

## Industrial Systems Biology and Metabolic Engineering of *Saccharomyces cerevisiae* A case study in succinic acid production

**JOSÉ MANUEL OTERO ROMERO**

*Department of Chemical and Biological Engineering  
CHALMERS UNIVERSITY OF TECHNOLOGY  
Göteborg, Sweden 2009*



THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Industrial Systems Biology and Metabolic Engineering of  
*Saccharomyces cerevisiae*  
A case study in succinic acid production

**JOSÉ MANUEL OTERO ROMERO**

Systems Biology  
Department of Chemical and Biological Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY  
Göteborg, Sweden 2009

Industrial Systems Biology and Metabolic Engineering of *Saccharomyces cerevisiae*  
A case study in succinic acid production

José Manuel Otero Romero

© JOSÉ MANUEL OTERO ROMERO, 2009.

ISBN 978-91-7385-346-0

Doktorsavhandlingar vid Chalmers tekniska högskola  
Ny serie nr 3027  
ISSN 0346-718X

PhD Thesis  
Systems Biology  
Department of Chemical and Biological Engineering  
Chalmers University of Technology  
SE-412 96 Göteborg  
Sweden  
Telephone + 46 (0)31-772 1000

Primary Funding Source:  
Merck Research Labs, Merck & Co., Inc  
770 Sumneytown Pike, West Point, PA 19486  
United States of America

Cover: Technical scheme of the integrated metabolic engineering cycle and systems biology approach prepared by José Manuel Otero Romero as part of this research

Printed by Chalmers Reproservice  
Göteborg, Sweden 2009

Dedicated to

Stephanie Praster Otero, the love of my life...

José Manuel Otero Sieira and Joaquina Otero Romero, the unwavering support of my life...

Melissa Anne Otero Romero, Psy. D., the best friend of my life...

*When, in 1930, I left the laboratory of Otto Warburg, I was confronted with the question of selecting a major field of study and I felt greatly attracted by the problem of the intermediary pathway of oxidations. These reactions represent the main energy source in higher organisms, and in view of the importance of energy production to living organisms (whose activities all depend on a continuous supply of energy) the problem seemed well worthwhile studying.*

-Hans A. Krebs  
*Nobel Lecture, December 11, 1953*

# Industrial Systems Biology and Metabolic Engineering of *Saccharomyces cerevisiae* A case study in succinic acid production

JOSÉ MANUEL OTERO ROMERO

Systems Biology  
Department of Chemical and Biological Engineering  
Chalmers University of Technology

## ABSTRACT

*Saccharomyces cerevisiae* is the most well characterized eukaryote, the preferred microbial cell factory for the largest industrial biotechnology product (bioethanol), and a robust commercially compatible scaffold to be exploited for diverse chemical production. Succinic acid is a highly sought after added-value chemical which is not overproduced in native *S. cerevisiae* strains. The genome-scale metabolic network reconstruction of *S. cerevisiae* enabled *in silico* gene deletion predictions. First, a multi-gene, non-intuitive, genetic engineering strategy guided by an evolutionary programming method to couple biomass formation through glycine-serine amino acid requirements to succinate production was proposed. Pursuing these targets, a multi-gene deletion strain was constructed, and directed evolution with selection was used to identify a succinate producing mutant. Physiological characterization coupled with integrated data analysis of transcriptome data in the metabolically engineered strain were used to identify 2<sup>nd</sup>-round metabolic engineering targets – overexpression of *ICL1*. The resulting strain represents a 30-fold improvement in succinate titer, and a 43-fold improvement in succinate yield on biomass, with only a 2.8-fold decrease in the specific growth rate compared to the reference strain. Further genome-scale metabolic modeling supplemented with pathway visualization, flux balance analysis, and model modifications to better simulate batch glucose conditions was performed. Identification of the top single and double gene deletion strategies, under aerobic and anaerobic conditions, resulted in three predictions with a 10-fold improvement in succinate yield on glucose compared to the reference: *MDH1*, *OAC1*, and *DIC1*. While  $\Delta mdh1$  and  $\Delta oac1$  strains failed to produce more succinate relative to the reference,  $\Delta dic1$  produced 0.02 C-mol C-mol-glucose<sup>-1</sup>, in close agreement with model predictions (0.03 C-mol C-mol-glucose<sup>-1</sup>). Pathway visualization, coupled with transcriptional profiling, suggested that succinate formation was coupled to mitochondrial redox balancing, and more specifically, reductive TCA cycle activity. The aforementioned metabolic engineering strategies were designed based on glucose supplementation and metabolism. Future *S. cerevisiae* microbial cell factories capable of fast and efficient xylose consumption for biorefinery compatibility, and succinic acid overproduction would be highly desirable. Metabolic engineering of *S. cerevisiae* for consumption of xylose aerobically without redirection of some carbon flux to overflow metabolites (ethanol, glycerol, acetate, xylitol) was accomplished by expression of *PsXYL1*, *PsXYL2*, and *PsXYL3* from the native xylose-metabolizing *Pichia stipitis*, and subsequent, directed evolution. The resulting *S. cerevisiae* strain showed xylose consumption at a specific rate of 0.31 g g-cell<sup>-1</sup> h<sup>-1</sup>, a specific growth rate of 0.18 h<sup>-1</sup>, and a biomass yield of 0.62 C-mol C-mol-xylose<sup>-1</sup>. Plasmid isolation and re-transformation confirmed the conferred phenotype resulted from a chromosomal modification. Transcriptional profiling confirmed a strongly up-regulated glyoxylate pathway enabling sustained respiratory metabolism. A proof-of-concept study was performed to determine if whole high-throughput genome sequencing could be used as a tool in metabolic engineering for direct identification of genotype to phenotype correlations. Therefore, whole genome sequencing of *S. cerevisiae* S288C and CEN.PK113-7D resulted in identification of 13,787 filtered SNPs in CEN.PK113-7D, with a total of 939 SNPs detected across 158 unique metabolic genes, 85 of which contained a total of 219 non-silent SNPs. *S. cerevisiae* CEN.PK113-7D exhibited significantly higher ergosterol content correlating with non-silent SNPs identified in *ERG8* and *ERG9*. The flux through the galactose uptake pathway was much lower in S288C compared with CEN.PK113-7D, correlating with the non-silent SNP enrichment in *GAL1* and *GAL10*. Inspection of the significantly differentially expressed genes between strains did not reveal an obvious gene cluster that would explain the significant physiological differences, strongly suggesting that genotype to phenotype correlation is manifested post-transcriptionally or post-translationally.

Keywords: metabolic engineering, industrial systems biology, *Saccharomyces cerevisiae*, DNA microarrays, industrial biotechnology, succinic acid, xylose, genome sequencing, *in silico* design.

# LIST OF PUBLICATIONS

This thesis is based on the following publications:

- I. Otero JM, Cimini D, Patil KR, Poulsen SG, Olsson L, Nielsen J: **Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory.** Paper submitted to *PLOS One*, 2009.
- II. Otero JM, Ågren R, Nielsen J: **Genome-scale modeling enables metabolic engineering of *Saccharomyces cerevisiae* for succinic acid production.** Paper submitted to *Metabolic Engineering*, 2009.
- III. Otero JM, Scalcinati G, Van Vleet JRH, Jeffries TW, Olsson L, Nielsen J: **Metabolic engineering of *Saccharomyces cerevisiae* for xylose consumption.** Paper submitted to *Microbial Cell Factories*, 2009.
- IV. Otero JM, Vongsangnak W, Asadollahi MA, Olivares-Hernández R, Maury J, Farinelli L, Barlocher L, Østerås M, Schalk M, Clark A, Nielsen J: **Whole genome sequencing of *Saccharomyces cerevisiae*: from genotype to phenotype for improved metabolic engineering applications.** *Manuscript*, 2009.
- V. Otero JM, Nielsen J: **Industrial Systems Biology.** *Biotechnology & Bioengineering* 2009, Paper accepted.

Throughout the course of this doctoral research contributions have been made to additional publications that are not included in this thesis. These publications are listed below:

- VI. Otero JM, Panagiotou G, Olsson L: **Fueling industrial biotechnology growth with bioethanol.** *Adv Biochem Eng Biotechnol* 2007, **108**:1-40.
- VII. Meijer S, Otero J, Olivares R, Andersen MR, Olsson L, Nielsen J: **Overexpression of isocitrate lyase-glyoxylate bypass influence on metabolism in *Aspergillus niger*.** *Metab Eng* 2009, **11**: 107-116.
- VIII. Otero JM, Papadakis E, Udatha DBRKG, Nielsen J, Panagiotou G: **Yeast biological networks unfold the interplay of antioxidants, genome, and phenotype.** *Paper submitted*, 2009.
- IX. Otero JM, Olsson L, Nielsen J: **Industrial biotech meets systems biology: petrochemical industry finding success by turning to biotechnology.** *Genetic Engineering News* 2007, **27**: 28-31.<sup>†</sup>
- X. Olsson L, Otero JM, Patil K, Nielsen J: **Fra petrokemisk til biobaseret ravsyreproduktion.** *Dansk Kemi* 2007, **88**: 24-26.<sup>†</sup>

<sup>†</sup>These papers were not subjected to scientific peer-review, but are invited technical commentaries.

# CONTRIBUTION SUMMARY

A summary of the contributions of José Manuel Otero (JMO) to each of the publications listed in the section **List of Publications** is provided below:

- I. All strain physiological characterization, directed evolution of strains, supervised construction of *ICL1* over-expression mutants, transcriptome measurements and analysis, paper preparation and submission.
- II. All strain physiological characterization, supervised genome-scale modeling, transcriptome measurements and analysis, paper preparation and submission.
- III. Supervised and assisted with strain evolution and physiological characterization, transcriptome analysis, paper preparation and submission. For this paper equal contribution authorship is shared between JMO and GS.
- IV. Genomic DNA isolation, SNP analysis, physiological characterization, transcriptome measurement and analysis, paper preparation and submission.
- V. Paper preparation and submission.
- VI. Paper preparation and submission.
- VII. Identification of isotopomer fragments, and assisted with summed fractional labeling calculation, flux simulations and analysis. Paper preparation related to flux analysis and review.
- VIII. Supervised and assisted with strain construction, physiological characterization, metabolomics profiling, transcriptome measurement and analysis, paper preparation, review, and submission. For this paper equal contribution authorship is shared between JMP and EP.
- IX. Paper preparation and submission.
- X. Contributed data and text to paper preparation.

# ABBREVIATIONS COMMONLY USED

*A. succiniciproducens*: *Anaerobiospirillum succiniciproducens*  
*A. succinogenes*: *Actinobacillus succinogenes*  
bbl: barrel (volumetric unit)  
DNA: Deoxyribonucleic acid  
*E. coli*: *Escherichia coli*  
FAD<sup>+</sup>: Flavin adenine dinucleotide  
FADH<sub>2</sub>: Flavin adenine dinucleotide di-hydrogen  
FBA: Flux balance analysis  
GRAS: Generally Regard As Safe  
GSMM: Genome-scale metabolic model  
Ka: acid dissociation constant  
 $K_{eq}$ : Equilibrium constant, [Products] [Reactants]<sup>-1</sup>  
LO: Lines of optimality  
*M. succiniciproducens*: *Mannheimia succiniciproducens*  
mRNA: messenger ribonucleic acid  
NAD<sup>+</sup>: Nicotinamide adenine dinucleotide  
NADH: Nicotinamide adenine dinucleotide hydrogen  
NADP<sup>+</sup>: Nicotinamide adenine dinucleotide phosphate  
NADPH: Nicotinamide adenine dinucleotide phosphate hydrogen  
ORF: Open reading frame  
PhPP: Phenotypic Phase Plane  
 $pK_a$ :  $-\log_{10} K_a$   
PPP: Pentose Phosphate Pathway  
RNA: Ribonucleic acid  
*S. cerevisiae*: *Saccharomyces cerevisiae* (baker's yeast)  
SNP: Single nucleotide polymorphism  
TCA: Tricarboxylic acid  
US DOE: United States Department of Energy  
US FDA: United States Food and Drug Administration  
USD: United States Dollars (\$)  
 $\Delta G$ : Gibbs free energy (J mol<sup>-1</sup>)  
 $\Delta G^o$ : Standard Gibbs free energy (J mol<sup>-1</sup>)  
 $\mu$ : Specific growth rate (h<sup>-1</sup>)  
 $\mu_{max}$ : Maximum specific growth rate (h<sup>-1</sup>)

## References to Thesis Papers in Summary:

Throughout the course of the summary presented in this thesis, the papers are referred to where appropriate by the designation, **Paper No.**, where the number referenced is in accordance with the roman numerals highlighted in the **List of Publications**. For example, if referring to the final paper, *Industrial Systems Biology*, then in the text will appear “Paper V”.

## Nomenclature:

Standard nomenclature for *S. cerevisiae* is used for designating genes, proteins and gene deletions: *ICL1*, Icl1p and *Δicl1*, respectively, for isocitrate lyase as an example. Deviations from this nomenclature in the publications where required are appropriately noted. Furthermore, the standard nomenclature, [ *i* ], is used to indicate concentration (moles l<sup>-1</sup>) of chemical species *i*.

# TABLE OF CONTENTS

<b>1.0 INDUSTRIAL BIOTECHNOLOGY.....</b>	<b>1</b>
<b>2.0 INDUSTRIAL SYSTEMS BIOLOGY .....</b>	<b>6</b>
<b>3.0 MICROBIAL METABOLISM .....</b>	<b>10</b>
3.1 A HISTORICAL PERSPECTIVE.....	10
<b>4.0 RECONSTRUCTED METABOLIC NETWORK MODELS.....</b>	<b>18</b>
4.1 INTRODUCTION.....	18
4.2 GENOME-SCALE RECONSTRUCTED NETWORK PROCESS.....	19
4.3 INDUSTRIAL SYSTEMS BIOLOGY APPLIED.....	23
<b>5.0 IN-DEVELOPMENT INDUSTRIAL BIOTECHNOLOGY PRODUCT: SUCCINIC ACID .....</b>	<b>24</b>
<b>6.0 <i>S. CEREVISIAE</i> FOR SUCCINIC ACID PRODUCTION .....</b>	<b>33</b>
6.1 MOTIVATIONS DRIVING A SUCCINIC ACID MICROBIAL CELL FACTORY .....	33
6.2 SUCCINIC ACID PRODUCTION IN <i>S. CEREVISIAE</i> – PRIOR WORK .....	35
6.3 TECHNICAL CHALLENGES ANTICIPATED .....	37
6.3.1 <i>Succinate Transport - Mitochondrial vs. Cytosolic</i> .....	37
6.3.2 <i>Redox and Co-Factor Balance</i> .....	38
6.3.3 <i>Cell Growth Regulation - Crabtree Effect</i> .....	40
<b>7.0 SUMMARY OF RESULTS.....</b>	<b>42</b>
7.1 PAPER I: INDUSTRIAL SYSTEMS BIOLOGY OF <i>SACCHAROMYCES CEREVISIAE</i> ENABLES NOVEL SUCCINIC ACID CELL FACTORY .....	42
7.2 PAPER II: GENOME-SCALE MODELING ENABLES METABOLIC ENGINEERING OF <i>SACCHAROMYCES CEREVISIAE</i> FOR SUCCINIC ACID PRODUCTION .....	46
7.3 PAPER III: METABOLIC ENGINEERING OF <i>SACCHAROMYCES CEREVISIAE</i> FOR XYLOSE CONSUMPTION.....	51
7.4 PAPER IV: WHOLE GENOME SEQUENCING OF <i>SACCHAROMYCES CEREVISIAE</i> : FROM GENOTYPE TO PHENOTYPE FOR IMPROVED METABOLIC ENGINEERING APPLICATIONS.....	56
7.5 RESULTS NOT INCLUDED IN PAPERS .....	62
7.5.1 <i>8D and 8D Evolved Carbon Balance</i> .....	62
7.5.2 <i>Carbon Dioxide Supplementation</i> .....	67
7.5.3 <i>Malonate Supplementation to S. cerevisiae CMB.GS010</i> .....	74
<b>8.0 CONCLUSIONS.....</b>	<b>76</b>
<b>9.0 PERSPECTIVES ON FUTURE DEVELOPMENTS IN INDUSTRIAL SYSTEMS BIOLOGY .....</b>	<b>79</b>
<b>10. ACKNOWLEDGEMENTS.....</b>	<b>82</b>
<b>11. REFERENCES.....</b>	<b>85</b>

# TABLE OF FIGURES

FIGURE 1: PRICE DATA FOR KEY CHEMICAL AND BIOTECHNOLOGY PROCESS RAW MATERIALS (OCT. 2005 – OCT. 2009).....	4
FIGURE 2: X-OMIC GLOSSARY .....	7
FIGURE 3: INDUSTRIAL SYSTEMS BIOLOGY .....	8
FIGURE 4: MICROBIAL METABOLISM.....	12
FIGURE 5: SUCCINIC ACID – A SUSTAINABLE BUILDING BLOCK CHEMICAL.....	25
FIGURE 6: CATABOLIC PATHWAY FOR GLUCOSE FERMENTATION BY <i>A. SUCCINICIPRODUCENS</i> AND <i>A. SUCCINOGENES</i> .....	30
FIGURE 7: GENETIC ENGINEERING OF CENTRAL CARBON METABOLISM IN <i>E.COLI</i> FOR SUCCINATE OVERPRODUCTION.....	31
FIGURE 8: CENTRAL CARBON METABOLISM OF <i>M. SUCCINICIPRODUCENS</i> AND <i>E. COLI</i> .....	32
FIGURE 9: ESTIMATED RAW MATERIALS AND OPERATING COSTS OF A BIO-BASED SUCCINIC ACID PRODUCTION PROCESS .....	34
FIGURE 10: <i>S. CEREVISIAE</i> CENTRAL CARBON METABOLISM .....	36
FIGURE 11: PROOF-OF-CONCEPT METABOLIC ENGINEERING FOR SUCCINATE OVERPRODUCTION .....	43
FIGURE 12: SUMMARY OF SUCCINATE MICROBIAL CELL FACTORY PERFORMANCE.....	45
FIGURE 13: EXPERIMENTAL AND SIMULATION COMPARATIVE DATA.....	47
FIGURE 14: EXPERIMENTAL AND SIMULATION COMPARATIVE DATA FOR REFERENCE, $\Delta OAC1$ , $\Delta MDH1$ , AND $\Delta DIC1$ STRAINS.....	49
FIGURE 15: TRANSCRIPTIONAL RESPONSE OF EVOLVED AND UNEVOLVED STRAINS.....	53
FIGURE 16: STRAIN CONSTRUCTION AND PLASMID RESCUE.....	55
FIGURE 17: PHYSIOLOGICAL CHARACTERIZATION OF <i>S. CEREVISIAE</i> S288C AND CEN.PK113- 7D .....	57
FIGURE 18: ERGOSTEROL MEASUREMETNS IN <i>S. CEREVISIAE</i> STRAINS S288C AND CEN.PK113- 7D .....	58
FIGURE 19: SNP ENRICHMENT IN <i>S. CEREVISIAE</i> METABOLISM.....	59
FIGURE 20: TRANSCRIPTOME RESPONSE TO GLUCOSE AND GALACTOSE OF <i>S. CEREVISIAE</i> STRAINS S288C AND CEN.PK113-7D .....	61
FIGURE 21: THEORETICAL ESTIMATION OF END OF FERMENTATION ACETALDEHYDE CONCENTRATION.....	64
FIGURE 22: SIGNIFICANT DIFFERNETIAL GENE EXPRESSION (8D EVOLVED VS. CEN.PK113-5D) .....	66
FIGURE 23: CARBOXYLATION REACTIONS LEADS TO SUCCINATE FORMATION .....	67
FIGURE 24: TABLE 1 FROM AGUILERA ET AL, 2005, SHOWING PHYSIOLOGICAL EFFECTS OF ELEVATED $\text{CO}_2$ IN FERMENTATION.....	68
FIGURE 25: MAXIMUM SPECIFIC GROWTH WITH DIFFERENT $\text{CO}_2$ SUPPLEMENTATION.....	69
FIGURE 26: RELATIVE YIELDS ON SUBSTRATE OF CEN.PK113-5D SUPPLEMENTED WITH $\text{CO}_2$ .....	71
FIGURE 27: RELATIVE YIELDS ON SUBSTRATE OF 8D EVOLVED SUPPLEMENTED WITH $\text{CO}_2$ .....	72
FIGURE 28: THERMODYANMIC ESTIMATION OF $\Delta G$ ( $\text{J mol}^{-1}$ ) OF SUCCINATE SPECIFIC ENZYMES .....	73
FIGURE 29: PHYSIOLOGICAL EFFECTS OF MALONATE SUPPLEMENTATION.....	75
FIGURE 30: INDUSTRIAL SYSTEMS BIOLOGY PIPELINE .....	80

## TABLE OF TABLES

TABLE 1: KEY SUCCINIC ACID MARKETS .....	27
TABLE 2: SUMMARY OF MAJOR MICROBIAL PRODUCTION PLATFORMS OF SUCCINIC ACID.....	29
TABLE 3: RELATIVE EFFECT ON THE MAXIMUM SPECIFIC GROWTH RATE FOR DIFFERENT CO <sub>2</sub> SUPPLEMENTATION CONCENTRATIONS COMPARED TO NO SUPPLEMENTATION.....	70

# Summary

## 1.0 Industrial Biotechnology

The term “industrial biotechnology” first widely appeared in the literature in the early 1980s when genetic engineering enabled by recombinant DNA technology was searching for applications beyond health care and biomedical research (Pass, 1981; Ferrández-García, 1982). Industrial biotechnology has emerged as a well-defined field with academic, state, and corporate participation. Formally, industrial biotechnology is the bioconversion, either through microbial fermentation or biocatalysis, of organic feedstocks extracted from biomass or their derivatives to chemicals, materials, and/or energy. Industrial biotechnology, often referred to as white biotechnology in Europe (Maury et al, 2005), aims to provide cost-competitive, environmentally friendly, self-sustaining alternatives to existing or newly proposed petrochemical processes. Processes that exploit industrial biotechnology have recently garnered increasing global attention with traditional petrochemical processing under scrutiny due to increasing raw material costs, environmental constraints, and decreasing self-sufficiency.

According to the Chemical & Engineering News survey of the top fifty chemical manufacturing corporations world-wide, including traditional petroleum companies with a chemical manufacturing business segment, the combined total sales in 2005 was \$665.6 billion, representing a 15% increase since 2004. Most recently, the A.D. Little consulting firm estimates the current, 2009, industrial biotechnology chemical market to be between \$70-\$100 billion, or 3-4% of global chemical sales, suggesting that 2009 chemical sales are estimated to be between \$2.3 to \$3.3 trillion (Reisch, 2009). The global scope of the sector is reaffirmed by the distribution of total sales across the United States, Europe, and Japan, of 29.3% (\$195.3 billion), 45.1% (\$300.5 billion), and 12.8% (\$86.9 billion), respectively, representing the top three geographical chemical sectors. For the previous five years (2000-2004), Dow Chemical (Midland, MI, USA), BASF (Ludwigshafen, Germany), and DuPont (Wilmington, DE, USA) have been the three largest chemical companies (based on total sales); however, in 2005 DuPont fell to the sixth position with Royal Dutch Shell (The Hague, The Netherlands), ExxonMobil (Irving, TX, USA), and TOTAL (Courbevoie, France) moving into the third, fourth, and fifth positions, respectively (Short, 2006). The fact that these three companies are traditional petroleum refiners is a direct reflection of the increasing crude petroleum prices resulting in significant price increases in chemical raw material feedstocks. For example, a 2005 corporate press release from DuPont noted the impact that increasing raw material prices have on the chemical industry:

*Higher crude oil prices have a global impact on fuel and feedstock costs. Higher prices for natural gas have an especially severe impact on costs for the North American chemical industry, which is highly dependent on natural gas as feedstock, while the rest of the world relies more heavily on oil derivatives. A \$10 increase in the price of a barrel of oil increases variable costs to the US chemical industry by about \$2.6 billion per year. (This includes fuel, power, and feedstock costs.) For natural gas a \$1 increase per mmbtu increases variable costs to the US chemical industry by \$3.7 billion per year (Energy Impact and Implications for Pricing, 2005).*

Dow Chemical and BASF were able to retain their top positions in large part to the fact that they have a substantial hydrocarbon and energy business sector (i.e., world leader in olefins and aromatics), and a major oil and gas production business sector, respectively, complementing their chemical business at times with increasing crude petroleum prices (Dow Corporate Report, 2005; BASF Financial Report, 2005).

Between January 1978 and January 2001, the average world cost of petroleum increased 69% from \$13/bbl<sup>1</sup> to \$22/bbl, while between January 2001 and January 2006, the cost increased 150% from \$22/bbl to \$55/bbl (Energy Information Administration, 2009). Significant increases in crude petroleum prices were fueled by increased energy demands, and limitations or threats to supply. Global energy consumption is projected to increase by 57% from 435 trillion MJ in 2002 to 681 trillion MJ in 2025 (D'Aquino, 2006). Global petroleum demand in 2004 was 82 million bbl/day, with a projected increase to 111 million bbl/day in 2025 (Petroleum Marketing Annual 2005, 2006). Furthermore, energy demands will continue to increase significantly as the emerging economies of China, India, and Russia continue to rapidly expand, reporting 2005 GDP growth rates of 9.9, 7.6, and 6.4%, respectively, compared to the average world GDP growth rate of 4.7% (World Fact Book, 2006).

Is the recent price increase in crude petroleum fueling the renewed interest in industrial biotechnology as a means of developing sustainable, cost-effective, and environmentally favorable processes? Numerous peer-review literature reports have outlined the role that industrial biotechnology can likely serve to exploit the benefits previously mentioned (Gavrilescu et al, 2005; Ragauskas et al, 2006; Herrera, 2004; Schubert, 2006). Furthermore, the A.D Little consulting firm suggests that the biotechnology chemical market could increase sevenfold by 2025 from today's levels, contributing about 17% of global chemical sales (Reisch, 2009). However, perhaps more convincing, is a review of the 2005 corporate annual reports of the six largest petro-chemical manufacturing companies, where five of the six clearly note increasing prices of raw materials and feedstocks have required industrial biotechnology to become a core position of their business units (quoted directly from the 2005 annual report of the indicated company):

- Royal Dutch Shell: *Shell is the world's largest marketer of biofuels and a leading developer of advanced biofuels technologies. During 2005, we entered a partnership with CHOREN Industries GmbH which will work towards the construction of the world's first commercial facility to convert biomass into high quality synthetic biofuel. This is in addition to our existing partnership with Iogen which is producing cellulose ethanol in Canada from plant waste. We are now working with Iogen and Volkswagen on a joint study to assess the economic feasibility of producing cellulose ethanol in Germany. These advanced biofuels can be used in today's cars and can cut carbon dioxide emissions by 90% compared with conventional fuels* (Royal Dutch Shell Corporate Report, 2005).
- DuPont: *In 2005, DuPont announced the creation of its newest Technology Platform, DuPont Bio-Based Materials. However, the first revolutionary products have already entered the marketplace. DuPont™ Sorona® is an innovative new polymer made with 1,3-propanediol (PDO). While PDO is currently made using a petroleum-based process, DuPont developed a way to make PDO from corn — a renewable resource — instead of petroleum. DuPont partnered with UK-based Tate & Lyle PLC to build the world's largest aerobic fermentation facility to produce Bio-PDO™ from corn. The facility, under construction in Loudon, Tennessee, will begin operation in late 2006. Production of Bio-PDO™ will consume 30 to 40 percent less energy per pound than petroleum-based PDO. So the production of 100 million pounds of bio-based material at the Loudon plant will save the equivalent of 10 million gallons of gasoline annually* (DuPont Annual Review, 2005).

---

<sup>1</sup> The unit *bbl* is an abbreviation for barrel, a common unit of measurement for petroleum equivalent to 42 US gallons or approximately 159 liters.

- Dow Chemical Company: *2005 sales rose to a new high of \$46.3 billion; however, feedstock and energy costs were approximately \$4 billion – representing almost 10% of total sales* (Dow Corporate Report, 2005).
- BASF: *We have combined the important technology-driven issues of the future in five growth clusters: energy management, raw material change, nanotechnology, plant biotechnology and white (industrial biotechnology)... By expanding white biotechnology, we aim to use our expertise in the areas of enzyme catalysis and fermentative manufacturing processes to develop new products and processes outside the current key areas of fine chemicals and intermediates* (BASF Financial Report, 2005).
- TOTAL: *TOTAL's efforts to expand its activities in the field of renewable energies are in line with our desire to prepare for the future of energy and to foster sustainable development. As regards biofuels, TOTAL already produces 170,000 metric tons per year of ETBE and our long-term diester supply contracts will be increasing strongly in the coming years, in conjunction with the food-processing industry. The Group is also launching research programs on second-generation biofuels. Within a few years, industry should be able to produce biofuels from a wide variety of biomass sources* (TOTAL Global Report, 2005).

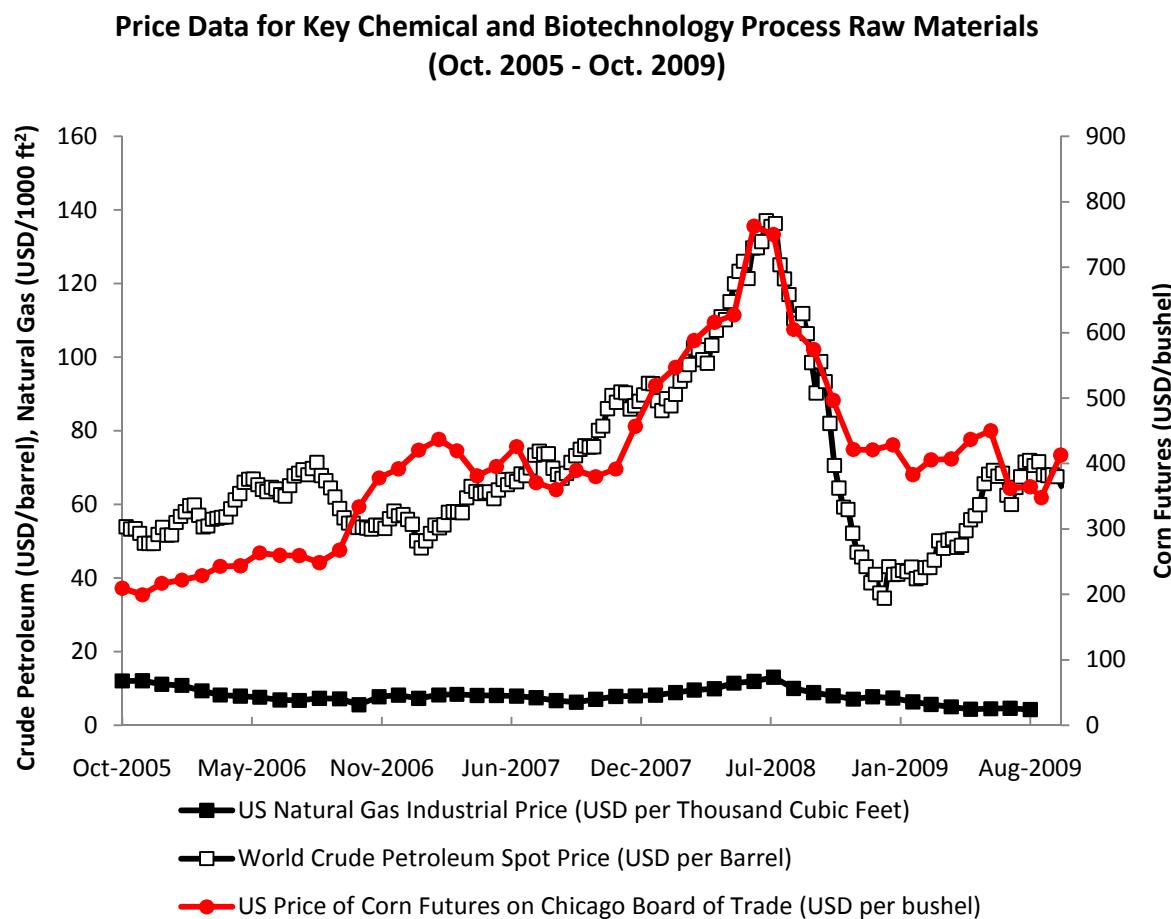
In the United States in 2004, approximately 6-8% of all petroleum supplies were consumed by the chemical industry for manufacturing, primarily in the form of natural gas, naphtha, and refinery gases (ethane, propane, and butane). The largest consumer of petroleum, as a sector, are transportation fuels forcing chemical companies to compete for refined petroleum products. Although academic centers and government research offices have long advocated and supported research in industrial biotechnology, commercialization will only be possible with the financial commitment of industry, beginning with the largest petrochemical manufacturing companies.

When a new industrial process is brought online, irrespective of the product, scale, or technology, there are inefficiencies and short-term operating and capital costs associated with start-up. As the process matures, and efficiencies are gained or introduced via subsequent process upgrades, the raw material cost as a percentage of the total operating cost increases across the life-time of the product, until it represents the largest cost fraction. This is particularly true for commodities and large volume products where profit margins are generally much lower relative to fine chemicals or pharmaceuticals, and hence, fine chemicals and pharmaceuticals have witnessed the largest penetration of industrial biotechnology (Hirche, 2006; Gavilescu, 2005). Therefore, the decision to develop bio-based process alternatives is highly dependent on the raw material cost, which in most cases is petroleum vs. biomass, either on a per mass or energy unit basis.

During the same time period that the cost of crude petroleum rose 150%, from January 2001 to 2005, the total number of bioethanol refineries in the US increased from 56 to 81, with total production capacity increasing from 6.6 billion liters per year to 13.8 billion liters per year. From January 2005 to 2006, the total number of refineries increased to 95 and output further increased to 14.3 billion liters per year – a greater than 200% and 150% increase since 2001, respectively. Total world production in 2005 was 46 billion liters, with the US and Brazil representing a combined 70% of the world's production. It should be further noted that by the end of 2005, 29 ethanol refineries and nine expansions of existing refineries were under construction, with a combined annual capacity of 5.7 billion liters. If you consider all of the US ethanol production capacity currently on-line, under expansion, and under construction, then the projected capacity is approximately 24 billion liters – approximately 85% of that required by the Renewable Fuels Standard by 2012 (From Niche to Nation: Ethanol Industry Outlook, 2006).

In October 2005, the doctoral research summarized and presented in this thesis was initiated and conceived in the context of the challenges aforementioned in petrochemical process

development. In 2005, bioethanol represented the largest industrial biotechnology product by volumetric production, with 46 billion liters produced world-wide. During the course of my PhD project, the economic, environmental, and socio-political pressures that initially drove industrial biotechnology process development have continued to accelerate the field. This is evidenced in part by the fact that in 2008 bioethanol production increased to 66 billion liters world-wide – a 43% increase since 2005. In 2005, there was little mention of cellulosic ethanol production, with the Iogen Corporation in Ottawa, Canada, highlighted for producing just over 3.7 million liters annually from wheat, oat and barely straw in a demonstration facility. Presently, there are 19 separate major cellulosic ethanol projects under development in the US alone, with a total estimated capacity of 1.5 billion liters annually, using feedstocks that include corn stover, wheat straw, switchgrass, grass seed, cellulosic urban wastes, sugarcane, softwood chips, wood, and poplar trees (Growing Innovation: Ethanol Industry Outlook, 2009).



**Figure 1: Price Data for Key Chemical and Biotechnology Process Raw Materials (Oct. 2005 – Oct. 2009).** The price data for US natural gas (USD/1000 ft<sup>3</sup>), world crude petroleum (USD/barrel), and US corn futures (USD/bushel) are presented from October 2005 to October 2009. The natural gas and crude petroleum prices are based on the archives of the US Energy Information Administration (Energy Information Administration, 2009), and the corn futures prices are based on the archives of the Chicago Board of Trades (Chicago Board-Trade Library, 2009).

Between October 2005 and October 2009, the raw material prices and stability of the key petrochemical process raw materials, crude petroleum and natural gas, are presented in Figure 1. Furthermore, the price of a key biotechnology process raw material, corn, is also presented in Figure 1. As previously discussed, there has been significant price instability in crude petroleum

and natural gas; however, as the possibilities of using corn feedstocks for biofuel production accelerated, so did the price of corn. Corn is a primary feedstock integrated into the world's food chain supply thereby any increase in the price of corn is directly passed onto the consumer. There have been numerous reviews addressing the global concern that industrial biotechnology process alternatives are driving the cost of the world's food supplies unsustainably higher (Kamm et al, 2007; Robertson et al, 2008; Octave et al, 2009). The largest industrial biotechnology product, bioethanol, continues to proliferate with increased presence of the enabling technologies of metabolic engineering and systems biology to specifically promote 2<sup>nd</sup> generation processes that utilize lignocellulose and other more desirable feedstocks (e.g., non-food). These disciplines have been applied to other chemical products as well for successful commercialization, suggesting a graduation of the field to *industrial systems biology* (Otero et al, 2007 and **Paper V**).

Industrial systems biology, more than ever, will be required to develop industrial biotechnology processes that not only depart from fossil fuel based feedstocks, but also carefully weigh the impact of using different sources of biomass (e.g., corn vs. lignocellulose). **Paper III** directly addresses the use of systems biology approaches to enable *S. cerevisiae* consumption of a lignocellulosic derived substrate, xylose. The matrix of different biomass sources and target chemical compounds for production will require advanced approaches and tools for developing microbial cell factories quickly. This thesis has demonstrated the use of several systems biology technologies integrated with microbial physiological characterization to ultimately enable and enhance metabolic engineering (**Papers I, II, III, and IV**). It is my vision that biorefineries will only be possible through the development of core platform microbial cell factories that can quickly be exchanged in and out of a processing facility based on the market demand. It is my hope and belief that the results and conclusions of this thesis facilitate in transforming that vision into an industrial reality.

## 2.0 Industrial Systems Biology

Systems biology is the quantitative collection, analysis, and integration of whole genome scale data sets enabling biologically relevant and often predictive mathematical models to be constructed. With genome sequences becoming readily available for production organisms, process development has been a benefactor of the scientific achievements in systems biology, particularly in the areas of transcriptomics, proteomics, metabolomics and fluxomics. Such developments today encompass a systems biology toolbox that may be further exploited for production of metabolic intermediates that often serve as desirable precursors in the petrochemical sector. Given the many definitions and extensive nomenclature that has evolved in systems biology, a glossary of *X-omic* terminology is highlighted in Figure 2.

*S. cerevisiae* today is the preferred industrial biotechnology production host, primarily as a result of proven industrial process robustness and exceptional physiological and *x-omics* characterization (Dien et al, 2003; Russo et al, 1995; Aldrio et al, 2006; Fisk et al, 2006). The *S. cerevisiae* genome sequence, consisting of 6,607 total open reading frames (4,845 verified; 951 uncharacterized; 811 dubious) (Fisk et al, 2006), was first made publicly available in 1996 largely through André Goffeau's coordination of the European yeast research community (Goffeau et al, 1996). Soon thereafter, in 1997 and 1998, the first cDNA spotted microarray exploring metabolic gene regulation, and the first commercial platform (Affymetrix) oligonucleotide microarray data exploring mitotic cell regulation were reported, respectively (DeRisi et al, 1997; Cho et al, 1998). The genome sequence coupled with extensive annotation based on fundamental biochemistry, peer-review literature, and available transcription data enabled publication of the first genome-scale metabolic model for *S. cerevisiae* in 2003 (Fürster et al, 2003). The genome-scale metabolic model represents an integration of extensive amounts of data into an annotated, defined, and uniform format permitting simulations of engineered genotypes to elicit desired phenotypes (Fürster et al, 2003; Famili et al, 2003).

Strain development has classically been dominated by random mutagenesis, facilitated by chemical mutagens and radiation, of a production host followed by screening and selection in controlled environments for a desired phenotype. Although this methodology has endured tremendous success, it has largely been end-product driven with minimal mechanistic understanding. Today, with the exponential increase in genome sequences of existing and future production hosts, coupled with tools from bioinformatics that enable integration and interrogation of *x-omic* data sets, it is possible to identify high-probability targeted genetic strategies to increase yield, titer, productivity, and/or robustness (Vemuri et al, 2005; Patil et al, 2004). It is also now possible to perform inverse metabolic engineering, where previously successful production systems may be *x-omically* characterized to elucidate key metabolic pathways and control points for future rounds of targeted metabolic engineering (Bro et al, 2004). In both *forward* and *inverse* metabolic engineering, systems level models and simulations are accelerating bio-based process development, resulting in reduced time to commercialization with significantly less resource commitment. As previously suggested, industrial systems biology encompasses many of the aforementioned methodologies and technologies. Figure 3 presents the enhanced metabolic engineering cycle (Bailey, 1991) with *industrial systems biology* applied to an industrial production host to develop a microbial cell factory.

Today, industrial biotechnologists are no longer considering singular products such as bioethanol, but rather diverse portfolios of petrochemical commodity, added-value, high added-value, and specialty chemicals to be produced using biotechnology. The term biorefinery was first defined in 1999, when it was suggested that lignocellulosic raw materials may be converted to a portfolio of bio-commodities via integrated unit processes, and offer competitive performance to existing petrochemical refineries (Lynd et al, 1999). If the biorefinery platform

**Systems Biology:** Systems biology is a multi-disciplinary approach relying on the integration and interrogation of diverse data types that share a common scaffold based on genomics and functional genomics characterization of a biological system, with particular focus on development of predictive and quantitative mathematical models.

**Metabolic Engineering:** Metabolic engineering is a field encompassing both *forward metabolic engineering* and *inverse metabolic engineering*, defined further below. Metabolic engineering is the gene-targeted, rational, and quantitative approach to redirection of metabolic fluxes to improve the yield, titer, productivity, and/or robustness associated with a specific metabolite in a biological system.

**Industrial Systems Biology:** The application of numerical or experimental methods developed as a result of individual or combined *x-ome* analysis to bioprocess development for commercialization. Bioprocess development encompasses strain or expression system improvements in terms of final product titer, yield, or productivity, or improvements in process robustness and efficiency.

**Forward Metabolic Engineering:** Defined as targeted metabolic engineering, it represents the linear progression from modelling to target gene identification to strain construction and characterization. Inherent to this strategy is specific and hypothesis driven genetic manipulations that are based on predictive metabolic modelling, from simple biochemical pathway stoichiometric balancing to more sophisticated kinetic models.

**Inverse Metabolic Engineering:** Also defined as reverse metabolic engineering, a host strain constructed via random or directed mutagenesis, and/or evolution is examined via systems biology tools to determine the genetic perturbation(s) that lead to the desired phenotype.

**X-omics:** A general term for referring to collection and analysis of any global data set whereby any type of informational pathway with reference back to the cell's genome is investigated. By definition, *x-ome* analysis and data collection requires the whole cell genetic sequence, preferably, annotated.

**Genomics:** The comprehensive study of the interactions and functional dynamics of whole sets of genes and their products, often species and strain specific.

**Transcriptomics:** The genome-wide study of mRNA expression levels in one or a population of biological cells for a given set of defined environmental conditions.

**Metabolomics:** The measurement of all metabolites, intracellular and extracellular, to access the complete metabolic response of an organism to an environmental stimulus or genetic modification. Here, a metabolite is defined as being any substrate or product participating in a reaction catalyzed by any gene product.

**Fluxomics:** The study of the complete set of fluxes that are measured or calculated in a given metabolic reaction network. A metabolic flux is defined as a quantitative measurement of the rate of conversion of reactants to products, where rate may be defined as the mass or concentration per unit time of reactant consumption and product formation.

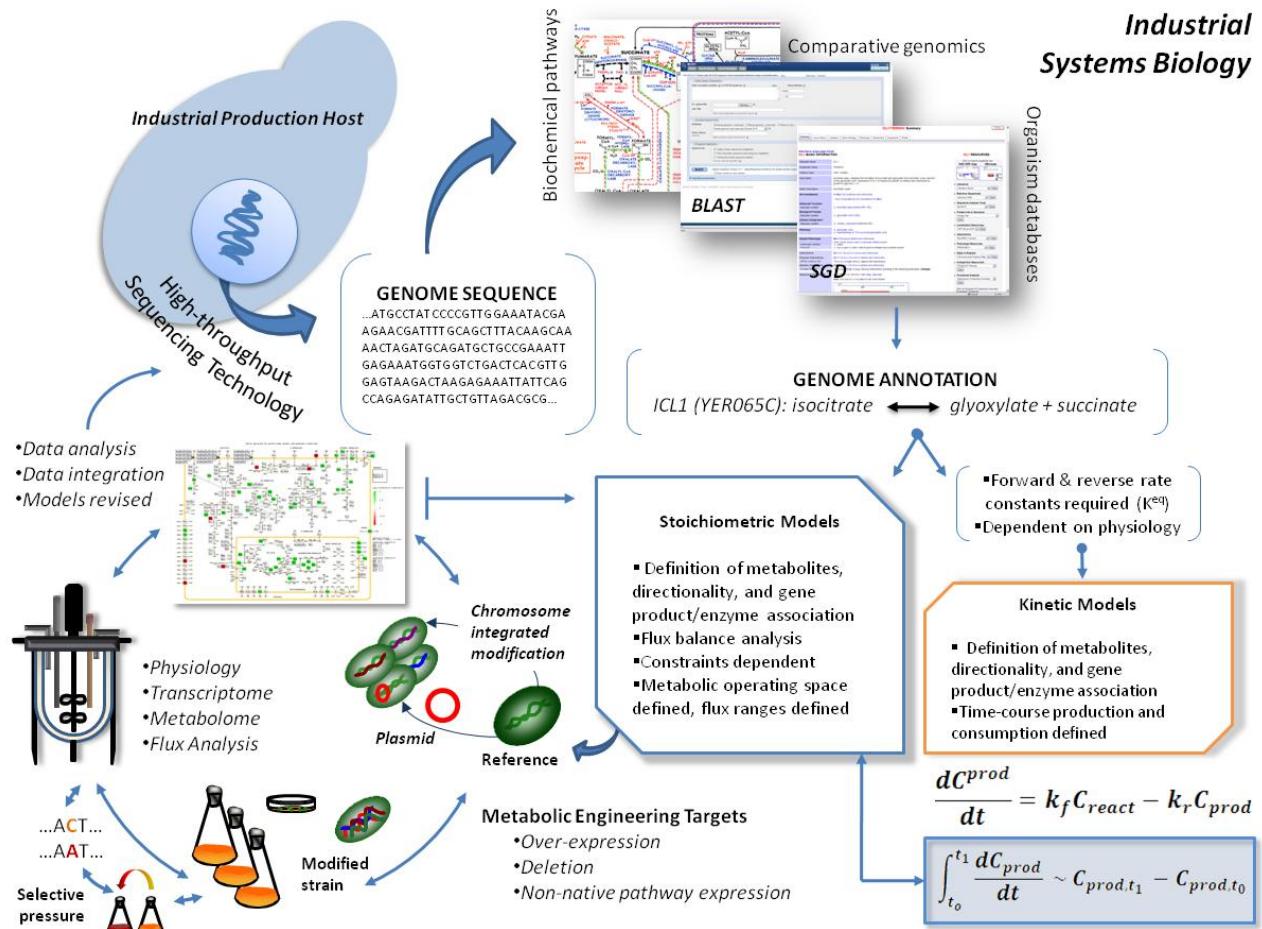
**Proteomics:** The large-scale analysis of the structure and function of proteins as well as of protein–protein interactions in a cell.

**Metagenomics:** The study of the genomes and associated *x-omes* in organisms recovered from the environment as opposed to laboratory cultures. Recovered *x-omes* may be from multiple organisms where the exact origin and tracing of *x-ome* information may not be known. Organisms recovered from the environment are often difficult to culture in controlled laboratory conditions, but may reveal interesting characteristics accessible through functional genomics.

**Figure 2: X-omic Glossary.** The glossary of terms defined above are used throughout the text, and furthermore, encountered in the broad literature that encompasses microbial metabolism, metabolic engineering, and systems biology.

model is to evolve from academic conception to industrial reality it will require two essential driving forces. First, the economic and socio-political landscape must continue to support and warrant the significant financial investment, favorable legislative policy, and consumer driven demand that will be required. Second, the advances and tools developed within systems biology

for metabolic engineering must be successfully applied in commercial environments. Several examples, such as bioethanol, have suggested that biorefineries are viable commercially; however, the diverse product streams that will be required continue to demand more sophisticated, native and non-native, multi-gene metabolic engineering approaches. These approaches may only be realized through advanced interrogation and integration of microbial metabolic space using systems biology tools.



**Figure 3: Industrial Systems Biology.** Industrial systems biology is a dynamic interaction between various disciplines and approaches. At the core is a platform technology based on a production host, for which a genome sequence is available, and subsequent annotations based on existing literature review, database query, comparative genomics, and experimental data, where available, are completed. The annotations may vary in types of functional genomics data assigned to specific fields; however, a standard skeleton syntax structure of defining a gene, the gene product (e.g., metabolic enzyme), the metabolites serving as reactants and products (including any co-factors and intermediates), and the resulting stoichiometry is often applied. This framework, referred to as a genome-scale metabolic network reconstruction, may then be used for stoichiometric or kinetic modeling. Often, because kinetics parameters such as the forward and reverse reaction rates at physiologically relevant conditions have not been experimentally determined for a significant fraction of the network, flux balance analysis (FBA) is used for predictive modeling as it only depends on the stoichiometry and network constraints (e.g., precise stoichiometric definition of biomass, ATP maintenance terms, glucose uptake rate). Once a high-probability of success metabolic engineering strategy has been identified, often requiring gene overexpression, deletion, or non-native pathway reconstruction, genetic engineering is performed on the production host, yielding a modified strain. The modified strain is initially characterized, and may undergo directed evolution or other non-targeted approaches to yield an improved phenotype. The resulting modified strain is then characterized under well-controlled fermentation conditions, where physiological parameters, such as maximum specific growth rate, substrate consumption rates, product yields and

titors, by-product formation, and morphology are determined. Furthermore, functional genomics characterization, often requiring transcriptome, proteome, metabolome, and fluxome measurements is completed. Bioinformatics, coupled with data integration, are then required for analysis of the resulting modified strain, and to identify opportunities for a second round of metabolic engineering. Furthermore, the analysis should lead to a revised model with improved predictive power that may yield promising strategies for further phenotype improvement. While this approach has often been referred to as the metabolic engineering cycle, we here compliment the traditional cycle to include integrative approaches and data sets from systems biology. Together, when applied to industrial biotechnology products, this is referred to as industrial systems biology.

## 3.0 Microbial Metabolism

Microbial metabolism from the perspective of an industrial biotechnology process engineer is a portfolio of available substrates, catalyzing reactions, and products that when controlled can be formulated into the collection of desirable pathways that will ultimately form a microbial cell factory. The section that follows aims to provide some context and historical perspective of microbial metabolism in the field of bioreaction and metabolic engineering.

### 3.1 A Historical Perspective

There have been extensive reviews regarding the application of random mutagenesis and directed evolution for novel development or enhancement of existing microbial cell factories for the production of a wide range of industrial biotechnology products (Parekh et al, 2000; Demain, 2000; Demain et al, 2008a; Demain et al, 2008b; Schmeisser et al, 2007). What is often referred to as “classical strain development” is dependent on the capability of inducing and promoting genetic diversity, under controlled laboratory conditions, in a desirable production host organism that can be selectively screened, isolated, cultured, and preserved based on a phenotypic criteria. Genetic diversity may be induced using mutagenic chemical agents, radiation, ultra-violet light exposure, intercalating agents, or through genetic recombination (Parekh, 2000). While resulting modified strains may then be further physiologically characterized, the specific and targeted genetic alterations that lead to the improved phenotype are not known, preventing any mechanistic understanding from being applied to future rounds of strain improvement. Furthermore, genetic alterations independent of the selective pressure or phenotype of interest may accumulate resulting in strain deficiencies and manifestation of undesirable phenotypes or reduced robustness.

Microbial metabolism, the working space of modern metabolic engineering, has been characterized and expanded upon as a scientific body of knowledge for nearly a century; however, if one were to establish a time-line, then perhaps 1932 serves as a suitable starting point, as it coincides with the *Science* publication of Professor Albert Jan Kluyver (1888-1956), entitled, “Microbial metabolism and its bearing on the cancer problem” (Kluyver, 1932). Similar to the recombinant DNA technology first pioneered in the early 1970s at Stanford University and the University of California at San Francisco (Williams et al, 1973), the first applications of microbial metabolism were related to human health and medicine. And yet, there was a third publication in 1932, appearing five months prior to Kluyver, *et al.*, that while it received significantly less attention at the time, offered one of the first literature examples of the infancy of the role microbial metabolism would play in industrial biotechnology, entitled, “An application of the autocatalytic growth curve to microbial metabolism,” and appeared in the *Journal of Bacteriology* (Pulley et al, 1932). In a twenty-two page manuscript, with four references, two notable observations are made. The first is a simple, yet governing observation of microbial growth kinetics, neatly summarized by the authors in the opening paragraph:

*Growth does not take place at a constant rate in living organisms. In bacterial cultures it is initially exceptionally slow, then increasingly rapid, and finally exceptionally slow. This is most conspicuous if the initial inoculum into fresh media is very small. We have found that the rate of accumulation of microbial metabolic products likewise is not constant but begins slowly, increases rapidly, and again slows down (Pulley et al, 1932).*

The second observation is that this manuscript represented a cross-disciplinary approach where kinetic differential reaction equations, first applied to monomolecular autocatalyzed chemical

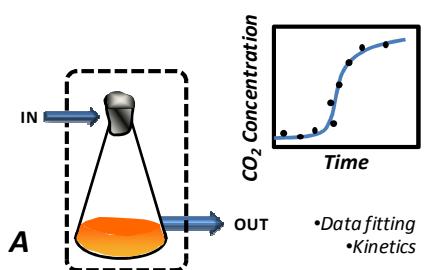
reactions, and later to the growth of plants and animals (Robertson et al, 1923), were used to fit nitrate and carbon dioxide accumulation data in soil bacteria and *Saccharomyces cerevisiae*, respectively (Pulley et al, 1932). It is the first clear example of where mathematical models were used to fit existing microbial metabolic data, and consequently, yield a predictive relationship between the accumulation of metabolic products, (e.g., nitrates and carbon dioxide), and time, governed by a reaction rate constant (See Figure 4). While Jacques Monod is often credited with the modern mathematical characterization of growth kinetics, commonly referred to as *Monod growth kinetics*, which includes a kinetic relationship for biomass formation as a function of substrate concentration and affinity, that relationship evolved from his seminal work on enzyme kinetics using  $\beta$ -galactosidase formation and activity in *Escherichia coli*, neatly summarized in a series of publications that culminated with a 1953 publication in *Nature* (Cohn et al, 1953). It was Pulley, *et al.*, nearly twenty years prior that was evaluating microbial metabolic relationships that would ultimately serve as the foundations for early industrial fermentations. During the same period numerous other scientists, including Meyerhof (Cohn et al, 1953; Meyerhof et al, 1949a; Meyerhof et al, 1949b; Meyerhof et al, 1949c; Meyerhof et al, 1949d; Meyerhof et al, 1949e; Meyerhof et al, 1949f; Meyerhof et al, 1949g), Embden (Kobayashi et al, 1954), Parnas (Mann, 1955), Warburg (Warburg et al, 1936; Krebs, 1972), Gori (Gori et al, 1939; Gori et al, 1952; Gori et al, 1954), Harden (Harden et al, 1911; Harden, 1913a; Harden et al, 1913b; Harden et al, 1914a; Harden et al, 1914b), and Neuberg (Gottschalk et al, 1956), were elucidating metabolic pathways, primarily involved in anaerobic fermentation, again with particular focus to medical applications. During the same time, Hans A. Krebs was elucidating primary components of central carbon metabolism, such as amino acid metabolism (e.g., glutamic acid, proline), and the citric acid cycle, today often referred to as the Krebs cycle (Krebs, 1935a; Krebs, 1935b; Krebs, 1937; Krebs, 1938a; Krebs et al, 1938b; Krebs et al, 1939; Krebs, 1940a; Krebs et al, 1940b; Krebs et al, 1952a; Krebs et al, 1952b; Krebs, 1948; Krebs, 1964), highlighting the cell's capability for aerobic metabolism and metabolite oxidation. It is interesting to note that baker's yeast once again played a critical role in the pursuit of this line of research, when  $^{14}\text{C}$ -labelled acetate was supplemented to *S. cerevisiae*, and oxidized. It was observed that the intracellular dicarboxylic acids (citric acid cycle intermediates) failed to incorporate the  $^{14}\text{C}$ , suggesting that the type of carbon source supplied lead to different modes of fermentation and oxidation (Krebs et al, 1952b). Krebs, *et al.*, did note that the results were not conclusive as intracellular transport resulting from compartmentalization may have created permeability barriers (Krebs et al, 1952b). In many respects, one could argue that this was the beginning of metabolic flux measurements for pathway elucidation<sup>2</sup>. In 1953, Hans A. Krebs shared the Nobel Prize in Physiology, along with Fritz A. Lipmann (credited with discovery of coenzyme-A) for their elucidation of intermediate metabolism. While focus remained on human health, in his Nobel address in 1953, Krebs emphasized in reference to a debate regarding acetate's oxidation and relation to the citric acid cycle:

*It is true that these results may not be looked upon as conclusive because permeability barriers might prevent the mixing of substances arising as intermediates with those that are present in other compartments of the cell, and at present it is best to regard the terminal pathway of oxidation in yeast, and certain other microorganisms, e.g. *E. coli*, as an open problem, even though the reactions of the cycle occur in these materials (Krebs, 1964).*

---

<sup>2</sup> Isotope labeling experiments widely appeared in the 1940s for studying the distribution of carbon atoms in fatty acids and acetoacetate, that were demonstrated to appear in the citric acid cycle, and consequently, that dicarboxylic acids are intermediates in the complete oxidation of fatty acids (Krebs, 1948).

Already then, baker's yeast and *E. coli*, two of the leading industrial biotechnology hosts, were being evaluated as model organisms, and one of the key distinguishing features of each organism that to this day impacts metabolic engineering strategies employed, was noted: compartmentalization.



### Microbial Metabolism Modelled Using Catalysis Chemistry

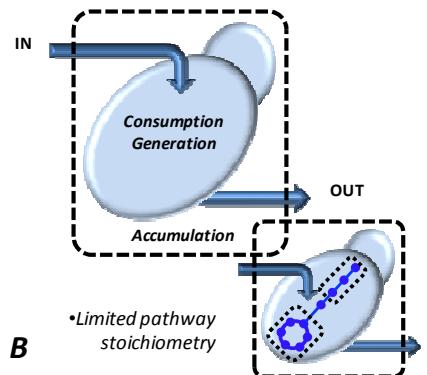
- Autocatalyzed chemical reaction chemistry applied to *S. cerevisiae* CO<sub>2</sub> production under exponential growth conditions

$$S \xrightleftharpoons{k_1 C} P, \text{ where } k_1 \text{ is reaction rate constant and } C \text{ is the catalyst}$$

$$\frac{dP}{dt} = k_1 C [S] - k_1 C [P], \text{ assume } C \ll P$$

$$\frac{dP}{dt} = k_1 P [S] - k_1 [P]^2 = k_1 [P]([S] - [P])$$

<1940s



### Microbial Metabolism Mass Balancing with Kinetics

- Microbial growth, substrate uptake, and production formation kinetics
- Isotope labeled substrates enables pathway flux estimations
- Central carbon metabolic pathways (e.g., citric acid cycle) identified

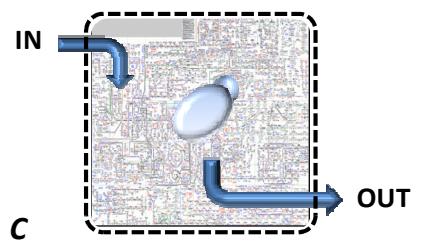
Biomass formation and substrate uptake kinetics:

$$\mu = \mu_{\max} \frac{S}{K_s + S}, \frac{dX}{dt} = \mu X - k_d X, \frac{dS}{dt} = \frac{1}{Y} \frac{dX}{dt}$$

Mass balancing of metabolites across biomass:

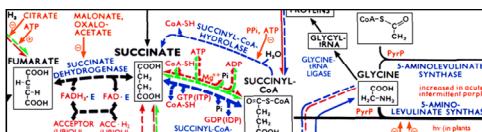
$$S_{in} F(t)_{in,i} - S_{out} F(t)_{out,i} + q_i X(t) = \frac{dS(t)_i}{dt}$$

1940-1960s



### Visual Reconstructed Metabolic Network

- Defined stoichiometric relationships, enzyme association, chemical structure, pathway connectivity, and where possible, regulation.



- 1000 enzymes visually represented
- >3000 enzymes classified [1]
- >30,000 proteins in a single mammalian cell [2]

>1960s

**Figure 4: Microbial Metabolism.** As suggested in panel A, microbial metabolism has been extensively investigated since the 1930s, when classical reaction rate expressions to describe autocatalyzed reaction chemistry were applied to carbon dioxide formation in *S. cerevisiae*. This approach required metabolite concentration profiles as functions of time, and the starting metabolite concentration to determine the reaction rate constant. This was a data-fitting approach with minimal focus on predictive power; however, semi-quantitatively explored the relationship between metabolite consumption and production rates, with relation to specific growth rate (Pulley, 1932). Continuing forward nearly twenty years, significant progress in elucidating glycolysis, the citric acid cycle, and fatty acid oxidation was made. During this time period, growth kinetics, the relationship between specific growth rate, substrate utilization, and product formation rate was further developed. This approach relied on classical mass balancing with emphasis on resolution of kinetic parameters (Panel B). From the 1960s onward, the approach of describing metabolism with mass balancing and kinetic parameters estimation vastly expanded to include the majority of metabolic space. A significant milestone was the publication of Biochemical Pathways that provided the first visual representation of major components of metabolism. With the 4<sup>th</sup> edition of Biochemical Pathways recently published, there are over 1000 enzyme catalyzed reactions depicted with specific annotation including stoichiometry, chemical structure, pathway connectivity, compartmentalization, and where possible, regulation (Michal, 1999). The reconstructed microbial network, particularly provided in a singular, graphical representation, permitted the first generation of metabolic engineering strategies to be devised based on rational, hypothesis-driven, strategies.

The foundations of modern microbial metabolism can continue to be traced through the literature, but the small sub-set of examples provided here intends only to provide perspective and context for the early industrial biotechnology processes. Prior to the development of recombinant DNA technology, there was a key development in the metabolism literature that to this day continues to impact metabolic engineering and systems biology approaches; however, it is often not discussed in this context. In 1965 the first edition of *Biochemical Pathways* was published by the Boehringer Mannheim GmbH (Mannheim, Germany) company, created and lead by Dr. Gerhard Michal. This was the first comprehensive, visual, graphic representation of metabolism, which was initially in wall chart form (See Figure 4); however, has recently been converted to book format (Michal, 1999)<sup>3</sup>. This integration of biological data that included stoichiometric relationships, chemical structure of all reactants, products, intermediates, and co-factors, assignment of enzymes to specific reactions, definition of compartments and transport, and where possible, regulatory data or interactions, was first comprised for central pathways such as glycolysis, the citric acid cycle, synthesis and degradation of fatty acids, amino acids, and nucleotides. Perhaps most importantly, it provided one of the first global visual representations of the reconstructed metabolic network, where connectivity between metabolites, pathways, and compartments could be realized. A systematic visual representation of the metabolic network permitted intuitive, hypothesis driven, metabolic engineering approaches to be developed, based on driving carbon flux in given directions, or re-directing other metabolic fluxes in desirable directions. This metabolic map frequents the walls of all major academic and industrial research centers of industrial biotechnology, and provided metabolic engineers a comprehensive overview of the landscape within which they were operating. Of course, the critical piece of information missing was the association between genes, gene products (e.g., functional enzymes), and the metabolic pathways on which they act. Furthermore, the *Biochemical Pathways* wall chart, while it has evolved to include organism-specific detail, has largely been treated as a summary of all metabolic reactions in the many different organisms studied. Again, it would ultimately be comparative genomics, coupled with bioinformatics efforts to create organism catered databases that could provide required specificity.

---

<sup>3</sup> As of this writing there have been four editions of the *Biochemical Pathways* wall chart, including the reference book, *Biochemical Pathways*, published in 2005, all continuing to be edited by Dr. Gerhard Michal. The wall chart continues to be made available by Roche Diagnostics GmbH (Mannheim, Germany), at [https://www.roche-applied-science.com/techresources/publications\\_req.jsp](https://www.roche-applied-science.com/techresources/publications_req.jsp).

### 3.2 Genome Sequencing & Functional Genomics

There are several reviews that have provided historical perspectives with respect to the formation of the multi-disciplinary field of systems biology, and its impact on metabolic engineering or more broadly industrial biotechnology, often focusing on milestone publications (Westerhoff et al, 2004; Oliver, 2006; Hermann et al, 2007; Rakors et al, 2007; Nielsen et al, 2008). A supplementary approach more focused on dissecting which milestones were critical for commercialization of industrial biotechnology, is to inspect the patent literature that in addition to cataloguing the specific scientific or technological achievement also suggests their industrial importance. The sequence of inventions described below, by no means exhaustive, provides a temporal context of some of the milestones in the fields of industrial biotechnology and recombinant DNA technology that ultimately culminated in the first major added-value product produced via extensive forward metabolic engineering.

In 1948, the Cold Spring Harbor Laboratory (Cold Spring Harbor, New York) was granted a US patent entitled, “Production of Penicillin”, which states, “It is an object of this invention to produce penicillin in extremely highly yields. Another object is to produce mutations of molds of the genus mycetes capable of yielding extremely large amounts of antibiotic substances (Demerec, 1948).” In 1982, Genentech, Inc. (South San Francisco, California), was granted a patent entitled, “Method for microbial polypeptide expression,” that cited somatostatin, an inhibitor of the secretion growth hormone, as an example polypeptide, and went on to describe, “Despite wide-ranging work in recent years in recombinant DNA research, few results susceptible to immediate and practical application have emerged. This has proven especially so in the case of failed attempts to express polypeptides and the like coded for by ‘synthetic DNA’, whether constructed nucleotide by nucleotide in the conventional fashion or obtained by reverse transcription from isolated mRNA (complimentary or ‘cDNA’). In this application we describe what appears to represent the first expression of a functional polypeptide product from a synthetic gene, together with related developments which promise widespread application (Riggs, 1982).” In 1985, the Purdue Research Foundation (West Lafayette, Indiana) was granted a patent entitled, “Direct fermentation of D-xylose to ethanol by a xylose-fermenting yeast mutant”, where in they claimed, “...a process for producing yeast mutants capable of utilizing D-xylose to ethanol in high yields is described (Gong, 1985)...”, the yeast mutants being *Candida sp.* XF 217 and *S. cerevisiae* SCXF 138. In 1997, the E.I. Du Pont de Nemours and Company (Wilmington, Delaware) was granted a US patent entitled, “Bioconversion of a fermentable carbon source to 1,3-propanediol by a single microorganism (Laffend et al, 1997).” This patent went on to comprehensively describe the metabolic pathways present in naturally producing microorganisms (e.g., *Citrobacter sp.*, *Clostridium sp.*, *Klebsiella sp.*), and which specific enzyme activities were both required for carbon flux redirection and redox balancing (e.g., NAD<sup>+</sup> regeneration). In one of the first major successes of modern metabolic engineering, it went on to describe the specific expression vectors and cloning techniques used to construct a recombinant strain of *E. coli* capable of high-yielding production of 1,3-propanediol<sup>4</sup> (Laffend et al, 1997; Nakamura et al, 2000). This rather brief survey of the extensive biotechnology patent literature suggests that for nearly forty years (1948-1985), development of industrial biotechnology processes was relegated to methods of mutant selection. However, within approximately ten additional years (1997), gene-targeted approaches were in use to construct microbial cell factories capable of producing high added-value chemicals.

---

<sup>4</sup> There have been several US patents issued to E.I. Du Pont de Nemours and Company, and collaborating enterprises, such as Genencor International, in the development of 1,3-propanediol; however, the US patent issued in 1995 represents the first of that series of patents (Laffend et al, 1997; Nakamura et al, 2000).

The history of industrial organic acid production begins in 1826 when the first commercial process for citric acid from imported Italian lemons was established in England (Papagianni, 2007). In 1919 the first industrial biotechnology process for citric acid production was launched in Belgium using *Aspergillus niger* (Papagianni, 2007), and during the 1930s and 1940s the bioprocess development and optimization, starting with media development, was prevalent (Papagianni, 2007). Another organic acid that was produced early was L-glutamic acid, more commonly known as monosodium L-glutamate (MSG), isolated by Professor Kikunae Ikeda of the Tokyo Imperial University as the principal component of umami (kelp) (Sano, 2009). In 1956, an industrial biotechnology process for L-glutamate was introduced using the production host *Corynebacterium glutamicum* (Sano, 2009). Similar to the history presented earlier regarding industrial biotechnology, there is a rich history of using microbial production platforms for synthesis of organic acids, particularly those used in the food and flavor sectors, that pre-dates biochemical reaction or metabolic engineering.

Those with industrial experience will recall that in the late 1980s and early 1990s, with recombinant DNA technology emerging from medical biotechnology, we witnessed increased expression of compounds previously produced via synthetic routes now becoming more prolific in production organisms (Demain, 2000; Moo-Young et al, 1984; Poppe et al, 1992; Schmid, 2003). In addition to *de novo* industrial biotechnology production of compounds, there was also increased efforts to enhance existing biotechnology processes. Examples of compounds in both categories, with their current estimated annual production include: L-glutamic acid (2,000,000 tons), citric acid (1,600,000), L-lysine (350,000 tons), lactic acid (250,000), food-processing enzymes (100,000 tons), vitamin C (80,000 tons), gluconic acid (87,000), antibiotics (35,000 tons), feed enzymes (20,000 tons), xanthan (10,000 tons), L-hydroxyphenylalanine (10,000 tons), Vitamin F (1000 tons), and Vitamin B12 (12 tons), to name a few (Gavrilescu et al, 2005; Bruggnik, 1996; Bruggnik et al, 2003; Eriksson, 1997; Sauer et al, 2007). This was made possible by the introduction of genetic sequences encoding for enzymes that were likely to catalyze desired reactions, or, the deletion of genes that would down-regulate undesired reactions and pathways. These approaches were largely hypothesis driven, resource intensive, and low-throughput, minimizing the probability of successfully identifying a genotype that would elicit a significantly improved phenotype. The real advantage of random mutagenesis, screening, and selection, was the relatively large experimental space that could be covered, even if mechanistic understanding was sacrificed. The other advantage was its track record – it worked. Fast-forward approximately ten years, and what has changed?

Although techniques that permitted manipulation of recombinant DNA existed, the annotated genome sequences of industrially relevant production hosts were not available. In the paper entitled, *Industrial Systems Biology* (**Paper V**), Figure 4 highlights the exponential increase in published genome sequences that first started in 1995 and have continued to expand through 2008. As of October 1, 2009, a total of 1117 published genome sequences, and 3355, 1187, and 112 bacterial, eukaryotic, and archaeal sequence projects are on-going, respectively (Kyrpides, 1999; Liolios et al, 2006). This genomic revolution was mainly driven by the medical research field, as illustrated in Table 1 of **Paper V**, which presents characteristics of those genomes sequenced between 1995 and 1999. It is seen that of the twenty-four sequences made available, only three could be considered to have broad applicability to the industrial biotechnology sector: *Saccharomyces cerevisiae*, *Escherichia coli*, and *Bacillus subtilis*, while the rest were driven by the medical community. One can even argue that sequencing of these three genomes was also mainly motivated by their medical relevance, either as a eukaryote model organism, pathogen or model pathogen. If we move beyond 1999, as indicated in Table 2 of **Paper V**, many more industrially important cell factories have been genome sequenced, and with the substantial reduction in sequencing costs even genome sequencing has become a tool to analyze cell factories with different phenotypes. The presence of complete genome sequences has clearly allowed better

targeting of genetic modifications, and information about the complete parts lists of a given cell factory is extremely valuable.

With genome sequences for several industrial model organisms in hand, it was the annotation of those sequences that bridged the gap between expanding knowledge-based databases (e.g., genome sequence collections) and the data-driven databases (e.g., application of the genome sequences for annotation, model development, and further understanding) (Viswanathan et al, 2008). The annotation of genome sequences has evolved into a well-defined discipline referred to as functional genomics, which focused on developing numerous experimental and theoretical tools for determination of gene function (Bruggeman et al, 2007). Functional genomics, through linking gene products (e.g., enzymes) to gene functions (e.g., reaction stoichiometry) has permitted the development of genome-scale models for various data types, such as reconstructed metabolic network models.

Following the release and annotation of a genome the next logical step is to evaluate the messenger RNA expression level on a whole genome scale, referred to as transcriptome analysis. Targeted metabolic engineering relies heavily on the assumption that a genetic perturbation – gene deletion, constitutive overexpression, regulated induction, or modulation – will confer a metabolic flux response. This stems from the central dogma of biology: DNA is transcribed to RNA and subsequently translated to polypeptides that give rise to phenotype. Prior to transcriptome analysis, genes were assumed to be expressed followed by post-translational regulation, with little understanding of interactions across gene loci (Schmid, 2003). In fact, transcriptome profiling of reference strains has provided a first approximation as to which pathways are active and equally important, inactive, assuming that up-regulated gene expression leads to up-regulated pathway activity. It has since been shown that this is not always true – elevated mRNA levels do not always translate to elevated protein levels or activity. It has also provided significant insight into alternative modes of regulation, such as transcription factor mediated as opposed to post-translational regulation. This has permitted narrowing of the experimental space that metabolic engineers need to consider, and made available new strategies to consider. Additionally, transcriptome profiling provides a quantitative *in vivo* assessment of several key metrics following a genetic perturbation relative to a reference case: (1) what is the net change in mRNA expression levels of the targeted gene(s), (2) what is the net change in mRNA expression levels of non-targeted gene(s), and (3) what is the net change in mRNA expression levels of either reference or constructed strains under specific environmental conditions. These questions aim to isolate which genes and pathways may serve as targets and/or explanations for observed or induced phenotypes. Measurement of the transcriptome, via readily available microarray technology, has evolved into a routinely measured data set for many industrially relevant organisms, including *E. coli* and *S. cerevisiae*, and is playing a central role in both *forward* and *reverse* metabolic engineering (Bro et al, 2004; Lynch et al, 2004; Gill, 2003).

Among the first applications of transcriptome measurements with industrial relevance to bioethanol production was establishing the baseline response of *S. cerevisiae* to diverse carbon substrates and medium compositions – essential for optimizing strains to given feedstocks and vice versa. Steady-state chemostat cultures were used to measure transcriptome responses under glucose, ethanol, ammonium, phosphate, and sulphate limitations (Wu et al, 2004). Results suggested that genes related to high-affinity glucose uptake, the TCA cycle, and oxidative phosphorylation were up-regulated in glucose-limiting conditions, while genes involved in gluconeogenesis and nitrogen catabolite repression were up-regulated in ethanol-grown cells (Wu et al, 2004). In a similar but earlier study, transcriptome measurements of *S. cerevisiae* grown using glucose-limited chemostats coupled with nitrogen, phosphorus, and sulphur limitations were performed (Boer et al, 2003). In total, 1881 transcripts (31% of the total 6,084 different open reading frames probed) were significantly up- or down-regulated between at least two conditions, and a total of 51 genes demonstrated a >10-fold higher or lower expression within a

given condition (Boer et al, 2003). The transcriptome profiles under each condition have provided genetic motifs that may be recognized and regulated by transcription factors, which may be used in metabolic engineering strategies that could cater to a specific growth medium composition.

With the experimental mechanics of collecting transcriptome becoming common place, attention and focus is now placed on data analysis methods and integration with other *x-ome* data sets. It has become abundantly clear that transcriptome data alone, unless used for purposes of environmental screening or quality control (i.e., confirming that an engineered genotype is producing the corresponding transcription profile), provides limited biological insight. Several efforts have emerged coupling transcriptome with metabolome and fluxome data (Phelps et al, 2002; Ideker et al, 2001; Erasmus et al, 2003; DeRisi et al, 1997). For example, elementary flux modes for three carbon substrates (glucose, ethanol, and galactose) were determined using the catabolic reactions from the genome-scale metabolic model of *S. cerevisiae*, and then used for gene deletion phenotype analysis. Control-effective fluxes were used to predict transcript ratios of metabolic genes for growth under each substrate, resulting in a high correlation between the theoretical and experimental expression levels of 38 genes when ethanol and glucose media were considered (Cakir et al, 2004). This example demonstrates that incorporating transcriptional functionality and regulation into metabolic networks for *in silico* predictions provides both more biologically representative models, and a means of bridging transcriptome and fluxome data. In another example, the topology of the genome-scale metabolic model constructed for *S. cerevisiae* is examined by correlating transcriptional data with metabolism. Specifically, an algorithm was developed enabling the identification of metabolites around which the most significant transcriptional changes occur (referred to as reporter metabolites) (Patil et al, 2005). Due to the highly connected and integrated nature of metabolism, genetic or environmental perturbations introduced at a given genetic locus will affect specific metabolites and then propagate throughout the metabolic network. Using transcriptome experimental data, predictions *a priori* of which metabolites are likely to be affected can be made, and serve as rational targets for additional inspection and metabolic engineering (Patil et al, 2005). This algorithm has been recently extended to include reporter reactions, whereby transcriptional data is correlated with the metabolic reactions of the reconstructed *S. cerevisiae* genome-scale metabolic network model to identify those reactions around which a genetic or environmental perturbation conferring transcriptional changes cluster (Cakir et al, 2006).

As more genomes continue to become available, and microarray technology continues to become more accessible with cost-effective customizable DNA microarrays now available, transcriptome data will continue to increase. Bioinformatics for data handling, integration of transcriptome with other *x-ome* data, and the development of various network models that rely on transcriptome data for biological interpretation will continue to develop. From an industrial biotechnology perspective transcriptome measurements and analysis have played a large role in reverse metabolic engineering – transcriptional surveying of a strain constructed either via random mutagenesis or directed evolution (Bro et al, 2004; Lynch et al, 2004; Gill et al, 2003; Sonderegger et al, 2004). For example, lysine production via *C. glutamicum* has undergone transcriptome and fluxome measurements to elucidate greater than 50 years of traditional metabolic engineering (random mutagenesis), providing new targets for improved strategies (Koffas et al, 2005; Wendisch, 2003; Wendisch et al, 2006). This effort, applied to other industrial biotechnology processes, is likely to intensify.

## 4.0 Reconstructed Metabolic Network Models

The preceding sections provided a foundation for how milestones in microbial metabolism, coupled with the more recent integration of functional genomics, has permitted the field of systems biology to be an extension and enhancement of the metabolic engineer's toolbox. This section discusses how reconstructed metabolic network models provide a simple stoichiometric mathematical framework for predictive simulations and organization of *x*-*omics* data sets.

### 4.1 Introduction

Even though genome sequencing and functional genomics have clearly facilitated the use of targeted genetic modifications for construction of cell factories with desirable phenotypes, the major step forward has been through the introduction of metabolic engineering, which is the enabling science for cell factory design and construction. Metabolic engineering involves the identification of specific and targeted genetic modifications (gene deletions, over-expression, or modulation) followed by implementation of these modifications via molecular biology tools that lead to re-direction of fluxes to enhance production or robustness of a given product or organism, respectively (Bailey, 1991; Stephanopoulos and Vallino, 1991; Nielsen, 2001; Kern et al, 2007; Tyo et al, 2007; Patnaik, 2008). A key technology in the successful application of metabolic engineering is the availability of a well annotated genome, including the quantitative tools that permit careful inspection and manipulation of the genome. Among those tools has been the recent development of genome-scale metabolic models (GSMMs). To develop a model of cellular metabolism that enables the prediction of concentration profiles as functions of time, the stoichiometry and kinetic reaction rates for each biochemical reaction in a cell at physiological conditions would be required. At present, this information is not available, neither via estimation or experimental measurement. Through careful annotation based on existing biochemical knowledge, literature review, and experimentation, however, it is possible to associate known genes with known biochemical reactions and their corresponding stoichiometry. The result is a biochemical model describing the formation and depletion of each metabolite that by providing mass-balance boundary conditions makes possible constraint based simulations of how the metabolic network operates at different conditions. In simpler terms, using basic stoichiometry these models can be used to predict the relationships between genes with function in the metabolic network operating in a cell. If cells are fed  $x$  grams of glucose it is possible through the use of linear programming and the biochemical model to predict the maximum  $y_i$  grams of formed product  $i$ . It is hereby also clear that GSMMs can be used to predict a theoretical landscape of genetic perturbations that can maximize product and biomass formation, even under different growth conditions (i.e., growth on alternative carbon sources). GSMMs have been developed for several model production organisms (See of Table 3 of **Paper V**), and were a major step in not only allowing model guided metabolic engineering, but also integration of different *x*-*ome* data for obtaining detailed metabolic characterization. GSMMs provide an appropriate scaffold for further expansion and data integration, owing to its easily manipulated mathematical framework. While that framework has been described previously, it warrants a brief review in the following section.

The first genome-scale *in silico* metabolic network model for *E. coli* was made available in 2000 and was among the first to demonstrate consistency between modeling predictions and *in vivo* physiology (Varma et al, 1994). Specifically, the model was used to explore the relationship between acetate, succinate, and oxygen uptake rates when attempting to maximize growth rate, to confirm the hypothesis that *E. coli* under acetate and succinate carbon limitations regulates its metabolic network to maximize growth rate. For industrial biotechnology process development,

it is often desirable to shift carbon flux from biomass to product formation, thereby maximizing the yield of product on substrate.

The first eukaryotic genome-scale metabolic model was reported in *S. cerevisiae* in 2003 based on its annotated genome sequence and a thorough examination of online pathway databases, biochemistry textbooks, and journal publications (Förster et al, 2003a). This genome-scale *in silico* model, by using a relatively simple synthetic medium, could predict 88% of the growth phenotypes correctly, indicating that this model in many cases can predict cellular behavior. In one step further, Duarte et al (2004) (Duarte et al, 2004a) used the *S. cerevisiae* genome-scale metabolic network constructed by Förster et al (2003) (Förster et al, 2003a) to generate a Phenotypic Phase Plane (PhPP) analysis that describes yeast's metabolic states at various levels of glucose and oxygen availability. Examination of the *S. cerevisiae* PhPP has led to the identification of two lines of optimality: LO<sub>growth</sub>, which represents optimal biomass production during aerobic, glucose-limited growth, and LO<sub>ethanol</sub>, which corresponds to both maximal ethanol production and optimal growth during microaerobic conditions. The predictions of the *S. cerevisiae* PhPP and genome scale model were compared to independent experimental data, and the results showed strong agreement between the computed and measured specific growth rates, uptake rates, and secretion rates. Thus, genome-scale *in silico* models can be used to systematically reconcile existing data available for *S. cerevisiae*, particularly now that yeast resources, databases, and tools for global analysis of genomic data have been expanded and made publicly available, such as the Saccharomyces Genome Database (Cherry et al, 1998).

## 4.2 Genome-Scale Reconstructed Network Process

Given the relatively large number of genome-scale reconstructions now available (See Table 2 of **Paper V**), a robust methodology has been established for *de novo* model construction. There are numerous reviews describing the process of genome-scale network reconstruction, including the initial biochemical annotation performed, the mathematical framework employed for describing metabolism, the resulting system of linear differential equations, the assumptions and constraints required for simplification, and ultimately numerical solution methods (Förster et al, 2003a; Edwards et al, 2002a; Edwards et al, 2002b, Förster et al, 2003b). Although the history of quantitative flux balance analysis may be traced with roots in various fields, particularly if one considers the previously discussed isotope-labeled substrate experiments performed (1950s), it is widely accepted that flux balance analysis became more wide-spread in the 1960s and 1970s. Early attempts focused on specific enzyme kinetics, such as characterization of yeast pyruvate kinase *in vitro* for calculation of glycolytic flux under anaerobic cultivation conditions (Barwell et al, 1972). Furthermore, the work of Michael Savageau and other groups in the development of systems analysis of biochemical processes, the broader framework for what today is commonly referred to as Biochemical Systems Theory, was emerging during the 1960s through a series of seminal publications in the *Journal of Theoretical Biology* (Savageau, 1969a; Savageau, 1969b; Savageau, 1970). Along similar lines two independent research groups, i.e. Kacser and Burns (1973) (Kacser et al, 1973) and Heinrich and Rapoport (1974) (Heinrich et al, 1973; Heinrich et al, 1974a; Heinrich et al, 1974b; Rapoport et al, 1974), developed a mathematical framework for quantitative analysis of how flux control in metabolic pathways is distributed, a concept that today is referred to as Metabolic Control Analysis. However, it should be noted that flux balance analysis, particularly with genome-scale resolution, has largely been developed since the late 1990s with significant computational tools and methodologies aimed at extracting more predictive power from the collection of models available (Covert et al, 2001; Price et al, 2003; Patil et al, 2004; Borodina et al, 2005; Breitling et al, 2008).

With over thirty genome-scale metabolic models constructed, and numerous others proposed, it is relevant to investigate the fundamental approach to flux balance analysis. The simplified mathematical framework presented here has been adapted from an excellent presentation of flux balance analysis (Edwards et al, 2001). To better illustrate the power of this methodology, let's define a hypothetical metabolic system, composed of unique metabolites  $A$ ,  $B$ ,  $C$ ,  $D$ , and  $E$ . Let us also define a two compartment biochemical reaction space (compartments 1 and 2). The resulting metabolic space is pictorially represented by Figure 5 of **Paper V**. The reactions and stoichiometry are clearly defined, and included in the stoichiometry is annotation of the compartmentalization. For purposes of this example,  $R_1$ ,  $R_2$ ,  $R_5$ ,  $R_6$ , and  $R_7$  will be referred to as catalyzing reactions, while  $R_3$ ,  $R_4$ ,  $R_8$ ,  $R_9$  are considered exchange transport reactions. It's important to note that nomenclature and approach may vary in model construction, but for example, internal transport of metabolite  $A_1$  from compartment one to compartment two, results in a new metabolite,  $A_2$ , being defined. Consequently this reaction is defined here as catalysis, because  $R_5$  may be modeled as the depletion of  $A_1$  to form  $A_2$  even though these metabolites are not in fact chemically unique. Each of these reactions, in a genome-scale network reconstruction would be further annotated by assigning function to a specific open reading frame (ORF), and subsequently a comprehensive list of all reactions, metabolites, and their assigned ORF are reconstructed, including identifying those reactions and metabolites that are unique (e.g., independent of compartmentalization, and representing novel chemical entities and their catabolic or anabolic reactions). The methodology then employed is derived from the classical principles of chemical engineering, where essentially a mass balance is performed across a defined system boundary<sup>5</sup>. A mass balance approach for a given metabolite may be considered of the qualitative form:

$$In - Out + Generation - Consumption = Accumulation \quad \text{Equation 1}$$

From a biochemical reaction perspective the above mass balance may be formalized mathematically, into the expression:

$$\frac{dX_i}{dt} = V_{syn,i} - V_{deg,i} - V_{gro,maint,i} \pm V_{trans,i} \quad \text{Equation 2}$$

In Equation 2, the accumulation of metabolite  $X_i$  with respect to time, is defined as the rate of synthesis ( $V_{syn}$ ), minus the rate of degradation ( $V_{deg}$ ), minus the rate of consumption related to growth or maintenance of existing biomass ( $V_{gro,maint}$ ), and then plus or minus the rate of transport ( $V_{trans}$ ) across a defined biological boundary (e.g., membrane). Equation 2, presently in scalar format, may then be written in matrix format, yielding Equation 3.

$$\frac{dX}{dt} = S \cdot V \pm V_{trans} \quad \text{Equation 3}$$

<sup>5</sup> In modern educational terms, mass and energy balancing is most commonly associated with the field of chemical engineering; however, in a biological context flux balance analysis originates from earlier enzymatic characterization and biochemical pathway analysis, as discussed above. However, in historical terms it's worth noting that mass and energy balances extend from the laws of mass and energy conservation, which were originally formalized in the 1700s by chemist, John Dalton, and experimentally demonstrated by chemist Mikhail Lomonosov (Roscoe et al, 1895).

Equation 3 represents a mass balance across all metabolites in the biochemical reaction space considered, having concentration  $X$ , and then defining a vector of all the metabolic reactions,  $V$ , and a stoichiometric matrix,  $S$ . Biological time-scales associated with changes in metabolite concentrations are often very rapid, and significantly faster than times scales associated with growth (e.g., for *S. cerevisiae* the doubling time is about 2 hours). It is therefore reasonable to assume that the concentrations of all the intracellular metabolites are in a steady state, yielding Equation 4.

$$\mathbf{0} = S \cdot V \pm V_{trans}$$

Equation 4

Equation 4 may be further simplified by considering that the rate of transport of all metabolites,  $X$ , maybe reduced to a constant value equivalent to the net transport of metabolites into or out of the bioreaction space. This simplification, converts  $V_{trans}$  to a constant term,  $b$ , a vector representing the net exchange flux of metabolites. This constant value,  $b$ , for each metabolite, in matrix format is expressed in Equation 5, noting the use of the identity matrix,  $I$ .

$$\mathbf{0} = S \cdot V + b \cdot I$$

Equation 5

Prior to further simplifications, it's appropriate to take Equation 5 and apply it to the system described in Figure 5 of **Paper V**. The resulting stoichiometric matrix,  $S$ , and vector describing all of the metabolic reactions,  $V$ , are presented below as Equation 6. Furthermore, the vector,  $b$ , representing the net exchange fluxes for each metabolite is also represented in Equation 6 (for simplification purposes the identity matrix is not included since  $b \cdot I = b$ ). For clarity purposes, the top row of each matrix serves as a column header. In the case of the vector,  $b$ , while all metabolites are designated, only certain metabolites have a net transport, previously defined as  $R_3$ ,  $R_4$ ,  $R_8$ , and  $R_9$ . Those metabolites are designated with the corresponding transport reaction ( $R_n$ ) as depicted in Figure 5 (**Paper V**).

$$\begin{array}{c}
 \left| \begin{array}{ccccc} A_1 & 0 & 0 & -1 & 0 & 0 \\ B_1 & 0 & 0 & 0 & -1 & 0 \\ C_1 & 0 & -1 & 0 & 0 & 1 \\ D_1 & 0 & -1 & 0 & 0 & 0 \\ E_1 & 0 & 1 & 0 & 0 & 0 \\ A_2 & -1 & 0 & 1 & 0 & 0 \\ B_2 & -1 & 0 & 0 & 1 & 0 \\ C_2 & 1 & 0 & 0 & 0 & -1 \end{array} \right| \bullet \left| \begin{array}{c} R_9 \\ R_1 \\ R_2 \\ R_5 \\ R_6 \\ R_7 \end{array} \right| + \left| \begin{array}{c} b_n \\ (R_3) A1 \\ (R_4) B_1 \\ C_1 \\ (R_8) D_1 \\ (R_9) E_1 \\ A_2 \\ B_2 \\ C_2 \end{array} \right| = 0
 \end{array}$$

Equation 6

Equation 5 may be further rearranged, particularly focusing on separation of those metabolites that have a net exchange flux. This is readily accomplished by definition of a new vector,  $b_n$ , that will only include rows of metabolites that have a net exchange flux (e.g., not all metabolites are transported across the system boundary). Consequently, the columns of the identity matrix will be reduced to the same number of rows as vector  $b_n$ , and be renamed to  $I_r$ . Furthermore, the stoichiometric matrix,  $S$ , may be sub-divided to include the stoichiometry for reactions that are related to metabolic catabolism or anabolism resulting in the net accumulation or transport of a metabolite, defined as  $S_r$ , and into reactions that constitute biomass formation and maintenance,

defined as  $S_{b,m}$ . Lastly, the modified stoichiometric matrices are multiplied by the corresponding flux vectors,  $V_{b,m}$  and  $V_r$ . The final resulting equation is included below, as Equation 7.

$$\mathbf{0} = \mathbf{S}_r \cdot \mathbf{V}_r + \mathbf{S}_{b,m} \cdot \mathbf{V}_{b,m} + \mathbf{b}_t \cdot \mathbf{I}_t$$

Equation 7

Using more familiar matrix notation, Equation 7 may be re-formulated to Equation 8 below.

$$\mathbf{S}' = \langle \mathbf{S}_r | \mathbf{S}_{b,m} | \mathbf{I}_t \rangle$$

$$\mathbf{V}' = \begin{matrix} \mathbf{V}_r \\ \mathbf{V}_{b,m} \\ \mathbf{b}_r \end{matrix}$$

$$\mathbf{S}' \cdot \mathbf{V}' = \mathbf{0}$$

Equation 8

The above form (Equation 8) is what is most commonly used in the literature to represent the flux balance of a stoichiometrically defined bioreaction space, similar to what is provided in Figure 5 of **Paper V**. From this point forward, additional constraints that are often specific to the bioreaction space being considered and the organism are included. These considerations will include but not be limited to the metabolite and reaction compartmentalization, the reversibility of the reactions, the net biomass equation (e.g., summation of all metabolite precursors, redox co-factors, and energy co-factors in stoichiometric quantities), the theoretical minimum and maximum metabolite fluxes, the minimum and maximum growth rates, the amount of ATP (or equivalent energy currency) required for maintenance, and the amount of starting fluxes for input exchange fluxes (e.g., glucose uptake rate). Lastly, an objective function, to be maximized or minimized, must be defined and typically takes the form of Equation 9, where  $Z$  is an objective function equal to the summation of the product of a unit vector,  $q_i$ , and the metabolic fluxes,  $V_i$ , where  $q_i$  is typically the growth rate flux or glucose uptake rate. Both of these fluxes serve as suitable maximization parameters for modeling *in vivo* microbial metabolism where under conditions of excess nutrients and limited substrate, the specific growth rate of microbes,  $\mu$ , will approach  $\mu_{max}$ . Note, that included in Equation 9 are the constraints on metabolites,  $V_i$ , which typically range from a minimum ( $a$ ) to a maximum ( $b$ ).

$$Z = \sum q_i \cdot V_i$$

$$\mathbf{a} \leq \mathbf{V}_i \leq \mathbf{b}$$

Equation 9

The resulting system of linear equations, for a given objective function, may be solved using linear programming methods, for which several numerical solution packages are available. The result is a solution space, which may be represented on a minimum of two or maximum of three dimensions, from which a specific phenotypic phase place is defined. Figure 7 of **Paper V** highlights what is commonly referred to as the phenotypic phase plane (PhPP), where a two-dimension or three-dimension solution space is considered for a simulation where the maximization of an objective function is considered under specific constraints, such as the optimization of growth rate under a constant glucose uptake rate ( $q_{gluc}$ ), and oxygen uptake rate ( $q_{O_2}$ ).

The approach described here in flux balance analysis has been applied to numerous organisms as described in Table 3 of **Paper V**, and in particular, has served as a critical tool in metabolic engineering approaches, and more recently, systems biology. Systems biology is the quantitative characterization of genetic, transcription, protein, metabolic, signaling and other informational pathway responses to a clearly defined perturbation of a biological system. The perturbation may be in terms of a genetic, chemical, or environmental stimulus. At the core of systems biology is the transformation of quantitative, typically large-scale data sets, into *in silico* models that provide both interpretation and prediction. GSMMs provide a framework of how *x-ome* data may be organized and over-laid on the metabolic network. As technologies have become more accessible for transcriptome (DNA oligonucleotide and cDNA microarrays), proteome (two-dimensional gel electrophoresis coupled to MS or direct MS analysis), fluxome (isotopically labeled substrates coupled to detection by GC-MS), and metabolome (numerous analytical methods including LC-MS and GC-MS) measurements, enormous data sets have been generated that require bioinformatics and quantitative models to be developed for data analysis, interpretation, and prediction. Industrial biotechnology is beginning to exploit the benefits of these tools realizing that metabolic engineering strategies for improved process development may first be screened *in silico*, producing a reduced, specific, and high-probability of success list of genetic perturbations that should be experimentally validated. The process is highly iterative, with strain construction and characterization providing new *x-ome* data that can be used to improve the models (i.e., experimental quality control of *in silico*) and metabolic engineering strategies.

### 4.3 Industrial Systems Biology Applied

There have been extensive reviews and literature describing industrial biotechnology, noting several prominent case studies (Hermann et al, 2007a; Takors et al, 2007; Dodds et al, 2007; Hermann et al, 2007b; Hatti-Kaul et al, 2007). Industrial systems biology is prevalent in two forms – either existing companies are building their own infrastructure through reshaping in-house competences or forming new process development groups with industrial systems biology capabilities and expertise, or they are out-sourcing process development to small, recently formed entities that specialize in industrial systems biology. Examples of such enterprises focused on providing industrial systems biology expertise to more traditional process development groups include METabolic EXplorer (France, Founded in 1999), Genomatica (USA, Founded in 2000), Fluxome Sciences (Denmark, Founded in 2002), and Microbia Precision Engineering (USA, a subsidiary of Ironwood Pharmaceuticals, formerly Microbia). Although small, these companies have significant collaborations with many of the major chemical manufacturing, nutraceutical, pharmaceutical, and petrochemical companies.

Industrial systems biology, while in its infancy, has already had significant impact on tangible products produced using industrial biotechnology. Although several products may be presented as case studies, perhaps a more appropriate context to gauge the impact of industrial systems biology is to consider three broader product classes, providing a key example in each. Industrial biotechnology products may be categorized into the following cross-sections: mature and developed, recently launched and rapidly growing, and in-development. Products representing each cross-section include bioethanol, 1,3-propanediol, and succinic acid, respectively. All three of these products are considered leading examples within their classes, and all three have been significantly impacted from the application of systems biology for development of commercialized microbial cell factories. **Paper V** provides an extensive review of bioethanol (also covered in Otero et al, 2007) and 1,3-propandiol, and here there will only be given a review of succinic acid production, as this has been the focus of the thesis.

## 5.0 In-Development Industrial Biotechnology Product: Succinic acid

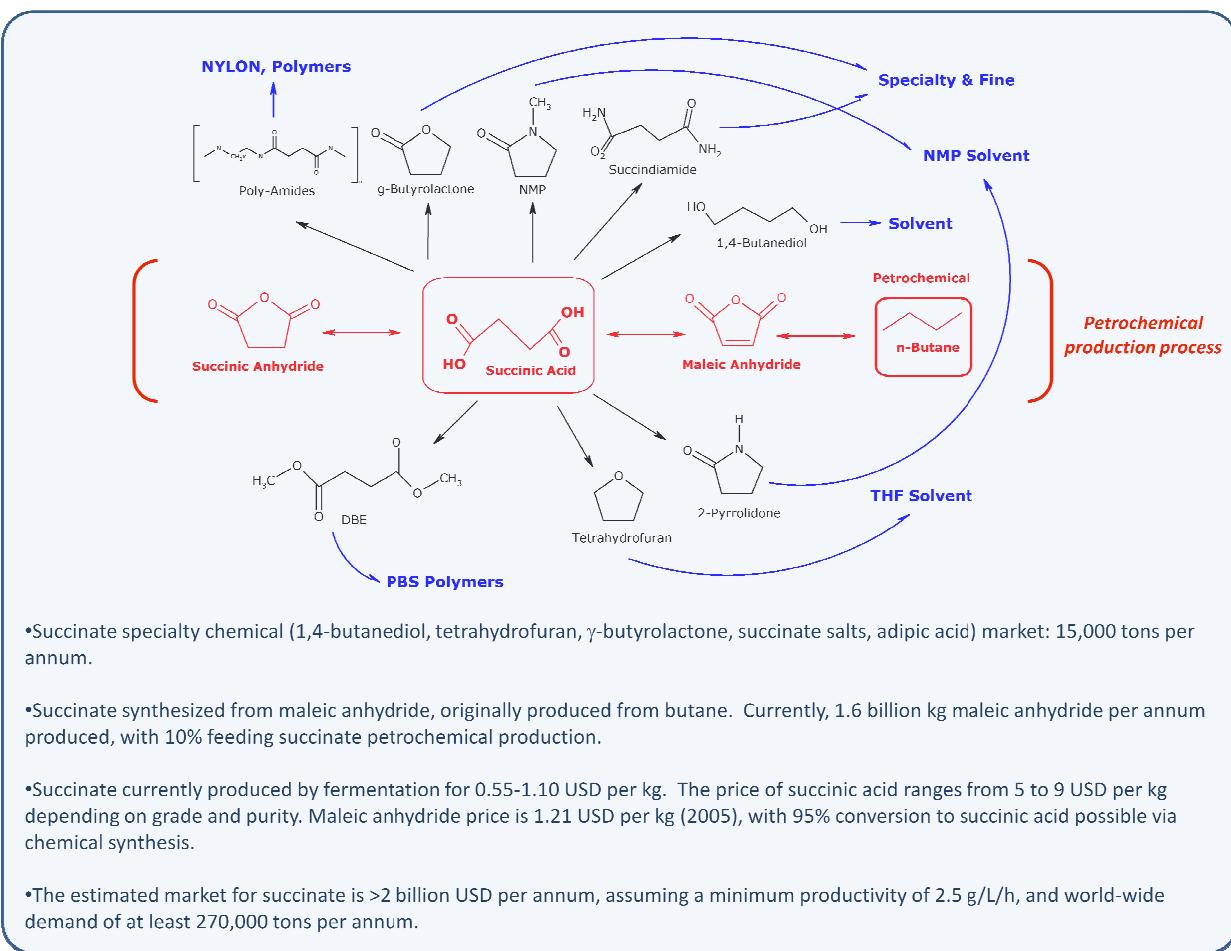
In 2004, based upon a critical analysis to identify the top building blocks that may be produced from biomass and subsequently converted to high-value bio-based chemicals, the US Department of Energy identified succinic acid ( $C_4H_6O_4$ ) as a top ten building block (US DOE, 2008). In the same year, 160 million kg of succinic acid were synthesized from petrochemical conversion of maleic anhydride. If bio-based succinic acid production becomes more commonplace global market demand is estimated to increase to 2 billion USD per annum with a total energy savings of 2.8 hundred thousand MWh year<sup>-1</sup>. Succinic acid is used in a variety of products and serves as a critical starting material or intermediate in the production of useful chemicals for solvents and polymers (See Figure 6).

Succinic acid is utilized in four major markets. First, and foremost, it is used as an additive in surfactants, detergents, extenders and foaming agents. Second, it is used as an ion chelator to prevent corrosion and pitting of metals that undergo electroplating. Third, it is widely used in the food industry as an acidulant/pH modifier, flavoring agent, and anti-microbial agent. Finally, it is used in a wide-variety of pharmaceutical products including therapeutics, antibiotics, amino acids, and vitamins. In 1999, the total amount of succinic acid produced was only 15K tons, with a total market size per annum of ~\$400 million (Zeikus et al, 1999). Succinic acid may serve as a precursor for high value-added specialty chemicals, including 1,4-butanediol (industrial solvent and raw material for polybutylene terephthalate resins), adipic acid (precursor to nylon), and tetrahydrofuran (solvent used in adhesives, printing inks, and magnetic tapes). Table 1 provides a detailed summary of the various chemical classes that succinic acid serves as either a feedstock or key intermediate for. Traditionally, the bulk of succinic acid has been produced by petrochemical conversion of maleic anhydride (Zeikus et al, 1999). In 2004, ~10% of the world's maleic anhydride capacity (1.6 million T year<sup>-1</sup>) was converted to succinic acid – 175K tons. This represents a 12-fold increase in the world-wide succinic acid demand over the previous 6 years (Kleff, 2004).

Succinic acid ( $C_4H_6O_4$ , MW 118.09 g mol<sup>-1</sup>,  $pK_{a_1}$  4.21,  $pK_{a_2}$  5.72, CAS Number: 110-15-6) is a polyprotic, dicarboxylic acid that occurs in nature. Succinate, the anion of succinic acid, is a citric acid cycle intermediate produced from the GTP coupled oxidation of succinyl-CoA by succinyl-CoA synthetase, and in many cases, as a by-product of the isocitrate lyase catalyzed conversion of isocitrate to glyoxylate. Succinic acid is then further oxidized to fumarate by succinate dehydrogenase, co-producing FADH<sub>2</sub>. There are numerous biomass based production platforms, all prokaryotic, including *Anaerobiospirillum succiniciproducens*, *Actinobacillus succinogenes*, *Succinivibrio dextrinosolvens*, *Corynebacterium glutanicum*, *Prevotella ruminocola*, a recently isolated bacterium from bovine rumen, *Mannheimia succiniciproducens*, and a metabolically engineered succinic acid over-producing *E. coli*. There have been several extensive reviews that detail the succinic acid market, and more specifically, comprehensively present the various metabolic engineering strategies coupled with application of systems biology that have been employed to date (Zeikus et al, 1999; Song et al, 2006; McKinlay et al, 2007; Jantama et al, 2008). Table 2 provides a short summary of the major microbial production hosts demonstrated to overproduce succinate. The two organisms that have been most significantly engineered from native isolations are *E. coli* and *M. succiniciproducens*.

*Anaerobiospirillum succiniciproducens* and *Actinobacillus succinogenes* have both been explored as platform technologies for fermentation based succinate production. *A. succiniciproducens* is a Gram-negative, obligate anaerobe that produce ~120 mol-succinate 100 mol-glucose<sup>-1</sup>; with further increases in yield when feed stocks such as whey are used (Lee et al, 2000). However, oxygen-free fermentations at industrially relevant scales are difficult to achieve, and even small levels of oxygen may cause genetic instability (Lin et al, 2004). *A. succinogenes* utilize a wide range

of sugar substrates and are not obligate anaerobes. Their volumetric yield is  $110 \text{ g l}^{-1}$  compared to  $65 \text{ g l}^{-1}$  by *A. succiniciproducens* (Zeikus et al, 1999). Both organisms utilize PEP carboxylase as their primary succinate producing pathway (See Figure 6). Productivity is further regulated by  $\text{CO}_2$  levels, where low  $\text{CO}_2$  levels ( $10 \text{ mol-CO}_2 / 100 \text{ mol-glucose}^{-1}$ ) will promote lactic acid production in *A. succiniciproducens*, and ethanol in *A. succinogenes*. At high  $\text{CO}_2$  levels ( $100 \text{ mol-CO}_2 / 100\text{-mol glucose}^{-1}$ ), succinate is the major product and only trace amounts of lactic acid or ethanol are produced in both organisms (Zeikus et al, 1999). At industrially relevant scales, it may be difficult to co-regulate dissolved  $\text{CO}_2$  and dissolved  $\text{O}_2$ . Furthermore, metabolic engineering or genetic manipulations are limited in each of these organisms due to the lack of a publicly available genome sequence and a GSSM for gene target identification.



**Figure 5: Succinic Acid – A Sustainable Building Block Chemical.** The above figure illustrates the diverse chemical synthesis opportunities offered by using succinic acid and succinic anhydride as a building block chemical. Presently, succinic acid is primarily used in the specialty chemical sector amounting to a total annual demand of 15,000 tons, and is produced from the conversion of maleic anhydride. Approximately 10% of the total maleic anhydride produced (1.6 billion kg per annum) is dedicated towards succinate production. Maleic anhydride is widely used in housing, construction, and automotive manufacturing industries, which as a result of their recent strong demand, has caused the price to increase from 0.95 USD per kg (2003) to 1.21 USD per kg (2005). This price increase has resulted in increased petrochemical maleic anhydride production facilities, where the raw material, *n*-butane is utilized. Consequently, there has been significant price pressure to design a cost-effective succinic acid industrial biotechnology process in the face of increasing maleic anhydride supplies. It's estimated that the market for succinate is in excess of 2 billion USD per annum, assuming a minimum productivity of  $2.5 \text{ g l}^{-1} \text{ h}^{-1}$  may be reached via fermentation.

More recently, significant progress in genetic reconstruction of the aerobic central metabolism of *Escherichia coli* has been accomplished. Under anaerobic fermentation *E. coli* produces relatively small amounts of succinate – 45 g l<sup>-1</sup> volumetric yield. Some genetic manipulations, including overexpression of phosphoenolpyruvate carboxylase (*PEPC*), overexpression of fumarate reductase (*FRD*), and overexpression of malic enzyme in the presence of inactivated pyruvate formate lyase (*PFL*) and lactate dehydrogenase (*LDH*) all resulted in improved succinate yields (Lin et al, 2004). However, all genetic manipulations were performed in an anaerobic fermentation model. It has been proposed that a dual-phase fermentation, encompassing an aerobic biomass generation phase, followed by anaerobic succinate production would be feasible; however, not practical (Vemuri et al, 2002a; Vemuri et al, 2002b). Aerobic fermentations often result in better carbon flux to biomass and faster cell growth yielding higher specific productivity. Furthermore, a dual-phase fermentation would likely result in electron-carrier imbalance, given that 2 mol-NADH mol-succinate<sup>-1</sup> are required and oxidative TCA cycle metabolism is largely responsible for replenishing NADH pools. However, under aerobic fermentation the primary carbon product is acetate. Therefore, Lin et al., in August 2004, reported genetic reconstruction of an aerobic central carbon metabolism in *E. coli* based upon glycolysis, TCA cycle, and the glyoxylate by-pass (See Figure 7). Theoretically, the maximum yield of succinate is 1.0 mol-succinate mol-glucose<sup>-1</sup>. Based upon five genetic mutations, *ΔsdhAB* (succinate dehydrogenase complex), *Δicd* (isocitrate dehydrogenase), *ΔpoxB* (pyruvate oxidase), *Δ(ackA-ptp)* (acetate kinase-phosphotransacetylase), and *ΔiclR* (glyoxylate operon *aceBAK* repressor), a succinate yield of  $0.304 \pm 0.03$  mol mol-glucose<sup>-1</sup> (wildtype,  $<0.05$  mol mol-glucose<sup>-1</sup> yield) was achieved while maintaining growth rates and final biomass levels comparable to wild-type. In bioreactor studies this yield increased to 0.7 mol mol-glucose<sup>-1</sup> (22 mM-succinate 63 mM-glucose<sup>-1</sup>). Additional increases in yield would be possible if accumulation of TCA cycle C<sub>6</sub> intermediates could be minimized. In a subsequent study from the same group, a four-mutation strain (precursor to the five mutation strain) was further evaluated – *ΔsdhAB*, *ΔpoxB*, *Δ(ackA-ptp)*, and *ΔiclR*. The primary difference between the two strains was inactivation of the phosphotransferase gene (*ptsG*), whose gene product catalyzes the reaction of glucose to glucose-6-phosphate, converting phosphoenolpyruvate to pyruvate as a by-product. Also, in the four-mutant strain there was overexpression of a *Sorghum* phosphoenolpyruvate carboxylase (*PEPC*), exhibiting resistance to malate feedback inhibition. The four-mutant strain, compared to the five-mutant strain, permitted aerobic succinate production via two pathways (TCA cycle and glyoxylate cycle) as opposed to when *ΔptsG* was inactivated and *PEPC* over-expressed (Lin et al, 2005) (See Figure 7). A succinate yield of 1.0 mol mol-glucose<sup>-1</sup> in aerobic fermentation was achieved, with an average specific productivity of 44.26 mg-succinate g-biomass<sup>-1</sup> h<sup>-1</sup> (Lin et al, 2005).

*E. coli* fermentations are controlled at pH ~7.0 using a dual-sided pH control strategy (e.g., 1.5N HNO<sub>3</sub> and 2N Na<sub>2</sub>CO<sub>3</sub>) (Lin et al, 2005). Succinic acid has pKa<sub>1</sub> and pKa<sub>2</sub> of 4.21 and 5.72, respectively, at 25°C (Sigma-Aldrich Product Information, 2005). Therefore, *E.coli* fermentations are not preferred for direct production of the associated acid, and would require additional unit operations for direct production of succinic acid. Furthermore, eukaryotes may be a preferred expression system to prokaryotes because of their resistance to phages and general increased robustness in industrial environments.

**Table 1: Key Succinic Acid Markets**

PRODUCT	CHEMICAL CLASS	PRODUCT DESCRIPTION	US ANNUAL MARKET SIZE (Limited availability)
Tetrahydrofuran (THF) <sup>1</sup>	Solvents, Adhesives, Inks	THF is a solvent and key ingredient of adhesives, printing inks, magnetic tape, thermoplastic urethane elastomers, and polyurethane fibers.	Total is 116 million kgs. Biomass based production of succinic acid could displace >10.4 million kg per year, at 3.41 USD per kg. This represents a minimum total market of >75 million USD.
1,4-Butaediol (BDO) <sup>1</sup>	Solvents, Resins, Chemical Intermediate	BDO is used in solvents, coating resins, and as an intermediate for producing other solvents and chemicals.	Total is 408,000 tons (2004). Biomass based production of succinic acid could displace >6.4 million kg per year, at 0.71 USD per kg. The 2004 price for BDO was 2.76 USD per kg, representing a total market >1 billion USD per annum.
Succinic salts <sup>2</sup>	Deicing Compounds, Coolants, Herbicides	There is a growing need to improve the performance of runway and wing deicing while reducing environmental and corrosion impacts. Many of the existing products serving airport-deicing operations are being heavily regulated because of their environmental toxicity. Succinic salts could replace 100% of airport deicers. Succinate salts lower the freezing point of water. This property coupled with corrosion inhibition properties, make succinate salts candidates for glycol alternatives.	
Succinate esters <sup>3</sup>	Fuel Additives	Succinate esters are excellent fuel oxygenates. Incorporation of diethyl succinate (DES) in diesel fuel results in emission reductions depending on the fuel grade. The DES is fully miscible with diesel fuel and requires no co-solvents or additional additives.	
Disuccinate esters <sup>3</sup>	Solvents	There is a market demand for alternative solvents given the highly volatile and chlorinated structure of existing solvents. Replacement solvents should be biodegradable and pose little threat of air pollution or ozone damage. Disuccinate esters are "green" solvents that have performance and environmental benefits.	
Diester succinates, specialty surfactants and other proprietary compounds. <sup>3</sup>	Personal care products	Diester succinates, specialty surfactants and other proprietary compounds are ingredients for the personal care "nature product" sector. Target products include nail polish remover, shampoos, and creams. The nail polish remover is safe, biodegradable, and non-volatile, eliminating toxic fumes and undesirable odors.	

1. Industrial Bioproducts: Today and Tomorrow (2003) US Department of Energy, Office of Energy Efficiency and Renewable Energy, Office of the Biomass Program, Washington, DC, USA.
2. Diversified Natural Products, Inc (2006) Available at <http://www.dnpco.com> (last accessed MAR 2006).
3. Brown, Robert C (2003) Bio-renewable Resources Engineering New Products from Agriculture. Iowa State Press, Ames, IA.

*M. succiniciproducens* MBEL55E is a capnophilic Gram-negative bacterium first isolated in 2002 from a bovine rumen in Korea that natively accumulates large amounts of succinic acid under glucose supplemented anaerobic (100% CO<sub>2</sub>) fermentation conditions (0.68 g-succinic acid g-glucose<sup>-1</sup>) (Lee et al, 2002). Shortly following the isolation, classical batch and continuous fermentation of sodium hydroxide treated wood hydrolysate was examined and resulted in a succinic acid productivity of 1.17 g l<sup>-1</sup> h<sup>-1</sup> (yield: 56%) and 3.19 g l<sup>-1</sup> h<sup>-1</sup> (yield: 55%), respectively (Lee et al, 2003). These were certainly the highest productivities reported at the time, and were particularly promising given the lignocellulosic feedstock used (Mixed substrate glucose and xylose, batch and continuous cultivations were also performed as controls, with similar productivities and yields resulting). In the same year, the 2,314,078 base pair genome sequence of *M. succiniciproducens* MBEL55E was reported co-currently with the genome-scale reconstructed metabolic network (Hong et al, 2004). The genome-scale reconstructed metabolic network, consisting of 373 reactions (121 reversible and 252 irreversible) and 352 metabolites, predicted, using MFA, a theoretical production of 1.71 and 1.86 moles of succinic acid for every mole of glucose under CO<sub>2</sub> and CO<sub>2</sub>-H<sub>2</sub> atmospheres, respectively (Hong et al, 2004). As a consequence of the simulations, they note, "Based on these findings, we now design metabolic engineering

strategies for the enhanced production of succinic acid; one such strategy will be increasing the PEP carboxylation flux while decreasing the fluxes to acetic, formic, and lactic acid.” (Hong et al, 2004). In 2006, the authors constructed a series of knock-out mutants of *M. succiniciproducens* MBEL55E that included disruption of three CO<sub>2</sub> catalyzing reactions (PEP carboxykinase, PEP carboxylase, malic enzyme) and disruption of four genes responsible for by-product formation of lactate, formate, and acetate ( $\Delta ldhA$ ,  $\Delta pflB$ ,  $\Delta pta$ , and  $\Delta ackA$  genes) (Lee et al, 2006a). Their results confirmed that a mutant capable of virtually no lactate, fumarate, or acetate formation was feasible, and that PEP carboxykinase was most critical for anaerobic growth and maximizing succinic acid production (Lee et al, 2006a). The resulting metabolically engineered strain, *M. succiniciproducens* LPK7 under batch fermentation conditions produced 0.97 mol-succinate mol-glucose<sup>-1</sup>, and under fed-batch fermentation conditions reached a maximum titer, productivity, and yield of 52.4 g l<sup>-1</sup>, 1.8 g l<sup>-1</sup> h<sup>-1</sup>, and 1.16 mol-succinate mol-glucose<sup>-1</sup>, respectively (Lee et al, 2006a). The theoretical carbon yield of succinate under excess reducing power and CO<sub>2</sub> carboxylation, is 2 mol-succinate mol-glucose<sup>-1</sup> ( $\Delta G^{\circ} = -317 \text{ kJ/mol}$ ) (McKinlay et al, 2007).

In 2006, which constituted one of the first examples of proteomics applied to industrial biotechnology process development, a proteome analysis of *M. succiniciproducens* was reported (Lee et al, 2006b). Using two-dimensional electrophoresis coupled with mass spectrometry, identification and characterization of 200 proteins distributed across whole cellular proteins (129), membrane proteins (48), and secreted proteins (30), was described. Characterization of cell growth and metabolite levels in conjunction with proteome measurements during the transition from exponential to stationary growth was carried out. Two interesting conclusions could be drawn from such analysis that was not possible *a priori*. First, a gene locus previously annotated as the succinate dehydrogenase subunit A (*SDHA*) is likely to be the fumarate reductase subunit A (*FRDA*) based on comparative proteome analysis supported by physiological data. Second, two novel enzymes were identified as likely metabolic engineering targets for future improvements in succinic acid production. PutAp and OadAp are enzymes responsible for acetate formation and conversion of oxaloacetate to pyruvate, respectively, and their deletion is likely to induce higher flux towards succinic acid through minimization of by-product formation (Lee et al, 2006b). This is a clear example of where proteome measurement and analysis not only provided novel information for future metabolic engineering strategies, but also served as a quality-control check for two critical assumptions. First, that genome annotation is error-free, and second, that mRNA expression directly correlates with protein expression and activity.

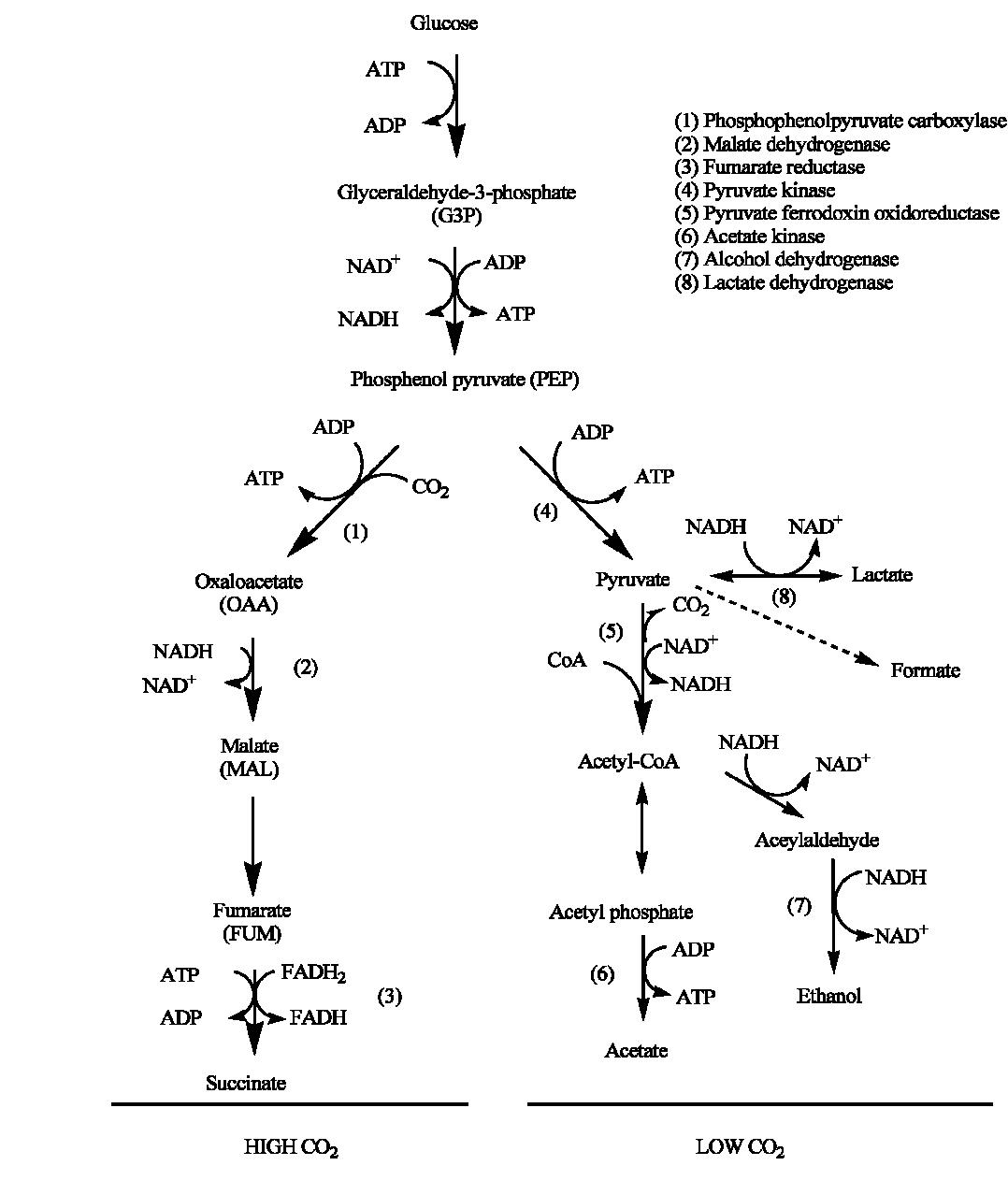
Most recently, in 2007, an updated genome-scale reconstructed network of *M. succiniciproducens* was presented that included 686 reactions and 519 metabolites based on reannotation and validation experiments (Kim et al, 2007). The refined reconstructed network, in conjunction with constraints-based flux analysis, was verified using comparative experimental data of the maximum specific growth rate and metabolic production formation rate for various MBEL55 mutants. In all simulation cases, the maximum specific growth rate was correctly predicted while the rate of succinic production, for a fixed glucose uptake rate, was in relatively good agreement (between 7.8 and 30.4%, depending on the genotype simulated *in vivo*). The model was further used to evaluate additional gene-deletion strategies likely to improve succinic acid production, and simulations were compared to strategies previously reported in genome-scale simulation of the *E. coli* reconstructed metabolic network (Kim et al, 2007; Burgard et al, 2003). Figure 8 provides a comparative analysis of the central carbon metabolism of *M. succiniciproducens* and *E. coli*. The comparative analysis of both genome-scale model simulation results suggested that the positive effect of various gene deletions on succinic acid production was more pronounced in *M. succiniciproducens* compared to *E. coli*, and that the metabolic performance, defined as the absolute flux of succinic acid production, was higher in *M. succiniciproducens* resulting from the higher observed glucose consumption rate under anaerobic conditions (Kim et al, 2007).

In approximately five years (2002-2007), a previously unknown microbe, *M. succiniciproducens*, was transformed into a leading microbial cell factory candidate for succinic acid production, as a result of the thorough application of systems biology tools: genome sequencing, genome-scale metabolic network reconstruction, fluxomics, proteomics, and subsequent model revision. It should be noted that similar approaches for *E. coli* and *A. succiniciproducens* have been reported; however, given the relative lack of *a priori* knowledge, short development time, and diversity of *x-ome* data collected and integrated, *M. succiniciproducens* remains a prominent example of successfully applied industrial systems biology.

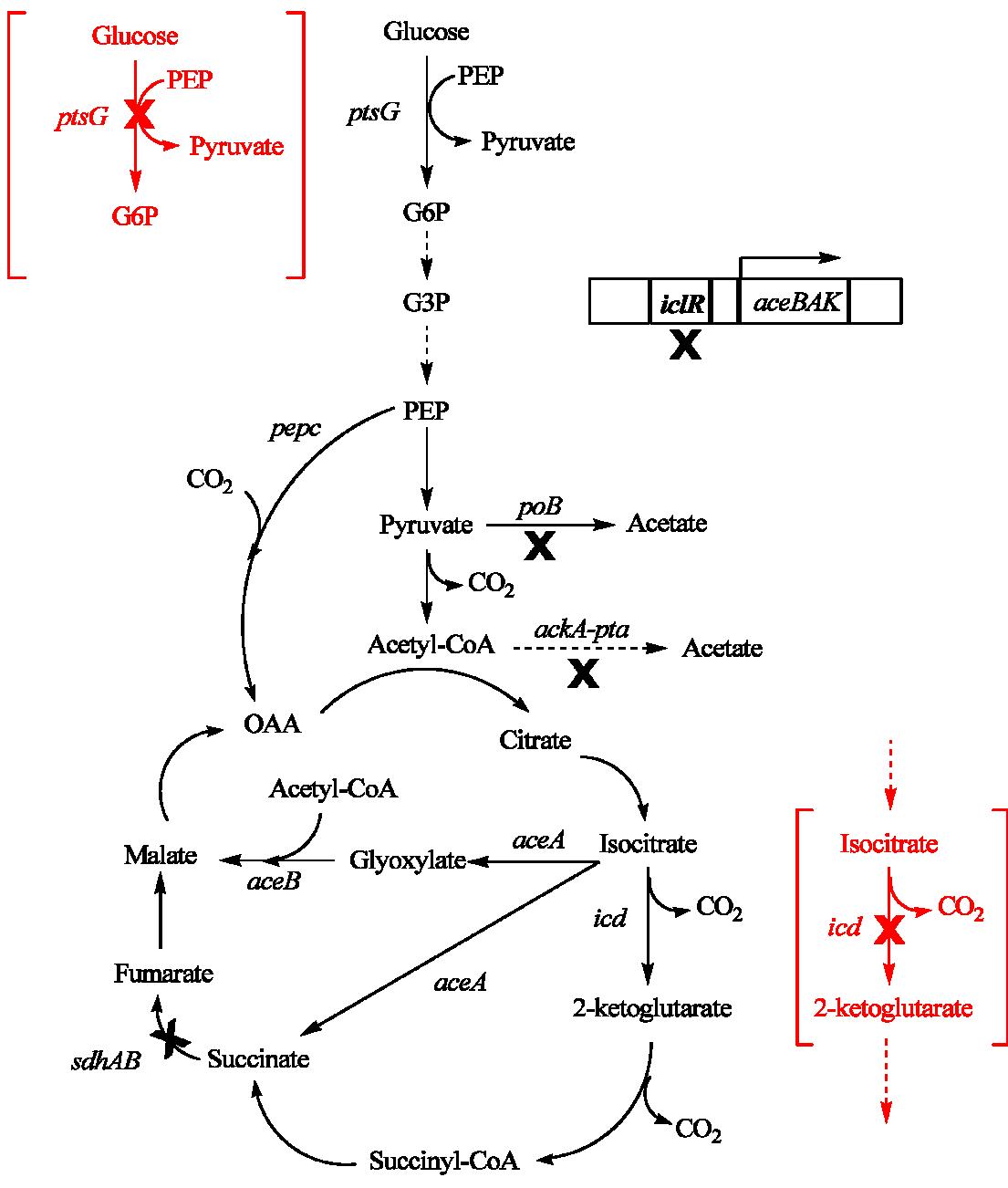
**Table 2: Summary of Major Microbial Production Platforms of Succinic Acid**

SOURCE	ORGANISM	FERMENTATION	YIELD <sup>A</sup> (g/g)	PRODUCTIVITY <sup>A</sup> (g/L/H)	TITER <sup>A</sup> (g/L)
[1]	<i>Actinobacillus succinogenes</i> , <i>Anaerobiospirillum succiniciproducens</i>	Obligate anaerobe, batch, corn steep liquor, yeast extract, pH 7	0.8	2.8	106
[2]	<i>Escherichia coli</i>	Dual-phase (an-/aerobic) fed-batch, yeast extract, tryptone, pH 7	1.1	1.3	99
[3]	<i>Escherichia coli</i>	Aerobic fed-batch, yeast extract, tryptone, pH 7	1.1	1.2	40
[4]	<i>Mannheimia succiniciproducens</i>	Facultative anaerobe, fed-batch, yeast extract, polypeptone, pH 6.1	0.7	3.9	52

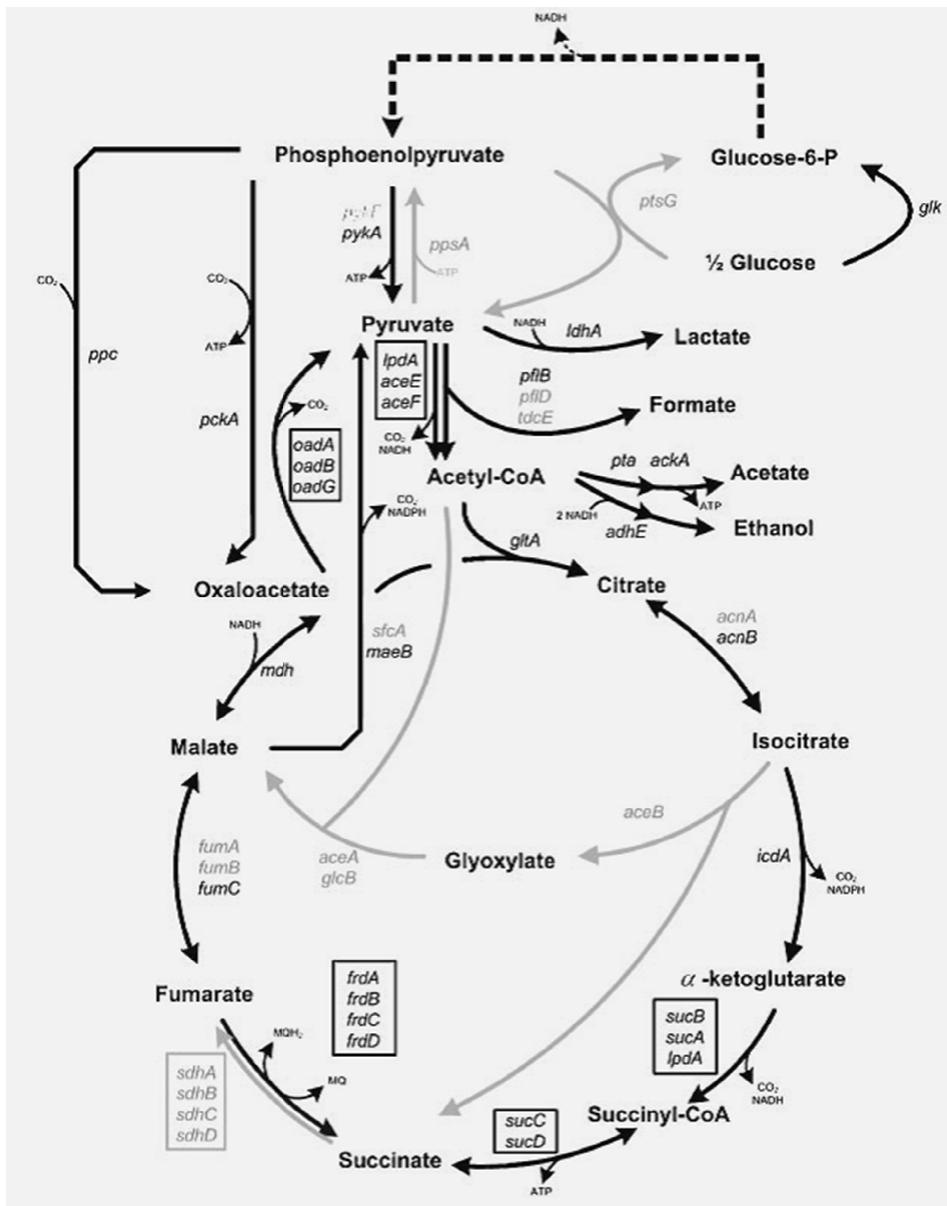
- A. The values of yield, productivity, and titer provided are the highest reported corresponding to the fermentation condition described. An in-depth review of the most current bio-based succinate production systems is available at: McKinlay JB, Vieille C, Zeikus JG “Prospects for a bio-cased succinate industry,” *Applied Microbiology and Biotechnology* vol. 76, no. 4, pp. 727-740, 2007.
1. Yedur S, Berglund KA, Dunuwila DD “Succinic acid production and purification,” US Patent No. 6,265,190; Filed Sep. 2, 1999; Issued Jul. 24, 2001.
  2. Vemuri GN, Eiteman MA, Altman E “Succinate production in dual-phase *Escherichia coli* fermentations depends on the time of transition from aerobic to anaerobic conditions,” *Journal of Industrial Microbiology and Biotechnology* vol. 28, no. 6, pp. 325-332, 2002.
  3. Sánchez AM, Bennett GN, San KY “Novel pathway engineering design of the anaerobic central metabolic pathway in *Escherichia coli* to increase succinate yield and productivity,” *Metabolic Engineering* vol. 7, no. 3, pp. 229-239, 2005.
  4. Lee SJ, Song H, Lee SY “Genome-based metabolic engineering of *Mannheimia succiniciproducens* for succinic acid production,” *Applied Environmental Microbiology* vol. 72, no. 3, pp. 1939-1948, 2006.



**Figure 6: Catabolic Pathway for Glucose Fermentation by *A. succiniciproducens* and *A. succinogenes*.** Succinate volumetric yield is 110 g/L and 65 g/L in *A. succinogenes* and *A. succiniciproducens*, respectively, and regulated by dissolved CO<sub>2</sub>. During high CO<sub>2</sub> concentrations, succinate is the major product, with minimal carbon directed towards lactic acid and ethanol. However, for low CO<sub>2</sub> concentrations, lactic acid production is up-regulated in *A. succiniciproducens*, while ethanol is up-regulated in *A. succinogenes*. The enzymes catalyzing the reactions indicated above are denoted by (#) in the top right corner (Zeikus et al, 1999).



**Figure 7: Genetic Engineering of Central Carbon Metabolism in *E. coli* for Succinate Overproduction.** Two strains were evaluated. The first strain employed five principal mutations (indicated by the black and red), while a second strain employed four mutations (indicated by the black only, with an additional mutation for over-expression of a malate feedback inhibition resistant to PEPCK) (Lin et al, 2005).



**Figure 8: Central Carbon Metabolism of *M. succiniciproducens* and *E. coli*.** The central carbon metabolic map is directly adapted from Kim et al, 2007 (Kim et al, 2007, Figure 2). Gene names in black color indicate enzymes that are present in both *M. succiniciproducens* and *E. coli*. Grey colored genes names exist only in *E. coli*. Gene names included in boxes form a multi-subunit enzyme complex. In this figure *PCKA* is shown to catalyze PEP to oxaloacetate. *S. cerevisiae* does not have PEP carboxylase. Please section 7.0 for further discussions on the thermodynamic unfavorability of this reaction.

## 6.0 *S. cerevisiae* for Succinic Acid Production

There has been extensive research and development completed on succinic acid production in different prokaryotic organisms. This section intends to introduce the opportunities available for *S. cerevisiae* based succinic acid production, and describe both the advantages and challenges.

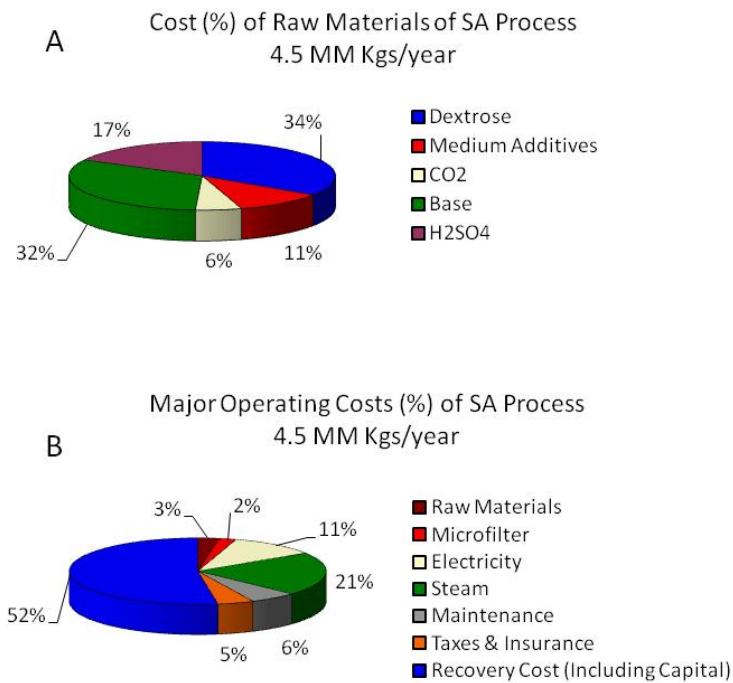
### 6.1 Motivations Driving a Succinic Acid Microbial Cell Factory

The significant work completed in prokaryotic hosts demonstrating high yields and productivities begs the simple question: why is metabolic engineering of succinic acid production in *S. cerevisiae* desirable? As with the majority of bio-based organic acid production previously described (Sauer et al, 2008), the overwhelming process cost is in downstream processing. In a strategic biorefinery analysis commissioned by the US Department of Energy (DOE), a theoretical biorefinery operation with bioethanol as the large volume commodity chemical and succinic acid as an added value co-product was considered. As stated previously, all prokaryotic organisms currently considered for bio-based production organisms (See Table 2) grow at neutral pH, thereby producing the succinate salt ( $pK_{a_1}$  and  $pK_{a_2} < 6$ ). Therefore, ammonia addition to the fermentation broth is required to form diammonium succinate, and subsequent sulfuric acid addition forms ammonium sulfate and succinic acid. The succinic acid is then polished by recrystallization using methanol. This purification, considered in the US DOE strategic analysis relies on two physical attributes: (1) succinic acid has limited solubility in water in the presence of sulfuric acid, thereby enabling separation of sulfates and succinic acid; and, (2) succinic acid is soluble in methanol while sulfates are not (Lynd et al, 2005).

Assuming a selling price of 2.68 USD, a total annual production of 45 MM kg year<sup>-1</sup>, a succinate yield on glucose of 0.9 g g<sup>-1</sup>, and succinate yield on dry feedstock of 0.4 g g<sup>-1</sup> it was determined that purification costs accounted for 39% of total operating costs. This represents a best case scenario and if an identical analysis considered for a total annual production of 4.5 MM kg year<sup>-1</sup> then purification costs are 52% of total operating costs. As discussed previously, over the long-term of any industrial biotechnology process, particularly for added-value and commodity chemicals, the raw materials become the dominant operating cost. Sulfuric acid and base required to operate at 4.5 MM kg year<sup>-1</sup> represents 49% of the total raw materials cost estimate, of which 17% is sulfuric acid (Lynd et al, 2005). Figure 9 provides a cost break-down of the total operating costs and raw material costs for the 4.5 MM kg year<sup>-1</sup> simulations.

*S. cerevisiae* as previously stated, and noted in **Papers I, II, III, and IV**, tolerates a relatively low pH range (3.5-5.50) with minimal impact on biomass growth, permitting direct expression of the associated succinic acid. Furthermore, the capability to operate a fermentation process at low pH significantly reduces the risk of contamination, of particular importance when considering large scale processing and the fact that steam represents the second largest fraction of the total operating cost after recovery. As previously described, many succinic acid derived final products and intermediates are used in human consumption, thereby we can leverage the Generally Regarded As Safe status by the US Food & Drug Administration of *S. cerevisiae*. Lastly, the ability of *S. cerevisiae* to grow on diverse feedstocks, including lignocellulose and its primary components (xylose and glucose), is a clear requirement to enable biorefinery flexibility. *S. cerevisiae* does not natively consume what is the second most abundant monosaccharide after glucose and the principal component of lignocellulose, xylose. The work described in **Paper III** further describes the benefits of metabolically engineering a *S. cerevisiae* strain for fast and efficient utilization of xylose with minimization of over-flow metabolites such as ethanol, acetate, glycerol, pyruvate, and xylitol (specifically observed only during xylose consumption), and maximization of carbon to biomass. It was anticipated that any successful microbial cell factory overproducing

succinate would employ a metabolic strategy coupling succinate production and biomass formation.



**Figure 9: Estimated Raw Materials and Operating Costs of a Bio-based Succinic Acid Production Process.** The above figures were reconstructed from the Strategic Biorefinery Analysis: Analysis of Biorefineries, commissioned by the US DOE during January 24, 2002 – July 1, 2002. This analysis considered the production of 4.5 MM kgs year<sup>-1</sup> of bio-based prokaryotic succinic acid production as an added-value co-product stream from a biorefinery producing commodity bioethanol. (A) The fractional cost, in percent, of the five major raw materials. (B) The fractional cost, in percent, of the major operating cost categories.

A final motivation, more holistically captured in **Paper V**, is the academic value in exploring the feasibility of redirection of carbon from C<sub>6</sub> substrates, such as glucose, to C<sub>4</sub> metabolic products, such as succinate. Central carbon metabolism in *S. cerevisiae* has been extensively investigated using a wide variety of substrates for determination of how glycolytic flux is distributed across C<sub>1</sub> (CO<sub>2,g</sub>), C<sub>2</sub> (ethanol, acetate), and C<sub>3</sub> (glycerol, pyruvate) products. For the *S. cerevisiae* CEN.PK113-7D strain cultivated under carbon-limited, aerobic, well-controlled batch fermentations, the distribution of carbon across biomass (CH<sub>1.596</sub>O<sub>0.396</sub>N<sub>0.216</sub>P<sub>0.0017</sub>S<sub>0.0024</sub>)<sup>6</sup>, C<sub>1</sub> (CO<sub>2</sub>), C<sub>2</sub> (ethanol, acetate), and C<sub>3</sub> (glycerol, pyruvate) products is 0.18, 0.14, 0.54, and 0.09 C-mol C-mol-glucose<sup>-1</sup>, respectively, with about 0.05 C-mol C-mol-glucose<sup>-1</sup> unaccounted for (data generated in this thesis for *S. cerevisiae* CEN.PK113-7D, 20 g l<sup>-1</sup> supplemented, aerobic batch fermentation). *S. cerevisiae* offers the unique advantage of being the most well characterized eukaryotic expression system, and therefore serves as a perfect model for exploring metabolic engineering and regulation of central carbon metabolism pathways, including those pertaining to succinate, and conserved across nearly all prokaryotes and eukaryotes.

<sup>6</sup> *S. cerevisiae* elemental composition determined while growing at glucose-limited conditions with excess nitrogen (MW 23.57 g C-mol-biomass<sup>-1</sup>) (Nielsen et al, 2003).

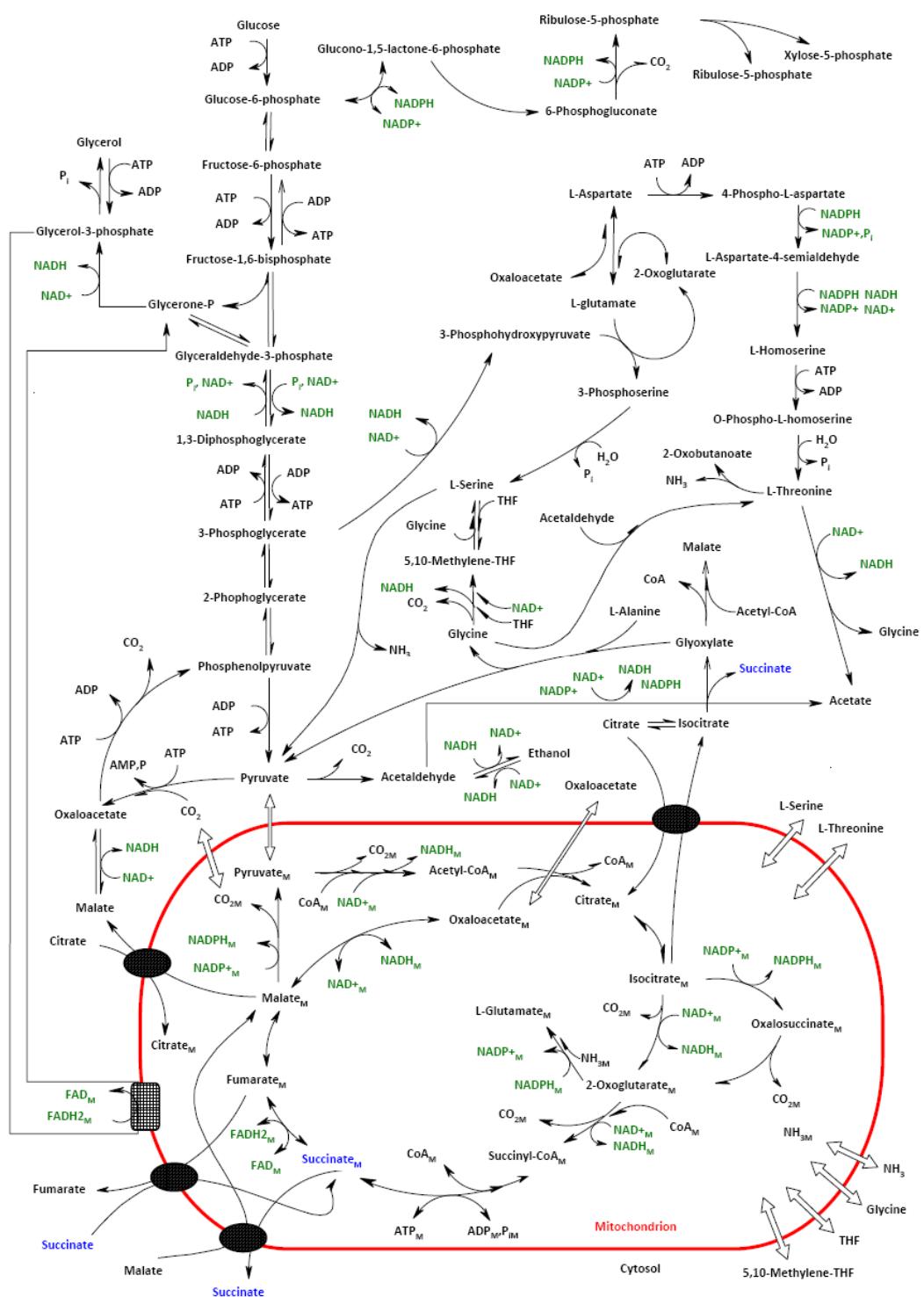
## 6.2 Succinic Acid Production in *S. cerevisiae* – Prior Work

*S. cerevisiae* primary carbon metabolism is governed by uptake of C<sub>6</sub> based sugars (e.g., glucose, galactose, fructose), and converting carbon to biomass, carbon dioxide, and C<sub>2</sub> and C<sub>3</sub> primary metabolites such as ethanol, acetate, glycerol, and pyruvate. Figure 10 provides an overview of the central carbon metabolism of *S. cerevisiae*. Relatively little research has been completed in an attempt to genetically modify *S. cerevisiae* to over-produce succinate – a C<sub>4</sub> organic acid TCA cycle intermediate.

Succinic acid production in genetically modified sake yeast strains has been demonstrated for modification of taste profiles, primarily focusing on multi-gene deletions of citric acid cycle enzymes aconitase (Aco1p), fumarate reductase (Osm1p),  $\alpha$ -ketoglutarate dehydrogenase (Kgd1p), fumarase (Fum1p), and succinate dehydrogenase (Sdh1p), resulting in <0.7 g l<sup>-1</sup> succinic acid on complex medium (Arikawa et al, 1999a; Arikawa et al, 1999b; Kubo et al, 2000). Much of this research has been led by the Food Technology Research Institute Nagano (Tokyo, Japan) where sake (Japanese rice wine) taste is highly influenced by organic acids including succinate. There has also been significant experimental work focused on elucidating the physiological role of cytosolic and mitochondrial fumarate reductase (Frds1p and Osm1p, respectively) in the context of facilitating anaerobic fermentation of *S. cerevisiae* (Camarasa et al, 2007; Arikawa et al, 1998; Enomoto et al, 2002). Significant effort has been applied to understand succinate formation in *S. cerevisiae* by exploring  $\Delta$ sdb1 and  $\Delta$ sdb3 deletion mutants, specifically using <sup>13</sup>C-NMR analysis of <sup>13</sup>C-labelled aspartate and glutamate supplemented anaerobic glucose fermentations, and DNA microarray analysis of aerobic and anaerobic glucose supplemented fermentations, respectively (Camarasa et al, 2003; Cimini et al, 2009). In both efforts, no significant succinate accumulation was observed through simple deletion of the primary succinate consuming reaction catalyzed by the succinate dehydrogenase complex. These efforts are referenced and further discussed in the contexts of **Paper I** and **Paper II**.

From an *in silico* approach, there has been one publication focusing on application of flux balance analysis (FBA) with the genome-scale metabolic network reconstruction of *S. cerevisiae* using an evolutionary programming method to couple biomass and succinate production (Patil et al, 2005). This approach highlighted several multi-gene deletion strategies for succinic acid overproduction, however, included no experimental validation of target predictions. Furthermore, this approach used a reduced genome-scale metabolic network reconstruction of iFF708, removing all duplicate and dead-end reactions. The limitations of this approach are discussed in **Paper II**.

Although not specific to succinic acid, an elegant example of malic acid production, that included engineering of pyruvate carboxylation (overexpression of PYC2), oxaloacetate reduction (overexpression of MDH3), and malate export (functional expression of the non-native SpMAE1), resulted in a *S. cerevisiae* strain capable of producing 59 g-malate l<sup>-1</sup> and 0.42 mol-malate mol-glucose<sup>-1</sup> (Zelle et al, 2008). Malate is an intermediate of the citric acid cycle resulting from the oxidation of fumarate to malate catalyzed by fumarate hydratase (encoded by FUM1). Malate is then further oxidized to oxaloacetate catalyzed by malate dehydrogenase co-producing a net 1-mole of NADH (encoded by MDH1 in the mitochondrion) (See Figure 10). The malate overproducing *S. cerevisiae* strain rather than utilizing the oxidative TCA cycle required metabolic engineering of the reductive TCA cycle to drive carbon from pyruvate to oxaloacetate (catalyzed by pyruvate carboxylase, encoded by PYC1 and PYC2, and co-producing a net ADP and P<sub>i</sub>) and then to malate. A similar approach, requiring yet further engineering and understanding of the reductive TCA cycle to convert malate to succinate would be beneficial, but a major hurdle with this strategy is the conversion of fumarate to succinate by fumarate reductase (encoded by OSM1) in the mitochondrion and which is thermodynamically favored in the direction of fumarate. This approach is discussed in significantly more detail in **Paper II**.



**Figure 10: *S. cerevisiae* Central Carbon Metabolism.** The major central carbon metabolic pathways including glycolysis, the TCA cycle, the glyoxylate by-pass, abbreviated parts of the PPP, and compartmentalization including cytoplasmic space and mitochondrion, are presented. The two major pathways where succinate metabolism is observed, the TCA cycle and the glyoxylate by-pass, are highlighted with succinate denoted in blue and a subscript M is included to signify mitochondrial pools of the metabolite. The black ellipses on the mitochondrial membrane represent transporters. This figure was reconstructed based on the genome-scale metabolic network reconstruction of Förster et al, 2003.

## 6.3 Technical Challenges Anticipated

Although impossible to predict all the technical challenges that will be encountered in pursuit of a succinic acid microbial cell factory in *S. cerevisiae*, a review of the primary literature suggests some areas of focus.

### 6.3.1 Succinate Transport - Mitochondrial vs. Cytosolic

Currently, *A. succiniciproducens*, *A. succinogenes*, *E. coli*, and *M. succiniciproducens* serve as the primary expression systems for over-production of succinate. A critical difference between these systems and *S. cerevisiae* is cellular compartmentalization (See Figure 10). Succinate, primarily produced via either the reductive or oxidative TCA-cycle, is present in the mitochondria while cytosolic succinate production is limited. Cytosolic succinate production from the direct conversion of isocitrate to succinate, yielding glyoxylate as a by-product, is significant only under aerobic non-glucose dependent growth (Moreira dos Santos et al, 2003; Regenberg et al, 2005) and further discussed in **Papers I, III, and IV**. It may be possible to explore metabolic engineering approaches that over-express cytosolic succinate production; however, preservation of an aerobic glucose-metabolism phenotype is desired to ensure that  $\mu_{max}$  (maximum specific growth rate) remains close to that of the wild-type  $\mu_{max}$  in order to ensure high volumetric productivity.

There is limited literature available on TCA-cycle intermediate transport, particularly for glucose-limited aerobic *S. cerevisiae* cultivations. The *ACR1* gene product, Acr1p, for *S. cerevisiae* grown on ethanol or acetate (expression subject to glucose repression), is a membrane protein that transports succinate and fumarate. Specifically, Acr1p transports succinate from the cytosol into the mitochondrial matrix, in exchange for fumarate. Under anaerobic growth succinate derived from isocitrate lyase (cytosol) is transported into the mitochondrial matrix to replenish C<sub>4</sub> organic acids of the TCA-cycle when oxaloacetate is transported to the cytosol for gluconeogenesis (Bojunga et al, 1998). Although the desire is to produce an aerobic strain grown on glucose, Acr1p and similar transport proteins may offer insight into developing an appropriate succinate transport model.

Dicarboxylate transport protein (DTP) genes for mitochondrial transport of organic acids in yeast were first identified in 1996 (Kakhniashvili et al, 1997). Specifically, DTP that exchanges malonate for malate, succinate, and phosphate was identified, overexpressed in *E. coli*, and characterized. The DTP gene from the yeast *S. cerevisiae* is a 298-residue basic protein that displays the mitochondrial transporter signature motif, three homologous 100-amino acid sequence domains, and six predicted membrane-spanning regions. The purified, overexpressed DTP was reconstituted in phospholipid vesicles, where kinetic properties ( $K_M = 1.55 \text{ mM}$ ,  $V_{MAX} = 3.0 \text{ } \mu\text{mol min}^{-1} \text{ mg protein}^{-1}$ ) and substrate specificity were determined (Kakhniashvili et al, 1997). It may be possible to genetically modify DTP genes to up-regulate succinate transport from the mitochondria to cytosol, and perhaps more challenging, to utilize similar gene products to shuttle succinate externally from the cytosol.

More recent work has investigated the kinetics of acetate, pyruvate, and succinate oxidation rates at pH ≥ 5.5 (no membrane diffusion of protonated acid species) in *S. cerevisiae* Y-503, by considering a model consisting of a plasma membrane dicarboxylate transporter, a mitochondrial dicarboxylate transporter and succinate dehydrogenase (Sdh complex) as a “succinate oxidase” system (Aliverdieva et al, 2006a; Aliverdieva et al, 2006b). Using antimycin A to completely inhibit succinate oxidation by ubiquinol oxidase, this three step oxidase system was the only pathway to succinate oxidation; further, the Sdh complex was the common step of oxidation of all three substrates in the TCA cycle. Measuring the cell and mitochondrial oxidation rates (nmol O<sub>2</sub> min<sup>-1</sup>) in the absence of glucose but extracellular presence of either acetate, pyruvate or

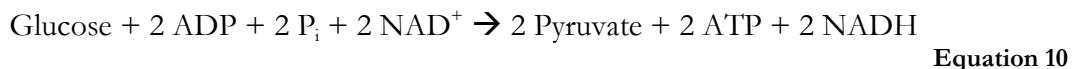
succinate, demonstrated that pyruvate and acetate oxidation rates were significantly higher than that of succinate. This suggests that succinate dehydrogenase is not the limiting step in succinate oxidation. Inhibition of the Sdh complex by 2-thienyltriflouroacetone was furthermore shown to affect the acetate oxidation rate much more severely than the succinate oxidation rate, another indication of the previous observation. More investigations of inhibition of the components of the system supported this view, and led to a final conclusion that the succinate oxidation rate is limited at the plasma membrane transporter step (Aliverdieva et al, 2006a; Aliverdieva et al, 2006b).

Any efforts to construct a *S. cerevisiae* microbial cell factory will have to consider mitochondrial to cytosolic transport of succinate, as well as cytosolic secretion of succinate extracellularly.

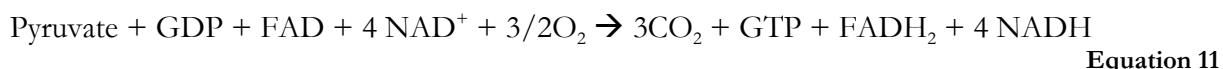
### 6.3.2 Redox and Co-Factor Balance

The redox potential is a function of pH, dissolved oxygen, and the redox states of compounds present in the medium. Correcting a redox imbalance was demonstrated to significantly increase succinate production in *E. coli* NZN111 (*FΔpfhl::Cam ldhA::Kan*), specifically through the use of more reduced carbon sources, and the supply of reducing power through the head-space gas (e.g., H<sub>2</sub>). In *E. coli* NZN111, eight moles of hydrogen reducing power are required for metabolism of two moles of pyruvate to two moles of succinate, while only four moles of hydrogen are produced for complete glycolysis of one mole of glucose (Hong et al, 2002). Similar considerations have to carefully be evaluated for *S. cerevisiae* succinate over-producing strains.

Furthermore, attention must be devoted to the balance of NADH, NAD(P)H, FADH<sub>2</sub>, and CoA co-factors. As previously described, the TCA cycle serves as a mechanism for the replenishment of NADH. In this context it is useful to recall that the overall reaction balance for glycolysis is:



Assuming aerobic conditions and relatively low glycolytic fluxes, the pyruvate carbon pool is directed towards oxidative TCA cycle metabolism. A pyruvate reaction balance, assuming complete oxidative TCA cycle metabolism is:



Note that if we assume that an over-producing succinate *S. cerevisiae* strain will have down-regulated succinate dehydrogenase activity (i.e., succinate to fumarate), then the overall pyruvate reaction scheme is:

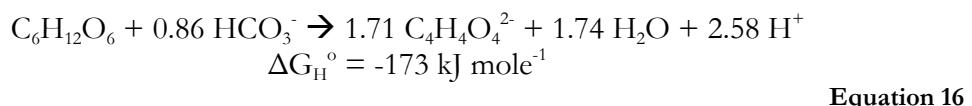


The final carbon species in Equation 12 is included to highlight that although Sdhp activity may be down-regulated, organic acid synthesis (e.g., fumarate, malate, oxaloacetate) must still be accounted for to close the carbon balance. In any case, the redox equivalents NADH and FADH<sub>2</sub> must be replenished. During aerobic conditions NADH and FADH<sub>2</sub> are oxidized during oxidative phosphorylation. The overall reaction scheme for oxidative phosphorylation is:

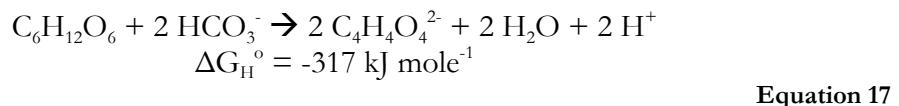


The number of moles of ATP produced from each oxygen atom consumed during oxidative phosphorylation is often referred to as the P/O ratio and this is different for oxidation of mitochondrial and cytosolic NADH. Furthermore, the P/O ratio between NADH and FADH<sub>2</sub> differs because NADH (enters at phosphorylation site I) enters the electron respiratory chain at an earlier stage than FADH<sub>2</sub> (enters at phosphorylation site II). Therefore, assuming a theoretical P/2e ratio for phosphorylation sites I, II, and III as 1.0, 0.5, and 1.0 respectively, then the theoretical P/O for NADH and FADH<sub>2</sub> is 2.5 and 1.5, respectively (van Gulik et al, 1995). In *S. cerevisiae*, phosphorylation site I is not active (Onishi et al, 1973), therefore the theoretical P/O for NADH and FADH<sub>2</sub> oxidation is 1.5 (van Gulik et al, 1995). Depending on the balance between mitochondrial and cytosolic redox equivalents, the theoretical P/O ratio is a maximum of 1.5 (Equation 13 to Equation 15). In *S. cerevisiae*, cytosolic NADH cannot pass the inner mitochondrial membrane where oxidative phosphorylation takes place, and reoxidation of cytosolic NADH is coupled to reduction of mitochondrial FAD to FADH<sub>2</sub> (Lei, 2001). Due to incomplete coupling between oxidation and phosphorylation (e.g., electron transport and energy generation), the observed P/O ratio is ~0.95-1.2, which is low compared to the theoretical value of 1.5 (van Gulik et al, 1995). However, this lower value is in agreement with the lack of phosphorylation site I activity. The P/O ratio to support both ATP generation for biomass production and maintenance is considered when determining the theoretical maximum yield of succinate on glucose,  $Y_{SSuc}$ .

**Paper II** provides further insights into using GSSM for simulation of succinate production in *S. cerevisiae*. Included in this paper is also a discussion of the theoretical maximum yield of succinate on glucose in *S. cerevisiae*. In bacterial systems, as previously discussed, the theoretical yield of succinate ( $\text{C}_4\text{H}_4\text{O}_4^{2-}$ ) on glucose ( $\text{C}_6\text{H}_{12}\text{O}_6$ ) is summarized below:



Implicit in Equation 16 is the fact that CO<sub>2</sub> carboxylation of pyruvate to form oxaloacetate via pyruvate carboxylase is thermodynamically favorable and readily observed in bacterial production hosts. In the presence of additional reducing power (e.g., H<sub>2</sub>g supplementation), the theoretical yield of succinate on glucose increases.



In *S. cerevisiae*, pyruvate carboxylase encoded by *PYC1* and *PYC2* (cytosolic pyruvate carboxylase isoforms) catalyzes the ATP-driven conversion of pyruvate and HCO<sub>3</sub><sup>-</sup> to oxaloacetate. *S. cerevisiae* expresses phosphoenolpyruvate (PEP) carboxykinase, encoded by *PCK1*, and facilitating gluconeogenesis by ATP-driven conversion of oxaloacetate to PEP, yielding CO<sub>2</sub>. In *S. cerevisiae* there is no PEP carboxylase (or the reversible PEP carboxykinase) as compared to bacterial systems where this reaction has been reported (Bazaes et al, 2007; Lu et al, 2009; Kim et al, 2007). Given that *S. cerevisiae* pyruvate carboxylase is minimally observed under batch glucose

fermentation conditions, the theoretical yield of succinate on glucose is governed by the bioreaction balance:



Equation 18

The potential for enhancing pyruvate carboxylation via pyruvate carboxylase, thereby driving reductive TCA cycle generation of succinic acid is discussed later.

**Paper II** expands on the simple bioreaction approach presented here and using GSMMs coupled with FBA considers the theoretical yield of succinate on glucose in *S. cerevisiae*. Assuming 1 mmol ATP g-DCW<sup>-1</sup> h<sup>-1</sup> maintenance cost and a 10 mmol glucose g-DCW<sup>-1</sup> h<sup>-1</sup> uptake rate, the maximum succinate yield was found to be 0.51 g g-glucose<sup>-1</sup>. This maximum yield is based on FBA when [H<sup>+</sup>] was balanced. The exact mechanism by which succinate is transported across the cytosolic membrane, as previously discussed, has yet to be clearly elucidated, with literature suggesting both dicarboxylic acid proton-coupling, and the absence of such coupling (Aliverdieva et al., 2006b). If [H<sup>+</sup>] is treated as an external metabolite (e.g., unconstrained), the maximum yield of succinate is 0.98 g g-glucose<sup>-1</sup>. Furthermore, if CO<sub>2</sub> uptake is permitted, enabling carboxylation reactions, the maximum theoretical yield is 1.124 g g-glucose<sup>-1</sup>. Given the lack of physiological characterization of succinate transport, and the relatively high impact of assumptions surrounding [H<sup>+</sup>] balancing, external [H<sup>+</sup>] was balanced throughout all simulations, and the maximum succinate yield was assumed to be 0.51 g g-glucose<sup>-1</sup> (0.52 C-mol C-mol-glucose<sup>-1</sup>). This represents a worst case scenario in terms of the theoretical potential for *S. cerevisiae* to stoichiometrically overproduce succinate. Clearly, bacterial production systems have a strong advantage in terms of a higher theoretical yield of succinate due to reductive TCA cycle activity (i.e., PEP carboxylase, malate dehydrogenase, and fumarate reductase). **Paper II** presents in depth a discussion and effort to exploit the native reductive TCA cycle capacity of *S. cerevisiae*.

### 6.3.3 Cell Growth Regulation - Crabtree Effect

In developing and characterizing a succinate over-producing *S. cerevisiae* strain, particularly using continuous and batch fermentations with diverse carbon sources (e.g., glucose and xylose), glucose response pathway physiology must be considered. *S. cerevisiae* similar to other eukaryotes undergoes a complex signaling cascade for detection of glucose, extracellular to intracellular signal transduction, and then amplification and modulation of that signal at the genomic, transcriptomic, proteomic, and metabolomic levels (Santangelo, 2006). A specific physiological response to glucose in *S. cerevisiae*, referred to as the Crabtree effect (Crabtree, 1929<sup>7</sup>), suggests that at low glycolytic fluxes assuming aerobic metabolism, all of the pyruvate carbon generated from glycolysis will be directed towards oxidative TCA cycle metabolism, and ultimately oxidative phosphorylation. Yet, when glycolytic flux increases and the pyruvate pool increases, carbon flux is directed towards overflow metabolites ethanol, acetate, and glycerol. In a review by Pronk, et al, the proposed mechanism for the Crabtree effect in yeast cells includes both long-term and short-term effects (Pronk et al, 1996). Specifically, the long-term effect is due to insufficient capacity to metabolize pyruvate under high extracellular glucose concentrations leading to the ethanol production as a result of insufficient respiratory capacity and repression of respiratory genes (Petrik et al, 1983; Rieger et al, 1983; Postma et al, 1989). The short-term effect is characterized by respiratory metabolism becoming saturated, again causing overflow metabolism

<sup>7</sup> Although the Crabtree effect is named after Herbert Grace Crabtree, the physiological effect was first discovered in tumor cells by Warburg in 1926 (Warburg, 1926). The later study of Crabtree also used tumor cells (Crabtree, 1929).

at the pyruvate metabolic node (Rieger et al, 1983; Van Urk et al, 1990). Related to the Crabtree effect is a phenomenon referred to as glucose repression (Ronne, 1995; Gancedo, 1998; Carlson, 1999). Specifically, extracellular glucose addition causes a greater than 3-fold change (either induction or repression) to approximately 20% of all genes. Moreover, 40% of all genes show at least a 2-fold change (Wang et al, 2004). Transcriptionally repressed genes are involved in gluconeogenesis, the glyoxylate cycle (succinate producing), peroxisomal  $\beta$ -oxidation, alternative carbon source metabolism, as well as many of the TCA cycle enzymes (succinate producing and consuming) (Ronne, 1995; Gancedo, 1998; Carlson, 1999; Lodi et al, 2001).

The complex signaling and regulatory network that *S. cerevisiae* employs in the central carbon metabolism response to extracellular glucose will be critical for any metabolic engineering strategies applied to *S. cerevisiae* for succinic acid production. To maximize specific and volumetric yield it is desirable to produce a strain that maintains a Crabtree negative phenotype – where the carbon flow will be directed towards maximizing succinate and biomass, with minimal losses to overflow metabolism and CO<sub>2</sub>. Metabolic engineering strategies will need to consider the impact of potentially high pyruvate pools due to partially inactivated TCA-cycle metabolism (i.e., succinate over-production may cause decreased flux through the TCA-cycle, causing pyruvate over-flow). Any metabolic engineering strategies proposed, particularly those considering modification of TCA-cycle reactions, should consider the impact of glucose repression.

One specific challenge, in the context of using GSMMs for metabolic engineering strategy development and prediction, is incorporating regulatory mechanisms. As previously discussed, GSMMs are exclusively stoichiometric and cannot distinguish between, for example, batch or continuous cultivations. Batch cultivations, supplemented with relatively high concentrations of glucose will exhibit Crabtree phenotypes (i.e., overflow metabolism) and glucose repression. Continuous cultivations, glucose-limited, will exhibit Crabtree negative phenotypes (i.e., respiratory metabolism induced) and glucose derepression. GSMMs for *S. cerevisiae* have conventionally been developed and tuned for continuous cultivation conditions (Förster et al, 2003a). However, industrially relevant processes are unlikely to use this mode of fermentation, and dominated by batch or fed-batch processing, thereby it will be critical to develop approaches and tools that can identify relevant metabolic engineering strategies and distinguish between batch (i.e., Crabtree) and continuous (i.e., Crabtree negative) cultivation conditions. The Crabtree effect and glucose repression are intimately intertwined with the results and discussion presented in **Paper I, II, III, and IV**. Specifically, **Paper II** directly addresses the challenges with using *in silico* simulations and GSMMs for batch *S. cerevisiae* conditions.

## 7.0 Summary of Results

A brief synopsis of each paper is provided below, including a summary of the major results produced from my PhD research. Both **Paper I** and **Paper II** directly address the metabolic engineering of *S. cerevisiae* for overproduction of succinic acid. Each of these studies present diverse approaches to increase carbon flux towards succinic acid, while sharing similar approaches that included GSMM predictions that were then investigated *in vivo* with strain construction and characterization, at both a physiological and *x-ome* level. **Paper II** also attempts to better understand what opportunities exist within the native *S. cerevisiae* metabolic network for reductive TCA cycle succinate production. **Paper III** was an effort to address the critical issue of feedstock flexibility, particularly the desire to engineer a microbial platform that can utilize xylose efficiently under aerobic conditions with no diversion of carbon to overflow metabolites. **Paper III** builds upon the concepts introduced in **Paper I** where succinate production was directly coupled to biomass formation, hence, the desire to engineer a strain that rapidly converted all xylose consumed to biomass. Finally, **Paper IV** was an effort to apply relatively novel technology, high-throughput genome sequencing for SNP detection, to the field of metabolic engineering by asking the question: can a direct link from genotype to phenotype be detected? **Paper IV** lays the foundation for better understanding what specific genetic modifications resulted from the successful directed evolution demonstrated in **Paper I** and **Paper III**. Together, all the papers presented intend to provide holistic approach to industrial systems biology and metabolic engineering of *S. cerevisiae* as a microbial cell factory for a desirable chemical compound.

### 7.1 Paper I: Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory.

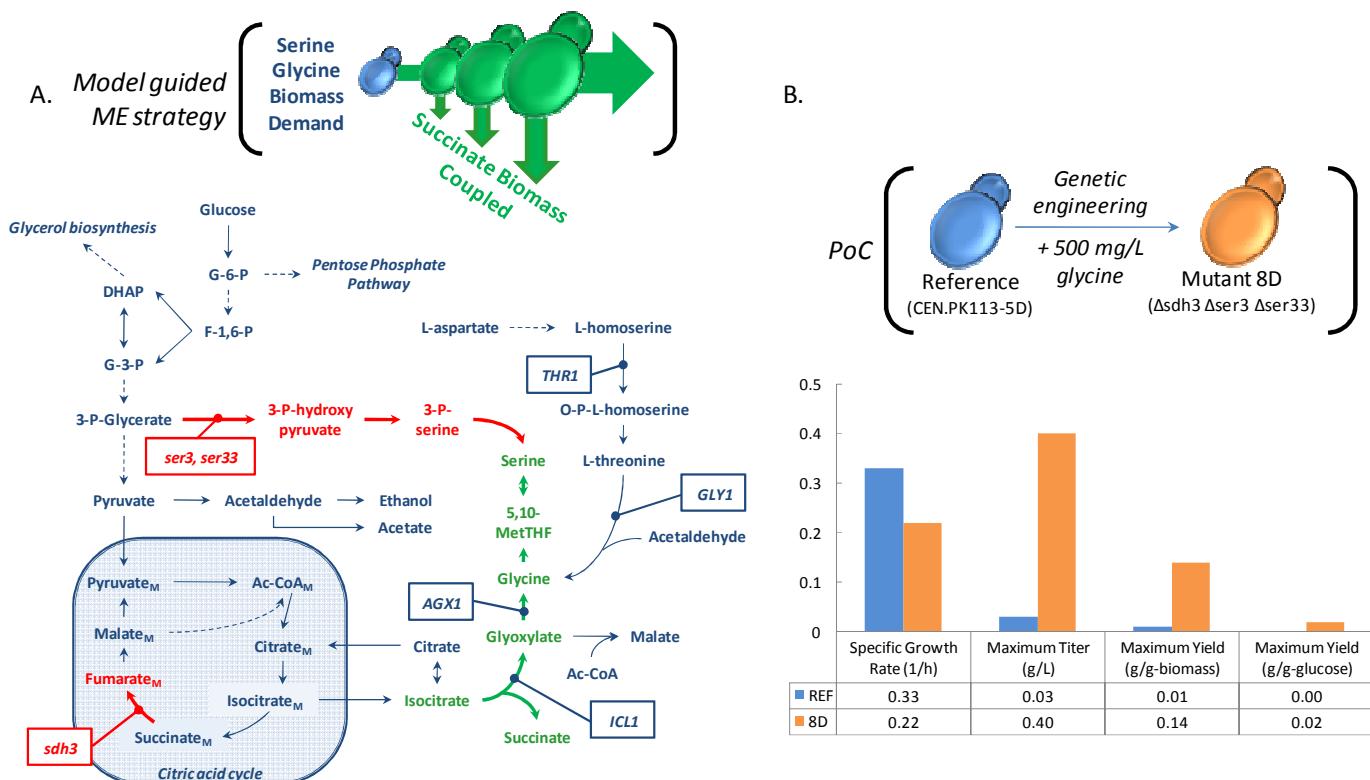
The genome-scale metabolic network reconstruction of *S. cerevisiae* permitted *in silico* prediction of gene deletions using an evolutionary programming method to couple biomass and succinate production (Patil et al, 2005). Glycine, serine, and threonine, all representing essential amino acids required for biomass formation, may be formed from both glycolytic and tricarboxylic acid cycle intermediates. Succinate formation results from the isocitrate lyase catalyzed conversion of isocitrate to equimolar glyoxylate and succinate, and from the  $\alpha$ -ketoglutarate dehydrogenase complex catalyzed conversion of  $\alpha$ -ketoglutarate to equimolar succinate. Succinate is subsequently depleted by the succinate dehydrogenase complex to equimolar fumarate (See Figure 10). The metabolic engineering strategy identified included deletion of the primary succinate consuming reaction encoded by *SDH3* (cytochrome b subunit of the succinate dehydrogenase complex, essential for function), and interruption of glycolysis derived serine by deletion of 3-phosphoglycerate dehydrogenase, Ser3p/Ser33p (isoenzymes). The remaining pathway for serine synthesis must originate from glycine, and glycine synthesis is largely derived from the alanine:pyruvate aminotransferase converting glyoxylate and alanine to glycine and pyruvate. With this strategy, glycine and serine biomass requirements are directly coupled to succinate formation via the glyoxylate cycle (See Figure 11).

The mutant resulting from the *in silico* strategy, referred to as 8D<sup>8</sup> ( $\Delta sdh3 \Delta ser3 \Delta ser33$ ), required supplementation with 500 mg l<sup>-1</sup> glycine to be able to grow. When evaluated in well controlled, aerobic, batch cultivations supplemented with 20 g l<sup>-1</sup> glucose in chemically defined

---

<sup>8</sup> The mutant 8D was constructed by Dr. Donatella Cimini. The mutant construction was guided by the metabolic modeling results described by Patil et al, 2005.

medium, it exhibited a 13-fold improvement in succinate secreted titer ( $0.03 \text{ v } 0.40 \text{ g l}^{-1}$ ), 14-fold improvement in succinate biomass yield ( $0.01 \text{ v } 0.14 \text{ g-succinate g-biomass}^{-1}$ ), and a modest 33% reduction in the specific growth rate. Thus, the *in silico* guided metabolic engineering strategy was shown to work, representing a proof-of-concept of the use of model guided metabolic engineering. However, in order to obtain a prototrophic strain directed evolution was employed to screen and select for 8D mutants that did not require glycine supplementation. Specifically, repeated shake flask cultivation and transfer in declining glycine concentration supplemented medium, from an initial  $500 \text{ mg l}^{-1}$  to  $0 \text{ mg l}^{-1}$  was performed. The resulting strain demonstrated a 7.7-fold improvement in succinate yield on biomass ( $0.09 \text{ v } 0.69 \text{ g-succinate g-biomass}^{-1}$ ), strongly suggesting the direct coupling of glycine formation from glyoxylate and succinate formation. The resulting strain had a relatively low specific growth rate,  $0.03 \text{ h}^{-1}$ , and was therefore subsequently cultivated in repeated shake flasks to improve the specific growth rate. Finally, a specific growth rate of  $0.14 \text{ h}^{-1}$  was reached, however, resulting in a decreased succinate yield on biomass ( $0.69 \text{ v } 0.27 \text{ g-succinate g-biomass}^{-1}$ ). The final strain, referred to as 8D Evolved, was shown to exhibit a 60-fold improvement in biomass coupled succinate production ( $0.01 \text{ v } 0.30 \text{ g-succinate g-biomass}^{-1}$ ), and 20-fold improvement in succinate titer ( $0.03 \text{ v } 0.60 \text{ g l}^{-1}$ ) relative to the reference strain when grown in aerobic batch cultivations.



**Figure 11: Proof-of-Concept Metabolic Engineering for Succinate Overproduction.** Panel A shows the central carbon metabolism of *S. cerevisiae*, and the model-guided metabolic engineering strategy for succinate overproduction. Legend: native reactions (blue solid line), lumped native reactions (blue dashed line), interrupted reactions (red solid line), up-regulated reactions (green solid line). Panel B demonstrates the proof of concept.

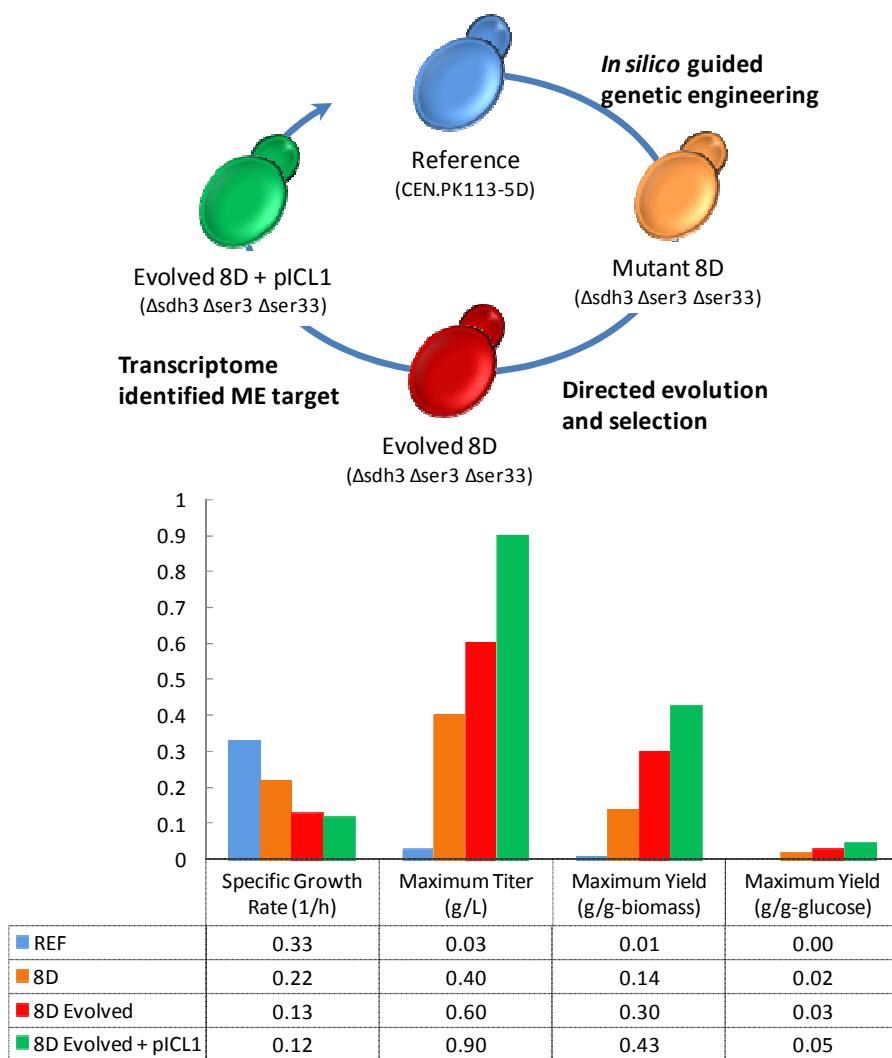
To investigate the apparent decoupling of succinate coupled biomass formation, and potentially identify second-round metabolic engineering strategies, the transcriptome was measured in aerobic, glucose-limited, mid-exponential phase grown batch cultivations of 8D Evolved and the reference strain. Of the total 2406 differentially expressed genes between the 8D Evolved and reference strain ( $p\text{-value}_{B-H} < 0.01$ ,  $|\log\text{-fold change}| > 0.5$ ,  $n=3$  biological replicates,

$n=2$  DNA microarray duplicates), 36 unique growth-related genes were identified suggesting that few of the genes with a significant change in transcription in 8D Evolved are due to changes in the specific growth rate. However, a total of 8 of the top 20  $p\text{-value}_{B,H}$  ranked differentially expressed genes identified from pair-wise comparison of 8D Evolved and the reference strain, are growth-related genes (*ARO9*, *SER3*, *JLP1*, *HMALPHA1*, *ARO10*, *MFALPHA2*, and two uncharacterized genes, *YPL033c* and *YLR267w*).

The top 2000 (there were no metabolic genes in the remaining 406 genes nor were there any biological process annotations available as determined by gene ontology, and therefore they were not included in further analysis) differentially expressed genes were selected for further analysis, and after removal of the 36 growth-related genes, a list of 1964 genes was submitted for metabolic pathway visualization and characterization. A total of 315 genes mapped to a specific metabolic pathway on the expression viewer, with a mean log-fold expression ratio value of  $0.3 \pm 1.3$  ( $n=315, \pm \text{SD}$ ).

Three biological insights were immediately apparent. First, *SDH3*, *SER3*, and *SER33* had negative log-fold expression ratios (log-fold change  $<-8.0$ ) confirming the gene deletions targeted in the 8D strain and the maintained low expression through the directed evolution. Second, when examining the glycine, serine, and threonine metabolism, *AGX1* was 4.3 log-fold change upregulated in the 8D Evolved strain, confirming significant upregulation of glycine synthesis from glyoxylate pools, as predicted by the original metabolic engineering strategy. However, there was no upregulation of *SHM2*, *SHM1*, the genes encoding pathways for L-serine formation from L-glycine pools. Most surprisingly *GLY1*, encoding threonine adolase, was significantly up-regulated (log-fold change 1.6). In the genome-scale metabolic network reconstructions of *S. cerevisiae* iFF708 and iND750, upon which the 8D metabolic engineering strategy is based, Gly1p encodes the reversible conversion of threonine to glycine and acetaldehyde (Förster et al, 2003a; Duarte et al, 2004b), leading to the prediction that threonine biosynthesis from glycolytic intermediates could be down-regulated, and provided for from glycine pools. This consequently leads to a greater biomass-coupled drive for glyoxylate synthesis from isocitrate, yielding equimolar succinate. Leveraging this over-all strategy, another *S. cerevisiae* mutant was constructed, referred to as 20G ( $\Delta sdb3$ ,  $\Delta ser3$ ,  $\Delta thr1$ ), where Thr1p, encoding homoserine kinase that is required for threonine biosynthesis, was deleted. However, this strain required threonine supplementation and after several extensive attempts at adaptive evolution, the threonine auxotrophy persisted, suggesting the reversibility of *GLY1* was limited with the adolase strongly favoring glycine formation. The significant up-regulation of *GLY1* therefore provides a strong hypothesis for why 8D Evolved had an attenuation of succinate production, even under increasing specific growth rate, suggesting a decoupling of biomass coupled succinate production.

The transcriptome not only provides for a global, rapid, and quantitative assessment of the predicted *in silico* metabolic engineering strategy and insight into the genetic engineering modifications that result from directed evolution and selection, but it also provides a source for identification of second round metabolic engineering targets not previously predicted. All tricarboxylic acid cycle genes are up-regulated, with the exception of *SDH3* (target gene deletion), and *ICL1*, providing a clear metabolic engineering target for up-regulation in the 8D Evolved strain. Therefore, native *ICL1* was PCR amplified and cloned into the 2  $\mu\text{m}$  ori plasmid containing the strong constitutive *TEF1* promoter and *CYC1* terminator, and then transformed into the reference, 8D, and 8D Evolved strain (strains transformed with the constructed plasmid pRS426T-*ICL1-C* are referred to as “with p*ICL1*”). All strains were evaluated in aerobic, glucose-supplemented batch fermentations, and only 8D Evolved with p*ICL1* exhibited a change in succinate production. Specifically, the succinate titer, succinate yield on biomass, and succinate yield on glucose were 0.90 g l<sup>-1</sup>, 0.43 g-succinate g-biomass<sup>-1</sup>, and 0.05 g-succinate g-glucose<sup>-1</sup>, respectively, representing a 1.5-fold, 1.4-fold, 1.7-fold improvement over 8D, respectively.



**Figure 12: Summary of Succinate Microbial Cell Factory Performance.** The specific growth rate ( $\text{h}^{-1}$ ), maximum succinate titer ( $\text{g l}^{-1}$ ), maximum succinate yield on biomass ( $\text{g g-biomass}^{-1}$ ), and maximum yield on glucose ( $\text{g g-glucose}^{-1}$ ) are reported for the reference strain, 8D, 8D evolved, and 8D evolved with pICL1.

The resulting strain, 8D Evolved with pICL1, represents a 30-fold improvement in succinate titer, and a 43-fold improvement in succinate yield on biomass, with only a 2.8-fold decrease in the specific growth rate compared to the reference strain (See Figure 12). Despite success of using simple stoichiometric-based calculations for driving metabolic engineering, it is interesting to note that regulatory mechanisms not captured in these models are likely playing a significant role in the succinate production observed. The biomass requirements for glycine and serine are 0.290 and 0.185 mmol g-DCW $^{-1}$ , respectively (Förster et al, 2003a). Assuming that all glycine, and all glycine and serine combined demands are supplied from the glyoxylate pool, then the theoretical production of succinate would amount to 0.034 and 0.056 g-succinate g-DCW $^{-1}$ , respectively. The 8D and 8D Evolved strains are producing 0.30 and 0.43 g-succinate g-biomass $^{-1}$ , respectively, suggesting a nearly 8-fold higher succinate production than required to meet biomass amino acid demands. A potentially 3<sup>rd</sup> metabolic engineering target would be deletion of *GLY1* to further minimize alternative biosynthetic routes of glycine production, thereby isolating all glycine production to be dependent on glyoxylate formation, and consequently succinate formation. Yet, it's clear that any increase in succinate formation would not be due to biomass

requirements, but rather regulatory (e.g., non-stoichiometric driven) mechanisms. Therefore, while the strategy presented and demonstrated here is likely to be a major component of an overall succinate production cell factory, complimentary strategies focusing on the other major succinate production pathway, TCA cycle, will be required.

Furthermore, this work clearly demonstrated that obvious genetic targets did not result in increased succinate formation. Specifically, deletion of the primary succinate consuming pathway ( $\Delta sdb3$ ) and constitutive over-expression of one of two of the primary succinate formation pathways ( $ICL1$ ) did not result in any increased succinate production (Cimini et al, 2009). It is further interesting to note that the 8D with pICL1 strain also did not result in any increased succinate production, but rather only in the 8D Evolved with pICL1 strain. The ability to measure transcriptome on a strain that underwent targeted genetic engineering and directed evolution was critical to identifying pICL1 as a 2<sup>nd</sup> metabolic engineering target, which would have been discarded if selected based on intuition.

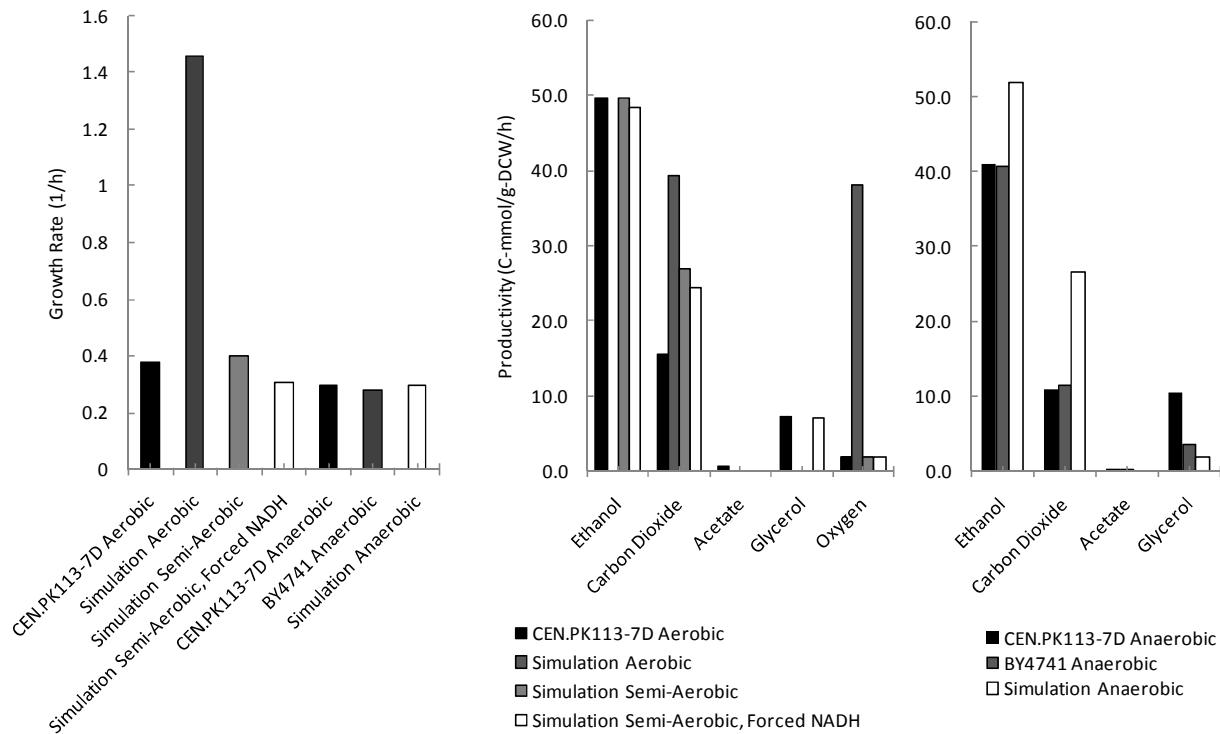
## 7.2 Paper II: Genome-scale modeling enables metabolic engineering of *Saccharomyces cerevisiae* for succinic acid production.

As previously described, from an *in silico* approach there has been one publication focusing on application of flux balance analysis (FBA) with the genome-scale metabolic network reconstruction of *S. cerevisiae* using an evolutionary programming method to couple biomass and succinate production (Patil et al, 2005). This approach highlighted several multi-gene deletion strategies for succinic acid overproduction, however, included no experimental validation of target predictions. Furthermore, this approach used a reduced genome-scale metabolic network reconstruction of iFF708, removing all duplicate and dead-end reactions. Attempts to reproduce those results using the complete iFF708 resulted in significantly reduced succinate yields on substrate than earlier found, and could only be obtained if a constraint preventing acetaldehyde secretion was imposed.

In **Paper II** we exploit FBA coupled with pathway visualization to explore succinic acid overproduction strategies as predicted by interrogation of the complete genome-scale metabolic reconstruction, iFF708. More specifically, we explore all single gene and double gene deletions under aerobic and anaerobic conditions, maximizing the objective function of growth rate with constrained glucose uptake rate, and observe the maximum succinate yield on substrate. The top three single gene deletion predictions, occurring under anaerobic glucose fermentation conditions, were experimentally evaluated in order to gain new insight into the predictive strength of *in silico* predictions. Furthermore, these three strains were physiologically and transcriptionally characterized with the objective to gain further knowledge on the C<sub>4</sub> acid production by *S. cerevisiae*.

Following the successful conversion of iFF708 to SBML format, and development of FBA and visualization tools, fermentation data of *S. cerevisiae* CEN.PK113-7D was used to evaluate the model's predictive power. Batch aerobic and anaerobic glucose fermentations performed in well-controlled 2L fermentations were compared to corresponding simulation conditions where the objective function, growth, was maximized while constraining glucose uptake rate, and for anaerobic conditions, constraining the oxygen uptake rate ( $r_{O_2}$ ) to 0 mmol-O<sub>2</sub> g-DCW<sup>-1</sup> h<sup>-1</sup>. It is found that there is a poor agreement with corresponding batch glucose aerobic experimental data due to the inability of the model to describe the Crabtree effect as discussed earlier (Åkesson et al, 2004). When  $r_{O_2}$  was constrained to experimentally determined fermentation values of 1.8 mmol-O<sub>2</sub> g-DCW<sup>-1</sup> h<sup>-1</sup>, referred to as semi-aerobic, the simulation accurately predicted the specific growth rate (0.38 vs. 0.40 h<sup>-1</sup>, experimental vs. simulation, respectively), ethanol yield (0.54 vs. 0.54 C-mol C-mol-glucose<sup>-1</sup>), and biomass yield (0.17 vs. 0.18 C-mol C-mol-glucose<sup>-1</sup>).

However, carbon dioxide (0.16 vs. 0.30 C-mol C-mol-glucose<sup>-1</sup>) and glycerol (0.08 vs 0.0 C-mol C-mol-glucose<sup>-1</sup>) yields were in poor agreement. Biomass formation as a result of glucose respiro-fermentative metabolism, with a high dependence on oxygen availability and glucose concentration, results in the formation of excess NADH (Nissen et al., 1997). Excess NADH, both cytosolic and mitochondrial, is a direct result of biomass required ATP generation, and compartmental redox balance is possible through cytosolic NADH dehydrogenases, the glycerol-3-phosphate shuttle, and mitochondrial redox shuttles (von Jagow et al, 1970; Luttik et al, 1998; Overkamp et al, 2000; Geertman et al, 2006). Glycerol formation results from redox balancing and NADH regeneration to NAD<sup>+</sup> in the cytosol and glycerol production can be reduced through expression of a cytosolic NADH oxidase (Vermuri et al, 2007). Improving the fit of the model to glycerol production can be accommodated by several means, but here we took a simple pragmatic approach by introducing an artificial conversion of NAD<sup>+</sup> → NADH, and then constraining this reaction to a flux such that the glycerol production is correctly described by the model. We chose this approach rather than simply constraining the glycerol flux as this was found to give better overall fit of the fluxes (See Figure 13).



**Figure 13: Experimental and Simulation Comparative Data.** Comparison of the specific growth rate and specific productivities for experimental data generated using the reference *S. cerevisiae* CEN.PK113-7D and BY4741 under aerobic and anaerobic glucose batch fermentations, and simulation data. For the condition, *Simulation Aerobic*, *Simulation Semi-Aerobic*, *Simulation Anaerobic*, the  $r_{O_2}$  was unconstrained (0-1000 mmol-O<sub>2</sub> g-DCW<sup>-1</sup> h<sup>-1</sup>), constrained to 1.8 mmol-O<sub>2</sub> g-DCW<sup>-1</sup> h<sup>-1</sup>, and constrained to 0 mmol-O<sub>2</sub> g-DCW<sup>-1</sup> h<sup>-1</sup>, respectively. The condition, *Simulation Semi-Aerobic, Forced NADH*, included the reaction FNADH constrained to 6 mmol-NADH g-DCW<sup>-1</sup> h<sup>-1</sup>. For aerobic experimental data the specific glucose uptake rate was 91.2 C-mmol g-DCW<sup>-1</sup> h<sup>-1</sup> for CEN.PK113-7D. For anaerobic experimental data the specific glucose uptake rate was 93.1 C-mmol g-DCW<sup>-1</sup> h<sup>-1</sup> for CEN.PK113-7D and 89.7 C-mmol g-DCW<sup>-1</sup> h<sup>-1</sup> for BY4741. For all simulation conditions the glucose uptake rate was constrained to 91.2 C-mmol g-DCW<sup>-1</sup> h<sup>-1</sup>.

As a consequence of introduction of this reaction (constraining FNADH to 6 mmol NADH g-DCW<sup>-1</sup> h<sup>-1</sup>), glycerol yield was 0.079 vs 0.078 C-mmol C-mmol-glucose<sup>-1</sup> (experimental vs. simulation, respectively). While simulated carbon dioxide yield were still higher than observed

experimentally, the semi-aerobic, forced NADH simulation condition exhibits strong alignment to experimentally determined specific growth rate and productivities.

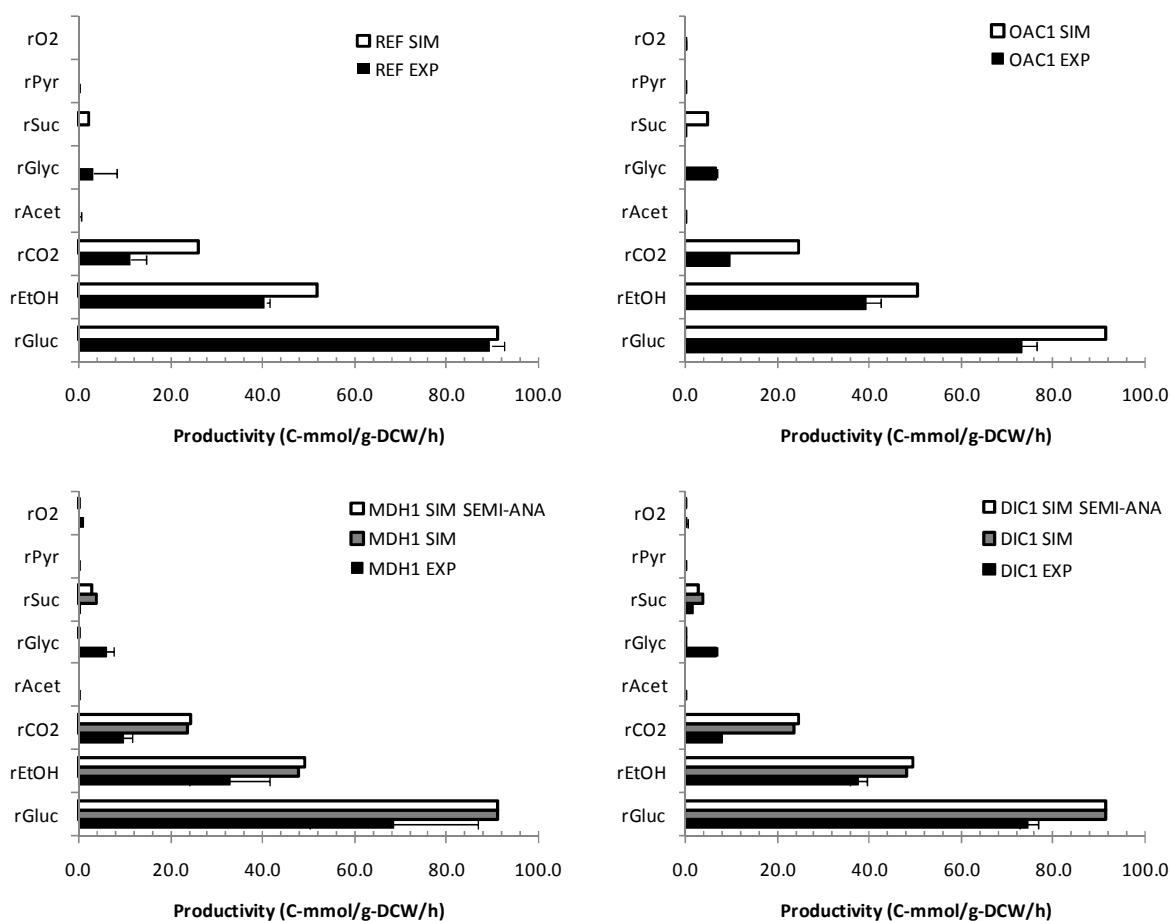
Under aerobic conditions there are no single gene deletions that result in increased succinate production. Interestingly, the reference case simulation under aerobic conditions with no gene deletions produces a small amount of succinate ( $0.003 \text{ C-mol C-mol-glucose}^{-1}$ ), which is not observed experimentally. If succinate excretion is constrained to zero, optimization of growth rate will result in growth while producing glycerol, under minimal amounts of oxygen, and then acetate under increasing amounts of oxygen. However, experimentally, both glycerol and acetate production are observed while succinate production is absent. Under aerobic conditions there is a strong sensitivity of succinate yield on substrate to  $r_{O_2}$  and for  $r_{O_2} > 2 \text{ mmol-O}_2 \text{ g-DCW}^{-1} \text{ h}^{-1}$  the succinate yield on substrate is zero.

At aerobic conditions double gene deletions only resulted in minor improvement of succinate production. Nearly all of the predictions required the deletion of the succinate dehydrogenase complex. Given the high degree of sensitivity of succinate production to  $r_{O_2}$ , anaerobic simulations offer the advantage of constraining this flux to zero, and these conditions can be tested reasonably well experimentally. Under anaerobic simulation conditions, a small amount of succinate is produced,  $0.003 \text{ C-mol C-mol-glucose}^{-1}$ , and if succinate production is constrained to zero, then the model predicts no growth. This is likely because the production of orotate from dihydroorotate, catalyzed by dihydroorotate dehydrogenase (encoded by *URA1*) required for pyrimidine synthesis, is coupled to the reduction of ubiquinone to ubiquinol. Under aerobic conditions oxygen serves as the final electron acceptor and enables ubiquinone regeneration, while under anaerobic conditions flavin adenine dinucleotide (FAD) serves as the electron acceptor for ubiquinone regeneration, and FAD must be regenerated from the transfer of electrons to fumarate, producing succinate. Given this proposed mechanism, the solution space for succinate production under anaerobic conditions rapidly approaches singularity with a high dependence on  $r_{O_2}$ . Given that experimentally it would be difficult to ensure 0 mmol  $O_2$ , potential gene deletions were therefore screened for micro aerobic conditions, where  $r_{O_2}$  was constrained to  $0.016 \text{ mmol-O}_2 \text{ g-DCW}^{-1} \text{ h}^{-1}$ , and determined to be the minimum  $r_{O_2}$  required for sustaining cell growth at the same rate if succinate production is constrained to zero or unconstrained.

A significant increase in the succinate yield, by a factor of approximately 10-fold from the reference case, can be obtained for the single gene deletions  $\Delta oac1$ ,  $\Delta mdh1$ , and  $\Delta dic1$  ( $0.033 \text{ C-mol C-mol-glucose}^{-1}$  vs.  $0.003 \text{ C-mol C-mol-glucose}^{-1}$ , single gene deletion vs. reference case simulation, respectively). Furthermore the significant increase in succinate yield on substrate resulted in nearly no impacts to growth rate ( $0.28 \text{ h}^{-1}$  vs.  $0.30 \text{ h}^{-1}$ , single gene deletion vs. reference case simulation, respectively). Physiologically, it was confirmed that  $\Delta oac1$ ,  $\Delta mdh1$ , and  $\Delta dic1$  are viable null mutants, and their annotation is well known, encoding for an inner mitochondrial membrane transporter (Oac1p), malate dehydrogenase (Mdh1p), and an inner dicarboxylate mitochondrial transporter (Dic1p), respectively (Cherry et al, 1998). Interestingly, further simulations of the best double gene deletions resulted in the same order of magnitude succinate yields on substrate compared to the aforementioned single gene deletions.

In order to explore and validate if the single gene deletions identified in silico result in more succinate production, the corresponding strains of the *S. cerevisiae* BY4741 background were cultivated anaerobically in 2L well controlled fermenters. There is a fair agreement between model predictions and experimental data. Focusing more closely on the specific succinate productivity, the reference case,  $\Delta mdh1$ , and  $\Delta oac1$  experimentally determined yields are significantly lower than expected based on model simulations. The  $\Delta dic1$  case, however, demonstrated a significantly higher yield of succinate compared to the reference case ( $0.02$  vs.  $0.00 \text{ C-mol C-mol-glucose}^{-1}$ ,  $\Delta dic1$  vs. reference, respectively), and was in-line with the in silico prediction ( $0.02$  vs.  $0.03 \text{ C-mol C-mol-glucose}^{-1}$ ,  $\Delta dic1$  experimental vs.  $\Delta dic1$  anaerobic

simulation, respectively). This represents a >10-fold improvement in succinate productivity based exclusively on a novel *in silico* prediction (See Figure 14).



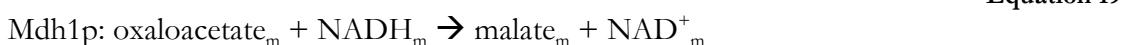
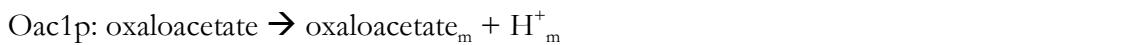
**Figure 14: Experimental and Simulation Comparative Data for Reference,  $\Delta oac1$ ,  $\Delta mdh1$ , and  $\Delta dic1$  Strains.** Summary of the specific growth rate (SGR) and specific consumption/productivity values for major carbon products (glucose, ethanol, carbon dioxide, acetate, glycerol, succinate, pyruvate, and oxygen) for both experimentally determined data of anaerobic batch glucose fermentations, and corresponding anaerobic simulation data of the BY4741 reference strain, and single gene deletion strains  $\Delta mdh1$ ,  $\Delta dic1$ , and  $\Delta oac1$ .

To gain further insight into the physiological performance of each strain identified via simulation results, genome-wide DNA microarray profiling was completed under anaerobic batch glucose fermentations. The number of differentially expressed genes for the  $\Delta oac1$  strain compared to the reference strain was very low, and consequently suggests that deletion of  $\Delta oac1$  causes virtually no transcriptional, and consequently, physiological differences compared to the reference BY4741 strain. The  $\Delta dic1$  and  $\Delta mdh1$  strains, compared to the reference strain, had 117 and 209 differentially expressed genes, respectively. Of these genes a total of 33% and 23% were up-regulated genes and 66% and 76% were down-regulated genes, for the  $\Delta dic1$  and  $\Delta mdh1$  strains, respectively. The average fold change of differentially expressed genes for the  $\Delta dic1$  strain, both up- and down-regulated, was  $\approx 2.5$ -fold greater than  $\Delta mdh1$ .

The differentially expressed gene sets for  $\Delta dic1$  and  $\Delta mdh1$  were submitted for gene ontology (GO) process annotation. Given the high degree of similarity in the GO process annotation for both the  $\Delta dic1$  and  $\Delta mdh1$  conditions, the complete list of differentially expressed genes were submitted for metabolic pathway annotation. What is immediately apparent is the

relatively small number of total metabolic pathway genes identified in  $\Delta dic1$  and  $\Delta mdb1$  compared to the reference, with a total of 10 and 20 genes identified as catalyzing metabolic reactions, respectively. Perhaps more striking is that there is an overlap of 9 metabolic pathway genes between both  $\Delta dic1$  and  $\Delta mdb1$ . The only differentially expressed gene present in the  $\Delta dic1$  condition, not present in the  $\Delta mdb1$  condition, is  $\Delta dic1$ .

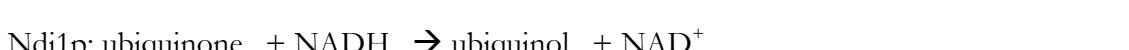
The metabolic engineering strategies identified through deletion of  $\Delta dic1$ ,  $\Delta mdb1$ , and  $\Delta oac1$ , suggest a common mechanism that was identified via visualization of the central carbon metabolism. As described earlier, mitochondrial redox balance must be maintained and while respiratory metabolic activity under anaerobic conditions is reduced compared to aerobic conditions, some activity is required to support glutamate/glutamine metabolism from  $\alpha$ -ketoglutarate (Camarasa et al, 2003; Camarsa et al, 2007), which produces NADH. During anaerobic metabolism, NAD<sup>+</sup> regeneration occurs via the following pathways, where the subscript  $m$  denotes mitochondrial:



In the cytosol malate is then converted to oxaloacetate, and the resulting NADH is converted to NAD<sup>+</sup> with the production of glycerol. The  $\Delta dic1$  strategy, relying on deletion of the mitochondrial dicarboxylate carrier Dic1p, catalyzes the following transport reaction, noting the intermediate transport of orthophosphate:



Assuming *DIC1* deletion, then the likely pathway is:



The  $\Delta dic1$  strategy relies heavily on the compartmental localization and function of Frds1p, soluble mitochondrial fumarate reductase, which continues to be poorly understood. However, recent work has suggested that a double deletion *S. cerevisiae* mutant,  $\Delta osm1 \Delta frds1$ , failed to grow

under batch glucose anaerobic conditions. Furthermore, during anaerobic growth, *FRDS1* expression in the wild-type was two to eight times higher than that of *OSM1*, suggesting that formation of succinate is strictly required for the reoxidation of FADH<sub>2</sub> and its expression may be oxygen-regulated (Camarasa et al, 2007). While neither *FRDS1* nor *OSM1* were significantly differentially expressed in the  $\Delta$ *mdh1* or  $\Delta$ *dic1* mutants compared to the reference strain, *FRDS1* was slightly up-regulated in the  $\Delta$ *dic1* mutant compared to the  $\Delta$ *mdh1* mutant ( $\log_{10}$  fold change 0.11 vs. -0.10, respectively). Lastly, as shown there was strong up-regulation of *CYC1* in both the  $\Delta$ *dic1* and  $\Delta$ *mdh1* mutants, suggesting that electron transport from ubiquinone cytochrome C oxidoreductase to cytochrome C oxidase was up-regulated, and required to facilitate electron transfer from NADH<sub>m</sub> to NAD<sup>+</sup><sub>m</sub>, and then from FADH<sub>2m</sub> to FADH<sub>m</sub> resulting in succinate formation. It has been well established that *CYC1* is both glucose repressed and regulated by the presence of oxygen and heme (Hörtner et al., 1982; Guarente et al, 1983; Boss et al, 1980; Guarente et al, 1984). Therefore, strong up-regulation during anaerobic batch glucose fermentations in combination with deletion of *DIC1* may have aided in the increased succinate formation observed. However, this does not explain the lack of succinate production observed in the  $\Delta$ *mdh1* mutant. It has been suggested that mitochondrial FADH<sub>2</sub> could be oxidized in the cytosol, which may provide an explanation for the failure of the  $\Delta$ *mdh1* and  $\Delta$ *oac1* mutants to produce any succinate (Enomoto et al, 2002). In any event, the strategies proposed here rely on the capacity for reductive TCA cycle activity under anaerobic conditions, and more specifically, the catalysis of fumarate to succinate via fumarate reductase. There is data suggesting that *S. cerevisiae* can exhibit this metabolic state (Camarasa et al, 2003; Camarasa et al, 2007).

### 7.3 Paper III: Metabolic engineering of *Saccharomyces cerevisiae* for xylose consumption.

Xylose is the second most abundant monosaccharide after glucose, and the most prevalent pentose sugar found in lignocelluloses. Significant efforts have focused on the metabolic engineering of *S. cerevisiae* enabling efficient xylose utilization for fuel bioethanol production under anaerobic conditions, and although several examples of success exist, there has yet to be engineered a strain that can consume xylose aerobically without redirection of some carbon flux to overflow metabolites including ethanol, glycerol, acetate, or xylitol. This study aims to metabolically engineer *S. cerevisiae* to exclusively consume xylose while maximizing carbon flux to biomass production. Such a platform may then be enhanced with complimentary metabolic engineering strategies that couple biomass production with high value-added chemicals.

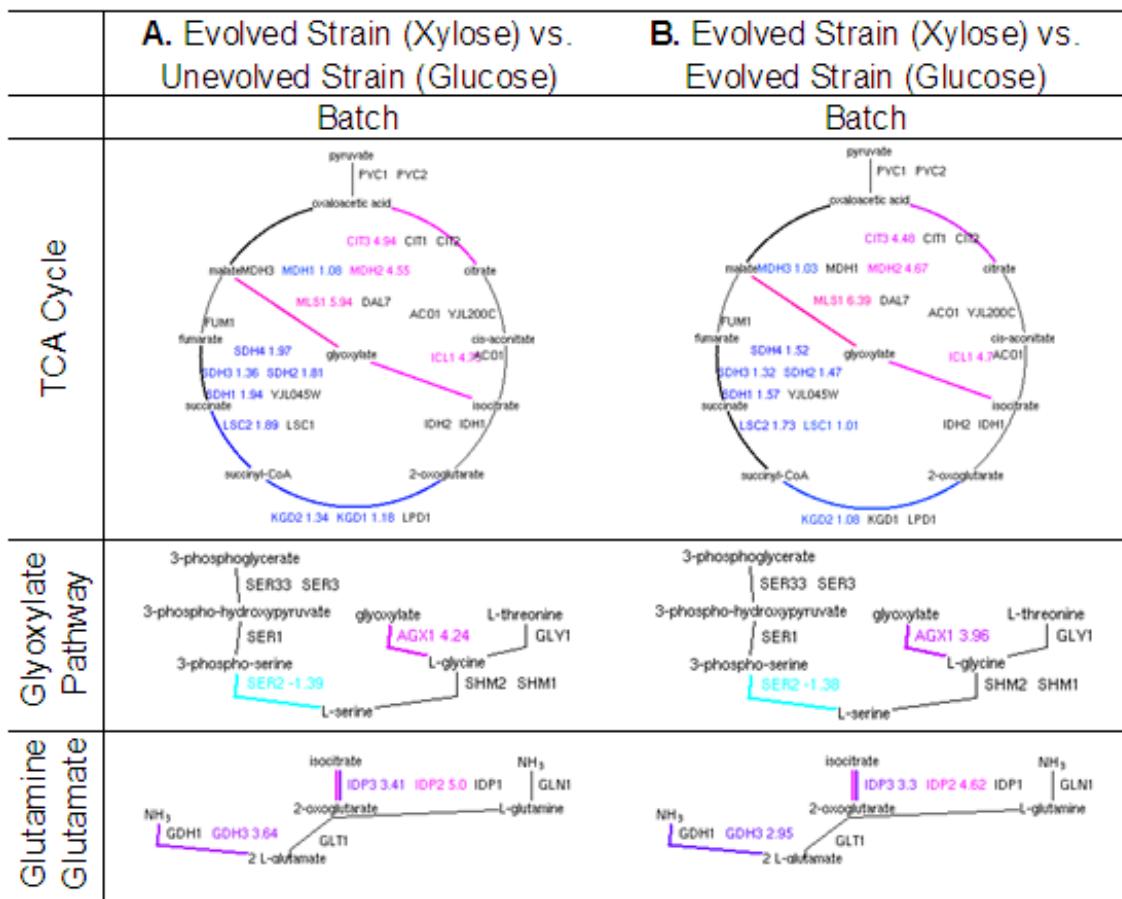
In **Paper III**, *S. cerevisiae* CEN.PK 113-3C, expressing *PsXYL1* (encoding xylose reductase, XR), *PsXYL2* (encoding xylitol dehydrogenase, XDH), and *PsXYL3* (encoding xylulose kinase, XK) from the native xylose-metabolizing yeast *Pichia stipitis*, was constructed (*S. cerevisiae* CMB.GS001), followed by a directed evolution strategy to improve xylose utilization rates. The xylose fermenting strains TMB3001, CPB.CR5, and CMB.GS001 were subjected to repetitive serial transfers in batch shake flask cultivations with minimal medium supplemented with 20 g l<sup>-1</sup> xylose. This approach targeted strain selection based on biomass formation rate, directly coupled to the xylose consumption rate. After four batch cultures, only strain CMB.GS001 demonstrated an appreciable improvement in xylose consumption. For all the other strains evaluated the residual xylose concentration measured in the culture was more than 18 g l<sup>-1</sup>. The initial total xylose consumption and biomass production of CMB.GS001 was 1.3 g l<sup>-1</sup> and 0.18 g dry cell weight, respectively. After serial cultivations over 10 cycles the xylose consumption for strain CMB.GS010 increased 15-fold to 20 g l<sup>-1</sup> and the biomass production increased 52-fold to 9.37 (g dry cell weigh) l<sup>-1</sup>. The initial specific growth rate of *S. cerevisiae* CMB.GS001 on xylose was 0.02 h<sup>-1</sup>. After these 10 transfers, covering a period of 500 h, the specific growth rate increased 9-fold

to 0.18 h<sup>-1</sup>. A total of 74 cell generations were produced across the ten cycles of directed evolution with the final 50-74 generations not yielding any improvement in specific growth rate.

Strain CMB.GS010 was physiologically characterized in stirred tank aerobic and anaerobic batch fermentations supplemented with 20 g l<sup>-1</sup> xylose or 20 g l<sup>-1</sup> glucose. The maximum specific xylose consumption rate was 0.31 g xylose (g dry cell weight)<sup>-1</sup> h<sup>-1</sup>, amongst the highest reported in the literature for aerobic growth on xylose of a *S. cerevisiae* strain carrying the genes encoding for XR, XDH and XK. Inoculated at an initial OD<sub>600</sub> of 0.01 (0.002g cell l<sup>-1</sup>), all the xylose was consumed within 60 h with biomass (62% C-mol C-mol xylose<sup>-1</sup>) and carbon dioxide (37% C-mol C-mol xylose<sup>-1</sup>) as the major fermentation products, noting the complete absence of xylitol production during the cultivation. CMB.GS010 was cultivated under anaerobic batch fermentation conditions with 20 g l<sup>-1</sup> xylose as the sole carbon source. After 100 h no growth or xylose consumption was observed. To ensure that the absence of growth was a direct consequence of the anaerobic environment, a recovery experiment was performed, where the culture was aerated quickly from an anaerobic to aerobic condition. Growth was immediately restored.

Transcriptome characterization was performed on the evolved strain (CMB.GS010) cultivated in batches with xylose and glucose as carbon sources, and continuous cultures with glucose as the sole carbon source; and the unevolved strain (CMB.GS001) with glucose as the sole carbon source in both batch and continuous cultivations (See Figure 15). The significant mRNA up-regulation of TCA cycle and glyoxylate pathways of the evolved strain on xylose compared to the unevolved or evolved strain on glucose under batch cultivations correlates well with the physiological observations that growth on xylose is dominated by respiratory metabolism. The glyoxylate pathway (*ICL1*, *MLS1*, *MDH1*, *MDH2*, *AGX1*) was significantly up-regulated in the evolved strain grown on xylose compared to the evolved strain grown on glucose or the unevolved strain grown on glucose. This pathway had a significantly higher log-fold change than succinate dehydrogenase and succinyl-CoA ligase (*SDH1*, *SDH2*, *SDH3*, *SDH4*, and *LSC2*, respectively), suggesting that the glyoxylate pathway plays an important role during respiratory metabolism of *S. cerevisiae*.

As an extension of the glyoxylate pathway, *IDP2* and *IDP3* were up-regulated significantly in all evolved strain batch xylose cultivations. Xylose metabolism requires the pentose phosphate pathway (PPP). This pathway involves the conversion of glucose-6-phosphate to 6-phosphogluconate, catalyzed by glucose-6-phosphate dehydrogenase (*ZWF1*), and further conversion to ribulose-5-phosphate with co-current production of CO<sub>2</sub>, catalyzed by 6-phosphogluconate dehydrogenase (*GND1*, *GND2*). The PPP is essential for generation of biomass precursors, which include D-ribose for nucleic acid biosynthesis, D-erythrose-4-phosphate for synthesis of aromatic amino acids, and NADPH for anabolic reactions (Jeffries, 2006). While the non-oxidative PPP satisfies D-ribose and D-erythrose-4-phosphate biomass precursor demands, cytosolic NADPH must still be generated, and the oxidative part of the pathway is by-passed during growth on xylose. Cytosolic isocitrate dehydrogenase (*Idp2*) catalyzes the oxidation of isocitrate to α-ketoglutarate, and is NADP<sup>+</sup> specific (Cherry et al, 1998). On both fermentable and non-fermentable carbon sources *Zwf1p* is constitutively expressed while *Idp2p* levels are glucose-repressed (Minard et al, 1998; Thomas et al, 1991). *Idp2p* levels have been demonstrated to be both elevated on non-fermentable carbon sources, and during the diauxic shift as glucose is depleted (Minard et al, 1998; Loftus et al, 1994; DeRisi et al, 1997). Furthermore, in *Δzwf1 Δadh6* *S. cerevisiae* mutants, it has been demonstrated that *IDP2* is up-regulated and generates enough NADPH to satisfy biomass requirements, noting that the NADP<sup>+</sup> specific cytosolic aldehyde dehydrogenase (*Adh6p*) catalyzing acetaldehyde conversion to acetate is the other major cytosolic source of NADPH (Minard et al, 2005). In the evolved strain *IDP2* and *IDP3* likely provide a source of NADPH to satisfy biomass requirements.



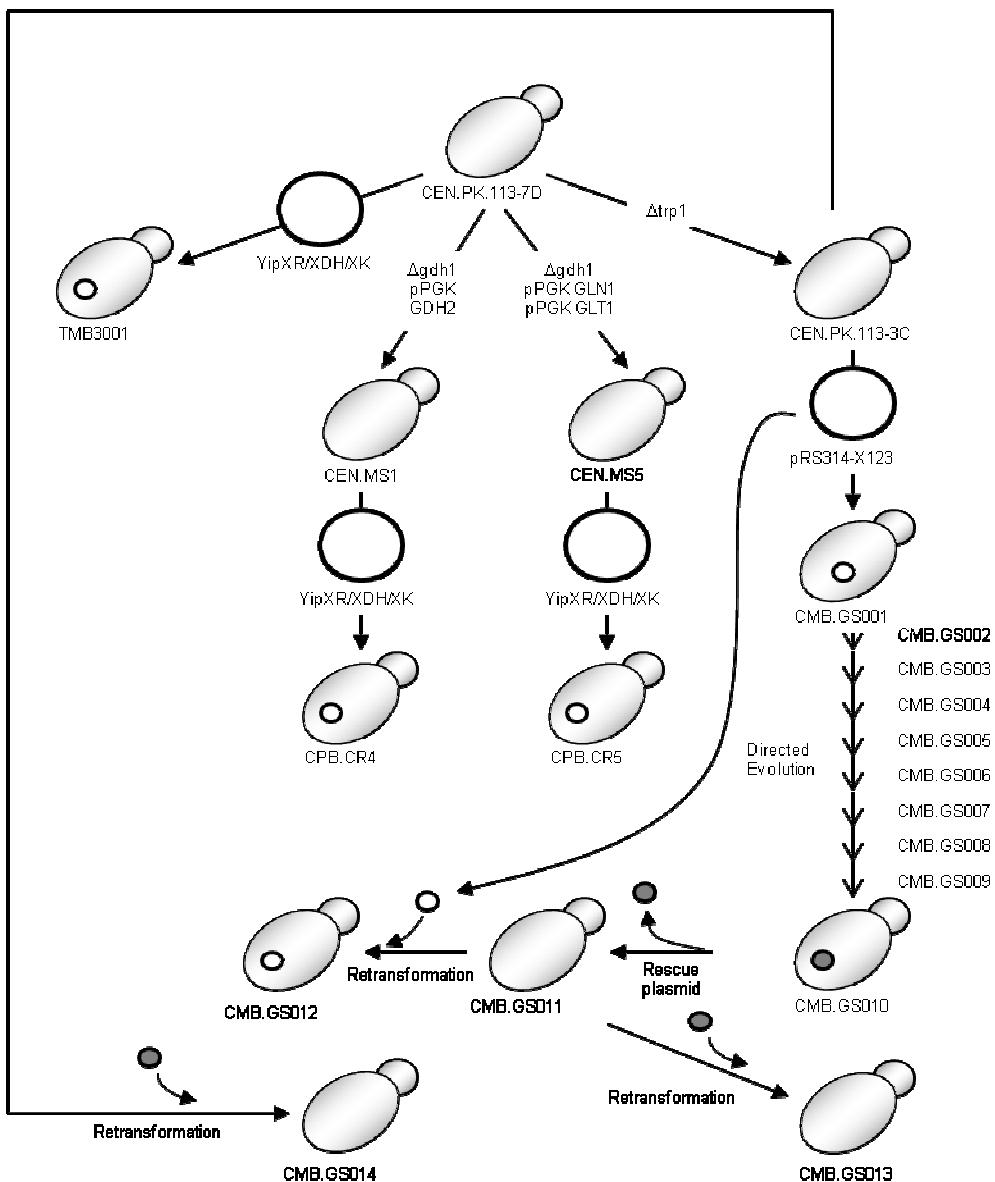
**Figure 15: Transcriptional Response of Evolved and Unevolved Strains.** Three central carbon metabolic pathways are presented: (1) tricarboxylic acid (TCA) cycle, (2) glyoxylate pathway, and (3) glutamine/glutamate synthesis. The log-fold change of significantly differentially expressed genes ( $\rho_{\text{adjusted}} < 0.01$ ,  $|\log\text{-fold change}| > 1$ ) is indicated next to the gene name. These metabolic maps are provided by the *Saccharomyces* Genome Database Pathway Expression Viewer. The comparative conditions evaluated include: (A) CMB.GS010 cultivated on batch xylose vs. CMB.GS001 cultivated on batch glucose, and (B) CMB.GS010 cultivated on batch xylose vs. CMB.GS010 cultivated on batch glucose. The terms evolved and CMB.GS010, and unevolved and CMB.GS001, are used interchangeably.

In order to investigate the possible causes of the dramatic increase in the specific growth rate of CMB.GS010 the plasmid pRS314-X123 was rescued by prolonged cultivation of CMB.GS010 on YPD medium followed by verification of plasmid loss by re-plating on minimal medium lacking tryptophan (See Figure 16). The resulting auxotrophic strain, named CMB.GS011, was transformed with pRS314-X123 (original un-evolved plasmid used to transform CMB.GS001) to obtain the strain CMB.GS012. Physiological characterization of CMB.GS012 and all subsequent strains was completed in semi-aerobic shake flasks with synthetic medium supplemented with 20 g l<sup>-1</sup> xylose. The maximum specific growth rate on xylose for CMB.GS012 was comparable to the evolved parental strain CMB.GS010. Furthermore, the plasmid extracted from CMB.GS010 was retransformed into CMB.GS011 to obtain strain CMB.GS013. CMB.GS013 exhibited the same specific growth rate as CMB.GS010 and CMB.GS012. The strain CMB.GS014 was created by transforming CEN.PK113-3C with the rescued plasmid from CMB.GS010 and exhibited specific growth rates similar to CMB.GS001. Consequently, one can conclude that the improved xylose consumption rate is a consequence of mutations in the genome and not in the plasmid carrying the properties needed for xylose metabolism.

The native xylose-fermenting strain *P. stipitis*, which is the source of the heterologous expressed enzymes, XR and XDH, does not produce xylitol during xylose fermentations (Skoog et al, 1990). Extensive xylitol formation has been observed in all the *S. cerevisiae* xylose consuming strains expressing these enzymes (Kötter et al, 1993; Tantirungkij et al, 1993; Walfridsson et al, 1995; Toivari et al, 2001; Ho et al, 1998; Eliasson et al, 2000) under anaerobic conditions. The production of xylitol has been shown to be the direct result of a redox imbalance of the NAD(P) cofactors between the XR and XDH (Roca et al, 2003; Eliasson et al, 2001; Wahlbom et al, 2002; Jeppsson et al, 2003; Verho et al, 2003; Träff-Bjerre et al, 2004; Watanabe et al, 2007). This imbalance has recently been successfully avoided by direct conversion of xylose to xylulose via the introduction of a bacterial isomerase (Kuyper et al, 2003; Kuyper et al, 2004). Xylitol formation is often described as being the major drawback of the XR-XDH strategy; however, in the engineered strain selected in this study the formation of xylitol was completely absent during all the xylose fermentations.

The absence of xylitol accumulation under oxidative conditions may be interpreted as a result of complete xylitol oxidation. Consistent with this assumption is that oxidation of xylitol to xylulose by XDH is limited by the availability of NAD<sup>+</sup>. Perhaps, the data in this study suggests, up-regulation of *IPD2* ensures sufficient NADPH production to drive xylitol catabolism.

The resulting metabolically engineered strain is a desirable platform for industrial production of biomass related products using xylose as a sole carbon source. To date, no comparable strategy expressing XR/XDH/XK has produced a strain capable of such fast aerobic growth with an absence of significant redirection of carbon flux to xylitol, glycerol, ethanol, or acetate.



**Figure 16: Strain Construction and Plasmid Rescue.** Schematic flow sheet of the construction of strains TMB3001, CPB.CR4, CPB.CRS, CMB.GS001 and CMB.GS010. Strain CEN.MS1 has been obtained deleting *GDH1* and over expressing *GDH2* in CEN.PK113-7D. Strain CEN.MS5 has been obtained deleting *GDH1* and over expressing *GLN1* and *GLT1* in CEN.PK113-7D. *GDH1* encodes for NADPH dependent glutamate dehydrogenase, *GDH2* encodes NADH dependent glutamate dehydrogenase, *GLN1* encodes glutamine synthetase, and *GLT1* encodes for glutamate synthase. Integrating vector YipXR/XDH/XK has been used to transform the strains CEN.PK113-7D, CEN.MS1, and CEN.MS5 yielding respectively, the strains TMB3001, CPB.CR4 and CPB.CRS. Centromeric plasmid pRS314-X123 was used to transform the parental strain CEN.PK113-7D yielding the strain CMB.GS001. Strain CMB.GS010 was derived from CMB.GS001 after cycles of repetitive culture selection in shake flasks. Strain CMB.GS011 was derived from CMB.GS010 after plasmid removal. Strain CMB.GS012 was obtained retransforming CMB.GS011 with the original plasmid pRS314-X123. Strain CMB.GS013 was obtained retransforming CMB.GS011 with the rescued plasmid. Strain CMB.GS014 was obtained retransforming CEN.PK 113-3C with the rescued plasmid. All strains with the designation “CMB.GS.xxx” were constructed and characterized as part of this thesis.

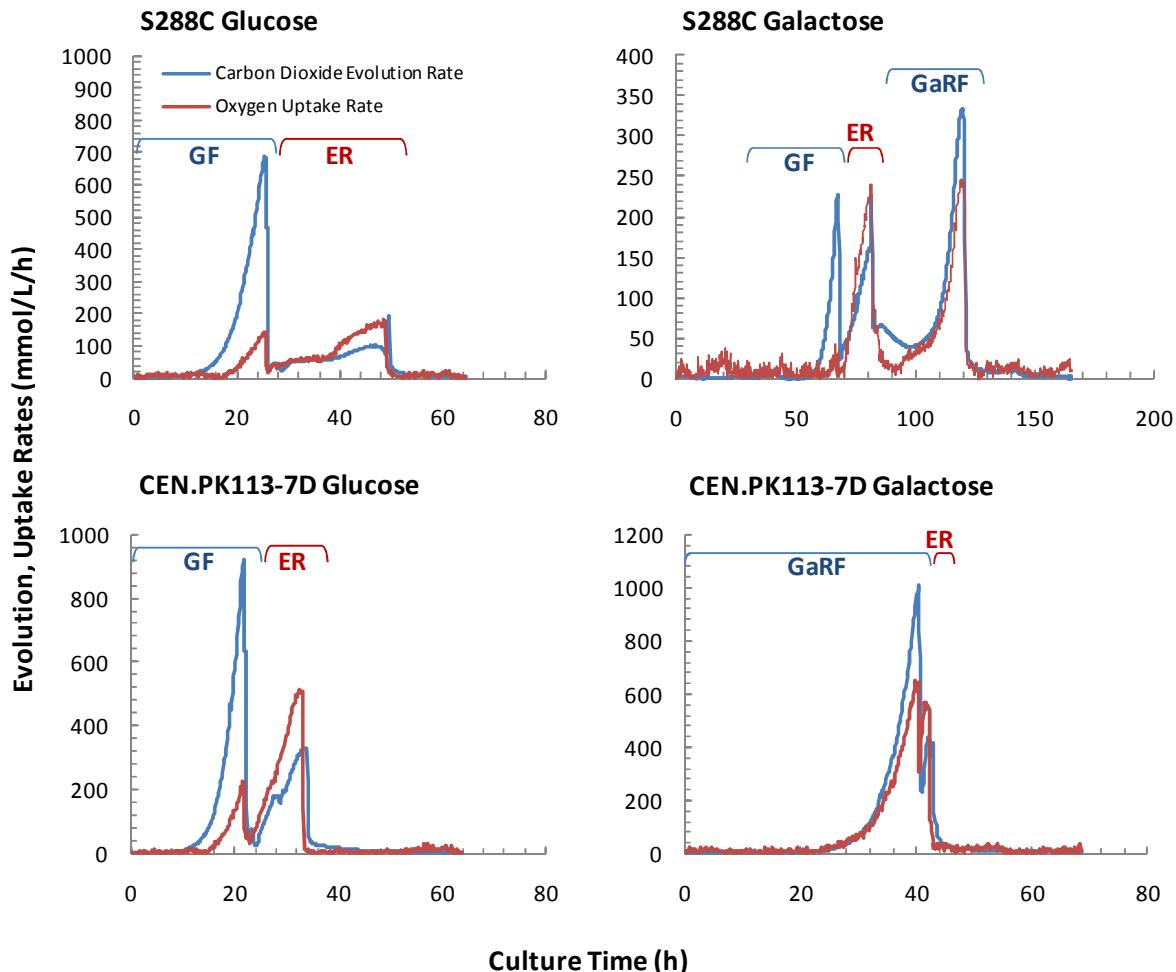
## 7.4 Paper IV: Whole genome sequencing of *Saccharomyces cerevisiae*: from genotype to phenotype for improved metabolic engineering applications.

Genome sequencing of industrially relevant organisms, including *S. cerevisiae* strain S288C, the first eukaryote genome sequence reported, provided a framework for gene annotation through functional genomics. Of particular relevance for metabolic engineering, an annotated genome sequence was a prerequisite for genome-scale metabolic network reconstructions (Goffeau, 1996; Förster et al, 2003a). Since the genome sequence of *S. cerevisiae* was released the technologies and costs associated with whole genome sequencing have advanced and decreased substantially, respectively, and this has opened the use of genome sequencing for advancing functional genomics, strain engineering, and other investigatory biology efforts (Srivatsan et al, 2008; Shendure et al, 2008; Morozova et al, 2008; Khavéjian et al, 2008; Warner et al, 2009). Furthermore, genomic hybridization to 25mer oligonucleotide tiling microarrays has been used to identify single nucleotide polymorphisms (SNPs) between S288C and the commonly used laboratory strain *S. cerevisiae* CEN.PK (Schacherer et al, 2007). These analyses lead to the identification of a total of 13,914 SNPs. However, this approach is unable to identify the exact nucleotide substitution, and consequently whether the transcribed SNP results in an amino acid substitution, presumably required to confer a change in enzyme and/or protein function.

The *S. cerevisiae* strains S288C and CEN.PK113-7D were physiologically and transcriptionally characterized in both batch fermentations with either glucose or galactose as carbon source (See Figure 17). On glucose, CEN.PK113-7D exhibited a 32% higher specific growth rate than S288C, correlating with the 33% higher specific glucose consumption rate. The CEN.PK113-7D extracellular metabolic specific productivity rates were 32.6%, 392%, and 17.9% higher for ethanol, acetate, and glycerol production compared to S288C, respectively, while the specific oxygen consumption rates were nearly equivalent ( $1.98 \text{ mmol-O}_2 \text{ g-DCW}^{-1} \text{ h}^{-1}$  for CEN.PK113-7D v.  $1.95 \text{ mmol-O}_2 \text{ g-DCW}^{-1} \text{ h}^{-1}$  for S288C). Following complete glucose fermentation, as indicated by the peak in carbon dioxide evolution rate (CER), both strains underwent a diauxic shift, clearly identified by the transition of the respiratory quotient (RQ) from  $>1$  to  $<1$ , and ethanol accumulation during growth on glucose ( $11.1 \text{ g l}^{-1}$  for CEN.PK113-7D v.  $11.3 \text{ g l}^{-1}$  for S288C). The ethanol respiratory (ER) phase was clearly distinguishable in the CEN.PK113-7D compared to S288C, as both the CER and the oxygen uptake rate (OUR) increased exponentially, corresponding with the increase in biomass ( $3.7$  to  $12.0 \text{ g-DCW l}^{-1}$ ). On the contrary, during the ER phase for S288C there was slow growth, clearly indicated by a slow increase in the CER and the OUR corresponding with a much lower increase in biomass ( $2.1$  to  $6.9 \text{ g-DCW l}^{-1}$ ). The significantly decreased ER phase in S288C compared to CEN.PK113-7D is also evident from the total time required to exhaust the ethanol (50 v. 33 h, respectively) (See Figure 17).

A similar characterization was performed using galactose as the carbon source. CEN.PK113-7D demonstrated a slight lag-phase compared to glucose fermentation; however, sustained a galactose specific growth rate of  $0.27 \text{ h}^{-1}$  and a galactose uptake rate of  $24.3 \text{ C-mmol g-DCW}^{-1} \text{ h}^{-1}$ , representing a 34% and 77% reduction, respectively, compared to glucose. Similarly, S288C was cultivated on galactose; however, a significant deficiency in the strain's ability to metabolize this carbon source was observed. A total of 25 h post-inoculation elapsed with no increase in biomass as compared to CEN.PK113-7D where after 6h post-inoculation two cell doublings were observed. At 25 h post-inoculation a glucose bolus of  $10 \text{ g l}^{-1}$  was added to promote growth, and rapidly, glucose fermentation, a diauxic shift, and ethanol respiration were observed. Both co-consumption of galactose and ethanol, and a galactose only respiro-fermentative growth phase was observed. During co-consumption the specific growth rate was  $0.14 \text{ h}^{-1}$ , while on galactose alone the specific growth rate was as low as  $0.02 \text{ h}^{-1}$ . Similarly, the

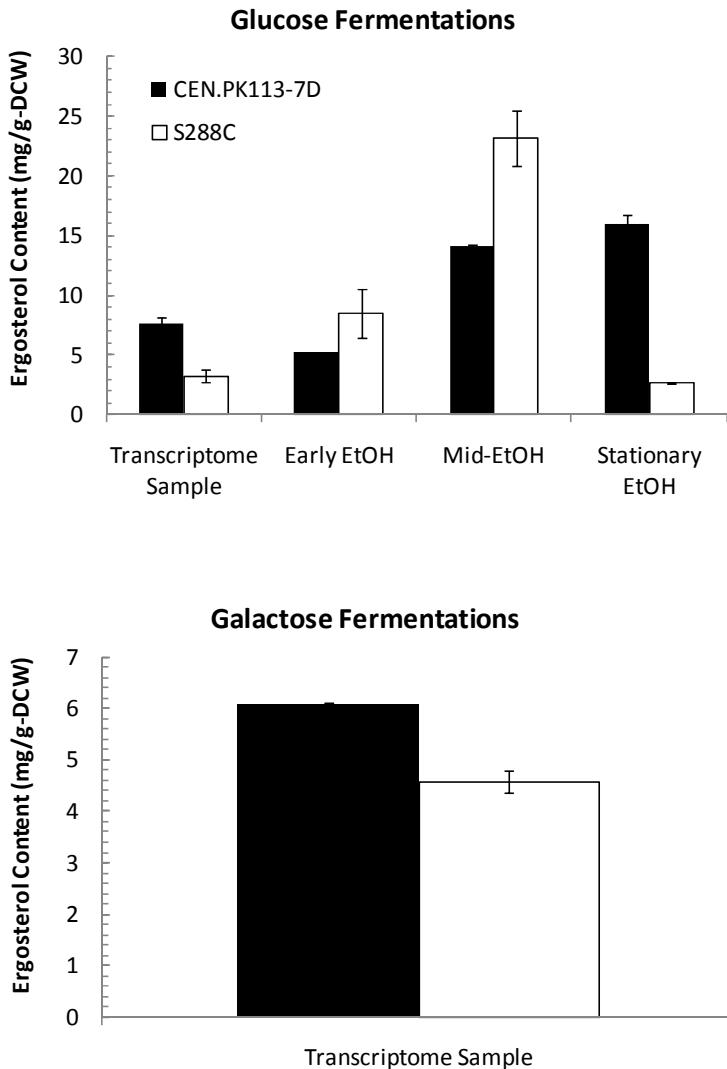
extracellular specific metabolite productivity rates were nearly zero when only galactose consumption was considered. Ethanol was consumed by 82 h post-inoculation, and in the period from 82 h to 128 h, only galactose consumption was observed, and biomass increased from 7.9 g-DCW L<sup>-1</sup> to 20.9 g-DCW L<sup>-1</sup>, representing a doubling time of 35 h compared to 2.6 h for CEN.PK113-7D (See Figure 17).



**Figure 17: Physiological Characterization of *S. cerevisiae* S288C and CEN.PK113-7D.** The plots above show the carbon dioxide evolution rate and oxygen uptake rate as a function of cultivation time for the strains S288C and CEN.PK113-7D supplemented with glucose and galactose, respectively. Glucose fermentation (GF), ethanol respiration (ER), galactose respiro-fermentation (GaRF). The black arrow in the S288C Galactose plot indicates when 20 g L<sup>-1</sup> glucose was supplemented (25h) when no growth was observed on galactose.

For each of the fermentations ergosterol measurements were performed at the same time samples were taken for transcriptome analysis. It is seen that during the mid-exponential phase of glucose fermentation (18-20h), the total ergosterol content was substantially higher in CEN.PK113-7D than in S288C ( $7.6 \pm 0.5$  mg g-DCW<sup>-1</sup> v.  $3.3 \pm 0.5$  mg g-DCW<sup>-1</sup>) (See Figure 18). However, during the diauxic shift and the ER phase S288C had significantly higher ergosterol content than CEN.PK113-7D, but post-ethanol metabolism CEN.PK113-7D again exhibited significantly higher ergosterol content ( $15.9 \pm 0.7$  mg g-DCW<sup>-1</sup> v.  $2.6 \pm 0.07$  mg g-DCW<sup>-1</sup>). For the galactose fermentations, the ergosterol content was only measured at the time of sampling for transcriptome analysis, which occurred at 78 h for S288C (co-consumption of ethanol and galactose observed), and 35 h for CEN.PK113-7D. The total ergosterol content on

galactose was  $6.1 \pm 0.04$  mg g-DCW $^{-1}$  and  $4.6 \pm 0.2$  mg g-DCW $^{-1}$  for CEN.PK113-7D and S288C, respectively. This is consistent with previous work where CEN.PK2-1C had very high ergosterol/erg-ester ( $20.0$  mg g-CDW $^{-1}$ ) and triacylglycerols content ( $15.2$  mg g-CDW $^{-1}$ ) compared to 9 other *S. cerevisiae* strains, including FY169 (ergosterol/erg-ester content:  $8.5$  mg g-CDW $^{-1}$ ; triacylglycerols content:  $2.4$  mg g-CDW $^{-1}$ ) which is isogenic to S288C (Daum et al, 1999; Winston et al, 1995).

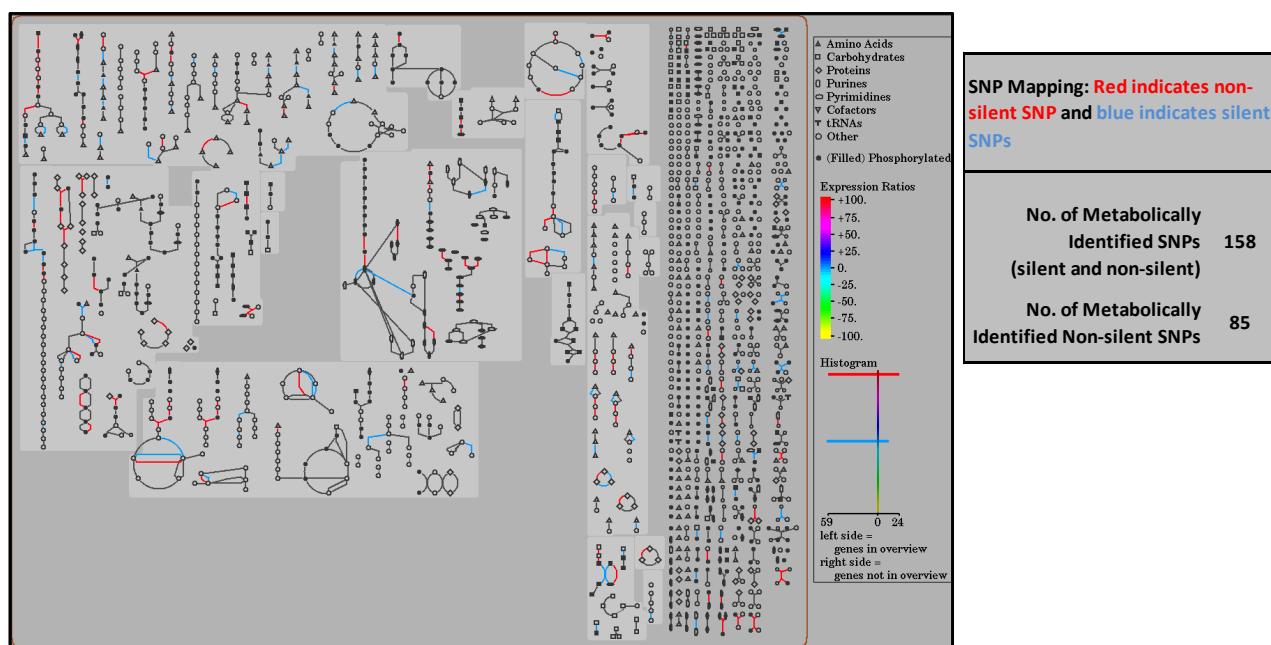


**Figure 18: Ergosterol Measurements in *S. cerevisiae* Strains S288C and CEN.PK113-7D.** Ergosterol content (mg g-DCW $^{-1}$ ) was measured for different samples taken during S288C and CEN.PK113-7D fermentations, supplemented with glucose and galactose. Transcriptome Sample was taken during the mid-exponential fermentation phase on glucose or respiration phase on galactose. For glucose fermentations, early ethanol, mid-ethanol, and stationary ethanol samples were taken post-diauxic shift to characterize the change in ergosterol during growth on ethanol. Error bars are  $\pm$  SD ( $n=2$ ).

Data from whole genome sequencing, including the number of reads, average coverage relative to the SGD reference genome, total number of non-ambiguous SNPs, and total number of filtered SNPs were generated. Not surprisingly, S288C had relatively few SNPs compared to CEN.PK113-7D given that the reference genome from SGD is based on S288C v 12.0 (Cherry et al, 1998). Furthermore, the 13,787 filtered SNPs identified using the MAQ software is in line with the previously estimated 13,914 SNPs for CEN.PK113-7D based upon DNA hybridization to 25mer oligonucleotide microarrays (Schacherer et al, 2007). A total of 782 metabolic genes as defined by SGD were used to query for SNPs in both the S288C and CEN.PK113-7D genome sequences. A total of 36 metabolic SNPs, 3 of which are non-silent, were identified across 14 independent metabolic genes (3 non-silent SNPs distributed across 3 metabolic genes) were

identified in S288C. A significantly higher number of metabolic SNPs, 939, were detected in CEN.PK113-7D and distributed across 158 unique metabolic genes, 85 of which contained a total of 219 non-silent SNPs (See Figure 19).

Differential gene expression between S288C and CEN.PK113-7D, cultivated on both glucose and galactose, was measured (See Figure 20). For the condition S288C v. CEN.PK113-7D cultivated on glucose, the top 272 differentially expressed genes, ranked according to  $p_{adj}$  value, were largely dominated by GO process terms response to stimuli and pheromone, and with the dominant metabolic process categories being trehalose metabolism, steroid metabolism, and amino acid transport. Some specific genes consistent with this categorization are *GSY1* (glycogen synthase) and *HMG1* (HMG-CoA reductase). For growth on galactose the top 501 differentially expressed genes, ranked according to  $p_{adj}$  value are characterized into GO process terms response to stimuli and stress, carbohydrate metabolism, and transport. Some specific metabolic genes expressed higher in S288C v. CEN.PK113-7D include *MDH2* (malate dehydrogenase), *FBP1* (fructose-1,6-bisphosphatase), *GAD1* (glutamate decarboxylase), *GDH3* (NADP<sup>+</sup> dependent glutamate dehydrogenase), *GSY1*, and *ICL1* (isocitrate lyase). Similarly, some specific metabolic genes that are lower expressed in S288C v. CEN.PK113-7D include *ARE2* (acyl-coA:sterol acetyltransferase), and *CYB5* (cytochrome b5).



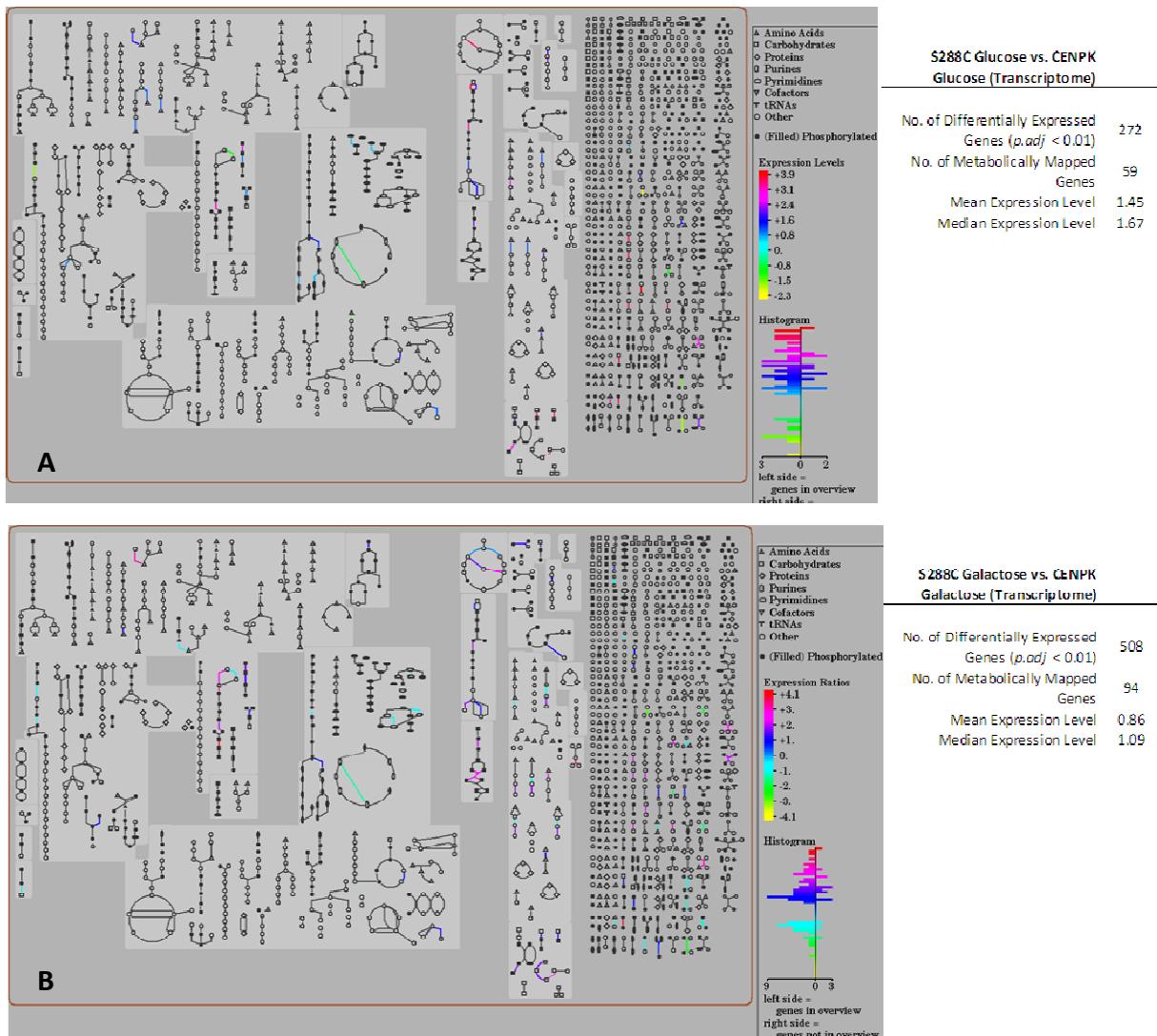
**Figure 19: SNP Enrichment in *S. cerevisiae* Metabolism.** The metabolic map produced using the Saccharomyces Genome Database (SGD) Expression Viewer (SRI International Pathway Tools version 12.0, based upon *Saccharomyces cerevisiae* S288C, version 12.0) was created using the SNP data produced for CEN.PK113-7D compared to S288C. Pathways in red indicate non-silent SNPs (85 genes) while those in blue indicate silent SNPs (73 genes). Note that number of genes does not necessarily coincide with number of pathways due to isoenzymes.

In an effort to characterize the non-silent metabolic SNPs identified in CEN.PK113-7D with biological significance, GO process categorization was performed and ranked according to significance ( $p < 0.01$ ). The most significant categories include carboxylic acid, organic acid, carbohydrate metabolism, followed by nitrogen, amino acid, lipid, aromatic compound, and glycoprotein metabolism.

A graphical overview of all silent and non-silent SNPs mapped to their specific metabolic pathways highlights two metabolic pathways, galactose uptake and ergosterol synthesis, where an enrichment of non-silent and silent SNPs was observed. The ergosterol biosynthetic pathway

had significant non-silent SNPs identified in *ERG8* and *ERG9*, and silent SNPs identified in *ERG20* and *HMG1*. Both *ERG8* and *ERG9* were not significantly differentially expressed, either in glucose or galactose, suggestive again that phenotypic observations, consistent with genome sequence variations, are not necessarily directly manifested at the transcriptome level. Both *ERG8* (encodes phosphomevalonate kinase) and *ERG9* (encodes squalene synthase) are essential cytosolic enzymes in the biosynthetic pathway of isoprenoids and sterols ( $\Delta erg8$  and  $\Delta erg9$ , were found to both be auxtrophic for ergosterol in the systematic deletion library), including ergosterol, from mevalonate (Tsay et al, 1991; Jennings et al, 1991; Cherry et al, 1998). The ergosterol biosynthetic pathway is highly regulated through feedback inhibition mechanisms and by several rate-controlling steps, including that catalyzed by HMG-CoA reductase, encoded by *HMG1* (Basson et al, 1988; Maury et al, 2005). Under both glucose and galactose, *HMG1* expression was significantly down-regulated in S288C compared to CEN.PK113-7D by 3.2-fold ( $p_{adj}$  value =  $3.3 \times 10^{-4}$ ) and 1.8-fold ( $p_{adj}$  value =  $8.6 \times 10^{-3}$ ), respectively, correlating with the significantly lower ergosterol content in S288C. Furthermore, *ERG9* has been previously identified as also having a regulatory role (Grabowska et al, 1998), consistent with the hypothesis that a non-silent SNP resulting in altered protein function could affect ergosterol synthesis. *ERG8* on the other hand has not been explicitly shown to have a regulatory function, yet, when the specific activity of  $0.06 \text{ } \mu\text{mol min}^{-1} \text{ mg}^{-1}$  is compared to other ergosterol synthetic enzymes such as *ERG13* (2.1 in *S. cerevisiae*), *ERG12* (0.77 in *S. cerevisiae*), *ERG20* (5.22 in *S. cerevisiae*), and especially the known regulator *HMG1/HMG2* (0.0035 in *S. cerevisiae*) it is suggestive that *ERG8* is likely a rate limiting step (Middleton et al, 1975; Gray et al, 1972; Tchen, 1958; Porter, 1985; Eberhardt et al, 1975; Rilling, 1985; Basson et al, 1986; Durr et al, 1960; Bloch et al, 1959). There were a large number of non-silent SNPs that encoded significant changes in amino acid classes, further suggestive that *ERG8* is a likely to be involved in control of flux through the ergosterol pathway. Lastly, the observation that neither *ERG8* nor *ERG9* were differentially expressed under glucose or galactose, suggests their potential affect on phenotype is likely post-transcriptional.

Similar to ergosterol biosynthesis, the flux through the galactose uptake pathway was much lower in S288C compared with CEN.PK113-7D, correlating with the non-silent SNP enrichment in *GAL1* and *GAL10*, and silent SNPs in *GAL7*. Neither *GAL1* (encodes galactokinase) nor *GAL10* (encodes UDP-glucose-4-epimerase) were significantly differentially expressed during growth on galactose; however, on glucose *GAL1* was significantly up-regulated ( $p_{adj}$  value =  $9.7 \times 10^{-4}$ ) 2.9-fold in CEN.PK113-7D. Both  $\Delta gal1$  and  $\Delta gal10$  mutants are unable to grow on galactose as sole carbon sources (Bhat et al, 1990; Bhat et al, 1992; Douglas et al, 1964). The significant number of non-silent SNPs in both essential galactose genes suggests obvious targets for explanation of why S288C is incapable of galactose respiratory metabolism. Furthermore, it should be noted that while S288C has been described as  $\Delta gal2$ , no SNPs were detected between CEN.PK113-7D and S288C, and CEN.PK113-7D was able to readily metabolize galactose meaning that a functional *GAL2* (encodes galactose permease, required for galactose utilization) is present in both S288C and CEN.PK113-7D.



**Figure 20: Transcriptome Response to Glucose and Galactose of *S. cerevisiae* Strains S288C and CEN.PK113-7D.** The metabolic map, produced using the Saccharomyces Genome Database (SGD) Expression Viewer (SRI International Pathway Tools version 12.0, based upon *Saccharomyces cerevisiae* S288C, version 12.0) was created using statistically significant log-fold expression values for S288C glucose vs. CEN.PK113-7D glucose (A), and S288C galactose vs. CEN.PK113-7D galactose (B).

In summary and perhaps not surprisingly, transcriptome analysis did not provide a clear hypothesis for major phenotypes observed, suggesting that genotype to phenotype correlations are manifested post-transcriptionally or post-translationally either through protein concentration and/or function (See Figure 19 and Figure 20). Clearly, future work must validate these correlations through genetic engineering of identified SNPs in either S288C or CEN.PK113-7D to see if desired phenotypes, such as increased galactose uptake or ergosterol synthesis in S288C, are observed. Future work must also expand on the metabolic SNP analysis presented to include all 13,787 SNPs, realizing that phenotypic observations may not necessarily be linked directly to metabolic SNPs, but rather SNPs affecting larger regulatory mechanisms and networks, such as those governed by transcription factors. Certainly, as CEN.PK113-7D and S288C continue to be exploited as laboratory strains, and *S. cerevisiae* more generally, for metabolic engineering applications, the integration of physiological characterization, transcriptome analysis, and metabolic SNP detection with the genome sequences provided here direct correlations between

observed phenotypes and genotypes can be established and hence allow for a wider use of genome sequencing in metabolic engineering.

## 7.5 Results Not Included in Papers

There are three major results that were not included in any of the papers included in this thesis. These results, while providing insight into the physiology of the various strains of *S. cerevisiae* considered in **Papers I, II, and III**, were supplementary to the primary findings reported. **Paper I** specifically noted that the carbon balance of the 8D and 8D Evolved mutants did not close; therefore an investigation was launched to determine the identity of the missing carbon product(s). **Paper I** and **Paper II**, as well as this thesis, highlight that one of the advantages of the bacterial succinate production platforms is incorporation of CO<sub>2</sub> through carboxylation reactions, specifically pyruvate carboxylase. Therefore, an investigation was performed to determine the effect of CO<sub>2</sub> supplementation on *S. cerevisiae* fermentations. **Paper III** noted that the objective was to construct a microbial cell factory capable of fast and efficient xylose utilization under aerobic conditions without any diversion of carbon to overflow metabolites. The mutant constructed, CMB.GS010, could potentially serve as an ideal platform for succinic acid overproduction; however, as clearly shown in the results, this mutant consumed xylose using respiratory metabolism. Therefore, an investigation was launched to determine if CMB.GS010 could sustain respiratory xylose metabolism with interruption of the oxidative TCA cycle – a requirement of any succinic acid overproducer.

In summary, the three results included: (1) detection and suggestion of the carbon species missing in the *S. cerevisiae* 8D and 8D Evolved strains, (2) the physiological effect of CO<sub>2</sub> supplementation on *S. cerevisiae* CEN.PK113-5D and 8D Evolved and (3) the physiological effect of malonate supplementation on *S. cerevisiae* CMB.GS010. These results are briefly presented and discussed. The materials and methods related to medium, cultivation, and analytical methods were identical to those described in **Papers I, II, III, and IV**. Each of the results described here while strongly suggestive require further investigation and development before they can be published.

### 7.5.1 8D and 8D Evolved Carbon Balance

In *S. cerevisiae* CEN.PK113-7D the carbon recovery on minimal medium supplemented with 20 g l<sup>-1</sup> glucose in aerobic batch fermentations was 95.7 ± 3.2% (*n*=3, ± SD). In *S. cerevisiae* 8D, 8D Evolved and 8D Evolved with pICL1 (See **Paper I** for detailed strain genotype and construction) grown under identical conditions, the carbon recovery was 76.3 ± 1.7%, 70.7 ± 2.1%, and 72.1 ± 1.8%, respectively (*n*=3, ± SD). During the course of fermentation of all the 8D based strains, there was a very strong odor noted in the off-gas towards the end of batch fermentation (<5 g l<sup>-1</sup> glucose). This odor was pronounced, and appeared to be an aldehyde. In particular, when compared with stock solutions of 5.0 g l<sup>-1</sup> acetaldehyde (Sigma-Aldrich Catalogue No. 402788, CAS No. 75-07-0, ≥99.5% ACS reagent pure), the odors were similar. Given the relatively large fraction of carbon not accounted for, and the potential opportunity for redirection of carbon flux through metabolic engineering to succinate, a small investigation was launched to detect the identity of the carbon species missing.

End of fermentation samples (See **Paper I**) were analyzed after filtration using a Solid Phase Microextraction (SPME) Method. The detailed methodology employed has been previously described (Asadollahi et al, 2008). In short, 2.0 ml of supernatant sample stored at -20°C in 2.0 mL Eppendorf tubes were thawed and immediately transferred to glass vials. Chemical analysis was performed by extracting volatile compounds from the head-space of the samples onto 100 µm polydimethyl siloxane (PDMS) fibers (Supelco, Bellefonte, PA). Post-thaw,

extraction was performed for 30 min at 70°C while mixing the sample with a small magnet. The analytes were thermally desorbed at 250°C from the SPME fiber onto the injector of a gas chromatograph in the splitless mode. The oven temperature was held initially at 45°C (1 min), then raised to 130°C by a ramp function of 10°C/min followed by a ramp of 3°C/min to 160°C. The oven temperature was further raised to 250°C at 10°C/min and maintained for 5 min to equilibrate. The samples were analyzed by GC-MS to determine the concentration of volatile compounds. GC-MS analyses were run on a Thermo Finnigan Focus GS coupled to a Focus DSQ quadropole mass spectrometer. Analytes from the SPME fiber sample were separated on an SLB-5 ms capillary column (30 m x 0.25 mm inner diameter, 0.25 µm film thickness; Supelco, Bellefonte, PA) using helium carrier gas (1.2 mL/min). Quantification of volatile compounds was carried out using standard curves of acetaldehyde generated after each analysis run. Specifically, acetaldehyde standard solutions (diluted in MilliQ water) at concentrations of 0.01, 0.1, 1.0, 5.0, 10.0, and 15.0 g l<sup>-1</sup> were analyzed. Furthermore, end of fermentation samples from 8D Evolved and CEN.PK113-7D batch fermentations were analyzed.

The results according to GC-MS confirmed a peak with identical retention time in the 8D Evolved samples to the acetaldehyde standard solutions (retention time, 1.5-2.0 min) and the same peak was absent in the CEN.PK113-7D fermentation samples. However, measurements and samples were difficult to reproduce given the very high volatility of the samples. Therefore, although there was strong evidence to suggest that in fact acetaldehyde concentrations were being detected in the 8D Evolved fermentations, the exact concentration was not estimated due to poor linearity and reproducibility using standard solutions (e.g., due to high volatility there was no detection of acetaldehyde in standard solutions <1.0 g l<sup>-1</sup>).

Therefore, assuming the identity of the missing carbon could partially be acetaldehyde, a theoretical calculation was performed to approximate the expected concentration of acetaldehyde at the end of 8D Evolved fermentations to recover 100% of the carbon. The thermodynamic approximation presented in Figure 21 considers two scenarios: a “low case” where 25% of carbon from glucose is diverted to acetaldehyde and a “high case” where 35% of carbon from glucose is diverted to acetaldehyde. The final concentration of acetaldehyde estimated to be in the liquid fermentation phase is 2.30 and 3.21 g l<sup>-1</sup> acetaldehyde, respectively. The estimated values are in-line with the GC-MS measurements. These concentrations of acetaldehyde are high for wild-type *S. cerevisiae* fermentations on glucose supplemented minimal medium with previous literature reporting detection of 0.1-0.2 g l<sup>-1</sup> acetaldehyde (Remize et al, 2000), (Romano et al, 1994), and furthermore describing a strong growth inhibition and toxicity to both elevated acetate and acetaldehyde concentrations (>0.5 g l<sup>-1</sup>).

$\begin{array}{c} \text{O} \\    \\ \text{CH}_3 - \text{C} - \text{H} \end{array}$	MW [g/mole] <b>Acetaldehyde</b> 44.05316 CH <sub>3</sub> CHO Molecular Formula			
		LOW CASE		
<b>PRIMARY INPUT</b>	Carbon Missing [C-mol/C-mol glucose]	0.250	Carbon Missing [C-mol/C-mol glucose]	0.350
	Acetaldehyde MW [g/C-mol]	22.027	Acetaldehyde MW [g/C-mol]	22.027
	Glucose MW [g/C-mol]	27.360	Glucose MW [g/C-mol]	27.360
	Convert to [g-Acetaldehyde/g-Glucose]	0.201	Convert to [g-Acetaldehyde/g-Glucose]	0.282
	Concentration of Glucose [g/L]	20.000	Concentration of Glucose [g/L]	20.000
	Convert to [g-Acetaldehyde/L]	4.025	Convert to [g-Acetaldehyde/L]	5.636
<b>PRIMARY OUTPUT</b>	<b>MAXIMUM FERMENTER CONCENTRATION OF ACETALDEHYDE [g/L]</b>	4.025	<b>MAXIMUM FERMENTER CONCENTRATION OF ACETALDEHYDE [g/L]</b>	5.636
	[mol/L]	0.091	[mol/L]	0.128
<hr/>				
<b>HENRY's LAW CONSTANT ESTIMATION</b>	Reactor Temp. [K]	303.000	$k_H = k_H^o \cdot e^{\left( \frac{-\Delta H_{soln}}{R} \left( \frac{1}{T} - \frac{1}{T^o} \right) \right)}$	
	Reactor Pressure [bar]	1.025	$\frac{-d \ln k_H}{d \left( \frac{1}{T} \right)} = \frac{\Delta H_{soln}}{R}$	
	Standard Temperature [K]	298.15K		
	Ideal Gas Constant, R, [L-atm/K-mol]	0.082		
	Henry's Law Constant, at standard temperature [M/atm]	14.264		
	d(lnK)/d(1/T) [K]	5371.429		
	Henry's Law Constant, at standard temperature [mol/L/atm]	14.264		
	Temp. Compensated Henry's Law Constant [mol/L/atm] (Apply Eqn. 1)	10.690		
<b>ACETALDEHYDE LOST IN VAPOR PHASE</b>	Partial Pressure Acetaldehyde [LOW], [atm]	0.977	Partial Pressure Acetaldehyde [HIGH], [atm]	1.368
	Concentration of Acetaldehyde in Vapor, Ideal Gas Law, [LOW] [g/L]	1.730	Concentration of Acetaldehyde in Vapor, Ideal Gas Law, [HIGH] [g/L]	2.422
<b>MASS BALANCE OF ACETALDEHYDE in LIQUID PHASE</b>	Acetaldehyde Remaining in Fermenter [LOW], [g/L]	2.296	Acetaldehyde Remaining in Fermenter [HIGH], [g/L]	3.214

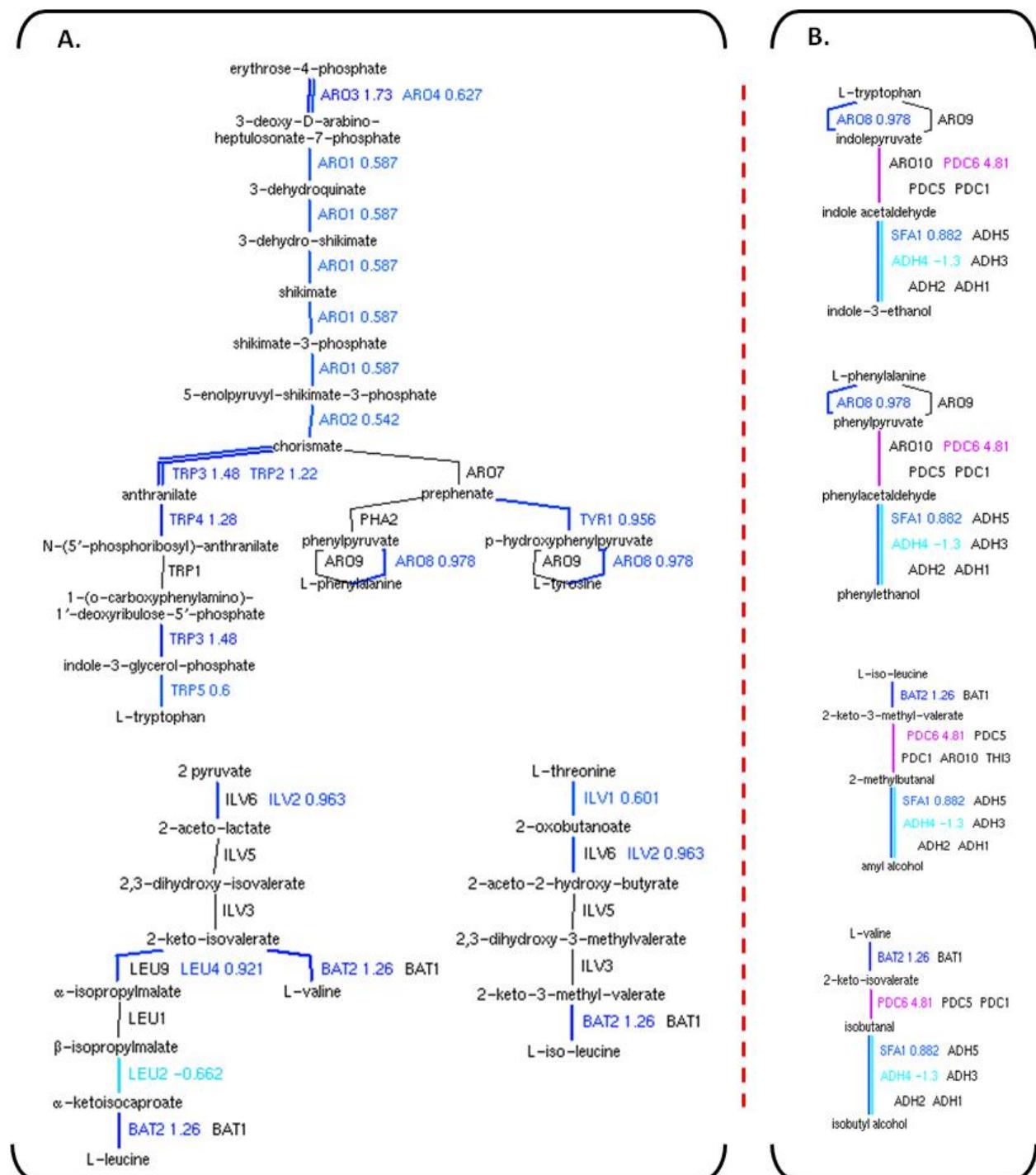
**Figure 21: Theoretical Estimation of End of Fermentation Acetaldehyde Concentration.** The thermodynamic analysis presented above estimate the acetaldehyde concentration remaining in the fermenter assuming 25% (LOW CASE) or 35% (HIGH CASE) of all glucose carbon is diverted to acetaldehyde.

In Paper I the transcriptome analysis of 8D Evolved compared to the reference strain, CEN.PK113-5D, is described under aerobic batch glucose fermentation conditions. Included in that transcriptome analysis was the metabolic mapping of statistically significant differentially expressed genes (See Figure 22).

In the 8D Evolved mutant amino acid bio-synthetic pathways, including L-tryptophan, L-phenylalanine, L-tyrosine, L-iso-leucine, L-valine, and L-leucine (note: *LEU2* down-regulated - 0.662 log<sub>2</sub>-fold, yet *ILV2*, *LEU4*, and *BAT2* are up-regulated), all exhibit a high functional enrichment of up-regulated pathway genes. Similarly, the four pathways for L-tryptophan, L-phenylalanine, L-iso-leucine, and L-valine degradation are up-regulated through *PDC6* (pyruvate decarboxylase iso-enzyme, 4.81 log<sub>2</sub>-fold). Within the same pathways, *ADH4* (alcohol dehydrogenase iso-enzyme IV) is significantly down-regulated (-1.3 log<sub>2</sub>-fold) suggesting a potential bottle-neck in the pathway. The intermediate metabolite that would likely accumulate as a result is indole acetaldehyde, phenylacetaldehyde, 2-methylbutanal, and isobutanal. Furthermore, *PDC6* is amongst three isoenzymes (*PDC1*, *PDC5*, *PDC6*) that convert pyruvate to equimolar acetaldehyde and CO<sub>2</sub>. It is interesting to note that previous studies have noted that *PDC6* is not expressed during glucose fermentation, but rather during growth on non-fermentable carbon courses (Hohmann, 1991a; Hohmann, 1991b).

Although significantly more analytical and quantitative analysis of the volatile species resulting in the 8D Evolved fermentations would be required for verification, there is significant

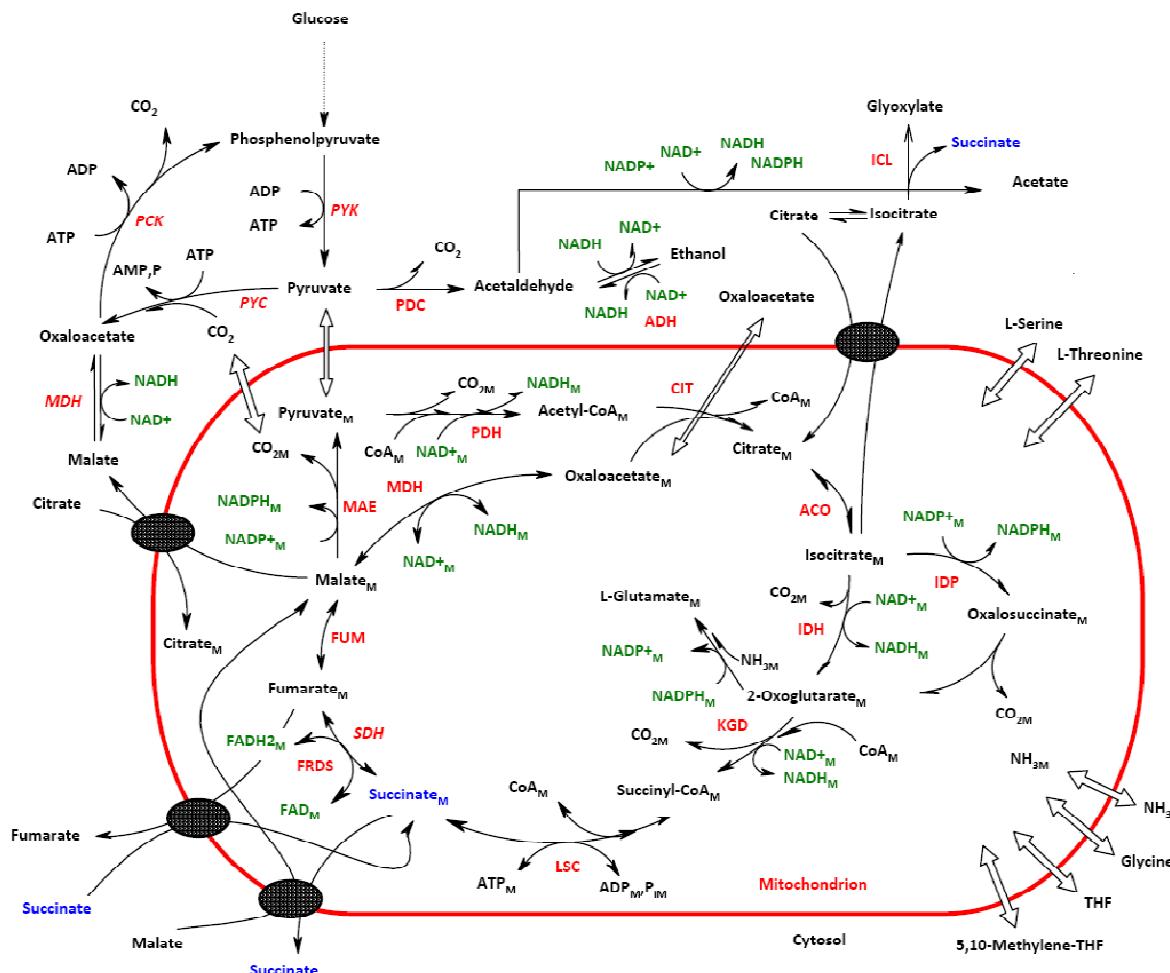
preliminary evidence to suggest production and accumulation of acetaldehyde and/or other related volatile compounds such as indole acetaldehyde, phenylacetaldehyde, 2-methylbutanal, and isobutanol. Furthermore, there is strong suggestion that future metabolic engineering targets, if the 8D Evolved strategy were to be further pursued, would include diversion of carbon away from acetaldehyde.



**Figure 22: Significant Differential Gene Expression (8D Evolved vs. CEN.PK113-5D).** The biosynthetic pathways of various amino acids and metabolic intermediates are shown with the  $\log_2$  fold change indicated next to the gene name. Panel A shows the significant up-regulation of L-tryptophan, L-phenylalanine, L-tyrosine, L-isoleucine, L-valine, and L-leucine biosynthetic pathways (note: *LEU2* down-regulated  $-0.662 \log_2$ -fold). Panel B shows the catabolism of L-tryptophan, L-phenylalanine, L-iso-leucine, and L-valine.

## 7.5.2 Carbon Dioxide Supplementation

As discussed in section 5.0, carboxylation of pyruvate via pyruvate carboxylase has proven critical for reductive TCA cycle succinate formation in bacterial expression hosts. Reactions and enzymes where carboxylation is observed in central carbon metabolism are indicated in Figure 23.



**Figure 23: Carboxylation Reactions Leads to Succinate Formation.** The metabolic pathways shown include pyruvate/phosphoenolpyruvate metabolism, the glyoxylate by-pass, and the TCA cycle. The enzymes listed include: PYK (pyruvate kinase), PYC (pyruvate carboxylase), PCK (phosphoenolpyruvate carboxykinase), PDC (pyruvate decarboxylase), MDH (malate dehydrogenase), ADH (Alcohol dehydrogenase), ICL (isocitrate lyase), PDH (pyruvate dehydrogenase), CIT (citrate synthase), ACO (aconitase), IDH (isocitrate dehydrogenase), IDP (NADP-specific isocitrate dehydrogenase), KGD (α-keto-glutarate dehydrogenase), LSC (succinyl-CoA ligase), SDH (succinate dehydrogenase), OSM/FRDS (fumarate reductase), MAE (malic enzyme).

In *S.cerevisiae*, the anaplerotic synthesis of oxaloacetate from pyruvate is catalyzed by pyruvate carboxylase (encoded by *PYK1* and *PYK2*). Pyruvate kinase (encoded by *PYK1* and *PYK2*) converts PEP to equimolar pyruvate, and PEP carboxykinase (encoded by *PCK1*) catalyzes the conversion of oxaloacetate to equimolar PEP. *PYK1* and *PYK2* consume  $\text{HCO}_3^-$ , while *PCK1* produces  $\text{HCO}_3^-$ .

Recently, the transcriptional response of *S.cerevisiae* to elevated levels of  $\text{CO}_2$  in chemostat culture has been investigated (Aguilera et al, 2005). In particular, three different cultivation conditions were investigated: aerobic glucose-limited, anaerobic glucose-limited, and aerobic nitrogen-limited cultivations. The effects of  $\text{CO}_2$  on cellular physiology were most pronounced in

aerobic glucose-limited chemostat cultures, while biomass and product yields in anaerobic cultures were not significantly affected by elevated CO<sub>2</sub> concentrations (See Figure 24). An intermediate effect was observed for nitrogen-limited cultures. In all experiments, elevated CO<sub>2</sub> supplementation was at the concentration of 79% v v<sup>-1</sup>. Under all three aeration and nutrient limitations only a small fraction of the genome showed a significant transcriptional response to elevated CO<sub>2</sub> concentration. In aerobic carbon-limited conditions under elevated CO<sub>2</sub> concentration, *PCK1* and *PYC1* were transcriptionally up-regulated (*PCK1* 10.5 fold change and *PYC1* 2.5 fold change, elevated CO<sub>2</sub>:reference). Enzymatic activity of the gene product, Pck1p, was correspondingly and strongly elevated, while Pyc1p was only slightly elevated.

In the same study, Table 1 of the original publication included for reference in Figure 24, suggested that elevated CO<sub>2</sub> concentrations lead to increased succinic acid production under both carbon and nitrogen limitations.

Table 1 Physiological effects of elevated CO <sub>2</sub> concentrations on glucose-grown chemostat cultures (pH 5, 30 °C, D = 0.10 h <sup>-1</sup> ) of <i>S. cerevisiae</i> CEN.PK113-7D										
CO <sub>2</sub> inlet gas (%)	Dissolved CO <sub>2</sub> <sup>a</sup>	Yield <sup>b</sup>	qO <sub>2</sub> <sup>c</sup>	qGlucose <sup>c</sup>	qEthanol <sup>c</sup>	qAcetate <sup>b</sup>	qPyruvate <sup>b</sup>	qGlycerol <sup>b</sup>	qSuccinate <sup>c</sup>	
Anaerobic	<0.01	0.46 ± 0.01	0.096 ± 0.002	—	6.52 ± 0.28	9.95 ± 0.37	0.02 ± 0.00	0.01 ± 0.00	0.87 ± 0.04	0.03 ± 0.00
Anaerobic	100	27.39 ± 0.04	0.091 ± 0.004	—	6.51 ± 0.67	10.39 ± 1.13	0.02 ± 0.00	0.02 ± 0.00	0.68 ± 0.08	0.03 ± 0.00
Aerobic, N-limited	0.05	0.46 ± 0.01	0.095 ± 0.002	2.70 ± 0.10	5.82 ± 0.14	7.99 ± 0.13	0.06 ± 0.01	0.10 ± 0.01	0.08 ± 0.01	0.04 ± 0.02
Aerobic, N-limited	79	21.70 ± 0.19	0.085 ± 0.001	2.56 ± 0.47	6.47 ± 0.24	9.78 ± 0.26	0.16 ± 0.01	0.22 ± 0.01	0.10 ± 0.01	0.18 ± 0.01
Aerobic, C-limited	0.05	0.22 ± 0.00	0.504 ± 0.005	2.71 ± 0.39	1.10 ± 0.07	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Aerobic, C-limited	79	22.30 ± 0.25	0.382 ± 0.045	4.73 ± 0.26	1.41 ± 0.18	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.13 ± 0.03

Results are the average ± standard deviation of three independent cultivations.  
<sup>a</sup> Calculated value according to Henry's law (mM).  
<sup>b</sup> Yield in biomass (g biomass (g consumed glucose)<sup>-1</sup>).  
<sup>c</sup> Expressed in mmol (g of biomass)<sup>-1</sup> h<sup>-1</sup>.

**Figure 24: Table 1 from Aguilera et al, 2005, Showing Physiological Effects of Elevated CO<sub>2</sub> in Fermentation.** In the above Table the authors indicate elevated succinate under aerobic, nitrogen-limited and aerobic, carbon limited conditions with elevated CO<sub>2</sub> concentrations

In the original study by Aguielera et al, 2005, there is no discussion of the increased succinic acid production. Therefore, it was decided to perform a physiological characterization of *S. cerevisiae* strains 8D Evolved and the reference, CENPK113-5D, in well-controlled, aerobic, batch glucose fermentations sparged with different CO<sub>2</sub> gas concentrations. Specifically, supplementations of gas at 0, 10, 20, 40, and 80% v v<sup>-1</sup> CO<sub>2</sub> for both strains were examined. For the reference strain, two additional concentrations were examined: 5% and 100% v v<sup>-1</sup> CO<sub>2</sub>. A detailed methodology for strain construction and physiological characterization is provided in **Paper I**.

A brief description of the physical chemistry and thermodynamic properties of CO<sub>2,g</sub> supplemented to a liquid phase is described by the equilibrium chemistry below.

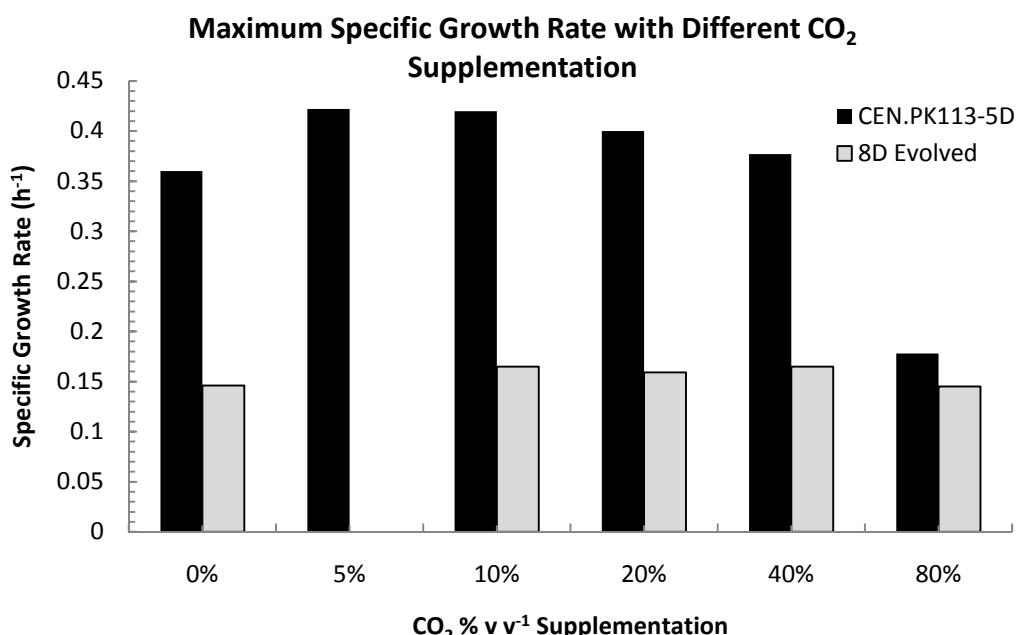


The concentration of dissolved CO<sub>2</sub> can be related to the external partial pressure of carbon dioxide by Equation 35:

$$[\text{CO}_2]_{\text{dissolved}} = H p_{\text{CO}_2}$$

Where  $p_{CO_2}$  is the partial pressure (atm) of dissolved  $CO_2$ , and  $H$  is Henry's law constant. For this investigation Henry's Law described above was used in order to calculate the concentration of  $CO_2$  in the fermenter. Given the relatively low solubility of  $CO_2$  in water ( $1.45 \text{ kg m}^{-3}$ ), it is not surprising using a Henry's Law estimation that the dissolved concentration of  $CO_2$  ranged from  $0.003 \text{ mol l}^{-1}$  ( $5\% \text{ v v}^{-1} CO_2$ ) to  $0.04 \text{ mol l}^{-1}$  ( $100\% \text{ v v}^{-1} CO_2$ ). The methodology and appropriate references for this estimation are provided in the thesis, *Evaluation of the Effect of High  $CO_2$  Concentration on Succinic Acid Production in *Saccharomyces cerevisiae** (Nestola, 2007). It should also be noted that utilization of sodium bicarbonate as a source of base for pH control would have been a desirable strategy to increase  $HCO_3^-$  availability for carboxylation reactions. The supplementation of  $CO_{2g}$  via sparging was specifically attempted because of the prior observations of Aguijera et al, 2005. If this research were to proceed then increased  $[CO_2]$  would be tested via supplementation of sodium bicarbonate to by-pass the poor solubility of  $CO_{2g}$  in water.

There were two significant results from these series of experiments. First, there was an enhancement in the specific growth rate in both 8D Evolved and CEN.PK113-5D when  $CO_2$  was supplemented (See Figure 25 and Table 3). Second, there was no significant increase in succinate production resulting from  $CO_2$  supplementation for CEN.PK113-5D (See Figure 26). Succinate production actually decreased from  $CO_2$  supplementation for 8D Evolved (See Figure 27).



**Figure 25: Maximum Specific Growth with Different  $CO_2$  Supplementation.** The maximum specific growth rate ( $\text{h}^{-1}$ ) for *S. cerevisiae* CEN.PK113-5D and 8D Evolved (Paper I) are reported for different  $CO_2$  supplementation values ( $\% \text{ v v}^{-1}$ ) in batch, aerobic, glucose-supplemented fermentations.

As indicated in Table 3, there was a maximum 17.2% and 13.0% improvement in the maximum specific growth rate of CEN.PK113-5D and 8D Evolved, respectively, when supplemented with  $5\% \text{ v v}^{-1} CO_2$  and  $10\%$  or  $40\% \text{ v v}^{-1} CO_2$ , respectively. In both strains, inhibition of growth was observed at  $80\% \text{ v v}^{-1} CO_2$ , although more pronounced in CEN.PK113-5D compared to 8D Evolved. In CEN.PK113-5D, it is interesting to note that there was a shift in carbon distribution from ethanol and glycerol to primarily biomass, acetate, and pyruvate. The

yield of biomass and acetate on glucose increased across the entire range of CO<sub>2</sub> supplementation. In the 8D Evolved strain, a very different shift in carbon distribution was observed, where across the entire range of CO<sub>2</sub> supplementation, yield of biomass and succinate decreased, while ethanol, acetate, and glycerol (40% v v<sup>-1</sup> CO<sub>2</sub>, exception) increased. Given the coupling of succinate production to biomass formation in 8D Evolved (**Paper I**), it is not surprising that both yields decreased. The genotype difference between 8D Evolved and CEN.PK113-5D, is the deletion of *SDH3*, *SER3*, and *SER33*. Given the changes in carbon distribution largely are concentrated in lower glycolysis and pyruvate metabolism, it is reasonable to speculate that increased concentrations of CO<sub>2</sub> in the presence of a functional TCA cycle (CEN.PK113-5D) lead to increased levels of biomass and acetate, while in the 8D Evolved strain with TCA cycle disruption, decreased carbon to biomass formation resulted in increased overflow metabolism (ethanol, acetate, glycerol). It should be noted that all experiments were executed only once, and that off-gas estimations of CO<sub>2</sub> and O<sub>2</sub> were not possible due to the over-saturation of the gas analyzer with the inlet feed stream that varied between (0-80% v v<sup>-1</sup> CO<sub>2</sub>). This prevented carbon recovery calculations from being performed. Furthermore, it should be noted that growth of CEN.PK113-5D at 100% v v<sup>-1</sup> CO<sub>2</sub> supplementation was completely inhibited.

**Table 3: Relative Effect on the Maximum Specific Growth Rate for Different CO<sub>2</sub> Supplementation Concentrations Compared to No Supplementation**

CO <sub>2</sub> Supplementation	CEN.PK113-5D	8D Evolved
[v v <sup>-1</sup> ]	[% ref $\mu_{max}$ ] <sup>1</sup>	[% ref $\mu_{max}$ ] <sup>1</sup>
5%	17.2	NA
10%	16.7	13.0
20%	11.1	8.9
40%	4.7	13.0
80%	-50.6	-0.7

**Notes:** 1. The relative increase or decrease, in percent, is relative to the maximum specific growth rate measured with no CO<sub>2</sub> supplementation.

In an effort to further understand and propose a potential mechanism as to why relatively small concentrations of CO<sub>2</sub> supplementation (<0.04 mol l<sup>-1</sup>) had significant effects on carbon metabolism, enzyme thermodynamics were considered. Specifically, the enzymes aconitase, succinyl-CoA ligase, isocitrate lyase, malate synthase, isocitrate dehydrogenase, alanine-glyoxylate aminotransferase, pyruvate carboxylase, PEP carboxylase, and threonine adolase were considered.

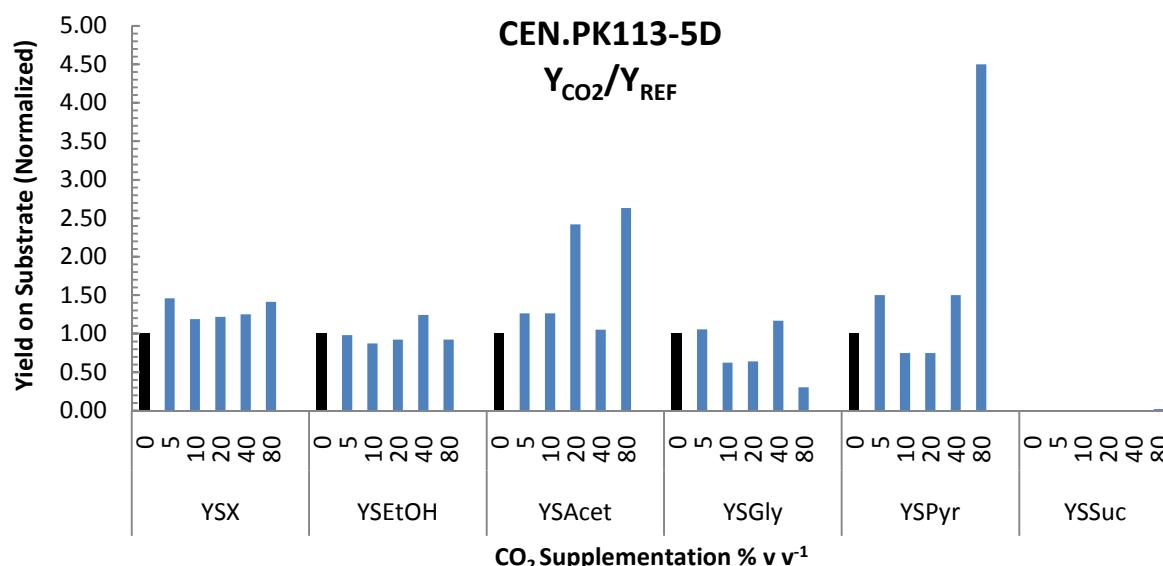
The Gibbs free energy is defined by Equation 36.

$$\Delta G = \Delta G^0 + RT \ln K_{eq} = \Delta G^0 + RT \ln \frac{[Products]}{[Reactants]}$$

Equation 36

Where  $K_{eq}$  is the equilibrium rate of reaction, and may be defined as the ratio of the concentration of the reaction products, [Products], divided by the concentration of reaction reactants, [Reactants]. Furthermore, R is the ideal gas constant and T is the temperature at which the reaction is considered. The challenge with biological systems is ascertaining physiologically relevant estimations of the standard Gibbs free energy of formation,  $\Delta G^0$ . The US National

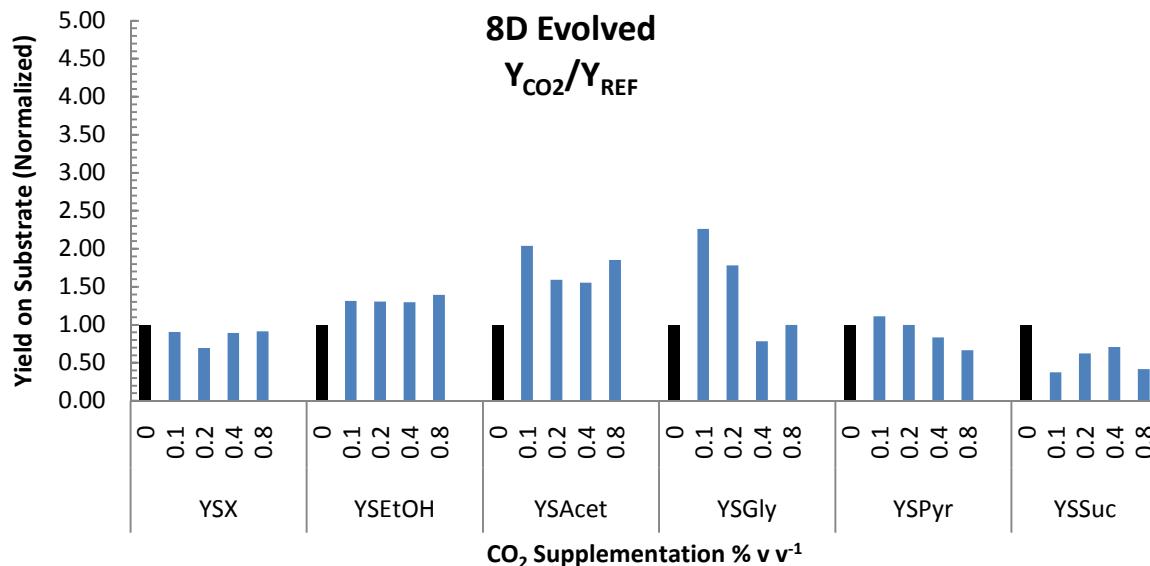
Institute of Standards and Technology supports the Standard Reference Database 74, which contains values related to thermodynamics of enzyme-catalyzed reactions (Goldberg et al, 2004). From this database, for each of the enzymes listed, the most physiologically relevant  $K_{eq}$  values were selected (pH 6.5-7.5, 20-30°C) and used to calculate  $\Delta G^\circ$ , assuming the reaction was at equilibrium ( $\Delta G=0$ ). Once a  $\Delta G^\circ$  was estimated, then a temperature of 30°C and range of  $K_{eq}$  values, defined as [Products]/[Reactants], was considered to estimate the  $\Delta G$ . This simple calculation provides a thermodynamic estimation of the spontaneity of the reaction, where  $\Delta G<0$  is thermodynamically favorable and the reaction proceeds forward spontaneously, while  $\Delta G>0$  is not thermodynamically favorable and the reaction will not proceed forward spontaneously (the reversible reaction is favored).



**Figure 26: Relative Yields on Substrate of CEN.PK113-5D Supplemented with CO<sub>2</sub>.** The yields of biomass (Y<sub>SX</sub>), ethanol (Y<sub>SEtOH</sub>), acetate (Y<sub>SAcet</sub>), glycerol (Y<sub>SGly</sub>), pyruvate (Y<sub>SPyr</sub>), and succinate (Y<sub>SSuc</sub>) on glucose in CEN.PK113-5D fermentations supplemented with CO<sub>2</sub> (% v v<sup>-1</sup>), are reported. The yields are normalized to the reference condition where 0% v v<sup>-1</sup> CO<sub>2</sub> was supplemented (black bar). The Y<sub>SSuc</sub> was not normalized as there is no production of succinate in CEN.PK113-5D under reference conditions. There was a negligible amount of succinate produced at 80% v v<sup>-1</sup>, hence the small bar indicated.

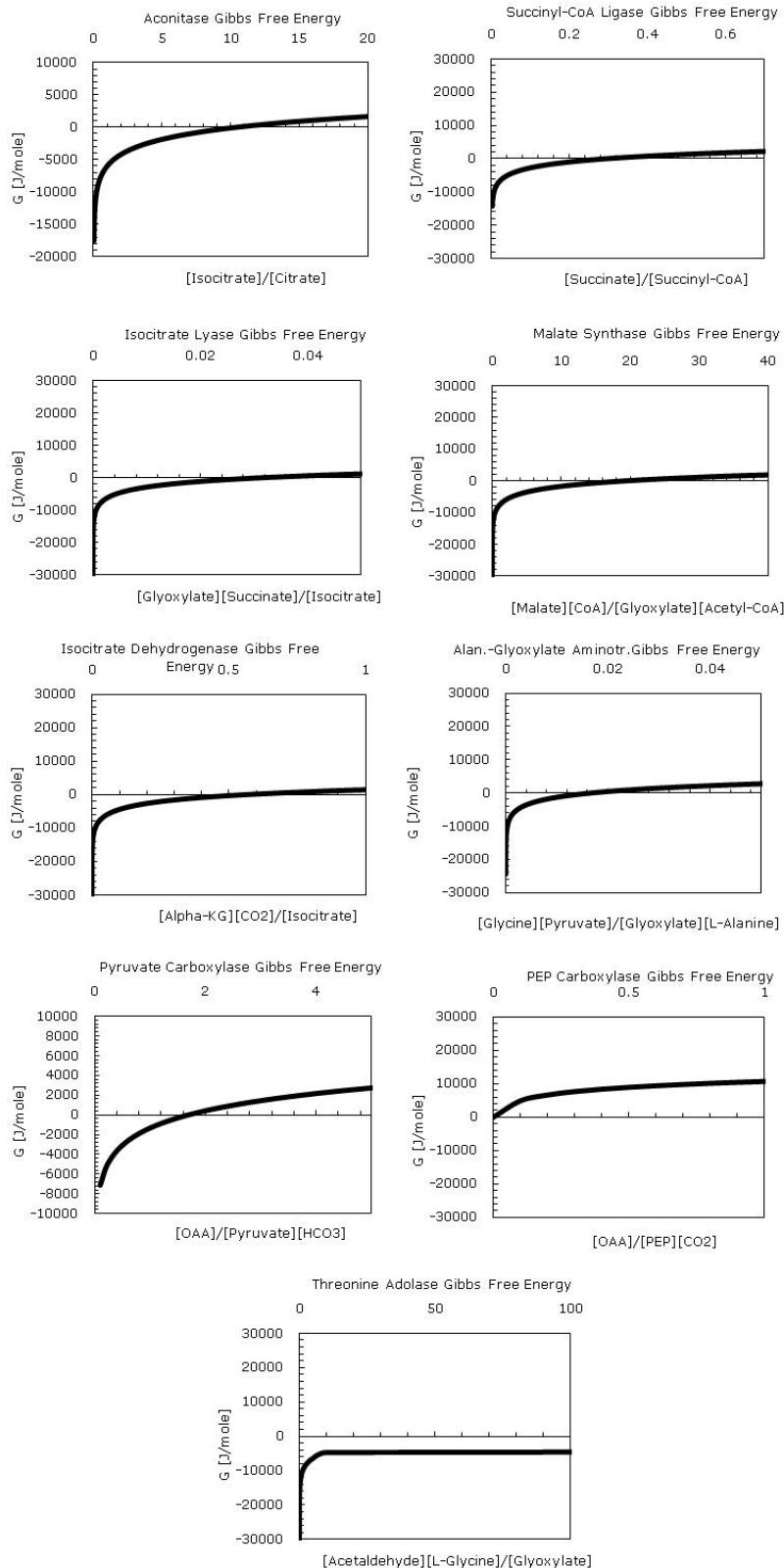
Figure 28 provides a summary of the  $\Delta G$  estimations for various TCA cycle, glyoxylate cycle, and glycolytic enzymes relevant to succinate metabolism. For purposes of understanding potential mechanisms for why CO<sub>2</sub> supplementation could affect central carbon metabolism, it is interesting to consider the reactions where CO<sub>2</sub> is either a product or reactant (Note: only considered for enzymes where physiologically relevant thermodynamic data were available). Those enzymes include isocitrate dehydrogenase (CO<sub>2</sub> is a product), pyruvate carboxylase (CO<sub>2</sub> is a reactant), and PEP carboxylase (CO<sub>2</sub> is a reactant; and this enzyme is not present in *S. cerevisiae*). Both isocitrate dehydrogenase and pyruvate carboxylase are thermodynamically favorable at [Products]/[Reactants] ratios of 0.53 and 1.7, respectively. Therefore, if CO<sub>2</sub> is supplemented externally, then isocitrate dehydrogenase becomes less favorable, and pyruvate carboxylase becomes more favorable. There is some suggestion that isocitrate dehydrogenase will be more sensitive to CO<sub>2</sub> supplementation based on the  $\Delta G$  response curve (See Figure 28). In 8D Evolved, isocitrate dehydrogenase activity is required for subsequent formation of succinate, glyoxylate, and serine. Therefore, a plausible hypothesis for why the biomass yield on glucose decreased with CO<sub>2</sub> supplementation specifically in the 8D Evolved strain is that isocitrate

dehydrogenase activity was less active. In CEN.PK113-5D, the same would be true except that TCA cycle activity is not disrupted, therefore, formation of oxaloacetate is favored, and further favored under CO<sub>2</sub> supplementation. CO<sub>2</sub> supplementation could drive increased flux to oxaloacetate, pyruvate and subsequently acetate. Further experimental exploration and validation would be required to determine if the hypotheses proposed are valid.



**Figure 27: Relative Yields on Substrate of 8D Evolved Supplemented with CO<sub>2</sub>.** The yields of biomass (YS<sub>X</sub>), ethanol (Y<sub>SEtOH</sub>), acetate (Y<sub>SAcet</sub>), glycerol (Y<sub>SGly</sub>), pyruvate (Y<sub>SPyr</sub>), and succinate (Y<sub>SSuc</sub>) on glucose in 8D Evolved fermentations supplemented with CO<sub>2</sub> (%v v<sup>-1</sup>), are reported. The yields are normalized to the reference condition where 0% v v<sup>-1</sup> CO<sub>2</sub> was supplemented (black bar).

Interestingly, PEP carboxylase is thermodynamically favorable at a [Products]/[Reactants] ratio of 0.01 (See Figure 28). This calculation confirms that PEP carboxylase is thermodynamically favorable in the reverse direction (e.g., PEP carboxykinase) and the widespread reporting in bacterial hosts that high PEP carboxylase activity is observed is thermodynamically unlikely (Bazaes et al, 2007; Lu et al, 2009; Kim et al, 2007). This analysis further suggests that perhaps an increase in [CO<sub>2</sub>] could increase thermodynamic favorability of PEP carboxylase; however, the reaction is still favored towards the formation of PEP (PEP carboxykinase) to support gluconeogenesis. As stated earlier, *S. cerevisiae* does not contain PEP carboxylase. A final interesting observation of the thermodynamic analysis performed is that under no [Product]/[Reactant] ratios is threonine adolase favorable in the direction of glycine formation from glyoxylate. The directionality of threonine adolase, encoded by *GLY1*, is discussed in **Paper I** and the experimental conclusions reached are verified by this thermodynamic evaluation. This is particularly relevant when considering future metabolic engineering strategies that aim to draw on glyoxylate formation, and consequently, succinate formation, for redirection from glycine to serine or threonine.



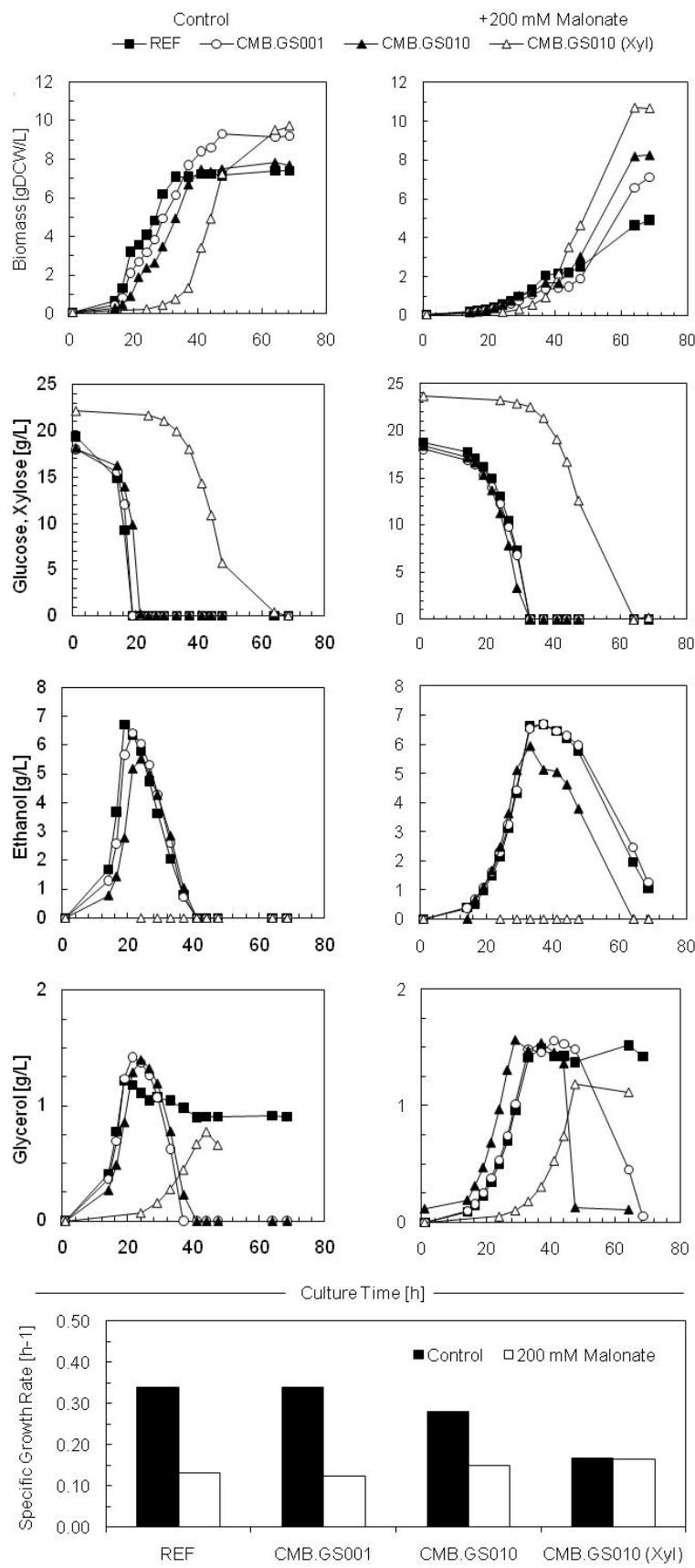
**Figure 28: Thermodynamic Estimation of  $\Delta G$  ( $J \text{ mol}^{-1}$ ) of Succinate Specific Enzymes.** Each individual plot is the Gibbs free energy,  $\Delta G$ , estimation of the metabolic enzyme indicated. The lower  $x$ -axis indicates the ratio of the enzyme catalyzed reaction  $[Products]/[Reactants]$ . The upper  $x$ -axis provides the relative ratio of  $[Products]/[Reactants]$  over which the simulation was considered.

### 7.5.3 Malonate Supplementation to *S. cerevisiae* CMB.GS010

**Paper III** describes the construction of strain CMB.GS010 for the fast and efficient utilization of xylose under aerobic conditions with minimal overflow metabolism to ethanol, glycerol, acetate, and xylitol. CMB.GS010 exhibited a strongly respiratory response. Therefore, if this strain should serve as an ideal platform for biomass coupled succinate formation, similar to 8D Evolved (**Paper I**), from xylose, it was required to determine if deletion of *SDH3* was feasible. Malonate is a well-established competitive chemical inhibitor of the succinate dehydrogenase complex (Sdh<sub>p</sub>) that competes with succinate and fumarate for the Sdh<sub>p</sub> dicarboxylate binding site (Oyedotun et al, 1999). A sterile solution of malonic acid was added after medium autoclavage at a final concentration of 200 mM. A preliminary screen revealed maximum Sdh<sub>p</sub> inhibition, as indicated by the ethanol consumption rate, with 200 mM malonate supplementation. The final pH of the medium was adjusted to 6.5 with 2M NaOH.

The evolved strain CMB.GS010, its parental strain CMB.GS001, and the reference strain CEN.PK.113-3C were investigated under aerobic shake flask conditions in the presence of 200 mM malonate in minimal medium supplemented with 20 g l<sup>-1</sup> glucose or 20 g l<sup>-1</sup> xylose (only for CMB.GS010). Cultivated on glucose all three strains exhibited similar growth profiles (See Figure 29). After complete glucose depletion, the diauxic shift was observed, and ethanol was respiration together with the re-assimilation of acetate and glycerol to produce biomass. When malonate was supplemented to the medium all three strains exhibited an initial exponential growth phase with glucose converted predominantly to ethanol at a reduced glucose consumption rate. Once glucose was exhausted, inhibition of ethanol respiration was detected, and growth was arrested for a minimum of 20 hours. After this lag-phase cells restarted assimilation of ethanol at a reduced rate. This is likely resulting from a de-inhibition of Sdh<sub>p</sub>, suggesting that malonate is either metabolized or degraded. Yet, when the evolved strain CMB.GS010 was cultured on xylose in the presence of malonate no significant difference in xylose consumption was detected. Xylose was completely consumed and no further significant effect on product formation was detected as compared to the control condition (See Figure 29). Under all conditions with malonate supplementation no succinic acid accumulation was detected.

The glyoxylate pathway was significantly up-regulated in the evolved strain cultivated on xylose compared to the evolved strain cultivated on glucose or the unevolved strain cultivated on glucose (**Paper III**). This pathway had a significantly higher log-fold change than succinate dehydrogenase and succinyl-CoA ligase, suggesting that the lower respiratory arm of the TCA cycle was being by-passed by the glyoxylate cycle. This is consistent with the physiological observation that malonate supplementation appeared to have no effect on the evolved strain cultivated on xylose compared to cultivation on glucose where respiration inhibition was observed (See Figure 29). Therefore, future metabolic engineering of CMB.GS010 for succinate formation by interruption of succinate dehydrogenase is likely favorable.



**Figure 29: Physiological Effects of Malonate Supplementation.** The left panels are aerobic shake flask cultivations on 20 g l<sup>-1</sup> glucose for the reference strain (CEN.PC 113-3C), CMB.GS001 (unevolved), CMB.GS010 (evolved), and 20 g l<sup>-1</sup> xylose only for CMB.GS010. The right panels are aerobic shake flask cultivations, identical to the left panels, supplemented with 200 mM malonate. Concentrations of glucose, xylose, biomass, ethanol, and glycerol are reported. The bottom bar graph panel shows the specific growth rate for each strain and condition.

## 8.0 Conclusions

Throughout the course of my PhD research industrial systems biology and metabolic engineering of *S. cerevisiae*, with particular attention to designing a microbial cell factory for overproduction of succinic acid, was utilized. While the strains described, specifically in **Papers I** and **II** require significantly more development prior to consideration for commercialization, the proof-of-concept has been established. Specifically, the genome-scale metabolic network reconstruction was used to guide mutant construction and understand the limitations of the native *S. cerevisiae* metabolic network (**Papers I** and **II**). Targeted genetic engineering was supplemented with directed evolution (**Papers I** and **III**), where a clear selective pressure was used for rapid screening and selection. Careful fermentation physiological characterization married to systems biology tools, including transcriptome analysis (**Papers I, II, III**) not only provided insight into why specific phenotypes were observed, but more importantly, enabled second-round metabolic engineering for improved succinic acid production (**Paper I**). Interestingly, **Papers I, II, and III** confirmed that while increased succinic acid production and xylose utilization was observed, these phenotypes could not be fully explained with stoichiometric modeling. This ignited the idea that high-throughput genome sequencing for SNP detection, specifically for metabolic engineering applications, where a genetically undefined mutant is compared to the reference starting strain, may help to elucidate direct genotype to phenotype relationships. **Paper IV** suggested that major phenotypic differences between two *S. cerevisiae* strains did exhibit SNPs that were suggestively correlated. **Paper IV** was the riskiest of the research strategies employed; however, this is often required when attempting to introduce a novel tool. High-throughput whole genome sequencing will become a routine systems biology tool in metabolic engineering, but still requires significant additional work to demonstrate *in silico* predictions and correlations *in vivo*. In short, the combination of **Papers I, II, III, and IV** robustly demonstrate proof-of-concept of the industrial systems biology supplemented metabolic engineering cycle first proposed in Figure 3.

More specifically, **Paper I**, demonstrated that chemical inhibition or targeted genetic engineering of the primary succinate consuming reactions, as well as over-expression of the primary succinate producing pathways, resulted in no succinate accumulation. A multi-gene, non-intuitive, genetic engineering strategy coupling biomass formation through glycine-serine amino acid requirements to succinate production resulted in a proof-of-concept microbial cell factory, 8D Evolved with pICL1. This strain exhibited a 30-fold improvement in succinate titer, 43-fold improvement in succinate yield on biomass, with only a 2.8-fold decrease in the specific growth rate compared to the reference strain. The succinate yield on biomass in the 8D and 8D Evolved strains was 8-fold higher than required to meet biomass amino acid demand, suggesting that regulatory mechanisms not captured in genome-scale metabolic network reconstruction are likely playing a significant role in the succinate production observed.

**Paper II** built-upon the experience from **Paper I**, and demonstrated that genome-scale metabolic network reconstructions, using pathway visualization and flux balance analysis, to predict succinic acid overproduction strategies assuming batch glucose conditions were feasible after model tuning with experimental data. A simple pragmatic approach of introducing an artificial conversion of  $\text{NAD}^+ \rightarrow \text{NADH}$ , and then constraining this reaction to a flux such that the glycerol production is correctly described by the model resulted in a better overall fit of the major carbon fluxes. Identification of the top single and double gene deletion strategies, under aerobic and anaerobic conditions, resulted in three predictions with a 10-fold improvement in succinate yield on glucose compared to the reference: *MDH1*, *OAC1*, and *DIC1*. Model validation was performed using knock-out strains cultivated anaerobically on glucose, coupled with physiological and genome-wide DNA microarray characterization. While  $\Delta mdh1$  and  $\Delta oac1$  strains failed to produce more succinate relative to the reference,  $\Delta dic1$  produced 0.02 C-mol C-

mol-glucose<sup>-1</sup>, in close agreement with model predictions (0.03 C-mol C-mol-glucose<sup>-1</sup>). Pathway visualization, coupled with transcriptional profiling, suggests that succinate formation is coupled to mitochondrial redox balancing, and more specifically, reductive TCA cycle activity.

All of the strains considered in **Papers I** and **II** were designed based on glucose supplementation and metabolism. It is clear that biorefinery compatible industrial biotechnology processes, including those considered for succinic acid production, will have to use diverse feedstocks including lignocellulose. Certainly, any future *S. cerevisiae* microbial cell factory capable of fast and efficient xylose consumption, and succinic acid overproduction would be highly desirable. In **Paper III** metabolic engineering of *S. cerevisiae* for consumption of xylose aerobically without redirection of some carbon flux to overflow metabolites (ethanol, glycerol, acetate, xylitol) was accomplished by expression of *PsXYL1*, *PsXYL2*, and *PsXYL3* from the native xylose-metabolizing *Pichia stipitis*, and subsequent, directed evolution. The resulting *S. cerevisiae* strain showed xylose consumption at a specific rate of 0.31 g g-cell<sup>-1</sup> h<sup>-1</sup>, a specific growth rate of 0.18 h<sup>-1</sup>, and a biomass yield of 0.62 C-mol C-mol-xylose<sup>-1</sup>. Plasmid isolation and re-transformation confirmed the conferred phenotype resulted from a chromosomal modification. Transcriptional profiling confirmed a strongly up-regulated glyoxylate pathway enabling sustained respiratory metabolism. Chemical inhibition of succinate dehydrogenase confirmed that this xylose consuming strain is a suitable candidate for further metabolic engineering of biomass coupled succinate production.

**Papers I** and **III** demonstrated the success of coupling systems biology approaches with well established methods of directed evolution. **Paper IV** was an attempt to use a relatively new technology and apply it to metabolic engineering for guiding and elucidating direct genotype to phenotype relationships. Whole high-throughput genome sequencing of *S. cerevisiae* S288C and CEN.PK113-7D resulted in identification of 13,787 filtered SNPs in CEN.PK113-7D, with a total of 939 SNPs detected across 158 unique metabolic genes, 85 of which contained a total of 219 non-silent SNPs. There were two central carbon metabolic pathways enriched with non-silent SNPs that also correlated with significant differences in phenotype. *S. cerevisiae* CEN.PK113-7D exhibited significantly higher ergosterol content during growth on glucose, and to a lesser extent, galactose. The ergosterol biosynthetic pathway had significant non-silent SNPs identified in *ERG8* and *ERG9*, and silent SNPs identified in *ERG20* and *HMG1*. The flux through the galactose uptake pathway was much lower in S288C compared with CEN.PK113-7D, correlating with the non-silent SNP enrichment in *GAL1* and *GAL10*, and silent SNPs in *GAL7*. More globally, the physiological characterization clearly suggests that S288C has a deficiency in metabolism of respiratory carbon sources, such as ethanol and galactose, when compared to CEN.PK113-7D. Inspection of the significantly differentially expressed genes between strains cultivated on glucose or galactose did not reveal an obvious gene cluster that would explain this significant physiological difference. Therefore, strongly suggestive that genotype to phenotype correlation is manifested post-transcriptionally or post-translationally either through protein concentration and/or function.

**Papers I, II, III, and IV** demonstrate that metabolic engineering supplemented with systems biology approaches, particularly transcriptomics, enabled rapid construction of microbial cell factories for succinic acid production and xylose utilization, and novel insight into central carbon metabolism flux distributions and regulation, particularly related to respiratory and TCA cycle metabolism. It is most important to realize that there is no single technology, methodology, or approach that will ensure rapid and successful microbial cell factory construction for chemical compounds. Rather, stoichiometric modeling constantly supplemented with *x*-omics characterization and traditional physiological characterization can help to reduce the design space for metabolic engineers. This enables focus on high probability of success strategies, and most importantly, a clear understanding of the limitations of the production host. It is clear, based on the conclusions described here that while *S. cerevisiae* can be engineered to overproduce succinic

acid, to reach the yields, titers, and productivities that would be required prior to commercialization, heterologous pathway expression will be required. This thesis serves as the required foundation upon which such a metabolic engineering strategy can be implemented.

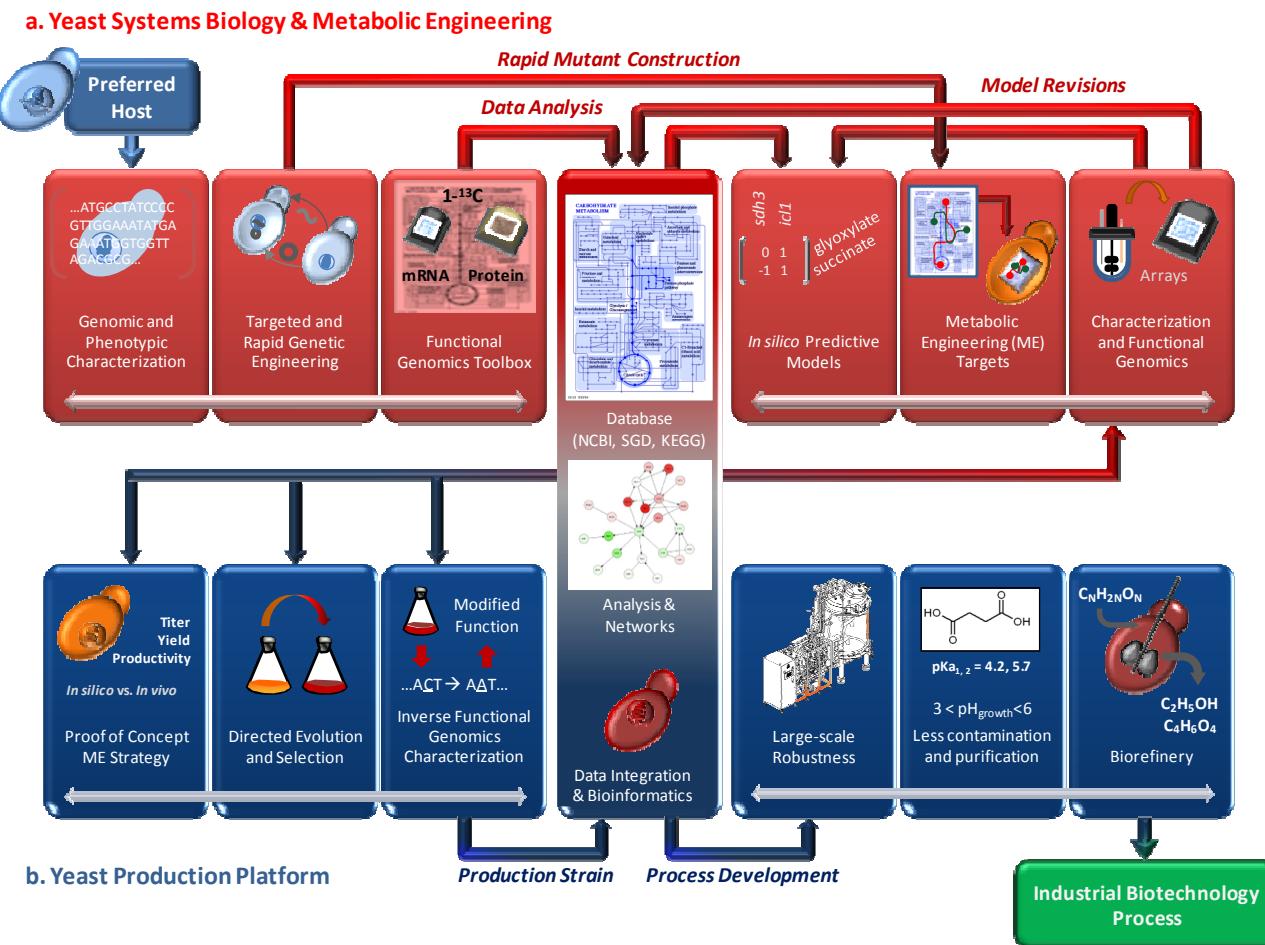
## 9.0 Perspectives on Future Developments in Industrial Systems Biology

**Paper V** provides a historical review and perspective on a new and emerging field, referred to here as *industrial systems biology*. Applying a mathematical framework to microbial metabolism, beginning in earnest as early as the 1930s, has provided a scaffold for large data sets, most recently associated with the emerging field of systems biology (transcriptomics, proteomics, fluxomics, metabolomics), to be integrated, interrogated, analyzed, and ultimately, reformulated into predictive models referred to as genome-scale metabolic reconstructed networks. These networks have offered metabolic engineers, in conjunction with accessible and easily applied recombinant DNA technology, the ability to define clear and high probability of success genetic targets for redirection of carbon flux from renewable, sustainable, and cost-effective substrates to high added-value and commodity chemical production. The construction of microbial cell factories to meet industrial biotechnology process development needs, previously relegated to classical methods of directed evolution, screening, selection, isolation, and propagation, are now being constructed faster and more efficiently through the use of systems biology toolboxes. *Industrial systems biology*, that includes the specific application of genome-scale technologies, both experimental and *in silico*, to industrial biotechnology process development. The impact of industrial systems biology is apparent over a broad cross-section of products, including succinic acid. Figure 30 provides a conceptual pipeline for industrial systems biology guided process development with proof of concept demonstrated in the development of the first over-producing succinic acid *S. cerevisiae* strain.

Although a large number of genome-scale metabolic network reconstructions are available, what is interesting to observe is the relatively poor coverage of microbial metabolism that these reconstructions offer. A close inspection of Table 2 in **Paper V** reveals that combined, all of the metabolic reconstructed networks have an average genome coverage of  $14.6 \pm 8.1\%$  ( $n=29$ ). If *S. cerevisiae*, the most well characterized eukaryote, is isolated as an example, the most recent metabolic reconstructed network has genome coverage of 13.6%, while 4691 of the 6608 total ORFs, 70.9%, have a verified function (Fisk et al, 2006; Hong et al, 2008). From a more general perspective, the problem of metabolic gap closing is exacerbated by the relatively large orphan metabolic activities, where 30-40% of the known metabolic activities that are classified by the Enzyme Commission have no associated genomic sequences in any organism (Breitling et al, 2008; Green et al, 2005; Lespinet et al, 2006). There is currently significant effort under-way to extend pathway reconstructions to regions of metabolism that are poorly understood or to a large degree, have been functionally neglected (Viswanathan et al, 2008; Breitling et al, 2008). Industrial biotechnology has largely focused on the production of added value and commodity chemicals; however, the largest expected growth sector is in the area of specialty and fine chemicals, where industrial biotechnology offers simpler routes for complex synthetic chemistry, or the possibility of *de novo* chemicals that may offer similar or enhanced application (Hirche, 2006; Gavrilescu et al, 2005). Specialty and fine chemical entities are typically present as metabolic intermediates in secondary and tertiary regions of metabolism, often poorly annotated, and rarely included in genome-scale network reconstructions. A clear example is lipid metabolism in *S. cerevisiae*, where a recent update to the existing genome-scale metabolic reconstruction, iN795, included 118 previously unreported lipid reactions relative to iND750 (See Table 3 of **Paper V**). Of those 118 lipid metabolism participating reactions, 28 were assigned to ergosterol esterification and lipid degradation – previously not represented (Nookaew et al, 2008).

Focusing specifically on continued metabolic engineering of *S. cerevisiae*, both **Paper I** and **II** confirm that the native portfolio of central carbon metabolism pathways is not likely to enable a commercially viable succinic acid microbial cell factory. Rather, introduction of non-native pathways, particularly those that could enable high rates of carboxylation of PEP and pyruvate for reductive TCA cycle succinate generation will be required. The use of genome-scale

metabolic models enabled rapid target identification and experimental verification of the potential of *S. cerevisiae* as a microbial cell factory for succinic acid, and would likely not have been possible using intuition. The potential industrial benefits of *S. cerevisiae* for organic acid production warrant further efforts using expression of heterologous pathways.



**Figure 30: Industrial Systems Biology Pipeline.** Panel A shows the yeast systems biology and metabolic engineering platform, which involves application of phenotypic characterization, targeted genetic engineering, and functional genomics for construction of mutants derived from a desirable industrial host. Functional genomics, coupled with manually curated databases and bioinformatics, permits genome-wide annotation that may be structured and mathematically framed into a genome-scale metabolic model (GSMM). GSMMs provide a scaffold for integrating genomics data, and can simulate operation of the metabolic network. The ability to scan a genome-wide solution space for targeted multi-gene modifications is required for rapid identification of metabolic engineering strategies. Following mutant construction, extensive high-throughput technologies such as transcriptomics, proteomics, fluxomics, and metabolomics can be used for mutant characterization. Panel B describes the yeast production platform, where the aforementioned data set enables proof of concept of predicted metabolic engineering strategies, and drives further strain improvement through directed evolution, selection, or other approaches that target specific genetic changes. Once an improved strain is isolated genomic tools may be used to inversely characterize non-specific modifications, leading to an updated and improved GSMM. The resulting production strain is then carried onto industrial process development where large-scale robustness and performance analysis, as well as integration with downstream processing, occur. At the core of integrating systems biology with process development is bioinformatics and network analysis. Industrial biotechnology processes that are cost-effective, sustainable, environmentally favorable, and reach a wide-range of chemical sectors will use a biorefinery model that can leverage platform technologies offered by using preferred industrial host organisms, such as yeast.

As with any mathematical framework that incorporates large collections of diverse biological data that are constantly being investigated, updated, re-annotated, re-analyzed, and debated, clear modeling objectives must be set forth. From an industrial biotechnology perspective, focused on identifying high yielding, robust, and easy to implement non-intuitive metabolic engineering strategies, microbial metabolic modeling must continue to expand upon constraint-based stoichiometric flux balance analysis that incorporates experimental verification, and subsequent model updating and expansion. Perhaps the emerging availability of kinetic parameters will enable fully dynamic metabolic reconstructions to be realized in the future, but for now, the full benefits of stoichiometric metabolic modeling have yet to be realized in constructing next generation microbial cell factories. Industrial systems biology is a new approach to a challenge of epic proportions: how do we develop processes for production of chemicals, materials, and energy that are cost-effective, renewable, sustainable, scalable, and environmentally-favorable?

## 10. Acknowledgements

The culmination of a doctoral thesis represents many years of study, learning, uncertainty, sacrifice, risk-taking, failure, success, and in the case for which the laboratory is not in their native country, a deep separation from loved friends and family. My doctorate was no different, and ultimately was possible because of two key ingredients: passion and people. I was determined to join a world-class, *high-risk high-reward*, multi-disciplinary laboratory where I was the most ignorant and naïve of my colleagues, enabling me to learn as much as possible, and push me to my limits. It is with tremendous pleasure I can say, this was what occurred, and I have no regrets.

I would first like to thank my supervisor, Professor dr. techn. Jens Nielsen for affording me the opportunity to be a Merck Doctoral Fellow at the Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark (September 2005 - June 2008), and the Systems Biology Group, Department of Chemical and Biological Engineering, Chalmers University of Technology (June 2008 – December 2009). Professor Nielsen represents a phenomenal scientific force that is analytical, fast, intense, and most importantly, both positive and passionate. Jens, I will never forget the openness, education, and guidance you provided throughout my experience, in both professional and personal endeavors. You are inspirational and as I have often witnessed and commented your unwavering support of your students no matter what the conditions makes you truly unique amongst a profession I fear may have lost this attribute. Jens, you have become a friend and mentor, and I hope this small *collaboration* is the first of many, both professional and personal.

I would like to thank my co-supervisor Professor Lisbeth Olsson for inviting me into her laboratory at both the Technical University of Denmark and later Chalmers University of Technology. As a doctoral student with perhaps a non-conventional background in terms of my prior bioprocess experience, Professor Olsson from the beginning taught me that academic research need not be divorced from industrial applications and relevance. Rather, as academic researchers we have a responsibility to ensure that our fields of study be practical, applicable, and perhaps most importantly, advances the simple challenges society faces. Lisbeth, you always took care of *me* first, and then research. I am thankful for all the support, discussions, and education provided, at times when doubt or frustration could have easily derailed my efforts. Similar to Jens, you have become a friend and mentor, and I hope our *collaboration* until now continues to grow, both professional and personal.

The Bioprocess Research & Development Division of Merck Research Labs, Merck & Co., Inc., was my first industrial research position after graduation from the Massachusetts Institute of Technology (Chemical Engineering, BS 2001; Biomedical Engineering, MEng 2002). I would like to thank Dr. Wayne K. Herber, Dr. John G. Aunins, Dr. Barry Buckland, Dr. Peter Salmon, Dr. John O. Konz, and Dr. David Robinson for supporting my Merck Doctoral Fellowship and selecting my candidacy amongst the most talented group of young scientists and engineers I have ever had the privilege of working with. I would also like to thank the former members of the Fermentation & Cell Culture Group, and the current members of the Vaccine & Biologics Process Development Group.

In particular, there are three individuals during my tenure at Merck that deserve special recognition. First, I am forever indebted and thankful to Dr. Hugh George, Senior Distinguished Research Fellow, retired, of Merck Research Labs. Dr. George, as he will always affectionately be referred too, provided my first introduction to *S. cerevisiae* and fermentation process development. Although often engulfed in the complexities of developing the HPV vaccine, Gardasil™, he consistently reminded me that our scientific edge must be our sharpest – regardless of the managerial or operational necessities that may consume us. Dr. George, you will always be my first yeast supervisor. I only hope I can one day carry on your example.

Second, I would like to profoundly thank my dear friend and colleague, Dr. Ashraf Amanullah, formerly a Senior Research Fellow, Merck Research Labs. Ashraf is truly the biochemical engineer's engineer, and provided invaluable advice, guidance, and most importantly, friendship when I first considered the thought of returning to graduate school, let alone considering a European laboratory. Ashraf continues to exude scientific leadership within industrial bioprocess development, and perhaps, this work gets me a little closer to the standards he has always insisted on. Ashraf, thanks for being that example when few existed.

Third, I would be remised if I did not provide sincere gratitude to Dr. Nedim E. Altaras, Associate Director, Merck Research Labs. Upon return from completing this doctoral research, Nedim has served as my supervisor, mentor, colleague, and friend. It is perhaps quite fitting that I have the privilege to be mentored by one of the first metabolic engineers hired by Merck Research Labs (Dr. Nedim E. Altaras completed his PhD research with Professor Douglas C. Cameron, University of Wisconsin, on the metabolic engineering of the 1,2-propanediol pathway). Nedim, affectionately called Nedo, I would not have been able to complete my thesis, continue to strive for success at Merck, and maintain my sanity without your constant mentoring and support. There were many times you would have been well within your right to not allow me the time you did to finish this thesis satisfactorily. I hope you find my training and research applicable and worth the investment, and that the conclusion of this chapter in my life represents the spring-board from which we will drive success at Merck – together.

One of the happiest memories I will take away from my PhD experience, and perhaps most valuable as I have the privilege to lead a scientific group in my future endeavors, will be the opportunity to teach and supervise undergraduate and graduate students. You can study the material of your research forever, but the ability to teach others is often the positive confirmation that you have in fact learned something. I thank *my students* for giving me the opportunity and honor to be *their teacher*. They included Gionata Scalcinati (MSc), Simon G. Poulsen (BSc), Martin Malthe Borch (BSc), Rasmus Ågren (MSc), Pier Nestola (MSc), and Manos Papadakis (MSc). This thesis and the results it describes would not have been possible without your hard work – I hope you feel it was worth it.

It is often said that your PhD research and experience will be more defined by the people that surround you, rather than the experiments that engulf you. I am forever thankful to those that were fellow PhD students at the time: Roberto Olivares-Hernández, Alessandro Fazio, Siavash Partow, Marta Papini, Jie Zhang, Margarita Salazar Peña, Dr. Donatella Cimini, Dr. Valeria Mapelli, Dr. Wanwipa Vongsangnak, Professor Mohammad Asadollahi, Dr. Songsak Wattanachaisaereekul, Dr. Intawat Nookae, Dr. Mikael R. Andersen, Dr. Susan Meijer, Dr. Ana Paula Oliveira, Dr. Kjeld Kjeldsen, and Dr. Jakob B. Nielsen. In addition, I would like to thank the entire staff, technicians, and faculty at the Center for Microbial Biotechnology, Technical University of Denmark (DTU). In particular, I would like to thank Professor Uffe H. Mortensen, who provided invaluable molecular biology knowledge in terms that even an engineer could understand. Building 223 at DTU, no matter what time of day or season of the year, always had its lights on and fermenters running, all impossible without the dedication of everyone who worked there. I thank you profoundly for the support. I also thank the faculty and staff at the Chalmers University of Technology, whom while I was never a student *in vitro*, always ensured my *in silico* status was upstanding.

As with most high-profile laboratories and large centers, the roles that Post-doctoral associates play are that of enablers, in every sense of the word. I would like to thank Dr. Jérôme Maury, Professor Goutham Vemuri, Professor Michael C. Jewett, Dr. Thomas Grotkjær, and Professor Gianni Panagiotou. We have shared many discussions, debates, evenings out, beers, and incredible times together.

Dr. Thomas, and later on, Dr. Jakob, you have become the closest of friends, and our trips in the US have become that of legends! I only hope we continue to bridge the small gap so many

others call the Atlantic Ocean. The doors are always open, and thank you for being the brothers and family I needed in Denmark.

Dr. Gianni, there are no words in the English and I suspect even the Greek language that could possibly explain the support, guidance, teaching, mentoring, and friendship you have extended to me. When I first started my PhD and we realized that perhaps we could come to enjoy each other's company, you told me that at the end of the process I would not remember the succinate titer of the strains, or the gene expression profiles, but rather the people that were around through the peaks and valleys of that work. You were right. Thank you for your brotherhood. I continue to swing from the trees while you build your Acropolis. This is only the beginning of a friendship that will last a life-time – I have no doubts.

There are many more collaborators, students, colleagues, and friends that could be included here by name. Know that my gratitude extends to all of you, and the limitations of space are only confined to this page – not in my heart. Thank you again.

At the onset, I dedicated this doctoral thesis to my wife, Stephanie P. Otero, to my parents, Jose M. and Joaquina Otero, and to my sister, Dr. Melissa A. Otero. It is therefore quite fitting that I close with a few words about the foundation upon which my life is built. Mom and Dad, you have served as role models and parents first, and friends second, for which I am eternally grateful. I can only hope that this accomplishment reaffirms the values and principles you have instilled in both Melissa and me. Thank you so much for the support and love you have shown every single day, even when those days were often sleepless and stressful. I continuously aspire to be like you, and hope this doctoral thesis is a small step in that direction.

Melissa, throughout this journey you more than anyone else could appreciate the battle that a doctorate can be. I am forever thankful for your support, unwavering smile, and constant love. It is most fitting that you would ultimately come to be the first Dr. Otero in our family – I was not one bit surprised nor do I think that honor should go to anyone else. Thank you for being the sister that I always wanted and at many times, needed.

Lastly, I'd like to extend a thank you to the love of my life, Stephanie. During the course of this doctorate, I had the privilege and honor to get married to the woman who has owned my heart for the better part of 11 years. Throughout all the sacrifice, time apart, sleepless nights in the lab, weekends spent writing, and stressful moments throughout, I know that what made them survivable is knowing you endured them with me. Completing a PhD is a wonderful academic achievement, but sharing it with someone like you makes it worth-while. Thank you for your unwavering support, love, and most of all, understanding.

## 11. References

- Adrio JL and Demain AL "Genetic improvement of processes yielding microbial products," *FEMS Microbiology Reviews* vol. 30, no. 2, pp. 187-214, 2006.
- Aguilera J, Petit T, de Winde JH, and Pronk JT "Physiological and genome-wide transcriptional responses of *Saccharomyces cerevisiae* to high carbon dioxide concentrations," *FEMS Yeast Research* vol. 5, no. 6-7, pp. 579-593, 2005.
- Åkesson M, Förster J, and Nielsen J "Integration of gene expression data into genome-scale metabolic models," *Metabolic Engineering* vol. 6, no. 4, pp. 285-293, 2004.
- Aliverdieva DA, Mamaev DV, Bondarenko DI, and Sholtz KF "Properties of yeast *Saccharomyces cerevisiae* plasma membrane dicarboxylate transporter," *Biochemistry (Moscow)* vol. 71, no. 10, pp. 1161-1169, 2006b.
- Aliverdieva DA, Mamaev DV, Lagutina LS, and Sholtz KF "Specific features of changes in levels of endogenous respiration substrates in *Saccharomyces cerevisiae* cells at low temperature," *Biochemistry (Moscow)* vol. 71, no. 1, pp. 39-45, 2006a.
- Arikawa Y, Enomoto K, Muratsubaki H, and Okazaki M "Soluble fumarate reductase isoenzymes from *Saccharomyces cerevisiae* are required for anaerobic growth," *FEMS Microbiology Letters* vol. 65, no. 1, pp. 111-116, 1998.
- Arikawa Y, Kuroyanagi T, Shimosaka M, Muratsubaki H, Enomoto K, Kodaira R, and Okazaki M "Effect of gene disruptions of the TCA cycle on the production of succinic acid in *Saccharomyces cerevisiae*," *Journal of Bioscience and Bioengineering* vol. 87, no. 1, pp. 28-36, 1999a.
- Arikawa Y, Misato K, Ritsuko K, Makoto S, Haruhiro M, Keichiro E, and Okazaki M "Isolation of sake yeast strains possessing various levels of succinate- and/or malate-producing abilities by gene disruption or mutation," *Journal of Bioscience and Bioengineering* vol. 87, no. 3, pp. 333-339, 1999b.
- Asadollahi MA, Maury J, Møller K, Nielsen KF, Schalk M, Clark A, and Nielsen J "Production of plant sesquiterpenes in *Saccharomyces cerevisiae*: effect of ERG9 repression on sesquiterpene biosynthesis," *Biotechnology and Bioengineering* vol. 99, no. 3, pp. 666-677, 2008.
- Bailey JE "Toward a science of metabolic engineering," *Science* vol. 252, no. 5013, pp. 1668-1675, 1991.
- Barwell CJ and Hess B "Application of kinetics of yeast pyruvate kinase in vitro to calculation of glycolytic flux in the anaerobic yeast cell," *Hoppe Seylers Z Physiol Chem* vol. 353, no. 7, pp. 1178-1184, 1972.
- BASF Financial Report (2005) BASF. Available at [www.bASF.com](http://www.bASF.com)
- Basson ME, Thorsness M, and Rine J "*Saccharomyces cerevisiae* contains two functional genes encoding 3-hydroxy-3-methylglutaryl-coenzyme A reductase," *Proceedings of the National Academy of Science USA* vol. 83, no. 15, pp. 5563-5567, 1986.
- Bhat PJ and Hopper JE "Overproduction of the GAL1 or GAL3 protein causes galactose-

independent activation of the GAL4 protein: evidence for a new model of induction of the yeast GAL/MEL regulon," *Molecular Cell Biology* vol. 12, no. 6, pp. 2701-2707, 1992.

Bhat PJ, Oh D, and Hopper JE "Analysis of the GAL3 signal transduction pathway activating GAL4 protein-dependent transcription in *Saccharomyces cerevisiae*," *Genetics* vol. 125, no. 2, pp. 281-291, 1990.

Bloch K, Chaykin S, Phillips AH, and De Waard A "Mevalonic acid pyrophosphate and isopentenylpyrophosphate," *Journal of Biological Chemistry* vol. 234, pp. 2595-2604, 1959.

Boer VM, de Winde JH, Pronk JT, and Piper MD "The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur," *Journal of Biological Chemistry* vol. 278, no. 5, pp. 3265-3274, 2003.

Bojunga N, Kötter P, and Entian KD "The succinate/fumarate transporter Acr1p of *Saccharomyces cerevisiae* is part of the gluconeogenic pathway and its expression is regulated by Cat8p," *Molecular Genetics and Genomics* vol. 260, no. 5, pp. 453-461, 1998.

Borodina I and Nielsen J "From genomes to *in silico* cells via metabolic networks," *Current Opinion in Biotechnology* vol. 16, no. 3, pp. 350-355, 2005.

Boss JM, Darrow MD, and Zitomer RS "Characterization of yeast iso-1-cytochrome c mRNA," *Journal of Biological Chemistry* vol. 255, no. 18, pp. 8623-8628, 1980.

Breitling R, Vitkup D, and Barrett MP "New surveyor tools for charting microbial metabolic maps," *Nature Reviews in Microbiology* vol. 6, no. 2, pp. 156-161, 2008.

Bro C and Nielsen J "Impact of 'ome' analyses on inverse metabolic engineering," *Metabolic Engineering* vol. 6, no. 3, pp. 204-211, 2004.

Bruggeman FJ and Westerhoff HV "The nature of systems biology," *Trends in Microbiology* vol. 15, no. 1, pp. 45-50, 2007.

Bruggink A, Straathof AJ, and van der Wielen LA "A 'Fine' chemical industry for life science products: green solutions to chemical challenges," *Advances in Biochemical Engineering and Biotechnology* vol. 80, pp. 69-113, 2003.

Bruggnik A "Biocatalysis and process integration in the synthesis of semi-synthetic antibiotics: biotechnology for industrial production of fine chemicals," *Chimia*, vol. 50, pp. 431-2, 1996.

Burgard AP, Pharkya P, and Maranas CD "Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization," *Biotechnology and Bioengineering* vol. 84, no. 6, pp. 647-657, 2003.

Cakir T, Kirdar B, and Ulgen KO "Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks," *Biotechnology and Bioengineering* vol. 86, no. 3, pp. 251-260, 2004.

Cakir T, Patil KR, Onsan Z, Ulgen KO, Kirdar B, and Nielsen J "Integration of metabolome data with metabolic networks reveals reporter reactions," *Molecular Systems Biology* vol. 2, no. 50, 2006.

Camarasa C, Faucet V, and Dequin S “Role in anaerobiosis of the isoenzymes for *Saccharomyces cerevisiae* fumarate reductase encoded by OSM1 and FRDS1,” *Yeast* vol. 24, no. 5, pp. 391-401, 2007.

Camarasa C, Grivet JP, and Dequin S “Investigation by <sup>13</sup>C-NMR and tricarboxylic acid (TCA) deletion mutant analysis of pathways for succinate formation in *Saccharomyces cerevisiae* during anaerobic fermentation,” *Microbiology* vol. 149, no. 9, pp. 2669-2678, 2003.

Carlson M “Glucose repression in yeast,” *Current Opinions in Microbiology* vol. 2, no. 2, pp. 202-207, 1999.

Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, Jia Y, Juvik G, Roe T, Schroeder M, Weng S, and Botstein D “SGD: Saccharomyces Genome Database,” *Nucleic Acids Research* vol. 26, no. 1, pp. 73-79, 1998.

Chicago Board-Trade Library (2009) Chicago Board of Traders. Available at <http://www.cmegroup.com>.

Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabrielian AE, Landsman D, Lockhart DJ, and Davis RW “A genome-wide transcriptional analysis of the mitotic cell cycle,” *Molecular Cell* vol. 2, no. 1, pp. 65-73, 1998.

Cimini C, Patil KR, Schiraldi C, and Nielsen J “Global transcriptional response of *Saccharomyces cerevisiae* to the deletion of SDH3,” *BMC Systems Biology* vol. 3, no. 17, 2009.

Cohn M, Monod J, Pollock MR, Spiegelman S, and Stanier RY “Terminology of enzyme formation,” *Nature* vol. 172, no. 4389, pp. 1096, 1953.

Cori CF, Schmidt G, and Cori GT “The Synthesis of a polysaccharide from glucose-1-phosphate in muscle extract,” *Science* vol. 89, no. 2316, pp. 464-465, 1939.

Cori GT “Enzymes and glycogen structure in glycogenesis,” *Z Kinderheilkd* vol. 10, no. 1-2, pp. 38-42, 1954.

Cori GT and Cori CF “Glucose-6-phosphatase of the liver in glycogen storage disease,” *Journal of Biological Chemistry* vol. 199, no. 2, pp. 661-667, 1952.

Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin II, Selkov E, and Palsson BØ “Metabolic modeling of microbial strains *in silico*,” *Trends in Biochemical Sciences* vol. 26, no. 3, pp. 179-186, 2001.

Crabtree HG “Observations on the carbohydrate metabolism of tumors,” *Journal of Biochemistry* vol. 23, no. 3, pp. 536-545, 1929.

D’Aquino R “Addressing Today’s Energy Challenge,” *Chemical Engineering Progress* vol. 102, no. 6, pp. 12-13, 2006

Daum G, Tuller G, Nemec T, Hrastnik C, Balliano G, Cattel L, Milla P, Rocco F, Conzelmann A, Vionnet C, Kelly DE, Kelly S, Schweizer E, Schüller HJ, Hojad U, Greiner E, and Finger K “Systematic analysis of yeast strains with possible defects in lipid metabolism,” *Yeast* vol. 15, no.

7, pp. 601-614, 1999.

Demain AL “Microbial biotechnology,” *Trends in Biotechnology* vol. 18, no. 1, pp. 26-31, 2000.

Demain AL and Adrio JL “Contributions of microorganisms to industrial biology,” *Mol Biotechnol* vol. 38, no. 1, pp. 41-55, 2008(b).

Demain AL and Adrio JL “Strain improvement for production of pharmaceuticals and other microbial metabolites by fermentation,” *Progress in Drug Research* vol. 65, pp. 251, 253-89, 2008(a).

Demerec M “Production of Penicillin,” *United States Patent and Trademark Office*, US Patent No. 2,445,758, granted 1948.

DeRisi JL, Iyler VR, and Brown PO “Exploring the metabolic and genetic control of gene expression on a genomic scale,” *Science* vol. 278, no. 5338, pp. 680-686, 1997.

Dien BS, Cotta MA, and Jeffries TW “Bacteria engineered for fuel ethanol production: current status,” *Applied Microbiology and Biotechnology* vol. 63, no. 3, pp. 258-266, 2003.

Dodds DR and Gross RA “Chemistry. Chemicals from biomass,” *Science* vol. 318, no. 5854, pp. 1250-1251, 2007.

Douglas HC and Hawthorne DC, “Enzymatic expression and genetic linkage of genes controlling galactose utilization in *Saccharomyces*,” *Genetics* vol. 49, pp. 837-844, 1964.

Dow Corporate Report (2005) The Dow Chemical Company. Available at [www.dow.com](http://www.dow.com)

Duarte NC, Herrgård MJ, and Palsson BØ “Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model,” *Genome Research* vol. 14, no. 7, pp. 1298-1309, 2004b.

Duarte NC, Palsson BØ, and Fu P “Integrated analysis of metabolic phenotypes in *Saccharomyces cerevisiae*,” *BMC Genomics* vol. 5, no. 1, pp. 63, 2004a.

DuPont Annual Review (2005) DuPont. Available at [www.dupont.com](http://www.dupont.com)

Durr IF and Rudney H “The reduction of beta-hydroxy-beta-methyl-glutaryl coenzyme A to mevalonic acid,” *Journal of Biological Chemistry* vol. 253, pp. 2572-2578, 1960.

Eberhardt NL and Rilling HC “Prenyltransferase from *Saccharomyces cerevisiae*. Purification to homogeneity and molecular properties,” *Journal of Biological Chemistry* vol. 250, no. 3, pp. 863-866, 1975.

Edwards JS, Covert M, and Palsson BØ “Metabolic modeling of microbes: the flux-balance approach,” *Environmental Microbiology* vol. 4, no. 3, pp. 133-140, 2002.

Edwards JS, Ibarra RE, and Palsson BØ “*In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data,” *Nature Biotechnology* vol. 19, no. 2, pp. 125-130, 2001.

Edwards JS, Ramakrishna R, and Palsson BØ “Characterizing the metabolic phenotype: a phenotype phase plane analysis,” *Biotechnology and Bioengineering* vol. 77, no. 1, pp. 27-36, 2002a.

Eliasson A, Christensson C, Wahlbom CF, and Hahn-Hägerdal B “Anaerobic xylose fermentation by recombinant *Saccharomyces cerevisiae* carrying XYL1, XYL2, and XKS1 in mineral medium chemostat cultures,” *Applied Environmental Microbiology* vol. 66, no. 8, pp. 3381-3386, 2000.

Eliasson A, Hofmeyr J-HS, Pedler S, and Hahn-Hägerdal B “The xylose reductase/xylitol dehydrogenase/xylulokinase ratio affects product formation in recombinant xylose-utilising *Saccharomyces cerevisiae*,” *Enzyme Microbiology and Technology* vol. 29, pp. 288-297, 2001.

Energy Impact and Implications for Pricing (2005) DuPont’s Economist’s Office, Press Release

Energy Information Administration (2009) US Department of Energy. Available at <http://www.eia.doe.gov>

Enomoto K, Arikawa Y, and Muratsubaki H “Physiological role of soluble fumarate reductase in redox balancing during anaerobiosis in *Saccharomyces cerevisiae*,” *FEMS Microbiology Letters* vol. 215, no. 1, pp. 103-108, 2002.

Erasmus DJ, van der Merwe GK, and van Vuuren HJ “Genome-wide expression analyses: Metabolic adaptation of *Saccharomyces cerevisiae* to high sugar stress,” *FEMS Yeast Research* vol. 3, no. 4, pp. 375-399, 2003.

Eriksson KE, editor. “Biotechnology in the pulp and paper industry,” *Advances in Biochemical Engineering and Biotechnology* vol. 57, pp. 339, 1997.

Famili I, Förster J, Nielsen J, and Palsson BØ “*Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network,” *Proceedings of the National Academy of Science USA* vol. 100, no. 23, pp. 13134-13139, 2003.

Ferrández-García F “Biotechnology. Genetic engineering. Chemico-pharmaceutical area,” *Rev Esp Fisiol* vol. 38 Suppl, pp. 353-366, 1982.

Fisk DG, Ball CA, Dolinski K, Engel SR, Hong EL, Issel-Tarver L, Schwartz K, Sethuraman A, Botstein D, and Cherry JM “*Saccharomyces cerevisiae* S288C genome annotation: a working hypothesis,” *Yeast* vol. 23, no. 12, pp. 857-865, 2006.

Förster J, Famili I, Fu P, Palsson BØ, and Nielsen J “Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network,” *Genome Research* vol. 13, no. 2, pp. 244-253, 2003.

Förster J, Famili I, Palsson BØ, and Nielsen J “Large-scale evaluation of in silico gene deletions in *Saccharomyces cerevisiae*,” *OMICS* vol. 7, no. 2, pp. 193-202, 2003b.

From Niche to Nation : Ethanol Industry Outlook 2006 (2006) Renewable Fuels Association Available at [www.ethanolRFA.org](http://www.ethanolRFA.org)

Gancedo JM “Yeast carbon catabolite repression,” *Microbiology and Molecular Biology Reviews* vol. 62, no. 2, pp. 334-361, 1998.

Gárdonyi M, Jeppsson M, Lidén G, Gorwa-Grausland MF, and Hahn-Hägerdal B “Control of xylose consumption by xylose transport in recombinant *Saccharomyces cerevisiae*,” *Biotechnology and Bioengineering* vol. 82, no. 7, pp. 818-824, 2003.

Gavrilescu M and Chisti Y “Biotechnology--a sustainable alternative for chemical industry,” *Biotechnology Advances* vol. 23, no. 7-8, pp. 471-499, 2005.

Geertman JM, van Maris AJ, van Dijken JP, and Pronk JT “Physiological and genetic engineering of cytosolic redox metabolism in *Saccharomyces cerevisiae* for improved glycerol production,” *Metabolic Engineering* vol. 8, no. 6, pp. 532-542, 2006.

Gill RT “Enabling inverse metabolic engineering through genomics,” *Current Opinions in Biotechnology* vol. 14, no. 5, pp. 484-490, 2003.

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, and Oliver SG “Life with 6000 genes,” *Science* vol. 274, no. 5287, pp. 546, 563-7, 1996.

Goldberg RN, Tewari YB, and Bhat TN “Thermodynamics of Enzyme-Catalyzed Reactions -a Database for Quantitative Biochemistry,” *Bioinformatics* vol. 20, no. 16, pp. 2874-2877, 2004.

Gong C-S “Direct Fermentation of D-xylose to Ethanol by a Xylose-Fermenting Yeast Mutant,” *United State Patent and Trademark Office*, US Patent No. 4,511,656, granted 1985.

Gottschalk A “Prof. Carl Neuberg,” *Nature* vol. 178, no. 4536, pp. 722-723, 1956.

Grabowska D, Karst F, and Szkopińska A “Effect of squalene synthase gene disruption on synthesis of polyprenols in *Saccharomyces cerevisiae*,” *FEBS Letters* vol. 434, no. 3, pp. 406-408, 1998.

Gray JC and Kekwick RG “The inhibition of plant mevalonate kinase preparations by prenyl pyrophosphates,” *Biochimica et Biophysica Acta* vol. 279, no. 2, pp. 290-296, 1972.

Green ML and Karp PD “Genome annotation errors in pathway databases due to semantic ambiguity in partial EC numbers,” *Nucleic Acids Research* vol. 33, no. 13, pp. 4035-4039, 2005.

Growing Innovation: 2009 Ethanol Industry Outlook (2009) Renewable Fuels Association Available at [www.ethanolRFA.org](http://www.ethanolRFA.org)

Guarente L and Mason T “Heme regulates transcription of the CYC1 gene of *S. cerevisiae* via an upstream activation site,” *Cell* vol. 32, no. 4, pp. 1279-1286, 1983.

Guarente L, Lalonde B, Gifford P, and Alani E “Distinctly regulated tandem upstream activation sites mediate catabolite repression of the CYC1 gene of *S. cerevisiae*,” *Cell* vol. 36, no. 2, pp. 503-511, 1984.

Harden A “The Enzymes of Washed Zymin and Dried Yeast (Lebedew). I. Carboxylase,” *Biochemical Journal* vol. 7, no. 2, pp. 214-217, 1913a.

Harden A and Norris RV "The Enzymes of Washed Zymin and Dried Yeast (Lebedeff). II. Reductase," *Biochemical Journal* vol. 8, no. 1, pp. 100-106, 1914a.

Harden A and Young WJ "The Enzymatic Formation of Polysaccharides by Yeast Preparations," *Biochemical Journal* vol. 7, no. 6, pp. 630-636, 1913b.

Harden A and Zilva SS "The Enzymes of Washed Zymin and Dried Yeast (Lebedeff). III. Peroxydase, Catalase, Invertase and Maltase," *Biochemical Journal* vol. 8, no. 3, pp. 217-226, 1914b.

Harden A, Thompson J, and Young WJ "Apparatus for Collecting and Measuring the Gases evolved during Fermentation," *Biochemical Journal* vol. 5, no. 5, pp. 230-235, 1911.

Hatti-Kaul R, Tornvall U, Gustafsson L, and Borjesson P "Industrial biotechnology for the production of bio-based chemicals--a cradle-to-grave perspective," *Trends in Biotechnology* vol. 25, no. 3, pp. 119-124, 2007.

Heinrich R and Rapoport TA "A linear steady-state treatment of enzymatic chains. General properties, control and effector strength," *European Journal of Biochemistry* vol. 42, no. 1, pp. 89-95, 1974a.

Heinrich R and Rapoport TA "A linear steady-state treatment of enzymatic chains. Critique of the crossover theorem and a general procedure to identify interaction sites with an effector," *European Journal of Biochemistry* vol. 42, no. 1, pp. 97-105, 1974b.

Heinrich R and Rapoport TA "Linear theory of enzymatic chains; its application for the analysis of the crossover theorem and of the glycolysis of human erythrocytes," *Acta Biol Med Ger* vol. 31, no. 4, pp. 479-494, 1973.

Hermann BG and Patel M "Today's and tomorrow's bio-based bulk chemicals from white biotechnology: a techno-economic analysis," *Applied Biochemistry and Biotechnology* vol. 136, no. 3, pp. 361-388, 2007a.

Hermann BG, Blok K, and Patel MK "Producing bio-based bulk chemicals using industrial biotechnology saves energy and combats climate change," *Environmental Science and Technology* vol. 41, no. 22, pp. 7915-7921, 2007b.

Herrera S "Industrial biotechnology – a chance at redemption," *Nature Biotechnology* vol. 22, no. 6, pp. 671-675, 2004

Hirche C "Trend Report No. 16: Industrial Biotechnology - White Biotechnology: A promising tool for the chemical industry," *ACHEMA 2006 28th International Exhibition-Congress on Chemical Engineering, Environmental Protection and Biotechnology*, no. 16, pp. 1-7, 2006.

Ho NW, Chen Z, and Brainard AP "Genetically engineered *Saccharomyces* yeast capable of effective cofermentation of glucose and xylose," *Applied Environmental Microbiology* vol. 64, no. 5, pp. 1852-1859, 1998.

Hohmann S "Characterization of PDC6, a third structural gene for pyruvate decarboxylase in *Saccharomyces cerevisiae*," *Journal of Bacteriology* vol. 173, no. 24, pp. 7963-7969, 1991a.

Hohmann S “PDC6, a weakly expressed pyruvate decarboxylase gene from yeast, is activated when fused spontaneously under the control of the PDC1 promoter,” *Current Genetics* vol. 20, no. 5, pp. 373-378, 1991b.

Hong EL, Balakrishnan R, Dong Q, Christie KR, Park J, Binkley G, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Krieger CJ, Livstone MS, Miyasato SR, Nash RS, Oughtred R, Skrzypek MS, Weng S, Wong ED, Zhu KK, Dolinski K, Botstein D, Cherry JM “Gene Ontology annotations at SGD: new data sources and annotation methods,” *Nucleic Acids Research* vol. 36, no. Database issue, pp. D577-81, 2008.

Hong SH and Lee SY “Importance of redox balance on the production of succinic acid by metabolically engineered *Escherichia coli*,” *Applied Microbiology and Biotechnology* vol. 58, no. 3, pp. 286-290, 2002.

Hong SH, Kim JS, Lee, In YH, Choi SS, Rih JK, Kim CH, Jeong H, Hur CG, and Kim JJ “The genome sequence of the capnophilic rumen bacterium *Mannheimia succiniciproducens*,” *Nature Biotechnology* vol. 22, no. 10, pp. 1275-1281, 2004.

Hörtner H, Ammerer G, Hartter E, Hamilton B, Rytka J, Bilinski T, and Ruis H “Regulation of synthesis of catalases and iso-1-cytochrome c in *Saccharomyces cerevisiae* by glucose, oxygen and heme,” *European Journal of Biochemistry* vol. 128, no. 1, pp. 179-184.

Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, and Hood L “Integrated genomic and phenotypic analyses of a systematically perturbed metabolic network,” *Science* vol. 292, no. 5518, pp. 929-934, 2001.

Jantama K, Haupt MJ, Svoronos SA, Zhang X, Moore JC, Shanmugam KT, and Ingram LO “Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* C that produce succinate and malate,” *Biotechnology and Bioengineering* vol. 99, no. 5, pp. 1140-1153, 2008.

Jeffries TW “Engineering yeasts for xylose metabolism,” *Current Opinions in Biotechnology* vol. 17, no. 3, pp. 320-326, 2006.

Jennings SM, Tsay YH, Fisch TM, and Robinson GW “Molecular cloning and characterization of the yeast gene for squalene synthetase,” *Proceeding of the National Academy of Science USA* vol. 88, no. 14, pp. 6038-6042, 1991.

Jeppsson M, Träff K, Johansson B, Hahn-Hägerdal B, and Gorwa-Grauslund MF “Effect of enhanced xylose reductase activity on xylose consumption and product distribution in xylose-fermenting recombinant *Saccharomyces cerevisiae*,” *FEMS Yeast Research* vol. 3, no. 2, pp. 167-175, 2003.

Kacser H and Burns JA “The control of flux,” *Symposia of the Society for Experimental Biology* vol. 27, pp. 65-104, 1973.

Kahvejian A, Quackenbush J, and Thompson JF “What would you do if you could sequence everything?,” *Nature Biotechnology* vol. 26, no. 10, pp. 1125-1133, 2008.

Kakhniashvili D, Mayor JA, Gremse DA, Xu Y, and Kaplan RS “Identification of a novel gene encoding the yeast mitochondrial diacarboxylate transport protein via overexpression,

purification, and characterization of its protein product," *Journal of Biological Chemistry* vol. 272, no. 7, pp. 4516-4521, 1997.

Kamm B and Kamm M "Biorefineries – Multi Product Processes," *Advances in Biochemical Engineering and Biotechnology* vol. 105, pp. 107-204, 2007.

Kern A, Tilley E, Hunter IS, Legisa M, and Glieder A "Engineering primary metabolic pathways of industrial micro-organisms," *Journal of Biotechnology* vol. 129, no. 1, pp. 6-29, 2007.

Kim TY, Kim HU, Park JM, Song H, Kim JS, and Lee SY "Genome-scale analysis of *Mannheimia succiniciproducens* metabolism," *Biotechnology and Bioengineering* vol. 97, no. 4, pp. 657-671, 2007.

Kleff S "Succinic acid production from biomass," *Symposium for Industrial Microbiology*, oral presentation publically available at <http://www.mbi.org/pres-archive.html> (last accessed Oct. 25, 2009), 2004.

Kluyver AJ "Microbial metabolism and its bearing on the cancer problem," *Science* vol. 76, no. 1980, pp. 527-532, 1932.

Kobayashi S and Fukumi H "Studies on the Embden-Meyerhof system and the Warburg-Lipman-Dickens pathway of the organisms of Pertussis-Parapertussis-Bronchisepticus group," *Japanese Journal of Medical Science and Biology* vol. 7, no. 4, pp. 377-384, 1954.

Koffas M and Stephanopoulos G "Strain improvement by metabolic engineering: lysine production as a case study for systems biology," *Current Opinions in Biotechnology* vol. 16, no. 3, pp. 361-366, 2005.

Kötter P and Ciriacy M "Xylose fermentation by *Saccharomyces cerevisiae*," *Applied Microbiology and Biotechnology* vol. 38, pp. 776-783, 1993.

Krebs HA "Metabolism of amino-acids: Deamination of amino-acids," *Biochemical Journal* vol. 29, no. 7, pp. 1620-1644, 1935a.

Krebs HA "Metabolism of amino-acids: The synthesis of glutamine from glutamic acid and ammonia, and the enzymic hydrolysis of glutamine in animal tissues," *Biochemical Journal* vol. 29, no. 8, pp. 1951-1969, 1935b.

Krebs HA "Micro-determination of alpha-ketoglutaric acid," *Biochemical Journal* vol. 32, no. 1, pp. 108-112, 1938a.

Krebs HA "Otto Heinrich Warburg, 1883-1970," *Biographical Memoirs of the Royal Society* vol. 18, pp. 629-699, 1972.

Krebs HA "The citric acid cycle and the Szent-Gyorgyi cycle in pigeon breast muscle," *Biochemical Journal* vol. 34, no. 5, pp. 775-779, 1940a.

Krebs HA "The role of fumarate in the respiration of *Bacterium coli commune*," *Biochemical Journal* vol. 31, no. 11, pp. 2095-2124, 1937.

Krebs HA "The tricarboxylic acid cycle," *Harvey Lecture* vol. Series 44, pp. 165-199, 1948.

Krebs HA and Cohen PP "Metabolism of alpha-ketoglutaric acid in animal tissues," *Biochemical Journal* vol. 33, no. 11, pp. 1895-1899, 1939.

Krebs HA and Eggleston LV "The oxidation of pyruvate in pigeon breast muscle," *Biochemical Journal* vol. 34, no. 3, pp. 442-459, 1940b.

Krebs HA and Holzach O "The conversion of citrate into cis-aconitate and isocitrate in the presence of aconitase," *Biochemical Journal* vol. 52, no. 3, pp. 527-528, 1952a.

Krebs HA, Gurin S, and Eggleston LV. *Biochemical Journal* vol. 51, pp. 614, 1952b.

Krebs HA, *Nobel Lectures in Physiology Medicine, 1942-1962*, Elsevier Publishing Company, Amsterdam, 1964.

Krebs HA, Salvin E, and Johnson WA "The formation of citric and alpha-ketoglutaric acids in the mammalian body," *Biochemical Journal* vol. 32, no. 1, pp. 113-117, 1938b.

Kubo Y, Takagi H, and Nakamori S "Effect of gene disruption of succinate dehydrogenase on succinate production in a sake yeast strain," *Journal of Bioscience and Bioengineering* vol. 90, no. 6, pp. 619-624, 2000.

Kuyper M, Harhangi HR, Stave AK, Winkler AA, Jetten MS, de Laat WT, de Ridder JJ, Op den Camp HJ, van Dijken JP, and Pronk JT "High-level functional expression of a fungal xylose isomerase: the key to the efficient ethanolic fermentation of xylose by *Saccharomyces cerevisiae*?", *FEMS Yeast Research* vol. 4, no. 1, pp. 69-78, 2003.

Kuyper M, Winkler AA, van Dijken JP, and Pronk JT "Minimal metabolic engineering of *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: a proof of principle," *FEMS Yeast Research* vol. 4, no. 6, pp. 655-664, 2004.

Kyprides NC "Genomes OnLine Database (GOLD 1.0): a monitor of complete and ongoing genome projects world-wide," *Bioinformatics* vol. 15, pp. 773-774, 1999.

Laffend L, Nagarajan V, and Nakamura C "Bioconversion of a fermentable carbon source to 1,3-propanediol by a single microorganism," *United States Patent and Trademark Office*, US Patent No. 5,686,276, granted 1997.

Lee JW, Lee SY, Song H, and Yoo J-S "The proteome of *Mannheimia succiniciproducens*, a capnophilic rumen bacterium," *Proteomics* vol. 6, no. 12, pp. 3550-3566, 2006.

Lee PC, Lee SY, Hong SH, and Chang HN "Batch and continuous cultures of *Mannheimia succiniciproducens* MBEL55E for the production of succinic acid from whey and corn steep liquor," *Bioprocess and Biosystems Engineering* vol. 26, no. 1, pp. 63-67, 2003.

Lee PC, Lee SY, Hong SH, and Chang HN "Isolation and characterization of a new succinic acid-producing bacterium, *Mannheimia succiniciproducens* MBEL55E, from bovine rumen," *Applied Microbiology and Biotechnology* vol. 58, no. 5, pp. 663-668, 2002.

Lee PC, Lee WG, Kwon S, Lee SY, and Chang HN "Batch and continuous cultivation of *Anaerobiospirillum succiniciproducens* for the production of succinic acid from whey," *Applied Microbiology and Biotechnology* vol. 54, no. 1, pp. 23-27, 2000.

Lee SJ, Song H, and Lee SY "Genome-based metabolic engineering of *Mannheimia succiniciproducens* for succinic acid production," *Applied Environmental Microbiology* vol. 72, no. 3, pp. 1939-1948, 2006a.

Lei F "Dynamics and nonlinear phenomena in continuous cultivations of *Saccharomyces cerevisiae*," Doctoral Thesis, Department of Chemical Engineering, Technical University of Denmark, pp. 0-231, 2001.

Lespinet O and Labedan B "Orphan enzymes could be an unexplored reservoir of new drug targets," *Drug Discovery Today* vol. 11, no. 7-8, pp. 300-305, 2006.

Lin H, Bennett GN, and San KY "Genetic reconstruction fo the aerobic central carbon metabolism in *Escherichia coli* for the absolute aerobic production of succinate," *Biotechnology and Bioengineering* vol. 89, no. 2, pp. 148-156, 2004a.

Lin H, Bennett GN, and San KY "Metabolic engineering of aerobic succinate productions systems in *Escherichia coli* to improve process productivity and achieve the maximum theoretical succinate yield," *Metabolic Engineering* vol. 7, no. 2, pp. 116-127, 2005.

Liolios K, Tavernarakis N, Hugenholtz P, and Kyrpides NC "The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide," *Nucleic Acids Research* vol. 34, pp. 332-340, 2006.

Lodi T, Saliola M, Donnini C, and Goffrini P "Three target genes for the transcriptional activator Cat8p of *Kluyveromyces lactis*: acetyl coenzyme A synthetase genes KlACS1 and KlACS2 and lactate permease gene KlJEN1," *Journal of Bacteriology* vol. 183, no. 18, pp. 5257-5261, 2001.

Loftus TM, Hall LV, Anderson SL, and McAlister-Henn L "Isolation, characterization, and disruption of the yeast gene encoding cytosolic NADP-specific isocitrate dehydrogenase," *Biochemistry* vol. 33, no. 32, pp. 9661-9667, 1994.

Luttik MA, Overkamp KM, Kötter P, de Vries S, van Dijken JP, and Pronk JT "The *Saccharomyces cerevisiae* NDE1 and NDE2 genes encode separate mitochondrial NADH dehydrogenases catalyzing the oxidation of cytosolic NADH," *Journal of Biological Chemistry* vol. 272, no. 38, pp. 24529-24534, 1998.

Lynch MD, Gill RT, and Stephanopoulos G "Mapping phenotypic landscapes using DNA micro-arrays," *Metabolic Engineering* vol. 6, no. 3, pp. 177-185, 2004.

Lynd LR, Wyman C, Laser M, Johnson D, and Landucci R "Strategic Biorefinery Analysis: Analysis of Biorefineries," Subcontract Report NREL/SR-510-35578, National Renewable Energy Laboratory, January 24, 2002 – July 1, 2002.

Lynd LR, Wyman CE, and Gerngross TU "Biocommodity Engineering," *Biotechnology Progress* vol. 15, no. 5, pp. 777-793, 1999.

Mann T “Prof. J. K. Parnas,” *Nature* vol. 175, no. 4456, pp. 532-533, 1955.

Maury J, Asadollahi MA, Møller K, Clark A, and Nielsen J “Microbial isoprenoid production: an example of green chemistry through metabolic engineering,” *Advances in Biochemical Engineering and Biotechnology* vol. 100, pp. 19-51, 2005.

McKinlay JB, Vieille C, and Zeikus JG “Prospects for a bio-based succinate industry,” *Applied Microbiology and Biotechnology* vol. 76, no. 4, pp. 727-740, 2007.

Meyerhof O “Further studies on the Harden-Young effect in alcoholic fermentation of yeast preparations,” *Journal of Biological Chemistry* vol. 180, no. 2, pp. 575-586, 1949a.

Meyerhof O and Green H “Synthetic action of phosphatase; equilibria of biological esters,” *Journal of Biological Chemistry* vol. 178, no. 2, pp. 655-667, 1949b.

Meyerhof O and Green H “Transphosphorylation by alkaline phosphatase in the absence of nucleotides,” *Science* vol. 110, no. 2863, pp. 503 1949c.

Meyerhof O and Oesper P “The enzymatic equilibria of phospho(enol)pyruvate,” *Journal of Biological Chemistry* vol. 179, no. 3, pp. 1371-1385, 1949d.

Meyerhof O and Wilson JR “Comparative study of the glycolysis and ATP-ase activity in tissue homogenates,” *Archives of Biochemistry and Biophysics* vol. 23, no. 2, pp. 246-255, 1949g.

Meyerhof O and Wilson JR “Studies on the enzymatic system of tumor glycolysis; glycolysis of free sugar in homogenates and extracts of transplanted rat sarcoma,” *Archives of Biochemistry and Biophysics* vol. 21, no. 1, pp. 1-21, 1949e.

Meyerhof O and Wilson JR “Studies on the enzymatic system of tumor glycolysis; comparative study of rat and mouse tumor homogenates,” *Archives of Biochemistry and Biophysics* vol. 21, no. 1, pp. 22-34, 1949f.

Michal G, *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*, 1<sup>st</sup> ed., John Wiley & Sons Inc. and Spektrum Akademischer Verlag Co-Publication, New York, 1999.

Middleton B and Tubbs PK “3-Hydroxy-3-methylglutaryl-CoA synthase from baker’s yeast,” *Methods in Enzymology* vol. 35, pp. 173-177, 1975.

Minard KI and McAlister-Henn L “Sources of NADPH in yeast vary with carbon source,” *Journal of Biological Chemistry* vol. 280, no. 48, pp. 39890-39896, 2005.

Minard KI, Jennings GT, Loftus TM, Xuan D, and McAlister-Henn L “Sources of NADPH and expression of mammalian NADP+-specific isocitrate dehydrogenases in *Saccharomyces cerevisiae*,” *Journal of Biological Chemistry* vol. 273, no. 47, pp. 31486-31493, 1998.

Moo-Young M, Glick BR, Bu'Lock JD, and Cooney CL “Biotechnology has been associated with the pharmaceutical and brewing industries,” *Biotechnology Advances* vol. 2, no. 2, pp. I-II, 1984.

Moreira dos Santos M, Gombert AK, Christensen B, Olsson L, and Nielsen J "Identification of *in vivo* enzyme activities in the catabolism of glucose and acetate by *Saccharomyces cerevisiae* by using  $^{13}\text{C}$ -labeled substrates," *Eukaryotic Cell* vol. 2, no. 3, pp. 599-608, 2003.

Morozova O and Marra MA "Applications of next-generation sequencing technologies in functional genomics," *Genomics* vol. 92, no. 5, pp. 255-264, 2008.

Nakamura C, Gatenby A, Hsu AK, La Reau R, Haynie L, Diaz-Torres M, Trimbur D, Whited G, Nagarajan V, Payne M, Picataggio S, and Nair R "Method for the production of 1,3-propanediol by recombinant microorganisms," *United States Patent and Trademark Office* vol. 1997, no. 6,013,494, 2000.

Nielsen J "Metabolic engineering," *Applied Microbiology and Biotechnology* vol. 55, no. 3, pp. 263-283, 2001.

Nielsen J and Jewett MC "Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*," *FEMS Yeast Research* vol. 8, no. 1, pp. 122-131, 2008.

Nielsen J, Villadsen J, and Lidén, *Bioreaction Engineering Principles*, Kluwer Academic/Plenum Publishers, New York, 2<sup>nd</sup> edition, 2003.

Nissen TL, Schulze U, Nielsen J, and Villadsen J "Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*," *Microbiology* vol. 143, No. 1, pp. 203-218, 1997.

Nookaew I, Jewett MC, Meechai A, Thammarongtham C, Laoteng K, Cheevadhanarak S, Nielsen J, and Bhumiratana S "The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism," *BMC Systems Biology* vol. 2, pp. 71, 2008.

Octave S and Thomas D "Biorefinery: Toward an industrial metabolism," *Biochimie* vol. 91, no. 6, pp. 659-664, 2009.

Office of Budget (2008) National Institutes of Health. The fiscal year 2008 budget available at <http://officeofbudget.od.nih.gov/ui/HomePage.htm>

Office of Budget (2008) US Department of Energy. The fiscal year 2008 budget available at <http://www.doe.gov/about/budget.htm>

Oliver SG "From genomes to systems: the path with yeast," *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 361, no. 1467, pp. 477-482, 2006.

Onishi T "Mechanisms of electron transport and energy conservation in the site I region of the respiratory chain," *Biochimica et Biophysica Acta* vol. 301, no. 2, pp. 102-128, 1973.

Otero JM, Panagiotou G, and Olsson L "Fueling Industrial Biotechnology Growth with Bioethanol," *Advances in Biochemical Engineering and Biotechnology* vol. 108, pp. 1-40, 2007.

Overkamp KM, Bakker BM, Kötter P, van Tuijl A, de Vries S, van Dijken JP, and Pronk JT “*In vivo* analysis of the mechanisms for oxidation of cytosolic NADH by *Saccharomyces cerevisiae* mitochondria,” *Journal of Bacteriology* vol. 182, no. 10, pp. 2823-2830, 2000.

Oyedotun KS and Lemire BD “The *Saccharomyces cerevisiae* succinate-ubiquinone oxidoreductase. Identification of Sdh3p amino acid residues involved in ubiquinone binding,” *Journal of Biological Chemistry* vol. 274, no. 34, pp. 23956-23962, 1999.

Papagianni M “Advances in citric acid fermentation by *Aspergillus niger*: Biochemical aspects, membrane transport and modeling,” *Biotechnology Advances* vol. 25, pp. 244-263, 2007.

Parekh S, Vinci VA, and Strobel RJ “Improvement of microbial strains and fermentation processes,” *Applied Microbiology and Biotechnology* vol. 54, no. 3, pp. 287-301, 2000.

Pass F “Biotechnology, a new industrial revolution,” *Journal of the American Academy of Dermatology* vol. 4, no. 4, pp. 476-477, 1981.

Patil KR and Nielsen J “Uncovering transcriptional regulation of metabolism by using metabolic network topology,” *Proceedings of the National Academy of Science USA* vol. 102, no. 8, pp. 2685-2690, 2005.

Patil KR, Åkesson M, and Nielsen J “Use of genome-scale microbial models for metabolic engineering,” *Current Opinions in Biotechnology* vol. 15, no. 1, pp. 64-69, 2004.

Patnaik R “Engineering complex phenotypes in industrial strains,” *Biotechnology Progress* vol. 24, no. 1, pp. 38-47, 2008.

Petrik M, Käppeli O, and Fiechter A “An expanded concept for the glucose effect in the yeast *Saccharomyces uvarum*: involvement of short- and long-term regulation,” *Journal Genetics and Microbiology* vol. 129, pp. 43-49, 1983.

Petroleum Marketing Annual 2005: Preliminary Report, Energy Information Administration (2006) US Department of Energy. Available at <http://www.eia.doe.gov>

Phelps TJ, Palumbo AV, and Beliaev AS “Metabolomics and microarrays for improved understanding of phenotypic characteristics controlled by both genomics and environmental constraints,” *Current Opinions in Biotechnology* vol. 13, no. 1, pp. 20-24, 2002.

Poppe L and Novak L, *Selective biocatalysis: a synthetic approach*, Wiley-VCH, Weinheim, 1992.

Porter JW “Mevalonate kinase,” *Methods in Enzymology* vol. 110, pp. 71-78, 1985.

Postma E, Verudyn C, Scheffers WA, and Van Dijken JP “Enzymic analysis of the Crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*,” *Applied Environmental Microbiology* vol. 55, no. 2, pp. 468-477, 1989.

Price ND, Papin JA, Schilling CH, and Palsson BØ “Genome-scale microbial *in silico* models: the constraints-based approach,” *Trends in Biotechnology* vol. 21, no. 4, pp. 162-169, 2003.

Pronk JT, Yde Steensma H, and Van Dijken JP “Pyruvate metabolism in *Saccharomyces cerevisiae*,” *Yeast* vol. 12, no. 16, pp. 1607-1633, 1996.

Pulley HC and Greaves JD “An Application of the Autocatalytic Growth Curve to Microbial Metabolism,” *Journal of Bacteriology* vol. 24, no. 2, pp. 145-168, 1932.

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJJ, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, and Templer R “The path forward for biofuels and biomaterials,” *Science* vol. 311, no. 5760, pp. 484-490, 2006

Rapoport TA, Heinrich R, Jacobasch G, and Rapoport S “A linear steady-state treatment of enzymatic chains. A mathematical model of glycolysis of human erythrocytes,” *European Journal of Biochemistry* vol. 42, no. 1, pp. 107-120, 1974.

Regenberg B, Grotkjær T, Winther O, Fausbøll A, Åkesson M, Bro C, Hansen LK, Brunak S, and Nielsen J “Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in *Saccharomyces cerevisiae*,” *Genome Biology* vol. 7, no. 11, pp. R107, 2006.

Reisch MS “Chemicals from Biorefineries,” *Chemical & Engineering News* vol. 87, no. 41, pp. 28-29, 2009.

Remize F, Andrieu E, and Dequin S “Engineering of the pyruvate dehydrogenase bypass in *Saccharomyces cerevisiae*: role of the cytosolic Mg(2+) and mitochondrial K(+) acetaldehyde dehydrogenases Ald6p and Ald4p in acetate formation during alcoholic fermentation,” *Applied Environmental Microbiology* vol. 66, no. 8, pp. 3151-3159, 2000.

Rieger M, Käppeli O, and Fiechter A “The role of limited respiration in the incomplete oxidation of glucose by *Saccharomyces cerevisiae*,” *Journal of Genetics and Microbiology*, vol. 129, pp. 653-661.

Riggs AD “Method for Microbial Polypeptide Expression,” *United States Patent and Trademark Office*, US Patent No. 4,366,246, granted 1982.

Rilling HC “Eukaryotic prenyltransferases,” *Methods in Enzymology* vol. 110, pp. 145-152, 1985.

Robertson GP, Dale VH, Doering OC, Hamburg SP, Melillo JM, Wander MM, Parton WJ, Adler PR, Barney JN, Cruse RM, Duke CS, Fearnside PM, Follett RF, Gibbs HK, Goldemberg J, Mladenoff DJ, Ojima D, Palmer MW, Sharpley A, Wallace L, Weathers KC, Wiens JA, and Wilhelm WW “Sustainable Biofuels Redux,” *Science* vol. 322, no. 5898, pp. 49-50, 2008.

Robertson TB, *The Chemical Basis of Growth Senescence*. J.B. Lippincott Company, Washington Square Press, Philadelphia, 1923.

Roca C, Nielsen J, and Olsson L “Metabolic engineering of ammonium assimilation in xylose-fermenting *Saccharomyces cerevisiae* improves ethanol production,” *Applied Environmental Microbiology* vol. 69, no. 8, pp. 4732-4736, 2003.

Romano P, Suzzi G, Turbanti L, and Polzinelli M “Acetaldehyde production in *Saccharomyces cerevisiae* wine yeasts,” *FEMS Microbiology Letters* vol. 118, no. 3, pp. 213-218, 1994.

Ronne H “Glucose repression in fungi,” *Trends in Genetics* vol. 11, no. 1, pp. 12-17, 1995.

Roscoe H and Harden A, *A New View of the Origin of Dalton's Atomic Theory*, Macmillan and Co., London, 1895.

Royal Dutch Shell Corporate Report (2005) The Royal Dutch Shell Company. Available at [www.shell.com](http://www.shell.com)

Russo S, Berkovitz Siman-Tov R, and Poli G "Yeasts: from genetics to biotechnology," *Journal of Environmental Pathology, Toxicology and Oncology* vol. 14, no. 3-4, pp. 133-157, 1995.

Sano C "History of glutamate production," *American Journal of Clinical Nutrition* vol. 90, no. 3, pp. 728S-732S, 2009.

Santangelo GM "Glucose signaling in *Saccharomyces cerevisiae*," *Microbiology and Molecular Biology Reviews* vol. 70, no. 1, pp. 253-282, 2006.

Sauer M, Porro D, Mattanovich D, and Branduardi P "Microbial production of organic acids: expanding markets," *Trends in Biotechnology* vol. 26, no. 2, pp. 100-108, 2008.

Savageau MA "Biochemical systems analysis. 3. Dynamic solutions using a power-law approximation," *Journal of Theoretical Biology* vol. 26, no. 2, pp. 215-226, 1970.

Savageau MA "Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions," *Journal of Theoretical Biology* vol. 25, no. 3, pp. 365-369, 1969a.

Savageau MA "Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation," *Journal of Theoretical Biology* vol. 25, no. 3, pp. 370-379, 1969b.

Schacherer J, Ruderfer DM, Gresham D, Dolinski K, Botstein D, and Kruglyak L "Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains," *PLOS One* vol. 2, no. 3, pp. e322, 2007.

Schmeisser C, Steele H, and Streit WR "Metagenomics, biotechnology with non-culturable microbes," *Applied Microbiology and Biotechnology* vol. 75, no. 5, pp. 955-962, 2007.

Schmid RD, *Pocket guide to biotechnology and genetic engineering*, Wiley-VCH, Weinheim, 2003.

Schubert C "Can biofuels finally take center stage?," *Nature Biotechnology* vol. 24, no. 7, pp. 777-784, 2006.

Shendure J and Ji H "Next-generation DNA sequencing," *Nature Biotechnology* vol. 26, no. 10, pp. 1135-1145, 2008.

Short PL "Global Top 50," *Chemical & Engineering News* vol. 84, no. 30, pp. 13-16, 2006.

Sigma-Aldrich Product Information. Succinic acid, Product Number S 3674. Sigma-Aldrich Catalog 2004-2005.

Skoog K and Hahn-Hägerdal B “Effect of oxygenation on xylose fermentation by *Pichia stipitis*,” *Applied Environmental Microbiology* vol. 56, no. 11, pp. 3389-3394, 1990.

Smedsgaard J and Nielsen J “Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics,” *Journal of Experimental Botany* vol. 56, no. 410, pp. 273-286, 2005.

Sonderegger M, Jeppsson M, Hahn-Hägerdal B, and Sauer U “Molecular basis for anaerobic growth of *Saccharomyces cerevisiae* on xylose, investigated by global gene expression and metabolic flux analysis,” *Applied Environmental Microbiology* vol. 70, no. 4, pp. 2307-2317, 2004.

Song H and Lee SJ “Production of succinic acid by bacterial fermentation,” *Enzyme and Microbial Technology* vol. 39, no. 3, pp. 352-361, 2006.

Srivatsan A, Han Y, Peng J, Tehranchi AK, Gibbs R, Wang JD, Chen R “High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies,” *PLOS Genetics* vol. 4, no. 8, pp. e1000139, 2008.

Stephanopoulos G and Vallino JJ “Network rigidity and metabolic engineering in metabolite overproduction,” *Science* vol. 252, no. 5013, pp. 1675-1681, 1991.

Takors R, Bathe B, Rieping M, Hans S, Kelle R, and Huthmacher K “Systems biology for industrial strains and fermentation processes--example: amino acids,” *Journal of Biotechnology* vol. 129, no. 2, pp. 181-190, 2007.

Tantirungkij M, Nakashima N, Seki T, Yoshida T “Construction of xylose-assimilating *Saccharomyces cerevisiae*,” *Journal of Fermentation and Bioengineering* vol. 75, pp. 83-88, 1993.

Tchen TT “Mevalonic kinase: purification and properties,” *Journal of Biological Chemistry* vol. 233, no. 5, pp. 1100-1103, 1958.

The World Fact Book (2006) Central Intelligence Agency, US. Available at <https://www.cia.gov/cia/publications/factbook/>

Thomas D, Cherest H, and Surdin-Kerjan Y “Identification of the structural gene for glucose-6-phosphate dehydrogenase in yeast. Inactivation leads to a nutritional requirement for organic sulfur,” *The EMBO Journal* vol. 10, no. 3, pp. 547-553, 1991.

Toivari MH, Aristidou A, Ruohonen L, and Penttilä M “Conversion of xylose to ethanol by recombinant *Saccharomyces cerevisiae*: importance of xylulokinase (XKS1) and oxygen availability,” *Metabolic Engineering* vol. 3, no. 3, pp. 236-249, 2001.

TOTAL Global Report (2005) TOTAL. Available at [www.total.com](http://www.total.com)

Träff-Bjerre KL, Jeppsson M, Hahn-Hägerdal B, and Gorwa-Grauslund MF “Endogenous NADPH-dependent aldose reductase activity influences production formation during xylose consumption in recombinant *Saccharomyces cerevisiae*,” *Yeast* vol. 21, no. 2, pp. 141-150, 2004.

Tsay YH and Ronbinson GW “Cloning and characterization of ERG8, an essential gene of *Saccharomyces cerevisiae* that encodes phosphomevalonate kinase,” *Molecular Cell Biology* vol. 11, no. 2,

pp. 620-631, 1991.

Tyo KE, Alper HS, and Stephanopoulos GN “Expanding the metabolic engineering toolbox: more options to engineer cells,” *Trends in Biotechnology* vol. 25, no. 3, pp. 132-137, 2007.

US Department of Energy (2008), “Top Value Added Chemicals from Biomass: Volume I”, available at <http://www1.eere.energy.gov/biomass/pdfs/35523.pdf>, last accessed May 5, 2008.

van Gulik WM and Heijnen JJ “A metabolic network stoichiometry analysis of microbial growth and product formation,” *Biotechnology and Bioengineering* vol. 48, no. 6, pp. 681-698, 1995.

Van Urk H, Voll WS, Scheffers WA, and Van Dijken JP “Transient-State Analysis of Metabolic Fluxes in Crabtree-Positive and Crabtree-Negative Yeasts,” *Applied Environmental Microbiology* vol. 56, no. 1, pp. 281-287, 1990.

Varma A and Palsson BØ “Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110,” *Applied Environmental Microbiology* vol. 60, no. 10, pp. 3724-3731, 1994.

Vemuri GN and Aristidou AA “Metabolic engineering in the -omics era: elucidating and modulating regulatory networks,” *Microbiology and Molecular Biology Reviews* vol. 69, no. 2, pp. 197-216, 2005.

Vemuri GN, Eiteman MA, and Altman E “Effects of growth rate mode and pyruvate carboxylase on succinic acid production by metabolically engineered strains of *Escherichia coli*,” *Applied Environmental Microbiology* vol. 68, no. 4, pp. 1715-1727, 2002a.

Vemuri GN, Eiteman MA, and Altman E “Succinate production in dual-phase *Escherichia coli* fermentations depends on the time of transition from aerobic to anaerobic conditions,” *Journal of Industrial Microbiology and Biotechnology* vol. 28, no. 6, pp. 325-332, 2002b.

Vemuri GN, Eiteman MA, McEwen JE, Olsson L, and Nielsen J “Increasing NADH oxidation reduces overflow metabolism in *Saccharomyces cerevisiae*,” *Proceedings of the National Academy of Science USA* vol. 104, no. 7, pp. 2402-2407, 2007.

Verho R, Londesborough J, Penttilä M, and Richard P “Engineering redox cofactor regeneration for improved pentose fermentation in *Saccharomyces cerevisiae*,” *Applied Environmental Microbiology* vol. 69, no. 10, pp. 5892-5897, 2003.

Viswanathan G, Seto J, Patil S, Nudelman G, and Sealfon S “Getting Started in Biological Pathway Construction and Analysis,” *PLoS Computational Biology* vol. 4, no. 2, pp. e16, 2008.

Viswanathan GA, Seto J, Patil S, Nudelman G, and Sealfon SC “Getting started in biological pathway construction and analysis,” *PLOS Computational Biology* vol. 4, no. 2, pp. e16, 2008.

von Jagow G and Klingenberg M “Hydrogen pathways in the mitochondrion of *Saccharomyces carlsbergensis*,” *European Journal of Biochemistry* vol. 12, no. 3, pp. 585-832, 1970.

Wahlbom CF and Hahn-Hägerdal B “Furfural, 5-hydroxymethyl furfural, and acetoin act as external electron acceptors during anaerobic fermentation of xylose in recombinant *Saccharomyces cerevisiae*,” *Biotechnology and Bioengineering* vol. 78, no. 2, pp. 172-178, 2002.

Walfridsson M, Hallborn J, Penttilä M, Keränen S, and Hahn-Hägerdal B “Xylose-metabolizing *Saccharomyces cerevisiae* strains overexpressing the TKL1 and TAL1 genes encoding the pentose phosphate pathway enzymes transketolase and transaldolase,” *Applied Environmental Microbiology* vol. 61, no. 12, pp. 4184-4190, 1995.

Wang Y, Pierce M, Schneper L, Güldal CG, Zhang X, Tavazoie S, and Broach JR “Ras and Gpa2 mediate one branch of a redundant glucose signaling pathway in yeast,” *PLoS Biology* vol. 2, no. 5, pp. E128, 2004.

Warburg O “Über den Stoffwechsel von Tumoren,” Springer, Berlin, 1926.

Warburg O and Christian W “Pyridin, the hydrogen-transferring component of the fermentation enzymes (pyridine nucleotide),” *Biochemische Zeitschrift* vol. 287, 1936.

Warner JR, Patnaik R, and Gill RT “Genomics enabled approaches in strain engineering,” *Current Opinions in Microbiology* vol. 12, no. 3, pp. 223-230, 2009.

Watanabe S, Saleh Aa, Pack SP, Annaluru N, Kodaki T, and Makino K “Ethanol production from xylose by recombinant *Saccharomyces cerevisiae* expressing protein engineered NADP+-dependent xylitol dehydrogenase,” *Journal of Biotechnology* vol. 130, no. 3, pp. 316-319, 2007.

Wendisch VF “Genome-wide expression analysis in *Corynebacterium glutamicum* using DNA microarrays,” *Journal of Biotechnology* vol. 104, no. 1-3, pp. 273-285, 2003.

Wendisch VF, Bott M, and Eikmanns BJ “Metabolic engineering of *Escherichia coli* and *Corynebacterium glutamicum* for biotechnological production of organic acids and amino acids,” *Current Opinion in Biotechnology* vol. 9, no. 3, pp. 268-274, 2006.

Westerhoff HV and Palsson BØ “The evolution of molecular biology into systems biology,” *Nature Biotechnology* vol. 22, no. 10, pp. 1249-1252, 2004.

Williams H, Boyer HW, and Helsinki DR “Size and base composition of RNA in supercoiled plasmid DNA,” *Proceedings of the National Academy of Science USA* vol. 70, no. 12, pp. 3744-3748, 1973.

Winston F, Dollard C, and Ricupero-Hovasse SL “Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C,” *Yeast* vol. 11, no. 1, pp. 53-55, 1995.

Wu J, Zhang N, Hayes A, Panoutsopoulou K, and Oliver SG “Global analysis of nutrient control of gene expression in *Saccharomyces cerevisiae* during growth and starvation,” *Proceedings of the National Academy of Science USA* vol. 101, no. 9, pp. 3148-3153, 2004.

Zeikus JG, Jain MK, and Elankovan P “Biotechnology of succinic acid production and markets for derived industrial products,” *Applied Microbiology and Biotechnology* vol. 51, pp. 545-552, 1999.

Zelle RM, de Hulster E, van Winden WA, de Waard P, Dijkema C, Winkler AA, Geertman JM, van Dijken JP, Pronk JT, and van Maris AJ “Malic acid production by *Saccharomyces cerevisiae*: engineering of pyruvate carboxylase, oxaloacetate reduction, and malate export,” *Applied Environmental Microbiology* vol. 74, no. 9, pp. 2766-2777, 2008.

The following undergraduate and graduate theses were carried out as extensions of the doctoral research presented in **Papers I, II, and III**, and section 7.5 Results Not Included in Papers. In addition, two theses were supervised in conjunction with **Paper VIII**, not presented in this thesis. The theses are included here as references.

Scalcnati G: **Production of Succinic Acid by Xylose Consumption in *Saccharomyces cerevisiae*.** MSc Thesis, Dipartimento di Biotecnologie e Bioscienze, Universita' degli Studi di Milano Bicocca, October 2007. Supervisors: José M. Otero, Lisbeth Olsson.

Nestola P: **Evaluation of the Effect of High CO<sub>2</sub> Concentration on Succinic Acid Production in *Saccharomyces cerevisiae*.** MSc Thesis, Facolta di Scienze Matematiche, Fisiche e Naturali Corso di Laurea Magistrale in Biotecnologie Industriali, Universita' degli Studi di Torino, March 2008. Supervisors: José M. Otero, Lisbeth Olsson.

Poulsen SG: **Metabolically Engineering *Saccharomyces cerevisiae* for Production of Succinic Acid.** BSs Thesis, Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark, July 2007. Supervisors: José M. Otero, Lisbeth Olsson.

Ågren R: **Interrogation and Visualization of the *S. cerevisiae* genome-scale network.** MSc Thesis, Department of Applied Microbiology, Lund Technical University, November 2008. Supervisors: José M. Otero, Jens Nielsen.

Borch MM: **Biotransformation of Ferulic Acid.** BSc Thesis, Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark, June 2007. Supervisors: José M. Otero, Gianni Panagiotou, Lisbeth Olsson.

Papadakis EM: **Yeast Systems Biology Approaches in Nutraceutical Discovery and Development.** MSc Thesis, Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark, September 2008. Supervisors: José M. Otero, Gianni Panagiotou, Jens Nielsen.

# PAPER I

**Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory.**

**José Manuel Otero**, Donatella Cimini, Kiran R. Patil, Simon G. Poulsen, Lisbeth Olsson, Jens Nielsen

Submitted to *PLOS One*, 2009

**Title:** Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory

**Authors:** José Manuel Otero<sup>1,2,4</sup>, Donatella Cimini<sup>3</sup>, Kiran R. Patil<sup>2</sup>, Simon G. Poulsen<sup>2</sup>, Lisbeth Olsson<sup>1,2</sup>, Jens Nielsen<sup>1,2\*</sup>

1. Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-41296 Gothenburg, Sweden.
2. Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark DK-2800, Kgs. Lyngby, Denmark.
3. Second University of Naples, Department of Experimental Medicine, Naples, Italy.
4. Current Address: Vaccine & Biologics Process Development, Bioprocess Research & Development, Merck Research Labs, West Point, PA, USA.

\*Corresponding author:

Professor Jens Nielsen  
Chalmers University of Technology  
Department of Chemical and Biological Engineering, Systems Biology  
Kemivägen 10  
SE-41296 Gothenburg, Sweden  
Phone: +46 (0) 31 772 8633  
Fax: +46 (0) 31 772 3801  
[nielsenj@chalmers.se](mailto:nielsenj@chalmers.se)

**Keywords:** metabolic engineering/*Saccharomyces cerevisiae*/succinic acid/DNA microarray/directed evolution

## **Abstract**

*Saccharomyces cerevisiae* is the most well characterized eukaryote, the preferred microbial cell factory for the largest industrial biotechnology product (bioethanol), and a robust commercially compatible scaffold to be exploited for diverse chemical production. Succinic acid is a highly sought after added-value chemical for which there is no native pre-disposition for production and accumulation in *S. cerevisiae*.

The genome-scale metabolic network reconstruction of *S. cerevisiae* enabled *in silico* gene deletion predictions using an evolutionary programming method to couple biomass and succinate production. Glycine and serine, both essential amino acids required for biomass formation, are formed from both glycolytic and TCA cycle intermediates. Succinate formation results from the isocitrate lyase catalyzed conversion of isocitrate, and from the  $\alpha$ -keto-glutarate dehydrogenase catalyzed conversion of  $\alpha$ -keto-glutarate. Succinate is subsequently depleted by the succinate dehydrogenase complex. The metabolic engineering strategy identified included deletion of the primary succinate consuming reaction, Sdh3p, and interruption of glycolysis derived serine by deletion of 3-phosphoglycerate dehydrogenase, Ser3p/Ser33p. Pursuing these targets, a multi-gene deletion strain was constructed, and directed evolution with selection used to identify a succinate producing mutant. Physiological characterization coupled with integrated data analysis of transcriptome data in the metabolically engineered strain were used to identify 2<sup>nd</sup>-round metabolic engineering targets. The resulting strain represents a 30-fold improvement in succinate titer, and a 43-fold improvement in succinate yield on biomass, with only a 2.8-fold decrease in the specific growth rate compared to the reference strain.

Intuitive genetic targets for either over-expression or interruption of succinate producing or consuming pathways, respectively, do not lead to increased succinate. Rather, we demonstrate how systems biology tools coupled with directed evolution and selection allows non-intuitive, rapid and substantial re-direction of carbon fluxes in *S. cerevisiae*, and hence shows proof of concept that this is a potentially attractive cell factory for over-producing different platform chemicals.

## Introduction

Industrial biotechnology is a promising alternative to traditional petrochemical production of chemicals focused on developing commercially sustainable and environmentally favorable processes (Otero *et al*, 2007). Metabolic engineering, the directed genetic modification of cellular reactions, aims to change the metabolic architecture of microorganisms to efficiently produce target chemicals (Nielsen *et al*, 2001). Although examples of metabolic engineering successes exist, there has yet to be developed a pipeline where preferred industrial hosts are rapidly engineered to produce a non-native accumulating target metabolite. Recent advances in systems biology has enabled *in silico* genome-scale metabolic network reconstructions to guide metabolic engineering strategies (Otero *et al*, 2007; Covert *et al*, 2001; Nielsen *et al*, 2008). Here we describe a pipeline where a microbial strain was constructed, physiologically characterized, and genomic tools were used to verify the predictions. This approach was repeated and complemented with traditional directed evolution and selection until a proof of concept microbial cell factory was reached. This pipeline resulted in the construction of a non-intuitive *Saccharomyces cerevisiae* cell factory over-producing succinic acid, a building block chemical.

*S. cerevisiae* is the most well characterized eukaryote and is unique in its broad application as an industrial production platform for a large portfolio of products including foods and beverages, bioethanol, vaccines, and therapeutic proteins (Otero *et al*, 2007). Many systems biology tools, including high-throughput genome sequencing, transcriptional profiling, metabolomics, carbon flux estimations, proteomics, *in silico* genome-scale modeling, and bioinformatics driven data integration were first applied to *S. cerevisiae* (Nielsen *et al*, 2008). Metabolic engineering has benefited from each of these tools; however, relatively few examples exist where cumulative integration has resulted in a generalized pipeline, in particular for the production of a target compound that the organism does not accumulate significantly naturally.

Succinic acid, systematically identified as butanedioic acid ( $\text{pK}_1 4.21$ ,  $\text{pK}_2 5.72$ ), is an added-value chemical building block, with an estimated 15,000 t/year world-wide demand predicted to expand to commodity chemical status with 270,000 t/year (McKinlay *et al*, 2007; Wilke *et al*, 2004), representing a potential >2 billion USD annual market. There are several elegant examples of bio-based production of succinate in *Anaerobiospirillum succiniciproducens*, *Actinobacillus succinogenes*, *Succinivibrio dextrinosolvens*, *Corynebacterium glutanicum*, *Prevotella ruminocola*, a recently isolated bacterium from bovine rumen, *Mannheimia succiniciproducens*, and a metabolically engineered succinic acid over-producing *E. coli* (McKinlay *et al*, 2007; Wilke *et al*, 2004; Zeikus *et al*, 1999; Song *et al*, 2006; Jantama *et al*, 2008; Lee *et al*, 2002). All of the hosts described are prokaryotic that grow at neutral pH, and consequently secrete the salt, succinate, requiring a cost-intensive acidification and precipitation to reach the desired succinic acid. This concern is not specific to succinic acid production, but rather universal when considering organic acid producing microbial cell factories (Sauer *et al*, 2008). *S. cerevisiae* represents a well-established, generally regarded as safe, robust, scalable (1L to 100,000L) industrial production host capable of growth on diverse carbon sources, chemically defined medium, both aerobic and anaerobic, and a wide pH operating range (3.0-6.0). Unlike the hosts described above, succinate is not naturally produced by *S. cerevisiae*; but as there are many factors of importance for the choice of a microbial cell factory it is not uncommon that the chosen cell factory lacks predisposition to produce the target chemical of choice (Adrio *et al*, 2006). As industrial biotechnology progresses forward and the concept of biorefineries are gaining increased importance, platform technologies including microbial cell factories that can plug-and-play into existing infrastructures must be developed (Lynd *et al*, 1999). *S. cerevisiae* is uniquely positioned as a

platform technology as it is already used widely for bioethanol production, but also because of the extensive library of genetic engineering tools, a very well annotated genome, many omics tools available, and well established complimentary approaches for directed evolution and selection. We therefore addressed the question whether it is possible to metabolically engineer *S. cerevisiae* such that the carbon fluxes are redirected towards succinic acid, and hereby establish proof-of-concept of using this yeast as a general cell factory platform for chemical production. The final strain emerging from this study requires significant further metabolic engineering and process development prior to consideration for commercialization, but the approach and integration of methods demonstrated supports the hypothesis that highly regulated central carbon metabolism in ideal production hosts can be reconfigured to produce target chemicals, relatively quickly and with minimal resources.

## Results

The genome-scale metabolic network reconstruction of *S. cerevisiae* permitted *in silico* prediction of gene deletions using an evolutionary programming method to couple biomass and succinate production (Patil *et al*, 2005). These results guided the metabolic engineering strategy described in Figure 1. Glycine, serine, and threonine, all representing essential amino acids required for biomass formation, may be formed from both glycolytic and tricarboxylic acid cycle intermediates. Succinate formation results from the isocitrate lyase, Icl1p, catalyzed conversion of isocitrate to equimolar glyoxylate and succinate, and from the  $\alpha$ -keto-glutarate dehydrogenase complex, Kgd1p/Kgd2p/Lpd1p, catalyzed conversion of  $\alpha$ -keto-glutarate to equimolar succinate, with a net production of CO<sub>2</sub>, NADH, and ATP. Succinate is subsequently depleted by the succinate dehydrogenase complex, Sdh1p/Sdh2p/Sdh3p/Sdh4p to equimolar fumarate with the net production of protonated ubiquinone. The metabolic engineering strategy identified included deletion of the primary succinate consuming reaction encoded by *Sdh3* (cytochrome b subunit of the succinate dehydrogenase complex, essential for function), and interruption of glycolysis derived serine by deletion of 3-phosphoglycerate dehydrogenase, Ser3p/Ser33p (isoenzymes). The remaining pathway for serine synthesis must originate from glycine, and glycine synthesis is largely derived from the alanine:pyruvate aminotransferase, Agx1p, converting glyoxylate and alanine to glycine and pyruvate. With this strategy, glycine and serine biomass requirements are directly coupled to succinate formation via the glyoxylate cycle. Substantial succinate accumulation (defined as >0.1 g L<sup>-1</sup>) in the culture broth is not observed in wild-type *S. cerevisiae*, and deletion of *sdh3* has not resulted in appreciable succinate accumulation (Cimini *et al*, 2009); a conclusion also found by chemical inhibition of the succinate dehydrogenase complex with titration of malonate (Supplementary Information 1), a chemical inhibitor of this complex (Aliverdieva *et al*, 2006).

The mutant resulting from the *in silico* strategy, referred to as 8D ( $\Delta$ *sdh3*  $\Delta$ *ser3*  $\Delta$ *ser33*), required supplementation with 500 mg L<sup>-1</sup> glycine to be able to grow. When evaluated in well controlled, aerobic, batch stirred tank reactors supplemented with 20 g L<sup>-1</sup> glucose in chemically defined medium, it exhibited a 13-fold improvement in succinate secreted titer (0.03 v 0.40 g L<sup>-1</sup>), 14-fold improvement in succinate biomass yield (0.01 v 0.14 g-succinate g-biomass<sup>-1</sup>), and a modest 33% reduction in the specific growth rate. Thus, the *in silico* guided metabolic engineering strategy was shown to work, representing a proof-of-concept of the use of model guided metabolic engineering. However, in order to obtain a prototrophic strain directed evolution was employed to screen and select for 8D mutants that did not require glycine supplementation. Specifically, repeated shake flask cultivation and transfer in declining glycine concentration supplemented medium, from an initial 500 mg L<sup>-1</sup> to 0 mg L<sup>-1</sup> in six increments (see Figure 2) was performed. The resulting strain demonstrated a 7.7-fold improvement in succinate yield on biomass (0.09 v 0.69 g-succinate g-biomass<sup>-1</sup>), strongly suggesting the direct coupling of glycine formation from glyoxylate and succinate formation. The resulting strain had a relatively low specific growth rate, 0.03h<sup>-1</sup>, and was therefore subsequently cultivated in repeated shake flasks and transferred across six shake flasks (only first three shake flasks shown in Figure 2) to improve the specific growth rate. Finally, a specific growth rate of 0.14h<sup>-1</sup> was reached, however, resulting in a decreased succinate yield on biomass (0.69 v 0.27 g-succinate g-biomass<sup>-1</sup>). The final strain, referred to as 8D Evolved, was shown to exhibit a 60-fold improvement in biomass coupled succinate production (0.01 v 0.30 g-succinate g-biomass<sup>-1</sup>), and 20-fold improvement in succinate titer (0.03 v 0.60 g L<sup>-1</sup>) relative to the reference strain when grown in aerobic batch cultivations.

To investigate the apparent decoupling of succinate coupled biomass formation, and potentially identify second-round metabolic engineering strategies, the transcriptome was measured in aerobic, glucose-limited, mid-exponential phase grown batch cultivations of 8D Evolved and the reference strain. Continuous cultivations, both carbon-limited and nitrogen-limited chemostats were attempted with the 8D Evolved mutant; however, in both cases steady-state was not reached and wash-out occurred, even at relatively low dilution rates ( $D=0.03\text{h}^{-1}$  compared to  $\mu_{max} = 0.14\text{h}^{-1}$ ). It was expected that 8D Evolved would not support cultivation in carbon-limited continuous culture due to the down-regulation of the TCA cycle ( $\Delta sdh3$ ), and consequently, reduced capacity for respiratory metabolism and oxidative phosphorylation. Therefore, batch cultivations were employed acknowledging the significant differences in specific growth rate ( $0.33$  v  $0.13\text{h}^{-1}$ ), and glucose uptake rate ( $90$  v  $26\text{ C-mmol g-DCW}^{-1}\text{ h}^{-1}$ ), while maintaining relatively similar biomass yields ( $0.18$  v  $0.19\text{ C-mol biomass C-mol glucose}^{-1}$ ).

Several studies have shown that significant differences in specific growth rate directly impact transcriptome interpretation, with anywhere between 268 and 2400 genes classified as potentially growth-related (Regenberg *et al*, 2006; Castrillo *et al* 2007; Fazio *et al* 2008). Previously generated continuous cultivation transcriptome data for both carbon-limited (glucose, respiratory growth) and nitrogen-limited (ammonium sulfate, respiro-fermentative growth) conditions at dilution rates of  $0.03$ ,  $0.1$ , and  $0.2\text{ h}^{-1}$  were therefore used to identify statistically differentially expressed growth-related genes (Fazio *et al*, 2008). A total of 6 and 7 differentially expressed genes were identified within the carbon-limited and nitrogen-limited data sets as being growth-related ( $p-value_{B-H}<0.1$ ,  $n=3$  at each dilution rate), respectively, and a total of 66 differentially expressed genes were identified when comparing carbon-limited and nitrogen-limited data sets, paired at each dilution rate ( $p-value_{B-H}<0.1$ ,  $n=3$  at each dilution rate). Of the total 2406 differentially expressed genes between the 8D Evolved and reference strain ( $p-value_{B-H}<0.01$ ,  $|log-fold change|>0.5$ ,  $n=3$  biological replicates,  $n=2$  DNA microarray duplicates), 36 unique growth-related genes were identified suggesting that few of the genes with a significant change in transcription in 8D Evolved are due to changes in the specific growth rate (see Figure 3). However, a total of 8 of the top 20  $p-value_{B-H}$  ranked differentially expressed genes identified from pair-wise comparison of 8D Evolved and the reference strain, are growth-related genes (*ARO9*, *SER3*, *JLP1*, *HMALPHA1*, *ARO10*, *MFALPHA2*, and two uncharacterized genes, *YPL033c* and *YLR267w*).

The top 2000 (there were no metabolic genes in the remaining 406 genes nor were there any biological process annotations available as determined by gene ontology, and therefore they were not included in further analysis) differentially expressed genes were selected for further analysis, and after removal of the 36 growth-related genes, a list of 1964 genes was submitted for metabolic pathway visualization and characterization to the Expression Viewer (Paley *et al*, 2006) available at the Yeast Genome Database (Fisk *et al*, 2006) (see Figure 3). The log-fold change of the 8D Evolved:Reference expression ratio was mapped onto the metabolic map of *S. cerevisiae* strain S288c, version 12.0, composed of 140 pathways, 925 enzymatic reactions, and a total of 675 compounds (see Supplementary Information 2). A total of 315 genes mapped to a specific metabolic pathway on the expression viewer, with a mean log-fold expression ratio value of  $0.3 \pm 1.3$  ( $n=315$ ,  $\pm SD$ ).

Three biological insights were immediately apparent (see Supplementary Information 2). First, *SDH3*, *SER3*, and *SER33* had negative log-fold expression ratios (log-fold change  $<-8.0$ ) confirming the gene deletions targeted in the 8D strain and the maintained low expression through the directed evolution. Second, when examining the glycine, serine, and threonine metabolism, *AGX1*

was 4.3 log-fold change upregulated in the 8D Evolved strain, confirming significant upregulation of glycine synthesis from glyoxylate pools, as predicted by the original metabolic engineering strategy. However, there was no upregulation of *SHM2*, *SHM1*, the genes encoding pathways for L-serine formation from L-glycine pools. Most surprisingly *GLY1*, encoding threonine adolase, was significantly up-regulated (log-fold change 1.6). In the genome-scale metabolic network reconstructions of *S. cerevisiae* iFF708 and iND750, upon which the 8D metabolic engineering strategy is based, Gly1p encodes the reversible conversion of threonine to glycine and acetaldehyde (Förster *et al*, 2003; Duarte *et al*, 2004), leading to the prediction that threonine biosynthesis from glycolytic intermediates could be down-regulated, and provided for from glycine pools. This consequently leads to a greater biomass-coupled drive for glyoxylate synthesis from isocitrate, yielding equimolar succinate. Leveraging this over-all strategy, another *S. cerevisiae* mutant was constructed, referred to as 20G ( $\Delta$ *sdh3*,  $\Delta$ *ser3*,  $\Delta$ *thr1*), where Thr1p, encoding homoserine kinase that is required for threonine biosynthesis, was deleted. However, this strain required threonine supplementation and after several extensive attempts at adaptive evolution, the threonine auxotrophy persisted, suggesting the reversibility of *Gly1* was limited with the adolase strongly favoring glycine formation (Supplementary Information 3). The significant up-regulation of *Gly1* therefore provides a strong hypothesis for why 8D Evolved had an attenuation of succinate production, even under increasing specific growth rate, suggesting a decoupling of biomass coupled succinate production. It should be noted that in the most recent update of the genome-scale metabolic reconstruction of *S. cerevisiae*, iIN800, the reversibility of *Gly1* was corrected based on this data (Nookaew *et al*, 2008).

The transcriptome not only provides for a global, rapid, and quantitative assessment of the predicted *in silico* metabolic engineering strategy and insight into the genetic engineering modifications that result from directed evolution and selection, but also provides a source for identification of second round metabolic engineering targets not previously predicted. Several targets were identified, but of particular interest was *ICL1*, encoding isocitrate lyase, converting isocitrate to glyoxylate and succinate in equimolar concentrations. All tricarboxylic acid cycle genes are up-regulated, with the exception of *SDH3* (target gene deletion), and *ICL1*, providing a clear metabolic engineering target for up-regulation in the 8D Evolved strain. Therefore, native *ICL1* was PCR amplified and cloned into the 2  $\mu$ m ori plasmid pRS426CT containing the strong constitutive *TEF1* promoter and *CYC1* terminator (Wattanachaisaereekul *et al*, 2008), and then transformed into the reference, 8D, and 8D Evolved strain (strains transformed with the constructed plasmid pRS426T-*ICL1-C* are referred to as “with p*ICL1*”). All strains were evaluated in aerobic, glucose-supplemented batch fermentations, and only 8D Evolved with p*ICL1* exhibited a change in succinate production (see Figure 4). Specifically, the succinate titer, biomass yield, and glucose yield were 0.90 g L<sup>-1</sup>, 0.43 g-succinate g-biomass<sup>-1</sup>, and 0.05 g-succinate g-glucose<sup>-1</sup>, respectively, representing a 1.5-fold, 1.4-fold, 1.7-fold improvement over 8D, respectively (see Figure 4).

## Discussion

A *S. cerevisiae* strain capable of succinate production, requiring redirection of carbon flux from typically produced C<sub>2</sub> (ethanol, acetate) and C<sub>3</sub> (glycerol, pyruvate) over-flow metabolites to the target C<sub>4</sub> succinic acid was achieved through metabolic engineering, requiring integration of systems biology methods and directed evolution. Clearly, the resulting strain (8D Evolved with pICL1), representing a proof-of-concept, still requires significant process development and further enhancement to compete commercially with existing bacterial platforms.

The resulting strain, 8D Evolved with pICL1, represents a 30-fold improvement in succinate titer, and a 43-fold improvement in succinate yield on biomass, with only a 2.8-fold decrease in the specific growth rate compared to the reference strain. Despite success of using simple stoichiometric-based calculations for driving metabolic engineering, it is interesting to note that regulatory mechanisms not captured in these models are likely playing a significant role in the succinate production observed. The biomass requirements for glycine and serine are 0.290 and 0.185 mmol g-DCW<sup>-1</sup> (Förster *et al*, 2003). Assuming that all glycine, and all glycine and serine combined demands are supplied from the glyoxylate pool, then the theoretical production of succinate would amount to 0.034 and 0.056 g-succinate g-DCW<sup>-1</sup>, respectively. The 8D and 8D Evolved strains are producing 0.30 and 0.43 g-succinate g-biomass<sup>-1</sup>, respectively, suggesting a nearly 8-fold higher succinate production than required to meet biomass amino acid demands. A potentially 3<sup>rd</sup> metabolic engineering target would be deletion of *GLY1* to further minimize alternative biosynthetic routes of glycine production, thereby isolating all glycine production to be dependent on glyoxylate formation, and consequently succinate formation. Yet, it's clear that any increase in succinate formation would not be due to biomass requirements, but rather regulatory (e.g., non-stoichiometric driven) mechanisms. Therefore, while the strategy presented and demonstrated here is likely to be a major component of an over-all succinate production cell factory, complimentary strategies focusing on the other major succinate production pathway, TCA cycle, will be required. Examples of malic acid production, that included engineering of pyruvate carboxylation (overexpression of *PYC2*), oxaloacetate reduction (overexpression of *MDH3*), and malate export (functional expression of the non-native *SpMAE1*), resulted in a *S. cerevisiae* strain capable of producing 59 g-malate L<sup>-1</sup> and 0.42 mol malate mol-glucose<sup>-1</sup> (Zelle *et al*, 2008). A similar approach, requiring yet further engineering and understanding of the reductive TCA cycle to convert malate to succinate is likely required, but a major hurdle with this strategy is the conversion of fumarate to succinate by fumarate reductase which is thermodynamically favored in the direction of fumarate.

The transcriptome analysis performed, specifically consideration of continuous culture data sets at different dilution rates to filter growth-related genes was an integral part of identifying the 2<sup>nd</sup> round of metabolic engineering targets. Although a relatively small number of growth related genes were identified, they were of high-value. For example, *ARO9* and *ARO10*, encoding key enzymatic conversion steps in aromatic amino acid metabolism may have incorrectly pointed towards tryptophan, tyrosine, and phenylalanine catabolism or phosphoenolpyruvate decarboxylase activity as metabolic areas of interest for understanding physiological differences between 8D Evolved and the reference strains. This approach may be extended to future efforts and other organisms, where continuous cultivation of engineered strains may not be possible, as in this case, or applied industrially where the dominant and preferred processing mode is batch.

Furthermore, this work clearly demonstrated that obvious genetic targets did not result in increased succinate formation. Specifically, deletion of the primary succinate consuming pathway (*Δsdh3*) and constitutive over-expression of one of two of the primary succinate formation pathways

(*ICL1*) did not result in any increased succinate production (See Supplementary Information 1) (Cimini *et al*, 2009). It is further interesting to note that the 8D with pICL1 strain also did not result in any increased succinate production, but rather only in the 8D Evolved with pICL1 strain. The ability to measure transcriptome on a strain that underwent targeted genetic engineering and directed evolution was critical to identifying pICL1 as a 2<sup>nd</sup> metabolic engineering target, which would have been discarded if selected based on intuition.

The approach employed represents an integration of diverse methods for rapid metabolic engineering proof-of-concept. The strain selection process thus need not be limited to considering organisms showing a predisposition to the production of the metabolite of interest, but rather, should include hosts most suitable for large-scale, robust, and biorefinery processing. With such hosts, carbon and redox flux redistribution requiring multi-gene approaches can be predicted, tested, and supplemented with directed evolution, screening, and selection. These strains are then genomically characterized and optimized until commercially viable titers, productivities, and yields are reached. It is only through whole-process optimization and elimination of severe constraints such as forced use of non-industrially favorable strains, that the promise of a bio-based economy may be fully realized.

## Materials and Methods

### Strain Construction

The reference strain *Saccharomyces cerevisiae* CEN.PK113-5D (*Mat a* *MAL2-8C* *SUC2* *URA3-52*) (van Dijken *et al*, 2000) was used for the construction of the *Δsdh3 Δser3 Δser33* knockout strain, referred to as the 8D mutant, and for the construction of the *Δsdh3 Δser3 Δthr1* knockout strain, referred to as the 20G mutant, through the cloning-free PCR-based allele replacement method previously described (Eredeniz *et al*, 1997). The upstream *SDH3* fragment was amplified by PCR from genomic DNA using the primers *SDH3\_Up\_Fw* (sequence 5'-CGAAATATGGTAAGAGAAAATG-3') and *SDH3\_Up\_Rv* (sequence 5'- CAGGGATGC GGCGCTGACGACATCG TTTATTATTCTTAGAGC-3'). Similarly, the downstream *SDH3* fragment was amplified using the primers *SDH3\_Dw\_Fw* (sequence 5'- CCGCTGCTAGGCGCGCCGTGCTTATGATTCTTAAGGCGACGC-3') and *SDH3\_Dw\_Rv* (sequence 5'- GTAATCTGTTATCGATAATCTGCC -3'). The upstream *THR1* fragment was amplified by PCR from genomic DNA using the primers *THR1\_Up\_Fw* (sequence 5'-GCAGTTC TTGCTCAGTAATCTTAG-3') and *THR1\_Up\_Rv* (sequence 5'-GCAGGGATGC GGCGCTGACCCATA TCTTCGAGATGATGACTC-3'). Similarly, the downstream *THR1* fragment was amplified using the primers *THR1\_Dw\_Fw* (sequence 5'-CCGCTGCTAGGCGCGCCGTGCATACTGTAATTGACC GTTAACGG-3') and *THR1\_Dw\_Rv* (sequence 5'-CCAATCATGGATGAACCAGTAATG-3'). The upstream *SER3* fragment was amplified by PCR from genomic DNA using the primers *SER3\_Up\_Fw* (sequence 5'- CTCACAATCGAGTAA TGCCTTG-3') and *SER3\_Up\_Rv* (sequence 5'- GCAGGGATGC GGCGCTGACCATTGCTGTCGA TTTTCTGTGG-3'). Similarly, the downstream *SER3* fragment was amplified using the primers *SER3\_Dw\_Fw* (sequence 5'- CCGCTGCTAGGCGCGCCGTGGGATAGAAGAATGCTTGAGGC-3') and *SER3\_Dw\_Rv* (sequence 5'- CGAATTGATTGTACCTGGTGC-3'). The upstream *SER33* fragment was amplified by PCR from genomic DNA using the primers *SER33\_Up\_Fw* (sequence 5'- GTACTCTTATGGGAGTCTTAGC -3') and *SER33\_Up\_Rv* (sequence 5'- GCAGGGATGC GGCGCTGACGCAGCTGAATAAGACATGTTAGG- 3'). Similarly, the downstream *SER33* fragment was amplified using the primers *SER33\_Dw\_Fw* (sequence 5'- GCAGGGATGC GGCGCTGACGCAGCTGAATAAGACATGTTAGG- 3') and *SER33\_Dw\_Rv* (sequence 5'-CTATT CTGGGTGGTCTTTACTGG- 3'). The lithium acetate transformation method was used (Gietz *et al*, 2002). As described previously, *URA3* from *Kluyveromyces lactis* was used as the selection marker in the transformation process (Eredeniz *et al*, 1997). With this approach transformants are easily selected on uracil depleted media supplemented with 5-fluoroorotic acid. The knockout was confirmed by restriction analysis followed by sequencing (MWG Biotech AG, Ebersberg, Germany).

The plasmid pRS426T-*ICL1-C* was constructed and transformed into 8D Evolved, described earlier and used for constitutive *S. cerevisiae* *ICL1* overexpression. The parent plasmid, pRS426CT (6347 bp), was previously constructed in our laboratory by inserting the strong constitutive *TEF1* promoter (gene encoding *S. cerevisiae* translation-elongation factor 1 $\alpha$ ) and the *CYC1* transcription terminator into pRS426 (Wattanachaisaereekul *et al*, 2008). This original backbone plasmid is a 5726 bp yeast episomal plasmid (YEp)-type shuttle vector with a high copy number of about 20 per cell (Christianson *et al*, 1992). The plasmid contains the 2  $\mu$ m ori and pUC ori for independent episomal replication in *S. cerevisiae* and *E. coli*, respectively, and *URA3* and *ampR* (*bla*, beta-lactamase) genes. The final plasmid size was 8074 bp, with 2484 containing the *TEF1* promoter, the *ICL1* insert, and the *CYC1* transcription terminator sequence, verified by sequencing (MWG Biotech AG, Ebersberg, Germany).

A total of eight primers were required for amplification of the native *ICL1* gene from the reference strain, sequencing of the constructed plasmid pRS426-*ICL1-C*, and PCR to verify plasmid presence in the transformed reference and 8D Evolved strains (referred to as 8D Evolved with p*ICL1*). The PCR amplification of *ICL1* was carried out using the Phusion™ High-Fidelity DNA Polymerase (Finnzymes Oy, Espoo, Finland) according to the manufacturer's protocol. The native *ICL1* was amplified from genomic DNA using the up- and downstream primers *ICL1\_Sp1* (sequence 5'-GCCTGCCA|CTAGTCAACGAAAAATGCCTATCCCCG-3'), and *ICL1\_Asp1* (sequence 5'-GCCTCGACCCGGGCTAGAGAAAGGCATTCTGCACGG-3'), respectively. The amplicon length was 1915 bp. The fragment was cut with restriction endonucleases (REN) *SpeI*, the restriction site of which was *de novo* introduced on primer *ICL1\_Sp1*, and *NgoMIV*, and then ligated with pRS426CT cut with *SpeI* and *XmaI*. By using the non-compatible RENs in either end of the insertion, the direction of the insert is secured and furthermore the sole parent plasmid *Xma* site is lost. This allowed for an *in vitro* pre-selection for the correct pRS426-*ICL1-C* construct prior to transformation.

The four sequencing primers for construct verification included *M13\_rev\_-29* (sequence 5'-CAGGAAACAGCTATGACC-3'), *ICL1\_In\_1f* (sequence 5'-CTGGTTGGCAGTGTCATCA-3'), *ICL1\_In\_2f* (sequence 5'-CATCCCACAGAGAACCCAAG-3'), and *M13\_uni\_-21* (sequence 5'-TGTAACAGACGGCCAGT-3'). The two primers used for plasmid verification via PCR (Taq DNA Polymerase of *Thermus aquaticus* from Sigma, St. Louis, MO, were *ICL1\_part\_Sense* (sequence 5'-TCCTGTTAGATTCTCAAATGGC-3') and *ICL1\_CYC\_Antisense* (sequence 5'-AAATTAAAGCCTTCGAGCGTCCC-3') and these were used for analytical PCRs according to the instruction manual's recommendations). Plasmid transformation of electrocompetent *E. coli* DH5 $\alpha$  were completed as described previously, as was plasmid transformation of the *S. cerevisiae* reference strain and 8D Evolved using the lithium acetate method (Wattanachaisaereekul *et al*, 2008; Eredeniz *et al*, 1997; Gietz *et al*, 2002).

### **Medium Formulation**

A chemically defined minimal medium of composition 5.0 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 3.0 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub>•7H<sub>2</sub>O, 1.0 mL L<sup>-1</sup> trace metal solution, 300 mg L<sup>-1</sup> uracil, 0.05 g L<sup>-1</sup> antifoam 204 (Sigma-Aldrich A-8311), and 1.0 mL L<sup>-1</sup> vitamin solution was used for all shake flask and 2L well-controlled fermentations (Verudyn *et al*, 1992). The trace element solution included 15 g L<sup>-1</sup> EDTA, 0.45 g L<sup>-1</sup> CaCl<sub>2</sub>•2H<sub>2</sub>O, 0.45 g L<sup>-1</sup> ZnSO<sub>4</sub>•7H<sub>2</sub>O, 0.3 g L<sup>-1</sup> FeSO<sub>4</sub>•7H<sub>2</sub>O, 100 mg L<sup>-1</sup> H<sub>3</sub>BO<sub>4</sub>, 1 g L<sup>-1</sup> MnCl<sub>2</sub>•2H<sub>2</sub>O, 0.3 g L<sup>-1</sup> CoCl<sub>2</sub>•6H<sub>2</sub>O, 0.3 g L<sup>-1</sup> CuSO<sub>4</sub>•5H<sub>2</sub>O, 0.4 g L<sup>-1</sup> NaMoO<sub>4</sub>•2H<sub>2</sub>O. The pH of the trace metal solution was adjusted to 4.00 with 2M NaOH and heat sterilized. The vitamin solution included 50 mg L<sup>-1</sup> d-biotin, 200 mg L<sup>-1</sup> para-amino benzoic acid, 1 g L<sup>-1</sup> nicotinic acid, 1 g L<sup>-1</sup> Ca-pantothenate, 1 g L<sup>-1</sup> pyridoxine HCl, 1 g L<sup>-1</sup> thiamine HCl, and 25 mg L<sup>-1</sup> m-inositol. The pH of the vitamin solution was adjusted to 6.5 with 2M NaOH, sterile-filtered and the solution was stored at 4°C. The final formulated medium, excluding glucose and vitamin solution supplementation, is adjusted to pH 5.0 with 2M NaOH and heat sterilized. For carbon-limited cultivations the sterilized medium is supplemented with 20 g L<sup>-1</sup> glucose, heat sterilized separately, and 1.0 mL L<sup>-1</sup> vitamin solution is added by sterile filtration (0.20 μm pore size Ministart®-Plus Sartorius AG, Goettingen, Germany). For cultures where glycine or threonine auxotrophic strains were cultivated the final culture medium was supplemented with glycine 500 mg L<sup>-1</sup> or 100 mg L<sup>-1</sup> threonine added by sterile filtration.

### **Shake Flask Cultivations and Stirred Tank Fermentations**

Shake flask cultivations were completed in 500 mL Erlenmeyer flasks with two diametrically opposed baffles and two side-necks with septums for sampling by syringe. Flasks were heat

sterilized with 100 mL of medium, inoculated with a single colony, and incubated at 30°C with orbital shaking at 150 RPM. Stirred tank fermentations were completed in well-controlled, aerobic, 2.2L Braun Biotech Biostat B fermentation systems with a working volume of 2L (Sartorius AG, Goettingen, Germany). The temperature was controlled at 30°C. The fermenters were outfitted with two disk-turbine impellers rotating at 600 RPM. Dissolved oxygen was monitored with an autoclavable polarographic oxygen electrode (Mettler-Toledo, Columbus, OH). During aerobic cultivation the air sparging flow rate was 2 vvm. The pH was kept constant at 5.0 by automatic addition of 2M KOH. Off-gas passed through a condenser to minimize the evaporation from the fermenter. The fermenters were inoculated from shake flask precultures to an initial OD<sub>600</sub> 0.01.

### **Fermentation Analysis**

**Off-gas Analysis:** The effluent fermentation gas was measured every 30 seconds for determination of O<sub>2</sub>(g) and CO<sub>2</sub>(g) concentrations by the off-gas analyzer Brüel and Kjær 1308 (Brüel & Kjær, Nærum, Denmark).

**Biomass Determination:** The optical density (OD) was determined at 600 nm using a Shimadzu UV mini 1240 spectrophotometer (Shidmazu Europe GmbH, Duisberg, Germany). Duplicate samples were diluted with deionized water to obtain OD<sub>600</sub> measurements in the linear range of 0-0.4 OD<sub>600</sub>. Samples were always maintained at 4°C post-sampling until OD<sub>600</sub> and dry cell weight (DCW) measurements were performed. DCW measurements were determined through the exponential phase, until stationary phase was confirmed according to OD<sub>600</sub> and off-gas analysis. Nitrocellulose filters (0.45 µm Sartorius AG, Goettingen, Germany) were used. The filters were pre-dried in a microwave oven at 150W for 10 min., and cooled in a dessicator for 10 min. 5.0 mL of fermentation broth were filtered, followed by 10 mL DI water. Filters were then dried in a microwave oven for 20 min. at 150W, cooled for 15 min. in a dessicator, and the mass was determined.

**Metabolite Concentration Determination:** All fermentation samples were immediately filtered using a 0.45 µm syringe-filter (Sartorius AG, Goettingen, Germany) and stored at -20°C until further analysis. Glucose, ethanol, glycerol, acetate, succinate, pyruvate, fumarate, citrate, oxalate, and malate were determined by HPLC analysis using an Aminex HPX-87H ion-exclusion column (Bio-Rad Laboratories, Hercules, CA). The column was maintained at 65°C and elution performed using 5 mM H<sub>2</sub>SO<sub>4</sub> as the mobile phase at a flow rate of 0.6 mL min.<sup>-1</sup>. Glucose, ethanol, glycerol, acetate, succinate, citrate, fumarate, malate, oxalate were detected on a Waters 410 differential refractometer detector (Shodex, Kawasaki, Japan), and acetate and pyruvate were detected on a Waters 468 absorbance detector set at 210 nm.

### **Transcriptomics**

**RNA Sampling and Isolation:** Samples for RNA isolation from the late-exponential phase of glucose-limited batch cultivations were taken by rapidly sampling 25 mL of culture into a 50 mL sterile Falcon tube with 40 mL of crushed ice in order to decrease the sample temperature to below 2°C in less than 10 seconds. Cells were immediately centrifuged (4000 RPM at 0°C for 2.5 min.), the supernatant discarded, and the pellet frozen in liquid nitrogen and it was stored at -80°C until total

RNA extraction. Total RNA was extracted using the FastRNA Pro RED kit (QBiogene, Carlsbad, USA) according to manufacturer's instructions after partially thawing the samples on ice. RNA sample integrity and quality was determined prior to hybridization with an Agilent 2100 Bioanalyzer and RNA 6000 Nano LabChip kit according to the manufacturer's instruction (Agilent, Santa Clara, CA).

**Probe Preparation and Hybridization to DNA Microarrays:** Messenger RNA (mRNA) extraction, cDNA synthesis, labeling, and array hybridization to Affymetrix Yeast Genome Y2.0 arrays were performed according to the manufacturer's recommendations (Affymetrix GeneChip® Expression Analysis Technical Manual, 2005-2006 Rev. 2.0). Washing and staining of arrays were performed using the GeneChip Fluidics Station 450 and scanning with the Affymetrix GeneArray Scanner (Affymetrix, Santa Clara, CA).

**Microarray Gene Transcription Analysis:** Affymetrix Microarray Suite v5.0 was used to generate CEL files of the scanned DNA microarrays. These CEL files were then processed using the statistical language and environment R v5.3 (R Development Core Team, 2007, [www.r-project.org](http://www.r-project.org)), supplemented with Bioconductor v2.3 (Bioconductor Development Core Team, 2008, [www.bioconductor.org](http://www.bioconductor.org)) packages Biobase, affy, gcrma, and limma (Smyth, 2005; Smyth, 2004). The probe intensities were normalized for background using the robust multiarray average (RMA) method only using perfect match (PM) probes after the raw image file of the DNA microarray was visually inspected for acceptable quality. Normalization was performed using the qspline method and gene expression values were calculated from PM probes with the median polish summary. Statistical analysis was applied to determine differentially expressed genes using the limma statistical package. Moderated *t*-tests between the sets of experiments were used for pair-wise comparisons. Empirical Bayesian statistics were used to moderate the standard errors within each gene and Benjamini-Hochberg's method was used to adjust for multi-testing. A cut-off value of adjusted *p*<0.05 was used for statistical significance. Furthermore, principal component analysis (PCA) was performed in order to elucidate the relative importance of substrate limitation (carbon vs. nitrogen) and growth rate (0.03 h<sup>-1</sup>, 0.1 h<sup>-1</sup>, 0.2h<sup>-1</sup>), previously described (Fazio *et al*, 2008), when compared with the gene expression of the reference and 8D Evolved strain. To select genes whose expression levels were related to these factors, the moderated *t*-statistics were followed up with *F*-distributions to yield a statistic referred to as *F<sub>g</sub>*, which is simply the usual *F*-statistic from linear model theory but with the posterior variance substituted for the sample variance in the denominator, as described elsewhere (Smyth, 2004). The cut-off value of adjusted *p*<0.1 was used for statistical significance.

All microarray data is MIAME compliant and the raw data has been deposited in ArrayExpress (<http://www.ebi.ac.uk/microarray-as/ae/>).

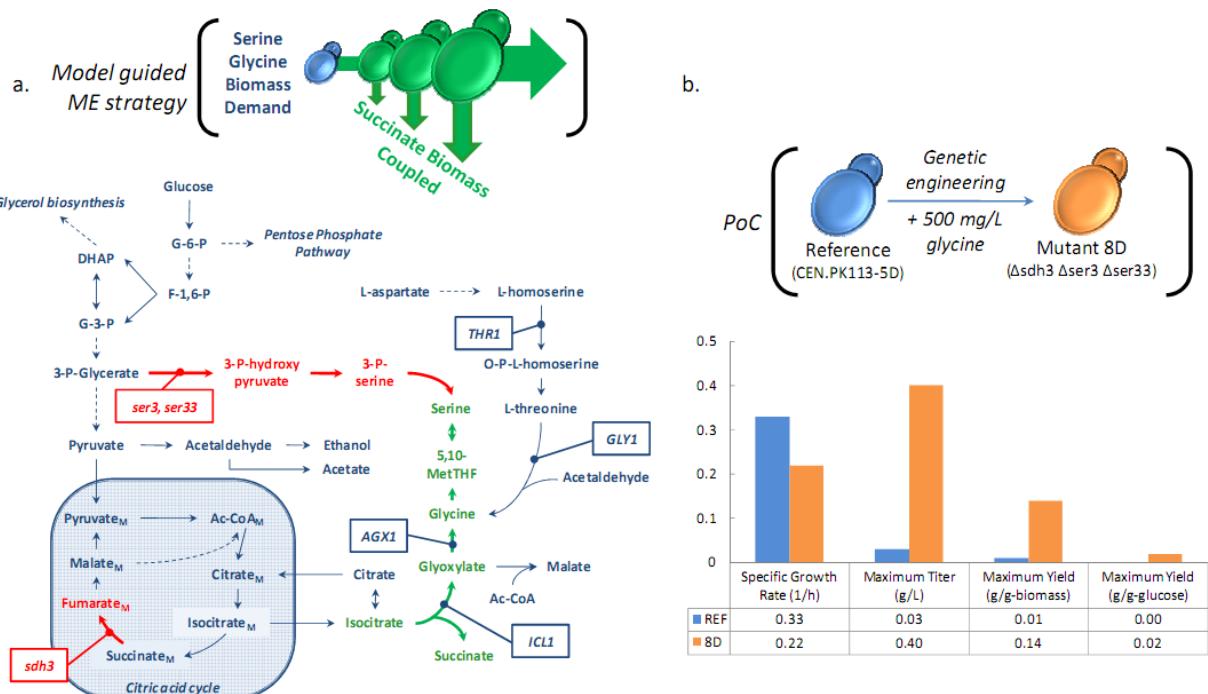
**Acknowledgements:** We are grateful to Gaëlle Lettier, Uffe Mortensen, Songsak Wattanachaisaereekul, Mohammad Asadollahi, and Jérôme Maury of the Center for Microbial Biotechnology, Technical University of Denmark, for their experimental assistance with yeast strain construction, plasmid construction, and yeast genetics methods. This work was in part supported by a Merck Doctoral Fellowship to J.M.O. from Merck Research Labs, Merck & Co., Inc.

**Author's Contributions:** JMO, DC, KRP, LO, and JN conceived and designed the experiments. JMO, DC, SGP performed the experimental work. JMO and JN wrote the manuscript. DC, KRP, SGP, LO, and JN reviewed the manuscript.

## References

1. Otero JM, Panagiotou G, Olsson L (2007) Fueling Industrial Biotechnology Through Bioethanol. *Adv. Biochem. Eng. Biotechnol.* **108**:1-40.
2. Nielsen J (2001) Metabolic engineering. *Appl. Microbiol. Biotechnol.* **55**(3):263-83.
3. Covert MW, Schilling CH, Famili I, Edwards JS, Goryanin II, Selkov E, Palsson BO (2001) Metabolic modeling of microbial strains *in silico*. *Trends Biochem Sci.* **26**(3):179-86.
4. Nielsen J, Jewett MC (2008) Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **8**(1):122-31.
5. McKinlay JB, Vieille C, Zeikus JG (2007) Prospects for a bio-based succinate industry. *Appl. Microbiol. Biotechnol.* **76**(4):727-40.
6. Wilke T, Vorlop KD (2004) Industrial bioconversion of renewable resources as an alternative to conventional chemistry. *Appl. Microbiol. Biotechnol.* **66**(2):131-42.
7. Zeikus JG, Jain MK, Elankovan P (1999) Biotechnology of succinic acid production and markets for derived industrial products. *Appl. Microbiol. Biotechno.* **51**:545-52.
8. Song H, Lee SY (2006) Production of succinic acid by bacterial fermentation. *Enzy. Microb. Techn.* **39**(3):352-61.
9. Jantama K, Haupt MJ, Svoronos SA, Zhang X, Moore JC, Shanmugam KT, Ingram LO (2008) Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* C that produce succinate and malate. *Biotechnol. Bioeng.* **99**(5):1140-53.
10. Lee PC, Lee SY, Hong SH, Chang HN (2002) Isolation and characterization of a new succinic acid-producing bacterium, *Mannheimia succiniciproducens* MBEL55E, from bovine rumen. *Appl. Microbiol. Biotechnol.* **58**(5):663-68.
11. Sauer M, Porro D, Mattanovich D, Braunduardi P (2008) Microbial production of organic acids: expanding the markets. *Trends Biotechnol.* **26**(2): 100-8.
12. Adrio JL, Deman AL (2006) Genetic improvement of processes yielding microbial products. *FEMS Microbiol. Rev.* **30**(2): 187-214.
13. Lynd LR, Wyman CE, Gerngross TU (1999) Biocommodity Engineering. *Biotechnol. Prog.* **15**(5):777-93.
14. Patil KR, Rocha I, Förster J, Nielsen J (2005) Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics.* **6**:308.
15. Cimini D, Patil KR, Schiraldi C, Nielsen J (2009) Global transcriptional response of *Saccharomyces cerevisiae* to the deletion of SDH3. *BMC Systems Biology* **3**(1):17.
16. Aliverdieva DA, Mamaev DV, Lagutina LS, Scholtz KF (2006) Specific features of changes in levels of endogenous respiration substrates in *Saccharomyces cerevisiae* cells at low temperature. *Biochemistry (Moscow).* **71**(1):39-45.
17. Regenberg B, Grotkær T, Winther O, Fausbøll A, Akesson M, Bro C, Hansen LK, Brunak S, Nielsen J (2006) Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in *Saccharomyces cerevisiae*. *Genome Biol.* **7**(11):R107.
18. Castrillo JI, Zeef LA, Hoyle DC, Zhang N, Hayes A, Gardner DC, Cornell MJ, Petty J, Hakes L, Wardleworth L, Rash B, Brown M, Dunn WB, Broadhurst D, O'Donoqhue K, Hester SS, Dunkley TP, Hart SR, Swainston N, Li P, Gaskell SJ, Paton NW, Lilley KS, Kell DB, Oliver SG (2007) Growth control of the eukaryote cell: a systems biology study in yeast. *J. Biol.* **6**(2):4.
19. Fazio A, Jewett MC, Daran-Lapujade P, Mustacchi R, Usaite R, Pronk JT, Worman CT, Nielsen J (2008) Transcription factor control of growth rate dependent genes in *Saccharomyces cerevisiae*: A three factor design. *BMC Genomics.* **9**:341.

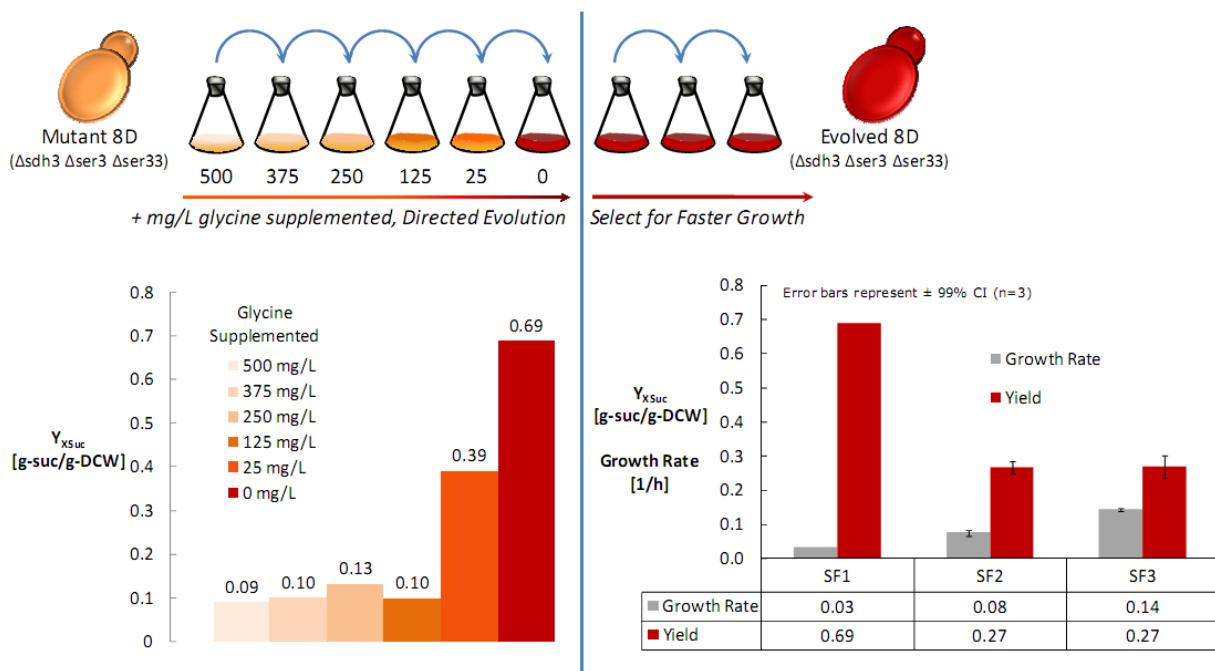
20. Paley SM, Karp PD (2006) The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acids Research*. **34**(13):3771-778.
21. Fisk DG, Ball CA, Dolinski K, Engel SR, Hong EL, Issel-Tarver L, Schwartz K, Sethuraman A, Botstein D, Cherry JM, *Saccharomyces* Genome Database Project (2006) *Saccharomyces cerevisiae* S288C genome annotation: a working hypothesis. *Yeast*. **23**(12): 857-65.
22. Förster J, Famili I, Fu P, Palsson BØ, Nielsen J (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res*. **13**(2): 244-53.
23. Duarte NC, Harrgård MJ, Palsson BØ (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res*. **14**(7): 1298-309.
24. Nookaew I, Jewett MC, Meechai A, Thammarongtham C, Laoteng K, Cheevadhanarak S, Nielsen J, Bhumiratana S (2008) The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Systems Biology*. **7**(2): 71.
25. Wattanachaisaereekul S, Lantz AE, Nielsen ML, Nielsen J (2008) Production of the polyketide 6-MSA in yeast engineered for increased malonyl-CoA supply. *Metab Eng* **10**(5): 246-254.
26. Eredeniz N, Mortensen UH, Rothstein R (1997) Cloning-free PCR-based allele replacement methods. *Genome Res* **7**:1174-1183.
27. Gietz RD, Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol*. **350**: 87-96.
28. Verudyn 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.
29. Christianson TW, Sikorski RS, Dante M, Sher JH, Hieter P (1992) Multifunctional yeast high-copy-number shuttle vectors. *Gen* **110**(1): 119-122.
30. Zelle RM, de Hulster E, van Winden WA, de Waard P, Dijkema C, et al. (2008) Malic acid production by *Saccharomyces cerevisiae*: Engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. *Appl Environ Microbiol* **74**(9): 2766-77.
31. van Dijken JP, Bauer J, Brambilla L, Duboc P, Francois JM, Gancedo C, Giuseppin ML, Heijnen JJ, Hoare M, Lange H, Madden EA, Niederberger P, Nielsen J, Parrou JL, Petit T, Porro D, Reuss M, van Riel N, Rizzi M, Steensma HY, Verrips CT, Vindeløv J, Pronk JT (2000) An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme and Microbial Technology* **26**(9-10): 706-714.
32. Smyth GK (2005) Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W (eds.), pp 397-420, New York: Springer.
33. Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3**, **1**(3).



**Figure 1.** Proof-of-concept: Successful metabolic engineering strategy guided by modeling.

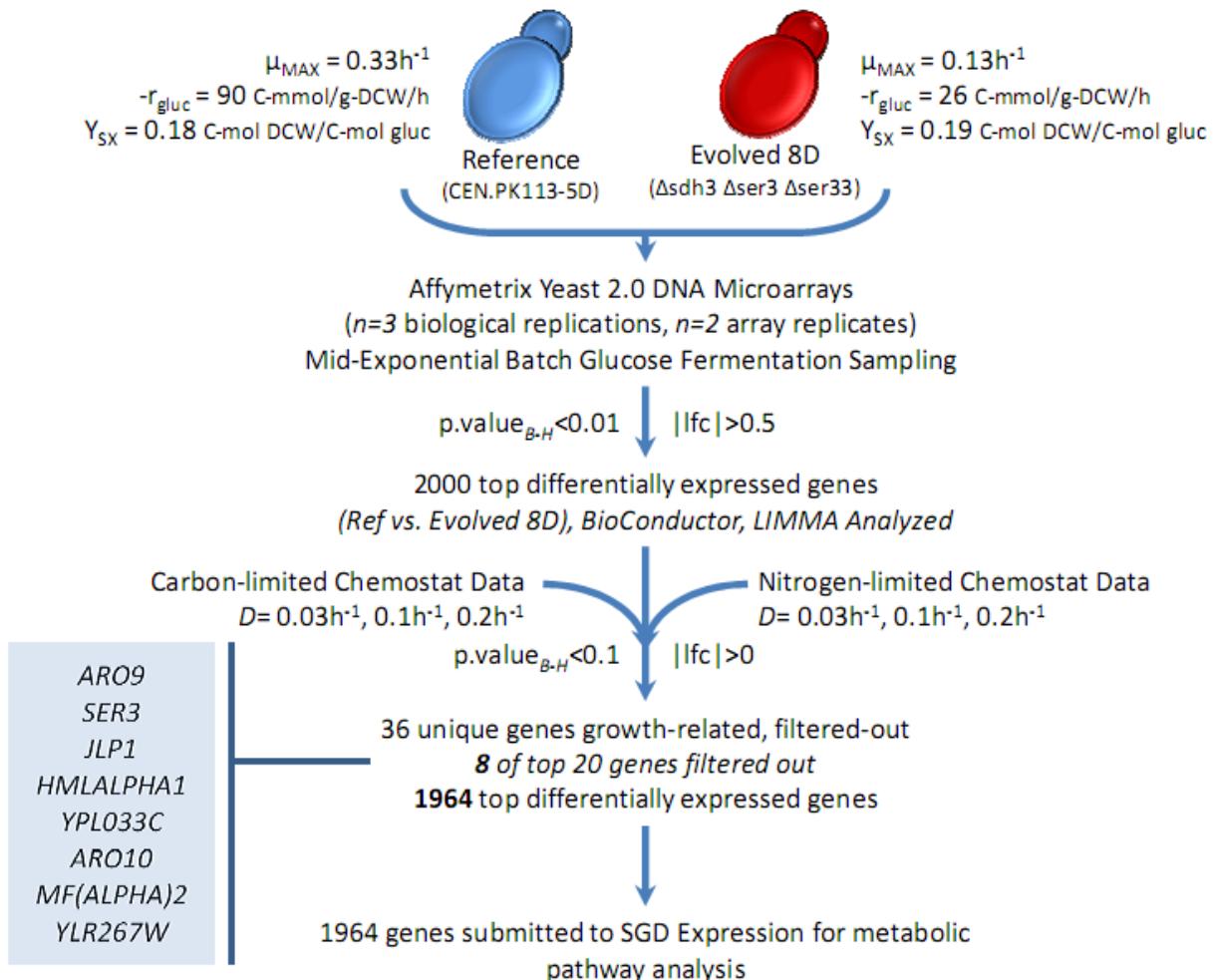
Panel a shows the central carbon metabolism of *S. cerevisiae*, and the model-guided metabolic engineering strategy for succinate over-production. Succinate production is directly coupled to biomass formation based on three gene deletions: *sdh3* (cytochrome b subunit of succinate dehydrogenase complex), and *ser3/ser33* (3-phosphoglycerate dehydrogenase isoenzymes). The remodeling of central carbon flux towards succinate requires minimizing the conversion of succinate to fumarate, and forcing the biomass-required amino acids L-glycine and L-serine to be produced from glyoxylate pools. Production of glyoxylate results from isocitrate conversion by *Icl1p*, producing equimolar succinate. As the biomass yield increases, the demand for L-glycine and L-serine increase proportionally, driving biomass-coupled succinate production. Legend: native reactions (blue solid line), lumped native reactions (blue dashed line), interrupted reactions (red solid line), up-regulated reactions (green solid line).

Panel b demonstrates the proof of concept. The reference strain and genetically engineered mutant strain, 8D, supplemented with 500 mg L<sup>-1</sup> glycine were physiologically characterized in 2L well-controlled stirred-tank fermentations. There was a 13.3X improvement in succinate titer.



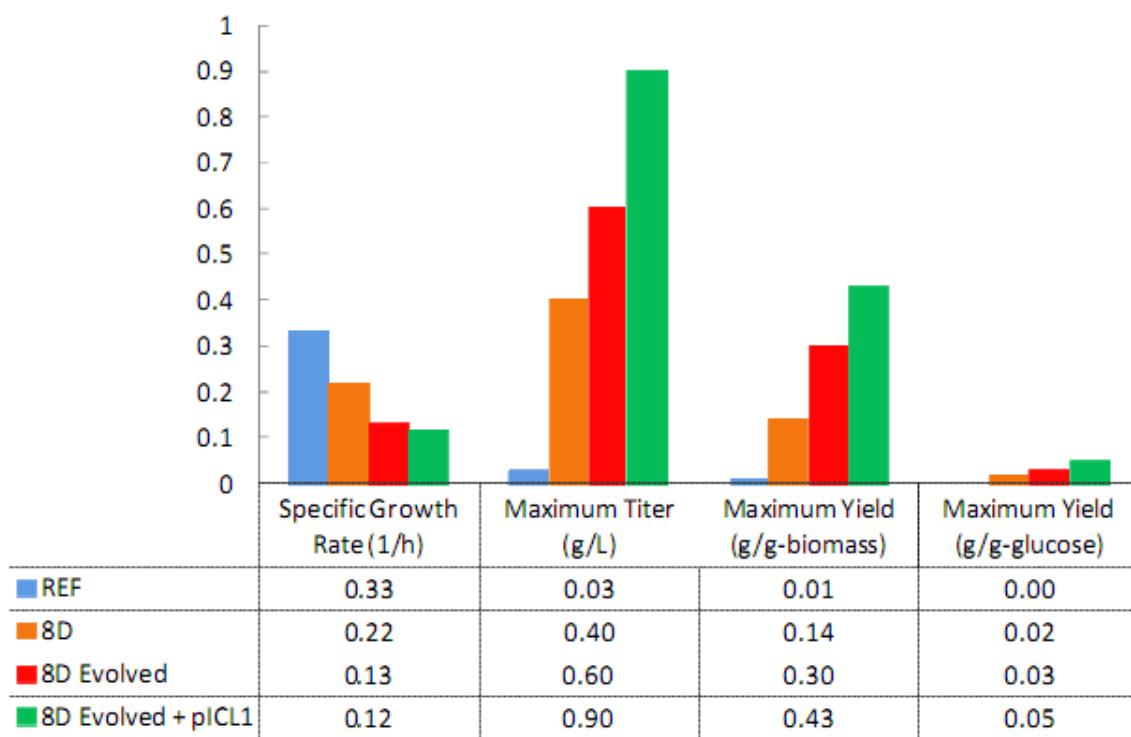
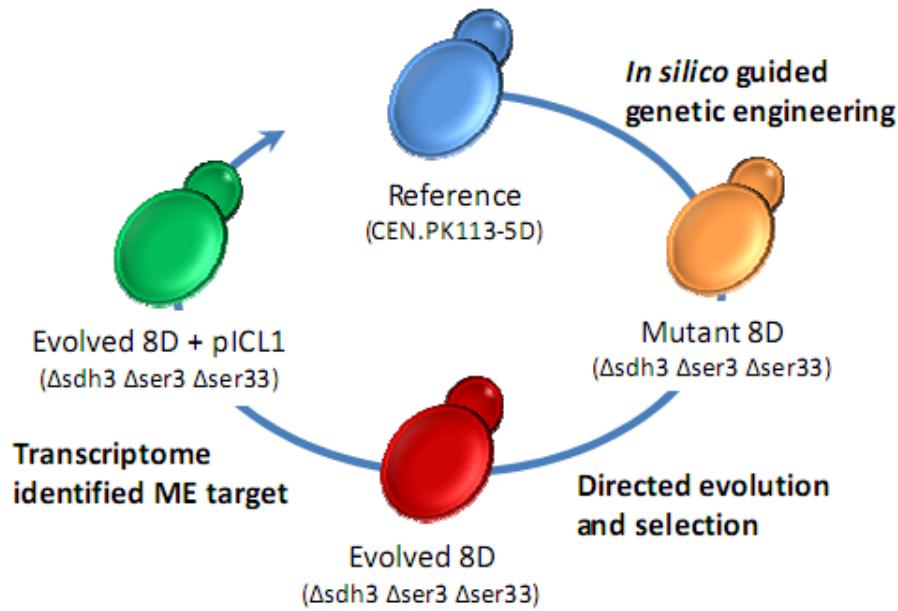
**Figure 2.** Metabolic engineering enhanced by directed evolutions.

Cell populations were transferred across six shake flask cultures until a glycine prototroph was isolated. Subsequently, successive cultures were used to select for faster growth. From the final shake flask (SF3) the strain Evolved 8D was isolated. The succinate yield on biomass is plotted for each shake flask culture, demonstrating a 7.8X increase. The right plot shows the profile of specific growth rate and succinate yield on biomass for the final selection of faster growing cells.



**Figure 3.** Transcriptome guided metabolic engineering – Analysis

Affymetrix Yeast 2.0 DNA microarrays were used for transcriptome analysis of each strain cultured in well-controlled glucose batch fermentations. The top 2000 differentially expressed genes had an adjusted  $p.value < 0.01$  and  $log\text{-fold change } (Ifc) > 0.5$ . A carbon-limited and nitrogen-limited chemostat transcriptome data set using the reference strain, surveyed at dilution rates ( $D$ ) of 0.03, 0.1, and  $0.2\text{h}^{-1}$  was used to determine which genes are growth-related under each condition. A total of 36 unique growth-related genes were identified from statistical analysis of each data set and with a total of 8 growth-related genes being among the top 20 differentially expressed genes between the reference and evolved 8D strain. After removal of the 36 genes, a total of 1964 genes were carried further for pathway analysis.



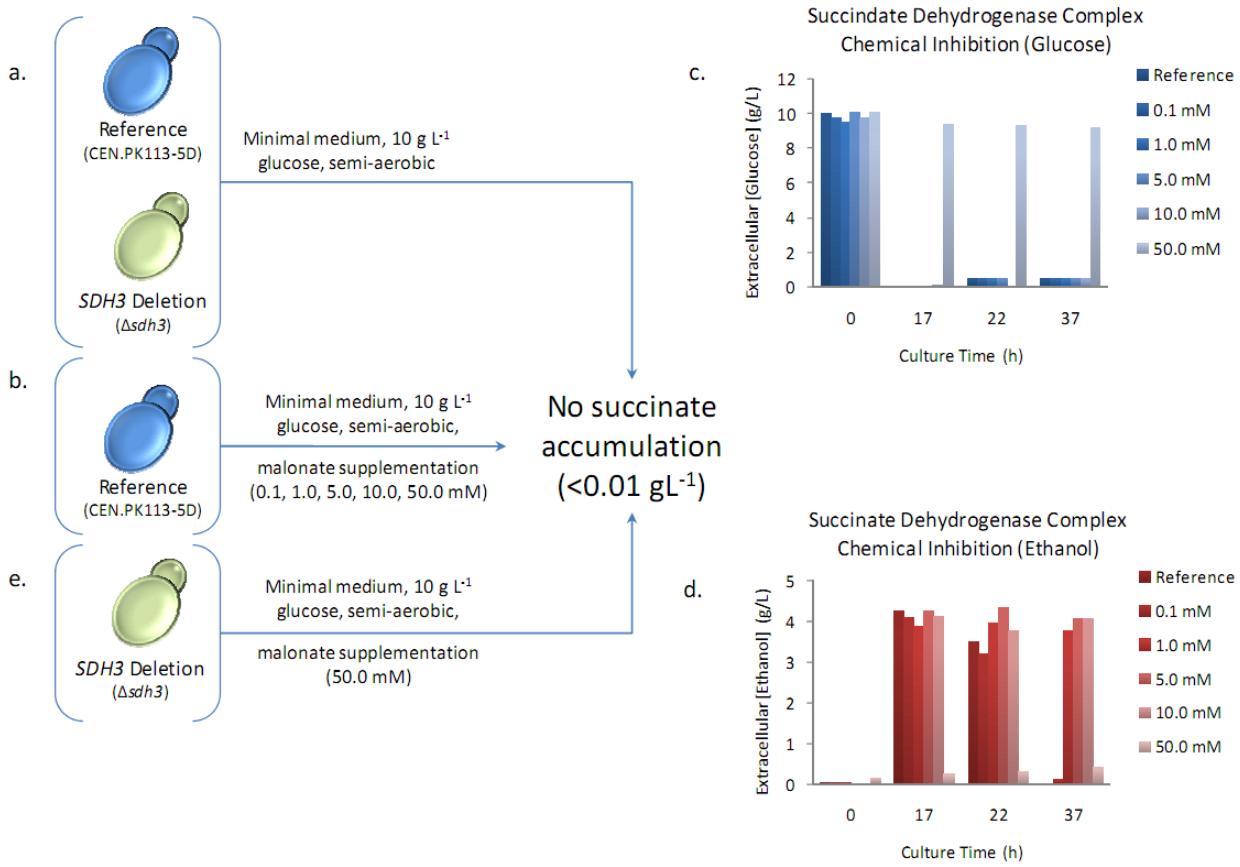
**Figure 4.** Summary of succinate microbial cell factory construction

The specific growth rate (1/h), maximum succinate titer (g/L), maximum succinate yield on biomass (g/g-biomass), and maximum yield on glucose (g/g-glucose) are reported for the reference strain, 8D, 8D evolved, and 8D evolved with pICL1. A 43-fold improvement in succinate yield on biomass was observed across the full cycle of metabolic engineering that included *in silico* guided approaches, directed evolution, and transcriptome based identification of a 2<sup>nd</sup> round of metabolic engineering targets.

## **Supplementary Information**

### **Table of Contents**

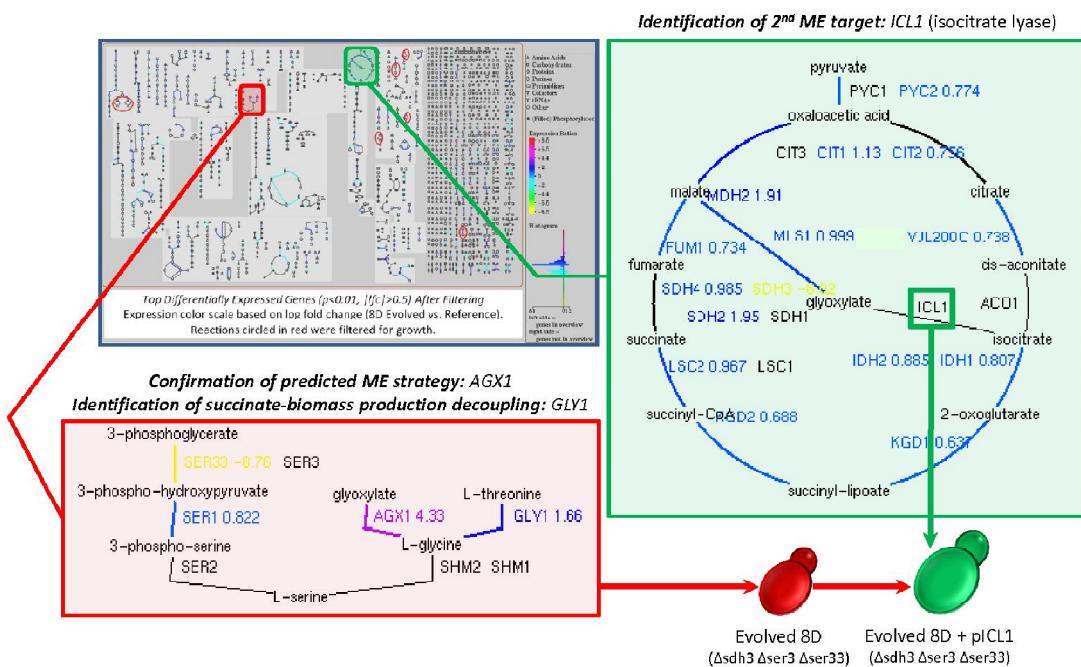
- 1. Supplementary Information 1:** Succinate dehydrogenase and succinate accumulation
- 2. Supplementary Information 2:** Transcriptome data and metabolic pathways
- 3. Supplementary Information 3:** Glycine, serine and threonine pathways for succinate production



## Supplementary Information 1

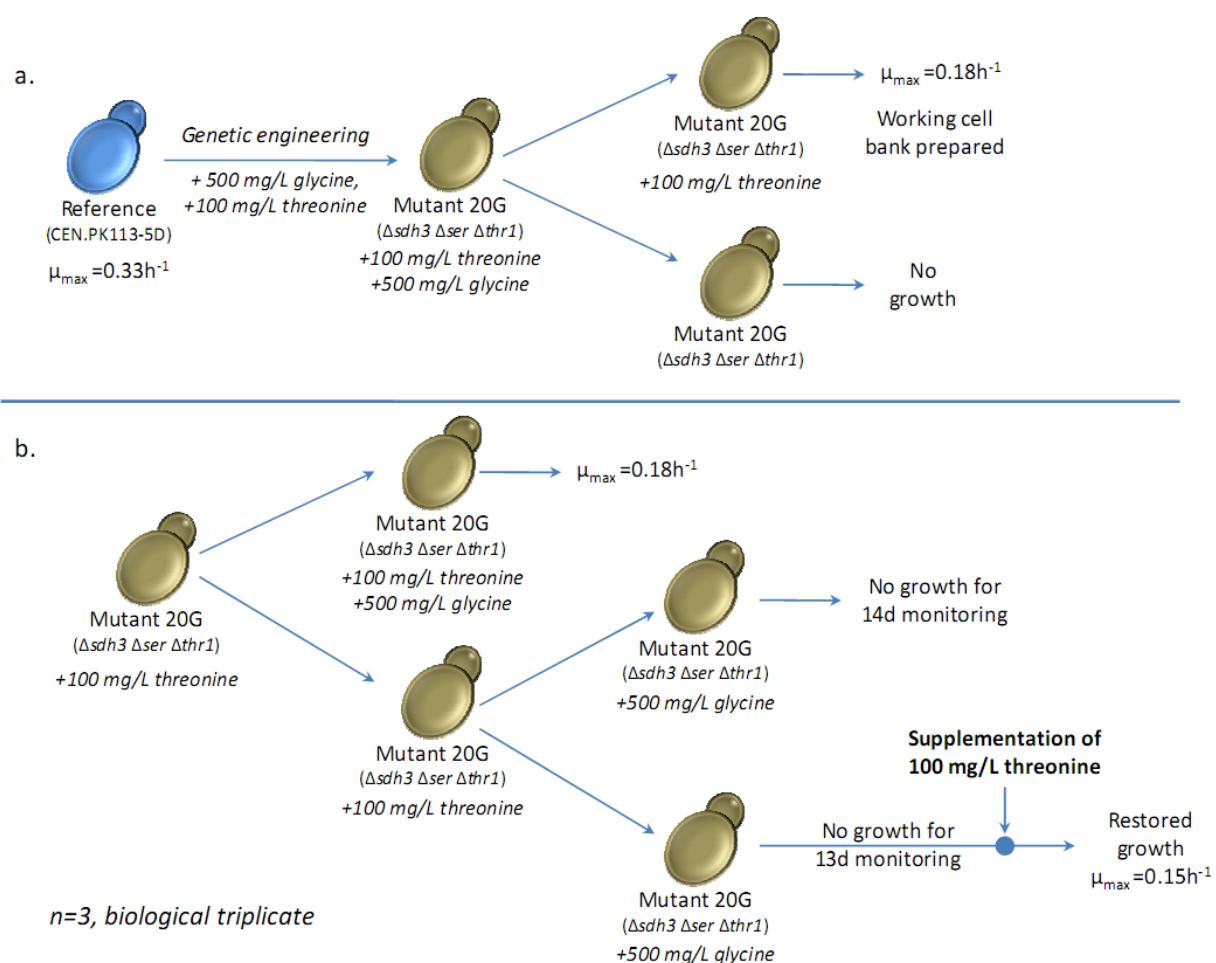
Inhibition of the succinate dehydrogenase complex with malonate supplementation in shake flask cultures was evaluated. The reference and *Δsdh3* strain, previously described (Cmini *et al*, 2008), were cultured in minimal media supplemented with 10 g L<sup>-1</sup> glucose and no succinate accumulation was detected (Panel a). The reference strain was cultured with 0.1, 1.0, 5.0, 10.0, and 50.0 mM malonate supplementation. Under no supplementation conditions succinate accumulation was observed (Panel b). In order to confirm that the concentration of malonate in the culture was effectively inhibiting succinate dehydrogenase activity, residual ethanol in the culture broth was monitored. Succinate dehydrogenase activity, as previously described, catalyzes the conversion of succinate to fumarte with net production of protonated ubquinone. Ethanol is a carbon source readily catabolized by *S. cerevisiae* using respiro-fermentative pathways and requiring succinate dehydrogenase activity. Panel c shows the residual glucose concentration in the culture broth at 0, 17, 22, and 37h post-inoculation for no supplementation of malonate (reference) and then 0.1, 1.0, 5.0, 10.0, and 50.0 mM malonate supplementation. These growth profiles were generated using the reference strain (CEN.PK113-7D). As expected, full catabolism of glucose was observed at all malonate concentrations with the exception of 50.0 mM, thereby considered an upper limit. Similarly, in panel d, is the ethanol concentration in the culture broth for the same malonate concentrations and sample times. At 37h, as expected, the reference strain had consumed nearly all ethanol produced during the glucose consumption phase. Malonate concentrations of 1.0, 5.0, and 10.0 mM malonate resulted in significant ethanol respiration inhibition compared to no supplementation and 0.1 mM malonate, confirming that respiro-fermentative catabolism was inhibited. Under no circumstances was succinate accumulation observed. Furthermore, the *Δsdh3*

strain was supplemented with 50.0 mM malonate to ensure no unexpected interaction between the genetic modification and malonate supplementation (panel e).



## Supplementary Information 2

A total of 1964 genes were submitted to the Saccharomyces Genome Database tool, Pathway Expression Viewer. The resulting Pathway Expression map shows the relative log-fold change of all *S. cerevisiae* metabolic reactions (Evolved 8D vs. Reference). Three key results are highlighted from the transcriptome. First, isocitrate lyase (*ICL1*) was amongst the few genes not up-regulated in the Evolved 8D strain, thereby becoming a 2<sup>nd</sup> round metabolic engineering target. Second, alanine:glyoxylate aminotransferase (*AGX1*) was 4.3 log-fold higher in the Evolved 8D strain, confirming the predicted model-guided strategy of up-regulated glycine formation from glyoxylate pools. Third, threonine adolase (*GLY1*) was 1.6 log-fold higher in the Evolved 8D strain. The genome-scale model reconstruction used for predictions annotated Gly1p as catalyzing the reversible conversion of threonine to glycine. This reaction has since been shown to be irreversible, converting threonine to glycine, consuming equimolar acetaldehyde. The transcriptome data suggests that the Evolved 8D strain demonstrated de-coupling of succinate and biomass production because alternative reactions (e.g., Gly1p) were supplying glycine pools.



### Supplementary Information 3

Panel a briefly describes the mutant construction of 20G,  $\Delta sdh3 \Delta ser3 \Delta thr1$ , from the reference strain and initially supplemented with 100 mg L<sup>-1</sup> threonine and 500 mg L<sup>-1</sup> glycine to satisfy the resulting auxotrophies. All growth challenges were evaluated in shake flasks supplemented with minimal medium, 300 mg L<sup>-1</sup> uracil, 10 g L<sup>-1</sup> glucose, and either threonine and/or glycine added, as indicated. The mutant 20G was not capable of sustaining growth in the absence of threonine, and therefore a working cell bank was prepared. Panel b describes the shake flask experiments and progression followed to evaluate the strain's ability to be evolved from threonine supplementation to glycine supplementation. When 20G culture was inoculated from threonine supplemented medium to glycine only supplemented medium, no growth was observed up to 14d post-inoculation (2 samples per day measuring OD<sub>600</sub>). On day 14, a shake flask culture of 20G only supplemented with glycine, was then supplemented with 100 mg L<sup>-1</sup> threonine, and growth was immediately restored.. It was therefore concluded that the mutant 20G was incapable of catalyzing glycine to threonine to satisfy threonine cellular demands, given that threonine synthesis was interrupted with the deletion of *thr1*. This experimental conclusion further supports that *Gly1* encoding threonine adolase, originally believed to reversibly catalyze the conversion of threonine to glycine, is irreversible and cannot meet threonine cellular demands from glycine pools.

## PAPER II

**Genome-scale modeling enables metabolic engineering of *Saccharomyces cerevisiae* for succinic acid production.**

**José Manuel Otero**, Rasmus Ågren, Jens Nielsen

Submitted to *Metabolic Engineering*, 2009

**Title:** Genome-scale modeling enables metabolic engineering of *Saccharomyces cerevisiae* for succinic acid production

**Authors:** José Manuel Otero<sup>1,2,3</sup>, Rasmus Ågren<sup>1,2</sup>, Jens Nielsen<sup>1,2\*</sup>

1. Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-41296 Gothenburg, Sweden.
2. Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark DK-2800, Kgs. Lyngby, Denmark.
3. Current Address: Vaccine & Biologics Process Development, Bioprocess Research & Development, Merck Research Labs, West Point, PA, USA.

\*Corresponding author:

Professor Jens Nielsen

Phone: +46 (0) 31 772 8633

Fax: +46 (0) 31 772 3801

[nielsenj@chalmers.se](mailto:nielsenj@chalmers.se)

**Keywords:** succinic acid, metabolic engineering, genome-scale modeling, DNA microarrays, *Saccharomyces cerevisiae*

## Abstract

Industrial biotechnology is attempting to leverage metabolic engineering and systems biology for rapid development of microbial cell factories designed to produce building block and high value chemicals, such as succinic acid. Until now only prokaryotic host organisms were considered, but *Saccharomyces cerevisiae* in addition to being the most well physiologically and genomically characterized eukaryote, is a proven commercial platform for the production of a wide range of products, owing to its robustness, scalability, feedstock flexibility, and wide pH range. In this work, we describe the application of a genome-scale metabolic network reconstruction, using pathway visualization and flux balance analysis, to predict succinic acid overproduction strategies assuming batch glucose conditions. Amongst the top single and double gene deletion strategies, under aerobic and anaerobic conditions, three predictions suggested a 10-fold improvement in succinate yield on glucose compared to the reference: *MDH1*, *OAC1*, and *DIC1*. Model validation was performed using knock-out strains cultivated anaerobically on glucose, coupled with physiological and genome-wide DNA microarray characterization. While  $\Delta mdh1$  and  $\Delta oac1$  strains failed to produce more succinate relative to the reference,  $\Delta dic1$  produced 0.02 C-mol/C-mol glucose, in close agreement with model predictions (0.03 C-mol/C-mol glucose). Pathway visualization, coupled with transcriptional profiling, suggests that succinate formation is coupled to mitochondrial redox balancing, and more specifically, reductive TCA cycle activity. While far from industrial titers, this proof-of-concept suggests that *in silico* predictions coupled with experimental validation can identify novel metabolic engineering strategies.

## Introduction

The chemical manufacturing industry is actively seeking cost-effective, environmentally friendly, renewable, and sustainable raw material feedstocks that will not only enable production of key chemical building blocks, but can serve as a platform for future products (Otero, et al, 2007). In 2004, the US Department of Energy identified succinic acid as an added-value chemical building block, with an estimated 15,000 t/year world-wide demand predicted to expand to commodity chemical status with 270,000 t/year, representing a potential >2 billion USD annual market (McKinlay, et al, 2007; Wilke, et al, 2004; US DOE, 2008). Within microbial metabolism succinate formation results from two routes: 1) the isocitrate lyase, Icl1p, catalyzed conversion of isocitrate to equimolar glyoxylate and succinate, and 2) from the  $\alpha$ -keto-glutarate dehydrogenase complex, Kgd1p/Kgd2p/Lpd1p, catalyzed conversion of  $\alpha$ -keto-glutarate to equimolar succinate, with a net production of CO<sub>2</sub>, NADH, and ATP. Succinate is subsequently depleted by the succinate dehydrogenase complex, Sdh1p/Sdh2p/Sdh3p/Sdh4p to equimolar fumarate with the net production of protonated ubiquinone (Cherry, et al, 1998).

Numerous industrial biotechnology efforts focused on metabolic engineering of prokaryotes to overproduce succinic acid, including *Anaerobiospirillum succiniciproducens*, *Actinobacillus succinogenes*, *Succinivibrio dextrinosolvens*, *Corynebacterium glutanicum*, *Prevotella ruminocola*, a recently isolated bacterium from bovine rumen, *Mannheimia succiniciproducens*, and a metabolically engineered succinic acid over-producing *E. coli*, have been presented (Zeikus, et al, 1999; Song, et al, 2006; McKinlay, et al, 2007; Jantama, et al, 2008; Lee, et al, 2002; Lee, et al, 2003; Lee, et al, 2006; Kim, et al, 2007). All of the aforementioned hosts grow at neutral pH, and consequently secrete the salt, succinate, requiring a cost-intensive acidification and precipitation to reach the desired succinic acid. This concern is not unique to succinic acid production, but rather broadly applicable when considering organic acid producing microbial cell factories (Sauer, et al, 2008). *Saccharomyces cerevisiae* represents a well-established, generally regarded as safe, robust, scalable (1L to 100,000L) industrial production host capable of growth on diverse carbon sources, chemically defined medium, both aerobic and anaerobic, and a wide pH operating range (3.0-6.0). Unlike the bacteria described above, succinate does not natively accumulate in *S. cerevisiae*; but as there are many factors of importance for the choice of a microbial cell factory it is not uncommon that the chosen cell factory lacks predisposition to produce the target chemical of choice (Adrio, et al, 2006). In fact, hopes of enabling future biorefineries, where platform technologies will be exploited to convert lignocellulosic feedstocks to a dynamic portfolio of high added-value chemicals, similar to petrochemical refineries today, will only be realized by harnessing the metabolic diversity of microbial cell factory platforms such as *S. cerevisiae* (Lynd, et al, 1999; Otero, et al, 2007).

Metabolic engineering is an applied science focussing on developing new or improving existing cell factories (Bailey, et al, 1991; Stephanopoulos, et al, 1991; Nielsen, et al, 2001; Tyo, et al, 2007). More specifically, those improvements are through the use of gene-targeted, rational, and quantitative approaches for redirection of metabolic fluxes to improve the yield, titer, productivity, and/or robustness associated with specific metabolites in a bio-reaction network (Nielsen, et al, 2007). The challenge is rapidly identifying and confirming the high probability of success genetic targets, which has inhibited the rate of commercial successes attributable to metabolic engineering. Genome-scale metabolic network reconstructions, extensively described and reviewed elsewhere (Edwards, et al, 2002(a); Edwards, et al, 2002(b); Forster, et al, 2003; Edwards, et al, 2001), provide a quantitative framework for stoichiometric biochemical models annotated with gene identity, coupled with mass-balance boundary conditions, to make possible constraint based simulations of how the metabolic network operates at different conditions. For *S. cerevisiae*, the most well physiologically and systems biology characterized eukaryote, there are at present five genome-scale metabolic models (iFF708, iND750, iLL672, iIN800, and the most recent *S. cerevisiae* consensus model), with the first, iFF708, appearing in 2003, comprised of 1175 reactions (842 unique), 584 metabolites, two compartments, and 708 structural open reading frames (ORF), representing 12.1% genome coverage (Forster, et al, 2003; Duarte, et al, 2004; Blank, et al, 2005; Nookaew, et al, 2005; Herrgard, et al, 2008).

There has been very limited work on metabolic engineering of *S. cerevisiae* for overproduction of succinic acid for industrial applications. Succinic acid production in genetically modified sake yeast strains has been demonstrated for modification of taste profiles, primarily focusing on multi-gene deletions of citric acid cycle enzymes aconitase (ACO1p), fumarate reductase (OSM1p),  $\alpha$ -ketoglutarate dehydrogenase (KGD1p), fumarase (FUM1), and succinate dehydrogenase (SDH1), resulting in <0.7 g/L succinic acid on complex medium (Arikawa, et al, 1999(a); Arikawa, et al, 1999(b); Kubo, et al, 2000). There has also been significant experimental work focused on elucidating the physiological role of cytosolic and mitochondrial fumarate reductase (FRDS1p and OSM1p, respectively) in the context of facilitating anaerobic fermentation of *S. cerevisiae* (Camarasa, et al, 2007; Arikawa, et al, 1998; Enomoto, et al, 2002). Significant effort has been applied to understand succinate formation in *S. cerevisiae* by exploring *SDH1* and *SDH3* deletion mutants, specifically using  $^{13}\text{C}$ -NMR analysis of  $^{13}\text{C}$ -labelled aspartate and glutamate supplemented anaerobic glucose fermentations, and DNA microarray analysis of aerobic and anaerobic glucose supplemented fermentations, respectively (Camarasa, et al, 2003; Cimini, et al, 2008). In both efforts, no significant succinate accumulation was observed through simple deletion of the primary succinate consuming reaction catalyzed by the succinate dehydrogenase complex.

From an *in silico* approach, there has been one publication focusing on application of flux balance analysis (FBA) with the genome-scale metabolic network reconstruction of *S. cerevisiae* using an evolutionary programming method to couple biomass and succinate production (Patil, et al, 2005). This approach highlighted several multi-gene deletion strategies for succinic acid overproduction, however, included no experimental validation of target predictions. Furthermore, this approach used a reduced genome-scale metabolic network reconstruction of iFF708, removing all duplicate and dead-end reactions.

In this work we exploit FBA coupled with pathway visualization to explore succinic acid overproduction strategies as predicted by interrogation of the complete genome-scale metabolic reconstruction, iFF708. More specifically, we explore all single gene and double gene deletions under aerobic and anaerobic conditions, maximizing the objective function of growth rate with constrained glucose uptake rate, and observe the maximum succinate yield on substrate. The top three single gene deletion predictions, occurring under anaerobic glucose fermentation conditions, were experimentally evaluated in order to gain new insight into the predictive strength of *in silico* predictions. Furthermore, these three strains were physiologically and transcriptionally characterized with the objective to gain further knowledge on the C4 acid production by *S. cerevisiae*.

## Materials & Methods

### Genome-scale Modeling and Visualization

The *Saccharomyces cerevisiae* genome-scale metabolic network reconstruction iFF708 consisting of 708 structural open reading frames (ORFs), 1175 metabolic reactions, two compartments, and 584 metabolites, was used for all simulations (Forster, et al, 2003) throughout this work, and is publically available for download<sup>1</sup>. The following compounds are necessary for growth in iFF708 and were unconstrained in all simulations: ammonia, phosphate, and sulfate. Ergosterol and zymosterol are necessary for growth under anaerobic conditions but were unconstrained in all simulations. Sodium, potassium, 1-hexadecene, and 1-octadecene were tested both under constrained and unconstrained simulations, and found to have no impact on any of the results. A maintenance ATP requirement of 1 mmol/g-DCW/h was assumed. Unless otherwise stated, all simulations conditions shared an identical objective function: maximizing growth under a limiting glucose uptake rate. The glucose uptake rate, based on experimentally determined glucose uptake rates of the *S. cerevisiae* CEN.PK113-7D under batch aerobic glucose fermentation conditions (See Table 2), was fixed to 15.2 C-mmol/g-DCW/h (91.2 mmol/g-DCW/h). For simulations referred to as *Semi-aerobic*, *Forced NADH* (see Figure 1), the objective was to accurately model the production of glycerol, which *in vivo* results from overflow metabolism of batch aerobic glucose fermentation. These simulations included the introduction of a theoretical reaction,  $FNADH: NAD^+ \rightarrow NADH$ , and were constrained to 6 mmol-NADH/g-DCW/h. The production of cytosolic NADH from NAD<sup>+</sup> and the constraint value selected were based on fitting existing experimental data of the major overflow metabolites (ethanol, glycerol, and acetate). It should be noted that the approach of introducing an NAD<sup>+</sup> reductase to simulate batch cultivation is merely used as a control to more accurately simulate batch fermentation conditions in *S. cerevisiae*. The physiological drivers and resulting employed approach are further discussed in the results and discussions. For simulations referred to as *Semi-Aerobic* or *Semi-Anaerobic*, the oxygen uptake rate,  $r_{O_2}$ , was constrained to 1.8-mmol O<sub>2</sub>/g-DCW/L and 0.016 O<sub>2</sub>/g-DCW/L, respectively. For simulations referred to as *Aerobic* and *Anaerobic*, the  $r_{O_2}$  was unconstrained (0-1000 mmol O<sub>2</sub>/g-DCW/L) or constrained to 0 mmol O<sub>2</sub>/g-DCW/L, respectively.

Flux balance analysis (FBA) has been previously employed using an in-house software package referred to as BioOpt, employing the LINDO Application Programming Interface (API) (Lindo Systems, Inc.; Chicago, IL) for linear optimization (Forster, et al, 2003; Famili, et al, 2003). In order to enable novel visualization and FBA tools to be developed, a script for converting iFF708 from BioOpt file format (text) to Systems Biology Mark-up Language (SBML) was developed within MATLAB® (The MathWorks, Inc.; Natick, MA). FBA tools developed and employed within MATLAB®, requiring the COBRA Toolbox (Becker, et al, 2007), SBML Toolbox (Keating, et al, 2006), and the open-source linear solver GLPK (Giorgetti, 2008), are described below.

The first tool, *runSimulationNoVisual.m*, is employed for running a single simulation where the user specifies the model (e.g., SBML version of iFF708), a list of constraints, and an objective function. The program's output includes a list of all exchange fluxes (metabolites taken up or excreted), and the yield and specific consumption/production rates of designated metabolites (glucose, ethanol, glycerol, acetate, succinate, pyruvate, oxygen, carbon dioxide, ergosterol, zymosterol, ammonia, phosphate, sulfate, glucosamine 6-phosphate). The second tool, *multipleGeneDeletion.m*, is used for screening of successful gene deletions. Specifically, the user specifies the model, a list of constraints, an objective function, a list of genes to be deleted, and the number of combinations of genes to be deleted (e.g., all combinations of two genes or all combinations of three genes). The third tool, *followChanged.m*, compares the results of two simulations and uses the constraints to display the reactions that are of interest. The program enables the user to screen for reactions involving specific metabolites, reactions with fluxes in a specified range, or reactions with fluxes that changed within a specified amount relative to the reference case.

<sup>1</sup> <http://www.sysbio.se>

Visualization of FBA results is an important feature of systematic interpretation of the response of the metabolic network to different constraints and/or objective functions. There are no tools available that enable rapid, automated, reaction-specific, visualization of FBA results. Due to the magnitude of simulation data generated, representing up to 708 reaction fluxes (assigned to structural ORFs), there is often a delicate balance between the level of detail displayed on metabolic maps and ensuring a scientifically interpretable tool, the primary objective of this approach. Therefore, construction of a succinate-specific pathway map was performed in CellDesigner™ (Funahashi, et al, 2003), focusing on central carbon and energy metabolism, and exchange reactions (See Figure 1). After the user-specified metabolic map is drawn, the user may link a reaction to the corresponding reaction in iFF708. Then, a fourth tool, *constructPathwayFromCelldesigner.m*, converts the CellDesigner™ drawn map into a MATLAB® compatible format. The fifth tool, *runSimulation.m*, may be then used to specify conditions for both a test and reference case, whereby the constraints and objective function for each case are specified, and the resulting simulation data is plotted onto the metabolic map, where the relative change in flux between the test and reference case is indicated by color. The visualization tool described here may be used to rapidly interpret FBA results amongst a large number of conditions, facilitating the extraction of biologically relevant information which until now has often been lacking. Figure 1 presents the over-all structure of the metabolic map.

### Strains

The reference strain *Saccharomyces cerevisiae* BY4741 (*MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0*) and the single deletion strains were all received from the European Saccharomyces Cerevisiae Archive for Functional Analysis (Frankfurt, Germany). The reference strain *Saccharomyces cerevisiae* CEN.PK113-7D (*Mat a URA3 HIS3 LEU2 TRP1 MAL2-8c SUC2*) was received from the Scientific Research and Development GmbH (Oberursel, Germany) (van Dijken, et al, 2000). The single gene-deletion knock-out strains used throughout this study and their corresponding genotype are presented in Table 1.

### Medium Formulation

A chemically defined minimal medium of composition 5.0 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 3.0 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.5 g L<sup>-1</sup> MgSO<sub>4</sub>•7H<sub>2</sub>O, 1.0 mL L<sup>-1</sup> trace metal solution, 300 mg L<sup>-1</sup> uracil, 800 mg L<sup>-1</sup> lysine, 200 mg L<sup>-1</sup> histidine, 200 mg L<sup>-1</sup> methionine, 0.05 g L<sup>-1</sup> antifoam 204 (Sigma-Aldrich A-8311), and 1.0 mL L<sup>-1</sup> vitamin solution was used for all shake flask and 2L well-controlled fermentations (Verudyn, et al, 1992). The trace element solution included 15 g L<sup>-1</sup> EDTA, 0.45 g L<sup>-1</sup> CaCl<sub>2</sub>•2H<sub>2</sub>O, 0.45 g L<sup>-1</sup> ZnSO<sub>4</sub>•7H<sub>2</sub>O, 0.3 g L<sup>-1</sup> FeSO<sub>4</sub>•7H<sub>2</sub>O, 100 mg L<sup>-1</sup> H<sub>3</sub>BO<sub>4</sub>, 1 g L<sup>-1</sup> MnCl<sub>2</sub>•2H<sub>2</sub>O, 0.3 g L<sup>-1</sup> CoCl<sub>2</sub>•6H<sub>2</sub>O, 0.3 g L<sup>-1</sup> CuSO<sub>4</sub>•5H<sub>2</sub>O, 0.4 g L<sup>-1</sup> NaMoO<sub>4</sub>•2H<sub>2</sub>O. The pH of the trace metal solution was adjusted to 4.00 with 2M NaOH and heat sterilized. The vitamin solution included 50 mg L<sup>-1</sup> d-biotin, 200 mg L<sup>-1</sup> para-amino benzoic acid, 1 g L<sup>-1</sup> nicotinic acid, 1 g L<sup>-1</sup> Ca-pantothenate, 1 g L<sup>-1</sup> pyridoxine HCl, 1 g L<sup>-1</sup> thiamine HCl, and 25 mg L<sup>-1</sup> myo-inositol. The pH of the vitamin solution was adjusted to 6.5 with 2M NaOH, sterile-filtered and the solution was stored at 4°C. The final formulated medium, excluding glucose and vitamin solution supplementation, is adjusted to pH 5.0 with 2M NaOH and heat sterilized. For carbon-limited cultivations the sterilized medium is supplemented with 20 g L<sup>-1</sup> glucose, heat sterilized separately, and 1.0 mL L<sup>-1</sup> vitamin solution is added by sterile filtration (0.20 µm pore size Ministart®-Plus Sartorius AG, Goettingen, Germany). For anaerobic fermentations a total of 4 g L<sup>-1</sup> ergosterol and 168 g L<sup>-1</sup> Tween 80 dissolved in pure ethanol was supplemented.

### Shake Flask Cultivations and Stirred Tank Fermentations

Shake flask cultivations were completed in 500 mL Erlenmeyer flasks with two diametrically opposed baffles and two side-necks with septums for sampling by syringe. Flasks were heat sterilized with 100 mL of medium, inoculated with a single colony, and incubated at 30°C with orbital shaking at 150 RPM. Stirred tank fermentations were completed in well-controlled, aerobic or anaerobic, 2.2L Braun Biotech Biostat B fermentation systems with a working volume of 2L (Sartorius

AG, Goettingen, Germany). The temperature was controlled at 30°C. The fermenters were outfitted with two disk-turbine impellers rotating at 600 RPM. Dissolved oxygen was monitored with an autoclavable polarographic oxygen electrode (Mettler-Toledo, Columbus, OH). During aerobic cultivation the air sparging flow rate was 1 vvm. During anaerobic cultivation nitrogen containing less than 5 ppm O<sub>2</sub> was used for sparging at a constant flow rate of 2 vvm, with less than 1% air saturated oxygen in the fermenter as confirmed by dissolved oxygen and off-gas analysis. The higher flow rate of 2 vvm was employed to ensure anaerobic conditions; however, it is acknowledged that ethanol stripping was likely to increase. The pH was kept constant at 5.0 by automatic addition of 2M KOH. Off-gas passed through a condenser to minimize the evaporation from the fermenter. The fermenters were inoculated from shake flask precultures to an initial OD<sub>600</sub> 0.005.

### **Fermentation Analysis**

**Off-gas Analysis:** The effluent fermentation gas was measured every 30 seconds for determination of O<sub>2</sub>(g) and CO<sub>2</sub>(g) concentrations by the off-gas analyzer Brüel and Kjær 1308 (Brüel & Kjær, Nærum, Denmark).

**Biomass Determination:** The optical density (OD) was determined at 600 nm using a Shimadzu UV mini 1240 spectrophotometer (Shimadzu Europe GmbH, Duisberg, Germany). Duplicate samples were diluted with deionized water to obtain OD<sub>600</sub> measurements in the linear range of 0-0.4 OD<sub>600</sub>. Samples were always maintained at 4°C post-sampling until OD<sub>600</sub> and dry cell weight (DCW) measurements were performed. DCW measurements were determined through the exponential phase, until stationary phase was confirmed according to OD<sub>600</sub> and off-gas analysis. Nitrocellulose filters (0.45 µm Sartorius AG, Goettingen, Germany) were used. The filters were pre-dried in a microwave oven at 150W for 10 min., and cooled in a desiccator for 10 min. 5.0 mL of fermentation broth were filtered, followed by 10 mL DI water. Filters were then dried in a microwave oven for 20 min. at 150W, cooled for 15 min. in a desiccator, and the mass was determined.

**Metabolite Concentration Determination:** All fermentation samples were immediately filtered using a 0.45 µm syringe-filter (Sartorius AG, Goettingen, Germany) and stored at -20°C until further analysis. Glucose, ethanol, glycerol, acetate, succinate, pyruvate, fumarate, citrate, oxalate, and malate were determined by HPLC analysis using an Aminex HPX-87H ion-exclusion column (Bio-Rad Laboratories, Hercules, CA). The column was maintained at 65°C and elution performed using 5 mM H<sub>2</sub>SO<sub>4</sub> as the mobile phase at a flow rate of 0.6 mL min.<sup>-1</sup>. Glucose, ethanol, glycerol, acetate, succinate, citrate, fumarate, malate, oxalate were detected on a Waters 410 differential refractometer detector (Shodex, Kawasaki, Japan), and acetate and pyruvate were detected on a Waters 468 absorbance detector set at 210 nm.

### **Transcriptomics**

**RNA Sampling and Isolation:** Samples for RNA isolation from the late-exponential phase of glucose-limited batch cultivations were taken by rapidly sampling 25 mL of culture into a 50 mL sterile Falcon tube with 40 mL of crushed ice in order to decrease the sample temperature to below 2°C in less than 10 seconds. Cells were immediately centrifuged (4000 RPM at 0°C for 2.5 min.), the supernatant discarded, and the pellet frozen in liquid nitrogen and it was stored at -80°C until total RNA extraction. Total RNA was extracted using the FastRNA Pro RED kit (QBiogene, Carlsbad, USA) according to manufacturer's instructions after partially thawing the samples on ice. RNA sample integrity and quality was determined prior to hybridization with an Agilent 2100 Bioanalyzer and RNA 6000 Nano LabChip kit according to the manufacturer's instruction (Agilent, Santa Clara, CA).

**Probe Preparation and Hybridization to DNA Microarrays:** Messenger RNA (mRNA) extraction, cDNA synthesis, labeling, and array hybridization to Affymetrix Yeast Genome Y2.0 arrays were performed according to the manufacturer's recommendations (Affymetrix GeneChip® Expression Analysis Technical Manual, 2005-2006 Rev. 2.0). Washing and staining of arrays were performed using the GeneChip Fluidics Station 450 and scanning with the Affymetrix GeneArray Scanner (Affymetrix, Santa Clara, CA).

**Microarray Gene Transcription Analysis:** Affymetrix Microarray Suite v5.0 was used to generate CEL files of the scanned DNA microarrays. These CEL files were then processed using the statistical language and environment R v5.3 (R Development Core Team, 2007, [www.r-project.org](http://www.r-project.org)), supplemented with Bioconductor v2.3 (Biconductor Development Core Team, 2008, [www.bioconductor.org](http://www.bioconductor.org)) packages Biobase, affy, gcrma, and limma (Smyth, et al, 2005). The probe intensities were normalized for background using the robust multiarray average (RMA) method only using perfect match (PM) probes after the raw image file of the DNA microarray was visually inspected for acceptable quality. Normalization was performed using the qspline method and gene expression values were calculated from PM probes with the median polish summary. Statistical analysis was applied to determine differentially expressed genes using the limma statistical package. Moderated t-tests between the sets of experiments were used for pair-wise comparisons. Empirical Bayesian statistics were used to moderate the standard errors within each gene and Benjamini-Hochberg's method was used to adjust for multi-testing. A cut-off value of adjusted  $p<0.05$  was used for statistical significance, unless otherwise specified (Smyth, et al, 2004). Gene ontology process annotation was performed by submitting differentially expressed gene (adjusted  $p<0.05$ ) lists to the Saccharomyces Genome Database GO Term Finder resource and maintaining a cut-off value of  $p<0.01$  (Cherry, et al, 1998).

## Results

The various modeling and visualization software tools developed to identify *in silico* metabolic engineering strategies for overproduction of succinic acid are described in Figure 1. These tools were specifically designed to leverage the original iFF708 genome scale metabolic network reconstruction of *S. cerevisiae*; however, they are broadly applicable to any genome scale metabolic model employing FBA for identification of novel strategies for overproduction of a product of interest. The CellDesigner file can together with the SBML format of the iFF708 model be downloaded from [www.sysbio.se](http://www.sysbio.se).

### Model Validation and Comparison to Experimental Data

Following the successful conversion of iFF708 to SBML format, and development of FBA and visualization tools described in Figure 1, fermentation data of *S. cerevisiae* CEN.PK113-7D, summarized in Table 2, was used to evaluate the model's predictive power. Batch aerobic and anaerobic glucose fermentations performed in well-controlled 2L fermentations were compared to corresponding simulation conditions where the objective function, growth, was maximized while constraining glucose uptake rate, and for anaerobic conditions, constraining the oxygen uptake rate ( $r_{O_2}$ ) to 0 mmol O<sub>2</sub>/g-DCW/h. Table 2 clearly demonstrates that under aerobic conditions, 96.3 ± 4.0% of all carbon is recovered, and distributed across ethanol (54%), acetate (1%), glycerol (8%), carbon dioxide (16%), and biomass (17%) formation. Figure 2 shows results of simulated carbon distributions and the specific growth rate, when oxygen was unconstrained (0-1000 mmol O<sub>2</sub>/g-DCW/h). It is found that there is a poor agreement with corresponding batch glucose aerobic experimental data due to the inability of the model to describe the Crabtree effect as discussed earlier (Åkesson, et al, 2004). When  $r_{O_2}$  was constrained to experimentally determined fermentation values of 1.8 mmol O<sub>2</sub>/g-DCW/h, referred to as semi-aerobic, the simulation accurately predicted the specific growth rate (0.38 vs. 0.40h<sup>-1</sup>, experimental vs. simulation, respectively), ethanol yield (0.54 vs. 0.54 C-mol/C-mol glucose), and biomass yield (0.17 vs. 0.18 C-mol/C-mol glucose). However, carbon dioxide (0.16 vs. 0.30 C-mol/C-mol glucose) and glycerol (0.08 vs 0.0 C-mol/C-mol glucose) yields were in poor agreement (See Supplementary Discussion 1). Biomass formation as a result of glucose respiro-fermentative metabolism, with a high dependence on oxygen availability and glucose concentration, results in the formation of excess NADH (Nissen, et al., 1997). Excess NADH, both cytosolic and mitochondrial, is a direct result of biomass required ATP generation, and compartmental redox balance is possible through cytosolic NADH dehydrogenases, the glycerol-3-phosphate shuttle, and mitochondrial redox shuttles (von Jagow, et al, 1970; Luttik, et al, 1998; Overkamp, et al, 2000; Geertman, et al, 2006). Glycerol formation results from redox balancing and NADH regeneration to NAD<sup>+</sup> in the cytosol, and glycerol production can be reduced through expression of a cytosolic NADH oxidase (Vermuri, et al, 2007). Improving the fit of the model to glycerol production can be accommodated by several means, but here we took a simple pragmatic approach by introducing an artificial conversion of NAD<sup>+</sup> → NADH, and then constraining this reaction to a flux such that the glycerol production is correctly described by the model. We chose this approach rather than simply constraining the glycerol flux as this was found to give better overall fit of the fluxes. Simulations exploiting this approach are referred to as *Semi-aerobic, Forced NADH*, and are also presented in Figure 2. As a consequence of introduction of this reaction (constraining FNADH to 6 mmol NADH/g-DCW/h), glycerol yield was 0.079 vs 0.078 C-mmol/C-mmol glucose (experimental vs. simulation, respectively). While simulated carbon dioxide yield were still higher than observed experimentally, the *Semi-Aerobic, Forced NADH* simulation condition exhibits strong alignment to experimentally determined specific growth rate and productivities (See Figure 2).

Model validation was initially performed using *S. cerevisiae* CEN.PK113-7D batch glucose aerobic fermentation data; however, realizing that succinate metabolic engineering strategies would likely require exploration of anaerobic metabolism, similar comparative analysis for anaerobic fermentations was performed. More specifically, the reference *S. cerevisiae* BY4741 was also included, noting that gene deletion strategies to be identified *in silico* could rapidly be evaluated *in vivo* using the systematic Yeast Knock-Out (YKO) library available from the *Saccharomyces* Gene

Deletion Project (Winzler, et al, 1999). Under anaerobic conditions, the carbon recovery for both strains CEN.PK113-7D and BY4741 are significantly less compared to aerobic conditions (Table 2); however, when evaluating experimental and simulation values for specific growth rate and specific productivities there is reasonable agreement. Specifically, for CEN.PK113-7D, BY4741, and anaerobic simulations the specific growth rate was 0.29, 0.27, and 0.29 h<sup>-1</sup>, respectively. For ethanol (41.0, 40.8, 51.9 C-mmol/g-DCW/h), glycerol (10.5, 3.5, 2.0 C-mmol/g-DCW/h), and carbon dioxide (11.1, 11.3, 26.5 C-mmol/g-DCW/h) specific productivities the agreement between experimental and model simulations were fair, but indicating that the lack of carbon recovery is likely a result of ethanol stripping and evaporation from the bioreactor. Emphasis was placed on ensuring simulation conditions and constraints captured experimentally observed metabolite production with less focus on matching exact flux values. This is consistent with the approach that genome-scale metabolic modeling, particularly coupled with the visualization tools described previously, would aid in identifying novel metabolic engineering strategies for which mechanistic understanding as to why certain flux distribution patterns were favorable compared to others could be provided.

### ***Gene Deletion Strategies for Succinate Overproduction***

Overproduction of succinic acid was evaluated *in silico* using the various simulation conditions previously described, including the model modifications and accompanying constraints required to more accurately predict experimentally observed batch physiology, under both aerobic and anaerobic conditions. Prior to investigating those results, the maximum theoretical yield of succinic acid was determined *in silico*. Assuming 1 mmol ATP/g-DCW/h maintenance cost and a 10 mmol glucose/g-DCW/h uptake rate, the maximum succinate yield is 0.51 g/g-glucose. This maximum yield is based on FBA when [H<sup>+</sup>] was balanced. The exact mechanism by which succinate is transported across the cytosolic membrane has yet to be clearly elucidated, with literature suggesting both dicarboxylic acid proton-coupling, and the absence of such coupling (Aliverdieva, et al., 2006). If [H<sup>+</sup>] is treated as an external metabolite (e.g., unconstrained), the maximum yield of succinate is 0.98 g/g-glucose. Furthermore, if carbon dioxide uptake is permitted, enabling carboxylation reactions, the maximum theoretical yield is 1.124 g/g-glucose. Given the lack of physiological characterization of succinate transport, and the relatively high impact of assumptions surrounding [H<sup>+</sup>] balancing, external [H<sup>+</sup>] was balanced throughout all simulations, and the maximum succinate yield was assumed to be 0.51 g/g-glucose (0.52 C-mol/C-mol glucose). This represents a worst case scenario in terms of the theoretical potential for *S. cerevisiae* to stoichiometrically overproduce succinate.

Under aerobic conditions there are no single gene deletions that result in increased succinate production (See Supplementary Data 1). Interestingly, the reference case simulation under aerobic conditions with no gene deletions produces a small amount of succinate (0.003 C-mol/C-mol glucose), which is not observed experimentally. If succinate excretion is constrained to zero, optimization of growth rate will result in growth while producing glycerol, under minimal amounts of oxygen, and then acetate under increasing amounts of oxygen. However, experimentally, both glycerol and acetate production are observed while succinate production is absent. Under aerobic conditions there is a strong sensitivity of succinate yield on substrate to  $r_{O_2}$  and for  $r_{O_2} > 2$  mmol O<sub>2</sub>/g-DCW/h the succinate yield on substrate is zero (to be discussed later).

At aerobic conditions double gene deletions only resulted in minor improvement of succinate production (maximum a factor of 5, data not shown). Nearly all of the predictions required the deletion of the succinate dehydrogenase complex (Sdh3p), which catalyzes the conversion of succinate to fumarate in the TCA cycle, and represents the primary succinate consumption reaction in *S. cerevisiae* central carbon metabolism. In addition to previous work suggesting that succinate dehydrogenase complex interruption does not lead to succinate accumulation (Camrasa, et al, 2003; Cimini, et al, 2009), Table 2 confirms that deletion of  $\Delta sdh3$  in the BY4741 strain also fails to accumulate succinate.

Given the high degree of sensitivity of succinate production to  $r_{O_2}$ , anaerobic simulations offer the advantage of constraining this flux to zero, and these conditions can be tested reasonably

well experimentally. Under anaerobic simulation conditions, a small amount of succinate is produced, 0.003 C-mol/C-mol glucose, and if succinate production is constrained to zero, then the model predicts no growth. This is likely because the production of orotate from dihydroorotate, catalyzed by dihydroorotate dehydrogenase (encoded by *URA1*) required for pyrimidine synthesis, is coupled to the reduction of ubiquinone to ubiquinol. Under aerobic conditions oxygen serves as the final electron acceptor and enables ubiquinone regeneration, while under anaerobic conditions flavin adenine dinucleotide (FAD) serves as the electron acceptor for ubiquinone regeneration, and FAD must be regenerated from the transfer of electrons to fumarate, producing succinate. Given this proposed mechanism, the solution space for succinate production under anaerobic conditions rapidly approaches singularity with a high dependence on  $r_{O_2}$ . Given that experimentally it would be difficult to ensure 0 mmol O<sub>2</sub>, potential gene deletions were therefore screened for micro aerobic conditions, where  $r_{O_2}$  was constrained to 0.016 mmol O<sub>2</sub>/g-DCW/h, and determined to be the minimum  $r_{O_2}$  required for sustaining cell growth at the same rate if succinate production is constrained to zero or unconstrained. Table 3 presents the top single gene deletions for succinate overproduction under both 0 and 0.016 mmol O<sub>2</sub>/g-DCW/h constraints. This approach is extended to include double gene deletions; however, only the conditions of 0.016 mmol O<sub>2</sub>/g-DCW/h are presented. Table 3 shows that a significant increase in the succinate yield, by a factor of approximately 10-fold from the reference case, can be obtained for the single gene deletions  $\Delta oac1$ ,  $\Delta mdh1$ , and  $\Delta dic1$  (0.033 C-mol/C-mol glucose vs. 0.003 C-mol/C-mol glucose, single gene deletion vs. reference case simulation, respectively). Furthermore the significant increase in succinate yield on substrate resulted in nearly no impacts to growth rate (0.28h<sup>-1</sup> vs. 0.30h<sup>-1</sup>, single gene deletion vs. reference case simulation, respectively). Physiologically, it was confirmed that  $\Delta oac1$ ,  $\Delta mdh1$ , and  $\Delta dic1$  are viable null mutants, and their annotation is well known, encoding for an inner mitochondrial membrane transporter (OAC1p), malate dehydrogenase (MDH1p), and an inner dicarboxylate mitochondrial transporter (DIC1p), respectively (Cherry, et al, 1998). Interestingly, further simulations of the best double gene deletions resulted in the same order of magnitude succinate yields on substrate compared to the aforementioned single gene deletions.

### ***Physiological Characterization of Gene Deletion Strains***

In order to explore and validate if the single gene deletions identified *in silico* result in more succinate production, the corresponding strains of the BY4741 background (See Table 1) were cultivated anaerobically in 2L well controlled fermenters. Fermentation results are presented in Table 2, and comparative analysis between simulation and experimental results are presented in Figure 3. It is seen that there is a fair agreement between model predictions and experimental data. Focusing more closely on the specific succinate productivity, the reference case,  $\Delta mdh1$ , and  $\Delta oac1$  experimentally determined yields are significantly lower than expected based on model simulations. The  $\Delta dic1$  case, however, demonstrated a significantly higher yield of succinate compared to the reference case (0.02 vs. 0.00 C-mol/C-mol glucose,  $\Delta dic1$  vs. reference, respectively), and was in-line with the *in silico* prediction (0.02 vs. 0.03 C-mol/C-mol glucose,  $\Delta dic1$  experimental vs.  $\Delta dic1$  anaerobic simulation, respectively). This represents a >10-fold improvement in succinate productivity based exclusively on a novel *in silico* prediction.

### ***Transcriptome Characterization of Gene Deletion Strains***

To gain further insight into the physiological performance of each strain identified via simulation results, genome-wide DNA microarray profiling was completed under anaerobic batch glucose fermentations. Table 4 provides an overall summary of the comparative transcriptome of differentially expressed genes between  $\Delta dic1$ ,  $\Delta mdh1$ , and  $\Delta oac1$  strains, each compared to the reference strain. The number of differentially expressed genes for the  $\Delta oac1$  strain compared to the reference strain was very low, and consequently suggests that deletion of  $\Delta oac1$  causes virtually no transcriptional, and consequently, physiological differences compared to the reference BY4741 strain. The  $\Delta dic1$  and  $\Delta mdh1$  strains, compared to the reference strain, had 117 and 209 differentially expressed genes, respectively. Of these genes a total of 33% and 23% were up-regulated genes and

66% and 76% were down-regulated genes, for the  $\Delta dic1$  and  $\Delta mdh1$  strains, respectively. The average fold change of differentially expressed genes for the  $\Delta dic1$  strain, both up- and down-regulated, was  $\approx 2.5$ -fold greater than  $\Delta mdh1$ . Given the relatively low differential expression for the  $\Delta oac1$  strain, no further transcriptional analysis was performed for this strain.

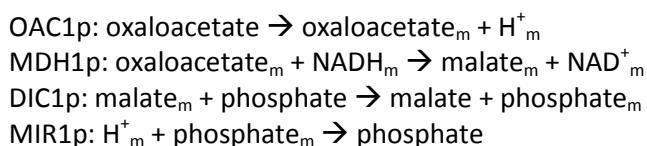
The differentially expressed genes sets for  $\Delta dic1$  and  $\Delta mdh1$  were submitted for gene ontology (GO) process annotation. Table 5 presents the statistically significant GO process annotation terms, showing a high degree of similarity for the two strains, with changes mainly in genes involved in energy metabolism and electron transport. It's particularly interesting to note that there is a large overlap for the two strains and there were only four GO process categories that were unique to  $\Delta mdh1$  as compared to  $\Delta dic1$ , and these are involved in sterol transport, lipid transport, generation of precursor metabolites and energy, and energy derivation by oxidation of organic compounds.

Given the high degree of similarity in the GO process annotation for both the  $\Delta dic1$  and  $\Delta mdh1$  conditions, the complete list of differentially expressed genes were submitted for metabolic pathway annotation using the SGD Pathway Expression Viewer and Reactome databases (Paley, et al., 2006; Matthew, et al., 2009). The results are presented in Supplementary Data 2 with color coding of genes according to their log-fold change and direction of expression relative to the reference case. What is immediately apparent is the relatively small number of total metabolic pathway genes identified in  $\Delta dic1$  and  $\Delta mdh1$  compared to the reference, with a total of 10 and 20 genes identified as catalyzing metabolic reactions, respectively. Perhaps more striking is that there is an overlap of 9 metabolic pathway genes between both  $\Delta dic1$  and  $\Delta mdh1$ . The only differentially expressed gene present in the  $\Delta dic1$  condition, not present in the  $\Delta mdh1$  condition, is  $\Delta dic1$ .

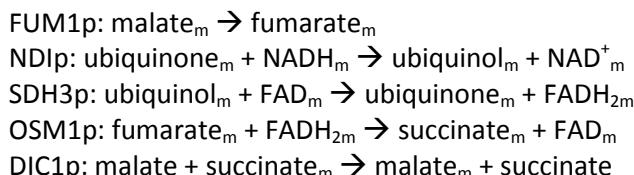
## Discussion

Succinic acid overproduction metabolic engineering strategies in *S. cerevisiae* are limited. Although a previous study to identify metabolic engineering strategies for succinic acid overproduction in *S. cerevisiae* using FBA on a reduced genome-scale model has been published (Patil, et al, 2005), attempts to reproduce those results using the complete iFF708 resulted in significantly reduced succinate yields on substrate than earlier found, and could only be obtained if a constraint preventing acetaldehyde secretion was imposed (data not shown). Here we evaluated specifically anaerobic growth conditions, and the top single gene deletion targets identified resulted in significantly higher succinate yields on substrate.

The metabolic engineering strategies identified through deletion of *Δdic1*, *Δmdh1*, and *Δoac1*, suggest a common mechanism that was identified via visualization of the central carbon metabolism. As described earlier, mitochondrial redox balance must be maintained and while respiratory metabolic activity under anaerobic conditions is reduced compared to aerobic conditions, some activity is required to support glutamate/glutamine metabolism from α-keto-glutarate (Camarasa, et al, 2003; Camarsa, et al, 2007), which produces NADH. During anaerobic metabolism, NAD<sup>+</sup> regeneration occurs via the following pathways, where the subscript *m* denotes mitochondrial:



In the cytosol malate is then converted to oxaloacetate, and the resulting NADH is converted to NAD<sup>+</sup> with the production of glycerol. If we consequently assume that *Δmdh1* were deleted, than NAD<sup>+</sup> regeneration would plausibly be completed via the following pathway reactions:



The above mechanism is highly dependent on several metabolic pathway assumptions, particularly that there are no other mitochondrial reactions capable of NAD<sup>+</sup><sub>*m*</sub> regeneration. Also, the *Δmdh1* strategy is highly sensitive to *r*<sub>O<sub>2</sub></sub>, as shown in Figure 4, because of the succinate production driven requirement for electron donation from ubiquinone to succinate and not oxygen. Even small values of *r*<sub>O<sub>2</sub></sub> (<0.1 mmol O<sub>2</sub>/g-DCW/h) result in no succinate production. If *COX1* (encoding subunit I of cytochrome C oxidase) and *RIP1* (encoding ubiquinol cytochrome C reductase) are deleted in combination with *MDH1*, to eliminate oxygen reactivity, the *r*<sub>O<sub>2</sub></sub> range across which succinate yield is observed for the *Δmdh1* strategy is extended to 0.6 mmol O<sub>2</sub>/g-DCW/h (See Figure 4). Furthermore, it should be noted that additional multi-gene deletion strategies leveraging the general *Δmdh1* strategy could be expanded. A simple triple gene deletion strategy of *Δmdh1 Δcat2 Δcit2* was simulated (data not shown), and resulted in further improved succinate yield on glucose (0.08 C-mol/C-mol glucose vs. 0.03 C-mol/C-mol glucose for only *Δmdh1*). *CAT2* and *CIT2* encode carnitine acetyl-CoA transferase and citrate synthase, respectively.

The *Δdic1* strategy, relying on deletion of the mitochondrial dicarboxylate carrier DIC1p, catalyzes the following transport reaction, noting the intermediate transport of orthophosphate:

Dic1: malate + succinate<sub>m</sub> → malate<sub>m</sub> + succinate  
 (malate + orthophosphate<sub>m</sub> → malate<sub>m</sub> + orthophosphate)  
 (succinate + orthophosphate<sub>m</sub> → succinate<sub>m</sub> + orthophosphate)

Assuming *DIC1* deletion, then the likely pathway is:

NDIp: ubiquinone<sub>m</sub> + NADH<sub>m</sub> → ubiquinol<sub>m</sub> + NAD<sup>+</sup><sub>m</sub>  
 SDH3p: ubiquinol<sub>m</sub> + FAD<sub>m</sub> → ubiquinone<sub>m</sub> + FADH<sub>2m</sub>  
 FRDS1p: fumarate + FADH<sub>2m</sub> → succinate + FAD<sub>m</sub>

*Net Reaction Stoichiometry:* fumarate + NADH<sub>m</sub> → succinate + NAD<sup>+</sup><sub>m</sub>

The *Δdic1* strategy relies heavily on the compartmental localization and function of FRDS1p, soluble mitochondrial fumarate reductase, which continues to be poorly understood. However, recent work has suggested that a double deletion *S. cerevisiae* mutant, *Δosm1 Δfrds1*, failed to grow under batch glucose anaerobic conditions. Furthermore, during anaerobic growth, *FRDS1* expression in the wild-type was two to eight times higher than that of *OSM1*, suggesting that formation of succinate is strictly required for the reoxidation of FADH<sub>2</sub> and its expression may be oxygen-regulated (Camarasa, et al, 2007). While neither *FRDS1* nor *OSM1* were significantly differentially expressed in the *Δmdh1* or *Δdic1* mutants compared to the reference strain, *FRDS1* was slightly up-regulated in the *Δdic1* mutant compared to the *Δmdh1* mutant ( $\log_{10}$  fold change 0.11 vs. -0.10, respectively). Lastly, as shown (See Supplementary Data 2) there was strong up-regulation of *CYC1* in both the *Δdic1* and *Δmdh1* mutants, suggesting that electron transport from ubiquinone cytochrome C oxidoreductase to cytochrome C oxidase was up-regulated, and required to facilitate electron transfer from NADH<sub>m</sub> to NAD<sup>+</sup><sub>m</sub>, and then from FADH<sub>2m</sub> to FADH<sub>m</sub> resulting in succinate formation. It has been well established that *CYC1* is both glucose repressed and regulated by the presence of oxygen and heme (Hortner, et al., 1982; Guarente, et al, 1983; Boss, et al, 1980; Guarente, et al, 1984). Therefore, strong up-regulation during anaerobic batch glucose fermentations in combination with deletion of *DIC1* may have aided in the increased succinate formation observed. However, this does not explain the lack of succinate production observed in the *Δmdh1* mutant. It has been suggested that mitochondrial FADH<sub>2</sub> could be oxidized in the cytosol, which may provide an explanation for the failure of the *Δmdh1* and *Δoac1* mutants to produce any succinate (Enomoto, et al, 2002). In any event, the strategies proposed here rely on the capacity for reductive TCA cycle activity under anaerobic conditions, and more specifically, the catalysis of fumarate to succinate via fumarate reductase. There is data suggesting that *S. cerevisiae* can exhibit this metabolic state (Camarasa, et al, 2003; Camarasa, et al, 2007).

A complete genome-scale metabolic reconstruction was used to predict single deletion strategies that could lead to increased succinate production that were physiologically feasible during anaerobic growth. Three of these strategies were validated *in vivo* and one, *Δdic1*, was identified to lead to a 10-fold improvement in succinate yield on substrate, in close agreement with the model prediction. Furthermore, pathway visualization, coupled with physiological characterization and transcriptome analysis were used to propose biological mechanisms. The mechanisms proposed rely heavily on intercompartmental transport reactions as well as redox balancing, both identified as the dominant GO process categories in the *Δdic1* succinate overproducing mutant. Further *in vivo* characterization of the transport reactions, and subsequent corresponding modifications to the genome-scale network reconstruction would be required for further improvements and understanding of metabolic engineering strategies.

## **Acknowledgements**

José Manuel Otero is a Merck Doctoral Fellow and acknowledges financial support from Merck Research Labs, Merck & Co., Inc.

## References

1. Adrio JL, Demain AL. 2006. Genetic improvement of processes yielding microbial products. FEMS Microbiol Rev 30: 187-214.
2. Aliverdieva DA, Mamaev DV, Bondarenko DI, Sholtz KF. 2006. Properties of yeast *Saccharomyces cerevisiae* plasma membrane dicarboxylate transporter. Biochemistry (Mosc.) 71: 1161-1169.
3. Arikawa Y, Enomoto K, Muratsubaki H, Okazaki M. 1998. Soluble fumarate reductase isoenzymes from *Saccharomyces cerevisiae* are required for anaerobic growth. FEMS Microbiology Letters 165: 111-116.
4. Arikawa Y, Kobayashi M, Kodaira R, Shimosaka M, Muratsubaki H, Enomoto K, Okazaki M. 1999(b). Isolation of Sake Yeast Strains Possessing Various Levels of Succinate- and/or Malate-Producing Abilities by Gene Disruption or Mutation. Journal of Bioscience and Bioengineering 87: 333-339.
5. Arikawa Y, Kuroyanagi T, Makoto S, Haruhiro M, Enomoto K, Kodaira R, Okazaki M. 1999(a). Effect of Gene Disruptions of the TCA Cycle on Production of Succinic Acid in *Saccharomyces cerevisiae*. Journal of Bioscience and Bioengineering 87: 28-36.
6. Bailey JE. 1991. Toward a science of metabolic engineering. Science 252: 1668-1675.
7. Becker SA, Fesit AM, Mo ML, Hannum G, Palsson BØ, Herrgård MJ. 2007. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox. Nat Protocols 2: 727-738.
8. Blank LM, Kuepfer L, Sauer U. 2005. Large-scale <sup>13</sup>C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. Genome Biol 6: R49.
9. Boss JM, Darrow MD, Zitomer RS. 1980. Characterization of yeast iso-1-cytochrome C mRNA. J Biol Chem 255: 8623-8628.
10. Camarasa C, Faucet V, Dequin S. 2007. Role in anaerobiosis of the isoenzymes for *Saccharomyces cerevisiae* fumarate reductase encoded by *OSM1* and *FRDS1*. Yeast 24: 391-401.
11. Camarasa C, Grivet J-P, Dequin S. 2003. Investigation by <sup>13</sup>C-NMR and tricarboxylic acid (TCA) deletion mutant analysis of pathways for succinate formation in *Saccharomyces cerevisiae* during anaerobic fermentation. Microbiology 149: 2669-2678.
12. Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, Jia Y, Juvik G, Roe T, Schroeder M, Weng S, Botstein D. 1998. SGD: Saccharomyces Genome Database. Nucleic Acids Res 26: 73-79.
13. Cimini D, Patil KR, Schiraldi C, Nielsen J. 2009. Global transcriptional response of *Saccharomyces cerevisiae* to the deletion of *SDH3*. BMC Systems Biology 3: 1-12.
14. Duarte NC, Palsson BØ, Fu P. 2004. Integrated analysis of metabolic phenotypes in *Saccharomyces cerevisiae*. BMC Genomics 5: 63.
15. Edwards JS, Covert M, Palsson BØ. 2002(a). Metabolic modeling of microbes: the flux-balance approach. Environ Microbiol 4: 133-140.
16. Edwards JS, Ibarra RU, Palsson BØ. 2001. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. Nat Biotechnol 19: 125-130.
17. Edwards JS, Ramakrishna R, Palsson BØ. 2002(b). Characterizing the metabolic phenotype: a phenotype phase plane analysis. Biotechnol Bioeng 77: 27-36.
18. Enomoto K, Arikawa Y, Muratsubaki H. 2002. Physiological role of soluble fumarate reductase in redox balancing during anaerobiosis in *Saccharomyces cerevisiae*. FEMS Microbiology Letters 215: 103-108.
19. Famili I, Förster J, Nielsen J, Palsson BØ. 2003. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc Natl Acad Sci USA 100: 13134-13139.

20. Fisk DG, Ball CA, Dolinski K, Engel SR, Hong EL, Issel-Tarver L, Schwartz K, Sethuraman A, Botstein D, Cherry JM; Saccharomyces Genome Database Project. 2006. *Saccharomyces cerevisiae* S288C genome annotation: a working hypothesis. Yeast 23: 857-65.
21. Förster J, Famili I, Fu P, Palsson BØ, Nielsen J. 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res 13: 244-253.
22. Förster J, Famili I, Palsson BØ, Nielsen J. 2003. Large-scale evaluation of *in silico* gene deletions in *Saccharomyces cerevisiae*. OMICS 7: 193-202.
23. Funahashi A, Tanimura N, Morohashi M, Kitano H. 2003. CellDesigner: A process diagram editor for gene-regulatory and biochemical networks. Biosilico1: 159-162.
24. Geertmann J-MA, van Maris AJA, van Dijken JP, Pronk JT. 2006. Physiological and genetic engineering of cytosolic redox metabolism in *Saccharomyces cerevisiae* for improved glycerol production. Metabolic Engineering 8: 532-542.
25. Giorgetti N. 2008. Hamtat fran GLPKMEX: A Matlab MEX Interface for the GLPK library. <http://glpkmex.sourceforge.net>, last accessed July 2008.
26. Guarente L, Lalonde B, Gifford P, Alani E. 1984. Distinctly regulated tandem upstream activation sites mediate catabolite repression of the *CYC1* gene of *S. cerevisiae*. Cell 36: 503-511.
27. Guarente L, Mason T. 1983. Heme regulated transcription of the *CYC1* gene of *S. cerevisiae* via an upstream activation site. Cell 32: 1279-1286.
28. Herrgård MJ, Swainston N, Dobson P, Dunn WB, Argas M, Blüthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novère N, Li P, Liebermeister W, Mo ML, Oliveira AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K, Spasić I, Weichert D, Brent R, Broomhead DS, Westerhoff HV, Kirdar B, Penttilä M, Klipp E, Palsson BØ, Sauer U, Oliver SG, Mendes P, Nielsen J, Kell DB. 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat Biotechnol 26: 1155-60.
29. Hörtner H, Ammerer G, Hartter E, Hamilton B, Rytka J, Bilinski T, Ruis H. 1982. Regulation and synthesis of catalases and iso-1-cytochrome C in *Saccharomyces cerevisiae* by glucose, oxygen, and heme. Eur J Biochem 128: 179-184.
30. Jantama K, Haupt MJ, Svoronos SA, Zhang X, Moore JC, Shanmugam KT, Ingram LO. 2008. Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* C that produce succinate and malate. Biotechnol Bioeng 99: 1140-1153.
31. Keating SM, Bornstein BJ, Finney A, Hucka M. 2006. SBMLToolbox: An SBML toolbox for MATLAB users. Bioinformatics 22: 1275-1277.
32. Kim TY, Kim HU, Park JM, Song H, Kim JS, Lee SY. 2007. Genome-scale analysis of *Mannheimia succiniciproducens* metabolism. Biotechnol Bioeng 97: 657-671.
33. Kubo Y, Takagi H, Nakamori S. 2000. Effect of gene disruption of succinate dehydrogenase on succinate production in a sake yeast strain. Journal of Bioscience and Bioengineering 90: 619-624.
34. Lee PC, Lee SY, Hong SH, Chang HN. 2002. Isolation and characterization of a new succinic acid-producing bacterium, *Mannheimia succiniciproducens* MBEL55E, from bovine rumen. Appl Microbiol Biotechnol 58: 663-668.
35. Lee PC, Lee SY, Hong SH, Chang HN. 2003. Batch and continuous cultures of *Mannheimia succiniciproducens* MBEL55E for the production of succinic acid from whey and corn steep liquor. Bioprocess Biosyst Eng 26: 63-67.
36. Lee SJ, Song H, Lee SY. 2006. Genome-based metabolic engineering of *Mannheimia succiniciproducens* for succinic acid production. Appl Environ Microbiol 72: 1939-1948.

37. Luttik MAH, Overkamp KM, Kotter P, de Vries S, van Dijken JP, Pronk JT. 1998. The *Saccharomyces cerevisiae* NDE1 and NDE2 genes encode separate mitochondrial NADH dehydrogenases catalyzing the oxidation of cytosolic NADH. *J Biol Chem* 273: 24529-24534.
38. Lynd LR, Wyman CE, Gerngross TU. 1999. Biocommodity Engineering. *Biotechnol Prog* 15: 777-793.
39. Matthews L, Gopinath G, Gillespie M, Caudy M, Croft D, de Bono B, Garapati P, Hemish J, Hermjakob H, Jassal B, Kanapin A, Lewis S, Mahajan S, May B, Schmidt E, Vastrik I, Wu G, Birney E, Stein L, D'Eustachio P. 2009. Reactome knowledgebase of biological pathways and processes. *Nucleic Acid Res* 37: D619-632.
40. McKinlay JB, Vieille C, Zeikus JG. 2007. Prospects for a bio-based succinate industry. *Appl Microbiol Biotechnol* 76: 727-740.
41. Nielsen J, Villadsen J, Liden G. 2003. *Bioreaction Engineering Principles*, 2<sup>nd</sup> edition. New York: Kluwer Academic/Plenum Publishers. 63 p.
42. Nielsen J. 2001. Metabolic engineering. *Appl Microbiol Biotechnol* 55: 262-283.
43. Nissen TL, Schulze U, Nielsen J, Villadsen J. 1997. Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*. *Microbiology* 143: 203-218.
44. Nookaew I, Jewett MC, Meechai A, Thammarongtham C, Laoteng K, Cheevadahanarak S, Nielsen J, Bhumiratana S. 2008. The genome-scale metabolic model iN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol* 2: 72.
45. Otero JM, Panagiotou G, Olsson L. 2007. Fueling Industrial Biotechnology Growth with Bioethanol. *Adv Biochem Eng Biotechnol* 108: 1-40.
46. Overkamp JM, Bakker BM, Kotter P, van Tuijl A, de Vries S, van Dijken JP, Pronk JT. 2000. *In vivo* analysis of the mechanism for oxidation of cytosolic NADH by *Saccharomyces cerevisiae* mitochondria. *J Bacteriol* 182: 2823-2830.
47. Paley SM, Karp PD. 2006. The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acid Res* 40: 3771-3778.
48. Patil KR, Rocha I, Förster J, Nielsen J. 2005. Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics* 6: 308.
49. Sauer M, Porro D, Mattanovich D, Branduardi P. 2008. Microbial production of organic acids: expanding the markets. *Trends Biotechnol* 26: 100-108.
50. Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* 3:3.
51. Smyth GK. 2005. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. New York: p 397-420.
52. Song H, Lee SJ. 2006. Production of succinic acid by bacterial fermentation. *Enzyme and Microbial Technology* 39: 352-361.
53. Stephanopoulos G, Vallino JJ. 1991. Network rigidity and metabolic engineering in metabolite overproduction. *Science* 252: 1675-1681.
54. Tyo KE, Alper HS, Stephanopoulos GN. 2007. Expanding the metabolic engineering toolbox: more options to engineer cells. *Trends Biotechnol* 25: 132-137.
55. US Department of Energy (DOE), Top Value Added Chemicals from Biomass: Volume I, available at <http://www1.eere.energy.gov/biomass/pdfs/35523.pdf>, last accessed May 5, 2008.
56. van Dijken JP, Bauer J, Brambilla L, Duboc P, Francois JM, Gancedo C, Giuseppin ML, Heijnen JJ, Hoare M, Lange H, Madden EA, Niederberger P, Nielsen J, Parrou JL, Petit T, Porro D, Reuss M, van Riel N, Rizzi M, Steensma HY, Verrips CT, Vindeløv J, Pronk JT. 2000. An

- interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. Enzyme and Microbial Technology 26(9-10): 706-714.
57. Vemuri GN, Eiteman MA, McEwen JE, Olsson L, Nielsen J. 2007. Increasing NADH oxidation reduces overflow metabolism in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 104: 2402-2407.
58. Verudyn 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.
59. von Jagow G, Klingenberg M. 1970. Hydrogen pathways in the mitochondrion of *Saccharomyces carlsbergensis*. Eur J Biochem 12: 585-832.
60. Wilke T, Vorlop KD. 2004. Industrial bioconversion of renewable resources as an alternative to conventional chemistry. Appl Microbiol Biotechnol 66: 131-42.
61. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentlemen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Véronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW. 1999. Functional Characterization of the *Saccharomyces cerevisiae* Genome by Gene Deletion and Parallel Analysis. Science 285: 901-906.
62. Zeikus JG, Jain M, Elankovan K, Elankovan P. 1999. Biotechnology of succinic acid production and markets for derived industrial products. Applied Microbiology and Biotechnology 51: 545-552.
63. Åkesson M, Förster J, Nielsen J. 2004. Integration of gene expression data into genome-scale metabolic models. Metabolic Engineering 6: 285-298.

**Table 1. *Saccharomyces cerevisiae* Strain Description and Genotype**

Strain Name	Strain Genotype	Source
<b>CEN.PK113-7D</b>	<i>MATa URA3 HIS3 LEU2 TRP1 SUC2 MAL2-8<sup>c</sup></i>	SRD GmbH <sup>1</sup>
<b>Reference (REF)</b>	BY4741: <i>MATa; his3Δ1; leu2Δ0' met15Δ0; ura3Δ0</i>	
<b>ΔMDH1</b>	BY4741: <i>MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YKL085w:kanMX4</i>	
<b>ΔOAC1</b>	BY4741: <i>MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YKL120w:kanMX4</i>	
<b>ΔDIC1</b>	BY4741: <i>MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0; YLR348c:kanMX4</i>	
<b>ΔSDH3<sup>2</sup></b>	BY4743: <i>MATa/α; his3Δ1/ his3Δ1; leu2Δ0/ leu2Δ0; met15Δ0/ met15Δ0; ura3Δ0/ ura3Δ0; YKL141w:kanMX4/YKL141w</i>	

1. Scientific Research and Development GmbH (Oberursel, Germany)
2. European Saccharomyces Cerevisiae Archive for Functional Analysis (Frankfurt, Germany)
3. All strains were haploid of mating typ a, with the exception of the ΔSDH3 strain, which is diploid and mating type a/α. A haploid strain of ΔSDH3 was reported as not viable.

**Table 2. Physiological Characterization**

<i>Strain</i>	CEN.PK113-7D	CEN.PK113-7D	Reference (BY4741)	<i>Δmdh1</i>	<i>Δdic1</i>	<i>Δoac1</i>	<i>Δsdh3</i>
<i>Conditions</i>	Aerobic ± Std. Dev. (n=2)	Anaerobic ± Std. Dev. (n=2)	Anaerobic ± Std. Dev. (n=4)	Anaerobic			
<i>Specific growth rate (h<sup>-1</sup>)</i>	0.38 ± 0.00	0.29 ± 0.01	0.28 ± 0.00	0.28 ± 0.01	0.24 ± 0.02	0.23 ± 0.00	0.26
<i>Productivities (C-mmol/g-DCW/h)</i>							
<i>r<sub>Gluc</sub></i>	91.2 ± 6.0	93.1 ± 4.0	89.7 ± 2.8	68.6 ± 18.2	74.5 ± 2.0	73.3 ± 3.0	68.5
<i>r<sub>EtOH</sub></i>	49.7 ± 6.6	41.0 ± 5.5	40.8 ± 0.8	32.9 ± 8.7	37.4 ± 1.9	39.2 ± 3.4	36.0
<i>r<sub>CO<sub>2</sub></sub></i>	15.4 ± 0.0	11.1 ± 0.0	11.3 ± 3.3	9.7 ± 2.1	7.9 ± 0.1	9.9 ± 0.1	12.0
<i>r<sub>Acet</sub></i>	0.7 ± 0.3	0.4 ± 0.1	0.3 ± 0.4	0.3 ± 0.1	0.2 ± 0.1	0.3 ± 0.1	0.4
<i>r<sub>Glyc</sub></i>	7.3 ± 3.4	10.5 ± 2.2	3.5 ± 4.8	5.9 ± 1.6	6.7 ± 0.3	6.7 ± 0.3	7.3
<i>r<sub>Suc</sub></i>	0 ± 0.0	0 ± 0.0	0.1 ± 0.0	0.1 ± 0.0	1.6 ± 0.1	0.1 ± 0.1	0.3
<i>r<sub>Pyr</sub></i>	0.4 ± 0.0	0.3 ± 0.0	0.1 ± 0.1	0.2 ± 0.1	0.1 ± 0.1	0.2 ± 0.1	0.2
<i>r<sub>O<sub>2</sub></sub></i>	1.8 ± 0.3	0 ± 0.0	0.1 ± 0.2	0.9 ± 0.0	0.2 ± 0.2	0.0 ± 0.2	0.4
<i>Carbon Recovery (%)</i>	96.3 ± 4.0	81.3 ± 3.2	79.0 ± 6.2	88.6 ± 6.7	84.7 ± 1.0	89.6	100.3

**Table 3. Top Gene Deletions under Anaerobic Constraints for Succinate Yield**

Simulation Conditions	Genotype	Specific Growth Rate [h <sup>-1</sup> ]	$\gamma_{SSuc}$ [C-mol/C-mol glucose]
TOP SINGLE GENE DELETIONS ANAEROBIC: Simulations with oxygen uptake rate constrained to 0 mmol/g-DCW/h	No deletions	0.29	0.003
	$\Delta oac1$	0.26	0.045
	$\Delta mdh1$	0.26	0.045
	$\Delta dic1$	0.26	0.044
	$\Delta fum1$	0.29	0.005
	$\Delta met22$	0.29	0.004
	$\Delta atp1$	0.29	0.003
	$\Delta zwf1$	0.29	0.003
TOP SINGLE GENE DELETIONS ANAEROBIC: Simulations with oxygen uptake rate set to 0.0163 mmol/g-DCW/h	No deletions	0.30	0.003
	$\Delta oac1$	0.28	0.033
	$\Delta mdh1$	0.28	0.033
	$\Delta dic1$	0.28	0.032
	$\Delta fum1$	0.30	0.005
	$\Delta met22$	0.30	0.004
	$\Delta aap1$	-	-
	$\Delta zwf1$	-	-
TOP DOUBLE GENE DELETIONS ANAEROBIC: Simulations with oxygen uptake rate constrained to 0.0163 mmol/g-DCW/h	No deletions	0.29	0.003
	$\Delta mdh1\Delta yat1$	0.25	0.061
	$\Delta mdh1\Delta cat2$	0.25	0.061
	$\Delta dic1\Delta yat1$	0.25	0.060
	$\Delta dic1\Delta cat2$	0.25	0.060
	$\Delta dic1\Delta cit2$	0.25	0.056
	$\Delta mdh1\Delta put2$	0.26	0.051
	$\Delta mdh1\Delta kgd1$	0.26	0.050
	$\Delta mdh1\Delta lsc2$	0.26	0.050
	$\Delta dic1\Delta oac1$	0.25	0.051
	$\Delta dic1\Delta lsc2$	0.26	0.049

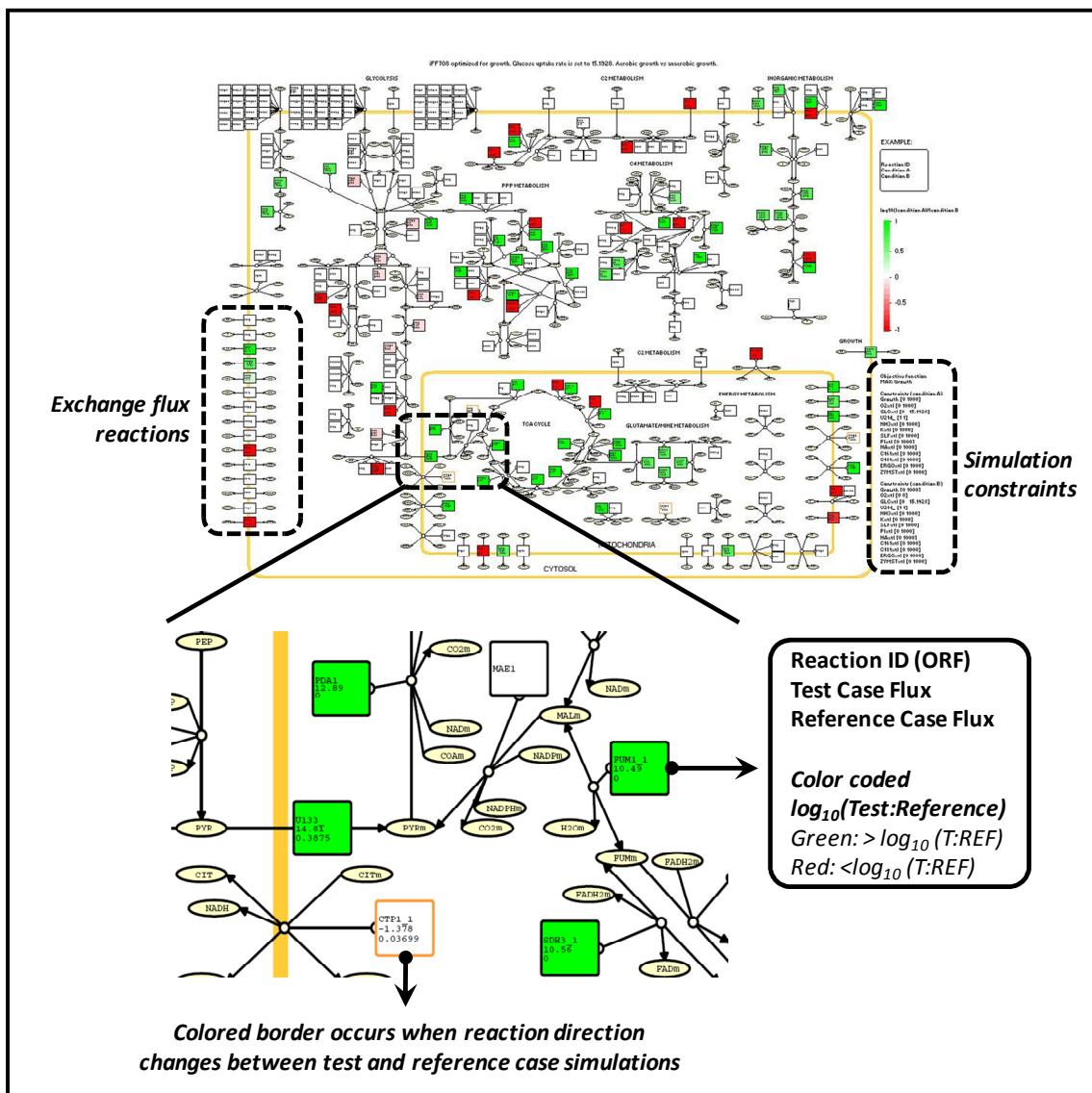
**Table 4. Summary of Differentially Expressed Genes**

<b>Comparative Transcriptome</b>	$\Delta\text{DIC1}$ vs. REF ( <i>n</i> =2)	$\Delta\text{MDH1}$ vs. REF ( <i>n</i> =2)	$\Delta\text{OAC1}$ vs. REF <sup>1</sup> ( <i>n</i> =2)
<b>No. Differentially Expressed Genes (<i>p-value<sub>B-H</sub></i>&lt;0.01)</b>	117	209	5
<i>Up-regulated</i>	39	49	3
<i>Down-regulated</i>	78	160	2
<b>Average Log-Fold Change (<math>\pm</math> Std. Dev)</b>			
<i>Up-regulated</i>	1.45 (1.61)	1.09 (0.60)	-
<i>Down-regulated</i>	-1.98 (1.53)	-1.55 (1.23)	-

**NOTES:** 1. For the comparison of  $\Delta\text{OAC1}$  vs. REF, the  $p-value_{BH}<0.1$  criteria was applied and resulted in only 5 differentially expressed genes. Given the low number of differentially expressed genes no average log-fold change is reported.

**Table 5. Process Gene Ontology Annotation of Differentially Expressed Genes of  $\Delta DIC1$ :REF and  $\Delta MDH1$ :REF**

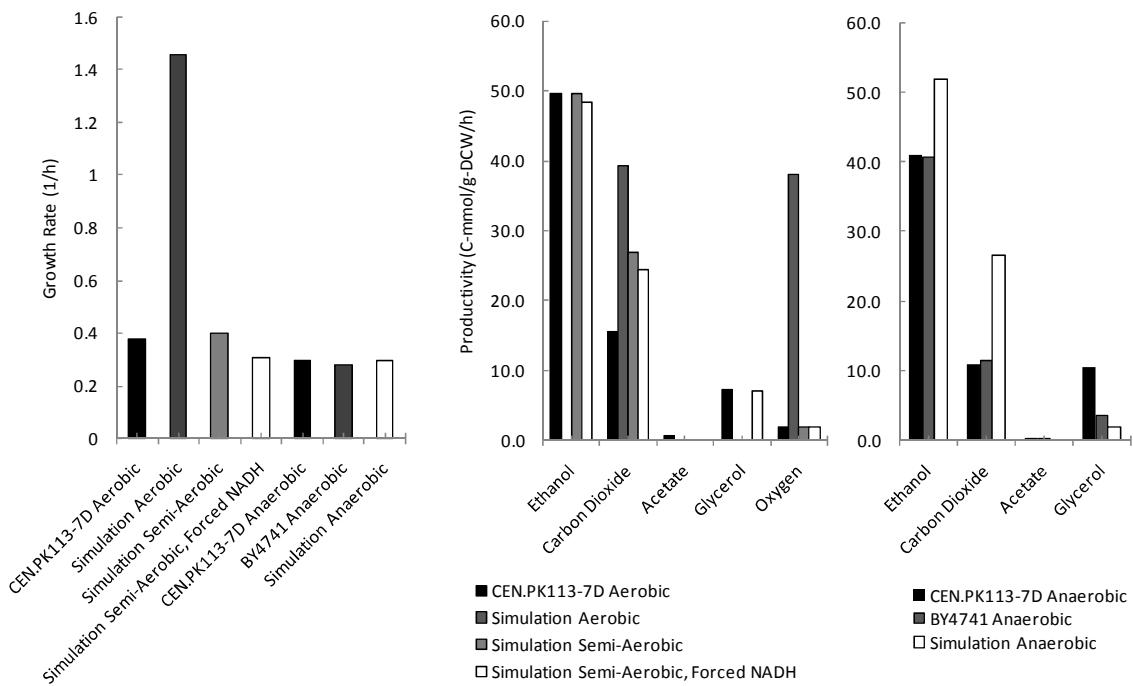
<i>Gene Ontology</i>	Genes Annotated in $\Delta dic1$	Genes Annotated in $\Delta mdh1$	<i>p-value</i> $\Delta dic1$	<i>p-value</i> $\Delta mdh1$
<i>mitochondrial electron transport, cytochrome C to oxygen</i>	COX4, COX6, CYC1, COX7, COX5A	COX4, COX6, CYC1, COX7, COX5A	1.20E-04	4.75E-03
<i>electron transport chain</i>	COX4, QCR10, COX6, CYC1, COX7, COX5A	COX4, QCR10, COX6, CYC1, COX7, COX5A, QCR2	5.80E-04	3.04E-03
<i>respiratory electron transport chain</i>	COX4, QCR10, COX6, CYC1, COX7, COX5A	COX4, QCR10, COX6, CYC1, COX7, COX5A, QCR2	5.80E-04	3.04E-03
<i>ATP synthesis coupled electron transport</i>	COX4, QCR10, COX6, CYC1, COX7, COX5A	COX4, QCR10, COX6, CYC1, COX7, COX5A, QCR2	5.80E-04	3.04E-03
<i>mitochondrial ATP synthesis coupled electron transport</i>	COX4, QCR10, COX6, CYC1, COX7, COX5A	COX4, QCR10, COX6, CYC1, COX7, COX5A, QCR2	5.80E-04	3.04E-03
<i>oxidation reduction</i>	COX4, QCR10, COX6, CYC1, COX7, COX5A	COX4, QCR10, COX6, CYC1, COX7, COX5A, QCR2	5.80E-04	3.04E-03
<i>sterol transport</i>	-	SWH1, SUT1, PDR11, DAN1, AUS1, HES1, SUT2	-	5.14E-05
<i>energy derivation by oxidation of organic compounds</i>	-	PET9, HOR2, BMH1, COX4, COX13, QCR10, COX6, PIG2, CYC1, MDH1, PET10, PUF3, NDE1, COX7, COX5A, QCR2	-	1.84E-03
<i>generation of precursor metabolites and energy</i>	-	PET9, HOR2, BMH1, HXK1, COX4, COX13, QCR10, COX6, PIG2, CYC1, MDH1, PET10, PUF3, NDE1, COX7, COX5A, PFK27, QCR2	-	2.65E-03
<i>lipid transport</i>	-	SWH1, SUT1, PDR11, DAN1, AUS1, HES1, FAA1, SUT2	-	2.94E-03



**Figure 1. Visualization of Succinate Specific Metabolic Pathways**

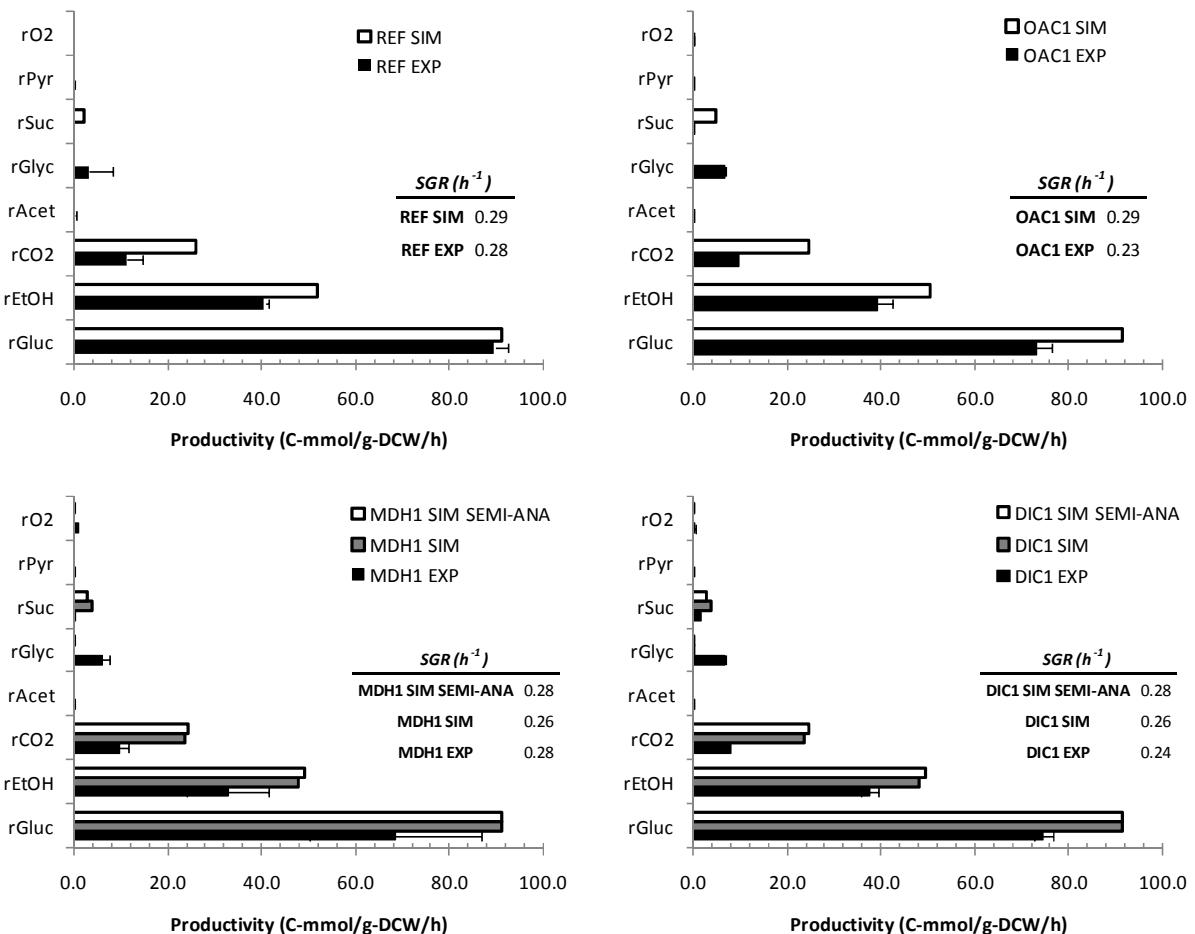
Illustrated above is the user-specified metabolic pathway map drawn in CellDesigner™, converted to a MATLAB® compatible format and upon-which FBA simulation data for a test and reference case can be overlaid. The above map represents a simulation of aerobic v. anaerobic glucose conditions with optimization for growth. Specific metabolic pathways, considered in two compartments (mitochondria, cytosol), portrayed to facilitate simulated strategies for succinate overproduction include: glycolysis, C<sub>2</sub>-metabolism, pentose phosphate pathway, tricarboxylic acid cycle, energy metabolism, relevant transporters, exchange flux reactions, and a reaction representing biomass formation (e.g., referred to as *growth*). The map also includes a convenient reaction specific information box that includes the reaction ID (ORF reference from original iFF708), the test case flux value, the reference case flux value,

and a  $\log_{10}$  ratio of both values (test:reference flux values). These boxes are correspondingly color-coded to provide an immediate visual summary of how test case fluxes compare to reference case fluxes. Furthermore, the border of the information box, if colored yellow, indicates a directional change in flux, specifically confirmed by a positive to negative, or vice versa, flux value change in the test case relative to the reference. The simulation clearly shows, as expected, that TCA cycle and respiratory metabolism is more active (indicated by green) as compared to fermentation (indicated by red).



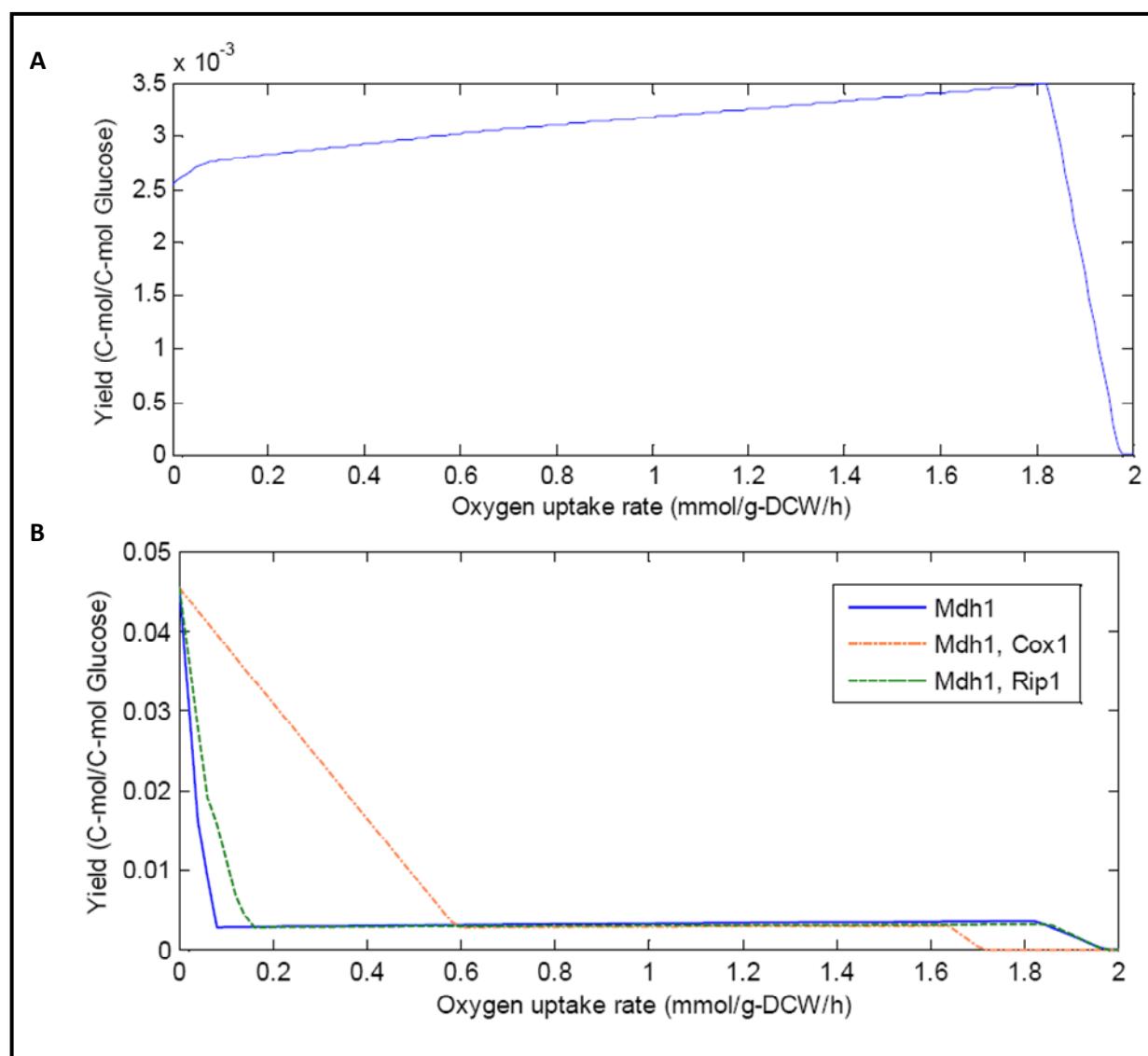
**Figure 2. Experimental and Simulation Comparative Reference Data**

Comparison of the specific growth rate and specific productivities for experimental data generated using the reference *S. cerevisiae* CEN.PK113-7D and BY4741 under aerobic and anaerobic glucose batch fermentations, and simulation data. For the condition, *Simulation Aerobic*, *Simulation Semi-Aerobic*, *Simulation Anaerobic*, the  $r_{O_2}$  was unconstrained (0-1000 mmol-O<sub>2</sub>/g-DCW/h), constrained to 1.8 mmol-O<sub>2</sub>/g-DCW/h, and constrained to 0 mmol-O<sub>2</sub>/g-DCW/h, respectively. The condition, *Simulation Semi-Aerobic, Forced NADH*, included the reaction FNADH constrained to 6 mmol-NADH/g-DCW/h. For aerobic experimental data the specific glucose uptake rate was 91.2 C-mmol/g-DCW/h for CEN.PK113-7D. For anaerobic experimental data the specific glucose uptake rate was 93.1 C-mmol/g-DCW/h for CEN.PK113-7D and 89.7 C-mmol/g-DCW/h for BY4741. For all simulation conditions the glucose uptake rate was constrained to 91.2 C-mmol/g-DCW/h.



**Figure 3. Experimental and Simulation Comparative Data for Reference,  $\Delta oac1$ ,  $\Delta mdh1$ , and  $\Delta dic1$  Strains**

Summary of the specific growth rate (SGR) and specific consumption/productivity values for major carbon products (glucose, ethanol, carbon dioxide, acetate, glycerol, succinate, pyruvate, and oxygen) for both experimentally determined data of anaerobic batch glucose fermentations, and corresponding anaerobic simulation data of the BY4741 reference strain, and single gene deletion strains  $\Delta mdh1$ ,  $\Delta dic1$ , and  $\Delta oac1$ . In general, the experimental data suggests a lower specific growth rate compared to the predicted growth rate, whether anaerobic simulations (referred to as *SIM*) or semi-anaerobic simulations (referred to as *SIM SEMI-ANA*) are considered. The simulation data for  $\Delta mdh1$  and  $\Delta dic1$  conditions attempt to highlight the significant sensitivity to relatively small changes in  $r_{O_2}$ , where the *SIM SEMI-ANA* simulation constrains  $r_{O_2}$  to 0.02 mmol  $O_2$ /g-DCW/L compared to 0 mmol  $O_2$ /g-DCWL/L, while impacting growth rate significantly. Both glucose and oxygen are consumed; however, are presented as positive values. Clearly, succinate production under simulation conditions was noted; however, only observed under the  $\Delta dic1$  experimental condition.



**Figure 4. Oxygen Sensitivity of Succinate Yield on Glucose**

Figure 4A is a plot of the succinate yield on glucose when  $r_{O_2}$  is constrained between 0 and 2 mmol O<sub>2</sub>/g-DCW/h, while maximizing growth as an objective function under constrained glucose uptake rate and no gene deletions (reference case). Figure 4B is a similar plot, although three independent genetic deletion combinations are considered: single gene deletion of *MDH1*, the double gene deletion of *MDH1* and *COX1*, and the double gene deletion of *MDH1* and *RIP1*. *MDH1* encodes malate dehydrogenase, *RIP1* encodes ubiquinol cytochrome C reductase, and *COX1* encodes subunit I of the cytochrome C oxidase.

**Supplementary Data 1. Top Single and Double Gene Deletions under Aerobic Constraints for Succinate Yield**

Simulation Conditions	Genotype	Specific Growth Rate [h <sup>-1</sup> ]	Y <sub>SSuc</sub> [C-mol/C-mol glucose]
<i>TOP SINGLE GENE DELETIONS</i> <i>AEROBIC: Simulations with oxygen uptake rate 1.778 mmol/g-DCW/h</i>	No deletions	0.40	0.003
	$\Delta fum1$	0.40	0.007
	$\Delta met22$	0.40	0.005
	$\Delta aus1$	0.40	0.004
<i>TOP SINGLE GENE DELETIONS</i> <i>AEROBIC BATCH: Simulations to model batch cultivations with oxygen uptake rate 1.778 mmol/g-DCW/h, and cytosolic NADH</i>	No deletions	0.30	0.003
	$\Delta fum1$	0.30	0.005
	$\Delta met22$	0.30	0.004
	$\Delta aus1$	0.30	0.003
<i>TOP DOUBLE GENE DELETIONS</i> <i>AEROBIC: Unconstrained oxygen uptake</i>	No deletions	1.45	0.000
	$\Delta fum1\Delta rip1$	0.98	0.016
	$\Delta sdh3\Delta ser1$	1.29	0.011
	$\Delta sdh3\Delta ser2$	1.29	0.011
	$\Delta sdh3\Delta fbp1$	1.28	0.011
	$\Delta sdh3\Delta tpi1$	1.28	0.011
	$\Delta sdh3\Delta lys20$	1.29	0.009
	$\Delta sdh3\Delta lys12$	1.29	0.009
	$\Delta sdh3\Delta lys4$	1.29	0.009
	$\Delta sdh3\Delta pgp1$	0.92	0.008
	$\Delta sdh3\Delta rpe1$	0.69	0.010

## Supplementary Data 2. Pathway Annotation of Differentially Expressed Genes

ΔDIC1 vs. REF (n=2)		ΔMDH1 vs. REF (n=2)		Enzyme Name	Reaction or Annotation
Standard Name	Systematic Gene Name	Standard Name	Systematic Gene Name		
HEM13	YDR044W	HEM13	YDR044W	Coproporphyrinogen III oxidase	Coproporphyrinogen III + Oxygen = Protoporphyrinogen IX + 2 CO <sub>2</sub> + 2H <sub>2</sub> O
SAM2	YDR502C	SAM2	YDR502C	S-adenosylmethionine synthetase	ATP + L-methionine + H <sub>2</sub> O = Pi + PPi + S-adenosyl-L-methionine
HXK1	YFR053C	HXK1	YFR053C	Hexokinase enzyme I	ATP + alpha-D-Glucose → ADP + alpha-D-glucose 6-phosphate
PDR11	YIL013C	PDR11	YIL013C	ATP-binding cassette (ABC) transporter	Mediates sterol uptake when sterol biosynthesis is compromised, regulated by Pdr1p, and required for anaerobic growth
CYC1	YJR048W	CYC1	YJR048W	Cytochrome c-, isoform 1	Electron carrier of the mitochondrial intermembrane space that transfers electrons from ubiquinone-cytochrome c oxidoreductase to cytochrome c oxidase during cellular respiration
CMK2	YOL016C	CMK2	YOL016C	Calmodulin-dependent protein kinase	May play a role in stress response
AUS1	YOR011W	AUS1	YOR011W	ATP-binding cassette (ABC) transporter	Involved in uptake of sterols and anaerobic growth
CRC1	YOR100C	CRC1	YOR100C	Mitochondrial inner membrane carnitine transporter	Required for carnitine-dependent transport of acetyl-CoA from peroxisomes to mitochondria during fatty acid beta-oxidation
CAR1	YPL111W	CAR1	YPL111W	Arginase	Catalyzes conversion of L-arginine to L-ornithine and urea, expression responds to both induction by arginine and nitrogen catabolite repression
DIC1	YLR348C	-	-	Mitochondrial dicarboxylate carrier	Catalyzes a dicarboxylate-phosphate exchange across the inner mitochondrial membrane, transports cytoplasmic dicarboxylates into the mitochondrial matrix
-	-	PET9	YBL030C	Mitochondrial ATP/ADP carrier	Exchanges cytosolic ADP for mitochondrially synthesized ATP
-	-	HEM3	YDL205C	Porphobilinogen deaminase	Catalyzes the conversion of 4-porphobilinogen to hydroxymethylbilane, the third step in heme biosynthesis; localizes to the cytoplasm and nucleus; expression is regulated by Hap2p-Hap3p, but not by levels of heme
-	-	CMK1	YFR014C	Calmodulin-dependent protein kinase	May play a role in stress response
-	-	VAM7	YGL212W	Component of the vacuole SNARE complex	Involved in vacuolar morphogenesis; SNAP-25 homolog; functions with a syntaxin homolog Vam3p in vacuolar protein trafficking
-	-	ERG9	YHR190W	Farnesyl-diphosphate farnesyl transferase (squalene synthase)	Catalyzes the reaction of two farnesyl pyrophosphate moieties to form squalene in the sterol biosynthesis pathway
-	-	MAD2	YJL030W	Component of the spindle-assembly checkpoint complex	Delays the onset of anaphase in cells with defects in mitotic spindle assembly; forms a complex with Mad1p
-	-	MDH1	YKL085W	Mitochondrial 32kDa dehydrogenase	Catalyzes interconversion of malate and oxaloacetate; involved in the tricarboxylic acid (TCA) cycle
-	-	YKU80	YMR106C	Yeast KU protein	Subunit of the telomeric Ku complex (Yku70p-Yku80p), involved in telomere length maintenance, structure and telomere position effect; relocates to sites of double-strand cleavage to promote nonhomologous end joining during DSB repair
-	-	CIR2	YOR356W	Mitochondrial protein with similarity to flavoprotein-type oxidoreductases	Found in a large supramolecular complex with other mitochondrial dehydrogenases
-	-	RLM1	YPL089C	MADS-box transcription factor	Component of the protein kinase C-mediated MAP kinase pathway involved in the maintenance of cell integrity; phosphorylated and activated by the MAP-kinase Slt2p
-	-	YMC1	YPR058W	Mitochondrial protein inner membrane transporter	Plays a role in oleate metabolism and glutamate biosynthesis; member of the mitochondrial carrier (MCF) family

**NOTES:** The dark and light red refer to up-regulated genes with a log-fold change of  $\geq 2.0$  and  $\geq 0.5$ , respectively, relative to the reference strain. The dark and light green refer to down-regulated genes

with a log-fold change of  $\leq 2.0$  and  $\leq 0.5$ , respectively, relative to the reference strain. All of the genes included on this list are statistically differentially expressed,  $p\text{-adjusted}_{B-H} < 0.01$  ( $n=2$ ).

## Supplementary Discussion 1

This work describes a simple methodology to stoichiometrically approach the typical overflow metabolite profile expected during aerobic batch glucose fermentation of *S. cerevisiae*, through the introduction of the artificial reaction *FNADH*. However, simulations proposed here consistently failed to align with carbon dioxide yield on substrates, and productivities, whereby the model produced approximately double the amount observed experimentally. While the relatively high carbon recovery observed experimentally in aerobic batch glucose fermentation suggests carbon dioxide measurements were accurate, it should be noted the theoretical ratio of carbon dioxide to ethanol production under purely fermentative glucose metabolism is 1:2, and experimentally under both aerobic, and anaerobic conditions in CEN.PK113-7D and BY4741 the ratio observed is 1:3 (Nielsen, et al, 2003). The original iFF708 ability to predict carbon dioxide production rates was validated experimentally with aerobic glucose-limited continuous cultivation data, and demonstrated excellent fit between dilution rates 0.1 and  $0.38\text{h}^{-1}$ , representing a broad span of respiratory quotients (RQ) (Famili, et al, 2003). This therefore suggests that carbon dioxide metabolism in CEN.PK113-7D and BY4741 under batch glucose fermentation conditions deviates from theoretical expectations, or when considered in the context of a highly interconnected network, not fully described by the stoichiometry of iFF708. It is not expected that the discrepancy in carbon dioxide predictive power would significantly alter the succinate overproduction strategies identified. It is further interesting to note that in the same work, the only data point not predicted by the original iFF708 was the glycerol production rate at the higher dilution rate,  $0.38\text{h}^{-1}$ , most representative of batch conditions (Famili, et al, 2003).

# **PAPER III**

## **Metabolic engineering of *Saccharomyces cerevisiae* for xylose consumption.**

**José Manuel Otero, Gionata Scalcinati, Jennifer R. H. Van Vleet, Thomas W. Jeffries, Lisbeth Olsson, Jens Nielsen**

Submitted to *Microbial Cell Factories*, 2009

**Title:** Metabolic engineering of *Saccharomyces cerevisiae* for xylose consumption

**Authors:** José Manuel Otero<sup>1,2,6,†</sup>, Gionata Scalcinati<sup>1,2,3,†</sup>, Jennifer R. H. Van Vleet<sup>4</sup>, Thomas W. Jeffries<sup>4,5</sup>, Lisbeth Olsson<sup>1,2\*</sup>, Jens Nielsen<sup>1,2</sup>

1. Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-41296 Gothenburg, Sweden.
2. Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark DK-2800, Kgs. Lyngby, Denmark.
3. Dipartimento di Biotecnologie e Bioscienze, Universita' Degli Studi Di Milano Biocca, Piazza della Scienza, 2, 20126 Milano, Italy
4. Department of Bacteriology, University of Wisconsin-Madison, 420 Henry Mall, Madison, Wisconsin, 53706-2398, USA
5. Institute for Microbial and Biochemical Technology, Forest Products Laboratory, One Gillford Pinchot Drive, Madison, Wisconsin 53726-2398, USA.
6. Current Address: Vaccine & Biologics Process Development, Bioprocess Research & Development, Merck Research Labs, West Point, PA, USA.

\*Contributed equally to this research

\*Corresponding author

Professor Lisbeth Olsson

Tel: +46 (0) 31 772 3805

Fax: +46 (0) 31 772 3801

E-mail: lisbeth.olsson@chalmers.se

**Keywords:** xylose, lignocellulose, *Saccharomyces cerevisiae*, transcriptome, metabolic engineering, directed evolution

## Abstract

### Background

Xylose is the second most abundant monosaccharide after glucose, and the most prevalent pentose sugar found in lignocelluloses. Industrial biotechnology aims to develop robust microbial cell factories, such as *Saccharomyces cerevisiae*, to produce an array of added value chemicals presently dominated by petrochemical processes. Significant efforts have focused on the metabolic engineering of *S. cerevisiae* enabling efficient xylose utilization for fuel bioethanol production under anaerobic conditions, and although several examples of success exist, there has yet to be engineered a strain that can consume xylose aerobically without redirection of some carbon flux to overflow metabolites including ethanol, glycerol, acetate, or xylitol. This study aims to metabolically engineer *S. cerevisiae* to exclusively consume xylose while maximizing carbon flux to biomass production. Such a platform may then be enhanced with complimentary metabolic engineering strategies that couple biomass production with high value-added chemicals.

### Results

In this study, *S. cerevisiae* CEN.PK 113-3C, expressing *PsXYL1* (encoding xylose reductase, XR), *PsXYL2* (encoding xylitol dehydrogenase, XDH), and *PsXYL3* (encoding xylulose kinase, XK) from the native xylose-metabolizing yeast *Pichia stipitis*, was constructed, followed by a directed evolution strategy to improve xylose utilization rates. The resulting strains were physiologically characterized under aerobic and anaerobic controlled batch fermentations supplemented with glucose, glucose/xylose mixtures, and xylose. The resulting *S. cerevisiae* strain was capable of consuming xylose at a specific rate of  $0.31 \text{ g g-cell}^{-1} \text{ h}^{-1}$ , a specific growth rate of  $0.18 \text{ h}^{-1}$ , and a biomass yield of  $0.62 \text{ C-mol/C-mol xylose}$ . Transcriptional profiling of this strain was employed to further elucidate pathway metabolism and physiology. The resulting strain produced only biomass and no by-products, with transcriptional profiling confirming a strongly up-regulated glyoxylate pathway enabling respiratory metabolism. Furthermore, plasmid isolation and re-transformation experiments confirmed the conferred phenotype resulted from a chromosomal modification.

### Conclusions

The resulting metabolically engineered strain is a desirable platform for industrial production of biomass related products using xylose as a sole carbon source. To date, no comparable strategy expressing XR/XDH/XK has produced a strain capable of such fast aerobic growth with an absence of significant redirection of carbon flux to xylitol, glycerol, ethanol, or acetate.

## Introduction

Xylose is the most abundant pentose sugar in lignocellulosic feedstocks, including hemicellulose, hardwoods, and crop residues, and is the second most abundant monosaccharide after glucose [1]. The demand for industrial biotechnology processes that leverage sustainable, environmentally favorable, and cost-effective raw materials as alternatives to petrochemical feedstocks is receiving unprecedented research focus [2]. *Saccharomyces cerevisiae* is a proven, robust, industrial production platform used for the expression of a wide range of therapeutic agents, food and beverage components, added value chemicals [3-4] and commodity chemicals (e.g., bioethanol) across large scales (>50,000L) [2]. *S. cerevisiae* offers numerous advantages including extensive physiological and systems biology characterization, Generally Regarded As Safe status from the US Food & Drug Administration, the ability to grow on minimal, chemically defined medium supplemented with alternative feedstock, and adequate growth across a wide pH range (pH 3-6). Wild-type *S. cerevisiae* is unable to efficiently utilize xylose as a primary substrate. The field has largely focused on metabolic engineering of *S. cerevisiae* for maximizing carbon flux from xylose to bioethanol under anaerobic conditions [2]. A microbial cell factory designed for broader biomass-coupled production of added value chemicals from xylose under aerobic conditions would be favored without loss of carbon to over-flow metabolites (ethanol, glycerol, xylitol).

Xylose uptake in *S. cerevisiae* is mediated by the hexose transporters encoded by the *HXT* gene family, but with significantly lower affinities ( $K_M = 137\text{-}300 \text{ mM}$ ) compared to glucose ( $K_M = 1.5\text{-}20 \text{ mM}$ ) [5-7]. Utilization of xylose in yeast and filamentous fungi is then characterized by a two-step oxido-reductive isomerization to xylitol via xylose reductase (XR, primarily NADPH consuming), and then xylitol conversion to xylulose via xylitol dehydrogenase (XDH, NADH producing) [8]. In bacteria, isomerization of xylose to xylulose occurs in a one step reaction catalyzed by xylose isomerase [9-10]. In yeast, fungi, and bacteria, the final conversion of xylulose to xylulose-5P via xylulose kinase (ATP consuming) is conserved.

Recombinant *S. cerevisiae* strains expressing the *Pichia stipitis* xylose reductase (*PsXYL1*) and *P. stipitis* xylitol dehydrogenase (*PsXYL2*) has lead to transformants that can oxidatively and exclusively consume xylose, although resulting in significant xylitol production [11-14]. While over-expression of the endogenous *XKS1* encoding xylulokinase improved the xylose utilization rate [15-16], xylitol formation persisted. The *S. cerevisiae* strain TMB3001 was amongst the first strains to integrate *PsXYL1* and *PsXYL2* and the endogenous *XKS1*, and demonstrate xylose consumption under first aerobic conditions, and then under anaerobic conditions where xylose was co-utilized with glucose [17]. There is a redox imbalance which results from recombinant co-expression of XR and XDH, and due to the lack of transhydrogenase activity in *S. cerevisiae*, and thereby inability to interconvert NADPH and NADH, there is a surplus formation of NADH and NADP<sup>+</sup>. A strategy employed to alleviate the redox imbalance was the introduction of a one-step xylose isomerase (XI) encoded by *XYLA* in *S. cerevisiae*, where xylose is converted to xylulose. Several attempts to express *XYLA* in *S. cerevisiae* from different microorganisms have not been successful, often resulting in low to inactive enzyme expression [18]. Expression of XI encoded by *ARAA* from the anaerobic fungus *Piromyces sp. E2* has resulted in very slow growth on xylose [10, 19]. Adaptation through prolonged chemostat continuous culture improved the xylose utilization of this XI expressing strain [20]. This strain was able to grow aerobically and anaerobically on xylose, and with further manipulation of the potentially limiting steps of the xylose pathway, growth and fermentation performance on xylose was further improved [21]. There have been numerous metabolic engineering efforts employed to alleviate the redox imbalance discussed above, and to further improve the xylose consumption rate. These efforts have been previously reviewed extensively [22-24].

Among the several possible bottlenecks investigated in xylose metabolism several limiting steps have been identified. The reduced ability of *S. cerevisiae* to grow efficiently on xylose has been attributed to the: (i) inefficient xylose uptake [11], (ii) insufficient level of expression of xylose transporters to enable significant sugar assimilation [25], (iii) redox imbalance generated in the first two steps of xylose metabolism involving the XDH and XR from *P. stipitis* [11, 26], (iv) the level of aeration [14, 27-28], (v) insufficient pentose phosphate pathway activity [11, 14], and (vi) the inability of pentose sugar metabolism to activate the lower part of Glycolysis [29-30].

Due to the previously described specificity of XR for NADPH and XDH for NAD<sup>+</sup> and the resulting redox imbalance, xylose metabolism is partially regulated by the availability of oxygen in both native and metabolically engineered yeasts [14, 27-28]. In the presence of oxygen excess NADH produced via NAD-dependent XDH can be respiration, and the NADPH demand for the XR reaction provided by the oxidative part of the pentose phosphate pathway. The level of oxygenation determines the split in carbon flux between biomass and ethanol production under aerobic conditions where xylose is mainly converted into biomass, while ethanol production is favored under anaerobic conditions [28]. The incomplete respiration of excess NADH under anaerobic conditions leads *S. cerevisiae* to produce and accumulate glycerol followed by xylitol. The xylose consumption rate and assimilation to biomass increases with increasing aeration level, relieving the accumulation of NADH yet still resulting in glycerol and xylitol formation [31-32].

This study aims to metabolically engineer *S. cerevisiae* such that it can consume xylose as the exclusive substrate while maximizing carbon flux to biomass production. Such a platform may then be enhanced with complimentary metabolic engineering strategies that couple biomass production with high value-added chemicals. In this study, *S. cerevisiae* CEN.PK 113-3C, expressing *PsXYL1* (encoding xylose reductase, XR), *PsXYL2* (encoding xylitol dehydrogenase, XDH), and *PsXYL3* (encoding xylulose kinase, XK) from the native xylose-metabolizing yeast *Pichia stipitis*, was constructed, followed by a directed evolution strategy to improve xylose utilization rates. The resulting strains were physiologically characterized under aerobic and anaerobic controlled batch fermentations supplemented with glucose, glucose/xylose mixtures, and xylose. Transcriptional profiling was employed to further elucidate pathway metabolism and physiology.

## Results

### Physiological Characterization of CMB.GS001

Batch cultivations of the xylose fermenting *S. cerevisiae* strains TMB3001, CPB.CR4, CPB.CR5, and CMB.GS001 were investigated in shake flask cultures in synthetic medium supplemented with 20 g l<sup>-1</sup> glucose or 20 g l<sup>-1</sup> xylose. The maximum specific growth rate ( $\mu_{\max}$ ) on each carbon source was determined (Supplementary Materials Table 1). The specific growth rate on glucose (0.29–0.30 h<sup>-1</sup>) for the recombinant strains was 16 % lower than the reference strain CEN.PK 113-7D (0.36 ± 0.01 h<sup>-1</sup>). During growth on xylose as the sole carbon source all the strains exhibit two distinct growth phases. An initial, rapid, growth phase interval (GPI) between 0 and 27 hours, and a second, significantly slower, growth phase interval (GPII) for the duration of the culture.

In contrast to the reference strain CEN.PK 113-7D, which cannot grow on xylose, the recombinant strains grew aerobically on xylose with a specific growth rate ranging from 0.16 h<sup>-1</sup> for TMB3001 to 0.07 h<sup>-1</sup> for CPB.CR5 during GPI (Supplementary Materials Table 1). During GPII a dramatic reduction in specific growth rate took place for all strains with values ranging from 0.009 h<sup>-1</sup> for CMB.GS001 to 0.005 h<sup>-1</sup> for TMB3001 and CPB.CR5. The xylose consumption rate was similar for all the strains, ranging from 0.06 g (g biomass)<sup>-1</sup> h<sup>-1</sup> for TMB3001 to 0.08 g (g biomass)<sup>-1</sup> h<sup>-1</sup> for CMB.GS001 (Supplementary Materials Table 1). In all cases less than 2 g l<sup>-1</sup> xylose was consumed.

The strain CMB.GS001 was grown under aerobic batch fermentation conditions with 20 g l<sup>-1</sup> xylose as the sole carbon source exhibiting slow growth on xylose with a maximum specific growth rate of 0.02 h<sup>-1</sup>. The maximum specific xylose consumption rate was 0.05 g xylose (g dry cell weight)<sup>-1</sup> h<sup>-1</sup>. The specific growth rate in the bioreactor was higher than in the shake flask cultures, due to a less efficient oxygen transfer in the shake flasks resulting in semi-aerobic conditions. With an inoculation OD<sub>600</sub> of 0.01 corresponding approximately to 0.002 g cell l<sup>-1</sup>, 1.4 g l<sup>-1</sup> xylose was consumed in 110 hours with biomass and carbon dioxide as the major products (Figure 1A). No other products were detected.

In order to investigate the effect of the inoculation cell density, the strain CMB.GS001 was cultivated with an initial OD<sub>600</sub> of 1 (corresponds to 0.2 g cell l<sup>-1</sup>) instead of 0.1. With this higher inoculation, growth was characterized by two distinct phases. The first growth phase (0–100 h) resulted in a maximum specific growth rate of 0.01 h<sup>-1</sup> and a specific xylose consumption rate of 0.03 g xylose (g dry cell weight)<sup>-1</sup> h<sup>-1</sup>. During this phase 1.7 g l<sup>-1</sup> xylose of the initial 20 g l<sup>-1</sup> were consumed with biomass and carbon dioxide as the major products (Figure 1B). After 100 h, a second phase characterized by significantly increased xylose consumption was observed, with a maximum specific growth rate of 0.12 h<sup>-1</sup> and a specific xylose consumption rate of 0.24 g xylose (g dry cell weight)<sup>-1</sup> h<sup>-1</sup>. During this phase the remaining xylose was completely converted to primarily biomass and carbon dioxide (Figure 1B), with ethanol and acetate formed in small amounts (0.26 combined C-mol/C-mol xylose, Table 2). Negligible amounts of glycerol and xylitol were detected (0.0053 combined C-mol/C-mol xylose, Table 2). Complete xylose utilization took more than 150 hours of total fermentation time (Figure 1B) with 94.9 C-mol/C-mol xylose carbon recovered (Table 2).

### Directed Evolution of CMB.GS001

Directed evolution was applied to select a spontaneous mutant with higher specific growth rate on xylose. The xylose fermenting strains TMB3001, CPB.CR5, and CMB.GS001 were subjected to repetitive serial transfers in batch shake flask cultivations with minimal medium supplemented with 20 g l<sup>-1</sup> xylose. This approach targeted strain selection based on biomass formation rate, directly coupled to the xylose consumption rate. After four batch cultures, only strain CMB.GS001 demonstrated an appreciable improvement in xylose consumption. For all the other strains evaluated the residual xylose concentration measured in the culture was more than 18 g l<sup>-1</sup> (Figure 2A). The initial total xylose consumption and biomass production of CMB.GS001 was 1.3 g l<sup>-1</sup> and 0.18 g dry cell weight, respectively. After serial cultivations over 10 cycles the xylose consumption for strain CMB.GS010 increased 15-fold to 20 g l<sup>-1</sup> and the biomass production increased 52-fold to 9.37 (g dry cell weight) l<sup>-1</sup> (Figure 2A). The initial specific growth

rate of *S. cerevisiae* CMB.GS001 on xylose was  $0.02\text{ h}^{-1}$ . After these 10 transfers, covering a period of 500 h (21 days), the specific growth rate increased 9-fold to  $0.18\text{ h}^{-1}$ . A total of 74 cell generations were produced across the ten cycles of directed evolution with the final 50–74 generations not yielding any improvement in specific growth rate (Figure 2B).

In order to investigate the possible causes of the dramatic increase in the specific growth rate of CMB.GS010 the plasmid pRS314-X123 was removed by prolonged cultivation of CMB.GS010 on YPD medium followed by verification of plasmid loss by re-plating on minimal medium lacking tryptophan. The resulting auxotrophic strain, named CMB.GS011, was transformed with pRS314-X123 (original un-evolved plasmid used to transform CMB.GS001) to obtain the strain CMB.GS012. Physiological characterization of CMB.GS012 and all subsequent strains was completed in semi-aerobic shake flasks with synthetic medium supplemented with  $20\text{ g l}^{-1}$  xylose. The maximum specific growth rate on xylose for CMB.GS012 was comparable to the evolved parental strain CMB.GS010. Furthermore, the plasmid extracted from CMB.GS010 was retransformed into CMB.GS011 to obtain strain CMB.GS013. CMB.GS013 exhibited the same specific growth rate as CMB.GS010 and CMB.GS012. The strain CMB.GS014 was created by transforming CEN.PK113-3C with the recovered plasmid from CMB.GS010 and exhibited specific growth rates similar to CMB.GS001. In short, CMB.GS012 and CMB.GS014 confirm that the improved xylose consumption rate similar to the original CMB.GS010 is a consequence of mutations in the genome and not in the plasmid carrying the properties needed for xylose metabolism, i.e. the recovered evolved plasmid or native unevolved plasmid did not confer any difference in xylose utilization.

## Physiological Characterization of CMB.GS010

### Batch Aerobic and Anaerobic Xylose Fermentation

Strain CMB.GS010 was physiologically characterized in stirred tank aerobic and anaerobic batch fermentations supplemented with  $20\text{ g l}^{-1}$  xylose or  $20\text{ g l}^{-1}$  glucose. The maximum specific xylose consumption rate was  $0.31\text{ g xylose (g dry cell weight)}^{-1}\text{ h}^{-1}$ , amongst the highest reported in the literature for aerobic growth on xylose of a *S. cerevisiae* strain carrying the genes encoding for XR, XDH and XK (Supplementary Materials Table 2). Inoculated at an initial  $\text{OD}_{600}$  of 0.01 ( $0.002\text{ g cell l}^{-1}$ ), all the xylose was consumed within 60 h with biomass (62% C-mol/C-mol xylose) and carbon dioxide (37% C-mol/C-mol xylose) as the major fermentation products, noting the complete absence of xylitol during the culture (Table 3). Compared to CMB.GS001, there were significant increases of 12 and 27% C-mol/C-mol xylose in biomass and carbon dioxide yields, respectively.

The xylose consumption rate was highest ( $0.31\text{ g xylose (g dry cell weight)}^{-1}\text{ h}^{-1}$ ) when the extracellular xylose concentration was above  $10\text{ g l}^{-1}$ , as demonstrated by the biomass concentration and peak carbon evolution rate (Figure 3A) subsequently decreasing to  $0.08\text{ g xylose (g dry cell weight)}^{-1}\text{ h}^{-1}$  until xylose exhaustion. To further investigate if xylose consumption is sensitive to changes in extracellular xylose concentration, CMB.GS010 was cultivated in semi-aerobic shake flasks with synthetic media supplemented with  $10\text{ g l}^{-1}$  xylose. Under this condition the strain exhibits a maximum specific growth rate of  $0.11\text{ h}^{-1}$  compared with  $0.18\text{ h}^{-1}$  when supplemented with  $20\text{ g l}^{-1}$  xylose. The reduced extracellular concentration of xylose to  $10\text{ g l}^{-1}$  resulted in an increased lag-phase (12 to 24 hours), and maximum specific xylose consumption rate of  $0.26\text{ g xylose (g dry cell weight)}^{-1}\text{ h}^{-1}$ .

CMB.GS010 was cultivated under anaerobic batch fermentation conditions with  $20\text{ g l}^{-1}$  xylose as the sole carbon source. After 100 h no growth or xylose consumption was observed (Figure 3B). To ensure that the absence of growth was a direct consequence of the anaerobic environment, a recovery experiment was performed, where the culture was aerated quickly from an anaerobic to aerobic condition. Growth was immediately restored to the above-described aerobic physiology.

### Batch Aerobic and Anaerobic Glucose Fermentation

The aerobic and anaerobic physiology of CMB.GS010 was evaluated in glucose supplemented batch fermentations to quantify the possible effects of directed evolution on the maximum specific growth rate

and the product yields compared to the reference strain CEN.PK113-7D. Comparable to previous results with CEN.PK 113-7D, during aerobic batch fermentation of CMB.GS010 on 20 g l<sup>-1</sup> glucose there was complete glucose conversion to primarily ethanol and carbon dioxide, with biomass, glycerol and acetate formed in smaller amounts. When glucose was depleted a diauxic shift takes place, after which ethanol, organic acids, and glycerol are consumed (Figure 3C). The main differences during aerobic growth on glucose between CMB.GS010 and the reference strain CEN.PK113-7D, were a reduction of the specific maximum growth rate from 0.36 h<sup>-1</sup> to 0.34 h<sup>-1</sup>, a 4-fold higher acetate production, and a small reduction in ethanol production (Table 2). Anaerobic cultivation of CMB.GS.010 and the reference strain CEN.PK113-7D on glucose supplemented medium also showed similar results. During anaerobic batch fermentation of CMB.GS010 on 20 g l<sup>-1</sup> glucose there was complete conversion to primarily ethanol and carbon dioxide, with biomass and glycerol formed in smaller amounts (Figure 3D and Table 2). The formation of ethanol and carbon dioxide stopped immediately after depletion of glucose in the medium. Reduction in the ethanol concentration after glucose was exhausted is attributed to evaporation (Figure 3D). The main difference during anaerobic growth on glucose was a reduction in the maximum specific growth rate from 0.34 to 0.29 h<sup>-1</sup>, and a small reduction in the ethanol yield (Table 2).

#### **Batch Aerobic and Anaerobic Mixed Substrate Fermentation**

In order to investigate the properties of the strain CMB.GS010 with respect to mixed sugar utilization the strain was grown aerobically in a mixture containing 10 g l<sup>-1</sup> glucose and 10 g l<sup>-1</sup> xylose. The results show that both sugars were completely consumed; however, with glucose remaining the preferred substrate. Three different growth phases can be identified (Figure 3E). During the first growth phase (0-20 h) cells consumed 10 g l<sup>-1</sup> glucose and 1.6 g l<sup>-1</sup> xylose in the same period (16% more carbon resulting from xylose consumption). The maximum specific growth rate was slightly lower compared with the  $\mu_{\max}$  for growth on glucose only (Table 2); however, the total biomass yield for the mixed sugar fermentation was higher than the comparable yield calculated from fermentation with glucose (Supplementary Materials Table 3)

Following glucose exhaustion there was a second growth phase (20-30 h) where the remaining xylose, 8 g l<sup>-1</sup> (0.27 C-moles l<sup>-1</sup>), was consumed in conjunction with the re-assimilation of ethanol produced during the glucose consumption phase. During this phase 0.17 C-moles l<sup>-1</sup> xylose and 0.15 C-moles l<sup>-1</sup> ethanol were consumed. In this phase the maximum specific growth rate decreased 2.5-fold from 0.32 to 0.13 h<sup>-1</sup>. The maximum xylose consumption rate during the first growth phase on glucose was 0.25 g (g biomass)<sup>-1</sup> h<sup>-1</sup>. Once glucose was depleted, the maximum xylose consumption rate was 0.18 g (g biomass)<sup>-1</sup> h<sup>-1</sup>. After ethanol re-assimilation, the xylose consumption continued until all the sugar was consumed in the third and final growth phase (>30 h) with a reduced maximum consumption rate of 0.06 g (g biomass)<sup>-1</sup> h<sup>-1</sup>. In contrast to the glucose consumption phase, the xylose–ethanol phase was characterized by a large production of biomass, corresponding to a 28%-increase in biomass yield (C-moles C-moles<sup>-1</sup>).

The fermentation characteristics of strain CMB.GS010 were also investigated under anaerobic growth on a medium containing 10 g l<sup>-1</sup> glucose and 10 g l<sup>-1</sup> xylose. The results show that only glucose was fully consumed (Figure 3F). During the first rapid exponential phase, glucose with a small fraction of xylose (less than 2 g l<sup>-1</sup>) was consumed, with ethanol and carbon dioxide as the major by-products. Biomass, glycerol, and acetate were formed in smaller amounts (Table 2). After the glucose consumption phase, a period of maintenance without detectable growth was observed where a small amount of xylose (1 g l<sup>-1</sup>) was consumed with concurrent production of glycerol (0.8 g l<sup>-1</sup>) and xylitol (0.6 g l<sup>-1</sup>). No differences were observed between the specific growth rate calculated during the period of simultaneous glucose and xylose consumption compared with the  $\mu_{\max}$  for growth on glucose only, while minor differences were observed in the product yields (Supplementary Materials Table 3).

A recovery experiment, similar to that described earlier was performed. Aeration of the fermenter after 100 h of anaerobic fermentation immediately resulted in the complete consumption of xylose and ethanol (Figure 3F).

### **Continuous Aerobic Xylose Fermentation**

A continuous fermentation of strain CMB.GS.010 was attempted for physiological characterization at a dilution rate ( $D$ ) of  $0.1\text{ h}^{-1}$ . The chemostat cultivation was started with an initial batch phase on minimal medium supplemented with  $10\text{ g l}^{-1}$  xylose. When the biomass reached a concentration of  $2\text{ g l}^{-1}$ , the chemostat was initiated by feeding  $10\text{ g l}^{-1}$  xylose at a constant rate of  $0.1\text{ l h}^{-1}$ . Under this condition the strain was unable to reach steady state, and the cell concentration decreased steadily. After 29 hours, only 72% of the starting cell concentration remained in the fermenter, and the concentration of xylose in the feed was increased to  $37\text{ g l}^{-1}$ , hypothesizing that concentration-dependent affinity for xylose transport was limiting. After 67 h a steady-state was reached, where 74% of the xylose in the feed ( $27\text{ g l}^{-1}$ ) was consumed. The specific xylose consumption rate was  $0.25\text{ g (g DCW)}^{-1}\text{h}^{-1}$ . Product yields expressed in C-mol C-mol $^{-1}$  xylose were:  $0.64 Y_{SX}$  (biomass),  $0.001 Y_{SE}$  (ethanol),  $0.001 Y_{SA}$  (acetate),  $0.002 Y_{SS}$  (succinic acid) and  $0.085 Y_{SC}$  (carbon dioxide). A carbon balance analysis suggests that the measured products accounted for only 80% of the consumed carbon. The steady-state was maintained for 70 h (10 residence times), after which the dilution rate was increased to  $0.25\text{ h}^{-1}$ . This increase is clearly above the maximum specific growth rate observed during the batch cultivations with  $20\text{ g l}^{-1}$  xylose, in an attempt to select a strain with an improved  $\mu_{max}$ . At this dilution rate growth was not supported and cell wash-out occurred.

### **Transcriptome Characterization**

Transcriptome characterization was performed on a total of five different cultivation conditions and with a biological replicate for each condition. The five conditions were: evolved strain (CMB.GS010) cultivated in batches with xylose and glucose as carbon sources, and continuous cultures with glucose as the sole carbon source; and the unevolved strain (CMB.GS001) with glucose as the sole carbon source in both batch and continuous cultivations. Table 4 gives an overview of the total number of differentially expressed genes between the different cultivation conditions selected to elucidate overall carbon flux distributions observed in CMB.GS010 compared to CMB.GS001. The specific comparisons made were focused on identifying fermentative vs. respiro-fermentative metabolism for growth on the different carbon sources and cultivation conditions. Supplementary Materials Table 4 provides a summary of the transcriptome study. Supplementary Materials Figure 2 is a principal component analysis (PCA) showing the clustering of the expression data after normalization for each condition. The evolved strain grown on xylose in both batch and chemostat conditions clustered closely together, with clear separation from the evolved strain grown on glucose under batch conditions. The evolved strain grown on batch glucose exhibited fermentative metabolism compared to the evolved xylose chemostat and batch cultures exhibiting respiratory metabolism. It is noteworthy that the unevolved strain grown under glucose chemostat conditions clustered separately from the evolved strain xylose fermentations, suggesting that while both conditions elicited a respiratory metabolic response the differentially expressed gene profiles are distinctive.

Supplementary Materials Figures 3 and 4 present the GO process terms identified from the significant differential gene expressions between the different conditions. Figure 4 is a schematic representation of the log-fold change gene expression of genes encoding enzymes of the central carbon metabolism often correlating with respiration, and it includes the tricarboxylic acid cycle (TCA cycle), glyoxylate pathway, and glutamine/glutamate metabolism. Figure 5 is a schematic representation of the log-fold change gene expression of the pentose phosphate pathway (PP pathway).

The significant mRNA up-regulation of TCA cycle and glyoxylate pathways of the evolved strain on xylose compared to the unevolved or evolved strain on glucose under batch cultivations correlates well with the physiological observations that growth on xylose is dominated by respiratory metabolism. The glyoxylate pathway (*ICL1*, *MLS1*, *MDH1*, *MDH2*, *AGX1*) was significantly up-regulated in the evolved strain grown on xylose compared to the evolved strain grown on glucose or the unevolved strain grown on glucose. This pathway had a significantly higher log-fold change than succinate dehydrogenase and succinyl-CoA ligase (*SDH1*, *SDH2*, *SDH3*, *SDH4*, and *LSC2*, respectively), suggesting that this pathway plays an important role during respiratory metabolism of *S. cerevisiae*.

The evolved strain cultivated in a xylose chemostat compared to the unevolved strain cultivated in a glucose chemostat exhibited up-regulated *ICL1* and *MDH2* expression; however, down-regulation of *DAL7* (log-fold change -3.6), which encodes malate synthase and suggests incomplete glyoxylate by-pass of the TCA cycle. These differences are consistent with the aforementioned PCA clustering comparison of the chemostat conditions on glucose and xylose. It is also worth noting that at these conditions there were no differential gene expression of *AGX1* or *SER2*, and *GDH3* was down-regulated. This mRNA expression pattern is consistent with the observation that the evolved strain cultivated in a xylose chemostat resulted in an incomplete carbon recovery, and when the dilution rate was increased wash-out quickly occurred. Finally, *IDP2* and *IDP3* were up-regulated significantly in all batch xylose cultivations with the evolved strain (Figure 4).

The evolved strain, when cultivated on xylose in a batch mode, is able to utilize the glyoxylate by-pass to efficiently respire the carbon source. Furthermore, the expression levels of *MDH2*, *PCK1*, and *FBP1* were up-regulated in the evolved strain cultivated on xylose compared to the evolved or unevolved strain cultivated on glucose, indicating some glyconeogenic activity (Supplementary Materials Figure 5). It should be mentioned though that these genes are also up-regulated at low dilution rates in glucose-limited chemostat cultures [34], and expression of these genes may therefore be associated with respiratory metabolism.

The mRNA expression profile of the evolved strain cultivated on xylose suggests a strong flux towards glucose-6-phosphate, requiring inspection of the pentose phosphate pathway (PP pathway). Figure 6 presents the two comparisons for which any differential gene expression in the PP pathway was detected. Independent of whether the evolved strain was cultivated on batch xylose or glucose, the *SOL* genes were significantly up-regulated (*SOL3* and *SOL4* encode 6-phosphogluconolactonase) compared to growth on glucose with the unevolved strain. However, the evolved strain cultivated on batch xylose, when compared to the unevolved on batch glucose, exhibited significant up-regulation of transketolase (encoded by *TKL2*, log-fold change 3.96). Transketolase (encoded by major isoform *TKL1* and minor isoform *TKL2*) in combination with transaldolase (encoded by *TAL1*) enables a reversible link between the non-oxidative PP pathway and glycolysis, allowing the cells to adapt their NADPH production and ribose-5-phosphate production to biomass demands [35]. The over-expression of *TAL1* and *TKL1* in *S. cerevisiae* over-expressing *PsXYL1* and *PsXYL2* has been previously demonstrated, and there was no influence on growth under either aerobic or anaerobic fermentation conditions in the *TKL1* over-expressed mutant [36]. *TKL2* over-expression was not considered and the authors concluded that transaldolase expression in *S. cerevisiae* is insufficient for effective utilization of PP pathway metabolites [36]. Furthermore, *TKL2* up-regulation has been correlated with carbon-limited chemostat culture [33], but this was not observed here.

## Discussion

The strain constructed in this work (CMB.GS010) was obtained through a combination of genetic modification (plasmid introduction), and the application of selective pressure (shake flask repetitive cultivation). Due to the native inability of *S. cerevisiae* to metabolize xylose, three essential genes for xylose uptake from *P. stipitis* (*PsXYL1*, *PsXYL2* and *PsXYL3*) were introduced. Using xylose uptake for selection a strain capable of fast aerobic xylose metabolism was obtained in a relatively short period of time (500 hours, corresponding to 74 cell generations). The 10-fold increase in the specific growth rate on xylose under aerobic growth conditions in only 21 days through repetitive shake flask cultures is evidence of the efficiency and simplicity of the method. Only the strain CMB.GS010 carrying the functional metabolic pathway reconstructed using the heterologous genes (*XYL1*, *XYL2* and *XYL3*) originating from *P. stipitis* exhibited enhanced xylose consumption. In fact, the other *S. cerevisiae* strains investigated (TMB.3001, CPB.CR4 and CPB.CR5) when modified to express the two heterologous enzymes, XR and XDH, in conjunction with the endogenous over-expresses XK activity, and subjected to the same selective pressure did not exhibit any appreciable improvement (Supplementary Materials Table 1) in xylose utilization. Furthermore, plasmid recovery and retransformation experiments confirm that the genetic modifications during adaptive evolution are present chromosomally in the host rather than any modifications to the plasmid.

Co-utilization of both sugars, glucose and xylose, is essential for an economically feasible conversion of lignocellulose to industrially relevant bio-products. Xylose is predominantly consumed after glucose exhaustion. This could be explained with a competitive inhibition model. Until now no transporters have been found in *S. cerevisiae* that can exclusively and specifically transport xylose. Nevertheless it is known that xylose competes with glucose for the same transporters [11, 50]. Interestingly, mRNA GO process term comparison of CMB.GS010 and CMB.GS001 cultivated on batch glucose indicated all significant terms being classified as transporters (Supplementary Materials Figure 5B). This suggests that genetic permutations in the strain CMB.GS010 resulting from directed evolution on xylose, and expressed independent of carbon source, are most likely related to transport.

The *S. cerevisiae* strain CMB.GS010 exhibited a high specific growth rate on xylose under aerobic conditions that exceeds published data on other *S. cerevisiae* strains metabolically engineered for xylose assimilation with XR, XDH and XK genes (Supplementary Materials Table 2). Strain CMB.GS010 clearly exhibited a respiratory metabolism on this sugar, similar to that observed for native *S. cerevisiae* cultivated in glucose-limited chemostats. Xylose utilization is almost entirely oxidative as indicated by the RQ coefficient ( $RQ < 1$ ), the high carbon fraction of xylose converted to biomass as compared to glucose metabolism, and the very low ethanol production. Furthermore, the physiological observations were supported by transcriptome data. The up-regulation of the glyoxylate pathway in the evolved strain grown on xylose compared to growth on glucose, or the un-evolved strain grown on glucose is in line with observations made at low dilution rates in glucose-limited chemostat cultures in wild-type *S. cerevisiae* [34].

As an extension of the glyoxylate pathway, *IDP2* and *IDP3* were up-regulated significantly in all evolved strain batch xylose cultivations (Figure 4). Xylose metabolism requires the pentose phosphate pathway (PPP). This pathway involves the conversion of glucose-6-phosphate to 6-phosphogluconate, catalyzed by glucose-6-phosphate dehydrogenase (*ZWF1*), and further conversion to ribulose-5-phosphate with co-current production of  $CO_2$ , catalyzed by 6-phosphogluconate dehydrogenase (*GND1*, *GND2*). The PPP is essential for generation of biomass precursors, which include D-ribose for nucleic acid biosynthesis, D-erythrose-4-phosphate for synthesis of aromatic amino acids, and NADPH for anabolic reactions [37]. While the non-oxidative PPP satisfies D-ribose and D-erythrose-4-phosphate biomass precursor demands, cytosolic NADPH must still be generated, and the oxidative part of the pathway is by-passed during growth on xylose. Cytosolic isocitrate dehydrogenase (*Idp2*) catalyzes the oxidation of isocitrate to  $\alpha$ -ketoglutarate, and is  $NADP^+$  specific [38]. On both fermentable and non-fermentable carbon sources *Zwf1p* is constitutively expressed while *Idp2p* levels are glucose-repressed [39-40]. *Idp2p* levels have been demonstrated to be both elevated on non-fermentable carbon sources, and during the diauxic shift as

glucose is depleted [39, 41-42]. Furthermore, in  $\Delta zwf1 \Delta adh6$  *S. cerevisiae* mutants, it was demonstrated that Idp2 is up-regulated and generates enough NADPH to satisfy biomass requirements, noting that the NADP<sup>+</sup> specific cytosolic aldehyde dehydrogenase (Adh6p) catalyzing acetaldehyde conversion to acetate is the other major cytosolic source of NADPH [43]. In the evolved strain *IDP2* and *IDP3* likely provide a source of NADPH to satisfy biomass requirements.

The native xylose-fermenting strain *P. stipitis*, which is the source of the heterologous expressed enzymes, XR and XDH, does not produce xylitol during xylose fermentations [27]. Extensive xylitol formation has been observed in all the *S. cerevisiae* xylose consuming strains expressing these enzymes [11, 13-17]. The production of xylitol has been shown to be the direct result of a redox imbalance of the NAD(P) cofactors between the XR and XDH [26, 44-49]. This imbalance has recently been successfully avoided by direct conversion of xylose to xylulose via the introduction of a bacterial isomerase [19-20]. Xylitol formation is often described as being the major drawback of the XR-XDH strategy; however, in the engineered strain selected in this study the formation of xylitol was completely absent during all the xylose fermentations.

The absence of xylitol accumulation under oxidative conditions may be interpreted as a result of complete xylitol oxidation. Consistent with this assumption is that oxidation of xylitol to xylulose by XDH is limited by the availability of NAD<sup>+</sup>. Perhaps, although not previously observed or described, the *P. stipitis* derived XR can effectively use NADH as a cofactor instead of NADPH, thereby eliminating the NADP<sup>+</sup>/NAD<sup>+</sup> imbalance, or as the data in this study suggests, up-regulation of *IDP2* ensures sufficient NADPH production to drive xylitol catabolism.

## Conclusion

In summary, a *S. cerevisiae* strain capable of consuming xylose at a rate of  $0.31 \text{ g g-cell}^{-1} \text{ h}^{-1}$ , a specific growth rate of  $0.18\text{h}^{-1}$ , and a biomass yield of  $0.62 \text{ C-mol biomass per C-mol xylose}$  was obtained through a combination of metabolic and directed evolution. Metabolic engineering encompassed targeted genetic engineering with expression of *PsXYL1*, *PsXYL2* and *PsXYL3* with adaptive evolutions and lead to an enhanced phenotype in the host (chromosomal) microbial cell factory. With no unwanted by-products, including xylitol, glycerol, ethanol, or acetate produced, and a strongly up-regulated glyoxylate pathway, this strain is a desirable platform for industrial production of biomass related products using xylose as a sole carbon source.

## Materials and methods

### *Saccharomyces cerevisiae* Strain Descriptions

All of the strains constructed in this study were derived from the reference *Saccharomyces cerevisiae* strain, CEN.PK 113-7D. The strains used in this study, including those constructed *de novo* are listed in Table 1. Included in the table are those strains that were modified using directed evolution, and are referred to as *evolved*. Table 1 also provides an overview of the plasmids utilized during the strain construction. Supplementary Materials Text 1 provides a more in-depth description of the strains listed in Table 1.

Strain CMB.GS001 was derived from the *S. cerevisiae* CEN.PK 113-3C wild type strain. This strain was transformed with the centromeric plasmid pRS314-X123, expressing *TRP1* encoding for N-(5' phosphoribosyl)-anthranilate isomerase. Into this plasmid *PsXYL1* encoding xylose reductase (*PsXRP*), *PsXYL2* encoding xylitol dehydrogenase (*PsXDHP*), and *PsXYL3* encoding xylulokinase (*PsXKp*) all derived from *P. stipitis* were cloned under the glyceraldehyde-3-phosphate dehydrogenase (*TDH3*) constitutive promoter and terminator [65].

Strains CMB.GS002-010 were evolved from CMB.GS.001 after cycles of repetitive culture selection in shake flasks. The three final digits of the strain identifier indicate from which cycle in the repetitive culture the strain originated, with the starting strain referred to as CMB.GS001.

A schematic flow sheet of the origin of each strain used in this study is specified in Supplementary Materials Figure 1. Stock cultures were grown at 30°C in 500 ml (working volume 100 ml) shake flasks on synthetic medium supplemented with 20 g l<sup>-1</sup> glucose, or 20 g l<sup>-1</sup> xylose for the xylose evolved strains CMB.GS001-010. When the late exponential phase was reached as determined by biomass optical density measurements at 600 nm (OD<sub>600</sub>), 30% (vol/vol) sterile glycerol was added, and 1.5 ml sterile cryovials were prepared and stored at -80°C.

### Yeast Strain Transformation

*Saccharomyces cerevisiae* strain CEN.PK113-3C was transformed with plasmid pRS314-X123 [65]. Cells were made competent for plasmid uptake using a traditional lithium acetate treatment [51]. A total of 1x10<sup>7</sup> cells/ml were transformed with 5µl of purified plasmid solution containing 100 ng of plasmid DNA. Transformants were selected using synthetic dextrose agar plates without tryptophan (ScD-trp).

### Directed Evolution and Selection of Strain CMB.GS010

Mutants of CMB.GS001 with higher specific growth rates on xylose were selected for by serial transfer of cells using repetitive cultures in shake flasks. Specifically, a 500 ml shake flask containing 100 ml of synthetic minimal medium with 20 g l<sup>-1</sup> xylose was inoculated with CMB.GS001. After 60 h, a new shake flask culture having the same medium composition was inoculated with cells from the preceding shake flask at an initial OD<sub>600</sub> of 0.025. This procedure was repeated for four iterations. Thereafter, the culture time was reduced to 48h. This 48 h cultivation was repeated for 6 iterations, after which strain CMB.GS010 was isolated. Cryovials were prepared following every cycle of repetitive culture as described in section 2.1.

### Medium Preparation

A synthetic minimal medium containing trace elements and vitamins was used for all shake flasks and stirred tank cultivations [50]. Fatty acids in the form of Tween 80 and Ergosterol were supplemented to anaerobic cultivations [51-52]. Tryptophan was supplemented for cultivations of CEN.PK113-3C to satisfy the auxotrophy. The compositions of the trace element, vitamin, and fatty acid solutions are included in Supplementary Materials List 1.

The medium used for stirred tank batch cultivations had the following composition: 5 g l<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 3 g l<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 0.5 g l<sup>-1</sup> MgSO<sub>4</sub>·7H<sub>2</sub>O, 1 ml l<sup>-1</sup> trace element solution, 1 ml l<sup>-1</sup> vitamin solution, 0.5 ml l<sup>-1</sup> antifoam 204 (Sigma A-8311), and 1.25 ml l<sup>-1</sup> Ergosterol/Tween 80 solution (final concentration 0.01 g

$\text{l}^{-1}$  Ergosterol and 0.42 g  $\text{l}^{-1}$  Tween 80). The fermentation medium was pH adjusted to 5.0 with 2 M NaOH and autoclaved. For the cultivations on glucose the concentration was 20 g  $\text{l}^{-1}$ , and for the cultivations on xylose the concentration was 20 g  $\text{l}^{-1}$ . For mixed sugar cultivation, 10 g  $\text{l}^{-1}$  glucose and 10 g  $\text{l}^{-1}$  xylose were used, yielding a final total sugar concentration of 20 g  $\text{l}^{-1}$ . Both the sugar solutions were added by sterile filtration using a cellulose acetate filter (0.20  $\mu\text{m}$  pore size Minisart®-Plus Satorius AG).

The medium used for shake flask cultivations had the same composition as described above, but the  $(\text{NH}_4)_2\text{SO}_4$  concentration was increased to 7.5 g  $\text{l}^{-1}$  and the  $\text{KH}_2\text{PO}_4$  to 14.4 g  $\text{l}^{-1}$  together with 20 g  $\text{l}^{-1}$  of glucose or xylose, and the pH was adjusted to 6.5 prior to autoclaving.

A yeast extract peptone dextrose (YPD) complex medium was used for yeast growth prior to transformation. The YPD medium had the following composition: 10 g  $\text{l}^{-1}$  yeast extract, 20 g  $\text{l}^{-1}$  peptone, 20 g  $\text{l}^{-1}$  glucose and 20 g  $\text{l}^{-1}$  agar is supplemented for Petri plate preparation.

A synthetic dextrose minus tryptophan medium (ScD-trp) was used as selective media post-transformation. The ScD-trp medium had the following composition: 7.25 g  $\text{l}^{-1}$  Dropout powder (J.T. Baker), 20 g  $\text{l}^{-1}$  agar and 20 g  $\text{l}^{-1}$  glucose.

### Inoculum Preparation

For shake flask cultivations a single colony isolate was selected from an YPD agar plate and suspended in 200  $\mu\text{l}$  of minimal medium. 100  $\mu\text{l}$  of this suspension was inoculated into a shake flask (glucose or xylose supplemented as specified), resulting in an  $\text{OD}_{600}$  of approximately 0.01. For stirred tank fermentations pre-cultures were prepared by inoculating 100 ml of medium (500 ml total volume Erlenmeyer shake flask) containing 20 g  $\text{l}^{-1}$  xylose or glucose with a 1.5 ml cryovial. The pre-culture was incubated at 30°C and grown to the late-exponential phase as determined by duplicate biomass  $\text{OD}_{600}$  measurements.

### Shake flask Cultivation

Cultivations were carried out in 500 ml baffled Erlenmeyer flasks with two diametrically opposite baffles and side necks for aseptic sampling by syringe. The flasks were prepared with 100 ml of medium as previously described and cultivated in a rotary shaker at 150 rpm (stroke length = 3 cm) with the temperature controlled at 30°C. The pH of the medium was adjusted to 6.5 with 2 M NaOH prior to sterilization.

### Stirred Tank Batch Fermentations

Stirred tank cultivations were performed in 2.2 liter Braun Biotech Biostat B fermentation systems with a working volume of 2 liters. The cultivations were operated at aerobic and anaerobic conditions with glucose and/or xylose as the carbon source. The fermentors were integrated with the Braun Biotech Multi-Fermenter Control System (MFCS) for data acquisition. The temperature was controlled at 30°C. The bioreactors were equipped with two disk-turbine impellers rotating at 600 rpm and a ring sparger. Dissolved oxygen was monitored using an autoclavable polarographic oxygen electrode. During aerobic fermentations the sparging flow rate of air was 2 vvm (volume per volume minute). During anaerobic cultivations nitrogen containing less than 5 ppm  $\text{O}_2$  was used for sparging at a flow rate of 2 vvm, with less than 1% air saturated oxygen in the fermenter as confirmed by the dissolved oxygen measurement and the off-gas analyzer. The pH was controlled constant at 5.0 by automatic addition of 2 M KOH. Off gas passed through a condenser cooled to 4°C to minimize evaporation. Fermentations were inoculated from shake flask precultures to a starting  $\text{OD}_{600}$  0.01, and  $\text{OD}_{600}$  1 for studies that evaluated the effect of inoculum concentration.

### Stirred Tank Continuous Fermentations

Aerobic, carbon limited, chemostat steady-state fermentations were operated at a dilution rate of 0.13  $\text{h}^{-1}$  (target: 0.1  $\text{h}^{-1}$ ) with a working volume of 1 L using the same stirred tank fermentation system

described for batch fermentations. The dissolved oxygen concentration was controlled above 40% air saturation by air sparging (1-3 vvm) and agitation (600-800 rpm). The working volume was kept constant by continuous removal of medium through a siphon connected to the effluent pump, where the weight of the fermenter and medium addition reservoir were continuously monitored. The chemostat fermentations were inoculated from shake flask precultures to starting OD<sub>600</sub> 0.01. The medium composition was the same as for the batch cultivations, but the xylose concentration was 10 g l<sup>-1</sup>. The bioreactor was operated as a batch cultivation for approximately 76 h with a total working volume of 2 l. Thereafter, the working volume was reduced to 1 l by removing 1 l of medium, the feed was started, and the fermentation was operated as a chemostat. After 75 h of continuous mode operation the xylose concentration in the feed was increased to 40 g l<sup>-1</sup>. Steady state was reached after at least 5 residence times, where the changes in growth conditions characterized by the measurement of CO<sub>2</sub> evolution rate, O<sub>2</sub> consumption rate, and biomass (measured with two different methods: OD<sub>600</sub> and dry weight) concentration were constant (<10% deviation). Control of substrate addition was performed by manual adjustment of the substrate feed pump based on substrate balance measurements. Chemostat cultures were routinely checked for potential bacterial and fungal infection by phase-contrast microscopy.

## Analysis

### Fermentation Off-gas Analysis

The effluent fermentation gas was measured every 30 seconds for determination of oxygen and carbon dioxide concentration by the off-gas analyzer Brüel and Kjaer 1308 [55], based on photoacoustic and magnetoacoustic detection techniques for CO<sub>2</sub> and O<sub>2</sub>, respectively.

### Cell Mass Determination

The optical density was determined at 600 nm using a Shimadzu UV mini 1240 spectrophotometer. Samples were diluted with deionized water to obtain OD<sub>600</sub> measurements in the linear range of 0-0.8 OD<sub>600</sub> units. The first OD<sub>600</sub> was measured 12 hours after inoculation, and afterward every 2-4 hours to be able to establish  $\mu_{\max}$  for the culture.

Dry weight measurements were determined throughout the exponential phase, until stationary phase was confirmed according to OD<sub>600</sub> and off-gas analysis. Nitrocellulose filters (0.45 µm Satorius AG) were used. The filters were pre-dried in a microwave oven at 150 W for 10 min and then cooled in a desiccator for 10 min. Five ml of cell medium were filtered and the residue was washed with deionized water. Filters were dried in a microwave oven for 15 min at 150 W, cooled for 15 min in a desiccator, and the mass was determined again using an analytical balance [56]. For both OD<sub>600</sub> and dry weight, duplicate measurements were made and biomass concentration (g-dry weight l<sup>-1</sup>) was determined based on dry weight.

### Extracellular Metabolite Analysis

During shake flask cultivations samples were taken at inoculation, mid-exponential phase, and early stationary phases. During fermentations samples were taken every 2 hours after 12 hours post-inoculation. Samples were filtered immediately using a 0.45 µm syringe-filter (Satorius AG) and stored at -20°C until further analysis. Glucose, xylose, ethanol, glycerol, acetate, succinate and xylitol concentrations were determined by HPLC analysis using an Aminex HPX-87H ion-exclusion column from Bio-Rad. The column was maintained at 65°C and elution performed using 5 mM H<sub>2</sub>SO<sub>4</sub> as the mobile phase at a flow rate of 0.6 ml/min. Glucose, xylose, ethanol, glycerol, acetate, succinate were detected on a Waters 410 differential refractometer detector (from Shodex, Kawasaki, Japan), whereas acetate, pyruvate and xylitol were detected on a Waters 468 absorbance detector set at 210 nm (the two detectors were connected in series). Using Dionex Chromeleon® software and six internal standards, concentration data were found from chromatograms and exported to Microsoft® Excel for processing.

## Transcriptomics

### RNA Sampling and Isolation

Samples for RNA isolation from the late-exponential phase of glucose-limited and xylose-limited batch, and continuous cultivations were taken by rapidly sampling 25 ml of culture into a 50 ml sterile Falcon tube with 40 ml of crushed ice in order to decrease the sample temperature below 2°C in less than 10 seconds. Cells were immediately centrifuged (4000 RPM at 0°C for 2.5 min.), the supernatant discarded, and the pellet frozen in liquid nitrogen and it was stored at -80°C until total RNA extraction. Total RNA was extracted using the RNeasy® Mini Kit (Qiagen, Valencia, CA) according to manufacturer's instructions after partially thawing the samples on ice. RNA sample integrity and quality was determined prior to hybridization with an Agilent 2100 Bioanalyzer and RNA 6000 Nano LabChip kit according to the manufacturer's instruction (Agilent, Santa Clara, CA).

### Probe Preparation and Hybridization to DNA Microarrays

Messenger RNA (mRNA) extraction, cDNA synthesis, labeling, and array hybridization to Affymetrix Yeast Genome Y2.0 arrays were performed according to the manufacturer's recommendations (Affymetrix GeneChip® Expression Analysis Technical Manual, 2005-2006 Rev. 2.0). Washing and staining of arrays were performed using the GeneChip Fluidics Station 450 and scanning with the Affymetrix GeneArray Scanner (Affymetrix, Santa Clara, CA).

### Microarray Gene Transcription Analysis

Affymetrix Microarray Suite v5.0 was used to generate CEL files of the scanned DNA microarrays. These CEL files were then processed using the statistical language and environment R v2.9.1 (R Development Core Team, 2007, [www.r-project.org](http://www.r-project.org)), supplemented with Bioconductor v2.3 (Bioconductor Development Core Team, 2008, [www.bioconductor.org](http://www.bioconductor.org)) packages Biobase, affy, gcrma, and limma [57]. The probe intensities were normalized for background using the robust multiarray average (RMA) method only using perfect match (PM) probes after the raw image file of the DNA microarray was visually inspected for acceptable quality. Normalization was performed using the qspline method and gene expression values were calculated from PM probes with the median polish summary. Statistical analysis was applied to determine differentially expressed genes using the limma statistical package. Moderated *t*-tests between the sets of experiments were used for pair-wise comparisons. Empirical Bayesian statistics were used to moderate the standard errors within each gene and Benjamini-Hochberg's method was used to adjust for multi-testing. A cut-off value of adjusted *p*<0.01 (referred to as *p<sub>adjusted</sub>*) was used for statistical significance, unless otherwise specified [58]. Gene ontology process annotation was performed by submitting differentially expressed gene (adjusted *p*<0.01) lists to the *Saccharomyces* Genome Database GO Term Finder resource and maintaining a cut-off value of *p*<0.01 for hypergeometric testing of cluster frequency compared to background frequency [59]. Metabolic pathway mapping was performed using Pathway Expression Viewer of the *Saccharomyces* Genome Database, where lists of differentially expressed genes (*p<sub>adjusted</sub>*<0.01, |log-fold change|>1) between two conditions were submitted [60].

**Competing interests**

The authors declare they have no competing interests.

**Author's contributions**

JMO, GS, JVV, JN, LO participated in the design of the study. JMO, GS performed the experimental work. JMO, GS wrote the manuscript. JVV, TWJ, LO, and JN edited the manuscript. All the authors have read and approved the final manuscript.

**Acknowledgements**

José M. Otero is a Merck Doctoral Fellow and acknowledges financial support from the Division of Bioprocess Research & Development, Merck Research Labs, Merck & Co., Inc. Part of this research has been financed by the Chalmers Foundation.

## References:

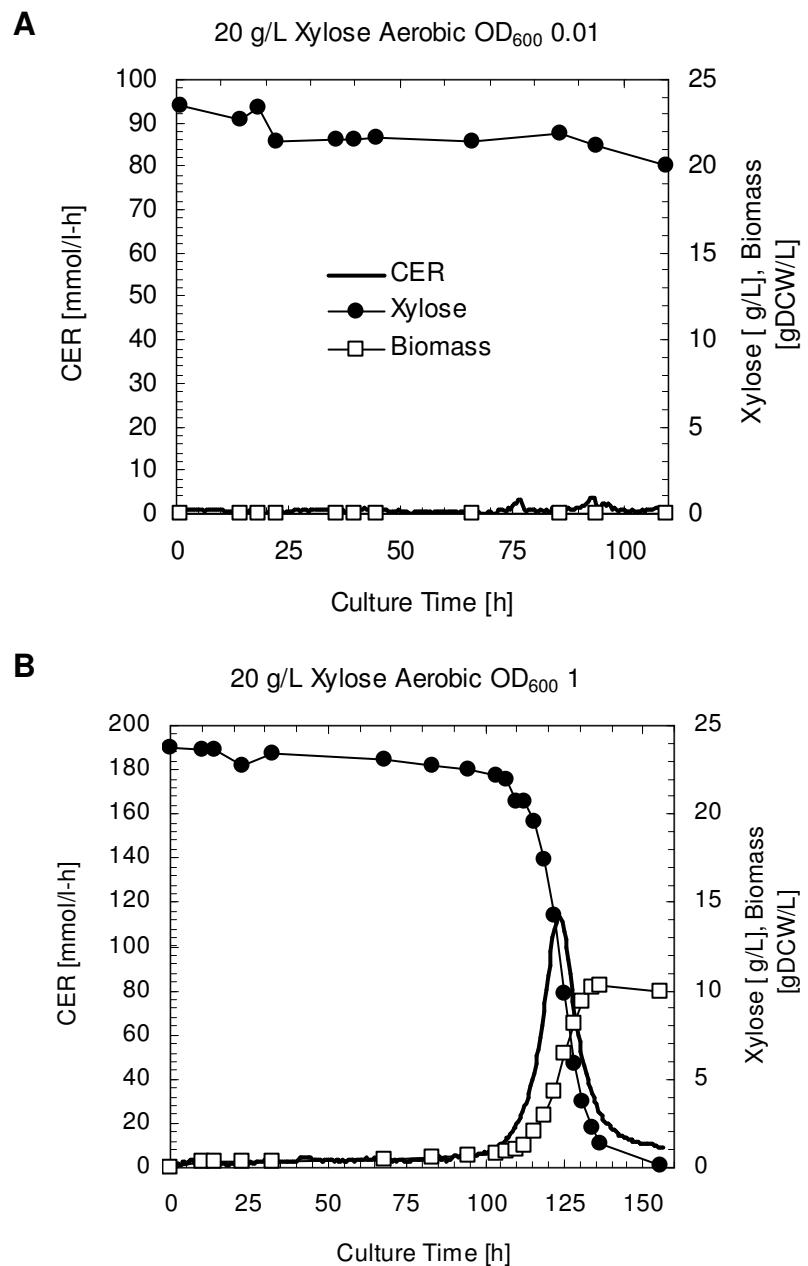
1. Olsson L, Jørgensen H, Krogh K, Roca C. **Bioethanol Production from Lignocellulosic Material.** In *Polysaccharides: Structural Diversity and Functional Versatility*. 2<sup>nd</sup> edition. Edited by Dumitriu S. New York: Marcel Dekker; 2004: 957-993.
2. Otero JM, Panagiotou G, Olsson L: **Fueling Industrial Biotechnology Growth with Bioethanol.** *Adv Biochem Eng Biotechnol* 2007, **108**:1-40.
3. Chemler JA, Yan Y, Koffas MA: **Biosynthesis of isoprenoids, polyunsaturated fatty acids and flavonoids in *Saccharomyces cerevisiae*.** *Microb Cell Fact* 2006, **5**: 20.
4. Ostergaard S, Olsson L, Nielsen J: **Metabolic engineering of *Saccharomyces cerevisiae*.** *Microbiol Mol Biol Rev* 2000, **64**: 34-50.
5. Batt CA, Carvalho S, Easson DD, Akedo M, Sinskey AJ: **Direct evidence for xylose metabolic pathway in *Saccharomyces cerevisiae*.** *Biotechno Bioeng* 1986, **28**: 549-553.
6. van Zyl C, Prior BA, Kilian SG, Kock JL: **D-xylose utilization by *Saccharomyces cerevisiae*.** *J Gen Microbio* 1989, **135**: 2791-2798.
7. Lee J, Kim MD, Ryu WY, Bisson FL, Seo HL: **Kinetic studies on glucose and xylose transport in *Saccharomyces cerevisiae*.** *Appl Microbiol Biotechnol* 2002, **60**: 186-191.
8. Webb SR, Lee H: **Regulation of D-xylose utilization by exoses in pentose-fermenting yeasts.** *Biotechnol Adv.* 1990, **8**: 685-697.
9. Mishra P, Singh A: **Microbial pentose utilization.** *Adv App Microbiol* 1993, **39**: 91-152.
10. Harhangi HR, Akhmanova AS, Emmens R, van der Drift C, de Laat WT, van Dijken PJ, Jetten MS, Pronk JT, Op den Camp HJ: **Xylose metabolism in the fungus *Piromyces sp.* strain E2 follow the bacterial pathway.** *Arch Microbiol* 2003, **180**: 134-141.
11. Kötter P, Ciriacy M: **Xylose fermentation by *Saccharomyces cerevisiae*.** *Appl Microbiol Biotechnol* 1993, **38**: 776-783.
12. Kötter P, Amore R, Hollenberg CP, Ciriacy M: **Isolation and characterization of *Pichia stipitis* xylitol dehydrogenase gene, XYL2, and construction of a xylose-utilizing *Saccharomyces cerevisiae* transformant.** *Current Genetics* 1990, **18**: 493-500.
13. Tantirungkij M, Nakashima N, Seki T, Yoshida T: **Construction of xylose-assimilating *Saccharomyces cerevisiae*.** *J Ferm Bioeng* 1993, **75**: 83-88.
14. Walfridsson M, Hallborn J, Penttilä M, Keränen S, Hahn-Hägerdal B: **Xylose-metabolizing *Saccharomyces cerevisiae* strains overexpressing TKL1 and TAL1 genes encoding the pentose phosphate pathway enzymes transketolase and transaldolase.** *Appl Environ Microbiol* 1995, **61**: 4184-4190.
15. Tiovári HM, Aristidou A, Ruohonen L, Penttilä M: **Conversion of xylose to ethanol by recombinant *Saccharomyces cerevisiae*: importance of xylulokinase (XKS1) and oxygen availability.** *Metab Eng* 2001, **3**: 236-249.
16. Ho NW, Chen Z, Brainard AP: **Genetically engineered *Saccharomyces* yeast capable of effective co-fermentation of glucose xylose.** *Appl Environ Microbiol* 1998, **64**: 1852-1859.
17. Eliasson A, Christensson C, Wahlbom FC, Hahn-Hägerdal B: **Anaerobic xylose fermentation by recombinant *Saccharomyces cerevisiae* carrying XYL1, XYL2, and XKS1 in mineral medium chemostat cultures.** *Appl Environ Microbiol* 2000, **66**: 3381-3386.
18. Amore R, Kötter P, Kuster C, Ciriacy M, Hollenberg CP: **Cloning and expression in *Saccharomyces cerevisiae* of the NAD(P)H-dependent xylose reductase-encoding gene (XYL1) from the xylose-assimilating yeast *Pichia stipitis*.** *Gene* 1991, **109**: 89-97.
19. Kuyper M, Harhangi HR, Stave AK, Winkler AA, Jetten MS, de Laat WT, den Ridder JJ, Op den Camp HJ, van Dijken JP, Pronk JT: **High-level functional expression of fungal xylose isomerase: the key to efficient ethanolic fermentation of xylose by *Saccharomyces cerevisiae*?** *FEMS Yeast Res* 2003, **4**: 69-78.
20. Kuyper M, Winkler AA, van Dijken JP, Pronk JT: **Minimal metabolic engineering of *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: proof of principle.** *FEMS Yeast Res* 2004, **4**: 655-

664.

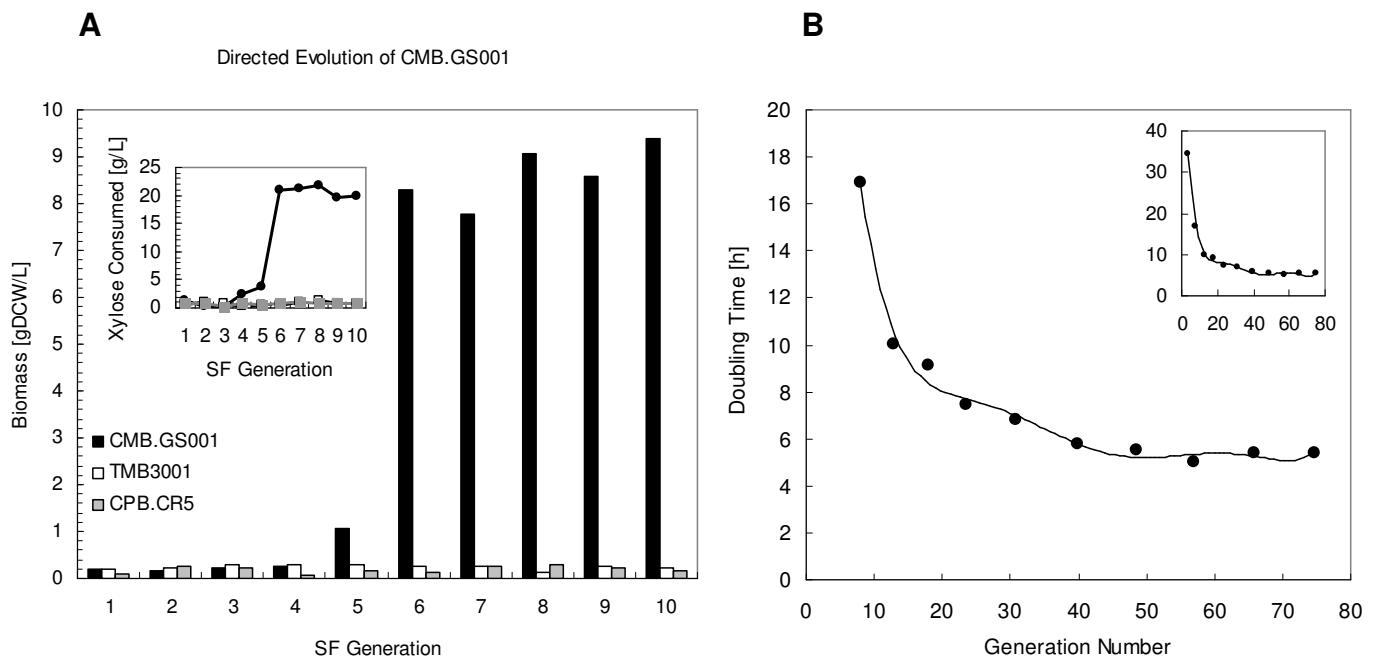
21. Kuyper M, Toirkens MJ, Diderich JA, Winkler AA, van Dijken JP, Pronk JT: **Evolutionary engineering of mixed-sugar utilization by a xylose-fermenting *Saccharomyces cerevisiae* strain.** *FEMS Yeast Res* 2005, **5**: 925-934.
22. van Maris AJ, Winkler AA, Kuyper M, de Laat WT, van Dijken JP, Pronk JT: **Development of efficient xylose fermentation in *Saccharomyces cerevisiae*: xylose isomerase as a key component.** *Adv Biochem Eng Biotechnol* 2007, **108**: 179-204.
23. Matsushika A, Inoue H, Kodaki T, Sawayama S: **Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: current state and perspectives.** *Appl Microbiol Biotechnol* 2009, **84**: 37-53.
24. Hahn-Hägerdal B, Karhumaa K, Jeppson M, Gorwa-Grauslund MF: **Metabolic engineering for pentose utilization in *Saccharomyces cerevisiae*.** *Adv Biochem Eng Biotechnol* 2007, **108**: 147-177.
25. Gárdonyi M, Jeppsson M, Lidén G, Gorwa-Grauslund MF, Hahn-Hägerdal B: **Control of xylose consumption by xylose transport in recombinant *Saccharomyces cerevisiae*.** *Biotechnol Bioeng* 2003, **82**: 818-824.
26. Roca C, Nielsen J, Olsson L: **Metabolic engineering of ammonium assimilation in xylose-fermenting *Saccharomyces cerevisiae* improves ethanol production.** *Appl Environ Microbiol* 2003, **69**: 4732-4736.
27. Skoog K, Hahn-Hägerdal B: **Effect of oxygenation on xylose fermentation by *Pichia stipitis*.** *Appl Environ Microbiol* 1990, **56**: 3389-3394.
28. du Preez JC: **Process parameters and environmental factors affecting D-xylose fermentation by yeast.** *Enzyme Microbiology and Technology* 1994, **16**: 944-956.
29. Boles E, Lehnert W, Zimmermann FK: **The role of NAD-dependent glutamate dehydrogenase in restoring growth on glucose of *Saccharomyces cerevisiae* phosphoglucose isomerase mutant.** *Eur J Biochem* 1993, **217**: 469-477.
30. Müller S, Boles E, May M, Zimmermann FK: **Different internal metabolites trigger the induction of glycolytic gene expression in *Saccharomyces cerevisiae*.** *J Bacteriol* 1995, **177**: 4517-4519.
31. Jin YS, Lapaza JM, Jeffries TW: ***Saccharomyces cerevisiae* engineered for xylose metabolism exhibits a respiratory response.** *Appl Environ Microbiol* 2004, **70**: 6816-6825.
32. Grotkjaer T, Christakopoulos P, Nielsen J, Olsson L: **Comparative metabolic network analysis of two xylose fermenting recombinant *Saccharomyces cerevisiae* strains.** *Metab Eng* 2005, **7**: 437-444.
33. Boer VM, de Winde JH, Pronk JT, Piper MD: **The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur.** *J Biol Chem* 2003, **8**: 3265-3274.
34. Regenberg B, Grotkjaer T, Winther O, Fausbøll A, Åkesson M, Bro C, Hansen LK, Brunak S, Nielsen J: **Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in *Saccharomyces cerevisiae*.** *Genome Biol* 2006, **7**: R107.
35. Senac T, Hahn-Hägerdal: **Effects of increased transaldolase activity on D-xylulose and D-glucose metabolism in *Saccharomyces cerevisiae* cell extracts.** *Appl Environ Microbiol* 1991, **57**: 1701-1706.
36. Walfridsson M, Hallborn J, Penttilä M, Keränen S, Hahn-Hägerdal B: **Xylose-metabolizing *Saccharomyces cerevisiae* strains overexpressing the TKL1 and TAL1 genes encoding the pentose phosphate pathway enzymes transketolase and transaldolase.** *Appl Environ Microbiol* 1995, **61**: 4184-4189.
37. Jeffries TW: **Engineering yeasts for xylose metabolism.** *Curr Opin Biotechnol* 2006, **17**: 320-326.
38. Cherry JM, Ball C, Weng S, Juvik G, Schmidt R, Adler C, Dunn B, Dwight S, Riles L, Mortimer RK, Botstein D: **Genetic and physical maps of *Saccharomyces cerevisiae*.** *Nature* 1997, **387**: 67-73.
39. Minard KI, Jennings GT, Loftus TM, Xuan D, McAlister-Henn L: **Sources of NADPH and expression of mammalian NADP+-specific isocitrate dehydrogenases in *Saccharomyces cerevisiae*.** *J Biol Chem* 1998, **273**: 31486-31493.
40. Thomas D, Cherest H, Surdin-Kerjan Y: **Identification of the structural gene for glucose-6-phosphate dehydrogenase in yeast.** *EMBO J* 1991, **10**: 547-553.
41. Loftus TM, Hall LV, Anderson SL, McAlister-Henn L: **Isolation, characterization, and disruption of the**

- yeast gene encoding cytosolic NADP-specific isocitrate dehydrogenase.** *Biochemistry* 1994, **33**: 9661-9667.
42. DeRisi JL, Iyer VR, Brown PO: **Exploring the metabolic and genetic control of gene expression on genomic scale.** *Science* 1997, **278**: 680-686.
  43. Minard KI, McAlister-Henn L: **Sources of NADPH in yeast vary with carbon source.** *J Biol Chem* 2005, **280**: 39890-39896.
  44. Eliasson A, Hofmeyr J-HS, Pedler S, Hahn-Hägerdal B: **The xylose reductase/xylitol dehydrogenase/xylulokinase ratio affects product formation in recombinant xylose-utilising *Saccharomyces cerevisiae*.** *Enzyme Microbiology and Technology* 2001, **29**: 288-297.
  45. Wahlbom CF, Hahn-Hägerdal B: **Furfural, 5-hydroxymethyl furfural, and acetoin act as external electron acceptors during anaerobic fermentation of xylose in recombinant *Saccharomyces cerevisiae*.** *Biotechnol Bioeng* 2002, **78**: 172-178.
  46. Jeppsson M, Träff-Bjerre KL, Johansson B, Hahn-Hägerdal B, Gorwa-Grauslund MF: **Effect of enhanced xylose reductase activity on xylose consumption and product distribution in xylose-fermenting recombinant *Saccharomyces cerevisiae*.** *FEMS Yeast Res* 2003, **3**: 167-175.
  47. Verho R, Londesborough J, Penttilä M, Richard P: **Engineering redox cofactor regeneration for improved pentose fermentation in *Saccharomyces cerevisiae*.** *Appl Environ Microbiol* 2003, **69**: 4732-4736.
  48. Träff-Bjerre KL, Jeppsson M, Hahn-Hägerdal B, Gorwa-Grauslund MF: **Endogenous NADPH-dependent aldose reductase activity influences products formation during xylose consumption in recombinant *Saccharomyces cerevisiae*.** *Yeast* 2003, **21**:141-150.
  49. Watanabe S, Saleh AA, Pack SP, Annaluru N, Kodaki T, Makino K: **Ethanol production from xylose by recombinant *Saccharomyces cerevisiae* expressing protein engineered NADP<sup>+</sup>-dependent xylitol dehydrogenase.** *J Biotechnol* 2007, **130**: 316-319.
  50. van Zyl WH, Eliasson A, Hobley T, Hahn-Hägerdal B: **Xylose utilization by recombinant strains of *Saccharomyces cerevisiae* on different carbon source.** *Appl Microbiol Biotechnol* 1999, **52**: 829-833
  51. Gietz RD, Woods RA: **Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method.** *Methods Enzymol* 2002, **350**: 87-96.
  52. Verduyn C, Postma E, Scheffers A, van Dijken P: **Effect of benzoic acid on metabolic fluxes in yeasts: a continuous culture study on the regulation of respiration and alcoholic fermentation.** *Yeast* 2002, **8**: 501-517.
  53. Andreasen AA, Stier TJB: **Anaerobic nutrition of *Saccharomyces cerevisiae*. I. Ergosterol requirement for growth in a defined medium.** *J Cellular Comp Physiol* 1953, **41**: 23-36.
  54. Andreasen AA, Stier TJB: **Anaerobic nutrition of *Saccharomyces cerevisiae*. II. Unsaturated fatty acid requirement for growth in a defined medium.** *J Cellular Comp Physiol* 1954, **43**: 271-281.
  55. Christensen LH, Schulze U, Nielsen J, Villadsen J: **Acoustic off gas analyzer for bioreactors: precision, accuracy and dynamics of detection.** *Chemical Eng Science* 1995, **50**: 2601-2610.
  56. Nielsen J, Olsson L: **On-Line *in situ* monitoring of biomass in submerged cultivations.** *Trends in Biotechnol* 1997, **15**: 517-522.
  57. Smyth GK: **Limma: linear models for microarray data.** In: *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Edited by Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W. New York: Springer; 2005: 397-420.
  58. Smyth GK: **Linear models and empirical Bayes methods for assessing differential expression in microarray experiments.** *Statistical Applications in Genetics and Molecular Biology* 2004, **3**: 1(3).
  59. Ball CA, Dolinski K, Dwight SS, Harris MA, Issel-Tarver L, Kasarskis A, Scafe CR, Sherlock G, Binkley G, Jin H, Kaloper M, Orr SD, Schroeder M, Weng S, Zhu Y, Botstein D, Cherry JM: **Integrating functional genomic information into *Saccharomyces* genome database.** *Nucleic Acids Res* 2000, **28**: 77-80.
  60. Ball CA, Jin H, Sherlock G, Weng S, Matese JC, Andrada R, Binkley G, Dolinski K, Dwight SS, Harris MA, Issel-Tarver L, Schroeder M, Botstein D, Cherry JM: **Saccharomyces Genome Database provides tools to survey gene expression and functional analysis data.** *Nucleic Acids Res* 2001, **29**: 80-81.

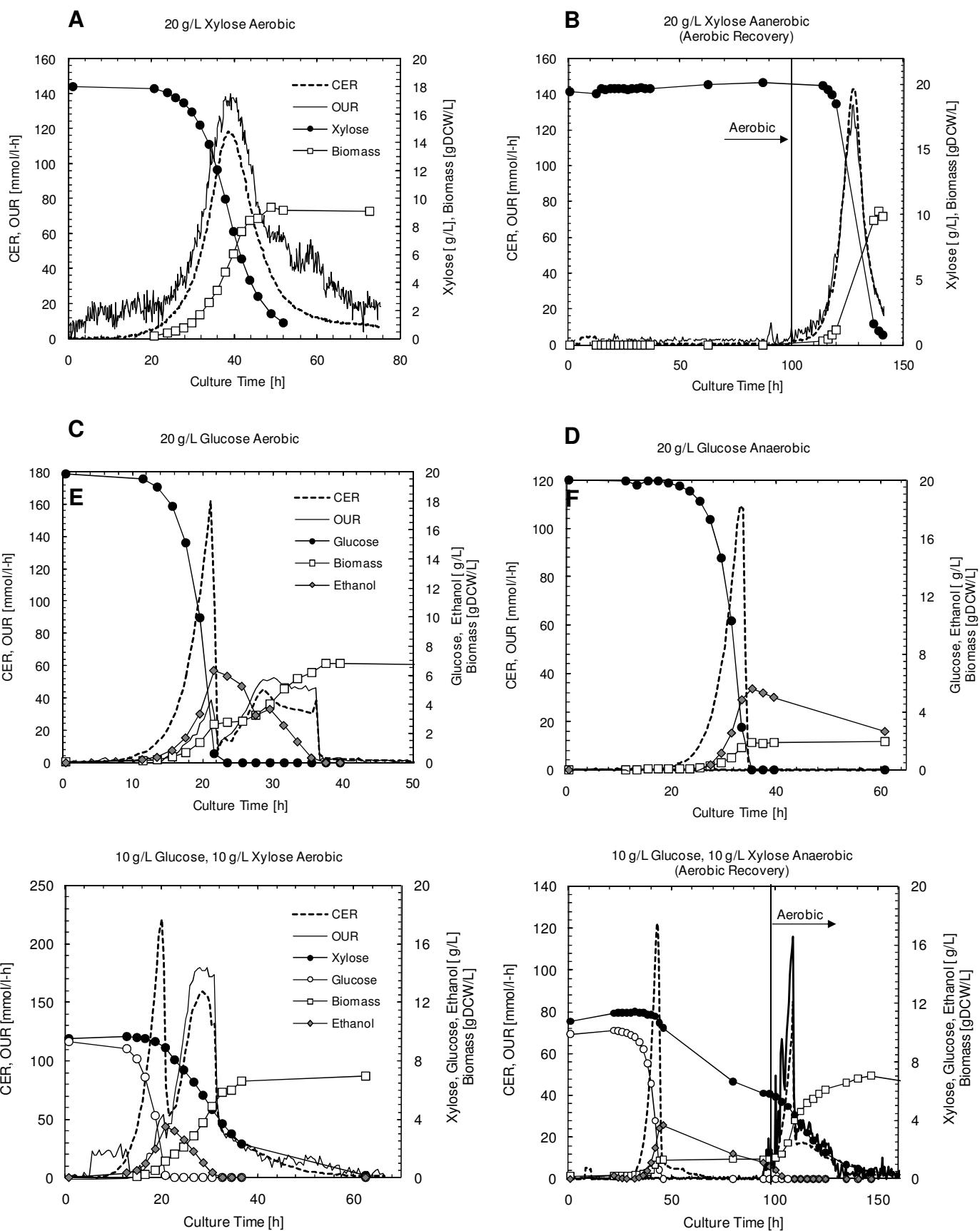
61. van Dijken JP, Bauer J, Brambilla L, Duboc P, Francois MJ, Gancedo C, Giuseppin MLF, Heijnen JJ, Hoare M, Lange HC, Madden EA, Niederberg P, Nielsen J, Parrou LJ, Petit T, Porro D, Reuss M, van Riel N, Rizzi M, Steensma HY, Verrips CT, Vindelov J, Pronk JT: **An interlaboratory comparison of physiological and genetic proprieties of four *Saccharomyces cerevisiae* strains.** *Enzyme and Microbial Technology* 2000, **26**: 706-741.
62. Guldener U, Heck S, Fieldler T, Beinhauer J, Hegemann JH: **A new efficient gene disruption cassette for repeated use in budding yeast.** *Nucleic Acids Res* 1996, **13**: 2519-2524.
63. Nissen TL, Kielland-Brandt MC, Nielsen J, Villadsen J: **Optimization of Ethanol Production in *Saccharomyces cerevisiae* by Metabolic Engineering of the Ammonium Assimilation.** *Metab Eng* 2000, **2**: 69-77.
64. Moreira dos Santos M, Thygesen G, Kötter P, Olsson L, Nielsen J: **Aerobic physiology of redox-engineered *Saccharomyces cerevisiae* strain modified in the ammonium assimilation for increased NADPH availability.** *FEM Yeast Res* 2003, **4**: 59-68.
65. Haiying N, Laplaza JM, Jeffries TW: **Transposon Mutagenesis to Improve the Growth of Recombinant *Saccharomyces cerevisiae* on D-Xylose.** *Appl Environ Microbiol* 2007, **73**: 2061-2066.
66. Karhumaa K, Fromanger R, Hahn-Hägerdal B, Gorwa-Grauslund MF: **High activity of xylose reductase and xylitol dehydrogenase improves xylose fermentation by recombinant *Saccharomyces cerevisiae*.** *Appl Microbiol Biotechnol* 2006, **73**: 1039-1046.
67. Karhumaa K, Garcia Sanchez R, Hahn-Hägerdal B, Gorwa-Grauslund MF: **Comparison of the xylose reductase-xylitol dehydrogenase and the xylose isomerase pathways for xylose fermentation by recombinant *Saccharomyces cerevisiae*.** *Microb Cell Fact* 2007, **6**: 5.
68. Wahlbom CF, Cordero Otero RR, van Zyl WH, Hahn-Hägerdal B, Jönsson LJ: **Molecular analysis of a *Saccharomyces cerevisiae* mutant with improved ability to utilize xylose shows enhanced expression of proteins involved in transport, initial xylose metabolism, and the pentose phosphate pathway.** *Appl Environ Microbiol* 2003, **69**: 740-746.
69. Wahlbom CF, van Zyl WH, Jönsson LJ, Hahn-Hägerdal B, Otero RR: **Generation of the improved recombinant xylose-utilizing *Saccharomyces cerevisiae* TMB 3400 by random mutagenesis and physiological comparison with *Pichia stipitis* CBS 6054.** *FEMS Yeast Res* 2003, **3**: 319-326.
70. Pitkänen JP, Rintala E, Aristidou A, Ruohonen L, Penttilä M: **Xylose chemostat isolates of *Saccharomyces cerevisiae* show altered metabolite and enzyme levels compared with xylose, glucose, and ethanol metabolism of the original strain.** *Appl Microbiol Biotechnol* 2005, **67**: 827-837.
71. Sonderegger M, Sauer U: **Evolutionary engineering of *Saccharomyces cerevisiae* for anaerobic growth on xylose.** *Appl Environ Microbiol* 2003, **69**: 1990-1998.
72. Jin YS, Alper H, Yang YT, Stephanopoulos G: **Improvement of xylose uptake and ethanol production in recombinant *Saccharomyces cerevisiae* through an inverse metabolic engineering approach.** *Appl Environ Microbiol* 2005, **71**: 8249-8256.
73. Karhumaa K, Hahn-Hägerdal B, Gorwa-Grausland MF: **Investigation of limiting metabolic steps in the utilization of xylose by recombinant *Saccharomyces cerevisiae* using metabolic engineering.** *Yeast* 2005, **22**: 359-368.



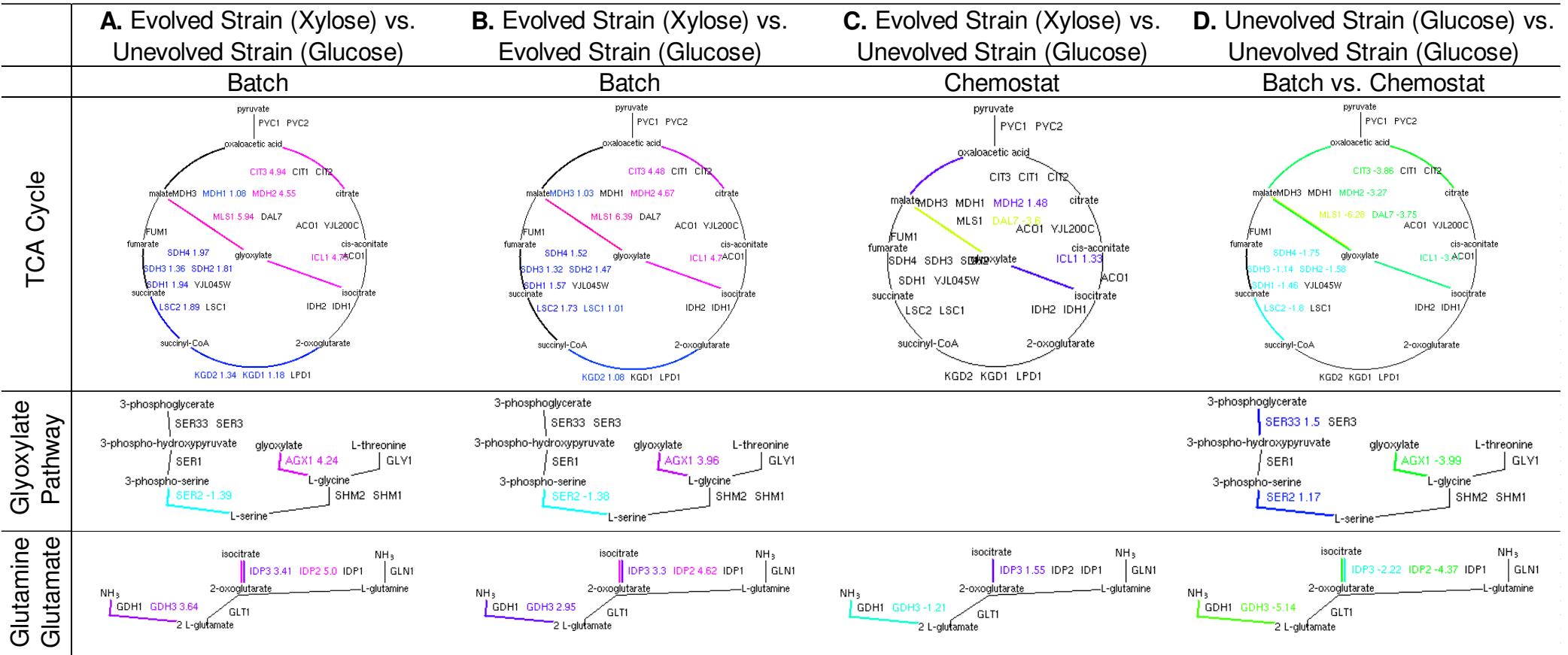
**Figure 1.** Aerobic fermentation growth profile of strain CMB.GS001 on defined minimal medium supplemented with  $20 \text{ g l}^{-1}$  xylose. Initial  $OD_{600}$  0.01 (A),  $OD_{600}$  1.0 (B). Carbon evolution rate (CER).



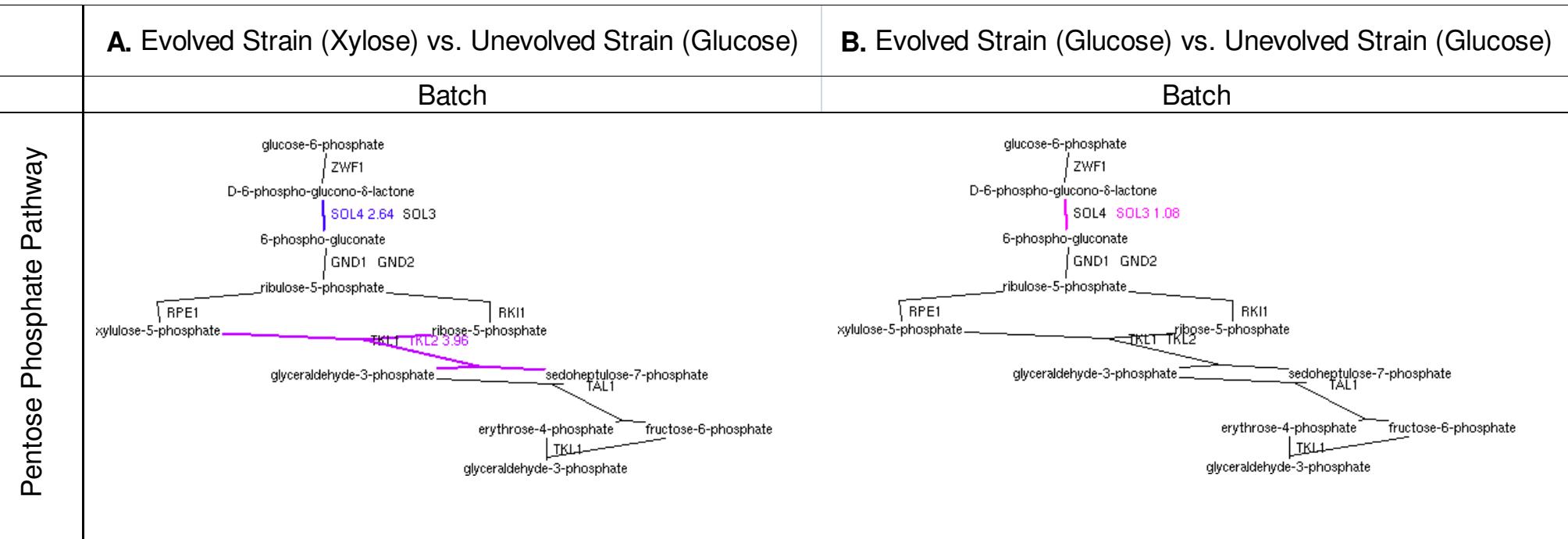
**Figure 2.** (A) Comparison in xylose consumption and biomass production during repetitive growth of *S. cerevisiae* TMB3001, CPB.CR5, and CMB.GS001 in shake flask cultures on synthetic medium with 20 g l<sup>-1</sup> xylose. Shake flask generation represents the number of specific shake flasks in the series of repetitive cultivations performed to select for mutants with higher specific growth rates and xylose utilization rates. (B) Doubling time during the serial transfers of *S. cerevisiae* CMB.GS002-010 in shake flask cultures on synthetic medium with 20 g l<sup>-1</sup> xylose as a function of the number of cell generations. Each data point represents the doubling time of a single shake flask culture estimated from OD<sub>600</sub> measurements. The small plot in the top right represents all 10 cycles, noting the initial doubling time of CMB.GS001 of 35h, and the rapid decrease to 10 h within less than 20 cell generations. For cell generations 50-74 there was no significant improvement in the specific growth rate.



**Figure 3.** For all plots presented, carbon evolution rate (CER), oxygen uptake rate (OUR), and xylose, glucose, ethanol and biomass concentrations as functions of cultivation time for CMB.GS010 (evolved strain). (A) 20 g l<sup>-1</sup> xylose aerobic batch culture. (B) 20 g l<sup>-1</sup> xylose anaerobic batch culture that was switched to aerobic at 100 h after no growth was observed. (C) 20 g l<sup>-1</sup> glucose aerobic batch culture. (D) 20 g l<sup>-1</sup> glucose anaerobic batch culture. (E) Mixed substrate cultivation with 10 g l<sup>-1</sup> glucose, 10 g l<sup>-1</sup> xylose aerobic batch culture. (F) Mixed substrate cultivation with 10 g l<sup>-1</sup> glucose, 10 g l<sup>-1</sup> xylose anaerobic batch culture that was switch to aerobic at 100 h after minimal growth was observed.



**Figure 4.** Three central carbon metabolic pathways are presented: (1) tricarboxylic acid (TCA) cycle, (2) glyoxylate pathway, and (3) glutamine/glutamate synthesis. The log-fold change of significantly differentially expressed genes ( $p_{\text{adjusted}} < 0.01$ ,  $|\log\text{-fold change}| > 1$ ) is indicated next to the gene name. These metabolic maps are provided by the *Saccharomyces* Genome Database Pathway Expression Viewer. The comparative conditions evaluated include: (A) CMB.GS010 cultivated on batch xylose vs. CMB.GS001 cultivated on batch glucose, (B) CMB.GS010 cultivated on batch xylose vs. CMB.GS010 cultivated on batch glucose, (C) CMB.GS010 cultivated on xylose in continuous culture (chemostat) vs. CMB.GS001 cultivated on glucose in continuous culture, (D) CMB.GS001 cultivated on batch glucose vs. CMB.GS001 cultivated on batch continuous culture. The terms evolved and CMB.GS010, and unevolved and CMB.GS001, are used interchangeably. If no pathway is shown for a given comparative condition then no significant differential gene expression was detected in that pathway.



**Figure 5.** The pentose phosphate (PP) pathway is presented. The log-fold change of significantly differentially expressed genes ( $p_{\text{adjusted}} < 0.01$ ,  $|\log\text{-fold change}| > 1$ ) is indicated next to the gene name. These metabolic maps are provided by the *Saccharomyces* Genome Database Pathway Expression Viewer. The comparative conditions evaluated include: (A) CMB.GS010 cultivated on batch xylose vs. CMB.GS001 cultivated on batch glucose, (B) CMB.GS010 cultivated on batch glucose vs. CMB.GS001 cultivated on batch glucose. The terms evolved and CMB.GS010, and unevolved and CMB.GS001, are used interchangeably.

**Table 1.** *S. cerevisiae* Strain and Plasmid Genotypes

Strain or Plasmid	Relevant Genotype	Origin/Reference
CEN.PK 113-7D	<i>MATa URA3 HIS3 LEU2 TRP1 SUC2 MAL2-8<sup>c</sup></i>	SRD GmbH <sup>a</sup>
CEN.PC 113-3C	<i>MATa URA3 HIS3 LEU2 trp1-289 SUC2 MAL2-8<sup>c</sup></i>	SRD GmbH <sup>a</sup>
TMB3001	<i>MATa SUC2 MAL2-8<sup>c</sup> pADH-XYL1 pPGK-XYL2 pPGK-XKS1</i>	[17]
CPB.CR4	<i>MATa URA3 HIS3 LEU2 TRP1 SUC2 MAL2-8<sup>c</sup> pADH-XYL1 pPGK-XYL2 pPGK-XKS1 gdh1Δ pPGK-GDH2</i>	[26]
CPB.CR5	<i>MATa URA3 HIS3 LEU2 TRP1 SUC2 MAL2-8<sup>c</sup> pADH-XYL1 pPGK-XYL2 pPGK-XKS1 gdh1Δ pPGK-GLT1 pPGK-GLN1</i>	[26]
CMB.GS001	<i>MATa URA3 HIS3 LEU2 TRP1 SUC2 MAL2-8<sup>c</sup> pTDH3-PsXYL1 pTDH3-PsXYL2 pTDH3-PsXYL3</i>	This study
CMB.GS002-010 <sup>b</sup>	<i>MATa URA3 HIS3 LEU2 TRP1 SUC2 MAL2-8<sup>c</sup> pTDH3-PsXYL1 pTDH3-PsXYL2 pTDH3-PsXYL3 Evolved</i>	This study
CMB.GS011	<i>MATa URA3 HIS3 LEU2 TRP1 SUC2 MAL2-8<sup>c</sup> Evolved</i>	This study
CMB.GS012	<i>MATa URA3 HIS3 LEU2 TRP1 SUC2 MAL2-8<sup>c</sup> pTDH3-PsXYL1 pTDH3-PsXYL2 pTDH3-PsXYL3 Evolved and retransformed with native plasmid</i>	This study
CMB.GS013	<i>MATa URA3 HIS3 LEU2 TRP1 SUC2 MAL2-8<sup>c</sup> pTDH3-PsXYL1 pTDH3-PsXYL2 pTDH3-PsXYL3 Plasmid recovered from CMB.GS010 and retransformed into CMB.GS011</i>	This study
CMB.GS014	<i>MATa URA3 HIS3 LEU2 TRP1 SUC2 MAL2-8<sup>c</sup> pTDH3-PsXYL1 pTDH3-PsXYL2 pTDH3-PsXYL3 Plasmid recovered from CMB.GS010 and retransformed into CEN.PK 113-3C</i>	This study
pRS314-X123	<i>pTDH3-PsXYL1 pTDH3-PsXYL2 pTDH3-PsXKS1 (TRP1, Centromeric)</i>	[65]
YlpXR/XDH/XK	<i>pADH-XYL1 pPGK-XYL2 pPGK-XKS1 (HIS3, Integrative)</i>	[17]

<sup>a</sup> Scientific Research and Development GmbH, Oberursel, Germany<sup>b</sup> The three final digits of the strain identifier indicate from which cycle in the directed evolution the strain originated, with the starting strain referred to as CMB.GS001

**Table 2. Physiological Characterization of CMB.GS001 and CMB.GS010**

Strain	CMB.GS010					CMB.GS001		CEN.PK113-7D	
Carbon source	Xylose		Glucose		Glucose/Xylose	Xylose		Glucose	
Conditions	Aerobic (S.D.,%) <sup>a</sup>	Aerobic	Anaerobic	Aerobic <sup>b</sup>	Anaerobic (S.D.,%) <sup>ab</sup>	Aerobic <sup>c</sup>	Aerobic <sup>d</sup>	Anaerobic <sup>e</sup>	
Specific growth rate (h <sup>-1</sup> )	0.18	0.34	0.29	0.32	0.29	0.12	0.36	0.34 (0)	
<i>Sugar consumed</i> (C-mol/L)									
Glucose	-	0.64	0.67	0.31	0.33 (0.7)	-	0.66	0.70	
Xylose	0.53 (4.2)	-	-	0.05	0.03 (0.5)	0.75	-	-	
<i>Sugar consumption rate</i> (g/g-cell/h)									
Glucose	-	2.31	2.98	3.62	1.58	-	2.36	n.a.	
Xylose	0.31	-	-	0.25	0.06	0.26	-	-	
<i>Biomass Yield</i> (Cmol/Cmol)									
	0.62	0.16	0.12	0.19	0.19	0.56	0.15	0.12	
<i>Carbon recovery (%)</i>									
	100.1	105.7	98.2	103.6	103.7	94.9	103.7	104.0	
<i>Products</i> (C-moles/L)									
Biomass	0.315 (2.1)	0.11	0.08	0.07	0.07 (0)	0.56	0.10	0.08	
CO <sub>2</sub>	0.195 (2.1)	0.25	0.25	0.13	0.08 (1.4)	0.36	0.21	0.20	
Ethanol	0.015 (0.7)	0.28	0.24	0.14	0.16 (0.7)	0.01	0.34	0.36	
Xylitol	0	0	0	0	0	0.002	0	0	
Glycerol	0.003 (0.1)	0.02	0.07	0.01	0.05 (0)	0.002	0.04	0.07	
Acetate	0.004 (0.5)	0.02	0	0.02	0.006 (0)	0.01	0.004	0.01	
Succinate	0	0	0	0	0	0	0	0	
Pyruvate	0	0	0	0	0	0	0	0	
<i>Productivities</i> (g/g-cell/h)									
Biomass	0.16 (3.2)	0.29	0.29	0.43	0.19 (2.1)	0.12	0.35	n.a.	
CO <sub>2</sub>	0.03 (0.1)	0.53	0.53	0.83	0.20 (4.2)	0.03	0.49	n.a.	
Ethanol	0.02 (1.7)	0.76	0.93	1.39	0.61 (1.4)	0.01	0.96	n.a.	
Xylitol	0	0	0	0	0	0.003	0	n.a.	
Glycerol	0.002 (0)	0.1	0.33	0.17	0.22(0)	0.005	0.18	n.a.	
Acetate	0	0.08	0.05	0	0.04 (0.7)	0	0.02	n.a.	
Succinate	0	0	0	0	0	0	0	n.a.	

<sup>a</sup> Values are the average of two independent batch fermentation performed in duplicate (n=2).<sup>b</sup> Values relative to the second phase of growth when xylose is primarily consumed.<sup>c</sup> Values indicated are for an inoculation OD<sub>600</sub> of 1.0 as opposed to 0.1. Specific growth rate for a fermentation inoculated at OD<sub>600</sub> 0.1 was 0.02 h<sup>-1</sup>.<sup>d</sup> Values from Otero JM, unpublished.<sup>e</sup> Adapted from [21]

n.a., data not available

**TABLE 3. Summary of Differential Gene Expression Based on Strain, Carbon Source, and Cultivation Condition**

<b>Strain (Carbon Source) Comparison</b>	Evolved Strain (Xylose) vs. Unevolved Strain (Glucose)	Evolved Strain (Glucose) vs. Unevolved Strain (Glucose)	Evolved Strain (Xylose) vs. Evolved Strain (Glucose)	Evolved Strain (Xylose) vs. Unevolved Strain (Glucose)	Unevolved Strain (Glucose) vs. Unevolved Strain (Glucose)
	Batch	Batch	Batch	Chemostat	Batch vs. Chemostat
Total No. of Genes Differentially Expressed	479	63	377	231	428
Average LFC ± 95% CI	1.21 ± 0.20	0.79 ± 0.37	1.20 ± 0.24	0.01 ± 0.24	-1.46 ± 0.20
No. of Genes Up-regulated	331	47	259	127	105
Average LFC ± 95% CI	1.48 ± 0.16	1.58 ± 0.20	2.52 ± 0.04	1.51 ± 0.10	1.53 ± 0.12
No. of Genes Down-regulated	148	16	118	104	323
Average LFC ± 95% CI	-1.58 ± 0.11	-1.54 ± 0.25	-1.69 ± 0.13	-1.81 ± 0.19	-2.43 ± 0.15
Total No. of Metabolic Pathway Differentially Expressed	116	8	93	58	107
Average LFC ± SD	2.04 ± 2.21	1.61 ± 0.67	1.93 ± 2.45	0.52 ± 1.80	-1.59 ± 2.20

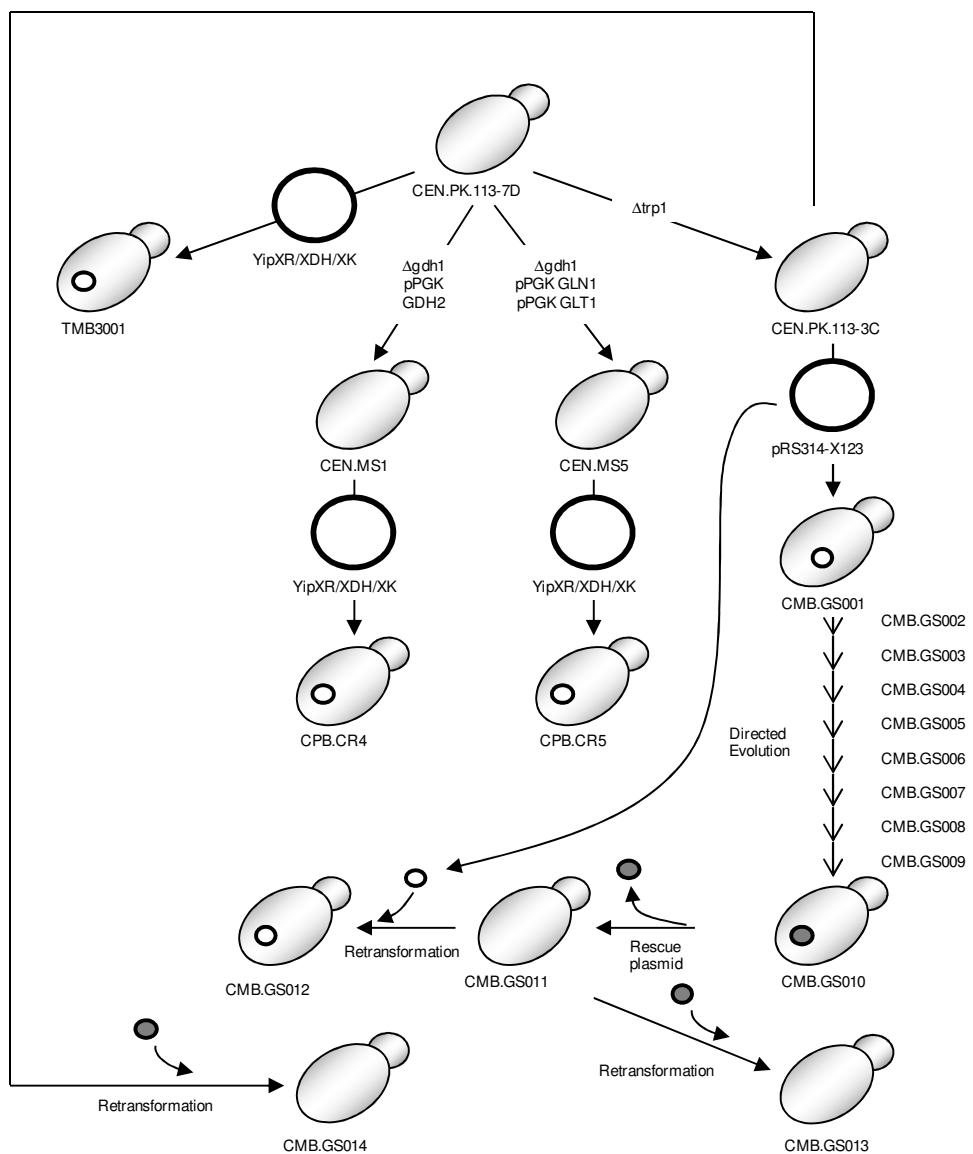
**Differential Gene Expression Statistics:**  $p_{adjusted} < 0.01$ ,  $|lfc| \geq 1$ ,  $n=2$

**Notes:** Confidence interval (CI); Standard deviation (SD).

## **Supplementary Materials**

### **Table of Contents:**

- 1. Supplementary Materials Figure 1:** Strain construction and plasmid rescue.
- 2. Supplementary Materials Text 1:** Strain construction genetic strategy.
- 3. Supplementary Materials List 1:** Fermentation medium recipe.
- 4. Supplementary Materials Table 1:** Shake flask strain characterization.
- 5. Supplementary Materials Table 2:** Xylose fermenting *S. cerevisiae* strains expressing XR, XDH, XK.
- 6. Supplementary Materials Table 3:** Production yield coefficients for aerobic and anaerobic batch fermentations.
- 7. Supplementary Materials Table 4:** Summary of transcriptome samples generated according to strain, carbon source, and cultivation condition.
- 8. Supplementary Materials Figure 2:** Principal component analysis of the normalized expression data.
- 9. Supplementary Materials Figure 3:** Gene ontology terms – Batch Evolved (Xylose) vs. Batch Evolved (Glucose), Batch Unevolved (Glucose) vs. Chemostat Unevolved (Glucose).
- 10. Supplementary Materials Figure 4:** Gene ontology terms – Batch Evolved (Xylose) vs. Batch Unevolved (Glucose), Batch Evolved (Glucose) vs. Batch Unevolved (Glucose), Chemostat Evolved (Xylose) vs. Chemostat Unevolved (Glucose).
- 11. Supplementary Materials Figure 5:** Metabolic pathway expression mapping of glycolysis and gluconeogenesis.



**Supplementary Materials Figure 1.** Schematic flow sheet of the construction of strains TMB3001, CPB.CR4, CPB.CR5, CMB.GS01 and CMB.GS10. Strain CEN.MS1 has been obtained deleting *GDH1* and over expressing *GDH2* in CEN.PK113-7D. Strain CEN.MS5 has been obtained deleting *GDH1* and over expressing *GLN1* and *GLT1* in CEN.PK113-7D. *GDH1* encodes for NADPH dependent glutamate dehydrogenase, *GDH2* encodes NADH dependent glutamate dehydrogenase, *GLN1* encodes glutamine synthetase, and *GLT1* encodes for glutamate synthase. Integrating vector YipXR/XDH/XK has been used to transform the strains CEN.PK113-7D, CEN.MS1, and CEN.MS5 yielding respectively, the strains TMB3001, CPB.CR4 and CPB.CR5. Centromeric plasmid pRS314-X123 was used to transform the parental strain CEN.PK113-7D yielding the strain CMB.GS001. Strain CMB.GS010 was derived from CMB.GS001 after cycles of repetitive culture selection in shake flasks. Strain CMB.GS011 was derived from CMB.GS010 after plasmid removal. Strain CMB.GS012 was obtained retransforming CMB.GS011 with the original plasmid pRS314-X123. Strain CMB.GS013 was obtained retransforming CMB.GS011 with the rescued plasmid. Strain CMB.GS014 was obtained retransforming CEN.PK 113-3C with the rescued plasmid.

## Supplementary Materials Text 1

Strain CEN.PK 113-3C carries a tryptophan auxotrophy by inactivation of *TRP1* encoding for N-(5' phosphoribosyl)-anthranilate isomerase, which catalyses the third step of the tryptophan biosynthetic pathway [59]. Strain TMB3001 features overexpression of xylose reductase (XR) and xylitol dehydrogenase (XDH) from *Pichia stipitis*, and the endogenous gene for xylolokinase (XKS), which have been integrated into the chromosome of CEN.PK 113-7D using the integrative plasmid YipXR/XDH/XK [17]. Strain CPB.CR4 has been constructed from the strain CEN.MS1 [64], originally derived from CEN.PK 113-7D, where *GDH1* encoding NADPH dependent glutamate dehydrogenase has been deleted using the loxP-KanMX-loxP disruption cassette [62] and *GDH2* encoding the NADH dependent glutamate dehydrogenase has been placed under a phosphoglycerate kinase (PGK) constitutive promoter [63]. The strain CEN.MS1 has been transformed with the plasmid YipXR/XDH/XK using the lithium acetate method [51] yielding the final strain CPB.CR4 [26]. Strain CPB.CR5 has been constructed from the strain CEN.MS5 [64], originally derived from the parental CEN.PK117-7D where *GDH1* has been deleted using the same method as in CEN.MS1. Furthermore, *GLN1* encoding glutamine synthetase and *GLT1* encoding glutamate synthase have been put under a PGK constitutive promoter [63]. The strain CEN.MS5 has been transformed with the plasmid YipXR/XDH/XK to obtain the strain CPB.CR5 [26].

## Supplementary Materials List 1

- **Trace element solution:** 15 g l<sup>-1</sup> EDTA, 0.45 g l<sup>-1</sup> CaCl<sub>2</sub>·2H<sub>2</sub>O, 0.45 g l<sup>-1</sup> ZnSO<sub>4</sub>·7 H<sub>2</sub>O, 0.3 g l<sup>-1</sup> FeSO<sub>4</sub>·7 H<sub>2</sub>O, 100 mg l<sup>-1</sup> H<sub>3</sub>BO<sub>3</sub>, 1 g l<sup>-1</sup> MnCl<sub>2</sub>·2H<sub>2</sub>O, 0.3 g l<sup>-1</sup> CoCl<sub>2</sub>·6 H<sub>2</sub>O, 0.3 g l<sup>-1</sup> CuSO<sub>4</sub>·5H<sub>2</sub>O, 0.4 g l<sup>-1</sup> NaMoO<sub>4</sub>·2H<sub>2</sub>O, 0.1 g l<sup>-1</sup> KI. The pH was adjusted to 4.00 with 2M NaOH and autoclaved.
- **Vitamin solution:** 50 mg l<sup>-1</sup> d-biotin, 200 mg l<sup>-1</sup> *para*-amino benzoic acid, 1 g l<sup>-1</sup> nicotinic acid, 1 g l<sup>-1</sup> Ca-pantothenate, 1 g l<sup>-1</sup> pyridoxine. HCl; 1 g l<sup>-1</sup> thiamine HCl and 25 g l<sup>-1</sup> m-inositol. The pH was adjusted to 6.5 with 2M NaOH and the solution was stored at 4°C. Vitamin solution is added after previous sterile filtration using cellulose acetate filters (0.45 µm pore size Minisart®-Plus Satorius AG).
- **Fatty acids:** 4 g l<sup>-1</sup> Ergosterol and 168 g l<sup>-1</sup> Tween 80 dissolved in pure ethanol.

**Supplementary Materials Table 1. Shake Flask Strain Characterization**

Aerobic specific growth rates in synthetic medium with 20g l<sup>-1</sup> glucose or xylose

Strain	Description	$\mu_{\max}$			Xylose consumption rate g (gDCW) <sup>-1</sup> h <sup>-1</sup>
		Glucose (h <sup>-1</sup> ) (S.D., %)	Xylose GPI	Xylose GPII	
TMB3001	<i>XYL1, XYL2, XKS1</i>	0.30 (1.5)	0.16 (0.1)	0.005 (0.2)	0.06
CPB.CR4	<i>XYL1, XYL2, XKS, Δgdh1, GDH2</i>	0.29 (0.2)	-	-	-
CPB.CR5	<i>XYL1, XYL2, XKS1, Δgdh1, GS-GOGAT</i>	0.30 (0.1)	0.07 (0.2)	0.005 (0.1)	0.07
CMB.GS001	<i>XYL1, XYL2, XYL3</i>	0.30 (0.1)	0.13 (0.2)	0.009 (0.2)	0.08

Values are the average of two independent experiments performed in duplicate (n=4). The relative standard deviation (S.D.) is given as a percentage of the average specific growth rate. Xylose consumption rate S.D. <5% and was calculated after 48h of growth. The xylose consumption rate in the shake flask cultures was calculated as the amount of xylose (g) consumed after 48 h, divided by time (h), and biomass (g dry cell weight), considering only starting and end point data due to the relatively slow growth on xylose.

Growth phase I (GPI), Growth phase II (GPII), Data not available (-).

**Supplementary Materials Table 2: Xylose Fermenting *S. cerevisiae* Strains Expressing XR, XDH, and XK**

Strain	Description	Carbon source	Directed Evolution Approach	$\mu_{\max}$ (h <sup>-1</sup> )	Reference
TMB3001	XYL1/XYL2/XKS1	50 g/L Xylose Shake flask Complex media	Wild-type	0.007	[66-67]
TMB3399	XYL1/XYL2/XKS1	20 g/L Xylose Bioreactor	Wild-type	0.03	[68-69]
TMB3400	XYL1-XYL2 (integrant)/XKS1/SOL3/GND1/TAL/TKL	20 g/L Xylose Bioreactor	Chemical mutagenesis	0.14	[68-69]
H2490	XYL1/XYL2/XKS1	30 g/L Xylose Bioreactor	Wild-type	0.05	[70]
H2490-4	XYL1/XYL2/XKS1 Chemostat isolated	30 g/L Xylose Bioreactor	Chemostat isolation	0.15	[70]
TMB3001C1	XYL1/XYL2/XKS1 Chemostat isolated	5 g/l Xylose SF	Chemostat isolation	0.12	[71]
YSX3	XYL1/XYL2/XKS1	20 g/l Xylose Shake flask Complex media	Wild-type	0.15	[31]
YSX3-TAL1I	XYL1/XYL2/XKS1/TAL1	Minimal media	None	0.12	[72]
TMB3057	XYL1/XYL2/XK/ TAL/TKL/RKI/RPE1/ΔGRE3	50 g/L Xylose Shake flask	None	0.16	[73]
TMB3055	XYL1/XYL2/XK/ TAL/TKL/RKI/RPE1/ΔGRE3 Evolved	50 g/L Xylose Shake flask	Plate/ SF repeated cultivation	0.17	[73]
CMB.GS010	PsXYL1/PsXYL2/PsXYL3 Evolved	20 g/L Xylose Bioreactor	SF repeated cultivation	0.18	This study

**Notes:** Shake flask (SF). *XYL1* encodes xylose reductase (XR). *XYL2* encodes xylitol dehydrogenase (XDH). *XKS1* encodes xylulokinase (XK). *SOL3* encodes 6-phosphogluconolactonase. *GND1* encodes 6-phosphogluconate dehydrogenase. *TAL1* encodes Transaldolase. *TKL1* encodes transketolase. *RKI1* encodes ibose-5-phosphate ketol-isomerase. *RPE1* encodes D-ribulose-5-phosphate 3-epimerase. *GRE3* encodes aldose reductase.

**Supplementary Materials Table 3: Product Yield Coefficients for Aerobic and Anaerobic Batch Fermentations**

Strain	Carbon source — g l <sup>-1</sup> —	Condition	Y <sub>SX</sub> <sup>a</sup>	Y <sub>SG</sub> <sup>b</sup>	Y <sub>SE</sub> <sup>c</sup>	Y <sub>SC</sub> <sup>d</sup>	Y <sub>SA</sub> <sup>e</sup>	X <sub>SP</sub> <sup>f</sup>	Y <sub>SXYL</sub> <sup>g</sup>
			— Cmol —	— Cmol <sup>-1</sup> —	— Cmol —	— Cmol <sup>-1</sup> —	— Cmol —	— Cmol <sup>-1</sup> —	— Cmol —
CMB.GS010	20 Xylose	Aerobic	0.62	0.01	0.01	0.09	0.00	0.00	0.00
		Anaerobic	-	-	-	-	-	-	-
	20 Glucose	Aerobic	0.15	0.04	0.43	0.16	0.03	0.01	-
		Anaerobic	0.12	0.11	0.41	0.12	0.02	0.01	-
	10 Glucose + 10 Xylose	Aerobic	0.18	0.05	0.50	0.16	0.00	0.01	-
		Anaerobic	0.16	0.14	0.50	0.10	0.03	0.01	-
CMB.GS001	20 Xylose	Aerobic	0.56	0.02	0.01	0.06	0.00	0.00	0.01
CEN.PK113-7D	20 Glucose	Aerobic	0.18	0.07	0.53	0.14	0.01	0.01	-
		Anaerobic	0.11	0.10	0.51	0.30	0.01	0.00	-

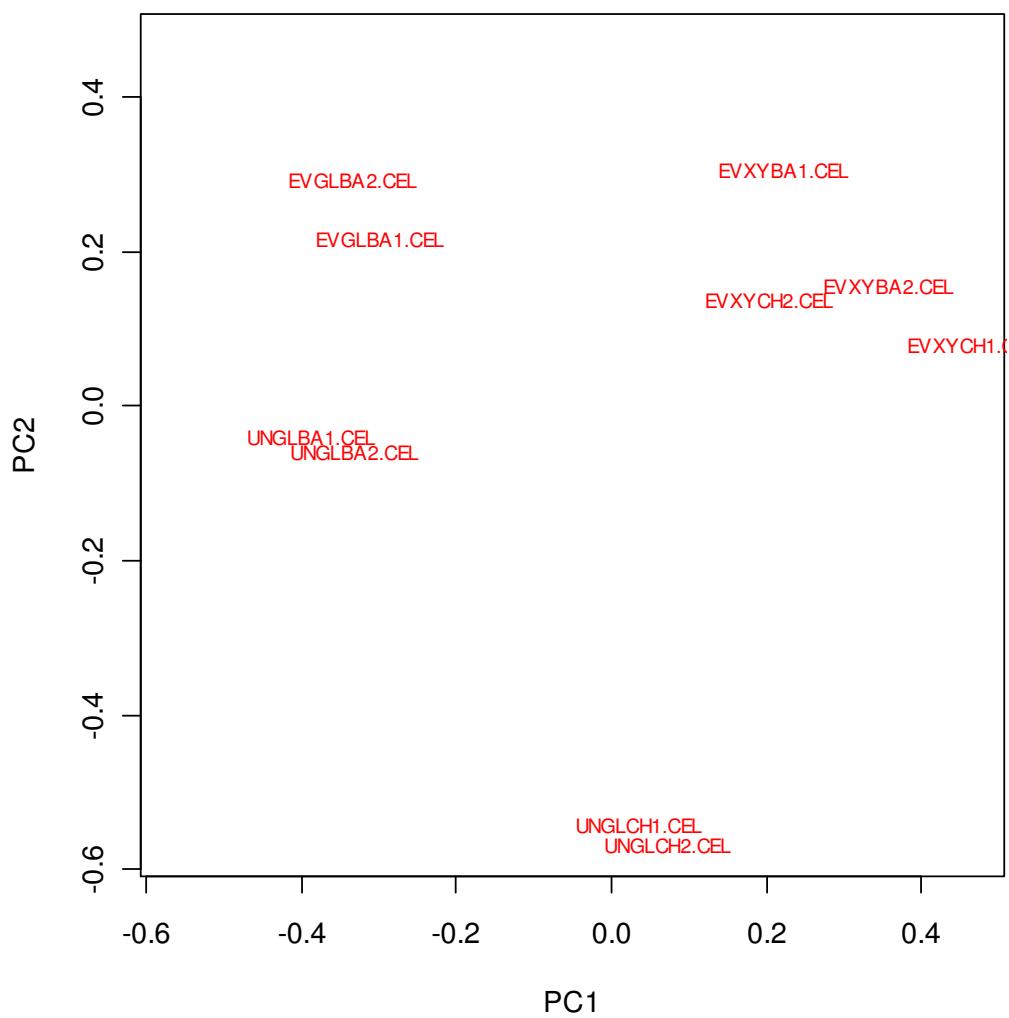
Yields reported are calculated considering only the exponential phase and the total consumed substrate.

<sup>a</sup>Biomass, <sup>b</sup>Glycerol, <sup>c</sup>Ethanol, <sup>d</sup>CO<sub>2</sub>, <sup>e</sup>Acetate, <sup>f</sup>Pyruvate and <sup>g</sup>Xylitol.

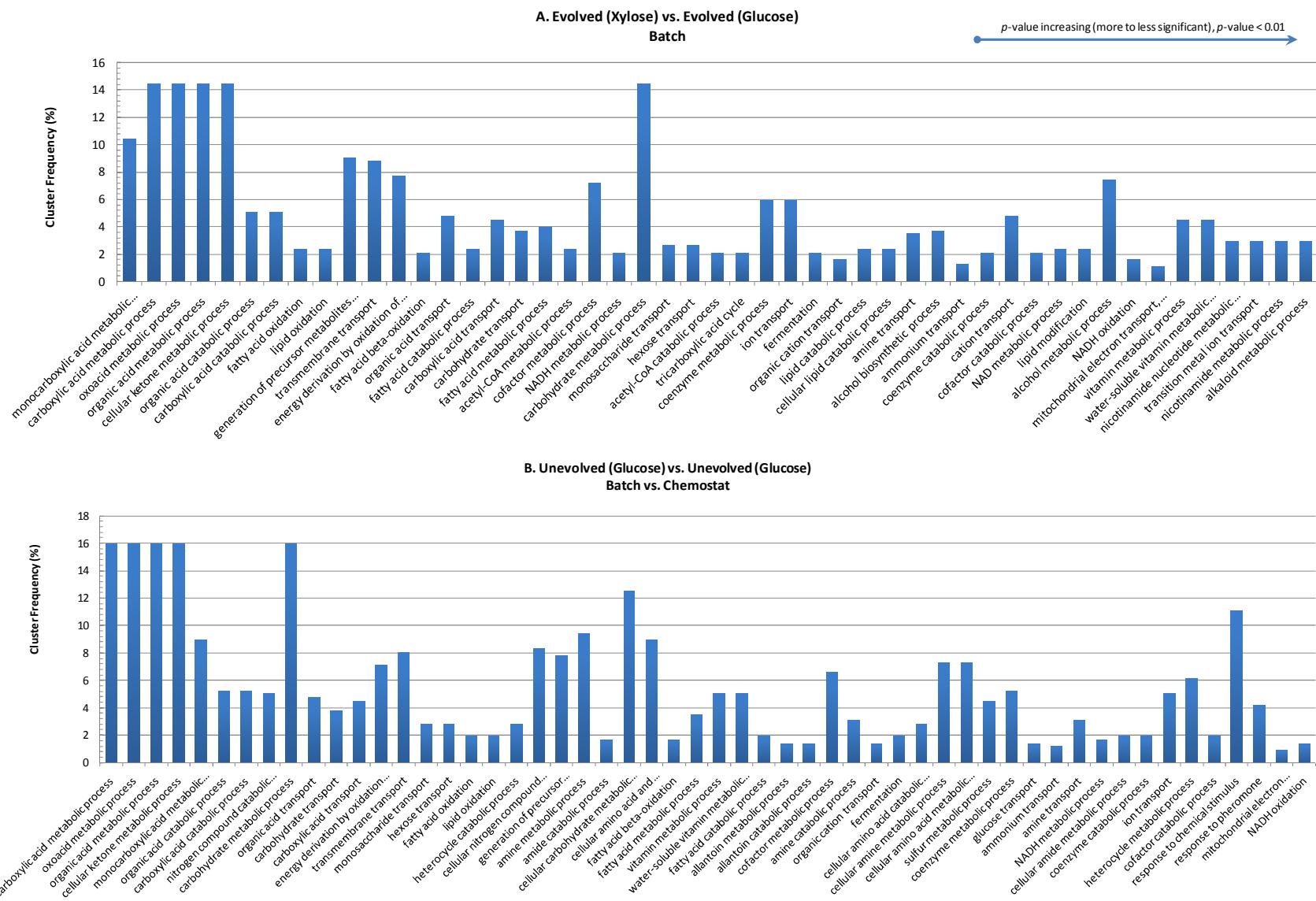
**Supplementary Materials Table 4: Summary of Transcriptome Samples Generated According to Strain, Carbon Source, and Cultivation Condition**

No.	Condition Identification	Strain	Carbon Source	Cultivation Condition	Naming Scheme (for R analysis)
1	Evolved strain xylose batch	CMB.GS.010	Xylose	Batch	EVXYBA1
2	Evolved strain xylose batch	CMB.GS.010	Xylose	Batch	EVXYBA2
3	Evolved strain xylose chemostat	CMB.GS.010	Xylose	Chemostat	EVXYCH1
4	Evolved strain xylose chemostat	CMB.GS.010	Xylose	Chemostat	EVXYCH2
5	Evolved strain glucose batch	CMB.GS.010	Glucose	Batch	EVGLBA1
6	Evolved strain glucose batch	CMB.GS.010	Glucose	Batch	EVGLBA2
7	Unevolved strain glucose batch	CMB.GS.001	Glucose	Batch	UNGLBA1
8	Unevolved strain glucose batch	CMB.GS.001	Glucose	Batch	UNGLBA2
9	Unevolved strain glucose chemostat	CMB.GS.001	Glucose	Chemostat	UNGLCH1
10	Unevolved strain glucose chemostat	CMB.GS.001	Glucose	Chemostat	UNGLCH2

### PCA Plot of Normalized Expression Data

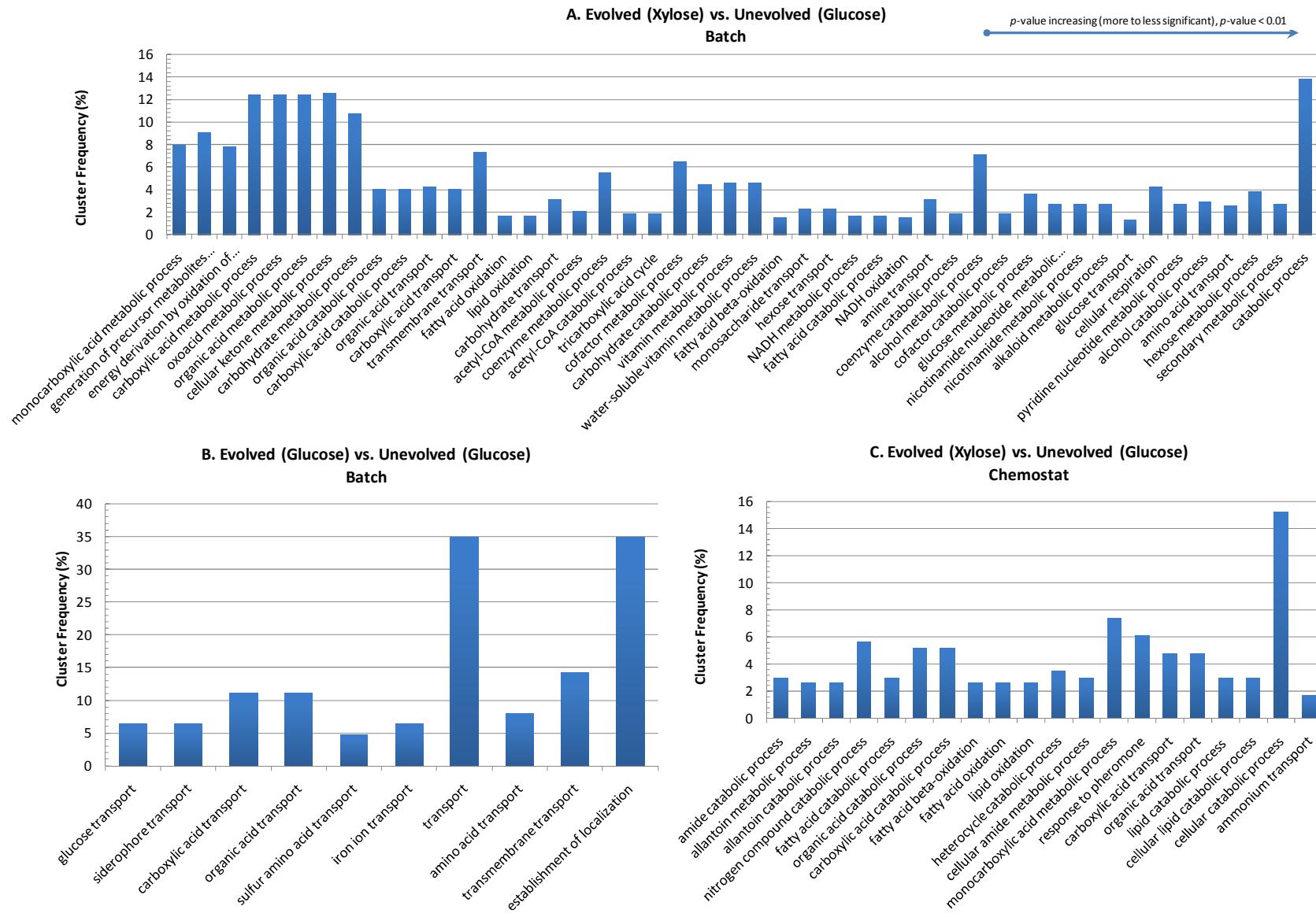


**Supplementary Materials Figure 2.** Principal component analysis (PCA) of the normalized expression data described in Supplementary Materials Table 4.

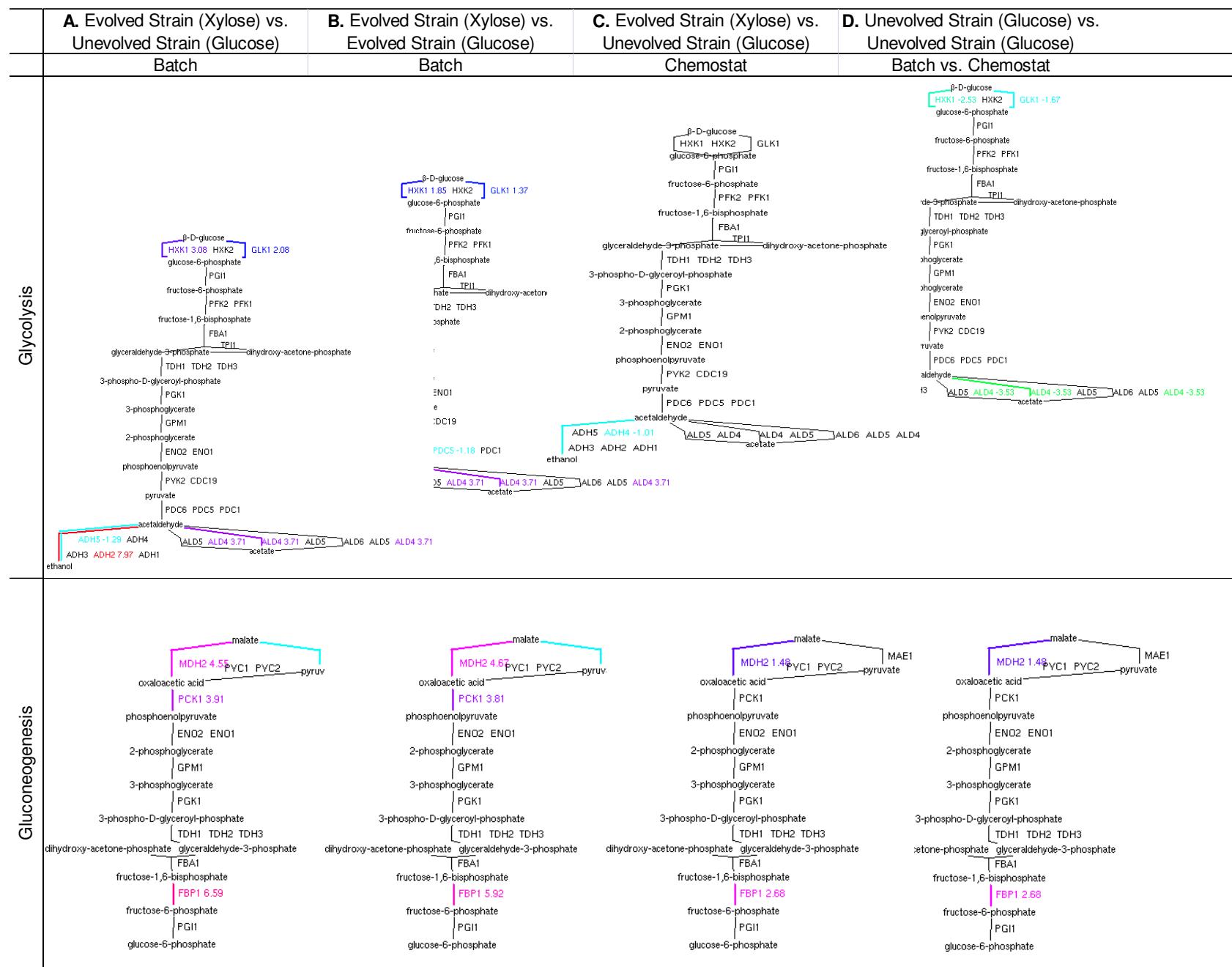


**Supplementary Materials Figure 3.** (A) Plot of the cluster frequency vs. gene ontology (GO) process terms for the significant differentially expressed genes of the CMB.GS010 (evolved) strain cultivated on batch xylose compared to batch glucose. (B) Similar plot of the cluster

frequency vs. GO process terms for the significant differentially expressed genes of the CMB.GS001 (unevolved) strain cultivated on batch glucose compared to a glucose-limited continuous culture (chemostat). The GO process terms are organized from most to least significant along the y-axis.



**Supplementary Materials Figure 4.** (A) Plot of the cluster frequency vs. gene ontology (GO) process terms for the significant differentially expressed genes of the CMB.GS010 (evolved) strain cultivated on batch xylose compared to strain CMB.GS001 (un-evolved) cultivated on batch glucose. (B) Similar plot of the cluster frequency vs. GO process terms for the significant differentially expressed genes of the CMB.GS010 strain cultivated on batch glucose compared to the strain CMB.GS001 cultivated on batch glucose. (C) Similar plot of the cluster frequency vs. GO process terms for the significant differentially expressed genes of the CMB.GS010 strain cultivated on xylose and strain CMB.GS001 cultivated on glucose in continuous cultivations (chemostats).



**Supplementary Materials Figure 5.** Two central carbon metabolic pathways are presented: (1) glycolysis, and (2) gluconeogenesis. The log-fold change of significantly differentially expressed genes ( $p_{\text{adjusted}} < 0.01$ ,  $|\log\text{-fold change}| > 1$ ) is indicated next to the gene name. These metabolic maps are provided by the Saccharomyces Genome Database Pathway Expression Viewer. The comparative conditions evaluated include: (A) CMB.GS010 cultivated on batch xylose vs. CMB.GS001 cultivated on batch glucose, (B) CMB.GS010 cultivated on batch xylose vs. CMB.GS010 cultivated on batch glucose, (C) CMB.GS010 cultivated on xylose in continuous culture (chemostat) vs. CMB.GS001 cultivated on glucose in continuous culture, (D) CMB.GS001 cultivated on batch glucose vs. CMB.GS001 cultivated on batch continuous culture. If no pathway is shown for a given comparative condition then no significant differential gene expression was detected in that pathway.

## PAPER IV

### Whole genome sequencing of *Saccharomyces cerevisiae*: from genotype to phenotype for improved metabolic engineering applications.

José Manuel Otero, Wanwipa Vongsangnak, Mohammad A. Asadollahi, Roberto Olivares-Hernández, Jérôme Maury, Laurent Farinelli, Loïc Barlocher, Magne Østerås, Michel Schalk, Anthony Clark, Jens Nielsen

*Manuscript, 2009*

**Title:** Whole genome sequencing of *Saccharomyces cerevisiae*: from genotype to phenotype for improved metabolic engineering applications

**Authors:** José Manuel Otero<sup>1,2,5</sup>, Wanwipa Vongsangnak<sup>1,2</sup>, Mohammad A. Asadollahi<sup>1,2,6</sup>, Roberto Olivares-Hernández<sup>1,2</sup>, Jérôme Maury<sup>2,7</sup>, Laurent Farinelli<sup>3</sup>, Loïc Barlocher<sup>3</sup>, Magne Østerås<sup>3</sup>, Michel Schalk<sup>4</sup>, Anthony Clark<sup>8</sup>, Jens Nielsen<sup>1,2\*</sup>

1. Department of Chemical and Biological Engineering, Chalmers University of Technology, SE-41296 Gothenburg, Sweden.
2. Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark DK-2800, Kgs. Lyngby, Denmark.
3. Fasteris SA, CH-1228-Plan-les-Ouates, Switzerland.
4. Firmenich SA, Corporate R & D Division, P.O. Box 239, CH-1211 Geneva 8, Switzerland.
5. Current Address: Vaccine & Biologics Process Development, Bioprocess Research & Development, Merck Research Labs, West Point, PA, 19846, USA.
6. Current Address: Biotechnology Group, Faculty of Advanced Sciences and Technologies, University of Isfahan, Isfahan 81746-73441, Iran.
7. Current Address: Fluxome Sciences A/S, Research & Development, DK-3660 Stenløse, Denmark.
8. Firmenich, Inc., Corporate R & D Division, P.O. Box 5880, Princeton, NJ, 08542, USA.

\*Corresponding author:

Professor Jens Nielsen

Phone: +46 (0) 31 772 8633

Fax: +46 (0) 31 772 3801

[nielsenj@chalmers.se](mailto:nielsenj@chalmers.se)

## Abstract

The needs for rapid and efficient microbial cell factory design and construction is possible through the enabling technology, metabolic engineering, which is now being facilitated with systems biology approaches. Metabolic engineering is often complimented by directed evolution, where selective pressure is applied to a partially genetically engineered strain to confer a desirable phenotype. The exact genetic modification or resulting genotype that leads to the improved phenotype is often not identified or understood to enable further metabolic engineering.

In this work we establish proof-of-concept that whole genome high-throughput sequencing can be used to identify single nucleotide polymorphisms (SNPs) between *S. cerevisiae* strains S288C and CEN.PK113-7D. S288C was the first eukaryote sequenced, serving as the reference genome for the *Saccharomyces* Genome Database, while CEN.PK113-7D is preferred laboratory strain for industrial biotechnology research. A total of 13,787 high-quality SNPs were detected, and when only considering metabolic genes (total: 782), 219 metabolism-specific SNPs are distributed across 158 metabolic genes, with 85 non-silent SNPs (e.g., encoding amino acid modifications). Amongst metabolic SNPs detected, there was pathway enrichment in the galactose uptake pathway (*GAL1*, *GAL10*) and ergosterol biosynthetic pathway (*ERG8*, *ERG9*). Physiological characterization confirmed a strong deficiency in galactose uptake and metabolism in S288C compared to CEN.PK113-7D, and similarly, ergosterol content in CEN.PK113-7D was significantly higher in both glucose and galactose supplemented cultivations compared to S288C. Furthermore, DNA microarray profiling of S288C and CEN.PK113-7D in both glucose and galactose batch cultures did not provide a clear hypothesis for major phenotypes observed, suggesting that genotype to phenotype correlations are manifested post transcriptionally or post-translationally either through protein concentration and/or function.

With an intensifying need for microbial cell factories that produce a wide array of target compounds, whole genome high-throughput sequencing for SNP detection can aid in better reducing and defining the metabolic landscape. This work demonstrates direct correlations between genotype and phenotype that provide clear and high-probability of success metabolic engineering targets.

## Introduction

Metabolic engineering is the enabling technology for identification of targeted genetic modifications such as gene deletions, over-expression, or modulation. The genetic engineering implemented in a host microbial cell factory ideally will lead to re-direction of fluxes to enhance production or robustness of a given product or organism, respectively (Bailey, 1991; Stephanopoulos & Vallino, 1991; Nielsen, 2001; Tyo, 2007; Patnaik, 2008). Metabolic engineering through systems biology has been complimented, and its application expanded in both scope and success. Systems biology is a multi-disciplinary approach to quantitative collection, analysis, and integration of whole genome scale data sets enabling construction of biologically relevant and often predictive mathematical models (Westerhoff & Palssøn, 2004; Oliver, 2006; Nielsen & Jewett, 2008). Genome sequencing of industrially relevant organisms, including *Saccharomyces cerevisiae* strain S288C, the first eukaryote genome sequence reported, provided a framework for gene annotation through functional genomics. More relevant to metabolic engineering, an annotated genome sequence was a prerequisite for genome-scale metabolic network reconstructions (Goffeau, 1996; Förster, 2003). Such reconstructions offer a biochemical model describing the formation and depletion of each metabolite that by providing mass-balance boundary conditions makes possible constraint based simulations of how the metabolic network operates at different conditions. In simpler terms, using basic stoichiometry these models can be used to predict the relationships between genes with function in the metabolic network operating in a cell. With nearly 14 years elapsing since the *S. cerevisiae* strain S288C genome sequence was made available, and more than 1,000 laboratories participating in functional genomics efforts, there are still 968 and 811 open reading frames (ORFs) classified as uncharacterized and dubious, respectively, according to the Saccharomyces Genome Database (SGD) (Cherry, 1997; Goffeau, 2000; Goffeau, 2004). Furthermore, since 2003 there have been published five major *S. cerevisiae* genome-scale metabolic network reconstructions, with the most recent models encompassing between 13-14% genome coverage (Nookaew, 2008; Herrgård, 2008). The opportunity to further extend genotype to phenotype annotation is abundant.

Industrial biotechnology is dominated by efforts to confer a desirable phenotype onto strains using different methods of directed evolution and random mutagenesis, requiring screening and selection. This approach, while providing little to no mechanistic understanding of which specific genetic perturbations lead to improved strains so they could be further exploited, has proven to be commercially successful as illustrated by the more than 1,000 fold improvement in penicillin titer by *Penicillium chrysogenum* (Nielsen, 1995). As industrial biotechnology applications expand, and the desire to custom-engineer microbial cell factories with novel architecture for native and heterologous metabolic pathways increases, the necessity on a genome-wide level to understand direct genotype to phenotype relationships has rapidly increased.

Within the same time period of approximately the last 10 years, the technologies and costs associated with whole genome sequencing have advanced and decreased, respectively. There are several excellent reviews of genome sequencing technologies, and their applications to functional genomics, strain engineering, and other investigatory biology efforts (Srivatsan, 2008; Shendure, 2008; Morozova, 2008; Khavérian, 2008; Warner, 2009). Prior work, specifically focused on characterizing genome-wide analysis of nucleotide polymorphisms in *S. cerevisiae* have utilized 25mer oligonucleotide microarrays (Affymetrix yeast tiling arrays) providing random and redundant coverage of the *S. cerevisiae* genome (Schacherer, 2007). This analysis included single nucleotide polymorphism (SNP) identification between S288C and the commonly used laboratory strain *S. cerevisiae* CEN.PK, where a total of 13,914 SNPs were identified. However, this approach is unable to identify the exact nucleotide substitution, and consequently whether the transcribed SNP results in an amino acid substitution, presumably required to confer a change in enzyme and/or protein function.

More recently a collaborative project, the Saccharomyces Genome Resequencing Project (SGRP) between the Sanger Institute and Institute of Genetics, University of Nottingham, completed the ABI sequencing of haploids of 37 *S. cerevisiae* strains to a coverage of 1-3X. Furthermore, Illumina-Solexa genome sequencing of four of the 37 *S. cerevisiae* strains, one of which included

S288C, was completed (Carter, 2008). This sequencing effort was focused on exploration of genomic variation in the context of evolution, thereby using multiple strains from different *Saccharomyces* species. It is a demonstration of a recent genome sequencing technology, referred to as Illumina-Solexa sequencing, compared to larger read methods such as Sanger or 454 sequencing. Illumina-Solexa sequencing is an ultra-high-throughput technology that performs sequencing by synthesis of random arrays of clonal DNA colonies attached to the surface of a flow cell. At each cycle of synthesis all four nucleotides, labelled with four different fluorescent dyes and blocked at the 3'-ends, are introduced in the flow cell for up to 36 such synthesis cycles. Then, segregation of each dye via independent filters enables image analysis to identify the corresponding nucleotide, and consequently, reconstruct the DNA sequence that likely generated each colony. This approach during this study generated short, 35 base pair (bp) reads (currently, the technology limitations are 76 bp or 2x76 bp paired-end reads), that must then be aligned to and assembled using a reference genome (Rougemont, 2008; Hernandez, 2008; Nikolaev, 2009).

In this work we propose that high-throughput genome sequencing of *S. cerevisiae* may serve as a commonplace tool, complimentary to transcriptomics and physiological characterization, to extract direct genotype to phenotype information. More specifically, we demonstrate that S288C, the strain utilized for the publically available *S. cerevisiae* genome sequence, exhibits atypical *S. cerevisiae* behavior related to central carbon metabolism as compared to CEN.PK113-7D, a common laboratory strain for industrial biotechnology applications (van Dijken, 2000). This behavior was characterized in well-controlled batch fermentations on glucose and galactose, complimented with transcriptome analysis. Finally, whole genome Illumina-Solexa sequencing of each strain was completed, and SNPs strictly related to metabolic genes were identified, characterized, and amino acid level analysis performed. There were clear correlations between physiology and metabolic pathway enrichment of non-silent SNPs observed, suggesting that genome-sequencing may assist in reducing the genetic target space for metabolic engineering applications. The analysis presented here serves as a foundation for comparative metabolic engineering SNP analysis, where in the future reference strains may be compared to their metabolically engineered derivatives that use directed evolution in order to answer the age-old question: what changed in our strain that makes it a preferred microbial cell factory?

## Results

### Physiological Characterization

The *S. cerevisiae* strains S288C and CEN.PK113-7D were physiologically characterized in both batch glucose and galactose supplemented fermentations. On glucose, CEN.PK113-7D exhibited a 32% higher specific growth rate than S288C, correlating with the 33% higher specific glucose consumption rate (see Table 2). The CEN.PK113-7D extracellular metabolic specific productivity rates were 32.6%, 392%, and 17.9% higher for ethanol, acetate, and glycerol production compared to S288C, respectively, while the specific oxygen consumption rates were nearly equivalent ( $1.98 \text{ O}_2 \text{ mmol g-DCW}^{-1} \text{ h}^{-1}$  for CEN.PK113-7D v.  $1.95 \text{ mmol-O}_2 \text{ g-DCW}^{-1} \text{ h}^{-1}$  for S288C). Following complete glucose fermentation, as indicated by the peak carbon dioxide evolution rate (CER), both strains underwent a diauxic shift, clearly identified by the transition of the respiratory quotient (RQ) from  $>1$  to  $<1$ , and ethanol accumulated during glucose fermentation ( $11.1 \text{ g L}^{-1}$  for CEN.PK113-7D v.  $11.3 \text{ g L}^{-1}$  for S288C) was respiro-fermented. The ethanol respiro-fermentation (ERF) phase (Figure 1) was clearly distinguishable in the CEN.PK113-7D compared to S288C, where both CER and oxygen uptake rates (OUR) linearly increased, corresponding with the increase in biomass ( $3.7$  to  $12.0 \text{ g-DCW L}^{-1}$ ). On the contrary, during the ERF phase for S288C there was a growth deficiency, clearly indicated by non-linear and significantly reduced CER and OUR rates, corresponding with a much lower increase in biomass ( $2.1$  to  $6.9 \text{ g-DCW L}^{-1}$ ). The significantly decreased ERF phase in S288C compared to CEN.PK113-7D is also evident from the total time required to exhaust the ethanol (50 v. 33 h, respectively).

A similar characterization was performed using batch galactose supplemented fermentations. CEN.PK113-7D demonstrated a slight lag-phase compared to glucose fermentation; however, sustained a galactose specific growth rate of  $0.27 \text{ h}^{-1}$  and galactose uptake rate of  $24.3 \text{ C-mmol g-DCW}^{-1} \text{ h}^{-1}$ , representing a 34% and 77% reduction, respectively, compared to glucose (see Table 2). All extracellular metabolic specific productivity rates were significantly decreased (ethanol, acetate, and glycerol were 93%, 6.8%, and 88% reduced compared to glucose, respectively), with the exception of OUR, which was 47% higher on galactose compared to glucose, leading to an effectively lower RQ of 1.5 compared to 11.9 during glucose cultivation. Furthermore, given the significantly lower RQ during the exponential phase of galactose fermentation, relatively little ethanol was produced ( $2.7 \text{ g L}^{-1}$ ), resulting in a short ERF phase (<5h) (See Figure 1). Similarly, S288C was cultivated on galactose; however, a significant deficiency in the strain's ability to metabolize this carbon source was observed. A total of 25 h post-inoculation elapsed with no increase in biomass as compared to CEN.PK113-7D where after 6h post-inoculation two cell doublings were observed. At 25 h post-inoculation a glucose bolus of  $10 \text{ g L}^{-1}$  was added to promote growth, and rapidly, glucose fermentation, a diauxic shift, and ethanol respiro-fermentation were observed (See Figure 1). Both co-consumption of galactose and ethanol, and a galactose only respiro-fermentative (GaRF) growth phase was observed. During co-consumption the specific growth rate was  $0.14 \text{ h}^{-1}$ , while on galactose only the specific growth rate was  $0.02 \text{ h}^{-1}$ . Similarly, the extracellular specific metabolite productivity rates were nearly zero when only galactose consumption was considered (See Table 2). Ethanol was consumed by 82 h post-inoculation, and in the period from 82 h to 128 h, only galactose consumption was observed, and biomass increased from  $7.9 \text{ g-DCWL L}^{-1}$  to  $20.9 \text{ g-DCW L}^{-1}$ , representing a doubling time of 35 h compared to 2.6 h for CEN.PK113-7D.

For each cultivation condition and strain, ergosterol measurements were performed and presented in Figure 2. At the same time of transcriptome sampling, which occurred during mid-exponential phase of glucose fermentation (18-20h), a total ergosterol of  $7.6 \pm 0.5 \text{ mg g-DCW}^{-1}$  and  $3.3 \pm 0.5 \text{ mg g-DCW}^{-1}$  for CEN.PK113-7D and S288C, respectively, was measured. Subsequently, the diauxic shift and ERF phase was characterized by two ergosterol samples during early and mid-ERF phase, and followed by a final (stationary) sample post-ethanol exhaustion. S288C ergosterol content was significantly higher during ethanol metabolism as compared to CEN.PK113-7D, but post-ethanol metabolism CEN.PK113-7D exhibited a significantly higher ergosterol content ( $15.9 \pm 0.7 \text{ mg g-DCW}^{-1}$  v.  $2.6 \pm 0.07 \text{ mg g-DCW}^{-1}$ ) as observed during glucose fermentation. For galactose

cultivations, ergosterol content was only measured during transcriptome sampling, which occurred at 78 h for S288C (co-consumption of ethanol and galactose observed), and 35 h for CEN.PK113-7D. The total ergosterol content on galactose was  $6.1 \pm 0.04$  mg g-DCW<sup>-1</sup> and  $4.6 \pm 0.2$  mg g-DCW<sup>-1</sup> for CEN.PK113-7D and S288C, respectively.

### **Transcriptome Characterization**

Differential gene expression between S288C and CEN.PK113-7D, cultivated on both glucose and galactose, is summarized in Table 5. The GO characterization (process, function, component) for the comparative conditions S288C v. CEN.PK113-7D cultivated on glucose and S288C v. CEN.PK113-7D cultivated on galactose, and divided into log-fold change (Ifc)  $>0$  and  $<0$ , is presented in Supplementary Materials Figure 1 and Figure 2. The metabolic pathway expression maps for each comparative condition are included in Supplementary Materials Figures 4 and 5. Lastly, all genes exhibiting statistically significant differential gene expression ( $p_{adj} < 0.01$ ) and having either a silent or non-silent SNP are included in Supplementary Materials Tables 1 and 2. The complete list of statistically significant differentially expressed genes is included as a spreadsheet in Supplementary File 1.

For the condition S288C v. CEN.PK113-7D cultivated on glucose, the top 272 differentially expressed genes, ranked according to  $p_{adj}$  value are characterized into GO process terms largely dominated by responses to stimuli and pheromone, with the dominant metabolic process categories being trehalose metabolism, steroid metabolism, and amino acid transport. Specific genes consistent with this categorization high in  $p_{adj}$  value rank and Ifc $>0$  are *GSY1* (glycogen synthase, Ifc 2.0,  $p_{adj}$  value rank 23) and for Ifc $<0$  is *HMG1* (HMG-CoA reductase, Ifc -1.7,  $p_{adj}$  value rank 14). For the condition S288C v. CEN.PK113-7D cultivated on galactose, the top 501 differentially expressed genes, ranked according to  $p_{adj}$  value are characterized into GO process terms response to stimuli and stress, carbohydrate metabolism, and transport. Specific metabolic genes noteworthy in this category, high in  $p_{adj}$  value rank amongst genes with Ifc $>0$  include *MDH2* (malate dehydrogenase, Ifc 2.8,  $p_{adj}$  value rank 8), *FBP1* (fructose-1,6-bisphosphatase, Ifc 4.2,  $p_{adj}$  value rank 15), *GAD1* (glutamate decarboxylase, Ifc 3.0,  $p_{adj}$  value rank 30), *GDH3*(NADP<sup>+</sup> dependent glutamate dehydrogenase, Ifc 3.2,  $p_{adj}$  value rank 32), *GSY1* (Ifc 1.4,  $p_{adj}$  value rank 41), and *ICL1* (isocitrate lyase, Ifc 2.7,  $p_{adj}$  value rank 54). Similarly, specific metabolic genes high in  $p_{adj}$  value rank amongst genes with Ifc $<0$  include *ARE2* (acyl-coA:sterol acetyltransferase, Ifc -2.3,  $p_{adj}$  value rank 10), and *CYB5* (cytochrome b5, Ifc -1.6,  $p_{adj}$  value rank 47).

### **Genome Sequencing, Metabolic SNP Identification**

Whole genome sequencing, including the number of reads, average coverage relative to the SGD reference genome, total number of non-ambiguous SNPs, and total number of filtered SNPs are presented in Table 3. Not surprisingly, S288C had relatively few SNPs compared to CEN.PK113-7D given that the reference genome from SGD is based on S288C v 12.0 (Cherry, 1997). Furthermore, the 13,787 filtered SNPs identified using the MAQ software is consistent with the previously estimated 13,914 SNPs for CEN.PK113-7D based upon DNA hybridization to 25mer oligonucleotide microarrays (Schacherer, 2007). Table 4 presents the results for metabolic SNP detection, where a total of 782 metabolic genes as defined by SGD were used to query for SNPs in both the S288C and CEN.PK113-7D genome sequences. A total of 36 metabolic SNPs, 3 of which are non-silent, were identified across 14 independent metabolic genes (3 non-silent SNPs distributed across 3 metabolic genes). A significantly higher number of metabolic SNPs, 939, were detected in CEN.PK113-7D and distributed across 158 unique metabolic genes, 85 of which contained a total 219 non-silent SNPs.

In an effort to characterize the non-silent metabolic SNPs identified in CEN.PK113-7D with biological significance, GO process categorization was performed and presented in Figure 3 ranked according to significance ( $p < 0.01$ ). The most significant categories include carboxylic acid, organic acid, carbohydrate metabolism, followed by nitrogen, amino acid, lipid, aromatic compound, and glycoprotein metabolism. Supplementary Materials Figure 3 presents the GO function and

component categorization, and as expected the highest significant concentration of non-silent SNPs ( $p<0.01$ ) distributed across a specific enzyme class is for transferases.

Furthermore, a graphical representation of all silent and non-silent SNPs mapped to their specific metabolic pathways is presented in Figure 4. Figure 5 highlights two metabolic pathways, galactose uptake and ergosterol synthesis, where an enrichment of non-silent and silent SNPs was observed. Specifically, *GAL1*, *GAL10*, *ERG8*, and *ERG9* contained non-silent SNPs, while *GAL7*, *ERG20* and *HMG1* contained silent SNPs. The specific SNPs are identified as well the resulting amino acid substitutions.

In addition to identifying SNP enriched metabolic pathways in CEN.PK113-7D, an analysis intended to determine the prevalence of the SNP across the top 10 homologous sequences resulting from a multi-alignment Pfam query was performed. To better quantify those results, the parameters *CEN.PK Match Frequency*, *Dominant AA Frequency*, *S288C Match Frequency*, and *Conservation Distance* were defined and calculated (see Supplementary Materials Figure 6). The *Conservation Distance*, bound between -1 and 1, is a measure of whether the SNP identified in CEN.PK113-7D is more prevalent amongst homologous Pfam sequences (maximum Conservation Distance = -1), or if S288C (reference SGD sequence) is more prevalent (maximum Conservation Distance = +1). Supplementary Materials Figure 7 presents the Conservation Distance across non-silent SNPs identified, with the average value of  $0.03 \pm 0.40$  ( $n = 219$ ), indicating that there is virtually no bias between S288C or CEN.PK113-7D as compared to their homologues. Extending this approach further, each amino acid polymorphism was characterized across a multi-alignment Pfam homologue search, and categorized according to standard amino acid properties (see Supplementary Materials Figure 6). For example, Figure 6 presents SNPs identified in *ERG8* at nucleotide positions 75 and 192. The resulting amino acid partially encoded by position 192 was 75% polar, 25% non-polar, 25% hydrophobic, and 75% hydrophilic looking across the top ten Pfam homologous sequences. Lastly, and of most relevance to understanding the amino acid functional changes resulting from a SNP, the same categorization is presented for the S288C v. CEN.PK113-7D sequence. For example, the SNP at position 192 of *ERG8* resulted in changing the encoded amino acid from non-polar (S288C) to polar (CEN.PK113-7D), and from hydrophobic (S288C) to hydrophilic (CEN.PK113-7D). This approach is extended to all the *ERG8* non-silent SNPs as an example of extending nucleotide level changes to amino acid functional changes (see Supplementary Materials Figure 8 for additional *ERG8* non-silent SNPs). Furthermore, Supplementary Materials Figure 9 highlights functional changes for all metabolic non-silent SNPs identified.

## Discussion

The physiological characterization clearly suggests that S288C has a deficiency in metabolism of respiro-fermentative carbon sources, such as ethanol and galactose, when compared to CEN.PK113-7D. Inspection of the significantly differentially expressed genes between strains cultivated on glucose or galactose did not reveal an obvious gene cluster that would explain this significant physiological difference. This is supported both by the GO characterization and pathway expression mapping.

In an effort to further investigate if larger regulatory mechanisms could be identified the list of genes exhibiting significant differential expression were submitted to the Yeast Search for Transcriptional Regulators And Consensus Tracking (YEASTRACT) curated repository of associations between transcription factors and target genes in *S. cerevisiae* (Teixeria, 2006; Monteiro, 2008). The transcription factor, Tec1p, was identified as directly regulating 21.1% of the total submitted gene list (See Table 5, 272 genes, S288C glucose v. CEN.PK113-7D glucose), and was 1.7-fold higher expressed in CEN.PK113-7D compared to S288C ( $p_{adj}$  value =  $7.2 \times 10^{-3}$ ). Tec1p was the only identified transcription factor to be significantly differentially expressed, and strongly regulates FLO11, a flocculin gene required for invasive growth, and pseudohyphal formation (Douglas, 2007). The transcription factors regulating the highest percentage of the differentially expressed genes, yet not being differentially expressed themselves, were Sok2p and Ste12p, with 32.5% and 21.5%, respectively, of submitted genes being directly regulated. Sok2p and Ste12p are transcription factors negatively regulating pseudohyphal differentiation (Cherry, 1997). A similar analysis was performed for galactose; however, similar results were obtained, with Sok2p and Ste12p directly regulating 23.1% and 17.4%, respectively, of the 501 differentially expressed genes (See Table 5). The transcription factors differentially expressed themselves were Msa1p and Msa2p, putative G1-specific cell cycle transcription activators, and Usv1p, a putative zinc finger transcription factor regulating growth on non-fermentable carbon sources. *USV1* expression was 2.2-fold higher in CEN.PK113-7D compared to S288C ( $p_{adj}$  value =  $3.6 \times 10^{-3}$ ). Although relatively little is known about Usv1p, it has been shown to be induced post-diauxic shift, consistent with the deficiency in post-diauxic shift metabolism observed in S288C (McCammon, 2003). With the exception of Usv1p, all transcription factors identified are more closely related to the significant difference in growth rates between strains rather than their respiro-fermentative metabolism. Metabolic SNPs identified and subsequent analysis did not identify clear correlations or pathway enrichment that could explain the lack of respiro-fermentative metabolism in S288C. Metabolic genes containing non-silent SNPs in CEN.PK113-7D, significantly differentially expressed on galactose, and related to oxidative metabolism included *ACS1* (Acetyl-CoA synthetase), *GAD1* (Glutamate decarboxylase), *YAT2* (Carnitine acetyltransferase), and *CCP1* (Mitochondrial cytochrome-c peroxidase) which were 7.8-fold, 3.2-fold, 4.5-fold, and 2.1-fold higher in CEN.PK113-7D, respectively (See Supplementary Materials Table 2).

There were two central carbon metabolic pathways enriched with non-silent SNPs that also correlated with significant differences in phenotype. *S. cerevisiae* CEN.PK113-7D exhibited significantly higher ergosterol content during growth on glucose, and to a lesser extent, galactose. This is consistent with previous work where CEN.PK2-1C had very high ergosterol/erg-ester (20.0 mg/g CDW) and triacylglycerols content (15.2 mg/g CDW) compared to 9 other *S. cerevisiae* strains, including FY169 (ergosterol/erg-ester content: 8.5 mg/g CDW; triacylglycerols content: 2.4 mg/g CDW) which is isogenic to S288C (Daum, 1999; Winston, 1995). The ergosterol biosynthetic pathway had significant non-silent SNPs identified in *ERG8* and *ERG9*, and silent SNPs identified in *ERG20* and *HMG1*. Both *ERG8* and *ERG9* were not significantly differentially expressed, either in glucose or galactose, suggestive again that phenotypic observations, consistent with genome sequence variations, are not necessarily directly manifested at the transcriptome level. Both *ERG8* (encodes phosphomevalonate kinase) and *ERG9* (encodes squalene synthetase) are essential cytosolic enzymes in the biosynthetic pathway of isoprenoids and sterols ( $\Delta$ *erg8* and  $\Delta$ *erg9*, were found to both be auxtrophic for ergosterol in the systematic deletion library), including ergosterol, from mevalonate (Tsay, 1991; Jennings, 1991; Cherry, 1997). The ergosterol biosynthetic pathway is highly

regulated through feedback inhibition mechanisms and by several rate-controlling steps, including that catalyzed by HMG-CoA reductase, encoded by *HMG1* (Basson, 1988; Maury, 2005). Under both glucose and galactose, *HMG1* expression was significantly down-regulated in S288C compared to CEN.PK113-7D by 3.2-fold ( $p_{adj}$  value =  $3.3 \times 10^{-4}$ ) and 1.8-fold ( $p_{adj}$  value =  $8.6 \times 10^{-3}$ ), respectively, correlating with the significantly less ergosterol content in S288C cultivated on glucose and to a lesser extent, on galactose. Furthermore, *ERG9* has been previously identified as also having a regulatory role (Grabowska, 1998), consistent with the hypothesis that a non-silent SNP resulting in altered protein function could affect ergosterol synthesis. *ERG8* on the other hand has not been explicitly shown to have a regulatory function, yet, when the specific activity of  $0.06 \mu\text{mol min}^{-1} \text{mg}^{-1}$  is compared to other ergosterol synthetic enzymes such as *ERG13* (2.1 in *S. cerevisiae*), *ERG12* (0.77 in *S. cerevisiae*), *ERG20* (5.22 in *S. cerevisiae*), and especially the known regulator *HMG1/HMG2* (0.0035 in *S. cerevisiae*) it is suggestive that *ERG8* is likely a rate limiting step (Middleton, 1975; Gray, 1972; Tchen 1958; Porter, 1985; Eberhardt, 1975; Rilling, 1985; Basson, 1986; Durr, 1960; Bloch, 1959). There were a large number of non-silent SNPs that encoded significant changes in amino acid classes, further suggestive that *ERG8* is a strong metabolic engineering target for understanding the significantly higher ergosterol content in CEN.PK113-7D. Lastly, the observation that neither *ERG8* nor *ERG9* were differentially expressed under glucose or galactose, suggests their potential affect on phenotype is likely post-translational.

Similar to ergosterol biosynthesis, the galactose uptake pathway phenotype in S288C was vastly down-regulated compared to CEN.PK113-7D, correlating with the non-silent SNP enrichment in *GAL1* and *GAL10*, and silent SNPs in *GAL7*. Neither *GAL1* (encodes galactokinase) nor *GAL10* (encodes UDP-glucose-4-epimerase) were significantly differentially expressed during growth on galactose; however, on glucose *GAL1* was significantly up-regulated ( $p_{adj}$  value =  $9.7 \times 10^{-4}$ ) 2.9-fold in CEN.PK113-7D. Both  $\Delta gal1$  and  $\Delta gal10$  mutants are unable to grow on galactose as sole carbon sources (Bhat, 1990; Bhat 1992; Douglas, 1964). The significant number of non-silent SNPs in both essential galactose genes suggests obvious targets for explanation of why S288C is incapable of galactose respiro-fermentative metabolism. Furthermore, it should be noted that while S288C has been described as  $\Delta gal2$  (See Table 1), no SNPs were detected between CEN.PK113-7D and S288C, and CEN.PK113-7D was able to readily metabolize galactose meaning a functional *GAL2* (encodes galactose permease, required for galactose utilization) is present in both S288C and CEN.PK113-7D.

A further metabolic engineering benefit of whole genome sequencing was the detection of a non-silent SNP resulting in a stop codon of *PAD1* (encodes phenylacrylic acid decarboxylase). Pad1p is essential for decarboxylation of aromatic carboxylic acids conferring resistance to cinnamic acid, and a non-silent SNP was detected at nucleotide position 294 (T to G), resulting in a stop codon (TAT → TAG) (Clausen M, 1994). Although Pad1 relevant phenotypes were not explored, the transcriptome response on glucose revealed significant differential expression of *PAD1* ( $p_{adj}$  value =  $1.5 \times 10^{-3}$ ), with 3.1-fold higher expression in S288C compared to CEN.PK113-7D. This is consistent with the stop codon detected in CEN.PK113-7D at position 294, noting that the total ORF genomic DNA sequence is 729 nucleotides, and therefore unlikely to be transcribed and detected.

In summary and perhaps not surprisingly, transcriptome analysis did not provide a clear hypothesis for major phenotypes observed, suggesting that genotype to phenotype correlations are manifested post-transcriptionally or post-translationally either through protein concentration and/or function. Clearly, future work must validate these correlations through genetic engineering of identified SNPs in either S288C or CEN.PK113-7D to see if desired phenotypes, such as increased galactose uptake or ergosterol synthesis in S288C, are observed. Future work must also expand on the metabolic SNP analysis presented to include all 13,787 SNPs, realizing phenotypic observations may not necessarily be linked directly to metabolic SNPs, but rather SNPs affecting larger regulatory mechanisms and networks, such as those governed by transcription factors. Certainly, as *S. cerevisiae* continues to be exploited, particularly for metabolic engineering applications, the integration of physiological characterization, transcriptome analysis, and metabolic SNP detection with high-throughput whole genome sequencing provides direct correlations between observed phenotypes and genotypes and offers high probability of success metabolic targets.

## **Acknowledgements**

This work was in part supported by a Merck Doctoral Fellowship to J.M.O.

## Materials and Methods

### Strain Description

The strains used in this study are presented and described in Table 1.

### Medium Formulation

A chemically defined minimal medium of composition  $5.0\text{ g L}^{-1}$   $(\text{NH}_4)_2\text{SO}_4$ ,  $3.0\text{ g L}^{-1}$   $\text{KH}_2\text{PO}_4$ ,  $0.5\text{ g L}^{-1}$   $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ ,  $1.0\text{ mL L}^{-1}$  trace metal solution,  $300\text{ mg L}^{-1}$  uracil,  $0.05\text{ g L}^{-1}$  antifoam 204 (Sigma-Aldrich A-8311), and  $1.0\text{ mL L}^{-1}$  vitamin solution was used for all shake flask and 2L well-controlled fermentations (Verudyn, 1992). The trace element solution included  $15\text{ g L}^{-1}$  EDTA,  $0.45\text{ g L}^{-1}$   $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ ,  $0.45\text{ g L}^{-1}$   $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ ,  $0.3\text{ g L}^{-1}$   $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ ,  $100\text{ mg L}^{-1}$   $\text{H}_3\text{BO}_4$ ,  $1\text{ g L}^{-1}$   $\text{MnCl}_2 \cdot 2\text{H}_2\text{O}$ ,  $0.3\text{ g L}^{-1}$   $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ ,  $0.3\text{ g L}^{-1}$   $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ ,  $0.4\text{ g L}^{-1}$   $\text{NaMoO}_4 \cdot 2\text{H}_2\text{O}$ . The pH of the trace metal solution was adjusted to 4.0 with 2M NaOH and heat sterilized. The vitamin solution included  $50\text{ mg L}^{-1}$  d-biotin,  $200\text{ mg L}^{-1}$  para-amino benzoic acid,  $1\text{ g L}^{-1}$  nicotinic acid,  $1\text{ g L}^{-1}$  Ca-pantothenate,  $1\text{ g L}^{-1}$  pyridoxine HCl,  $1\text{ g L}^{-1}$  thiamine HCl, and  $25\text{ mg L}^{-1}$  m-inositol. The pH of the vitamin solution was adjusted to 6.5 with 2M NaOH, sterile-filtered and the solution was stored at  $4^\circ\text{C}$ . The final formulated medium, excluding glucose and vitamin solution supplementation, is adjusted to pH 5.0 with 2M NaOH and heat sterilized. For carbon-limited cultivations the sterilized medium is supplemented with  $40\text{ g L}^{-1}$  glucose or  $40\text{ g L}^{-1}$  galactose, heat sterilized separately, and  $1.0\text{ mL L}^{-1}$  vitamin solution is added by sterile filtration (0.20  $\mu\text{m}$  pore size Ministart®-Plus Sartorius AG, Goettingen, Germany).

### Shake Flask Cultivations and Stirred Tank Fermentations

Shake flask cultivations were completed in 500 mL Erlenmeyer flasks with two diametrically opposed baffles and two side-necks with septums for sampling by syringe. Flasks were heat sterilized with 100 mL of medium, inoculated with a single colony, and incubated at  $30^\circ\text{C}$  with orbital shaking at 150 RPM. Stirred tank fermentations were completed in well-controlled, aerobic, 2.2L Braun Biotech Biostat B fermentation systems with a working volume of 2L (Sartorius AG, Goettingen, Germany). The temperature was controlled at  $30^\circ\text{C}$ . The fermenters were outfitted with two disk-turbine impellers rotating at 600 RPM. Dissolved oxygen was monitored with an autoclavable polarographic oxygen electrode (Mettler-Toledo, Columbus, OH). During aerobic cultivation the air sparging flow rate was 2 vvm. The pH was kept constant at 5.0 by automatic addition of 2M KOH. Off-gas passed through a condenser to minimize the evaporation from the fermenter. The fermenters were inoculated from shake flask precultures to an initial  $\text{OD}_{600}$  0.01.

### Fermentation Analysis

**Off-gas Analysis:** The effluent fermentation gas was measured every 30 seconds for determination of  $\text{O}_2(\text{g})$  and  $\text{CO}_2(\text{g})$  concentrations by the off-gas analyzer Brüel and Kjær 1308 (Brüel & Kjær, Nærum, Denmark).

**Biomass Determination:** The optical density (OD) was determined at 600 nm using a Shimadzu UV mini 1240 spectrophotometer (Shidmazu Europe GmbH, Duisberg, Germany). Duplicate samples were diluted with deionized water to obtain  $\text{OD}_{600}$  measurements in the linear range of 0-0.4  $\text{OD}_{600}$ . Samples were always maintained at  $4^\circ\text{C}$  post-sampling until  $\text{OD}_{600}$  and dry cell weight (DCW) measurements were performed. DCW measurements were determined through the exponential phase, until stationary phase was confirmed according to  $\text{OD}_{600}$  and off-gas analysis. Nitrocellulose filters (0.45  $\mu\text{m}$  Sartorius AG, Goettingen, Germany) were used. The filters were pre-dried in a microwave oven at 150W for 10 min., and cooled in a dessicator for 10 min. 5.0 mL of fermentation broth were filtered, followed by 10 mL DI water. Filters were then dried in a microwave oven for 20 min. at 150W, cooled for 15 min. in a desiccator, and the mass was determined.

**Metabolite Concentration Determination:** All fermentation samples were immediately filtered using a 0.45 µm syringe-filter (Sartorius AG, Goettingen, Germany) and stored at -20°C until further analysis. Glucose, ethanol, glycerol, acetate, succinate, pyruvate, fumarate, citrate, oxalate, and malate were determined by HPLC analysis using an Aminex HPX-87H ion-exclusion column (Bio-Rad Laboratories, Hercules, CA). The column was maintained at 65°C and elution performed using 5 mM H<sub>2</sub>SO<sub>4</sub> as the mobile phase at a flow rate of 0.6 mL min.<sup>-1</sup>. Glucose, ethanol, glycerol, acetate, succinate, citrate, fumarate, malate, oxalate were detected on a Waters 410 differential refractometer detector (Shodex, Kawasaki, Japan), and acetate and pyruvate were detected on a Waters 468 absorbance detector set at 210 nm. Ergosterol measurements were made according to previous published methods (Asadollahi, 2009).

### **Genome Sequencing**

**DNA Isolation:** A standard 500 mL shake flask, supplemented with 10 g L<sup>-1</sup> glucose and inoculated with a single colony of *S. cerevisiae* S288C or CEN.PK113-7D, was permitted to grow for 24-48h at 30°C until visual inspection confirmed a high optical density. A total of 5 mL culture was aliquoted into 15 mL sterile tubes (one per extraction), centrifuged (4000 RCF) for 5 min, washed with 2 mL deionized water, and pelleted. Cell pellets were resuspended in 0.5 mL lysis buffer. Lysis buffer consisted of 0.1M Tris pH 8.0, 50 mM EDTA, and 1% SDS final concentration. The lysis buffer suspension was transferred to a 1.5 mL FastPrep screw cap tube, to which 200 µL acid-washed glass beads (250-500 µm) and 25 µL 5M NaCl was added. A FastPrep™ FP120 (QBiogene, Irvine, CA) was used for cell lysis, with two cycles of 20s disruption and 1 min on ice. The resulting cell suspension was centrifuged (13,000 RCF) for 10 min., and the resulting clear liquid, approximately 350 µL, avoiding white cell debris and beads, was aspirated with a pipette and transferred to 1.5 mL microcentrifuge tubes. 400 µL chloroform (TE-saturated) was added to each tube, mixed, and a chloroform extraction performed. 1 mL 99% ethanol was added to the resulting suspension, mixed, centrifuged (13,000 RCF) for 6 min., ethanol decanted, and then resuspended in 70% ethanol. The resulting suspension was centrifuged (13,000 RCF) for 6 min., ethanol decanted, and pellet permitted to dry for 25-60 min. The pellet was then resuspended in 50 µL 2 mM Tris, incubated for 10 min. at 37°C, and stored at -20°C.

**Illumina Genome Sequencing and SNP Analysis:** Isolated DNA from *S. cerevisiae* S288C and CEN.PK113-7D was shipped to Fasteris SA (Geneva, Switzerland). Fasteris SA, utilizing the Solexa technology according to the manufacturer's recommendations (Illumina). The whole-genome sequencing was performed mid-2007 on a Genome Analyzer "classic" instrument with sequencing kits version 1 and base calling on the Solexa Pipeline (v0.2.2.5). Included in the sequencing was confirmation of SNP detection using two independent approaches: the Mapping and Assembly with Quality (MAQ) software package (<http://maq.sourceforge.net>) and the Edena software package (<http://www.genomic.ch/edena.php>). The MAQ software, designed for building map assemblies from short reads generated by the Illumina-Solexa 1G Genetic Analyzer, can specifically fast align short base pair reads (35 bp) to the reference genome, in this case, the *S. cerevisiae* S288C v12.0 sequence available at the Saccharomyces Genome Database (SGD). The *maq assemble -m* command was used to call the consensus sequences from read mapping, with the value *-m* set to 1, which specifies the maximum numbers of mismatches allowed for a read to be used in consensus calling. The *maq.pl SNPfilter* command was used for high-quality SNP identification, where specifically, SNPs that are covered by few reads (specified by *-d*), by too many reads (specified by *-D*), near to a potential indel (specified by *-w*), falling in a possible repetitive region (specified by *-Q*), or having low-quality neighboring bases (specified by *-n*) are ruled out. The threshold values applied to SNP detection for the CEN.PK113-7D sequence relative to the reference sequence were *d*>5, *D*<255, *w*<1.5, *Q*>50, and *n*>40. These threshold parameter values were tested such that the amount of coverage and proportion of genome with aligned sequences was maximized, and a graphical representation of SNPs was produced to confirm results. The Edena approach for *de novo* assembly

has been previously described (Hernandez, 2008). The Edena assembly results for both whole genome sequencing and specific to SNP detection for metabolic genes are presented in Supplementary Materials Table 3. The Edena produced assembly and SNP detection results was found to have poor coverage compared to MAQ software results; however, serves as an independent verification of the SNPs detected. For purposes of the subsequent SNP analysis only the MAQ software results are used. The FASTA files of each genome sequence are available upon request.

All metabolic genes containing SNPs, both silent and non-silent, were manipulated within the software BioEdit v7.08 (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). Specifically, the ORF genomics nucleotide sequence available on SGD ([www.yeastgenome.org](http://www.yeastgenome.org)) were imported into BioEdit, and the sequences modified with the identified SNP, creating a new CEN.PK113-7D sequence for that ORF relative to the original S288C strain. Both the S288C and CEN.PK113-7D nucleotide sequences were then translated in fix full frames, and amino acid polymorphisms were identified, leading to the categorization of each SNP as either being silent or non-silent. Subsequent physiological characterization of the gene and all relevant amino acid information from UnitProt were managed in a spreadsheet using Microsoft Excel. Multi-sequence Pfam alignments were performed using a custom BioPerl script and the UNIX operating environment. Calculations and characterization described in Supplementary Materials Figure 6, related to amino acids, were then performed using Microsoft Excel.

### ***Transcriptomics***

***RNA Sampling and Isolation:*** Samples for RNA isolation from the late-exponential phase of glucose-limited and galactose-limited batch cultivations were taken by rapidly sampling 25 mL of culture into a 50 mL sterile Falcon tube with 40 mL of crushed ice in order to decrease the sample temperature to below 2°C in less than 10 seconds. Cells were immediately centrifuged (4000 RCF at 0°C for 2.5 min.), the supernatant discarded, and the pellet frozen in liquid nitrogen and it was stored at -80°C until total RNA extraction. Total RNA was extracted using the FastRNA Pro RED kit (QBiogene, Carlsbad, USA) according to manufacturer's instructions after partially thawing the samples on ice. RNA sample integrity and quality was determined prior to hybridization with an Agilent 2100 Bioanalyzer and RNA 6000 Nano LabChip kit according to the manufacturer's instruction (Agilent, Santa Clara, CA).

***Probe Preparation and Hybridization to DNA Microarrays:*** Messenger RNA (mRNA) extraction, cDNA synthesis, labeling, and array hybridization to Affymetrix Yeast Genome Y2.0 arrays were performed according to the manufacturer's recommendations (Affymetrix GeneChip® Expression Analysis Technical Manual, 2005-2006 Rev. 2.0). Washing and staining of arrays were performed using the GeneChip Fluidics Station 450 and scanning with the Affymetrix GeneArray Scanner (Affymetrix, Santa Clara, CA).

***Microarray Gene Transcription Analysis:*** Affymetrix Microarray Suite v5.0 was used to generate CEL files of the scanned DNA microarrays. These CEL files were then processed using the statistical language and environment R v5.3 (R Development Core Team, 2007, [www.r-project.org](http://www.r-project.org)), supplemented with Bioconductor v2.3 (Bioconductor Development Core Team, 2008, [www.bioconductor.org](http://www.bioconductor.org)) packages Biobase, affy, gcrma, and limma (Smyth, 2005). The probe intensities were normalized for background using the robust multiarray average (RMA) method only using perfect match (PM) probes after the raw image file of the DNA microarray was visually inspected for acceptable quality. Normalization was performed using the qspline method and gene expression values were calculated from PM probes with the median polish summary. Statistical analysis was applied to determine differentially expressed genes using the limma statistical package. Moderated t-tests between the sets of experiments were used for pair-wise comparisons. Empirical Bayesian statistics were used to moderate the standard errors within each gene and Benjamini-Hochberg's method was used to adjust for multi-testing. A cut-off value of adjusted  $p < 0.01$  was used

for statistical significance (Smyth, 2000). Statistically significant differential gene expression lists were then submitted to the GO Term Finder (version 0.83) of the Saccharomyces Genome Database (SGD) for GO process, function, and component statistically significant identification ( $p<0.01$ ). Furthermore, the same differential gene expression lists were submitted to the Expression Viewer (Pathway Tools version 12.0 generated by SRI International on SGD) for metabolic pathway mapping and identification (Cherry, 1997).

## References

1. Asadollahi MA, Maury J, Møller K, Nielsen KF, Schalk M, Clark A, Nielsen J (2008) Production of plant sesquiterpenes in *Saccharomyces cerevisiae*: effect of *ERG9* repression on sesquiterpene biosynthesis. *Biotechnol Bioeng* 99(3): 666-677.
2. Bailey JE (1991) Toward a science of metabolic engineering. *Science* 252(5013): 1668-1675.
3. Basson ME, Thorsness M, Finer-Moore J, Stroud RM, Rine J (1988) Structural and functional conservation between yeast and human 3-hydroxy-3-methylglutaryl coenzyme A reductases, the rate-limiting enzyme of sterol biosynthesis. *Mol Cell Biol* 8(9): 3797-3808.
4. Basson ME, Thorsness M, Rine J (1986) *Saccharomyces cerevisiae* contains two functional genes encoding 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *Proc Natl Acad Sci USA* 83(15): 5563-5567.
5. Bhat PJ, Hopper JE (1992) Overproduction of the GAL1 or GAL3 protein causes galactose-independent activation of the GAL4 protein: evidence for a new model of induction for the yeast GAL/MEL region. *Mol Cell Biol* 12(6): 2701-2707.
6. Bhat PJ, Oh D, Hopper JE (1990) Analysis of the GAL3 signal transduction pathway activating GAL4 protein-dependent transcription in *Saccharomyces cerevisiae*. *Genetics* 125(2): 281-291.
7. Bloch K, Chaykin S, Phillips AH, de Waard A (1959) Mevalonic acid pyrophosphate and isopentenylpyrophosphate. *J Biol Chem* 234: 2595-2604.
8. Carter D (2008) Saccharomyces Genome Resequencing Project: User Manual. Available at: <http://www.sanger.ac.uk/Teams/Team118//sgrp/>. Last accessed September 4, 2009.
9. Cherry JM, Ball C, Weng S, Juvik G, Schmidt R, Adler C, Dunn B, Dwight S, Riles L, Mortimer RK, Botstein D (1997) Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* 387: 67-73.
10. Clausen M, Lamb CJ, Megnet R, Doerner PW (1994) PAD1 encodes phenylacrylic acid decarboxylase which confers resistance to cinnamic acid in *Saccharomyces cerevisiae*. *Gene* 142(1): 107-112.
11. Daum G, Tuller G, Nemec T, Hrastnik C, Balliano G, Cattel L, Milla P, Rocco F, Conzelmann A, Vionnet C, Kelly DE, Kelly S, Schweizer E, Schüller H-J, Hojad U, Greiner E, Finger K (1999) Systematic analysis of yeast strains with possible defects in lipid metabolism. *Yeast* 15(7): 601-614.
12. Douglas HC, Hawthorne DC (1964) Enzymatic expression and genetic linkage of genes controlling galactose utilization in *Saccharomyces*. *Genetics* 49: 837-844.
13. Douglas LM, Li L, Yang Y, Dranginis AM (2007) Expression and characterization of the flocculin Flo11/Muc1, a *Saccharomyces cerevisiae* mannoprotein with homotypic properties of adhesion. *Eukaryot Cell* 6(12): 2214-2221.
14. Durr IF, Rudney H (1960) The reduction of beta-hydroxy-beta-methyl-glutaryl coenzyme A to mevalonic acid. *J Biol Chem* 235: 2572-2578.
15. Eberhardt NL, Rilling HC (1975) Prenyltransferase from *Saccharomyces cerevisiae*. Purification to homogeneity and molecular properties. *J Biol Chem* 250(3): 863-866.
16. Förster J, Famili I, Fu P, Palssøn, Nielsen J (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* 13(2): 244-53.
17. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 genes. *Science* 274(5287): 546, 563-7.
18. Grabowska D, Karst F, Szkopinska A (1998) Effect of squalene synthase gene disruption on synthesis of polyprenols in *Saccharomyces cerevisiae*. *FEBS Lett* 434(3): 406-408.
19. Gray JC, Kekwick RG (1972) The inhibition of plant mevalonate kinase preparations by prenyl pyrophosphates. *Biochim Biophys Acta* 279(2): 290-296.
20. Hernandez D, François P, Farinelli L, Osterås M, Schrenzel J (2008) *De novo* bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Res* 18(5): 802-809.
21. Herrgård MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Blüthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novère N, Li P, Liebermeister W, Mo ML, Oliveira AP,

- Petranovic D, Pettifer S, Simeonidis E, Smallbone K, Spasić I, Weichert D, Brent R, Broomhead DS, Westerhoff HV, Kirdar B, Penttilä M, Klipp E, Palsson BØ, Sauer U, Oliver SG, Mendes P, Nielsen J, Kell DB (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol* 26(10):115-1160.
22. Jennings SM, Tsay TH, Fisch TM, Robinson GW (1991) Molecular cloning and characterization of the yeast gene for squalene synthesis. *Proc Natl Acad Sci USA* 88: 6038-6042.
23. Johnston M, Andrew S, Brinkman R, Cooper J, Ding H, Dover J, Favello A, Fulton L, Gattung S, and et al (1994) Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome VIII. *Science* 265(5181): 2077-82.
24. Kahvejian A, Quanckenbush J, Thompson JF (2008) What would you do if you could sequence everything? *Nat Biotechnol* 26(10): 1125-1133.
25. Maury J, Asadollahi MA, Møller K, Clark A, Nielsen J (2005) Microbial isoprenoid production: an example of green chemistry through metabolic engineering. *Adv Biochem Engin/Biotechnol* 100: 19-51.
26. McCammon MT, Epstein CB, Przybyla-Zawislak B, McAlister-Henn L, Butow RA (2003) Global transcription analysis of Krebs tricarboxylic acid cycle mutants reveals an alternating pattern of gene expression and effects of hypoxic and oxidative genes. *Mol Biol Cell* 14(3): 958-972.
27. Middleton B, Tubbs PK (1975) 3-Hydroxy-3-methylglutaryl-CoA synthase from baker's yeast. *Methods Enzymol* 35: 173-177.
28. Monteiro PT, Mendes N, Teixeira MC, d-Orey S, Tenreiro S, Mira N, Pais H, Francisco AP, Carvalho AM, Lourenço A, Sá-Correia I, Oliveria AL, Freitas AT (2008) YEASTRACT-DICOVERER: new tools to improve the analysis of transcriptional regulatory association in *Saccharomyces cerevisiae*. *Nucl Acids Res* 36: D132-D136.
29. Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92: 255-264.
30. Mortimer RK, Johnston JR (1986) Genealogy of principal strains of the yeast genetic stock center. *Genetics* 113(1): 35-43.
31. Nielsen J (1995) Physiological engineering aspects of *Penicillium chrysogenum*. Polyteknisk Forlag, Lyngby, p. 13.
32. Nielsen J (2001) Metabolic engineering. *Appl Microbiol Biotechnol* 55(3): 263-283.
33. Nielsen J and Jewett MC (2008) Impact of systems biology on metabolic engineering of *Saccharomyces cerevisiae*. *FEMS Yeast Res* 8(1): 122-131.
34. Nikolaev SI, Iseli C, Sharp AJ, Robyr D, Rougemont J, Gehrig C, Farinelli L, Antonarakis SE (2009) Detection of genomic variation by selection of a 9 Mb DNA region and high throughput sequencing. *PLOS One* 4(8): e6659.
35. Nookaei I, Jewett MC, Meechai A, Theammarongtham C, Laoteng K, Cheevadhanarak S, Nielsen J, Bhumriatana S (2008) The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol* 2:71.
36. Oliver SG (2006) From genomes to systems: the path with yeast. *Philos Trans R Soc Lond B Biol Sci* 361(1467): 477-482.
37. Porter JW (1985) Mevalonate kinase. *Methods Enzymol* 110: 71-78.
38. Rilling HC (1985) Eukaryotic prenyltransferases. *Methods Enzymol* 110: 145-152.
39. Rougemonet J, Amzallag A, Iseli C, Farinelli L, Xenarios I, Naef F (2008) Probabilistic case calling of Solexa sequencing data. *BMC Bioinformatics* 9: 431.
40. Schacherer J, Ruderfer DM, Gresham D, Dolinski K, Botstein D, Kruglyak (2007) Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. *PLOS One* 3: e322.
41. Shendure J and Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26(10): 1135-1145.
42. Smyth GK (2000) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* 3, 1(3).

43. Smyth GK (2005) Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds.), Springer, New York, pages 397-420.
44. Srivatsan A, Han Y, Peng J, Tehranchi AK, Gibbs R, Wang JD, Chen R (2008) High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. *PLOS Genetics* 4(8): e1000139.
45. Stephanopoulos G and Vallino JJ (1991) Network rigidity and metabolic engineering in metabolite overproduction. *Science* 252(5013): 1675-1681.
46. Tchen TT (1958) Mevalonic kinase: purification and properties. *J Biol Chem* 233(5): 1100-1103.
47. Teixeria MC, Monteiro P, Jain P, Tenreiro S, Fernandes AR, Mira NP, Alenquer M, Freitas AT, Oliveira AL, Sá-Correia I (2006) The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucl Acids Res* 34: D446-D451.
48. Tsay YH, Robinson GW (1991) Cloning and characterization of ERG8, an essential gene of *Saccharomyces cerevisiae* that encodes phosphomevalonate kinase. *Mol Cell Biol* 11(2): 620-631.
49. Tyo KE, Alper HS, Stephanopoulos GN (2007) Expanding the metabolic engineering toolbox: more options to engineer cells. *Trends Biotechnol* 25(3): 132-137.
50. van Dijken JP, Bauer J, Brambilla L, Duboc P, François JM, Gancedo C, Giuseppin ML, Heijnen JJ, Hoare M, Lange HC, Madden EA, Niederberger P, Nielsen J, Parrou JL, Petit T, Porro D, Reuss M, van Riel M, Rizzi M, Steensma HY, Verrips CT, Vindeløv J, Pronk JT (2000) An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb Technol* 26 (9-10): 706-714.
51. Verudyn, C. et al. 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 (1992).
52. Warner JR, Patnaik R, Gill RT (2009) Genomics enabled approaches in strain engineering. *Curr Opin Microbiol* 12(3): 223-230.
53. Westerhoff HV and Palssøn BO (2004) The evolution of molecular biology into systems biology. *Nat Biotechnol* 22(10): 1249-1252.
54. Winston F, Dollard C, Ricupero-Hovasse SL (1995) Construction of a set of convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast* 11(1): 53-55.

**Table 1.** Description of *Saccharomyces cerevisiae* strains

Strain Name	Genotype	Source	References
S288C	<i>MATα SUC2 gal2 mal mel flo1 flo8-1 hap1 ho bio1 bio6</i> First <i>S. cerevisiae</i> sequenced and stored in SGD. Strain not capable of anaerobic galactose metabolism ( <i>gal2</i> ), and contains a mutated copy of <i>HAP1</i> with a Ty1 insertion in the carboxy terminus. Strain resequenced as part of SGRP. Strain is a prototrophic haploid.	ATCC® <sup>A</sup>	Mortimer, 1986 Johnston, 1994 Goffeau 1996 Cherry, 1997
CEN.PK 113-7D	<i>MATa URA3 HIS3 LEU2 TRP1 SUC2 MAL2-8<sup>C</sup></i> CEN.PK strain family was constructed as part of an interdisciplinary German research project ("Stofflässe in Mikroorganismen"). Strain is a prototrophic haploid. The strain was obtained from Dr. P. Kötter (J.W. Universität, Frankfurt, Germany).	SRD GmbH <sup>B</sup>	Cherry, 1997 van Dijken, 2000

**NOTES:** A. American Type Culture Collection (ATCC®).

B. Scientific Research and Development (SRD) GmbH.

**Table 2.** Physiological characterization of *S. cerevisiae* strains S288c and CEN.PK113-7D

STRAIN SUBSTRATE	S288c Glucose		CEN.PK113-7D Glucose		S288c Galactose		S288c Galactose/Ethanol		CEN.PK113-7D Galactose	
	Mean	$\pm SD$ (n=2)	Mean	$\pm SD$ (n=2)	Mean	$\pm SD$ (n=2)	Mean	$\pm SD$ (n=2)	Mean	$\pm SD$ (n=2)
$\mu\text{-max (h}^{-1}\text{)}$	0.31	0.01	0.41	0.01	0.02	0.00	0.14	0.01	0.27	0.00
<b>Carbon Recovery (%)</b>	96.6	1.9	95.5	3.9	n/a	n/a	n/a	n/a	79.6	2.6
<b>Specific Productivity or Consumption Rate<sup>A</sup></b>										
- $r_{\text{gluc}}$ or - $r_{\text{gal}}$	79.35	5.48	105.15	0.24	1.21	0.57	4.50	0.70	24.28	0.33
$r_{\text{CO}_2}$	18.36	0.52	23.62	0.87	0.11	0.01	0.58	0.12	4.31	0.24
$r_{\text{EtOH}}$	37.59	4.39	49.88	0.34	0.01	0.01	-3.97	0.62	3.50	0.26
$r_{\text{Acet}}$	0.24	0.02	1.18	0.04	0.00	0.00	-0.04	0.02	1.10	0.11
$r_{\text{Glyc}}$	6.08	0.94	7.17	2.64	0.00	0.00	-0.58	0.05	0.89	0.09
$r_{\text{Pyr}}$	0.47	0.04	0.69	0.06	0.00	0.00	0.00	0.00	0.04	0.01
$r_{\text{Suc}}$	0.01	0.01	0.03	0.05	0.00	0.00	0.00	0.00	0.00	0.00
$r_X$	13.85	1.11	17.79	0.02	0.78	0.05	5.85	0.20	9.49	0.89
- $r_{\text{O}_2}$	1.95	0.07	1.98	2.75	0.08	0.01	0.97	0.20	2.91	0.22

NOTES: A. (C-mmol/g-DCW/h). The term "n/a" refers to not applicable.

**Table 3.** Illumina Genome Sequencing Results

Sequencing Parameter	S288C	CEN.PK113-7D
<b>No. of Reads</b>	5,301,907	6,603,200
<b>No. of Aligned Reads</b>	5,176,155	6,431,119
<b>Total Bases<sup>A</sup> (bp)</b>	181,165,425	217,579,460
<b>Calculated Average Coverage</b>	15X	18X
<b>Genome Percent Reference Coverage (%)</b>	99.9	99.5
<i>MAQ Software Determination</i>		
<b>No. of Contigs</b>	660	1434
<b>Total Gap Size (bp)</b>	10,403	74,619
<b>Total No. of SNPs</b>	3,032	27,868
<b>Total No. of Non-Ambiguous SNPs</b>	1,013	24,663
<b>Total No. of Filtered SNPs<sup>B</sup></b>	311	13,787

**Notes:** Basepairs (bp). (A) Each read is 35 basepairs in length. (B) Filtered SNPs determined based on cut-off criteria within the Mapping and Assembling with Quality (MAQ) software environment.

**Table 4.** Metabolic SNP Detection

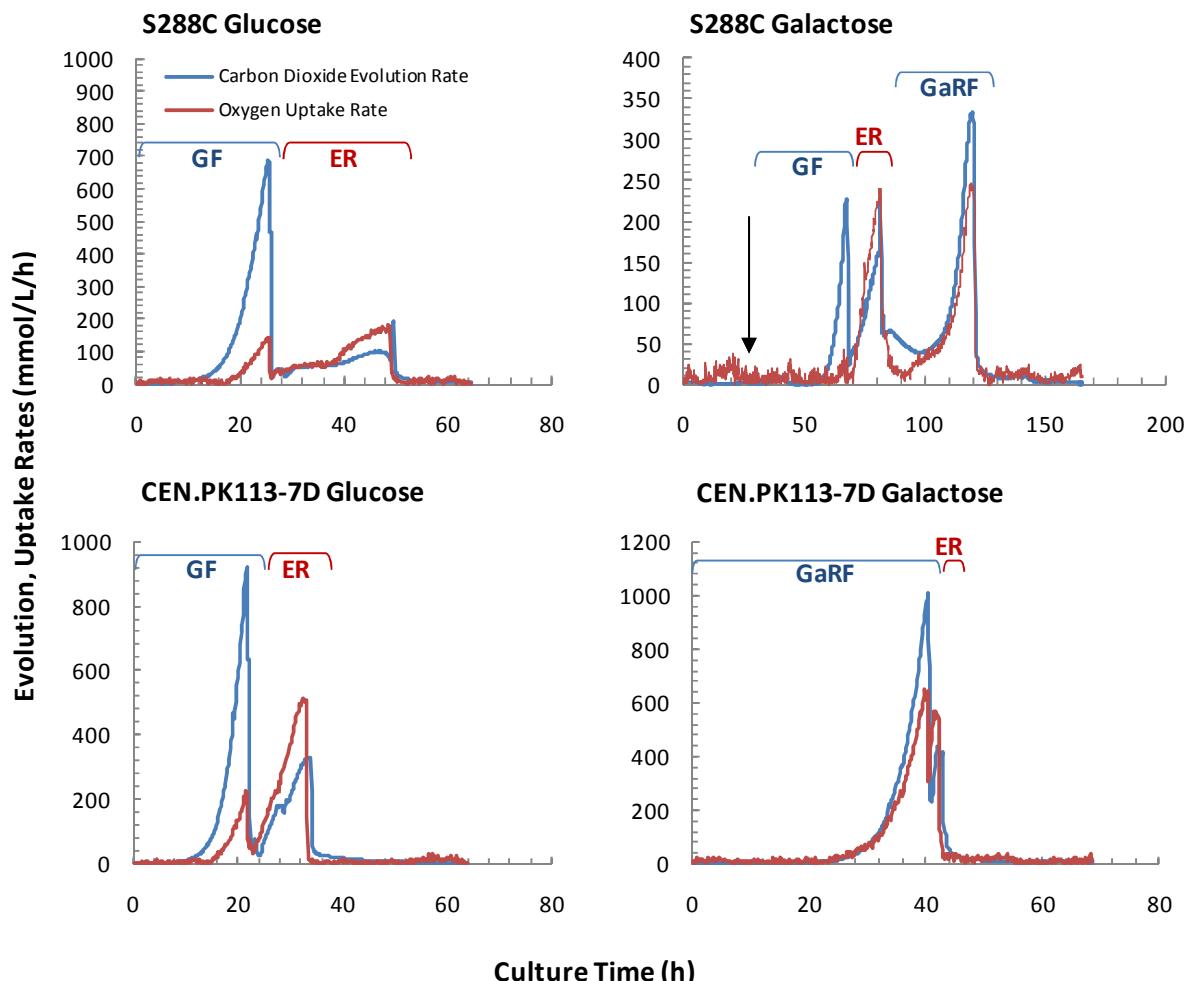
Metabolic SNP Detection Parameter	S288C	CEN.PK113-7D
<b>Total No. of Metabolic Genes Considered<sup>A</sup></b>	782	782
<b>Total No. of Metabolic Bases (bp)</b>	1.16M	1.16M
 <i>MAQ Software Determination</i>		
<b>No. of Reads</b>	5,301,907	6,603,200
<b>No. of Aligned Reads</b>	477,565	623,400
<b>Total Gap Size (bp)</b>	0	0
<b>Metabolic Genome Percent Reference Coverage (%)</b>	99.7	99.4
<b>Total No. of Filtered SNPs</b>	36	939
 <i>Metabolism Focused Detection</i>		
<b>Total No. of Metabolic SNPs Detected</b>	36	939
<b>Total No. Non-silent Metabolic SNPs Detected</b>	3	219
<b>Percent of SNPs Detected Non-silent (%)</b>	8.3	23.3
<b>Total No. of Metabolic Genes Containing SNP</b>	14	158
<b>Total No. of Metabolic Genes with Non-silent SNP</b>	3	85

**Notes:** Basepairs (bp). (A) The total number of genes classified as metabolic was based on the Saccharomyces Genome Database, Strain S288C, version 12.0. The “S288C” designation in Table 4 refers to the resequencing of *S. cerevisiae* S288C using Illumnia sequencing technology in this work.

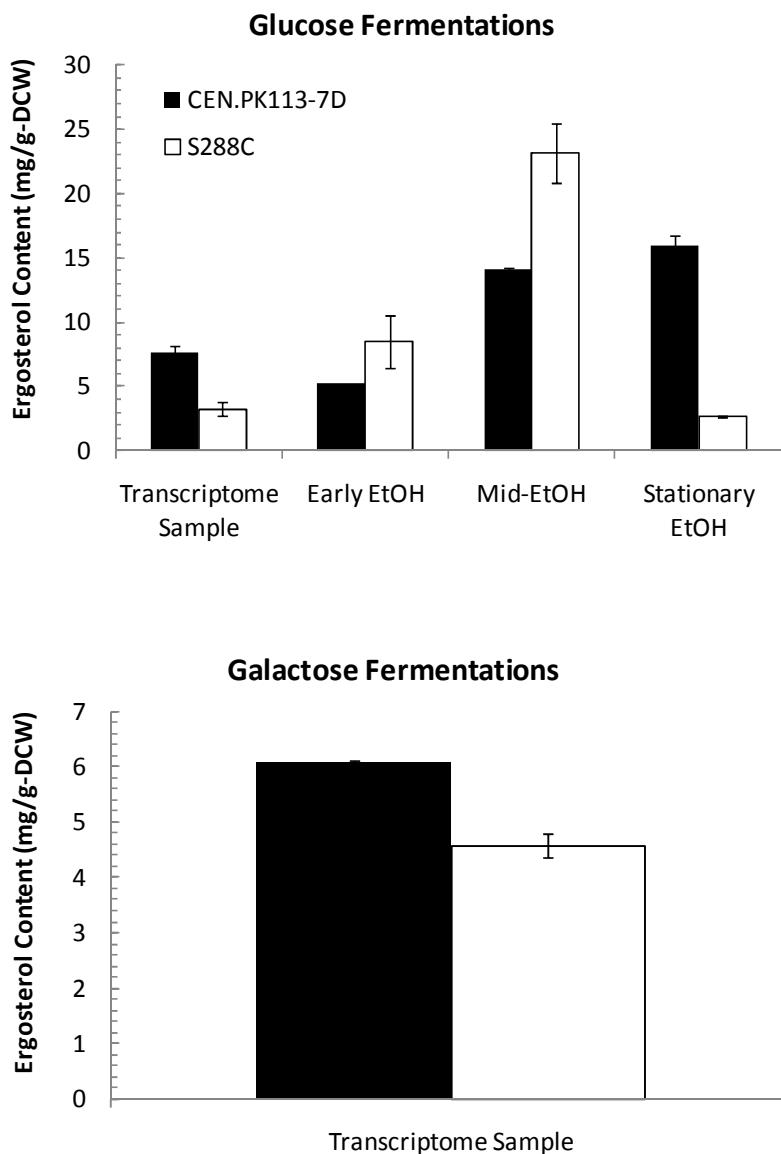
**Table 5.** Summary of Differential Gene Expression

Summary of Differential Expression ( $p_{adj} < 0.01$ )	S288C v. CEN.PK113-7D	
	Glucose (n=2)	Galactose (n=2)
Total No. of Differentially Expressed Genes	272	501
No. of Genes LFC>0	204	337
LFC ± SD	$2.13 \pm 1.41$	$1.81 \pm 1.17$
No. of Genes with SNPs Detected	13	17
No. of Genes with Non-silent SNPs	7	10
No. of Genes LFC<0	68	164
LFC ± SD	$-2.12 \pm 1.32$	$-1.53 \pm 1.05$
No. of Genes with SNPs Detected	4	4
No. of Genes with Non-silent SNPs	1	0

**Notes:** Single nucleotide polyphormism (SNP). Standard deviation (SD). Non-silent SNPs defined as a nucleotide modification results in a translated amino acid modification.

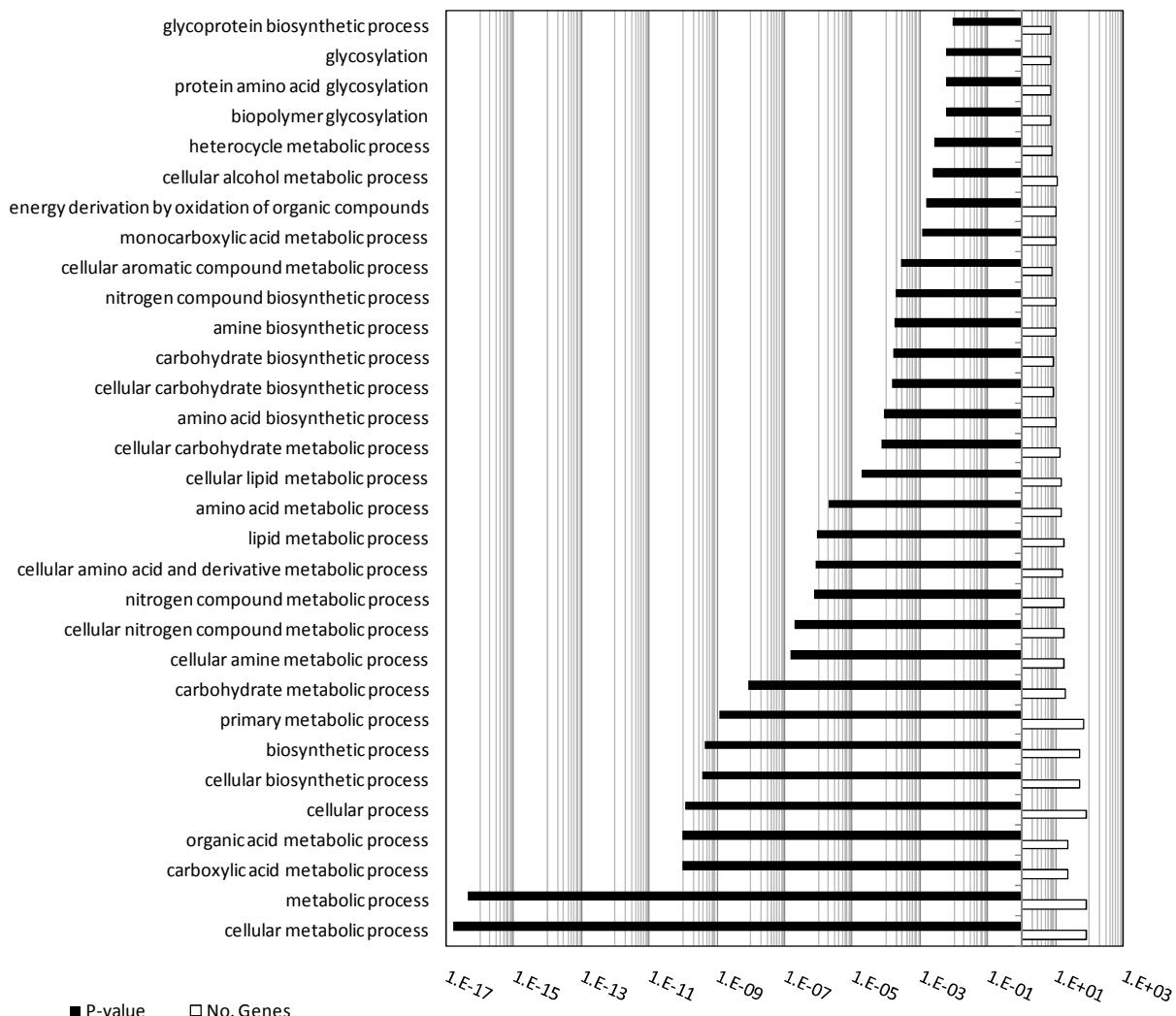


**Figure 1.** The plots above show the carbon dioxide evolution rate and oxygen uptake rate as a function of cultivation time for the strains S288C and CEN.PK113-7D supplemented with glucose and galactose, respectively. Glucose fermentation (GF), ethanol respiration (ER), galactose respiro-fermentation (GaRF). The black arrow in the S288C Galactose plot indicates when 20 g L<sup>-1</sup> glucose was supplemented (25h) when no growth was observed on galactose.

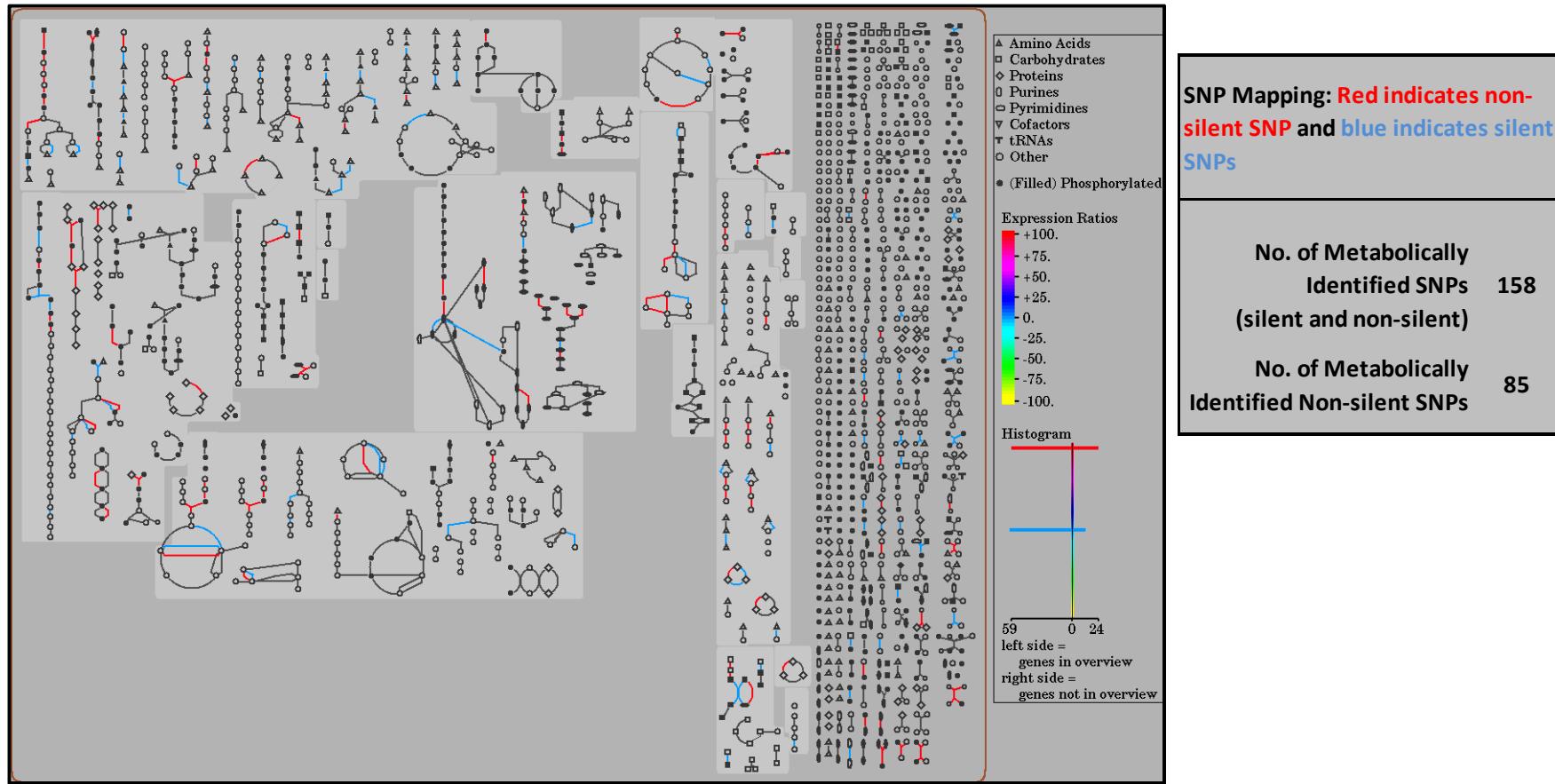


**Figure 2.** Ergosterol content (mg/g-DCW) was measured for different samples taken during S288C and CEN.PK113-7D fermentations, supplemented with glucose and galactose. Transcriptome Sample was taken during the mid-exponential fermentation phase on glucose or respiration phase on galactose. For glucose fermentations, early ethanol, mid-ethanol, and stationary ethanol samples were taken post-diauxic shift to characterize the change in ergosterol during growth on ethanol. Error bars are  $\pm$  SD ( $n=2$ ).

### GO PROCESS, Non-Silent SNP Characterization

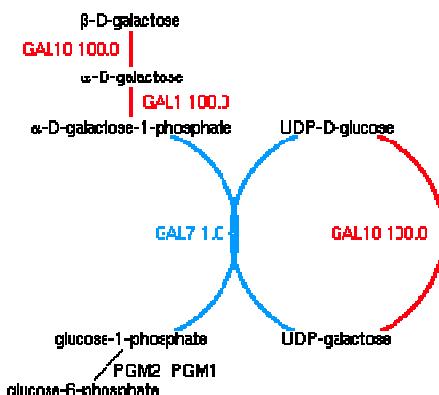


**Figure 3.** Gene ontology (GO) process terms for the non-silent SNPs identified in CEN.PK113-7D compared to S288C. The x-axis in log-scale displays both the significance of each category ( $p < 0.01$ , symbol: solid black), and the number of genes from the total of 85 containing non-silent SNPs (symbol: solid white). GO process characterization performed using the Saccharomyces Genome Database (SGD).



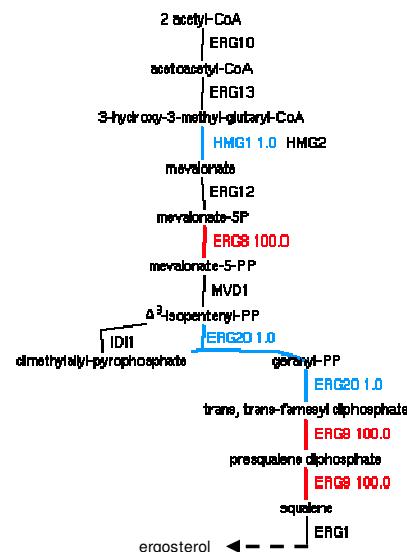
**Figure 4.** The metabolic map produced using the Saccharomyces Genome Database (SGD) Expression Viewer (SRI International Pathway Tools version 12.0, based upon *Saccharomyces cerevisiae* S288C, version 12.0) was created using the SNP data produced for CEN.PK113-7D compared to S288C. Pathways in red indicate non-silent SNPs (85 genes) while those in blue indicate silent SNPs (73 genes). Note that number of genes does not necessarily coincide with number of pathways due to isoenzymes.

### Galactose Uptake Pathway



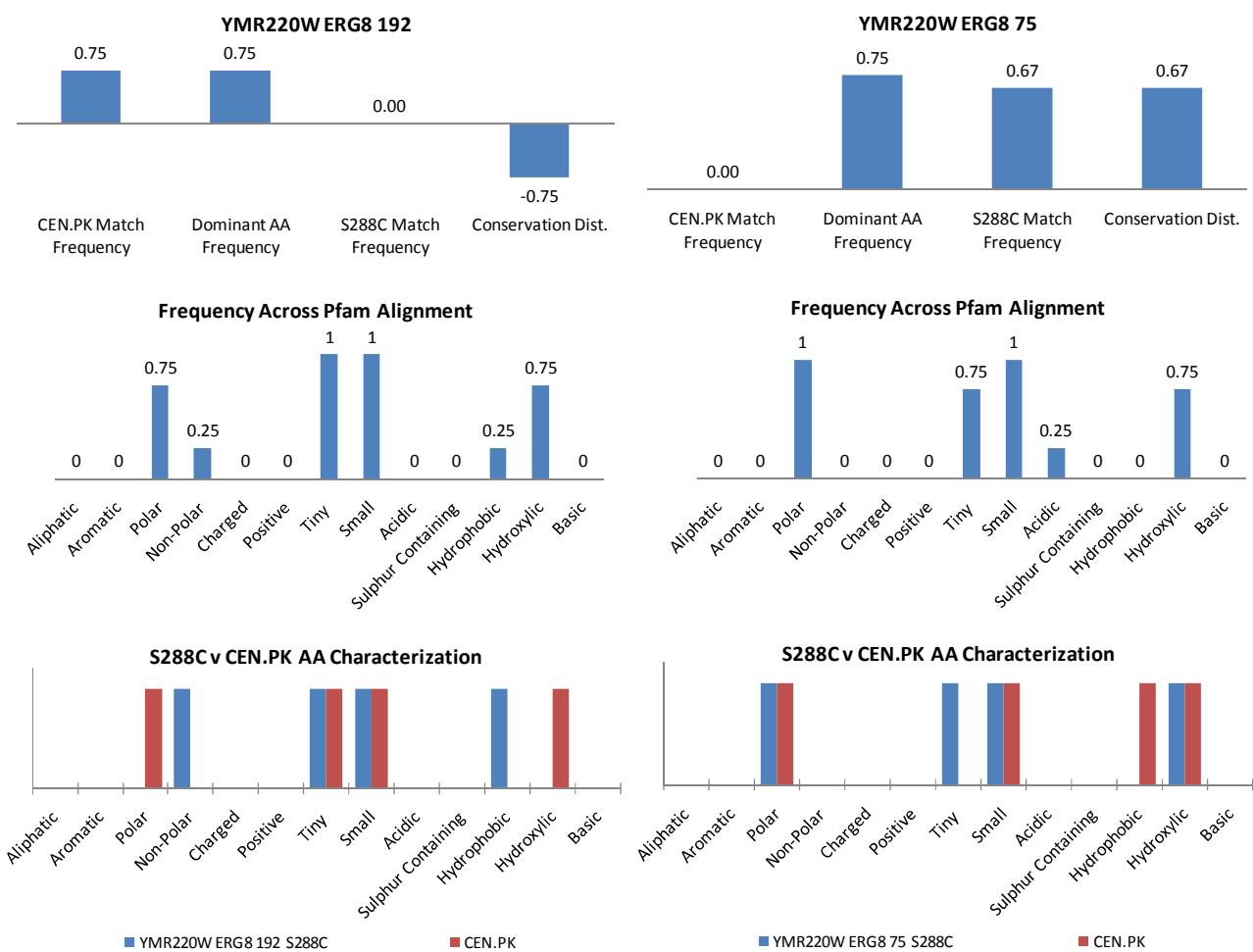
- Non-silent SNP (Red)
- Silent SNP (Blue)

### Ergosterol Synthesis Pathway



Gene Name (Systematic Name)	Nucleotide Position	Nucleotide (S288C)	Nucleotide (CEN.PK113-7D)	Amino Acid Position	Amino Acid (S288C)	Amino Acid (CEN.PK113-7D)
ERG8 (YMR220W)	146	G	A	49	G	E
	224	G	C	75	S	T
	574	G	T	192	A	S
	739	G	A	247	D	N
	306	A	G	102	K	K
	387	T	C	129	D	D
	759	T	A	253	I	I
	879	A	G	293	P	P
	1038	A	G	346	E	E
	856	G	A	286	G	S
ERG9 (YHR190W)	1125	G	A	375	Q	Q
	1128	T	C	376	F	F
	890	T	C	297	L	P
GAL1 (YBR020W)	1241	C	A	414	A	E
	1089	A	G	363	A	A
	1257	C	T	419	D	D
	1530	A	T	510	L	L
	787	C	A	263	Q	K
GAL10 (YBR019C)	1554	G	C	518	C	G
	1798	T	G	599	C	G
	1590	G	A	530	T	T
	1677	A	G	559	S	S
	1827	G	A	609	K	K

**Figure 5.** Two pathways with a significant number of SNPs, both silent (blue and denoted with 1.0 next to gene in pathway map) and non-silent (red and denoted with 100.0 next to gene in pathway map) are included: galactose uptake pathway and ergosterol synthesis pathway. Both standard single letter codes for nucleotides and amino acids are utilized.

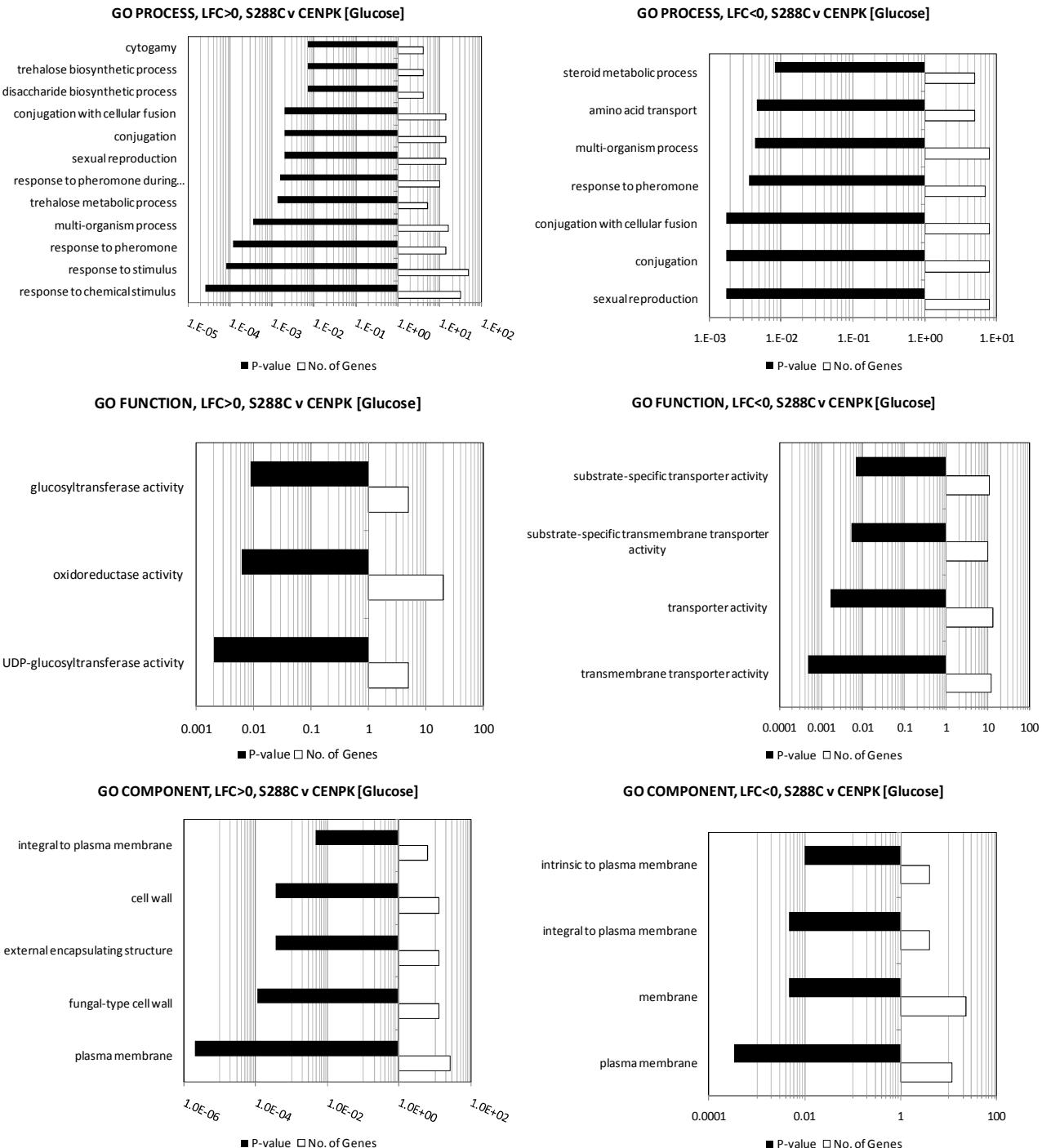


**Figure 6.** The gene *ERG8* of the ergosterol synthesis pathway contains a total of four non-silent SNPs, two of which, located at nucleotide positions 192 and 75, are analyzed here. The top plots show the CEN.PK Match Frequency, Dominant AA Frequency, S288C Match Frequency, and Conservation Distance. The middle plots show the frequency (fraction) of each categorization across the amino acid sequences resulting from Pfam multi-sequence alignment. The bottom plots shows the characterization of the original S288C amino acid (symbol: red bar) and the CEN.PK113-7D amino acid (symbol: blue bar). The gene *ERG8* contained a total of 4 non-silent SNPs, and Supplementary Materials Figure 8 includes the other 2 non-silent SNPs (nucleotide positions 49 and 247).

## **Supplementary Materials**

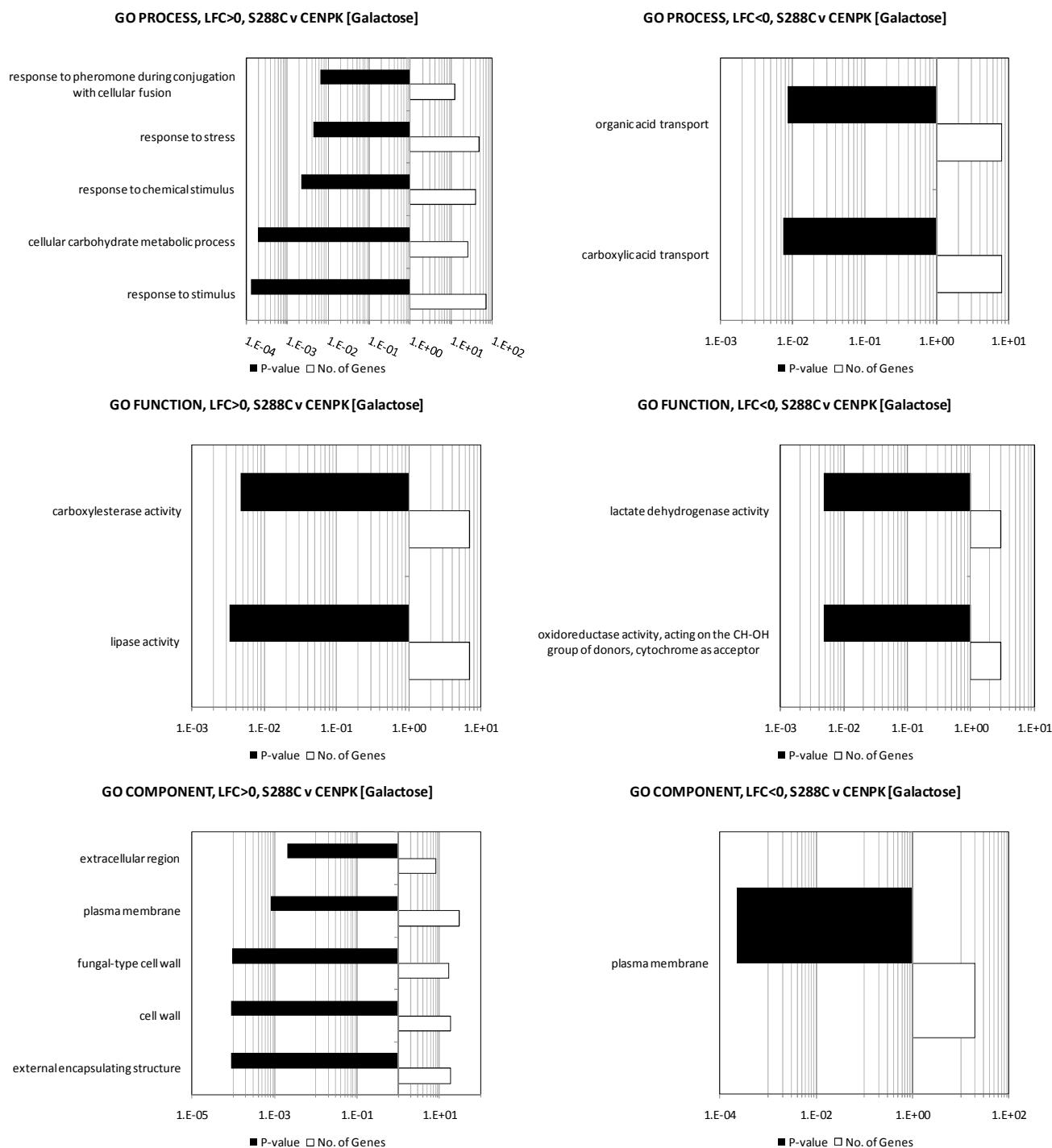
### **Table of Contents:**

- 1. Supplementary Materials Figure 1:** Gene ontology terms for S288C v. CENPK (Glucose).
- 2. Supplementary Materials Figure 2:** Gene ontology terms for S288C vs. CEN.PK (Galactose).
- 3. Supplementary Materials Figure 3:** Gene ontology terms for SNP characterization.
- 4. Supplementary Materials Figure 4:** Metabolic pathway expression mapping for S288C Glucose vs. CENPK Glucose.
- 5. Supplementary Materials Figure 5:** Metabolic pathway expression mapping for S288C Galactose vs. CENPK Galactose.
- 6. Supplementary Materials Table 1:** Transcriptome – S288C v CENPK Glucose.
- 7. Supplementary Materials Table 2:** Transcriptome – S288C v CENPK Galactose.
- 8. Supplementary Materials Figure 6:** Methodology for SNP characterization at amino acid level.
- 9. Supplementary Materials Figure 7:** Conservation distance.
- 10. Supplementary Materials Figure 8:** SNP analysis of ERG8 at nucleotide positions 49 and 247.
- 11. Supplementary Materials Table 3:** EDENA determination for *de novo* assembly of S288C and CEN.PK113-7D sequences.
- 12. Supplementary Materials Figure 9:** Amino acid characterization of SNPs detected between S288C and CEN.PK113-7D.

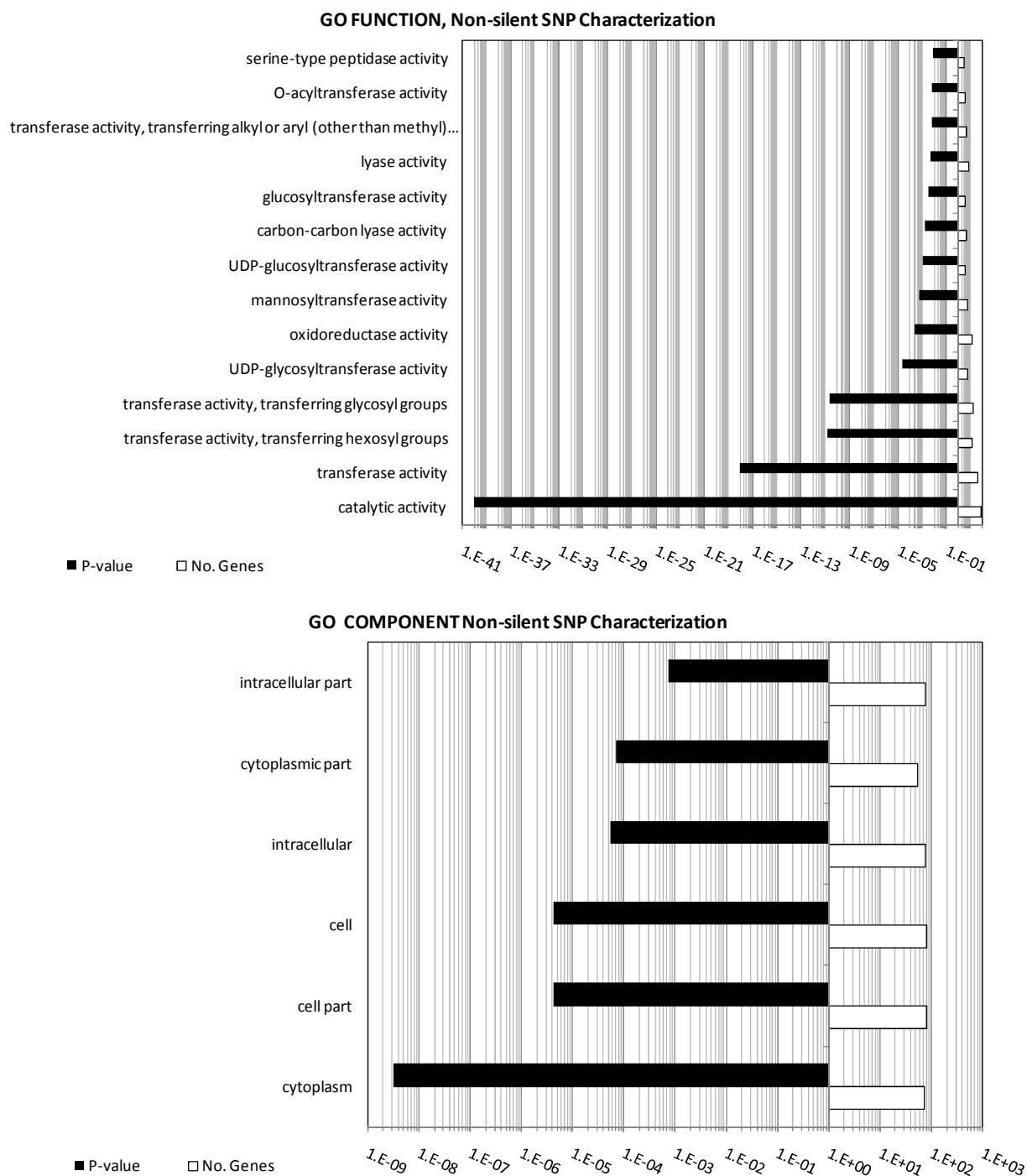


**Supplementary Materials Figure 1.** Gene ontology (GO) process, function, and component terms for differentially expressed genes of S288C vs. CEN.PK113-7D cultivated on glucose. The x-axis in log-scale displays both the significance of each category ( $p < 0.01$ , symbol: solid black), and the number of differentially expressed genes in each GO category (symbol: solid white). The GO terms were

determined separately for genes expressing positive and negative log-fold change (LFC). GO characterization performed using the *Saccharomyces* Genome Database (SGD).



**Supplementary Materials Figure 2.** Gene ontology (GO) process, function, and component terms for differentially expressed genes of S288C vs. CEN.PK113-7D cultivated on galactose. The x-axis in log-scale displays both the significance of each category ( $p < 0.01$ , symbol: solid blank), and the number of differentially expressed genes in each GO category (symbol: solid white). The GO terms were determined separately for genes expressing positive and negative log-fold change (LFC). GO characterization performed using the *Saccharomyces Genome Database* (SGD).



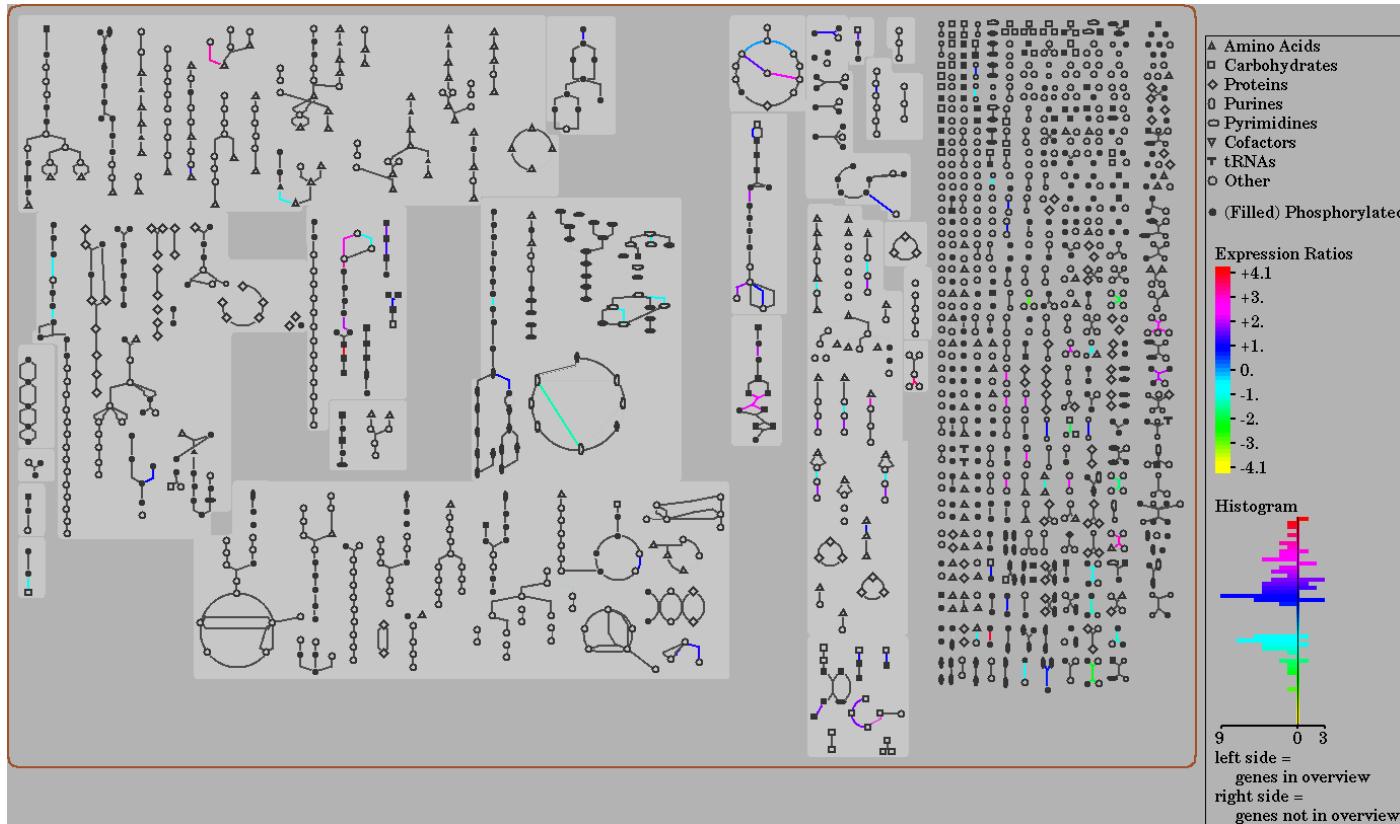
**Supplementary Materials Figure 3.** Gene ontology (GO) function and component terms for the non-silent SNPs identified in CEN.PK113-7D compared to S288C. The x-axis in log-scale displays both the significance of each category ( $p < 0.01$ , symbol: solid black), and the number of genes from the total of 85 containing non-silent SNPs (symbol: solid white). GO process characterization performed using the Saccharomyces Genome Database (SGD).



**S288C Glucose vs. CENPK  
Glucose (Transcriptome)**

No. of Differentially Expressed Genes ( $p.\text{adj} < 0.01$ )	272
No. of Metabolically Mapped Genes	59
Mean Expression Level	1.45
Median Expression Level	1.67

**Supplementary Materials Figure 4.** The metabolic map produced using the *Saccharomyces* Genome Database (SGD) Expression Viewer (SRI International Pathway Tools version 12.0, based upon *Saccharomyces cerevisiae* S288C, version 12.0) was created using statistically significant log-fold expression values for S288C glucose vs. CEN.PK113-7D glucose.



**S288C Galactose vs. CENPK  
Galactose (Transcriptome)**

No. of Differentially Expressed Genes ( $p.adj < 0.01$ )	508
No. of Metabolically Mapped Genes	94
Mean Expression Level	0.86
Median Expression Level	1.09

**Supplementary Materials Figure 5.** The metabolic map produced using the Saccharomyces Genome Database (SGD) Expression Viewer (SRI International Pathway Tools version 12.0, based upon *Saccharomyces cerevisiae* S288C, version 12.0) was created using statistically significant log-fold expression values for S288C galactose vs. CEN.PK113-7D galactose.

**Supplementary Materials Table 1**

**Transcriptome: S288C v CENPK GLUCOSE ( $p_{adj} < 0.01$ )**

Systematic Gene Name	logFC	p <sub>adj</sub> value	Standard Gene Name	Description <sup>1</sup>	Silent SNP	Non-silent SNP
YBR001C	1.67	3.55E-04	<i>NTH2</i>	Putative neutral trehalase, required for thermotolerance and may mediate resistance to other cellular stresses	Y	
YMR250W	2.04	7.12E-04	<i>GAD1</i>	Glutamate decarboxylase, converts glutamate into gamma-aminobutyric acid (GABA) during glutamate catabolism; involved in response to oxidative stress	Y	
YML100W	2.61	7.41E-04	<i>TSL1</i>	Large subunit of trehalose 6-phosphate synthase (Tps1p)/phosphatase (Tps2p) complex, which converts uridine-5'-diphosphoglucose and glucose 6-phosphate to trehalose, homologous to Tps3p and may share function	Y	
YBR020W	1.55	9.72E-04	<i>GAL1</i>	Galactokinase, phosphorylates alpha-D-galactose to alpha-D-galactose-1-phosphate in the first step of galactose catabolism; expression regulated by Gal4p	Y	
YDR538W	1.61	1.55E-03	<i>PAD1</i>	Phenylacrylic acid decarboxylase, confers resistance to cinnamic acid, decarboxylates aromatic carboxylic acids to the corresponding vinyl derivatives; homolog of <i>E. coli</i> UbiX	Y	
YLR258W	2.77	6.34E-03	<i>GSY2</i>	Glycogen synthase, similar to Gsy1p; expression induced by glucose limitation, nitrogen starvation, heat shock, and stationary phase; activity regulated by cAMP-dependent, Snf1p and Pho85p kinases as well as by the Gac1p-Glc7p phosphatase	Y	
YHL012W	0.85	8.07E-03	<i>n/a</i>	Putative protein of unknown function, has some homology to Ugp1p, which encodes UDP-glucose pyrophosphorylase	Y	

YFR015C	2.05	7.41E-05	<i>GSY1</i>	Glycogen synthase with similarity to Gsy2p, the more highly expressed yeast homolog; expression induced by glucose limitation, nitrogen starvation, environmental stress, and entry into stationary phase	Y
YJL172W	1.47	5.17E-04	<i>CPS1</i>	Vacuolar carboxypeptidase yscS; expression is induced under low-nitrogen conditions	Y
YJL166W	2.99	6.09E-04	<i>QCR8</i>	Subunit 8 of ubiquinol cytochrome-c reductase complex, which is a component of the mitochondrial inner membrane electron transport chain; oriented facing the intermembrane space; expression is regulated by Abf1p and Cpf1p	Y
YHR216W	1.36	1.08E-03	<i>IMD2</i>	Inosine monophosphate dehydrogenase, catalyzes the first step of GMP biosynthesis, expression is induced by mycophenolic acid resulting in resistance to the drug, expression is repressed by nutrient limitation	Y
YAL062W	1.66	3.61E-03	<i>GDH3</i>	NADP(+)-dependent glutamate dehydrogenase, synthesizes glutamate from ammonia and alpha-ketoglutarate; rate of alpha-ketoglutarate utilization differs from Gdh1p; expression regulated by nitrogen and carbon sources	Y
YMR101C	1.74	5.89E-03	<i>SRT1</i>	Cis-prenyltransferase involved in synthesis of long-chain dolichols (19-22 isoprene units; as opposed to Rer2p which synthesizes shorter-chain dolichols); localizes to lipid bodies; transcription is induced during stationary phase	Y
YJR078W	-1.50	4.12E-03	<i>BNA2</i>	Putative tryptophan 2,3-dioxygenase or indoleamine 2,3-dioxygenase, required for the de novo biosynthesis of NAD from tryptophan via kynurenine; expression is upregulated upon telomere uncapping; regulated by Hst1p and Aft2p	Y

YGR287C	-1.06	4.80E-03	<i>n/a</i>	Protein of unknown function that may interact with ribosomes, based on co-purification experiments; has similarity to alpha-D-glucosidase (maltase); authentic, non-tagged protein detected in purified mitochondria in high-throughput studies	Y
YKL029C	-1.10	4.12E-03	<i>MAE1</i>	Mitochondrial malic enzyme, catalyzes the oxidative decarboxylation of malate to pyruvate, which is a key intermediate in sugar metabolism and a precursor for synthesis of several amino acids	Y
YML075C	-1.69	3.32E-04	<i>HMG1</i>	One of two isozymes of HMG-CoA reductase that catalyzes the conversion of HMG-CoA to mevalonate, which is a rate-limiting step in sterol biosynthesis; localizes to the nuclear envelope; overproduction induces the formation of karmellae	Y

**NOTES:** 1. Descriptions adopted from Saccharomyces Genome Database (SGD).

**Supplementary Materials Table 2**

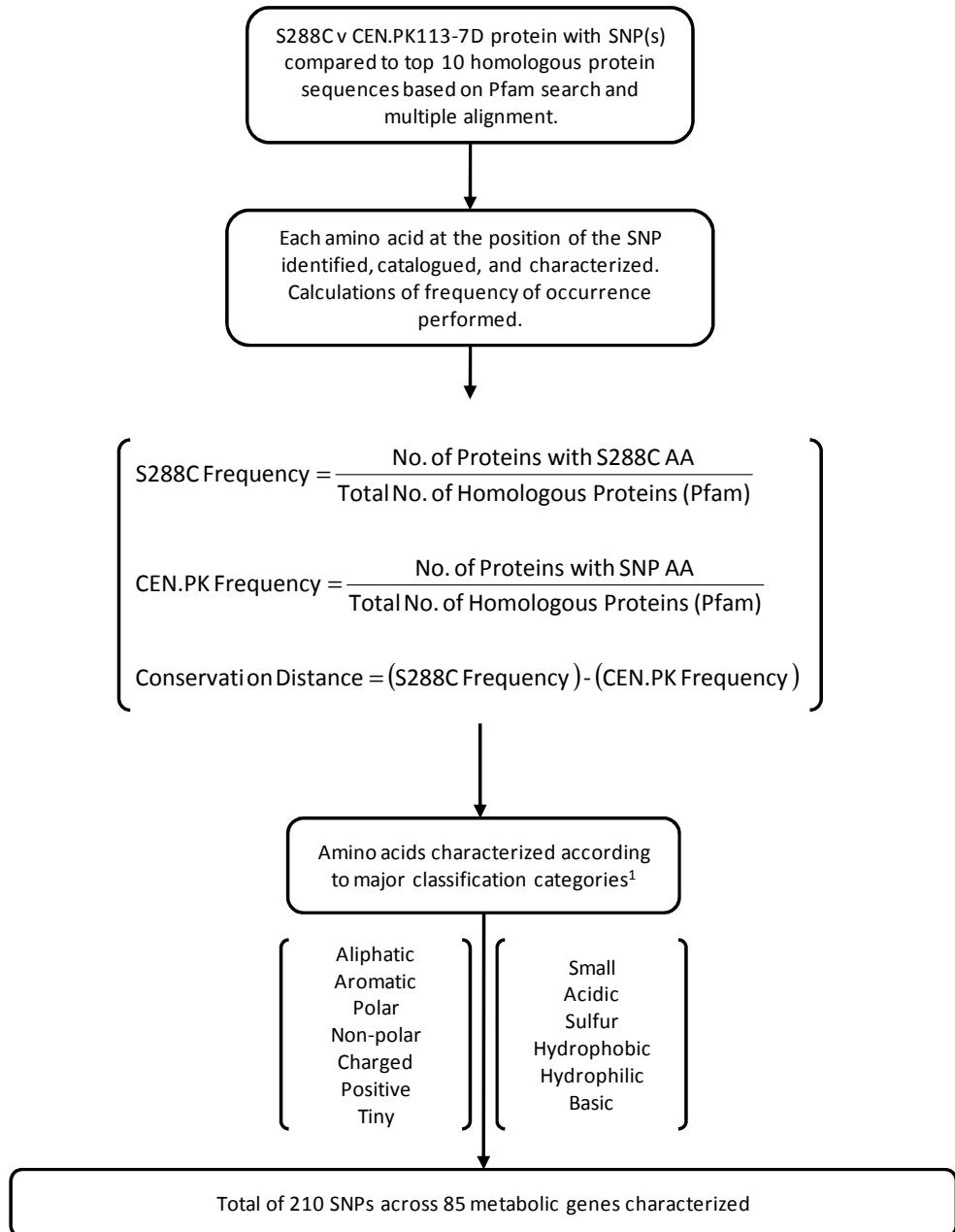
**Transcriptome: S288C v CENPK GALACTOSE ( $p_{adj} < 0.01$ )**

Systematic Gene Name	logFC	$p_{adj}$ value	Standard Gene Name	Description	Silent SNP	Non-silent SNP
YMR250W	2.96	1.54E-04	<i>GAD1</i>	Glutamate decarboxylase, converts glutamate into gamma-aminobutyric acid (GABA) during glutamate catabolism; involved in response to oxidative stress	Y	
YBR001C	1.66	3.81E-04	<i>NTH2</i>	Putative neutral trehalase, required for thermotolerance and may mediate resistance to other cellular stresses	Y	
YHL012W	1.36	9.38E-04	<i>n/a</i>	Putative protein of unknown function, has some homology to Ugp1p, which encodes UDP-glucose pyrophosphorylase	Y	
YML100W	2.34	1.08E-03	<i>TSL1</i>	Large subunit of trehalose 6-phosphate synthase (Tps1p)/phosphatase (Tps2p) complex, which converts uridine-5'-diphosphoglucose and glucose 6-phosphate to trehalose, homologous to Tps3p and may share function	Y	
YPL268W	1.27	2.52E-03	<i>PLC1</i>	Phospholipase C, hydrolyzes phosphatidylinositol 4,5-bisphosphate (PIP2) to generate the signaling molecules inositol 1,4,5-triphosphate (IP3) and 1,2-diacylglycerol (DAG); involved in regulating many cellular processes	Y	
YAL054C	1.70	2.54E-03	<i>ACS1</i>	Acetyl-coA synthetase isoform which, along with Acs2p, is the nuclear source of acetyl-coA for histone acetylation; expressed during growth on nonfermentable carbon sources and under aerobic conditions	Y	
YKR066C	1.04	3.22E-03	<i>CCP1</i>	Mitochondrial cytochrome-c peroxidase; degrades reactive oxygen species in mitochondria, involved in the response to oxidative stress	Y	

YLR258W	2.91	3.93E-03	<i>GSY2</i>	Glycogen synthase, similar to Gsy1p; expression induced by glucose limitation, nitrogen starvation, heat shock, and stationary phase; activity regulated by cAMP-dependent, Snf1p and Pho85p kinases as well as by the Gac1p-Glc7p phosphatase	Y
YDR530C	0.76	6.36E-03	<i>APA2</i>	Diadenosine 5',5"-P1,P4-tetraphosphate phosphorylase II (AP4A phosphorylase), involved in catabolism of bis(5'-nucleosidyl) tetraphosphates; has similarity to Apa1p	Y
YER024W	2.17	8.48E-03	<i>YAT2</i>	Carnitine acetyltransferase; has similarity to Yat1p, which is a carnitine acetyltransferase associated with the mitochondrial outer membrane	Y
YAL062W	3.29	1.72E-04	<i>GDH3</i>	NADP(+)-dependent glutamate dehydrogenase, synthesizes glutamate from ammonia and alpha-ketoglutarate; rate of alpha-ketoglutarate utilization differs from Gdh1p; expression regulated by nitrogen and carbon sources	Y
YFR015C	1.44	3.21E-04	<i>GSY1</i>	Glycogen synthase with similarity to Gsy2p, the more highly expressed yeast homolog; expression induced by glucose limitation, nitrogen starvation, environmental stress, and entry into stationary phase	Y
YJL166W	3.16	4.40E-04	<i>QCR8</i>	Subunit 8 of ubiquinol cytochrome-c reductase complex, which is a component of the mitochondrial inner membrane electron transport chain; oriented facing the intermembrane space; expression is regulated by Abf1p and Cpf1p	Y
YER065C	2.71	4.85E-04	<i>ICL1</i>	Isocitrate lyase, catalyzes the formation of succinate and glyoxylate from isocitrate, a key reaction of the glyoxylate cycle; expression of ICL1 is induced by growth on ethanol and repressed by growth on glucose	Y

YDR058C	2.16	3.19E-03	<i>TGL2</i>	Protein with lipolytic activity towards triacylglycerols and diacylglycerols when expressed in <i>E. coli</i> ; role in yeast lipid degradation is unclear	Y
YHR216W	0.89	5.40E-03	<i>IMD2</i>	Inosine monophosphate dehydrogenase, catalyzes the first step of GMP biosynthesis, expression is induced by mycophenolic acid resulting in resistance to the drug, expression is repressed by nutrient limitation	Y
YHR018C	1.37	9.04E-03	<i>ARG4</i>	Argininosuccinate lyase, catalyzes the final step in the arginine biosynthesis pathway	Y
YGR287C	-1.72	4.97E-04	<i>n/a</i>	Protein of unknown function that may interact with ribosomes, based on co-purification experiments; has similarity to alpha-D-glucosidase (maltase); authentic, non-tagged protein detected in purified mitochondria in high-throughput studies	Y
YJL172W	-1.04	1.95E-03	<i>CPS1</i>	Vacuolar carboxypeptidase <i>yscS</i> ; expression is induced under low-nitrogen conditions	Y
YKL029C	-1.07	3.55E-03	<i>MAE1</i>	Mitochondrial malic enzyme, catalyzes the oxidative decarboxylation of malate to pyruvate, which is a key intermediate in sugar metabolism and a precursor for synthesis of several amino acids	Y
YML075C	-0.74	8.58E-03	<i>HMG1</i>	One of two isozymes of HMG-CoA reductase that catalyzes the conversion of HMG-CoA to mevalonate, which is a rate-limiting step in sterol biosynthesis; localizes to the nuclear envelope; overproduction induces the formation of karmellae	Y

**NOTES:** 1. Descriptions adopted from *Saccharomyces Genome Database (SGD)*.

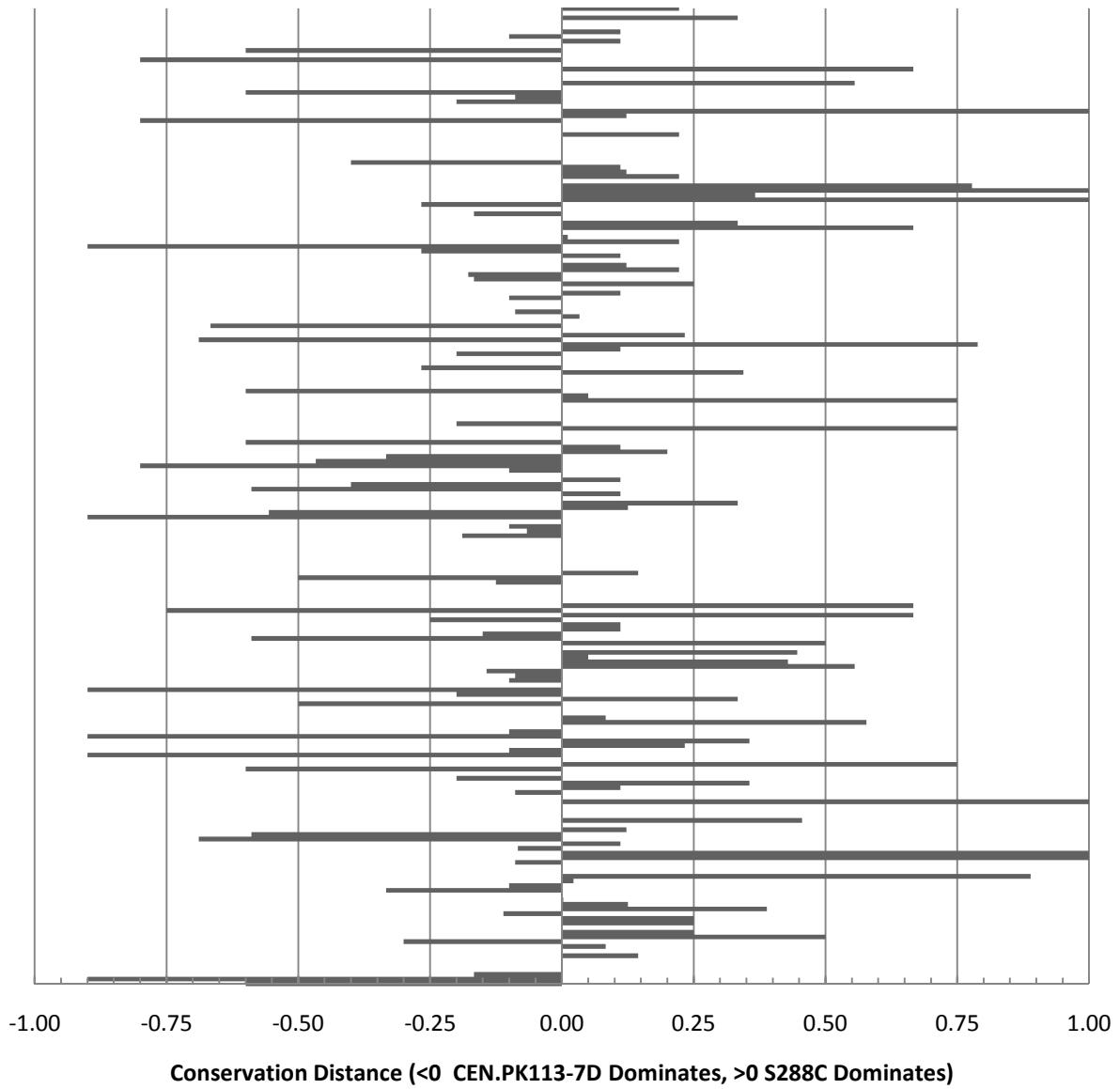


1. Lehninger Principles of Biochemistry, 4<sup>th</sup> edition (David L. Nelson, Michael M. Cox, eds.)

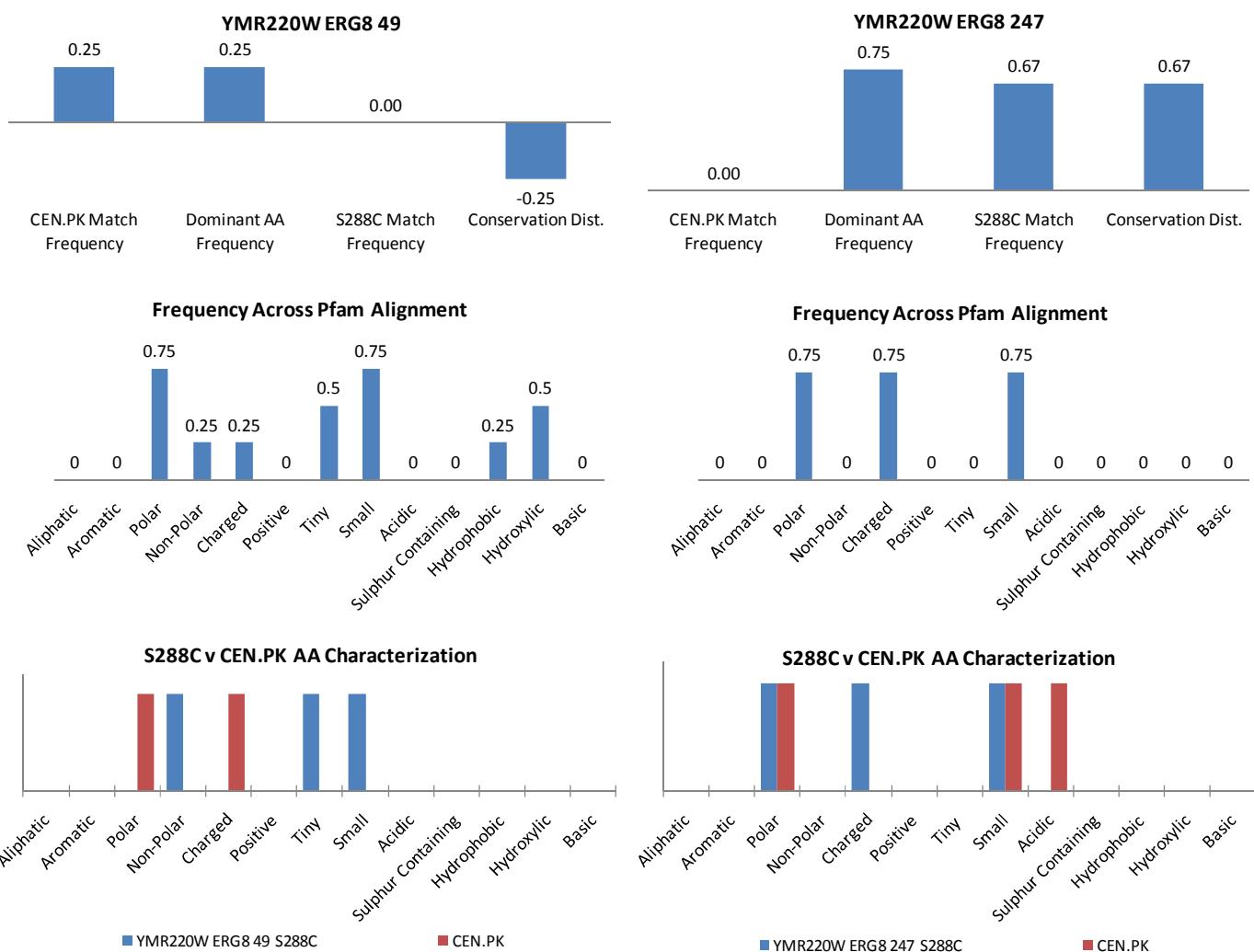
**Supplementary Materials Figure 6.** The above flow-diagram describes the bioinformatics approach taken to estimate the likelihood of occurrence of a non-silent SNP in CEN.PK113-7D or S288C. Specifically, the top 10 homologous protein sequences based on Pfam search and multiple alignments were determined. Each amino acid the SNP position identified was catalogued and characterized. Specifically, indices referred to as the *S288C Frequency*, *CEN.PK Frequency*, and *Conservation Distance* was calculated. The *Conservation Distance*, bound from -1 to 1, is a convenient measure of whether the nucleotide detected in CEN.PK113-7D or S288C is dominant compared to homologous sequences. Resulting amino acids were then characterized according to their physical chemistry properties.

## Conservation Distance

Conservation Distance Average  $\pm$  SD =  $0.03 \pm 0.40$



**Supplementary Materials Figure 7.** The *Conservation Distance*, previously described in Supplementary Materials Figure 6, is plotted for all 210 non-silent SNPs.



**Supplementary Materials Figure 8.** The gene *ERG8* of the ergosterol synthesis pathway contains a total of four non-silent SNPs, two of which, located at nucleotide positions 49 and 247, are analyzed here. The top plots show the CEN.PK Match Frequency, Dominant AA Frequency, S288C Match Frequency, and Conversation Distance. The middle plots show the frequency (fraction) of each categorization across the amino acid sequences resulting from Pfam multi-sequence alignment. The bottom plots shows the characterization of the original S288C amino acid (symbol: red bar) and the CEN.PK113-7D amino acid (symbol: blue bar).

**Supplementary Materials Table 3.**

Genome Sequencing Parameter	S288C	CEN.PK113-7D
<i>EDENA Software Determination for de novo Assembly (Secondary Check of SNP Detection)</i>		
<b>Reads Base Length (bp)</b>	35	35
<b>No. of Reads</b>	5,301,907	6,603,200
<b>No. of Unique Reads (Aligned)</b>	4,387,286	5,045,108
<b>No. of Contigs</b>	12,775	16,436
<b>Total Base Length (bp)</b>	4,440,488	7,326,814
<b>Average Contig Length (bp)</b>	345	446
<b>Maximum Contig Length (bp)</b>	2,031	2,734
<b>Genome Reference Coverage (%)</b>	36	59.9
<b>Total No. of SNPs Detected</b>	696	13,984
 <b>Metabolic SNP Detection</b>		
<i>EDENA Software Determination for de novo Assembly (Secondary Check of SNP Detection)</i>		
<b>Reads Base Length (bp)</b>	35	35
<b>No. of Reads</b>	5,301,907	6,603,200
<b>No. of Unique Reads (Aligned)</b>	4,387,286	5,054,108
<b>No. of Contigs</b>	12,775	16,436
<b>Total Base Length (bp)</b>	4,400,488	7,326,814
<b>Average Contig Length (bp)</b>	345	446
<b>Maximum Contig Length (bp)</b>	2,031	2,734
<b>Genome Reference Coverage (%)</b>	41.7	65.7
<b>Total No. of SNPs Detected</b>	71	1133

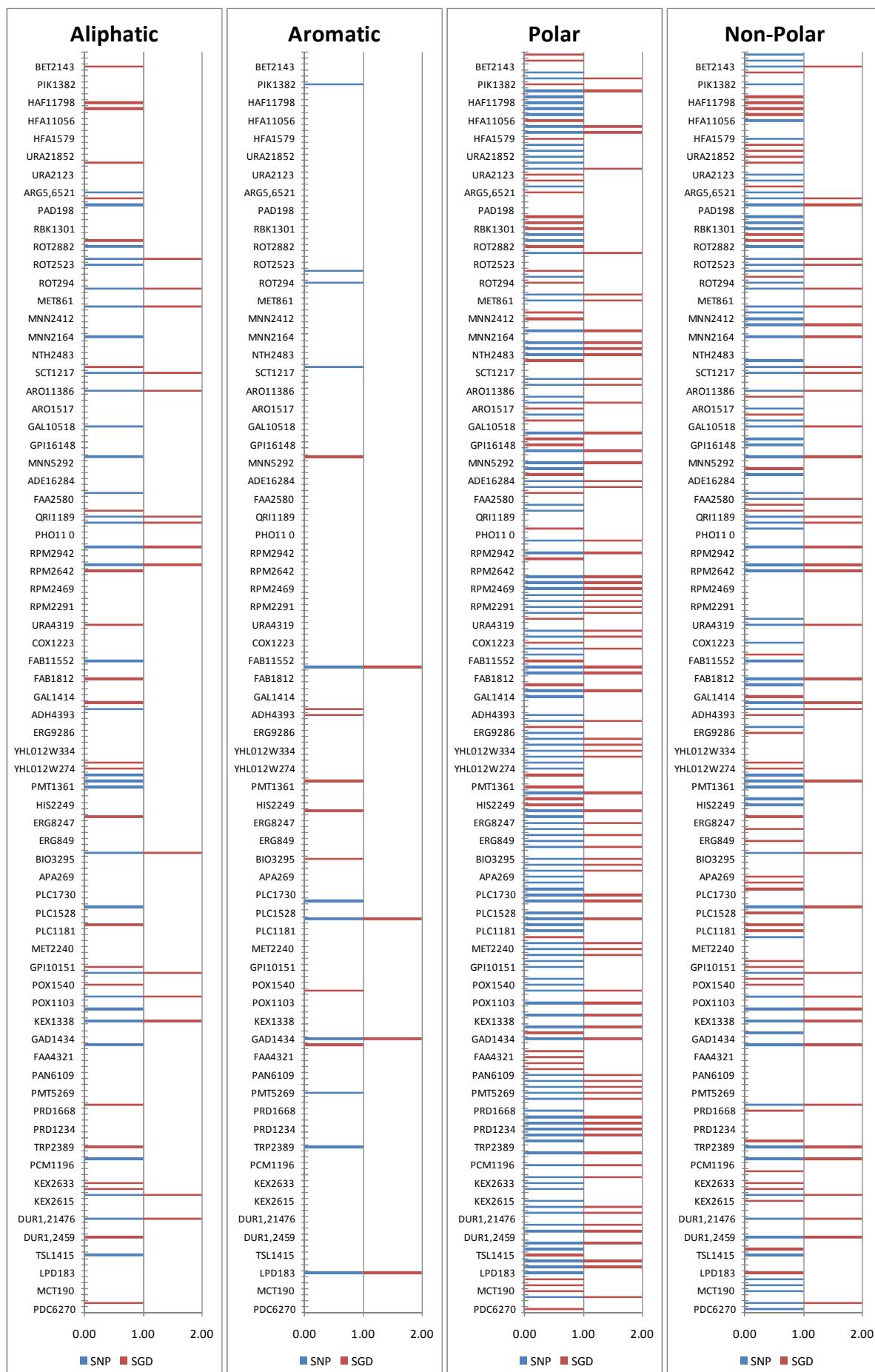
## **Supplementary Materials Figure 9**

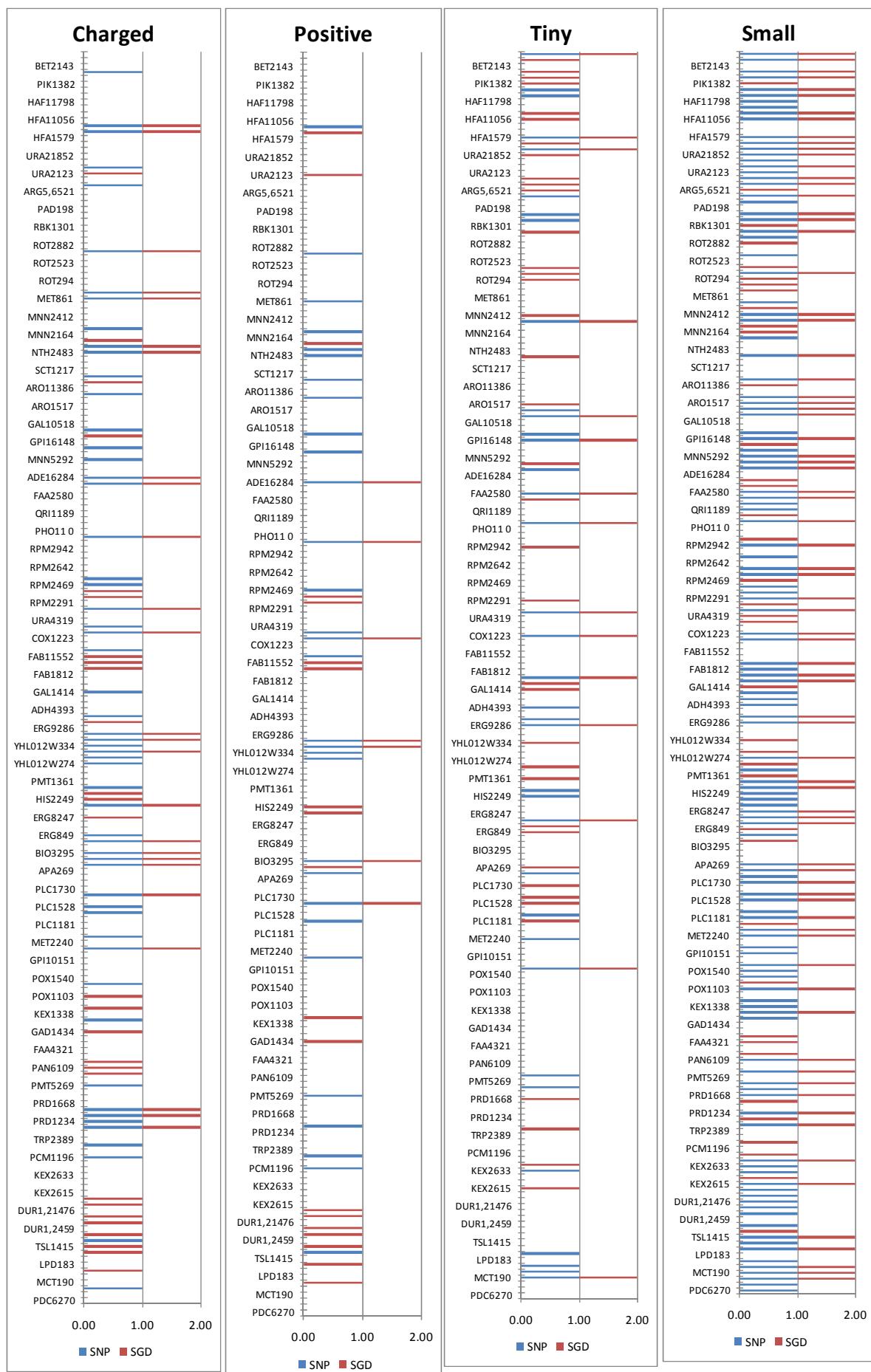
### ***Brief Description of Amino Acid Characterization – CEN.PK113-7D vs S288C Profiles***

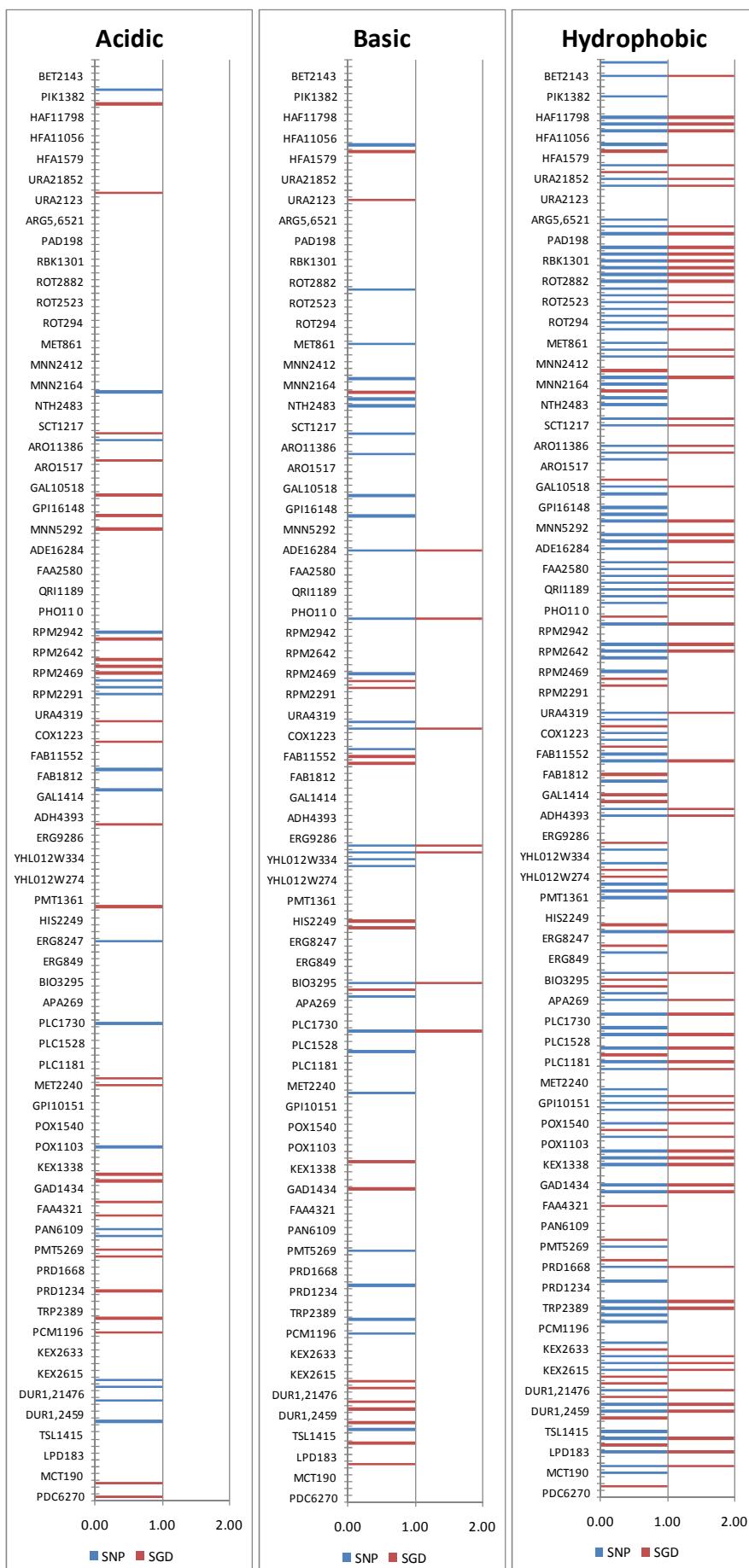
In the figures that follow, the same amino acid characterization as presented earlier was performed for only the CENPK sequence (referred to as “SNP”), and the S288c sequence (referred to as “SGD”). The scoring system was a simple binary assignment, where 0 indicated the amino acid did not fall into the category or 1 indicating that it did. Therefore, the plots that follow should have the following interpretation:

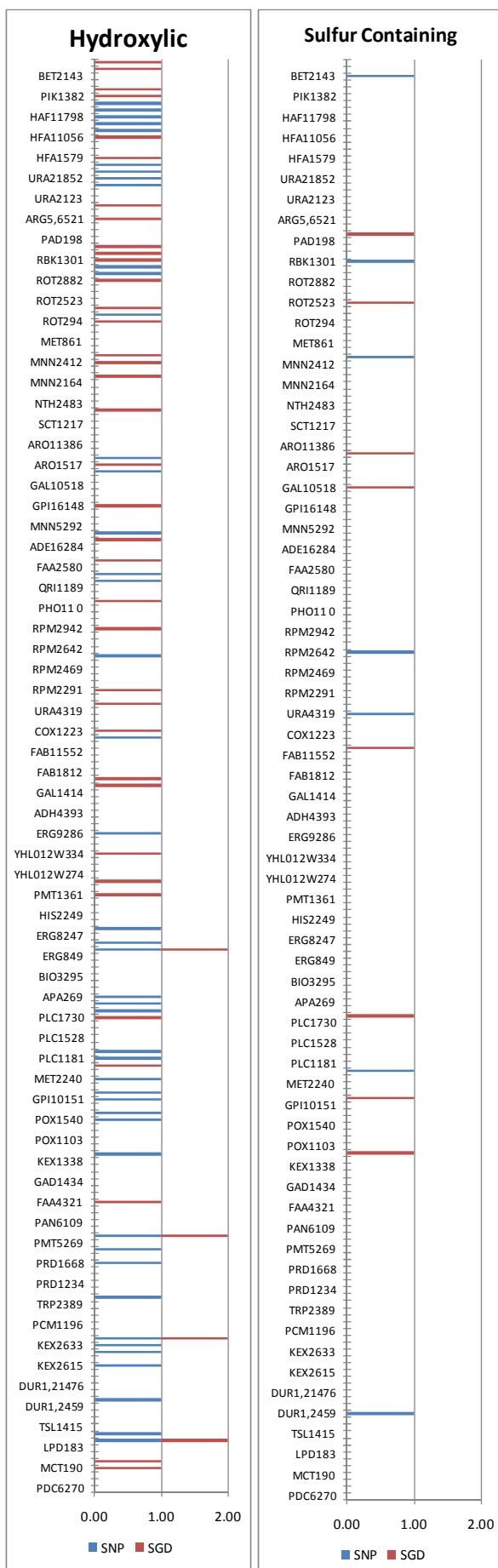
- *If the value of the individual amino acid is zero, than neither the SNP (CEN.PK113-7D) or SGD (S288C) amino acid qualify for that category*
- *If the value of the individual amino acid is 1, then only one of the sequences – SNP (CEN.PK113-7D) or SGD (S288C) – fall into the category, suggesting a change in characterization. The color codes should then be used to determine which sequence falls into the category.*
- *If the value of the individual amino acid is 2, then both sequences fall into that functional category.*

Note that on the y-axis is the individual non-silent SNPs. The naming nomenclature used is “GeneSNP”, such that for the example “BET2143”, the gene is *BET2* and the SNP is at nucleotide position 143. Due to the large number of SNPs each individual name is not included. Therefore, the plots that follow intend to provide a visual perspective for how many non-silent SNPs resulted in a functional amino acid change.









# **PAPER V**

**Industrial systems biology.**

**José Manuel Otero, Jens Nielsen**

Accepted in *Biotechnology & Bioengineering*, 2009

**Title:** Industrial Systems Biology

**Authors:** José Manuel Otero, Jens Nielsen<sup>\*</sup>

*Systems Biology, Department of Chemical and Biological Engineering, Chalmers University of Technology,  
SE-412 96 Göteborg, Sweden*

\*Corresponding author:

Professor Jens Nielsen

Phone: +46 (0) 31 772 8633

Fax: +46 (0) 31 16 0062

[nielsenj@chalmers.se](mailto:nielsenj@chalmers.se)

## **Abstract**

The chemical industry is currently undergoing a dramatic change driven by demand for developing more sustainable processes for the production of fuels, chemicals, and materials. In biotechnological processes different microorganisms can be exploited, and the large diversity of metabolic reactions represents a rich repository for the design of chemical conversion processes that lead to efficient production of desirable products. However, often microorganisms that produce a desirable product, either naturally or because they have been engineered through insertion of heterologous pathways, have low yields and productivities, and in order to establish an economically viable process it is necessary to improve the performance of the microorganism. Here metabolic engineering is the enabling technology. Through metabolic engineering the metabolic landscape of the microorganism is engineered such that there is an efficient conversion of the raw material, typically glucose, to the product of interest. This process may involve both insertion of new enzymes activities, deletion of existing enzyme activities, but often also de-regulation of existing regulatory structures operating in the cell. In order to rapidly identify the optimal metabolic engineering strategy the industry is to an increasing extent looking into the use of tools from systems biology. This involves both *x-ome* technologies such as transcriptome, proteome, metabolome, and fluxome analysis, and advanced mathematical modeling tools such as genome-scale metabolic modeling. Here we look into the history of these different techniques and review how they find application in industrial biotechnology, which will lead to what we here define as *industrial systems biology*.

## Introduction

The term “industrial biotechnology” first widely appeared in the literature in the early 1980s when genetic engineering, propelled by recombinant DNA technology, was searching for applications beyond health care and medical biotechnology [1-2]. Industrial biotechnology today represents an established field with significant government, corporate, and academic investment. Formally, industrial biotechnology is the bioconversion, either through microbial fermentation or cell-free biocatalysis, of organic feedstocks extracted from biomass or their derivatives to chemicals, materials, and/or energy. Biomass is the result of photosynthetic carbon fixation by plants to form organic polymers that may be digested, enzymatically or chemically, to carbohydrate, protein, and lipid monomers. Industrial biotechnology, often referred to as white biotechnology in Europe [3], aims to provide cost-competitive, environmentally friendly, sustainable alternatives to existing or newly proposed petrochemical processes. Processes that exploit industrial biotechnology have recently garnered increasing global attention with traditional petrochemical processing under scrutiny due to increasing raw material costs, environmental constraints, and decreasing sustainability.

Industrial biotechnology has experienced unprecedented growth with bio-based production processes representing 5% of the total chemical production sales volume. By 2010, several studies have estimated that the total fraction will increase to 20%, representing \$310 billion of a projected total sales volume of \$1,600 billion. Industrial biotechnology will continue to capture significant sales volume percentages in the arenas of basic chemicals and commodities (2 to 15%), specialty and added-value chemicals (2 to 20%), and polymers (1 to 15%). However, the greatest percentage gain is likely to occur in the fine chemical market (16 to 60%), where industrial biotechnology platforms enable complex chemistry that are presently produced via complex synthetic or combinatorial routes [4]. Furthermore, industrial biotechnology is enabling new products including novel therapeutic agents such as polyketides, and specialty chemicals not previously identified such as the diverse polyunsaturated fatty acids and biopolymers produced by microalgae [5].

In its relatively short history, industrial biotechnology commercialization of fermentation processes for antibiotics (penicillin production by *Penicillium chrysogenum*; annual market size exceeding US \$1.5 billion), vitamins (L-ascorbic acid production by the Reichstein process and biocatalysis by *Gluconobacter oxydans*; annual market size exceeding US \$600 million), organic acids (citric acid production by *Aspergillus sp.*; annual market size exceeding US \$1.5 billion), and amino acids (L-glutamate and L-lysine production by *Corynebacterium glutamicum*; annual production exceeding 600,000 tons) are well established and successful [5]. In each of these examples, host organisms well suited for production of the target compound were naturally isolated. Furthermore, under controlled environments, random mutagenesis followed by screening, selection, and traditional bioprocess development were used to enhance production yields, titers, productivities, and robustness. Despite the fact that this method provides little to no mechanistic understanding of which specific genetic perturbations lead to improved strains so that they could be further exploited, it has proven to be commercially successful as illustrated by the more than 1,000 fold improvement in penicillin titer by *P. chrysogenum* [6].

The significant increases in research and development, and commercialization at industrial scales of biotechnological processes may be attributed to several key factors, which can be grouped into four broader factors that are important to consider in connection with development of a new bio-based process: (1) Process Economics, (2) Biotechnology Process Development, (3) Environmental Impact and (4) Sustainability and Self-Sufficiency. Each of these broad factors involve several identifiable and quantitative

drivers fueling the application of industrial biotechnology to processes previously exclusive to the petrochemical industry or for the production of new chemicals. Figure 1 outlines an overview of how each of these four factors needs to be considered and evaluated before development of any industrial biotechnology process.

This review aims to provide a historical perspective of industrial biotechnology process development, and in particular, focus on the rapid deployment of metabolic engineering and systems biology technologies that first emerged from academic research groups driven by the human health and medical biotechnology sectors. Specifically, mature, recently launched and in-development examples of products that have benefited from systems biology will be highlighted for motivation. Based on such examples, and *de novo* processes presently in proof-of concept, we here define a new term, *industrial systems biology*, acknowledging that tools established in the rapidly growing field of systems biology, often applied to metabolic engineering, are prevalent in two forms. Enterprises are reshaping existing or forming new process development groups with industrial systems biology capabilities and expertise, or they are out-sourcing process development to small, recently formed entities that specialize in industrial systems biology. Examples of such enterprises focused on providing industrial systems biology expertise to more traditional process development groups include METabolic EXplorer ([www.metabolic-explorer.com](http://www.metabolic-explorer.com), France, Founded in 1999), Genomatica ([www.genomatica.com](http://www.genomatica.com), USA, Founded in 2000), Fluxome Sciences ([www.fluxome.com](http://www.fluxome.com), Denmark, Founded in 2002), Amyris Biotechnologies ([www.amyris.com](http://www.amyris.com), USA, Founded in 2003) and Microbia Precision Engineering ([www.microbia.com](http://www.microbia.com), USA, a subsidiary of Ironwood Pharmaceuticals, formerly Microbia). Although relatively small (<50 million USD 2009 total revenue), these companies have significant collaborations with many of the major chemical manufacturing, nutraceutical, pharmaceutical, and petrochemical companies.

### **Industrial Systems Biology**

Systems biology is the quantitative analysis, often through the use of predictive mathematical models, of biological systems. Often it involves collection, analysis, and integration of whole genome scale data sets with the objective to gain a quantitative phenotypic description of the biological system. With genome sequences becoming readily available for production organisms, process development has been a benefactor of the scientific achievements in systems biology, particularly in the areas of transcriptomics, proteomics, metabolomics and fluxomics. Such developments today encompass a systems biology toolbox that may be further exploited for production of metabolic intermediates that often serve as desirable precursors in the petrochemical sector. Figure 2 provides a road-map for how industrial systems biology may be applied to microbial cell factories of interest.

The examples to be discussed will largely focus on upstream process development, with particular attention paid to the metabolic engineering strategy employed, and how functional genomics data and analysis provided clear advantages. The target products cited will draw examples from numerous fermentation organisms.

It will be of little surprise that the largest industrial biotechnology product in the world, recently garnering unprecedented corporate, social, and government support, is bioethanol. In 2008, total world production was 65.6 billion liters, with the production volume and the total number of refineries built between 2007 and 2008 in the United States increasing by 9.1 billion liters and 29, respectively [7]. Perhaps most indicative of the role systems biology and metabolic engineering is playing in biofuels is the reportedly 26 cellulosic or 2<sup>nd</sup> generation bio-ethanol pilot plants and projects in development as of 2008,

with the largest facility operated by Verenium Corporation (Cambridge, MA) producing nearly 4 million L annually [8]. *S. cerevisiae* today is the preferred bioethanol production host, amongst other industrial biotechnology products, primarily as a result of proven industrial process robustness and exceptional physiological and *x*-*omics* characterization [9-12]. The *S. cerevisiae* genome sequence, consisting of 6,607 total open reading frames (4,807 verified; 989 uncharacterized; 811 dubious) [13], was first made publicly available in 1996 largely through André Goffeau's coordination of the European yeast research community [14]. Soon thereafter, in 1997 and 1998, the first cDNA spotted microarray exploring metabolic gene regulation, and the first commercial platform (Affymetrix) microarray data exploring mitotic cell regulation were reported, respectively [15-16]. The genome sequence coupled with extensive annotation based on fundamental biochemistry, peer-review literature, and available transcription data enabled publication of the first genome-scale metabolic model for *S. cerevisiae* in 2003 [17]. The genome-scale metabolic model represents an integration of extensive amounts of data into an annotated, defined, and uniform format permitting simulations of engineered genotypes to elicit desired phenotypes [17-18].

Strain development has classically been dominated by random mutagenesis of a production host followed by screening and selection in controlled environments for a desired phenotype. Although this methodology has endured tremendous success, it has largely been end-product driven with minimal mechanistic understanding. Today, with the exponential increase in genome sequences of existing and future production hosts, coupled with tools from bioinformatics that enable integration and interrogation of *x*-*omic* data sets, it is possible to identify high-probability targeted genetic strategies to increase yield, titer, productivity, and/or robustness [19-21]. It is also now possible to perform inverse or reverse metabolic engineering, where previously successful production systems may be *x*-*omically* characterized to elucidate key metabolic pathways and control points for future rounds of targeted metabolic engineering [7, 22]. In both *forward* and *inverse* metabolic engineering, systems level models and simulations are accelerating bio-based process development, resulting in reduced time to commercialization with significantly less resource commitment.

Today, industrial biotechnologists are no longer considering singular products, but rather diverse portfolios of petrochemical commodity, added-value, high added-value, and specialty chemicals to be produced using biotechnology. The term biorefinery was first defined in 1999, when it was suggested that lignocellulosic raw materials may be converted to numerous bio-commodities via integrated unit processes, and offer competitive performance to existing petrochemical refineries [23]. If the biorefinery platform model is to evolve from academic conception to industrial reality it will require two essential driving forces. First, the economic and socio-political landscape must continue to support and warrant the significant financial investment, favorable legislative policy, and consumer driven demand that will be required. Second, the advances and tools developed within systems biology for metabolic engineering must be successfully applied in commercial environments. Several examples, such as bioethanol, have suggested that biorefineries are viable commercially; however, the diverse product streams that will be required continue to demand more sophisticated, native and non-native, multi-gene metabolic engineering approaches. These approaches may only be realized through advanced interrogation and integration of microbial metabolic space using systems biology tools.

### **Engineering Microbial Metabolism**

There have been extensive reviews regarding the application of random mutagenesis and directed evolution for novel development or enhancement of existing microbial cell factories for the production of a

wide range of industrial biotechnology products [24-29]. What is often referred to as “classical strain development” is dependent on the capability of inducing and promoting genetic diversity, under controlled laboratory conditions, in a desirable production host organism that can be selectively screened, isolated, cultured, and preserved based on a phenotypic criteria. Genetic diversity may be induced using mutagenic chemical agents, radiation, ultra-violet light exposure, intercalating agents, or through genetic recombination [24]. While resulting modified strains may then be further physiologically characterized, the specific and targeted genetic alterations that lead to the improved phenotype are not known, preventing any mechanistic understanding from being applied to future rounds of strain improvement.

First explored in prokaryotes and introduced to the extended biotechnology community in the seminal 1991 *Science* publication, “Toward a science of metabolic engineering”, authored by the late Professor James E. Bailey (1944-2001), metabolic engineering was a natural evolution from the earlier discipline of bioreaction engineering that largely focused on deterministic systems of stoichiometric relationships between mass (e.g., carbon, nitrogen) and energy (e.g., ATP, reducing equivalents) [30]. In the same 1991 *Science* issue, Stephanopoulos and Vallino specifically described the challenges of metabolic engineering applicability to the overwhelming goal at the time: metabolite overproduction [31]. As the authors originally observed, “In order to enhance the yield and productivity of metabolite production, researchers have focused almost exclusively on enzyme amplification or other modifications of the product pathway. However, overproduction of many metabolites requires significant redirection of flux distributions in the primary metabolism, which may not readily occur following product deregulation because metabolic pathways have evolved to exhibit control architectures that resist flux alterations at branch points. This problem can be addressed through the use of some general concepts of metabolic rigidity, which include a means for identifying and removing rigid branch points within an experimental framework.” Consequently, metabolic engineering is an enabling science that encompasses gene-targeted, rational, and quantitative approaches for redirection of metabolic fluxes to improve the yield, titer, productivity, and/or robustness associated with specific metabolites in a bio-reaction network.

In order to fully appreciate the advances in microbial metabolic engineering, particularly noting the relatively short time period within which this field has developed, a historical perspective is required. Figure 3 presents a time-line highlighting research milestones that formed the basis for early bio-reaction engineering, including the early identification, classification, and characterization of pathways. The figure is not all inclusive, but rather illustrates the significant steps in microbial physiology characterization, initially single enzyme function and kinetic modeling, leading to elucidation of whole pathways.

### **Impact of Genome Sequencing and Functional Genomics**

It is interesting to recall that in the late 1980s and early 1990s, with recombinant DNA technology emerging from medical biotechnology, we witnessed expression of compounds previously produced via synthetic routes now being attempted in production organisms [25, 31-33], and some may argue that there is a certain element of *déjà vu* with respect to the current expectations of industrial biotechnology. Back then, just as now the construction of new production strains was made possible by the introduction of genetic sequences encoding for enzymes that were likely to catalyze desired reactions, or, the deletion of genes that would down-regulate undesired reactions and pathways. However, a major difference from the late 1980s and now is that, although techniques that permitted manipulation of recombinant DNA existed, the annotated genome sequences of industrially relevant production hosts were not available. Furthermore, the whole arsenal of analytical techniques available to us today, much due to development of

systems biology tools in the medical field, and it was therefore not possible, as today, to perform detailed phenotypic and metabolic analysis of metabolically engineered strains. This represents a major difference from these early endeavors in metabolic engineering. One thing that clearly has driven the introduction of systems biology into the field of industrial biotechnology has been the sequencing of industrially relevant microorganisms. Figure 4 highlights the exponential increase in published genome sequences that first started in 1995 and have continued to expand through 2009 with a total of 992 published genome sequences, and 2523, 1029, and 96 bacterial, eukaryotic, and archaeal sequence projects on-going, respectively. This genomic revolution was mainly driven by the medical research field, as illustrated in Table 1, which presents characteristics of those genomes sequenced between 1995 and 1999. It is seen that of the twenty-four sequences made available, only three could be considered to have broad applicability to the industrial biotechnology sector: *Saccharomyces cerevisiae*, *Escherichia coli*, and *Bacillus subtilis*, while the rest were driven by the medical community. One can even argue that sequencing of these three genomes was also mainly motivated by their medical relevance, either as a eukaryote model organism, pathogen or model pathogen. If we move beyond 1999, many more industrially important cell factories have been genome sequenced, and with the substantial reduction in sequencing costs genome sequencing has even become a tool to analyze cell factories with different phenotypes. The presence of complete genome sequences has clearly allowed better targeting of genetic modifications, and information about the complete parts lists of a given cell factory is extremely valuable.

With genome sequences for several industrial model organisms in hand, it was the annotation of those sequences that bridged the gap between expanding knowledge-based databases (e.g., genome sequence collections) and the data-driven databases (e.g., application of the genome sequences for annotation, model development, and further understanding) [34]. The annotation of genome sequences has evolved into a well-defined discipline referred to as functional genomics, which focused on developing numerous experimental and theoretical tools for determination of gene function [35]. Functional genomics, through linking gene products (e.g., enzymes) to gene functions (e.g., reaction stoichiometry) has permitted the development of genome-scale models for various data types, such as reconstructed metabolic network models.

### Metabolic Models

Even though genome sequencing has clearly facilitated the use of targeted genetic modifications for construction of cell factories with desirable phenotypes, the major value contributed by genome-sequences is that they provide a parts list, which can further be used for reconstruction of genome-scale metabolic models (GSMMs). To develop a model of cellular metabolism that enables the prediction of concentration profiles as functions of time, the stoichiometry and kinetic reaction rates for each biochemical reaction in a cell at physiological conditions would be required. At present, this information is not available, neither via estimation nor via experimental measurement. Through careful annotation based on existing biochemical knowledge, literature review, and experimentation; however, it is possible to associate known genes with known biochemical reactions and their corresponding stoichiometry. The result is a biochemical model describing the formation and depletion of each metabolite that by providing mass-balance boundary conditions makes possible constraint based simulations of how the metabolic network operates at different conditions. In simpler terms, using basic stoichiometry these models can be used to predict the relationships between genes with function in the metabolic network operating in a cell. It is hereby also clear that GSMMs can be used to predict a theoretical landscape of genetic perturbations that can

maximize product and biomass formation, even under different growth conditions (i.e., growth on alternative carbon sources). GSMMs have been developed for several model production organisms (see Table 2), and these models represent a major step in not only allowing model guided metabolic engineering, but also integration of different *x-ome* data for obtaining detailed metabolic characterization. GSSMs provide an appropriate scaffold for further expansion and data integration, owing to its easily manipulated mathematical framework.

There are numerous reviews describing the process of genome-scale network reconstruction, including the initial biochemical annotation performed, the mathematical framework employed for describing metabolism, the resulting system of linear differential equations, the assumptions and constraints required for simplification, and ultimately numerical solution methods [17, 36-38]. The simplified mathematical framework presented here has been adapted from an excellent presentation of flux balance analysis [39]. To better motivate the power of this methodology, let's define a hypothetical metabolic system, composed of unique metabolites *A*, *B*, *C*, *D*, and *E*. Let us also define a two compartment biochemical reaction space (compartments 1 and 2). The resulting metabolic space may be pictorially represented by Figure 5. The reactions and stoichiometry are clearly defined, and included in the stoichiometry is annotation of the compartmentalization. Each of these reactions, in a genome-scale network reconstruction would be further annotated by assigning function to a specific open reading frame (ORF), and subsequently a comprehensive list of all reactions, metabolites, and their assigned ORF are reconstructed, including identifying those reactions and metabolites that are unique (e.g., independent of compartmentalization, and representing novel chemical entities and their catabolic or anabolic reactions). The methodology then employed is derived from the classical principles of chemical engineering, where essentially a mass balance is performed across a defined system boundary, i.e.

$$\frac{dX}{dt} = S \cdot V \pm V_{trans} \quad \text{Equation 1}$$

Equation 1 represents a mass balance across all metabolites in the biochemical reaction space considered, having concentration  $X$ , and then defining a vector of all the metabolic reactions,  $V$ , and a stoichiometric matrix,  $S$ . Biological time-scales associated with changes in metabolite concentrations are often very rapid, and significantly faster than times scales associated with growth (e.g., for *S. cerevisiae* the doubling time is about 2 hours). It is therefore reasonable to assume that the concentrations of all the intracellular metabolites are in a steady state, yielding equation 2.

$$0 = S \cdot V \pm V_{trans} \quad \text{Equation 2}$$

Equation 2 may be further simplified by considering that the rate of transport of all metabolites,  $X$ , maybe reduced to a constant value equivalent to the net transport of metabolites into or out of the bioreaction space. This simplification, converts  $V_{trans}$  to a constant term,  $b$ , a vector representing the net exchange flux of metabolites. This constant value,  $b$ , for each metabolite, in matrix format is expressed in equation 3, noting the use of the identity matrix,  $I$ .

$$0 = S \cdot V + b \cdot I \quad \text{Equation 3}$$

From this point forward, additional constraints that are often specific to the bioreaction space being considered and the organism are included. These considerations will include, but not be limited to the metabolite and reaction compartmentalization, the reversibility of the reactions, the net biomass equation (e.g., summation of all metabolite precursors, redox co-factors, and energy co-factors in stoichiometric

quantities), the theoretical minimum and maximum metabolite fluxes, the minimum and maximum growth rates, the amount of ATP (or equivalent energy currency) required for maintenance, and the amount of starting fluxes for input exchange fluxes (e.g., glucose uptake rate). Lastly, an objective function, to be maximized or minimized, must be defined and typically takes the form of equation 4, where  $Z$  is an objective function equal to the summation of the product of a unit vector,  $\mathbf{q}_i$ , and the metabolic fluxes,  $\mathbf{V}_i$ , where  $\mathbf{q}_i$  is typically the growth rate flux or glucose uptake rate. Both of these fluxes serve as suitable maximization parameters for modeling *in vivo* microbial metabolism where under conditions of excess nutrients and limited substrate, the specific growth rate of microbes,  $\mu$ , will approach  $\mu_{\max}$ . Note, that included in equation 4 are the constraints on metabolites,  $\mathbf{V}_i$ , which typically range from a minimum ( $a$ ) to a maximum ( $b$ ).

$$Z = \sum q_i \cdot V_i \\ a \leq V_i \leq b \quad \text{Equation 4}$$

The resulting system of linear equations, for a given objective function, may be solved using linear programming methods, for which several numerical solution packages are available. The result is a solution space, which may be represented as a cone in the multidimensional flux space. Figure 6 highlights what is commonly referred to as the phenotypic phase plane (PhPP), where a two-dimension or three-dimension solution space is considered for a simulation where the maximization of an objective function is considered under specific constraints, such as the optimization of growth rate under a constant glucose uptake rate ( $q_{\text{gluc}}$ ), and oxygen uptake rate ( $q_{O_2}$ ).

The approach described here in flux balance analysis has been applied to numerous organisms as described in Table 2, and in particular, has served as a critical tool in metabolic engineering approaches, and more recently, systems biology. At the core of systems biology is the transformation of quantitative, typically large-scale data sets, into *in silico* models that provide both interpretation and prediction. GSMMs provide a framework of how *x-ome* data may be organized and overlaid on the metabolic network. As technologies have become more accessible for transcriptome (DNA oligonucleotide and cDNA microarrays), proteome (Two-dimensional gel electrophoresis coupled to MS or direct MS analysis), fluxome (isotopically labeled substrates coupled to detection by GC-MS), and metabolome (numerous analytical methods including LC-MS and GC-MS) measurements, enormous data sets have been generated that require bioinformatics and quantitative models to be developed for data analysis, interpretation, and prediction. Industrial biotechnology is beginning to exploit the benefits of these tools realizing that metabolic engineering strategies for improved process development may first be screened *in silico*, producing a reduced, specific, and high-probability of success list of genetic perturbations that should be experimentally validated. The process is highly iterative, with strain construction and characterization providing new *x-ome* data that can be used to improve the models (i.e., experimental quality control of *in silico*) and metabolic engineering strategies (See Figure 2).

### **Industrial Systems Biology: Case Stories**

There have been extensive reviews and literature describing industrial biotechnology, noting several prominent case studies [40-44]. Although several products may be presented as case studies, perhaps a more appropriate context to gauge the impact of industrial systems biology is to consider three broader product classes, providing a key example in each. Industrial biotechnology products may be categorized into the following cross-sections: mature and developed, recently launched and rapidly growing, and in-

development. Products representing each cross-section include bioethanol, 1,3-propanediol, and succinic acid, respectively. All three of these products are considered leading examples within their classes, and all three have been significantly impacted from the application of systems biology for development of commercialized microbial cell factories.

### ***Mature and Developed: Bioethanol***

As highlighted previously, the largest industrial biotechnology product in the world, both in terms of volumetric production and total sales, is bioethanol, which continues to endure phenomenal expansion fueled by unprecedented corporate, social, and political support [7]. For the foreseeable future bioethanol will continue to expand as the most mature and developed biotechnology product, with unprecedented demand for more advanced metabolic engineering strategies and application of systems biology tools to enhance all segments of bioethanol process development.

The producers of bioethanol are using a variety of fermentation platforms; however, *S. cerevisiae* is among the more popular serving as a credit to its robustness for large-scale (>300,000L) fermentation processes. The use of metabolic models for optimization of bioethanol production in *S. cerevisiae* has been demonstrated in two studies. A simple metabolic model was used to identify the deletion of NADPH-dependent glutamate dehydrogenase and overexpression of the NADH-dependent glutamate dehydrogenase, which resulted in increased ethanol production coupled with a 40% reduction in the production of the by-product glycerol [45]. In a second study a genome-scale metabolic model was used to identify a new target for improving bioethanol production by insertion of an NADPH-forming glyceraldehyde dehydrogenase. This resulted in increased bioethanol with reduced glycerol formation [46]. In both of these examples there was performed modification of the redox metabolism, and this has in general shown effective for improving many bioprocesses as the redox co-factors NADH and NADPH participate in a larger number of reactions.

With the experimental mechanics of collecting transcriptome becoming common place, attention and focus is now placed on data analysis methods and integration with other *x-ome* data sets. In another example, the topology of the genome-scale metabolic model constructed for *S. cerevisiae* is examined by correlating transcriptional data with metabolism. Specifically, an algorithm was developed enabling the identification of metabolites around which the most significant transcriptional changes occur (referred to as reporter metabolites) [47]. Due to the highly connected and integrated nature of metabolism, genetic or environmental perturbations introduced at a given genetic locus will affect specific metabolites and then propagate throughout the metabolic network. Using transcriptome experimental data, predictions *a priori* of which metabolites are likely to be affected can be made, and serve as rational targets for additional inspection and metabolic engineering. This algorithm has been recently extended to include reporter reactions, whereby metabolome data is correlated with the metabolic reactions of the reconstructed *S. cerevisiae* genome-scale metabolic network model to identify those reactions around which a genetic or environmental perturbation confer metabolite changes [48].

There have been several examples where flux measurements and analysis has significantly contributed to bioethanol strain development, particularly with respect to engineering xylose and pentose consuming fermentations. As highlighted earlier, bioethanol conversion from cellulosic biomass sources will be critical to meeting renewable fuel standards, and xylose represents the most abundant pentose sugar in hemicellulose, hardwoods and crop residues, and the second most abundant monosaccharide after glucose [49]. *S. cerevisiae* fails to consume pentose sugars efficiently, compared to glucose, and therefore

significant research has occurred in metabolically engineering such strains. For example, Grotkær et al (2005) compared the flux profile of two recombinant *S. cerevisiae* strains, TMB3001 and CPB.CR4, both expressing xylose reductase (XR) and xylitol dehydrogenase (XDH) from *P. stipitis*, and the native xylulokinase (XK), but CPB.CR4 included a *GDH1* deletion and *GDH2* being put under a *PGK* promoter [50]. Expression of XR, XDH, and XK lead to highly inefficient xylose utilization due to a co-factor imbalance, where excess NADH must be regenerated via xylitol production resulting in reduced ethanol yield. Therefore, metabolic engineering of the ammonium assimilation through deletion of the NADPH-dependent glutamate dehydrogenase (*GDH1*) and over-expression of the NADH-dependent glutamate dehydrogenase (*GDH2*) resulted in a 16% higher ethanol yield due to a 44% xylitol reduction [50-51]. Using a reverse metabolic engineering approach, metabolic flux analysis was used to characterize the intracellular fluxes for both strains based on experimental data of anaerobic continuous cultivations using a growth limited feed of <sup>13</sup>C labeled glucose, confirming that XR activity shifted from being mostly NADPH to partly NADH dependent in the CPB.CR4 strain. Furthermore, the analysis revealed, unexpectedly, activation of the glyoxylate cycle in CPB.CR4 generating the question of whether glyoxylate cycle activation may be preferred for ethanol yield. It was only through flux measurements and analysis, based on a reduced reconstructed metabolic network that the distribution of carbon believed to have been altered via targeted genetics could be confirmed.

The examples of where various *x-ome* technologies have been integrated with genome-scale reconstructions of microbial metabolism to elucidate previously poorly understood phenotypes, or for *de novo* prediction of metabolic engineering strategies in the context of bioethanol process development, are extensive [7, 52].

#### **Recently Launched: 1,3-Propanediol**

An example of a product recently launched, previously produced via petrochemical conversion and now made possible by industrial biotechnology, is 1,3-propanediol (PDO), produced by DuPont's new technology platform, DuPont Bio-Based Materials, and called Bio-PDO™. PDO is a critical intermediate in the production of polymers composed of terephthalic acid and PDO, commonly used for apparel, fiber, and carpet industries, and serves as an intermediate for DuPont's new polymer platform, Sorona®. DuPont has partnered with Tate & Lyle PLC to produce Bio-PDO™ using a proprietary fermentation platform based on *E.coli* conversion of D-glucose to dihydroxyacetone phosphate (DHAP), further to glycerol, and finally to PDO. The development of the microbial cell factory was completed in collaboration with Genencor International. The manufacturing facility was completed in 2006 with production beginning in November 2006, and using corn as the principle feedstock it will produce 45 million kg/year at full capacity [53]. A life-cycle assessment of the production of nylon-6 polymer versus the production of renewably sourced Sorona® with Bio-PDO™ results in 30% less energy usage and 63% less greenhouse gas emissions (including bio-based content stored in the product) [54]. The estimated demand for polymers composed of PDO is 500,000 to 1,000,000 tons per annum [55].

PDO is amongst the oldest fermentation products, first identified by August Freund in 1881 in a glycerol-fermenting mixed culture of *Clostridium pasteurianum* [56], and later quantitatively analyzed at the microbiology school of Delft [57] and continued at Ames, Iowa [58]. In native PDO producing organisms (*Citrobacter*, *Clostridium*, *Enterobacter*, *Klebsiella*, and *Lactobacillus* species) briefly mentioned, PDO formation is the result of anaerobic fermentation of glycerol where excess reducing equivalents in the form of NADH are regenerated (NAD<sup>+</sup>) via glycerol dehydratase (*dhaB1-B3*) activity followed by 1,3-propanediol

oxidoreductase (*dhaT*) activity [55]. Owing to the significant cost-benefit of utilizing glucose feedstocks, a metabolic engineering strategy requiring the heterologous expression of pathways forcing carbon flux redirection from DHAP to 1,3-propanediol. The key genetic modifications implemented in an *E.coli* microbial scaffold, exhibiting no accumulation of PDO and relatively little accumulation of glycerol (as compared to acetate or succinate), as detailed in Nakamura, et al., 2003, included:

- Over-expression of the non-native (*S. cerevisiae*) glycerol-3-phosphate dehydrogenase (DAR1, alias GPD1) and glycerol-3-phosphate phosphatase (GPP2), for glycerol accumulation.
- Over-expression of the non-native (*K. pneumonia*) glycerol dehydratase (*dhaB1*, *dhaB2*, *dhaB3*) and the B<sub>12</sub> reactivating factors (*dhaBX*, *orfX*) for conversion of glycerol to 3-hydroxypopionaldehyde.
- Over-expression of the native oxidoreductase (*yghD*) to complete the conversion of 3-hydroxypopionaldehyde to PDO; however, using NADPH as opposed to NADH.
- Deletion of the native glycerol kinase (*glpK*), glycerol dehydrogenase (*gldA*), and triosephosphate isomerase (*tpi*) were essential in ensuring a maximized carbon flux from glucose to DHAP, with minimal reversion of carbon back to glyceraldehydes-3-phosphate and consequently, the TCA cycle and respiration.
- Down-regulation of the native phosphotransferase system (PTS) and glyceraldehydes dehydrogenase (*gap*), the former replaced with an exclusively ATP dependent phosphorylation (elimination of phosphoenolpyruvate dependence) using the native galactose permease (*galP*) and glucokinase (*glk*), and the later augmented with a reconstituted native *tpi*. The reconstitution of *tpi* in the background of *gap* deletion provides a useful flux control point.

The integration of the above modifications, coupled with additional strategies that remain proprietary, resulted in an *E. coli* under fed-batch conditions capable of producing PDO with a final yield of 51% (w-PDO/w-glucose), titer of 135 g/L, and productivity of 3.5 g/L/h [55].

Bio-PDO™ is amongst the first success stories for metabolic engineering and industrial biotechnology in the added value chemical industry. Yet, it should be realized that greater than 10 years of development and significant resources were invested to reach this milestone. Furthermore, based on the available literature and conference presentations, the process described here required an enormous development of recombinant DNA technology, enzyme characterization and profiling, and classical bioreaction pathway analysis that was unavailable during development. Consequently, many of the systems biology tools, unavailable at the time, were not employed during the metabolic engineering strategy design and implementation.

Today, metabolic modeling of microbial metabolism is being applied to the Bio-PDO™ to confirm the expected phenotype of intracellular flux distributions, and identify potential opportunities for second generation metabolic engineering strategies. Specifically, metabolic flux analysis (MFA) was developed to enable dynamic measurement of intracellular flux distributions using isotopically labeled [1-<sup>13</sup>C]-glucose supplemented to fed-batch fermentations of *E.coli* K12 (strain over-producing Bio-PDO™) in a ratio of 3:1 naturally enriched [U-<sup>13</sup>C]-glucose. A detailed metabolic reconstructed network of *E. coli* metabolism was completed, and included 75 reactions and 74 metabolites, encompassing 5 substrates (i.e., glucose, citrate, O<sub>2</sub>, NH<sub>3</sub>, SO<sub>4</sub>), 5 products (i.e., PDO, biomass, CO<sub>2</sub>, acetate, and ATP), and 63 balanced intracellular metabolites [59]. However, previously developed MFA suffers from the limitation of assuming an isotopic steady state owing to the relatively short time-scales of intracellular metabolite concentration pools. In the research of Antoniewicz, et al., 2007, the authors extended the scope of flux resolution from steady-state to dynamic environments through a modeling strategy that employed derivatives of isotopomer spectral

analysis, and classical MFA based on the mass isotopomer distributions of amino acids determined by elementary metabolite unit (EMU) modeling [59-60]. For the first time, the time-profile *in vivo* fluxes of the fed-batch industrial process for Bio-PDO™ production were resolved, consisting of 82 redundant measurements across 20 distinct time-points. Intracellular flux distributions were found to change extensively over the course of the fed-batch profile, noting a decrease in the split ratio between glycolysis and the pentose phosphate pathway (PPP) of 70/30 at 20h to 50/50 at 43h, and a decrease in the flux from glyceraldehydes-3-phosphate to 3-phosphoglycerate of approximately 21% across the same time interval. The flux from DHAP to glycerol-3-phosphate and ultimately PDO increased by approximately 10% and remained relatively constant at 132. However, the efflux of PDO had a large variation increasing from 78 (18.6h) to 138 (28.6h) and then decreased to 130 (40.7h) (all flux values were normalized to the glucose uptake rate of 100). During the same time interval the primary energy producing pathway, the TCA flux, remained relatively constant at 46. Metabolic modeling, in the context of flux estimations and a reconstructed metabolic network, provide verification that the genetic engineering believed to confer a desired phenotype does so. Furthermore, these data provide an *in vivo* opportunity to assess opportunities for further metabolic engineering, such as targeting the discrepancy between the efflux of PDO compared to the intracellular PDO formation flux. Yet another potential metabolic engineering target is that the metabolic model was incapable of accounting for all of the net ATP produced (176 excess ATP flux for P/O ratio of 3, and 123 excess ATP flux with a P/O ratio of 2) compared to the ATP consuming reactions. Given the modifications made to the *E. coli* K12 PTS and the fact that the TCA cycle flux, the primary source of ATP under aerobic conditions, remained constant, there may be opportunities to redirect excess ATP to higher biomass formation, consequently increasing the productivity of PDO.

PDO production using industrial biotechnology will continue to accelerate, mature, and it's feasible that PDO will reach commodity chemical status as market demand and sustainable, cost-effective supply, increase. Furthermore, it should be noted that microbial metabolic modeling, primarily in the form of MFA, is also being pursued in alternative production organisms, such as *Klebsiella pneumoniae* [61-65].

#### ***In Development: Succinic Acid***

In 2004, based upon a critical analysis to identify the top building blocks that may be produced from biomass and subsequently converted to high-value bio-based chemicals, the US Department of Energy identified succinic acid ( $C_4H_2O_2(OH)_2$ ) as a top ten building block [66]. In the same year, 160,000 tons of succinic acid were synthesized from petrochemical conversion of maleic anhydride (10% of total worldwide maleic anhydride production). If bio-based succinic acid production becomes more commonplace global market demand is estimated to increase to 2 billion USD per annum with a total energy savings of 2,872,096 MWh/year [67]. Succinic acid is used in a variety of products and serves as a critical starting material or intermediate in the production of useful chemicals for solvents and polymers.

There are numerous biomass based production platforms, all prokaryotic, including *Anaerobiospirillum succiniciproducens*, *Actinobacillus succinogenes*, *Succinivibrio dextrinosolvens*, *Corynebacterium glutanicum*, *Prevotella ruminocola*, a recently isolated bacterium from bovine rumen, *Mannheimia succiniciproducens*, and a metabolically engineered succinic acid over-producing *E. coli*. There have been several extensive reviews that detail the succinic acid market, and more specifically, comprehensively present the various metabolic engineering strategies coupled with application of systems biology that have been employed to date [68-71]. The two organisms that have been most significantly

engineered from native isolations are *E. coli* and *M. succiniciproducens*, and for illustrative purposes, only *M. succiniciproducens* will be highlighted as it epitomizes the potential for industrial systems biology.

*M. succiniciproducens* MBEL55E is a capnophilic (e.g., thriving in high CO<sub>2</sub> concentrations) Gram-negative bacterium first isolated in 2002 from a bovine rumen in Korea that natively accumulates large amounts of succinic acid under glucose supplemented anaerobic (100% CO<sub>2</sub>) fermentation conditions (0.68 g-succinic acid/g-glucose) [72]. Shortly following the isolation, classical batch and continuous fermentation of sodium hydroxide treated wood hydrosylate was examined and resulted in a succinic acid productivity of 1.17 g/L/h (yield: 56%) and 3.19 g/L/h (yield: 55%), respectively [73]. These were certainly the highest productivities reported at the time, and were particularly promising given the lignocellulosic feedstock used (Mixed substrate glucose and xylose, batch and continuous cultivations were also performed as controls, with similar productivities and yields resulting). In the same year, the 2,314,078 base pair genome sequence of *M. succiniciproducens* MBEL55E was reported co-currently with the genome-scale reconstructed metabolic network [74]. The genome-scale reconstructed metabolic network, consisting of 373 reactions (121 reversible and 252 irreversible) and 352 metabolites, predicted, using MFA, a theoretical production of 1.71 and 1.86 moles of succinic acid for every mole of glucose under CO<sub>2</sub> and CO<sub>2</sub>-H<sub>2</sub> atmospheres, respectively [74]. As a consequence of the simulations, they note, “Based on these findings, we now design metabolic engineering strategies for the enhanced production of succinic acid; one such strategy will be increasing the PEP carboxylation flux while decreasing the fluxes to acetic, formic, and lactic acid.” [74]. In 2006, the authors constructed a series of knock-out mutants of *M. succiniciproducens* MBEL55E that included disruption of three CO<sub>2</sub> catalyzing reactions (PEP carboxykinase, PEP carboxylase, malic enzyme) and disruption of four genes responsible for by-product formation of lactate, formate, and acetate (*ldhA*, *pflB*, *pta*, and *ackA* genes) [75]. Their results confirmed that a mutant capable of virtually no lactate, fumarate, or acetate formation was feasible, and that PEP carboxykinase was most critical for anaerobic growth and maximizing succinic acid production [75]. The resulting metabolically engineered strain, *M. succiniciproducens* LPK7 under batch fermentation conditions produced 0.97 mol succinic acid per mol glucose, and under fed-batch fermentation conditions reached a maximum titer, productivity, and yield of 52.4 g/L, 1.8 g/L/h, and 1.16 mol succinic acid per mol glucose, respectively [159]. The theoretical carbon yield of succinate under excess reducing power and CO<sub>2</sub> carboxylation, is 2 mol succinic acid per mol glucose ( $\Delta G^\circ = -317 \text{ kJ/mol}$ ) [70].

In 2006, which constituted one of the first examples of proteomics applied to industrial biotechnology process development, proteome analysis of *M. succiniciproducens* was reported [76]. Using two-dimensional electrophoresis coupled with mass spectrometry, identification and characterization of 200 proteins distributed across whole cellular proteins (129), membrane proteins (48), and secreted proteins (30), was described. Characterization of cell growth and metabolite levels in conjunction with proteome measurements during the transition from exponential to stationary growth was carried out. Two interesting conclusions could be drawn from such analysis that was not possible *a priori*. First, a gene locus previously annotated as the succinate dehydrogenase subunit A (*sdhA*) is likely to be the fumarate reductase subunit A (*frdA*) based on comparative proteome analysis supported by physiological data. Second, two novel enzymes were identified as likely metabolic engineering targets for future improvements in succinic acid production. PutA and OadA are enzymes responsible for acetate formation and conversion of oxaloacetate to pyruvate, respectively, and their deletion is likely to induce higher flux towards succinic acid through minimization of by-product formation [76]. This is a clear example of where proteome measurement and analysis not only provided novel information for future metabolic engineering

strategies, but also served as a quality-control check for two critical assumptions. First, that the genome annotation is correct, and second, that mRNA expression directly correlates with protein expression and activity.

Most recently, in 2007, an updated genome-scale reconstructed network of *M. succiniciproducens* was presented that included 686 reactions and 519 metabolites based on reannotation and validation experiments [77]. The refined reconstructed network, in conjunction with constraints-based flux analysis, was verified using comparative experimental data of the maximum specific growth rate and metabolic production formation rate for various MBEL55 mutants. In all simulation cases, the maximum specific growth rate was correctly predicted while the rate of succinic production, for a fixed glucose uptake rate, was in relatively good agreement (between 7.8 and 30.4%, depending on the genotype simulated *in vivo*). The model was further used to evaluate additional gene-deletion strategies likely to improve succinic acid production, and simulations were compared to strategies previously reported in genome-scale simulation of the *E. coli* reconstructed metabolic network [77-78]. The comparative analysis of both genome-scale model simulation results suggested that the positive effect of various gene deletions on succinic acid production was more pronounced in *M. succiniciproducens* compared to *E. coli*, and that the metabolic performance, defined as the absolute flux of succinic acid production, was higher in *M. succiniciproducens* resulting from the higher observed glucose consumption rate under anaerobic conditions [77].

In approximately five years (2002-2007), a previously unknown microbe, *M. succiniciproducens*, was transformed into a leading microbial cell factory candidate for succinic acid production, as a result of the thorough application of systems biology tools: genome sequencing, genome-scale metabolic network reconstruction, fluxomics, proteomics, and subsequent model revision. It should be noted that similar approaches for *E. coli* and *A. succiniciproducens* have been reported; however, given the relative lack of *a priori* knowledge, short development time, and diversity of *x-ome* data collected and integrated, *M. succiniciproducens* remains a prominent example of successfully applied industrial systems biology.

## Conclusions and Perspectives

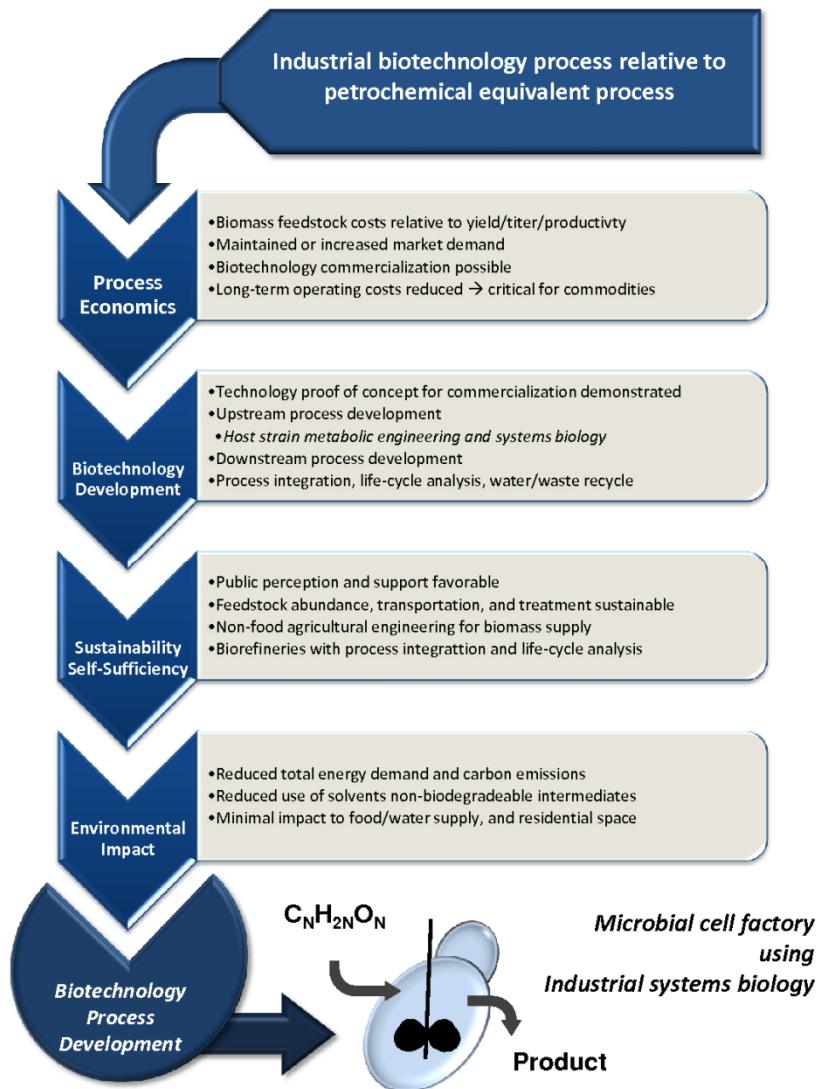
Applying a mathematical framework to microbial metabolism, beginning in earnest as early as the 1930s, has provided a scaffold for large data sets, most recently associated with the emerging field of systems biology (transcriptomics, proteomics, fluxomics, metabolomics), to be integrated, interrogated, analyzed, and ultimately, reformulated into predictive models referred to as genome-scale metabolic reconstructed networks. These networks, presently available for thirty microorganisms and growing, have offered metabolic engineers, in conjunction with accessible and easily applied recombinant DNA technology, the ability to define clear and high probability of success genetic targets for redirection of carbon flux from renewable, sustainable, and cost-effective substrates to high added-value and commodity chemical production. The construction of microbial cell factories to meet industrial biotechnology process development needs, previously relegated to classical methods of directed evolution, screening, selection, isolation, and propagation, are now being constructed faster and more efficiently through the use of systems biology toolboxes. Here then, we define a new term, *industrial systems biology*, that includes the specific application of genome-scale technologies, both experimental and *in silico*, to industrial biotechnology process development. The impact of industrial systems biology is apparent over a broad cross-section of products, which may be classified as mature and developed (e.g., bioethanol), recently launched and rapidly growing (e.g., 1,3-propanediol), and in-development (e.g., succinic acid).

Although a large number of genome-scale metabolic network reconstructions are available, what is interesting to observe is the relatively poor coverage of microbial metabolism that these reconstructions offer. A close inspection of Table 2 reveals that combined, all of the metabolic reconstructed networks have an average genome coverage of  $14.6 \pm 8.1\%$  ( $n=29$ ). If *S. cerevisiae*, the most well characterized eukaryote, is isolated as an example, the most recent metabolic reconstructed network has genome coverage of 13.6%, while 4691 of the 6608 total ORFs, 70.9%, have a verified function [13, 79]. From a more general perspective, the problem of metabolic gap closing is exacerbated by the relatively large orphan metabolic activities, where 30-40% of the known metabolic activities that are classified by the Enzyme Commission have no associated genomic sequences in any organism [80-82]. There is currently significant effort under-way to extend pathway reconstructions to regions of metabolism that are poorly understood or to a large degree, have been functionally neglected [34, 80]. Industrial biotechnology has largely focused on the production of added value and commodity chemicals; however, the largest expected growth sector is in the area of specialty and fine chemicals, where industrial biotechnology offers simpler routes for complex synthetic chemistry, or the possibility of *de novo* chemicals that may offer similar or enhanced application [4-5]. Specialty and fine chemical entities are typically present as metabolic intermediates in secondary and tertiary regions of metabolism, often poorly annotated, and rarely included in genome-scale network reconstructions. A clear example is lipid metabolism in *S. cerevisiae*, where a recent update to the existing genome-scale metabolic reconstruction, iN795, included 118 previously unreported lipid reactions relative to iND750 (See Table 3). Of those 118 lipid metabolism participating reactions, 28 were assigned to ergosterol esterification and lipid degradation – previously not represented [83].

In addition to refined annotation and extension of metabolic models into uncharted metabolic pathways, the computational methods used for predictive simulations and model analysis are rapidly improving. For example, systematic evaluation of a diverse range of objective functions, including, maximization of biomass yield, maximization of ATP yield, minimization of the overall intracellular flux, maximization of ATP yield per flux unit, maximization of biomass yield per flux unit, minimization of glucose consumption rate, minimization of the redox potential, minimization of ATP producing fluxes, maximization of ATP producing fluxes, and minimization of reaction steps, when compared to experimental *in vivo*  $^{13}\text{C}$ -determined fluxes can provide insight to which optimization function best represents the metabolic network [84]. Yet another example employed a bi-level programming framework for identification of optimal gene deletions resulting in overproduction of a desired product by stoichiometrically including a drain towards biomass formation, thereby coupling production and biomass formation. This approach, called OptKnock, revealed non-intuitive metabolic engineering strategies for succinate, lactate, and 1,3-propanediol production, and in particular, provided strategies that lend themselves to improvement via directed evolution, where growth selection and adaptation are now directly linked to growth [78]. These, and other examples, are pushing the limits of high-value non-intuitive metabolic engineering strategies that may be deciphered from genome-scale reconstructed networks [85-95].

As with any mathematical framework that incorporates large collections of diverse biological data that are constantly being investigated, updated, re-annotated, re-analyzed, and debated, clear modeling objectives must be set forth. From an industrial biotechnology perspective, focused on identifying high yielding, robust, and easy to implement non-intuitive metabolic engineering strategies, microbial metabolic modeling must continue to expand upon constraint-based stoichiometric flux balance analysis that incorporates experimental verification, and subsequent model updating and expansion. Perhaps the

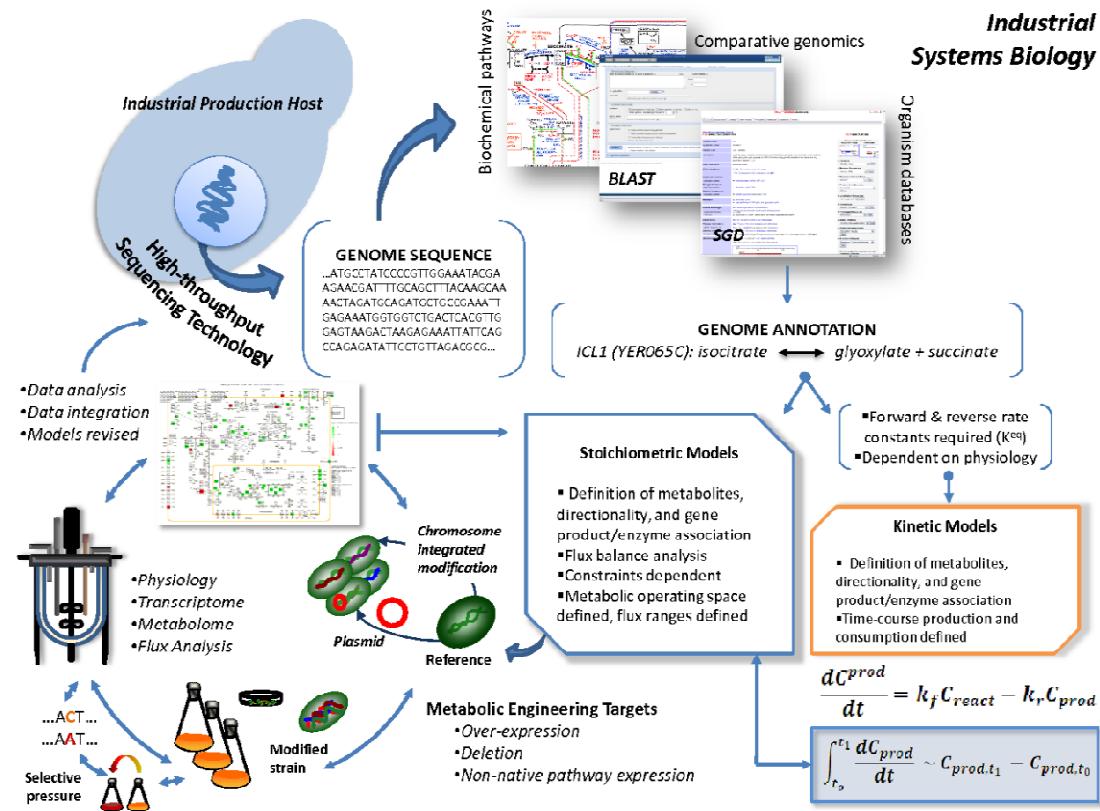
emerging availability of kinetic parameters will enable fully dynamic metabolic reconstructions to be realized in the future, but for now, the full benefits of stoichiometric metabolic modeling have yet to be realized in constructing next generation microbial cell factories. Industrial systems biology is a new approach to a challenge of epic proportions: how do we develop processes for production of chemicals, materials, and energy that is cost-effective, renewable, sustainable, scalable, and environmentally-favorable?



**Figure 1. Industrial Biotechnology Drivers**

The above figure summarizes the four key factors that are often evaluated when considering substitution of a petrochemical process with a biotechnology process, or its implementation for production of a novel chemical. Process economics, as compared to petrochemical equivalents or other benchmarking processes, are critical in establishing commercial viability, with particular focus paid to long-term operating costs. Next, biotechnology development costs, resources, and development efforts are considered, with initial analysis focused on establishing pilot-plant scale proof-of-concept. The final two factors to be critically evaluated include sustainability and self-sufficiency, and environmental impact. Sustainability and self-sufficiency not only relate to process specific considerations, such as feedstock availability, or the opportunities for further expansion through biorefinery integration, but also includes focus on public

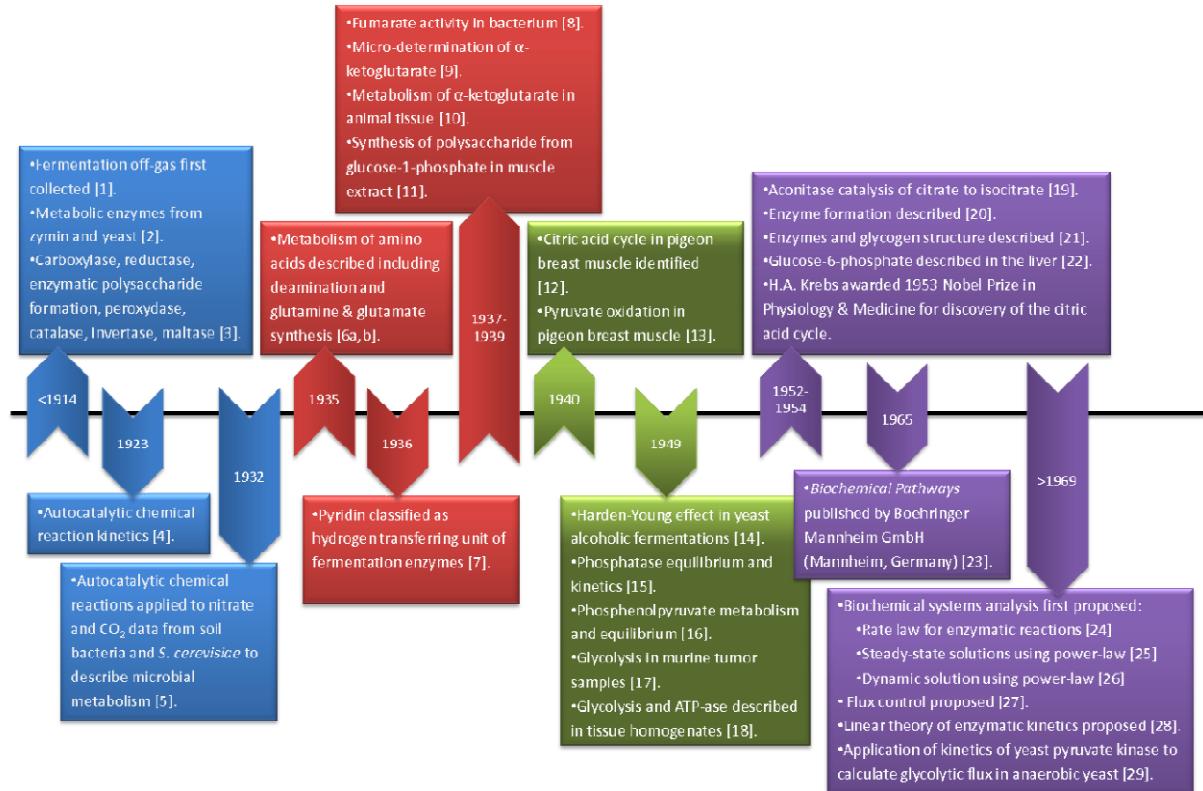
perception and the socio-political landscape. A careful consideration of these four general sectors, will ultimately determine whether proceeding with biotechnology process development is warranted or not. While not immediately obvious to most research and development scientists or engineers, it is critical to not divorce the impact these considerations may have on process development, particularly in designing strategies for construction of a microbial cell factory. It is such analysis that often defines the constraints, boundaries, targets, and viable metabolic engineering strategies, including which systems biology approaches should be exploited to experimentally demonstrate proof-of-concept.



**Figure 2. Industrial Systems Biology**

Industrial systems biology is a dynamic interaction between various disciplines and approaches. At the core is a platform technology based on a production host, for which a genome sequence is available, and subsequent annotations based on existing literature review, database query, comparative genomics, and experimental data, where available, are completed. The annotations may vary in types of functional

genomics data assigned to specific fields; however, a standard skeleton syntax structure of defining a gene, the gene product (e.g., metabolic enzyme), the metabolites serving as reactants and products (including any co-factors and intermediates), and the resulting stoichiometry is often applied. This framework, referred to as a genome-scale metabolic network reconstruction, may then be used for stoichiometric or kinetic modeling. Often, because kinetics parameters such as the forward and reverse reaction rates at physiologically relevant conditions have not been experimentally determined for a significant fraction of the network, flux balance analysis (FBA) is used for predictive modeling as it only depends on the stoichiometry and network constraints (e.g., precise stoichiometric definition of biomass, ATP maintenance terms, glucose uptake rate). Once a high-probability of success metabolic engineering strategy has been identified, often requiring gene overexpression, deletion, or non-native pathway reconstruction, genetic engineering is performed on the production host, yielding a modified strain. The modified strain is initially characterized, and may undergo directed evolution or other non-targeted approaches to yield an improved phenotype. The resulting modified strain is then characterized under well-controlled fermentation conditions, where physiological parameters, such as maximum specific growth rate, substrate consumption rates, product yields and titers, by-product formation, and morphology are determined. Furthermore, functional genomics characterization, often requiring transcriptome, proteome, metabolome, and fluxome measurements is completed. Bioinformatics, coupled with data integration, are then required for analysis of the resulting modified strain, and to identify opportunities for a second round of metabolic engineering. Furthermore, the analysis should lead to a revised model with improved predictive power that may yield promising strategies for further phenotype improvement. While this approach has often been referred to as the metabolic engineering cycle, we here compliment the traditional cycle to include integrative approaches and data sets from systems biology. Together, when applied to industrial biotechnology products, this is referred to as industrial systems biology.



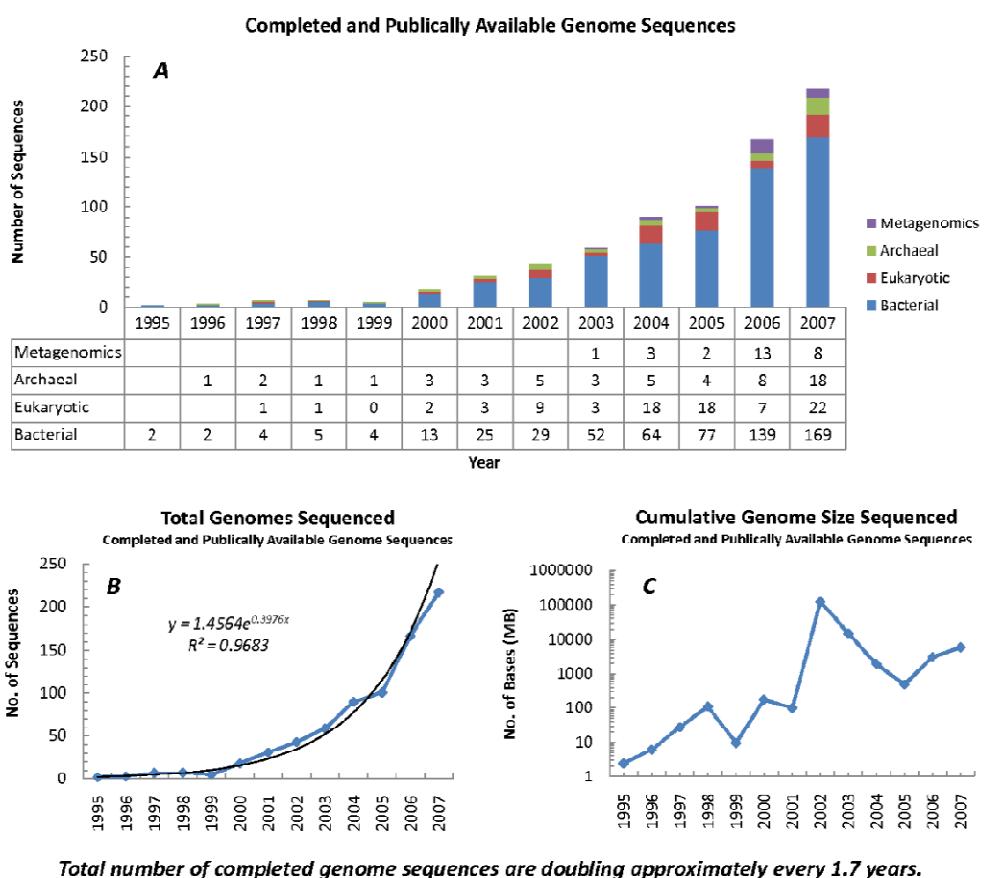
**Figure 3. Historical Time-line of Microbial Metabolism**

A historical time-line of major milestones in the field of microbial metabolism beginning in 1911 when the collection of fermentation off-gas was first described. The next approximately 60 years were dominated largely by the identification and characterization of central carbon metabolism, largely in the context of enzyme mass action kinetics, and was particularly highlighted by the seminal work of Hans A. Krebs, who was awarded the 1953 Nobel Prize in Physiology & Medicine for discovery of the citric acid cycle. The end of this time period is

dominated by the work of Michael Savageau and other groups in the development of systems analysis of biochemical processes, the broader framework for what today is commonly referred to as Biochemical Systems Theory, emerging during the 1960s through a series of seminal publications in the *Journal of Theoretical Biology*. Along similar lines two independent research groups, i.e. Kacser and Burns (1973) and Heinreich and Rappoport (1973), developed a mathematical framework for quantitative analysis of how flux control in metabolic pathways is distributed; a concept that today is referred to as Metabolic Control Analysis.

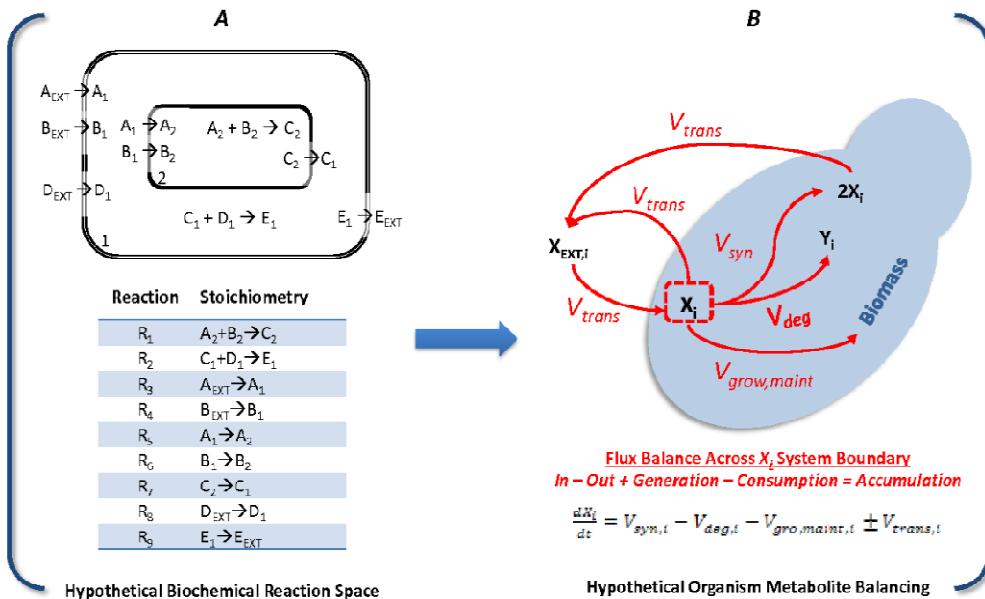
1. Harden A, Thompson J and Young WJ. 1911. Apparatus for Collecting and Measuring the Gases evolved during Fermentation. *Biochem J* 5: 230-235.
2. Harden A and Young WJ. 1913. The Enzymatic Formation of Polysaccharides by Yeast Preparations. *Biochem J* 7: 630-636.
3. Harden A and Zilva SS. 1914. The Enzymes of Washed Zymin and Dried Yeast (Lebedeff). III. Peroxydase, Catalase, Invertase and Maltase. *Biochem J* 8: 217-226.
4. Robertson TB. 1923. The Chemical Basis of Growth Senescence. J.B. Lippincott Company, Washington Square Press, Philadelphia.
5. Pulley HC and Greaves JD. 1932. An Application of the Autocatalytic Growth Curve to Microbial Metabolism. *J Bacteriol* 24: 145-168.
6. (a) Krebs HA. 1935. Metabolism of amino-acids: Deamination of amino-acids. *Biochem J* 29: 1620-1644 (a).  
(b) Krebs HA. 1935. Metabolism of amino-acids: The synthesis of glutamine from glutamic acid and ammonia, and the enzymic hydrolysis of glutamine in animal tissues. *Biochem J* 29: 1951-1969.
7. Warburg O and Christian W. 1936. Pyridin, the hydrogen-transferring component of the fermentation enzymes (pyridine nucleotide). *Biochemische Zeitschrift* 287.
8. Krebs HA. 1937. The role of fumarate in the respiration of *Bacterium coli commune*. *Biochem J* 31: 2095-2124.
9. Krebs HA. 1938. Micro-determination of alpha-ketoglutaric acid. *Biochem J* 32: 108-112.
10. Krebs HA and Cohen PP. 1939. Metabolism of alpha-ketoglutaric acid in animal tissues. *Biochem J* 33: 1895-1899.
11. Cori CF, Schmidt G and Cori GT. 1939. The synthesis of polysaccharide from glucose-1-phosphate in muscle extract. *Science* 89: 464-465.
12. Krebs HA. 1940. The citric acid cycle and the Szent-Gyorgyi cycle in pigeon breast muscle. *Biochem J* 34: 775-779.
13. Krebs HA and Eggleston LV. 1940. The oxidation of pyruvate in pigeon breast muscle. *Biochem J* 34: 442-459.
14. Meyerhof O. 1949. Further studies on the Harden-Young effect in alcoholic fermentation of yeast preparations. *J Biol Chem* 180: 575-586.
15. Meyerhof O and Green H. 1949. Synthetic action of phosphatase; equilibria of biological esters. *J Biol Chem* 178: 655-667.
16. Meyerhof O and Oesper P. 1949. The enzymatic equilibria of phospho(enol)pyruvate. *J Biol Chem* 179: 1371-1385.

17. Meyerhof O and Wilson JR. 1949. Studies on the enzymatic system of tumor glycolysis; glycolysis of free sugar in homogenates and extracts of transplanted rat sarcoma. *Arch Biochem* 21: 1-21.
18. Meyerhof O and Wilson JR. 1949. Comparative study of the glycolysis and ATP-ase activity in tissue homogenates. *Arch Biochem* 23: 246-255.
19. Krebs HA and Holzach O. 1952. The conversion of citrate into cis-aconitate and isocitrate in the presence of aconitase. *Biochem J* 52: 527-528.
20. Cohn M, Monod J, Pollock MR, Spiegelman S and Stanier RY. 1953. Terminology of enzyme formation. *Nature* 172: 1096.
21. Cori GT. 1954. Enzymes and glycogen structure in glycogenesis. *Z Kinderheilkd* 10: 38-42.
22. Cori GT and Cori CF. 1952. Glucose-6-phosphatase of the liver in glycogen storage disease. *J Biol Chem* 199: 661-667.
23. Michal G. 1999. Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, 1<sup>st</sup> ed., John Wiley & Sons Inc. and Spektrum Akademischer Verlag Co-Publication, New York.
24. Savageau MA. 1969. Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *J Theor Biol* 25: 365-369.
25. Savageau MA. 1969. Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation. *J Theor Biol* 25: 370-379.
26. Savageau MA. 1979. Biochemical systems analysis. 3. Dynamic solutions using a power-law approximation. *J Theor Biol* 26: 215-226.
27. Kacser H and Burns JA. 1973. The control of flux. *Symp Soc Exp Biol* 27: 65-104.
28. Heinrich R and Rapoport TA. 1973. Linear theory of enzymatic chains; its application for the analysis of the crossover theorem and of the glycolysis of human erythrocytes. *Acta Biol Med Ger* 31: 479-494.
29. Barwell CJ and Hess B. 1972. Application of kinetics of yeast pyruvate kinase in vitro to calculation of glycolytic flux in the anaerobic yeast cell. *Hoppe Seylers Z Physiol Chem* 353: 1178-1184.



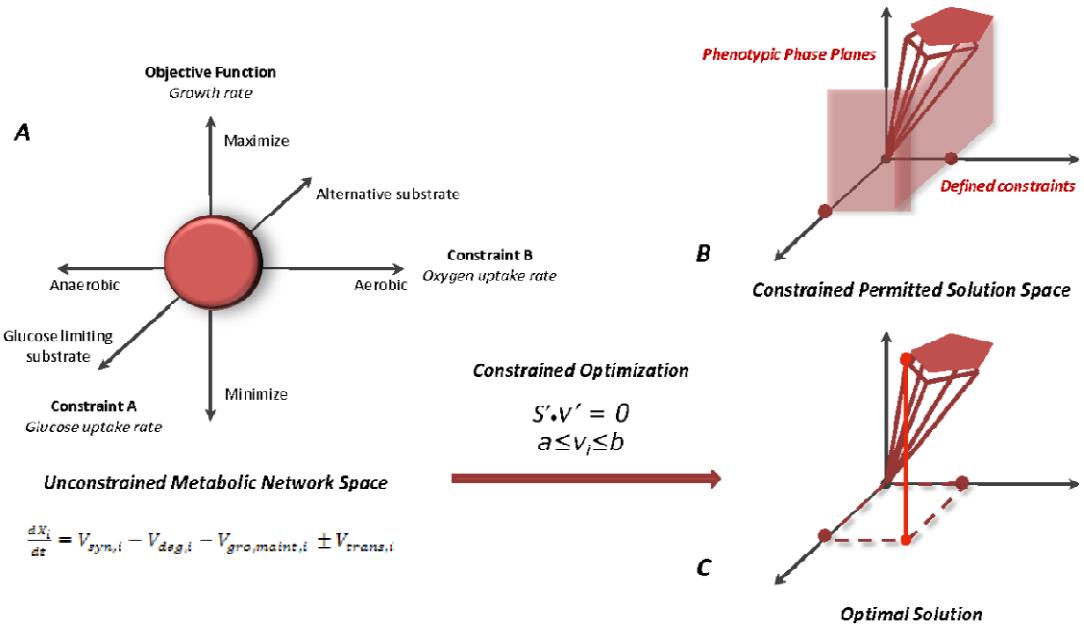
**Figure 4. Total Completed and Publically Available Genome Sequences** The above figure presents a summary of the characteristics of all completed and publically available genome sequences on a per year basis between 1995 and 2007. The data is adapted from the Genomes OnLine Database [180]. A. A plot of the number of completed genome sequences, broken down according to organism classification (Archael, Eukaryotic, Bacterial, and Metagenomics). Although an increasing number of eukaryotic and archaeal genomes and metagenomes have been sequenced, the overwhelming majority of organisms sequenced continue to be bacterial. B. The number of genome sequences has doubled approximately every 1.7 years, although data from 2006, 2007 and 2008 (not shown) suggests that this genome sequencing rate is declining. C. cumulative size of the genomes sequenced. Specifically, the size of each genome sequenced was summed across all the genomes sequenced in a given year (the coverage of each genome sequenced is not considered in this calculation). The cumulative genome size increased robustly until 1998; however, between 1999 and 2001, there was marked decrease, culminating with a large increase in 2002. Between 2002 and 2005 the cumulative size decreased significantly. Between 2005 and 2007 there has been an increase, although still below the 2002 levels. The data suggest that while there has been an increase in the number of total completed genome

sequences publically available, the size of those genomes has not been increasing. This is consistent with the observation that smaller genomes, such as those represented by metagenomes and bacterial organisms have dominated most recent sequencing efforts.



**Figure 5. Biochemical Reaction Space**

Depiction of both a hypothetical biochemical reaction space (A), and a hypothetical organism metabolite balancing (B). In panel A, a two-compartment bioreaction chemical space is suggested, with unique metabolites  $A$ ,  $B$ ,  $C$ ,  $D$ , and  $E$ , considered in a network of exchange transport reactions, catalysis reactions, and internal transport reactions. Catalysis reactions in this context are defined as catabolic or anabolic reactions, and include internal transport reactions. This may be most reasoned by noting that while  $A_1$  and  $A_2$  are chemically identical metabolites (e.g., metabolite  $A$  is considered a single unique metabolite), for modeling purposes they are considered independent metabolites, and the transport of  $A$  from compartment 1 ( $A_1$ ) to compartment 2 ( $A_2$ ) is equivalent to the conversion of  $A_1$  to  $A_2$ . Therefore,  $R_1$ ,  $R_2$ ,  $R_5$ ,  $R_6$ , and  $R_7$  are in fact considered metabolic reactions. Reactions  $R_3$ ,  $R_4$ ,  $R_8$ , and  $R_9$  are then considered exchange transport reactions, where the external metabolite ( $A_{EXT}$ ) enters the system. In panel B, the principles applied to the theoretical biochemical reaction space are then framed in the context of cellular metabolism. A flux ( $V_{n,i}$ ) balance across metabolite  $X_i$  is considered, and four categorical fluxes considered include: transport fluxes ( $V_{trans,i}$ ), synthesis fluxes ( $V_{syn,i}$ ), degradation fluxes ( $V_{deg,i}$ ), and a flux representing a depletion of metabolite  $X_i$  to satisfy growth and maintenance requirements ( $V_{gro,maint,i}$ ). These fluxes may be summed to determine the accumulation of metabolite  $X_i$  within the system boundary considered with respect to time. However, for most flux balance analysis applications the time-scales of dynamic changes in metabolite pools are often significantly faster than the time-scales associated with growth, therefore, a steady-state assumption is often applied ( $dX_i/dt = 0$ ).



**Figure 6. Flux Balance Analysis and Phenotypic Phase Planes**

Phenotypic phase planes that result from flux balance analysis that has origins in performing a steady-state mass balance of metabolites across a defined system boundary. As depicted in panel A, the resulting reconstructed metabolite network may be mapped onto two- or three-dimensions. The y-axis represents an objective function, which can be either maximized or minimized, while the other two dimensions (x-axis, z-axis) represent flux constraints. Common constraints include the glucose and oxygen uptake rates, which will create a bound solution space where the objective function, often growth rate, can be maximized as shown in panel B. The resulting phenotypic phase plane, shown in panel C, then yields an optimal solution that satisfies the optimization criteria.

**TABLE 1. Characteristics of Publicly Available Genome Sequences Published Between 1995 and 1999**

DATE	DOMAIN <sup>1</sup>	ORGANISM	RELEVANCE	SIZE(kb)	NUMBER OF ORFs	PUBLICATION
Jul-95	B	<i>Haemophilus influenzae</i>	Medical, Human Pathogen	1830	1657	<i>Science</i> 269,496-512
Oct-95	B	<i>Mycoplasma genitalium</i>	Medical, Human Pathogen, Animal Pathogen	580	477	<i>Science</i> 270,397-403
Jun-96	B	<i>Synechocystis sp.</i>	Biotechnological, Environmental, Ocean carbon cycle	3573	3172	<i>DNA Res</i> 3,109-136
Sep-96	A	<i>Methanocaldococcus jannaschii</i>	Biotechnological, Energy production	1664	1729	<i>Science</i> 273,1058-1073
Nov-96	B	<i>Mycoplasma pneumoniae</i>	Medical, Human Pathogen	816	689	<i>NAR</i> 24,4420-4449
May-97	E	<i>Saccharomyces cerevisiae</i>	Model organism	12069	5860	<i>Nature</i> 387,5-105
Aug-97	B	<i>Helicobacter pylori</i>	Medical, Human Pathogen	1667	1576	<i>Nature</i> 388,539-547
Sep-97	B	<i>Escherichia coli</i>	Medical	4639	4243	<i>Science</i> 277,1453-1474
Nov-97	A	<i>Methanothermobacter thermoautotrophicus</i>	Biotechnological, Energy production	1751	1873	<i>J.Bacteriology</i> 179,7135-7155
Nov-97	B	<i>Bacillus subtilis</i>	Biotechnological	4214	4105	<i>Nature</i> 390,249-256
Nov-97	A	<i>Archaeoglobus fulgidus</i>	Biotechnological	2178	2420	<i>Nature</i> 390,364-370
Dec-97	B	<i>Borrelia burgdorferi</i>	Medical, Human Pathogen	910	851	<i>Nature</i> 390,580-586
Mar-98	B	<i>Aquifex aeolicus</i>	Biotechnological	1551	1529	<i>Nature</i> 392,353-358
Apr-98	A	<i>Pyrococcus horikoshii(shinkaij)</i>	Biotechnological	1738	1955	<i>DNA Research</i> 5,55-76
Jun-98	B	<i>Mycobacterium tuberculosis</i>	Medical, Human Pathogen, Animal Pathogen	4411	4402	<i>Nature</i> 393,537-544
Jul-98	B	<i>Treponema pallidum pallidum</i>	Medical, Human Pathogen	1138	1036	<i>Science</i> 281,375-388
Oct-98	B	<i>Chlamydia trachomatis</i>	Medical, Human Pathogen, Animal Pathogen	1042	895	<i>Science</i> 282,754-759
Nov-98	B	<i>Rickettsia prowazekii</i>	Medical, Biothreat, Human Pathogen	1111	835	<i>Nature</i> 396,133-140
Dec-98	E	<i>Caenorhabditis elegans</i>	Model organism	100272	23209	<i>Science</i> 282,2012-2018
Jan-99	B	<i>Helicobacter pylori</i>	Medical, Human Pathogen	1643	1491	<i>Nature</i> 397,176-180
Apr-99	B	<i>Chlamydophila pneumoniae</i>	Medical, Human Pathogen	1230	1052	<i>Nat Genet</i> 21,385-389
Apr-99	A	<i>Aeropyrum pernix</i>	Biotechnological	1669	1700	<i>DNA Research</i> 6,83-101
May-99	B	<i>Thermotoga maritima</i>	Biotechnological, Energy production, Evolutionary	1860	1858	<i>Nature</i> 399,323-329

Nov-99

B

*Deinococcus radiodurans*

Environmental,  
Bioremediation

3060

2637

Science 286,1571-1577

---

**Notes:** 1. Bacterial (B), Eukaryotic (E), Archaeal (A)

**TABLE 2. Genome-Scale Metabolic Reconstructed Networks**

Organism	Genome Sequenced	Reference	Genome Dimensions		Metabolic Network Characteristics					Reference	
			Size (kB)	Total ORFs	Total Reactions (Unique) <sup>a</sup>	Total Metabolites (Unique) <sup>b</sup>	Total Genes (Enzymes) <sup>c</sup>	Percent Genome Covered <sup>d</sup>	Compartments <sup>e</sup>		
<i>Escherichia coli</i>	<i>Escherichia coli</i> K-12 MG1655	Blattner, et al, <i>Science</i> , 1997	4639	4243	(627)	(438)	660	15.6	Cytoplasm, Extracellular	iJE660	Edwards, et al, <i>PNAS</i> , 2000
	<i>Escherichia coli</i> K-12 MG1655	Blattner, et al, <i>Science</i> , 1997	4639	4243	(931)	(625)	904	21.3	Cytoplasm, Extracellular	iJR904	Reed, et al, <i>Genome Biology</i> , 2003
	<i>Escherichia coli</i> K-12 MG1655	Blattner, et al, <i>Science</i> , 1997	4639	4243	2077 (1339)	1668 (1039)	1260	29.7	Cytoplasm, Periplasm, Extracellular	iAF1260	Feist, et al, <i>Molecular Systems Biology</i> , 2007
<i>Saccharomyces cerevisiae</i>	<i>Saccharomyces cerevisiae</i> S288C	Goffeau, et al, <i>Nature</i> , 1997	12069	5860	1175 (842)	584	708	12.1	Cytoplasm, Mitochondria, Extracellular	iFF708	Förster, et al, <i>Genome Research</i> , 2003
	<i>Saccharomyces cerevisiae</i> S288C	Goffeau, et al, <i>Nature</i> , 1997	12069	5860	1489 (1149)	646	750	12.8	Cytoplasm, Mitochondria, Peroxisome, Nucleus, Endoplasmic reticulum, Golgi apparatus, Vacuole, Extracellular	iND750	Duarte, et al, <i>Genome Research</i> , 2004
	<i>Saccharomyces cerevisiae</i> S288C	Goffeau, et al, <i>Nature</i> , 1997	12069	5860	1038	636	672	11.5	Cytoplasm, Mitochondria, Extracellular	iLL672	Blank and Kuepfer, et al, <i>Genome Biology</i> , 2005
	<i>Saccharomyces cerevisiae</i> S288C	Goffeau, et al, <i>Nature</i> , 1997	12069	5860	1431	1013	795	13.6	Cytoplasm, Mitochondria, Extracellular	iIN795	Nookaew, et al, <i>BMC Syst Biol</i> , 2008

	<i>Saccharomyces cerevisiae</i> S288C	Goffeau, et al, <i>Nature</i> , 1997	12069	5860	1761	1168	832	14.2	15 compartments	NA	Herrgard, et al, <i>Nat Biotechnol</i> , 2008
<i>Haemophilus influenzae</i>	<i>Haemophilus influenzae</i>	Fleischmann, et al, <i>Science</i> , 1995	1830	1657	461	451	400	24.1	Cytoplasm, Extracellular	iCS400	Schilling, et al, <i>J Theor Biol</i> , 2000
<i>Helicobacter pylori</i>	<i>Helicobacter pylori</i> 26695	Tomb, et al, <i>Nature</i> , 1997	1667	1576	388	403	291	18.5	Cytoplasm, Extracellular	iCS291	Schilling, et al, <i>J Bacteriol</i> , 2002
	<i>Helicobacter pylori</i> 26695	Tomb, et al, <i>Nature</i> , 1997	1667	1576	476	485	341	21.6	Cytoplasm, Extracellular	iIT341	Thiele, et al, <i>J Bacteriol</i> , 2005
<i>Plasmodium falciparum</i>	<i>Plasmodium falciparum</i> 3D7	Gardner, et al, <i>Nature</i> , 2002	22900	5268	697	525	(816)	NA	Cytoplasm, Extracellular	iY816	Yeh, et al, <i>Genome Research</i> , 2004
<i>Mannheimia succinifaciens</i>	<i>Mannheimia succinifaciens</i> MBEL55E	Hong, et al, <i>Nat Biotechnol</i> , 2004	2314	2380	373	352	329	0.1	Cytoplasm, Extracellular	iSH329	Hong, et al, <i>Nat Biotechnol</i> , 2004
	<i>Mannheimia succinifaciens</i> MBEL55E	Hong, et al, <i>Nat Biotechnol</i> , 2004	2314	2380	686 (638)	519	425	17.9	Cytoplasm, Extracellular	iTK425	Kim, et al, <i>Biotechnol Bioeng</i> , 2007
<i>Methanococcus jannaschii</i>	<i>Methanococcus jannaschii</i> DSM 2661	Bult, et al, <i>Science</i> , 1996	1664	1729	609	510	(436)	NA	Cytoplasm, Extracellular	iST436	Tsoka, et al, <i>Archaea</i> , 2003
<i>Streptomyces coelicolor</i>	<i>Streptomyces coelicolor</i> A3(2) M145	Bentley, et al, <i>Nature</i> , 2002	8667	7769	971 (700)	500	711	9.2	Cytoplasm, Extracellular	iIB711	Borodina, et al, <i>Genome Research</i> , 2005
<i>Aspergillus niger</i>	<i>Aspergillus niger</i> CBS 513.88	Pel, et al, <i>Nat Biotechnol</i> , 2007	33900	14165	355	284	20	0.1	Cytoplasm, Mitochondria, Glyoxysome, Extracellular	iHD20	David, et al, <i>Eur J Biochem</i> , 2003
	<i>Aspergillus niger</i> CBS 513.88 and <i>Aspergillus niger</i> ATCC 9029 <sup>g</sup>	Pel, et al, <i>Nat Biotechnol</i> , 2007	33900	14165	2443	2349	(988)	NA	Cytoplasm, Extracellular	iS988	Sun, et al, <i>Genome Biology</i> , 2007
	<i>Aspergillus niger</i> CBS 513.88 and <i>Aspergillus niger</i> ATCC 1015 <sup>h</sup>	Pel, et al, <i>Nat Biotechnol</i> , 2007	33900	14165	2240 (1190)	1045 (782)	871	6.1	Cytoplasm, Mitochondria, Extracellular	iMA871	Andersen, et al, <i>Molecular Systems Biology</i> , 2008

<i>Aspergillus nidulans</i>	<i>Aspergillus nidulans</i> FGSC A4	Galagan, et al, <i>Nature</i> , 2005	31000	9500	1213 (794)	732 (551)	666	7.0	Cytoplasm, Mitochondria, Glyoxysome, Extracellular	iHD666	David, et al, <i>Genome Biology</i> , 2006
<i>Aspergillus oryzae</i>	<i>Aspergillus oryzae</i> RIB40	Machida, et al, <i>Nature</i> , 2005	37000	12074	(1679)	1040	1184	9.8	Cytoplasm, Mitochondria, Extracellular	iWV1184	Vongsangnak, et al, <i>BMC Genomics</i> , 2008
<i>Lactococcus lactis</i>	<i>Lactococcus lactis</i> IL1403	Bolotin, et al, <i>Genome Research</i> , 2001	2365	2321	621	509 (422)	358	30.6	Cytoplasm, Extracellular	iAO358	Oliveira, et al, <i>BMC Microbiol</i> , 2005
<i>Lactobacillus plantarum</i>	<i>Lactobacillus plantarum</i> WFCs1	Kleerebezem, et al, <i>PNAS</i> , 2003	3308	3009	704	670	210 (710)	23.6	Cytoplasm, Extracellular	iBT710	Teusink, et al, <i>AEM</i> , 2005
<i>Bacillus subtilis</i>	<i>Bacillus subtilis</i> 168	Kunst, et al, <i>Nature</i> , 2007	4214	4105	1020	988	844	20.6	Cytoplasm, Extracellular	iY0844	Oh, et al, <i>J Biol Chem</i> , 2007
	<i>Bacillus subtilis</i> 168	Kunst, et al, <i>Nature</i> , 2007	4214	4105	563	NA	534	13.0	Cytoplasm, Extracellular	iAG534	Goelzer, et al, <i>BMS Systems Biology</i> , 2008
<i>Staphylococcus aureus</i>	<i>Staphylococcus aureus</i> N315 (MRSA)	Kuroda, et al, <i>Lancet</i> , 2001	2813	2588	640	571	619	23.9	Cytoplasm, Extracellular	iSB619	Becker, et al, <i>BMC Bioinformatics</i> , 2005
<i>Corynebacterium glutamicum</i>	<i>Corynebacterium glutamicum</i> Nakagawa	Ikeda, et al, <i>Appl Microbiol Biotechnol</i> , 2001	3309	2993	(446)	411	446	14.9	Cytoplasm, Extracellular	iKK446	Kjeldsen, et al, <i>Biotechn &amp; Bioeng</i> , 2008
<i>Mycobacterium tuberculosis</i>	<i>Mycobacterium tuberculosis</i> H37Rv (lab strain)	Cole, et al, <i>Nature</i> , 1998	4411	4402	939	828	661	15.0	Cytoplasm, Extracellular	iNJ661	Jamshidi, et al, <i>BMC Systems Biology</i> , 2007
	<i>Mycobacterium tuberculosis</i> H37Rv (lab strain)	Cole, et al, <i>Nature</i> , 1998	4411	4402	849	739	726 (723)	16.5	Cytoplasm, Extracellular	iDB726	Beste, et al, <i>Genome Biology</i> , 2007
<i>Methanoscincina barkeri</i>	<i>Methanoscincina barkeri</i> Fusaro	Maeder, et al, <i>J Bacteriol</i> , 2006	4837	3606	619	558	692	19.2	Cytoplasm, Extracellular	iAF692	Feist, et al, <i>Molecular Systems Biology</i> , 2006

<i>Rhizobium etli</i>	<i>Rhizobium etli</i> CFN42	Gonzalez, et al, <i>PNAS</i> , 2006	4381	4035	387	371	363	9.0	Cytoplasm, Extracellular	iRO36 3	Resendis-Antonio, et al, <i>PLOS Comp Biol</i> , 2007
<i>Homo sapiens</i>	<i>Homo sapiens</i>	Lander, et al, <i>Nature</i> , 2001	3200000	26966	3311	2712	1496	5.5	Cytoplasm, Mitochondrion, Golgi apparatus, Endoplasmic reticulum, Lysosome, Peroxisome, Nucleus, Extracellular	<i>Homo sapien s</i> Recon 1	Duarte, et al, <i>PNAS</i> , 2007
<i>Homo sapiens</i> Mitochondria <sup>i</sup>	NA	NA	NA	NA	189 (153)	230	(298)	NA	Cytoplasm, Mitochondrion, Extracellular	iTV298	Vo, et al, <i>J Biol Chem</i> , 2004
<i>Mus musculus</i> Cardiomyocyte	<i>Mus musculus</i> C57BL/6J	Mouse Genome Sequencing Consortium, et al, <i>Nature</i> , 2002	2500000	24174	1220	872	473	2.0	extracellular space, cytosol, mitochondria	iKS473	Sheikh, et al, <i>Biotechnol Prog</i> , 2005

- Notes** Genome-scale reconstructed metabolic networks report several different parameters characterizing the network. Absolute consistency amongst reconstructions is not feasible; however, where possible common parameters have been defined providing dimensions of the reconstructed networks.
- a** Total reactions, as defined by the authors, includes intracellular, extracellular, and exchange reactions. Where available, unique reactions are defined as the total number of reactions absent of any isoenzyme catalyzed reactions, where the reaction stoichiometry is identical.
  - b** Total metabolites, as defined by the authors, includes all reactants, products, co-factors, catalysts, and intermediates involved in any stoichiometric reaction. Unique metabolites are defined as those unique in chemical structure, since a fraction of metabolites with identical chemical structure may be found in multiple compartments.
  - c** In most reconstructed networks the number of genes included in the model, as defined by the open reading frames (ORFs) producing a gene product that catalyzes a defined stoichiometric reaction, are provided. However, several reconstructions only include the gene product (e.g., enzymes) with no indication of ORF association.
  - d** In most reconstructed networks the percent of the sequenced genome annotated by the model is provided. Here, the percent genome covered is calculated based upon the total number of ORFs from the originally sequenced organism, and the total genes included in the reconstruction. Values may differ slightly from the original publication of the reconstructed network if the sequenced genome of the organism has been updated.
  - e** In most reconstructed networks the compartmentalization used is provided. However, in several models no mention of compartments is provided, therefore, annotation of this parameter is provided here based on inspection of the model.
  - f** The common nomenclature used for model identification is *i-First name-Last name-Number of ORFs represented*. In several cases there has been deviation from this nomenclature and the model name used in the original publication is provided for consistency.

- g** Strain *Aspergillus niger* ATCC 9029 was genome sequenced, 3-fold coverage, by Integrated Genomics (Chicago, USA); however, this genome sequence is not presently, publically available.
- h** Strain *Aspergillus niger* ATCC 1015 was genome sequenced by the Department of Energy's Joint Genome Institute; however, this genome sequence is not presently listed as completed.
- i** This reconstructed model was not based on a genome sequence, but rather included analysis of a proteome network constructed for human mitochondria.

## References

1. Pass F. 1981. Biotechnology, a new industrial revolution. *J Am Acad Dermatol* 4:476-477.
2. Ferrandiz-Garcia F. 1982. Biotechnology. Genetic engineering. Chemico-pharmaceutical area. *Rev Esp Fisiol* 38:353-366.
3. Maury J, Asadollahi MA, Moller K, Clark A and Nielsen J. 2005. Microbial isoprenoid production: an example of green chemistry through metabolic engineering. *Adv Biochem Eng Biotechnol* 100:19-51.
4. Hirche C. 2006. Trend Report No. 16: Industrial Biotechnology - White Biotechnology: A promising tool for the chemical industry. *ACHEMA 2006 28th International Exhibition-Congress on Chemical Engineering, Environmental Protection and Biotechnology* 16:1-7.
5. Gavrilescu M and Chisti Y. 2005. Biotechnology--a sustainable alternative for chemical industry. *Biotechnology Advances* 23:471-499.
6. Nielsen J. 1995. Physiological Engineering Aspects of *Penicillium chrysogenum*. Polyteknisk Forlag, Lyngby, p. 13.
7. Otero JM, Panagiotou G, and Olsson L. 2007. Fueling Industrial Biotechnology Growth with Bioethanol. *Adv Biochem Eng Biotechnol* 108:1-40.
8. Renewable Fuels Association, *Growing Innovation: Ethanol Industry Outlook 2009*, Available at [http://www.ethanolrfa.org/objects/pdf/outlook/RFA\\_Outlook\\_2009.pdf](http://www.ethanolrfa.org/objects/pdf/outlook/RFA_Outlook_2009.pdf), last visited May 3, 2009.
9. Dien BS, Cotta MA and Jeffries TW. 2003. Bacteria engineered for fuel ethanol production: current status. *Appl Microbiol Biotechnol* 63:258-266.
10. Russo S, Berkovitz Siman-Tov R and Poli G. 1995. Yeasts: from genetics to biotechnology. *J Environ Pathol Toxicol Oncol* 14:133-157.
11. Adrio JL and Demain AL. 2006. Genetic improvement of processes yielding microbial products. *FEMS Microbiol Rev* 30: 187-214.
12. Smedsgaard J and Nielsen J. 2005. Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics. *J Exp Bot* 56: 273-286.
13. Fisk DG, Ball CA, Dolinski K, Engel SR, Hong EL, Issel-Tarver L, Schwartz K, Sethuraman A, Botstein D and Cherry JM. 2006. *Saccharomyces cerevisiae* S288C genome annotation: a working hypothesis. *Yeast* 23: 857-865.
14. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H and Oliver SG. 1996. Life with 6000 genes. *Science* 274: 546, 563-7.
15. DeRisi JL, Iyer VR and Brown PO. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278: 680-686.
16. Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabrielian AE, Landsman D, Lockhart DJ and Davis RW. 1998. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell* 2: 65-73.
17. Forster J, Famili I, Fu P, Palsson BO and Nielsen J. 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* 13: 244-253.
18. Famili I, Forster J, Nielsen J and 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 USA* 100: 13134-13139.
19. Vemuri GN and Aristidou AA. 2005. Metabolic engineering in the -omics era: elucidating and modulating regulatory networks. *Microbiol Mol Biol Rev* 69: 197-216.
20. Patil KR, Akesson M and Nielsen J. 2004. Use of genome-scale microbial models for metabolic engineering. *Curr Opin Biotechnol* 15: 64-69.

21. Stephanopoulos G. 1999. Metabolic fluxes and metabolic engineering. *Metab Eng* 1: 1-11.
22. Bro C and Nielsen J. 2004. Impact of 'ome' analyses on inverse metabolic engineering. *Metab Eng* 6: 204-211.
23. Lynd LR, Wyman CE and Gerngross TU. 1999. Biocommodity Engineering. *Biotechnol Prog* 15: 777-793.
24. Parekh S, Vinci VA and Strobel RJ. 2000. Improvement of microbial strains and fermentation processes. *Appl Microbiol Biotechnol* 54: 287-301.
25. Demain AL. 2000. Microbial biotechnology. *Trends Biotechnol* 18: 26-31.
26. Demain AL and Adrio JL. 2008. Strain improvement for production of pharmaceuticals and other microbial metabolites by fermentation. *Prog Drug Res* 65: 251, 253-89.
27. Demain AL and Adrio JL. 2008. Contributions of microorganisms to industrial biology. *Mol Biotechnol* 38: 41-55.
28. Schmeisser C, Steele H and Streit WR. 2007. Metagenomics, biotechnology with non-culturable microbes. *Appl Microbiol Biotechnol* 75: 955-962.
29. Patnaik R. 2008. Engineering complex phenotypes in industrial strains. *Biotechnol Prog*. 24: 38-47.
30. Bailey JE. 1991. Toward a science of metabolic engineering. *Science* 252: 1668-1675.
31. Moo-Young M, Glick BR, Bu'Lock JD and Cooney CL. 1984. Biotechnology has been associated with the pharmaceutical and brewing industries. *Biotechnol Adv* 2: I-II.
32. Poppe L and Novak L. 1992. Selective biocatalysis: a synthetic approach. Wiley-VCH, Weinheim.
33. Schmid RD. 2003. Pocket guide to biotechnology and genetic engineering. Wiley-VCH, Weinheim.
34. Viswanathan G, Seto J, Patil S, Nudelman G and Sealfon S. 2008. Getting Started in Biological Pathway Construction and Analysis. *PLoS Computational Biology* 4: e16.
35. Bruggeman FJ and Westerhoff HV. 2007. The nature of systems biology. *Trends Microbiol* 15: 45-50.
36. Edwards JS, Covert M and Palsson BO. 2002. Metabolic modeling of microbes: the flux-balance approach. *Environ Microbiol* 4: 133-140.
37. Edwards JS, Ramakrishna R and Palsson BO. 2002. Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol Bioeng* 77: 27-36.
38. Forster J, Famili I, Palsson BO and Nielsen J. 2003. Large-scale evaluation of in silico gene deletions in *Saccharomyces cerevisiae*. *OMICS* 7: 193-202.
39. Edwards JS, Ibarra RU and Palsson BO. 2001. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* 19: 125-130.
40. Hermann BG and Patel M. 2007. Today's and tomorrow's bio-based bulk chemicals from white biotechnology: a techno-economic analysis. *Appl Biochem Biotechnol* 136: 361-388.
41. Takors R, Bathe B, Rieping M, Hans S, Kelle R and Huthmacher K. 2007. Systems biology for industrial strains and fermentation processes--example: amino acids. *J Biotechnol* 129: 181-190.
42. Dodds DR and Gross RA. 2007. Chemistry: Chemicals from biomass. *Science* 318: 1250-1251.
43. Hermann BG, Blok K and Patel MK. 2007. Producing bio-based bulk chemicals using industrial biotechnology saves energy and combats climate change. *Environ Sci Technol* 41: 7915-7921.
44. Hatti-Kaul R, Tornvall U, Gustafsson L and Borjesson P. 2007. Industrial biotechnology for the production of bio-based chemicals--a cradle-to-grave perspective. *Trends Biotechnol* 25: 119-124.
45. Nissen TL, Kielland-Brandt MC, Nielsen J and Villadsen J. 2000. Optimization of ethanol production in *Saccharomyces cerevisiae* by metabolic engineering of the ammonium assimilation. *Metab Eng* 2: 69-77.
46. Bro C, Regenberg B, Forster J and Nielsen J. 2006. In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metab Eng* 8: 102-111.
47. Patil KR and Nielsen J. 2005. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci USA* 102: 2685-2689.

48. Cakir T, Patil KR, Onsan Zi, Ulgen KO, Kirdar B and Nielsen J. 2006. Integration of metabolome data with metabolic networks reveals reporter reactions. *Mol Syst Biol* 2: 50.
49. Olsson L, Jørgensen H, Krogh K and Roca C. 2004. Bioethanol Production from Lignocellulosic Material. *Polysaccharides: Structural Diversity and Functional Versatility*, Second Edition pp. 957-993.
50. Grotkjaer T, Christakopoulos P, Nielsen J and Olsson L. 2005. Comparative metabolic network analysis of two xylose fermenting recombinant *Saccharomyces cerevisiae* strains. *Metab Eng* 7: 437-444.
51. Roca C, Nielsen J and Olsson L. 2003. Metabolic engineering of ammonium assimilation in xylose-fermenting *Saccharomyces cerevisiae* improves ethanol production. *Appl Environ Microbiol* 69: 4732-4736.
52. Antoni D, Zverlov VV and Schwarz WH. 2007. Biofuels from microbes. *Appl Microbiol Biotechnol* 77: 23-35.
53. E. I. du Pont de Nemours and Company website specifically focused on the Sorona® product family. Available at [http://www2.dupont.com/Sorona/en\\_US/](http://www2.dupont.com/Sorona/en_US/), last accessed May 5, 2008.
54. E. I. du Pont de Nemours and Company 2007 Annual Review. Available at [http://library.corporate-ir.net/library/73/733/73320/items/283770/DD\\_2007\\_AR\\_v2.pdf](http://library.corporate-ir.net/library/73/733/73320/items/283770/DD_2007_AR_v2.pdf), last accessed May 5, 2008.
55. Nakamura CE and Whited GM. 2003. Metabolic engineering for the microbial production of 1,3-propanediol. *Curr Opin Biotechnol* 14: 454-459.
56. Freund A. 1881. Ueber die bildung und darstellung von trimethylene-alkohol aus glycerin. *Monatsheft fur Chemie* 2: 636-641.
57. Braak HR. 1928. Onderzoeken over Vergisting van Glycerine. Delft.
58. Mickelson MN and Werkman CH. 1940. The Dissimilation of Glycerol by Coli-Aerogenes Intermediates. *J Bacteriol* 39: 709-715.
59. Antoniewicz MR, Kraynie DF, Laffend LA, Gonzalez-Lergier J, Kelleher JK and Stephanopoulos G. 2007. Metabolic flux analysis in a nonstationary system: fed-batch fermentation of a high yielding strain of *E. coli* producing 1,3-propanediol. *Metab Eng* 9: 277-292.
60. Antoniewicz MR, Kelleher JK and Stephanopoulos G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab Eng* 9: 68-86.
61. Mu Y, Teng H, Zhang DJ, Wang W and Xiu ZL. 2006. Microbial production of 1,3-propanediol by *Klebsiella pneumoniae* using crude glycerol from biodiesel preparations. *Biotechnol Lett* 28: 1755-1759.
62. Cheng KK, Zhang JA, Liu DH, Sun Y, Yang MD and Xu JM. 2006. Production of 1,3-propanediol by *Klebsiella pneumoniae* from glycerol broth. *Biotechnol Lett* 28: 1817-1821.
63. Zhang Q, Teng H, Sun Y, Xiu Z and Zeng A. 2008. Metabolic flux and robustness analysis of glycerol metabolism in *Klebsiella pneumonia*. *Bioprocess Biosyst Eng* 31: 127-135.
64. Chi N, Liu C, Liu Y, Zhang Q and Zheng X. 2003. [Cloning and expressing of 1,3-propanediol oxidoreductase-encoding gene]. *Wei Sheng Wu Xue Bao* 43: 717-721.
65. Liu HJ, Zhang DJ, Xu YH, Mu Y, Sun YQ and Xiu ZL. 2007. Microbial production of 1,3-propanediol from glycerol by *Klebsiella pneumoniae* under micro-aerobic conditions up to a pilot scale. *Biotechnol Lett* 29: 1281-1285.
66. US Department of Energy, "Top Value Added Chemicals from Biomass: Volume I", available at <http://www1.eere.energy.gov/biomass/pdfs/35523.pdf>, last accessed May 5, 2008.
67. Wisconsin Biorefining Development Initiative, "Biobased Products: Succinic acid" , available at <http://wisbiorefine.org/prod/sacid.pdf>, last accessed May 31, 2009.
68. Zeikus JG, Jain MK and Elankovan P. 1999. Biotechnology of succinic acid production and markets for derived industrial products. *Applied Microbiology and Biotechnology* 51: 545-552.
69. Song H and Lee SJ. 2006. Production of succinic acid by bacterial fermentation. *Enzyme and Microbial Technology* 39: 352-361.

70. McKinlay JB, Vieille C and Zeikus JG. 2007. Prospects for a bio-based succinate industry. *Appl Microbiol Biotechnol* 76: 727-740.
71. Jantama K, Haupt MJ, Svoronos SA, Zhang X, Moore JC, Shanmugam KT and Ingram LO. 2008. Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* C that produce succinate and malate. *Biotechnol Bioeng* 99: 1140-1153.
72. Lee PC, Lee SY, Hong SH and Chang HN. 2002. Isolation and characterization of a new succinic acid-producing bacterium, *Mannheimia succiniciproducens* MBEL55E, from bovine rumen. *Appl Microbiol Biotechnol* 58: 663-668.
73. Lee PC, Lee SY, Hong SH and Chang HN. 2003. Batch and continuous cultures of *Mannheimia succiniciproducens* MBEL55E for the production of succinic acid from whey and corn steep liquor. *Bioprocess Biosyst Eng* 26: 63-67.
74. Hong SH, Kim JS, Lee SY, In YH, Choi SS, Rih JK, Kim CH, Jeong H, Hur CG and Kim JJ. 2004. The genome sequence of the capnophilic rumen bacterium *Mannheimia succiniciproducens*. *Nat Biotechnol* 22: 1275-1281.
75. Lee SJ, Song H and Lee SY. 2006. Genome-based metabolic engineering of *Mannheimia succiniciproducens* for succinic acid production. *Appl Environ Microbiol* 72: 1939-1948.
76. Lee JW, Lee SY, Song H and Yoo J-S. 2006. The proteome of *Mannheimia succiniciproducens*, a capnophilic rumen bacterium. *Proteomics* 6: 3550-3566.
77. Kim TY, Kim HU, Park JM, Song H, Kim JS and Lee SY. 2007. Genome-scale analysis of *Mannheimia succiniciproducens* metabolism. *Biotechnol Bioeng* 97: 657-671.
78. Burgard AP, Pharkya P and Maranas CD. 2003. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 84: 647-657.
79. Hong EL, Balakrishnan R, Dong Q, Christie KR, Park J, Binkley G, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Krieger CJ, Livstone MS, Miyasato SR, Nash RS, Oughtred R, Skrzypek MS, Weng S, Wong ED, Zhu KK, Dolinski K, Botstein D, Cherry JM. 2008. Gene Ontology annotations at SGD: new data sources and annotation methods. *Nucleic Acids Res* 36: D577-81.
80. Breitling R, Vitkup D and Barrett MP. 2008. New surveyor tools for charting microbial metabolic maps. *Nat Rev Microbiol* 6: 156-161.
81. Green ML and Karp PD. 2005. Genome annotation errors in pathway databases due to semantic ambiguity in partial EC numbers. *Nucleic Acids Res* 33: 4035-4039.
82. Lespinet O and Labedan B. 2006. Orphan enzymes could be an unexplored reservoir of new drug targets. *Drug Discov Today* 11: 300-305.
83. Nookaew I, Jewett MC, Meechai A, Thammarongtham C, Laoteng K, Cheevadhanarak S, Nielsen J, and Bhumiratana S. 2008. The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol* 7:71.
84. Schuetz R, Kuepfer L and Sauer U. 2007. "Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol* 3: 119.
85. Joyce AR and Palsson BO. 2008. Predicting gene essentiality using genome-scale in silico models. *Methods Mol Biol* 416: 433-457.
86. Joyce AR and Palsson BO. 2007. Toward whole cell modeling and simulation: comprehensive functional genomics through the constraint-based approach. *Prog Drug Res* 64: 265, 267-309.
87. Joyce AR, Reed JL, White A, Edwards R, Osterman A, Baba T, Mori H, Lesely SA, Palsson BO and Agarwalla S. 2006. Experimental and computational assessment of conditionally essential genes in *Escherichia coli*. *J Bacteriol* 188: 8259-8271.
88. Joyce AR and Palsson BO. 2006. The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol* 7: 198-210.

89. Rocha I, Forster J and Nielsen J. 2008. Design and application of genome-scale reconstructed metabolic models. *Methods Mol Biol* 416: 409-431.
90. Kim HU, Kim TY and Lee SY. 2008. Metabolic flux analysis and metabolic engineering of microorganisms. *Mol Biosyst* 4: 113-120.
91. Kim TY, Sohn SB, Kim HU and Lee SY. 2008. Strategies for systems-level metabolic engineering. *Biotechnol J* 3: 612-623.
92. Notebaart RA, Teusink B, Siezen RJ and Papp B. 2008. Co-regulation of metabolic genes is better explained by flux coupling than by network distance. *PLoS Comput Biol* 4: e26.
93. Notebaart RA, van Enckevort FH, Francke C, Siezen RJ and Teusink B. 2006. Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics* 7: 296.
94. Zhao J, Ding GH, Tao L, Yu H, Yu ZH, Luo JH, Cao ZW and Li YX. 2007. Modular co-evolution of metabolic networks. *BMC Bioinformatics* 8: 311.
95. Satish Kumar V, Dasika MS and Maranas CD. 2007. Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics* 8: 212.

## CHALMERS UNIVERSITY OF TECHNOLOGY

SE 412 96 Göteborg, Sweden

Telephone: +46-(0)31 772 10 00

[www.chalmers.se](http://www.chalmers.se)

The chemical industry is currently undergoing a dramatic change driven by demand for developing more sustainable processes for the production of fuels, chemicals, and materials. In biotechnological processes different microorganisms can be exploited, and the large diversity of metabolic reactions represents a rich repository for the design of chemical conversion processes that lead to efficient production of desirable products. However, often microorganisms that produce a desirable product, either naturally or because they have been engineered through insertion of heterologous pathways, have low yields and productivities, and in order to establish an economically viable process it is necessary to improve the performance of the microorganism. Here metabolic engineering is the enabling technology. Through metabolic engineering the metabolic landscape of the microorganism is engineered such that there is an efficient conversion of the raw material, typically glucose, to the product of interest. This process may involve both insertion of new enzymes activities, deletion of existing enzyme activities, but often also de-regulation of existing regulatory structures operating in the cell. In order to rapidly identify the optimal metabolic engineering strategy the industry is to an increasing extent looking into the use of tools from systems biology. This involves both *x-ome* technologies such as transcriptome, proteome, metabolome, and fluxome analysis, and advanced mathematical modeling tools such as genome-scale metabolic modeling.

*Saccharomyces cerevisiae* is the most well characterized eukaryote, the preferred microbial cell factory for the largest industrial biotechnology product (bioethanol), and a robust commercially compatible scaffold to be exploited for diverse chemical production. Succinic acid is a highly sought after added-value chemical for which there is no native production and accumulation in *S. cerevisiae*. Intuitive genetic targets for either over-expression or interruption of succinate producing or consuming pathways, respectively, do not lead to increased succinate. Rather, I demonstrate how systems biology tools coupled with directed evolution and selection allows for non-intuitive, rapid and substantial re-direction of carbon fluxes in *S. cerevisiae*. Hence, I show proof of concept that this is a potentially attractive cell factory for over-producing different platform chemicals. This approach is extended to further engineer *S. cerevisiae* to express biorefinery desirable phenotypes. With a clear presentation of the history of different systems biology tools, a review how they find application in industrial biotechnology, and novel demonstration of their potential in my thesis, I define a new term, *industrial systems biology*. Industrial systems biology facilitating metabolic engineering presents new opportunities for bio-based processes to provide sustainable, environmentally favorable, and cost-effective alternatives. The microbial cell factory is at the heart of these processes.



## JOSÉ MANUEL OTERO ROMERO

Systems Biology

Department of Chemical and Biological Engineering  
Chalmers University of Technology