THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Engineering yeast shikimate pathway towards production of aromatics: Rational design of a chassis cell using systems and synthetic biology

Nils Jonathan Helmuth Averesch

Dipl. Ing.

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2016

School of Chemical Engineering

Abstract

Aromatic building blocks are amongst the most important bulk feedstocks in the chemical industry. As these compounds are commonly derived from petrochemistry, obtaining them is becoming more and more a matter of costs and sustainability. Biochemistry gives rise to a wealth of compounds that can potentially replace current petroleum-based chemicals or be used for novel materials. The aromatic compounds *para*-aminobenzoic acid (pABA) and *para*-hydroxybenzoic acid (pHBA) and the aromatics derived compound *cis,cis*-muconic acid (ccMA) can be precursors for, but are not limited to, terephthalic or/and adipic acid. These are essential feedstocks for the production of PET and nylon. The three compounds can be derived from the shikimate pathway, an anabolic pathway leading to the biosynthesis of aromatic amino acids, present in certain prokaryotes and eukaryotes, including fungi. By combination of metabolic modelling with genetic engineering, a microbial production system based on the yeast *Saccharomyces cerevisiae* can be designed, which effectively channels flux into the target compounds.

In order to develop a competitive bio-based process, yields, titers and rates need to be maximized. While productivity or rates in a process can be altered using genetic engineering, carbon yield, and pathway feasibility are stoichiometrically and thermodynamically predetermined. Both limitations need to be considered when designing a microbial production system.

For formation of adipic acid and precursors many bio-based routes exists. More rational than just picking one for *in vivo* studies rather all available biochemical pathways were examined *in silico* using metabolic modelling. To compare theoretical yields and reaction thermodynamics an interface that allowed network-embedded thermodynamic analysis of elementary flux modes was developed. This allowed distinguishing between thermodynamically feasible and infeasible flux distributions. Feasible maximum theoretical product carbon yields were substantially different in *E. coli* and *S. cerevisiae* metabolic models and ranged from 32% to 92%. Further, many pathways appeared to be restricted by a thermodynamic equilibrium lying on the substrate side, some even infeasible.

The only routes that deliver significant product yields and were thermodynamically favoured were shikimate pathway based. Being currently of strong scientific interest, recent implementations of these pathways in *E. coli* and *S. cerevisiae* were evaluated and strain construction strategies optimized, using the concept of constrained minimal cut-sets. Especially in *S. cerevisiae* a single non-obvious knock-out target allowed coupling of growth to product formation; in particular, the deletion of the pyruvate kinase reaction resulted in a minimum yield constraint of 28%.

Though unique to shikimate pathway, this strategy is transferrable to other products which are derived from chorismate and also involve the formation of pyruvate as a by-product. This applies to pHBA. With further optimizations, the strategy was applied *in vivo* for the production of this compound in *S. cerevisiae*. As the pyruvate kinase knock-out entails a growth defect on glucose, a synthetic circuit was used, which allowed conditional knock-down and activation of the determined genetic modifications and by this dynamic control of the phenotype. Thus, production could be separated from growth and it could be proven that the *in silico* determined genetic intervention strategy holds valid *in vivo*, resulting in a 1.1 mM final product titer.

Further, production of pABA from shikimate pathway in *S. cerevisiae* was investigated and optimized: Several alleles from different yeast strains of the genes (*ABZ1* and *ABZ2*) for pABA formation from chorismate were screened, using a strain genetically engineered to channel flux to chorismate. *ABZ1* of AWRI1631 and *ABZ2* of QA23 delivered the highest pABA production. To further increase production, the impact of carbon-source on product yield was investigated *in silico* using metabolic modelling. It was found that combined glycerol-ethanol was a superior carbon-source than glucose, glycerol or ethanol alone, especially when employed in molar ratios between 1:2 and 2:1. This could be confirmed *in vivo* with carbon yields reaching up to 2.64%. A fed-batch process on glycerol-ethanol delivered total aromatics titers as high as 1.73 mM.

It could be shown that feasibility and viability of adipic acid production greatly depends on the pathway and the organism chosen for engineering. Weaknesses of existing strain construction strategies for ccMA production could be identified and a radical optimization strategy was determined.

Inferences from metabolic modelling were proven experimentally for pHBA and pABA production. The obtained concentrations and yields are the highest in *S. cerevisiae* to date and among the highest in a microbial production system, underlining the potential of yeast as a cell factory for renewable aromatic feedstocks.

Declaration by author

This thesis is composed of my original work and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Peer-reviewed papers:

Quorum-Sensing Linked RNA interference for Dynamic Metabolic Pathway Control in *Saccharomyces cerevisiae*. Thomas C. Williams, **Nils J. H. Averesch**, Gal Winter, Manuel R. Plan, Claudia E. Vickers, Lars K. Nielsen, Jens O. Krömer. *Metab. Eng.* **2015**, 10.1016/j.ymben.2015.03.008

Tailoring strain construction strategies for muconic acid production in *S. cerevisiae* and *E. coli*. **Nils J. H. Averesch**, Jens O. Krömer. *Metab. Eng. Commun.* **2014**, 10.1016/j.meteno.2014.09.001

In vivo instability of chorismate causes substrate loss during fermentative production of aromatics. Gal Winter, **Nils J. H. Averesch**, Dariela A. Nunez-Bernal, Jens O. Krömer. *Yeast.* **2014**, 10.1002/yea.3025

Production of *para*-aminobenzoic acid from different carbon-sources in engineered *Saccharomyces cerevisiae*. **Nils J. H. Averesch**, Gal Winter, Jens O. Krömer. Submitted to *Microb. Cell. Fact.*, under review

Combined elementary flux mode and network-embedded thermodynamic analysis of microbial adipic acid production. **Nils J. H. Averesch**, Verónica S. Martínez, Lars K. Nielsen, Jens O. Krömer. Previously submitted to *Metab. Eng.*, under revision

Conference abstracts:

Model-guided strain engineering enables coupling of product formation to growth. **Nils J. H. Averesch**, Jens O. Krömer, Gal Winter. *NZMS*, Wellington, **2014**

Combining Elementary Mode Analysis with a Network Embedded Thermodynamic Approach for Analysis of Microbial Adipic Acid Production. **Nils J. H. Averesch**, Verónica S. Martínez, Jens O. Krömer. *ME-X*, Vancouver, **2014**

Publications included in this thesis

Combined elementary flux mode and network-embedded thermodynamic analysis of microbial adipic acid production. **Nils J. H. Averesch**, Verónica S. Martínez, Lars K. Nielsen, Jens O. Krömer. Previously submitted to *Metab. Eng.*, under revision – incorporated as Chapter 2.1.

Contributor	Statement of contribution
Nils J.H. Averesch (Candidate)	Designed the experiments (70%) Conducted the experiments (60%) Analysed the data (60%) Drafted the paper (100%)
Verónica S. Martínez	Designed the experiments (20%) Conducted the experiments (35%) Analysed the data (30%) Reviewed and edited the paper draft (20%)
Lars K. Nielsen	Reviewed the paper draft (10%)
Jens O. Krömer	Designed the experiments (10%) Conducted the experiments (5%) Analysed the data (10%) Reviewed and edited the paper draft (70%)

Tailoring strain construction strategies for muconic acid production in *S. cerevisiae* and *E. coli*. **Nils J. H. Averesch**, Jens O. Krömer. *Metab. Eng. Commun.* **2014**, 10.1016/j.meteno.2014.09.001 – incorporated as Chapter 2.2.

Contributor	Statement of contribution
Nils J.H. Averesch (Candidate)	Jens O. Krömer and Nils J.H. Averesch jointly conceived the study and edited the manuscript. Nils J.H. Averesch designed the metabolic networks, conducted the analysis, and drafted the manuscript. Both authors read and approved the final manuscript.
Jens O. Krömer	

Quorum-Sensing Linked RNA interference for Dynamic Metabolic Pathway Control in *Saccharomyces cerevisiae*. Thomas C. Williams, **Nils J. H. Averesch**, Gal Winter, Manuel R. Plan, Claudia E. Vickers, Lars K. Nielsen, Jens O. Krömer. *Metab. Eng.* **2015**, 10.1016/j.ymben.2015.03.00 – incorporated as Chapter 2.3.

Contributor	Statement of contribution
Thomas C. Williams	Designed the experiments (60%) Conducted the experiments (90%) Analysed the data (70%) Drafted the paper (80%)
Nils J.H. Averesch (Candidate)	Designed the experiments (20%) Conducted the experiments (10%) Analysed the data (20%) Drafted the paper (20%) Reviewed and edited the paper draft (10%)
Gal Winter	Designed the experiments (5%) Reviewed and edited the paper draft (10%)
Manuel R. Plan	Analysed the data (10%)
Claudia E. Vickers	Designed the experiments (2.5%) Reviewed and edited the paper draft (10%)
Lars K. Nielsen	Designed the experiments (2.5%) Reviewed and edited the paper draft (10%)
Jens O. Krömer	Designed the experiments (10%) Reviewed and edited the paper draft (60%)

Production of *para*-aminobenzoic acid from different carbon-sources in engineered *Saccharomyces cerevisiae*. **Nils J. H. Averesch**, Gal Winter, Jens O. Krömer. Submitted to *Microb. Cell. Fact.*, under revision – incorporated as Chapter 2.4.

Contributor	Statement of contribution
Nils J.H. Averesch (Candidate)	Designed the experiments (80%) Conducted the experiments (95%) Analysed the data (75%) Drafted the paper (100%)
Gal Winter	Designed the experiments (10%) Conducted the experiments (5%) Analysed the data (5%) Reviewed and edited the paper draft (10%)
Jens O. Krömer	Designed the experiments (10%) Analysed the data (20%) Reviewed and edited the paper draft (90%)

Contributions by others to the thesis

Program for NET analysis of EFMs scripted by Verónica S. Martínez

Synthetic quorum sensing circuit developed by Thomas C. Williams

Analytics (HPLC) performed by Manuel R. Plan (Metabolomics Australia)

Statement of parts of the thesis submitted to qualify for the award of another degree

None

Acknowledgements

This thesis was partially funded by the Australian Research Council (DE120101549), tuition fees and living allowance during candidature were provided by the UQI scholarship and CEMES funding.

Thanks to my friends and family and everybody who helped and/or supported me in any possible way during the time of PhD candidature.

In particular, my gratitude goes to Jens O. Krömer for competent supervision whenever I need it (regardless of whether I was aware of that or not) but also the freedom to prioritise my research efforts by myself. Further, I appreciate the always open door whatever the matter.

Further gratitude goes to Gal Winter, Nicolas Lekieffre, Verónica S. Martínez, Thomas C. Williams, Manuel R. Plan, Lars K. Nielsen, Oliver Kayser, Frauke Kracke and Gabriele A. König-Averesch.

Stay Punk – Not More, Not Less

Keywords

Yeast, shikimate, pABA, pHBA, muconic, adipic, EMA, NET-analysis, knock-out strategy

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code	ANZSRC description	ANZSRC % share
060113	Synthetic Biology	20
060114	Systems Biology	30
100305	Industrial Microbiology (incl. Biofeedstocks)	50

Fields of Research (FoR) Classification

FoR code	FOR description	FoR % share
0601	Biochemistry and Cell Biology	25
0605	Microbiology	25
1003	Industrial Biotechnology	50

Table of Contents

List of Figures & Tables

List of Abbreviations

2-OXO = 2-oxoglutarate, 2,3-Ddha-CoA = 2,3-didehydroadipyl-CoA, AAA = aromatic amino acids, ADH = acetaldehyde dehydrogenase, ANTH = anthranilate, CDM = chemically defined medium, ccMA = *cis,cis*-muconic acid, cMCs = constrained minimal cut-sets, CoA = Coenzym A, DAHP = 3-deoxy-arabinoheptulosonate 7-phosphate, DHBA = 2,3-dihydroxybenzoate, DHQ = 3-dehydroquinate, DHS = 3-dehydroshikimate, E4P = erythrose 4-phosphate, EDP = Entner-Doudoroff pathway, EFM = elementary flux modes, EMA = elementary mode analysis, EPSP = 5-enolpyruvylshikimate-3-phosphate, ETH = Ethanol, F6P = fructose-6-phosphate, FA = fatty acid, FAS = fatty acid synthesis, FRD = fumarate reductase, FUM = fumarate, FVA = flux variability analysis, G6P = glucose-6-phosphate, G6PD = glucose-6-phosphate dehydrogenase, GA3P = glyceraldehyde 3-phosphate, GLC = Glucose, GLY = Glycerol, GPDH = glycerol-3-phosphate dehydrogenase, EDA = 2-dehydro-3-deoxy-phosphogluconate aldolase, LDH = lactate dehydrogenase, LOD = level of detection, MAE = malic enzyme, MAL = malate, MDH = malate dehydrogenase, NET-analysis = network embedded thermodynamic analysis, NET-EMA = network embedded thermodynamic elementary mode analysis, OAA = oxaloacetate, OD_{660} = optical density at 660 nm, ORF = open reading frame, P = phosphate, PP = pyrophosphate, pABA = *para*-aminobenzoic acid, PCA = protocatechuate, PCK = phosphoenolpyruvate carboxykinase, PEA = phenethyl alcohol, PEP = phosphoenol pyruvate, PET = polyethylene terephthalate, pHBA = *para*-hydroxybenzoate, PHE = phenylalanine, $pH_{ext.}$ = extracellular pH, $pH_{int.}$ = intracellular pH, PPP = pentose phosphate pathway, PPS = phosphoenolpyruvate synthase, PPY = phenyl pyruvate, PTS = phosphotransferase system, PYK = pyruvate kinase, PYR = pyruvate, QS = quorum sensing, RNAi = RNA interference, S3P = shikimate-3-phosphate, SA = salicylate, SFL = summed fractional labelling, SNP = single-nucleotide polymorphism, SUCC = succinate, SUCC-CoA = succinyl coenzyme A, TCA-cycle = tricarboxylic acid cycle, TRP = tryptophan, TYR = tyrosine, WT = wild-type, Y_{max} = maximum theoretical carbon yield, Y_{min} = minimum theoretical carbon yield, $\Delta_f G$ = Gibbs free energy of formation; $\Delta_r G$ = Gibbs free energy of reaction; $\Delta_r G'^{\circ}$ = transformed standard Gibbs free energy of reaction

Main text of Thesis

1. Introduction

There is a worldwide push in the chemical industry to move from fossil fuel dependency to renewable feedstocks. This not only serves the goal to move towards a more sustainable society but also helps to maintain current standards of living. For this the availability and commercial viability of synthetic materials is essential. Biotechnology is anticipated to be a key to solving these problems. In the polymer industry aromatic and aromatic-derived compounds are among the most important building blocks. The challenge is the efficient and fast production of the products of interest in biological systems. Many microorganisms are well studied and understood, their biochemical pathways are extremely versatile, giving rise to a wealth of diverse organic compounds. Further, they have a rapid metabolism, allowing them to grow to high density rapidly. This makes them very suitable for biotechnology and attractive for production on industrial scale. Here the focus is on the development of a microbial production system for shikimate pathway derivatives as chemical feedstocks, using a systems and synthetic biology approach.

1.1 General Introduction

1.1.1 Industrial significance of bio-derived building blocks

There is currently a strong interest in the metabolic engineering community to develop biological replacements for fossil-fuel based feedstocks for chemical industry. It has been predicted that bio-feedstocks could increase to 17% of the global chemical business by 2025, equivalent to 425 billion $US. Recent developments spawn bio-replacement chemicals of the classes of diamines, dialcohols and dicarboxylates for nylon and polyesters products. Unfortunately to date, the majority of these bio-replacement compounds are hardly cost competitive. For many more important compounds, there is also not even a bio-process available. This highlights the major hurdle for biotechnological production: outperforming existing chemical synthesis. The cost advantage in comparison to petrochemical-based adipic acid has been predicted to possibly reach 20 – 30% (depending on the feedstock used) (Guzman, 2010), which indicates investment into the cleaner bio-production, though this cost advantage will strongly depend on the pathway, substrates, and current oil price.

This project targets aromatic and di-carboxylic organic acids, which can be derived from the shikimate metabolic pathway present in certain micro-organisms for example yeasts,

or bacteria. In particular target compounds are (i) adipic acid, which can be derived from muconic acid (ccMA) via aromatic intermediates of the shikimate pathway and (ii) aromatic precursors like *para*-hydroxybenzoic acid (pHBA) and *para*-aminobenzoic acid (pABA).

1.1.1.1 The polymer precursor adipic acid and potential precursors

Adipic acid (hexanedioic acid, Figure 1) is a main precursor for the production of nylon-6,6 and polyurethanes (Musser, 2000). The 2010 global production capacity of adipic acid has been estimated to be about 2.6 million t with an annual growth rate of about 3% (Centi *et al.*, 2009). About 65% of the adipic acid is used for the production of nylon-6,6, 24% for polyurethanes (IHS Chemical, 2012, IHS Chemical, 2012) and the remainder for a broad range of other applications, including resins, reactants for plasticizers, food acid, and lubricants (MCGroup, 2011). Adipic acid is mainly produced through benzene reduction to cyclohexane followed by a two-step oxidation via cyclohexanol/cyclohexanone and reaction with nitric acid (EPA, 1994, Musser, 2005). This method achieves very high yields, however, the precursors are all petrochemistry based and thus non-sustainable. Also, considerable amounts of NO_x are generated, which have an additional environmental impact; especially the greenhouse gas N_2O is a large by-product in this process. Adipic acid production accounts for up to 10% of the total global anthropogenic N_2O emissions per year and significantly contributes to global warming and depletion of the ozone layer (Sato *et al.*, 1998, Mainhardt & Kruger, 2000, Alini *et al.*, 2007).

Biotechnological production of adipic acid offers a promising alternative to chemical synthesis. Recently emerging papers and reviews emphasize the strong scientific and commercial interest to develop a biological replacement for fossil fuel based adipic acid (Chen & Nielsen, 2013, Polen *et al.*, 2013, Van Duuren & Wittmann, 2014, Xie *et al.*, 2014, Bart & Cavallaro, 2015, Bart & Cavallaro, 2015, Deng *et al.*, 2016). Basically, there are three approaches that can be distinguished: (i) direct production of adipic acid, (ii) production of the unsaturated adipic acid precursor *cis,cis*-muconic acid (Figure 1) and (iii) production of D-glucaric acid. Muconic and glucaric acid be converted to adipic acid via chemo-catalytic hydrogenation, which can only be conducted under high pressure of H_2 in a Platinum catalysed reaction (Polen *et al.*, 2013, Deng *et al.*, 2016). Nevertheless for the conversion of *cis,cis*-muconic acid (ccMA) to adipic acid an efficient conversion yield of 97% [$mol_{adipate}/mol_{ccMA}$] has been reported (Niu *et al.*, 2002). The approach to derive the precursor ccMA from shikimate pathway is currently of particular scientific attraction as ccMA can also be used for the production of terephthalic acid (Burk *et al.*, 2011) and options to obtain it are abundant (Xie *et al.*, 2014).

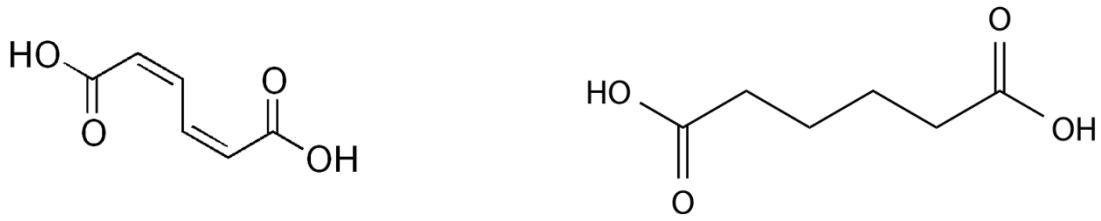


Figure 1: *cis,cis*-muconic acid and adipic acid

1.1.1.2 The potential polymer precursors pHBA and pABA

"Polyesters" like PET (polyethylene terephthalate) contribute 18% to the world polymer production. PET is composed of ethylene glycol and terephthalic acid (J. Sheehan, 2000). Terephthalic acid production accounts for < 50 Mio t/a equivalent to 80 billion $US value. Potential bio-replacements for terephthalic acid are *para*-hydroxybenzoic acid (pHBA) and *para*-aminobenzoic acid (pABA). pHBA and pABA are natural metabolites in *S. cerevisiae* and can both be derived from shikimate pathway.

pHBA (Figure 2) is used in liquid crystal polymers and has an estimated market value of approximately $150 million per annum (Krömer *et al.*, 2013). It is the base material for parabens, preservatives in the cosmetic and pharmaceutical industry (Kluczyk *et al.*, 2002). pABA (Figure 2) is a compound with versatile applicability, it is already being used use as a crosslinking agent for resins and dyes, as a precursor in the pharmaceutical industry and as a therapeutic itself (e.g. for the drug POTABA®). It is also a good UV filter as it absorbs UVB radiation (Krömer *et al.*, 2013). Further pHBA and pABA could be chemically converted to terephthalic acid (e.g. via Sandmeyer reaction) making it a potential feedstock for polyethylene terephthalate (PET) production. PET is used for packaging as well as clothing and recently also in the auto industry (Research, 2014). The global market volume for PET packaging was nearing 16 Mt valued at 48.1 billion USD in 2014 (Pira, 2014). This is expected to rise to 19.9 Mt and 60 billion USD in 2019 (Pira, 2014). The global market volume for bio-based PET exceeded 540 kt in 2012 and is expected to rise sharply till 2020, with a doubling in 2015 and a tripling in 2016 in respect to 2014. While current bio-based PET is claimed to be up to 30% sustainable (Heinz, 2011) in fact only the monoethylene glycol share of the copolymer is derived from renewable feedstocks, while the terephthalic acid component, the major part of PET by volume and cost, is still petroleum based. Therefore industrial interest in developing a fully sustainable PET product is high and substantial investments are being made in order to establish bio-based terephthalic acid production (Burk *et al.*, 2011, Osterhout *et al.*, 2013, Research, 2014).

pABA may also be chemically converted to *para*-phenylenediamine (e.g. via Kochi- or Hunsdiecker reaction followed by nucleophilic substitution), which is (besides terephthalic acid) the second building block of the aramid fibre Kevlar®. Global demand for these materials, which combine high strength with low weight as well as chemical and heat resistance properties, was estimated to 74.5 kt in 2014 and is expected to rise to 110 kt in 2020 with an estimated value of 4.7 billion USD (Markets, 2014).

Figure 2: *para*-hydroxybenzoic acid (pHBA) and *para*-aminobenzoic acid (pABA)

1.1.2 Yeast in Biotechnology

Yeast, the oldest organism used in biotechnology, is currently regaining industrial relevance (Branduardi & Porro, 2012, Mattanovich *et al.*, 2014), as shown by new biotech-processes being developed based on yeast, e.g. artemisinin / farnesene by Amyris, vanillin by Evolva and dodecanedioic & adipic acid by Verdezyne (Hansen *et al.*, 2009, de Jong *et al.*, 2012, Paddon *et al.*, 2013, Polen *et al.*, 2013, Paddon & Keasling, 2014), and even old ones being moved to yeast e.g. lactic acid by Cargill (Datta & Henry, 2006). This is due to the manyfold cost and operational advantages for commercial processes that the use of these very robust microbes offer: Above all yeasts are tolerant to low pH and temperature and unsusceptible to phage contamination (Fletcher *et al.*, 2015). This also applies to the baker's yeast *S. cerevisiae*, which is popular in both, food industries and at laboratory scale, as it is well-studied and non-hazardous (Agency, 2012). Recent advances in metabolic engineering of *S. cerevisiae* have shown that systems and synthetic biology tools are advanced and readily available nowadays, allowing rational construction of capable production strains (Li & Borodina, 2014, Fletcher *et al.*, 2015). Through combination of metabolic modelling and genetic engineering *S. cerevisiae* can be analysed and tailored *in silico* as well as *in vivo* (Kim *et al.*, 2012, Borodina & Nielsen, 2014). This allows even very complex natural product biosynthetic pathways to be reconstructed in this organism (Galanie *et al.*, 2015, Kayser & Averesch, 2015).

1.1.2.1 Yeast shikimate pathway and derived compounds

The shikimate pathway is the central metabolic route leading to the formation of tryptophan (TRP), tyrosine (TYR) and phenylalanine (PHE). It starts with the condensation of intermediates of glycolysis and pentose phosphate pathway (PPP), phosphoenol pyruvate (PEP) and erythrose 4-phosphate (E4P), respectively, which enter the pathway through a series of condensation and redox reactions via 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP), 3-dehydroquinate (DHQ), 3-dehydroshikimate (DHS) to shikimate. Shikimate is subsequently converted to the central branch point metabolite chorismate under ATP hydrolysis and introduction of a second PEP. From there the pathway branches off to anthranilate and prephenate leading to aromatic amino acid (AAA) and pABA synthesis, the latter being a precursor for folate metabolism (Botet *et al.*, 2007). The pathway exclusively exists in plants and microorganisms. An overview of the pathway and it's regulation in the baker's yeast *Saccharomyces cerevisiae* is given in Figure 3.

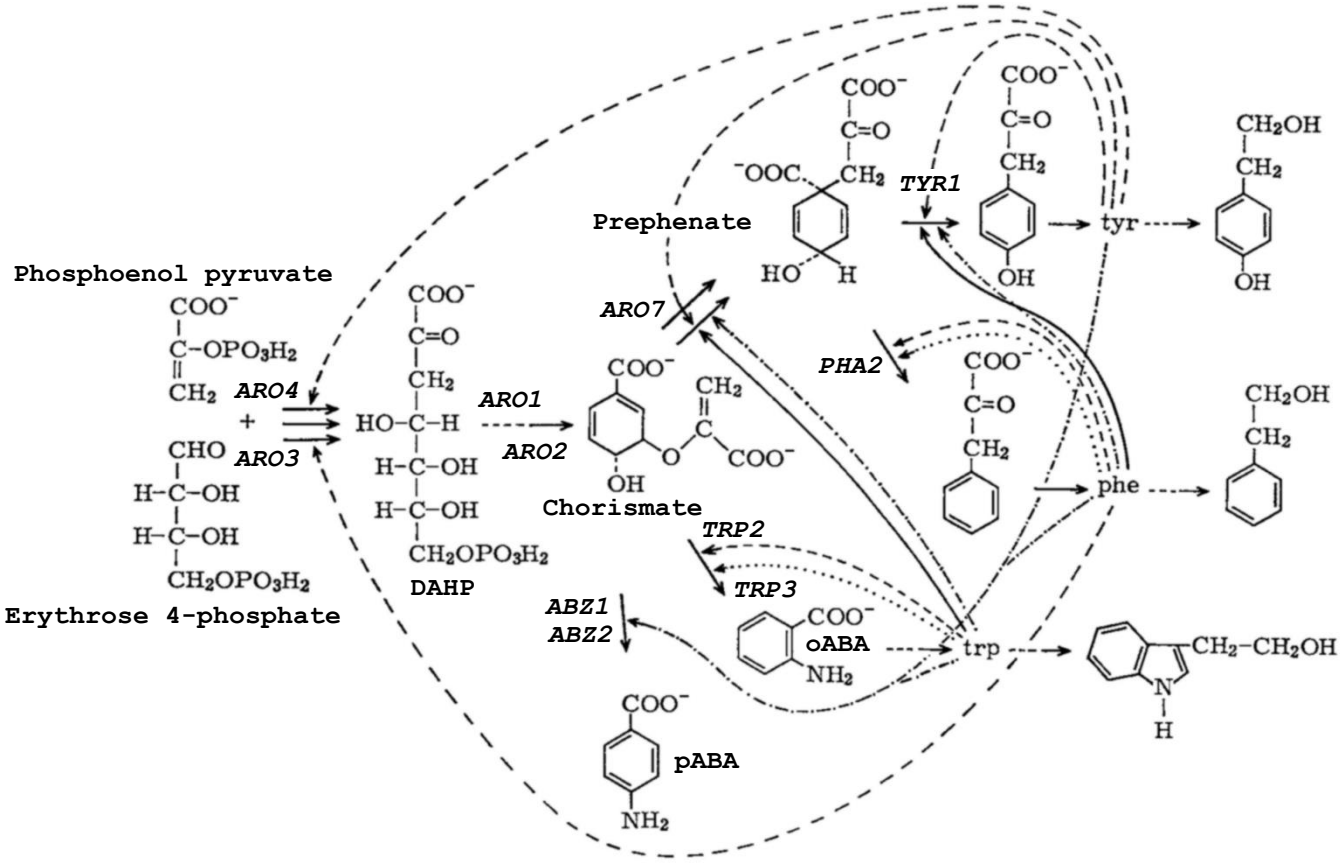


Figure 3: Biosynthesis of aromatic amino acids in *S. cerevisiae* including regulatory elements - - - feedback inhibition, ⋯ repression, ⎯ activation, -·-·- induction (reproduced from (Lingens *et al.*, 1967)).

The shikimate pathway also leads to folate and ubiquinone (coenzyme Q) synthesis via the precursor pABA (Marbois *et al.*, 2010). Outgoing from chorismate, pABA is synthesised in two subsequent reactions catalysed by aminodeoxychorismate synthase and aminodeoxychorismate lyase, the respective yeast genes are *ABZ1* and *ABZ2*. The former has similarity to *Escherichia coli* pABA synthase components PabAp and PabBp, indicating that it is a bi-functional enzyme (Edman *et al.*, 1993). The aminodeoxychorismate synthase transfers an amino-group from glutamine to the *para*-position of chorismate, thus forming glutamate and 4-amino-4-deoxychorismate. The aminodeoxychorismate lyase then cleaves the ester-bond between the ring and the C3-body releasing pyruvate and pABA. Generally pABA is a compound with a relatively low concentration in the cell, nevertheless, it is essential to the organism. This is for example exploited when screening sulfa-drugs, as these compete with pABA in folate biosynthesis (Castelli *et al.*, 2001). In other organisms, like for example *E. coli*, the first step in ubiquinone biosynthesis is the formation of pHBA in a single step bioconversion directly from the central shikimate pathway intermediate chorismate. The reaction is catalysed by chorismate pyruvate-lyase which is encoded by the gene *ubiC* (Nichols & Green, 1992). It is remarkable that pHBA can also be detected in yeast though no chorismate lyase analogue has been described to date (Marbois *et al.*, 2010, Pierrel *et al.*, 2010), indicating that either the responsible enzyme has yet to be discovered, or the compound is obtained from a different precursor. It has been proposed that this precursor is rather 4-hydroxyphenylpyruvate then catechol, giving rise to a second pathway to ubiquinone synthesis in yeast (Marbois *et al.*, 2010).

1.1.2.2 Regulation of the yeast shikimate pathway

Regulation of the shikimate pathway in *S. cerevisiae* is manifold (Lingens *et al.*, 1967) and yet to be fully understood. In general *GCN4*, the transcriptional activator of amino acid biosynthetic genes, controls AAA biosynthesis, tightly regulating expression at transcriptional level in response to amino acid starvation (Braus, 1991).

The initial step of the shikimate pathway is catalysed by two DAHP synthase isozymes (Meuris, 1973, Kunzler *et al.*, 1992), encoded by *ARO4* and *ARO3* and regulated through feedback inhibition by TYR and PHE, respectively. Moreover, the *ARO4* coded enzyme is feedback-inhibited not only by TYR but also high concentrations of PHE and TRP (Kunzler *et al.*, 1992). The five subsequent bioconversions are catalysed by a penta-functional protein, a conglomerate of mono-functional domains, which is expressed from the gene *ARO1* (Duncan *et al.*, 1987).

The gene *ARO7* expresses chorismate mutase which catalyses the conversion of chorismate to prephenate where the biosynthesis route to TYR and PHE branches off, it is also regulated by *GCN4* (Schmidheini *et al.*, 1990). However, this step appears to proceed also spontaneously, as a study on the deletion of *ARO7* shows (Winter *et al.*, 2014). The next step to tyrosine is catalysed by prephenate dehydrogenase, encoded by *TYR1*. The expression is dependent on PHE levels, regulating the equilibrium between the two products of this pathway, TYR and PHE (Mannhaupt *et al.*, 1989).
The initial step to tryptophan biosynthesis is catalysed by anthranilate synthase, encoded by *TRP2* which forms a multifunctional hetero-oligomeric complex with indole-3-glycerol phosphate synthase encoded by *TRP3*. As well *TRP3* (Prantl *et al.*, 1985) as also *TRP2* (Krömer *et al.*, 2013) may have second functions, as deletion leads to a heavy growth rate reduction regardless of the supplementation of TRP.

1.1.3 Previous approaches on production of aromatics and derived compounds

The aromatic nature of the shikimate pathway intermediates gives rise to the idea to derive bio-replacements for fossil-fuel-derived aromatics. By metabolic engineering abundance of certain suitable compounds may be increased, while draining flux from other branches off the pathway. This was used previously to produce ccMA as well as pHBA and folate (metabolic successor of pABA).

1.1.3.2 Bio-based production of adipic acid and potential precursors

After early work on muconic acid production in the mid 1990s (Frost & Draths, 1996, Frost & Draths, 1997) the target adipic acid has recently moved back into the spotlight with several companies (Genomatica, Verdezyne) filing patents (Boussie *et al.*, 2010, Burgard *et al.*, 2010, Burgard *et al.*, 2012, Picataggio & Beardslee, 2012, Raemakers-Franken *et al.*, 2012) and the metabolic engineering research community re-focussing on this compound (Niu *et al.*, 2002, IHS Chemical, 2012, Weber *et al.*, 2012, Curran *et al.*, 2013, Sun *et al.*, 2013, Lin *et al.*, 2014, Sun *et al.*, 2014, Yu *et al.*, 2014, Babu *et al.*, 2015, Clomburg *et al.*, 2015, Deng & Mao, 2015, Horwitz *et al.*, 2015, Reizman *et al.*, 2015, Sengupta *et al.*, 2015, Vardon *et al.*, 2015). Considering the extensive work that has been done it is not surprising that there is a wide variety of proclaimed metabolic routes to produce adipic acid or the precursors muconic or glucaric acid (Figure 4). Each major pathway has several different sub-routes that utilize alternative biochemical conversions, which can significantly influence titre, yield and productivity.

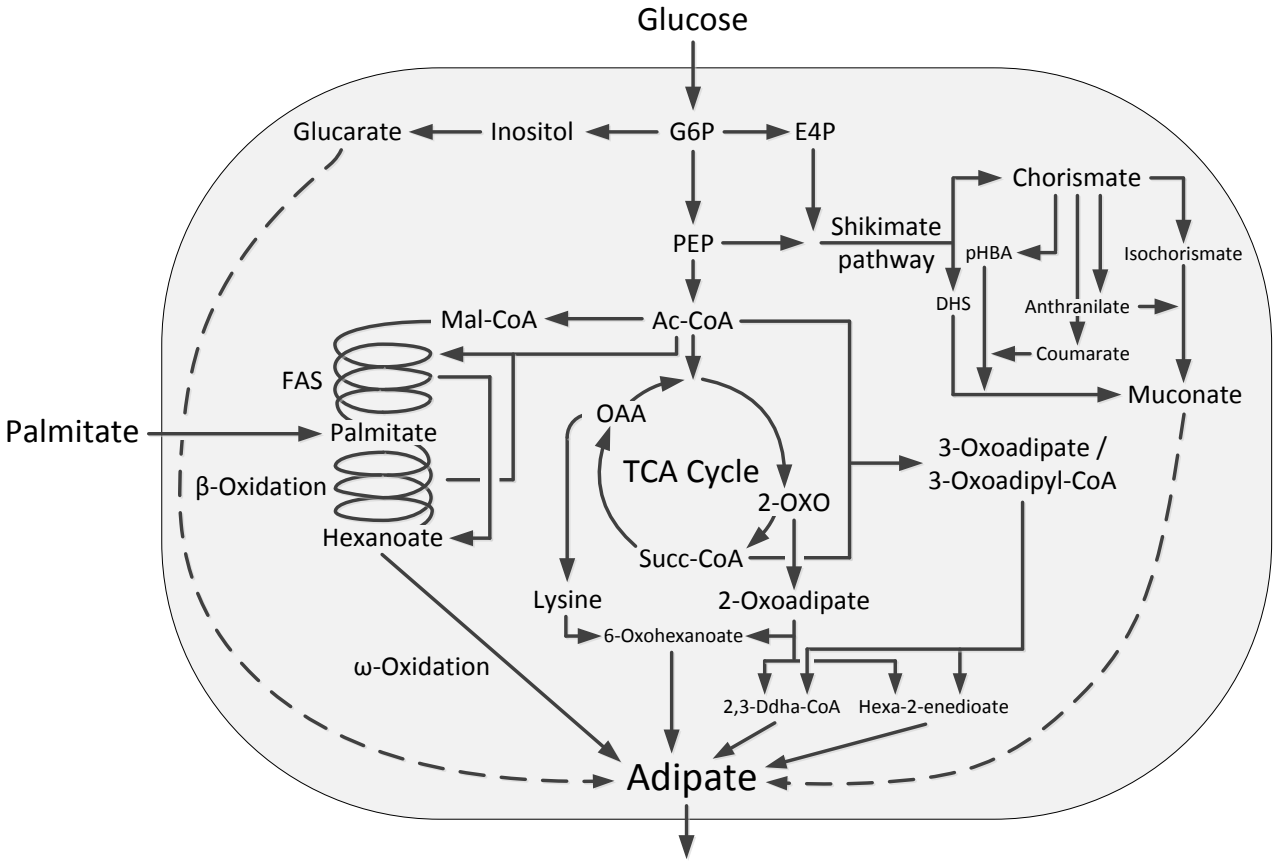


Figure 4: Overview of metabolic routes to adipic acid and its precursors muconic and glucaric acid. Biochemical reduction of muconic and glucaric acid to adipic acid is indicated as hypothetical.

Production of muconic acid was initially reported in *Pseudomonas putida* deriving it from benzoate. Titers of 7.4 g/L with a maximum yield of 91% (Schmidt & Knackmuss, 1984), 13.5 g/L with productivities ranging from 0.8 to 5.5 g/(L×h) (Choi *et al.*, 1997), and 32 g/L (Bang & Choi, 1995) were achieved. Though effective in terms of maximum titer, yield, and productivity, benzoate can however hardly be considered a sustainable feedstock.

De novo production of muconic acid from glucose via the shikimate pathway (Figure 5) has been pioneered in *E. coli* (Draths & Frost, 1994). The optimized process showed high maximum titers (36.8 g/L) after 48 h and yields of 22% (mol/mol) (Niu *et al.*, 2002). The pathway was established in two steps outgoing from dehydroshikimate (DHS), via protocatechuate and catechol (Figure 5 a) by means of the *aroZ* gene from *Klebsiella pneumoniae* encoding dehydroshikimate dehydratase, the *aroY* gene from *K. pneumoniae* encoding protocatechuate (PCA) decarboxylase and the *catA* gene from *Acinetobacter calcoaceticus* encoding catechol 1,2-dioxygenase. The levels of the entry metabolites to shikimate pathway PEP and E4P were increased and a feedback inhibition resistant DAHP synthase was overexpressed. Further, blocking the pathway below DHS through deletion of the DHS dehydrogenase gene *aroE*, directed carbon flux to ccMA.

The *E. coli* result astonishingly remains unrivaled: Though no less elaborate, recent approches employing the same pathway (Figure 5a) in the *S. cerevisiae* were so far unable to achieve comparable titers (1.56 mg/L (Weber *et al.*, 2012), 141 mg/L (Curran *et al.*, 2013)). In (Weber *et al.*, 2012) partial deletion of *ARO1* (the *aroE* analogous domain) blocked the conversion of 3-dehydroshikimate into shikimate. A DHS dehydratase from *Bacillus thuringiensis* (*aroZ*), a PCA decarboxylase composed of three different subunits encoded by the genes B, C and D taken from *K. pneumoniae* (*aroY*) and a catechol 1,2-dioxygenases from *Acinetobacter radioresistens* (*catA*) composed the pathway. The bottleneck here is the first step: Only very low PCA levels < 7 mg/L could be detected and therefore limited the overall titre (Weber *et al.*, 2012). In (Curran *et al.*, 2013) the DHS dehydratase was taken from *Podospora anserina* and the PCA decarboxylase from *Enterobacter cloacae*. A catechol 1,2-dioxygenase from *Candida albicans* completed the pathway. Knock-out of *ARO3* and over-expression of a feedback-resistant *ARO4* increased overall flux to aromatics. Further optimizations on that were the deletion of *ZWF1* and over-expression of *TKL1*. The former intended to increase flux to the shikimate pathway in order to increase conversion of G6P to E4P, by blocking the channelling of G6P into the oxidative branch of the pentose phosphate pathway. The latter favoured the conversion of pentoses into E4P by increased transketolase levels.

Also, production of muconic acid from glucose and glycerol in *E. coli* via a different branch off the shikimate pathway (Figure 5b) could only achieve 390 mg/L muconic acid (Sun *et al.*, 2013). Here another route for the production of *cis,cis*-muconic acid from shikimate pathway was established downstream of chorismate, branching off from anthranilate on the way to TRP synthesis. This was done by cloning an anthranilate 1,2-dioxygenase from *Pseudomonas aeruginosa* (encoded by the genes *antA antB antC*, cloned as an operon) and catechol 1,2-dioxygenase from *Pseudomonas putida* (*catA1*) in *E. coli*. TRP biosynthesis was blocked and over-expression of glutamine synthase resulted in a strengthened glutamine regeneration system favouring anthranilate formation. Two more variations of this pathway (Figure 5c) for the production of ccMA have been developed by this group, both involve the isomerization of chorismate to isochorismate. One proceeds via 2,3-dihydroxybenzoate (Sun *et al.*, 2014), the other one via salicylate (Lin *et al.*, 2014). The routes delivered respective titers of 480 and 1500 mg/L. In the first case (Figure 5c$_I$) this was achieved using a strain with a deletion in *entE*, engineering to channel flux to chorismate as described in another study (Lin *et al.*, 2013), to overexpress the *E. coli* genes *entCBA* (coding for isochorismate synthase, isochorismatase and 2,3-dihydro-2,3-DHBA dehydrogenase) along with a *Klebsiella pneumoniae* 2,3-DHBA decarboxylase and

the same catechol 1,2-dioxygenase as in the previous study (Sun *et al.*, 2013). In the second case (Figure 5c_{II}) an “off the shelf” phenylalanine overproducing *E. coli* strain NST 74 (ATCC31884) with deletions in *pheA* and *tyrA* was amended with the same modifications to increase precursor abundance as described previously (Lin *et al.*, 2013). In addition, the genes for isochorismate synthase from *E. coli* (*entC*), isochorismate pyruvate lyase from *Pseudomonas fluorescens* (*PchB*), an optimized salicylate 1-monoxygenase from *Pseudomonas putida* (*nahG*^{opt}) and the catechol 1,2-dioxygenase as described before (Sun *et al.*, 2013) were overexpressed.

Another route to ccMA leads from chorismate via pHBA to PCA, from here the steps coincide with the DHS-pathway (Figure 5d). It has been shown in *E. coli* that PCA and catechol can be produced this way, with the highest yield being that of catechol at a titre of 451 mg/L (Pugh *et al.*, 2014) (equivalent to 3.69% total carbon product yield). Engineering was also based on the phenylalanine over-producing strain NST 74. Overexpression of chorismate lyase, encoded by *ubiC*, initiated the pathway, a pHBA hydroxylase (*pobA*) from *Pseudomonas aeruginosa* and a protocatechuate decarboxylase from *Enterobacter cloacae* completed it. To further increase flux to the target compounds the chorismate mutase/prephenate dehydratase (*pheA*) was disrupted. Recently the complete pathway outgoing from glucose until ccMA was established (Sengupta *et al.*, 2015). The engineered *E. coli* strain was capable of producing almost 170 mg/L ccMA. The strain had deletions in the genes *ptsH*, *ptsI*, *crr*, and *pykF* to boost shikimate pathway, further, the *E. coli* genes (*aroF*^{FBR}, *aroE*, *aroL* and *ubiC*) for DAHP synthase, shikimate 5-dehydrogenase, shikimate kinase II and chorismate lyase were overexpressed along with a *Pseudomonas putida* pobA, *Klebsiella pneumoniae aroY*, and *Acinetobacter* sp. ADP1 *catA* coding for 4-hydroxybenzoate hydrolyase, protocatechuate decarboxylase, and catechol 1,2-dioxygenase.

A different approach to sustainable ccMA production proceeds outgoing from lignin-derived species (Deng *et al.*, 2016). Here a ccMA titer of 13.5 g/L was obtained with a yield of 74%. This was done by engineering the metabolism of a *Pseudomonas putida* KT2440 strain. In a pathway that uses downstream AAA-metabolism intermediates like coumarate as substrates a route to ccMA was constructed that partially coincides in the last steps with the route via pHBA (Figure 5d).

Another alternative is the production of glucaric acid from sugar. In *E. coli* 2.5 g/L glucaric acid (Moon *et al.*, 2010) could be produced. This was recently further improved by 42% (Reizman *et al.*, 2015).

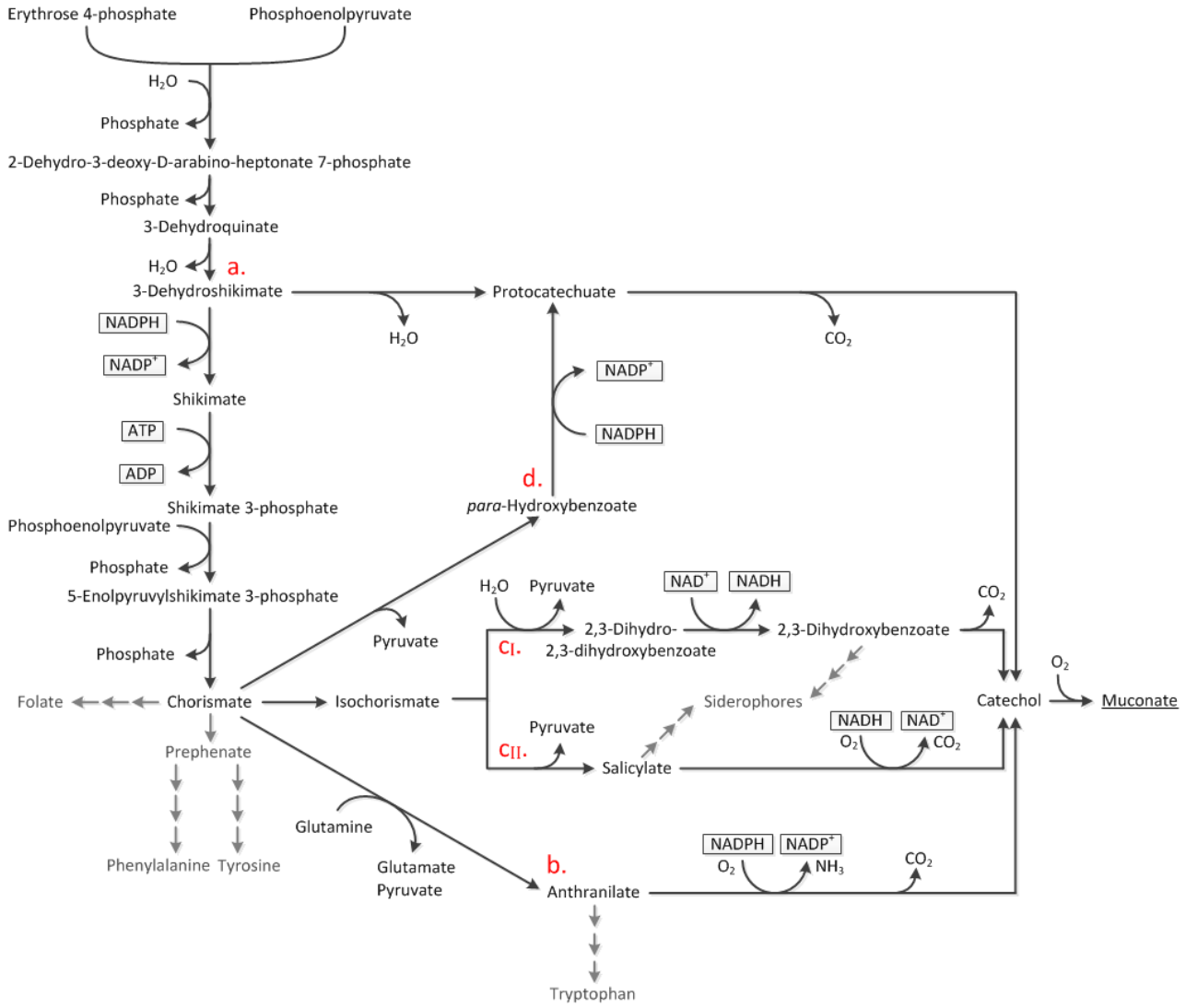


Figure 5: Shikimate pathway with routes to muconic acid: Shikimate pathway outgoing from the precursors E4P and PEP with routes to ccMA including co-factors in each step. Adjacent pathways not involved in ccMA synthesis are greyed out. Triple arrows indicate multiple steps. The DHS-pathway (a) proceeds via protocatechuate outgoing from 3-dehydroshikimate, while the ANTH-pathway (b) branches off further downstream from chorismate via anthranilate. The DHBA-pathway (c_I) and the SA-pathway (c_{II}) both proceed via isochorismate. A route linking the chorismate derived routes with the DHS-pathway is the pHBA-pathway (d). Although all routes lead to the formation of catechol prior to ring opening release of ccMA, the involved co-factors, and metabolites to catechol formation are essentially different.

The direct bio-production of adipic acid was reported for the first time in the pathogenic yeast *Candida tropicalis* using fatty acids and/or sugars as substrates: A 400 L fermenter in a pilot plant produced up to 15 kg adipic acid per week. This corresponds to 37.5 g/L per week or a productivity of approximately 0.223 g/(L×h) (chemicals-technology.com, 2012). In total, an annual production of 800 kg has been reported (Becker *et al.*, 2015). Production of adipic acid by reversed fatty acid oxidation and analogous bioconversions

has since also been reported in *E. coli* from glycerol and glucose. Titers were 12 µg/L (Babu *et al.*, 2015) and 0.5 g/L (combined C_6, C_8, and C_{10} dicarboxylic acids) (Clomburg *et al.*, 2015).

TCA-cycle based pathways are mostly covered in patents (Boussie *et al.*, 2010, Burgard *et al.*, 2010, Burgard *et al.*, 2012, Raemakers-Franken *et al.*, 2012) and are *E. coli* based. One pathway establishing adipic acid production via the reverse adipate degradation achieved 639 ± 34 µg/L (Yu *et al.*, 2014). Recently the same pathway was reported in *Thermobifida fusca* to reach 2.23 g/L (0.045 g/g glucose yield) (Deng & Mao, 2015).

1.1.3.1 Bio-based production of pHBA and pABA

Initial bio-based production of pHBA was established in plants like *N. tabacum* (Siebert *et al.*, 1996, Köhle *et al.*, 2003). More recently two different possibilities to derive pHBA from sugarcane have been explored (McQualter *et al.*, 2005). Microbial production of pHBA from glucose was first reported using *Klebsiella pneumonia* to derive it from chorismate reaching 120 mg/L (Müller *et al.*, 1995). This was done by introducing the *E. coli ubiC* gene, encoding for chorismate lyase, on a plasmid into a strain deficient in the ability to produce AAAs. More recently bacterial production of pHBA has been improved in *Pseudomonas putida* (Verhoef *et al.*, 2010, Meijnen *et al.*, 2011) reaching yields of 16.3 C-mol%, although these processes are not at commercially viable levels yet.

Fermentative production of pHBA in *E. coli* has been patented (Johnson *et al.*, 1998) and reported (Barker & Frost, 2001) deriving the compound via shikimate pathway from glucose, where a final concentration of 12 g/L and a total carbon yield of 13% [mol/mol] was achieved. This was done by overexpression of a DAHP synthase (*aroF*^{fbr}) feedback inhibition insensitive to AAAs and elevation of expression levels of enzymes in the pathway. The conversion of chorismate to pHBA was accomplished by the overexpression of the *ubiC* encoded chorismate lyase. It was found that the DAHP synthase was feedback inhibited by pHBA while DAHP synthase over-expression reduced chorismate lyase expression levels (Barker & Frost, 2001). An active transport of the aromatic out of the cell (Van Dyk *et al.*, 2004) may contribute to achieving these comparatively high titers of extracellular pHBA.

Production of pHBA in the *S. cerevisiae* strain S288c achieved a titer of 0.65 mM (Krömer *et al.*, 2013). This was done by overexpressing the *E. coli ubiC* gene in a double knock-out strain (Δ*ARO7* & Δ*TRP3*). Overexpressing an *ABZ1* gene, coding for aminodeoxychorismate synthase, from the wine yeast strain AWRI1631, instead of the

ubiC gene, resulted in the production of pABA with a final titer of 0.25 mM (Krömer *et al.*, 2013).

Recently also the production of pABA in *E. coli* has been reported, reaching maximum titers of 4.8 g/L at a yield of 21% (Koma *et al.*, 2014). This was accomplished by employing the *E. coli aroF*^{fbr} (DAHP synthase) and *pabC* (aminodeoxychorismate lyase) genes in combination with a *Corynebacterium efficiens pabAB* (aminodeoxychorismate synthase).

Production of folates has been accomplished in many different bacteria, with lactic acid bacteria leading the way (Sybesma *et al.*, 2003). Also, wild-type yeast has been reported to produce considerable amounts of folate (10.8 mg/100 g dry matter) under the right circumstances (Patring *et al.*, 2006).

Utilizing the same biochemical conversion steps as for production of pHBA recently a non-natural pathway for production of terephthalic acid has been patented: Only outgoing from a different precursor (OH group at C2 in E4P replaced with CH_3) it rather leads to the di-acid than the hydroxyl-acid (Osterhout *et al.*, 2013).

1.2 Research Objectives

1.2.1 Development of a chassis cell for production of target compounds

The first objective is the development of a mutant strain of *S. cerevisiae*, which is characterized by enhanced flux to shikimate pathway and impaired AAA synthesis. This helps to channel flux into the desired compounds. It also allows studying the function of the target genes in a new context, leading to new insights and deepened understanding of shikimate pathway. In particular, *ARO7* and *TRP3* are disrupted and a feedback resistant version of *ARO4* (Fukuda *et al.*, 1992) is introduced (Figure 6).

1.2.2 *In silico* analysis of biological production systems

Parallel to strain engineering metabolic network modelling is used to describe the cell, which is incorporated as the second objective. This helps to understand the impact of genetic modifications on cellular metabolism, additionally, it also aids in tailoring the next generation production strains: On the one hand identification of thermodynamic limitations of alternative product-pathways enables identification of the most promising route. On the other hand determination of non-obvious knock-out targets in central metabolism and gained insight into carbon utilization is used to further increase product yield. Both will help to prioritize research efforts.

1.2.3 Establishing production of target compounds

The third objective is to establish production of the target aromatic compounds. Against the backdrop of the previously achieved base level accumulation (Krömer *et al.*, 2013) new approaches, identified in the previous step (objective 1 & 2), are used to elevate titre, yield, and productivity. For the production of pHBA (blue in Figure 6) the identified knock-out targets are applied utilizing a synthetic circuit that allows their dynamic implementation. For the production of pABA (red in Figure 6) different *ABZ1* and *ABZ2* genes from wine yeast strains are introduced and compared in order to identify the best enzymes. This is based on the knowledge of a dysfunctionality of *ABZ1* genes of laboratory yeasts (Ambroset *et al.*, 2011). Correlating performance of the different enzymes with the sequences of the different genes may also allow drawing of conclusions towards mutations that affect the activity. Further the *in silico* determined impact of the carbon-source is investigated in order to increase product yield.

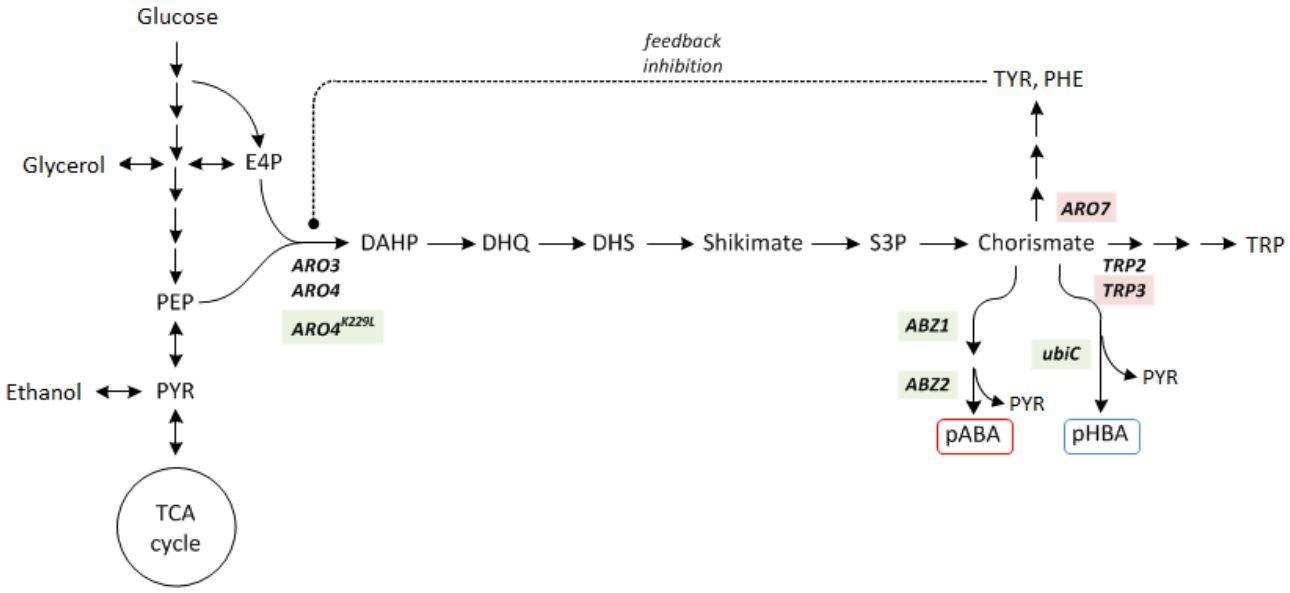


Figure 6: Overview of the simplified shikimate pathway and primary mandatory genetic alterations to obtain pHBA (blue) and pABA (red) from genetically engineered shikimate pathway in *S. cerevisiae*. Dashed line: Feedback inhibition of initial step by final metabolites, to be abolished. Highlighted in red are designated knock-out targets, green additional genes to be (over)expressed. The central intermediates are: 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP), 3-dehydroquinate (DHQ), 3-dehydroshikimate (DHS), shikimate, shikimate-3-phosphate (S3P), chorismate, phenylalanine (PHE), tyrosine (TYR), tryptophan (TRP) and *para*-aminobenzoate (pABA). Important genes and the respective enzymes involved are: *ARO3*/*ARO4*: 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase, *ARO7*: chorismate mutase, *TRP2*: anthranilate synthase, *TRP3*: indole-3-glycerol-phosphate synthase, *ABZ1*: aminodeoxychorismate synthase, *ABZ2*: aminodeoxychorismate lyase.

2 Research Outcomes

2.1 *In silico* analysis of microbial adipic acid production reveals stoichiometric and thermodynamic constraints[1]

The first instalment to develop a microbial production system for polymer feedstocks was a holistic analysis of the feasibility of proposed potential metabolic routes. This allows prioritizing research efforts based on fundamental thermodynamic and stoichiometric limitations.

2.1.1 Abstract

In the course of the rapidly advancing bio-production of industrial relevant chemicals, adipic acid, a nylon-6,6 precursor, has emerged as a very popular target. In the present study, 16 different production routes to adipic acid were compared using a novel tool for network-embedded thermodynamic analysis of elementary flux modes. The tool distinguishes between thermodynamically feasible and infeasible modes, which impacted theoretical yields and revealed limitations of metabolite concentrations. In particular, maximum product carbon yields ranged from 32% to 99% on glucose and/or palmitate in *E. coli* and *S. cerevisiae* metabolic models, heavily depending on the metabolic route. More importantly, metabolite concentrations appeared to be thermodynamically restricted in several pathways. In consequence, the number of feasible flux distributions was reduced, in some cases even deeming the whole pathways infeasible. This highlights the importance of pathway choice. Routes that were thermodynamically favourable over a large concentration and pH range were shikimate pathway based. The low pH capability of *S. cerevisiae* was found to benefit the thermodynamic equilibrium towards product formation, in particular in case of glucaric acid production. This emphasizes the significance of the host organism. Furthermore, reversibility of the mitochondrial malate dehydrogenase appeared to be controversial and a major restriction in *S. cerevisiae* metabolism.

[1] This chapter is based on a manuscript previously submitted to *Metabolic Engineering*, paper under revision in 2016

2.1.2 Introduction

The wealth of pathways to adipic acid (Figure 4) creates the challenge of finding the most promising route, a problem that has so far been ignored. No comparison of the overall potential in terms of maximum theoretical product yields and thermodynamic feasibility has been described in the literature.

In general, optimization of a fermentation process requires yields, rates and titres to be maximized (Vickers *et al.*, 2012). One possibility to assess the yields of a metabolic network is elementary flux mode (EFM) analysis (EMA) (Schuster & Hilgetag, 1994). EMA provides all feasible steady state flux distributions of a metabolic network in a single analysis and enables the assessment of maximum product yields as well as the determination of the ideal flux distribution in a network during optimal production. The approach was previously applied to the development of a range of fermentation products, such as glutamate, methionine, succinate, isobutanol and other compounds (Diniz *et al.*, 2006, Krömer *et al.*, 2006, Chen *et al.*, 2009, Chen *et al.*, 2010, Li *et al.*, 2012, Gruchattka & Kayser, 2015).

Productivity or rates in a process are controlled by capacity- and kinetics-based regulation. While capacity can nowadays be routinely altered using gene knock-out / knock-in and over-expressions, a reaction only occurs in the thermodynamically feasible direction, i.e. where the Gibbs energy (Δ_rG) is negative (Mavrovouniotis, 1993, Kümmel *et al.*, 2006, Henry *et al.*, 2007, Fleming *et al.*, 2009). Δ_rG is a function of the reactants concentrations and the standard Gibbs energy of reaction ($\Delta_rG°$); which depends on the standard Gibbs energy of formation of the reactants ($\Delta_fG°$) and their stoichiometric coefficients in the respective reaction. The range of the Δ_rG of a biological reaction can be estimated from physiological ranges of metabolite concentrations. In reverse, individual irreversible reactions may dictate the range of feasible metabolite concentrations and therefore the direction of otherwise reversible reactions. A methodology that can be applied to examine a full metabolic network at once is network-embedded thermodynamic analysis (NET-analysis) (Kummel *et al.*, 2006). NET-analysis can thus assess whether metabolite concentrations are feasible with respect to the reaction directionalities and assign reaction directionalities based on physiological ranges of metabolite concentrations and thermodynamic constraints (Martínez *et al.*, 2014). This allows the prediction of a pathway's thermodynamic feasibility based on metabolites concentration ranges. In practical terms, this could

mean that a stoichiometrically highly efficient route may be thermodynamically infeasible (Boghigian *et al.*, 2010).

As every EFM basically represents a unique metabolic network itself the thermodynamic feasibility of each single flux distribution is not given. Therefore we developed "*NExT-EMA*", a tool that allows channelling of EFMs (computed with *EFMTool* (Terzer & Stelling, 2008) or other implementations for calculation of flux modes, e.g. (van Klinken & Willems van Dijk, 2015)), into NET-analysis, based on *NExT* (Martínez & Nielsen, 2014).

The concept of thermodynamic analysis of elementary flux modes is not entirely new. In the first attempt, the approach was combined with a genetic algorithm to design an *E. coli* metabolic network optimized for product and biomass formation (Boghigian *et al.*, 2010). Full NET-analysis of EFMs was applied for the first time in combination with flux variability analysis (FVA) to gain insight into *S. cerevisiae* metabolism on a system-level (Jol *et al.*, 2012). This allowed insight into intracellular fluxes and combinations of reactions that could not operate under the investigated conditions were identified. Recently *tEFMA*, a tool that integrates evaluation of the thermodynamic feasibility of elementary during enumeration of these, was developed in order to facilitate the analysis of large-scale metabolic networks (Gerstl *et al.*, 2015). However, the tool was only demonstrated with a simple *E. coli* network (Gerstl *et al.*, 2015).

Previous tools were not applicable to our problem, due to certain restrictions and disadvantages. Therefore, we designed *NExT-EMA* with ease of use in mind while not sacrificing universal applicability: It requires considerably less input (no FVA data needed) and is therefore readily applied and transferred to various metabolisms. Further, it is based on the most recent and advanced thermodynamic data (Noor *et al.*, 2013) and NET-analysis tool (Martínez & Nielsen, 2014) and allows the analysis of fully compartmentalized complex eukaryotic networks (unlike *tEFMA*). Even then computational costs were not an issue (cf. section 2.1.3.10).

NExT-EMA was used to analyse *E. coli* and *S. cerevisiae* metabolic networks comparing the yields and feasibility of all currently known routes to adipic acid and its biological precursors *cis,cis*-muconic acid and glucaric acid from glucose and/or fatty acids (Figure 4). The present work highlights the importance of pathway choice to maximize the potential of the process.

The achievable product titre in a process can be further limited by several factors. Adipic acid could be directly toxic to the cells, which so far has not been described, or the achievable extracellular concentration could be limited by the efficiency and mechanism of the product export from the cells (Maris *et al.*, 2004). This is particularly important for pathways that have reactions with $\Delta_rG'°$ close to zero as limitations in export can lead to accumulation of intermediates and thus cause a reaction to stagnate because of thermodynamic infeasibility. Therefore, also different efflux systems for adipic acid in the cell envelope were evaluated in detail.

2.1.3 Materials and Methods

2.1.3.1 Biochemical pathways to adipic acid

Six fundamentally different pathways featuring unique intermediates and enzymatic conversions were compiled from patents (Frost & Draths, 1996, Frost & Draths, 1997, Boussie *et al.*, 2010, Burgard *et al.*, 2010, Burgard *et al.*, 2012, Picataggio & Beardslee, 2012, Raemakers-Franken *et al.*, 2012) and scientific publications (Draths & Frost, 1994, Weber *et al.*, 2012, Curran *et al.*, 2013, Sun *et al.*, 2013, Lin *et al.*, 2014, Sun *et al.*, 2014, Yu *et al.*, 2014, Babu *et al.*, 2015, Clomburg *et al.*, 2015, Deng & Mao, 2015, Sengupta *et al.*, 2015, Vardon *et al.*, 2015). In total 16 different main routes for carbon-flow were considered. Each route was implemented in two different organism models: *E. coli* and *S. cerevisiae*.

2.1.3.2 Muconic acid production via the shikimate pathway

It is difficult to define separate routes for the production of ccMA from shikimate pathway as many options exist regarding host organism, biochemical route, heterologous enzymes, carbon-source and fermentation method (Zhang *et al.*, 2015). Basically, five different routes were distinguished; three of these contain alternative interconversions (Figure 7). The routes were named according to key intermediates / branch off from shikimate pathway: dehydroshikimate, anthranilate, isochorismate, pHBA and coumarate. To be able to compare this pathway among the others for production of adipic acid a hypothetical reductase enzyme that uses two redox equivalents (NADPH) was assumed for the biochemical reduction of the two double bonds in *cis,cis*-muconic acid (similar to a fumarate reductase-like activity).

2.1.3.3 Glucaric acid production via inositol

D-glucaric acid can be derived from inositol metabolism. From the central metabolite glucose-6-phosphate, the start point of the glycolysis and the pentose phosphate pathway, myo-inositol is formed in two steps of molecular transformation. In two more bioconversions D-glucaric acid is obtained (Figure 7). In order to compare this pathway among the others, a theoretical dehydratase utilizing four redox equivalents (NADH) to dehydrate D-glucaric acid to adipic acid by removing the four hydroxyl-groups was assumed (similar to a fumarase like activity with subsequent reduction of the double bonds).

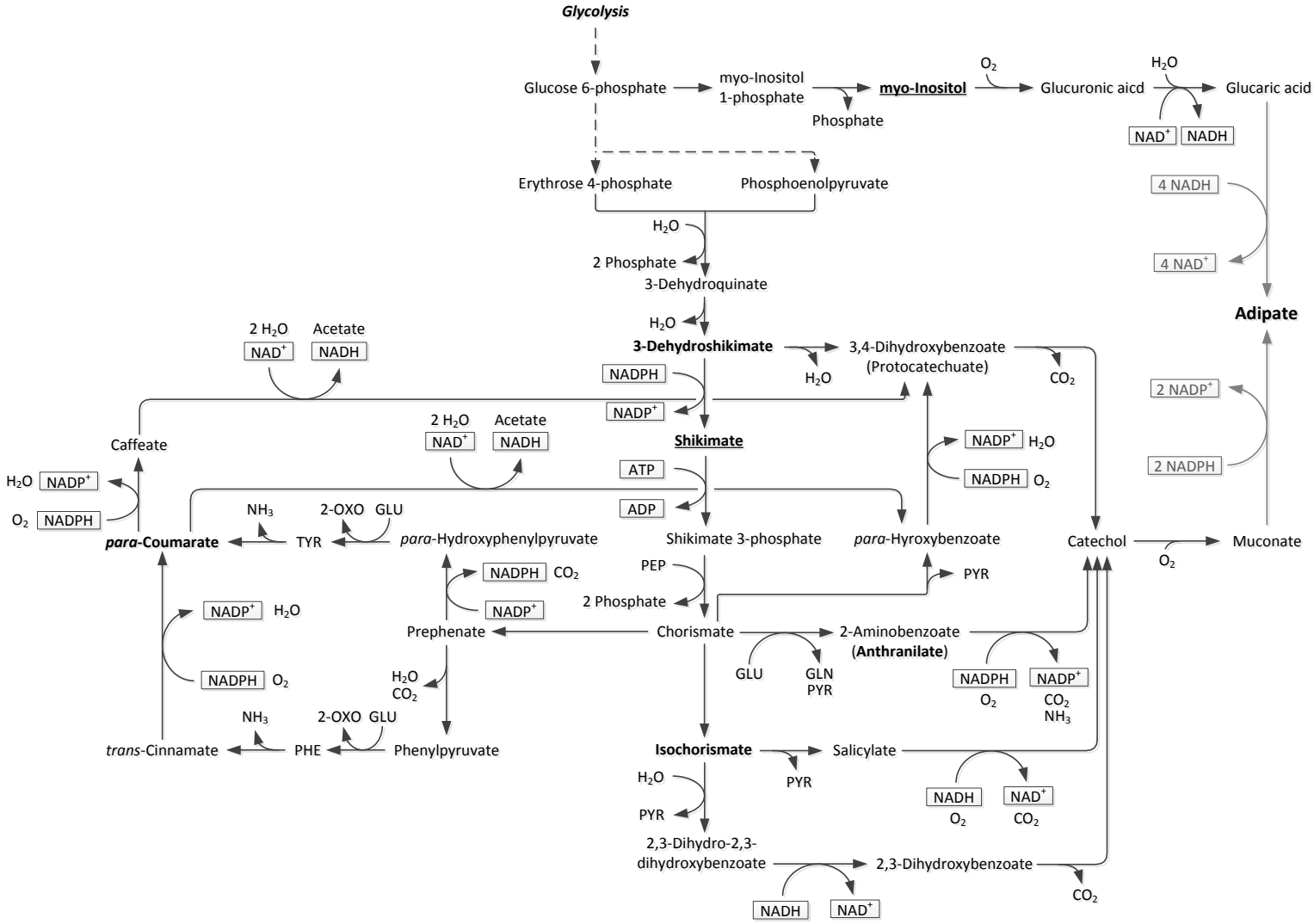


Figure 7: Glycolysis based pathways for biological production of adipic acid. Intermediates that define a pathway are in bold and underlined; central intermediates of a separate route are in bold only. Dashed lines indicate multiple reaction steps.

2.1.3.4 Production via the 3-oxoadipate pathway

The 3-oxoadipate pathway, reported as “reverse adipate degradation” (Burgard *et al.*, 2010, Burgard *et al.*, 2012, Yu *et al.*, 2014, Babu *et al.*, 2015), leads over two different sub-routes to the intermediates hexa-2-enedioate & 2,3-didehydroadipyl-CoA (Figure 8). The former is directly converted to adipic acid, the latter is first reduced to adipyl-CoA. Hence, four different sub-routes were considered for this pathway. For conversion of adipyl-CoA to adipic acid two options were assumed: one step release of adipic acid in a hydrolase reaction via an adipyl-CoA thioesterase (Babu *et al.*, 2015) or two steps utilizing promiscuous phosphate butyryl transferase and butyryl kinase enzymes (Yu *et al.*, 2014). As succinyl-CoA is mitochondrial in *S. cerevisiae* the whole pathway was assumed to be mitochondrial.

2.1.3.5 Production via the 2-oxoadipate pathway

Similarly to the 3-oxoadipate pathway also for the production of adipic acid via 2-oxoadipate multiple routes have been described (Burgard *et al.*, 2010, Burgard *et al.*, 2012), partially involving identical metabolites. All routes were compiled into one interlinked pathway (Figure 8); out of these nine different possible routes for carbon-flow were identified for analysis, in accordance with the respective patents. While every route is essentially unique, the overall stoichiometry may come down to the same lumped reaction in some cases. Therefore, carbon yields can be identical while thermodynamics may differ. For conversion of adipyl-CoA to adipic acid the same reactions as for the 3-oxoadipate pathways were assumed.

2.1.3.6 Production via lysine degradation

Generation of adipic acid from lysine (Figure 8) is closely related to 2-oxoadipate metabolism. In most bacteria lysine is firstly synthesized from aspartate and then broken down via 6-oxohexanoate, which enters the 2-oxoadipate pathway. In most prokaryotes, including *E. coli*, production of lysine from aspartate proceeds via a succinylase, while some other prokaryotes, like *C. glutamicum*, use a dehydrogenase. Both options were included in the *E. coli* model. Yeasts synthesize lysine naturally via 2-oxoadipate partially in the mitochondria, rather than through the aspartate biosynthesis. Therefore only the last steps of this pathway coincide in the prokaryotic and eukaryotic models.

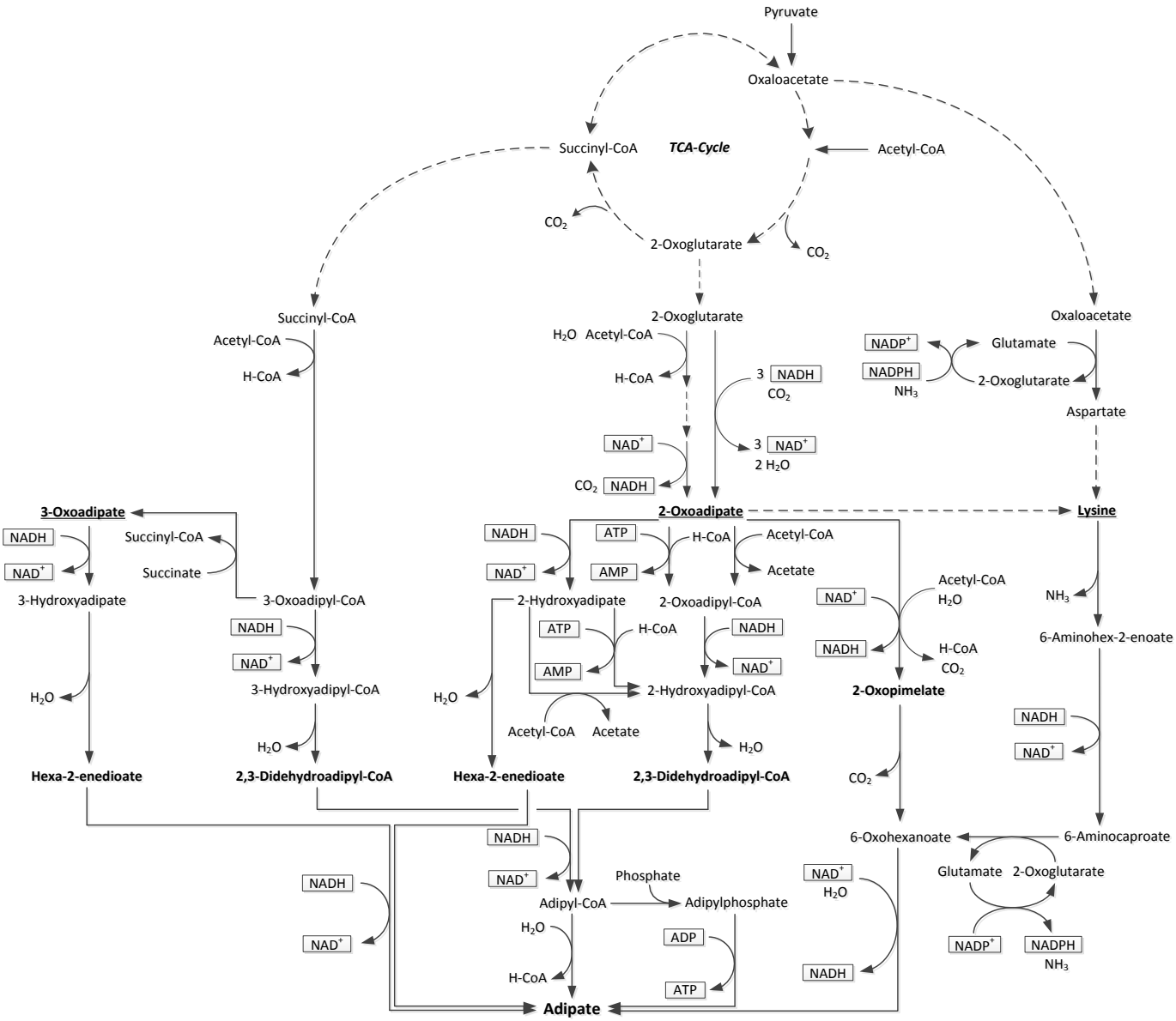


Figure 8: TCA-cycle based pathways for biological production of adipic acid. Intermediates that define a pathway are in bold and underlined; central intermediates of a separate route are in bold only. Dashed lines indicate multiple reaction steps.

2.1.3.7 Production from fatty acid metabolism

A unique variant to obtain adipic acid from engineered microorganisms is fatty acid metabolism. Inferred from the molecular structure this seems natural as adipic acid consists of a saturated C_6-backbone with terminal α,ω-carboxy groups, making it a basically short chain dicarboxylic fatty acid. Rather than distinguishing between different routes this variant of adipic acid production is determined by the type of carbon-source used and the point of entry into the pathway cascade: according to a patent, (Picataggio & Beardslee, 2012) both, fatty acids (FA) and glucose can be substrates. If glucose is the sole substrate, synthesis proceeds either directly outgoing from central metabolism through ω-oxidation to adipate, or via fatty acid synthesis (FAS) and subsequent breakdown of the FA through β-oxidation and carboxylation via ω-oxidation. If FAs are a carbon-source, these can directly enter the β-oxidation and proceed through ω-oxidation to adipic acid. For the pathways described above (Figure 9) a mixture of FAs is utilized either solely or in combination with glucose to feed the yeast *Candida tropicalis* for the production of adipic acid (Picataggio & Beardslee, 2012). The pathways were transferred to a model of the budding yeast *S. cerevisiae* for a more universal significance and comparability with the other pathways. Moreover, the composition of the fatty acid mixture used in the patent is not explicitly defined and also might vary between batches. Therefore, the most common saturated fatty acid palmitate (C_{16}) was assumed as a representative of the FAs mixture. For the conversion of adipyl-CoA to adipate three possibilities were examined, according to described enzyme classes (Burgard *et al.*, 2010, Burgard *et al.*, 2012): an adipyl-CoA hydrolase (EC 3.1.2.-), an adipyl-CoA synthase (EC 6.2.1.-) / phosphotransadipylase/adipate kinase (EC 2.3.1.-/2.7.2.-) and an adipyl-CoA:acetyl-CoA transferase (EC 2.3.1.-).

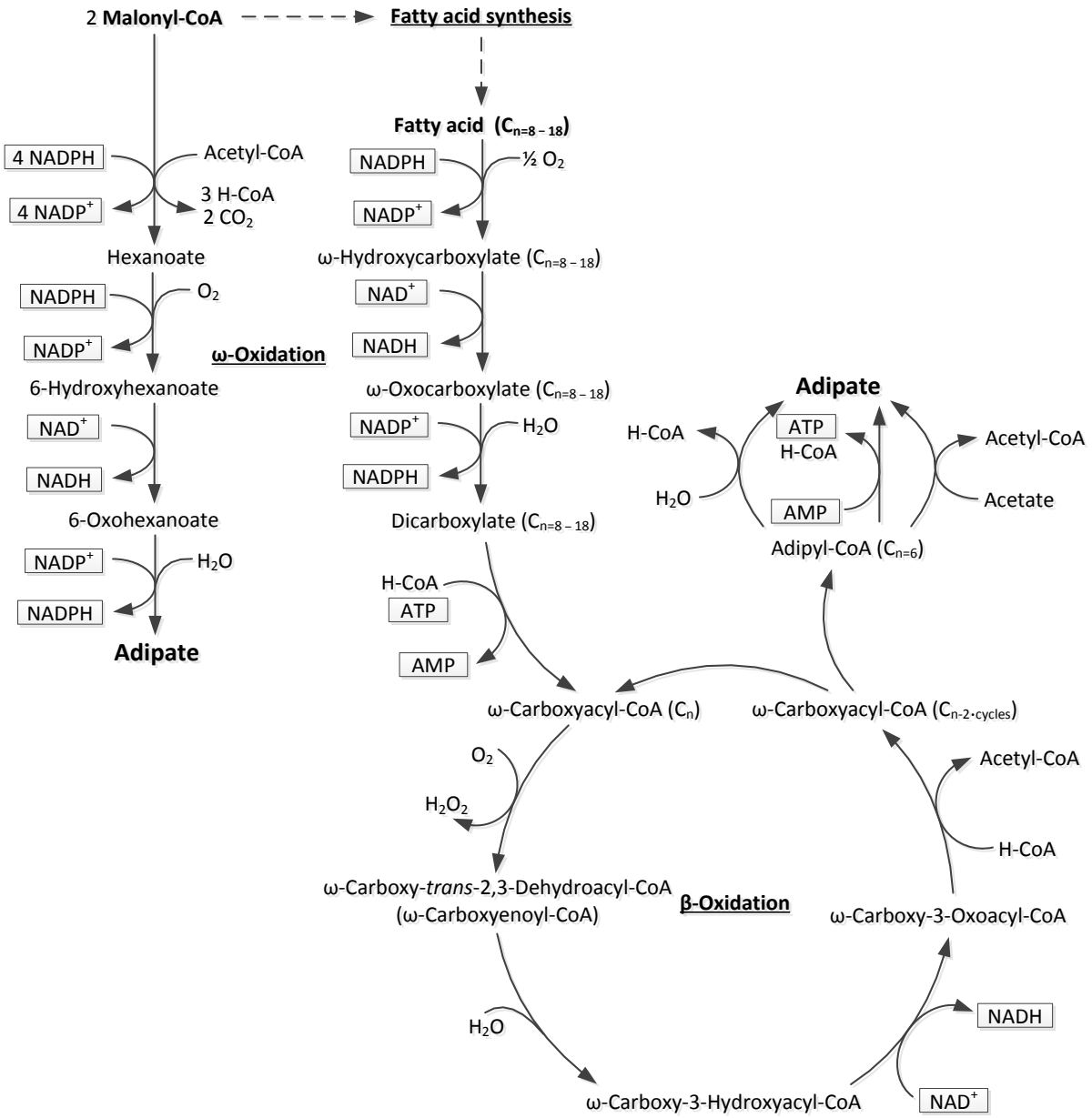


Figure 9: FA metabolism based pathways for biological production of adipic acid. Different pathways are written in bold and underlined. Rather than distinguishing between different routes to adipic acid this variant for adipic acid production is determined by the type of carbon-source used (not apparent from the figure) and the points of entry into the pathway cascade, which are indicated in bold.

2.1.3.8 Stoichiometric networks of central carbon metabolism

Metabolic networks of *E. coli* and *S. cerevisiae* were based on previously published ones (Averesch & Krömer, 2014), expanded by a peroxisomal compartment (Hiltunen *et al.*, 2003) and additional transport reactions (Rottensteiner &

Theodoulou, 2006) as well as uptake of palmitate as an alternative substrate to glucose to realize FA metabolism based production pathways. While glucose transport in *E. coli* and *S. cerevisiae* can be considered non-limiting under most circumstances (Bisson & Fraenkel, 1984), the exact uptake mechanism for FA into the cell is currently unclear. Given that the lipophilic character of the uncharged alkane residue dominates over the carboxylic acid function (Hamilton, 1998) transport should purely depend on the concentration gradient, oil/water interface, and partitioning coefficient. Due to the lack of data, we assumed free and non-limiting palmitate influx. *Ab initio* reaction directionalities were inferred from experimentally curated database data (EcoCyc, SGD). H_2O was not balanced (since water is necessarily ubiquitous in a biological system) for enumeration of elementary flux modes (to reduce the number of irrelevant modes).

However, water was naturally respected for thermodynamic analysis. Further, reactions that consume or produce CO_2, needed to be balanced with one equivalent of H_2O on the opposing side of the reaction to balance oxygen, as in the system the total CO_2 (co2tot) is distributed among the species HCO_3^-, CO_3^{2}, CO_2 and H_2CO_3 in the aqueous phase (Alberty, 2003). Global carbon balancing reactions (*E. coli* R1 – R13 & *S. cerevisiae* R1 – R11) are irrelevant for NET-analysis and were therefore disregarded for determination of reaction directionalities, as was the biomass equation (determination of Δ_rG not applicable, reaction assumed feasible). The full networks are included as supplementary material (files 2.1.1 – 2.1.4) as an example for use with the *NExT-EMA* program.

2.1.3.9 Principles, data, and assumptions for thermodynamic analysis

Thermodynamic data of compounds ($\Delta_fG°$), including pseudoisomers, was obtained using the website version (http://equilibrator.weizmann.ac.il) of the eQuilibrator tool (Flamholz *et al.*, 2012). Experimental data (Alberty, 2003, Alberty, 2006) was always preferred over component contribution method (Noor *et al.*, 2013) derived values, if neither was available it was derived by means of group contribution method (Jankowski *et al.*, 2008) principles. The different sources of data on Gibbs energy of formation were handled as previously reported (Martínez *et al.*, 2014), in order to avoid potential inconsistencies. Physical data (pH, ionic strength, electrical potential, volume) for the different compartments (extracellular, cytosol, mitochondrial matrix & intermembrane space, peroxisome) were largely adopted from previous analyses

(Martínez *et al.*, 2014), unless indicated otherwise: an ionic strength of 0.15 M was considered for all compartments (Kümmel *et al.*, 2006), respective extracellular and cytosolic pH levels of 7.0 and 7.6 were used for *E. coli* (Zilberstein *et al.*, 1984). For *S. cerevisiae* at an extracellular pH of 6.7 a cytosolic pH of 7.2 and mitochondrial pH levels of 7.5 (matrix) and 6.4 (intermembrane space) were used respectively (Orij *et al.*, 2009). As peroxisomal pH is highly disputed (latest theories indicate variability due to a Donnan equilibrium (Antonenkov & Hiltunen, 2012), suggesting a variable pH gradient (Haraldsdóttir *et al.*, 2012)) two scenarios with the maximum and minimum (Haraldsdóttir *et al.*, 2012) reported pHs were investigated. Electrical potentials were estimated accordingly: based on the proton motive force the membrane potential was related to the pH gradient as described previously (Haraldsdóttir *et al.*, 2012). Volume of the peroxisomal compartment was adjusted depending on the pathway / carbon-source used: while under conditions of growth on glucose the peroxisomes typically only make up 1 – 2% of the total volume of the cells, it increases with FAs and certain other substrates, in some cases reaching up to 80% (Veenhuis *et al.*, 2000). Further, when applying more accurate constraints through limited metabolite concentrations, the significance of the results is improved. Where available for *E. coli* and *S. cerevisiae* metabolism, maximum and minimum metabolite concentrations were applied. These were adopted from previous analyses (Henry *et al.*, 2007, Jol *et al.*, 2012) or limited between 0.0001 mM and 10 mM (Kümmel *et al.*, 2006, Martínez *et al.*, 2014).

2.1.3.10 Network embedded thermodynamic analysis of elementary flux modes

EFMs were calculated using the *EFMTool* 4.7.1, freely available for academic use (http://www.csb.ethz.ch/tools/efmtool) (Terzer & Stelling, 2008). The EFMs were then analysed with NET-analysis, using a modified version of *NExT* (Martínez & Nielsen, 2014) (tool available upon request); the basic workflow is outlined in Figure 10. NET-analysis was performed at physiological conditions and was limited to modes that involved the formation of the product of interest (modes with no flux to adipic acid are irrelevant here as the adipic acid yield will always be 0). Computations were performed with MATLAB R2015b (The MathWorks, Natick, USA) using the built in fmincon solver and the ILOG CPLEX optimizer (IBM, New York, USA) on a desktop

computer (Intel Core i7-2600 CPU @ 3.4 GHz, 32 GB RAM).[2] The maximum carbon yields of the different elementary modes were determined with by drawing carbon balances around the transport reactions into and out of the balance area.

The results of the thermodynamic analysis of the EFMs of the different networks were further analysed. First, feasible and infeasible EFMs were compared by means of the directionality of the reactions belonging to these two groups, i.e. the reactions that always had the same direction in all feasible or all infeasible EFMs were identified. By comparing these two groups of reactions potential combinations of reactions (patterns) that caused infeasibility could be determined. Then the respective network was recursively analysed with *NExT* using an algorithm, which considered all reactions reversible, except the reactions that were always irreversible in the infeasible EFMs but not in the feasible ones. In case no clear differences were found between the directionality of feasible and infeasible EFMs, the constraints of the first infeasible EFM were used.

The troubleshooting algorithm identified the reaction(s) of a model which became thermodynamically infeasible due to their conflicting directionalities under the specified conditions of metabolite concentrations and reaction directionalities. The algorithm consists of an iterative process where first a reaction that, if removed, would render a model thermodynamically feasible, is identified. Then it is verified that the directionality in the particular reaction is sufficient to render the model infeasible. These steps are repeated until only the smallest set of reactions is found, which is required to generate infeasibility. Finally, the infeasible EFMs were crosschecked to verify that removing the identified pattern constrains would make them feasible. If infeasible EFMs remained after the removal of the constrains, the iterative process was repeated with the remaining infeasible EFMs.

The analysis of the feasible and infeasible EFMs, the identification of infeasible patterns and the final check of the infeasible EFMs were scripted in MATLAB (tool available upon request).

---

[2] Enumeration of EFMs proceeded at a minimum speed of 3,100 modes/sec when using *EFMTool* and at an average speed of 50,000 modes/sec when using the new *FluxModeCalculator* (van Klinken & Willems van Dijk, 2015). *NExT* reached a minimum speed of 120 modes/sec in case of the most complex *S. cerevisiae* network when relying on the fmincon solver. A different solver, as for example LINDO™, may accelerate the process significantly, especially in combination with parallel computing. In case the number of EFMs of a network becomes too high to fit into memory for at once analysis, the respective file can be split and analysed step-by-step. This was only necessary for the most complex *S. cerevisiae* network, which had more than 40 million EFMs.

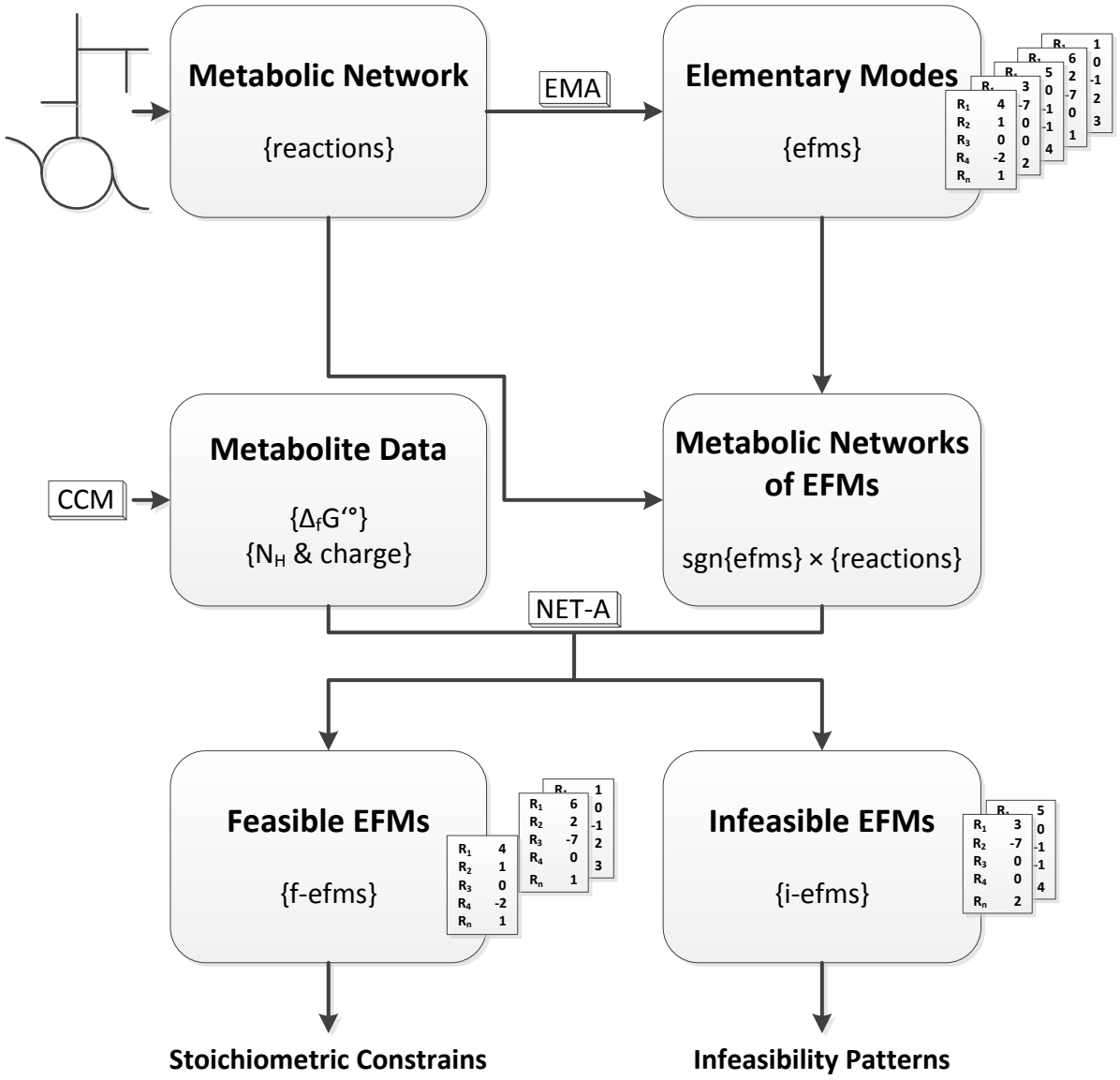


Figure 10: Flowchart for network-embedded thermodynamic analysis of elementary flux modes. EMA = elementary mode analysis; CCM = component contribution method; NET-A = network-embedded thermodynamic analysis; sgn = signum function: Reactions were constrained in NET-analysis according to the direction (= sign) of the flux determined by EMA, for 0 flux the reaction was unconstrained.

2.1.3.11 Definition of pathway feasibility limits by Gibbs free energies of reactions

Besides the thermodynamic feasibility analysis of the *EFMs*, the metabolic networks were also subject to a separate full NET-analysis at different physiological conditions in order to evaluate the thermodynamic equilibria of the different pathways in depth (results are included in supplementary files 3 & 4). In addition to the transformed standard Gibbs free energy ($\Delta_r G'°$) also the Gibbs free energies at boundary reactant

concentrations (Δ_rG_{min} and Δ_rG_{max}) were determined for each reaction. In detail the Δ_rG_{min} was determined using highest substrate and lowest product concentrations and the Δ_rG_{max} using lowest substrate and highest product concentrations, analogous to what has been described previously (Martínez *et al.*, 2014). A reaction was considered feasible if at least the Δ_rG_{min} was negative. If further also the $\Delta_rG'°$ was negative, the thermodynamic equilibrium lies on the substrate side and the reaction is thermodynamically favoured in the written direction. A unique metabolic route to adipic acid was considered favoured if these two criteria applied to all reactions on the route. However, if a single reaction had a positive Δ_rG_{min} the whole route was considered infeasible. All routes that were neither favoured nor infeasible were classified thermodynamically restricted. Further, pH can influence this equilibrium in both ways, depending on which side of an equation the protons take part in the reaction and for transport reactions in which direction a proton gradient exists. In order to investigate this, the Gibbs free energies at lowest feasible physiological pH values were studied separately (*S. cerevisiae* extracellular: 3.5, cytosol: 4.4 (Imai & Ohno, 1995, Guldfeldt & Arneborg, 1998); *E. coli* extracellular: 5.9 cytosol: 6.8 / 6.1 (Diez-Gonzalez & Russell, 1997)). The pH of other compartments was scaled relatively to the cytosolic one, e.g. where no mitochondrial pH was determined a constant ΔpH of + 0.3 in respect to the cytosol was assumed.

2.1.3.12 Assessment of cellular transport processes

Different scenarios for the export of the dicarboxylic acids from the cytosol to the medium were investigated in detail. In particular, passive transport via diffusion, proton-mediated symport, and active transport coupled with ATP hydrolysis were compared.

As uncharged molecules can virtually travel freely over a cell membrane internal and external undissociated forms can be assumed to be equivalent. Knowing that in both environments the different protonation states are in equilibrium passive diffusion of the uncharged compound over the cell membrane can be described by Eq. (1):

$$\frac{c_{int}}{c_{ext}} = \frac{1 + (10^{pH_{int} - pK_{a1}}) \times (1 + 10^{pH_{int} - pK_{a2}})}{1 + (10^{pH_{ext} - pK_{a1}}) \times (1 + 10^{pH_{ext} - pK_{a2}})} \quad (1)$$

Similarly for a dicarboxylate/proton anionic symport Eq. (2) can be formulated, derived from (Maris *et al.*, 2004) it also includes a factor that accounts for the efflux of protons:

$$\frac{c_{int}}{c_{ext}} = \left(10^{\left(n \times \Delta pH - \frac{(n-1)\times \Delta\psi}{Z}\right)}\right) \times \frac{1 + (10^{pK_{a2} - pH_{int}}) \times (1 + 10^{pK_{a1} - pH_{int}})}{1 + (10^{pK_{a2} - pH_{ext}}) \times (1 + 10^{pK_{a1} - pH_{ext}})} \quad (2)$$

With n = number of transported protons, ΔpH and Δψ describing the difference in pH and redox potential across the membrane and Z = ln(10) T R F^{-1}. While R (gas constant) = 8.31446 J $(mol\ K)^{-1}$ and F (faraday constant) = 96,485.333 s A mol^{-1}, T depends on the (optimum) cultivation temperature (in Kelvin) of the respective microorganism.

In the case of active transport driven by ATP hydrolysis, intracellular and extracellular concentrations of a di-acid correlate in a way where Eq. (1) is amended by a factor accounting for the additional driving force, as described by (Maris *et al.*, 2004), resulting in Eq. (3):

$$\frac{c_{int}}{c_{ext}} = \left(10^{\left(\frac{\Delta G_{ATP}}{Z \times F}\right)}\right) \times \frac{1 + (10^{pH_{int} - pK_{a1}}) \times (1 + 10^{pH_{int} - pK_{a2}})}{1 + (10^{pH_{ext} - pK_{a1}}) \times (1 + 10^{pH_{ext} - pK_{a2}})} \quad (3)$$

Where $\Delta_r G_{ATP}$ is the energy obtained from ATP hydrolysis (ATP + H_2O → ADP + P_i) calculated using *NExT* (Martínez *et al.*, 2014) and adjusted for the respective compartments properties.

2.1.4 Results and Discussion

Elementary mode analysis of the different routes to adipic acid production revealed fundamental differences in maximum possible carbon yields between the available biochemical options. Thermodynamic analysis of the EFMs enabled the exclusion of infeasible modes resulting in a more realistic determination of the theoretical yields. Estimation of the Gibbs free energy of reactions provided additional insight into thermodynamic restrictions affecting metabolite concentrations of the pathways and further aided in their classification. Detailed evaluation of different product export mechanisms put recently reported titres and pathway thermodynamics into perspective.

2.1.4.1 Stoichiometric limitations of the different pathways to adipic acid

Elementary flux mode analysis showed a large spread of maximum theoretical carbon yields for the different routes (Table 4 & Table 2). The highest maximum yields (Y_{max}) were achieved for the conversion of glucose to adipic acid via the 3-oxoadipate pathways, which reached to over 90% in both *E. coli* and *S. cerevisiae*. The high yield is possible because the precursor, succinyl-CoA, can be derived from

a reverse operating TCA cycle where no CO_2 is formed, and no CO_2 is formed in the subsequent reactions of the two routes (Figure 8). The route with the second highest yield in *E. coli*, reaching 90%, was derived from lysine degradation, however not so in *S. cerevisiae*: here the Y_{max} was only 66%. The reasons are differences in the pathways leading to the synthesis of lysine rather than the breakdown to adipic acid. In particular, partial mitochondrial localization of the pathway in yeast requires additional energy for transport of metabolites, also the fact that 2-oxoglutarate initiates the pathway has a unique impact (see below). Carbon yields of the 2-oxoadipate pathways were not much lower than 90% in *E. coli*; the yield was slightly reduced because the formation of the precursor 2-oxoglutarate from TCA-cycle involves the production of CO_2. In *S. cerevisiae,* however, the yields were very low and without biomass adipic acid formation was stoichiometrically impossible. This means that product formation is coupled to growth and a direct competition exists, which resulted in a reduced product yield (cf. yield vs. biomass plots, supplementary files 2.1.5 & 2.1.6). This is because 2-oxoglutarate is a precursor for 2-oxoadipate pathways: *S. cerevisiae* does not have transhydrogenases, therefore, the only sink for the NADPH, which is produced during 2-oxoglutarate formation from isocitrate, is biomass. Shikimate pathway based routes greatly differed among one another in their Y_{max}, which ranged from 57% – 75% in both organisms. In general the further downstream the route branches off from the shikimate pathway the lower the yield, as additional bioconversions require more resources. The Y_{max} of the inositol route was (almost) as high as the highest shikimate pathway based one. FA metabolism based routes were a special case: the routes also showed a wide distribution of possible yields, however greatly depended on the carbon-source and on the presence of anabolic and catabolic FA metabolism pathways. Highest yields were obtained on a glucose-feed with direct synthesis of a C6-body (hexanoate) and subsequent ω-oxidation. The overall lowest Y_{max} of 37.5% was obtained when adipic acid was produced from a C16-FA-feed via β-oxidation. This can be explained by the incapability to recycle excess acetyl-CoA, which is generated during β-oxidation of the C16-FA palmitate. When FA biosynthesis was implemented alongside with β-oxidation, acetyl-CoA could be recycled and the yield improved to values comparable to the sugar based variant. This effect was not present when using solely glucose as carbon-source. The assumption that FA biosynthesis and oxidation could be metabolically active at the same time, however, opposes natural

transcriptional control (Hardie & Pan, 2002) and substantial regulatory redesign of FA metabolism would be necessary in order to enable simultaneous FA synthesis and ω-, β-oxidation. A combined glucose- and C16-feed could also not improve the yield significantly. Further, it is noteworthy that when assuming adipic acid to be the final product of β-oxidation, meaning it was only allowed to proceed to a C6-body instead of full breakdown of the FA into C2-bodys, a minimum yield constraint of 37.5% was introduced (cf. yield vs. biomass plots, supplementary file 2.1.5 & 2.1.6).
Finally, it is noteworthy that shikimate pathway and FA metabolism based pathways are both limited to aerobic conditions due to the nature of the product-pathways (oxygen is an essential substrate).

2.1.4.2 Feasibility of elementary flux modes and impact on maximum carbon yields

NET-analysis was used to distinguish between thermodynamically feasible and infeasible EFMs. In some cases, this affected the theoretical maximum product carbon yields severely (Table 4 & Table 2, for detailed information on the feasibility of the flux modes of each network, refer to supplementary files 2.1.3 & 2.1.4). Especially for *S. cerevisiae* infeasible flux distributions were identified in every network. Most of the time the infeasibility could be attributed to two reactions in the central carbon metabolism occurring in a certain pattern: a mitochondrial malate dehydrogenase (R66) that proceeded in the forward direction (the direction it is written in), limited the NADH[m]/NAD[m] ratio to 0.15. This subsequently affected all flux distributions where the malic enzyme (R67) was also active in the forward direction, rendering all modes with this pattern infeasible. Although the malate dehydrogenase (MDH1) has been reported to be reversible, and even favoured in forward direction (according to SGD), these findings are in accordance with previous ones (Jol *et al.*, 2012). The theoretical carbon yields were unaffected by this co-factor ratio restriction (cf. supplementary files 2.1.5 & 2.1.6, yield vs. biomass plots of feasible and infeasible modes).
Infeasibility of the 3-oxoadipate pathway was caused additionally by two other patterns. The drop in maximum product carbon yield by 1% of the 2,3-didehydroadipyl-CoA route, of which over half of the EFMs were infeasible, was caused by a ratio restriction of ATP[m] / (ADP[m] × P[m]), linked to the directions of the mitochondrial succinyl-CoA synthetase and the ATP synthase (R62 & R98). The route via hexa-2-enedioate was completely infeasibility due to a contradiction of the

3-oxoadipyl-CoA[m] concentration in the reactions of the succinyl-CoA:acetyl-CoA transferase / β-ketoadipyl-CoA thiolase and 3-oxoadipyl-CoA transferase (R107 & R108): In order for R107 to proceed in forward direction the maximum concentration of 3-oxoadipyl-CoA[m] could be no higher than 0.12 mM. This did not comply with the subsequent R108, which needed 3-oxoadipyl-CoA[m] at a minimum concentration of 0.33 mM in order to proceed in forward direction.

The fraction of infeasible EFMs was greatest for FA based routes (70%) that involved the pathway of FA β-oxidation. Peroxisomal phosphate transport (R85) into the cytosol was responsible for the infeasibility and only EFMs, where phosphate could be balanced within the peroxisome were feasible. This was only possible in one of the three alternatives for release of adipate, namely release of adipate by the adipyl-CoA synthetase (EC 6.2.1.-) which allows peroxisomal phosphate to be recirculated. However the Δ_rG_{min} of peroxisomal phosphate export appeared to be very close to 0 (< than 1 kJ/mol, which is less than the error of the component contribution method (Noor *et al.*, 2013)) and strongly dependent on pH (cf. section 3.3), therefore this infeasibility criterion may not be critical and the respective flux distributions may be considered feasible after all.

For *E. coli* the different physiological conditions and the greater range of metabolite concentrations did not constrain the malate dehydrogenase, therefore, in most cases 100% of the flux modes were feasible. However some pathways were still affected by thermodynamic constraints: all product forming EFMs of the inositol pathway were infeasible in *E. coli* due to the $\Delta_rG'^{\circ}$ of the uronolactonase reaction (R79) which was largely positive. This is surprising as an *in vivo* route for the production of glucaric acid has been reported previously, reaching significant titers in the g/L range (Reizman *et al.*, 2015). A possible explanation is that the uronolactonase is dependent on a yet unknown co-factor. It is also possible that the reaction is coupled with another one, involving channelling of substrates or products to or from another enzyme, which seems particularly likely, as all other reactions in the pathway are thermodynamically greatly favoured. Additionally, according to KEGG an alternative branch of the inositol pathway may exist, which would avoid the problematic reaction by means of a detour via gulonate.

Further, a portion of the product forming EFMs of the 2-oxopimelate route, part of the 2-oxoadipate pathway, were infeasible in *E. coli*. The explanation for the infeasibility involves a combination of reactions causing restrictions in metabolite concentrations:

the pattern that was found in all modes that were infeasible includes the reactions of the glyceraldehyde-3-phosphate dehydrogenase, 3-phosphoglycerate phosphatase, phosphoglycerate mutase and phosphopyruvate hydratase (R23, R24, R25, R26) proceeding in forward direction, a reverse operating ATP synthase (R71) and R77 in the pathway to adipic acid. These six reaction directionalities together evoke a maximum NADH[c]/NAD[c] ratio constraint of 0.024. This ratio-limit does not allow R77 to proceed in forward direction, thus making the set of all of these reaction directionalities together infeasible. The consequence is a thinning out of low biomass modes (cf. yield vs biomass plots, supplementary files 2.1.5 & 2.1.6) and a slight reduction in the Y_{max} with biomass formation (cf. Table 1). Although there is in the given pathway no impact on the overall maximum theoretical yields, this yet demonstrates the potential of NET-EMA: When investigating a particular pathway *in vivo*, narrowing the metabolite constraints with experimentally acquired data may generate a more accurate description of the constraints that limit the actual system.

Product forming EFMs of the pathway to adipic acid via lysine were also partially infeasible in *E. coli*: subsequently, the Y_{max} dropped to 87%. The cause was the non-enzymatic reaction in the dehydrogenase branch of the lysine biosynthesis (R86 in the network, KEGG reaction number R04336), which was infeasible, although many organisms, like e.g. *C. glutamicum*, use this biosynthetic pathway. This might explain why organisms like *E. coli* utilize the succinylase branch, although the carbon yield is slightly lower. Another possibility is that the reaction is enzymatically catalysed after all - hypothetical co-factor participation would thus lead to a different $\Delta_rG'°$ and the reaction could become feasible.

Unlike the previously published *tEFMA* for calculation of thermodynamically feasible EFMs (Gerstl *et al.*, 2015, Gerstl *et al.*, 2015). *NExT-EMA* provides additional insight into metabolism as it allows identification of the cause of infeasibility of certain flux distributions. That this can be pivotal for strain and process design has been shown for *S. cerevisiae* by means of the 3-oxoadipate pathways and FA based routes and for *E. coli* by means of the 2-oxopimelate route and lysine degradation pathway. Further, the most recent implementation for enumeration of EFMs, *FluxModeCalculator* (van Klinken & Willems van Dijk, 2015), significantly reduces computational times and costs. In combination with *NExT-EMA* this enables the analysis of very large scale metabolic networks on non-workstation computers (which was so far not possible with EFMTool or *tEFMA*).

Table 1: Numbers of EFMs, Y_{max} [C-mol/C-mol] and impact of thermodynamics on Y_{max} for pathways to production of adipic acid in *E. coli*. For details on the reactions that comprise a route refer to Figure 7 - Figure 9 and the full networks in supplementary files 2.1.1 & 2.1.3.

Super-pathway Pathway	Route	# total EFMs	# product forming EFMs	# feasible product forming EFMs	max. yield (w/o biomass)	max. yield (w/ biomass)	max. feasible yield (w/o biomass)	max. feasible yield (w/ biomass)
Glycolysis Shikimate	dehydroshikimate	663,899	230,271	230,271	74.31	70.04	74.31	70.04
	anthranilate	890,948	457,320	457,320	61.14	55.25	61.14	55.25
	isochorismate	906,377	472,749	472,749	69.48	66.09	69.48	66.09
	pHBA	671,485	237,857	237,857	62.57	52.91	62.57	52.91
	coumarate	2,237,284	1,803,656	1,803,656	59.78	57.59	59.78	57.59
	inositol	571,656	138,028	0	74.31	68.76	0	0
TCA-cycle 3-Oxoadipate	hexa-2-enedioate	583,044	149,416	149,416	92.31	91.62	92.31	91.62
	2,3-didehydroadipyl	808,921	375,293	375,293	92.31	91.62	92.31	91.62
2-Oxoadipate	hexa-2-enedioate	527,563	93,935	93,935	88.30	83.09	88.30	83.09
	2,3-didehydroadipyl	1,422,346	988,718	988,718	86.29	82.60	86.29	82.60
	2-oxopimelate	641,999	208,371	202,794	88.30	83.09	88.30	82.90
	lysine	734,628	301,000	197,966	90.00	85.67	87.10	81.48
Fatty acid metabolism Glucose-feed	ω-oxidation	633,496	199,868	199,868	74.12	65.47	74.12	65.47
	C6-FAS ω-oxidation	669,858	236,230	236,230	78.75	76.27	78.75	76.27
	C16-FAS ω- & β-oxidation	685,554	251,926	251,926	50.47	46.95	50.47	46.95
C16-feed	ω- & β-oxidation	92,858	70,937	70,937	37.50	37.50	37.50	37.50
	C16-FAS ω- & β-oxidation	73,392	73,392	73,392	73.05	69.08	73.05	69.08
Glucose- & C16-feed		1,542,307	1,108,679	1,108,679	73.05	69.08	73.05	69.08

Table 2: Numbers of EFMs, Y_{max} [C-mol/C-mol] and impact of thermodynamics on Y_{max} for pathways to production of adipic acid in *S. cerevisiae*. For details on the reactions that comprise a route refer to Figure 7 - Figure 9 and the full networks in supplementary files 2.1.2 & 2.1.4.

Super-pathway Pathway	Route	# total EFMs	# product forming EFMs	# feasible product forming EFMs	max. yield (w/o biomass)	max. yield (w/ biomass)	max. feasible yield (w/o biomass)	max. feasible yield (w/ biomass)
Glycolysis Shikimate	dehydroshikimate	1,448,593	323,048	304,332	75.00	66.08	75.00	66.08
	anthranilate	2,291,909	1,166,364	1,103,210	60.84	59.85	60.84	59.85
	isochorismate	2,032,318	906,773	858,520	68.84	68.57	68.84	68.57
	pHBA	1,855,105	729,560	688,924	62.20	60.60	62.20	60.60
	coumarate	2,753,013	1,627,468	1,545,842	57.14	56.66	57.14	56.66
	inositol	1,746,121	620,576	587,138	74.31	72.22	74.31	72.22
TCA-cycle 3-Oxoadipate	hexa-2-enedioate	3,339,064	1,809,176	0	91.57	90.07	0	0
	2,3-didehydroadipyl	6,936,422	5,406,534	2,743,166	91.57	90.07	90.58	90.07
2-Oxoadipate	hexa-2-enedioate	2,749,090	1,219,202	1,140,296	0	38.09	0	38.09
	2,3-didehydroadipyl	20,481,631	17,310,438	16,145,432	0	38.09	0	38.09
	2-oxopimelate	4,519,321	1,752,471	1,630,396	0	32.00	0	32.00
	lysine	2,045,778	515,890	482,831	66.66	49.83	66.66	49.83
Fatty acid metabolism Glucose-feed	ω-oxidation	9,162,674	6,395,824	5,989,820	60.00	54.46	60.00	54.46
	C6-FAS & ω-oxidation	9,282,723	6,515,873	6,065,029	66.67	65.49	66.67	65.49
	C16-FAS, ω- & β-oxidation	18,259,286	9,958,736	3,097,508	48.26	46.71	48.26	46.48
C16-feed	ω- & β-oxidation	1,005,516	692,487	129,201	37.50	37.50	37.50	37.50
	C16-FAS, ω- & β-oxidation	596,952	596,946	173,139	68.38	66.00	68.38	66.00
Glucose- & C16-feed		41,853,761	33,553,211	10,085,903	68.57	60.82	68.57	60.82

2.1.4.3 Pathway feasibility limits and impact of pH on Gibbs free energies of reactions

For pathways with $\Delta_r G'°$s close to zero, a change in compartment pH due to environmental changes can impact the pathway thermodynamic feasibility. Further, the $\Delta_r G'°$ at low and high ratios between products and substrates can provide an indication on a pathways performance. Therefore, after establishing the differences in achievable carbon yields, the boundary $\Delta_r G'°$s of each reaction of the different networks were determined at different physiological conditions in a separate thermodynamic analysis (results are included in supplementary files 2.1.3 & 2.1.4). As NET-analysis of the full set of reactions in a pathway's network features the highest constraints the results represent a "worst case scenario". Also, the reader should be aware that the condition $\Delta_r G < 0$ is necessary but not sufficient for a reaction to occur *in vivo*.

Determination of the different $\Delta_r G'°$, $\Delta_r G'°_{min}$ and $\Delta_r G'°_{max}$ revealed that a range of pathways contain reactions with a thermodynamic equilibrium that lies on the substrate side. Favourable were only the routes via dehydroshikimate, anthranilate, pHBA and 2-oxopimelate in both, *E. coli* and *S. cerevisiae*, all other pathways contained at least one reaction with a $(\Delta_r G'°) > 0$ (metabolite independent standard Gibbs energy of reaction). Further, the malate dehydrogenase reaction appeared to be completely irreversible in the forward direction at low pH conditions in *S. cerevisiae*. At the very lowest boundary of the pH range, the inositol pathway became also feasible in *E. coli*. Especially in *S. cerevisiae* the improvement was larger, as in the yeast generally a lower physiological pH is possible than in *E. coli*. Also, the $\Delta_r G'°$ of the dehydrogenase reaction in the lysine pathway dropped with lower pH; however it failed to become feasible in *E. coli* due to its narrow pH range. As the change of pH can also have the adverse effect, depending on which side of the equation the protons take part in, the other pathways did in general not benefit from low pH, especially for the 2- and 3-oxoadipate pathways most thermodynamic equilibria deteriorated. In case of the FA metabolism based pathways, no definite effect was present, as some reactions changed in one and others in the opposite direction. Peroxisomal pH, however, had a profound impact on routes involving β-oxidation, which were only feasible at high cytosolic pH in combination with an alkaline peroxisome. Although a high peroxisomal pH restricts the ATP/AMP antiporter, efflux of phosphate from the peroxisome is thermodynamically impossible at low peroxisomal pH. This supports studies arguing for an alkaline peroxisome (van Roermund *et al.*, 2004) and the postulation of a proton gradient (Rottensteiner & Theodoulou, 2006) that can potentially drive transporters.

The phosphate transport problem and the malate dehydrogenase pattern are also good examples showing the advantages of combined elementary mode and thermodynamics-based network analysis: separate NET-analysis of the complete networks is biased, as it gives insight into the range of the Gibbs energy of reactions, but is not sufficient in order to determine if a certain metabolic state (i.e. EFM) is feasible or infeasible. On the one hand, the reactions in question are often not located in the pathway itself and therefore not obvious problems, unless combined with flux analysis. On the other hand a critical Gibbs energy of reaction does not necessarily deem all EFMs infeasible. Even EFMs that include the critical reaction may or may not be feasible due to the presence or absence of other constraints, which are pivotal for the particular reaction. A reduced network (that is any EFM), which has consequently less constraints than the full set of reactions, can therefore still be feasible.

Another restriction of the FA based routes is related to the nature of these cyclic pathways: In normal metabolism, the product of FA β-oxidation, acetyl-CoA, is not accumulated but metabolised in various reactions. In the case of adipic acid overproduction acetyl-CoA will ultimately accumulate (unless FA biosynthesis and oxidation occur simultaneously), which will lead to a product concentration that will shift the thermodynamic equilibrium and stop these reactions. This may explain why the process (Picataggio and Beardslee, 2012) could only achieve production levels corresponding to a mg L^{-1} magnitude (chemicals-technology.com, 2012, Becker *et al.*, 2015).

Overall, several routes contain reactions that have an extremely narrow concentration range beyond which the reactions become infeasible. Therefore, a careful balancing of the pathways that are thermodynamically restricted will be a crucial engineering target; otherwise thermodynamics will limit the production via these pathways. Pathway balancing has to go hand in hand with the engineering of highly efficient product export, as is explained in the following.

2.1.4.4 Evaluation of different scenarios for product export

A crucial factor that can severely affect the thermodynamic equilibria of the reactions in the product-pathway and the achievable product titres in a process is the efficiency and mechanism of the product export from the cells (Maris *et al.*, 2004). This is particularly important for pathways that have reactions with $\Delta rG'^{\circ}$ close to zero, as limitations in export can lead to accumulation of intermediates and thus cause a reaction to stagnate because of thermodynamic infeasibility. Therefore, the impact of different efflux systems for adipic acid in the cell envelope was evaluated in detail.

Diffusion over the cell-membrane is theoretically possible, as long as the driving force (the intracellular concentration of product) is high enough. Making use of Eq. (1) to estimate adipic acid transport out of the cell it becomes clear that c_{ext} will always be magnitudes smaller than c_{int} as long as pH_{int} is higher than pH_{ext} (in most cases given due to the obligatory proton gradient across the membrane). With respective values for pH_{int} of 7.2 (Orij *et al.*, 2009) and 7.6 (Zilberstein *et al.*, 1984) for *S. cerevisiae* and *E. coli*, respectively, and when assuming a physiologically very high intracellular concentration of adipic acid of 100 mM, a maximum extracellular concentration of 10.3 mM (≈ 1.5 g L^{-1}) for *S. cerevisiae* and 6.4 mM (≈ 0.9 g L^{-1}) for *E. coli* may be achieved by passive diffusion. This becomes substantially worse when assuming a proton symporter (Eq. (2)) as also protons are transported against the pH gradient (at c_{int} = 100 mM c_{ext} may only reach 0.33 mM for *S. cerevisiae* and 0.07 mM for *E. coli*). In contrast, an active ABC-like transport system (Eq. (3)) can achieve extracellular concentrations at the solubility limit with only minimal intracellular concentrations (0.08 µM for *S. cerevisiae*, 0.13 µM for *E. coli*).

Given the thermodynamic constraints that many pathways face in terms of feasible intracellular concentration ranges, product export seems to be one of the most relevant aspects. As demonstrated above, the only scenario suitable to achieve competitive product titers is active transporter but to date no such mechanism has been reported for adipic acid. In fact, many of the currently published titers for adipic and muconic acid lay within the feasible range of passive diffusion or proton symport. This includes works in *S. cerevisiae* (Weber *et al.*, 2012, Curran *et al.*, 2013) and *E. coli* (Sun *et al.*, 2013, Sun *et al.*, 2014, Yu *et al.*, 2014, Babu *et al.*, 2015, Sengupta *et al.*, 2015). Exceptions are the early work of the Frost lab on muconic acid production from glucose in *E. coli* (Draths & Frost, 1994, Frost & Draths, 1997, Niu *et al.*, 2002) and routes via salicylate (Lin *et al.*, 2014) and from lignin-derived species (Deng *et al.*, 2016) as well as the production of adipic acid with *Thermobifida fusca* (Deng & Mao, 2015). The higher muconic acid titers indicate the presence of some kind of active transport mechanism. Identification of this transporter could be a significant step towards efficient muconic and adipic acid production. While the shikimate pathway based routes via dehydroshikimate and anthranilate are thermodynamically favoured over a wide range of intermediate concentrations and pH levels, recent demonstrations of these pathways *in vivo* only achieved muconic acid titers up to 141 mg L^{-1} (Curran *et al.*, 2013) in *S. cerevisiae* and 389.96 ± 12.46 mg L^{-1} (Sun *et al.*, 2013) in *E. coli*, which is still far from the previously reported titre of 36.8 g L^{-1} (Niu *et al.*, 2002) via the dehydroshikimate route in *E. coli*. The analysis of intracellular metabolite concentrations in these strains may help to answer the question of export limitation.

2.1.6 Conclusion

Known metabolism currently offers 16 different solutions to the problem of biological adipic acid production. Ten of these routes lead directly to adipic acid, while six routes would yield the precursors *cis,cis*-muconic or glucaric acid. Here, we present the first comprehensive evaluation of differences in theoretical yields and thermodynamic feasibility.

NExT-EMA was developed to perform NET-analysis on elementary flux modes and distinguish between thermodynamically feasible and infeasible solutions. The applicability of *NExT-EMA* to real world problems and advantages over existing methods was demonstrated by analysing adipic acid production in prokaryotic and eukaryotic models.

The analysis illustrates that high adipic acid yields in a biological system are theoretically achievable but significant differences exist between yields and thermodynamic feasibility of the different pathways. Examples are the infeasible inositol pathway and the partially infeasible lysine degradation in *E. coli*, as well as the infeasible route via hexa-2-enedioate of the 3-oxoadipate pathway in *S. cerevisiae*. In some cases extreme differences between the organisms were found, in particular the 2-oxoadipate based routes delivered high yields in *E. coli* but featured the lowest in *S. cerevisiae*. In both organisms the routes via 3-oxoadipate achieve the highest yields but are limited by a thermodynamic equilibrium on the substrate side. The only thermodynamically favoured routes are those based on the shikimate pathway (dehydroshikimate, anthranilate and pHBA routes) and the 2-oxopimelate route, which can achieve moderate to high yields. However, recent publications of the dehydroshikimate and anthranilate routes show so far at best mediocre product titres, hinting at kinetic and/or product export limitations. Low pH capability influenced the thermodynamic equilibrium of the critical reaction in the inositol pathway in favour of product formation. Yields of FA based pathways appeared to be strongly dependent on the applied carbon-source and capability of acetyl-CoA recycling as a by-product of β-oxidation. In *S. cerevisiae* peroxisomal pH was found to have a crucial thermodynamic impinge on FA energy metabolism and the reversibility of the malate dehydrogenase was identified as a crucial element, strongly depending on the redox-cofactor ratio and therefore the feasibility of a large number of modes.

The minimization of intracellular intermediate accumulation and active product export were identified as pivotal for any process to be successful. Despite the theoretical nature of our analysis or perhaps exactly due to that, we were able to identify fundamental predetermined restrictions for adipic acid formation, providing a profound guidance for the development of worthwhile bio-production.

2.2 Tailoring strain construction strategies for muconic acid production in *S. cerevisiae* and *E. coli*[3]

From the holistic feasibility analysis in chapter 2.1 the undertaking arose to analyse the shikimate pathway based routes in depth regarding optimizations, in order to find a constraint that could warrant a minimum efficiency in respect to yield and therefore a significant improvement. Especially non-obvious knock-out targets, for application *in vivo* in different microbial production systems, were determined (Averesch & Krömer, 2014).

2.2.1 Abstract

There is currently a strong interest to derive the biological precursor *cis,cis*-muconic acid from shikimate pathway-branches to develop a biological replacement for adipic acid. Pioneered by the Frost laboratory more than two decades ago, this concept has regained interest. However, recent approaches suffer from low product titers and yields. Therefore, an *in silico* comparison of all strain construction strategies was conducted to highlight stoichiometric optimizations. Using elementary mode analysis new knock-out strategies were determined in *S. cerevisiae* and *E. coli*. The strain construction strategies are unique to each pathway-branch and organism, allowing significantly different maximum and minimum yields. The maximum theoretical product carbon yields on glucose ranged from 86% (dehydroshikimate-branch) to 69% (anthranilate-branch). In most cases a coupling of growth to product formation was possible. Especially in *S. cerevisiae* chorismate-routes a minimum yield constraint of 46.9% could be reached. The knock-out targets are non-obvious, and not-transferable, highlighting the importance of tailored strain construction strategies.

2.2.2 Introduction

2.2.2.1 Muconic acid production via the shikimate pathway

Biotechnological production of ccMA recently experienced a renaissance being implemented in yeast (Weber *et al.*, 2012, Curran *et al.*, 2013) and a new route being introduced in *E. coli* (Sun *et al.*, 2013, Lin *et al.*, 2014, Pugh *et al.*, 2014, Sun *et al.*, 2014). From the shikimate pathway five stoichiometrically unique routes lead to ccMA (cf. Figure 5). In this study the routes were named according to the key intermediates in the respective pathways: dehydroshikimate = DHS-pathway, anthranilate = ANTH-pathway,

[3] This chapter is based on a paper published in *Metabolic Engineering Communications* in 2014

2,3-dihydroxybenzoate = DHBA-pathway, salicylate = SA-pathway and *para*-hydroxybenzoate = pHBA-pathway. The latter four have in common to proceed via chorismate and are therefore collectively referred to as chorismate derived routes. All routes branch of from shikimate pathway at different intermediates (protocatechuate, anthranilate, isochorismate, *para*-hydroxybenzoate) and differ in their co-factors and co-metabolites, which makes a significant difference for stoichiometric carbon yields as well as in terms of energy demand and redox equivalents required for ccMA synthesis.

First reported in *E. coli* (Frost & Draths, 1996) production of ccMA from glucose branching off the shikimate pathway from dehydroshikimate (Figure 5a) was recently also reported in *S. cerevisiae* (Weber *et al.*, 2012, Curran *et al.*, 2013). The latter approaches however only achieved 23590- and 261-fold lower titers and marginal carbon yields (0.01% and 0.9%) than in the earliest study. Also novel pathways leading to ccMA further downstream in the shikimate pathway in *E. coli* via the anthranilate branch (Sun *et al.*, 2013) (Figure 5b), utilizing the siderophore-biosynthesis pathway via isochorismate (Figure 5c), deriving ccMA from 2,3-dihydroxybenzoic acid (c_I) (Sun *et al.*, 2014) and from salicylic acid (c_{II}) degradation (Lin *et al.*, 2014) only achieved a 94-, 77- and 25-fold less ccMA titers on a glucose and glycerol co-feed (4%, 4.95% and 15.48% total carbon product yields respectively). Another shikimate pathway based route, which is a hybrid of the other pathways, producing ccMA from glucose via pHBA (Figure 5d) in *E. coli* (Sengupta *et al.*, 2015) achieved only a titer that was 216-fold below the benchmark, at a yield of 2.155%.

While all these routes feature a sophisticated rearrangement of bioconversions on the way to ccMA, they mostly focus on the shikimate pathway locally and are based on obvious knock-out targets and modifications to shift flux distribution within the pathway. The only rational approach to incorporate modifications that are not located within the shikimate pathway used flux balance analysis to pursue further optimizations (Curran *et al.*, 2013): When flux was directed through glucose-6-phosphate dehydrogenase (G6PD) the transketolase would form fructose-6-phosphate (F6P) and glyceraldehyde-3-phosphate from erythrose-4-phosphate and xylulose-5-phosphate. This resulted in a decreased theoretical carbon yield of 60.9% (Curran *et al.*, 2013). Based on these findings pentose phosphate pathway was modified: Deletion of *ZWF1*, the gene coding for G6PD (R24 in supplementary file 2.2.2), would result in carbon flux to enter pentose phosphate pathway via transketolase (R31 in supplementary file 2.2.2). The gene coding for this (*TKL1*) was over-expressed in order to favour this. Thus, by blocking the channelling of G6P into PPP and favouring the conversion of pentoses into E4P, flux to the shikimate pathway would increase through higher E4P levels derived from F6P.

Other attempts that can be seen as more general approaches to metabolic engineering are strengthening of glutamine regeneration in order to favour anthranilate formation (R65 in supplementary file 2.2.3) (Sun *et al.*, 2013) and overexpression of the phosphoenolpyruvate synthase and transketolase 1 (Lin *et al.*, 2014) to channel carbon flux to the shikimate pathway. Yet the beneficial impact of these modifications cannot be easily foreseen without holistic insight into the metabolism. E.g. transamination is a very common process in many pathways and global effects are unknown.

Given the fact that the shikimate pathway is deeply anchored in central metabolism a more rational approach considering the whole metabolism at once could open unknown opportunities to increase production with less effort by e.g. identifying one or two essential knock-out targets in order to significantly minimize the possible metabolic adjustments of the organism and force it into a state of product formation. It may also help to identify crucial elements of the metabolism which serves a better understanding of its peculiarities. Enroute to ensure product formation on a global level, *in silico* modelling by means of RobustKnock was used to determine target reactions that would improve ccMA production via the pHBA pathway, if deleted. In particular, the phosphotransferase system (PTS) and the pyruvate kinase were identified and applied as knock-out targets, in order to increase PEP availability (Sengupta *et al.*, 2015).

Many computational approaches exist to determine knock-out targets, e.g. the bi-level programming frameworks OptKnock (Burgard *et al.*, 2003), the more recent RobustKnock (Tepper & Shlomi, 2010) or OptFlux (Rocha *et al.*, 2010), which combines strain optimization tasks with pathway analysis. A recent example for model-guided network engineering on yeast shikimate pathway (Brochado & Patil, 2013) shows the potential of these approaches. The issue with these tools is that only one particular solution is computed, resulting in a narrow solution space that draws an incomplete picture of the cells resulting metabolism, thus these tools are considered biased (Lewis *et al.*, 2012).

This can explain the issues of the optimization strategy in (Curran *et al.*, 2013): Elimination of a route leading to a reduced maximum yield (deletion of *ZWF1*) is necessary, but may not be sufficient in order to create a metabolism that operates at high yields. It is still unforeseeable in which other states the cell may be able to operate, potentially evading a high-yield metabolism.

This can be addressed by determining the complete solution space e.g. with elementary mode analysis (EMA) (Schuster & Hilgetag, 1994). In EMA, an array of flux distributions can be compared, which makes it possible to assess the reduction of metabolic freedom (minimum / maximum product / biomass yields) in favour of product formation. This

delivers a more universal description of the metabolisms degree of freedom after application of the respective deletions. Outgoing from this, a minimum efficiency network may be designed through minimizing the metabolic functionality. That this can be applied *in vivo* was shown in (Trinh *et al.*, 2008).

Elementary Mode Analysis (EMA) by means of EFMTool (Terzer & Stelling, 2008), was combined with a holistic approach to compute constrained minimal cut-sets (cMCs) (Hadicke & Klamt, 2011). These cut-sets represent all available genetic intervention strategies to introduce a constraint that allows a minimum efficiency greater than zero with respect to the desired product. That this concept can be easily transferred and applied to various strain optimization problems shows e.g. the study in (Gruchattka *et al.*, 2013). A maximum of five gene deletions was allowed in order to keep the deletion strategy realistic with respect to experimental effort.

2.2.3 Methods

2.2.3.1 Elementary flux mode analysis

EFMs were calculated using EFMTool 4.7.1 (Terzer & Stelling, 2008) freely available from http://www.csb.ethz.ch/tools/efmtool in MATLAB R2012b (The MathWorks, Natick, USA) on the HPC cluster (Barrine) of the University of Queensland using a node with four eight-core X7550 CPUs @ 2 GHz and 1024 GB of 1066 MHz RAM running a PBSPro batch system. Each mode represents a feasible, steady-state, flux distribution of a metabolic network. The maximum carbon yields of the different elementary modes were calculated in MATLAB by drawing carbon balances around the transport reactions into and out of the network. The yield for each product in each mode is defined by:

$$Y_P = v_p \times n_p / v_s \times n_s$$

v_p and v_s represent the flux rates for products and substrates leaving and entering the balance area, n_p and n_s are the numbers of carbon atoms in the product and substrate. Yields were converted to percent; for a graphical depiction of the distribution of modes the biomass yields were plotted against the yields of the desired product for a given network.

2.2.3.2 Construction of metabolic networks

Stoichiometric networks of *E. coli* and *S. cerevisiae* were compiled based on literature (Krömer *et al.*, 2006, Jol *et al.*, 2012, Krömer *et al.*, 2013) and metabolic pathway databases (Kanehisa & Goto, 2000, Cherry *et al.*, 2012, Keseler *et al.*, 2013, Caspi *et al.*, 2014, Kanehisa *et al.*, 2014) and can be found in the supplementary material (supplementary file 2.2.1). The two different branches to ccMA via shikimate pathway were

compiled from literature (Draths & Frost, 1994, Sun *et al.*, 2013) and named according to its branch off metabolite (DHS or ANTH). These networks contain central carbon metabolism, which is only comprised of glycolysis, EDP (*E. coli* only), PPP, TCA-cycle, glyoxylate cycle, anaplerosis, glutamate / glutamine interconversion, electron transport chain, and biomass formation which was broken down to central metabolism derived precursors (supplementary file 2.2.1, R95 in *S. cerevisiae* network and R76 in *E. coli* network). Amino acid biosynthesis was not included in the model; this concerns also routes to AAAs. In the case of the shikimate pathway, this already corresponds to obvious knock-out targets where production of AAAs and folate precursors is inhibited to direct flux to the sole product ccMA.

2.2.3.3 Determination of knock-out targets

For determination of knock-out targets, the method of constrained minimal cut-sets was used, as described in (Hadicke & Klamt, 2011). The desired modes were defined depending on the Y_{max} of each network: *S. cerevisiae* DHS-pathway modes with a yield > 60%, *S. cerevisiae* ANTH-pathway modes with a yield > 50%, *E. coli* DHS-pathway modes with a yield > 70%, *E. coli* ANTH-pathway modes with a yield > 45%. Of these "desired modes" at least 10% had to remain after application of the determined cut-sets. All remaining “target modes” were to be abolished. Cut-sets were calculated allowing five or less knock-outs with respect to the feasibility of the targets (Ruckerbauer *et al.*, 2014): E.g. reactions that are catalysed by the same enzyme cannot be separate knock-out targets (for instance transketolase), neither can spontaneous reactions be targets (ATP hydrolysis). Biomass formation was also defined to be not a knock-out target, as all sets that abolish growth are infeasible. The reactions that are excluded from the cut-set are highlighted in supplementary file 2.2.1.

2.2.4 Results

Not only the knock-out targets but also the availability of a feasible strategy depends as much on the pathway as on the organism’s individual metabolism. This resulted in significant differences in the maximum and minimum yields that could be achieved in the different scenarios.

The maximum yields of the pathways were essentially different in both organisms. The highest theoretical carbon yield (Y_{max}) of the product in respect to the substrate glucose was obtained for the DHS-pathway with 85.7% (0.676 $g_{ccMA}/g_{glucose}$). For chorismate derived pathways most yields were significantly lower and varied between 68.8%

(0.5427 $g_{ccMA}/g_{glucose}$) for the ANTH-pathway in *E. coli* and 80% (0.631 $g_{ccMA}/g_{glucose}$) for the DHBA-pathway in *S. cerevisiae*. When comparing the two organisms, *S. cerevisiae* performed generally slightly better than *E. coli*. This is most significant when comparing the Y_{max} that also allows growth (formation of biomass), e.g. for the DHS-pathway Y_{max} with biomass is 5.35% lower in *E. coli* than in *S. cerevisiae* (cf. Figure 12 E vs. Figure 11 E). All yields can be found in the figures of the respective networks EFM distribution or in supplementary table 1.

For comparability with the results obtained in (Sun *et al.*, 2013, Lin *et al.*, 2014, Sun *et al.*, 2014) the $Y_{max'}$ on glucose and glycerol co-feed were calculated for *E. coli*. These can also be found in supplementary table 2. These were slightly higher (e.g. 89.6% for the DHS-pathway and 73.3% for the ANTH pathway). Carbon-sources different to glucose were not subject to further analysis.

The respective approaches failed to reach these maxima by a big margin. The dashed green vertical lines in Figure 11 A and B indicate the yields reached in (Weber *et al.*, 2012) and (Curran *et al.*, 2013) respectively. For the approaches in *E. coli* the yields are indicated in the same way in supplementary figure 5 A for (Niu *et al.*, 2002), supplementary figure 5 D for (Sun *et al.*, 2013), supplementary figure 5 F for (Sun *et al.*, 2014) and supplementary figure 5 H for (Lin *et al.*, 2014). In order to optimize flux distribution of the metabolisms to favour a higher yield, knock-out targets were determined for each pathway in both organisms. The respective Y_{max} and Y_{min} of the knock-out networks are indicated on the respective figures or can be found in the supplementary tables 3 & 4.

2.2.4.1 Two knock-out targets favour high yield metabolism in S. cerevisiae DHS-pathway

The effect of the deletion of *ZWF1*, previously applied *in vivo* (Curran *et al.*, 2013), was investigated. As can be seen from Figure 11 B the deletion of *ZWF1*, encoding for G6PD, (cf. supplementary file 2.2.1, R24) eliminates a large number of high biomass + low product yield modes, as desired. This effect can be expanded when combining a G6PD deficiency with a deletion of the glycerol-3-phosphate dehydrogenases (GPDH) (R43) as can be seen in Figure 11 D. Although this removed the bulk of modes with a product yield less than 40%, still a large number of modes with zero product yield remains. No cut-sets with five or less knock-out targets could be determined that would result in a minimum efficiency different from zero in respect to the product.

2.2.4.2 Two essential knock-outs enforce yield maximization in S. cerevisiae chorismate derived pathways

For the routes via chorismate a coupling of central metabolism to product formation was possible, allowing the introduction of a minimum yield constraint to the metabolism. In particular deletion of the pyruvate kinases (PYK) (cf. supplementary file 2.2.1, R23) leaves no other source for cytosolic pyruvate formation than the release of pyruvate during product formation. The effect can be seen in Figure 11 G exemplary for the ANTH-pathway, where a Y_{min} of 28.4% was obtained (DHBA-, SA- and pHBA pathway are analogue, cf. supplementary figure 6 & 7. When combined with a deletion of the GPDH (R43) (which eliminated glycerol formation) Y_{min} increased even further to 46.9% (Figure 11 H). Y_{max} was unaffected, remaining at 65.8% with / 70.4% without formation of biomass (compare Figure 11 E with H).

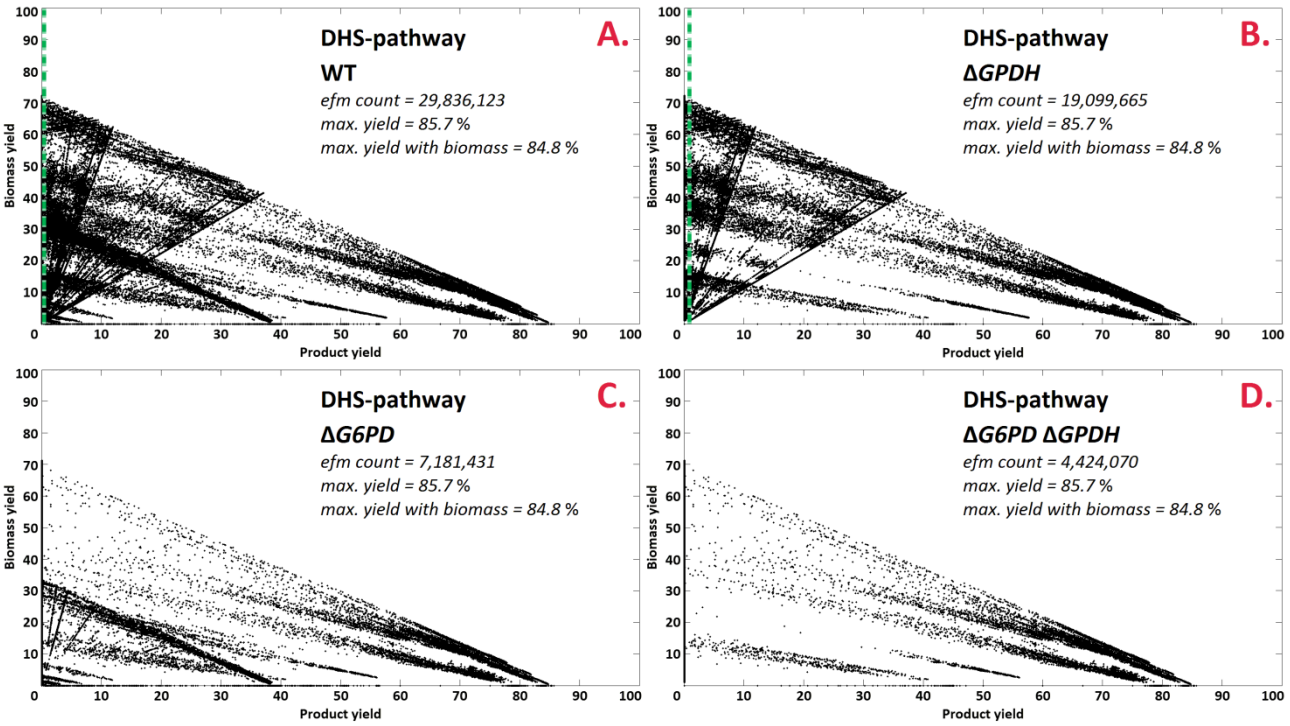


Figure 11: Yield vs. biomass plots of knock-out strategies for *S. cerevisiae*: Product vs. biomass yield plots of the EFM distribution of *S. cerevisiae* DHS-pathway networks. For each pathway four scenarios are shown, comparing the wild-type with the determined knock-out metabolism, key data, as well as respective knock-outs, are indicated on the charts. Each point in a chart corresponds to the specific product and biomass yield of the respective elementary flux mode. Yields are carbon yields in %. A dashed vertical line indicates currently achieved product yields in the respective approaches.

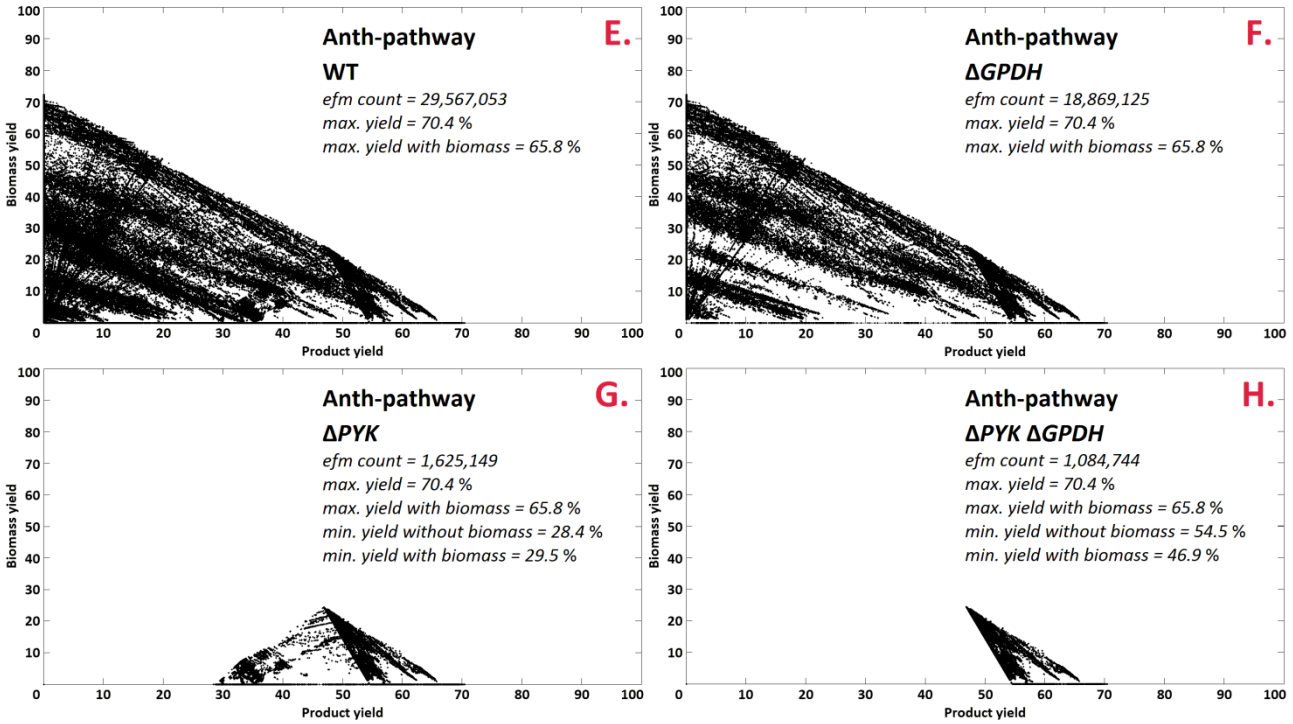


Figure 11 continued: Yield vs. biomass plots of knock-out strategies for *S. cerevisiae*: Product vs. biomass yield plots of the EFM distribution of *S. cerevisiae* ANTH-pathway networks. For each pathway four scenarios are shown, comparing the wild-type with the determined knock-out metabolism, key data, as well as respective knock-outs, are indicated on the charts. Each point in a chart corresponds to the specific product and biomass yield of the respective elementary flux mode. Yields are carbon yields in %.

2.2.4.3 Constraining the minimum yield diminishes Y_{max} in E. coli DHS-pathway

Four minimal cut-sets could be found that result in two different flux distributions with a minimum efficiency greater than zero. The knock-out of glucose-6-isomerase (GPI) and phosphogluconate dehydrogenase (PGD) (cf. supplementary file 2.2.1, R18 & R33) was essential, while the final scenarios could be obtained with the knock-out of either phosphoglycerate mutase (PGM) (R25) or enolase (ENO) (R26) (Figure 12 C) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (R23) or phosphoglycerate kinase (PGK) (R24) (Figure 12 D). This results in Y_{min}s of 71.6% and 73.3% respectively, while Y_{max} was reduced to just below 75%. No cut-set with five or less knock-outs that did not diminish Y_{max} could be found.

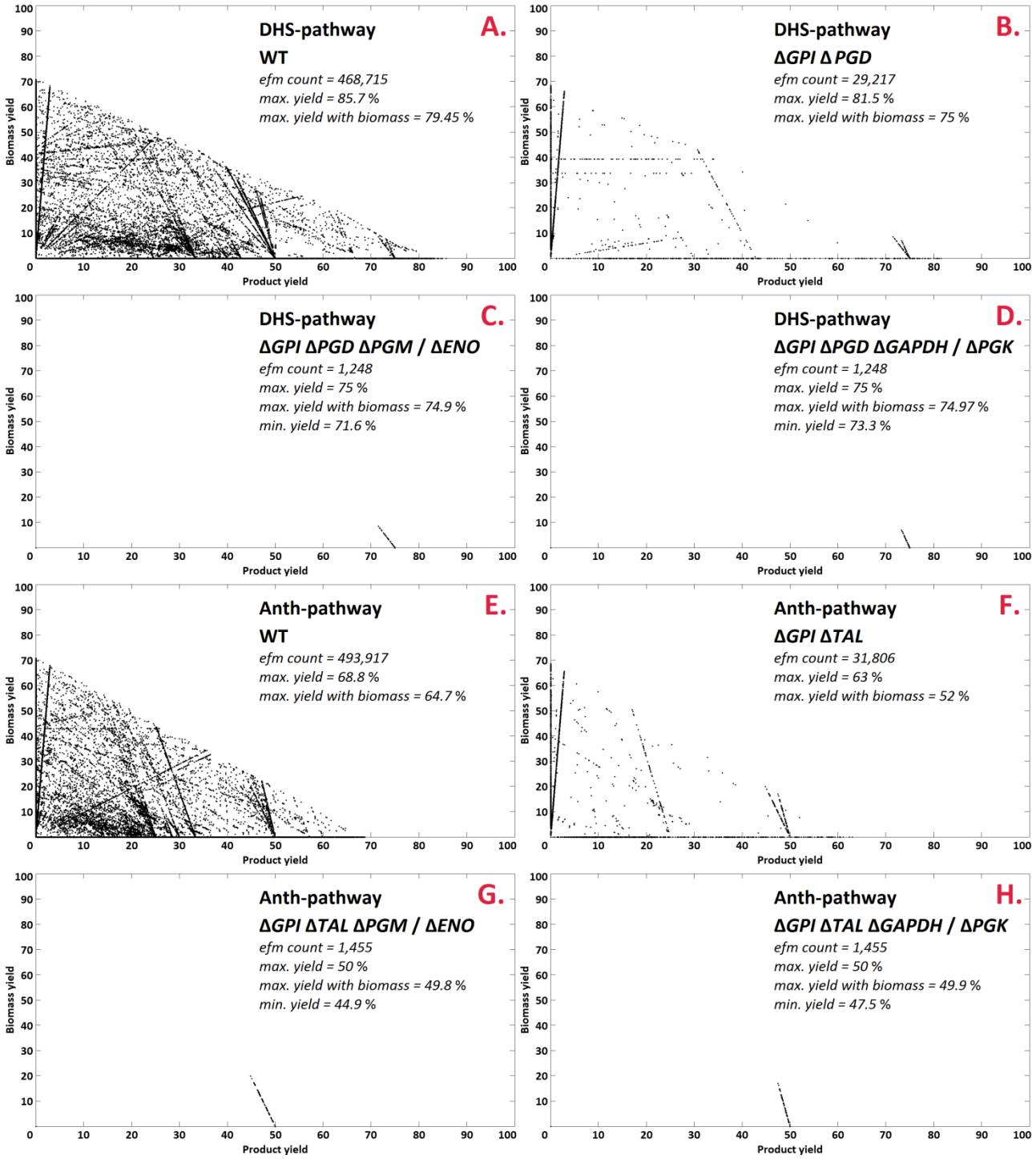


Figure 12: Yield vs. biomass plots of knock-out strategies for *E. coli*: Product vs. biomass yield plots of the EFM distribution of *E. coli* networks. For each pathway four scenarios are shown, comparing the wild-type with the determined knock-out metabolism, key data, as well as respective knock-outs, are indicated on the charts. Each point in a chart corresponds to the specific product and biomass yield of the respective elementary flux mode. Yields are carbon yields in %.

2.2.4.4 Alternative knock-out strategies for preservation of Y_{max} in E. coli

For *E. coli* ANTH-pathway four cut-sets consisting of three knock-out targets analogous to the DHS-pathway could be found: The knock-out of the GPI and the transaldolase (TAL) (cf. supplementary file 2.2.1, R18 & R37) is essential, while the same redundant targets as in the DHS-pathway, PGM or ENO (R25 & R26) and GAPDH or PGK (R23 & R24), result in alternative flux distributions (Figure 12 G & H). Y_{min} was 44.9% and 47.5%, respectively. However, enforcing a Y_{min} also greatly reduced the number of modes and, above all, Y_{max}, which was just below 50%, almost 20% less than in the original metabolism.

In search for knock-out targets that would preserve Y_{max} and result in a similar pattern as for *S. cerevisiae,* a second strategy was determined, resulting in a cut-set with five knock-out targets. In particular, the pyruvate forming reactions of the PTS (cf. supplementary file 2.2.1, R16), PYK (R27), 2-dehydro-3-deoxy-phosphogluconate aldolase (EDA) (R32) and the malic enzymes (MAE) (R54 & R55) were identified as compulsory knock-out targets (Figure 13 G). In addition, deletion of the fumarate reductase (FRD) (R47) allowed Y_{min} to be raised from 5.2% to 25% (Figure 13 H). This knock-out strategy was also applicable to the other chorismate derived pathways (cf. supplementary figure 8 & 9).

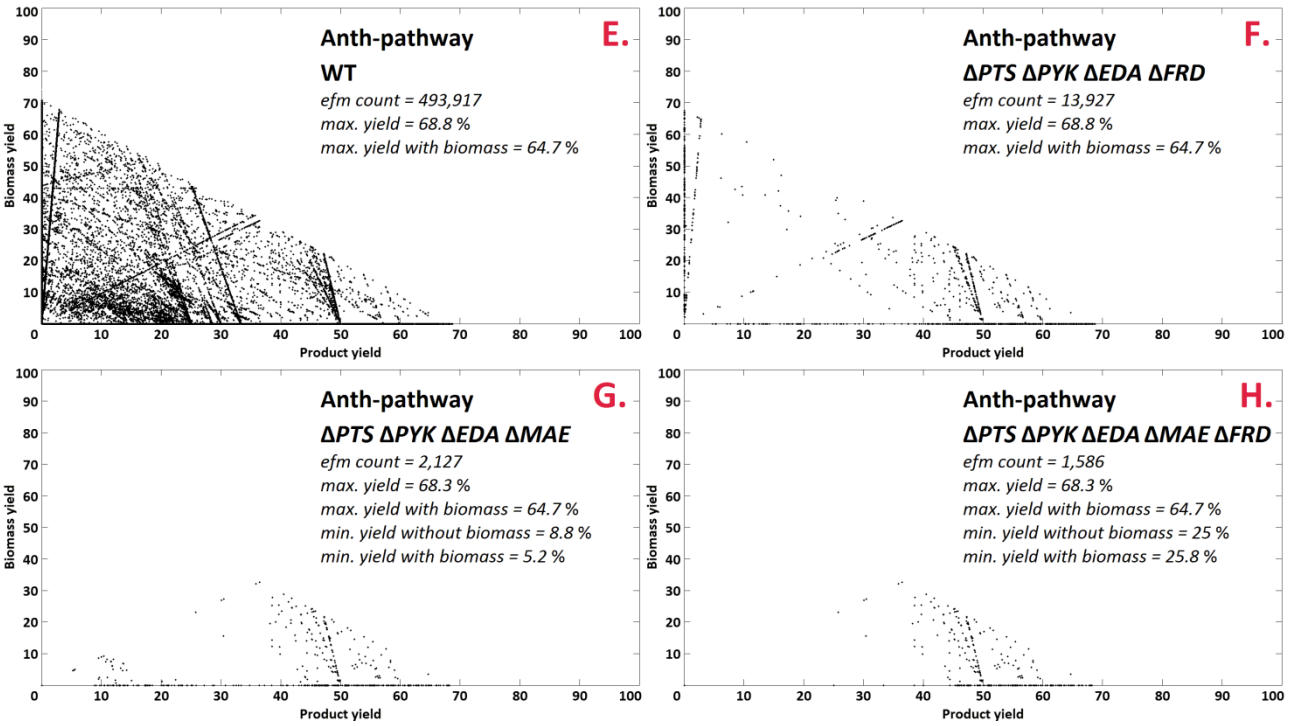


Figure 13: Yield vs. biomass plots of alternative knock-out strategy for *E. coli* ANTH-pathway: Product vs. biomass yield plots of the EFM distribution of *E. coli* networks. Four scenarios are shown, comparing the wild-type with the determined knock-out metabolism, key data, as well as respective knock-outs, are indicated on the charts. Each point in a chart corresponds to the specific product and biomass yield of the respective elementary flux mode. Yields are carbon yields in %.

2.2.5 Discussion

The advantage of the DHS-pathway over the chorismate derived pathways in terms of higher yield can be explained by the higher energy demand of the latter. In particular, the DHS-pathway has no need for co-factors other than O_2 and the shikimate-pathway substrates PEP and E4P (cf. Figure 5). Chorismate derived pathways have direct demand for two NADPH and one ATP (cf. Figure 5). In addition, a second PEP is required, which is equivalent to the demand for two ATP (if regeneration from PYR is assumed, cf. supplementary file 2.2.3 R28 & R74). Indirectly the ANTH-pathway needs another ATP, as the transamination step (supplementary file 2.2.2 R104b, supplementary file 2.2.3 R84b) requires regeneration of glutamate to glutamine (supplementary file 2.2.2 R86, supplementary file 2.2.3 R65). This also explains the advantage of the pHBA-pathway and the isochorismate derived routes, as these do not involve a transamination. In addition, in the DHBA-pathway, which features the highest yield of all chorismate derived pathways, an additional redox equivalent in form of NADH is gained.

In order to explain the low and extremely low product yields obtained by (Curran *et al.*, 2013) and (Weber *et al.*, 2012), distribution of modes has to be considered with the objective of a microorganism in mind: Naturally it will always try to maximize its growth (here the biomass). Looking at the product vs. biomass yield plots (Figure 11 – Figure 13) it becomes clear that maximization of growth involves drain of flux from product formation, which is thus minimized: The two extremes high product yield and high biomass yield are mutually exclusive.

The lowest carbon yield is obtained by Weber et al. (0.01%). Looking at the distribution of modes of the respective network (Figure 11 A) this is no surprise, as the applied strategy does not constrain the Y_{min}. (Curran *et al.*, 2013) apply the *ZWF1* knock-out, taking out the G6PD (Figure 11 B) in order to favour a higher carbon yield. This knock-out is also confirmed in this study as an essential requirement to optimize the pathway. When the elementary flux modes (EFMs) for this network are analysed (Figure 11 B) though, it becomes evident that this is not sufficient: Still every distribution of a product carbon yield between 0 and 85% is possible while the majority features a low yield. From a stoichiometric viewpoint, this can explain why the *in vivo* product carbon yield of Curran et al. (0.9%), though almost 100-fold higher than in any other study in *S. cerevisiae*, still remains below the potential of the pathway. To overcome this problem it needs to be approached from a different angle, in order to actually make use of the microorganism's goal to maximize its growth, as explained in the following.

2.2.5.1 Metabolic peculiarities of S. cerevisiae facilitate knock-out strategies

During the elementary mode analysis (EMA) it became obvious, that out of the two precursors for the shikimate pathway, PEP and E4P, the former is crucial. This has already been realized by (Niu *et al.*, 2002) where a difference in theoretical maximum yields of 43% was determined, depending on the availability of PEP. For the production of one mol ccMA, one mol of PEP is consumed in the DHS- and two moles in the chorismate derived branches of the shikimate pathway while only one E4P is needed. In the pathways via chorismate per mol ccMA one mol PYR is released again. If PYR production in central metabolism is taken out, flux to product needs to be proportionally high in order to obtain enough PYR to feed central metabolism. Why is this knock-out strategy now so applicable to and effective in *S. cerevisiae*? The explanation lies in the nature of the eukaryote's metabolism. It is possible because in yeast central metabolism no pyruvate producing reactions other than PYK and MAE exist. Apart from the products of shikimate pathway only the biosynthesis of glycine from glyoxylate, the homocysteine production through transsulfuration and the degradation of alanine and serine also produce pyruvate (Cherry *et al.*, 2012). The anabolic pathways (glycine and homocysteine) are linked to biomass formation themselves and therefore cannot provide additional PYR in order to increase biomass formation. Alanine itself is made from PYR and therefore cannot be regarded an alternative source of PYR while the serine degradation pathway is used for the degradation of extracellular serine as the sole nitrogen source. In addition, it is unique to the shikimate pathway that PEP is a substrate while all other pyruvate producing reactions start from different metabolites. Also, the *S. cerevisiae* compartmentalization is an advantage, as the malic enzyme is localized in the mitochondria (Boles *et al.*, 1998) it does not have a direct influence on the cytosolic shikimate pathway. Thus, coupling of central metabolism to product formation is *in silico* possible with deletion of only one target (PYK) when using pathways via chorismate for production. Not only the increase in availability of PEP as an educt (which is the main focus in common strain construction strategies) but also the formation of PYR as a by-product is essential in order to realize this knock-out strategy. By taking out the conversion of PEP to PYR in central metabolism the shikimate pathway replaces this function, thus enforcing product formation.

2.2.5.2 Feasibility of theoretical knock-out targets for practical application

In the strain construction strategy pictured in this study for pathways via chorismate in *S. cerevisiae*, cytosolic PYR can only be derived in combination with product formation. This is based on current knowledge, assuming that PYR cannot readily cross the mitochondrial

membrane back into the cytosol. While the structure responsible for the uptake of pyruvate into the mitochondria has recently been revealed (Bricker *et al.*, 2012, Herzig *et al.*, 2012), the reverse, that is the export of PYR from mitochondria into the cytosol, has not been reported to date. Although studies of a PYK knock-out mutant indicate that PYR formed by MAE in the mitochondria can have an influence on cytosolic central metabolism to some extent (Boles *et al.*, 1998), growth of the mutant is only possible on lactate or ethanol and glycerol as carbon-source (Boles *et al.*, 1997). Thus, even if transport of mitochondrially formed PYR into the cytosol would be possible, in the case of glucose feed, carbon would have to pass through glycolysis and the shikimate pathway with product coupling first, before malate would be available to MAE. Only in the case of ethanol or lactate feed flux could enter the TCA-cycle downstream of PYR, thus allowing circumvention of product formation.

A possible concern when arguing that product formation is coupled to central metabolism is that formation of PYR does not occur during the final step of ccMA formation. E.g. in the case of the ANTH-pathway it could be sufficient for the organism to proceed to anthranilate in order to obtain PYR, thus rather accumulating anthranilate than ccMA. From a thermodynamic point of view, this is unlikely as the $\Delta_rG'°$s for the two subsequent reactions from anthranilate to ccMA are extremely negative (Flamholz *et al.*, 2012), meaning the reaction equilibrium is largely on the product side. A high anthranilate concentration would even boost this, favouring the pathway to proceed to the end product ccMA. The same applies for the SA-pathway, while for the DHBA- and the pHBA-pathway, which both proceed via different isomers of dihydroxybenzoic acid, the $\Delta_rG'°$ of the reaction to catechol is close to 0. This may give preference to the ANTH- and the SA-pathway, although their yields are lower.

Further concerns are related to *in vivo* application difficulties of the identified knock-out targets in *S. cerevisiae*. Deletion of the paralogs, which code for the two PYK isozymes (see supplementary table 5 for more details), results in a severe growth defect(Boles *et al.*, 1997). The same applies for the two GPDH isozymes (Hubmann *et al.*, 2011). For GPDH a knock-down strategy exists that retains a viable metabolism while almost completely inhibiting glycerol formation (Hubmann *et al.*, 2011). When combining this with an overexpression of the alternate PYR synthesis route (that is the product-pathway) this should enable growth again. In terms of PYK also regulatory factors play a role, as not both isozymes are always expressed to the same level at the same time (Boles *et al.*, 1997), so that deletion of *CDC19* only could be sufficient.

2.2.5.3 Metabolic peculiarities of E. coli hinder knock-out strategies

The knock-out strategy for *E. coli* DHS-pathway and the first one for ANTH-pathway similarly introduce a minimum yield constraint: Biomass formation is limited by forcing all flux through the EDP. For this the deletion of GPI is essential in both cases and blocks the entrance (R18) of glycolysis. In the DHS-pathway the oxidative part of PPP is blocked by knock-out of PGD (R33) (Figure 12 B); for the ANTH-pathway deletion of TAL (R37) has a similar effect (Figure 12 F). The four alternative knock-outs PGM, ENO, GAPDH and PGK (R23 – 26) constrain flux to the biomass precursor 3-PG from either side, resulting in a very narrow distribution of modes. This is very unrefined, resulting in a massive reduction of Y_{max}. Further, it also limits the cell's metabolic freedom to adjust to its environmental settings. Therefore, the value of this strategy for *in vivo* application is questionable. This extremely constraining knock-out strategy also explains the reduction of Y_{max} and underlines that not always the most minimal intervention strategy is necessarily also the best. A longer cut-set can sometimes also be of advantage, as outlined in the following.

An alternate knock-out strategy can preserve Y_{max} in the ANTH-pathway (and the other chorismate derived pathways) and results in similar coupling of growth to product formation as in *S. cerevisiae*. However, the PEP-pyruvate node in *E. coli* is significantly different from *S. cerevisiae* and three additional PYR producing reactions (PTS, EDA, MAE) besides PYK hamper the application of the same knock-out strategy. A knock-out of the phosphotransferase system is known in the literature (Escalante *et al.*, 2012) allowing a change of glucose uptake mechanism especially if combined with overexpression of hexose permease (Hernandez-Montalvo *et al.*, 2003). Also, combinations of PTS and PYK knock-out (Lee *et al.*, 2005) have been applied successfully, especially in respect to the production of AAAs (Gosset *et al.*, 1996, Meza *et al.*, 2012, Papagianni, 2012). Also very recently the knock-out of PTS and PYK by means of *ptsH*, *ptsI*, *crr*, and *pykF* has been described and applied for ccMA production from pHBA (Sengupta *et al.*, 2015). Of the remaining targets (EDA and MAE) each of the single knock-outs is possible (Baba *et al.*, 2006). However, more combinations of these knock-out targets have to our knowledge not been described so far. This could be due to the extensive mutations in central metabolism having unpredictable consequences, even though modelling predicts that the resulting metabolism should be viable (see supplementary table 5 for more details on genes coding for the respective enzymes).

Further, the formation of succinate as a by-product has been identified as one of the major features in modes with low product yields (Figure 13 G). Succinate, a product of mixed acid fermentation in *E. coli*, is in these modes formed by R47 (cf. supplementary file 2.2.1).

It can drain carbon flux from shikimate pathway, as it is obtained from PEP in four steps via oxaloacetate, malate and fumarate (Thakker *et al.*, 2012). The enzyme identified as knock-out target (FRD) is only active under anaerobic conditions while the TCA-cycle is inactive (aerobically succinate is only produced if the glyoxylate cycle is active) (Cecchini *et al.*, 2002, Thakker *et al.*, 2012). A knock-out may not be necessary as aerobic fermentation is obligatory for ccMA formation (in the last step oxygen is incorporated in the ring structure of catechol between the two hydroxyl groups, thus releasing ccMA).

One possible approach to facilitate critical knock-outs is the knock-down of the respective target genes. sRNAs in *E. coli* can be utilized for posttranscriptional repression of a gene (Sharma *et al.*, 2013). Although more tricky in prokaryotes, as the lack of a nucleus reduces the chance for the interaction of the sRNAs with the mRNA between transcription and translation, it is still a promising approach. It results in attenuation rather than full silencing, which can also be important to retain a basal activity of critical target genes. This could be combined with an inducible system (a wealth of approaches exists *E. coli* (Baneyx, 1999)) in order to further facilitate the cultivation of the strain.

So far it has only been discussed which reactions need to be taken out additionally in respect to yeast in order to also couple product formation to central metabolism. But why is a yield reduction observed in the knock-out strategy for the DHS-pathway and the first ANTH-pathway knock-out strategy in *E. coli* (Figure 12)? This is due to limited possibilities to recycle PYR, which exits shikimate pathway, back into central metabolism. Of the replenishing reactions that fill up the PEP-pyruvate-oxaloacetate node in *E. coli* (Sauer & Eikmanns, 2005, Papagianni, 2012) only the phosphoenolpyruvate carboxykinase (PCK) and the PPS remain. Thus the formation of PEP can either happen from PYR at the expense of two ATP equivalents (PPS) or from oxaloacetate (PCK). In contrast to *S. cerevisiae, E. coli* does not possess a pyruvate carboxylase which incorporates one CO_2 into PYR to form oxaloacetate.

Only two options remain to convert PYR to OAA: One way is to further metabolise the PYR via acetyl-CoA. This leads to decarboxylation. The acetyl-CoA can then only be assimilated via the glyoxylic shunt and one-third of the carbon is lost as CO_2. The other possibility for PYR conversion to OAA is the reverse operation of MAE in concert with the malate dehydrogenase and PCK, which would lead to a carbon neutral but ATP-dependent conversion of PYR back to PEP. Because of the lower energy demand and the absence of CO_2 formation, the final option has a higher carbon yield. The *in vivo* observation that the MAE is up-regulated in a fully PYK deficient *E. coli* mutant and the PCK is also highly active (Emmerling *et al.*, 2002) supports this.

2.2.6 Conclusion

When using pathways via chorismate it was possible to apply a minimum yield constraint with as little as one knock-out in *S. cerevisiae* without any loss in Y_{max}. Although the Y_{max} of the DHS-pathway is higher, only the chorismate derived pathways allow for this strain construction strategy. In *E. coli,* five knock-outs are necessary to achieve a similar picture without a drop in Y_{max}. Therefore, yeast seems to be the favourable organism for production of ccMA in terms of knock-out strategies.

This study not only highlights the importance of pathway choice, but also clearly points out how differences in central metabolism of the host organism can affect a synthetic pathway.

2.3 Application of *in silico* determined knock-out targets for production of *para*-hydroxybenzoic acid[4]

The approach described in chapter 2.2 is transferrable to other shikimate pathway derived metabolites that also involve the formation of pyruvate as a by-product. For the production of pHBA from shikimate pathway in *S. cerevisiae*, the strategy was applied *in vivo* using gene silencing under the control of a synthetic circuit (Williams *et al.*, 2013) that enabled dynamic metabolic pathway control (Williams *et al.*, 2015).

2.3.1 Abstract

Some of the most productive metabolic engineering strategies involve genetic modifications that cause a severe metabolic burden on the host cell. Constraining genetic modifications, including those limiting growth, can be more effective if they are 'switched on' after a population growth phase has been completed. This can be achieved by means of dynamic regulation using a synthetic quorum sensing circuit previously established in *S. cerevisiae.* The circuit autonomously triggers gene expression at a high population density, and was linked with an RNA interference module to enable target gene silencing. The circuit was used to control flux through the shikimate pathway for the production of *para*-hydroxybenzoic acid (pHBA). Dynamic RNA repression allowed gene knock-downs which were identified by elementary flux mode analysis as highly productive but with low biomass formation to be implemented after a population growth phase, resulting in the highest published pHBA titer in yeast (1.1 mM).

2.3.2 Introduction

The metabolic engineering which is required to over-produce many commercially valuable bio-products at high yields can impose a severe metabolic burden on native cellular processes (Bentley *et al.*, 1990, Lee *et al.*, 2008), limiting population growth and therefore product titer. This is because engineered pathways compete with native metabolism for carbon flux, ATP, and redox cofactors, and can result in the accumulation of toxic compounds (Keasling, 2008, Lee *et al.*, 2008). Furthermore, many of the most productive gene deletion or overexpression strategies identified by *in silico* modelling are not feasible for constitutive application *in vivo* due to the limitations they impose on cell growth or survival. This study demonstrates that this problem can be overcome by largely separating

[4] This chapter is based on a paper published in *Metabolic Engineering* in 2015

growth phases from production using an autonomous genetic regulatory circuit, which enables dynamic modulation of gene expression.

The number of well-characterized genetic induction systems in *S. cerevisiae* is limited. Most popular are the GAL promoters and the heterologous doxycycline responsive Tet promoters (Blount *et al.*, 2012). These systems are inherently limited by the requirement for the addition of expensive inducers to the media, and in the case of the GAL promoters by the repression of gene expression on glucose. Furthermore, they do not truly separate growth from production since cell-cycle continues when gene expression is induced. In this study, the use of a recently developed synthetic quorum sensing circuit (Williams *et al.*, 2013) for the dynamic control of metabolism and the separation of growth and production phases in yeast via a cell-cycle arrest phenotype was investigated.

In nature, many microorganisms are able to coordinate population-wide gene expression according to population density via cell-to-cell communication with extracellular signalling molecules. This phenomenon, termed 'quorum sensing' (QS) holds promise as a mechanism for separating growth from production in industrial microorganisms for the production of growth-limiting compounds (Choudhary & Schmidt-Dannert, 2010) and QS has been explored for application in biotechnology (March & Bentley, 2004, Tsao *et al.*, 2010, Song *et al.*, 2011). A previously developed QS circuit enables the spatiotemporal control of gene expression, and is activated using an 'AND' gate system requiring both high population density and the presence of low concentrations of AAAs in the media. One limitation of the pheromone QS circuit is the lack of a mechanism for repression or silencing of gene expression. To enable the dynamic repression of gene expression the QS circuit was interfaced with an RNA interference (RNAi) module (Drinnenberg *et al.*, 2009). This RNAi module has the potential to repress the translation of any mRNA via the transcription of simple hairpin RNA constructs which target mRNAs for degradation according to simple base-pair complementarity.

As a mechanism for dynamic pathway control and product formation the pheromone QS system was used to apply a novel, *in silico* determined, knock-out strategy for production of *para*-hydroxybenzoic acid (pHBA) in *S. cerevisiae*, which would result *in vivo* in a severe growth defect on glucose if applied constitutively (Averesch & Krömer, 2014). It is not only unique to shikimate pathway derived products that involve the formation of pyruvate as a by-product, but also to the organism *S. cerevisiae* (Averesch & Krömer, 2014), thus making the RNAi interfaced S circuit the ideal system to test it.

2.3.3 Materials and Methods[5]

2.3.3.1 Elementary flux mode analysis

EFMs were calculated using EFMTool 4.7.1 (Terzer & Stelling, 2008) freely available from http://www.csb.ethz.ch/tools/efmtool in MATLAB R2014a (The MathWorks, Natick, USA) on a desktop computer (Intel Core i7-4770 CPU @ 3.4 GHz and 32 GB RAM). The maximum carbon yield of the different elementary modes was calculated in MATLAB by drawing carbon balances around the transport reactions into and out of the network. The stoichiometric network analysis of *S. cerevisiae* was based on a previously published study (Averesch & Krömer, 2014). The metabolic network can be found in the supplementary material, it comprises the central carbon metabolism as well as the chorismate biosynthesis and a chorismate lyase (*ubiC*) reaction that catalyses pHBA formation from chorismate.

2.3.4 Results and Discussion

2.3.4.1 Strain design and metabolic network modelling

The most promising targets for gene knock-out to enhance flux towards pHBA carry a severe growth penalty. In earlier studies the knock-out of *ARO7* (encoding chorismate mutase) and *TRP3* (encoding anthranilate synthase) has been described as a way to increase chorismate availability for pHBA formation (Krömer *et al.*, 2013). However, this strategy also causes phenylalanine, tyrosine and tryptophan auxotrophy and the associated growth defect, even when cells are fed these amino acids (Winter *et al.*, 2014). To avoid the auxotrophies associated with gene deletion during the growth phase, these mutations were selected for conditional knock-down using a novel QS circuit (Williams *et al.*, 2013) combined with RNAi. However given that the QS circuit is activated by the presence of tryptophan in the growth medium, and considering that tryptophan causes strong feedback inhibition of anthranilate synthase (Miozzari*, et al.*, 1978, Graf*, et al.*, 1993), targeting of anthranilate synthase (*TRP3*) was not necessary.

In addition to the targeting of *ARO7* an advanced genetic intervention strategy that would more radically affect metabolism on a global cellular level and ensure pHBA production was identified. Using a metabolic modelling approach a stoichiometric model of central carbon metabolism, that already had the chorismate mutase (*ARO7*) and anthranilate

[5] For materials and methods of work performed by other contributors to this study (media, growth conditions, strains and plasmids, analytics, quantitative real-time PCR, enzyme assays and statistical analysis) refer to the full manuscript in the supplementary.

synthase (*TRP3*) knock-outs applied, was subject to elementary flux mode analysis. A single target reaction was identified to significantly constrain the system: The knock-out of the pyruvate kinase reaction (catalysed by the homolog gene products of *CDC19* and *PYK2*) would result in a minimum efficiency of the metabolism in respect to product formation. This becomes evident from Figure 14, where the biomass yield of all possible flux distributions is plotted against the carbon yield: While in Figure 14 A the pHBA yield is not restricted and can be zero, the knock-out of the pyruvate kinase (Figure 14 B) results in the introduction of a minimum yield constraint. The underlying fundamentals of this effect are the same as explained in-depth in chapter 2.2. As an inevitable side effect, maximum biomass yield is also affected, being more than halved. Consequentially *CDC19* mutants are growth impaired when glucose is the sole carbon-source (Sprague, 1977). Thus, traditional gene deletion is not feasible for glucose-based fermentations, which are favoured at large scale due to feedstock cost and availability. It was, therefore, essential to implement pyruvate kinase repression dynamically along with chorismate mutase (*ARO7*) repression.

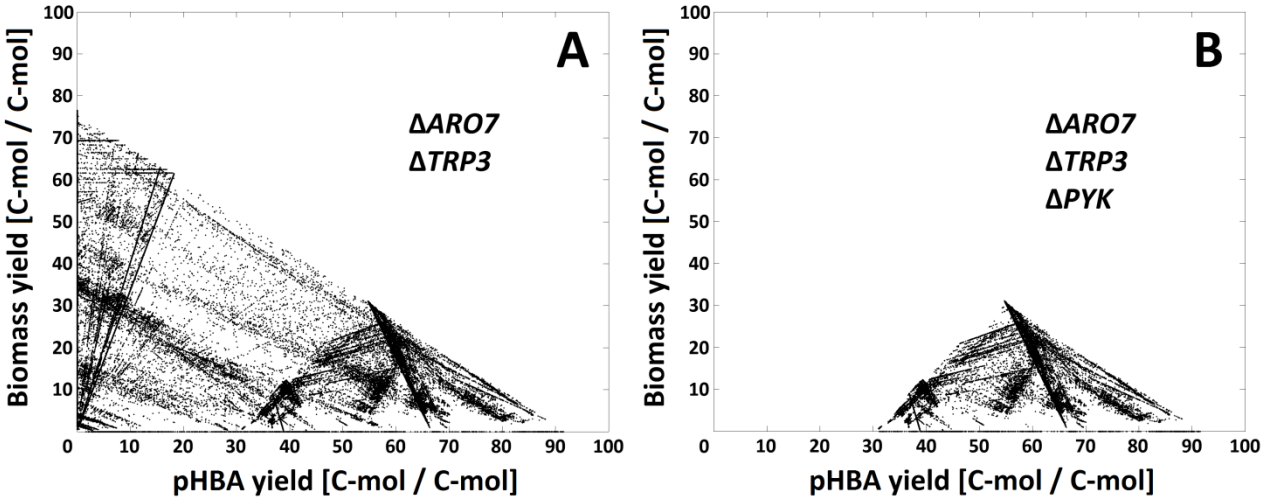


Figure 14: Elementary flux mode distribution of knock-out networks. Each point in a plot represents a feasible flux distribution (elementary mode) the cell can operate in. Elementary modes are shown with relative contributions to biomass (y-axes) and product formation (x-axes) for a network of *S. cerevisiae* on glucose as the carbon-source with (A) chorismate mutase and anthranilate synthase knock-outs and with (B) pyruvate kinase also removed. The deletion of the pyruvate kinase function results in coupling of growth (biomass formation) to product formation. This confers a minimum productivity to the metabolism in respect to pHBA formation, even in a state of growth arrest (biomass yield = 0).

2.3.4.2 Interfacing pHBA production with the QS regulatory network and RNAi

Argonaute (*AGO1*) and Dicer (*DCR1*) genes from the related yeast *S. castelii* were imported into *S. cerevisiae*, so that mRNAs could be targeted for degradation via the transcription of double-stranded hairpin RNA with complementarity to the target (Drinnenberg *et al.*, 2009).

Expression of the relevant genes to establish transcription of hairpin RNA constructs and enhanced pHBA production was required to be controlled by the pheromone responsive *FUS1J2* promoter (Ingolia & Murray, 2007). When an AAA such as tryptophan is present in the medium the α-pheromone peptide is expressed from the *ARO9* promoter and secreted from the cell (Williams *et al.*, 2013). At a high population density the extracellular pheromone triggers gene expression from the *FUS1J2* promoter and causes cells to arrest growth in the G1 phase of the cell-cycle (cf. full manuscript in supplementary for a detailed description of the circuit). This mechanism serves the decoupling of population growth and production phases, allowing growth limiting engineering strategies to be implemented. During the pheromone auto-induced growth arrest, pHBA production is switched on (Figure 16) via the expression of the relevant pathway genes and RNAi repression constructs from the *FUS1J2* promoter. Specifically, several pathway enzymes were dynamically expressed along with the knock-down targets (Figure 15). Previous work has identified TKL1p and the feedback-inhibition resistant ARO4^{K229L}p as important enzymes for the production of shikimate-derived compounds (Luttik *et al.*, 2008, Curran *et al.*, 2013). Finally ubiCp, the enzyme that catalyses pHBA production from chorismate (Siebert *et al.*, 1994), was also expressed under QS regulation to dynamically control production.

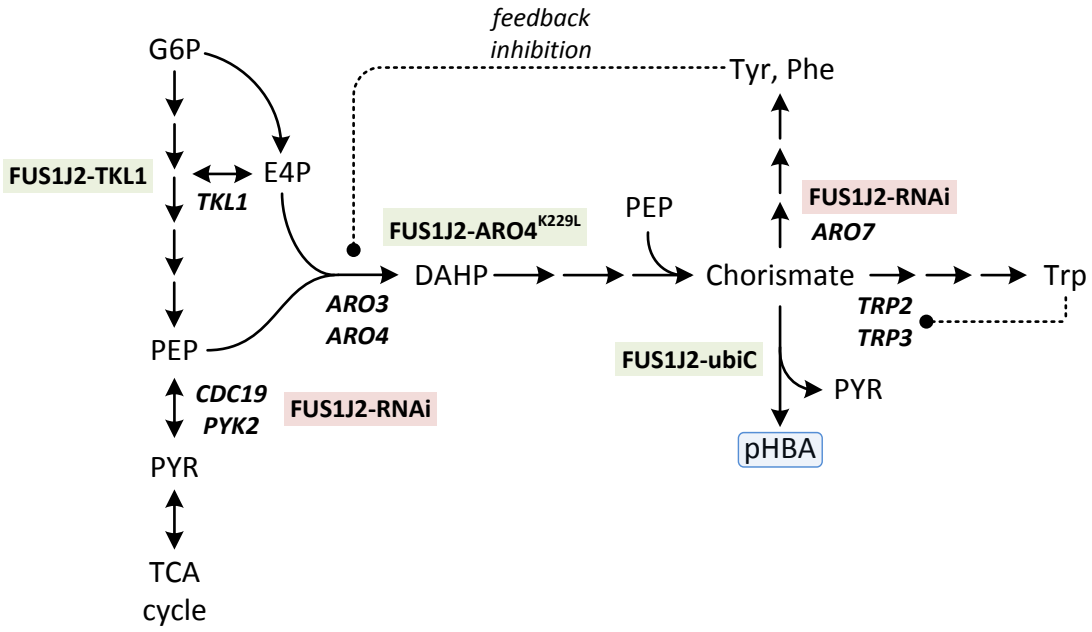


Figure 15: The shikimate pathway feeds AAA biosynthesis using PEP and E4P. The supply of these precursors was increased by overexpressing *TKL1* and targeting *CDC19* for repression using RNA interference. In order to enhance flux through the shikimate pathway in response to pheromone, a feedback resistant version of *ARO4* (K229L) was placed under the control of the *FUS1J2* promoter. *S. cerevisiae* has no known native enzyme for pHBA formation, therefore, a codon optimised version of the *E. coli* chorismate pyruvate lyase (*ubiC*) was incorporated under the control of the *FUS1J2* promoter to catalyse pHBA formation from chorismate. Chorismate availability was increased via the expression of an RNAi repression construct for *ARO7*.

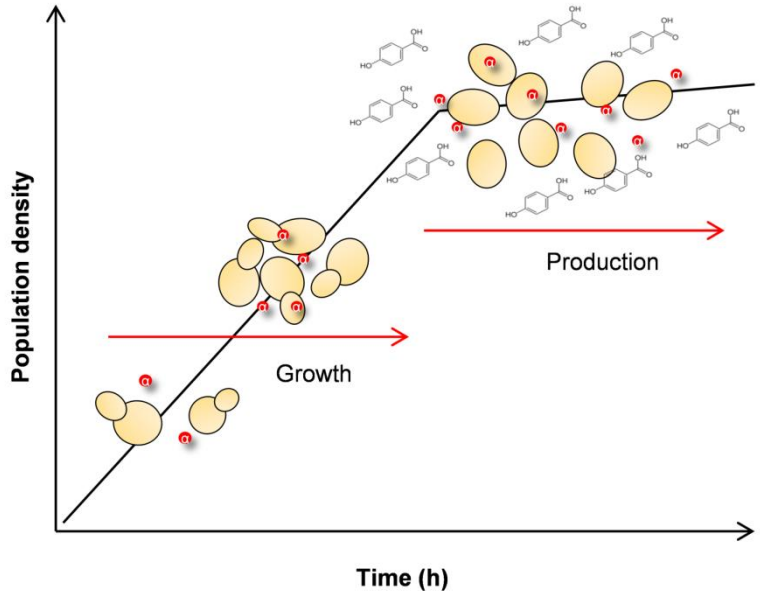


Figure 16: As a quorum sensing population grows over time the concentration of α-pheromone increases. At a high population density a critical α-pheromone concentration is reached which triggers gene expression from pheromone responsive promoters. Cells arrest growth and switch on the production of pHBA as part of a coordinated production phase (figure by (Williams *et al.*, 2015)).

2.3.4.3 Stepwise verification of modification for improved pHBA production

pHBA titers were measured with and without QS activated in the same strain after the introduction of each genetic modification (Figure 17). In every case pHBA production was significantly lower without the activation of the QS circuit, demonstrating the dynamic functionality and range of the engineered regulation.

QS-mediated expression of a codon-optimised *E. coli* chorismate lyase (*ubiC*) in strain PHB04 resulted in pHBA production reaching 46 µM. This was a 77% increase over the tryptophan-treated control strain (PHB03), which contained only the QS and RNAi modules and produced endogenous levels of pHBA (Figure 17). The addition of purified α-pheromone to yeast cultures has previously been shown to result in cell death while deletion of the *FUS1* gene (involved in facilitating membrane fusion of mating partners) reduced mortality from 30% of a population to 5% (Zhang *et al.*, 2006). Given that the QS system relies on the production of α-pheromone it was possible that cell death in response to pheromone would limit pHBA production. Therefore increased pHBA production in a quorum sensing strain which had the *FUS1* gene deleted was expected (PHB05). Interestingly, pHBA accumulation was not significantly different from PHB04. This suggests that cell death in response to pheromone was not a limiting factor in this system.

The additional QS-mediated expression of a feedback resistant 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase enzyme (*ARO4*^{K229L}) increased PHBA titer a further 5.7-fold to 297 µM (Figure 17; strain PHB06). This is consistent with previous data showing that over-expression of *ARO4*^{K229L} enhances flux through the shikimate pathway 4-5 fold (Luttik *et al.*, 2008).

All strains carried an RNAi module (in the form of Argonaute and Dicer genes). When an *ARO7*-targeted RNAi hairpin was expressed, pHBA titer increased 80% (in respect to PHB06) to 534 µM (Figure 17, strain PHB07). This suggests that conditional repression of Aro7p during the production phase increases chorismate availability for the ubiCp enzyme (Figure 15).

It has previously been demonstrated that the overexpression of *TKL1* re-wires pentose phosphate pathway flux to increase erythrose 4-phosphate supply to the shikimate pathway (Curran *et al.*, 2013). When pheromone-responsive *TKL1* expression was introduced, pHBA titer increased by further 49% to 795 µM (strain PHB08). Expression of *TKL1* also increased the pHBA titer in the non-QS-activated control cultures, although this effect was not as prominent in subsequent strains. This effect can be attributed to the combinatorial leaky expression of pathway components from the *FUS1J2* promoter.

In order to fully exploit the use of dynamic regulation, the *CDC19* gene was conditionally repressed. *CDC19* encodes a pyruvate kinase enzyme; Δ*CDC19* strains are not viable for growth on glucose (Giaever *et al.*, 2002). QS-linked RNAi-mediated repression of *CDC19* resulted in a further 35% increase in pHBA titer over the previous strain (PHB08), to 1070 µM (Figure 17; strain PHB10). This proved that the *in silico* determined knock-out strategy is valid *in vivo*. This final strain represented a 37-fold increase in titer over the base strain (PHB03), and a yield of 4.93 mg/$g_{Glucose}$.

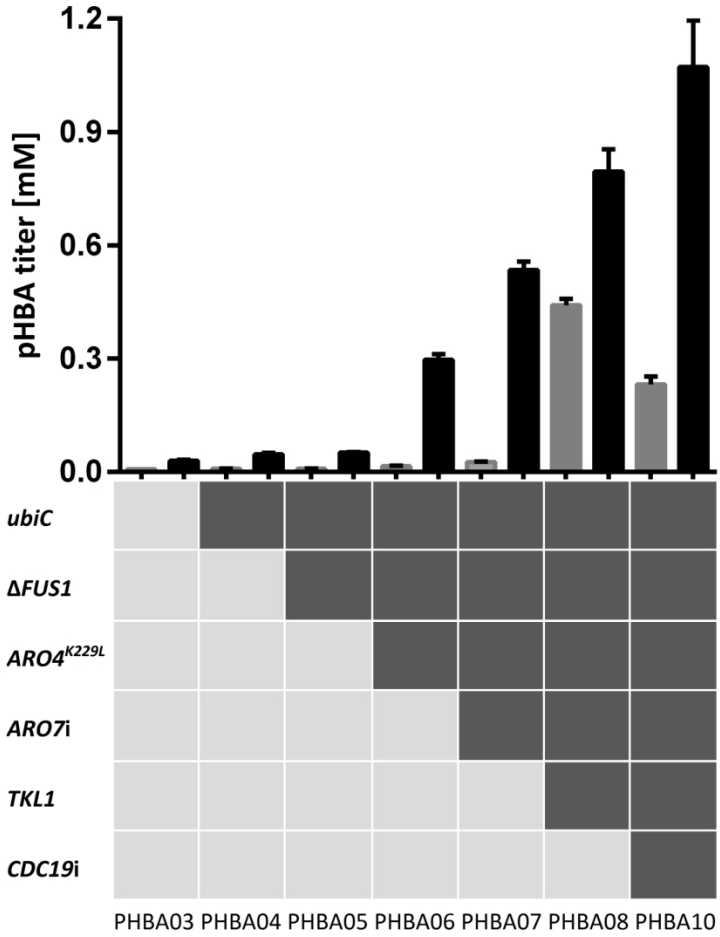


Figure 17: pHBA titers of shake-flask fermentations using quorum sensing mediated modulation of target genes. Black bars indicate active QS while grey bars indicate inactive QS. Light squares represent the absence of the listed modification while tinted squares indicate the presence. *ubiC*, *ARO4*^{K229L} and *TKL1* were up-regulated, while *ARO7*i and *CDC19*i represent the introduction of RNAi hairpins which target the corresponding genes for repression. Δ*FUS1* indicates the absence of the gene (deletion). Strain names are listed below each column. The control strain PHB03 only had a QS circuit and RNAi module (figure reproduced from (Williams *et al.*, 2015)).

2.3.4.4 Validation of separated growth and production

Correlating growth and production profiles of PHB10 over the course of the fermentation verified that the rate of pHBA production was highest when cell division arrested due to induction of QS by means of a sharp increase in pHBA concentration (data not shown, cf. full manuscript in supplementary). Without QS activation the same strains grew to a higher population density. However QS populations continued to increase in optical density due to increasing cell size, but ultimately reached a much lower level than non-QS populations due to the arrest of cell division. This is due to cells continuing to grow larger than their counterparts in non-QS populations while arrested in G1 phase (Bardwell, 2005).

2.3.4.5 Level of repression and residual enzyme activity of knock-down targets

RNAi module functionality was examined by means of the level of target gene mRNA using RT-qPCR. *ARO7* and *CDC19* mRNA levels were compared between the final production strain (PHB10) and a strain with no hairpin RNA expression constructs but which is otherwise identical (PHB11, cf. full manuscript in supplementary).

Expression of the *ARO7* RNAi construct appeared to be leaky, as mRNA levels of *ARO7* were already significantly lower compared to the control, prior to induction. This was surprising given that the *FUS1J2* promoter is known to have very low basal expression levels in the absence of pheromone (Ingolia & Murray, 2007) but reflects the sensitivity of the RNAi system, which was previously noted for its extreme leakiness (Drinnenberg *et al.*, 2009). *CDC19* mRNA levels were not significantly different compared to the control. This difference can be attributed to the fact that *CDC19* transcript was highly abundant: mRNA levels were 132 times higher than those observed for *ARO7* prior to the activation of QS, so that leaky expression was insufficient to repress *CDC19* (before QS activation).

When QS response was fully activated, there was significant further repression of *ARO7* mRNA, and repression of *CDC19* mRNA was evident. Late in the fermentation *ARO7* and *CDC19* repression decreased. This was most likely due to weaker hairpin expression from the *FUS1J2* promoter as cells become gradually desensitized to pheromone over time (Bardwell, 2005).

Further enzyme activity assays were used to assess whether the RNAi system was effectively decreasing protein levels for the target genes, as protein abundance and mRNA levels are often not well correlated in yeast (Greenbaum *et al.*, 2003). Chorismate mutase (ARO7p) activity was significantly reduced to 17.6% of that observed in the control strain. Pyruvate kinase (result of CDC19p and PYK2p) activity was reduced to 11.1% of the control (cf. full manuscript in supplementary for details).

2.3.4.6 Correlation of in vivo results with metabolic modelling

Although the dynamic repression of the essential *CDC19* gene resulted in a significant increase in pHBA titer, the yield was nowhere near that predicted by elementary mode analysis of the pyruvate kinase knock-out (Figure 14, ≥ 30% carbon yield when pyruvate kinase reaction is removed) whereas the carbon yield obtained using RNAi knock-down of *CDC19* was 0.75% [C-mol/C-mol] of glucose). This can be attributed to several factors, first of all the dynamic RNAi approach did not completely remove target mRNA from the system, even though there was a dramatic repression effect. Although *CDC19* was effectively repressed by the hairpin construct, there was still residual pyruvate kinase

activity. This may be attributed to the expression of the *CDC19* paralog *PYK2*, which is transcriptionally de-repressed under low glucose conditions (Boles *et al.*, 1997). Rather than a knock-out this knock-down was anticipated and intended to support the viability of the induced cells. Other explanations for the discrepancy between theoretical and actual yields regard the assumptions taken by the modelling approach. Namely, steady-state growth with glucose as the sole carbon-source, further enzyme kinetics and product inhibition are not considered. These assumptions are violated by the dynamic nature of the system, the fact that little is known about any metabolic perturbations associated with the unique pheromone response phenotype and G1 cell-cycle arrest, and the shift from glucose to ethanol consumption. In fact, when assuming ethanol as sole carbon-source in elementary mode analysis, the minimum yield is reduced to 1.5% (Figure 18), which is significantly closer to what has been observed. This makes sense, as the knock-downs, and therefore, the pHBA production phenotype, are only fully active after a certain growth phase, where presumably a major fraction of glucose already has been consumed (without significant amounts of pHBA being produced) and / or converted into ethanol, which serves as carbon-source after the diauxic shift.

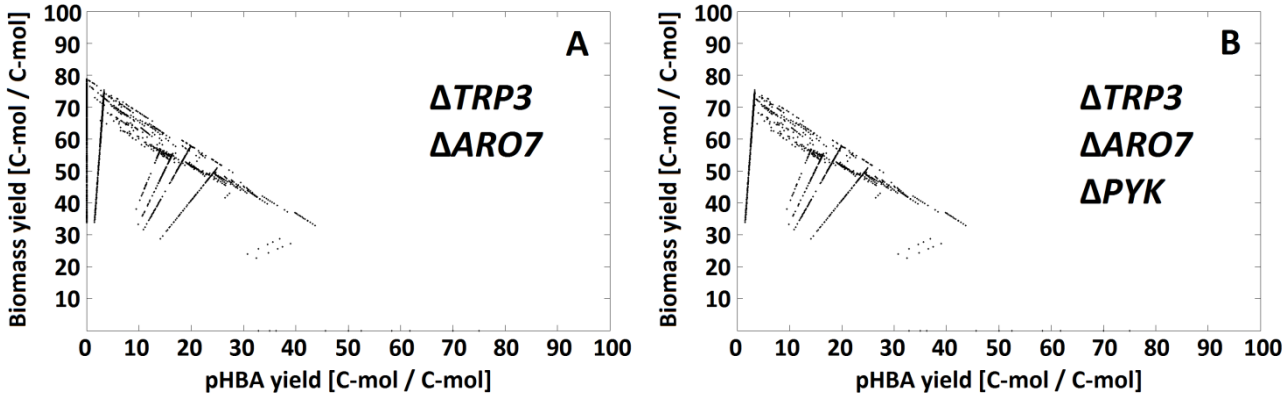


Figure 18: Elementary flux mode distribution of knock-out networks. Each point in the plot represents a feasible flux distribution (elementary mode) the cell can operate in. Elementary modes are shown with relative contributions to biomass (y-axes) and product formation (x-axes) for a network of *S. cerevisiae* on ethanol as carbon-source with (A) chorismate mutase and anthranilate synthase knock-outs and with (B) pyruvate kinase also removed. The coupling of growth (biomass formation) to product formation by deletion of the pyruvate kinase function confers a minimum productivity to the metabolism in respect to pHBA formation, even in a state of growth arrest (biomass yield = 0).

The reason, why on ethanol still a minimum constraint, although lower than on glucose, exists, lies in the fact that *S. cerevisiae* is a compartmentalized organism. This means that firstly not all central metabolism reactions, especially those of TCA-cycle, are present in mitochondria and cytosol, and secondly not all metabolites are transportable between the compartments. If now ethanol feeds via acetyl-CoA directly into the TCA-cycle, intermediates can transfer into the mitochondrion and PYR can again be derived via the malic enzyme. This explains the lower minimum yield. However, only the mitochondrial (although the major) share of the PYR required for biomass formation can be obtained this way. PYR is not easily transportable from the mitochondria back into the cytosol, and no other reaction that delivers PYR exists in the cytosol (Averesch & Krömer, 2014). PEP, however, can be formed in the cytosol by phosphoenolpyruvate carboxykinase. Conversion of cytosolic PEP to PYR can still only proceed via shikimate pathway in combination with product formation. This explains why still a minimum yield constraint exists. For details on the biochemical and transport reactions present in *S. cerevisiae* central metabolism refer to the metabolic network, supplementary file to the manuscript incorporated in this chapter and further the metabolic maps, supplementary files to the manuscript incorporated in chapter 2.2.

2.3.5 Conclusion

Strain construction strategies that evoke a tight metabolic constraint in order to maximize product formation inevitably achieve this on the cost of growth. Separating growth from production using dynamic regulation of metabolism is a promising approach to implement genetic modifications that are difficult or inconvenient to be implemented constitutively. The principal barrier to realizing dynamic metabolic control in yeast has been the lack of suitable mechanisms for both activating and repressing gene expression. The QS-linked RNAi circuit is capable of achieving these objectives, allowing both the expression and/or repression of genes that would normally prevent biomass accumulation (and hence product titer if controlled constitutively). Using this system, populations maintain normal growth prior to the activation of quorum sensing and the cell-cycle arrested production phase. The maximum titer of 1.07 mM (148 mg/L) was only obtained when combining shikimate pathway modifications with the *in silico* determined global genetic intervention strategy, proving the *in vivo* validity of the determined targets. This 37-fold improvement over the base strain also represents the highest recorded pHBA level in *S. cerevisiae*, with previous efforts using extended bioreactor fermentation resulting in a titer 0.65 mM (Krömer *et al.*, 2013).

2.4 Production of *para*-aminobenzoic acid in *Saccharomyces cerevisiae*[6]

For the production of pABA the same genetic intervention strategy that couples product formation to growth holds valid, however, application *in vivo* appeared to be hampered significantly by the insufficiency of flux to pABA and therefore PYR availability. Thus, focus was here on improving the channelling of flux through shikimate pathway and maximizing pABA formation.

2.4.1 Abstract

Biological production of the aromatic compound *para*-aminobenzoic acid (pABA) is of great interest to the chemical industry. Besides its application in pharmacy and as a crosslinking agent for resins and dyes pABA is a potential precursor for the high-volume aromatic feedstocks terephthalic acid and *para*-phenylenediamine. The yeast *S. cerevisiae* synthesises pABA in the shikimate pathway: Outgoing from the central shikimate pathway intermediate chorismate, pABA is formed in two enzyme-catalysed steps, encoded by the genes *ABZ1* and *ABZ2*. In this study *S. cerevisiae* metabolism was genetically engineered for the overproduction of pABA. Using *in silico* metabolic modelling an observed impact of carbon-source on product yield was investigated and exploited to optimize production.

A strain that incorporated the feedback resistant $ARO4^{K229L}$ and deletions in the *ARO7* and *TRP3* genes, in order to channel flux to chorismate, was used to screen different *ABZ1* and *ABZ2* genes for pABA production. In glucose based shake-flasks fermentations the highest titer (600 µM) was reached when over-expressing the *ABZ1* and *ABZ2* genes from the wine yeast strains AWRI1631 and QA23, respectively. *In silico* metabolic modelling indicated a metabolic advantage for pABA production on glycerol and combined glycerol-ethanol carbon-sources. This was confirmed experimentally, the empirical ideal glycerol to ethanol uptake ratios of 1:2 – 2:1 correlated with the model. A ^{13}C tracer experiment determined that up to 32% of the produced pABA originated from glycerol. Finally, in fed-batch bioreactor experiments pABA titers of 1.73 mM (237 mg/L) and carbon yields of 2.64% could be achieved.

[6] This chapter is based on a manuscript revised and resubmitted to *Microbial Cell Factories* in 2016, paper currently under review

In this study, a combination of genetic engineering and *in silico* modelling has proven to be a complete and advantageous approach to increase pABA production. Especially the enzymes that catalyse the last two steps towards product formation appeared to be crucial to direct flux to pABA. A stoichiometric model for carbon-utilization proved useful to design carbon-source composition, leading to increased pABA production. The reported pABA concentrations and yields are, to date, the highest in *S. cerevisiae* and the second highest in a microbial production system, underlining the great potential of yeast as a cell factory for renewable aromatic feedstocks.

2.4.2 Introduction

2.4.2.1 Significance of bio-derived aromatic feedstocks for chemical industry

Microbial metabolic pathways give rise to many compounds that can potentially substitute currently petroleum-based chemicals with bio-derived ones or replace them with bio-based alternatives. Using metabolic engineering these capabilities are already being exploited to create new organisms that use abundant and renewable feedstocks to efficiently produce an expanding spectrum of valuable chemicals (Jang *et al.*, 2012, Chen & Nielsen, 2013, Becker & Wittmann, 2015). In particular aromatics have great potential for bio-based production as the shikimate pathway gives rise to a wealth of aromatics and derived compounds, with diverse applications in different industries, including the chemical one (Hansen *et al.*, 2009, Curran *et al.*, 2013, Krömer *et al.*, 2013, Jendresen *et al.*, 2015, Rodriguez *et al.*, 2015, Williams *et al.*, 2015). The shikimate pathway intermediate *para*-aminobenzoic acid (pABA) is one of these aromatics with versatile applicability.

2.4.2.2 Aromatic compounds derived from yeast shikimate pathway

pABA is naturally produced by yeast as an intermediate in the shikimate pathway, the central metabolic route leading to the formation of the AAAs phenylalanine, tyrosine, and tryptophan (Braus, 1991). A simplistic reproduced of the pathway can be found in Figure 6. Outgoing from chorismate the AAAs and pABA are formed, synthesis of the latter is comprised of two key enzymes encoded by the genes *ABZ1* (aminodeoxychorismate synthase) and *ABZ2* (aminodeoxychorismate lyase). Regulation of the shikimate pathway is tightly controlled by its end-products,

foremost by inhibition of its initial step by phenylalanine and tyrosine (Luttik *et al.*, 2008). However, a point mutation in the *ARO4* gene (leading to a changed amino acid sequence, $ARO4^{K229L}$) has been reported that makes the gene product largely resistant to feedback inhibition (Luttik *et al.*, 2008).

Despite the biosynthesis route to pABA being present in *S. cerevisiae*, common laboratory strains like S288c and CEN.PK derived ones, suffer a growth deficit if pABA is omitted from media, supposedly due to reduced functionality of the *ABZ1* gene (Ambroset *et al.*, 2011). In the comparative study that examined this, it was also found that the wine yeast EC1118 carries a more functional version of this gene (Ambroset *et al.*, 2011). In fact, a previous study on pABA production in S288c showed that overproduction could be achieved using the *ABZ1* gene from the wine yeast AWRI1631, which is identical to the EC1118 one (Krömer *et al.*, 2013). The study identified yeast as a suitable host, with regard to *in silico* determined theoretical maximum yields (0.53 $g/g_{Glucose}$) and *in vivo* toxicity limits (50% growth rate reduction at 0.62 g/L) (Krömer *et al.*, 2013). However, the production was moderate (250 µM), raising the question, if *ABZ2* might be currently limiting pABA production in yeast. Recently overproduction was reported in *E. coli* (4.8 g/L) where also the choice of the bacterial analogues to *ABZ1* and *ABZ2* (*pabA* & *pabB* / *pabAB* and *pabC*) was shown to be crucial (Koma *et al.*, 2014).

In the present study overproduction of pABA in *S. cerevisiae* from different carbon-sources is improved using genetic engineering and metabolic modelling. In particular, the metabolic bottleneck to pABA formation from chorismate and the use of glycerol as an alternative carbon-source were investigated. Glycerol is a by-product of both the biodiesel and the bioethanol industries and is considered an inexpensive and sustainable feedstock that does not directly compete with resources of the food industry (da Silva *et al.*, 2009, Clomburg & Gonzalez, 2013). In this study a mixed glycerol-ethanol (GLY/ETH) feed with a high glycerol-share was used to maximize pABA production and yield.

2.4.3 Materials and Methods

2.4.3.1 Metabolic network analysis

Metabolic modelling was conducted by means of elementary mode analysis using EFMTool (Terzer & Stelling, 2008). The framework was modified from what has been described before (Averesch & Krömer, 2014). In particular, metabolic networks were amended with additional reactions for pABA formation and glycerol / ethanol uptake. The networks are given in supplementary file 2.4.2.

2.4.3.2 Media and strain maintenance

S. cerevisiae strains were propagated on yeast extract peptone dextrose (YPD) medium, 1 L was composed of 10 g yeast extract, 20 g peptone and 20 g dextrose (D-glucose) in H_2O. For strain construction the antibiotics G418 (geneticin) and clonNAT (nourseothricin) were added into the medium to final concentrations of 200 mg/L and 100 mg/L, respectively. Also for strain construction as well as maintenance selective synthetic complete medium was used. It was made up from 6.8 g/L yeast nitrogen base without amino acids (Y0626 Sigma) and 1.92 g/L drop-out supplement without uracil (Y1501 Sigma) with 20 g/L glucose or 10 g/L glycerol and 30 g/L ethanol as carbon-source. Solid media contained 20 g/L agar.

For cultivation of yeast chemically defined minimal medium (CDM) was used, 1 litre was composed of: 2 g $(NH_4)_2SO_4$, 1 g $(NH_4)_2HPO_4$, 5.28 g NaH_2PO_4, 0.5 g $MgSO_4 \cdot 7\ H_2O$, 0.015 g EDTA, 1 g KCl, 0.15 g $CaCl_2 \cdot 2\ H_2O$, 0.85 g Na_2HPO_4 amended with trace elements and vitamins in aqueous solution. Trace elements (per 1 litre H_2O): 4.5 mg $ZnSO_4 \cdot 7\ H_2O$, 3 mg $Fe(III)Cl_3 \cdot 6\ H_2O$, 1 mg H_3BO_3, 0.85 mg $MnSO_4 \cdot H_2O$,0.4 mg $NaMoO_4 \cdot 2\ H_2O$, 0.3 mg $CoCl_2 \cdot 6\ H_2O$, 0.3 mg $CuSO_4 \cdot 5\ H_2O$, 0.1 mg KI. Vitamins (per 1 litre): 3 mg *myo*-inositol, 1 mg calcium D-pantothenate, 0.3 mg thiamine hydrochloride, 0.075 mg pyridoxine hydrochloride, 0.0015 mg D-biotin. In the case of the knock-out strain 76 mg/L tyrosine, phenylalanine and tryptophan were supplemented, while for the background strain 0.2 mg/L pABA were essential for growth. The pH was adjusted to 6, carbon-sources were glucose (GLC) glycerol (GLY), ethanol (ETH) or a glycerol-ethanol mix (GLY/ETH) with amounts as given in Figure 21 for the respective experiments. For the tracer experiments 20 g/L fully labelled U-^{13}C glucose (Cambridge Isotope Laboratories, Tewksbury, MA) or 2 g/L of fully labelled ^{13}C ethanol (99%, ICON isotopes, Summit, NJ) and 20 g/L unlabelled

glycerol were used as carbon-source. Medium for the constant-feed bioreactor experiment initially contained the two-fold ammonium source and 20 g/L glycerol + 10 g/L ethanol. The feed solution was CDM +ETH with the two-fold ammonium source and 10% ethanol (v/v). For the pulse-feed fermentation, the C-source was changed to 10 g/L glycerol + 30 g/L ethanol. The three feed pulses were of different composition to suit the demands of the fermentation: The first was a 0.76 g/L AAA stock solution, the second CDM +ETH with 100 g/L ethanol and 0.38 g/L AAAs, the third was GLY+ETH 1:2 (w/w).

E. coli plasmid maintenance strains were cultivated in lysogeny broth (LB) supplemented with 100 mg/L ampicillin. Strains were grown at 30°C with 200 rpm orbital shaking. All cryo-cultures were stored in 25% glycerol at -80°C.

2.4.3.3 Strain and plasmid construction

Strain construction was based on the *S. cerevisiae* strain CEN.PK113-5D. *ABZ1* and *ABZ2* genes originated from different sequenced and commercially available wine yeast strains (

Table **5**). Cloning was routinely conducted following standard protocols (White, 2008) and using enzymes obtained from New England Biolabs to manufacturer instructions. In particular Q5® Hot Start High-Fidelity DNA Polymerase and High-Fidelity (HF®) Restriction Endonucleases were used. For sub-cloning, the *E. coli* strain DH5α was utilized. Plasmids and PCR products were purified using the Thermo Scientific GeneJET Plasmid Miniprep and PCR Purification Kits. Cloned PCR products were verified by sequencing (Sanger service by AGRF, Brisbane, AU).

The *TRP3* deletion was carried out by replacing the target gene via homologous recombination, integrating the natMX antibiotic marker as described in (Güldener *et al.*, 1996) using the lithium acetate method to transform the cells (Gietz & Woods, 2002). Successful integrations / deletions were verified by PCR using primers binding inside and outside of the target / marker gene (verA – verD) as outlined in (Güldener *et al.*, 2002). The same principles were used for *ARO7* deletion / $ARO4^{K229L}$ integration with a few differences: After incorporating the point mutation in codon 229 of the original *ARO4* (as already described in (Williams *et al.*, 2015)), which changes the respective amino acid residue from Lys to Leu and renders the protein feedback inhibition resistant, it was cloned to pCV3. The plasmid pCV3-

ARO4^{K229L} was used directly to obtain an integration cassette comprised of the *ARO4*^{K229L} gene fused with the *POL4* terminator and the kanMX marker that would integrate at the *ARO7* locus, thus replacing it and putting the *ARO4*^{K229L} under the control of the *ARO7* promoter, simplifying experimental efforts. The constructed strain (CEN.PK113-5D Δ*trp3*::natMX Δ*aro*7::kanMX-*ARO4*^{K229L}, PABA0) served as the host for the pABA expression plasmids.

Plasmid vectors for expression of the genes for pABA formation from chorismate were constructed based on pSP-G1 (Partow *et al.*, 2010). The *ABZ1* genes were always cloned to the *TEF1* promotor's multi-cloning site while the *ABZ2* genes were cloned into the *PGK1* promotor's polylinker. The primers used to amplify the respective integration / deletion cassettes and cloning inserts were purchased from Integrated DNA Technologies, the sequences can be found in Table 3. For cloning, standard desalted primers were used, while for deletion / integration PAGE-purified primers proved to be significantly more efficient.

Vectors were constructed in a way that allowed conclusive comparison of the different genes against each other without having to compare each gene separately. Plasmids used in this study are given in Table 4. The plasmids were transformed to the PABA0 strain, genotypes of the final transgenic production strains (CEN.PK113-5D Δ*trp*3::natMX Δ*aro*7::kanMX-*ARO4*^{K229L} pSP-G1-*ABZ1-ABZ2*, PABA1 – PABA5) can be found in Table 5.

Table 3: Primers used in this study. Annealing sites in primers for amplification of integration cassettes are shown in red, while overhangs for homologues recombination have been marker blue. For cloning primers the restriction sites are indicated in bold letters.

Name	5' to 3' sequence
TRP3 del for	GCTGGTCAGTATTAACACATTACCAAATCTGTTTGGTTCTACATAGAACGCCCAGCTGAAGCTTCGTACGC
TRP3 del rev	CTTCTATAGGCTACTATACATGAAGGCGTTCGCCCTTACAACGTTAAGTTGCATAGGCCACTAGTGGATC
ARO7p-3'–ARO4-5' for	CGATTCCGTTTTGCTAACAAATAGCACTCAGCATCCTGCATAAAATTGGTATAAGATATGAGTGAATCTCCAATGT TCGCTGCCAACG
ARO7t-5'–TEFt-3' rev	GCGTAGTAATCACTCGGCAATGTGGAATTGTTACCGTGATAGCCTTCATGCTTCGACACTGGATGGCGGCGTTA GTATCG
TRP3 verA	GTTCGACAAGGAGCTGGTCAG
TRP3 verB	CAAGCAGGATGGTAGGGACG
TRP3 verC	ACGTCAGTGGAGGCACTTGG
TRP3 verD	TACATGAAGGCGTTCGCCC
nat verB	GAGGTCACCAACGTCAAC
nat verC	GTAAGCCGTGTCGTCAAG
ARO7 verA	CTTTATCGTAAATTTTCAATGG
ARO7 verD	GGTTATTGGGTTGATAATTCCT
ARO4 verB	GTTGAGATTCGGTGGTTCTGGC

ARO4 verC	TTGGCTGATTTGGTCTCCTTCG
KmR verB	TGTACGGGCGACAGTCACAT
KmR verC	GTGTTCCTGCGCCGGTTGC
ARO7 verA2	CACCTATTGCGCCGCTCGCGGAATACAATTAC
ARO7 verB	GTAATTAACTTCTGGGGCATAAGGCGCC
ARO7 verC	GGTTCTGTTGCCACTAGAGATATAGAATG
ARO7 verD2	CCAAGACTTTATATGCCAATGGAGGGATAGGCGC
$ABZ1_{AWRI1631}$ for BglII	CTGA**AGATCT**AAAAAAATGCTGTCCGATACAATTGAC
$ABZ1_{AWRI1631}$ rev SacI	AGTC**GAGCTC**CTACATGAAAATTTGTAAGTTGC
$ABZ1_{QA23}$ for BglII	CTGA**AGATCT**AAAAAAATGTCATCTTCTGTGATTAGTG
$ABZ1_{QA23}$ rev SacI	AGTC**GAGCTC**CTACATGAAAATTTGCAAGTTGC
$ABZ2_{S288c}$ for BamHI	CTGA**GGATCC**AAAAAAATGTCACTAATGGACAATTG
$ABZ2_{S288c}$ rev XhoI	AGTC**CTCGAG**TCAATATTTTGTCTTCACTGTTCCC
$ABZ2_{AWRI1631}$ for BamHI	CTGA**GGATCC**AAAAAAATGTTCAACGAAAGGCAATGC
$ABZ2_{AWRI1631l}$ rev XhoI	AGTC**CTCGAG**TCAATATTTTGTCTTCACTGTTC
$ABZ2_{QA23}$ for BamHI	CTGA**GGATCC**AAAAAAATGAACGWATTTTTTTTATTAGATGAAC
$ABZ2_{QA23}$ rev XhoI	AGTC**CTCGAG**TCAATATTTTGTCTTCACTGTTCC

Table 4: Plasmids used in this study as a template for amplification of deletion markers and over-expressing of heterologous genes.

Name	Details	Origin
pUG74	PCR template vector, source for loxP-flanked natMX gene deletion marker	(Hegemann & Heick, 2011)
pCV3-*ARO4*^{K229L}	*ARO4*^{K229L} expression vector, source for *ARO4*^{K229L}–kanMX integration cassette	(Williams *et al.*, 2015)
pSP-G1	Ura^{+} selectable double expression vector, contains *TEF1–PGK1* bidirectional promoter	(Partow *et al.*, 2010)
pSP-G1-*ABZ1*$_{AWRI1631}$-*ABZ2*$_{AWRI1631}$½	pSP-G1 with *ABZ1* and *ABZ2* cloned under control of *TEF1* and *PGK1* promoter respectively	this study
pSP-G1-*ABZ1*$_{AWRI1631}$-*ABZ2*$_{AWRI1631}$*full*	pSP-G1 with *ABZ1* and *ABZ2* cloned under control of *TEF1* and *PGK1* promoter respectively	this study
pSP-G1-*ABZ1*$_{AWRI1631}$-*ABZ2*$_{S288c}$	pSP-G1 with *ABZ1* and *ABZ2* cloned under control of *TEF1* and *PGK1* promoter respectively	this study
pSP-G1-*ABZ1*$_{AWRI1631}$-*ABZ2*$_{QA23}$	pSP-G1 with *ABZ1* and *ABZ2* cloned under control of *TEF1* and *PGK1* promoter respectively	this study
pSP-G1-*ABZ1*$_{QA23}$-*ABZ2*$_{QA23}$	pSP-G1 with *ABZ1* and *ABZ2* cloned under control of *TEF1* and *PGK1* promoter respectively	this study

Table 5: *Saccharomyces cerevisiae* strains used in this study as a source of gDNA as PCR templates and host organism for genetic engineering.

Strain	Genotype / Plasmid	Note	Origin
S288c	MATα SUC2 gal2 mal2 mel flo1 flo8-1 hap1 ho bio1 bio6	Reference strain, source of $ABZ2_{S288c}$ http://www.yeastgenome.org/strain/S288C/overview	(Mortimer & Johnston, 1986)
AWRI1631		Wine yeast strain, source of $ABZ1_{AWRI1631}$, $ABZ2_{AWRI1631}$*full* and $ABZ2_{AWRI1631}$½ http://www.yeastgenome.org/strain/AWRI1631/overview	(Borneman *et al.*, 2008)
Lalvin QA23		Wine yeast strain, source of $ABZ1_{QA23}$ and $ABZ2_{QA23}$ http://www.yeastgenome.org/strain/LalvinQA23/overview	(Borneman *et al.*, 2011)
CEN.PK113-5D	MATa ura3-52 MAL2-8^{c} SUC2	Background strain used to construct the production strains http://www.yeastgenome.org/strain/CENPK/overview	Euroscarf
PABA0	CEN.PK113-5D Δ*trp*3::natMX Δ*aro*7::$ARO4^{K229L}$-kanMX	Indole-3-glycerol-phosphate synthase and chorismate mutase knock-out strain with feedback resistant 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase	this study
PABA1	PABA0 pSP-G1-$ABZ1_{AWRI1631}$-$ABZ2_{AWRI1631}$½	PABA0 strain over-expressing AWRI1631 aminodeoxychorismate synthase and AWRI1631 aminodeoxychorismate lyase	this study
PABA2	PABA0 pSP-G1-$ABZ1_{AWRI1631}$-$ABZ2_{AWRI1631}$*full*	PABA0 strain over-expressing AWRI1631 aminodeoxychorismate synthase and AWRI1631 aminodeoxychorismate lyase	this study
PABA3	PABA0 pSP-G1-$ABZ1_{AWRI1631}$-$ABZ2_{S288c}$	PABA0 strain over-expressing AWRI1631 aminodeoxychorismate synthase and S288c aminodeoxychorismate lyase	this study
PABA4	PABA0 pSP-G1-$ABZ1_{AWRI1631}$-$ABZ2_{QA23}$	PABA0 strain over-expressing AWRI1631 aminodeoxychorismate synthase and QA23 aminodeoxychorismate lyase	this study
PABA5	PABA0 pSP-G1-$ABZ1_{QA23}$-$ABZ2_{QA23}$	PABA0 strain over-expressing QA23 aminodeoxychorismate synthase and QA23 aminodeoxychorismate lyase	this study

2.4.3.4 Strain cultivation

250 mL Erlenmeyer baffled cell culture vent cap polycarbonate flasks (Corning®, Corning, NY) were used for shake-flask experiments (culture volume being 10% of total) using a Multitron 3-stack incubation shaker with 25 mm shaking throw at 200 rpm (INFORS HT, Bottmingen, Schweiz) for incubation of shake-flasks. Experiments were carried out in biological triplicates, single colonies from solid medium were used to inoculate primary pre-cultures on liquid medium (CDM). After overnight growth, a secondary pre-culture was inoculated to a defined OD_{660}. Cells from log-phase were used to inoculate the main-culture to an OD_{660} of 0.8 in the case of *ABZ1*/*ABZ2* performance screening on glucose and 0.2 for all other shake-flask experiments. Growth was monitored by means of OD_{660} and pH, collecting supernatant for analysis every three hours during exponential growth and twice a day in stationary phase (after 9 h in case of GLC, after 24 h in case of GLY/ETH).

Fed-batch bioreactor experiments were carried out in duplicates using a BIOSTAT® B (Sartorius, Göttingen, Germany). Shake-flask pre-cultures on GLY/ETH were kept exponential (by transferring cells from cultures approaching end log-phase to fresh medium, while adjusting volume to not exceed a final OD of 5 in order to suffice oxygen demand) until enough cells were collected to inoculate 2x400 mL medium to an OD of 1.8 or 4.4 for the respective experiments. Cells were washed twice with fresh medium, resuspended in 24 mL thereof and transferred into the bioreactors. Feeding was started when initial ethanol was nearing depletion (inferred by retardation of growth by means of OD). In case of the pulse-feed experiment the three doses were added to the fermentation (10-fold concentrated AAA stock added 1:10, concentrated CDM +ETH added 1:10, GLY/ETH mix added 1:38.5) at once at the indicated time points (Figure 24). In the case of constant-feed experiment, flow was set to 100 µL/min, corresponding to the approximate ethanol uptake rate (predetermined in shake-flasks), adjusted to the actual OD. In the final stage (cf. Figure 24) the feed was increased by 3.5-fold. Aeration was minimized (0.15 L/min in case of the pulse-feed and 0.16 L/min in case of the continuous-feed experiment) in order to minimize stripping and foaming. However, the system was controlled to be aerobic with dissolved oxygen no lower than 80%. The pH was controlled at 6 using 1 M H_3PO_4 and 10% NH_4OH. All fermentations were conducted at 30°C.

2.4.3.5 Metabolite analysis

Using an advancement of the method described before (Williams *et al.*, 2015) extracellular concentrations of aromatic compounds (pABA, anthranilate, pHBA, PHE, TYR, TRP, PEA, tryptophol and folate) were determined. Specifically, cells were removed from 1 mL culture by centrifugation at 4°C, for 5 min. at 13,000×g. Clarified supernatant was stored in a microcentrifuge tube at -20°C until analysis. Metabolites were analysed by RP-HPLC using an Agilent 1200 HPLC system. In brief, samples were kept at 4°C in a high-performance auto-sampler (Agilent HiP-ALS, G1367B). 30 µL was injected into a C18 column (Zorbax Extend C18, 3.5 µm, 4.6 x 150 mm, Agilent PN: 763953-902) with a guard column (SecurityGuard Gemini C18, Phenomenex PN: AJO-7597). The column temperature was kept at 35°C in a thermostatted column compartment (Agilent TCC, G1316B), and analytes were eluted using gradient chromatography at 1 mL/min. flow rate. The gradient method was as follows: 5 – 35% B from 0 – 5 min., 35 – 40% B from 5 – 14 min., 100% B from 15 – 18 min., and 5% B from 18.1 – 20 min. using a binary pump (Agilent Bin Pump, G1312A). Solvent B was 0.045% TFA in 80% acetonitrile in water while Solvent A was 0.045% TFA in water. Solvents were degassed using an in-line degasser (Agilent Degasser, G1379B). Analytes of interest were monitored using a diode array detector (Agilent DAD SL, G1315C) at 206 and 254 nm wavelengths. Spectral scans were also performed on each eluting peaks from 190 – 400 nm to confirm their spectral fingerprint and purity. Chromatograms were integrated using the Agilent ChemStation chromatography software.

Extracellular glucose, glycerol, ethanol and organic acids were measured as described previously (Dietmair *et al.*, 2010) in cell-free supernatants collected as mentioned earlier.

The metabolite data were analysed and integrated with the fermentation data using Microsoft Excel. To determination yields accurately, evaporation was respected: For shake-flasks evaporation of total liquid was estimated gravimetrically, using a predetermined correlation between ethanol concentration and evaporation of ethanol from CDM +GLY/ETH overtime loss of ethanol during fermentation was estimated incrementally. The errors of the yields were based on maximum and minimum yields determined based on boundaries of ethanol evaporation, as this exceeded standard deviation between the triplicates. For the bioreactor experiments, the method described in (Duboc & von Stockar, 1998) was employed to determine evaporation of water and stripping of ethanol. Yields (in percent C-mol/C-mol) were calculated from the evaporation corrected titers while only the metabolised fraction of the carbon-source was respected.

^{13}C labelled tracer cultivations were performed on fully labelled glucose (D-Glucose–$^{13}C_6$ 99 atom%; Isotec, Germany) or on a mixture of fully labelled ethanol ($^{13}C_2$ 99 atom%; Icon Isotopes, USA) and naturally enriched glycerol. Amino acid labellings were obtained from cell hydrolysates (filtrate of cell pellet incubated with 250 µL 1 M HCl @ 100°C overnight) and analysed using gas chromatography – mass spectrometry (GC-MS) described previously (Winter *et al.*, 2014). For this 0.2 mL of supernatant sample or 0.05 mL of cell-hydrolysate were dried under a stream of N_2 and subsequently derivatized with N-methyl-tert-butyldimethylsilyl-trifluoroacetamide (MBDSTFA; Macherey-Nagel, Düren, Germany) as described before (Krömer *et al.*, 2013, Winter *et al.*, 2014). For pABA the ion cluster m/z = 308-315 was analysed. It represents the m-57 fragment (full carbon backbone of pABA) of MBDSTFA derivatized pABA. The volatile PEA was extracted into hexane (equal volume to sample, vortex, decant) and directly injected into GC-MS. For PEA, GC-MS settings were modified from the above, by applying a 1:10 split injection. The ion cluster m/z 122 – 130 contained the full carbon backbone of PEA and was used to calculate enrichment. Generally, labelling distributions were corrected for all naturally occurring stable isotopes in the derivatized molecules that do not belong to the biological carbon backbone using the software IsoCor (Millard *et al.*, 2012). The corrected mass distributions were then used in IsoCor to calculate the mean enrichment of compounds using relative summed fractional labelling (SFL) calculated as follows:

$$SFL = (0\times I_{m+0} + 1\times I_{m+1} + 2\times I_{m+2} \ldots + n\times I_{m+n}) / n$$

Where *I* is the intensity of the (*m+i*) ion in the ion cluster and *n* the number of carbon atoms in the analysed mass fragment.

2.4.4 Results and Discussion

2.4.4.1 Designing a strain for pABA production - screening of different ABZ1 and ABZ2 genes connects pABA production to gene sequences

The production of pABA in *S. cerevisiae* requires diversion of carbon flux towards pABA synthesis through the elimination of competing metabolic pathways and enhancement of pABA synthesis through over-expression of key enzymes. Therefore, the *TRP3* and *ARO7* genes were deleted and the feedback inhibition resistant $ARO4^{K229L}$ gene was introduced, in order to overcome product inhibition and increase flux to the shikimate pathway (Curran *et al.*, 2013, Rodriguez *et al.*, 2015, Williams *et al.*, 2015). In the resulting genetically modified base strain "PABA0" (cf. Figure 6 & Table 5) $ARO4^{K229L}$ was integrated at the *ARO7* locus, thus disrupting the *ARO7* gene and putting the $ARO4^{K229L}$ under control of *ARO7*'s native promoter in order to achieve a positive feedback loop for high flux to shikimate pathway due to AAA starvation.

PABA0 was used as a chassis cell to screen different alleles of *ABZ1* and *ABZ2* for their effect on pABA production. To start with, sequences were compared *in silico*, by means of the alignment of the 43 different *ABZ1* and *ABZ2* nucleotide and amino acid sequences, derived from SGD (www.yeastgenome.org). With the aim to cover the greatest possible variety, the alleles of the lab yeast (S288c) and three wine yeasts (AWRI1631, AWRI796, QA23) were chosen. These *ABZ1* and *ABZ2* genes were compared in detail regarding sequence similarities and differences (Figure 19, cf. supplementary file 2.4.1 for full sequence analysis).

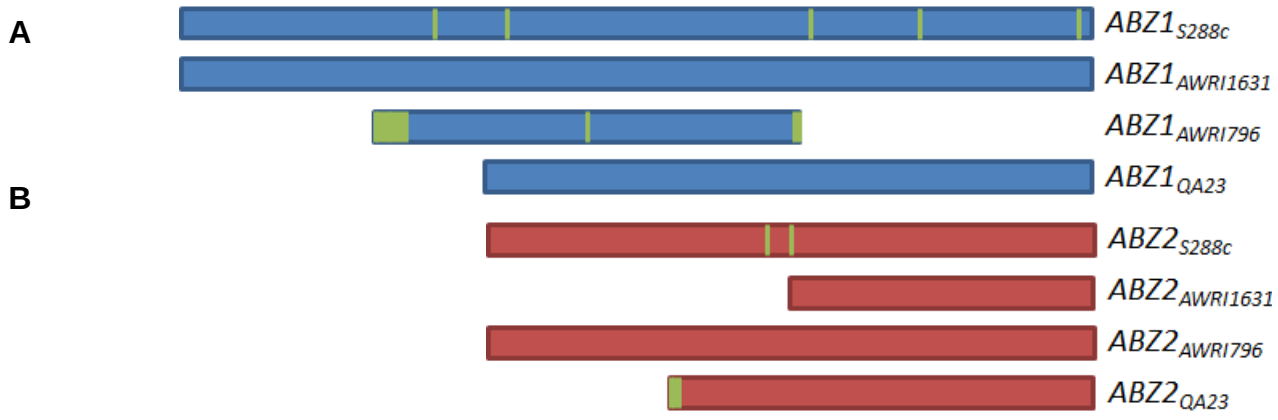


Figure 19: Simplified depiction of the alignment of *ABZ1* and *ABZ2* protein sequences. Amino acid residues unique to a certain sequence are indicated in green. Sequences are shown from N-terminus to C-terminus.

The *ABZ1* genes (Figure 19A) distinguished not only by SNPs, but also appeared in different ORFs. So did the *ABZ2* genes, furthermore it was found that the region upstream of the annotated ORF of $ABZ2_{AWRI1631}$ coincides with the nucleotide sequence present in the genome of AWRI796. This means that the full ORF of $ABZ2_{AWRI796}$ is present also in AWRI1631. This may indicate a false annotation of either gene. Therefore, both possible ORFs (cf. Figure 19B) were included for screening, the putatively falsely annotated gene is referred to as $ABZ2_{AWRI1631}½$ (which corresponds to the original annotation of $ABZ2_{AWRI1631}$) the other one as $ABZ2_{AWRI1631}$full (which corresponds to the original annotation of $ABZ2_{AWRI796}$). The yeast strain Lalvin QA23 provided other substantially different *ABZ1* and *ABZ2* genes. All four *ABZ2* genes and the *ABZ1* genes of AWRI1631 and QA23 were chosen for *in vivo* comparison.

The strains “PABA1” – “PABA5” (cf. Table 5), which over-expressed the different *ABZ1* & *ABZ2* genes, were compared in shake-flask experiments with the aim to find the best set of *ABZ1* and *ABZ2* genes for high pABA production. PABA0 already exhibited baseline pABA production with respect to the background strain (CEN.PK113-5D), indicating that the modifications so far already directed flux towards pABA, even without increased expression of *ABZ1* and *ABZ2*. This was supported by the observation that the strain grew without supplementary pABA. Over-expression of *ABZ1* and *ABZ2* led in all cases to higher pABA production than PABA0. The highest production was achieved by the combination of $ABZ1_{AWRI1631}$ and $ABZ2_{QA23}$, which showed a ~35-fold increase in pABA concentration compared to PABA0. Further, it can be noted that the full $ABZ2_{AWRI1631}$ gene in PABA3 led to a higher production than the truncated version in PABA2. Finally, PABA5, harbouring $ABZ1_{QA23}$ in combination with $ABZ2_{QA23}$, was inferior compared to PABA2, PABA3, and PABA4.

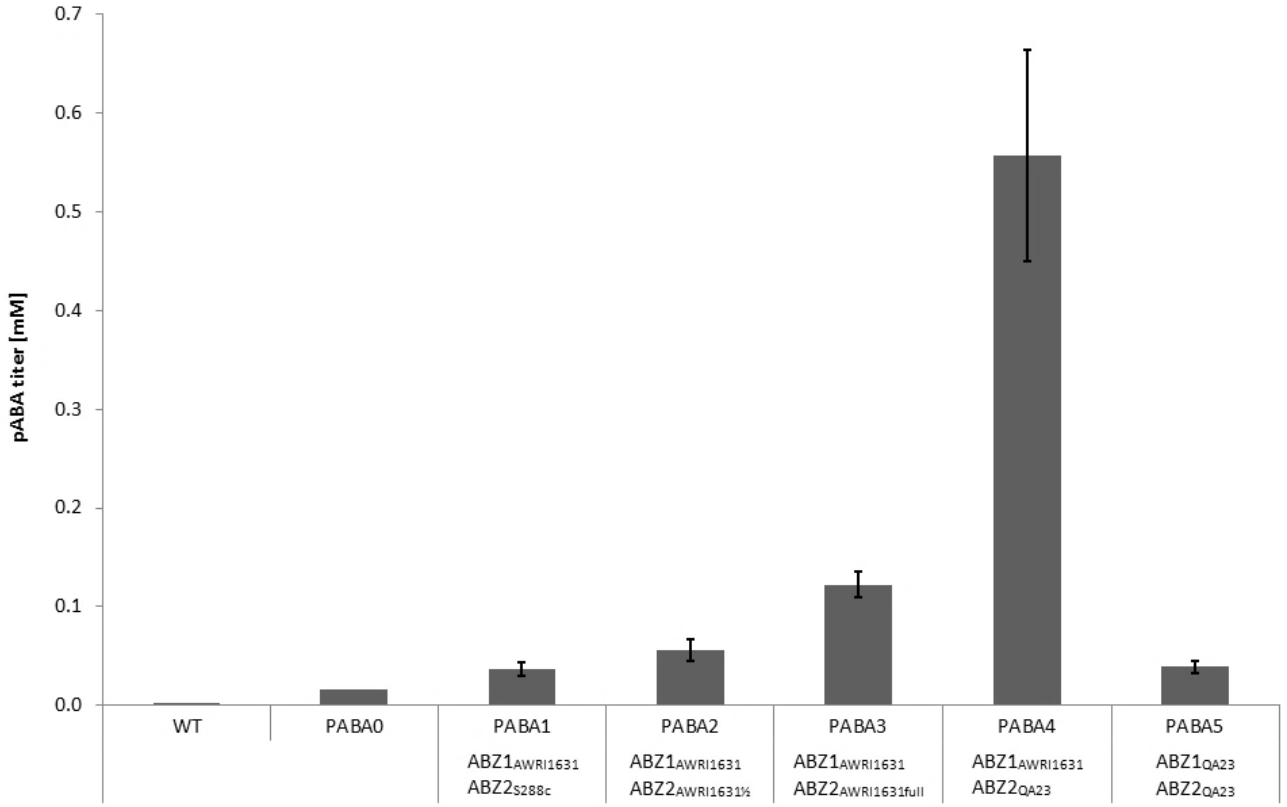


Figure 20: Production of pABA in shake flask experiments of *S. cerevisiae* strains. WT = CEN.PK113-5D, PABA0 = CEN.PK113-5D Δ*trp3* Δ*aro7* *ARO4*^{K229L}. PABA1 – PABA5 are based on PABA0, with over-expression of additional *ABZ1* and *ABZ2* alleles on a plasmid expression vector as indicated on the figure (for full genotype cf. Table 5).

Inferred from the marginal pABA production of strain PABA1 it appears that also the *ABZ2* gene of S288c and *S. cerevisiae* strains with an identical sequence is impaired. The most obvious explanation seems to be that the two differences in the amino acid sequence of *ABZ2*$_{S288c}$ in respect to *ABZ2*$_{AWRI796}$ (cf. Figure 19) reduce the functionality of the enzyme. With PABA3 producing nearly twice the amount of pABA compared to PABA2 it is evident that *ABZ2*$_{AWRI1631}$ was annotated falsely in the AWRI1631 genome, and the larger ORF (corresponding to *ABZ2*$_{AWRI796}$) is the correct one. Comparing the sequence of *ABZ2*$_{QA23}$ in PABA4, which excel in pABA production, to the other genes it seems very likely that some relation between the unique C-terminus of *ABZ2*$_{QA23}$ and its effect on pABA production exists. Comparing PABA4 to PABA5 shows that also the choice of *ABZ1* has a great influence, as only combining *ABZ1*$_{AWRI1631}$ with *ABZ2*$_{QA23}$ allowed greatly increased pABA titers. It seems likely that *ABZ1*$_{QA23}$ is dysfunctional due to the 5'-end truncation of the gene. However, in general, the differences that arise for pABA production by employing different *ABZ* genes may not be only due to differences in activity of the enzymes but also enzyme concentration. This may involve more complex regulation on transcriptional and translational levels for example in relation to codon usage and/or mRNA stability.

Investigating this in detail may provide advanced knowledge that will allow further optimizations. Relating ligand binding sites to the mutations in amino acid sequences of the different ABZ genes could, for instance, allow conclusions towards enzyme functionality (cf. supplementary file 2.4.1).

2.4.4.2 In silico analysis of carbon-source composition leads to enhanced pABA production in vivo

During characterization experiments of the PABA strains it could be observed that the production of pABA was higher later in the fermentation when glucose was depleted and glycerol & ethanol were reassimilated (cf. section 2.4.4.3). To understand this phenotype *in silico* modelling was used: Elementary mode analysis was performed analogous to a previous study (Averesch & Krömer, 2014), in order to compare biomass and product carbon-yields on the substrates glucose, glycerol, ethanol and glycerol + ethanol. The achievable maximum carbon yield for pABA from ethanol is generally lower than from glucose or glycerol (Table 6, cf. supplementary file 2.4.2 for distribution of elementary flux modes) while glycerol allows a higher theoretical pABA yield than glucose. The highest possible carbon yield for pABA with and without biomass formation can be achieved *in silico* from a combination of glycerol and ethanol as feed, which may explain our experimental observation of high pABA formation in the second fermentation phase when reutilizing ethanol and/or glycerol.

Table 6: Theoretical maximum product carbon yields using different substrates.

Substrate	Glucose	Ethanol	Glycerol	Glycerol+Ethanol
Maximum pABA yield [%]	84	75	90	92
Maximum pABA yield with biomass [%]	81	50	86	87

The analysis of the product yield from all elementary modes on GLY/ETH-feed allowed the determination of an ideal ratio between glycerol and ethanol. Figure 21 shows that the majority of feasible flux distributions feature a GLY:ETH ratio lower than 10 while the highest product yields are obtained with GLY:ETH ratios between 0.5 and 2 (magnified area in Figure 21).

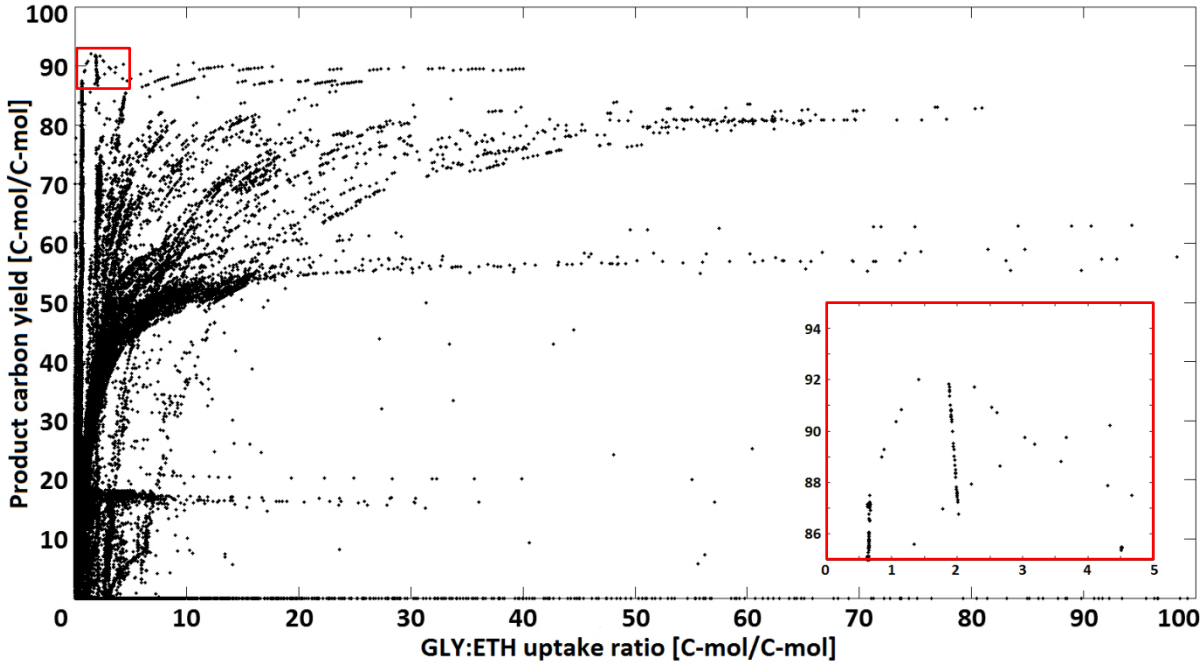


Figure 21: Distribution of pABA yields in dependence of the glycerol to ethanol ratio in elementary flux modes in a window with a GLY:ETH ratio ≤ 100 [C-mol/C-mol]. Each point in the chart corresponds to the specific product carbon yields [%] and the substrate ratio of the respective elementary flux mode. The insert at the bottom right corner is a magnification of the area in the top left corner where the modes with the highest yields are located.

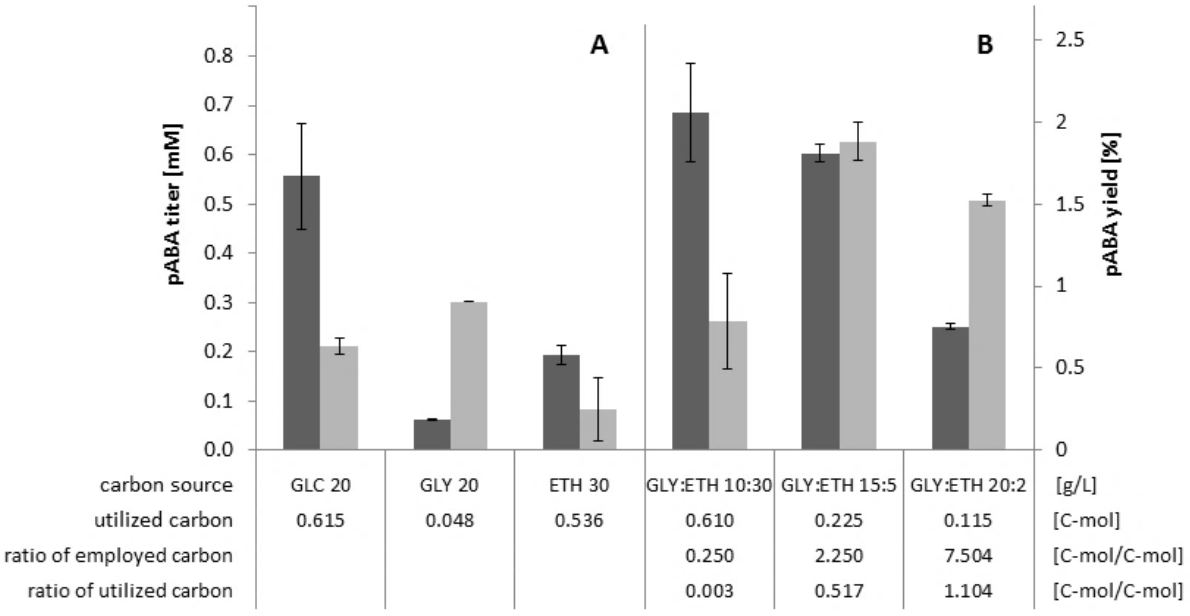
Metabolic network modelling revealed that stoichiometry of pABA production is improved when glycerol is the carbon-source. Increased Y_{max} (compared to glucose) can be explained with additional reducing power ($FADH_2$) gained during glycerol metabolisation, which can be used to regenerate ATP. The initial step of ethanol utilization also reduces an extra redox co-factor (NADH); however on ethanol alone very limited metabolic freedom exists, as all flux to upstream pathways (glycolysis, pentose phosphate pathway, shikimate pathway) has to be derived from acetyl-CoA via phosphoenolpyruvate carboxykinase and malic enzyme (cf. EMA networks in supplementary file 2.4.2 for reactions). In particular for the formation of the shikimate pathway precursor PEP from ethanol, two more mol ATP are required in respect to glucose and four more in respect to glycerol.

This is compensated on a combined GLY/ETH-feed: Separate carbon-flow to the upper and lower parts of central metabolism allows full advantage of higher availability of reduced redox equivalents, resulting in a higher biomass yield and more modes close to the stoichiometric limit (line between max. biomass and max. product yield, cf. supplementary file 2.4.2 for biomass vs. product yield plots). In addition, the majority of

modes with low product / biomass formation are abolished, as the two major by-products on glucose (GLY/ETH) are substrates now.

In order to reproduce the observed higher pABA production of glucose batches after the diauxic shift (when re-utilizing glycerol and/or ethanol) and to test the prediction of the elementary flux mode analysis, the impact of the carbon-sources glucose, glycerol, ethanol and glycerol+ethanol mixtures was tested *in vivo* using the PABA4 strain in shake-flask experiments. Comparing single carbon-sources the pABA titer was found to be highest on glucose, the carbon yield, however, was higher on glycerol (Figure 22A). The product carbon yield from ethanol can be noted as lowest, despite a high uncertainty due to high evaporation (cf. section 2.4.3.5). Concentrations of C-source where chosen aiming at comparable amounts of C-mol, e.g. when using 30 g/L ethanol even in case of < 50% substrate loss by evaporation an equivalent amount of C-mol as in the glucose experiment was available over the course of the fermentation.

In general, the highest pABA yields and also titers of all trialled carbon-sources were obtained when combining glycerol and ethanol. Different concentrations and ratios were tested to determine the ideal ratio of glycerol to ethanol (Figure 22B). It was found that initial GLY:ETH ratios > 3:1 g/g allowed for uptake ratios > 0.5 C-mol/C-mol, delivering the highest pABA yields. Enforcing an increase in uptake ratio only worked up to a certain degree: A 20:2 g/g GLY:ETH feed-ratio only allowed an uptake ratio of 1.1 C-mol/C-mol and while the pABA yield was only slightly reduced, the titer was more than halved.



carbon source	GLC 20	GLY 20	ETH 30	GLY:ETH 10:30	GLY:ETH 15:5	GLY:ETH 20:2	[g/L]
utilized carbon	0.615	0.048	0.536	0.610	0.225	0.115	[C-mol]
ratio of employed carbon				0.250	2.250	7.504	[C-mol/C-mol]
ratio of utilized carbon				0.003	0.517	1.104	[C-mol/C-mol]

Figure 22: pABA titers (dark grey) and yields (light grey) achieved by PABA4 on different carbon-sources (A) and GLY:ETH different ratios (B). Titers are evaporation corrected and represent the maximum after either carbon-source was used up and/or no further product formation occurred. Utilized carbon is the amount of C-mol from either carbon-source that was taken up and actually

metabolised. Yields and carbon usage ratios are based on the evaporation corrected titers while only respecting the actually metabolised fraction of the carbon-source.

The results indicate that both, GLY and ETH, were utilized and contributed to pABA production, however in different ways. Therefore, the contribution of the carbon from the two substrates to pABA formation on a combined GLY/ETH carbon-source was investigated in a tracer experiment with 1,2-^{13}C labelled ETH and natural GLY. Enrichment analysis of pABA revealed that in total up to 31% were derived from GLY (mean enrichment 69%). Further analysis of the labelling of amino acids revealed that glycerol was mainly metabolised in the upper part of glycolysis and pentose phosphate pathway, with little to no C^{12}-carbon reaching the TCA-cycle (mean enrichment of threonine, aspartate, and glutamate 92.2%, 92.9%, 93.9%, respectively). The activity of the threonine aldolase pathway for glycine biosynthesis could be seen by a glycine enrichment of 79.5%, while serine was only labelled 59.8% and clearly was also derived from glycerol. Since the alanine enrichment reached 85.2% it seems that glycerol did not reach the pyruvate pool to a large extent. Ethanol finally also reached the pentose phosphate pathway and subsequently the biosynthesis of histidine (44.1% enriched) and the aromatics. The fact that ethanol served as a precursor for pentose phosphate pathway derived compounds points towards a limitation in the supply of C_3 molecules from the glycerol uptake. During the tracer experiment the GLY:ETH uptake ratio was 1.1 C-mol/C-mol. It is striking that the overall glycerol usage is higher than what ends up in pABA (GLY:ETH ratio = 0.47 C-mol/C-mol). This allows the conclusion that glycerol contributes to pABA production not directly by the supply of carbon. Therefore, a benefit in energy metabolism is the only possible explanation, the inferences from EMA support this theory.

2.4.4.3 Carbon-sources composition controls aromatic product spectrum in vivo

To characterize pABA production on GLY/ETH in more detail, the metabolic profiles over the course of fermentations on GLC and GLY/ETH were compared (Figure 23): On GLC the maximum pABA titer was reached only after almost 130 h while a major part of pABA was produced during re-assimilation of ethanol / glycerol. On GLY/ETH the maximum pABA titer was reached within 78 h while the rate of pABA production was highest when ethanol was still available ($r_{0\text{-}46h}$ = 39.52 µmol/g×h vs. $r_{46\text{-}78h}$ = 7.3 µmol/g×h). Albeit pABA titer was only slightly higher on GLY/ETH than on GLC, productivity was significantly greater ($r_{GLY/ETH}$ = 26.7 µmol/g×h vs. r_{GLC} = 7.2 µmol/g×h).

Further, a dependency of the product spectrum on the carbon-source could be observed. The major by-product was PEA, which is derived from phenylpyruvate, either as an intermediate of PHE degradation or *de novo* from prephenate (formation of phenylpyruvate has been observed *in vivo* even though *ARO7* was knocked out (Winter *et al.*, 2014)) On glucose PEA titers even exceeded pABA titers. However on GLY/ETH formation of PEA was reduced (3-fold) in favour of increased selectivity for pABA. Further on GLY/ETH it was observed that PEA production early in the fermentation correlated directly with the decline of PHE (cf. supplementary file 2.4.3). Tracer experiments showed that growing the strain solely on U-^{13}C glucose led to no significant labelling of PEA (mean enrichment 3%) at the end of the exponential growth (where PHE was depleted), proving that *de novo* synthesis had not occurred. In stationary phase, however, PEA appeared to be partially ^{13}C enriched (mean enrichment 29%), indicating *de novo* synthesis from chorismate (cf. supplementary file 2.4.3 for details). No other aromatic by- or downstream products of the shikimate pathway were found, in particular, no accumulation of AAAs was observed. The tryptophan degradation product tryptophol was identified in GC-MS samples of the tracer experiments but appeared to be unlabelled and hence originated from tryptophan degradation only (not *de novo* synthesized). Further dihydrofolate and folate (which pABA is a precursor for) were also not found to be present in the media in detectable concentrations (LOD < 4 µM).

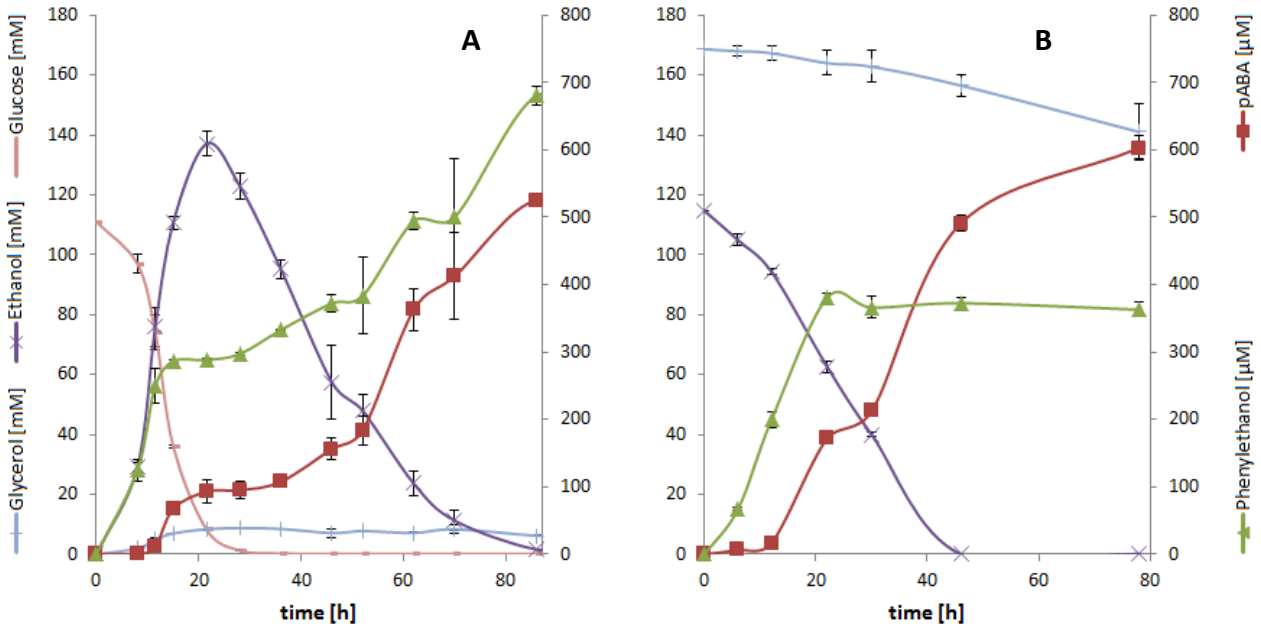


Figure 23: Time course of substrate consumption and product formation of main aromatic products on GLC (A) and GLY/ETH (B). Titers are corrected for evaporation of H_2O to reflect comparable metabolite concentrations.

The advantageous stoichiometry for pABA formation on GLY/ETH can explain the initial observation that a major fraction of pABA production on glucose only occurred late in the fermentation after the diauxic-shift: When glucose was consumed ethanol and glycerol were reassimilated, and the major amount of aromatics formed. The fact that production and growth only proceeded until ethanol was depleted indicates that ethanol was essential in order to sustain a viable metabolism of the cell. This, together with the clues from EMA and the tracer experiments, can explain the observed "sweet-spot" of the GLY:ETH ratio for pABA production: Ethanol concentration has to be limited in order to not dominate carbon supply for the cell (due to faster utilization kinetics) but apparently also needs to be high enough to support metabolism and allow sufficient glycerol utilization. The *in vivo* determined ideal glycerol to ethanol uptake-ratio of 0.5 C-mol/C-mol, which delivered the highest pABA yields, also correlates with the ideal ratio predicted by modelling (optimum between 0.5 and 2 C-mol/C-mol). Interestingly, further enforcing higher glycerol uptake affected pABA titer negatively and also slightly diminished the yield. This was most likely due to the impaired glycerol uptake mechanism of laboratory yeast (Swinnen *et al.*, 2013), presumably resulting in an imbalanced metabolism of the cell if ethanol share in the carbon-source is too low.

While ETH was always depleted in the course of the fermentation, of the initial amount of GLY only a minor fraction was utilized. Evaporation of ETH was a major reason for the decline of ethanol concentration during the fermentation, accounting for almost half thereof (depending on the ethanol concentration). Biomass yields on GLY/ETH were however at least equivalent to the ones GLC (data not shown). This also correlates with results from EMA where maximum biomass yield was highest on GLY/ETH (cf. supplementary file 2.4.2). Carbon usage may be improved by elevating glycerol uptake / utilization, therefore allowing for greater and quicker glycerol utilization, which would also aid in limiting ethanol evaporation.

2.4.4.4 Controlling by-product formation and maximizing pABA production

To maximize pABA production a fed-batch bioreactor experiment with a high starting cell density was conducted. Therefore a GLY:ETH ratio which delivered the highest pABA titers in shake-flask experiments (10:30 g/L) was chosen. Feed pulses were employed when growth slowed down (cf. section 2.4.3.4 for process details) to replenish depleted substrates. Figure 24A shows the development of substrates and products / biomass throughout the fermentation. In this experiment, the highest pABA (1728 ±10 µM ≈ 237 mg/L) and overall aromatics titers (cf. Table 7) were achieved. pABA share of total

aromatics, however, was only 52% (by titer in mM) due to substantial PEA formation (especially shortly after AAAs were added, therefore presumably derived from PHE degradation). Further small amounts of *para*-hydroxybenzoic acid (pHBA) (< 50 µM) were detected. Growth appeared to be limited by the availability of AAAs. Glycerol utilization was marginal (ratio of used GLY/ETH < 0.05 C-mol/C-mol), negatively impacting the carbon yield (Table 7). Stripping appeared to be a significant reason for ethanol reduction.

To limit stagnation of growth due to immediate degradation of abundant AAAs and to further improve pABA production and yield, an experiment with a continuous feed stream was designed. This also intended to explore if higher GLY utilization could be enforced through the limitation of ETH availability. In theory, this would compensate for the slower kinetics of GLY uptake and minimize substrate loss by evaporation. Utilization of a greater share of GLY while minimizing ETH usage was anticipated to benefit product yield (cf. section 2.4.4.2). The fed-batch bioreactor experiment was conducted with an ETH-feed in the order of the cells ETH uptake rate (cf. section 2.4.3.4 for process details). As can be seen from Figure 24B stripping of ETH was successfully minimized. Also, the PEA titer could be successfully reduced, in comparison to the previous bioreactor experiment (cf. Table 7), which benefited the fraction of pABA (> 83% by titer in mM) of total aromatics. Final pABA titer could not be improved, due to the continuously diluting feed stream, however, the highest product carbon yield (2.64%) in this study could be reached. The early formation of PEA largely corresponded to the decline of the initial PHE in the medium (data not shown). The steady supply of AAAs successfully maintained growth throughout the fermentation while no more PHE was degraded to PEA. Also, pHBA was detected again, reaching up to 100 µM. GLY utilization could be increased (ratio of used GLY/ETH ≈ 0.07 C-mol/C-mol), benefiting the pABA yield (Table 7). Interestingly the highest GLY utilization was observed towards the end of the fermentation when ethanol concentration peaked due to increased feed rate.

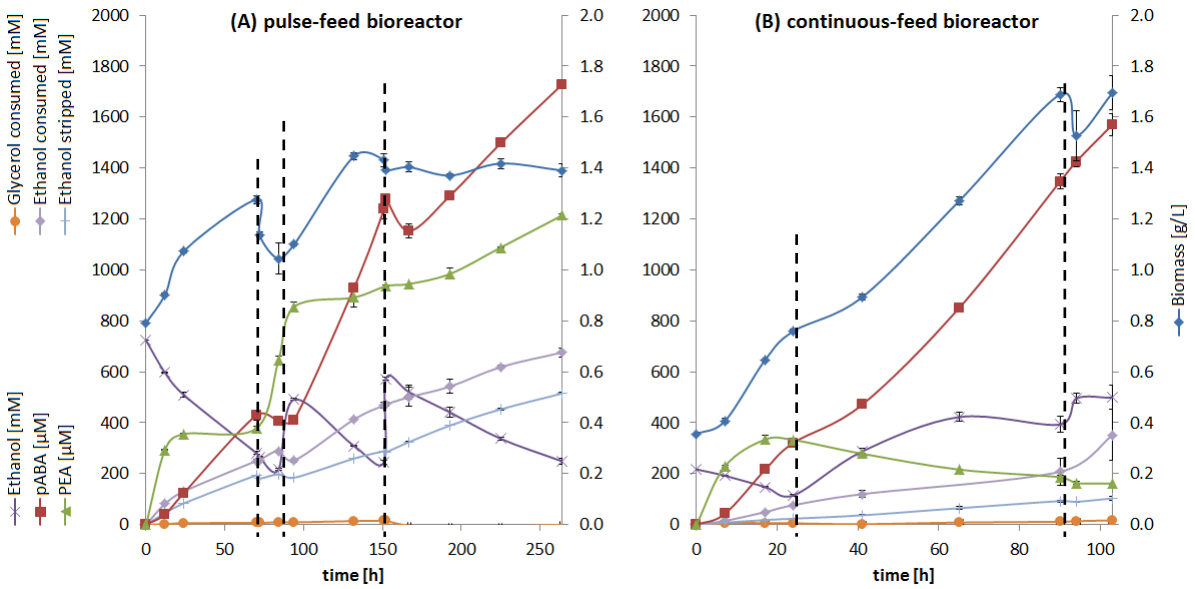


Figure 24: Substrate uptake and aromatic product formation over time in the bioreactor fermentation on GLY/ETH with (A) feed-pulse and (B) continuous-feed. Profiles of consumed substrates have been adjusted / corrected for dilution to reflect actual substrate uptake. Dashed vertical lines indicate a feed-pulse and activation / increase of the feed respectively.

Sharply increasing PEA titers after each spike with AAAs in the pulse-feed bioreactor compared to the continuous-feed bioreactor correlate with the results from the tracer experiments, where most PEA on GLY/ETH was derived from PHE degradation and little to no PEA was formed *de novo*. The same correlation can be seen in the initial hours of the continuous-feed bioreactor experiment. However additional PHE, contained in the gradually added feed-solution in this experiment (and thus added to the fermentation over time), was apparently not degraded to PEA. This also explains why growth in the continuous-feed bioreactor was largely sustained, while the pulse-feed bioreactor suffered from impaired growth, although sufficient carbon-source was available throughout the fermentation: High concentrations of AAAs were rapidly degraded into higher alcohols, thus depleting quickly. This highlights the advantage of the continuous-feed bioreactor. With improved growth also glycerol usage increased, especially towards the end of the continuous-feed bioreactor experiment where growth was fastest. This is in accordance with the theory that ethanol is needed in order to support glycerol utilization.

Table 7: Maximum titers and product carbon yields of key aromatic compounds and total aromatics obtained from batch bioreactor fermentations. Yields are based on dilution / evaporation corrected concentrations to account for actual substrate utilization.

Pulse-feed fermentation

Compound	pABA	PEA	Total[#]
Titer [mM]	1.73	1.21	3.31 / 1.75 *de novo*
[mg/L]	237	148	444
Yield [%]	1.52	0.4*	1.9*

Continuous-feed fermentation

Compound	pABA	PEA	Total[#]
Titer [mM]	1.57	0.16	1.9 / 1.67 *de novo*
[mg/L]	215	19.5	270
Yield [%]	2.64	0*	2.8*

[#] Total aromatics titer is comprised of pABA, PEA, pHBA, and tryptophol

* Carbon yields of PEA and total aromatics were estimated to reflect only the *de novo* synthesized fraction

In both bioreactor experiments, the final pABA titer was reduced because of dilution by the feed stream. When accounting for dilution of the medium due to the feed volumes in both experiments, it appeared that the theoretical final pABA production was higher for the continuous-feed experiment (1.57 mM × 2.12) than for the pulse-feed experiment (1.73 mM × 1.356). This influenced the final pABA yield (Table 7) and can be explained with the greater substrate specific yield coefficients (Table 8) of the experiment with the continuous-feed.

The greater pABA yield and product specific yield coefficients also show the advantage of the continuous-feed. Also, the productivity was significantly greater, with the final titer reached more than 2.56 times faster. However, the biomass specific yield coefficient for pABA was greater in the pulse-feed experiment in the second and third phase. This means that pABA production was still high, even though growth was slowing down, possibly hinting at a strategy that might in the future allow separation of production from growth.

Table 8: Yield coefficients for pABA in respect to biomass and the two substrates glycerol and ethanol in the bioreactor experiments. Values are based on dilution / evaporation corrected concentrations to account for actual substrate utilization.

	Pulse-feed fermentation			Continuous-feed fermentation
	phase 1	phase 2	phase 3	
$Y_{pABA/Biomass}$ [mmol/(g_{CDW})]	1	2.5	100	1
$Y_{pABA/GLY}$ [mmol/mol]	69	169.5	188.7	270.27
$Y_{pABA/ETH}$ [mmol/mol]	1.9	4.9	4.4	11.3

The results from the bioreactor experiments represent the upper limits of the currently best strain for pABA production, which are constrained by substrate utilization: If growth is limited (either by means of low ethanol concentrations or depletion of AAA) then glycerol is hardly utilized. If ethanol and AAAs are present in excess, more glycerol is utilized, but flux overflow and degradation of AAAs diverts carbon to PEA and results in loss of substrate due to evaporation. Balancing this represents the first major challenge. The defective glycerol utilization mechanism in *S. cerevisiae* laboratory strains, which diminishes productivity and increases ethanol demand, may be alleviated by over-expression of *STL1*, $GUT1_{CBS6412\text{-}13A}$ and *GUT2* as described before (Swinnen *et al.*, 2013). Improved glycerol uptake and flux into central metabolism could help to not only enhance growth but also production on glycerol. It may allow reduction of the ethanol share of the carbon-source, simultaneously reducing process costs and possibly also resulting in an improved yield while utilizing a more sustainable carbon-source.

The final challenge is to further optimize the formation of pABA by means of increasing the activity of the ABZ1p and ABZ2p in order to drain flux faster and more efficiently from shikimate pathway. Initial conclusions drawn from the screening of the different *ABZ* genes towards their sequence may help to prioritize efforts to further optimize the enzymes. These would then also reduce the risk of flux overflow towards PEA, which would become critical again once carbon-source utilization has been optimized.

2.4.4.5 Metabolic bottlenecks of carbon-source utilization and aromatics production

It was recently shown that ethanol stress reduced the production of higher alcohols of AAA, including PEA (Avbelj *et al.*, 2015). This was believed to happen via upregulation of the genes *ARO8*, *ARO9* and *ARO10*. However, these genes evoke in combination the formation of PEA via catabolism of AAAs. Thus, this can in the present study not fully explain the observed accumulation of PEA from glucose to concentrations greatly higher

than the amount of supplemented AAA. Therefore, another reason for reduced PEA production on ethanol is proposed in addition to the described effects:

It was recently demonstrated that the conversion of chorismate to prephenate / phenylpyruvate in a Δ*aro7* strain proceeds also spontaneously (Winter *et al.*, 2014), the reaction rate is merely a matter of substrate and product concentrations, temperature and pH. Given that pABA production is kinetically limited, in a situation of high flux into the shikimate pathway during faster glucose metabolism, chorismate could be accumulating, leading to an overflow into PEA (a similar flux overflow resulting in an accumulation of PHE and TYR has been reported when producing pABA in *E. coli* (Koma *et al.*, 2014)). During growth on GLY/ETH overall conversion rates are lower, thus, also flux arriving at chorismate is lower, so that the kinetics of pABA formation are not overwhelmed and chorismate does not overflow into PEA. Thus, the bottleneck of carbon-source utilization compensates the bottleneck towards pABA formation.

2.4.5 Conclusion

With respective titers of 1.9 mM and 1.73 mM for aromatic compounds in general and pABA, in particular, this study demonstrates the potential of *S. cerevisiae* as a microbial cell factory. By means of pABA production, it could be shown that the variability in different *ABZ1* and *ABZ2* sequences of sequenced yeast strains has a great impact on their functionality. Two genes could be identified that in combination allow a significant improve in pABA production. Further this study demonstrates the potential of metabolic engineering when rationally making use of metabolic modelling and genetic engineering: As predicted by modelling it could be shown that production of pABA from a combined glycerol-ethanol carbon-source was more efficient than from glucose. The benefits are twofold: On the one hand, stoichiometry allowed a higher yield (as predicted by modelling); on the other hand kinetic limitations (that otherwise result in flux overflow) were less influential resulting in increased selectivity for the desired compound pABA. This allowed a more than 6-fold increase in product titer and an almost 8-fold increase in yield over the previous benchmark (Krömer *et al.*, 2013). The reported results are the highest pABA titer and yield in *S. cerevisiae* to date and the second highest in a biological system.

The future challenge in order to further increase productivity, titer and yield is to direct flux more efficiently to pABA. In particular, further improvement of the *ABZ1* and *ABZ2* genes is essential in order to take pABA production to the next level. Enhancing glycerol uptake and metabolisation could then also help to increase the pABA yield (and possibly also titer) as well as reduce process costs.

3 General Conclusion and Future Perspectives

3.1 Main Conclusion

A metabolic modelling tool was developed that allowed holistic *in silico* analysis of the pathways for production of adipic acid and precursors for the hexanedioic acid in the cell envelope. Shikimate pathway based routes turned out to be amongst the most promising ones. In consequence existing strain construction strategies for muconic acid production from shikimate pathway were examined in more detail, identifying weaknesses and tailoring optimizations *in silico*.

Transferring the *in silico* determined strain construction strategy for coupling of growth to product formation to an *in vivo* system, production of pHBA was established using a novel synthetic circuit for dynamic regulation of gene expression. This verified *in vivo* validity of the *in silico* determined knock-out strategy.

Further, a chassis strain was constructed that effectively channelled carbon flux to the target products. This also resulted in a deeper understanding of the shikimate pathway, in particular, the chorismate (*ARO7*) mutase reaction was found to proceed also uncatalysed, due to the instability of chorismate.

Screening of different genes for production of pABA identified *ABZ1* and *ABZ2* genes that allowed significantly increased product titers. These originated from the wine yeasts AWRI1631 and QA23, respectively. In combination with *in silico* designed composition of carbon-source this improved the product carbon yield while minimizing by-products by avoiding loss of carbon due to the leaky *ARO7* knock-out.

3.2 Future Perspectives

The integrated picture of the different research outcomes defines how research efforts need to be focused in the future. This largely relates to the application of the gained insight and knowledge *in vivo*. The overarching goal is to achieve a significant increase in aromatics production, the major chance to trigger such a breakthrough is the application of the PYK knock-out. Current strains, however, are simply not efficient / effective enough to allow this (unpublished data). Therefore, in first instance efforts need to focus on enhancing the formation of the respective aromatic products. This is particularly important for the production of pABA, which is hindered by several factors.

Codon optimization of $ABZ1_{AWRI1631}$ & $ABZ2_{QA23}$ may help to increase expression of these, which was identified as a problem (unpublished data). The prokaryotic functional analogous *pabAB* & *pabC* are another set of enzymes which seem very promising (Koma

et al., 2014) for enhancing pABA formation. Finally, the design of a fusion protein (Chen *et al.*, 2013) of the aminodeoxychorismate synthase & aminodeoxychorismate lyase (e.g. *pabABC*) may channel conversion of 4-amino-4-deoxychorismate and thus streamline pABA production.

Further, with increasing reaction rates towards the formation of pABA, the supply of substrates for these reaction may also become limiting. Therefore strengthening glutamine regeneration, as shown to be beneficial in an analogous study (Sun *et al.*, 2013), may require overexpression of *GLN1*. Increasing chorismate levels is more complex and needs to be carefully balanced with downstream product formation due to the high instability of chorismate (Winter *et al.*, 2014). Nevertheless, overexpression of *aroL* has been shown to significantly increase flux through shikimate pathway (Rodriguez *et al.*, 2015). Preliminary results (unpublished data) show that this is also valid for the production of pHBA. To increase overall carbon flux into shikimate pathway, alternatively or additionally to $ARO4^{K229L}$ the mutant $ARO4^{Q166K}$ (Galanie *et al.*, 2015) could be employed, going further into protein engineering a potential double mutant $ARO4^{Q166K,K229L}$ may be worthwhile investigating.

A different approach, developed from modelling (unpublished data), is to use N-acetylglucosamine as combined nitrogen and carbon-source, as shown to be possible in *S. cerevisiae* (Wendland *et al.*, 2009). This could be a promising approach to enforce pABA production as pABA would provide a suitable sink for the surplus NH_3, thus introducing a minimum yield efficiency.

Knock-outs that were done to channel flux to target products appeared to be problematic (excessive amounts of AAAs were required) and imposed constraints for growth, which hamper process scale-up and titer maximization. Constitutive knockdowns and promoter attenuation (Blazeck *et al.*, 2012, Curran *et al.*, 2014) might alleviate these restrictions; there may, however, be more elegant solutions. For pABA production mostly the *TRP3* knock-out constrained growth while the *ARO7* knock-out is already more like a knockdown (Winter *et al.*, 2014). Nevertheless at least partial reconstruction of the strain(s) seems inevitable. This opens the opportunity to implement all acquired knowledge and tackle the problem more rationally from a different angle: The idea is to first develop an advanced yeast strain generally optimized for production of aromatics, by exploiting the fact that shikimate pathway can replace the pyruvate kinase function. This requires minimization of auxotrophies in order to maximize growth. Further, for the PYK knock-out to function, maximization of PYR production is necessary, in order to meet the demand of central metabolism; this may be achieved by combining pABA, oABA & pHBA production. New

strains would ideally include all optimizations of the previous ones while omitting identified weaknesses. In particular, strains to be constructed could include the following genetic modifications: Δ*aro7*, $ARO4^{K229L}$ or/and $ARO4^{Q166K}$, *ubiC* or $ubiC^{CCSS}$, *GLN1*, *aroL,* and $ABZ1_{AWRI1631}$-$ABZ2_{QA23}$ or *pabABC*.

In the next step PYK single knock-out strains (Δ*pyk2* or Δcdc19) with low pyruvate kinase activity may already achieve an increase in aromatics production, as well as gain tolerance to the mutation(s). Here it is extremely important to reduce auxotrophies, ideally limiting them only to PYR. For adaptive evolution of the strains in repeated batch fermentations (Zhang *et al.*, 2015) a combined carbon-source of GLC+GLY/ETH or GLC+PYR could be used, while reducing the GLY/ETH or PYR share in each step. In fact, in preliminary experiments a Δ*cdc19* mutant, derive from one of the aromatics producing strains, was reproducibly able to regain substantial growth on GLC over six cycles of repeated batch cultivation (unpublished data). This may finally allow construction of fully pyruvate kinase activity deficient strain(s), which achieve tight coupling of growth to product formation.

Evolutionary force may then unfold its full potential to improve not only fitness but actually the overall aromatics yield of the strain(s). Evolved strain(s) could then be characterized and sequenced to determine the full phenotype and genotype (possibly identifying new feedback inhibition resistant enzymes). These strain(s) may then serve as the basis for production of different aromatics, either by tailoring different production platforms using the determined mutations or by directly engineering the evolved strain(s) into a distinct direction. Promising target products may include anthranilic acid (oABA), muconic acid via 2,3-dihydroxybenzoic acid and vanillic acid via pHBA, but are not limited to these.

In case a fully pyruvate kinase deficient strain remains problematic, silencing of *CDC19* and/or *PYK2* along with constitutive knock-outs of either one may be worthwhile. For this use of a CRISPRi system (DiCarlo *et al.*, 2013, Horwitz *et al.*, 2015) may be the better choice than RNAi, as inhibiting transcription rather than translation allows tighter control of the system. CRISPRi also features a less complex synthetic circuit with fewer heterologous enzymes. Both advantages would significantly reduce the metabolic burden, which is important for the Δ*pyk* phenotype to focus resources.

Transferring the concept to another organism with a more robust secondary metabolism may be worthwhile in the long run, i.e. employing (filamentous) fungi that are particularly capable of producing aromatics and/or Crabtree negative (Wittmann *et al.*, 2002, Simone Maria Menegatti de Oliveira *et al.*, 2013, Kim *et al.*, 2014). Applicability of the PYK knock-out would, however, have to be proven again for each unique organism.

Bibliography / List of References

Agency UEP (2012) *Saccharomyces cerevisiae* Final Risk Assessment. Vol. 2015 p.^pp.

Alberty RA (2003) *Thermodynamics of Biochemical Reactions*. Wiley-VCH, Hoboken.

Alberty RA (2006) Appendix 2: Tables of Transformed Thermodynamic Properties. *Biochemical Thermodynamics,*p.^pp. 409-424. John Wiley & Sons, Inc.

Alini S, Basile F, Blasioli S, Rinaldi C & Vaccari A (2007) Development of new catalysts for N_2O-decomposition from adipic acid plant. *Applied Catalysis B: Environmental* **70**: 323–329.

Ambroset C, Petit M, Brion C*, et al.* (2011) Deciphering the Molecular Basis of Wine Yeast Fermentation Traits Using a Combined Genetic and Genomic Approach. *G3* **1**: 263-281.

Antonenkov VD & Hiltunen JK (2012) Transfer of metabolites across the peroxisomal membrane. *BBA-Mol Basis Dis* **1822**: 1374-1386.

Avbelj M, Zupan J, Kranjc L & Raspor P (2015) Quorum-Sensing Kinetics in Saccharomyces cerevisiae: A Symphony of ARO Genes and Aromatic Alcohols. *J Agr Food Chem* **63**: 8544-8550.

Averesch NJH & Krömer JO (2014) Tailoring strain construction strategies for muconic acid production in *S. cerevisiae* and *E. coli*. *Metab Eng Commun* **1**: 19-28.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL & Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**: 2006 0008.

Babu T, Yun EJ, Kim S, Kim DH, Liu KH, Kim SR & Kim KH (2015) Engineering Escherichia coli for the production of adipic acid through the reversed β-oxidation pathway. *Process Biochem*.

Baneyx F (1999) Recombinant protein expression in *Escherichia coli*. *Curr Opin Biotechnol* **10**: 411-421.

Bang S-G & Choi CY (1995) DO-stat fed-batch production of *cis,cis*-muconic acid from benzoic acid by *Pseudomonas putida* BM014. *Journal of Fermentation and Bioengineering* **79**: 381-383.

Bardwell L (2005) A walk-through of the yeast mating pheromone response pathway. *Peptides* **26**: 339-350.

Barker JL & Frost JW (2001) Microbial synthesis of *p*-hydroxybenzoic acid from glucose. *Biotechnol Bioeng* **76**: 376-390.

Bart JCJ & Cavallaro S (2015) Transiting from Adipic Acid to Bioadipic Acid. 1, Petroleum-Based Processes. *Industrial & Engineering Chemistry Research* **54**: 1-46.

Bart JCJ & Cavallaro S (2015) Transiting from Adipic Acid to Bioadipic Acid. Part II. Biosynthetic Pathways. *Industrial & Engineering Chemistry Research* **54**: 567-576.

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

Becker J, Lange A, Fabarius J & Wittmann C (2015) Top value platform chemicals: bio-based production of organic acids. *Curr Opin Biotech* **36**: 168-175.

Bentley WE, Mirjalili N, Andersen DC, Davis RH & Kompala DS (1990) Plasmid-encoded protein: The principal factor in the “metabolic burden” associated with recombinant bacteria. *Biotechnology and Bioengineering* **35**: 668-681.

Bisson LF & Fraenkel DG (1984) Expression of kinase-dependent glucose uptake in *Saccharomyces cerevisiae*. *J Bacteriol* **159**: 1013-1017.

Blazeck J, Garg R, Reed B & Alper HS (2012) Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. *Biotechnol Bioeng* **109**: 2884-2895.

Blount BA, Weenink T & Ellis T (2012) Construction of synthetic regulatory networks in yeast. *FEBS letters* **586**: 2112-2121.

Boghigian BA, Shi H, Lee K & Pfeifer BA (2010) Utilizing elementary mode analysis, pathway thermodynamics, and a genetic algorithm for metabolic flux determination and optimal metabolic network design. *BMC Syst Biol* **4**: 49.
Boles E, de Jong-Gubbels P & Pronk JT (1998) Identification and characterization of MAE1, the *Saccharomyces cerevisiae* structural gene encoding mitochondrial malic enzyme. *J Bacteriol* **180**: 2875-2882.
Boles E, Schulte F, Miosga T, Freidel K, Schluter E, Zimmermann FK, Hollenberg CP & Heinisch JJ (1997) Characterization of a glucose-repressed pyruvate kinase (Pyk2p) in *Saccharomyces cerevisiae* that is catalytically insensitive to fructose-1,6-bisphosphate. *J Bacteriol* **179**: 2987-2993.
Borneman AR, Forgan AH, Pretorius IS & Chambers PJ (2008) Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. *FEMS Yeast Res* **8**: 1185-1195.
Borneman AR, Desany BA, Riches D, Affourtit JP, Forgan AH, Pretorius IS, Egholm M & Chambers PJ (2011) Whole-Genome Comparison Reveals Novel Genetic Elements That Characterize the Genome of Industrial Strains of *Saccharomyces cerevisiae*. *PLoS Genet* **7**: e1001287.
Borodina I & Nielsen J (2014) Advances in metabolic engineering of yeast Saccharomyces cerevisiae for production of chemicals. *Biotechnol J* **9**: 609-620.
Botet J, Mateos L, Revuelta JL & Santos MA (2007) A chemogenomic screening of sulfanilamide-hypersensitive *Saccharomyces cerevisiae* mutants uncovers *ABZ2*, the gene encoding a fungal aminodeoxychorismate lyase. *Eukaryot Cell* **6**: 2102-2111.
Boussie TR, Dias EL, Fresco ZM & Murphy VJ (2010) Production of adipic acid and derivates from carbohydrate-containing materials. Vol. A1 p.^pp. Rennovia, Inc., US20100317822A1.
Boussie TR, Dias EL, Fresco ZM & Murphy VJ (2010) Production of adipic acid and derivates from carbohydrate-containing materials. Vol. US20100317822A1 p.^pp. Rennovia, Inc., US.
Branduardi P & Porro D (2012) Yeasts in Biotechnology. *Yeast,*p.^pp. 347-370. Wiley-VCH Verlag GmbH & Co. KGaA.
Braus GH (1991) Aromatic amino acid biosynthesis in the yeast *Saccharomyces cerevisiae*: a model system for the regulation of a eukaryotic biosynthetic pathway. *Microbiol Rev* **55**: 349-370.
Bricker DK, Taylor EB, Schell JC*, et al.* (2012) A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, Drosophila, and humans. *Science* **337**: 96-100.
Brochado AR & Patil KR (2013) Overexpression of O-methyltransferase leads to improved vanillin production in baker's yeast only when complemented with model-guided network engineering. *Biotechnol Bioeng* **110**: 656-659.
Burgard AP, Pharkya P & Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* **84**: 647-657.
Burgard AP, Pharkya P & Osterhout RE (2010) Microorganisms for the production of adipic acid and other compounds. Vol. US007799545B2 p.^pp. Genomatica, Inc., US.
Burgard AP, Pharkya P & Osterhout RE (2010) Microorganisms for the production of adipic acid and other compounds. Vol. US7799545B2 p.^pp. Genomatica, Inc., US.
Burgard AP, Pharkya P & Osterhout RE (2012) Microorganisms for the production of adipic acid and other compounds. Vol. US8088607B2 p.^pp. Genomatica, Inc., US.
Burgard AP, Pharkya P & Osterhout RE (2012) Microorganisms for the production of adipic acid and other compounds. Vol. US008088607B2 p.^pp. Genomatica, Inc., US.
Burk MJ, Osterhout RE & Sun J (2011) Semi-synthetic terephthalic acid via microorganisms that produce muconic acid. Vol. A1 p.^pp. US.
Caspi R, Altman T, Billington R*, et al.* (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* **42**: D459-471.

Castelli LA, Nguyen NP & Macreadie IG (2001) Sulfa drug screening in yeast: fifteen sulfa drugs compete with p-aminobenzoate in *Saccharomyces cerevisiae*. *FEMS Microbiol Lett* **199**: 181-184.
Cecchini G, Schroder I, Gunsalus RP & Maklashina E (2002) Succinate dehydrogenase and fumarate reductase from *Escherichia coli*. *Biochim Biophys Acta* **1553**: 140-157.
Centi G, Trifiro F, Perathoner S & Cavani F (2009) *Sustainable Industrial Processes*. Weinheim: Wiley-VCH, Avenel.
chemicals-technology.com (2012) Verdezyne Adipic Acid Pilot Plant, Carlsbad, California, United States of America. Vol. 2013 p.^pp. Net Resources International.
Chen N, Du J, Liu H & Xu Q (2009) Elementary mode analysis and metabolic flux analysis of L-glutamate biosynthesis by *Corynebacterium glutamicum*. *Annals of Microbiology* **59**: 317-322.
Chen X, Zaro JL & Shen W-C (2013) Fusion protein linkers: Property, design and functionality. *Adv Drug Deliver Rev* **65**: 1357-1369.
Chen Y & Nielsen J (2013) Advances in metabolic pathway and strain engineering paving the way for sustainable production of chemical building blocks. *Curr Opin Biotechnol*.
Chen Y & Nielsen J (2013) Advances in metabolic pathway and strain engineering paving the way for sustainable production of chemical building blocks. *Curr Opin Biotech* **24**: 965-972.
Chen Z, Liu H, Zhang J & Liu D (2010) Elementary mode analysis for the rational design of efficient succinate conversion from glycerol by *Escherichia coli*. *J Biomed Biotechnol* **2010**: 518743.
Cherry JM, Hong EL, Amundsen C*, et al.* (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* **40**: D700-705.
Choi WJ, Lee EY, Cho MH & Choi CY (1997) Enhanced production of *cis,cis*-muconate in a cell-recycle bioreactor. *Journal of Fermentation and Bioengineering* **84**: 70-76.
Choudhary S & Schmidt-Dannert C (2010) Applications of quorum sensing in biotechnology. *Applied Microbiology and Biotechnology* **86**: 1267-1279.
Clomburg JM & Gonzalez R (2013) Anaerobic fermentation of glycerol: a platform for renewable fuels and chemicals. *Trends Biotechnol* **31**: 20-28.
Clomburg JM, Blankschien MD, Vick JE, Chou A, Kim S & Gonzalez R (2015) Integrated engineering of β-oxidation reversal and ω-oxidation pathways for the synthesis of medium chain ω-functionalized carboxylic acids. *Metab Eng* **28**: 202-212.
Curran KA, Leavitt JM, Karim AS & Alper HS (2013) Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. *Metab Eng* **15**: 55-66.
Curran KA, Crook NC, Karim AS, Gupta A, Wagman AM & Alper HS (2014) Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat Commun* **5**.
da Silva GP, Mack M & Contiero J (2009) Glycerol: a promising and abundant carbon source for industrial microbiology. *Biotechnol Adv* **27**: 30-39.
Datta R & Henry M (2006) Lactic acid: recent advances in products, processes and technologies — a review. *J Chem Technol Biot* **81**: 1119-1129.
de Jong B, Siewers V & Nielsen J (2012) Systems biology of yeast: enabling technology for development of cell factories for production of advanced biofuels. *Curr Opin Biotechnol* **23**: 624-630.
Deng Y & Mao Y (2015) Production of adipic acid by the native-occurring pathway in *Thermobifida fusca* B6. *J Appl Microbiol* **119**: 1057-1063.
Deng Y, Ma L & Mao Y (2016) Biological production of adipic acid from renewable substrates: Current and future methods. *Biochem Eng J* **105, Part A**: 16-26.
DiCarlo JE, Norville JE, Mali P, Rios X, Aach J & Church GM (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res*.
Dietmair S, Timmins NE, Gray PP, Nielsen LK & Krömer JO (2010) Towards quantitative metabolomics of mammalian cells: Development of a metabolite extraction protocol. *Anal Biochem* **404**: 155-164.

Diez-Gonzalez F & Russell JB (1997) The ability of *Escherichia coli* O157:H7 to decrease its intracellular pH and resist the toxicity of acetic acid. *Microbiology* **143 (Pt 4)**: 1175-1180.
Diniz SC, Voss I & Steinbuchel A (2006) Optimization of cyanophycin production in recombinant strains of *Pseudomonas putida* and *Ralstonia eutropha* employing elementary mode analysis and statistical experimental design. *Biotechnol Bioeng* **93**: 698-717.
Draths KM & Frost JW (1994) Environmentally Compatible Synthesis of Adipic Acid from D-Glucose. *J Am Chem Soc* **116**: 399-400.
Drinnenberg IA, Weinberg DE, Xie KT, Mower JP, Wolfe KH, Fink GR & Bartel DP (2009) RNAi in Budding Yeast. *Science* **326**: 544-550.
Duboc P & von Stockar U (1998) Systematic errors in data evaluation due to ethanol stripping and water vaporization. *Biotechnol Bioeng* **58**: 428-439.
Duncan K, Edwards RM & Coggins JR (1987) The pentafunctional arom enzyme of *Saccharomyces cerevisiae* is a mosaic of monofunctional domains. *Biochem J* **246**: 375–386.
Edman JC, Goldstein AL & Erbe JG (1993) *Para*-aminobenzoate synthase gene of *Saccharomyces cerevisiae* encodes a bifunctional enzyme. *Yeast* **9**: 669-675.
Emmerling M, Dauner M, Ponti A, Fiaux J, Hochuli M, Szyperski T, Wuthrich K, Bailey JE & Sauer U (2002) Metabolic flux responses to pyruvate kinase knockout in *Escherichia coli*. *J Bacteriol* **184**: 152-164.
EPA (1994) Background Report AP-42, Section 6.2, Adipic Acid Production. p.^pp. 6. U.S. Environmental Protection Agency, OAQPS/TSD/EIB, Research Triangle Park.
Escalante A, Salinas Cervantes A, Gosset G & Bolivar F (2012) Current knowledge of the *Escherichia coli* phosphoenolpyruvate-carbohydrate phosphotransferase system: peculiarities of regulation and impact on growth and product formation. *Appl Microbiol Biotechnol* **94**: 1483-1494.
Flamholz A, Noor E, Bar-Even A & Milo R (2012) eQuilibrator--the biochemical thermodynamics calculator. *Nucleic Acids Res* **40**: D770-775.
Fleming RM, Thiele I & Nasheuer HP (2009) Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to *Escherichia coli*. *Biophys Chem* **145**: 47-56.
Fletcher E, Krivoruchko A & Nielsen J (2015) Industrial systems biology and its impact on synthetic biology of yeast cell factories. *Biotechnol Bioeng* n/a-n/a.
Frost JW & Draths KM (1996) Synthesis of adipic acid from Biomass-derived carbon sources. Vol. US005487987A p.^pp. Purdue Research Foundation, US.
Frost JW & Draths KM (1997) Bacterial cell transformation for production of *cis,cis*-muconic acid and catechol. Vol. US005616496A p.^pp. Purdue Research Foundation, US.
Fukuda K, Asano K, Ouchi K & Takasawa S (1992) Feedback-insensitive mutation of 3-deoxy-d-arabino-hepturosonate-7-phosphate synthase caused by a single nucleotide substitution of *ARO4* structural gene in *Saccharomyces cerevisiae*. *J Ferment Bioeng* **74**: 117-119.
Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M & Smolke CD (2015) Complete biosynthesis of opioids in yeast. *Science* **349**: 1095-1100.
Gerstl MP, Jungreuthmayer C & Zanghellini J (2015) tEFMA: computing thermodynamically feasible elementary flux modes in metabolic networks. *Bioinformatics*.
Gerstl MP, Ruckerbauer DE, Mattanovich D, Jungreuthmayer C & Zanghellini J (2015) Metabolomics integrated elementary flux mode analysis in large metabolic networks. *Scientific Reports* **5**: 8930.
Giaever G, Chu A, Ni L*, et al.* (2002) Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**: 387-391.
Gietz RD & Woods RA (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* **350**: 87-96.
Gosset G, Yong-Xiao J & Berry A (1996) A direct comparison of approaches for increasing carbon flow to aromatic biosynthesis in *Escherichia coli*. *J Ind Microbiol* **17**: 47-52.

Greenbaum D, Colangelo C, Williams K & Gerstein M (2003) Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* **4**: 117.
Gruchattka E & Kayser O (2015) *In Vivo* Validation of *In Silico* Predicted Metabolic Engineering Strategies in Yeast: Disruption of α-Ketoglutarate Dehydrogenase and Expression of ATP-Citrate Lyase for Terpenoid Production. *PLoS ONE* **10**: e0144981.
Gruchattka E, Hadicke O, Klamt S, Schutz V & Kayser O (2013) *In silico* profiling of *Escherichia coli* and *Saccharomyces cerevisiae* as terpenoid factories. *Microb Cell Fact* **12**: 84.
Güldener U, Heck S, Fielder T, Beinhauer J & Hegemann JH (1996) A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res* **24**: 2519-2524.
Güldener U, Heinisch J, Koehler GJ, Voss D & Hegemann JH (2002) A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acids Res* **30**: e23.
Guldfeldt LU & Arneborg N (1998) Measurement of the effects of acetic acid and extracellular pH on intracellular pH of nonfermenting, individual *Saccharomyces cerevisiae* cells by fluorescence microscopy. *Appl Environ Microbiol* **64**: 530-534.
Guzman Dd (2010) Bio-adipic acid prepares for entry. *ICIS Chemical Business* **278**: 22.
Hadicke O & Klamt S (2011) Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metab Eng* **13**: 204-213.
Hamilton JA (1998) Fatty acid transport: difficult or easy? *J Lipid Res* **39**: 467-481.
Hansen EH, Moller BL, Kock GR, Bunner CM, Kristensen C, Jensen OR, Okkels FT, Olsen CE, Motawia MS & Hansen J (2009) *De novo* biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). *Appl Environ Microbiol* **75**: 2765-2774.
Haraldsdóttir HS, Thiele I & Fleming RMT (2012) Quantitative Assignment of Reaction Directionality in a Multicompartmental Human Metabolic Reconstruction. *Biophys J* **102**: 1703-1711.
Hardie DG & Pan DA (2002) Regulation of fatty acid synthesis and oxidation by the AMP-activated protein kinase. *Biochem Soc Trans* **30**: 1064-1070.
Hegemann JH & Heick SB (2011) Delete and repeat: a comprehensive toolkit for sequential gene knockout in the budding yeast *Saccharomyces cerevisiae*. *Methods Mol Biol* **765**: 189-206.
Heinz (2011) Driving Sustainable Growth. p.^pp.
Henry CS, Broadbelt LJ & Hatzimanikatis V (2007) Thermodynamics-based metabolic flux analysis. *Biophys J* **92**: 1792-1805.
Hernandez-Montalvo V, Martinez A, Hernandez-Chavez G, Bolivar F, Valle F & Gosset G (2003) Expression of galP and glk in a *Escherichia coli* PTS mutant restores glucose transport and increases glycolytic flux to fermentation products. *Biotechnol Bioeng* **83**: 687-694.
Herzig S, Raemy E, Montessuit S, Veuthey JL, Zamboni N, Westermann B, Kunji ER & Martinou JC (2012) Identification and functional expression of the mitochondrial pyruvate carrier. *Science* **337**: 93-96.
Hiltunen JK, Mursula AM, Rottensteiner H, Wierenga RK, Kastaniotis AJ & Gurvitz A (2003) The biochemistry of peroxisomal β-oxidation in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol Rev* **27**: 35-64.
Horwitz Andrew A, Walter Jessica M, Schubert Max G*, et al.* (2015) Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. *Cell Systems* **1**: 88-96.
Hubmann G, Guillouet S & Nevoigt E (2011) Gpd1 and Gpd2 fine-tuning for sustainable reduction of glycerol formation in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **77**: 5857-5867.
IHS Chemical (2012) Bio-Based Adipic Acid. Vol. 2013 p.^pp. IHS Chemical.
IHS Chemical (2012) Chemical Economics Handbook: Adipic Acid. Vol. 2013 p.^pp. IHS Chemical.

Imai T & Ohno T (1995) The relationship between viability and intracellular pH in the yeast *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **61**: 3604-3608.
Ingolia NT & Murray AW (2007) Positive-Feedback Loops as a Flexible Biological Module. *Current Biology* **17**: 668-677.
J. Sheehan R (2000) Terephthalic Acid, Dimethyl Terephthalate, and Isophthalic Acid. *Ullmann's Encyclopedia of Industrial Chemistry,* p.^pp. Wiley-VCH Verlag GmbH & Co. KGaA.
Jang YS, Kim B, Shin JH, Choi YJ, Choi S, Song CW, Lee J, Park HG & Lee SY (2012) Bio-based production of C2-C6 platform chemicals. *Biotechnol Bioeng* **109**: 2437-2459.
Jankowski MD, Henry CS, Broadbelt LJ & Hatzimanikatis V (2008) Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys J* **95**: 1487-1499.
Jendresen CB, Stahlhut SG, Li M, Gaspar P, Siedler S, Förster J, Maury J, Borodina I & Nielsen AT (2015) Novel highly active and specific tyrosine ammonia-lyases from diverse origins enable enhanced production of aromatic compounds in bacteria and yeast. *Appl Environ Microb*.
Johnson BF, Amaratunga M & Lobos JH (1998) Method for increasing total production of 4-hydroxybenzoic acid by biofermentation. Vol. A p.^pp. General Electric Company, US.
Jol SJ, Kummel A, Terzer M, Stelling J & Heinemann M (2012) System-level insights into yeast metabolism by thermodynamic analysis of elementary flux modes. *PLoS Comput Biol* **8**: e1002415.
Kanehisa M & Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**: 27-30.
Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M & Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* **42**: D199-205.
Kayser O & Averesch N (2015) *Technische Biochemie: Die Biochemie und industrielle Nutzung von Naturstoffen*. Springer-Verlag.
Keasling JD (2008) Synthetic Biology for Synthetic Chemistry. *ACS Chem Biol* **3**: 64-76.
Keseler IM, Mackie A, Peralta-Gil M*, et al.* (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res* **41**: D605-612.
Kim I-K, Roldão A, Siewers V & Nielsen J (2012) A systems-level approach for metabolic engineering of yeast cell factories. *FEMS Yeast Res* **12**: 228-248.
Kim T-Y, Lee S-W & Oh M-K (2014) Biosynthesis of 2-phenylethanol from glucose with genetically engineered *Kluyveromyces marxianus*. *Enzyme Microb Technol* **61–62**: 44-47.
Kluczyk A, Popek T, Kiyota T, de Macedo P, Stefanowicz P, Lazar C & Konishi Y (2002) Drug evolution: *p*-aminobenzoic acid as a building block. *Curr Med Chem* **9**: 1871-1892.
Köhle A, Sommer S, Li S-M, Schilde-Rentschler L, Ninnemann H & Heide L (2003) Secondary metabolites in transgenic tobacco and potato: high accumulation of 4-hydroxybenzoic acid glucosides results from high expression of the bacterial gene ubiC. *Mol Breeding* **11**: 15-24.
Koma D, Yamanaka H, Moriyoshi K, Sakai K, Masuda T, Sato Y, Toida K & Ohmoto T (2014) Production of P-aminobenzoic acid by metabolically engineered *Escherichia coli*. *Biosci Biotechnol Biochem* **78**: 350-357.
Krömer JO, Wittmann C, Schroder H & Heinzle E (2006) Metabolic pathway analysis for rational design of *L*-methionine production by *Escherichia coli* and *Corynebacterium glutamicum*. *Metab Eng* **8**: 353-369.
Krömer JO, Nunez-Bernala D, Averesch NJH, Hampe J, Varela J & Varela C (2013) Production of aromatics in *Saccharomyces cerevisiae*—A feasibility study. *J Biotechnol* **163**: 184-193.
Kummel A, Panke S & Heinemann M (2006) Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol Syst Biol* **2**: 2006 0034.
Kümmel A, Panke S & Heinemann M (2006) Systematic assignment of thermodynamic constraints in metabolic network models. *BMC Bioinformatics* **7**: 512.
Kunzler M, Paravicini G, Egli CM, Irniger S & Braus GH (1992) Cloning, primary structure and regulation of the *ARO4* gene, encoding the tyrosine-inhibited 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Saccharomyces cerevisiae*. *Gene* **113**: 67-74.

Lee SJ, Lee DY, Kim TY, Kim BH, Lee J & Lee SY (2005) Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and in silico gene knockout simulation. *Appl Environ Microbiol* **71**: 7880-7887.
Lee SK, Chou H, Ham TS, Lee TS & Keasling JD (2008) Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current Opinion in Biotechnology* **19**: 556-563.
Lewis NE, Nagarajan H & Palsson BO (2012) Constraining the metabolic genotype–phenotype relationship using a phylogeny of *in silico* methods. *Nat Rev Micro* **10**: 291-305.
Li M & Borodina I (2014) Application of synthetic biology for production of chemicals in yeast *Saccharomyces cerevisiae*. *FEMS Yeast Res*.
Li S, Huang D, Li Y, Wen J & Jia X (2012) Rational improvement of the engineered isobutanol-producing Bacillus subtilis by elementary mode analysis. *Microbial Cell Factories* **11**: 1-12.
Lin Y, Shen X, Yuan Q & Yan Y (2013) Microbial biosynthesis of the anticoagulant precursor 4-hydroxycoumarin. *Nat Commun* **4**.
Lin Y, Sun X, Yuan Q & Yan Y (2014) Extending shikimate pathway for the production of muconic acid and its precursor salicylic acid in *Escherichia coli*. *Metab Eng* **23**: 62-69.
Lingens F, Goebel W & Uesseler H (1967) Regulation der Biosynthese der aromatischen Aminosauren in *Saccharomyces cerevisiae*. *Europ J Bioch* **1**: 363–374.
Luttik MA, Vuralhan Z, Suir E, Braus GH, Pronk JT & Daran JM (2008) Alleviation of feedback inhibition in *Saccharomyces cerevisiae* aromatic amino acid biosynthesis: quantification of metabolic impact. *Metab Eng* **10**: 141-153.
Mainhardt H & Kruger D (2000) N_2O Emissions from Adipic acid and Nitric acid Production. p.^pp.
Mannhaupt G, Stucka R, Pilz U, Schwarzlose C & Feldmann H (1989) Characterization of the prephenate dehydrogenase-encoding gene, TYR1, from *Saccharomyces cerevisiae*. *Gene* **85**: 303-311.
Marbois B, Xie LX, Choi S, Hirano K, Hyman K & Clarke CF (2010) para-Aminobenzoic acid is a precursor in coenzyme Q6 biosynthesis in Saccharomyces cerevisiae. *J Biol Chem* **285**: 27827-27838.
March JC & Bentley WE (2004) Quorum sensing and bacterial cross-talk in biotechnology. *Current Opinion in Biotechnology* **15**: 495-502.
Maris AJAv, Konings WN, Dijken JPv & Pronk JT (2004) Microbial export of lactic and 3-hydroxypropanoic acid: implications for industrial fermentation processes. *Metab Eng* **6**: 245-255.
Markets Ra (2014) Aramid Fibers (Para and Meta) - A Global Market Overview Vol. 2015 p.^pp.
Martínez VS & Nielsen LK (2014) NExT: Integration of Thermodynamic Constraints and Metabolomics Data into a Metabolic Network. *Metabolic Flux Analysis,* Vol. 1191 (Krömer JO, Nielsen LK & Blank LM, eds.), p.^pp. 65-78. Springer New York.
Martínez Verónica S, Quek L-E & Nielsen Lars K (2014) Network Thermodynamic Curation of Human and Yeast Genome-Scale Metabolic Models. *Biophys J* **107**: 493-503.
Mattanovich D, Sauer M & Gasser B (2014) Yeast biotechnology: teaching the old dog new tricks. *Microb Cell Fact* **13**: 34.
Mavrovouniotis ML (1993) Identification of localized and distributed bottlenecks in metabolic pathways. *Proc Int Conf Intell Syst Mol Biol* **1**: 275-283.
MCGroup (2011) Adipic acid 2011 World Market Outlook and Forecast. Vol. 2013 p.^pp. Merchant Research & Consulting ltd.
McQualter RB, Chong BF, Meyer K, Van Dyk DE, O'Shea MG, Walton NJ, Viitanen PV & Brumbley SM (2005) Initial evaluation of sugarcane as a production platform for *p*-hydroxybenzoic acid. *Plant Biotechnol J* **3**: 29-41.
Meijnen J-P, Verhoef S, Briedjlal AA, Winde JH & Ruijssenaars HJ (2011) Improved *p*-hydroxybenzoate production by engineered *Pseudomonas putida* S12 by using a mixed-substrate feeding strategy. *Appl Microbiol Biot* **90**: 885-893.

Meuris P (1973) Feedback inhibition of the DAHP synthetases by tRNA in *Saccharomyces cerevisiae*. *Molec gen Genet* **121**: 207-218.
Meza E, Becker J, Bolivar F, Gosset G & Wittmann C (2012) Consequences of phosphoenolpyruvate:sugar phosphotranferase system and pyruvate kinase isozymes inactivation in central carbon metabolism flux distribution in *Escherichia coli*. *Microb Cell Fact* **11**: 127.
Millard P, Letisse F, Sokol S & Portais J-C (2012) IsoCor: correcting MS data in isotope labeling experiments. *Bioinformatics* **28**: 1294-1296.
Moon TS, Dueber JE, Shiue E & Prather KL (2010) Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*. *Metab Eng* **12**: 298-305.
Mortimer RK & Johnston JR (1986) Genealogy of principal strains of the yeast genetic stock center. *Genetics* **113**: 35-43.
Müller R, Wagener A, Schmidt K & Leistner E (1995) Microbial production of specifically ring-13C-labelled 4-hydroxybenzoic acid. *Appl Microb Biotech* **43**: 985-988.
Musser MT (2000) Adipic Acid. *Ullmann's Encyclopedia of Industrial Chemistry,*p.^pp. Wiley-VCH Verlag GmbH & Co. KGaA.
Musser MT (2005) Adipic Acid. *ULLMANN'S Encyclopedia of Industrial Chemistry,*(Wiley-VCH, ed.) p.^pp. Wiley-VCH, Weinheim.
Nichols BP & Green JM (1992) Cloning and sequencing of *Escherichia coli* ubiC and purification of chorismate lyase. *J Bacteriol* **174**: 5309-5316.
Niu W, Draths KM & Frost JW (2002) Benzene-free synthesis of adipic acid. *Biotechnol Prog* **18**: 201-211.
Noor E, Haraldsdóttir HS, Milo R & Fleming RMT (2013) Consistent Estimation of Gibbs Energy Using Component Contributions. *PLoS Comput Biol* **9**: e1003098.
Orij R, Postmus J, Ter Beek A, Brul S & Smits GJ (2009) In vivo measurement of cytosolic and mitochondrial pH using a pH-sensitive GFP derivative in *Saccharomyces cerevisiae* reveals a relation between intracellular pH and growth. *Microbiology* **155**: 268-278.
Osterhout RE, Burgard AP & Burk MJ (2013) Microorganisms and processes for producing terephthalic acid and its salts. Vol. A3 p.^pp. US.
Paddon CJ & Keasling JD (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat Rev Micro* **12**: 355-367.
Paddon CJ, Westfall PJ, Pitera DJ*, et al.* (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**: 528-532.
Papagianni M (2012) Recent advances in engineering the central carbon metabolism of industrially important bacteria. *Microb Cell Fact* **11**: 50.
Partow S, Siewers V, Bjorn S, Nielsen J & Maury J (2010) Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast* **27**: 955-964.
Patring J, Hjortmo S, Jastrebova J, Svensson U, Andlid T & Jägerstad IM (2006) Characterization and quantification of folates produced by yeast strains isolated from kefir granules. *Eur Food Res Technol* **223**: 633-637.
Picataggio S & Beardslee T (2012) Biological methods for preparing adipic acid. Vol. A1 p.^pp. Verdezyne, Inc., US.
Picataggio S & Beardslee T (2012) Biological methods for preparing adipic acid. Vol. US20120021474A1 p.^pp. Verdezyne, Inc., US.
Pierrel F, Hamelin O, Douki T, Kieffer-Jaquinod S, Mühlenhoff U, Ozeir M, Lill R & Fontecave M (2010) Involvement of Mitochondrial Ferredoxin and Para-Aminobenzoic Acid in Yeast Coenzyme Q Biosynthesis. *Chem Biol* **17**: 449-459.
Pira S (2014) PET Packaging Industry News. Vol. 2015 p.^pp.
Polen T, Spelberg M & Bott M (2013) Toward biotechnological production of adipic acid and precursors from biorenewables. *J Biotechnol* **167**: 75-84.

Prantl F, Strasser A, Aebi M, Furter R, Niederberger P, Kirschner K & Huetter R (1985) Purification and characterization of the indole-3-glycerolphosphate synthase/anthranilate synthase complex of *Saccharomyces cerevisiae*. *Eur J Biochem* **146**: 95-100.
Pugh S, McKenna R, Osman M, Thompson B & Nielsen DR (2014) Rational engineering of a novel pathway for producing the aromatic compounds p-hydroxybenzoate, protocatechuate, and catechol in *Escherichia coli*. *Process Biochem*.
Raemakers-Franken PC, Schürmann M, Trefzer AC & De Wildeman SMA (2012) Preparation of adipic acid. Vol. US20120028320A1 p.^pp. DSM IP Assets B.V., US.
Reizman IMB, Stenger AR, Reisch CR, Gupta A, Connors NC & Prather KLJ (2015) Improvement of glucaric acid production in *E. coli* via dynamic control of metabolic fluxes. *Metab Eng Commun* **2**: 109-116.
Research GV (2014) Global Bio-based Polyethylene Terephthalate (PET) Market. Vol. 2015 p.^pp.
Rocha I, Maia P, Evangelista P, Vilaca P, Soares S, Pinto JP, Nielsen J, Patil KR, Ferreira EC & Rocha M (2010) OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Syst Biol* **4**: 45.
Rodriguez A, Kildegaard KR, Li M, Borodina I & Nielsen J (2015) Establishment of a yeast platform strain for production of p-coumaric acid through metabolic engineering of aromatic amino acid biosynthesis. *Metab Eng* **31**: 181-188.
Rottensteiner H & Theodoulou FL (2006) The ins and outs of peroxisomes: Co-ordination of membrane transport and peroxisomal metabolism. *BBA-Mol Cell Res* **1763**: 1527-1540.
Ruckerbauer DE, Jungreuthmayer C & Zanghellini J (2014) Design of optimally constructed metabolic networks of minimal functionality. *PLoS One* **9**: e92583.
Sato K, Aoki M & Noyori R (1998) A "Green" route to adipic acid: direct oxidation of cyclohexenes with 30 percent hydrogen peroxide. *Science* **281**: 1646-1647.
Sauer U & Eikmanns B (2005) The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol Rev* **29**: 765.
Schmidheini T, Mosch HU, Graf R & Braus GH (1990) A GCN4 protein recognition element is not sufficient for GCN4-dependent regulation of transcription in the ARO7 promoter of *Saccharomyces cerevisiae*. *Mol Gen Genet* **224**: 57-64.
Schmidt E & Knackmuss H-J (1984) Production of *cis,cis*-muconate from benzoate and 2-fluoro-*cis,cis*-muconate from 3-fluorobenzoate by 3-chlorobenzoate degrading bacteria. *Appl Microb Biotech* **20**: 351-355.
Schuster S & Hilgetag C (1994) On Elementary Flux Modes in Biochemical Reaction Systems at Steady State. *J Biol Systems* **2**: 165-182.
Sengupta S, Jonnalagadda S, Goonewardena L & Juturu V (2015) Metabolic Engineering of a Novel Muconic Acid Biosynthesis Pathway via 4-Hydroxybenzoic Acid in *Escherichia coli*. *Appl Environ Microb* **81**: 8037-8043.
Sharma V, Sakai Y, Smythe KA & Yokobayashi Y (2013) Knockdown of recA gene expression by artificial small RNAs in Escherichia coli. *Biochem Biophys Res Commun* **430**: 256-259.
Siebert M, Severin K & Heide L (1994) Formation of 4-hydroxybenzoate in Escherichia coli: characterization of the ubiC gene and its encoded enzyme chorismate pyruvate-lyase. *Microbiology* **140**: 897-904.
Siebert M, Sommer S, Li S-m, Wang Z-x, Severin K & Heide L (1996) Accumulation of 4-Hydroxybenzoate Glucosides as a Result of the Expression of the Bacterial *ubiC* Gene in Tobacco. *Plant Physiol* **112**: 811–819.
Simone Maria Menegatti de Oliveira, Simone Damasceno Gomes, Luciane Sene, Silvia R. Machado Coelho, Ana Cláudia Barana, Marney Pascoli Cereda, Divair Christ & Piechontcoski J (2013) Production of 2-phenylethanol by *Geotrichum fragrans*, *Saccharomyces cerevisiae* and *Kluyveromyces marxianus* in cassava wastewater. *JFAE* **11**: 158-163.
Song H, Payne S, Tan C & You L (2011) Programming microbial population dynamics by engineered cell–cell communication. *Biotechnology Journal* **6**: 837-849.

Sun X, Lin Y, Yuan Q & Yan Y (2014) Biological production of muconic acid via a prokaryotic 2,3-dihydroxybenzoic acid decarboxylase. *ChemSusChem* **7**: 2478-2481.
Sun X, Lin Y, Huang Q, Yuan Q & Yan Y (2013) A Novel Muconic Acid Biosynthetic Approach by Shunting Tryptophan Biosynthesis via Anthranilate. *Appl Environ Microbiol* **79**: 4024-4030.
Swinnen S, Klein M, Carrillo M, McInnes J, Nguyen HT & Nevoigt E (2013) Re-evaluation of glycerol utilization in *Saccharomyces cerevisiae*: characterization of an isolate that grows on glycerol without supporting supplements. *Biotechnol Biofuels* **6**: 157.
Sybesma W, Starrenburg M, Tijsseling L, Hoefnagel MH & Hugenholtz J (2003) Effects of cultivation conditions on folate production by lactic acid bacteria. *Appl Environ Microbiol* **69**: 4542-4548.
Tepper N & Shlomi T (2010) Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics* **26**: 536-543.
Terzer M & Stelling J (2008) Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics* **24**: 2229-2235.
Thakker C, Martinez I, San KY & Bennett GN (2012) Succinate production in *Escherichia coli*. *Biotechnol J* **7**: 213-224.
Trinh CT, Unrean P & Srienc F (2008) Minimal *Escherichia coli* Cell for the Most Efficient Production of Ethanol from Hexoses and Pentoses. *Appl Environ Microb* **74**: 3634-3643.
Tsao C-Y, Hooshangi S, Wu H-C, Valdes JJ & Bentley WE (2010) Autonomous induction of recombinant proteins by minimally rewiring native quorum sensing regulon of *E. coli*. *Metab Eng* **12**: 291-297.
Van Duuren JBJH & Wittmann C (2014) First and Second Generation Production of Bio-Adipic Acid. *Bioprocessing of Renewable Resources to Commodity Bioproducts,* p.^pp. 519-540. John Wiley & Sons, Inc.
Van Dyk TK, Templeton LJ, Cantera KA, Sharpe PL & Sariaslani FS (2004) Characterization of the *Escherichia coli* AaeAB efflux pump: a metabolic relief valve? *J Bacteriol* **186**: 7196-7204.
van Klinken JB & Willems van Dijk K (2015) FluxModeCalculator: an efficient tool for large-scale flux mode computation. *Bioinformatics*.
van Roermund CWT, de Jong M, IJlst L, van Marle J, Dansen TB, Wanders RJA & Waterham HR (2004) The peroxisomal lumen in *Saccharomyces cerevisiae* is alkaline. *J Cell Sci* **117**: 4231-4237.
Vardon DR, Franden MA, Johnson CW, Karp EM, Guarnieri MT, Linger JG, Salm MJ, Strathmann TJ & Beckham GT (2015) Adipic acid production from lignin. *Energ Environ Sci* **8**: 617-628.
Veenhuis M, Salomons FA & Van Der Klei IJ (2000) Peroxisome biogenesis and degradation in yeast: A structure/function analysis. *Microsc Res Techniq* **51**: 584-600.
Verhoef S, Ballerstedt H, Volkers RJM, Winde JH & Ruijssenaars HJ (2010) Comparative transcriptomics and proteomics of *p*-hydroxybenzoate producing *Pseudomonas putida* S12: novel responses and implications for strain improvement. *Appl Microbiol Biot* **87**: 679-690.
Vickers CE, Klein-Marcuschamer D & Krömer JO (2012) Examining the feasibility of bulk commodity production in *Escherichia coli*. *Biotechnol Lett* **34**: 585-596.
Weber C, Bruckner C, Weinreb S, Lehr C, Essl C & Boles E (2012) Biosynthesis of *cis,cis*-muconic acid and its aromatic precursors, catechol and protocatechuic acid, from renewable feedstocks by *Saccharomyces cerevisiae*. *Appl Environ Microbiol* **78**: 8421-8430.
Wendland J, Schaub Y & Walther A (2009) N-Acetylglucosamine Utilization by Saccharomyces cerevisiae Based on Expression of Candida albicans NAG Genes. *Appl Environ Microb* **75**: 5840-5845.
White S (2008) Gene cloning principles and applications by Julia Lodge, Pete Lund, and Steve Minchin. *Biochem Mol Biol Edu* **36**: 170-170.
Williams TC, Nielsen LK & Vickers CE (2013) Engineered Quorum Sensing Using Pheromone-Mediated Cell-to-Cell Communication in *Saccharomyces cerevisiae*. *ACS Synth Biol* **2**: 136-149.

Williams TC, Averesch NJ, Winter G, Plan MR, Vickers CE, Nielsen LK & Krömer JO (2015) Quorum-sensing linked RNA interference for dynamic metabolic pathway control in *Saccharomyces cerevisiae*. *Metab Eng* **29**: 124-134.
Winter G, Averesch NJ, Nunez-Bernal D & Krömer JO (2014) *In vivo* instability of chorismate causes substrate loss during fermentative production of aromatics. *Yeast* **31**: 333-341.
Wittmann C, Hans M & Bluemke W (2002) Metabolic physiology of aroma-producing *Kluyveromyces marxianus*. *Yeast* **19**: 1351-1363.
Xie N-Z, Liang H, Huang R-B & Xu P (2014) Biotechnological production of muconic acid: current status and future prospects. *Biotechnol Adv* **32**: 615-622.
Yu J-L, Xia X-X, Zhong J-J & Qian Z-G (2014) Direct biosynthesis of adipic acid from a synthetic pathway in recombinant *Escherichia coli*. *Biotechnol Bioeng* **111**: 2580-2586.
Zhang H, Li Z, Pereira B & Stephanopoulos G (2015) Engineering *E. coli-E. coli* cocultures for production of muconic acid from glycerol. *Microb Cell Fact* **14**: 134.
Zhang N-N, Dudgeon DD, Paliwal S, Levchenko A, Grote E & Cunningham KW (2006) Multiple Signaling Pathways Regulate Yeast Cell Death during the Response to Mating Pheromones. *Mol Biol Cell* **17**: 3409-3422.
Zhang Y, Liu G, Engqvist MKM, Krivoruchko A, Hallström BM, Chen Y, Siewers V & Nielsen J (2015) Adaptive mutations in sugar metabolism restore growth on glucose in a pyruvate decarboxylase negative yeast strain. *Microbial Cell Factories* **14**: 1-11.
Zilberstein D, Agmon V, Schuldiner S & Padan E (1984) *Escherichia coli* intracellular pH, membrane potential, and cell growth. *J Bacteriol* **158**: 246-252.

Appendices[7]

Supplementary files to chapter 2.1

Supplementary files 2.1.1 & 2.1.2: EMA networks

Excel files containing the stoichiometric networks of *E. coli* and *S. cerevisiae* for Elementary Mode Analysis. Sheets in the excel files are sorted by pathway, with the last sheet containing additional information regarding the literature the pathway was compiled from, key intermediates in order to identify the pathways, as well as maximum product carbon yields.

Supplementary files 2.1.3 & 2.1.4: NET-analysis networks

Excel files containing the stoichiometric networks as well as physical data for compartments and metabolites of *E. coli* and *S. cerevisiae* for Network Embedded Thermodynamic Analysis and results of the different thermodynamic analysis's, including detailed information on the infeasibility of the flux modes of each network.

Supplementary files 2.1.5 & 2.1.6: Yield plots

Product carbon yield vs. biomass plots of feasible and infeasible modes from EM-NET analysis of *E. coli* and *S. cerevisiae*. Each point in a chart corresponds to the specific product and biomass yield of the respective elementary flux mode. Yields are carbon yields in %.

Supplementary files to chapter 2.2

Supplementary file 2.2.1 – EMA networks

Excel file containing the reactions of the central metabolism of each organism supplemented by the respective pathway branch. Knock-out targets are highlighted in red. The file contains two excel sheets (*S. cerevisiae* & *E. coli*), each containing different networks for the respective pathways. The determined knock-out targets are indicated in red for each cut-set, potential overexpression targets in green. Reactions that were excluded from identification as knock-out target are highlighted yellow.

[7] For electronic copies of the supplementary files refer to the publications the respective chapters are based on or contact NJHA directly at: nils.averesch@uq.net.au

Supplementary file 2.2.2 – map of S. cerevisiae metabolic network

Graphical depiction of the metabolites interactions in the *S. cerevisiae* network. The reactions are identified by their number according to the network file (supplementary file 1). For better clarity “phosphate” (P) was left out from reactions in the figure unless involved in transport. Contribution to biomass formation (R95) is indicated with dashed arrows for the respective compounds. The two alternative routes to product formation are labelled with “a” to “d” according to Figure 5. Mitochondrial processes are highlighted in blue (mitochondrial only) and light blue (reactions exist in mitochondria and cytosol).

Supplementary file 2.2.3 – map of E. coli metabolic network

Graphical depiction of the metabolites interactions in the *E. coli* network. The reactions are identified by their number according to the network file (supplementary file 1). For better clarity “phosphate” (P) was left out from reactions in the figure unless involved in transport. Contribution to biomass formation (R78) is indicated with dashed arrows for the respective compounds. The two alternative routes to product formation are labelled with “a” to “d” according to Figure 5. Key metabolites for the knock-out strategy are highlighted in grey.

Supplementary file 2.2.4 – additional tables comparing yields and knock-out targets, figures (EFM distribution) for knock-out networks of chorismate-derived routes and glucose & glycerol co-feed

Supplementary table 1: Comparison of maximum product yields of wild-type networks for *S. cerevisiae* and *E. coli*

Supplementary table 2: Comparison of maximum product yields on glucose & glycerol co-feed of wild-type networks for *E. coli*

Supplementary table 3: Comparison of maximum and minimum product yields of knock-out strategies for *S. cerevisiae*

Supplementary table 4: Comparison of maximum and minimum product yields of knock-out strategies for *E. coli*

Supplementary table 5: Extended information on knock-out targets - enzymes and corresponding coding genes

Supplementary figure 5: EFM distribution of glucose & glycerol co-feed for *E. coli* wild-type networks

Supplementary figure 6: EFM distribution of knock-out strategies for *S. cerevisiae* isochorismate derived routes

Supplementary figure 7: EFM distribution of knock-out strategies for *S. cerevisiae* pHBA-pathway

Supplementary figure 8: EFM distribution of alternative knock-out strategy for *E. coli* isochorismate derived routes

Supplementary figure 9: EFM distribution of alternative knock-out strategy for *E. coli* pHBA-pathway

Supplementary files to chapter 2.3

Manuscript incorporated in chapter 2.3

Supplementary files to chapter 2.4

***Supplementary file 2.4.1: Sequence analysis & discussion of* ABZ1 *&* ABZ2**

XLS-file containing sequence alignments & codon usage analysis of the different examined *ABZ1* and *ABZ2* genes

Supplementary file 2.4.2: EMA networks & EFM plots for pABA production from different carbon-sources

XLS-file containing stoichiometric models for elementary mode analysis & result figures of product vs. biomass yield plots

Supplementary file 2.4.3: Design & evaluation of tracer experiments

PDF-document with additional information regarding results and discussion of the ^{13}C enrichment experiments