RWTH Aachen University

Institute of Applied Microbiology

Univ. Prof. Dr.-Ing. Lars M. Blank

Improved genome-scale model of *Ogataea polymorpha* based on experimental and computational approaches

Bachelor thesis

Author:	Constantin Schedel
First examiner:	Univ. Prof. Dr.-Ing. Lars M. Blank
Second examiner:	Dr. rer. nat. Martin Zimmermann
Supervisor:	Dr. Ulf Liebal

Date: 21 February 2018

Statutory Declaration in Lieu of an Oath

I hereby declare in lieu of an oath that I have completed the present bachelor thesis entitled “Improved genome-scale model of *Ogataea polymorpha* based on computational and experimental approaches” independently and without illegitimate assistance from third parties. I have used no other than the specified sources and aids. In case that the thesis is additionally submitted in an electronic format, I declare that the written and electronic versions are fully identical. The thesis has not been submitted to any examination body in this, or similar, form.

Aachen, 21 February 2018 Constantin Schedel (345337)

Abstract

Ogataea polymorpha is a thermotolerant, methylotrophic yeast and a promising host to generate platform chemicals from methanol, derived e.g. from carbon capture and utilization streams. The development of the organism into a production-strain with industrial applications requires additional strain design with the support of metabolic modelling on the basis of a genome-scale model (GEM). However, the current reconstruction, drafted from a published reconstruction of the related yeast *P. pastoris* and KEGG annotations, needs to be refined. A phenotype microarray (PM) experiment was performed to test the suitability of 190 metabolites as carbon-sources. To calculate the non-growth associated maintenance (NGAM) costs of growth on glucose and methanol, a chemostat-experiment was performed. During the research on inconsistent phenotypes between PM experiment and simulation, new reactions and metabolites were annotated with identifiers and gene-protein reaction relationships. Compared to the *P. pastoris* reference model, the GEM gained 35 new substrates and failed to metabolise eight. After refinement, the usage of 170 out of 190 substrates was correctly reproduced by the GEM. A new NGAM could not be implemented, because not enough measuring points could be generated. To further improve the GEM, substrate-specific biomass-compositions and growth-associated maintenance costs must be characterised. Multiple degradation-pathways must be identified in the metabolic pathways of the yeast. Additionally, the gene-protein-relationships need to be updated to represent *O. polymorpha* and allow quick identification of beneficial mutants.

Contents

List of Figures

List of Tables

List of Abbreviations

AOX	Alcohol oxidase
ATP	Adenosinetriphosphate
a.u.	Arbitrary unit
CDW	Cell dry weight
Chebi	Chemical entities of biological interest (Hastings et al., 2013)
COBRA-Toolbox	Constraint-based reconstruction and analysis-Toolbox
D	Dilution-rate
DAS	Dihydroxyacetone synthase
DHA	Dihydroxyacetone
DNA	Desoxyribonucleic acid
DO	Dissolved oxygen
E. coli	*Escherichia coli*
FBA	Flux balance analysis
FBP	Fructose 1,6-bisphosphate
g	gramm
GAP	Glyceraldehyde-3-phosphate
GCMS	Gas chromatography-mass spectrometry
GEM	Genome-scale metabolic model
GENRE	Genome-scale network reconstruction
Glc	Glucose
GPR	Gene-protein reaction relationship
h	Hour
H. polymorpha	*Hansenula polymorpha*
HPLC	High-performance liquid chromatography
IMP	Inosinemonophosphate
JGI	Joint Genome Institute – Mycocosm (Riley et al., 2016)
KEGG	Kyoto Encyclopedia of Genes and Genomes (Kanehisa et al., 2017; Kanehisa et al., 2000; Kanehisa et al., 2016)
l	Liter
lb	Lower bound
MeOH	Methanol
MBP	Mega base-pairs
NADH	Nicotinamide adenine dinucleotide
NGAM	Non-growth associated maintenance
OD	Optical density
O. polymorpha	*Ogataea polymorpha*
PM	Phenotype microarray
P. pastoris	*Pichia pastoris*
R	Coefficient of determination / R^2
RNA	Ribonucleic acid
S-Matrix	Stoichiometric matrix
T	Transmittance
t_c	Complete runtime of phenotype array cultivation (72 hours)
TCA-Cycle	Tricarbonacid cycle
ub	Upper bound
V	Volume
X5P	Xylolose 5-phosphate

1. Introduction

One of the overarching challenges in biotechnology is the development of engineered microbial strains. The production yields of many microorganisms is often times limited by their natural selection towards maximum responsiveness, rather than maximum overproduction of specific chemical components (Burgard et al., 2003). Additionally, the availability of whole-genome sequencing techniques led to the advanced understanding of many organisms in the last decade. These genomes, alongside increasing computational performances have opened the path towards advanced bioinformatic tools and the reconstruction of metabolic networks at a genome-scale level (Lopes et al., 2017). Such networks enable the simulation of metabolic networks and the interaction between internal cellular objectives and the optimization of chemical overproduction targets (Burgard et al., 2003). OptKnock, a bilevel optimization framework, suggested knockouts for chemical overproduction of target metabolites in *E. coli* that matched with published mutant strains back in 2003 (Burgard et al., 2003). Computational approaches have the benefit to suggest not only knockouts of competing pathways, but also complex mechanisms to compensate lost functions and metabolic fluxes (Burgard et al., 2003). In the context of biotechnological application of *Ogataea polymorpha,* there is no functional genome-scale model to benefit from these possibilities yet. This thesis will contribute to establish a genome-scale model of *Ogataea polymorpha*.

1.1 *In silico* modelling

In silico tools are an important resource and aid in studying the metabolic capabilities of organisms. One such tool, called genome scale network reconstructions (GENRE), is a collection of biochemical transformations, enzymes and their associated genes for a specific organism, linked by genome annotations or primary literature (Hamilton et al., 2014). Over the past decades, many GENREs have been created, the most comprehensive being that of *Escherichia coli* K12 (Liao et al., 2011). *E. coli* is the most studied microorganism, resulting in more knowledge about its genome annotations, functional characterisation and growth behaviour that can be used than for any other organism. This, allowed the creation of a better refined network of biochemical reactions from the gene protein-reaction relationships (GPRs) collected for *E.coli* (Liao et al., 2011).

Over all, GENREs can be used for contextualization of high-throughput data, guidance of metabolic engineering, directing hypothesis-driven discovery, interrogation of multi-species relationships, and network property discovery (Oberhardt et al., 2009). By adding a stoichiometric matrix to a reconstruction, a mathematical representation called a Genome scale metabolic model (GEMs) is created. GEMs allow a computational evaluation and manipulation of phenotypes (Hamilton et al., 2014).

1.2 Genome-scale metabolic model

A genome-scale metabolic model (GEM) is a mathematical representation of a GENRE. It stores all metabolites and mass- and charge balanced reactions of a target organism in a stoichiometric matrix. By adding constraints, such as thermodynamic constraints, limitations to metabolic fluxes as well as objectives, a GEM can be used to simulate phenotype expressions with the help of a variety of tools (Thompson et al., 2016).

The stoichiometric Matrix (S-Matrix) is the focus of a GEM. All stored metabolites are represented by the rows and all reactions are represented by the columns of the matrix. Metabolites that are not active in a specific reaction have zero in the matrix-intersection of their row and the column of the reaction. Active metabolites display the stoichiometry of the reaction. Therefore, metabolites that are used as substrates are represented by negative values while products have positive values. Because the model must differentiate between extra- and intracellular space, as well as all possible compartments of eukaryotic organisms, it is important to separate them in simulations as distinct reaction-spaces. For that reason, a unique metabolite occupies multiple S-matrix rows for each reaction space they occur in. Transport between compartments, including exchange reactions into and out of the cell round up the capabilities of the GEM.

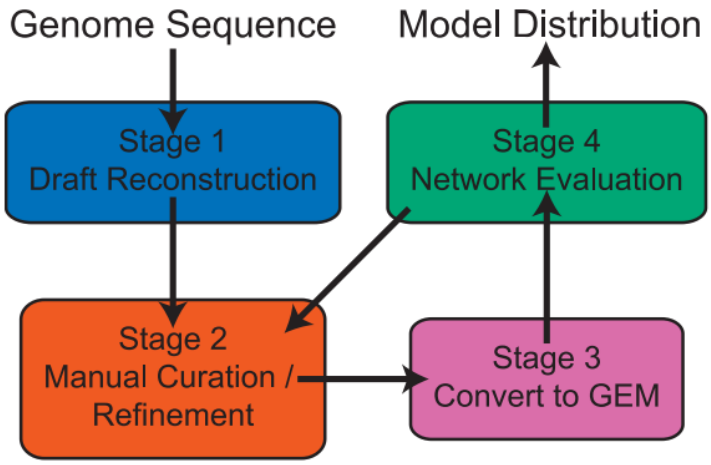


Figure 1. Schematic steps of network reconstruction process (Hamilton et al., 2014).

The construction of a GEM is separated into a total of 96 steps, summarised in four stages (Figure 1, Thiele et al. (2010)). Stage 1 sums up the primary annotation of information gathered into a GENRE and implementing it into a S-Matrix. Next, the information must be verified. This includes updating and refining GPRs, curating reactions to ensure thermodynamic and stoichiometrically accuracy and linking them to unique identifiers such as EC-numbers or a ChEBI-Identifier (Hastings et al., 2013), as well as their biochemical-pathways. Linking all possible interactions of an enzyme with multiple metabolites or all enzymes that can carry out a specific reaction, are two of these steps. Transport reactions must be added to ensure the functionality of the GEM, even if there are no annotations or primary literature for protein-channels to be found (Thiele et al., 2010).

Furthermore, a GEM needs a biomass equation to enable the simulation of growth phenotypes. This equation is non-enzymatic and contains all macromolecules and cofactors that make up the cell dry weight of the organism (Feist et al., 2010). All lipids, carbohydrates, proteins, RNAs or DNAs as well as cofactors are grouped together and included into this reaction as collections. The biomass equation also includes an ATP-hydrolysis, called the growth-associated maintenance (GAM). For more accurate growth predictions, a non-growth associated ATP maintenance reaction can be added as well (NGAM). This reaction simulates the energy costs of metabolites with unknown biosynthesis and degradation costs. (Hamilton et al., 2014). Because macromolecule composition and maintenance costs can adapt on different culture conditions, especially carbon sources (Tomàs-Gamisans et al., 2016) correct annotation of those are essential for correct phenotype calculations.

Next, the curated GENRE gets converted into a GEM. There are different software platforms that have been created for *in silico* experimentation using GEMs, among them the COBRA toolbox for Matlab (Schellenberger et al., 2011). In Stage 4 of the network reconstruction, the created reconstruction is evaluated with experimental data. For this purpose, flux balance analysis is the most frequently used way to simulate phenotypes (Hamilton et al., 2014). Inconsistent phenotypes include dead ends that block further analysis and hint at gaps in the network or demand transport of those dead-end metabolites outside the model. The prediction of growth rates on predefined substrates is another common step to validate biomass compositions and the pathways leading to their precursors (Hamilton et al., 2014).

1.3 COBRA Toolbox

The Constraint based Reconstruction and Analysis (COBRA) Toolbox is an *in silico* toolbox, available for Matlab and can be used to simulate, analyse and predict a variety of metabolic phenotypes using GEMs. First released in 2007, it can be used to construct bottom-up GEMs, based on experimental and bibliomic data (Schellenberger et al., 2011). One analysis tool for the COBRA toolbox is FBA, simulating steady-state fluxes through the metabolic model and minimizing or maximising the output of a predefined goal.

1.4 Flux balance analysis (FBA)

Flux balance analysis is a linear optimizer, used to calculate metabolic flows through a metabolic network. It simulates a maximal or a minimal optimization for a specific reaction of the model, predicting the phenotype of the modelled organism. Usually, these phenotypes are the production of important metabolites or growth-phenotypes under predefined conditions (Orth et al., 2010). FBA uses two kinds of restraints to achieve its goal. First, the stoichiometric coefficients of the S-matrix impose mass-balance constraints on the metabolic flows through the model. In addition, the production of a metabolite must be equal to the consumption of the same metabolite (Orth et al., 2010). This pseudo-steady-state results in a solution-vector v that must satisfy the equation (Figure 2):

$$S \cdot v = 0 \tag{1}$$

Additionally, the fluxes represented in the solution-vector are restricted by flux-bounds. These restrictions, represented by upper and lower bounds on every reaction, are used to restrict

unidirectional reactions as well as to define specific culture-conditions. Both restrictions combined define the space that can be used to distribute fluxes and optimize a specific objective (Orth et al., 2010).

The solution-vector v is not unique for a S-matrix. This is based on the shape of the S-matrix, which usually has more reactions than metabolites and thus contains more variables than equations in the calculation (Orth et al., 2010). The constraints are necessary to define the solution-space in which the solution can lie, but the number of solutions is still possibly infinite. In the COBRA toolbox, FBA can be tasked to give either the maximum or the minimum solution of this matrix-space. In consequence, the phenotypes calculated with FBA are restricted to the end result and do not evaluate the flux-compositions leading to the optimal solution. Phenotypes predicted with FBA have the flaw, that enzymatic activities cannot be considered. There are approaches to improve GEMs with the addition of enzymatic activities, but a complete artificial representation of a cell is not achievable with current technology (Lopes et al., 2017).

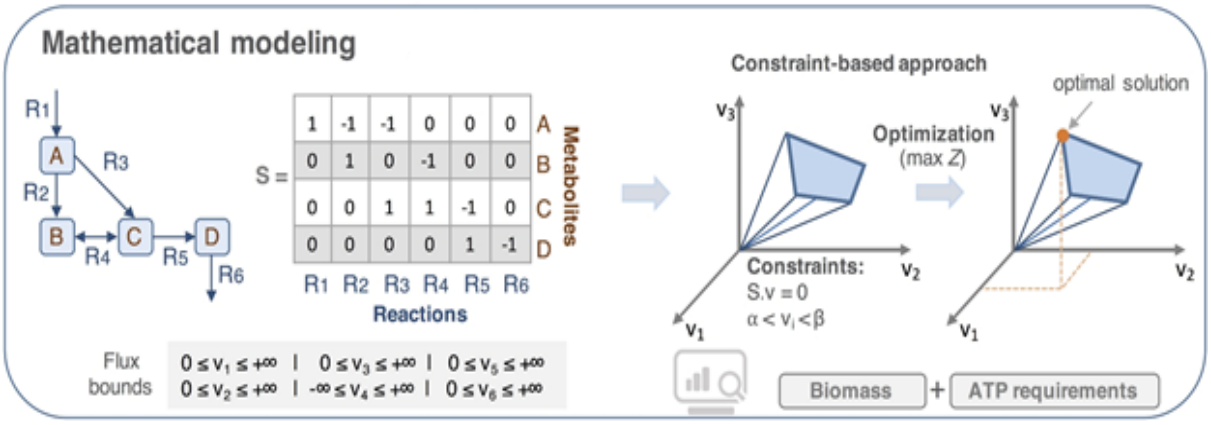


Figure 2: The conceptual basis of constraint-based modelling. Adapted from (Lopes et al., 2017)

1.5 Methylotrophy

The Utilization of C1 components as the sole carbon- and energy source is widely distributed in the evolutionary tree of microorganisms, meaning this skill is found among prokaryotes, as well as eukaryotes. Whereas prokaryotes developed ways to utilize a wide variety of C1 molecules with a variety of different pathways, eukaryotes utilize only methanol and methane from this spectrum as a carbon source (Yurimoto et al., 2011). Furthermore, this eukaryotic methylotrophs consist only of the yeasts *Candida, Pichia, Ogataea, Kuraishia*, and *Komagataella (Yurimoto et al., 2011).* The potential applications of methylotrophic yeasts have been studied since the 1970, showing that they are equipped with a strong, methanol inducible AOX Promotor. In addition, they have been model organisms for the biogenesis of peroxisomes (Yurimoto et al., 2011).

1.6 *Ogataea polymorpha*

Target of this thesis is to update a GEM of the methylotrophic yeast *O. polymorpha*. This yeast, formerly know under the synonym *Pichia angusta* and still often referred to as *Hansenula polymorpha*, is a member of the *Saccharomycetaceae* family (Figure 3). It occurs ubiquitous but most commonly in soil, spoiled orange juice, maize meal and the guts of various insect species (Ramezani-Rad et al., 2003). Colonies have a white colour and never grow filamentous. Isolates are homothallic and reproduce by budding.

O. polymorpha, together with the other methylotrophic yeasts *Pichia pastoris* and *Candida boindii*, are used as production platforms for industrial protein production such as pharmaceuticals (Ravin et al., 2013). Its most well-known strains are *H. polymorpha* DL-1, CBS4732 and NCYC495 (Ramezani-Rad et al., 2003), although the species includes several phylogenetically distinct strains that were reclassified as *O. polymorpha* and *O. parapolymorpha* (Ramezani-Rad et al., 2003; Suh et al., 2010). It is also known to be quite thermotolerant, with a growth-optimum at 30 °C to 37 °C and observed glucose-fermentations up to 45 °C (Ryabova et al., 2003).

The GEM of *O. polymorpha* is based on the most complex published GEM of *P. pastoris*, iMT2016 (Jente, 2017). Some of the many qualities of iMT1026 were its multiple substrate specific biomass compositions, ranging from glucose to mixed compositions of methanol and glycerol, as well as a refined network of annotations (Tomàs-Gamisans et al., 2016). A

consequence of using a model of a close relative of *O. polymorpha* (Figure 1) was, that the GPRs in the model are GPRs for the genome of *P. pastoris*.

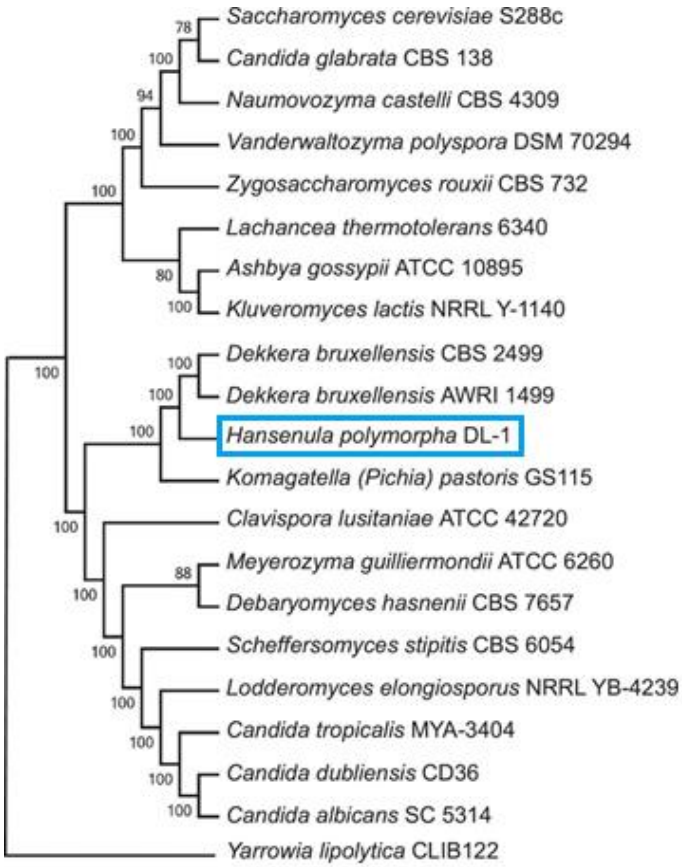


Figure 3: Phylogenetic position of *O. polymorpha* (formerly *H. polymorpha*) within Hemiascomycetes. Adapted from (Ravin et al., 2013).

1.7 Physiology of methylotrophic yeasts

The assimilation of methanol in methylotrophic yeasts is performed by the Xylolose-5-pathway (Figure 4). Initially, methanol is oxidised in the peroxisomes by alcohol oxidase (AOX) to formaldehyde and hydrogen peroxide. Formaldehyde is further fixed with xylose 5-phosphate (X5P) by the enzyme dihydroxyacetone synthase (DAS), producing dihydroxyacetone (DHA) and glyceraldehyde 3-phosphate (GAP) (Yurimoto et al., 2011). Both metabolites are transported into the cytosol. To form a cycle and regenerate X5P, a portion of the GAP is used to synthesise fructose 1,6-bisphosphate (FBP). Another part of the formaldehyde is oxidised to CO_2 by the cytosolic dissimilation pathway (Yurimoto et al., 2011). This way, methylotrophic

yeasts utilize a molecule that is seen as a core building block of a carbon-neutral future (Olah, 2005).

Formaldehyde and hydrogen peroxide are highly toxic, so growth on methanol as the sole carbon source requires the yeasts to emphasize on the methanol-degradation pathway. The enzymes that metabolize these compounds are all found in high amounts in the peroxisomes of the yeasts, linking production and consumption tightly together and reducing chances to harm the surrounding cell (Yurimoto et al., 2011). In case of *O. polymorpha*, methanol-limited continuous cultures consist of up to 80% of the cell volume in peroxisomes (Van Dijk et al., 2000). Furthermore, the AOX and DAS proteins are so prominent in the peroxisomes that they constitute to up to 60% of the cell protein in total (Van Dijk et al., 2000).

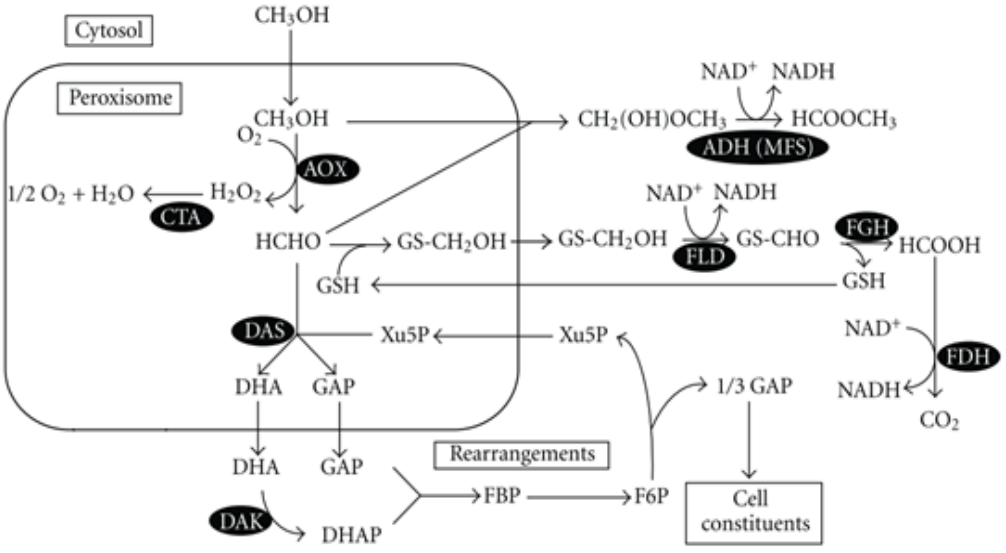


Figure 4: Xylolose-5-Pathway and Methanol degradation of methylotrophic yeasts. Adapted from (Yurimoto et al., 2011)

1.8 Biolog Phenotype Microarrays

Phenotype Microarrays (PMs) represent a technology to assay thousands of cellular phenotypes at once and thus collecting information about the underlying genotypes. This knowledge can be further used to improve the predictions and complexity of GENREs by comparing it to the growth phenotypes of its GEM counterpart. The first three of these arrays consist of carbon- and nitrogen-sources. PMs are based on 96 well microarrays with 95 unique metabolites checked for each array, and a negative-control on the upper-left corner.

Unlike transcriptomics and proteomics methods that are based on molecular information, PM arrays analyse the phenotype that are expressed from the organism. If an organism can utilize a metabolite as a carbon-source, NADH is used as a carrier for electrons to the electron-transport chain, resulting in a higher NADH turnover in the cell. PMs are built around a tetrazolium-dye that gets reduced if enough NADH is present in the cell. The reduction is visualised by a shift in the absorption spectrum of the dye to purple, making the reduction measurable via photometric methods. The optimal wavelengths for the dye used for yeasts and for biomass were obtained. 490 nm is the optical maximum of the dye and to avoid interferences, 750 nm were proposed for biomass yield (B. Bochner, Biolog personal communication). By selecting sub-circuits within the metabolic pathways, this technique allows to gather information about enzyme reactions that are expressed in the living organism without analysing its transcriptions or enzyme composition (Bochner et al., 2001).

1.9 Chemostat

Chemostats are bioreactors, capable of steady-state fermentations and are used for studying the impact of growth rates on cellular processes or for studying adaptive evolution (Ziv et al., 2013). A microbiological culture is considered to be in steady-state, if the growth-rate of the culture remains constant for one period of growth in an otherwise constant environment (Ziv et al., 2013). Steady-states can be achieved in a chemostat, by defining a constant dilution rate D. By selecting a constant flow F of new feed-medium to dilute a medium with the fixed volume V and an additional flow of downstream medium, the dilution rate is defined as (Novick et al., 1950):

$$D = \frac{F}{V} \tag{2}$$

The constant transport of cells out of the reactor creates an artificial mortality-rate for the microorganisms. This and the constant flow of new medium enables them to grow in a continuing exponential-phase, resulting in a constant growth-rate μ after multiple population doublings (Novick et al., 1950).

$$\mu = D \tag{3}$$

This means, that during a steady-state, all cellular mechanisms and extracellular influences lead to a constant exterior environment that can be used for multiple applications (Ziv et al., 2013). District characteristics, such as biomass yields and uptake rates of a growth-limiting carbon-sources, can be quantified. A successful chemostat must have a realization to ensure sterility-maintenance, mixing and aeration of the culture and constant flowrates (Ziv et al., 2013). Changes in the environmental conditions can lead to changes in the growth-rate and therefore disrupt steady-states.

1.10 Aim of this work

The genome scale model of *O. polymorpha* was first drafted from the *P. pastoris* model iMT1026 and had to be adjusted to fit *O. polymorpha* instead. Aim of this work was the improvement of the draft version. Two experiments were planned to improve the growth-phenotype predictions of the model. The first experiment was the calculation of the non-growth associated maintenance cost on methanol and glucose as the sole carbon sources. The second experiment was the analysis of growth-phenotypes on a variety of carbon- and nitrogen-sources, using phenotype micro-arrays. The data gathered from the experiment was then used to compare experimental and *in silico* growth-phenotypes. Observed differences were used as an approach to refine the metabolic capabilities of the model.

2. Material & Methods

2.1 Material

2.1.1 Machines & Equipment

Table 1: List of Equipment

Machine/ Equipment	Manufacturer
DASbox© Mini Bioreactor system	Eppendorf AG (Hamburg, Germany)
DASGIP© MP8	Eppendorf AG (Hamburg, Germany)
Eclipse E400	Nikon Instruments Europe BV (Amsterdam, Netherlands)
ES Series Lab Freezer and Refrigerator	Thermo Fischer Scientific Inc. (Dreieich, Germany)
Growth Profiler 1152	EnzyScreen BV (Heemstede, Netherlands)
Heraeus™ MegafugeTM 16R Centrifuge	Thermo Fischer Scientific Inc. (Dreieich, Germany)
Heraeus™ Pico 17 entrifuge	Thermo Fischer Scientific Inc. (Dreieich, Germany)
HPLC System Gold	Beckman Coulter (Brea, USA)
IKA-Combimag REOO magnetic stirrer	IKA©-Werke GmbH & Co. KG (Staufen, Germany)
Incubator IPP 800	Memmert GmbH & Co. KG (Schwabach, Germany)
Multitron Incubation shaker	Infors AG (Bottmingen, Switzerland)
MSC-Advantage™ Class II Biological Safety Cabinets	Thermo Fischer Scientific Inc. (Dreieich, Germany)
NewClassic Balances MS-series	Mettler-Toledo AG (Greifensee, Switzerland)
PIPETMAN Classic™	Gilson, Inc. (Middleton, USA)
Precision balance 510, 572	KERN & SOHN (Balingen, Germany)
Ultrospec™ 10 Cell density meter	GE Healthcare (Chalfont St Giles, UK)
Pipet-Lite Multi Pipette L12-300 XLS+	Mettler-Toledo GmbH (Gießen, Germany)
PowerWave HT Microplate Spectrophotometer	BioTek Instruments, Inc. (Winooski, USA)

2.1.2 Chemicals

All chemicals used were of analytical grade and purchased form Merck KGaA (Darmstadt, Germany), Sigma Aldrich Chemie GmbH (Steinheim, Germany), Carl Roth GmbH Co KG (Karlsruhe, Germany) and BIOLOG Life Science Institute Forschungslabor und Biochemica-Vertrieb GmbH (Bremen, Germany).

2.1.3 Media

Liquid media:

These media were used for pre cultivations

YEP medium: 10 g/L YE, 20 g/L peptone, 20g/L glucose with 200 mg/L gentamycin (Gent) and 100 mg/L nourseothricin (Nour) when needed

These media were used for main cultivation

According to Biolog:

Recipe for 1x PM1 & PM2 Inoculating Fluids from Stock Solutions: IFY-0 20 nm, Dye mix D, F or H 0.32 ml, sterile water 3.18 ml, cell suspension 0.5 ml

Recipe for 1x PM3 Inoculating Fluids from Stock Solutions: IFY-0 60 ml, Dye mix D, F or H 0.96 ml, 2.4 M D-glucose 3.0 ml, sterile water 6.54 ml, cell suspension 1.5 ml

IFY-0 and Dye mix H were provided by Biolog

Cell suspension is composed of cells with a turbidity of 62% in a Yeast Nutrient Supplement (NS) solution of 10 ml 48 mM L-leucine and 90 ml sterile water.

Adapted from Yan D. et al:

Methanol-Chemostat medium: 3 g/L KH_2PO_4, 6.6 g/L K_2SO_4, 0.5 g/L Mg_2SO_4, 2 g/L $(NH_4)_2SO_4$, 2% Methanol, 1 ml Vitamin solution, 0.1 ml Trace-elements and 0.4 g/L L-Leucine

Glucose-Chemostat medium: 3 g/L KH_2PO_4, 6.6 g/L K_2SO_4, 0.5 g/L Mg_2SO_4, 2 g/L $(NH_4)_2SO_4$, 2% Glucose, 1 ml Vitamin solution, 0.1 ml Trace-elements and 0.4 g/L L-Leucine

1000x trace elements:

0.15 g/L Na_2EDTA, 0.045 g/L $ZnSO_4 \cdot 7\ H_2O$, 0.01 g/L $MnCl_4 \cdot 2\ H_2O$, 3 mg/L $CoCl_4 \cdot 6\ H_2O$, 3 mg/L $CuSO_4 \cdot 5\ H_2O$, 4 mg/L $Na_2MoO_4 \cdot 2\ H_2O$, 45 mg/L $CaCl_2 \cdot 2\ H_2O$, 30 mg/L $FeSO_4 \cdot 7\ H_2O$, 10 mg/L H_3BO_3, 1 mg/L KI

1000x Vitamin solution

0.05 g/L D-biotin, 1.0 g/L Ca-D-pantothenate, 1.0 g/L nicotinic acid, 25.00 g/L *myo*-inositol, 1.0 g/L thiamine-hydrochloride 1.0 g/L pyridoxine hydrochloride, 0.2 g/L p-aminobenzoic acid (4-aminobenzoic acid)

2.2 Methods

2.2.1 Simulations

The draft-reconstruction for *O. polymorpha* called Opol-model is based on a published reconstruction model, created for the methylotrophic yeast *Pichia pastoris* (Tomàs-Gamisans et al., 2016). The GEM iMT1026 itself was the result of merging three earlier *P. pastoris* GEMs (iPP668, PpaMBEL1254 and iLC915) into one single model (Tomàs-Gamisans et al., 2016). Hence most GPRs are not yet updated to fit the *O. polymorpha* genome and are assumed from iMT1026. The same goes to empirical data, such as biomass compositions and NGAM, as well.

Phenotype-simulations were done with the COBRA-toolbox and the linear programming solver Gurobi (Gurobi Optimization, 2016). The boundaries, set for FBA were as follows: boundaries for active reactions were set to -1000 to 1000. Unidirectional reactions were set to 0 for either ub or lb. Most exchange reactions were deactivated. Waste-metabolites were able to leave the model. Only the uptake of H_2O, H^+, phosphate, oxygen, sulfate, Fe^{2+}, K^+, a nitrogen-source and a carbon-source was enabled. Under standard-conditions, these were NH_4 and MeOH. For example, if a growth-phenotype for different uptake rates were to be analysed, the 40:60 biomass-composition for glycerol and methanol was set as the target for model-optimization, and the lb for the carbon-source was adjusted. The S-matrix was then solved to extract the solution with the highest possible growth-phenotype under the set conditions. If not stated otherwise, the standard uptake-rate of 10 $\frac{\text{mmol}}{\text{gCDW}\cdot\text{h}}$ was chosen as the lb.

$$\begin{aligned}\text{Biomass composition } 40{:}\,60\ (\text{Glycerol g: MeOH g}){:}\ & 89.69\ H_2O + 89.69\ \text{ATP} \\ & + 0.057\ \text{LIPIDS} + 0.368\ \text{PROTEIN} + 0.33\ \text{CARBOHYDRATES} \\ & + 0.065\ \text{RNA} + 0.0019\ \text{DNA} + \text{cof} \\ & \rightarrow 89.69\ H^+ + 89.69\ p_i + 89.69\ \text{ADP} + \text{Biomass}\end{aligned}$$

2.2.2 Biolog Phenotype Arrays

The PM inoculation was prepared according to Biolog guidelines with *O. polymorpha* NCYC 495. The number of cells needed from a 24 h preculture for 20 ml NS-solution with 48x 62% transmission (T) was calculated. Necessary were information about the end-volume, end-OD and start-OD. The cells were centrifuged at 5000 rpm for 5 minutes, the pellet was washed with 0.9 % NaCl-solution and centrifuged again at 5000 rpm for 5 minutes. Biotic PM1 and PM2 were inoculated from the same media with a cell density of 62% T in each well. Biotic

PM3 arrays were inoculated parallel to PM1 and PM2 with PM3 medium. All Microarrays were placed into the plate-reader and measured at 490 nm for dye reduction and 750 nm for biomass-gain to generate t0 samples.

In total, samples from 6 points were collected from each biotic Microarray. Each at 490 nm and 750 nm. Apart from t0, the measurements took place after 24 h, 48 h, 52 h and 72 h. After each run in the plate reader, the Microarrays were placed in a growth-profiler for cultivation. For the measurements, the lids were taken off and the arrays were measured unsterile. This was done consequently to strong condensation on the inside of the lids that were noticed during the first run with one biotic and an abiotic PM1 array. Additionally, an abiotic Microarray was cultivated alongside the biotic arrays. It was prepared the same way biotic arrays were prepared, but instead of adding cell suspension, NS-solution without cells was added to complete the media. The growth-profiler was set at 37 °C and 150 rpm. To minimize contact with the heated floor of the growth profiler, all plates were placed on top of empty reusable microarrays.

2.2.3 Evaluation of PM data

The recommended evaluation process of Phenotype Microarray data included machines and software that was not available. For each array, a threshold value for dye reductions at 490 nm was created by statistical analysis (Chapter 3.2.3) and used to exclude substrates with a total slope between t0 h and t72 h of the threshold value or lower. Replicates were then checked for consistency. Only substrates with 66% of replicates surpassing the threshold were assumed to display growth. Next step was a manual curation step of the remaining substances. They were checked for inconsistencies in the abiotic arrays and in the slopes between consecutive measurements. If the correlation between abiotic 490 nm/750 nm values differed greatly from the abiotic negative-control, the metabolites were excluded.

The substances of the PM tests were checked of being metabolites in the *O. polymorpha* draft-model. If a substance was found with an exchange reaction, it was tested for *in silico* growth as the sole active carbon source. The list of positive *in silico* growth-phenotypes was compared to the growth phenotypes in the PM test afterwards.

2.2.4 Updating Metabolites and Reactions

For improving the *O. polymorpha* draft-model, I created a workflow (Figure 5) based upon the schematic steps for creating a GEM. First, a summary was prepared that showed all metabolites on PM1 and PM2 that were considered grown but failed to display growth in the model. This summary was annotated to include all KEGG-identifiers for the metabolites, and each metabolite was checked for associated EC-numbers and enzymes in the KEGG-Database. If reactions were found, they were compared to the JGI-database of *O. polymorpha.* This was done in two ways: (1) if a metabolite was part of a complex metabolic network, the KEGG-networks inside he JGI-database were used to determine associated reactions. (2) If that was not the case, for example if no metabolic pathway was annotated for a metabolite, the associated EC-numbers were compared to the JGI genomes if matches outside the KEGG-networks were available (Riley et al., 2016).

Reactions identified this way had to be compared to the list of reactions inside the model, as well as linked to GPRs of the JGI database. GPRs were added to the list of reactions before comparing it to the model. In the draft-model, most gene protein relationships were still GPRs for *P. pastoris* and should be updated. All missing reactions were then added to the draft-model with a GPR, a network -if available- and reaction-boundaries. New metabolites were added to the list of metabolites. If additional exchange-reactions were necessary for growth-phenotypes, they were added as well. Next, the remaining metabolites were checked for published primary literature that could be used as further evidence for the observed growth on PM. Additionally, promising pathways were checked in JGI for other reactions using the same EC-number. If such evidence was found, the reactions were added with their GPRs as well.

After the first refinement and manual curation was done, the updated model was loaded into matlab using the COBRA toolbox to test, if the phenotypes were changed correctly. If the Opol-model was now capable of growing on a metabolite as the sole carbon source, the curation was deemed as successful. If not, the model was searched for mistakes again. Additional reactions, like exchange-reactions for a gap-product, were added accordingly to the workflow. If no knowledge about possible reactions in *O. polymorpha*, related yeasts or reactions in general could be found, or if growth could not be achieved by the curated model, all possible approaches were saved for later curation.

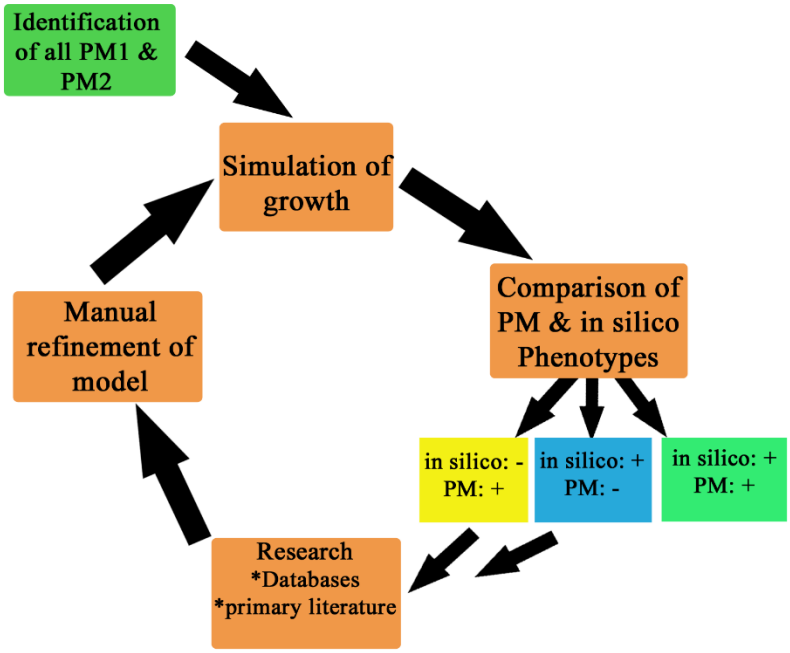


Figure 5: Workflow for the refinement and curation of the *O. polymorpha* draft-model based on the growth phenotypes of *O. polymorpha* on the PM1 & PM2 arrays.

2.2.5 DAS-Box

The four reactors of the DASbox were used to run two different duplicate fermentation-approaches. Two *O. polymorpha* NCYC 495 cultures were fed methanol as the sole carbon-source, the other two were fed glucose. To allow different dilution rates in each approach, all reactors were supplied by separate pumps and separate tubes. Medium was shared between cultures of the same approach. All compounds were autoclaved separately and assembled to the complete system under a clean-bench. The system was kept sterile throughout.

Each bioreactor was monitored by a pH-electrode, a dissolved oxygen (DO)-electrode and a temperature-electrode embedded in the caps (Figure 6). Airflow was saturated with water and passed to the bottom of the medium. An exhaust was placed on one of the cap-openings. Two rotors were placed a centimetre apart of each other in the bottom third of the reactor. Two tubes in the reactor were used for inoculation and as an effluent tube for waste-medium. Another opening held the temperature-electrode. The tubes for feed-medium ended at the cap where the medium dropped into the reactor. Three tubes were used to connect each bioreactor with a feed-medium, a collector for downstream-medium and a 7.5 % solution of ammonia. The flow-rates

through all 12 tubes could be changed and monitored separately. Before starting a run, the pH- and DO-electrode and all pumps were calibrated according to the guidelines by Eppendorf.

During the fermentation process, the DASbox controlled the temperature at 37 °C and pH at 4. The DASbox automatically fed ammonia-solution to the medium if the pH sank below the threshold. DO was monitored but not used to influence rotation speed or ventilation. Both were set at constant rates and were not changed during runs. The volume was kept constant by selecting a higher pump rate for Downstream than for Feed. The end of the exit tube was placed slightly above the media, resulting in a constant level and thus a near constant volume.

Each run started with a 24-hour batch-phase to stimulate growth and to start the chemostat with a high cell density. During the run, the reactors were checked every 24 hours for foam formation, increases in the liquid volume, OD of the culture and contamination. Foam formation was prevented with single drops of Antifoam 204. Daily microscope samples were taken to validate OD and to check for contamination and status of the yeasts inside the culture. Unlike the glucose feed medium that was prepared for multiple days, methanol-feed medium was changed daily to ensure a constant methanol-concentration. The cultures were checked for steady-state, i.e. in constant ODs and DO levels over multiple days. If possible, three samples instead of one were collected if steady-state was assumed. The media-flow rates for feed and downstream were adjusted for a new dilution-rate after the samples were collected.

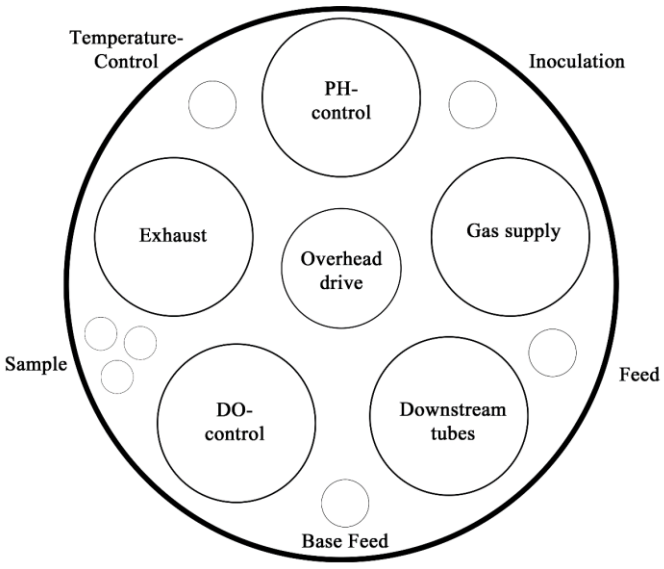


Figure 6: Arrangement of modules on the cap of the DASbox. In total, 5 bigger openings and 5 smaller openings could be equipped with modules. In the middle was an opening for the overhead drive.

2.2.6 Relationship between OD and g Cell Dry weight of *O. polymorpha*

A 24 h preculture of *O. polymorpha* NCYC 495 was checked for its OD and used to fill a falcon-tube with 10 ml of the cell suspension. Centrifugation was performed at 3000 rpm for 15 minutes. The pellet was resuspended in 1.5 ml 0.9 % NaCl-solution, placed into a standard reaction tube and dried at 70 °C for 24 h. The weight of the tube was determined before adding the cell suspension and after the drying. This was done for triplicates with three different ODs. Afterwards, the weight differences were plotted against the original OD of the 10 ml suspensions. The result ($0.181\ \mathrm{g \cdot CDW \cdot l^{-1} \cdot OD^{-1}}$ was calculated of the arithmetic mean of 0.188, 0.184 and $0.170\ \mathrm{g \cdot CDW \cdot l^{-1} \cdot OD^{-1}}$.

2.2.7 Hardware & Software

For this thesis, the cluster of the RWTH Aachen and FASTX 2 were used. MatLab, the Cobra Toolbox, and the linear problem solver Gurobi were used to simulate the metabolism of *O. polymorpha.* fmincon was used as a non-linear programming solver for manual calculations. The formula minimized by fmincon was $y = 1000 - x$. Furthermore, they were used for the statistical analysis and the creation of most diagrams in this thesis. Excel was used for the manual curation of the GENRE and the remaining diagrams.

2.2.8 Databases

KEGG, the Kyoto Encyclopaedia of Genes and Genomes, is a database that provides its users with detailed molecular-level information. Some of its compartments are KEGG-Pathway, KEGG-Reaction, KEGG-Enzyme and KEGG-Compound. Pathway and Enzyme were essential parts of this thesis, because they were used as a primary source of manual curation. KEGG-Enzyme was primarily used on basis of EC-numbers that catalogue enzymes into groups related by their catalytic capabilities (Kanehisa et al., 2017; Kanehisa et al., 2000; Kanehisa et al., 2016). KEGG reactions were usually displayed bidirectional and limited information was given if a reaction was restricted in a living cell. Unidirectional reactions were often times displayed by the appropriate KEGG-pathway. If neither reaction nor enzyme displayed a unidirectionality, new reactions were annotated as bidirectional.

The Joint Genome Institute – Mycocosm database (JGI-Mycocosm) was used as a knowledgebase for the strain *Ogataea polymorpha* NCYC 495 leu1.1 v2.0 (Riley et al., 2016). Its most

prominently used feature was JGI-KEGG. V.1.0 had a genome assembly size of 8.97 MBP, seven scaffolds and 0.0% of scaffold length in gaps. Since September 2010, the database is in v.2.0 with a refined genome annotation (Riley et al., 2016). EC-numbers were the primary enzyme-identifier in both KEGG and JGI and were used to search for GPRs.

During the research for the annotations as well as additional pathways for the degradation of PM metabolites, two more databases were used. The first, MetaCyc metabolic pathway database is a curated database with information for 2609 pathways and genomes of thousands of organisms (Caspi et al., 2007). Unlike JGI, MetaCyc has no annotated yeast-genomes and was therefore not prioritised as a source of information. The main function used was the pathway database. It was primarily used to find detailed information about pathways that were not annotated for *O. polymorpha* in JGI. The second, Chemical Entities of Biological Interest (ChEBI) is focused on chemical compounds and was used primarily to improve the added information about new metabolites (Hastings et al., 2013).

3. Results

A genome scale metabolic-reconstruction includes thousands of reactions and metabolites. This can be seen in the Opol_model of the yeast *Ogataea polymorpha*. At the start of this thesis, it consisted of 2226 reactions and 1734 metabolites. One main goal was to refine the model, based on found phenotypes in the phenotype microarrays. The other main goal was to determine *O. polymorpha* specific ATP maintenance consumptions for improvement of the growth-phenotype prediction. Before the results of these tasks are presented, I will examine two growth-related factors inside the model and their sensitivity to change.

3.1 Sensitivity Analysis

3.1.1 Glucose is *in silico* a more effective carbon-source than MeOH

The first test in the *O. polymorpha* model had the goal to analyse growth rates on varying glucose and methanol uptake-rates. The exchange reactions were varied from 0 to 20 $\frac{\text{mmol}}{\text{gCDW}\cdot\text{h}}$ in 0.5 $\frac{\text{mmol}}{\text{gCDW}\cdot\text{h}}$ intervals respectively and plotted afterwards (Figure 1). The results showed, that growth solely on glucose resulted in higher growth rates than on the same amount of MeOH, but the differences in carbon per mol for the two metabolites demand further analysis. 20 $\frac{\text{mmol}}{\text{gCDW}\cdot\text{h}}$ of both metabolites resulted in the highest overall growth (< 2.5 g CDW). It can also be shown, that changes in the uptake rates for a constant growth-phenotype are linearly connected. The borderlines between growth-intervals (Figure 7) visualise this behaviour. Further calculations were made to calculate the slope in uptake-changes under constant growth-phenotypes for a growth-interval between 0.05 and 0.5 $\frac{\text{gCDW}}{\text{h}}$ (Addendum 7.1.1). The mean of 10 slopes was -0.1549 ± 0.0005, meaning that the equivalent of 1 mmol Glucose was 6.46 mmol of MeOH. Since Glucose is a C6-body and methanol is a C1-body, additional 0.46 mmol of MeOH are necessary to enable the same growth-phenotype.

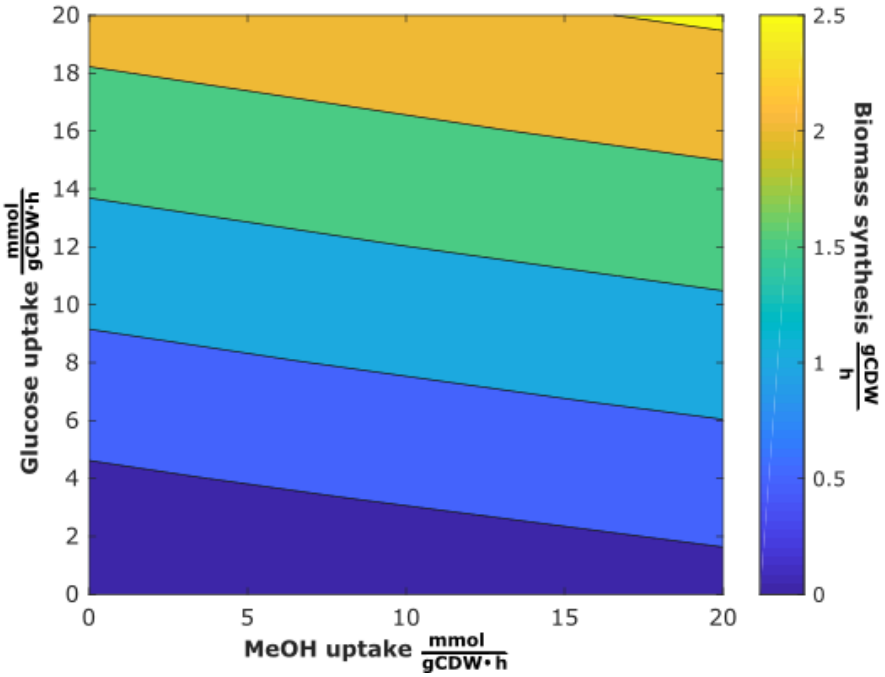


Figure 7: Growth yields of *O. polymorpha* on a mixture of methanol and glucose. Methanol- and Glucose-uptake are plotted on the x- and y-axis. The biomass synthesis on the z-axis is represented by a colour code. Each colour represents a growth interval.

3.1.2 *In silico* growth is most sensitive to Lipid changes

A second test was made to analyse the effect of different biomass-compositions on growth of *O. polymorpha*. I defined the carbon source as a 10 $\frac{\text{mmol}}{\text{gCDW·h}}$ methanol uptake, varied 4 components of the models biomass synthesis by ± 10% independendly from each other and compared the slopes (Figure 8 and Table 2). First, all 4 groups of tested biomolecules (carbohydrates, lipids, proteins and RNA) showed, that a higher demand resulted in a lower growthrate, because less biomass could be assembled if mor eprecursors were nessesary. DNA has not been manipulated, because different conditions do not affect the genome itself and thus the amount of DNA in a cell. Second, the slope for lipid-sensitivity was far higher than the other 3 slopes. A 10% change in lipid-demand resulted in a 30% biomass decrease. In comparison, the lowest slope was found for RNA with an 8% decrease under the same conditions.

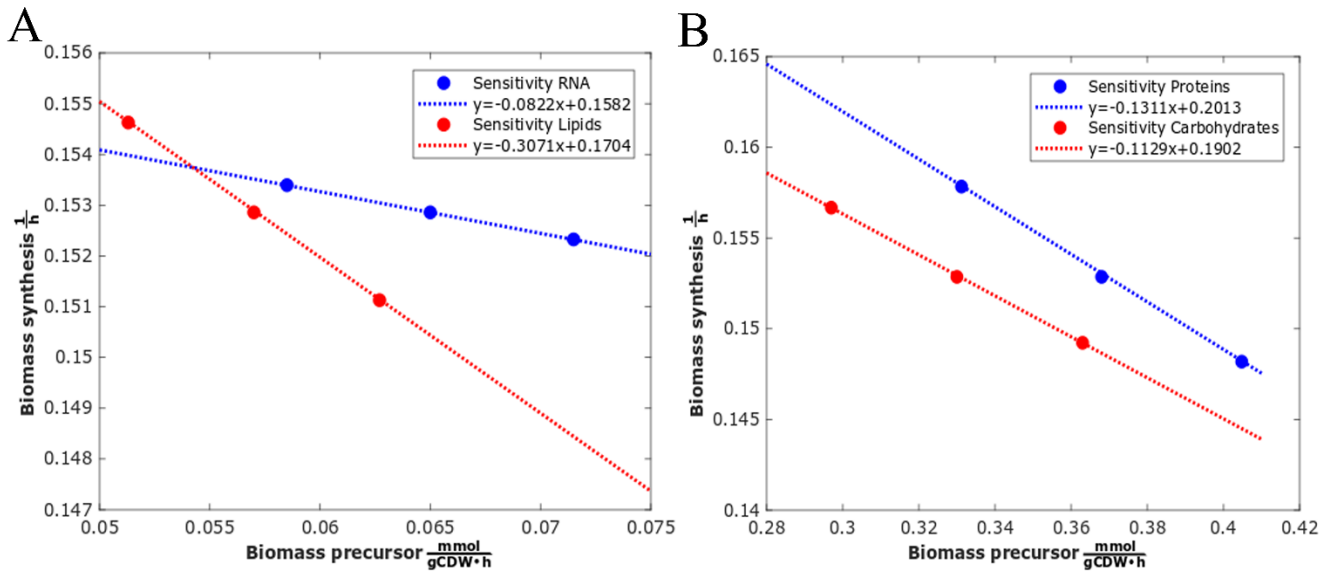


Figure 8: Reconstruction-draft Opol-model growth rates on different biomass compositions. Four compounds of biomass, lipids (A), RNA (A), Proteins (B) and Carbohydrates (B) were varied by ± 10% to analyse the sensitivity on biomass synthesis.

Table 2: Sensitivities of the reconstruction-draft Opol-model on altering biomass-composition

altered Biomass-Precursors	Slope/Sensitivity $[\frac{gCDW}{mmol}]$	stoichiometric coefficient $[\frac{mmol}{gCDW \cdot h}]$
Carbohydrates	0.1129	0.330
Lipids	0.3071	0.057
Proteins	0.1311	0.368
RNA	0.0822	0.065

3.2 Biolog Phenotype Arrays

3.2.1 *O. polymorpha* was able to grow on PM3 without a known source of nitrogen

The first step in the evaluation of the PM arrays was a comparison of the raw data from all three different phenotype microarrays. In this comparison, it was obvious that the results of PM1 and PM2 were comparable, whereas PM3 showed overall higher ODs. Especially the results of the PM3 negative controls were troublesome (Addendum 7.2.1). They did not show the lowest values on the test, few other wells with substrates had end-ODs as low as one-third of the negative controls. Comparing these observations with the abiotic test of PM3 showed, that this did not happen as an interaction between medium and substrates. The yeast seemed to grow without an available nitrogen source (Addendum 7.2.2). To investigate into this

hypothesis, I tested if *O. polymorpha* could grow on IFY-0, Glucose, Biolog H-Dye-mix and Leucine without an additional nitrogen-source. Additionally, I tested if excluding the dye-mix changed the behaviour. This assumption was based on the Inoculation manual that differentiates the carbon- and nitrogen tests solely on adding glucose for PM3. Four reaction tubes, two with the same medium-composition as prepared for the inoculation of PM3 and two with the same medium but the dye-mix, were inoculated from a master-plate, incubated at 37 °C for 24 hours and afterwards tested for growth using the standard OD-sensor. The yeast grew in all four tubes to ODs between 1.2 a.u. and 1.6. a.u. Because these values are distinctly higher than during the inoculation, PM3 was excluded from further analysis.

3.2.2 Detection of abiotic dye reduction

The remaining PM1 and PM2 arrays had to be analysed to establish a way to define growth and biotic dye reduction. First step was to find out if the absorptions at 490 nm and 750 nm represent independent features of the system. The start and end values gathered from the abiotic microarrays were plotted separately. Each 490 nm OD was plotted against the 750 nm OD of the same well and time (Figure 9 shows the plots of PM1). Noticeably, the bulk of wells on both plots were aligned linearly. At t0 (A), only a group of four wells did not show around a 1:1 ratio between 490 nm and 750 nm. Instead, their ODs at 490 nm were far higher. On the other hand, the variety after 120 h (tEnd) was far higher than at the start, but the bulk of wells stayed at a near 1:1 ratio, even if they spread up to ODs from 0.1 a.u. to 1.2 a.u. A linear regression fitted only at tEnd, because the group of variating metabolites had a bigger impact in the start measurement. Tests for correlation were done for abiotic PM1 and abiotic PM2 in parallel (Figure 9, Addendum 7.2.5 and 7.2.6). Both showed similar behaviour, although PM2 ended after 72 h and PM1 after 120 h. PM2 had more variating wells as well (Addendum 7.25 and 7.2.6). In both arrays, the metabolites that variated at t0 were the same that variated at tEnd. The results were interpreted as a correlation between the measurements at 490 nm and 750 nm, if no abiotic dye reduction occurred. The abiotic active metabolites were assumed to be the group of metabolites, that were clearly separable from the bulk.

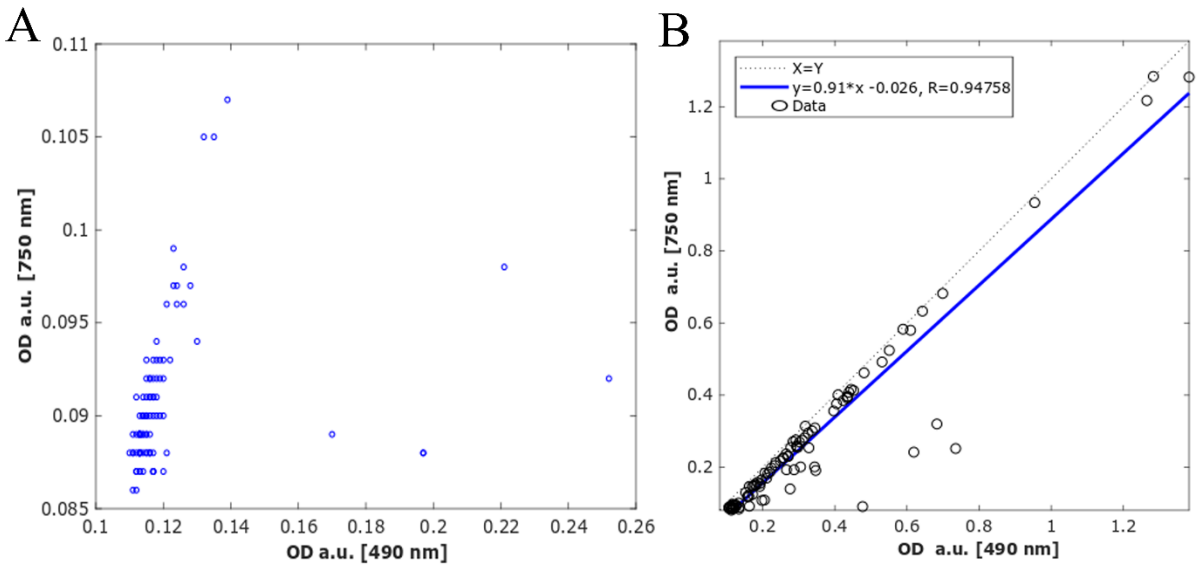


Figure 9: Abiotic PM1 ODs at 750 nm plotted against 490 nm ODs of the same time. Two times (A: t0 (0 h) and B: tEnd (120 h)) were compared. A linear regression was made for tEnd.

3.2.3 Growth thresholds were established with statistical analysis

As a result of the correlation between the two measured wavelengths, only the wavelength representing dye reduction was analysed to establish a way to define growth on the microarrays. First, all slopes between t0 end tEnd (72 h) were calculated. Instead of dividing the totals by 72 to receive values for $\frac{\text{a.u.}}{\text{h}}$, a variable t_c was introduced to represent the duration of the experiment. The slopes were then grouped into 0.02 $\frac{\text{a.u.}}{\text{t}_\text{c}}$ intervals for each microarray (Figure 10). In total, most wells showed absorptions at the far low end of the spectrum. Whereas only individual wells were part of intervals above 0.25 $\frac{\text{a.u.}}{\text{t}_\text{c}}$, the interval with the highest number of wells was the second lowest between 0.02 $\frac{\text{a.u.}}{\text{t}_\text{c}}$ and 0.04 $\frac{\text{a.u.}}{\text{t}_\text{c}}$. This applied to all five analysed microarrays (2x PM1 and 3x PM2) (Figure 10). This means that up to more than a third of all wells of a PM had dye reductions very close to the negative controls. All negative controls were localised in the lowest interval (0.00 $\frac{\text{a.u.}}{\text{t}_\text{c}}$ – 0.02 $\frac{\text{a.u.}}{\text{t}_\text{c}}$).

A threshold had to be established that systematically separated wells with positive dye reduction, from wells with negative dye reduction. Because the negative controls were all in the lowest interval, this meant that it could be possible that all wells above that had positive dye reduction. On the other hand, the dye could be reduced abiotically, by NADH released from cell digestion or from exposure to light. All this and water evaporation could all slightly

influence the results. Additionally, it would not be expected, that more than 90 % of all substrates show growth. A threshold, based on the distribution should be created.

Regarding the range of absorbance in Figure 10, the highest numbers were clear signs of active dye reduction in the wells. To investigate further into the lower end, these high-end wells had to be cut. For this, the arithmetic mean of the total of wells in a PM was computed and all substrates above this mean were marked as positive dye reduction. Now, the overall range of remaining wells for all 5 microarrays was below 0.3 $\frac{\text{a.u.}}{t_c}$ (Figure 11).

Next, the goal was to determine a threshold for the remaining wells. When the slopes of the lower half (below ~ 0.3 $\frac{\text{a.u.}}{t_c}$) are binned into intervals of 0.01 $\frac{\text{a.u.}}{t_c}$, the histogram of the relative frequencies of the slopes was approximately distributed as a skew normal distribution (Figure 11). Furthermore, the interval harbouring the most wells was now the interval between 0.02 $\frac{\text{a.u.}}{t_c}$ and 0.03 $\frac{\text{a.u.}}{t_c}$. The threshold for positive dye reduction was now defined as:

$$\text{dye reducton} < \text{arythmetic mean (x)} + \text{standart deriviation (x)} \qquad (4)$$

With x being the number of wells per interval. Therefore, a 68 % confidence interval was selected to be the threshold.

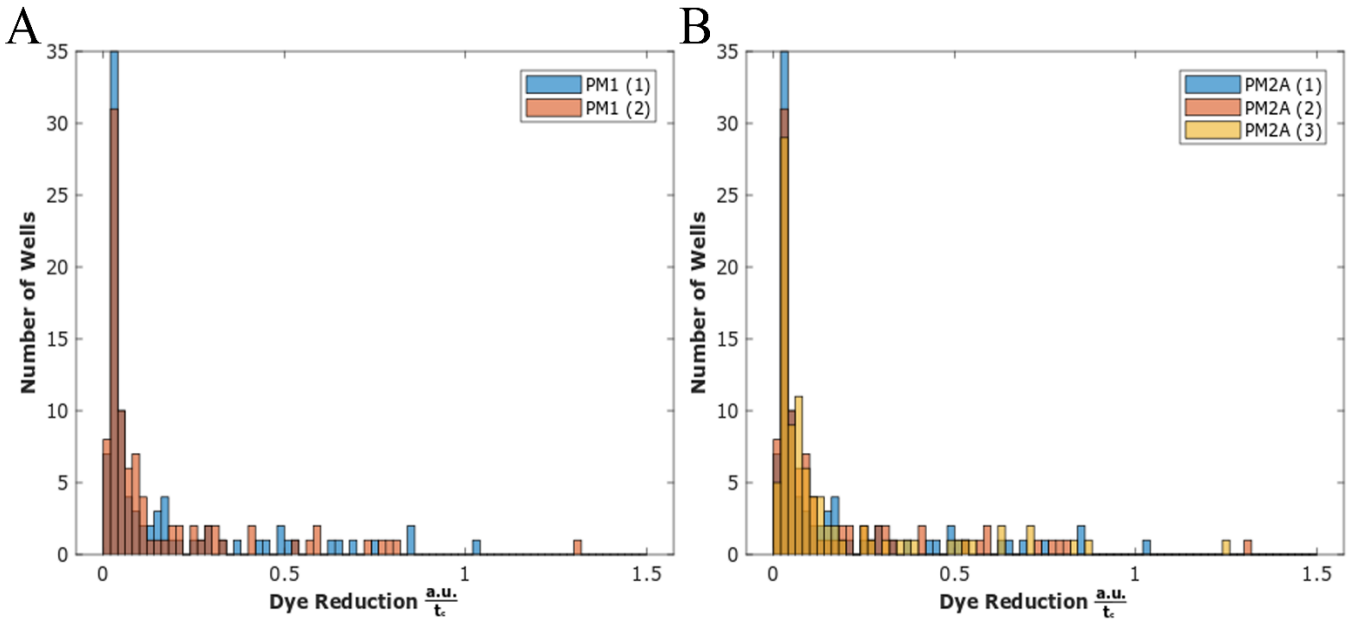


Figure 10: PM OD-Histograms. PM1 duplicate (A) and PM2 triplicate (B) slopes between t0 and tEnd were collected in bins of 0.02 $\frac{\text{a.u.}}{t_c}$.

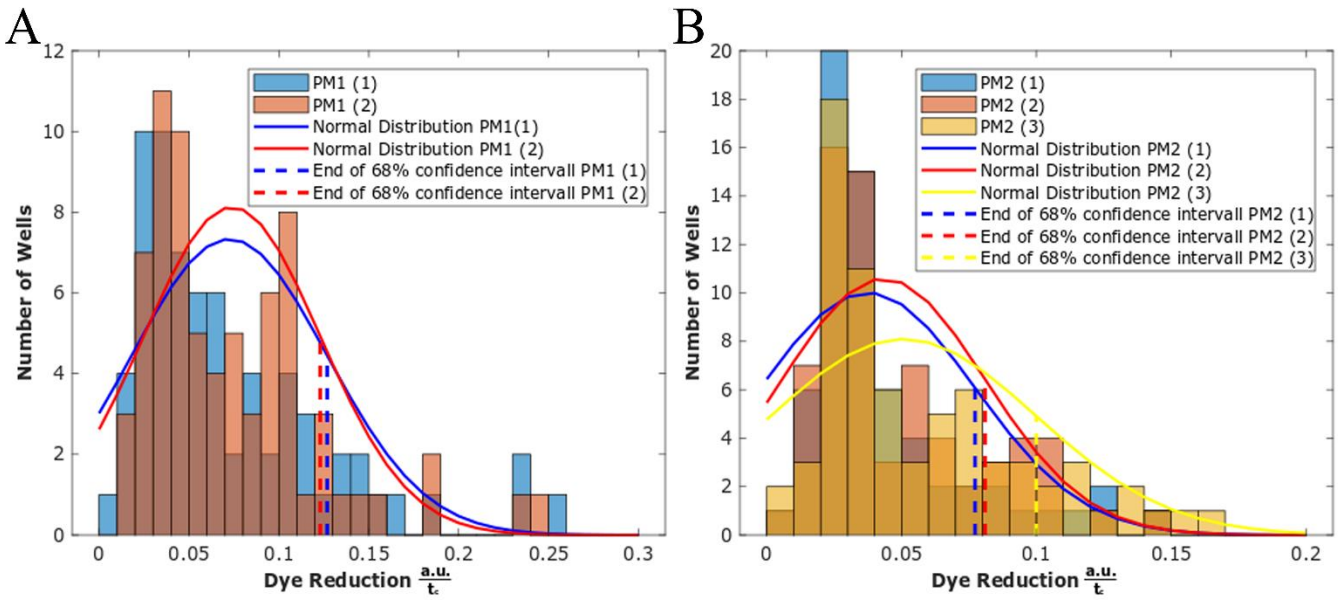


Figure 11: Statistical Analysis of PM OD-Histograms. PM1 duplicate (A) and PM2 triplicate (B) substrates with slopes below the arithmetic means were collected in bins of 0.01 $\frac{\text{a.u.}}{t_c}$. For every approach, a normal distribution for the arithmetic mean and standard derivation of the approach was projected on top of the histogram. The 68 % confidence interval for each approach is marked by a dashed line.

3.2.4 Most substrates showed the same phenotypes and many new phenotypes can be investigated for implementation into the model

Before a comparison between experimental and *in silico* phenotypes could be done, the lists of each array had to be merged and manually curated. At this point, 5 arrays in total – two PM1 and three PM2 arrays- were left. Therefore, all PM1 metabolites that did not grow on both approaches and PM2 metabolites with growth on one single array were excluded, resulting in a minimum 66% threshold. Substrates with extraordinary behaviour on the abiotic arrays were checked and excluded when seen fit (Addendum 7.2.3). If a metabolite had suspicious difference in its behaviour that can be linked to contamination, for example if large changes in the OD occurred at different time segments, it was excluded too (Addendum 7.2.4).

Now, the final lists of experimental growth-phenotypes on the PM-arrays were compared to the *in silico* growth-phenotypes of the same substrate (Figures 12 and 13 and addendum 7.2.8). The *in silico* phenotypes were created with FBA on media without additional carbon-sources. In total, four different combinations of phenotypes were observed. If the GEM generated a growth yield on a substrate, the experiment either confirmed (++ / green) or negated it (+-, blue). On the other hand, there were many substrates that had no positive in silico growth but had grown on the arrays (-+ / yellow). The fourth combination was, that neither in silico nor experimental tests showed growth (-- / white). Regarding the proportions of each group, (-+ / yellow) was the second most frequent with 35 metabolites, following 13 (++ / green) and 8 (+- / blue) metabolites. 134 substrates were (-- / white). In total, 147 out of 190 phenotypes had corresponding growth effects between simulation and experiment.

PM1	1	2	3	4	5	6	7	8	9	10	11	12
A					Succinic acid			L-Proline		D-Trehalose	D-Mannose	Dulcitol
B		D-Sorbitol	Glycerol				DL-a-Glycerol Phosphate	D-Xylose	L-Lactic acid		D-Mannitol	L-Glutamic acid
C				D-Ribose		L-Rhamnose	D-Fructose	Acetic acid	a-D-Glucose	Maltose		
D	L-Asparagine					a-Ketoglutaric acid					Sucrose	
E	L-Glutamine						a-Hydroxybutyric acid		Adonitol	Maltotriose		
F		Citric acid			Fumaric acid							Inosine
G												L-Malic acid
H					D-Psicose							

Figure 12: Comparison of PM1 growth-phenotypes. In total, there were 96 wells in 8 columns and 8 columns. Rows are marked with numbers from 1 to 12, columns with letters from A to H. Metabolites with positive growth-phenotypes *in silico* and experimental are marked green, only *in silico* blue and only experimental yellow. Metabolites that fall under one of those categories have their names displayed. All others are left blank.

PM2A	1	2	3	4	5	6	7	8	9	10	11	12
A										Laminarin		
B			b-D-Allose	Amygdalin		D-Arabitol	L-Arabitol			i-Erythritol	D-Fucose	
C	Gentiobiose			D-Melezitose	Maltitol			3-Methylglucose				Palatinose
D	D-Raffinose						Turanose	Xylitol				
E				Citramalic acid					Hydroxy-butyric acid			
F					Oxalomalic acid				Sorbic acid			
G		L-Alaninamide		L-Arginine					L-Isoleucine	L-Leucine		
H	L-Ornithine			L-Valine					Dihydroxy-acetone			

Figure 13: Comparison of PM2 growth-phenotypes. In total, there were 96 wells in 8 columns and 8 columns. Rows are marked with numbers from 1 to 12, columns with letters from A to H. Metabolites with positive growth-phenotypes *in silico* and experimental are marked green, only *in silico* blue and only experimental yellow. Metabolites that fall under one of those categories have their names displayed. All others are left blank.

3.3 Refinement of the GEM

The results of the Phenotype-Microarrays had to be compaired with primary literature to refine and curate the Opol-model. Two groups obtained from the PM data were used for the refinement. The first being metabolites without positive growth-phenotypes in the model that showed growth effects in the experiment and the second one being the eight metabolites, that had only *in silico* growth effects. The JGI-Database for *O. polymorpha* includes a KEGG-Pathway section with annotated, enzyme catalysated reactions found for the yeast. This tool was used as the first step for finding related reactions for the metabolites.

3.3.1 A variety of reactions were annotated based on JGI knowledge

Of the reactions using L-proline as a metabolite, three were annotated with two EC numbers for *O. polymorpha* (Figure 14). The first reaction was catalysed by the proline iminopeptidase, hydrolysing N-terminal proline from a polypeptide-chain. This reation was not yet included in the Opol-model. Therefore, it was implemented with the EC-number, pathway and GPR of the enzyme, but it was deactivated right afterwards. The reason for this was, that the model does not differentiate the metabolite $\mathrm{Peptide_n}$ from $\mathrm{Peptide_{(n-1)}}$. L-Proline could be created without a cost, unless the simulation surpresses loops. The second and third reaction found for *O. polymorpha* were both catalysed by the L-proline oxidase (EC 1.5.1.2), leading to Pyrroline-5-carboxylate and reducing either $NADP^+$ or NAD^+ in the process. Both reactions were already part of the Opol-model, making 3.4.11.5 the only new addition.

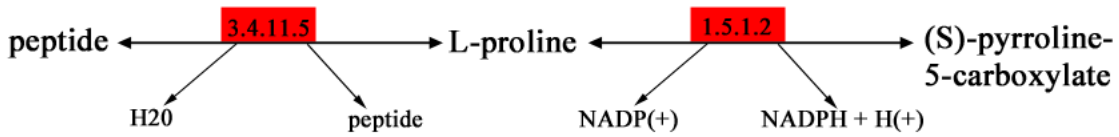


Figure 14: Annotated reactions for L-proline utilization. The reactions are an extract of the arginine and proline metabolism. Red EC-numbers mark reactions annotated for *O. polymorpha* NCYC 495 leu1.1 v2.0 in JGI.

Adonitol dehydrogenase (EC 1.1.1.56, Figure 15) oxidises adonitol under reduction of either NAD^+ or $NADP^+$ to D-ribulose. The ketopentose can be further processed into D-Ribulose-5-phosphate, connecting adonitol to the pentose phosphate-pathway. Only the Redox-reaction involving NAD^+ was found for *O. polymorpha*. Adonitol, also known as ribitol, was not yet part of the Opol-model and required an exchange reaction to be utilized. More metabolites that lacked exchange reactions were sn-Glycerol-3-phosphate and L-glutamine. All of them were already part of the model and had no further reactions to annotate. The growth-phenotypes on PM1 showed, that the yeast was capable of growing on both metabolites and the changes led to positive growth-simulations for L-glutamine and sn-Glycerol-3-phosphate.

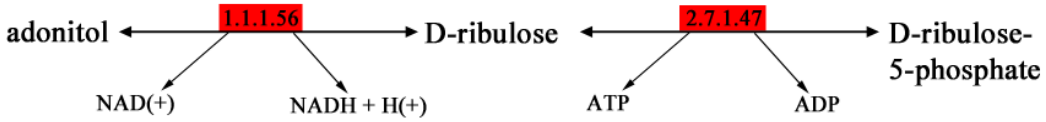


Figure 15: Annotated reactions for D-ribulose utilization. The reactions are an extract of the pentose and glucoronate interconventions. Red EC-numbers mark reactions annotated for *O. polymorpha* NCYC 495 leu1.1 v2.0 in JGI.

Before the manual refinement, three reactions were annotated in the Opol-model for the utilization of inosine (Figure 16). (1) The enzyme 5'-mononucleotidase acted on the ester-bond of IMP. (2) Inosine and ammonia could be used to produce adenosine and reaction (3) led to hypoxanthine and α-D-ribose-1-phosphate via the inosine phosphorylase. A new reaction was found, that also led to hypoxanthine. This reaction is catalysed by purine nucleosidase with D-ribose as the second product. Indeed, D-ribose itself was part of the PM1 array, but it was excluded because abiotic activity. However, this does not dirctly lead to the question if D-ribose can be part of the model, only if it can be utilized as a carbon-source. After the update, a positive growth-phenotype was simulated for inosine.

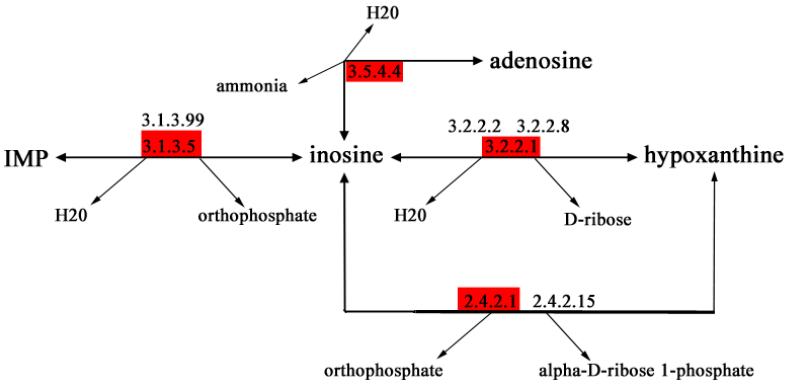


Figure 16: Annotated reactions for inosine utilization. The reactions are an extract of the purine metabolism. Red EC-numbers mark reactions annotated for *O. polymorpha* NCYC 495 leu1.1 v2.0 in JGI.

The next metabolite with experimental growth but missing in the model was Amygdalin, a cyanogenic glycoside composed of a nitrile-group and two D-glucose-units. JGI showed, that *O. polymorpha* NCYC leu1.1 v.2.0 had a GPR for the enzyme beta-glucosidase. This enzyme catalysed two reactions, leading from amygdalin to prunasin and further to mandelonitrile (Figure 17). All three metabolites had to be newly implemented. Each reaction also separated a glucose-unit from amygdalin and prunasin respectively. However, the production of mandelonitril led to another problem. FBA demands metabolic flux compensations for all produced metabolites and mandelonitrile is not pat of any additional pathway of the model, making it a gap-product. This led to the implementation of an additional exchange reaction for the export of mandelonitril besides the exchange reaction for amygdalin. Altough two glucose-units were produced for one unit of amygdalin, the model was not capable of simulating a growth-phenotype on amygdalin as the sole carbon source. No additional information was obtainable for utilization of the metabolite.

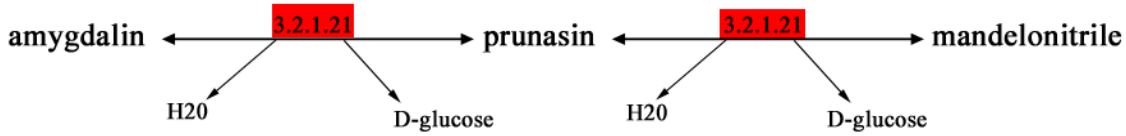


Figure 17: Annotated reactions for amygdalin utilization. The reactions are an extract of the Cyanoaino acid metabolism. Red EC-numbers mark reactions annotated for *O. polymorpha* NCYC 495 leu1.1 v2.0 in JGI.

Dihydroxyacetone, a metabolite on PM2, was found to be part of the glycerolipid metabolism and the link between glycerole and glyceronephosphate (Figure 18). JGI however had only one

of the two directions annotated for the yeast. The glycerone kinase phosphorylates glycerone to glyceronephosphate under consumption of one ATP. Glyceronephosphate on the other hand is part of the glycolysis and an isomer to glyeraldehyde-3-phosphate. No information was obtainable for the connection to glycerole. These changes were found to be insufficient to allow growth-simulation on Dihydroxyacetone.

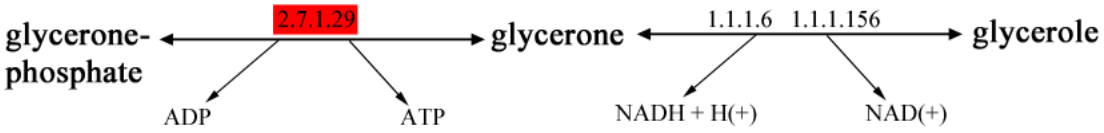


Figure 18: Annotated reactions for glycerone utilization. The reactions are an extract of the Glycerolipid metabolism. Red EC-numbers mark reactions annotated for *O. polymorpha* NCYC 495 leu1.1 v2.0 in JGI.

Only a few reactions were annotated inside the butanoate metabolism of *O. polymorpha* NCYC leu1.1. The largest path of connected reactions inside the pathway led from L-glutamate to 4-aminobutanoate and across succinate-semialdehyde to succinate (Figure 19). 4 -Hydroxybutanoic acid, also linked to succinatesemialdehyde, was not found for the organism. A genome-search for the EC-numbers of the two reaction towards succinatesemialdehyde showed, that the genome had genes for the enzyme with the EC-number 1.1.1.-. In regards of the positive growth-phenotype on the PM array and the close connection to succinate, the assumption was made, that this enzyme was capable of unspecifically catalysing the nessesary reaction. Alltogether, two reactions were newly annotated. The first beeing 1.1.1-, marked as an unspecific addition, and the second being 1.2.1.16, a reaction also not yet implemented into the model.

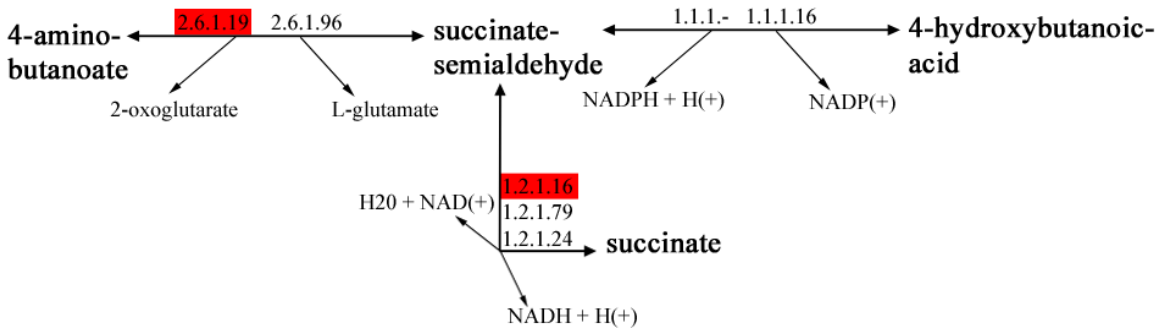


Figure 19: Annotated reactions for 4-hydroxybutanoic acid utilization. The reactions are an extract of the butanoate metabolism. Red EC-numbers mark reactions annotated for *O. polymorpha* NCYC 495 leu1.1 v2.0 in JGI.

3.3.2 Maltase was found to catalayse a variety of saccharides in *O. polymorpha*

The enzyme maltase (EC-number 3.2.1.20), also known as alpha-glucosidase, hydrolyses O- and S-glycosyl compounds. The EC-number is not linked to a single enzyme in KEGG, but to muliple enzymes that mainly hydrolyse (1→4)-alpha-glycosidic-bounds and occur in the galactose- as well as the starch and sucrose metabolisms . In JGI, maltase was found to catalyse two new reactions for metabolites on the PM-arrays (Figures 20 and 21). Maltose and sucrose both get split by the enzyme to form either two glucose units or one glucose and one fructose unit. This newly annotated reaction was essential for their utility as carbon-sources. Besides this information, a paper proposed, that the maltase of *O. polymorpha* is capable of hydrolysing even more saccharides. In this paper, the catalytic efficiencies of maltase for multipe other PM metabolites were presented. This showed, that the enzyme MAL1 of *O. polymorpha* can ultilize turanose, maltotriose, melezitose, palatinose and α-methylglycoside (Viigand et al., 2016). Of these five metabolites, four had strong growth effects on the PM except α-methylglycoside. Additionally, of the five metabolites, melezitose and α-methylglycoside were the ones with the lowest catalytic efficiency and the second was the only one without a growth-phenotype in both literature and PM-experiment (Viigand et al., 2016). This information was used to annotate four additional reactions for the maltase GPR. All six implemented or updated metabolites simulated growth afterwards.

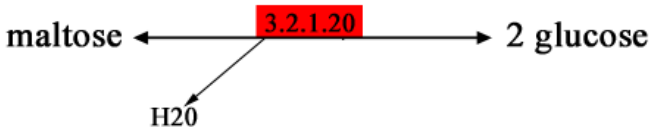


Figure 20: Annotated reactions for maltose utilization. The reactions are an extract of the starch and sucrose metabolism. Red EC-numbers mark reactions annotated for *O. polymorpha* NCYC 495 leu1.1 v2.0 in JGI.

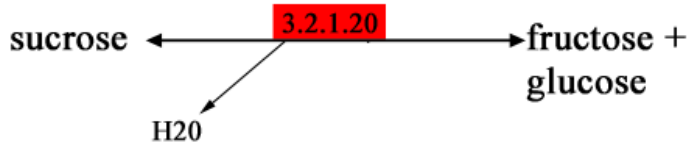


Figure 21: Annotated reactions for sucrose utilization. The reactions are an extract of the starch and sucrose metabolism. Red EC-numbers mark reactions annotated for *O. polymorpha* NCYC 495 leu1.1 v2.0 in JGI.

3.3.3 Contradictory growth information on dulcitol

On the PM1 array, dulcitol did not surpass the treshhold for both dublicates (Addendum 7.2.8). However, the negative-duplicate fell below the treshhold, because because its OD sank from 52 h to 72 h. This was taken as a reason to further analyse the metabolic pathway of dulcitol, despite the threshhold. A reaction was annotated in the Galactose-metabolism pathway of *O. polymorpha* that tied dulcitol, also called galactitol, to D-galactose, catalysed by the aldehyde reductase (Figure 22). This finding was implemented into the model. D-Galactose itself had no further reactions annotated for the yeast. Furthermore, it was also a substrate on the PM1 array and showed no growth-phenotype. Earlier investigation on the metabolism of *O. polymorpha* reassured this result. Lisanne Jente showed, that the yeast does not have the nessesary enzyme for D-galactose degradiation (Jente, 2017). This led to the hypothesis, that an alternative degradation-path for dulcitol migh be available. Since the enyzme catalysing D-galactose to dulcitol is found to be present in *O. polymorpha*, an alternative pathway for dulcitol would enable the model to utilize the saccharide as well. This would contradict the other observations, unless the yeast is capable of utilizing intracellular D-galactose, but can not transport it into the cell. In KEGG and MetaCyc, two distict pathways are found to degradate dulcitol and connect it to the glycolysis and the pentose-phosphat pathway (Figure 23). They first one leads to D-frutose-6-phosphate and was investigated for *Agrobacterium tumefaciens* (Wichelecki et al., 2015) and the second one leads to GAP and DHAP and was found for *E. coli* (Nobelmann et al., 1996).

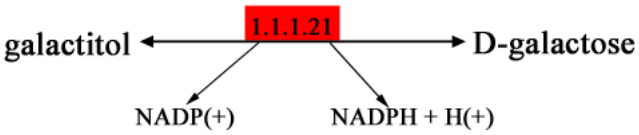


Figure 22: Annotated reactions for sucrose utilization. The reactions are an extract of the Galactose metabolism. Red EC-numbers mark reactions annotated for *O. polymorpha* NCYC 495 leu1.1 v2.0 in JGI.

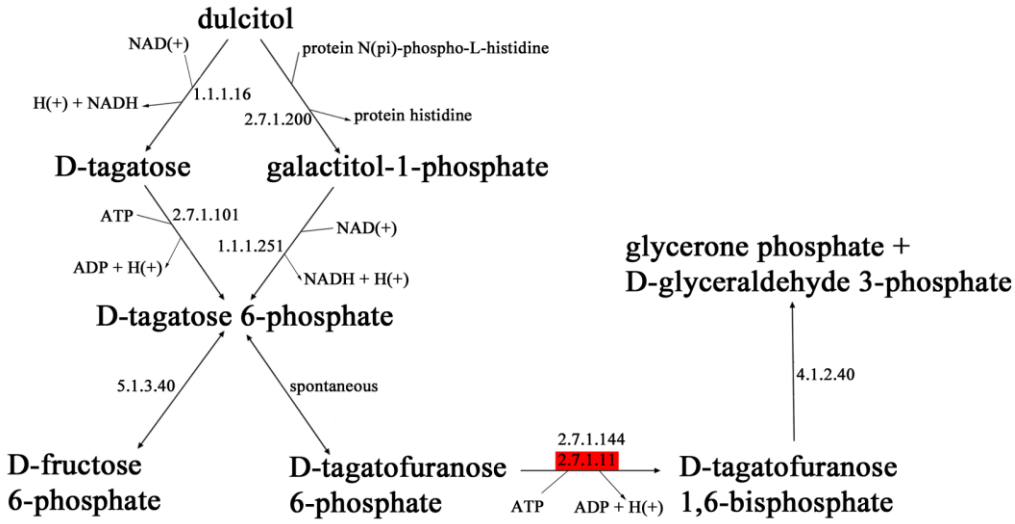


Figure 23: Possible degradiation-paths for dulcitol. The reactions are an extract of the galactose metabolism. Red EC-numbers mark reactions annotated for *O. polymorpha* NCYC 495 leu1.1 v2.0 in JGI. Directions were based on MetaCyc.

3.3.4 Several metabolites were found to be blocked by gap-metabolites

FBA requires stoichiometric balance for each metabolite to simulate steady-state. Therefore, gap-metabolites in otherwise functional metabolic pathways block the simulation. Multiple PM metabolites where found to have annotated reactions in the model or where updated to have them, but could not be ultilized via FBA. By enabling the exchange of uracil and xanthine, the growth-phenotypes of L-glutamic acid, L-asparagine, L-arginine and D-arabitol could be simulated. Uracil was found by searching for a gap-metabolites in the degradiation of L-glutamic acid and activating possible candidates. This also enabled most of the other mentioned metabolites. Other metabolites that were already annotated, metabolites with often times multiple reactions linked to their degradiation but no positive growth-phenotypes are likely to have other gap-metabolites blocking them as well. This case is assumed for L-leucine, L-valine, L-arabitol, dihydroxyacetone and amygdalin. Mandelonitril, the newly implemented

degradiation-product of amygdalin, was given an active exchange-reaction but it nonetheless did not lead to growth, despite the direct connection to D-glucose. Dihydroxyacetone was first excluded in the abiotic test, but was later investigated because it was the substrate with the highest continuing ODs at 490 nm during the PM2 experiment, besides being annotated in the model.

3.3.5 Most (+-)-Metabolites were annotated for *O. polymorpha*

Only eight metabolites of 190 were placed in the group of metabolites with *in silico* growth-simulation, but no evidence for actual growth. Of those eight metabolites, five were part of the TCA-cycle and therefore core metabolites of aerobic metabolism (Figure 24). All essential reactions of the TCA-cycle were annotated at least once in the JGI-database as well as the Opol-model, but this information does not directly indicate if the yeast can also assimilate these organic acids. Another noteworthy observation regarding the (+-)-metabolites was, that they were heavily linked to abiotic dye-reduction. Of the five organic acids, only keto-glutaric acid was outstanding during the abiotic test. Of the remaing three however, every single one was excluded from the experiment because of the abiotic test. D-Ribose, one of the remaining three, was found to be annotated in both JGI and the Opol-model. Additional literature for the assimilation capabilities of this metabolite were found as well (Eggeling et al., 1978; Suh et al., 2010). In this literature, D-xylose was also tested, showing a growthrate of 0.07 h^{-1} (Eggeling et al., 1978). D-Xylose was excluded from the analysis because of abiotic interaction between dye and substrate (Addendum 7.2.3). This was also the case for D-ribose. In total, two of three abiotic-active metabolites were placed in this group. 2-Oxoglutarate showed a high difference in 490 nm and 750 nm absorption towards tEnd as well, but did not surpass the threshhold for growth-phenotypes. Without information on a possible utilization, it was hypothesised that 2-oxoglutarate could not be utilized by *O. polymorpha* and was a remnand of the iMT1026-model.

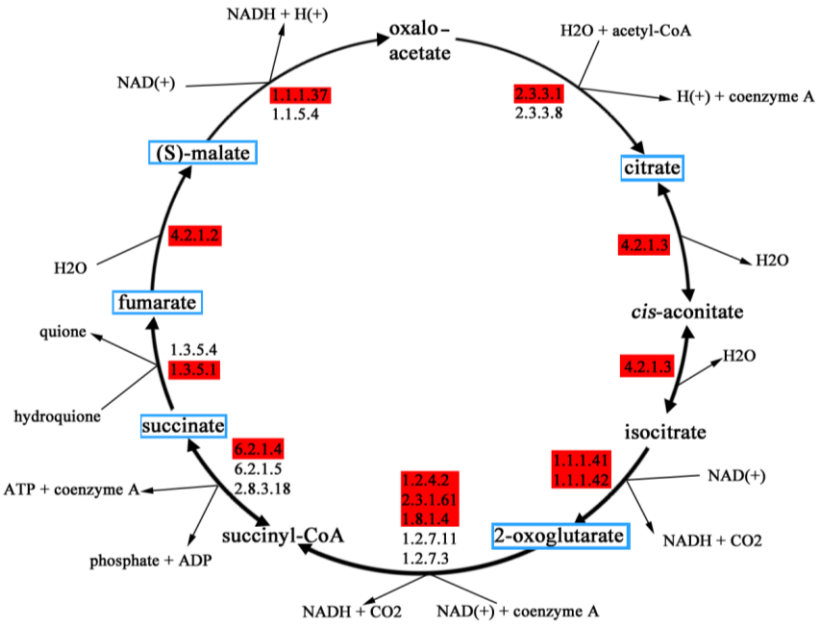


Figure 24: Extract of the TCA cycle. Red EC-numbers mark reactions annotated for *O. polymorpha* NCYC 495 leu1.1 v2.0. Five metabolites are highlighted blue. Those were tested for growth on the PM arrays but did not grow.

3.3.6 Multiple metabolites had either unannotated pathways or no pathways at all

Not all metabolites, that investigated had reactions anotated for *O. polymorpha*. Multiple metabolites were even found to have no connected reactions at all. These metabolites were laminarin, gentiobiose, maltitol, sorbic acid, 3-methylglucose and L-alaninamide. The metabolites that were found to be catalysed by maltase (Chapter 3.3.2) also had no annotations in KEGG, but primary literature supported the experimental growth (Viigand et al., 2016). On the other hand, six metabolites were connected to reactions that did not relate to the *O. polymorpha* genome (Figure 25). For these metabolites ((A) A-allose, (B) D-psicose, (C) erythritol, (D) raffinose, (E) citramalic acid and (F) 3-oxomalate, Figure 25) no primary literature was found. The number of reactions nessesary to test if the yeast is capable of utilizing this methods reach from a single reaction ((B) D-psicose) to four ((C) erythritol). In case of raffinose, the left pathway would leed to melibiose, a metabolite that itsef did not grow on the PM test. For that reason, it is hypothesized that the right pathway is more likely than the left.

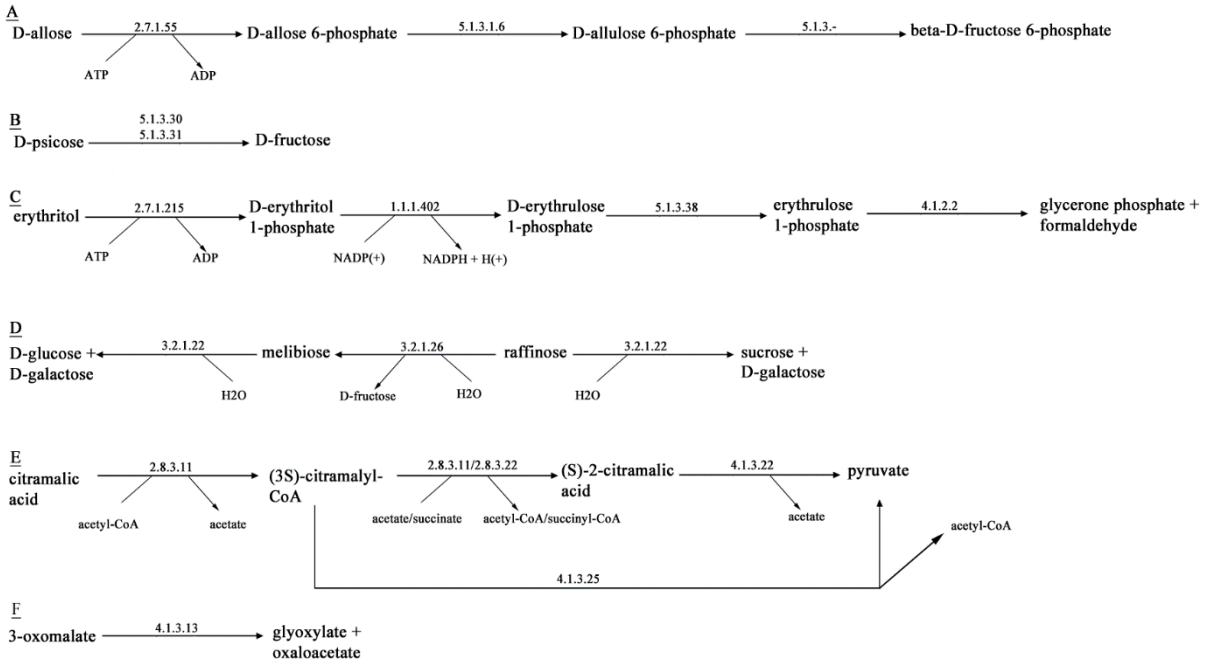


Figure 25: Possible utilization-pathways for metabolites without annotated reactions in JGI. The reactions were taken from KEGG. (A) is a possible pathway for A-allose, (B) for D-psicose, (C) for erythritol, (D) for raffinose, (E) for citramalic acid and (F) for 3-oxomalate.

3.4 Determination of the NGAM

Two runs were performed to establish NGAMs of *O. polymorpha* on glucose and methanol as the sole carbon sources. During both runs, the glucose fermentations had be be aborted after under 150 hours because of contaminations. As a result, not enough steady-state data could be collected for further calculations for the NGAM of glucoseFor methanol, three individual steady-states that were achieved and recorded, running for as much as 240 h (Figure 26). For two of them, triplicates were taken for the calculation of the OD during the steady-state. The third steady-state was measured twice (Table 3). The definition of steady-state was only met for two of the three periods. The third (blue) had to be stopped because of a contamination after 240 hours (Figure 26). During the runs, difficulties in maintaining a constant volume emerged at multiple occasions in all four reactors. Both extremes of levels above and below the pipe-opening of the downstream-pump were observed. A solution was not found during the experiment.

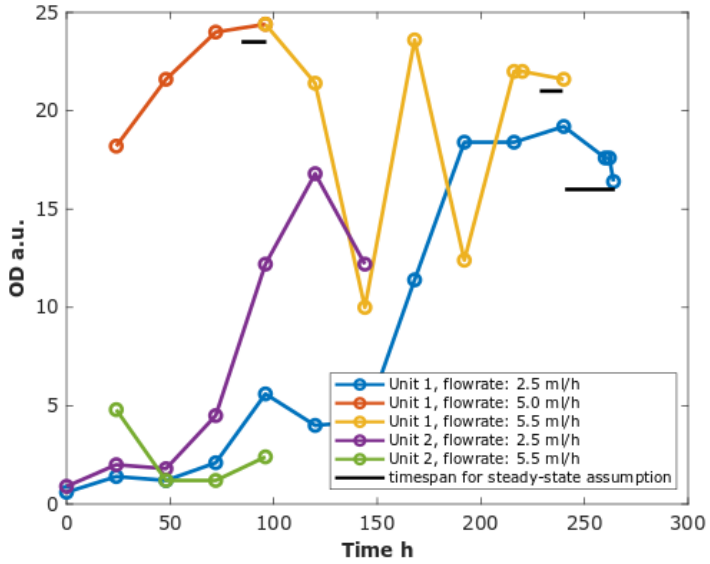


Figure 26: Chemostat methanol-fermentations for *O. polymorpha*. The fermentations were performed in two runs (run 1: blue and violet, run 2: red, yellow and green) in a DASbox reactor. The culutres were kept at 60 ml volume and flowrates as stated in the legend. Samples were taken every 24 h to measure the OD and check for contaminations. The duration nessesary for a steady-state at the specific dilusion-rate is marked by a black bar for the specific fermentation. Three steady-states were achieved during the runs, two of them with triplicate measurement-points (yellow and blue) and one with dublicates (red).

Table 3: ODs of steady-states during the chemostat-cultivations of O. polymorpha on methanol as the sole carbon source.

Flow rate $\left[\frac{ml}{h}\right]$	**ODs [a.u.]**	**Arithmetic mean [a.u.]**	**Total time intervall [h]**
2.5	17.6, 17.6, 16.4	17.2	4
5.0	24.0, 24.4	24.2	24
5.5	22.0, 22.0, 21.6	21.9	24

To calculate the carbon-source uptake during the steady-states, additional information had to be acquired. First, the samples had to be tested for remains of the carbon-source. This was necessary to calculate the actual amount of methanol that was consumed per hour. The samples were analysed in a HPLC (Felix Küttner) and no excess carbon-source was detected. Therefore, the yeasts were able to assimilate all of the 2 % methanol of the feed-medium. Secondly, the correlation between ODs and gCDW needed to be determined. This knowledge was used to calculate the uptake rate of methanol for every steady-state (Table 4). The volume was assumed to be 60 ml for all the reactors, because the shifting levels could not be measured during the

fermentations. Biomass synthesis/ growth yield on the other hand was defined with Formula 3 as the dilution rate of the chemostat.

Figure 27 shows, the experimental uptake against the biomass synthesis along with simulated results using an NGAM of 2.81 $\frac{\text{mmo ATP}}{\text{gCDW·h}}$. The simulation showed a direct linear correlation for the two factors. The linear correlation coefficient for the three experimental measurement points is 0.9243, supporting the linearity of the simulation. Compared to the simulation, the correlation for the experiment had a steeper slope and a higher y-intercept. Simulating the NGAM using y-intercept of the experimental linear correlation and as an uptake rate resulted in a methanol NGAM of 10 ± 8 $\frac{\text{mmol ATP}}{\text{gCDW·h}}$ (Table 5).

Table 4: Calculation of the methanol uptake during the steady-states for three dilution rates. The Volume was defined as 60 ml for each run. Molecular weight of methanol was defined as 32.04 $\frac{\text{mol}}{\text{g}}$. The gCDW arose from experimental determination and was defined to be 0.1807 $\frac{\text{gCDW}}{\text{l} \cdot \text{OD}}$ for *O. polymorpha*.

Flow rate $\left[\frac{\text{ml}}{\text{h}}\right]$	Dilution rate $\left[\frac{1}{\text{h}}\right]$	Uptake $\left[\frac{\text{mmol}}{\text{h}}\right]$	Real uptake $\left[\frac{\text{mmol}}{\text{gCDW·h}}\right]$
2.5	0.042	1.23	6.61
5.0	0.083	2.47	9.40
5.5	0.092	2.71	11.44

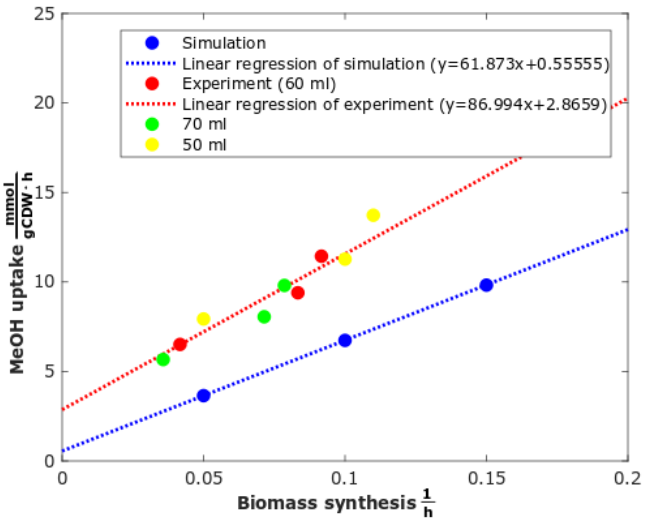


Figure 27: Plot of the methanol uptake against the biomass synthesis of *O. polymorpha* during steady-state in a chemostat fermentation. Besides the experimental data (red), a simulation from the Opol-model was added as a comparison (blue) and alternative measurement points calculated with 70 ml (green) and 50 ml (yellow). A linear regression was calculated for both experimental and simulated data. The y-interception of the linear regression is defined as the amount of carbon source consumed by a culture to maintain itself.

Table 5: NGAMs simulated from the y-intercepts of Figure 27. The MeOH uptake was defined as the y-intercept value, biomass synthesis was set to $0\,\frac{1}{h}$ and the model was tasked to maximise the NGAM for these conditions.

Carbon-Source	**NGAM** $[\frac{\textbf{mmol ATP}}{\textbf{gCDW·h}}]$
Glucose (simulation)	2.80
Methanol (experimental)	10 ± 8

4. Discussion

4.1 Sensitivity Analysis in the unrefined Opol-model

The test for growth behaviour on methanol and glucose showed, that the yield for glucose was higher than for an equivalent amount of methanol. All conditions were kept constant during the simulation, therefore, any difference must arise from the utilization of the two metabolites. As a result, the model predicts a higher utilization cost of methanol in comparison to glucose. This would result in a higher predicted GAM, because less energy can be obtained from the same amount of carbon. A GAM calculated for *P. pastoris* on methanol was 166.77 mmol ATP $\cdot$ $gCDW^{-1} \cdot h^{-1}$ (Tomàs-Gamisans et al., 2017) compared to a value of glucose of 89.69 mmol ATP $\cdot$ $gCDW^{-1} \cdot h^{-1}$ (Tomàs-Gamisans et al., 2016), which would fit the prediction for the *O. polymorpha* GAM.

Four out of five macromolecule groups that are used to describe a growth-phenotype were analysed for their sensitivity. Only lipids showed an outstanding influence on the growth simulation, although the value of the stoichiometric coefficient is much lower compared to carbohydrates and proteins (Table 2). Lipids must bind either more energy or more carbon per unit of macromolecules. Furthermore, not only changes in the biomass composition but changes in the lipid composition can affect simulated growth phenotypes. Thomàs-Gamisans et al. have found no significant changes in the total lipid content of different *P. pastoris* biomass compositions, but changes in the lipid composition itself (Tomàs-Gamisans et al., 2017).

4.2 Biolog Phenotype arrays

The cut-off for the statistical analysis of growth / no growth decisions in the phenotype micro arrays was used to maximize the representation of possible substrate metabolism (minimizing type II error). The cut-off was determined by statistical analysis of low dye-reduction phenotypes, here defined to display values below the arithmetic mean of an array (Chapter 3.2.3). These low dye-reduction wells followed an approximal normal distribution and the 68 % confidence interval was chosen as cut-off for positive substrate usage. If a 95 % confidence interval was chosen instead, ten metabolites more would have been excluded on PM1 and PM2 (Addendum 7.2.7). Of these ten metabolites, one had an *in silico* growth-phenotype and four more

were found in JGI. The remaining five however were difficult to evaluate, because many of them had no connected reactions in KEGG. All metabolites that displayed *in vivo* growth-phenotypes had to be checked for annotations before implementation into the model either way, so the remaining metabolites could represent false-positives or a lack of knowledge for these metabolites.

4.3 Curation of the Opol-model

Explaining the growth of dulcitol and non-growth of galactose is challenging, because the two metabolites are chemically highly similar. First of all, both Jente (2017) and Suh et al. (2010) demonstrated, that galactose cannot be utilized by *O. polymorpha*. Since no further reactions were leading away from galactose that would explain and enable utilization of intracellular galactose, the new reaction was not enough to enable a growth-phenotype for dulcitol. Another pathway is possible to enable dulcitol-utilization, but that would demand further investigation (Chapter 3.3.3, Figure 23). No primary literature could be acquired to support another pathway and therefore, it would have to be investigated *in vivo* by first confirming the utilization of dulcitol in a shake flask cultivation and later identifying degradation metabolites by GCMS. The same method would be used for other new metabolites as well (Figure 25). Additionally, a possible way to utilize dulcitol would also enable the utilization of intracellular galactose, a factor that would change the predicted growth-phenotypes for other, galactose-producing metabolites.

More results of the PM experiment can be compared to published data on assimilation capabilities of *O. polymorpha* (Suh et al., 2010). All in all, most phenotypes were shared, but the differences were very relevant for the conditions of the PM test (Table 6). Three metabolites were on this list, that showed growth-phenotypes but were excluded because of abiotic activity. Of these three, two were found in JGI (D-ribose and D-xylitol) and one was not tested, because it was part of the model, but did not allow a growth-simulation (D-arabinose). All three however would have been considered as positive growth-phenotypes, so they do not in fact disagree with the literature. However, two other metabolites did show differences. L-Lactate was part of the model and it would have been cut off by a 95 % confidence interval during the statistical analysis (Addendum 7.2.7). The other, raffinose, grew after 24 h on one and after 72 h on another array and had no annotated reactions in JGI. Two metabolites were excluded due to strong and temporally separated OD-changes that were interpreted as contaminations

(Addendum 7.2.4). The OD-changes for raffinose were more in line with other metabolites, but the behaviour could be another, weaker contamination. Altogether, the comparison demonstrates a good reflection on the PM-results. On the one hand, L-lactate can be considered as a sign, that the cut off might be too low, but on the other hand it was linked to annotated enzymes in JGI and could have been cut wrongfully.

Table 6: Comparison of growth-phenotypes found on the PM-arrays and assimilations found by (Suh et al., 2010).

Metabolite	Assimilation (Suh et al., 2010)	PM-experiment
Sucrose	+	+
D-Arabinose	+	- / abiotic
L-Arabinitol	+	+
L-Lactate	-	+
Succinate	+	+
D-Ribose	+	- / abiotic
D-Xylose	+	- / abiotic
Maltose	+	+
Melibiose	-	-
Melezitose	+	+
Erythritol	+	+
Xylitol	+	+
Citrate	-	-
Raffinose	-	+

The maltase of *O. polymorpha* is a sign, that enzymes may catalyse reactions for more substrates than a general database, like KEGG, accounts for. Multiple of the metabolites, turanose, maltotriose, melezitose and palatinose, would not have been annotated otherwise. This leads to two assumptions pointing back to the evaluations: First, positive metabolites on the PM test without annotated reactions might be degraded by other enzymes, that are specific for similar molecules. Second, the substrate-specificity of homologous enzymes can differ from one organism to another, even as closely related as *P. pastoris* and *O. polymorpha* (Viigand et al., 2016). With that in mind, if multiple enzymes have the same EC-number and are active in the same compartment, they could potentially catalyse other reactions as well. In case of 4 -hydroxybutanoic acid, one reaction was nessesary to connect it to the model. The EC-number of the enzyme was found in JGI but only for another enzyme. The assumption was made that either this enzyme is capable of utilizing 4 -hydroxybutanoic acid as a unspecific

substrate or the actual enzyme was not annotated. This needs to be further tested to evaluate the hypothesis.

Next, the non-growth of important metabolites of the TCA-cycle under aerobic conditions can be used to determine further weaknesses of the experiment. On the one hand, the necessity of a validated mean of transport was not an essential part of this thesis. Information about transporters were not available for the most metabolites and not part of the *O. polymorpha* database. If a growth-phenotype was observed, it was assumed that there must be a mean of transportation. On the other hand, metabolites without growth-phenotypes but with annotated reactions could mean, that there was no transporter. In addition, the outcome of an experiment is strongly depended on the conditions. For example: succinate, fumarate and malate did not grow on PM1. However, experiments show that *O. polymorpha* can grow on all three of those metabolites (Addendum 7.4) and for *E. coli*, it was demonstrated that utilization of succinate was dependent on whether anaerobic conditions were met or not (Pos et al., 1998; Reynolds et al., 1983). As a result, the conditions on the arrays may affect the ability of the yeast to utilize tested metabolites. One of those conditions is the cultivation time. *O. polymorpha* was observed to have long lag-phases on malate and even longer on fumarate (Addendum 7.4). The cultivation time during the PM-experiment was not long enough to detect this behaviour. PM arrays can be used for a wide variety of cultures, from bacteria to tissue, meaning the pH should be around pH 7 because of the varying demands. This would also indicate towards longer lag-phases for succinate, further supporting the theory (Addendum 7.4). In summary, many factors must be considered when deciding whether an observed non-growth under this experimental condition is a reason to deduce on the metabolism of the yeast.

In total, 147 out of 190 metabolites were found to be predicted correctly by the Opol-model. Of the remaining 43 metabolites, 16 were corrected in the model to fit the experimental-phenotypes. Seven out of eight in silico-only metabolites were assumed to be false-negatives due to the experimental conditions or abiotic-activity exclusion (D-ribose and D-xylose). All of them combined are 170 out of 190 correct predictions for the updated model. Of the remaining 20 metabolites, seven were also found to have either one or more reactions included in the model but could not be simulated. Only twelve in total had no connections in the JGI database or primary literature. Five out of ten metabolites that would have been cut off with a 95 % confidence interval were part of these twelve metabolites. This could mean that the cut off was low enough to allow few false-positives and not to cut many positives as false negatives. Since 50 % of the questionable metabolites did have annotations, the number of false-negatives

would have increased with the 95 % confidence interval. However, the information in ChEBI and KEGG for the unannotated metabolites show that not all of them are relevant for the usefulness of the Opol-model. After all, the PM-arrays were made not solely for yeasts.

Having said that, it would imply that only one metabolite, 2-hydroxybutyric acid, was usable to simulate growth but had no *in vivo* growth-phenotype and was not found in JGI or primary literature, too. One single metabolite that could be excluded from the GEM, because only *P. pastoris* could utilize it, while multiple additions to the model were be made. Looking at the genomes of both yeasts, neither the size nor the amount of predicted protein-coding genes was reflecting that difference. The *P. pastoris* genome is around 8.35 MPB long and has up to 5313 protein-coding genes (De Schutter et al., 2009; Kuberl et al., 2011). In case of *O. polymorpha*, the numbers are 8.97 MPB and 5325 genes (Ravin et al., 2013; Riley et al., 2016). This does not correlate with the growth-phenotypes and emphasises the peculiarity of the situation.

4.4 NGAM of *O. polymorpha*

During the fermentations, multiple challenges occurred that lowered the chances of promising results. One major limitation was the bacterial contamination that occurred at the latest after 120 h in the glucose-limited fermentations and after 264 h in the methanol-limited fermentations (Figure 27, Addendum 7.4.1). However, it was possible to run the chemostat for 250 h and collect steady-states for two different dilutions. A third run with the experience of the earlier fermentations was not possible because of time-limitations. Another factor that interfered with the result generation was the inability of the system to maintain a constant volume. Both higher and lower culture levels were observed during the run, although the downstream pump should have removed excess medium. Controlling the culture-volume with a sensor would have been an option, but all available instrument-slots were already occupied with other essential tools (Figure 6). With that in mind, the volume needed to calculate the y-axis interception could have fluctuated (Figure 27). The changes ± 10 ml made to the calculations would have changed the NGAM, but all values were still far above the *P. pastoris* maintenance uptake and therefore, the glucose NGAM of 2.81 $\frac{\text{mmol ATP}}{\text{gCDW·h}}$ (Tomàs-Gamisans et al., 2016). Because it was impossible to generate duplicates for the NGAM measurements, only an assumption can be made, that the maintenance cost of *O. polymorpha* on methanol-limited growth conditions is higher, than that of glucose-limited *P. pastoris*. This would correlate with the predictions

made from the growth-simulations (Chapters 3.1.1 and 4.1). However, Thomàs-Gamisans also demonstrated, that the methanol-NGAM of iMT2026 is 0.44 $\frac{\text{mmol ATP}}{\text{gCDW}\cdot\text{h}}$, a value far lower than that of the glucose-NGAM (Tomàs-Gamisans et al., 2017).

The Opol-model is still a draft version of this exact model and uses the same utilization path that methylotrophic yeasts have in common. If the predicted NGAM of *O. polymorpha* lies higher than on glucose, it is probably a result of the very high maintenance uptake in the experiment. In addition, the slope in Figure 27 represents the carbon-source requirement for growth and can therefore be used to predict the GAM of *O. polymorpha* in relationship to *P. pastoris*. The slope indicates towards a higher growth-related maintenance on methanol. This would correlate with the methanol-GAM of *P. pastoris* that is higher on methanol than on glucose (Tomàs-Gamisans et al., 2017).

5. Conclusion & Outlook

The aim of this thesis was to improve the Opol-reconstruction draft. During the refinement, phenotype experiments were conducted and a variety of different carbon-sources to simulate growth-phenotypes were added. The updated model consists of 1752 metabolites and 2275 reactions and correctly predicted 170 out of 190 *in vivo* growth-phenotypes. However, the knowledge gathered in databases for the yeast did not account for all the metabolites that were tested. Multiple metabolites must be tested individually to gather more information if *O. polymorpha* is truly capable of utilizing them. Furthermore, most GPRs must be updated, because they originate from the *P. pastoris* model iMT1026. A parallel examination of nitrogen-sources was not realizable with PM-arrays, which keeps this task open for later work.

The fermentations to obtain non-growth associated maintenance reactions for the improvement of simulated growth-phenotypes on glucose and methanol did not result in statistically convincing NGAMs. However, the obtained steady-states already led to the calculation of a maintenance-cost at zero growth. In the future, more data will decrease the impact of deviations during the fermentation and allow for a better prediction of the methanol and glucose NGAMS. With that in mind, the prediction of *O. polymorpha* specific GAM and biomass-compositions are the next step to improve growth-simulations with the Opol-model.

6. References

Bochner, B.R., Gadzinski, P., Panomitros, E., (2001) Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome research 11, 1246-1255.

Burgard, A.P., Pharkya, P., Maranas, C.D., (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnology and bioengineering 84, 647-657.

Caspi, R., Foerster, H., Fulcher, C.A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., et al., (2007) The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic acids research 36, D623-D631.

De Schutter, K., Lin, Y., Tiels, P., Van Hecke, A., Glinka, S., Rouze, P., Van de Peer, Y., et al., (2009) Genome sequence of the recombinant protein production host Pichia pastoris. Nature biotechnology 27.

Eggeling, L., Sahm, H., (1978) Derepression and partial insensitivity to carbon catabolite repression of the methanol dissimilating enzymes inHansenula polymorpha. European journal of applied microbiology and biotechnology 5, 197-202.

Feist, A.M., Palsson, B.O., (2010) The biomass objective function. Current opinion microbiology 13.

Gurobi Optimization, I., (2016) Gurobi Optimizer Reference Manual, http://www.gurobi.com, 07.02.2018

Hamilton, J.J., Reed, J.L., (2014) Software platforms to facilitate reconstructing genome-scale metabolic networks. Environmental microbiology 16, 49-59.

Hastings, J., de Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., et al., (2013) The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Research 41, D456-D463.

Jente, L., (2017) Comparing O. polymorpha and P. pastoris and constructing a genome-scale metabolic model. RWTH-Aachen, Bachlor-Thesis

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K., (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research 45, D353-D361.

Kanehisa, M., Goto, S., (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 28, 27-30.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research 44, D457-D462.

Kuberl, A., Schneider, J., Thallinger, G.G., Anderl, I., Wibberg, D., Hajek, T., Jaenicke, S., et al., (2011) High-quality genome sequence of Pichia pastoris CBS7435. J Biotechnol 154.

Liao, Y.-C., Huang, T.-W., Chen, F.-C., Charusanti, P., Hong, J.S., Chang, H.-Y., Tsai, S.-F., et al., (2011) An experimentally validated genome-scale metabolic reconstruction of Klebsiella pneumoniae MGH 78578, iYL1228. Journal of bacteriology 193, 1710-1717.

Lopes, H., Rocha, I., (2017) Genome-scale modeling of yeast: chronology, applications and critical perspectives. FEMS Yeast Research 17, fox050-fox050.

Nobelmann, B., Lengeler, J.W., (1996) Molecular analysis of the gat genes from Escherichia coli and of their roles in galactitol transport and metabolism. Journal of bacteriology 178, 6790-6795.

Novick, A., Szilard, L., (1950) Description of the chemostat. Science 112, 715-716.

Oberhardt, M.A., Palsson, B.Ø., Papin, J.A., (2009) Applications of genome-scale metabolic reconstructions. Molecular systems biology 5, 320.

Olah, G.A., (2005) Beyond oil and gas: the methanol economy. Angewandte Chemie International Edition 44, 2636-2639.

Orth, J.D., Thiele, I., Palsson, B.Ø., (2010) What is flux balance analysis? Nature biotechnology 28, 245-248.

Pos, K.M., Dimroth, P., Bott, M., (1998) The Escherichia coli citrate carrier CitT: a member of a novel eubacterial transporter family related to the 2-oxoglutarate/malate translocator from spinach chloroplasts. Journal of bacteriology 180, 4160-4165.

Ramezani-Rad, M., Hollenberg, C.P., Lauber, J., Wedler, H., Griess, E., Wagner, C., Albermann, K., et al., (2003) The Hansenula polymorpha (strain CBS4732) genome sequencing and analysis. FEMS yeast research 4, 207-215.

Ravin, N.V., Eldarov, M.A., Kadnikov, V.V., Beletsky, A.V., Schneider, J., Mardanova, E.S., Smekalova, E.M., et al., (2013) Genome sequence and analysis of methylotrophic yeast Hansenula polymorpha DL1. BMC genomics 14, 837.

Reynolds, C.H., Silver, S., (1983) Citrate utilization by Escherichia coli: plasmid-and chromosome-encoded systems. Journal of bacteriology 156, 1019-1024.

Riley, R., Haridas, S., Wolfe, K.H., Lopes, M.R., Hittinger, C.T., Göker, M., Salamov, A.A., et al., (2016) Comparative genomics of biotechnologically important yeasts. Proceedings of the National Academy of Sciences, 201603941.

Ryabova, O.B., Chmil, O.M., Sibirny, A.A., (2003) Xylose and cellobiose fermentation to ethanol by the thermotolerant methylotrophic yeast Hansenula polymorpha. FEMS yeast research 4, 157-164.

Schellenberger, J., Que, R., Fleming, R.M., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., et al., (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. Nature protocols 6, 1290-1307.

Suh, S.-O., Zhou, J.J., (2010) Methylotrophic yeasts near Ogataea (Hansenula) polymorpha: a proposal of Ogataea angusta comb. nov. and Candida parapolymorpha sp. nov. FEMS yeast research 10, 631-638.

Thiele, I., Palsson, B.O., (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature protocols 5.

Thompson, R.A., Dahal, S., Garcia, S., Nookaew, I., Trinh, C.T., (2016) Exploring complex cellular phenotypes and model-guided strain design with a novel genome-scale metabolic model of Clostridium thermocellum DSM 1313 implementing an adjustable cellulosome. Biotechnology for biofuels 9, 194.

Tomàs-Gamisans, M., Ferrer, P., Albiol, J., (2016) Integration and validation of the genome-scale metabolic models of Pichia pastoris: a comprehensive update of protein glycosylation pathways, lipid and energy metabolism. PloS one 11, e0148031.

Tomàs-Gamisans, M., Ferrer, P., Albiol, J., (2017) Fine-tuning the P. pastoris iMT1026 genome-scale metabolic model for improved prediction of growth on methanol or glycerol as sole carbon sources. Microbial biotechnology.

Van Dijk, R., Faber, K.N., Kiel, J.A., Veenhuis, M., van der Klei, I., (2000) The methylotrophic yeast Hansenula polymorpha: a versatile cell factory. Enzyme and microbial technology 26, 793-800.

Viigand, K., Visnapuu, T., Mardo, K., Aasamets, A., Alamäe, T., (2016) Maltase protein of Ogataea (Hansenula) polymorpha is a counterpart to the resurrected ancestor protein ancMALS of yeast maltases and isomaltases. Yeast 33, 415-432.

Wichelecki, D.J., Vetting, M.W., Chou, L., Al-Obaidi, N., Bouvier, J.T., Almo, S.C., Gerlt, J.A., (2015) ATP-binding cassette (ABC) transport system solute-binding protein-guided identification of novel d-altritol and galactitol catabolic pathways in Agrobacterium tumefaciens C58. Journal of Biological Chemistry 290, 28963-28976.

Yurimoto, H., Oku, M., Sakai, Y., (2011) Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis. International journal of microbiology 2011.

Ziv, N., Brandt, N.J., Gresham, D., (2013) The use of chemostats in microbial systems biology. Journal of visualized experiments: JoVE.

7. Addendum

7.1 Sensitivity analysis

7.1.1 Calculations for uptake changes

Biomass synthesis [h^{-1}]	Methanol uptake [$\frac{mmol}{gCDW \cdot h}$]	Glucose uptake [$\frac{mmol}{gCDW \cdot h}$]	Glc/MeOH
0.05	3.65	0.57	0.16
0.10	6.74	1.05	0.16
0.15	9.82	1.52	0.16
0.20	12.91	2.00	0.15
0.25	16.00	2.48	0.15
0.40	19.09	2.96	0.15
0.35	22.17	3.43	0.15
0.40	25.26	3.90	0.15
0.45	28.35	4.38	0.15
0.50	31.44	4.86	0.15

7.2 PM analysis

7.2.1 PM3 negative controls

Time [h]	PM3.1 ODs [a.u.] (490 nm/750 nm)	PM3.2 ODs [a.u.] (490 nm/750 nm)	PM3.abiotic ODs [a.u.] (490 nm/750 nm)
0	0.115 0.087	-	0.112 0.088
24	0.497 0.341	0.47 0.319	-
48	0.554 0.395	0.524 0.37	-
52	0.552 0.393	0.532 0.38	-
72	0.576 0.413	0.581 0.423	0.065 0.044

7.2.2 Growth without additional nitrogen-source

Inoculation from master-plate. 5 ml IFY-0 medium with leucin and glucose. 24 h cultivation at 37 °C.

+ Dye		- Dye	
1 [a.u.]	2 [a.u.]	1 [a.u.]	2 [a.u.]
1.6	1.5	1.3	1.2

7.2.3 Excluded substrates (abiotic activity)

For most metabolites, a correlation between 490 nm and 750 nm was observable. tEnd was 120 hours for PM1 and 72 hours for PM2. Not all metabolites in this list were excluded from further analysis.

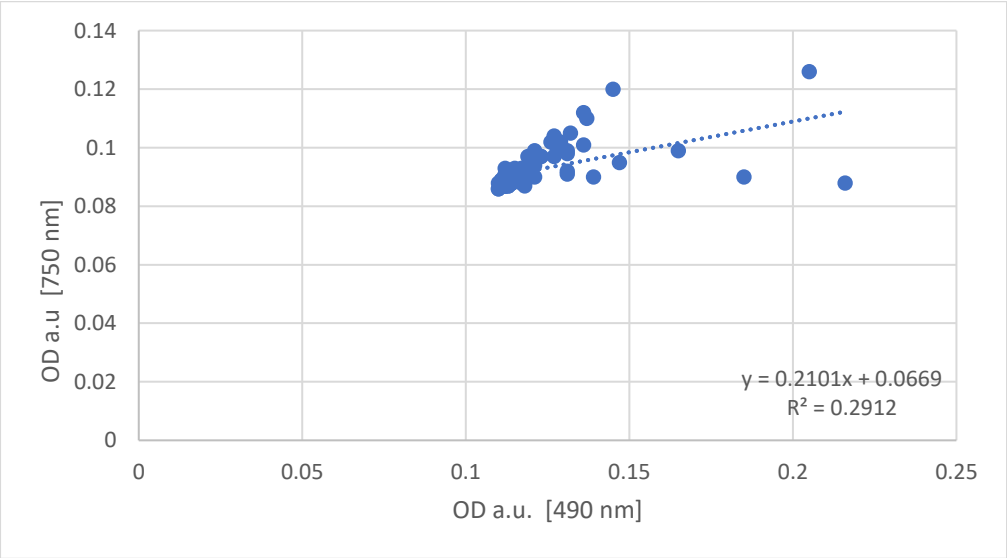
Metabolite	Name	ODs t0 [a.u.] (490 nm/750 nm)	ODs tEnd [a.u.] (490 nm/750 nm)	Surpassed threshhold	Included in model
PM1 – A2	L-Arabinose	0.17 0.089	0.276 0.14	-	+
PM1 – B8	D-Xylose	0.197 0.088	0.477 0.091	+	+
PM1 – C4	D-Ribose	0.252 0.092	0.735 0.252	+	+

PM1 – C7	D-Fructose	0.128 0.097	0.205 0.109	+	+
PM1 – D6	a-Ketoglutaric acid	0.116 0.09	0.198 0.108	-	+
PM1 – F10	Glyoxylic acid	0.113 0.089	0.683 0.32	-	+
PM1 – H6	L-Lyxose	0.197 0.088	0.619 0.242	+	-
PM2 – A9	Inulin	0.136 0.112	0.131 0.064	-	-
PM2 – A12	Pectin	0.205 0.126	0.177 0.069	-	+
PM2 – B5	D-Arabinose	0.185 0.09	0.418 0.205	+	+
PM2 – B9	Desoxy-2-ribose	0.139 0.09	0.341 0.166	+	+
PM2 – E5	D-Glucosamine	0.165 0.099	0.43 0.181	+	-
PM2 – E12	5-keto-D-Gluconic acid	0.147 0.095	0.463 0.19	+	+
PM2 – F9	Sorbic acid	0.131 0.092	0.392 0.239	-	-
PM2 – H9	Dihydroxyacetone	0.216 0.088	0.657 0.257	+	+

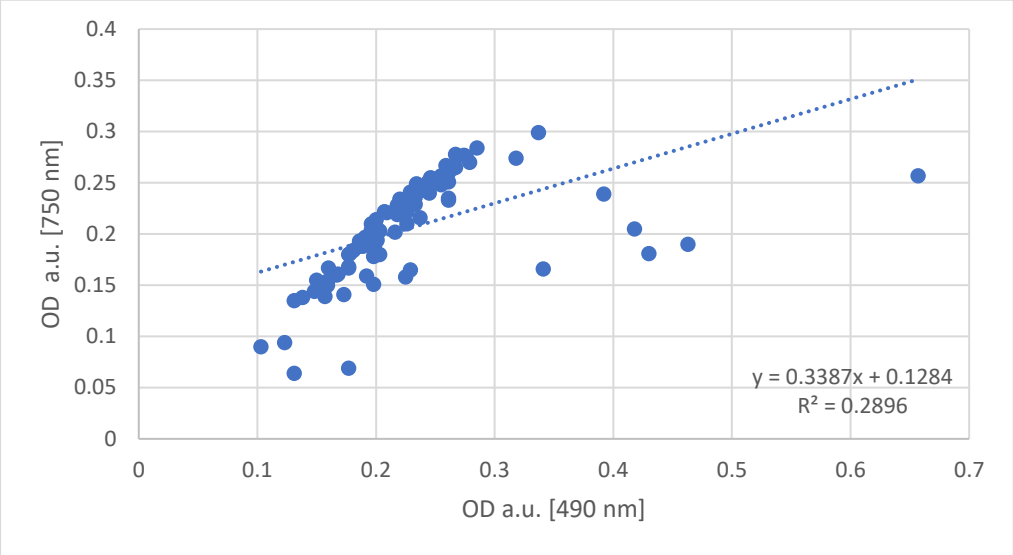
7.2.4 Excluded substrates (contamination)

Metabolite	Name	Reason
PM2 – E8	b-Hydroxybutyric acid	Growth on 2.3 after 48 h, and strong growth on 2.1 after 72 h
PM2 – H3	L-Pyroglutamic acid	Strong growth on 2.2 after 72 h and on 2.3 after 48 h

7.2.5 PM2 abiotic correlation (t0)



7.2.6 PM2 abiotic correlation (tEnd)



7.2.7 Metabolites additionally excluded in a 95 % confidence interval

Metabolite	Name	Annotated Reactions
PM1 – B07	Glycerol phosphate	+
PM1 – B09	L-Lactic acid	+
PM1 – H05	D-Psicose	-
PM2 – B03	b-D-Allose	-
PM2 – B11	D-Fucose	-
PM2 – C01	Gentiobiose	-
PM2 – C08	3-Methylglucose	-
PM2 – G04	L-Arginine	+
PM2 – H01	Ornithine	+
PM2 – H4	L-Valine	+

7.2.8 List of all metabolites that were actively compaired

PM1-Metabollites	growth in silico	growth in experiment	growthratio on arrays
L-Proline	-	+	2/2
Dulcitol	-	+	1/2
DL-a-Glycerol Phosphate	-	+	2/2
L-Lactic acid	-	+	2/2
L-Glutamic acid	-	+	2/2
Maltose	-	+	2/2
L-Asparagine	-	+	2/2
Sucrose	-	+	2/2
L-Glutamine	-	+	2/2
Adonitol	-	+	2/2
Maltotriose	-	+	2/2
Inosine	-	+	2/2
D-Psicose	-	+	2/2

Succinic acid	+	-	0/2
D-Xylose	+	-	2/2
D-Ribose	+	-	2/2
a-Ketoglutaric acid	+	-	2/2
a-Hydroxybutyric acid	+	-	0/2
Citric acid	+	-	0/2
Fumaric acid	+	-	0/2
L-Malic acid	+	-	0/2
D-Trehalose	+	+	2/2
D-Mannose	+	+	2/2
D-Sorbitol	+	+	2/2
Glycerol	+	+	2/2
D-Mannitol	+	+	2/2
L-Rhamnose	+	+	2/2
D-Fructose	+	+	2/2
Acetic acid	+	+	2/2
a-D-Glucose	+	+	2/2

PM2- Metabolites	**growth in silico**	**growth in experiment**	**growthratio on arrays**
Laminarin	-	+	3/3
b-D-Allose	-	+	2/3
Amygdalin	-	+	3/3
D-Arabitol	-	+	3/3
L-Arabitol	-	+	3/3
i-Erythritol	-	+	3/3
D-Fucose	-	+	2/3
Gentiobiose	-	+	2/3
D-Melezitose	-	+	3/3
Maltitol	-	+	3/3
3-Methylglucose	-	+	2/3
Palatinose	-	+	3/3
D-Raffinose	-	+	2/3
Turanose	-	+	3/3
Citramalic acid	-	+	3/3
g-Hydroxybutyric acid	-	+	3/3
Oxalomalic acid	-	+	2/3
Sorbic acid	-	+	3/3
L-Alaninamide	-	+	3/3
L-Arginine	-	+	3/3
L-Leucine	-	+	3/3
L-Ornithine	+	+	2/3
L-Valine	-	+	2/3
Dihydroxyacetone	-	+	3/3
Xylitol	+	+	3/3
L-Isoleucine	+	+	3/3

7.3 Additions to the model

7.3.1 List of all implemented reactions, transporters and exchange reactions

Reaction Name	EC-number	GPR
ATP:glycerone phosphotransferase	2.7.1.29	e_gw1.6.200.1
(R)-Prunasin beta-D-glucohydrolase	3.2.1.21	
Amygdalin beta-glucosidase	3.2.1.21	fgenesh1_pg.3_#_632
Amygdalin transporter		
Amygdalin exchange		
Inosine ribohydrolase	3.2.2.1	fgenesh1_kg.2_#_370_#_isotig03402
Ribitol:NAD+ 2-oxidoreductase	1.1.1.56	estExt_Genewise1Plus2.C_2_t10234
Adonitol transporter		
Adonitol exchange		
Sucrose glucohydrolase	3.2.1.20	fgenesh1_kg.1_#_333_#_isotig01325
Maltose glucohydrolase	3.2.1.20	fgenesh1_kg.1_#_333_#_isotig01325
Dulcitol transporter		
Dulcitol exchange		
Galactitol:NADP+ 1-oxidoreductase	1.1.1.21	fgenesh1_kg.1_#_232_#_isotig00929
Galactitol:NAD+ 1-oxidoreductase	1.1.1.21	fgenesh1_kg.1_#_232_#_isotig00929
L-Arabinose isomerase	5.3.1.4	fgenesh1_kg.2_#_369_#_isotig04558
Proline iminopeptidase	3.4.11.5	fgenesh1_kg.3_#_2_#_isotig02527
Uridine kinase [UTP, ITP, dATP, dGTP, dTTP, dCTP, dUTP]	2.7.1.48	fgenesh1_kg.5_#_229_#_isotig02607
Maltotriose Exchange		
Maltotriose Transporter		
alpha-Glucosidase [maltotriose]	3.2.1.20	fgenesh1_kg.1_#_333_#_isotig01325
D-Melezitose Exchange		
D-Melezitose Transporter		
alpha-Glucosidase [melezitose]	3.2.1.20	fgenesh1_kg.1_#_333_#_isotig01325
Palatinose Exchange		
Palatinose transporter		
alpha-Glucosidase [palatinose]	3.2.1.20	fgenesh1_kg.1_#_333_#_isotig01325
Turanose Exchange		
Turanose transporter		
alpha-Glucosidase [turanose]	3.2.1.20	fgenesh1_kg.1_#_333_#_isotig01325
g-Hydroxybutyric acid Exchange		
g-Hydroxybutyric acid Transporter		
unspecific Dehydrogenase	1.1.1.-	fgenesh1_kg.5_#_332_#_isotig04945
Succinate-semialdehyde dehydrogenase [NAD+]	1.2.1.16	fgenesh1_kg.4_#_229_#_isotig04080
Glycerol 3-phosphate Exchange		
Glycerol 3-phosphate Transporter		
Dihydroxyacetone Exchange		
Dihydroxyacetone Transporter		
L-Arabitol:NAD+ 1-oxidoreductase	1.1.1.21	fgenesh1_kg.1_#_232_#_isotig00929
Mandelonitril Exchange		
Mandelonitril transport		

7.3.2 List of all added metabolites

Metabolite	Neutral formula	charge	compartment	KEGG ID	ChEBI ID
Dulcitol	C6H14O6	0	c	C01697	chebi:16813
Dulcitol	C6H14O6	0	e	C01697	chebi:16813
Adonitol	C5H12O5	0	c	C00474	chebi:15963
Adonitol	C5H12O5	0	e	C00474	chebi:15963
Amygdalin	C20H27NO11	0	e	C08325	chebi:17019
Amygdalin	C20H27NO11	0	c	C08325	chebi:17019
Mandelonitrile	C8H7NO		c	C00561	
Prunasin	C14H17NO6	0	c	C00844	chebi:17396
Dihydroxyacetone	C3H6O3	0	e	C00184	chebi:16016
Maltotriose	C18H32O16	0	c	C01835	chebi:27931
Maltotriose	C18H32O16	0	e	C01835	chebi:27931
D-Melezitose	C18H32O16	0	c	C08243	chebi:6731
D-Melezitose	C18H32O16	0	e	C08243	chebi:6731
Palatinose	C12H22O11	0	c	C01742	chebi:18394
Palatinose	C12H22O11	0	e	C01742	chebi:18394
Turanose			c	G03588	
Turanose			e	G03588	
g-Hydroxybutyric acid	C4H8O3	0	c	C00989	chebi:30830
g-Hydroxybutyric acid	C4H8O3	0	e	C00989	
Glycerol 3-phosphate	C3H9O6P	-2	e	C00093	chebi:57597
Mandelonitrile	C8H7NO		e	C00561	

7.3.3 List of all actively compared metabolites and the results of the refinements on *in silico* growth-phenotypes

Substrate	*in silico*	experimental	Corrected phenotypes	Found Sources/ Updates
L-Proline	-	+	+	JGI
Dulcitol	-	+	-	JGI
DL-a-Glycerol Phosphate	-	+	+	Exchange reaction
L-Lactic acid	+	+		(Suh et al., 2010)
L-Glutamic acid	-	+	+	model
Maltose	-	+	+	JGI and (Viigand et al., 2016)
L-Asparagine	-	+	+	model
Sucrose	-	+	+	JGI and (Viigand et al., 2016)
L-Glutamine	-	+	+	exchange reaction
Adonitol	-	+	+	JGI
Maltotriose	-	+	+	(Viigand et al., 2016)
Inosine	-	+	+	JGI
D-Psicose	-	+	-	
Succinic acid	+	-	+	JGI
D-Xylose	+	-	-	JGI
D-Ribose	+	-	+	JGI and (Suh et al., 2010)
a-Ketoglutaric acid	+	-	+	JGI
a-Hydroxybutyric acid	+	-	-	
Citric acid	+	-	+	JGI

Fumaric acid	+	-	+	JGI
L-Malic acid	+	-	+	JGI
D-Trehalose	+	+		
D-Mannose	+	+		
D-Sorbitol	+	+		
Glycerol	+	+		
D-Mannitol	+	+		
L-Rhamnose	+	+		
D-Fructose	+	+		
Acetic acid	+	+		
a-D-Glucose	+	+		
Laminarin	-	+	-	
b-D-Allose	-	+	-	
Amygdalin	-	+	-	JGI
D-Arabitol	-	+	+	
L-Arabitol	-	+	-	
i-Erythritol	-	+	-	(Suh et al., 2010)
D-Fucose	-	+	-	
Gentiobiose	-	+	-	
D-Melezitose	-	+	+	(Viigand et al., 2016)
Maltitol	-	+	-	
Palatinose	-	+	+	(Viigand et al., 2016)
Turanose	-	+	+	(Viigand et al., 2016)
Citramalic acid	-	+	-	
g-Hydroxybutyric acid	-	+	-	
Oxalomalic acid	-	+	-	
Sorbic acid	-	+	-	
L-Alaninamide	-	+	-	
L-Arginine	-	+	+	
L-Leucine	-	+	-	
L-Valine	-	+	-	
Dihydroxyacetone	-	+	+	
Xylitol	+	+		(Suh et al., 2010)
L-Isoleucine	+	+		
L-Ornithine	+	+		
D-Raffinose	-	+	-	(Suh et al., 2010)
3-Methylglucose	-	+	-	

7.4 Growth-experiments in collaboration with Felix Küttner

7.4.1 Growth of H.p. dyku80 in Yan et al. @200rpm/37°C/CGQ for 168h on Succinate salt as c-source.

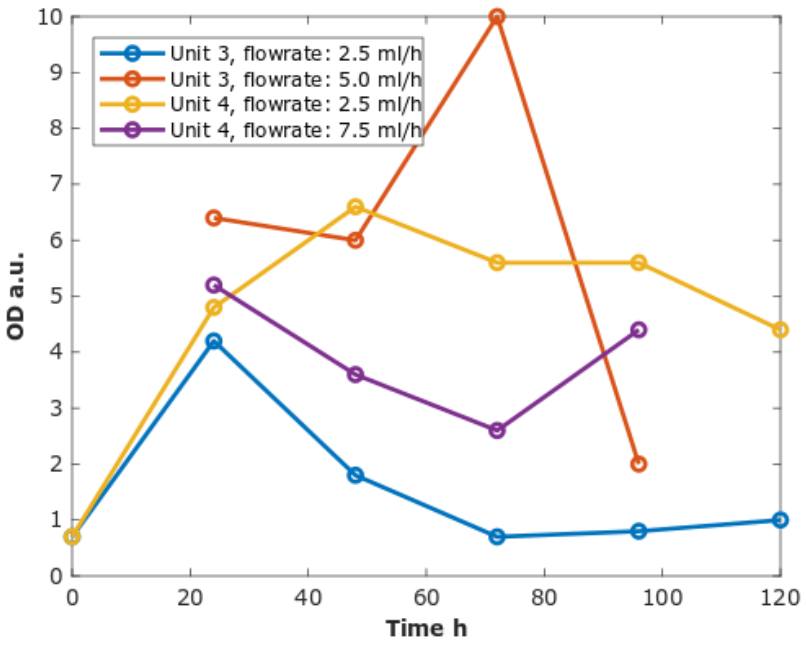
Strain	stress additive (Succinate [g/L])	used Medium	Final OD	Start pH	End pH	Start exp.	µ max [1/h]
Δ yku80	180 mM Succinate	Yan et. al	1 ± 0	6.46	6.70	---	---
Δ yku80	180 mM Succinate pH5	Yan et. al	2.1 ± 0.1	5.27	5.48	79	0.02
Δ yku80	180 mM Succinic acid	Yan et. al	3,6 ± 0	2.93	3.66	17	0.01

7.4.2 Growth of H.p. dyku80 in Yan et al. @200 rpm, 37°C cultivation for 144h on180 mM Malate and fumarate as sole carbon source.

Strain	c-source	OD 0h	pH 0h	OD	Devia-tion	pH 270h	Devia-tion	exp. Phase start [h]	µ max 1/h	Devia-tion
dyku80	180 mM Mal	1.2	5.37	2.4	0.1	6.33	0.02	49	0.010	0.004
dyku80	180 mM Fum	1.1	5.95	2.5	0.2	7.31	0.02	90	0.014	0.002

7.5 Determination of the NGAM

7.5.1 Chemostat methanol-fermentations for *O. polymorpha*



Acknowledgements

I would like to thank:

Prof. Dr. Lars M. Blank for being the primary examiner of this thesis and for giving me the opportunity to perform my bachelor thesis at his institute.

Dr. Martin Zimmermann for being the secondary examiner of this thesis.

My supervisor, Dr. Ulf Liebal, for his engagement support and advice. Thank you for introducing me into system biology and modelling. You were always there if I needed help or to discuss a topic, and your constructive feedback has always encouraged me to improve.

Felix Küttner for the great time I had working with and learning from him.

Finally, I would like to thank all my friends and family for their support.