CHALMERS
UNIVERSITY OF TECHNOLOGY

Predicting metabolic strategies in *Saccharomyces cerevisiae* with a kinetically constrained FBA model

Master's thesis in Complex Adaptive Systems

Avlant Nilsson

Department of Chemical and Biological Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2014

Predicting metabolic strategies in *Saccharomyces cerevisiae* with a kinetically constrained FBA model

Avlant Nilsson

Department of Chemical and Biological Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2014

Predicting metabolic strategies in *Saccharomyces cerevisiae* with a kinetically constrained FBA model
AVLANT NILSSON

Department of Chemical and Biological Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone +46 (0)31-772 1000

Chalmers Reproservice
Gothenburg, Sweden 2014

Predicting metabolic strategies in *Saccharomyces cerevisiae* with a kinetically constrained FBA model

AVLANT NILSSON
Department of Chemical and Biological Engineering
Chalmers University of Technology

Abstract

Metabolism is central to all life. It provides the energy and the building blocks from which the cells are constructed and maintained. Synthetic biologists often make genetic alterations to the enzymes involved in metabolism to improve product yields. Drastic changes in metabolism are linked to several diseases, e.g. cancer. It is therefore desirable to understand and quantitatively predict cell metabolism.

Flux balance analysis (FBA) is a successful mathematical approach for predicting the metabolic activity of a cell. It makes use of the stoichiometry of the biochemical reactions and the rates of nutrient uptake. These relations are used to generate self consistent sets of metabolic fluxes, i.e. rates of metabolic conversion over the reactions. Amongst these it is common to select the set that has the highest growth rate as the predicted set. This has been shown to agree well with experimental data.

One problem with the standard FBA approach is that it does not constrain the flux levels. In the living cell fluxes are constrained by the fact that they are performed by a finite amount of enzymes. The enzyme levels are limited by the amount of energy available for enzyme production and a limited space for enzymes to occupy. It has been shown that taking such limits in to account can improve the prediction powers of FBA.

In this master thesis a modified version of FBA has been developed that uses the fluxes and enzyme kinetic parameters to estimates the weight of the participating enzymes. The total protein weight is constrained to experimentally observed levels. This allows prediction of the maximum growth rate for different substrates and shifts in metabolic strategy to fermento respiration at high growth rates. This might become of use to metabolic engineers in predicting if a potential pathway might decrease cell fitness.

Keywords: FBA, Resource management, Cost, Crowding, Crabtree effect, Galactose, Cell size, *Saccharomyces cerevisiae*, Yeast

Acknowledgments

I want to thank my main supervisor professor Jens Nielsen, PhD, dr.tech. for entrusting me with this project and for interesting discussions resulting in several of the founding ideas of the thesis, the significance of the logarithmic distribution of enzyme turnover values, the alternative cost of a limited protein pool and that enzyme cost might explain the lower growth rate on galactose.

I also want to thank my unofficial supervisors PhD Students Antonio Marras and Amir Feizi for early discussions and references for the projects formulation, for discussions on the biomass equation and the nature of enzyme complexes and their active sites.

I want to thank my opponent Anna Emanuelsson for her engaged reading and professional opposition.

Thanks to Petri-Jaan Lahtvee, PhD for his input on protein half life and for references criticizing the direct link between fluxes and enzyme concentrations. Thanks also to Eduard Kerkhoven, PhD for his suggestions on how a biomass equation can be extracted from a larger model. Thanks to Rasmus Ågren, PhD for developing RAVEN that after sever pruning forms the main source code in the project. Thanks to associate professor Dina Petranovic, PhD for her engaged teaching and for tailoring the bachelor project that provided me with insights in the methods of systems biology. Thanks also to PhD Student Joel Wilsson for early discussions about the ME models.

I want to thank the Swedish state and Chalmers University of Technology for paying for and organizing my university education.

Thanks to Oscar Svensson for three years of successful cooperation with lab work during our bachelor studies.

And special thanks to Irene Lobo and my family and friends for their support.

Avlant Nilsson, Göteborg, June 2014

Contents

1 Background **1**

1.1 What is *Saccharomyces cerevisiae*? 2

1.2 The Functions of *Saccharomyces cerevisiae* by Mass 5

1.2.1 Central Carbon Metabolism 6

1.3 The Crabtree Effect 8

1.4 Growth Described by the Monod Equation 9

1.5 The differences in enzymatic activity amongst proteins 10

1.5.1 The Effect of Concentration on Enzymatic Activity . . 10

1.6 Literature Review 11

1.6.1 Flux Balance Analysis 11

1.6.2 Kinetic Models 12

1.6.3 Whole Cell Models 13

1.6.4 Metabolic Dilution 13

1.6.5 Crowding Models 14

1.6.6 Prediction of Metabolic Strategies 16

1.6.7 The Meaning of Optimality in a Cellular Context . . . 16

1.6.8 Linear Programming 18

1.6.9 Criticism and Problems with the Crowding Models . . 18

1.7 Purpose, Scope, Limitations and Goal 19

1.7.1 Limitations 20

1.7.2 Ethical and Environmental benefits 21

2 Methods **23**

2.1 A Kinetically Constrained Version of FBA 24

2.1.1 Reconstruction of the Stoichiometric Network 24

2.1.2 Biomass Equation 24

2.1.3 Linear Programming Solver and Modeling Framework 25

2.1.4 Predicting and Constraining the Total Amount of Protein . . . 26
2.1.5 Isozymes, Multifunctional Enzymes and Complexes . . 26
2.1.6 Matrix formulation . . . 27
2.2 Growth Dependent Experimental Data as Input . . . 28
2.3 Sensitivity Analysis . . . 29
2.4 Estimation of the c-Parameter from Experimental Data . . . 29
2.5 Estimating the CCMp-Parameter from Proteomics Data . . . 30
2.6 Modeling of Other Constraints . . . 30
2.6.1 Growth and Ribosomes . . . 31
2.6.2 The Relation Between Size and Growth . . . 33
2.6.3 The Effects of Cell Size on Uptake . . . 34
2.6.4 The Effects of Cell Size on the Cell Wall . . . 34
2.7 Data Visualization . . . 34
2.8 Notes on Modeling . . . 35

3 Materials 37

3.1 Kcat Values . . . 37
3.2 Pathway Annotation Data . . . 39
3.3 Protein Weights and Complexes Data . . . 39
3.4 Chemostat Experiment Data . . . 40
3.5 Protein Abundance Data . . . 40
3.6 Data on Cell Size . . . 41

4 Results 43

4.1 Estimation of Model Parameters . . . 43
4.2 Results from the Model . . . 43
4.2.1 Glucose Uptake Saturation . . . 44
4.2.2 The Switch Between 4 Different Metabolic Strategies . 44
4.2.3 The Sum of Fluxes and Specific Activity Explain the Crabtree Effect . . . 47

4.2.4 Lower Predicted Growth Rate for Galactose 49
4.2.5 High Predicted Protein Content in the Crabtree negative *Kluyveromyces marxianus* 50
4.2.6 Prediction of Protein Abundance 50
4.2.7 Stability 51
4.3 Model of other Constraints 53
4.3.1 The Ribosomes Constrain the Space for CCM 53
4.3.2 Uptake Rate as a Function of Size 55
4.3.3 Cell Wall as a Function of Size 56

5 Discussion **59**
5.1 Possible Reasons for the Inaccurate Prediction of Enzyme Concentrations 59
5.2 The Low Enzyme Utilization at Low Growth Rate 60
5.3 Effects of Cell Size 61
5.4 The Crabtree Effect 61
5.5 Probable Effects of Gene Deletion 63
5.6 Practical Considerations With Regards to Modeling 63
5.7 Consequences for Bio Engineering 64

6 Outlook **65**
6.1 Extending the Model to the Whole Metabolic Network 65
6.2 A Tool for Synthetic Biologists 65
6.3 Predicting Izosymes 65
6.4 Improving the Enzyme Concentration Predictions 66
6.5 Maintenance Cost and Protein Degradation 66

7 Conclusion **67**

References **69**

A Model Reconstruction **79**

A.1 Inclusion of the Reactions of Oxidative Phosphorylation . . . 79

A.2 Inclusion of the Galactose Pathway 80

A.3 Extracting Substrate Requirements for Biomass Formation from a Larger Model 81

A.4 A Modular Biomass Equation 83

A.5 Amino Acid Composition 84

B Kcat and the c-Parameter **87**

B.1 The c-Parameter 87

B.2 Kcat Values Used in the Study 89

B.2.1 Glycolysis 89

B.2.2 Pentose Phosphate Pathway 90

B.2.3 Fermentation Pathways 91

B.2.4 TCA Cycle 91

B.2.5 Oxidative Phosphorylation 92

B.2.6 Other 93

B.2.7 Galactose 93

B.3 Cross Verification of Kcat 94

C Experimental Data **97**

C.1 Chemostat Data 97

C.2 Composition Data 98

C.3 Internal Metabolite Concentration 99

D The Number of Subunits in the Complexes **101**

E Functional Annotation of the Proteins **103**

E.1 From Protein Abundance 103

E.2 Choice of Function Categories 104

E.2.1 Lumping of Kegg Categories 105

E.2.2 Text Search Terms Used for Categorization 106

F Limits on Growth From Size and Protein Synthesis 109

F.1 Ribosome Amount Predicted from Growth Rate 109

F.2 Tables Of Cell Sizes Dependency on Growth Rate 112

F.3 Literature Data on Cell Wall for Biomass Estimation 113

G Original Project Description 115

1 Background

The cell is the basic unit of life. Improving the understanding of the cell has therefore potential to improve quality of life through many cell related applications e.g. treatment of diseases, production of biofuels, increasing productivity in agriculture. The cell is also a fascinating topic in its own right. Systems Biology is a field of study that focuses on the interactions between the components of a cell. These interactions give rise to complex behavior.

Making computer models of cells can be a method to improve understanding of their complex behavior. A model integrates knowledge in a structured framework and allows the modeler to generate verifiable predictions of cell behavior. A more elaborate discussion on modeling follows below (Section 2.8).

Saccharomyces cerevisiae (S. cerevisiae) is a popular organism for modeling purposes. Being eukaryote, i.e. cells are compartmentalized like mammalian cells, but at the same time unicellular, gives it an intermediate level of complexity, suitable for studies. As a result there are large amounts of data available. The motivations for using *S. cerevisiae* will be further elaborated below (Section 1.1).

Fast growing organisms, such as *S. cerevisiae*, have a streamlined resource management system. Metabolites, i.e. food, are taken up from the surroundings at high rates and the cell culture can double its weight in less than 2 hours. This is performed by a network of biochemical reactions that transform substrate, e.g. glucose, ammonia, into forms more useful for the cell, e.g. ATP, proteins. These reactions are in general facilitated by enzymes (i.e. catalytic proteins) that enhance the speed of selected reactions several orders of magnitude compared with the spontaneous reaction speed. The enzymes are in turn produced from metabolites by other enzymes. This generates a complex codependency which can be described and predicted by modeling.

There exist an upper limit on the growth rate for all organisms (Section 1.4). Many causes of this limit have been suggested; metabolite uptake rate [80], ribosome translation speed [41], metabolic speed [8], etc. For a dynamic system a single limiting factor is improbable. This since the optimal response to a limit is to decrees the non-limiting factors to avoid over capacity and give space to more of the limiting factor, e.g increase ribosomes and decrease metabolism. This is likely to result in a situations where all

factors reach a balance where each in turn would appear to be limiting to an observer. Because of rigidity, e.g. the discrete amount of DNA strands, such adjustments might however not always be possible. These limits might interact with each other to form a complex limiting landscape (Section 2.6).

It has been observed that yeast switches metabolic strategy at high growth rates, the Crabtree-effect (Section 1.3). A similar shift has also been noticed in fast growing cancer cell, the Warburg effect [74].

Modeling of metabolic networks is commonly performed using flux balance analysis (FBA). These models predicts the fluxes of metabolites over the biochemical reactions. This is done using the equations of the reactions and the uptake rate of metabolites (Section 1.6). The standard version of FBA does however not take the codependency between metabolism and enzymes amount into account [47]. It can therefore predict impossible flux levels in some cases.

Some of the limitations of FBA can be overcome by models that consider the physical nature of the enzymes [8, 17, 1]. These models constrain the fluxes by incorporating physical phenomena e.g. enzymatic speed, weight, transcription, translation. Some of these models are very complex [51] and consist of almost 100 000 equations. Yet they are currently only implemented for the simpler prokaryote organisms, i.e. cells without compartmentalization, e.g. bacteria. Other models are very simplistic [36] and only include the cost of enzymes indirectly by minimizing the cellular activity. The current state of the modeling field will be reviewed in a section below (Section 1.6).

The purpose of this master thesis is to model how the behavior of *S. cerevisiae* changes under different growth rates. This will be done through a set of models involving a modified version of FBA that takes the limit of metabolic fluxes imposed by enzymes into account. The models attempt to tread a middle ground between complexity and simplicity. The purpose and goal of the thesis will be further discussed below together with the scope and limitations (Section 1.7).

1.1 What is *Saccharomyces cerevisiae*?

Saccharomyces cerevisiae (*S. cerevisiae*) is a yeast. It is commonly used in baking and brewing and has been an integrated part of human culture for thousands of years. It is one of the most popular organisms in biotechnology and serves as a work horse in many industrial fermentation processes. Its

current use in industrial service alows increased understanding and discoveries to be directly translated into increased productivity. Both *S. cerevisiae* and humans belong to the eukaryote domain of life. In many cases discoveries in yeast can be translated to discoveries in humans, since many proteins are similar i.e. homologs.

S. cerevisiae is a microrganism and therefore behaves in a comparably simple way. Many aspects of a mammalian system are almost non-existent, for instance cell-cell interactions and cell differentiation [57, 39]. The latter implies that there seldom are several versions of the same protein, i.e. alternative splicing, which means that one gene codes for one protein.

One noticeable property of *S. cerevisiae* is that it is surrounded with a thick cell wall made out of mainly carbohydrates. The thickness of the wall might be as large as 5% of the cell's radius, and it can take up more than 20% of its dry weight [40].

The reproduction of *S. cerevisiae* also carries some peculiarities. Most unicellular organisms divide in half and become two cells of equal sizes, fission. *S. cerevisiae* however reproduces by budding, a process where the mother cell forms a growing bud that becomes the daughter cell. The daughter cells are typically smaller than the mother cells, approximately 60% of the mother cell's size [24]. For each budding the mother cell gets a scar. By counting these scars the age of the cells can be determined.

The yeast cell can reproduce both asexually and sexually. The most common form is asexual reproduction. In this case the cell carries a single copy of DNA, i.e a haploid cell, which is copied to the daughter creating a genetic clone. A less common form involves 2 or more copies of DNA, a polyploid cell in which case the offspring has a different genetic markup than the mother.

S. cerevisiae is a so called model organism, meaning that it is better known, less complex and more convenient for experimental studies and manipulations than other organisms. Some motivations for modeling *S. cerevisiae* are that it has a fully sequenced genome and that the function of most of its proteins are known[1]. It also has relatively few genes, around 6000, compared with the human genome with 30 000 genes [39]. There are many documented experimental procedures for gathering data on yeast resulting in a rich collection of available measurements. There exist data on its regulation, i.e. transcriptomics, protein abundance, i.e. proteomics, and metabolite concentrations, i.e. metabolomics. A consequence of an efficient genetic toolbox

[1] At the time of writing approximately 17% of the genes were annotated with "unknown function" in Saccharomyces Genome Database (SGD) [19].

that permits easy modifications of the genome (gene knock outs, over expression, replacements etc.) as well as libraries of genetically altered strains [39].

In experimental settings yeast is commonly grown in chemostats or in batch cultures. In batch cultures all nutrients are added at the beginning of the experiment and the cells grow at maximum rate until the nutrients are depleted. In the case of the chemostat, cells are given nutrients at the same rate as they are being consumed (Figure 1). Cells are removed at the same rate as they are growing which gives a constant biomass concentration in the reactor. This setting allows growth rate dependent phenomena to be studied.

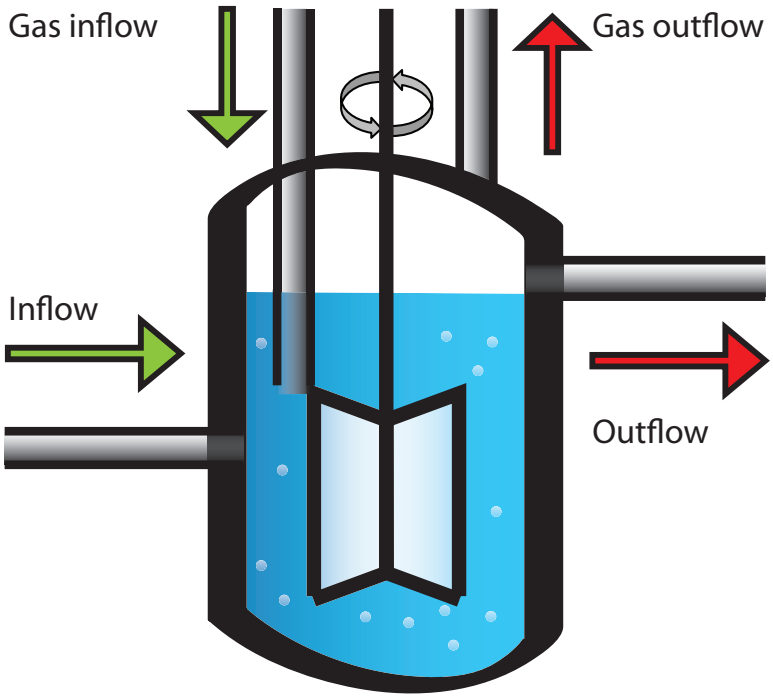

Figure 1: Schematic illustration of a chemostat. There is a constant inflow of water with dissolved nutrients and a constant outflow of water with unprocessed nutrients and cells. If the chemostat is ideally stirred the concentration of cells and nutrients in the outflow is the same as in the reactor. To replace the cells flowing out of the reactor, the specific growth rate of the cells becomes equal to the the dilution rate $\frac{outflow}{reactor\,volume}$. If the dilution rate is higher than the maximum growth rate of the cells, they are washed out.

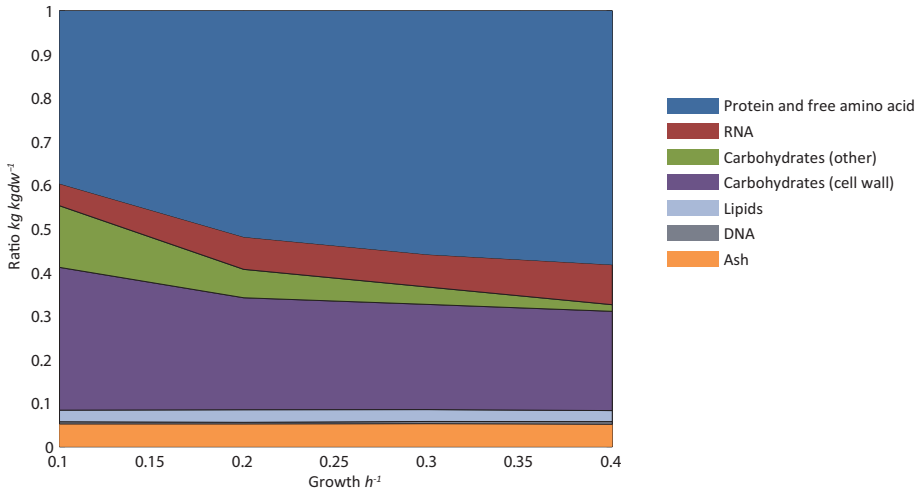

Figure 2: The distribution of the building blocks of a cell for increasing specific growth rate [63].

1.2 The Functions of *Saccharomyces cerevisiae* by Mass

In order to survive and reproduce a cell must uphold a range of functions. A protecting membrane and cell wall, DNA for storing the genetic information, RNA for transcribing and translating it in to proteins that catalyse biochemical reactions. Depending on the growth conditions the cell will prefer a certain cellular composition (Figure 2). As growth rate increases there is in general less stored reserves of carbohydrates and more protein and RNA.

Protein plays the most active role in the cell. The function of each protein is in general very specific, e.g. catalyzing a certain reaction. With the aid of protein abundance measurements, i.e. proteomics, and protein functional annotation and categorization it is possible to get a quantitative picture of the distribution of the total protein mass for a given condition (Figure 3).

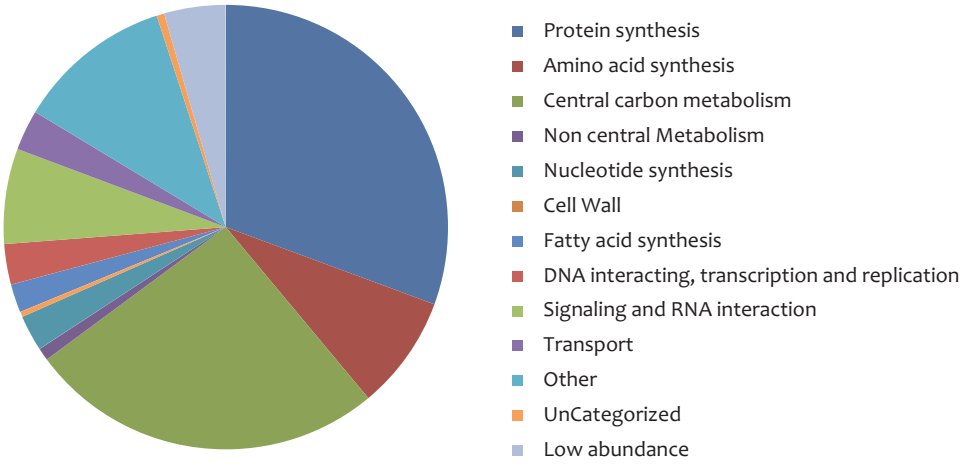

Figure 3: The share of the total protein mass for each functional category at high growth rates and with growth medium including amino acids (Appendix E).

Protein synthesis is the most abundant category at high growth rates, which reflects that self replication is the dominant activity. In addition there is also amino acid synthesis at 8%, reflecting that the cells were grown on media which already contained amino acids.

Aside from constituting a third of the protein of the cell, protein synthesis involves all of the RNA, or 10% of the total weight. Approximately 80% of this is directly in the protein producing ribosomes, rRNA, and 15% indirectly by carrying amino acids to the ribosomes, tRNA. The remaining 5% are dedicated to carrying the information on protein production from DNA, mRNA.

The energy and building blocks required for growth are given by the central carbon metabolism. Approximately one quarter of the proteins are involved in this metabolism, it is further discussed below (Section 1.2.1).

1.2.1 Central Carbon Metabolism

The central carbon metabolism is a cellular function that is highly conserved between organisms. In a network of enzymatic reactions it converts nutrients, e.g. glucose, to a range of metabolites and to energy. The output is used by the other proteins of the cell to carry out their functions. Central carbon metabolism involves several well known metabolic pathways such as glycolysis, the pentos phosphate pathway, the TCA cycle and oxidative

phosphorylation, where the two latter take place in a separate membrane enclosed compartment, i.e. mitochondria.

S. cerevisiae has a particularly inefficient respiration chain, the total central carbon metabolism can produce approximately 16.5 ATP per glucose molecule, compared with literature values for other organisms of up to 30 ATP. This is in part a consequence of a difference in the first step of the respiration chain. In many organisms this is performed by a complex of several proteins seated in the mitochondrial membrane, where the location and structure allows it to pump hydrogen atoms against the proton gradient, ultimately generating ATP. In yeast, however, this function is performed by a single small protein that only associates to the membrane and does not contribute to the gradient.

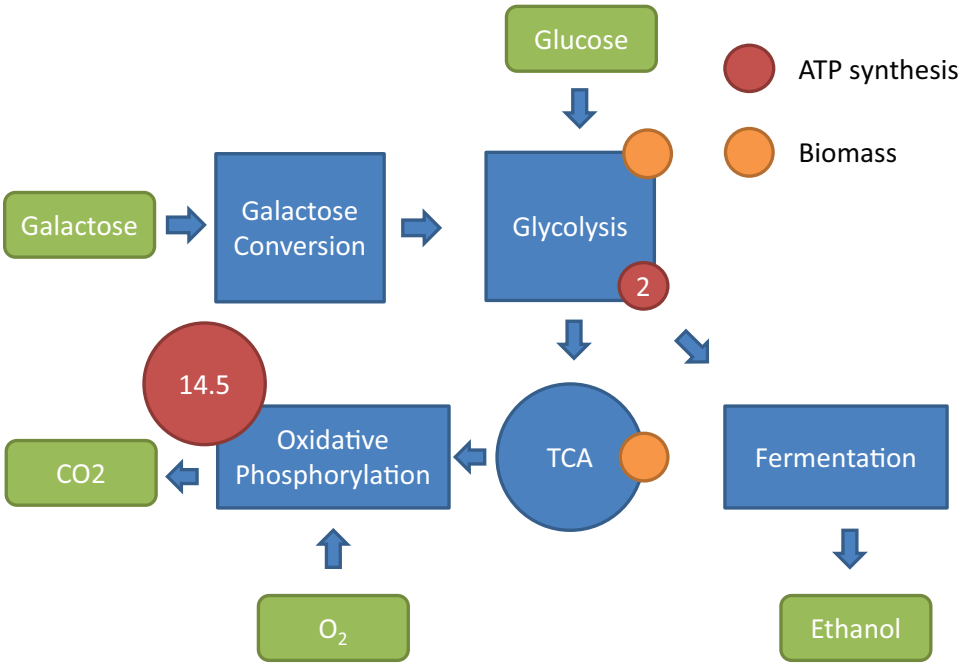

Figure 4: Glucose enters glycolysis, some sub steps provide precursors for the lipid, carbohydrate and amino acid synthesis that make up the biomass equation. Glycolysis results in the net production of 2 ATP and some energy rich products. These energy rich products are either excreted through the fermentation pathway or enter the mitochondria and the TCA cycle in which they generate precursors to the biomass equation and simpler energy rich products. These products are further used for energy production by oxidative phosphorylation, resulting in the production of 14.5 ATP. If Galactose is used as substrate it first has to be converted to enter glycolysis.

1.3 The Crabtree Effect

The Crabtree effect refers to a shift in metabolic strategy from respirative to fermentative growth at high growth rates [74]. The shift in metabolic strategy decreases the biomass yield severely i.e. the biomass/substrate ratio decreases [22]. Instead substrate is converted to ethanol. The underlying reason for this shift has been unknown [22], but recently it has been suggested that it could arise from molecular crowding [47, 74]. The proposed mechanism being that the space taken up by the enzymes increases with growth rate because more are needed to handle the increasing fluxes. Since cells only have a finite volume, the enzymes of oxidative phosphorylation might be down regulated to make space for more glycolytic enzymes.

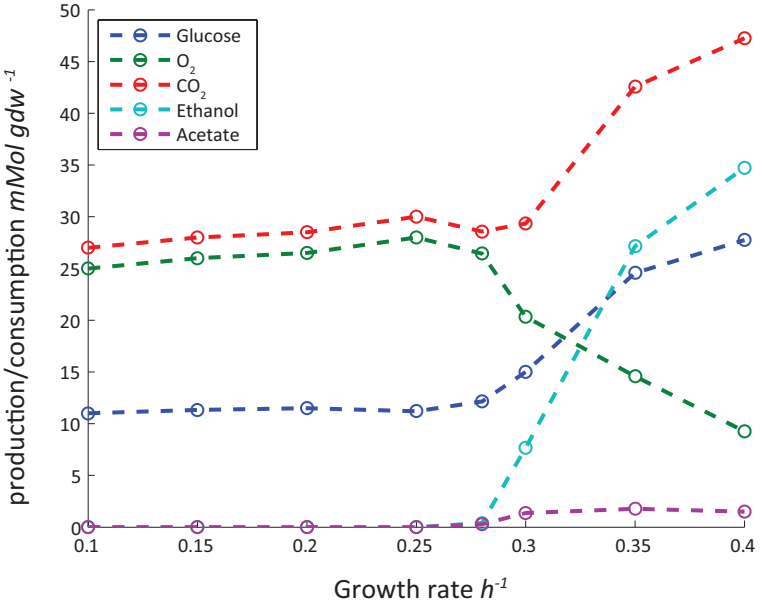

Figure 5: The amount of biomass per consumed substrate remains more or less constant at low growth rates. At a certain growth rate the cell switches metabolic strategy followed by an increase of substrate consumption and a corresponding increase in ethanol production. Data from literature [75].

The Crabtree effect is rather specific for *S. cerevisiae*. Other organisms, e.g. *Kluyveromyces marxianus* can grow at high growth rates, $\mu = 0.49$, without experiencing the Crabtree effect. At these rates the protein content is very high, as much as 71% of the dry weight [26].

1.4 Growth Described by the Monod Equation

All organisms have an upper limit on their growth rate. The growth rate, μ, of an organism is typically controlled by some limiting substrate, $[S]$. Increasing the substrate increases the growth of the organism until the maximal growth rate, μ_{max}, is reached. This behavior is described by the Monod equation (Equation 1).

$$\mu = \mu_{max} \frac{[S]}{k_s + [S]} \tag{1}$$

At low substrate concentrations the growth increases linearly with substrate, and at high concentrations the relation is logarithmic, reflecting a sinking margin benefit of substrate (Figure 6). The description is intuitive and the Monod equation fits experimental data of many microorganisms well.

The maximum growth rate depends on the substrate. For the sugar galactose the growth rate is 0.17-0.23 [14] compared with 0.46 [26] for growth on glucose.

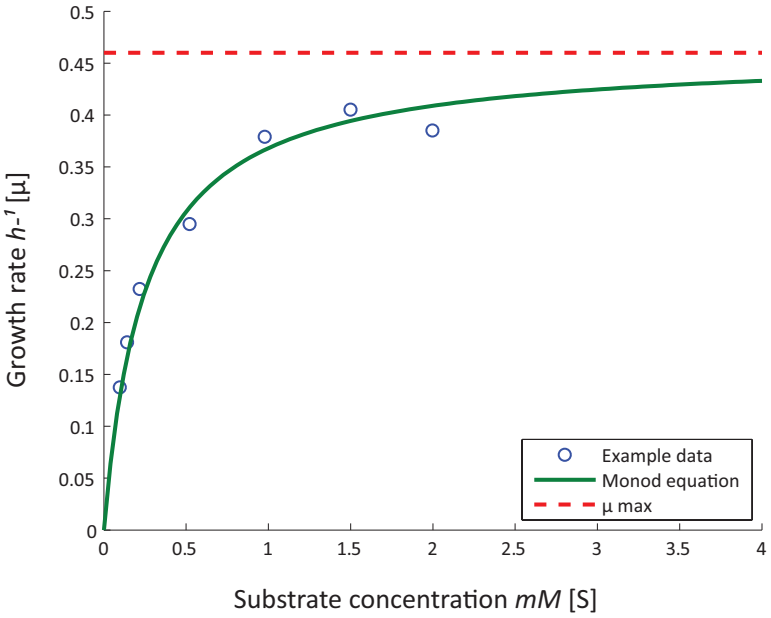

Figure 6: An example of the Monod equation.

1.5 The differences in enzymatic activity amongst proteins

Also enzymes have a maximal rate by which they catalyze biochemical reactions. In the simplest form, enzymatic activity follows Michaelis Menten kinetics. In similarity to the Monod equation, the rate of enzymes increase linearly at low concentrations and saturates at some maximal catalytic rate for high concentrations.

The maximal amount of molecules per second that can be catalyzed by a single enzyme (turnover) is given by the k_{cat} value. The values of k_{cat} are available in databases and literature and span several orders of magnitude. One of the most famous enzymes catalase can catalyze over a million molecules each second [60]. For the enzymes of central carbon metabolism, the k_{cat} values lie between 1 and $10^4\, s^{-1}$ with a median speed of $180\, s^{-1}$ (Section 3.1).

1.5.1 The Effect of Concentration on Enzymatic Activity

The rate by which enzymes operate in cellular conditions will always be lower than their k_{cat} since not all enzymes will be bound to a substrate. The concentration of substrate necessary for half of the enzymes to be occupied is given by km. This value varies for different enzymes, the median km in brenda is 60 μmol [60]. This value is in the same order of the average molarity [2]. A study of absolute metabolite concentrations has also shown that most metabolite concentration are above their km value[11], for central carbon metabolism the values were around their km.

A high enzyme utilization requires large metabolite pools. However there are limitations to the the pool size given by the solvent capacity of water and by the amount of undesired reactions that might take place at high concentrations[6]. It has also been suggested that osmotic pressure might be limiting [69]. For reactions with small thermodynamic driving force the concentration of substrate must be considerably larger than the concentration of product to avoid wasteful backward flux. If several of these reaction occur in sequence the concentration drop at each step can add up to a substantial decrease in substrate concentration [69]. A trade off might exist between enzyme utilization and metabolite concentration[69].

[2] Metabolites correspond to 3% of the dry weight in *E. coli* [53]. The mass to water ratio of the cell is approximately 0.34. This gives a total metabolite concentration of 10 g/liter. Assuming that all metabolites have the molar weight of glucose (180 g/mol) the total molarity is 0,056. Assuming there exists approximately 1000 metabolites the average concentration becomes 55 μmol.

The use of km in a cellular context might in some cases be misleading since crowding from macromolecules affects the effective concentration through the excluded volume [53], macromolecules reduce the space for metabolites to occupy which increases the concentration in the space that they can occupy.

1.6 Literature Review

There exist many models of cells with different scopes and purposes, e.g. constraint based models, kinetic models and whole cell models. Constraint based models aim to predict the cells' states by reducing the set of possible states using constraints [45]. Commonly it is the metabolic network that is the target of the model. There are many versions of constraint based models, a recent review article lists over 100 different variants [45].

This project uses a modified FBA model (Section 1.6.1). There are several relevant modifications to FBA which are assumed to make it more realistic and thereby more accurate (Section 1.6.3, 1.6.4 and 1.6.5).

1.6.1 Flux Balance Analysis

Flux balance analysis is a constraint based model with increasing popularity (Figure 7). The model describes the fluxes of metabolites over the cells' reactions using reaction stoichiometry, reaction directionality and substrate uptake rates as constraints. Since there is an influx of metabolites there also has to be a sink. This is commonly taken to be the biomass equation, a set of metabolites that when drained contribute to growth.

Commonly the biomass equation will involve the cellular components, e.g. amino acids, lipids, carbohydrates and energy in the form of ATP. Solutions are found by assuming that the concentration of each metabolite does not change over time, a steady state. The best description of the system is commonly assumed to be the set of fluxes that optimizes some biologically relevant sub set of fluxes, e.g. maximizes biomass or ATP production or minimizes glucose uptake. In practice this is done by setting up the reactions in a stoichiometric matrix and searching for an optima with linear programming (Section 1.6.8). Several modifications and extensions to standard FBA exist, adding different constraints such as thermodynamic constraints, constraints from gene expression data and kinetic constraints [56].

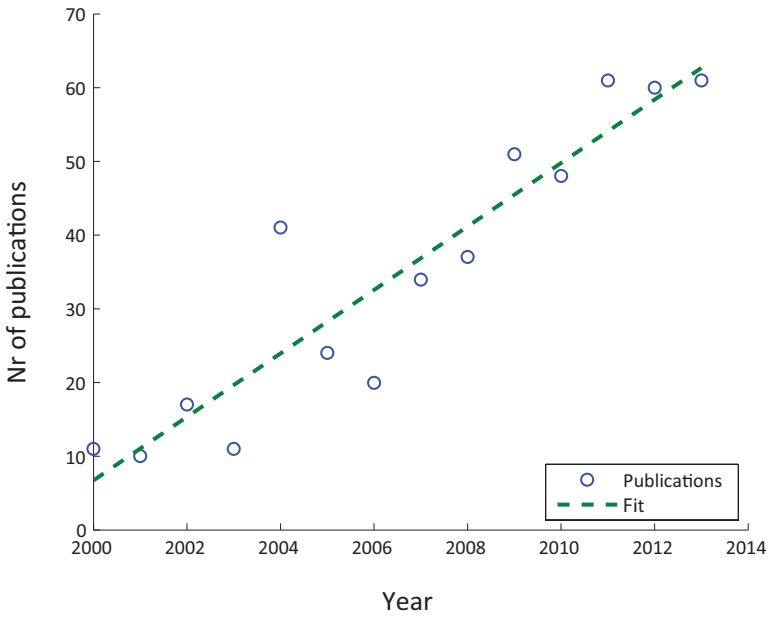

Figure 7: Results for the term "flux balance analysis" in PubMed. The data suggest that there are currently 60 people working directly with FBA, assuming one publication each year per active scientist. The slope of the linear fit is 4.3.

Central to FBA is the assumption that the concentrations of metabolites does not change. This also means that the fluxes are time invariant. This is a strong and incorrect assumption that simplifies the modeling. It might however be considered justified if seen as an average over the cell population or the single cell cycle, since no metabolite gets depleted or reaches infinity for continuous cultures.

1.6.2 Kinetic Models

Whilst FBA describes the time averaged behavior of the cell, kinetic models aim to describe the time dynamic behavior [65]. They normally involve solving a set of ordinary differential equations and kinetics data from measurements of enzyme activity. Because of their demand on experimental data and the complexity of the model, they are often limited in scope, e.g. a kinetic model was used to predict enzymes abundance in glycolysis [66].

1.6.3 Whole Cell Models

Whole cell models [38] aim to describe the whole cell as a system in order to predict metabolite functions, life cycle stages or other behavior and phenotypes. These have become more common as a result of improved data and integrated modeling methods, e.g. commonly FBA is integrated with other modules [38, 31].

They may also be done on a more conceptual level with few components and without experimental data to demonstrate system properties [47]. One such model has conceptually shown that low and high growth rates might give rise to different metabolic strategies. This was done by setting up a system with a pathway that was efficient but slow and one that was fast but wasteful. The same model showed that it is expected to have a larger fraction of the proteins devoted to substrate uptake at low substrate concentration and growth rates, and that a maximum growth rate can be expected based on the rates of the involved units.

The recently developed Metabolism with gene Expression (ME) models extend FBA beyond metabolism to also include the transcription and translation processes [43, 70, 51]. Along sides with the metabolites they therefore also include the concentration of the enzymes responsible for the metabolic reactions. The included enzymes are used to regulate the fluxes and therefore their cost is taken into consideration by the model. The ME models are very complex involving as many as 80 000 reactions.

1.6.4 Metabolic Dilution

Standard FBA does not take metabolite dilution in to account [12]. When the cells grow at a certain rate, the pools of metabolite must grow at the same rate. This leads to a drain of metabolites after each reaction which has the effect that reactions will proceed seemingly slower. An effect that increases with growth rate [3]. Since models are normally fitted to experimental results with a maintenance cost given in ATP [27] neglecting this will not affect growth predictions. It will however result in an over estimation of reactions involved in ATP production and an underestimation other reactions. Taking metabolite dilution in to account requires knowledge of the metabolic pool

[3] The growth rate for yeast will be at the most 0.46 and the metabolite concentration is 3% of the cells' weight, not taking this into account might affect the results by as much as 1.4%

sizes which requires experimental data or a model. The dilution effect will therefore not be considered in this thesis.

1.6.5 Crowding Models

Crowding models acknowledge that all fluxes must be limited since they are mediated by proteins that only can have a limited concentration (Table 1). The models require a prediction of protein concentrations from fluxes, which is commonly done using k_{cat} (Section 1.5). A full scale metabolic model of *E. coli* has been made with a limitation on enzyme concentration [8]. In this model the concentration of each protein was approximated by multiplying the flux with values drawn from a gamma distribution. This was interpreted as each flux contributing to a protein volume which is limited by the finite volume of a cell. They also sanity checked their predictions using k_{cat} values from databases.

The MOMENT model predicts the enzyme concentrations based on database k_{cat} values [1]. The results are interpreted in terms of each proteins weight being limited by the total weight of proteins. They find that the trend for maximum growth rate of *E. coli* grown on different substrates can be predicted. They show that the accuracy of the prediction is dependent on using the real k_{cat} values. Randomly drawn k_{cat} values gives a much worse prediction pearson correlation 0.2 compared with 0.47. This shows that the effect does not arise only from the metabolic network, but also from the weights. Which indicates that using an average turnover for all reactions, which has been suggested [31], might give incorrect results.

A resent article studies the effect of crowding on a limited membrane surface, which applies to membrane bond proteins, e.g. transporters and the proteins of the electron transport chain in the case of *E. coli* [80].

It has since earlier been noticed that a minimization of the total flux can give good predictions [36]. This relates to crowding since minimizing flux also decreases the amount of proteins used. There have also been attempts at minimizing the metabolic network while preforming the same tasks [48].

Standard FBA only predicts the net flux of reactions. If there is a large backward flux over the enzyme, the net flux will underestimate the total amount of enzyme required for the flux [36]. This can possibly be resolved by using ΔG values [69]. A tug of war simulation has been made where optimal enzyme utilization is traded for high metabolite concentrations [69].

Table 1: Table of crowding models.

Name	**Description**	
FBA with Molecular Crowding (FBAwMC)	Calculates the **volume** of enzymes that are required for a certain flux distribution, and limits it. For each reaction a crowding coefficient is calculated describing its contribution to the total enzyme volume.	[8]
FBA with Membrane Economics (FBA^{ME})	Calculates the **area** of membrane that is required for a certain distribution of fluxes over the membrane bond enzymes, and limits it. The relative membrane cost of en enzyme is calculated as the fraction of the membrane space that would be required for 1 mol of reaction.	[80]
Metabolic Modeling with Enzyme kinetics (MOMENT)	Calculates the total **weight** required for a certain flux distribution, and limits it. For each reaction the required enzyme mass is calculated. Since FBAwMC converts mass to volume with a constant this a equivalent but possibly more intuitive formulation.	[1]
ME models, Whole-Cell Computational Model	These models involve some crowding limits since they predict enzyme concentrations using kcat.	[43] [70] [51] [38]

1.6.6 Prediction of Metabolic Strategies

The existence of different metabolic strategies has been experimentally observed and conceptually modeled [47] (Section 1.6.3). A model that describes this has been developed testing the cost and performance of 5 different metabolic strategies [17]. Both the fixed and dynamic cost for each strategy was calculated and the performance was calculated under different conditions. The model showed that different strategies are indeed optimal under different conditions. This has also been shown cross species, where a model was able to predict which glycolysis pathway different species were likely to use [25]. The model compared the energy production of the pathways with the consumption related to the involved proteins. The Crabtree effect in S. cerevisiae could be reproduced for some realizations of randomly sampled kcat values and molecular volumes [74]. The order in which metabolites are taken up by *E. coli* has also been predicted [8].

1.6.7 The Meaning of Optimality in a Cellular Context

In the case of modeling it is common to describe the objective, e.g. growth, of a cell by an objective function. Given an accurate description of the input to the function and assuming a correct understanding of what is beneficial for a cell, it could be expected that the cell has been selected for optimality. By definition optimality for a system occurs when there is no way to make a change that increases benefit. Optimality is dependent on the objective of the system. For complex systems in complex environments the objective will typically also be intricate.

By keeping the surroundings of an organism constant and simple it is possible to influence what is optimal. After only a few hundred generations of evolution organisms conform relatively well with the simple types of objectives used by modelers [44]. But also when considering simple objectives there is some controversy, it has been suggested that there is a trade of between biomass yield and speed of growth [1]. Another complication is that organisms in the same culture can acquire different objectives, such as in the case of free riders [30].

A common definition of optimality is the specific growth rate, the speed of biomass growth. A growth function is constructed where the metabolic precursors for growth related macromolecules are drawn together with the energy demands to assemble them. The composition of macromolecules is normally taken from dry weight measurements of cells. Since the compo-

sition changes demanding on the condition, different functions can be used for different conditions [50].

There is a conflict between optimality and flexibility [61]. It has been discussed in terms of a trade of between changing configuration totally at some cost, or preforming slightly worse. This was studied by comparing the predictions from several different objectives and it was shown that only a objective that weighted these objectives could explain the behavior. It has however also been argued that a more complete model can incorporates these different objectives into a single objective [51].

In a non-constant environment optimality becomes less dependable since events that are unforeseen might change what is optimal. Such events might be a sudden increase in glucose rendering the investment in transporter proteins excessive. Another event might be a decrease in extracellular amino acids, rendering the amino acid syntheses under dimensioned. The cost of uncertainty is hard to predict for the cells since it involves uncertainty. A possible solution is to maintain the ability to quickly adjust the system [42], some degree of sub optimality is also likely to remain because of the coarseness of the regulation and a rough fitness surface. Having an excessive regulatory system might in it self be costly (Figure 8).

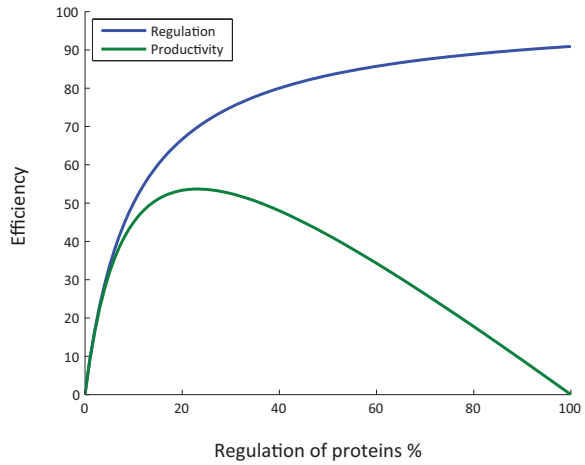

Figure 8: Illustration of the tradeoff between optimal efficiency and space requiring regulation. The efficiency is assumed to increase with diminishing returns, modeled as Michaelis Menten. As regulatory enzymes take up space and resources there is less space for the productive enzymes. Assuming that work is the product of amount and efficiency there must exist an optimal regulation level between 0 and 100% of proteins dedicated to regulation.

Instead of modeling with basis in optimality it is possible to model with constraints taken directly from experimental data [45]. In the absence of an objective the model takes the role of descriptive rather than predictive, it can however be used to calculate the non-measured fluxes.

1.6.8 Linear Programming

Linear program is commonly used to find solutions to FBA problems. It is a way to find the minimum value for a linear combination c of values in a vector x. In the case of FBA the vector represents the fluxes. The values of x may be bounded by an upper $x_{ub}$and lower x_{lb}bound. The vector has to fulfill a set of inequality constraints b that arise from different linear combinations of the elements of the vector. The constraints are commonly written on matrix from S (Equation 2). By combining two inequality constraints it is possible to get an equality constraint.

$$\begin{aligned} S \times x &\leqslant b, \\ min(x \times c) \end{aligned} \tag{2}$$

There exist several algorithms for finding the minimum, e.g. simplex algorithm, interior point algorithm. Because of the linearity there can only exist one optimal value. If the problem is consistent and bounded, the solvers are guarantied to find the global minimum.

1.6.9 Criticism and Problems with the Crowding Models

A problem that has been noticed by the ME modelers is that the crowding models will under predicts protein concentration at low growth rates [51]. This is solved in the ME model by switching to a protein maximization strategy at low growth rates. Although this appears to give a good prediction of protein concentration it is rather ad hock.

A set of criticisms [23, 42, 20, 32, 72] is related to there not being a strong correlation between changes in gene expression in terms of mRNA and a corresponding change in flux. A lot of the criticism applies mainly to the steps between gene expression and protein expression where interfering phenomena such as gene silencing, mRNA and protein degradation etc might occur. Some of the criticism however also applies to crowding models and

the steps between protein concentration and fluxes. Interfering phenomena might be product inhibition, substrate limitation, alosteric regulation, pH dependence, phosphorylation, etc. the considerations that are taken into account in kinetic models (Section 1.6.2). Although a certain protein concentration is a necessary condition for a certain flux, there might exist a range of concentrations above the minimum [42].

The for mentioned criticism is generally directed towards predicting fluxes from enzyme concentration, rather than the opposite. The criticism is often based on observations from over expression of a set of proteins, with no noticeable change in the fluxes. Since proteins are involved in pathways, other proteins might block increased flux through insufficient substrate concentration from preceding steps, or a product build up at subsequent steps. Over expression will also increase crowding which might block increased flux. One article [42] notes that in the cases were fluxes change it is mainly as an result of changes in growth rate. This is however expected since a decrease in an important flux will ultimately decrease growth.

A common criticism towards optimality is that it is commonly possible to improve fitness by gene deletion or other kinds of engineering [42]. That gene deletion can improve growth could be seen as an argument for crowding being a limit to growth.

1.7 Purpose, Scope, Limitations and Goal

The purpose of this master thesis is to investigate what will be the consequence of limiting metabolic fluxes by the amount of proteins that are required to carry them. The main focus will be on the effects on metabolic strategies, how the metabolic behavior of *S. cerevisiae* changes with different growth rates and for different substrates.

This has already been done with good results in *E. coli* with real data and conceptually with random data for *S. cerevisiae* (Section 1.6). There has however, to the best of my knowing, not been made such a model that uses real kinetic data for *S. cerevisiae.*

The model is mainly intended to have a theoretical value in determining if the introduction of these limitations will change the behavior of the system. But some qualitative and order-of-magnitude quantitative results are also expected. Some of the biological questions that the thesis will try to address:

1. Can shifts in metabolic strategy such as the Crabtree effect be explained by protein related flux limitation?

2. Can the lower specific growth rate observed for growth on galactose be explained in a similar way?
3. The concentration of all enzymes involved in the model will be predicted as a side effect of the main method (Section 2). Do these predictions agree with literature values?
4. Since higher growth rates require more flux, the protein concentration is expected to increase. Can this be related to changes in the mean cell size?
5. Large cells gives a lower uptake per cell volume. Can this impose an upper bound on the cell size?
6. Central carbon metabolisms share of the total protein mass is approximately 25%. Does competition over space from the ribosomes explain why this share is not larger?
7. Since low growth rates require very little flux the proteins are expected to be underutilized. Can this be related to lower metabolite concentrations?

1.7.1 Limitations

To keep the project relatively simple it will focus on one organism, *Saccharomyces cerevisiae*. There will be no time dynamics, the steady state assumption from FBA will be used. The study will be limited to the central carbon metabolism. The other cellular components are assumed to be independent of CCM and will only be modeled through their drain on the metabolites in CCM and their limiting effect on the protein pool. This includes transport reaction that are intimately linked to CCM.

Although cells have an extensive regulatory network that controls the cell, it will not be modeled. It is thereby implicitly assumed that the cell can reach all possible metabolic states, i.e. enzyme concentrations, through regulation. And that the regulation is made in such way as to reach optimality. Optimality will be seen in terms of maximizing growth rate or minimizing substrate utilization for a given growth rate, which is equivalent [27]. It should be emphasis that although regulation is not explicitly modeled, all the metabolic states reached by the system must do so through regulation. The model can in that sense be viewed as a prediction of which regulations the cell will impose.

The gathering of stoichiometry data from databases and literature has already been performed for S. cerevisiae [3, 50, 27], only minor changes will be made to these reconstructions. To allow different metabolic strategies to

emerge the network however needs to be of a sufficient size. And the drain of metabolites from other cellular functions to be reasonably accurate.

No physical experiments will be performeds, the project relies fully on the experimental work already available in literature.

The result of this project will hopefully be a demonstration of how a simple extension to FBA can allow modelers as well as experimentalists to get more accurate predictions when planing pathway additions and substitutions, as well as in understanding pathology in biological systems.

1.7.2 Ethical and Environmental benefits

The choice to only use *in silico* models, i.e. computer models of cells, allows a lot of experiments to be done at high pace and at a low economical and environmental cost. The use of *in silico* models allows genetic experiments to be carried out with out specific permission and without risk for experimentalists and research animals. Improvements of *in silico* models might in the long term decrease the research cost in GMO development and thereby increase its usage. This might have positive effects on the productivity of the world, but could also result in increased genetic homogeneity which might make the ecological system more vulnerable.

2 Methods

FBA was the main method used in this project. It is well founded and in broad practical use [50]. Some modifications were made to the FBA formulation to incorporate the enzyme concentration prediction and limitation. Several systems with this purpose are already in use [8, 1, 51], the framework was however built extending on the Reconstruction, Analysis and Visualization of Metabolic Networks (RAVEN) framework for FBA [3]. The motivation being that the changes were relatively simple to implement, and that no potential additional features were introduced that might affect the result (Section 2.1).

The results depend on a big set of experimental determined k_{cat} values from literature. To investigate how measurement errors might affect the result, simulations with perturbations were made (Section 2.3). To further aid the analysis of the data, a visualization tool was constructed (Section 2.7).

The proteins of central carbon metabolism are only a part of the cell, $\approx$ 25% of the proteins at high growth rates. The interactions with the other functions of the cell were taken in to account by analyzing proteomics data (Section 2.5) and by extracting the biomass equation from a larger model (Section 2.1.2).

FBA gives an answer to what is optimal given a certain biomass composition and substrate uptake. For changing growth rates both of these factors are subject to change. A wrapping framework that allowed a systematic investigation of the predictions of FBA under different conditions was developed. This framework did not rely on any specific theoretical foundation as it is simply a convenient and systematic way to test different parameters. The framework allowed the limitations on protein concentrations to vary with the condition, and thereby to relate to experimentally measured protein concentrations.

It has been theorized [54, 80] that substrate uptake rates might be a limiting factor at high growth rates. A simple model testing this alternative hypothesis was investigated (Section 2.6.3).

It is desirable to model cells in a computer both from an applied and theoretical perspective. However care must be taken to avoid common modeling mistakes (Section 2.8).

2.1 A Kinetically Constrained Version of FBA

A FBA model requires a stoichiometric description of a metabolic network (Section 2.1.1), a biomass equation (Section 2.1.2) and a linear programming solver (Section 2.1.3). The main modification in the kinetically constrained version of the FBA model, is the prediction of the amount of proteins and thereby their mass (Section 2.1.4). This requires data on protein kinetics and weights as well, as data on the protein composition of enzymatic complexes (Section 2.1.5).

2.1.1 Reconstruction of the Stoichiometric Network

The stoichiometric model that was used in this project was based on the model "smallYeast, (Central carbon metabolism for yeast)" from the RAVEN toolbox [3]. The smallYeast model is a scaled down version of the iFF708 [27] model, developed for use in the RAVEN workshop. The model includes glycolysis, the pentose phosphate pathway, the citric acid cycle and a simplified version of oxidative phosphorylation. In total it amounts to 54 reactions. For the purpose of this project some changes were made to the model; the oxidative phosphorylation was extended (Appendix A.1) and the galactose pathway was added (Appendix A.2) amounting to 79 reactions.

The reconstructed model produces 16.5 ATP per glucose molecule under optimal conditions which is in agreement with the iFF708 model and slightly lower than the 18.4 ATP predicted by the "smallYeast" model.

2.1.2 Biomass Equation

The biomass equation represents the drain of metabolites required for growth. It is commonly calculated from experimentally determined dry weight composition. This composition however varies with both growth rate and nutrient source requiring different biomass compositions to be used for different conditions [50].

The biomass equation also includes an experimentally fitted maintenance parameter representing ATP consumption for protein turnover, PH stabilization, etc.[27, 22]. Since this parameter is not explicitly measured, it needs to be calculated from the biomass yield at each growth rate. Calculating the maintenance for each growth rate was considered to be beyond the scope of this thesis and a fixed value of 35.36 ATP was taken from literature [27].

Some simple calculations of the maintenance cost indicates that this might be an under estimate for some conditions (Appendix A.4).

A biomass equation was generated based on the iFF708 model. This was done by running simulations for biomass production in the iFF708 model, where the model was allowed to use all metabolites in smallYeast as substrates. This simulation calculated the drain of on the metabolites of CCM imposed by the biomass equation. Care was taken to prevent "unnatural" fluxes to arise as a result of the broad availability of metabolites (Appendix A.3).

Although the extracted biomass equation is in good agreement with iFF708 in terms of predicted growth rate and fluxes, it relies on the assumption that these drains cannot change in response to shifts in metabolic strategy, etc. This is likely to be incorrect. Correcting for this would however require the modeling of pathways outside of CCM which is beyond the scope of the thesis. The alternative of constraining only the fluxes of CCM within a larger model was avoided since it could allow the optimization algorithm to bypass the constraints through alternative pathways, giving incorrect predictions.

To simplify experiments of varying biomass composition, a modular version of the biomass equation was generated where each biological component was given its own reaction, e.g. protein, RNA. This was done by dividing the IFF708 model's biomass equation into components and running simulations, maximizing each in turn. The total biomass equation was recreated as the superposition of these components. This is should not be completely accurate since there exist synergy effects in formation of the different components, e.g. RNA formation creates amino acids as a byproduct, that are utilized by the protein formation. The pathways for forming the different components are however relatively independent and the resulting fluxes and predicted growth rates were almost identical.

2.1.3 Linear Programming Solver and Modeling Framework

The RAVEN [3] tool was used as a basis for the FBA solving framework. It is written for MATLAB and uses MOSEK (MOSEK ApS, Copenhagen, Denmark) as the linear programming solver. In RAVEN the stoichiometric matrix is stored as a list of equations in an excel sheet. Modifications were made to the source code to allow k_{cat} and weight parameters to be read by the software.

2.1.4 Predicting and Constraining the Total Amount of Protein

The amount of proteins N required for each flux F normalized by kg dry weight was calculated (Equation 3) using k_{cat} and a factor c, $0 < c < 1$ relating the speed of the enzyme to the theoretical maximum. In this project the value 0.5 was used for all reactions, it will be refereed to as the c parameter (Section 2.4).

$$N = \frac{F}{c \times k_{cat}} \tag{3}$$

The amount of amino acids, w, required for a given flux, was calculated from the size of the protein, A, and N (Equation 4).

$$w = \frac{A}{c \times k_{cat}} \times F \tag{4}$$

Since backward fluxes are commonly represented by negative values in FBA, they would result in a negative protein abundance. The lower bound for all fluxes was therefore set to zero and all reversible reactions were copied and one of them reversed by multiplication with -1.

To prevent that both forward and backward reactions take place at the same time each flux was associated with an small $\epsilon = 10^{-6}$ drain on ATP, making lower fluxes more favorable, but with a negligible effect on the system as a whole.

Although central carbon metabolism was the only modeled cellular function, it does not involve all of the cells' proteins. The total weight of protein in CCM, $protein_{CCM}$ has to be some fraction of the total protein weight $protein_{total}$ (Equation 5). This fraction was taken as 0.20 and referred to as the $CCMp$ parameter (Section 2.5).

$$0 \leq CCMp = \frac{protein_{CCM}}{protein_{total}} \leq 1 \tag{5}$$

2.1.5 Isozymes, Multifunctional Enzymes and Complexes

Standard FBA deals with fluxes over reactions rather than over proteins. Isozymes are different proteins that catalyze the same reaction. For simplic-

ity a single enzyme representing all isozymes was chosen. When comparing the predicted abundance with literature values all isozymes were included.

In some cases the same protein can catalyze several reactions. If the reactions share the same catalytic site, the abundance can be added, this was the case for the proteins TKL1 and TKL2 in the model. If the reactions take place on different sites, adding the abundance might lead to an overestimation of protein. This was the case for the protein GAL10 in the model. This was solved by not predicting protein abundance for the fastest of the reactions.

The enzymes that form the biochemical reactions are in most cases in the form of complexes. In most cases these complexes are of a simple dimeric type. It is not in general necessary to take this into account since these complexes tend to have a number of active sites that scale with the oligomer number i.e. dimers have two active sites, tetramers have 4 active sites[53]. This means that normalization on the number of active sites yields monomeric structures.

Some of the complexes have a large number of subunits per active site. For these the stoichiometry was taken as input and the weight per active site was calculated (Appendix D).

2.1.6 Matrix formulation

The constraint matrix was implemented as in standard FBA with an additional row of constraints relating fluxes to amounts of enzymes (Figure 9). For each reaction with a k_{cat} value an element was added to the extra row relating the flux to the required amount of amino acids (equation 4). If the reaction did not have a k_{cat} value it was set to 0. The row was constrained by setting the upper bound to $CMCp \times total\,amount\,of\,amino\,acid$, the lower bond was set to 0.

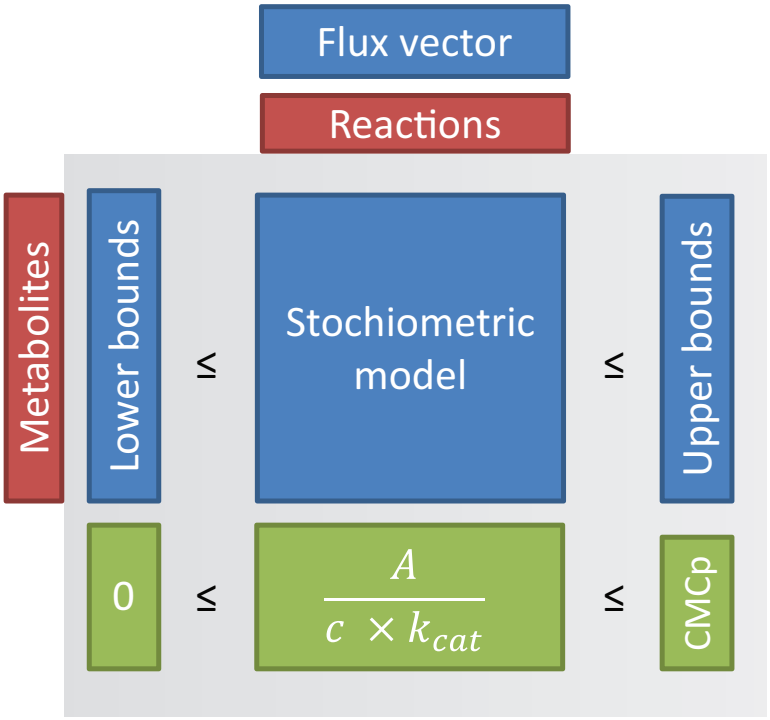

Figure 9: Conceptual layout of the linear problem. A flux vector fulfilling the constraints given by the matrix is sought. The vector gives the fluxes over the reactions. The enzyme concentrations are proportional to the fluxes with a factor $(c \times k_{cat})^{-1}$. The amount of amino acids for each protein is given by the size A and the concentration. The sum of these values is bounded by the protein constraint that is calculated from the $CMCp$ parameter. The green boxes indicate extensions from standard FBA.

2.2 Growth Dependent Experimental Data as Input

The total amount of amino acid for CCM was in general estimated from the CCMp parameter and protein concentration at high growth rates. To better predict the observed data at lower growth rates the amino acid amount was recalculated from experimentally measured protein concentrations at the given growth rate (Section 2.5). In addition the growth equation was updated to reflect the experimentally measured values. For the intermediate growth rates the values from the two closest measurements was interpolated.

2.3 Sensitivity Analysis

The model relies on experimental data and the improbable assumption that the c parameter is the same for all reactions. It is therefore important to investigate the stability of the results. Assuming that the k_{cat} and c parameter are correct on average it is of interest to investigate how perturbations around these values affects the system. Since c always is multiplied with k_{cat} it suffices to test the effect of different k_{cat}. The stability of the predicted maximum growth rate and the composition of the excreted metabolites was investigated as follows:

1. Each value of k_{cat} was perturbed by multiplying it with a random value, r. The values were drawn from a uniform distribution placed symmetrically around 1 with the size a, $[1-a, 1+a]$, eg [0.5 1.5].
2. Simulations were run for growth maximization and the maximum growth rate was recorded together with the ethanol, acetate and glycerol production.
3. Steps 1 and 2 were repeated 1000 times.
4. Steps 1-3 were repeated different values of a, 0.1, 0.2 and 0.5.
5. Histograms of the growth rate were plotted and the amount of experiments where each product was produced was counted.

The sensitivity of the individual parameters was tested by changing the parameter whilst keeping the others unchanged. The values were perpetuated by multiplying the k_{cat} value with a factor between 0.1 and 10.

2.4 Estimation of the c-Parameter from Experimental Data

It is common to assume that metabolite concentrations are near saturation levels [8, 51] and that enzymes therefore will operate at near maximum speed. However experiments [11] show that metabolite concentrations are close to km for CCM [4]. At these concentrations the true reaction speed, $kcat_{true}$, will be proportional to the maximum speed, $kcat_{max}$ (Equation 6) with a coefficient $c = 0.5$. It is very likely that the c value is different for each enzyme. Estimating the c value for each enzyme is however beyond the scope of this thesis and an average c of 0.5 will be used.

[4] In these experiments the total concentration of the metabolome was measured to 300 mM which appears to be high. Extracellular glucose concentrations is around 0,104 -1,133 mM [13]. Which can be compared with the first step of glycolysis that has a km of 0.04 mM which is well below external levels [60].

$$c = \frac{k_{cat\,true}}{k_{cat\,max}} \tag{6}$$

Several factors count into the c parameter. Cells might not be fully active through the entire cell cycle. The optimal pH and temperature might not be present. There might exist molecules interfering with the process. The enzyme might have almost the same concentration as the substrate [**?**]. A study [73] tests the effect of cell like conditions on the maximum performance of enzymes. The average decrease in performance for the 12 measured enzymes was on average 20% but varied a lot (Appendix B.1).

The value of c can be estimated from experimental data [73]. This can be done by measuring the maximum enzymatic activity in cell free extracts, i.e. a solute extracted from cells that has the same internal ratio of proteins as in the cell (Appendix B.1). The maximum enzymatic activity in this extract corresponds to the flux that would be carried if $c = 1$. By comparing measured fluxes to the maximum a value of c can be estimated (Equation 7).

$$\frac{F_{messured}}{F_{max}} = \frac{k_{cat\,true} \times N}{k_{cat\,max} \times N} = \frac{k_{cat\,true}}{k_{cat\,max}} = c \tag{7}$$

2.5 Estimating the CCMp-Parameter from Proteomics Data

The cellular composition was determined (Appendix E) using protein abundance data [79], this data was a weighted mean of several different experiments. Each abundance was multiplied by the size of the protein giving a distribution of the the amino acids. The proteins belonging to the CCM model were compared with the total abundance giving the CCMp parameter.

2.6 Modeling of Other Constraints

The amount of proteins in CCM is limited in the model by the experimentally determined CCMp parameter (Section 2.5). This limit could however be caused by some other bottleneck that makes increased protein concentrations unnecessary. To determine if CCM could have a larger share of the cellular protein several biological phenomena were investigated (Figure

10). The studied phenomena were ribosomes increase with growth (Section 2.6.1), increasing cell sizes effect on uptake (Section 2.6.3), and increasing cell sizes effect on cell wall and free protein (Section 2.6.4).

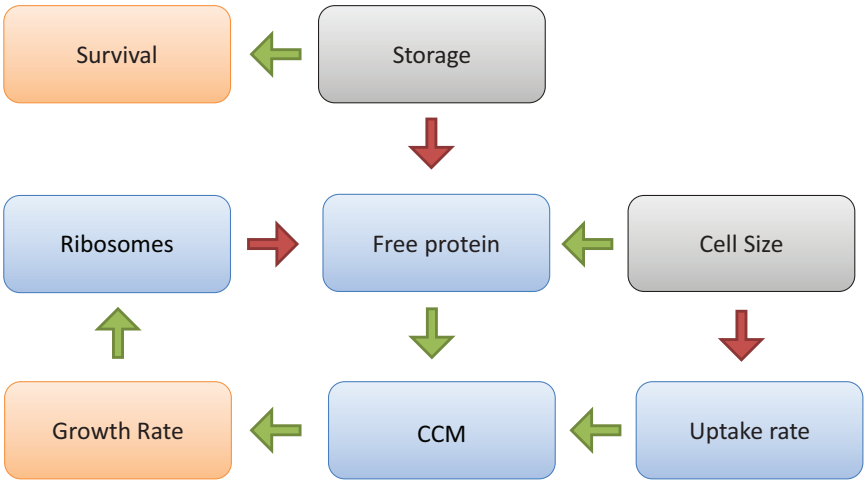

Figure 10: Conceptual layout of the models interaction with other cell related phenomena. The objective of the cell is survival and growth. The higher the flux through CCM the higher growth is possible. A high growth rate requires high amounts of ribosomes. This however decreases the free protein available for CCM and other purposes. By increasing the cell size the free protein levels can increase, but this affects uptake rate negativity putting a constraint on CCM. If growth rates are low and the free protein is not limiting CCM the cell can store carbohydrates improving its fitness in case of starvation.

2.6.1 Growth and Ribosomes

The amount of ribosomes increases with growth rate [54]. The structure of yeasts ribosome is known [10] with regards to both RNA and protein content. The amount of protein directly in ribosomes can be estimated from the RNA amount. The abundance of amino acids dedicated to protein Synthesis was calculated form RNA abundance.

1. The ratio of ribosomal RNA (rRNA) out of the total RNA (rRNA/RNA) was estimated to 0.815 based on *E. coli* data [5], this is in line with an old study on yeast [78].
2. The amount of rRNA was calculated from this ratio and experimental RNA levels.

[5] table 2.1 in Physical Biology of the cell [53]

3. The amount of ribosomes was calculated based on rRNA and the number of nucleotides per ribosome (Table 2).
4. The total amount of amino acid was calculated based on the number of ribosomes above and the amount of amino acids for each ribosome (Table 2).
5. The total amount of amino acid in protein synthesis was calculated based on the ratio of ribosomal amino acids and total protein synthesis amino acids, 2.54 (Figure 11).

Table 2: The composition of nucleotides and amino acids in the ribosomes of *E. coli* and *S. cerevisiae*.

	Nucleotides	Amino acids	Reference
E. coli	4567	7459	[54] (Bio numbers ID 101175 -76)
S. cerevisiae	5500	13000	[10]

Additionally an attempt to calculate the amount of ribosomes directly from the growth rate from the translation rate. It was however found that the literature values for ribosomal translation speed were more than 30% lower than what is actually required to maintain growth. Correcting for this, gives an under prediction of the ribosomal amount at low growth rates, possibly indicating a lower translation speed or that the ribosomes are occupied with protein degradation (Appendix F).

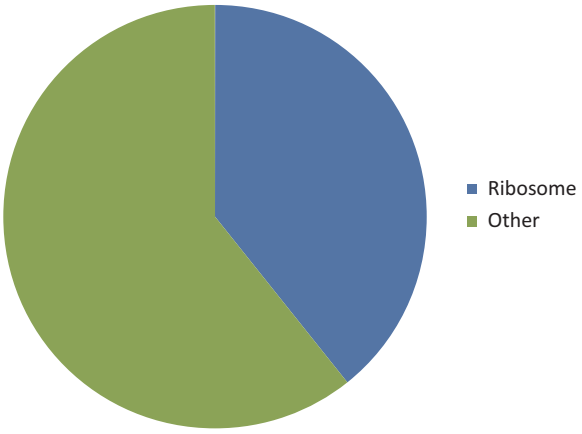

Figure 11: The amino acid abundance of ribosomal protein and other protein for protein synthesis. The data shows that amino acid abundance for protein synthesis is in total 2.54 times higher than the ribosomal abundance (Appendix E).

2.6.2 The Relation Between Size and Growth

Two important quantities are expected to follow from cell size. Maximum glucose uptake per kg dry weight. Which is limited by the surface volume ratio. And the mass percentage in the cell wall.

The mean volume of yeast cells is thought to increase with growth rate [40]. The volume as a function of growth rate from a glucose and nitrogen limited chemostat experiment was investigated [13] (Appendix F.2).

For nitrogen limited cells there exist a good linear fit for the volume, v, with respect to the growth rate μ (Figure 12), which gives an empirical formula (Equation 8). The relationship holds in the range μ=0.1 to $\mu = 0.3$ where the volume increases from 22.9 to 37.3 μm^3. The volume does not appear to fall below this value at lower growth rates.

$$v = 76.1\mu + 15.8 \tag{8}$$

For glucose limited cells the volume shows no simple dependency on growth rate, the mean size is 26,4 μm^3, and the volume is in general lower than for nitrogen limitation.

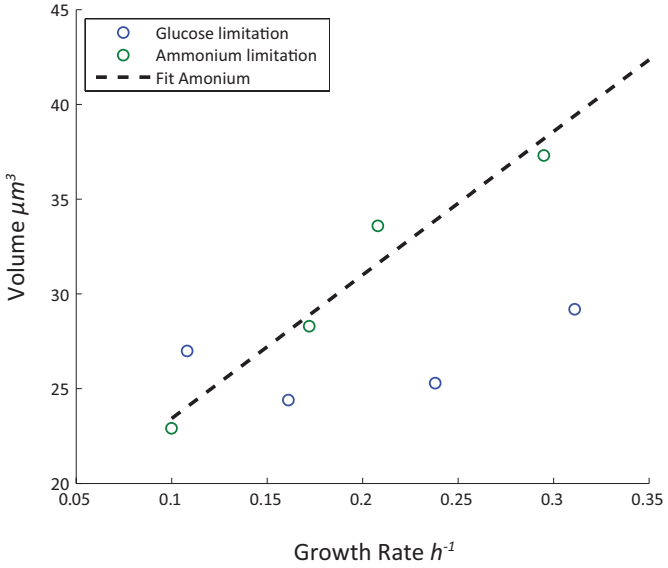

Figure 12: The volume as a function of growth rate from glucose and ammonium limited experiments [13]. The linear fit is for ammonium limitation.

2.6.3 The Effects of Cell Size on Uptake

The maximum possible glucose uptake per kg dry weight decreases with cell size as a consequence of decreasing surface/volume ratio. According to proteomics data [79] there are approximately $2.2 \times 10^{-5}\, mol$ glucose transporter per kg dry weight at high growth rates. The highest literature value for transporter activity is $237 s^{-1}$ ([54] bio numbers id 101738) which gives a maximum uptake of $18.77\, mol\, h^{-1}$. The literature activity value is however estimated from measured glucose uptake rates in living cells, so higher theoretical maximum values might be expected.

The amount of transporters per kg dry weight is expected to scale with the surface volume ratio [2]. From this the predicted maximum uptake rate can be estimated from the volume of the cell. This is based on the reference size of approximately 56 μm^3 [2] and the reference surface volume ratio 1.27. The maximum uptake is calculated assuming that the uptake scales with the surface volume ratio.

Using the empirical volume and growth function (Equation 8) the growth rate was estimated for each volume. The predicted consumption was calculated for an anaerobic growing strain with a biomass yield of 0.11 [49] and a fermenting strain with a biomass yield of 0.2 [35].

2.6.4 The Effects of Cell Size on the Cell Wall

The mass of cell wall associated carbohydrates is expected to decrease with the cell size since the surface volume ratio decreases. The ratio of carbohydrates in the cell wall was estimated using the relative surface volume ratio estimated from the empirical relation between cell size and growth (Equation 8) and the measurements of the content at $\mu = 0.4$ as a starting point (Appendix F.3).

2.7 Data Visualization

A simple tool for visualizing the metabolic network with color coded data was developed using MATLABs biograph toolkit. The graph was created from the stochiometric matrix as follows:

1. Reactions without interest for the graph were removed, e.g. the biomass equation.
2. Ubiquitous metabolites, e.g. ATP and NADH, were removed.

3. An adjacency matrix was created for the reactions by setting all reactions sharing a metabolite as adjacent.
4. The data to present was logarithmized, shifted to 0 and normalized to max 1.
5. A graph was drawn with the biograph toolkit and nodes were color coded by the data.

The data of main interest was k_{cat}, weight and the k_{cat}/weight ratio, i.e. the specific activity.

2.8 Notes on Modeling

Models have great possibilities to aid researchers in their decision making and understanding [68]. For this to work optimally it requires relatively simple, well documented, interesting models with defined scopes. There is often a trade-of between simplicity and accuracy, as well as between generality and accuracy. Models that are too complicated might never be of practical use e.g. a model that requires a lot of parameters to be known and configured. Although models with kinetic constraints by nature will require more parameters than without. It has been an explicit goal to keep the number of parameters low and the number of free parameters very limited, sacrificing accuracy for simplicity and generality. Ir-reproducibility might also arise from models relying too heavily on random sampling with insufficient statistical analysis. In this thesis random sampling will only be used to test the stability of the results.

From a theoretical perspective the act of building models might be of interest even if the models are not expected to generate predictions with practical applications. The model has potential to become a scaffold onto which the knowledge of a field is combined. The model building might even encourage future knowledge generation by accentuating where the blank spaces of knowledge reside, e.g. the periodic table. The storage of information in the form of a model is a compact and persistent format of knowledge allowing work to be built on earlier work, e.g. using a stochiometric model without knowing all the details in its construction.

The act of modeling will commonly bring clarity as the modeler is forced to articulate the idea in a formal way exposing strengths and weaknesses. Modeling typically involves integration of data from many sources, contradictions in the underlying data can thereby be exposed. Resolving these problems with the aid of models can improve the quality of the data in a

circle of knowledge [27]. An important consequence of modeling is the set of unresolved questions that might result in biologically relevant hypotheses.

Using models however has its limitations [68]. The most obvious practical disadvantage is that they can be misleading. A faulty model might misguide a researcher, this might especially be the case if it gives good predictions within a limited range of conditions. Because of the time persistent nature of models, these mistakes can live on for a considerable amount of time. A similar problem might be over-parametrized models that can be fitted to predict any outcome.

3 Materials

The thesis combines data from a large set of sources. The leading resources for each data type is used when possible. Turnover values, k_{cat}, were mainly collected from the Brenda database and peer reviewed literature (Section 3.1). Data on which pathway proteins belong to was collected from the KEGG database (Section 3.2). Protein weights and the structure of complexes were taken from the UniProt database (Section 3.3). Protein abundance data was taken from the PaxDb database (Section 3.5). Cellular composition was taken from two studies (Section 2.5). And the dependency of cell size on growth rate was taken from a high throughput study (Section 3.6).

3.1 Kcat Values

Brenda [60] is the main database for k_{cat} values (Figure 13). Data availability, consistency and quality was not always the best. Data was not always available for the organism of choice. In those cases data from the genetically closest organism was used, the organism was identified in each case using blast [38]. The temperature was not the same in all experiments which is known to affect the values by as much as a factor 2^6. Data was, when possible, taken at 30^oC.

In some cases results were found in other sources, e.g. an article collecting reported k_{cat} values [4] or directly from individual experiments [66, 28].

In other cases the k_{cat} was calculated from specific activity. This value is normally given in μmol converted substance, a, per mg purified protein and minute. From this k_{cat} can be calculated (Equation 9) using the weight of protein, w, in mg per mol.

$$k_{cat} = \frac{a \times w}{60} \tag{9}$$

In 4 cases (8%) data was estimated with other methods. In one case the value of a functionally similar protein was used. In one case the k_{cat} value was set to the highest value from another specie without homology. And in

[6] http://antoine.frostburg.edu/chem/senese/101/kinetics/faq/temperature-and-reaction-rate.shtml

two cases the median of all data was used (Appendix B.2). In general data availability was lowest for the TCA cycle (Appendix B).

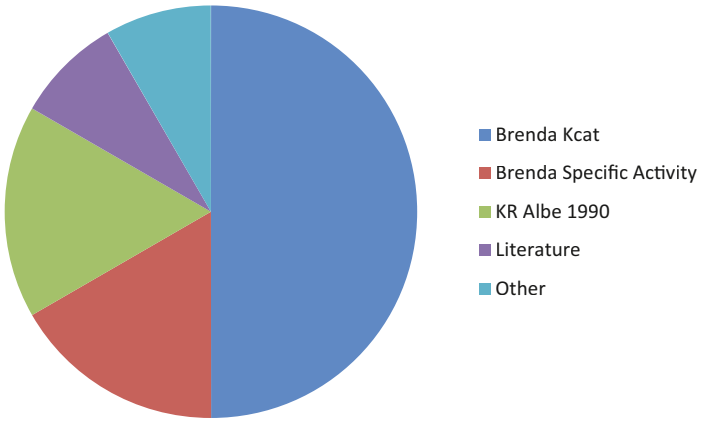

Figure 13: The 48 k_{cat} values were drawn from several sources, mainly Brenda [60] 50% + 17% and an article collecting k_{cat} values from different sources [4] 17%.

The databases often contained several suggested k_{cat} values. Many of these values were from mutation experiments where the k_{cat} was decreased. It is likely that some of the reported experiments might be missing important co-factors which will be available in the cell [73]. In other cases the value of k_{cat} was calculated from the weight of dimers, since the experiments were preformed before the actual protein weight was known (Appendix B.2.7). In these cases the k_{cat} value was recalculated using the specific activity.

The highest reported value for yeast in Brenda was normally used, as suggested by literature [38]. If the values were very different from other reported values the source was checked. The use of the maximum value makes the data sensitive to experimental measurement errors. It has therefore also been suggested to use the median value [1]. This however would make the results dependent on several experiments and the amount of sources to verify would increase dramatically. Using the median makes the results sensitive to double reporting, since there in many cases are experiments reporting several similar values.

Since the system uses inverse k_{cat} values, a few proteins with low k_{cat} values dominate the protein approximation (Figure 14). This is expected since the logarithm of protein abundance is approximately normally distributed (Section 15). Specific care was therefore taken when handling enzymes with

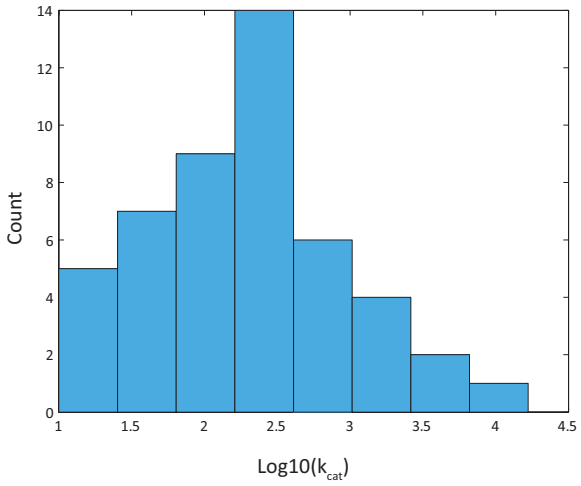

Figure 14: The values of k_{cat} in CCM are logarithmically distributed, the log distribution is however not necessary normal $pval \approx 0.43$ for a χ^2 test. There are relatively few low kcat values 16% are below 50. Most, 66%, are in the range 50-1000.

low k_{cat} and some k_{cat} values based on incorrect protein weights were found in the database (Appendix B.2).

3.2 Pathway Annotation Data

The Kyoto Encyclopedia of Genes and Genomes (KEGG) [52] database contains pathway annotation for many genes. This allows a simple mapping of protein abundance data to cellular function. A script for retrieving KEGG data and mapping the genes was developed.

3.3 Protein Weights and Complexes Data

The Universal Protein Resource (UniProt) [71] is a database of protein data. The weights and the complex stoichiometry was in general gathered from there (Section D). A simple script for retrieving all the weight data from a UniProt xml was developed.

For the purpose of future extension, the weight of the protein was taken in terms of number of amino acid residues, AA, rather than $dalton$. A conversion factor of 110 corresponding to the average weight of an amino acid (Appendix A.5) can in general be used $AA \times 110 = dalton$.

UniProt was also the main resource for the stoichiometry of protein complexes (Appendix D).

3.4 Chemostat Experiment Data

Data from chemostat experiments was retrieved from literature [35, 55] (Appendix C.1). The experiment measures the in fluxes of O_2, glucose and the out fluxes of CO_2, ethanol, acetate and glycogen. One of the studies [55] focuses on the execration of acetate and showed that this occurs at lower growth rates than ethanol execration.

The cellular composition of cells was taken from two studies [63, 49] (Appendix C). These experiments were preformed for nitrogen limited aerobic cells [63] and carbon limited anaerobic cells [49], which are different conditions from the carbon limited areobic cells that are investigated in this project. The data was therefore used with some caution.

3.5 Protein Abundance Data

Data on protein abundance was taken from PaxDb [79]. The values where composed of a weighted average of several experiments. The data was in general sampled in the exponential phase of batch cultures grown on rich medium. Representing the cellular composition at high growth rates. The logarithm of the data was normally distributed (Figure 15).

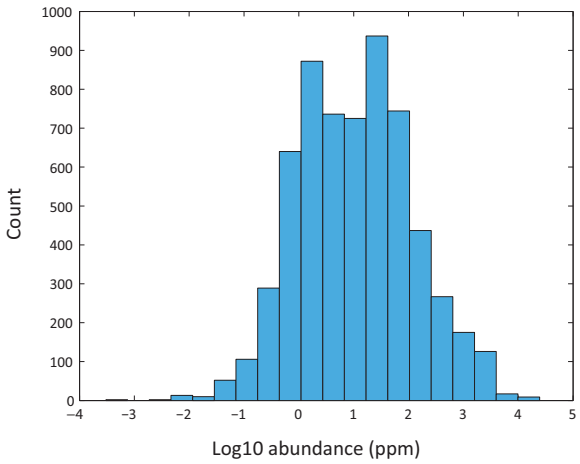

Figure 15: The abundance of protein has a log normal distribution, $pval \approx 10^{-44}$ for a χ^2 test. The more than 500 proteins with zero expression were excluded.

3.6 Data on Cell Size

Volume as a function of growth rate from glucose and nitrogen limited chemostat experiments was taken from literature [13] (Appendix F.2).

4 Results

The model is able to identify experimentally observed shifts in metabolic strategy and predict a maximum growth rate consistent with experimental values for two different substrates (Section 4.2). The results are insensitive to perturbations (Section 4.2.7).

Other constrains such as maximum uptake rate and the space taken by ribosomes are shown to become important at high growth rates (Section 4.3).

The chosen modeling parameters are in reasonable agreement with experimental observation (Section 4.1).

4.1 Estimation of Model Parameters

Central carbon metabolism composed 25% of the proteome by weight. The proteins used in the model corresponded to 20%. The protein share of the dry weight in the exponential phase of batch conditions was taken as 57% based on literature [62], which fits well with extrapolation from chemostat data (Appendix C.2). This gave a constraint of 1.04 mol amino acids in central carbon metabolism per kg dry weight.

The value of c was calculated for 12 different enzymes using experimental data. The value of c was 0.43 ± 0.08 for batch conditions. For low growth rates, 0.1, in chemostat conditions the value was 0.08 ± 0.01. approximately 20% of the value in batch conditions. The estimate shows that chosen c value is consistent with experimental data at high growth rates.

4.2 Results from the Model

The model predicts a maximum growth rate in agreement with literature values (Section 4.2.1). It predicts 4 qualitatively different metabolic strategies including the Crabtree effect, at different growth rates (Section 4.2.2). The model predicts a lower maximum growth rate on galactose in agreement with literature values (Section 4.2.4). The model is consistent with the observation that protein content is high in a Crabtree negative yeast (Section 4.2.5). The predicted abundance of the individual enzymes is not in good agreement with experimental values (Section 4.2.6), the model is however insensitive to perturbations (Section 4.2.7)

4.2.1 Glucose Uptake Saturation

The model predicts a maximal substrate uptake level. The predicted growth rate increases when the substrate uptake limit is increased. At a certain level. 15 mol per kg dry weight and hour. further increase does not improve growth rate. and the model does not predict higher glucose uptake. At this point the protein kinetics is limiting and there is no change in configuration that can make use of additional substrate.

As one could expect the maximum uptake and growth rate is proportionate to the CCMp parameter. Since the CCMp is also proportionate to the c parameter the results can be interpreted as a change in the c parameter. The predicted growth and uptake rates are in good agreement with experimental values for chemostat at high growth rates (Appendix C.1).

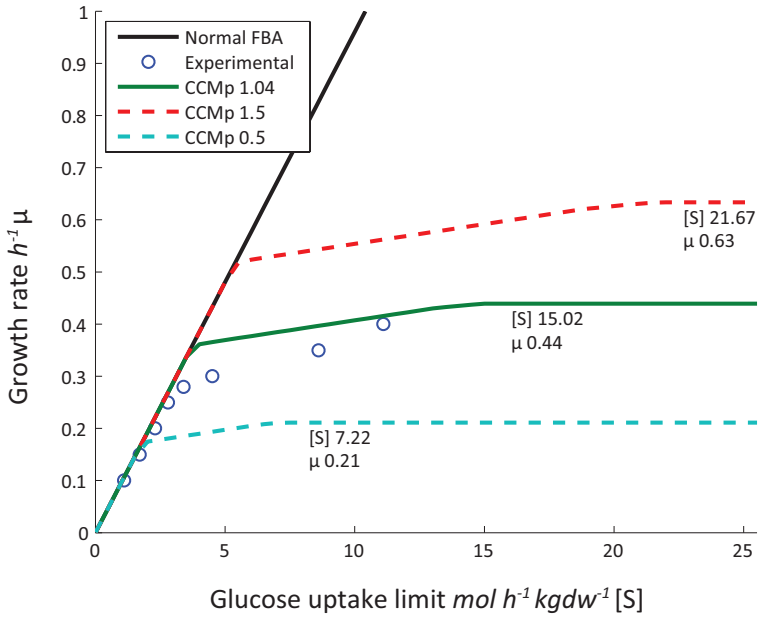

Figure 16: The uptake rate saturates at a level that depends on the CCMp parameter. Experimental data from literature [75].

4.2.2 The Switch Between 4 Different Metabolic Strategies

The model predicts 4 qualitatively different metabolic strategies (Figure 17) in the stages lending up to saturation:

1. A fully aerobic growth (**aerobic),** that is predicted by standard FBA.

2. An acetate excreting growth (**acetate**), that has been discussed in literature [55].
3. The respiro fermenting growth (**crabtree**) with negative correlation between glucose uptake and respiration [34].
4. A non acetate producing growth (**fermenting**).

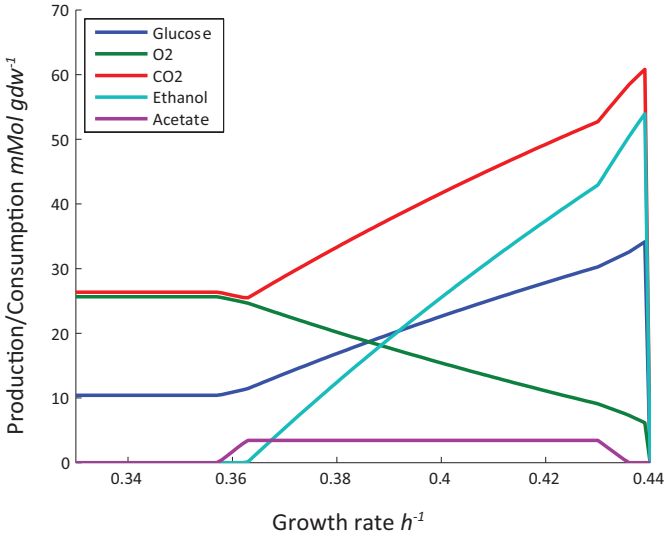

Figure 17: The model predicts 4 different metabolic strategies. Fully aerobic growth (up to 0.35), acetate excreting growth (up to 0.36), growth excreting ethanol and acetate (up till 0.43) and growth excreting only alcohol. For growth rates above 0.44 their is no strategy since this is the highest predicted growth rate.

At a certain growth rate the efficient **aerobic** metabolic strategy becomes to space demanding and the cell shifts to less efficient fermenting strategies changing many phenotypes of the system. A shift from the high yield strategy of 16.5 ATP per glucose to a mixture with the low yield strategy of 2 ATP per glucose. The shift follows from the the weight of proteins required for growth being lower for the **fermenting** strategy compared with the **arobic** strategy (Table 3). At high growth rates the **arobic** strategy would require more protein than is available and other strategies that are less optimal with regards to glucose utilization emerge.

Table 3: The predicted demand of enzymatic protein to uphold fluxes corresponding to growth by 0.1 h^{-1} for the **arobic** and the **fermentative** metabolic strategies. And the corresponding uptake of glucose.

Strategy	**Amino acids**	**Glucose**
	mol	*mol*
Arobic	0.291	1.04
Fermenting	0.237	3.41

The internal fluxes change as the strategy shifts (Figure 4). As one would expect the ATP synthase (ATP1) activity steadily decreases as the strategy shifts from **Aerobic** to **Fermenting**, mirroring the ethanol production. That the model still predicts some activity from ATP in **fermenting** growth strategy might follow from the definition of the protein production in the biomass equation. This reaction produces NADP in this model, it might be that ATP production is the best strategy to balance this. In a model including protein synthesis this will maybe be resolved in another way.

The fluxes through the pentos phosphate are lower during the **acetate** and **crabtree** growth, the NADPH required for growth is instead generated through acetate production. At the fermenting stage the the mitochondrial NADPH is produced by MAE, possibly reflecting that there is no shunt for NADPH from the cytosol to the mitocondria in the model.

Table 4: Normalized fluxes for key reactions. The values are first normalized by the growth rate and then divided by the maximum value for all growth rates.

	Aerobic	**Acetate**	**Crabtree**	**Fermenting**
Growth	0.3	0.36	0.4	0.439
PYC	0.8175	0.8175	0.8175	1
CIT	1	0.9382	0.4355	0.0804
MAE1	0	0	0	1
ZWF	1	0.6778	0.3435	1
ENO	0.2199	0.2406	0.6309	1
HXK	0.305	0.3193	0.6628	1
ethOUT	0	0	0.4734	1
acOUT	0	0.4907	1	0
ATP1	1.0000	0.9845	0.6032	0.2343

The model predicts a transition to fermentation at a higher growth rate than is given in most experiments. Running the model with experimentally determined biomass composition introduces the **crabtree** effect earlier (Figure 18). The model fits the data relatively well for the interval 0.2 to 0.35. Beyond this the model runs in to the protein bond and does not give predictions since the experimentally determined protein level is to low for the model to run.

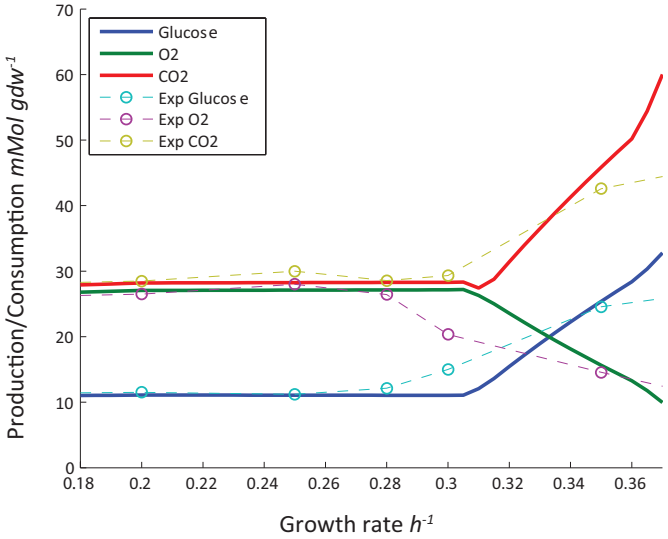

Figure 18: Experimental data compared with predicted data for some of the phenotypes of the Crabtree effect. For this comparison the model was run using experimentally determined composition data for the biomass equation and experimentally determined protein concentration with the CCMp parameter (Appendix C).

4.2.3 The Sum of Fluxes and Specific Activity Explain the Crabtree Effect

The higher protein requirement for the **aerobic** strategy is to one part an effect of the sum of fluxes (Table 5). Although the **aerobic** strategy produces 8.25 times more ATP per substrate, the sum of fluxes per ATP is only 0.7 times higher for the fermenting strategy. The specific activity of the glycolytic and fermentation enzymes therefore only need to be slightly higher in these pathways for them to be the lightest per ATP.

Table 5: Sum of fluxes for the main metabolic strategies, the galactose strategy will be discussed below (Section 4.2.4).

Strategy	**$\sum$Flux** $\frac{mol\ flux}{mol\ substrate}$	**ATP** $\frac{mol\ ATP}{mol\ substrate}$	**Flux per ATP** $\frac{mol\ flux}{mol\ ATP}$
Glycolysis	15		
Fermentation	4		
TCA + Oxidative Posphorylation	78.5		
Galactose	5		
Glycolysis + Fermentation	19	2	9.5
Glycolysis + TCA + Oxidative Posphorylation	93.5	16.5	5.67
Galactose + Glycolysis + Fermentation	24	2	12

The metabolic network was visualized with normalized specific activity color coded (Figure 19). The distribution of slow enzymes appears to be relatively evenly distributed over the network. It is mainly ATP1, KGD and RIP1 that make oxidative phosphorylation slow. The enzymes for acetate production are slower than ethanol producing possibly contributing to making it the favored excretion pathway.

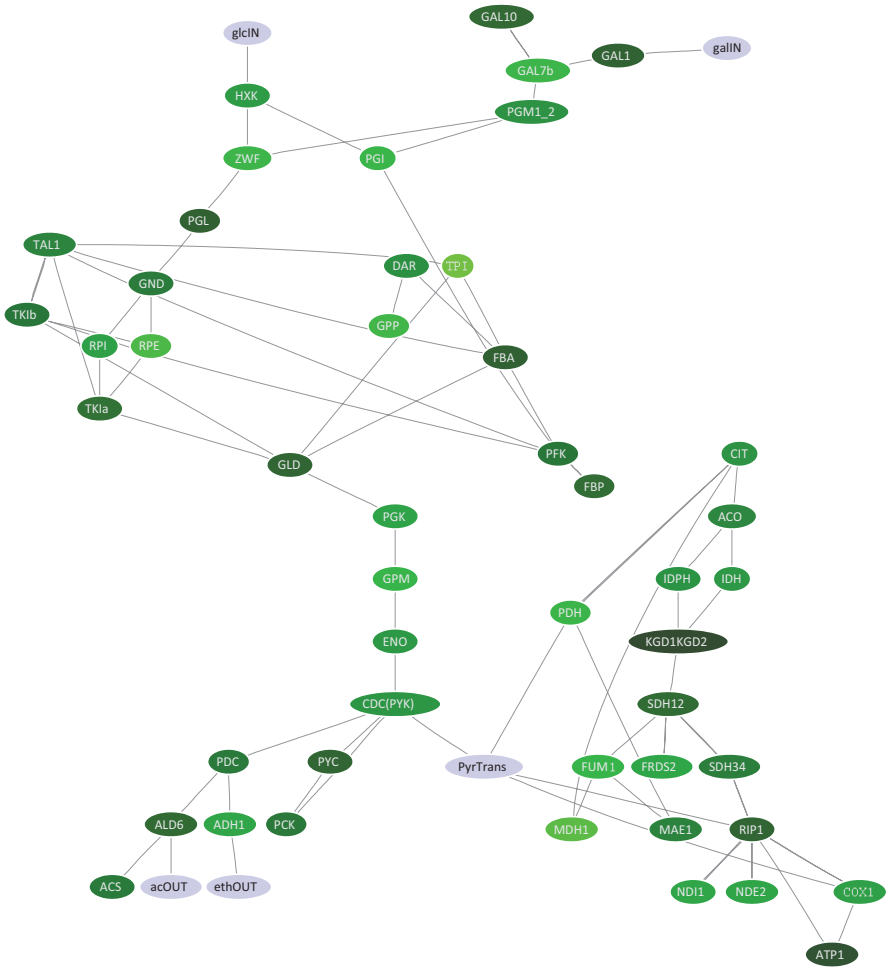

Figure 19: Visualization of the the enzymes specific activity in the network. Darker green indicates lower specific activity.

4.2.4 Lower Predicted Growth Rate for Galactose

The model predicts a significantly lower maximum growth rate for galactose than for glucose (Figure 20). The additional steps that are added on to glycolysis from the galactose conversion (Table 5) makes it more bulky, 0.31 mol amino acid are required to grow 0.1. It is mainly the proteins GAL1 and GAL10 that are responsible for this (Figure 19). The growth rate predicted by the model with experimental composition is in line with literature values [14].

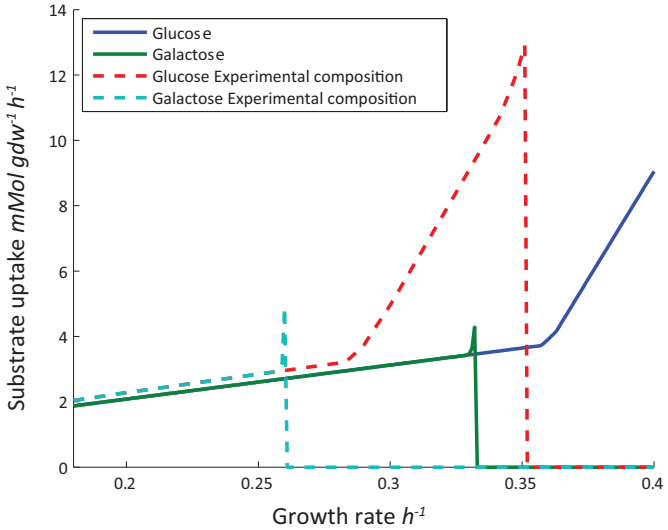

Figure 20: The predicted substrate uptake for glucose and galactose with standard model and the experimental composition model. The maximum growth rate for galactose is significantly lower than for glucose in both cases. The model predicts that a short transition to fermentation would be optimal. This realm is however very thin.

4.2.5 High Predicted Protein Content in the Crabtree negative *Kluyveromyces marxianus*

The model predicts the existence of a Crabtree effect in the Crabtree negative *Kluyveromyces marxianus,* assuming that its metabolic network and the specific activity of the enzymes is the same as for *S. cerevisiae.* However, the high protein content observed in *Kluyveromyces marxianus*, 72%, makes the transition take place at growth rates that are similar to the observed maximum growth rate. The model predicts that the acetate producing strategy is dominant at this growth rate which is consistent with experimental observations [26].

4.2.6 Prediction of Protein Abundance

The prediction of protein abundance in different pathways is better for the constrained model compared with unconstrained FBA (Figure 21), which is expected since the unconstrained model does not predict the shift in metabolic strategy.

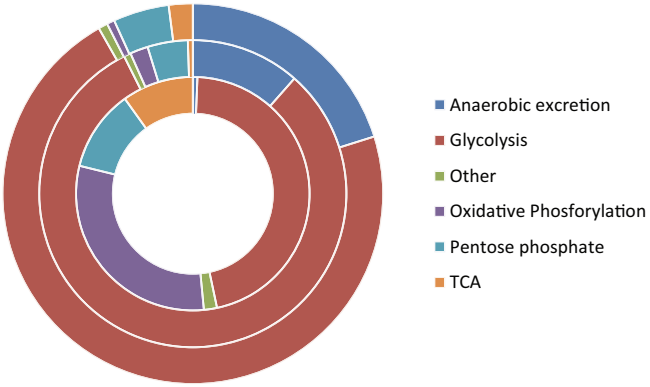

Figure 21: The relative abundance of proteins by pathway, experimental (outer), model predictions at high growth rate (middle) and predictions from standard FBA (inner).

The prediction of the abundance of individual proteins is not very good. A few proteins with low k_{cat} values tend to dominated the prediction (Figure 22).

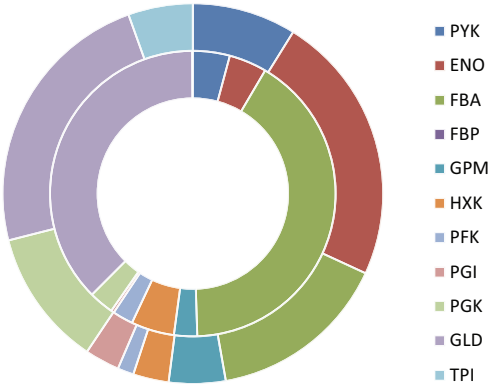

Figure 22: The relative abundance of enzymes in glycolysis, experimental (outer) and model prediction (inner).

4.2.7 Stability

The predictions of the model are relatively insensitive to perturbations of the k_{cat} values. The **fermenting** or **crabtree** strategy is predicted in all realizations for random perturbations of the k_{cat} values up to 20% (Figure

23). For perturbations up to 50% these strategies are predicted in 90% of the realizations. The predicted growth rates of the perturbations are distributed around the predicted value and the spread scales with the level of perturbation.

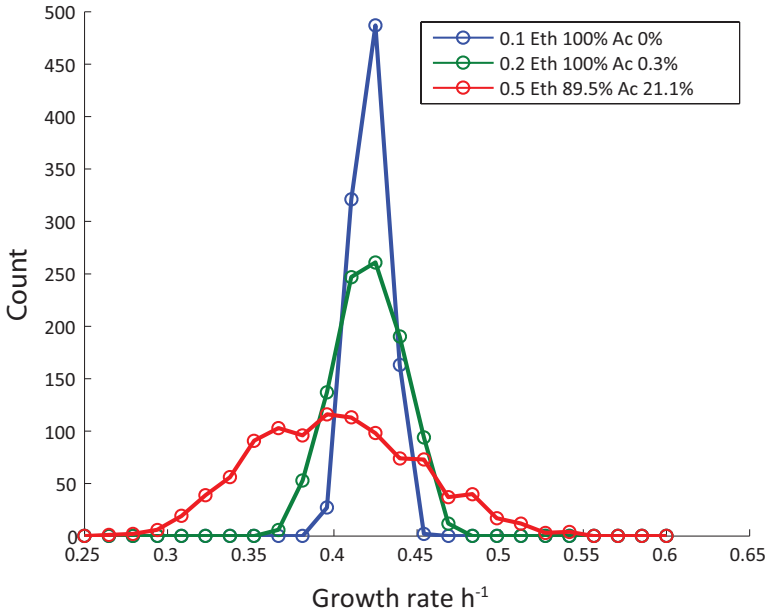

Figure 23: Histograms of the stability of the predictions of the system with increasing degrees of perturbation. Each k_{cat} is perturbed by a random value within the perturbation threshold, the distribution is the result of 1000 simulations with such perturbations.

The system is relatively insensitive to perturbations of the individual k_{cat} values (Figure 24). Only 11 k_{cat} values, 16%, have a significant effect on predicted growth rate and metabolic strategy when perpetuated up to 10 times their estimated value. The phenotype is strongly affected in 5 cases, and arises from changes in ATP syntheases (ATP1), Phosphofructokinase (PFK), Fructose-1,6-bisphosphatase (FBP), Triosephosphate dehydrogenase (GLD) and Pyruvate decarboxylase (PDC).

The 3 glycolytic enzymes PFK, FBP and GLD affect the system when their k_{cat} value is lowered, by increasing the bulkiness of glycolysis and making the **fermenting** strategy impossible. This results in a lower predicted growth rate. The change in PDC increases the bulkiness of the anaerobic pathway and acts in a similar way. Increased k_{cat} for ATP1 affects the system by making the **aerobic** strategy less bulky than the **fermenting**, resulting in higher predicted growth rates. The switch in strategy occurs already at 3

times higher k_{cat} values.

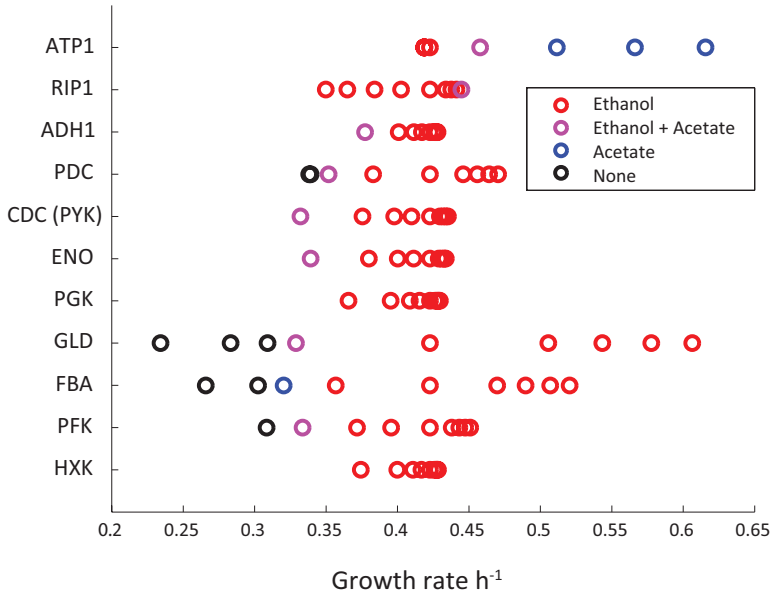

Figure 24: The results of perturbing an individual k_{cat} value by multiplication with 1/10, 1/5, 1/3, 1/2, 1, 2, 3, 5, 10. Only reactions with a difference between maximum an minimum growth rate of at least 10% or a change in the phenotype are shown. The phenotypes of the metabolic strategies are **fermentative** (red), **crabtree** (pink), **acetate** (blue) and **aerobic** (black).

4.3 Model of other Constraints

Limitations from non-modeled phenomena is expected to affect the cells. The increasing amount of proteins involved in protein synthesis at high growth rate limits the space for CCM (Section 4.3.1). The maximum uptake rate as a function of cell size shows that the uptake is far from saturated at low growth rates. As cells become larger with higher growth rate, the uptake might impose a limit (Section 4.3.2). The decreasing amount of carbohydrates in the cell wall as a consequence of increased size is shown to only have a small effect on the biomass composition (Section 4.3.3)

4.3.1 The Ribosomes Constrain the Space for CCM

The amount of protein required for protein synthesis increases with increasing growth rate (Figure 25). The total amount of protein increases faster up to a growth rate of 0.3, increasing the free space. After this the total

protein increases slower than the protein required for protein synthesis and the free space decreases.

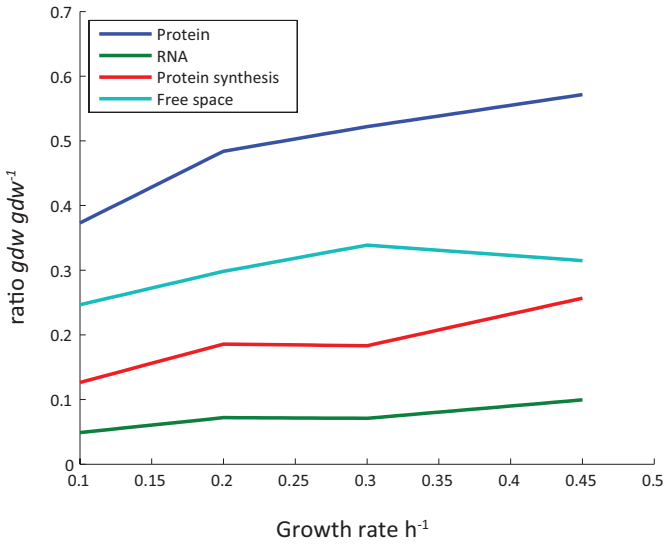

Figure 25: The amount of protein not involved in protein synthesis (Free space) increases up until a growth rate of 0.3. The proteins involved in protein synthesis is assumed to be proportional to RNA. Data from regression of composition data [63] (Appendix C.2).

If the space for CCM is proportional to the free space, then the space for CCM starts to decrease around the growth rate of 0.3, which is when the **acetate** and **crabtree** strategies are observed to begin. The model also predicts that the space reserved for CCM reaches 100% utilization at a growth rates of 0.3, using experimental protein amounts. The utilization of the free space increases more or less linearly from 43% at the growth rate 0.1.

Under the assumption that a certain amount of protein is fixed, e.g. DNA binding, and does not increase with increasing cell size, the free space can become at best constant (Figure 26). Assuming a larger fixed cost would lead to negative free space at low growth rates. Even when assuming a large amount of fixed protein the free space would eventually decrease at higher growth rates, reflecting the ever growing demand for space by the ribosomes.

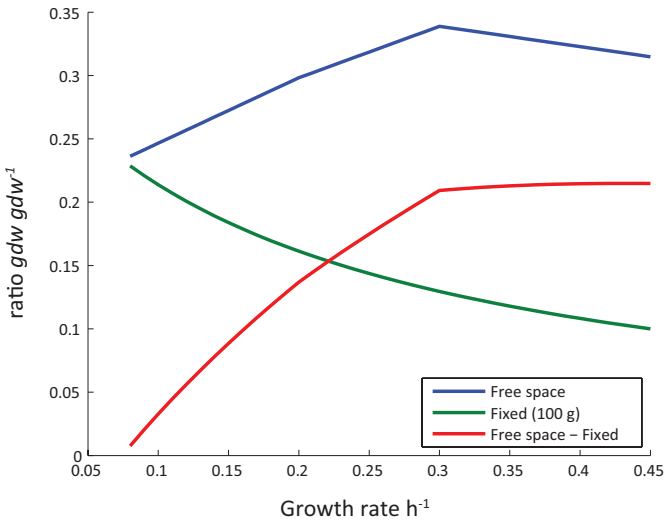

Figure 26: Even by assuming that a large fraction of the proteins are independent of the cell size, the space available for CCM only stabilizes. Data from regression of composition data [63] (Appendix C.2).

4.3.2 Uptake Rate as a Function of Size

Increasing growth rate is assumed to result in increasing cell size. The surface volume ratio decreases at high growth rates decreasing the maximum uptake rate. This rate might be limiting at high growth rates (Figure 27). At low growth rates it might still be influential as the true uptake is likely to be a function of the maximum, but it will not be limiting. The volume for the predicted requirement curves is estimated from the growth rate vs volume equation (Equation 8). The predicted growth rate for the reference value is approximately 10% higher than predicted indicating that the extrapolation of the equation does not give very accurate results.

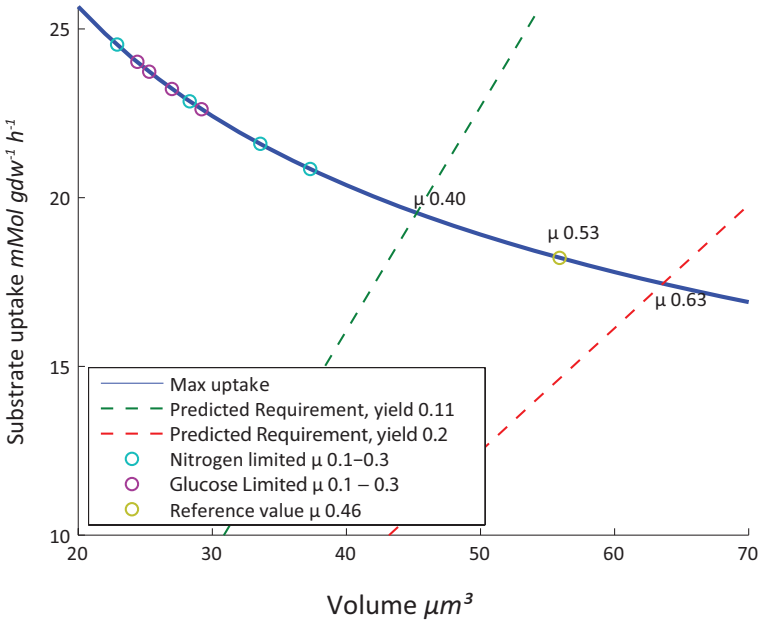

Figure 27: The predicted maximum uptake rate depends on the volume of the cell. The reference value [2] is marked with a yellow circle. At this size the cell volume is approximately 56 μm^3 [2] and the surface volume ratio 1.27. The blue line scales the predicted uptake at this size to other volumes with the relative surface volume ratio. The predicted consumption for a respiro fermenting (yield 0.2) and a fermenting cell (yield 0.11) are shown with dashed lines. The predicted growth rate is shown at the intersection. Cell volumes from nitrogen and glucose limited experiments are marked on the line with circles.

4.3.3 Cell Wall as a Function of Size

Assuming a constant thickness of the cell wall its share of the biomass is expected to decrease with the surface volume ratio. The model shows a reasonably good fit in the range $\mu = 0.2$ to $\mu = 0.4$. For $\mu = 0.1$ the fit is not so good. probably reflecting that the measurement of "Other carbohydrates" includes storage molecules. The decrease is relatively small. Even at volumes that would be associated with a growth of 0.7 h^{-1} the share barely goes below 0.2, indicating that decreasing the cell wall only gives a small contribution to the observed increase of protein with growth.

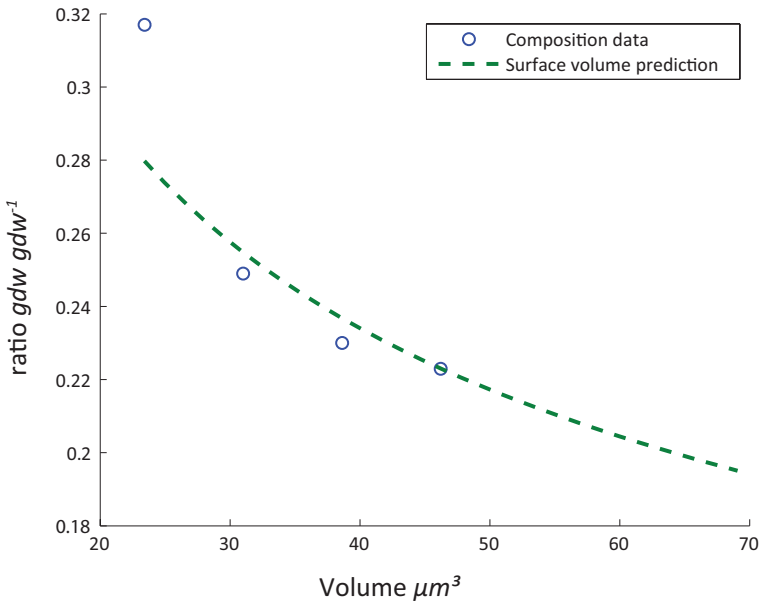

Figure 28: An estimate of the ratio of biomass in the cell wall based on the empirical formula for cell volume (Equation 8) and composition data at different growth rates (Appendix C.2). The Mannan together with "Other carbohydrates" is assumed to make up the cell wall.

5 Discussion

The purpose of the thesis was to investigate if constrains on the fluxes by the amount of involved enzymes can affect predictions. It has been shown that these constrains become important at high growth rates, and that they affects what is the most optimal metabolic strategy. Both the Crabtree effect (Section 5.4) and the lower maximum growth rate on galactose can be explained by the model.

The prediction of the individual enzyme concentrations is not so good, this might have several reasons (Section 5.1). The sensitivity analysis indicates that this probably is not a problem for the qualitative results.

At low growth rates the model under-predicts the total amount of enzymes in central carbon metabolism, as the CCMp constraint is not enforced (Section 5.2). This might be caused by defacto lower space for CCM at these rates or by lower enzymatic efficiency.

It has been observed that cells tend towards larger sizes at higher growth rates. This is probably a method to increase active protein concentrations. As size increases the maximum uptake rate decreases, possibly putting a limit on the increase of size and thereby growth (Section 5.3).

Gene deletion phenotype prediction, e.g. gene essentiality, is a common technique for testing FBA models. Such experiments have not been preformed in this project but the model is expected to preform well for such tests (Section 5.5).

Some considerations for future models (Section 5.6) and some consequences for bio engineering (Section 5.7) are also discussed.

5.1 Possible Reasons for the Inaccurate Prediction of Enzyme Concentrations

The predictions of the individual enzyme concentrations does not fit very well with experimental data. This is likely an effect of the c parameter being different for each enzyme. For reversible reactions the relation in concentration between substrate and product is important. Several such reactions in a row would result in lower concentrations at the final metabolite. A method to predict the enzyme concentrations directly from these considerations has been suggested in literature [69]. However implementing and evaluating this method was beyond the scope of this project.

The level of backwards flux might have a huge impact on the predicted amount since the model only takes net fluxes in to account. This might be the reason that the model under-predicts the amount of the ethanol producing ADH1. This enzyme plays an important role in ethanol uptake and its therefore likely that it is very sensitive to substrate and product concentrations.

Metabolite dilution was not taken in to account in the model. Enzymes with large product pools will therefore appear to be slower than enzymes with small pools. The pool sizes are however relatively small compared with the total flux so the effect should be very limited.

Apart from differences in the c parameter the metabolites might differ due to slight differences in metabolic strategy. From abundance data it appears as if the model over-predicts oxidative phosphorylation activity. This might be a response to the NADH formed by the protein synthesis pseudo-reaction (Section A.4).

5.2 The Low Enzyme Utilization at Low Growth Rate

The model predicts that CCM only takes up 40% of the available space at low growth rates. Since the proteins still exists this must be inaccurate. In part be explained by the CCMp parameter over predicting the available space. A consequence of lower levels of non ribosomal protein and the existence of fixed costs (Section 4.3.1). It can also be interpreted as the efficiency of the proteins being 40%, then the enzymes have a lower c at low growth rates, and they therefore take more space per flux. This is commonly assumed in literature [73]. A reason for this might be lower average metabolite concentrations.

There exists some experimental evidence that the average metabolite concentration is lower at low growth rates (Section C.3, Figure 31) [13]. This could possibly be linked to the observed lower extracellular glucose concentration at lower growth rates.

It could be expected that metabolite concentrations are higher under nitrogen limited growth, since CCM can be saturated and only protein formation will be underutilized. This could explain why the protein concentrations from the nitrogen composition experiments where too low for the model at high growth rates.

5.3 Effects of Cell Size

The results suggest that cells might become larger to increase their protein concentration. The demand for higher protein levels is a consequence of increased demand for both CCM and protein synthesis. For high growth rates (above 0.3) the increase in protein corresponds to the increase in protein synthesis making little further space for CCM. The results also suggest that the increase in cell size might be bounded by the maximum uptake rate.

It is likely that the surface volume ratio puts a pressure on cells to become as small as possible, since the maximum uptake capacity increases and thereby the uptake even at sub maximal sizes. The increases in size and subsequent decrease in uptake efficiency is likely the reason for the high residual glucose concentrations observed at higher dilution rates.

Although it is true that growth maximization for a certain glucose level and glucose minimization for a certain growth level is equivalent, it is not strictly necessary for the organisms to minimize glucose consumption. The fact that this gives good predictions might be a consequence of competition for glucose that decreases the total glucose levels and favors a low glucose strategy.

If large size is indeed disfavored, it could explain why *S. cerevisiae* has a budding replication system. This would allow the mother as well as the daughter cell to be smaller.

The cells appear to never become smaller than a certain size. This might reflect a disposition towards carbohydrate storage rather than increased uptake at low growth rates. It could also reflect that the fixed costs for DNA replication etc have to be maintained by a certain amount of energy producing units.

The surface volume ratio analysis assumes that the density is constant. By using vacuoles the cell could increase the surface without increasing the active volume to the same extent and thereby achieve a higher surface/(active volume) ratio. This has not been taken in to account in the analysis.

5.4 The Crabtree Effect

The Crabtree metabolic strategy is the most enzyme efficient. It can however be avoided by increasing the enzyme concentration, which is the case for *Kluyveromyces marxianus.* One reason for this strategy to be disfavored

might be that nitrogen is scarce in nature. Since the ethanol can later be digested the long term benefits of high substrate utilization might be limited.

A further motivation is that these protein levels can not be reached without decreasing the thickness of the cell wall (Section 4.3.3). This might compromise the stability of the organism.

That the Crabtree effect is predicted to occur also in the Crabtree negative *Kluyveromyces marxianus* might be explained by some other limitation being dominant. This limitation could occur at lower growth rates than where the Crabtree effect is expected to take place. These limitation could arise from large cell sizes and therefore substrate uptake limitation. The low biomass concentration (1.31 g DW l^{-1}) and the high residual glucose concentration (7.25 g l^{-1}) obtained in experiments [26], might indicate that there is an uptake limitation caused by size, or that the intracellular concentrations of metabolites needs to be very high.

The Crabtree effect can be induced at lower growth rates by stress [55]. This could be a consequence of increased metabolism for stress related ATP consumption and thereby an increasing load on CCM. It could also be the effect of increase in stress related proteins decreasing the space for CCM.

A major cause of the Crabtree effect is the comparatively low ATP yield from oxidative phosphorylation. The low yield is a consequence of the lack of a proton pumping Complex 1 (Section 1.2.1). It is however possible that this makes the protein lighter and faster, which decreases the miss fitting. Since humans have a proton pumping Complex 1 it is possible that the the Warburg effect has a different origin than the Crabtree. One possibility could be limited oxygen availability. A recent study however suggests that molecular crowding also could be the cause of the Warburg effect [76]. Assuming that the Warburg effect is required for fast growth, pharmaceuticals decreasing the speed of the glycolytic or fermenting enzymes could be a potential therapy for cancer, this has also been suggested in literature [18].

It is important to stress that the model does not take regulation in to account. This implies that the results are in no conflict with results showing that the Crabtree effect is a consequence of regulation. It is in fact expected that the cell needs to regulate it self to undergo the necessary metabolic strategy shift. Since the Crabtree effect be avoided by increasing the protein levels, the results may also be consistent with the suggestion that the Crabtree effect is an act of chemical warfare.

The Crabtree metabolic strategy requires less mitocondrial enzymes, it is therefore likely that the mitochondria contributes less to the total dry weight

at high growth rates.

5.5 Probable Effects of Gene Deletion

The deletion of non essential genes has been shown to increase growth under some conditions [42]. This can be explained by the model, as this would give more space for growth improving proteins. The fact that yeast normally expresses these growth limiting genes might be that they increase robustness, fitness for other conditions. It can not be expected that yeast has been fully optimized for life in scientific or industrial conditions [42].

Gene deletions results in rerouting of the metabolic pathways. It has not been tested, but it is probable that a models taking enzyme kinetics in to account will give better estimates on the growth effects of gene deletions, since rerouting of the fluxes around the deleted pathways will increase the protein load, predicting lower growth rates.

5.6 Practical Considerations With Regards to Modeling

This thesis generally assumes that there is 1 active site per protein. This should in general be fine since k_{cat} experiments normally will calculate the activity per participating protein.

The setup requires that the modeler notices when a protein is a part of a complex. It also requires that the stoichiometry of the complex is known. A further complication that has not been taken in to account is the additional structural requirements for some proteins e.g. proteins required for assembly, the lipids of mitochondria or other compartments etc. Since the proteins participation in complexes have a large effect on the specific activity it is advisable to include them. The lack of systematic information on complexes is a hinder for effective modeling. Reliable complex databases would simplify the process.

Because of the sensitivity to low k_{cat} a few enzymes tend to take up most of the space. The data for slow enzymes needs to be analyzed carefully since many database values predate genome sequencing and might rely on incorrect protein weights. This could be circumvented by using specific activity directly. The sensitivity analysis shows that although careful selection is preferred, the same qualitative results can be achieved with modest data quality. The specific activity can however not be set to the same value for all enzymes as seen by the flux calculations (Table 5).

Since large fluxes result in high predicted protein content, lower fluxes will automatically be predicted at high growth rates. This improves the likely hood of finding a unique solution to the optimization problem [67] and resolves loops [7].

5.7 Consequences for Bio Engineering

Assuming that protein is limited, experimentalists should consider gene deletion as their main strategy for archiving their objectives. Viable strains with several knockouts could be the bases for new high productive strains. When including new proteins the specific activity of the new genes should be considered.

Since k_{cat} is known to increase with temperature, creating temperature tolerant strains should make higher growth rates possible. It has also been shown in an ANOVA study that higher temperature improves growth [5].

6 Outlook

The model has only been implemented for CCM, several other pathways such as the amino acid and protein synthesis pathways might be included with good results (Section 6.3).

6.1 Extending the Model to the Whole Metabolic Network

Since cost is a systems phenomena, neglecting parts of the system might affect predictions. Integrating FBA into a framework that takes uptake processes, metabolite concentrations and translation speed might give even better predictions. Such models might be able to take cell size as a constraint, which might be useful since it is experimentally relatively simple to determine.

6.2 A Tool for Synthetic Biologists

Since the model relies on k_{cat} values that can be determined in vitro and protein weights that are available in UniProt. It should be relatively simple to extend existing FBA models with these constraints. Such models might become of use to metabolic engineers in predicting if a potential pathway might decrease cell fitness. And if their pathway is predicted to be a favored metabolic strategy.

6.3 Predicting Izosymes

Izosymes have different amino acid compositions, weight and kinetic parameters. It is possible that there is a tradeoff between expensive amino acids from nutrient perspective, and expensive amino acids from the weight of the protein synthesis machinery and production of NADH. The choice of isozymes might therefore reflect a trade of between these properties. The choice of amino acid as the weight definition in the model is a preparation for a model that predicts the amino acid composition of the cell from the enzymes being used.

Another possible trade of with regards to weight is between high specific activity but low affinity. This has to some extent been shown for glucose transporters [32].

6.4 Improving the Enzyme Concentration Predictions

Allowing a prediction of the individual c parameters would drive the model close to becoming a kinetic model. But using the Gibbs free energy formalism ΔG might give a rough suggestion to what the metabolite concentrations might be and thereby if the enzymes will be operating at full speed.

If the prediction of protein abundance can be improved it would be possible to run the simulation backwards using protein abundance data. This could in general be useful when running experiments predicting the true c parameters.

6.5 Maintenance Cost and Protein Degradation

The maintenance cost in the biomass equation is likely to change with growth rate (Section A.4). This has not been taken in to consideration, but might have a quantitative impact on the results. It is largely known which phenomena causing the maintenance cost to occur. So modeling it should be possible.

The attempts at modeling the protein degradation as a part of the maintenance cost was hindered by inconsistent data on protein half life. Literature suggest that the median half life is 40 min[9] or 10 h [33]. It was further hindered by nonphysically low literature estimates on ribosomal translation speed (Section F.1).

Cell size is also likely to affect maintenance cost since intracellular pH is defended at a smaller relative surface when the cells are large (Section A.4).

7 Conclusion

Models with kinetic constraints show superior prediction power compared with unconstrained FBA. They are able to predict a maximum growth rate. They predict a lower maximum growth rate for substrates with expensive additional pathways. The also can predict shifts between several different metabolic strategies, aerobic, acetate producing, respiro fermenting (Crabtree) and fermenting.

The model shows that the Crabtree effect can be the favored metabolic strategy when there is a limited protein amount. The effect is a network property that arises as a result of a lower protein weight per produced ATP for the fermentation pathways. This effect is probably specific for organisms that have a low ATP yield for oxidative phosphorylation. And that have a disposition towards low protein concentrations. The cell wall in *Saccharomyces cerevisiae* create such a disposition.

The results of the model demonstrate that growth maximization can be a viable concept in predicting the preferred metabolic strategies of an organism given that all relevant constraints are in place.

Simple calculations indicate that increase in cell size with increased growth may be caused by the increased demand for CCM and ribosomal protein. The decreasing uptake rate with increasing cell size may put an upper bound to the increase in cell size.

Kinetically constrained models may become a simple supplement to FBA, that will allow modelers as well as experimentalists to get more accurate predictions when planing pathway additions and substitutions to organisms.

References

[1] Roi Adadi, Benjamin Volkmer, Ron Milo, Matthias Heinemann, and Tomer Shlomi. Prediction of Microbial Growth Rate versus Biomass Yield by a Metabolic Network with Kinetic Parameters, 2012.

[2] J Adams and P E Hansche. Population studies in microorganisms. I. Evolution of diploidy in Saccharomyces cerevisiae. *Genetics*, 76:327–338, 1974.

[3] Rasmus Agren, Liming Liu, Saeed Shoaie, Wanwipa Vongsangnak, Intawat Nookaew, and Jens Nielsen. The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum. *PLoS Computational Biology*, 9, 2013.

[4] K R Albe, M H Butler, and B E Wright. Cellular concentrations of enzymes and their substrates. *Journal of theoretical biology*, 143:163–195, 1990.

[5] F. N. Arroyo-L??pez, Sandi Orli??, Amparo Querol, and Eladio Barrio. Effects of temperature, pH and sugar concentration on the growth parameters of Saccharomyces cerevisiae, S. kudriavzevii and their interspecific hybrid. *International Journal of Food Microbiology*, 131:120–127, 2009.

[6] Daniel E Atkinson. *Cellular Energy Metabolism and Its Regulation.* New York: Academic, 1977.

[7] Daniel A Beard, Shou-dan Liang, and Hong Qian. Energy balance for analysis of complex metabolic networks. *Biophysical journal*, 83:79–86, 2002.

[8] Q K Beg, A Vazquez, J Ernst, M A de Menezes, Z Bar-Joseph, A-L Barabási, 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*, 104:12663–12668, 2007.

[9] Archana Belle, Amos Tanay, Ledion Bitincka, Ron Shamir, and Erin K O'Shea. Quantification of protein half-lives in the budding yeast proteome. *Proceedings of the National Academy of Sciences of the United States of America*, 103:13004–13009, 2006.

[10] A. Ben-Shem, N. Garreau de Loubresse, S. Melnikov, L. Jenner, G. Yusupova, and M. Yusupov. The Structure of the Eukaryotic Ribosome at 3.0 A Resolution, 2011.

[11] Bryson D Bennett, Elizabeth H Kimball, Melissa Gao, Robin Osterhout, Stephen J Van Dien, and Joshua D Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature chemical biology*, 5:593–599, 2009.

[12] Tomer Benyamini, Ori Folger, Eytan Ruppin, and Tomer Shlomi. Flux balance analysis accounting for metabolite dilution. *Genome biology*, 11:R43, 2010.

[13] Viktor M Boer, Christopher A Crutchfield, Patrick H Bradley, David Botstein, and Joshua D Rabinowitz. Growth-limiting intracellular metabolites in yeast growing under diverse nutrient limitations. *Molecular biology of the cell*, 21:198–211, 2010.

[14] Christoffer Bro, Steen Knudsen, Birgitte Regenberg, Lisbeth Olsson, and Jens Nielsen. Improvement of galactose uptake in Saccharomyces cerevisiae through overexpression of phosphoglucomutase: example of transcript analysis as a tool in inverse metabolic engineering. *Applied and environmental microbiology*, 71:6465–6472, 2005.

[15] P. M. Bruinenberg, J. P. Van Dijken, and W. A. Scheffers. A Theoretical Analysis of NADPH Production and Consumption in Yeasts, 1983.

[16] André B Canelas, Nicola Harrison, Alessandro Fazio, Jie Zhang, Juha-Pekka Pitkänen, Joost van den Brink, Barbara M Bakker, Lara Bogner, Jildau Bouwman, Juan I Castrillo, Ayca Cankorur, Pramote Chumnanpuen, Pascale Daran-Lapujade, Duygu Dikicioglu, Karen van Eunen, Jennifer C Ewald, Joseph J Heijnen, Betul Kirdar, Ismo Mattila, Femke I C Mensonides, Anja Niebel, Merja Penttilä, Jack T Pronk, Matthias Reuss, Laura Salusjärvi, Uwe Sauer, David Sherman, Martin Siemann-Herzberg, Hans Westerhoff, Johannes de Winde, Dina Petranovic, Stephen G Oliver, Christopher T Workman, Nicola Zamboni, and Jens Nielsen. Integrated multilaboratory systems biology reveals differences in protein metabolism between two reference yeast strains. *Nature communications*, 1:145, 2010.

[17] Ross P Carlson. Metabolic systems cost-benefit analysis for interpreting network structure and regulation. *Bioinformatics (Oxford, England)*, 23:1258–1264, 2007.

[18] Zhao Chen, Weiqin Lu, Celia Garcia-Prieto, and Peng Huang. The Warburg effect and its cancer therapeutic implications, 2007.

[19] J Michael Cherry, Eurie L Hong, Craig Amundsen, Rama Balakrishnan, Gail Binkley, Esther T Chan, Karen R Christie, Maria C Costanzo, Selina S Dwight, Stacia R Engel, Dianna G Fisk, Jodi E Hirschman, Benjamin C Hitz, Kalpana Karra, Cynthia J Krieger, Stuart R Miyasato, Rob S Nash, Julie Park, Marek S Skrzypek, Matt Simison, Shuai Weng, and Edith D Wong. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic acids research*, 40:D700–5, 2012.

[20] Pascale Daran-Lapujade, Mickel L A Jansen, Jean-Marc Daran, Walter van Gulik, Johannes H de Winde, and Jack T Pronk. Role of transcriptional regulation in controlling fluxes in central carbon metabolism of Saccharomyces cerevisiae. A chemostat culture study. *The Journal of biological chemistry*, 279:9125–9138, 2004.

[21] K. M. Davies, C. Anselmi, I. Wittig, J. D. Faraldo-Gomez, and W. Kuhlbrandt. Structure of the yeast F1Fo-ATP synthase dimer and its role in shaping the mitochondrial cristae, 2012.

[22] Iman Famili, Jochen Forster, Jens Nielsen, and Bernhard O Palsson. Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proceedings of the National Academy of Sciences of the United States of America*, 100:13134–13139, 2003.

[23] M. E. Feder and J. C. Walser. The biological limitations of transcriptomics in elucidating stress and stress responses. In *Journal of Evolutionary Biology*, volume 18, pages 901–910, 2005.

[24] Francisco Ferrezuelo, Neus Colomina, Alida Palmisano, Eloi Garí, Carme Gallego, Attila Csikász-Nagy, and Martí Aldea. The critical size is set at a single-cell level by growth rate to attain homeostasis and adaptation, 2012.

[25] Avi Flamholz, Elad Noor, Arren Bar-Even, Wolfram Liebermeister, and Ron Milo. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences of the United States of America*, 110:10039–44, 2013.

[26] Gustavo Graciano Fonseca, Andreas Karoly Gombert, Elmar Heinzle, and Christoph Wittmann. Physiology of the yeast Kluyveromyces

marxianus during batch and chemostat cultures with glucose as the sole carbon source. *FEMS Yeast Research*, 7:422–435, 2007.

[27] Jochen Förster, Iman Famili, Patrick Fu, Bernhard ØPalsson, and Jens Nielsen. Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome research*, 13:244–253, 2003.

[28] Kathrin Förster, Paola Turina, Friedel Drepper, Wolfgang Haehnel, Susanne Fischer, Peter Gräber, and Jan Petersen. Proton transport coupled ATP synthesis by the purified yeast H+-ATP synthase in proteoliposomes. *Biochimica et Biophysica Acta - Bioenergetics*, 1797:1828–1837, 2010.

[29] 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*, 262:3754–3761, 1987.

[30] Anisha Goel, Meike Tessa Wortel, Douwe Molenaar, and Bas Teusink. Metabolic shifts: a fitness perspective for microbial cell factories, 2012.

[31] Anne Goelzer, Vincent Fromion, and Gérard Scorletti. Cell design in bacteria as a convex optimization problem. *Automatica*, 47:1210–1218, 2011.

[32] J??rg Hauf, Friedrich K. Zimmermann, and Susanne M??ller. Simultaneous genomic overexpression of seven glycolytic enzymes in the yeast Saccharomyces cerevisiae. In *Enzyme and Microbial Technology*, volume 26, pages 688–698, 2000.

[33] Andreas O. Helbig, Pascale Daran-Lapujade, Antonius J. A. van Maris, Erik A. F. de Hulster, Dick de Ridder, Jack T. Pronk, Albert J. R. Heck, and Monique Slijper. The diversity of protein turnover and abundance under nitrogen-limited steady-state conditions in Saccharomyces cerevisiae, 2011.

[34] Jan Heyland, Jianan Fu, and Lars M Blank. Correlation between TCA cycle flux and glucose uptake rate during respiro-fermentative growth of Saccharomyces cerevisiae. *Microbiology*, 155:3827–3837, 2009.

[35] P I M V a N Hoek, Johannes P V a N Dijken, and Jack T Pronk. Effect of Specific Growth Rate on Fermentative Capacity of Baker's Yeast. *Society*, 64:4226–4233, 1998.

[36] Hermann-Georg Holzhütter. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *European journal of biochemistry / FEBS*, 271:2905–2922, 2004.

[37] Carola Hunte, Juergen Koepke, Christian Lange, Tanja Ro??manith, and Hartmut Michel. Structure at 2.3 ?? resolution of the cytochrome bc1 complex from the yeast Saccharomyces cerevisiae co-crystallized with an antibody Fv fragment. *Structure*, 8:669–684, 2000.

[38] JonathanÂ R. Karr, JayoditaÂ C. Sanghvi, DerekÂ N. Macklin, MiriamÂ V. Gutschow, JaredÂ M. Jacobs, Benjamin Bolival, Nacyra Assad-Garcia, JohnÂ I. Glass, and MarkusÂ W. Covert. A Whole-Cell Computational Model Predicts Phenotype from Genotype, 2012.

[39] Vikram Khurana and Susan Lindquist. Modelling neurodegeneration in Saccharomyces cerevisiae: why cook with baker's yeast? *Nature reviews. Neuroscience*, 11:436–449, 2010.

[40] Frans M Klis, Chris G de Koster, and Stanley Brul. Cell Wall-Related Bionumbers and Bioestimates of Saccharomyces cerevisiae and Candida albicans. *Eukaryotic cell*, 13:2–9, 2014.

[41] Stefan Klumpp, Matthew Scott, Steen Pedersen, and Terence Hwa. Molecular crowding limits translation and cell growth. *Proceedings of the National Academy of Sciences of the United States of America*, 110:16754–9, 2013.

[42] Karl Kochanowski, Uwe Sauer, and Victor Chubukov. Somewhat in control-the role of transcription in regulating microbial metabolic fluxes, 2013.

[43] Joshua A. Lerman, Daniel R. Hyduke, Haythem Latif, Vasiliy A. Portnoy, Nathan E. Lewis, Jeffrey D. Orth, Alexandra C. Schrimpe-Rutledge, Richard D. Smith, Joshua N. Adkins, Karsten Zengler, and Bernhard O. Palsson. In silico method for modelling metabolism and gene product expression at genome scale, 2012.

[44] Nathan E Lewis, Kim K Hixson, Tom M Conrad, Joshua A Lerman, Pep Charusanti, Ashoka D Polpitiya, Joshua N Adkins, Gunnar Schramm, Samuel O Purvine, Daniel Lopez-Ferrer, Karl K Weitz, Roland Eils, Rainer König, Richard D Smith, and Bernhard ØPalsson. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Molecular systems biology*, 6:390, 2010.

[45] Nathan E. Lewis, Harish Nagarajan, and Bernhard O. Palsson. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods, 2012.

[46] 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*, 2:e146v4, 2014.

[47] Douwe Molenaar, Rogier van Berlo, Dick de Ridder, and Bas Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular systems biology*, 5:323, 2009.

[48] Ettore Murabito, Evangelos Simeonidis, Kieran Smallbone, and Jonathan Swinton. Capturing the essence of a metabolic network: a flux balance analysis approach. *Journal of theoretical biology*, 260:445–452, 2009.

[49] T L Nissen, U Schulze, J Nielsen, and J Villadsen. Flux distributions in anaerobic, glucose-limited continuous cultures of Saccharomyces cerevisiae. *Microbiology*, 143 (Pt 1:203–218, 1997.

[50] Intawat Nookaew, Michael C Jewett, Asawin Meechai, Chinae Thammarongtham, Kobkul Laoteng, Supapon Cheevadhanarak, Jens Nielsen, and Sakarindr Bhumiratana. The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: a scaffold to query lipid metabolism. *BMC systems biology*, 2:71, 2008.

[51] Edward J O'Brien, Joshua a Lerman, Roger L Chang, Daniel R Hyduke, and Bernhard ØPalsson. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular systems biology*, 9:693, 2013.

[52] H Ogata, S Goto, K Sato, W Fujibuchi, H Bono, and M Kanehisa. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids research*, 27:29–34, 1999.

[53] Rob Phillips, Jane Kondev, and Julie Theriot. *Physical biology of the cell*. Garland Science, 2009.

[54] Rob Phillips and Ron Milo. A feeling for the numbers in biology. *Proceedings of the National Academy of Sciences of the United States of America*, 106:21465–21471, 2009.

[55] E Postma, C Verduyn, W A Scheffers, and J P Van Dijken. Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of Saccharomyces cerevisiae. *Applied and environmental microbiology*, 55:468–477, 1989.

[56] Jennifer L. Reed. Shrinking the Metabolic Solution Space Using Experimental Datasets, 2012.

[57] Christoph Ruckenstuhl, Didac Carmona-Gutierrez, and Frank Madeo. The sweet taste of death: glucose triggers apoptosis during yeast chronological aging. *Aging (Albany, NY. Online)*, 2:643–649, 2010.

[58] Naoyoshi Sakaki, Rieko Shimo-Kon, Kengo Adachi, Hiroyasu Itoh, Shou Furuike, Eiro Muneyuki, Masasuke Yoshida, and Kazuhiko Kinosita. One rotary mechanism for F1-ATPase over ATP concentrations from millimolar down to nanomolar. *Biophysical journal*, 88:2047–2056, 2005.

[59] R Scherrer, E Berlin, and Gerhardt. Density, porosity, and structure of dried cell walls isolated from Bacillus megaterium and Saccharomyces cerevisiae. *Journal of bacteriology*, 129:1162–1164, 1977.

[60] Ida Schomburg, Antje Chang, Sandra Placzek, Carola Söhngen, Michael Rother, Maren Lang, Cornelia Munaretto, Susanne Ulas, Michael Stelzer, Andreas Grote, Maurice Scheer, and Dietmar Schomburg. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic acids research*, 41:D764–72, 2013.

[61] R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer. Multidimensional Optimality of Microbial Metabolism, 2012.

[62] U Schulze, G Lidén, J Nielsen, and J Villadsen. Physiological effects of nitrogen starvation in an anaerobic batch culture of Saccharomyces cerevisiae. *Microbiology*, 142 (Pt 8:2299–2310, 1996.

[63] Ulrik Schulze. *Anaerobic physiology of Saccharomyces cerevisiae*. PhD thesis, Lyngby, Technical University of Denmark, 1995.

[64] Christopher A Sellick and Richard J Reece. Contribution of amino acid side chains to sugar binding specificity in a galactokinase, Gal1p, and a transcriptional inducer, Gal3p. *The Journal of biological chemistry*, 281:17150–17155, 2006.

[65] K Smallbone, N Malys, H L Messiha, J A Wishart, and E Simeonidis. {B}uilding a kinetic model of trehalose biosynthesis in {S}accharomyces cerevisiae. *Meth. Enzymol.*, 500:355–370, 2011.

[66] Kieran Smallbone, Hanan L Messiha, Kathleen M Carroll, Catherine L Winder, Naglis Malys, Warwick B Dunn, Ettore Murabito, Neil Swainston, Joseph O Dada, Farid Khan, Pinar Pir, Evangelos Simeonidis, Irena Spasić, Jill Wishart, Dieter Weichart, Neil W Hayes, Daniel Jameson, David S Broomhead, Stephen G Oliver, Simon J Gaskell, John E G McCarthy, Norman W Paton, Hans V Westerhoff, Douglas B Kell, and Pedro Mendes. A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes. *FEBS letters*, 587:2832–41, 2013.

[67] Kieran Smallbone and Evangelos Simeonidis. Flux balance analysis: a geometric perspective. *Journal of theoretical biology*, 258:311–315, 2009.

[68] Ralf Steuer. Computational Models of Metabolism: Stability and Regulation in Metabolic Networks. *Advances*, 142:105–251, 2009.

[69] Naama Tepper, Elad Noor, Daniel Amador-Noguez, Hulda S Haraldsdóttir, Ron Milo, Josh Rabinowitz, Wolfram Liebermeister, and Tomer Shlomi. Steady-State Metabolite Concentrations Reflect a Balance between Maximizing Enzyme Efficiency and Minimizing Total Metabolite Load. *PloS one*, 8:e75370, 2013.

[70] Ines Thiele, Ronan M T Fleming, Richard Que, Aarash Bordbar, Dinh Diep, and Bernhard O. Palsson. Multiscale Modeling of Metabolism and Macromolecular Synthesis in E. coli and Its Application to the Evolution of Codon Usage. *PLoS ONE*, 7, 2012.

[71] UNIPROT. Activities at the Universal Protein Resource (UniProt). *Nucleic acids research*, 42:D191–8, 2014.

[72] Kaspar Valgepea, Kaarel Adamberg, Andrus Seiman, and Raivo Vilu. Escherichia coli achieves faster growth by increasing catalytic and translation rates of proteins. *Molecular bioSystems*, 9:2344–58, 2013.

[73] Karen Van Eunen, Jildau Bouwman, Pascale Daran-Lapujade, Jarne Postmus, André B. Canelas, Femke I C Mensonides, Rick Orij, Isil Tuzun, Joost Van Den Brink, Gertien J. Smits, Walter M. Van Gulik, Stanley Brul, Joseph J. Heijnen, Johannes H. De Winde, M. Joost Teixeira De Mattos, Carsten Kettner, Jens Nielsen, Hans V. Westerhoff, and

Barbara M. Bakker. Measuring enzyme activities under standardized in vivo-like conditions for systems biology. *FEBS Journal*, 277:749–760, 2010.

[74] Milan JA van Hoek and Roeland MH Merks. Redox balance is key to explaining full vs. partial switching to low-yield metabolism, 2012.

[75] P van Hoek, J P van Dijken, and J T Pronk. Effect of specific growth rate on fermentative capacity of baker's yeast. *Appl.Environ.Microbiol.*, 64:4226–4233, 1998.

[76] Alexei Vazquez and Zolt??n N. Oltvai. Molecular crowding defines a common origin for the warburg effect in proliferating cells and the lactate threshold in muscle physiology. *PLoS ONE*, 6, 2011.

[77] Tobias von der Haar. A quantitative estimation of the global translational activity in logarithmically growing yeast cells. *BMC systems biology*, 2:87, 2008.

[78] C Waldron and F Lacroute. Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. *Journal of bacteriology*, 122:855–865, 1975.

[79] 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, 2012.

[80] Kai Zhuang, Goutham N Vemuri, and Radhakrishnan Mahadevan. Economics of membrane occupancy and respiro-fermentation. *Molecular systems biology*, 7:500, 2011.

A Model Reconstruction

The stoichiometric model that was used in this project was based on the model "smallYeast, (Central carbon metabolism for yeast)" from the RAVEN toolbox [3]. The smallYeast model is a scaled down version of the iFF708 model developed for use in the RAVEN workshop. The model includes Glycolysis, the Pentose phosphate pathway, the citric acid cycle and a simplified version of oxidative phosphorylation which amounts to 54 reactions (Table 6).

For the purpose of this project some changes were made to the model. The Oxidative phosphorylation was extended (Section A.1), exchange of phosphor was permitted, the biomass equation was separated in to modules (Section A.4) and the amino acid composition for biomass formation was altered (Section A.5).

Table 6: The number of reactions for each pathway included in the SmallYeast and the CCM model. The category other mainly includes reactions for Acetyl-CoA formation.

Pathway	**SmallYeast**	**CCM model**
Galactose Pathway	0	6
Exchange Reactions	8	10
Compartment transport	3	6
Oxidative phosphorylation	2	6
TCA	10	10
Ethanol and Acetate formation	5	5
Pentos phosphate pathway	8	8
Glycolytic Reactions	11	11
Growth	2	12
Other	5	5
Total	54	79

A.1 Inclusion of the Reactions of Oxidative Phosphorylation

In order to give an accurate representation of the speed and size of the oxidative phosphorylation the model was extended to involve the major

reactions in the electron transport chain. The reactions were taken from the iFF708 model and normalized on the limiting substrate, i.e. the one that k_{cat} is given for. To be able to embed the pathway some transport reactions were included and some reactions and metabolites were moved to the mitochondria in agreement with the iFF708 model. This resulted in a model which aggregated with the iFF708 in terms of ATP production (Table 7).

Table 7: Mol ATP generated from 1 mol glucose for the different models. For the purpose of the test a conversion reaction of mitochondrial NADPH to NADH was added to the smallModel to overcome the lack of NADPH sink when only producing ATP.

Model	**Anaerobic**	**Aerobic**
	ATP/glu	ATP/glu
smallModel	2	18.4
Reconstructed smallModel	2	16.5
iFF708	2	16.5

The NADPH reducing enzyme NCP1 which is present in the IFF708 model was not included since it is claimed to be non-mitochondrial [71]. Although it is significantly expressed under normal conditions (46600 copies per cell [71]) its absence will not affect the models performance under normal conditions since a decision can be made between NADH and NADPH production in the mitochondria. This was included in the model by splitting the NADPH forming reaction IDP1 in to the NADH forming IDH and NADPH forming IDPH in agreement with the iFF708 model.

The Succinyl forming LSC1LSC2 reaction was bundled with the FADH2 forming SDH reaction since these reactions take place in one complex and the intermediates are not used by any other reactions in the model.

A.2 Inclusion of the Galactose Pathway

The galctose pathway involves 4 reactions that converts galactose to alpha-D-glucose 6-phosphate. The product enters glycolysis below hexokinase. There exists a cycle involving UDP-glucose 4-epimerase which has to carry flux for the Galactose-1-phosphate uridylyltransferase to operate. The reactions were taken from uniprot and KEGG as the iFF708 model allows UDP-glucose 4-epimerase to be bypassed.

A.3 Extracting Substrate Requirements for Biomass Formation from a Larger Model

The biomass equation in the smallYeast model does not appear to include growth related maintenance and therefore the model overestimates biomass formation (Table 8). A new biomass equation was therefore generated based on the iFF708 model using the metabolites in smallYeast as input. To prevent "unnatural" fluxes to arise as a result of the broad availability of metabolites, the stoichiometric matrix was truncated before simulation. This was done as follows:

1. A simulation was run in the iFF708 model with 1 mol glucose as input, maximizing growth, giving a flux distribution f.
2. The stoichiometric matrix was truncated by removing all reactions with 0 flux and setting all reactions irreversibly in the direction of the flux.
3. All exchange reactions were removed.
4. For each metabolite in the smallYeast model a reaction was added that produced or consumed the metabolite. The metabolites that were consumed were set to the ones consumed in the original smallYeast model. These correspond mainly to metabolites that come in pairs that carry energy (eg. ATP/ADP), reduction power (eg. NADH/NAD) or carbon (AC-COA/COA) and can not by themselves be synthesized by the smallYeast model[7].
5. Additionally reactions producing ammonia and sulfate were added since these are required for biomass formation but are not formed by central carbon metabolism and therefore are not present in the smallYeast model.
6. The amount of ATP consumed was coupled to the amount of ADP produced. This was done to ensure that RNA and DNA formation did not utilize ATP or ADP as a substrate.
7. A simulation was run using this modified matrix with the biomass formation constrained to 1 and the objective set to minimize all fluxes. The objective guaranties that only metabolites that exit central carbon metabolism are depleted. This since formation of any interior

[7] (a) Interestingly biomass formation consumes reduced NADPH whilst producing even more of the reduced NADH, it is however established that this is what yeast does [15].

metabolites would require additional reactions and thus not minimize the fluxes.

8. The fluxes over the added exchange reactions were taken as the growth requirements for biomass formation.

This generated a model ("CCM model") that was similar to the iFF708 model with respect to biomass growth (Table 8) and flux profile (Figure 29). All fluxes were within ±5% between the models apart from 4 fluxes, ACS, CIT, ACO and MDH that were up to 20% lower in the central carbon model.

Table 8: The biomass growth for 1 mol glucose predicted by the reconstructed smallModel and the reconstructed smallModel with modified biomass equation, CCM model.

Model	**Biomass growth**
	Growth per mol glu
Reconstructed smallModel	0.1209
CCM model	0.0913
iFF708	0.0912

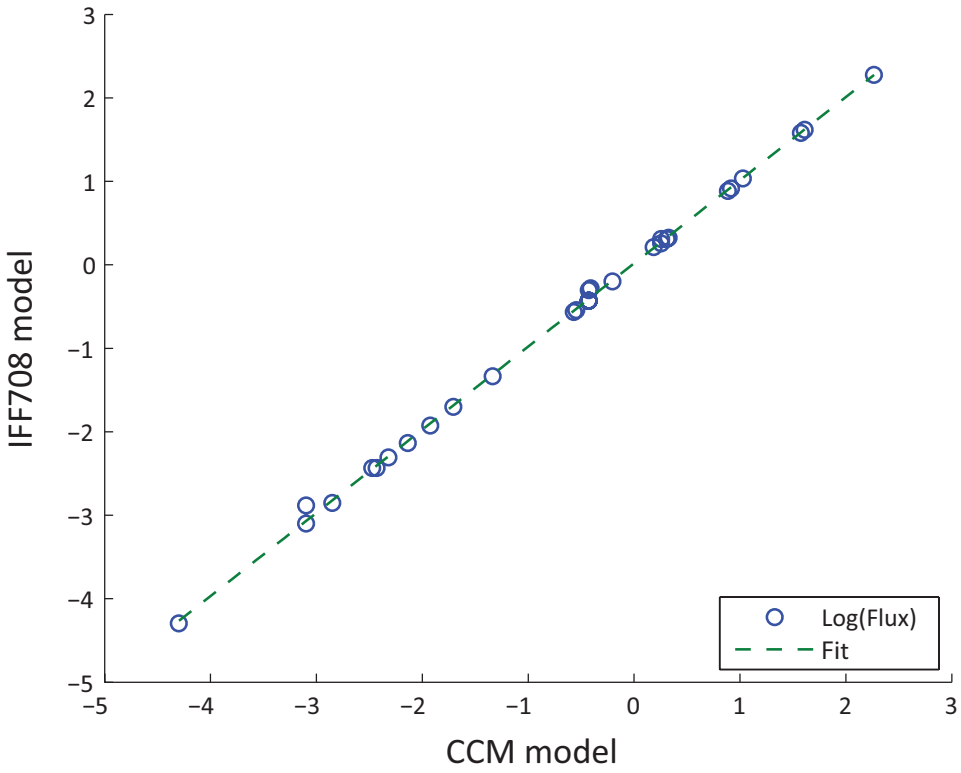

Figure 29: Log log plot of the fluxes over the reactions in the CCM model and the corresponding reactions in iFF708.

A.4 A Modular Biomass Equation

The flux distribution predicted by FBA depends on the composition of biological component (e.g. ratio of proteins, lipids, carbohydrates), since each composition gives its own biomass equation.

A modular version of the biomass equation was generated where each biological component was given its own reaction. This was done by dividing the IFF708 models biomass equation in to components and running a simulation maximizing each in turn (as in section A.3). The total biomass equation was recreated as the superposition of these components. This is not completely accurate since there normally exist synergy effects in formation of the different components e.g. RNA formation creates amino acids as a byproduct that are utilized by the protein formation. The reactions for forming the components are however relatively independent and the resulting fluxes and predicted biomass growth were almost identical (data not shown).

This modular biomass equation allows a systematic evaluation of the out-

come of different biomass compositions. These effects of such changes have been assumed to be small [27]. The effects of the changing biomass composition are however relatively large. If the fixed maintenance costs are ignored there is a 33% higher maintenance cost for the composition associated with low growth rates than the high (Table 9). Including a fixed maintenance cost of 1 ATP per hour and kg dw decreases the difference to at most 12%.

Table 9: Predicted maintenance drain of ATP to ensure a biomass yield of 0.51. Biomass composition taken from literature [63]. Only the main contributions protein and carbohydrates are shown. The fixed ATP consumption is calculated based on the consumption of 1 mol ATP per hour and kg dw [27].

Growth Rate	**Protein**	**Carbohydrates**	**Total**	**Fixed**	**Growth**
h^{-1}	Weight %		Mol ATP per kg dw		
0.1	38	45	40.37	10	30.37
0.2	50	31	34.09	5	29.09
0.3	53	27	35.04	3.33	31.66
0.4	57	23	30.08	2.5	28.3

A.5 Amino Acid Composition

Protein synthesis is an important part of the biomass equation and each amino acid gives a different contribution. Measurements of the amino acid composition i.e. the relative abundance of each amino acid are therefore included in the biomass equation. Through proteomics data the amino acid composition of the cell can be calculated. The composition was calculated using abundance for each protein [79] and protein the protein composition for each protein [71] these were multiplied and the sums of each amino acid was normalized with the sum of all amino acids. This approach does not take in to account systematic differences in the composition between the synthesized proteins and the mature proteins through cleavage, this effect could be expected to be accentuated for methionine since it also is the starting codon and therefore present in the sequence for all proteins.

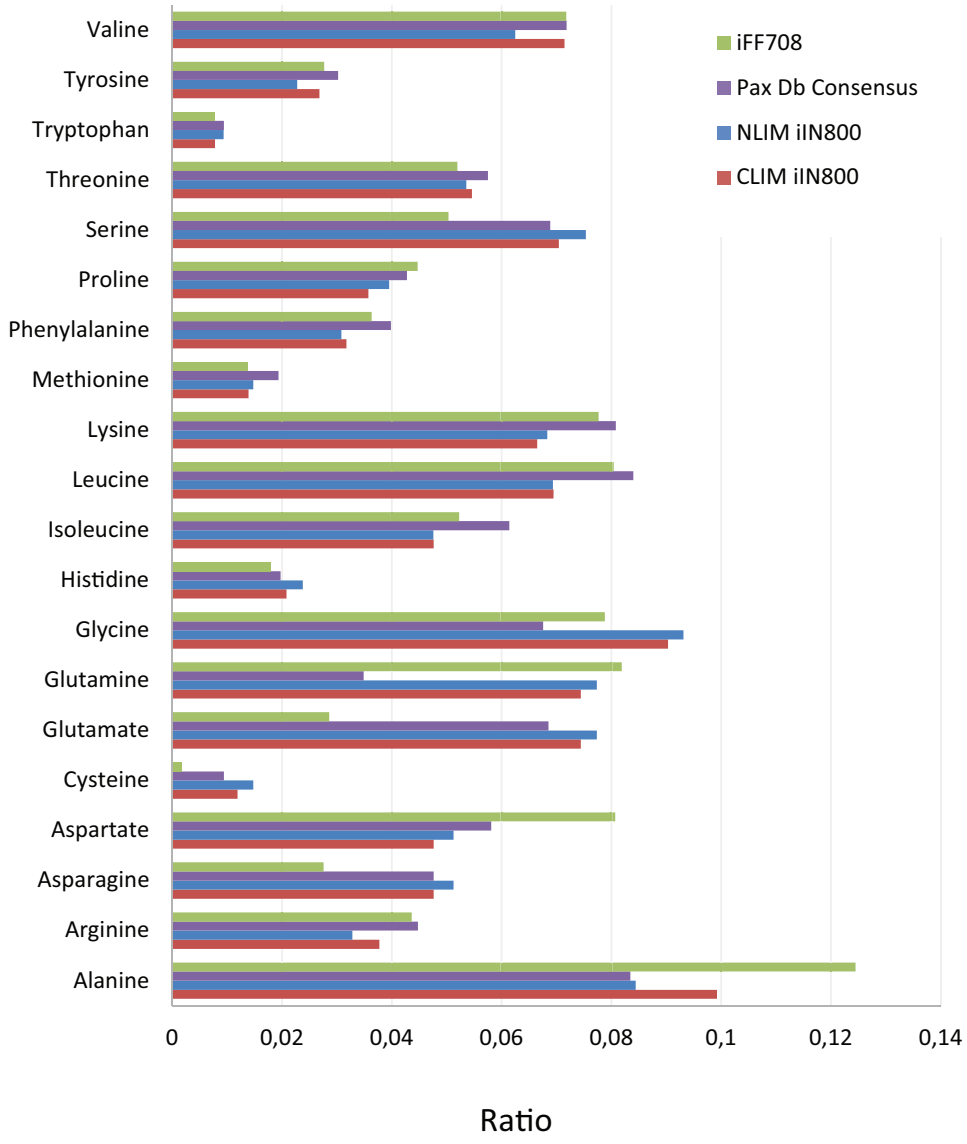

Figure 30: Comparison of the amino acid composition between calculations from proteomics data and data used in the models iFF708 and iNN800. The iFF708 diverges stronger than iNN800 from the calculated data. There is a big difference between the models and the calculated data for the amino acid Glutamine.

The composition used in the iIN800 and iFF708 models was compared with the calculated composition. There were big differences between the models and the calculated data. Most notable this was for the iFF708 model (Figure 30). All models showed a 100% larger estimate for glutamine, a fact that might be of interest for future studies. As expected the calculated ratio of

Methionine was higher (approximately 25%) than in the models.

The effects of the different compositions on mean amino acid weight and predicted biomass growth were relatively small for the tests (Table 10), the estimated mean weight differed by 2% and the weight adjusted yield differed by 0.9%.

Table 10: The effect on weight and biomass yield of different biomass compositions. The average amino acid weight is given for the condensed state i.e. the weight of 1 H_2O is removed from the weights of the amino acids. The yield was calculated by inserting the amino acid composition in the iFF708 model and running a simulation with 1 mol glucose per kg dry weight and hour. The weight adjusted simulation sets the total protein content to 450 g per kg dry weight thereby adjusting the total amount of mol amino acid synthesized by the average weight.

Amino acid composition	**Weight** g/mol	**Yield**	**Weight adjusted**
Nitrogen limited iIN800	108.81	0.0971	0.0925
Carbon limited iIN800	108.54	0.0971	0.0924
iFF708	108.60	0.0967	0.0920
Pax Db Consensus	110.88	0.0953	0.0928
Difference (max/min)	2.1 %	1.9 %	0.9%

B Kcat and the c-Parameter

The c parameter was estimated from experimental data (Section B.1). The values of *kcat* where taken from databases and literature (Section B.2).

B.1 The c-Parameter

A study[73] tests the effect of cell like conditions on the maximum performance of enzymes. The average decrease in performance for the 12 measured enzymes was 20% (Table 11).

Table 11: The reduction of catalytic activity from optimized values given by an in vivo like assay compared with an optimized assay for 2 stains and 2 conditions. For two enzymes (marked with *) the in vivo like assay is more optimal this is discussed in the study [73].

Name	**In vivo like assay**	**Std**
HXK	0.52	0.21
PGI	0.37	0.05
PFK	0.49	0.05
ALD	0.73	0.27
GAPDH	0.44	0.12
PGK	1.23*	0.40
GPM	0.82	0.26
ENO	0.68	0.13
PYK	0.86	0.23
PDC	1.98*	0.81
ADH	0.94	0.62
TPI	0.13	0.05
Total	0.78	0.57

A study[16] uses cell free extracts. and measures the maximum turnover of substrate per g protein in the extract. This is equivalent to measuring what the maximum flux F_{max} could be (Equation 10).

$$k_{cat} \times N \times f = F_{max} \tag{10}$$

Where N is the amount of the given protein in a g and f is a factor reflecting that cell free extracts will give an over representation of cytosolic enzymes. It is taken to be 1.43 for batch cultures and 2 for chemostat cultures reflecting a larger amount of membrane and cell wall bound enzymes in this condition.

The real flux F_r can be estimated using FBA with experimentally measured uptake rates for each condition. Division of these two expressions of F gives an expression for c (Equation 11).

$$c = \frac{F_{max}}{f \times F_r} \tag{11}$$

The value of c was calculated for 12 different enzymes (Table 12). The value of c was 0.43 ± 0.08 for batch conditions. For chemostat conditions the value was 0.08 ± 0.01. approximately 20% of the value in batch conditions.

Table 12: The measured maximum possible flux compared with the matching FBA estimated flux. The data comes from two strains and two different laboratories [16]. For one value (marked with *) the estimated flux is higher than the measured maximum this is discussed in the study [73].

Name	**Batch**	**Std**	**Chemostat**	**Std**
HXK	0.64	0.21	0.08	0.02
PGI	0.17	0.03	0.03	0.01
PFK	NA	-	0.23	0.04
ALD	0.38	0.15	0.09	0.05
GAPDH	0.22	0.08	0.03	0.01
PGK	0.14	0.03	0.03	0.01
GPM	0.15	0.04	0.03	0.01
ENO	0.76	0.14	0.16	0.04
PYK	0.16	0.04	0.05	0.02
PDC	1.53*	0.36	0.22	0.06
ADH	0.61	0.38	0.03	0.01
TPI	0.01	0.01	0.00	-
Total	0.43	0.08	0.08	0.01

B.2 Kcat Values Used in the Study

Brenda [60] is the main database for *kcat* values.

B.2.1 Glycolysis

The k_{cat} values of Glycolysis (Table 13). There is a good order of magnitude agreement between the used values and the human and max values. The value for FBA could not be found in the database or literature. It was therefore estimated from the value of *E coli* a closes homologue, 49% identical and 19% similar [19]. The *E coli* value was divided by 1.4 to reflect the decrease in temperature [8]. The value for PFK was taken from literature [66], as the value in a secondary source was [4] 4 times higher than max reflecting that the activity of the complex was measured.

[8] http://antoine.frostburg.edu/chem/senese/101/kinetics/faq/temperature-and-reaction-rate.shtml

Table 13: The k_{cat} valus for glycolysis used (Used) in this study. together with the highest reported k_{cat} from all species and conditions (Max) and the k_{cat} for human (Human). The value for FBA (marked with a *) was estimated from E.coli. Sources are abbreviated as follows
B =Brenda [60]
B(s) = Brenda specific activity [60]. calculated from specific activity and weight in uniprot.
KRA = k_{cat} from a survey of literature k_{cat} [4].

Name	**EC**	**Used** s^{-1}	**Max** s^{-1}	**Human** s^{-1}	**Source**
HXK	2.7.1.1	276	483.3	101	B(s)
PGI	5.3.1.9	1338.3	2765	3330	KRA
PFK	2.7.1.11	210	185	357	[66]
FBP	3.1.3.11	24	175	23.5	B(s)
FBA	4.1.2.13	14.2*	64.5	59.7	B
TPI	5.3.1.1	16700	68330		B
GLD	1.2.1.12	16.7	70		B
PGK	2.7.2.3	354	2633	2633	B
GPM	5.4.2.11	530	3200		B
ENO	4.2.1.11	230	230	81.68	B
CDC	2.7.1.40	232	3204	1182	B

B.2.2 Pentose Phosphate Pathway

The k_{cat} values of the pentose phosphate pathway (Table 14). There is a good order of magnitude agreement between the used values and max values when such exist. A value for PGL could only be found in a non per reviewed paper [46].

Table 14: The k_{cat} valus for pentos phosphate pathway, see Table 13 for descriptions and abbreviations. The value for PGL (marked with a *) comes from a non per reviewed paper.

Name	**EC**	**Used** s^{-1}	**Max** s^{-1}	**Human** s^{-1}	**Source**
ZWF	1.1.1.49	1443.3	3500	571	KRA
PGL	3.1.1.31	10*			[46]
GND	1.1.1.44	70	98	0.217	KRA
RPI	5.3.1.6	200	97		KRA
RPE	5.1.3.1	1300	7100		B
TKIa	2.2.1.1	56.7		6.3	B
TAL1	2.2.1.2	69.2	13	8.9	KRA
TKIb	2.2.1.1	69		6.3	B

B.2.3 Fermentation Pathways

The k_{cat} values of the Fermentation pathways (Table 15). There is a good order of magnitude agreement between the used values and max values when such exist.

Table 15: The k_{cat} valus for germentation pathways, see Table 13 for descriptions and abbreviations. For one value ADH1 the k_{cat} was only given for the backwards reaction, an external source was therefore used.

Name	**EC**	**Used** s^{-1}	**Max** s^{-1}	**Human** s^{-1}	**Source**
DAR	1.1.1.8	134,2	390		KRA
GPP	3.1.3.21	833	833		B
PDC	4.1.1.1	73,1	486		B
ADH1	1.1.1.1	340	3500	34,8	[29]
ALD6	1.2.1.3	31	19,7	67,5	B(s)

B.2.4 TCA Cycle

The k_{cat} values for the TCA cycle (Table 16). There was limited information available for these enzymes. For two enzymes ACO and FRDS2 the median

value of all data was used, the FRDS2 reaction is however not used in the simulations run in this project. The value of IDH was taken as the value IDPH because of the similarity between the reactions, this value also happens to be the median value. For the reaction MDH1 the k_{cat} values from other organisms appeared to be so high that it would have little effect on the simulation, the value was therefore taken as the maximum value.

Table 16: The k_{cat} valus for TCA cycle, see Table 13 for descriptions and abbreviations. For ACO and FRDS2 the median was used (marked with a *). For IDPH the value of IDH was used (marked with **). For MDH1 the highest value in Brenda was used (marked with ***).

Name	**EC**	**Used** s^{-1}	**Max** s^{-1}	**Human** s^{-1}	**Source**
PYC	6.4.1.1	60	60		B
CIT	2.3.3.1	188	174		B(s)
ACO	4.2.1.3	180*	3,3		Median
IDH	1.1.1.41	178	73	36,8	KRA
IDPH	1.1.1.42	178**	255		Similarity
KGD1KGD2	1.2.4.2	27,9			B(s)
SDH12	1.3.5.1	60	260		B
FRDS2	1.3.1.6	180*	658		Median
FUM1	4.2.1.2	1020	658		B(s)
MDH1	1.1.1.37	4729***	4729		BrendaMax

B.2.5 Oxidative Phosphorylation

The k_{cat} values for the TCA cycle (Table 17). There is a good order of magnitude agreement between the used values and the max values. For the values of ATP synthase a recent study [28] was used. The values in this study are a factor 2 higher than previous studies. This might in part be an effect of the high PH gradient, $\Delta 3.2$ units or 133 mV. Which can be compared with approximately $\Delta 1$ or 43 mV normally found in cellular systems (BIONID 107274[54]). The reported rates of ATP synthase is however only about 15-20% of ATP hydrolysis [58]. This means that the speed of ATP syntheses is $40\,s^{-1}$ compared with the maximum measured $270\,s^{-1}$ for prokaryote ATP synthase.

Table 17: The k_{cat} valus for Oxidative Phosforylation , see Table 13 for descriptions and abbreviations. For ATP1 a recent study was used

Name	**EC**	**Used** s^{-1}	**Max** s^{-1}	**Human** s^{-1}	**Source**
NDI1	1.6.5.3 (1.6.5.9)	500	550		B
NDE2	1.6.5.3 (1.6.5.9)	500	550		B
SDH34	1.3.5.1	60	260		B
RIP1	1.10.2.2	220	459,5		B
COX1	1.9.3.1	1500	2000	80	B
ATP1	3.6.3.14	120	55		[28]

B.2.6 Other

The k_{cat} values for the other pathways (Table 18). For ACS the database value for Saccharomyces appeared to be suspiciously low, 1% of max value and the closest homolog was used *Salmonella enterica*, 47% identical 18% similar.

Table 18: The k_{cat} valus for the other pathways, see Table 13 for descriptions and abbreviations. For MAE1 there was no value and the E coli value was used (marked with *). This reaction is however not used in the simulations run in the project, neither is PCK. ACS is from homolog (marked with **).

Name	**EC**	**Used** s^{-1}	**Max** s^{-1}	**Human** s^{-1}	**Source**
MAE1	1.1.1.38	134,4*	134,4		B
PDH	1.2.4.1	3866	486	69,9	KRA
ACS	6.2.1.1	95.1**	144,9		B
PCK	4.1.1.49	65	65		B

B.2.7 Galactose

The k_{cat} values of galactose (Table 19). There is a good order of magnitude agreement for human for GAL1 and GAL10 . The max values are an order of magnitude larger for GAL1 and GAL10. These values are not the largest values in the database since these are calculated using an incorrect protein weight, probably reflecting a dimer form. The GAL1 value is the

second largest and taken from a recent study [64]. The k_{cat} for Gal10 was recalculated using uniprot protein weight.

Table 19: The k_{cat} valus for the Galactose pathway, see Table 13 for descriptions and abbreviations.

Name	**EC**	**Used** s^{-1}	**Max** s^{-1}	**Human** s^{-1}	**Source**
GAL1	2.7.1.6	22,3	146	8,7	[64]
GAL10	5.1.3.2	40,8	760	36	B(s)
PGM1_2	5.4.2.2	214	398		B(s)
GAL7	2.7.7.12	987	987	98	B

B.3 Cross Verification of Kcat

The k_{cat} values was calculated for 12 different enzymes (Table 20) using maximum flux data [16] and equation 10. The parameter N was calculated from protein abundance studies [79] and cellular composition data [63]. And f was taken as 1.43 (Section B). The calculated k_{cat} was compared with the tabulated k_{cat} values. The average ratio of database and estimated k_{cat} was 1.1 indicating a slight average overestimation of k_{cat} values. According to these calculations FBA and GLD have the most underestimated k_{cat} values and ENO the most overestimated.

Table 20: The measured maximum possible flux compared with the matching FBA estimated flux. The data comes from two strains and two different laboratories [16]. For one value (marked with *) the estimated flux is higher than the measured maximum this is discussed in the study [73].

Name	**EC**	**Database**	**Mean Estimate**	**Std**	**Ratio**
HXK	2.7.1.1	276	296.7	68.6	0.93
PGI	5.3.1.9	1338.3	878.5	425.6	1.52
PFK	2.7.1.11	210	183.6		1.14
FBA	4.1.2.13	14.2	89.8	71.9	0.16
GLD	1.2.1.12	16.7	173.4	73.4	0.10
PGK	2.7.2.3	354	626.3	119.8	0.57
GPM	5.4.2.11	530	1414.6	316.7	0.37
ENO	4.2.1.11	230	53.0	12.1	4.34
CDC	2.7.1.40	232	661.2	86.5	0.35
PDC		73.1	64.2	26.5	1.14
ADH		340	259.8	87.4	1.31
TPI	5.3.1.1	16700	12814.8	14979.7	1.30
Average					1.10

C Experimental Data

C.1 Chemostat Data

Data from chemostat experiments of the cells in and out fluxes was retrieved from literature. (Table 21)

Table 21: Measured [35] input and out fluxes from chemostat experiments as a function of growth rate. D. The fluxes are given in $\mu mol \cdot g^{-1} = q$. The yield refers to the ratio weight of substrate uptake per weight biomass produced. The recovery refers to the measured influx of carbon molecules compared with the measured carbon molecules in the out flux.

D	Yield	O_2	CO_2	Glucose	Ethanol	Acetate	Gglycerol	Recovery
h^{-1}	$g \cdot g^{-1}$	q	q	q	q	q	q	$cmol \cdot cmol$
0.1	0.48	2.5	2.7	1.1	0	0	0	0.96
0.15	0.49	3.9	4.2	1.7	0	0	0	1.024
0.2	0.48	5.3	5.7	2.3	0	0	0	1.009
0.25	0.48	7	7.5	2.8	0	0	0	1.026
0.28	0.46	7.4	8	3.4	0.11	0.08	0	0.97
0.3	0.37	6.1	8.8	4.5	2.3	0.41	0	0.991
0.35	0.23	5.1	14.9	8.6	9.5	0.62	0.05	0.994
0.4	0.2	3.7	18.9	11.1	13.9	0.6	0.15	0.979

Table 22: Measured [55] input and out fluxes from chemostat experiments as a function of growth rate. D. The values for ethanol and acetate are given with their concentration in the chemostat. Acetate production proceeds ethanol production.

D	Yield	O_2	CO_2	Ethanol	Acetate
h^{-1}	$g \cdot g^{-1}$	q	q	mM	mM
0.25	0.43	9.4	9.0	0	0.15
0.30	0.36	12.3	12.2	0	0.48
0.32	0.25	10.2	15.3	40	1.50
0.33	0.23	9.7	16.9	60	1.80

C.2 Composition Data

Measurements of cellular composition data from 3 conditions and two organisms (Table 23, 24, 25).

Table 23: Measured [63] biomass composition of *Saccharomyces Cerevisiae* for increasing growth rates. The values are g per $100g$ dry weight.

Growth rate	0.1	0.2	0.3	0.4
Protein	37.30	48.40	52.20	55.50
Free Amino	1.10	1.90	1.10	1.60
DNA	0.50	0.40	0.50	0.60
RNA	4.90	7.20	7.10	9.00
Lipid	2.60	2.80	2.60	2.50
Glycogen	10.80	4.50	2.20	0.50
Trehalose	2.90	1.80	1.60	1.00
Mannan	16.10	13.60	11.10	9.00
Other carbohydrates	15.60	11.30	11.90	13.30
Ash	5.00	5.00	5.00	5.00

Table 24: Measured [49] biomass composition of *Saccharomyces Cerevisiae*. for increasing growth rates The values are g per $100g$ dry weight.

Growth rate	0.1	0.2	0.3	0.4
Protein	45.0	50.0	55.5	60.1
Free amino acids	1.1	1.3	1.1	2.0
DNA	0.4	0.4	0.5	0.6
RNA	6.2	8.2	10.1	12.1
Lipid	2.9	3.0	3.8	3.4
Glycogen	8.4	4.2	0.6	0.0
Trehalose	0.8	0.2	0.0	0.0
Mannan	13.1	12.9	12.0	13.3
Other carbohydrates	18.4	15.4	12.6	3.7
Ash	5.0	5.0	5.0	5.0

Table 25: Measured [26] biomass composition of *Kluyveromyces marxianus* for increasing growth rates. The values are g per $100g$ dry weight.

Growth rate	0.1	0.25	0.5
Protein	37.0 ± 1.5	52.9 ± 1.0	71.9 ± 2.7
DNA	0.2 ± 0.1	0.5 ± 0	0.6 ± 0.1
RNA	4.9 ± 0.3	7.8 ± 0.3	10.6 ± 0.1
Lipid	5.1 ± 0	5.1 ± 0	5.1 ± 0
Carbohydrate	49.5 ± 1.1	31.3 ± 0.9	9.6 ± 0.8
Ash	2.6 ± 0.1	2.3 ± 0.2	2.6 ± 0.1

C.3 Internal Metabolite Concentration

Internal metabolites (Figure 31).

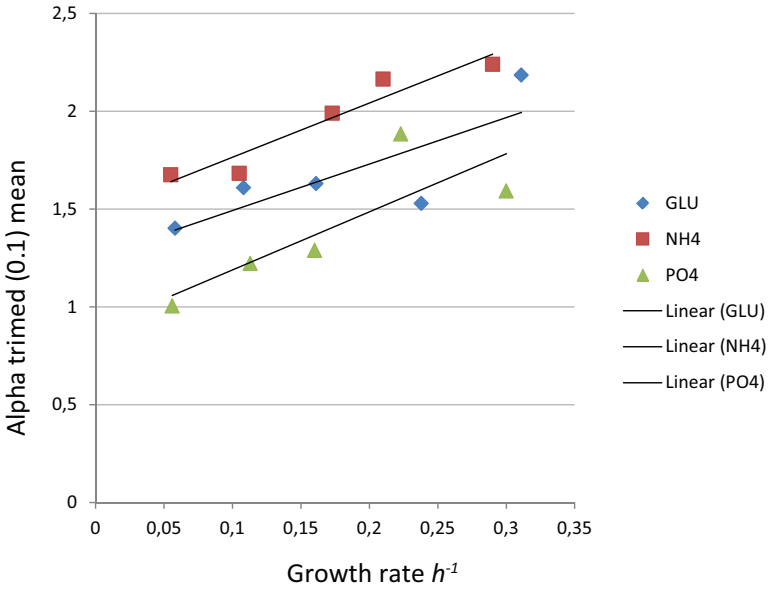

Figure 31: Some experiments [13] show that the internal metabolite concentration increases with growth rate. [13].

D The Number of Subunits in the Complexes

The number of subunits and amino acid residues in the protein complexes (Table 26).

Table 26: Table of the number of subunits and the estimated weight in amino acids, AA, of the complex.

Reaction name	**RXNID**	**Subunits**	**AA**	**Reference**
Alpha-ketoglutarate dehydrogenase	KGD1	3	1976	[71]
Pyruvate dehydrogenase complex	PDA1	4	1572	[71]
Phosphofructokinase	PFK	2	1946	[71]
NADH-coenzyme Q oxidoreductase (Complex I)	NDI1	1	513	[71]
Succinate-Q oxidoreductase (Complex II)	SDH12, SDH34	4	906, 379	[71]
Q-cytochrome c oxidoreductase (Complex III)	RIP1	20	4490	[37]
Cytochrome c oxidase (Complex VI)	COX1	12	1995	[71]
ATP synthase (Complex V)	ATP1	26	5854	[21]

E Functional Annotation of the Proteins

The cellular composition was determined using protein abundance data [79], this data is a weighted mean of several different experiments. A set of relevant categories were chosen and the data was categorized as follows:

1. The 6624 proteins were ordered by abundance and the 5% with the lowest abundance were categorized as low abundant. This corresponded to more than 4600 proteins, out of which more than 500 had no recorded abundance at all. After this coarse filtering less than 2000 proteins remained.
2. The remaining proteins were mapped to the pathway they belong to in the Kyoto Encyclopedia of Genes and Genomes (KEGG)[52]. There normally existed several pathways that covered the chosen categories (Appendix E.2), the pathways were therefore lumped to match the categories. Many proteins were categorized as belonging to several pathways. To avoid double counting, categories were ranked and the gene was taken to belong to the category of highest rank. This ranking created a bias in favor of metabolism and against signaling and DNA replication. The mapping by KEGG pathway corresponded to 950 proteins and covered 75% of the total protein abundance.
3. For the remaining 1000 proteins, text search was performed in the protein description using search terms associated with cellular functions. This led to a further categorization of 400 proteins corresponding to 10% of the total abundance.
4. The remaining 600 proteins were categorized as other and corresponded to 10% of the total abundance.
5. The 2000 high abundant proteins were sorted by name and each category was manually scanned for misannotations. Some 50 proteins were adjusted e.g “37S ribosomal protein MRP1, mitochondrial” was re categorized from “lipid synthesis” to “Protein synthesis”.
6. The abundance of each protein was multiplied by its sequence length giving an abundance weighted by amino acid content.

E.1 From Protein Abundance

The total amino acid abundance for each category was calculated using abundance data and protein sequence length. A comparison of these abundances gave a measure of how large fraction of the total protein biomass that is dedicated to each function. Up to 30% of the proteins were fount

to be involved in making more proteins. Additionally 8% of the proteins were dedicated to making the building blocks of proteins, amino acids. For most growth conditions this is probably an underestimate since the cells used in the abundance study were grown on a media (YPD) that already contains amino acids. According to this categorization 26% of the proteins were involved in central carbon metabolism.

The high abundance in the protein and energy production pathways are not merely a consequence of the functions consisting of many different proteins (Table 27). The median amino acid abundance of the proteins are more than twice as high for theses functions compared with the median of all functions.

Table 27: The median amino acid abundance for the proteins in each category compared with the median abundance of the proteins in all the data (excluding the low abundant data).

Category	Count	Average E+05	Median	Max E+06
Protein synthesis	523	2,2	171	7.3
Amino acid synthesis	98	3,1	284	2.7
Central carbon metabolism	139	6,9	198	11
Non central Metabolism	62	0.6	76	0.3
Nucleotide synthesis	27	3,6	199	4,5
Cell Wall	19	0.7	95	0.5
Fatty acid synthesis	28	2,7	164	2.4
DNA interacting, transcription and replication	133	0.8	76	1.2
Signaling and RNA interaction	179	1,4	92	5.7
Transport	103	1,1	94	1.5
Other	597	0.7	76	1.4
UnCategorized	59	0.4	32	0.3

E.2 Choice of Function Categories

The cellular function of highest interest to this project was central carbon metabolism. This since it will be the constraint used for the proteins involved. However the remaining functions might also be of interest since they give a picture of what the alternative cost of central carbon metabolism

might be i.e. why the cell does not dedicate 1%unit more to central carbon metabolism and 1%unit less to some other function.

Protein synthesis is one of these irreplaceable functions. A decrease in protein production would affect growth negatively, the same holds for amino acid synthesis. Other functions that can not be decreased without affecting growth are the ones involved in biomass formation e.g. nucleotide and fatty acid synthesis and the cell wall. This might also apply to some extent to non central Metabolism.

Signaling and RNA interaction was of particular interest since its effects are explicitly neglected in this project. DNA interacting, and Replication is interesting since it is expected to represent a fixed cost, and transport is expected to scale with the surface. The motivation for adding a uncharacterized category rather than adding these to the other category is to illustrate that there still exists .

E.2.1 Lumping of Kegg Categories

The KEGG pathways were were lumped by cellular function.

```
1 Central carbon metabolism        sce00010 sce00620
    sce00020 sce00190 sce00052 sce00051 sce00500
    sce00030 sce00040
2 Amino acid synthesis     sce00250 sce00260 sce00270
    sce00280 sce00290 sce00300 sce00310 sce00330
    sce00340 sce00350 sce00360 sce00380 sce00400
    sce00410
3 Protein synthesis        sce03050 sce04141 sce03010
    sce03008 sce00970 sce00520 sce00562 sce00563
    sce00564 sce00590 sce00591 sce00592 sce00670
    sce00760 sce00770 sce00860 sce04120 sce00130
4 Fatty acid synthesis     sce00061 sce00062 sce00071
    sce01040 sce00561 sce00565 sce00600 sce04146
5 Transport        sce02010 sce03060 sce04130 sce04144
    sce04145
6 Nucleotide synthesis     sce00230 sce00240
7 DNA interacting, transcription and Replication
    sce04113 sce04111 sce03440 sce03030 sce03450
    sce03430 sce03410 sce03420
8 Signaling/Transcription sce03022 sce03013 sce03015
    sce03018 sce03020 sce04140 sce04011 sce04070
    sce04122 sce03040
9 Non central Metabolism  sce00072 sce00100 sce00430
    sce00450 sce00460 sce00480 sce00680 sce00730
    sce00740 sce00750 sce00510 sce00511 sce00513
    sce00514 sce00630 sce00640 sce00650 sce00910
    sce00920 sce00780 sce00785 sce00790 sce00909
    sce00900
```

Figure 32: The cellular functions and the corresponding KEGG pathways ids they were mapped to. In the case of protein belonging to several pathways the protein was assigned to function where it first was encountered in order from top to bottom.

E.2.2 Text Search Terms Used for Categorization

Search terms were used to map Uniprot protein names to cellular category (Table 28).

Table 28: The search terms used for mapping Uniprot names to cellular functions.

Category	**Search terms**
Protein synthesis	“Chaperone”, “Ribosome”, “tRNA”, “Translation”
Cell Wall	“Wall”
DNA interacting, transcription and Replication	“Transcription”, “Bud”, “Chromatin”, “DNA”, “Histone”
Signaling and RNA interaction	“Receptor”
Transport	“Transport”, “Vesicle”
Uncharacterized	“Uncharacterized”

F Limits on Growth From Size and Protein Synthesis

The limits imposed by CCM could be caused by some other bottleneck that makes increased protein concentrations unnecessary.

F.1 Ribosome Amount Predicted from Growth Rate

There exists a correlation between the amount of RNA and the amount of ribosomes. The amount ribosomes can therefore be estimated from cellular composition data (Tabel 29). From this the percentage of the protein mass that is ribosomal can be calculated to between 8.5-13%, which gives a total protein synthesis mass of 21-33%, which fits well with abundance data (30%). For prokaryotes ribosomes concentration grows linearly with growth rate [54].

Table 29: Estimates of ribosome content from measurements of cellular RNA content at different growth rates for nitrogen limited (a) and carbon limited (b) *S. cerevisiae*.

a)

S. cerevisiae, nitrogen limited	**Growth Rate**			
	0.1	**0.2**	**0.3**	**0.4**
RNA [kg]	0.049	0.072	0.071	0.090
Protein [kg]	0.373	0.484	0.522	0.555
Estimated amount of ribosomes [10^{-5}mol]	2.24	3.29	3.25	4.12
Percent of total protein in Ribosomes [%]	8.44	9.55	8.73	10.41
Estimated percent of protein in protein synthesis [%]	21.47	24.32	22.23	26.51

b)

S. cerevisiae, carbon limited	**Growth Rate**			
	0.1	**0.2**	**0.3**	**0.4**
RNA [kg]	0.063	0.082	0.101	0.121
Protein [kg]	0.45	0.5	0.555	0.601
Estimated amount of ribosomes [10^{-5}mol]	2.88	3.75	4.62	5.53
Percent of total protein in Ribosomes [%]	8.99	10.53	11.69	12.93
Estimated percent of protein in protein synthesis [%]	22.89	26.81	29.75	32.91

There exists a correlation between the amount of ribosomes and the growth rate (Eq 12).

$$\mu \times A = R \times v_{rib} \tag{12}$$

Where μ is the specific growth rate, A is the total amount of amino acids in a kg, R is the amount of ribosomes and v_{rib} is the ribosomal speed at that growth rate. Or simply as growth rate increases more ribosomes are needed.

The maximal translation speed of a ribosome is suggested to be 10 amino acids per second and ribosome for *S. cerevisiae* [77]. For *E. coli* it is assumed to be 20 amino acids a second [54, 77]. It can however be shown that the ribosomal translation speed given in literature is too low to sustain growth at high growth rates (Figure 33). Adjusting the translation speed for growth to be sustainable corresponds to a 38% increase in E. coli and between 24-54% increase in S. cerevisiae. These figures are calculated assuming that no protein degradation takes place, which is probable to be incorrect.

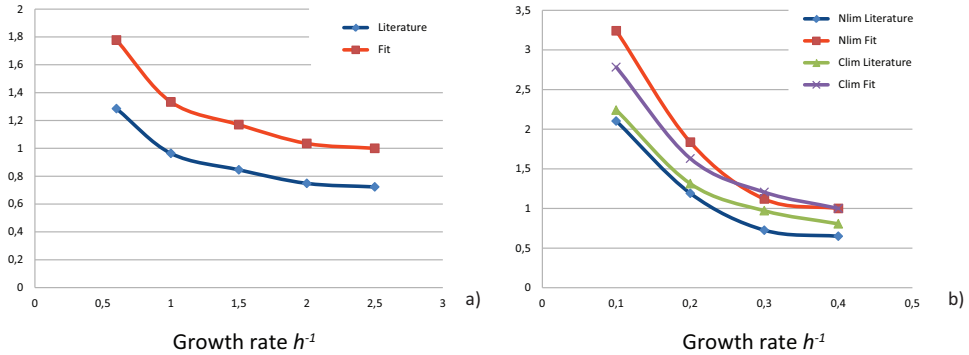

Figure 33: Theoretical maximal protein production divided by the production required to sustain growth for *E. coli* a) Nitrogen limited S. cerevisiae and carbon limited S. cerevisiae. b). The theoretical maximum was calculated using parameters for ribosome speed from literature and with a value fitted to allow growth (Table 29). The literature values where 20 AA/s for *E. coli* and 10 AA/s for *S. cerevisiae*. The fitted values where 27.66 AA/ for *E. coli*, 15.41 AA/s for nitrogen limited and 12.41 AA/s for carbon limited *S. cerevisiae*.

From the equation for ribosome amount (Eq 12) the minimum amount of ribosomes required can be calculated if the growth rate μ, the maximum translation speed v_{rib} and the total mass of amino acids and RNA T is known (Eq 13).

$$A = \frac{T \times v_{rib}}{\mu w_r + v_{rib} w_a} \qquad (13)$$
$$R = \frac{T \times \mu}{\mu w_r + v_{rib} w_a}$$

This makes use of the conservation of mass (Eq 14).

$$R \times w_r = T - A \times w_A \tag{14}$$

Where $w_A$is the mol weight of amino acids, $0.108\,kg\,mol^{-1}$. And w_r is the RNA weight for a mol of ribosomes $2.19 \times 10^3\,kg\,mol^{-1}$. The w_r involves the rRNA/RNA ratio R_{ratio} the number of nucleotides per ribosome R_n and the mol weight of nucleotides w_n (Eq 15).

$$w_r = \frac{R_n \times w_n}{R_{ratio}} = \frac{5500 \times 0.324}{0.815} = 2.19 \times 10^3\,kg\,mol^{-1} \tag{15}$$

This allows predictions of RNA content at different growth rates (Figure 34).

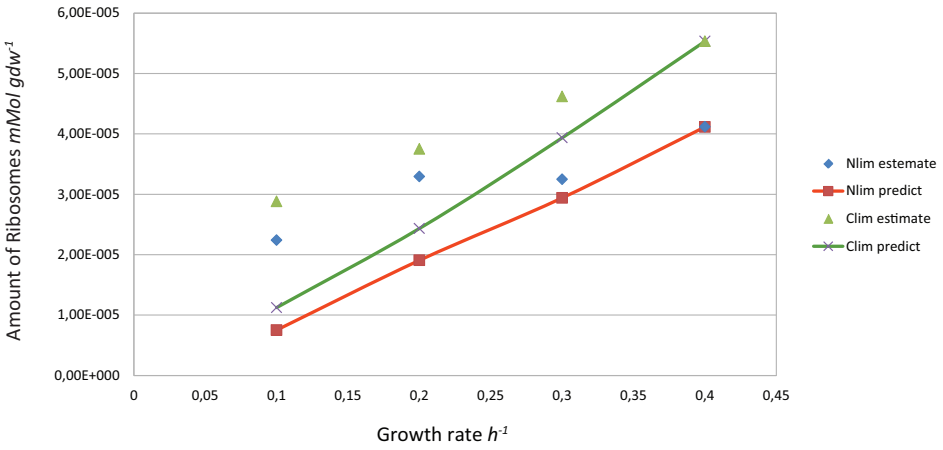

Figure 34: Estimated ribosome amount based on RNA content and predicted ribosome amount based on RNA+Protein content and eq 15. There is an underestimate of the amount of ribosomes for low growth rates.

F.2 Tables Of Cell Sizes Dependency on Growth Rate

Volume as a function of growth rate from glucose and nitrogen limited chemostat experiments (Tabel 30, 31).

Table 30: Volume as a function of growth rate from a glucose limited chemostat experiment [13]. Radius. area and surface/volume ratio are calculated assuming yeast is a sphere.

Growth rate	**Volume**	**Stdev**	**Surface / volume**
h^{-1}	μm^3	μm^3	$\frac{\mu m^2}{\mu m^3}$
0.108	27.0	0.03	1.61
0.161	24.4	0.06	1.67
0.238	25.3	0.05	1.65
0.311	29.2	0.07	1.57

Table 31: Volume as a function of growth rate from a nitrogen limited chemostat experiment [13]. Radius. area and surface/volume ratio are calculated assuming yeast is a sphere.

Growth rate	**Volume**	**Stdev**	**Surface / volume**
h^{-1}	μm^3	μm^3	$\frac{\mu m^2}{\mu m^3}$
0.100	22.9	0.3	1.70
0.172	28.3	0.3	1.59
0.208	33.6	0.3	1.50
0.295	37.3	1.3	1.45

F.3 Literature Data on Cell Wall for Biomass Estimation

The mass of cell wall associated carbohydrates is expected to decrease with the cell size since the surface volume ratio decreases. A simple model describing this was created based on literature values (Table 32) of cell wall thickness, water content and the density of carbohydrates.

The carbohydrate content was calculated as follows:

1. The volume of the cell at a given growth rate was estimated (equation 8).
2. The dry weight of the cell was calculated from the total volume.
3. The total volume of the cell wall was calculated from the wall thickness and assuming sphere or cube shape of the cell.
4. The weight of carbohydrates in the cell wall was calculated from the volume and divided by the dry weight.

Table 32: Cell wall data.

Parameter	**Value**	**Source**
Density carbohydrate $\frac{g}{cm^3}$	≈ 1.45 (1.35-1.55)	[59]
Water content wt/wt	68%	[59]
cm^3 wall / g carbohydrate	$\frac{1}{1.45} + \frac{0.68}{0.32} = \frac{1}{0,36}$	
Cell wall thickness μm	0.11	[40]
Carbohydrate content (dry weight)	94.5%	[40]
Density cell $\frac{g}{cm^3}$	1.11	[40]
Dry weight of wet weight	34%	[40]

G Original Project Description

The project description as of 2014-01-21.

MASTER THESIS PROJECT DESCRIPTION

TITLE

Including the cost of enzyme production in a flux balance analysis of Saccharomyces cerevisiae.

BACKGROUND

Flux balance analysis (FBA)[1] is a way to model a metabolic network. The model describes the concentration of the metabolites using the stoichiometry of the reactions. Solutions are found assuming steady state with regards to metabolite concentration. The solution that optimizes some biologically relevant objective function (e.g. biomass production) is assumed to be a good description of the system. Recently FBA has been extended with the so called ME models[2]. Among other things a ME model includes the concentration of the enzymes responsible for the biochemical reactions. This has highlighted that traditional FBAs do not account for the cost of using a reaction and therefor have a tendency to utilize many reactions.

PURPOSE

The purpose is to test to which extent neglecting the cost of enzymes has affected the predictive powers of FBA and if predictions can be improved if the cost is included.

GOAL

The goal is to develop a modified version of FBA that incorporates cost of enzyme production and that allows comparison with normal FBA. The goal is also to investigate the predictive powers of this model.

LIMITATIONS

The metabolic network investigated will be limited to the central carbon metabolism of Saccharomyces cerevisiae.

---

[1] http://en.wikipedia.org/wiki/Flux_balance_analysis

[2] http://www.nature.com/ncomms/journal/v3/n7/full/ncomms1928.html