh[c] <> etoh[e]
G3PP	glyc3p+h2o <> pi + glyc

**Appendix 5.6:** Metabolic flux analysis results.**Table S5.4:** Metabolic flux analysis-2 (MFA-2). All the values are in mmol/(gDCW·h) ± sd.

Reaction	Normoxia	O <sub>2</sub> -Limited	Hypoxia
PGI	0.45 ± 0.04	0.66 ± 0.04	1.11 ± 0.05
PFK + FBA	0.68 ± 0.02	0.80 ± 0.02	1.11 ± 0.03
TPI	0.68 ± 0.02	0.80 ± 0.02	1.11 ± 0.03
GAPDH+PGK+GPM+ENO+PYK	1.49 ± 0.04	1.71 ± 0.05	2.25 ± 0.05
G6PDH+6PGDH	0.40 ± 0.05	0.46 ± 0.04	0.45 ± 0.06
RPI	0.15 ± 0.02	0.13 ± 0.02	0.07 ± 0.02
RPE	0.25 ± 0.03	0.32 ± 0.03	0.38 ± 0.04
TK(1)	0.13 ± 0.02	0.11 ± 0.02	0.04 ± 0.02
TA	0.14 ± 0.02	0.12 ± 0.02	0.04 ± 0.02
TK(2)	0.13 ± 0.02	0.10 ± 0.02	0.02 ± 0.02
PDH	0.72 ± 0.04	0.63 ± 0.03	0.67 ± 0.05
CITL	0.68 ± 0.05	0.58 ± 0.04	0.62 ± 0.06
ACO	0.64 ± 0.04	0.55 ± 0.04	0.59 ± 0.06
ICITDH	0.60 ± 0.04	0.52 ± 0.04	0.56 ± 0.06
AKGDH + SUDH + FHM	0.46 ± 0.04	0.39 ± 0.04	0.44 ± 0.06
MALDH	0.43 ± 0.04	0.37 ± 0.04	0.42 ± 0.06
PYRCK	0.40 ± 0.02	0.34 ± 0.02	0.33 ± 0.01
PDC	0.13 ± 0.05	0.47 ± 0.04	1.01 ± 0.07
ADDH + ACETCoA	0.15 ± 0.05	0.20 ± 0.04	0.19 ± 0.06
G3PDH	0.00 ± 0.01	0.00 ± 0.01	0.00 ± 0.00
HXT+HXK	1.00 ± 0.01	1.28 ± 0.03	1.72 ± 0.03
ALH + ET	0.00 ± 0.00	0.29 ± 0.02	0.82 ± 0.05
ARAT	0.00 ± 0.00	0.12 ± 0.01	0.32 ± 0.01
OAA_MT	0.30 ± 0.02	0.25 ± 0.02	0.24 ± 0.01
RC	4.44 ± 0.19	3.92 ± 0.13	3.93 ± 0.26
Oxidative Phosphorylation	4.40 ± 0.28	4.98 ± 0.28	6.54 ± 0.35
Lipid	0.34 ± 0.12	0.45 ± 0.11	0.44 ± 0.16
Carbohydrate	0.90 ± 0.08	1.21 ± 0.11	1.19 ± 0.11
Glycogen	0.37 ± 0.04	0.22 ± 0.02	0.22 ± 0.02
Trehalose	0.03 ± 0.01	0.01 ± 0.00	0.01 ± 0.01
RNA	0.19 ± 0.02	0.18 ± 0.02	0.17 ± 0.02
DNA	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Protein	1.55 ± 0.10	1.62 ± 0.10	1.59 ± 0.10
NH <sub>4</sub> <sup>+</sup> Assimilation	0.50 ± 0.03	0.52 ± 0.04	0.53 ± 0.05
SO <sub>4</sub> <sup>2-</sup> Assimilation	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
Metal Assimilation	0.16 ± 0.01	0.15 ± 0.03	0.14 ± 0.03
ATP maintenance	2.30 ± 0.23	2.30 ± 0.23	2.30 ± 0.23

**Table S5.5:** Metabolic flux analysis-3 (MFA-3). All the values are in mmol/(gDCW·h) ± sd.

Reaction	Normoxia	O <sub>2</sub> -Limited	Hypoxia
PGI	0.45 ± 0.03	0.66 ± 0.03	1.11 ± 0.05
PFK + FBA	0.68 ± 0.01	0.80 ± 0.01	1.11 ± 0.03
TPI	0.68 ± 0.01	0.80 ± 0.01	1.11 ± 0.03
GAPDH+PGK+GPM+ENO+PYK	1.49 ± 0.02	1.70 ± 0.03	2.25 ± 0.05
G6PDH+6PGDH	0.40 ± 0.04	0.46 ± 0.04	0.45 ± 0.06
RPI	0.15 ± 0.01	0.14 ± 0.01	0.07 ± 0.02
RPE	0.25 ± 0.02	0.32 ± 0.03	0.38 ± 0.04
TK(1)+TA	0.14 ± 0.01	0.11 ± 0.01	0.04 ± 0.02
TK(2)	0.12 ± 0.01	0.09 ± 0.01	0.02 ± 0.02
PDH	0.72 ± 0.04	0.63 ± 0.03	0.67 ± 0.05
CITL	0.68 ± 0.04	0.58 ± 0.03	0.62 ± 0.06
ACO	0.64 ± 0.04	0.55 ± 0.03	0.59 ± 0.06
ICITDH	0.60 ± 0.04	0.52 ± 0.03	0.56 ± 0.06
AKGDH + SUDH + FHM	0.46 ± 0.04	0.39 ± 0.03	0.44 ± 0.06
MALDH	0.43 ± 0.04	0.37 ± 0.03	0.42 ± 0.06
PYRCK	0.40 ± 0.01	0.34 ± 0.01	0.33 ± 0.01
PDC	0.13 ± 0.04	0.47 ± 0.04	1.01 ± 0.07
ADDH + ACETCoA	0.15 ± 0.04	0.19 ± 0.04	0.19 ± 0.06
G3PDH	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
HXT+HXK	1.00 ± 0.01	1.28 ± 0.02	1.72 ± 0.03
ALH + ET	0.00 ± 0.00	0.29 ± 0.02	0.82 ± 0.05
ARAT	0.00 ± 0.00	0.12 ± 0.01	0.32 ± 0.01
OAA_MT	0.30 ± 0.27	0.25 ± 0.01	0.24 ± 0.01
RC	4.44 ± 0.11	3.91 ± 0.13	3.93 ± 0.26
Oxidative Phosphorylation	4.40 ± 0.08	4.60 ± 0.28	5.71 ± 0.36
Lipid	0.34 ± 0.03	0.45 ± 0.10	0.44 ± 0.16
Carbohydrate	0.90 ± 0.01	1.21 ± 0.11	1.19 ± 0.11
Glycogen	0.37 ± 0.02	0.22 ± 0.02	0.22 ± 0.02
Trehalose	0.03 ± 0.00	0.01 ± 0.00	0.01 ± 0.01
RNA	0.19 ± 0.09	0.18 ± 0.02	0.17 ± 0.02
DNA	0.00 ± 0.03	0.00 ± 0.00	0.00 ± 0.00
Protein	1.55 ± 0.00	1.62 ± 0.10	1.59 ± 0.10
NH <sub>4</sub> <sup>+</sup> Assimilation	0.50 ± 0.01	0.52 ± 0.04	0.53 ± 0.05
SO <sub>4</sub> <sup>2-</sup> Assimilation	0.01 ± 0.23	0.01 ± 0.00	0.01 ± 0.00
Metall Assimilation	0.16 ± 0.01	0.15 ± 0.03	0.14 ± 0.03
ATP maintenance	2.30 ± 0.23	2.30 ± 0.23	2.30 ± 0.23

**Table S5.6:** Metabolic flux analysis-4 (MFA-4). All the values are in mmol/(gDCW·h) ± sd.

Reaction	Normoxia	O <sub>2</sub> -Limited	Hypoxia
PGI	0.45 ± 0.03	0.66 ± 0.03	1.11 ± 0.05
PFK + FBA	0.68 ± 0.01	0.80 ± 0.01	1.11 ± 0.03
TPI	0.68 ± 0.01	0.80 ± 0.01	1.11 ± 0.03
GAPDH+PGK+GPM+ENO+PYK	1.49 ± 0.02	1.70 ± 0.03	2.25 ± 0.05
G6PDH+6PGDH	0.40 ± 0.04	0.46 ± 0.04	0.45 ± 0.06
RPI	0.15 ± 0.01	0.14 ± 0.01	0.07 ± 0.02
RPE	0.25 ± 0.02	0.32 ± 0.03	0.38 ± 0.04
TK(1)+TA	0.14 ± 0.01	0.11 ± 0.01	0.04 ± 0.02
TK(1)	0.12 ± 0.01	0.09 ± 0.01	0.02 ± 0.02
TK(3)	0.13 ± 0.01	0.09 ± 0.01	0.02 ± 0.02
PDH	0.72 ± 0.04	0.63 ± 0.03	0.67 ± 0.05
CITL	0.68 ± 0.04	0.58 ± 0.03	0.62 ± 0.06
ACO	0.64 ± 0.04	0.55 ± 0.03	0.59 ± 0.06
ICITDH	0.60 ± 0.04	0.52 ± 0.03	0.56 ± 0.06
AKGDH + SUDH + FHM	0.46 ± 0.04	0.39 ± 0.03	0.44 ± 0.06
MALDH	0.43 ± 0.04	0.37 ± 0.03	0.42 ± 0.06
PYRCK	0.40 ± 0.01	0.34 ± 0.01	0.33 ± 0.01
PDC	0.13 ± 0.04	0.47 ± 0.04	1.01 ± 0.07
ADDH + ACETCoA	0.15 ± 0.04	0.19 ± 0.04	0.19 ± 0.06
G3PDH	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
HXT+HXK	1.00 ± 0.01	1.28 ± 0.02	1.72 ± 0.03
ALH + ET	0.00 ± 0.00	0.29 ± 0.02	0.82 ± 0.05
ARAT	0.00 ± 0.00	0.12 ± 0.01	0.32 ± 0.01
OAA_MT	0.30 ± 0.27	0.25 ± 0.01	0.24 ± 0.01
RC	4.44 ± 0.11	3.91 ± 0.13	3.93 ± 0.26
Oxidative Phosphorylation	4.40 ± 0.08	4.60 ± 0.28	5.71 ± 0.36
Lipid	0.34 ± 0.03	0.45 ± 0.10	0.44 ± 0.16
Carbohydrate	0.90 ± 0.01	1.21 ± 0.11	1.19 ± 0.11
Glycogen	0.37 ± 0.02	0.22 ± 0.02	0.22 ± 0.02
Trehalose	0.03 ± 0.00	0.01 ± 0.00	0.01 ± 0.01
RNA	0.19 ± 0.09	0.18 ± 0.02	0.17 ± 0.02
DNA	0.00 ± 0.03	0.00 ± 0.00	0.00 ± 0.00
Protein	1.55 ± 0.00	1.62 ± 0.10	1.59 ± 0.10
NH <sub>4</sub> <sup>+</sup> Assimilation	0.50 ± 0.01	0.52 ± 0.04	0.53 ± 0.05
SO <sub>4</sub> <sup>2-</sup> Assimilation	0.01 ± 0.23	0.01 ± 0.00	0.01 ± 0.00
Metal Assimilation	0.16 ± 0.01	0.15 ± 0.03	0.14 ± 0.03
ATP maintenance	2.30 ± 0.23	2.30 ± 0.23	2.30 ± 0.23

**Table S5.7:** Metabolic flux analysis-5 (MFA-5). All the values are in mmol/(gDCW·h) ± sd.

Reaction	Normoxia	O <sub>2</sub> -Limited	Hypoxia
PGI	0.46 ± 0.03	0.68 ± 0.02	1.12 ± 0.05
PFK + FBA	0.66 ± 0.01	0.76 ± 0.01	1.10 ± 0.03
TPI	0.67 ± 0.01	0.76 ± 0.01	1.10 ± 0.03
GAPDH+PGK+GPM+ENO+PYK	1.48 ± 0.02	1.63 ± 0.03	2.24 ± 0.05
G6PDH+6PGDH	0.39 ± 0.03	0.44 ± 0.03	0.45 ± 0.06
RPI	0.12 ± 0.01	0.11 ± 0.01	0.06 ± 0.02
RPE	0.26 ± 0.02	0.33 ± 0.02	0.39 ± 0.04
TK(1)+TA	0.15 ± 0.02	0.10 ± 0.01	0.03 ± 0.04
TA r	0.04 ± 0.02	0.01 ± 0.01	0.00 ± 0.03
TK(1)	0.10 ± 0.02	0.09 ± 0.01	0.03 ± 0.04
TK(3)	0.12 ± 0.01	0.08 ± 0.01	0.02 ± 0.02
PDH	0.72 ± 0.03	0.62 ± 0.02	0.67 ± 0.05
CITL	0.68 ± 0.03	0.57 ± 0.03	0.62 ± 0.06
ACO	0.64 ± 0.03	0.55 ± 0.03	0.59 ± 0.06
ICITDH	0.61 ± 0.03	0.53 ± 0.03	0.57 ± 0.06
AKGDH + SUDH + FHM	0.46 ± 0.03	0.41 ± 0.02	0.44 ± 0.06
MALDH	0.44 ± 0.03	0.41 ± 0.02	0.42 ± 0.06
PYRCK	0.40 ± 0.01	0.31 ± 0.01	0.33 ± 0.01
PDC	0.13 ± 0.04	0.45 ± 0.04	1.00 ± 0.07
ADDH + ACETCoA	0.15 ± 0.04	0.19 ± 0.03	0.19 ± 0.06
G3PDH	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
HXT+HXK	1.01 ± 0.01	1.28 ± 0.02	1.72 ± 0.03
ALH + ET	0.00 ± 0.00	0.28 ± 0.02	0.82 ± 0.05
ARAT	0.00 ± 0.00	0.13 ± 0.01	0.32 ± 0.01
OAA_MT	0.30 ± 0.01	0.35 ± 0.02	0.24 ± 0.01
RC	4.44 ± 0.16	3.91 ± 0.11	3.93 ± 0.25
Oxidative Phosphorylation	4.37 ± 0.26	4.60 ± 0.26	5.71 ± 0.35
Lipid	0.33 ± 0.10	0.43 ± 0.09	0.44 ± 0.15
Carbohydrate	0.90 ± 0.08	1.21 ± 0.11	1.19 ± 0.11
Glycogen	0.37 ± 0.03	0.22 ± 0.02	0.22 ± 0.02
Trehalose	0.03 ± 0.01	0.01 ± 0.00	0.01 ± 0.01
RNA	0.18 ± 0.02	0.17 ± 0.02	0.17 ± 0.02
DNA	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Protein	1.55 ± 0.09	1.61 ± 0.09	1.59 ± 0.10
NH <sub>4</sub> <sup>+</sup> Assimilation	0.50 ± 0.03	0.52 ± 0.04	0.53 ± 0.05
SO <sub>4</sub> <sup>2-</sup> Assimilation	0.01 ± 0.00	0.01 ± 0.00	0.01 ± 0.00
Metal Assimilation	0.16 ± 0.01	0.15 ± 0.03	0.14 ± 0.03
ATP maintenance	2.30 ± 0.23	2.30 ± 0.23	2.30 ± 0.23



## Chapter 6: *Pichia pastoris* central carbon metabolites adaptability<sup>e</sup>

---

<sup>e</sup> Manuscript in preparation as: Marc Carnicer, Krisitn Baumann, Angela ten Pierick, Zhen Zeng, Reza M. Seifar, Jan van Dam, Joseph J. Heijnen, Pau Ferrer, Walter van Gulik and Joan Albiol. ***Pichia pastoris* central carbon metabolism adaptation to recombinant protein production and oxygen availability.**

## Background

Systematic global quantitative analysis provides a unique platform to characterize the impact of environmental and genetic perturbations on the cell's physiology, also enabling to assess the role of different levels of regulation on the overall metabolic phenotype of an organism. In addition, the study of metabolic network properties and their regulation *in vivo* is a key field in systems biology and, together with proteomic, transcriptomic or fluxomic studies, has become a tool towards strain optimization [9–12]. Between the main omics technologies, metabolomics is expected to play a significant role in bridging the phenotype–genotype gap, since it amplifies changes in the proteome and provides a better representation of the phenotype of an organism than other methods [120].

Recently, we have identified a beneficial effect of low oxygen availability on the expression of a human Fab fragment in *Pichia pastoris* [100]. Furthermore, transcriptional profiling analyses revealed that the availability of oxygen was strongly affecting ergosterol biosynthesis, central carbon metabolism and stress responses, in particular the unfolded protein response [121]. However, no specific analysis of the oxygen availability and the foreign protein secretion impact on the intracellular metabolite level has been performed.

Therefore, in this study, a wide range metabolomic analysis of different central carbon metabolites were carried out at different oxygenation conditions using a recombinant *P. pastoris* strain secreting a Fab antibody fragment. The results were compared to its reference (non-producing) strain. Moreover, in order to gain further insights on the mechanisms of metabolome's adaptation to limited oxygen availability and recombinant protein production conditions, integration of transcriptomic data from previous analogous studies from our group will be explored.

Overall, this study aimed at improving the understanding of the tight regulation mechanisms governing the variation in intracellular metabolite levels of the central carbon metabolism as a result of the different oxygenation conditions employed, as well as assessing the potential impact of heterologous protein secretion on *P. pastoris*' metabolome.

## Materials and Methods

### ***Strain and cultivation conditions***

Analytical grade reagents were supplied by Sigma-Aldrich. HPLC-grade methanol and ethanol were supplied by J.T. Baker.

In this study, the *P. pastoris* strain X-33 pGAPZ $\alpha$ A Fab3H6 [39], secreting a the light and heavy chains of a human monoclonal antibody Fab fragment under the constitutive GAP promoter and the *S. cerevisiae* alpha-mating factor leader, was used as expressing strain. Values from a *P. pastoris* wild type strain grown in equivalent conditions were used for comparison reasons (Chapter 5).

The cultivation conditions and the culture media used are described in Chapter 5. Briefly, two glucose-limited chemostat cultures at  $0.1 \text{ h}^{-1}$  at different oxygenation conditions were carried out changing the oxygen enrichment of the inlet gas. Initially, the oxygen concentration in the inlet gas stream corresponded to normal air (therefore 20.95 % v/v) leading to a totally normoxic condition (i.e.  $p\text{O}_2 > 20$  %, fully respiratory metabolism). Inlet gas oxygen levels were subsequently stepwise reduced by replacing different air proportions with nitrogen. Thereby creating either oxygen limited or hypoxic conditions in the bioreactor which are characterized by different ethanol and arabinitol production rates [47, 52, 76]. In order to have biological replicates, two chemostat cultures for each oxygen condition were performed. Besides, on the different cell cultures performed, the experimental data consistency was verified applying standard procedures to elemental conservation relations as constraints [38]. For all chemostat cultivations, the statistical consistency test was passed at 95% confidence level, indicating that there was no proof for gross measurement errors.

### **Sampling**

The different chemostat experimental conditions were maintained for 5 residence times before sampling. As previously reported, this cultivation time is enough to reach a metabolic steady state in *P. pastoris* [105]. For each steady state condition, duplicate samples for intracellular metabolite measurement were taken using the previously described optimized protocol for the direct measurement of *P. pastoris* metabolome [105].

### **Metabolite analysis**

The concentrations of the principal metabolites from *P. pastoris* metabolome were quantified using two different analysis platforms, LC-MS/MS and GC-MS. G6P, F6P, FBP, 2PG+3PG, PEP, citrate,  $\alpha$ KG, succinate, fumarate, malate, G1P, 6PG, M6P, T6P, G3P S7P and UDP-glc were determined by anion-exchange LC-MS/MS [30]. AMP, ADP and ATP were determined by ion-pair reverse-phase LC-MS/MS [106]. Intracellular glucose, 3PG, 2PG, DHAP, R5P, Rbu5P, Xu5P, E4P, Trehalose and isocitrate were determined by GC-MS [107]. Quantification was based on isotope dilution mass spectrometry (IDMS) [29]. In total, 33 metabolites, with a wide variety of chemical and physical properties, were analyzed. However, only 30 of these could actually be determined because the concentrations of glyoxylate, mannitol-1-phosphate and fructose-2,6-bisphosphate were found to be below the detection limit in all the analyzed samples. Moreover, it is worth to be noticed that the 2PG and 3PG were quantified together by LC-MS but also separately by GC-MS.

### **Network-embedded thermodynamic (NET) analysis**

The network-embedded analyses were performed using the cytosol model described in Chapter 5. Briefly, cytosol was the only cellular compartment considered in the analysis as no information on the specific volumes of the different cellular compartments was available.

The feasible ranges of the quantified intracellular metabolites were calculated using the respective experimental mean and standard deviation. The lower and upper limits were defined according to a confidence interval of 0.80 of a Student-T distribution. In addition, the unmeasured metabolites were specifically constrained as performed in previous work to do not lose thermodynamic restriction [96, 103] (Chapter 5).

For the *P. pastoris* cytosol, the assumed ionic strength was 0.15 M while the pH was assumed to be 7.2 (D. Mattanovich personal communication). This methodology of NET analysis has been programmed in the 'anNET' software developed by Zamboni and co-workers [104].

### **Metabolic Flux Analysis (MFA)**

Intracellular flux distributions were calculated from the measured specific production and consumption rates (see Results section) using the improved stoichiometric model and global approach described in Chapter 5, the MFA-5.

### **Regulation Theory**

It is well known that the metabolic fluxes can be regulated at hierachycal level by changes in enzyme capacities resulting from changes on gene expression and mRNA translational modifications. Moreover, each metabolic rate can also be regulated at metabolic level as a result of altered metabolite levels. In order to evaluate the contribution of each of those regulatory levels on the final metabolic flux, Daran-Lapujade and co-workers [17] studied how the mRNA, protein,  $V_{max}$  and metabolic fluxes correlated under different culture conditions in *S. cerevisiae*. Equation 6.1 summarizes the global regulatory analysis formalism:

$$\frac{\Delta \log J}{\Delta \log J} = \frac{\Delta \log f(\mathbf{e})}{\Delta \log J} + \frac{\Delta \log g(\mathbf{X}, \mathbf{K})}{\Delta \log J} = 1 \quad (6.1)$$

Where  $\mathbf{J}$  is the metabolic flux at steady state at each specific metabolic reaction,  $\mathbf{e}$  is the enzyme concentration;  $\mathbf{X}$  is a vector of the concentration of the different metabolites which may affect the reaction and  $\mathbf{K}$  is a vector including all the metabolic reaction parameters. The specific enzyme capacity, taken as function of the enzyme concentration, can be in turn decomposed by taking into account separately the possible transcriptional and translational regulatory levels as shown in equation 6.2 (where the original log expression has been changed for  $\log_2$  with no loss in generality):

$$\frac{\Delta \log_2 J_i}{\Delta \log_2 J_i} = \frac{\Delta \log_2 [mRNA_i]}{\Delta \log_2 J_i} + \frac{\Delta \log_2 \left( \frac{V_{max,i}}{[mRNA_i]} \right)}{\Delta \log_2 J_i} + \frac{\Delta \log_2 g_i(X, K)}{\Delta \log_2 J_i} = \rho_{i,mRNA,flux} + \rho_{i,translati on,flux} + \rho_{i,met,flux} = 1 \quad (6.2)$$

Where  $[mRNA_i]$  and  $V_{max,i}$  represent the mRNA concentrations and the enzyme capacity of the  $i$  reaction. The  $\rho_{i,mRNA,flux}$  represent the changes in mRNA amounts. The  $\rho_{i,translati on,flux}$  shows the enzyme capacities modifications not related to the transcriptional level (i.e. mRNA processing or protein stability). The  $\rho_{i,met,flux}$  includes all metabolic effects on the specific  $i$  reaction.

Several approaches have been reported for the determination of the metabolite levels impact on the metabolic fluxes based on different existing kinetic approximations relating intracellular metabolite levels with metabolic fluxes (reviewed by Heijnen,[19]). Regardless of the approximation used, these modeling approaches share several kinetic assumptions (Eq. 6.3).

$$\frac{J_i}{e_i} = k_{i,cat} * \prod_j^n \left( \frac{X_j}{P_j + X_j} \right) * h_i(Y, Q) * \left( 1 - \frac{\Gamma_i}{K_{i,eq}} \right) \quad (6.3)$$

Where  $X_j$  and  $P_j$  are the different substrates and kinetic parameters which influence the reaction catalyzed by the enzyme  $i$ ,  $e_i$ . The  $h$  function determines the allosteric regulation of the  $e_i$  by the  $Y$  metabolites with the  $Q$  parameters. The  $K_{i,eq}$  show the *in vivo* equilibrium constant and  $\Gamma_i$  the mass action ratio of the enzyme  $i$ .

Thereafter, the product operator term is named metabolic substrate regulation, the  $h_i$  function is referred as metabolic allosteric regulation and the thermodynamic multiplicative term as thermodynamic force. This equation intends to be equivalent to the  $g$  function in equation 6.2. Therefore, following the same regulatory strategy, the different factor which may affect the final flux could be extended subdividing the metabolic regulation in substrate, allosteric and thermodynamic regulation (Eq. 6.4).

$$\frac{\Delta \log_2 g_i(X, K)}{\Delta \log_2 J_i} = \frac{\Delta \log_2 \prod_j^n \left( \frac{X_j}{P_j + X_j} \right)}{\Delta \log_2 J_i} + \frac{\Delta \log_2 h_i(Y, Q)}{\Delta \log_2 J_i} + \frac{\Delta \log_2 \left( 1 - \frac{\Gamma_i}{K_{i,eq}} \right)}{\Delta \log_2 J_i} \quad (6.4)$$

In order to graphically visualize the obtained intracellular quantification together with the central carbon metabolism and amino acid biosynthetic pathways the VANTED software was used [122].

## Results and Discussion

Nowadays, *P. pastoris* is recognized as a well established and powerful expression system for many recombinant protein production processes. Nevertheless, this host still possesses a number of limitations, particularly those related to folding and conformational stress due to protein over-

secretion, which eventually may impose a metabolic burden to the cell [46]. It is therefore of interest to investigate potential interactions between folding stress and central metabolism.

In this study, a model protein, the 3H6 antibody fragment [39], was over-expressed under the control of a constitutive pGAP promoter at different oxygen availability conditions. This experimental setup was analogous to the previously used by Baumann and co-workers in a recent transcriptomic, proteomic and fluxomic profiling study of this organism [52].

As expected, the specific Fab productivities obtained in the present study at different oxygen conditions were higher at lower levels of oxygen availability (Table 1), and were in the range of the previously reported productivities [47, 52, 76]. Moreover, the balanced rates of the different chemostat cultures using either the control (Chapter 5) or Fab-expressing strains are shown in Table 6.1.

**Table 6.1:** Overview of the macroscopic growth parameters. The Control Strain values are taken from Chapter 5.

	Control Strain			Expressing Strain		
	Normoxic	O <sub>2</sub> -limited	Hypoxic	Normoxic	O <sub>2</sub> -limited	Hypoxic
q <sub>Fab</sub> <sup>c</sup>				40 ± 5	82 ± 2	74 ± 9
q <sub>Glc</sub> <sup>a</sup>	-1.00 ± 0.02	-1.28 ± 0.03	-1.72 ± 0.05	-1.01 ± 0.02	-1.37 ± 0.03	-1.56 ± 0.04
q <sub>O<sub>2</sub></sub> <sup>a</sup>	-2.35 ± 0.06	-2.01 ± 0.07	-2.01 ± 0.15	-2.44 ± 0.07	-1.99 ± 0.08	-1.81 ± 0.13
q <sub>CO<sub>2</sub></sub> <sup>a</sup>	2.43 ± 0.06	2.55 ± 0.06	3.21 ± 0.14	2.52 ± 0.07	2.68 ± 0.07	2.94 ± 0.12
q <sub>X</sub> <sup>a</sup>	3.57 ± 0.15	3.83 ± 0.18	3.77 ± 0.23	3.55 ± 0.15	3.77 ± 0.18	3.58 ± 0.22
q <sub>EtOH</sub> <sup>a</sup>		0.31 ± 0.02	0.84 ± 0.06		0.41 ± 0.03	0.83 ± 0.06
q <sub>Ara</sub> <sup>a</sup>		0.13 ± 0.01	0.33 ± 0.01		0.19 ± 0.01	0.24 ± 0.02
RQ <sup>b</sup>	1.03 ± 0.04	1.27 ± 0.05	1.60 ± 0.14	1.03 ± 0.04	1.34 ± 0.06	1.63 ± 0.13

<sup>a</sup> mmol/(g Biomass·h)

<sup>b</sup> mol CO<sub>2</sub>/mol O<sub>2</sub>

<sup>c</sup> µg Fab/(g Biomass·h)

The substrate and product rates obtained at normoxic and O<sub>2</sub>-limited conditions showed no significant differences compared with the corresponding rates previously reported for *P. pastoris* [52], thereby reflecting equivalent metabolic states. Conversely, the hypoxic condition tested appeared to be less stringent compared with the corresponding cultivation condition previously reported, as previously reported (Chapter 5).

### ***Thermodynamic consistency and metabolome data reliability improvement***

The intracellular metabolite quantifications have to follow the second law of thermodynamics given specific direction of the metabolic fluxes. In an initial NET analysis, the 3H6 Fab expressing strain presented thermodynamic inconsistency using the classical non-oxidative pentose phosphate pathway similar to the control *P. pastoris* strain (Chapter 5). Nevertheless, using a non classical stoichiometry in the non-oxidative pentose phosphate pathway (MFA-5; Chapter 5) the intracellular metabolite levels could be validated (See Appendix for detailed information).

In addition, the NET analysis also allowed the detection of any significant metabolite measurement error to constrain the metabolite concentration ranges. This improves the reliability of the experimental dataset. Figure 6.1 shows the obtained results from the NET analysis of some metabolites from the expressing strain (Normoxia, Oxygen-limited and Hypoxia) (See appendix for complete metabolite data set).

Globally, in 14 out of 21 measured metabolite the NET analysis could not further constrain the metabolite concentration ranges being it in agreement with the metabolic fluxes directions. This result highlights the high quality of the metabolite measurements obtained. Nevertheless, only few differences were detected between the expressing and control strain metabolome fingerprint (Figures 3-6).

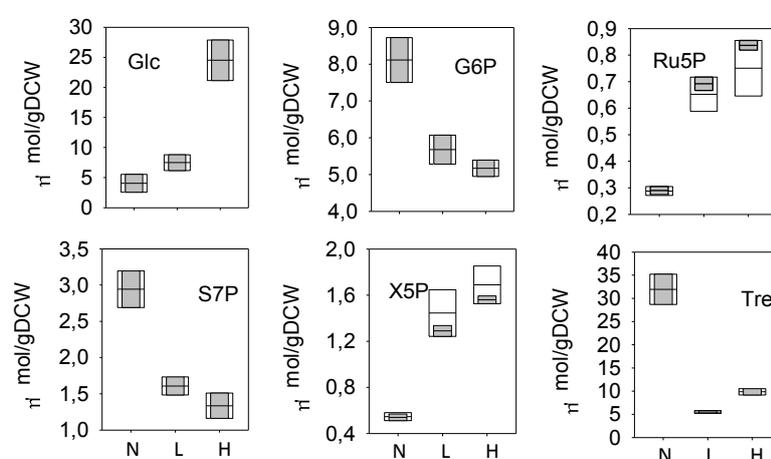


Figure 6.1: **Concentration range of the measured and the most probable ranges of the non measured ones.** The white bars represent *a priori* considered metabolite ranges while the light grey bars show the corrected values after performing a network-embedded thermodynamic (NET) analysis. N: Normoxia; L: Oxygen-limited; H: Hypoxia.

In addition, the  $\text{NAD}^+/\text{NADH}_{\text{min}}$  ratios was calculated for the Fab expressing strain as well as performed for the control strain are shown in Table 6.2. As an outcome, looking at the normoxia and the oxygen-limited conditions from both strains, the Fab production led to higher  $\text{NAD}^+/\text{NADH}_{\text{min}}$  ratios indicating a possible NADH cost of the recombinant protein production. Nevertheless, that was not the case comparing the hypoxic conditions. Thereby, no general impact could be found in the redox state of the cell regarding the Fab production (Table 6.2).

### **Thermodynamic regulation**

The NET analysis of the metabolome data set could also be used to study the substrate-product relation changes related to the thermodynamic regulation of the fluxes (Equation 6.4). In Figure 6.2 there are represented the transformed gibbs energy resulted from the NET analysis of the different metabolome data sets of *P. pastoris* including the control (Chapter 5) and expressing strain. From the metabolic

control analysis it is accepted that the reactions operating near equilibrium usually have a large sensitivity of the reaction rate towards variations in metabolite concentrations [98]. On the other hand, reactions with large values of Gibbs energy of reaction do not necessarily have an impact on flux control, however they are more likely to impose flux control being an active regulation sites [123].

**Table 6.2:** Ratios of the oxidation states at different oxygen conditions for *P. pastoris* expressing and control strain\* (Chapter 5).

		NAD <sup>+</sup> /NADH <sub>min</sub>		
		pH	7.0	7.2
Control Strain*	Normoxia	233	101	28
	Oxygen Limited	6.1	2.7	0.7
	Hypoxia	13.7	5.9	1.6
Expressing Strain	Normoxia	272	118	33
	Oxygen Limited	14.5	6.3	1.7
	Hypoxia	10.7	4.6	1.3

As expected, between the enzymes with higher transformed gibbs energy were the well know glycolysis key enzymes, PYK, PFK and HKX, as well as the two enzymes of the oxidative pentose phosphate pathway, G6PDH and 6PGDH. In addition, the enzymes involved in the ethanol fermentative pathway, PDC and ADH, also presented an active regulation together with PYRCK, Trehalase and G3PDH. Unexpectedly, TK(3) reaction, which is described as a reversible reaction, in all the conditions analyzed presented a large transformed gibbs energy suggesting only one directionality.

Looking the Equation 3, all the reactions presenting active regulation are insensible to the substrate-product ratio variation (thermodynamic force > 0.99, Equation 6.1). Therefore, any metabolic flux variation through that enzymes have to be modulated changing the catalytic activity ( $e_i \cdot k_{cat}$  in Equation 1) or due to its kinetic. Nevertheless, the reaction with transformed gibbs energy closer to zero could be influenced by the substrate-product ratio. Therefore, the thermodynamic of these reactions could have an impact over the flux as recently demonstrated to be for *S. cerevisiae* [96].

In this study, the thermodynamic regulation coefficients (Eq 4) of the reactions taken into account in the NET analysis were calculated. In the procedure, the most probable transformed Gibbs free energy for each reaction was assumed to be the value in the middle of the Gibbs energy window ranges. Since the oxygen impact on the cell physiology is much stronger than the effect of the recombinant protein production [52], the strains were considered as biological replicates to describe the *P. pastoris* oxygen limitation adaptation for the regulatory coefficient calculations. Therefore, all the oxygen conditions and strains data sets were used to describe the thermodynamic force impact over the flux.

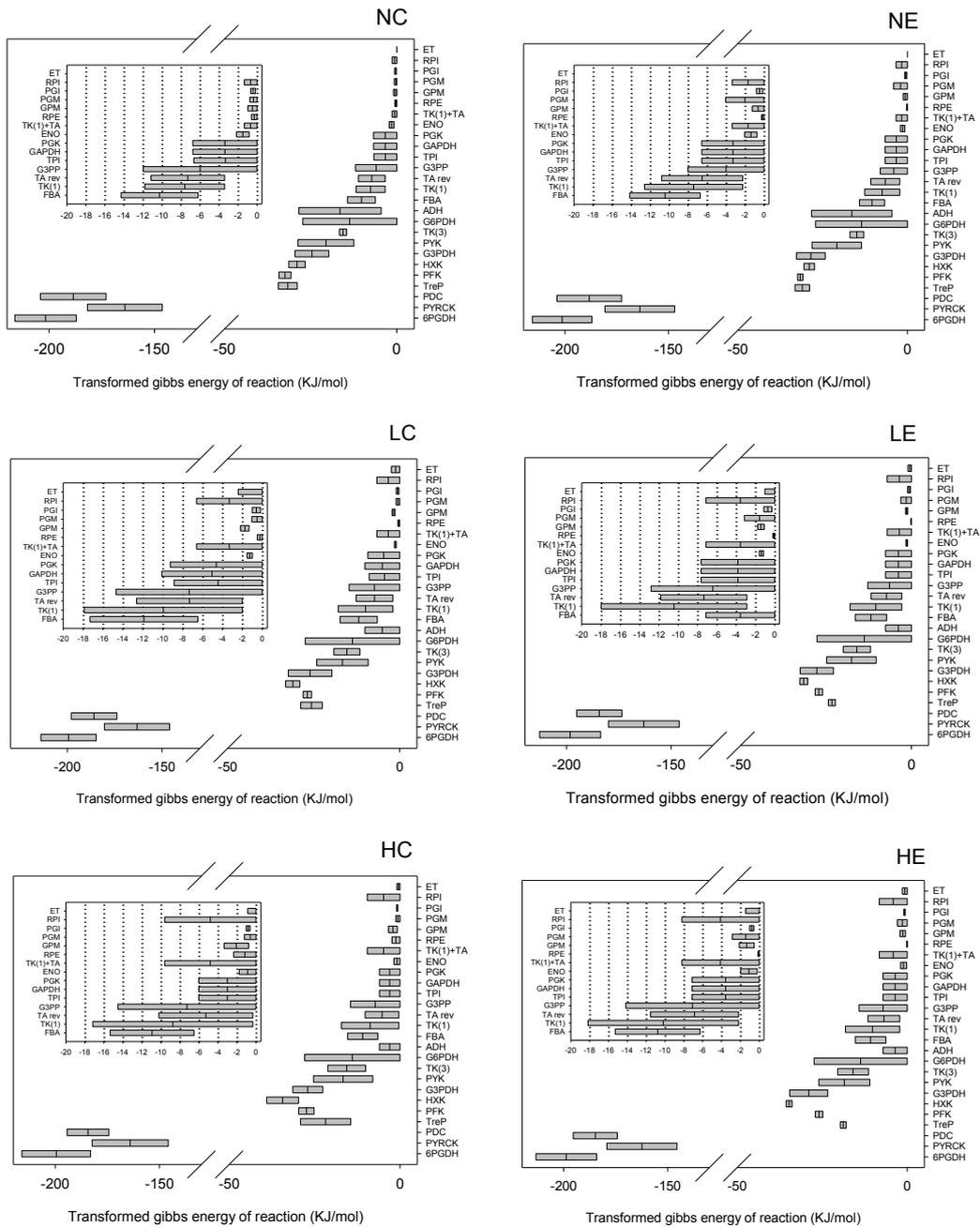


Figure 6.2: Transformed gibbs energy of the control (C) and expressing (E) strain growing under different oxygen conditions. N: Normoxia; L: Oxygen-limited; H: Hypoxia.

From the beginning, the reactions which presented in all the data sets middle Gibbs energy over 10 KJ/mol (thermodynamic force > 0.99) were discarded being them insensible to the substrate-product ratio variation (6PGDH, PYRCK, PDC, TreP, PFK, HXK, G3PDH, PYK, TK(3), TK(1) and G6PDH)

On the other side, Table 6.3 shows the thermodynamic regulation coefficients of the rest of the reaction as well as the p-value from Student-T test whether the coefficients were significantly different than zero. Significant thermodynamic regulation was considered for p-values < 0.10. For RPI and TK(1)+TA coefficient calculation only the values from normoxia and oxygen-limited conditions were

used due to different trend was found under hypoxic conditions (Appendix). Besides, the reaction classification developed for *S. cerevisiae* regarding the substrate-product ratio variation was added for comparison purpose [96].

**Table 6.3:** Thermodynamic regulation coefficients from *P. pastoris* under different oxygen availability conditions. \*Coefficient corresponding to the nomoxic – oxygen limited transition. \*\* Data taken from [96]. NE: Near-equilibrium reaction; PE: Pseudo-equilibrium reaction.

	$\rho_{\text{Thm.flux},i}$	p-value	S.c**
PGI	0.66 ± 0.08	0.001	NE
PGM	-0.24 ± 1.31	0.86	NE
FBA	0.01 ± 0.01	0.20	NE
TPI	0.20 ± 0.21	0.40	PE
GAPDH	0.21 ± 0.30	0.52	PE
PGK	0.22 ± 0.29	0.49	PE
GPM	1.90 ± 0.74	0.06	NE
ENO	-0.62 ± 0.23	0.05	NE
RPE	1.10 ± 2.27	0.65	PE
RPI*	-7.41 ± 1.61	0.04	NE
TA rev	0.02 ± 0.02	0.40	-
TK(1)+TA*	-2.48 ± 0.72	0.07	NE - PE
ADH	-0.72 ± 0.27	0.06	NE
ET	-0.42 ± 0.41	0.42	-

For the reactions belonging to the upper glycolysis, only PGI showed significant thermodynamic regulation in agreement with the metabolic flux variation (positive coefficient). However, for PGM, FBA and TPI the substrate-product variation was not significant to modify the reaction fluxes. In the lower glycolysis reactions, the first reactions of the pathway presented no thermodynamic influence counterstained to the GPM and ENO reactions. Nevertheless, opposite regulation pattern was determined comparing GPM and ENO reactions. Specifically, GPM regulation was in agreement with the metabolic flux variation (positive coefficient) although ENO thermodynamics reduced the thermodynamic force of the reaction under higher metabolic fluxes (negative coefficient). This evidence opens the possible thermodynamic regulation of the ENO reaction in *P. pastoris* for controlling the glycolysis flux.

In the Pentose Phosphate pathway, for RPE reaction the flux under oxygen limitation conditions was increased significantly due to the arabinitol production under these conditions [52]. However, no thermodynamic regulation was determined towards this product formation. Contrarily, the RPI flux was reduced as oxygen availability became lower independently of its substrate-product ratio variation which promoted higher flux. In addition, the metabolite channeling recently described for *P. pastoris* (Chapter 5) showed also a thermodynamic regulation in disagreement with the flux trend regarding the oxygen availability. These values may be explained by *P. pastoris* physiological pressures aimed to

minimize of carbon losses as arabinitol in this part of the metabolism. On the other hand, TA rev was not significantly regulated when shifting between conditions.

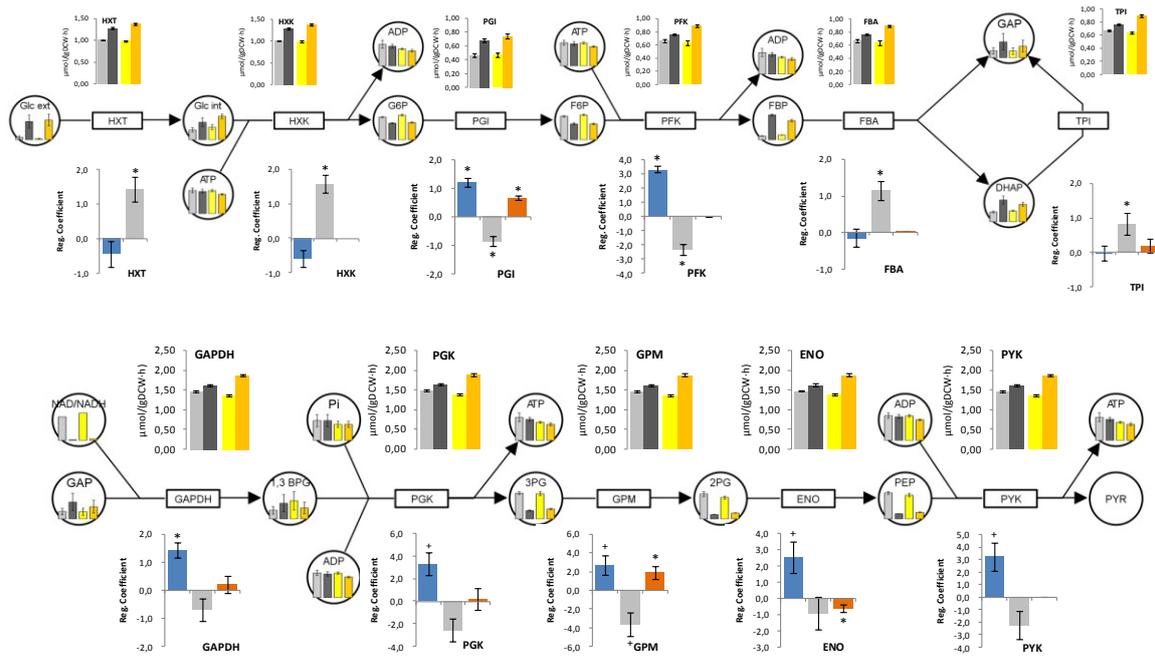
Looking the ethanol metabolic pathway, it was only active under oxygen limitation thereby, the thermodynamic coefficient in Table 6.3 showed the variation between the two conditions where the oxygen was limited. As outcomes, the ethanol diffusion through the cell membrane was not limiting the ethanol production under the different conditions due to no differences between the intracellular and extracellular pools ratio could be observed (no thermodynamic regulation of ET). Nevertheless, similarly to the arabinitol production, the cell tried to reduce the ethanol secretions through the negative regulation of the ADH reaction.

On the other hand, Table 6.3 shows also the classification developed for *S. cerevisiae* regarding the influence (near equilibrium; NE) or not (pseudo equilibrium; PE) of the substrate-product ratio on the metabolic flux changes [96]. Comparing the obtained results from this study and the recently published for *S. cerevisiae* grown under different growth rates, quite agreement could be observed between the two yeast regulation. In addition, it is well to notice that, even *S. cerevisiae* and *P. pastoris* were grown changing different culture parameter, both studies observed a glycolysis flux increase due to the ethanol production suggesting possible agreement in the metabolome impact. Thereby, giving sense to the concordance on the thermodynamic regulation of the different reactions presented in Table 6.3.

### **Global central carbon metabolism regulation**

The oxygen availability has strong impact in the central metabolism due to his crucial role as electron acceptor. In this study, it was clear that the thermodynamic regulation it was not the only mechanism controlling the metabolic fluxes under the different oxygenation conditions. In previous studies, different omics levels were studied showing significant variation between oxygen conditions [52]. Therefore, to gain an overview of the potential regulation of other levels, the  $\rho_{i,mRNA,flux}$  regulatory coefficients were calculated for the central carbon metabolism using the mRNA data from the previously performed study [52]. Since hypoxic condition of this study was not exactly equivalent to that used in previous transcriptomic studies, only normoxic and oxygen limiting conditions were compared for both expressing and control strains.

In the calculations, in order to obtain a global transcriptional regulatory coefficient for each metabolic reaction, the transcriptional levels corresponding to different isoforms of a given enzyme were pooled together (See appendix for more detailed information on the calculation). Moreover, taking into account the theory of regulation equation (Eq. 1), we could calculate which had to be the rest of omic level contribution for obtaining the proper flux ( $\rho_{i,rest,flux} = 1 - \rho_{i,mRNA,flux} - \rho_{i,Thm,flux}$ ). Figures 6.3-6.6 show the different intracellular metabolite levels of different parts of the *P. pastoris* metabolism as well as the metabolic fluxes and the calculated regulatory coefficients for each reaction.

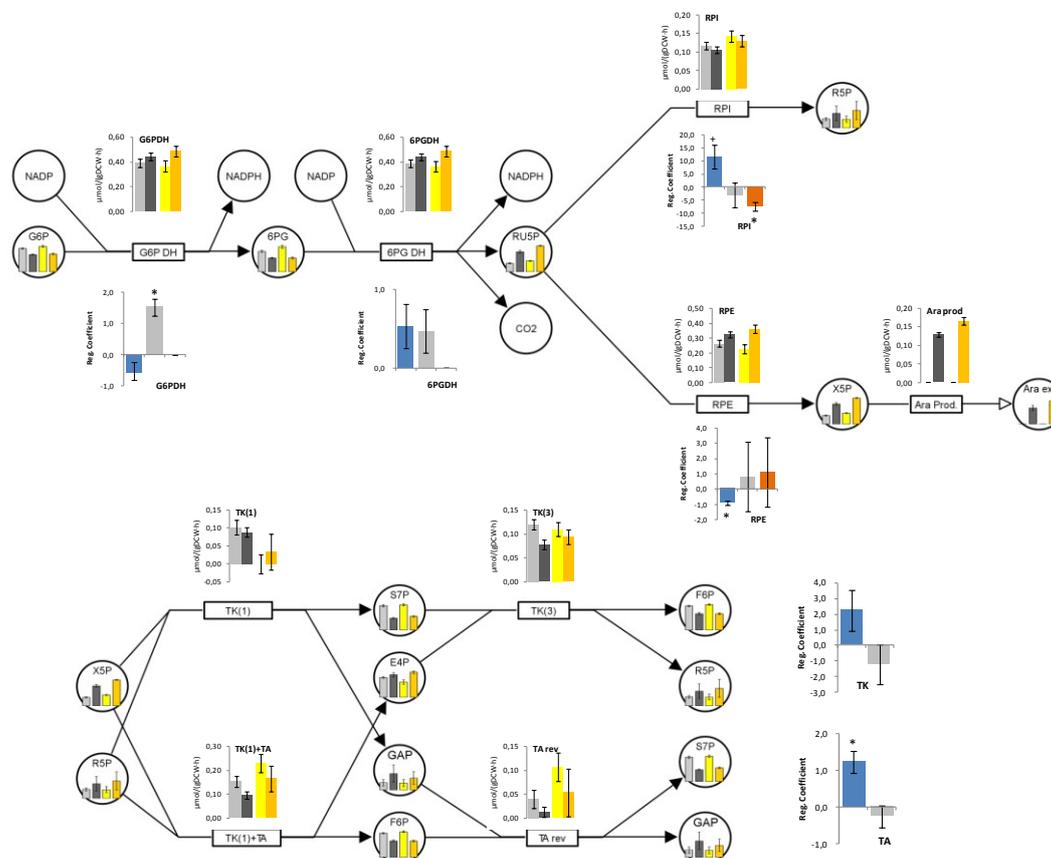


**Figure 6.3: Glycolysis *P. pastoris* metabolism.** The circles represent the different metabolites linked by the metabolic reactions. The bar diagrams inside the circles represent the intracellular metabolite quantifications. The upper bar diagrams represent the metabolic flux of each reaction and the lower bar diagram the regulation coefficients from Normoxia to Oxygen-limited condition. **Metabolic flux and intracellular metabolite graphs:** Black-Grey bars: Control strain; Yellow bars: Expressing strain; dark to light colours scale indicates Normoxic to Oxygen-limited conditions. **Regulation coefficient graphs:** Blue bars: transcriptional coefficient; Orange bars; thermodynamic coefficient; Grey bars: rest of omic level regulation. The error bars represent the SE of the quantification. \* and + symbols indicate p-values lower than 0.10 and between 0.10 and 0.15 respectively from Student-T test whether the coefficients were significantly different than zero. Each graphic has its own scale.

### Glycolysis

Daran-Lapujade and co-workers concluded that the glycolytic flux in *S. cerevisiae* is predominantly regulated at posttranscriptional level [17]. Focusing in Figures 6.3 and as already pointed out in previous studies [52, 124], *P. pastoris* seems to behave differently. The two first reactions of the glycolysis, the HXT and HXK, were mainly regulated by changes not related with the transcriptional or thermodynamic level. However, for the PGI reaction, its metabolic flux change was mediated positively (in agreement with the metabolic flux changes) by changes in the PGI mRNA levels and by changes in the F6P/G6P ratios. In addition, the PGI had also a negative contribution to the final flux changes (oppositely to the metabolic flux changes) from the rest of omic levels. For the PFK reaction, the regulation was similar to the PGI although, for this reaction, no thermodynamic regulation was determined due to the middle Gibbs energy were over 10 KJ/mol in all the conditions. On the other hand, the FBA and TPI reactions did not present significant transcriptional or thermodynamic regulation.

On the lower glycolysis, although only was strictly significant for the GAPDH (p-value < 0.10), a general positive agreement between the metabolic flux changes and the mRNA levels could be determined. On the other side, the thermodynamic regulation, as seen in previous section, only was significant in the GPM and ENO reactions.



**Figure 6.4: Pentose phosphate pathway *P. pastoris* metabolism.** The circles represent the different metabolites linked by the metabolic reactions. The bar diagrams inside the circles represent the intracellular metabolite quantifications. The upper bar diagrams represent the metabolic flux of each reaction and the lower bar diagram the regulation coefficients from Normoxia to Oxygen-limited condition. **Metabolic flux and intracellular metabolite graphs:** Black-Grey bars: Control strain; Yellow bars: Expressing strain; dark to light colours scale indicates Normoxic to Oxygen-limited conditions. **Regulation coefficient graphs:** Blue bars: transcriptional coefficient; Orange bars; thermodynamic coefficient; Grey bars: rest of omic level regulation. The error bars represent the SE of the quantification. \* and + symbols indicate p-values lower than 0.10 and between 0.10 and 0.15 respectively from Student-T test whether the coefficients were significantly different than zero. Each graphic has its own scale.

Interestingly, when we looked at the reaction with significant regulation from the rest of omic levels, a positive agreement between the intracellular substrate levels trend and the rest of regulation contribution was seen allowing a possible kinetic contribution to the metabolic flux changes. For

example, in the FBA reaction, the significant positive contribution from the rest of the omics level agrees with the significant increase of FBP availability for the FBA enzyme.

### Pentose phosphate pathway

The pentose phosphate pathway was highly affected by the arabinitol production characteristic of *P. pastoris* growing at low oxygen supply [52, 76]. The oxidative part of the pathway is the major source of NADPH production for biomass biosynthesis in *P. pastoris*. Besides, the G6PDH and 6PGDH are enzymes highly allosterically regulated by the intracellular levels of NADPH and ATP which could be related with the significant regulation of the rest of omic levels in the G6PDH reaction (Figure 6.4).

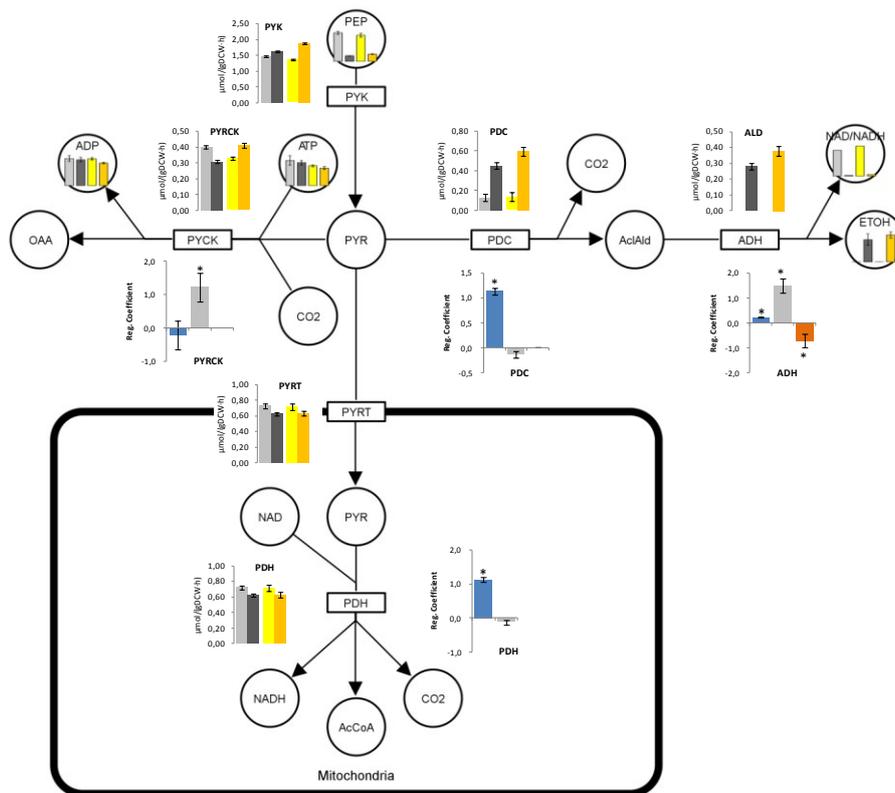


Figure 6.5: **Pyruvate node *P. pastoris* metabolism.** The circles represent the different metabolites linked by the metabolic reactions. The bar diagrams inside the circles represent the intracellular metabolite quantifications. The upper bar diagrams represent the metabolic flux of each reaction and the lower bar diagram the regulation coefficients from Normoxia to Oxygen-limited condition. **Metabolic flux and intracellular metabolite graphs:** Black-Grey bars: Control strain; Yellow bars: Expressing strain; dark to light colours scale indicates Normoxic to Oxygen-limited conditions. **Regulation coefficient graphs:** Blue bars: transcriptional coefficient; Orange bars; thermodynamic coefficient; Grey bars: rest of omic level regulation. The error bars represent the SE of the quantification. \* and + symbols indicate p-values lower than 0.10 and between 0.10 and 0.15 respectively from Student-T test whether the coefficients were significantly different than zero. Each graphic has its own scale.

In the non oxidative pentose phosphate pathway, the RPI and the RPE are the enzymes which connect the 3 pentose phosphate metabolites. Interestingly, they presented opposite transcriptional regulation,

being for RPI positive and for RPE negative, entailing a physiological pressure for reducing the metabolic flux to the arabinitol production. In addition, looking the variation of the intracellular metabolite levels of the 3 pentoses, they presented the same trend as the extracellular arabinitol suggesting that these pools are highly affected by the cell needs of NADH re-oxidation (Figure 6.4).

For TK and TA regulation coefficient calculation, all the reaction where these enzymes were involved were taken into account from where only a significant transcriptional regulation could be determined for the TA enzyme as well as the already explained thermodynamic influence in the TK(1)+TA reaction (Table 6.3).

#### Pyruvate node

In the pyruvate node (Figure 6.5), there was a clear impact of the ethanol production under oxygen limited condition. The first enzyme of the alcoholic fermentation pathway, the PDC, was found to be highly transcriptionally induced, meanwhile the ADH reaction presented significant transcriptional and thermodynamic regulation as well as from the other omic levels. On the other hand, the anapleratic pathway which produced cytosolic oxalacetate seemed to be not significantly influenced by the PYRCK mRNA levels being the other omic levels the main source of variation.

In previous work, the decrease of the pyruvate import to the mitochondria was already described due to the lower oxygen availability [100] which regulate the transcriptional levels of the PDH reaction (Figure 6.5). Unfortunately, no pyr specific intracellular metabolite quantifications could be determined for *P. pastoris* [105] which would have been important in the understanding of the intracellular metabolites trends in this part of the metabolism.

#### TCA cycle

The TCA cycle was highly influenced by the oxygen availability in *P. pastoris* as already described by Baumann and co-workers [52]. Overall, the metabolic flux changes from normoxic to oxygen limited conditions were significant mediated by changes in the mRNA levels of the different enzymes involved in the pathway (Figure 6). However, there was a clear exception, the ICITDH which presented a significant regulation from the rest of omic levels. Unfortunately, no thermodynamic regulation could be studied due to the lack of specific distribution of many TCA metabolites between the different cellular compartments.

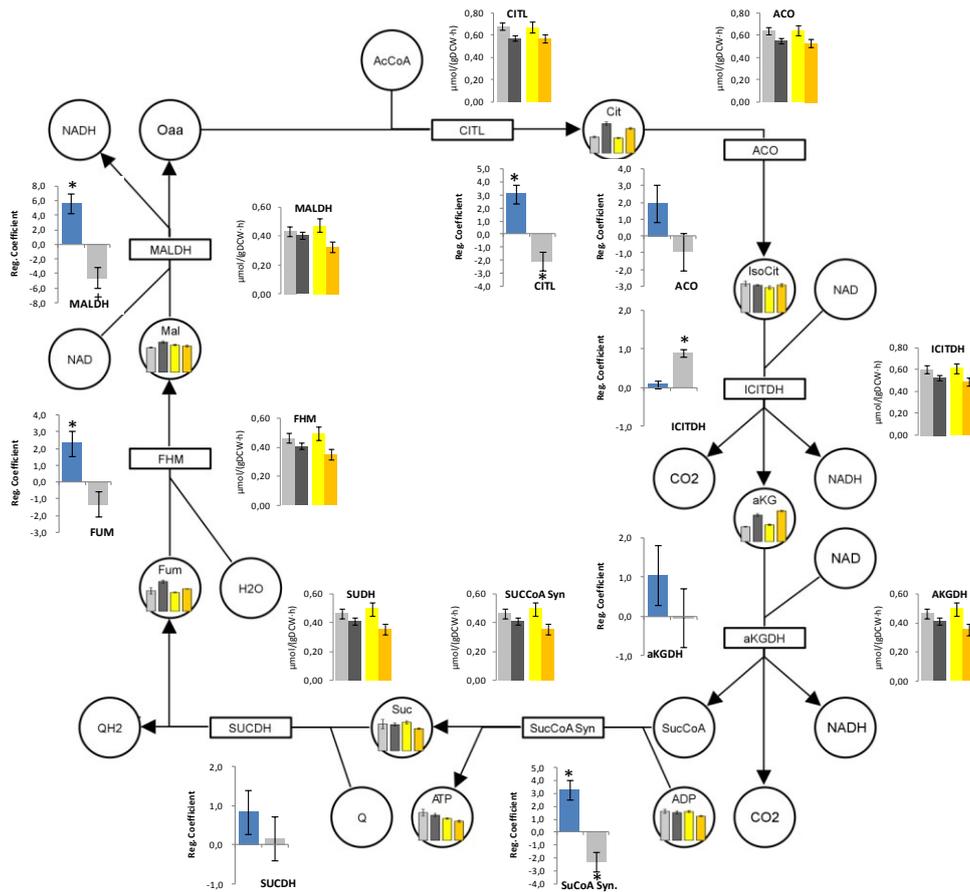


Figure 6.6: **TCA cycle metabolic pathway of *P. pastoris***. The circles represent the different metabolites linked by the metabolic reactions. The bar diagrams inside the circles represent the intracellular metabolite quantifications. The upper bar diagrams represent the metabolic flux of each reaction and the lower bar diagram the regulation coefficients from Normoxia to Oxygen-limited condition. **Metabolic flux and intracellular metabolite graphs:** Black-Grey bars: Control strain; Yellow bars: Expressing strain; dark to light colours scale indicates Normoxic to Oxygen-limited conditions. **Regulation coefficient graphs:** Blue bars: transcriptional coefficient; Grey bars: rest of omic level regulation. The error bars represent the SE of the quantification. \* and + symbols indicate p-values lower than 0.10 and between 0.10 and 0.15 respectively from Student-T test whether the coefficients were significantly different than zero. Each graphic has its own scale.

## Conclusions

In this study, a physiological explanation of the intracellular metabolite fingerprint under different experimental conditions, combining different “omic” level analyses, was obtained by a multi-level analysis of *Pichia pastoris*. It demonstrates the importance of the system biology approach in order to understand the cell physiological behavior under different environmental conditions. In addition, the understanding of how these metabolite levels could regulate the different intracellular metabolite fluxes could help in future strain improvements.

Overall, the results obtained by combining different “omics” analyses provide a metabolic fingerprint of the oxygen availability impact and recombinant protein production, thus improving the physiological knowledge of biotechnologically relevant yeast such as *P. pastoris*.

**Appendix 6.1.** Physiological parameter comparison

The chemostat cultures performed in this study intended to simulate the same environmental conditions as the ones used in previous works working using the same strains [47, 52, 76]. In Table S6.1, it is shown the consumption and production rates calculated in this study as well as the ones quantified in the Baumann and co-workers study [52]. In both studies, the *P. pastoris* cells were grown in constant environmental conditions for, at least, 5 residence times. Later on, the microorganisms were step wise oxygen constrained by means of reducing the oxygen concentration in the inlet gas flow inducing a fermentative metabolism as could be verified by the arabinitol and ethanol production. Besides, the respiratory coefficient was also an indicator of the oxygen limitation environment as this parameter is highly influenced by the oxygen availability in the reactor.

For all the normoxic conditions, the respiratory quotient (RQ) ratios were close to unit as expected from a totally respiratory metabolism indicating high reproducibility of that environmental condition. For the oxygen-limited conditions, the RQ ratios were around the 1.2-1.3 entailing some oxygen constriction.

On the other hand, the hypoxic conditions resulted in a completely different oxygen limitation environment as could be observed comparing the RQ ratios. The hypoxic cultures from Baumann and co-workers [52] not only presented a stronger oxygen limitation as could be observed in the higher RQ ratios, but also the cultures could only be run for 3.5 residence times. Overall, only the normoxic and oxygen-limited conditions had comparable rates which were not the case for hypoxic conditions.

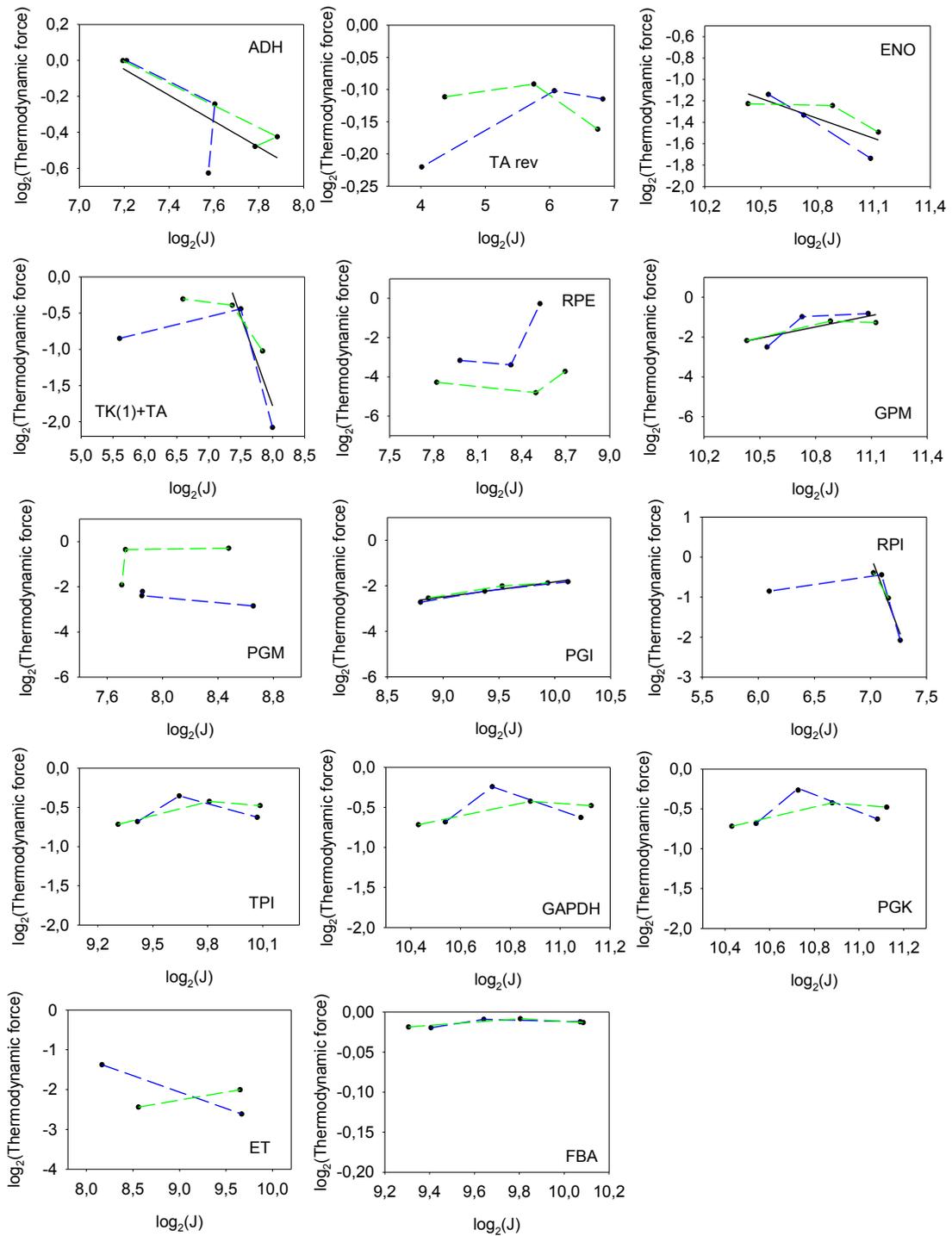
**Table S6.1:** Physiological parameter comparison of *P. pastoris* strains under different oxygen conditions. The rates values were calculated in mmol/(gDCW\*h).

mmol/(gDCW*h)	Control Strain			Expressing Strain			
	Normoxic	O <sub>2</sub> -limited	Hypoxic	Normoxic	O <sub>2</sub> -limited	Hypoxic	
This Study	q <sub>Glc</sub>	-1.00 ± 0.02	-1.28 ± 0.03	-1.72 ± 0.05	-1.01 ± 0.02	-1.37 ± 0.03	-1.56 ± 0.04
	q <sub>O<sub>2</sub></sub>	-2.35 ± 0.06	-2.01 ± 0.07	-2.01 ± 0.15	-2.44 ± 0.07	-1.99 ± 0.08	-1.81 ± 0.13
	q <sub>CO<sub>2</sub></sub>	2.43 ± 0.06	2.55 ± 0.06	3.21 ± 0.14	2.52 ± 0.07	2.68 ± 0.07	2.94 ± 0.12
	q <sub>X</sub>	3.57 ± 0.15	3.83 ± 0.18	3.77 ± 0.23	3.55 ± 0.15	3.77 ± 0.18	3.58 ± 0.22
	q <sub>E<sub>10</sub>H</sub>		0.31 ± 0.02	0.84 ± 0.06		0.41 ± 0.03	0.83 ± 0.06
	q <sub>Ara</sub>		0.13 ± 0.01	0.33 ± 0.01		0.19 ± 0.01	0.24 ± 0.02
	RQ	1.03 ± 0.04	1.27 ± 0.05	1.60 ± 0.14	1.03 ± 0.04	1.34 ± 0.06	1.63 ± 0.13
Baumann et al. 2010	q <sub>Glc</sub>	-0.99 ± 0.04	-1.27 ± 0.05	-1.85 ± 0.08	-0.95 ± 0.04	-1.31 ± 0.05	-1.69 ± 0.04
	q <sub>O<sub>2</sub></sub>	-2.20 ± 0.12	-1.79 ± 0.24	-0.28 ± 0.01	-2.26 ± 0.12	-1.54 ± 0.06	-0.44 ± 0.07
	q <sub>CO<sub>2</sub></sub>	2.27 ± 0.12	2.18 ± 0.24	1.71 ± 0.07	2.29 ± 0.12	2.09 ± 0.06	1.73 ± 0.06
	q <sub>X</sub>	3.61 ± 0.24	3.94 ± 0.30	3.92 ± 0.35	3.25 ± 0.22	3.76 ± 0.27	3.75 ± 0.18
	q <sub>E<sub>10</sub>H</sub>		0.31 ± 0.02	1.16 ± 0.06		0.33 ± 0.02	0.98 ± 0.05
	q <sub>Ara</sub>		0.10 ± 0.00	0.48 ± 0.09		0.21 ± 0.01	0.42 ± 0.02
	RQ	1.03 ± 0.08	1.22 ± 0.21	6.14 ± 0.40	1.01 ± 0.08	1.36 ± 0.07	3.90 ± 0.62

**Appendix 6.2.** Thermodynamic regulation coefficient calculation

The thermodynamic regulation coefficient was defined as the contribution of the thermodynamic force changes to the final differential flux. In Figure S1, the  $\log_2 J_i$  versus the  $\log_2$  (thermodynamic force) for each reaction is represented. The data used for these graphics was taken from the normoxic and oxygen-limited conditions of the expressing and control strains as was assumed to be biological replicates (See main manuscript). Nevertheless, despite it was assumed equal regulation between the strains, in each graphic the data points coming from the same cells were connected with a differential dashed line, blue for the control strain and green for the expressing.

The thermodynamic regulation coefficients ( $\rho_{i,Thm,flux}$ ) were determined from a lineal regression. In addition, the p-values from a two-tailed t-test to evaluate if the coefficients were equal to zero were calculated allowing to determine the confidence of the coefficient. In Figure S1, for enzymes with a p-value lower than 0.1, a lineal regression was depicted over the dashed lines from each strain. In those cases, the lineal regression equation and the dashed lines were qualitatively compared to validate equal regulation between the strains. As a result, 6 out of 15 reactions analyzed showed significant thermodynamic regulation regarding the oxygen availability.

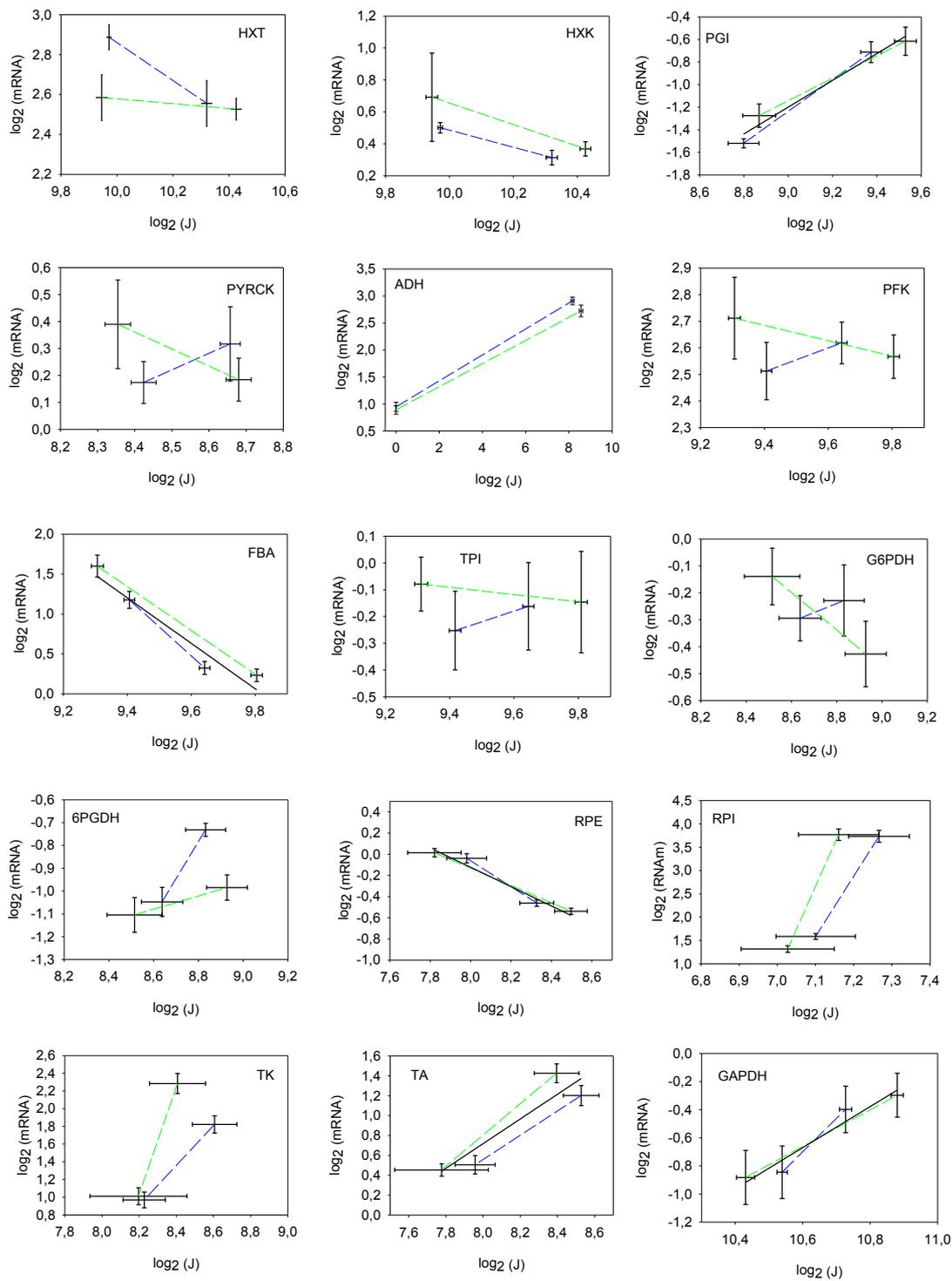


**Figure S6.1:** Thermodynamic regulation coefficients of the central carbon metabolism enzymes of *P. pastoris*. Blue dashed line = control strain data; Green dashed line = expressing strain data; Black straight line = linear regression equation.

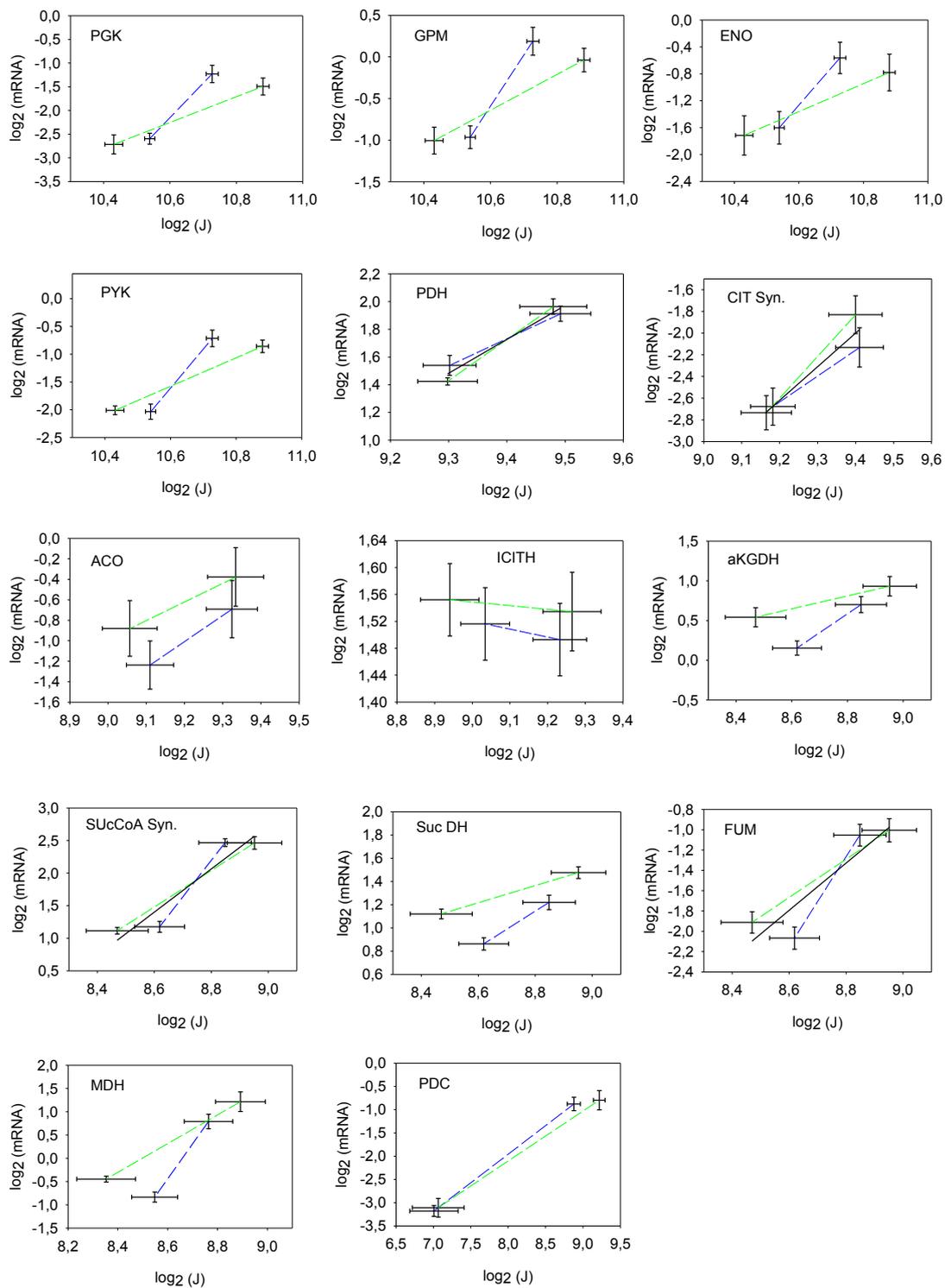
**Appendix 6.3.** Transcriptional regulation coefficient calculation

The transcriptional regulation coefficient was defined as the contribution of the mRNA changes to the final differential flux. In Figure S6.2 the  $\log_2 J_i$  versus the  $\log_2 [\text{mRNA}_i]$  for each reaction is represented. The data used for these graphics was taken from the normoxic and oxygen-limited conditions of the expressing and control strains as was assumed to be biological replicates (See main manuscript). Nevertheless, despite it was assumed equal regulation between the strains, in each graphic the data points coming from the same cells were connected with a differential dashed line, blue for the control strain and green for the expressing.

The mRNA regulation coefficients ( $\rho_{i,\text{mRNA},\text{flux}}$ ) were determined from a lineal regression. In addition, the p-values from a two-tailed t-test to evaluate if the coefficients were equal to zero were calculated allowing to determine the confidence of the coefficient. In Figure S6.2, for enzymes with a p-value lower than 0.1, a lineal regression was depicted over the dashed lines from each strain. In those cases, the lineal regression equation and the dashed lines were qualitatively compared to validate equal regulation between the strains. As a result, 10 out of 29 enzymes analyzed showed significant transcriptional regulation regarding the oxygen availability.



**Figure S6.2:** Transcriptional regulation coefficients of the central carbon metabolism enzymes of *P. pastoris*. Blue dashed line = control strain data; Green dashed line = expressing strain data; Black straight line = linear regression equation.



**Figure S2 continued:** Transcriptional regulation coefficients of the central carbon metabolism enzymes of *P. pastoris*. Blue dashed line = control strain data; Green dashed line = expressing strain data; Black straight line = lineal regression equation.



## Chapter 7: *Pichia pastoris* free amino acids adaptability<sup>f</sup>

---

<sup>f</sup> Manuscript submitted as: Marc Carnicer, Angela ten Pierick, Jan van Dam, Joseph J. Heijnen, Joan Albiol, Walter van Gulik and Pau Ferrer. **Quantitative metabolomics analysis of amino acid metabolism in recombinant *Pichia pastoris* under different oxygen availability conditions.**



## Background

Recombinant protein overproduction often results in a metabolic burden. Such effect may be reflected on process parameters such maximum growth rate, biomass yield or specific substrate consumption of yeast cells [125–129], thus suggesting a potential impact on the cell's energy metabolism, possibly derived from higher maintenance requirements [130]. Furthermore, production of recombinant proteins may cause cellular stress due to unfolded proteins and unsuitable or inefficient secretion [46], which, in turn, may negatively affect cell growth, even at relatively low expression levels [127, 131], that is, at the product yields range where effects derived from increased energy and precursor demands for protein synthesis on cell growth should be theoretically negligible. Interestingly, limited but significant alterations in the carbon flux distribution over the central metabolism have been recently shown [132–135]. Furthermore, amino acid supplementation of the growth medium has been shown to partially unburden cellular metabolism during recombinant protein production in yeast [134, 136–138]. Interestingly, recent amino acid supplementation studies in *P. pastoris* has provided novel evidence that the adaptation of the central metabolism to recombinant protein production can not only be explained by an increased drain of precursors for protein synthesis [134]. Indeed, amino acids are not only important precursors for protein synthesis but also participate in the regulation of major metabolic pathways. Glutamic and Aspartic acid for instance, are components of the aspartate/malate shuttle [82] and their concentrations control the rate of oxidation of glycolytic NADH.

We have previously reported the beneficial impact of hypoxia conditions on recombinant protein production [47]. The physiological bases of this beneficial effect were further investigated in a recent multilevel study including transcriptome, proteome and metabolic flux analyses [52]. These studies allowed to gather information on the biological processes involved in the adaptation to hypoxia and their relation with extracellular recombinant protein production in *P. pastoris*. However, the potential effect of oxygen availability and/or recombinant protein production on the intracellular metabolite levels (particularly of the amino acid precursors used for protein synthesis) remains to be investigated.

This chapter investigates the potential impact of foreign protein expression and secretion on amino acid metabolism, the free intracellular amino acids pools were analyzed in carbon limited chemostat cultivations at a fixed growth rate and different oxygenation conditions, using a recombinant *P. pastoris* strain secreting a Fab antibody fragment. The results are compared to the reference (non-producing) strain. Moreover, in order to gain further insights in the regulation of amino acid biosynthetic pathways, integration of transcriptome data from previous analogous studies from our group have also been explored. Overall, this study aimed at improving the understanding of the regulation mechanisms governing the variation in intracellular amino acid levels as a result of the different oxygenation conditions employed, as well as gaining further insight in the potential interactions between energy

metabolism and amino acid metabolism and, how such interactions may be perturbed by heterologous protein secretion in *P. pastoris*.

## Methods

### ***Strain and cultivation conditions***

Analytical grade reagents were supplied by Sigma-Aldrich. HPLC-grade methanol and ethanol were supplied by J.T. Baker.

In this study, the *P. pastoris* strain X-33 pGAPZαA Fab3H6 [39], secreting the light and heavy chains of a human monoclonal antibody Fab fragment under the constitutive GAP promoter and the *S. cerevisiae* alpha-mating factor leader, was used as expressing strain. A strain with an integrated empty-vector was used as reference strain. The experimental set up was as described in [105]. Briefly, glucose-limited chemostat cultures at a dilution rate of  $0.1 \text{ h}^{-1}$  at different oxygenation conditions were carried out by changing the oxygen content of the inlet gas. Initially, the oxygen concentration in the inlet gas stream corresponded to normal air (20.95 % v/v) leading to a totally normoxic condition (i.e.  $pO_2 > 20 \%$ , fully respiratory metabolism). Inlet gas oxygen levels were subsequently stepwise reduced by replacing different air proportions with nitrogen. Thereby creating either oxygen limited or hypoxic conditions in the bioreactor which are characterized by different ethanol and arabinitol production rates [47, 52, 76]. Two chemostats were performed for each metabolic steady state and strain.

### ***Sampling***

The different chemostat experimental conditions were maintained for 5 residence times before sampling. As previously reported, this cultivation time is enough to reach a metabolic steady state in *P. pastoris* [105](Chapter 4). For each steady state condition, duplicate samples for intracellular metabolite measurement were taken using the previously described optimized protocol for the direct measurement of *P. pastoris* metabolome [105].

### ***Metabolite analysis***

The intracellular concentrations of Ala, Val, Leu, Ile, Thr, Asn, Asp, Met, Trp, His, Pro, Glu, Gln, Orn, Lys, Tyr and Phe were determined by GC-MS [33, 34]. Glycine levels were also measured in this analysis. However, the obtained results were not included as the values were found to be inconsistent [105]. Quantification was based on isotope dilution mass spectrometry (IDMS) [29]. In total, 17 amino acids were analyzed plus Glycine.

In order to graphically visualize the obtained intracellular quantification together with the central carbon metabolism and amino acid biosynthetic pathways the VANTED software was used [122].

### ***Antibody fragment quantification***

Fab amounts in soluble cell extracts and in culture broths were performed by means of a sandwich ELISA assay as previously described [47].

### ***Principal component analysis***

Reference values for each amino acid pool level were calculated as mean of all measurements independently of the strain or oxygen availability. Afterwards, for each condition, the specific amino acid measurements were divided by the reference value to obtain a free scale value of its variation. These relative fold changes for each condition were illustrated through PCA using the xlstat plug-in software for Excel.

### ***Determination of energy and degree of reduction for amino acid synthesis***

The degree of reduction of each amino acid were calculates using the amino acid molecular formulas as explained elsewhere [139]. Energy costs for the biosynthesis of each amino acid were taken from previously published data for *S. cerevisiae* growing under aerobic conditions [140] due to both yeast share equal amino acid biosynthetic metabolic pathways [60, 79, 80].

## **Results and Discussion**

### ***Growth and recombinant protein secretion in recombinant *P. pastoris****

The burden caused by recombinant protein production in yeast and, in particular, *P. pastoris*, has been recently suggested to impact the central metabolism even at relatively low expression levels, i.e. where increased precursor (amino acids) demands for recombinant protein production may be negligible [134, 135, 141]. Integration of transcriptomic and metabolomic data of recombinant cells under different environmental conditions may help to understand the metabolic adaptations of the cell's central metabolism to protein production under different environmental conditions, particularly in relation to amino acid metabolism. In this study, intracellular amino acid pools of a recombinant *P. pastoris* strain expressing the 3H6 antibody fragment [39] under the control of a constitutive pGAP promoter were measured during growth in chemostat cultures under different oxygen availability conditions. The macroscopic growth parameters for both the control and Fab-producing *P. pastoris* strains during growth at three different oxygen supply rates have been recently reported elsewhere [52, 76]. A summary of these data is given in Table 7.1. Coherent with previous studies [52, 76], the adaptation from normoxic (fully aerobic) to hypoxic conditions lead to a shift from fully respiratory to respiro-fermentative metabolism, as well as increased secreted recombinant product productivities. Importantly, this experimental setup for the normoxic and oxygen-limiting conditions was analogous to

the one previously used by Baumann and co-workers in a transcriptomic, proteomic and fluxomic profiling study of this organism [52].

**Table 7.1:** Summary of macromolecular culture parameters.

	Control Strain			Expressing Strain		
	Normoxic	O <sub>2</sub> -limited	Hypoxic	Normoxic	O <sub>2</sub> -limited	Hypoxic
q <sub>Fab</sub> <sup>a</sup>				40 ± 5	82 ± 2	74 ± 9
q <sub>Glc</sub> <sup>b</sup>	-1.00 ± 0.02	-1.28 ± 0.03	-1.72 ± 0.05	-1.01 ± 0.02	-1.37 ± 0.03	-1.56 ± 0.04
q <sub>O<sub>2</sub></sub> <sup>b</sup>	-2.35 ± 0.06	-2.01 ± 0.07	-2.01 ± 0.15	-2.44 ± 0.07	-1.99 ± 0.08	-1.81 ± 0.13
q <sub>CO<sub>2</sub></sub> <sup>b</sup>	2.43 ± 0.06	2.55 ± 0.06	3.21 ± 0.14	2.52 ± 0.07	2.68 ± 0.07	2.94 ± 0.12
q <sub>X</sub> <sup>b</sup>	3.57 ± 0.15	3.83 ± 0.18	3.77 ± 0.23	3.55 ± 0.15	3.77 ± 0.18	3.58 ± 0.22
q <sub>EtOH</sub> <sup>b</sup>		0.31 ± 0.02	0.84 ± 0.06		0.41 ± 0.03	0.83 ± 0.06
q <sub>Ara</sub> <sup>b</sup>		0.13 ± 0.01	0.33 ± 0.01		0.19 ± 0.01	0.24 ± 0.02
RQ <sup>c</sup>	1.03 ± 0.04	1.27 ± 0.05	1.60 ± 0.14	1.03 ± 0.04	1.34 ± 0.06	1.63 ± 0.13

<sup>a</sup> µg Fab / (g Biomass·h)

<sup>b</sup> mmol / (g Biomass·h)

<sup>c</sup> mol CO<sub>2</sub> / mol O<sub>2</sub>

Conversely, the hypoxic condition tested appeared to be less stringent compared with the corresponding cultivation condition previously reported, as indicated by a lower RQ as well as lower specific rates of ethanol and arabinitol production. This was further supported by the fact that, in the mentioned previous studies only 3.5 residence times could be accomplished in the hypoxic condition [47], whereas in this study the hypoxic condition could be extended up to a minimum of 5 residence times, which is the period is required to reach a true metabolomic steady state [105].

### **Global analysis of intracellular amino acid pools**

To obtain a global overview of the effects of different oxygen supply conditions and recombinant protein secretion on the measured intracellular free amino acid pool sizes, we subjected the relative changes in metabolite pool sizes (see Materials and Methods) to Principal Component Analysis (PCA) (Figure 7.1; see also appendix for full data from PCA analysis). PCA projection demonstrated that the maximum variability in the data set clearly differentiated between different oxygenation conditions (precisely, between normoxia and the two oxygen-restricted conditions), with the first component (PC1) covering 71.3 % of the data variance. The second principal component (PC2), which explained only 19.2 % of the total amino acid pools variance, clearly discriminated between the Fab expressing and the reference strain, indicating a limited impact of the antibody fragment production on the *P. pastoris* amino acid metabolome. Ala, Trp and Asp were the amino acids with the highest contribution (24.0%, 20.6% and 19.7% respectively) to the variance in PC2.

Overall, these data reflect a higher impact of oxygen availability rather than recombinant protein production on the global physiologic response of *P. pastoris*, consistent with previous transcriptomic and proteomic analyses [52]. To help decipher the potential dependence of the amino acid metabolism

on oxygen availability and the burden caused by recombinant protein secretion, a more detailed analysis was performed focusing on each of these two factors separately.

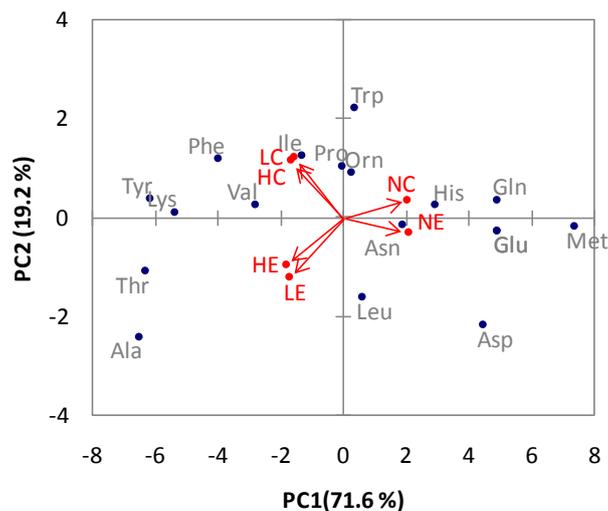


Figure 7.1: **Principal component analysis (PCA) of amino acid data.** Principal component analysis of the intracellular amino acid pools data in a 2D graph of PC1 and PC2. The biplot shows amino acid data (scores) as labeled dots and treatment effect (loadings) as vectors for the expressing (E) and control (C) strain at different concentrations in the inlet gas (Normoxia (N); Oxygen-limited (L) and Hypoxia (H)). Vectors that are close together are highly correlated in terms of the observed amino acid pool sizes for each treatment, while vectors that are orthogonal are poorly correlated. PC1 correlates well with the change in oxygen conditions, whereas PC2 appears to be correlated with the strain type.

### ***Recombinant Protein Production Effect***

The amino acid pools provide building blocks for protein synthesis. Therefore, an impact of Fab production on these metabolites was *a priori* expected. In fact, previous studies using amino acids for media supplementation or complex extracts have proven to have a positive effect on recombinant protein production [133, 138, 142].

In order to analyze more specifically the potential effects of recombinant protein production on the *P. pastoris* amino acid metabolism, a comparison was made of the free amino acid pools of the reference and the Fab producing strains grown at each oxygenation condition. The metabolites levels measured in the control strain grown at each oxygenation condition were taken as reference values to be compared with the metabolites of the Fab-expressing strain grown at the corresponding condition, thereby obtaining concentration ratios for each metabolite of the Fab-expressing relative to the control strain at each culture condition (Figure 7.2). In addition, statistical analysis allowed to identify those metabolite ratios that were significantly higher or lower than 1 (that is, those metabolite pool sizes that varied significantly between both strains).

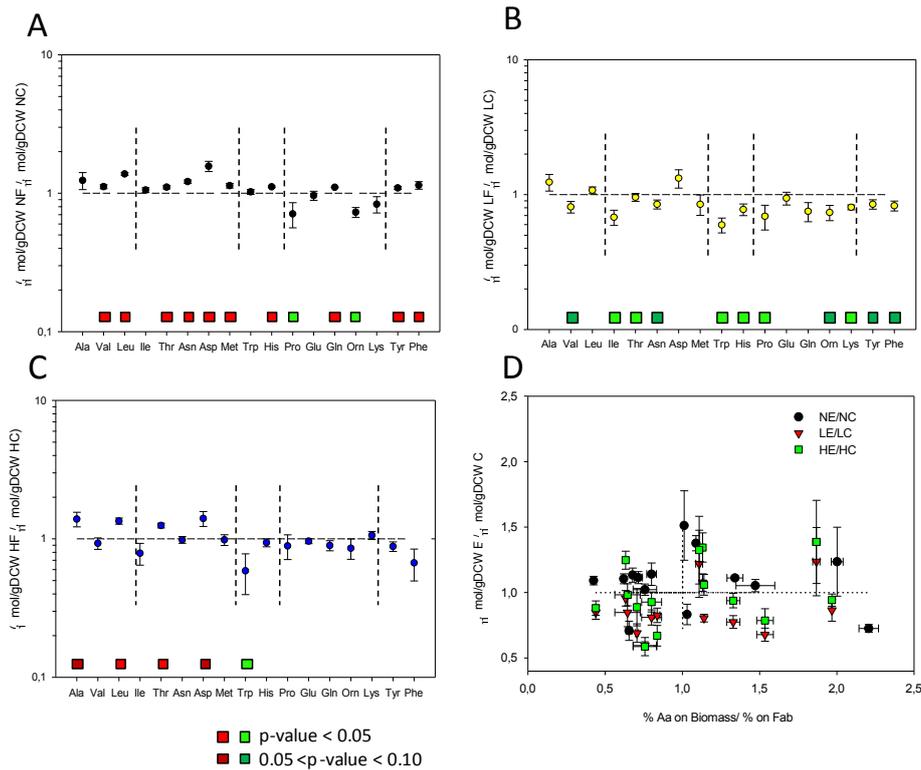


Figure 7.2: **Amino acid pools comparison of *P. pastoris* control (C) and Fab expressing (E) strains under different oxygen conditions.** The values are the average and the standard error of calculated ratios. Horizontal dashed line represents a ratio of 1. In A, B and C vertical dashed line separate metabolites belonging to the same metabolic pathway. The squares over the metabolite names denote the results from the T-Student test. Green and red colors indicate lower and higher ratios respectively. **A:** normoxic condition; **B:** O<sub>2</sub>-limited condition; **B:** hypoxic conditions. **D:** Biomass/Fab amino acid proportion ratios influence on the intracellular amino acid pool changes. N: Normoxic, L: O<sub>2</sub>-limited; H: Hypoxic

The impact of recombinant protein production on amino acid pool sizes under normoxic conditions is shown in Figure 7.2A. Overall, 10 out of 17 amino acid pools increased their intracellular levels significantly, while only 2 of them had an opposite behaviour. In particular, most of the free amino acids from the Ala, Asp and Phe families presented significantly increased levels in the Fab-expressing strain. Nevertheless, the Glu family showed different trends depending on the amino acid, being Orn and Pro significantly decreased, while Gln increased. Increased amino acid pools in the Fab-expressing strain could be an indication of an increased carbon flux through the corresponding biosynthetic pathways. Alternatively, an increased pool size of a given amino acid could indicate that such increase is required to sustain the same or similar flux in the Fab-producing strain. This would occur, for example, if in some of the amino acid biosynthetic pathways their corresponding enzyme levels decreased or they were partially inhibited, allowing for the redistribution of the carbon fluxes in these pathways, thereby compensating differences between the Fab product and the average proteome amino acid

compositions. However, such impact on metabolic fluxes was not observed under these conditions, probably due to the fact that Fab product yields accounted only for a very small fraction of the cell's total protein, considering that both strains had an equivalent biomass composition under a given oxygenation condition [76].

Conversely, generally decreased amino acid pool ratios were found in the Fab-expressing strains under oxygen-limiting conditions, compared to the reference strain (Figure 7.2B 11/17 were reduced, respectively), that is, an opposite behaviour to that observed under normoxic conditions. We observed the same trend when comparing the Fab-expressing strain growing under hypoxic condition; however, the data variance was higher and, therefore, the observed changes were not statistically significant. As discussed above, this observation might be associated to the increased specific heterologous protein productivity found under oxygen limitation, as the Fab producing strain might require higher metabolic fluxes of amino acids for protein synthesis. Nevertheless, this hypothesis is highly unlikely, since Fab production levels were relatively low in relation to the total cell protein [76] and, therefore, cannot explain *per se* the apparent drain of precursor pools. Besides, this trend was not uniform within each of the amino acid families. For instance, in the Ala family, Ala and Leu pools increased while Val decreased in the Fab-expressing strain. Also, in the Asp amino acid family only Asp levels varied significantly. Moreover, no correlation was found between the relative abundance of each amino acid in the Fab antibody fragment and the observed changes in the corresponding free amino acid pools (Figure 7.2D). Nevertheless the specific Fab production correlated inversely (correlation value of  $-0.93$ ) with the average of the metabolite ratios between Fab-expressing and reference strains at each culture condition. This may reflect a global re-adjustment of the free amino acid pools to compensate for the recombinant protein overproduction. Although such readjustment could not be the result from a direct drain of building blocks at higher Fab synthesis levels, other phenomena related with recombinant protein production might provide some explanations. For instance, recent studies on recombinant protein secretion using  $^{34}\text{S}$  labeling strategies with the same strain used in this study have revealed that about 58% of the Fab protein produced intracellularly is actually degraded within the cell, and only 35 % is secreted [143]. In addition to protein degradation, increased energetic demands related with the cost of the folding, refolding and secretion processes of the Fab product could result in an overall readjustment of amino acid metabolism. In fact, when the Fab producing strain was cultivated at lower temperatures (20°C), the unfolded protein stress response was reduced, leading to a reduced metabolic burden and higher specific productivities compared with cells grown at 25°C and 30°C, in which an increase in energy demand was evidenced by an up-regulation of the TCA cycle, slightly higher in Fab-secreting strains [141]. Besides, increased maintenance requirements associated with heterologous protein production (e.g. unfolded protein stress response) may cause additional energy demands. Interestingly, recent amino acid supplementation studies using a recombinant *P. pastoris* strain revealed that such supplementation partially relieved the metabolic burden from recombinant

protein production. Furthermore, de novo amino acid synthesis in cells growing on different combinations of amino acids supplementations was inversely correlated with the corresponding energetic cost for most amino acids [134].

The potential dependence between fold changes observed amongst amino acid pools when comparing the Fab-expressing to the reference strain and their corresponding energy and redox costs were analyzed (Figure 7.3). Interestingly, lower fold changes in amino acid pools with the highest energy cost (that is, aromatic amino acids family) was found under oxygen reduced conditions (that is, under higher Fab productivities), but not under normoxia. This might point at the hypothesis that the cell adjusts its overall amino acid metabolism to minimize the energetic burden caused by Fab production. Nevertheless, no trend was observed regarding the C-mol degree of reduction of each amino acid (Figure 7.3).

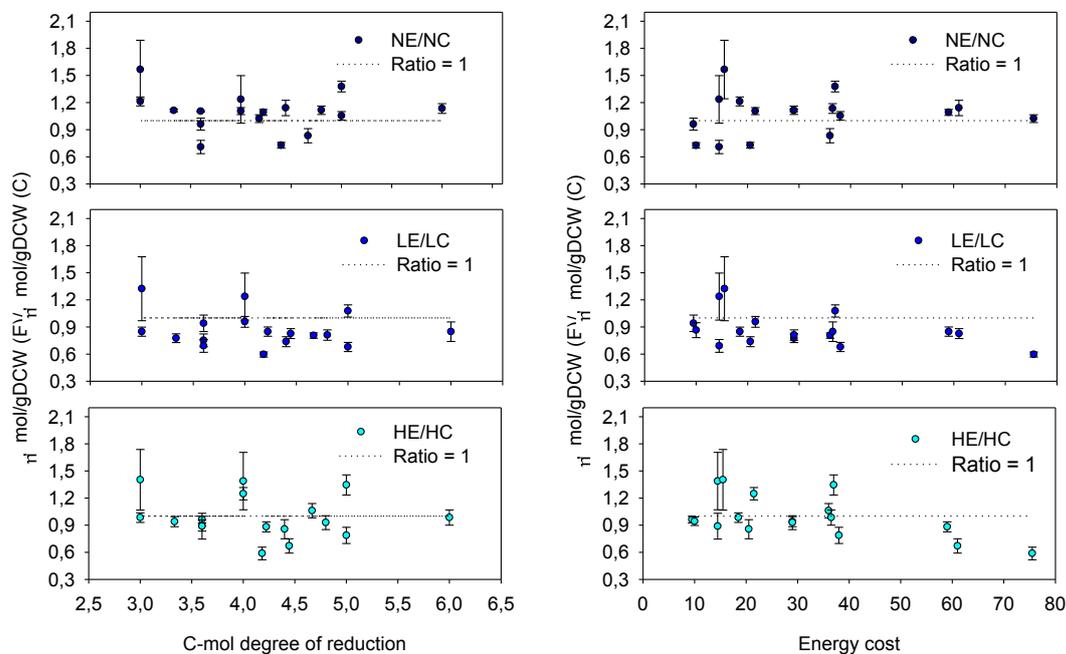


Figure 7.3 **Amino acid C-mol degree of reduction and energy costs influence on the expressing/control amino acid pool ratios.** The values are the average and the standard error of calculated ratios. Horizontal dashed line represents a ratio of 1. The energy cost for an amino acid is defined as the number of high energy phosphate bounds ( $\sim\text{PO}_4$ ) that are required for its synthesis. The amino acids energy costs were taken from *S. cerevisiae* published data [140]. The C-mol degree of reduction of each amino acid was calculated from each molecular formula.

### **Oxygen availability effect**

The oxygen availability has a strong impact on central metabolism due to his crucial role as electron acceptor, as already inferred from the growth parameters (Table 7.1). Also, previous transcriptomic studies under analogous cultivation conditions revealed amino acid metabolism as one of the major

cellular processes regulated by oxygen availability [52]. Interestingly, the number of genes in the GO group of amino acid metabolism that were downregulated under hypoxic conditions was substantially higher in the Fab-producing strain.

To investigate the potential correlation between observed changes in free amino acid pools and proteome amino acid composition upon oxygen availability changes, we related the observed fold changes for each free amino acid pool with the change in the relative abundance of the corresponding amino acid in the cell's proteome amino acid composition (Figure 7.4). The effect of oxygen availability on the cell's amino acid metabolism could be already inferred from previous measurements of the cell's protein amino acid composition [76]. For instance, the relative abundance of amino acids derived from oxalocetate, pyruvate, phosphoenolpyruvate and 3PG were increased and those derived from aKG were reduced under hypoxic conditions.

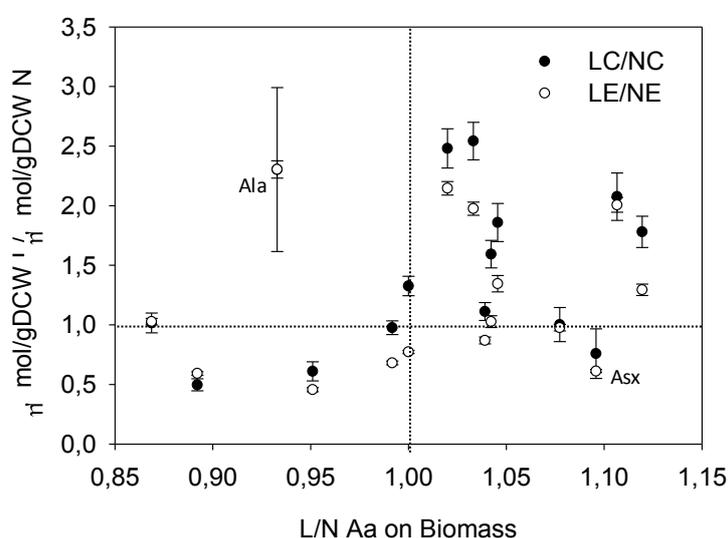


Figure 7.4: **Impact of the differential biomass demand over the intracellular amino acid pools under different oxygen conditions.** The values are the average and the standard error of calculated ratios. Horizontal and vertical dashed lines divided the graph in four spaces to facilitate the interpretation. The biomass demands under normoxia and oxygen-limited conditions were taken from [76]. Asx: Asp+Asn

Notably, free amino acid pools seemed to correlate with the relative abundance in the cell proteome for most amino acids, with the clear exception of alanine and, to a much lesser extent, Asx. For these cases, a possible explanation would be stronger regulation from other part of the metabolism as for example their intracellular metabolite precursor levels. However, further studies need to be performed to validate it.

To obtain an overview of the potential correlations between transcriptional changes in amino acid biosynthetic genes and intracellular amino acid levels (free pools as well as proteome amino acids) upon a change in oxygen availability, transcriptomic, metabolomic and cell protein compositional data

were integrated into biosynthetic pathway maps (Figures 7.5 and 7.6). The specific amino acid composition of the whole protein extracts at different oxygen conditions was included in order to show the biosynthetic demand for each specific biosynthetic pathway. Since hypoxic condition of this study was not exactly equivalent to that used in previous transcriptomic studies, only normoxic and oxygen limiting conditions were compared for both E and C strains.

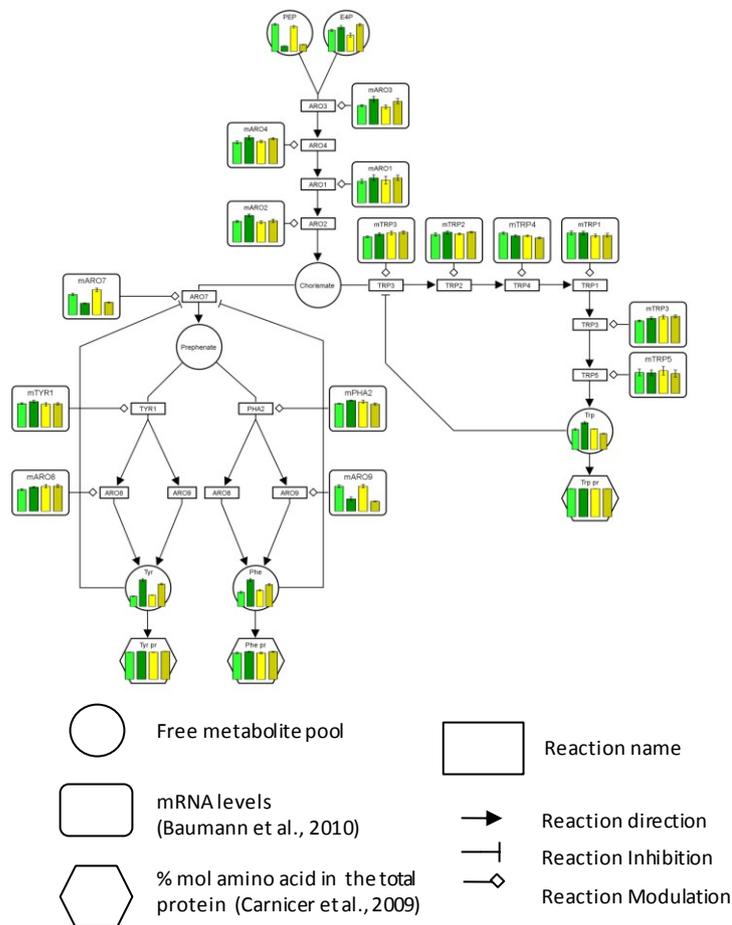


Figure 7.5: **Aromatic amino acid biosynthetic pathway comparison under different oxygen conditions.** The metabolite level bars in the graphs are the average and the standard error from at least 4 measurements. The control and the Fab-expressing strain are represented by green and yellow bars respectively. Oxygenation conditions from normoxic to oxygen-limited conditions are represented as light to dark color scale. The mRNA data are indicated as m plus the specific reaction name. Each graphic has its own scale

A direct comparison of the relative changes observed in the free amino acid levels between different oxygen availability conditions and the corresponding changes observed in the whole protein extract (that is, free intracellular amino acid pools plus cell protein amino acids) indicated that the latter were less pronounced.

The integrated data seem to reflect the oxygen-dependent transcriptional regulation of amino acid biosynthesis pathways. In particular, transcriptional levels of several key regulatory enzymes in the

biosynthetic pathways correlated inversely with the carbon flow through them, as well as with their corresponding end-metabolite levels, probably denoting the negative feedback control of the pathway. For instance, the levels of free tyrosine and phenylalanine were substantially increased under oxygen-limiting conditions, concomitantly with a reduction of the mRNA levels of *ARO7*, while an opposite pattern was observed for the free tryptophan pools and *TRP3* transcriptional levels (Figure 7.5). This pattern was also observed in other amino acid biosynthetic pathways such as methionine.

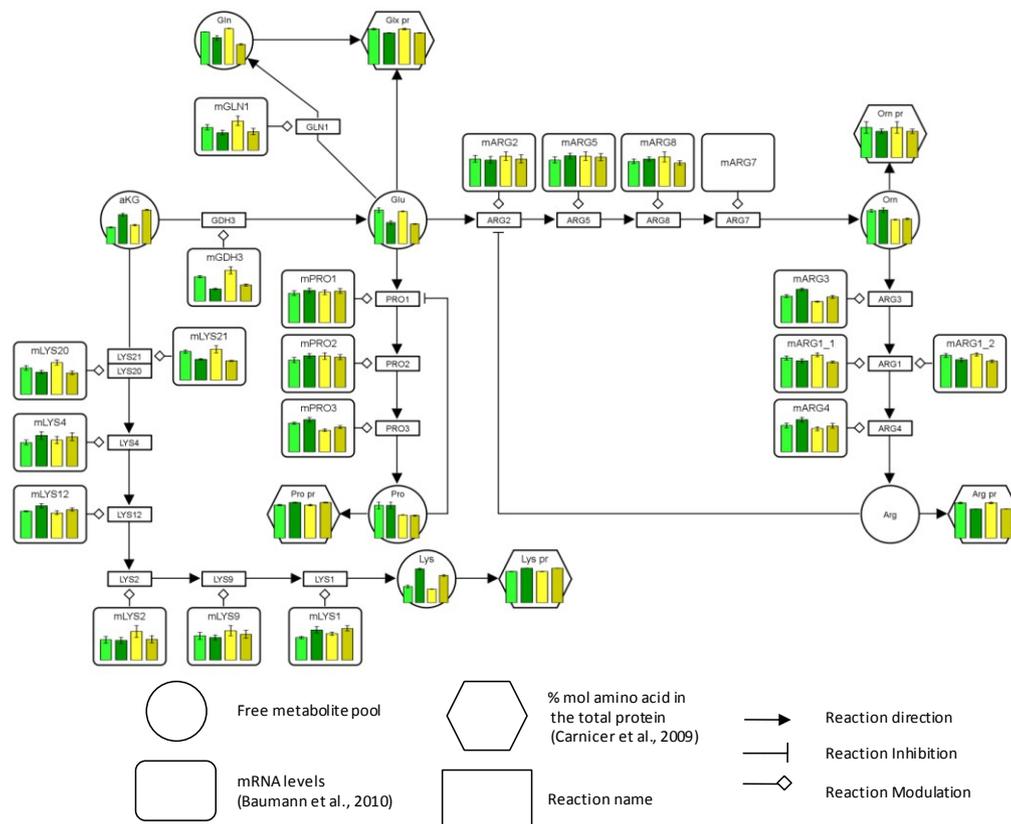


Figure 7.6: **Glutamate biosynthetic pathway comparison under different oxygen conditions.** The metabolite levels bars in the graphs are the average and the standard error from at least 4 measurements. The control and the Fab-expressing strain are represented by black-grey and dark-light yellow bars, respectively. Oxygenation conditions from normoxic to hypoxic are represented in a dark-to-light color scale. The mRNA data are indicated as m plus the specific reaction name. Each graphic has its own scale

Similarly, in the glutamate amino acid family (Figure 7.6), the carbon flow to lysine was increased under oxygen limiting conditions, concomitantly with a reduction in transcript levels of several genes of its pathway, particularly *LYS20*, the first reaction of the pathway. In addition, the carbon flux to the Glu and Gln biosynthetic pathway was significantly reduced at lower oxygen availability. This pattern correlated directly with the reduced transcriptional levels of *GDH3* and *GLN1*.

## Conclusions

Previous transcriptomic studies [52] pointed at the impact of oxygen availability and recombinant protein production on amino acid metabolism. The analysis at the metabolomic level further confirms such impact. In particular, our data point at a major impact of oxygen availability rather than recombinant protein production on the free amino acid pools, coherent with previous transcriptional analyses. Notably, changes in free amino acid pools observed at different oxygenation conditions generally correlated directly with the changes in relative abundances in the corresponding amino acids of the cell's proteome, with alanine being the major exception. In contrast, the impact of recombinant protein production on the free amino acids pools depended on the oxygenation state. Importantly, the observed changes did not correlate with the difference in amino acid composition of the recombinant product and the cell's proteome, but rather to the energetic costs (specifically, for those amino acids with highest energy costs), thereby suggesting a possible dependence between mitochondrial metabolism and amino acid anabolism as a potential target to modulate the metabolic burden caused by recombinant protein production.

In addition, simple combination of metabolomic and transcriptional data suggests that their systematic integration into genomic-scale metabolic models should allow gaining further knowledge on the regulation of central and amino acid metabolism, as well as identifying metabolic bottlenecks limiting enhanced recombinant protein production.

## **Chapter 8: Outlook and future directions**



The aim of this research project was to contribute to the understandings of the *Pichia pastoris* physiology under different stress conditions. Specifically, the oxygen availability and recombinant protein production were the two parameters studied throughout this thesis.

Since the beginning, the available studies using metabolomics and fluxomics approaches have experienced a significant increase highlighting the potential of both techniques and the scientific community efforts to put them available. In this context, my contribution has been the development of a *P. pastoris* intracellular metabolite quantification method opening the metabolomic world for this microorganism.

Nevertheless, despite the methodological advances, many uncertainties still remain and much more are forthcoming. Pentose phosphate pathway could be an example. This pathway has become very important pathway in the thesis for many reasons and quite significant improvements have been done in describing its behaviour. However, further investigations need to be addressed on this topic.

In this chapter, rather than repeating the conclusions from previous ones, regarding each of the challenges which were dealt with, I chose to use it to address different questions which came to my mind along the thesis. Some of them I asked myself when preparing the result and discussion sections of the different chapters. Others are simply possibilities which could be interesting to try.

### **Different recombinant protein production strategy**

Since the beginning of the thesis, the used recombinant *Pichia pastoris* cell produces the antibody fragment 3H6 as a complex recombinant protein model. Using this system, and comparing it to non-recombinant cells, many results have been produced gaining knowledge about the impact of this production all over *P. pastoris* physiology [52](Chapters 6 and 7). Moreover, other research groups were also working, with the same cells or with other cell factories, in order to characterize the 3H6 antibody fragment production impact [46, 141]. Nevertheless, what would happen with other recombinant protein model? Or, changing the expression system?

Our expressing *P. pastoris* strain usually produce between 40-100  $\mu\text{g Fag}/(\text{gDCW}\cdot\text{h})$  depending on the environmental conditions. Since the first results, this low specific productivity led to quite low or no one impact independently of omic level investigated. Nevertheless, what would the results look like, for example, using a simpler model allowing, probably, higher productivities?

This question is partly addressed by Joel Jordà Murria, a colleague from my group, who is investigating the metabolic flux profiling of *P. pastoris* related to a Lipase recombinant protein production. However, the culture and the induction strategy were not the same which make the comparison not straight forward.

Another possibility would have been to change the expression system. In our expression cassette, the light and heavy chains were under control of the glyceraldehyde 3-phosphate promoter which was transcriptionally upregulated under oxygen limitation. Thereby, our recombinant protein expression was influenced by the oxygen availability in the bioreactor which agreed with the higher specific productivities under shortage of oxygen. However, now what needs to be investigated is if the oxygen, independently of the transcriptomical impact, improves the recombinant protein production.

## **Metabolite intracellular transport and compartmentalization**

Eukaryotic cells are more complex due to its internal architecture. This higher cell complexity has been a challenge which I had to cope with along the thesis in different chapters.

In the metabolic flux analysis, the well characterisation of the stoichiometry reactions taken place between the compartments it is a matter of outstanding importance when you work with eukaryotic cells. Different reactions could lead completely different results being both agree with obtained experimental data. In *P. pastoris*, part of the TCA cycle is described to work not only in the mitochondria but also in the cytosol [60]. These results opened many transport possibilities between the TCA metabolites intermediates which needed to be characterized. Moreover, a redox shuttle could occur between the cytosol and mitochondria using these TCA metabolites [82]. Using the  $^{13}\text{C}$ -NMR ratios these problem could be partly bypassed assuming that the main metabolite which is transported through the mitochondrial membrane was oxalacetate although, it is not what was happening in reality.

On the other side, the eukaryotic cell architecture became also important when developing the intracellular metabolite quantification method for *P. pastoris*. Since today, there is not any method or protocol which could quantify metabolites differentiating between compartments. Despite, different tricks to bypass this problem are already described. Different possibilities would be to use reaction indicators as performed for the redox cytosol state in *S. cerevisiae* [116] or thermodynamic restrictions similarly to Chapter 5 [103]. Nevertheless, the optimum would be to applied some cell fractionation together with the quenching protocol which it is unclear when or whether it will be possible to effectively combine [144].

On the other hand, in order to gain knowledge in the cell compartmentalization an accurate determination of the volume fraction of each compartment would be interesting as described for *S. cerevisiae* [145].

## ***Pichia pastoris* genome-scale metabolic models and $^{13}\text{C}$ -kinetic models**

Nowadays, different kind of modelling approaches exist depending on the methodology/approach used. There are the bottom-up models represented, for example, by the genome-scale metabolic models and the top-down models being exemplified by the  $^{13}\text{C}$ -kinetic modelling

Genome-scale metabolic models have been described as one of the most powerful tools for metabolic engineering [146]. Recently, two genome-scale models have been performed for *P. pastoris* [79, 80]. So, both models can be used for performing rational genetic changes improving the change of success. Today, there are software which could be used to manage these models (ex: COBRA Matlab toolbox) although they are not straight forward to use and understand. However, the forthcoming projects have to be adapted to the new available tools.

On the other hand, oppositely to the genome-scale models, the  $^{13}\text{C}$ -kinetic models are addressed to determine the metabolic regulation of the fluxes. These models are based in fitting kinetic equations to labelled metabolite trends obtained experimentally. These models are nowadays centred mainly to the central carbon metabolism due to the impossibility to accurately analyse all the intracellular metabolites as well as the increasing number of parameters in bigger models. Developing these kind of models would help to improve future metabolic engineered strains when the genetics does not work.

## Hierarchical regulation

The flux through a pathway or an enzyme can be changed by many different regulatory mechanisms. These mechanisms can be dissected as performed in chapter 6 using the regulation analysis theory (Eq. 8.1).

$$\frac{\Delta \log J}{\Delta \log J} = \frac{\Delta \log V_{max}}{\Delta \log J} + \frac{\Delta \log g(X,K)}{\Delta \log J} = 1 \quad (\text{Eq. 8.1})$$

This theory mainly divides the different regulatory mechanisms in two groups depending on what they regulate. One group includes the changes aimed to modify the maximum catalytic activity of a reaction representing the flux hierarchical regulation. Oppositely, the other group embraces the metabolic flux regulation.

In chapter 6, the transcriptional regulation coefficients of the central metabolic pathways were determined being for many enzymes significant. Nevertheless, no data was available regarding the maximum catalytic activity changes under the different tested conditions. Recently, it has been developed an *in vivo*-like methodologies to determine the enzyme activities for *S. cerevisiae* [147] and for *Lactococcus lactis* [148] opening the possibility to determine the total hierarchical regulation.

This analysis not only could be informative to determine which kind of regulation is dominant for each enzyme but also very useful in designing future metabolic engineering strategies due to you will know what to be modified.

## Bibliography

1. Lee SY, Lee D-Y, Kim TY: **Systems biotechnology for strain improvement.** *Trends in Biotechnology* 2005, **23**:349-58.
2. Porro D, Gasser B, Fossati T, et al.: **Production of recombinant proteins and metabolites in yeasts: when are these systems better than bacterial production systems?** *Applied Microbiology and Biotechnology* 2011, **89**:939-48.
3. Pscheidt B, Glieder A: **Yeast cell factories for fine chemical and API production.** *Microbial Cell Factories* 2008, **7**:25.
4. Cereghino JL, Cregg JM: **Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*.** *FEMS Microbiology Reviews* 2000, **24**:45-66.
5. Sauer M, Porro D, Mattanovich D, Branduardi P: **Microbial production of organic acids: expanding the markets.** *Trends in Biotechnology* 2008, **26**:100-8.
6. Ekins A, Martin VJJ: **Cellular Metabolism.** In *The Metabolic Pathway Engineering Handbook*. 2010.
7. Lim S-J, Jung Y-M, Shin H-D, Lee Y-H: **Amplification of the NADPH-related genes *zwf* and *gnd* for the oddball biosynthesis of PHB in an *E. coli* transformant harboring a cloned *phbCAB* operon.** *Journal of Bioscience and Bioengineering* 2002, **93**:543-9.
8. Petranovic D, Nielsen J: **Can yeast systems biology contribute to the understanding of human disease?** *Trends in Biotechnology* 2008, **26**:584-90.
9. Gulik WM van: **Fast sampling for quantitative microbial metabolomics.** *Current Opinion in Biotechnology* 2010, **21**:27-34.
10. Bolten CJ, Kiefer P, Letisse F, Portais J-C, Wittmann C: **Sampling for metabolome analysis of microorganisms.** *Analytical Chemistry* 2007, **79**:3843-9.
11. Oldiges M, Lütz S, Pflug S, et al.: **Metabolomics: current state and evolving methodologies and tools.** *Applied Microbiology and Biotechnology* 2007, **76**:495-511.
12. Douma RD, Verheijen PJT, Laat WT de, Heijnen JJ, Gulik WM van: **Dynamic gene expression regulation model for growth and penicillin production in *Penicillium chrysogenum*.** *Biotechnology and Bioengineering* 2010, **106**:608-18.
13. Sauer U: **Metabolic networks in motion: <sup>13</sup>C-based flux analysis.** *Molecular Systems Biology* 2006, **2**:62.

14. Eunen K van, Rossell S, Bouwman J, Westerhoff H van, Bakker BM: **Quantitative Analysis of Flux Regulation Through Hierarchical Regulation Analysis**. In *Methods in Enzymology*. 1st edition. Elsevier Inc.; 2011, **500**:571-595.
15. Kahn D, Westerhoff HV: **The regulatory strength: how to be precise about regulation and homeostasis**. *Acta Biotheoretica* 1993, **41**:85-96.
16. Rossell S, Weijden CC van der, Lindenbergh A, et al.: **Unraveling the complexity of flux regulation: a new method demonstrated for nutrient starvation in *Saccharomyces cerevisiae***. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**:2166-71.
17. Daran-Lapujade P, Rossell S, Gulik WM van, et al.: **The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels**. *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**:15753-8.
18. Kuile BH ter, Westerhoff HV: **Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway**. *FEBS letters* 2001, **500**:169-71.
19. Heijnen JJ: **Approximative kinetic formats used in metabolic network modeling**. *Biotechnology and Bioengineering* 2005, **91**:534-45.
20. Rossell S, Weijden CC van der, Kruckeberg AL, Bakker BM, Westerhoff HV: **Hierarchical and metabolic regulation of glucose influx in starved *Saccharomyces cerevisiae***. *FEMS Yeast Research* 2005, **5**:611-9.
21. Wiechert W, Möllney M, Petersen S, Graaf a a de: **A universal framework for <sup>13</sup>C metabolic flux analysis**. *Metabolic Engineering* 2001, **3**:265-83.
22. Marx A, Eikmanns BJ, Sahm H, Graaf AAD, Eggeling L: **Response of the Central Metabolism in *Corynebacterium glutamicum* to the use of an NADH-Dependent Glutamate Dehydrogenase**. *Metabolic Engineering* 1999, **1**:35-48.
23. Smolders GJ, Meij J van der, Loosdrecht MC van, Heijnen JJ: **Stoichiometric model of the aerobic metabolism of the biological phosphorus removal process**. *Biotechnology and Bioengineering* 1994, **44**:837-48.
24. Schmidt K, Marx A, Graaf AAD, et al.: **Tracer Experiments and Metabolite Balancing for Metabolic Flux Analysis: Comparing Two Approaches**. *Biotechnology and Bioengineering* 1998, **58**:254-257.
25. Winden W a van, Verheijen PJT, Heijnen JJ: **Possible pitfalls of flux calculations based on <sup>13</sup>C-labeling**. *Metabolic Engineering* 2001, **3**:151-62.

26. Maaheimo H, Fiaux J, Cakar ZP, et al.: **Central carbon metabolism of *Saccharomyces cerevisiae* explored by biosynthetic fractional  $^{13}\text{C}$  labeling of common amino acids.** *European Journal of Biochemistry* 2001, **268**:2464-79.
27. Jouhten P, Rintala E, Huuskonen A, et al.: **Oxygen dependence of metabolic fluxes and energy generation of *Saccharomyces cerevisiae* CEN.PK113-1A.** *BMC Systems Biology* 2008, **2**:60.
28. Szyperski T, Glaser RW, Hochuli M, et al.: **Bioreaction network topology and metabolic flux ratio analysis by biosynthetic fractional  $^{13}\text{C}$  labeling and two-dimensional NMR spectroscopy.** *Metabolic Engineering* 1999, **1**:189-97.
29. Mashego MR, Wu L, Dam JC van, et al.: **MIRACLE: mass isotopomer ratio analysis of  $\text{U-}^{13}\text{C}$ -labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites.** *Biotechnol Bioeng* 2004, **85**:620-8.
30. Dam JC van, Eman M, Frank J, et al.: **Analysis of glycolytic intermediates in *Saccharomyces cerevisiae* using anion exchange chromatography and electrospray ionization with tandem mass spectrometric detection.** *Analytica Chimica Acta* 2002, **460**:209-218.
31. Dettmer K, Aronov PA, Hammock BD: **Mass spectrometry-based metabolomics.** *Mass Spectrometry Reviews* 2007, **26**:51- 78.
32. Lange HC, Eman M, Zuijlen G van, et al.: **Improved rapid sampling for in vivo kinetics of intracellular metabolites in *Saccharomyces cerevisiae*.** *Biotechnology and Bioengineering* 2001, **75**:406-15.
33. Canelas AB, Ras C, Pierick A ten, et al.: **Leakage-free rapid quenching technique for yeast metabolomics.** *Metabolomics* 2008, **4**:226-239.
34. Canelas AB, Pierick A ten, Ras C, et al.: **Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics.** *Anal Chem* 2009, **81**:7379-89.
35. Villas-Bôas SG, Højer-Pedersen J, Akesson M, Smedsgaard J, Nielsen J: **Global metabolite analysis of yeast: evaluation of sample preparation methods.** *Yeast* 2005, **22**:1155-69.
36. Bolten CJ, Wittmann C: **Appropriate sampling for intracellular amino acid analysis in five phylogenetically different yeasts.** *Biotechnology Letters* 2008, **30**:1993-2000.
37. Taymaz-Nikerel H, Mey M de, Ras C, et al.: **Development and application of a differential method for reliable metabolome analysis in *Escherichia coli*.** *Analytical Biochemistry* 2009, **386**:9-19.
38. Verheijen PJT: **Data Reconciliation and Error Detection.** In *The Metabolic Pathway Engineering Handbook*. edited by Smolke CD2010:8.1-8.13.

39. Gach JS, Maurer M, Hahn R, et al.: **High level expression of a promising anti-idiotypic antibody fragment vaccine against HIV-1 in *Pichia pastoris***. *J Biotechnol* 2007, **128**:735-46.
40. Gasser B, Sauer M, Maurer M, Stadlmayr G, Mattanovich D: **Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts**. *Applied and Environmental Microbiology* 2007, **73**:6499-507.
41. Daly R, Hearn MTW: **Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production**. *Journal of Molecular Recognition* 2005, **18**:119-38.
42. Strausberg RL, Strausberg SL: **Overview of protein expression in *Saccharomyces cerevisiae***. In *Current protocols in protein science*. 2001, **Chapter 5**:Unit5.6.
43. Cregg JM, Tolstorukov I, Kusari A, et al.: **Expression in the yeast *Pichia pastoris***. In *Methods Enzymol*. 1st edition. Elsevier Inc.; 2007, **463**:169-89.
44. Haynes CM, Titus E a, Cooper A: **Degradation of misfolded proteins prevents ER-derived oxidative stress and cell death**. *Molecular Cell* 2004, **15**:767-76.
45. Malhotra JD, Miao H, Zhang K, et al.: **Antioxidants reduce endoplasmic reticulum stress and improve protein secretion**. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:18525-30.
46. Gasser B, Saloheimo M, Rinas U, et al.: **Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview**. *Microb Cell Fact* 2008, **7**:11.
47. Baumann K, Maurer M, Dragosits M, et al.: **Hypoxic fed-batch cultivation of *Pichia pastoris* increases specific and volumetric productivity of recombinant proteins**. *Biotechnol Bioeng* 2008, **100**:177-83.
48. Förster J, Famili I, Fu P, Palsson BØ, Nielsen J: **Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network**. *Genome Research* 2003, **13**:244-53.
49. Feist AM, Palsson BØ: **The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli***. *Nature Biotechnology* 2008, **26**:659-67.
50. Nookaew I, Jewett MC, Meechai A, et al.: **The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism**. *BMC Systems Biology* 2008, **2**:71.
51. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ: **Reconstruction of biochemical networks in microorganisms**. *Nature Reviews. Microbiology* 2009, **7**:129-43.
52. Baumann K, Carnicer M, Dragosits M, et al.: **A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions**. *BMC Syst Biol* 2010, **4**:141.

53. Semke J, Mahdavi J, Mathis M: **Automatic recording apparatus for use in chromatography of amino acids.** *Fed Proc* 1958, **17**:1107-15.
54. Herbert D, Phipps PJ, Strange RE: **Chemical analysis of microbial cells.** *Methods in Microbiology* 1971, **5b**:209-344.
55. Lange HC, Heijnen JJ: **Statistical reconciliation of the elemental and molecular biomass composition of *Saccharomyces cerevisiae*.** *Biotechnology and Bioengineering* 2001, **75**:334-44.
56. Roustan J-louis, Sablayrolles J-marie: **Trehalose and glycogen in wine-making yeasts : methodological aspects.** *Biotechnology Letters* 2002, **24**:1059-1064.
57. Hara a, Radin NS: **Lipid extraction of tissues with a low-toxicity solvent.** *Analytical Biochemistry* 1978, **90**:420-6.
58. Mey M De, Lequeux G, Maertens J, et al.: **Comparison of DNA and RNA quantification methods suitable for parameter estimation in metabolic modeling of microorganisms.** *Analytical Biochemistry* 2006, **353**:198-203.
59. Benthin S, Nielsen J, Villadsen J: **A simple and reliable method for the determination of cellular RNA content.** *Biotechnology Techniques* 1991, **5**:39-42.
60. Solà A, Maaheimo H, Ylönen K, Ferrer P, Szyperski T: **Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*.** *Eur J Biochem* 2004, **271**:2462-70.
61. Gasser B, Maurer M, Gach JS, Kunert R, Mattanovich D: **Engineering of *Pichia pastoris* for Improved Production of Antibody Fragments.** *Biotechnology and Bioengineering* 2006, **94**:353-361.
62. Stephanopoulos GN, Aristidou AA, Nielsen J: *Metabolic Engineering: Principles and Methodologies.* Academic Press; 1998.
63. Schutter K de, Lin Y-C, Tiels P, et al.: **Genome sequence of the recombinant protein production host *Pichia pastoris*.** *Nat Biotechnol* 2009, **27**:561-6.
64. Mattanovich D, Graf AB, Stadlmann J, et al.: **Genome, secretome and glucose transport highlight unique features of the protein production host *Pichia pastoris*.** *Microb Cell Fact* 2009, **8**:29.
65. Unnikrishnan I, Miller S, Meinke M, LaPorte DC: **Multiple positive and negative elements involved in the regulation of expression of *GSY1* in *Saccharomyces cerevisiae*.** *The Journal of Biological Chemistry* 2003, **278**:26450-7.
66. Goffeau A, Barrell BG, Bussey H, et al.: **Life with 6000 Genes.** *Science* 1996, **274**:546-567.

67. Noorman HJ, Romein B, Luyben KC, Heijnen JJ: **Classification, error detection, and reconciliation of process information in complex biochemical systems.** *Biotechnology and Bioengineering* 1996, **49**:364-76.
68. Heijden RT van der, Heijnen JJ, Hellinga C, Romein B, Luyben KC: **Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates.** *Biotechnology and Bioengineering* 1994, **43**:3-10.
69. Heijden RT van der, Romein B, Heijnen JJ, Hellinga C, Luyben KC: **Linear constraint relations in biochemical reaction systems: II. Diagnosis and estimation of gross errors.** *Biotechnology and Bioengineering* 1994, **43**:11-20.
70. Tokuoka K: **Sugar- and salt-tolerant yeasts.** *Journal of Applied Microbiology* 1993, **74**:101-110.
71. Passoth V, Fredlund E, Druvefors UA, Schnürer J: **Biotechnology, physiology and genetics of the yeast *Pichia anomala*.** *FEMS Yeast Research* 2006, **6**:3-13.
72. Fiaux J, Çakar ZP, Sonderegger M, et al.: **Metabolic-Flux Profiling of the Yeasts *Saccharomyces cerevisiae* and *Pichia stipitis*.** *Eukaryotic Cell* 2003, **2**:170-180.
73. Sauer U, Hatzimanikatis V, Bailey JE, et al.: **Metabolic fluxes in riboflavin-producing *Bacillus subtilis*.** *Nature Biotechnology* 1997, **15**:448-452.
74. Szyperski T: **Biosynthetically directed fractional <sup>13</sup>C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism.** *European Journal of Biochemistry* 1995, **232**:433-48.
75. Sauer U, Lasko DR, Fiaux J, et al.: **Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism.** *Journal of Bacteriology* 1999, **181**:6679-88.
76. Carnicer M, Baumann K, Töplitz I, et al.: **Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels.** *Microb Cell Fact* 2009, **8**:65-78.
77. Stephanopoulos GN, Aristidou AA, Nielsen J: **Review of Cellular metabolism.** In *Metabolic Engineering: Principles and Methodologies*. 1997, **16**.
78. Solà A, Jouhten P, Maaheimo H, et al.: **Metabolic flux profiling of *Pichia pastoris* growth on glycerol / methanol mixtures in chemostat cultures at low and high dilution rates.** *Microbiology* 2007, **153**:281-90.
79. Sohn SB, Graf AB, Kim TY, et al.: **Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production.** *Biotechnol J* 2010, **5**:705-715.
80. Chung BKS, Selvarasu S, Camattari A, et al.: **Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement.** *Microb Cell Fact* 2010, **9**:50-64.

81. Fischer E: **High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived  $^{13}\text{C}$  constraints.** *Analytical Biochemistry* 2004, **325**:308-316.
82. Bakker BM, Overkamp KM, Maris AJA van, et al.: **Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*.** *FEMS Microbiol Rev* 2001, **25**:15-37.
83. Dijken JP van, Weusthuis R a, Pronk JT: **Kinetics of growth and sugar consumption in yeasts.** *Antonie van Leeuwenhoek* 1993, **63**:343-52.
84. Roermund CWT van, Hetteema EH, Berg M van den, Tabak HF, Wanders RJ a: **Molecular characterization of carnitine-dependent transport of acetyl-CoA from peroxisomes to mitochondria in *Saccharomyces cerevisiae* and identification of a plasma membrane carnitine transporter, Agp2p.** *The EMBO Journal* 1999, **18**:5843-52.
85. Strijbis K, Roermund CWT van, Hardy GP, et al.: **Identification and characterization of a complete carnitine biosynthesis pathway in *Candida albicans*.** *The FASEB Journal* 2009, **23**:2349-59.
86. Frick O, Wittmann C: **Characterization of the metabolic shift between oxidative and fermentative growth in *Saccharomyces cerevisiae* by comparative  $^{13}\text{C}$  flux analysis.** *Microbial Cell Factories* 2005, **4**:30.
87. Fredlund E, Blank LM, Schnürer J, Sauer U, Passoth V: **Oxygen- and Glucose-Dependent Regulation of Central Carbon Metabolism in *Pichia anomala*.** *Applied and Environmental Microbiology* 2004, **70**:5905-5911.
88. Fredlund E, Broberg A, Boysen ME, Kenne L, Schnürer J: **Metabolite profiles of the biocontrol yeast *Pichia anomala* J121 grown under oxygen limitation.** *Applied Microbiology and Biotechnology* 2004, **64**:403-9.
89. Tang YJ, Garcia Martin H, Myers S, et al.: **Advances in analysis of microbial metabolic fluxes via  $^{13}\text{C}$  isotopic labeling.** *Mass Spectrometry Reviews* 2009, **28**:362- 375.
90. Douma RD, Jonge LP de, Jonker CTH, et al.: **Intracellular metabolite determination in the presence of extracellular abundance: Application to the penicillin biosynthesis pathway in *Penicillium chrysogenum*.** *Biotechnology and Bioengineering* 2010, **107**:105-15.
91. Wu L, Mashego MR, Dam JC van, et al.: **Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly  $^{13}\text{C}$ -labeled cell extracts as internal standards.** *Analytical Biochemistry* 2005, **336**:164-71.
92. Wittmann C, Krömer JO, Kiefer P, Binz T, Heinze E: **Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria.** *Analytical Biochemistry* 2004, **327**:135-9.
93. Wellerdiek M, Winterhoff D, Reule W, Brandner J, Oldiges M: **Metabolic quenching of *Corynebacterium glutamicum*: efficiency of methods and impact of cold shock.** *Bioprocess and Biosystems Engineering* 2009, **32**:581-92.

94. Tredwell GD, Edwards-Jones B, Leak DJ, Bundy JG: **The Development of Metabolomic Sampling Procedures for *Pichia pastoris*, and Baseline Metabolome Data.** *PLoS ONE* 2011, **6**:1-16. e16286.
95. Eastmond P, Graham IA: **Trehalose metabolism: a regulatory role for trehalose-6-phosphate?** *Current Opinion in Plant Biology* 2003, **6**:231-235.
96. Canelas AB, Ras C, Pierick A ten, Gulik WM van, Heijnen JJ: **An in vivo data-driven framework for classification and quantification of enzyme kinetics and determination of apparent thermodynamic data.** *Metabolic Engineering* 2011, **13**:294-306.
97. Abbott BJ, Laskin a I, McCoy CJ: **Effect of growth rate and nutrient limitation on the composition and biomass yield of *Acinetobacter calcoaceticus*.** *Applied Microbiology* 1974, **28**:58-63.
98. Nielsen J: **Metabolic control analysis of biochemical pathways based on a thermokinetic description of reaction rates.** *Biochemistry Journal* 1997, **321**:133-138.
99. Heyland J, Fu J, Blank LM: **Correlation between TCA cycle flux and glucose uptake rate during respiro-fermentative growth of *Saccharomyces cerevisiae*.** *Microbiology* 2009, **155**:3827-37.
100. Baumann K, Carnicer M, Dragosits M, et al.: **A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions.** *BMC Syst Biol* 2010, **4**:141.
101. Kleijn RJ, Winden W van, Gulik WM van, Heijnen JJ: **Revisiting the <sup>13</sup>C-label distribution of the non-oxidative branch of the pentose phosphate pathway based upon kinetic and genetic evidence.** *The FEBS Journal* 2005, **272**:4970-82.
102. Selivanov V a, Puigjaner J, Sillero A, et al.: **An optimized algorithm for flux estimation from isotopomer distribution in glucose metabolites.** *Bioinformatics* 2004, **20**:3387-97.
103. Kümmel A, Panke S, Heinemann M: **Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data.** *Molecular Systems Biology* 2006, **2**:0034.
104. Zamboni N, Kümmel A, Heinemann M: **anNET: a tool for network-embedded thermodynamic analysis of quantitative metabolome data.** *BMC Bioinformatics* 2008, **9**:199.
105. Carnicer M, Canelas AB, Pierick A, et al.: **Development of quantitative metabolomics for *Pichia pastoris*.** *Metabolomics* 2012, **8**:284-298.
106. Seifar RM, Ras C, Dam JC van, et al.: **Simultaneous quantification of free nucleotides in complex biological samples using ion pair reversed phase liquid chromatography isotope dilution tandem mass spectrometry.** *Analytical Biochemistry* 2009, **388**:213-9.

107. Cipollina C, Pierick A ten, Canelas AB, et al.: **A comprehensive method for the quantification of the non-oxidative pentose phosphate pathway intermediates in *Saccharomyces cerevisiae* by GC-IDMS.** *Journal of Chromatography. B* 2009, **877**:3231-6.
108. Orij R, Brul S, Smits GJ: **Intracellular pH is a tightly controlled signal in yeast.** *Biochimica et Biophysica Acta* 2011, **1810**:933-944.
109. Stephanopoulos GN, Aristidou AA, Nielsen J: **Comprehensive Models for Cellular Reactions.** In *Metabolic Engineering: Principles and Methodologies*. 1997.
110. Wisselink HW, Cipollina C, Oud B, et al.: **Metabolome, transcriptome and metabolic flux analysis of arabinose fermentation by engineered *Saccharomyces cerevisiae*.** *Metabolic Engineering* 2010, **12**:537-51.
111. Klimacek M, Krahulec S, Sauer U, Nidetzky B: **Limitations in xylose-fermenting *Saccharomyces cerevisiae*, made evident through comprehensive metabolite profiling and thermodynamic analysis.** *Applied and Environmental Microbiology* 2010, **76**:7566-74.
112. Nelson DL, Cox MM: *Lehninger. Principles of Biochemistry*. Fourth Edi. Omega; 2011.
113. Ozcan S, Johnston M: **Function and regulation of yeast hexose transporters.** *Microbiology and Molecular Biology Reviews* 1999, **63**:554-69.
114. Does AL, Bisson LF: **Comparison of glucose uptake kinetics in different yeasts.** *Journal of Bacteriology* 1989, **171**:1303-1308.
115. Urk H van, Postma E, Scheffers W a, Dijken JP van: **Glucose transport in crabtree-positive and crabtree-negative yeasts.** *Journal of General Microbiology* 1989, **135**:2399-406.
116. Canelas AB, Gulik WM van, Heijnen JJ: **Determination of the cytosolic free NAD/NADH ratio in *Saccharomyces cerevisiae* under steady-state and highly dynamic conditions.** *Biotechnology and Bioengineering* 2008, **100**:734-43.
117. Veech RL: **The determination of the redox states and phosphorylation potential in living tissues and their relationship to metabolic control of disease phenotypes.** *Biochemistry and Molecular Biology Education* 2006, **34**:168-179.
118. Pitkanen J, Aristidou AA, Salusjärvi L, Ruohonen L, Penttilä M: **Metabolic flux analysis of xylose metabolism in recombinant *Saccharomyces cerevisiae* using continuous culture.** *Metabolic Engineering* 2003, **5**:16-31.
119. Hahn-Hägerdal B, Karhumaa K, Fonseca C, Spencer-Martins I, Gorwa-Grauslund MF: **Towards industrial pentose-fermenting yeast strains.** *Applied Microbiology and Biotechnology* 2007, **74**:937-53.

120. Cascante M, Marin S: **Metabolomics and fluxomics approaches**. In *Essays in Biochemistry*. Portland Press; 2008:67-82.
121. Baumann K, Adelantado N, Lang C, Mattanovich D, Ferrer P: **Protein trafficking, ergosterol biosynthesis and membrane physics impact recombinant protein secretion in *Pichia pastoris***. *Microb Cell Fact* 2011, **10**:93.
122. Junker BH, Klukas C, Schreiber F: **VANTED: a system for advanced data analysis and visualization in the context of biological networks**. *BMC Bioinformatics* 2006, **7**:109.
123. Wang L, Birol I, Hatzimanikatis V: **Metabolic control analysis under uncertainty: framework development and case studies**. *Biophysical Journal* 2004, **87**:3750-63.
124. Baumann K, Dato L, Graf AB, et al.: **The impact of oxygen on the transcriptome of recombinant *S. cerevisiae* and *P. pastoris* - a comparative analysis**. *BMC Genomics* 2011, **12**:218.
125. Cos O, Ramón R, Montesinos JL, Valero F: **Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast *Pichia pastoris* under different promoters: a review**. *Microb Cell Fact* 2006, **5**:17.
126. Görgens JF, Zyl WH van, Knoetze JH, Hahn-Hägerdal B: **The metabolic burden of the PGK1 and ADH2 promoter systems for heterologous xylanase production by *Saccharomyces cerevisiae* in defined medium**. *Biotechnol Bioeng* 2001, **73**:238-45.
127. Cos O, Serrano A, Montesinos JL, et al.: **Combined effect of the methanol utilization (Mut) phenotype and gene dosage on recombinant protein production in *Pichia pastoris* fed-batch cultures**. *J Biotechnol* 2005, **116**:321-35.
128. Vigentini I, Brambilla L, Branduardi P, et al.: **Heterologous protein production in *Zygosaccharomyces bailii*: physiological effects and fermentative strategies**. *FEMS Yeast Res* 2005, **5**:647-52.
129. Krogh AM, Beck V, Christensen LH, et al.: **Adaptation of *Saccharomyces cerevisiae* expressing a heterologous protein**. *J Biotechnol* 2008, **137**:28-33.
130. Hensing MC, Rouwenhorst RJ, Heijnen JJ, Dijken JP van, Pronk JT: **Physiological and technological aspects of large-scale heterologous-protein production with yeasts**. *Antonie van Leeuwenhoek* 1995, **67**:261-79.
131. Resina D, Bollók M, Khatri NK, et al.: **Transcriptional response of *P. pastoris* in fed-batch cultivations to *Rhizopus oryzae* lipase production reveals UPR induction**. *Microb Cell Fact* 2007, **6**:21.
132. Çelik E, Çalik P, Oliver SG: **A structured kinetic model for recombinant protein production by Mut<sup>+</sup> *Pichia pastoris***. *Chemical Engineering Science* 2009, **64**:5028-5035.

133. Heyland J, Fu J, Blank LM, Schmid A: **Quantitative physiology of *Pichia pastoris* during glucose-limited high-cell density fed-batch cultivation for recombinant protein production.** *Biotechnol Bioeng* 2010, **107**:357-68.
134. Heyland J, Fu J, Blank LM, Schmid A: **Carbon metabolism limits recombinant protein production in *Pichia pastoris*.** *Biotechnol Bioeng* 2011, **108**:1942-53.
135. Jordà J, Jouhten P, Cámara E, et al.: **Metabolic flux profiling of recombinant protein secreting *Pichia pastoris* growing on glucose:methanol mixtures.** *Microb Cell Fact* 2012:Submitted.
136. Görgens JF, Zyl WH van, Knoetze JH, Hahn-Hägerdal B: **Amino acid supplementation improves heterologous protein production by *Saccharomyces cerevisiae* in defined medium.** *Appl Microbiol Biotechnol* 2005, **67**:684-91.
137. Görgens JF, Passoth V, Zyl WH van, Knoetze JH, Hahn-Hägerdal B: **Amino acid supplementation, controlled oxygen limitation and sequential double induction improves heterologous xylanase production by *Pichia stipitis*.** *FEMS Yeast Res* 2005, **5**:677-83.
138. Hahn-Hägerdal B, Karhumaa K, Larsson CU, et al.: **Role of cultivation media in the development of yeast strains for large scale industrial use.** *Microb Cell Fact* 2005, **4**:31.
139. Stephanopoulos GN, Aristidou AA, Nielsen J: **Material Balances and Data Consistency.** In *Metabolic Engineering: Principles and Methodologies*. 1997.
140. Raiford DW, Heizer EM, Miller RV, et al.: **Do amino acid biosynthetic costs constrain protein evolution in *Saccharomyces cerevisiae*?** *J Mol Evol* 2008, **67**:621-30.
141. Dragosits M, Stadlmann J, Albiol J, et al.: **The Effect of Temperature on the Proteome of Recombinant *Pichia pastoris* research articles.** *J Proteome Res* 2009, **8**:1380-1392.
142. Görgens JF, Planas J, Zyl WH van, Knoetze JH, Hahn-Hägerdal B: **Comparison of three expression systems for heterologous xylanase production by *S. cerevisiae* in defined medium.** *Yeast* 2004, **21**:1205-17.
143. Pfeffer M, Maurer M, Köllensperger G, et al.: **Modeling and measuring intracellular fluxes of secreted recombinant protein in *Pichia pastoris* with a novel <sup>34</sup>S labeling procedure.** *Microb Cell Fact* 2011, **10**:47.
144. Canelas AB: **Towards quantitative metabolomics and in vivo kinetic modeling in *Saccharomyces cerevisiae*.** *Thesis* 2010.
145. Uchida M, Sun Y, Mcdermott G, et al.: **Quantitative analysis of yeast internal architecture using soft X-ray tomography.** *Yeast* 2011:227-236.
146. Oud B, Maris AJ a van, Daran J-M, Pronk JT: **Genome-wide analytical approaches for reverse metabolic engineering of industrially relevant phenotypes in yeast.** *FEMS Yeast Research* 2011.

147. Eunen K van, Bouwman J, Daran-Lapujade P, et al.: **Measuring enzyme activities under standardized in vivo-like conditions for systems biology.** *The FEBS Journal* 2010, **277**:749-60.
148. Goel A, Santos F, Vos WMD, Teusink B, Molenaar D: **A standardized assay medium to measure enzyme activities of *Lactococcus lactis* while mimicking intracellular conditions.** *Applied and Environmental Microbiology* 2011.

## List of Scientific Contributions

### Scientific Articles

1. Resina, David; Mauer, Michael; Cos, Oriol; Arnau, Carolina; Carnicer, Marc; Marx, Hans; Gasser, Brigitte; Valero, Francisco; Mattanovich, Diethard; Ferrer, Pau. **Engineering of bottlenecks in *Rizopus oryzae* production in *Pichia pastoris* using the nitrogen source-regulated *FLD1* promoter**. New Biotechnology. 2009, 25: 396-403.
2. Carnicer, Marc; Baumann, Kristin; Töplitz, Isabel; Sánchez-Ferrando, Francisco; Mattanovich, Diethard; Ferrer, Pau and Albiol, Joan. **Macromolecular and elemental composition analysis and extracellular metabolite balances of *Pichia pastoris* growing at different oxygen levels**. Microbial Cell Factories. 2009, 8: 65-78.
3. Baumann, Kristin<sup>§</sup>; Carnicer, Marc<sup>§</sup>; Dragosits, Martin; Graf, Alexandra B; Stadlmann, Johannes; Jouhten, Paula; Maaheimo, Hannu; Gasser, Brigitte; Albiol, Joan; Mattanovich, Diethard and Ferrer, Pau. **A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions**. BMC Systems Biology. 2010, 4: 141-163.(<sup>§</sup> Equal Contribution)
4. Carnicer, M., Canelas, A. B., Pierick, A., Zeng, Z., Dam, J., Albiol, J., Ferrer, P., Heijnen, J. J., and Gulik, W. **Development of quantitative metabolomics for *Pichia pastoris***. Metabolomics 2012, 8:284-298.
5. Marc Carnicer, Angela ten Pierick, Jan van Dam, Joseph J. Heijnen, Joan Albiol, Walter van Gulik and Pau Ferrer. **Quantitative metabolomics analysis of amino acid metabolism in recombinant *Pichia pastoris* under different oxygen availability conditions**. (Submitted)
6. Marc Carnicer, Krisitn Baumann, Angela ten Pierick, Zhen Zeng, Reza M. Seifar, Jan van Dam, Joseph J. Heijnen, Pau Ferrer, Walter van Gulik and Joan Albiol. ***Pichia pastoris* central carbon metabolism adaptation to recombinant protein production and oxygen availability**. (In preparation)
7. Marc Carnicer, Angela ten Pierick, Zhen Zeng, Reza M. Seifar, Jan van Dam, Joseph J. Heijnen, Pau Ferrer, Walter van Gulik and Joan Albiol. **Oxygen impact on *Pichia pastoris* metabolome: a thermodynamic study**. (In preparation)

### Oral Presentation

1. Carnicer, M., Baumann, K., Pierick, A., Zeng, Z., Seifar, R., Dam, J., Albiol, J., Ferrer, P., Heijnen, J. J., and Gulik, W. **Integration of transcriptomic and metabolomic data reveals distinct traits in the metabolic regulation of recombinant *Pichia pastoris*' central carbon metabolism**. Pichia-2012. Alpbach, Austria. 2012.

### Poster Presentation

1. Carnicer, M., Jouhten, P., Töplitz, I., Baumann, K., Sánchez-Ferrando, F., Mattanovich, D., Maaheimo, H., Ferrer P. and Albiol J. ***Metabolic Flux profiling of Pichia pastoris expressing an antibody fragment under different oxygen conditions.*** 5<sup>th</sup> Conference on Recombinant Protein Production. Alguero, Italy. 2008.
2. Carnicer, M., Jouhten, P., Töplitz, I., Baumann, K., Sánchez-Ferrando, F., Mattanovich, D., Maaheimo, H., Ferrer P. and Albiol J.. ***Metabolic Flux profiling of Pichia pastoris expressing an antibody fragment under different oxygen conditions.*** The 14<sup>th</sup> European Congress on Biotechnology. Barcelona, Spain. 2009.
3. Carnicer, M., Canelas, A. B., Pierick, A., Zeng, Z., Dam, J., Albiol, J., Ferrer, P., Heijnen, J. J., and Gulik, W. ***Systematic evaluation of metabolite leakage to obtain an optimal quenching protocol for Pichia pastoris.*** 4<sup>th</sup> Conference of Physiology of Yeast and Filamentous Fungy (PYFF). Rotterdam, The Netherlands. 2010.
4. Carnicer, M., Canelas, A. B., Pierick, A., Zeng, Z., Seifar, R., Dam, J., Albiol, J., Ferrer, P., Heijnen, J. J., and Gulik, W. ***Impact of antibody fragment production on Pichia pastoris metabolome.*** Trends in Metabolomics - Analytics and Applications. Frankfurt, Germany. 2011. Poster presentation.