THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Metabolic Engineering of Central Carbon Metabolism in *Saccharomyces cerevisiae*

The contribution of systems biology to physiological studies

MARTA PAPINI

Systems and Synthetic Biology

Department of Chemical and Biological Engineering

CHALMERS UNIVESITY OF TECHNOLOGY

Göteborg, Sweden 2012

ABSTRACT

Saccharomyces cerevisiae is one of the most well characterized yeast of large industrial interest due to several attractive features such as its capability to efficiently convert glucose into ethanol and carbon dioxide at high flux, its amenability to genetic modification and the presence of extensive knowledge databases. For these reasons, it is often considered a suitable cell factory for the production of different classes of compounds. The central carbon metabolism of *S. cerevisiae* has been object of numerous studies aiming at elucidating the complex mechanisms underlying the tight cellular balance arising as a consequence of a wide variety of regulatory pathways and phenomena, such as the Crabtree effect. To engineer efficient cell factories, a deep knowledge of cellular metabolism and its regulatory mechanisms is of fundamental importance to further de-regulate regulatory circuits hampering the reaching of desired characteristics. In this perspective, metabolic engineering and systems biology can supply valid and more efficient approaches for a global understanding of the yeast cell. Although the central carbon metabolism of *S. cerevisiae* has been object of numerous investigations, the design of intuitive metabolic engineering strategies has often encountered several hurdles due to the tight regulation exerted by the cell. In this doctoral thesis, the contribution of systems biology and metabolic engineering to gaining new insight into the central carbon metabolism of *S. cerevisiae* is addressed. Different metabolic engineering approaches to re-wire the glycolytic flux are presented. While the first and most direct approach is based on a deletion in the lower part of glycolysis through the construction of a phosphoglycerate mutase (*Δgpm1*) mutant, a more elaborated approach is described in the expression of the *Aspergillus nidulans* phosphoketolase pathway in *S. cerevisiae*. Fermentation technology as well as tools within systems biology, such as DNA microarrays and ^{13}C flux analysis, were used as tools for the characterization of the recombinant phenotypes, highlighting the challenges faced by the re-wiring of essential pathways, thus indicating the robustness and the primary role in metabolism of the glycolytic pathway.

To undertake a different approach to investigate the central carbon metabolism of *S. cerevisiae*, a high-throughput based comparison with the Crabtree negative yeast *Scheffersomyces stipitis* (*Pichia stipitis*) was performed. Integrative, system-level analysis of the two yeasts growing aerobically under glucose excess and glucose limitation conditions contributed to gain insight into a different regulation of the central carbon metabolism of the two yeasts. What emerges from the different works performed is that physiological studies based on metabolic engineering benefit from systems biology methodologies such as transcriptomics, fluxomics and metabolomics, supporting the characterization of recombinant and wild-type strains and helping to bridge the gap between genotype and phenotype. As both microarrays and RNA-seq have been used to characterize transcriptomes of different yeast strains, an attempt to address and compare the performances of the two transcriptomic platforms is presented in the last chapter of this thesis where a technical comparison between the two methodologies is described, addressing the contribution of the different steps involved in RNA-seq analysis to obtain biologically meaningful data.

Keywords: Metabolic engineering, system biology, *Saccharomyces cerevisiae*, DNA microarrays, RNA-sequencing, 13C-based flux analysis, fermentation technology, *Scheffersomyces stipitis*

LIST OF PUBLICATIONS

This thesis is based on the following publications:

I. **Phosphoglycerate mutase knock-out mutant *Saccharomyces cerevisiae*: physiological investigation and transcriptome analysis**

Papini M, Nookaew I, Scalcinati G, Siewers V, Nielsen J.

Biotechnol J. 2010 Oct;5 (10):1016-27.

II. **Physiological characterization of recombinant *Saccharomyces cerevisiae* expressing the *Aspergillus nidulans* phosphoketolase pathway: validation of activity through ^{13}C-based metabolic flux analysis**

Papini M, Nookaew I, Siewers V, Nielsen J.

Appl Microbiol Biotechnol. 2012 Feb 26. [Epub ahead of print]

III. ***Scheffersomyces stipitis*: a comparative systems biology study with the Crabtree yeast *Saccharomyces cerevisiae***

Papini M, Nookaew I, Uhlén M, Nielsen J.

Manuscript submitted

IV. **From reads to differential gene expression: a comprehensive comparison of RNA-seq based transcriptome analysis and cross comparison with microarrays**

Nookaew I#, Papini M#, Porputtapong N, Scalcinati G, Fagerberg L, Uhlén M, Nielsen J.

# The authors contributed equally to this work

Manuscript submitted

V. **Systems biology of industrial microorganisms**

Papini M, Salazar M, Nielsen J.

Adv Biochem Eng Biotechnol. 2010;120:51-99.

ABBREVIATIONS COMMONLY USED

^{13}C: carbon 13

Acetyl-CoA: acetyl coenzyme A

ATP: Adenosine-5'-triphosphate

CO_2: carbon dioxide

DNA: deoxyribonucleic acid

DNA-seq: DNA-sequencing

FAD+: flavin adenine dinucleotide

NAD+: nicotinamide adenine dinucleotide

NADH: reduced form of nicotinamide adenine dinucleotide

NADP+: nicotinamide adenine dinucleotide phosphate

NADPH: reduced form of nicotinamide adenine dinucleotide phosphate

NGS: next generation sequencing

ORF: open reading frame

PHK: phosphoketolase

PPP: pentose phosphate pathway

RNA-seq: RNA sequencing

RNA: ribonucleic acid

TCA: tricarboxylic acid cycle

μmax: specific growth rate (h^{-1})

TABLE OF CONTENTS

1 INTRODUCTION

1.1 *S. cerevisiae* cell factory and Industrial Biotechnology

This doctoral thesis is about the contribution of metabolic engineering and systems biology to the global understanding of central carbon metabolism of the yeast *Saccharomyces cerevisiae*. *S. cerevisiae*, commonly known as bakers' yeast, is probably the most studied and characterized yeast, belonging to the class of *Saccharomycetes*. The word *Saccharomyces* comes from the greek σακχαρ= sugar and μύκης= mold, indicating a fungus that can grow on sugar. *S. cerevisiae* is very well known for its ability to ferment, under fully aerobic conditions, glucose to ethanol and carbon dioxide, thus been classified as a Crabtree positive yeast. Because of this property, it has been used since ancient times for the production of alcoholic beverages and food (Gelinas, 2009). Besides its usage in brewing, wine production and baking products (Yeh, *et al.*, 2009), the yeast *S. cerevisiae* is commonly used as a model system for the study of numerous cellular processes such as apoptosis and ageing (Petranovic & Nielsen, 2008). The availability of advanced molecular biology techniques and the establishment of robust fermentation technologies together with its GRAS (Generally Regarded as Safe) status, made this yeast an attractive platform for the production of both chemicals and pharmaceuticals compounds, paving the way for the concept of cell factory. In this regards, with the arising of metabolic engineering as a science to install in microorganisms recombinant pathways to produce molecules of industrial relevance (Bailey, 1991), *S. cerevisiae*, together with *Corynebacterium glutamycum* and *Escherichia coli*, was one of workhorses for metabolic engineering applications driven by industrial interest. The interest of the industrial biotech market in *S. cerevisiae* is tremendous as this microorganism has the potential to be in the near future the workhorse for production of sustainable chemicals. The reasons underlying the choice of *S. cerevisiae* as a suitable host for the production of chemical compounds are several; it can tolerate relatively low pH, thus allowing the recovery of the acid in its carboxylated form, decreasing the associated downstream costs, it has a very high glycolytic flux thus through appropriate engineering strategies it is possible to obtain high productivities; additionally, several studies have shown that it is possible to expand the substrate range of *S. cerevisiae* for utilization of naturally abundant pentose sugars like xylose and arabinose (Ha, *et al.*, 2011). This capability can be directly coupled to the production of chemicals, opening new possibilities for the development of sustainable processes, thus conferring to this microorganism an additional advantageous feature (Lynd, *et al.*, 1999). Numerous works aim at the engineering of *S. cerevisiae* as cell factory for the production of different classes of compound, from bioethanol to organic acids, such as succinic (Raab, *et al.*, 2010) and lactic acid (Porro, *et al.*, 1995), biofuels and recombinant therapeutic proteins such as insulin (Kjeldsen, 2000) and the antimalarial precursor artemisinic acid through the engineering of the isoprenoid pathway (Ro, *et al.*, 2006). Currently, *S. cerevisiae* is the preferred cell factory for one of the world largest industrial production: bioethanol (Otero & Nielsen, 2010) and it is used during continuous cultivations for the production of insulin. Significant efforts have been focused to engineer this yeast for efficient and competitive pentose fermentation (Kuyper, *et al.*, 2004, Van Vleet & Jeffries, 2009).

With the sequencing of the *S. cerevisiae* genome in 1996 (Goffeau, *et al.*, 1996) and the era of "genomics revolution", several "omics" techniques started to be routinely applied (Goffeau, 2000). Studies based on hybridization arrays (DeRisi, *et al.*, 1997) (Daran-Lapujade, *et al.*, 2009) and tiling arrays (Juneau, *et al.*, 2007) led to an increased understanding of yeast and, together with the development of genome-scale metabolic models (GEMs), contributed to the identification of new, non intuitive targets for metabolic engineering applications (Nissen, *et al.*, 2000) (Bro, *et al.*, 2006) (Grotkjaer, *et al.*, 2005). The last decade, thanks to the development of next-generation sequencing (NSG) such as DNAseq and RNAseq, saw the establishment of these technologies as a efficient tools to unravel the complexity of genomes and transcriptomes (Nagalakshmi, *et al.*, 2008, Daran-Lapujade, *et al.*, 2009). The availability of a complete genome sequence and whole-genome transcriptomic information has contributed significantly to the understanding of *S. cerevisiae*, allowing the identification of genetic variation among strains, splicing versions and UTRs, offering a new perspective in analyzing differential expression of mRNAs.

1.2 Systems Biology and Metabolic Engineering

As *S. cerevisiae* has been used since ancient times for production of goods, there has always been an increasing attention to manipulate the yeast cell in order to improve its performances and to develop strains with new traits and specific characteristics to be utilized in industrial processes. Traditionally, strain improvement was carried out through random mutagenesis followed by screening of the desired phenotype among a high number of mutants. Despite great success, these methods are time-consuming and generate mutation that often are not beneficial and do not contribute to confer the desired characteristics; additionally, even in the successful cases, no understanding of the cellular phenotype obtained can be gained. In respect to that, the development of the science of metabolic engineering provided a valid alternative to traditional mutagenesis for the engineering of the cell factories, allowing the rationale engineering of the desired cellular phenotype(s) through recombinant DNA technologies (Bailey, 1991). Classical metabolic engineering strategies are based on the identification of rate-limiting steps in a certain pathway (e.g. pathway for the generation of a desired molecule) and on the attempt to alleviate this regulation by overexpressing the genes coding for the responsible enzymes. Through metabolic engineering it is also possible to confer new traits by recruiting heterologous activities and overexpressing in the production host. In this respect, the concept of metabolic engineering can be broaden and defined as direct improvement of cellular phenotypes through the introduction, deletion, and modification of metabolic pathways (Ostergaard, *et al.*, 2000). Despite holding great potential, metabolic engineering had to face several challenges related to the engineering of complex phenotypes. This is because the development of efficient strains requires the combination in the production host of different characteristics such as increased production rate of the desired product together with increased growth rate and decreased formation of by-products. These features cannot be easily conferred as the cell machinery is subjected to a tight regulation. The advent of inverse metabolic engineering accounted for an iterative cycle that encompasses 1) the identification, construction or calculation of the desired

phenotype 2) the determination of factors that contribute to the desired phenotype and 3) creating that phenotype in another strain through genetic engineering or environmental manipulation (Bailey, *et al.*, 2002). Inverse metabolic engineering has developed as a cycle where previous successful recombinant host can be further characterized to elucidate its characteristic in order to detect new targets for improved strategies (Otero, 2009).

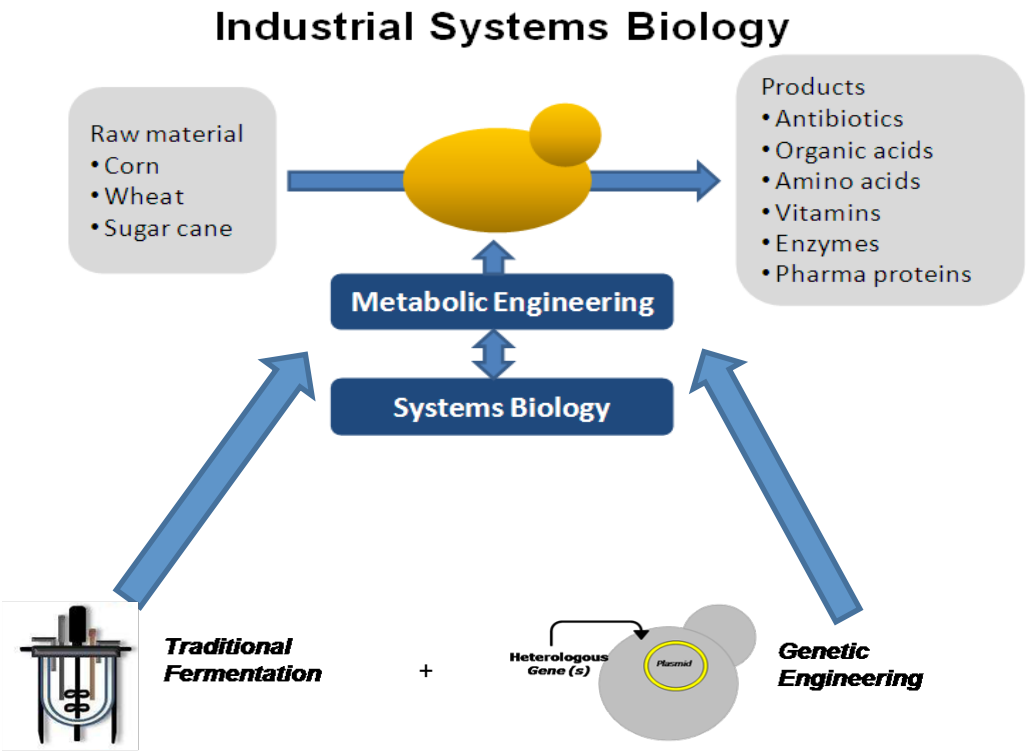

Fig. 1 The contribution of Systems Biology and Metabolic Engineering to the design of the cell factories. The purpose of industrial systems biology is to engineer efficient *S. cerevisiae* cell factories able to convert the raw substrate into different products. The strain engineering is accomplished through metabolic engineering strategies based on genetic engineering and fermentation technology, while systems biology support the screening and characterization of the recombinant cell phenotype, allowing the identification of bottlenecks for further improvement

In the last decade, the development of systems biology accounted for improvements in metabolic engineering strategies. Systems biology is a very broad research field and several the definition have been given (Nielsen & Jewett, 2008), however what clearly emerge from the different pioneering works on systems biology is the common aim to furnishing a better understanding of the functioning of the living cell giving by providing a holistic view rather than explaining single, isolated processes and to explain quantitatively biological systems through mathematical modeling (Nielsen & Vidal, 2010). Systems biology is a highly interdisciplinary approach and encompasses multiple techniques, from X-omic technologies to the application of metabolic mathematical modeling for the prediction of cellular function. The continuous development in systems biology techniques allowed to better understand cellular phenotype and contributed to gain insight into the metabolic regulation underlying physiological processes, by integrating experimental analysis and modeling

efforts (Heinemann & Sauer, 2010). Hereby systems biology aims at mapping regulatory structure and further de-regulate them (Papini, *et al.*, 2010) by combining omics techniques and mathematical modeling on the engineered and wild-type phenotypes in order to identify new, non intuitive targets to *ad-hoc* design metabolic engineering strategies. A representation of the contribution of systems biology and metabolic engineering to industrial biotechnology can be found in Fig.1

1.2.1 The systems biology toolbox

In this doctoral thesis, systems biology is regarded to as a support discipline in direct and inverse metabolic engineering strategies to gain insight into the phenotype of recombinant and wild-type strains. Systems biology is the quantitative collection, analysis and integration of whole-genome–scale data (Otero & Nielsen, 2010) thus, within systems biology, different methodologies are applied to generate high-throughput data that can be used for the understanding of cellular phenotypes. The integration of the large amount of data generated is the main challenge that systems biology has to face in order to reach a holistic view of the functioning of the cell. In this regards, the reconstruction of Genome-Scale Metabolic Models (GEMs) led to a significant step-forward, allowing for the reconstruction of models of cellular metabolism describing formation and depletion of each metabolites based on mass-balance constraints. These models allow prediction of cellular behavior, using basic stoichiometric information. The most applied techniques in systems biology are briefly introduced in Table 1.

Table 1. The Systems Biology Toolbox

Tool	***Description***	***Methods***	***Applications***
Genomics	Comprehensive study of the genome of a certain organisms (McKusick, 1997).	• Sanger sequencing • Next Generation Sequencing (SOLIDA, Illumina, 454)	Genome sequence information and gene function (functional genomics)
Transcriptomics	Study of the level of mRNAs expressed by the cell under a given condition (Gentalen & Chee, 1999, Devaux, *et al.*, 2001).	• Hybridization arrays (Affymetrix Arrays) • Tiling Arrays • RNA sequencing (Huber, *et al.*, 2006)	Analysis of differential expression levels of the entire set of genes. Identification of UTR, intergenic region and splicing variation.
Proteomics	Analysis of the entire set of proteins present in a cell in a certain condition (Humphery-Smith & Blackstock, 1997, Salih, 2005).	• 2D –PAGE gels • Protein Arrays • GC-MS/ LC-MS • Phosphoproteomics	Identification of function of proteins; protein-protein interaction and activation by phosphorylation.
Metabolomics	The measurement of all metabolites (intra and extracellular) present in the cell in a certain moment. (Villas-Boas, *et al.*, 2005, Villas-Boas, *et al.*, 2006, Mapelli, *et al.*, 2008)	• Intra or extracellular metabolome analysis. Different quenching and extraction methods are available. • GC-MS /LC-MS	Identification and quantification of metabolites as key compounds to elucidate a certain metabolic behavior.
Fluxomics	Measurement of the ensemble of metabolic fluxes	• Flux distribution is based on ^{13}C labeling	Flux analysis provide information about activity of

	active in the cell under a certain condition (Wiechert, 2001, Wittmann, 2007).	and analysis of enrichment patterns of proteinogenic amino-acids • GC-MS / NMR	pathways (Nielsen, 2003) and topology of the metabolic network.

The techniques mostly applied through this PhD research are flux analysis, metabolome analysis and transcriptomics. As the principles and methods of transcriptome analysis will be addressed in details in the last chapter of the thesis, here I would like to give a brief overview about flux analysis.

1.2.2 Flux analysis

The expression of metabolic fluxes in a cell represents the final outcome of cellular regulation in response to a certain condition. For this reason, analysis of how metabolic fluxes are distributed in the cell provides useful and realistic information about its metabolic state, bridging the gap to what is observed at the gene transcription level and the level of the gross phenotype.

Differently from others methods, the distribution of intracellular fluxes in a cell cannot be analyzed directly but, by feeding the cell with labeled substrates, it is possible to observe how the labeled material is distributed over the metabolic network and then calculate the flux distribution. Generally, measurement of fluxes relies on assumptions. For instance, we assume that all fluxes into a certain intracellular metabolite pool are in balance with the fluxes out of that pool; this implies that the balances about each metabolite impose a number of constraints on the system. If we have J fluxes and K metabolites we will have F degree of freedom F=J –K, thus by measuring the F fluxes the remaining fluxes can be calculated (Nielsen, 2003).

Despite in the early stage of metabolic flux analysis (MFA) determination of flux distribution was based on the measurement of exchange fluxes (stoichiometric MFA), this method does not provide good precision and accuracy. Nowadays, the most common way for flux measurement relies on carbon labeling experiments (CLE) (Wiechert, 2001). By feeding the cell with ^{13}C-labeled glucose, the label becomes distributed over the entire metabolic network, allowing the measurement of enrichment patterns of proteinogenic amino-acids through NMR or GC-MS. Once spectral data are available, it is necessary to combine them with biochemical transition of the cellular reactions as the fate of the isotopomers is determined by the fate of the carbon atoms of each metabolite. From the analysis of enrichment patterns of a certain metabolite it is possible to established which pathway are active. To evaluate the distribution of fluxes in the metabolic network, mathematical models describing how the labeled material is distributed are required (metabolic network model), increasing the mathematical complexity of this method (Wiechert & de Graaf, 1996). Based on these models and the information experimentally obtained, it is possible to run simulations until reaching of an optimal solution describing the intracellular network (Wiechert, *et al.*, 1997, Wiechert, *et al.*, 2001). For this purpose several algorithms and models are present; here it was used the model established and validated by Gombert *et al.* (Gombert, *et al.*, 2001, Grotkjaer, *et al.*, 2005).

Despite being an extremely powerful technique, providing realistic information about the metabolic state at a certain condition, intracellular flux analysis is not extensively used. A reason for that can be found in the

complexity of the calculations required for flux estimation from enrichment patterns and the lack of general algorithms for the analysis and statistical evaluation.

2. 1 *S. cerevisiae* central carbon metabolism

The metabolism of *S. cerevisiae* has been the object of studies since the XVII century (1680) when the Dutch naturalist Anton van Leeuwenhoek started to study its morphology. In 1785 the French Charles Cagnaird de la Tour discovered that fermentation was not a chemical but indeed a microbial process, as it was later clearly proven by Louis Pasteur, who defined the process of fermentation as "respiration without air"(Pasteur, 1876) (Barnett, 2000). Since then, *S. cerevisiae* has been extensively characterized in light of its capability to convert glucose to ethanol and, due to its industrial relevance, ethanol fermentation is one of the most well characterized and exploited microbial processes.

The term central carbon metabolism refers to the set of anabolic and catabolic reactions needed by the cell for the biosynthesis of precursor and to generate energy. A schematic representation of *S. cerevisiae* central carbon metabolism can be found in Fig. 2. In *S. cerevisiae*, the central carbon metabolism encompasses both catabolic and anabolic pathways. Among the catabolic ones we have a) **glycolysis**, essential for the break-down of sugar, generating ATP and reducing power, b) **pentose-phosphate pathway** (PPP) used by the cell to generate NADPH used in bioreductive biosynthesis c) **tricarboxylic acid cycle** (TCA) or Krebs cycle, being an amphybolic pathway and needed to produce pyruvate, generating CO_2, $FADH_2$ and NADH. Linked to the Krebs cycle is d) the **glyoxylate cycle** which have an anaplerotic role in sustaining the replenishment intermediate for the correct operation of the TCA cycle and allowing for growth on C2 and C3 compounds. The reduced carriers generated through the TCA (NADH and $FADH_2$) are oxidized via the e) **electron transport chain**, located on the inner mitochondrial membrane. The electron transport chain pumps protons out of the inner mitochondrial membrane, creating a proton gradient used by the ATPase to synthesize ATP through a process called **oxidative phosphorylation**. f) **Gluconeogenesis** is one of the main anabolic pathway used by the cell to generate hexose phosphates during growth on C2 and C3 substrate.

All the pathways in the central carbon metabolism contribute to supply precursors for biomass formation, sustaining growth. In this regard, the different pathways are finely tuned to meet the exact need for building blocks and Gibbs free energy. This is because all the pathways interact among each other as they share cofactors (such as ATP, ADP and redox equivalents), substrates/ products (metabolites) and compartment, leading to a tight regulation of the cell. Despite this regulation, *S. cerevisiae* shows "metabolic uncoupling": the lack of coupling between anabolic requirements and catabolic energy production, meaning that catabolic activity is higher than what required for growth (Larsson, *et al.*, 1995); this phenomenon has been observed both during glucose-excess and under steady state conditions and brings in an additional level of complication when describing cellular phenotypes.

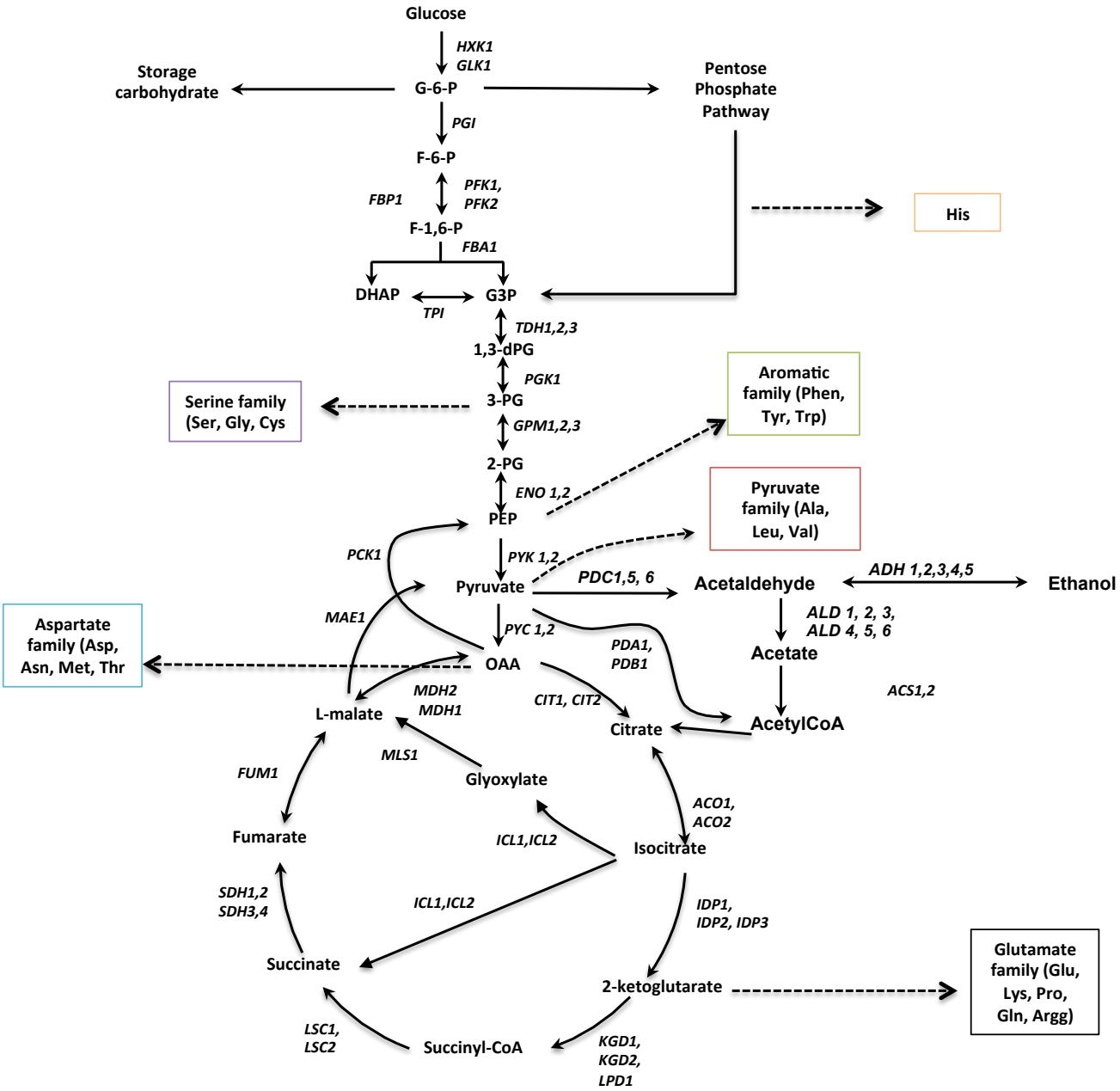

Fig. 2 Pathways involved in the central carbon metabolism of *S. cerevisiae*

What is extremely intriguing and most fascinating about the yeast metabolism', is the fact that different pathways are activated or down-regulated by the cell to satisfy its needs according to different environmental conditions, reflecting different metabolic states. At this point the question arises about how, when and why a certain pathway is active, introducing the concept of metabolic regulation. This is indeed a rather complicated question and some of the conditions necessary to activate a certain pathway will be addressed in the next paragraphs. It is worth to mention that we are still far from having a global understanding of the interconnection of regulatory mechanisms by which the cell machinery finely tunes its metabolism however, combining approaches giving detailed snapshots together with studies aiming at providing a global view of

the cellular processes can contribute to gain insight into these mechanisms and lead to a more complete understanding.

2.2. Fermentation, respiration and regulatory phenomena

S. cerevisiae is a Crabtree positive yeast, facultatively anaerobe, able to perform alcoholic fermentation of glucose under fully aerobic conditions. When growing aerobically, *S. cerevisiae* can both respire and ferment. Glucose can be fully oxidized, under respiratory conditions, to biomass and CO_2 or, under respiro-fermentative conditions, primarily to CO_2 and ethanol. As the growth of *S. cerevisiae* can be supported by mitochondrial oxidative phosphorylation or by substrate-level phosphorylation leading to ethanol fermentation, the activation of pathways contributing to respiro-fermentative or respiratory metabolism will depend on environmental factors such as the carbon source (and the amount) available (e.g. repressing / non-repressing and fermentable /non-fermentable sugars) and/or on the presence or absence of oxygen. On the contrary, anaerobic growth is supported exclusively by fermentative processes through substrate-level phosphorylation; under this condition, since no oxygen is present, the activation of specific "hypoxic" genes is required (Kwast, *et al.*, 1998). As mentioned above, the metabolic state of *S. cerevisiae* is dependent on the sugar present in the cultivation medium (raffinose and ethanol are, for instance, non-fermentable carbon sources) and, for glucose and other fermentable carbon source (such as fructose, maltose, mannose and galactose), the fermentation rate is dependent on the amount of glucose present (de Jong-Gubbels, *et al.*, 1995). The consumption of different carbon sources is differently regulated and leads to different metabolic conditions (e.g. raffinose and ethanol can be only respired, whereas glucose is both respired and fermented); additionally, some fermentable sugars, such as mannose and galactose, show higher respiration rates, compared to glucose, probably as results of inhibition of fermentation (Kappeli, 1986).

It is well established that during aerobic growth on medium containing glucose concentrations above a certain threshold, 0.5-0.8 mM (Verduyn, *et al.*, 1992), *S. cerevisiae* produces ethanol as a result of the Crabtree effect. This phenomenon is accompanied by an increase in the production of CO_2 and the secretion, in smaller quantities, of other metabolites such as acetate, pyruvic acid (van Dijken, *et al.*, 1993) and glycerol, needed to maintain the redox balance (Walker, 1998). Dissimilation of glucose to pyruvate through glycolysis is stoichiometrically linked to NADH formation, which is then re-oxidized through ethanol fermentation (Bakker, *et al.*, 2001). Alcoholic fermentation of glucose is used by yeast as a way to re-oxidize excess of cytosolic NADH (Overkamp, *et al.*, 2000)and it has recently been shown that when *S. cerevisiae* is supplied with an alternative heterologous oxidase, reduced aerobic ethanol fermentation is observed (Vemuri, *et al.*, 2007), thus suggesting that, in *S. cerevisiae*, the onset of fermentation is a consequence of the limited respiratory capacity.

Under anaerobic conditions, the only source of ATP generation is substrate-level phosphorylation. Anaerobically, *S. cerevisiae* shows fully fermentative metabolism and ethanol is the main product found, however formation of glycerol, carbon dioxide and smaller amount of others by-products (acetate, pyruvic

acid) are observed. The energy gain from alcoholic fermentation is lower than that obtained by respiratory dissimilation of glucose and the cell respond to a lower ATP yield under fermentative conditions by increasing glycolytic fluxes.

During fully respiratory growth on glucose, the pyruvate generated through the glycolytic pathway is oxidized to acetyl-CoA and further through the tricarboxylic acid cycle (TCA) to produce CO_2 and H_2O. Electrons generated in the oxidation process are captured by $FADH_2$ and NADH which are re-oxidized through mitochondrial respiration, leading to the generation of additional ATP. Several studies demonstrated that the respiratory capacity of *S. cerevisiae* is lower than other microorganisms and the theoretical P/O ratio (number of ADP molecules phosphorylated per pair of electron transferred to oxygen) is approximately 2 (Famili, *et al.*, 2003), but the operational P/O ratio is lower and probably in the range of 1-1.5 (Verduyn, *et al.*, 1991). The fully respiratory dissimilation of glucose yields more than 10 ATP per glucose, which can support a biomass yield of 0.5 g biomass per g glucose. The aerobic ATP yield is higher than what can be theoretically obtained through ethanol fermentation (Bakker, *et al.*, 2001) and the fermentation of glucose to ethanol and CO_2 brings a biomass yield 0.1 g biomass per g glucose. An interesting parameter used to determine the metabolic state of *S. cerevisiae* is the respiratory quotient RQ defined as the moles of CO_2 produced per mole of O_2 consumed. Under fully respiratory conditions, the RQ is around 1, whereas when shifting towards respiro-fermentative metabolism, the specific oxygen uptake rate becomes smaller and the RQ assumes values greater than 1.

To study the metabolism of *S. cerevisiae* under respiratory or fermentative conditions, different tools to keep the cell in the respiro-fermentative or in the respiratory state have been developed. During batch cultivations in the first phase of growth, respiration is repressed and the glucose consumed is fermented to ethanol and smaller amounts of by-products such as glycerol, acetate and pyruvate. When glucose is depleted, *S. cerevisiae* is able to efficiently adapt its metabolism and consume the ethanol produced in the first phase to gain energy, a phenomenon called diauxic growth. It was shown (Ephrussi, *et al.*, 1956) that the slow respiration rate of glucose observed during batch cultures is due to inhibition of the activity and/or the synthesis of the respiratory enzymes *cytochrome a*, *b* and *c* complex while the higher respiratory rates of sugars such as galactose and mannose are consequences of the de-repression of cytochrome complex synthesis (Fiechter, 1975). During chemostat cultivations at low dilution rates, when the glucose concentration is low (<100 mg L^{-1}), *S. cerevisiae* shows fully respiratory metabolism. It was suggested that when glucose is fed to the culture at low rates, its breakdown occurs oxidatively (Kappeli, 1986) and, under these conditions, the cells have enough time to adapt to respiratory conditions (Fiechter, 1975) (Barford & Hall, 1979) which results in a purely respiratory metabolism, yielding high amount of biomass with no production of other metabolites. When the dilution rate is increased above the critical one, fermentation is used in addition to respiration as *S. cerevisiae* shift its metabolism to respire-fermentative mode until reaching the wash-out.

Several regulatory phenomena are present in yeast species to respond to environmental changes such as the presence of oxygen or glucose. The mechanisms regulating the response to a certain condition took the name

from the scientist who discovered and characterized them. Differences in facultatively fermentative yeasts with respect to the onset of ethanol fermentation due to the presence oxygen, are ascribed to the phenomena known as ***Custer Effect*** and ***Pasteur Effect***. The Pasteur effect can be described as the decrease of fermentation efficiency in presence of air (Wyman, 2000), while the Custer effect (not observable in *S. cerevisiae*) indicates inhibition of fermentation by anaerobiosis.

An overview of the most characterized regulatory phenomena occurring in different yeasts can be found in Table 2.

Table 2. A brief overview of regulatory phenomena in different yeast species

Effect	Description	Remarks
Pasteur Effect	Inhibition of fermentation by presence of oxygen, associated with a decreased affinity for sugar uptake under aerobic conditions (Lagunas, 1981).	Insignificant during growth on glucose, mannose and galactose. Resting cells of *S. cerevisiae* show limited Pasteur effect.
Kluyver Effect	Yeasts such as *Candida utilis, Kluyveromyces wickerhamii and Debaryomyces yamadae* can ferment glucose only anaerobically. Maltose, lactose and sucrose can not be fermented (Kaliterna, *et al.*, 1995, Kaliterna, *et al.*, 1995).	Probably due to the slower uptake of sugars anaerobically.
Custer Effect	The yeast from the species *Dekkera* and *Brettanomyces* ferment glucose to ethanol faster under aerobic conditions; transient inhibition of fermentation by anaerobiosis (Scheffer.Wa, 1966).	Oxygen stimulates ethanol production due to a lack of intracellular NAD^+ (Walker, 1998).
Crabtree Effect	Suppression of respiration by high glucose concentration.	Observed also in tumor cells.

It was initially believed that alcoholic fermentation of sugars was taking place only during anaerobic growth. In 1929, Herbert Crabtree discovered in tumor cells what was named the **Crabtree effect**: the suppression of respiration due to high glucose conditions. It was only in 1966 that De Deken, detecting in yeast excess of CO_2 and ethanol formation during aerobic growth on glucose (Alexander, 1990), recognized in *S. cerevisiae* the same mechanism (De Deken, 1966). De Deken attributed the Crabtree effect to the inhibition of synthesis or activity of respiratory enzymes. He observed that at low glucose concentrations (6×10^{-3} M) "the rate of respiratory adaptation was increased with increasing glucose concentrations and with a decrease in the fermentation rate". On the contrary, at high glucose concentration, the fermentation rate increased while the rate of respiratory adaptation decreased.

It is thus clear that in *S. cerevisiae*, the Crabtree effect contributes to determine the extent of fermentation and respiration. What causes the Crabtree effect has been object of numerous investigations throughout decades of yeast physiological research and can by ascribed to a series of factors such as the regulation of

mitochondrial genes involved in respiration, the mode of sugar transport and the overflow through pyruvate decarboxylase (PDC). Two Crabtree effects can be described in yeast: one is the long-term effect, e.g. the production of ethanol at high growth rates regardless of the fermentation mode, and the short-term effect: the production of ethanol upon transition from glucose limitation to glucose excess (Vanurk, *et al.*, 1988). The long-term effect is ascribed to an insufficient respiratory capacity due to the repression of respiratory genes (Postma, *et al.*, 1989), whereas the second arises when respiratory metabolism becomes saturated, causing overflow at the level of pyruvate (Pronk, *et al.*, 1996).

In the field of yeast physiology, there has always been a constant interest in understanding what determines the onset of fermentative metabolism and numerous efforts have been focused on characterizing in details the metabolic states of *S. cerevisiae*. The metabolic shift from fermentative to oxidative growth has been investigated through different methods, from traditional physiological studies to microarrays-based works, as well as through metabolome and flux distribution analysis. Transcript levels during fermentative and respiratory conditions have been characterized based on expression arrays (DeRisi, *et al.*, 1997), showing differential expression of a wide number of gene families (over 400 transcripts changed by more than 2-fold). The metabolic shift observed when *S. cerevisiae* cells are transferred from a fermentable to a non-fermentable carbon source has also been addressed through microarray experiments (Kuhn, *et al.*, 2001), showing a decrease in translation efficiency. Transcriptome studies demonstrated to be a useful tool to compare expression patterns between different conditions however, applying microarray studies to look at gene expression provide limited biological insights regulatory phenomena. In regards to this, analysis of intracellular fluxes has proven to be a valid tool in characterizing cellular phenotypes as they reflect the final outcome of the regulation of the cell. Analysis of intracellular metabolic fluxes has been applied to characterize the shift towards fermentative conditions, providing significant information about how the fluxes are distributed in the cell as a result of regulatory phenomena. In a work by Frick and Wittman, the authors describe the re-arrangement in metabolic fluxes as a response to different dilution rates, through ^{13}C glucose labeling (Frick & Wittmann, 2005). *S. cerevisiae* was cultivated in chemostat and the dilution rate increased until the critical one, when yeast switches its metabolism from fully respiratory towards the respiro-fermentative. The shift towards respiro-fermentative conditions is characterized by a 5-fold increase in the glucose uptake rate, together with a decrease in the pentose phosphate pathway (PPP) flux and decrease in the fluxes through the tricarboxylic acid cycle (TCA). As the PPP is required for NADPH generation, the finding of a reduced flux through the PPP well correlates with the reduced biomass yield, suggesting that, under respiratory conditions, the PPP pathway is enough to sustain anabolic NADPH demand. An increased flux through the pyruvate carboxylase (PYC) was reported, indicating an increased formation of oxaloacetate (OAA) supply. Interestingly, the flux through pyruvate decarboxylase (PDC) was found to assume significant values also during fully oxidative conditions, probably indicating the fundamental role of this reaction in providing acetyl-CoA supply; this finding is in agreement with the increased acetyl-CoA transport in the mitochondria observed. However, the flux through the acetyl-CoA synthase (ACS) remains constant even at high dilution rates when acetate secretion increased.

The capability of shifting from fermentation to respiration under aerobic conditions, is one of the most fascinating and studied phenomenon of *S. cerevisiae* metabolism, however a complete understanding of the regulation systems and mechanisms underlying the rearrangement of flux distribution and gene expression analysis is still under discussion. Recently, it has been suggested that the switch between fermentation and respiration is only dependent on the rate of glucose uptake (Huberts, *et al.*, 2012). This hypothesis is well in agreement with the work performed by Elbing *et al.* (Elbing, *et al.*, 2004) who suggested that it is the rate of glucose uptake to determine the flux through glycolysis and the onset of ethanol fermentation. Based on several studies, it was suggested that the switch to fermentative conditions is subjected to flux dependent regulation and that overflow metabolism does not play a major role in the distribution of fluxes between respiratory and fermentative metabolism (Huberts, *et al.*, 2012).

2.3 Glucose repression

The Crabtree effect is closely related to glucose repression. Glucose repression is a phenomenon present in *S. cerevisiae* that shares similarities to the bacterial carbon catabolite repression. *S. cerevisiae* can grow on a wide variety of carbon sources but glucose and fructose are, amongst all, the preferred ones. The term glucose-repression (Ronne, 1995, Carlson, 1998, Gancedo, 1998) indicates the condition when the presence of glucose triggers the repression of enzymatic activities necessary for the growth on other carbon sources (Santangelo, 2006). Extracellular glucose addition causes remarkable rearrangement in global gene expression levels, leading to a higher than 2-fold change (bidirectional) of approximately 40% of genes (Wang, *et al.*, 2004) and to a higher than 3 fold-change of only 20% of the genes.

The overall picture of this phenomenon is rather complex as different players and regulators take part in the glucose repression cascade, exerted through different signaling routes. Differently from the bacterial catabolite repression, the yeast glucose repression cascade is mediated by two different proteins: the protein kinase Snf1, encoded by the gene SNF1 (Sucrose Non-Fermenting 1), acting on Mig1 and Adr1, and through the PKA pathway via Hxk2 (Herrero, *et al.*, 1995, Ronne, 1995). Adr1 acts on ethanol and glycerol metabolism, β-oxidation of fatty acids and amino-acid metabolism while Mig1 is a transcriptional repressor controlling the metabolism of alternative carbon sources.

Under high glucose concentrations, Snf1 is inactive and the transcriptional repressor Mig1 remains in the nucleus (Ronne, 1995), repressing the transcription of several glucose-repressed genes. At low glucose concentration, Snf1 is active and can modulate the activity Mig1 which can translocate in the cytosol and relieve the glucose-repression cascade (Gancedo, 1998, Christensen, *et al.*, 2009). Mig1 is differentially phosphorylated in response to glucose and its activity is controlled by phosphorylation; it was shown that Snf1 is required for *in vivo* phosphorylation of Mig1 and that mutation on Mig1 in different recognition sites for Snf1 eliminated the differential phosphorylation of Mig1 to glucose (Treitel, *et al.*, 1998).

Snf1 is required for the transcription of genes involved not only in the metabolism of alternative carbon sources, but also involved in sporulation, glycogen storage, peroxisome biogenesis and on respiratory genes

through Hap4; Snf1 represses the pathway consuming ATP (e.g. lipid biosynthesis) and activates pathways generating ATP such as respiration, β-oxidation of fatty acids; additionally induces glycogen accumulation and modulates the general cellular stress response. Snf1 has a catalytic subunit and two regulatory subunits and it is activated by phosphorylation on a threonine residue (McCartney & Schmidt, 2001). Recently it has been shown that it is ADP the metabolite able to activate Snf1 in response to glucose limitation and protect its inactivation against dephosphorylation by Glc7, similarly to what observed in the mammalian cell AMPK (Mayer, *et al.*, 2011). Glucose is sensed on the cell membrane through Snf3 and Rgt2. Despite these two proteins are annotated among glucose transporters, as they share similarities to genes of *HXT* family, they cannot transport glucose. Snf3 and Rgt2 act through the repressor protein Rgt1, binding to the promoter of *HXT* genes and repressing their transcription.

Besides the Mig1 mediated pathway, glucose repression is exerted also via the mediator protein Hxk2. At least two regulatory pathways are regulated via Hxk2, one regulating Mig1 and the other acting via PKA pathway to promote cell growth and division. The involvement of Hxk2 in the PKA pathway was confirmed based on a study based on Hxk2 mutants (Westergaard, *et al.*, 2007); activation of this pathway regulates cellular processed such as cell division, growth control and nutrient availability. It was proposed by Ahuatzi *et al.* that Hxk2 has also a regulatory role in the glucose repression cascade as the authors showed that Hx2 interacts *in vivo* with both Mig1 and the catalytic subunit of Snf1 (Ahuatzi, *et al.*, 2007).

2.3.2 Glucose transport

As mentioned in the above paragraph, the mechanism of glucose repression is related to the transport of glucose inside the cell. *S. cerevisiae* can detect the amount of glucose present in the media and, according to it, activate different hexose transporters. In this regard, glucose transporters play an important role in glucose repression as they act on delivering glucose to the glycolytic pathway, thus contributing to determine the extent of fermentation and respiration (Ye, *et al.*, 1999).

In *S. cerevisiae* two uptake systems are present: one is a low-affinity and constitutive, the other is glucose-repressed with high affinity. The genes coding for glucose transporters are 20 and they are named *HXT1-17* (*HXT12* has 2 pseudogenes), *SNF3* and *RGT2* (Ozcan & Johnston, 1999). All these proteins share homologies in the transmembrane domains while the cytosolic part show remarkable differences (Boles & Hollenberg, 1997). Among all the hexose transporters, two are sensitively different from the others: Snf3p and Rgt2p are regulatory proteins involved in sensing glucose concentration but they do not have the capability of transporting glucose into the cell. The mechanism by which Snf3p and Rgt2p sense the presence of glucose has not been characterized in details but it is known that they act through the protein Rgt1. When glucose levels are high, the transcriptional activator Rgt1 becomes active and contribute to the induced expression of *HXT1* (Johnston, 1999). Hxt1 and Hxt3 are known to have low affinity and transport glucose into the cell when high concentrations of glucose are present, whilst Hxt2, -4, -6, -7 have high affinity for glucose. In response to low concentrations of glucose (0.1%), the transcription of *HXT2*-4 is induced, whereas *HXT1*

transcription remains low. Transcription of *HXT1* is instead induced by higher concentrations of glucose (4%) when *HXT2* and *HXT4* are repressed (Ozcan & Johnston, 1999). It was shown that the capacity of glucose transporters can influence the growth rate and the glycolytic flux (Diderich, *et al.*, 1999).
It was firstly in the yeast *Kluyveromyces lactis* that it was found that lower activity of glucose transporter relieved glucose repression (Weirich, *et al.*, 1997). This pattern has also been observed in *S. cerevisiae* where strains with null mutations in *HXT1* and *HXT7*, showed relieved glucose repression of invertase and maltase, as similarly observed for mutations in *HXT2*.

2.4 Levels of regulation of the cell

As previously mentioned, the expression of pathways of the yeast metabolism is tightly regulated in order to fuel cellular processes and maintain homeostasis; the fine-tuning of cellular metabolism is necessary in order to convert substrates into Gibbs free energy, redox power and precursor metabolites and ensure that no depletion of intracellular metabolites pools occur. The tuning of cell metabolism takes place through the activation or down-regulation of metabolic pathways at different levels in order to cope with the different requirements of the cell under different environmental conditions.
According to central dogma of molecular biology, the code expressed in the DNA (genes) is transcribed into mRNA (transcript), which is successively translated into a protein that gives rise to a certain phenotype. As a first approximation, it was assumed that increased expression of genes was reflected by an increased activity of the pathway encoded by these genes. With the development of studies based on quantifying mRNA levels and enzyme activities first, and flux analysis later, it became clear that this relationship does not always hold, and elevated mRNA levels do not necessarily translate in higher fluxes through a certain pathway or elevated enzyme activities. Translated proteins do most of the times require post-translation modification to become active, implying that there is no direct correlation between mRNA abundance, protein levels and fluxes through a reaction, as a consequence of different levels of cellular control of kinetic properties of the enzymes taking part in the metabolism reactions.
Several are the mechanisms that have evolved to ensure metabolic regulation of the cell. Metabolic regulation can occur at different levels: 1) Gene level, as the abundance of the mRNA coding for a specific gene is of course dependent on its transcription; the transcription of a certain gene can be induced, repressed or de-repressed in several fashions, however expression of a certain mRNA is also dependent on its stability (Santiago, *et al.*, 1987) and on the rate of degradation, an example is the respiratory gene *CYC1* (Yun & Sherman, 1995) 2) Enzyme levels, once the mRNA is translated into protein, this does not implies that the specific protein (enzyme) will be active. Activation of enzymes in certain pathway obeys to specific mechanism of regulation such as feedback inhibition and allosteric activation. Enzyme-level regulation is by far the most common in the yeast cells as it allows a fast response to environmental changes. The activity of a certain enzyme is also dependent on multiple factors among which cofactor availability. Examples of enzymes regulations are phosphofructo-1-kinase (Pf1k) and pyruvate kinase (Pyk1) which are know to be

allosterically regulated 3) Compartment-level, the localization of a certain protein in a compartment (mitochondria, peroxisomes etc.) is crucial for its activity and transport mechanisms within compartments play an essential role in the regulation of some metabolic pathways.
Ultimately, the understanding of the control of flux is one of the aims in metabolic engineering. In regards to that, Metabolic Control Analysis (MCA) can be considered an important discipline to address the contribution of metabolic parameters in determining a certain flux (Stephanopoulos, *et al.*, 1998). Through metabolic control analysis it is possible to address the control strength by the different enzymes in a pathway and quantify the response of fluxes and metabolite pools to changes in the kinetics of enzymes (Hatzimanikatis & Bailey, 1996).

The regulation of the central carbon metabolism of *S. cerevisiae* has always been a matter of outstanding interest however, despite the numerous studies addressed at establishing the possible levels of regulation, the manner in which the different pathways are regulated is still matter of debate. To unravel the complexity of the regulation of *S. cerevisiae*, the utilization of different fermentation-based technologies such as batch and chemostat cultivations have contributed to elucidate different regulatory mechanisms arising as a consequence of different metabolic conditions. During batch cultivations, glucose repression occurs and the concentrations of metabolites and products change during the time. In this respect, chemostat cultivations can be considered suitable tools for physiological studies since it is possible to fix a specific growth rate and establishing a steady-state by keeping the cell in the respiratory state. Under this condition, the physiology of the microorganism remains constant (Weusthuis, *et al.*, 1994) and it is possible to manipulate the specific parameters of interest in order to study long term responses (Piper, *et al.*, 2002).
The expression of certain patterns of genes is dependent on the carbon source and the composition of the medium (minimal vs complex). High transcript levels of glyoxylate cycle and gluconeogenic genes are, for instance, observed during growth on C2 compounds and *MAL* genes induced upon growth on maltose however, despite these intuitive examples, there is not a direct correlation between a certain conditions and the gene expression patterns. A few works describe the correlation between mRNAs and protein abundance, showing a linear correlation with a low correlation coefficient (Gygi, *et al.*, 1999, Pradet-Balade, *et al.*, 2001, Konig, *et al.*, 2011), however a generally better correlation has been identified for some metabolic enzymes (Olivares-Hernandez, *et al.*, 2010).

2.4.1 Regulation of the glycolytic pathway

Glycolysis is a cytosolic pathway necessary for the break down one molecule of glucose into two pyruvates, producing one molecules of ATP and two of NADH; additionally glycolysis has also a primary role in generating precursors for biomass (3-phosphoglycerate, 3PG, is precursor for serine amino-acids and phosphoenolpyruvate, PEP, for the aromatic family). The end product of glycolysis, pyruvate, represents an

important branchpoint in the central carbon metabolism of yeast. Through pyruvate decarboxylase (PDC), pyruvate can be converted into acetaldehyde and, successively to ethanol by alcohol dehydrogenase (ADH) with consumption of one NADH. Acetaldehyde can also be converted into acetate by the acetaldehyde dehydrogenase (AcDH) with reduction of one $NAD(P)^+$ to NAD(P)H. Cytosolic acetate can be converted into acetyl-CoA by the acetate synthase (ACS) and used as precursor for biosynthesis of secondary metabolites and in lysine biosynthesis. The overall stoichiometry for the EMP pathway can be found in equation 1.

$$Glc+2ADP+2P+2NAD^+ \rightarrow 2\ PYR + 2ATP + 2NADH + 2\ H_2O + 2H^+ \quad \textit{eq. 1}$$

At the beginning of the 90s it was believed that glycolytic gene were regulated at the transcriptional level, depending on the carbon source, (Moore, *et al.*, 1991) through the action of Rap1p, a transcriptional regulator of glycolytic genes, and through the protein Gcr1 (Tornow, *et al.*, 1993). However, despite several glycolytic genes are subjected to transcriptional regulation, it was found that the Pyruvate kinase gene *PYK1* and the Phosphofructokinase 2 gene *PFK2* were regulated at translational level (Moore, *et al.*, 1990). Several glycolytic enzymes are subjected to allosteric regulation; for instance, presence of ATP inhibits the activity of hexokinase (*HXK2*) that catalyze the first step of glycolysis, thus controlling the glycolytic flux (Larsson, *et al.*, 2000). It has also been observed that the presence of glycolytic metabolites affects the transcription of different glycolytic genes hypothesizing the role of these "messenger metabolites" in controlling the glycolytic flux. In 2008 it was shown that low concentrations of glucose-6-phosphate (G6P) and fructose 6-phosphate (F6P) stimulated the respiratory flux and that this effect was strongly antagonized by fructose 1,6-bisphosphate (F1-6bP) (Diaz-Ruiz, *et al.*, 2008), thus suggesting that F1-6bP inhibits the respiration rates in Crabtree positive yeasts. Through the construction of functional chimeras of the Hxt transporter, it was shown that, in these mutants, it is rate of glucose uptake that controls the glycolytic rate and, ultimately, ethanol production (Elbing, *et al.*, 2004). In this work the authors demonstrated that in wild-type *S. cerevisiae* strain at high glucose concentrations during batch cultivations, the glucose uptake capacity does not control the glycolytic flux. Differently, the strains with chimeric Hxt transporters showed a decrease in glycolytic rates resulting in a decrease of ethanol production.

Microarrays studies have contributed to the understanding of gene expression patterns during cultivation on different carbon sources. In a chemostat study on synthetic medium on different carbon sources (glucose, maltose, ethanol and acetate) only a few genes (180) were found to have significant changes on the different carbon sources, contrary to what found by DeRisi *et al* (DeRisi, et al., 1997). This behavior is likely to depend on the different regulation during growth on complex or minimal medium. Among the differentially expressed genes, 13 transcripts were up-regulated as response to glucose alone (Daran-Lapujade, *et al.*, 2004) and, among these, only four are involved in central carbon metabolism (*SUC2*, *SUC4*, *HXT4* and *GIP2*).

In a work from 2001, the authors sought to determine whether control over the metabolic network was occurring at transcriptional and translational level (hierarchical regulation) or through metabolic regulation

(ter Kuile & Westerhoff, 2001). By analyzing glycolytic enzymes through mathematical modeling, they concluded that regulation is almost never completely hierarchical.

Transcriptomic data were compared with flux distributions of central carbon metabolism based on a genome scale metabolic model, showing that, *in vivo*, there is little correlation between transcript levels, flux profiles and enzyme activity (Daran-Lapujade, *et al.*, 2004). Particularly, genes from the tricarboxylic acid cycle, glycolysis and pentose phosphate pathway did not find correlation, at least in magnitude, with changes in fluxes. The authors found that whereas the glycolytic flux was reduced by 3-fold, only *HXK1* and *TDH1* showed a decreased level in transcript, indicating that fluxes through glycolysis are controlled at post-transcriptional level.

Several other studies highlight the absence of correlation between mRNA and proteins, mRNA and fluxes and enzyme levels and fluxes. In a study from 2007, Daran-Lapujade *et al.* applied regulation analysis on the glycolytic pathway to demonstrate that glycolysis is regulated at post-transcriptional level (Daran-Lapujade, *et al.*, 2007). The authors measured enzyme activities, mRNA levels and metabolic fluxes showing that the role of regulation of mRNAs in controlling the glycolytic flux is marginal but are indeed translation efficiency and degradation that control the activity of the entire pathway.

2.4.2 Regulation of gluconeogenesis and glyoxylate cycle

In *S. cerevisiae*, growth on C2 carbon sources requires the activation of gluconeogenic pathway and the glyoxylate cycle, causing a substantial reprogramming of gene expression. Gluconeogenesis can be seen as a reversed glycolysis and its main function is to generate glucose-6-phosphate (G6P) from C2 and C3 compounds and glucogenic amino-acids. The glyoxylate cycle is a modification of the TCA cycle, used to convert C2 compounds into C4 dicarboxylic acids bypassing oxidative decarboxylation. In the presence of glucose above 1%, e.g. during batch cultivations, synthesis of the enzymes of this pathway is repressed while this bypass is active during fully respiratory conditions. The localization of this pathway is still a matter of debate and it seems to be able to operate both in the peroxisomes and in the cytosol.

The activation of gluconeogenesis and glyoxylate cycle is mediated by Cat8p, necessary for de-repression of gluconeogenic enzymes (Hedges, *et al.*, 1995). Gluconeogenic mRNAs are known to be repressed by very low glucose concentrations and induced after the diauxic shift or upon growth on non-fermentable carbon sources (Holzer, 1989, Turcotte, *et al.*, 2010). The expression of gluconeogenic genes was found to co-regulate with the expression of some respiratory genes as respiration is necessary during gluconeogenesis to gain energy through oxidative phosphorylation, explaining the reason why respiro-deficient mutant are not able to grow on non-fermentable carbon sources. Gluconeogenic genes are controlled by the transcriptional factor Rds2p (Soontorngun, *et al.*, 2007). Rds2p is activated through phosphorylation by Snf1 and has been found to induce transcription of the main gluconeogenic genes *PCK1* (phosphoenolpyruvate carboxykinase) and *FBP1* (fructose-1,6-bisphosphatase). *FBP1* is degraded when glucose starved cells are replaced with fresh glucose, showing a degradation pattern similar to that observed for the malate dehydrogenase gene *MDH2*

(Cytoplasmic malate dehydrogenase) involved both in glyoxylate cycle and gluconeogenesis. The genes *FBP1* and *MDH2* are controlled by Snf1p through an induction mechanism involving the regulators Cat8 and Sip4 (Schuller, 2003). Degradation of *FBP1* and *MDH2* occur through phosphorylation by hexokinase or glucokinase in presence of glucose (Hung, *et al.*, 2004). Beside transcriptional control, allosteric control of the main enzymatic activities of gluconeogenic pathway is also observed (Heinisch, *et al.*, 1996).

2.4.3 Regulation of tricarboxylic acid cycle (TCA cycle)

The TCA or Krebs cycle leads to the formation of NADH, CO_2 and $FADH_2$. For each mole of pyruvate oxidized, 4 mol of NADH, 1 of $FADH_2$, $3CO_2$ and 1 GTP are generated. The overall stoichiometry for a complete TCA cycle can be found in equation 2.

$$pyr + 3H_2O + GDP + 2P + 4NAD^+ + FAD \rightarrow 3CO_2 + GTP + 4NADH + FADH_2 + 4H^+ \quad eq.\ 2$$

The TCA cycle is known to be regulated at the transcriptional level and to be repressed by glucose (Liu & Butow, 1999). High glucose concentrations cause transcriptional repression of the genes from the TCA cycle and respiration and only mRNAs corresponding to *CIT2* (citrate synthase) and *IDP1* (NADP-specific isocitrate dehydrogenase) have been reported to be over-expressed at high glucose concentrations (Yin, *et al.*, 2003).

Studies based on ^{13}C labeling fluxes during batch cultivations have shown that, under these conditions, the TCA cycle mainly operates as two branches (Gombert, *et al.*, 2001) however, irrespective of the high glucose concentration, the flux through the TCA was found to increase with a decrease in glucose uptake rate. Using glucose repression negative mutant (*hxk2*, *grr1*) and respiration mutants (*hap2*) it was observed that the glucose uptake rate, and not the specific growth rate, inversely correlates with the activity of the TCA (Heyland, *et al.*, 2009). Some enzymes of the TCA (such as the citrate synthase *CIT1*) are transcriptionally regulated by the Hap2/3/4/5 transcriptional complex as a response to the presence of non-fermentable carbon sources (Liu & Butow, 1999). Despite the transcriptional, carbon source dependent regulation of several genes, the Krebs cycle is also highly regulated at enzymic level through substrate concentration and feedback inhibition. Enzymes such as the isocitrate dehydrogenase (*IDH*) are activated by AMP and inhibited by ATP (Hu, *et al.*, 2006), whereas low level of $NADH/NAD^+$ favor the activity of the main enzymes involved in the control of the cycle. The major regulatory sites in the TCA cycle are at the level of citrate synthase, isocitrate dehydrogenase and α-ketoglutarate dehydrogenase (Coleman & Bhattacharjee, 1975, Gadde & McCammon, 1997). The TCA cycle is used by the cell to generate precursors for biosynthesis of amino-acids (oxaloacetate, OAA, for aspartate family and α-ketoglutarate for glutamate family), additionally, the TCA cycle is also involved in catabolism of fatty acids for energy production through the metabolite acetyl-CoA (Strijbis & Distel, 2010).

2.4.4 Regulation of mitochondrial respiration

Mitochondrial respiration is the process that leads to the generation of energy (ATP) through a process called oxidative phosphorylation. Oxidative phosphorylation is the process most commonly used by eukaryotes to utilize oxygen under aerobic conditions, re-oxidizing reduced cofactors to H_2O (Rosenfeld & Beauvoit, 2003). The NADH generated in the mitochondrial matrix is oxidized by a NADH:ubiquinone oxidoreductase, also called internal NADH dehydrogenase, localized on the inner mitochondrial membrane (Bakker, *et al.*, 2001). Yeast mitochondria can oxidize cytosolic NADH directly through the external NADH dehydrogenase or cytosolic NADH can also be oxidized by the respiratory chain via glycerol-3-phosphate shuttle. These systems couple oxidation of cytosolic NADH to the respiratory chain. All the NADH oxidation pathways converge to the ubiquinone pool which donates (Zitomer, *et al.*, 1979) electrons to the cytochrome c through the bc1 complex; oxidation of cytochrome c by molecular oxygen is catalyzed by cytochrome c oxidase (de Vries & Marres, 1987). Additionally, the succinate dehydrogenase complex is directly coupled to the respiratory chain as the $FADH_2$ producing during oxidation of succinate to fumarate acts as an electron donor for ubiquinase (Cimini, *et al.*, 2009).

Respiratory genes are subjected to transcriptional regulation and, for some of them, expression is reduced in response to environmental conditions such as anaerobiosis and high glucose concentration (Marykwas & Fox, 1989) (e.g. *CYC1*, Cytochrome c isoform 1, is up-regulated 5 fold when growing on non-fermentable carbon source (Zitomer, *et al.*, 1979)). The activation of genes encoding components of the mitochondrial respiratory chain is transcriptionally regulated by the Hap2/3/4/5 complex (Fendt & Sauer, 2010). Most of the enzymes are also subjected to post-transcriptional regulation. It was observed that respiratory enzymes and mitochondrial formation are repressed indirectly by glucose through the presence of "high-energy substance obtained by the metabolism of glucose"(Polakis, *et al.*, 1965). This observation was expanded by Larsson *et al.* (Larsson, *et al.*, 1997), who reported increased glycolytic rates under energy excess conditions, thus clarifying the role of ATP in regulating glycolysis. During batch cultivations, the cell obtains the energy needed for growth through glycolysis, repressing the other pathways necessary for energy production such as the respiratory chain. Scheffler *et al.* (Scheffler, *et al.*, 1998) suggested that it is the mRNA turnover of mitochondrial proteins required for oxidative phosphorylation and respiration that controls repression of respiration and that genes such as succinate dehydrogenase *SDH2* are controlled post-transcriptionally.

2.4.5 Metabolism of non-fermentable carbon sources

Despite glucose being the preferred carbon source, C2 and C3 compounds such as glycerol, ethanol and acetate can be used by *S. cerevisiase* as alternative carbon sources. As described in the previous

paragraphs, in order for the cell to be able to metabolize these compounds it necessary the activation of gluconeogenic pathway and the glyoxylate cycle. The cell use different ways to transport C2 and C3 compounds; glycerol is transported into the cell through the symporter *STL1* and converted into glycerol-3-phosphate, which is successively converted into dihydroxyacetone-phosphate (DHAP) that can successively enter the glycolytic or gluconeogenic pathway. The genes responsible for the two conversions, the glycerol kinase *GUT1* and the glycerol-3-phosphate dehydrogenase *GUT2*, are induced during growth on glycerol and ethanol but repressed during growth on glucose. Differently, acetate and ethanol enter the cell through facilitated diffusion, even though acetate carriers have been identified (Casal, *et al.*, 1999). Ethanol is metabolized to acetaldehyde through the alcohol dehydrogenase encoded by the gene *ADH2* and to acetate by the aldehyde dehydrogenase (*ALD6*). Acetate instead is converted into acetyl-CoA by acetyl-CoA synthase *ACS1* (Turcotte, *et al.*, 2010).

2.5 The role of metabolic engineering and systems biology in physiological studies

What emerges from the previous chapters is the complex regulation of *S. cerevisiae* metabolism. Despite having been object of numerous investigations, not all the factors determining the onset regulatory phenomena and their exact molecular mechanisms have been elucidated. In the course of yeast physiological studies, different approaches have been undertaken to increase the understanding of of the functioning of the cell such as applying intuitive and direct genetic modifications on the system and study the response through, e.g. fermentation and calculation of physiological parameters or measurement of key enzymes levels. Such approaches have contributed to significantly increase the understanding of the yeast cell and if, on one hand, their strength lies in furnishing a concrete answer to a detailed question, they often fail to provide a holistic overview of the functioning of the living cell, limiting the answer to the *a priori* knowledge of the system under study. In the last two decades, the development of metabolic engineering and robust tools within systems biology accounted for more efficient approaches to understand the functioning of the cell. This allowed for an improvement of physiological studies, contributing to an increased understanding of recombinant phenotypes through high-throughput analysis.

At the beginning of the '80, it was believed that phosphofructokinase (Pfk) was the rate limiting step in glycolysis and, in order to manipulate the glycolytic flux, efforts have been focused on modulating the expression of the gene coding for this enzyme (Heinisch, 1986). Such approaches did not encounter large success, as well as most of other approaches aiming at modulating the glycolytic flux by increasing the level of expression of combinations of other enzymes of the glycolytic pathway (Schaaff, *et al.*, 1989). The limited success of these studies, is mainly due to the erroneous knowledge about regulation of metabolic fluxes as it was thought that the regulation of one pathway could have simply be controlled by one enzyme. Recombinant *S. cerevisia* strains with simultaneous overexpression of enzymes of the lower glycolytic pathway

were characterized under glucose-limiting, glucose excess conditions and their capacity to respond to an increased ATP demand through glucose pulse was also evaluated (Peter Smits, *et al.*, 2000). It was shown by the authors that the steady state cells respond to the glucose pulse by accelerating their specific carbon dioxide production rate two-fold faster than the control, suggesting and enhanced fermentative capacity of the recombinant strain.

An example about the contribution of metabolic engineering and systems biology to the understanding of yeast physiology can be found in the different attempts to modulate the shift from respiration to fermentation. The presence of different metabolic modes in *S. cerevisiae* is a factor of fundamental importance to consider when designing strategies to engineer the cell factories. Particularly, depending on the specific application, one might want to privilege respiratory metabolism rather than fermentation, or vice-versa.

To eliminate the onset of fermentative metabolism one could either block the pathway of oxidation of pyruvate to ethanol or stimulate the respiratory pathway. Several works were focused on modulating the expression of regulators controlling respiratory processes through, e.g. the overexpression of the gene *HAP4*; *HAP4* encodes for the activator subunit of the Hap respiratory complex, normally repressed in the presence of glucose and induced during growth on glucose under respiratory conditions or when *S. cerevisiae* is growing on non-fermentable carbon sources. The overexpression of *HAP4* resulted in the redirection of the carbon flux towards TCA and respiration. During batch conditions a decrease in ethanol formation and an increase in the specific growth rate were also observed (Blom, *et al.*, 2000). Nevertheless, overexpression of *HAP4* was shown to increase the biomass yield on glucose during nitrogen-limited chemostat cultures but not in batch cultures on glucose (van Maris, *et al.*, 2001) and, during aerobic glucose-limited chemostat cultures, overexpression of *HAP4* increased the specific growth rate at which aerobic fermentation sets on.

Another approach to modulate the onset of fermentation is to block the ethanol pathway by deleting the structural genes encoding pyruvate decarboxylase (PDC complex). Flikweert *et al.* (Flikweert, *et al.*, 1999) engineered a *Δpdc* mutant, showing that such a mutant strain is not able to grow on minimal medium with glucose as sole carbon source not even during fully respiratory conditions (low dilution rate chemostat) and addition of small amounts of C2 compounds such as ethanol or acetate is necessary. As the minimal amount of acetate required for growth experimentally measured was found to correlate with the amount of cytosolic acetyl-CoA theoretically required, it was suggested that the pyruvate decarboxylase reaction is essential in providing acetyl-CoA. Evolutionary engineering of this mutant during chemostat cultivations decreasing acetate concentrations followed by glucose selection, resulted in a *Δpdc* strain able to grow without addition of C2 compounds. The evolved strain was capable of growing on minimal medium during batch cultivations with glucose as sole carbon source, showing an increased pyruvate yield (van Maris, *et al.*, 2004) thus indicating the successful approach of combining metabolic engineering strategies with directed evolution. Transcriptome analysis together with analysis of upstream regions of genes showing up-regulation, indicated a partial relief of repression of genes with a possible Mig1 binding sites in their UAS, suggesting that a relief

of glucose repression can contribute to ethanol tolerance. Transcriptome analysis performed on nitrogen-limited chemostat cultures showed several differences in gene pattern expression, especially in the genes coding for hexose transportes, however the specific causes at the origin of the incapability of *Δpdc* strain to grow on glucose are not fully understood. The *Δpdc* study, albeit being almost a decade old, represents a good example of how the combination of different approaches (metabolic engineering, physiology, evolutionary engineering and transcriptome studies) resulting in an integrative study can fit the reverse metabolic engineering cycle, leading to a better understanding of engineered phenotypes and, most importantly, to a continuous improvement in the design process.

In the last decade, high-throughput techniques such as microarray analysis, metabolite profiling and flux analysis based on ^{13}C labeling have been widely applied in physiological studies. The application of these approaches contributed not only to improve the characterization of recombinant cellular phenotypes but also to the design of more efficient and non-intuitive strategies to engineer the cellular metabolism.

Systems biology–based approaches also contributed to the understanding of physiological phenomena and a good example can be found in the Snf1 regulatory network (Usaite, *et al.*, 2009). In this study, three different knock-out mutants with defect in the glucose repression cascade were compared based on mRNA levels, proteins and metabolites. Using computational tools the authors succeeded in identified the role exerted by Snf1p in regulating energy metabolism. In a previous work, Westergaard *et al.* (Westergaard, *et al.*, 2007) applied transcriptome analysis to explore the phenotypical differences observed among different mutants in the glucose repression cascade. Analysis of differential gene expression allowed the identification of profound effects on TCA cycle, respiration and ATP synthesis coupled to proton transports of the *HXT2* and *GRR1* deletion mutants, thus suggesting close interaction between the two pathways for glucose sensing and the one acting through Snf1.

The main challenge of the application of systems biology in identifying regulatory network and elucidate cellular function is linked to the limited capacity of integration of the large amount of data generated. Several efforts have been focused on integrating transcriptome with metabolome and fluxome data (Ideker, *et al.*, 2001, Phelps, *et al.*, 2002, Erasmus, *et al.*, 2003). An excellent solution to this challenge is represented by the work performed by Moxley *et al.* (Moxley, *et al.*, 2009), where the authors developed a model-base approach to correlate transcriptomic with metabolic flux data, developing a flux change predictor capable to predict metabolic response and flux rewiring of a *Δgcn4* mutant based on mRNA levels. Similar approaches are likely to increase in the future, providing more accurate physiological data to integrate in biologically representative models for description and prediction of cellular phenotypes.

3. Results

The objective of my research was to gain new knowledge on the central carbon metabolism of *Saccharomyces cerevisiae*. Particularly, I dedicated my research to the design of metabolic engineering strategies to manipulate the flux through the glycolytic pathway and to address the abilities of different high-throughput approach in describing cellular phenotypes, evaluating the contribution of different systems biology tools in metabolic engineering applications and physiological studies.

Glycolysis represents one of the core pathway of the central carbon metabolism and, as the rate through the glycolytic flux contributes to determine the shift towards fermentative metabolism, we sought to manipulate the glycolytic pathway through different strategies. The most direct approach to limit flux through glycolysis can be found in my first study (**paper I**; "Phosphoglycerate mutase knock-out mutant *Saccharomyces cerevisiae*: physiological investigation and transcriptome analysis". Papini *et al.* 2010) where I describe the construction of a mutant deleted in the lower part of glycolysis by deleting the gene *GPM1* coding for glycero-phosphomutase. The *Δgpm1* mutant shows a severely impaired growth and a remarkably increased stress response. The effects of the deletion on global gene expression levels are investigated through microarray analysis, uncovering the impact of this deletion on central carbon metabolism, the global stress response and carbon catabolite repression.

In my second study (**paper II**; "Physiological characterization of recombinant *Saccharomyces cerevisiae* expressing the *Aspergillus nidulans* phosphoketolase pathway: validation of activity through ^{13}C-based metabolic flux analysis". Papini *et al.* 2012) an alternative strategy to re-direct carbon fluxes in the central carbon metabolism by expressing in *S. cerevisiae* the *Aspergillus nidulans* phosphoketolase (PHK) pathway is presented. The introduction of the phosphoketolase (PHK) route has the potential to re-wire the carbon flow towards acetate/acetyl-CoA formation, which could theoretically be further directed towards respiratory processes via the glyoxylate cycle/TCA cycle. The strain expressing the heterologous route was physiologically characterized and the intracellular flux distributions were resolved, proving the capability of *S. cerevisiae* to utilize the PHK pathway, as it was suggested based on an increased acetate yield observed during batch cultivations. Nevertheless, in order to evaluate whether the PHK could provide an alternative route for glucose catabolism we introduced the PHK pathway in the *Δgpm1* strain. Unfortunately, despite different evolutionary approaches, selection on glucose of a *Δgpm1* strain expressing the PHK route was not achieved as the recombinant strain showed to be highly sensitive to the presence of glucose.

To further characterize the shift towards fermentative metabolism, a systems-level comparison between the Crabtree negative yeast *Scheffersomyces stipitis* and *S. cerevisiae* was performed (**paper III**; "*Scheffersomyces stipitis*: a comparative systems biology study with the Crabtree yeast *Saccharomyces cerevisiae*". Papini M, Nookaew I, Uhlén M, Nielsen J; Manuscript submitted). In *S. stipitis*, differently from *S. cerevisiae,* the onset of fermentation is determined by a decrease in oxygen levels. High-throughput analysis through mRNAs expression, metabolome profiling and flux distributions analysis, was applied in order to compare the different response

of *S. cerevisiae* and *S. stipitis* to glucose excess and glucose limitation conditions and to identify possible differences in the regulatory network.

After having used both microarray analysis and RNA-seq for physiological studies, an evaluation of the different methodologies for transcriptome analysis was performed (**paper IV**; "From reads to differential gene expression: a Comprehensive Comparisons of RNA-Seq based Transcriptome Analysis and cross comparison with microarrays". Nookaew I#, Papini M#, Pornputtapong N, Scalcinati G, Fagerberg L, Uhlén M, Nielsen J; manuscript submitted). In this study different computational methods for analysis of RNA-seq data were compared among each other and cross-compared with a microarray platform. Case study for this work was the transcriptional response of *S. cerevisiae* laboratory strain CEN.PK113-7D to respiro-fermentative (batch cultivations) and fully respiratory conditions (chemostat).

3.1 Phosphoglycerate mutase knock-out mutant *Saccharomyces cerevisiae*: physiological investigation and transcriptome analysis

Glycolysis is a particularly interesting pathway in the central carbon metabolism as it is one of the most well-conserved routes for glucose catabolism among different organisms. *S. cerevisiae* can rapidly switch its metabolism from fermentative to respiratory conditions by increasing its glycolytic flux and fermentative enzymes (van den Brink, *et al.*, 2008). This metabolic flexibility is a consequence of the evolution of this yeast in competitive environments in order to rapidly utilize the sugars available. Several efforts have been focused on studying the response of different glycolytic mutants, but the characterization of deletion mutants is usually challenged by their grow defects. Mutants deleted in glycolytic genes usually fail to grow on glucose, being moderately or strongly inhibited by its presence (Lam & Marmur, 1977). Traditionally, glycolytic mutants have been isolated by screening for a phenotype showing no growth on glucose or inhibition in media containing glucose (Clifton, *et al.*, 1978). Although several studies focused on the deletion of genes in the upper part of glycolysis, in this study, we describe the physiological and metabolic response to the deletion in the lower part of glycolysis by deleting the gene *GPM1*. The gene *GPM1* encodes for the glycero-phosphomutase, responsible for the inter-conversion of 3-phospho-glycerate into 2-phosphoglycerate. It is known that this gene is constitutively expressed during growth on glucose and its promoter shares similarities to the promoters of other glycolytic genes such as *PYK1*(pyruvate kinase), *ENO1* (enolase) and *PDC1*(pyruvate decarboxylase) (Rodicio, et al., 1993*)*. The expression of *GPM1* is regulated by the glycolytic transcriptional modulator *GCR1*. During the sequencing of the yeast genome, 2 other copies of *GPM1* were identified and named *GPM2* and *GPM3* however, it was shown that expression of these genes in a mutant deleted in *gpm1* cannot complement the mutation on glucose nor affect the growth of the deletion mutant on other carbon sources (Heinisch, *et al.*, 1998). As this gene is necessary for both glycolysis and gluconeogenesis, a strain with non-functional copy of this gene can only grow on a mix of ethanol and glycerol (Ciriacy & Breitenbach, 1979). In agreement with previous results on solid medium, the *Δgpm1* mutant constructed in this study was

able to grow only on a mixture of ethanol 2,5% and glycerol 3%, being strongly inhibited by the addition of glucose, even in minimal amounts. This effect is due to carbon catabolite repression and inactivation exerted by the presence of glucose (Rodicio & Heinisch, 1987).

We analyzed the physiological pattern of the *Δgpm1* mutant during batch cultivation in minimal medium on glycerol and ethanol, comparing it to the reference strain CEN.PK 113-5D. The main physiological effect showed by the *Δgpm1* mutant, besides a longer lag phase, was a lower specific growth rate μmax (almost half of the reference strain). The calculated physiological parameters are reported in Table 3, showing that acetate yield on substrate (YsAc) is slightly increased in the *Δgpm1* mutant, while CO_2 ($YsCO_2$) yield on substrate was slightly decreased. The growth curves of both strains are reported in Fig. 3. Interestingly, as it can be observed in the figure, the *Δgpm1* mutant did not seem to show glycerol consumption.

Table 3. Physiological parameters of the reference strain CEN.PK 113-5D and the *Δgpm1* mutant under aerobic batch cultivations with 20gL^{-1} as carbon source

	Reference strain	***Δgpm1***
Ysx	0.442±0.02	0.459±0.02
$YsCO_2$	0.178±0.03	0.16±0.008
YsAc	0.025±0.01	0.067±0.06
YsSucc	0.005±0.002	0.001±0.001
μmax (hr^{-1})	0.172±0.001	0.097±0.009

As it was previously observed that glycerol is needed for growth, we used flux balance analysis (FBA) in combination with the genome-scale metabolic model ilN800 (Nookaew, *et al.*, 2008) to estimate the theoretical requirement of glycerol. In the theoretical evaluation, we simulated the cellular phenotype with two different approaches. In the first case we assumed that no glycerol was required and predicted the specific growth rate, which was lower to that experimentally measured. In the other case, we constrained the growth rate to that observed experimentally and predicted instead the specific glycerol uptake rate. Based on these values, we calculated the glycerol requirement for biomass formation to be extremely low (0.062 C-moles/C-moles for the reference strain and 0.051 for the *Δgpm1*), explaining why we did not observe significant glycerol consumption. The theoretical values are reported in Table 4.

In order to capture global changes as response to the deletion of *GPM1,* we used microarray analysis to compare gene expression levels of the reference strain and of the *Δgpm1* mutant. For this purpose the two strains were cultivated during aerobic batch cultures and the RNA extracted during mid exponential phase was hybridized onto array chips. The analysis of microarray data confirmed, as previously suggested, that the other isoenzymes of *GPM1*, *GPM2* and *GPM3* could not compensate the role of *GPM1*, as we did not find increased expression of *GPM2* nor *GPM3* compared to the reference strain.

Table 4. Theoretical evaluation of glycerol consumption

Strain	*Glycerol*	*Growth rate (h^{-1})*	*Glycerol Uptake rate (mmol/gDW/h)*	*Glycerol requirement for biomass formation (C-moles/ C-moles)*
CEN.PK 113-5D	+	0.172	1.298	0.062
Δgpm1	+	0.093	0.580	0.051

In Fig.4 the expression of genes involved in central carbon metabolism is reported, according to their p-values and fold-change. Analysis of microarray data uncovered the overexpression of the key genes of the gluconeogenic pathway (mainly *FBP1* and *TDH1*, but also *PCK1*), the pentose phosphate pathway genes *TKL1* and, to a lower extent, *GND2*. Genes of the glyoxylate cycle were also found to have increased expression in the *Δgpm1* mutant, compared to the reference strain. The minor isoform of succinate dehydrogenase gene *SDH1b* showed to be up-regulated in the mutant. This gene is responsible for the oxidation of succinate and for transport of electrons to ubiquinione and it was shown to be capable of restoring respiratory capacity of mutants with respiratory defect on non-fermentable substrate (Colby, *et al.*, 1998). Among the most differentially expressed genes, we found the gene *NQM1*, corresponding to an unspecific transaldolase which is induced upon diauxic shift, together with *NCA3*, a modulator of the

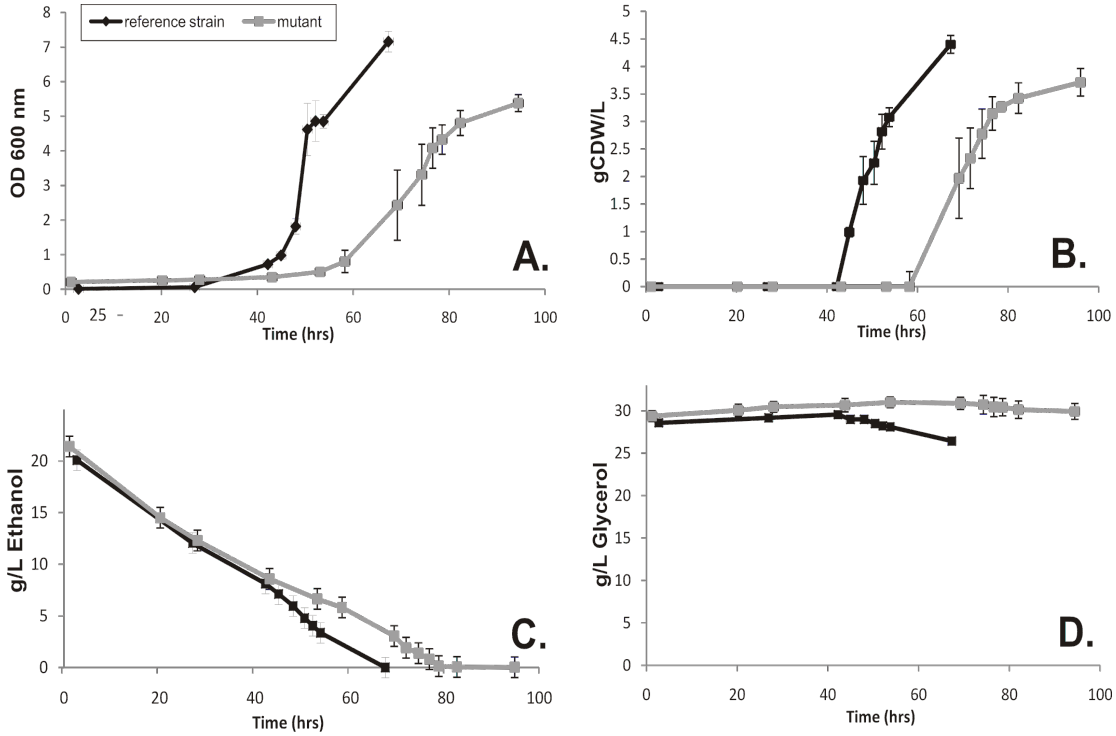

Fig. 3 Dry cell weight and ethanol consumption patterns of the reference strain CEN.PK 113-5D and the *Δgpm1* mutant. The strain is growing during aerobic cultivations in minimal medium with 3 % v/v glycerol and 2,5 % v/v ethanol as carbon sources.

mitochondrial expression of the Fo-F1 ATP synthase to have increased expression in the *Δgpm1* mutant. The overexpression of the transaldolase might be in agreement with previous results (Breitenbachschmitt, *et al.*, 1984) indicating that gluconeogenesis can be partially fed by the pentose phosphate pathway.

Interestingly, genes involved in the metabolism of storage carbohydrates such as trehalose and glycogen were also found to have increased expression in the *Δgpm1* mutant. This result could be interpreted as a consequence of the high stress conditions of the mutant due to energy limitation and might be linked to the overexpression of several genes involved in fatty acid β-oxidation, particularly Fox2, Pox1 and Pot1 as both pathways are activated during energy limitation conditions. Besides looking directly at the genes of the central carbon metabolism, we used the Reporter Feature Algorithm (Patil & Nielsen, 2005) to detect key regulatory nodes around which the transcriptional response is more significant (Oliveira, *et al.*, 2008). Based on the algorithm it was possible to identify significant Gene Ontology terms (GO) and Reporter Transcription Factors, thus pointing towards the pathways most affected by changes in gene expression.

The results found based on the Reporter Feature algorithm showed consistency with what was observed at single-gene level and are shown in Fig.5, where GO terms and reporter transcription factors are reported according to their p-values. GO terms involved in stress response, PPP pathway, TCA and gluconeogenesis were found to be over-represented in the*Δgpm1* mutant. The only terms found to have increased representation in the reference strain were those related to ribosome biogenisis and rRNA maturation, in agreement with its higher growth rate. Analysis of reporter Transcription Factors (TFs) showed the TFs around which significant changes in transcription occurs. Most of them were found to be involved in the yeast stress response (Arr1, Cin5, Crz1, Hot1, Hsf1, Msn2-4, Rpn4 and Yap1-5-6). TFs regulating amino acids metabolism (Leu3 and Met4) and nutrient limitations (Gcn4, Gis1) were also found to have increased significance in the *Δgpm1* mutant, together with Skn7 and Sko1. These TFs are involved in the osmotic stress response and are also known to interact with the Tup1-Cyc8 complex. This complex, also known as Ssn6-Tup1 (Treitel, *et al.*, 1998), is a repressor involved in regulating the expression of wide variety of genes family in response to glucose signal, oxygen presence and DNA damage, being recruited by Mig1p to the promoter of glucose repressed genes (Treitel & Carlson, 1995). Interestingly, in the *Δgpm1* mutant, the transcription factor Mig1 was found among the transcription factors with increased significance. Reporter TF involved in β-oxidation of fatty acids such as Oaf1 and Pip2 were also identified, in agreement with what was observed at single-gene level. The increased significance of TFs involved in β-oxidation pathway in the *Δgpm1* mutant relates to the over-representation of Mig1, confirming the previously established role of Mig1 in regulating energy generating pathways. Genes under the control of Hap4 were found to have increased significance; Hap4 is the regulatory subunit of the Hap respiratory complex, induced upon shift towards respiratory conditions and coordinating the expression of several respiratory genes.

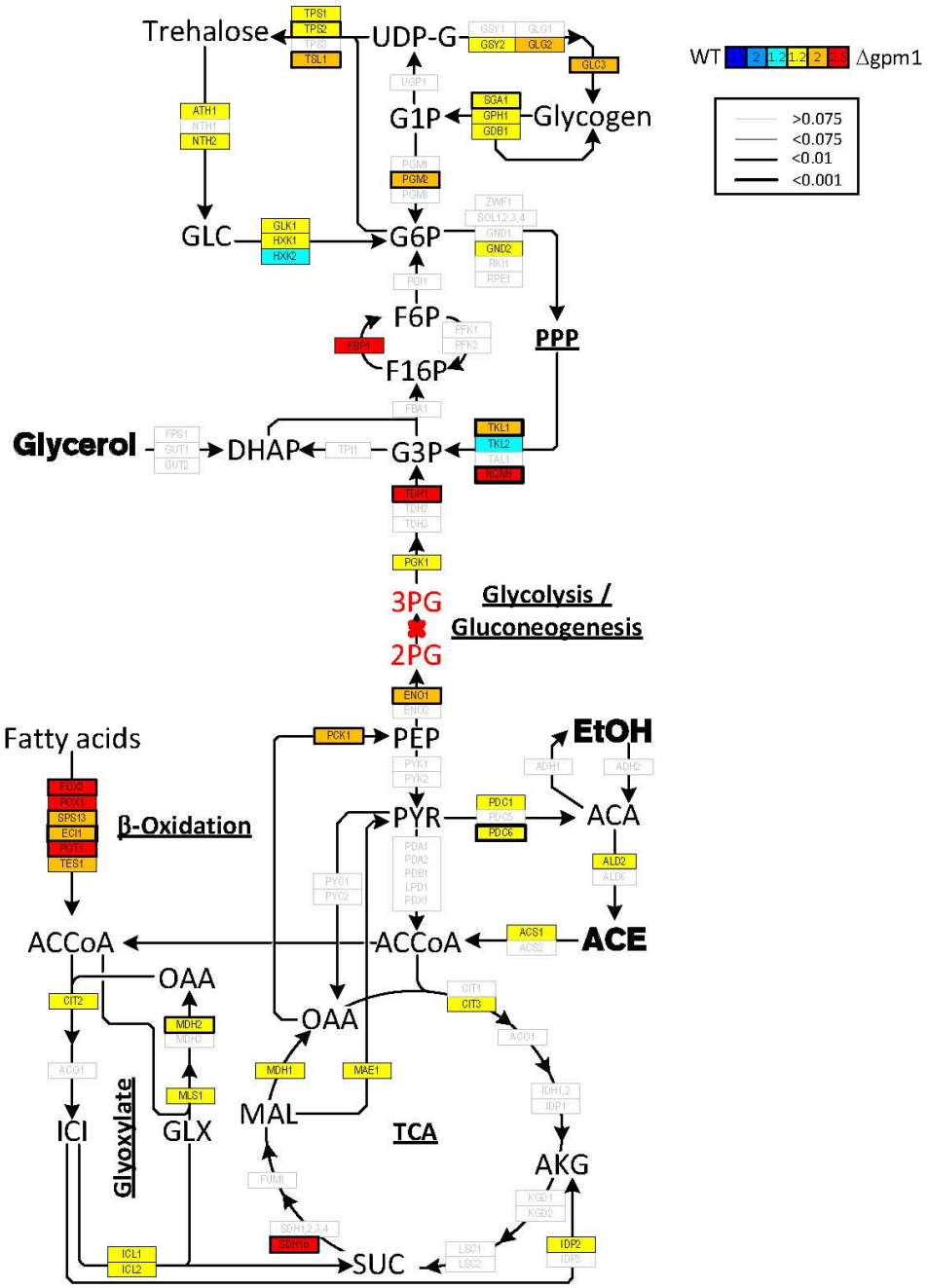

Fig. 4 Expression levels of genes of central carbon metabolism of the *Δgpm1* mutant and the reference strain CEN.PK113-5D. Color indicates fold-change while thickness of the line represents Q-values

This finding, together with the overexpression observed at single gene level of *NCA3* and *SDH1b*, might suggest that the *Δgpm1* mutant is attempting to increase its respiratory capacity. Furthermore the transcription factor Adr1 was also found to be over-represented in the *Δgpm1* mutant; Adr1 is a transcriptional activator of ethanol metabolism, fatty acids metabolism and meiosis, negatively regulated during growth on glucose and activated during aerobic oxidation of non-fermentable carbon sources, probably via Snf1 (Young, *et al.*, 2003, Tachibana, *et al.*, 2005).

On the contrary, in the reference strain, among the TFs regulating genes around which significant up-regulation occurs we found *SFP1*, involved in ribosome biogenesis and, to a significantly lower extent, we identified *STB5*, involved in oxidative stress.

The integration of the results based on different analysis of microarrays data showed the increased representation in the *Δgpm1* mutant of terms involved in the gluconeogenic pathway and respiratory processes. These findings, together with the increased significance of Mig1, suggest that the *Δgpm1* mutant is less catabolite-repressed than the reference strain and that it is attempting to increase its respiratory capacity to compensate for the energy imbalance caused by the interruption in the glycolytic flux. In agreement with this hypothesis is the finding, confirmed by the integrated analysis, about the up-regulation of genes involved in trehalose metabolism and fatty acids oxidation. This result can be explained based on the energy imbalance and the increased stress conditions of the *Δgpm1* mutant as trehalose is accumulated in yeast as result of stress response (Mager & Ferreira, 1993). Additionally, it was previously shown that growth on non-fermentable carbon source requires gluconeogenic production of glucose-6-phosphate, necessary for the biosynthesis of trehalose (Thomas, *et al.*, 1991). As the mutant cannot have a net ATP gain from the glycolytic flux, it is very likely attempting to compensate its energy imbalance by modulating the metabolism of storage carbohydrate and increasing its respiratory capacity.

The integrated study on the *Δgpm1* mutant addressed several aspects of the role of *GPM1*. Similarly to others glycolytic mutants, the *Δgpm1* showed a severe growth defect and increased stressed levels both through physiological data (longer lag phase, retarded growth) and through microarray data. The overexpression of gluconeogenic genes and, to a less extent, some of the genes from the glyoxylate cycle suggest that in this mutant ethanol is respired whereas glycerol, despite being required in minimal amounts, is needed to fuel gluconeogenesis. In regards to this, the simulations performed based on the genome-scale metabolic model contributed to address the amount of glycerol needed for growth. Through the analysis of reporter transcription factor analysis it was possible to establish the up-regulation of TFs involved in carbon catabolite repression such as Hap4 and Mig1. These observations point to the confirmed role of microarray to identify hidden spot of metabolism where significant transcriptional changes occur. Interesting is the finding that the *Δgpm1* mutant showed increased level of expression of the gene *NCA3*, a gene encoding for one of the subunits of the Fo-F1 ATPase complex. Together with the over-representation of Hap4 and the overexpression of the minor subunit of the succinate dehydrogenase complex *SDH1b*, this is in agreement with the previous hypothesis that the mutant is trying to increase its respiratory capacity in order to compensate for the energy imbalance caused by the block in the glycolytic flux. This hypothesis correlates well with the increased oxidation of fatty acids observed and changes in the metabolism of reserve carbohydrates, suggesting that the *Δgpm1* mutant in trying to cope with the energy imbalance by modulating pathways leading to energy generation. The results presented, demonstrate that changes in fluxes through the lower glycolytic flux have a profound effect on to the entire re-programming of the cell metabolism through a substantial re-arrangement of gene expression patterns. It would be extremely interesting to determine the

flux distribution of this mutant in order to establish the contribution of gluconeogenic pathway and respiration in sustaining the growth, however, the inability to grow on glucose as sole carbon sources sets limitation to this investigation.

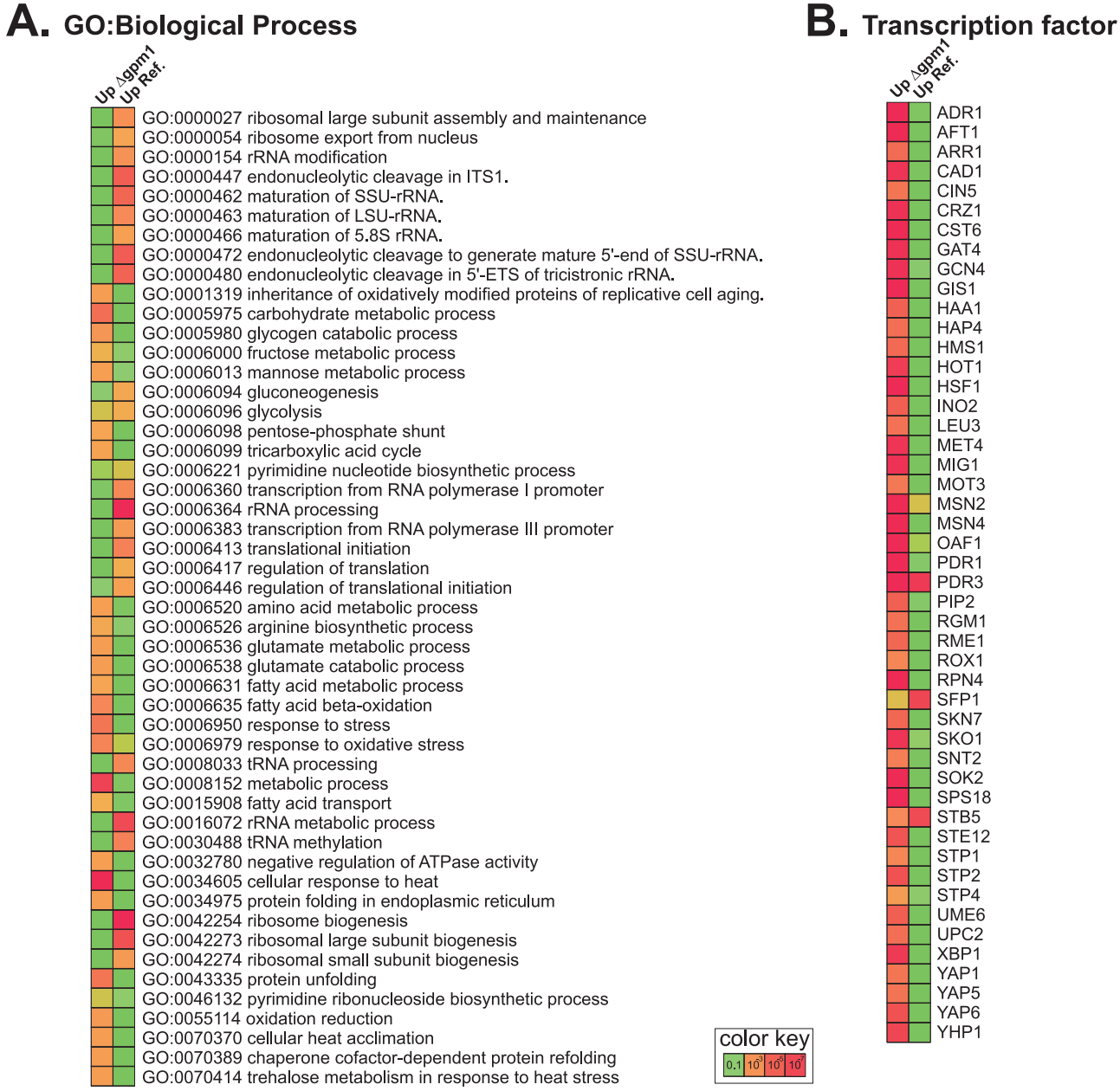

Fig. 5 GO terms (A) and Reporter Transcription Factors (B) of reference strain and *Δgpm1* mutant.

3.2 Expression in *Saccharomyces cerevisiae* of a recombinant route for glucose catabolism: the *Aspergillus nidulans* phosphoketolase pathway

The phosphoketolase pathway (PHK) is a route for carbon dissimilation used both by microorganisms of the Eukarya and Bacteria domains. Utilization of the phosphoketolase pathway provides an alternative route for carbon catabolism to the Embden-Meyerhof-Parnas pathway as it allows the shunt of carbon from the pentose phosphate pathway (PPP) intermediate xylulose-5-phosphate to acetate or acetyl-CoA. In the first step of this route, xylulose-5-phosphate is converted into acetyl-phosphate and glyceraldehyde-3-phosphate by an enzyme called xylulose-5-phosphate phosphoketolase, encoded by the gene *xpk* and found both in eukaryotes and prokaryotes.

The fate of acetyl-phosphate is different depending on the microorganism and the condition. In bacterial species such as *Bacillus subtilis*, homofermentative lactic acid bacteria and bifidobacteria, as well as in *Escherichia coli*, acetyl-phosphate is converted into acetyl-CoA through the action of the enzyme phosphotransacetylase, encoded by the gene *pta*. In filamentous fungi such as *Aspergiullus nidulans*, the second reaction of the route is catalyzed by the action of the enzyme acetate kinase, responsible for converting acetyl-phosphate into acetate, however acetate kinase is also found in bacteria such as *Chlamydomonas* and *Phytophtora* where they could act in the opposite direction to the *PTA* pathway under different growth conditions (S., 2006). In the fungal route, the use of the PHK pathway is energetically advantageous compared to the utilization of the PPP as 1 ATP can be gained in the second reaction of the pathway. However, several bacterial species such as *B. subtilis* possess both enzymes catalyzing the last step as they can convert acetyl-CoA to acetate via the acetyl-phosphate intermediate. In *B. subtilis*, the concomitant activity of these two enzymes is relevant for the production of acetate, one of the most important by-products of carbon metabolism (Presecan-Siedel, *et al.*, 1999) and the expression of these genes is regulated by carbon catabolite repression. Other bacteria can use this pathway in the opposite direction, for acetate assimilation or to generate ATP from excess of cytosolic acetyl-CoA. Xylose fermenting yeasts can also use this pathway when growing on pentose sugars such as xylose to generate acetate and ATP (Sonderegger, *et al.*, 2004). The different routes of the phosphoketolase pathway are represented in Fig. 6.

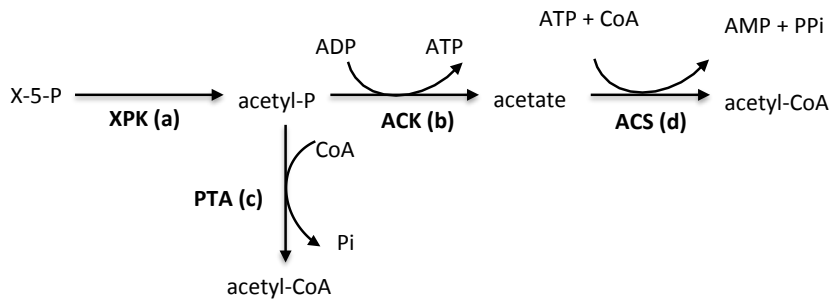

Fig.6 Different routes of phosphoketolase pathway. *Bacillus subtilis* and some lactic acid bacteria utilize reactions (a) and (c). *Aspergillus nidulans* during growth on xylose utilize reaction (a) and (b). XPK= phosphoketolase, ACK=acetate kinase, PTA=phosphotransacetylase, ACS= acetyl-CoA synthase

It was shown in *A. nidulans* that the overexpression of the endogenous *xpkA* gene increased the growth rate and the biomass yield during growth on xylose and ethanol, however, no significant effects were reported during growth on glucose (Panagiotou, *et al.*, 2008), indicating that the PHK route offers the possibility to generate cytosolic acetyl-CoA from the PPP during growth on xylose but there is no requirement for an efficient PHK during growth on glucose.

In this study, we sought to evaluate whether the phosphoketolase route could be used by *S. cerevisiae* for glucose catabolism and whether this could confer, as reported for filamentous fungi, any growth advantage. To this end, the PHK pathway from *A. nidulans* was expressed in *S. cerevisiae*. A final motivator for this study was to evaluate if the PHK pathway could be used as an alternative route to the EMP pathway for glucose catabolism. To accomplish and evaluate this strategy, we overexpressed in *S. cerevisiae* the genes from the *A. nidulans* PHK pathway (*xpkA* and *ack*). To force the flux towards the pentose phosphate pathway we sought to engineer the redox metabolism of *S. cerevisiae* by expressing the transhydrogenase gene *sth* from *Azotobacter vinelandii*. Non-membrane transhydrogenases are a class of cytosolic enzymes independent of proton translocation, present in several bacteria such as *E. coli*, *Pseudomonas* spp. and A. *vinelandii*, interconverting NADH into NADPH and vice-versa. In our work, we cloned the gene from *A. vinelandii* as this was previously expressed in *S. cerevisiae* (Nissen, *et al.*, 2001) showing that this enzyme is active and generating surplus NADH from NADPH, leading to increased oxoglutarate production. Additionally, to sustain growth and ensure the functionality of anabolic pathways in case the glycolytic flux was reduced, we over-expressed the endogenous PEP-carboxykinase gene from *S. cerevisiae* (*PCK1*), having a major anaplerotic role in *S. cerevisiae* metabolism (Zelle, *et al.*, 2010), in order to ensure availability of phosphoenolpyruvate (PEP) for biosynthesis. The overall strategy is represented in Fig. 7.

Among the different options and methods available to realize this strategy, we decided to overexpress the four genes by cloning on one multicopy vector under the control of two constitutive promoters (pTEF1 and pPGK1). The vector constructed, represented in Fig.8 and named pMPa, was introduced in CEN.PK 113-5D, generating the recombinant strain MP003.

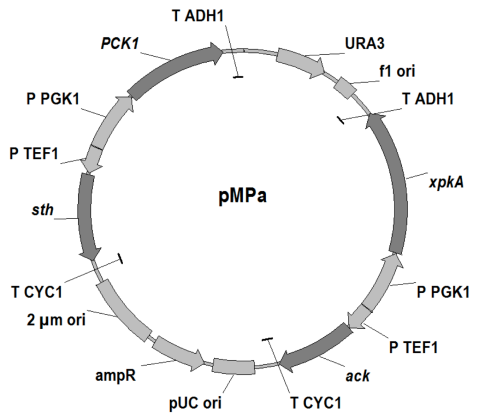

Fig. 8 The vector pMPa, carrying the genes *xpkA*, *sth*, *ack* and *PCK1*. PGK1 and TEF1 are the promoters used with the terminators ADH1 and CYC1

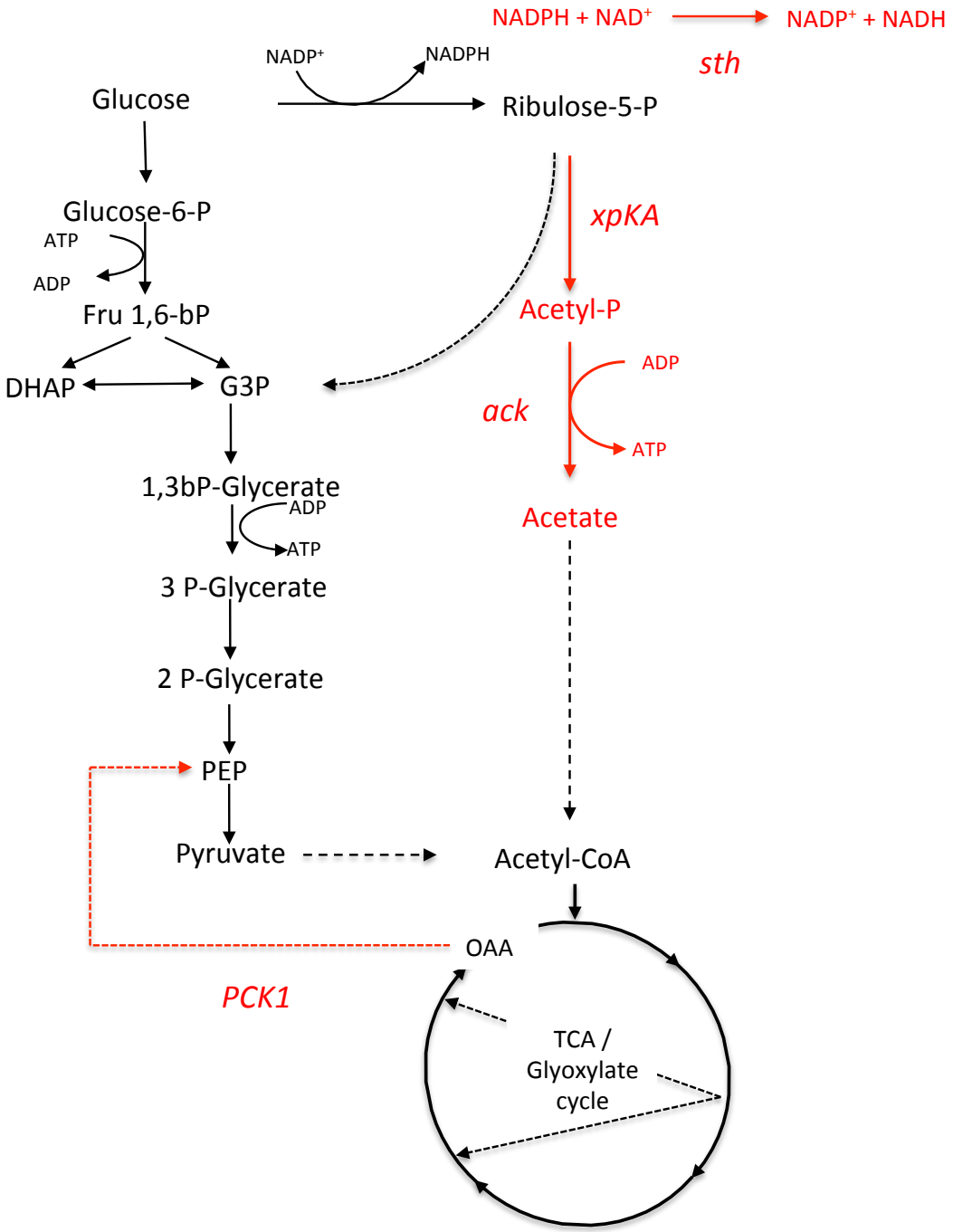

Fig.7 Recombinant PHK route expressed in *S. cerevisiae*. In red recombinant activities are shown. xpKa = xylulose 5-phosphate phosphoketolase (*A. nidulans*); *ack*= acetate kinase (*A. nidulans*); *sth*= transhydrogenase (*A. vinelandii*); *PCK1*= PEP carboxykinase (*S. cerevisiae*). The transformation of these activities into the strain CEN.PK113-5D generated the strain MP003.

The metabolism of the strain MP003, expressing the PHK pathway, was characterized during aerobic batch fermentation with 20 gL^{-1} glucose as carbon source and the calculated physiological parameters are shown in Table 5. The strain MP003 showed, compared with the wild-type, a lower specific maximum growth rate and a slower glucose uptake rate. Interestingly, the yield of acetate on substrate was found to be slightly

higher, whereas no significant difference is observed for the ethanol yield. Overall, the physiological analysis of the strain MP003 did not show remarkable differences from the wild-type and the main effects observed were slower growth, probably due to the expression of the recombinant pathway, and an increased acetate yield.

Table 5. Physiological parameters of the wild-type and the recombinant strain MP003. Yields and rates are calculated on the exponential phase of batch aerobic cultivation on minimal medium having 20gL^{-1} glucose as carbon source

Strain	μmax [h^{-1}]	Ysx [g /g Glc]	r Glc [g/ g DCW/ h]	YsEtOH [C-mol/ C-mol Glc]	YsAc [C-mol/ C-mol Glc]
WT	**0.410**	**0.17**	**2.72**	**0.45**	**0.016**
SDev	0.002	0.013	0.065	0.005	0.006
MP003	**0.330**	**0.15**	**2.60**	**0.46**	**0.019**
SDev	0.019	0.008	0.090	0.006	0.001

The moderate increase found in the acetate yield might suggest an increased flow of carbon towards acetate formation as a consequence of the expression of the PHK route, thus indicating that the recombinant *S. cerevisiae* strain MP003 is using the PHK route. In order to validate the activity of the PHK route in the recombinant strain MP003 and analyze the effects of the expression of the PHK pathway on the global metabolic network, we resolved intracellular flux distribution based on ^{13}C labeling experiments. Quantification of metabolic fluxes is the most direct measurement of the activity of the cell metabolism and its application can provide significant information about the metabolic state of the cell, especially in metabolic engineering studies. Through ^{13}C flux labeling it is possible to determine how metabolic fluxes are distributed in the living cell and how the different parts of the metabolism interact. Upon feeding with ^{13}C labeled glucose and analysis through GC-MS it is possible to determine the specific enrichment patterns of proteinogenic amino-acids; by combining the spectra information with carbon transition of the biochemical reactions and a mathematical model describing the cellular metabolism, it is possible, through metabolic flux analysis (MFA), to determine and quantify the different metabolic fluxes present in the cell. We based our calculations on a model previously established (can be found in the Appendix I) by including the heterologous PHK route as a reaction going from xylulose-5-phosphate to acetate. It is then possible, as described in Thykaer and Nielsen (Thykaer & Nielsen, 2007), to calculate the fraction of carbon going through the PHK pathway based on the fractional enrichment pattern of valine.

The fluxes of the recombinant strain MP003 and of the wild-type were resolved during chemostat cultivations at dilution rate 0.1 h^{-1} and are shown in Fig.8 and the values from measured and calculated SFL (summed fractional labeling) are reported in the Appendix II. Under chemostat cultivations, the only difference reported by the two strains is found in the biomass yield on substrate Ysx (0.41 g/g Glc for MP003 and 0.46 g/g Glc for the wild-type strain). As it can be observed in this figure, the phosphoketolase pathway

was found to be active in the recombinant strain, albeit a low value, proving the ability of the strain MP003 to channel the carbon through this heterologous route. Besides proving the activity of the PHK pathway in *S. cerevisiae* during growth on glucose, flux distribution analysis was used to address the effects of the recombinant expression on the central carbon metabolism. Interestingly, the glycolytic flux does not show differences compared to the wild-type, despite the flux from G6P to F6P is found to be increased to compensate for a lower fraction of carbon going to the PPP. These findings suggest that the catabolism of carbon through the PHK route does not interfere with the glycolytic flux, thus indicating that the activity of the PHK route is not sufficient to replace glucose catabolism through the glycolytic pathway. The finding of a lower flux of carbon going to the PPP was unexpected as it was expected that the over-expression of the transhydrogenase would have encouraged the carbon flux towards the PPP. The lower value might be due to several reasons, such as a possible cofactor imbalance arising as a result of the overexpression of the transhydrogenase in combination with the expression of the PHK pathway or, on the contrary, it might indicate the preference of *S. cerevisiae* to metabolize glucose exclusively via glycolysis, thus attempting to decrease the flux through towards the PPP to lower the carbon flux through the PHK route. Nevertheless the lower flux towards the PPP might reflect, to a certain extent, the lower biomass yield on substrate of the strain MP003.

Remarkable differences between the two strains are found at the pyruvate branchpoint: in the strain MP003, a lower fraction of the carbon is channeled towards acetate formation via pyruvate decarboxylase (PDC) and the acetaldehyde dehydrogenase (AcDH). Interestingly, in the strain MP003, the flux through acetyl-CoA synthase (ACS) is also found to be slightly lower. The lower flux through the PDC is balanced by a substantial increase in the pyruvate transport into the mitochondrial and by an increase through the pyruvate dehydrogenase (PDH) flux, converting mitochondrial pyruvate to acetyl-CoA. In agreement with these results we observed an increased in MP003 flux through the citrate synthase reaction and in fluxes through the TCA cycle. The increase fluxes through the TCA cycle might indicate an increased respiratory capacity of the recombinant strain, however, no physiological difference based on O_2 consumption or CO_2 production were observed. The different flux distribution at the level of PDC and PDH indicates that the strain MP003 mainly supplies the mitochondrial acetyl-CoA pool via PDH and that the acetyl-CoA is used to fuel the TCA, showing higher values than the wild-type. The two main anaplerotic fluxes, malate dehydrogenase (MAE) and the decarboxylation of pyruvate to oxaloacetate (PYC) showed lower value in the recombinant strain MP003. The finding, in the recombinant strain MP003, of a (slightly) lower flux through ACS might indicate that the acetate formed via the PHK pathway cannot be converted into acetyl-CoA and used to fuel the TCA cycle. Based on ^{13}C labeling it is not possible to resolve the flux through the PCK1 reaction and we thus have no indication about increased activity of this enzyme.

In this study we expressed in *S. cerevisiae* the fungal PHK pathway together with the *A. vinelandii* transhydrogenase and the endogenous anaplerotic enzyme PCK1. Flux distribution analysis through ^{13}C-labeling demonstrated the activity of flux through the phosphoketolase pathway during chemostat

cultivations. Despite being active, the phosphoketolase pathway showed a low value, probably as a consequence of the lower carbon flux towards the PPP, indicating that the expression of the transhydrogenase gene did not encourage the flux towards the PPP. Additionally, our results indicate that, despite *S. cerevisiae* can utilize the PHK pathway, the glycolytic flux is a preferred route for carbon dissimilation.

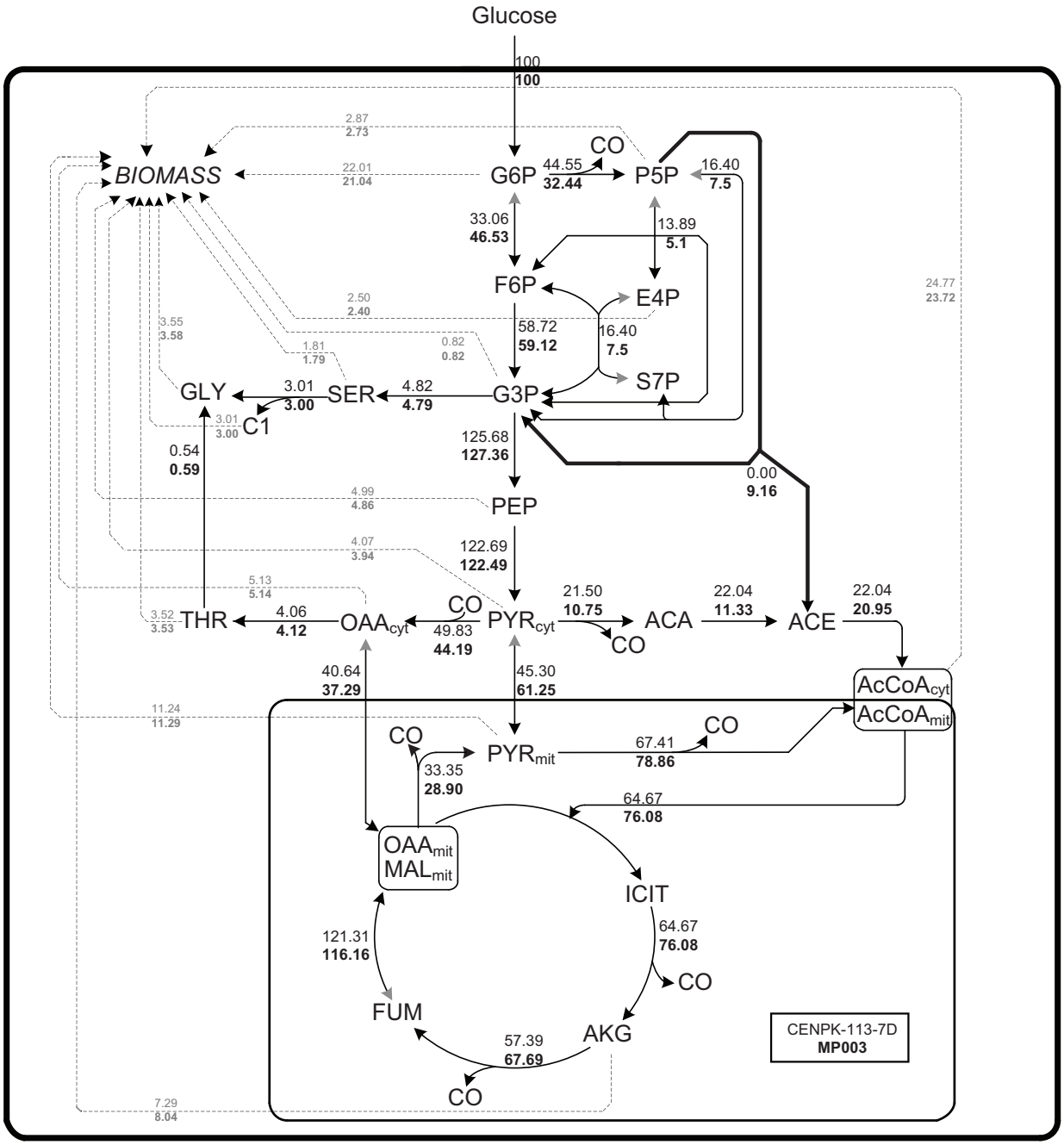

Fig. 8 Intracellular flux distribution based on 13C labeling during chemostat cultivations at dilution rate 0.1h^{-1} . Flux showed are net fluxes calculated based on the model that can be found in the Appendix I.

No remarkable physiological effects were observed on overflow metabolism towards ethanol formation in the recombinant strain, probably reflecting the consistency found in the glycolytic flux, however, fluxes during chemostat cultivations showed a slightly increased activity of the TCA cycle in the recombinant strain.

The expression of the PHK pathway in *S. cerevisiae* has remarkable consequences on the reprogramming of the yeast metabolism, particularly at the level of the pyruvate branchpoint. The recombinant strain expressing the PHK showed preference for the utilization of PDH reaction to generate mitochondrial acetyl-CoA and showed increased activity of TCA enzymes. On the contrary, anaplerotic reactions showed lower values compared to the wild-type. Contrary to what thought when designing the strategy, the acetate formed via the PHK route could not be converted, at least during chemostat cultivations, into acetyl-CoA.

Our results indicate that the PHK route, despite being used by *S. cerevisiae*, does not interfere with glucose dissimilation through the glycolytic pathway. However, expression of this route led to a moderate increase in acetate production during batch cultivations. In this respect, this strategy holds relevance for the future metabolic engineering strategies having acetate as a building block for their biosynthesis (biofuels, isoprenoids, polyketides etc.).

It would be interesting to test the expression of the PHK pathway as stand-alone route for glucose catabolism; this was the main idea behind the work that will be presented in paragraph 3.2.2.

3.2.1 Expression in *S. cerevisiae* of a hybrid phosphoketolase route

Additionally to the complete *A. nidulans* phosphoketolase route, the *B. subtilis* gene *pta* encoding for the phosphotransacetylase reaction, catalyzing the second step of the PHK pathway, was also evaluated. This gene was cloned, together with the other genes (*xpK* from *A. nidulans*, *PCK1* and *sth* from *A. vinelandii*) on the same vector described in the previous chapter, thus replacing the *ack* gene. The vector was named pMPb. This vector, was transformed into the strain CEN.PK 113-5D generating the recombinant strain MP004. The expression of this pathway in *S. cerevisiae* should lead to increased acetyl-CoA formation as the *pta* gene catalyzes the reaction converting acetyl-phosphate into acetyl-CoA with the release of a phosphate group.

The physiological parameters of this strain were evaluated during aerobic batch cultivations on 20 g L^{-1} glucose as carbon source. The strain MP004 showed a specific growth rate (μmax) of 0,33 h^{-1} and a biomass yield on substrate (Ysx) of 0,15 C-moles/ C-moles, similarly to what observed for MP003. However, compared to MP003, the strain MP004 reported a lower yield of acetate (YsAc) 0,017 in C-moles/ C-moles. The activity of the hybrid phosphoketolase route expressed by the strain MP004 was addressed through intracellular flux analysis, similarly to what done for the recombinant strain MP003. Interestingly, no flux through the PHK reaction was observed, indicating the incapability of *S. cerevisiae* to utilize this route. This might be due to different thermodynamic drives behind the bacterial reaction or to inactivation or incomplete folding of the bacterial enzyme.

3.2.2 Expression of phosphoketolase pathway in the *Δgpm1* mutant

Intracellular flux analysis of the strain MP003 showed that the expression of the *A. nidulans* PHK route in *S. cerevisiae* resulted in the capability of the yeast to utilize the heterologous pathway, despite low values of fluxes were reported. Additionally, ^{13}C based flux analysis addressed the effects of the expression of the PHK pathway on the reprogramming of *S. cerevisiae* metabolism however, no effects on the glycolytic pathway were observed. In order to establish whether the expression of the PHK in *S. cerevisiae* could function as a stand-alone route for glucose catabolism and be used as alternative route to the glycolytic pathway, we introduced the PHK route in a strain with non-functional glycolysis. For this purpose, we used the *Δgpm1* mutant as a background for the expression of the *A. nidulans* PHK pathway, introducing the vector pMPa and generating the strain MP001. In Table 6, the strains constructed are summarized.

Table 6. Summary of the recombinant *S. cerevisiae* strains engineered in this thesis

Background	Vector	Recombinant strain
CEN.PK 113-5D	pMPa	MP003
CEN.PK 113-5D *Δgpm1*	pMPa	MP001
CEN.PK 113-5D	pMPb	MP004

In light of the growth defect on glucose of the *Δgpm1* mutant, the recombinant strain MP001 was selected on a mix of glycerol 3% and ethanol 2,5 % as carbon sources, as used for the *Δgpm1*. The growth curve of the recombinant strain MP001, compared to the reference *Δgpm1*, is shown in Fig. 9.

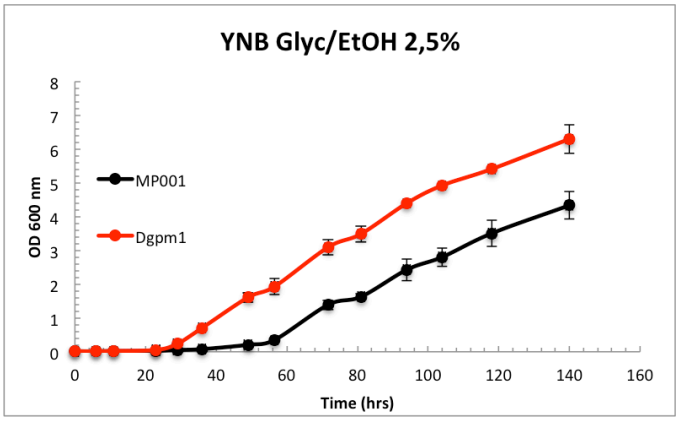

Fig. 9 Growth curves of the strain MP001 and the reference strain *Δgpm1*. Based on shake flasks cultivations on minimal medium Glycerol 3% and Ethanol 2,5%

The strains MP001 showed a longer lag phase compared to the reference *Δgpm1* strain, however the growth rates (μmax) of the two strains were similar: 0,35 h^{-1} for the *Δgpm1* mutant and 0,33 h^{-1} for the strain MP001. As our initial goal was to evaluate if the introduction of the PHK pathway could function as a stand-alone route for glucose catabolism, we tested whether the introduction of this pathway could restore the growth on glucose of the *Δgpm1* mutant. As previously described, the deleted strain showed to be strongly inhibited by the presence of glucose but we thought that installing the PHK route might have supported glucose catabolism. Unfortunately, MP001 did not show growth on glucose media alone (neither on complex nor synthetic medium) and a similar behavior was observed when glucose was added (0,1 %) to the mix of glycerol and ethanol during shake flask cultivation, probably as a consequence of the strong inhibition effect exerted by glucose. Addition of L-carnitine to the medium, known to mediate the acetyl-CoA shuttle into the mitochondria, did not help to restore growth on glucose. These results suggested that the expression of the PHK route could not rescue the growth defect of a strain deleted in the lower part of glycolysis, indicating that the PHK pathway cannot be utilized by *S. cerevisiae* as a stand-alone route for glucose catabolism.

3. 4. *Scheffersomyces stipitis*: a comparative systems biology study with the Crabtree positive yeast *Saccharomyces cerevisiae*

This paper reports the systems biology-based comparison between the Crabtree negative *Scheffersomyces stipitis* and *S. cerevisiae*. *S. stipitis* is a homotallic, mainly found in haploid form yeast which was firstly isolated in the gut of passalid beetles, commonly known with the name *Pichia stipitis*. As this yeast is Crabtree negative, the onset of ethanol fermentation is regulated by the availability of oxygen, specifically to a decrease in the oxygen tension, differently from *S. cerevisiae* where the glucose excess triggers the onset of ethanol fermentation. *S. stipitis* has greater respiratory capacity than *S. cerevisiae*, due to the presence of an alternative respiration system (STO), a SHAM sensitive terminal oxidase, donating electrons directly to O_2 from ubiquinone, branching out before the cytochrome C complex. *S. stipitis* is widely known for its capability of ferment pentose sugars such as xylose and arabinose with high rates and, in light of this feature, it has been exploited as a source of genes for the engineering of pentose assimilation pathway in other microorganisms. For this reason a considerable amount of research has been performed during growth on xylose, however a few works address the details of its glucose metabolism. In this work, the metabolism of *S. stipitis* was investigate during aerobic cultivations on glucose under batch and chemostat conditions and compared to *S. cerevisiae*. It has been extensively mentioned in the previous chapters that *S. cerevisiae*, as a consequence of the Crabtree effect, presents different metabolic states in response to glucose excess conditions, showing very distinct patterns between batch cultivation, when the strain shows respiro-fermentative metabolism, and chemostat cultivations, when *S. cerevisiae* shows fully respiratory metabolism. In this perspective, the comparison of yeasts with different metabolic behaviors can be regarded to as an interesting case-study to

address the contribution of high-throughput, indirect approaches in describing details of cellular phenotypes and capturing similarities and differences.

The metabolism of the two yeasts was extensively characterized using a global approach based on metabolome profiling, flux distribution, analysis of mRNA levels through RNA sequencing (RNA-seq) and physiological parameters. The results obtained were compared, showing how systems biology can contribute to increase the understanding of yeasts species less characterized than *S. cerevisiae* and to identify differences in metabolic networks of the two yeasts. To analyze gene expression levels, rather than using the microarrays platform we performed RNA-seq. RNA-seq is a powerful technique that revolutionized the field of transcriptomics, allowing for a higher dynamic range of expression and bringing qualitative and quantitative improvements to transcriptomic studies; additional technical information on this technique will be addressed in the next chapter.

The genome of *S. stipitis* was sequenced in 2007; it is bigger in size than *S. cerevisiae* (15.4 Mb vs 12.1 Mb) (Jeffries, *et al.*, 2007), however the number of predicted ORF is similar (5.841 vs 5.807) as *S. stipitis* have higher percentage of exons. Genome analysis revealed high homology with *Debaryomyces hansenii*, with whom *S. stipitis* shares 151 gene families not found in other genomes. Other yeasts evolutionary related to *S. stipitis* are *Schizosaccharomyces pombe* and *Candida intermedia*, with whom *S. stipitis* shares a number of genes for sugar transporters. Recently, the transcriptome of *S. stipitis* was sequenced in a study aiming at comparing differential expression of genes during growth on different carbon sources (glucose and on xylose). However the behavior of *S. stipitis* growing on glucose has not been addressed in details. In this paper, the results obtained by growing *S. stipitis* on glucose under aerobic batch and chemostat condition are presented.

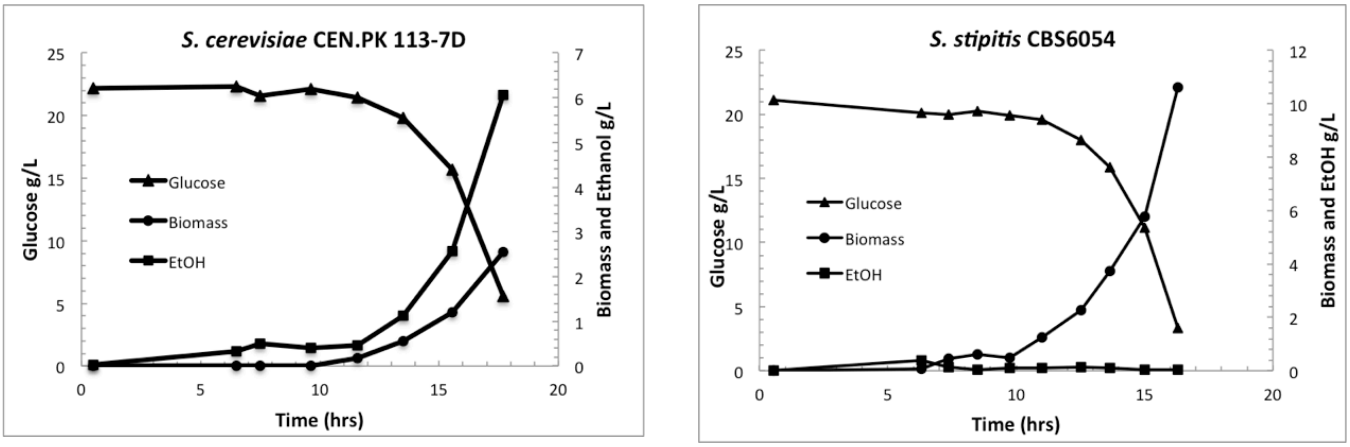

Fig.10 Growth curves of *S. stipitis* CBS6054 and *S. cerevisiae* CEN.PK113-7D. Aerobic batch cultivations with 20 gL-1 glucose as carbon source

The growth curve of *S. stipitis* CBS6054 is shown in Fig.10 and the calculated physiological parameters are reported in Table 7. As expected, *S. stipitis* did not show metabolites secretion during aerobic batch cultivations and the only products formed were biomass (Ysx 0,54 g/g glucose, compared to a for wild-type *S.*

cerevisiae strain CEN.PK 113-7D Ysx= 0,17 g/ g glucose) and CO_2. Similar yields were reported by *S. stipitis* under chemostat cultivations, while *S. cerevisiae*, as a result of purely respiratory metabolism, showed a higher biomass yield (Ysx = 0,51 g/ g glucose) during chemostat cultivations.

Based on physiological analysis, we suggested a constant respiratory mode of *S. stipitis* during aerobic cultivations, however, we wanted to address details of this behavior and elucidate to what extent the metabolic pattern observed for *S. stipitis* was similar to that shown by *S. cerevisiae* during fully respiratory conditions. Additionally, we sought to identify eventual differences in *S. stipitis* growing under the two cultivation conditions. To confirm what observed through physiological analysis, we resolved intracellular flux distributions of *S. stipitis* and *S. cerevisiae* growing under both conditions and compared the results. Fiaux *et al.* have previously characterized the intracellular flux network of *S. stipitis* (Fiaux, *et al.*, 2003) however, fractional labeling was detected through 2D-NMR and a different method and model for flux calculations was used, limiting the resolution of the metabolic network to the calculation of precursors abundance.

Table 7. Physiological parameters of *S. stipitis* and *S. cerevisiae*. Calculated on the exponential phase in aerobic batch cultivation with glucose 20gL^{-1} as carbon sources

	S. cerevisiae	*S. stipitis*
μmax [h^{-1}]	0.40	0.47
Glucose consumption rate [C-mmol /g DW/ h]	84.5	26.7
Ysx [g /g]	0.17	0.55
YsEtOH [g /g]	0.33	0.003
YsPyr [g /g]	0.003	0.004
YsAc [g /g]	0.01	0.001

Here, we estimated the distribution of fluxes under the two conditions based on the analysis of enrichment patterns of proteinogenic amino-acids, using the same model described for *S. cerevisiae*. The resolved intracellular network of the two yeasts is reported in Fig. 11 A and 11 B.

If we look at the figures cultivation-wise, it is possible to capture immediately that while in *S. cerevisiae* there is a substantial difference in the distribution of the metabolic network under the two conditions, this behavior is not observed for *S. stipitis* where a strong consistency in flux values between the two cultivations conditions is found, supporting the results observed during physiological characterization about the presence in *S. stipitis* of one main respiratory mode under both conditions, due to the absence of fermentative mode during glucose excess conditions. Although the flux distribution presented by *S. stipitis* showed similarities to the respiratory mode observed for *S. cerevisiae* during chemostat conditions, it was also possible to identify some differences. The oxidative part of the PPP, showed a higher flux in *S. stipitis*; this result might indicate an increased demand of NADPH necessary for biomass formation in agreement with the higher biomass yield of *S. stipitis*. Other differences are found at the pyruvate branchpoint: the flux through the pyruvate carboxylase reaction (PYC) which in *S. cerevisiae* shows low values during batch of *S. cerevisiae* and higher values during chemostat

cultivations, shows a constant value in *S. stipitis* which is higher than that observed for *S. cerevisiae* during batch but lower than that reported during oxidative growth. The flux through the pyruvate decarboxylase (PDC) reaction, leading to acetaldehyde formation from pyruvate, shows low values in *S. stipitis*, similar to what reported for *S. cerevisiae* during oxidative growth, when ethanol production does not occur. Additionally, the flux through the acetyl-CoA synthase (ACS) is slightly lower in *S. stipitis* than in *S. cerevisiae* under oxidative growth. The flux through the pyruvate dehydrogenase (PDH) reaction and the pyruvate import into the mitochondria show higher values in *S. stipitis* under both conditions than these found during fully respiratory conditions in *S. cerevisiae*.

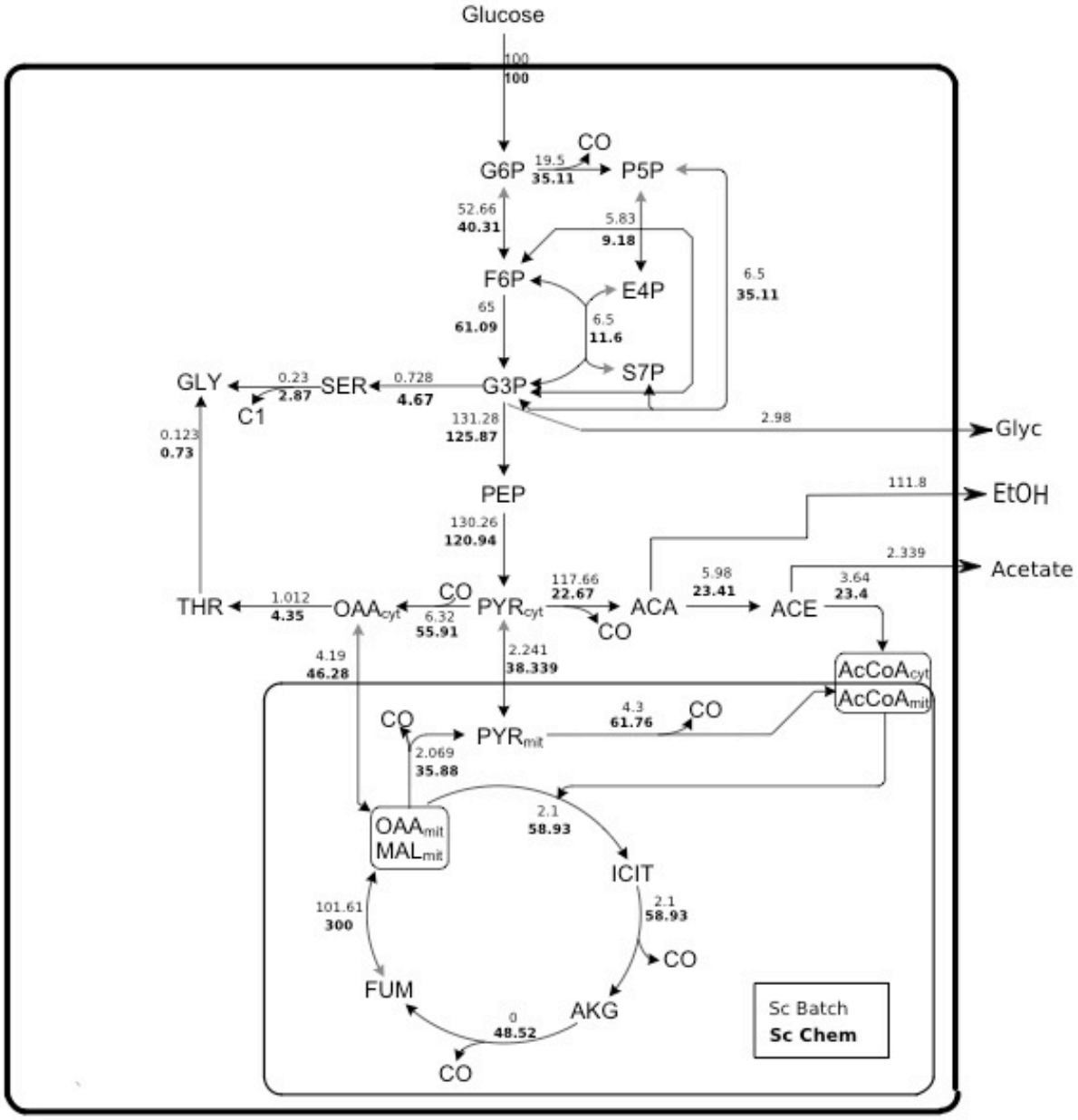

Fig. 11 A Intracellular flux distribution of *S. cerevisiae* under batch and chemostat cultivations based on ^{13}C-labeling

In accordance with the increased mitochondrial acetyl-CoA formation through the PDH found in *S. stipitis*, the fluxes through the TCA cycle also reported increased values compared to those reported for *S. cerevisiae* under respiratory conditions. These results might suggest that, in *S. stipitis*, acetyl-CoA used to fuel the TCA is mainly generated through the pyruvate dehydrogenase reaction and a remarkably lower fraction is produced via the PDC bypass.

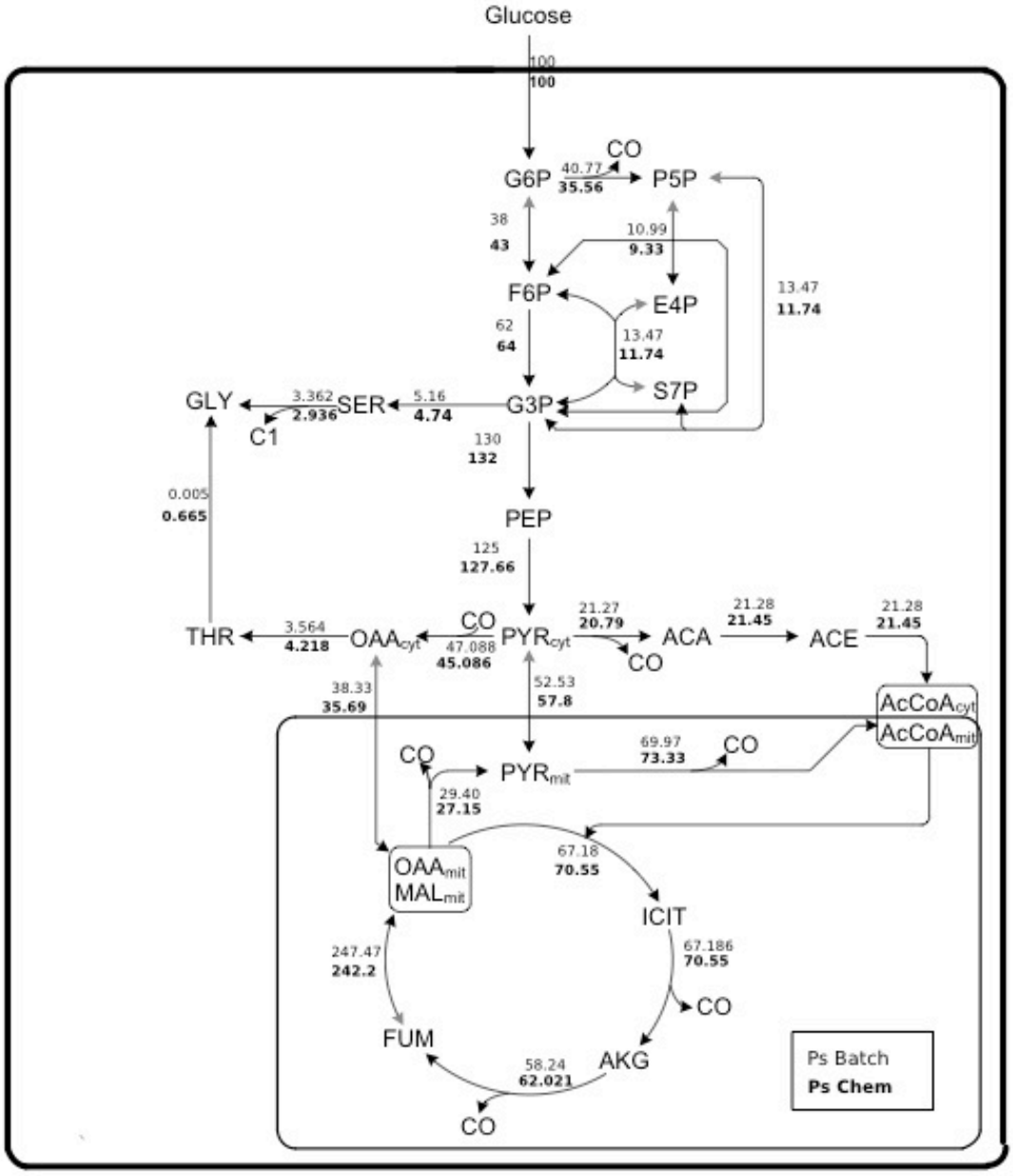

Fig. 11 B Intracellular flux distribution of *S. stipitis* under batch and chemostat cultivations based on ^{13}C-labeling.

The flux through the malic enzyme (MAE) in *S. stipitis* is lower than that observed for *S. cerevisiae*; this result, linked to the lower flux found through the pyruvate carboxylate (PYC), might indicate that anaplerotic reactions are regulated differently in the two yeasts. Overall, analysis of flux distribution based on ^{13}C-labeling indicated the presence of respiratory metabolism for *S. stipitis* under both batch and chemostat

conditions showing similarities to the respiratory mode showed by *S. cerevisiae* during chemostat conditions, even though several differences have been found.

To further characterize the differences observed between the two yeasts during batch cultivations, we performed metabolome profiling (Villas-Boas, *et al.*, 2005, Mapelli, *et al.*, 2008). Analysis of intracellular metabolites, also known as metabolic fingerprinting, allows the quantification of the metabolites present in the cell under a certain condition and can thus furnish detailed information about its metabolic state. For this study, intracellular metabolites from batch cultures of *S. cerevisiae* and *S. stipitis* were analyzed and compared. After statistical analysis, we identified 114 metabolites to be differently represented between the two yeasts based on fold-change. The most significant differences are shown in Fig. 12 where metabolites are listed according to the pathway to which they belong.

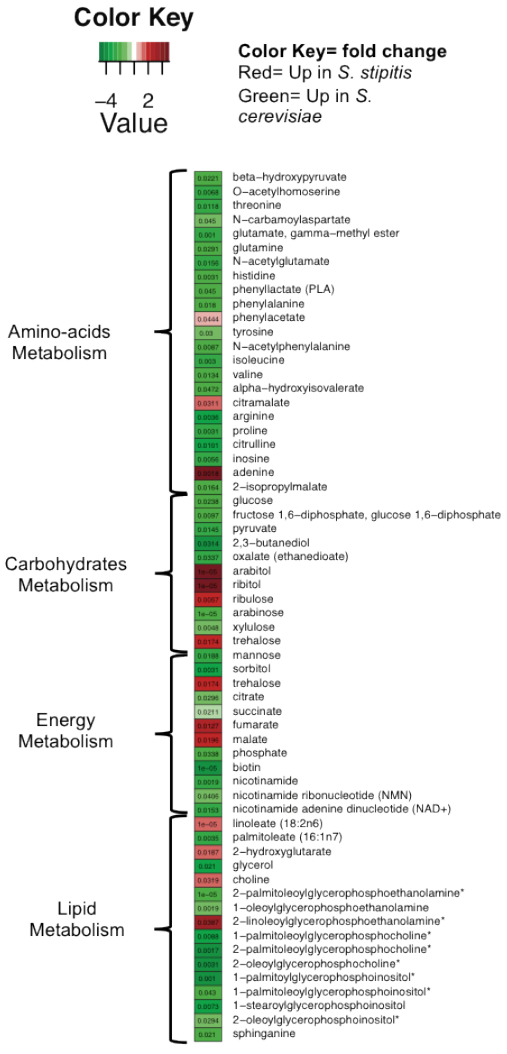

Fig. 12 Results from metabolome analysis. The color indicate fold-change based on protein-normalized levels.

Several differences were found in the content of amino-acids as most of them are present in slightly lower amounts in *S. stipitis*; differently, adenine is present in remarkably higher amount, probably indicating the

higher specific growth rate and biomass yield of *S. stipitis* compared to *S. cerevisiae*. Several differences are present in the super-pathway of carbohydrates. In *S. stipitis*, high amount of sugar-alcohols such as arabitol and ribitol are found. The abundance of these metabolites, together with the increase found in ribulose-5-phosphate levels, might suggest that these polyols can be synthesized in *S. stipitis* from the pentose-phosphate pathway intermediate. It is known that *S. stipitis* can metabolize sugars such as xylose and arabinose and it seems likely that the same pathway used for their assimilation is used in the reverse direction to generate polyols. The presence of these sugars might have, as observed for *Aspergillus niger* under oxygen-limiting conditions, a role in maintaining the osmotic balance or might be used as storage carbohydrates (Diano, *et al.*, 2006). In the carbohydrate super-pathway, a strong increase in the trehalose levels is found for *S. stipitis*. This finding might be connected, together with an increase in adenine level, to the increased biomass yield on substrate of *S. stipitis*, in agreement with previous results (Van Urk, *et al.*, 1990) suggesting that, in Crabtree negative yeasts, higher rates of biomass production can be ascribed to accumulation of reserve carbohydrates. Energy metabolism also shows some differences: the TCA intermediates fumarate and malate are found to be present in higher amounts in *S. stipitis*, while citrate and succinate are slightly lower. Additionally, intracellular metabolome analysis uncovered several differences in lipids compositions. The reason for a high fold-change of fatty acid such as linoleate (18:2) and the phospholipid 2-linoleoyl glycero-phosphoetanolamine is due to the fact that these lipids are naturally not abundant in *S. cerevisiae* (only extremely low amount were detected). Choline content is also found increased in *S. stipitis*. Together, the differences found in metabolites involved in lipid metabolism point towards a different composition of fatty acids between the two yeasts.

To gain further insight into *S. stipitis* growing on glucose, we looked at gene expression levels under the two cultivation conditions and compare it to *S. cerevisiae*. Very little is known about regulatory pathways in *S. stipitis* and only a few works describe gene expression levels (comparing expression on different carbon sources) (Jeffries & Van Vleet, 2009), through custom-designed array chips. Recently, the transcriptome of *S. stipitis* has been sequenced in a study comparing expression levels on glucose and xylose, identifying only 214 ORF whose expression was affected by the carbon source (Yuan, *et al.*, 2011). In this work, transcript levels were analyzed in a high-throughput fashion using RNA-seq. From RNA-seq we obtained 5.97 millions reads, of which >95% mapped to the genomes of *S. stipitis*. To capture differences in the transcriptional network of the two yeasts, we used RNA-seq data to describe, in a comparative fashion, differences in expression patterns of genes under batch and chemostat conditions. For this purpose, we used RNA-seq data both qualitatively, identifying significant Gene Ontology terms and Reporter Metabolites based on the Reporter Feature algorithm (Patil & Nielsen, 2005), but also in a quantitative fashion to determine the expression levels of genes involved in central carbon metabolism condition-wise. The Gene Ontology (GO) terms considered significant after statistical test are reported in Fig. 13 A, while Fig. 13 B shows the reporter metabolites around which the most relevant transcriptional differences occur, according to the super-pathway to which they belong. In this figure directional p-values are reported, where terms up-regulated in batch (negative p-value) are shown in green, while up-regulation in chemostat (positive p-value) is shown in red. Here it is

possible to compare the differential representation of each term/reporter metabolite condition-wise, thus identifying differences in expression patterns within the two conditions between *S. cerevisiae* and *S. stipitis*.

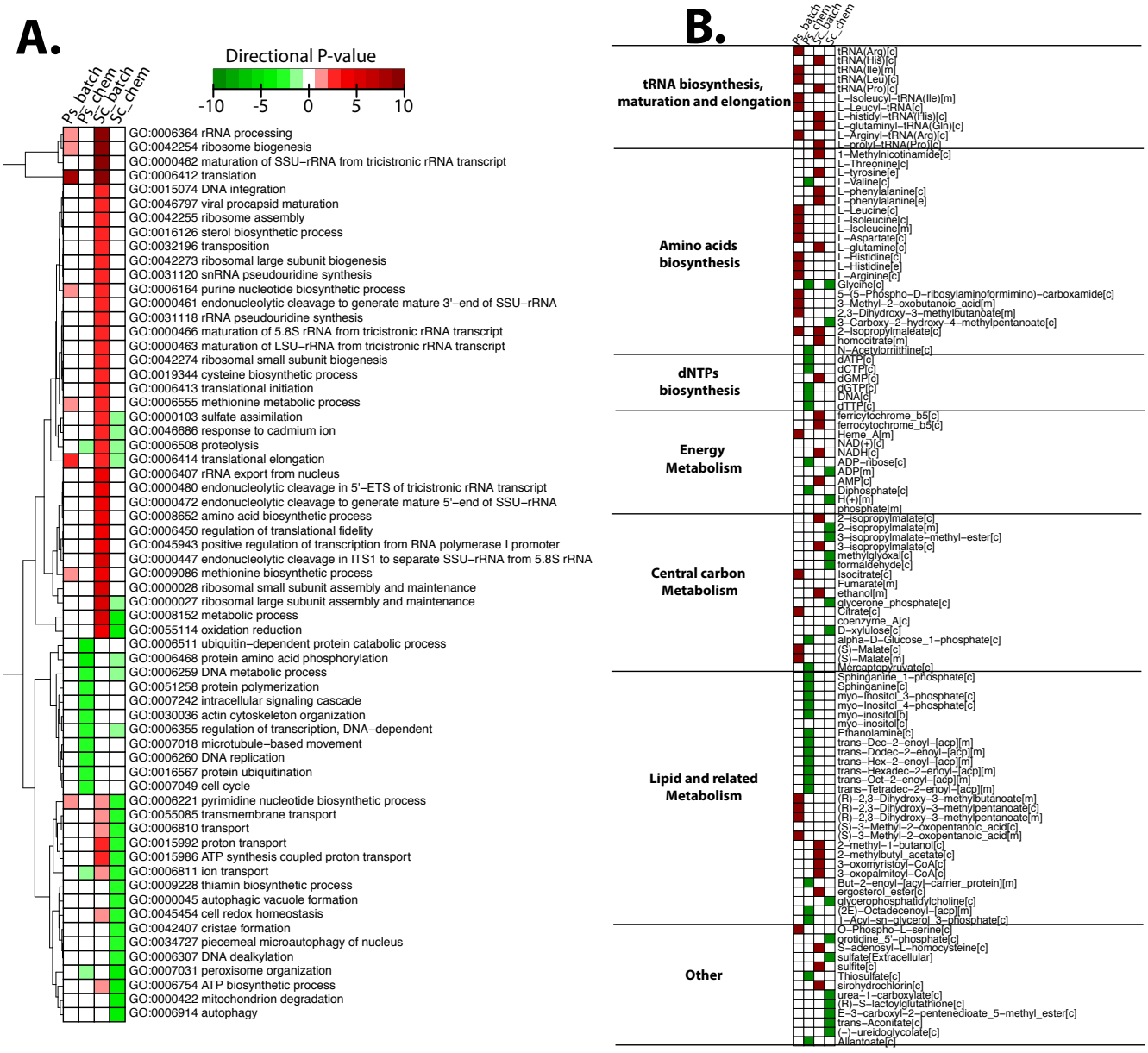

Fig. 13 A) Gene Ontology terms B) Reporter Metabolites identified on RNA-seq data based on the Reporter Feature algorithm. The color represents the directional p-value, red = up in batch, green = up in chemostat

The transcriptional response of *S. cerevisiae* to the shift from respiro-fermentative to respiratory conditions has been extensively characterized through microarray studies, showing significant changes in expression profiles of a considerable number of genes. In agreement with previous results, RNA-seq analysis showed the over-representation, during batch cultivations of *S. cerevisiae*, of a large number of families. A different pattern is observed for *S. stipitis*: comparing the two cultivations conditions it is possible to observe that not many terms are remarkably changed; in *S. stipitis*, mainly growth-related terms (such rRNA elongation, ribosome biogenesis, translation process) are found to be over-represented during batch cultivations, however both

yeasts show an increase in metabolites related to tRNA metabolism during batch conditions (Fig. 13B), as a result of the higher growth rate during batch cultivations. In *S. stipitis* we found several reporter metabolites involved in amino-acid biosynthesis to be up-regulated in batch, while dNTPs biosynthesis appears to be over-represented during chemostat cultivations. Additionally, within metabolites of central carbon metabolism, we found isocitrate, citrate and malate to be increased in *S. stipitis* during batch cultivations.

Besides looking at the global response to glucose excess conditions, RNA-seq data were used to capture differences in transcripts abundances of genes involved in the central carbon metabolism in order to provide a quantitative comparison between the transcriptome of the two yeasts. Similarly to qualitative analysis, our aim was to identify different patterns of expression between the two yeasts in each condition. Expression patterns were analyzed based on FPKM (fragments per kilobase per million sequenced reads), a parameter used to quantify the expression level of a certain mRNA after having normalized for its length and the total number of reads (Oshlack, *et al.*, 2010).

The differential expression of genes involved in the central carbon metabolism is shown in Fig. 14, where $\log_2$ FPKM values are reported. Glycolytic genes show different expression between the two yeasts. Interestingly, the expression of the glucose transporters *HXK1* and *GLK1* is different between the two yeasts: while their expression is glucose-repressed in *S. cerevisiae,* showing lower values during batch cultivations, in *S. stipitis* their expressions do not change as much between the two conditions. The genes for glyceraldehyde-3-phosphate dehydrogenase *TDH1* and *TDH2* are expressed in remarkable lower amount compared to *S. cerevisiae*. Minor differences in the glycolytic genes are found in *PGI*, *FBA1*, *PFK1-2* and *TPI* genes, which show slightly lower expression in *S. stipitis*. PPP genes (not shown in the figure) did not show significant changes under the two conditions for none of the yeasts, in agreement with previous results about their housekeeping function (Nogae & Johnston, 1990, Schaaff-Gerstenschlager & Zimmermann, 1993). Genes involved in the TCA and glyoxylate cycle do show different expression: while in *S. cerevisiae* most of TCA genes are down-regulated during batch conditions (except *IDP1*), in *S. stipitis* this trend is not found and no substantial change in FPKM values is observed between the two conditions. Interestingly, the expression of the main glyoxylate genes *MLS1* and *ICL1* is lower in *S. stipitis* than in *S. cerevisiae* during chemostat cultivations. Genes of the PDH complex *PDA1* and *PDB1* are expressed in slightly higher amount in *S. stipitis*, instead mRNAs coding for genes of the PDC complex (*PDC1* and *PDC2* in *S. stipitis* and *PDC1* and *PDC5* in *S. cerevisiae*) are less abundant in *S. stipitis*. *ADH1* and *ADH2* were found to be less expressed in *S. stipitis* than in *S. cerevisiae*, however a remarkable response to glucose excess conditions was found in *S. stipitis* for ADH2.

Based on RNA-seq data it was possible to uncover a different expression of genes involved in central carbon metabolism of the two yeasts, indicating that transcriptional regulation of central carbon metabolism to glucose excess conditions is substantially different, qualitatively and quantitatively, between *S. stipitis* and in *S. cerevisiae*.

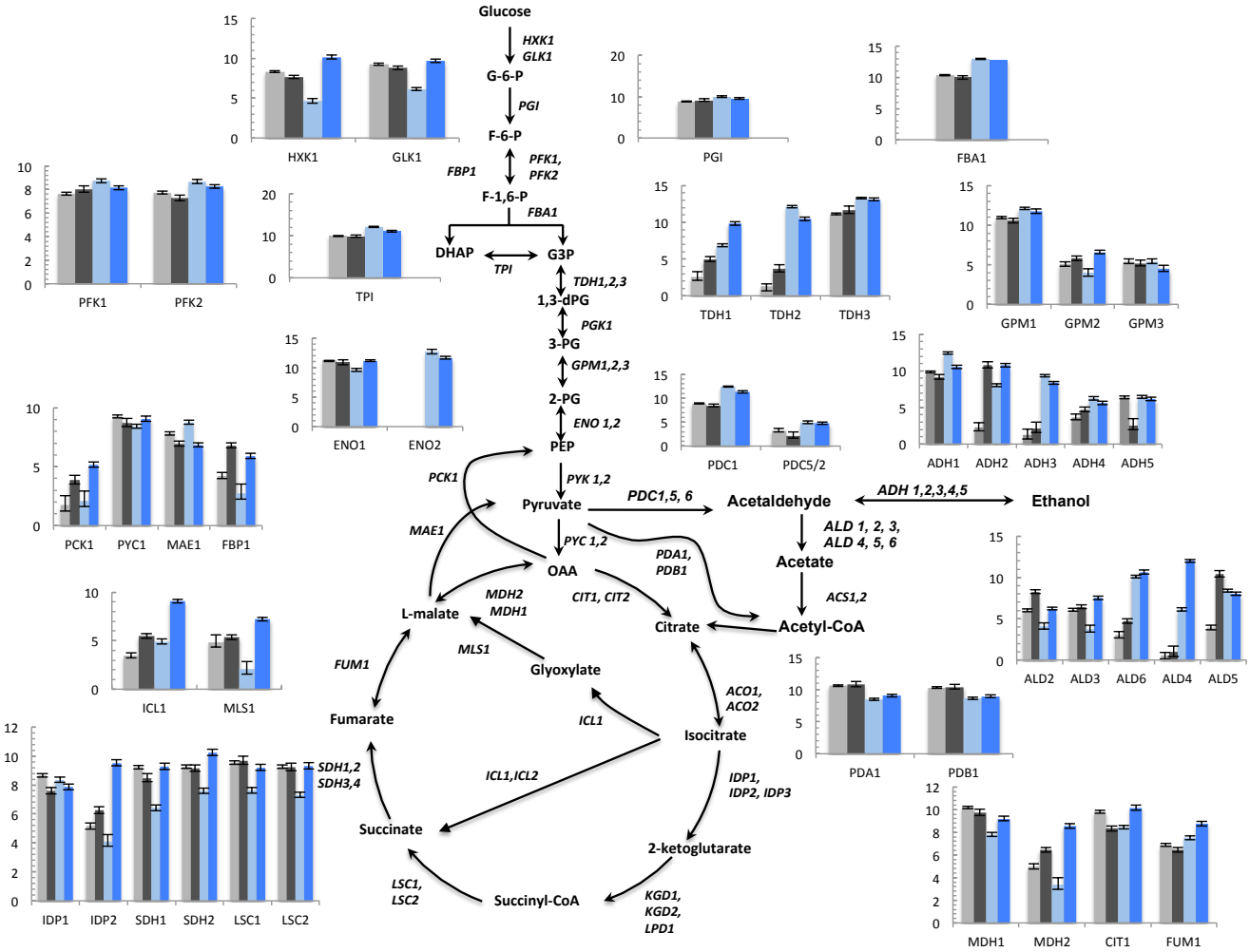

Fig. 14 mRNAs levels of genes involved in central carbon metabolism. Expression level is shown as $\log_2$ FPKM of RNA-seq data processed through Stampy and Cuffdiff. Light grey= *S. stipitis* batch, dark grey= *S. stipitis* chemostat; light blue= *S. cerevisiae* batch, dark blue= *S. cerevisiae* chemostat

Since little is known about transcriptional regulation in *S. stipitis* and RNA-seq highlighted a different regulation of the transcriptional network compared to *S. cerevisiae*, we looked at transcription factors (TF) of the two yeasts performing *in-silico* analysis of transcription factors. For this purpose, we used the Fungal Transcription Factors Database to obtain the protein sequence of the known TFs that we used to reconstruct orthologs families through the software OrthoMCL (Li, *et al.*, 2003). Based on this reconstruction, we identified unique and shared families of transcription factors. Out of 115 unique families of TFs found in *S. stipitis*, only 9 have been characterized and have an established function. The rest are putative or unknown proteins, sharing similarities mainly with *Candida* spp., *Spathospora passalidarum* and *D. hansenii*; however, even in these yeasts, the function of these TFs is unknown. Among the identified ones, we found TFs involved in asparagine metabolism, arginine metabolism, lactose metabolism and fluconazole resistance. Additionally, transcriptional activator of nitrogen regulated genes and TFs involved in controlling transcriptional processes were identified. Interestingly, the subunit A of histone-like transcriptional activator of respiratory functions (Gene ID: 4851321), was uniquely identified in *S. stipitis*.

The analysis of unique TFs in *S. cerevisiae* was facilitated by the larger availability of information present for this yeast. Among the unique families of TFs, a list of those involved in central carbon metabolism can be found in Table 8. Interestingly, gluconeogenic regulators such as *RMD5* and *VID22* were found unique for *S. cerevisiae,* as well as the phospholipids biosynthesis pathway regulators *INO2* and *INO4*. The TF *SGC1*, known to control the expression of genes of the glycolytic pathway, identified as suppressor of defects caused by mutation in *gcr1*, was found uniquely present in *S. cerevisiae*. Additionally, *RGT1*, regulating expression of several glucose transporters of the *HXT* family was also uniquely identified in *S. cerevisiae*. These findings, together with differential expression of genes of central carbon metabolism, support the hypothesis that the two yeasts have different regulatory mechanisms controlling gene expression under the two cultivation conditions. Unfortunately, due to the limited amount of information available for *S. stipitis*, it was not possible to address further the functions of the unique TFs identified in the Crabtree negative yeast.

Table 8. Unique TFs involved in central carbon metabolism identified in the Crabtree yeast *S. cerevisiae*

ORF	***Name***	***Description***
YDR123C **YOL108C**	*INO2* *INO4*	Positive regulatory required for depression of the phospholipid biosynthetic enzymes, regulated by OPI1.
YOR344C	*SGC1*	Serine-rich protein binding E-boxes of glycolytic genes and contributes to their activation. Has been found to suppress the *gcr1* requirement for enolase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase, phosphoglycerate mutase, and pyruvate kinase gene expression. It is necessary for maximal enolase expression
YKL038W	*RGT1*	Transcription factor that regulates expression of several glucose transporter (HXT) genes in response to glucose.
YGL209W	*MIG2*	Regulatory protein, involved in glucose repression of the *SUC* genes.
YER028C	*MIG3*	Transcriptional repressor controlled by Snf1 involved in controlling the transcription of *SIR* genes. Also involved in the response to toxic agents.
YLR373C	*VID22*	Glycosylated integral membrane protein involved in fructose-1,6-bisphosphatase (FBPase) transport and degradation.
YDR255C	*RMD5*	Conserved protein involved in the degradation of the gluconeogenic enzyme fructose-1,6-bisphosphatase; also required for sporulation. Negative regulator of gluconeogenesis.
YJL103C	*GSM1*	Putative zinc cluster protein of unknown function; proposed to be involved in the regulation of energy metabolism.

The systems-biology based comparison performed in this work contributed to the understanding of the metabolism of *S. stipitis* during growth on glucose under different cultivations conditions. Our study confirmed the presence of one main respiratory mode for *S. stipitis* regardless of the glucose concentration. Despite the differences in genome evolution and metabolism of these yeasts, the respiratory mode of *S. stipitis* shows remarkable similarities to the metabolic patterns observed for *S. cerevisiae* during respiratory conditions. Through an integrating systems-level approach, it was possible to address more in depth differences and similarities between the main metabolic mode of *S. stipitis* and the oxidative growth of *S. cerevisiae* and to identify transcription factors involved in controlling different response to glucose excess. Additionally, sequencing of RNA uncovered different patterns of expression of several genes involved in the central carbon metabolism between the two yeast species. Despite the non-direct relationship between mRNA levels, metabolites and fluxes, the attempt to integrate data coming from these different techniques presented in this work, uncovered a robust consistency in identifying a one-mode phenotype of *S. stipitis*, sharing similarities to the patterns found in *S. cerevisiae* during oxidative growth.

3.5 From reads to differential gene expression: a comprehensive comparison of RNA-seq based transcriptome analysis and cross comparison with microarrays

After having used RNA-seq to compare the expression patterns of genes from *S. cerevisiae* and *S. stipitis*, we wanted to address the performances of RNA-seq in describing transcriptomic data and compare it to microarray analysis. We therefore analyzed the contribution of the different steps involved in analysis of RNA-seq data and compared with the results obtained through microarrays analysis. Case study for this work was the metabolism of *S. cerevisiae* under batch and chemostat cultivation conditions. These conditions reflect different metabolic states arising as response to glucose excess conditions, allowing the detection of differences in transcript levels. Additionally, these conditions have been deeply characterized and an extensive amount of data is present as reference.

The development of next generation sequencing (NGS) accounted for a substantial improvement in the field of transcriptomic studies, offering the possibility to analyze genome-wide expression. RNA-seq is a technology that allows for high-throughput analysis of transcriptomes and, as a consequence of its increasing application in the last years, methodologies and tools for data analysis are undergoing a fast development, accounting for a decrease in the cost of RNA-seq to the point that this technique can start to be routinely used in academic labs. Through RNA-seq, it is possible to reconstruct the transcriptome of a (micro)organisms based on a reference genome or without a priori knowledge of its genome sequence (*de novo* sequencing). RNA-seq can provide significant information about splicing variation, smRNA and miRNA and allows for the analysis of non-translated regions controlling gene expression such as UTR and bidirectional promoters. Additionally, through RNA-seq, it is possible to identify SNVs and detect allele-specific forms of splicing. RNA-seq can also be used to quantify gene expression as an alternative technique to microarray analysis. The main advantage of RNA-seq compared to hybridization array, is that RNA-seq is free of background hybridization and allows for a higher dynamic range; furthermore, with RNA-seq, dependency on probe design is also avoided. Besides RNA-seq and hybridization arrays, tiling arrays have also been used to characterize expression levels of genes. These chips are high-density DNA microarrays based on overlapping probes covering the entire genome (Mader, *et al.*, 2011) but differently to conventional microarrays, tiling arrays allow for the analysis of differential expression within the same sample, however a main limitation of tiling arrays is that they have a high level of noise.

While the genome of *S. cerevisiae* was released as early as 1996 (Goffeau, *et al.*, 1996) the first RNA-seq based transcriptome study on *S. cerevisiae*, performed on the strain BY4742, was published in 2008 (Nagalakshmi, *et al.*, 2008). This pioneer work in the field, addressed some fundamental properties of the yeast transcriptome, indicating its complexity.

As the interest of RNA-seq as a technology for replacement of microarrays in transcriptomic studies is increasing, it is of outmost importance to evaluate the performances of the two methodologies in identifying differential expression and to address the consistency of results coming from the two platforms. Marioni *et al.*

(Marioni, *et al.*, 2008) report a comparison with microarray data, identifying high reproducibility within replicates and a correlation based on fold-change between RNA-seq data and microarrays ≥ 0.7.
In our study, we evaluated the performances of the two platforms (microarray and RNA-seq) in detecting transcripts, identifying expression levels and differential gene expression (DGE). Additionally, reproducibility within replicates is also addressed. We started our analysis by addressing the properties of different methods for aligning short-reads obtained from sequencing, comparing different read-aligners and, successively, we analyze the effects of different statistical methods for DGE identification, thus highlighting the contribution of the different steps involved in the analysis of data generated through RNA-seq.
Differently from previous genome-wide transcriptome studies, mainly based on the strain BY4742, we used the *S. cerevisiae* strain CEN.PK113-7D. This strain is widely used as a laboratory strain in biotechnological research and its genome has been sequenced in 2010 (Otero, *et al.*, 2010), showing a genetic variation on ORF, compared to the strain s288c, of 14%. A significant amount of SNVs (single nucleotide variation) were detected in metabolic genes, especially in the galactose and ergosterol biosynthetic pathways. Despite S288c is one of the most well annotated and characterized laboratory strain, whose genome is used as reference for the yeast database SGD, the genetic variation found might lead to inconsistencies when the data available for this strain are used to interpret the behavior of other *S. cerevisiae* strains.

Additionally to RNA-seq, we also sequenced the DNA (DNA-seq) of the strain CEN.PK113-D. The results obtained are reported in Table 9, in comparison to the reference strain S288c, showing that 28139 SNVs and 3520 INDELs (insertion/ deletion) were identified between the two strains. Around 60% of SNVs and 22% of INDEL were found to be in the ORFs and a significant number were also present in the upstream (-1000 bp) region. The results obtained uncovered a higher genetic variation than that reported by Otero *et al.* (Otero, *et al.*, 2010). As the Affymetrix probes are designed on the genome of S288c, we wanted to investigate the presence of SNVs or INDEL on the probes that might lead to problems during hybridization. For this purpose, we mapped the Affymetrix probes on the sequenced CEN.PK113-7D genome, identifying 2472 and 119 probes with SNVs and INDEL, respectively.

Table 9. Results from DNA-seq of the strain CEN.PK 113-7D compared to S288c. Summary of genetic variations (SNV and INDEL) between *S. cerevisiae* CENPK 113-7D and reference strain S288c. The numbers show the genetic variation on specific region; the numbers in parenthesis represent the number of features containing at least one genetic variation in their ORF, upstream sequence, or probe.

Region	SNVs	INDELs
In the genome	28139	3520
1000 bp upstream	15494 (2469)	2704 (1175)
Open reading frame ORF	17139 (2269)	702 (335)
Microrray probe (total 640011)	2868 (2472)	151(119)

To analyze the effects of different aligners and statistical methods on RNA-seq data and compare the results with those obtained through the microarray technology, *S. cerevisiae* was cultivated in triplicates under batch and chemostat cultivations and the RNA extracted from each biological replicate was, in parallel, sequenced on the "Genome-Analyzer" (Illumina) and hybridized to "Yeast Genome 2.0" microarray chips (Affymetrix).

To process the short-reads obtained from the sequencing process, there are two common approaches: one is to map based on a reference genome whereas the other, called *de novo* assembly, allows the sequencing without a reference genome. We initially performed assembly using as reference the genome of S288c. When mapping to a reference genome, different methods to map the short-reads are available. The short-reads from sequencing process were aligned using three available aligners: Gsnap, Stampy and Tophat and the pairwise-correlation within each replicate determined. Initially, we calculated the pairwise-correlation based on intensities (FPKM for RNA-seq data and normalized signal for microarray analysis), as shown in Fig.15 A. The sample-wise comparison with different aligners, showed a correlation of ≥ 0.94, while the cross-comparison between the two methodologies showed also high (but lower than the previous) correlation (≥ 0.81). Additionally it was possible to identify a good reproducibility among replicates. Besides looking at specific intensities, we sought to assess how the different aligners perform when fold-change (batch vs chemostat) is considered, as shown in Fig. 15 B. Based on fold-change, the three different aligners (with and without removal of potential duplicates arising during library preparation) showed high pairwise correlation values (≥ 0.99) and the cross-comparison with array data was substantially improved by using fold-change (≥ 0.93).

Addressing differential expression is of critical importance in transcriptome studies, therefore we wanted to evaluate the influence of different statistical methods to capture differential gene expression (DGE) and compare the results obtained with these methods with the DGE obtained based on analysis of microarray data. After processing the short-reads with the Stampy package, we compared the performances of 5 different statistical methods to capture differential expression: DEseq, BaySeq, edgeR, noiSeq and Cuffdiff. The capabilities of the different statistical methods to detect DGE (where DGE is considered as differential gene expression between the two conditions: batch vs chemostat) after a cutoff Q-value < 1e-5 (noiSeq Pr >0.875) are shown in Fig. 16 A. Interestingly, all the statistical methods for differential expression based on RNA-seq were able to capture 962 DGE. The number of uniquely identified DGE through Cufflink, baySeq, DESeq and noiSeq is relatively low, however edgeR identified a higher number of DGE. When comparing the DGE identified with RNA-seq and with microarrays (Fig. 16 B), 828 ORF were commonly detected by all methods, while145 DGE were identified by microarray but not captured by RNA-seq. Cufflink, noiSeq and BaySeq only led to a small number of uniquely identified DGE, while DESeq captured a high number of DGE (347).

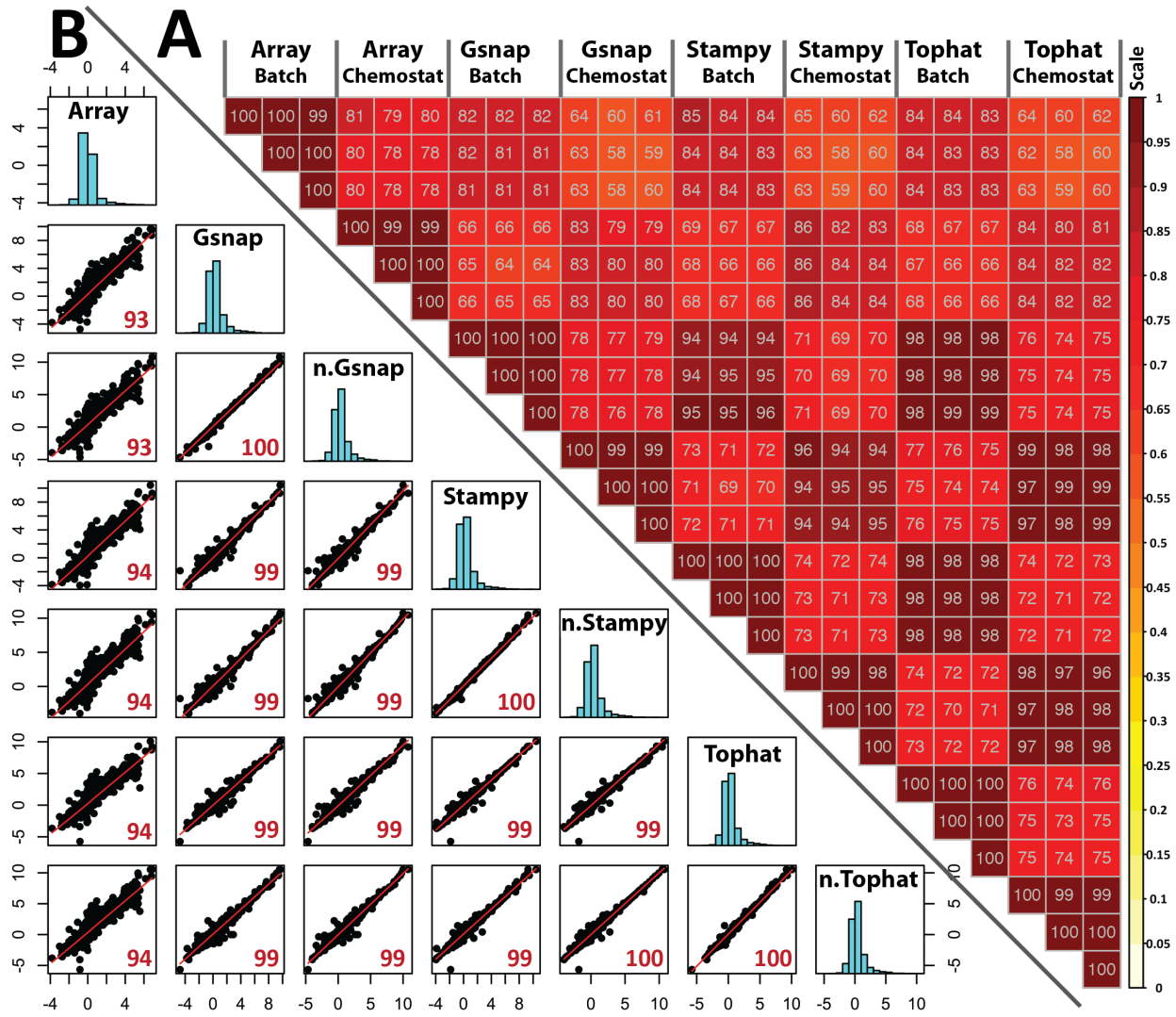

Fig. 15 Sample-wise correlation of transcriptome data from microarray and RNAseq with different processing methods. A) Pair-wise correlation of different biological replicates from batch and chemostat cultivations (for microarray analysis the normalized signals and for RNAseq analysis the FPKM valued were used). The color intensities (scale in the side bar) and the numbers indicate the degree of pair-wise correlation. **B)** Lower-left triangle matrix: scatter plot based on fold changes of gene expression (average values, batch vs chemostat). The red numbers indicate the level of pairwise correlation between different methods. On the diagonal of the triangle matrix, the distribution of fold changes of each processing methods is presented as histograms. Array = microarray, *Gsnap, Stampy and Tophat* = process quality reads by Gsnap, Stampy and Tophat aligners after removal of potential PCR duplicates, *n.Gsnap , n Stampy, nTophat*= process quality reads by Gsnap aligner without removing potential PCR duplicates.

After having addressed the performances of different statistical methods to capture DGE after aligning with Stampy, we sought to determine the impact of processing methods on DGE identification. To this end, we sought to determine whether using different alignment software, specifically Gsnap, Stampy and Tophat, could influence the detection of DGE and, to this purpose, performed statistical analysis on the reads

processed with the three aligners using Cuffdiff. The results obtained were compared to the DGE identified through microarrays and are shown in Fig.16 C. Here, it is possible to observe the high consistency of Cuffdiff in identifying DGE on the reads aligned with Tophat and Stampy, while processing with Gsnap led to a higher number of identified DGE. As shown in the Venn's diagram, 1130 DGE were identified by all the platforms regardless of the aligner used; however, while 364 were uniquely identified through microarray analysis, 512 DGE were identified through RNA-seq with all the three read-aligners but not through microarrays.

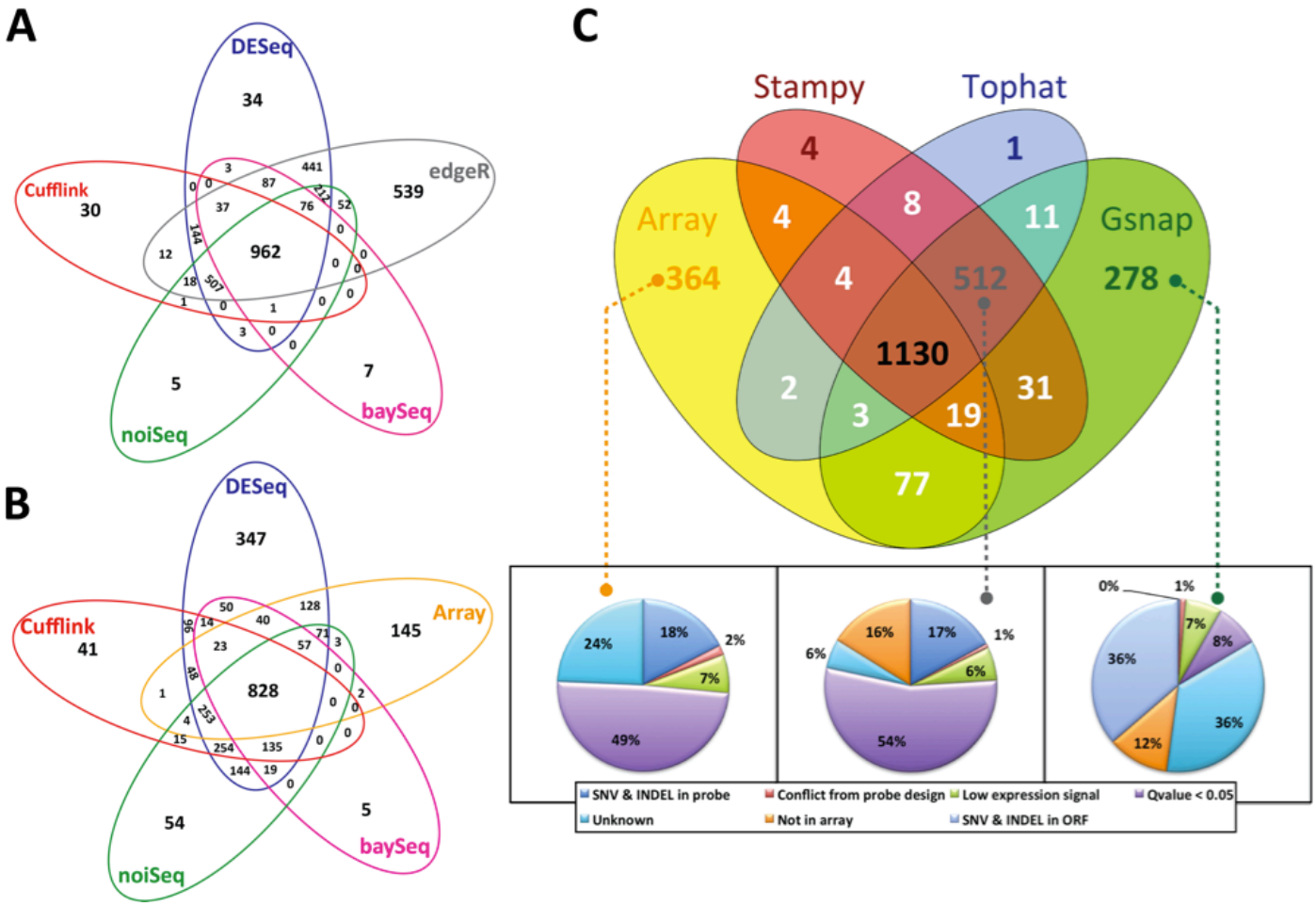

Fig. 16 Comparisons of number of differential gene expression (DGE) identified by different statistical methods for analysis of RNA-seq data and cross comparison with differential gene expression identified from microarray data. **A)** Venn diagram of the comparison of DGE based on RNA-seq data (result from Stampy aligner) through five different statistical methods: Cuffdiff, DESeq, noiSeq, edgeR and baySeq. **B)** Venn diagram of the cross comparison of DGE based on RNA-seq data (result from Stampy aligner) identified through Cuffdiff, noiSeq and DEseq method versus differential gene expression from microarray data. **C)** Venn diagram of the cross-comparison of DGE based on RNA-seq data identified through Cuffdiff method, using the three different aligners. The potential factors underlying the differences in genes identified with each method are presented as percentages pie chart. A cutoff Q-value < 1e-5 for all method except noiSeq $p > 0.875$ was used for the analysis.

Interestingly, 16% of the DGE uniquely identified by using all the aligners, showed that these differentially expressed ORF are not present on the microarray chips, while 17% showed SNVs and/or INDELs. When

decreasing the stringency of the cutoff Q-value, the number of DGE commonly identified with the different read-aligners is reduced by 49% in microarray and by 54% in RNA-seq data. Surprisingly, processing of data with Gsnap aligner allowed for the unique identification of 278 DGE. More than 36% of the DGE identified with Gsnap contain SNVs or INDEL in their ORF while it was not possible to address the reason of unique DGE identification for 36% of them. Based on these results, we can suggest that Gsnap does not perform as well as the other read-aligners when analyzing transcriptomes with SNVs or INDELs compared to the reference.

As alternative approach, we used *de novo* assembly using the software Trinity to eliminate the potential effects due to the presence of SNVs and INDEL with the reference genome S288c (Robertson, *et al.*, 2010, Grabherr, *et al.*, 2011). This approach is computationally more demanding than the traditional one but it allows the assembly of a transcriptome even when a reference genome is not available. We compared the results obtained with *de novo* sequencing to these obtained with the traditional approach and with microarrays. Similarly to what we found with the reference approach, the correlation observed among biological replicates is remarkably high (≥ 0.98). The intensity-based correlation among different samples was also in good agreement with the results obtained mapping without a reference genome (≥ 0.87) and only slightly reduced when comparing to the intensities obtained with microarrays (≥ 0.86); pairwise correlation based on fold-change was also remarkably high (0.96 when comparing to the results obtained with the reference genome and 0.91 with microarrays). As done with the traditional mapping approach, we compared the *de novo* approach with the reference approach and with microarrays in identifying DGE. *De novo* processed data were analyzed for differential expression through Cuffdiff and DESeq. We found that the majority of transcripts (5617) could be captured with all methods, while only 67 genes could not be detected through *de novo* assembly and this was mainly due to their low expression values. Overall, *de novo* sequencing approach for DGE identification gave results well in agreement with these obtained through microarrays and when mapping on a reference genome.

To enhance the utilization of RNA-seq in transcriptomic studies, we wanted to address the impact of using different statistical methods to confer biological meanings to RNA-seq data and compare with microarrays. To this purpose, we applied a holistic approach to detect changes in the global response of yeast metabolism under batch and chemostat conditions using the Reporter Feature algorithm. The statistical methods that are compared for analysis of RNA-seq data are edgeR, baySeq, DESeq, Trinity (*de novo* assembly) and Cuffdiff. Based on the Reporter Feature algorithm, we identified 38 significant Gene Ontology terms below the cutoff p-value 0.05, shown in Fig. 17.

Interestingly, all the methods showed consistency in identifying GO terms with high p-values related to growth (consequence of the increased growth during batch cultivations), despite the term *translation elongation GO0006414* reported different p-values based on the different methods (particularly with Cuffdiff and DESeq). On the contrary, some of the terms known to be relevant during fully respiratory growth do not

always show consistency among the different methods. Specifically, edgeR showed inconsistencies in capturing fatty acid beta-oxidation terms (as well as DESeq), fatty acid metabolic process and TCA, while BaySeq failed to identify increased p-values of ATP-coupled proton translocation. IMP biosynthesis also showed high p-values for DESeq, arrays and BaySeq while p-value was lower for Cuffdiff, Trinity and edgeR. Overall when referring to microarray data, edgeR and BaySeq were found to produce the most discrepant data.

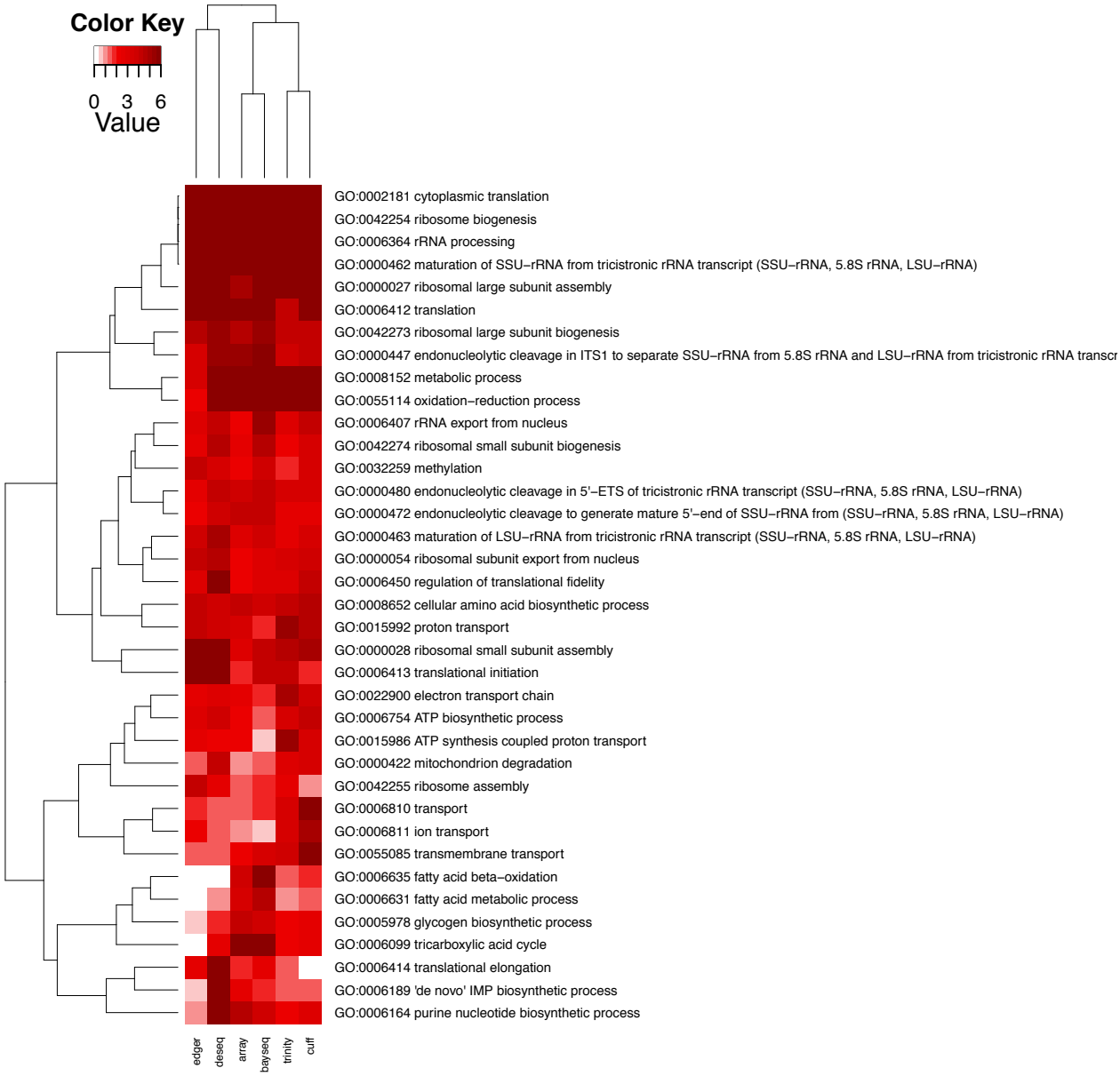

Fig. 17. Results of GO terms based on microarrays data and RNA-seq data analyzed through edgeR, DESeq, BaySeq, Cuffdiff and *de novo* assembly based on Trinity

In this study we addressed the properties of the genome of the strain CEN.PK113-7D by comparing it to the well-characterized *S. cerevisiae* strain S288c, identifying a higher genetic variation than previously reported. For the first time, we provided a comprehensive cross-comparison between the two platforms used for

transcriptomic studies: microarrays and RNA-seq, addressing the contribution of the different steps involved in analysis of RNA-seq data. Different read-aligners and statistical methods for DGE identification were compared and cross-compared to the microarray platform. From our work we concluded that Stampy has the highest accuracy in detecting transcripts with genetic variation while Gsnaps, which requires a minor amount of time to process the short-reads, showed to perform better when a reference genome is available; in regards to that, Tophat represents a good compromise between computational times and accuracy. Overall a good correlation between the RNA-seq and microarrays platforms was observed, both when comparing sample-wise intensities and when detecting fold-change. The different statistical methods for DGE identification of RNA-seq data perform similarly good in detecting DGE, although edgeR showed some inconsistencies, identifying an increased number of DGE compared to the other methods. The effects of different aligners in assessing DGE upon statistical analysis were also evaluated and compared to microarrays. The results based on this comparison are well in agreement, showing that the majority of DGE identified with the microarray platform were mainly due to the stringency of the cutoff applied or to the presence of genetic variation. Using *de novo* assembly showed high consistency to the traditional mapping approach and to microarray in detecting transcripts, expression values and DGE. Additionally, we showed that all the methods, beside edgeR, performed well in identifying GO terms differentially expressed between the two biological conditions, indicating consistency in identifying biological meaning.

Technical aspects of RNA sequencing (RNA-seq)

Through RNA-seq, complementary DNAs (cDNAs) generated from the RNA of interest are directly sequenced using next-generation sequencing technologies (Nagalakshmi, *et al.*, 2010). The total RNA is extracted from the yeast cell and (usually) enriched for poly-A. Poly-A RNA is firstly converted into double-stranded cDNA that is successively fragmented and used for library preparation, according to the Illumina library preparation protocol, resulting in a library of cDNA fragments with adaptors attached to one or both ends. Each molecule is then sequenced at the core facility through the Illumina Genome-Analyzer. Other sequencing platforms are available, such as SOLid or Roche 454.

Results of the sequencing are millions of short reads (25bp to 300 bp) that can be taken from one end or from both ends of each cDNA fragment (known as paired-end) (Martin & Wang, 2011).

Several steps are necessary to analyzed the short reads obtained through the sequencing run and attribute biological meaning (Garber, *et al.*, 2011):

1. Read mapping: align reads to a reference genome. This process can be accomplished using tables-based methods or Burrows Wheeler transform or Seed methods. A software package widely used is **TopHat**, able to map non-junction reads using *Bowtie*. Other read-aligners are Stampy and Gsnaps. Alternatively, it is possible to perform *de novo* assembly.
2. Transcriptome reconstruction: mapped-reads have to be assembled into biologically meaningful units (gene-level, exon-level or transcript-level). This can be done based on genome-guided reconstruction or genome-independent assembly (*de novo* assembly).
3. Expression quantification: statistical testing is applied and data are normalized, leading to a gene list with associated p-values. In this study, we used within-library normalization methods that allow the quantification of the expression level of each gene relatively to the other genes found in the same sample. The number of reads mapped on a gene depends on its abundance and on its length. For gene with single isoforms, it is possible to use the **Cufflink** software to obtain a parameter called FPKM that describes the abundance of a transcript, weighted by the average of gene length
4. Biological meaning has to be attributed to the reconstructed list of gene. This task can be accomplished using methods similar to those applied for microarray analysis, such as the identification of Gene Ontology terms.

4. Conclusions and perspectives

The research I carried out during my PhD studies is focused on metabolic engineering of the central carbon metabolism of *S. cerevisia*e. Particularly, in this thesis, it was addressed how systems biology can serve as a tool to gain new knowledge of cellular phenotypes of wild-type and recombinant yeasts and evaluate different metabolic engineering strategies. In this regards, I sought to apply different approaches to engineer the central carbon metabolism of *S. cerevisiae* and evaluate the consequences of the introduced perturbation on the reprogramming of cellular activities through different high-throughput techniques.

Two distinct metabolic engineering strategies, described in **paper I** and **paper II**, were used to attempt to manipulate the glycolytic flux, one of the core pathways of central carbon metabolism. In **paper I** a direct and more intuitive approach to limit the glycolytic flux is presented, showing that blocking the lower part of glycolysis leads to a severe growth defects and to the inability to grow on glucose. The study on the *Δgpm1* mutant showed the elevated stressed response of the recombinant strain and demonstrated that the other isoenzymes (*GPM2* and *GPM3*) could not compensate the interruption in the glycolytic pathway. Based on this study it was found that the mutant is trying to increase its respiratory capacity, very likely in order to compensate for the energy imbalance resulting from the deletion in the lower glycolytic flux; additionally, based on microarray analysis, it was possible to demonstrate that the *Δgpm1* mutant up-regulates processes involved in energy generation and analysis of reporter transcription factor highlighted that the mutants is less catabolite repressed.

Paper II represents an attempt to re-wire the fluxes in the central carbon metabolism through the expression of a heterologous route: the *A. nidulans* phosphoketolase (PHK) pathway. Based on ^{13}C-labeling analysis it was possible to show that the introduction of PHK in *S. cerevisiae* led to an active flux through this reaction, proving the capability of the recombinant strain to utilize this alternative route for glucose dissimilation and increasing the yield of acetate during batch cultivations. Study of the fluxome of the recombinant strain addressed the impact of the introduction of the PHK pathway on the reprogramming of central carbon metabolism, highlighting significant differences at the pyruvate branchpoint and a slightly increased activity of the TCA cycle. However, the installation of the recombinant route did not contribute to reduce the glycolytic flux. What emerges from both approaches is the tremendous robustness of glycolysis as a pathway optimized to maintaining metabolites pools, cofactors balance and energy demands. The work performed demonstrated that the yeast cell machinery strives to ensure the functional robustness of the central carbon metabolism as a result of a long evolution processes, thus highlighting the challenges faced by metabolic engineering approached aimed at re-wiring the central carbon metabolism of *S. cerevisiae*. To overcome the rigidity due to the tight regulation of the cell, more elaborated approaches aiming at tuning the regulatory network are required, thus supporting the importance of a global understanding and a holistic view of regulatory phenomena. A further attempt to manipulate the tight balance of the central carbon metabolism is found in the (unpublished) approach aiming at replacing the glycolytic pathway with the

phosphoketolase route in the *Δgpm1* mutant. The impossibility to evolve the recombinant strain to grow on glucose indicates the strong preference of the yeast cell towards glycolytic pathway as a route for energy generation and for maintaining the balance among different metabolic pathways. Despite the success of numerous metabolic engineering strategies for production of compounds having as precursors metabolites of the central carbon metabolism, what can be concluded based on the work presented here, is that forcing the replacement of endogenous primary pathway with heterologous routes it is by far more challenging than installing heterologous route branching out from primary pathways.

Besides the above-described strategies to re-wire the central carbon metabolism, a different approach to provide a global view of the yeast cell have also been undertaken. In **paper III** a comprehensive and genome-wide study was performed to characterize the onset of respiro-fermentative metabolism in *S. cerevisiae*. This was accomplished by comparing *S. cerevisiae* to the Crabtree negative yeast *Scheffersomyces stipitis* through an integrative approach based on traditional physiology, X-omics technologies (fluxome, metabolome), high-throughput genomics (RNA-sequencing) and *in silico* analysis of transcription factors. Analysis of intracellular metabolic fluxes confirmed the presence of a respiratory mode of a constant *S. stipitis* regardless of glucose concentration, showing similarities, as well as a few differences, to the respiratory metabolism showed by *S. cerevisiae* under chemostat cultivations. Interestingly, metabolome analysis showed an increased content in polyols and a different composition in fatty acids. RNA-seq analysis uncovered differences in transcript levels of different genes involved in central carbon metabolism thus suggesting a different transcriptional regulation between the two yeasts. This pattern was confirmed by *in silico* analysis of transcription factors that identified in *S. cerevisiae* the presence of unique transcription factors controlling the glycolytic and gluconeogenic pathway as well as glucose transporters.

Different studies during my PhD research were based on analyzing gene expression profiles both through microarrays and RNA-seq. Built on these experiences, in **paper IV**, a technical comparison between the two methodologies is presented, addressing the contribution of the different steps involved in attributing biological meaning to shorts-read obtained from NGS. Interestingly, the pairwise correlation observed with the microarray platform is remarkably high, both based on intensities and (even higher) on fold-change, indicating the robustness of the array methodology. The use of different read-aligners to identify differential expression showed high consistency among the different methods and with microarrays, indicating that most of the differentially expressed gene that were uniquely captured through RNA-seq were due to genetic variation in the probe design. Based on our results, we concluded that data coming from the two platforms showed consistent results and can be considered valid to investigate yeast metabolism through differential gene expression. Nevertheless, it is a clear advantage that RNA-seq offers, compared to microarrays, to analyze UTR regions, identifying spliced isoforms and smRNAs, offering a higher dynamic range and thus having great potential of increasing the understanding of gene expression regulation.

Overall, the work performed in this thesis conveys at identifying the contribution of metabolic engineering in physiological studies, addressing the role of systems biology and high-throughput technologies in characterizing recombinant and wild-type strains. The robustness and the optimality of the glycolytic

pathway as result of natural selection is what clearly emerges from the different metabolic engineering strategies performed in this thesis. Systems-level approach contributed to the identification of regulatory differences in the regulatory networks between *S. cerevisiae* and *S. stipitis* during growth on glucose. Additionally, a comprehensive comparison between different methods for transcriptome analysis is provided.

Acknowledgements

The accomplishment of a PhD is an evolution process accompanied by development, improvement and growth. I owe this progress and major step in my life to my Supervisor and PI Prof. Jens Nielsen, who was the only reason why I decided to move to Denmark first and Sweden later. I am extremely thankful for your constant guidance, passion, scientific lead and for sharing your deep knowledge about yeast metabolism with me; as ambitious person as I am, I enjoyed tremendously to work from the beginning on a high-risk project. A special thanks goes to Prof. Dina Petranovic for her ability to encourage me and help me to progress as a scientist and as a person during my PhD research.

I would like to thank my closest collaborators, Dr. Verena Siewers for the molecular biology support and consultancy during strain engineering, Dr. Intawat Nookaew for performing most of the bioinformatics and statistical analysis, together with Tobias Östersund, always ready to supply statistical explanations. Natapol Pornputtapong is also kindly acknowledged for developing the excellent transcriptome browser for support in the CEN.PK113-7D transcriptome project.

I would like to thank all my colleagues and all the technical and administrative staff here at Systems and Synthetic Biology, Chemical and Biological Engineering Department and the administrative staff at the Chemical and Biological Engineering Department, particularly I would like to acknowledge Erica Dahlin and Martina Butorac. A special thanks goes to my closest colleagues Siavash Partow, Juan Octavio Valle Rodriguez, Kuk-ki Hong, Zihe Liu, Dr. Jie Zhang, Dr. Rahul Kumar, Dr. Margarita Salazar, Dr. Roberto Olivares Hernandez, and Dr. Wanwipa Vongsagnak.

There is one person that, despite not being physically present during my time in Sweden, contributed significantly to my professional development, this person is Dr. Jose' Manuel Otero. Manny, when I first met you I was impressed by your honesty, determination in working and will to succeed and ultimately, it was the ambition of being like you one day that motivated me and gave me the strength to put everything under discussion, to go beyond my limits and always perform research at my best.

Dr. Johansson, you have a special place in my life. Despite not being a scientist, you understood what science means for me more than any other scientists. It is only now that I fully understand what you meant by saying that there is a substantial difference between being a PhD student and a PhD. Arch. Fausto, you played a major role in the accomplishment of my PhD and I want to express my thankfulness for always standing by my side, for discussing Greek and Latin literature, philosophy, ethics, for answering my existential doubts about Seneca and Kant during any time of night and day. You reminded me where I am from, thus where I am going. Finally I would like to thank the two persons that contributed most to make me the person and scientist I am today: mom and dad. From you I learned that nothing comes for free and I had to work hard and honestly to accomplish my goals but, first of all, you taught me the value of setting a goal. Mom thanks for your unconditional love and support in work and life, without your wise reasoning and recommendations I would never have accomplished this. Dad, you had to leave me already in the first stage of my PhD research but your teachings and thoughts kept me going and motivated me throughout these years, memory

just can't be erased. To Mrs. Laura Garagnani: what I am now, my high ambitions, my perseverance, my strong will, my obstinacy and my stubbornness are your inheritance, despite more than half a century between us, we have so much in common that I can only say I am thankful from everything I learned from you. Eleonora Ravello de Santi, thanks for your support and for always sharing your optimism and positive throughts with me.

Finally, I would like to thank the love of my life, Gionata. We decided to follow together the same path and shared our lives, aware but not scared by the obstacles in front of us. I am grateful for having had the opportunity to share this fantastic experience with you. As long as I have you on my side, I can do everything and I am ready to take up new challenges.

Ultimately, a big thanks goes to my training and sparring partners at GBG MMA and to all my coaches, I learned so much from you all. It is absolutely true what I learned there, that as long as you can stand and think, you have a chance to succeed.

REFERENCES

Ahuatzi D, Riera A, Pelaez R, Herrero P & Moreno F (2007) Hxk2 regulates the phosphorylation state of Mig1 and therefore its nucleocytoplasmic distribution. *J Biol Chem* **282**: 4485-4493.

Bailey JE (1991) Toward a science of metabolic engineering. *Science* **252**: 1668-1675.

Bailey JE, Sburlati A, Hatzimanikatis V, Lee K, Renner WA & Tsai PS (2002) Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. *Biotechnol Bioeng* **79**: 568-579.

Bakker BM, Overkamp KM, van Maris AJA, Kotter P, Luttik MAH, van Dijken JP & Pronk JT (2001) Stoichiometry and compartmentation of NADH metabolism in Saccharomyces cerevisiae. *FEMS Microbiol Rev* **25**: 15-37.

Barford JP & Hall RJ (1979) Investigation of the significance of a carbon and redox balance to the measurement of gaseous metabolism of Saccharomyces cerevisiae. *Biotechnol Bioeng* **21**: 609-626.

Barnett JA (2000) A history of research on yeasts 2: Louis Pasteur and his contemporaries, 1850-1880. *Yeast* **16**: 755-771.

Blom J, De Mattos MJ & Grivell LA (2000) Redirection of the respiro-fermentative flux distribution in Saccharomyces cerevisiae by overexpression of the transcription factor Hap4p. *Appl Environ Microbiol* **66**: 1970-1973.

Boles E & Hollenberg CP (1997) The molecular genetics of hexose transport in yeasts. *FEMS Microbiol Rev* **21**: 85-111.

Breitenbachschmitt I, Schmitt HD, Heinisch J & Zimmermann FK (1984) Genetic and Physiological Evidence for the Existence of a 2nd Glycolytic Pathway in Yeast Parallel to the Phosphofructokinase-Aldolase Reaction Sequence. *Molecular & General Genetics* **195**: 536-540.

Bro C, Regenberg B, Forster J & Nielsen J (2006) In silico aided metabolic engineering of Saccharomyces cerevisiae for improved bioethanol production. *Metab Eng* **8**: 102-111.

Carlson M (1998) Regulation of glucose utilization in yeast. *Curr Opin Genet Dev* **8**: 560-564.

Casal M, Paiva S, Andrade RP, Gancedo C & Leao C (1999) The lactate-proton symport of Saccharomyces cerevisiae is encoded by JEN1. *Journal of Bacteriology* **181**: 2620-2623.

Christensen TS, Oliveira AP & Nielsen J (2009) Reconstruction and logical modeling of glucose repression signaling pathways in Saccharomyces cerevisiae. *BMC Syst Biol* **3**: 7.

Cimini D, Patil KR, Schiraldi C & Nielsen J (2009) Global transcriptional response of Saccharomyces cerevisiae to the deletion of SDH3. *BMC Syst Biol* **3**.

Ciriacy M & Breitenbach I (1979) Physiological effects of seven different blocks in glycolysis in Saccharomyces cerevisiae. *Journal of Bacteriology* **139**: 152-160.

Clifton D, Weinstock SB & Fraenkel DG (1978) Glycolysis mutants in Saccharomyces cerevisiae. *Genetics* **88**: 1-11.

Colby G, Ishii Y & Tzagoloff A (1998) Suppression of sdh1 mutations by the SDH1b gene of Saccharomyces cerevisiae. *Yeast* **14**: 1001-1006.

Coleman JS & Bhattacharjee JK (1975) Regulation of Citrate Synthase Activity of Saccharomyces-Cerevisiae. *Antonie Van Leeuwenhoek Journal of Microbiology* **41**: 249-256.

Daran-Lapujade P, Daran JM, van Maris AJ, de Winde JH & Pronk JT (2009) Chemostat-based micro-array analysis in baker's yeast. *Adv Microb Physiol* **54**: 257-311.

Daran-Lapujade P, Jansen ML, Daran JM, van Gulik W, de Winde JH & Pronk JT (2004) Role of transcriptional regulation in controlling fluxes in central carbon metabolism of Saccharomyces cerevisiae. A chemostat culture study. *J Biol Chem* **279**: 9125-9138.

Daran-Lapujade P, Rossell S, van Gulik WM, *et al.* (2007) The fluxes through glycolytic enzymes in Saccharomyces cerevisiae are predominantly regulated at posttranscriptional levels. *Proc Natl Acad Sci U S A* **104**: 15753-15758.

De Deken RH (1966) The Crabtree effect: a regulatory system in yeast. *J Gen Microbiol* **44**: 149-156.

de Jong-Gubbels P, Vanrolleghem P, Heijnen S, van Dijken JP & Pronk JT (1995) Regulation of carbon metabolism in chemostat cultures of Saccharomyces cerevisiae grown on mixtures of glucose and ethanol. *Yeast* **11**: 407-418.

de Vries S & Marres CA (1987) The mitochondrial respiratory chain of yeast. Structure and biosynthesis and the role in cellular metabolism. *Biochimica Et Biophysica Acta* **895**: 205-239.

DeRisi JL, Iyer VR & Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**: 680-686.

Devaux F, Marc P & Jacq C (2001) Transcriptomes, transcription activators and microarrays. *FEBS Lett* **498**: 140-144.

Diano A, Bekker-Jensen S, Dynesen J & Nielsen J (2006) Polyol synthesis in Aspergillus niger: influence of oxygen availability, carbon and nitrogen sources on the metabolism. *Biotechnol Bioeng* **94**: 899-908.

Diaz-Ruiz R, Averet N, Araiza D, Pinson B, Uribe-Carvajal S, Devin A & Rigoulet M (2008) Mitochondrial oxidative phosphorylation is regulated by fructose 1,6-bisphosphate. A possible role in Crabtree effect induction? *J Biol Chem* **283**: 26948-26955.

Diderich JA, Teusink B, Valkier J, Anjos J, Spencer-Martins I, van Dam K & Walsh MC (1999) Strategies to determine the extent of control exerted by glucose transport on glycolytic flux in the yeast Saccharomyces bayanus. *Microbiology* **145 (Pt 12)**: 3447-3454.

Elbing K, Larsson C, Bill RM, *et al.* (2004) Role of hexose transport in control of glycolytic flux in Saccharomyces cerevisiae. *Appl Environ Microbiol* **70**: 5323-5330.

Ephrussi B, Slonimski PP, Yotsuyanagi Y & Tavlitzki J (1956) Variations Physiologiques Et Cytologiques De La Levure Au Cours Du Cycle De La Croissance Aerobie. *Comptes Rendus Des Travaux Du Laboratoire Carlsberg-Serie Physiologique* **26**: 87-102.

Erasmus DJ, van der Merwe GK & van Vuuren HJ (2003) Genome-wide expression analyses: Metabolic adaptation of Saccharomyces cerevisiae to high sugar stress. *FEMS Yeast Res* **3**: 375-399.

Famili I, Forster J, Nielson J & Palsson BO (2003) Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc Natl Acad Sci U S A* **100**: 13134-13139.

Fendt SM & Sauer U (2010) Transcriptional regulation of respiration in yeast metabolizing differently repressive carbon substrates. *BMC Syst Biol* **4**: 12.

Fiaux J, Cakar ZP, Sonderegger M, Wuthrich K, Szyperski T & Sauer U (2003) Metabolic-flux profiling of the yeasts Saccharomyces cerevisiae and Pichia stipitis. *Eukaryot Cell* **2**: 170-180.

Fiechter A (1975) Continuous cultivation of yeasts. *Methods Cell Biol* **11**: 97-130.

Flikweert MT, de Swaaf M, van Dijken JP & Pronk JT (1999) Growth requirements of pyruvate-decarboxylase-negative Saccharomyces cerevisiae. *FEMS Microbiol Lett* **174**: 73-79.

Frick O & Wittmann C (2005) Characterization of the metabolic shift between oxidative and fermentative growth in Saccharomyces cerevisiae by comparative 13C flux analysis. *Microb Cell Fact* **4**: 30.

Gadde DM & McCammon MT (1997) Mutations in the IDH2 gene encoding the catalytic subunit of the yeast NAD(+)-dependent isocitrate dehydrogenase can be suppressed by mutations in the CIT1 gene encoding citrate synthase and other genes of oxidative metabolism. *Archives of Biochemistry and Biophysics* **344**: 139-149.

Gancedo JM (1998) Yeast carbon catabolite repression. *Microbiology and Molecular Biology Reviews* **62**: 334-+.

Garber M, Grabherr MG, Guttman M & Trapnell C (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods* **8**: 469-477.

Gelinas P (2009) Inventions on baker's yeast strains and specialty ingredients. *Recent Pat Food Nutr Agric* **1**: 104-132.

Gentalen E & Chee M (1999) A novel method for determining linkage between DNA sequences: hybridization to paired probe arrays. *Nucleic Acids Res* **27**: 1485-1491.

Goffeau A (2000) Four years of post-genomic life with 6,000 yeast genes. *FEBS Lett* **480**: 37-41.

Goffeau A, Barrell BG, Bussey H, *et al.* (1996) Life with 6000 genes. *Science* **274**: 546, 563-547.

Gombert AK, Moreira dos Santos M, Christensen B & Nielsen J (2001) Network identification and flux quantification in the central metabolism of Saccharomyces cerevisiae under different conditions of glucose repression. *Journal of Bacteriology* **183**: 1441-1451.

Grabherr MG, Haas BJ, Yassour M, *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**: 644-652.

Grotkjaer T, Christakopoulos P, Nielsen J & Olsson L (2005) Comparative metabolic network analysis of two xylose fermenting recombinant Saccharomyces cerevisiae strains. *Metab Eng* **7**: 437-444.

Gygi SP, Rochon Y, Franza BR & Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19**: 1720-1730.

Ha SJ, Galazka JM, Kim SR, *et al.* (2011) Engineered Saccharomyces cerevisiae capable of simultaneous cellobiose and xylose fermentation. *Proc Natl Acad Sci U S A* **108**: 504-509.

Hatzimanikatis V & Bailey JE (1996) MCA has more to say. *J Theor Biol* **182**: 233-242.

Hedges D, Proft M & Entian KD (1995) CAT8, a new zinc cluster-encoding gene necessary for derepression of gluconeogenic enzymes in the yeast Saccharomyces cerevisiae. *Mol Cell Biol* **15**: 1915-1922.

Heinemann M & Sauer U (2010) Systems biology of microbial metabolism. *Curr Opin Microbiol* **13**: 337-343.

Heinisch J (1986) Construction and physiological characterization of mutants disrupted in the phosphofructokinase genes of Saccharomyces cerevisiae. *Curr Genet* **11**: 227-234.

Heinisch JJ, Boles E & Timpel C (1996) A yeast phosphofructokinase insensitive to the allosteric activator fructose 2,6-bisphosphate. Glycolysis/metabolic regulation/allosteric control. *J Biol Chem* **271**: 15928-15933.

Heinisch JJ, Muller S, Schluter E, Jacoby J & Rodicio R (1998) Investigation of two yeast genes encoding putative isoenzymes of phosphoglycerate mutase. *Yeast* **14**: 203-213.

Herrero P, Galindez J, Ruiz N, Martinez-Campa C & Moreno F (1995) Transcriptional regulation of the Saccharomyces cerevisiae HXK1, HXK2 and GLK1 genes. *Yeast* **11**: 137-144.

Heyland J, Fu J & Blank LM (2009) Correlation between TCA cycle flux and glucose uptake rate during respiro-fermentative growth of Saccharomyces cerevisiae. *Microbiology* **155**: 3827-3837.

Holzer H (1989) Proteolytic catabolite inactivation in Saccharomyces cerevisiae. *Revis Biol Celular* **21**: 305-319.

Hu G, Lin AP & McAlister-Henn L (2006) Physiological consequences of loss of allosteric activation of yeast NAD(+)-specific isocitrate dehydrogenase. *Journal of Biological Chemistry* **281**: 16935-16942.

Huber W, Toedling J & Steinmetz LM (2006) Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics* **22**: 1963-1970.

Huberts DH, Niebel B & Heinemann M (2012) A flux-sensing mechanism could regulate the switch between respiration and fermentation. *FEMS Yeast Res* **12**: 118-128.

Humphery-Smith I & Blackstock W (1997) Proteome analysis: genomics via the output rather than the input code. *J Protein Chem* **16**: 537-544.

Hung GC, Brown CR, Wolfe AB, Liu J & Chiang HL (2004) Degradation of the gluconeogenic enzymes fructose-1,6-bisphosphatase and malate dehydrogenase is mediated by distinct proteolytic pathways and signaling events. *J Biol Chem* **279**: 49138-49150.

Ideker T, Thorsson V, Ranish JA*, et al.* (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**: 929-934.

Jeffries TW & Van Vleet JR (2009) Pichia stipitis genomics, transcriptomics, and gene clusters. *FEMS Yeast Res* **9**: 793-807.

Jeffries TW, Grigoriev IV, Grimwood J*, et al.* (2007) Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast Pichia stipitis. *Nat Biotechnol* **25**: 319-326.

Johnston M (1999) Feasting, fasting and fermenting. Glucose sensing in yeast and other cells. *Trends in Genetics* **15**: 29-33.

Juneau K, Palm C, Miranda M & Davis RW (2007) High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proc Natl Acad Sci U S A* **104**: 1522-1527.

Kaliterna J, Weusthuis RA, Castrillo JI, Vandijken JP & Pronk JT (1995) Coordination of Sucrose Uptake and Respiration in the Yeast Debaryomyces-Yamadae. *Microbiology-Uk* **141**: 1567-1574.

Kaliterna J, Weusthuis RA, Castrillo JI, Vandijken JP & Pronk JT (1995) Transient Responses of Candida-Utilis to Oxygen Limitation - Regulation of the Kluyver Effect for Maltose. *Yeast* **11**: 317-325.

Kappeli O (1986) Regulation of carbon metabolism in Saccharomyces cerevisiae and related yeasts. *Adv Microb Physiol* **28**: 181-209.

Kjeldsen T (2000) Yeast secretory expression of insulin precursors. *Appl Microbiol Biotechnol* **54**: 277-286.

Konig J, Zarnack K, Luscombe NM & Ule J (2011) Protein-RNA interactions: new genomic technologies and perspectives. *Nat Rev Genet* **13**: 77-83.

Kuhn KM, DeRisi JL, Brown PO & Sarnow P (2001) Global and specific translational regulation in the genomic response of Saccharomyces cerevisiae to a rapid transfer from a fermentable to a nonfermentable carbon source. *Mol Cell Biol* **21**: 916-927.

Kuyper M, Winkler AA, van Dijken JP & Pronk JT (2004) Minimal metabolic engineering of Saccharomyces cerevisiae for efficient anaerobic xylose fermentation: a proof of principle. *FEMS Yeast Res* **4**: 655-664.

Kwast KE, Burke PV & Poyton RO (1998) Oxygen sensing and the transcriptional regulation of oxygen-responsive genes in yeast. *J Exp Biol* **201**: 1177-1195.

Lagunas R (1981) Is Saccharomyces-Cerevisiae a Typical Facultative Anaerobe. *Trends Biochem Sci* **6**: 201-203.

Lam KB & Marmur J (1977) Isolation and characterization of Saccharomyces cerevisiae glycolytic pathway mutants. *Journal of Bacteriology* **130**: 746-749.

Larsson C, Pahlman IL & Gustafsson L (2000) The importance of ATP as a regulator of glycolytic flux in Saccharomyces cerevisiae. *Yeast* **16**: 797-809.

Larsson C, Vonstockar U, Marison I & Gustafsson L (1995) Metabolic Uncoupling in Saccharomyces-Cerevisiae. *Thermochimica Acta* **251**: 99-110.

Larsson C, Nilsson A, Blomberg A & Gustafsson L (1997) Glycolytic flux is conditionally correlated with ATP concentration in Saccharomyces cerevisiae: a chemostat study under carbon- or nitrogen-limiting conditions. *Journal of Bacteriology* **179**: 7243-7250.

Li L, Stoeckert CJ, Jr. & Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* **13**: 2178-2189.

Liu Z & Butow RA (1999) A transcriptional switch in the expression of yeast tricarboxylic acid cycle genes in response to a reduction or loss of respiratory function. *Mol Cell Biol* **19**: 6720-6728.

Liu ZC & Butow RA (1999) A transcriptional switch in the expression of yeast tricarboxylic acid cycle genes in response to a reduction or loss of respiratory function. *Mol Cell Biol* **19**: 6720-6728.

Lynd LR, Wyman CE & Gerngross TU (1999) Biocommodity engineering. *Biotechnol Prog* **15**: 777-793.

Mader U, Nicolas P, Richard H, Bessieres P & Aymerich S (2011) Comprehensive identification and quantification of microbial transcriptomes by genome-wide unbiased methods. *Curr Opin Biotechnol* **22**: 32-41.

Mager WH & Ferreira PM (1993) Stress response of yeast. *Biochem J* **290 (Pt 1)**: 1-13.

Mapelli V, Olsson L & Nielsen J (2008) Metabolic footprinting in microbiology: methods and applications in functional genomics and biotechnology. *Trends Biotechnol* **26**: 490-497.

Marioni JC, Mason CE, Mane SM, Stephens M & Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* **18**: 1509-1517.

Martin JA & Wang Z (2011) Next-generation transcriptome assembly. *Nat Rev Genet* **12**: 671-682.

Marykwas DL & Fox TD (1989) Control of the Saccharomyces cerevisiae regulatory gene PET494: transcriptional repression by glucose and translational induction by oxygen. *Mol Cell Biol* **9**: 484-491.

Mayer FV, Heath R, Underwood E, *et al.* (2011) ADP regulates SNF1, the Saccharomyces cerevisiae homolog of AMP-activated protein kinase. *Cell Metab* **14**: 707-714.

McCartney RR & Schmidt MC (2001) Regulation of Snf1 kinase. Activation requires phosphorylation of threonine 210 by an upstream kinase as well as a distinct step mediated by the Snf4 subunit. *J Biol Chem* **276**: 36460-36466.

McKusick VA (1997) Genomics: structural and functional studies of genomes. *Genomics* **45**: 244-249.

Moore PA, Bettany AJ & Brown AJ (1990) Multiple copies of the pyruvate kinase gene affect yeast cell growth. *J Gen Microbiol* **136**: 2359-2366.

Moore PA, Sagliocco FA, Wood RM & Brown AJ (1991) Yeast glycolytic mRNAs are differentially regulated. *Mol Cell Biol* **11**: 5330-5337.

Moxley JF, Jewett MC, Antoniewicz MR, *et al.* (2009) Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. *Proc Natl Acad Sci U S A* **106**: 6477-6482.

Nagalakshmi U, Waern K & Snyder M (2010) RNA-Seq: a method for comprehensive transcriptome analysis. *Curr Protoc Mol Biol* **Chapter 4**: Unit 4 11 11-13.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M & Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**: 1344-1349.

Nielsen J (2003) It is all about metabolic fluxes. *Journal of Bacteriology* **185**: 7031-7035.

Nielsen J & Jewett MC (2008) Impact of systems biology on metabolic engineering of Saccharomyces cerevisiae. *FEMS Yeast Res* **8**: 122-131.

Nielsen J & Vidal M (2010) Systems biology of microorganisms. *Curr Opin Microbiol* **13**: 335-336.

Nissen TL, Kielland-Brandt MC, Nielsen J & Villadsen J (2000) Optimization of ethanol production in Saccharomyces cerevisiae by metabolic engineering of the ammonium assimilation. *Metab Eng* **2**: 69-77.

Nissen TL, Anderlund M, Nielsen J, Villadsen J & Kielland-Brandt MC (2001) Expression of a cytoplasmic transhydrogenase in Saccharomyces cerevisiae results in formation of 2-oxoglutarate due to depletion of the NADPH pool. *Yeast* **18**: 19-32.

Nogae I & Johnston M (1990) Isolation and characterization of the ZWF1 gene of Saccharomyces cerevisiae, encoding glucose-6-phosphate dehydrogenase. *Gene* **96**: 161-169.

Nookaew I, Jewett MC, Meechai A, *et al.* (2008) The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol* **2**: 71.

Olivares-Hernandez R, Usaite R & Nielsen J (2010) Integrative Analysis Using Proteome and Transcriptome Data From Yeast to Unravel Regulatory Patterns at Post-Transcriptional Level. *Biotechnol Bioeng* **107**: 865-875.

Oliveira AP, Patil KR & Nielsen J (2008) Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. *BMC Syst Biol* **2**: 17.

Oshlack A, Robinson MD & Young MD (2010) From RNA-seq reads to differential expression results. *Genome Biol* **11**: 220.

Ostergaard S, Olsson L & Nielsen J (2000) Metabolic engineering of Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **64**: 34-50.

Otero JM (2009) Industrial Systems Biology and Metabolic Engineering of Saccharomyces cerevisiae. Thesis, Chalmers University of Technology, Göteborg, Sweden.

Otero JM & Nielsen J (2010) Industrial systems biology. *Biotechnol Bioeng* **105**: 439-460.

Otero JM, Vongsangnak W, Asadollahi MA, *et al.* (2010) Whole genome sequencing of Saccharomyces cerevisiae: from genotype to phenotype for improved metabolic engineering applications. *BMC Genomics* **11**: 723.

Overkamp KM, Bakker BM, Kotter P, van Tuijl A, de Vries S, van Dijken JP & Pronk JT (2000) In vivo analysis of the mechanisms for oxidation of cytosolic NADH by Saccharomyces cerevisiae mitochondria. *Journal of Bacteriology* **182**: 2823-2830.

Ozcan S & Johnston M (1999) Function and regulation of yeast hexose transporters. *Microbiology and Molecular Biology Reviews* **63**: 554-+.

Panagiotou G, Andersen MR, Grotkjaer T, Regueira TB, Hofmann G, Nielsen J & Olsson L (2008) Systems analysis unfolds the relationship between the phosphoketolase pathway and growth in Aspergillus nidulans. *PLoS One* **3**: e3847.

Papini M, Salazar M & Nielsen J (2010) Systems biology of industrial microorganisms. *Adv Biochem Eng Biotechnol* **120**: 51-99.

Pasteur L (1876) *Études sur la bière, ses maladies*. Gauthier-Villars, Paris,.

Patil KR & Nielsen J (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci U S A* **102**: 2685-2689.

Peter Smits H, Hauf J, Muller S, *et al.* (2000) Simultaneous overexpression of enzymes of the lower part of glycolysis can enhance the fermentative capacity of Saccharomyces cerevisiae. *Yeast* **16**: 1325-1334.

Petranovic D & Nielsen J (2008) Can yeast systems biology contribute to the understanding of human disease? *Trends Biotechnol* **26**: 584-590.

Phelps TJ, Palumbo AV & Beliaev AS (2002) Metabolomics and microarrays for improved understanding of phenotypic characteristics controlled by both genomics and environmental constraints. *Curr Opin Biotechnol* **13**: 20-24.

Piper MDW, Daran-Lapujade P, Bro C, Regenberg B, Knudsen S, Nielsen J & Pronk JT (2002) Reproducibility of oligonucleotide microarray transcriptome analyses - An interlaboratory comparison using chemostat cultures of Saccharomyces cerevisiae. *Journal of Biological Chemistry* **277**: 37001-37008.

Polakis ES, Bartley W & Meek GA (1965) Changes in the activities of respiratory enzymes during the aerobic growth of yeast on different carbon sources. *Biochem J* **97**: 298-302.

Porro D, Brambilla L, Ranzi BM, Martegani E & Alberghina L (1995) Development of metabolically engineered Saccharomyces cerevisiae cells for the production of lactic acid. *Biotechnol Prog* **11**: 294-298.

Postma E, Verduyn C, Scheffers WA & Van Dijken JP (1989) Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of Saccharomyces cerevisiae. *Appl Environ Microbiol* **55**: 468-477.

Pradet-Balade B, Boulme F, Beug H, Mullner EW & Garcia-Sanz JA (2001) Translation control: bridging the gap between genomics and proteomics? *Trends Biochem Sci* **26**: 225-229.

Presecan-Siedel E, Galinier A, Longin R, Deutscher J, Danchin A, Glaser P & Martin-Verstraete I (1999) Catabolite regulation of the pta gene as part of carbon flow pathways in Bacillus subtilis. *Journal of Bacteriology* **181**: 6889-6897.

Pronk JT, Steensma HY & vanDijken JP (1996) Pyruvate metabolism in Saccharomyces cerevisiae. *Yeast* **12**: 1607-1633.

Raab AM, Gebhardt G, Bolotina N, Weuster-Botz D & Lang C (2010) Metabolic engineering of Saccharomyces cerevisiae for the biotechnological production of succinic acid. *Metab Eng* **12**: 518-525.
Ro DK, Paradise EM, Ouellet M, *et al.* (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**: 940-943.

Robertson G, Schein J, Chiu R, *et al.* (2010) De novo assembly and analysis of RNA-seq data. *Nat Methods* **7**: 909-912.

Rodicio R & Heinisch J (1987) Isolation of the yeast phosphoglyceromutase gene and construction of deletion mutants. *Mol Gen Genet* **206**: 133-140.

Rodicio R, Heinisch JJ & Hollenberg CP (1993) Transcriptional control of yeast phosphoglycerate mutase-encoding gene. *Gene* **125**: 125-133.

Ronne H (1995) Glucose Repression in Fungi. *Trends in Genetics* **11**: 12-17.

Rosenfeld E & Beauvoit B (2003) Role of the non-respiratory pathways in the utilization of molecular oxygen by Saccharomyces cerevisiae. *Yeast* **20**: 1115-1144.

S. I-SCMSRSK (2006) Acetate kinase: not just a bacterial enzyme. *TRENDS in Micoriobology* **14**: 249-253.
Salih E (2005) Phosphoproteomics by mass spectrometry and classical protein chemistry approaches. *Mass Spectrom Rev* **24**: 828-846.

Santangelo GM (2006) Glucose signaling in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **70**: 253-282.

Santiago TC, Bettany AJ, Purvis IJ & Brown AJ (1987) Messenger RNA stability in Saccharomyces cerevisiae: the influence of translation and poly(A) tail length. *Nucleic Acids Res* **15**: 2417-2429.

Schaaff I, Heinisch J & Zimmermann FK (1989) Overproduction of glycolytic enzymes in yeast. *Yeast* **5**: 285-290.

Schaaff-Gerstenschlager I & Zimmermann FK (1993) Pentose-phosphate pathway in Saccharomyces cerevisiae: analysis of deletion mutants for transketolase, transaldolase, and glucose 6-phosphate dehydrogenase. *Curr Genet* **24**: 373-376.

Scheffer.Wa (1966) Stimulation of Fermentation in Yeasts by Acetoin and Oxygen. *Nature* **210**: 533-&.

Scheffler IE, de la Cruz BJ & Prieto S (1998) Control of mRNA turnover as a mechanism of glucose repression in Saccharomyces cerevisiae. *Int J Biochem Cell Biol* **30**: 1175-1193.

Schuller HJ (2003) Transcriptional control of nonfermentative metabolism in the yeast Saccharomyces cerevisiae. *Curr Genet* **43**: 139-160.

Sonderegger M, Schumperli M & Sauer U (2004) Metabolic engineering of a phosphoketolase pathway for pentose catabolism in Saccharomyces cerevisiae. *Appl Environ Microbiol* **70**: 2892-2897.

Soontorngun N, Larochelle M, Drouin S, Robert F & Turcotte B (2007) Regulation of gluconeogenesis in Saccharomyces cerevisiae is mediated by activator and repressor functions of Rds2. *Mol Cell Biol* **27**: 7895-7905.

Stephanopoulos G, Aristidou AA & Nielsen JH (1998) *Metabolic engineering : principles and methodologies*. Academic Press, San Diego.

Strijbis K & Distel B (2010) Intracellular acetyl unit transport in fungal carbon metabolism. *Eukaryot Cell* **9**: 1809-1815.

Tachibana C, Yoo JY, Tagne JB, Kacherovsky N, Lee TI & Young ET (2005) Combined global localization analysis and transcriptome data identify genes that are directly coregulated by Adr1 and Cat8. *Mol Cell Biol* **25**: 2138-2146.

ter Kuile BH & Westerhoff HV (2001) Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett* **500**: 169-171.

Thomas D, Cherest H & Surdin-Kerjan Y (1991) Identification of the structural gene for glucose-6-phosphate dehydrogenase in yeast. Inactivation leads to a nutritional requirement for organic sulfur. *EMBO J* **10**: 547-553.

Thykaer J & Nielsen J (2007) Evidence, through C13-labelling analysis, of phosphoketolase activity in fungi. *Process Biochemistry* **42**: 1050-1055.

Tornow J, Zeng X, Gao W & Santangelo GM (1993) GCR1, a transcriptional activator in Saccharomyces cerevisiae, complexes with RAP1 and can function without its DNA binding domain. *EMBO J* **12**: 2431-2437.

Treitel MA & Carlson M (1995) Repression by SSN6-TUP1 is directed by MIG1, a repressor/activator protein. *Proc Natl Acad Sci U S A* **92**: 3132-3136.

Treitel MA, Kuchin S & Carlson M (1998) Snf1 protein kinase regulates phosphorylation of the Mig1 repressor in Saccharomyces cerevisiae. *Mol Cell Biol* **18**: 6273-6280.

Turcotte B, Liang XB, Robert F & Soontorngun N (2010) Transcriptional regulation of nonfermentable carbon utilization in budding yeast. *FEMS Yeast Res* **10**: 2-13.

Usaite R, Jewett MC, Oliveira AP, Yates JR, 3rd, Olsson L & Nielsen J (2009) Reconstruction of the yeast Snf1 kinase regulatory network reveals its role as a global energy regulator. *Mol Syst Biol* **5**: 319.

van den Brink J, Canelas AB, van Gulik WM, Pronk JT, Heijnen JJ, de Winde JH & Daran-Lapujade P (2008) Dynamics of glycolytic regulation during adaptation of Saccharomyces cerevisiae to fermentative metabolism. *Appl Environ Microbiol* **74**: 5710-5723.

van Dijken JP, Weusthuis RA & Pronk JT (1993) Kinetics of growth and sugar consumption in yeasts. *Antonie Van Leeuwenhoek* **63**: 343-352.

van Maris AJ, Bakker BM, Brandt M, *et al.* (2001) Modulating the distribution of fluxes among respiration and fermentation by overexpression of HAP4 in Saccharomyces cerevisiae. *FEMS Yeast Res* **1**: 139-149.

van Maris AJ, Geertman JM, Vermeulen A, *et al.* (2004) Directed evolution of pyruvate decarboxylase-negative Saccharomyces cerevisiae, yielding a C2-independent, glucose-tolerant, and pyruvate-hyperproducing yeast. *Appl Environ Microbiol* **70**: 159-166.

Van Urk H, Voll WS, Scheffers WA & Van Dijken JP (1990) Transient-state analysis of metabolic fluxes in crabtree-positive and crabtree-negative yeasts. *Appl Environ Microbiol* **56**: 281-287.

Van Vleet JH & Jeffries TW (2009) Yeast metabolic engineering for hemicellulosic ethanol production. *Curr Opin Biotechnol* **20**: 300-306.

Vanurk H, Mak PR, Scheffers WA & Vandijken JP (1988) Metabolic Responses of Saccharomyces-Cerevisiae Cbs 8066 and Candida-Utilis Cbs 621 Upon Transition from Glucose Limitation to Glucose Excess. *Yeast* **4**: 283-291.

Vemuri GN, Eiteman MA, McEwen JE, Olsson L & Nielsen J (2007) Increasing NADH oxidation reduces overflow metabolism in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **104**: 2402-2407.

Verduyn C, Stouthamer AH, Scheffers WA & Vandijken JP (1991) A Theoretical Evaluation of Growth Yields of Yeasts. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **59**: 49-63.

Verduyn C, Postma E, Scheffers WA & Van Dijken JP (1992) Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* **8**: 501-517.

Villas-Boas SG, Rasmussen S & Lane GA (2005) Metabolomics or metabolite profiles? *Trends Biotechnol* **23**: 385-386.

Villas-Boas SG, Moxley JF, Akesson M, Stephanopoulos G & Nielsen J (2005) High-throughput metabolic state analysis: the missing link in integrated functional genomics of yeasts. *Biochem J* **388**: 669-677.

Villas-Boas SG, Noel S, Lane GA, Attwood G & Cookson A (2006) Extracellular metabolomics: a metabolic footprinting approach to assess fiber degradation in complex media. *Anal Biochem* **349**: 297-305.

Walker GM (1998) *Yeast physiology and biotechnology*. J. Wiley & Sons, Chichester ; New York.

Wang Y, Pierce M, Schneper L, Guldal CG, Zhang X, Tavazoie S & Broach JR (2004) Ras and Gpa2 mediate one branch of a redundant glucose signaling pathway in yeast. *PLoS Biol* **2**: E128.

Weirich J, Goffrini P, Kuger P, Ferrero I & Breunig KD (1997) Influence of mutations in hexose-transporter genes on glucose repression in Kluyveromyces lactis. *European Journal of Biochemistry* **249**: 248-257.

Westergaard SL, Oliveira AP, Bro C, Olsson L & Nielsen J (2007) A systems biology approach to study glucose repression in the yeast Saccharomyces cerevisiae. *Biotechnol Bioeng* **96**: 134-145.

Weusthuis RA, Pronk JT, van den Broek PJ & van Dijken JP (1994) Chemostat cultivation as a tool for studies on sugar transport in yeasts. *Microbiol Rev* **58**: 616-630.

Wiechert W (2001) 13C metabolic flux analysis. *Metab Eng* **3**: 195-206.

Wiechert W & de Graaf AA (1996) In vivo stationary flux analysis by 13C labeling experiments. *Adv Biochem Eng Biotechnol* **54**: 109-154.

Wiechert W, Siefke C, de Graaf AA & Marx A (1997) Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis. *Biotechnol Bioeng* **55**: 118-135.

Wiechert W, Mollney M, Petersen S & de Graaf AA (2001) A universal framework for 13C metabolic flux analysis. *Metab Eng* **3**: 265-283.

Wittmann C (2007) Fluxome analysis using GC-MS. *Microb Cell Fact* **6**: 6.

Wyman CE (2000) Process engineering for new biocommodity technologies. *Abstracts of Papers of the American Chemical Society* **219**: U524-U524.

Ye L, Kruckeberg AL, Berden JA & van Dam K (1999) Growth and glucose repression are controlled by glucose transport in Saccharomyces cerevisiae cells containing only one glucose transporter. *Journal of Bacteriology* **181**: 4673-4675.

Yeh LT, Charles AL, Ho CT & Huang TC (2009) A novel bread making process using salt-stressed Baker's yeast. *J Food Sci* **74**: S399-402.

Yin ZK, Wilson S, Hauser NC, Tournu H, Hoheisel JD & Brown AJP (2003) Glucose triggers different global responses in yeast, depending on the strength of the signal, and transiently stabilizes ribosomal protein mRNAs. *Mol Microbiol* **48**: 713-724.

Young ET, Dombek KM, Tachibana C & Ideker T (2003) Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. *J Biol Chem* **278**: 26146-26158.

Yuan T, Ren Y, Meng K, *et al.* (2011) RNA-Seq of the xylose-fermenting yeast Scheffersomyces stipitis cultivated in glucose or xylose. *Appl Microbiol Biotechnol* **92**: 1237-1249.

Yun DF & Sherman F (1995) Initiation of translation can occur only in a restricted region of the CYC1 mRNA of Saccharomyces cerevisiae. *Mol Cell Biol* **15**: 1021-1033.

Zelle RM, Trueheart J, Harrison JC, Pronk JT & van Maris AJ (2010) Phosphoenolpyruvate carboxykinase as the sole anaplerotic enzyme in Saccharomyces cerevisiae. *Appl Environ Microbiol* **76**: 5383-5389.

Zitomer RS, Montgomery DL, Nichols DL & Hall BD (1979) Transcriptional regulation of the yeast cytochrome c gene. *Proc Natl Acad Sci U S A* **76**: 3627-3631.

APPENDIX 1

% Glucose uptake

v1: GLC <-> G6P
abcdef <-> abcdef

% EMP-pathway
v2: G6P <-> F6P
abcdef <-> abcdef

v3: F6P <-> G3P + G3P
abcdef <-> cba + def

v4: G3P <-> PEP
abc <-> abc

v5: PEP <-> PYRCYT
abc <-> abc

% PP-pathway and Phosphoketolase

v6: G6P <-> CO + P5P
abcdef <-> a + bcdef

vx1: P5P <-> G3P + ACE
abcde <-> abc + de

v7: P5P + P5P <-> S7P + G3P
abcde + fghij <-> fgabcde + hij

v8: S7P + G3P <-> F6P + E4P
abcdefg + hij <-> abchij + defg

v9: P5P + E4P <-> F6P + G3P
abcde + fghi <-> abfghi + cde

% Formation of AcCoA in the cytosol

v14: ACE <-> ACCOACYT
ab <-> ab

% Anaplerotic reaction (cytosolic)

v15: PYRCYT + CO <-> OAACYT
abc + d <-> abcd

% TCA-cycle

v16: PYRMIT <-> ACCOAMIT + CO
abc <-> bc + a

v17: OAAMIT + ACCOAMIT <-> ICIT
abcd + ef <-> dcbafe

v18: ICIT <-> AKG + CO
abcdef <-> abcef + d

v19: AKG <-> FUM + CO
abcde <-> bcde + a

% variable scrambling fraction

v20: FUM + FUM <-> OAAMIT + OAAMIT
abcd + efgh <-> abcd + hgfe

v21: OAAMIT <-> FUM
abcd <-> abcd

% Transports

v22: OAAMIT <-> OAACYT
abcd <-> abcd

v24: PYRCYT <-> PYRMIT
abc <-> abc

% Threonine / Serine / Glycine metabolism (all enzymes assumed to be cytosolic)

v25: G3P <-> SER
abc <-> abc

v26: SER <-> GLY + C1
abc <-> ab + c

v27: OAACYT <-> THR
abcd <-> abcd

v28: THR <-> GLY + ACA
abcd <-> ab + cd

% Malic enzyme (oxaloacetate decarboxylation, mitochondrial)

v29: OAAMIT <-> PYRMIT + CO
abcd <-> abc + d

% Drain of intermediates to macromolecules

v30: G6P <-> G6POUT
abcdef <-> abcdef

v31: P5P <-> P5POUT
abcde <-> abcde

v32: E4P <-> E4POUT
abcd <-> abcd

v33: G3P <-> G3POUT
abc <-> abc

v34: PEP <-> PEPOUT
abc <-> abc

v35: PYRMIT <-> PYRMITOUT
abc <-> abc

v36: PYRCYT <-> PYRCYTOUT
abc <-> abc

v37: OAACYT <-> OAACYTOUT
abcd <-> abcd

v38: AKG <-> AKGOUT
abcde <-> abcde

v39: ACCOACYT <-> ACCOACYTOUT
ab <-> ab

v40: ACCOAMIT <-> ACCOAMITOUT
ab <-> ab

v41: SER <-> SEROUT
abc <-> abc

v42: GLY <-> GLYOUT
ab <-> ab

v43: C1 <-> C1OUT
a <-> a

v44: THR <-> THROUT
abcd <-> abcd

net 0.0 200.0; exch = 0.0;

% CO_2-evolution

v48: CO_2 <-> CO_2OUT
a <-> a