CHALMERS
UNIVERSITY OF TECHNOLOGY

A kinetically-constrained FBA-model of the synthesis of aromatic amino acid-derived products in *Saccharomyces cerevisiae*

Master Of Science Thesis in Biotechnology

Markus Janasch

Department for Biology and Biological Engineering
Systems and Synthetic Biology Group
CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2015

A kinetically-constrained FBA-model of the synthesis of aromatic amino acid-derived products in *Saccharomyces cerevisiae*

Master Of Science Thesis in Biotechnology

Supervisor/Examiner: Jens Nielsen

Department for Biology and Biological Engineering
Systems and Synthetic Biology Group
Chalmers University Of Technology
SE-412 96 Göteborg
Sweden
Telephone: +46(0)31 772 2759 (Department Administrator Anna Bolling)

Gothenburg, Sweden 2015

Acknowledgements

The following words of thanks are dedicated to all the people, without whom this thesis would not exist in the way it does today.

First of all, I want to thank my supervisor and examiner **Prof. dr. tech. Jens Nielsen** for entrusting me with this thesis and for giving me the opportunity to be part of the Systems and Synthetic Biology Research Group at Chalmers TH.

I also want to thank **Avlant Nilsson** for his time and encouraging talks, and for providing many scripts, which for the purpose of this thesis could be modified and employed in the analysis and **Benjamin Sanchez Barja** for interesting discussions and helpful suggestions.
Furthermore I want to thank the whole research group for creating a comfortable working atmosphere.

Special thanks goes to **my family** for supporting me throughout my whole time in Gothenburg for my Master Studies.

Finally, I want to thank my old and new friends, team mates and class mates in Germany and in Sweden. You are the best, thank you!

Abstract

During the last decades the research interest for creating a more sustainable and environmentally-friendly society and industry has increased dramatically. The employment of microorganisms to create valuable compounds such as fuels and chemicals from renewable resources has gained high popularity. With directed modifications in the metabolism of these microorganisms, metabolic engineering seeks to optimize the properties of these organisms for an efficient bioprocess to produce valuable compounds. One characteristic of metabolic engineering is the extensive use of computational models to predict the behavior of the metabolism and thereby identifying suitable targets for genetic interventions in an *in silico* to *in vivo* progress.

Flux Balance Analysis (FBA) is a popular analysis method for metabolic models, as it only requires the underlying stoichiometric network of the metabolism modelled. With FBA the optimal flux distribution, given a certain objective to be maximized, can be calculated with linear programming algorithms. As FBA only takes the stoichiometry and reaction directionality into account, further physiological constraints have to be included in the framework to increase the predictive strength of the simulations. In this thesis, an expanded form of FBA, that includes kinetic enzyme parameters as additional constraints on the system, was used to analyze a metabolic model of the popular industrially-used organisms *Saccharomyces cerevisiae*, that included, additionally to the native metabolism, also recombinant enzymatic steps to form the plant secondary metabolites resveratrol and naringenin from the aromatic amino acids phenylalanine and tyrosine. Resveratrol and naringenin have been found to be beneficial to human health by offering anti-inflammatory, anti-carcinogenesis and anti-oxidant properties.

The kinetically-constrained model was successfully used to estimate the impact of introducing recombinant pathways on the protein pool in the cell as well as to identify the enzymatic steps offering the highest control over the flux towards these products. This might be used in metabolic engineering to estimate the efficiency of metabolic pathways in the context of the whole metabolism form a protein cost point of view.

Keywords: FBA, constraint-based modelling, *S. cerevisiae*, aromatic amino acids, resveratrol naringenin, flux control, recombinant pathways, protein mass distribution

Contents

Abbreviations

AAA Aromatic Amino Acid

Brenda Braunschweig Enzyme Database

EC Number Enzyme Commission Bumber

FBA Flux Balance Analysis

FCC Flux control Coefficient

GRAS Generally regarded as safe

KEGG Kyoto Encyclopedia of Genes and Genomes

MCA Metabolic Control Analysis

RAVEN Reconstruction, Analysis and Visualization of Metabolic Networks

TCA Cycle Tricarboxylic acid cycle (citric acid cycle)

1 Introduction

1.1 Metabolic Engineering and Modelling

In the context of the increasing need for sustainable and feasible bioprocesses as alternatives to the industry based on fossile resources, the development of so-called cell factories gained popularity in science and industry. Cell factories are microorganisms being employed to produce certain chemicals of industrial relevance like fuels and pharmaceuticals from renewable resources, often with modified metabolism to improve the efficiency. This includes increased productivity (higher yield/titer/production rate), increased robustness and even the expression of metabolic pathways from other organisms to form new products.
With the approach of metabolic engineering of directed genetic changes under the utilization of modern analysis tools and the promising advances in modelling, cell factories are being promised a bright future in a bio economy [1].

The characteristic of metabolic engineering is the close interaction between laboratory experiments and computational analysis. Figure 1.1 displays the iterative cycle of metabolic engineering, where metabolic modelling marks the connection between the analysis of propierties of the metabolic network and the synthesis of a new, modified system with improved properties. Metabolic modelling seeks to formulate the results into mathematical frameworks which can be used to simulate and predict the cellular behavior and the changes that lead to increased performances by identifying suitable targets for genetic modifications to increase the flux through certain pathways. The proposed modifications are implemented and analyzed in laboratory experiments, with the results being used to adapt and improve the models, leading to a better understanding of the biological system and thereby optimized methods to optimize it for industrial purposes.

One popular method in computational analysis of metabolic networks is flux balance analysis (FBA), as it only requires the underlying stoichiometric structure of the model and a small amount of parameters to calculate the flux distribution through the network assuming a certain objective to be maximized [2]. Because of its simplicity and assumptions the predictive strength of FBA models is limited and the incorporation of more further parameters has been proposed [3; 4; 5] to allow for presumingly more realistic predictions.

By including protein kinetics into the FBA framework as described in [5], the shifts between metabolic strategies in the unicellular yeast *Saccharomyces cerevisiae* could be predicted using

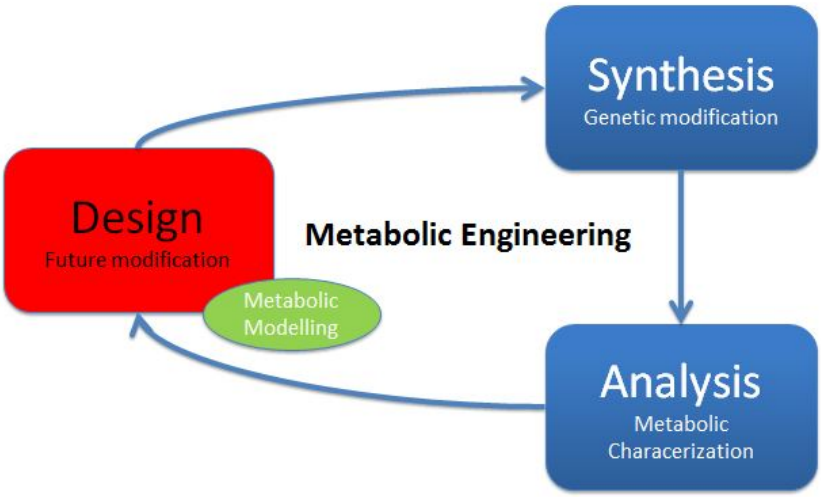

Figure 1.1: Iterative Cycle of Metabolic Engineering.

a stoichiometric model of the central carbon metabolism of the organism (Avlant Nilsson, unpublished data).
The work presented here expands this framework by including more parts of the metabolism and additionally reactions that are not naturally found in *S. cerevisiae* for producing industrially valuable compounds. This moves the impact of the work from only understanding the native cellular behavior towards a recombinant, modified system as it is common in metabolic engineering.
The recombinant pathways included in the model form the plant secondary metabolites resveratrol and naringenin from the aromatic amino acids tyrosine and phenylalanine. Both were found to be beneficial for human health due to their aromatic structure and the associated properties [6].

1.2 Aim and Objective of this Master Thesis

The aim of this work was to develop a constraint-based flux model of the metabolism of *S. cerevisiae* in an kinetically-constrained FBA framework, that includes the natural and recombinant pathways to simulate the formation of the industrially valuable compounds resveratrol and naringenin.

The incorporation of kinetic parameters of the enzymes responsible for the catalysis of each step enables the usage of metabolic pathway analysis methods such as metabolic control analysis [7] to identify enzymatic steps that exert control over the flux. Furthermore, it allows for an estimation of the mass of individual enzymes required for a certain production or growth rate and thereby the impact of introducing recombinant pathways on the protein pool in the cell.
With this approach, suitable targets for genetic intervention may be identified and recombinant

pathways may be ranked according to their impact on the protein content and composition in the future.
The main research questions that were investigated in this thesis include:

Q1: Can the dynamic metabolite changes in batch fermentations be explained by protein-kinetic constraints?

Q2: How much does the introduction of a recombinant pathways impact the protein pool available for the native metabolism?

Q3: How does the protein mass distribution change under maximum yield conditions for the desired products resveratrol and naringenin?

Q4: Which enzymatic steps offer the highest control over the flux towards biomass synthesis and desired production formation?

During the process of this work several tasks were carried out in trying to find answers for the stated research questions. An existing stoichiometric network model of the central carbon metabolism of *S. cerevisiae* was expanded by reactions of the necessary to simulate growth on different carbon sources as well as to synthesize the recombinant products. Furthermore an extensive literature and database review was carried out to collect the kinetic information for each enzymatic steps.
A mathematical framework for the analysis was created on basis of the RAVEN toolbox [8] for MatLab, enabling the expanded FBA approach used in this work.
The resulting model simulations were compared to experimental data to investigate their validity. Finally, the model was used to estimate protein mass distributions under several conditions and flux controlling steps in the network were identified.
No experiments were carried out during this work and the experimental parameters and values are taken from research articles.

1.3 Structure of the Report

The report for this master thesis is divided into 7 chapters:

- chapter 2 presents all the theoretical background about the modelled organism, the included metabolism, the recombinant pathways and the analysis theory
- chapter 3 explains the methodology for the analysis in a detailed way, including the network reconstruction as well as the main analysis methods
- chapter 4 describes all the necessary information needed to create and validate the kinetically-constrained FBA framework
- chapter 5 covers all the results from this work
- chapter 6 puts the results into perspective and covers the discussion of the results
- chapter 7 concludes the work and gives suggestions for future advances

2 Theoretical Background

The following chapter provides the background information about the investigated organism *Saccharomyces cerevisiae* and its importance for biotechnology, the metabolism of interest, namely the synthesis of aromatic amino acids and derived products, as well as the basic aspects of the theoretical part of metabolic engineering covered in this work, including flux balance analysis frameworks and metabolic control analysis.

2.1 The yeast *Saccharomyces cerevisiae*

2.1.1 General physiological and morphological characteristics

The unicellular organism *Saccharomyces cerevisiae* (*S. cerevisiae*), better known as "baker's yeast", is probably the most investigated and characterized organism in the highly diverse group of yeasts and belongs to the class of *Saccharomycetes*. It has been used for the production of food and beverages like wine and beer since ancient times due to its ability to ferment glucose and other sugars to ethanol and CO_2, even under aerobic conditions ("Crabtree effect").
Antonie van Leeuwenhoek first described yeast from fermenting beer in 1680. For a long time it was unclear if these microorganisms were a by-product or the cause of the fermentation process, until Louis Pasteur proved the latter in 1886 [9]. The name *Saccharomyces* means "sugar yeast" and was first used in 1837 for budding yeast in beer and wine, and incorporates currently seven species [10]. The natural habitats of *S. cerevisiae* are diverse and include tree barks, soil and (decaying) leafs [10].

Beside its metabolism, another distinct characteristic of *S. cerevisiae* is its cell cycle. It performs an asymmetrical mitotic division into a mother and a daughter cell, which is approximately 60 % in size of the mother cell [11] . This process is referred to as budding and marks one of the most well-known features in biology. Every division leaves a budding scar on the mother cell and enables the determination of the replicative age of the cell by counting these scars under a microscope.

S. cerevisiae evolved being able to withstand various changes in the environmental conditions of their natural habitats such as nutritional shortage, temperature, salinity, low pH and osmotic stress. The cells has several molecular responses to handle with these stress conditions, mostly leading to a cell cycle arrest while the damage these conditions cause are being sorted out [12]. Protection against physical damage is provided by a thick cell wall, which can make up to 25 %

to 30 % of the cell dry weight. It consists of carbohydrates (about 60 % of cell wall dry weight, mostly 1,3 β -Glucan) and glycoproteins for various functions such as sensing and protection [13].
Growth rate and nutritional situation have a severe influence on the biomass composition of *S. cerevisiae*. At a growth rate of 0.1 1/h the protein content differs between 37.5 to 45 % of the cell dry weight and at high growth rates of 0.4 between 55.5 and 60.1 % cell dry weight [14; 15].

In 1996 the genome of *S. cerevisiae* was sequenced [16] as the first eukaryotic organism, revealing about 6000 genes (humans about 30000) with most of their functionality known by now.
Despite being a microorganism, *S. cerevisiae* belongs to the domain of eukaryotes. Eukaryotic cells are defined by having specialized cell compartments separated by lipid membranes, called organelles, to separate certain cellular processes. Since human cells are eukaryotic as well, many discoveries in yeast can be translated to human cell biology because many genes and their functions are highly conserved amongst species of the same domain.
With the lesser complexity and the easier cultivation, *S. cerevisiae* is therefore a very potent model organism for studying molecular processes in human cells like cell division, apoptosis and even cancer and other diseases [17].

2.1.2 *S. cerevisiae* in industrial biotechnology

It is not only because of its long historical use that *S. cerevisiae* is one of the most popular organisms in industrial bioprocesses nowadays, but several characteristics paved the way for the successful implementation of *S. cerevisiae* as a production host and may enable even further improvements.
With its cell wall and the stress response mechanisms it is naturally a robust organism. The ability to withstand low pH values, produced acids can be recovered in their carboxylated from which reduces the downstream costs. Additionally, studies have shown the possibility to expand the substrate range for *S. cerevisiae*, allowing the cells to grow on naturally abundant sugars like xylose and arabinose, thus enabling a more sustainable process. There are also efforts to modify *S. cerevisiae* to utilize the cheap and sustainable lignocellulose from wood [18; 19]

The cultivation of *S. cerevisiae* is relatively uncomplicated, since it is nonpathogenic and categorized as GRAS (generally regarded as safe). Furthermore, the loss of cultivations due to virus infections are, in opposite to bacterial bioprocesses, nonexistent. Because of the long historical usage, there are also established technologies and fermentation strategies to optimize the conditions for bioprocesses using *S. cerevisiae*.

Due to the broad understanding of physiology and molecular biology of the organism, there exists a diverse and efficient genetic toolbox, including numerous vectors and transformation strategies, utilizing amongst others the natural homologous recombination activity of *S. cerevisiae*.
Recent developments in synthetic biology and genetic engineering such as the CRISP-Cas system increase the efficiency, accuracy and speed of genetic modifications even further [19; 20] With the methods of synthetic biology and metabolic engineering on the rise, the metabolism can

be further optimized towards a certain product and whole new pathways can be implemented to increase the product range. This includes the successful production of the malaria drug precursor amorphadiene [21] and vanillin [22] and studies aiming for the production of secondary metabolites such as opiates, polyketides, isoprenoids and fatty acid derived products [19].

Since the sequencing of the genome in 1996, several databases with other "omics" measurement data followed. The increasing amount of data and advanced bioinformatics tools enabled the development and establishment of computational models concerning metabolism and its control, thus enabling the iterative interactions in metabolic engineering between design, implementation and analysis in combination with systems biology [23].

2.2 Synthesis of Aromatic Amino Acids and Derived Products

Amino acids are a class of organic compounds characterized by an amine (-NH_2) and a carboxylic acid (-COOH) functional group as well as a side-chain, specifying each amino acid. In physiological roles they can occur freely as metabolites in biosynthetic pathways, or as proteinogenic amino acids, which are the building blocks of peptides and proteins. They are combined to polymers via peptide bonds under the release of water. Proteinogenic amino acids carry both functional groups on their first carbon-atom, the alpha-carbon, and are therefore named alpha-amino acids.
Depending on their side chains, amino acids can be grouped differently according to their physicochemical properties in polar/apolar, charged/uncharged or hydrophilic/hydrophobic.
The biosynthesis of amino acids in *S. cerevisiae* is highly connected with nitrogen fixation [9; 24]. Inorganic nitrogen (mostly in form of ammonia, NH_3) is bound via a reductive transamination with 2-oxoglutarate to form glutamate. Glutamate, a amino acid itself, serves as the starting point for all proteinogenic amino acids in yeast as it transfers the amino group into the metabolism. Another source of nitrogen can be the uptake of glutamine, which can be catabolized into glutamate. The synthesis if the other amino acids is variously grouped according to their structure and is connected to the central carbon metabolism at various points [9; 24].
The group of aromatic amino acids consists of phenylalanine, tyrosine and tryptophan. All of them include a 6-carbon ring structure in their side-chains. In opposite to humans, *S. cerevisiae* and plants are able to synthesize aromatic amino acids on their own via the chorismate pathway from pyruvate and D-erythrose-4-phosphate. Chorismate (from greek "separating") marks the branch point in the pathway, where one branch leads towards phenylalanine and tyrosine and another towards tryptophan, in regards of amino acids. Chorismate serves also as a metabolite for the synthesis of folate and ubiquinone [9]. The last two branches will not be covered in this work and special emphasis will be put on the synthesis of phenylalanine and tyrosine.

The pathway to chorismate consists of seven reactions, carried out by only four enzymes (see Figure A.1 in the appendix). It begins with a condensation reaction of phosphoenolpyruvate from glycolysis and D-erythrose-4-phosphate from the pentose phosphate pathway, giving the 7-carbon sugar 3-deoxy-D-arabino-heptulosonate-7-phosphate, DAHP. This reaction is catalyzed

by DAHP-synthase, encoded by the genes ARO3 or ARO4. These are inhibited by tyrosine and phenylalanine, respectively.
The following five reactions, which are catalyzed by five separate enzymes in *Escherichia coli*, are carried out by a single pentafunctional enzyme in *S. cerevisiae* (ARO1). The reactions are assumed to be not tightly channeled and occur not in sequence of the domains.
The reactions begin with a ring closure by dephosphorylation in domain C, giving 3-dehydroquinate. A dehydration in domain E then leads to 3-dehydroshikimate, followed by a reduction to shikimate in domain D using NADPH. Under employment of ATP, shikimate gets phosphorylated in domain B to shikimate-3-phosphate. Finally, a pyruvoyl-group from another phosphoenolpyruvate molecule is added under the removal of phosphate, resulting in 5-enolpyruvyl-3-phosphate. Chorismate is then formed in another dephosphorylation reaction, catalyzed by chorismate synthase (ARO2).
For the branch to phenylalanine and tyrosine (see Figure A.2 in the appendix), the pyruvoyl-group gets rearranged from an O- to a C-linkage by chorismate mutase(ARO7), which is inhibited by tyrosine and activated by tryptophan.
The direct ketoacid precursors for phenylalanine and tyrosine are formed separately. Prephenate dehydratase (Pha2) gives phenylpyruvate, which by transamination (ARO8/9) reacts to phenylalanine. Prephenase dehydrogenase (Tyr1) gives 4-hydroxyphenylpyruvate, the substrate for the transamination (ARO8/9) that forms tyrosine [9].

Whereas in *S. cerevisiae* aromatic amino acids are present at rather low level [6] and are mainly used for protein synthesis, they are widely used in plant secondary metabolism and serve as precursors for benzylisoquinoline alkaloids, stilbenes and various flavonoids (Phenylalanine, Tyrosine) as well as various alkaloids (Tryptophan) [6].

In this thesis the synthesis of resveratrol and naringenin as recombinant pathways in *S. cerevisiae* will be covered (see Figure 2.1). Both of them inherit an anti-oxidant activity due to their aromatic structure. The commercial production from plants is limited by their low growth rate and low production efficiency as well as complex and expensive extraction methods [25; 26]. The implementation of *S. cerevisiae* as a cell factory for these compounds is therefore a promising alternative.
Resveratrol is a stilbene used as an anti-aging compound and naturally found in grapes, peanuts and a various berries as well as food products derived from these [25]. It has been identified as a potentially valuable compound for industrial production because of its anti-inflammatory, -aging, and -carcinogenesis effects [27].

Naringenin is a branch point metabolite and precursor for many flavonoids with antibacetrial, antiviral and anti-tumor properties [6; 26]. The pathways leading to each of the compounds are chosen according to literature and availability of kinetic parameters as well as experimental data.
Whereas phenylalanine is catalyzed towards coumarate via the intermediate cinnamate, tyrosine is converted directly into coumarate. Coenzyme A is added to coumarate, before it reacts with three molecules of malonyl-CoA towards either naringenin chalcone and then further to

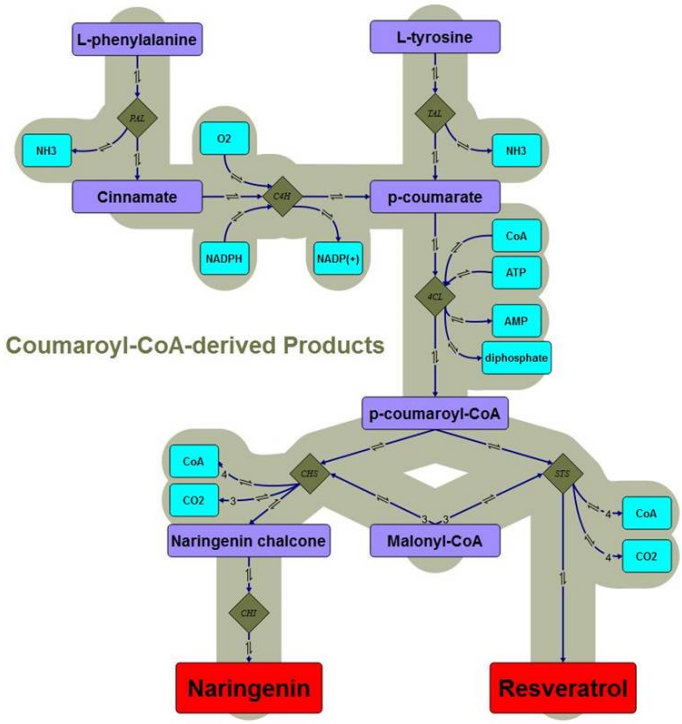

Figure 2.1: Pathways for the production of resveratrol and naringenin from phenylalanine and tyrosine. The enzymes used are found in Table B.7 in the appendix.

naringenin or directly into resveratrol. Malonyl-CoA is naturally synthesized in *S. cerevisiae* by acetyl-CoA-carboxylase from acetyl-CoA and CO_2.

2.3 Metabolic Modelling

2.3.1 Flux Balance Analysis

In the context of more detailed understanding of biological systems and their optimization for industrial processes, mathematical descriptions of metabolic functions increased in popularity for biotechnological research. Attempts to model systematic metabolic dynamics however require data on kinetics of many enzymes and regulatory features, which are usually difficult to obtain. These limitations can be overcome if only the distribution of the metabolic steady-state fluxes is of interest, based on the pathway stoichiometry, reaction directionality and the requirements for growth. This constrained-based modelling of metabolic networks is known as Flux Balance Analysis (FBA) and relies on measuring the rates of production/consumption of metabolites in the growth medium, so-called exchange fluxes, serving as the constraints. The internal fluxes can then be fitted from these exchange fluxes by linear regression, as long as the metabolic network, e.g. the set of possible enzymatic reactions, is known. The basis of constrained based modelling is therefore the underlying stoichiometric network as well as the exchange fluxes, which in combination form a solution space including all possible flux distributions allowed by these

constraints, representing only physiological feasible phenotypes. By taken even more constraints into account, such as co-factor balancing and reaction directionality, the solution space can be reduced even more [28]. This demands that the metabolic network with the related constraints describes the physiology and the limitations at the steady-state in a realistic manner, including biomass synthesis, maintenance requirements and secretion of important products.

In order to analyze a metabolic network, the collected information has to be put into an appropriate mathematical framework. As the name suggests, FBA is based on creating mass balances around each intracellular metabolite in the network.

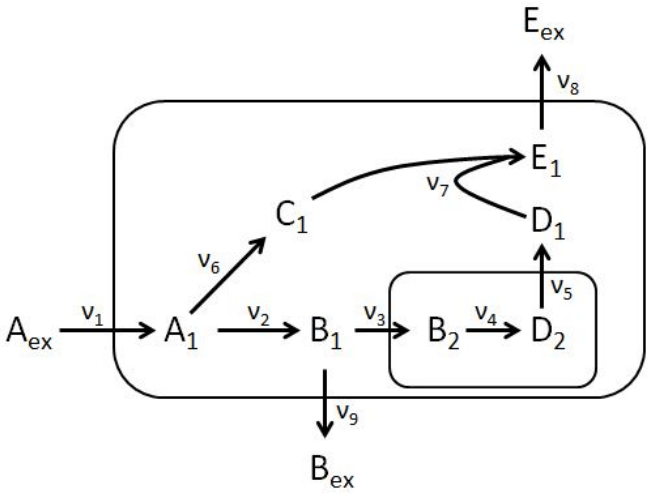

Figure 2.2: Example of a simple stoichiometric metabolic network

For each intracellular metabolite in the example in Figure 2.2, the following mass balance equation can be formulated:

$$\frac{\mathrm{dX_i}}{\mathrm{dt}} = \nu_{\mathrm{in,i}} - \nu_{\mathrm{out,i}} + \nu_{\mathrm{generated,i}} - \nu_{\mathrm{consumed,i}} - \mu[\mathrm{X_i}] \tag{2.1}$$

According to (2.1), the change in the concentration of metabolite X_i is dependent on its uptake rate ($\nu_{in,i}$), minus the rate of its excretion ($\nu_{out,i}$), plus the rate of its intracellular generation ($\nu_{generated,i}$), minus its intracellular consumption ($\nu_{generated,i}$). The last term accounts for the decrease in concentration of the metabolite ($[\mathrm{X_i}]$) following a increase in cell size during growth (with specific growth rate μ), and is referred to as the dilution term ($\mu[\mathrm{X_i}]$).
As an example, the mass balance around one metabolite in the network in Figure 2.2, A, is derived:

$$\frac{\mathrm{dA_1}}{\mathrm{dt}} = \nu_1 - \nu_2 - \nu_6 - \mu[\mathrm{A_1}] \tag{2.2}$$

Note that producing reactions have a positive sign and consuming reactions have a negative sign in the mass balance equation, representing the stoichiometry of the network.

This dynamic balance equation would require the kinetics of enzymatic reactions to be solved, which are not available for every reaction and might differ under different conditions. However, with the assumptions of FBA one is able to bypass this problem.
The concentration of intracellular metabolites is believed to be very low in comparison to the fluxes. With this assumption the dilution term can be neglected ($\mu[\mathrm{X_i}] = 0$) [7]. Furthermore it is assumed, that metabolism, e.g. the change in metabolite concentrations, occur way faster than changes in the environment or in growth behavior. The concentrations of intracellular metabolites are therefore believed to be in steady-state relative to growth and the dynamics of the systems can be ignored ($\frac{\mathrm{dX_i}}{\mathrm{dt}} = 0$).
With these assumptions the mass balance of A_1 gets reduced to:

$$0 = \nu_1 - \nu_2 - \nu_6 \tag{2.3}$$

This kind of mass balance can be created for every intracellular metabolite, which can be combined in a matrix formulation for the whole system:

$$0 = \mathrm{S} \cdot \nu \tag{2.4}$$

The so-called stoichiometric matrix S contains all the stoichiometric information about the intracellular metabolites, as the name suggests. The rows represent the metabolites, the columns represent the reactions. The vector ν contains all fluxes, e.g. all the reaction rates. For the small example network the S-matrix looks as follows. For a better understanding the corresponding metabolites and rates are shown as well.

$$S = \begin{array}{c} \\ A_1 \\ B_1 \\ B_2 \\ D_2 \\ D_1 \\ C_1 \\ E_1 \end{array} \begin{array}{c} \begin{array}{ccccccccc} \nu_1 & \nu_2 & \nu_3 & \nu_4 & \nu_5 & \nu_6 & \nu_7 & \nu_8 & \nu_9 \end{array} \\ \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \end{pmatrix} \end{array}$$

By solving the linear equation (2.4), the intracellular fluxes can be estimated to get the flux-vector ν, forming the solution space. Generally the number of reactions will be higher than the number of metabolites in the system, which means the number of unknown variables is higher than the number of equations and the system is underdetermined. This is symbolized by the degrees of freedom, the number created by subtracting the number of intracellular metabolites from the number of fluxes. In the example system there are two degrees of freedom.
Following this circumstance there exist no unique solution for the flux distribution, but many possible ones that satisfy equation 2.4. The mechanisms through which the flux distribution is

chosen in nature is a complex interplay of enzymatic and genetic regulation, and is not known in detail. The solution space containing all feasible metabolic flux distributions, defined by the S-matrix, resembles the metabolic genotype of the investigated organism. By deleting reactions or introducing new ones via genetic engineering this solution space can be modified [2]

To solve the problem of the existence of more than one possible flux distribution it is assumed that the cell is striving to meet a certain metabolic objection [2]. All other fluxes are then chosen to maximize or minimize this particular objective. A commonly used objective for microorganisms is the maximization of growth, since under optimal conditions it is a good approximation for the selective driving force. In nature, the fastest growing organism generally outcompetes the other ones. Even though being a good approximation, the objective depends on the environment and there might be competing objective functions in the cell [29].

$$\textbf{Maximize:} \quad \mathrm{c} \cdot \nu \tag{2.5}$$

$$\textbf{Subject to:} \quad 0 = \mathrm{S} \cdot \nu \tag{2.6}$$

For models using the FBA-approach, the objective is defined as the linear combination of the flux-vector ν and the c-vector, containing coefficients for each reaction representing the objective. The flux distribution that maximizes the objective, in the constraints defined by the S-matrix and the lower and upper limit of exchange fluxes, is found using linear programming.

The method can be applied to very large networks, incorporating all enzymatic and transport reactions of a cellular system related to their genes. These genome-scale models have shown to be capable of quantifying the fluxes through certain pathways and thereby identifying possible intervention targets for gene deletions [30].

2.3.2 Extended Flux Balance Models

Classical FBA as described above takes only the reaction stoichiometry and the directionality into account. To increase the predictive power of these models, several approaches were developed to include further physiological constraints. These constraints may reduce the solution space further and may eliminate infeasable flux distributions.
FBA with molecular crowding (FBAwMC) take into account the limited volume of the cell and the thereby limited maximal concentration of enzymes required for fluxes according to their size. To estimate the amount of enzyme required for a certain flux kinetic enzyme data is required. In [3], a concentration-limited metabolic model of *E. coli* was successfully used to evaluate the metabolic activity and substrate uptake via FBAwMC.
The MOMENT (MetabOlic Modeling with ENzyme kineTics) framework marks an extension to the FBAwMC model and seeks to quantify the protein mass required for a certain flux, constrained by the fraction of the total protein that can be used up by metabolic enzymes. For this method the turnover number k_{cat} of each enzyme is used. It defines the number of molecules

that can be turned over by one molecule of enzyme in a certain time, usually given in [1/s]. A study on *E. coli* predicted growth rates under certain conditions better than FBAwMC [5]. The study also revealed better results when kinetic data from databases were used instead of values drawn from a gamma distribution. A yet unpublished study by Avlant Nilsson used a similar approach to explain the crabtree effect in *S. cerevisiae* with a kinetically-constrained metabolic model of the central carbon metabolism. Other approaches tend to limit the available membrane surface for membrane bound proteins such as transporters [31].
To account for a more realistic reaction directionality and to model backwards fluxes through enzymes in a physiological manner, a thermodynamic framework using ΔG values and metabolite concentrations was described for a genome-scale model of *E. coli* [4].

2.3.3 Metabolic Control Analysis and Flux Control Coefficients

The information in the following section is mainly taken from [7].
FBA is suitable to estimate the flux distribution around metabolic branch points according to the flux boundaries and the objective function, but gives no information about the parameters that are responsible for the control of the metabolic flux. One of the most important goals in the analysis of metabolic networks is to identify these parameters since they can assist in determining suitable targets for the directed modification of the underlying metabolic network.

In opposite to many proposed qualitative influences on control such as covalent and cooperative activators and inhibitors for enzymes, MCA offers a more rigor view on the relationship between flux control and enzymatic kinetics. The basis of MCA is the linear perturbation of the nonlinear behavior of enzyme kinetics.
Even though local in nature and extrapolations should be handled very carefully, MCA can be very useful to quantify the influence of individual enzymatic reactions on the metabolic flux control and thereby elucidating the concept of rate-limiting steps in metabolic pathways, their influences on intracellular metabolite concentrations as well as on the overall systemic metabolic behavior [7].

As with FBA, MCA only applies strictly to steady-state, which in MCA is defined by the enzymes activities catalyzing the individual reactions as well as the concentrations of the substrate of the first reaction and the product of the last reaction in the pathway. These system defining factors are called system parameters, in order to distinguish them from system variables. The latter are determined by the values of the system parameters and include the metabolic flux itself or concentrations of intermediary metabolites.

The central question of MCA is how the system parameters affect the system variables, or in other words, how sensitive a system variable is to changes in the systems parameters. These sensitivities can be expressed by a set of coefficients, most prominent being the control coefficients.
In the case of this work, the system variable of interest is the metabolic flux through a pathway, which is influenced by the individual enzyme activities of each step. This describes the most important control coefficients, the flux control coefficents (FCC). They give information about

the relative change in the steady-state flux caused by an infinitesimal change in an enzyme activity in the pathway, summarizing the extent of systemic flux control by each enzymatic step.

$$FCC^J = \frac{E}{J}\frac{dJ}{dE} = \frac{d\ln J}{d\ln E} \tag{2.7}$$

Since FCCs are defined as being the relative change of the flux divided by the relative change in enzyme activity, the factors have to be normalized with the total values of flux and individual enzyme activity. They are therefore dimensionless and the magnitude is not related to the flux or the unit of the enzyme activity. In linear pathways they can take values between 0 and 1, where 1 represents the highest flux control. For branched pathways negative values are possible. As stated above, the enzyme with the largest FCC has the most control over the flux at this particular steady-state, since the flux is affect the most by the change of the activity of this enzyme. This concludes, that an increase in enzyme activity of this enzyme would increase in the largest overall flux increase, pointing out a possible target for modifications.

2.4 Cultivation Modes

Commonly used fermentation processes can be differentiated according their operation mode into three different types.
For batch cultivations all nutrients are added once at the beginning. The cells grow until the nutrients are depleted, which means that the cells experience a continuously changing environment due to increasing cell and product concentration and decreasing nutrients, while the volume of the fermentation stays the same.
In the case of *S. cerevisiae*, a high initial concentration of glucose forces the production of ethanol under aerobic conditions, the so-called "Crabtree effect". The ethanol produced during this phase can then be used by the cells as a carbon- and energy-source when all the glucose is consumed. This increases the biomass concentration even further.
In the second form, the fed-batch, a continuous inflow of fresh media supports the cells with nutrients for a longer time, which might increase the biomass further.
Lastly, in continuous fermentations a inflow of fresh media equals the outflow of reactor volume. With proper stirring it is assumed that this operation mode enables steady-state conditions for the cells, thus allowing to create optimal conditions throughout the whole process. The long operation times and increased contamination risks as well as an overall more complicated experimental setup are the major downsides of continuous cultures.

In relation to the type of operation mode the experimental data for the modelling were generated from, special handling is required. Since in FBA a steady-state is assumed, experimental data from continuous fermentations are easiest to use for this form of model analysis. Nonetheless,

the other two forms are commonly used in industrial and experimental processes as well.
In the exponential growth phase of batch fermentations the cells usually work at maximum rate and the changes in the environment as well as cell physiology is relatively stable and therefore assumed to be at a steady-state.
The employment of dynamic FBA as described in [32] enables a rough estimation of the dynamic behavior using FBA (see section 3.3).

3 Methodology

The main method employed in this work was flux balance analysis in a modified version to incorporate enzyme-kinetic data and the molecular weight of the corresponding proteins as additional constraints, which successfully has been used prior to this work sources.

As the computational interface for performing the FBA and all related analysis methods of the model network the RAVEN (**R**econstruction, **A**nalysis and **V**isualization of M**e**tabolic **N**etworks) toolbox [8] for MATLAB was used. Only slight modifications to the source code were needed to allow the incorporating the metabolic network stoichiometric network model with all the additional data in form of an MS excel file and constraining the protein amount. The toolbox already inherited the algorithm for solving linear optimization problems, for which the MOSEK (MOSEK ApS, Copenhagen, Denmark) framework as the linear programming solver is used.
The FBA approach needs an underlying metabolic network model resembling the stoichiometry and reaction directionality as well as the allowed exchange of metabolites. With the expanded framework, an additional constraint had to be defined as the upper limit of protein content in the cell and furthermore the fraction of this total protein content that was represented by the enzymatic reactions that have corresponding protein data associated.
Since the stoichiometric network investigated did not only include endogenous but heterologous reactions, factors like the energy maintenance, protein abundances and exchange reaction rates were likely to differ from wild-type literature data.
To evaluate the impact of the experimental data used to build up the model, the sensitivity to perturbations in the data used is being invested by recording changes in relevant (exchange) fluxes.
An important tool in metabolic engineering to modify the fluxes in a metabolic network is the exploration of enzymes that offer the highest influence on the flux through a certain pathway. For this purpose a metabolic control analysis was performed to identify the flux control coefficients.

The following chapter covers the development and description of the metabolic model used in this thesis, the implementation of the expanded FBA using the protein data and the methods used in analyzing the model behavior and the impact on the protein pool of introducing a recombinant pathway.

3.1 Reconstructing the Metabolic Network

The model reconstructed for this work marks an extension of the "smallYeast" model form the RAVEN toolbox tutorial [8]. The latter is derived from the genome-scale iFF708 model [33] and includes the central carbon metabolism (glycolysis, citric acid cycle, pentose-5-phosphate pathway) as well as a simplified reactions resembling oxidative phosphorylation and a biomass reaction, amounting in 54 reactions overall.

For the extended model, referred to as AAA model (Aromatic Amino Acid) from now on, several reactions and related metabolites were added to the original smallYeast framework to describe oxidative phosphorylation more realistically (see section 3.1.1) and to allow the model to synthesize the aromatic amino acids tyrosine and phenylalanine (see section 3.1.2) as well as the products naringenin and resveratrol (see section 3.1.3). The additional reactions from yeast were implemented from the Yeast Pathway Database [34] and the KEGG database [35] and in agreement with the iFF708 model. The downstream pathway to form the desired products were taken from research articles and implemented according to the KEGG database and MetaCyc database [36].
Furthermore, the biomass equation was adapted to the added metabolites and the free uptake and excretion of phosphate and ammonia were added, amounting in overall **170** reactions. All reactions that contribute to the protein constraint are listed in the appendix with the related protein kinetics and molecular weight and assumed subunit structure.

3.1.1 Inclusion of the Oxidative Phosphorylation

Since oxidative phosphorylation marks an essential part of energy metabolism in yeast under industrial conditions, a realistic representation was implemented in the AAA model by including the major reactions of the respiratory electron transport chain. The five main reactions as well as associated compartment transport were added according to the iFF708 model.

3.1.2 Adding reactions for aromatic amino acids and derived products

One of the major parts of the model expanding was to include to reactions for the formation of the aromatic amino acids tyrosine and phenylalanine. The implementation covers all the reactions of the chorismate pathway and the synthesis of the amino acids as described in detail in section 2.2.
The reactions for the formation of the products resveratrol and naringenin (see section 2.2) could not be extracted from a more complex yeast model like the iFF708 model because they are not endogenous to *S. cerevisiae*, with exception of the malonyl-CoA forming Acetyl-CoA Carboxylase *ACC1*. The reactions were implemented according to MetaCyc and KEGG pathways described in several research articles [25; 26].

3.1.3 Glyoxylate cycle for growth on ethanol

To allow the model to simulate a realistic growth on ethanol, the reactions of the glyoxylate shortcut were added to the model. The glyoxylate cycle is a crucial part of gluconeogenesis to create important precursor metabolites under the growth on non-fermentative carbon-sources such as ethanol. The glyoxylate shortcut was implemented in the model with two reactions complementing the TCA cycle, forming malate from isocitrate via glyoxylate under the formation of succinate.

3.1.4 Biomass equation and synthesis of biomass constituents

The biomass equation in stoichiometric metabolic models resembles the drain for each metabolite required for accumulation of biomass, representing the growth of the cell. It also contains a factor representing the consumption of energy in form of ATP for maintenance processes in the cell such as pH stabilization, protein turnover, osmolarity and other non-growth-related processes. This factor as well as the composition of the biomass equation is subject to changes related to growth rate and nutrition, but was set to a value of 35.36 [33] for this thesis. To fit the experimental batch data the ATP maintenance factor was modified.(see section 5.2)
The biomass equation itself was build up modularely to represent the synthesis of protein, RNA, DNA, lipid and carbohydrate biomass on their own, resulting in eight modular reactions whose superposition forms the final biomass equation together with the maintenance mentioned above. This modular set up allowed for easy implementation of changes in the actual composition of the biomass. The metabolites included in each module and their stoichiometric coefficients were taken from [33]. The reactions to synthesize the biomass metabolites in each module were derived from the iFF708 model by an algorithm developed by Avlant Nilsson (Unpublished data), which was modified to use the metabolites in the AAA model and will not be described further. The algorithm allowed the synthesis of each biomass metabolite only from the metabolites included in the AAA model, resulting in a "lumped" reaction representing the drain of the metabolites in the AAA model. This resulted in 38 lumped reactions to form the biomass constituents used in the eight modular equations.
In order to allow the lumped reactions for most of the amino acids to be valid, two reactions to form cytoplasmic 2-oxoglutarate were implemented as well as a tricarboxylate transporter that allows the antiporter exchange of mitochondrial and cytoplasmic citrate and isocitrate.
All the metabolites from the heterologous pathway were not included, since it is believed that they do not contribute in a significant way to the biomass but are instead reaction completely into the final products.

3.1.5 Reversible Reactions

In stoichiometric network analysis, the backwards flux through reversible reactions is usually represented through a negative flux value. With the extended FBA approach used in this study,

this would result in a negative protein abundance which is physiologically infeasible. To avoid this case every reversible reaction was split up into two reactions, one for the forward and one for the backwards flux, whereas the forwards reaction the forwards reaction defined in the KEGG database [35].

3.1.6 Isoenzymes and subunits

For many enzymes there exist several structural forms, so called isoenzymes, that catalyze the same reaction. From their slightly different structure result different properties like a different catalytic activity or substrate affinity etc. They also differ in expression levels, depending on the physiological state and growth conditions the cell is in. This allows a more sophisticated control on metabolism. Furthermore some isoenzymes might be specific for a certain cellular compartment.
Since the k_{cat} values are mostly taken from databases according to their EC (Enzyme Commission) number which only defines the reaction and does not specify the isoenzyme catalyzing the reaction, the same value is taken for every isoenzyme.
To include the same reaction several times to accord for every isoenzyme would not have solved this problem since the linear optimization algorithm would always have choosen only the reaction with the highest k_{cat}, given a similar molecular weight of the enzyme. It was therefore chosen to only include the enzyme with the highest abundance according to proteomics data to be responsible for the catalysis of the related reaction.

Many functional enzymes in their physiological state consist of several polypeptide subunits, forming dimers, tetramers or larger complexes of the same or different subunits. To take this into account the molecular weight of each enzymes was calculated according to the number and weight of their subunits.

3.2 Kinetically constrained Flux Balance Analysis

The original framework of FBA constraints only the upper and lower boundaries of the fluxes, mostly defining the input or forcing a certain output of the model. In the kinetically constrained version used in this thesis each flux of an enzymatic reaction is related to the kinetic parameters of the catalyzing protein and its molecular weight, enabeling the prediction of the amount of protein mass required for a certain flux. By setting boundaries on the protein mass allowed in a cell an additional constrained is set on the framework, which is expected to make predictions more realistic.

3.2.1 Expanding the FBA framework

The constraint matrix used in standard FBA (see section 2.4.1) was expanded by an additional row, relating the flux of each reaction with a k_{cat} value to the protein mass required. The lower boundary was set to 0, since negative protein constraints are not defined in this framework. The upper boundary is set to the specific fraction of the total protein mass in the cell covered by the model (see section 3.2.2).

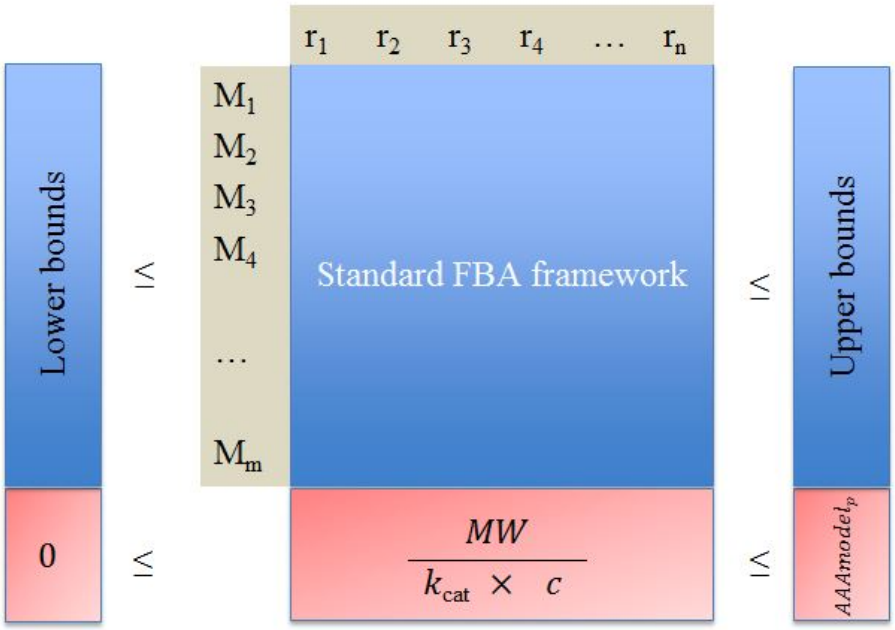

Figure 3.1: Expansion of the standard FBA framework by a kinetic protein constraint on every reaction catalyzed by an enzyme.

3.2.2 Determining the protein constraint

The required amount of proteins P for each flux F was determined according to equation 3.1 by the turnover number k_{cat} multiplied by a factor c ($0 \leq c \leq 1$), which relates the actual speed of the catalysis to the theoretical maximum. This c parameter is very likely to vary for every enzyme in relation to the growth conditions, it was assumed that on an average of all the enzymes used in this work the variations even out and therefore one c value for every reaction was chosen.

$$F = \frac{P}{c * k_{cat}} \tag{3.1}$$

The enzymatic reactions in the model cover only a fraction of the total protein pool in the cell. The lumped reactions and the transport reactions between compartments and the extracellular medium are not related to protein kinetics and therefore not constrained by the framework.

To represent the fraction of protein covered by the model and to simulate the competition around the protein pool by the heterologous pathways, a factor $AAAmodel_P$ was introduced.

$$0 \leq AAAmodel_P = \frac{Protein_{AAAmodel}}{Protein_{total}} \leq 1 \tag{3.2}$$

The $AAAmodel_P$ factor was determined by protein abundance data (see Materials 4.3) to be $AAAmodel_P = 0.235$.

3.3 Iterative Flux Balance Analysis for Batch Cultivations)

One of the major assumptions of FBA is the steady-state of all fluxes and metabolite concentrations in the metabolic network (see section 2.4.1). In order to simulate the dynamics of batch fermentations an iterative algorithm was implemented as described in [2] and compared to experimental data.
The basis of this method is to use FBA to calculate the flux distribution and thereby the consumption and secretion of relevant metabolites for short time steps delta T, in which the steady-state assumption is expected to be valid. With the consumption and secretion as well as the growth rate predicted by FBA the changes in each time step can be calculated, given initial values. The results of each time step are then used as the initial values for the next time step.
The algorithm was implemented in MatLab following [2] and included the following steps:

1. The substrate concentration of the current time step is the resulting concentration of the last time step. For the first time step the initial concentration from the experimental data was used.

$$\mathrm{X_{t,initial}} = \mathrm{X_{t-1,result}} \tag{3.3}$$

$$[\mathrm{S}]_{\mathrm{t,initial}} = [\mathrm{S}]_{\mathrm{t-1,result}} \tag{3.4}$$

2. To set the upper limit for substrate uptake in the FBA framework the amount of available substrate for each unit of biomass was calculated.

$$\text{Substrate uptake limit} = \frac{[\mathrm{S}]_{\mathrm{t,initial}}}{\mathrm{X_t} \cdot \Delta \mathrm{t}} \tag{3.5}$$

3. With the FBA framework the flux distribution for the network model is calculated, giving the actual substrate uptake rate $\nu_{S,in}$, metabolite secretion rates $\nu_{P,out}$ and the growth rate μ.

4. By using the actual uptake and production rates the concentrations for the next step are calculated using the differential equations (see 2.5).

$$\frac{\mathrm{dX}}{\mathrm{dt}} = \mu \mathrm{X} \;\rightarrow\; \mathrm{X_{t,result}} = \mathrm{X_{t,initial}} \cdot \mathrm{e}^{\mu \Delta \mathrm{t}} \tag{3.6}$$

$$\frac{\delta[\mathrm{S}]}{\delta \mathrm{t}} = -\nu_{\mathrm{S,in}} \cdot \mathrm{X} \rightarrow\; [\mathrm{S}]_{\mathrm{t,result}} = [\mathrm{S}]_{\mathrm{t,initial}} + \frac{\nu_{\mathrm{S,in}}}{\mu} \mathrm{X_{t,result}} (1 - \mathrm{e}^{\mu \Delta \mathrm{t}}) \tag{3.7}$$

$$[\mathrm{P}]_{\mathrm{t,result}} = [\mathrm{P}]_{\mathrm{t,initial}} + \nu_{\mathrm{P,out}} \mathrm{X_t} \tag{3.8}$$

The algorithm was used to predict the changes in concentrations of biomass, glucose, ethanol and resveratrol, according to the experimental data. During the presence of glucose it was assumed to be the only substrate taken up by the cell. When glucose was depleted, growth on ethanol was allowed.

3.4 Determination of the Flux Control

As mentioned above, one of the major goals of computational metabolic engineering is to identify the factors exert the biggest influence on the flux through a certain pathway. To identify suitable intervention points to optimize the flux in a metabolic network Metabolic Control Analysis has shown to be a suitable tool (see section 2.3.2).
The method relates the changes in pathway flux to the perturbation of each single kinetic parameter and thereby quantifying the influence of each enzymatic step, following the steps implemented in MatLab:

1. Solve the linear optimization problem with the orginial k_{cat} (p_{orig}) values and record the resulting flux through the objective function (F_{orig}).
2. Multiply the k_{cat} of the ith reaction with the factor 1.1 (p_{pert}), which corresponds to a perturbation of 10 %, keep the k_{cat} values for the other reactions unchanged.
3. Solve the linear optimization problem with the one perturbed k_{cat} (p_{pert})and record the flux through the objective function (F_{pert}).
4. Calculate the flux control coefficient for the ith enzymatic step according to formula 2.7 (see section 2.3.3)

$$FCC_i = \frac{\ln(p_{pert} - p_{orig})}{\ln(F_{pert} - F_{orig})} \tag{3.9}$$

5. Repeat steps 2-4 for every enzymatic reaction with protein constraint.
6. Repeat steps 1-5 for different objectives such as growth rate or production rates

4 Materials

The model developed this thesis relies on experimental data from a high variety of sources, most of it taken from databases where specific parameters of biological interest are collected.
The values for the enzyme kinetics (k_{cat}) were mostly taken from the Brenda database or identified by peer-reviewing scientific literature (see section 4.1). The weight of the polypeptides was collected from the Saccharomyces Genome Database, with expection of the heterologous proteins for the downstream pathways from tyrosine and phenylalanine. The information about the subunit structures of the enzymes are mainly extracted from uniprot, paxDB or by peer-reviewing literature (see section 4.2). Abundance data from pax DB was used to estimate the amount of cellular protein covered by the model (see section 4.3). The total amount of protein in the cell dependent on growth conditions was taken from several research articles.
Experimental chemostat data was used to validate the correctness of the main exchange fluxes simulated by the model (see section 4.4.1). The batch data used for comparison with the simulations from the dynamic FBA (see section 3.3) was taken from yet unpublished experiments (see section 4.4.2).
A complete list of all individual protein data used in this thesis is attached in the appendix.

4.1 Turnover values

The most of the k_{cat} values were retrieved from the enzyme database BRENDA (**Br**aunschweig **En**zyme **Da**tabase) [37]. Determining the kinetic properties in own experiments was beyond the scope of this work.
BRENDA offers information about enzyme properties from several organisms collected from various sources. The database was accessed by the EC number related to every reaction. The EC number defines the type of enzymatic reaction and therefore all isoenzymes exert the same EC number (see 3.1.6).
For *S. cerevisiae*, data availability and consistency varied a lot from enzyme to enzyme. If the information was not available for *S. cerevisiae*, the amino acid sequences of the enzymes available were compared via pblast on uniprot and judged on their degree of similarity.
Since the kinetic performance of enzymes depends on physical conditions such as pH and temperature, k_{cat} values were preferably chosen for a pH around 7 because it is the usual physiological condition and a temperature of 30 °C.
Often BRENDA offers a variety of k_{cat} values for a single reaction, of which most were obtained from mutated enzymes with a changed amino acids sequence, decreasing or increasing the k_{cat}.

As suggested in [38], the highest value was chosen, which might underestimate the actual amount of protein needed for a certain flux. If the median would be used as suggested in [5], the values might be more prone to experimental errors, since the more data sources would be involved.

If the suggested k_{cat} values differed in a substantial amount, the individual sources were investigated for conditions and strains used. Even though it is assumed that the *invitro* experiments mimic the physiological conditions found in the cell, some important co-factors, maybe not yet identified, might be missing.

In addition to BRENDA, other sources were employed. Research articles collecting enzymatic data such as [39] or individual articles covering information about the enzymes of interest were used in an extensive amount as well.
These articles usually did not include the k_{cat} value directly, but the specific activity a of the enzymes. Whereas the k_{cat} is usually in the units [1/S], giving the number of substrate molecules "turned over" per molecule of enzyme per second, the specific activity usually relates the amount of substrate (µmol) per weight of enzymes (mg) per minute, [µmol/(mg · min)]. The k_{cat} value can easily be calculated via formula 4.1, using the molecular weight w of the enzyme and dividing by 60 to transform from minutes to seconds.

$$k_{cat} = \frac{a \cdot w}{60} \tag{4.1}$$

Formula 4.1 shows the importance of the accurate molecular weight of the enzyme when calculating the k_{cat} from the specific activity. If the enzyme in its physiological state is a tetramer, consisting of four subunits, the molecular weight of the enyzme complex is the sum of all of its subunits. Sometimes the structure of enzymes was still not known entirely.
For the special case of the pentafunctional enzyme encoded by ARO1, which catalyzes five reactions in the chorismate pathway (see 2.2), the reactions were lumped together and the lowest k_{cat} from the five reactions was chosen, since it defines the required abundance of the protein for a certain flux.

4.2 Protein Weight and Structure

Data on the polypeptide weights was taken from the Saccharomyces Genome Database [34]. For the information about the underlying subunit structure the Universal Protein Resource (UniProt) Database [40] was the main source but as for the k_{cat} values, individual research articles had to be used for enzymes with uncertain complexity. For some enzymes only a suspected subunit composition could be found.

4.3 Protein Abundance and total Content

The data on protein abundance to determine the fraction of the total protein that is covered by enzymes in the AAA model was taken from paxDB [41]. This database offers protein abundance in ppm (parts per million) integrated and weighted from several studies under several conditions. For this work the abundance from batch experiments on minimal medium were used. The abundances of the proteins from *S. cerevisiae* used in the model were added and resulted in a value of $AAAmodel_P = 0.235$.

For the heterologous proteins the weight was estimated by multiplying the length of the amino acids sequences with the average weight of amino acids (0.11 kDa/amino acid).

The total protein in the cell differs also from condition to condition and depends also on the growth rate (see section 2.1.1). The total protein content of the cell was set to 0.55 $g_{protein}/g_{biomass,dw}$ during growth on glucose. For the growth on ethanol, a protein content of 0.49 $g_{protein}/g_{biomass,dw}$ was set, which might mark an overestimation [42].

4.4 Experimental data

4.4.1 Experimental Chemostat Data

Experimental data von [43] was used to validate the simulation for the prediction of shifts in the physiology of the metabolism. The steady-state fluxes of the consumption of glucose and oxygen as well as the excretion of ethanol, CO_2 and acetate were measured in dependency of the growth rate (Table 4.1).

Table 4.1: Experimental chemostat data for consumption and excretion fluxes as functions of the growth rate, calculated from [43]. Growth rate in [1/h], all other fluxes in [mmol/(h · g_{dw})]

Growth rate μ	r_{O_2}	r_{CO_2}	$r_{glucose}$	$r_{ethanol}$	$r_{acetate}$
0.025	0.8	0.8	0.3	0	0
0.05	1.3	1.4	0.6	0	0
0.1	2.5	2.7	1.1	0	0
0.15	3.9	4.2	1.7	0	0
0.2	5.3	5.7	2.3	0	0
0.25	7	7.5	2.8	0	0
0.28	7.4	8	3.4	0.11	0.08
0.3	6.1	8.8	4.5	2.3	0.41
0.35	5.1	14.9	8.6	9.5	0.62
0.4	3.7	18.9	11.1	13.9	0.6

4.4.2 Experimental Batch Data for Resveratrol Production

The concentrations for the simulation of the dynamics of the batch fermentation were calculated from unpublished data from Mingji Li (minl@biosustain.dtu.dk).

Table 4.2: Experimental batch data for cell density and concentrations of glucose, ethanol and resveratrol.

time h	Cell density g/L	[Glucose] mmol/L	[Ethanol] mmol/L	[Resveratrol] µmol/L
0	0.02754	166.23	0	0
6	0.28084	146.40	49.34	0
8	0.6528	113.09	84.26	0
10	1.4892	59.07	177.59	0.2
12	2.6112	0.005	251.2	0.50
15	3.009	0.05	214.44	0.80
18	4.0494	0.08	165.02	3.20
21	5.4468	0.03	46.94	8.70
22	6.5178	0.05	9.92	12.45
24	6.477	0.023	9.39	11.85
36	6.6708	0	0.67	11.95

5 Results

The AAA model was successfully tested for its ability to simulate certain physiological behaviors in relation to the growth rate, including the crabtree effect under high growth rates with glucose as the substrate (see section 5.1).
By adding the required reactions, it was used to simulate the trends of the dynamic changes of cell density and metabolite concentrations, including the production of resveratrol, according to experimental batch data, showing also the ability to simulate the possibility of growth with ethanol as a carbon- and energy-source (see section 5.2)

With the model being able to produce the desired products in a realistic manner, the protein mass occupied by the recombinant pathways, competing for the available protein pool, was evaluated (see section 5.3).

The enzymatic steps offering the highest impact on cellular growth and production of each product were determined by MCA through calculating the flux control coefficients (see section 5.4).

5.1 Simulating physiological behavior

By including kinetic enzyme parameters and an upper limit on available protein into the FBA framework metabolic behavior for *S. cerevisiae* could be simulated and reproduced (Figure 5.1). Fo this simulation the flux through recombinant pathways was not included. The resulting exchange fluxes are in acceptable agreement with experimental chemostat data, when the acetate excretion is limited to 0.6 mmol/(h $\cdot$ g_{dW}) and the activity of the ATP Synthase complex is increased by 33 %. Considerable deviations occure at high growth rates, displaying an overprediction of CO_2 and ethanol excretion as well as increased glucose consumption (Table 5.1).

Table 5.1: Comparison between the experimental fluxes and the simulated fluxes at a growth rate of 0.4 1/h

Name	Flux, experimental mmol/(h $\cdot$ g_{dw})	Flux, simulated mmol/(h $\cdot$ g_{dw})	Deviation %
O_2 consumption	3.7	2.75	74.3
CO_2 excretion	18.9	23.36	123.6
Glucose consumption	11.1	13.27	123.6
Ethanol excretion	13.9	20.38	146.6

As to be seen in Figure 5.1, at lower growth rates the model simulates consumption and secretion rates in good agreement with experimental data. The metabolic shift from efficient oxidative phosphorylation towards less efficient fermentation is predicted at a growth rate of about 0.28 1/h, shown by the decreased consumption of O_2 and the secretion of ethanol and acetate. By including the basic reactions for gluconeogenesis as the glyoxylate shortcut, the AAA model is able simulate growth on ethanol as well (data not shown). The predicted maximum growth rates for glucose and ethanol as the carbon/energy source were 0.44 1/h and 0.15 1/h, respectively.

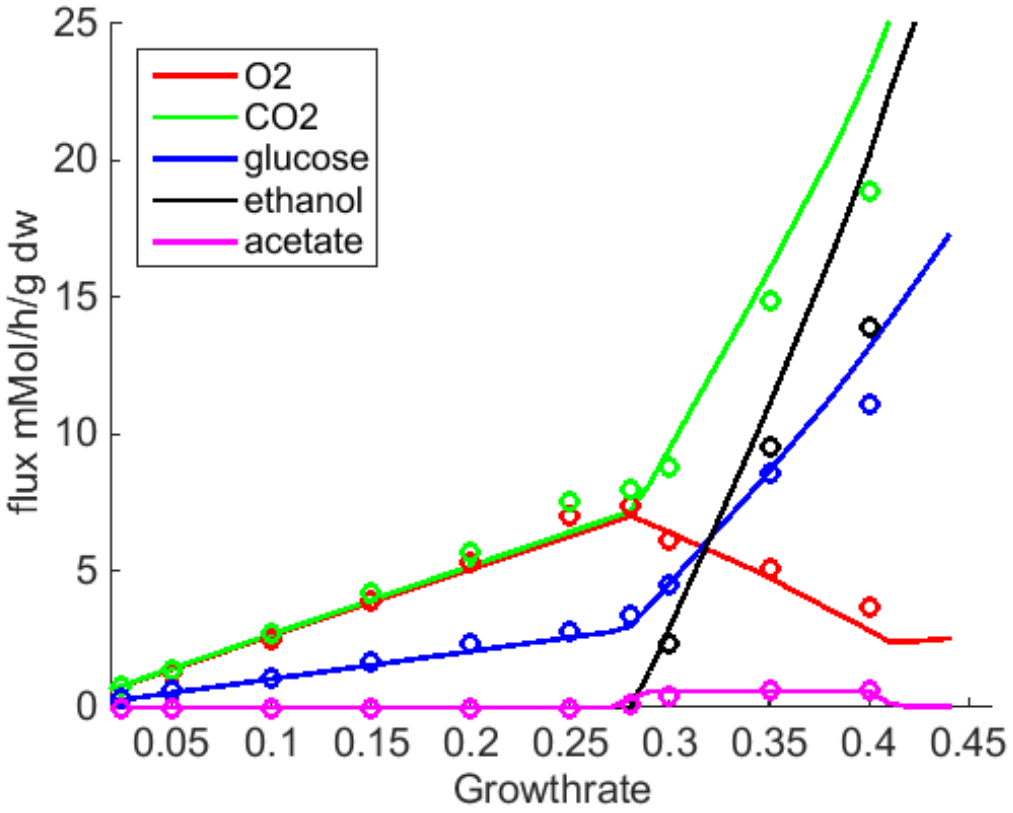

Figure 5.1: Exchange fluxes of O_2, CO_2, glucose, ethanol and acetate in dependency of the growth rate.

5.2 Iterative FBA for resveratrol production in a batch fermentation

To simulate the dynamic changes of the cell density and the concentrations of glucose, ethanol and resveratrol in a batch culture, a FBA was used in an iterative algorithm (see section 3.3). The simulation created by the model with the included recombinant pathway for resveratrol was compared to experimental values provided by Mingji Li (unpublished data, see section 4.4.2). The initial concentrations for calculating the first simulation step were taken from the same dataset.

In the experimental setup, the genes for the resveratrol producing pathway were under the control of strong, constitutively expressed promoters, resulting in a higher ATP demand than in wild-type cells due to protein synthesis. To account for the higher energy demand, the ATP

maintenance was set to 8 mmol/(h · g_{dw}) to fit the experimental data.
The production of resveratrol was coupled to the biomass production with a factor for each growth phase, calculated from the experimental data. In this case it was assumed, that the production correlates with growth, giving a factor of 0.25 µmol/h resveratrol production per g_{dw}/h biomass for glucose growth and 3.5 µmol/h resveratrol production per g_{dw}/h biomass for growth on ethanol. The results of the simulation are displayed in Figure 5.2.

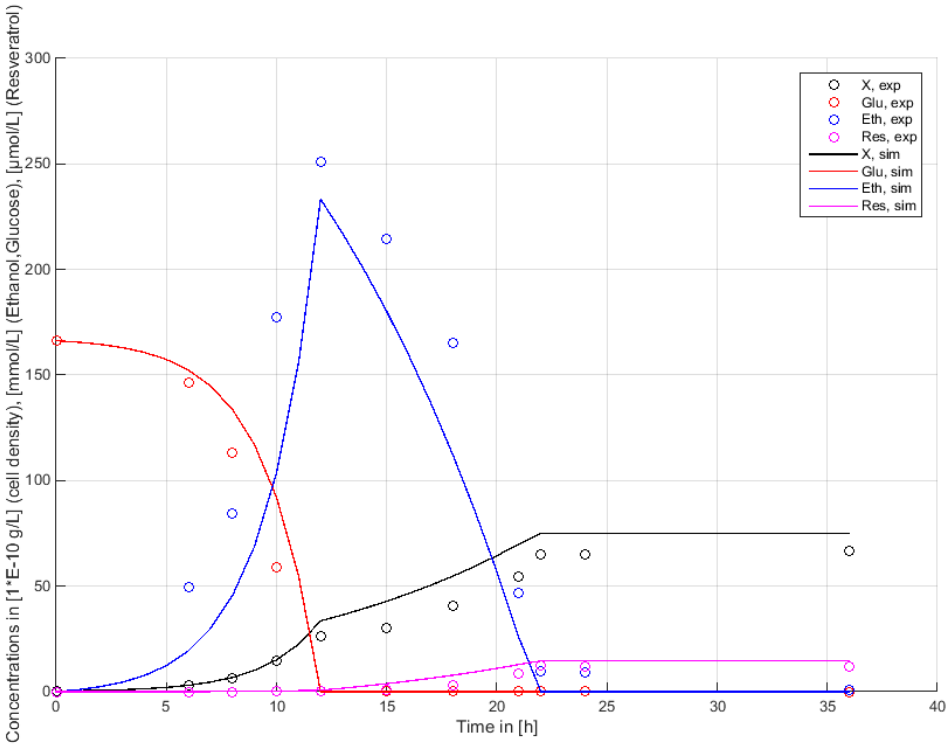

Figure 5.2: Dynamic changes in cell density and concentrations of ethanol, glucose and resveratrol, comparison between experimental and simulated data.

The iterative algorithm with FBA successfully predicts the trends in the dynamic behavior of the concentrations. The model shows the typical phases in aerobic batch cultivations with the exponential growth under glucose excess in combination with ethanol production. When the glucose is consumed, the cells can use the ethanol to further increase the biomass, which is also shown in the simulated predictions.
The simulated growth rate on ethanol is with 0.4 1/h in good agreement with the experimental data, where the growth rate was calculated to be 0.39 1/h. Even though clearly simulating the trends in concentrations, the AAA model underpredicts both the glucose consumption and the ethanol excretion during growth on glucose (0 h - 12 h). This contradicts the findings the results from the simulations of the physiological behavior, where under high growth rates the ethanol production was overpredicted compared to experimental data. This results in an overall lower ethanol concentration in the simulation in comparison to the experimental batch data.
The growth on ethanol (12 h - 22 h) results in a comparable cell density when all ethanol is consumed, but the growth phase itself follows a more linear character compared to the experimental data, even though the growth rates were comparable with 0.098 1/h and 0.082 1/h for

experimental and simulated growth rate, respectively.
The final resveratrol concentrations were comparably similar with 11.95 µmol/L (experimental) to 14.6 µmol/L (simulated).

5.3 Theoretical maximum production yields

The model predicted a theoretical maximum yield for resveratrol production of 241.4 mg_{res}/g_{glc} and 512.9 mg_{res}/g_{eth}. The higher yield on ethanol is in agreement with the experimental data. The highest yields extracted from the experimental data were 1.86 mg_{res}/g_{glc} and 16 mg_{res}/g_{eth} (Unpublished data from Mingji Li, not shown).

For naringenin, the predicted maximum yields for naringenin were 287.9 mg_{nar}/g_{glc} and 611.1 mg_{nar}/g_{eth}. The trend is in agreement with experimental data by [26], were a yield on glucose of 2 mg_{nar}/g_{glc} was achieved and a the yield on ethanol was calculated to be 10 mg_{nar}/g_{eth}.

In both cases the predicted yield was higher under the growth on ethanol, which is in agreement with the experimental data. The values of the theoretical maximum yield are though overestimations, since under these conditions the simulated biomass production was 0, which is presumably not the physiological reality.

5.4 Impact of recombinant pathways on the protein pool

Assuming a constant total protein pool for the cell, the impact of introducing reactions of a heterologous pathway on this protein pool was investigated. By calculating the flux distribution and thereby the protein mass for every enzyme flux for the model with and without the production, the amount of protein "stolen" from the original network by the recombinant pathway was determined. For the comparison the growth on ethanol was used, since it was the phase with the higher resveratrol production compared to the glucose growth. The five highest changes in cellular protein concentration of enzymes of the native *S. cerevisiae* metabolism are shown in Table 5.2. Positive percentages indicate an increase, negative percentages a decrease in individual enzyme concentration compared to biomass production without coupled growth. The reactions leading to the production of tyrosine (TYR, ARO8) and malonyl-CoA (ACC) as the substrates for resveratrol production are increased.
The impact on the total protein pool was determined to be 0.4 % for resveratrol production during ethanol growth, meaning that 0.4 % of the protein pool for the endogenous pathways is "stolen" by the recombinant pathway.

A similar approach was used for simulating the impact of the protein pool for the production of naringenin. The same assumption for ATP maintenance was made. The value of 0.004 mmol/(h*g_{dw}) was used as the factor to connect growth with product excretion. The maximum yield of naringenin on ethanol was higher but for a better comparison with the protein mass changes for the

Table 5.2: Calculated protein masses from the model prediction for coupled production of resveratrol in comparison with growth only (no coupled production).

Name	Caculated mass, no production $ng_{protein}/g_{biomass,dw}$ *$g_{protein}/g_{biomass,dw}$	Calculated mass, resveratrol production $ng_{protein}/g_{biomass,dw}$ *$g_{protein}/g_{biomass,dw}$	Change %
TYR	45.7	46.8	2.4
ARO8, Tyrosine	4849.4	4965.9	2.4
ACC	182.16	185.0	1.57
TKL1 (reverse)	0.813	0.733	-9.88
TAL (reverse)	0.573	0.517	-9.88
$Total_{AAA}$	0.1151*	0.1147*	**-0.4**

resveratrol production case, a similar factor for the coupling was chosen. and the protein mass was compared as described above.

Table 5.3: Calculated protein masses from the model prediction for coupled production of naringenin in comparison with growth only (no coupled production).

Name	Caculated mass, no production $mg_{protein}/mg_{biomass,dw}$ *$g_{protein}/g_{biomass,dw}$	Calculated mass, naringenin production $mg_{protein}/mg_{biomass,dw}$ *$g_{protein}/g_{biomass,dw}$	Change %
TYR	45.7	47.2	3.26
ARO8 (Tyr)	4849.4	5007.5	3.26
ACC	182.16	186.36	2.3
ARO7	182.16	176	1.05
TKL1 (reverse)	0.813	0.725	-10.83
TAL (reverse)	0.573	0.511	-10.83
$Total_{AAA}$	0.1151*	0.1150*	**-0.16**

As for the production of resveratrol, the enzymes catalyzing the production of the substrates for the downstream pathway towards naringenin are increased the most (TYR, ARO8, ACC). Even though the pathway towards naringenin includes an additional catalytic step and therefore an additional enzyme compared to the resveratrol pathway and has a higher yield on biomass, the impact on the total protein is with 0.16 % lower than for the resveratrol production with comparable growth-to-production coupling.
Negative percentages were seen for the backwards reactions of TAL and TKL1, both reactions of the pentose-phosphate-pathway and responsible for biomass formation during growth on ethanol.

To investigate the predicted protein mass for achieving the theoretical maximum yields for resveratrol and naringenin on ethanol, the simulation was repeated with the objective on the

production of each product individually. The simulated results are displayed in Table C.1 in the appendix.
As expected, the individual protein masses generally shift away from the protein responsible for growth with only a few exceptions, as no growth was predicted under the maximum yield conditions for resveratrol and naringenin production. The remaining percentages for the endogenous metabolism are 32.7 % and 57.4 % for resveratrol and naringenin production, respectively. The proteins of the recombinant pathways as well as the proteins for the production of their precursors are in higher abundance compared to the biomass optimized prediction.
It is notable, that the predicted individual protein masses are always higher in the naringenin-producing simulation. When comparing the final enzymatic steps, VINST for resveratrol production takes with 77.88 $mg_{protein}/g_{biomass,dw}$ way more of the total protein mass than CHS and CHI with 48.26 $mg_{protein}/g_{biomass,dw}$ and 1 $ng_{protein}/g_{biomass,dw}$, respectively, for naringenin production.

5.5 Metabolic control analysis and flux control

To investigate which enzymatic step the highest control over the flux through a certain pathway exerts, MCA was performed (see section 3.4). The production of each product was assumed as related to growth and the main production occurred during growth on ethanol. The MCA was therefore performed by simulating these conditions. The growth rate with the coupled production of either resveratrol or naringenin was chosen as the objective to be optimized.
The results of the MCA are displayed in Table 5.4. The kinetic protein parameter is displayed as the specific activity, to show the relation between turnover number and weight of the enzyme.
The ATP Synthase, a crucial enzyme for ATP production under growth on ethanol, exerts the highest FCC, since it is the main source of ATP during growth on ethanol. Other enzymes of the respiratory chain (CBC, SDH) show a high control over the flux as well as they are also involved in the ATP production. ARO8 marks the final reactions in the synthesis of tyrosine and phenylalanine and influences therefore both, the growth and the synthesis of resveratrol or naringenin.
The other enzymes displayed are part of the gluconeogenesis and therefore important for growth on ethanol as well.

Furthermore, the FCCs of the direct flux towards each of the product were calculated and are displayed in table 5.5 and table 5.6. Note that the FCCs of two different pathways are not to be compared in total value, but rather their relation to the other FCCs of the same pathway. The most control over the flux for both production pathways is exerted by the final steps towards the final product (VINST for resveratrol, CHS for naringenin) as well as the final step in the synthesis of the precursor tyrosine (ARO8). The model simulates the production of the desired products only from tyrosine, due to the optimization algorithm.
The flux is also to some extend controlled by the activity of enzymes of the respiratory chain (ATP Synthase, CBC, SDH) since they are the main source of ATP required for the product

Table 5.4: Results of MCA for growth and coupled production of resveratrol and naringenin during ethanol growth, ten highest shown.

Name	$FCC_{Resveratrol}$	$Flux_{Resveratrol}$ mmol/(h · g_{dw})	$FCC_{Naringenin}$	$Flux_{Naringenin}$ mmol/(h · g_{dw})	Specific Activity mmol/(mg · min)
ATP Synthase	0.450	18.22	0.452	18.29	11
ARO8 (Phe)	0.133	0.02	0.134	0.02	0*
ARO8 (Tyr)	0.123	0.014	0.124	0.015	0*
CBC	0.084	8.12	0.084	8.15	26
ALD2	0.052	3.92	0.052	3.94	21
ICL	0.039	1.44	0.039	1.45	10
KGD	0.033	0.94	0.033	0.94	8
SDH	0.026	2.5	0.026	2.75	25
ACS	0.024	4.18	0.024	4.19	48
MLS	0.011	1.44	0.011	1.45	36

synthesis. The fifth-highest FCC is in both cases exerted by ACC as it synthesises malonyl-CoA, an important precursor that gets included and the downstream end of the recombinant pathways.

Table 5.5: FCCs for production of naringenin during ethanol growth, 13 highest shown. * indicates very low specifc activities.

Name	FCC	Flux mmol/(h · g_{dw})	Specific Activity mmol/(min · $mg_{protein}$)
ARO8 (Tyr)	0.422	0.08485	0*
CHS	0.408	0.08485	0*
ATP Synthase	0.063	254.537	11
CBC	0.012	114.542	26
ACC	0.012	0.25454	6
ALD2	0.010	0.7533	21
ICL	0.008	0.28282	10
4CL	0.006	0.08485	4
ACS	0.005	0.82018	48
ARO1	0.004	0.08485	6
TYR	0.004	0.08485	6
SDH	0.003	0.28282	25
ARO1	0.003	0.08485	8

Table 5.6: FCCs for production of resveratrol during ethanol growth, 13 highest shown. * indicates very low specifc activities.

Name	FCC	Flux mmol/(h · g_{dw})	Specific Activity mmol/(min · $mg_{protein}$)
VINST	0.666	0.04727	0*
ARO8 (Tyr)	0.233	0.04727	0*
ATP Synthase	0.035	141.817	11
CBC	0.007	0.63818	26
ACC	0.007	0.14182	6
ALD2	0.005	0.40969	21
ICL	0.004	0.15757	10
ACS	0.003	0.45697	48
4CL	0.003	0.04727	4
SDH	0.002	0.15757	25
ARO1	0.002	0.04727	6
ARO2	0.002	0.04727	8
TYR	0.002	0.04727	6

6 Discussion

In the following chapter the results of this thesis will be put into perspective of the assumptions that were made and the general limitations of the created metabolic network model. The different parts of the thesis will be discussed separately.

6.1 General limitations and assumptions

The results of the simulations of any kind of mathematical modelling approach are in general highly dependent on the data values these frameworks are based on. In the case of models simulating biological properties, especially concerning metabolism and physiology, experimentally determined parameters are of outmost importance and define the liability of the model predictions.
The main method used in this work was FBA, which in itself inherits already the assumptions of a steady state for metabolism as well as a neglectable dilution during growth.

The model created for this thesis relies on the kinetic parameters, subunits structure and correct molecular weight of more than 60 enzymes. For many of these the underlying structure and thereby molecular weight and number of catalyzing active sites is known, but for some the molecular interactions, also with other enzymes, is not fully understood and might differ from condition to condition. Even though experimental methods for investigating the kinetic properties of enzymes seek to mimic the conditions as to be found *in vivo*, the cellular composition and molecular interactions remain complex and important affecting factors might not be identified yet. The modelling approach used in this thesis does not take into account the regulatory mechanisms that influence the enzyme activity or the expression level in relation to the growth conditions like the kind of carbon source or the supply of nutrients and thereby the growth rate. These conditions play also an important role in other factors influencing the predictions by the model. The biomass composition, most importantly the cellular protein content, changes in dependency of the environmental conditions. This includes the protein abundances the $AAAmodel_P$-factor is based on (see section 3.2.2).
For some conditions, the availability of experimental data is also very limited.

Since the investigation and integration of all the influencing factors was beyond the scope of this thesis, the results of this work are based on several assumptions concerning the experimental parameters.
With the central carbon metabolism, oxidative phosphorylation and the biosynthesis of tyrosine

and phenylalanine the model only includes a small part of the overall metabolism of *S. cerevisiae*. Other reactions are concentrated in only a few reactions taking care of the biomass synthesis. Not all exchange reactions with the environment *S. cerevisiae* can perform are included. As described in section 3.2.2, the fraction of proteins covered in the model was estimated to be about 23.5 %. The kinetics of the enzymes in the remaining 76.5 % are not included in the model but might have a big impact on the behavior. Since there was no protein abundance data for growth on ethanol, the same $AAAmodel_P$-factor was assumed as for the growth on glucose, which is likely to be only a rough estimation. The $AAAmodel_P$-factor is also highly dependent on the expression level of the related proteins, which also changes with the conditions.
The turnover numbers used to put the kinetic constraint on the FBA framework were all assumed to work at half of their maximum reaction speed ($c = 0.5$, see section 3.2.2) under all conditions. This is highly likely to be incorrect, since under certain conditions some enzymes might be inhibited or activated by co-factors and other effectors, which are not covered in the modelling framework used in this thesis. With the regulation not directly covered it was therefore assumed, that the resulting optimized flux distribution is achieved by the cell due to regulation processes that lead this particular optimized flux distribution concerning growth or production of resveratrol or naringenin.

Since there was only very limited data for cellular biomass composition during growth on ethanol, the biomass composition was assumed to be the same as for growth on glucose. As for the protein content, a 10 % lower protein content was assumed during growth on ethanol, but further analysis on the cellular composition of *S. cerevisiae* during growth on ethanol have to be performed to adapt the model and increase the predictive strength.

6.2 Reproducing experimental data

Considering the limitations in availability of data and the assumptions that had to be made because of these, the simulated predictions are in good agreement with the experimental data they were compared to.

The shifts in metabolic strategies by *S. cerevisiae* could successfully be predicted and reproduced (see section 5.1). The decreased oxygen consumption as well as the excretion of ethanol is explainable by the size of the protein complexes of the respiration chain, which would take too much space under high growth rates. Instead, the metabolism switches to the more wasteful (steep increase in glucose consumption) fermentation to fulfill the increased ATP demand under high growth rates, because there are more mol of ATP per gram of protein synthesized in the fermentation pathway compared to the respiratory chain. A more detailed study on predicting metabolic strategies in *S. cerevisiae* with a kinetically-constrained model of the central carbon metabolism was performed by Avlant Nilsson (unpublished data). The AAA model is able to predict the trends in these metabolic shifts as well, showing that the inclusion of further parts of the metabolism, such as the synthesis of aromatic amino acids as presented in this thesis, in the kinetically-constrained framework is possible in principle.

The increased ethanol and CO_2 excretion as well as the increased glucose consumption in comparison to the experimental data indicate too low turnover numbers for the responsible reactions, since more metabolites have to be reacted to sustain the ATP demand for growth in comparison to the experimental data.

The iterative FBA algorithm used for simulating the dynamic batch behavior was able to simulate the trends of the experimental data well after adjusting the non-growth related ATP maintenance to 8 mmol/(h · g_{dw}) (see section 5.2) . In literature the value is assumed to be relatively low at 1 mmol/(h · g_{dw}) [7], but under stress conditions it can reach values of 17 mmol/(h · g_{dw}) in anaerobic growth [44].
The experimental batch data for the resveratrol production were taken from a study (Mingji Li, unpublished data) where the recombinant genes were put under the control of constitutively strongly expressing promoters. The additional protein synthesis puts the cell under certain stress conditions as well, explaining the high non-growth related ATP maintenance.
Even though underpredicting the glucose consumption and ethanol production during growth on glucose compared to the experimental data, which contradicts the findings concerning the shifts in metabolic strategies, where ethanol production was overpredicted during high growth rates. The expression of the recombinant pathway seem to impair the system furthermore, additionally to the higher non-growth related ATP maintenance.
Since the model does not take the molecular mechanism into account that take place during the shift into growth on ethanol and would naturally cause a lag phase, the biomass is overpredicted by the model during this shift. During growth on ethanol, the model predicts the increase in cell density more linearly in comparison with the experimental data. For biomass synthesis on ethanol *S. cerevisiae* uses gluconeogenesis [9], a partly backwards glycolytic pathway. The model uses the same turnover number for both reaction directions, which is likely to be incorrect and suspected to be responsible for the too linear growth.
The coupling of growth rate and production of resveratrol fitted the experimental data well, considering the overprediction of biomass concentration which, due to the coupling, causes the overprediction of the final resveratrol titer. The maximum yield of resveratrol and naringenin was calculated to be higher on ethanol than on glucose, which is in agreement with the experimental data ([26] for naringenin). This justifies the different coupling factors for growth on glucose or ethanol, respectively. It is believed, that the acetyl-CoA pool in the cytoplasm is increased during growth on ethanol [42] which leads to an increased production of resveratrol. Acetyl-CoA is catalyzed into malonyl-CoA, which is an important precursor for either resveratrol or naringenin.

6.3 Impact on the protein pool

By calculating the flux distribution with the kinetically-constrained FBA framework, the mass of each involved enzyme required to enable the flux could be determined. For the coupled production, the enzyme mass distribution is changed towards production of the aromatic amino acids,

as expected. TYR and ARO8 catalyze the last two steps in the synthesis of tyrosine. Because of the linear programming algorithm, the downstream products are exclusively formed from tyrosine. The pathway contains one enzymatic step less towards the final products compared to the synthesis from phenylalanine. Furthermore, ACC is found in higher abundance under production conditions. As mentioned before, it is responsible for the supply of malonyl-CoA and thereby important for the final product synthesis. The shift towards more TYR, ARO8 and ACC follows the forced production of resveratrol or naringenin. The negative percentages, meaning the reduction in protein mass in comparison, for the backwards reactions in TAL and TKL1 show the tradeoff between growth and production. As parts of the pentose-phosphate-pathway, the enzymes are responsible for biomass production under growth on ethanol. In this function they metabolize molecules that are precursors used in the recombinant pathways. Thus, reduction shifts the metabolism away from growth and towards production of the desired compound.

For the maximum yields of resveratrol and naringenin on ethanol, the protein mass distribution shifts away from growth and towards enzymes responsible for the production of the final molecule and its precursors (Table C.1). The higher mass for most of the individual enzymes for the naringenin production is in agreement with the higher calculated theoretical maximum yield of 611.1 mg_{nar}/g_{eth} compared to 512.9 mg_{res}/g_{eth} (see section 5.3).
As described in section 5.4, the final steps of each recombinant pathway take a major fraction of the available protein mass. More of the available protein mass is occupied by the final enzyme, VINST, in resveratrol production. Due to its low turnover number of 0.029 1/s, a lot of enzyme is needed to enable the flux. In comparison, for naringenin production the second last step (CHS) inherits way more protein than the last step (CHI). This is also explainable with the turnover numbers, where CHS as the second last step has a turnover number of only 0.042 and CHI as the last step inherits a turnover number of 833 1/s. This indicates that for naringenin the second last step in the pathway is a limiting step in the final synthesis as described in literature [26]. The total values for the individual enzyme masses represent a unrealistic, only theoretical maximum. Since no biomass is produced and other important parts of the cell are not included in the model, the protein masses are highly likely to be overestimated.

6.4 Flux control coefficients

As expected, enzymatic steps with either a very low specific activity or the ones that carry a high flux exert the highest control over the flux.
For the coupled biomass and product synthesis, the main flux control is on the ATP Synthase and other enzymes of the respiratory chain, which are needed for ATP production under growth on ethanol. Phenylalanine and tyrosine synthesis are both catalyzed by ARO8 in the final step. They are needed for biomass synthesis as well as product formation. The very low turnover numbers for these steps cause therefore a strong impact when perturbed.
No flux notable flux control is exerted by the enzymes of the recombinant pathways, which

is probably due to the rather small coupling factor of 0.0035 and 0.004 for resveratrol and naringenin, respectively.

If only the flux through the pathways towards each of the products is considered, the highest flux control is exerted by the seemingly limiting steps for the formation of the final product, as they inherit the lowest turnover numbers. ARO8 would thereby a promising modification target of the native metabolism of *S. cerevisiae*, but product inhibition of tyrosine and phenylalanine on ARO8 and ARO9 might be challenging due to the complex interactions and high control [9]. ATP Synthase and other enzymes of the respiratory chain still exerts the third highest control over the flux, showing the still high demand of ATP required for the production.
Although being a good indicator for flux control, there are several limitations that have to be considered interpreting results of MCA. FCCs do not take into account if the enzyme is a regulatory or a control enzyme. The influence of activators and inhibitors is not described, for example the feedback inhibition of the final product on the first step of the pathway, as for the aromatic amino acid synthesis [9]. This means the first enzyme is a regulatory enzyme, but usually has only a small FCC. Furthermore, FCCs only have little prediction value, since they only apply under steady-state conditions, which may be a good approximation, but give no information about the dynamics of the system. Because of the nonlinear character of enzymatic reaction kinetics, large perturbations of the activities are hard to describe in the MCA-approach.

7 Conclusion And Future Aspects

In this thesis the kinetically-constrained FBA framework was successfully implemented and used to simulate the flux distribution in a recombinant metabolic network for *S. cerevisiae*. By including kinetic parameters and a constraint on the total available protein mass, the model is able to simulate metabolic shifts and the dynamic metabolite changes in batch fermentations as well as the synthesis of products derived from aromatic amino acids.
The impact of recombinant pathways competing for the same protein pool as the native metabolism can be evaluated to certain extends. Furthermore the protein distribution for maximum yields using ethanol as the carbon source could be estimated. Finally, by employing the protein kinetics, the enzymatic steps offering the highest control over the flux towards biomass and product synthesis could be calculated.
As discussed, the results are likely to be biased by the limitations and the assumptions of the model in this work. Further analysis of recombinant systems concerning their enzyme activities, gene expressions and biomass composition and actual protein abundances need to be carried out to increase the predictive strength and liability of these models.

Nonetheless, kinetically-constrained models can offer a helpful and rather easily implementable addition to standard FBA methods to constrain the solution space further towards more realistic predictions under certain conditions.
The shown method might be used to evaluate the protein impact of recombinant pathways on the system and thereby identifying which enzymatic step is a suitable target for optimization. Furthermore, the method could be used to rank several pathways leading to the same product by their impact on the protein pool, helping to decide between certain enzymes or even whole pathways against others. A pathway with fewer steps might have a higher impact on the total protein pool because its enzymes exert too low turnover numbers and therefore more molecules would be needed to enable the flux.

The next step in this development would be to investigate and include the kinetic enzyme parameters of the whole system on a genome scale. Furthermore the inclusion of other metabolite concentrations by metabolomic analysis and thereby the addition of thermodynamic constraints are suspected to increase the validity of these models. The introduction of recombinant pathways into a already existing network might impair the system on several levels which are not yet covered in detail in metabolic models. Complex regulation mechanisms, protein degradation and signalling pathways play an important role in further optimization of biological systems and thereby an important role in creating efficient cell factories for establishing a sustainable bio-based economy.

Bibliography

[1] J. Nielsen. Metabolic engineering. *Applied Microbiology and Biotechnologyl*, 55(3):263–283, 2001.

[2] A. Varma and B. O. Palsson. Metabolic flux balancing - basic concepts, scientific and practical use. *Bio-Technology*, 12(10):994–998, 1994.

[3] Q. K. Beg, A. Vazquez, J. Ernst, M. A. de Menezes, Z. Bar-Joseph, A. L. Barabasi, and Z. N. Oltvai. Intracellular crowding defines the mode and sequence of substrate uptake by escherichia coli and constrains its metabolic activity. *Proceedings of the National Academy of Sciences of the United States of America*, 104(31):12663–12668, 2007.

[4] C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis. Thermodynamics-based metabolic flux analysis. *Biophysical Journal*, 92(5):1792–1805, 2007.

[5] R. Adadi, B. Volkmer, R. Milo, M. Heinemann, and T. Shlomi. Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. *Plos Computational Biology*, 8(7), 2012.

[6] Anastasia Krivoruchko and Jens Nielsen. Production of natural products through metabolic engineering of saccharomyces cerevisiae. *Current Opinion in Biotechnology*, 25:7–15, 2015.

[7] G. Stephanopoulos. Metabolic engineering. *Biotechnology and Bioengineering*, 58(2-3):119–120, 1998.

[8] R. Agren, L. M. Liu, S. Shoaie, W. Vongsangnak, I. Nookaew, and J. Nielsen. The raven toolbox and its use for generating a genome-scale metabolic model for penicillium chrysogenum. *Plos Computational Biology*, 9(3), 2013.

[9] Dan G. Fraenkel. *Yeast Intermediate Metabolism*. CSH Press, 2011.

[10] P. J. Boynton and D. Greig. The ecology and evolution of non-domesticated saccharomyces species. *Yeast*, 31(12):449–462, 2014.

[11] F. Ferrezuelo, N. Colomina, A. Palmisano, E. Gari, C. Gallego, A. Csikasz-Nagy, and M. Aldea. The critical size is set at a single-cell level by growth rate to attain homeostasis and adaptation. *Nature Communications*, 3, 2012.

[12] Ian W. Dawes. Stress responses. In J. Richard Dickinson and Micheal Schweizer, editors, *The Metabolism and Molecular Physiology of Saccharomyces cerevisiae*, chapter 9, pages 376–438. CRC Press, 2004.

[13] Frans M. Klis, Piet De Groot, Stanley Brul, and Klaas Hellingwerf. Molecular organization and biogenesis of the cell wall. In J. Richard Dickinson and Micheal Schweizer, editors, *The Metabolism and Molecular Physiology of Saccharomyces cerevisiae*, chapter 5, pages 117–139. CRC Press, 2004.

[14] Ulrik Schulze. Anaerobic physiology of saccharomyces cerevisiae. *Ph.D. Thesis, Lungby, Technical University of Denmark*, 1995.

[15] U. Schulze, G. Liden, J. Nielsen, and J. Villadsen. Physiological effects of nitrogen starvation in an anaerobic batch culture of saccharomyces cerevisiae. *Microbiology-Uk*, 142:2299–2310, 1996.

[16] A. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. Life with 6000 genes. *Science*, 274(5287):546–&, 1996.

[17] D. Petranovic and J. Nielsen. Can yeast systems biology contribute to the understanding of human disease? *Trends in Biotechnology*, 26(11):584–590, 2008.

[18] S. J. Ha, J. M. Galazka, S. R. Kim, J. H. Choi, X. M. Yang, J. H. Seo, N. L. Glass, J. H. D. Cate, and Y. S. Jin. Engineered saccharomyces cerevisiae capable of simultaneous cellobiose and xylose fermentation. *Proceedings of the National Academy of Sciences of the United States of America*, 108(2):504–509, 2011.

[19] M. Kavscek, M. Strazar, T. Curk, K. Natter, and U. Petrovic. Yeast as a cell factory: current state and perspectives. *Microbial Cell Factories*, 14, 2015.

[20] F. David and V. Siewers. Advances in yeast genome engineering. *Fems Yeast Research*, 15(1), 2015.

[21] P. J. Westfall, D. J. Pitera, J. R. Lenihan, D. Eng, F. X. Woolard, R. Regentin, T. Horning, H. Tsuruta, D. J. Melis, A. Owens, S. Fickes, D. Diola, K. R. Benjamin, J. D. Keasling, M. D. Leavell, D. J. McPhee, N. S. Renninger, J. D. Newman, and C. J. Paddon. Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proceedings of the National Academy of Sciences of the United States of America*, 109(3):111–118, 2012.

[22] A. R. Brochado and K. R. Patil. Overexpression of o-methyltransferase leads to improved vanillin production in baker's yeast only when complemented with model-guided network engineering. *Biotechnology and Bioengineering*, 110(2):656–659, 2013.

[23] J. Nielsen and M. C. Jewett. Impact of systems biology on metabolic engineering of saccharomyces cerevisiae. *Fems Yeast Research*, 8(1):122–131, 2008.

[24] J. Richard Dickinson. Nitrogen metabolism. In J. Richard Dickinson and Micheal Schweizer, editors, *The Metabolism and Molecular Physiology of Saccharomyces cerevisiae*, chapter 4, pages 104–116. CRC Press, 2004.

[25] S. Y. Shin, S. M. Jung, M. D. Kim, N. S. Han, and J. H. Seo. Production of resveratrol from tyrosine in metabolically engineered saccharomyces cerevisiae. *Enzyme and Microbial Technology*, 51(4):211–216, 2012.

[26] F. Koopman, J. Beekwilder, B. Crimi, A. van Houwelingen, R. D. Hall, D. Bosch, A. J. A. van Maris, J. T. Pronk, and J. M. Daran. De novo production of the flavonoid naringenin in engineered saccharomyces cerevisiae. *Microbial Cell Factories*, 11, 2012.

[27] K. P. L. Bhat, J. W. Kosmeder, and J. M. Pezzuto. Biological effects of resveratrol. *Antioxidants & Redox Signaling*, 3(6):1041–1064, 2001.

[28] H. P. J. Bonarius, G. Schmid, and J. Tramper. Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. *Trends in Biotechnology*, 15(8):308–314, 1997.

[29] R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer. Multidimensional optimality of microbial metabolism. *Science*, 336(6081):601–604, 2012.

[30] E. Simeonidis and N. D. Price. Genome-scale modeling for metabolic engineering. *Journal of Industrial Microbiology & Biotechnology*, 42(3):327–338, 2015.

[31] K. Zhuang, G. N. Vemuri, and R. Mahadevan. Economics of membrane occupancy and respiro-fermentation. *Molecular Systems Biology*, 7, 2011.

[32] A. Varma and B. O. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type escherichia-coli w3110. *Applied and Environmental Microbiology*, 60(10):3724–3731, 1994.

[33] J. Forster, I. Famili, P. Fu, B. O. Palsson, and J. Nielsen. Genome-scale reconstruction of the saccharomyces cerevisiae metabolic network. *Genome Research*, 13(2):244–253, 2003.

[34] J. M. Cherry, C. Adler, C. Ball, S. A. Chervitz, S. S. Dwight, E. T. Hester, Y. K. Jia, G. Juvik, T. Roe, M. Schroeder, S. A. Weng, and D. Botstein. Sgd: Saccharomyces genome database. *Nucleic Acids Research*, 26(1):73–79, 1998.

[35] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 27(1):29–34, 1999.

[36] P. D. Karp, M. Riley, M. Saier, I. T. Paulsen, S. M. Paley, and A. Pellegrini-Toole. The ecocyc and metacyc databases. *Nucleic Acids Research*, 28(1):56–59, 2000.

[37] Antje Chang, Ida Schomburg, Lisa Jeske Placzek, Marcus Ulbrich, Mel Xiao, Christoph W. Sensen, and Dieter Schomburg. Brenda in 2015: exciting developments in its 25th year of existence. *Nucleic Acids Research*, 43:D439–D446, 2014.

[38] J. R. Karr, J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-Garcia, J. I. Glass, and M. W. Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2):389–401, 2012.

[39] K. R. Albe, M. H. Butler, and B. E. Wright. Cellular concentrations of enzymes and their substrates. *Journal of Theoretical Biology*, 143(2):163–195, 1990.

[40] M. Pruess, M. J. Martin, C. O'Donovan, A. Bairoch, C. Wu, and R. Apweiler. Uniprot - the universal protein resource. *Molecular & Cellular Proteomics*, 3(10):S259–S259, 2004.

[41] M. Wang, M. Weiss, M. Simonovic, G. Haertinger, S. P. Schrimpf, M. O. Hengartner, and C. von Mering. Paxdb, a database of protein abundance averages across all three domains of life. *Molecular & Cellular Proteomics*, 11(8):492–500, 2012.

[42] Patricia De Jong-Gubbels, Peter Vanrolleghem, Sef Heijnen, Johannes P. Van Dijken, and Jack T. Pronk. Regulation of carbon metabolism in chemostat cultures of saccharomyces cerevisiae grown on mixtures of glucose and ethanol. *YEAST*, 11:407–418, 1995.

[43] P. Van Hoek, J. P. Van Dijken, and J. T. Pronk. Effect of specific growth rate on fermentative capacity of baker's yeast. *Applied and Environmental Microbiology*, 64(11):4226–4233, 1998.

[44] Cornelis Verduyn, Erik Postma, W. Alexander Scheffers, and Johannes P. Van Dijken. Energetics of saccharomyces cerevisiae in anaerobic glucose-limited chemostat cultures. *Journal of General Microbiology*, 136:405–412, 1990.

[45] Lyris M. F. De Godoy, Jesper V. Olsen, Jürgen Cox, Michael L. Nielsen, Nina C. Hubner, Florian Fröhlich, Tobias C. Walther, and Matthias Mann. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature Letters*, 455:1251–1255, 2008.

[46] L. Jacob, V. Beecken, L. J. Bartunik, M. Rose, and H. D. Bartunik. Purification and crystallization of yeast hexokinase isoenzymes - characterization of different forms by chromatofocusing. *Journal of Chromatography*, 587(1):85–92, 1991.

[47] Haijuan Li and Frank Jordan. Effects of substitution of tryptophan 412 in the substrate activation pathway of yeast pyruvate decarboxylase. *Biochemistry*, 38:10004–10012, 1999.

[48] Hanan L. Messiha, Edward Kent, Naglis Malys, Kathleen M. Carroll, Neil Swainston, Pedro Mendes, and Kieran Smallbone. Enzyme characterisation and kinetic modelling of the pentose phosphate pathway in yeast. *PeerJ PrePrints*, 2014.

[49] A. J. Ganzhorn, D. W. Green, A. D. Hershey, R. M. Gould, and B. V. Plapp. Kinetic characterization of yeast alcohol dehydrogenases - amino-acid residue-294 and substrate-specificity. *Journal of Biological Chemistry*, 262(8):3754–3761, 1987.

[50] C. Ingram-Smith, B. I. Woods, and K. S. Smith. Characterization of the acyl substrate binding pocket of acetyl-coa synthetase. *Biochemistry*, 45(38):11482–11490, 2006.

[51] Hirabaya.T and T. Harada. Isolation and properties of alpha-ketoglutarate dehydrogenase complex from bakers yeast (saccharomyces-cerevisiae). *Biochemical and Biophysical Research Communications*, 45(6):1369–&, 1971.

[52] H. Muratsubaki and K. Enomoto. One of the fumarate reductase isoenzymes from saccharomyces cerevisiae is encoded by the osm1 gene. *Archives of Biochemistry and Biophysics*, 352(2):175–181, 1998.

[53] J. S. Keruchenko, I. D. Keruchenko, K. L. Gladilin, V. N. Zaitsev, and N. Y. Chirgadze. Purification, characterization and preliminary-x-ray study of fumarase from saccharomyces-cerevisiae. *Biochimica Et Biophysica Acta*, 1122(1):85–92, 1992.

[54]

[55] H. Durchschlag, G. Biedermann, and H. Eggerer. Large-scale purification and some properties of malate synthase from bakers-yeast. *European Journal of Biochemistry*, 114(2):255–262, 1981.

[56] S. Devries and L. A. Grivell. Purification and characterization of a rotenone-insensitive nadh - q6 oxidoreductase from mitochondria of saccharomyces-cerevisiae. *European Journal of Biochemistry*, 176(2):377–384, 1988.

[57] J. J. Kessl, K. H. Ha, A. K. Merritt, B. B. Lange, P. Hill, B. Meunier, S. R. Meshnick, and B. L. Trumpower. Cytochrome b mutations that modify the ubiquinol-binding pocket of the cytochrome bc(1) complex and confer anti-malarial drug resistance in saccharomyces cerevisiae. *Journal of Biological Chemistry*, 280(17):17142–17148, 2005.

[58] K. Forster, P. Turina, F. Drepper, W. Haehnel, S. Fischer, P. Graber, and J. Petersen. Proton transport coupled atp synthesis by the purified yeast h+-atp synthase in proteoliposomes. *Biochimica Et Biophysica Acta-Bioenergetics*, 1797(11):1828–1837, 2010.

[59] G. H. Braus. Aromatic amino-acid biosynthesis in the yeast saccharomyces-cerevisiae - a model system for the regulation of a eukaryotic biosynthetic-pathway. *Microbiological Reviews*, 55(3):349–370, 1991.

[60] J. R. Coggins, M. R. Boocock, S. Chaudhuri, J. M. Lambert, J. Lumsden, G. A. Nimmo, and D. D. S. Smith. The arom multifunctional enzyme from neurospora-crassa. *Methods in Enzymology*, 142:325–342. H8898 year = 1987.

[61] I. Iraqui, S. Vissers, M. Cartiaux, and A. Urrestarazu. Characterisation of saccharomyces cerevisiae aro8 and aro9 genes encoding aromatic aminotransferases i and ii reveals a new aminotransferase subfamily. *Molecular and General Genetics*, 257(2):238–248, 1998.

[62] G. Mannhaupt, R. Stucka, U. Pilz, C. Schwarzlose, and H. Feldmann. Characterization of the prephenate dehydrogenase-encoding gene, tyr1, from saccharomyces-cerevisiae. *Gene*, 85(2):303–311, 1989.

[63] Yechun Wang, Hankuil Yi, Melissa Wang, Oliver Yu, and Joseph M. Jez. Structural and kinetic analysis of the unnatural fusion protein 4-coumaroyl-coa ligase::stilbene synthase. *Journal Of The American Chemical Society*, 133:20684–20687, 2011.

Appendix

The appendix includes visualizations of the biosynthesis pathway for phenylalanine and tyrosine (see section 2.2), the information about all enzymes included in the model as well as the results of the protein mass distribution for the maximum yield of resveratrol and naringenin on ethanol (see section 5.4).

A Biosynthesis of aromatic amino acids

Appendix A contains the visualization of the synthesis of phenylalanine and tyrosine in *S. cerevisiae* from phosphoenolpyruvate and D-erythrose-4-phosphate via the chorismate pathway. Chorismate marks a branchpoint between the synthesis of tryptophan and the other two aromatic amino acids phenylalanine and tyrosine.

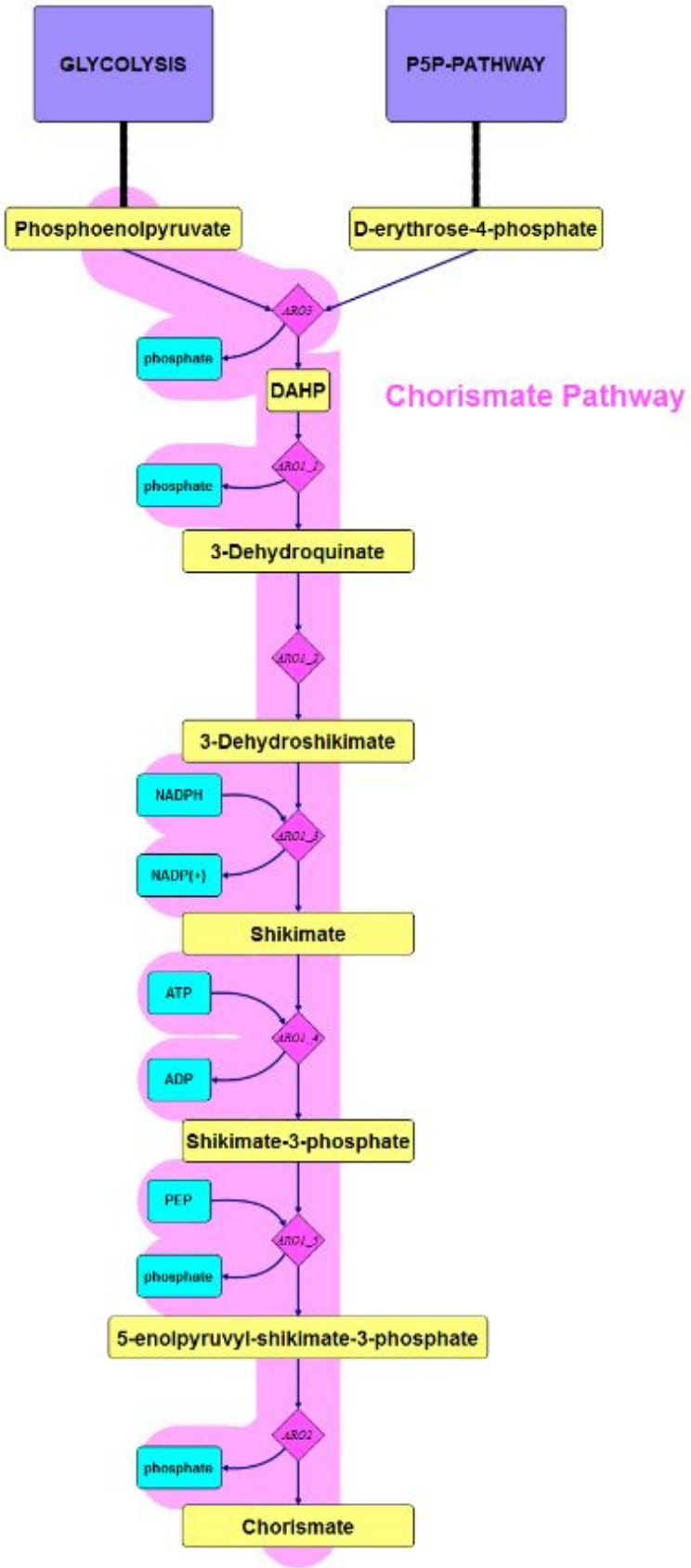

Figure A.1: Chorismate pathway in *S. cerevisiae*. The reactions ARO1_1 - ARO1_5 are catalyzed by the pentafunctional enzyme ARO1.

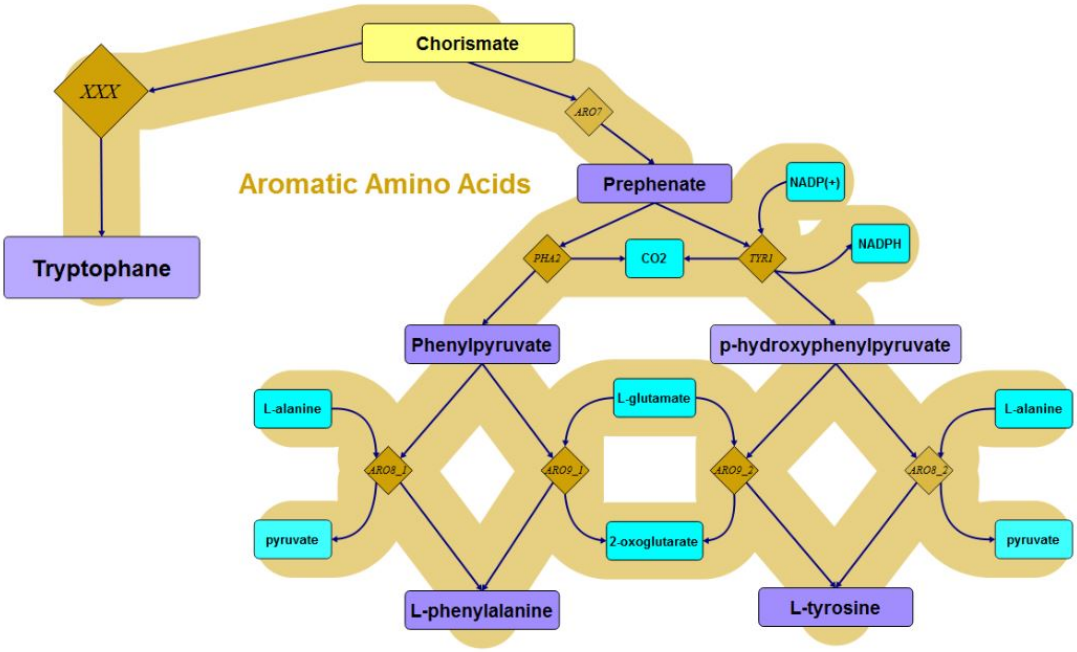

Figure A.2: Synthesis of aromatic amino acids. The synthesis of tryptophan was not covered in detail in this thesis and is therefore shortened to a single reaction named "XXX".

B Enzyme data used in the study

The major source for the turnover numbers (k_{cat}) was the enzyme database Brenda [37] ("Brenda") as well as the collection displayed in [39] ("Albe"). For other enzymes individual research literature was reviewed. The weight was calculated by adding up the weights of the subunits of each enzyme, whose weight was taken from the yeast genome database [34] or calculated from the length of the polypeptide and the average weight for amino acids of 0.11 g/AA. The * attached to k_{cat} sources indicates that the k_{cat} was calculated from the specific activity with the molecular weight for the whole molecule. The + indicates that for very similar reactions no separate k_{cat} was found and therefore the same k_{cat} was assumed for the reactions. If no value for *S. cerevisiae* was found, values from other organisms were used ("Brenda other").
The abundance data to calculate the fraction of the total protein pool was mostly taken from [45] or the integrated abundance in paxDB [41].
All k_{cat} values are in 1/s, all protein weights are in kDa, all abundances are in ppm.

Table B.1: Protein parameters for glycolysis used in this thesis.

Name	EC number	k_{cat}	Weight	Abundance	Source k_{cat}
HXK	2.7.1.1	557.63	107.86	5513	[46]
PGI	5.3.1.9	1379.03	122.58	7190	Albe
PFK	2.7.1.11	850.28	850.28	4763	Albe*
FBP	3.1.3.11	186.2	153.04	14.4	Albe*
FBA	4.1.2.13	142.6	79.22	21823	Albe*
TPI	5.3.1.1	8930	53.58	4010	Albe*
TDH	1.2.1.12	369.42	143	31024	Albe*
PGK	2.7.2.3	354	44.74	21615	Brenda
GPM	5.4.2.11	2356.05	110.44	16460	Brenda*
ENO	4.2.1.11	312	93.82	21615	Albe*
PYK	2.7.1.4	1324.56	218.2	8059.6	Albe*
PDC	1.1.1.1	245.96	245.96	26013.46	[47]*

Table B.2: Protein parameters for Pentose-phosphate-pathway used in this thesis.

Name	EC number	k_{cat}	Weight	Abundance	Source k_{cat}
ZWF	1.1.1.49	1443.3	57.52	512	Albe
SOL	3.1.1.31	10	27.78	219.2	[48]
GND	1.1.1.44	74.96	107.08	3138.59	Albe*
RKI	5.3.1.6	335	113	121	[48]
RPE	5.1.3.1	1300	25.96	153	Brenda
TKL1	2.2.1.1	105.78	147.6	3678	Albe*
TAL	2.2.1.2	75.29	74.06	2265	Albe*
TKL2	2.2.1.1	107.53	150.04	3678	Albe*

Table B.3: Protein parameters for glycerol formation and fermentation used in this thesis.

Name	EC number	k_{cat}	Weight	Abundance	Source k_{cat}
GPD	1.1.1.49	134.2	42.85	898	Albe
GPP	3.1.1.31	833	27.94	-	Brenda
ADH	1.1.1.44	1360	147.4	18810.42	[49]
ALD6	5.3.1.6	22	54.41	7128	Brenda*
ALD2	5.1.3.1	19	55.18	79.4	Brenda*

Table B.4: Protein parameters for the citric acid and the glyoxylate cycle used in this thesis.
For ACO, LSC, MAE and MDH no values for *S. cerevisiae* could be found. For ACO the value was calculated from the specific activity of *Yarrowia lipolytica*(77.7 % identity). For LSC the value from *Acetobacter aceti* was used and for MAE the value from *Escherichia coli*. **For MDH the highest k_{cat} value in Brenda was used, since for all other organisms in Brenda the value was very high as well.

Name	EC number	k_{cat}	Weight	Abundance	Source k_{cat}
ACS	6.2.1.1	65.4	75.48	1096	[50]
PDH	1.2.4.1	3866.67	237.58	2166.2	Albe
PYC	6.4.1.1	240	520.6	1409	Brenda
CIT1	2.3.3.1	293.3	53.36	557	Albe
CIT2	2.3.3.1	293.3	51.41	133	Albe
ACO	4.2.1.3	100	85.37	6264	Brenda other
IDP1	1.1.1.42	57.2^{+}	96.392	765	Albe
IDP2	1.1.1.42	57.2^{+}	93.12	52.9	Albe
IDH	1.1.1.41	187.66	316.28	1048	Albe
KGD	1.2.4.2	30	232.56	502.2	[51]*
LSC	6.2.1.5	201	81.94	883	Brenda other
OSM	1.3.1.6	16.34	50.84	724	[52]*
FUM	4.2.1.2	4075.6	212.64	397	[53]*
MDH	1.1.1.37	4729	71.3	506	Brenda**
MAE	1.1.1.38	134.4	74.38	231	Brenda other
PCK	4.1.1.49	252	243.88	139	[54]
SDH	1.3.5.1	60	142.8	216.23	Brenda
ICL	4.1.3.1	10.4	62.4	546	Brenda*
MLS	2.3.3.9	113.79	188.37	132	[55]

Table B.5: Protein parameters for oxidative phosphorylation used in this study.

Name	EC number	k_{cat}	Weight	Abundance	Source k_{cat}
NDI	1.6.5.3	500	57.26	66.4	[56]
NDE	1.6.5.3	500	62.79	127	[56]
CBC	1.10.2.2	220	501.24	1052.26	[57]
COX	1.9.3.1	1500	225.89	333.21	Brenda
ATPSyn	3.6.3.14	240	1283.6	3780.49	[58]

Table B.6: Protein parameters for aromatic amino acid synthesis used in this study.
For ARO2 and for the pentafunctional ARO1 the value was taken from *Neurospora crassa*(67 % and 51.1 % identity, respectively). The value of the slowest of the five reactions was used. For ARO8 and ARO9, no separate values were found and the same value was assumed for both enzymes. For PHA the value of *E. coli* was employed.

Name	EC number	k_{cat}	Weight	Abundance	Source k_{cat}
ARO3	1.6.5.3	10	41.07	466	[59]
ARO1	1.6.5.3	36.6	349.46	1549	Other, [60]
ARO2	1.6.5.3	21.71	163.36	983	Brenda other*
ARO7	1.6.5.3	361	59.48	113	Brenda
ARO8	1.9.3.1	0.053^{+}	56.17	2431	[61]
ARO9	3.6.3.14	0.053^{+}	58.52	917	[61]
PHA	1.10.2.2	1932	38.22	9.92	Brenda other
TYR	1.9.3.1	5.1	50.92	28.6	[62]

Table B.7: Protein parameters for resveratrol and naringenin synthesis used in this study.
ACC is naturally found in *S. cerevisiae*. The values for PAL, C4H and 4CL were taken from *Arabidopsis thaliana*. The k_{cat} for VINST was taken from *Vitis vinifera*, TYRAL from *Rhodobacter capsulatus*, CHS from *Hypericum androseamum* and CHI from *Glycine max*.

Name	EC number	k_{cat}	Weight	Abundance	Source k_{cat}
ACC	6.4.1.2	25	250.32	3200	Brenda*
PAL	4.3.1.24	115.8	78.73	-	Brenda
C4H	1.14.13.11	1.72	57.8	-	Brenda
TYRAL	4.3.1.23	27.7	53	-	Brenda
4CL	6.2.1.12	3.96	61.05	-	Brenda
VINST	2.3.1.95	0.029	86	-	[63]*
CHS	2.3.1.74	0.042	43	-	Brenda
CHI	4.4.1.6	833	19	-	Brenda

C Protein Mass Distribution

Table C.1 shows the protein mass distribution calculated from the flux distribution and the protein parameters for the maximum yields of biomass, resveratrol and naringenin on ethanol (see section 5.4).

Table C.1: Calculated protein mass distribution from the model for maximum yield for resveratrol and naringenin production on ethanol compared to maximum biomass yield on ethanol.

Name	For max. biomass $ng_{protein}/g_{biomass,dw}$	For max. resveratrol $ng_{protein}/g_{biomass,dw}$	For max. naringenin $ng_{protein}/g_{biomass,dw}$
PGI reverse	11.64	0	0
FBP	176.90	7.195137196	12.91402294
FBA reverse	119.57	4.863278176	8.72874057
TPI reverse	1.30	0.052524878	0.094273043
TDH reverse	174.77	13.55468483	24.32830759
PGK reverse	57.06	4.425542422	7.943080837
GPM reverse	22.78	1.641404749	2.946036748
ENO reverse	146.14	10.52966671	18.8989249
PYK	13.87	0	0
RKI	7.52	2.952891132	5.299927268
RPE reverse	0.45	0.174813567	0.313760024
TKL1	0	12.21508783	21.92396339
TKL1 reverse	1.39	0	0
TAL	0	8.611131268	15.45548663
TAL reverse	0.98	0	0
TKL2 reverse	29.71	24.42987273	43.84738307
GPD	0.61	0	0
ADH reverse	289.61	27.51505417	49.3847485
ALD6	1173.06	64.95178621	116.5771874
ALD2	6382.90	661.0209778	1186.417971
ACS	2929.18	317.3588252	569.6040309
CIT1	249.59	15.57070818	27.94671973
CIT2	6.82	0	0
ACO	1231.92	74.73414675	134.1348274
IDP2	63.89	0	0

Table C.1 Continued: Calculated protein mass distribution from the model for maximum yield for resveratrol and naringenin production on ethanol compared to maximum biomass yield on ethanol.

Name	For max. biomass $ng_{protein}/g_{biomass,dw}$	For max. resveratrol $ng_{protein}/g_{biomass,dw}$	For max. naringenin $ng_{protein}/g_{biomass,dw}$
IDH	911.47	0	0
KGD	4117.62	0	0
LSC	216.54	0	0
OSM	0.87	0	0
FUM	71.21	4.567380665	8.197655875
MDH	32.19	2.639757367	4.737906488
MAE	38.86	0	0
PCK	585.65	84.72068229	152.0589261
ICL	4868.06	525.2487765	942.7304256
MLS	1343.15	144.921523	260.1089906
SDH	3248.92	208.3486814	373.9497355
NDI	886.62	71.17926303	127.7544279
CBC	10374.68	807.7777059	1449.820836
COX	685.74	53.39180076	95.82901913
ATPSyn	54656.53	4213.680516	7562.825472
ARO3	83.35	107.8598363	193.5896929
ARO1	193.77	250.7560621	450.0636264
ARO2	152.71	197.615477	354.6854959
ARO7	2.98	4.32711873	7.766427384
PHA	0.20	0	0
ARO8 (Phe)	10896.61	0	0
TYR	78.20	262.2124284	470.6258164
ARO8 (Tyr)	8300.63	27833.22998	49955.81887
ACC	311.80	788.8816424	1415.905681
TYRAL	0	50.24943169	90.18901184
4CL	0	404.8792652	726.6880364
VINST	0	77881.71514	0
CHS	0	0	48258.81941
CHI	0	0	1.075142742