



Universidade do Minho
Escola de Engenharia

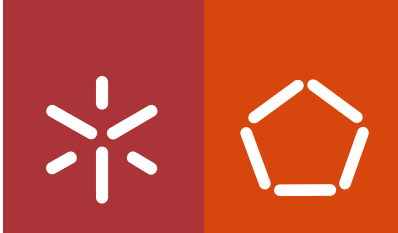
Rui Miguel Pinheiro da Silva Pereira

Improvement of *in silico* strain engineering methods in *Saccharomyces cerevisiae*

Esta investigação foi financiada pela Fundação para a Ciência e Tecnologia através da concessão de uma bolsa de doutoramento (SFRH/BD/51111/2010), co-financiada pelo POPH - QREN - Tipologia 4.1 -Formação Avançada - e comparticipados pelo Fundo Social Europeu (FSE) e por fundos nacionais do Ministério da Ciência, Tecnologia e Ensino Superior (MCTES).



Rui Miguel Pinheiro da Silva Pereira Improvement of *in silico* strain engineering methods in *Saccharomyces cerevisiae*



Universidade do Minho
Escola de Engenharia

Rui Miguel Pinheiro da Silva Pereira

**Improvement of *in silico* strain engineering
methods in *Saccharomyces cerevisiae***

PhD thesis in Bioengineering

This work was executed under the supervision of:
Professor Isabel Cristina A. Pereira da Rocha
Professor Jens Nielsen

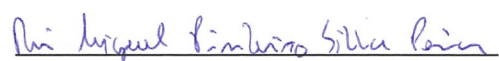
July 2015

STATEMENT OF INTEGRITY

I hereby declare having conducted my thesis with integrity. I confirm that I have not used plagiarism or any form of falsification of results in the process of the thesis elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

University of Minho, July 21st 2015


Rui Miguel Pinheiro da Silva Pereira

AGRADECIMENTOS

Para concluir este doutoramento, tive o privilégio de contar com a ajuda e apoio de diversas pessoas, tanto ao nível científico como motivacional. A todos o meu muito obrigado por tornarem possível a elaboração desta tese.

Em primeiro lugar quero agradecer à minha orientadora, Professora Isabel Rocha, por me ter acolhido como bolseiro de investigação e posteriormente por me ter incentivado a ingressar no doutoramento. Obrigado por todos os seus ensinamentos, pela paciência para ouvir as minhas dúvidas e pela orientação do trabalho que consta nesta tese. Agradeço também a sua disponibilidade para me animar quando o trabalho corria menos bem e por sempre mostrar confiança que era possível recuperar o tempo perdido. Há 6 anos atrás não tinha sequer ouvido falar em modelos à escala genómica e sem a sua ajuda não teria sido possível chegar a este nível.

I would also like to express my sincere gratitude to Professor Jens Nielsen for taking me as his PhD student, for all his advices and for teaching me so much about yeast and science in general. My time in Chalmers had its ups and downs, but your positive attitude was vital to keep me motivated to chase my goals. Thank you for making me feel welcome in your group and for the freedom you gave me to make decisions about my work.

Thank you Doctor Verena Siewers for all your patience to listen to all my theories about why experiments did not work and for all your help in making them work. Your constant availability for a quick question or a small meeting was of the utmost importance for me to learn so much. I also want to thank Doctor Yun Chen for his help planning the 3-HP work and for all the fruitful discussions about the results.

O meu obrigado ao Professor Eugénio Ferreira que, juntamente com a Professora Isabel Rocha, me acolheu no seu grupo como bolseiro de investigação e que me incentivou a ingressar no programa doutoral em Bioengenharia.

Também agradeço à minha instituição de acolhimento, Centro de Engenharia Biológica da Universidade do Minho, onde encontrei todas as condições para realizar este trabalho científico e à Fundação para a Ciência e Tecnologia e ao programa MIT Portugal pela atribuição da bolsa de doutoramento (SFRH/BD/51111/2010).

Quero agradecer aos membros do grupo de trabalho BisBII pelos bons momentos passados durante o doutoramento, pelas aventuras durante os fins-de-semana de grupo e pelas discussões e conselhos sobre o trabalho realizado nesta tese. Um obrigado especial ao Paulo Vilaça e ao Paulo Maia pela preciosa ajuda durante as minhas aventuras pela programação e pelo apoio constante no uso do OptFlux.

O meu agradecimento a todos os membros do Laboratório de Bioprocessos pelos bons momentos que lá passei, pela vossa ajuda e boa disposição. Um obrigado especial à Ana Guimarães e à Andrea Silva pela ajuda e motivação que sempre me deram.

I also want to express my gratitude to everyone in SysBio for welcoming me and for all your help during my visits to Sweden. Thank you to all the research engineers, administrative staff and fellow co-workers for your help during my time there. Special thanks go to António Roldão, Il Kwon, Rahul Kumar, Stefan Tippman and José Martínez for everything that you taught me and for your support during the lab work. Thanks to my

office mates for a relaxed working environment and for introducing me to Swedish concepts, such as fika and afterwork.

Aos meus colegas e amigos do programa doutoral em Bioengenharia MIT Portugal por um ano curricular cheio de bons momentos, em especial à Joana Rodrigues, Catarina Lima, Patrícia Pereira e ao Joaquim Barbosa. À Joana Rodrigues um obrigado especial pela presença constante durante esta aventura e pela tua amizade e motivação.

Obrigado também à Cristiana, Ester e Florência pela amizade de longa data e pelos bons momentos que passamos juntos. A todos os amigos de Biologia Aplicada, obrigado por partilharem as vossas histórias do doutoramento e me fazerem sentir melhor.

To Martina, a very special thank you for your love and for making me happy, you made Sweden feel like home (hvala ti puno ljubica moja).

Quero agradecer à minha família por todo o apoio que me deram e por me incentivarem a prosseguir nos momentos de maior desânimo. Quero agradecer especialmente aos meus pais e à minha irmã por todo o carinho e ajuda que me deram e por nunca me deixarem sentir sozinho, mesmo estando fora de Portugal.

ABSTRACT

The buildup of knowledge about microbial metabolism and the development of genome engineering techniques gave rise to the rational modification of microorganisms in order to use them to biosynthesize chemicals of industrial interest. Recently, the construction of genome-scale metabolic models (GSMMs) allowed the design of strain engineering strategies *in silico*. This thesis focused on the study and improvement of *in silico* strain engineering methodologies using *Saccharomyces cerevisiae* as a case study organism.

Firstly, in order to investigate the accuracy of the GSMMs available for *S. cerevisiae*, their capacity to simulate the intracellular fluxes in central metabolism was tested. The results revealed that the simulations contained relevant errors in important areas of the central metabolism. A careful manual curation of the feasibility of all reactions producing or consuming NADH / NADPH resulted in the improvement of many fluxes in central metabolic pathways when compared to fluxes measured experimentally.

The lack of a simulation method that could predict in quantitative terms the phenotype of strains with complex engineered genotypes, led to the development of a novel simulation method called turnover dependent phenotypic simulation (TDPS). This method was designed with the goal of simulating the majority of the genetic modifications usually implemented in engineered strains. The assumption that the production turnover of a metabolite can be used as an indication of its abundance was used in the formulation of TDPS in order to take into account the availability of resources when modelling genetic modifications. TDPS was validated using metabolically engineered *S. cerevisiae* strains available in the literature by comparing the production yields of the target metabolite.

TDPS was then applied to the optimization of the availability of cytosolic acetyl-CoA in *S. cerevisiae*, by using an evolutionary algorithm to search for sets of genetic alterations that could improve the production yield of 3-hydroxypropionic acid (3-HP) derived from acetyl-CoA. Although the yields obtained experimentally were considerably lower than the simulations suggested, a positive effect on the 3-HP yield was observed for the down-regulation of the pyruvate dehydrogenase complex and the deletion of *ACHI* (succinyl-CoA:acetate CoA-transferase).

SUMÁRIO

O progresso que tem sido feito na área da fisiologia microbiana, juntamente com o desenvolvimento de técnicas de engenharia genética, permitiu a criação de estirpes microbianas modificadas racionalmente com o intuito de otimizar a produção de compostos de interesse industrial. Mais recentemente, a construção de modelos metabólicos à escala genómica (MMEG) proporcionou o desenho de estirpes modificadas *in silico*. Esta tese focou-se no estudo e melhoramento de metodologias de manipulação de estirpes *in silico*, usando *Saccharomyces cerevisiae* como caso de estudo.

De forma a investigar a precisão dos MMEG disponíveis para *S. cerevisiae*, a sua capacidade para simular os fluxos intracelulares foi testada. Os resultados mostraram que os fluxos simulados continham erros em áreas importantes do metabolismo central e que a curação manual das reacções envolvidas no metabolismo de NADH e NADPH resulta em melhorias significativas nos fluxos metabólicos centrais.

A ausência de um método de simulação que conseguisse prever quantitativamente o fenótipo de estirpes com genótipos complexos, levou ao desenvolvimento de um método novo designado por *turnover dependent phenotypic simulation (TDPS)*. Este método foi concebido com o objectivo de simular a maior parte das modificações genéticas normalmente implementadas em estirpes modificadas. A formulação do TDPS teve como base o uso do nível de produção de um metabolito como indicador da sua abundância, de forma a modelar as modificações genéticas em função da disponibilidade de recursos. A validação deste método foi feita usando dados da literatura sobre estirpes geneticamente modificadas de *S. cerevisiae*, através da comparação dos rendimentos simulados e reais.

O método de simulação TDPS foi posteriormente aplicado na optimização da produção de acetil-CoA no citosol de *S. cerevisiae*, usando um algoritmo evolucionário para procurar conjuntos de alterações genéticas que aumentassem a produção de ácido 3-hidroxiopropiónico derivado de acetil-CoA. Apesar dos rendimentos experimentais serem mais baixos que as simulações sugeriam, observou-se um efeito positivo da sub-regulação do complexo da piruvato desidrogenase e da eliminação do gene *ACHI* (succinil-CoA:acetato CoA-transferase).

TABLE OF CONTENTS

CHAPTER 1 Introduction	1
1.1 Context and Motivation	3
1.2 Research aims	4
1.3 Outline of the thesis	5
1.4 References	6
CHAPTER 2 Strain engineering using genome-scale metabolic models	7
2.1 Introduction	9
2.2. Strain engineering with GSMMs	11
2.2.1 Phenotype simulation methods	13
2.2.2 Target discovery methods	17
2.2.3 Auxiliary network analysis methods	24
2.3. Strain Engineering of <i>S. cerevisiae</i> using GSMMs	25
2.3.1 Genome scale models of <i>S. cerevisiae</i>	26
2.3.2 Examples of application	27
2.4. Conclusions	28
References	29
CHAPTER 3 Improving the flux distributions simulated with genome-scale metabolic models of <i>S. cerevisiae</i>	41
3.1 Introduction	43
3.2 Methods	44
3.2.1 Model retrieval and pre-processing	44
3.2.2 Simulations	45
3.2.3 Flux variability analysis	45
3.2.4 Optimizations of metabolic engineering targets	45

3.3 Results and discussion	46
3.3.1 Genome-scale metabolic model comparison	46
3.3.2 Emulation of cofactor concentrations through reaction constraints	49
3.3.3 Flux variability analysis	55
3.3.4 Mutant phenotype prediction	56
3.4 Conclusion	60
References	62
CHAPTER 4 TDPS - Turnover dependent phenotypic simulation: a quantitative constraint-based simulation method that accommodates all main strain design strategies	67
4.1 Introduction	69
4.2 Turnover Dependent Phenotypic Simulation	72
4.2.1 Modeling genetic modifications	72
4.2.2 Objective function reinforces network rigidity	74
4.2.3 Algorithm flowchart	76
4.3 Methods	79
4.3.1 Implementation of the TDPS algorithm	79
4.3.1.1 Pre-calculations	79
4.3.1.2 Splitting reversible reactions in two half-reactions	80
4.3.1.3 TDPS implementation	81
4.3.1.4 TDPS_FBA implementation	83
4.3.1.5 Solver	83
4.3.2 Toy model simulations	83
4.3.3 Case study simulations	84
4.3.3.1 Genome Scale Models and reference flux distributions	84
4.3.3.2 Parametrization of the TDPS simulation of mutant phenotypes	84
4.4. Results and discussion	85
4.4.1 TDPS toy-model validation	85

4.4.2 TDPS validation with experimental results	88
4.4.2.1 Case-study 1: PHB production in <i>S. cerevisiae</i>	89
4.4.2.2 Case-study 2: 3-HP production in <i>S. cerevisiae</i>	92
4.4.2.3 Case-study 3: L-malic acid production in <i>S. cerevisiae</i>	97
4.4.2.4 Case-study 4: santalene production in <i>S. cerevisiae</i>	101
4.4.2.5 Case-study 5: vanillin β -D-glucoside production in <i>S. cerevisiae</i>	102
4.5 Conclusions and future perspectives.....	105
References.....	106
CHAPTER 5 Optimization of the cytosolic acetyl-CoA pool in <i>S. cerevisiae</i> using <i>in silico</i> driven metabolic engineering	111
5.1 Introduction.....	113
5.2 Materials and methods	115
5.2.1 Model and software.....	115
5.2.2 Optimizations of 3-HP production using TDPS	116
5.2.3 Flux distribution under growth on ethanol.....	116
5.2.4 PCR and DNA handling.....	116
5.2.5 Media used and strain handling	117
5.2.6 Plasmid construction.....	118
5.2.7 Strain construction	119
5.2.8 Batch cultivations.....	123
5.2.9 Analytical methods	124
5.3 Results and discussion	124
5.3.1 <i>In silico</i> optimization of 3-HP production under glucose limitation	124
5.3.2 Translation of the simulation results into strain engineering strategies.....	127
5.3.3 Strain analysis in bioreactors	129
5.3.3.1 Strain analysis under respiro-fermentative glucose metabolism	130
5.3.3.2 Oxidative growth on fermentation products	132

5.3.4 Integration of Ach1p role into C2 metabolism	135
5.4 Conclusions.....	138
References.....	138
CHAPTER 6 General conclusions.....	145
6.1 General conclusions	147
6.2 Recommended future work.....	148
APPENDIX	151
Appendix A: Manual curation of the genome-scale models of <i>S. cerevisiae</i>	153
Appendix B: Modifications applied to the model Yeast 6 for simulation and optimization purposes	166
Appendix C: Construction of 3-hydroxypropionic acid producing strains.....	171
References.....	175

LIST OF FIGURES

Figure 2.1- Different categories adopted for classifying the computational methods used for strain engineering with genome-scale metabolic models.....	12
Figure 2.2- Simplified representation of the differences between the objective functions used in FBA, MOMA, LMOMA and PSEUDO.....	15
Figure 3.1- Distribution of fluxes in the central metabolism predicted by pFBA for the models: <i>i</i> FF708 (green), <i>i</i> MM904 (blue), <i>i</i> TO977 (orange), Yeast 6 (black).....	47
Figure 3.2- Comparison between experimentally determined fluxes and the predictions of pFBA for the modified models. <i>i</i> FF708 (green), <i>i</i> MM904 (blue), <i>i</i> TO977 (yellow), Yeast 6 (black), ¹³ C-MFA fully aerobic chemostat at a dilution rate of 0.1 h ⁻¹ (red)[50] , ¹³ C-MFA fully aerobic chemostat at a dilution rate of 0.1 h ⁻¹ (grey)[51] .	52
Figure 3.3- Comparison between the experimentally determined fluxes (x-axis) and pFBA simulation of each model (y-axis). For each graph it is shown: the linear regression equation, the correlation coefficient (R ²), the Spearman correlation coefficient, the Pearson correlation coefficient and the sum of differences. The experimental fluxes used in the correlations were obtained by averaging both sets of data [46, 47].....	54
Figure 3.4- Flux distributions simulated with linear MOMA for the mevalonate mutant (A) and the acetate mutant (B) using the improved flux distribution as the reference. The wild-type flux values are shown in green, the mutant values are shown in red and the knock-outs are signaled by a red cross..	59
Figure 4.1- Flux calculations performed by the objective function implemented in TDPS using the up-regulation of reaction R ₂ as an example.....	75
Figure 4.2- Flowchart describing the calculations performed by the TDPS algorithm (the detailed implementation is provided in section 4.3.1)..	77
Figure 4.3- Flux distributions obtained with TDPS for different types of genetic modifications using a toy-model. For up-regulations the C parameter was set to 2 and for down-regulation to 0.5. The reaction activation penalty constant used was 1.	86
Figure 5.1- Schematic representation of the flux modifications suggested by the TDPS algorithm. Up-regulations are shown in green, down-regulations in yellow and deletions	

in red. MCR is referred as an up-regulation but in this case this terminology is used to denote an increase of activity in absolute terms, not in comparison to a wild-type.. 126

Figure 5.2- 3-HP yields on glucose at the end of the batch cultivation for the strains engineered to improved cytosolic acetyl-CoA supply. The values shown are the average of at least two biological replicates and the error bars represent the standard deviations.

..... 130

Figure 5.3- Flux distributions for growth on ethanol of wild-type *S. cerevisiae* (green) and *ACH1* deletion mutant (red) obtained using FBA..... 136

Figure C1- Flux distribution in the central metabolism for the wild-type (green) and for the strain desgins obtained with TDPS A-D (red). The wild type flux distribution was obtained using the cofactor constraints described in chapter 3 and the mutant strains' flux distributions were calculated with TDPS..... 173

LIST OF TABLES

Table 2.1- Phenotype simulation methods available for genome-scale metabolic models	14
Table 2.2- Computational methods available for target discovery using genome-scale metabolic models	17
Table 2.3- List of the available genome-scale models for <i>S. cerevisiae</i>	26
Table 2.4- Experimental implementation of strain engineering strategies obtained using genome-scale metabolic models	28
Table 3.1- List of selected GSMM available for <i>S.cerevisiae</i>	46
Table 3.2- List of modified reactions in each model (reactions are identified by their IDs on the model).	50
Table 3.3- List of variable fluxes obtained using flux variability analysis (reactions are identified by the same IDs from Figure 3.1 and Figure 3.2).....	55
Table 3.4- Description of the strain designs obtained with OptFlux for mevalonate and acetate production using the improved reference flux distribution (the standard names of the deleted genes are shown).	57
Table 4.1- Summary of constraint-based strain design methods that include in their formulation up/down-regulation or activations	70
Table 4.2 - Constraints applied in TDPS to the flux values (V_n) for each type of genetic modification	74
Table 4.3- TDPS validation case-study using PHB production in <i>S. cerevisiae</i>	90
Table 4.4- TDPS validation case-study using 3-hydroxypropionic acid (3-HP) production in <i>S. cerevisiae</i>	94
Table 4.5- TDPS validation case-study using L-malic acid (MA) production in <i>S. cerevisiae</i>	100
Table 4.6- TDPS validation case-study using santalene production in <i>S. cerevisiae</i>	102
Table 4.7- TDPS validation case-study using vanillin-glucoside (VG) production in <i>S. cerevisiae</i>	104
Table 5.1- Plasmids used in this work.....	118
Table 5.2- Strains used in this work.....	120

Table 5.3 - Metabolic engineering strategies obtained using TDPS for 3-HP optimization	125
Table 5.4 - Genetic engineering strategies chosen to mimic the flux regulations obtained in the simulations	127
Table 5.5 - Physiological properties during respiro-fermentative metabolism of the strains engineered for 3-HP production.....	131
Table 5.6 - Physiological properties during oxidative growth on the fermentation products of the strains engineered for 3-HP production.....	133
Table A1 - List of reactions modified during the curation process of the model <i>iFF708</i>	153
Table A2 - List of reactions modified during the curation process of the model <i>iMM904</i>	156
Table A3 - List of reactions modified during the curation process of the model <i>iTO977</i>	159
Table A4 - List of reactions modified during the curation process of the model Yeast 6.	163
Table B1 - List of general changes performed in the model Yeast 6.....	166
Table B2 - List of changes applied to the model Yeast 6 to allow polyhydroxybutyrate (PHB) production.....	168
Table B3 - List of changes applied to the model Yeast 6 to allow 3-hydroxypropionic acid (3-HP) production	168
Table B4 - List of changes applied to the model Yeast 6 to allow santalene production	168
Table B5 - List of changes applied to the model Yeast 6 to allow vanillin β -D-glucoside production	169
Table B6 - List of changes applied to the model Yeast 6 for optimizing 3-hydroxypropionic acid (3-HP) production.....	169
Table B7 - List of changes applied to the model Yeast 6 to simulated growth on C ₂ carbon sources.....	170
Table C1 - Primers used in the work performed in chapter 5	171

LIST OF ABBREVIATIONS

<i>3DSD</i>	3-dedhydroshikimate dehydratase
3-HP	3-hydroxypropionic acid
3PG	3-phosphoglycerate
A-ALD	Acetylating acetaldehyde dehydrogenase
<i>AAT2</i>	Cytosolic aspartate aminotransferase
<i>ACAR</i>	Aryl carboxylic acid reductase
ACC	Acetyl-CoA carboxylase
AcCoA	Acetyl-coenzyme A
Acetyl-CoA	Acetyl-coenzyme A
ACH1	Succinyl-CoA:acetate CoA-transferase
ACONT	Aconitase
ACS	Acetyl-CoA synthetase
ACSc	Cytosolic acetyl-CoA synthetase
ACSp	Peroxisomal acetyl-CoA synthetase
ADE3	Cytosolic C1-tetrahydrofolate synthase
ADH	Alcohol dehydrogenase
ADP	Adenosine 5'-diphosphate
Akg	2-oxoglutarate
AKGD	Alpha-ketoglutarate dehydrogenase
ALD	Aldehyde dehydrogenase
ALD2/6	NAD ⁺ /NADP ⁺ - dependent cytosolic aldehyde dehydrogenase
ALDc	Cytosolic aldehyde dehydrogenase
ALDm	Mitochondrial aldehyde dehydrogenase
AMP	Adenosine 5'-monophosphate
ATP	Adenosine 5'-triphosphate
B_n^{Neg}	Binary variable
B_n^{Pos}	Binary variable
BPCY	Biomass product coupled yield
<i>C</i>	Regulation parameter for TDPS
Cit	Citrate
CoA	Coenzyme A
CSM	Complete supplement mixture
CSm	Mitochondrial citrate synthase
CSp	Peroxisomal citrate synthase
dFBA	Dynamic flux balance analysis
DHAP	Dihydroxyacetone-phosphate
D_n	Directionality parameter for reaction <i>n</i>

DNA	Deoxyribonucleic acid
EDTA	Ethylene diamine tetraacetic acid
FBA	Flux balance analysis
FCF	Flux coupling finder
$F_{m,n}$	Non-normalized fraction of the production turnover
FPP	Farnesyl pyrophosphate
$F_{m,n}^R$	Reference turnover fraction value of reaction n for metabolite m
FUM	Mitochondrial fumarase
FUMc	Cytosolic fumarase
FVA	Flux variability analysis
G3P	Glyceraldehyde-3-phosphate
G3PD1ir	Glycerol-3-phosphate dehydrogenase
G3PT	Glycerol-1-phosphatase
GAPD	Glyceraldehyde-3-phosphate dehydrogenase
gCDW	Grams of cell dry weight
GDBB	Genetic design through branch and bound
GDH1	NADPH-dependent glutamate dehydrogenase
GDH2	NADH-dependent glutamate dehydrogenase
GDH3	NADPH-dependent glutamate dehydrogenase
GDLS	Genetic design through local search
GHMT2r	Serine hydroxymethyltransferase
Glx	Glyoxylate
GM	Set of genetic modifications
GM_{DR}	Set of reactions that should be down-regulated
GM_{KO}	Set of reactions that should be knocked-out
GM_{UR}	Set of reactions that should be up-regulated
GND	6-phosphogluconate dehydrogenase
GSMM	Genome-scale metabolic model
h	Number of active reactions consuming metabolite m
HEX	Hexokinase
HIS	L-histidine
HPLC	High-performance liquid chromatography
$hsOMT$	O-methyltransferase
ICD	Mitochondrial isocitrate dehydrogenase
Icit	Isocitrate
ICL	Isocitrate lyase
IDP2	Cytosolic NADP ⁺ -specific isocitrate dehydrogenase
I_m	set of inactive consumers of metabolite m in the reference flux distribution
KanMX	Kanamycin resistance marker
kb	Kilobase
K_{cat}	Enzymatic turnover number
K_m	Enzymatic half-saturation constant

LMOMA	Linear minimization of metabolic adjustment
<i>loxP</i>	Cre recombination site
<i>m</i>	Metabolite index
<i>M</i>	Set of all metabolites
MA	L-malic acid
malCoA	Malonyl-coenzyme A
malonyl-CoA	Malonyl-coenzyme A
MALS	Malate synthase
<i>MCR</i>	Malonyl-CoA reductase
<i>mcr</i>	Malonyl-CoA reductase gene from <i>Chloroflexus aurantiacus</i>
MDH	Mitochondrial malate dehydrogenase
MDHc	Cytosolic malate dehydrogenase
ME	Mitochondrial malic enzyme
MFA	Metabolic flux analysis
MILP	Mixed integer linear programming
MiMBI	Minimization of metabolites balance
M_n	Set containing all the precursors of reaction n
MOMA	Minimization of metabolic adjustment
MTD1	NAD ⁺ -dependent 5,10-methylenetetrahydrofolate dehydrogenase
n	Reaction index
N	Set of all reactions
NAD ⁺	β-nicotinamide adenine dinucleotide
NADH	β-nicotinamide adenine dinucleotide, reduced form
NADP ⁺	β-nicotinamide adenine dinucleotide 2'-Phosphate
NADPH	β-nicotinamide Adenine dinucleotide 2'-phosphate, reduced form
Oaa	Oxaloacetate
<i>OF</i>	Objective function value
ORF	Open reading frames
P	Penalty constant for activated reactions
PAL	Protocatechuic aldehyde
PCK	Phosphoenolpyruvate carboxykinase
PCR	Polymerase chain reaction
PDC	Pyruvate decarboxylase
PDC ⁻	Pyruvate decarboxylase negative
PDH	Pyruvate dehydrogenase
PDH bypass	Pyruvate dehydrogenase bypass
PDHc	Pyruvate dehydrogenase complex
PEP	phosphoenolpyruvate
pFBA	Parsimonious flux balance analysis
PFF	Proportional flux forcing
PFK	Phosphofructokinase
PFL	Pyruvate formate lyase
PGI	Phosphoglucose isomerase
PGK	3-phosphoglycerate kinase
PGMT	phosphoglucomutase

PHB	Polyhydroxybutyrate
P_m	set of reactions that can produce metabolite m
PP3	Lump of reactions producing glyceraldehyde-3-phosphate
PP6	Lump of reactions producing fructose-6-phosphate
PPP	Pentose phosphate pathway
PSP	Phosphoserine phosphatase
PYC	Pyruvate carboxylase
PYK	Pyruvate kinase
R_m^+	Set of active reactions consuming metabolite m in a reference state
R^2	Correlation coefficient
RI	Refractive Index
R_m	Set of all the reactions that can consume metabolite m
ROOM	Regulatory on/off minimization
rpm	Rotations per minute
S	Stoichiometric matrix
SAM	S-adenosyl methionine
<i>SanSyn</i>	Santalene synthase from <i>Clausena lansium</i>
SBML	Systems biology markup language
SCAT	Succinyl-CoA:acetate CoA-transferase
SD	Synthetic dextrose
Ser	L-serine
$S_{m,n}$	Stoichiometric coefficient of metabolite m in reaction n
Succ	Succinate
SucCoa	Succinyl-CoA
SUCD	Succinate dehydrogenase
SUCOAS	Succinyl-CoA ligase
TDPS	Turnover dependent phenotypic simulation
<i>tHMG1</i>	Truncated version of HMG-CoA reductase gene
THRS	Threonine synthase
T_m	Production turnover of the metabolite m
TPI	Triose phosphate isomerase; Metabolites: 13dPG
T_m^R	Reference production turnover for metabolite m
U	Set of reversible reactions
<i>UGT</i>	UDP-glycosyltransferase
URA	Uracil
V	Flux vector
VG	Vanillin β -D-glucoside
V^{LB}	Vector of lower flux bounds
V_n	Flux value of the reaction n
V_n^{Neg}	Flux value in the reverse direction of reaction n
V_n^{Pos}	Flux value in the forward direction of reaction n
V_n^R	reference flux value in reaction n
V^{UB}	Vector of upper flux bounds
VVM	Gas volume flow per unit of liquid volume per minute
$X_{m,n}$	Normalized turnover fraction value of reaction n for metabolite m

YPD Yeast extract, peptone and dextrose
YPGal Yeast extract, peptone and galactose

CHAPTER 1

Introduction

The buildup of knowledge about microbial metabolism and the development of sophisticated genome engineering techniques gave rise to the rational modification of microorganisms in order to use them to biosynthesize chemicals of industrial interest. Recently, the possibility of easily sequencing whole genomes allowed the reconstruction of metabolic models at the genome scale, which can be applied to the design of strain engineering strategies *in silico*.

This thesis focuses on the study and improvement of *in silico* strain engineering methodologies using *S. cerevisiae* as a case study organism. Firstly, the accuracy of the genome-scale models was examined in detail and their flux predictions were improved by manual curation. Secondly, a new simulation methodology was developed to allow the simulation of complex strain designs. Finally, the improved models and the novel simulation method were applied to a practical case study aimed at optimizing the production of acetyl-CoA availability in the cytosol of *S. cerevisiae*.

1.1 Context and Motivation

The use of fermentation to alter the properties of certain foods has been part of our society for thousands of years. For example, the production of cheese, yogurt, bread, beer, wine and vinegar all have an underlying fermentation process. Although the consumption of fermented foods has a long history, it was only in the 19th century that Louis Pasteur described the role of microorganisms in the different fermentation processes. Advances in the understanding of microbial fermentation allowed several chemicals to be produced industrially, including citric acid from *Aspergillus niger*, penicillin from *Penicillium chrysogenum* and L-glutamic acid from *Corynebacterium glutamicum*. The discovery that the genetic information was encoded in the DNA of a living organism [1] followed by the development of recombinant DNA technology [2], were crucial steps for what would become modern biotechnology. Nowadays, there are several classes of industrial products synthesized using biological processes, including: organic acids (e.g., citric acid and lactic acid), alcohols (e.g., 1,3-propanediol and ethanol), hydrocarbons (e.g. farnesene), amino acids (e.g., L-glutamic acid and L-lysine), antibiotics (e.g., penicillins and cephalosporins) and biopharmaceuticals (e.g, insulin and interferon α) [3].

During the early days of biotechnology, the choice of an organism to produce a compound of interest started by screening naturally occurring species for a desired phenotype. For example, the industrial production of penicillin started by the discovery, by Alexander Fleming, that some species of Fungi secrete compounds that inhibit the proliferation of bacteria. Furthermore, the strain improvement processes were mostly based on random mutagenesis followed by screening of strains with enhanced production levels. These techniques were labor intensive but they were the best alternative for strain improving before advanced genetic engineering procedures were developed.

The buildup of knowledge about microbial metabolism and the development of more sophisticated genome engineering techniques gave rise to the rational modification of microorganisms to improve their phenotypical properties towards a certain goal. This methodology is usually referred to as metabolic engineering [4, 5] and has been the subject of a vast number of publications over the past twenty years [6]. A few examples of rational strain designs that resulted in industrial processes include the production of artemisinin (anti-malarial) [7], farnesene (biofuel) and 1,3-propanediol (polymer) [8].

More recently the possibility of sequencing whole genomes allowed the reconstruction of metabolic models at the genome scale. These models can be used to simulate the phenotype of an organism and analyze the change in behavior when genetic modifications are applied. The use of genome-scale models for searching for strain engineering strategies has been growing for the last fifteen years and a considerable amount of methodologies have been developed for this purpose. However, despite the large number of methodologies available for strain optimization purposes, their use is still quite uncommon, which is partly derived from the lack of accuracy of the simulations and unreliable results in quantitative terms. In order to increase the use and acceptance of *in silico* strain engineering methodologies in the metabolic engineering field, it is necessary to carefully analyze the existing methodologies and improve them to the point where researchers will trust simulations over their rationally derived hypothesis.

Baker's yeast, *Saccharomyces cerevisiae*, is a model organism with plenty of physiological information readily available (Saccharomyces Genome Database [9]) and it is the most used eukaryotic organism in metabolic engineering studies [6]. It is a microorganism easy to engineer and cultivate, for which there are plenty of genome engineering tools available [10]. The full genome of *S. cerevisiae* was the first to be made available for a eukaryotic organism [11] and the corresponding genome-scale metabolic model was also the first of its kind for a eukaryote [12]. The accumulated knowledge available for *S. cerevisiae*, in combination with the large amount of literature regarding rational strain engineering strategies, makes it one of the best choices for a case-study regarding *in silico* strain design methods.

1.2 Research aims

The main goal of this thesis was to improve the methodologies used for *in silico* strain engineering by using *S. cerevisiae* as the case study organism. To achieve that goal the accuracy of the genome-scale models was first examined in detail by comparing how well they could predict the metabolic flux distributions in central metabolic pathways. Afterwards, the models were curated manually in order to improve their flux predictions. Using that information, the focus was then the development of a simulation method that could replicate the results of experimental strain designs available from the literature. For that purpose, a new methodology was developed that allows the simulation of complex

strain designs composed of gene deletions and up-/down-regulations. The final goal was the application of the improved models and simulation method to a practical case study aimed at optimizing the production of acetyl-CoA availability in the cytosol of *S. cerevisiae*.

1.3 Outline of the thesis

This thesis is organized in six different chapters: a general introduction and contextualization of the subjects explored in this thesis is presented in the current chapter, the research aims stated above are explored in chapters 2-5 and the final conclusions and recommendations for further work are given in chapter 6. The four chapters covering the research aims were organized as follows:

- In chapter 2, a comprehensive list of computational methods devoted to strain design was collected from the literature and a summary of their main features was compiled. Furthermore, the full list of genome-scale models for *S. cerevisiae* was assembled and an exhaustive search of their applications to engineer the metabolism of this microorganism was performed.
- A comparison of the simulation accuracy of selected genome-scale models of *S. cerevisiae* was performed in chapter 3, in order to verify if faulty flux predictions could lead to erroneous predictions of gene knock-outs. Improvements were applied to selected models by curating the metabolic reactions involved in NADH and NADPH metabolism.
- In chapter 4 a novel simulation method entitled turnover dependent phenotypic simulation is presented with the goal of simulating quantitatively the phenotype of strains with complex genotypes in a resource conscious manner. The developed method was validated using metabolically engineered *S. cerevisiae* strains available in the literature by comparing the production yields of the target metabolites between the simulations and experiments.
- In chapter 5 the turnover dependent phenotypic simulation method (chapter 4) was used in combination with an evolutionary algorithm to search for sets of genetic alterations that would improve the production yield of 3-hydroxypropionic acid derived from acetyl-CoA. The *in silico* suggested modifications were implemented sequentially in *S. cerevisiae*, and the resulting strains were

physiologically characterized in batch fermentations to test the validity of the simulations.

1.4 References

1. Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171:737–8.
2. Cohen SN, Chang ACY, Boyer HW, Helling RB (1973) Construction of Biologically Functional Bacterial Plasmids In Vitro. *Proc Natl Acad Sci* 70:3240–3244.
3. Chotani G, Dodge T, Arbige M (2012) Industrial Biotechnology: Discovery to Delivery. In: Kent JA (ed) *Handb. Ind. Chem. Biotechnol.* SE - 30. Springer US, pp 1131–1182
4. Bailey JE, Birnbaum S, Galazzo JL, et al. (1990) Strategies and Challenges in Metabolic Engineering. *Ann N Y Acad Sci* 589:1–15.
5. Bailey J (1991) Toward a science of metabolic engineering. *Science* (80-) 252:1668–1675.
6. Woolston BM, Edgar S, Stephanopoulos G (2013) Metabolic engineering: past and future. *Annu Rev Chem Biomol Eng* 4:259–88.
7. Paddon CJ, Westfall PJ, Pitera DJ, et al. (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 496:528–32.
8. Nakamura CE, Whited GM (2003) Metabolic engineering for the microbial production of 1,3-propanediol. *Curr Opin Biotechnol* 14:454–459.
9. Cherry JM, Hong EL, Amundsen C, et al. (2012) *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* 40:D700–5.
10. Jensen MK, Keasling JD (2014) Recent applications of synthetic biology tools for yeast metabolic engineering. *FEMS Yeast Res.*
11. Goffeau A, Aert R, Agostini-Carbone ML, et al. (1997) The yeast genome directory. *Nature* 387:5.
12. Förster J, Famili I, Fu P, et al. (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* 13:244–53.

CHAPTER 2

Strain engineering using genome-scale metabolic models

The advent of complete genomic sequences allowed the creation of models containing the full set of biochemical transformations present in an organism. These Genome-Scale Metabolic Models (GSMMs) can be used to simulate the flux distribution of an organism in different environmental and genetic conditions. Therefore, their use in metabolic engineering can be of great help to characterize modified strains and search for possible genetic targets that optimize a certain desirable trait.

Here we explore the computational methodologies available to aid in the strain engineering process using GSMMs by dividing them in three categories according to their practical application. Furthermore, we show how GSMMs have been used to predict interesting genetic targets and create improved strains of *S. cerevisiae*. Although the experimental implementation of strain designs obtained *in silico* has shown good results, this type of strain engineering methodology is still not widespread in the metabolic engineering field.

2.1 Introduction

The genome encodes the information necessary to synthesize all the enzymes present in a cell and consequently the diversity of chemical transformations therein. In 1995, *Haemophilus influenza* became the first free-living organism to have its genome completely sequenced [1], which enabled the reconstruction, a few years later, of the first model of metabolism at the genome scale [2]. With several sequencing projects finishing in the following years, additional genome-scale metabolic models (GSMMs) became available, including for *Escherichia coli* [3] and for *Saccharomyces cerevisiae* [4], with the latter being the first GSMM available for a eukaryotic organism. By 2004 a total of five GSMMs were already available [5] and that number grew to more than 100 as of 2014 [6].

The construction of a GSMM starts by assigning a metabolic function to each gene coding for an enzymatic reaction that is present in the genomic sequence. Furthermore, many additional curation steps are required until the finalized model is ready for simulation purposes. The full process of GSMM reconstruction has been described in detail in several publications [7–10] and software tools that can help in the reconstruction process are available [11–16]. These software tools can be of great help in the reduction of the total time necessary to reconstruct a GSMM and can prove extremely valuable for annotating genomes of less studied organisms. Besides the stoichiometry and reversibility for all chemical reactions that can occur in a certain organism, GSMMs can include additional details, such as the kinetic parameters for each enzymatic reaction [17]. However, the availability of kinetic information is very scarce, which makes it very challenging to gather these data at the genome-scale. Furthermore, dynamical modelling is also computationally intensive, which makes it unpractical for simulating genome-scale networks.

Since the biochemical information included in most GSMMs is limited to the stoichiometry and reversibility of all reactions, the application of these models is restricted to steady-state modelling of intracellular fluxes [18, 19]. The use of stoichiometric models to estimate the flux distribution in a biological network dates back as far as 1969 [20], but further experiments were not reported until the mid-1980's [21, 22] and the field only started to expand in the early 1990's (reviewed in [23]). Stoichiometric modelling [19, 24, 25], also referred in the literature as constraint-based

modelling [18], of genome-scale metabolic networks is an expanding field and has been applied in different areas of the biological sciences, including: the contextualization of high-throughput data, strain engineering, study of host-symbiont relationships, metabolism of pathogenesis and the search for cancer drug targets (reviewed in [7, 18, 26, 27]). A thorough review of the field of constraint-based modelling was published in 2014 [18], which counted over 600 publications in this area of knowledge.

The mathematical foundation for modeling stoichiometric networks is based on the assumption that the system is in a steady state, which allows mass balance constraints to be imposed on metabolite pools as shown in equation 2.1 [19, 22, 23]:

$$S \cdot V = 0 \quad \text{Equation 2.1}$$

Where S is a matrix containing the stoichiometric coefficients for all metabolic reactions (one row for each metabolite and one column per reaction) and V is the flux vector. In other words, equation 2.1 can be converted into a set of linear mass balance equations that impose the production rate of each metabolite to be equal to its consumption rate. This steady-state assumption is based on the fast nature of metabolic changes when compared to growth or environmental changes [23]. Although most metabolites in the system must be completely balanced, an exception is made for selected metabolites that can be absorbed/excreted from/to the medium. In addition to equation 2.1, it is usually assumed that the flux values should be constrained within certain limits, which defines the reversibility of reactions and the magnitude of the input fluxes:

$$V^{LB} \leq V \leq V^{UB} \quad \text{Equation 2.2}$$

Where V^{LB} is the vector of lower flux bounds and V^{UB} is the vector of upper flux bounds.

Mass balance and flux constraints (equations 2.1 and 2.2) are usually not enough to obtain a singular flux distribution because the number of variables (fluxes) is often much larger than the number of mass balance equations (metabolites), which makes the system of equations underdetermined. Using metabolic flux analysis [28] additional flux constraints (such as measured metabolite consumption and excretion rates) can be imposed to try to find the flux distribution that better approximates the experimental values. Alternatively, linear programming can be used to optimize a certain cellular goal (e.g. biomass yield or ATP production), which is the principle of Flux Balance Analysis (FBA) [29]. One

common strategy to simulate the flux distribution of microbial cells with FBA is to maximize the flux through a pseudo-reaction describing the consumption of biosynthetic precursors and energy requirements for synthesizing a specific amount of cellular material [10, 30, 31]. Depending on the type of organism and the specific condition that needs to be simulated, other objective functions can be used [31, 32] and even deduced from experimental data [33].

The simulation of flux phenotypes with FBA or other constraint-based simulation methods provides an interesting alternative for testing the effects of genetic modifications on the phenotypic behavior of an organism [34]. These methods can be of great help in designing strains with desired phenotypic characteristics because the time required for a simulation is considerably faster than the corresponding experimental implementation (reviewed in [35–39]). This chapter provides a detailed analysis of the computational methods available in the literature that can help designing strain engineering strategies by modelling metabolism at the genome scale. Furthermore, the application of *in silico* strain design methods to *Saccharomyces cerevisiae* was explored in detail to provide a context of how GSMMs can help in practice the strain engineering process.

2.2. Strain engineering with GSMMs

Rational metabolic engineering [40] procedures are the result of accumulated biochemical information about a certain organism, which allows the formulation of hypotheses about the effects that a certain genetic modification will provoke on the phenotypical behavior of a cell. One big difficulty of rational strain designs is that metabolic networks are composed of hundreds of interconnected metabolic reactions, which can lead to unforeseeable results. To model the effect of genetic modifications at the system scale, the most suitable methodology is constraint-based modelling of genome-scale networks [18], which allows the calculation of flux distributions in steady state conditions for a certain genotype.

As mentioned in the introduction, most of the GSMMs available are stoichiometric representations of metabolism and their foundation lies on the mass balances of metabolites. Although the core assumption is quite simple, several methods of variable complexity have been developed with the goal of helping to analyze and modify the

metabolic capabilities of the cell. Previously, some compilations of the available constraint-based methodologies have been made by classifying the methods phylogenetically [34], by the types of modifications they allow [38, 39], or by the underlying objective function [37]. The classification used here is based on the utility of each method for different strain engineering tasks. Figure 2.1 shows the three categories we used to group the different strain engineering constraint-based methods available in the literature: the first category gathers the methods that return a flux phenotype for a given set of genetic modifications (phenotype simulation), the second group of methods is composed of strain optimization methods that yield a set of strain engineering modifications that should optimize a certain metabolic engineering goal (target discovery), and the third category includes methods that can help in strain engineering but are not related with the search or analysis of genetic targets (auxiliary network analysis).

Although the methods on each category should serve different practical purposes, some of them could conceptually belong to several categories simultaneously. For example, most target discovery methods have an underlying simulation routine that could also be placed in the phenotype simulation category. However, quite often the simulation part of the target discovery methods is just a version of FBA used to search for genetic targets that optimize the desired product. Therefore, each method was classified into a single category from Figure 2.1, according to its main purpose of application.

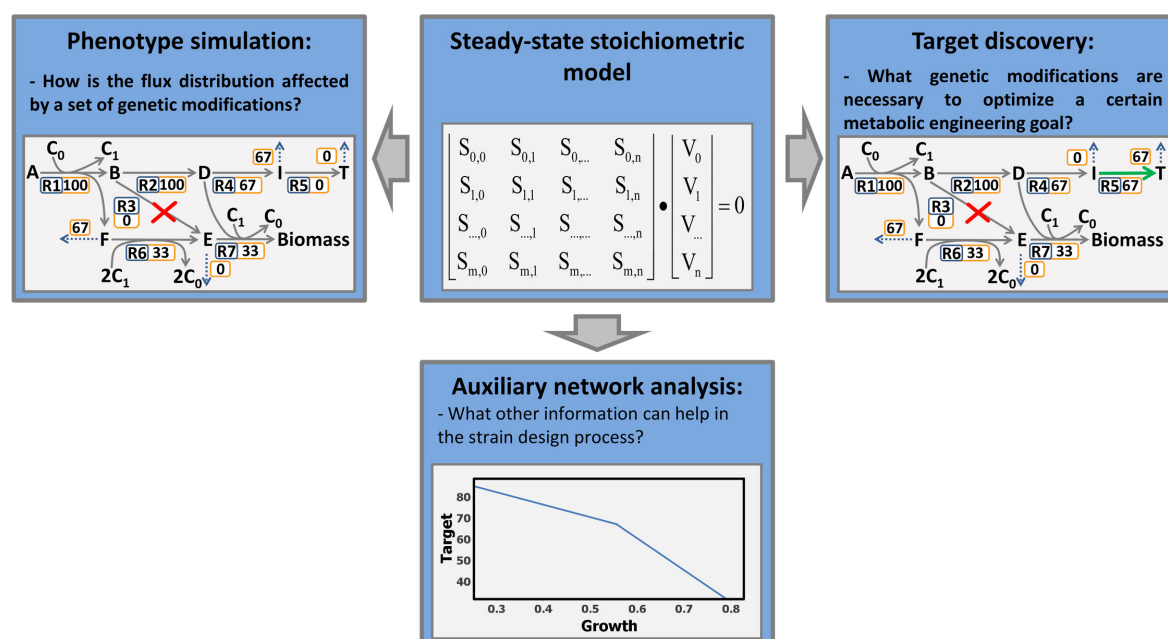


Figure 2.1- Different categories adopted for classifying the computational methods used for strain engineering with genome-scale metabolic models.

2.2.1 Phenotype simulation methods

One of the issues usually encountered when designing metabolically engineered strains with desired phenotypical traits is the appearance of unexpected effects. Given the complexity of metabolic pathways, it is often hard to understand what caused the undesired behavior, resulting in additional experiments to elucidate the underlying causes. The estimation of fluxes using labelled substrates (reviewed in [41, 42]) is one of the possibilities to characterize the fluxes of a mutant organism, but it is a troublesome and expensive procedure. Alternatively, the effects of a set of genetic modifications can be easily simulated using constraint-based computational methods, which can help pinpoint the reasons behind an undesirable phenotype. Furthermore, a careful analysis of the simulated fluxes might yield additional non-obvious genetic targets that can help achieving the desired goal.

Table 2.1 shows a list of the available methods capable of simulating flux phenotypes using GSMMs and a short description of their main attributes. FBA was the first method developed for simulating stoichiometric models and its invention predates the appearance of GSMMs. Briefly, FBA finds a distribution of fluxes that meets a certain objective formulated with linear programming. The most common objective function used to simulate the fluxes of microbial GSMMs is the maximization of the biomass yield, which is based on the presumption that evolution has selected for phenotypes that use limited resources efficiently for growing and dividing [10, 30, 31]. One common drawback of the flux distributions obtained with FBA is that they are usually not unique, i.e., for the same value of the objective function there are several reactions for which the metabolic fluxes can vary. Parsimonious FBA (pFBA) was developed to further refine the flux distribution obtained with FBA by removing most futile loops from the network, which can be achieved by minimizing the total sum of fluxes in the network or the number of active reactions [43, 44]. These assumptions are based on the efficiency of metabolic networks, which are expected to perform a certain task with a minimal amount of resources expended.

Table 2.1- Phenotype simulation methods available for genome-scale metabolic models

Method	Brief description	References
FBA	Linear programming simulation method that can estimate the flux distribution in a genome-scale network by assuming a certain cellular objective. Flux distributions of microorganisms are usually calculated by assuming maximal biomass formation. Can be used to predict wild-type phenotypes and the growth outcome of gene knock-outs and their associated flux distribution.	[2, 21, 22, 29]
pFBA	Bi-level formulation intended to refine the flux distribution obtained with FBA by minimizing the number of active enzymes or the total sum of fluxes.	[43, 44]
MOMA	MOMA assumes that the flux distribution of a deletion mutant should stay as similar as possible to the wild-type organism. Therefore, MOMA computes the flux distribution for a mutant metabolic network by minimizing the Euclidean distance between a reference set of fluxes (for the wild-type strain) and the mutant fluxes.	[45]
LMOMA	Same as MOMA but the minimization in LMOMA is referent to the Manhattan distance instead of the Euclidean distance.	[46]
PSEUDO	This formulation is similar to MOMA because it also minimizes the Euclidean distance between the mutant organism and the wild-type. However, instead of using a single flux distribution as the reference, PSEUDO uses a degenerate optimal region of the flux space delimited by a minimal threshold imposed on biomass production.	[47]
ROOM	This formulation computes the flux phenotype of a deletion mutant by minimizing the number of reactions that are activated or inactivated in comparison to a reference flux distribution.	[48]
MiMBI	This formulation assumes that a deletion mutant will attempt to minimize the changes in the turnover of all metabolites (sum of producing and consuming fluxes) in comparison to a reference flux distribution.	[49]
Under/overexpression plugin for OptFlux	This formulation emulates the over and under expression of genes by imposing flux constraints based on a reference flux distribution. The mutant's flux distribution can be simulated by choosing any of the available objective functions in OptFlux (e.g. FBA, MOMA and LMOMA).	[50, 51]

Some of the uses given to FBA include the prediction of gene essentiality [2, 3, 52, 53], the growth rate of a knock-out mutant after directed evolution [54] and sets of knock-outs that couple the production of a certain product to growth [55, 56]. The use of an objective function in FBA that assumes maximum biomass formation usually results in initial growth levels of engineered strains lower than the simulations suggest. However, it has been shown that after directed evolution, the growth and production levels of engineered strains may converge to the values simulated with FBA [54–56]. Although FBA is mostly used on its own, some complementary methods have been developed to include regulatory information [57–60], transcriptomic data [61] and thermodynamic constraints [62, 63].

As discussed above for FBA, the assumption of optimality for biomass formation can lead to unrealistic predictions of the growth phenotype in knock-out mutants. In order to predict more accurately how a metabolic network reacts to perturbation, several more

objective functions have been developed (Table 2.1). Segrè et al. developed a formulation to simulate the effects of knock-outs on a metabolic network designated Minimization Of Metabolic Adjustment (MOMA) [45]. This methodology assumes that when an organism is faced with a gene deletion, it will try to minimize the adjustments of the flux values in comparison with a wild-type reference. MOMA formulates this objective function mathematically as the minimization of the Euclidean distance between the mutant set of fluxes compared to the reference wild-type fluxes (Figure 2.2). The growth predictions of gene knock-outs simulated with MOMA are more conservative than the results obtained with FBA and it has been shown that MOMA can predict more accurately gene essentiality in some cases [45, 64]. MOMA has also been used for prediction of metabolic engineering targets and many of them have been implemented experimentally with varying degrees of success [65–69].

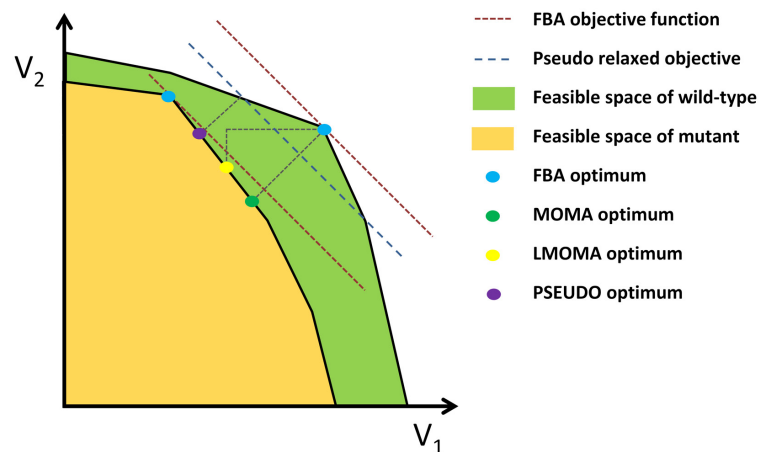


Figure 2.2- Simplified representation of the differences between the objective functions used in FBA, MOMA, LMOMA and PSEUDO. V_1 and V_2 represent two arbitrary fluxes in a network and the green area delimits the feasible flux space for the wild-type organism, while the orange area is correspondent to the mutant feasible space. The FBA objective function is represented by a dashed brown line and the optimal value of V_1 and V_2 in an FBA simulation is shown as a blue dot. The different distance metrics (grey dashed lines) are shown in the 2D plot along with the optimal objective function values for each method (colored dots). The relaxed region of the flux space used in PSEUDO is represented as a dashed blue line. Adapted from Segrè et al. [45].

In addition to the normal MOMA formulation, two additional variations are available in the literature: linear MOMA (LMOMA) [46] and PSEUDO [47]. One common issue usually encountered in the flux distributions computed with MOMA is that this formulation favors a large number of small flux changes in detriment of a few large changes in the metabolic network. This is caused by the quadratic formulation used to calculate the flux distance in MOMA and can be solved by using LMOMA, which uses the Manhattan distance between the reference and perturbed network to find the flux

phenotype of the knock-out mutant (Figure 2.2). Another issue that might arise from using MOMA/LMOMA is the importance given to a reference flux distribution. Usually, the reference set of fluxes is calculated using FBA or pFBA [65–67] and any error in this flux distribution will be propagated to all the predictions. The methodology developed in PSEUDO can tackle this issue by not using a single flux distribution as a reference but a region of the flux space delimited by a minimum threshold imposed on the biomass yield (Figure 2.2) [47]. The authors of this methodology reported some improvements in comparison to MOMA and FBA flux predictions [47].

Shlomi et al. pursued the same concept of minimal metabolic adjustment but in a different perspective [48]. Instead of minimizing the flux differences between the mutant and the wild-type organism, the methodology entitled ROOM (Regulatory On/Off Minimization) minimizes the number of reactions that are activated or deactivated in a mutant in comparison to a reference flux distribution (Table 2.1). The assumption behind ROOM is that, when faced with a set of knock-outs, a cell will adjust its internal fluxes by making the minimum amount of regulatory changes, i.e., the magnitude of the fluxes can change, but the set of active enzymes should be similar to the wild-type organism. The predictions obtained for ROOM were closer to FBA than MOMA and revealed that MOMA is better at estimating transient metabolic adaptations, while FBA and ROOM can better predict the phenotype of an evolved knock-out mutant [48].

Another issue encountered in the formulation of MOMA and LMOMA was the dependence of the mutant phenotype on the scale of the stoichiometry of the metabolic reactions [49]. By using different stoichiometric representations of a metabolic network that are biochemically equivalent, Brochado et al. showed that the simulation outcome of MOMA/LMOMA was sensitive to the stoichiometric representation chosen for the network [49]. Since biochemically equivalent networks should produce the same results, the authors propose a new methodology entitled Minimization of Metabolites Balance (MiMBI). The formulation underlying MiMBI solves the stoichiometry dependence of other algorithms by using the metabolite turnovers as the variables in the objective function. Instead of minimizing the changes in the fluxes in comparison to a reference network, MiMBI minimizes the changes in the turnovers of all metabolites in the network. As a consequence, MiMBI provides more robust results, which are not dependent on the numerical stoichiometric representation chosen to describe a metabolic network.

The final method shown in Table 2.1 was included as a representative of how other genetic modifications can be incorporated in the simulation methods discussed. The under/ overexpression plugin [50, 51] can be used in combination with the simulation methods described in Table 2.1 to predict the effect of up- or down-regulating certain fluxes in the network. This methodology uses a wild-type flux distribution as a reference to estimate how over or underexpressing a certain gene will affect the corresponding fluxes. After applying constraints to the modified fluxes, this method calculates the flux distribution of the mutant by using one of the objective functions available in OptFlux.

2.2.2 Target discovery methods

While the simulation of phenotypes *in silico* can be of great help for analyzing the internal fluxes of a few rationally designed strains, it would take too long to manually discover combinations of genetic modifications that meet a certain metabolic engineering objective. In order to search for interesting genetic targets, several methods have been developed to search among the large number of different strain engineering strategies for the ones that result in the desired phenotype. A compilation of the methods available in the literature for discovering potential metabolic engineering targets using GSMMs is shown in Table 2.2.

Table 2.2- Computational methods available for target discovery using genome-scale metabolic models

Method	Brief description	References
OptKnock	Bi-level optimization method that has an FBA inner problem that optimizes the biomass yield and an outer problem that maximizes the production of a target compound. Given a maximum number of allowed knock-outs, this method returns a single optimal strain design.	[70]
Robustknock	FBA simulations can suffer from alternative optimal solutions and, as a consequence, the inner problem of OptKnock might return a strain design with variable production levels of the target metabolite. In RobustKnock the strain designs returned are guaranteed to be robust.	[71]
ReacKnock	Similar formulation to OptKnock but the Karush-Kuhn-Tucker method was used to reformulate the bi-level optimization problem to a single level one.	[72]
Objective tilting	FBA simulations can suffer from alternative optimal solutions and, as a consequence, OptKnock and OptGene strain designs might suffer from variable production levels. This method applied modifications to OptGene and OptKnock to guarantee a robust production level.	[73]
OptORF	Bi-level optimization method that has an FBA inner problem that optimizes the biomass yield and an outer problem that maximizes the production of a target compound. Transcription regulation further constraints the search for strain designs composed of gene deletions and over-expressions.	[74]
OptSwap	Bi-level optimization method that has an FBA inner problem that optimizes the biomass yield and an outer problem that maximizes the production of a target compound. This method returns an optimal strain design composed of reactions with changed cofactor specificity and knock-outs.	[75]

Table 2.2- Computational methods available for target discovery using genome-scale metabolic models (continuation)

Method	Brief description	References
FaceCon/ ShadowCon	These modules can be added to optimizations methods such as OptKnock and OptORF to increase the diversity of strain designs obtained. Possible applications include designing strains with by-product inhibition, variable coupling constraints and co-production of two metabolites.	[76]
FastPros	This method uses the shadow prices to find strain designs composed of knock-outs that result in the overproduction of a target metabolite under biomass maximization. The optimal combination of knock-outs can be found with other methodologies (e.g. OptKnock) by restricting the search space to the best knock-out candidates obtained with FastPros.	[77]
BiMOMA	Bi-level optimization method that has a MOMA inner problem that minimizes the distance to a wild-type reference and an outer problem that maximizes the production of a target compound. Given a maximum number of allowed knock-outs, this method returns a single optimal strain design.	[78]
OptGene	Optimization based on evolutionary algorithms that was designed to find sets of knock-outs using one of the available simulation methods for the inner problem (authors tested FBA and MOMA). OptGene returns several near optimal strain designs that maximize a target compound.	[51, 79]
SA/SEA	Evolutionary and simulated annealing algorithms were designed to search for near optimal sets of gene deletions that maximize the production of a target compound. Similar to OptGene, but sets were used to represent genetic alterations to allow strain designs with flexible size.	[51, 80]
CiED	Optimization based on evolutionary algorithms that was designed to find sets of knock-outs using FBA. Adopts a different heuristic from OptGene that is capable of retaining beneficial mutations and performs an assessment of the frequency that each modification appears in the optimal designs.	[81]
GDMO	Multi-objective optimization algorithm that searches for gene deletions/nutrient availability that maximize the production of a target metabolite in addition to other objectives (e.g. biomass formation). GDMO returns a pareto front with the best tradeoff between the objectives and uses FBA as the underlying flux calculation method.	[82]
GDLS	Optimization method based on local search that offers a much faster alternative to global search methods (e.g. OptKnock). Can be used for finding sets of gene deletions of larger sizes in feasible computation times.	[83]
GDBB	This method uses a truncated branch and bound algorithm that can replace the bi-level formulation present in similar strain design methods (e.g. OptKnock and GDLS). GDBB can find near-optimal solutions (knock-outs or others) in comparatively much shorter times.	[84]
OptReg	Bi-level optimization method that has an FBA inner problem that optimizes the biomass yield and an outer problem that maximizes the production of a target compound. OptReg finds the optimal strain design composed of reaction deletions and up/down-regulations.	[85]
EMILiO	This method uses successive linear programming to find flux constraints (upper and lower bounds) that maximize the production of a target compound (while maximizing biomass with FBA). Genetic modifications are modelled by constraining fluxes to specific optimal levels instead of pre-calculated levels like in OptReg.	[86]
Redirector	This method manipulates the objective function of FBA to model genetic up- and down-regulations. It is used to optimize the production of a target metabolite by using GDLS to find near-optimal strain designs composed of up- and down-regulations.	[87]
OptStrain	Optimization method that first searches for heterologous reactions which need to be added to a certain host in order to allow the production of the target product. Subsequently, OptKnock is used to find reaction deletions that optimize the product formation.	[88]
SimOptStrain	Similar to OptStrain, but the addition of heterologous reactions and search for reaction knock-outs is done simultaneously instead of sequentially.	[78]

Table 2.2- Computational methods available for target discovery using genome-scale metabolic models (continuation)

Method	Brief description	References
OptForce	Unlike most strain optimization methods, OptForce does not maximize biomass formation to find the flux distribution of the mutant strain. In this case, OptForce finds the minimal set of fluxes that must be forced (up/down-regulated and knocked-out) to make sure the target metabolite is produced at optimal levels.	[89]
CosMos	Similar to OptForce, but the fluxes can be forced to any value instead of predetermined intervals.	[90]
k-OptForce	Similar to OptForce but kinetic information is added to the network where available. k-OptForce finds the minimal set of fluxes that must be forced (up/down-regulated and knocked-out) as well as the minimal set of enzyme parameters that must be changed to optimize the target metabolite.	[91]
OP-Synthetic	This method searches for changes on the flux limits calculated with FVA [92] by increasing step-by-step the production of a target metabolite. When a flux bound violates the wild-type flux limits it is assumed that this reaction is a possible target for up/down-regulation or deletion depending on the direction of change compared to the reference interval.	[93]
FSEOF	Methodology that scans the changes in the fluxes that are predicted by FBA when the production of the target metabolite is forced to increase step-by-step. The fluxes that increase with the enforced production of the target are assumed as possible up-regulation candidates.	[94]
FVSEOF	While FSEOF searches for changes in the fluxes predicted by FBA, FVSEOF searches for changes in the flux bounds calculated with FVA. This methodology also returns possible candidates for up-regulation.	[95]
SIMUP	Bi-level optimization method with an FBA inner problem that optimizes the biomass yield and an outer problem that promotes the co-utilization of two sugars by searching for knock-out combinations that are lethal when only one substrate is available.	[96]
FOCAL	Bi-level method that finds media conditions and knock-out combinations that make a target flux coupled to another (e.g. growth). Can also be applied to strain engineering for substrate co-utilization or other strategies that require flux coupling.	[97]
DySScO	First the GDLS method is used to find strain designs composed of knock-outs that optimize a target product. Then, each strain is tested with dFBA [98] to estimate their titer and productivity in batch cultures.	[99]

Among the methods shown in Table 2.2, the first method developed for target discovery was OptKnock and its formulation consisted in the search for combinations of knock-outs that result in an increased production of a target metabolite [70]. OptKnock is a bi-level optimization method that uses FBA to calculate the phenotype of a certain combination of knock-outs by assuming maximum biomass formation. The result returned by OptKnock is the best combination of knock-outs that maximize the engineering goal, while taking into account a maximum number of knock-outs and a minimum biomass formation rate. Some of the results obtained with this optimization algorithm have been used to find knock-out targets that resulted in the construction of strains with improved production levels of lactic acid, 1,4-butanediol and 2,3-butanediol [55, 100, 101].

One of the problems encountered in OptKnock was the possibility that the combination of knock-outs returned would result in a strain with variable production rates *in silico* [71, 73]. This variability is caused by the non-uniqueness of the flux distribution obtained from the FBA simulation layer, which can result in a non-robust production flux for the target metabolite. The mutants suffering from non-robustness usually can produce alternative byproducts instead of the target metabolite without impacting the maximum value of the biomass yield simulated with FBA. This robustness issue was addressed in the formulation of RobustKnock [71] and in the utilization of objective function tilting [73], which resulted in improved optimization methods that returned strain designs with guaranteed production levels *in silico*. An additional issue has also been reported regarding the transformation method used to convert the bi-level problem of OptKnock into a single level optimization problem [72]. Xu et al. tried to solve this issue by using the Karush-Kuhn-Tucker method to reformulate the bi-level optimization problem into a single level one, which resulted in a new method designated ReacKnock [72].

In addition to the above-mentioned optimization methods, there are additional variations of the OptKnock formulation that include special features that can be of great help in the design of metabolically engineered strains. Besides searching for metabolic gene knock-outs, OptORF can also use regulatory information available for the metabolic network in order to find modifications at the regulatory level that improve the production of a target metabolite [74]. Furthermore, there is an optimization method (OptSwap) that maximizes the production of a metabolite of interest by suggesting changes in cofactor (NADH and NADPH) specificity of oxidoreductase enzymes in addition to knocking out sets of reactions [75].

Table 2.2 also includes software modules, such as FaceCon/ ShadowCon, that can be added to optimization methods that search for the globally optimal strain design (e.g. OptKnock and OptORF) in order to introduce variability and include additional desirable traits in the strain designs obtained [76]. Furthermore, to shorten the computational time needed to find globally optimal solutions, the number of possible knock-out targets can be reduced beforehand with FastPros [77].

All optimization methods discussed so far have FBA at the center of their bi-level formulation. However, there is also a version of OptKnock that has the MOMA simulation method included in the inner simulation layer [78]. The method, designated by

the authors as BiMOMA searches for knock-outs that maximize a certain metabolic engineer goal, but instead of maximizing biomass formation, it simulates the mutant phenotypes by minimizing the distance to a wild-type flux distribution.

One important drawback of the target discovery methods that search for a global optimum solution like OptKnock is that, as the number of allowed genetic modifications increases, the total searchable space of strain designs grows exponentially, which makes the computation time required to solve the problem impractical. This severely limits the maximum number of genetic modification that can be included in the strain designs computed with OptKnock and similar methods. One of the possibilities to solve this limitation is to use evolutionary algorithms or other nature-inspired heuristics to find strain designs with desired phenotypes. The first evolutionary optimization method published was OptGene [79] and since then many other became available (SEA [51, 80], CiED [81], GDMO [82]). Evolutionary algorithms can search for strain designs with larger sets of genetic modifications, can be used in combination with multiple simulations methods (such as FBA, MOMA or LMOMA) and require less computation time to reach near-optimal design solutions. Further advantages of evolutionary algorithms include the possibility of optimizing non-linear objective functions (e.g. resembling productivity) and the fact that several near-optimal strain designs are suggested by the algorithm, instead of a single global optimum like in OptKnock-based methods. One possible drawback of this type of optimization is that the near-optimal strain designs found by the evolutionary algorithm might be distant from the global optimum. In parallel with OptKnock, the use of evolutionary algorithms to find interesting genetic targets has also resulted in the construction of several strains with improved production levels of metabolites of interest [56, 69, 81, 102].

In addition to evolutionary algorithms, there are also other methods that have been developed to help in the search for strain designs of larger sizes. Genetic Design through Local Search (GDLS) was developed as an alternative methodology to find near-optimal sets of knock-outs using a local search methodology [83]. When compared to evolutionary algorithms, the GDLS heuristic was able to achieve higher *in silico* yields for certain case-studies [83]. Also developed as an alternative for exact algorithms that search for globally optimal solutions, the Genetic Design through Branch and Bound (GDBB) uses a truncated branch and bound algorithm that outperforms such methods in terms of computational time required [84].

Although most computational methods that search for interesting genetic targets were initially focused on knock-outs, many of them have included in their formulation additional types of genetic modifications. OptReg was the first method to allow the optimization of strains composed of up- and down-regulations in addition to knock-outs [85]. Similarly to OptKnock, the formulation of OptReg originates a bi-level problem with FBA as the inner simulation layer. The methodology used by OptReg to model up/down-regulation is based on restraining the fluxes of the target reactions to values over or below the levels calculated beforehand. These flux constraints are applied to the FBA simulation layer and biomass formation is maximized to obtain a flux distribution for the mutant strain. OptReg returns a single optimal strain design, which represents the combination of genetic modifications that results in the best production yield of the target metabolite.

Strain designs consisting of knock-outs and up/down-regulations can also be obtained with an optimization method developed by Yang et al. [86]. This target discovery method, designated EMILiO, models up/down-regulations using a formulation that allows a more flexible manipulation of the flux values when compared to OptReg [86]. In contrast to OptReg and EMILiO, Rockwell et al. developed an optimization method named Redirector that does not manipulate flux values to model up- and down-regulations [87]. Instead, this novel method manipulates the objective function of the FBA simulation to model how up- and down-regulations affect the cellular phenotype. Redirector uses GDLS as the optimization layer in order to find near-optimal combinations of genetic modifications that increase the production of the target metabolite.

Some experimental strains designs include in their list of genetic modifications the addition of heterologous genes, which can enable the synthesis of non-native metabolites of interest. OptStrain [88] and SimOptStrain [78] are able to search for the minimal amount of gene additions necessary to synthesize non-native compound of interest. Furthermore, these methods can optimize the production of the target compounds by searching for sets of knock-outs that result in an improved phenotype.

Most of the target discovery methods discussed above use FBA to simulate the phenotype of the mutant strains being optimized. These methods search for sets of genetic modifications that make the production of a target metabolite coupled to growth, which should force the mutated organism to excrete the desired target as a byproduct of growth.

However, there are some methods that search for sets of genetic modifications that “force” the production of a target metabolite independently of the growth levels. OptForce is an optimization method that finds the minimal set of fluxes that must be constrained (up/down-regulated and knocked-out) to make sure the target metabolite is produced at optimal levels [89]. In biological terms, the flux modulations suggested by OptForce must be met quantitatively in order to guarantee that the phenotype will be as the simulations predict. Optimization methods similar to OptForce have been developed with an increased flexibility in the flux modulations (CosMos) [90] or the inclusion of optimized sets of enzymatic parameters (k-OptForce) [91]. The application of the OptForce algorithm has resulted in the construction of an *E. coli* strain with improved malonyl-CoA levels and increased naringenin production [103].

Another class of optimization methods that does not use biomass maximization as the objective function is composed of the methods OP-Synthetic [93], FSEOF [94] and FVSEOF [95]. All these methods use similar strategies to find genetic targets of interest by analyzing the flux changes in the network when the production of the target metabolite is progressively enforced. The FSEOF and FVSEOF optimization methods return a set of up-regulation targets suggested by the flux analysis, while the OP-Synthetic algorithm recommends up/down-regulations and knock-outs that should result in an increased production of the metabolite of interest. FSEOF has been used to search for up-regulation targets that resulted in the construction of an *E. coli* strain with improved lycopene production [94].

The application of target discovery algorithms is not limited to the increase in production of metabolites of interest. In fact, there are also some methods that can predict genetic alterations that promote the co-utilization of two substrates simultaneously (SIMUP [96], FOCAL [97]). These methods can be helpful in the optimization of industrial strains that will be grown on complex feedstocks, such as lignocellulosic biomass. Another optimization method, DySScO, can be used to find knock-out strains that have important industrial characteristics such as high productivity and titer of a metabolite of interest [99].

2.2.3 Auxiliary network analysis methods

In addition to the two categories of computational methods described above (phenotype simulation and target discovery), we also considered a third group that includes additional methodologies which can also help in the design of engineered strains. One of the methods we included in this category is the analysis of flux variability (FVA) [92, 104], which returns the interval of values that each flux in the metabolic network can assume. After a set of genetic modifications is obtained from a target discovery method, the resulting mutant strain can be analyzed with FVA to test the robustness of the fluxes of interest. FVA is usually employed by forcing the value of the objective function of the FBA simulation (e.g. biomass) to be equal to its optimal value, and then the objective function becomes the minimization or maximization of each individual flux. FVA can also be used to create a graph that relates the maximum yield of the product of interest for different levels of biomass formation. The analysis of the results can reveal the best tradeoff between biomass formation and target production.

The relationship between two fluxes can also be studied by using phenotype phase planes [105, 106] and the flux coupling finder (FCF) methods [107]. One of the applications of a phenotype phase plane is to verify how the ratio of availability of two nutrients can affect biomass growth. This type of analysis can lead to the optimization of culture conditions based solely on the stoichiometry of the metabolic network. Regarding FCF, this method can be used to study how two fluxes in the network depend on each other. One possible use for this method is to search for equivalent reaction knock-outs, which can prove helpful when one of the required knock-outs from a strain design is difficult to implement experimentally.

Another helpful computational method for optimizing the culture conditions of a microorganism was developed by Suthers et al. [108], which consists in the search for the minimal set of metabolites necessary to sustain growth. This methodology can prove helpful in the design of culture media for less studied microorganisms or for determining auxotrophies of mutant strains.

Dynamic FBA (dFBA) allows testing the behavior of a strain in non-steady state environmental conditions, such as batch or fed-batch [98]. This method can help choosing between different strain designs to conclude about which one would be the most suited for an industrial bioprocess.

Additional analysis can be carried on the sensitivity of a flux distribution to changes in the turnover of a metabolite (flux-sum [109]), on the minimal set of genes necessary to sustain a certain growth level [110] or for pathway design (e.g., BNICE [111], SimPheny Biopathway Predictor[100], GEM-Path [112], DESHARKY [113] and retrosynthesis algorithm [114]). The resulting pathways obtained from pathway design algorithms can then be included in GSMMs in order to test them and optimize the production of the heterologous target.

Many of the methods described for simulating phenotypes, discovering metabolic engineering targets and analyzing GSMMs can be found compiled within software packages, such as the COBRA toolbox [115], OptFlux [51] and the BioMet toolbox [116]. This kind of software packages can be of great help for users that are giving their first steps in this field, especially for people with an experimental background that want to try some of the algorithms available for analyzing and designing strains.

2.3. Strain Engineering of *S. cerevisiae* using GSMMs

Baker's yeast, *S. cerevisiae*, is a model organism with plenty of physiological information readily available (Saccharomyces Genome Database [117]) and it is the most used eukaryotic organism in metabolic engineering studies [118]. It is a microorganism easy to engineer and cultivate, for which there are plenty of genome engineering tools available [119]. The full genome of *S. cerevisiae* was the first to be made available for an eukaryotic organism [120] and the corresponding genome-scale metabolic model was also the first of its kind for a eukaryote [4].

Most strain engineering studies in *S. cerevisiae* still rely on the use of rational approaches to optimize the production of compounds of interest [118, 121, 122]. However, with the development of a vast number of computational methods for *in silico* strain design (section 2.2) the number of metabolic engineering studies based on modelling approaches will tend to increase.

2.3.1 Genome scale models of *S. cerevisiae*

Since the first GSMM of *S. cerevisiae* was published in 2003 [4], several authors have been trying to expand and/or improve the metabolic reconstruction of this microorganism (as reviewed in [123, 124]). A list of the genome-scale models available in the literature for *S. cerevisiae* is given in Table 2.3, focusing on model size and other relevant parameters. The numbers show that over the years there was a tendency to include more Open Reading Frames (ORFs) and metabolic reactions in the models. Furthermore, new compartments were added and the elemental balance of the reactions was implemented in some models. The high number of models available for *S. cerevisiae* also points out to the relevance of this microorganism for the metabolic engineering community.

Table 2.3- List of the available genome-scale models for *S. cerevisiae*

Model	Year	ORFs included	Reactions included	Other information
<i>i</i> FF708 [4]	2003	708	1175	3 compartments (2 metabolic)
<i>i</i> ND750 [125]	2004	750	1489	8 compartments (7 metabolic) Elementally balanced
<i>i</i> LL672 [126]	2005	672	1038	3 compartments (2 metabolic)
<i>i</i> MH805 [127]	2006	805 (55 regulatory genes)	1489	8 compartments (7 metabolic) Elementally balanced
<i>i</i> IN800 [128]	2008	800	1446	3 compartments (2 metabolic)
Yeast 1 [129]	2008	832	1857	15 compartments (7 metabolic) Elementally balanced
<i>i</i> MM904 [52]	2009	904	1412	8 compartments (7 metabolic) Elementally balanced
Yeast 4 [130]	2010	924	2576	15 compartments (7 metabolic) Elementally balanced
Kinetic model [17]	2010	832	956	2 compartments (1 metabolic)
Yeast 5 [131]	2012	918	2110	15 compartments (7 metabolic) Elementally balanced
<i>i</i> TO977 [132]	2013	977	1566	4 compartments (3 metabolic)
Yeast 6 [133]	2013	900	1888	15 compartments (7 metabolic) Elementally balanced
Yeast 7 [134]	2013	916	3493	16 compartments (7 metabolic) Elementally balanced

Most of the available genome-scale models for *S. cerevisiae* are purely stoichiometric, but there are two exceptions. The *i*MH805 [127] model is derived from the *i*ND750 [125], but it contains 55 genes that code for transcription factors, which can be used for studying regulation at the genome-scale. Furthermore, there is one kinetic model at genome-scale for yeast that was derived from Yeast 1 [129] using a strategy described in [17, 135]. More recently, a new generation of genome-scale models that include gene expression

values in their formulation, named ME-models, have become available for some microorganism, but none is yet available for *S. cerevisiae* [136].

2.3.2 Examples of application

As stated above, the use of strain engineering strategies derived from GSMMs is still the exception and not the rule in metabolic engineering. Table 2.4 shows the full extent of the application of the tools described in this chapter to engineer *S. cerevisiae*. Nevertheless, the results show some promising implementations, and some of them resulted in the increase of production of high value compounds such as cubebol, vanillin and amorphadiene.

Regarding the use of GSMMs for strain engineering, five out of the nine reports shown in Table 2.4 used *iFF708*, which reveals a tendency to use models that suffered more validation rather than the most recent models. Three of the metabolic engineering strategies shown were obtained with OptGene, showing a clear preference of this algorithm over OptKnock and recent optimization strategies. Table 2.4 also shows an example of the use of FBA to calculate the maximum yields of two pathways, which proved to be helpful in choosing the strategy less dependent on oxygen availability (3-hydroxypropionic acid case-study). This kind of data might seem irrelevant at the lab scale, but for industrial processes the costs involved in aerating bioreactors might be crucial in determining if a biotechnological process is lucrative.

Another interesting observation from the results shown in Table 2.4 is the number of experimental implementations that include the engineering of cofactor abundance. Three of the nine strain designs have attempted to increase the availability of NADPH by deleting one of the main consumers of this metabolite [69, 102, 137]. In the first case this resulted in a reduced accumulation of an unwanted byproduct, glycerol, while in the other two the increase of NADPH availability was important to improve the flux to the target metabolites.

Table 2.4- Experimental implementation of strain engineering strategies obtained using genome-scale metabolic models

Target	Brief description	Reference
Ethanol	<i>iFF708</i> was used to test some strategies aimed at reducing glycerol formation and increasing the ethanol yield in anaerobic conditions. The best strategy consisted in the expression of a NADP ⁺ -dependent glyceraldehyde-3-phosphate dehydrogenase, which resulted in a decrease of 40 % in glycerol accumulation and an increase of 3 % on the ethanol yield when growing on glucose. The ethanol yield increased 25% when <i>S. cerevisiae</i> was grown on xylose and glucose.	[137]
Cubebol	OptGene was used to find gene knock-outs that increased the production of cubebol using MOMA in the <i>iFF708</i> model. As a result, the deletion of <i>GDH1</i> increased by 85 % the cubebol titer.	[102]
Formic acid	FBA was used with <i>iND750</i> to simulate the best combination of three gene deletions that would increase the production of formic acid on a formate dehydrogenase negative basal strain. The deletion of <i>ALT2</i> , <i>ZWF1</i> and <i>FUM1</i> resulted in a 16.5-fold increase in formic acid yield.	[138]
Vanillin	OptGene was used to find gene knock-outs that increased the production of vanillin using MOMA in the model <i>iFF708</i> . The deletion of <i>PDC1</i> and <i>GDH1</i> resulted in up to 2-fold improvement in the vanillin yield on glucose limited chemostats.	[69]
2,3-Butanediol	OptKnock was used to find gene knock-outs that increased the production of 2,3-Butanediol using the model <i>iMM904</i> . The deletion of <i>ADH1</i> , <i>ADH3</i> and <i>ADH5</i> resulted in an increase of 55-fold in the 2,3-Butanediol production yield.	[101]
Succinate	A strategy composed of three gene deletions (<i>SDH3</i> , <i>SER3</i> and <i>SER33</i>) was taken from a previous study that used OptGene to optimize succinate formation using FBA and the model <i>iFF708</i> . The triple deletion strain showed a 13-fold improvement in the succinate yield.	[56, 79]
Succinate	<i>iFF708</i> was used to test all the single and double gene deletions that would improve succinate production using FBA as the simulation method. The best result obtained was for the single deletion of <i>DIC1</i> , which increased the succinate yield to 0.02 C-mol/C-mol. This yield was very similar to the <i>in silico</i> prediction of 0.03 C-mol/C-mol glucose.	[139]
Amorphadiene	<i>iMM904</i> was used in MOMA simulations in order to find single gene knock-outs with improved production of amorphadiene. Up to 10-fold increase in the production yield was observed for some of the single gene knock-outs implemented experimentally.	[140]
3-hydroxypropionic acid	Two different metabolic routes were evaluated for the production of 3-hydroxypropionic acid using the model <i>iTO977</i> . The β -alanine dependent pathway was found to be able to achieve higher theoretical yields and was less dependent on oxygen availability.	[141]

2.4. Conclusions

In the last fifteen years we have witnessed an explosion in the availability of GSMMs and in several types of computational methods to analyze them. FBA paved the way for the prediction of flux phenotypes using stoichiometric models and more complex methodologies with improved characteristics have been developed ever since. The

application of these constraint-based methodologies to metabolic engineering opened a new field of *in silico* strain design, which will certainly grow fueled by successful case-studies (Table 2.4) and with the improvement of the simulation accuracy.

The number of experimental implementations of strain engineering strategies obtained using genome-scale metabolic models is still a small percentage of the total amount, but this can change if more attention is paid to quality instead of the quantity of methods published each year. As shown in section 2.2, there is a huge number of different computational methods available to choose from and to pick the appropriate method for certain metabolic engineering task can prove challenging even to an experienced user. Another relevant issue is that most of the computational methods discussed in this chapter are not available in a user friendly format. Some of these issues are solved by using software packages that compile many of the strain engineering methods into a single location (e.g., COBRA toolbox [115], OptFlux [51] and the BioMet toolbox [116]). With available manuals and installation guides this type of software can deliver complex strain design methods to a target audience that might lack the technical skills to use them unaided.

References

1. Fleischmann RD, Adams MD, White O, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269:496–512.
2. Edwards JS, Palsson BO (1999) Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J Biol Chem* 274:17410–6.
3. Edwards JS, Palsson BO (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A* 97:5528–5533.
4. Förster J, Famili I, Fu P, et al. (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* 13:244–53.
5. Price ND, Reed JL, Palsson BØ (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol* 2:886–97.
6. Monk J, Nogales J, Palsson BO (2014) Optimizing genome-scale network reconstructions. *Nat Biotechnol* 32:447–52.

7. Rocha I, Förster J, Nielsen J (2008) Design and application of genome-scale reconstructed metabolic models. *Methods Mol Biol* 416:409–31.
8. Feist AM, Herrgård MJ, Thiele I, et al. (2009) Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 7:129–43.
9. Hamilton JJ, Reed JL (2014) Software platforms to facilitate reconstructing genome-scale metabolic networks. *Environ Microbiol* 16:49–59.
10. Thiele I, Palsson BØ (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* 5:93–121.
11. Notebaart RA, van Enkevort FHJ, Francke C, et al. (2006) Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics* 7:296.
12. Agren R, Liu L, Shoaie S, et al. (2013) The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput Biol* 9:e1002980.
13. Henry CS, DeJongh M, Best AA, et al. (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* 28:977–82.
14. Arakawa K, Yamada Y, Shinoda K, et al. (2006) GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes. *BMC Bioinformatics* 7:168.
15. Sun J, Zeng A-P (2004) IdentiCS--identification of coding sequence and *in silico* reconstruction of the metabolic network directly from unannotated low-coverage bacterial genome sequence. *BMC Bioinformatics* 5:112.
16. Dias O, Rocha M, Ferreira EC, Rocha I (2010) Merlin: metabolic models reconstruction using genome-scale information. In: Banga JR, Bogaerts P, Impe JFM, Dochain D SIL (ed) Proc. 11th Int. Symp. Comput. Appl. Biotechnol. (CAB 2010). Belgium: Oude Valk College, pp 120–125
17. Smallbone K, Simeonidis E, Swainston N, Mendes P (2010) Towards a genome-scale kinetic model of cellular metabolism. *BMC Syst Biol* 4:6.
18. Bordbar A, Monk JM, King ZA, Palsson BO (2014) Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet* 15:107–20.
19. Gombert AK, Nielsen J (2000) Mathematical modelling of metabolism. *Curr Opin Biotechnol* 11:180–186.
20. Shapiro HM (1969) Input-output models of biological systems: Formulation and applicability. *Comput Biomed Res* 2:430–445.
21. MR W (1984) Metabolic maps for the Apple II. *Biochem Soc Trans* 12:1093–1094.

22. Fell DA, Small JR (1986) Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem J* 238:781–6.
23. Varma A, Palsson BO (1994) Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/Technology* 12:994–998.
24. Llaneras F, Picó J (2008) Stoichiometric modelling of cell metabolism. *J Biosci Bioeng* 105:1–11.
25. Patil KR, Akesson M, Nielsen J (2004) Use of genome-scale microbial models for metabolic engineering. *Curr Opin Biotechnol* 15:64–9.
26. Milne CB, Kim P-J, Eddy JA, Price ND (2009) Accomplishments in genome-scale in silico modeling for industrial and medical biotechnology. *Biotechnol J* 4:1653–70.
27. Oberhardt MA, Palsson BØ, Papin JA (2009) Applications of genome-scale metabolic reconstructions. *Mol Syst Biol* 5:320.
28. Christensen B, Nielsen J (2000) Metabolic network analysis. A powerful tool in metabolic engineering. *Adv Biochem Eng Biotechnol* 66:209–31.
29. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? *Nat Biotechnol* 28:245–8.
30. Feist AM, Palsson BO (2010) The biomass objective function. *Curr Opin Microbiol* 13:344–9.
31. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol* 3:119.
32. Nolan RP, Fenley AP, Lee K (2006) Identification of distributed metabolic objectives in the hypermetabolic liver by flux and energy balance analysis. *Metab Eng* 8:30–45.
33. Burgard AP, Maranas CD (2003) Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol Bioeng* 82:670–7.
34. Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* 10:291–305.
35. Simeonidis E, Price ND (2015) Genome-scale modeling for metabolic engineering. *J Ind Microbiol Biotechnol*. doi: 10.1007/s10295-014-1576-3
36. Garcia-Albornoz MA, Nielsen J (2013) Application of Genome-Scale Metabolic Models in Metabolic Engineering. *Ind Biotechnol* 9:203–214.
37. Long MR, Ong WK, Reed JL (2015) Computational methods in metabolic engineering for strain design. *Curr Opin Biotechnol* 34:135–141.

38. Kim B, Kim WJ, Kim DI, Lee SY (2014) Applications of genome-scale metabolic network model in metabolic engineering. *J. Ind. Microbiol. Biotechnol.*
39. Xu C, Liu L, Zhang Z, et al. (2013) Genome-scale metabolic model in guiding metabolic engineering of microbial improvement. *Appl Microbiol Biotechnol* 97:519–39.
40. Bailey J (1991) Toward a science of metabolic engineering. *Science* (80-) 252:1668–1675.
41. Niefenführ S, Wiechert W, Nöh K (2014) How to measure metabolic fluxes: a taxonomic guide for (13)C fluxomics. *Curr Opin Biotechnol* 34C:82–90.
42. Sauer U (2006) Metabolic networks in motion: 13C-based flux analysis. *Mol Syst Biol* 2:62.
43. Ponce de León M, Cancela H, Acerenza L (2008) A strategy to calculate the patterns of nutrient consumption by microorganisms applying a two-level optimisation principle to reconstructed metabolic networks. *J Biol Phys* 34:73–90.
44. Lewis NE, Hixson KK, Conrad TM, et al. (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* 6:390.
45. Segrè D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A* 99:15112–7.
46. Becker SA, Feist AM, Mo ML, et al. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* 2:727–38.
47. Wintermute EH, Lieberman TD, Silver PA (2013) An objective function exploiting suboptimal solutions in metabolic networks. *BMC Syst Biol* 7:98.
48. Shlomi T, Berkman O, Ruppin E (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A* 102:7695–700.
49. Brochado AR, Andrejev S, Maranas CD, Patil KR (2012) Impact of stoichiometry representation on simulation of genotype-phenotype relationships in metabolic networks. *PLoS Comput Biol* 8:e1002758.
50. Gonçalves E, Pereira R, Rocha I, Rocha M (2012) Optimization approaches for the in silico discovery of optimal targets for gene over/underexpression. *J Comput Biol* 19:102–14.
51. Rocha I, Maia P, Evangelista P, et al. (2010) OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Syst Biol* 4:45.
52. Mo ML, Palsson BO, Herrgård MJ (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol* 3:37.

53. Joyce AR, Palsson BØ (2008) Predicting gene essentiality using genome-scale in silico models. *Methods Mol Biol* 416:433–57.
54. Fong SS, Palsson BØ (2004) Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nat Genet* 36:1056–8.
55. Fong SS, Burgard AP, Herring CD, et al. (2005) *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* 91:643–8.
56. Otero JM, Cimini D, Patil KR, et al. (2013) Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory. *PLoS One* 8:e54144.
57. Chandrasekaran S, Price ND (2010) Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 107:17845–50.
58. Covert MW, Schilling CH, Palsson B (2001) Regulation of gene expression in flux balance models of metabolism. *J Theor Biol* 213:73–88.
59. Covert MW, Xiao N, Chen TJ, Karr JR (2008) Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. *Bioinformatics* 24:2044–50.
60. Shlomi T, Eisenberg Y, Sharan R, Ruppin E (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol* 3:101.
61. Van Berlo RJP, de Ridder D, Daran J-M, et al. (2011) Predicting metabolic fluxes using gene expression differences as constraints. *IEEE/ACM Trans Comput Biol Bioinform* 8:206–16.
62. Fleming RMT, Thiele I (2011) von Bertalanffy 1.0: a COBRA toolbox extension to thermodynamically constrain metabolic models. *Bioinformatics* 27:142–3.
63. Beard DA, Liang S, Qian H (2002) Energy balance for analysis of complex metabolic networks. *Biophys J* 83:79–86.
64. Snitkin ES, Dudley AM, Janse DM, et al. (2008) Model-driven analysis of experimentally determined growth phenotypes for 465 yeast gene deletion mutants under 16 different conditions. *Genome Biol* 9:R140.
65. Park JH, Lee KH, Kim TY, Lee SY (2007) Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation. *Proc Natl Acad Sci U S A* 104:7797–802.
66. Alper H, Jin Y-S, Moxley JF, Stephanopoulos G (2005) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab Eng* 7:155–64.

67. Jung YK, Kim TY, Park SJ, Lee SY (2010) Metabolic engineering of *Escherichia coli* for the production of polylactic acid and its copolymers. *Biotechnol Bioeng* 105:161–71.
68. Scalcinati G, Partow S, Siewers V, et al. (2012) Combined metabolic engineering of precursor and co-factor supply to increase alpha-santalene production by *Saccharomyces cerevisiae*. *Microb Cell Fact* 11:117.
69. Brochado AR, Matos C, Møller BL, et al. (2010) Improved vanillin production in baker's yeast through *in silico* design. *Microb Cell Fact* 9:84.
70. Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* 84:647–57.
71. Tepper N, Shlomi T (2010) Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics* 26:536–43.
72. Xu Z, Zheng P, Sun J, Ma Y (2013) ReacKnock: identifying reaction deletion strategies for microbial strain optimization based on genome-scale metabolic network. *PLoS One* 8:e72150.
73. Feist AM, Zielinski DC, Orth JD, et al. (2010) Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. *Metab Eng* 12:173–86.
74. Kim J, Reed JL (2010) OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst Biol* 4:53.
75. King ZA, Feist AM (2013) Optimizing Cofactor Specificity of Oxidoreductase Enzymes for the Generation of Microbial Production Strains—OptSwap. *Ind Biotechnol* 9:236–246.
76. Tervo CJ, Reed JL (2014) Expanding metabolic engineering algorithms using feasible space and shadow price constraint modules. *Metab Eng Commun* 1:1–11.
77. Ohno S, Shimizu H, Furusawa C (2014) FastPros: screening of reaction knockout strategies for metabolic engineering. *Bioinformatics* 30:981–7.
78. Kim J, Reed JL, Maravelias CT (2011) Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. *PLoS One* 6:e24162.
79. Patil KR, Rocha I, Förster J, Nielsen J (2005) Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics* 6:308.
80. Rocha M, Maia P, Mendes R, et al. (2008) Natural computation meta-heuristics for the *in silico* optimization of microbial strains. *BMC Bioinformatics* 9:499.

81. Fowler ZL, Gikandi WW, Koffas MAG (2009) Increased malonyl coenzyme A biosynthesis by tuning the *Escherichia coli* metabolic network and its application to flavanone production. *Appl Environ Microbiol* 75:5831–9.
82. Costanza J, Carapezza G, Angione C, et al. (2012) Robust design of microbial strains. *Bioinformatics* 28:3097–104.
83. Lun DS, Rockwell G, Guido NJ, et al. (2009) Large-scale identification of genetic design strategies using local search. *Mol Syst Biol* 5:296.
84. Egen D, Lun DS (2012) Truncated branch and bound achieves efficient constraint-based genetic design. *Bioinformatics* 28:1619–23.
85. Pharkya P, Maranas CD (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab Eng* 8:1–13.
86. Yang L, Cluett WR, Mahadevan R (2011) EMILiO: a fast algorithm for genome-scale strain design. *Metab Eng* 13:272–81.
87. Rockwell G, Guido NJ, Church GM (2013) Redirector: designing cell factories by reconstructing the metabolic objective. *PLoS Comput Biol* 9:e1002882.
88. Pharkya P, Burgard AP, Maranas CD (2004) OptStrain: a computational framework for redesign of microbial production systems. *Genome Res* 14:2367–76.
89. Ranganathan S, Suthers PF, Maranas CD (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol* 6:e1000744.
90. Cotten C, Reed JL (2013) Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering. *Biotechnol J* 8:595–604.
91. Chowdhury A, Zomorodi AR, Maranas CD (2014) k-OptForce: integrating kinetics with flux balance analysis for strain design. *PLoS Comput Biol* 10:e1003487.
92. Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* 5:264–276.
93. Liu H, Li Y, Wang X (2015) OP-Synthetic: identification of optimal genetic manipulations for the overproduction of native and non-native metabolites. *Quant Biol* 2:100–109.
94. Choi HS, Lee SY, Kim TY, Woo HM (2010) *In silico* identification of gene amplification targets for improvement of lycopene production. *Appl Environ Microbiol* 76:3097–105.
95. Park JM, Park HM, Kim WJ, et al. (2012) Flux variability scanning based on enforced objective flux for identifying gene amplification targets. *BMC Syst Biol* 6:106.

96. Gawand P, Hyland P, Ekins A, et al. (2013) Novel approach to engineer strains for simultaneous sugar utilization. *Metab Eng* 20:63–72.
97. Tervo CJ, Reed JL (2012) FOCAL: an experimental design tool for systematizing metabolic discoveries and model development. *Genome Biol* 13:R116.
98. Mahadevan R, Edwards JS, Doyle FJ (2002) Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J* 83:1331–40.
99. Zhuang K, Yang L, Cluett WR, Mahadevan R (2013) Dynamic strain scanning optimization: an efficient strain design strategy for balanced yield, titer, and productivity. DySScO strategy for strain design. *BMC Biotechnol* 13:8.
100. Yim H, Haselbeck R, Niu W, et al. (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat Chem Biol* 7:445–52.
101. Ng CY, Jung M-Y, Lee J, Oh M-K (2012) Production of 2,3-butanediol in *Saccharomyces cerevisiae* by *in silico* aided metabolic engineering. *Microb Cell Fact* 11:68.
102. Asadollahi MA, Maury J, Patil KR, et al. (2009) Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through *in silico* driven metabolic engineering. *Metab Eng* 11:328–34.
103. Xu P, Ranganathan S, Fowler ZL, et al. (2011) Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab Eng* 13:578–87.
104. Burgard AP, Vaidyaraman S, Maranas CD Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnol Prog* 17:791–7.
105. Varma A, Boesch BW, Palsson BO (1993) Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl Environ Microbiol* 59:2465–73.
106. Edwards JS, Ramakrishna R, Palsson BO (2002) Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol Bioeng* 77:27–36.
107. Burgard AP, Nikolaev E V, Schilling CH, Maranas CD (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res* 14:301–12.
108. Suthers PF, Dasika MS, Kumar VS, et al. (2009) A genome-scale metabolic reconstruction of *Mycoplasma genitalium*, iPS189. *PLoS Comput Biol* 5:e1000285.
109. Chung BKS, Lee D-Y (2009) Flux-sum analysis: a metabolite-centric approach for understanding the metabolic network. *BMC Syst Biol* 3:117.

110. Burgard AP, Maranas CD (2001) Probing the performance limits of the *Escherichia coli* metabolic network subject to gene additions or deletions. *Biotechnol Bioeng* 74:364–75.
111. Hatzimanikatis V, Li C, Ionita JA, et al. (2005) Exploring the diversity of complex metabolic networks. *Bioinformatics* 21:1603–9.
112. Campodonico MA, Andrews BA, Asenjo JA, et al. (2014) Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path. *Metab Eng* 25:140–158.
113. Rodrigo G, Carrera J, Prather KJ, Jaramillo A (2008) DESHARKY: automatic design of metabolic pathways for optimal cell growth. *Bioinformatics* 24:2554–6.
114. Carbonell P, Planson A-G, Fichera D, Faulon J-L (2011) A retrosynthetic biology approach to metabolic pathway design for therapeutic production. *BMC Syst Biol* 5:122.
115. Schellenberger J, Que R, Fleming RMT, et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* 6:1290–307.
116. Cvijovic M, Olivares-Hernández R, Agren R, et al. (2010) BioMet Toolbox: genome-wide analysis of metabolism. *Nucleic Acids Res* 38:W144–9.
117. Cherry JM, Hong EL, Amundsen C, et al. (2012) *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* 40:D700–5.
118. Woolston BM, Edgar S, Stephanopoulos G (2013) Metabolic engineering: past and future. *Annu Rev Chem Biomol Eng* 4:259–88.
119. Jensen MK, Keasling JD (2014) Recent applications of synthetic biology tools for yeast metabolic engineering. *FEMS Yeast Res*.
120. Goffeau A, Aert R, Agostini-Carbone ML, et al. (1997) The yeast genome directory. *Nature* 387:5.
121. Dai Z, Liu Y, Guo J, et al. (2014) Yeast synthetic biology for high-value metabolites. *FEMS Yeast Res*. doi: 10.1111/1567-1364.12187
122. Nielsen J (2014) Synthetic biology for engineering acetyl coenzyme a metabolism in yeast. *MBio* 5:e02153–14–.
123. Nookaew I, Olivares-Hernández R, Bhumiratana S, Nielsen J (2011) Genome-scale metabolic models of *Saccharomyces cerevisiae*. *Methods Mol Biol* 759:445–63.
124. Osterlund T, Nookaew I, Nielsen J (2012) Fifteen years of large scale metabolic modeling of yeast: developments and impacts. *Biotechnol Adv* 30:979–88.

125. Duarte NC, Herrgård MJ, Palsson BØ (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* 14:1298–309.
126. Kuepfer L, Sauer U, Blank LM (2005) Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res* 15:1421–30.
127. Herrgård MJ, Lee B-S, Portnoy V, Palsson BØ (2006) Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res* 16:627–35.
128. Nookaew I, Jewett MC, Meechai A, et al. (2008) The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol* 2:71.
129. Herrgård MJ, Swainston N, Dobson P, et al. (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol* 26:1155–60.
130. Dobson PD, Smallbone K, Jameson D, et al. (2010) Further developments towards a genome-scale metabolic model of yeast. *BMC Syst Biol* 4:145.
131. Heavner BD, Smallbone K, Barker B, et al. (2012) Yeast 5 - an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network. *BMC Syst Biol* 6:55.
132. Österlund T, Nookaew I, Bordel S, Nielsen J (2013) Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst Biol* 7:36.
133. Heavner BD, Smallbone K, Price ND, Walker LP (2013) Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database (Oxford)* 2013:bat059.
134. Aung HW, Henry SA, Walker LP (2013) Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism. *Ind Biotechnol* 9:215–228.
135. Smallbone K, Simeonidis E, Broomhead DS, Kell DB (2007) Something from nothing: bridging the gap between constraint-based and kinetic modelling. *FEBS J* 274:5576–85.
136. King ZA, Lloyd CJ, Feist AM, Palsson BO (2015) Next-generation genome-scale models for metabolic engineering. *Curr Opin Biotechnol* 35:23–29.
137. Bro C, Regenberg B, Förster J, Nielsen J (2006) *In silico* aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metab Eng* 8:102–111.

138. Kennedy CJ, Boyle PM, Waks Z, Silver PA (2009) Systems-level engineering of nonfermentative metabolism in yeast. *Genetics* 183:385–97.
139. Agren R, Otero JM, Nielsen J (2013) Genome-scale modeling enables metabolic engineering of *Saccharomyces cerevisiae* for succinic acid production. *J Ind Microbiol Biotechnol* 40:735–47.
140. Sun Z, Meng H, Li J, et al. (2014) Identification of novel knockout targets for improving terpenoids biosynthesis in *Saccharomyces cerevisiae*. *PLoS One* 9:e112615.
141. Borodina I, Kildegaard KR, Jensen NB, et al. (2014) Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in *Saccharomyces cerevisiae* via β -alanine. *Metab Eng* 27:57–64.

CHAPTER 3

Improving the flux distributions simulated with genome-scale metabolic models of *S. cerevisiae*

A genome-scale metabolic model is a comprehensive description of cellular metabolism and can be used to evaluate genotype-phenotype relationships. There is therefore much interest in using these models for evaluating gene knock-outs that optimize the production of specific metabolites that may have industrial interest. However, to obtain a good correlation between simulations and *in vivo* results, a thorough validation of the flux predictions should be made.

In order to investigate the accuracy of simulated intracellular fluxes, *Saccharomyces cerevisiae* was chosen as the case-study because of its industrial relevance. The results revealed that steady-state simulations of the models available contained mistakes in important areas of central metabolism, such as the pentose phosphate pathway. Using cofactor abundance in aerobic conditions as guidance, constraints were applied to all the metabolic reactions including NADH of NADPH, resulting in several improvements in the pentose phosphate pathway and other areas related to NADPH metabolism. Those improvements were shown to have a positive impact on the simulation of gene knock-outs obtained for the production of acetate and mevalonate.

When genome-scale metabolic models are used for the simulation of gene knock-out phenotypes, erroneous flux predictions might compromise the accuracy of the results. Therefore, a careful curation of the wild-type network can improve dramatically the simulations, resulting in a better correlation with experimental evidence.

The information presented in this Chapter is being prepared for submission to a peer reviewed journal:

Pereira R., Nielsen J., Rocha I., Improving the flux distributions simulated with genome-scale metabolic models of *S. cerevisiae*.

3.1 Introduction

In metabolic engineering, Genome-Scale Metabolic Models (GSMMs) can be used to predict the effect of different environmental conditions and genetic modifications on the metabolism of microorganisms of industrial interest [1–6]. If used together with optimization algorithms, it is possible to find an optimal set of genetic modifications leading to a desired phenotype [7–9]. Frequently the validation of these models is done by comparing the predictions obtained solely with physiological data (e.g., growth rates and gene essentiality) and rarely with *in vivo* flux distributions. This is mainly because of the lack of data available on fluxomics and due to the existence of several alternative simulation methods that may provide different solutions in terms of the fluxes.

Another pertinent problem appears when the algorithms used to simulate phenotypes require a reference (wild-type) flux distribution, such as MOMA [10], ROOM [11] or MiMBL [12]. The easiest way of calculating the reference flux distribution for a GSMM is to use the fluxes obtained from Flux Balanced Analysis (FBA) [13, 14] or Parsimonious Flux Balanced Analysis (pFBA) [15] simulations [16–18]. However, little is known about the accuracy of the internal flux distribution calculated with these methods. Segrè et al. [10] have reported that even small deviations in the wild-type flux distribution used for MOMA simulations can have an impact on the simulation outcome. Furthermore, Brochado et al. [12] also noted that the reference flux distribution had a significant impact when MiMBL was used to simulate single gene deletion mutants.

Several cases of *in vivo* implementation of MOMA results are reported in the literature with varying degrees of success [16–19, 1]. One of the possible reasons for the inconsistent predictions might be the use of a reference flux distribution with severe deviations from *in vivo* fluxes. This may be solved by using experimental fluxes to constrain the reference network, as described by Kuepfer et al. [20], but only when data is available for that particular condition. In cases where experimental data are absent, the only alternative would be to curate the wild-type flux distribution using the available knowledge. In fact, it has been shown before that significant improvements in the internal fluxes predicted by FBA could be achieved for a GSMM of *Arabidopsis thaliana*. Briefly, the problems encountered in the oxidative part of the pentose phosphate pathway were solved by the authors by accounting for maintenance NADPH and ATP costs [21, 22].

Saccharomyces cerevisiae is one of the most studied microorganisms and a considerable amount of information about its metabolism is available. It has proven to be quite versatile in terms of industrial applications, making it very attractive for metabolic engineering. Several GSMMs have been published for *S. cerevisiae* over the last ten years [20, 23–28] and their prediction of growth rate and exchange fluxes using FBA is generally quite good. These facts make *S. cerevisiae* very attractive for investigating in detail the capacity of GSMMs and simulation methods to predict the intracellular flux distribution of the central carbon metabolism.

In this work we aim at analyzing the reliability of the flux distributions predictions obtained with *S. cerevisiae* GSMMs. We begin by comparing the performance of several models available in the literature for *S. cerevisiae* and propose modifications to improve their internal flux prediction under fully aerobic conditions. Furthermore, the improved models are compared to *in vivo* flux estimations for validations, and used to find metabolic engineering strategies unattainable with the original versions.

3.2 Methods

3.2.1 Model retrieval and pre-processing

The models *iFF708* [23] and *iTO977* [25] were obtained in the SBML format from the authors' website (<http://129.16.106.142/models.php?c=S.cerevisiae/>). The model *iMM904* [24] was downloaded and converted to SBML using the script provided by the authors in the supplementary material. Yeast 6.06 [29] was downloaded in SBML from the project's website: <http://sourceforge.net/projects/yeast/files/>.

Each model was imported into OptFlux 3.07 [30] and the *in silico* environmental conditions were set to mimic minimal growth medium supplemented with glucose under fully aerobic conditions (ammonia: unconstrained uptake, phosphate: unconstrained uptake, sulfate: unconstrained uptake, oxygen: unconstrained uptake, glucose: 1.15 mmol/gCDW·h).

3.2.2 Simulations

All simulations were run within OptFlux 3.07 using IBM ILOG CPLEX optimization studio (academic) as the linear and quadratic programming solver. For the calculation of the internal flux distributions Parsimonius Flux Balanced Analysis (pFBA) [15] was the chosen method, while the maximization of biomass production was the objective function.

In genome-scale metabolic models it is common to find duplicated reactions and other stoichiometrically equivalent pathways that catalyze the same chemical transformation. The phosphorylation of glucose, for example, can be carried by hexokinases or glucokinases. To simplify the visualization of the fluxes through the central metabolism of *S. cerevisiae*, the fluxes through equivalent reactions were summed into a single value. Using the example above, all the reactions phosphorylating glucose to glucose-6-phosphate present in each model were treated as a single entity. Each model was inspected manually to create an equivalence map between the reactions in the central metabolism and the reactions in the model.

3.2.3 Flux variability analysis

To verify the uniqueness of the flux distributions calculated by FBA, each reaction's variability was tested by following the methodology reported in [31, 32]. Briefly, the value of the objective function (biomass growth) was fixed at its maximum and each flux was minimized or maximized. However, since the models tested contained many reactions in the central metabolism that are duplicated, if one of them is minimized the flux could go through the other. Furthermore, if the duplicated reactions are reversible, futile cycles can arise. To avoid these irrelevant glitches in this analysis, when one reaction's variability was tested, all the other "equivalent partners" were disabled.

3.2.4 Optimizations of metabolic engineering targets

In order to find possible knock-out targets, the evolutionary optimization algorithm included in OptFlux was used [7, 33]. A linear version [34] of the MOMA algorithm [10], where the Euclidian distance of the fluxes is replaced by the Manhattan distance, was the chosen method to simulate the mutant phenotypes. The number of possible knock-out targets was reduced by removing: essential reactions, dead ends and other reactions that

cannot be active under growth on glucose. Two different objective functions were used in the evolutionary algorithm: the Biomass Product Coupled Yield (BPCY) [7] or the maximization of the target flux assuming a minimum biomass growth threshold (yield). The evolutionary algorithm was run at least three times for each target, setting the number of solution evaluations to 50,000 and the maximum number of knock-outs allowed to 5.

3.3 Results and discussion

3.3.1 Genome-scale metabolic model comparison

Since the first GSMM of *S. cerevisiae* was published in 2003 [23], several authors have expanded and improved the metabolic reconstruction of this microorganism (as reviewed in [35, 36]). A summary of the GSMMs available in the literature is given in Table 3.1, focusing on model size and other relevant parameters. The numbers show that over the years more ORFs and metabolic reactions have been included in the models. Furthermore, compartments were added and elemental balance of the reactions was implemented in some models.

Table 3.1- List of selected GSMM available for *S.cerevisiae*

Model	Year	ORFs included	Reactions included	Other information
<i>i</i> FF708 [23]	2003	708	1175	3 compartments (2 metabolic)
<i>i</i> ND750 [28]	2004	750	1489	8 compartments (7 metabolic) Elementally balanced
<i>i</i> LL672 [20]	2005	672	1038	3 compartments (2 metabolic)
<i>i</i> IN800 [27]	2008	800	1446	3 compartments (2 metabolic)
<i>i</i> MM904 [24]	2009	904	1412	8 compartments (7 metabolic) Elementally balanced
<i>i</i> TO977 [25]	2013	977	1566	4 compartments (3 metabolic)
Yeast 6 [29]	2013	900	1888	15 compartments (7 metabolic) Elementally balanced

To validate GSMMs, their authors usually use large scale essentiality data (lethal/viable) and extracellular flux predictions. However, not too much attention is paid to the intracellular flux distribution simulated with FBA [13, 37, 38].

Selected models from Table 3.1 were simulated under the same environmental conditions, and their intracellular flux distribution around the central metabolism was compared and

evaluated. We chose *iFF708* for this analysis because it is still widely used for phenotype simulation [1, 4–6]. Furthermore, the three most recent models from Table 3.1 were also included to verify how the models evolved on the flux prediction side.

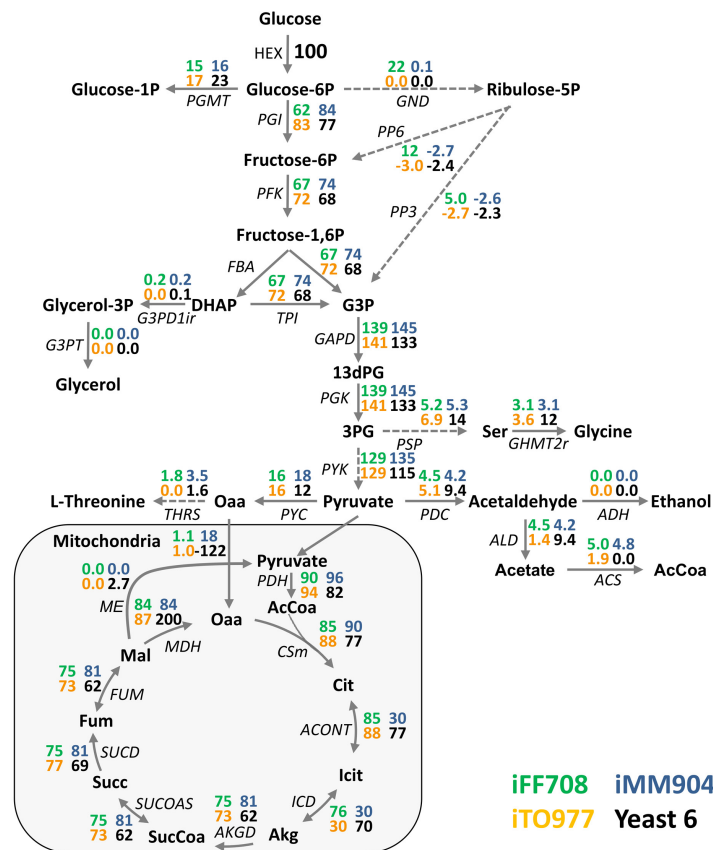


Figure 3.1- Distribution of fluxes in the central metabolism predicted by pFBA for the models: *iFF708* (green), *iMM904* (blue), *iTO977* (orange), Yeast 6 (black). Reactions: ACONT- aconitase, ACS- acetyl-CoA synthetase, ADH- alcohol dehydrogenase, AKGD- alpha-ketoglutarate dehydrogenase, ALD- aldehyde dehydrogenase, CSM- citrate synthase, FBA- fructose 1,6-bisphosphate aldolase, FUM- fumarase, G3PD1ir- glycerol-3-phosphate dehydrogenase, G3PT- glycerol-1-phosphatase, GAPD- glyceraldehyde-3-phosphate dehydrogenase, GHMT2r- serine hydroxymethyltransferase, GND- 6-phosphogluconate dehydrogenase, HEX- hexokinase, ICD- mitochondrial isocitrate dehydrogenase, MDH- mitochondrial malate dehydrogenase, ME- mitochondrial malic enzyme, PDC- pyruvate decarboxylase, PDH- pyruvate dehydrogenase, PFK- phosphofruktokinase, PGI- phosphoglucose isomerase, PGK- 3-phosphoglycerate kinase, PGMT- phosphoglucomutase, PP3- sum of the non-oxidative reactions of the pentose phosphate pathway producing glyceraldehyde-3-phosphate, PP6- sum of the non-oxidative reactions of the pentose phosphate pathway producing fructose-6-phosphate, PSP- phosphoserine phosphatase, PYC- pyruvate carboxylase, PYK- pyruvate kinase, SUCD- succinate dehydrogenase, SUCOAS- succinyl-CoA ligase, THRS- threonine synthase, TPI- triose phosphate isomerase. Metabolites: 13dPG- 1,3-diphosphoglycerate, 3PG- 3-phosphoglycerate, AcCoA- acetyl-CoA, Akg- 2-oxoglutarate, Cit- citrate, DHAP- dihydroxyacetone-phosphate, Fum- fumarate, G3P- glyceraldehyde-3-phosphate, Icit- isocitrate, Mal - L-malate, Oaa- oxaloacetate, Ser- L-serine, Succ- succinate, SucCoa- succinyl-CoA.

The first step in the analysis was a simple comparison of the maximum growth rate achievable under the same environmental conditions. With a fixed glucose uptake rate of $1.15 \text{ mmol}/(\text{gCDW}\cdot\text{h})$ ($0.207 \text{ g}/\text{gCDW}\cdot\text{h}$) and assuming a biomass yield on Glucose of $0.5 \text{ gCDW}/\text{g glucose}$ [39], the maximum growth rate expected should be 0.10 h^{-1} . Three of the models predicted a maximum growth rate close to 0.11 h^{-1} , but the simulations performed with Yeast 6 could achieve 0.16 h^{-1} . Taking a closer look at this issue revealed an unreasonably high flux through the oxidative phosphorylation reactions that was culminating in the production of large amounts of ATP. The reason behind this odd behavior was a cycle of proton export from the mitochondria, originated by the aspartate:proton symporter (reaction `r_1117` in the model). Using several transporters and metabolic reactions, Yeast 6 was capable of creating a proton gradient across the mitochondrial membrane without energy input. The solution to this problem was to disable the transporter activity in the direction from the mitochondria to the cytosol. After this simple modification, the growth dropped to reasonable levels. The authors of Yeast 6 actually reported that this model was capable of using ATP Synthase in anaerobic conditions [29], which can be explained by the same cycle reported here.

In Figure 3.1 the fluxes through the central metabolic pathways are shown for the corrected Yeast 6 and the other selected models. One of the most noticeable differences that can be observed is the absence of flux in the oxidative Pentose Phosphate Pathway (PPP) in the most recent models (*iMM904*, *iTO977* and Yeast 6). No flux through the oxidative PPP suggests problems in the reactions producing and/or consuming NADPH present in the most recent models. An analysis of the fluxes revealed that in *iTO977* and *iMM904* all the NADPH required for anabolism was coming from the cytosolic NADP-specific isocitrate dehydrogenase, which explained the lack of flux in the PPP in these models. In the case of Yeast 6, more inconsistencies were found around the NADPH metabolism. In this model, the main reaction fixing ammonia was the NADH-dependent glutamate dehydrogenase instead of the NADPH dependent variant. Furthermore, the NADPH required for most anabolic processes was being produced by the cytosolic aldehyde dehydrogenase and the cytosolic C1-tetrahydrofolate synthase. It was thus evident that any improvement in the internal flux distributions simulated with the GSMMs of *S. cerevisiae* would involve changes in the reactions producing and consuming NADPH.

3.3.2 Emulation of cofactor concentrations through reaction constraints

Given all the indications pointing towards problems with the metabolism of NADPH we decided to analyze in detail all the metabolic reactions that include this cofactor. Since NADH can replace NADPH as the electron donor in many reactions of *S. cerevisiae*, it was decided to also include it in the analysis.

One of the issues encountered in the model comparison was that the cytosolic isocitrate dehydrogenase was acting as the main source of NADPH in the cytosol for two of the models, *iMM904* and *iTO977*. Although the gene coding for this enzyme is described as repressed by glucose [40], under glucose limiting conditions it is not likely that it will be fully repressed. However, it was shown by Satrustegui et al. [41] that the ratio of NADPH/NADP⁺ can act as an inhibitor of the NADP⁺-dependent isocitrate dehydrogenase. Furthermore, under growth on glucose this ratio is high enough to inhibit this enzyme. Additionally, the concentration of these cofactors can also affect the change in Gibbs free energy of this reaction, which should also favor the direction of NADPH consumption.

Since under fully aerobic glucose limited conditions the ratio of NADPH/NADP⁺ is usually very high, in order to drive anabolic reactions, and the ratio of NADH/NAD⁺ is usually very low, in order to promote fast metabolic oxidation of the substrate [41–44], an attempt was made to generalize the findings of Satrustegui et al. [41] for other reactions in the network. With these ratios in mind, all cytosolic reactions that could produce NADPH or consume NADH were checked manually to evaluate their feasibility. The curation process is documented in Appendix A (tables A1-A4) and the final list of modified reactions is shown in Table 3.2.

During the manual curation there were some modifications applied to reinforce cofactor availability constraints that resulted in a lethal phenotype. In each case the reason for the lack of growth was analyzed using information available in the literature and the other models as comparison.

One of the lethal candidate reactions for inactivation in the model *iFF708* was the squalene epoxidase (model ID: ERG1), which was capable of producing NADPH in this model. However, this reaction is described to consume NADPH [45], supported by the fact that the other models all had the correct stoichiometry. Therefore, this reaction was

corrected in the model *iFF708* and removed from the list of reactions to be inactivated. Another problem was found when the cytosolic isocitrate dehydrogenase (model ID: *IDP2_1*) was inactivated. Again, the inactivation of this enzyme resulted in a lethal phenotype, which was traced back to the lack of mitochondrial transporters for 2-oxoglutarate. Using literature information and Yeast 6 model as an example, two transporters for this metabolite were added to the model: citrate/2-oxoglutarate antiporter (*YHM2*) and malate/2-oxoglutarate antiporter (*ODC1*, *ODC2*) [46, 47].

Table 3.2- List of modified reactions in each model (reactions are identified by their IDs on the model).

Model	Reversibility constrained		Inactivated	
	NADH	NADPH	NADH	NADPH
<i>iFF708</i>	MDH3, MDH2, ADH2, ADH1, ADH5, ADH4, BIO2, SFA1_2, SFA1_1, LYS1, TDH1, TDH2, TDH3	HMG2, HMG1, ECM17, MET10, LYS9	U45_, PRO3_3, FAS1_4, FOX2, PRO2_1, LYS2_2, HOM6_1	TYR1, IDP3_1, ARA1_2, ALD6, IDP2_1
<i>iMM904</i>	BTDD-RR, FALDH, 2HBO, ALCD2x, SACCD2, MDH, GAPD	HMGCOAR, SACCD1, SULR	AASAD2, G5SD2, HPROa, ALCD25xi, ALCD2ir, ALCD26xi, ALCD22xi, FMNRx, ALCD23xi, C22STDSx, LNS14DMx, HSDxi, ALCD24xi	LSERDHr, ATHRDHr, PPND2, GLYCDy, ARAB1D2, ICDHy, ALDD20y, ALDD2y
<i>iTO977</i>	BIO2, LYS1, r275, ADH1, SFA1_1, MDH2	HMG1, FAS2_1_2, FAS2_2_2, FAS2_3_2, FAS2_4_2, FAS2_5_2, FAS2_6_2, FAS2_7_2, FAS2_8_2, FAS1_1_2, FAS1_2_2, FAS1_3_2, FAS1_4_2, FAS1_5_2, FAS1_6_2, FAS1_7_2, FAS1_8_2, LYS9, ECM17	LYS2_2, PRO2_1, U45_, PRO3_3, HOM6_1, PGA3	ARA1_2, r581, r572, r671, IDP2_1, r253, r833, ALD6, TYR1
Yeast 6.06	r_0714, r_0486, r_0470, r_0003	-	r_0169, r_0166, r_0179, r_0182, r_0186, r_0441, r_1010, r_2115	r_0173, r_0177, r_0234, r_0659, r_0676, r_0690, r_0321, r_0939

In the model *iMM904* a similar issue was encountered when the cytosolic isocitrate dehydrogenase (model ID: *R_ICDHy*) was inactivated. As before, the addition of the 2-oxoglutarate transporters solved this issue and restored normal growth. However, following the addition of these transporters there was an abnormal increase in biomass formation caused by an artificial cycle of proton export from the mitochondria, similar to the one observed for the original Yeast 6 and discussed in section 3.3.1. It was solved by changing the reversibility of three mitochondrial transporters: aspartate:proton symporter (model ID: *R_ASPT2m*) was allowed only to transport into the mitochondria; oxaloacetate:proton symporter (model ID: *R_OAA2m*) was allowed only to transport

into the mitochondria; and mitochondrial dicarboxylate carrier (model ID: R_MALtm) was allowed only to transport L-malate into the mitochondria. These changes restored normal growth levels and no further issues were detected in the network.

Regarding Yeast 6, the list of reactions to inactivate also included two candidates that were essential for growth: sterol dehydrogenase (model ID: r_0234) and prephenate dehydrogenase (model ID: r_0939). Regarding the first reaction, *iMM904* included an alternative reaction that used NAD^+ instead of NADP^+ . According to Baudry et al. [48] this enzyme should use NAD^+ as a cofactor, so a NAD^+ dependent reaction was added to Yeast 6. For prephenate dehydrogenase, all the other models included a NAD^+ dependent variant for this enzyme, and the only reference found for this enzyme in *S. cerevisiae* actually refers to a possible NAD^+ dependent enzyme [49]. Thus, a NAD^+ dependent variant for this reaction was also added to Yeast 6.

The corrected models containing the restrictions from Table 3.2 were used for simulations and the resulting flux distributions are shown in Figure 3.2. In addition, we also included experimentally determined flux values obtained from labeled glucose chemostats [50, 51]. An overall analysis of Figure 3.2 revealed a dramatic improvement in the flux through the oxidative PPP in the models *iMM904*, *iTO977* and Yeast 6.

Focusing on *iFF708* (Figure 3.2), there was also an increase in the flux through the oxidative PPP to levels closer to the experimental fluxes. Since many anabolic reactions from the original model were capable of consuming either NADH or NADPH, when the NADH consuming equivalent was inactivated (Table 3.2), the requirements for NADPH increased. As a consequence, the flux through the PPP had to compensate for the increased requirements for this cofactor. Another visible difference is the increased transport of oxaloacetate into the mitochondria, which can be explained by the inactivation of the cytosolic isocitrate dehydrogenase (model ID: IDP2_1) and the addition of the mitochondrial 2-oxoglutarate transporters. These modifications shifted 2-oxoglutarate production from the cytosol to the mitochondria, resulting in extra drainage of citric acid cycle intermediaries. Therefore, the anaplerotic flux of oxaloacetate into the mitochondria had to increase.

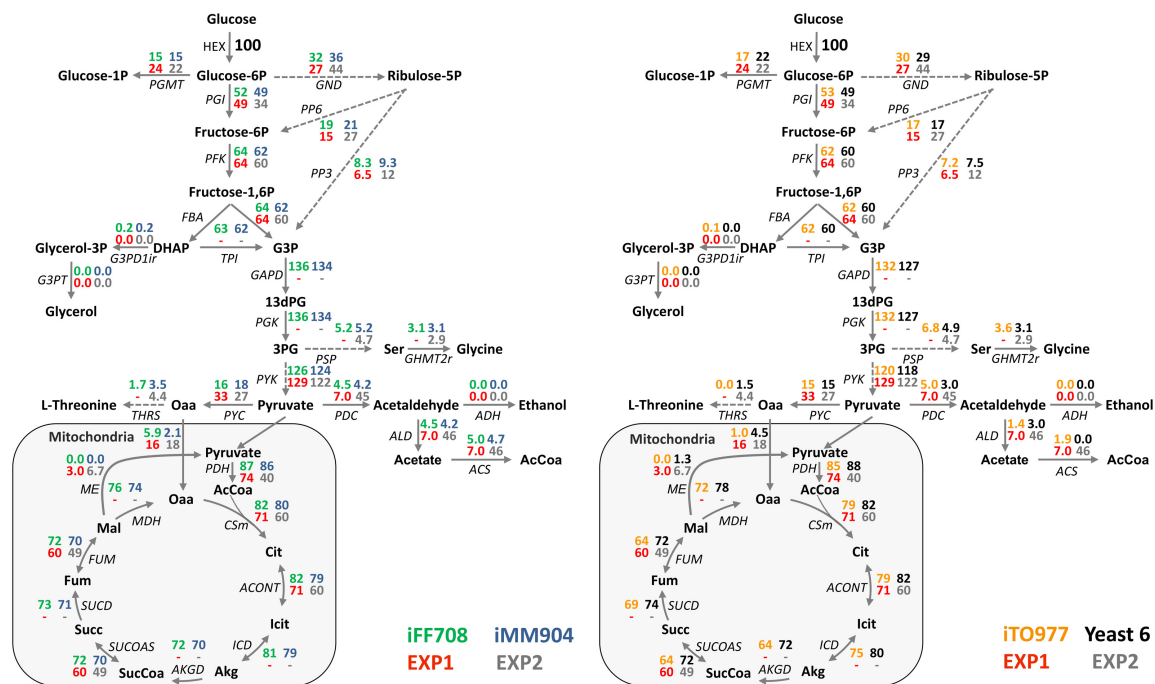


Figure 3.2- Comparison between experimentally determined fluxes and the predictions of pFBA for the modified models. *iFF708* (green), *iMM904* (blue), *iTO977* (yellow), Yeast 6 (black), ^{13}C -MFA fully aerobic chemostat at a dilution rate of 0.1 h^{-1} (red)[50], ^{13}C -MFA fully aerobic chemostat at a dilution rate of 0.1 h^{-1} (grey)[51]. Reactions: ACONT- aconitase, ACS- acetyl-CoA synthetase, ADH- alcohol dehydrogenase, AKGD- alpha-ketoglutarate dehydrogenase, ALD- aldehyde dehydrogenase, CSm- citrate synthase, FBA- fructose 1,6-bisphosphate aldolase, FUM- fumarase, G3PD1ir- glycerol-3-phosphate dehydrogenase, G3PT- glycerol-1-phosphatase, GAPD- glyceraldehyde-3-phosphate dehydrogenase, GHMT2r- serine hydroxymethyltransferase, GND- 6-phosphogluconate dehydrogenase, HEX- hexokinase, ME- mitochondrial malic enzyme, PDC- pyruvate decarboxylase, PDH- pyruvate dehydrogenase, PFK- phosphofruktokinase, PGI- phosphoglucose isomerase, PGK- 3-phosphoglycerate kinase, PGMT- phosphoglucomutase, PP3- sum of the non-oxidative reactions of the pentose phosphate pathway producing glyceraldehyde-3-phosphate, PP6- sum of the non-oxidative reactions of the pentose phosphate pathway producing fructose-6-phosphate, PSP- phosphoserine phosphatase, PYC- pyruvate carboxylase, PYK- pyruvate kinase, SUCD- succinate dehydrogenase, SUCOAS- succinyl-CoA ligase, THRS- threonine synthase, TPI- triose phosphate isomerase. Metabolites: 13dPG- 1,3-diphosphoglycerate, 3PG- 3-phosphoglycerate, AcCoA- acetyl-CoA, Akg- 2-oxoglutarate, Cit- citrate, DHAP- dihydroxyacetone-phosphate, Fum- fumarate, G3P- glyceraldehyde-3-phosphate, Icit- isocitrate, Mal - L-malate, Oaa- oxaloacetate, Ser- L-serine, Succ- succinate, SucCoa- succinyl-CoA.

Regarding *iMM904*, there was a significant improvement in all of the PPP fluxes, allied to enhancements in the citric acid cycle. The reasons for such an improvement in the flux distribution of the PPP can be attributed to two separated factors: an increase in the anabolic requirement for NADPH and a decrease in the production of NADPH in alternative pathways. The modifications shown in Table 3.2 addressed both of these issues and resulted in a flux distribution that is quite similar to experimentally estimated fluxes. Furthermore, the fluxes in the non-oxidative PPP changed direction when compared to Figure 3.1. Because in the original model there was no flux in the oxidative

PPP, the intermediates of this pathway needed for biomass production were being produced from fructose-6-phosphate and glyceraldehyde-3-phosphate formed in glycolysis, which resulted in inverted fluxes in the non-oxidative PPP. Concerning the improvements in the citric acid cycle, the original *iMM904* predicted that a considerable part of the flux through aconitase and isocitrate dehydrogenase was taking place in the cytosol (fluxes not shown). This flux detour was abolished when the cytosolic isocitrate dehydrogenase (model ID: R_ICDH_y) was inactivated and the 2-oxoglutarate transporters were added, resulting in a standard operation of the citric acid cycle inside the mitochondria.

Similarly to *iMM904*, *iTO977* also showed improvements in PPP and citric acid cycle. In this case, the reason for the zero flux through the PPP was solely the activity of the cytosolic isocitrate dehydrogenase (model ID: IDP2_1), which was supplying all NADPH required for growth. Furthermore, this issue was also responsible for the low flux through mitochondrial isocitrate dehydrogenase. Consequently, the cofactor constraints applied to the model solved both of these problems.

Finally, Yeast 6 also benefited from the restrictions applied. Although the most prominent change was the flux in the oxidative PPP, there were also improvements in the citric acid cycle, L-serine biosynthesis and oxaloacetate transport. Furthermore, the nitrogen metabolism (not shown in Figure 3.2) also exhibited large improvements. The ammonium assimilation pathway switched from NADH dependent to NADPH dependent, which is consistent with the importance of NADPH-dependent glutamate dehydrogenase for growth on glucose [52, 53]. Despite not being related with the modifications applied, Yeast 6 was the only model that could predict flux through the malic enzyme.

After the manual curation process, some of the fluxes in the central carbon metabolism were still incorrectly simulated by pFBA. One of those cases was the absence of flux through the canonical pathway for L-threonine production in the model *iTO977* (THRS in Figure 3.2). The reason for the absence of flux in this reaction is the presence of a threonine aldolase (model ID: GLY1) in *iTO977* with incorrect reversibility constraints, causing L-threonine to be synthesized from glycine and acetaldehyde. In the case of Yeast 6, no flux is predicted in acetyl-CoA synthetase (ACS in Figure 3.2), which seems strange given the essentiality of acetyl-CoA synthesis in the cytosol. Again, an erroneous reversibility was responsible for this glitch. In this model, a cytosolic acetyl-CoA

hydrolase (model ID: r_0110) was able to catalyze the reverse reaction, joining acetate and coenzyme A with no energy input.

The qualitative improvements observed in Figure 3.2 could also be seen when the correlation between the experimental and simulated fluxes was analyzed (Figure 3.3). In the model *iFF708* all the parameters analyzed improved slightly with the exception of Pearson coefficient. The modest improvements observed are consistent with the fact that *iFF708* was the model with the best initial flux distribution. However, for the other three models the improvements are much more pronounced, and in all cases the correlation coefficients show a much better prediction of the central carbon fluxes.

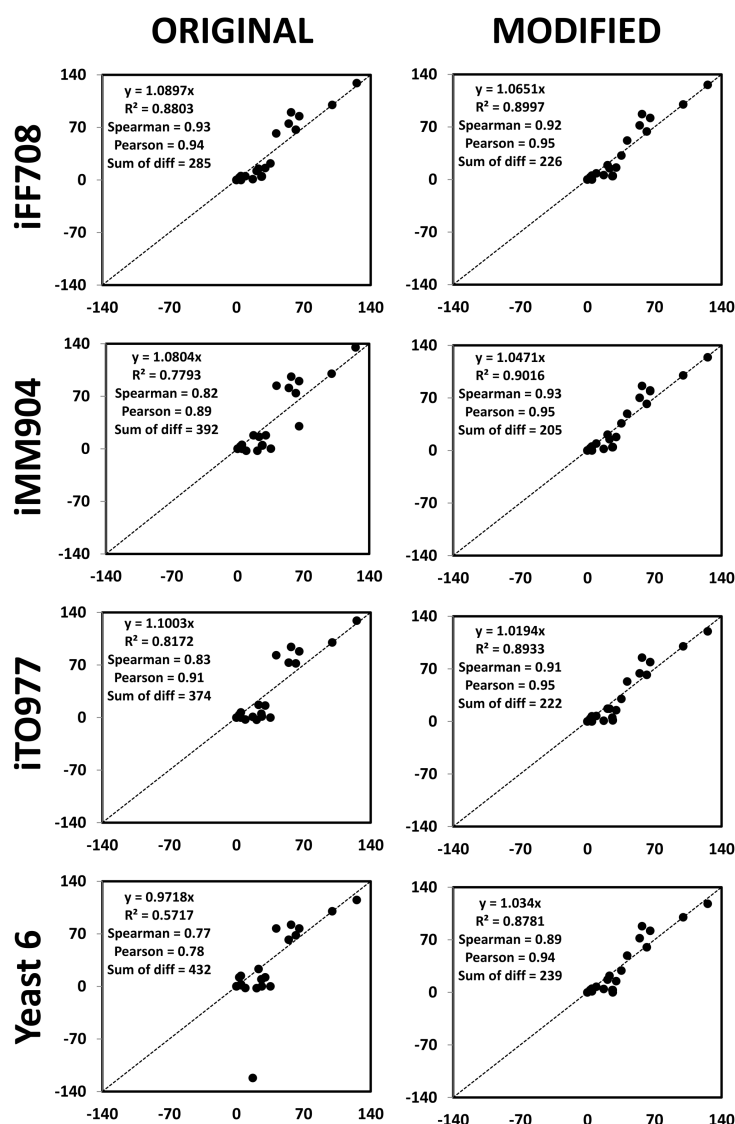


Figure 3.3 - Comparison between the experimentally determined fluxes (x-axis) and pFBA simulation of each model (y-axis). For each graph it is shown: the linear regression equation, the correlation coefficient (R^2), the Spearman correlation coefficient, the Pearson correlation coefficient and the sum of differences. The experimental fluxes used in the correlations were obtained by averaging both sets of data [46, 47].

3.3.3 Flux variability analysis

When FBA is used to simulate flux distributions in large-scale models, part of the network can be underdetermined. This means that a subset of reactions can vary without having an impact on the objective function (growth rate in the case). To avoid random variability between simulations and reduce the formation of futile cycles in the network, pFBA [15] is normally used after an FBA simulation. In summary, first the value of the objective function is fixed at its optimum and then the sum of all fluxes is minimized. Biologically, this minimization assumes that the cell will try to conserve resources by controlling the total amount of enzymes available.

Although pFBA solves some of the problems regarding the uniqueness of the fluxes, it is still important to verify if a particular flux can be variable or not, i.e., if there are alternative flux distributions for the same FBA optimum. In order to test if the fluxes calculated are unique, each of the reactions represented in Figure 3.1 and Figure 3.2 was tested. The reactions that showed more than 10% variation in comparison to the value obtained with pFBA are listed in Table 3.3.

Table 3.3- List of variable fluxes obtained using flux variability analysis (reactions are identified by the same IDs from Figure 3.1 and Figure 3.2).

Variability	<i>iFF708</i>		<i>iMM904</i>		<i>iTO977</i>		Yeast 6	
	Original	Modified	Original	Modified	Original	Modified	Original	Modified
Variable minimum	G3PD1ir, GHMT2r, OAAat, FUM	G3PD1ir, GHMT2r, FUM	G3PD1ir, OAAat, ACONT, ICD, MDH	G3PD1ir, PDC, OAAat, ACONT	PSP, GHMT2r, ALD, OAAat, ICD, AKGD, FUM, MDH	G3PD1ir, GHMT2r, ALD, OAAat, AKGD, FUM, MDH	PYC, OAAat, ICD, AKGD, SUCOAS, FUM	G3PD1ir, GHMT2r
Variable maximum	G3PD1ir, GHMT2r, OAAat, SUCD, FUM	G3PD1ir, GHMT2r, OAAat, SUCD, FUM	G3PD1ir, OAAat, ACONT, SUCD, FUM, MDH	G3PD1ir, ACONT, SUCD	G3PD1ir, GHMT2r, OAAat, AKGD, SUCD, FUM, MDH	G3PD1ir, GHMT2r, OAAat, AKGD, SUCD, FUM	PDC, ALD, MDH	G3PD1ir, SUCD

Abbreviations: ACONT- aconitase, AKGD- alpha-ketoglutarate dehydrogenase, ALD- aldehyde dehydrogenase, FUM- fumarate, G3PD1ir- glycerol-3-phosphate dehydrogenase, GHMT2r- serine hydroxymethyltransferase, ICD- mitochondrial isocitrate dehydrogenase, MDH- mitochondrial malate dehydrogenase, OAAat- mitochondrial oxaloacetate transport, PDC- pyruvate decarboxylase, PSP- phosphoserine phosphatase, PYC- pyruvate carboxylase, SUCD- succinate dehydrogenase, SUCOAS- succinyl-CoA ligase.

The list of variable fluxes from Table 3.3 revealed that one of the reactions that showed the most variability was glycerol-3-phosphate dehydrogenase (G3PD1ir). The cause for this variability was found to be related to NADH oxidation in the mitochondria. When this reaction is maximized it can be used to transport NADH into the mitochondria using the glycerol phosphate shuttle. However, the easiest and shortest way to oxidize the NADH should be the external NADH dehydrogenase. Nevertheless, both alternatives are stoichiometrically equivalent and either of them can be used without affecting the maximum *in silico* growth rate.

Many of the reactions from the citric acid cycle are also present in Table 3.3. The reason for such variation is either the presence of equivalent reactions in the cytosol or the existence of a reverse reaction. In the case of malate dehydrogenase (MDH), similar enzymes exist in the mitochondria and cytosol, which together can form a futile cycle. Inside the mitochondria, fumarate reductase and succinate dehydrogenase can also form a futile cycle, resulting in variability for these fluxes. The variability for the other reactions in Table 3.3 was also investigated, but all and all of them could be explained by one of the examples discussed above. Therefore, we can assume that the changes between the original and modified models (Figure 3.1 and Figure 3.2) were the result of the restrictions applied.

3.3.4 Mutant phenotype prediction

Many methods used for phenotype simulation of gene knock-outs rely on the minimization of the adjustment required after a perturbation. Using a distribution of fluxes for the wild-type organism as a reference, knock-out simulation methods try to minimize the magnitude of changes in flux distributions (MOMA) [10], the number of activated/deactivated reactions (ROOM) [11] or changes in metabolite turnovers (MiMBL) [12]. The reference fluxes used for the calculations are usually obtained from an FBA or pFBA simulation of the wild-type model, with growth maximization used as the objective function [16–18]. However, as shown in Figure 3.1, the accuracy of this reference is dependent on the model used, and erroneous values can lead to large differences on the simulation outcome.

Using the model Yeast 6 as an example we tried to verify the impact of the improved flux distribution shown in Figure 3.2 on knock-out simulations using linear MOMA as the simulation method. Two organic acids (mevalonate and acetate) were used as case studies in order to compare the effect of using the original or the curated flux distribution on the production levels of these metabolites. Table 3.4 shows two examples of mutant strains obtained using an evolutionary algorithm included in OptFlux, which searched for sets of knock-outs that increase the production levels of the target metabolites. Since some of the knock-outs implemented during the optimization process could alter the abundance of NADPH or NADH, the constraints derived from the curation process (Table 3.2) might not be applicable to the mutant strains simulated. Therefore, we used the improved flux distribution (Figure 3.2) in the mutant simulations performed, but did not applied the cofactor constraints (Table 3.2) to the mutant strains being optimized. In other words, because the mutant strains being tested might need to adapt to changes in cofactor metabolism, it would not be appropriate to continue assuming the same high NADPH/NADP⁺ and low NADH/NAD⁺.

Table 3.4. Description of the strain designs obtained with OptFlux for mevalonate and acetate production using the improved reference flux distribution (the standard names of the deleted genes are shown).

Target	Inactivated reactions	Product Yield (g/g glucose)
Mevalonate	<i>MTDI</i> : NAD ⁺ + 5,10-methylene-THF -> NADH + 5,10-methenyl-THF <i>LSC1</i> or <i>LSC2</i> : ADP(m) + HPO ₄ ²⁻ (m) + Succinyl-CoA(m) -> CoA(m) + Succinate(m) + ATP(m)	Corrected reference: 0.056
	<i>GDH1</i> and <i>GDH3</i> : H ⁺ + NH ₄ + NADPH + 2-oxoglutarate -> H ₂ O + L-glutamate + NADP ⁺	Original reference: 0.0075
Acetate	<i>ZWF1</i> : D-glucose-6-phosphate + NADP ⁺ -> H ⁺ + NADPH + 6-phosphogluconolactone	Corrected reference: 0.045
	<i>IDP2</i> : NADP ⁺ + Isocitrate -> CO ₂ + NADPH + 2-oxoglutarate <i>GDH2</i> : NAD ⁺ + H ₂ O + L-glutamate <-> H ⁺ + NADH + NH ₄ + 2-oxoglutarate <i>AAT2</i> : L-aspartate + 2-oxoglutarate <-> L-glutamate + Oxaloacetate	Original reference: 0.0

Table 3.4 shows a selection of the strain designs obtained with OptFlux for the optimization of mevalonate and acetate production by using the improved flux distribution during the strain optimization process. One strain design for each case study was included in Table 3.4, with detailed information about the gene knock-outs necessary to achieve the corresponding production yields. The results shown for the original reference were obtained by simulating the same strain designs with the flux distribution for Yeast 6 show in Figure 3.1.

For the mevalonate producing mutant three different metabolic reactions needed to be inactivated (corresponding to four genes) to achieve a yield of 0.056 g/g glucose. A detailed inspection of the changes in the flux distribution (Figure 3.4) revealed that the deletion of the NADPH-dependent glutamate dehydrogenase (*GDH1/GDH3*) causes a surplus in the availability of NADPH. The deletion of the NAD⁺-dependent 5,10-methylenetetrahydrofolate dehydrogenase (*MTDI*) avoids the transfer of electrons from NADPH to NAD⁺, forcing the NADPH to be used somewhere else. With the inactivation of the succinyl-CoA synthetase (gene name - *LSCI*, ID in Figure 3.4 - SUCOAS) the citric acid cycle is broken and part of the flux is redirected through pyruvate decarboxylase (PDC). Most of this flux returns to central metabolism using the enzymes from glyoxylate cycle. Nevertheless, this deletion causes an excess of acetyl-CoA in the mitochondria, which can be condensed into hydroxymethylglutaryl-CoA and transported to the cytosol. In that compartment, this metabolite can be converted to mevalonate while consuming NADPH, which alleviates the excess of this cofactor. In summary, two deletions are required to increase the amount of NADPH available in the cytosol, while the other increases the flux to hydroxymethylglutaryl-CoA (precursor of mevalonate). If the original reference flux distribution would be used, the amount of mevalonate excreted by this mutant would be around seven times lower (Table 3.4), which shows the importance of an accurate distribution of fluxes for phenotype simulation.

Experimental evidence is available to support part of this strategy presented here for mevalonate production. The inactivation of *GDH1* was followed by Nissen *et al.* to alter cofactor abundance in anaerobic conditions and decrease glycerol accumulation [54]. Furthermore, Scalcinati *et al.* also used the *GDH1* deletion to increase the flux through the mevalonate pathway as part of a metabolic engineering strategy to increase α -santalene production [19]. The production of vanillin and cubebol also benefited from the increased NADPH pool resulting from *GDH1* deletion [1, 5]. Given the experimental evidences supporting the *in silico* predictions, it seems that the improved reference led to better phenotypical predictions in the central carbon metabolism.

Regarding the acetate producing mutant (Table 3.4), there is also a large difference in the acetate yield depending on which reference flux distribution is used. A quick analysis of the deletions required revealed that, similarly to the mevalonate producing mutant, NADPH metabolism is involved. However, in this case, the key to produce acetate is

forcing the cytosolic NADP⁺-dependent aldehyde dehydrogenase (gene name – *ALD6*, ID in Figure 3.4 - ALD) to be the source of cytosolic NADPH (Figure 3.4).

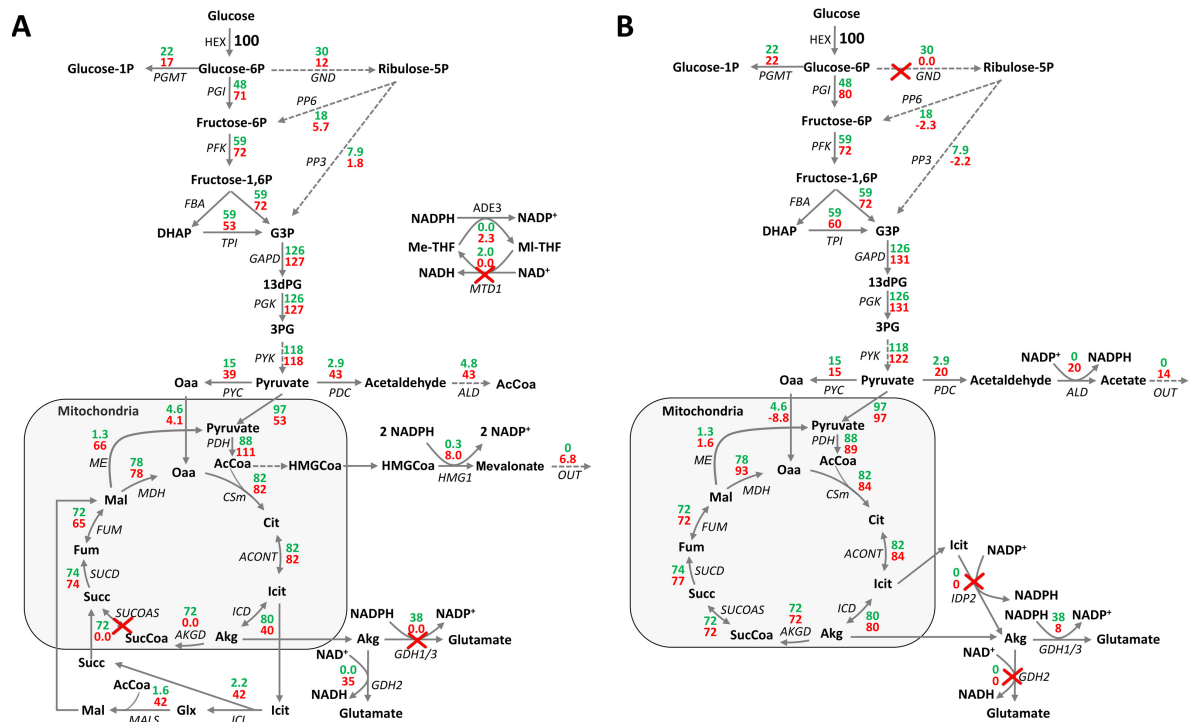


Figure 3.4.- Flux distributions simulated with linear MOMA for the mevalonate mutant (A) and the acetate mutant (B) using the improved flux distribution as the reference. The wild-type flux values are shown in green, the mutant values are shown in red and the knock-outs are signaled by a red cross. Reactions: ACONT- aconitase, ADE3- cytosolic C1-tetrahydrofolate synthase, AKGD- alpha-ketoglutarate dehydrogenase, ALD- aldehyde dehydrogenase, CSm- citrate synthase, FBA- fructose 1,6-bisphosphate aldolase, FUM- fumarase, GAPD- glyceraldehyde-3-phosphate dehydrogenase, GDH1/3- NADP⁺-dependent glutamate dehydrogenase, GDH2- NAD⁺-dependent glutamate dehydrogenase, GND- 6-phosphogluconate dehydrogenase, HEX- hexokinase, HMG1- HMG-CoA reductase, ICD- mitochondrial isocitrate dehydrogenase, ICL- isocitrate lyase, IDP2- cytosolic NADP-specific isocitrate dehydrogenase, MALS- malate synthase, MDH- mitochondrial malate dehydrogenase, ME- mitochondrial malic enzyme, MTD1- NAD-dependent 5,10-methylenetetrahydrofolate dehydrogenase, PDC- pyruvate decarboxylase, PDH- pyruvate dehydrogenase, PFK- phosphofruktokinase, PGI- phosphoglucose isomerase, PGK- 3-phosphoglycerate kinase, PGMT- phosphoglucomutase, PP3- sum of the non-oxidative reactions of the pentose phosphate pathway producing glyceraldehyde-3-phosphate, PP6- sum of the non-oxidative reactions of the pentose phosphate pathway producing fructose-6-phosphate, PYC- pyruvate carboxylase, PYK- pyruvate kinase, SUCD- succinate dehydrogenase, SUCOAS- succinyl-CoA ligase, THRS- threonine synthase, TPI- triose phosphate isomerase; Metabolites: 13dPG- 1,3-diphosphoglycerate, 3PG- 3-phosphoglycerate, AcCoA- acetyl-CoA, Akg- 2-oxoglutarate, Cit- citrate, DHAP- dihydroxyacetone-phosphate, Fum- fumarate, G3P- glyceraldehyde-3-phosphate, Glx- glyoxylate, Icit- isocitrate, HMGCoA- 3-hydroxy-3-methylglutaryl-CoA, Mal- L-malate, Oaa- oxaloacetate, Succ- succinate, SucCoA- succinyl-CoA.

To reach the acetate producing phenotype, glucose-6-phosphate dehydrogenase (gene name – *ZWF1*, ID in Figure 3.4 - GND) must be deleted so that the oxidative PPP is inactivated. Since this pathway is the main source of NADPH in the cytosol, the level of

this cofactor becomes severally reduced. Because the cytosolic NADP⁺-specific isocitrate dehydrogenase (*IDP2*) could be a possible source of NADPH it was also inactivated in this mutant. The final two deletions act on the other side of NADPH metabolism, i.e., on its consumption. The deletion of the NADH-dependent glutamate dehydrogenase (*GDH2*) is necessary to “force” the use of the NADPH dependent enzyme (*GDH1/GDH3*). With the same goal, deleting the cytosolic aspartate aminotransferase (*AAT2*) also increases the flux through the NADPH dependent glutamate dehydrogenase. All four deletions trigger rearrangements on the NADPH metabolism that culminate in an increased flux through the cytosolic NADP⁺-dependent aldehyde dehydrogenase (gene name – *ALD6*, ID in Figure 3.4 - ALD). Since this reaction becomes the main source of NADPH in the cytosol, the extra flux results in acetate accumulation.

Experimental evidences support that the inactivation of Glucose-6-phosphate dehydrogenase (*ZWF1*) in *S.cerevisiae* can decrease the availability of cytosolic NADPH, resulting in reduced growth on glucose minimal medium [55], methionine auxotrophy [56] and increased sensitivity to oxidizing agents [57]. The overexpression of the cytosolic NADP⁺-dependent aldehyde dehydrogenase (*ALD6*) has been shown to solve the methionine auxotrophy by increasing the availability of cytosolic NADPH [58]. Furthermore, even in the absence of these two genes there is evidence that the cytosolic NADP⁺-specific isocitrate dehydrogenase (*IDP2*) can also fulfill this role [59]. The compensation between these enzymes as the cytosolic NADPH source supports the behavior of the acetate producing mutant in Table 3.4 and indicates that the model was predicting accurately the NADPH metabolism.

3.4 Conclusion

An analysis of some of the *S. cerevisiae* GSMMs available in the literature revealed that the internal flux distributions predicted by FBA under fully aerobic conditions included some inconsistencies in central areas of the metabolism. The oxidative pentose phosphate pathway, for example, did not carry any flux in three out of the four models tested. Since the differences observed were originating from reactions containing NADH and NADPH, a strategy was implemented to improve the flux prediction of FBA simulations.

The results showed that many improvements could be achieved by applying constraints on reactions involved in the metabolism of NADPH and NADH. Although the flux distributions obtained are only valid under fully aerobic growth on glucose we showed that considerable improvements were possible, especially in the pentose phosphate pathway and in the metabolism of NADPH. Furthermore, the cofactor restrictions led to the discovery of structural errors in many of the models, leading to considerable improvements in central metabolic pathways.

When GSMMs are used for phenotype simulation in metabolic engineering, the internal flux distribution can be of vital importance. Here we showed that a careful curation of the network under fully aerobic conditions produced knock-out suggestions that are biological meaningful. Furthermore, the same knock-outs would result in quite distinct results if the original flux distribution is used, which shows the importance of the wild-type flux distribution for phenotype simulation methods.

Moreover, as many metabolic engineering strategies rely on unbalancing cofactor abundance, either NADH or NADPH, the results presented here demonstrate that, in order to have good predictions and a better design of improved mutants, it is of vital importance to have a good baseline prediction of flux distributions for the wild-type organism. Moreover, it is important to start considering to include comparisons of model predictions with *in vivo* flux distributions in the model construction and validation pipeline.

One of the surprising facts found was that the oldest of the models analyzed showed the best prediction of central carbon fluxes. This might be explained by the increased effort of manual curation spent on its construction [23]. *iFF708* is still preferred by many authors in metabolic engineering projects over the more recent GSMMs because of its better description of central metabolism [1, 4–6]. Since the modifications proposed here revealed that similar flux distributions can be obtained with more recent models, it is now possible to take advantage of better compartmentalization, elementally balance reactions and increased genome coverage, without sacrificing the quality of central metabolism fluxes.

References

1. Brochado AR, Matos C, Møller BL, et al. (2010) Improved vanillin production in baker's yeast through *in silico* design. *Microb Cell Fact* 9:84.
2. Fowler ZL, Gikandi WW, Koffas MAG (2009) Increased malonyl coenzyme A biosynthesis by tuning the *Escherichia coli* metabolic network and its application to flavanone production. *Appl Environ Microbiol* 75:5831–9.
3. Choi HS, Lee SY, Kim TY, Woo HM (2010) *In silico* identification of gene amplification targets for improvement of lycopene production. *Appl Environ Microbiol* 76:3097–105.
4. Bro C, Regenbreg B, Förster J, Nielsen J (2006) *In silico* aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metab Eng* 8:102–111.
5. Asadollahi MA, Maury J, Patil KR, et al. (2009) Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through in silico driven metabolic engineering. *Metab Eng* 11:328–34.
6. Otero JM, Cimini D, Patil KR, et al. (2013) Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory. *PLoS One* 8:e54144.
7. Patil KR, Rocha I, Förster J, Nielsen J (2005) Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics* 6:308.
8. Ranganathan S, Suthers PF, Maranas CD (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol* 6:e1000744.
9. Kim J, Reed JL, Maravelias CT (2011) Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. *PLoS One* 6:e24162.
10. Segrè D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A* 99:15112–7.
11. Shlomi T, Berkman O, Ruppin E (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A* 102:7695–700.
12. Brochado AR, Andrejev S, Maranas CD, Patil KR (2012) Impact of stoichiometry representation on simulation of genotype-phenotype relationships in metabolic networks. *PLoS Comput Biol* 8:e1002758.
13. Fell DA, Small JR (1986) Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem J* 238:781–6.

14. Varma A, Palsson BO (1994) Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/Technology* 12:994–998.
15. Lewis NE, Hixson KK, Conrad TM, et al. (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* 6:390.
16. Park JH, Lee KH, Kim TY, Lee SY (2007) Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation. *Proc Natl Acad Sci U S A* 104:7797–802.
17. Alper H, Jin Y-S, Moxley JF, Stephanopoulos G (2005) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab Eng* 7:155–64.
18. Jung YK, Kim TY, Park SJ, Lee SY (2010) Metabolic engineering of *Escherichia coli* for the production of polylactic acid and its copolymers. *Biotechnol Bioeng* 105:161–71.
19. Scalcinati G, Partow S, Siewers V, et al. (2012) Combined metabolic engineering of precursor and co-factor supply to increase alpha-santalene production by *Saccharomyces cerevisiae*. *Microb Cell Fact* 11:117.
20. Kuepfer L, Sauer U, Blank LM (2005) Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res* 15:1421–30.
21. Williams TCR, Poolman MG, Howden AJM, et al. (2010) A genome-scale metabolic model accurately predicts fluxes in central carbon metabolism under stress conditions. *Plant Physiol* 154:311–23.
22. Cheung CYM, Williams TCR, Poolman MG, et al. (2013) A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. *Plant J* 75:1050–61.
23. Förster J, Famili I, Fu P, et al. (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* 13:244–53.
24. Mo ML, Palsson BO, Herrgård MJ (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol* 3:37.
25. Österlund T, Nookaew I, Bordel S, Nielsen J (2013) Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst Biol* 7:36.
26. Herrgård MJ, Swainston N, Dobson P, et al. (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol* 26:1155–60.

27. Nookaew I, Jewett MC, Meechai A, et al. (2008) The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol* 2:71.
28. Duarte NC, Herrgård MJ, Palsson BØ (2004) Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* 14:1298–309.
29. Heavner BD, Smallbone K, Price ND, Walker LP (2013) Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database (Oxford)* 2013:bat059.
30. Rocha I, Maia P, Evangelista P, et al. (2010) OptFlux: an open-source software platform for *in silico* metabolic engineering. *BMC Syst Biol* 4:45.
31. Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* 5:264–276.
32. Burgard AP, Vaidyaraman S, Maranas CD Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnol Prog* 17:791–7.
33. Rocha M, Maia P, Mendes R, et al. (2008) Natural computation meta-heuristics for the *in silico* optimization of microbial strains. *BMC Bioinformatics* 9:499.
34. Becker SA, Feist AM, Mo ML, et al. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* 2:727–38.
35. Nookaew I, Olivares-Hernández R, Bhumiratana S, Nielsen J (2011) Genome-scale metabolic models of *Saccharomyces cerevisiae*. *Methods Mol Biol* 759:445–63.
36. Osterlund T, Nookaew I, Nielsen J (2012) Fifteen years of large scale metabolic modeling of yeast: developments and impacts. *Biotechnol Adv* 30:979–88.
37. Edwards JS, Palsson BO (2000) Metabolic flux balance analysis and the *in silico* analysis of *Escherichia coli* K-12 gene deletions. *BMC Bioinformatics* 1:1.
38. Kauffman KJ, Prakash P, Edwards JS (2003) Advances in flux balance analysis. *Curr Opin Biotechnol* 14:491–6.
39. Van Dijken JP, Weusthuis RA, Pronk JT (1993) Kinetics of growth and sugar consumption in yeasts. *Antonie Van Leeuwenhoek* 63:343–52.
40. Haselbeck RJ, McAlister-Henn L (1993) Function and expression of yeast mitochondrial NAD- and NADP-specific isocitrate dehydrogenases. *J Biol Chem* 268:12116–22.
41. Satrustegui J, Bautista J, Machado A (1983) NADPH/NADP⁺ ratio: regulatory implications in yeast glyoxylic acid cycle. *Mol Cell Biochem* 51:123–127.

42. Dijken JP, Scheffers WA (1986) Redox balances in the metabolism of sugars by yeasts. *FEMS Microbiol Lett* 32:199–224.
43. Canelas AB, van Gulik WM, Heijnen JJ (2008) Determination of the cytosolic free NAD/NADH ratio in *Saccharomyces cerevisiae* under steady-state and highly dynamic conditions. *Biotechnol Bioeng* 100:734–43.
44. Voet D, Voet JG (2011) Introduction to Metabolism. *Biochemistry*, 4th ed. Wiley, pp 560–562
45. Leber R, Landl K, Zinser E, et al. (1998) Dual Localization of Squalene Epoxidase, Erg1p, in Yeast Reflects a Relationship between the Endoplasmic Reticulum and Lipid Particles. *Mol Biol Cell* 9:375–386.
46. Castegna A, Scarcia P, Agrimi G, et al. (2010) Identification and functional characterization of a novel mitochondrial carrier for citrate and oxoglutarate in *Saccharomyces cerevisiae*. *J Biol Chem* 285:17359–70.
47. Palmieri L, Agrimi G, Runswick MJ, et al. (2001) Identification in *Saccharomyces cerevisiae* of two isoforms of a novel mitochondrial transporter for 2-oxoadipate and 2-oxoglutarate. *J Biol Chem* 276:1916–22.
48. Baudry K, Swain E, Rahier A, et al. (2001) The effect of the *erg26-1* mutation on the regulation of lipid metabolism in *Saccharomyces cerevisiae*. *J Biol Chem* 276:12702–11.
49. Mannhaupt G, Stucka R, Pilz U, et al. (1989) Characterization of the prephenate dehydrogenase-encoding gene, *TYR1*, from *Saccharomyces cerevisiae*. *Gene* 85:303–11.
50. Jouhten P, Rintala E, Huuskonen A, et al. (2008) Oxygen dependence of metabolic fluxes and energy generation of *Saccharomyces cerevisiae* CEN.PK113-1A. *BMC Syst Biol* 2:60.
51. Gombert AK, Moreira dos Santos M, Christensen B, Nielsen J (2001) Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. *J Bacteriol* 183:1441–51.
52. Dubois E, Grenson M, Wiame J-M (1974) The Participation of the Anabolic Glutamate Dehydrogenase in the Nitrogen Catabolite Repression of Arginase in *Saccharomyces cerevisiae*. *Eur J Biochem* 48:603–616.
53. Avendaño A, Deluna A, Olivera H, et al. (1997) *GDH3* encodes a glutamate dehydrogenase isozyme, a previously unrecognized route for glutamate biosynthesis in *Saccharomyces cerevisiae*. *J Bacteriol* 179:5594–7.
54. Nissen TL, Kielland-Brandt MC, Nielsen J, Villadsen J (2000) Optimization of Ethanol Production in *Saccharomyces cerevisiae* by Metabolic Engineering of the Ammonium Assimilation. *Metab Eng* 2:69–77.

55. Blank LM, Kuepfer L, Sauer U (2005) Large-scale ^{13}C -flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol* 6:R49.
56. Thomas D, Cherest H, Surdin-Kerjan Y (1991) Identification of the structural gene for glucose-6-phosphate dehydrogenase in yeast. Inactivation leads to a nutritional requirement for organic sulfur. *EMBO J* 10:547–53.
57. Noga I, Johnston M (1990) Isolation and characterization of the *ZWF1* gene of *Saccharomyces cerevisiae*, encoding glucose-6-phosphate dehydrogenase. *Gene* 96:161–169.
58. Grabowska D, Chelstowska A (2003) The *ALD6* gene product is indispensable for providing NADPH in yeast cells lacking glucose-6-phosphate dehydrogenase activity. *J Biol Chem* 278:13984–8.
59. Minard KI, McAlister-Henn L (2005) Sources of NADPH in yeast vary with carbon source. *J Biol Chem* 280:39890–6.

CHAPTER 4

TDPS - Turnover dependent phenotypic simulation: a quantitative constraint-based simulation method that accommodates all main strain design strategies

The uncertain relation between genotype and phenotype makes strain engineering a laborious trial and error procedure. Constraint-based modelling methodologies can expedite the strain engineering process by helping in the search for interesting genetic modification targets. Although the search for gene knock-outs is fairly established with *in silico* methodologies, most strain design methods still model gene up/down-regulations by forcing the corresponding flux values to pre-calculated levels without having in consideration the availability of resources.

The method described in this chapter, Turnover Dependent Phenotypic Simulation (TDPS), was designed with the goal of simulating quantitatively the phenotype of strains with complex genotypes in a resource conscious manner. In TDPS the flux values are never forced to pre-calculated levels because the modelling methodology used takes into account the availability of resources in the network by assuming that the production turnover of a metabolite can be used as an indication of its abundance. TDPS was validated using metabolically engineered *S. cerevisiae* strains available in the literature by comparing the simulated and experimental production yields of the target metabolite.

The information presented in this Chapter is being prepared for submission to a peer reviewed journal:

Pereira R., Vilaça P., Nielsen J., Rocha I. TDPS - Turnover dependent phenotypic simulation: a quantitative constraint-based simulation method that accommodates all main strain design strategies.

4.1 Introduction

In strain engineering, the relation between a genetic alteration and its effects on the phenotypical behavior of an organism is rarely straightforward, which usually results in a trial and error engineering procedure. Constraint-based modeling of metabolic fluxes is a fast alternative to test the effect of genetic alterations at the genome-scale level and several methods have been developed for this purpose (reviewed in chapter 2 and reference [1]). When compared with other modeling approaches, constraint-based modelling can offer a genome-scale representation of metabolism without the need for kinetic information [2]. However, even with all the advances in constraint-based modelling, the simulation of common strain engineering designs involving up and down-regulation of a few genes can still be challenging.

The list of strain design methodologies gathered in Table 4.1 represent a selection of the most complete in terms of the variety of genetic modifications included in their formulation. Although there are plenty of additional computational methods devoted to strain design (reviewed in chapter 2 and reference [1]), the majority of them is focused on gene knock-outs and do not include in their formulation up- and down-regulations. Furthermore, many of these methods were designed to be used in an optimization context, i.e., they can indicate possible metabolic engineering targets, but cannot simulate quantitatively the effects that a set of genetic modifications would have on the organism's phenotype. Since the simulation of strain phenotypes can be difficult, the validation of strain design methods is usually made by comparing the genetic targets obtained *in silico*, using optimization algorithms, with previously described metabolically engineered strains [3–7]. Alternatively, it is also possible to validate *in silico* strain designs by implementing them experimentally [8–10], but this process is quite time and resource consuming. Given the amount of rational strain designs available in the literature, it would be quite interesting to use them to validate the prediction capabilities of strain design algorithms.

Among the options shown in Table 4.1, the most oriented for simulation purposes is the under/over-expression plugin [7] included in OptFlux [11], which allows the simulation of deletions and up/ down-regulations of genes/reactions according to user specified levels. The flux constraints created by the under/over-expression plugin can then be used in combination with the diverse simulation methodologies included in OptFlux (FBA [12, 13], pFBA [14], MOMA [15], LMOMA [16], ROOM [17] and MIMBL [18]). The main

disadvantage of this algorithm is that up/down-regulations are restricted to metabolic reactions that have a non-zero flux value in the reference flux distribution. As a consequence, it is impossible to simulate gene/reaction activations or the introduction of heterologous genes.

Table 4.1– Summary of constraint-based strain design methods that include in their formulation up/down-regulation or activations

Method (year)	Focus	Types of modifications allowed	Brief description
OptReg (2005) [3]	Optimization (target discovery) ^a	Up-regulation (activation ^c), down-regulation and deletion	Gene regulations are modeled by constraining the flux bounds to predefined intervals
OptForce (2010) [4]	Optimization (target discovery)	Up-regulation (activation), down-regulation and deletion	Gene regulations are not modeled ^b
EMILiO (2011) [5]	Optimization (target discovery)	Up-regulation (activation), down-regulation and deletion	Gene regulations are modeled by optimizing the upper and lower flux bounds of individual fluxes
Under/over expression plugin for OptFlux (2012) [7, 11]	Simulation and optimization (target discovery)	Up-regulation, down-regulation and deletion	Gene regulations are modeled by constraining the fluxes to a multiple of a reference (wild-type) flux
CosMos (2013) [6]	Optimization (target discovery)	Up-regulation (activation), down-regulation and deletion	Similar to OptForce, but the flux constraints can take any value in a continuous interval
Redirector (2013) [9]	Optimization (target discovery) ^a	Up-regulation (activation) and down-regulation	Gene regulations are modeled by adding the target fluxes to the objective function
Proportional Flux Forcing (2014) [10]	Simulation and optimization (target discovery)	Activation and deletion	Reaction activation/addition is modeled in a turnover dependent formulation

^a Although this method was not designed to simulate phenotypes its formulation can be used for this purpose

^b This strain optimization method inverts the optimization problem by analyzing which changes of flux are important to force flux to the target metabolite

^c Since one of the methods from this list can only up-regulate reactions that are active in a reference state, the activation of a reaction is defined here as the possibility of up-regulating inactive reactions

Other methods have different strategies to simulate the activation of reactions: OptReg [3] constrains the target flux to a value higher than a fraction of the maximum, EMILiO [5] manipulates the target flux to any feasible value, and Redirector [9] adds the flux variable to the objective function of the linear problem. However, these methodologies do not take into consideration the availability of resources when they “force” the flux through the activated metabolic reaction. Therefore, even if the substrate for the activated reaction is not being produced in the wild-type network, there will be a cascade of activations in the rest of the network to supply it. Understandably, no biological meaning can be derived

from flux simulations where a whole pathway can be activated by simply forcing flux through the terminal metabolic step.

In contrast, the framework for reaction activation proposed by Ip et al., denominated Proportional Flux Forcing (PFF) [10], offers a much more attractive alternative to the methods mentioned above. This algorithm makes the flux through an activated reaction dependent on the abundance of its substrate. Since constraint-based flux simulations do not take into account metabolite concentrations, the abundance is estimated from the metabolite production turnover (sum of the fluxes producing the precursor metabolite) [10]. To simulate the flux through an activated reaction, PFF makes the target flux value proportional to the turnover of the precursor metabolite and simulates the distribution of fluxes using FBA [12, 13]. The authors developed this strategy in order to find gene deletions that increase the turnover (availability) of the precursor metabolite and decrease the flux in reactions competing for the same precursor. However, this strategy is limited to a single reaction activation and to knock-outs, which makes it inadequate for the simulation of more complex metabolically engineered strains.

Here, we propose a new approach to simulate the flux distribution of complex strain designs using a new constraint-based modelling method called Turnover Dependent Phenotypic Simulation (TDPS). TDPS was designed with the goal of simulating quantitatively the phenotype of strains with complex genotypes in a resource conscious manner. Besides gene deletions and down-regulations, TDPS can also simulate the up-regulation of metabolic reactions, as well as the introduction of heterologous genes or the activation of “dormant” reactions. In TDPS the flux values are never forced to pre-calculated levels because the modelling methodology used takes into account the availability of resources in the network by assuming that the production turnover of a metabolite can be used as an indication of its abundance. Furthermore, a newly developed objective function that promotes network rigidity was implemented in conjunction with the turnover dependent genetic constraints, in order to allow TDPS to predict the flux rearrangements in mutant strains. Based on Stephanopoulos and Vallin’s [19] work on the rigidity of metabolic networks, this objective function reinforces network rigidity by minimizing the change in the split ratios between a reference network and the disturbed one. The goal was to keep the overall flux pattern in the network close to the wild-type even when selected split ratios are being manipulated. The complete simulation method was validated using phenotypic data of several strains available in the literature.

4.2 Turnover Dependent Phenotypic Simulation

Motivated by the lack of simulation-oriented constraint-based methods capable of simulating complex strain designs at the genome-scale level, we developed the TDPS algorithm. Unlike other algorithms that manipulate flux values to emulate genetic modifications, TDPS uses a strategy similar to PFF [10] to take into consideration the availability of precursors for the regulated reactions. TDPS can simulate most types of genetic modifications used for strain engineering purposes and does it in a resource conscious manner. Therefore, the flux distributions obtained can be interpreted biologically and can be used to improve the physiological knowledge about the metabolic network. Furthermore, TDPS allows a quantitative approach to *in silico* strain design that can be validated with data available from metabolically engineered strains.

4.2.1 Modeling genetic modifications

In comparison with other methodologies, constraint-based modeling is a simple and fast alternative to estimate the flux distribution within a metabolic network [2]. However, it has the disadvantage of only modelling flux values in steady-state conditions, which leaves out the explicit modelling of metabolite and enzyme concentrations [2]. Given the absence of metabolite concentrations and enzymatic kinetic parameters, the most common strategy to simulate gene up and down-regulation is to force the regulated fluxes to a specific value or interval. While this strategy can help understand the system wide changes caused by the up/down-regulation of a certain reaction, it has no biological significance because a forced flux does not take into consideration the availability of precursors.

One possibility to infer the availability of a metabolite using constraint-based modelling is by calculating the magnitude of the fluxes producing or consuming it (turnover) [18, 20]. Here, we use the concept of production turnover (T_m), to refer to the sum of all fluxes producing metabolite m multiplied by the corresponding stoichiometric coefficients. Using T_m as an indicator for metabolite abundance, the activation of a reaction n can be modelled assuming that a certain fraction of T_m will be consumed by the activated reaction (as suggested by Ip et al. [10]). In this way, the flux through reaction n (V_n) becomes proportional to the T_m and can even be zero if no reaction in the network is producing metabolite m , i.e., if T_m is equal to zero. Unlike other methods from Table 4.1,

this heuristic does not force the flux to a certain value or interval, instead makes it dependent on the availability of the precursor.

This concept has been extended in TDPS, as shown in Table 4.2, by assuming that the up-regulation of a reaction n consuming metabolite m can be seen as an increase in the fraction ($X_{m,n}$) of the T_m that has to be consumed by reaction n in comparison to a reference fraction value. In the same line of thought, a down-regulated reaction can be modelled by decreasing the fraction ($X_{m,n}$) of the T_m that can be consumed by reaction n . The reference fraction values mentioned above are calculated from a wild-type flux distribution representative of normal growing conditions. For example, if reaction n is responsible for 25 % of the consumption of metabolite m under normal conditions, it can be up-regulated by forcing the flux to be at least 50 % of T_m . It is important to note that T_m is not a reference T_m value, but a variable in the simulation, which means that the flux is not forced to be above a certain value, but to be at least 50 % of whatever value T_m will have in the final simulation. The inactivation of a metabolic reaction is modeled as in all constraint-based methods by setting the flux value (V_n) through the inactivated reaction to zero.

Another innovation of TDPS in comparison to PFF is the possibility of regulating reactions with more than one reactant (Table 4.2). If the up-regulated/activated reaction has multiple reactants, the created constraints assume that the flux through reaction n (V_n) will have to be higher than the lowest restriction imposed by the availability of precursors, which can be formulated with an OR operator (Table 4.2). In accordance, the flux through a down-regulated reaction with multiple reactants also has to be lower than the lowest constraint imposed by the availability of precursors. In this case, an AND operator is used instead (Table 4.2).

The methodologies described in Table 4.2 try to mimic *in vivo* enzyme activity at two levels: firstly, the abundance of the substrate is modelled using metabolite production turnovers (T_m); secondly the fraction values ($X_{m,n}$) can be interpreted as the result of a combined effect of enzymatic kinetic parameters (K_m and K_{cat}) and enzyme concentration. This is a simplification of reality that does not encompass other cellular events such as the different types of regulation, but it is considerably less computationally intensive than other modelling alternatives and can be used to model metabolism at the genome scale.

Table 4.2 - Constraints applied in TDPS to the flux values (V_n) for each type of genetic modification

Type of modification	Flux Constraints
Inactivation	$V_n = 0$
Up-regulation/Activation	$V_n \cdot S_{m_0,n} > X_{m_0,n} \cdot T_{m_0} \vee \dots \vee V_n \cdot S_{m_i,n} > X_{m_i,n} \cdot T_{m_i}, m \in M_n, M_n = \{m_0, \dots, m_i\}, V_n \geq 0$
Down-regulation	$V_n \cdot S_{m_0,n} < X_{m_0,n} \cdot T_{m_0} \wedge \dots \wedge V_n \cdot S_{m_i,n} < X_{m_i,n} \cdot T_{m_i}, m \in M_n, M_n = \{m_0, \dots, m_i\}, V_n \geq 0$

M_n - Subset containing all the precursors of reaction n , $S_{m,n}$ -Stoichiometric coefficient of metabolite m in reaction n , T_m - Production turnover of metabolite m in the simulation, V_n - Flux value of reaction n in the simulation, $X_{m,n}$ - Turnover fraction value (see text)

4.2.2 Objective function reinforces network rigidity

As shown in Table 4.2, TDPS uses metabolite turnover fractions ($X_{m,n}$) to model genetic up- and down-regulations. The constraints applied serve the purpose of directing a certain fraction (higher or lower than a threshold) of the available precursor to the modified metabolic reaction. In order to obtain a flux distribution for the mutant organism it is also necessary to define a cellular goal that represents how the network would react to the disturbances applied. Since the constraints applied manipulate split ratios, it would be appropriate to define an objective function that represents how a cell controls the partition of flux at important nodes. Stephanopoulos and Vallin [19] have introduced the idea that some of the nodes in the network have rigid flux ratios, which have evolved with the objective of supplying biomass growth with relatively stable ratios of building blocks. Therefore, the concept of turnover fractions ($X_{m,n}$) was extended to create an objective function that promotes stability in the split flux ratios over the entire network.

The rigidity based objective function was formulated as the minimization of two different terms (Figure 4.1): the first term represents the difference between the amount of metabolite m that reaction n is consuming and the amount it was supposed to consume according to the calculated turnover fractions ($X_{m,n}$); the second term is a penalty for the activation of reactions that were inactive in the reference flux-distribution. While the first term promotes the stability of the split ratios, the second term prevents flux from getting dispersed into pathways that were not active in the reference flux distribution. The $X_{m,n}$ values used in the objective function are calculated using a wild-type (reference) flux distribution as the base and are modified depending on the genetic modifications applied

(Figure 4.1). The detailed calculations regarding these parameters are presented in section 4.2.3.

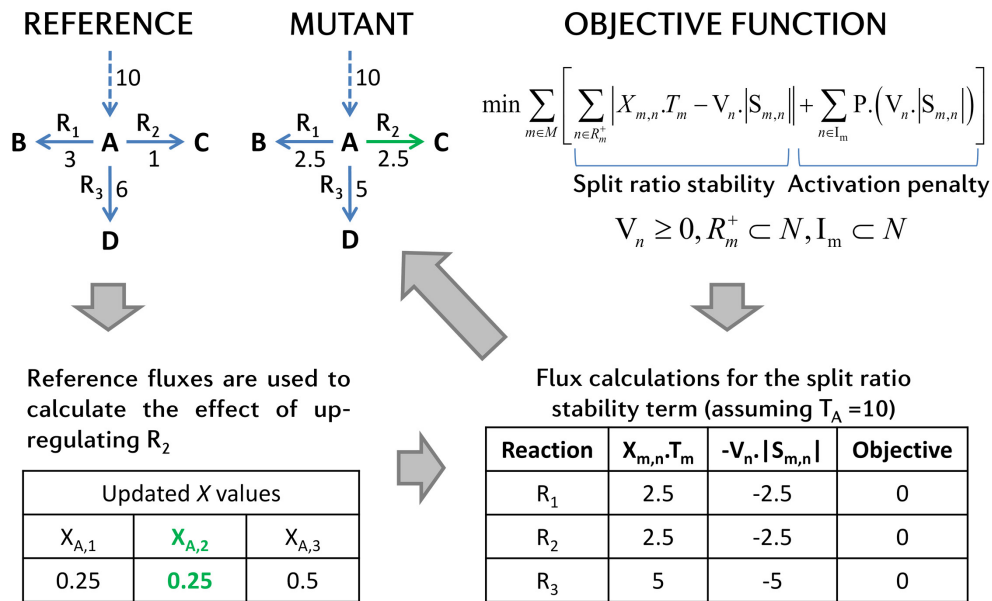


Figure 4.1- Flux calculations performed by the objective function implemented in TDPS using the up-regulation of reaction R_2 as an example. I_m - set of inactive reactions in the reference flux distribution that can consume metabolite m , M - set of all metabolites, N - set of all reactions, P - penalty constant for activated reactions, R_m^+ - set containing all the active reactions consuming metabolite m in the reference flux distribution, $S_{m,n}$ -stoichiometric coefficient of metabolite m in reaction n , T_m - production turnover of metabolite m in the simulation, V_n - flux value of reaction n in the simulation, $X_{m,n}$ - normalized fraction value.

In the example shown in Figure 4.1, the up-regulation of R_2 was achieved by increasing the fraction ($X_{A,2}$) of the production turnover of A (T_A) that reaction R_2 was forced to consume (from a value of 0.1 in the wild-type to 0.25 in the mutant). As a consequence the $X_{m,n}$ values for R_1 and R_3 decreased accordingly to compensate for the up-regulation of R_2 (see section 4.2.3 for the details about the calculations). In the simplistic network shown in Figure 4.1, the fluxes can easily be readjusted to comply with all the modified fraction values dictated by the algorithm. However, if the modified fraction values were imposed rigidly in a genome-scale metabolic network, the result would most likely be an unfeasible flux distribution. Therefore, the first term of the objective function (Figure 4.1) was formulated as the minimization of the difference between the amount of the total T_m that each reaction should be consuming (as dictated by $X_{m,n}$ values) and the real value.

The big advantage of an objective function that uses a reference set of split ratios in its formulation rather than flux values (such as MOMA [15] or LMOMA [16]) is that the network can adjust to a genetic modification without an implicit penalty in the objective value. While simulation methods that promote flux stability always try to keep the

network exactly like the reference, the objective function of TDPS allows fluxes to be flexible. Nevertheless, the flux flexibility is limited by the minimization of the disturbances in the split ratios computed from the reference network. As shown in Figure 4.1, the distribution of fluxes changed to accommodate the up-regulation of R_2 without penalizing the objective value. Furthermore, the objective function promoted the stability of the split ratios between R_1 and R_3 , i.e., in the mutant network the ratio between the flux in R_1 and R_3 remained the same as in the reference flux distribution (1:2). This shows how TDPS tries to keep the ratios between the non-modified fluxes stable when a disturbance is introduced in a metabolic node.

4.2.3 Algorithm flowchart

TDPS is capable of simulating complex strain designs using two separate components: a newly developed modelling heuristic that can handle most types of genetic modifications (Table 4.2) and an innovative objective function that promotes flux ratio rigidity after network disturbances (Figure 4.1). Figure 4.2 shows how these two components were integrated into a robust simulation algorithm designed specifically to allow quantitative phenotypical analysis of complex metabolically engineered strains. The first step in the simulation process is to choose a Genome Scale Metabolic Model (GSMM) and define the environmental conditions by specifying a set of exchange fluxes (carbon source, nitrogen source, oxygen availability, etc.). These parameters are then used to compute a reference (wild-type) flux distribution using pFBA [14] as the simulation method. It is worthy of note that the reference flux distribution obtained with pFBA might change depending on the linear programming solver used. Since pFBA minimizes the total sum of fluxes in the network, assuming maximum biomass growth, it cannot determine a unique flux distribution if there are parallel pathways with equivalent stoichiometry in the network. Therefore, depending on the solver used, the reference flux distributions can have small variations in the fluxes through parallel pathways. In order to assure the reproducibility of the results obtained with TDPS, the reference flux distribution used should always be kept constant (the same is valid when other simulation tools that require a reference flux distribution are used, such as MOMA).

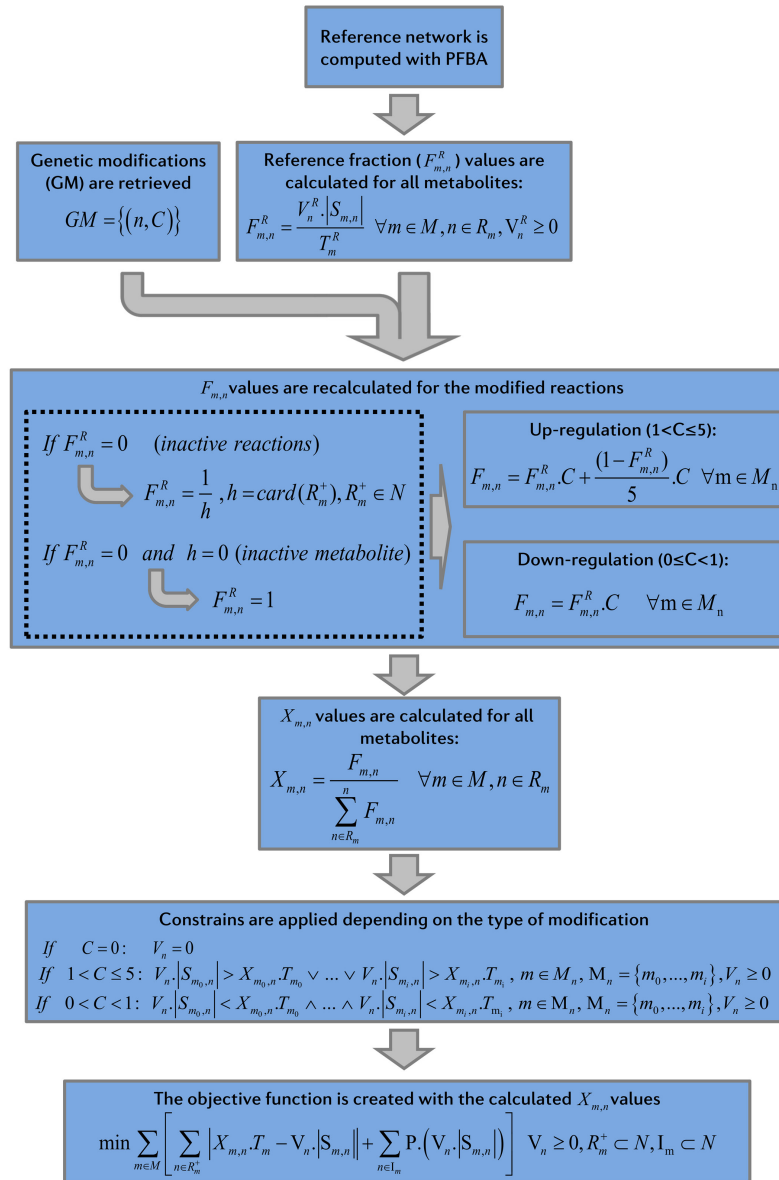


Figure 4.2- Flowchart describing the calculations performed by the TDPS algorithm (the detailed implementation is provided in section 4.3.1). C - regulation parameter, $F_{m,n}$ - fraction of the production turnover consumed by reaction n (before normalization), $F_{m,n}^R$ - fraction of the production turnover consumed by reaction n in the reference network, GM - set of genetic modifications, h - number of active reactions consuming metabolite m , I_m - set of inactive reactions in the reference flux distribution that can consume metabolite m , M - set of all metabolites, M_n - subset containing all the precursors of reaction n , N - set of all reactions, P - penalty constant for activated reactions, R_m - set containing all the reactions that can consume metabolite m , R_m^+ - set containing all the active reactions consuming metabolite m in the reference flux distribution, $S_{m,n}$ -stoichiometric coefficient of metabolite m in reaction n , T_m - production turnover of metabolite m in the simulation, T_m^R - production turnover of metabolite m in the reference flux distribution, V_n - flux value of reaction n in the simulation, V_n^R - flux value for reaction n in the reference flux distribution, $X_{m,n}$ - normalized fraction value.

Using a reference flux distribution, the TDPS algorithm then starts the calculations necessary to obtain the $X_{m,n}$ values required to formulate the genetic modification constraints (Table 4.2) and the objective function (Figure 4.1). To avoid confusing

nomenclatures of the variables, the intermediate turnover fraction values used in the calculations shown in Figure 4.2 were named F values. First, the wild-type or reference fraction values ($F_{m,n}^R$) are calculated using the reference flux distribution, by computing the fraction of the reference production turnover for metabolite m (T_m^R) that each of the consumer reactions (R_m) is using (Figure 4.2). The set of desired genetic modifications (GM) is then used to modify the $F_{m,n}^R$ values according to the regulation parameter C associated with each modification. If a reaction is active in the reference flux distribution ($F_{m,n}^R > 0$) and it is targeted for up-regulation ($1 < C \leq 5$), then the algorithm increases the fraction value for all its precursors (M_n) according to the formula shown in Figure 4.2. The up-regulation formula is composed of two terms: the first term increases the modified fraction value ($F_{m,n}$) up to five times the reference fraction and the second term is important to allow significant up-regulation when the magnitude of $F_{m,n}^R$ in the wild-type organism is so low that multiplying it by 5 would not change its value significantly. If an active reaction is a target for down-regulation ($0 < C < 1$) then its $F_{m,n}$ value is decreased by direct multiplication with the C parameter. In both cases the magnitude of the C value should mimic the severity of the regulation applied to the organism and it is up to the user to find the value that best describes the genetic modification undertaken.

When the reaction targeted for modification is inactive in the reference flux distribution ($F_{m,n}^R = 0$), it is necessary to estimate the fraction of the total production turnover that this reaction will consume when activated. TDPS uses the number of reactions that compete for the same substrate (h) to estimate the average $F_{m,n}^R$ value for each precursor required by the activated reaction. The number of reactions that compete for the same substrate (h) refers to the active consumers of metabolite m (R_m^+) in the reference flux distribution and is calculated using the cardinality ($card$) of R_m^+ . With the estimated $F_{m,n}^R$ value, TDPS can then apply the same rules described above for up- and down-regulation, depending on the value of the C parameter. Although the up-regulation of an inactive reaction is quite straightforward to understand, it is also biologically relevant to down-regulate inactive reactions. For example, in heavily engineered strains it is very likely that some of the modifications applied will result in a regulatory response that induces the activation of reactions that are usually off. These activated reactions can be important down-regulation

targets and it would not be appropriate to exclude them. Therefore, the estimated $F_{m,n}^R$ value is valid for both up and down-regulations.

There is also the possibility that an inactive reaction ($F_{m,n}^R = 0$) contains one or several precursors that are not being actively consumed by any reaction in the network ($h = 0$). In this case the TDPS algorithm assumes that the activated reaction will be the only consumer for those metabolites ($F_{m,n}^R = 1$).

After manipulating the $F_{m,n}$ values in accordance to the genetic modifications, the sum of all the fraction values for a certain metabolite is no longer guaranteed to be equal to one. To normalize them back to unitary fractions, each $F_{m,n}$ value is divided by the summed fraction values for the respective metabolite, which yields the $X_{m,n}$ values (Figure 4.2). Using the rules described in Table 4.2, the $X_{m,n}$ values are then employed in the creation of flux constraints for all the reactions present in the *GM* set. Finally, the $X_{m,n}$ values are used in the creation of the split ratio stability term of the objective function (Figure 4.2), which is then minimized by the solver in order to produce the mutant flux distribution.

4.3 Methods

4.3.1 Implementation of the TDPS algorithm

The TDPS method was implemented in the JAVA programming language within the libraries of the OptFlux software [11]. OptFlux is an *in silico* metabolic engineering framework that allows the user to import and manipulate GSMMs and perform strain optimization/simulation tasks. With the integration of TDPS in OptFlux it is possible to access additional functionalities such as strain optimization algorithms and flux visualization tools, which expands the possible applications of TDPS.

4.3.1.1 Pre-calculations

The reference flux distributions that are required for the execution of TDPS were calculated with pFBA using biomass maximization as the objective function [14]. Using

the reference flux distribution, the fraction values ($F_{m,n}^R$) were calculated for all reactions (n) consuming metabolite m ($n \in R_m$):

$$F_{m,n}^R = \frac{V_n^R \cdot |S_{m,n}|}{T_m^R} \quad \forall m \in M, n \in R_m, V_n^R \geq 0 \quad \text{Equation 4.1}$$

Where V_n^R is the reference flux value in reaction n , $S_{m,n}$ is the stoichiometric coefficient of metabolite m in reaction n and M is the set of all the metabolites in the network. The reference production turnovers (T_m^R) were calculated for all metabolites in the network ($m \in M$) using the formula:

$$T_m^R = \sum_{n \in P_m} V_n^R \cdot |S_{m,n}| \quad \forall m \in M, n \in P_m, V_n^R \geq 0 \quad \text{Equation 4.2}$$

Where P_m is set of reactions that can produce metabolite m .

4.3.1.2 Splitting reversible reactions in two half-reactions

During the simulations it was necessary to create variables for the production turnovers of the mutant flux distribution. These variables are required in the formulation of the flux constraints derived from the genetic modifications and for the formulation of the objective function. However, the existence of reversible reactions in metabolic models makes the creation of a variable describing the production turnover of a metabolite using linear programming quite challenging. Since the flux variable can be either positive or negative it was necessary to split all reversible reactions in two positive half-reactions in order to have flux variables that are specific for the forward or reverse direction:

$$V_n = V_n^{Pos} - V_n^{Neg} \quad \forall n \in U \quad \text{Equation 4.3}$$

$$0 \leq V_n^{Pos} \leq V_n^{UB} \cdot B_n^{Pos} \quad B_n^{Pos} \in \{0,1\} \quad \text{Equation 4.4}$$

$$V_n^{LB} \cdot B_n^{Neg} \leq -V_n^{Neg} \leq 0 \quad B_n^{Neg} \in \{0,1\} \quad \text{Equation 4.5}$$

$$B_n^{Pos} + B_n^{Neg} \leq 1 \quad \text{Equation 4.6}$$

Where V_n^{Pos} is the flux value in the forward direction of reaction n during the simulation, V_n^{Neg} is the flux value in the reverse direction of reaction n during the simulation, U is the set containing all reversible reactions in the network, V_n^{LB} is the lower flux bound of reaction n , V_n^{UB} is the upper flux bound of reaction n , and B_n^{Pos} and B_n^{Neg} are the binary variables used to prevent both half-reactions from being active simultaneously. In addition to the variables created for the half-reactions it was also necessary to prevent both half-reactions to be active simultaneously so that futile cycles could not arise. Binary variables were used to create additional constraints so that only one of the half-reactions could be active at a time (Equation 4.6). As a consequence of using binary variables, the simulations performed with TDPS become a Mixed Integer Linear Programming (MILP) problem.

4.3.1.3 TDPS implementation

The mathematical implementation of the objective function shown in Figure 4.1 required a minor modification in the split ratio stability term to avoid that a flux distribution full of zeros would be a possible solution for the minimization problem. Therefore, the reference turnover value was included in the split ratio stability term to ensure that the turnover in the mutant flux distribution stayed close to the reference value. Using the modified objective function, TDPS was formulated mathematically as follows:

$$\min \sum_{m \in M} \left[\sum_{n \in R_m^+, n \notin U} |X_{m,n} \cdot (2T_m - T_m^R) - V_n \cdot |S_{m,n}|| + \sum_{n \in I_m, n \notin U} P \cdot (V_n \cdot |S_{m,n}|) + \sum_{n \in R_m^+, n \in U, S_{m,n} < 0} |X_{m,n} \cdot (2T_m - T_m^R) - V_n^{Pos} \cdot |S_{m,n}|| + \sum_{n \in I_m, n \in U, S_{m,n} < 0} P \cdot (V_n^{Pos} \cdot |S_{m,n}|) + \sum_{n \in R_m^+, n \in U, S_{m,n} > 0} |X_{m,n} \cdot (2T_m - T_m^R) - V_n^{Neg} \cdot |S_{m,n}|| + \sum_{n \in I_m, n \in U, S_{m,n} > 0} P \cdot (V_n^{Neg} \cdot |S_{m,n}|) \right], V_n \geq 0 \quad \text{Equation 4.7}$$

$$\text{s.t. } T_m = \sum_{n \in P_m, n \notin U} V_n \cdot |S_{m,n}| + \sum_{n \in P_m, n \in U, S_{m,n} > 0} V_n^{Pos} \cdot |S_{m,n}| + \sum_{n \in P_m, n \in U, S_{m,n} < 0} V_n^{Neg} \cdot |S_{m,n}| \quad \forall m \in M, V_n \geq 0 \quad \text{Equation 4.8}$$

$$V_n^{LB} \leq V_n \leq V_n^{UB} \quad \text{Equation 4.9}$$

$$V_n = V_n^{Pos} - V_n^{Neg} \quad \forall n \in U \quad \text{Equation 4.10}$$

$$0 \leq V_n^{Pos} \leq V_n^{UB} \cdot B_n^{Pos} \quad B_n^{Pos} \in \{0,1\} \quad \text{Equation 4.11}$$

$$V_n^{LB} \cdot B_n^{Neg} \leq -V_n^{Neg} \leq 0 \quad B_n^{Neg} \in \{0,1\} \quad \text{Equation 4.12}$$

$$B_n^{Pos} + B_n^{Neg} \leq 1 \quad \text{Equation 4.13}$$

$$V_n > \min \left(\frac{X_{m_0,n} \cdot T_{m_0}}{|S_{m_0,n}|}, \dots, \frac{X_{m_i,n} \cdot T_{m_i}}{|S_{m_i,n}|} \right), m \in M_n, M_n = \{m_0, \dots, m_i\}, n \notin U, V_n \geq 0, n \in GM_{UR} \quad \text{Equation 4.14}$$

$$V_n^{Pos} > \min \left(\frac{X_{m_0,n} \cdot T_{m_0}}{|S_{m_0,n}|}, \dots, \frac{X_{m_i,n} \cdot T_{m_i}}{|S_{m_i,n}|} \right), m \in M_n, M_n = \{m_0, \dots, m_i\}, n \in U, n \in GM_{UR}, D_n = 1 \quad \text{Equation 4.15}$$

$$V_n^{Neg} > \min \left(\frac{X_{m_0,n} \cdot T_{m_0}}{|S_{m_0,n}|}, \dots, \frac{X_{m_i,n} \cdot T_{m_i}}{|S_{m_i,n}|} \right), m \in M_n, M_n = \{m_0, \dots, m_i\}, n \in U, n \in GM_{UR}, D_n = -1 \quad \text{Equation 4.16}$$

$$V_n < \min \left(\frac{X_{m_0,n} \cdot T_{m_0}}{|S_{m_0,n}|}, \dots, \frac{X_{m_i,n} \cdot T_{m_i}}{|S_{m_i,n}|} \right), m \in M_n, M_n = \{m_0, \dots, m_i\}, n \notin U, V_n \geq 0, n \in GM_{DR} \quad \text{Equation 4.17}$$

$$V_n^{Pos} < \min \left(\frac{X_{m_0,n} \cdot T_{m_0}}{|S_{m_0,n}|}, \dots, \frac{X_{m_i,n} \cdot T_{m_i}}{|S_{m_i,n}|} \right), m \in M_n, M_n = \{m_0, \dots, m_i\}, n \in U, n \in GM_{DR}, D_n = 1 \quad \text{Equation 4.18}$$

$$V_n^{Neg} < \min \left(\frac{X_{m_0,n} \cdot T_{m_0}}{|S_{m_0,n}|}, \dots, \frac{X_{m_i,n} \cdot T_{m_i}}{|S_{m_i,n}|} \right), m \in M_n, M_n = \{m_0, \dots, m_i\}, n \in U, n \in GM_{DR}, D_n = -1 \quad \text{Equation 4.19}$$

$$V_n = 0 \quad n \in GM_{KO} \quad \text{Equation 4.20}$$

$$\sum_{n \in N} S_{m,n} \cdot V_n = 0 \quad \forall m \in M \quad \text{Equation 4.21}$$

Where GM_{UR} is the set of reactions that should be up-regulated, GR_{DR} is the set of reactions that should be down-regulated, GM_{KO} is the set of reactions that should be knocked-out and D_n is the direction of reaction n that should be modified if a reversible reaction is targeted for up- or down-regulation. As shown in equations 4.7 and 4.8, the number of terms present in the equations had to be increased to adjust for the existence of reversible reactions in the metabolic model. This was achieved by replacing V_n by the appropriate half-reaction (V_n^{Pos} or V_n^{Neg}) for all reversible reactions ($n \in U$). The flux constraints formulated in equations 4.14 - 4.19 also take into consideration which direction of reaction n should be modified by checking a directionality parameter (D_n)

provided by the user. For modifying the forward reaction of n , D_n must be set to 1 and for the reverse reaction to -1.

4.3.1.4 TDPS_FBA implementation

TDPS_FBA was implemented to test the robustness of the predictions obtained with the TDPS algorithm. Initially, a normal TDPS simulation was performed and the optimal value of the objective function (OF) was stored. Subsequently, a new TDPS problem was formulated with all the initial constraints but including a new one:

$$\sum_{m \in M} \left[\sum_{n \in R_m^+, n \notin U} |X_{m,n} \cdot (2T_m - T_m^R) - V_n \cdot |S_{m,n}|| + \sum_{n \in I_m, n \notin U} P \cdot (V_n \cdot |S_{m,n}|) + \sum_{n \in R_m^+, n \in U, S_{m,n} < 0} |X_{m,n} \cdot (2T_m - T_m^R) - V_n^{Pos} \cdot |S_{m,n}|| + \sum_{n \in I_m, n \in U, S_{m,n} < 0} P \cdot (V_n^{Pos} \cdot |S_{m,n}|) + \sum_{n \in R_m^+, n \in U, S_{m,n} > 0} |X_{m,n} \cdot (2T_m - T_m^R) - V_n^{Neg} \cdot |S_{m,n}|| + \sum_{n \in I_m, n \in U, S_{m,n} > 0} P \cdot (V_n^{Neg} \cdot |S_{m,n}|) \right] \leq T \cdot OF, V_n \geq 0 \quad \text{Equation 4.22}$$

Where T is the tolerance constant used to relax the initial objective function value. The objective function used in TDPS was then replaced in TDPS_FBA by the maximization of the biomass production flux.

4.3.1.5 Solver

The academic version of IBM ILOG CPLEX optimization studio V12.5.1 64bit was used as the LP and MILP solver in all simulations performed with TDPS. Given the size and complexity of the problems generated by TDPS, the CPLEX parameter “NumericalEmphasis” was activated in all simulations performed to avoid occasional numerical instability issues. All other CPLEX parameters were used as predefined by the manufacturer.

4.3.2 Toy model simulations

TDPS was used to simulate the effects of different types of genetic modifications on a toy model using a penalty constant for activated reactions (P) equal to 1, the C parameter for up-regulation was assumed to be 2 and for down-regulation it was assumed to be 0.5. The reference flux distribution used in the simulations was calculated using pFBA by maximizing the biomass reaction R7.

4.3.3 Case study simulations

4.3.3.1 Genome-scale models and reference flux distributions

The consensus GSMM for *S. cerevisiae* version 6.06 [21] was downloaded in the SBML format from the project's website: <http://sourceforge.net/projects/yeast/files/> and modified according to Appendix B (Tables B1-B5). The Yeast 6.06 GSMM was imported into OptFlux 3.07 [11] and all stoichiometric coefficients were multiplied by 1000 in order to help solving occasional numerical instability issues reported by CPLEX.

The case-studies that were collected from the literature to validate TDPS contained data obtained with two types of cultivation methods, chemostat and batch cultures. In order to simulate glucose limited chemostat conditions with a dilution rate of 0.1 h^{-1} , the glucose uptake rate was set to $1.15 \text{ mmol}/(\text{gCDW}\cdot\text{h})$ and the uptake rates of ammonia, phosphate, sulfate and oxygen were unconstrained. The modifications applied in chapter 3 to improve the reference flux distribution were also applied here during the computation of the reference flux distribution.

With the purpose of simulating batch cultures, the glucose uptake rate was estimated from batch cultivations of *S. cerevisiae* CENPK (initial glucose concentration of 20 g/L) and set to $11.4 \text{ mmol}/(\text{gCDW}\cdot\text{h})$. To mimic respiro-fermentative metabolism, the oxygen uptake rate was limited to $1.3 \text{ mmol}/(\text{gCDW}\cdot\text{h})$ in order to obtain a final biomass yield on glucose characteristic of *S. cerevisiae* (0.12 g/g) [22]. During the computation of this flux distribution, the constraints described in chapter 3 were also applied, but in this case only the ones related to NADPH metabolism. Since during respiro-fermentative metabolism the abundance of NADH is altered in comparison to glucose limited conditions, the NADH related constraints used in chapter 3 were not used for calculating this reference set of fluxes.

4.3.3.2 Parametrization of the TDPS simulation of mutant phenotypes

The simulation of mutant phenotypes was performed with TDPS using a penalty constant for activated reactions (P) of 50. This value was found to provide a good balance between the maintenance of the split ratio values (first term of the objective function in Figure 4.1) and the inhibition of reaction activation (second term of the objective function in Figure 4.1) in the conditions that Yeast 6.06 was tested.

The drain reactions were excluded from the objective function formulation because they do not represent any metabolic entity and are only present in the model to allow exchange of metabolites with the medium. Some metabolites were also excluded from the TDPS calculations because they are available in excess and it would not be logical that they were included in the resource based formulation used to model genetic modifications. In chemostat conditions the metabolites excluded were: H_2O , H^+ , SO_4^{2-} , NH_4^+ , HPO_4^{2-} , Fe^{2+} and O_2 . Coenzyme A was also included in the list of metabolites to be excluded because it has a passive role in metabolism, i.e., it is always recycled in its original form and its availability should not be a limiting factor for metabolic fluxes at steady-state. In batch conditions, because oxygen availability was limiting, O_2 was re-included in the calculations.

Given the lack of knowledge regarding which values for the C parameter would be more appropriate to simulate the genetic modifications extracted from the literature, a random number generator was used to create virtual mutant populations of at least 500 individuals with diversified phenotypes for the same set of genetic conditions. For each up-regulation 500 random numbers between 1 and 5 were attributed to the C parameter ($1 < C \leq 5$), while for each down-regulation 500 random numbers between 0 and 1 ($0 < C < 1$) were generated and for reaction inactivation the C parameter was set to zero. The final values presented for each virtual strain include the average production yield obtained for the target metabolite, the minimum and the maximum production yields obtained for the randomly generated population.

TDPS_FBA simulations were carried by relaxing 10 % ($T = 1.1$) the optimal value of the objective function obtained with TDPS and maximizing the biomass formation with the added constraint.

4.4. Results and discussion

4.4.1 TDPS toy-model validation

TDPS was developed to simulate mutant phenotypes of engineered strains using GSMMs. However, GSMMs are difficult to visualize due to their size and it would be challenging to validate some of the aspects of TDPS using models with over 1000 metabolic reactions. Therefore, a small toy-model was created to verify if all the components of

TDPS were working as expected. Figure 4.3 shows the toy-model developed, along with the changes in the flux distribution resulting from different types of genetic modifications that can be simulated with TDPS.

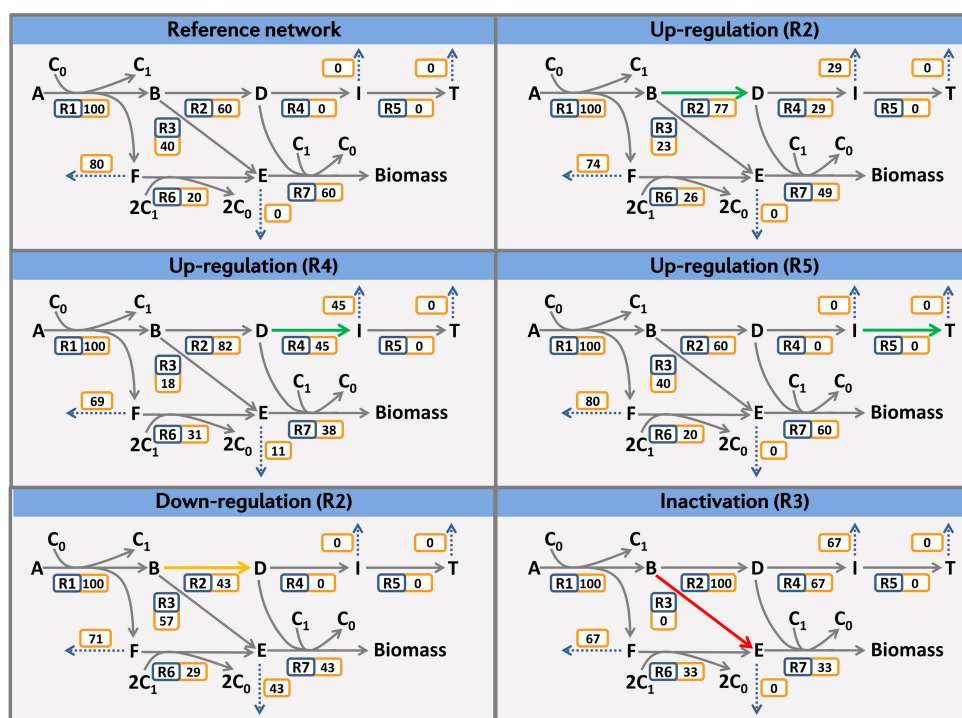


Figure 4.3- Flux distributions obtained with TDPS for different types of genetic modifications using a toy-model. For up-regulations the C parameter was set to 2 and for down-regulation to 0.5. The reaction activation penalty constant used was 1.

The first step in the formulation of TDPS consists in the computation of a wild-type flux distribution using pFBA [14] (Figure 4.2) to serve as a reference for the calculations performed by the algorithm. In order to obtain the reference flux distribution for the toy-model (Figure 4.3), the biomass production (R7) was maximized and a fixed substrate (metabolite A) uptake rate of 100 flux units was assumed. After examining the wild-type set of fluxes, we selected three testing targets for up-regulation (R2, R4, and R5), one for down-regulation (R2) and one for inactivation (R3). The up-regulations were simulated by setting the C parameter to 2 and for the down-regulation a value of 0.5 was used instead.

The up-regulation of R2 was selected as an example for the amplification of a reaction active in the reference flux distribution. By defining the C parameter strength to 2, the constraints generated by the algorithm resulted in an increase of 17 flux units in R2 when comparing to the reference value (Figure 4.3). The flux increase occurred at the expense

of a symmetric decrease in the flux of R3, which is the only reaction competing for the same precursor (metabolite B). Although the absolute flux increase does not appear very high, the flux ratio between R2 and R3 increased from 3:2 to almost 8:2. As a result of up-regulating R2, the flux in reaction R4 was also activated in order to dispose of the excessive production of metabolite D, which cannot be excreted or used for biomass formation (Figure 4.3). The fact that R5 remained inactive shows that the penalty term for activated reactions (Figure 4.1) is working as expected, because only the indispensable reactions were activated.

One of the features of TDPS is to make the flux through an up-regulated/activated reaction dependent on the availability of its precursors. In order to test this feature, two inactive reactions with differently available precursors were selected as targets for up-regulation (R4 and R5). Regarding the up-regulation of R4, Figure 4.3 shows how the availability of metabolite D results in a significant flux flowing through the activated reaction. On the contrary, when reaction R5 is up-regulated, there are no changes in the flux distribution in comparison with the reference network (Figure 4.3). The contrasting up-regulation phenotypes observed are the result of a considerable difference between the production turnovers of metabolites D and I. Therefore, simulating the activation of a metabolic reaction in TDPS does not imply that there will be any changes in the network, which highlights the advantage of using TDPS to simulate reaction activations in comparison to other methods that manipulate flux values directly (OptReg [3] and EMILiO [5]) or even the objective function (Redirector [9]).

Although the up-regulation of R5 on its own resulted in a phenotype with zero flux over that reaction, TDPS is capable of modeling genetic modifications in an integrated manner. Therefore, if R4 is up-regulated simultaneously with R5, the production turnover for metabolite I will increase and R5 will become active as a consequence (data not shown). The fact that TDPS allows genetic modifications to affect each other is a consequence of the strategy used to handle production turnovers. In TDPS, the metabolite production turnover used to create the flux constraints is not a constant obtained from the reference flux distributions, but a variable inside the MILP problem. Consequently, TDPS can be used to find metabolic engineering strategies that are composed of genetic alterations that work synergistically to reach the desired production goal.

To illustrate the application of the TDPS algorithm to down-regulations, reaction R2 was selected as a down-regulation target. As shown in Figure 4.3, the attenuation of R2 resulted in a decrease of 17 flux units, which caused a concomitant increase of the same magnitude in the flux through R3. This down-regulation limited the amount of metabolite D available in the network, which reduced the biomass formation rate. In terms of split ratios, the down-regulation of R2 resulted in a decrease from 3:2 to 1.5:2.

TDPS handles gene deletions in the same way as other simulation methods available, so the only innovation introduced in this case is how the objective function (Figure 4.1) calculates the flux distribution of the inactivation mutant. In Figure 4.3 we show how TDPS predicts a drop in biomass formation in response to a decrease in the availability of metabolite E and an increase in reaction R6 in order to recycle excess of metabolite C₁.

Given the size of the toy-model, it was not possible to observe the objective function promoting stability at the flux ratio level. Since the model is quite small and we assumed a fixed substrate uptake rate, there are not many degrees of freedom in the network to allow significant flux flexibility in the network. Nevertheless, the toy-model served the purpose of testing how TDPS handled each type of genetic modification in an easy to visualize scale.

4.4.2 TDPS validation with experimental results

After the initial validation using a toy-model, the simulation method was put to the test with physiological data gathered from the literature. Given the importance of the reference flux distribution in the simulation outcome of strain design methods (chapter 3), it would not be advisable to test TDPS using models of uncertain quality. Since the genome-scale models for *S. cerevisiae* were previously curated in an effort to improve their wild-type flux distribution during growth on glucose (chapter 3), this microorganism was chosen as the case-study. Besides the availability of curated flux distributions, there was also the advantage of *S. cerevisiae* being a model organism with abundant physiological data available, as well as plenty of strain engineering strategies accessible in the literature.

In order to select strain engineering strategies appropriate for the validation of the TDPS algorithm, we restricted the case-study search to *S. cerevisiae* strains grown aerobically on minimal media with glucose as the sole carbon source. Furthermore, to validate the

steady-state simulations obtained with TDPS it would be preferred to work only with data derived from glucose limited chemostat experiments. However, chemostat fermentations are labor intensive and as a consequence the majority of metabolic engineering studies available in the literature use batch cultures in bioreactors or shake-flasks for strain characterization instead. To include both batch and chemostat data in the validation of TDPS, we created two different reference flux distributions: the first one describes glucose-limited conditions characteristic of chemostat cultures growing at a dilution rate of 0.1 h^{-1} and the second one simulates respiro-fermentative conditions typical of the glucose consumption phase in batch fermentations.

Besides providing a reference flux distribution it is also necessary to select appropriate C parameters (see section 4.2) for each genetic modification simulated in TDPS. However, the C parameter is dependent on the expression levels and kinetic characteristics of the enzyme used, which means that it would not be appropriate to specify the same value for all the modifications under analysis. In order to get an overview of the range of phenotypes that the TDPS algorithm can predict, each strain was simulated by attributing 500 random values to each C parameter within the ranges defined in the methods section. Besides the average value we also show the minimum and maximum values obtained in the simulations to illustrate the simulation range of possible phenotypes.

To simulate the phenotypes for the strains obtained in the literature, two different versions of TDPS were used: the regular version as it is described in section 4.2 and a relaxed version (TDPS_FBA) where the optimal value of the objective function is allowed to relax and is included as an additional constraint in a new problem (described in section 4.3.1.4). TDPS_FBA was useful to determine how robust the phenotypes obtained with TDPS are, i.e., to check if the production yields obtained with TDPS can vary without compromising the objective function value considerably.

4.4.2.1 Case-study 1: PHB production in *S. cerevisiae*

The first case-study selected to validate TDPS consists in the production of polyhydroxybutyrate (PHB) in *S. cerevisiae* using the genes of the *phaCAB* operon from *Ralstonia eutropha* [23, 24]. In this case-study Kocharin et al. determined the PHB production in batch cultures (bioreactor and shake-flask) of a strain expressing three heterologous genes from *R. eutropha* (SCKK005) [23]. The authors also improved the

availability of cytosolic acetyl-CoA by expressing a deregulated acetyl-CoA synthetase (acs_{SE}^{L641P}) from *Salmonella enterica* and by up-regulating the cytosolic aldehyde dehydrogenase (*ALD6*) and alcohol dehydrogenase (*ADH2*) endogenous of *S. cerevisiae*, which resulted in the strain SCKK006 that was characterized in batch cultures (bioreactor and shake-flask) [23] as well as in glucose limited chemostats [24].

Table 4.3 shows the PHB production yields calculated from the results of Kocharin et al. [23, 24] along with the corresponding yields simulated with TDPS. The simulated yields shown for batch conditions in Table 4.3 were obtained specifically for a respiro-fermentative metabolism in glucose, which makes their value comparable only with the glucose consumption phase of bioreactor cultivations. Since the shake-flask yields were calculated at the end of the fermentation, the values shown include both the respiro-fermentative phase on glucose and the growth on fermentation products phase. Therefore, the shake-flask yields shown in Table 4.3 cannot be compared directly with the simulated yields obtained with TDPS but can still be helpful in the validation of the algorithm.

Table 4.3- TDPS validation case-study using PHB production in *S. cerevisiae*

Strain	Genotype relevant for simulations	<i>In vivo</i> yields (mg PHB/g glucose)	TDPS yields (mg PHB/g glucose)
SCKK005 [23]	Gene additions: <i>phaA</i> (EC 2.3.1.9), <i>phaB</i> (EC 1.1.1.36), <i>phaC</i> (EC 2.3.1.-)	Bioreactor cultivation - Glucose phase: 0.020 mg/g - Final yield: 0.097 mg/g Shake Flask - Final yield: 0.5 mg/g	Batch simulation: TDPS: 1.0 mg/g Min: 0.26 mg/g Max: 1.6 mg/g TDPS_FBA: 1.0 mg/g Min: 0.30 mg/g Max: 1.6 mg/g
SCKK006 [23, 24]	Gene additions: <i>phaA</i> (EC 2.3.1.9), <i>phaB</i> (EC 1.1.1.36), <i>phaC</i> (EC 2.3.1.-), acs_{SE}^{L641P} (EC 6.2.1.1) Gene up-regulations: <i>ALD6</i> (EC 1.2.1.4), <i>ADH2</i> * (EC 1.1.1.1)	Bioreactor cultivation - Glucose phase: 0.13 mg/g - Final yield: 2.3 mg/g Shake Flask - Final yield: 13 mg/g	Batch simulation: TDPS: 21 mg/g Min: 11 mg/g Max: 31 mg/g TDPS_FBA: 21 mg/g Min: 9.9 Max: 30 mg/g
		Chemostat (D = 0.1 h⁻¹) Steady state yield: 2.6 mg/g	Chemostat simulation: TDPS: 50 mg/g Min: 12 mg/g Max: 91 mg/g TDPS_FBA: 10 mg/g Min: 2.6 mg/g Max: 17 mg/g

**ADH2* up-regulation was not included in the simulations (see text)

Regarding strain SCKK005 (Table 4.3), the average yield simulated with TDPS (1.0 mg/g) was two orders of magnitude higher than the experimental value obtained in the glucose phase of the bioreactor fermentation (0.020 mg/g). Furthermore, even the

minimum value simulated with TDPS (0.26 mg/g) was still one order of magnitude above the experimental value (0.020 mg/g). However, the final PHB yield reported for shake-flask cultivations (0.5 mg/g) was around 5 times higher than the corresponding value in bioreactors (0.097 mg/g). Assuming that the yield in the glucose phase in the shake-flask cultivations would also be 5 times higher than the bioreactor counterpart, we estimated the respiro-fermentative yield in shake-flasks to be close to 0.1 mg/g. When this estimation is compared to the lowest value simulated with TDPS (0.26 mg/g), the difference between simulated and experimental values decreases significantly.

In comparison with SCKK005, the strain SCKK006 includes the additional expression of acs_{SE}^{L641P} from *S. enterica* and the up-regulation of *ALD6* and *ADH2* in order to improve the availability of acetyl-CoA in the cytosol of *S. cerevisiae*. Since the up-regulation of *ADH2* was implemented to improve ethanol utilization after glucose is exhausted, this gene was not included in the simulations performed with TDPS. Looking at Table 4.3, the improvement in the PHB production yield observed for the bioreactor cultivation was of 6.5 times in the glucose phase and 23 times overall. Furthermore, in shake-flask cultivations the final yield improvement was of 26 times compared to strain SCKK005. TDPS predicted on average an improvement of 21 times in the production yield of PHB in respiro-fermentative conditions. Although this value is above the observed improvement for the glucose phase, it is quite similar to the overall improvement obtained in bioreactor and shake-flask cultivations. As seen for strain SCKK005, the production levels predicted by TDPS for strain SCKK006 (21 mg/g) are two orders of magnitude above the experimental value determined in bioreactor cultivations (0.13 mg/g). Even if we consider the minimum value simulated with TDPS (11 mg/g) the difference is still quite high. Again, only the shake-flask yields can approximate the values simulated with TDPS.

The consistently higher PHB production levels determined computationally can be attributed to an overestimation of the reference fraction value (F^R) in the calculations performed by the TDPS algorithm. As described in section 4.2, when a reaction is activated in the metabolic network, given the absence of reference data, the F^R is estimated from other reactions in the network that consume the same precursor. In this case the estimated F^R value for the acetoacetyl-CoA reductase (*phaB*) was very high, which caused the PHB yields to be overestimated independently of the C parameter associated with this up-regulation. In biological terms, this means that the performance of

the PHB biosynthetic enzymes introduced in *S. cerevisiae* was below the estimation made by TDPS. Consequently, the simulations assumed that the capacity of the heterologous enzymes to compete for precursors with the endogenous enzymes was higher than in reality, which resulted in the differences observed. Nevertheless, the magnitude of the variation between simulations and experiments was considerably lower in shake-flask fermentations, which might indicate a better activity of the PHB biosynthetic enzymes in these conditions.

The final data presented in Table 4.3 refers to experiments performed with the strain SCKK006 in chemostat cultures. As seen for the batch cultures, the simulated yield (50 mg/g) is higher than the experimentally determined value (2.6 mg/g) but in this case the difference is only around one order of magnitude. Looking at the interval of simulated yields, the minimum yield predicted by TDPS (12 mg/g) is only 5 times higher than the experimental value. Furthermore, the minimum yield predicted with TDPS_FBA (relaxed version of TDPS) is actually coincident with the experimentally determined value (2.6 mg/g). When compared with the results obtained for the batch cultures, the yields simulated in glucose limited conditions are much closer to the simulated values. The better apparent performance of the TDPS algorithm in glucose limited conditions might be a consequence of the superior activity of the PHB producing enzymes under these cultivation conditions in comparison with respiro-fermentative metabolism. Alternatively, the assumptions made in the creation of a reference flux distribution for respiro-fermentative conditions might be inadequate for an accurate simulation of production yields in batch cultures. However, the limited data available on chemostat conditions did not allow us to draw a confident conclusion about the differences observed between the conditions tested.

4.4.2.2 Case-study 2: 3-HP production in *S. cerevisiae*

The biosynthesis of 3-hydroxypropionic acid (3-HP) is not native to *S. cerevisiae* but it can be achieved from malonyl-CoA by a double step reduction. In the work of Chen et al. and Shi et al. [25, 26], the authors introduced into *S. cerevisiae* the bi-functional malonyl-CoA reductase (*mcr*) from *Chloroflexus aurantiacus* in order to allow the production of 3-HP from malonyl-CoA in this microorganism. Furthermore, the authors also created a series of strains with improved 3-HP production yields that we used here as case-studies to validate the simulation capability of TDPS. In Table 4.4 the results obtained from both

publications are shown, together with the simulated yields calculated with the TDPS simulation method.

Strains HPY01 and HPY15 include the expression of the *mcr* gene from *C. aurantiacus* and showed a basal production level of 3-HP that relied on the natural availability of malonyl-CoA in *S. cerevisiae*. As shown in Table 4.4, the 3-HP production yield in the glucose phase of the bioreactor cultivation was very close to the simulated yield obtained with TDPS (2.4 mg/g vs. 1.3 mg/g), which indicates that the simulation method estimated quite accurately the availability of malonyl-CoA in the cell. Although the experimental yield was not inside the simulated interval (0.44 – 2.2 mg/g), the predictions given by TDPS were still quite remarkable for this strain when compared with the PHB case-study. In this case, the production yields in shake-flask were comparable to the bioreactor yields and did not offer any additional insight.

Strain HPY04 was engineered for an improved availability of malonyl-CoA by up-regulating the acetyl-CoA carboxylase gene (*ACCI*) [25]. The shake-flask cultivation of this strain resulted in an overall increase in the 3-HP production yield of 1.7 times in comparison with a predicted improvement of 2.2 times given by TDPS (Table 4.4). We could also estimate approximately the 3-HP yield during the glucose phase in shake-flasks by assuming that 41% of the final yield is produced during the respiro-fermentative phase (value calculated from strain HPY18). The resulting value, 3.2 mg/g, was inside in the simulated interval (0.42 - 4.5 mg/g) and was quite close to the average value generated with TDPS (2.8 mg/g).

In addition to strain HPY04, the *ACCI* up-regulation genotype was also implemented in strain HPY18, but in this case the *ACCI* gene was further modified to improve its enzymatic activity [26]. The 3-HP yields of strain HPY18 during cultivation in bioreactors improved 2.4 times in the glucose consumption phase and 3.4 times overall in comparison with strain HPY15 (Table 4.4). The 2.4 times improvement observed in the glucose phase is remarkably similar to the average improvement of 2.2 times obtained from the TDPS simulations. In regard to the production yield in the glucose phase of the bioreactor cultivation, the value reported for strain HPY18 (5.8 mg/g) falls outside the interval predicted by TDPS (0.42 - 4.5 mg/g). However, if we consider the TDPS_FBA results, the simulated interval was expanded just enough to include the experimental value.

Table 4.4- TDPS validation case-study using 3-hydroxypropionic acid (3-HP) production in *S. cerevisiae*

Strain	Genotype relevant for simulations	<i>In vivo</i> yields (mg 3-HP/g glucose)	TDPS yields (mg 3-HP/g glucose)
HPY01 [25] / HPY15 [26]	Gene additions: <i>mcr</i> (EC 1.2.1.75 + EC 1.1.1.298)	Shake Flask - Final yield: 4.6 mg/g Bioreactor cultivation - Glucose phase: 2.4 mg/g - Final yield: 4.1 mg/g	Batch simulation: TDPS: 1.3 mg/g Min: 0.44 mg/g Max: 2.2 mg/g TDPS_FBA: 1.3 mg/g Min: 0.45 mg/g Max: 2.2 mg/g
HPY04 [25] / HPY18 [26]	Gene additions: <i>mcr</i> (EC 1.2.1.75 + EC 1.1.1.298) Gene up-regulations: <i>ACCI</i> (EC 6.4.1.2)	Shake Flask - Final yield: 7.7 mg/g Bioreactor cultivation** - Glucose phase: 5.8 mg/g - Final yield: 14 mg/g	Batch simulation: TDPS: 2.8 mg/g Min: 0.42 mg/g Max: 4.5 mg/g TDPS_FBA: 3.3 mg/g Min: 0.51 Max: 5.8 mg/g
HPY05 [25]	Gene additions: <i>mcr</i> (EC 1.2.1.75 + EC 1.1.1.298), <i>acs_{SE}^{L641P}</i> (EC 6.2.1.1) Gene up-regulations: <i>ALD6</i> (EC 1.2.1.4), <i>ADH2*</i> (EC 1.1.1.1)	Shake Flask - Final yield: 7.0 mg/g	Batch simulation: TDPS: 0.88 mg/g Min: 0.27 mg/g Max: 1.8 mg/g TDPS_FBA: 1.0 mg/g Min: 0.33 Max: 1.9 mg/g
HPY06 [25]	Gene additions: <i>mcr</i> (EC 1.2.1.75 + EC 1.1.1.298), <i>acs_{SE}^{L641P}</i> (EC 6.2.1.1) Gene up-regulations: <i>ALD6</i> (EC 1.2.1.4), <i>ADH2*</i> (EC 1.1.1.1), <i>ACCI</i> (EC 6.4.1.2)	Shake Flask - Final yield: 11 mg/g	Batch simulation: TDPS: 19 mg/g Min: 5.6 mg/g Max: 49 mg/g TDPS_FBA: 17 mg/g Min: 5.4 mg/g Max: 35 mg/g
HPY09 [25]	Gene additions: <i>mcr</i> (EC 1.2.1.75 + EC 1.1.1.298), <i>gapN</i> (EC 1.2.1.9)	Shake Flask - Final yield: 6.0 mg/g	Batch simulation: TDPS: 1.2 mg/g Min: 0.42 mg/g Max: 2.1 mg/g TDPS_FBA: 1.3 mg/g Min: 0.43 mg/g Max: 2.1 mg/g
HPY11 [25]	Gene additions: <i>mcr</i> (EC 1.2.1.75 + EC 1.1.1.298), <i>acs_{SE}^{L641P}</i> (EC 6.2.1.1), <i>gapN</i> (EC 1.2.1.9) Gene up-regulations: <i>ALD6</i> (EC 1.2.1.4), <i>ADH2*</i> (EC 1.1.1.1), <i>ACCI</i> (EC 6.4.1.2) Gene deletions: <i>MLS1</i> (EC 2.3.3.9)	Shake Flask - Final yield: 23 mg/g	Batch simulation: TDPS: 116 mg/g Min: 14 mg/g Max: 160 mg/g TDPS_FBA: 80 mg/g Min: 6.6 mg/g Max: 130 mg/g

**ADH2* up-regulation was not included in the simulations (see text)

** In the strain used in this cultivation (HPY18) the up-regulation of *ACCI* was boosted by inhibiting its post-translational regulation [26]

The remaining strains presented in Table 4.4 were only characterized by the authors in shake-flask cultures, which hampered the comparison of the overall yields against the simulations. However, the overall improvement in comparison to HPY01 should still indicate if the TDPS simulations were capturing the cellular adaptation when faced with the additional disturbances. Strain HPY05 has an improved acetyl-CoA availability when compared to HPY01 because of the additional expression of a deregulated acetyl-CoA synthetase from *S. enterica* (*acs_{SE}^{L641P}*) and the up-regulation of *ALD6* and *ADH2*. As explained for the PHB case-study, the up-regulation of *ADH2* was not included in any of the simulations because the amplification of this gene was targeted for activation only after the glucose in the media was exhausted. The performance of the strain HPY05 in shake flask revealed an improvement of 1.5 times compared to HPY01 (Table 4.4). However, the yield predicted by TDPS actually decreased in comparison with the control strain. This result is quite odd judging by the improvement observed in the simulation of a similar strain from the PHB case-study (SCKK006). Since the up-regulation of the acetyl-CoA boosting genes was shown to be functional *in silico* for the PHB producing strain SCKK006, the problem with the simulation of strain HPY05 must be in the limited conversion of acetyl-CoA to malonyl-CoA. This theory is supported by the simulations obtained in strain HPY06, which showed that the combined action of the acetyl-CoA boosting strategy with the up-regulation of *ACC1* results in an improved 3-HP yield *in silico* when compared to each strategy on its own.

The yield reported for strain HPY06 improved 2.4 times when compared with the production yield of strain HPY01 (Table 4.4), which was quite low compared to the 15 times improvement predicted by TDPS. Such a considerable difference might be the consequence of a bottleneck in one of the up-regulated genes. As shown by Shi et al. the attenuation of the post-translational regulation of Acc1p resulted in a higher flux in this enzyme (HPY18) [26] and in an improvement of 1.8 times compared with a strain with the up-regulation of the wild-type *ACC1* (HPY04). Therefore, it is likely that the up-regulation of a modified *ACC1* gene in strain HPY06 would result in a bigger improvement in 3-HP production, which should reduce the difference between the experimental yield and the value simulated with TDPS. Although the shake-flask yield cannot be directly compared with the simulated values, its value is included within the interval simulated with TDPS.

Strain HPY09 was engineered for improved NADPH availability by introducing in *S. cerevisiae* a NADP⁺-dependent glyceraldehyde-3-phosphate dehydrogenase (*gapN*) from *Streptococcus mutans* [25]. The objective of expressing *gapN* was to enhance the activity of the *mcr* gene, which requires two molecules of NADPH per molecule of 3-HP produced. Although the 3-HP yield improved 1.3 times in strain HPY09 cultivated in shake-flasks, the TDPS simulation failed to predict the advantage of the increased availability of NADPH (Table 4.4). However, when the expression of *gapN* was combined with all the other genetic modifications, we observed a cooperative effect in the TDPS simulations (strain HPY11 in Table 4.4). The shake-flask yield for strain HPY11 improved 5 times over the strain HPY01, which was still considerably lower than the 89 fold average improvement predicted by TDPS. Even if we consider the more conservative estimation obtained with TDPS_FBA for the improvement in the minimum simulated value, the fold change would still be 15 times, which is still considerably higher than the values reported experimentally. Regarding the production yield predicted by TDPS for strain HPY11, the average value (116 mg/g) was significantly higher than the overall yield reported in shake-flask (23 mg/g). However, the minimum values shown for TDPS (14 mg/g) and TDPS_FBA (6.6 mg/g) represent only 61 % and 29 % of the total 3-HP yield reported in shake-flasks (23 mg/g). When compared to the value reported for the percentage of 3-HP produced during the glucose phase for strain HPY18 (41 %), these simulated yields can be considered a reasonable estimation for observed experimental yield

In general, the results obtained for the less engineered strains were quite satisfactory, and the production yields simulated with TDPS were very close to the values reported experimentally. However, for the more engineered strains there was a clear over-estimation of the average 3-HP yields by TDPS. As discussed above for strain HPY06, this might be the result of a bottleneck in the 3-HP production pathway, which is reducing the experimental yields in comparison with the average predictions made by TDPS. The acetyl-CoA carboxylase (*Acc1p*) has been reported before to be one of the bottlenecks [26], but other enzymes might also be working at limited capacity. The existence of a bottleneck is the most likely suspect for the divergence observed between the experimental results and the simulated values and it became more pronounced with the increase in the number of genetic modifications applied. Nevertheless, the minimum values simulated with TDPS and TDPS_FBA indicate that the introduction of enzymes

with limited capacity can be simulated with these algorithms by assuming more conservative C parameters during the simulations.

The simulation of the 3-HP productions yields in TDPS showed that this algorithm was successful in predicting the ranking between metabolically engineered strains. Among all the strains shown in Table 4.4, only two of them failed to be ranked correctly among the others (HPY05 and HPY09). Although the simulations of strains HPY05 and HPY09 did not show a direct improvement of the 3-HP production yield, when combined with additional genetic modifications, there was a clear positive effect of the underlying strategy. For example, in the specific case of the improvement of NADPH availability, it was shown that increasing this cofactor on its own (strain HPY09) was not relevant because the limiting substrate *in silico* was malonyl-CoA. However, when combined with a strategy to improve malonyl-CoA availability, the same NADPH improvement strategy had a positive impact on 3-HP production. This observation indicates that the strategy used in TDPS to model genetic modifications in a manner dependent on the limiting substrate is working as intended.

4.4.2.3 Case-study 3: L-malic acid production in *S. cerevisiae*

The production of L-malic acid (MA) in the cytosol of *S. cerevisiae* can be achieved from oxaloacetate and NADH using the malate dehydrogenase enzyme endogenous of this microorganism. Zelle et al. used a pyruvate decarboxylase negative (PDC⁻) *S. cerevisiae* as the platform strain to optimize L-malic acid production [27]. Since PDC⁻ *S. cerevisiae* cannot grow in the absence of C₂ compounds (ethanol, acetate, etc.), Zelle et al. used an evolved PDC⁻ yeast that could grow on minimal media with glucose as the single carbon source both in batch and glucose limited chemostats [28]. The adaptation behind the C₂-independent phenotype could not be determined but it has been shown separately that the up-regulation of the threonine aldolase gene (*GLY1*) can supply enough C₂ intermediates to cure the growth phenotype of a PDC⁻ strain in glucose limited chemostats [29].

We used the TDPS method to simulate the phenotype of a PDC⁻ strain in order to test how the growth rate would be affected in respiro-fermentative conditions with glucose as the single carbon source. The resulting phenotype was characterized by an 86 % reduction in the biomass yield on glucose with a considerable increase in the production of glycerol and pyruvate (data not shown). Interestingly, *in silico*, the lack of C₂ intermediates in the

cytosol was compensated by an increase in flux of the threonine aldolase (*GLY1*), which is in accordance with the work of van Maris et al. [29]. Since the authors in reference [27] used an evolved PDC⁻ strain with unknown physiological adaptations, there was not enough information available to calculate a proper reference flux distributions for the evolved strain. Therefore, we opted for simulating all the genetic modifications, including the PDC knock-outs, simultaneously. Although this was not the optimal scenario, it should still give an idea regarding the performance of the engineered strains.

Table 4.5 shows the MA production yields of the mutant strains constructed and characterized by Zelle et al. [27] in comparison with the respective simulations performed with TDPS. Regarding the single up-regulation strains (A-C), TDPS could not replicate the better performance reported for strains B and C in comparison to strain A. The simulations showed that up-regulating the L-malic acid transporter (*SpMAE1*) was more effective than up-regulating the pyruvate carboxylase (*PYC2*) or the malate Dehydrogenase gene (*MDH3*). *In silico*, the advantage of up-regulating the transporter in comparison with the other two genes lies in the connectivity of MA with other pathways in *S. cerevisiae*. While the up-regulation of *SpMAE1* results in the prompt transport of MA out of the cell, the up-regulation of *MDH3* just increases the abundance of this metabolite inside the cell, which can then enter other metabolic pathways without being excreted.

For the double up-regulation strains (D-F), the MA production ranking predicted by TDPS was also not coincident with the results obtained experimentally (Table 4.5). The experimental yields reported for strains D and F were very close to the values predicted by TDPS, but the performance of strain E was seriously underestimated by the algorithm. The failed prediction for strain E made it the worst performing mutant among the double up-regulation strains simulated with TDPS, which is exactly the opposite result obtained for strain E experimentally. *In silico*, the strains sharing the up-regulation of *PYC2* showed the best MA yields (strains D and F), which indicates that *PYC2* up-regulation is quite important for MA production. However, the experimental results reported for strain E seem to indicate that *PYC2* up-regulation is the least important of the modifications tested. These results suggest that oxaloacetate is quite limiting *in silico* when compared with the experimental results reported for the evolved PDC⁻ strain. This is also corroborated by the experimental yields reported for the single up-regulation strains (A-C), which showed that *PYC2* up-regulation had the least impact on MA production,

indicating that oxaloacetate availability is not the most limiting factor *in vivo* for the evolved PDC⁻ strain. The experimental yield reported for strain E (173 mg/g) was not inside the intervals simulated with TDPS (14 – 53 mg/g) or TDPS_FBA (38 – 140 mg/g), but these results would probably improve if a reference distribution specific for the evolved PDC⁻ strain would be used in the simulations.

When all the three up-regulations were combined into a single strain it was possible to verify that both experimental and simulated production yields for MA ranked first among all strains shown in Table 4.5. The experimental yield reported for strain G (312 mg/g) was above the simulated interval predicted by TDPS (94 – 190 mg/g) but it was included in the relaxed interval obtained with TDPS_FBA (55 – 530 mg/g). The big difference in the yields predicted by TDPS and TDPS_FBA is a direct consequence of the severe growth phenotype caused by deletion of the PDC genes, which caused harsh changes in the flux distribution of the mutant strain. Given the large flux readjustment, allowing the objective function of TDPS to relax by 10 % in this strain meant that a significant portion of the metabolic pathways could readapt, causing the simulation outcome to change pronouncedly between simulation methods.

The inferior performance of TDPS in this case-study was most likely caused by the severe flux phenotype triggered by the deletion of all PDC genes. While the flux changes predicted by TDPS were based solely on the objective function included in TDPS, the PDC⁻ strain used in the experiments was evolved in the lab with the goal of improving its growth phenotype and autotrophy for C₂ compounds. Therefore, other genetic adaptations were probably acquired by the PDC⁻ mutant during the laboratory evolution experiments. These unknown adaptations were not included in the simulations, and consequently the results obtained here could have been affected. This issue can be ameliorated by using a reference flux distribution that better represents the evolved PDC⁻ strain used as the platform for building the MA overproducing strains analyzed in this case-study. However, in order to do that we would need more information concerning the fluxes of the evolved PDC⁻ strain.

Table 4.5- TDPS validation case-study using L-malic acid (MA) production in *S. cerevisiae*

Strain	Genotype relevant for simulations	<i>In vivo</i> yields (mg MA/g glucose)	TDPS yields (mg MA/g glucose)
A [27]	Gene up-regulations: <i>PYC2</i> (EC 6.4.1.1) Gene deletions: <i>PDC1, PDC2, PDC6</i> (EC 4.1.1.1)	Shake Flask - Glucose phase: 13.6 mg/g	Batch simulation: TDPS: 0 mg/g Min: 0 mg/g Max: 0 mg/g TDPS_FBA: 2.8 mg/g Min: 0 mg/g Max: 43 mg/g
B [27]	Gene up-regulations: <i>MDH3</i> (EC 1.1.1.37) Gene deletions: <i>PDC1, PDC2, PDC6</i> (EC 4.1.1.1)	Shake Flask - Glucose phase: 36 mg/g	Batch simulation: TDPS: 0 mg/g Min: 0 mg/g Max: 0 mg/g TDPS_FBA: 0 mg/g Min: 0 mg/g Max: 0 mg/g
C [27]	Gene up-regulations: <i>SpMAE1</i> (malate transporter) Deletion: <i>PDC1, PDC2, PDC6</i> (EC 4.1.1.1)	Shake Flask - Glucose phase: 34 mg/g	Batch simulation: TDPS: 11 mg/g Min: 7.8 mg/g Max: 12 mg/g TDPS_FBA: 54 mg/g Min: 27 mg/g Max: 71 mg/g
D [27]	Gene up-regulations: <i>PYC2</i> (EC 6.4.1.1), <i>MDH3</i> (EC 1.1.1.37) Gene deletions: <i>PDC1, PDC2, PDC6</i> (EC 4.1.1.1)	Shake Flask - Glucose phase: 53 mg/g	Batch simulation: TDPS: 42 mg/g Min: 0 mg/g Max: 100 mg/g TDPS_FBA: 44 mg/g Min: 0 mg/g Max: 160 mg/g
E [27]	Gene up-regulations: <i>MDH3</i> (EC 1.1.1.37), <i>SpMAE1</i> (malate transporter) Gene deletions: <i>PDC1, PDC2, PDC6</i> (EC 4.1.1.1)	Shake Flask - Glucose phase: 173 mg/g	Batch simulation: TDPS: 36 mg/g Min: 14 mg/g Max: 53 mg/g TDPS_FBA: 90 mg/g Min: 38 mg/g Max: 140 mg/g
F [27]	Gene up-regulations: <i>PYC2</i> (EC 6.4.1.1), <i>SpMAE1</i> (malate transporter) Gene deletions: <i>PDC1, PDC2, PDC6</i> (EC 4.1.1.1)	Shake Flask - Glucose phase: 56 mg/g	Batch simulation: TDPS: 44 mg/g Min: 5.9 mg/g Max: 87 mg/g TDPS_FBA: 107 mg/g Min: 34 mg/g Max: 240 mg/g
G [27]	Gene up-regulations: <i>PYC2</i> (EC 6.4.1.1), <i>MDH3</i> (EC 1.1.1.37), <i>SpMAE1</i> (malate transporter) Gene deletions: <i>PDC1, PDC2, PDC6</i> (EC 4.1.1.1)	Shake Flask - Glucose phase: 312 mg/g	Batch simulation: TDPS: 130 mg/g Min: 94 mg/g Max: 190 mg/g TDPS_FBA: 413 mg/g Min: 55 mg/g Max: 530 mg/g

Regardless of the issues observed for the individual strains, TDPS was still capable of predicting the better overall performance of the double up-regulation strains (D-F) in comparison to the ones with a single up-regulation (A-C). Furthermore, the best performing strain was also correctly predicted by the simulation algorithm and its production yield was within the interval simulated with TDPS_FBA.

4.4.2.4 Case-study 4: santalene production in *S. cerevisiae*

Santalene is an isoprenoid that belongs to the sesquiterpene class (15 carbon atoms) and can be synthesized from farnesyl pyrophosphate (FPP), which is a precursor for sterol biosynthesis in *S. cerevisiae*. Scalcinati et al. introduced in *S. cerevisiae* the santalene synthase (*SanSyn*) from *Clausena lansium* in order to characterize metabolic engineered strains with improved production of sesquiterpenes [30]. The strain SCIGS28 was built by introducing a codon optimized *SanSyn* into *S. cerevisiae* together with the up-regulation of a truncated version of HMG-CoA reductase (*tHMG1*) that lacks feedback regulation [30]. Further optimization of santalene production was accomplished in strain SCIGS29 by down-regulating squalene synthase (*ERG9*) and deleting the lipid phosphate phosphatase (*LPPI*) [30]. Since *LPPI* deletion addressed the unspecific hydrolysis of FPP by the respective enzyme and this metabolic activity was not present in the Yeast 6 model, this modification was not included in the simulations performed.

The santalene production yields reported for the sesquiterpene producing strains are shown in Table 4.6 along with the simulation values obtained with TDPS. The average simulated yield obtained for strain SCIGS28 (1.5 mg/g) was almost triple the value reported experimentally (0.57 mg/g), but the interval of simulated values (0.52 – 2.5 mg/g) still encompasses the experimental yield. The fact that the experimental yield reported for strain SCIGS28 is closer to the lower simulated value can indicate that the *SanSyn* enzyme used for santalene production has poor kinetic properties. Alternatively, the *in silico* availability of FPP might have been overestimated, which would also result in the differences observed between the average values.

Table 4.6- TDPS validation case-study using santalene production in *S. cerevisiae*

Strain	Genotype relevant for simulations	<i>In vivo</i> yields (mg santalene/g glucose)	TDPS yields (mg santalene/g glucose)
SCIGS28 [30]	Gene additions: <i>SanSyn</i> (EC 4.2.3.50) Gene up-regulations: <i>tHMG1</i> (EC 1.1.1.34)	Chemostat ($D = 0.1 \text{ h}^{-1}$) Steady state yield: 0.57 mg/g	Chemostat simulation: TDPS: 1.5 mg/g Min: 0.52 mg/g Max: 2.5 mg/g TDPS_FBA: 1.5 mg/g Min: 0.52 mg/g Max: 2.5 mg/g
SCIGS29 [30]	Gene additions: <i>SanSyn</i> (EC 4.2.3.50) Gene up-regulations: <i>tHMG1</i> (EC 1.1.1.34) Gene down-regulations: <i>ERG9</i> (EC 2.5.1.21)	Chemostat ($D = 0.1 \text{ h}^{-1}$) Steady state yield: 1.58 mg/g	Chemostat simulation: TDPS: 1.1 mg/g Min: 0 mg/g Max: 2.8 mg/g TDPS_FBA: 1.8 mg/g Min: 0.41 mg/g Max: 3.8 mg/g

Regarding the performance of the strain SCIGS29, the simulations performed using TDPS failed to predict the reported increase of 2.8 times in the production of santalene (Table 4.6). The yield simulated with TDPS (1.1 mg/g) actually decreased in comparison with strain SCIGS28 (1.5 mg/g). This drop in santalene production was caused by a severe decrease in biomass formation triggered by the down-regulation of *ERG9*. Since ergosterol is an essential growth component of *S. cerevisiae*, the down-regulation of *ERG9* can affect the *in silico* growth dramatically. Although we expected that the down-regulation of *ERG9* would be compensated *in silico* by an increase in the flux towards FPP, in reality this was not observed in the simulations performed with TDPS. On the other hand, the secondary growth objective function used in TDPS_FBA was able to compensate partly the drop in biomass yield by increasing the production turnover of FPP. As a consequence, the simulations obtained with TDPS_FBA resulted in a 20 % increase in the average santalene yield (1.8 mg/g) along with a 36 % increase in the maximum value simulated (3.8 mg/g). Although these values are still below the improvement observed experimentally, the experimental yield (1.58 mg/g) is within the simulated interval (0.41 mg/g – 3.8 mg/g).

4.4.2.5 Case-study 5: vanillin β -D-glucoside production in *S. cerevisiae*

Vanillin β -D-glucoside (VG) is a flavoring compound of industrial importance that can be biosynthesized from 3-dehydroshikimate (an intermediate of aromatic amino acid synthesis). Brochado et al. used a *S. cerevisiae* strain containing four heterologous genes

necessary to produce VG in this microbe: 3-dehydroshikimate dehydratase (*3DSD*), aryl carboxylic acid reductase (*ACAR*), O-methyltransferase (*hsOMT*) and UDP-glycosyltransferase (*UGT*) [31]. In addition to these four gene additions, the basal VG producing strain (VG0) has two additional deletions: the gene coding for a NADPH-dependent medium chain alcohol dehydrogenase (*ADH6*) and the gene coding for exo-1,3-beta-glucanase (*EXG1*). Both deletions were implemented to address unspecific degradation of vanillin and VG by the mentioned enzymes. These two deletions were not included in the simulations performed with TDPS because the unspecific activities of Adh6p and Exg1p were not present in the Yeast 6 model.

Table 4.7 shows the VG production yields reported by Brochado et al. [31] in comparison with the values simulated with TDPS. For strain VG0, the average VG yield obtained with TDPS (1.6 mg/g) was 60 % lower than the value reported experimentally (4.0 mg/g). The difference decreased to 18 % if the experimental value was compared to the maximum value simulated with TDPS (3.3 mg/g). Furthermore, the difference was less pronounced in the simulations performed with TDPS_FBA and the maximum value simulated with this method for strain VG0 (3.8 mg/g) was only 5% lower than the experimentally determined value. Although the average yields obtained in the simulations were underestimated, the maximum simulated yields were reasonably good approximations to the experimental value reported for strain VG0.

A closer analysis of the additional metabolites excreted *in silico* by the strain VG0 revealed that one of the intermediaries in the production of VG, protocatechuic aldehyde (PAL), was being produced with a yield of 82 mg/g (data not shown). Experimentally, PAL was also produced by VG0 with a yield of 7 mg/g, which indicates that the conversion of PAL into vanillin is a limiting step. This step is catalyzed by the enzyme O-methyltransferase and requires S-Adenosyl methionine (SAM) as the methyl group donor. In the TDPS simulations performed for strain VG0, we observed that SAM was the limiting precursor in the production of VG. The limited availability of this metabolite resulted in less than 1% of the total PAL molecules being converted to VG *in silico*. Judging by the accumulation of PAL reported for strain VG0 in chemostat experiments, the limitation of SAM might also be a contributing factor. Although other explanations for the PAL accumulation cannot be ruled out, these results showed the potential of the TDPS algorithm in the prediction of limited precursor availability using stoichiometric models.

Table 4.7- TDPS validation case-study using vanillin-glucoside (VG) production in *S. cerevisiae*

Strain	Genotype relevant for simulations	<i>In vivo</i> yields (mg VG/g glucose)	TDPS yields (mg VG/g glucose)
VG0 [31]	Gene additions: <i>3DSD</i> (EC 4.2.1.118), <i>ACAR</i> (EC 1.2.1.30), <i>hsOMT</i> (EC 2.1.1.-), <i>UGT</i> (EC 2.4.1.136)	Chemostat ($D = 0.1 \text{ h}^{-1}$) Steady state yield: 4.0 mg/g	Chemostat simulation: TDPS: 1.6 mg/g Min: 0.42 mg/g Max: 3.3 mg/g TDPS_FBA: 2.2 mg/g Min: 0.74 mg/g Max: 3.8 mg/g
VG2 [31]	Gene additions: <i>3DSD</i> (EC 4.2.1.118), <i>ACAR</i> (EC 1.2.1.30), <i>hsOMT</i> (EC 2.1.1.-), <i>UGT</i> (EC 2.4.1.136) Gene deletions: <i>PDC1</i> (EC 4.1.1.1)	Chemostat ($D = 0.1 \text{ h}^{-1}$) Steady state yield: 5.6 mg/g	Chemostat simulation: TDPS: 1.6 mg/g Min: 0.45 mg/g Max: 3.5 mg/g TDPS_FBA: 2.2 mg/g Min: 0.73 mg/g Max: 3.8 mg/g
VG4 [31]	Gene additions: <i>3DSD</i> (EC 4.2.1.118), <i>ACAR</i> (EC 1.2.1.30), <i>hsOMT</i> (EC 2.1.1.-), <i>UGT</i> (EC 2.4.1.136), <i>GDH2</i> (EC 1.4.1.2) Gene deletions: <i>PDC1</i> (EC 4.1.1.1), <i>GDH1</i> (EC 1.4.1.4)	Chemostat ($D = 0.1 \text{ h}^{-1}$) Steady state yield: 4.2 mg/g	Chemostat simulation: TDPS: 1.5 mg/g Min: 0.39 mg/g Max: 3.3 mg/g TDPS_FBA: 2.2 mg/g Min: 0.74 mg/g Max: 3.9 mg/g

To improve the performance of strain VG0, the gene coding for pyruvate decarboxylase 1 (*PDC1*) was deleted by the authors in [31], originating strain VG2. *PDC1* was chosen as a target by the authors because the strain VG0 produced significant amounts of ethanol, which was decreasing the yields of biomass and product formation. The ethanol production observed was justified by Brochado et al. as the consequence of a decrease in the respiratory capacity of VG0, possibly caused by the toxicity of one or several intermediates present in the VG biosynthetic pathway [31]. Since the constraint-based modeling approach used in TDPS does not take into account any toxicity phenomena, no ethanol production was observed in the simulations performed for VG0. Therefore, we did not expect that TDPS would predict correctly the outcome of any genetic modifications targeted at solving phenotypes induced by the toxicity of the intermediates. As anticipated, the experimental results revealed a 40 % improvement in the VG production yield experimentally, which we could not replicate in the simulations performed with TDPS (Table 4.7).

The final strain analyzed in this case-study (VG4) included the replacement of the canonical NADPH-dependent glutamate dehydrogenase (*GDH1*) with a NADH

dependent variant (*GDH2*). The switch of the cofactor specificity in this reaction, which is responsible for ammonia assimilation, was intended to increase the availability of NADPH in the cytosol, so that more flux could flow in the ACAR reaction. However, the resulting strain showed no improvement both *in silico* and experimentally when compared with VG0 (Table 4.7). In fact, when compared with VG2, the strain VG4 showed a decrease in the production yield of VG in chemostat cultures. As discussed above the production of VG is limited *in silico* by the availability of SAM and no further improvements should be observed unless this bottleneck is addressed.

4.5 Conclusions and future perspectives

Constraint-based modelling methodologies can be of great help during the laborious trial and error strain design procedure by suggesting promising metabolic engineering targets. However, most *in silico* strain design methods were developed with an optimization oriented goal and are not equipped for simulating quantitatively the flux changes in complex mutant strains. One of biggest drawbacks present in most of the methods currently available is that they force the flux values in activated/heterologous reactions without considering the availability of resources.

TDPS was developed with the goal of simulating the majority of the genetic modifications usually implemented in engineered strains. Besides gene deletions and down-regulations, TDPS can also simulate the up-regulation of metabolic reactions as well as the introduction of heterologous genes or the activation of “dormant” reactions. The methodology implemented inside TDPS takes into account the availability of resources in the network by assuming that the production turnover of a metabolite can be used as an indication of its abundance. Consequently, the flux values are never forced to pre-calculated levels, allowing the objective function to work in harmony with the genetic constraint in order to predict the flux rearrangements in mutant strains.

To our knowledge this was the first time that a constraint-based method was validated quantitatively using metabolically engineered *S. cerevisiae* strains available in the literature by comparing the simulated and experimental production yields of the target metabolite. Overall, the estimated yields obtained with TDPS were quite close to the experimentally reported values, although the performance oscillated among the different

case-studies tested. The fact that most experimental results available in the literature were obtained in batch cultures (PHB, 3-HP and L-malic acid) complicated the comparison of the simulated steady-state values with the experimental yields. Furthermore, some simulation inconsistencies were also observed in several cases: when a strain with an uncertain genetic background was used (L-malic acid), when there were flux changes resulting from product toxicity (vanillin β -D-glucoside) and when some of the modifications had a large impact on biomass formation (L-malic acid and santalene). The results also showed that TDPS might be helpful in finding metabolic bottlenecks (namely the ones caused by limitations in the availability of precursors from other pathways), but further experiments would be required to confirm such findings. Although this work was not focused on the search for metabolic engineering targets, TDPS could also be used for this purpose when combined with strain optimization algorithms.

Overall, the results obtained are very promising but could eventually be improved by using an objective function that better describes the cellular adaptation when faced with serious disturbances. In laboratory conditions, mutant microorganisms are continuously being selected for better growth, first in media plates and then on each cultivation step until the final characterization experiment. Judging by the better performance of the hybrid objective function tested (TDPS_FBA) in some case-studies, we believe that if we used an objective function that incorporates a term to promote better growth rates, the simulation results would improve. It would also be interesting to expand the case-studies used to validate TDPS to other organisms in order to increase the number of strains characterized in chemostats, which could help fine-tuning the performance of the algorithm.

References

1. Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. *Nat Rev Microbiol* 10:291–305.
2. Bordbar A, Monk JM, King ZA, Palsson BO (2014) Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet* 15:107–20.
3. Pharkya P, Maranas CD (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab Eng* 8:1–13.

4. Ranganathan S, Suthers PF, Maranas CD (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol* 6:e1000744.
5. Yang L, Cluett WR, Mahadevan R (2011) EMILiO: a fast algorithm for genome-scale strain design. *Metab Eng* 13:272–81.
6. Cotten C, Reed JL (2013) Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering. *Biotechnol J* 8:595–604.
7. Gonçalves E, Pereira R, Rocha I, Rocha M (2012) Optimization approaches for the *in silico* discovery of optimal targets for gene over/underexpression. *J Comput Biol* 19:102–14.
8. Xu P, Ranganathan S, Fowler ZL, et al. (2011) Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab Eng* 13:578–87.
9. Rockwell G, Guido NJ, Church GM (2013) Redirector: designing cell factories by reconstructing the metabolic objective. *PLoS Comput Biol* 9:e1002882.
10. Ip K, Donoghue N, Kim MK, Lun DS (2014) Constraint-based modeling of heterologous pathways: Application and experimental demonstration for overproduction of fatty acids in *Escherichia coli*. *Biotechnol Bioeng* 111:2056–66.
11. Rocha I, Maia P, Evangelista P, et al. (2010) OptFlux: an open-source software platform for *in silico* metabolic engineering. *BMC Syst Biol* 4:45.
12. Fell DA, Small JR (1986) Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem J* 238:781–6.
13. Varma A, Palsson BO (1994) Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/Technology* 12:994–998.
14. Lewis NE, Hixson KK, Conrad TM, et al. (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* 6:390.
15. Segrè D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A* 99:15112–7.
16. Becker SA, Feist AM, Mo ML, et al. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* 2:727–38.
17. Shlomi T, Berkman O, Ruppin E (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A* 102:7695–700.

18. Brochado AR, Andrejev S, Maranas CD, Patil KR (2012) Impact of stoichiometry representation on simulation of genotype-phenotype relationships in metabolic networks. *PLoS Comput Biol* 8:e1002758.
19. Stephanopoulos G, Vallino J (1991) Network rigidity and metabolic engineering in metabolite overproduction. *Science* (80-) 252:1675–1681.
20. Chung BKS, Lee D-Y (2009) Flux-sum analysis: a metabolite-centric approach for understanding the metabolic network. *BMC Syst Biol* 3:117.
21. Heavner BD, Smallbone K, Price ND, Walker LP (2013) Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database* (Oxford) 2013:bat059.
22. Van Dijken J., Bauer J, Brambilla L, et al. (2000) An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb Technol* 26:706–714.
23. Kocharin K, Chen Y, Siewers V, Nielsen J (2012) Engineering of acetyl-CoA metabolism for the improved production of polyhydroxybutyrate in *Saccharomyces cerevisiae*. *AMB Express* 2:52.
24. Kocharin K, Nielsen J (2013) Specific growth rate and substrate dependent polyhydroxybutyrate production in *Saccharomyces cerevisiae*. *AMB Express* 3:18.
25. Chen Y, Bao J, Kim I-K, et al. (2014) Coupled incremental precursor and co-factor supply improves 3-hydroxypropionic acid production in *Saccharomyces cerevisiae*. *Metab Eng* 22:104–9.
26. Shi S, Chen Y, Siewers V, Nielsen J (2014) Improving production of malonyl coenzyme A-derived metabolites by abolishing Snf1-dependent regulation of Acc1. *MBio* 5:e01130–14.
27. Zelle RM, de Hulster E, van Winden WA, et al. (2008) Malic acid production by *Saccharomyces cerevisiae*: engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. *Appl Environ Microbiol* 74:2766–77.
28. Van Maris AJA, Geertman J-MA, Vermeulen A, et al. (2004) Directed Evolution of Pyruvate Decarboxylase-Negative *Saccharomyces cerevisiae*, Yielding a C2-Independent, Glucose-Tolerant, and Pyruvate-Hyperproducing Yeast. *Appl Environ Microbiol* 70:159–166.
29. Van Maris AJA, Luttik MAH, Winkler AA, et al. (2003) Overproduction of threonine aldolase circumvents the biosynthetic role of pyruvate decarboxylase in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 69:2094–9.
30. Scalcinati G, Partow S, Siewers V, et al. (2012) Combined metabolic engineering of precursor and co-factor supply to increase alpha-santalene production by *Saccharomyces cerevisiae*. *Microb Cell Fact* 11:117.

31. Brochado AR, Matos C, Møller BL, et al. (2010) Improved vanillin production in baker's yeast through *in silico* design. *Microb Cell Fact* 9:84.

CHAPTER 5

Optimization of the cytosolic acetyl-CoA pool in *S. cerevisiae* using *in silico* driven metabolic engineering

Acetyl-coenzyme A (acetyl-CoA) is the biosynthetic precursor of many important metabolites for the chemical industry (e.g. biopolymers, isoprenoids and lipids). The production yield of acetyl-CoA derived products in *S. cerevisiae* is usually limited by the low levels of this metabolite in the cytosol and, understandably, many metabolic engineering strategies have tried to address this limitation. However, the creation of a cell factory engineered for optimal acetyl-CoA availability has been hampered by the complexity of the native pathways present in *S. cerevisiae*.

Genome-scale metabolic modeling offers a whole cell perspective of metabolism and may be used to simulate flux changes in response to perturbations in the network. In order to optimize the flux to acetyl-CoA in *S. cerevisiae*, the turnover dependent phenotypic simulation method (chapter 4) was used in combination with an evolutionary algorithm to search for sets of genetic alterations that improve the production yield of 3-hydroxypropionic acid (3-HP) derived from acetyl-CoA. The *in silico* suggested modifications were implemented sequentially in *S. cerevisiae*, and the resulting strains were physiologically characterized in batch fermentations to test the validity of the simulations. Although the yields obtained experimentally were considerably lower than the simulations suggested, a positive effect on the 3-HP yield was observed for the down-regulation of the pyruvate dehydrogenase complex and the deletion of *ACH1* (succinyl-CoA:acetate CoA-transferase).

The information presented in this Chapter is being prepared for submission to a peer reviewed journal:

Pereira R., Chen Y., Rocha I., Siewers V., Nielsen J., Optimization of the cytosolic acetyl-CoA pool in *S. cerevisiae* using *in silico* driven metabolic engineering.

5.1 Introduction

Acetyl-coenzyme A (acetyl-CoA) is a central metabolic node in most living organisms and it is the precursor for many industrially important bio-based products [1]. In *S. cerevisiae* the two major sources of acetyl-CoA are the pyruvate dehydrogenase complex (PDHc) in the mitochondria and acetyl-CoA synthetase (ACS) in the cytosol and peroxisome [2]. One additional mitochondrial source has been reported, Ach1p, but it appears to be solely important for growth on acetate [3]. Ach1p was initially reported to be an acetyl-CoA hydrolase [4] but additional characterization revealed that its major role is to transfer the coenzyme A moiety from succinyl-CoA to acetate in the mitochondria of *S. cerevisiae* [3].

In the mitochondria, acetyl-CoA produced by the PDHc has a major catabolic role under non-fermentative conditions because it fuels the citric acid cycle, which contributes substantially to ATP production in the cell. Additionally, mitochondrial acetyl-CoA is also necessary for the anabolic synthesis of cellular components such as L-leucine and 2-oxoglutarate derived amino-acids [5]. In the cytosol, ACS is the exclusive source of acetyl-CoA and plays an essential anabolic role [6] in the synthesis of fatty acids and L-lysine [5, 7]. Furthermore, cytosolic acetyl-CoA is also the precursor of the mevalonate pathway, which is responsible for the synthesis of important cellular components such as isoprenoids and sterols.

When *S. cerevisiae* grows on ethanol or other C₂ carbon sources, the availability of pyruvate is very limited and PDHc cannot fulfil the need of acetyl-CoA in the mitochondria. In this case, the cytosolic/peroxisomal acetyl-CoA synthetase becomes the central supply of acetyl-CoA for the cell [2]. However, acetyl-CoA cannot be directly transported between compartments in the absence of carnitine [8] and the glyoxylate cycle is required to synthesize C₄ and C₆ intermediates that can be transported to the mitochondria and enter the citric acid cycle [2].

To improve the availability of acetyl-CoA in the cytosol of *S. cerevisiae*, several strategies have been implemented over the years. In the work of Shiba et al. [9], the authors up-regulated the PDH bypass by over-expressing the aldehyde dehydrogenase gene *ALD6* and introducing a deregulated ACS from *Salmonella enterica* [10] in *S. cerevisiae*. The result was an increase in the synthesis of acetyl-CoA in the cytosol that

boosted the production of amorphaadiene and mevalonate. Similar strategies were implemented for the improvement of polyhydroxybutyrate [11], α -santalene [12], 1-butanol [13] and 3-hydroxypropionic acid (3-HP) production [14]. In parallel with the PDH bypass amplification, it has also been shown that deleting the peroxisomal citrate synthase (*CIT2*) or the malate synthase (*MLS1*), which participate in the glyoxylate cycle, has a positive impact on the yield of acetyl-CoA derived products [12–14].

Different enzymes have also been tested to improve the synthesis of acetyl-CoA in the cytosol and offer alternative ways to optimize its production. Kozak et al. showed that an acetylating acetaldehyde dehydrogenase (A-ALD) can replace the joint activity of the native aldehyde dehydrogenase (ALD) and ACS in *S. cerevisiae* [15]. The advantage of this system is that while the sequential activity of ALD and ACS consumes 2 ATP equivalents, A-ALD has no ATP requirements and can increase the maximum theoretical yield of acetyl-CoA derived products. In the same work, it was also shown that pyruvate formate lyase (PFL) can replace the activity of all the enzymes in the PDH bypass of *S. cerevisiae* under anaerobic conditions [15]. Additionally it has also been demonstrated that it is possible to express a functional PDH complex in the cytosol of *S. cerevisiae* derived from either *Enterococcus faecalis* [16], *Escherichia coli* or *S. cerevisiae* [17]. Furthermore, additional heterologous pathways have also been tested in *S. cerevisiae* to produce cytosolic acetyl-CoA: ATP-citrate lyase can produce acetyl-CoA and oxaloacetate from citrate [17] and the coordinated activity of phosphoketolase and phosphotransacetylase can synthesize acetyl-CoA from D-xylulose-5-phosphate [18, 19].

Despite all the efforts to optimize acetyl-CoA production in the cytosol of *S. cerevisiae*, the product yields obtained are still quite far from the theoretical maximum. Therefore, there is a significant room to improve acetyl-CoA cell factories and to clarify additional aspects of acetyl-CoA metabolism in *S. cerevisiae*.

The application of Genome Scale Metabolic Models (GSMMs) to strain engineering is of industrial importance because it promises to be a cost efficient methodology to build and improve microbial cell factories [20, 21]. They have been applied successfully in the past for strain engineering of *S. cerevisiae*, for instance in the improvement of sesquiterpene [22] and vanillin production [23]. Given the laborious and expensive nature of strain engineering, it is important that simulation results have a high success rate when implemented experimentally. In order to maximize the chances of success of an *in silico*

strain engineering project, special attention should be paid to the quality of the GSMM used and the accuracy of the simulation method applied. Nevertheless, even failed strain engineering strategies can still be useful to offer new insights into the cell's physiology and can offer good case studies to improve GSMMs and simulation algorithms.

Here we tested the application of the Turnover Dependent Phenotypic Simulation (TDPS) method (chapter 4) in the optimization of the acetyl-CoA supply in the cytosol of *S. cerevisiae*. The model chosen, Yeast 6 [24], was manually curated in chapter 3 to improve the accuracy of central carbon fluxes. In order to assess the availability of acetyl-CoA we used 3-hydroxypropionic acid (3-HP) production as a case study. The production of this compound has been shown to benefit from an increase in acetyl-CoA availability and given its extracellular nature it is easy to quantify [14]. Furthermore, the recent development of a deregulated Acc1p makes the conversion of acetyl-CoA to malonyl-CoA (direct precursor of 3-HP) more efficient and independent of the glucose concentration in the medium [25]. To test the validity of the results obtained from the TDPS simulations, we implemented three different *in silico* strategies to improve acetyl-CoA availability. The strains were then characterized in bioreactors to compare their behavior with the predictions from TDPS.

5.2 Materials and methods

5.2.1 Model and software

The GSMM Yeast 6.06 [24] was downloaded in SBML from the project's website: <http://sourceforge.net/projects/yeast/files/> and modified according to Appendix B (Table B1). The model was imported into OptFlux 3.07 [26] and the *in silico* environmental conditions were set to mimic minimal growth media supplemented with glucose under fully aerobic conditions (ammonia: unconstrained uptake, phosphate: unconstrained uptake, sulfate: unconstrained uptake, oxygen: unconstrained uptake, glucose: 1.15 mmol/(gCDW·h)). All simulations were run within OptFlux 3.07 [26] using IBM ILOG CPLEX Optimization Studio (Academic) as the linear and MILP programming solver. The "NumericalEmphasis" parameter in CPLEX was activated in order to solve the occasional numerical instability issues observed in MILP problems.

5.2.2 Optimizations of 3-HP production using TDPS

The turnover dependent phenotypic simulation (chapter 4) was used to simulate the mutant phenotypes, while the evolutionary algorithm included in OptFlux was in charge of the strain optimization layer [27]. pFBA [28] was used for the calculation of the reference (wild-type) flux distributions, with the maximization of biomass production as the objective function. The cofactor constraints applied in chapter 3 were also used here to improve the quality of the reference flux distribution in regard to NADPH and NADH metabolism. The Yeast 6 model was modified according to Appendix B (Table B6) to account for the heterologous malonyl-CoA reductase activity and 3-HP excretion.

The total number of possible strain engineering targets was reduced by removing: reactions without associated genes, dead ends and other reactions that cannot be active under growth on glucose, transporting reactions and others included in pathways of little relevance for 3-HP production. The objective function used in the evolutionary algorithm was the Biomass Product Coupled Yield (BPCY) [27] and the C parameter (see chapter 4) could assume any value in the set $\{0, 0.25, 0.5, 2, 5\}$, which represent two levels of up-regulation (2x and 5x), two levels of down-regulation (0.5x and 0.25x) and the inactivation is represented by zero. The reaction activation penalty constant was set to 50 and the evolutionary algorithm was run five times, setting the number of solution evaluations to 50,000 and the maximum number of strain modifications allowed to 10.

5.2.3 Flux distribution under growth on ethanol

The same environmental conditions mentioned in 5.2.1 were used here with the exception that glucose was replaced by ethanol as the sole carbon source (ethanol specific consumption rate: 3.68 mmol/(gCDW·h) [29]). Two modifications had to be implemented to make the glyoxylate cycle functional in the yeast 6 model as described in Appendix B (Table B7). For the calculation of the flux distributions pFBA [28] was used, with the maximization of biomass production as the objective function.

5.2.4 PCR and DNA handling

Oligonucleotides were purchased from Sigma-Aldrich (St. Louis, USA) and a *PDC1* gene fragment (see below) was purchased from Integrated DNA Technologies (Leuven, Belgium). High fidelity PCR was performed with Phusion polymerase from Thermo

Scientific (Waltham, USA), fusion PCR was performed with PrimeSTAR HS DNA polymerase from Takara Bio Europe (Otsu, Japan) and colony PCR was performed with DreamTaq DNA polymerase from Thermo Scientific. *E. coli* plasmid extraction, PCR product purification, and DNA gel extraction were performed with the respective GeneJET kits from Thermo Scientific. *S. cerevisiae* plasmid extraction was performed with Zymoprep Yeast Plasmid Miniprep II from Zymo Research (Irvine, USA). All restriction enzymes were obtained in fast-digestion format from Thermo Scientific. PCR reactions were performed in a S1000 Thermal Cycler from Bio-Rad (Hercules, USA) using the protocols recommended by the manufacturers for each polymerase. Yeast genomic DNA for colony PCR and gene cloning was prepared as described in [30].

E. coli DH5 α was used for plasmid isolation and maintenance using the competence and transformation procedures developed by Inoue et al. [31] as described in [32]. *S. cerevisiae* transformation was performed using the lithium acetate method as described by Gietz and Woods in [33]. Yeast recombinational cloning [34, 35] procedures used for plasmid construction were performed as described by Bessa et al. [36].

5.2.5 Media used and strain handling

Selection and maintenance in *E. coli* of the plasmids constructed/used in this work was performed in LB medium containing 10 g/L of peptone, 10 g/L of NaCl, 5 g/L of yeast extract and supplemented with 100 mg/L of ampicillin sodium salt. The solid version of this medium also included 16 g/L of agar. All cultivations of *E. coli* were performed at 37 °C and 200 rpm agitation (for liquid cultures).

Recombinant yeast strains were selected for HIS⁺ and URA⁺ phenotypes on Synthetic Dextrose (SD) medium containing 6.7 g/L of yeast nitrogen base without amino acids (ForMedium, Norfolk, United Kingdom), 0.77 g/L of complete supplement mixture (CSM lacking uracil or histidine) (ForMedium) and 20 g/L of glucose. Simultaneous selection of HIS⁺ and URA⁺ phenotypes was done in the same medium but containing 0.75 g/L of complete supplement mixture lacking uracil and histidine (ForMedium). For KanMX marker selection the cells were cultivated in YPD medium, containing 10 g/L yeast extract, 20 g/L of peptone, 20 g/L of glucose and supplemented with 200 mg/L G418 disulphate (ForMedium). To loop out the KanMX marker the cells were cultivated in YPGal medium which contained 20 g/L of galactose instead of glucose. To make SD,

YPGal or YPD plates, 20 g/L of agar was included in the medium formulation. All cultivations of *S. cerevisiae* were performed at 30 °C with 200 rpm agitation in the case of liquid cultures.

Strains were preserved for long-term storage after overnight cultivation in selective media by adding glycerol to a final concentration of 15 % and keeping them in a -80 °C freezer.

5.2.6 Plasmid construction

All plasmids used and constructed in this study are listed in Table 5.1, as well as their main features and source.

$ACCI^{(ser659ala, ser1157ala)}$ was obtained from Shi et al. [25] and cloned into the *NotI* and *SacI* sites of the pYC1 plasmid [14]. The resulting plasmid pYC5 contains the $ACCI^{(ser659ala, ser1157ala)}$ gene controlled by the *PGK1* promoter and a codon-optimized version of the malonyl-CoA reductase gene (*mcr*) from *Chloroflexus aurantiacus* [14] under the control of the *TEF1* promoter.

Table 5.1- Plasmids used in this work

Plasmid	Features	Source
pYC1	P_{TEF1} - <i>mcr</i> - T_{CYC1} <i>URA3</i>	[14]
pYC5	P_{TEF1} - <i>mcr</i> - T_{CYC1} P_{PGK1} - $ACCI^{(ser659ala, ser1157ala)}$ - T_{ADHI} <i>URA3</i>	This work
pIYC04	P_{TEF1} - T_{ADHI} P_{PGK1} - T_{CYC1} <i>HIS3</i>	[12]
pIYC05	P_{TEF1} - <i>acsSE</i> ^{L641P} - T_{ADHI} P_{PGK1} - <i>ALD6</i> - T_{CYC1} <i>HIS3</i>	[12]
p3HP02	P_{PGK1} - <i>PDC1</i> (5LS30)- T_{CYC1} P_{TEF1} - <i>acsSE</i> ^{L641P} - T_{ADHI} P_{PGK1} - <i>ALD6</i> - T_{CYC1} <i>HIS3</i>	This work
p3HP03	P_{PGK1} - <i>PDC1</i> (5LS30)- T_{CYC1} P_{TEF1} - <i>acsSE</i> ^{L641P} - T_{ADHI} P_{PGK1} - <i>ALD2</i> - T_{CYC1} <i>HIS3</i>	This work
XI-5	Vector containing homology regions for site XI-5 <i>URA3</i>	[38]
X-2	Vector containing homology regions for site X-2 <i>URA3</i>	[38]
pUG6	Vector containing the KanMX marker	[39]
pDel1	<i>loxP</i> P_{TEF1} -KanMX P_{GALI}	[Alexandra Bergman, unpublished]
pDel2	P_{GALI} - <i>cre</i> <i>loxP</i>	

The coding sequence of the *PDC1* 5LS30 gene [37] was constructed by introducing 5 codon substitutions on the CENPK113-11C genomic DNA template. The gene was constructed in three fragments: the first fragment was amplified with primers 1 and 2 (a list of primers used in this chapter is given in Table C1 of Appendix C) to introduce the first codon substitution, the second fragment was amplified with primers 3 and 4 to introduce the second codon substitution and the final fragment (last 750 base pairs) was

ordered from IDT technologies to include the other three codon substitutions. The three fragments were fused by PCR using primers 1 and 6, which resulted in the final sequence containing all the codon substitutions.

To construct plasmid p3HP02, the mutated *PDC1* coding sequence was cloned by gap repair into the *AscI* site of plasmid pIYC05 [12] under the control of a *PGK1* promoter (amplified from pIYC05 with primers 7 and 8) and a *CYC1* terminator (amplified from pIYC05 with primers 9 and 10). The final plasmid contains *PDC1* 5LS30 controlled by a *PGK1* promoter, followed by a codon-optimized version of the acetyl-CoA synthetase gene from *S. enterica* with the L641P mutation (*acs_{SE}^{L641P}*) under the control of a *TEF1* promoter and finally the *ALD6* gene also under the control of a *PGK1* promoter. The assembled plasmid was extracted from *S. cerevisiae* and transformed in *E. coli* in order to be amplified and confirmed by restriction digestion and sequencing.

To build the plasmid p3HP03, the *ALD6* coding sequence present in plasmid p3HP02 was replaced by *ALD2* using gap repair cloning. First, the coding sequence of *ALD2* was amplified from CEN.PK113-11C genomic DNA using primers 11 and 12. The primers contained regions homologous to the *PGK1* promoter and *CYC1* terminator in order to promote homologous recombination with the flanking regions of *ALD6*. CEN.PK113-11C was transformed with p3HP02 digested with *XhoI* (an *XhoI* site is located between the *CYC1* terminator and the *ALD6* coding sequence) and the *ALD2* DNA fragment, which resulted in the insertion of *ALD2* between the *PGK1* promoter and the *CYC1* terminator. The resulting plasmid, p3HP03, was extracted from *S. cerevisiae* and transformed in *E. coli* in order to be amplified and confirmed by restriction digestion and sequencing.

5.2.7 Strain construction

The strain CEN.PK113-11C, auxotrophic for histidine and uracil, was used as the base for the construction of all strains used in this work. The complete list of strains is shown in Table 5.2 with the corresponding genotype and source.

Table 5.2- Strains used in this work

Strain	Genotype	Source
CEN.PK113-11C	<i>MATa SUC2 MAL2-8^c ura3-52 his3-Δ1</i>	P. Kötter, University of Frankfurt, Germany
HPRA	(CEN.PK113-11C) <i>PTC5Δ::loxP</i>	This work
HPRB	(CEN.PK113-11C) <i>PTC5Δ::loxP ACH1Δ::KanMX</i>	This work
HPR00	(CEN.PK113-11C) $P_{TEFI-mcr}-T_{CYCI} P_{PGKI-ACCI}^{(ser659ala, ser1157ala)}-T_{ADHI} URA3$	This work
HPR01	(CEN.PK113-11C) $P_{TEFI-mcr}-T_{CYCI} P_{PGKI-ACCI}^{(ser659ala, ser1157ala)}-T_{ADHI} URA3 HIS3$	This work
HPR02	(CEN.PK113-11C) $P_{TEFI-mcr}-T_{CYCI} P_{PGKI-ACCI}^{(ser659ala, ser1157ala)}-T_{ADHI} URA3 P_{PGKI-PDC1} (5LS30)-T_{CYCI} P_{TEFI-acsSE}^{L641P}-T_{ADHI} P_{PGKI-ALD2}-T_{CYCI} HIS3$	This work
HPR03	(CEN.PK113-11C) $P_{TEFI-mcr}-T_{CYCI} P_{PGKI-ACCI}^{(ser659ala, ser1157ala)}-T_{ADHI} URA3 P_{PGKI-PDC1} (5LS30)-T_{CYCI} P_{TEFI-acsSE}^{L641P}-T_{ADHI} P_{PGKI-ALD4c}-T_{CYCI} HIS3$	This work
HPR04	(HPRA) $P_{TEFI-mcr}-T_{CYCI} P_{PGKI-ACCI}^{(ser659ala, ser1157ala)}-T_{ADHI} URA3 P_{PGKI-PDC1} (5LS30)-T_{CYCI} P_{TEFI-acsSE}^{L641P}-T_{ADHI} P_{PGKI-ALD2}-T_{CYCI} HIS3$	This work
HPR05	(HPRA) $P_{TEFI-mcr}-T_{CYCI} P_{PGKI-ACCI}^{(ser659ala, ser1157ala)}-T_{ADHI} URA3 P_{PGKI-PDC1} (5LS30)-T_{CYCI} P_{TEFI-acsSE}^{L641P}-T_{ADHI} P_{PGKI-ALD4c}-T_{CYCI} HIS3$	This work
HPR06	(HPRB) $P_{TEFI-mcr}-T_{CYCI} P_{PGKI-ACCI}^{(ser659ala, ser1157ala)}-T_{ADHI} URA3 P_{PGKI-PDC1} (5LS30)-T_{CYCI} P_{TEFI-acsSE}^{L641P}-T_{ADHI} P_{PGKI-ALD2}-T_{CYCI} HIS3$	This work
HPR07	(HPRB) $P_{TEFI-mcr}-T_{CYCI} P_{PGKI-ACCI}^{(ser659ala, ser1157ala)}-T_{ADHI} URA3 P_{PGKI-PDC1} (5LS30)-T_{CYCI} P_{TEFI-acsSE}^{L641P}-T_{ADHI} P_{PGKI-ALD4c}-T_{CYCI} HIS3$	This work

Strain HPRA was constructed by deleting the coding sequence of the phosphatase 2C (*PTC5*) in CEN.PK113-11C. Briefly, two overlapping DNA fragments containing the KanMX marker and the Cre recombinase were amplified from the plasmids pDel1 and pDel2 [Alexandra Bergman, unpublished] and fused to up- and downstream genomic sequences flanking *PTC5*. The upstream sequence was amplified from genomic DNA with primer pair 13/14 and the downstream region was amplified with primers 15/16. The DNA fragment containing the KanMX marker was amplified with primers 54/55 using pDel1 as template and the fragment containing the Cre recombinase gene was amplified with primers 56/57 using pDel2 as template. The upstream DNA fragment was fused to the KanMX fragment with primers 13/55 and the downstream DNA fragment was fused to the Cre recombinase fragment with primers 16/56. Both fusion fragments were transformed into CEN.PK113-11C and the deletion was confirmed by colony PCR using primer pairs 21/55 and 22/56. The fully assembled cassette was flanked by *loxP* sites and included the KanMX marker and a galactose inducible Cre recombinase. Therefore, to

loop out the whole cassette, the mutants confirmed by PCR were cultured in YPGal medium. After an overnight culture in YPGal medium the cells were plated in YPGal plates and the loopout was confirmed by colony PCR using primers 21 and 22.

Strain HPRB was obtained by deleting the coding sequence of succinyl-CoA:acetate CoA-transferase (*ACH1*) in strain HPRA. Briefly, two overlapping fragments [40] of the KanMX marker were amplified from the plasmid pUG6 [39] and fused to up- and downstream genomic sequences flanking *ACH1*. The upstream sequence was amplified from genomic DNA with the primer pair 23/24, while the downstream region was amplified with primers 25/26. The 5' fragment of the KanMX marker was amplified with primers 17/18 using pUG6 as template and the 3' overlapping fragment was amplified with primers 19/20 using the same template. The upstream DNA fragment was fused to the 5' KanMX fragment with primers 18/23 and the downstream DNA fragment was fused to the 3' KanMX fragment with primers 19/26. Both fusion fragments were transformed into HPRA and the deletion was confirmed by colony PCR using primers 18 and 27.

To obtain HPR00, the 12 kilobase (kb) *MreI/AscI* fragment from pYC5 was integrated in site 5 of chromosome XI [38] from CEN.PK113-11C. The *MreI/AscI* fragment, which contains $P_{TEFI-mcr}-T_{CYCI}$ and $P_{PGKI-ACCI}^{(ser659ala+ser1157ala)}-T_{ADHI}$, was co-transformed into CEN.PK113-11C with upstream and downstream DNA fragments designed to promote homologous recombination in the correct integration site. The upstream DNA fragment was amplified by PCR from the XI-5 vector [38] using primers 28/29 and includes: a genome homology region upstream of the integration site and the *ADHI* terminator. The downstream DNA fragment was amplified from the same vector using primers 30/31 and includes: a genome homology region downstream of the integration site, the *URA3* marker gene from *Kluyveromyces lactis* flanked by direct repeats and the *CYCI* terminator. The integration was confirmed by colony PCR with the primer pairs: 40/41, 38/42 and 43/44.

To create strain HPR01, the 3.8 kb *MreI/ApaLI* fragment from pIYC04 [12] was integrated in site number 2 of the chromosome X [38] from HPR00. The *MreI/ApaLI* fragment, which contains $P_{TEFI}-T_{ADHI}$, $P_{PGKI}-T_{CYCI}$ and the *HIS3* marker, was co-transformed into HPR00 with upstream and downstream DNA fragments designed to promote homologous recombination in the correct integration site. The upstream

fragment included: a genome homology region upstream of the integration site amplified from X-2 plasmid with primers 32/33 and a region homologous to the 5' end of *MreI/ApaLI* fragment amplified from p3HP03 with primers 34/35. Both fragments were fused by PCR using primers 32/35. The downstream fragment included: a region homologous to the downstream part of the integration site amplified from X-2 plasmid with primers 36/37 and a region homologous to the 3' end of *MreI/ApaLI* fragment amplified from p3HP03 with primers 38/39. The integration was confirmed by colony PCR with the primer pairs: 35/45 and 38/46.

To build strain HPR02 the same strategy was implemented to integrate the *MreI/ApaLI* fragment from p3HP03 into site number 2 of chromosome X [38] of the strain HPR00. The *MreI/ApaLI* fragment, which contains P_{PGKI} -*ALD2*- T_{CYC1} , P_{TEFI} -*acsSE*^{L641P}- T_{ADHI} , P_{PGKI} -*PDC1*(5LS30)- T_{CYC1} and the *HIS3* marker, was co-transformed into HPR00 with upstream and downstream DNA fragments designed to promote homologous recombination in the correct integration site. The upstream fragment included: a genome homology region upstream of the integration site amplified from X-2 plasmid with primers 32/33 and a region homologous to the 5' end of the *MreI/ApaLI* fragment amplified from p3HP03 with primers 34/35. Both fragments were fused by PCR using primers 32/35. The downstream fragment included: a genome homology region downstream of the integration site amplified from X-2 plasmid with primers 36/37 and a region homologous to the 3' end of the *MreI/ApaLI* fragment amplified from p3HP03 with primers 38/39. The integration was confirmed by colony PCR with the primer pairs: 35/45, 38/46, 11/39, 47/48 and 6/7.

The construction of strain HPR03 was hampered by the impossibility of cloning the truncated *ALD4* sequence (*ALD4c*) into the *ALD6* locus in p3HP02. Whenever the putative plasmid was transformed into *E. coli* there were few transformants that never contained the right construct. To solve this issue, the *ALD4c* sequence was amplified from the genomic DNA using the primer pair 49/50 and fused to a *PGKI* promoter (amplified with primers 51/52) and a *CYC1* terminator (amplified with primers 38/39) using primers 51 and 39. The complete *ALD4c* cassette was co-transformed with the 9.3 kb *ApaLI/XbaI* fragment from p3HP02 (the *XbaI* site is located in the middle of the *ALD6* sequence) and with the same upstream and downstream fragments created for the construction of HPR02. The integration was confirmed by colony PCR with the primer pairs: 35/45, 38/46, 39/53, 47/48 and 6/7.

Strains HPR04 and HPR06 were created by integrating the 12 kb *MreI/AscI* fragment isolated from pYC5 and the *MreI/ApaI* fragment from p3HP03 into the genome of HPPRA and HPRB, respectively. The same strategies described in the construction of strain HPR02 were used to integrate the two DNA fragments. To build strains HPR05 and HPR07 the same integration procedures described for strain HPR03 were followed using the basal strains HPPRA and HPRB as the template.

5.2.8 Batch cultivations

Batch fermentations were performed in the minimal medium described by Verduyn et al. [41] containing 20 g/L of glucose, 5 g/L of $(\text{NH}_4)_2\text{SO}_4$, 3 g/L of KH_2PO_4 , 0.5 g/L of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.05 mL/L of antifoam, 1 mL/L of a vitamin solution and 1 mL/L of a trace metal solution. The trace metal solution contained 15 g/L of Na_2EDTA , 4.5 g/L of $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 0.84 g/L of $\text{MnCl}_2 \cdot 2\text{H}_2\text{O}$, 0.3 g/L of $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, 0.3 g/L of $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0.4 g/L of $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, 4.5 g/L of $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 3 g/L of $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 1 g/L of H_3BO_3 and 0.1 g/L of KI. The vitamin solution contained 0.05 g/L of biotin, 1 g/L of calcium pantothenate, 1 g/L of thiamin hydrochloride, 1 g/L of pyridoxine hydrochloride, 1 g/L of nicotinic acid, 0.2 g/L of p-aminobenzoic acid and 25 g/L of myo-inositol. The medium used to prepare the pre-cultures in shake-flasks is the same as above with the following modifications: no antifoam, 7.5 g/L of $(\text{NH}_4)_2\text{SO}_4$, 14.4 g/L of KH_2PO_4 and the pH was adjusted to 6.5 with NaOH before autoclaving.

Each fermenter was inoculated with an initial OD_{600} of 0.05 using a pre-culture obtained by cultivating a single colony of the desired strain in 25 mL of minimal medium inside a 100 mL shake-flask at 200 rpm in an orbital shaker kept at 30°C. The batch fermentations were performed in a DASGIP® Parallel Bioreactor System for Microbial Applications with 8 simultaneous bioreactors placed on a temperature controlled BioBlock. The operating volume for the fermentations was 0.5 L, the temperature set-point was maintained at 30 °C, the airflow was set at 1 VVM, the pH was controlled by feedback controlled addition of 2 M KOH, the dissolved oxygen was kept above 30 % of saturation by feedback control of the stirring speed from 400 rpm until a maximum of 1200 rpm. The concentration of O_2 and CO_2 in exhaust gas was monitored by a DASGIP® GA4 exhaust analyzer.

5.2.9 Analytical methods

Cell dry weight was determined by filtering 5 mL of culture broth through a 0.45 μm pore filter from Sartorius. The initial weight of the filters was determined after drying them in a microwave oven for 30 min at 150 W and letting them cool in a desiccator. The filters were then used to filter 5 mL of culture broth and washed twice with 5 mL of deionized water. Again, the filters were dried for 30 min at 150 W in a microwave oven and stored in a desiccator before being weighted.

Samples from the bioreactors were taken at precise sampling times and filtered using syringe nylon filters with a pore size of 0.45 μm into HPLC vials and stored in a $-20\text{ }^{\circ}\text{C}$ freezer until being analyzed. HPLC analysis was performed in a Dionex UltiMate 3000 systems using a refractive index detector (RI-101 from Shodex) and an absorbance detector set at 210 nm (Ultimate 3000 VWD from Dionex). The samples were analyzed using an Aminex HPX-87H column from Bio-Rad, which was kept at $65\text{ }^{\circ}\text{C}$ and 0.5 mM H_2SO_4 was used as the mobile phase with a flow rate of 0.5 mL/min. Quantitative analysis of glucose, 3-HP, glycerol, acetate and ethanol was performed by injecting a mixture of standards with known concentrations of each metabolite. Calibration curves were calculated using the peak areas of the RI detector for glucose, glycerol and ethanol and of the absorbance detector for 3-HP and acetate.

5.3 Results and discussion

5.3.1 *In silico* optimization of 3-HP production under glucose limitation

Strain optimization strategies are often complex and involve different types of genetic alterations. We used TDPS (chapter 4) in combination with an evolutionary algorithm to search for combinations of up/down-regulations and deletions that would optimize the production of 3-HP under steady-state conditions. The goal was to find metabolic engineering strategies that optimize cytosolic acetyl-CoA availability and can be applied to other products that share this precursor. Table 5.3 shows the details of four different virtual strains (A-D) obtained *in silico* from TDPS simulations and that have increased 3-HP production and also a viable growth phenotype.

Table 5.3 - Metabolic engineering strategies obtained using TDPS for 3-HP optimization

Strain	<i>In silico</i> genotype	TDPS	
		3-HP yield on glucose	Biomass yield on glucose
A	Up-regulation: 5x ACC, 2x MCR	0.023 g/g	0.36 g/g
B	Up-regulation: 2x PDC, 2x ALDc, 2x ACS, 2x ACC, 2x MCR	0.20 g/g	0.18 g/g
C	Up-regulation: 2x PDC, 2x ALDc, 2x ACS, 5x ACC, 2x MCR; Down-regulation: 0.5x PDH	0.33 g/g	0.13 g/g
D	Up-regulation: 2x PDC, 2x ALDc, 2x ACS, 5x ACC, 2x MCR; Down-regulation: 0.5x PDH; Deletion: SCAT	0.37 g/g	0.054 g/g

Metabolic reactions: PDC- pyruvate decarboxylase, ALDc- cytosolic NAD⁺ aldehyde dehydrogenase, ACS- acetyl-CoA synthetase, ACC- Acetyl-CoA carboxylase, MCR- malonyl-CoA reductase, PDH- pyruvate dehydrogenase complex, SCAT- succinyl-CoA:acetate CoA-transferase

With the up-regulation of acetyl-CoA carboxylase (ACC) and expression of malonyl-CoA reductase (MCR), strain A represents the basal level of 3-HP obtainable with the endogenous supply of cytosolic acetyl-CoA under fully oxidative growth on glucose. Although MCR is referred in Table 5.3 as an up-regulation, this terminology is used to denote an increase of activity in absolute terms, not in comparison to a wild-type organism (*S. cerevisiae* does not possess MCR activity endogenously). To improve the 3-HP production yield, the optimization algorithm suggested, in strain B, the up-regulation of three genes that constitute the pyruvate dehydrogenase bypass (PDH bypass). The combined action of pyruvate decarboxylase (PDC), aldehyde dehydrogenase (ALDc) and acetyl-CoA synthetase (ACS) is able to deviate flux away from the oxidative metabolism of pyruvate (inside the mitochondria) and boost the amount of acetyl-CoA present in the cytosol. Strain B shows an improvement in the 3-HP production yield of 8-fold, when compared to strain A, at the expense of a 50% drop in biomass formation.

Further improvement of strain B can be achieved by down-regulating the pyruvate dehydrogenase complex in the mitochondria as shown in strain C (Table 5.3). This enzymatic complex is responsible for most of the pyruvate consumption and its attenuation provokes an increase of flux through pyruvate decarboxylase and subsequent reactions of the PDH bypass (strain C). The severe decrease in oxidative metabolism is responsible for the biomass yield to drop to 36% of strain A (Table 5.3). The increased

flux to acetyl-CoA can then be channeled into 3-HP, by ACC and MCR, and results in a 65% increase of the 3-HP production yield over strain B.

The final step in the improvement of 3-HP was achieved by preventing that part of the acetaldehyde formed in the cytosol would be channeled into the mitochondria, where it could be converted to acetyl-CoA by the action of mitochondrial aldehyde dehydrogenase and succinyl-CoA:acetate CoA-transferase (SCAT). The inactivation of SCAT in strain D results in an increase in the flux through the PDH bypass, which further improves the 3-HP yield by 12% (16-fold improvement overall) while reducing the biomass yield to 15% of strain A. Figure 5.1 shows the metabolic context of all the modifications suggested by the simulations and the detailed information about flux rearrangements is provided in the Appendix C (Figure C1).

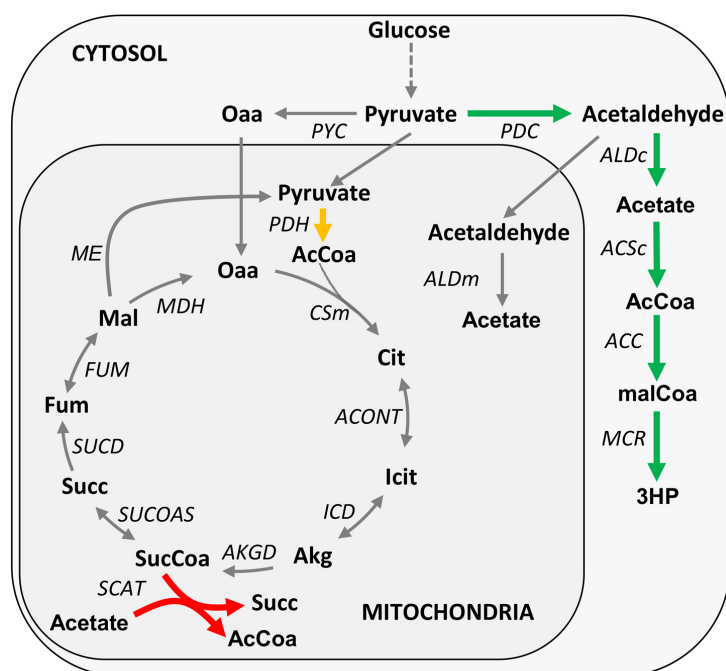


Figure 5.1- Schematic representation of the flux modifications suggested by the TDPS algorithm. Up-regulations are shown in green, down-regulations in yellow and deletions in red. MCR is referred as an up-regulation but in this case this terminology is used to denote an increase of activity in absolute terms, not in comparison to a wild-type. Reactions: ACC- acetyl-CoA carboxylase, ACONT- aconitase, ACSc- acetyl-CoA synthetase, AKGD- alpha-ketoglutarate dehydrogenase, ALDc- cytosolic aldehyde dehydrogenase, ALDm- mitochondrial aldehyde dehydrogenase, CSm- citrate synthase, FUM- fumarase, ICD- mitochondrial isocitrate dehydrogenase, MCR- malonyl-CoA reductase, MDH- mitochondrial malate dehydrogenase, ME- mitochondrial malic enzyme, PYC- pyruvate carboxylase, PDH- pyruvate dehydrogenase complex, PDC- pyruvate decarboxylase, SCAT- succinyl-CoA:acetate CoA-transferase, SUCD- succinate dehydrogenase, SUCOAS- succinyl-CoA ligase; Metabolites: 3HP- 3-hydroxypropionic acid, AcCoA- acetyl-CoA, Akg- 2-oxoglutarate, Cit- citrate, Fum- fumarate, Icit- isocitrate, Mal - L-malate, malCoA- malonyl-CoA, Oaa- oxaloacetate, Succ- succinate, SucCoA- succinyl-CoA.

5.3.2 Translation of the simulation results into strain engineering strategies

When it comes to implementation of strain optimization strategies that were obtained from *in silico* metabolic modelling, there are many interpretations that can be derived for the same virtual strain. Although gene deletions are quite straightforward to implement, up/down-regulations may have different interpretations, i.e., change in gene expression, modification of protein regulation or use of enzymes with attenuated/improved kinetic parameters. Depending on the type of regulation present in the host, it is of utmost importance to choose the right strategy to maximize the impact of each modification at the flux level. In order to convert the *in silico* genotypes from Table 5.3 into genetic engineering operations, each target was analyzed in detail to determine the best option to achieve the desired phenotype. Table 5.4 summarizes all the genetic modifications conducted to mimic the flux changes observed at the simulation level.

Table 5.4- Genetic engineering strategies chosen to mimic the flux regulations obtained in the simulations

<i>In silico</i> modification	Genetic modification	
	Expression	Regulation
Up-regulation MCR ^a	Integration of a heterologous Malonyl-CoA reductase from <i>Chloroflexus aurantiacus</i> (<i>mcr</i>) [14] under a strong promoter	-
Up-regulation ACC	Integration of a mutated <i>ACC1</i> from <i>S. cerevisiae</i> [25] under a strong promoter	The mutated (ser659ala,ser1157ala) <i>Acc1p</i> is no longer regulated at the protein level by Snf1p-mediated phosphorylation and shows improved activity
Up-regulation PDC	Integration of a mutated <i>PDC1</i> from <i>S. cerevisiae</i> [37] under a strong promoter	The mutated <i>Pdc1</i> contains five point mutations that improved its activity (lower $S_{0.5}$) and reduced its inhibition by phosphate
Up-regulation ACS	Integration of a mutated <i>acs</i> from <i>S. enterica</i> (<i>acs</i> _{SE} ^{L641P}) [10, 12] under a strong promoter	Regulation by acetylation at the protein level was abolished in the mutated <i>Acs</i> (<i>leu641pro</i>)
Up-regulation ALDc	- Integration of <i>ALD2</i> under a strong promoter - Integration of <i>ALD4</i> (without mitochondrial targeting signal) [42] under a strong promoter	-
Down-regulation of PDH	Deletion of <i>PTC5</i> [43, 44]	Activation of the PDH complex by the phosphatase <i>Ptc5p</i> was abolished, resulting in lower activity
Deletion SCAT	Deletion of <i>ACHI</i>	-

^a MCR is referred as an up-regulation but in this case this terminology is used to denote an increase of activity in absolute terms, not in comparison to a wild-type.

Metabolic reactions: **PDC-** pyruvate decarboxylase, **ALDc-** cytosolic NAD⁺ aldehyde dehydrogenase, **ACS-** acetyl-CoA synthetase, **ACC-** Acetyl-CoA carboxylase, **MCR-** malonyl-CoA reductase, **PDH-** pyruvate dehydrogenase complex, **SCAT-** succinyl-CoA:acetate CoA-transferase

As shown in Table 5.4, it was necessary to introduce a heterologous gene encoding malonyl-CoA reductase in *S. cerevisiae*, because 3-HP is not naturally produced in this microorganism. Based on the work of Chen et al., we decided to use a Mcr from *C. aurantiacus* because it has shown increased activity compared to other enzymes when expressed in yeast [14]. Regarding the up-regulation of ACC, it was previously shown by Shi et al. [25] that efficient up-regulation of Acc1p activity in *S. cerevisiae* is achieved by two point mutations to abolish regulation by Snf1p. In order to bypass Acc1p regulation and increase the flux through this enzyme, the mutated *ACCI* was integrated in the genome under the control of a strong constitutive promoter.

The great improvement in 3-HP production shown for strain B (Table 5.3) is the result of an efficient up-regulation of the PDH bypass. In order to replicate this phenotype *in vivo*, the three enzymes that constitute the PDH bypass need to have good kinetic properties and low regulation. The first of the three enzymes, pyruvate decarboxylase (Pdc1p), has a substrate affinity constant for pyruvate (K_m) ten times higher than the PDH complex [45]. Consequently, in glucose limited cultures growing at low growth rates ($< 0.20 \text{ h}^{-1}$) the over-expression of Pdc1p has no impact on the flux distribution around the pyruvate node [46]. Furthermore, even under glucose abundance conditions the over-expression of Pdc1p on its own does not have a significant effect on the ethanol formation rate [47]. Given the kinetic limitations of this enzyme, it was unlikely that a simple over-expression would change the flux ratio in the pyruvate node in favor of the PDH bypass. Therefore, we opted to use a mutated Pdc1p (5LS30) evolved by Stevenson et al. [37] that showed a 3-fold reduction in the pyruvate half-saturation constant ($S_{0.5}$) and 4-fold reduction in phosphate inhibition (Table 5.4).

Additionally, the simulation results suggested the up-regulation of a NAD^+ dependent aldehyde dehydrogenase (ALDc) for the second step of the PDH bypass. The analysis of the flux distribution suggested that the use of a NAD^+ dependent enzyme under glucose limitation may have a positive impact on 3-HP production by increasing the amount of ATP available in cell, which is required for acetyl-CoA (2 ATP equivalents) and malonyl-CoA synthesis (1 ATP). In *S. cerevisiae*, Ald4p and Ald6p are the major enzymes with acetaldehyde dehydrogenase activity [48]. While Ald6p is cytosolic and NADP^+ dependent, Ald4p is localized in the mitochondria and can use both NAD^+ and NADP^+ [48]. There are additional aldehyde dehydrogenases in *S. cerevisiae* (Ald2p, Ald3p and Ald5p) but their activity towards acetaldehyde is much lower [48, 49]. As shown in Table

5.4, Ald2p and a cytosolic version of Ald4p [42] were over-expressed separately in order to test which of them would result in higher flux through the PDH bypass.

The final enzyme in the PDH bypass, acetyl-CoA synthetase (ACS), was up-regulated by integrating a copy of a mutant version of this enzyme from *S. enterica* into the yeast genome [10]. This enzyme is acetylation-insensitive and was shown to have a positive influence on the production of metabolites that have acetyl-CoA as their precursor [9–12, 14].

In order to decrease the rate of oxidative metabolism of pyruvate, it was suggested by the simulations that the PDH complex should be down-regulated. As mentioned above this type of modification can have many interpretations and not all of them may produce the desired effect. Based on the work of Gey et al. [43], there are two phosphatases that contribute to the dephosphorylation of the PDH complex, Ptc5p and Ptc6p, and both of them contribute to the activation of this enzymatic complex. Therefore, the down-regulation of the PDH-complex can be achieved by deleting either of the corresponding genes [43]. According to published results, the deletion of *PTC5* results in a less severe phenotype (56 % reduction in activity) than *PTC6* inactivation (94 % reduction in activity) [43, 44]. Therefore, since the simulation results showed that a reduction of 50 % in the activity of the PDH complex would be optimal, we chose to delete *PTC5*.

The final strain engineering operation was the inactivation of succinyl-CoA:acetate CoA-transferase inside the mitochondria. According to the literature, *ACH1* is the only gene in *S. cerevisiae* coding for this activity and its deletion should result in the desired effect [3].

5.3.3 Strain analysis in bioreactors

The constructed strains (HPR01-HPR07) were cultivated under the controlled conditions of a bioreactor and the extracellular metabolite concentrations were measured at two time-points: after glucose was exhausted, which represents the respiro-fermentative metabolism of *S. cerevisiae*; and after all the ethanol was consumed, which represents fully oxidative metabolism of the products of fermentation (ethanol, glycerol and acetate).

Figure 5.2 shows the overall 3-HP yields on glucose at the end of the batch cultivation, i.e., after the O₂ consumption and biomass production dropped to zero. When the yield of HPR01 is compared to the corresponding simulation result in Table 5.3 (strain A), we can

see that the results are quite similar in terms of magnitude (0.013 g/g versus 0.023 g/g). However, the strains engineered for improved acetyl-CoA availability performed worse than what would have been expected from the simulation results. Furthermore, the best constructed strain (HPR06) showed an increase of 52 % over the basal strain, which is much lower than the 16-fold improvement predicted.

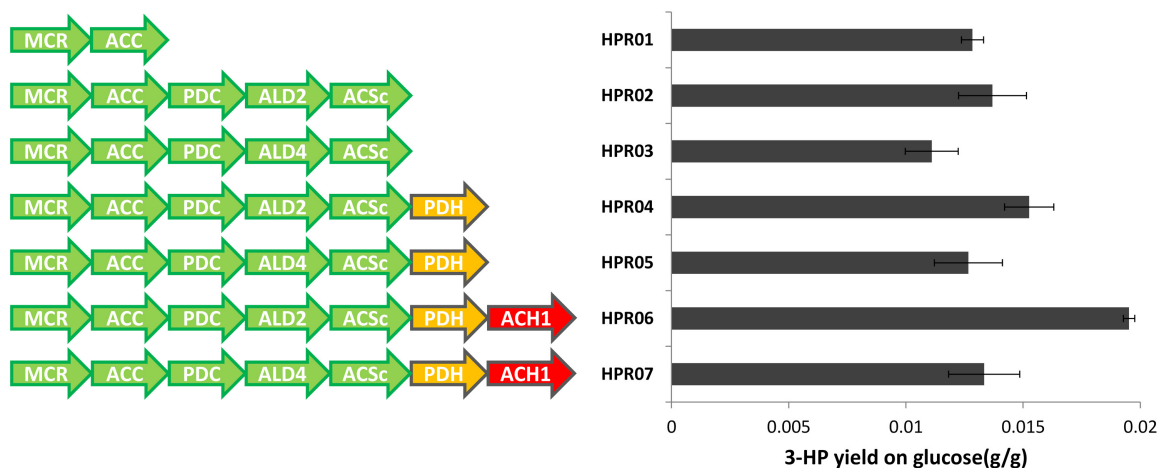


Figure 5.2- 3-HP yields on glucose at the end of the batch cultivation for the strains engineered to improved cytosolic acetyl-CoA supply. The values shown are the average of at least two biological replicates and the error bars represent the standard deviations. MCR- Expression of the malonyl-CoA reductase gene from *C. aurantiacus*, ACC- up-regulation of *ACC1*, PDC- up-regulation of *PDC1*, ALD2- up-regulation of *ALD2*, ALD4- expression of a truncated *ALD4* lacking the mitochondrial targeting signal, ACSc- expression of a mutated *acs* from *S. enterica*, PDH- deletion of *PTC5* to decrease the activity of the Pyruvate Dehydrogenase complex, ACH1- deletion of the succinyl-CoA:acetate CoA-transferase (*ACH1*).

5.3.3.1 Strain analysis under respiro-fermentative glucose metabolism

The simulation results showed that the most important step in the 3-HP optimization process was the supply of cytosolic acetyl-CoA using the PDH bypass. As discussed in section 5.3.2 the success of this strategy relies on the capacity of the three enzymes from this pathway to redirect flux from pyruvate oxidation to acetyl-CoA production. Under respiro-fermentative conditions the flux through PDC should not be limiting, because there is a considerable amount of acetaldehyde being produced by this enzyme [50]. Therefore, the limiting step for 3-HP production under respiro-fermentative conditions should lie downstream of PDC.

The next step of the PDH bypass is the oxidation of acetaldehyde to acetate catalyzed by an aldehyde dehydrogenase. In this work, we tested two NAD^+ -dependent aldehyde dehydrogenases, Ald2p and Ald4cp, in order to compare which one would result in a

higher flux through this pathway. As shown in Table 5.5, all strains over-expressing Ald4cp (HPR03, HPR05 and HPR07) accumulated around 3-fold more acetate extracellularly. The increased NAD⁺-dependent acetaldehyde dehydrogenase activity in the cytosol is also corroborated by the 2-fold increase in glycerol production in the strains expressing Ald4cp. Since glycerol production is used by the cell as a “redox valve” to dispose of excessive NADH [51], the increase in NADH originating from acetate formation was compensated (partially) by glycerol formation. The strains over-expressing Ald2p (HPR02, HPR04, HPR06) also showed increase production of acetate in comparison with the control strain (HPR01), but in lower amounts than observed for the Ald4cp strains.

Table 5.5- Physiological properties during respiro-fermentative metabolism of the strains engineered for 3-HP production.

Strain	Acetate (g/L)	Glycerol (g/L)	Ethanol (g/L)	Dry weight (g/L)	3HP (g/L)	μ_{max} (h ⁻¹)
HPR01	0.22 ± 0.00	1.06 ± 0.03	9.25 ± 0.44	2.40 ± 0.21	0.18 ± 0.01	0.35 ± 0.01
HPR02	0.30 ± 0.02	1.15 ± 0.10	9.28 ± 0.42	2.53 ± 0.09	0.20 ± 0.01	0.32 ± 0.03
HPR03	0.88 ± 0.03	2.21 ± 0.10	8.31 ± 0.30	2.35 ± 0.09	0.16 ± 0.01	0.31 ± 0.02
HPR04	0.31 ± 0.02	1.01 ± 0.03	9.27 ± 0.10	2.25 ± 0.09	0.21 ± 0.00	0.36 ± 0.01
HPR05	0.92 ± 0.02	2.16 ± 0.01	8.26 ± 0.00	2.14 ± 0.18	0.18 ± 0.00	0.32 ± 0.01
HPR06	0.35 ± 0.02	0.86 ± 0.06	9.29 ± 0.16	2.25 ± 0.08	0.21 ± 0.03	0.36 ± 0.02
HPR07	1.05 ± 0.07	2.38 ± 0.16	8.38 ± 0.02	1.82 ± 0.02	0.16 ± 0.01	0.29 ± 0.02

The samples were taken right after the glucose in the medium was exhausted and the values are given in g/L for the extracellular metabolite concentrations and for cell dry weight. The maximum growth rate (μ_{max}) is given in h⁻¹. The values shown are the average of at least two biological replicates ± the standard deviations.

Interestingly, the strains expressing Ald4cp (HPR03, HPR05 and HPR07) accumulate less 3-HP than the Ald2p over-expressing counterparts, meaning that it is very unlikely that ALD activity is the limiting step for 3-HP production. One possible explanation for this counter-intuitive phenotype may be the decreased availability of NADPH caused by the competition between the native NADP⁺-dependent Ald6p and the introduced NAD⁺-dependent Ald4cp. Since 3-HP production requires 2 NADPH molecules per 3-HP molecule, decreasing NADPH formation in the cytosol may result in a lower activity of Mcr and lower 3-HP production. In fact, it has been shown before that increasing the availability of NADPH has a positive impact on 3-HP production [14].

Judging by the levels of acetate accumulation for the strains expressing Ald4cp, ALD activity does not seem to be the reason for the low performance of these strains. ACS is much more likely to be the limiting step in acetyl-CoA production, even though the strains constructed include an unregulated version of this enzyme from *S. enterica* (section 5.3.2). In previous studies, the acs_{SE}^{L641P} was expressed from multi-copy plasmids in *S. cerevisiae* [9, 11, 12, 14], while here only a single copy was integrated into the genome. Given the heterologous nature of this enzyme it is probable that the single copy integration method used may result in limiting amounts of ACS activity. Unfortunately, our attempts of using strains based on expression from two plasmids (p3HP02 + pYC5) instead of genomic integrations resulted in severe phenotypic instability, which prevented long term culture of those strains (data not shown).

Another possibility for the lack of effect of the PDH bypass amplification on 3-HP production can also be related with potential additional post-transcriptional regulation of Acc1p or insufficient enzyme levels of Acc1p and/or Mcr. On one hand, we integrated a single copy of the mutated *ACCI* and of *mcr*, which might be limiting under these conditions. On the other hand it is also possible that Acc1p is subject to additional regulation mechanisms that have not been reported yet. Another factor that was not analyzed is the availability of biotin and of biotin:apoprotein ligase (Bpl1p) to activate Acc1p. In fact, it has been shown in *E. coli* that additional levels of both of these components boosted the activity of *Photorhabdus luminescens* Acc [52].

The down-regulation of the PDH complex under respiro-fermentative metabolism should have rather little impact since it is already repressed by elevated glucose concentrations [53, 54]. As expected there were no significant differences between HPR02/3 and their PDH down-regulated equivalent counterparts HPR04/5. Furthermore, the deletion of *ACHI* also had no significant effect on 3-HP production, although it appears to be detrimental to biomass formation when combined with Ald4cp expression. As shown in Table 5.5 for strain HPR07 the drop in biomass formation is coincident with the highest acetate and glycerol formation among all strains.

5.3.3.2 Oxidative growth on fermentation products

When *S. cerevisiae* is grown on ethanol as the sole carbon source several changes occur in central carbon metabolism. Ethanol is first oxidized to acetaldehyde and then converted

to acetyl-CoA using the enzymes in the PDH bypass. Acetyl-CoA can then enter the glyoxylate cycle to supply energy and biomass precursors to the cell [2]. The production of acetyl-CoA is so vital in these conditions that an additional acetyl-CoA synthetase (Acs1p) becomes active to accommodate the increased flux to acetyl-CoA [55, 56]. Therefore, when compared to respiro-fermentative metabolism there should be a significant improvement in the availability of cytosolic acetyl-CoA in cells growing on ethanol. However, as shown in Table 5.6, the 3-HP production levels were quite low taking into consideration the increased flux through acetyl-CoA.

Table 5.6- Physiological properties during oxidative growth on the fermentation products of the strains engineered for 3-HP production.

Strain	Acetate (g/L)	Glycerol (g/L)	Ethanol (g/L)	Dry weight (g/L)	3HP (g/L)	μ_{max} (h ⁻¹)
HPR01	-0.22 ± 0.00	-0.98 ± 0.02	-9.25 ± 0.44	3.77 ± 0.07	0.08 ± 0.03	0.10 ± 0.01
HPR02	-0.30 ± 0.02	-1.01 ± 0.09	-9.28 ± 0.42	3.57 ± 0.19	0.07 ± 0.01	0.08 ± 0.00
HPR03	-0.88 ± 0.03	-1.04 ± 0.12	-8.31 ± 0.30	3.45 ± 0.16	0.06 ± 0.04	0.10 ± 0.01
HPR04	-0.30 ± 0.02	-0.91 ± 0.01	-9.27 ± 0.10	3.58 ± 0.06	0.10 ± 0.02	0.10 ± 0.01
HPR05	-0.91 ± 0.02	-1.00 ± 0.07	-8.26 ± 0.00	3.49 ± 0.07	0.08 ± 0.02	0.12 ± 0.01
HPR06	-0.35 ± 0.01	-0.43 ± 0.10	-9.29 ± 0.16	2.66 ± 0.14	0.18 ± 0.01	0.05 ± 0.01
HPR07	-0.32 ± 1.34	-0.21 ± 0.07	-8.38 ± 0.02	1.71 ± 0.47	0.11 ± 0.02	0.03 ± 0.01

Changes in metabolite concentrations and cell dry weight (compared to the end of the glucose phase) are given in g/L. Negative values represent the consumption and positive represent the production of each species. The maximum growth rate (μ_{max}) is given in h⁻¹. The values shown are the average of at least two biological replicates ± the standard deviations.

Comparing the performance of the strains HPR02/3 to the control strain HPR01 reveals that as observed in respiro-fermentative conditions the up-regulation of the PDH bypass enzymes does not lead to an improvement in the 3-HP production levels (Table 5.6). Taking into consideration that we expressed an unregulated ACS from *S. enterica* and that Acs1p should also be active in the cytosol, it appears that the limiting step in 3-HP production is located downstream of acetyl-CoA metabolism. Even if Acs1p is mostly active in the peroxisome in order to fuel the glyoxylate cycle [2], the additional activity from the introduced Acs_{SE}^{L641P} should be supplying additional acetyl-CoA in the cytosol. Assuming that acetyl-CoA is in fact not limiting under growth on ethanol, means that the activity of Acc1p and/or Mcr are the bottleneck in this pathway.

Judging by the similar levels of biomass formation between HPR01 and HPR02/3, it appears that the amount of acetyl-CoA being diverted to other malonyl-CoA derived

metabolites (such as lipids) is not substantial enough to affect the final cell dry weight. However, we did not measure the accumulation of any additional metabolites derived from malonyl-CoA and consequently cannot conclude about a less pronounced increase in the flux through Acc1p. In fact, it has been shown before that the expression of the mutated Acc1p used in this study results in an increase of fatty acid ethyl esters accumulation in *S. cerevisiae* [25]. As mentioned before, it is possible that insufficient levels of Acc1p or lack of biotin or Bpl1p activity are limiting the flux through the acetyl-CoA carboxylase from reaching higher fluxes.

Another possibility for the low 3-HP production levels observed could be that insufficient enzyme activity of Mcr is limiting the flux to 3-HP. Since Mcr is competing for malonyl-CoA with endogenous enzymes of *S. cerevisiae*, this phenotype could be explained by a low substrate affinity of Mcr towards malonyl-CoA in comparison with other consuming enzymes. If other enzymes are very efficient at metabolizing malonyl-CoA, the flux to 3-HP might be limited because of the low activity of Mcr caused by a low intracellular concentration of malonyl-CoA. This hypothesis is supported by an increase in 3-HP production in the ethanol phase when extra *mcr* is expressed from a multi-copy plasmid in strain HPR00 (data not shown).

In contrast to the differences observed under respiro-fermentative conditions, under growth on ethanol the expression of Ald4cp did not influence 3-HP formation (Table 5.6). When strain HPR03 (Ald4cp) is compared to HPR02 (Ald2p) the only differences observed are the amount of ethanol and acetate consumed, which were accumulated in different amounts during the glucose consumption phase. This lack of effect is supported by the finding that the deletion of *ALD6* can be compensated by Ald4p activity [42], which indicates that the cofactor specificity of this enzyme is not paramount for normal growth on ethanol.

Regarding the behavior of the PDH down-regulation mutants (HPR04 and HPR05) a small increase was observed in 3-HP production (43 % and 33 %) compared to the original strains (HPR02 and HPR03). However, the increase is not statistically significant and therefore quite difficult to interpret. This may indicate that PDH has a small role in ethanol or glycerol utilization and that its down-regulation increases the flux through the PDH bypass. Nevertheless, the lack of significant phenotypical changes in the *PTC5* deletion strains suggests that the PDH complex is not very relevant for the consumption

of ethanol or glycerol. These findings are supported by the lack of phenotype of a PDH-negative mutant grown on ethanol limited chemostats [29].

Of all genetic modification implemented, *ACH1* deletion resulted in the most severe phenotype when the cells were growing in the mixture of fermentation products. Strain HPR06 showed a prolonged growth phase on ethanol, which lasted 23 hours in comparison to the 12.5 hours of HPR04 (data not shown), and a 50% decrease in the maximum growth rate (Table 5.6). During the initial period of ethanol consumption this strain accumulated extracellular acetate (data not shown) as observed by Fleck and Brock in their *ACH1* deletion mutant [3]. The acetate “overflow” phenotype suggests that Ach1p is responsible for consuming part of the acetate originating from ethanol oxidation. As shown in Table 5.6, HPR06 shows an improvement in 3-HP production of over 2-fold when compared to HPR01. This phenotype is probably caused by the increased acetyl-CoA production in the cytosol as a result of the inactivation of a competing pathway for acetate consumption.

Unexpectedly, HPR07, which differs from HPR06 in the aldehyde dehydrogenase used (Ald4cp vs. Ald2p), showed an even longer growth phase on ethanol (data not shown) and difficulty in fully consuming the acetate accumulated (Table 5.6). This odd phenotype may be caused by an imbalance of activity between the cytosolic aldehyde dehydrogenase and acetyl-CoA synthetase, which results in acetate induced stress and reduced biomass yield and 3-HP production (Table 5.6). A thorough analysis of the *ACH1* deletion phenotype and its interaction with Ald4cp expression is given in section 5.3.4.

5.3.4 Integration of Ach1p role into C₂ metabolism

As show by Fleck and Brock [3] and reinforced by the phenotypes observed (Table 5.6), Ach1p has an important role in the metabolism of C₂ carbon sources by *S. cerevisiae*. At the same time, the PDH bypass and the glyoxylate cycle also have a critical part in C₂ metabolism [2]. In order to estimate the contribution of each pathway for ethanol metabolism we used flux balance analysis [57, 58] and a genome scale metabolic model of *S. cerevisiae* to estimate the fluxes in the central carbon metabolism. Figure 5.3 shows the computed fluxes in the most relevant reactions for the wild-type *S. cerevisiae* (green) and for the *ACH1* deletion mutant (red). The flux values were estimated assuming

maximum biomass yield under a fixed ethanol influx. Therefore, the flux distribution represents the most efficient metabolic state achievable using the enzymatic transformations present in the model.

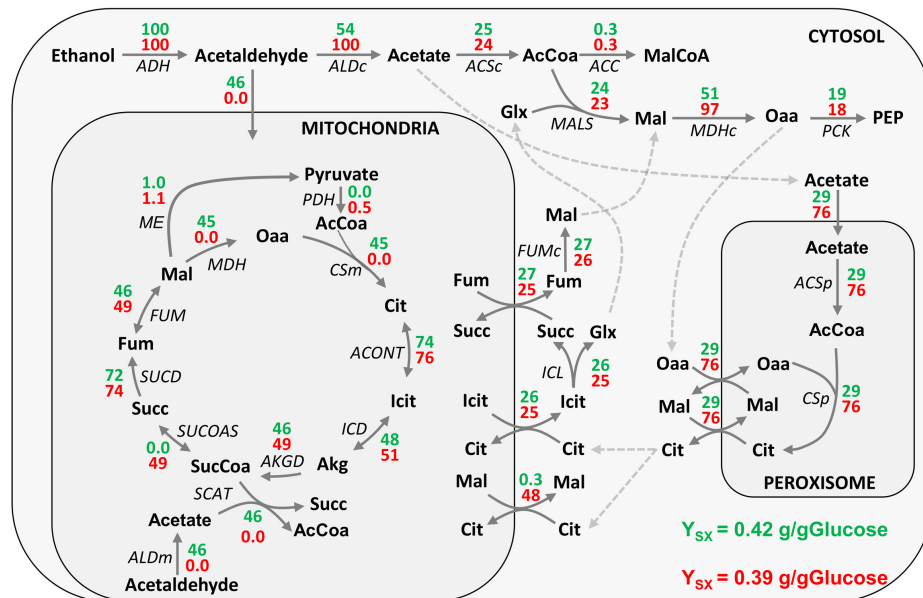


Figure 5.3- Flux distributions for growth on ethanol of wild-type *S. cerevisiae* (green) and *ACH1* deletion mutant (red) obtained using FBA. Reactions: ACC- acetyl-CoA carboxylase, ACONT- aconitase, ACSc- cytosolic acetyl-CoA synthetase, ACSp- peroxisomal acetyl-CoA synthetase, ADH- alcohol dehydrogenase, AKGD- alpha-ketoglutarate dehydrogenase, ALDc- cytosolic aldehyde dehydrogenase, ALDm- mitochondrial aldehyde dehydrogenase, CSm- mitochondrial citrate synthase, CSp- peroxisomal citrate synthase, FUM- mitochondrial fumarase, FUMc- cytosolic fumarase, ICD- mitochondrial isocitrate dehydrogenase, ICL- isocitrate lyase, MALS- malate synthase, MDH- mitochondrial malate dehydrogenase, MDHc- cytosolic malate dehydrogenase, ME- mitochondrial malic enzyme, PCK- phosphoenolpyruvate carboxykinase, PDH- pyruvate dehydrogenase complex, SCAT- succinyl-CoA:acetate CoA-transferase, SUCD- succinate dehydrogenase, SUCOAS- succinyl-CoA ligase; Metabolites: AcCoA- acetyl-CoA, Akg- 2-oxoglutarate, Cit- citrate, Fum- fumarate, Glx- glyoxylate, Icit- isocitrate, Mal - L-malate, malCoA- malonyl-CoA, Oaa- oxaloacetate, PEP- phosphoenolpyruvate, Succ- succinate, SucCoa- succinyl-CoA.

As shown in Figure 5.3 for the wild-type flux distribution, there is a split of flux at the acetaldehyde node that diverges either to acetyl-CoA synthetase (ACS) in the cytosol/peroxisome or to succinyl-CoA:acetate CoA-transferase (SCAT) in the mitochondria. Since Ach1p is more energy efficient (one ATP equivalent needed) than ACS (two ATP equivalents needed) in the activation of acetate to acetyl-CoA, it would be logical that Ach1p is the preferred route if we disregard their occurrence in different compartments. However, given the small differences in the simulated yields between wild-type and *ACH1* deletion mutant (Figure 5.3) it is not likely that there is a significant selective pressure on the use of Ach1p based on efficiency only. Instead, it is much more likely that Ach1 is used to maximize the capacity of the cell to metabolize C₂ carbon

sources and achieve higher growth rates. This theory is supported by the observation that *ACH1* deletion mutants have a longer growth phase on ethanol, lower maximum growth rate (Table 5.6) and excrete acetate in the early growth period (this work and [3]).

The reason for the existence of two options for activating acetate to acetyl-CoA may be related to the limitation of acetyl-CoA synthetase to support very high fluxes. As shown in Figure 5.3, Ach1p (SCAT) is responsible for activating 46% of the acetate available, which is then used for synthesizing citrate and fueling the citric acid cycle. If *ACH1* is deleted the overall flux through ACS (cytosolic + peroxisomal) almost doubles, and more citrate has to be produced by the peroxisomal citrate synthase. Since the additional citrate has to be transported across two membranes (peroxisomal and mitochondrial) in order to be metabolized in the mitochondria, it might also act as a metabolic bottleneck in *ACH1* deletion mutants.

Fleck and Brock [3] suggested that the growth retardation phenotype of the *ACH1* deletion strain on acetate was caused by acetate related toxicity in the mitochondria. However, it has been reported by Chan et al. [59] that if the flux through ACS is artificially up-regulated in *S. enterica* growing on acetate, this results in slow growth caused by ATP/ADP depletion and AMP accumulation. This phenotype can be rescued by expressing an ADP-forming acetyl-CoA synthetase because it reduces substantially the energetic burden of acetate consumption. Judging by the flux changes observed in Figure 5.3 for the *ACH1* deletion mutant it is very likely that a higher flux in the cytosolic/peroxisomal ACS also results in ATP depletion and accumulation of AMP in *S. cerevisiae*. Furthermore, when Ald4cp was over-expressed in HPR07, we observed additional growth retardation (Table 5.6) when compared to an isogenic strain expressing the less active Ald2p (HPR06). This phenotype is consistent with the hypothesis that higher flux through ACS leads to the depletion of ATP and an energetic imbalance in the cell.

Putting all the evidences together, Ach1p seems to act on two fronts: to help the cell metabolizing a considerable part of the acetate and preventing excessive flux through ACS which might cause an energy imbalance. Further clarification of its role will need additional experiments similar to the ones mentioned above for *S. enterica* [59].

5.4 Conclusions

In this study, we used *in silico* metabolic modeling to discover new strain engineering targets that would improve acetyl-CoA availability in the cytosol and increase the production levels of 3-HP. Although the yields obtained experimentally are considerably lower than the ones suggested in the simulations, we observed a positive effect of PDH down-regulation and *ACH1* deletion on the 3-HP yield in batch cultivations (Figure 5.2). Overall, the results would benefit from additional studies regarding the regulation of the acetyl-CoA and malonyl-CoA nodes in order to improve the flux through the PDH bypass and acetyl-CoA carboxylase.

We also observed that Ach1p might have a more important role in the metabolism of C₂ carbon sources than previously acknowledged (Figure 5.3). In addition to the increase in 3-HP production, the inactivation of *ACH1* also caused a considerable growth slowdown in the ethanol consumption phase of the batch fermentation. The reason for such a severe phenotype is not fully understood and further studies are required to clarify the exact mechanism responsible for the growth retardation. However, FBA analysis of the flux readjustment needed in an Ach1p negative *S. cerevisiae* indicates that a higher flux through ACS may cause ATP depletion and consequently an energetic imbalance. A similar response to a high flux through ACS has been described in *S. enterica* and supports this theory.

References

1. Nielsen J (2014) Synthetic biology for engineering acetyl coenzyme a metabolism in yeast. *MBio* 5:e02153–14–.
2. Chen Y, Siewers V, Nielsen J (2012) Profiling of cytosolic and peroxisomal acetyl-CoA metabolism in *Saccharomyces cerevisiae*. *PLoS One* 7:e42475.
3. Fleck CB, Brock M (2009) Re-characterisation of *Saccharomyces cerevisiae* Ach1p: fungal CoA-transferases are involved in acetic acid detoxification. *Fungal Genet Biol* 46:473–85.
4. Lee FJ, Lin LW, Smith JA (1990) A glucose-repressible gene encodes acetyl-CoA hydrolase from *Saccharomyces cerevisiae*. *J Biol Chem* 265:7413–8.

5. Flikweert MT, de Swaaf M, van Dijken JP, Pronk JT (1999) Growth requirements of pyruvate-decarboxylase-negative *Saccharomyces cerevisiae*. FEMS Microbiol Lett 174:73–9.
6. Berg MA, Steensma HY (1995) ACS2, a *Saccharomyces cerevisiae* Gene Encoding Acetyl-Coenzyme A Synthetase, Essential for Growth on Glucose. Eur J Biochem 231:704–713.
7. Flikweert MT, Van Der Zanden L, Janssen WM, et al. (1996) Pyruvate decarboxylase: an indispensable enzyme for growth of *Saccharomyces cerevisiae* on glucose. Yeast 12:247–57.
8. Van Roermund CW, Hettema EH, van den Berg M, et al. (1999) Molecular characterization of carnitine-dependent transport of acetyl-CoA from peroxisomes to mitochondria in *Saccharomyces cerevisiae* and identification of a plasma membrane carnitine transporter, Agp2p. EMBO J 18:5843–52.
9. Shiba Y, Paradise EM, Kirby J, et al. (2007) Engineering of the pyruvate dehydrogenase bypass in *Saccharomyces cerevisiae* for high-level production of isoprenoids. Metab Eng 9:160–8.
10. Starai VJ, Gardner JG, Escalante-Semerena JC (2005) Residue Leu-641 of Acetyl-CoA synthetase is critical for the acetylation of residue Lys-609 by the Protein acetyltransferase enzyme of *Salmonella enterica*. J Biol Chem 280:26200–5.
11. Kocharin K, Chen Y, Siewers V, Nielsen J (2012) Engineering of acetyl-CoA metabolism for the improved production of polyhydroxybutyrate in *Saccharomyces cerevisiae*. AMB Express 2:52.
12. Chen Y, Daviet L, Schalk M, et al. (2013) Establishing a platform cell factory through engineering of yeast acetyl-CoA metabolism. Metab Eng 15:48–54.
13. Krivoruchko A, Serrano-Amatriain C, Chen Y, et al. (2013) Improving biobutanol production in engineered *Saccharomyces cerevisiae* by manipulation of acetyl-CoA metabolism. J Ind Microbiol Biotechnol 40:1051–6.
14. Chen Y, Bao J, Kim I-K, et al. (2014) Coupled incremental precursor and co-factor supply improves 3-hydroxypropionic acid production in *Saccharomyces cerevisiae*. Metab Eng 22:104–9.
15. Kozak BU, van Rossum HM, Benjamin KR, et al. (2014) Replacement of the *Saccharomyces cerevisiae* acetyl-CoA synthetases by alternative pathways for cytosolic acetyl-CoA synthesis. Metab Eng 21:46–59.
16. Kozak BU, van Rossum HM, Luttik MAH, et al. (2014) Engineering acetyl coenzyme A supply: functional expression of a bacterial pyruvate dehydrogenase complex in the cytosol of *Saccharomyces cerevisiae*. MBio 5:e01696–14.
17. Lian J, Si T, Nair NU, Zhao H (2014) Design and construction of acetyl-CoA overproducing *Saccharomyces cerevisiae* strains. Metab Eng 24:139–49.

18. Sonderegger M, Schümperli M, Sauer U (2004) Metabolic engineering of a phosphoketolase pathway for pentose catabolism in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 70:2892–7.
19. De Jong BW, Shi S, Siewers V, Nielsen J (2014) Improved production of fatty acid ethyl esters in *Saccharomyces cerevisiae* through up-regulation of the ethanol degradation pathway and expression of the heterologous phosphoketolase pathway. *Microb Cell Fact* 13:39.
20. Milne CB, Kim P-J, Eddy JA, Price ND (2009) Accomplishments in genome-scale *in silico* modeling for industrial and medical biotechnology. *Biotechnol J* 4:1653–70.
21. Garcia-Albornoz MA, Nielsen J (2013) Application of Genome-Scale Metabolic Models in Metabolic Engineering. *Ind Biotechnol* 9:203–214.
22. Asadollahi MA, Maury J, Patil KR, et al. (2009) Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through *in silico* driven metabolic engineering. *Metab Eng* 11:328–34.
23. Brochado AR, Matos C, Møller BL, et al. (2010) Improved vanillin production in baker's yeast through *in silico* design. *Microb Cell Fact* 9:84.
24. Heavner BD, Smallbone K, Price ND, Walker LP (2013) Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database (Oxford)* 2013:bat059.
25. Shi S, Chen Y, Siewers V, Nielsen J (2014) Improving production of malonyl coenzyme A-derived metabolites by abolishing Snf1-dependent regulation of Acc1. *MBio* 5:e01130–14.
26. Rocha I, Maia P, Evangelista P, et al. (2010) OptFlux: an open-source software platform for *in silico* metabolic engineering. *BMC Syst Biol* 4:45.
27. Patil KR, Rocha I, Förster J, Nielsen J (2005) Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics* 6:308.
28. Lewis NE, Hixson KK, Conrad TM, et al. (2010) Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* 6:390.
29. Pronk JT, Wenzel TJ, Luttik MA, et al. (1994) Energetic aspects of glucose metabolism in a pyruvate-dehydrogenase-negative mutant of *Saccharomyces cerevisiae*. *Microbiology* 140 (Pt 3):601–10.
30. Lööke M, Kristjuhan K, Kristjuhan A (2011) Extraction of genomic DNA from yeasts for PCR-based applications. *Biotechniques* 50:325–8.
31. Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* 96:23–8.

32. Sambrook J, Russell DW (2001) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
33. Gietz RD, Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* 350:87–96.
34. Orr-Weaver TL, Szostak JW (1983) Yeast recombination: the association between double-strand gap repair and crossing-over. *Proc Natl Acad Sci U S A* 80:4417–21.
35. Oldenburg K (1997) Recombination-mediated PCR-directed plasmid construction *in vivo* in yeast. *Nucleic Acids Res* 25:451–452.
36. Bessa D, Pereira F, Moreira R, et al. (2012) Improved gap repair cloning in yeast: treatment of the gapped vector with Taq DNA polymerase avoids vector self-ligation. *Yeast* 29:419–23.
37. Stevenson BJ, Liu J-W, Ollis DL (2008) Directed evolution of yeast pyruvate decarboxylase 1 for attenuated regulation and increased stability. *Biochemistry* 47:3013–25.
38. Mikkelsen MD, Buron LD, Salomonsen B, et al. (2012) Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab Eng* 14:104–11.
39. Güldener U, Heck S, Fielder T, et al. (1996) A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res* 24:2519–24.
40. Erdeniz N, Mortensen UH, Rothstein R (1997) Cloning-Free PCR-Based Allele Replacement Methods. *Genome Res* 7:1174–1183.
41. Verduyn C, Postma E, Scheffers WA, Van Dijken JP (1992) Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 8:501–17.
42. Mukhopadhyay A, Wei B, Weiner H (2013) Mitochondrial NAD dependent aldehyde dehydrogenase either from yeast or human replaces yeast cytoplasmic NADP dependent aldehyde dehydrogenase for the aerobic growth of yeast on ethanol. *Biochim Biophys Acta* 1830:3391–8.
43. Gey U, Czupalla C, Hoflack B, et al. (2008) Yeast pyruvate dehydrogenase complex is regulated by a concerted activity of two kinases and two phosphatases. *J Biol Chem* 283:9759–67.
44. Krause-Buchholz U, Gey U, Wünschmann J, et al. (2006) YIL042c and YOR090c encode the kinase and phosphatase of the *Saccharomyces cerevisiae* pyruvate dehydrogenase complex. *FEBS Lett* 580:2553–60.
45. Postma E, Verduyn C, Scheffers WA, Van Dijken JP (1989) Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 55:468–77.

46. Van Hoek P, Flikweert MT, van der Aart QJ, et al. (1998) Effects of pyruvate decarboxylase overproduction on flux distribution at the pyruvate branch point in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 64:2133–40.
47. Schaaff I, Heinisch J, Zimmermann FK Overproduction of glycolytic enzymes in yeast. *Yeast* 5:285–90.
48. Wang X, Mann CJ, Bai Y, et al. (1998) Molecular Cloning, Characterization, and Potential Roles of Cytosolic and Mitochondrial Aldehyde Dehydrogenases in Ethanol Metabolism in *Saccharomyces cerevisiae*. *J Bacteriol* 180:822–830.
49. Navarro-Aviño JP, Prasad R, Miralles VJ, et al. (1999) A proposal for nomenclature of aldehyde dehydrogenases in *Saccharomyces cerevisiae* and characterization of the stress-inducible *ALD2* and *ALD3* genes. *Yeast* 15:829–42.
50. Gombert AK, Moreira dos Santos M, Christensen B, Nielsen J (2001) Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. *J Bacteriol* 183:1441–51.
51. Dijken JP, Scheffers WA (1986) Redox balances in the metabolism of sugars by yeasts. *FEMS Microbiol Lett* 32:199–224.
52. Leonard E, Lim K-H, Saw P-N, Koffas M a G (2007) Engineering central metabolic pathways for high-level flavonoid production in *Escherichia coli*. *Appl Environ Microbiol* 73:3877–86.
53. WENZEL TJ, LUTTIK MAH, BERG JA, STEENSMA HY (1993) Regulation of the *PDA1* gene encoding the E1alpha subunit of the pyruvate dehydrogenase complex from *Saccharomyces cerevisiae*. *Eur J Biochem* 218:405–411.
54. Oliveira AP, Ludwig C, Picotti P, et al. (2012) Regulation of yeast central metabolism by enzyme phosphorylation. *Mol Syst Biol* 8:623.
55. Van den Berg MA, de Jong-Gubbels P, Kortland CJ, et al. (1996) The Two Acetyl-coenzyme A Synthetases of *Saccharomyces cerevisiae* Differ with Respect to Kinetic Properties and Transcriptional Regulation. *J Biol Chem* 271:28953–28959.
56. De Jong-Gubbels P, van den Berg MA, Steensma HY, et al. (1997) The *Saccharomyces cerevisiae* acetyl-coenzyme A synthetase encoded by the *ACS1* gene, but not the *ACS2*-encoded enzyme, is subject to glucose catabolite inactivation. *FEMS Microbiol Lett* 153:75–81.
57. Fell DA, Small JR (1986) Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem J* 238:781–6.
58. Varma A, Palsson BO (1994) Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/Technology* 12:994–998.
59. Chan CH, Garrity J, Crosby HA, Escalante-Semerena JC (2011) In *Salmonella enterica*, the sirtuin-dependent protein acylation/deacylation system (SDPADS)

maintains energy homeostasis during growth on low concentrations of acetate. *Mol Microbiol* 80:168–83.

CHAPTER 6

General conclusions

This final chapter includes the general conclusions about the work that is included in this thesis, as well as additional suggestions of future work to be carried out to answer some questions that were raised by the results obtained here.

6.1 General conclusions

The main goal of this thesis was to improve the methodologies used for *in silico* strain engineering by using *Saccharomyces cerevisiae* as the case study organism. The main conclusions resulting from the work developed here are as follows:

- A review of the literature, regarding the availability of genome-scale modeling methods that can be used for strain engineering, revealed that there is a considerable variety and quantity to choose from. However, the abundance of methods is somehow contradictory with the number of experimental implementations of strain engineering strategies obtained using genome-scale metabolic models in *S. cerevisiae*.
- An analysis of some of the genome-scale metabolic models available for *S. cerevisiae* has shown that the internal flux distributions predicted with parsimonious flux balanced analysis (pFBA) under fully aerobic conditions included some inconsistencies in central areas of the metabolism. Using cofactor abundance in aerobic conditions as guidance, constraints were applied to all the metabolic reactions including NADH of NADPH, resulting in several improvements in the pentose phosphate pathway and other parts of NADPH metabolism. Those improvements were shown to have a positive impact on the simulation of gene knock-outs obtained for the production of acetate and mevalonate.
- The lack of a simulation method that could predict in quantitative terms the phenotype of strains with complex engineered genotypes, led us to develop a novel simulation method called turnover dependent phenotypic simulation (TDPS). This method was designed with the goal of simulating the majority of the genetic modifications usually implemented in engineered strains and its formulation takes into account the availability of resources in the network by assuming that the production turnover of a metabolite can be used as an indication of its abundance.

- TDPS was validated quantitatively using metabolically engineered *S. cerevisiae* strains available in the literature by comparing the production yields of the target metabolite between the simulations and experiments. Overall, the estimated yields obtained with TDPS were quite close to the experimentally reported values, although the performance oscillated among the different case-studies tested. To the best of our knowledge, this was the first time that a simulation method was validated in quantitative terms using experimental data.
- TDPS was used in combination with the manually curated models in order to search for new strain engineering targets that would improve acetyl-CoA availability in the cytosol and increase the production levels of 3-hydroxypropionic acid (3-HP). Although the yields obtained experimentally were considerably lower than the ones suggested in the simulations, we observed a positive effect of the pyruvate dehydrogenase down-regulation and *ACH1* (succinyl-CoA:acetate CoA-transferase) deletion on the 3-HP yield in batch cultivations.
- It was also observed that Ach1p might have a more important role in the metabolism of C₂ carbon sources than previously acknowledged because its inactivation caused a considerable growth slowdown in the ethanol consumption phase of the batch fermentation. A modified genome-scale model of *S. cerevisiae* was used to predict how the deletion of *ACH1* affects the internal fluxes of cells growing on ethanol and some hypotheses were formulated to explain the phenotypes observed.

6.2 Recommended future work

As it is usually the case, most research projects lead to more questions than answers. Although this thesis provided relevant insights into the field of *in silico* strain engineering, a few questions remain unanswered and consequently we provide a few topics where additional research would be recommendable:

- The results obtained with TDPS were very promising but could still be improved in our opinion by using an objective function that better describes the cellular

adaptation when faced with serious disturbances. Therefore, further efforts should be made to test additional objective functions. In addition, it would also be recommendable to expand the case-studies used to validate TDPS to other organisms, in order to help fine-tuning the performance of the algorithm.

- The experimental results obtained for the engineered 3-HP producing strains would benefit from additional studies regarding the regulation of the acetyl-CoA and malonyl-CoA nodes in order to improve the flux through the pyruvate dehydrogenase bypass and acetyl-CoA carboxylase. Therefore, the corresponding enzymes should be further investigated for limited availability of cofactors and the existence of regulatory phenomena.
- The reason for the severe phenotype observed for the *ACHI* deletion in *S. cerevisiae* during the ethanol phase of a batch culture was not fully understood. Therefore further studies are required to clarify the exact mechanism responsible for the growth retardation. A good starting point would be to verify if the higher flux through acetyl-CoA synthetase (as suggested by FBA simulations) is creating an energetic imbalance.
- Although the strains engineered for improved availability of acetyl-CoA were designed for glucose limited conditions, the experiments reported in this thesis only encompassed batch cultivations. Therefore, in order to understand better the physiology of those strains, it would be interesting to perform chemostat experiments and compare the results obtained with the simulated values.
- As a final remark, it is important to point out that in order to increase the use and acceptance of *in silico* strain engineering methodologies in the metabolic engineering field, it is necessary to carefully analyze the existing methodologies and improve them to the point that researchers will trust simulations over their rationally derived hypothesis. To do so, further experimental studies should be carried out to validate strain engineering computational methods, in order to reveal their advantages and disadvantages.

APPENDIX

Appendix A: Manual curation of the genome-scale models of *S. cerevisiae*

Table A1- List of reactions modified during the curation process of the model *iFF708*

Reaction ID	Stoichiometric equation	Genes	Comment	verdict
MDH3	Malate + NAD+ \rightleftharpoons NADH + Oxaloacetate	YDL078C	part of the glyoxylate cycle (not needed in glucose growth)	reversibility constrained
MDH2	Malate + NAD+ \rightleftharpoons NADH + Oxaloacetate	YOL126C	part of the glyoxylate cycle (not needed in glucose growth)	reversibility constrained
U45_	D-Mannitol_1-phosphate + NAD+ \rightleftharpoons beta-D-Fructose_6-phosphate + NADH	NA (Mannitol-1-phosphate 5-dehydrogenase)	dead end (D-Mannitol_1-phosphate only present in this reaction)	inactivated
ADH2	Ethanol + NAD+ \rightleftharpoons Acetaldehyde + NADH	YMR303C	not active with a dilution rate of 0.10 h ⁻¹	reversibility constrained
ADH1	Ethanol + NAD+ \rightleftharpoons Acetaldehyde + NADH	YOL086C	not active with a dilution rate of 0.10 h ⁻¹	reversibility constrained
PRO3_3	L-1-Pyrroline-3-hydroxy-5-carboxylate + NADH \Rightarrow trans-4-Hydroxy-L-proline + NAD+	YER023W	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
ADH5	Ethanol + NAD+ \rightleftharpoons Acetaldehyde + NADH	YBR145W	not active with a dilution rate of 0.10 h ⁻¹	reversibility constrained
ADH4	Ethanol + NAD+ \rightleftharpoons Acetaldehyde + NADH	YGL256W	not active with a dilution rate of 0.10 h ⁻¹	reversibility constrained
FAS1_4	Acyl-[acyl-carrier_protein] + NAD+ \rightleftharpoons 2,3-Dehydroacyl-[acyl-carrier_protein] + NADH	YKL182W	dead end (general compound present in the reaction: acyl-CoA)	inactivated
BIO2	Dethiobiotin + 2 NAD+ + Hydrogen_sulfide + ATP \rightleftharpoons Biotin + 2 NADH + AMP + Pyrophosphate	YGR286C	reaction is coupled to ATP hydrolysis	reversibility constrained
SFA1_2	Ethanol + NAD+ \rightleftharpoons Acetaldehyde + NADH	YDL168W	not active with a dilution rate of 0.10 h ⁻¹	reversibility constrained
SFA1_1	Formaldehyde + Glutathione + NAD+ \rightleftharpoons S-Formylglutathione + NADH	YDL168W	not active with a dilution rate of 0.10 h ⁻¹ (formaldehyde is not produced in these conditions)	reversibility constrained
LYS1	N6-(L-1,3-Dicarboxypropyl)-L-lysine + NAD+ \rightleftharpoons L-Lysine + 2-Oxoglutarate + NADH	YIR034C	used in the direction of L-lysine synthesis (NADH producing)	reversibility constrained

Table A1- List of reactions modified during the curation process of the model *iFF708* (continuation)

Reaction ID	Stoichiometric equation	Genes	Comment	Verdict
TDH1	D-Glyceraldehyde_3-phosphate + Orthophosphate + NAD+ \rightleftharpoons NADH + 3-Phospho-D-glyceroyl_phosphate	YJL052W	part of glycolysis (NADH producing)	reversibility constrained
TDH2	D-Glyceraldehyde_3-phosphate + Orthophosphate + NAD+ \rightleftharpoons NADH + 3-Phospho-D-glyceroyl_phosphate	YJR009C	part of glycolysis (NADH producing)	reversibility constrained
TDH3	D-Glyceraldehyde_3-phosphate + Orthophosphate + NAD+ \rightleftharpoons NADH + 3-Phospho-D-glyceroyl_phosphate	YGR192C	part of glycolysis (NADH producing)	reversibility constrained
FOX2	(3S)-3-Hydroxyacyl-CoA + NAD+ \rightleftharpoons 3-Oxoacyl-CoA + NADH	YKR009C	dead end (general compound present in the reaction: 3-oxoacyl-CoAs)	inactivated
GPD2	Glycerone_phosphate + NADH \Rightarrow sn-Glycerol_3-phosphate + NAD+	YOL059W	thermodynamically favorable in the forward direction (dG=-31.5 KJ/mol)*	no changes applied
GPD1	Glycerone_phosphate + NADH \Rightarrow sn-Glycerol_3-phosphate + NAD+	YDL022W	thermodynamically favorable in the forward direction (dG=-31.5 KJ/mol)*	no changes applied
GLT1	2-Oxoglutarate + L-Glutamine + NADH \Rightarrow NAD+ + 2 L-Glutamate	YDL171C	thermodynamically favorable in the forward direction (dG=-40 KJ/mol)*	no changes applied
PRO2_1	alpha-D-Glutamyl_phosphate + NADH \Rightarrow NAD+ + Orthophosphate + L-Glutamate_5-semialdehyde	YOR323C	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
NDE1	NADH + Ubiquinone-9M \Rightarrow UbiquinolM + NAD+	YMR145C	Important role in the oxidation of cytosolic NADH (part of the oxidative phosphorylation)	no changes applied
NDE2	NADH + Ubiquinone-9M \Rightarrow UbiquinolM + NAD+	YDL085W	Important role in the oxidation of cytosolic NADH (part of the oxidative phosphorylation)	no changes applied
LYS2_2	L-2-Amino adipate + NADH + ATP \Rightarrow L-2-Amino adipate_6-semialdehyde + NAD+ + AMP + Pyrophosphate	YBR115C	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
HOM6_1	L-Aspartate_4-semialdehyde + NADH \Rightarrow NAD+ + L-Homoserine	YJR139C	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
TYR1	Prephenate + NADP+ \Rightarrow 3-(4-Hydroxyphenyl)pyruvate + CO2 + NADPH	YBR166C	capable of using NADH or NADPH in the model, more likely to use NADH	inactivated
IDP3_1	Isocitrate + NADP+ \Rightarrow NADPH + Oxalosuccinate	YNL009W	cytosolic NADP ⁺ -dependent isocitrate dehydrogenase [1]	inactivated
ARA1_2	D-Arabinose + NADP+ \Rightarrow D-Arabinono-1,4-lactone + NADPH	YBR149W	capable of using NADH or NADPH in the model, more likely to use NADH	inactivated
HMG2	(R)-Mevalonate + CoA + 2 NADP+ \rightleftharpoons (S)-3-Hydroxy-3-methylglutaryl-CoA + 2 NADPH	YLR450W	Hydrolysis of CoA and consumption of 2 NADPH; thermodynamically favorable in the forward direction (dG=-18.6 kJ / mol)*	reversibility constrained

Table A1- List of reactions modified during the curation process of the model *iFF708* (continuation)

Reaction ID	Stoichiometric equation	Genes	Comment	Verdict
HMG1	(R)-Mevalonate + CoA + 2 NADP+ \rightleftharpoons (S)-3-Hydroxy-3-methylglutaryl-CoA + 2 NADPH	YML075C	Hydrolysis of CoA and consumption of 2 NADPH; thermodynamically favorable in the forward direction ($dG=-18.6$ kJ / mol)*	reversibility constrained
ZWF1	alpha-D-Glucose_6-phosphate + NADP+ \rightleftharpoons D-Glucono-1,5-lactone_6-phosphate + NADPH	YNL241C	Oxidative pentose phosphate pathway	no changes applied
ERG1	Squalene + Oxygen + NADP+ \Rightarrow (S)-2,3-Epoxy-squalene + NADPH	YGR175C	wrong stoichiometry (see text)	stoichiometry corrected
ALD6	Acetaldehyde + NADP+ \Rightarrow NADPH + Acetate	YPL061W	capable of using NADH or NADPH in the model, more likely to use NADH	inactivated
GND2	6-Phospho-D-gluconate + NADP+ \Rightarrow NADPH + CO ₂ + D-Ribulose_5-phosphate	YGR256W	Oxidative pentose phosphate pathway	no changes applied
GND1	6-Phospho-D-gluconate + NADP+ \Rightarrow NADPH + CO ₂ + D-Ribulose_5-phosphate	YHR183W	Oxidative pentose phosphate pathway	no changes applied
YBR006WU GA2	Succinate_semialdehyde + NADP+ \Rightarrow Succinate + NADPH	YBR006W	thermodynamically favorable in the forward direction ($dG=-63$ KJ/mol)*	no changes applied
ADE3_1	5,10-Methylenetetrahydrofolate + NADP+ \rightleftharpoons 5,10-Methenyltetrahydrofolate + NADPH	YGR204W	inconclusive data regarding the reversibility of the reaction	no changes applied
ECM17	Sulfite + 3 NADPH \rightleftharpoons Hydrogen_sulfide + 3 NADP+	YJR137C	thermodynamically favorable in the reverse direction ($dG=103$ KJ/mol)*	reversibility constrained
IDP2_1	Isocitrate + NADP+ \Rightarrow NADPH + Oxalosuccinate	YLR174W	cytosolic NADP ⁺ -dependent isocitrate dehydrogenase [1]	inactivated
MET10	Sulfite + 3 NADPH \rightleftharpoons Hydrogen_sulfide + 3 NADP+	YFR030W	thermodynamically favorable in the reverse direction ($dG=103$ KJ/mol)*	reversibility constrained
LYS9	L-Glutamate + L-2-Amino adipate_6-semialdehyde + NADPH \rightleftharpoons N6-(L-1,3-Dicarboxypropyl)-L-lysine + NADP+	YNR050C	used in the direction of L-lysine synthesis (NADPH consuming)	reversibility constrained

* The change in Gibbs free energy was calculated using the website: <http://equilibrator.weizmann.ac.il/>. Standard conditions were used for every metabolite except NADPH, NADP+, NADH and NAD+. For these cofactors the concentrations were obtained from references [1] and [2]: [NADPH]= 263 μ M; [NADP+]= 35.5 μ M; [NADH]= 120 μ M; [NAD+]= 1810 μ M

Table A2- List of reactions modified during the curation process of the model *iMM904*

Reaction ID	Stoichiometric equation (model IDs)	Genes	Comment	Verdict
BTDD-RR	[c] : btd-RR + nad \rightleftharpoons actn-R + h + nadh	YAL060W	not active with a dilution rate of 0.10 h ⁻¹ (acetoin is not produced in these conditions)	reversibility constrained
FALDH	[c] : fald + gthrd + nad \rightleftharpoons Sfglutth + h + nadh	YDL168W	not active with a dilution rate of 0.10 h ⁻¹ (formaldehyde is not produced in these conditions)	reversibility constrained
2HBO	[c] : 2hb + nad \rightleftharpoons 2obut + h + nadh	NA	not active with a dilution rate of 0.10 h ⁻¹ (2-Hydroxybutyrate is not produced in these conditions)	reversibility constrained
ALCD2x	[c] : etoh + nad \rightleftharpoons acald + h + nadh	YDL168W	not active with a dilution rate of 0.10 h ⁻¹ (ethanol is not produced in these conditions)	reversibility constrained
SACCD2	[c] : h2o + nad + saccrp-L \rightleftharpoons akg + h + lys-L + nadh	YIR034C	used in the direction of L-lysine synthesis (NADH producing)	reversibility constrained
MDH	[c] : mal-L + nad \rightleftharpoons h + nadh + oaa	YOL126C	part of the glyoxylate cycle (not needed in glucose growth)	reversibility constrained
GAPD	[c] : g3p + nad + pi \rightleftharpoons 13dpg + h + nadh	YJL052W or YJR009C or YGR192C	part of Glycolysis (NADH producing)	reversibility constrained
AASAD2	[c] : L2aadp + atp + h + nadh \rightarrow L2aadp6sa + amp + nad + ppi	YBR115C and YGL154C	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
G5SD2	[c] : glu5p + h + nadh \rightarrow glu5sa + nad + pi	YOR323C	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
G3PD1ir	[c] : dhap + h + nadh \rightarrow glyc3p + nad	YDL022W	thermodynamically favorable in the forward direction (dG=-31.5 KJ/mol)*	no changes applied
HPROa	[c] : 1p3h5c + (2) h + nadh \rightarrow 4hpro-LT + nad	YER023W	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
ALCD25xi	[c] : h + nadh + pacald \rightarrow 2phetoh + nad	YBR145W, YGL256W, YDL168W, YMR303C, YOL086C	not active with a dilution rate of 0.10 h ⁻¹ (2-phenylethanol is not produced in these conditions)	inactivated
ALCD2ir	[c] : acald + h + nadh \rightarrow etoh + nad	YOL086C or YGL256W or YBR145W	not active with a dilution rate of 0.10 h ⁻¹ (ethanol is not produced in these conditions)	inactivated

Table A2- List of reactions modified during the curation process of the model *iMM904* (continuation)

Reaction ID	Stoichiometric equation (model IDs)	Gene	Comment	Verdict
ALCD26xi	[c] : h + id3acald + nadh --> ind3eth + nad	YGL256W or YDL168W or YMR303C or YBR145W or YOL086C	not active with a dilution rate of 0.10 h ⁻¹ (Indole-3-ethanol is not produced in these conditions)	inactivated
ALCD22xi	[c] : 2mbald + h + nadh --> 2mbtoh + nad	YBR145W or YMR303C or YDL168W or YGL256W or YOL086C	not active with a dilution rate of 0.10 h ⁻¹ (2-methyl-1-butanol is not produced in these conditions)	inactivated
FMNRx	[c] : fmn + h + nadh --> fmnh2 + nad	YLR011W	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
ALCD23xi	[c] : 2mpal + h + nadh --> ibutoh + nad	YGL256W or YBR145W or YDL168W or YOL086C or YMR303C	not active with a dilution rate of 0.10 h ⁻¹ (isobutyl-alcohol is not produced in these conditions)	inactivated
C22STDSx	[c] : ergtrol + h + nadh + o2 --> ergtetrol + (2) h2o + nad	(((YIL043C and YNL111C) and YMR015C) or ((YNL111C and YMR015C) and YKL150W))	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
LNS14DMx	[c] : (2) h + lanost + (3) nadh + (3) o2 --> 44mctr + for + (4) h2o + (3) nad	(((YNL111C and YHR007C) and YKL150W) or ((YIL043C and YNL111C) and YHR007C))	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
HSDxi	[c] : aspsa + h + nadh --> hom-L + nad	YJR139C	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
GLUSx	[c] : akg + gln-L + h + nadh --> (2) glu-L + nad	YDL171C	thermodynamically favorable in the forward direction (dG=-40 KJ/mol)*	no changes applied
ALCD24xi	[c] : 3mbald + h + nadh --> iamoh + nad	((((YDL168W or YGL256W) or YMR303C) or YOL086C) or YBR145W)	not active with a dilution rate of 0.10 h ⁻¹ (Isoamyl-alcohol is not produced in these conditions)	inactivated
NADH2-u6cm	h[c] + nadh[c] + q6[m] --> nad[c] + q6h2[m]	YMR145C or YDL085W	Important role in the oxidation of Cytosolic NADH (part of the oxidative phosphorylation)	no changes applied
MTHFD	[c] : mlthf + nadp <=> methf + nadph	YGR204W	inconclusive data regarding the reversibility of the reaction	no changes applied
HMGCOAR	[c] : coa + mev-R + (2) nadp <=> (2) h + hmgcoa + (2) nadph	YLR450W or YML075C	Hydrolysis of CoA and consumption of 2 NADPH; thermodynamically favorable in the forward direction (dG=-18.6 kJ / mol)*	reversibility constrained

Table A2- List of reactions modified during the curation process of the model *iMM904* (continuation)

Reaction ID	Stoichiometric equation (model IDs)	Gene	Comment	Verdict
SACCD1	[c] : L2aadp6sa + glu-L + h + nadph <=> h2o + nadp + sacrp-L	YNR050C	used in the direction of L-lysine synthesis (NADPH consuming)	reversibility constrained
LSERDHR	[c] : nadp + ser-L <=> 2amsa + h + nadph	YMR226C	dead end (2-aminomalonate semialdehyde only present in this reaction)	inactivated
SULR	[c] : (3) h2o + h2s + (3) nadp <=> (5) h + (3) nadph + so3	YFR030W or YJR137C	thermodynamically favorable in the reverse direction (dG=103 KJ/mol)*	reversibility constrained
ATHRDHR	[c] : athr-L + nadp <=> 2aobut + h + nadph	YMR226C	dead end	inactivated
GND	[c] : 6pgc + nadp --> co2 + nadph + ru5p-D	(YGR256W or YHR183W)	Oxidative pentose phosphate pathway	no changes applied
G6PDH2	[c] : g6p + nadp --> 6pgl + h + nadph	YNL241C	Oxidative pentose phosphate pathway	no changes applied
PPND2	[c] : nadp + pphn --> 34hpp + co2 + nadph	YBR166C	capable of using NADH or NADPH in the model, more likely to use NADH	inactivated
SSALy	[c] : h2o + nadp + succsal --> (2) h + nadph + succ	YBR006W	thermodynamically favorable in the forward direction (dG=-63 KJ/mol)*	no changes applied
GLYCDy	[c] : glyc + nadp --> dha + h + nadph	YOR120W	thermodynamically favorable in the reverse direction (dG=25 KJ/mol)*	inactivated
SHCHD	[c] : dscl + nadp --> h + nadph + scl	YBR213W	enzyme is reported in the databases (KEGG and SGD) as NAD dependent: EC 1.3.1.76	stoichiometry corrected
ARAB1D2	[c] : arab-D + nadp --> Dara14lac + h + nadph	YBR149W	only active is Arabinose is supplied (dead end)	inactivated
ICDHy	[c] : icit + nadp --> akc + co2 + nadph	YLR174W	cytosolic NADP ⁺ -dependent isocitrate dehydrogenase [1]	inactivated
ALDD20y	[c] : h2o + id3acald + nadp --> (2) h + ind3ac + nadph	YPL061W	dead end (indole-3-acetate cannot be metabolized)	inactivated
ALDD2y	[c] : acald + h2o + nadp --> ac + (2) h + nadph	YPL061W	other models have NAD-dependent alternative- see text	inactivated and new reaction was added: R_ALDD2x

* The change in Gibbs free energy was calculated using the website: <http://equilibrator.weizmann.ac.il/>. Standard conditions were used for every metabolite except NADPH, NADP⁺, NADH and NAD⁺. For these cofactors the concentrations were obtained from references [1] and [2]: [NADPH]= 263 μM; [NADP⁺]= 35.5 μM; [NADH]= 120 μM; [NAD⁺]= 1810 μM

Table A3- List of reactions modified during the curation process of the model *iTO977*

Reaction ID	Stoichiometric equations	Gene	Comment	Verdict
GLT1	2-oxoglutarate[Cytosol] + L-glutamine[Cytosol] + NADH[Cytosol] => 2 L-glutamate[Cytosol] + NAD(+)[Cytosol]	YDL171C	thermodynamically favorable in the forward direction (dG=-40 KJ/mol)*	no changes applied
BIO2	ATP[Cytosol] + H2S[Cytosol] + 2 NAD(+)[Cytosol] + dethiobiotin[Cytosol] <=> AMP[Cytosol] + 2 NADH[Cytosol] + biotin[Cytosol] + diphosphate[Cytosol]	YGR286C	reaction is coupled to ATP hydrolysis	reversibility constrained
LYS2_2	ATP[Cytosol] + L-2-aminoadipate[Cytosol] + NADH[Cytosol] => AMP[Cytosol] + L-2-aminoadipate 6-semialdehyde[Cytosol] + NAD(+)[Cytosol] + diphosphate[Cytosol]	YBR115C:YGL154C	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
PRO2_1	D-alpha-glutamyl phosphate[Cytosol] + NADH[Cytosol] => L-glutamate 5-semialdehyde[Cytosol] + NAD(+)[Cytosol] + phosphate[Cytosol]	YOR323C	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
U45_	D-mannitol 1-phosphate[Cytosol] + NAD(+)[Cytosol] <=> NADH[Cytosol] + beta-D-fructofuranose 6-phosphate[Cytosol]	NA (Mannitol-1-phosphate 5-dehydrogenase)	dead end (D-Mannitol_1-phosphate only present in this reaction)	inactivated
PRO3_3	L-1-pyrroline-3-hydroxy-5-carboxylate[Cytosol] + NADH[Cytosol] => NAD(+)[Cytosol] + trans-4-hydroxy-L-proline[Cytosol]	YER023W	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
HOM6_1	L-aspartate 4-semialdehyde[Cytosol] + NADH[Cytosol] => L-homoserine[Cytosol] + NAD(+)[Cytosol]	YJR139C	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
LYS1	L-saccharopine[Cytosol] + NAD(+)[Cytosol] <=> 2-oxoglutarate[Cytosol] + L-lysine[Cytosol] + NADH[Cytosol]	YIR034C	used in the direction of L-lysine synthesis (NADH producing)	reversibility constrained
r275	NAD(+)[Cytosol] + ergosta-5,7,24(28)-trienol[Cytosol] <=> Ergosta-5,7,22,24,(28)-tetraen-3beta-ol[Cytosol] + O2[Cytosol] + NADH[Cytosol]	YKL150W:YMR015C:YNL111C	capable of using NADH or NADPH in the model, more likely to use NADPH	reversibility constrained
ADH1	NAD(+)[Cytosol] + ethanol[Cytosol] <=> NADH[Cytosol] + acetaldehyde[Cytosol]	YBR145W;YDL168W;YGL256W;YMR303C;YOL086C	not active with a dilution rate of 0.10 h-2	reversibility constrained
SFA1_1	NAD(+)[Cytosol] + formaldehyde[Cytosol] + glutathione[Cytosol] <=> NADH[Cytosol] + S-formylglutathione[Cytosol]	YDL168W	not active with a dilution rate of 0.10 h-1 (formaldehyde is not produced in these conditions)	reversibility constrained

Table A3- List of reactions modified during the curation process of the model *iTO977* (continuation)

Reaction ID	Stoichiometric equations	Gene	Comment	Verdict
MDH2	NAD(+)[Cytosol] + malate[Cytosol] <=> NADH[Cytosol] + oxaloacetate[Cytosol]	YDL078C	part of the glyoxylate cycle (not needed in glucose growth)	reversibility constrained
PGA3	NADH[Cytosol] + 2 ferricytochrome b5[Cytosol] => NAD(+)[Cytosol] + 2 ferrocytochrome b5[Cytosol]	YML125C	dead end (ferrocytochrome b5 only present in this reaction)	inactivated
GPD1	NADH[Cytosol] + glycerone phosphate[Cytosol] => NAD(+)[Cytosol] + sn-glycerol 3-phosphate[Cytosol]	YDL022W;YOL059W	thermodynamically favorable in the forward direction (dG=-31.5 KJ/mol)*	no changes applied
SCS7	NADH[Cytosol] + P-ceramide[Cytosol] + oxygen[Cytosol] => D-ceramide[Cytosol] + NADP(+)[Cytosol]	YMR272C	no information in databases about this reactions	no changes applied
NDH1	NADH[Cytosol] + ubiquinone-9[Mitochondrion] => NAD(+)[Cytosol] + ubiquinol[Mitochondrion]	YDL085W;YMR145C	Important role in the oxidation of cytosolic NADH (part of the oxidative phosphorylation)	no changes applied
HMG1	(R)-mevalonate[Cytosol] + 2 NADP(+)[Cytosol] + coenzyme A[Cytosol] <=> (S)-3-hydroxy-3-methylglutaryl-CoA[Cytosol] + 2 NADPH[Cytosol]	YLR450W;YML075C	Hydrolysis of CoA and consumption of 2 NADPH; thermodynamically favorable in the forward direction (dG=-18.6 kJ / mol)*	reversibility constrained
FAS2_1_2	NADPH[Cytosol] + acetoacetyl-[acp][Cytosol] <=> (R)-3-hydroxybutanoyl-[acp][Cytosol] + NADP(+)[Cytosol]	YPL231W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
FAS2_2_2	3-oxohexanoyl-[acp][Cytosol] + NADPH[Cytosol] <=> (R)-3-hydroxyhexanoyl-[acp][Cytosol] + NADP(+)[Cytosol]	YPL231W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
FAS2_3_2	3-oxooctanoyl-[acp][Cytosol] + NADPH[Cytosol] <=> (R)-3-hydroxyoctanoyl-[acp][Cytosol] + NADP(+)[Cytosol]	YPL231W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
FAS2_4_2	3-oxodecanoyl-[acp][Cytosol] + NADPH[Cytosol] <=> (R)-3-hydroxydecanoyl-[acp][Cytosol] + NADP(+)[Cytosol]	YPL231W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
FAS2_5_2	3-oxododecanoyl-[acp][Cytosol] + NADPH[Cytosol] <=> (R)-3-hydroxydodecanoyl-[acp][Cytosol] + NADP(+)[Cytosol]	YPL231W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
FAS2_6_2	3-oxotetradecanoyl-[acp][Cytosol] + NADPH[Cytosol] <=> (R)-3-hydroxytetradecanoyl-[acp][Cytosol] + NADP(+)[Cytosol]	YPL231W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
FAS2_7_2	3-oxohexadecanoyl-[acp][Cytosol] + NADPH[Cytosol] <=> (R)-3-hydroxypalmitoyl-[acp][Cytosol] + NADP(+)[Cytosol]	YPL231W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained

Table A3- List of reactions modified during the curation process of the model *iTO977* (continuation)

Reaction ID	Stoichiometric equations	Gene	Comment	Verdict
FAS2_8_2	3-oxostearoyl-[acp][Cytosol] + NADPH[Cytosol] \Leftrightarrow 3-hydroxyoctadecanoyl-[acp][Cytosol] + NADP(+)[Cytosol]	YPL231W	fatty acid synthesis- reaction occurs in the forward direction	reversibility constrained
ADE3_1	5,10-methylenetetrahydrofolate[Cytosol] + NADP(+)[Cytosol] \Leftrightarrow 5,10-methenyltetrahydrofolate[Cytosol] + NADPH[Cytosol]	YGR204W	inconclusive data regarding the reversibility of the reaction	no changes applied
GND1	6-phospho-D-gluconate[Cytosol] + NADP(+)[Cytosol] \Rightarrow CO2[Cytosol] + D-ribulose 5-phosphate[Cytosol] + NADPH[Cytosol]	YGR256W;YHR183W	Oxidative pentose phosphate pathway	no changes applied
r173	2-methylpropanal[Cytosol] + NADP(+)[Cytosol] \Rightarrow NADPH[Cytosol] + isobutyl alcohol[Cytosol]	YMR318C;YCR105W	wrong stoichiometry-reaction should consume NADPH	stoichiometry corrected
ARA1_2	D-arabinose[Cytosol] + NADP(+)[Cytosol] \Rightarrow D-arabinono-1,4-lactone[Cytosol] + NADPH[Cytosol]	YBR149W	capable of using NADH or NADPH in the model, more likely to use NADH	inactivated
FAS1_1_2	NADPH[Cytosol] + trans-but-2-enoyl-[acp][Cytosol] \Leftrightarrow NADP(+)[Cytosol] + butyryl-[acp][Cytosol]	YKL182W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
FAS1_2_2	NADPH[Cytosol] + trans-hex-2-enoyl-[acp][Cytosol] \Leftrightarrow NADP(+)[Cytosol] + hexanoyl-[acp][Cytosol]	YKL182W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
FAS1_3_2	NADPH[Cytosol] + trans-oct-2-enoyl-[acp][Cytosol] \Leftrightarrow NADP(+)[Cytosol] + octanoyl-[acp][Cytosol]	YKL182W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
FAS1_4_2	NADPH[Cytosol] + trans-dec-2-enoyl-[acp][Cytosol] \Leftrightarrow NADP(+)[Cytosol] + decanoyl-[acp][Cytosol]	YKL182W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
FAS1_5_2	NADPH[Cytosol] + trans-dodec-2-enoyl-[acp][Cytosol] \Leftrightarrow NADP(+)[Cytosol] + dodecanoyl-[acp][Cytosol]	YKL182W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
FAS1_6_2	NADPH[Cytosol] + trans-tetradec-2-enoyl-[acp][Cytosol] \Leftrightarrow NADP(+)[Cytosol] + tetradecanoyl-[acp][Cytosol]	YKL182W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
FAS1_7_2	NADPH[Cytosol] + trans-hexadec-2-enoyl-[acp][Cytosol] \Leftrightarrow NADP(+)[Cytosol] + hexadecanoyl-[acp][Cytosol]	YKL182W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
FAS1_8_2	NADPH[Cytosol] + trans-octadec-2-enoyl-[acp][Cytosol] \Leftrightarrow NADP(+)[Cytosol] + stearoyl-[acp][Cytosol]	YKL182W	fatty acid synthesis - reaction occurs in the forward direction	reversibility constrained
r581	FMN[Cytosol] + NADP(+)[Cytosol] \Rightarrow NADPH[Cytosol] + Reduced FMN[Cytosol]	YPL069C	capable of using NADH or NADPH in the model, more likely to use NADH	inactivated

Table A3- List of reactions modified during the curation process of the model *iTO977* (continuation)

Reaction ID	Stoichiometric equations	Gene	Comment	Verdict
r572	FMN[Cytosol] + NADP(+)[Cytosol] => NADPH[Cytosol] + Reduced FMN[Cytosol]	YLR011W	capable of using NADH or NADPH in the model, more likely to use NADH	inactivated
ZWF1	NADP(+)[Cytosol] + alpha-D-glucose 6-phosphate[Cytosol] <=> 6-O-phosphono-D-glucono-1,5-lactone[Cytosol] + NADPH[Cytosol]	YNL241C	Oxidative pentose phosphate pathway	no changes applied
r671	NADP(+)[Cytosol] + glycerol[Cytosol] => Dihydroxyacetone[Cytosol] + NADPH[Cytosol]	YOR120W	thermodynamically favorable in the reverse direction (dG=25 KJ/mol)*	inactivated
IDP2_1	NADP(+)[Cytosol] + isocitrate[Cytosol] => NADPH[Cytosol] + oxalosuccinate[Cytosol]	YLR174W	cytosolic NADP ⁺ -dependent isocitrate dehydrogenase [1]	inactivated
r253	L-Allo-threonine[Cytosol] + NADP(+)[Cytosol] => L-2-Amino-3-oxobutanoate[Cytosol] + NADPH[Cytosol]	YMR226C	dead end (L-allo-threonine only present in this reaction)	inactivated
r833	NADP(+)[Cytosol] + L-serine[Cytosol] => 2-Aminomalonate semialdehyde[Cytosol] + NADPH[Cytosol]	YMR226C	dead end (2-aminomalonate semialdehyde only present in this reaction)	inactivated
ALD6	NADP(+)[Cytosol] + acetaldehyde[Cytosol] => NADPH[Cytosol] + acetate[Cytosol]	YPL061W	capable of using NADH or NADPH in the model, more likely to use NADH	inactivated
TYR1	NADP(+)[Cytosol] + prephenate[Cytosol] => 3-(4-hydroxyphenyl)pyruvate[Cytosol] + CO2[Cytosol] + NADPH[Cytosol]	YBR166C	capable of using NADH or NADPH in the model, more likely to use NADH	inactivated
LYS9	L-2-aminoadipate 6-semialdehyde[Cytosol] + L-glutamate[Cytosol] + NADPH[Cytosol] <=> L-saccharopine[Cytosol] + NADP(+)[Cytosol]	YNR050C	used in the direction of L-lysine synthesis (NADPH consuming)	reversibility constrained
YBR006W	NADP(+)[Cytosol] + succinate semialdehyde[Cytosol] => NADPH[Cytosol] + succinate[Cytosol]	YBR006W	thermodynamically favorable in the forward direction (dG=-63 KJ/mol)*	no changes applied
ECM17	3 NADPH[Cytosol] + sulfite[Cytosol] <=> H2S[Cytosol] + 3 NADP(+)[Cytosol]	YFR030W;YJR137C	thermodynamically favorable in the reverse direction (dG=103 KJ/mol)*	reversibility constrained
AYR1	NADP(+)[Cytosol] + acylglycerone phosphates[Cytosol] => NADPH[Cytosol] + acyl-sn-glycerol 3-phosphates[Cytosol]	YIL124W	wrong stoichiometry-reaction should consume NADPH (EC 1.1.1.101)	stoichiometry corrected

* The change in Gibbs free energy was calculated using the website: <http://equilibrator.weizmann.ac.il/>. Standard conditions were used for every metabolite except NADPH, NADP⁺, NADH and NAD⁺. For these cofactors the concentrations were obtained from references [1] and [2]: [NADPH]= 263 μM; [NADP⁺]= 35.5 μM; [NADH]= 120 μM; [NAD⁺]= 1810 μM

Table A4- List of reactions modified during the curation process of the model Yeast 6

Reaction ID	Stoichiometric equations	Gene	Comment	Verdict
r_0169	$H_c + NADH_c + \text{phenylacetaldehyde}_c \rightarrow NAD_c + 2\text{phenylethanol}_c$	YOL086C or YBR145W) or YDL168W	not active with a dilution rate of 0.10 h ⁻¹ (2-phenylethanol is not produced in these conditions)	inactivated
r_0166	$H_c + NADH_c + 2\text{methylbutanal}_c \rightarrow NAD_c + 2\text{methylbutanol}_c$	YOL086C or YBR145W) or YDL168W	not active with a dilution rate of 0.10 h ⁻¹ (2-methyl-1-butanol is not produced in these conditions)	inactivated
r_0179	$H_c + NADH_c + 3\text{methylbutanal}_c \rightarrow NAD_c + \text{isoamylol}_c$	((YOL086C or YBR145W) or YDL168W)	not active with a dilution rate of 0.10 h ⁻¹ (isoamylol is not produced in these conditions)	inactivated
r_0182	$H_c + NADH_c + \text{isobutyraldehyde}_c \rightarrow NAD_c + \text{isobutanol}_c$	((YOL086C or YBR145W) or YDL168W)	not active with a dilution rate of 0.10 h ⁻¹ (isobutanol is not produced in these conditions)	inactivated
r_0186	$H_c + NADH_c + \text{indol}_3\text{ylacetaldehyde}_c \rightarrow NAD_c + \text{tryptophol}_c$	((YOL086C or YBR145W) or YDL168W)	not active with a dilution rate of 0.10 h ⁻¹ (tryptophol is not produced in these conditions)	inactivated
r_0770	$H_c + NADH_c + \text{ubiquinone}_6_m \rightarrow NAD_c + \text{ubiquinol}_6_m$	(YMR145C or YDL085W)	important role in the oxidation of cytosolic NADH (part of the oxidative phosphorylation)	no changes applied
r_0771	$ATP_c + NADH_c \rightarrow H_c + ADP_c + NADPH_c$	(YJR049C or YEL041W)	<i>de novo</i> production of NADPH	no changes applied
r_0714	$NAD_c + S\text{malate}_c \rightleftharpoons H_c + NADH_c + \text{oxaloacetate}_c$	YOL126C	part of the glyoxylate cycle (not needed in growth on glucose)	reversibility constrained
r_2115	$H_c + NADH_c + \text{acetaldehyde}_c \rightarrow NAD_c + \text{ethanol}_c$	YOL086C or YBR145W	not active with a dilution rate of 0.10 h ⁻¹ (ethanol is not produced in these conditions)	inactivated
r_0441	$H_c + NADH_c + FMN_c \rightarrow NAD_c + FMNH2_c$	YLR011W	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
r_0486	$NAD_c + \text{phosphate}_c + \text{glyceraldehyde}_3\text{phosphate}_c \rightleftharpoons H_c + NADH_c + 1_3\text{bisphospho}_D\text{glycerate}_c$	((YJR009C or YJL052W) or YGR192C)	part of Glycolysis (NADH producing)	reversibility constrained
r_0470	$NAD_c + H2O_c + L\text{glutamate}_c \rightleftharpoons H_c + NADH_c + \text{ammonium}_c + 2\text{oxoglutarate}_c$	YDL215C	should degrade glutamate to ammonia and alpha-ketoglutarate (SGD)	reversibility constrained
r_0472	$H_c + NADH_c + L\text{glutamine}_c + 2\text{oxoglutarate}_c \rightarrow NAD_c + 2.0*L\text{glutamate}_c$	YDL171C	thermodynamically favorable in the forward direction (dG=-40 KJ/mol)*	no changes applied

Table A4- List of reactions modified during the curation process of the model Yeast 6 (continuation)

Reaction ID	Stoichiometric equations	Gene	Comment	Verdict
r_0491	$H_c + NADH_c + \text{dihydroxyacetone_phosphate_c} \rightarrow \text{glycerol_3_phosphate_c} + NAD_c$	(YOL059W or YDL022W)	thermodynamically favorable in the forward direction ($dG=-31.5$ KJ/mol)*	no changes applied
r_0003	$NAD_c + R_R_2_3_butanediol_c \rightleftharpoons H_c + NADH_c + R_acetoin_c$	YAL060W	not active with a dilution rate of 0.10 h ⁻¹ (acetoin is not produced in these conditions)	reversibility constrained
r_1010	$H_r + \text{oxygen_r} + \text{squalene_r} + NADH_r \rightarrow H_2O_r + S_2_3_epoxysqualene_r + NAD_r$	YGR175C	capable of using NADH or NADPH in the model, more likely to use NADPH	inactivated
r_0173	$H_2O_c + NADP_c + \text{acetaldehyde_c} \rightarrow 2.0 * H_c + NADPH_c + \text{acetate_c}$	YPL061W	capable of using NADH or NADPH in the model, more likely to use NADH	inactivated
r_0177	$\text{indol_3_ylacetaldehyde_c} + H_2O_c + NADP_c \rightarrow 2.0 * H_c + \text{indole_3_acetate_c} + NADPH_c$	YPL061W	dead end (indole-3-acetate cannot be metabolized)	inactivated
r_0234	$NADP_c + \text{zymosterol_intermediate_1c_c} \rightarrow H_c + \text{carbon_dioxide_c} + NADPH_c + \text{zymosterol_intermediate_2_c}$	YGL001C	other models have NAD-dependent alternative- see text	inactivated and new reaction was added: r_234x
r_0659	$NADP_c + \text{isocitrate_c} \rightarrow \text{carbon_dioxide_c} + NADPH_c + 2_oxoglutarate_c$	YLR174W	cytosolic NADP ⁺ -dependent isocitrate dehydrogenase [1]	inactivated
r_0676	$NADP_c + L_allothreonine_c \rightarrow H_c + L_2_amino_3_oxobutanoate_c + NADPH_c$	YMR226C	dead end	inactivated
r_0690	$NADP_c + L_serine_c \rightarrow H_c + NADPH_c + L_alpha_formylglycine_c$	YMR226C	dead end (alpha_formylglycine only present in this reaction)	inactivated
r_0732	$NADP_c + 5_10_methylentetrahydrofolate_c \rightleftharpoons NADPH_c + 5_10_methenyl_THF_c$	YGR204W	inconclusive data regarding the reversibility of the reaction	no changes applied
r_0321	$NADP_c + D_arabinose_c \rightarrow H_c + NADPH_c + D_arabinono_1_4_lactone_c$	YBR149W	capable of using NADH or NADPH in the model, more likely to use NADH	inactivated
r_1023	$H_2O_c + NADP_c + \text{succinic_semialdehyde_c} \rightarrow 2.0 * H_c + \text{succinate_c} + NADPH_c$	YBR006W	thermodynamically favorable in the forward direction ($dG=-63$ KJ/mol)*	no changes applied
r_0466	$D_glucose_6_phosphate_c + NADP_c \rightarrow H_c + NADPH_c + 6_O_phosphono_D_glucono_1_5_lactone_c$	YNL241C	Oxidative pentose phosphate pathway	no changes applied

Table A4- List of reactions modified during the curation process of the model Yeast 6 (continuation)

Reaction ID	Stoichiometric equations	Gene	Comment	Verdict
r_0889	$\text{NADP_c} + 6\text{_phospho_D_gluconate_c} \text{ ----->} \\ \text{carbon_dioxide_c} + \text{NADPH_c} + \text{D_ribulose_5_phosphate_c}$	(YGR256W or YHR183W)	Oxidative pentose phosphate pathway	no changes applied
r_0939	$\text{prephenate_c} + \text{NADP_c} \text{ ----->} \\ 3\text{_4_hydroxyphenylpyruvate_c} + \text{carbon_dioxide_c} + \\ \text{NADPH_c}$	YBR166C	other models have NAD-dependent alternative- see text	inactivated and new reaction was added: r_939x

* The change in Gibbs free energy was calculated using the website: <http://equilibrator.weizmann.ac.il/>. Standard conditions were used for every metabolite except NADPH, NADP+, NADH and NAD+. For these cofactors the concentrations were obtained from references [1] and [2]: [NADPH]= 263 μM; [NADP+]= 35.5 μM; [NADH]= 120 μM; [NAD+]= 1810 μM

Appendix B: Modifications applied to the model Yeast 6 for simulation and optimization purposes

Table B1- List of general changes performed in the model Yeast 6

Modification	Details	Source/Justification
Added reaction r_4039	acetate_m + succinyl_CoA_m<---->succinate_m + acetyl_CoA_m	imported from Yeast 7.11 (more recent consensus model)
Modified r_0718	NADP_m + S_malate_m<---->NADPH_m + pyruvate_m + carbon_dioxide_m	updated according to Yeast 7.11 (more recent consensus model)
Modified r_0773	NADH_m + H_m + ubiquinone_6_m<----> NAD_m + ubiquinol_6_m	updated according to Yeast 7.11 (more recent consensus model)
Modified r_0226	4 H_c + 1 ADP_m + 1 phosphate_m <----> 3 H_m + 1 H2O_m + 1 ATP_m	updated according to Yeast 7.11 (more recent consensus model)
Modified r_2129	H_m <----> H_c	updated according to Yeast 7.11 (more recent consensus model)
Modified r_1110	ADP_c + ATP_m <----> ATP_c + ADP_m	updated according to Yeast 7.11 (more recent consensus model)
Modified r_0470	NAD_c + H2O_c + L_glutamate_c <----> H_c + NADH_c + ammonium_c + 2_oxoglutarate_c	updated according to Yeast 7.11 (more recent consensus model)
Removed r_0303	citrate_c <----> H2O + cis_aconitate_c	updated according to Yeast 7.11 (more recent consensus model)
Removed r_0338	ubiquinone_6_m + S_dihydroorotate_c <----> ubiquinol_6_m + orotate_c	updated according to Yeast 7.11 (more recent consensus model)
Removed r_0339	oxygen_c + S_dihydroorotate_c <----> hydrogen_peroxide_c + orotate_c	updated according to Yeast 7.11 (more recent consensus model)
Removed r_2127	NAD_c + S_dihydroorotate_c <----> NADH_c + orotate_c	updated according to Yeast 7.11 (more recent consensus model)
Modified r_1254	H_e + pyruvate_e <----> H_c + pyruvate_c	updated according to Yeast 7.11 (more recent consensus model)
Gene rule update r_0530	(YPL252C AND YDR376W) or (YPL252C AND YDR376W AND YER141W)	updated according to Yeast 7.11 (more recent consensus model)
Updated biomass reaction r_4041	Add 1.0E-6 * chitin_c and 1.0E-6* heme_a_m to reactants	updated according to Yeast 7.11 (more recent consensus model)
Gene rule update r_0076	(YER037W or YGL224C)	updated according to Yeast 7.11 (more recent consensus model)
Gene rule update r_0078	(YER037W or YGL224C)	updated according to Yeast 7.11 (more recent consensus model)
Gene rule update r_1619	YER037W	updated according to Yeast 7.11 (more recent consensus model)
Removed r_0333	phosphate_c + 2_deoxyuridine_c <----> uracil_c + 2_deoxy_D_ribofuranose_1_phosphate_c	updated according to Yeast 7.11 (more recent consensus model)

Table B1- List of general changes performed in the model Yeast 6 (continuation)

Modification	Details	Source/Justification
Removed r_0944	phosphate_c + adenosine_c <----> alpha_D_ribose_1_phosphate_c + adenine_c	updated according to Yeast 7.11 (more recent consensus model)
Removed r_0945	phosphate_m + adenosine_m <----> adenine_m + alpha_D_ribose_1_phosphate_m	updated according to Yeast 7.11 (more recent consensus model)
Removed r_0946	phosphate_c + 2_deoxyadenosine_c <----> 2_deoxy_D_ribofuranose_1_phosphate_c + adenine_c	updated according to Yeast 7.11 (more recent consensus model)
Removed r_0947	phosphate_c + 2_deoxyguanosine_c <----> guanine_c + 2_deoxy_D_ribofuranose_1_phosphate_c	updated according to Yeast 7.11 (more recent consensus model)
Removed r_0948	phosphate_c + 2_deoxyinosine_c <----> hypoxanthine_c + 2_deoxy_D_ribofuranose_1_phosphate_c	updated according to Yeast 7.11 (more recent consensus model)
Removed r_0952	phosphate_c + xanthosine_c <----> alpha_D_ribose_1_phosphate_c + 9H_xanthine_c	updated according to Yeast 7.11 (more recent consensus model)
Removed r_1044	phosphate_c + thymidine_c <----> 2_deoxy_D_ribofuranose_1_phosphate_c + thymine_c	updated according to Yeast 7.11 (more recent consensus model)
Added reaction r_4045	H2O_c + uridine_c <----> uracil_c + D_ribose_c	imported from Yeast 7.11 (more recent consensus model)
Gene rule update r_0888	(YMR105C or YKL127W)	updated according to Yeast 7.11 (more recent consensus model)
Gene rule update r_0907	((YMR278W or YMR105C) or YKL127W)	updated according to Yeast 7.11 (more recent consensus model)
Modified r_0110	H_c + coenzyme_A_c + acetate_c <----> H2O_c + acetyl_CoA_c	As corrected in chapter 3
Added reaction r_0234x	NAD_c + zymosterol_intermediate_1c_c <----> H_c + NADH_c + carbon_dioxide_c + zymosterol_intermediate_2_c	As corrected in chapter 3
Added reaction r_0939x	prephenate_c + NAD_c <----> NADH_c + 3_4_hydroxyphenylpyruvate_c + carbon_dioxide_c	As corrected in chapter 3
Inactivated r_1840	3_hydroxy_3_methylglutaryl_CoA_c <----> 3_hydroxy_3_methylglutaryl_CoA_m	No data supporting such reaction
Modified r_1117	H_c + L_aspartate_c <----> H_m + L_aspartate_m	As corrected in chapter 3

Table B2- List of changes applied to the model Yeast 6 to allow polyhydroxybutyrate (PHB) production

Modification	Details	Source/Justification
Added reaction R_phaA	2 acetyl_CoA_c -----> coenzyme_A_c + acetoacetyl_CoA_c	Acetoacetyl-CoA thiolase from <i>Ralstonia eutropha</i>
Added reaction R_phaB	H_c + NADPH_c + acetoacetyl_CoA_c -----> 3hbcoa_c + NADP_c	Acetoacetyl-CoA reductase from <i>Ralstonia eutropha</i>
Added reaction R_phaC	3hbcoa_c -----> phb_c + coenzyme_A_c	PHB synthase from <i>Ralstonia eutropha</i>
Added reaction R_EX_PHB_e	PHB_c ----->	PHB drain reaction required for simulation purposes

Table B3- List of changes applied to the model Yeast 6 to allow 3-hydroxypropionic acid (3-HP) production

Modification	Details	Source/Justification
Added reaction R_MCR	2* H_c + 2* NADPH_c + 1* malonyl_CoA_c -----> 1* coenzyme_A_c + 2* NADP_c + 1* 3hp_c	Malonyl-CoA reductase from <i>Chloroflexus aurantiacus</i>
Added reaction R_EX_3hp_e	M_3hp_c ----->	3-HP drain reaction required for simulation purposes
Added reaction R_gapN	glyceraldehyde_3_phosphate_c + H2O_c + NADP_c -----> H_c + NADPH_c + 3_phosphoglycerate_c	Glyceraldehyde-3-phosphate dehydrogenase from <i>Streptococcus mutans</i>

Table B4- List of changes applied to the model Yeast 6 to allow santalene production

Modification	Details	Source/Justification
Added reaction R_SANSYN	farnesyl_diphosphate_c -----> diphosphate_c + santalene_c	Santalene synthase from <i>Clausena lansium</i>
Added reaction R_EX_santalene_e	santalene_c ----->	Santalene drain reaction required for simulation purposes

Table B5- List of changes applied to the model Yeast 6 to allow vanillin β -D-glucoside production

Modification	Details	Source/Justification
Added reaction r_3DSD	3_dehydroshikimate_c -----> H2O_c + pac_c	3-dehydroshikimate dehydratase from <i>Podospora pausiceta</i>
Added reaction r_ACAR	ATP_c + pac_c + NADPH_c -----> H2O_c + diphosphate_c + AMP_c + pal_c + NADP_c	Aryl-aldehyde dehydrogenase from <i>Nocardia sp.</i>
Added reaction r_hsOMT	pal_c + S_adenosyl_L_methionine_c -----> S_adenosyl_L_homocysteine_c + vanillin_c	O-methyltransferase from <i>Homo sapiens</i>
Added reaction r_UGT	UDP_D_glucose_c + vanillin_c -----> UDP_c + vanillin_glucoside_c	UDP-glycosyltransferase from <i>Arabidopsis thaliana</i>
Added reaction r_ex_pac	pac_c ----->	Protocatechuic acid drain reaction required for simulation purposes
Added reaction r_ex_pal	pal_c ----->	Protocatechuic aldehyde drain reaction required for simulation purposes
Added reaction r_ex_vanillin_glucosid e	vanillin_glucoside_c ----->	Vanillin glucoside drain reaction required for simulation purposes

Table B6- List of changes applied to the model Yeast 6 for optimizing 3-hydroxypropionic acid (3-HP) production

Modification	Details	Source/Justification
Inactivated r_0113	ATP_m + coenzyme_A_m + acetate_m -----> acetyl_CoA_m + AMP_m + diphosphate_m	It is not likely that Acp1p is active in the mitochondria [3, 4]
Updated biomass reaction r_4041	Removed 1.0E-6 * chitin_c and 1.0E-6* heme_a_m from reactants	Low stoichiometric coefficients interfere with the optimization procedure
Added reaction R_MCR	2* H_c + 2* NADPH_c + 1* malonyl_CoA_c -----> 1* coenzyme_A_c + 2* NADP_c + 1* 3hp_c	Malonyl-CoA reductase from <i>Chloroflexus aurantiacus</i>
Added reaction R_EX_3hp_e	M_3hp_c ----->	3-HP drain reaction required for simulation purposes

Table B7- List of changes applied to the model Yeast 6 to simulated growth on C₂ carbon sources

Modification	Details	Source/Justification
Added reaction r_1635p	M_acetate_c <----> M_acetate_x	Acetate transporter to the peroxisome/Diffusion of acetate to the peroxisome
Added reaction r_0113	coenzyme_A_x + acetate_x + ATP_x -----> acetyl_CoA_x + diphosphate_x + AMP_x	Acs1p is possibly localized in the peroxisome [5]. It is required for an active glyoxylate cycle in the model Yeast 6

Appendix C: Construction of 3-hydroxypropionic acid producing strains

Table C1- Primers used in the work performed in chapter 5

#	Name	Sequence (5'-3')
1	PDC1_Fw	ATGTCTGAAATTACTTTGGGTAAATATTTG
2	PDC1_mut_Rv1	ATCTGTCAATTCAGCTGGGGCGGTAGTAATGTCAGTGATCATAGCAGTG
3	PDC1_mut_Fw2	ACCGCCCCAGCTGAAATTGACAGATGTATCAGAACCCTTACGTCACCCAAAGACCAG
4	PDC1_mut_Rv2	CGAATTTTCATTTGGACACCTG
5	PDC1_FR_FW	ATCAGAAAACGCCACTTTCCC
6	PDC1_Rv	TTATTGCTTAGCGTTGGTAGCA
7	PGK1_AscI_site	GGTCGCTCCAATTCCTAGTTCGTTTGGCGCGCCTGGAAGTACCTTCAAAGAATGG
8	PGK1_PDC1	ATTTACCCAAAAGTAATTCAGACATTGTTTGGATCCTTGTTTTATATTTG
9	CYC1_PDC1	TGCTGCTACCAACGCTAAGCAATAAATCCGCTCTAACCGAAAAGG
10	CYC1_AscI_site	TTCGCTATTACGCCAGCTGGATAAAGGCGCGCCCTTCGAGCGTCCCAAAAC
11	RPLC_ALD2_Fw	ACAACAAATATAAAAACAAGGATCCAAAACAATGCCTACCTGTATACTGATATCG
12	RPLC_ALD2_Rv	TAGCTAGCCGCGGTACCAAGCTTACTCGAGTTAGTTGTCCAAAGAGAGATTTATG
13	Ppp1_UP_Fw	AGTGGCTTCCCTTGCTTTC
14	Ppp1_Up_Rv	ATACGAACGGTAATTAAGGGTTGTCAGTTAAGGGAGACATGGAGAGA
15	Ppp1_Down_Fw	GCTATACGAACGGTATATCAGATCCTGACGATATAAACTTATGCTGAGG
16	Ppp1_Down_Rv	GATTGTGGTTTTGTAACGTCG
17	KanMX_Fw	CTGAAGCTTCGTACGCTG
18	KanMX_IntRv	TCACCATGAGTGACGACTG
19	KanMX_IntFw	TTCCAACATGGATGCTGAT
20	KanMX_Rv	CCACTAGTGGATCTGATATCACC
21	Ppp1_Out_Fw	GGTGGCATCTTCGAAATATAAC
22	Ppp1_Out_rv	ACCAACTAACGCAATGATGA
23	ACH1_UpFw	CAAACATAACCACGATCCAAACG
24	ACH1_UpRv	CTGCAGCGTACGAAGCTTCAGGCATACCTAACTCTCTGCTTTAAACAA
25	ACH1_Down_Fw	AGGTGATATCAGATCCACTAGTGGAGAATGCTTTCAAGTTCCACACC
26	ACH1_Down_Rv	GCTTACCAATCCTTCCACCAC
27	ACH1_out_Fw	GCAGAGATTATCGCCATCAACTACTA
28	pXI5_UP_Fw	GCGGAGAAGTCGTTGATAG
29	ADH1_Fw	ACGAAAAGATGGAAAAGGGTCAAATCGTTGGTAGATACGTTGTTGACACTTC
30	Cyc1_Fw_long	CGAGTAAGCTTGGTACCGCGGCTAGCTAAGATCCGCTCTAACCGAAAAGGA
31	pXI5_Down_Rv	GATCATAGATCCGGCACTTAGAGA
32	X2_UP_Fw	CGTCTATGAGGAGACTGTTAGTTGG
33	X2_UP_Rv	TCACCAAGCTCTTAAAACGGGAATTTATGGTGCACACGTGACCACTTCGAGAGC
34	His3_Fw	GTGCACCATAAAATCCCGTT
35	His3_Rv	CAGTGGTGTGATGGTCGTCT

Table C1- Primers used in the work performed in chapter 5 (continuation)

36	X2_Down_Fw	AGGTTTTGGGACGCTCGAAGATCCTCCGGATCCCTGCATAATCGGCCTCACA
37	X2_Down_rv	CTCGCCAAGGCATTACCATC
38	Cyc1_fw	CTCGAGTAAGCTTGGTACCGC
39	Cyc1_Rv	GATCCGGAGGATCTTCGAGC
40	check_XI5_fw	GGTAGTTGTTCTACACGCCA
41	ACC1_test_fw	CATCAGGTAGGCGAAGCATCA
42	check_XI5_rv	GCGAAAGCGTGGACTAGGTA
43	MCR1shortF	TTGGAAGTTATGGATGGTTCTGAC
44	MCR1shortR	ATTACTCTGGACCTCTTGCTCTA
45	check_X2_fw	ACAATGTAGTGGTAGTAGCAGCA
46	check_X2_Rv	TTCCATCGTCGGTTGTACGC
47	ACSse_Fw	CCTGCGAATATCGCTGACAGG
48	tADH1_Rv	GGCAAGGTAGACAAGCCGACAAC
49	RPLC_ALD4_Fw	ACAACAAATATAAAACAAGGATCCAAAACAATGTCACACCTTCCTATGACAGTG
50	RPLC_ALD4_Rv	TAGCTAGCCGCGGTACCAAGCTTACTCGAGTTACTCGTCCAATTTGGCAC
51	TEF1_rv	GTTTTGCGCCGCTTGTAAT
52	PGK1_rv	CATTGTTTGGATCCTTGTTTTA
53	ALD4_check_fw	CCGTGCCTTCTCTAATGGGT
54	DEL1_fw	GACAACCCTTAATTACCGTTCGTATAATG
55	DEL1_rv	GACATTTAGGTTTTTCTCCTTGACG
56	DEL2_fw	CTCTCGAGAACCCTTAATCG
57	DEL2_rv	GGATCTGATATACCGTTCGTATAGC

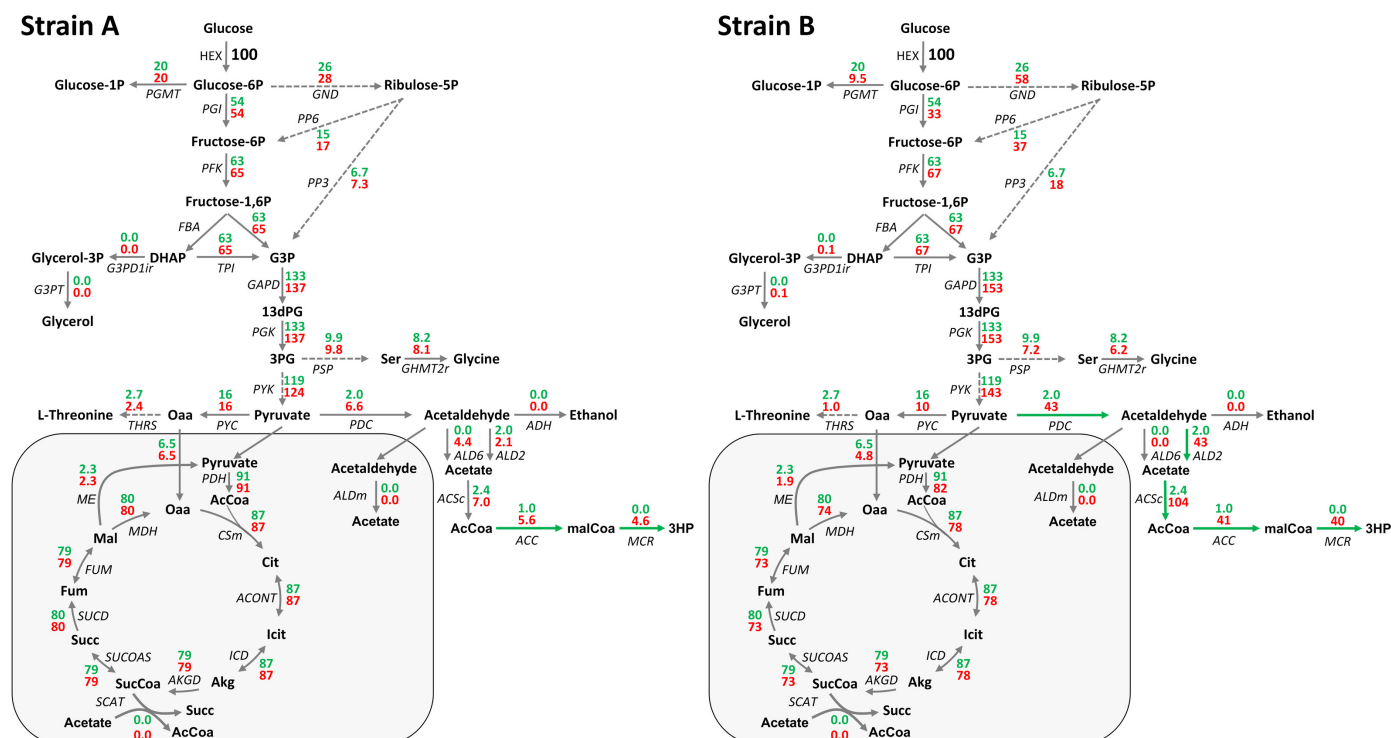


Figure C1- Flux distribution in the central metabolism for the wild-type (green) and for the strain designs obtained with TDPS A-D (red). The wild type flux distribution was obtained using the cofactor constraints described in chapter 3 and the mutant strains' flux distributions were calculated with TDPS. Reactions: ACC- acetyl-CoA carboxylase, ACH1- succinyl-CoA:acetate CoA-transferase, ACONT- aconitase, ACSc- acetyl-CoA synthetase, AKGD- alpha-ketoglutarate dehydrogenase, ADH- alcohol dehydrogenase, ALD2/6- $\text{NAD}^+/\text{NADP}^+$ cytosolic aldehyde dehydrogenase, ALDm- mitochondrial aldehyde dehydrogenase, CSm- citrate synthase, FBA- fructose 1,6-bisphosphate aldolase, FUM- fumarase, G3PD1ir- glycerol-3-phosphate dehydrogenase, G3PT- glycerol-1-phosphatase, GAPD- glyceraldehyde-3-phosphate dehydrogenase, GHMT2r- serine hydroxymethyltransferase, GND- 6-phosphogluconate dehydrogenase, HEX- hexokinase, ICD- mitochondrial isocitrate dehydrogenase, MCR- malonyl-CoA reductase, MDH- mitochondrial malate dehydrogenase, ME- mitochondrial malic enzyme, PDC- pyruvate decarboxylase, PDH- pyruvate dehydrogenase, PFK- phosphofruktokinase, PGI- phosphoglucose isomerase, PGK- 3-phosphoglycerate kinase, PGMT- phosphoglucomutase, PP3- sum of the non-oxidative reactions of the pentose phosphate pathway producing glyceraldehyde-3-phosphate, PP6- sum of the non-oxidative reactions of the pentose phosphate pathway producing fructose-6-phosphate, PSP- phosphoserine phosphatase, PYC- pyruvate carboxylase, PYK- pyruvate kinase, SUCD- succinate dehydrogenase, SUCOAS- succinyl-CoA ligase, THRS- threonine synthase, TPI- triose phosphate isomerase; Metabolites: 13dPG- 1,3-diphosphoglycerate, 3HP- 3-hydroxypropionic acid, 3PG- 3-phosphoglycerate, AcCoA- acetyl-CoA, Akg- 2-oxoglutarate, Cit- citrate, DHAP- dihydroxyacetone-phosphate, Fum- fumarate, G3P- glyceraldehyde-3-phosphate, Icit- isocitrate, Mal - L-malate, malCoA- malonyl-CoA, Oaa- oxaloacetate, Ser- L-serine, Succ- succinate, SucCoa- succinyl-CoA.

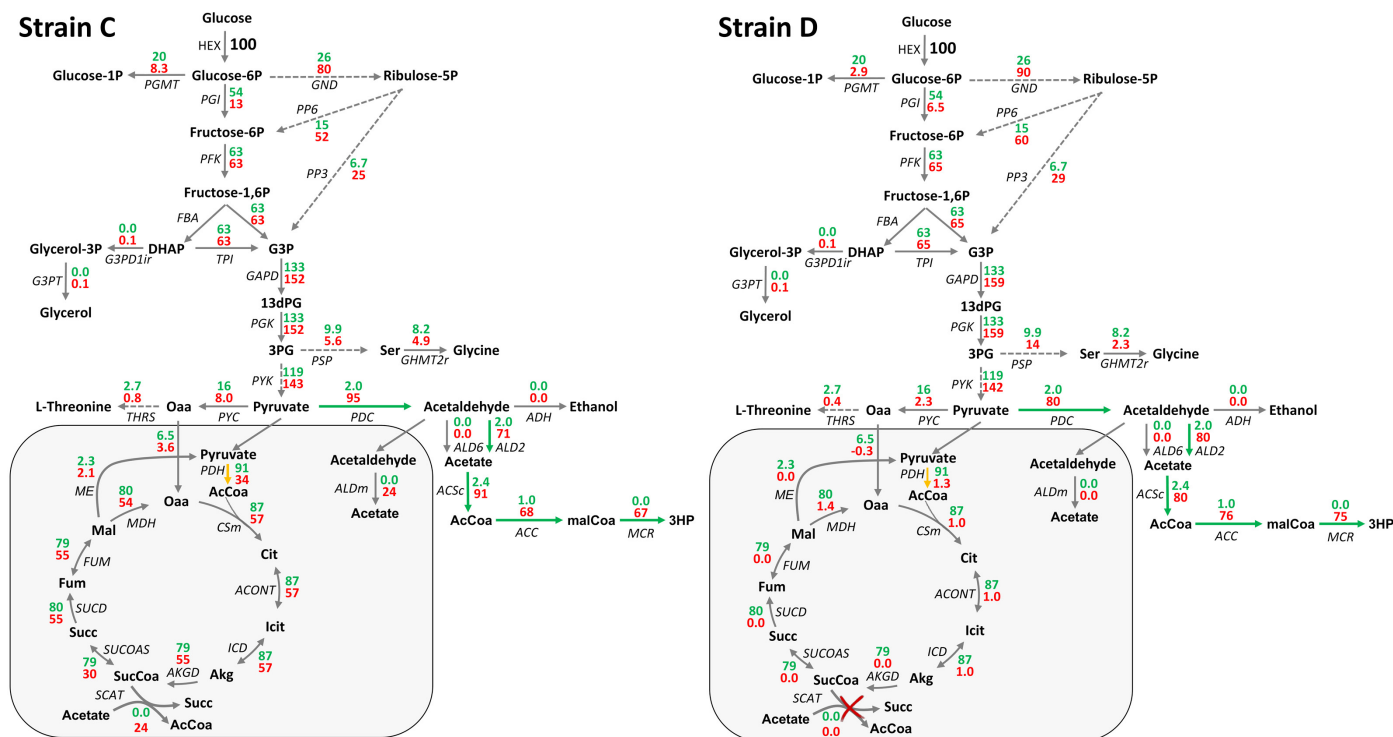


Figure C1- Flux distribution in the central metabolism for the wild-type (green) and for the strain designs obtained with TDPS A-D (red). The wild type flux distribution was obtained using the cofactor constraints described in chapter 3 and the mutant strains' flux distributions were calculated with TDPS. Reactions: ACC- acetyl-CoA carboxylase, ACH1- succinyl-CoA:acetate CoA-transferase, ACONT- aconitase, ACSc- acetyl-CoA synthetase, AKGD- alpha-ketoglutarate dehydrogenase, ADH- alcohol dehydrogenase, ALD2/6- NAD⁺/NADP⁺ cytosolic aldehyde dehydrogenase, ALDm- mitochondrial aldehyde dehydrogenase, CSm- citrate synthase, FBA- fructose 1,6-bisphosphate aldolase, FUM- fumarase, G3PD1ir- glycerol-3-phosphate dehydrogenase, G3PT- glycerol-1-phosphatase, GAPD- glyceraldehyde-3-phosphate dehydrogenase, GHMT2r- serine hydroxymethyltransferase, GND- 6-phosphogluconate dehydrogenase, HEX- hexokinase, ICD- mitochondrial isocitrate dehydrogenase, MCR- malonyl-CoA reductase, MDH- mitochondrial malate dehydrogenase, ME- mitochondrial malic enzyme, PDC- pyruvate decarboxylase, PDH- pyruvate dehydrogenase, PFK- phosphofruktokinase, PGI- phosphoglucose isomerase, PGK- 3-phosphoglycerate kinase, PGM1- phosphoglucosmutase, PP3- sum of the non-oxidative reactions of the pentose phosphate pathway producing glyceraldehyde-3-phosphate, PP6- sum of the non-oxidative reactions of the pentose phosphate pathway producing fructose-6-phosphate, PSP- phosphoserine phosphatase, PYC- pyruvate carboxylase, PYK- pyruvate kinase, SUCD- succinate dehydrogenase, SUCOAS- succinyl-CoA ligase, THRS- threonine synthase, TPI- triose phosphate isomerase; Metabolites: 13dPG- 1,3-diphosphoglycerate, 3HP- 3-hydroxypropionic acid, 3PG- 3-phosphoglycerate, AcCoA- acetyl-CoA, Akg- 2-oxoglutarate, Cit- citrate, DHAP- dihydroxyacetone-phosphate, Fum- fumarate, G3P- glyceraldehyde-3-phosphate, Icit- isocitrate, Mal - L-malate, malCoA- malonyl-CoA, Oaa- oxaloacetate, Ser- L-serine, Succ- succinate, SucCoA- succinyl-CoA.(continuation).

References

1. Satrustegui J, Bautista J, Machado A (1983) NADPH/NADP⁺ ratio: regulatory implications in yeast glyoxylic acid cycle. *Mol Cell Biochem* 51:123–127.
2. Canelas AB, van Gulik WM, Heijnen JJ (2008) Determination of the cytosolic free NAD/NADH ratio in *Saccharomyces cerevisiae* under steady-state and highly dynamic conditions. *Biotechnol Bioeng* 100:734–43.
3. Huh W-K, Falvo J V, Gerke LC, et al. (2003) Global analysis of protein localization in budding yeast. *Nature* 425:686–91.
4. Chen Y, Siewers V, Nielsen J (2012) Profiling of cytosolic and peroxisomal acetyl-CoA metabolism in *Saccharomyces cerevisiae*. *PLoS One* 7:e42475.
5. De Virgilio C, Bürckert N, Barth G, et al. (1992) Cloning and disruption of a gene required for growth on acetate but not on ethanol: the acetyl-coenzyme A synthetase gene of *Saccharomyces cerevisiae*. *Yeast* 8:1043–51.