

---

**Understanding metabolic regulation  
and cellular resource allocation through  
optimization**

---

**Dissertation**

zur Erlangung des Grades

**Doctor rerum naturalium**

Fachbereich Mathematik und Informatik

Freie Universität Berlin

Alexandra-Mirela REIMERS

Berlin, 2017

Betreuer: Prof. Dr. Alexander BOCKMAYR  
Fachbereich Mathematik und Informatik  
Freie Universität Berlin

Zweitgutachterin: Prof. Dr. Dr. h.c. Edda KLIPP  
Institut für Biologie  
Humboldt Universität zu Berlin

Tag der Disputation: 12. Oktober 2017

# Contents

<b>Abstract</b>	<b>9</b>
<b>1 Introduction</b>	<b>11</b>
1.1 Metabolic networks and notation . . . . .	12
1.2 Metabolic modeling . . . . .	14
1.2.1 Dynamic modeling . . . . .	14
1.2.2 Constraint-based modeling . . . . .	15
1.3 Cellular resource allocation . . . . .	19
1.4 Structure of this thesis . . . . .	20
<b>2 Resource allocation formalisms</b>	<b>23</b>
2.1 Resource allocation principles in a self-replicator . . . . .	23
2.1.1 Growth rate and dilution . . . . .	24
2.1.2 Mass balance equations . . . . .	25
2.1.3 Kinetics of rate equations . . . . .	26
2.1.4 Membrane integrity . . . . .	26
2.1.5 Total proteome is limited . . . . .	27
2.1.6 Additional constraints . . . . .	27
2.1.7 The nonlinear optimization problem . . . . .	28
2.1.8 Results and extensions . . . . .	28
2.2 Resource balance analysis . . . . .	29
2.2.1 Relaxing steady-state constraints and imposing quota . . . . .	29
2.2.2 Kinetics of rate equations are replaced by linear constraints . . . . .	29
2.2.3 Modeling volume vs. modeling density . . . . .	30

## Contents

---

2.2.4	Nutrient uptake . . . . .	31
2.2.5	The quadratically constrained optimization problem . . . . .	31
2.2.6	Mathematical and biological implications . . . . .	32
2.2.7	Extension through metabolism and gene expression models	32
2.3	Dynamic enzyme-cost flux balance analysis . . . . .	36
2.3.1	Mass balance equations and quasi-steady-state of metabolism . . . . .	36
2.3.2	Enzyme capacity and nonnegativity constraints . . . . .	37
2.3.3	Imposing quota production . . . . .	38
2.3.4	Objective functions . . . . .	38
2.3.5	The dynamic optimization problem, solving strategies and main results . . . . .	39
2.4	Conditional flux balance analysis . . . . .	40
2.5	Conclusions . . . . .	41
<b>3</b>	<b>Steady states and upper bounds on growth rates from resource allocation models</b>	<b>43</b>
3.1	Introduction . . . . .	43
3.1.1	Classical derivation based on the quasi-steady-state assumption . . . . .	44
3.1.2	The perspective based on long time periods . . . . .	45
3.2	The steady-state assumption for long time periods . . . . .	47
3.2.1	Modeling assumptions . . . . .	47
3.2.2	Average fluxes . . . . .	48
3.2.3	Violation of the steady-state condition for finite time T . . . . .	49
3.2.4	Escherichia coli . . . . .	49
3.2.5	Saccharomyces cerevisiae . . . . .	49
3.2.6	Homo sapiens (HeLa cells) . . . . .	50
3.3	Applications to yield optimization . . . . .	50
3.4	Kinetic constraints . . . . .	51
3.4.1	Average concentrations can be inconsistent with average fluxes . . . . .	52
3.4.2	Linear kinetic constraints remain consistent . . . . .	53

3.5	Conclusions . . . . .	54
3.5.1	Average fluxes satisfy the steady-state assumption . . . . .	54
3.5.2	Pitfalls of averaging . . . . .	55
<b>4</b>	<b>Building and encoding a metabolic resource allocation model</b>	<b>57</b>
4.1	Model prerequisites . . . . .	58
4.2	Building the protein production reactions . . . . .	60
4.2.1	The case of enzymes encoded by one gene only . . . . .	60
4.2.2	The case of isoenzymes . . . . .	61
4.2.3	The case of enzyme complexes encoded by several genes . . . . .	63
4.2.4	The ribosome . . . . .	63
4.2.5	Compartmentalization . . . . .	64
4.3	Setting up quota compounds and storage . . . . .	64
4.3.1	Initial quota compound amounts . . . . .	65
4.3.2	The case of noncatalytic proteins . . . . .	67
4.3.3	Storage . . . . .	69
4.4	Assigning reaction turnover rates . . . . .	69
4.5	Validating the model using experimental data . . . . .	70
4.6	Units . . . . .	72
4.6.1	Metabolic reactions . . . . .	72
4.6.2	Reactions producing quota compounds . . . . .	72
4.6.3	Reactions producing enzymes . . . . .	73
4.7	Debugging the model . . . . .	74
4.8	The SBML representation of a metabolic resource allocation model	75
4.8.1	Compartments . . . . .	75
4.8.2	Species . . . . .	75
4.8.3	Gene products . . . . .	78
4.8.4	Reactions . . . . .	78
4.9	Conclusions . . . . .	80

## Contents

---

<b>5 Metabolic resource allocation in practice: numerical concerns and solving strategies</b>	<b>81</b>
5.1 Artefacts of model formulations . . . . .	81
5.1.1 deFBA dependency on the discount factor $\varphi$ in the objective	81
5.1.2 Linear versus exponential growth in deFBA . . . . .	82
5.1.3 deFBA dependency on choice of simulation endpoint . . . . .	84
5.2 Numerical considerations and scaling . . . . .	86
5.2.1 The midpoint rule discretization . . . . .	86
5.2.2 Cyclicity and discretization problems . . . . .	87
5.2.3 cFBA with dilution term . . . . .	89
5.3 Connections between cFBA and RBA . . . . .	91
5.3.1 Modeling assumptions (with dilution) . . . . .	92
5.3.2 Average fluxes and average concentrations . . . . .	93
5.3.3 Violation of the steady-state condition by dilution . . . . .	94
5.3.4 Bounds for cFBA from RBA . . . . .	95
5.4 Condition numbers, time scale separation and scaling . . . . .	96
5.4.1 Recap on matrix condition numbers . . . . .	97
5.4.2 Condition numbers for typical deFBA and cFBA problems . . . . .	98
5.5 Convexity of the binary search in cFBA . . . . .	99
5.6 Implementation . . . . .	100
5.6.1 Variables . . . . .	100
5.6.2 Constraint matrix . . . . .	101
5.6.3 Bounds on variables . . . . .	102
5.6.4 Shift experiments . . . . .	102
5.6.5 Short term deFBA . . . . .	102
<b>6 Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth</b>	<b>103</b>
6.1 Introduction . . . . .	104
6.2 Model building . . . . .	106
6.2.1 The genome-scale reconstruction . . . . .	106
6.2.2 The macromolecules of autocatalytic growth . . . . .	106

6.3	Model constraints and objective . . . . .	108
6.3.1	Constraints . . . . .	109
6.3.2	The optimization objective . . . . .	114
6.3.3	The quadratic program and binary search . . . . .	115
6.3.4	Further model specifications . . . . .	116
6.4	Results . . . . .	117
6.4.1	Growth under constant light . . . . .	117
6.4.2	Adaptations to different light intensities . . . . .	118
6.4.3	Sensitivity analysis . . . . .	119
6.4.4	A day in the life of <i>Synechococcus elongatus</i> PCC 7942 . . . . .	123
6.4.5	Metabolite partitioning during diurnal growth . . . . .	127
6.4.6	The dynamics of glycogen accumulation . . . . .	129
6.4.7	Glycogen accumulation for variable day lengths . . . . .	130
6.5	Discussion . . . . .	131
<b>7</b>	<b>Optimal resource allocation in yeast under dynamically changing environments</b>	<b>135</b>
7.1	Model building . . . . .	136
7.1.1	Choice of reconstruction and model reduction . . . . .	136
7.1.2	The macromolecules of autocatalytic growth . . . . .	139
7.2	Model constraints and objective . . . . .	140
7.2.1	Constraints . . . . .	140
7.2.2	Objective . . . . .	142
7.2.3	deFBA vs. short term deFBA . . . . .	143
7.2.4	Sudden shifts . . . . .	143
7.3	Results . . . . .	143
7.3.1	Growth in constant environmental conditions . . . . .	143
7.3.2	Overflow metabolism is an optimal behavior from a resource allocation perspective . . . . .	144
7.3.3	Diauxie is an optimal behavior from a resource allocation perspective . . . . .	146
7.3.4	Adaptations of wildtype yeast to environmental shifts . . . . .	147
7.3.5	Adaptations of yeast mutants to oxygen availability shifts . . . . .	155
7.4	Discussion . . . . .	156

## Contents

---

<b>8 Perspectives: inferring gene regulation from dynamic resource allocation models</b>	<b>159</b>
8.1 Formalizing the question . . . . .	160
8.1.1 Toy model for inference . . . . .	160
8.1.2 The general case . . . . .	161
8.2 The simplest case: a linear model . . . . .	163
8.3 Evaluating the learned regulation . . . . .	165
8.4 Discussion . . . . .	170
<b>9 Conclusions</b>	<b>173</b>
<b>Appendix</b>	<b>177</b>
A The steady-state assumption . . . . .	177
B The SBML representation of resource allocation models: illustration using a toy example . . . . .	179
C <i>Synechococcus elongatus</i> 7942 model . . . . .	192
D <i>Saccharomyces cerevisiae</i> model . . . . .	198
E Regulation inference using deFBA time courses . . . . .	207
<b>Bibliography</b>	<b>209</b>
<b>Notation index</b>	<b>231</b>
<b>Glossary</b>	<b>233</b>
<b>Index</b>	<b>237</b>
<b>Zusammenfassung</b>	<b>241</b>
<b>Curriculum vitae</b>	<b>243</b>
<b>Selbstständigkeitserklärung</b>	<b>245</b>
<b>Acknowledgements</b>	<b>247</b>



# Abstract

This thesis is a contribution to the field of systems biology, where mathematical and computational models are used to study large biological networks such as the metabolism or the signaling pathways of living organisms. These models are simplified representations of the studied biological systems and come in different granularities and abstraction levels, depending on the size of the networks and on the modeling formalisms.

One of the largest networks studied within systems biology is the metabolism, which comprises all the biochemical reactions happening inside a cell. Until recently, such large metabolic networks have been studied mainly in isolation and under stationary conditions, without considering the environment dynamics or the enzymatic resources needed to catalyze all the biochemical reactions. This has been mainly done using constraint-based analysis and optimization. While proven to be very successful in predicting cellular behavior in some cases, this approach is not suited for microorganisms living under changing environments. Two examples are cyanobacteria, whose metabolism is adapted to the daily changes in the sunlight availability, and yeasts living in large bioreactors and thus moving in an environment governed by local heterogeneities.

This thesis builds on top of recent developments in dynamic resource allocation formalisms for metabolism, which use tools from dynamic optimization and optimal control. We focus on modeling and understanding resource allocation in large (sometimes genome-scale) metabolic models.

After giving an overview of existing tools for the study of metabolic resource allocation, the thesis presents a new mathematical derivation of the widely used steady-state assumption for metabolic networks and shows how this can be used to provide upper bounds on dynamic resource allocation solutions. In preparation for the case studies, we present a guide for generating a dynamic resource allocation model using information from online databases, as well as guidelines and useful problem transformations. All the theory developed so far is then applied in two case studies. One of them investigates the cyanobacterium *Synechococcus elongatus* PCC 7942. This is the first genome-scale dynamic resource allocation study. It gives insight into the temporal organization of enzyme synthesis processes following light availability and shows that the linear pattern of glycogen accumulation throughout the day period is an optimal behavior that arises as a tradeoff between several conflicting resource allocation objectives. The second case study concerns the yeast *Saccharomyces cerevisiae*. We aim to understand what mechanisms enable some of the cells to survive environmental transitions. We show that overflow metabolism and diauxie, which are phenomena widely spread in nature, are optimal behaviors from a resource allocation perspective. Moreover, we investigate how one can use resource allocation models to understand how yeast adapts to oxygen and nutrient availability shifts. We end with a perspectives chapter which provides some preliminary results for using time courses from dynamic resource allocation models to infer the regulatory structures that implement these optimal behaviors.



# Chapter 1

## Introduction

“Most of an organism, most of the time, is developing from one pattern into another, rather than from homogeneity into a pattern. One would like to be able to follow this more general process mathematically also. The difficulties are, however, such that one cannot hope to have any very embracing *theory* of such processes, beyond the statement of the equations. It might be possible, however, to treat a few particular cases in detail with the aid of a digital computer. This method has the advantage that it is not so necessary to make simplifying assumptions as it is when doing a more theoretical type of analysis.” (Alan Mathison Turing, *The chemical basis of morphogenesis*, 1952)

Sixty-five years after the publication of *The chemical basis of morphogenesis* by Alan Turing, we are sitting on top of a plethora of computational techniques that we developed to help us understand biology. And one cannot help noticing how right Turing was. We still do not have a unifying theory for understanding biology and life, not even for the simplest bacteria. However, computer models of biological systems indeed helped and still help us understand better how living organisms grow and evolve.

Recent techniques in genomics, such as whole-genome sequencing (Ng and Kirkness, 2010), whole transcriptome shotgun sequencing (Morin et al., 2008) or mass spectrometry-based proteomics (Aebersold and Mann, 2003) have provided detailed information about the structure of several organisms. However, having the complete genome sequence of an organism is not the end of the story, but rather the beginning of a whole series of analyses starting with the identification of genes, regulatory and signaling structures, and hopefully ending with an understanding of how everything encoded in the genome comes together into what can be observed by the naked eye and by experiments. Microarrays provide us with data about genes that are expressed by the organism in different growth and/or disease conditions. But they do not tell us how the activity

## Chapter 1. Introduction

---

of those genes determines the phenotype. Similarly, mass spectrometry-based proteomics gives us information about what proteins are present and in which concentrations, but does not tell us anything about their function. Therefore, there is a growing need for understanding the integrated behavior of cellular subsystems such as the metabolism and the gene regulatory structure. This is how systems biology emerged. Scientists in this field try to understand design principles of biology by looking at organisms as a whole, rather than analyzing individual subsystems (Kitano, 2002).

A wide part of systems biology is concentrated on understanding growth phenotypes via the study of genome-scale metabolic networks. As whole-genome sequencing advanced, genome-scale reconstructions (Francke et al., 2005; Notebaart et al., 2006) of the metabolism of many organisms became available (Förster et al., 2003a; Feist et al., 2007; Thiele et al., 2013). As a matter of fact, more than 2600 functional draft reconstructions have been generated up to now (Büchel et al., 2013) and many of them can be retrieved from online databases such as BioModels (Le Novère et al., 2006; Li et al., 2010; Chelliah et al., 2015; Juty et al., 2015). These reconstructions contain information about all enzymatic and spontaneous reactions that can happen in the organism's cells. Since metabolic genes can often be mapped to the enzymes they encode and enzymes to the reactions they catalyze, one can easily obtain information about the impact of a gene knock-out only by performing simulations on a metabolic network in which the corresponding reaction is deleted (Deutscher et al., 2006). Moreover, metabolic network study can help in biotechnology and strain design by predicting knock-out strategies that couple growth of an organism to production of biotechnologically useful by-products (Hädicke and Klamt, 2011; Burgard et al., 2003). Genome-scale metabolic networks also help in the search for cancer drug targets by identifying pairs of synthetic lethal genes (Jerby and Ruppín, 2012). And these are only some of the application areas of metabolic modeling.

But before we go on with studying metabolism, let us first see what a metabolic model is.

### 1.1 Metabolic networks and notation

The main structures through which we will discuss metabolism in this thesis are (genome-scale) metabolic networks. They are directed hypergraphs that describe the biochemical reactions happening inside a cell. Figure 1.1 shows a toy example of such a network and establishes nomenclature for its components.

## 1.1 Metabolic networks and notation

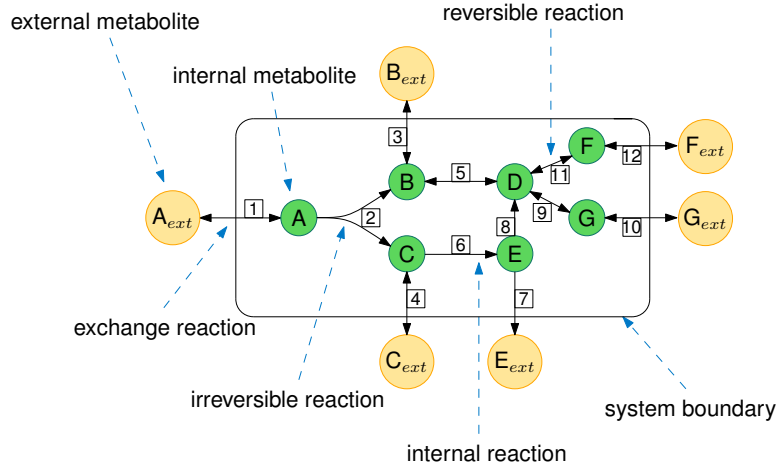


Figure 1.1: **Example metabolic network.** Assume that  $A_{ext}$ ,  $B_{ext}$ ,  $C_{ext}$ ,  $E_{ext}$ ,  $F_{ext}$ , and  $G_{ext}$  are metabolites present outside the cell and  $A - G$  are metabolites present in the cell. The figure shows the metabolic network representation of the reaction system:  $A_{ext} \leftrightarrow A$ ,  $A \rightarrow B + C$ ,  $B_{ext} \leftrightarrow B$ ,  $C \leftrightarrow C_{ext}$ ,  $B \leftrightarrow D$ ,  $C \rightarrow E$ ,  $E \rightarrow E_{ext}$ ,  $E \rightarrow D$ ,  $D \leftrightarrow G$ ,  $G \leftrightarrow G_{ext}$ ,  $D \leftrightarrow F$ ,  $F \leftrightarrow F_{ext}$ .

Because of its similarity with graphs, a metabolic network can be represented mathematically via a stoichiometric matrix  $S$ , which is nothing else than an incidence matrix that allows for encoding the hyperedges as well as the stoichiometry of metabolites in reactions. The rows of  $S$  correspond to the internal metabolites and the columns to the internal and exchange reactions. For instance, the stoichiometric matrix of the network depicted in figure 1.1 is:

$$S = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{matrix} \\ \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \\ G \end{matrix} & \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{pmatrix} \end{matrix}.$$

We call the tuple  $(\mathcal{M}, \mathcal{R}, S)$  a metabolic network, where  $\mathcal{R} := \mathcal{R}^M \cup \mathcal{R}^T$  is the set of internal reactions (denoted by  $\mathcal{R}^M$ ) and exchange reactions (denoted by  $\mathcal{R}^T$ ) (1 – 12 in the example in figure 1.1),  $\mathcal{M}$  is the set of internal metabolites (A-G in the example in figure 1.1), and  $S$  is the stoichiometric matrix. We note that, in classical metabolic network analysis, external metabolites are typically not modeled, but they will become important later on in this thesis.

Often, information about the reversibility of reactions and lower and upper bounds on reaction fluxes is available and can thus be incorporated into the

## Chapter 1. Introduction

---

mathematical description. We denote by  $Irr \subseteq \mathcal{R}$  the set of irreversible reactions, by  $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathcal{R}}$  the lower and upper reaction rate (flux) bounds.

We use  $S_{*r}$  to denote the column corresponding to reaction  $r$  and  $S_{m*}$  to denote the row corresponding to metabolite  $m$  respectively.  $S_{mr}$  then denotes the stoichiometric coefficient of metabolite  $m$  in reaction  $r$ . We will use sets of indices to denote submatrices:  $S_{*Irr}$  for example will denote the submatrix of  $S$  corresponding to all irreversible reactions  $Irr$ .

Of particular importance for the metabolic modeling is also the set of enzymes that catalyze the reactions, which we denote as  $\mathcal{E}$ . We use  $\mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_{|\mathcal{R}|})^\top$  to denote the fluxes (the rates at which substrates are converted into products, also known as reaction rates),  $\mathbf{c} = (\mathbf{c}_1, \dots, \mathbf{c}_{|\mathcal{M}|})^\top$  to denote metabolite concentrations,  $\mathbf{e} = (\mathbf{e}_1, \dots, \mathbf{e}_{|\mathcal{E}|})^\top$  to denote the set of enzyme concentrations, where  $\mathbf{x}^\top$  denotes the transpose of the vector  $\mathbf{x}$ . We use  $\mu$  to denote the growth rate of the organism. Since some of the methods we will use involve molar amounts rather than concentrations, we use  $\mathbf{n}_i$  to denote the molar amount of compound  $i$ , irrespective of whether it is a metabolite or an enzyme.

Please note that the fluxes  $\mathbf{v}$ , the concentrations  $\mathbf{c}, \mathbf{e}$ , the molar amounts  $\mathbf{n}$ , and the growth rate  $\mu$  can in some cases depend on time. Furthermore,  $\mathbf{v}(t) \in \mathbb{R}^{|\mathcal{R}|}$ ,  $\mathbf{c}(t) \in \mathbb{R}_{\geq 0}^{|\mathcal{M}|}$ ,  $\mathbf{e}(t) \in \mathbb{R}_{\geq 0}^{|\mathcal{E}|}$ , and  $\mathbf{n} \in \mathbb{R}_{\geq 0}^{|\mathcal{M} \cup \mathcal{E}|}$  are vectors and thus written in boldface, as all other vectors appearing in this thesis. Equality or inequality relationships between vectors and constants apply elementwise, i.e.,  $\mathbf{a} \geq 0$  indicates that every entry in the vector  $\mathbf{a}$  is greater or equal to zero.

## 1.2 Metabolic modeling

### 1.2.1 Dynamic modeling

A classical and well-known way of modeling biochemical reaction systems is via systems of ordinary differential equations that describe the rate of change of metabolite concentrations  $\mathbf{c}$  (Klipp et al., 2008). Assuming a constant cell volume, the rate of change of the concentration  $\mathbf{c}_i$  of a metabolite  $i$  is given by the difference between its rate of production and the rate at which it is consumed:

$$\frac{d\mathbf{c}_i(t)}{dt} = \mathbf{v}_{production} - \mathbf{v}_{consumption}.$$

This can be written for metabolic networks using the stoichiometric matrix  $S$ :

$$\frac{d\mathbf{c}(t)}{dt} = S\mathbf{v}(t).$$

The reaction rates  $\mathbf{v}$  are usually nonlinear functions of the substrate and enzyme concentrations<sup>1</sup> and we can write them as:

$$\mathbf{v}(t) = f(\mathbf{e}(t), \mathbf{c}(t)).$$

The deterministic kinetic rate laws  $f : \mathbb{R}_{\geq 0}^{|\mathcal{E}|} \times \mathbb{R}_{\geq 0}^{|\mathcal{M}|} \rightarrow \mathbb{R}^{|\mathcal{R}|}$  are largely unknown, and even if they are known, they depend on kinetic parameters that need to be determined experimentally. For example, if we assume Michaelis-Menten kinetics (Michaelis and Menten, 1913) for a simple irreversible reaction  $j : A \rightarrow B$ , the rate law is given by

$$\mathbf{v}_j = \mathbf{e}_j k_{cat} \frac{\mathbf{c}_A}{\mathbf{c}_A + K_M}, \quad (\text{MM})$$

where  $K_M$  is the so-called Michaelis constant, which is a constant specific to each reaction, and  $k_{cat}$  is the turnover rate of the respective enzyme and is defined as the maximum number of molecules of substrate that the enzyme can convert to product per catalytic site per unit of time.

To the basic kinetic model presented above, various degrees of regulatory information can be added, including but not limited to enzyme production dynamics via the ribosomes depending on transcription factor activities or various types of enzyme activity inhibition.

At genome-scale, such modeling approaches are largely limited by the huge number of reactions, the nonlinearity of the kinetic rate laws and the missing information about kinetic parameters. Moreover, the number of enzymological studies has reduced significantly since 1998 (based on statistics on entries in the BRENDA database (Schomburg et al., 2013)), and there is little hope that this situation will become better in the near future (Holzhütter, 2004).

### 1.2.2 Constraint-based modeling

To overcome the computational limitations and data requirements of kinetic modeling of reaction networks, constraint-based modeling has emerged as yet another popular approach (Price et al., 2004; Bordbar et al., 2014). One of the core assumptions of this type of modeling is that the internal metabolites are at steady-state, i.e., their production and consumption fluxes are balanced and thus their level is assumed to be constant.

This steady-state assumption can be motivated from two different perspectives. In the time-scales perspective, we use the fact that metabolism is much faster than other cellular processes such as gene expression. Hence, the steady-state assumption is derived as a quasi-steady-state approximation of the metabolism

<sup>1</sup>There are other additional factors that can influence the reaction rate such as (allosteric) enzyme inhibitors, pH or temperature. Depending on the level of detail of the model, these can also be included or left out.

## Chapter 1. Introduction

---

that adapts to the changing cellular conditions. However, the time scale separation argument is not necessarily needed to derive the steady-state assumption. A second perspective, on which we will focus in chapter 3, states that, on the long run, no metabolite can accumulate or deplete. In contrast to the first perspective it is not immediately clear how this perspective can be captured mathematically and what assumptions are required to obtain the steady-state condition. We leave the answers to this question for chapter 3, and now focus on what are the consequences of assuming intracellular metabolites to be at steady-state.

This assumption essentially eliminates the need for kinetic rate laws and parameters. It can be written as:

$$S\mathbf{v} = 0.$$

We will see in chapter 3 that one way to understand  $\mathbf{v}$  is as an average over time of  $\mathbf{v}(t)$ .

This transforms the problem into a linear equation system which is much easier to analyze. It however comes at several expenses: the solution space of this system, even after adding constraints about reaction directionality and bounds, is a polyhedron and hence the system is underdetermined. Additionally, a hidden assumption comes into place, namely that all enzymes are present and in the correct concentrations to sustain the resulting fluxes.

Additional information about the reversibility of reactions, inferred, for example, from thermodynamic considerations, can be incorporated by adding the constraint  $\mathbf{v}_{Irr} \geq 0$ .

To limit the size of the system, assumptions about the cellular behavior are made. Namely, the cell is assumed to optimize a particular biological objective. In the first article to introduce constraint-based modeling for metabolic networks, the assumption was that the cell tends to optimize its growth rate in the form of biomass production (“It is expected that the metabolic phenotype of wild type strains is defined by a tendency to optimize their growth rates, at least in nutritionally rich environments such as those found in the majority of bioprocesses.”, Varma and Palsson (1994), p. 995).

The maximization of biomass production is usually done by adding a pseudoreaction to the stoichiometric matrix that consumes all metabolites (amino acids, fatty acids, carbohydrates, ATP, cofactors etc.) assumed to be needed for cellular growth, and flux through this reaction is maximized. This approach is known as flux balance analysis (Varma and Palsson, 1994; Orth et al., 2010), in short FBA, and is shown below:

$$\begin{aligned} \max_{\mathbf{v} \in \mathbb{R}^{|\mathcal{R}|}} \mathbf{v}^{biomass} \\ \text{s.t. } S\mathbf{v} = 0 \\ \mathbf{v}_{Irr} \geq 0 \\ \mathbf{l} \leq \mathbf{v} \leq \mathbf{u}, \end{aligned}$$



where  $\mathbf{l}, \mathbf{u}$  are lower and upper bounds on the reaction fluxes that may come either from information about maximal reaction rates or as a need to limit growth yield as explained below.

Later on, a series of other objectives have been introduced. For an evaluation of objective functions for predicting intracellular fluxes in the bacterium *Escherichia coli* we refer the reader to (Schuetz et al., 2007). It turns out that not only there is no universally assumed cellular objective, but some of them may even be conflicting. Take, for instance, the fact that the cell is assumed to turn off “jobless” pathways to save on enzyme costs, but, at the same time, it is also assumed to be minimizing response time to changes in the environment, i.e., have the enzymes already produced to be able to react immediately. This is how multiobjective modeling of metabolism came to life (Schuetz et al., 2012).

When dealing with multiobjective problems, we no longer talk about unique optimal objective values and unique growth yields, but about sets of optimal solutions known as the Pareto front. This is the set of solutions that cannot be improved with respect to one objective without making another objective worse. All these solutions are optimal with respect to the given objectives, but they represent different compromises between the individual objectives. In addition, multiobjective problems should not be confused with problems that consider a single objective function with multiple weighted terms. In the latter case, only one of the Pareto solutions is found instead of the Pareto front, and this solution may be very sensitive to the weighting of the individual objective terms. Indeed, it has been shown via a combination of experimental and computational methods that, for the bacterium *E. coli*, the metabolism works close to the Pareto-optimal surface defined by competing objectives (Schuetz et al., 2012). In this case, the Pareto front is determined by three objectives: maximization of biomass yield, maximization of ATP yield and minimization of total flux through the network.

The advantage of the constraint-based formulation is that it gives rise to a linear optimization problem, also known as linear program (LP), which can be solved very fast. Since unlimited nutrient uptake would give rise to unlimited growth, usually a constraint that limits nutrient uptake is added, such as  $v_{\text{glucose uptake}} \leq 1$ , meaning that the cell is only allowed to take up one unit of e.g. glucose per unit time, i.e., the flux through the glucose uptake reaction is at most one.

By introducing the limits on nutrient uptake however, we do not optimize growth rate anymore, but growth yield. That is, FBA does not actually predict how fast the organism can grow, but how efficiently it can turn the nutrient into growth. This is however not what microorganisms living in rich environments do. Rather, at high substrate concentration the organisms opt for maximizing growth rate and, in doing this, they sometimes choose yield-suboptimal strategies. Only when the substrate becomes limiting do microorganisms use yield-optimal strategies (Molenaar et al., 2009).

## Chapter 1. Introduction

---

To give some examples, yeast cells use fermentation even in the presence of oxygen to process glucose at high extracellular glucose concentrations (van Dijken et al., 1993). However, fermentation is yield-suboptimal since it produces a total of only two ATP molecules per molecule of glucose. The yield-optimal strategy would be to use respiration, which results, depending on the organism, in about 30 molecules of ATP per glucose molecule.

On the other hand, although it has a high ATP yield, aerobic conversion of glucose employs large pathways with many enzymes. To run all these pathways at a high rate, the cell has to make the effort of producing all the necessary enzymes. In addition, the production of all these enzymes in high amounts results in molecular crowding of the cell, which can impede other cellular processes by hindering diffusion.

For a comparison, in fermentation, the whole Krebs cycle and oxidative phosphorylation are replaced by only three reactions needing three enzymes for catalysis.

Not only yeast, but many other biological systems are known for not acting yield-optimally. Other examples are cancer cells that display the Warburg effect (Warburg, 1956), or the lactic acid fermentation of *Bacillus subtilis* (Sonenshein, 2007). These examples, together with additional theoretical and experimental observations (Schuster et al., 2008), point out that using growth yield as a cue for growth rate may lead to false predictions. Thus, FBA is not always the way to go for predicting metabolic flux.

Another important point is that maximization of yield comes together with a hidden assumption: enzymes are present in the correct concentrations for achieving optimal yield reaction rates. However, in reality, before using an enzyme, the organism has to produce it. Producing enzymes in very high concentrations comes at two expenses: cellular resources have to be invested in enzymes, and increased molecular crowding that can hinder other biological processes.

While modeling the former is the very topic of this thesis and will be discussed extensively by giving a state of the art of metabolic resource allocation in chapter 2, the latter we briefly explain here.

The main idea of accounting for molecular crowding as a way to limit reaction flux has been addressed in (Beg et al., 2007) and in (Shlomi et al., 2011) via imposing extra constraints on the solution space. The constraints can be formulated either on the total enzyme mass, as done in (Shlomi et al., 2011), or on the total enzyme volume, as done in (Beg et al., 2007), following roughly the same principles. One typically imposes a constraint that limits the total sum of fluxes, weighted by the mass or the volume of respective metabolic enzymes. To impose these constraints one typically needs additional data such as enzyme molecular weights or volumes, which may not always be available in the literature.

In addition to the FBA disadvantages already explained, we note that it is rarely the case that FBA reports a unique optimal flux distribution. More often, there are several flux vectors with the same optimal growth yield. While imposing thermodynamic constraints can help to further shrink the optimal solution space (Beard et al., 2002, 2004; Schellenberger et al., 2011; Hamilton et al., 2013; Reimers, 2014) by further constraining some reaction reversibilities, it is rarely the case that this results in a unique solution. For this reason, typically additional methods like flux variability analysis (Mahadevan and Schilling, 2003) need to be used to check the optimal solution variability, and flux modules (Kelk et al., 2012; Müller and Bockmayr, 2014; Reimers et al., 2015) can be used to visualize which alternate pathways give rise to the variability.

### 1.3 Cellular resource allocation

Another view in metabolic modeling is that microorganisms are autocatalytic systems. They invest resources in terms of nutrients and time into enzymes that they then use to catalyze reactions to again produce more enzymes. In addition, ribosomes need to be present to translate the enzymes, and the ribosomes themselves need to be produced, following the same principle. A question of prioritization then arises: in which components should the cell invest such that it grows optimally?

This is why, in recent years, the systems biology of metabolism has moved more and more from classical metabolic network study towards the study of growth as a result of an optimized cellular economy.

The ideas that a cell minimizes the total enzyme concentration needed for a fixed steady-state flux and that there exists a competition among reactions for available enzyme resources are however not new, but go back to (Brown, 1991) and (Klipp and Heinrich, 1999). Moreover, similar ideas have been used to study activation of metabolic pathways and enzyme allocation under a constraint of limited total enzymatic capacity starting with (Klipp et al., 2002). This was then further investigated using experimental (Zaslaver et al., 2004) as well as mathematical techniques (Oyarzún et al., 2009; Bartl et al., 2010) and was even brought forward in a multiobjective dynamic optimization study (de Hijas-Liste et al., 2014).

In their article, (Molenaar et al., 2009) pointed out that the limited total proteome constraint does not apply only to metabolism, but to all cellular subsystems, which have to share resources among each other. In this study a small dynamic model of a self-replicating system is used to explain how overflow metabolism arises by means of tradeoffs between different growth strategies. Further on, Goelzer et al. (2011) introduced resource balance analysis (RBA), as a means of predicting the cell composition of bacteria in a specific (constant) environment through a convex optimization problem that takes into account the bioenergetic costs of running a pathway. More or less in parallel, Lerman

## Chapter 1. Introduction

---

et al. (2012) introduced the idea of an integrated model of metabolism and gene expression (ME model) as a means to explore the relationship between genotype and phenotype using biochemical representations of transcription and translation processes. Their research group then continued with an ME model of the model organism *Escherichia coli* (O'Brien et al., 2013). Also experimental studies focused on relating absolute protein abundances to how metabolic pathways balance production costs and activity requirements (Li et al., 2014).

These steady-state resource allocation formalisms have then been combined with the dynamic optimization ideas in (Klipp et al., 2002; Oyarzún et al., 2009; de Hijas-Liste et al., 2014), to understand how resources are distributed in a dynamically changing environment by means of a dynamic enzyme-cost flux balance analysis (deFBA) and conditional flux balance analysis (cFBA) (Waldherr et al., 2015; Rügen et al., 2015).

Such dynamic resource allocation models have a wide area of applicability. One such an example is the study of microorganisms growing in industry-scale bioreactors. There, the organism has to balance resources not only in order to be able to grow optimally, but also in order to survive transitions through local heterogeneities of the reactor. The ability to take such transitions into account within metabolism has been shown to be crucial for survival (van Heerden et al., 2014).

In addition, a perfect study case is the metabolism of phototrophic organisms, who live in regular day-night light conditions. Dynamic resource allocation modeling helps us understand how they organize their synthesis and storage processes following light availability.

The main underlying hypothesis of such studies is that organisms have been shaped through evolution by reoccurring changes in their environment and that the observed patterns in their metabolism are the result of optimized growth in this changing environment. Thus, along the lines of (Klipp, 2009), we can check using optimization principles and models if what we observe in nature are indeed optimal behaviors.

### 1.4 Structure of this thesis

Following the recent developments in optimal dynamic resource allocation formalisms, this thesis focuses on modeling and understanding resource allocation in large (sometimes genome-scale) metabolic models.

In chapter 2 we start with a state of the art overview on metabolic resource allocation formalisms.

We then present in chapter 3 a new mathematical derivation of the widely used steady-state assumption that motivates its successful use in many applications. This derivation can then be used to quickly provide upper bounds on growth rates in dynamic resource allocation problems.

Chapter 4 then provides a protocol on how to generate a dynamic resource allocation model. Before we turn our attention to applications, we present in chapter 5 some guidelines, useful problem transformations and instructions for numerically solving the resulting dynamic resource allocation problems.

Having set the stage this way, we proceed in chapter 6 with an in depth study at genome-scale of the metabolic resource allocation in the cyanobacterium *Synechococcus elongatus* PCC 7942. This is the first genome-scale dynamic resource allocation study and it presents insight into the temporal organization of synthesis processes following light availability. Moreover, it shows that the linear pattern of the accumulation of storage (glycogen) throughout the day period is an optimal behavior that arises as a tradeoff between several conflicting resource allocation objectives.

We continue in chapter 7 with a study of yeast living in dynamically changing environments. This is motivated by the observation that, in industrial-scale bioreactors, yeast cells live in a changing environment governed by local heterogeneities. We therefore want to understand what mechanisms enable some of the cells to survive such environment transitions, and why other subpopulations do not make it through the transitions and die out. We show that overflow metabolism and diauxie, phenomena that are not exclusively observed in yeast but widely spread in nature, are optimal behaviors from a resource allocation perspective. Moreover, we investigate how one can use resource allocation models to understand how yeast adapts to oxygen and nutrient availability shifts.

A next step, which we explore in chapter 8, is the perspective of using time courses from dynamic resource allocation models to infer the regulatory structures that implement these optimal behaviors.

We conclude by summarizing the results of the thesis and listing possible next steps in using dynamic resource allocation models to understand metabolism and its regulation.



## Chapter 2

# Resource allocation formalisms

This chapter aims at giving an overview of the current metabolic resource allocation formalisms and exemplifies them using the toy model proposed in (Molenaar et al., 2009). This will help us better understand the methods used for the case studies in chapters 6 and 7.

### 2.1 Resource allocation principles in a self-replicator

In their article, (Molenaar et al., 2009) start by noting that there are several common patterns that can be observed in the physiology of unicellular organisms, such as increase of cell size and ribosomal content with increasing growth rate, or a shift to energetically inefficient metabolism at high growth rates. Moreover, such patterns are not only observed in unicellular organisms, but also in tumor cells for example. They argue that such patterns hint at the existence of design principles that result in the optimization of evolutionary fitness.

This idea of fitness optimality is not new. It has already been applied successfully in the study of cellular subsystems like metabolism. There, global metabolic reaction rates (fluxes) are predicted using a network of biochemical reactions and assuming that cells have evolved towards optimizing growth yield in a constant environment. While the resulting predictions have been shown to fit experimental measurements relatively well (Edwards et al., 2001; Ibarra et al., 2002), such models fail to predict the usage of inefficient metabolic routes and “spilling” of resources at high nutrient concentrations, a phenomenon known in the literature as overflow metabolism. The reason, according to (Molenaar et al., 2009), lies in the modeling of only one subsystem (metabolism), rather than the subsystem in its context, where the cell would have to produce catalytic units before these can be used to catalyze reactions and would have to balance the costs and benefits of the macromolecules it produces together with the effects of such macromolecules on the overall growth rate.

## Chapter 2. Resource allocation formalisms

---

To prove their theory, (Molenaar et al., 2009) propose a self-replicator model that contains most cellular subsystems, including production of macromolecules such as membrane, the ribosome, or transporters. We have reproduced this self-replicator model in figure 2.1, and we will use it to explain the state of the art of resource allocation formalisms in the next sections. In addition, we present in table 2.1 the list of species and reactions of this model.

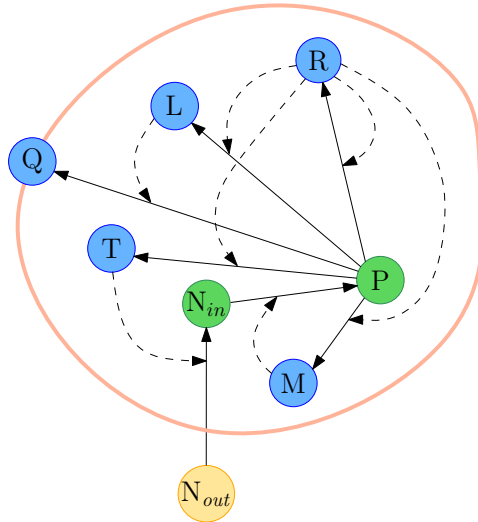


Figure 2.1: Self-replicator toy model proposed by (Molenaar et al., 2009) in their resource allocation study. Extracellular nutrients are depicted in yellow, intracellular metabolites are shown in green, and macromolecules in blue. Reactions are shown in continuous arrows, while catalysis relationships are in dashed arrows. The model contains all subsystems of a cell:  $N_{out}$ , extracellular nutrient source;  $N_{in}$ , intracellular nutrient;  $P$ , metabolic precursors;  $L$ , lipid synthesis enzyme;  $Q$ , lipid membrane;  $R$ , ribosome;  $T$ , transporter.

(Molenaar et al., 2009) consider the self-replicator system in a growth medium with infinite volume and a constant substrate concentration  $c_{N_{out}}$ . Furthermore, they assume the main objective of the system is to maximize the growth rate  $\mu$  under a balanced growth condition, where each component grows at the same rate. The main limitations of growth are constraints imposed by physics and chemistry that we will detail in the following subsections.

### 2.1.1 Growth rate and dilution

To understand the growth rate  $\mu$ , we need to look at the volume of the system, which is assumed to grow exponentially, following the equation

$$V(t) = V_0 \cdot e^{\mu \cdot t}, \quad (2.1)$$



## 2.1 Resource allocation principles in a self-replicator

Table 2.1: List of species, reactions, and catalysis relationships for the model in figure 2.1. We note that we have a one to one correspondence between reactions and their catalyzing enzymes and we therefore denote them by the same letter.

External metabolites:		$N_{out}$
Internal metabolites:	$\mathcal{M} = \{N_{in}, P, Q\}$	
Enzymes:	$\mathcal{E} = \{L, M, R, T\}$	
Reactions	Catalyzed by	Turnover rate
$N_{out} \rightarrow N_{in}$	T	7
$N_{in} \rightarrow P$	M	5
$P \rightarrow Q$	L	5
$P \rightarrow L$	R	3
$P \rightarrow M$	R	3
$P \rightarrow R$	R	3
$P \rightarrow T$	R	3

where  $V_0$  is the initial volume of the bacterial population. By taking the derivative in equation 2.1, we obtain

$$\mu = \frac{1}{V(t)} \frac{dV(t)}{dt}, \quad (2.2)$$

consistent with the definition in (Heinrich and Schuster, 1996). The case where the growth rate is not constant is discussed in section 5.3.

Following the definition of concentration of a molecule  $i$  as number of moles  $\mathbf{n}_i(t)$  per volume  $V(t)$ ,  $\mathbf{c}_i(t) = \frac{\mathbf{n}_i(t)}{V(t)}$ , we obtain then that

$$\begin{aligned} \frac{d}{dt} \mathbf{c}_i(t) &= \frac{d}{dt} \frac{\mathbf{n}_i(t)}{V(t)} \\ &= \frac{d\mathbf{n}_i(t)}{dt} \frac{1}{V(t)} - \frac{dV(t)}{dt} \frac{1}{V(t)} \frac{\mathbf{n}_i(t)}{V(t)} \\ &= \underbrace{\frac{d\mathbf{n}_i(t)}{dt} \frac{1}{V(t)}}_{\text{production}} - \underbrace{\mu \frac{\mathbf{n}_i(t)}{V(t)}}_{\text{dilution by growth}} \end{aligned}$$

### 2.1.2 Mass balance equations

The authors of (Molenaar et al., 2009) assume an exponential growth model for the biomass of the system where  $\mu$  is the constant specific growth rate. In addition, they assume a steady-state in the form of balanced growth, i.e., the rates of production and dilution by growth of each intracellular species are balanced as

$$\begin{aligned} 0 &= \frac{d\mathbf{c}(t)}{dt} = \sum \mathbf{v}^{\text{synthesis}}(t) - \sum \mathbf{v}^{\text{dilution}}(t) = S_{\mathcal{M}^*} \mathbf{v}(t) - \mu \mathbf{c}(t), \\ 0 &= \frac{d\mathbf{e}(t)}{dt} = \sum \mathbf{v}^{\text{synthesis}}(t) - \sum \mathbf{v}^{\text{dilution}}(t) = S_{\mathcal{E}^*} \mathbf{v}(t) - \mu \mathbf{e}(t), \end{aligned}$$

## Chapter 2. Resource allocation formalisms

---

where  $S$  is the stoichiometric matrix. Therefore, they assume that the fluxes and the concentrations are constant and thus we omit their time dependency in this section and write simply  $\mathbf{v}$ ,  $\mathbf{c}$ , and  $\mathbf{e}$ .

The system can modulate the growth rate  $\mu$ , by modulating the production rate of each protein. That means that the fractions of ribosomes used for the synthesis of each protein can be changed and for each  $i \in \{L, M, R, T\}$  we have

$$\frac{d\mathbf{e}_i(t)}{dt} = S_{i*} \mathbf{v}(t) - \mu \mathbf{e}_i(t) = \alpha_i \mathbf{v}_R - \mu \mathbf{e}_i = 0,$$

where  $\mathbf{v}_R$  is the rate at which the total ribosome pool synthesizes proteins,  $\alpha_i$  is the fraction of this pool dedicated to the synthesis of protein  $i$ , and the stoichiometry  $S_{i*}$  is omitted because all coefficients are one.

### 2.1.3 Kinetics of rate equations

We have already seen in the previous section that the flux values  $\mathbf{v}$  play an important role in the growth of the system. In this model, they are all assumed to follow irreversible Michaelis-Menten kinetics, as described in table 2.2.

Table 2.2: Rate equations for the reactions of the self-replicator.  $k_{cat}^i$  is the turnover rate of enzyme  $i$ , which can be found in table 2.1.  $K_M$  is the Michaelis constant and is assumed to be equal to 1 for all reactions.

Reaction	Rate equation
$N_{out} \rightarrow N_{in}$	$\mathbf{v}_T = \frac{k_{cat}^T \cdot \mathbf{e}_T \cdot \mathbf{c}_{N_{out}}}{K_M + \mathbf{c}_{N_{out}}}$
$N_{in} \rightarrow P$	$\mathbf{v}_M = \frac{k_{cat}^M \cdot \mathbf{e}_M \cdot \mathbf{c}_{N_{in}}}{K_M + \mathbf{c}_{N_{in}}}$
$P \rightarrow Q$	$\mathbf{v}_L = \frac{k_{cat}^L \cdot \mathbf{e}_L \cdot \mathbf{c}_P}{K_M + \mathbf{c}_P}$
$P \rightarrow i \in \{L, M, R, T\}$	$\mathbf{v}_R = \frac{k_{cat}^R \cdot \mathbf{e}_R \cdot \mathbf{c}_P}{K_M + \mathbf{c}_P}$

### 2.1.4 Membrane integrity

An important observation is that the lipids Q, which we call from now on quota metabolites, will in fact never be produced in a solution that is optimal with respect to maximizing  $\mu$  given the constraints above. They are simply a drain of resources that does not contribute to the autocatalytic nature of the system. However, in reality, the membrane lipids play an important role in keeping the integrity of a cell. Therefore, a constraint is imposed to make sure that the membrane of the self-replicator is not only made of transporters, but a lipid quota is

## 2.1 Resource allocation principles in a self-replicator

---

also present. This is done by imposing that the amount of transporters needs to be at most equal to that of membrane lipids, i.e.,

$$\mathbf{c}_Q \geq \mathbf{e}_T.$$

Just as the lipids, also DNA and RNA are important components without catalytic role that are essential for replicating the cell. They are not considered explicitly in this model, but they will later on play a role in metabolism and expression models (ME models), resource balance analysis (RBA), dynamic enzyme-cost flux balance analysis (deFBA), and conditional flux balance analysis (cFBA), which we explain in later chapters. We call these quota metabolites, to highlight that a certain quota in the total biomass has to be dedicated to them.

### 2.1.5 Total proteome is limited

The volume of a cell is obviously not infinite, and that means there is also a limit to how much protein can be contained inside a cell. Therefore, the authors of (Molenaar et al., 2009) impose a constraint on the total amount of proteins as

$$\sum_{i \in \{L, M, R, T\}} \mathbf{e}_i \leq 1.$$

As mentioned in the introduction, the idea of such a constraint is not new. A version of this constraint has been already introduced in (Klipp et al., 2002) and later on used in flux balance analysis with molecular crowding (Beg et al., 2007), as well as in other metabolic resource allocation formalisms as detailed in the following sections.

### 2.1.6 Additional constraints

There are also more obvious constraints that we need for the self-replicator model. For instance, species concentrations should be nonnegative, which can be expressed as

$$\mathbf{c}, \mathbf{e} \geq 0.$$

Additional volume constraints are used in the article. We do not detail them because they do not bring any additional understanding of the resource allocation.

Finally, the ribosome fractions used for protein synthesis should sum up to 1,

$$\sum_{i \in \{L, M, R, T\}} \alpha_i = 1.$$

### 2.1.7 The nonlinear optimization problem

Putting together the constraints above, we obtain a nonlinear optimization problem (NLP) over the variables  $\mu$ ,  $\mathbf{c}$ ,  $\mathbf{e}$ ,  $\mathbf{v}$ ,  $\alpha$  for every fixed  $\mathbf{c}_{N_{out}}$ :

$$\begin{aligned} & \max_{\alpha, \mu, \mathbf{c}, \mathbf{e}, \mathbf{v}} \mu \\ & \text{s.t. } S_{i^*} \cdot \mathbf{v} - \mu \cdot \mathbf{c}_i = 0 \quad \forall i \in \{N_{in}, P, Q\} \end{aligned} \quad (2.3)$$

$$\alpha_i \cdot \mathbf{v}_R - \mu \cdot \mathbf{e}_i = 0 \quad \forall i \in \{L, M, R, T\} \quad (2.4)$$

$$\mathbf{v}_T = \frac{k_{cat}^T \cdot \mathbf{e}_T \cdot \mathbf{c}_{N_{out}}}{K_M + \mathbf{c}_{N_{out}}} \quad (2.5)$$

$$\mathbf{v}_M = \frac{k_{cat}^M \cdot \mathbf{e}_M \cdot \mathbf{c}_{N_{in}}}{K_M + \mathbf{c}_{N_{in}}} \quad (2.6)$$

$$\mathbf{v}_i = \frac{k_{cat}^i \cdot \mathbf{e}_i \cdot \mathbf{c}_P}{K_M + \mathbf{c}_P} \quad \forall i \in \{L, R\} \quad (2.7)$$

$$\mathbf{c}_Q \geq \mathbf{e}_T \quad (2.8)$$

$$\sum_{i \in \{L, M, R, T\}} \mathbf{e}_i \leq 1 \quad (2.9)$$

$$\alpha, \mathbf{c}, \mathbf{e} \geq 0 \quad (2.10)$$

$$\sum_{i \in \{L, M, R, T\}} \alpha_i = 1. \quad (2.11)$$

The authors solve the optimization problem using GAMS (<https://www.gams.com/>) in combination with the KNITRO solver. This solver only guarantees finding local solutions to the NLP. The authors then check the global solution optimality using the LINDOGlobal solver. Such solvers can use branch and cut methods to break the NLP into subproblems which are either infeasible, optimal or that are in turn split again. At additional computational cost, they can automatically linearize some of the nonlinear relationships. They guarantee global optimality within a user-set tolerance. Please note that using an NLP solver for such a small model may be feasible in terms of the solving time, but it will definitely not be an option anymore if the problem grows to hundreds or even thousands of variables.

### 2.1.8 Results and extensions

Although it is a simple model, the self-replicator and its extensions presented in (Molenaar et al., 2009) are capable of displaying many behaviors observed in real organisms. The authors show how growth of the ribosome pool with increasing growth rate, overflow metabolism, or growth strategies on two substrates arise as tradeoffs between the costs and benefits of proteome allocation.

The model has inspired a series of resource allocation studies, of which we detail resource balance analysis (RBA), metabolism and gene expression models (ME

models), dynamic enzyme-cost flux balance analysis (deFBA), and conditional flux balance analysis (cFBA) in the following sections.

## 2.2 Resource balance analysis

Introduced by (Goelzer et al., 2011), RBA exploits the same idea as the resource allocation model in (Molenaar et al., 2009): modeling the cell as an interdependence of several subsystems that complete different tasks contributing to growth and that share common resources. There are however several differences to the work of (Goelzer et al., 2011), which we will detail below.

To begin with, the authors of RBA give a formal definition of the growth rate  $\mu$  as a function of the volume of the cell, as we have already detailed it in the previous section. As in our steady-state assumption including dilution effects in section 5.3 and as in the article of (Molenaar et al., 2009), they note that at steady-state the production and dilution of all cellular macromolecules should balance in order to maintain their concentrations constant. Furthermore, they also assume that the fluxes are constant.

### 2.2.1 Relaxing steady-state constraints and imposing quota

However, they relax the steady-state constraint compared to equations (2.3) and (2.4), and allow overproduction of metabolic precursors and macromolecules, but not of metabolic intermediates. Assuming the stoichiometric matrix also contains the macromolecule production reactions, equations (2.3) and (2.4) are replaced by

$$\begin{aligned}
 S_{P^*} \cdot \mathbf{v} &\geq 0, \\
 S_{N_{in}^*} \cdot \mathbf{v} &= 0, \\
 S_{i^*} \cdot \mathbf{v} - \mu \cdot \mathbf{e}_i &\geq 0, & \forall i \in \{L, M, R, T\}, \\
 S_{Q^*} \cdot \mathbf{v} - \mu \cdot \mathbf{c}_Q &\geq 0.
 \end{aligned}$$

We note also that only dilution of macromolecules by cell growth is modeled in RBA, but not dilution of internal metabolites  $N_{in}$  and  $P$ . In chapter 5 we estimate the error that we make by neglecting metabolite dilution via cell growth.

Furthermore, by including  $Q$  in the latter equation and fixing the required amount  $\mathbf{c}_Q \geq 0$ , quota metabolites are required to also grow at the rate  $\mu$ .

### 2.2.2 Kinetics of rate equations are replaced by linear constraints

Reaction rates in RBA are not modeled using kinetic rate laws as in (Molenaar et al., 2009). Instead the authors assume that enzymes are substrate-saturated and that the reaction rate is given by the product of the turnover rate and the enzyme amount. As reasoning, the authors cite a PhD thesis written in French

## Chapter 2. Resource allocation formalisms

---

that is not available online, but which suggests that “at steady state, the enzymes operate at (or close to) saturation” (Goelzer et al., 2011). Although not mentioned in the RBA article, in (Bennett et al., 2009) absolute metabolite concentrations in *Escherichia coli* are compared to the  $K_M$  values of their degrading enzymes, and the results indeed indicate that most enzymes analyzed (83%) are more than 50% saturated with substrate. Moreover, 59% of the analyzed enzymes process substrates in a concentration that is more than 10-fold higher than their  $K_M$ . These results indicate that, for most reactions  $r$  catalyzed by an enzyme  $E$ , the reaction flux is somewhere in the interval  $[\frac{1}{2}k_{cat}^E \cdot \mathbf{e}_E, k_{cat}^E \cdot \mathbf{e}_E]$ , assuming simple Michaelis-Menten kinetics. The equality assumption that  $\mathbf{v}_r = k_{cat}^E \cdot \mathbf{e}_E$  in RBA is still quite strong, since it forces flux whenever the enzyme is present although regulatory effects may impose that only a fraction of this flux is present at times.

Nevertheless, this assumption simplifies the equations for fluxes (2.5)-(2.7) to

$$|\mathbf{v}_i| = k_{cat}^i \cdot \mathbf{e}_i, \forall i \in \{T, M, L, R\}.$$

This is the first time such an approximation is introduced in order to linearize kinetic expressions. The absolute value is present because some reactions may be reversible, and in that case their flux values, albeit negative, should still be constrained by the enzyme amount. This usage of the absolute value in the constraint introduces a nonlinearity into the optimization problem. However, this is easily resolved by introducing extra variables  $\mathbf{v}^{max} \geq 0$ , and requiring that

$$\mathbf{v}_i - \mathbf{v}_i^{max} \leq 0 \quad \text{and} \quad -\mathbf{v}_i - \mathbf{v}_i^{max} \leq 0, \forall i \in \{T, M, L, R\}.$$

Note that this transformation relaxes the original equality constraint, into  $\mathbf{v}_r \leq k_{cat}^E \cdot \mathbf{e}_E$ , and thus the saturation hypothesis is given up.

It should additionally be noted that, instead of using the fractions  $\alpha$  to model competition of the protein synthesis reactions for the ribosome, RBA uses an upper bound on the sum of enzyme production fluxes in combination with the rate equation linearization as

$$\sum_{i \in \{T, M, L, R\}} \mathbf{v}_{R_i} \leq k_{cat}^R \cdot \mathbf{e}_R,$$

where  $\mathbf{v}_{R_i}$  denotes the synthesis flux for protein  $i$ .

### 2.2.3 Modeling volume vs. modeling density

As in (Molenaar et al., 2009), RBA also does not explicitly model volume. Instead, a density constraint is used that requires the intracellular density to remain constant to ensure diffusion is not impaired. In detail, the weighted sum of the densities of the macromolecules is bounded by the mean density  $\bar{D}$  of the cell as

$$\sum_{i \in \{T, L, M, R\}} \rho_i \cdot \mathbf{e}_i + \rho_Q \cdot \mathbf{c}_Q \leq \bar{D},$$

where  $\rho_i$  is the density of macromolecule  $i$ . This effectively means that the saturation assumption from above is given up.

This constraint is only imposed for the macromolecules, since metabolite concentrations are not modeled explicitly and metabolites are assumed to be small enough to not impair diffusion. The main role of this constraint is to provide an upper bound on total proteome, similar to constraint (2.9).

### 2.2.4 Nutrient uptake

The RBA article does not explicitly mention how extracellular nutrient uptake is modeled. We assume that they use a Michaelis-Menten rate law as in (Molenaar et al., 2009), and fix the parameters  $k_{cat}^T, K_M$ , as well as the substrate concentration  $\mathbf{c}_{N_{out}}$ . If this is the case, the uptake rate  $\mathbf{v}_T$  depends linearly on the amount of enzyme  $\mathbf{e}_T$ , while the rest of the terms are fixed, as

$$\mathbf{v}_T = \frac{k_{cat}^T \cdot \mathbf{e}_T \cdot \mathbf{c}_{N_{out}}}{K_M + \mathbf{c}_{N_{out}}}.$$

### 2.2.5 The quadratically constrained optimization problem

By replacing the kinetic expressions for the fluxes with linear terms that only depend on enzyme concentrations and  $k_{cat}$ , RBA casts the resource allocation problem into a quadratic program, with  $\mu$  as common variable in all quadratic terms. Putting together all constraints, RBA for the toy model in figure 2.1 is given by:

$$\max_{\mu, \mathbf{c}, \mathbf{e}, \mathbf{v}^{max}} \mu$$

$$\text{s.t. } S_{P^*} \cdot \mathbf{v} \geq 0 \quad (2.12)$$

$$S_{N_{in}^*} \cdot \mathbf{v} = 0 \quad (2.13)$$

$$S_{i^*} \cdot \mathbf{v} - \mu \cdot \mathbf{e}_i \geq 0 \quad \forall i \in \{L, M, R, T\} \quad (2.14)$$

$$S_{Q^*} \cdot \mathbf{v} - \mu \cdot \mathbf{c}_Q \geq 0 \quad (2.15)$$

$$\mathbf{v}_i^{max} = k_{cat}^i \cdot \mathbf{e}_i \quad \forall i \in \{L, M, R, T\} \quad (2.16)$$

$$\mathbf{v}_i - \mathbf{v}_i^{max} \leq 0 \quad \forall i \in \{L, M, R, T\} \quad (2.17)$$

$$-\mathbf{v}_i - \mathbf{v}_i^{max} \leq 0 \quad \forall i \in \{L, M, R, T\} \quad (2.18)$$

$$\sum_{i \in \{T, M, L, R\}} \mathbf{v}_{R_i} \leq k_{cat}^R \cdot \mathbf{e}_R \quad (2.19)$$

$$\mathbf{v}_T = \frac{k_{cat}^T \cdot \mathbf{e}_T \cdot \mathbf{c}_{N_{out}}}{K_M + \mathbf{c}_{N_{out}}} \quad (2.20)$$

$$\sum_{i \in \{T, L, M, R\}} \rho_i \cdot \mathbf{e}_i + \rho_Q \cdot \mathbf{c}_Q \leq \bar{D} \quad (2.21)$$

$$\mathbf{v}^{max}, \mathbf{c}, \mathbf{e}, \mu \geq 0. \quad (2.22)$$

## Chapter 2. Resource allocation formalisms

---

Since  $\mu$  is common to all quadratic terms, instead of solving a quadratic problem one can use a half-interval search (also known as binary search) over  $\mu$ , and for every fixed  $\mu$  solve a linear feasibility problem. Briefly, one starts with a very small fixed value of  $\mu$  and successively doubles it until the resulting feasibility problem becomes infeasible. Once this is the case, a typical binary search can be used to search the optimal value for  $\mu$ , which is now in the interval between the last feasible  $\mu$  and its double. Usually one has to set a tolerance for the smallest change in  $\mu$  at which this procedure should stop.

### 2.2.6 Mathematical and biological implications

In addition to providing a resource allocation framework that is efficiently solvable also for large-scale systems, the authors prove mathematically several important properties about the framework, of which we mention here a few.

The article provides a proof that, if the resulting feasibility problem is feasible for a fixed  $\mu^*$ , then it is also feasible for any other value of  $\mu$  in the interval  $[0, \mu^*]$ . They also prove that there exists a finite optimal  $\mu^*$  and that every value larger than this will result in an infeasible problem.

The authors show that the growth rate  $\mu$  increases when the fixed required non-catalytic biomass quota  $c_Q$  is decreased. This fact has also been proven experimentally in (Fischer and Sauer, 2005) by deleting the inductor of expression for flagellar proteins in *Bacillus subtilis*, which are proteins only required for mobility, with no catalytic role. The resulting mutant strain displayed a faster growth rate than the wild type.

Last but not least, (Goelzer et al., 2011) show that increasing the turnover rates reduces the necessary amounts of enzymes for running the same flux and thus increases growth rate, and that “cheap” pathways (in terms of synthesis investment) are nearly always preferred (e.g. uptake amino acids rather than *de novo* synthesis). They also point out that an “expensive” pathway may be preferred sometimes if it reduces cost of producing a co-metabolite needed somewhere else in the system.

(Goelzer et al., 2011) apply RBA to a model of *Bacillus subtilis* metabolism, with 342 genes, 277 enzymes, 54 transporters, and 358 metabolic reactions and show that the modular configuration of this metabolic network is a function of the medium composition.

### 2.2.7 Extension through metabolism and gene expression models

Up to now we have seen how RBA takes into account resource investment into metabolic pathways by modeling protein translation explicitly. One can however go one level of detail higher, as in metabolism and gene expression models (ME models) (Lerman et al., 2012; O’Brien et al., 2013), and also model the transcription process. One can then provide limits on the number of proteins that



can be translated from a given messenger RNA (mRNA) before this mRNA is degraded or passed on to daughter cells. The synthesis of mRNA has to then match the costs of degradation and dilution via growth, which can be achieved through coupling constraints, as we will detail in this section. We, however, do not aim here at a full description of ME models, but more at a high level view on the extension they provide to RBA.

In figure 2.2 we show the transcription and translation processes, as they are modeled in ME models. In addition to metabolites, enzymes, and quota components, ME models also explicitly incorporate RNA as mRNA, rRNA, and tRNA. Furthermore, they model production, usage, dilution to daughter cells, and degradation of enzymes and RNA molecules through so-called “coupling constraints”.

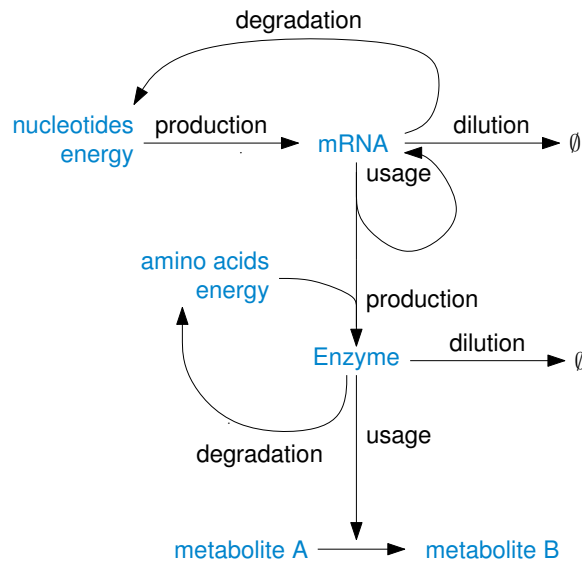


Figure 2.2: Scheme of the transcription and translation processes as they are modeled in ME models. Metabolites and macromolecules are depicted in blue, while reactions and degradation/dilution by growth processes are shown in black.

### Growth rate vs. doubling time

Just as RBA, ME models include dilution via growth using the growth rate  $\mu$ . However, sometimes the constraints are expressed as a function of the doubling time  $T_d$ , which can be obtained from the growth rate  $\mu$  as

$$T_d := \frac{\ln 2}{\mu}.$$

## Chapter 2. Resource allocation formalisms

---

### Dilution and degradation of mRNA

For each enzyme modeled, the ME model additionally keeps track of the mRNA molecule which is used to translate the enzyme. That means, that for every enzyme  $i \in \{T, M, L, R\}$  we model the production, dilution, and degradation fluxes involving  $\text{mRNA}_i$  using coupling constraints. The first coupling constraint relates dilution and degradation of such mRNAs by imposing that a certain mRNA can be degraded a maximum number of times before it is passed on to a daughter cell. In short,

$$\mathbf{v}_i^{\text{dilution}} \geq a_{\max} \cdot \mathbf{v}_i^{\text{degradation}}, \forall i \in \{\text{mRNA}_T, \text{mRNA}_M, \text{mRNA}_L, \text{mRNA}_R\}.$$

The parameter  $a_{\max}$  is the ratio of the mean lifetime of an mRNA molecule  $\tau_{\text{mRNA}}$  and the doubling time  $T_d$ ,

$$a_{\max} := \frac{\tau_{\text{mRNA}}}{T_d}.$$

In the simulations  $\tau_{\text{mRNA}}$  is typically set to 5 minutes, since 80% of mRNAs in *E. coli* have been shown to have half-lives between 3 and 8 minutes (Bernstein et al., 2002).  $T_d$ , as a function of the growth rate  $\mu$ , comes into this constraint as a variable which is instantiated at each new iteration of the binary search explained for RBA. The fluxes  $\mathbf{v}_i^{\text{dilution}}$ ,  $\mathbf{v}_i^{\text{degradation}}$  are variables of the LPs solved at each binary search iteration.

Although the authors do not explicitly mention this fact and since the system is assumed to grow exponentially, we expect that the expression for the dilution flux is given by

$$\mathbf{v}_i^{\text{dilution}} = \mu \cdot \mathbf{c}_i, \forall i \in \{\text{mRNA}_T, \text{mRNA}_M, \text{mRNA}_L, \text{mRNA}_R\}.$$

### Limited mRNA usage for translation

The second coupling constraint relates the mRNA layer to the enzyme layer and thus imposes a bound on the number of times an mRNA can be used for translation before it has to be degraded. Concretely,

$$\mathbf{v}_i^{\text{degradation}} \geq b_{\max} \cdot \mathbf{v}_i^{\text{usage}}, \forall i \in \{\text{mRNA}_T, \text{mRNA}_M, \text{mRNA}_L, \text{mRNA}_R\}.$$

The parameter  $b_{\max}$  is dependent on the ribosome translation rate in amino acids per second, the ribosome footprint (how many ribosomes fit on the transcript), the length of the transcript, and the mean lifetime of the mRNA molecule  $\tau_{\text{mRNA}}$ . To give an example, let us look at an mRNA that is 1000 nucleotides long. Since the space a ribosome takes on the transcript is about 20 nucleotides long, it follows that approximately 50 ribosomes can fit on this mRNA. Moreover, since three nucleotides form a codon that translates an amino acid, the resulting protein will be 333 amino acids long. Assuming we model a bacterium, the maximal

translation rate of the ribosome will be about 20 amino acids per second (Young and Bremer, 1976). Given these numbers, the maximum translation rate is given by

$$k_{cat}^R = 50 \text{ ribosomes} \cdot \frac{20 \text{ amino acids}}{\text{second} \cdot \text{ribosome}} \cdot \frac{1 \text{ protein}}{333 \text{ amino acids}} \simeq 3 \frac{\text{proteins}}{\text{second}}.$$

$b_{max}$  is then given by  $\frac{1}{k_{cat}^T \cdot \tau_{mRNA}}$ , which for the example protein would be 1/900. Naturally,  $b_{max}$  needs to be computed for every modeled mRNA in this fashion.

Although the authors do not explicitly mention it, we assume that the usage flux for the mRNA is the same as the production flux for its corresponding protein, and thus

$$\mathbf{v}_{mRNA_i}^{\text{usage}} = \mathbf{v}_i^{\text{production}}, \forall i \in \{T, M, L, R\}.$$

### Limited enzyme usage for catalysis

In the case of enzymes, the authors of the ME model formalism assume that the degradation flux is negligible compared to the dilution flux,

$$\mathbf{v}_i^{\text{degradation}} \ll \mathbf{v}_i^{\text{dilution}}.$$

Thus, as a third coupling constraint, the enzyme layer is related to the metabolism layer by imposing a total bound on how many times an enzyme is used for catalysis before it is passed on to a daughter cell, similar to the coupling constraint of the previous layer,

$$\mathbf{v}_i^{\text{dilution}} \geq c_{max} \cdot \mathbf{v}_i^{\text{usage}}, \forall i \in \{T, M, L, R\}.$$

Again, although the authors do not explicitly mention the fact, we assume that the usage flux for the enzyme is related to the fluxes of the reactions it catalyzes as

$$\mathbf{v}_i^{\text{usage}} = \frac{1}{k_{cat}^i} \sum_{j \in \nu_i} \mathbf{v}_j,$$

where  $\nu_i$  gives the set of metabolic reactions catalyzed by the enzyme  $i$ , and the underlying assumption is that the enzyme catalyzes these reactions at the same turnover rate  $k_{cat}^i$ .

The parameter  $c_{max}$  is then dependent on how fast the enzyme is ( $k_{cat}^i$ ) and on the doubling time as

$$c_{max} = \frac{1}{k_{cat}^i \cdot T_d}.$$

All these coupling constraints between the different modeled layers of the system are then added to the RBA formulation and the whole problem is solved as explained for RBA.

We note that the number of parameters of the model substantially increases by adding the mRNA layer and that reasonable estimates of all these parameters may not always be available in the literature.

## 2.3 Dynamic enzyme-cost flux balance analysis

We have already seen some formalisms for predicting the optimal metabolic resource allocation in a constant given environment. However, cells very rarely live in a constant environment outside of the laboratory. Instead, they are often faced with the choice between being a “generalist” and a “specialist”. This means a choice between being robust against changes in the environment, expressing enzymes that are not needed in that environment, at the expense of a lower growth rate versus achieving a maximum growth rate in that environment at the expense of little robustness and the risk of not being able to adjust to sudden changes.

In this respect, a different sort of resource allocation model is needed, that allows tracking the effect of environmental changes on the metabolic resource allocation. For this purpose, dynamic enzyme-cost flux balance analysis (Waldherr et al., 2015), in short deFBA, has been introduced.

### 2.3.1 Mass balance equations and quasi-steady-state of metabolism

A first observation we make is that as opposed to the model in (Molenaar et al., 2009) and to RBA and ME models, deFBA does not model a cell, but a population. As such, the growth rate  $\mu$  does not appear explicitly throughout the deFBA formalism.

The authors of deFBA do not make any assumptions about the growth of the system, but instead begin by writing the mass balance equations of the system. However, these are now using molar amounts  $\mathbf{n}(t)$  instead of concentrations. Furthermore, molar amounts and fluxes are no longer considered constant, but time-dependent, and thus we use  $\mathbf{n}(t)$  and  $\mathbf{v}(t)$  to denote them.

As before, we will exemplify all constraints using the model in figure 2.1:

$$\dot{\mathbf{n}}_{N_{out}}(t) = -\mathbf{v}_T(t) \quad (2.23)$$

$$\dot{\mathbf{n}}_{N_{in}}(t) = \mathbf{v}_T(t) - \mathbf{v}_M(t) \quad (2.24)$$

$$\dot{\mathbf{n}}_P(t) = \mathbf{v}_M(t) - \mathbf{v}_L(t) - \sum_{i \in \{L, M, R, T\}} \mathbf{v}_{R_i}(t) \quad (2.25)$$

$$\dot{\mathbf{n}}_Q(t) = \mathbf{v}_L(t) \quad (2.26)$$

$$\dot{\mathbf{n}}_i(t) = \mathbf{v}_{R_i}(t) \quad \forall i \in \{L, M, T, R\}. \quad (2.27)$$

The authors then note that the system displays two time-scales: metabolite amounts are changing very fast compared to extracellular concentrations and macromolecule amounts. In addition, macromolecule production reactions have much larger stoichiometric coefficients compared to metabolic reactions.

These observations allow the authors to use Tikhonov’s theorem (Khalil, 2002) and do a quasi-steady-state approximation for the dynamics of the internal

## 2.3 Dynamic enzyme-cost flux balance analysis

metabolite amounts. They do not prove that the conditions for applying Tikhonov's theorem hold, since the conditions "are hard to check in realistic networks because enzyme kinetics are not always known and because it may not be possible to solve for the steady-state of [internal metabolite amounts]" ((Waldherr et al., 2015), p. 472). Instead, as is typically the case in the constraint-based modeling, the authors assume that a stable quasi-steady-state exists based on biophysical insight, but they also note that this need not always be the case.

After applying the quasi-steady-state approximation for the internal metabolite amounts, the system becomes

$$\dot{\mathbf{n}}_{N_{out}}(t) = -\mathbf{v}_T(t) \quad (2.28)$$

$$0 = \mathbf{v}_T(t) - \mathbf{v}_M(t) \quad (2.29)$$

$$0 = \mathbf{v}_M(t) - \mathbf{v}_L(t) - \sum_{i \in \{L, M, R, T\}} \mathbf{v}_{R_i}(t) \quad (2.30)$$

$$\dot{\mathbf{n}}_Q(t) = \mathbf{v}_L(t) \quad (2.31)$$

$$\dot{\mathbf{n}}_i(t) = \mathbf{v}_{R_i}(t) \quad \forall i \in \{L, M, R, T\}. \quad (2.32)$$

### 2.3.2 Enzyme capacity and nonnegativity constraints

So far the reaction fluxes  $\mathbf{v}(t)$  in the system above are free variables. However, in reality, they depend on several quantities, among which we mention kinetic parameters, enzyme molar amount and substrate concentrations. Internal metabolite concentrations are however no longer modeled explicitly because of the quasi-steady-state approximation. The authors therefore employ only enzyme capacity constraints based on the maximum velocities, which are dependent only on enzyme amounts and turnover rates ( $k_{cat}$ ).

The capacity constraints are similar to RBA, and imposed at each point in time as

$$\mathbf{v}_i(t) \leq k_{cat}^i \mathbf{n}_i(t) \quad \forall i \in \{T, M, L\} \quad (2.33)$$

$$\sum_{i \in \{T, L, M, R\}} \frac{\mathbf{v}_{R_i}(t)}{k_{cat}^{R_i}} \leq \mathbf{n}_R(t), \quad (2.34)$$

for all  $t \geq 0$ . Please note that these upper bounds on reaction fluxes are time-dependent and change with possible increases in enzyme amounts at each time point.

It is also important to note that, if an enzyme catalyzes several reactions, as is the case of the ribosome, then the total flux through those reactions, weighted by the turnovers, is bound by the enzyme amount. Furthermore, as in the case of RBA, the bounds are typically used for the absolute values of the fluxes, but since all reactions are irreversible in our toy model, we omit this detail here.

## Chapter 2. Resource allocation formalisms

---

Last but not least, additional biomass-independent bounds are added to ensure nonnegativity of macromolecule amounts and of fluxes through irreversible reactions (in this toy model case all reactions):

$$\mathbf{n}(t), \mathbf{v}(t) \geq 0, \quad \forall t \geq 0. \quad (2.35)$$

### 2.3.3 Imposing quota production

In a similar fashion to RBA, production of quota components is imposed as a path constraint. At each time point the amount of  $Q$  is required to make up at least a certain fraction  $q$  of the total biomass, as

$$\mathbf{n}_Q(t) \geq q \sum_{i \in \{L, M, T, R, Q\}} \mathbf{n}_i(t), \quad \forall t \geq 0$$

Note that in our simple toy model we have assumed that all macromolecules have the same molecular weight. This is not the case in general, and thus in a real biological model the sum of the macromolecules above would be weighted by the molecular weights of the individual macromolecules to obtain the total biomass, as we will see later in the models in chapters 6 and 7.

### 2.3.4 Objective functions

Since deFBA does not explicitly model the growth rate, the authors explore several possible objective functions, some of which are inspired from classical FBA. One of them is maximization of the total biomass at the end of the simulation time,

$$\max \sum_{i \in \{L, M, T, R, Q\}} \mathbf{n}_i(t_f), \quad (2.36)$$

where  $t_f$  denotes the end time. In a small model the authors show that this objective function gives rise to a large variability in the optimum.

A second objective is the minimization of the final time such that at the end of the simulation all substrate has been used, i.e., the substrate should be consumed as quickly as possible,

$$\begin{aligned} \min t_f \\ \text{s.t. } \mathbf{n}_{N_{out}}(t_f) = 0. \end{aligned} \quad (2.37)$$

A third objective is the maximization of the integral of the biomass over the whole simulation time, discounted by a negative exponential, as

$$\max \int_0^{t_f} \sum_{i \in \{L, M, T, R, Q\}} \mathbf{n}_i(t) e^{-\varphi t} dt. \quad (2.38)$$

The authors show that a solution obtained in this way is unique for a small toy model.

The factor  $\varphi$  in the negative exponential is used “to reduce the effect of the terminal time  $t_f$  on the objective function, since the objective function value is uniformly bounded for varying terminal times, provided that the discount parameter is larger than the maximal growth rate”. However, we do not a priori know the growth rate, so it is unclear how one could set  $\varphi$  for an entirely new model. Moreover, we will see in chapter 5 that the choice of  $\varphi$  can be crucial for the resulting system dynamics.

### 2.3.5 The dynamic optimization problem, solving strategies and main results

Putting together all the constraints described above, we obtain a dynamic optimization problem:

$$\max_{\mathbf{n}(t), \mathbf{v}(t)} \int_0^{t_f} \sum_{i \in \{L, M, T, R, Q\}} \mathbf{n}_i(t) e^{-\varphi t} dt \quad (2.39)$$

$$\text{s.t. } \dot{\mathbf{n}}_{N_{out}}(t) = -\mathbf{v}_T(t) \quad \forall t \geq 0 \quad (2.40)$$

$$0 = \mathbf{v}_T(t) - \mathbf{v}_M(t) \quad \forall t \geq 0 \quad (2.41)$$

$$0 = \mathbf{v}_M(t) - \mathbf{v}_L(t) - \sum_{i \in \{L, M, R, T\}} \mathbf{v}_{R_i}(t) \quad \forall t \geq 0 \quad (2.42)$$

$$\dot{\mathbf{n}}_Q(t) = \mathbf{v}_L(t) \quad \forall t \geq 0 \quad (2.43)$$

$$\dot{\mathbf{n}}_i(t) = \mathbf{v}_{R_i}(t) \quad \forall i \in \{L, M, R, T\}, \forall t \geq 0 \quad (2.44)$$

$$\mathbf{v}_i(t) \leq k_{cat}^i \mathbf{n}_i(t) \quad \forall i \in \{T, M, L\}, \forall t \geq 0 \quad (2.45)$$

$$\sum_{i \in \{T, L, M, R\}} \frac{\mathbf{v}_{R_i}(t)}{k_{cat}^{R_i}} \leq \mathbf{n}_R(t) \quad \forall t \geq 0 \quad (2.46)$$

$$\mathbf{n}(t), \mathbf{v}(t) \geq 0 \quad \forall t \geq 0 \quad (2.47)$$

$$\mathbf{n}_Q(t) \geq q \sum_{i \in \{L, M, T, R, Q\}} \mathbf{n}_i(t) \quad \forall t \geq 0 \quad (2.48)$$

The authors suggest to discretize the dynamic variables  $\mathbf{n}(t), \mathbf{v}(t)$  using a Radau numerical scheme and thus cast the problem into a linear programming problem which can then be solved efficiently. We will see in chapter 5 how exactly to do this and that other discretization schemes such as the midpoint rule are possible and have the advantage of more simplicity in the implementation.

To show how the method performs, the authors then apply it to a toy model of core carbon metabolism adapted from (Covert et al., 2001). This model is shown to reproduce important biological processes such as the diauxic shift. Furthermore, the authors display the dynamically changing biomass composition that, as opposed to steady-state constraint-based models, can in this framework adapt to changes in the environment.

The deFBA method thus sets the stage for a series of computational studies that can now be performed in order to understand adaptation of organisms to dynamic environments.

### 2.4 Conditional flux balance analysis

Conditional flux balance analysis (cFBA), introduced by (Rüger et al., 2015), can be seen as a variant of deFBA, which has been developed as a dynamic resource allocation formalism to understand phototrophic growth in a periodic day-night environment.

As such, most of the ingredients of a deFBA model can also be found in the cFBA model. We will therefore not detail again the constraints, but instead explain the main difference between the two approaches.

The main assumption behind cFBA is the fact that the studied organism grows in a chemostat and that the cellular composition after a full diurnal period is invariant, that is

$$\mathbf{n}(t_f) = \alpha \cdot \mathbf{n}(0),$$

where  $\alpha$  is the amount by which the culture has grown in one period, and is itself a variable of the dynamic optimization problem.

Given this condition, the organism is assumed to maximize the multiplication factor  $\alpha$  by which it grows in a full period. We observe that this assumption introduces a quadratic term in the dynamic optimization problem. However, as  $\alpha$  is the only quadratic term, the authors perform a binary search as in RBA to obtain the maximum factor  $\alpha$ , where at each step they solve a linear program.

A second minor difference between deFBA and cFBA are the initial conditions. While as part of deFBA an RBA problem is solved to obtain the initial amounts of macromolecules, in the case of cFBA, these initial amounts are themselves variables, and an additional constraint is imposed that limits their total sum weighted by molecular weights, as in the case of constraint (2.21) in RBA. The quota fraction constraint is then only set for the initial time point and throughout the time period these quota components are required to grow with the same factor  $\alpha$  as the catalytic biomass.

Last but not least, cFBA a priori sets a step size and hence a number of time steps and then builds the linear programming problem by combining the constraints for each time point following the explicit Euler scheme. The procedure is mathematically equivalent to modeling macromolecule amounts by ODEs as in deFBA and using an explicit Euler discretization of the dynamic quantities. On the other hand, in deFBA the discretization is done using a more complicated collocation method. Because of the mathematical equivalence, in this thesis, we treat cFBA as a dynamic optimization method, just as deFBA, with an explicit Euler discretization scheme of the dynamic variables.



We note that, due to the exponential nature of the growth process described in cFBA, the use of an explicit Euler scheme may not be a very good choice, as explicit numerical methods have been shown to behave badly on numerically stiff problems (Deuffhard and Bornemann, 2013). Thus, we would recommend that a more stable numerical scheme such as implicit midpoint rule be used for such dynamic optimization problems as cFBA. We will see in chapter 5 that, even when such an implicit scheme is used, we may need to reformulate the optimization problem in order to obtain numerically stable results.

## 2.5 Conclusions

With the advent of all the methods for studying resource allocation described above, it is now possible to computationally explore several interesting biological questions.

We use cFBA to study the implications of the diurnal environment on the metabolic resource allocation in *Synechococcus elongatus* PCC 7942 in chapter 6 and we use deFBA to understand the optimal resource allocation in *Saccharomyces cerevisiae* under dynamically changing environments in chapter 7.

However, before we come to the applications, we detail in the next chapter how one constructs a metabolic resource allocation model and in chapter 5 what are the numerical difficulties of such resource allocation formalisms and how one solves such problems in practice.



## Chapter 3

# Steady states and upper bounds on growth rates from resource allocation models

The work presented in this chapter has been done in collaboration with Arne Reimers and is published in *Journal of Theoretical Biology* under (Reimers and Reimers, 2016), available at <http://dx.doi.org/10.1016/j.jtbi.2016.06.031>.

Note that, in contrast to the previous chapter where RBA and deFBA were introduced, the stoichiometric matrix here only contains the rows for the internal metabolites and metabolic and exchange reactions. Similarly, the flux vectors only contain the entries that correspond to the metabolic and exchange reactions. How enzyme concentrations change based on fluxes is not stated here because our results do not require this extension.

Before we continue the study of resource allocation models, we provide in this chapter a derivation of the steady-state assumption using flux averages over time. This helps us prove that FBA models provide upper bounds on biomass production rates obtained from deFBA models.

### 3.1 Introduction

As introduced in section 1.2.2, a key assumption for modeling metabolic networks is that production and consumption of internal metabolites must balance (steady-state assumption). This assumption lies at the core of many metabolic network analysis techniques such as FBA (Varma and Palsson, 1994; Orth et al., 2010), elementary flux mode analysis (Schuster and Hilgetag, 1994), metabolic

### Chapter 3. Steady states and upper bounds on growth rates from resource allocation models

---

control analysis (Heinrich and Schuster, 1998) or gene intervention studies (Burgard et al., 2003; Hädicke and Klamt, 2011).

Given the stoichiometric matrix  $S$  of a metabolic network, we call a vector of reaction rates (fluxes)  $\mathbf{w}$  a steady-state flux if it satisfies

$$S\mathbf{w} = 0. \quad (\text{SS})$$

In this chapter we provide a new, mathematically sound derivation of the steady-state condition using flux averages over time. This derivation does not require any underlying theory on dynamics, like oscillations, in metabolic networks. While the biological motivation of our approach (Fell, 1997; Schuster and Fell, 2007; Knoke et al., 2008; Steuer and Junker, 2009; Palsson, 2015), is well known, the mathematical foundation presented here strengthens the existing approaches that study metabolism using steady-state fluxes.

The steady-state assumption, as used in metabolic network analysis is usually mathematically derived from a quasi-steady-state perspective. This perspective is however not always applicable, as pointed out by (Song and Ramkrishna, 2009). Therefore, our mathematical derivation presented here does not use the quasi-steady-state argument. We nevertheless outline the quasi-steady-state perspective below for the sake of comparison.

#### 3.1.1 Classical derivation based on the quasi-steady-state assumption

To illustrate the differences between the existing theory and our new derivation, we first recall from section 2.3.1 how the steady-state assumption is mathematically derived in the quasi-steady-state perspective.

Given a kinetic model

$$\dot{\mathbf{c}}(t) = S\mathbf{v}(t), \quad \mathbf{v}(t) = f(\mathbf{e}(t), \mathbf{c}(t)) \quad (\text{KM1})$$

that describes the dynamics of the internal metabolite concentrations  $\mathbf{c}$ , reaction rates  $\mathbf{v}$  and enzyme concentrations  $\mathbf{e}$  given the kinetic rate laws  $f$ , we assume that the dynamics of the metabolism can be approximated by a quasi-steady-state solution with respect to the enzyme dynamics.

A quasi-steady-state solution of (KM1) is a tuple of time-dependent functions  $(\mathbf{c}, \mathbf{v}, \mathbf{e})$  such that

$$0 = S\mathbf{v}(t), \quad \mathbf{v}(t) = f(\mathbf{e}(t), \mathbf{c}(t)) \quad \text{for all } t \geq 0. \quad (\text{QSS})$$

Please note that in this model enzyme and metabolite concentrations can still change over time while fluxes transition from one metabolic steady-state to another, and are therefore not constant.

Indeed, as Varma and Palsson put it, “this assumption is based on the fact that metabolic transients are typically rapid compared to cellular growth rates and environmental changes. The consequence of this assumption is that all metabolic fluxes leading to the formation and degradation of any metabolite must balance” (Varma and Palsson (1994), p. 994). Similar reasons for assuming a quasi-steady-state for metabolism are obtained by comparing the time scale of metabolic processes (fast) to those of e.g. transcriptional regulation or cell cycle (slow) (Heinrich and Schuster, 1996; Moreira dos Santos et al., 2004; Almqvist et al., 2014). Hence, it is assumed that at every time point the metabolite concentrations have converged to a steady-state and thus the quasi-steady-state assumption (QSS) follows (Schilling et al., 1999; Voss et al., 2003; Waldherr et al., 2015).

There are, however, situations when the quasi-steady-state assumption cannot be applied (Song and Ramkrishna, 2009; Behre and Schuster, 2009). This means the derivation above cannot be used. Therefore, the main result of this chapter is a derivation, which does not need this assumption.

Before we continue with our new mathematical approach, it is worth noting the difference between the steady-states in (QSS) and the global steady-state used in classical metabolic network analysis tools such as FBA.

Given (QSS), for every time point  $t$ ,  $\mathbf{v}(t)$  is a steady-state flux. Therefore, we consider the quasi-steady-state assumption a *time-local* property. From this the steady-state condition  $S \cdot \mathbf{w} = 0$  as used in classical metabolic network analysis is derived. This simplification allows for an efficient analysis of metabolic networks, since metabolite concentrations and time do not need to be modeled anymore. For example, it is used in methods such as FBA to predict optimal biomass yields.

In FBA we use only one steady-state flux to describe the whole growth cycle. This is what we call a *time-global* steady-state flux. However, metabolic fluxes are not constant in time. For instance, during the cell cycle the cell goes through different phases ( $G_1$ ,  $S$ ,  $G_2$  and  $M$ ) during which the metabolic activity is different. Therefore, the metabolism can be considered to use different time-local steady-state fluxes that follow the division cycle. Since the sum of steady-state fluxes yields another steady-state flux, i.e., if  $S\mathbf{w} = 0$  and  $S\mathbf{v} = 0$ , then  $S(\mathbf{w} + \mathbf{v}) = 0$ , by combining the time-local steady-state fluxes we can obtain a time-global steady-state flux for the whole growth cycle.

#### 3.1.2 The perspective based on long time periods

However, we do not need time-local steady-states to obtain a time-global steady-state. For example the steady-state assumption is also often motivated by stating that no metabolite can accumulate or deplete on the long run (Fell, 1997). We provide here a general mathematical framework based on this idea.

### Chapter 3. Steady states and upper bounds on growth rates from resource allocation models

In particular, we will generalize the following approach used in (Steuer and Junker, 2009; Knoke et al., 2008, 2010). They observe that, if after a time  $T$  no net change  $\Delta \mathbf{c}(T) = 0$  has occurred in the metabolite concentrations, we obtain  $S \int_0^T \mathbf{v}(t) dt = 0$ . Hence, in this case, the average flux

$$\tilde{\mathbf{v}}(T) := \frac{1}{T} \int_0^T \mathbf{v}(t) dt \quad (\text{AVGV})$$

is also a steady-state flux. In contrast to the fluxes derived via the quasi-steady-state assumption, it applies globally over the time interval  $[0, T]$ . In particular, in cases where the quasi-steady-state assumption is not entirely justified (Song and Ramkrishna, 2009), one can still obtain a time-global steady-state.

Building upon the ideas in section 1.5.2 of (Steuer and Junker, 2009), we observe that, if we consider a long enough time period  $T$ , we do not necessarily need to come back to the same concentration, but in order to obtain an average steady-state flux we only require that the concentrations stay bounded (see figure 3.1). While this is implied by physical laws, it should also happen because accumulation of metabolites in very high amounts is toxic for a cell. Therefore, on the long run, to avoid such toxicity, every metabolite should be produced, on average, at the same rate at which it is consumed (Fell, 1997).

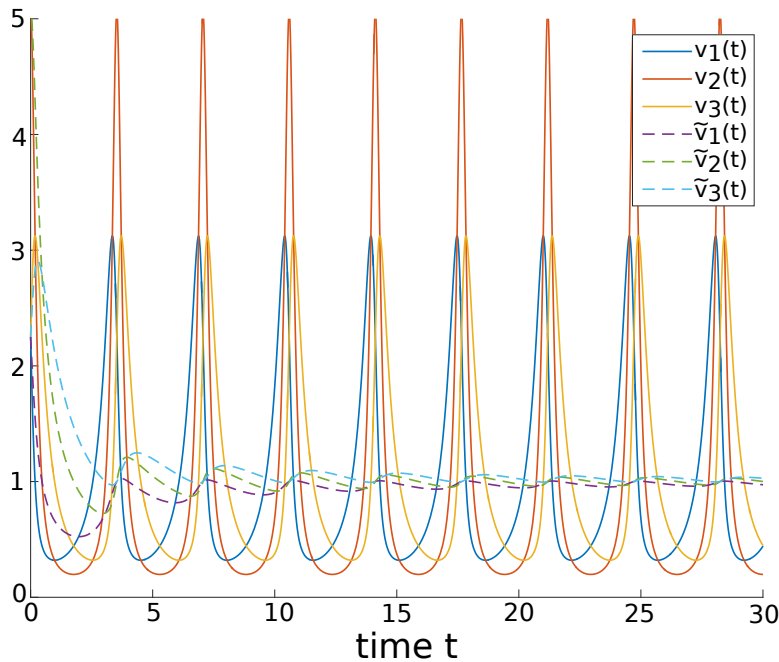


Figure 3.1: Fluxes  $v_i$  and average fluxes  $\tilde{v}_i$  for the example system discussed in section 3.4. While the fluxes continue oscillating indefinitely, the average fluxes converge to a steady-state.

Based on these observations, we present in section 3.2 a mathematical perspective on the steady-state assumption that does not need the quasi-steady-state

## 3.2 The steady-state assumption for long time periods

---

argument, but instead considers flux averages over time. Using this model we compute for how long we have to observe the system to obtain a sufficiently good steady-state distribution on the example of three model organisms.

The mathematical framework presented here is not only another justification for the steady-state assumption, but can also be used to mathematically show when FBA gives upper bounds on yield and growth rate as shown in section 3.3.

However, there are also some caveats when dealing with the steady-state assumption for long time periods. In section 3.4 we present a simple, artificial mass-action system, where the constraints implied by kinetic rate laws are violated by the average concentrations and fluxes. We conclude by posing the question whether a metabolic system can be more efficient by utilizing oscillations than with simple steady-state fluxes.

### 3.2 The steady-state assumption for long time periods

Since no metabolite can accumulate or deplete indefinitely, it follows intuitively that production and consumption of all metabolites must balance. We will now formulate this argument mathematically.

#### 3.2.1 Modeling assumptions

Our result applies to a very general setting. We essentially only ignore stochastic effects and thus require the following modeling assumptions:

- We assume that the volume stays constant and changes in concentrations are only reaction-driven i.e., we do not yet consider dilution of metabolites due to cell growth. The case when the volume can change is discussed in section 5.3. While enzyme concentrations can be varied arbitrarily (e.g. due to regulatory control), metabolite concentrations and fluxes have to satisfy the following relationship, mentioned already in the introduction, for every  $t \geq 0$ :

$$\dot{\mathbf{c}}(t) = \mathbf{S}\mathbf{v}(t), \quad \mathbf{v}(t) = f(\mathbf{e}(t), \mathbf{c}(t)). \quad (\text{KM1})$$

- Concentrations are measured in number of molecules per fixed volume. Hence, enzyme and metabolite concentrations are bounded, i.e., there exist  $c^{\max}$  and  $e^{\max}$  with  $\|\mathbf{c}(t)\| \leq c^{\max}$ ,  $\|\mathbf{e}(t)\| \leq e^{\max}$  for every time  $t \geq 0$ , where  $\|\mathbf{x}\|$  denotes the Euclidean norm of the vector  $\mathbf{x}$ .
- We assume that the function  $f$  that represents the kinetic rate laws is continuous.
- We assume that  $\mathbf{c}(t)$  is differentiable and  $\mathbf{e}(t)$  is a continuous function of time.

### Chapter 3. Steady states and upper bounds on growth rates from resource allocation models

---

Since  $f$  is continuous, and  $\mathbf{c}(t)$  and  $\mathbf{e}(t)$  are bounded and continuous, it follows that  $\mathbf{v}(t)$  must also be bounded and continuous (Rudin, 1976).

#### 3.2.2 Average fluxes

For a given time period  $T$ , we define the average fluxes  $\tilde{\mathbf{v}}$ , as introduced above, as:

$$\tilde{\mathbf{v}}(T) := \frac{1}{T} \int_0^T \mathbf{v}(t) dt. \quad (\text{AVGV})$$

To mathematically analyze long time periods, we consider the case when  $T \rightarrow \infty$ . Unfortunately, it can happen that

$$\bar{\mathbf{v}} := \lim_{T \rightarrow \infty} \tilde{\mathbf{v}}(T) \quad (3.1)$$

does not exist (see Appendix B in (Reimers and Reimers, 2016) for an example). For simplicity, we assume in the following that the limit exists. Even in the case when the limit does not exist, the results hold in a similar fashion as we have shown in Appendix A of (Reimers and Reimers, 2016).

In (Reimers and Reimers, 2016) we have proved that average fluxes are steady-state fluxes:

**Theorem 3.2.1**  $S\bar{\mathbf{v}} = 0$ .

PROOF Since the concentrations are non-negative and bounded, i.e.,  $0 \leq \mathbf{c}(t)$  and  $\|\mathbf{c}(t)\| \leq c^{\max}$  for all  $t \geq 0$ , it follows by the main theorem of integration that

$$\left\| \int_0^T \dot{\mathbf{c}}(t) dt \right\| = \|\mathbf{c}(T) - \mathbf{c}(0)\| \leq c^{\max} \quad (3.2)$$

$$\Rightarrow \left\| \frac{1}{T} \int_0^T \dot{\mathbf{c}}(t) dt \right\| \leq \frac{c^{\max}}{T} \quad (3.3)$$

$$\Rightarrow \lim_{T \rightarrow \infty} \left\| \frac{1}{T} \int_0^T \dot{\mathbf{c}}(t) dt \right\| \leq \lim_{T \rightarrow \infty} \frac{c^{\max}}{T} = 0. \quad (3.4)$$

Since the norm is only zero for the zero point, it follows that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \dot{\mathbf{c}}(t) dt = 0. \quad (3.5)$$

The rest of the proof follows by the definitions (KM1) and (AVGV). ■



## 3.2 The steady-state assumption for long time periods

### 3.2.3 Violation of the steady-state condition for finite time T

For obtaining the statement of theorem 3.2.1 we have assumed that  $T \rightarrow \infty$ . However, in practice we do not run the experiments for infinitely long time. We are therefore interested in how large do we have to choose  $T$  so that the fluxes violate the steady-state condition by at most  $\varepsilon$ .

We observe that with  $T := \frac{c^{\max}}{\varepsilon}$  we get

$$\|S\bar{\mathbf{v}}(T)\| = \left\| \frac{1}{T} \int_0^T S\mathbf{v}(t) dt \right\| = \left\| \frac{1}{T} \int_0^T \dot{\mathbf{c}}(t) dt \right\| = \frac{\|\mathbf{c}(T) - \mathbf{c}(0)\|}{T} \leq \frac{c^{\max}}{T} = \varepsilon. \quad (3.6)$$

We consider three organisms: *Escherichia coli*, *Saccharomyces cerevisiae*, and *Homo sapiens* (HeLa cells) and compute in each case the averaging time  $T$  so that we obtain a relative violation of the steady-state condition of at most 1% in the approximation of fluxes.

### 3.2.4 Escherichia coli

For *E. coli* the average glucose uptake flux is  $1.63 \cdot 10^{-18}$  mol/(s · cell) (Jain and Srivastava, 2009; Loferer-Krößbacher et al., 1998). Since we would like to have a violation of at most 1%, our  $\varepsilon$  is then  $0.01 \cdot 10^{-18} = 10^{-20}$  mol/(s · cell). The maximum metabolite concentration measured in this organism is 96 mM (Bennett et al., 2009). We will therefore consider  $c^{\max}$  as

$$\begin{aligned} c^{\max} &= 100 \text{ mmol/L} \\ &= 10^{-1} \text{ mol/L} \cdot 0.6 \cdot 10^{-15} \text{ L/cell} \\ &= 0.6 \cdot 10^{-16} \text{ mol/cell,} \end{aligned}$$

where  $1 \text{ cell} = 0.6 \cdot 10^{-15} \text{ L}$  is the volume of an *E. coli* cell (Kubitschek, 1990). Therefore, if we average the fluxes of this organism over a period  $T = \frac{c^{\max}}{\varepsilon} = 6000$  s  $\approx 2$  h the steady-state condition will be violated by at most 1%.

Note that this means that we would have to average over six to eight generations in the case of very fast growing *E. coli*. This is reasonable considering the fact that we need to average out fluctuations arising from the cell cycle.

### 3.2.5 Saccharomyces cerevisiae

In the case of *S. cerevisiae*, the average intracellular fluxes are around  $1.38 \cdot 10^{-18}$  mol/(s · cell) (Stewart et al., 2010; Mitchison, 1958). Our  $\varepsilon$  in this case is therefore again  $10^{-20}$  mol/(s · cell). According to (Canelas et al., 2008) and (Finka and Goloubinoff, 2013), we can choose  $c^{\max} = 10^{-16}$  mol/cell. Thus, the minimum time period for averaging so that we violate the steady-state condition by at most 1% is  $T = 10000$  s  $\approx 3$  h.

## Chapter 3. Steady states and upper bounds on growth rates from resource allocation models

---

### 3.2.6 Homo sapiens (HeLa cells)

Finally, HeLa cells have a glucose uptake flux of about  $4.5 \cdot 10^{-17}$  mol/(s · cell) (Mojena et al., 1985), and thus we choose  $\varepsilon = 10^{-19}$  mol/(s · cell). By the findings of (Mojena et al., 1985), we can choose  $c^{\max} = 10^{-15}$  mol/cell. Thus, in this case  $T = 10^4$  s  $\approx$  3 h.

## 3.3 Applications to yield optimization

We have seen that, given a kinetic model, the average fluxes satisfy the steady-state assumption. This also applies to optimal control problems, such as those resulting deFBA and cFBA. However, since we know that the average flux is a steady-state flux we can build a much simpler FBA model to bound the results of the optimal control problem. Let us consider the optimum of the following FBA problem:

$$\begin{aligned} v_{FBA}^* &:= \max_{\mathbf{v}} \mathbf{v}_{biomass} \\ \text{s.t. } \mathbf{S}\mathbf{v} &= 0 \\ \mathbf{l} &\leq \mathbf{v} \leq \mathbf{u}, \end{aligned} \quad (\text{FBA1})$$

where  $\mathbf{v}_{biomass}$  is the flux through the biomass reaction. Consider in addition the optimum of the following optimal control problem formulated based on the kinetic model (KM1):

$$\begin{aligned} \bar{v}^* &:= \max_{\mathbf{v}, \mathbf{c}, \mathbf{e}} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{v}_{biomass}(t) dt \\ \text{s.t. } \dot{\mathbf{c}}(t) &= \mathbf{S}\mathbf{v}(t) & \forall t \geq 0 \\ \mathbf{v}(t) &= f(\mathbf{e}(t), \mathbf{c}(t)) & \forall t \geq 0 \\ \mathbf{l} &\leq \mathbf{v}(t) \leq \mathbf{u} & \forall t \geq 0 \end{aligned} \quad (\text{OCP1})$$

Note that, since the change of enzyme amounts in deFBA is given only by the product of  $\mathbf{S}$  and  $\mathbf{v}$ , maximizing the integral of enzyme production fluxes above is the same as maximizing the integral of the enzyme amounts. Hence, our objective in (OCP1) is the same as objective (2.38) in section 2.3.4.

Furthermore, we note that the existence of the limit in the objective of (OCP1) is an implicit constraint of the optimization problem.

We observe that the FBA optimum is an upper bound for all steady-state solutions and hence also for average fluxes:

**Corollary 3.3.1**  $v_{FBA}^* \geq \bar{v}^*$ .

PROOF We recall from (3.1) that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{v}_{biomass}(t) dt = \bar{\mathbf{v}}_{biomass}. \quad (3.7)$$

If the limit  $\bar{\mathbf{v}}$  in (3.1) exists, then by theorem 3.2.1,  $\bar{\mathbf{v}}$  satisfies  $S\bar{\mathbf{v}} = 0$ . It is also easy to see that  $\bar{\mathbf{v}}$  also satisfies  $\mathbf{l} \leq \bar{\mathbf{v}} \leq \mathbf{u}$ . Hence  $\bar{\mathbf{v}}$  is a feasible solution of (FBA1).

If  $\bar{\mathbf{v}}$  does not exist, we instead can use theorem 3 in Appendix A of (Reimers and Reimers, 2016) and any accumulation point will be a feasible solution of (FBA1). ■

We observe that (OCP1) is more general than a typical deFBA problem, since it allows for changes in metabolite concentrations. Furthermore, we also note that we can add additional constraints to (OCP1) that link  $\mathbf{e}(t)$  with  $\mathbf{v}(t)$  and the solution of (OCP1) will still be a feasible solution of (FBA1).

We will see in section 5.3 that we can prove a similar result for RBA and cFBA, where dilution by growth is also included in the model.

### 3.4 Kinetic constraints

In the case of resource allocation models such as the one in (Molenaar et al., 2009) that we described in section 2.1, additional constraints next to the steady-state condition are employed. For example, assume we want to use the actual kinetic rate laws encoded by  $f$  to also constrain the average steady-state solution by the average enzyme and substrate concentrations. Our kinetically constrained steady-state model will then have the form:

$$S\mathbf{w} = 0, \quad \mathbf{w} = f(\mathbf{e}, \mathbf{c}) \quad \mathbf{e}, \mathbf{c} \geq 0 \quad (\text{KSS})$$

Can the results from the previous section also be applied to this model?

Assume we have measured average fluxes  $\bar{\mathbf{v}}$  and enzyme concentrations  $\bar{\mathbf{e}}$  in an experiment (or from a simulation of the dynamic model). In the previous sections we have found that  $S\bar{\mathbf{v}} = 0$ . Can we also always find concentrations  $\hat{\mathbf{c}}$  such that we get a feasible solution to the kinetically constrained steady-state model (KSS)? If we cannot, then kinetically constrained steady-state models may be overconstrained. The answer is not easy, since in the next subsection we will observe that  $\hat{\mathbf{c}} = \bar{\mathbf{c}}$  does not always give a feasible solution.

To formulate the problem we define the average enzyme concentrations

$$\bar{\mathbf{e}}(T) := \frac{1}{T} \int_0^T \mathbf{e}(t) dt. \quad (\text{AVGE})$$

Again, the average enzyme concentrations might not exist. But for simplicity, we assume here that

$$\bar{\mathbf{e}}(T) := \lim_{T \rightarrow \infty} \bar{\mathbf{e}}(T) \quad (3.8)$$

exists. For the general case we refer the reader to Appendix A of (Reimers and Reimers, 2016).

We can now formulate the problem as:

### Chapter 3. Steady states and upper bounds on growth rates from resource allocation models

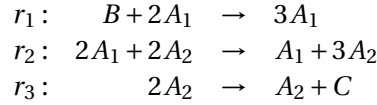
---

**Problem 3.4.1** Does there always exist a  $\hat{\mathbf{c}} \in \mathbb{R}_{\geq 0}^{\mathcal{M}}$  such that  $\bar{\mathbf{v}} = f(\bar{\mathbf{e}}, \hat{\mathbf{c}})$ ?

In the following subsection we illustrate the difficulties posed by problem 3.4.1 using a toy example.

#### 3.4.1 Average concentrations can be inconsistent with average fluxes

We consider the following toy metabolic network.



For simplicity we assume that the system is subject to mass-action kinetics and enzyme concentrations have no effect.  $B$  and  $C$  are boundary metabolites and are kept at a constant concentration of 1. Considering all kinetic constants to be 1, we get the following system of ordinary differential equations:

$$\begin{aligned} \dot{\mathbf{c}}_1(t) &= \mathbf{v}_1(t) - \mathbf{v}_2(t) & \mathbf{v}_1(t) &= \mathbf{c}_1(t)^2 \\ \dot{\mathbf{c}}_2(t) &= \mathbf{v}_2(t) - \mathbf{v}_3(t) & \mathbf{v}_2(t) &= \mathbf{c}_1(t)^2 \mathbf{c}_2(t)^2 \\ & & \mathbf{v}_3(t) &= \mathbf{c}_2(t)^2, \end{aligned} \quad (\text{Ex1})$$

where  $\mathbf{c}_1(t)$  and  $\mathbf{c}_2(t)$  denote the concentrations of metabolites  $A_1$  and  $A_2$  respectively, and  $\mathbf{v}_1(t), \mathbf{v}_2(t), \mathbf{v}_3(t)$  denote the fluxes through  $r_1, r_2, r_3$ .

The only steady-state solutions of this system are  $\mathbf{c}_1 = \mathbf{0} = \mathbf{c}_2$  and  $\mathbf{c}_1 = 1 = \mathbf{c}_2$ . If we do not start in such a steady-state, the system will oscillate. This can be seen as follows. The function

$$H(\mathbf{x}) = \mathbf{x}_1 + \frac{1}{\mathbf{x}_1} + \mathbf{x}_2 + \frac{1}{\mathbf{x}_2}, \quad (3.9)$$

where  $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^\top \in \mathbb{R}^2$ , is a Hamiltonian of the considered ODE system, since the derivative  $\frac{d}{dt} H(\mathbf{c}(t))$  is zero for all  $t \geq 0$ . We observe that the system cannot explode since for any  $(\mathbf{c}_1, \mathbf{c}_2) > 0$  the Hamiltonian has a finite constant value and hence both  $\mathbf{c}_1(t)$  and  $\mathbf{c}_2(t)$  stay bounded for all  $t \geq 0$ . Furthermore, we observe that, for any other starting point that is not a steady-state, the Hamiltonian has a value different from 4, which is the minimum achieved at  $\mathbf{c}_1 = \mathbf{c}_2 = 1$ .

The trajectory of this system (obtained from numerical integration) with the starting point  $(\mathbf{c}_1, \mathbf{c}_2) = (2, 2)$  is shown in figure 3.2. As average concentrations and fluxes we approximated numerically:

$$\bar{\mathbf{c}} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{c}(t) dt \approx (0.82, 0.82) \quad (3.10)$$

$$\bar{\mathbf{v}} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{v}(t) dt \approx (1.00, 1.00, 1.00) \quad (3.11)$$

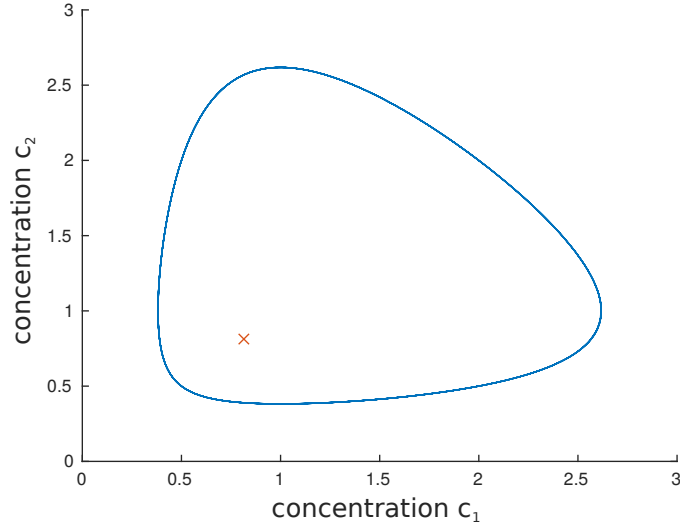


Figure 3.2: Trajectory of our toy model with starting point  $\mathbf{c}_1(0) = 2, \mathbf{c}_2(0) = 2$ . The cross marks the average concentration for  $T \rightarrow \infty$ .

We therefore conclude that in the toy example, the average concentrations are not compatible with the average fluxes i.e.,  $\bar{\mathbf{v}} \neq f(\bar{\mathbf{c}})$  where  $f$  denotes the kinetic rate laws of the toy system. In particular, the average concentrations do not even correspond to any steady-state flux distribution. This has also been observed and mathematically analyzed by (Knoke et al., 2010) for oscillations of  $\text{Ca}^{2+}$  in non-excitable cells using Jensen's inequality (Jensen, 1906).

Please note that problem 3.4.1 remains open since there exists a concentration vector  $(\mathbf{c}_1, \mathbf{c}_2) = (1, 1)$  in the toy example that is consistent with the average fluxes.

### 3.4.2 Linear kinetic constraints remain consistent

Because of the problem described above, we consider kinetic constraints as used in RBA, deFBA and cFBA as

$$\mathbf{v}_i(t) \leq k_{cat}^+ \mathbf{e}_i(t). \quad (\text{ubMM})$$

For reversible reactions, we get the additional bound

$$\mathbf{v}_i(t) \geq -k_{cat}^- \mathbf{e}_i(t), \quad (\text{lbMM})$$

where  $k_{cat}^+, k_{cat}^-$  are the turnover rates for forward and the reverse direction of the reaction respectively.

Therefore, we assume in the following that for a set of reactions  $\mathcal{K} \subseteq \mathcal{R}$  we have constants  $\mathbf{k}^+, \mathbf{k}^-$  such that

$$-\mathbf{k}_i^- \mathbf{e}_i(t) \leq \mathbf{v}_i(t) \leq \mathbf{k}_i^+ \mathbf{e}_i(t) \quad \text{for all } t \geq 0 \text{ and } i \in \mathcal{K}. \quad (3.12)$$

### Chapter 3. Steady states and upper bounds on growth rates from resource allocation models

---

This form gives us enough flexibility to also constrain the average fluxes using the average enzyme concentrations.

Assuming the average enzyme concentrations  $\bar{\mathbf{e}}$  exist, for the linear kinetic constraints as used in section 3.4.2, we get the following result:

**Proposition 3.4.2** *Assume  $\bar{\mathbf{v}}$ ,  $\bar{\mathbf{e}}$  exist. Then it holds that*

$$-\mathbf{k}_i^- \bar{\mathbf{e}}_i \leq \bar{\mathbf{v}}_i \leq \mathbf{k}_i^+ \bar{\mathbf{e}}_i \quad \text{for all } i \in \mathcal{K}. \quad (3.13)$$

PROOF For the first inequality we have

$$\mathbf{k}_i^+ \bar{\mathbf{e}}_i - \bar{\mathbf{v}}_i = \lim_{T \rightarrow \infty} \mathbf{k}_i^+ \bar{\mathbf{e}}_i(T) - \bar{\mathbf{v}}_i(T) \quad (3.14)$$

$$= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{k}_i^+ \mathbf{e}_i(t) - \mathbf{v}_i(t) dt \geq 0, \quad (3.15)$$

because  $\mathbf{v}_i(t) \leq \mathbf{k}_i^+ \mathbf{e}_i(t)$  for all  $t \geq 0$ . Thus,  $\bar{\mathbf{v}}_i \leq \mathbf{k}_i^+ \bar{\mathbf{e}}_i$ . The same argument also applies to show that  $-\mathbf{k}_i^- \bar{\mathbf{e}}_i \leq \bar{\mathbf{v}}_i$ . ■

The case when  $\bar{\mathbf{v}}$ ,  $\bar{\mathbf{e}}$  do not exist is discussed in (Reimers and Reimers, 2016). This result allows us to include enzyme capacity constraints of the type used in RBA, deFBA and cFBA into (FBA1).

## 3.5 Conclusions

One of the main arguments against using the steady-state assumption in models of biochemical reaction networks is that, if one assumes steady-state, oscillations that are biologically important will not be observed in the simulation results (Goldbeter, 1997; Sowa et al., 2014; Papagiannakis et al., 2017).

### 3.5.1 Average fluxes satisfy the steady-state assumption

However, in many cases we might not be interested in these oscillations, because they increase the complexity of the model, or even make it computationally intractable. In these cases, where only average fluxes over long time periods are of interest (e.g. if we are interested in predicting the lethality of gene knockout experiments), we have shown that the steady-state assumption can still be applied, i.e., we still get a reasonably good description of the metabolic system by computing a steady-state flux. In particular, it is also valid for oscillating systems that are not at steady-state at any point in time. The only condition is that the system is averaged over a long enough time period.

For example, FBA computes an upper bound on the biomass yield, which applies to all steady-state solutions. Hence it also applies to the average fluxes. Therefore, the system cannot obtain a higher yield using oscillations.

We showed that an average over three hours is sufficient to obtain fluxes that only slightly violate the steady-state assumption for *E. coli*, *S. cerevisiae*, and HeLa cells. Since the estimate was rather pessimistic, much shorter averaging times might be sufficient in practice. Furthermore, the violation we obtained lies within the error range of current measurement technology for concentrations and fluxes.

### 3.5.2 Pitfalls of averaging

We have seen that, if only average fluxes over long time periods are of interest, the steady-state assumption, is clearly a good model. While this adds another argument why methods like FBA can indeed predict the growth rates of some organisms accurately (Edwards et al., 2001; Harcombe et al., 2013), the integration of nonlinear constraints should be done with care.

For instance, average concentrations might be inconsistent with the average fluxes. Thus, we cannot exclude that there exists a chemical reaction system (candidates are described in Knoke et al. (2010); Sowa et al. (2014); Gottstein et al. (2014)) where an oscillation can induce a higher average flux than the flux that would be possible by assuming steady-state. We have, however, shown that linear constraints, such as those imposed by enzyme availability and enzyme capacity, do not introduce inconsistencies for average concentrations and fluxes.

Since these linear kinetic constraints remain consistent for average concentrations and fluxes, we could show using averaging why FBA provides upper bounds on biomass yields obtained from resource allocation models. In addition, we will see in section 5.3 how, in a similar fashion, RBA provides upper bounds on growth rates from cFBA models. This is a very useful result when debugging errors in the resource allocation model construction. We now turn our attention in the next chapter to how one constructs such resource allocation models.





## Chapter 4

# Building and encoding a metabolic resource allocation model

The section on SBML encoding presented in this chapter has been done in collaboration with Henning Lindhorst within the European project ROBUSTYEAST. The work presented in this chapter has been published in the journal *Metabolites* under (Reimers et al., 2017b) and can be found at <http://dx.doi.org/10.3390/metabo7030047>. Furthermore, the developed SBML standard has been made available on Fairdomhub as a standard operating procedure (SOP) at <https://www.fairdomhub.org/sops/304>.

In this chapter we present a guide for generating a complete (genome-scale) metabolic resource allocation model, as well as a proposal for how to represent such models in Systems Biology Markup Language (SBML) format. These models lie at the core of resource allocation studies in metabolic networks. To the best of our knowledge, no guidelines for generating such a model have been published up to now, although the idea of metabolic resource allocation studies has been present in the field of systems biology for several years.

Therefore, we focus in this chapter on a step-by-step guide towards constructing such a model, summarized in figure 4.1, with a focus on the deFBA formalism which is described in chapter 2. We detail here all the necessary information as well as which databases may be used to retrieve it (see table 4.1). To facilitate exchange among researchers, we furthermore propose a standard for specifying such models in SBML format using the Flux Balance Constraints extension (Olivier and Bergmann, 2015). Last but not least, we provide some guidelines for diagnosing possible problems or errors that may arise during the model building process.

## 4.1 Model prerequisites

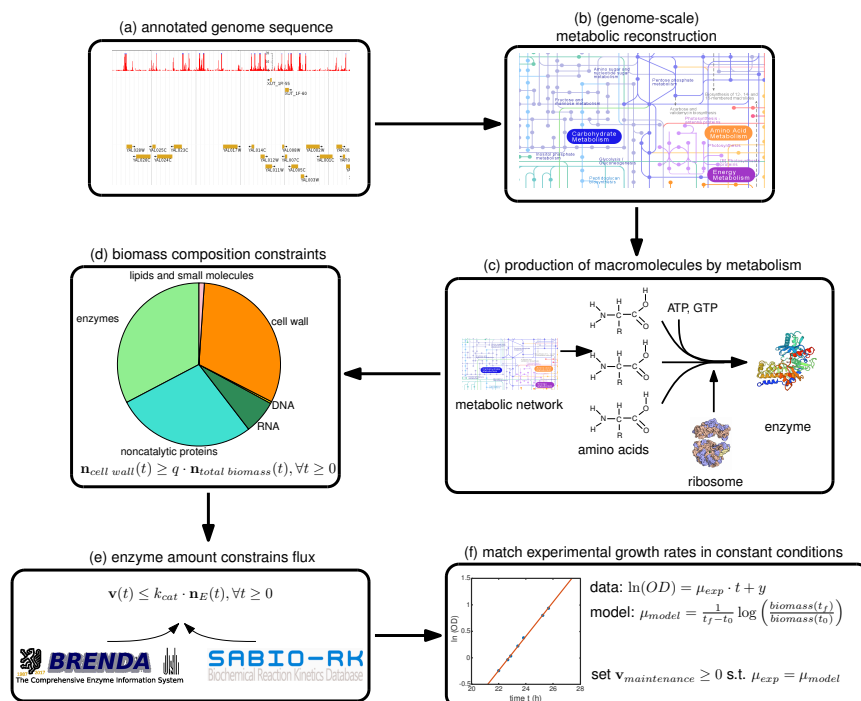


Figure 4.1: **Roadmap for generating a deFBA model.** From an annotated genome sequence (a) of the organism of interest the metabolic network (b) is reconstructed following instructions in (Thiele and Palsson, 2010). Given the gene-reaction mapping and the annotated genome sequence, the enzymes and ribosomes (c), and their synthesis reactions are added to the stoichiometric matrix (see Section 4.2). Next, the biomass composition constraints (d) should be set up using information from the biomass objective function of the metabolic network model (see Section 4.3). Then reaction turnover rates (e) sourced from literature and online databases should be added (see Section 4.4). Lastly but most importantly, the deFBA model should be fine tuned to match experimental growth rates (f) obtained in the laboratory (see Section 4.5). Images retrieved from: (a) <http://goo.gl/aBNfPz> (Skinner et al., 2009; Cherry et al., 2012); (b) [http://www.genome.jp/kegg-bin/show\\_pathway?map01100](http://www.genome.jp/kegg-bin/show_pathway?map01100) (Kanehisa and Goto, 2000); (c) [http://www.genome.jp/kegg-bin/show\\_pathway?map01100](http://www.genome.jp/kegg-bin/show_pathway?map01100) (Kanehisa and Goto, 2000), <http://pdb101.rcsb.org/motm/10> (Berman et al., 2000), <https://swissmodel.expasy.org/repository/uniprot/P04806> (Kiefer et al., 2008).

Table 4.1: Databases where the necessary information needed to build a metabolic resource allocation model can be found.

Resource	Link	Reference
<b>Annotated genome sequences</b>		
Genbank	<a href="https://www.ncbi.nlm.nih.gov/genbank/">https://www.ncbi.nlm.nih.gov/genbank/</a>	Benson et al. (2013)
UniProtKB	<a href="http://www.uniprot.org/">http://www.uniprot.org/</a>	UniProt Consortium (2014)
<b>Metabolic network reconstructions</b>		
BiGG	<a href="http://bigg.ucsd.edu/">http://bigg.ucsd.edu/</a>	King et al. (2016)
BioModels	<a href="https://www.ebi.ac.uk/biomodels-main/">https://www.ebi.ac.uk/biomodels-main/</a>	Juty et al. (2015)
ModelSEED	<a href="http://modelseed.org/">http://modelseed.org/</a>	Devoid et al. (2013)
KEGG	<a href="http://www.genome.jp/kegg/pathway.html">http://www.genome.jp/kegg/pathway.html</a>	Kanehisa and Goto (2000)
Pathway Tools	<a href="http://bioinformatics.ai.sri.com/ptools/">http://bioinformatics.ai.sri.com/ptools/</a>	Karp et al. (2016)
<b>Enzyme subunit stoichiometry</b>		
UniProtKB	<a href="http://www.uniprot.org/">http://www.uniprot.org/</a>	UniProt Consortium (2014)
<b>Ribosome composition</b>		
Ribosomal Protein Gene Database	<a href="http://ribosome.med.miyazaki-u.ac.jp/">http://ribosome.med.miyazaki-u.ac.jp/</a>	Nakao et al. (2004)
KEGG	<a href="http://www.genome.jp/kegg/">http://www.genome.jp/kegg/</a>	Kanehisa and Goto (2000)
<b>Quantitative proteomics datasets</b>		
MaxQuant	<a href="http://maxqb.biochem.mpg.de/mxldb/">http://maxqb.biochem.mpg.de/mxldb/</a>	Schaab et al. (2012)
Proteomaps	<a href="https://www.proteomaps.net/index.html">https://www.proteomaps.net/index.html</a>	Liebermeister et al. (2014)
<b>Turnover rates</b>		
BRENDA	<a href="http://www.BRENDA-enzymes.org/">http://www.BRENDA-enzymes.org/</a>	Schomburg et al. (2013)
SABIO-RK	<a href="http://sabio.villa-bosch.de/">http://sabio.villa-bosch.de/</a>	Wittig et al. (2012)

## **Chapter 4. Building and encoding a metabolic resource allocation model**

---

The most important prerequisite for building a metabolic resource allocation model is a high quality (genome-scale) metabolic reconstruction of the organism of interest. As we will see in the following sections, it is very important that key-elements of this reconstruction such as the gene-reaction mapping or the biomass composition and coefficients are as accurate as possible.

A description of how exactly to come up with such a genome-scale metabolic reconstruction is out of the scope of this chapter. To date, more than 2600 functional draft reconstructions have been generated (Büchel et al., 2013) and many of them can be retrieved from online databases such as BioModels (Le Novère et al., 2006; Li et al., 2010; Chelliah et al., 2015; Juty et al., 2015). If there exists no genome-scale metabolic reconstruction for the organism of interest, but the full genome sequence of the organism is available, the protocol of (Thiele and Palsson, 2010) can be followed in order to generate the metabolic network reconstruction.

Note that, depending on the organism, it may not be possible to simulate a complete genome-scale deFBA model due to the size of the resulting linear program. While RBA allows simulation of steady-state resource allocation in genome-scale networks (Goelzer and Fromion, 2017), the dynamic approaches like deFBA are constrained to smaller sizes. Networks with up to 500 metabolic reactions can be successfully simulated with deFBA as demonstrated in (Reimers et al., 2017a). If the starting metabolic network is too large, tools such as the minimal network finder in (Röhl and Bockmayr, 2017) or the procedures in (Erdrich et al., 2015; Ataman et al., 2017) can be used to reduce the size of the network while at the same time keeping desired functionalities.

Another key prerequisite of a deFBA model are the annotated amino acid sequences of all the genes present in the model, as we will explain next.

### **4.2 Building the protein production reactions**

Any good quality genome-scale metabolic reconstruction comes with a gene-reaction mapping. Such a mapping describes exactly what genes are involved in the catalysis of each reaction. In addition, it also offers information about isoenzymes, i.e., enzymes that differ in amino acid sequence but catalyze the same reaction. In the following subsections we will explain how to make use of the gene-reaction mapping in order to construct the protein production reactions for a deFBA model.

#### **4.2.1 The case of enzymes encoded by one gene only**

To build any protein production reaction using the gene-reaction mapping the key ingredient is a database of all genes and their corresponding amino acid sequences. This can be obtained as a FASTA file from online databases such as Genbank (Benson et al., 2013) or UniProt (UniProt Consortium, 2017). A

## 4.2 Building the protein production reactions

---

FASTA file is a file formatted such that it represents either nucleotide or peptide sequences using single-letter codes. The advantage of using UniProt is that, through the Java API (<http://www.ebi.ac.uk/uniprot/japi/usage.html>), one can automatically access the sequences, as well as information about the Enzyme Commission number (EC number) or subunit stoichiometry for enzymes. There also exist organism-specific databases where this information can be retrieved. To give some examples, in the case of *Saccharomyces cerevisiae* one could obtain such a FASTA file also from the Saccharomyces Genome Database (Cherry et al., 2012), while for cyanobacteria one could use Cyanobase (Nakao et al., 2010).

To then build the enzyme production reaction for an enzyme encoded by one gene only, we look up the corresponding gene in the FASTA file, compute its amino acid count, and set the amino acids with their respective counts as reactants for the production reaction. Additional reactants are then the energy cofactors needed to grow the peptide chain: per amino acid added, one ATP is hydrolyzed into AMP and PP<sub>i</sub>, and two GTP molecules are hydrolyzed into GDP and P<sub>i</sub>. Last but not least, for the translation initiation a molecule of 10-formyltetrahydrofolate (fTHF) is converted into a molecule of tetrahydrofolate (THF) (Nelson et al., 2008).

Another important factor that comes into play is then also whether the enzyme is a monomer, i.e., if only one copy of the corresponding gene product is needed to build the enzyme. If this is not the case, then the amount of copies of the gene product required to produce the enzyme has to be accounted for in the production reaction. This information can be often retrieved from the UniProt database. As an example, if the enzyme is a homotrimer, i.e., three copies of the gene product are needed to build it, then the stoichiometries of the amino acids and the energy cofactors in the production reaction have to be multiplied by three. The importance of taking into account such information can be seen in the cost of producing such an enzyme. Not taking into account such extra costs can result in wrong predictions in the use of alternative pathways.

### 4.2.2 The case of isoenzymes

Isoenzymes usually arise as a result of partial genome duplication and subsequent point mutations or insertion/deletion events in the course of evolution. They usually have different kinetic properties and are subject to different regulatory influences. Isoenzymes are important features of metabolism that allow fine-tuning of reactions rates in a way that satisfies the exact needs of the organism at different stages of development or of the cell cycle. An example of such pairs of enzymes are glucokinase and hexokinase, which both catalyze the conversion of a hexose into a hexose 6-phosphate. While hexokinase is inhibited by glucose 6-phosphate, glucokinase is not. This different regulatory feature, together with its lower affinity for glucose, allows glucokinase to play different

## Chapter 4. Building and encoding a metabolic resource allocation model

---

roles in cells of specific organs.

Isoenzymes also play a special role in a deFBA model. To see this, let us assume we have two enzymes  $e_1$  and  $e_2$  that catalyze the same reaction  $r$ . In the deFBA model we would then build two enzyme production reactions with different amino acid requirements for the two enzymes and then the sum of the amounts of these enzymes will bound the flux through reaction  $r$  together with the corresponding turnover rates.

Having mentioned turnover rates, it is important to keep in mind that usually isoenzymes are assigned the same EC number. This means that, when searching turnover rates for isoenzymes we are usually bound to find the same values in databases such as BRENDA (Schomburg et al., 2013), although in reality the turnover rates may be different. This is partially due to the EC number being a bad standard, but also partly due to lack of biochemical data on different enzyme isoforms. This fact results in isoenzymes most often having the same catalytic efficiency in a deFBA model.

Therefore, the only distinguishing feature of isoenzymes from the perspective of resource allocation, is their amino acid cost. The amino acid cost is not a defined concept. It can be understood in several ways: the length of the enzyme sequence in amino acids or the amount of energy in the form of ATP needed to build all the amino acids in the enzyme or the number of carbon molecules invested in building all the amino acids for an enzyme. If we think of the amino acid cost of an enzyme in terms of the length of the sequence, then if the two isoenzymes have the same length, which of them will be in the end be produced in the deFBA model strongly depends on the requirements of the rest of the system in terms of amino acids at that particular time point. On the other hand, if the lengths of the two isoenzymes are different, very often the deFBA model will predict that only the enzyme with the shorter sequence will be produced and used. This points out two aspects: (i) the system can be simplified by simply not considering at all the more expensive enzyme and this way sparing in the number of variables and the solving time, and (ii) it may be that the regulatory structure of the cell actually has a strong impact on whether these enzymes are produced in reality, but since deFBA ignores this regulatory structure it may mispredict the usage of these enzymes.

Last but not least, which isoenzyme is used may also be dependent on the extracellular conditions in which the cell lives. In this case, even if usually shorter (and hence very likely cheaper) enzymes are preferred, it can still be that the environment turns this decision around and gives rise to a different resource allocation scheme. Therefore, if they do not make a significant difference in the complexity of the resulting linear program, it is recommended that all isoenzymes are represented in the resource allocation model.

The way the isoenzyme production reactions are built depends on their subunit structure. If each of them is only encoded by one gene, then we proceed as in

## 4.2 Building the protein production reactions

---

subsection 4.2.1 above. If several genes are needed to encode one of the isoenzymes, then the respective isoenzyme production reaction will have all amino acids of all involved gene products as substrate requirements, as described in subsection 4.2.3 below.

### 4.2.3 The case of enzyme complexes encoded by several genes

Suppose we want to build the production reaction for an enzyme encoded by the combination of three gene products,  $g_1$ ,  $g_2$ , and  $g_3$ . The metabolic model encodes this as the gene rule for the respective reaction being “ $g_1$  AND  $g_2$  AND  $g_3$ ”. In this case, we compute the amino acid counts for all three genes, add them up and use them as reactants as described in subsection 4.2.1, while making sure we also adapt the ATP and GTP requirements.

This strategy is however only correct in the case in which the peptides encoded by each gene participate in the enzyme as monomers. This is not always the case, and often peptides encoded by one gene participate in enzymes as dimers or trimers, as is the case for example in ATPase, where subunit alpha participates as trimer. Therefore, it is often the case that, before we add up the amino acid counts for all genes, we have to multiply some of the counts with the factor with which they participate in the enzyme.

Information about the stoichiometry of individual peptides within enzymes is unfortunately not readily incorporated in genome-scale metabolic reconstructions and can often only be found through extensive literature research or by querying the UniProt database. Failing to incorporate such stoichiometric information is very likely to give rise to false computational predictions due to the different tradeoffs arising from incorrect enzyme energetic cost.

### 4.2.4 The ribosome

In deFBA models, the ribosome is assumed to simply have catalytic function just as any other enzyme. The amount of ribosome bounds the combined fluxes through the protein production reactions via different turnover rates.

#### Ribosome production

Considering the ribosome to be an “enzyme” encoded by several genes, its production reaction can be modeled as described in subsection 4.2.3 above. As opposed to usual enzymes for which it may be difficult to find the stoichiometry of individual peptides, the ribosomes are rather well studied and information about their composition can be found for many organisms in the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) or in the Ribosomal Protein Gene Database (Nakao et al., 2004). In addition to the ribosomal proteins, also the ribosomal RNA needs to be taken into account for the

## Chapter 4. Building and encoding a metabolic resource allocation model

---

production reaction. Information about this can also be found in the KEGG resource.

### Ribosome translation rate

The ribosome translation rate is a key parameter in a deFBA model and it has very high impact on the tradeoffs that govern the choice of one model behavior over another. This parameter directly affects the enzymatic cost of different pathways.

Ribosome translation rates vary between prokaryotes and eukaryotes and they even vary with growth rate within the same organism (Bremer and Dennis, 2008). They are usually measured in attached amino acids per second, and hence the efficiency of the ribosome for building different enzymes is dependent on this parameter, but also on the respective enzymes' lengths. For instance, if we consider the translation of one enzyme of 100 amino acids by a bacterial ribosome with a rate of 15 amino acids per second then, assuming the enzyme does not compete with other proteins for the ribosome, the enzyme will be translated at a maximal rate of  $3600 \cdot 15/100 = 540$  enzymes per hour. In general the formula for computing the  $k_{cat}$  of the ribosome for the production of a protein is thus given by

$$k_{cat} = 3600 \frac{r}{l},$$

where  $r$  is the ribosome rate in amino acids per second, and  $l$  is the length of the protein in amino acids. In this way we can assign all protein production turnover rates, including those of the protein quota, by using the protein amino acid length, which is the sum of the stoichiometries of the amino acids in the protein production reaction.

### 4.2.5 Compartmentalization

Eukaryotic cells, as opposed to prokaryotic ones, are usually compartmentalized, with compartments such as cytosol, mitochondrion etc. This compartmentalization plays a role in the way enzymes are built, in the sense that there may be identical enzymes that are active in the cytosol as well as in the mitochondrion for example. In this case, two production reactions should be used, one for each compartment, since an enzyme that is in the cytosol cannot catalyze a reaction in the mitochondrion.

## 4.3 Setting up quota compounds and storage

Although the catalytic biomass is the main part of the model that is responsible for the autocatalytic cycle, there are several noncatalytic macromolecules that are also needed in a full cell model. Examples are DNA, RNA, cell wall or membrane. Without accounting for the growth and duplication of these components,



### 4.3 Setting up quota compounds and storage

we would not be modeling a whole copy of a cell and we would be neglecting a significant biosynthetic energy requirement. We recall that we have named these macromolecules *quota* metabolites, to reflect the fact that a certain quota of the total biomass needs to be dedicated to them.

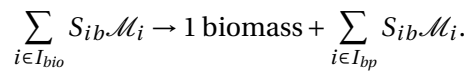
In the case of a periodic environment as is the case for cyanobacteria, algae or plants for example, the production of quota compounds can be enforced by imposing initial quota amounts and requiring that these amounts grow at the same rate as the catalytic biomass, as done in the cFBA model in chapter 6.

In case the environment is not periodic, the simulation can be started by imposing these quota components to make up a certain percentage of total initial biomass, and that these percentages are then kept throughout all the time points of the solution. The question then arises: what are appropriate initial amounts for these quota components?

#### 4.3.1 Initial quota compound amounts

A good place to look for initial quota amounts is the biomass reaction of the metabolic network reconstruction. The stoichiometric coefficients for the reactants of this artificial reaction describe the average composition of the modeled cell. To better understand this, let us take a look at the biomass reaction of the Yeast 6.06 model which we have reproduced in table 4.2.

We observe here the main biomass components: proteins (in the form of charged transfer RNAs), storage (glycogen and trehalose), DNA (in the form of dAMP, dCMP, dGMP, dTMP), RNA (AMP, CMP, GMP, UMP), cell wall (mannan and  $\beta$ -D-glucan), membrane (lumped lipid), other small molecules, and the ATP energy needed for polymerization. For ease of understanding later on, we denote the biomass reaction by  $b$ , the indices of biomass components (reactants of biomass reaction) as  $I_{bio}$ , and the indices of biomass byproducts (products of biomass reaction) as  $I_{bp}$ . Then the biomass reaction has the form



In general, the reactant stoichiometries for the biomass reaction are chosen such that, when weighted by the corresponding molecular weights (MW), they add up to 1, i.e.,

$$\sum_{i \in I_{bio}} S_{ib} MW_i - \sum_{i \in I_{bp}} S_{ib} MW_i = 1.$$

In other words, we can look at the reactant stoichiometries, multiplied by the molecular weights, as the fractions  $q_i := S_{ib} MW_i$  of the respective metabolites within one cell, or equivalently one gram dry weight of cells. Hence, these can also be used as initial quota amounts.

## Chapter 4. Building and encoding a metabolic resource allocation model

Table 4.2: Biomass reaction of the Yeast 6.06 model.

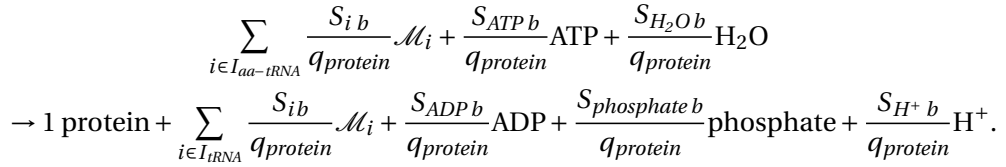
Reactants	Stoichiometry	Products	Stoichiometry
Ala-tRNA(Ala)	0.4588	tRNA(Ala)	0.4588
Arg-tRNA(Arg)	0.1607	tRNA(Arg)	0.1607
Asn-tRNA(Asn)	0.1017	tRNA(Asn)	0.1017
Asp-tRNA(Asp)	0.2975	tRNA(Asp)	0.2975
Cys-tRNA(Cys)	0.0066	tRNA(Cys)	0.0066
Gln-tRNA(Gln)	0.1054	tRNA(Gln)	0.1054
Glu-tRNA(Glu)	0.3018	tRNA(Glu)	0.3018
Gly-tRNA(Gly)	0.2904	tRNA(Gly)	0.2904
His-tRNA(His)	0.0663	tRNA(His)	0.0663
Ile-tRNA(Ile)	0.1927	tRNA(Ile)	0.1927
Leu-tRNA(Leu)	0.2964	tRNA(Leu)	0.2964
Lys-tRNA(Lys)	0.2862	tRNA(Lys)	0.2862
Met-tRNA(Met)	0.0507	tRNA(Met)	0.0507
Phe-tRNA(Phe)	0.1339	tRNA(Phe)	0.1339
Pro-tRNA(Pro)	0.1647	tRNA(Pro)	0.1647
Ser-tRNA(Ser)	0.1854	tRNA(Ser)	0.1854
Thr-tRNA(Thr)	0.1914	tRNA(Thr)	0.1914
Trp-tRNA(Trp)	0.0284	tRNA(Trp)	0.0284
Tyr-tRNA(Tyr)	0.1020	tRNA(Tyr)	0.1020
Val-tRNA(Val)	0.2646	tRNA(Val)	0.2646
ATP	59.2760	ADP	59.2760
H <sub>2</sub> O	59.2760	phosphate	58.70001
(1→3)-β-D-glucan	1.1348	H <sup>+</sup>	59.3050
(1→6)-β-D-glucan	1.1348	biomass	1
glycogen	0.5185		
trehalose	0.0234		
mannan	0.8079		
riboflavin	0.00099		
lipid	1		
sulphate	0.0200		
dAMP	0.0036		
dCMP	0.0024		
dGMP	0.0024		
dTMP	0.0036		
AMP	0.0460		
CMP	0.0447		
GMP	0.0460		
UMP	0.0599		

### 4.3 Setting up quota compounds and storage

To help reduce the number of quota compounds we model and also the size of the resulting resource allocation problems, we can build helper spontaneous reactions that consume the metabolites above and produce the needed quota compounds. As an example, all charged transfer RNAs (whose indices we denote by  $I_{aa-tRNA} \subset I_{bio}$ ) would be consumed to produce one merged protein quota compound, and release all the uncharged tRNAs (indices denoted by  $I_{tRNA} \subset I_{bp}$ ). In setting up this reaction, we should make sure that we adjust the stoichiometric coefficients in such a way that, multiplied with the corresponding amino acids' molecular weights, they add up to one, i.e., we need to divide them by

$$q_{protein} = \sum_{i \in I_{aa-tRNA}} S_{ib} MW_i.$$

Thus, the protein quota building reaction will read



For the protein quota in Yeast 6.06, we can compute from table 4.2  $q_{protein} = 0.466298$ . Therefore, the corresponding quota production reaction would then look as shown in table 4.3. The necessary ATP needed for polymerization, which is the fraction  $q_{protein}$  of the total ATP consumed in the original biomass reaction, is also part of this reaction.

After setting up this reaction, the required initial amount of protein quota is  $q_{protein}$ . However, of these proteins, some are modeled explicitly as enzymes, and in the next section we will see how to adjust the  $q_{protein}$  to only require the proteins that are not modeled as enzymes or ribosome. For the rest of the quota compounds (DNA, RNA, cell wall, membrane, other small molecules) we would proceed in a similar fashion as for the proteins, with the sole difference that their  $q$  would not need to be adjusted once computed.

#### 4.3.2 The case of noncatalytic proteins

The noncatalytic proteins quota is a special case because, in typical biomass reactions, we only have one protein component which encompasses all protein content present in one gram dry weight of cells. However, we need to distinguish in a deFBA model between proteins with catalytic role, and those without, and thus to find a way of splitting this percentage, since obviously the cell contains other proteins besides those we model individually as enzymes.

One way to do this is find a (genome-scale) quantitative proteomics dataset. If not already scaled, we normalize the protein amounts in the dataset to add up to

## Chapter 4. Building and encoding a metabolic resource allocation model

Table 4.3: Helper reaction for producing the protein quota in the Yeast 6.06 model.

Reactants	Stoichiometry	Products	Stoichiometry
Ala-tRNA(Ala)	0.9839201541	tRNA(Ala)	0.9839201541
Arg-tRNA(Arg)	0.3446294001	tRNA(Arg)	0.3446294001
Asn-tRNA(Asn)	0.2181008711	tRNA(Asn)	0.2181008711
Asp-tRNA(Asp)	0.6380040232	tRNA(Asp)	0.6380040232
Cys-tRNA(Cys)	0.0141540388	tRNA(Cys)	0.0141540388
Gln-tRNA(Gln)	0.2260357111	tRNA(Gln)	0.2260357111
Glu-tRNA(Glu)	0.6472255939	tRNA(Glu)	0.6472255939
Gly-tRNA(Gly)	0.6227777087	tRNA(Gly)	0.6227777087
His-tRNA(His)	0.1421837537	tRNA(His)	0.1421837537
Ile-tRNA(Ile)	0.4132550429	tRNA(Ile)	0.4132550429
Leu-tRNA(Leu)	0.6356450167	tRNA(Leu)	0.6356450167
Lys-tRNA(Lys)	0.6137705931	tRNA(Lys)	0.6137705931
Met-tRNA(Met)	0.1087287529	tRNA(Met)	0.1087287529
Phe-tRNA(Phe)	0.2871554242	tRNA(Phe)	0.2871554242
Pro-tRNA(Pro)	0.3532076054	tRNA(Pro)	0.3532076054
Ser-tRNA(Ser)	0.3975998181	tRNA(Ser)	0.3975998181
Thr-tRNA(Thr)	0.4104671262	tRNA(Thr)	0.4104671262
Trp-tRNA(Trp)	0.060905258	tRNA(Trp)	0.060905258
Tyr-tRNA(Tyr)	0.2187442365	tRNA(Tyr)	0.2187442365
Val-tRNA(Val)	0.5674482841	tRNA(Val)	0.5674482841
ATP	36.3823134562	ADP	36.3823134562
H <sub>2</sub> O	36.3823134562	phosphate	36.3823134562
		H <sup>+</sup>	36.3823134562
		protein	1

one. We observe that in this dataset we find two types of proteins corresponding to our model: those modeled explicitly as enzymes, which sum up to a fraction  $f_e$ , and those not present explicitly in our model, that we call quota proteins, and which sum up to  $1 - f_e$  after the normalization of the dataset.

Since we want to adapt the protein quota to only account for the noncatalytic proteins, we adjust  $q_{protein}$  by multiplying it with the fraction  $1 - f_e$  of noncatalytic proteins in the dataset.

An important point here is the growth rate at which the cells used for the quantitative proteomics measurement were growing. It has been shown already in several studies that, for instance, the total amount of ribosomes grows linearly with the growth rate and that partitioning of proteome strongly varies with growth rate and growth conditions (Klumpp et al., 2009; Scott et al., 2010, 2014). Since deFBA models an autocatalytic system that grows exponentially, quantitative

proteomics datasets from exponentially growing cultures should be used if available.

### 4.3.3 Storage

Besides the catalytic and quota macromolecules, most organisms also have storage macromolecules. For some of them, like cyanobacteria, the storage is essential to survive over the night period, when no energy from the sun is present. For others, like yeast, the storage is useful to survive through periods of starvation, as well as sudden changes in the nutrient landscape, when new enzymes and transporters have to be synthesized in order to survive. Therefore, it is important that a resource allocation model takes such storage macromolecules into account.

Taking a second look at the biomass reaction of Yeast 6.06 in table 4.2, we observe that glycogen and trehalose are also consumed. We have however not added them as quota compounds, because these are the storage macromolecules of yeast. Therefore, we allow these macromolecules to accumulate in a deFBA model, the same way as proteins. Moreover, they are part of the total biomass used as objective and to constrain the quota amounts. However, there is no reason why we should impose an initial amount for the storage. We still allow that initial storage amount is present by constraining the total biomass at the beginning of the simulation to be equal to 1 gram. In this way, storage may be present at the beginning, but at the expense of having less catalytic biomass.

## 4.4 Assigning reaction turnover rates

Turnover rates are necessary parameters in a resource allocation model. They are involved in the capacity constraints that bound reaction fluxes using the amount of their catalyzing enzymes. A recent study has shown that turnover numbers reported in online databases are a good enough approximation of *in vivo* turnover numbers (Davidi et al., 2016).

The two main databases for retrieving turnover numbers are BRENDA (Schomburg et al., 2013) and SABIO-RK (Wittig et al., 2012). While BRENDA offers both manually curated as well as text mining data, SABIO-RK only offers data that was either manually extracted from the literature or directly submitted by experimenters. As a result, BRENDA offers a larger amount of turnover rates than SABIO-RK. Both databases offer automated retrieval options. Some simple rules of thumb for retrieving turnover rates from these databases are that one should filter for wild type, non-recombinant values, and, if possible, should make sure that the measurements were done at (nearly) physiological pH and temperature values.

Although large amounts of biochemical data are now available, not all turnover rates for the organism of interest can be found. We recommend in this case, if

## Chapter 4. Building and encoding a metabolic resource allocation model

turnover rates for a given enzyme from other organisms are found, that these should be used. The question then arises: which of the available turnover rates should be used? Should it be a mean or a median of all found turnover rates, or the turnover rate from the organism that has the most sequence similarity with the target organism within that protein?

To answer this question, we have automatically retrieved wild type turnover rates from the BRENDA database for all enzymes from three organisms: *Saccharomyces cerevisiae*, *Escherichia coli*, and *Bacillus subtilis*. In a second iteration, we retrieved turnover rates of all enzymes from all other organisms, excluding the organism of interest, and computed the mean, median, and best sequence match with the organism of interest  $k_{cat}$ . The best sequence match was obtained by computing the alignment score using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970) with the BLOSUM62 scoring matrix (Henikoff and Henikoff, 1992). Since we mostly want that the order of magnitude of the turnover rates is correct, we computed the Pearson correlation coefficients between the logarithms of  $k_{cat}$  values from the organism of interest and the logarithms of the mean, median and best sequence match  $k_{cat}$  values obtained from other organisms. Only values corresponding to the same enzymes were compared. The resulting correlation coefficients are displayed in table 4.4.

Table 4.4: Pearson correlation coefficients between the logarithms of  $k_{cat}$  values from the organism of interest and the logarithms of the mean, median and best sequence match  $k_{cat}$  values obtained from other organisms.

Organism	median	mean	best sequence match
<i>Saccharomyces cerevisiae</i>	0.701	0.650	0.526
<i>Escherichia coli</i>	0.808	0.756	0.606
<i>Bacillus subtilis</i>	0.762	0.708	0.679

We observe that, in the cases we have analyzed, the medians of all turnover rates enzyme-wise is the best approximation for the actual turnover numbers in the organism of interest. Moreover, the order of magnitude correlation coefficients are very high and the P-values we get are all in the order of  $10^{-14}$  or lower, indicating that indeed these median turnover rates from other organisms are good enough approximations of the real  $k_{cat}$  values if no specific data is available for the organism of interest.

To give an idea of the spread of the turnover rate data, we show in figure 4.2 a plot of the  $k_{cat}$  values in yeast versus the median  $k_{cat}$  values from other organisms.

### 4.5 Validating the model using experimental data

Once the model has been constructed, a first step before further investigations is its validation using experimental data. This can stretch from fairly basic match-

## 4.5 Validating the model using experimental data

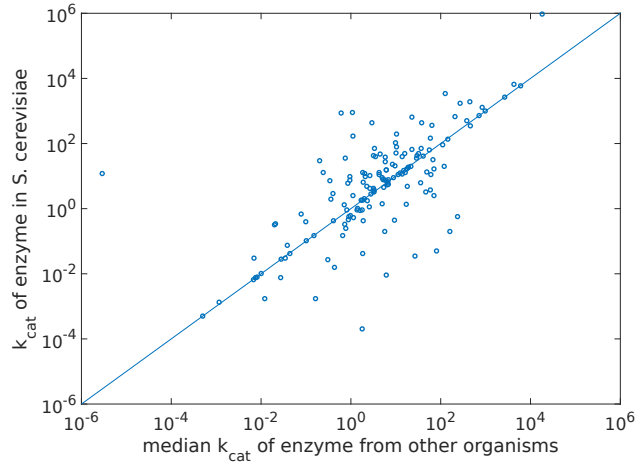


Figure 4.2: Turnover rates in yeast versus the median  $k_{cat}$  values from other organisms.

ing of growth rates obtained from batch experiments to matching of reaction fluxes if data is available.

In general, the growth rate obtained in the model under constant conditions should provide an upper bound on the growth rate measured in the lab, since the model gives the optimal behavior of the metabolism, which may not always be observed in the lab. To check that this is the case, we need data from an exponentially growing batch culture of the organism of interest at saturating nutrient concentrations. In this case, it is sufficient to compute the growth rate of the culture  $\mu_{exp}$  as the slope of the logarithm of the optical density (OD) measurements versus time, as in figure 4.1. In addition, we also need to compute the instantaneous growth rate of the model (under the same conditions as in the experiments), which we define as

$$\mu_{model}(t) := \frac{1}{t_{i+1} - t_i} \ln \left( \frac{biomass(t_{i+1})}{biomass(t_i)} \right), \quad \forall t \in (t_i, t_{i+1}). \quad (4.1)$$

We observe that, if nutrients are saturating,  $\mu_{model}(t)$  is constant, and hence we will refer to it in this case as simply  $\mu_{model}$ .

We note that this definition is nearly identical to the growth rate definition in equation (2.2). It is however based on mass and not on volume, and thus assumes that the mass grows exponentially. Therefore, with a slight abuse of notation and by assuming the density of the cell remains constant, we will use the same letter  $\mu$  to denote the mass-based growth rate.

If  $\mu_{model}$  is smaller than  $\mu_{exp}$ , the problem lies very likely in the  $k_{cat}$  values and these should then be checked manually. If this is not the case,  $\mu_{model}$  can be

## Chapter 4. Building and encoding a metabolic resource allocation model

---

tuned to  $\mu_{exp}$  by forcing flux through the ATP maintenance reaction<sup>1</sup>, which hydrolyzes ATP into ADP and phosphate. This makes sense also from a biological perspective, since we cannot claim that our resource allocation model covers all energy expending processes in an actual cell, so the ATP maintenance reaction serves the purpose of modeling this remaining energy expenditure.

### 4.6 Units

One very important aspect in the model building process are the units, as mismatches between them nearly always give rise to either infeasible or incorrect computational results. Therefore, in this section we will elaborate on the units of the quantities in a deFBA model.

#### 4.6.1 Metabolic reactions

Metabolic reactions  $r$ , including the exchange reactions, are in general of the form

$$r : \sum_i S_{ir} X_i \rightarrow \sum_j S_{jr} X_j,$$

where  $X_i, X_j$  are metabolic reactants resp. products and  $S_{ir}, S_{jr}$  are their stoichiometric coefficients in reaction  $r$ .

As it is usually the case in metabolic network reconstructions, the unit of the flux through this reaction is  $\frac{mmol}{gDW \cdot h}$ . This information can be found in the SBML file of the metabolic network reconstruction. However, methods such as deFBA use molar amounts, and thus the  $gDW$  part is not present, so the fluxes there are in  $\frac{mmol}{h}$ .

Since the derivative of the molar amount of a metabolite participating in this reaction has to have the same unit as the flux, it follows that the internal and external metabolite amounts are measured in  $mmol$ .

The unit of the turnover rates of these reactions is  $h^{-1}$ .

#### 4.6.2 Reactions producing quota compounds

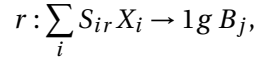
In some metabolic network reconstructions artificial reactions are added to form each biomass component. These biomass components are then in turn consumed by the biomass reaction. This is for instance the case for the *Synechococcus elongatus* model in chapter 6. These reactions have the general

---

<sup>1</sup>Note that the flux we need to force through maintenance strongly depends on how detailed the model is and how much lumping of reactions has been done. We therefore cannot provide an order of magnitude approximation for how much flux should be forced. Modelers should however be aware that the need for a large ATP maintenance forced flux to match experimental growth rates is an indication of model errors or poor model quality.



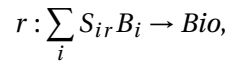
form



where  $X_i$  are internal metabolites,  $B_j$  is an artificial metabolite that represents a quota macromolecule (e.g. membrane), and  $S_{ir}$  are corresponding stoichiometric coefficients, as explained in section 4.2. The coefficients  $S_{ir}$  are chosen such that, when multiplied with the molecular weights of the metabolites  $X_i$ , their sum amounts to 1 gram.

The unit of the flux through such a reaction is then  $h^{-1}$ . The amounts of the metabolites  $X_i$  are measured in *mmol*, and the amount of  $B_j$  is measured in *g*. The reactions producing quota compounds are artificial reactions, and hence in the deFBA model they are marked as spontaneous, i.e., no enzyme is needed for their catalysis.

The biomass reaction of the metabolic reconstruction then has the form



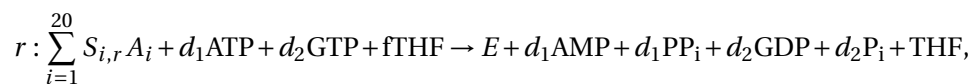
where  $B_i$  are the artificial metabolites that represent quota macromolecules and *Bio* is then the biomass itself, measured in grams. The coefficients  $S_{ir}$  are then chosen such that they represent the fractions of the respective macromolecules in one gram dry weight of cells. The flux through this reaction is measured in  $h^{-1}$ .

While this biomass reaction is not of direct interest in the case of deFBA models, its coefficients  $S_{ir}$  are used as initial quota compound amounts and the whole simulation is set such that at the beginning we are looking at one gram dry weight of cells.

A special case arises here: the quota reaction producing nonmetabolic proteins. This reaction is not spontaneous, but it is catalyzed by the ribosome and competes this way for the ribosome with the enzyme production reactions. Thus, we have to compute a turnover rate for it as described in section 4.3 above. The unit of this turnover rate however is not  $h^{-1}$ , but  $\frac{1}{mmol \cdot h}$ .

### 4.6.3 Reactions producing enzymes

Enzyme production reactions have the form



where  $A_1, \dots, A_{20}$  are the 20 proteinogenic amino acids,  $E$  is the produced enzyme, and  $S_{i,r}$ ,  $d_1 := \sum_{i=1}^{20} S_{i,r}$ ,  $d_2 := 2 \cdot d_1$  are the stoichiometric coefficients of the amino acids and energy requirements, set up as described in section 4.2.

The flux, molar amounts, and turnover rate units for such reactions are the same as the ones for metabolic reactions.

### 4.7 Debugging the model

Even if one follows very carefully the steps we have described above, it is rarely the case that the model building process ends here. Typically, one needs additional efforts for “debugging” errors.

One example of such an error is that, when simulating growth in a simple environment without any limiting nutrients using cFBA, the resulting LP problem is infeasible. If this happens, one should try to analyze the simplest possible case: constant environment, one discretization time step. As we will see in chapter 5, this is equivalent to solving an RBA model. This results in the smallest problem size, since the problem size typically grows linearly with the number of discretization points. In this simplest setting, one can then use the tools of LP solvers to compute an irreducible inconsistent subsystem (IIS) and get an idea about which constraints are involved in the infeasibility. An IIS is a subset of the constraints and variable bounds of the LP. Removing any constraint of the IIS from the LP produces a feasible result. Standard commercial LP solvers like CPLEX or Gurobi offer tools for computing IIS. The type of constraints involved in the IIS could then help diagnose the infeasibility.

A similar problem can be found in a deFBA model simulated in a simple environment without nutrient limitations, with the difference that the resulting LP is not infeasible, but instead no biomass growth is obtained. One can approach this problem in a similar way as above, by forcing a small flux through the ribosome production reaction using a positive lower bound for instance. This will then very likely result in an infeasible LP that can be approached, as above, using the IIS.

However, not too much time should be spent on exploring the IIS because it may happen that there is actually nothing wrong with the model. In other words, the numerical instabilities may be the reason for an infeasibility. It is therefore worth trying to solve the LP also with a more accurate LP solver that can perform iterative refinement, such as SoPlex (Wunderling, 1996; Gleixner et al., 2012, 2016). If a feasible solution is obtained this way, then numerical problems are the cause of the infeasibility.

Last but not least, it is also worth checking that all quota compounds can be produced. If this is not the case, this may be a reason for the infeasibility. To check that each quota compound can be produced, simply require them one by one and solve the resulting LPs.

It can sometimes also happen that the LP problem is feasible, but the resulting time courses reveal problems of the solution. We mention here two types of problems: (i) non-smoothness/instabilities of the time courses of storage/biomass components, and (ii) significantly high amounts of a few of the enzymes. In the former case, the discretization of the dynamic variables may be faulty, as we will see in the next chapter. In the latter case, it may be that

## 4.8 The SBML representation of a metabolic resource allocation model

---

the turnover numbers of those enzymes are not correct. Typically few enough enzymes show this problem, so one can simply check those turnover rates manually and look up alternatives in the databases.

### 4.8 The SBML representation of a metabolic resource allocation model

For metabolic network models used together with constraint-based modeling, it is standard to define them in the systems biology markup language (SBML), an XML-based way of representing models ([http://sbml.org/Main\\_Page](http://sbml.org/Main_Page)). However, there is so far no specific way of representing resource allocation models, which come with several extra ingredients in addition to the metabolic network part. Therefore, in this section we propose a way of making use of the existing SBML capabilities for representing resource allocation models. We illustrate this proposal using a toy resource allocation model listed in table 4.5 for which we attach the SBML representation in Appendix B.

#### 4.8.1 Compartments

We keep any compartments present from the original reconstruction of the metabolic network. If possible, we place the gene products in the compartments in which they are acting. This means for instance that enzymes are located where the reactions they are catalyzing are happening. Species without any real physical location in the model, e.g. quota compounds, can be placed arbitrarily in any compartment.

#### 4.8.2 Species

Each species (metabolite or macromolecule) must contain the fields:

- `id`
- `compartment`
- `constant` (true or false)
- `boundaryCondition` (true or false)
- `initialAmount`

Depending on the function of the species in the model, it also contains a `ram:species` annotation.

We distinguish between limiting extracellular metabolites ( $N_1$  and  $N_2$ ), nonlimiting extracellular metabolites ( $O_2$ ), intracellular metabolites ( $N$ ,  $AA$ ,  $ATP$ ), storage ( $Stor$ ), and biomass components ( $ETrans1$ ,  $ETrans2$ ,  $Complex_1$ ,  $EMetab1$ ,  $EMetab2$ ,  $EStor$ ,  $S$ ,  $R$ ).

The parameters `constant` and `boundaryCondition` are used to express whether a species is intracellular (`constant = "false"` `boundaryCondition =`

## Chapter 4. Building and encoding a metabolic resource allocation model

Table 4.5: List of species, reactions, and catalysis relationships for the toy model whose SBML representation can be found in Appendix B. The reaction that consumes AA and ATP but produces nothing is the maintenance reaction and is considered spontaneous. The reaction that produces the storage compound Stor has different turnover numbers for the forward (f) and reverse (r) directions. Complex\_1 is a transporter complex composed of ETrans1 and ETrans2. S is a structural quota component and R is the ribosome.

External metabolites:	N <sub>1</sub> , N <sub>2</sub> , O <sub>2</sub>		
Internal metabolites:	N, AA, ATP		
Macromolecules:	Stor, ETrans1, ETrans2, Complex_1, EMetab1, EMetab2, EStor, S, R		

Reactions		Catalysed by	Turnover rate
N <sub>1</sub> + O <sub>2</sub>	↔ N	ETrans1	1800
N <sub>2</sub>	↔ N	Complex_1	2400
N	→ AA + ATP	EMetab1	2000
N	→ AA + ATP	EMetab2	2500
N	→ AA + 2 ATP	EMetab2	2000
50 AA + 60 ATP	→ ∅		
200 AA + 300 ATP	↔ Stor	EStor	f: 25, r: 30
100 AA + 450 ATP	→ ETrans1	R	10
60 AA + 270 ATP	→ ETrans2	R	16.67
160 AA + 720 ATP	→ Complex_1	R	6.25
200 AA + 900 ATP	→ EMetab1	R	5
160 AA + 720 ATP	→ EMetab2	R	6.25
150 AA + 500 ATP	→ EStor	R	5
1500 AA + 200 ATP	→ S	R	10
1000 AA + 4500 ATP	→ R	R	1

“false”), whether it is part of the environment and assumed to be present in abundance (constant = “true” boundaryCondition = “true”), or whether the species is a limiting nutrient (constant = “false” boundaryCondition = “true”). We show in Table 4.6 examples on how to set the mandatory species fields for each of these metabolite types.

Table 4.6: Examples for setting species fields for each species type.

	id	compartment	constant	boundaryCondition	initialAmount
N <sub>1</sub>	N1	external	false	false	10
O <sub>2</sub>	O2	external	true	true	10
N	N	metabolites	false	false	0
Stor	Stor	storage	false	false	0
R	R	bio	false	false	0.03364

## 4.8 The SBML representation of a metabolic resource allocation model

The biomass and storage species have an additional annotation field for storing their molecular weight, their objective weight, the biomass percentage and their function as a species in the deFBA model. The biomass percentage attribute contains the fractions of the quota components that need to be reinforced at each time point. We distinguish between the species types “storage”, “enzyme”, and “quota”. Species without a `ram:species` annotation are considered to be either external species or metabolic species in steady-state. Below we show the annotation fields for the ribosome and the structural component.

Ribosome R:

```
<species id="R" name="Ribosome" compartment="cytosol"
  initialAmount="0.03364"
constant="false">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops/304">
      <ram:species ram:molecularWeight="weight_R" ram:
        objectiveWeight="weight_R"
        ram:biomassPercentage="zero" ram:speciesType="enzyme"/>
    </ram:RAM>
  </annotation>
</species>
```

Structural component S:

```
<species id="S" name="Structrual biomass component" compartment
  ="bio"
initialAmount="0.7499" constant="false">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops/304">
      <ram:species ram:molecularWeight="weight_S" ram:
        objectiveWeight="weight_S" ram:biomassPercentage="
        bp_S" ram:speciesType="storage"/>
    </ram:RAM>
  </annotation>
</species>
```

Note that "weight\_R", "weight\_S", "zero", "bp\_S" are ids of parameters defined in the list of parameters of the SBML model.

### Guideline to ensure uniqueness of macromolecule ids

There are some enzymes that can act in different compartments of the cell. An example is fumarase, which catalyzes reactions both in the cytosol and in the mitochondrion in yeast. While we include the respective compartments for the species in their description, a common error is to give both enzymes the same

## Chapter 4. Building and encoding a metabolic resource allocation model

---

id. Hence, we suggest to name enzymes in a specific pattern combining e.g. their name and their respective location.

If the enzyme is acting in only one compartment we choose its id in the format “Main\_id\_[acting-compartment]”. If the enzyme is a transporter between two compartments we choose “Main\_id\_[compartment1]\_[compartment2]”. If the enzyme is translated from only one gene (e.g. ETrans1), this represents the main id. For enzyme complexes made of multiple gene products we suggest simply using “Complex\_number” as main id (e.g. Complex\_1). Of course, the user can choose these ids freely, but following these suggestions can help with easier evaluation of the model in deFBA implementations.

### 4.8.3 Gene products

Each macromolecule is not only present as a species, but also as a gene product, with the fields `id`, `label`, and `associatedSpecies`. The `id` is either the gene name for single gene macromolecules (e.g. ETrans1) or `Complex_number` (e.g. Complex\_1) for complexes as explained above. The `label` field includes the recipe for the creation of the enzyme(-complex). The `associatedSpecies` is the id of the biomass species associated with this gene product and used for bounding the flux. An example for Complex\_1 is:

```
<fb:geneProduct fb:id="Complex_1" label="1*GTRANS1 AND 1*
  GTRANS2" associatedSpecies="Complex_1"/>
```

### 4.8.4 Reactions

All reactions must contain the fields:

- `id`
- `reversible` (true or false)
- `fast` (true or false)
- `listOfReactants`
- `listOfProducts`

Depending on the kind of reaction, they also include an `fb:geneProductAssociation` and a `ram:reaction` annotation. Additionally, we recommend adding the EC number, if known, to the reactions in form of a note:

```
<notes>
  <body xmlns="http://www.w3.org/1999/xhtml">
    <p>EC Number: x.y.z.t</p>
  </body>
</notes>
```

The `id` must be unique for each reaction. For reactions producing biomass we recommend starting the `id` with “synth\_” for easier reading and reduced chance

## 4.8 The SBML representation of a metabolic resource allocation model

of assigning the same id multiple times. Furthermore, this makes it easier to distinguish between the reactions from the original metabolic network model and the deFBA additions. The `fast` attribute can also be used to distinguish biomass production reactions (`fast = "false"`) from the rest of the reactions (`fast = "true"`).

The `fbc:geneProductAssociation` is used to map catalysis relationships between the enzymes (which are also gene products, see above) and the reactions. Each `fbc:geneProductAssociation` consists of at most one `fbc:geneProductRef`, which gives the id of the associated gene product. Reactions that do not have a gene product association are considered spontaneous (e.g. the ATP maintenance reaction), and hence their capacity is not bound using a  $k_{cat}$ .

The forward and reverse  $k_{cat}$  values for each reaction can be found in the annotation, in the fields `kcatForward` and `kcatBackward` respectively. It is typically the case that the values in the `kcatForward` and `kcatBackward` are defined as parameters, and in these fields the ids of the respective parameters are stored, as we will see in the examples below.

### The maintenance reaction

As we have seen above, maintenance reactions are typically part of metabolic resource allocation models, since these models do not account for all energy expenditures of the cell. The maintenance reactions we consider here are growth-associated, i.e., the flux forced through them is dependent on the total biomass at each time point. Thus, we add an annotation field, `maintenanceScaling`, which specifies what is the factor  $f$  such that

$$v_{\text{maintenance}}(t) \geq f \cdot \text{biomass}(t).$$

For typical reactions this field is “zero” as in the case of the metabolic reaction

```
<reaction id="Metab1_2" reversible="false" fast="true">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops/304">
      <ram:reaction ram:kcatForward="kcat2" ram:kcatBackward="zero
        " ram:maintenanceScaling="zero"/>
    </ram:RAM>
  </annotation>
  <fbc:geneProductAssociation fbc:id="Emetab2">
    <fbc:geneProductRef fbc:geneProduct="Emetab2" />
  </fbc:geneProductAssociation>
  <listOfReactants>
    <speciesReference species="N" stoichiometry="1" constant="
      true"/>
  </listOfReactants>
</reaction>
```

## Chapter 4. Building and encoding a metabolic resource allocation model

---

```
</listOfReactants>
<listOfProducts>
<speciesReference species="AA" stoichiometry="1" constant="
  true"/>
<speciesReference species="ATP" stoichiometry="1" constant="
  true"/>
</listOfProducts>
</reaction>
```

For the maintenance reaction we have the representation

```
<reaction id="Maintenance" reversible="false" fast="true">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops/304">
      <ram:reaction ram:kcatForward="zero" ram:kcatBackward="zero"
        ram:maintenanceScaling="main"/>
    </ram:RAM>
  </annotation>
  <listOfReactants>
    <speciesReference species="AA" stoichiometry="50" constant="
      true"/>
    <speciesReference species="ATP" stoichiometry="60" constant="
      true"/>
  </listOfReactants>
  <listOfProducts>
  </listOfProducts>
</reaction>
```

### 4.9 Conclusions

In this chapter we have provided a guide for generating a metabolic resource allocation model. Together with the step by step guidelines, the links to the relevant databases are listed and some guidelines for finding errors that may arise in the model building process are provided. Last but not least, we have proposed a standard for exchanging such models using SBML and the flux balance constraints package.

We will present in detail in chapters 6 and 7 models that have been generated following these guidelines together with the new insights into metabolic resource allocation we have obtained using them.

However, before we proceed, we dedicate the next chapter to techniques for solving deFBA problems, numerical concerns, and actual implementation details.



## Chapter 5

# Metabolic resource allocation in practice: numerical concerns and solving strategies

Assuming that we have generated a full and correct dynamic resource allocation model following the instructions in the previous chapter, the question then comes: how do we actually solve this? What discretization scheme do we use? How do we handle numerically the large difference in order of magnitude of the turnover rates and of the enzyme production and metabolic fluxes?

A lot of the work that goes into studying resource allocation is actually spent on programming the tools for solving such problems. We discuss in this chapter how everything was set up and what problems needed to be overcome in order to obtain the results presented in chapters 6 and 7.

We start by presenting some problems encountered when applying deFBA and cFBA and propose ways to work around these problems. Next, we address numerical matters in solving deFBA and cFBA problems and present a reformulation of cFBA that alleviates discretization problems. Finally, we offer a concise description for the software written to solve deFBA and cFBA problems.

### 5.1 Artefacts of model formulations

#### 5.1.1 deFBA dependency on the discount factor $\varphi$ in the objective

As already mentioned in chapter 2, a practical concern with deFBA is setting the discount factor  $\varphi$  used in the objective function (2.38). Not only do simulations depend very strongly on the choice of  $\varphi$  as shown in figure 5.1, but there is no way of deciding which value to use for this parameter. Although the deFBA framework itself does not depend on using this particular objective function, we

## Chapter 5. Metabolic resource allocation in practice: numerical concerns and solving strategies

note that this is the objective used to simulate the fact that the modeled organism has evolved to grow as fast as possible in the given environment. Thus, this is the objective function used in most applications.

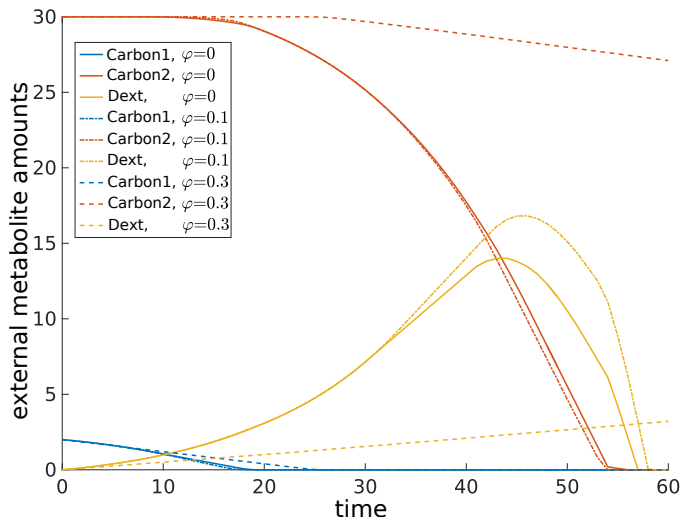


Figure 5.1: External metabolite time courses from the toy deFBA model in (Waldherr et al., 2015) using  $\varphi = 0$ ,  $\varphi = 0.1$ , and  $\varphi = 0.3$  as objective discount factors. Note that both values 0.1 and 0.3 have been used in the simulations in the deFBA article (Waldherr et al., 2015).

The way we have circumvented this problem is simply by eliminating the negative exponential discount altogether ( $\varphi = 0$ ) and instead using as objective

$$\max \int_0^{t_f} \sum_i \mathbf{b}_i \mathbf{n}_i(t) dt, \quad (5.1)$$

where  $i$  denotes a biomass component and the vector  $\mathbf{b}$  contains the molecular weights of the biomass components.

### 5.1.2 Linear versus exponential growth in deFBA

Another difficulty arising in solutions of deFBA problems is that, depending on the extracellular nutrient levels and on the end time  $t_f$  of the simulation, sometimes a linear growth of the biomass is predicted instead of an exponential growth. In theory, any differentiable function, and hence also the exponential, can be *locally* approximated by a linear function. We note, however, that this behavior is not only observed locally, but over a large portion of the simulation period, corresponding to several consecutive time points spread over several hours.

Linear growth is characterized by a linear shape of the total biomass curve over time. We have noticed this problem arising in three situations:

## 5.1 Artefacts of model formulations

- when the extracellular nutrients run out during the simulation,
- when the simulation end time is too small, and a larger  $t_f$  would result in exponential growth,
- towards the end of the simulation time a switch from exponential to linear growth is sometimes observed.

We believe that this is an artefact of the objective function. As shown in figure 5.2, it takes some time before the linear curve is overtaken by an exponential curve, and thus, before this intersection time the linear curve is simply optimal while the exponential one is not.

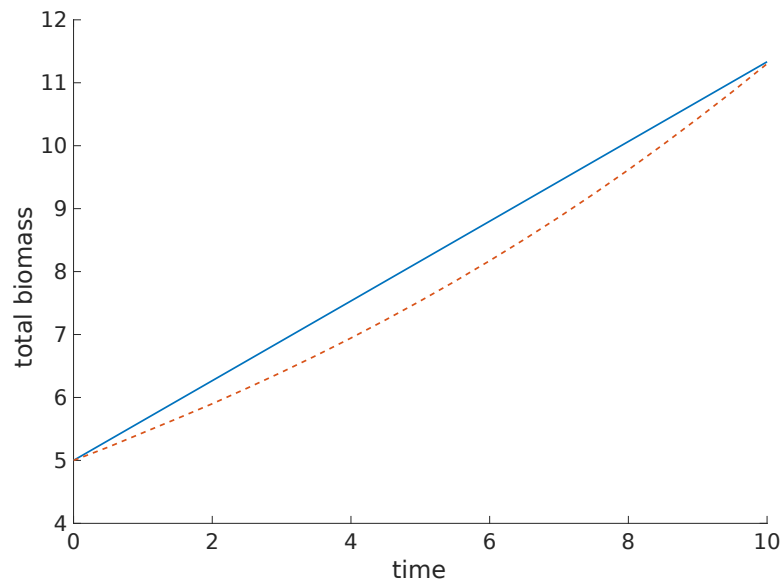


Figure 5.2: The total biomass accumulation in the toy model from (Waldherr et al., 2015), using an end time  $t_f = 10$  (continuous curve) and  $t_f = 60$  (dashed curve). The initial extracellular nutrient amounts are the same as in figure 5.1 above. The discount factor has been set as  $\varphi = 0$ .

The linear growth results from the system producing only the cheapest biomass component as shown in figure 5.3.

## Chapter 5. Metabolic resource allocation in practice: numerical concerns and solving strategies

---

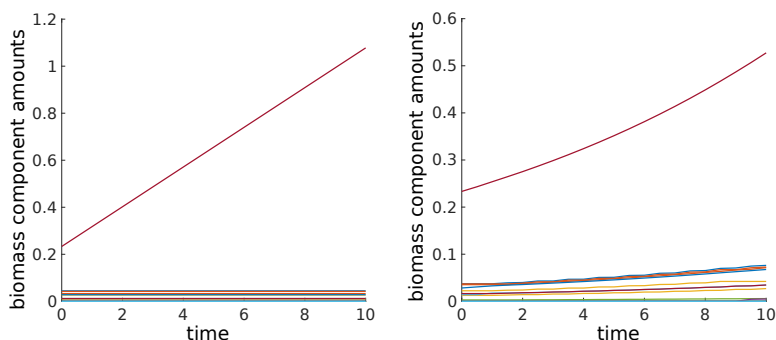


Figure 5.3: Time courses of individual biomass components corresponding to figure 5.2. The left plot shows the time courses for  $t_f = 10$  (linear growth), while the right plot shows those for  $t_f = 60$  (exponential growth).

There is no easy solution to this problem, and we recommend that the modeler simply makes sure that nutrients do not run out during the simulation and that, if a linear growth curve is obtained, then the simulation end time  $t_f$  should be increased.

In a starvation scenario, i.e., when the nutrients run out before the end of the simulation, we observe that deFBA with the biomass integral as objective function (equation (5.1)) predicts linear growth characterized by production of only the “cheapest” (in terms of required resources) biomass component. What is observed in reality, however, is a much more complex behavior: the cell degrades certain proteins to make other catalytic components from the resulting building blocks in an attempt to survive as long as possible. Simply producing the cheapest component in such a scenario is nothing that would enable the cell to survive longer. Therefore, the use of deFBA with the objective (5.1) should be done with care when modeling a cell in starvation conditions.

### 5.1.3 deFBA dependency on choice of simulation endpoint

Another issue we found is the dependency of deFBA solutions on the choice of the end time  $t_f$ . We illustrate this in figure 5.4, where the same model has been simulated using  $t_f = 60$  and  $t_f = 150$ . In this simulation the discounted objective  $J_3$  from the original manuscript (Waldherr et al., 2015), corresponding to objective (2.38) in chapter 2, has been used. Thus, as the authors claim, the quantitative differences should not be due to the existence of alternative optima.

## 5.1 Artefacts of model formulations

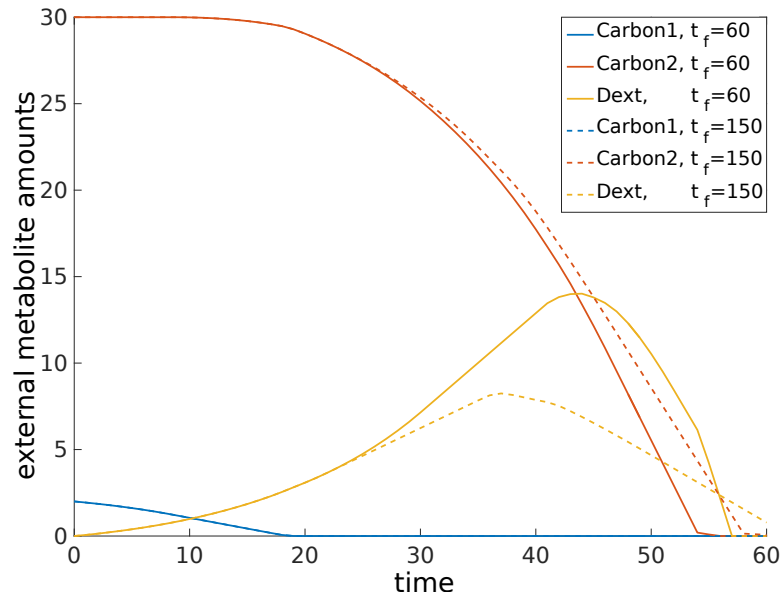


Figure 5.4: External metabolite time courses from the original deFBA toy model in (Waldherr et al., 2015), scenario 1 (carbon switch). The end time points of the simulations were 60 minutes and 150 minutes. We used one discretization point per minute and  $\varphi = 0$ , which is equivalent to using objective (5.1) above.

As has been already proposed in (Lindhorst et al., 2016), this problem can be overcome by using the short term deFBA (st-deFBA). st-deFBA uses a finite prediction horizon and, as the simulations proceed, this moves forward in time. Only the first time point of the result of each simulation is stored as part of the solution and used as input for the next simulation, where the prediction horizon window had moved by one step. The prediction horizon is chosen as the earliest point in time at which the exponential strategy leads to a better objective value than the linear growth strategy. For the exact details we refer to (Lindhorst et al., 2016). However, using short term deFBA comes at the expense of additional simulation time since at each step in the solution a new LP is solved.

We have considered in the work presented here that the end time dependency is small enough to be neglected and have used deFBA as in (Waldherr et al., 2015).

Obviously, one can argue that this end time point dependency does not exist in cFBA. However, cFBA is only suited for organisms displaying balanced growth in a periodic environment such as cyanobacteria or algae. For other organisms, like yeast, there is no reason why we should have periodic biomass composition constraints.

cFBA comes however, with a different difficulty related to optimality as we will see in section 5.5.

## **5.2 Numerical considerations and scaling**

### **5.2.1 The midpoint rule discretization**

To numerically solve deFBA and cFBA problems, the dynamic variables for the macromolecule and external metabolite amounts need to be discretized in time. The authors of deFBA propose a Gauß-Legendre method with Radau collocation points of order 2 and 3 using a Lagrange polynomial interpolation function, while the authors of cFBA use an iterative scheme that turns out to be mathematically equivalent to the simplest discretization method, explicit Euler.

The explicit Euler method can be shown to be numerically unstable, in particular for stiff equations. This means that the numerical solution obtained using explicit Euler becomes very large for equations where the exact solution does not. This is the reason why this numerical method should not be used in practice. On the other hand, a Gauß-Legendre method with Radau collocation points of order 2 and 3 using a Lagrange polynomial interpolation function is not only difficult to implement, but it also susceptible to interpolation errors known as Runge's phenomenon. These are oscillations that appear at the edges of the interpolation interval when using identically spaced interpolation points (Runge, 1901).

A much simpler and stable choice is to use the implicit midpoint rule, which is the simplest collocation method (after the Euler schemes). In short, this states that, given a differential equation

$$y'(t) = f(t, y(t)),$$

and a step size  $h$ , we can discretize it as

$$y_k = y_{k-1} + h \cdot f\left(t_{k-1} + \frac{h}{2}, \frac{y_k + y_{k-1}}{2}\right),$$

where  $y_k$  is the numerical solution at time point  $t_k$ .

This method allows us to preserve a simple implementation and, at the same time, guarantees numerical stability even for stiff equations. Therefore, we have decided to use the implicit midpoint rule in the studies presented in chapters 6 and 7.

Specifically, fixing a number of discretization points  $N = \frac{t_f}{h}$ , we discretize the dynamic variables for macromolecule and extracellular metabolite amount ( $\mathbf{n}$ ) at each discretization point  $t_k$ , while the fluxes  $\mathbf{v}$  and the derivative values  $\dot{\mathbf{n}}$  are evaluated always in the middle of the discretization intervals at  $\frac{t_k + t_{k-1}}{2}$ . This

## 5.2 Numerical considerations and scaling

means that for the dynamic optimization problem in section 2.3.5 we obtain:

$$\max_{\mathbf{n}, \dot{\mathbf{n}}, \mathbf{v}} \sum_{k=1}^N \sum_{i \in \{L, M, T, R, Q\}} \mathbf{n}_i(t_k) \quad (5.2)$$

$$\text{s.t. } \dot{\mathbf{n}}_{\text{Sout}} \left( \frac{t_k + t_{k-1}}{2} \right) = -\mathbf{v}_T \left( \frac{t_k + t_{k-1}}{2} \right) \quad (5.3)$$

$$0 = \mathbf{v}_T \left( \frac{t_k + t_{k-1}}{2} \right) - \mathbf{v}_M \left( \frac{t_k + t_{k-1}}{2} \right) \quad (5.4)$$

$$0 = \mathbf{v}_M \left( \frac{t_k + t_{k-1}}{2} \right) - \mathbf{v}_L \left( \frac{t_k + t_{k-1}}{2} \right) - \sum_{i \in \{L, M, R, T\}} \mathbf{v}_{R_i} \left( \frac{t_k + t_{k-1}}{2} \right) \quad (5.5)$$

$$\dot{\mathbf{n}}_Q \left( \frac{t_k + t_{k-1}}{2} \right) = \mathbf{v}_L \left( \frac{t_k + t_{k-1}}{2} \right) \quad (5.6)$$

$$\dot{\mathbf{n}}_i \left( \frac{t_k + t_{k-1}}{2} \right) = \mathbf{v}_{R_i} \left( \frac{t_k + t_{k-1}}{2} \right), \quad \forall i \in \{L, M, T, R\} \quad (5.7)$$

$$\mathbf{v}_i \left( \frac{t_k + t_{k-1}}{2} \right) \leq k_{cat}^i \frac{\mathbf{n}_i(t_k) + \mathbf{n}_i(t_{k-1})}{2}, \quad \forall i \in \{L, M, T\} \quad (5.8)$$

$$\sum_{i \in \{T, L, M, R\}} \frac{\mathbf{v}_{R_i} \left( \frac{t_k + t_{k-1}}{2} \right)}{k_{cat}^{R_i}} \leq \frac{\mathbf{n}_R(t_k) + \mathbf{n}_R(t_{k-1})}{2} \quad (5.9)$$

$$\mathbf{n}(t_k), \mathbf{v} \left( \frac{t_k + t_{k-1}}{2} \right) \geq 0 \quad (5.10)$$

$$\mathbf{n}_Q(t_k) \geq q \sum_{i \in \{L, M, T, R, Q\}} \mathbf{n}_i(t_k) \quad (5.11)$$

$$\mathbf{n}(t_k) = \mathbf{n}_{t_{k-1}} + h \dot{\mathbf{n}} \left( \frac{t_k + t_{k-1}}{2} \right), \quad (5.12)$$

for all  $k \in \{1, \dots, N\}$ .

This practically means that the LP problem has, for each macromolecule  $2N$  different variables ( $N$  concentration and  $N$  derivative variables) and for each reaction flux  $N$  different variables related using the constraints above.

### 5.2.2 Cyclicity and discretization problems

In this section we point out that also the midpoint rule can lead to strong numerical errors. This can in particular be seen in the case of using a cyclic objective and maximizing the multiplication factor  $\alpha$  as done in cFBA. The following example illustrates this. Assume we are looking at a system with only one reaction



which has flux  $\nu$  and the reaction is catalyzed by  $P$  with a turnover rate  $k_{cat}$ . Then we want to find the maximum multiplication factor  $\alpha$  as follows:

## Chapter 5. Metabolic resource allocation in practice: numerical concerns and solving strategies

---

$$\begin{aligned}
 & \max_{v, \mathbf{n}_P} \alpha \\
 & \text{s.t. } \dot{\mathbf{n}}_P = v \\
 & \quad v \leq k_{cat} \mathbf{n}_P \\
 & \quad \mathbf{n}_P(t_f) = \alpha \mathbf{n}_P(0)
 \end{aligned}$$

We take the simplest case and use only one time step in the discretization, the length of which is  $t_f$ . That means that the derivative  $\dot{\mathbf{n}}_P$  and the flux  $v$  are constant, and the expression for  $\mathbf{n}_P(t_f)$  is given by

$$\mathbf{n}_P(t_f) = \mathbf{n}_P(0) + t_f \dot{\mathbf{n}}_P \quad (5.13)$$

and the capacity bound for  $v$  becomes

$$v \leq k_{cat} \frac{\mathbf{n}_P(t_f) + \mathbf{n}_P(0)}{2},$$

since in the midpoint rule we evaluate  $\mathbf{n}_P$  in the middle of the time interval  $[0, t_f]$  and approximate the lacking value of  $\mathbf{n}_P\left(\frac{t_f}{2}\right)$  by  $\frac{\mathbf{n}_P(t_f) + \mathbf{n}_P(0)}{2}$ .

To get the largest multiplication factor  $\alpha$ , it is obvious that a flux  $v$  will be chosen such that the flux capacity constraint  $v \leq k_{cat} \mathbf{n}_P$  is satisfied with equality, therefore

$$v = k_{cat} \frac{\mathbf{n}_P(t_f) + \mathbf{n}_P(0)}{2}.$$

Replacing  $\mathbf{n}_P(t_f)$  by means of equation (5.13), we get

$$v = k_{cat} \frac{\mathbf{n}_P(0) + t_f \dot{\mathbf{n}}_P + \mathbf{n}_P(0)}{2} = k_{cat} \left( \mathbf{n}_P(0) + \frac{t_f}{2} \dot{\mathbf{n}}_P \right).$$

Since  $\dot{\mathbf{n}}_P = v$ , we have

$$v = k_{cat} \left( \mathbf{n}_P(0) + \frac{t_f}{2} v \right) \Leftrightarrow v = \frac{k_{cat} \mathbf{n}_P(0)}{1 - k_{cat} \frac{t_f}{2}}.$$

Substituting back in equation (5.13), we get

$$\mathbf{n}_P(t_f) = \mathbf{n}_P(0) + t_f \frac{k_{cat} \mathbf{n}_P(0)}{1 - k_{cat} \frac{t_f}{2}},$$

and from the cyclicity constraint then

$$\mathbf{n}_P(t_f) = \mathbf{n}_P(0) + t_f \frac{k_{cat} \mathbf{n}_P(0)}{1 - k_{cat} \frac{t_f}{2}} = \alpha \mathbf{n}_P(0).$$



## 5.2 Numerical considerations and scaling

From the last equation we can then obtain the multiplication factor  $\alpha$  as

$$\alpha = 1 + t_f \frac{k_{cat}}{1 - k_{cat} \frac{t_f}{2}}, \quad (5.14)$$

which points out that  $\alpha$  depends on the end point of the simulation  $t_f$ . In particular, the multiplication factor becomes infinite as  $k_{cat} \frac{t_f}{2}$  approaches 1, that is when the time step approaches  $\frac{2}{k_{cat}}$ .

This points out an unintuitive effect of the combination between the differential equation discretization and the cyclicity of the system. We believe that this is not restricted to cyclic problems, but is a general issue of deFBA and cFBA. In the cyclic case it is only particularly easy to exhibit the artefact. In the following section we show how one can avoid this problem by reformulating the problem and introducing a dilution term in the differential equation for  $\mathbf{n}_p$ .

### 5.2.3 cFBA with dilution term

#### Problem reformulation

We start with a model

$$\begin{aligned} \dot{\mathbf{n}}_{\mathcal{E}}(t) &= S_{\mathcal{E}*} \mathbf{v}(t) & \forall t \geq 0 \\ 0 &= S_{\mathcal{M}*} \mathbf{v}(t) & \forall t \geq 0 \\ A\mathbf{v}(t) &\leq B\mathbf{n}_{\mathcal{E}}(t) & \forall t \geq 0 \\ \alpha \mathbf{n}_{\mathcal{E}}(0) &= \mathbf{n}_{\mathcal{E}}(t_f), \end{aligned}$$

where  $S_{\mathcal{E}*}$  are the enzyme-producing stoichiometries and  $S_{\mathcal{M}*}$  describes the metabolism.  $A, B$  can be arbitrary matrices to constrain fluxes by enzyme amounts.

With a slight abuse of notation, for arbitrary  $\mu > 0$  we define the enzyme concentrations  $\mathbf{e}$  and fluxes  $\mathbf{w}$  subject to dilution by  $\mu$  as

$$\begin{aligned} \mathbf{e}(t) &:= \frac{\mathbf{n}_{\mathcal{E}}(t)}{V(0) \exp(\mu t)} \\ \mathbf{w}(t) &:= \frac{\mathbf{v}(t)}{V(0) \exp(\mu t)}, \end{aligned}$$

where  $V(0)$  is the initial volume of the cell population.

## Chapter 5. Metabolic resource allocation in practice: numerical concerns and solving strategies

---

This yields

$$\begin{aligned}\dot{\mathbf{e}}(t) &= \frac{d}{dt} \frac{\mathbf{n}_{\mathcal{E}}(t)}{V(0) \exp(\mu t)} = \frac{\dot{\mathbf{n}}_{\mathcal{E}}(t)}{V(0) \exp(\mu t)} - \mu \frac{\mathbf{n}_{\mathcal{E}}(t)}{V(0) \exp(\mu t)} \\ \Rightarrow \dot{\mathbf{e}}(t) &= S_{\mathcal{E}*} \frac{\mathbf{v}(t)}{V(0) \exp(\mu t)} - \mu \mathbf{e}(t) \\ \Rightarrow \dot{\mathbf{e}}(t) &= S_{\mathcal{E}*} \mathbf{w}(t) - \mu \mathbf{e}(t) \\ 0 &= S_{\mathcal{M}*} \mathbf{w}(t) \\ A\mathbf{w}(t) &\leq B\mathbf{e}(t) \\ \alpha \mathbf{e}(0) V(0) &= V(0) \exp(\mu t_f) \mathbf{e}(t_f)\end{aligned}$$

Hence, with  $\alpha = \exp(\mu \cdot t_f)$ , we get  $\mathbf{e}(0) = \mathbf{e}(t_f)$ .

This way we can transform a cFBA problem formulation into a formulation that includes dilution by growth. By doing the transformation backwards, we can transform a steady-state solution of the dilution model ( $\mathbf{e}, \mathbf{w}$  variables) to a uniform growth solution of the original problem ( $\mathbf{n}, \mathbf{v}$  variables).

Note that bounds on concentrations and fluxes are still possible. Moreover, it is even possible that the constraint matrices  $A$  and  $B$  depend on time.

### Discretization error in the dilution formulation

The dilution formulation of the problem is no longer affected by the discretization errors described in the previous section. To see this, we consider again the example in section 5.2.2, where we are looking at a system with only one reaction



which has flux  $w$  and the reaction is catalyzed by  $P$  with a turnover rate  $k_{cat}$ . In the transformed problem formulation we want to find the maximum  $\mu$  using the following optimization problem:

$$\begin{aligned}\max \mu \\ \text{s.t. } \dot{\mathbf{e}}_P &= w - \mu \mathbf{e}_P\end{aligned}\tag{5.15}$$

$$w \leq k_{cat} \mathbf{e}_P\tag{5.16}$$

$$\mathbf{e}_P(0) = \mathbf{e}_P(t_f)\tag{5.17}$$

We consider again the simplest case, using one time step of length  $t_f$  for discretization. This means that, by applying the discretization, we obtain

$$\mathbf{e}_P(t_f) = \mathbf{e}_P(0) + t_f \cdot \dot{\mathbf{e}}_P \stackrel{(5.17)}{\Leftrightarrow} \mathbf{e}_P(0) = \mathbf{e}_P(0) + t_f \cdot \dot{\mathbf{e}}_P \Leftrightarrow \dot{\mathbf{e}}_P = 0.\tag{5.18}$$

The bound (5.16) is applied taking the midpoint rule into account, as in section 5.2.2, using the value of  $\mathbf{e}_P$  at the middle of the time interval, i.e.,

$$w \leq k_{cat} \frac{\mathbf{e}_P(0) + \mathbf{e}_P(t_f)}{2} \stackrel{(5.17)}{\Leftrightarrow} w \leq k_{cat} \mathbf{e}_P(0).$$

### 5.3 Connections between cFBA and RBA

To get the largest  $\mu$ , it is obvious that  $w$  will be chosen such that constraint (5.16) is satisfied with equality and we thus obtain

$$w = k_{cat} \mathbf{e}_P(0). \quad (5.19)$$

From equations (5.15), (5.18), and (5.19), we observe that

$$w = \mu \mathbf{e}_P(0) \quad \text{and} \quad w = k_{cat} \mathbf{e}_P(0),$$

which shows that

$$\mu = k_{cat}.$$

We thus observe that the expression for  $\mu$  no longer depends on the end time  $t_f$  as it did in (5.14). Moreover, we observe that, with

$$\alpha = \exp(\mu \cdot t_f) \Leftrightarrow \mu = \frac{\log(\alpha)}{t_f},$$

the values for  $\mu$  and  $\alpha$  are compatible when taking the limit  $t_f \rightarrow 0$  in equation (5.14),

$$\begin{aligned} \lim_{t_f \rightarrow 0} \frac{\log(\alpha)}{t_f} &\stackrel{(5.14)}{=} \lim_{t_f \rightarrow 0} \frac{\log\left(1 + t_f \frac{k_{cat}}{1 - k_{cat} \frac{t_f}{2}}\right)}{t_f} \\ &\stackrel{\text{l'Hôpital}}{=} \lim_{t_f \rightarrow 0} \frac{\frac{1}{1 + t_f \frac{k_{cat}}{1 - k_{cat} \frac{t_f}{2}}} \cdot \frac{k_{cat} \left(1 - k_{cat} \frac{t_f}{2}\right) + t_f \frac{k_{cat}^2}{2}}{\left(1 - k_{cat} \frac{t_f}{2}\right)^2}}{1} \\ &= k_{cat}. \end{aligned}$$

#### Conclusion

It is not coincidental that we have chosen  $\mu$  for the transformation.  $\mu$  is the actual growth rate of the exponentially growing system.

To keep the simplicity of explanations in chapter 6, we consider this dilution transformation an implementation detail. We still describe the model using the original cFBA formulation, although in the solving software we implement the dilution formulation.

### 5.3 Connections between cFBA and RBA

After the dilution transformation, we observe that the cFBA problem formulation looks strikingly similar to RBA, and that cFBA with one time step is in fact equivalent to RBA. Indeed, we show in this section that RBA provides an upper bound on growth rates obtained using cFBA.

## Chapter 5. Metabolic resource allocation in practice: numerical concerns and solving strategies

---

We recall from chapter 2 that concentration is defined as the number of molecules  $\mathbf{n}$  of substance present in a certain volume  $V$  of a solution, i.e.,  $\mathbf{c}(t) = \frac{\mathbf{n}(t)}{V(t)}$ . While in section 3.2 we assumed a constant volume, we now allow the volume to change over time. This happens for example in the case of cell growth, when the total volume of all cells grows.

The metabolic network usually consumes and produces metabolites as specified in the stoichiometric matrix  $S$ . The product  $S_{\mathcal{M}*}\mathbf{v}(t)$  then gives the change in metabolite concentrations when the volume stays constant. This product reflects the net production of each metabolite by the metabolic network. In case the volume changes, we also obtain a dilution term, as in RBA and as in section 5.2.3 above, so

$$\begin{aligned}\dot{\mathbf{c}}(t) &= S_{\mathcal{M}*}\mathbf{v}(t) - \mu(t)\mathbf{c}(t), \\ \dot{\mathbf{e}}(t) &= S_{\mathcal{E}*}\mathbf{v}(t) - \mu(t)\mathbf{e}(t),\end{aligned}\tag{KM2a}$$

where the growth rate  $\mu(t)$  is defined as

$$\mu(t) := \frac{\dot{V}(t)}{V(t)}.\tag{5.20}$$

We observe that

$$\tilde{\mu}(T) := \frac{1}{T} \int_0^T \mu(t) dt.\tag{AVGM}$$

is the average growth rate:

**Proposition 5.3.1** *It holds for all  $T \geq 0$  that  $V(T) = V(0) \exp(\tilde{\mu}(T)T)$ .*

PROOF We define  $W(t) := \ln(V(t))$  for all  $t \geq 0$ . It follows that

$$\dot{W}(t) = \frac{\dot{V}(t)}{V(t)} = \mu(t)\tag{5.21}$$

$$\Rightarrow \int_0^T \mu(t) dt = \int_0^T \dot{W}(t) dt = W(T) - W(0)\tag{5.22}$$

$$\Rightarrow \frac{V(T)}{V(0)} = \exp(W(T) - W(0)) = \exp\left(\int_0^T \mu(t) dt\right).\tag{5.23}$$

The proposition follows by definition of  $\tilde{\mu}$ . ■

### 5.3.1 Modeling assumptions (with dilution)

For ease of notation, we use  $\mathbf{z} := (\mathbf{c}, \mathbf{e})^\top$  and treat enzyme and metabolite concentrations together in the results we derive here. Following the derivation above based on (KM2a), we now consider the new kinetic model

$$\begin{aligned}\dot{\mathbf{z}}(t) &= S\mathbf{v}(t) - \mu(t)\mathbf{z}(t), \\ \mathbf{v}(t) &= f(\mathbf{z}(t)),\end{aligned}\tag{KM2}$$

where  $f$  denotes the kinetic rate law functions. The system (KM2) now also models dilution of internal metabolites and enzymes via cell growth. The rest of the assumptions are the same as in section 3.2.1. In addition we assume that the growth rate  $\mu(t)$  is positive, bounded and continuous for all time points  $t \geq 0$ .

### 5.3.2 Average fluxes and average concentrations

For a given time period  $T$ , we additionally define the average concentrations  $\bar{\mathbf{z}}$  as:

$$\bar{\mathbf{z}}(T) := \frac{1}{T} \int_0^T \frac{\mu(t)}{\bar{\mu}(T)} \mathbf{z}(t) dt. \quad (\text{AVGZ})$$

Note that we scale the concentrations by the growth rate  $\mu$ . The motivation for this is that, in order to avoid depletion of metabolite pools, it is much more important to overproduce metabolites in fast-growing periods than in slow-growing ones. We observe that  $\bar{\mathbf{z}}(T)$  can be considered an average over growth rates rather than over time since

$$\bar{\mathbf{z}}(T) = \frac{\int_0^T \mu(t) \mathbf{z}(t) dt}{\int_0^T \mu(t) dt}. \quad (5.24)$$

As in the case for the average fluxes  $\bar{\mathbf{v}}$ , it can happen that

$$\bar{\mathbf{z}} := \lim_{T \rightarrow \infty} \bar{\mathbf{z}}(T), \quad (5.25)$$

$$\bar{\mu} := \lim_{T \rightarrow \infty} \bar{\mu}(T) \quad (5.26)$$

do not exist. Hence we again assume, for simplicity, that the limits  $\bar{\mathbf{v}}$ ,  $\bar{\mathbf{z}}$ , and  $\bar{\mu}$  exist and refer the reader to the appendix of (Reimers and Reimers, 2016) for the case where they do not exist.

We observe in the following that, if we consider dilution, the steady-state condition for the average fluxes changes slightly:

**Theorem 5.3.2**  $S\bar{\mathbf{v}} = \bar{\mathbf{z}}\bar{\mu}$ .

PROOF Following the same arguments as in the proof of theorem 3.2.1, we obtain that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T S\mathbf{v}(t) - \mu(t)\mathbf{z}(t) dt = 0. \quad (5.27)$$

We can then use equations (AVGV), (5.24), and (AVGM) to arrive at the theorem

## Chapter 5. Metabolic resource allocation in practice: numerical concerns and solving strategies

---

statement as

$$0 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T S\mathbf{v}(t) dt - \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mu(t)\mathbf{z}(t) dt \quad (5.28)$$

$$\stackrel{\text{(AVGV)}}{=} S\bar{\mathbf{v}} - \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mu(t)\mathbf{z}(t) dt \quad (5.29)$$

$$\stackrel{\text{(5.24)}}{=} S\bar{\mathbf{v}} - \lim_{T \rightarrow \infty} \bar{\mathbf{z}}(T) \frac{1}{T} \int_0^T \mu(t) dt \quad (5.30)$$

$$\stackrel{\text{(AVGM)}}{=} S\bar{\mathbf{v}} - \bar{\mathbf{z}}\bar{\mu}. \quad (5.31)$$

■

One way to understand theorem 5.3.2 is to think of  $\bar{\mathbf{z}}$  as the composition of the biomass of an RBA model.  $\bar{\mu}$  is then the flux through the biomass production reactions. Because  $\bar{\mathbf{z}}$  includes in our case all macromolecules and all metabolites present in the cell, all these have to be duplicated upon cellular division. This way, the dilution term enforces the production of all macromolecules present in the cell.

We observe here a shortcoming in the formulation of RBA and cFBA, which ignore metabolite dilution by growth. We check the impact of this shortcoming in the next section.

### 5.3.3 Violation of the steady-state condition by dilution

In the following we estimate the order of magnitude of this violation relative to the fluxes. For this purpose let us assume that the RBA model approximates  $\bar{\mathbf{z}}$  with a total error of 1 mM. We again use the three organisms *Escherichia coli*, *Saccharomyces cerevisiae* and *Homo sapiens* as examples. The detailed calculations for the three organisms are presented in Appendix A. Other examples where the same trend can be observed can be found in (Stephanopoulos et al., 1998).

#### **Escherichia coli**

*E. coli* has an average cell volume of  $0.6 \mu\text{m}^3$  (Kubitschek, 1990), a dry weight of 0.489 pg (Loferer-Krößbacher et al., 1998), an average growth rate on glucose of  $0.9 \text{ h}^{-1}$  (Andersen and von Meyenburg, 1980), and a glucose uptake rate of  $12 \text{ mmol}/(\text{gDW} \cdot \text{h})$  (Jain and Srivastava, 2009). Using these values it follows that an approximation error of 1 mM for  $\bar{\mathbf{z}}$  implies a violation of the steady-state condition in the order of  $10^{-4}$  relative to the fluxes.

#### **Saccharomyces cerevisiae**

A similar result is obtained in the case of *S. cerevisiae*, which has an average growth rate on glucose of  $0.4 \text{ h}^{-1}$  (Waldron and Lacroute, 1975), average intra-

cellular fluxes of 0.5 mmol/(gDW · h) (Stewart et al., 2010), a dry weight of approximately  $10^{-11}$  g = 10 pg (Mitchison, 1958), and a volume of  $20 \mu\text{m}^3$  (Tyson et al., 1979). These values imply that an approximation error of 1 mM for  $\bar{\mathbf{z}}$  leads to a violation of the steady-state condition in the order of  $10^{-3}$  relative to the fluxes.

### Homo sapiens (HeLa cells)

In the case of HeLa cells we obtain a similar order of magnitude for the violation. HeLa cells have an average growth rate of  $0.06 \text{ h}^{-1}$  (Kumei et al., 1989), glucose uptake flux of about 18 nmol/(min · mg protein) (Mojena et al., 1985), approximately 150 pg protein (Finka and Goloubinoff, 2013), and a volume of  $2600 \mu\text{m}^3$  (Luciani et al., 2001; Finka and Goloubinoff, 2013). With these values, an approximation error of 1 mM for  $\bar{\mathbf{z}}$  implies a steady-state condition violation in the order of  $10^{-3}$  relative to the fluxes.

#### 5.3.4 Bounds for cFBA from RBA

Let  $\bar{\mu}^*$  be the optimum of the following optimal control problem formulated based on the kinetic model (KM2):

$$\begin{aligned} \bar{\mu}^* &:= \max_{\mathbf{v}(t), \mathbf{z}(t), \mu} \mu \\ \text{s.t.} \quad & \dot{\mathbf{z}}(t) = \mathbf{S}\mathbf{v}(t) - \mu(t)\mathbf{z}(t) & \forall t \geq 0 \\ & -K^-\mathbf{z}(t) \leq \mathbf{v}(t) \leq K^+\mathbf{z}(t) & \forall t \geq 0 \\ & \mathbf{l}(t) \leq \mathbf{v}(t) \leq \mathbf{u}(t) & \forall t \geq 0, \end{aligned} \quad (\text{OCP2})$$

where  $\mathbf{l}(t)$  and  $\mathbf{u}(t)$  are time-dependent functions that can be used to enforce limits from the environment, and  $K^-, K^+$  are matrices containing the turnover rates for the forward and reverse directions of the reactions.

Let  $\mu_{RBA}^*$  denote the optimum of the following RBA problem:

$$\begin{aligned} \mu_{RBA}^* &:= \max_{\mathbf{v}, \mathbf{z}, \mu} \mu \\ \text{s.t.} \quad & \mathbf{S}\mathbf{v} = \mathbf{z}\mu \\ & -K^-\mathbf{z} \leq \mathbf{v} \leq K^+\mathbf{z} \\ & \bar{\mathbf{l}} \leq \mathbf{v} \leq \bar{\mathbf{u}}, \end{aligned} \quad (\text{RBA2})$$

where  $\bar{\mathbf{l}}, \bar{\mathbf{u}}$  are defined as

$$\begin{aligned} \bar{\mathbf{l}} &:= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{l}(t) dt, \\ \bar{\mathbf{u}} &:= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{u}(t) dt, \end{aligned}$$

assuming the limits exist.

## Chapter 5. Metabolic resource allocation in practice: numerical concerns and solving strategies

---

In this case it also holds that the RBA optimum gives an upper bound for the growth rate  $\tilde{\mu}^*$ :

**Corollary 5.3.3** *It holds that  $\mu_{RBA}^* \geq \tilde{\mu}^*$ .*

PROOF Assume the limits  $\bar{\mathbf{v}}$  and  $\bar{\mathbf{z}}$  exist. Then, by theorem 5.3.2,  $S\bar{\mathbf{v}} = \bar{\mathbf{z}}\mu$ .

Since  $\mu$  does not change over time in (OCP2), it follows that

$$\tilde{\mu}(T) = \mu$$

and thus we have

$$\bar{\mathbf{z}} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{z}(t) dt.$$

By the linearity of the enzyme capacity constraint in (OCP2), we obtain

$$-K^- \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{z}(t) dt \leq \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{v}(t) dt \leq K^+ \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{z}(t) dt,$$

which is equivalent to

$$-K^- \bar{\mathbf{z}} \leq \bar{\mathbf{v}} \leq K^+ \bar{\mathbf{z}}.$$

It is also easy to see that  $\bar{\mathbf{v}}$  satisfies  $\bar{\mathbf{I}} \leq \bar{\mathbf{v}} \leq \bar{\mathbf{u}}$ . Thus, we observe that  $\bar{\mathbf{v}}, \bar{\mathbf{z}}, \mu$  is a feasible solution of (RBA2).

If  $\bar{\mathbf{v}}$  and  $\bar{\mathbf{z}}$  do not exist, we instead can use theorem 3 in Appendix A of (Reimers and Reimers, 2016) and any accumulation point pair  $\bar{\mathbf{v}}, \bar{\mathbf{z}}$  will be a feasible solution of RBA2. ■

Since cFBA assumes a constant growth rate over the whole cycle, similar to (OCP2), we therefore conclude that RBA provides upper bounds on growth rates from cFBA models. This fact can conveniently be used when debugging implementation errors of dynamic resource allocation models as described in chapter 4.

### 5.4 Condition numbers, time scale separation and scaling

To derive the quasi-steady-state approximation, the inventors of deFBA make use of a transformation that reflects the fact that macromolecule production reactions are much slower than metabolic reactions and that the macromolecule amounts in general change much more slowly than metabolite amounts. This transformation effectively means that the macromolecule production fluxes are scaled by a small dimensionless factor  $\epsilon$ .

This transformation also has practical consequences for the solving implementation. Since macromolecule production fluxes are much smaller than metabolic fluxes, we obtain a large order of magnitude difference between these variables inside the LP. This large difference gives rise in general to numerical



## 5.4 Condition numbers, time scale separation and scaling

---

inaccuracies and rounding errors, as explained in (Klotz, 2014). By scaling the macromolecule production fluxes, the macromolecule concentrations and their derivatives, we thus avoid such rounding errors. For ease of notation, we also consider this an implementation detail and do not include it in the problem descriptions in this thesis, but only in the software implementation.

Despite the  $\varepsilon$  scaling, additional numerical inaccuracies can stem from the  $k_{cat}$  values used. A quick look at figure 4.2 will reveal that these parameters spread over several orders of magnitude. The only kind of scaling that can help improve the numerical stability in this case is the individual scaling of problematic rows of the constraint matrices.

To get an intuition about the numerical condition of typical dynamic metabolic resource allocation problems, we compute the condition number of the basis matrices corresponding to typical solutions of the problems in chapters 6 and 7.

### 5.4.1 Recap on matrix condition numbers

Let us first have a look at the derivation of the condition number, which we have adapted from (Klotz, 2014). Given the linear programming problem

$$\max \mathbf{a}^T \mathbf{x} \quad (5.32)$$

$$\text{s.t. } A\mathbf{x} = \mathbf{b} \quad (5.33)$$

$$\mathbf{x} \geq 0, \quad (5.34)$$

we look at the subset  $B$  of the rows of the matrix  $A$  that is present in the basis. The solution at this point is thus given by

$$\mathbf{x}_B = A_{B^*}^{-1} \mathbf{b}.$$

Considering a small perturbation  $\Delta \mathbf{b}$  of the right hand side  $\mathbf{b}$ , we are interested in the change  $\Delta \mathbf{x}_B$  this perturbation will induce in the solution  $\mathbf{x}_B$ . After a couple of computations and after applying the Cauchy-Schwarz inequality, we obtain the upper bound

$$\|\Delta \mathbf{x}_B\| \leq \|A_{B^*}^{-1}\| \cdot \|\Delta \mathbf{b}\|. \quad (5.35)$$

Similarly,

$$\|\mathbf{b}\| \leq \|A_{B^*}\| \cdot \|\mathbf{x}_B\|. \quad (5.36)$$

Multiplying the left and right hand sides of (5.35) and (5.36) together and rearranging terms we obtain

$$\frac{\|\Delta \mathbf{x}_B\|}{\|\mathbf{x}_B\|} \leq \|A_{B^*}\| \cdot \|A_{B^*}^{-1}\| \left( \frac{\|\Delta \mathbf{b}\|}{\|\mathbf{b}\|} \right),$$

## Chapter 5. Metabolic resource allocation in practice: numerical concerns and solving strategies

---

which gives an upper bound on the relative change in the solution upon the perturbation  $\|\Delta \mathbf{b}\|$ . For more details of the derivation we refer to (Klotz, 2014).

We observe that the condition number of the basis matrix, which is a submatrix of the constraint matrix  $A$ , defined as

$$\kappa(A_{B^*}) = \|A_{B^*}\| \cdot \|A_{B^*}^{-1}\|$$

is a factor in the upper bound on the error we obtain upon a perturbation. This number is a property of the modeled system rather than a result of rounding errors. Since ill-conditioning of a problem is typically defined as a small perturbation in the input giving rise to a large error in the output, the condition number allows us to assess the ill-conditioning of a problem given (submatrices of) its constraint matrix  $A$ .

The rounding errors that occur in finite precision computing environments can also be seen as perturbations to the entries of the constraint matrix or the right hand side. These errors can be increased by a large condition number. To give an intuition, assume  $\|\Delta \mathbf{b}\|$  is in the order of  $10^{-16}$  and  $\kappa(A_{B^*})$  is in the order of  $10^{13}$ . Then rounding errors up to  $10^{-16} \cdot 10^{13} = 10^{-3}$  can arise in the calculations. In such cases, algorithms of LP solvers can run into numerical difficulties.

### 5.4.2 Condition numbers for typical deFBA and cFBA problems

The basis matrix of a typical solution of the *Synechococcus elongatus* model in chapter 6 has a condition number in the order of  $10^{11}$ , while that of a typical solution of the *Saccharomyces cerevisiae* model in chapter 7 has a condition number in the order of  $10^9$ . Note that these condition numbers are obtained after the scaling by  $\varepsilon$  explained above.

In the light of the example above, these values are obviously large enough to present a problem for standard commercial solvers such as CPLEX or Gurobi, which offer feasibility check precision of only up to  $10^{-9}$ .

Therefore, to minimize the potential rounding errors appearing, we have decided to use an open source LP solver, SoPlex (Wunderling, 1996; Gleixner et al., 2012, 2016), with higher precision (up to  $10^{-16}$ ) than standard commercial solvers, and which can perform iterative refinement of the solution, going up to arbitrary precision.

This increase in precision comes however at the expense of increased solving time. In addition, the solving time is further increased in the software written for this thesis by the lack of an interface of SoPlex for MATLAB. This results in each LP having to be written to file by the MATLAB code and then read by the solver. With large LPs such as those resulting from dynamic resource allocation problem, this becomes a real challenge.

## 5.5 Convexity of the binary search in cFBA

While deFBA problems become linear programs after the discretization and hence are guaranteed to arrive at a global optimum, this is not the case for cFBA problems because of the quadratic terms in the constraints.

Similar to RBA and ME models, cFBA uses a binary search procedure to arrive at the optimal multiplication factor  $\alpha^+$  which corresponds to the optimal growth rate  $\mu^+$ . However, we need to prove mathematically that this binary search is guaranteed to arrive at the optimal growth rate. To prove this, one has to show that, if the organism can grow at a growth rate  $\mu^+$ , then it can also grow at any growth rate  $\mu \in [0, \mu^+]$ . While biologically this is intuitively correct, we also show here how it works out mathematically.

We can only prove the convexity of a slightly modified problem, where we weaken the constraint

$$\dot{\mathbf{e}}(t) = S_{\mathcal{E}^*} \mathbf{w}(t) - \mu^+ \mathbf{e}(t)$$

by only requiring an inequality, as

$$\dot{\mathbf{e}}(t) \leq S_{\mathcal{E}^*} \mathbf{w}(t) - \mu^+ \mathbf{e}(t).$$

At a steady-state, as in the case of RBA, this would mean that production of macromolecules has to be greater or equal to dilution by growth, i.e.,

$$S_{\mathcal{E}^*} \mathbf{w}(t) - \mu^+ \mathbf{e}(t) \geq 0,$$

which is the same as constraint (2.14) in the RBA formulation.

**Theorem 5.5.1** *Given a feasible solution  $(\mathbf{w}(t)^+, \mathbf{e}(t)^+, \mu^+)^{\top}$  with  $\mu^+ > 0$  of the optimization problem*

$$\begin{aligned} & \max_{\mathbf{w}(t), \mathbf{e}(t), \mu} \mu \\ & \text{s.t. } \dot{\mathbf{e}}(t) \leq S_{\mathcal{E}^*} \mathbf{w}(t) - \mu \mathbf{e}(t) \\ & \quad \mathbf{0} = S_{\mathcal{M}^*} \mathbf{w}(t) \\ & \quad A \mathbf{w}(t) \leq B \mathbf{e}(t) \\ & \quad \mathbf{e}(0) = \mathbf{e}(t_f) \end{aligned} \tag{OCP3}$$

for every  $\mu \geq 0, \delta\mu > 0$  with  $\mu^+ = \mu + \delta\mu$ ,  $(\mathbf{w}(t)^+, \mathbf{e}(t)^+, \mu)^{\top}$  is also a feasible solution of (OCP3).

PROOF Since

$$\dot{\mathbf{e}}(t)^+ \leq S_{\mathcal{E}^*} \mathbf{w}(t)^+ - \mu^+ \mathbf{e}(t)^+$$

is the only constraint where  $\mu^+$  appears, we only need to check that this is satisfied also for the smaller growth rate  $\mu$ . With  $\mu^+ = \mu + \delta\mu$ , for  $\delta\mu > 0$ , we obtain

$$\dot{\mathbf{e}}(t)^+ \leq S_{\mathcal{E}^*} \mathbf{w}(t)^+ - (\mu + \delta\mu) \mathbf{e}(t)^+ \leq S_{\mathcal{E}^*} \mathbf{w}(t)^+ - \mu \mathbf{e}(t)^+,$$

since the term  $\delta\mu \cdot \mathbf{e}(t)^+$  is positive. ■

## Chapter 5. Metabolic resource allocation in practice: numerical concerns and solving strategies

---

From a biological perspective, weakening the constraint as above is equivalent to overproducing amino acids or energy, and most organisms have mechanisms to dissipate excess energy such as futile cycles or wastage of ATP by an ATPase. However, by weakening the constraint for the cyanobacterium model in chapter 6, we obtain a slightly increased growth rate in the reference day case (3.68% increase in  $\mu$ ). This points out that indeed using the equality version is a restriction to the system and this may be one reason why we were not able to prove a stronger convexity statement.

### 5.6 Implementation

In this section we give a very brief description about implementing deFBA (and its cFBA variant) and what are necessary parts of this implementation. For an example implementation focused on deFBA we refer the reader to the github repository at <https://github.com/alexandra-m-reimers/deFBA>.

#### 5.6.1 Variables

The variables of a deFBA problem are the fluxes  $\mathbf{v}$ , the amounts for the extracellular metabolites and macromolecules  $\mathbf{n}$ , and their derivatives  $\dot{\mathbf{n}}$ . Each of these variables in turn has  $N$  instances, one for each time point of the discretization. We can therefore imagine  $\mathbf{v}$  for instance as a matrix with  $N$  rows and  $|\mathcal{R}|$  columns. The same then applies for  $\mathbf{n}$  and  $\dot{\mathbf{n}}$ . Therefore, a deFBA numerical solution is a structure made of these three different matrices. At this point it is clear that a necessary ingredient of the implementation are two functions that can convert from the solution vector to the variable matrices (*toStruct*) and back (*toVector*) for ease of access of the results as well as for easily formulating the constraints.

In addition to  $\mathbf{v}$ ,  $\dot{\mathbf{n}}$ ,  $\mathbf{n}$ , if one does not want to split reversible reactions, then helper variables for imposing the capacity bounds are needed. As an example, assume reversible reaction  $r_1$  and irreversible reaction  $r_2$  are both catalyzed by enzyme  $e_1$ , with turnover rates  $k_1$  and  $k_2$  respectively. Then the capacity bound for them is given by

$$\frac{|\mathbf{v}_{r_1}(t)|}{k_1} + \frac{\mathbf{v}_{r_2}(t)}{k_2} \leq \mathbf{n}_{e_1}(t).$$

Obviously, we need to reformulate this constraint since LP solvers cannot handle absolute values. One way to do this, is to introduce helper variables  $\mathbf{E}$  and

formulate this constraint as

$$\begin{aligned}
 \mathbf{v}_{r_1}(t) &\leq k_1 \mathbf{E}_{r_1}(t) \\
 -\mathbf{v}_{r_1}(t) &\leq k_1 \cdot \mathbf{E}_{r_1}(t) \\
 \mathbf{v}_{r_2}(t) &\leq k_2 \mathbf{E}_{r_2}(t) \\
 \mathbf{E}_{r_1}(t) + \mathbf{E}_{r_2}(t) &\leq \mathbf{n}_{e_1}(t) \\
 \mathbf{E}(t) &\geq 0
 \end{aligned}$$

This formulation allows us the flexibility of also having different turnover rates for the forward and the reverse direction of a reaction.  $\mathbf{E}$  has the same structure and dimensions as the vector  $\mathbf{v}$ .

However, another special case to be handled is also the case where another enzyme  $e_2$  catalyzes  $r_2$  and another reaction  $r_3$ . We already notice that we would need a second, different  $\mathbf{E}_{r_2}(t)$  variable for implementing this constraint as explained above. Therefore, we need to look at  $\mathbf{E}$  as a three dimensional matrix, with one dimension for the reactions, one for the enzymes, and one for the time points. We therefore talk for example about variables like  $\mathbf{E}_{r_1}^{e_1}(t)$ . At this point it becomes even more clear that the functions *toStruct* and *toVector* are crucial for avoiding implementation mistakes.

### 5.6.2 Constraint matrix

The constraint matrix of a deFBA (or cFBA) problem contains several constraint types:

- differential equations equality constraint for external metabolites (e.g. (5.3)) and macromolecules (e.g. (5.6)-(5.7))
- steady-state equality constraint for internal metabolites (e.g. (5.4)-(5.5))
- enzyme capacity inequality constraint (e.g. (5.8)-(5.9))
- discretization equality constraint for relating  $\dot{\mathbf{n}}$  to  $\mathbf{n}$  (e.g. (5.12))
- inequality constraint to impose quota macromolecule production (e.g. (5.11)).

For ease of implementation and debugging, we recommend that a function is written for constructing the submatrix of the constraint matrix and the subvector of the right hand side corresponding to each type of constraint. Then these submatrices and subvectors can be merged and passed to the solver. In a similar fashion, a function should be written to build the objective vector.

To build all these constraint submatrices, a key element is a function that, given a variable type (e.g.  $\mathbf{v}$ ), the reaction (resp. component) index, and the time point returns the index of this variable in the solution vector (*getIndex*).

## Chapter 5. Metabolic resource allocation in practice: numerical concerns and solving strategies

---

### 5.6.3 Bounds on variables

Using the function *getIndex*, positivity constraints and bounds on irreversible or exchange reactions can be formulated for each time point. In a similar fashion, dynamic bounds as the light uptake in chapter 6 can be formulated.

### 5.6.4 Shift experiments

In case one wants to perform nutrient shift experiments as in chapter 7, some of the constraints need to be adjusted. For example, assume that at time point 3 one nutrient source  $y_1$  is removed. Obviously, the constraint

$$\mathbf{n}_{y_1}(3) = \mathbf{n}_{y_1}(2) + h \cdot \dot{\mathbf{n}}_{y_1} \quad (2.5)$$

no longer holds. In such cases, this discretization constraint needs to be removed. In addition, the constraint that  $\mathbf{lb}_{y_1}(3) = \mathbf{ub}_{y_1}(3) = 0$  needs to be added, where  $\mathbf{lb}$ ,  $\mathbf{ub}$  are the lower and upper bound vectors of the LP.

The best way to implement such constraints is to have a structure as part of the model that stores the variable type ( $\mathbf{n}$  in this case), the index of the shifted compound (index of  $y_1$ ) and the time point at which the shift occurs (3 in this case). This structure then has one field for each shift.

Shifts in nonlimiting extracellular metabolites (e.g.  $O_2$  in yeast), for which we do not model an explicit amount but instead the amount taken up is only bounded by the turnover rate and the transporter amount, are easier to implement since we do not need to drop any constraint, but only impose flux bounds at the respective time points.

### 5.6.5 Short term deFBA

In case one wants to implement the short term deFBA (Lindhorst et al., 2016), which we have introduced briefly in section 5.1.3, a loop around a typical deFBA run is sufficient. At each step of the loop a deFBA problem is solved over the given prediction horizon ( $t_f$ ) and the values of the first time point of the solution are stored and passed on as initial values to the next iteration. A way of finding the minimum prediction horizon such that exponential growth is preferred to linear growth (cf. section 5.1.2) is described in (Lindhorst et al., 2016).

In case shift experiments are implemented, with each iteration the shift time point comes “closer” and the time point values for the shifts need to be updated also.

## Chapter 6

# Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

The work presented in this chapter has been done in collaboration with Henning Knoop and Ralf Steuer. This work has been published in Proceedings of the National Academy of Sciences (PNAS) under (Reimers et al., 2017a) and can be found at <http://dx.doi.org/10.1073/pnas.1617508114>.

Cyanobacteria are an integral part of the Earth's biogeochemical cycles and a promising resource for the synthesis of renewable bioproducts from atmospheric CO<sub>2</sub>. Growth and metabolism of cyanobacteria are tied to the diurnal rhythm of light availability. So far, however, insight into the stoichiometric and energetic constraints of cyanobacterial diurnal growth is limited. In this chapter we investigate the optimal allocation of cellular resources during diurnal phototrophic growth using a genome-scale metabolic reconstruction of the cyanobacterium *Synechococcus elongatus* PCC 7942. We formulate phototrophic growth as an autocatalytic process and solve the resulting time-dependent resource allocation problem using cFBA. Based on a narrow and well-defined set of parameters, we obtain an *ab initio* prediction of growth properties over a full diurnal cycle. The computational model allows us to study the optimality of metabolite partitioning during diurnal growth. The cyclic pattern of glycogen accumulation, an emergent property of the model, has timing characteristics that are in qualitative agreement with experimental findings. Our model gives insight into the dynamic resource allocation problem of phototrophic diurnal growth. Furthermore, it serves as a general framework

## Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

---

to assess the optimality of metabolic strategies that evolved in phototrophic organisms under day-night conditions.

### 6.1 Introduction

Cyanobacterial photoautotrophic growth requires a highly coordinated distribution of cellular resources to different intracellular processes, including the *de novo* synthesis of proteins, ribosomes, lipids, and other cellular components. For unicellular organisms, the optimal allocation of limiting resources is a key determinant of evolutionary fitness. The protein economy and its implications for bacterial growth laws have been studied extensively. However, this was done almost exclusively for heterotrophic organisms under stationary environmental conditions (Molenaar et al., 2009; Scott et al., 2010; Flamholz et al., 2013; Vázquez-Laslop and Mankin, 2014; Hui et al., 2015; Burnap, 2015; Weiße et al., 2015).

For photoautotrophic organisms, including cyanobacteria, growth-dependent resource allocation is further subject to light-dark cycles that partition cellular metabolism into distinct phases. Recent experimental results have demonstrated the relevance of time-specific synthesis for cellular survival and growth (Shultzaberger et al., 2015; Diamond et al., 2015; Lambert et al., 2016). Nonetheless, the consequences of a day-night environment on the cellular resource allocation problem are insufficiently understood, and computational approaches developed for heterotrophic organisms are not straightforwardly applicable to diurnal phototrophic growth (Westermarck and Steuer, 2016).

Here, we use dynamic optimization to study optimal diurnal resource allocation for phototrophic growth. We are mainly interested in the stoichiometric and energetic constraints that shape the relationship between the maximal growth rate and the relative partitioning of metabolic, photosynthetic, and ribosomal proteins during a full day-night cycle. We aim to obtain a prediction of the properties that arise from a narrow and well-defined set of parameters and assumptions about cyanobacterial growth – and to contrast these properties with experimentally observed behavior.

We assemble and numerically evaluate an autocatalytic genome-scale model of cyanobacterial growth, based on a metabolic reconstruction of the cyanobacterium *Synechococcus elongatus* PCC 7942, which has been provided by Henning Knoop. Our model significantly improves upon previous computational analyses of diurnal phototrophic growth (Knoop et al., 2013; Cheung et al., 2014; Knies et al., 2015; Rügen et al., 2015). Key aspects of the model are depicted in figure 6.1.

Our main question is: What are the rate and temporal order of synthesis reactions such that we obtain maximal growth of a cyanobacterial cell in a day-night



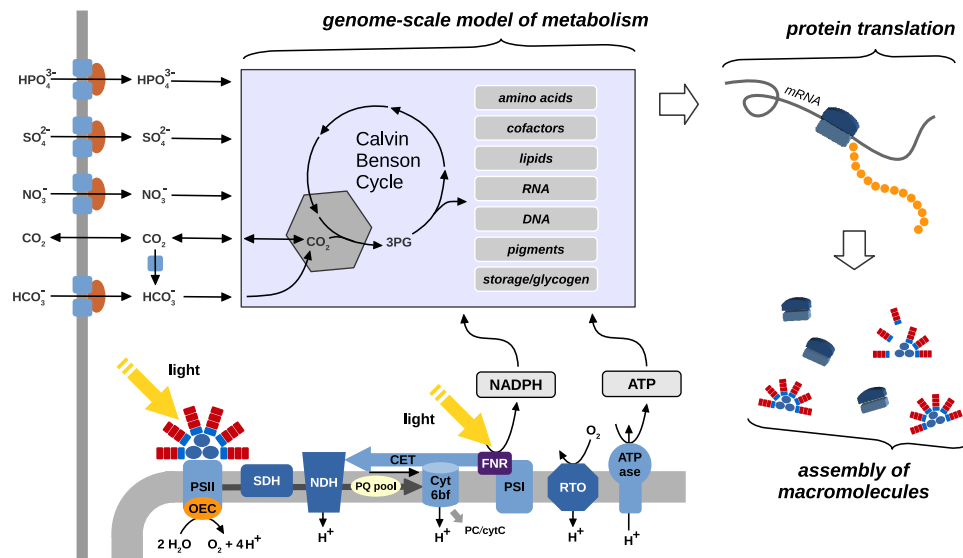


Figure 6.1: **An autocatalytic growth model of *Synechococcus elongatus* PCC 7942.** Energy and reducing agents from the photosynthetic light reactions drive the fixation of inorganic carbon via the Calvin-Benson cycle, as well as the subsequent synthesis of cellular macromolecules. The synthesis of macromolecules is modeled as described in chapter 4. The capacity of each metabolic reaction depends on the availability of its catalyzing enzymes. Enzymes are translated using their amino acids, which are themselves the products of metabolism. The abundances of all macromolecules relevant to cellular growth (enzymes, transporters, photosynthetic and respiratory protein complexes, phycobilisomes, and ribosomes) are time-dependent quantities that are governed by differential equations, as explained in chapter 2.

environment? To answer this, we build a resource allocation model of *Synechococcus elongatus* PCC 7942 as explained in chapter 4. Our key results include:

- a prediction of the timing of intracellular synthesis reactions that is in good agreement with experimental observations about metabolite partitioning during diurnal growth,
- limits on the estimated maximal rate of phototrophic growth that are close to observed experimental values, suggesting a highly optimized metabolism,
- a predicted optimal timing of glycogen accumulation that is in good agreement with recent experimental data.

## **6.2 Model building**

### **6.2.1 The genome-scale reconstruction**

The genome-scale reconstruction provides a manually curated stoichiometric description of all metabolic reactions relevant to cellular growth: Photons are absorbed by light-harvesting antennae, the phycobilisomes, attached primarily to photosystem II (PSII). The energy derived from absorbed photons drives water splitting at the oxygen-evolving complex (OEC) and, via the photosynthetic electron transport chain (ETC), regenerates cellular ATP and NADPH. The ETC consists of a set of large protein complexes, PSII, the cytochrome  $b_6f$  complex (Cyt $b_6f$ ), photosystem I (PSI), and ATP synthase (ATPase), embedded within the thylakoid membrane.

The metabolic network was reconstructed by Henning Knoop using the reconstruction of *Synechocystis* sp. PCC 6803 (Knoop et al., 2013) as a scaffold. Main differences to the reconstruction of *Synechocystis* sp. PCC 6803 are:

- a smaller genome size, 3.57 megabases for *Synechocystis* sp. PCC 6803 versus 2.8 megabases for *Synechococcus elongatus* PCC 7942
- no known tocopherol synthesis
- no known PHB and cyanophycin pathways
- no known echinenone synthesis (carotenoid)
- no known delta 6 and 15 fatty acid desaturases
- no annotated urea metabolism
- methionine synthesis is annotated
- only alternative synthesis pathway for branched chain amino acids (Wu et al., 2010)
- incomplete TCA cycle that cannot operate in cyclic mode
- neither the bypass of (Zhang and Bryant, 2011), nor the GABA shunt (Knoop et al., 2013) is annotated
- malate dehydrogenase and glutamate dehydrogenase are not annotated

### **6.2.2 The macromolecules of autocatalytic growth**

Given the metabolic network reconstruction, we use the gene-reaction mapping together with the sequence annotations to describe the production of each metabolic enzyme from the metabolic network, as detailed in chapter 4.

#### **Ribosomes**

Ribosome synthesis is modeled analogously to enzyme production. We assembled a list of the ribosomal proteins, their corresponding genes, and the ribosomal RNA, based on the KEGG resource. Table A1 in Appendix C provides the ribosome composition of *Synechococcus elongatus* PCC 7942.

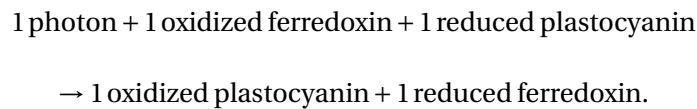
The ribosome translation rate of *Synechococcus elongatus* PCC 7942 has not been measured directly, we therefore assume a rate similar to that of *Escherichia coli*, namely 15 amino acids per second (Young and Bremer, 1976).

### Synthesis of photosystems and the electron transport chain

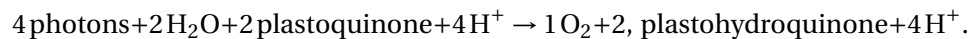
The photosynthetic electron transport chain consists of a number of protein complexes. Their location is of lesser importance for the resource allocation problem, so only the ETC of the thylakoid membrane is considered. The protein complexes of the ETC constrain flux through the ETC analogously to metabolic enzymes that constrain biochemical flux.

#### The photosynthetic electron transport chain

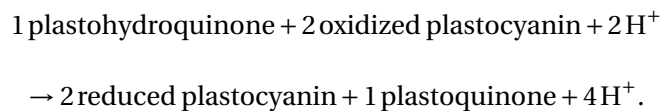
The reactions of photosystem I have been merged together into a single overall reaction:



Similarly, the reactions of photosystem II have been merged together into the overall reaction:



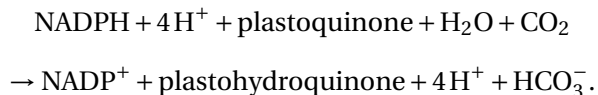
The reactions of the cytochrome b<sub>6</sub>f complex have been merged together into:



The NADPH dehydrogenase complex (NDH I) is known to participate in a variety of reactions within respiration, cyclic electron transport around PSI and CO<sub>2</sub> uptake (Ma and Ogawa, 2015). Its precise role is still not fully understood. In our model, NDH I catalyzes the following two reactions:



and



The gene compositions of PSI, PSII, NDH I and Cytb<sub>6</sub>f and the corresponding stoichiometries are provided in tables A2-A4 in Appendix C. We note that pigments are necessary compounds for the synthesis of the photosystems. For pigments whose stoichiometries are not known, a separate quota metabolite is included to enforce their presence in the model biomass.

## Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

---

### Phycobilisomes

Phycobilisomes (PBS) are protein complexes that act as light harvesting antennae (Thomas et al., 1993). The captured energy is transferred to the PSII chlorophyll. In *Synechococcus elongatus* PCC 7942, phycobilisomes are essential for the correct functioning of PSII (Bhalerao et al., 1995). PBS attach and detach from the photosystems and in this way regulate how much light is absorbed and transferred to the photosystems (Liu et al., 2013). Since PBS are assumed to be predominantly associated with PSII, we only consider energy transfer from PBS to PSII.

From a structural perspective, the phycobilisomes are made of two cylinders that form the core and six light harvesting antennae of variable length (Bhalerao et al., 1995; Campbell et al., 1998). The antenna length influences the efficiency of the phycobilisome: the longer the antenna, the more efficient the light harvesting. In times of high light intensity, the antennae can be shortened and the proteins that belonged to them are degraded back into individual amino acids.

Thus, we model the photosystem II and phycobilisomes in individual states, according to how long the antennae are. We consider as base state a PSII to which the core of the phycobilisome is attached. Cyanobacteria with such phycobilisomes are able to survive at very low growth rates (Bhalerao et al., 1995). We then model transitions to other states where the antennae size is increased by reactions that “consume” the respective proteins and produce a complex with longer antennae. The respective transitions are shown in figure 6.2. The gene composition of a phycobilisome is provided in table A3 in Appendix C.

### Respiratory chain

The gene composition of ATPase, Cytochrome c oxidase, and succinate dehydrogenase are detailed in table A4 in Appendix C.

The model encompasses a total of 465 macromolecules and 1112 reactions, including 645 metabolic and exchange reactions, 616 metabolic genes, as well as 467 reactions describing the synthesis of macromolecules.

## 6.3 Model constraints and objective

Before we list all the constraints, we would like to make two observations:

- **Glycogen** is the main storage component of cyanobacteria. It is typically accumulated during the day and consumed over the night as energy source to ensure survival. In our model we denote it by  $G$  and allow it to accumulate.
- **Ribosomes**, denoted by  $R$ , are considered as enzymatic component, and thus  $R \in \mathcal{E}$ .

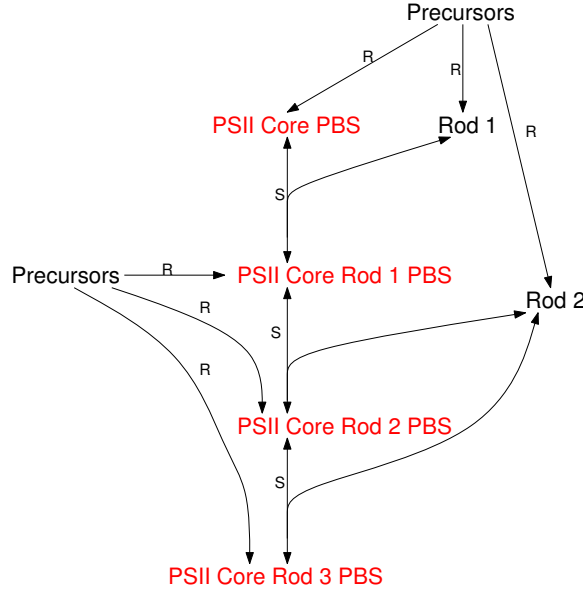


Figure 6.2: PSII-phycolibosome complexes and the transitions between them. The PSII-phycolibosome complex reactions labeled with R are assumed to be catalyzed by the ribosome, while reactions labeled with S are assumed to be spontaneous. The following abbreviations are used: PSII Core PBS - Complex of photosystem II and core phycobilisome, PSII Core Rod 1 PBS - Complex of photosystem II, core phycobilisome, and 6 antennae of length 1, PSII Core Rod 2 PBS - Complex of photosystem II, core phycobilisome, and 6 antennae of length 2, PSII Core Rod 3 PBS - Complex of photosystem II, core phycobilisome, and 6 antennae of length 3, Rod 1 - rod protein made of phycocyanin, CpcA, CpcB and CpcG, Rod 2 - rod protein made of phycocyanin, CpcA, CpcB and CpcC.

### 6.3.1 Constraints

#### Steady-state versus time-dependent quantities

As in described in chapter 2, we assume metabolism to be at steady-state, whereas amounts of macromolecules are time-dependent quantities. In particular, every metabolite  $\mathcal{M}_i$  is assumed to be produced at all time points at the same rate as it is consumed. Therefore, we obtain the constraint

$$\frac{d\mathbf{n}_{\mathcal{M}}(t)}{dt} = S_{\mathcal{M}*}\mathbf{v}(t) = 0, \quad (6.1)$$

for all internal metabolites at all time points  $t$ .

Enzymes and ribosomes have to be synthesized by cellular metabolism and their dynamics are governed by a system of differential equations at all time points  $t$ ,

$$\frac{d\mathbf{n}_{\mathcal{E}}(t)}{dt} = S_{\mathcal{E}*}\mathbf{v}(t). \quad (6.2)$$

## Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

---

Similarly, quota metabolites, which we denote by the set  $\mathcal{Q}$ , have to be synthesized using precursors from cellular metabolism, therefore all time points  $t$  we have

$$\frac{d\mathbf{n}_{\mathcal{Q}}(t)}{dt} = S_{\mathcal{Q}*}\mathbf{v}(t). \quad (6.3)$$

To account for basic cell maintenance in the absence of light, we allow the glycogen  $G$  to accumulate and be consumed. Therefore, glycogen amount is allowed to vary and obeys the differential equation

$$\frac{d\mathbf{n}_G(t)}{dt} = S_{G*}\mathbf{v}(t). \quad (6.4)$$

So, putting all differential equations together, we obtain

$$\frac{d\mathbf{n}_i(t)}{dt} = S_{i*}\mathbf{v}(t), \quad \forall i \in \mathcal{E} \cup \mathcal{Q} \cup G. \quad (6.5)$$

### Enzyme amounts constrain reaction rates

Enzyme amounts constrain reaction rates within the metabolic network as explained in chapter 2. Turnover numbers for metabolic enzymes are retrieved from the BRENDA database (Schomburg et al., 2013). A known problem in this context is the fact that recorded values for a specific enzyme spread over several orders of magnitude. In our model we deal with this problem by using the median value of all wild type turnover numbers reported for an enzyme. For enzymes with no annotated values for  $k_{cat}$ , we follow the strategy used by (Shlomi et al., 2011): we use the median value of all the known wild type turnover rates we found as  $k_{cat}$ . An alternative strategy to deal with unknown  $k_{cat}$  values has also been provided in section 4.4.

Turnover numbers for the 7 macromolecules of the ETC are sourced from the primary literature and listed in table 6.1.

Table 6.1: Turnover numbers sourced from the literature for the macromolecules of the electron transport chain.

Compound	Catalytic efficiency	Reference
PSI	500 s <sup>-1</sup>	(Vermaas, 2001)
PSII	1000 s <sup>-1</sup>	(Vermaas, 2001)
NDH-1	130 s <sup>-1</sup>	(Teicher and Scheller, 1998)
Cytb6f	200 s <sup>-1</sup>	(Vermaas, 2001)
Cyt c oxidase	670 s <sup>-1</sup>	(Howitt and Vermaas, 1998)
SDH	1300 s <sup>-1</sup>	(Cooley and Vermaas, 2001)
ATPase	1000 s <sup>-1</sup>	(Nitschmann and Peschek, 1986)

### 6.3 Model constraints and objective

For irreversible enzyme-catalyzed reactions, our constraint then reads

$$\mathbf{v}_j(t) \leq k_{cat}^j \mathbf{n}_{\mathcal{E}_j}(t), \quad (6.6)$$

for all irreversible reactions  $j$  at all time points  $t$ . In the case of reversible reactions, both directions are constrained

$$\mathbf{v}_j(t) \leq k_{cat}^{j+} \mathbf{n}_{\mathcal{E}_j}(t), \quad \mathbf{v}_j(t) \geq -k_{cat}^{j-} \mathbf{n}_{\mathcal{E}_j}(t), \quad (6.7)$$

where  $k_{cat}^{j+}$  and  $k_{cat}^{j-}$  is the turnover rates for the forward and reverse direction respectively. We impose these two constraints at each time point  $t$  for each reversible reaction  $j$ . The constraints apply only for enzyme-catalyzed reactions. Rates of spontaneous reactions remain free of these bounds.

In case several reactions are catalyzed by the same enzyme, their total flux weighted by the inverse of the respective turnover numbers is bound by the enzyme amount. Such a situation happens, for instance, in the case of the ribosome, as detailed in section 2.2.2.

Note that enzyme production reactions are irreversible and hence the bound is only applied for the forward direction of the reaction, while the reverse direction is prohibited using a lower flux bound of zero at all time points.

#### Amounts of quota compounds

The model includes eight quota compounds: DNA, RNA, cell wall, pigments, noncatalytic proteins, lipids, cofactors and vitamins, ions. These are modeled dynamically and their initial values are set to be equal to their corresponding amounts in 1 gram dry weight of *Synechococcus elongatus* PCC 7942 cells and displayed in table A5 in Appendix C. We denote their initial amounts as the vector  $\mathbf{q}_0$ , and therefore we require that

$$\mathbf{n}_{\mathcal{Q}}(t_0) = \mathbf{q}_0. \quad (6.8)$$

We additionally impose that, together with metabolic proteins and glycogen, the initial composition vector adds up to one gram, and thus obtain the additional constraint

$$\sum_{i \in \mathcal{Q}} \mathbf{n}_i(t_0) + \sum_{i \in \mathcal{E}} \mathbf{n}_i(t_0) \cdot MW_i + \mathbf{n}_G(t_0) = 1. \quad (6.9)$$

We note that protein amounts are expressed in *mmol*, while the glycogen and the quota amounts are expressed in *grams*. This why in the constraint above we multiply the enzyme amounts with their molecular weights  $MW_i$ .

## Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

---

### Adjustment of pigment quota

Chlorophyll,  $\beta$ -carotene and phylloquinone are important ingredients of the photosystems. Therefore, the reactions building the photosystems already incorporate these pigments. To account for this, the general pigment quota requirements have to be adjusted. Without the adjustment, the original biomass requirements are

$$\begin{aligned} &0.841 \text{ mmol Chlorophyll a} + 0.136 \text{ mmol } \beta\text{-Carotene} \\ &+ 0.321 \text{ mmol Zeaxanthin} + 0.064 \text{ mmol } \gamma\text{-Carotene} \\ &+ 0.068 \text{ mmol Phylloquinone} \rightarrow 1 \text{ g Pigment,} \end{aligned}$$

and the initial pigment quota would be 0.0244 g.

Since chlorophyll a,  $\alpha$ -carotene and phylloquinone are no longer part of the quota compounds, we identify the factor  $f$  such that

$$f \cdot 0.321 \text{ mmol} \cdot MW_{\text{Zeaxanthin}} + f \cdot 0.064 \text{ mmol} \cdot MW_{\gamma\text{-Carotene}} = 1 \text{ g Pigment}$$

and then change the initial pigment quota to  $\frac{0.0244}{f}$  g. We obtain  $f = 4.609$ , and therefore the pigment quota formation equation becomes

$$1.479 \text{ mmol Zeaxanthin} + 0.295 \text{ mmol } \gamma\text{-Carotene} \rightarrow 1 \text{ g Pigment,}$$

and the initial pigment quota is 0.0053 g.

### Periodicity of the system

We consider diurnal growth as a periodic system. Hence, we enforce that all macromolecule amounts at the end of the time period are multiples of their amounts at the beginning of the time period. We thus obtain the constraint

$$\alpha \cdot \mathbf{n}(t_0) = \mathbf{n}(t_f). \quad (6.10)$$

This constraint ensures balanced growth of the whole system as already described by (Rüger et al., 2015).

### Non-catalytic proteins and constraints on initial protein amounts

To account for proteins without catalytic activity within our model, we include additional proteins as a quota compound, the synthesis of which is also catalyzed by the ribosome.

According to previous quantitative proteomics data by (Guerreiro et al., 2014), the proteins included in our model, make up a fraction of 45% of the total proteome of *Synechococcus elongatus* PCC 7942. In the original biomass reaction of



the metabolic model, proteins make up 0.51 g in a gram dry weight of cells. Of these, 45% are catalytic proteins, and the remaining 55% represent quota proteins.

The model is able to choose the initial distribution of catalytic proteins, but needs to obey constraint (6.9). For fast growing cells of *Synechococcus elongatus* PCC 7942 no experimental estimates are available. We conjecture that the quota of non-catalytic proteins for fast growing *Synechococcus elongatus* PCC 7942 is significantly lower than 55%. Growth rate increases significantly with a decreasing amount of quota compounds (see figure 6.7 and discussion). We note that minimal models of (heterotrophic) cellular growth typically also include a growth-independent fraction of protein, typically of the order of 50% (Scott et al., 2014).

#### Light uptake

Light availability is modeled by a half-wave rectified sine function that mimics the day-night cycle

$$l(t) = \begin{cases} l_{max} \sin\left(\frac{2\pi t}{t_f}\right) & \text{if } \sin\left(\frac{2\pi t}{t_f}\right) \geq 0 \\ 0 & \text{else,} \end{cases} \quad (6.11)$$

where  $t_f = 24$  h and  $l_{max}$  the maximum light intensity that occurs  $t_f/4$  hours after dawn.

The amount of light absorbed by the system is proportional to the combined amount of photosystem I and photosystem II-phycobilisome complexes at the respective time point multiplied by their respective effective cross sections:

$$\mathbf{v}_{P700}(t) \leq \sigma_{PSI} \cdot \mathbf{n}_{PSI}(t) \cdot l(t) \quad (6.12)$$

$$\mathbf{v}_{P680}(t) \leq \sigma_{PSII} \cdot \mathbf{n}_{PSII}(t) \cdot l(t), \quad (6.13)$$

where  $\mathbf{v}_{P700}$  and  $\mathbf{v}_{P680}$  are the fluxes of absorbed photons respectively, and  $\sigma_{PSI}$  and  $\sigma_{PSII}$  are the cross sections of photosystems-phycobilisome complexes. We do not distinguish between photons of different wavelength.

The cross section of PSI is assumed to be equal to  $0.5 \text{ nm}^2$ , independent of the status of phycobilisomes. The cross section of PSII depends on the length of the rods of its attached phycobilisome. A PSII with only the core of the phycobilisome is assumed to have a cross section of  $0.1 \text{ nm}^2$ , if the rods of the phycobilisome have length one, then the cross section is  $0.33 \text{ nm}^2$ , rods of length two give a cross section of  $0.67 \text{ nm}^2$ , and full length rods give a cross section for PSII of  $1 \text{ nm}^2$  (Mackenzie et al., 2004). The effective cross sections only affect absorbed light versus incoming light intensity and do not qualitatively affect simulation results.

## Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

---

### Maintenance

Similar to conventional FBA models, we assume that there are other processes that require energy and are not considered by our model. Therefore, the model contains a non-growth associated maintenance reaction that hydrolyzes ATP into ADP and  $P_i$ . We enforce a lower bound of  $0.13 \text{ mmol}\cdot\text{h}^{-1}$  for the flux through this reaction at each time point in order to account for energy consumption of general maintenance.

### Discretization of time points across the diurnal cycle

As detailed in chapter 5, we discretize time using the implicit midpoint rule. Derivatives  $\dot{\mathbf{n}}$  and fluxes  $\mathbf{v}$  are evaluated at the middle of the discretization intervals, while macromolecule amounts  $\mathbf{n}$  are evaluated at the ends of the discretization intervals.

#### 6.3.2 The optimization objective

As optimization objective, we assume that the cell has evolved to grow as much as possible within a full diurnal period, that is, the regulatory system has evolved such that the multiplication factor  $\alpha$  involved in constraint (6.10) is maximal. Even if this assumption of optimality turns out to be incorrect, the optimal solution with respect to the assumption is still of high interest to compare with experimentally observed behavior. It is only by knowledge of optimal solutions of the resource allocation problem that suboptimal behavior, or incorrect assumptions and parameters, can be identified.

We start our simulation with 1 gram dry weight and track the changes in the cellular composition and growth over day-night cycles.

### 6.3.3 The quadratic program and binary search

The optimization problem is given by

$$\begin{aligned}
 & \max_{\alpha, \mathbf{v}, \mathbf{n}, \dot{\mathbf{n}}} \alpha \\
 & \text{s.t. } S_{\mathcal{M}*} \mathbf{v} \left( \frac{t_j + t_{j-1}}{2} \right) = 0, \\
 & \dot{\mathbf{n}}_i \left( \frac{t_j + t_{j-1}}{2} \right) = S_{i*} \mathbf{v} \left( \frac{t_j + t_{j-1}}{2} \right), \quad \forall i \in \mathcal{E} \cup \mathcal{Q} \cup \mathcal{G}, \\
 & \sum_{i \in \mathcal{V}_k} \frac{\mathbf{v}_i \left( \frac{t_j + t_{j-1}}{2} \right)}{k_{cat}^{i+}} \leq \frac{\mathbf{n}_k(t_j) + \mathbf{n}_k(t_{j-1})}{2}, \quad \forall k \in \mathcal{E}, \\
 & - \sum_{i \in \mathcal{V}_k \setminus \text{Irr}} \frac{\mathbf{v}_i \left( \frac{t_j + t_{j-1}}{2} \right)}{k_{cat}^{i-}} \leq \frac{\mathbf{n}_k(t_j) + \mathbf{n}_k(t_{j-1})}{2}, \quad \forall k \in \mathcal{E}, \\
 & \mathbf{n}_{\mathcal{Q}}(t_0) = \mathbf{q}_0, \\
 & \alpha \mathbf{n}(t_0) = \mathbf{n}(t_f), \\
 & \sum_{i \in \mathcal{Q}} \mathbf{n}_i(t_0) + \sum_{i \in \mathcal{E}} \mathbf{n}_i(t_0) \cdot \text{MW}_i + \mathbf{n}_G(t_0) = 1, \\
 & \mathbf{v}_{\text{light}} \left( \frac{t_j + t_{j-1}}{2} \right) \leq l \left( \frac{t_j + t_{j-1}}{2} \right), \\
 & \mathbf{v}_{P700} \left( \frac{t_j + t_{j-1}}{2} \right) \leq \sigma_{PSI} \cdot \frac{\mathbf{n}_{PSI}(t_j) + \mathbf{n}_{PSI}(t_{j-1})}{2} \cdot l \left( \frac{t_j + t_{j-1}}{2} \right), \\
 & \mathbf{v}_{P680} \left( \frac{t_j + t_{j-1}}{2} \right) \leq \sigma_{PSII} \cdot \frac{\mathbf{n}_{PSII}(t_j) + \mathbf{n}_{PSII}(t_{j-1})}{2} \cdot l \left( \frac{t_j + t_{j-1}}{2} \right), \\
 & \mathbf{v}_{\text{maintenance}} \left( \frac{t_j + t_{j-1}}{2} \right) \geq 0.13, \\
 & \mathbf{n}(t_j) = \mathbf{n}(t_{j-1}) + (t_j - t_{j-1}) \cdot \dot{\mathbf{n}} \left( \frac{t_j + t_{j-1}}{2} \right), \\
 & \mathbf{v}_{\text{Irr}} \left( \frac{t_j + t_{j-1}}{2} \right), \mathbf{n}(t_j) \geq 0.
 \end{aligned}$$

for all  $j \in \{1, \dots, N\}$ , where  $\mathcal{V}_k$  denotes the set of reactions that are catalyzed by enzyme  $k$ , and  $N$  is the number of discretization points.

As pointed out in chapter 5, this formulation of the problem gives rise to discretization errors, so in the implementation of the problem the dilution formulation described in section 5.2.3 has been used.

We notice that, because  $\alpha$  and  $\mathbf{n}$  are both variables in the model, our program contains a quadratic constraint, namely the one that guarantees the periodicity of the system. Since we aim to maximize  $\alpha$ , we run a binary search and for each new value of  $\alpha$  we test the feasibility of the resulting linear program as in (Goelzer et al., 2011; O'Brien et al., 2013; Rügen et al., 2015).

## Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

---

Due to the numerical condition of our problem, we chose to solve the individual feasibility linear programs using the SoPlex 2.2.1 optimization package (Wunderling, 1996; Gleixner et al., 2012, 2016), as explained in chapter 5.

All code is provided at

<https://sourceforge.net/projects/cfba-synpcc7942/>.

### 6.3.4 Further model specifications

- **Growth rate and multiplication factor:** As a reminder, relationships between growth rate  $\mu$  (unit  $h^{-1}$ ), the multiplication factor  $\alpha$  (unitless), and the division time  $T_D$  (unit  $h$ ) are

$$\alpha = \exp(\mu \cdot 24h), \quad \mu = \frac{\ln \alpha}{24h}, \quad T_D = \frac{\ln(2)}{\mu}. \quad (6.14)$$

- **Light intensity and the cross sections of photosystems:** Figure 6.3 indicates that the values used as photosystem cross section and sourced from (Mackenzie et al., 2004) are too low. However, since a modification of parameters with hindsight would violate our aim of an *ab initio* prediction of emergent properties, we decided to keep figure 6.3 unchanged. However, the absolute value of the incoming light intensity only impacts the model via the effective cross sections. A change of the cross sections results in a shift of incoming light intensity, with no further impact on any simulations or model-derived property. Since the cross section is quadratic as a function of diameter, small changes in the effective diameter may result in significant changes with respect to the saturating light intensity.
- **Constraints and range of applicability:** The evaluation of our model is based on the assumption of a stationary culture in a periodic environment. This equation primarily holds on the culture level. That is, under stationary conditions, we expect average cell composition to be invariant with respect to a full diurnal cycle (either in a turbidostat setting or via serial dilution).

However, equation (6.10) must not necessarily hold for an individual cell. Nonetheless, equation (6.10) is still a valid assumption for our analysis, based on the following arguments: Firstly, cyanobacterial growth happens on diurnal time scales. Typical division times are approximately 24h. Faster rates, up to  $2.5 - 3h$  are only observed under highly optimized conditions. Secondly, our main interest are (metabolic) synthesis reactions related to diurnal growth. We conjecture that such cellular temporal programs go beyond the timespan of a single cell cycle. Indeed, it has been shown that the phase of cellular oscillations persists also after division events (Elowitz and Leibler, 2000). It is therefore reasonable to assume

that an evolved temporal metabolic program to optimize resource allocation reflects the external (light) conditions, according to equation (6.10), rather than, for example, an individual cell cycle. Our assumption implies that a cell has evolved to synthesize glycogen according to the global resource allocation problem considered herein, even though dusk might happen only after 1 – 2 division events.

We do not expect our analysis to capture cellular resource allocation for very slow division times. While equation (6.10) certainly remains applicable, other constraints than the energetic implications of *de novo* protein synthesis, such as protein turnover and repair, become relevant, and eventually dominant. Such additional constraints may be included within the model, but are outside the scope of the current analysis.

- **Possible improvements of the model:** We consider our analysis to be a reasonable first installment to evaluate the energetic and stoichiometric implications of diurnal phototrophic growth. Nonetheless, the model allows for a number of improvements to evaluate specific environmental conditions in future analysis. In particular, carbon limitation might be considered which requires a more detailed representation of carbon cycling processes and the carboxysome. Right now the cost of carbon cycling is part of general maintenance. Likewise, limitations of other factors, in particular nitrogen, may be included. We also expect that a more detailed representation of photodamage, as a result of high light intensity, should be considered in future installments of the model. In each case, quantitative information about the respective processes exists. A particular challenge, however, is to formulate the respective processes such that a solution of the respective dynamic optimization problem remains computationally feasible.

## 6.4 Results

### 6.4.1 Growth under constant light

Prior to evaluating diurnal dynamics, we investigate light-limited growth under constant light. Our aim is to compare the results to conventional flux balance analysis and thereby to verify the consistency of the model. Solving the global resource allocation problem, we obtain the multiplication factor  $\alpha$  and the growth rate  $\mu = \log(\alpha)/24h$  as a function of the light intensity, as well as the cellular composition for different growth rates. Key results are shown in figure 6.3.

To compare model properties with previous results obtained using FBA, we use a reference light intensity  $I = 150 \mu\text{mol photons } s^{-1} m^{-2}$  resulting in the absorption of  $15.9 \text{ mmol photons } gDW^{-1} h^{-1}$ , a growth rate of  $\mu = 0.03 h^{-1}$  (multiplication factor  $\alpha \approx 2$ ), and an oxygen evolution rate of  $1.92 \text{ mmol } gDW^{-1} h^{-1}$ . These values are in quantitative agreement with values previously estimated using FBA (Nogales et al., 2012; Knoop et al., 2013). In particular, performing a

## Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

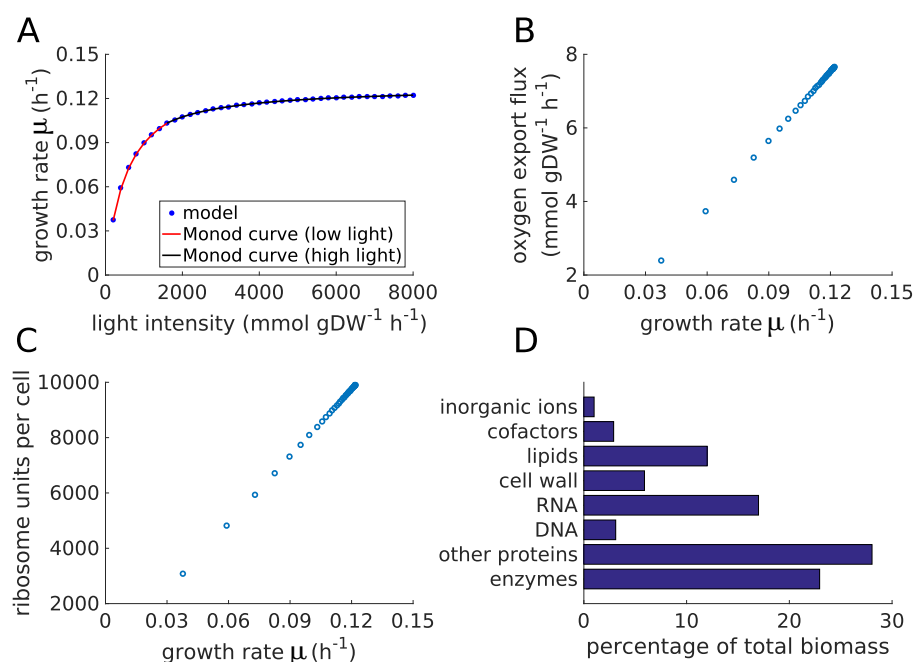


Figure 6.3: **Balanced growth under constant light.** A: The resulting growth rate  $\mu$  as a function of light intensity is consistent with a Monod growth law. B: Oxygen evolution as a function of the growth rate  $\mu$ . C: Ribosome content per cell as a function of  $\mu$ , assuming a cell dry mass of  $1.5 \text{ pg}$ . D: The cellular composition for a light intensity of  $150 \mu\text{mol photons } m^{-2} s^{-1}$ .

conventional FBA on the metabolic reconstruction of *Synechococcus elongatus* PCC 7942 using a static biomass objective function (BOF) and a light uptake of  $15.8 \text{ mmol absorbed photons } gDW^{-1} h^{-1}$  results in an oxygen evolution rate of  $1.92 \text{ mmol } gDW^{-1} h^{-1}$  and a growth rate of  $\mu = 0.03 \text{ h}^{-1}$ . We note that quantitative agreement between results obtained from the resource allocation problem and conventional FBA cannot be expected *a priori*—as the former are also based on capacity constraints induced by kinetic parameters whose values are globally sourced from databases. In contrast to the static pre-assigned BOF used in FBA, also the cellular composition of the autocatalytic model is an emergent result of the global resource allocation problem (figure 6.3D). The allocated abundance of catalytic macromolecules is in quantitative agreement with previously reported BOFs (Nogales et al., 2012; Knoop et al., 2013).

### 6.4.2 Adaptations to different light intensities

When solving the model for different (but constant) light intensities, the growth rate, as well as the oxygen evolution rate, increase with increasing light intensity (figures 6.3A and 6.3B). We note that light uptake depends on the cross section of PSII, reported to be  $\sigma_{PSII} \approx 1 \text{ nm}^2$  (Mackenzie et al., 2004). The results

shown in figure 6.3 indicate that the reported value significantly underestimates the actual effective cross section. Similar to results obtained for models of heterotrophic growth (Molenaar et al., 2009; Scott et al., 2010; Weiße et al., 2015), the relative amount of ribosomes linearly increases with increasing growth rate (figure 6.3C). We observe that growth as a function of light saturates at a growth rate of  $\mu_{max} = 0.1281 \text{ h}^{-1}$  (multiplication factor  $\alpha \approx 18$ , doubling time  $T_d \approx 5.4 \text{ h}$ ), estimated using a Monod growth equation (figure 6.3A). The maximal doubling time obtained from the model is slightly lower but still within the range of the fastest published doubling time of *Synechococcus elongatus* PCC 7942, reported by (Yu et al., 2015) as  $T_d = 4.9 \pm 0.7 \text{ h}$  ( $\mu \approx 0.14 \text{ h}^{-1}$ ).

### 6.4.3 Sensitivity analysis

Since the model primarily provides an upper bound for the maximal specific growth rate, we performed a sensitivity analysis of growth rate as a function of estimated parameters, in particular with respect to the catalytic efficiencies  $k_{cat}$ .

While the sensitivity with respect to the catalytic efficiencies of individual enzymes is low (figures 6.4-6.6), a major determinant of maximal growth rate is the ratio of non-catalytic proteins (figures 6.7-6.8). Based on recent proteomics data for slow growing cells of *Synechococcus elongatus* PCC 7942 (Guerreiro et al., 2014), the relative quota of non-catalytic proteins was determined to be 55% of total protein mass. No experimental estimates exist for fast growing cells. If the true ratio for fast growing cells is assumed to be  $\sim 20\%$ , the resulting growth rate is  $\mu_{max} \approx 0.20 \text{ h}^{-1}$ , corresponding to a doubling time of  $T_D = 3.5 \text{ h}$ , and slightly exceeding the reported maximal growth rate of *Synechococcus elongatus* PCC 7942.

## Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

---

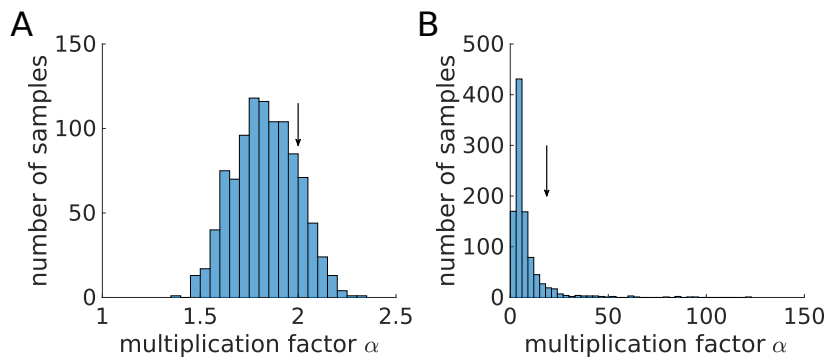


Figure 6.4: **Sensitivity of the multiplication factor to  $k_{cat}$  values for constant light (A: low light, B: high light).** To test for the importance of individual  $k_{cat}$ , we randomized the assignment between the  $k_{cat}$  and their respective enzymes. That is, enzymes are assigned a  $k_{cat}$  drawn from the original distribution (random sampling with replacement). The randomization is motivated by the assumption that there is no systematic bias in BRENDA as far as the overall distribution of  $k_{cat}$  is concerned, but individual assignments may be erroneous. The figure shows the distribution of the multiplication factor  $\alpha$  after randomization. A: For a light intensity of  $150 \mu\text{mol photons s}^{-1} \text{m}^{-2}$  we observe a low sensitivity. The median  $\alpha_{median} = 1.83$  of the distribution is close to the original value of  $\alpha \approx 1.99$  (before randomization), which is indicated by the arrow. B: For a light intensity of  $6000 \mu\text{mol photons s}^{-1} \text{m}^{-2}$  we observe a highly skewed distribution. The median of the distribution  $\alpha_{median} = 5.0$  is much lower than the original reference value of  $\alpha \approx 18.6$  ( $\mu \approx 0.12$ ), which is indicated by the arrow. The fraction of randomized  $\alpha$  larger than the reference value is 7.2%. This does not allow claiming a significant difference between original and randomized growth rates. Nonetheless, models with randomly assigned  $k_{cat}$  seem to have consistently lower growth rates than the original model. This fact is more pronounced at high light intensities. We therefore hypothesize that the assignments of  $k_{cat}$  are not random, but evolutionarily selected to allow for higher growth rates. The hypothesis requires further investigation. With respect to overall sensitivity, we conclude that, at low light intensities, randomization of  $k_{cat}$  has no major effect on growth rate.



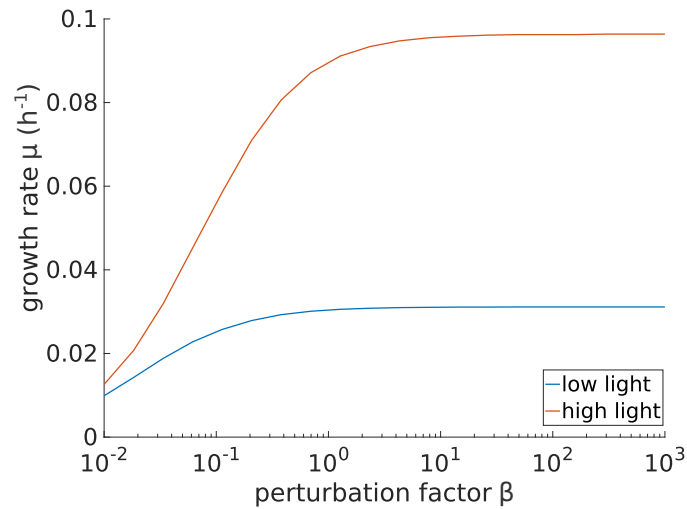


Figure 6.5: **Dependency of the growth rate  $\mu$  on RuBisCO turnover rate in constant light.** The turnover rate of RuBisCO  $k_{cat}^{mod} = \beta \cdot k_{cat}$  in the model was multiplied with a perturbation factor  $\beta$  spanning several orders of magnitude. The growth rate does not increase significantly even for  $\beta \approx 100$ . The simulation was run for a constant light intensity of  $150 \mu\text{mol photons s}^{-1} \text{m}^{-2}$  (low light) and of  $1000 \mu\text{mol photons s}^{-1} \text{m}^{-2}$  (high light).

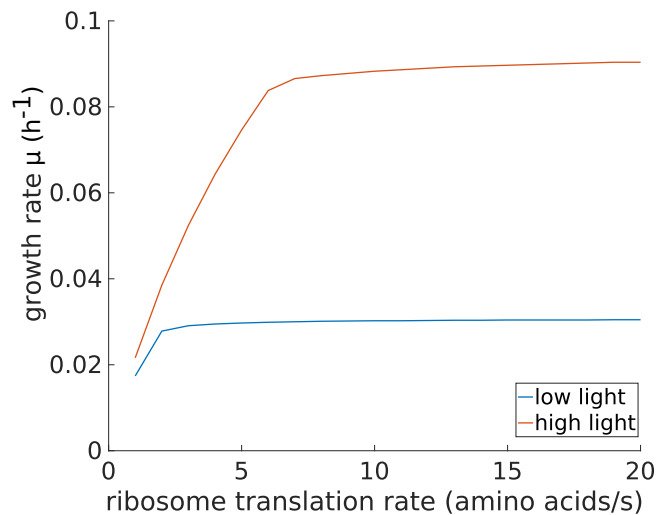


Figure 6.6: **Dependency of the growth rate  $\mu$  on the ribosome translation rate in constant light.** The reference value for the ribosome translation rate used in the simulations is 15 amino acids per second. The simulation was run for a constant light intensity of  $150 \mu\text{mol photons s}^{-1} \text{m}^{-2}$  (low light) and of  $1000 \mu\text{mol photons s}^{-1} \text{m}^{-2}$  (high light).

## Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

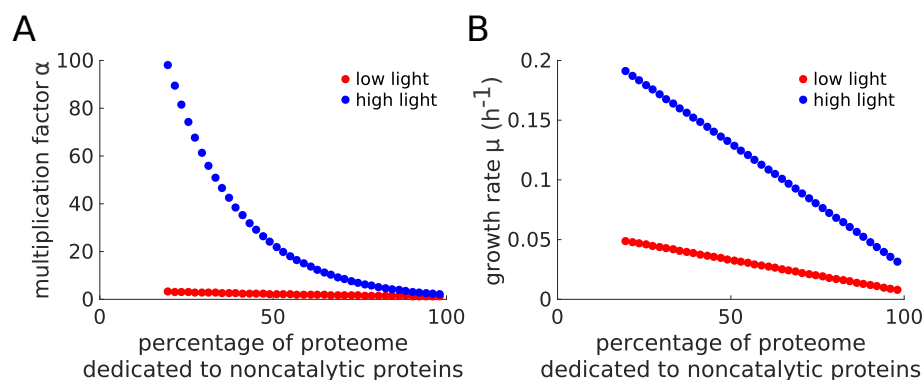


Figure 6.7: **Sensitivity of the growth rate on the ratio of non-catalytic proteins in the proteome.** Shown is the multiplication factor  $\alpha$  (A) and the growth rate  $\mu$  (B) as a function of the percentage of non-catalytic proteins in the proteome. As expected, the growth rate increases for a lower amount of quota proteins. We conjecture that the amount is variable and considerably lower for fast growing cells. We note that the protein complexes of the ETC, phycobilisomes, and proteins of central metabolism (including RuBisCO) are assumed to constitute the bulk of the proteome and are all included as catalytic proteins within our model.

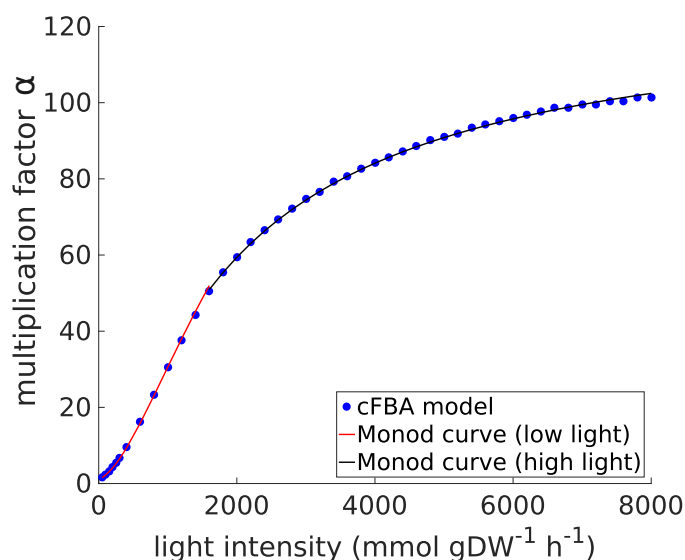


Figure 6.8: **Balanced growth under constant light when the percentage of non-catalytic proteins is set to 20% of the proteome.** Shown is the multiplication factor  $\alpha$  as a function of light intensity. The maximal growth rate obtained in this case is  $\mu = 0.2019 h^{-1}$ .

#### 6.4.4 A day in the life of *Synechococcus elongatus* PCC 7942

Going beyond constant light conditions, our main interest is a solution of the resource allocation problem for diurnal light. The light intensity is modeled as a sinusoidal half-wave with a peak value of  $600 \mu\text{mol photons s}^{-1} \text{ m}^{-2}$ , except otherwise noted. We seek to identify a time-dependent flux pattern (or patterns) that maximizes the overall growth of the cell during a diurnal cycle. Growth rates and overall cellular composition (figure 6.9) depend on the peak light intensity. The results (figures 6.10-6.11) are qualitatively similar to the case of constant light already depicted in figure 6.3.

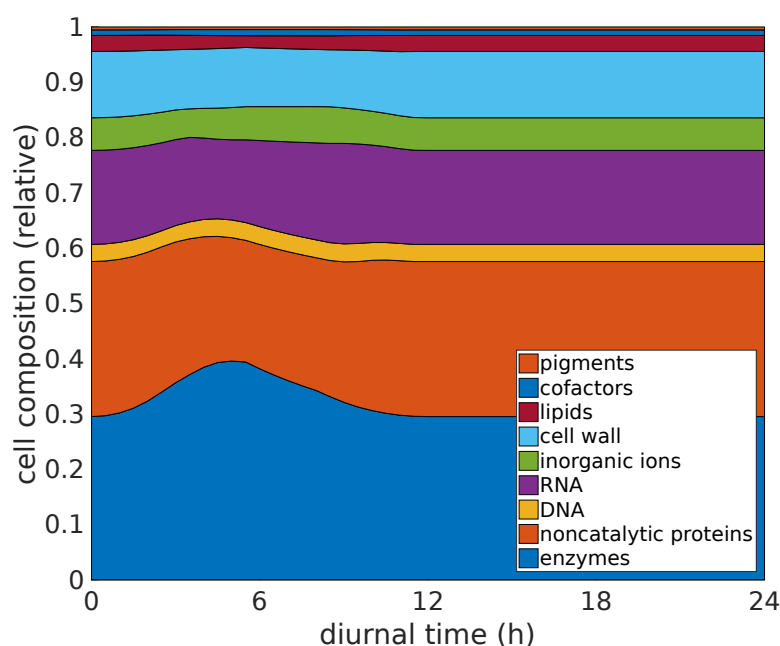


Figure 6.9: **Biomass composition over a full diurnal cycle.** Shown is the relative biomass composition over a full diurnal cycle as a result of the resource allocation problem. The simulations were run for a peak light intensity of  $600 \mu\text{mol photons s}^{-1} \text{ m}^{-2}$ .

Figure 6.12 shows the resulting metabolic flux pattern for a reference day as a function of diurnal time, the relative flux values normalized to the RuBisCO carbon fixation flux, as well as selected examples of fluxes corresponding to different functional categories.

**Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth**

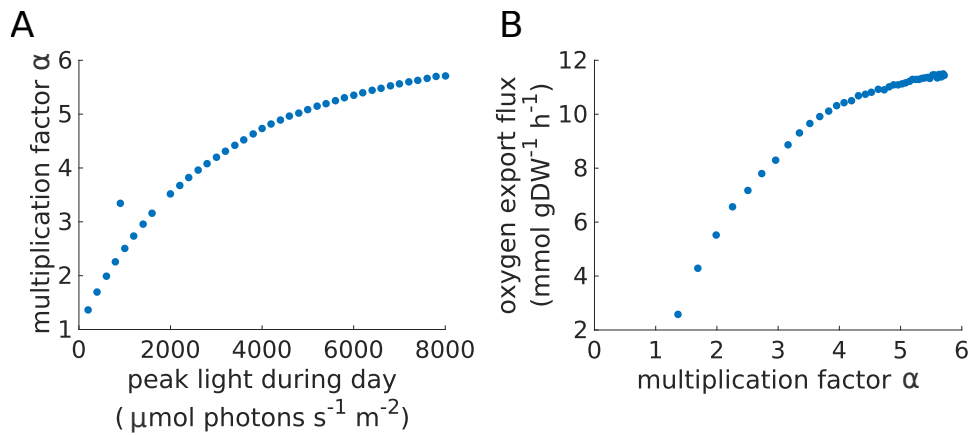


Figure 6.10: **Growth in diurnal light conditions.** A: Dependency of the multiplication factor  $\alpha$  on the peak light intensity. Growth increases with increasing light. B: Oxygen export flux at noon as a function of the multiplication factor  $\alpha$ . The plots are in good qualitative agreement with the results shown for constant light in figure 6.3.

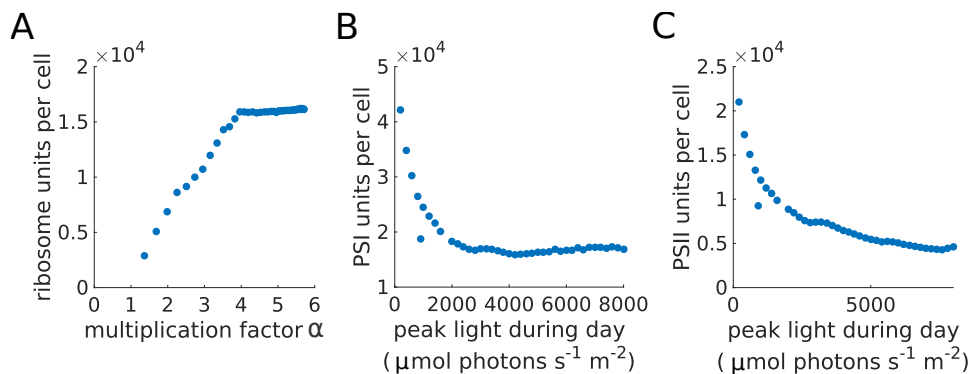
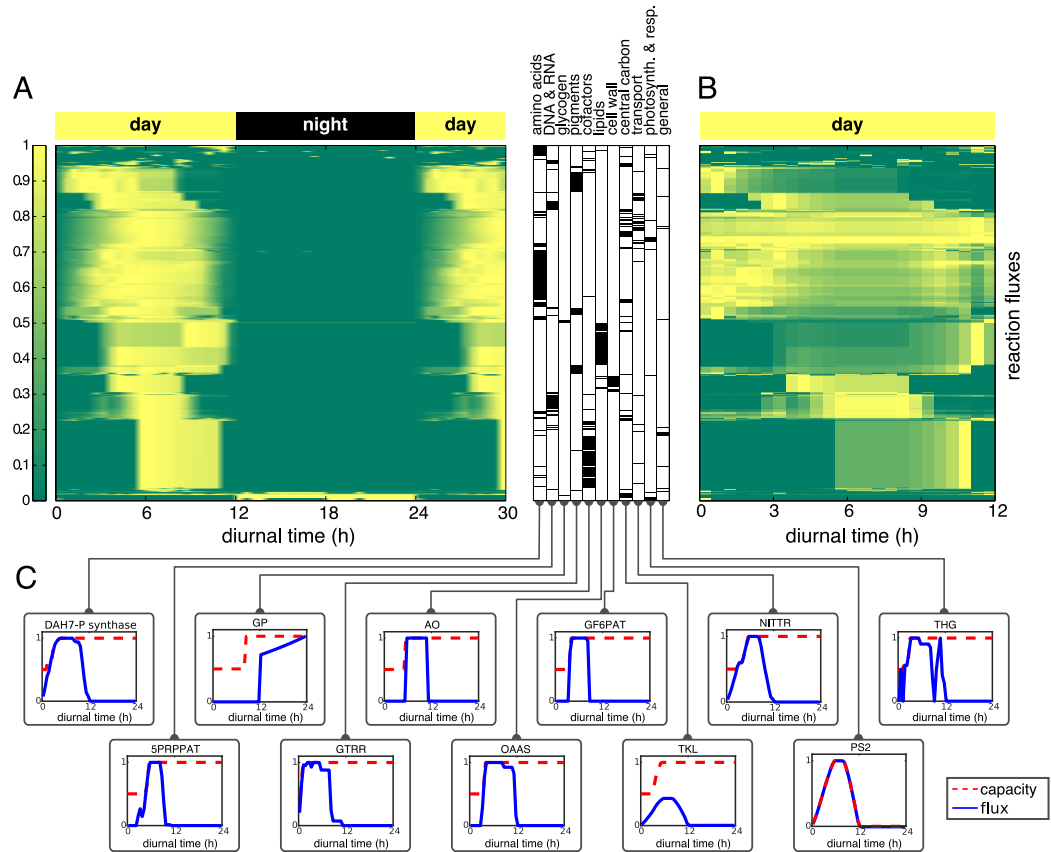


Figure 6.11: **Growth in diurnal light conditions.** A: The ribosome content per cell as function of the multiplication factor  $\alpha$ . B: Dependency of the number of PSII per cell on peak light intensity. C: Dependency of the number of PSI per cell on peak light intensity. The numbers were inferred from the number per dry weight, assuming a cell mass of  $1.5 \text{ pg}$ .



**Figure 6.12: A solution of the resource allocation problem over a full diurnal cycle.** A: Normalized metabolic fluxes as a function of diurnal time. To indicate the periodicity of the solution, the beginning of the subsequent light period is also shown (hours 24-30). We observe a highly coordinated metabolic activity over a diurnal period. Metabolism is organized into distinct temporal phases, ranging from synthesis of amino acids and pigments, to synthesis of lipids, DNA/RNA and peptidoglycan, to synthesis of co-factors. The functional category of each flux is indicated by a black bar in the table adjacent to the plot. B: Excerpt of the light period. The flux values are normalized to RuBisCO flux and scaled to the unit interval. The normalization emphasizes relative carbon partitioning rather than the dependence of metabolic fluxes on total light input. C: Selected metabolic fluxes and their corresponding capacity bounds. Dashed red lines indicate the enzymatic capacities (proportional to enzyme amount). Photosynthesis and reactions of central metabolism closely follow light availability (e.g. transketolase TKL or PSII). Metabolic activity during the early light period is dominated by amino acid synthesis (e.g. DAH7-P for synthesis of aromatic amino acids, nitrate uptake NITTR, and GTRR towards chlorophyll). Later, metabolic activity shifts to DNA and RNA synthesis (e.g. 5PRPPAT) and lipid synthesis (e.g. 3-OA-ACP-syn.), followed by synthesis of peptidoglycan (e.g. GF6PAT) and co-factors (e.g. L-Asp-O. towards nicotinamide adenine dinucleotide). During night, glycogen is utilized via the glycogen phosphorylase (GP). Abbreviations: DAH7-P synthase, 3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase; GP, glycogen phosphorylase; AO, aspartate oxidase; GF6PAT, glucosamine-fructose-6-phosphate amino-transferase; NITTR, nitrate transporter; THG, transhydrogenase; 5PRPPAT, 5'-phosphoribosylpyrophosphate amidotransferase; GTRR, glutamyl-tRNA reductase; OAAS, 3-oxoacyl-ACP synthase; TKL, transketolase; PSII, photosystem II.

## Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

The solution of the resource allocation problem exhibits a highly coordinated metabolic activity over a diurnal period: oxygenic photosynthesis is active during the light period. Inorganic carbon is imported and assimilated via the Calvin-Benson cycle (anabolism), respiratory components (catabolism) are active during the dark period. We highlight two results of the model.

First, we observe that growth is dynamic. The instantaneous growth rate, defined in equation (4.1), follows a specific pattern over the light period. The growth rate is low in the early morning, increases over the course of the light period and decreases again towards dusk (figure 6.13). Such a dynamic growth rate over a diurnal cycle was recently reported for the cyanobacterium *Synechocystis* sp. PCC 6803 (Angermayr et al., 2016), and is also observed when the sinusoidal light input is replaced by the square-wave light-dark cycle typically employed in experiments.

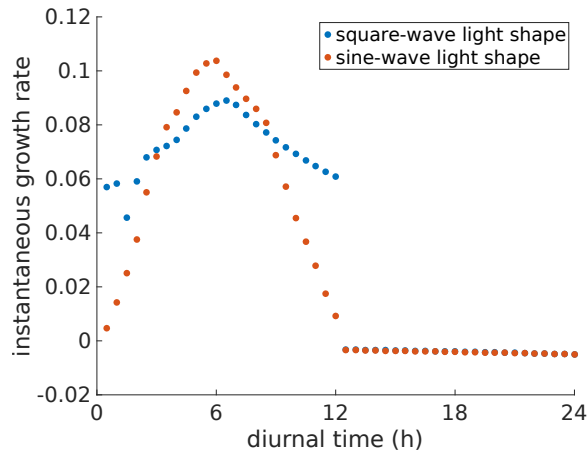


Figure 6.13: **Instantaneous growth rate ( $h^{-1}$ ) for reference day using sine-wave and square-wave light input.** The instantaneous growth rate is defined as  $\frac{1}{\Delta t} \cdot \log\left(\frac{biomass(t_i)}{biomass(t_{i-1})}\right)$ , where  $biomass(t_i)$  is the biomass amount at time point  $t_i$  and  $\Delta t = t_i - t_{i-1}$  (here  $\Delta t = 0.5h$ ). The instantaneous growth rate has recently been measured for *Synechocystis* sp. PCC 6803 under diurnal conditions (see figure 1A in Angermayr et al. (2016)). Therein, after a lag phase, the growth rate increases until shortly after noon, and decreases again towards the evening. Measurements were done under a square-wave light input. The initial lag phase might be due to the sudden light increase at dawn, such effects are not captured by our model. We observe a similar curve under both, sinusoidal, as well as square-wave light input.

Secondly, we observe that growth is exclusive to the light period. During the night phase, glycogen is used for cellular maintenance, serving as a substrate for respiration via the oxidative pentose phosphate pathway and ultimately cytochrome C oxidase. While this observation is unsurprising given the known data on *S. elongatus* 7942 and other cyanobacteria, we emphasize that cessa-

tion of metabolic activity at dusk is not self-evident but already the result of a cellular tradeoff. Limited metabolic activity at night implies not using already existing capacity in terms of enzymes and ribosomes during dark. The benefit of avoiding or reducing the timespan of this idle cellular capacity, however, would require additional storage compounds that must be synthesized prior to the dark period. This, in turn, would entail additional enzymatic costs in terms of storage synthesis capacity. To demonstrate this tradeoff, we conducted an *in silico* experiment using a reduced cost for glycogen synthesis and utilization by neglecting the synthesis costs of the respective enzymes. In this case, cellular synthesis also prevails during the night phase, driven by an increased amount of stored glycogen at dusk (figure 6.14). The experiment, however, also reveals that the amount of necessary additional storage capacity is significant. The high storage requirements explain why such behavior is not observed for the original parameters.

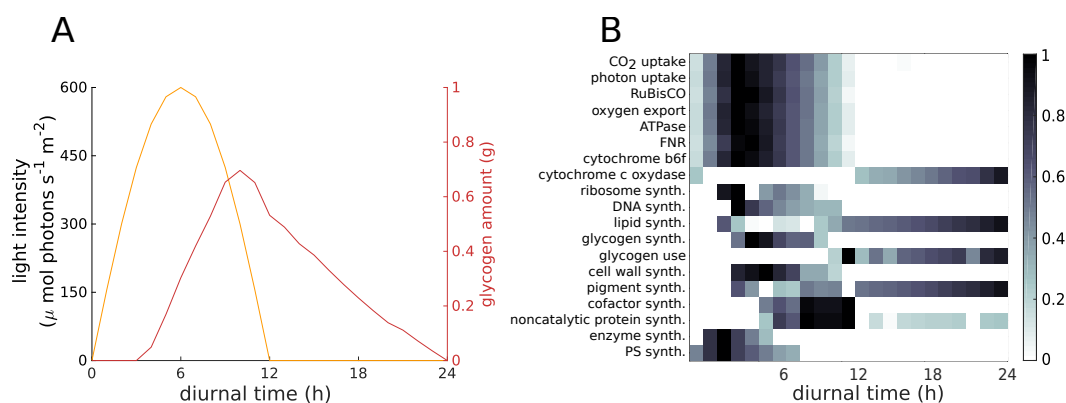


Figure 6.14: **A hypothetical scenario in which the synthesis and breakdown of glycogen requires no enzymatic costs.** Glycogen synthesis and use are implemented using spontaneous reactions. As expected, under these conditions it is energetically more favorable to use synthesis reactions also during the night phase, thereby lowering capacity requirements during the light phase, at the expense of increased glycogen storage. A: Timing and dynamics of glycogen accumulation over the day. The plot shows that synthesis reactions during the night period require a significant amount of glycogen. B: Metabolic activity of key reactions. Time courses are color-coded and normalized to the unit interval.

### 6.4.5 Metabolite partitioning during diurnal growth

The results obtained from the resource allocation problem can be compared to known experimental observations about metabolite partitioning during diurnal growth. While no <sup>13</sup>C flux measurements over a full diurnal cycle have been conducted yet, several studies have investigated the transcriptome, proteome and physiology of *S. elongatus* PCC 7942 and other cyanobacteria over a full

## Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

---

24 h diurnal cycle (Lehmann et al., 2013; Guerreiro et al., 2014; Diamond et al., 2015; Saha et al., 2016). Published transcriptome studies of cyanobacteria typically show global rhythms in gene expression, including a significant reduction of the transcriptional output during the dark period (Ito et al., 2009; Lehmann et al., 2013; Beck et al., 2014). While environmental light-dark cycles are sufficient to drive transcriptional and metabolic rhythms, metabolic rhythms are also influenced by the cyanobacterial circadian clock and persist in cultures under constant light (Markson et al., 2013; Pattanayak et al., 2014; Diamond et al., 2015). Several transcriptomic studies described a broad temporal order of diurnal growth, typically distinguishing between genes peaking at dawn versus genes peaking at dusk (Ito et al., 2009; Diamond et al., 2015). The former set includes genes associated with the Calvin-Benson cycle, as well as genes associated with amino acid synthesis. The latter set includes genes associated with glycogen mobilization and the oxidative pentose phosphate pathway (Diamond et al., 2015). The observed temporal order is consistent across different cyanobacterial strains. For the cyanobacterium *Synechocystis* sp. PCC 6803, (Saha et al., 2016) report an upregulation of PSI and PSII transcripts, genes for amino acid metabolism, and genes from the Calvin-Benson cycle at the beginning of the light period, whereas essential genes for glycogen catabolism showed upregulation in the dark period. (Behrenfeld et al., 2008) report metabolism is dominated by amino acid synthesis between sunrise and noon in a synchronized culture of *Prochlorococcus* PCC 9511.

The results of the resource allocation problem, shown in figure 6.12, closely replicate this temporal order. Metabolism at dawn is dominated by amino acid synthesis and synthesis of pigments. Photosynthetic activity and reactions of central metabolism closely follow light availability. Later in the day, DNA and RNA are synthesized, then lipid, peptidoglycan, and co-factors follow. We note that synthesis of enzymes can precede reaction flux, for example for the glycogen phosphorylase (plot GP in figure 6.12C). Overall, the differences in activity between light and dark metabolism, the initiation of amino acid metabolism at dawn, and storage metabolism are in qualitative agreement with reported metabolite partitioning during diurnal growth.

However, several caveats limit a direct comparison of individual reaction rates to currently available data. Current transcriptomics data are predominantly measured under constant light conditions using synchronized cultures. Such measurements aim at discerning the effects of the circadian clock versus light-driven regulation and conceal actual resource allocation. Furthermore, gene expression is not necessarily indicative of metabolic flux. Measurements also typically involve slow growing cultures, with reduced requirements for *de novo* protein synthesis. Correspondingly, proteomics measurements typically exhibit reduced diurnal variability (Waldbauer et al., 2012; Guerreiro et al., 2014) compared to transcriptomics data. The qualitative agreement of results obtained from the computational model with transcriptomics data may therefore provide



further incentives for experimental studies to focus on *de novo* protein synthesis in fast growing cultures – and thereby to enable a quantitative comparison of the timing of protein synthesis.

#### 6.4.6 The dynamics of glycogen accumulation

Glycogen is the main storage compound in cyanobacteria. Cells accumulate glycogen during the light phase and mobilize it as a source of carbon and energy during the night. It was recently shown that the timing of glycogen accumulation is under tight control of the cyanobacterial circadian clock and disruption of the clock results in altered glycogen dynamics (Diamond et al., 2015). We therefore investigate the dynamics of glycogen accumulation in the context of the global resource allocation problem. We note that our simulation does not impose any ad hoc constraints on the kinetics and timing of glycogen synthesis. Rather, accumulation of glycogen is a consequence of optimal resource allocation. Figure 6.15 shows the time course of glycogen accumulation obtained from the global resource allocation problem over a diurnal period, as well as the predicted carbon partitioning ratio during the light period.

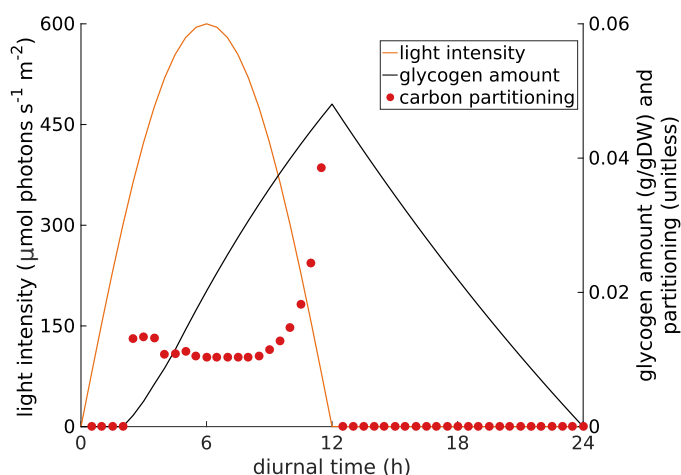


Figure 6.15: Timing and dynamics of glycogen accumulation over a full diurnal cycle. After a brief lag phase, cells accumulate glycogen during the light period and mobilize it as a source of carbon and energy during the night. Shown are absolute amounts of stored glycogen per gram dry weight, as well as the carbon partitioning ratio. We define the carbon partitioning ratio as the ratio between the glycogen synthesis flux and the carbon assimilation flux (RuBisCO), weighted by the corresponding stoichiometries.

The constant rate of glycogen accumulation through the light period is in excellent agreement with recent data on *S. elongatus* PCC 7942 (Diamond et al., 2015) and *Synechocystis* sp. PCC 6803 (Saha et al., 2016). To verify the robustness of our approach, and since most experimental studies use a square-wave light in-

## Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

tensity, rather than a sinusoidal function, we also investigate glycogen accumulation using a square-wave light function, resulting in a similar functional form (figure 6.16).

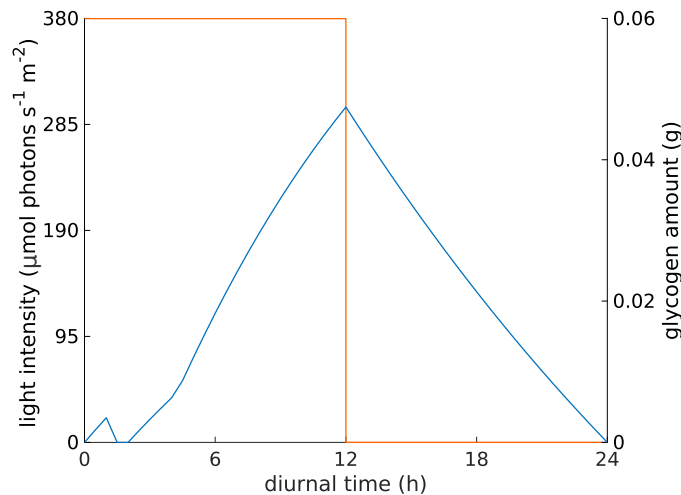


Figure 6.16: **Glycogen under a square light wave.** The overall shape of the curve remains robust. We note that results for a square-wave input have to be interpreted with caution. Our approach is based on the assumption that *S. elongatus* evolved an appropriate regulatory scheme that coordinates its metabolism in their natural diurnal environments. For experiments in an environment different from the natural environment, the evolved regulatory scheme is still in place. We therefore expect to observe a superposition of the effects of the new environment and the original (evolved) regulatory scheme.

We emphasize that the linear slope is not self-evident, but emerges as a trade-off between at least two conflicting objectives: (i) minimal withdrawal of carbon during the early growth period, when the resources could be invested in creating enzymatic and translational capacity that can be used throughout the day versus (ii) investing a minimal amount of carbon resources into the enzymes needed for glycogen synthesis. While the former objective favors withdrawal of carbon for glycogen synthesis later in the day, the later justifies a constant carbon withdrawal for storage synthesis throughout the light period. To the best of our knowledge, the initial lag phase has not yet been noted experimentally, but might provide a stimulus for further experimental evaluation.

### 6.4.7 Glycogen accumulation for variable day lengths

To highlight glycogen accumulation as a systemic property, we also investigate the minimal amount of accumulated glycogen for different photoperiods. Figure 6.17A shows the resulting time courses for different lengths of day versus night periods. Figure 6.17B shows the minimal amount of glycogen required at

the end of the light period.

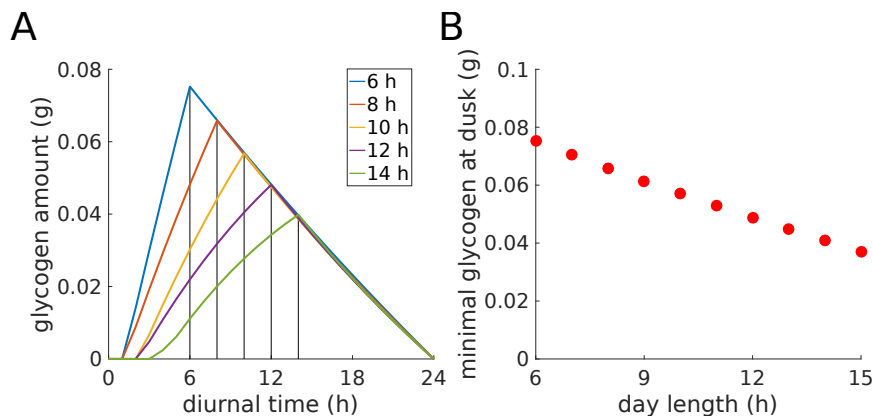


Figure 6.17: **Glycogen for different day lengths.** A: Time courses of glycogen accumulation for different lengths of the light period (day length). B: Minimal glycogen requirements for different day lengths. Peak glycogen content is always observed at dusk. While the minimal amount of glycogen required at dusk exhibits a clear bound, cells can accumulate more glycogen with no discernible effects on overall growth yield. Higher glycogen at dusk implies increased metabolic activity shortly before dawn at the expense of slightly reduced synthesis reactions during the light period.

The overall form of glycogen accumulation remains largely identical. We note, however, that the amount of glycogen required at dusk exhibits a certain plasticity. Firstly, if the night period is doubled, slightly less than twice the glycogen is required to sustain night metabolism. Secondly, while the minimal amount of glycogen required at dusk exhibits a clear bound, cells can accumulate more glycogen with no discernible effects on overall growth. In this case, certain synthesis tasks, in particular lipid synthesis, can be postponed to the end of the night period, thereby requiring less enzyme capacity during the day at the expense of an increased glycogen storage at dusk.

## 6.5 Discussion

Phototrophic growth under diurnal conditions requires a precise coordination of metabolic processes – and the resulting constraints and tradeoffs are challenging to describe using constraint-based analysis and conventional FBA (Henson, 2015). We have developed a genome-scale resource allocation model that allowed us to investigate the stoichiometric and energetic constraints of diurnal phototrophic growth in the context of a global resource allocation problem. We focused on the net stoichiometric and energetic implications of diurnal growth on the *de novo* synthesis of proteins and other cellular macromolecules.

Our aim was a prediction of optimal diurnal resource allocation: how are

## Chapter 6. Cellular tradeoffs and optimal resource allocation during cyanobacterial diurnal growth

---

metabolism and the synthesis reactions of cellular macromolecules organized over a full diurnal cycle? What is the optimal timing of glycogen accumulation during the light phase? From the perspective of cellular resource allocation, these questions can be asked without detailed knowledge about regulatory mechanisms and their corresponding kinetic parameters.

The results obtained from the computational model allowed us to pinpoint several energetic tradeoffs and constraints related to diurnal growth. Overall, the model-derived time courses are in good qualitative agreement with previous experimental observations about flux partitioning in *S. elongatus* PCC 7942 and other cyanobacteria. Growth is dynamic and takes place during the light phase. Carbon fixation and the reactions of central metabolism largely follow light availability. The synthesis of amino acids and pigments dominate metabolism during the early light period, whereas other synthesis reactions peak at later time points. In the absence of light, almost all metabolic activity ceases, and cellular metabolism is dominated by respiratory activity. While well-known experimentally, we emphasize that in the computational model the cessation of metabolic activity during darkness is a consequence of a tradeoff between the cost of unused enzymatic capacity during darkness versus the cost of additional storage that would be required for synthesis reactions to take place in the absence of light. In this respect, the function of the storage compound glycogen is analogous to a cellular battery or capacitor: we expect that if glycogen synthesis and utilization would not entail additional enzymatic (and other) costs, synthesis reactions would continue during the night. Within our computational framework we confirmed this hypothesis using an *in silico* experiment with modified enzyme synthesis costs.

For our reference parameters, the predicted timing characteristics of glycogen accumulation matched recent experimental observations (Diamond et al., 2015; Saha et al., 2016). The emergent dynamics of glycogen accumulation point to the role of the circadian oscillator to modulate metabolite partitioning during growth. It has been shown recently that the circadian oscillator controls the timing of glycogen accumulation such that it occurs at a constant rate through the light period. A disrupted clock results in increased glycogen accumulation early in the day (Diamond et al., 2015). The disrupted pattern of glycogen accumulation is indeed significantly different from the optimal profile predicted here. A manifest hypothesis is therefore that cyanobacterial growth is organized according to a temporal program that evolved to maximize growth in a periodic environment, and that the circadian clock is a regulatory circuit that modulates the transcriptional program of the cell to approach this metabolic optimum. Misalignments between metabolism, clock and environmental cycles will therefore result in impaired growth, as observed by (Lambert et al., 2016).

Recent analyses (Goelzer et al., 2015; Price et al., 2016; Yang et al., 2016) reveal several interesting differences with respect to our results. The model-derived growth rates for *E. coli* corresponding to an optimally allocated proteome were

consistently higher than the experimentally measured rates. In the model, the latter could be supported with 95% less proteome under certain conditions (Yang et al., 2016). This was attributed to cellular “bet hedging” in (generalist) wild-type *E. coli* against unknown environmental challenges. Furthermore, significant protein production without detectable growth benefit was observed experimentally (Yang et al., 2016; Price et al., 2016).

In contrast, the model-derived maximal growth rates we report here are within the range of the (maximal) growth rates observed for *S. elongatus* PCC 7942, even though actual growth rates observed in a laboratory are typically significantly slower (Kuan et al., 2015).

We note that the resource allocation problem only provides upper bounds for the growth rate, assuming an optimally-allocated metabolism, and does not incorporate several detrimental factors, such as light damage and possible photoinhibition. Thus, the close correspondence between observed and model-derived values suggests that cyanobacterial metabolism operates close to optimality – at least in experiments designed for rapid growth.

The model-derived maximal growth rates should be interpreted as an order-of-magnitude approximation, not as a precise estimate. For example, a major unknown factor is the relative amount of non-catalytic proteins, estimated to be up to 55% of total protein for slow growing cells (Guerreiro et al., 2014). We hypothesize that for fast growing cells this percentage is considerably lower. The impact of non-catalytic proteins as (condition-specific) niche-adaptive proteins on the maximal growth rate was already discussed in (Burnap, 2015). The question to what extent slow growing cyanobacteria perform cellular bet hedging similar to *E. coli* and how the allocation of proteome to non-growth-related processes correlates with resistance to adverse environmental and sudden stress conditions remains a question for further research – with implications for biotechnological strain design.



## Chapter 7

# Optimal resource allocation in yeast under dynamically changing environments

The work presented in this chapter has been done within the European project ROBUSTYEAST, a collaboration of

- Freie Universität Berlin (Alexander Bockmayr)
- Vrije Universiteit Amsterdam (Frank Bruggeman)
- École polytechnique fédérale de Lausanne (Vassily Hatzimanikatis)
- Katholieke Universiteit Leuven (Steffen Waldherr)

This project aims at revealing the principles for robustness of metabolism to extracellular nutrient dynamics in yeast using systems biology approaches. Experimental data used in this chapter was provided by Johan van Heerden.

Microbial strains used in biotechnological production pipelines at industry-scale need to fulfill two key requirements: they need to be able to produce the products of interest at high yield, and they need to be robust to the local heterogeneities arising in large industrial bioreactors, especially to transient limitations in nutrient sources and oxygen.

For the former requirement of high yield several systems biology approaches have been developed to obtain very productive strains (Hädicke and Klamt, 2011; Kim et al., 2011; Jungreuthmayer and Zanghellini, 2012; Trinh, 2012; Gruchattka et al., 2013; Erdrich et al., 2014; von Kamp and Klamt, 2014; Klamt and Mahadevan, 2015). However, such strain optimization strategies report strains that are optimized for steady-state conditions, and not for the dynamical conditions present in industry-scale bioreactors. As a result, these strains are

## Chapter 7. Optimal resource allocation in yeast under dynamically changing environments

---

very rarely capable of overcoming the nutrient and oxygen shifts arising from imperfect mixing. Indeed it has recently been shown that individual cells respond very differently to changes in nutrient availability. This gives rise to a subpopulation adapting to the new conditions and resuming growth, while another subpopulation dies out, and thus induces a huge productivity loss (van Heerden et al., 2014).

We, in the ROBUSTYEAST consortium, believe that one element that makes a difference in survival is the choice of the organism between being a generalist and being a specialist. That means a choice between investing resources into proteins that are unnecessary at the moment but may become useful later versus only producing the proteins needed at the moment. Besides that, we want to understand how does the organism adapt to the new environment.

For this purpose, we built a dynamic resource allocation model of the yeast *Saccharomyces cerevisiae*. As a start, we validated the growth rates obtained from our model using experimental data provided by Johan van Heerden, as explained in section 7.3.1. We have obtained a very good agreement between model and data, without any parameter changes.

We continued by studying how wildtype yeast adapts to changes in the environment.

Next, we have designed a series of studies of *in silico* yeast knock-out mutant strains in dynamic environments. The idea behind this is that, mutants who are viable in constant conditions, but do not survive environment shifts are missing a gene that is vital for surviving the transition. Thus, we provide a list of candidate genes critical for survival in dynamic conditions. This list remains to be validated once data from the experimental partners in the ROBUSTYEAST project is available.

### 7.1 Model building

#### 7.1.1 Choice of reconstruction and model reduction

The first step in building a resource allocation model for yeast was choosing a metabolic reconstruction to start with. Since 2003, more than 25 genome-scale metabolic network reconstructions of yeast have been published (Förster et al., 2003b; Duarte et al., 2004; Krieger et al., 2004; Kuepfer et al., 2005; Herrgård et al., 2006; Hjersted and Henson, 2006; Herrgård et al., 2008; Nookaew et al., 2008; Mo et al., 2009; Aho et al., 2010; Dobson et al., 2010; Zomorodi and Maranas, 2010; Costenoble et al., 2011; Matsuda et al., 2011; Swainston et al., 2011; Szappanos et al., 2011; Celton et al., 2012; Heavner et al., 2012; Aung et al., 2013; Heavner et al., 2013; Österlund et al., 2013), some of which were reviewed in (Heavner and Price, 2015). The genome-scale models range in size between iND750 (Duarte et al., 2004) with 1061 metabolites and 1266 reactions and Yeast 7 (Aung et al., 2013) with 2220 metabolites and 3498 reactions.



Based on this wide variety, one would most likely choose the newest consensus model, Yeast 7, as this contains the most recent and comprehensive annotations. However, the size of this model is too large to allow a dynamic resource allocation analysis using deFBA. One can reduce this model, while preserving desired functionalities, as described in chapter 4 using a network reducer. However, after running the minimal network finder tool for more than 24 hours, we still could not obtain a reduction. Therefore, we have decided to use Yeast 6 (Heavner et al., 2013), which we were able to reduce using the minimal network finder from (Röhl and Bockmayr, 2017).

We focus, just as the experimental partners in ROBUSTYEAST, on growth in minimal Verduyn medium, whose composition we detail in table A6 in Appendix D. Additionally, we would like to study yeast both in aerobic and anaerobic conditions. According to (Ishtar Snoek and Yde Steensma, 2007), yeast needs sterols and unsaturated fatty acids in the medium to grow anaerobically. Therefore, additional medium components were added. For a complete list of these components see table A7 in Appendix D.

The network reducer in (Röhl and Bockmayr, 2017) uses a mixed integer linear programming approach to compute the minimal subnetwork of a given network that still preserves some user-given functionalities. For instance, one functionality could be that the reduced network retains 99% of the original network's flux through the biomass reaction in a fixed environment. Another functionality could be that certain metabolites or reactions should be kept in the reduced network.

Therefore, to reduce the Yeast 6 model, we have defined a list a functionalities which should be kept in the reduced model:

- The reduced model should display at least 99% of the Yeast 6 biomass flux in the following environmental conditions:
  - aerobic glucose
  - anaerobic glucose
  - aerobic galactose
  - anaerobic galactose
- The reduced model should retain the fermentation pathway.

We obtained a reduced network with 454 metabolites, 438 reactions, and 429 genes, which can be found at <https://ndownloader.figshare.com/files/8653207>. As already seen in chapter 6, it is possible to study networks of this size using dynamic resource allocation tools.

After the reduction, we wanted to include as much as possible from the new information in Yeast 7 in our reduced network. Moreover, we made sure that

## Chapter 7. Optimal resource allocation in yeast under dynamically changing environments

---

the network can consume possible by-products of overflow metabolism such as ethanol or glycerol, as is the case in reality. Therefore, the following additional changes to the network have been made. Note that the reactions in Yeast 6 are a subset of those in Yeast 7, and the reaction IDs are the same.

- reaction r\_1117 has been constrained to be irreversible (Pereira et al., 2016)
- reactions r\_2111 and r\_2133 (biomass) have been removed
- putrescine cycling through reactions r\_1250 and r\_1251 has been removed
- reaction r\_0364 has been constrained to be irreversible as it has a strongly negative Gibbs free energy at physiological pH and reactant concentrations:  $\Delta_r G'^m = -55.4 \text{ kJ/mol}$  computed using the eQuilibrator resource (Flamholz et al., 2012)
- reactions r\_1130 and r\_1131 have been constrained to be irreversible (Marobbio et al., 2006)
- reactions r\_0735 and r\_0736 have been constrained to be irreversible (Miziorko, 2011)
- reaction r\_0143 has been constrained to be irreversible as it has a strongly negative Gibbs free energy at physiological pH and reactant concentrations:  $\Delta_r G'^m = -48.7 \text{ kJ/mol}$  computed using the eQuilibrator resource (Flamholz et al., 2012)
- reaction r\_1245 has been constrained to be irreversible (Hamel et al., 2004)
- reaction r\_0962 has been constrained to be irreversible (Portela et al., 2006)
- $\text{H}^+$  was removed from reaction r\_0507 for charge balancing (Aung et al., 2013)
- GDH2 (r\_0470) has been set to irreversible in the backwards direction (Aung et al., 2013)
- mitochondrial ADP/ATP transporter (r\_1110) does not cotransport protons (Klingenberg, 1980; Voza et al., 2004)
- ATP synthase (r\_0226) moves 4 cytoplasmic protons (Aung et al., 2013)
- NADH:ubiquinone oxidoreductase (r\_0773) is not proton translocating (Velázquez and Pardo, 2001)
- malic enzyme (r\_0718) is not  $\text{NAD}^+$ -, but  $\text{NADP}^+$ -dependent (Boles et al., 1998; dos Santos et al., 2004)
- changed proton stoichiometry for cytochrome oxidase (r\_0438) (Postmus et al., 2011) and for ubiquinol cytochrome-c reductase (r\_0439) (Trumpower, 1990)
- L-aminoadipate-semialdehyde dehydrogenase (r\_0678) requires ATP (Zabriskie and Jackson, 2000)
- metabolites exchanged through the mitochondrial transporter r\_1099 were changed to 2-oxoadipate and 2-oxoglutarate and the reaction was set as reversible (Palmieri et al., 2001)

- alcohol dehydrogenase (r\_0163, ethanol to acetaldehyde), acetaldehyde transport (r\_1632), aldehyde dehydrogenase (r\_0174, r\_0175), acetyl-CoA synthetase (r\_0113), adenylate kinase (r\_0149), and cytosolic glycerol 3-phosphate dehydrogenase (r\_0491), which had not been retained by the reduction, were added back to ensure the model can consume overflow metabolism by-products
- fructose biphosphatase (r\_0449) and phosphoenolpyruvate carboxykinase (r\_0884), which had not been retained by the reduction, were added back to ensure the model can run gluconeogenesis
- glucan 1,4- $\alpha$ -glucosidase (r\_0463), glycogen phosphorylase (r\_0511), and  $\alpha,\alpha$ -trehalase (r\_0194), which had not been retained by the reduction, were added back to ensure the model can consume the storage products glycogen and trehalose

### 7.1.2 The macromolecules of autocatalytic growth

Starting from the reduced network, a deFBA model of yeast has been constructed, following the guidelines in chapter 4.

#### Ribosomes

The ribosome turnover rate was set to 10 attached amino acids per second according to data in (Karpinets et al., 2006; Milo and Phillips, 2015), while the list of ribosomal proteins was sourced from the Ribosomal Protein Gene Database (Nakao et al., 2004) and is displayed in table A8 in Appendix D.

#### Enzyme subunit stoichiometry

The subunit stoichiometry of all enzymes has been sourced manually from UniProtKB (UniProt Consortium, 2014) and from the *Saccharomyces* Genome Database (Cherry et al., 2012). The SBML file, generated in the format described in section 4.8, is in the process of being uploaded to FairdomHub.

#### Storage

The two main storage molecules of yeast are glycogen and trehalose, which we denote by  $G$  and  $TH$ . Just as in the *Synechococcus* model in the previous chapter, these are allowed to accumulate and may be consumed if available.

#### Noncatalytic biomass

The noncatalytic (quota) biomass has been organized into six quota metabolites: noncatalytic proteins, RNA, DNA, cell wall, lipids and small molecules. The production reactions for these quota metabolites are shown in table A9 in Appendix D.

## 7.2 Model constraints and objective

### 7.2.1 Constraints

#### Steady-state versus time-dependent quantities

As before, we assume internal metabolites are at a quasi-steady-state with respect to the macromolecules and environment dynamics. As in chapter 6, we have

$$\frac{d\mathbf{n}_{\mathcal{M}}(t)}{dt} = S_{\mathcal{M}*}\mathbf{v}(t) = 0, \quad (7.1)$$

and

$$\frac{d\mathbf{n}_i(t)}{dt} = S_{i*}\mathbf{v}(t), \quad \forall i \in \mathcal{E} \cup \mathcal{Q} \cup G \cup TH. \quad (7.2)$$

#### Enzyme amounts constrain reaction rates

Enzyme amounts constrain reaction rates within the metabolic network as explained in chapter 2. Turnover numbers for metabolic enzymes were retrieved from the BRENDA database (Schomburg et al., 2013) and manually sourced from the literature. The turnover numbers for glucose and galactose uptake were set to  $200 \text{ s}^{-1}$  according to data in (Ye et al., 2001).

For irreversible enzyme-catalyzed reactions  $j$ , our constraint then reads

$$\mathbf{v}_j(t) \leq k_{cat}^j \mathbf{n}_{\mathcal{E}_j}(t), \quad (7.3)$$

at all time points  $t$ . In the case of reversible reactions, both directions are constrained

$$\mathbf{v}_j(t) \leq k_{cat}^{j+} \mathbf{n}_{\mathcal{E}_j}(t), \quad \mathbf{v}_j(t) \geq -k_{cat}^{j-} \mathbf{n}_{\mathcal{E}_j}(t), \quad (7.4)$$

where  $k_{cat}^{j+}$  and  $k_{cat}^{j-}$  are the turnover rates for the forward and reverse direction respectively. We impose these two constraints at each time point  $t$  for each reversible reaction  $j$ . The constraints apply only for enzyme-catalyzed reactions. Rates of spontaneous reactions remain free of these bounds.

In case several reactions are catalyzed by the same enzyme, their total flux weighted by the inverse of the respective turnover numbers is bound by the enzyme amount. Such a situation happens, for instance, in the case of the ribosome, as detailed in the previous chapters. Enzyme synthesis reactions are irreversible and hence the bound is only applied for the forward directions of the reactions, while the reverse directions are prohibited using a lower flux bound of zero at all time points.

#### Extracellular metabolites

We distinguish between nonlimiting extracellular metabolites (present in saturation amounts) and part of the Verduyn medium

- O<sub>2</sub>
- H<sup>+</sup>
- H<sub>2</sub>O
- CO<sub>2</sub>
- ammonium
- phosphate
- sulphate
- (R)-pantothenate
- 14-demethylsterol
- ergosta-5,7,22,24(28)-tetraen-3beta-ol
- fecosterol
- myo-inositol
- lanosterol
- zymosterol

and nutrient sources they may deplete

- glucose
- galactose
- ethanol

The amounts of the latter are modeled dynamically following the uptake by the metabolic network as

$$\frac{d\mathbf{n}_{\mathcal{X}}(t)}{dt} = S_{\mathcal{X}*}\mathbf{v}(t), \quad (7.5)$$

where  $\mathcal{X}$  denotes the set of extracellular metabolites available in limited amounts. The uptake of extracellular metabolites present at saturation is only constrained by the existing transporter capacity and its catalytic efficiency.

### Amounts of quota compounds

The initial values for the six quota compounds are set to be equal to their corresponding amounts in 1 gram dry weight of yeast cells and displayed in table A10 in Appendix D. We denote their initial amounts as the vector  $\mathbf{q}_0$ , and require that these proportions are kept as lower bounds for quota amounts throughout the simulation time as

$$\mathbf{n}_{\mathcal{Q}}(t) \geq \mathbf{q}_0 \sum_{i \in \mathcal{E} \cup \mathcal{Q} \cup G \cup TH} MW_i \mathbf{n}_i(t). \quad (7.6)$$

According to the data in (De Godoy et al., 2008), the catalytic proteins which are modeled explicitly in our model sum up to a fraction of 0.40561 of the total proteome of yeast grown in minimal medium. In the biomass reaction of Yeast 6 we find that proteins make up 0.466298 grams in one gram dry weight of cells. Thus, we have set the noncatalytic protein quota requirement to  $(1 - 0.40561) \cdot 0.466298 = 0.27716 \text{ g/gDW}$ .

## Chapter 7. Optimal resource allocation in yeast under dynamically changing environments

---

As before, together with metabolic proteins, glycogen, and trehalose, the initial composition vector adds up to one gram, and thus we obtain the additional constraint

$$\sum_{i \in \mathcal{Q}} \mathbf{n}_i(t_0) + \sum_{i \in \mathcal{E}} \mathbf{n}_i(t_0) \cdot MW_i + \mathbf{n}_G(t_0) + \mathbf{n}_{TH}(t_0) = 1. \quad (7.7)$$

### Maintenance

No ATP maintenance has been enforced since the growth rates obtained in the model are close to those observed in experiments, as we will show in the results section.

### Time discretization

As detailed in chapter 5, we discretize time using the implicit midpoint rule. Derivatives  $\dot{\mathbf{n}}$  and fluxes  $\mathbf{v}$  are evaluated at the middle of the discretization intervals, while macromolecule amounts  $\mathbf{n}$  are evaluated at the ends of the discretization intervals.

### Environment shifts

To analyze how yeast deals with environment shifts, we distinguish between two types of shifts:

- **Shifts in extracellular metabolites present at saturation.** In case we model that an extracellular metabolite that is not a nutrient is no longer present in the environment for a time period, we simply block the respective transporter by imposing its lower and upper flux bounds to be zero.
- **Shifts in nutrient sources.** In case we model shifts in nutrient sources, such as a shift from glucose to galactose medium, we impose at the shift time point  $t_{shift}$  the desired bounds on the respective extracellular nutrient amounts. In addition, for the nutrient removed from the medium, we need to remove the constraint that links its amount at  $t_{shift}$  with its amount at the time point right before  $t_{shift}$ , as this constraint is no longer satisfied.

### 7.2.2 Objective

Since yeast does not follow a light availability periodicity as was the case for cyanobacteria, we do not use cFBA, but deFBA to study dynamic resource allocation in this model. Thus, as detailed in chapters 2 and 5, we use the objective

$$\max_{\mathbf{n}(t), \dot{\mathbf{n}}(t), \mathbf{v}(t)} \int_0^{t_f} \sum_{i \in \mathcal{E}} MW_i \mathbf{n}_i(t) dt. \quad (7.8)$$

To avoid the side effects of including storage and noncatalytic compounds in the objective, analyzed in (Waldherr and Lindhorst, 2017), we only include enzymes in the objective. Note that, since quota amounts are required to have a certain fraction of total biomass at each time point, they are implicitly maximized using this objective.

### 7.2.3 deFBA vs. short term deFBA

Since through a global optimization approach such as deFBA the system already knows at the start of the simulation time whether and when an environment shift will occur, an important method for these studies was short term deFBA, which we briefly explained in chapter 2. This way, by only using a limited prediction horizon, the system only detects a nutrient shift shortly before the shift actually happens, thereby mimicking the behavior of real cells (Kochanowski et al., 2013).

### 7.2.4 Sudden shifts

Sudden environment shifts, for example from environment A to environment B, are implemented using two chained deFBA problems, one for each environment. The initial biomass amounts for the second deFBA problem, i.e., in environment B, are set to be equal to the last computed biomass amounts in environment A, i.e., to  $\mathbf{n}_{\mathcal{E} \cup \mathcal{G} \cup \mathcal{H}}(t_f^A)$  where  $t_f^A$  is the total simulation time in environment A.

Since the condition number of a typical deFBA problem for yeast was lower, indicating a better conditioning of the LP problem, all simulations were performed using the CPLEX Optimization Studio version 12.2.

## 7.3 Results

Prior to studying yeast under dynamic environments, we validate our model against growth rate experimental data obtained in the same (constant) environment.

### 7.3.1 Growth in constant environmental conditions

Figures 7.1-7.3 provide a comparison between the experimental and model-derived growth rates in aerobic glucose, aerobic galactose, and anaerobic glucose minimal Verduyn medium. We observe that the model provides a relatively tight upper bound on the experimental growth rates. This is what we expect, as the model provides an optimal solution that may not be reached in the lab. We have therefore decided to not add any additional ATP maintenance requirement to the model.

## Chapter 7. Optimal resource allocation in yeast under dynamically changing environments

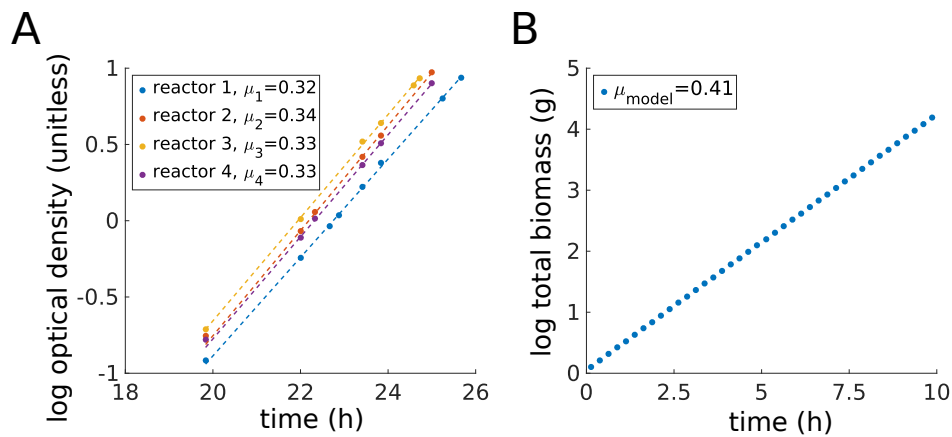


Figure 7.1: **Comparison between the experimental and model-derived growth rates in aerobic glucose Verduyn medium.** A. Log optical density of four batch cultures of yeast grown in aerobic minimal Verduyn medium enriched with sterols and glucose. The derived growth rates were obtained by fitting a line through the data points. B. Log of model-derived total biomass for the same medium computed using deFBA. The model-derived growth rate was obtained by computing the instantaneous growth rate.

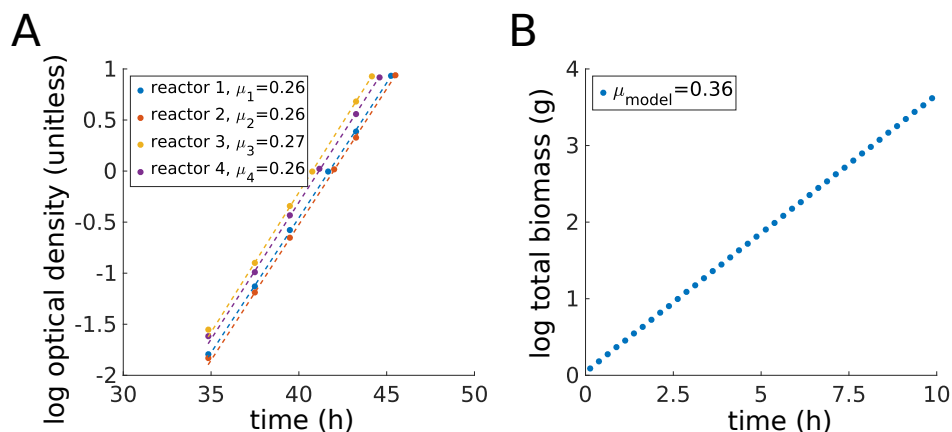


Figure 7.2: **Comparison between the experimental and model-derived growth rates in aerobic galactose Verduyn medium.** A. Log optical density of four batch cultures of yeast grown in aerobic minimal Verduyn medium enriched with sterols and galactose. The derived growth rates were obtained by fitting a line through the data points. B. Log of model-derived total biomass for the same medium computed using deFBA. The model-derived growth rate was obtained by computing the instantaneous growth rate.

### 7.3.2 Overflow metabolism is an optimal behavior from a resource allocation perspective

To display even more the very good qualitative agreement between model and data, we show in figure 7.4 the extracellular glucose and ethanol concentrations



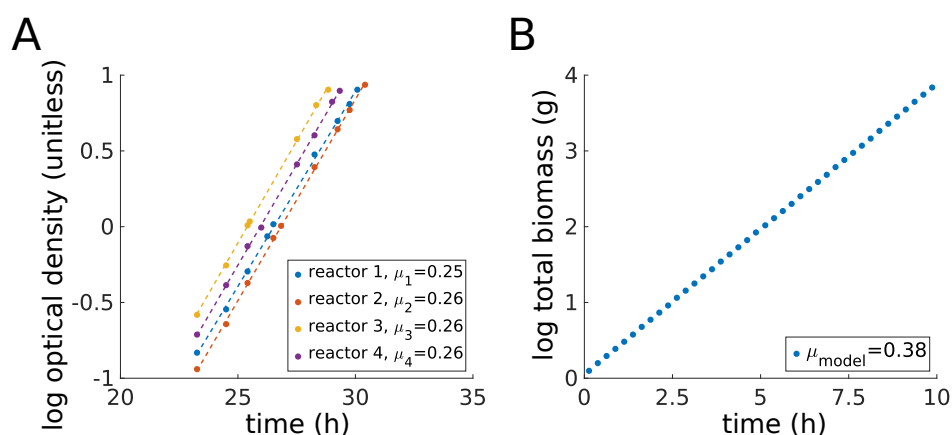


Figure 7.3: **Comparison between the experimental and model-derived growth rates in anaerobic glucose Verduyn medium.** A. Log optical density of four batch cultures of yeast grown in anaerobic minimal Verduyn medium enriched with sterols and glucose. The derived growth rates were obtained by fitting a line through the data points. B. Log of model-derived total biomass for the same medium computed using deFBA. The model-derived growth rate was obtained by computing the instantaneous growth rate.

in aerobic glucose conditions, measured using high performance liquid chromatography, as well as their model-derived extracellular amount counterparts. We note that a quantitative comparison of extracellular metabolite amounts is not possible here due to unit mismatch between data (concentrations) and model-derived quantities (molar amounts).

In addition to the good qualitative agreement, figure 7.4 also shows that the model ferments glucose even in the presence of oxygen, phenomenon also known in the literature as “overflow metabolism”. This indicates that, although often termed “wasteful”, fermentation in aerobic conditions is the best strategy for optimizing growth from a resource allocation perspective, consistent with the conclusions in (Molenaar et al., 2009).

To understand the growth advantage offered by fermentation over respiration, we have simulated the model under the same conditions as before, but with the ethanol export reaction blocked. This way, we have essentially blocked the whole fermentation pathway, as ethanol cannot be accumulated inside the cell due to the quasi-steady-state constraint on internal metabolites. The resulting instantaneous growth rate stabilized at a value of  $0.37 \text{ s}^{-1}$ , which is 10% lower than the growth rate obtained when fermentation was allowed ( $0.41 \text{ s}^{-1}$ ).

However, the ideal strategy in aerobic glucose conditions seems to not be exclusively fermentation. Instead, a small flux through the respiratory pathway can also be observed. This brings a small growth advantage and is also the reason why the model derived growth rate in anaerobic glucose conditions is slightly

## Chapter 7. Optimal resource allocation in yeast under dynamically changing environments

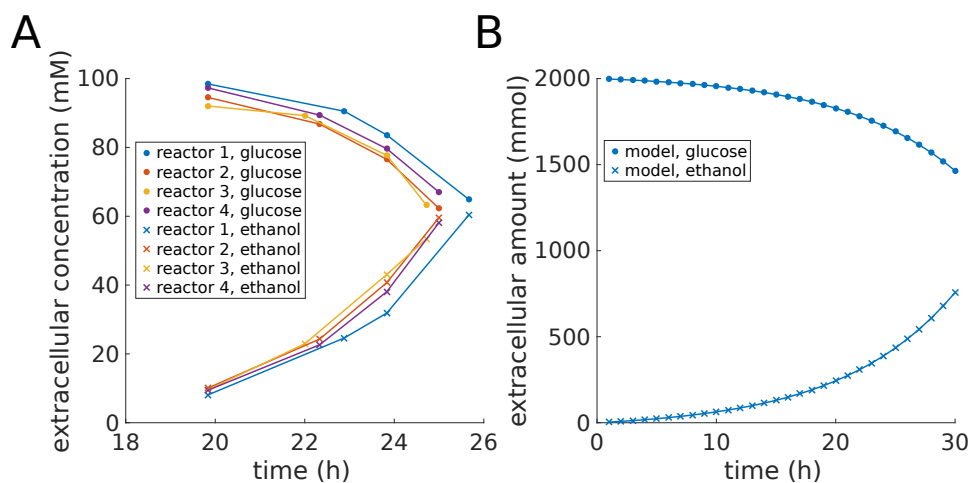


Figure 7.4: **Comparison between the experimental and model-derived extracellular glucose and ethanol amounts in aerobic glucose Verduyn medium.** A. Extracellular glucose and ethanol concentrations measured in the laboratory by high performance liquid chromatography, corresponding to the reactors in figure 7.1. B. Model-derived extracellular glucose and ethanol amounts computed using deFBA.

lower than the growth rate obtained when oxygen is present (figures 7.1 and 7.3).

### 7.3.3 Diauxie is an optimal behavior from a resource allocation perspective

Another well-known phenomenon in yeast is the diauxie. This manifests when yeast grows in a batch culture with a mixture of two sugars as nutrient source (typically glucose-galactose). Instead of metabolizing both sugars simultaneously, some yeasts, including *Saccharomyces cerevisiae*, consume glucose first, and then galactose, resulting in two separate growth phases (New et al., 2014; Siegal, 2015).

The diauxie is often considered to be a consequence of selection to minimize expression of “useless” metabolic pathways when a nutrient that can be more efficiently used is present in the medium (Magasanik, 1961; MacLean, 2007). This means that also from a resource allocation perspective it must be more efficient to metabolize the sugars sequentially rather than at the same time. We have tested this hypothesis using our resource allocation model. Indeed, although the model does not account for any regulatory mechanisms, and although deFBA has global knowledge and can thus “foresee” that glucose will run out, the sugars are still consumed sequentially, as shown in figure 7.5.

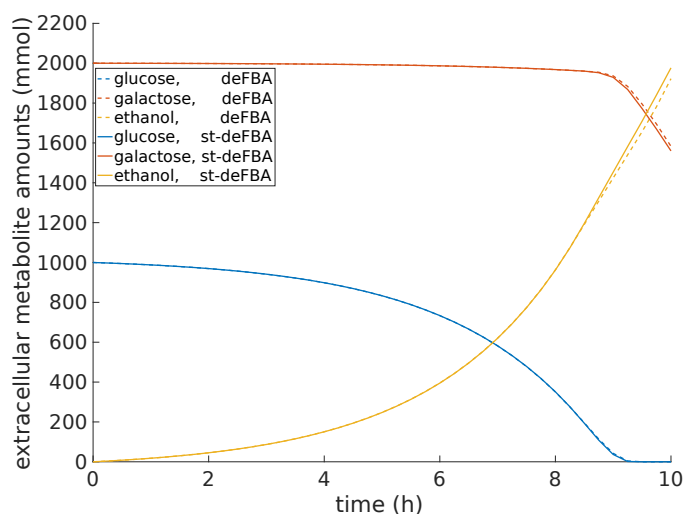


Figure 7.5: **Model-derived extracellular metabolite time courses showing the preferential use of glucose over galactose.** The time courses have been obtained using deFBA (dashed curves) and short term deFBA (st-deFBA, continuous curves) with a prediction horizon of one hour for Verduyn medium enriched with sterols, glucose, and galactose, under aerobic conditions.

### 7.3.4 Adaptations of wildtype yeast to environmental shifts

Following our study of diauxie, with its slow environmental changes, we turned our attention to what happens when yeast is subjected to sudden changes in the environment. We investigated two scenarios: (i) sudden unavailability/availability of oxygen, and (ii) sudden changes in the nutrient source, i.e., shifts from glucose to galactose.

#### Oxygen availability shifts

We investigated slow fluctuations in oxygen availability, happening at intervals of four doubling times ( $T_d \approx 1.5 h$  in aerobic glucose conditions). As opposed to fast fluctuations, where we expect metabolic flexibility to play a crucial role, the slow fluctuations will be governed by yeast's ability to undergo transcriptional adaptation.

As such, we have designed our *in silico* experiment over a total period of 24 hours, during which we alternate oxygen supply as shown in figure 7.6.

Since we are also interested in sudden changes (local, immediate) for the system and it is important that these changes are not a priori known, as is the case in the global resource allocation problem, we use both deFBA and short term deFBA to study the impact of these shifts.

In the case of deFBA, we observe the behavior of an organism that is adapted to

## Chapter 7. Optimal resource allocation in yeast under dynamically changing environments

---

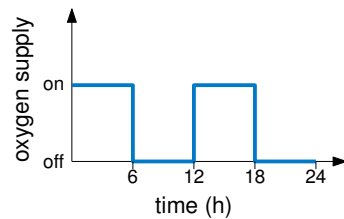


Figure 7.6: **Timing of shifts in oxygen availability.** The shifts are implemented in the model using temporary zero bounds on the oxygen uptake flux during the period when the oxygen supply is off.

a dynamic environment, so it expects the shifts to arise and has prior knowledge of them.

Short term deFBA shows us how an organism that senses these shifts for the first time shortly before they happen (a maximum of one hour prediction horizon) behaves.

A third case is when the organism is a specialist, adapted to growth in constant environment, and is then confronted for the first time with a sudden environmental shift, without any prior knowledge. For implementing the third case a sequence of deFBA problems was solved, one for each environment window (i.e., the windows 0–6, 6–12, 12–18, and 18–24 hours). For each environment window we solve a deFBA problem, imposing as initial condition the biomass composition at the end of the previous window. For the first window, the initial biomass composition is free, but has to obey the constraints on quota and total mass mentioned in section 7.2.1.

Our first observation is that in all cases growth is dynamic (figure 7.7). In sudden shifts, as well as with a short prediction horizon (1 hour resp. 30 minutes), we observe adaptation periods before the maximal growth rate is achieved in each shift window.

The results in figure 7.7 highlight two facts. First, even if the system has global knowledge about when shifts will occur, there is still a small adaptation period before the maximal growth rate in the new environment is attained (deFBA curve). In case of small prediction horizons ahead of the shift, enzymes needed in the new environment are synthesized gradually after the shift at the expense of a lower growth rate. The simulation with sudden environment shifts displays the longest adaptation period in the new environment.

Second, an increase in the prediction horizon (and hence in the preparation time for the new environment) results in a slight increase in growth rate, but is not sufficient for yeast to be fully prepared for the new conditions.

Another important observation is that the system maintains a linear relationship between the growth rate and the amount of ribosome units throughout the

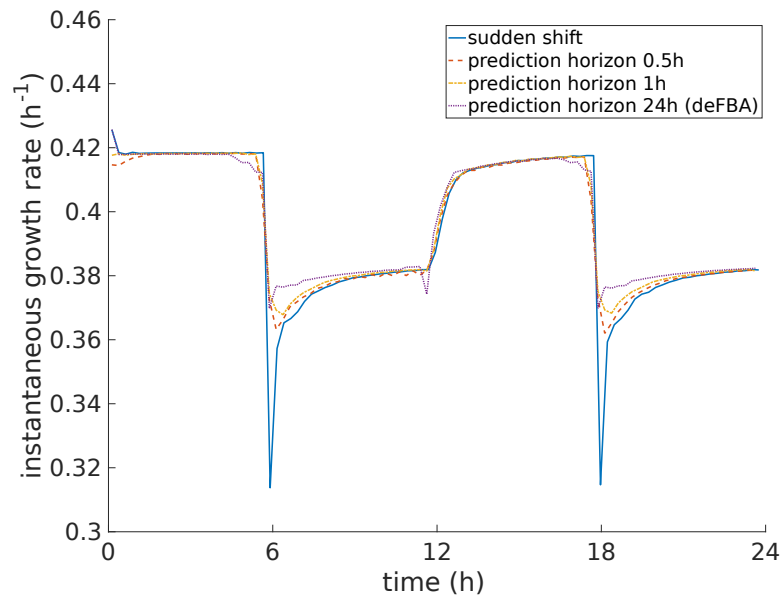


Figure 7.7: **Dynamics of the growth rate following the shifts in oxygen availability from figure 7.6.** Shown are the growth rate profiles obtained from a sudden shift (two sequential deFBA problems with a sudden environment change, continuous curve), using short term deFBA with prediction horizon of half an hour (dashed curve) and an hour (dash-dotted curve), as well as the growth rate when the system has global knowledge of the shifts (one deFBA problem, dotted curve) in Verduyn medium enriched with glucose and sterols.

shifts (figure 7.8). With increasing growth rates, we see an increase in the fraction of the total proteome that is dedicated to the ribosomes, consistent with observations in mechanistic models of heterotrophic growth (Molenaar et al., 2009; Scott et al., 2010; Weiße et al., 2015), as well as with the results in chapter 6. The points off the main line correspond to pairs of growth rate and ribosomal proteome fractions right after the shifts.

To understand which new enzymes are synthesized during the adaptation to anaerobic growth following a sudden oxygen removal, we have investigated the enzyme synthesis fluxes (per unit biomass) that are higher in the interval 6 – 12 *h* compared to the interval 0 – 6 *h* (figure 7.9).

## Chapter 7. Optimal resource allocation in yeast under dynamically changing environments

---

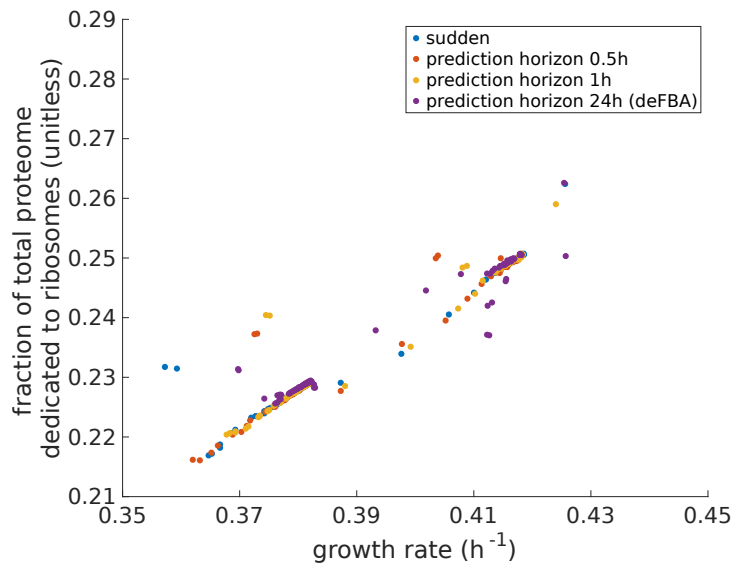


Figure 7.8: **The linear relationship between growth rate and the fraction of the total proteome that is dedicated to the ribosomes over the shift period.** Results correspond to the growth curves in figure 7.7.

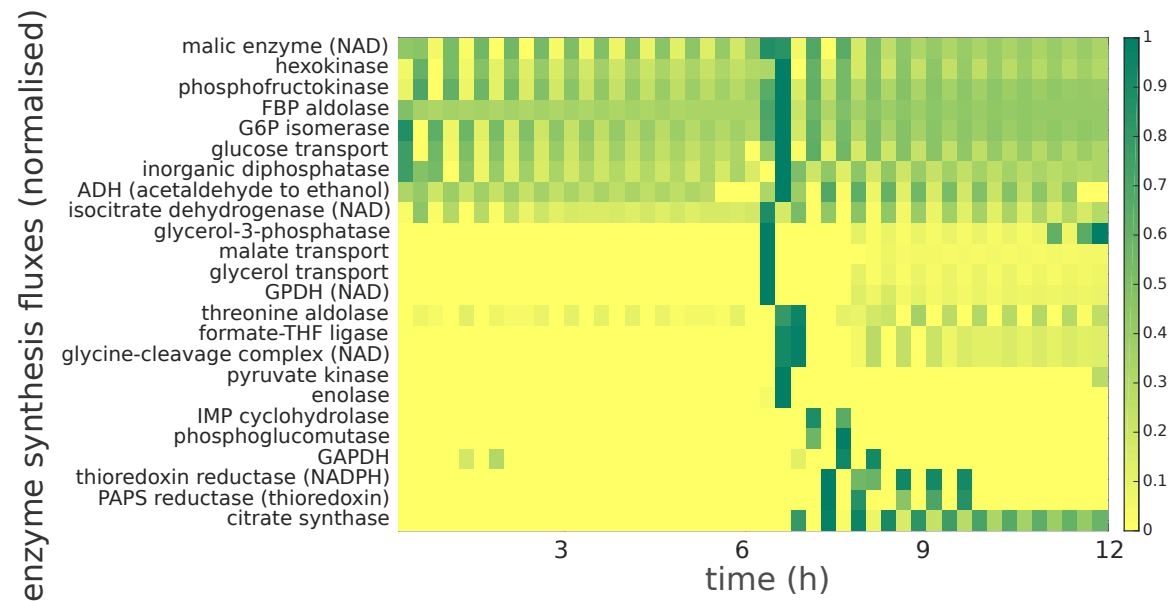


Figure 7.9: **Normalized enzyme synthesis fluxes per unit biomass that are higher in the interval 6 – 12 h compared to the interval 0 – 6 h for the case when the organism is faced with the sudden shift for the first time.** Results correspond to the first half of the continuous growth curve in figure 7.7. Abbreviations: NAD, nicotinamide adenine dinucleotide; FBP, fructose-bisphosphate; G6P, glucose 6-phosphate; ADH, alcohol dehydrogenase; GPDH, glycerol-3-phosphate dehydrogenase; THF, tetrahydrofolate; IMP, inosine monophosphate; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; PAPS, phosphoadenylyl-sulfate reductase.

## Chapter 7. Optimal resource allocation in yeast under dynamically changing environments

---

We observe a strong increase in the synthesis fluxes for enzymes belonging to the pentose phosphate pathway, fermentation, glycerol metabolism, as well as some glycolysis and TCA cycle enzymes at the beginning of the adaptation period. The increases in the pentose phosphate pathway and fermentation are very likely due to these pathways taking over the flux that can no longer run through respiration.

In addition, the impossibility to respire results in an  $\text{NAD}^+/\text{NADH}$  redox imbalance. According to (Albers et al., 1996; Ansell et al., 1997; Bakker et al., 2001; Pählman et al., 2001) fermentation is redox inert, i.e., the NADH produced in glycolysis is reoxidized by converting pyruvate to ethanol and  $\text{CO}_2$ . However, biomass production generates excess NADH, which has to be re-oxidized by redirecting part of the assimilated glucose to the formation and export of glycerol. Therefore, we conclude that the adaptation period is due to enzymes of the glycerol and pentose phosphate pathways having to be synthesized before maximal growth in anaerobic conditions is possible.

### Nutrient shifts

A recent study has shown that the length of the growth lag observed when yeast cells are switched from the preferred carbon source (glucose) to alternative nutrients like maltose, galactose, or ethanol differs between wild yeast strains (New et al., 2014). Moreover, the study points out that, after pregrowth on alternative nutrients followed by growth on glucose, the length of the growth lag increases with increasing time spent by the culture in the glucose medium. The lag is believed to be due to the time needed for the cells to synthesize the enzymes needed in the new environment.

In addition to testing whether our model can reproduce this behavior, we wanted to identify which enzymes are synthesized during the lag period. Thus, we have conducted an *in silico* experiment that subjects our model to such nutrient shifts. After identifying the biomass composition in aerobic galactose conditions, we have set this as initial condition and simulated aerobic growth on glucose for different time periods ranging from two to ten hours. After this, the system was shifted again to aerobic galactose conditions, while setting as initial biomass amount the biomass composition of the last time point in glucose.

The growth rate curves obtained from these shift experiments are shown in figure 7.10. We observe that the longer the cells are growing on glucose, the better they adapt their growth rate to this environment. Consistent with the findings in (New et al., 2014), we find also that the longer the time spent feeding on glucose is, the larger the growth rate drop when the cells are shifted again to galactose and the longer the time they need to adapt to galactose again.

While in both constant glucose and constant galactose environments yeast uses a respiro-fermentation strategy, we note that growth in constant galactose con-



ditions displays a higher rate of respiration compared to growth in constant glucose conditions. This is also the reason why we see the growth rate difference in figures 7.1 and 7.2.

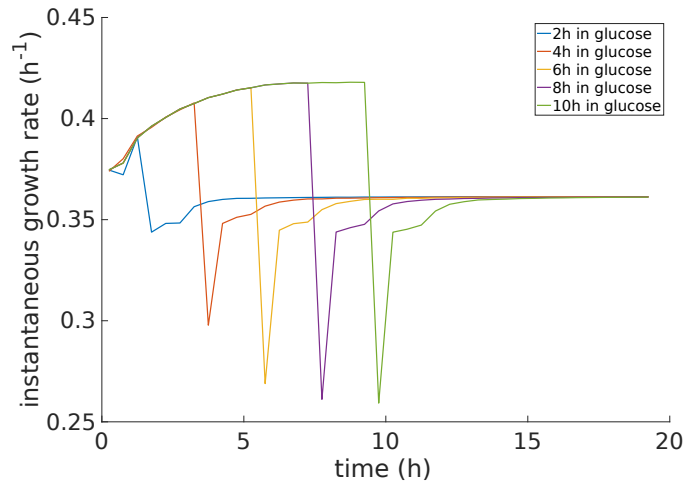


Figure 7.10: **Growth rates following nutrient shift from glucose to galactose.** The longer the time spent feeding on glucose, the larger is the growth rate drop when the cells are shifted again to galactose and the longer the time they need to adapt to galactose again.

To find the enzymes that are synthesized during the lag period, we have chosen the solution of the system that spent four hours in glucose before the switch to galactose. We have identified the enzyme synthesis fluxes that are on average higher right after the shift compared to their average before the shift. Figure 7.11 shows these normalized enzyme synthesis fluxes per unit biomass. We notice that the large majority of these enzymes are part of the TCA cycle and the respiration pathways. This is consistent with the findings of (Fendt and Sauer, 2010), who show that during galactose growth there is an increase in TCA cycle fluxes compared to glucose.

Shifting flux from fermentation to respiration however requires a large investment of resources into respiratory enzymes, which are not only larger but also much more numerous than the enzymes required by fermentation. Thus, the longer the cell population has spent in glucose, the more enzymes for respiration it has to produce and therefore the longer it needs for adaptation.

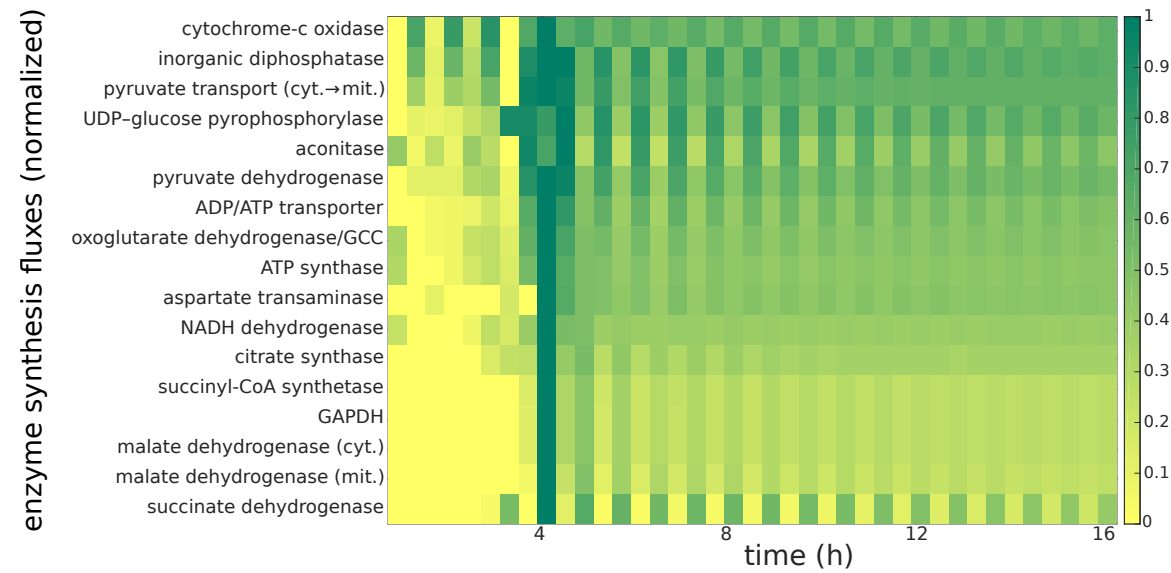


Figure 7.11: **Normalized enzyme synthesis fluxes per unit biomass that are on average higher right after the shift from glucose to galactose.** Results correspond to the nutrient shift after 4 hours of growth in glucose and are obtained using two chained deFBA problems with a sudden shift from glucose to galactose conditions in aerobic Verduyn medium enriched with sterols. Abbreviations: cyt., cytosol; mit., mitochondrion; UDP, uridine-5'-diphosphate; GCC, glycine cleavage complex; GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

### 7.3.5 Adaptations of yeast mutants to oxygen availability shifts

To find candidate genes that are involved in adaptation to sudden environment shifts of the type found in large-scale bioreactors we have screened all viable single knock-out mutants. These are mutants that survive in both constant aerobic and anaerobic glucose conditions. We have subjected these mutants to sudden oxygen availability shifts following the pattern in figure 7.6. We expect that the mutants that were viable in constant conditions but are no longer viable in the shift conditions lack a gene that is essential for surviving the shifts.

We have found fifteen genes for which the corresponding single knock-out mutants display this behavior. These genes are listed in table 7.1, together with the affected enzymes.

Table 7.1: Genes for which the corresponding single knock-out mutants are viable under constant aerobic and anaerobic glucose conditions, but not viable under oxygen shift conditions. We conjecture that these genes are essential for survival in dynamic oxygen availability conditions.

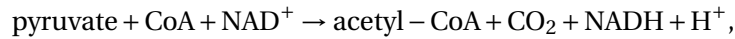
Gene	Affected enzyme
YBR221C	pyruvate dehydrogenase
YER178W	pyruvate dehydrogenase
YGR193C	pyruvate dehydrogenase
YNL071W	pyruvate dehydrogenase
	pyruvate dehydrogenase
YFL018C	oxoglutarate dehydrogenase
	glycine cleavage complex
YDR345C	glucose transporter isoenzyme 4
YJR158W	glucose transporter isoenzyme 13
YFL011W	glucose transporter isoenzyme 7, D-galactose transporter isoenzyme 1
YNL318C	D-galactose transport isoenzyme 5
YKR072C	phosphopantothoenoylcysteine decarboxylase isoenzymes 2 and 4
YML008C	S-adenosyl-methionine delta-24-sterol-c-methyltransferase
YNR013C	phosphate transport isoenzyme 4
YPL117C	isopentenyl-diphosphate D-isomerase

Of the enzymes in table 7.1, pyruvate dehydrogenase is used by the wildtype both in constant aerobic and anaerobic glucose conditions as well as in oxygen shift conditions. The glycine cleavage complex is active in the wildtype only in constant anaerobic glucose conditions and in the anaerobic period in case of an oxygen shift. All other enzymes in table 7.1 are not used in constant or oxygen shift conditions by the wildtype. However, mutants may have a different set of active reactions compared to the wildtype in both the aerobic and anaerobic part, where these enzymes are important.

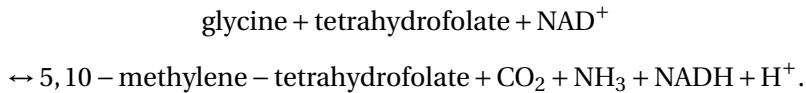
## Chapter 7. Optimal resource allocation in yeast under dynamically changing environments

---

The pyruvate dehydrogenase complex catalyzes the overall reaction



while the glycine cleavage complex catalyzes the reaction



We observe that both reactions have an impact on the  $\text{NAD}^+/\text{NADH}$  balance of the cell. As discovered in section 7.3.4, the main enzymes produced upon an oxygen availability shift are mainly responsible for stabilizing the  $\text{NAD}^+/\text{NADH}$  redox balance in the new conditions. In this context, we conjecture that the additional  $\text{NAD}^+/\text{NADH}$  imbalance induced by turning off pyruvate dehydrogenase or the glycine cleavage complex makes it impossible for the cell to survive the oxygen shift.

The issue deserves further experimental investigation. A computational investigation possibility is to relax the steady-state constraints for  $\text{NAD}^+$  and  $\text{NADH}$  within the resource allocation model, and to track changes in the balance between the two compounds. Additional insight may also be gained by the use of a kinetic model, where one can track changes in the  $\text{NAD}^+$  and  $\text{NADH}$  concentrations.

Since the rest of the enzymes in table 7.1 are not used in constant or oxygen shift conditions by the wildtype, they may appear there as a result of alternate optima within the first quarter of the shift period. One could argue that, since the corresponding reactions are not used in the wildtype, then the wildtype solution is a feasible solution for the mutant as well. However, if the use of an additional constraint for knocking out one gene results in a different yet still optimal solution in the first quarter of the shift period compared to the wildtype solution, then it is possible that this new solution can no longer be continued in the second quarter of the shift period, as opposed to the optimum used by the wildtype simulation. Unfortunately, running the variability analysis for a deFBA solution is very computationally intensive for such large models, so this hypothesis cannot be further investigated.

## 7.4 Discussion

In this chapter we have discussed the building and analysis of a dynamic resource allocation model for the yeast *Saccharomyces cerevisiae*. We have investigated growth under constant conditions, as well as environmental shifts concerning oxygen and nutrient availability.

We have seen that the resource allocation model provides an upper bound on the experimentally derived growth rates for the investigated constant conditions. In addition, the simulation under constant conditions also pinpointed

that overflow metabolism, i.e., the choice of yeast to ferment glucose even under aerobic conditions although this results in a suboptimal ATP yield, is an optimal behavior from a resource allocation perspective.

We were also able to prove, using the resource allocation model, that the diauxie is an optimal behavior. Both the overflow metabolism and the diauxie are behaviors that are widely spread in nature. Organisms have different regulatory structures that help implement these behaviors and we have shown here why these regulatory structures have evolved: to maximize growth rate by optimizing the allocation of limited resources.

The second part of this chapter was concerned with the study of resource allocation in dynamic environments. The motivation for this study comes from biotechnology, and is concerned with the decreased productivity of strains optimized for steady-state conditions when they are faced with the local heterogeneities of imperfectly mixed industry-scale bioreactors.

We have shown here that dynamic resource allocation models are a powerful tool for understanding why some of the strains do not survive in dynamic environments. In the case of yeast, shifts in oxygen availability were shown to cause imbalances in the  $\text{NAD}^+/\text{NADH}$  redox balance in our simulations. Additional knock-out simulations helped us show that two of the key enzymes involved in survival of oxygen availability shifts are the pyruvate dehydrogenase complex and the glycine cleavage complex. We conjecture that knock-out of these genes renders the  $\text{NAD}^+/\text{NADH}$  imbalance resulting from the shift impossible to restore. Additional experimental evidence is however needed to fully confirm this finding.

Nutrient shifts among preferred (glucose) and alternate (galactose) carbon sources were also investigated. We have shown that the recent observations in (New et al., 2014), that the longer yeast has spent feeding on the preferred nutrient the longer the growth rate lag upon switch to the alternate nutrient, can be predicted and explained using resource allocation arguments. According to our simulations, the lag is due to the time yeast needs to synthesize the large and numerous respiratory and TCA cycle enzymes it needs to grow optimally on the alternate carbon source.

However, we have also discovered some caveats of our methods in investigating sudden environment shifts. We have seen that, due to the possible existence of alternate optima, some genes were predicted to be essential for surviving oxygen availability shifts although the wildtype does not use or produce these genes before, during or after the shift. We believe that a next step in the analysis should be the development of efficient methods for analyzing deFBA solution variability. Currently, computing the solution variability is impossible due to the time and computational effort needed. A possible help may come from using flux coupling information (Burgard et al., 2004). For instance, information about fully

## **Chapter 7. Optimal resource allocation in yeast under dynamically changing environments**

---

coupled reactions, which are reactions whose fluxes always differ only by a factor, may help reduce the number of optimization problems to be solved.

An additional next step in understanding how organisms adapt to dynamic environments would be to try to infer the regulatory rules responsible for the transitions. While for yeast many of the regulatory structures are known, for other organisms they are still to be discovered. One idea would be to use time courses from dynamic resource allocation models for the inference. Since the only limit is the computational power, one can hope to be able to gather enough simulation data from such models to be able to infer the regulation using machine learning or other approaches.

To pursue this idea, we present in the next chapter some preliminary work using a toy model where we try to infer the catabolite repression mechanism, which is the regulatory structure behind the diauxie.

## Chapter 8

# Perspectives: inferring gene regulation from dynamic resource allocation models

We have seen that deFBA solves a global dynamic resource allocation problem, and thus at early time points in the solution it has knowledge of what will happen later and adjusts its resource allocation accordingly. This is not truly the case in biology, where cells can at most sense gradual changes in the environment (Kochanowski et al., 2013), but definitely do not have knowledge of these changes hours before they happen. We have also seen that an alternative method we could use is short term deFBA, where the prediction horizon determines how much in advance can the system sense a change in the environment. However, this prediction horizon has to be chosen in such a way that exponential growth is still the optimal strategy, as we have seen in section 5.1.2. Depending on the system, it can still be that this minimal prediction horizon is much larger than what is typically observed in real cells.

A second alternative would be an iterative RBA, where we fix a time step and at each iteration solve RBA problems after which we update the extracellular nutrient levels according to the predicted fluxes and the time step. However, in this case at each iteration we would have the same objective and gradual changes in the environment would not be translated into gradual changes in the biomass composition. This means we expect the biomass composition to remain constant, resulting in the system being “surprised” when some nutrients run out.

So the question still remains: how can we model such processes without knowledge of future events but also without the cell being entirely unprepared for environment changes? A strategy would be to do something similar to an iterative RBA, but change the objective function depending on the environment state. Thus, we can try to infer the different objective functions given environment

## Chapter 8. Perspectives: inferring gene regulation from dynamic resource allocation models

---

information using machine learning techniques and data from time courses of deFBA models. This will result in a control function that can be understood as a regulatory network and can be applied iteratively.

In this chapter we present some first steps in this direction. We note that, even if the learned regulatory structure is not close to what we expect regulatory networks to look like, it is still a benefit since it can be applied without knowledge of what will happen at later time steps.

### 8.1 Formalizing the question

Before we begin formalizing the inference problem, we remark that the only type of regulation we can infer from resource allocation models is transcriptional regulation. This means, we can only infer connections of the type

change in environment  $\rightarrow$  production of enzyme  $e$  is upregulated/downregulated.

There are several reasons why we are limited only to this type of regulation. First, resource allocation models cannot incorporate any kind of information related to posttranslational modifications (phosphorylation, methylation, reversible binding of cofactors) of enzymes. Second, internal metabolite amounts are assumed to be at steady-state, and thus not modeled explicitly. Hence, we cannot use these to infer metabolic regulation of the type end-product inhibition. Third, we would need a resolution of minutes or seconds in the time courses in order to account for posttranslational regulation, which is currently not possible in dynamic resource allocation models.

#### 8.1.1 Toy model for inference

Before we try to infer metabolic regulation in a genome-scale model, it is important to see that our tools work on toy models. Therefore, we limit ourselves in this chapter to inference of regulation in a toy model, and leave the genome-scale case as a future perspective.

An ideal test case for learning regulation is the core carbon toy network from (Covert et al., 2001), which we reproduce in figure 8.1.

In the (Covert et al., 2001) article the network is studied in combination with a set of regulatory rules. In addition, the network has been used to illustrate deFBA in (Waldherr et al., 2015; Lindhorst et al., 2016), and hence a deFBA model for it already exists. We provide the macromolecule production reactions and the turnover rates of this deFBA model in table 8.1.

The model, as described in (Waldherr et al., 2015; Lindhorst et al., 2016) and shown in figure 8.2, is capable of reproducing the diauxic shift behavior. This means it preferentially consumes  $\text{Carb1}_{ext}$  and only when this carbon source



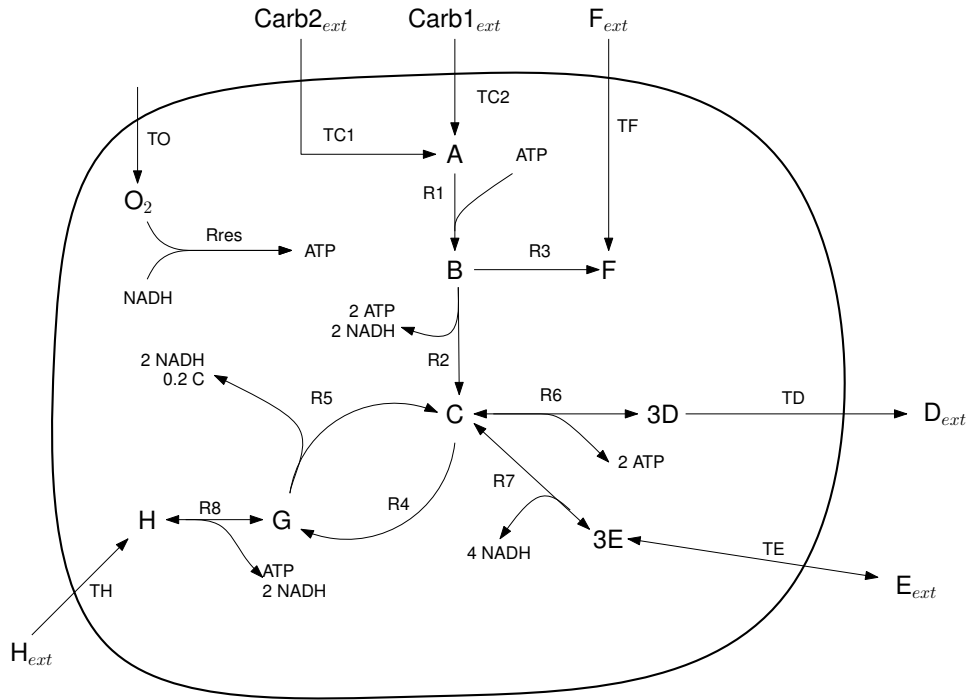


Figure 8.1: **Core carbon network for inferring regulation.** The network is reproduced from (Covert et al., 2001).

is exhausted it starts consuming Carb2<sub>ext</sub>. The diauxie is known to be due to catabolite repression (Deutscher, 2008), that is inhibition of the synthesis of the transporter for other carbon sources in the presence of the preferred carbon source.

This phenomenon is an ideal test case for our inference methods. We expect that an inhibitory connection between Carb1<sub>ext</sub> and the synthesis flux for TC2 is learned.

### 8.1.2 The general case

In this inference problem, we would like to find connections between extracellular metabolite amounts and enzyme synthesis fluxes.

Thus, we are given two matrices containing (several) time series obtained from a deFBA model with different initial values (rows correspond to the  $N$  time points, while columns correspond to metabolite amounts or fluxes):  $X \in \mathbb{R}^{K \times |\mathcal{X}|}$  contains the time courses of extracellular metabolite amounts,  $Y \in \mathbb{R}^{K \times |\mathcal{R}^E|}$  (where  $\mathcal{R}^E$  is the set of enzyme synthesis reactions) contains the time courses of the enzyme production fluxes, and  $K$  is the number of rows of the two matrices. In the most general case we want to learn a function  $g : \mathbb{R}^{|\mathcal{X}|} \rightarrow \mathbb{R}^{|\mathcal{R}^E|}$  that relates

## Chapter 8. Perspectives: inferring gene regulation from dynamic resource allocation models

Table 8.1: Metabolic, transport, and macromolecule production reactions and turnover rates for the model in figure 8.1 taken from (Lindhorst et al., 2016).

Reaction	Enzyme	Enzyme molecular weight	$k_{cat}$
Transport reactions			
Carb1 <sub>ext</sub> → A	TC1	4	3000
Carb2 <sub>ext</sub> → A	TC2	15	2000
F <sub>ext</sub> → F	TF	4	3000
→ O <sub>2</sub>	S	7.5	1000
D → D <sub>ext</sub>	S	7.5	1000
E <sub>ext</sub> ↔ E	S	7.5	1000
H → H <sub>ext</sub>	TH	4	3000
Metabolic reactions			
A + ATP → B	ER1	5	1800
B → C + 2 ATP + 2 NADH	ER2	5	1800
B → F	ER3	20	1800
C → G	ER4	5	1800
G → 0.8 C + 2 NADH	ER5	5	1800
C ↔ 2 ATP + 3 D	ER6	10	1800
C + 4 NADH → 3 E	ER7	10	1800
G + ATP + 2 NADH ↔ H	ER8	40	1800
NADH + O <sub>2</sub> → ATP	ERres	5	1800
Biomass reactions			
400 H + 1600 ATP → TC1	R	60	2.5
1500 H + 6000 ATP → TC2	R	60	0.67
400 H + 1600 ATP → TF	R	60	2.5
400 H + 1600 ATP → TH	R	60	2.5
250 H + 250 C + 250 F + 1500 ATP → S	R	60	3
500 H + 2000 ATP → ER1	R	60	2
500 H + 2000 ATP → ER2	R	60	2
2000 H + 8000 ATP → ER3	R	60	0.5
500 H + 2000 ATP → ER4	R	60	2
500 H + 2000 ATP → ER5	R	60	2
1000 H + 4000 ATP → ER6	R	60	1
1000 H + 4000 ATP → ER7	R	60	1
4000 H + 16000 ATP → ER8	R	60	0.25
500 H + 2000 ATP → ERres	R	60	2
4500 H + 1500 C + 21000 ATP → R	R	60	0.2

entries in  $X$  to entries in  $Y$ , i.e.

$$Y_{i*} = g(X_{i-1*}), \quad \forall i \in \{2, \dots, K\}.$$

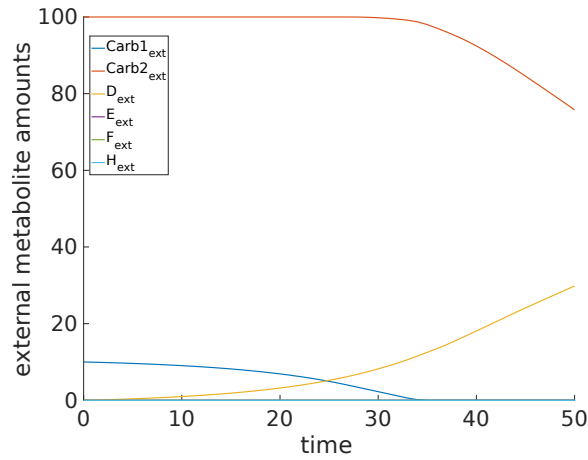


Figure 8.2: Extracellular metabolite time courses for the deFBA model in table 8.1, showing the diauxic shift under aerobic conditions.

Note that we use the enzyme production fluxes and not the enzyme amounts because the latter only increase in a deFBA model, while the fluxes can increase and decrease and thus mimic up- and downregulation.

### 8.2 The simplest case: a linear model

While in general regulatory interactions are nonlinear and best described by sigmoid functions (see (Jamshidi, 2012) and references therein), we discuss here the most simple possible case and assume  $g$  to be a linear function, i.e.

$$Y_{i*} = g(X_{i-1*}) := MX_{i-1*} + \epsilon_i, \quad \forall i \in \{2, \dots, K\}, \quad (8.1)$$

where  $\epsilon$  denotes the error terms, i.e., the difference between the predicted enzyme production fluxes  $\hat{\mathbf{v}}_{\mathcal{Q}^E}$  and the actual enzyme production fluxes  $Y$  used as training and obtained from the resource allocation model,

$$\epsilon := Y - \hat{\mathbf{v}}_{\mathcal{Q}^E} = Y - MX.$$

The matrix  $M$ , which can also be seen as the adjacency matrix of a regulatory network, then gives the influences of extracellular metabolite amounts on enzyme synthesis fluxes, with positive entries upregulating the production of the corresponding enzymes and negative entries downregulating it.

In statistics, the task of learning  $M$  is called multiple linear regression and is one of the simplest machine learning tools (Friedman et al., 2001). The method finds, for the system in equation 8.1, a matrix  $M$  such that the square error  $\epsilon^\top \epsilon$  is minimized (thus the name least squares regression).

One important property we would like to have for  $M$  is sparsity. This comes from two motivations: it is unlikely that any component has more than three

## Chapter 8. Perspectives: inferring gene regulation from dynamic resource allocation models

---

regulators due to crowding effects, and the more edges the regulatory network has the more we are prone to overfitting the data.

Shrinkage methods such as Ridge or Lasso regression minimize the regression coefficients in  $M$  by imposing a penalty on their size. This is done by introducing an additional term in the objective which penalizes either the 2- (in the case of Ridge) or the 1-norm (in the case of Lasso) of the coefficients using a Lagrange multiplier parameter  $\lambda$ . Thus, the objectives of Ridge and Lasso regression for the system in equation 8.1 are

$$\min_M \epsilon^T c - \lambda \|M\|_p,$$

with  $p = 2$  and  $p = 1$  respectively. A third shrinkage method, the elastic net, uses both the 1- and 2-norm penalties in the objective.

Such shrinkage methods are the easiest way of obtaining sparsity in the resulting network  $M$ . Thus, we use least squares, Ridge, Lasso, and elastic net regression for estimating the regulation network in this section.

We use as training data time courses of the toy deFBA model in figure 8.1 with the initial extracellular metabolite amounts in table 8.2 under aerobic conditions. We infer the four linear models using the python package scikit-learn (<http://scikit-learn.org/stable/index.html>).

Table 8.2: Initial extracellular metabolite amounts for the deFBA model to generate the training data. The deFBA problems were solved using the final time  $t_f = 70$  with  $N = 70$  discretization points for the dynamic variables.

Condition	Carb1 <sub>ext</sub>	Carb2 <sub>ext</sub>	D <sub>ext</sub>	E <sub>ext</sub>	F <sub>ext</sub>	H <sub>ext</sub>
1	2	120	0	0	0	0
2	10	110	0	0	0	0
3	20	100	0	0	0	0
4	30	90	0	0	0	0
5	40	80	0	0	0	0
6	50	70	0	0	0	0
7	60	60	0	0	0	0
8	70	50	0	0	0	0
9	80	40	0	0	0	0
10	90	30	0	0	0	0
11	100	20	0	0	0	0
12	110	10	0	0	0	0
13	120	2	0	0	0	0

After training the linear models we obtain four matrices,  $M_{LS}$ ,  $M_{Lasso}$ ,  $M_{Ridge}$ , and  $M_{EN}$ , which describe the influences of extracellular metabolite levels on enzyme production fluxes. We show the values in  $M_{LS}$ ,  $M_{Lasso}$ ,  $M_{Ridge}$ , and  $M_{EN}$  in

tables A11-A14 in Appendix E. We note that, in the case of the shrinkage methods, the influences of the Lagrange multiplier values on the results vary, with the Lasso being the most sensitive to the values of  $\lambda$  and the Ridge being the least sensitive. We have chosen the values in such a way that the smallest regression coefficients are used, while at the same time ensuring a good prediction behavior, as we will see next.

### 8.3 Evaluating the learned regulation

After learning the regulation, we would like to know how close the system that includes regulation is to the original deFBA time courses. We expect a lower growth rate, since deFBA shows us the behavior of the system assuming an ideal regulatory system, i.e., assuming the error  $\epsilon$  in equation (8.1) is zero.

To evaluate our learned regulation, we have developed an iterative procedure where at each time point the learned regulation is applied to determine the target enzyme production fluxes  $\hat{\mathbf{v}}_{\mathcal{R}^E}$ . Then, a linear program (LPreg) is used to obtain a solution that is as close as possible to these target synthesis fluxes, while obeying a steady-state for the internal metabolite amounts, irreversibilities, and enzyme capacity constraints.

Note that  $\hat{\mathbf{v}}_{\mathcal{R}^E}(t_{i+1})$  is a prediction of the new target fluxes given the extracellular metabolite amounts  $\mathbf{n}_{\mathcal{X}}(t_i)$ .  $\hat{\mathbf{v}}_{\mathcal{R}^E}(t_{i+1})$  is different from the training data  $Y$ .

We start by presenting the algorithm to compute such an iterative regulation enzyme-cost solution:

---

**Algorithm 1** Algorithm to compute an iterative regulation enzyme-cost solution.

---

**Input:** vector  $\mathbf{n}(t_0)$  of initial enzyme and external metabolite amounts, learned function  $g$  s.t.  $\hat{\mathbf{v}}_{\mathcal{R}^E}(t_{i+1}) = g(\mathbf{n}_{\mathcal{X}}(t_i))$ , step size  $\Delta t$ , end time  $t_f$ .

$i := 0$

**while**  $t_i < t_f$  **do**

$t_{i+1} := t_i + \Delta t$ .

Compute target enzyme production fluxes  $\hat{\mathbf{v}}_{\mathcal{R}^E}(t_{i+1}) := g(\mathbf{n}_{\mathcal{X}}(t_i))$ .

Find  $\mathbf{v}(t_{i+1})$  from  $\hat{\mathbf{v}}_{\mathcal{R}^E}(t_{i+1})$  and  $\mathbf{n}(t_i)$  using (LPreg).

**if**  $\exists \mathbf{v}(t_{i+1})$  **then**

Compute new amounts  $\mathbf{n}(t_{i+1}) = \mathbf{n}(t_i) + \mathbf{S}\mathbf{v}(t_{i+1})\Delta t$ .

**else**

Break.

**end if**

$i := i + 1$ .

**end while**

**Return**  $\mathbf{n}(t), \mathbf{v}(t) \forall t \in \{t_1, \dots, t_{i-1}\}$

---

If the organism cannot grow in the given environment using the resources in

## Chapter 8. Perspectives: inferring gene regulation from dynamic resource allocation models

---

$\mathbf{n}(t_0)$ , then Algorithm 1 will return infeasible. If this is not the case, then Algorithm 1 will run either until it has reached the given end time  $t_f$ , or until it is not possible to find a new flux vector using (LPreg) because the extracellular nutrients are exhausted.

To compute enzyme production fluxes that are as close as possible to the target fluxes  $\hat{\mathbf{v}}_{\mathcal{R}^E}(t_{i+1})$  given the resource allocation model, the enzyme and extracellular metabolite amounts  $\mathbf{n}(t_i)$ , and the step size  $\Delta t$  we solve the following linear program:

$$\begin{aligned}
 & \max_{\gamma(t_{i+1}) \in \mathbb{R}, \mathbf{v}(t_{i+1}) \in \mathbb{R}^{|\mathcal{R}|}} \gamma(t_{i+1}) \\
 & \text{s.t. } \gamma(t_{i+1}) \cdot \hat{\mathbf{v}}_{\mathcal{R}^E}(t_{i+1}) \leq \mathbf{v}_{\mathcal{R}^E}(t_{i+1}) \\
 & \quad S_{\mathcal{M}^*} \mathbf{v}(t_{i+1}) = 0 \\
 & \quad \sum_{i \in \mathcal{V}_k} \frac{\mathbf{v}_i(t_{i+1})}{k_{cat}^{i+}} \leq \mathbf{n}_k(t_i) \quad \forall k \in \mathcal{E} \\
 & \quad - \sum_{i \in \mathcal{V}_k \setminus \text{Irr}} \frac{\mathbf{v}_i(t_{i+1})}{k_{cat}^{i-}} \leq \mathbf{n}_k(t_i) \quad \forall k \in \mathcal{E} \quad (\text{LPreg}) \\
 & \quad \mathbf{n}_{\mathcal{X}}(t_i) - S_{\mathcal{X}^*} \mathbf{v}(t_{i+1}) \Delta t \geq 0 \\
 & \quad \mathbf{v}_{\text{Irr}}(t_{i+1}) \geq 0
 \end{aligned}$$

The linear program (LPreg) computes a flux vector  $\mathbf{v}_{\mathcal{R}^E}(t_{i+1})$  such that internal metabolites are kept at steady-state, enzyme capacity constraints are obeyed, no more external metabolite amounts are used than available in the given step size, and irreversible reactions only go forward. This flux vector is the vector that contains the enzyme production fluxes closest to the target values  $\hat{\mathbf{v}}_{\mathcal{R}^E}(t_{i+1})$  obtained by applying the regulation function, and at the same time maximizes growth by maximizing the enzyme production fluxes.

The (LPreg) is solved at each iteration of Algorithm 1 and for each of these iterations a different  $\gamma(t_{i+1})$  is computed. The role of  $\gamma(t_{i+1})$  in (LPreg) is to bring  $\mathbf{v}_{\mathcal{R}^E}(t_{i+1})$  as close as possible to  $\hat{\mathbf{v}}_{\mathcal{R}^E}(t_{i+1})$ . Moreover, a closer look at (LPreg) will reveal that, by maximizing  $\gamma(t_{i+1})$ , we implicitly maximize  $\mathbf{v}_{\mathcal{R}^E}(t_{i+1})$  and hence we also maximize biomass production.

Other ways that could be used to bring  $\mathbf{v}_{\mathcal{R}^E}(t_{i+1})$  as close as possible to  $\hat{\mathbf{v}}_{\mathcal{R}^E}(t_{i+1})$  would be minimizing the 1-, 2- or infinity-norm of their difference as

$$\min_{\mathbf{v}(t_{i+1}) \in \mathbb{R}^{|\mathcal{R}|}} \left\| \mathbf{v}_{\mathcal{R}^E}(t_{i+1}) - \hat{\mathbf{v}}_{\mathcal{R}^E}(t_{i+1}) \right\|_c,$$

with  $c \in \{1, 2, \infty\}$  and eliminating the constraint involving  $\gamma(t_{i+1})$ . We have tried these objective functions as well, but the resulting time course predictions were worse than when using the  $\gamma$  formulation in (LPreg).

Since the function  $g$  representing the regulation was learned from the time courses of an optimal resource allocation model, by being as close as possible

### 8.3 Evaluating the learned regulation

---

to the values predicted by  $g$ , we are as close as possible to the optimal resource allocation, while at the same time taking into account the regulatory part. We note that this iterative scheme is setting a different objective function at each step by allowing a different  $\gamma(t_{i+1})$  and requiring different target enzyme synthesis fluxes at each time point.

We have applied Algorithm 1 for the four linear models  $M_{LS}$ ,  $M_{Lasso}$ ,  $M_{Ridge}$ , and  $M_{EN}$  to evaluate how close the iterative regulation enzyme-cost solution is to the deFBA solution. We show the resulting dynamics of enzyme synthesis fluxes and extracellular metabolite amounts in figures 8.3 and 8.4, respectively.

We observe that in all four cases there is a significant delay in the time courses of the iterative regulation enzyme-cost solution compared to the deFBA time courses. However, the overall qualitative behavior is strikingly similar, suggesting that the linear predictors are a good model.

On the other hand, the matrices  $M_{LS}$ ,  $M_{Lasso}$ ,  $M_{Ridge}$ , and  $M_{EN}$  are far from sparse or intuitive. They all display a negative influence of  $\text{Carb1}_{ext}$  on the production of the transporter for  $\text{Carb2}_{ext}$ , which is why we do see preferential uptake of  $\text{Carb1}_{ext}$  in figure 8.4 for all four models. However, this connection comes along with other influences in  $M_{LS}$ ,  $M_{Lasso}$ ,  $M_{Ridge}$ , and  $M_{EN}$  which are hard to understand.

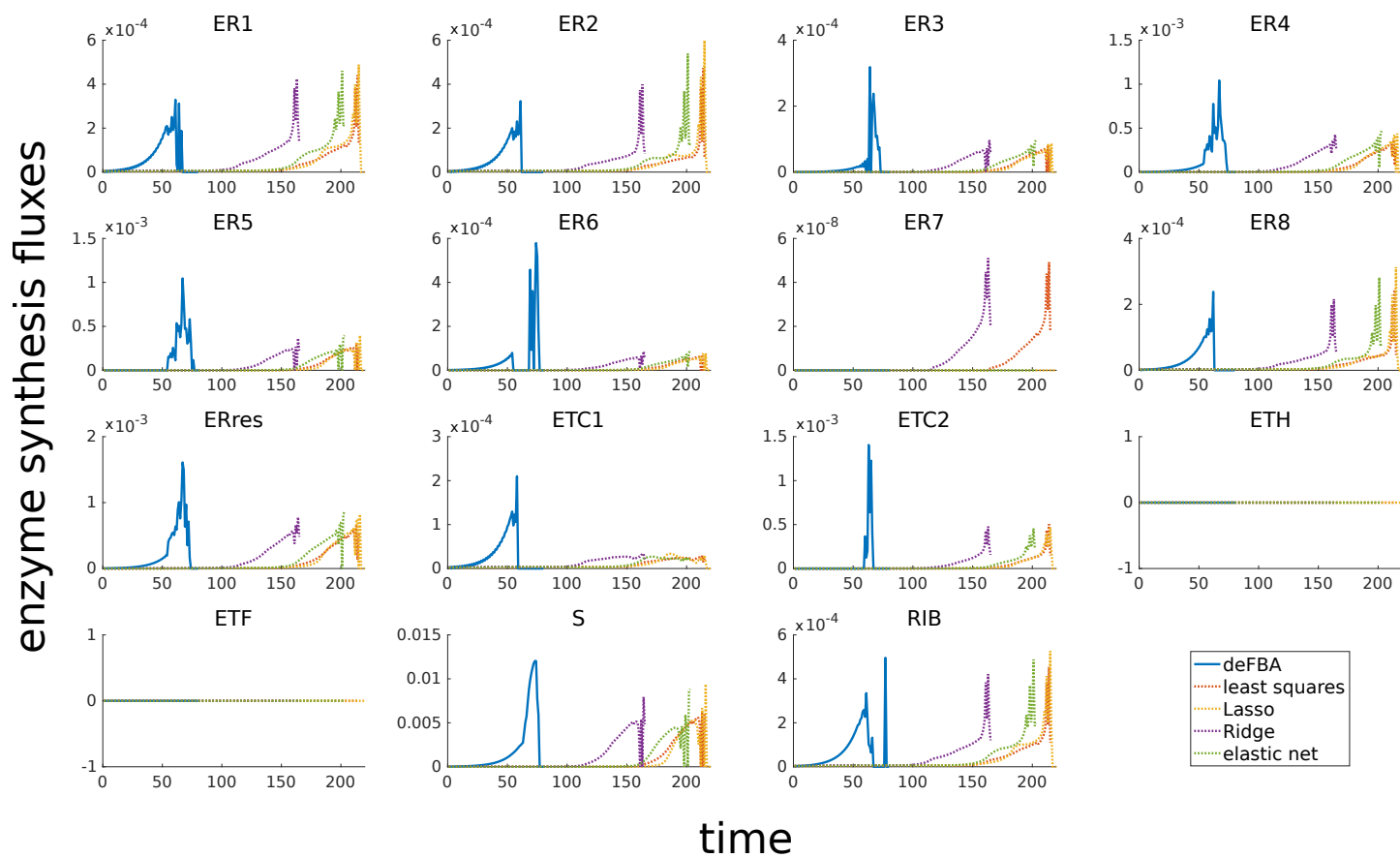


Figure 8.3: **Evaluation of the learned linear regulation models.** Shown are the enzyme synthesis time courses for the deFBA model using  $\mathbf{c}_{\text{Carb1}_{ext}}(0) = 120$ ,  $\mathbf{c}_{\text{Carb2}_{ext}}(0) = 100$ ,  $\mathbf{c}_{\text{D}_{ext}}(0) = \mathbf{c}_{\text{E}_{ext}}(0) = \mathbf{c}_{\text{F}_{ext}}(0) = \mathbf{c}_{\text{H}_{ext}}(0) = 0$  as initial conditions (continuous curves), and the time courses obtained using the iterative enzyme-cost regulation in Algorithm 1 for the different linear models (dotted curves).



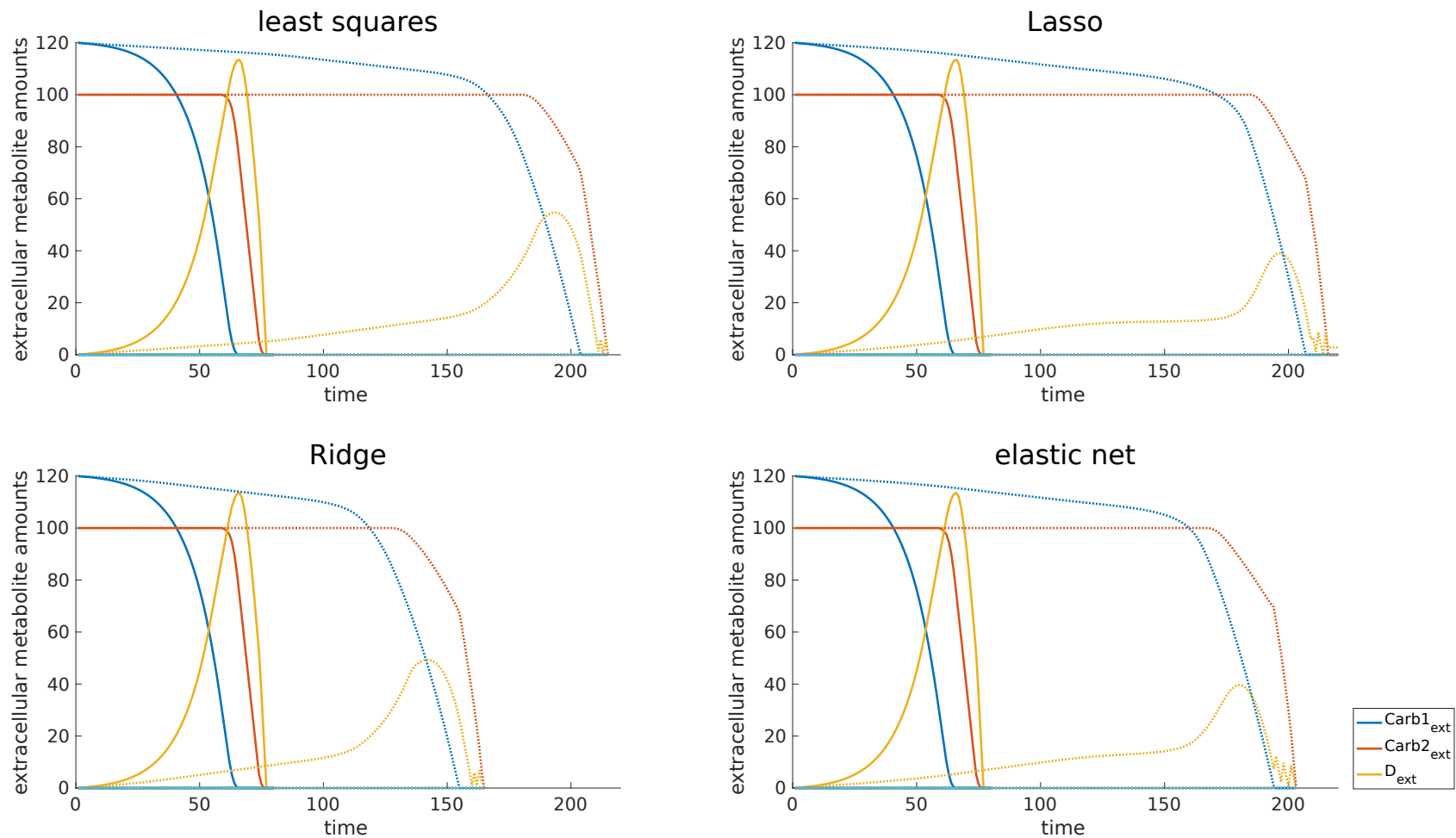


Figure 8.4: **Evaluation of the learned linear regulation models.** Shown are the extracellular metabolite time courses for the deFBA model using  $\mathbf{c}_{\text{Carb1}_{ext}}(0) = 120$ ,  $\mathbf{c}_{\text{Carb2}_{ext}}(0) = 100$ ,  $\mathbf{c}_{\text{D}_{ext}}(0) = \mathbf{c}_{\text{E}_{ext}}(0) = \mathbf{c}_{\text{F}_{ext}}(0) = \mathbf{c}_{\text{H}_{ext}}(0) = 0$  as initial conditions (continuous curves), and the time courses obtained using the iterative enzyme-cost regulation in Algorithm 1 for the different linear models (dotted curves).

## **8.4 Discussion**

This chapter presented some first steps in the direction of inferring regulatory structures using time courses from dynamic resource allocation models. We have seen that it is possible to infer matrices that implement the regulation responsible for diauxie using simple machine learning tools such as linear regression. Moreover, we have presented here an iterative algorithm that applies the learned regulation while taking into account enzyme capacity constraints. The benefit of this iterative procedure compared to deFBA is that, when making the prediction for the current time point, the model no longer has knowledge of what will happen in the future.

We have seen that, although we only used linear models, we can rediscover the general shape of the enzyme synthesis time courses predicted by deFBA. In addition, although we observe a lag period compared to deFBA, the extracellular metabolite time courses resulting from this procedure display the preferential use of one carbon source over the alternative.

However, the inferred matrices are hard to interpret and do not necessarily reflect the general structure of regulation networks encountered in biology. What we would expect to see is rather a sparse network with three layers: an input layer consisting of extracellular metabolite time courses and exchange reaction fluxes which then influence a middle layer of transcription factors, which then in turn influence an “output” layer, namely the enzyme synthesis fluxes. This resembles more the structure of another very popular machine learning tool: an artificial neural network with one hidden layer which represents the transcription factors.

We are thus tempted to assume that, if we were to use a neural network instead of a linear regression model, we might be able to get even closer to the deFBA time courses than we are in figures 8.3 and 8.4. By using a logistic activation function for the middle layer, we would be even closer to the sigmoid shape of activation and inhibition used in ODE models of gene regulatory networks (Klipp et al., 2008).

On the other hand, this can come at several expenses. First, we would need much more training data than we have used in this perspectives chapter. Second, we would be fitting more parameters and we would be in more danger of overfitting our data. Third, finding global optima when inferring neural networks is hard since we are no longer in the well-studied linear case, but we are actually fitting nonlinear functions. Last, there would always be the questions of how large should the hidden layer be, what values should we use for the initial guess, how should we ensure sparsity?

It may be helpful to already enforce some regulatory links to be present in the network, based on prior knowledge that we can obtain from databases such as Yeasttract (Teixeira et al., 2014). This is not trivial since it comes at the expense of

having to implement the solving routine to incorporate additional constraints on neural network edges.

An important extension to the iterative scheme and the learning procedure would also be to include the enzyme amounts and the import fluxes for nonlimiting extracellular metabolites as explanatory variables in the inference problem. An example would be, for the toy model in figure 8.1, the import flux for oxygen. Since oxygen is not assumed to be limiting, its extracellular amount is not modeled, but only the import flux. Thus, for learning regulatory mechanisms responsible for oxygen shifts of the type in chapter 7, it is important to allow connections between the oxygen import flux and the enzyme synthesis fluxes.



## Chapter 9

# Conclusions

This PhD thesis is dedicated to the study of dynamic resource allocation in metabolic networks. In the beginning of my PhD studies the state of the art in this topic was limited to the existence of two related methods: dynamic enzyme-cost flux balance analysis (Waldherr et al., 2015) and conditional flux balance analysis (Rügen et al., 2015). These methods had been applied, as a proof of concept, to a small core carbon network adapted from (Covert et al., 2001) and to a minimal model of phototrophic growth, respectively. In addition, there were two related formalisms for steady-state resource allocation studies: resource balance analysis (Goelzer et al., 2011) and gene expression models (Lerman et al., 2012). These had already been successfully applied at genome scale for the study of *Bacillus subtilis*, *Thermatoga maritima*, and *Escherichia coli*.

However, there were no guidelines for how to construct a dynamic genome-scale resource allocation model. Thus, my first contribution was to find the ingredients of such models, how they come together and depend on each other, what are the units of each ingredient, and where to find all the data needed to build such a model. This is now formulated as a guide and presented in chapter 4. In addition to this guide, chapter 4 also provides a standard for encoding such models in SBML, for ease of exchange and reproducibility.

When I started this PhD there was also no software prototype for solving dynamic resource allocation problems at genome-scale. While implementing this software I had to tackle all the mathematical difficulties presented in chapter 5. This started with solution dependencies on simulation end point, and continued with numerical discretization of the dynamic variables and problem reformulation, giving upper bounds on the model-derived growth rates, suitably scaling the resulting linear programs, analyzing the convexity of the quadratically constrained cFBA problem and thus justifying that binary search is a suitable solving approach, and ending with the actual software implementation and solver interfaces.

## Chapter 9. Conclusions

---

All the results in chapter 5 made it possible to study dynamic resource allocation at genome-scale for the first time. This was done in two case studies, on *Synechococcus elongatus* in chapter 6, and on *Saccharomyces cerevisiae* in chapter 7.

The case study on *Synechococcus elongatus* pinpointed several energetic trade-offs and constraints related to diurnal growth in cyanobacteria. Among the key results was the observation that growth is dynamic and takes place during the light phase. This was shown to be an optimal behavior that arises as a trade-off between the cost of not using existing enzymes and ribosomes during dark and the additional storage costs that would be required to continue using these enzymatic resources at night. Moreover, the linear pattern of glycogen accumulation throughout the day was shown to be an optimal behavior. The close match between the model-derived and experimental growth rates pointed out that cyanobacterial metabolism operates close to optimality.

The study of dynamic resource allocation in *Saccharomyces cerevisiae* subjected to environmental changes showed that two seemingly suboptimal behaviors that are widely spread in nature, overflow metabolism and diauxie, are in fact consequences of an optimal allocation of enzyme resources. Moreover, we have discovered that, in case of oxygen availability fluctuations, a key property that allows some of the individuals to survive is their ability to rewire their metabolism and to restore their  $\text{NAD}^+/\text{NADH}$  redox balance. Finally, we have shown that the experimentally observed increasing growth rate lag that arises upon a nutrient shift after increasing time feeding on the preferred carbon source is due to the time yeast needs to synthesize the large and numerous respiratory and TCA cycle enzymes needed to grow optimally on the alternate carbon source.

The natural next steps would be to use the time courses resulting from these dynamic resource allocation studies to infer the regulatory structures that coordinate the observed optimal behaviors. Chapter 8 provided some preliminary results in this direction using machine learning and an iterative regulation scheme combined with resource allocation. I believe that this idea has a lot of potential and that more effort should be dedicated to alternative inference tools besides linear models, with the natural next choice being neural networks.

Finally, another point where resource allocation studies may provide further understanding of biology are microbial community interactions. In such a microbial community, the species cooperate by exchanging metabolites and this way grow faster than if they would not cross-feed. However, as pointed out in (Gottstein et al., 2016), existing approaches either fail to predict the interactions in the first place as is the case for FBA, or easily become computationally intractable due to the nonlinear kinetic description of the exchange fluxes, as is the case for dynamic FBA. One of the main challenges arising in such community models is also modeling organisms that grow at different rates, but still cooperate and survive together. I argue that, in the end, the cooperative behavior

---

is due to sharing resources and that dynamic resource allocation studies using deFBA may provide solutions to the challenges encountered so far.





# Appendix

## A The steady-state assumption

### Calculation of dilution fluxes

In the following calculations all equalities should be considered as rough approximations.

### Calculation for *E. coli*

<b>Volume:</b>	$1 \text{ cell} = 0.6 \mu\text{m}^3 = 0.6 \cdot 10^{-15} \text{ dm}^3 = 0.6 \cdot 10^{-15} \text{ L}$
<b>Dry weight:</b>	$1 \text{ cell} = 0.489 \text{ pgDW} = 4.89 \cdot 10^{-13} \text{ gDW}$
<b>Flux:</b>	$ v  = 12 \text{ mmol}/(\text{gDW} \cdot \text{h}) = 12 \cdot 10^{-3} \text{ mol}/(\text{gDW} \cdot \text{h})$ $= 12 \cdot 10^{-3} \cdot 4.89 \cdot 10^{-13} \text{ mol}/(\text{cell} \cdot \text{h}) = \frac{12 \cdot 4.89 \cdot 10^{-16} \text{ mol}}{3600 \text{ s} \cdot \text{cell}}$ $= \mathbf{1.63 \cdot 10^{-18} \text{ mol}/(\text{s} \cdot \text{cell})}$
<b>Concentration :</b>	$1 \text{ mM} = 10^{-3} \text{ mol}/\text{L} = 10^{-3} \text{ mol}/\text{L} \cdot 0.6 \cdot 10^{-15} \text{ L}/\text{cell}$ $= 0.6 \cdot 10^{-18} \text{ mol}/\text{cell}$
<b>Growth:</b>	$\mu = 0.9/\text{h} = \frac{0.9}{3600 \text{ s}} = 2.5 \cdot 10^{-4} \text{ s}^{-1}$ $\mu \cdot 1 \text{ mM} = \mathbf{1.5 \cdot 10^{-22} \text{ mol}/(\text{s} \cdot \text{cell})}$

---

### Calculation for *S. cerevisiae*

**Volume:**  $1 \text{ cell} = 20 \mu\text{m}^3 = 2 \cdot 10^{-14} \text{L}$

**Dry weight:**  $1 \text{ cell} = 10^{-11} \text{gDW}$

**Flux:**  $|v| = 0.5 \text{mmol}/(\text{h} \cdot \text{gDW}) = 0.5 \cdot 10^{-3} \text{mol}/(\text{h} \cdot \text{gDW})$   
 $= 0.5 \cdot 10^{-3} \cdot 10^{-11} \text{mol}/\text{h} = \frac{0.5 \cdot 10^{-3} \cdot 10^{-11} \text{mol}}{3600 \text{s} \cdot \text{cell}}$   
 $= \mathbf{1.38 \cdot 10^{-18} \text{mol}/(\text{s} \cdot \text{cell})}$

**Concentration :**  $1 \text{mM} = 10^{-3} \text{mol}/\text{L} = 10^{-3} \text{mol}/\text{L} \cdot 2 \cdot 10^{-14} \text{L}/\text{cell}$   
 $= 2 \cdot 10^{-17} \text{mol}/\text{cell}$

**Growth:**  $\mu = 0.4/\text{h} = \frac{0.4}{3600 \text{s}} = 1.11 \cdot 10^{-4} \text{s}^{-1}$   
 $\mu \cdot 1 \text{mM} = \mathbf{2.22 \cdot 10^{-21} \text{mol}/(\text{s} \cdot \text{cell})}$

### Calculation for *H. sapiens* (HeLa)

**Volume:**  $1 \text{ cell} = 2600 \mu\text{m}^3 = 2.6 \cdot 10^{-12} \text{L}$

**Protein weight:**  $1 \text{ cell} = 150 \text{pgProtein} = 1.5 \cdot 10^{-7} \text{mgProtein}$

**Flux:**  $v = 18 \text{nmol}/(\text{min} \cdot \text{mgProtein}) = 18 \cdot 10^{-9} \text{mol}/(\text{min} \cdot \text{mgProtein})$   
 $= 18 \cdot 10^{-9} \cdot 1.5 \cdot 10^{-7} \text{mol}/(\text{min} \cdot \text{cell}) = \frac{18 \cdot 10^{-9} \cdot 1.5 \cdot 10^{-7} \text{mol}}{60 \text{s} \cdot \text{cell}}$   
 $= \mathbf{4.5 \cdot 10^{-17} \text{mol}/(\text{s} \cdot \text{cell})}$

**Concentration :**  $1 \text{mM} = 10^{-3} \text{mol}/\text{L} = 10^{-3} \text{mol}/\text{L} \cdot 2.6 \cdot 10^{-12} \text{L}/\text{cell}$   
 $= 2.6 \cdot 10^{-15} \text{mol}/\text{cell}$

**Growth:**  $\mu = 0.06/\text{h} = \frac{0.06}{3600 \text{s}} = 1.67 \cdot 10^{-5} \text{s}^{-1}$   
 $\mu \cdot 1 \text{mM} = \mathbf{4.34 \cdot 10^{-20} \text{mol}/(\text{s} \cdot \text{cell})}$

## B The SBML representation of resource allocation models: illustration using a toy example

---

### B The SBML representation of resource allocation models: illustration using a toy example

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level3/version1/core"
  level="3" version="1" xmlns:fb="http://www.sbml.org/sbml/
  level3/version1/fbc/version2" fbc:required="false">
  <model id="resalloc1" name="resalloc1" fbc:strict="false">
    <notes>
      <body xmlns="http://www.w3.org/1999/xhtml">
        <p>This model uses the ram standard 1.0</p>
      </body>
    </notes>

    <listOfCompartments>
      <compartment id="ext" name="external" constant="true"/>
      <compartment id="cytosol" constant="true"/>
    </listOfCompartments>

    <listOfSpecies>
      <species id="N1" compartment="ext" initialAmount="10"
        constant="false" boundaryCondition="false"
        hasOnlySubstanceUnits="true"/>
      <species id="N2" compartment="ext" initialAmount="10"
        constant="false" boundaryCondition="false"
        hasOnlySubstanceUnits="true"/>
      <species id="O2" compartment="ext" constant="true"
        boundaryCondition="true" hasOnlySubstanceUnits="true"
        initialAmount="1000"/>
      <species id="AA" compartment="cytosol" initialAmount="0"
        constant="false" boundaryCondition="false"
        hasOnlySubstanceUnits="true"/>
      <species id="ATP" compartment="cytosol" initialAmount="0"
        constant="false" boundaryCondition="false"
        hasOnlySubstanceUnits="true"/>
      <species id="N" compartment="cytosol" initialAmount="0"
        constant="false" boundaryCondition="false"
        hasOnlySubstanceUnits="true"/>
      <species id="Stor" name="Storage molecules" compartment="
        cytosol" initialAmount="0.0" constant="false"
        boundaryCondition="false" hasOnlySubstanceUnits="true"
      >
    </listOfSpecies>
    <annotation>
```

---

```

    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:species ram:molecularWeight="weight_0" ram:
        objectiveWeight="zero" ram:biomassPercentage="
          zero" ram:speciesType="storage"/>
    </ram:RAM>
  </annotation>
</species>
<species id="Complex_1" name="Transporter complex"
  compartment="cytosol" initialAmount="
    0.004166667083123639" constant="false"
  boundaryCondition="false" hasOnlySubstanceUnits="true"
  >
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:species ram:molecularWeight="weight_1" ram:
        objectiveWeight="weight_11" ram:biomassPercentage
          ="zero" ram:speciesType="enzyme"/>
    </ram:RAM>
  </annotation>
</species>
<species id="Emetab1" name="Generic metabolic enzyme"
  compartment="cytosol" initialAmount="
    1.9442790116926745" constant="false" boundaryCondition
    ="false" hasOnlySubstanceUnits="true">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:species ram:molecularWeight="weight_2" ram:
        objectiveWeight="weight_12" ram:biomassPercentage
          ="zero" ram:speciesType="enzyme"/>
    </ram:RAM>
  </annotation>
</species>
<species id="Emetab2" name="Specialized metabolic enzyme"
  compartment="cytosol" initialAmount="
    0.008000000399798693" constant="false"
  boundaryCondition="false" hasOnlySubstanceUnits="true"
  >
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">

```

## B The SBML representation of resource allocation models: illustration using a toy example

---

```
<ram:species ram:molecularWeight="weight_3" ram:
  objectiveWeight="weight_13" ram:biomassPercentage
  ="zero" ram:speciesType="enzyme"/>
</ram:RAM>
</annotation>
</species>
<species id="Estor" name="Metabolic enzyme for handling
  storage" compartment="cytosol" initialAmount="0.0"
  constant="false" boundaryCondition="false"
  hasOnlySubstanceUnits="true">
<annotation>
  <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
    /304">
    <ram:species ram:molecularWeight="weight_4" ram:
      objectiveWeight="weight_14" ram:biomassPercentage
      ="zero" ram:speciesType="enzyme"/>
    </ram:RAM>
  </annotation>
</species>
<species id="Etrans1" name="Transporter enzyme"
  compartment="cytosol" initialAmount="
  0.005555555555555556" constant="false"
  boundaryCondition="false" hasOnlySubstanceUnits="true"
  >
<annotation>
  <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
    /304">
    <ram:species ram:molecularWeight="weight_5" ram:
      objectiveWeight="weight_15" ram:biomassPercentage
      ="zero" ram:speciesType="enzyme"/>
    </ram:RAM>
  </annotation>
</species>
<species id="Etrans2" name="Cofactor of transporter enzyme
  " compartment="cytosol" initialAmount="0.0" constant="
  false" boundaryCondition="false" hasOnlySubstanceUnits
  ="true">
<annotation>
  <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
    /304">
    <ram:species ram:molecularWeight="weight_6" ram:
      objectiveWeight="weight_16" ram:biomassPercentage
      ="zero" ram:speciesType="enzyme"/>
    </ram:RAM>
```

---

```

    </annotation>
  </species>
  <species id="R" name="Ribosome" compartment="cytosol"
    initialAmount="0.033641975308641964" constant="false"
    boundaryCondition="false" hasOnlySubstanceUnits="true"
  >
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:species ram:molecularWeight="weight_7" ram:
        objectiveWeight="weight_17" ram:biomassPercentage
          ="zero" ram:speciesType="enzyme"/>
    </ram:RAM>
  </annotation>
</species>
<species id="S" name="Structrual biomass component"
  compartment="cytosol" initialAmount="
  0.7499999999999999" constant="false" boundaryCondition
  ="false" hasOnlySubstanceUnits="true">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:species ram:molecularWeight="weight_8" ram:
        objectiveWeight="weight_18" ram:biomassPercentage
          ="biomp" ram:speciesType="quota"/>
    </ram:RAM>
  </annotation>
</species>
</listOfSpecies>

<listOfParameters>
  <parameter constant="true" id="kcatt" value="1800" />
  <parameter constant="true" id="kcatC" value="2400" />
  <parameter constant="true" id="kcat1" value="2000" />
  <parameter constant="true" id="kcat2" value="2500" />
  <parameter constant="true" id="zero" value="0.0" />
  <parameter constant="true" id="kcatR1" value="10" />
  <parameter constant="true" id="kcatR2" value="16.67" />
  <parameter constant="true" id="kcatR3" value="5" />
  <parameter constant="true" id="kcatR4" value="6.25" />
  <parameter constant="true" id="kcatR5" value="1" />
  <parameter constant="true" id="kcatR6" value="10" />
  <parameter constant="true" id="kcatC1" value="6.25" />
  <parameter constant="true" id="kcatSf" value="25" />

```

## B The SBML representation of resource allocation models: illustration using a toy example

```
<parameter constant="true" id="kcatSb" value="30" />
<parameter constant="true" id="main" value="0.2" />
<parameter constant="true" id="biomp" value="0.15" />
<parameter constant="true" id="weight_0" value="0.2" />
<parameter constant="true" id="weight_1" value="0.16" />
<parameter constant="true" id="weight_2" value="0.2" />
<parameter constant="true" id="weight_3" value="0.16" />
<parameter constant="true" id="weight_4" value="0.15" />
<parameter constant="true" id="weight_5" value="0.1" />
<parameter constant="true" id="weight_6" value="0.06" />
<parameter constant="true" id="weight_7" value="1.0" />
<parameter constant="true" id="weight_8" value="1.5" />
<parameter constant="true" id="weight_10" value="0.2" />
<parameter constant="true" id="weight_11" value="0.16" />
<parameter constant="true" id="weight_12" value="0.2" />
<parameter constant="true" id="weight_13" value="0.16" />
<parameter constant="true" id="weight_14" value="0.15" />
<parameter constant="true" id="weight_15" value="0.1" />
<parameter constant="true" id="weight_16" value="0.06" />
<parameter constant="true" id="weight_17" value="1.0" />
<parameter constant="true" id="weight_18" value="1.5" />
</listOfParameters>

<fbc:listOfGeneProducts>
  <fbc:geneProduct fbc:id="Etrans1" fbc:label="1*GTRANS1"
    fbc:associatedSpecies="Etrans1"/>
  <fbc:geneProduct fbc:id="Etrans2" fbc:label="1*GTRANS2"
    fbc:associatedSpecies="Etrans2"/>
  <fbc:geneProduct fbc:id="Complex_1" fbc:label="1*GTRANS1
    AND 1*GTRANS2" fbc:associatedSpecies="Complex_1" />
  <fbc:geneProduct fbc:id="Emetab1" fbc:label="1*GMETAB1"
    fbc:associatedSpecies="Emetab1"/>
  <fbc:geneProduct fbc:id="Emetab2" fbc:label="1*GMETAB2"
    fbc:associatedSpecies="Emetab2"/>
  <fbc:geneProduct fbc:id="R" fbc:label="1*GRIBOS" fbc:
    associatedSpecies="R"/>
  <fbc:geneProduct fbc:id="Estor" fbc:label="1*GSTORF" fbc:
    associatedSpecies="Estor"/>
</fbc:listOfGeneProducts>

<listOfReactions>
  <reaction id="Maintenance" reversible="false" fast="true">
    <annotation>
```

---

```

    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:reaction ram:kcatForward="zero" ram:
        kcatBackward="zero" ram:maintenanceScaling="main"
        />
    </ram:RAM>
  </annotation>
  <listOfReactants>
    <speciesReference species="AA" stoichiometry="50"
      constant="true"/>
    <speciesReference species="ATP" stoichiometry="60"
      constant="true"/>
  </listOfReactants>
</reaction>
<reaction id="PMetab1" reversible="false" fast="false">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:reaction ram:kcatForward="kcatR3" ram:
        kcatBackward="zero" ram:maintenanceScaling="zero"
        />
    </ram:RAM>
  </annotation>
  <fbc:geneProductAssociation fbc:id="Ribosome">
    <fbc:geneProductRef fbc:geneProduct="R" />
  </fbc:geneProductAssociation>
  <listOfReactants>
    <speciesReference species="AA" stoichiometry="200"
      constant="true"/>
    <speciesReference species="ATP" stoichiometry="800"
      constant="true"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="EMetab1" stoichiometry="1"
      constant="true"/>
  </listOfProducts>
</reaction>
<reaction id="Metab1_1" reversible="false" fast="true">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:reaction ram:kcatForward="kcat1" ram:
        kcatBackward="zero" ram:maintenanceScaling="zero"
        />

```



## B The SBML representation of resource allocation models: illustration using a toy example

---

```
</ram:RAM>
</annotation>
<fbc:geneProductAssociation fbc:id="Emetab1">
  <fbc:geneProductRef fbc:geneProduct="Emetab1" />
</fbc:geneProductAssociation>
<listOfReactants>
  <speciesReference species="N" stoichiometry="1"
    constant="true"/>
</listOfReactants>
<listOfProducts>
  <speciesReference species="AA" stoichiometry="1"
    constant="true"/>
  <speciesReference species="ATP" stoichiometry="1"
    constant="true"/>
</listOfProducts>
</reaction>
<reaction id="Trans2" reversible="true" fast="false">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:reaction ram:kcatForward="kcatC" ram:
        kcatBackward="kcatC" ram:maintenanceScaling="zero
        "/>
    </ram:RAM>
  </annotation>
  <fbc:geneProductAssociation fbc:id="Complex_1">
    <fbc:geneProductRef fbc:geneProduct="Complex_1"/>
  </fbc:geneProductAssociation>
  <listOfReactants>
    <speciesReference species="N2" stoichiometry="1"
      constant="true"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="N" stoichiometry="1"
      constant="true"/>
  </listOfProducts>
</reaction>
<reaction id="Metab1_2" reversible="false" fast="true">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:reaction ram:kcatForward="kcat2" ram:
        kcatBackward="zero" ram:maintenanceScaling="zero"
        />
    </ram:RAM>
  </annotation>
</reaction>
```

---

```

    </ram:RAM>
  </annotation>
  <fbc:geneProductAssociation fbc:id="Emetab2">
    <fbc:geneProductRef fbc:geneProduct="Emetab2" />
  </fbc:geneProductAssociation>
  <listOfReactants>
    <speciesReference species="N" stoichiometry="1"
      constant="true"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="AA" stoichiometry="1"
      constant="true"/>
    <speciesReference species="ATP" stoichiometry="1"
      constant="true"/>
  </listOfProducts>
</reaction>
<reaction id="Trans1" reversible="true" fast="false">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:reaction ram:kcatForward="kcat" ram:
        kcatBackward="kcat" ram:maintenanceScaling="zero
          "/>
    </ram:RAM>
  </annotation>
  <fbc:geneProductAssociation fbc:id="Etrans1">
    <fbc:geneProductRef fbc:geneProduct="Etrans1" />
  </fbc:geneProductAssociation>
  <listOfReactants>
    <speciesReference species="N1" stoichiometry="1"
      constant="true"/>
    <speciesReference species="O2" stoichiometry="2"
      constant="true"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="N" stoichiometry="1"
      constant="true"/>
  </listOfProducts>
</reaction>
<reaction id="Storing" reversible="true" fast="false" fbc:
  lowerFluxBound="kcatSf" fbc:upperFluxBound="kcatSb">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">

```

## B The SBML representation of resource allocation models: illustration using a toy example

---

```
<ram:reaction ram:kcatForward="kcatSf" ram:
  kcatBackward="kcatSb" ram:maintenanceScaling="
  zero"/>
</ram:RAM>
</annotation>
<fbc:geneProductAssociation fbc:id="Estor">
  <fbc:geneProductRef fbc:geneProduct="Estor" />
</fbc:geneProductAssociation>
<listOfReactants>
  <speciesReference species="AA" stoichiometry="200"
    constant="true"/>
    <speciesReference species="ATP" stoichiometry="
    300" constant="true"/>
</listOfReactants>
<listOfProducts>
  <speciesReference species="Stor" stoichiometry="1"
    constant="true"/>
</listOfProducts>
</reaction>
<reaction id="PEtrans1" reversible="false" fast="false">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
    /304">
      <ram:reaction ram:kcatForward="kcatR1" ram:
        kcatBackward="zero" ram:maintenanceScaling="zero"
        />
    </ram:RAM>
  </annotation>
  <fbc:geneProductAssociation fbc:id="Ribosome">
    <fbc:geneProductRef fbc:geneProduct="R" />
  </fbc:geneProductAssociation>
  <listOfReactants>
    <speciesReference species="AA" stoichiometry="100"
      constant="true"/>
    <speciesReference species="ATP" stoichiometry="400"
      constant="true"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="Etrans1" stoichiometry="1"
      constant="true"/>
  </listOfProducts>
</reaction>
<reaction id="PEtrans2" reversible="false" fast="false">
  <annotation>
```

---

```

    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:reaction ram:kcatForward="kcatR2" ram:
        kcatBackward="zero" ram:maintenanceScaling="zero"
        />
    </ram:RAM>
  </annotation>
  <fbc:geneProductAssociation fbc:id="Ribosome">
    <fbc:geneProductRef fbc:geneProduct="R" />
  </fbc:geneProductAssociation>
  <listOfReactants>
    <speciesReference species="AA" stoichiometry="60"
      constant="true"/>
    <speciesReference species="ATP" stoichiometry="240"
      constant="true"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="Etrans2" stoichiometry="1"
      constant="true"/>
  </listOfProducts>
</reaction>
<reaction id="PComplex_1" reversible="false" fast="false">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:reaction ram:kcatForward="kcatC1" ram:
        kcatBackward="zero" ram:maintenanceScaling="zero"
        />
    </ram:RAM>
  </annotation>
  <fbc:geneProductAssociation fbc:id="Ribosome">
    <fbc:geneProductRef fbc:geneProduct="R" />
  </fbc:geneProductAssociation>
  <listOfReactants>
    <speciesReference species="AA" stoichiometry="160"
      constant="true"/>
    <speciesReference species="ATP" stoichiometry="640"
      constant="true"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="Complex_1" stoichiometry="1"
      constant="true"/>
  </listOfProducts>
</reaction>

```

## B The SBML representation of resource allocation models: illustration using a toy example

---

```
<reaction id="PMetab2" reversible="false" fast="false">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops/304">
      <ram:reaction ram:kcatForward="kcatR4" ram:kcatBackward="zero" ram:maintenanceScaling="zero" />
    </ram:RAM>
  </annotation>
  <fbc:geneProductAssociation fbc:id="Ribosome">
    <fbc:geneProductRef fbc:geneProduct="R" />
  </fbc:geneProductAssociation>
  <listOfReactants>
    <speciesReference species="AA" stoichiometry="160" constant="true"/>
    <speciesReference species="ATP" stoichiometry="640" constant="true"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="EMetab2" stoichiometry="1" constant="true"/>
  </listOfProducts>
</reaction>
<reaction id="PEstor" reversible="false" fast="false">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops/304">
      <ram:reaction ram:kcatForward="kcatR3" ram:kcatBackward="zero" ram:maintenanceScaling="zero" />
    </ram:RAM>
  </annotation>
  <fbc:geneProductAssociation fbc:id="Ribosome">
    <fbc:geneProductRef fbc:geneProduct="R" />
  </fbc:geneProductAssociation>
  <listOfReactants>
    <speciesReference species="AA" stoichiometry="150" constant="true"/>
    <speciesReference species="ATP" stoichiometry="600" constant="true"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="Estor" stoichiometry="1" constant="true"/>
  </listOfProducts>
</reaction>
```

---

```

    </listOfProducts>
  </reaction>
  <reaction id="PR" name="Ribosome expression" reversible="
    false" fast="false">
    <annotation>
      <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
        /304">
        <ram:reaction ram:kcatForward="kcatR5" ram:
          kcatBackward="zero" ram:maintenanceScaling="zero"
        />
      </ram:RAM>
    </annotation>
    <fbc:geneProductAssociation fbc:id="Ribosome">
      <fbc:geneProductRef fbc:geneProduct="R" />
    </fbc:geneProductAssociation>
    <listOfReactants>
      <speciesReference species="AA" stoichiometry="1000"
        constant="true"/>
      <speciesReference species="ATP" stoichiometry="4000"
        constant="true"/>
    </listOfReactants>
    <listOfProducts>
      <speciesReference species="R" stoichiometry="1"
        constant="true"/>
    </listOfProducts>
  </reaction>
  <reaction id="Metab2" reversible="false" fast="true">
    <annotation>
      <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
        /304">
        <ram:reaction ram:kcatForward="kcat1" ram:
          kcatBackward="zero" ram:maintenanceScaling="zero"
        />
      </ram:RAM>
    </annotation>
    <fbc:geneProductAssociation fbc:id="Emetab2">
      <fbc:geneProductRef fbc:geneProduct="Emetab2" />
    </fbc:geneProductAssociation>
    <listOfReactants>
      <speciesReference species="N" stoichiometry="1"
        constant="true"/>
    </listOfReactants>
    <listOfProducts>

```

## B The SBML representation of resource allocation models: illustration using a toy example

---

```
<speciesReference species="AA" stoichiometry="1"
  constant="true"/>
<speciesReference species="ATP" stoichiometry="2"
  constant="true"/>
</listOfProducts>
</reaction>
<reaction id="PS" name="Structural biomass production"
  reversible="false" fast="false">
  <annotation>
    <ram:RAM xmlns:ram="https://www.fairdomhub.org/sops
      /304">
      <ram:reaction ram:kcatForward="kcatR6" ram:
        kcatBackward="zero" ram:maintenanceScaling="zero"
        />
    </ram:RAM>
  </annotation>
  <fbc:geneProductAssociation fbc:id="Ribosome">
    <fbc:geneProductRef fbc:geneProduct="R" />
  </fbc:geneProductAssociation>
  <listOfReactants>
    <speciesReference species="AA" stoichiometry="1500"
      constant="true"/>
    <speciesReference species="ATP" stoichiometry="200"
      constant="true"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="S" stoichiometry="1"
      constant="true"/>
  </listOfProducts>
</reaction>
</listOfReactions>
</model>
</sbml>
```

## C *Synechococcus elongatus* 7942 model

Table A1: Ribosome composition of *S. elongatus* 7942.

Ribosomal protein	Gene	Stoichiometry	Molecular weight (Da)
Small ribosomal subunit			
S1	Synpcc7942_0694	1	34591
S2	Synpcc7942_2530	1	28400
S3	Synpcc7942_2226	1	27718
S4	Synpcc7942_1487	1	23206
S5	Synpcc7942_2216	1	19330
S6	Synpcc7942_0012	1	12346
S7	Synpcc7942_0886	1	17756
S8	Synpcc7942_2219	1	14682
S9	Synpcc7942_2205	1	14937
S10	Synpcc7942_0883	1	12179
S11	Synpcc7942_2210	1	13712
S12	Synpcc7942_0887	1	13991
S13	Synpcc7942_2211	1	13979
S14	Synpcc7942_0446	1	11752
S15	Synpcc7942_2299	1	10308
S16	Synpcc7942_1772	1	9555
S17	Synpcc7942_2223	1	9347
S18	Synpcc7942_1123	1	8307
S19	Synpcc7942_2228	1	10238
S20	Synpcc7942_1520	1	10892
S21	Synpcc7942_1774	1	7029
Large ribosomal subunit			
L1	Synpcc7942_0633	1	25855
L2	Synpcc7942_2229	1	31706
L3	Synpcc7942_2232	1	22452
L4	Synpcc7942_2231	1	23203
L5	Synpcc7942_2220	1	20016
L6	Synpcc7942_2218	1	19198
L7/L12	Synpcc7942_0631	2x (2-3)=4-6	13151
L9	Synpcc7942_2559	1	16679
L10	Synpcc7942_0632	1	18810
L11	Synpcc7942_0634	1	14903
L13	Synpcc7942_2206	1	17072
L14	Synpcc7942_2222	1	13318
L15	Synpcc7942_2215	1	15289
L16	Synpcc7942_2225	1	16139

Continued on next page



Table A1 – continued from previous page

Ribosomal protein	Gene	Stoichiometry	Molecular weight (Da)
L17	Synpcc7942_2208	1	13262
L18	Synpcc7942_2217	1	13047
L19	Synpcc7942_2541	1	13444
L20	Synpcc7942_1277	1	13316
L21	Synpcc7942_1219	1	13870
L22	Synpcc7942_2227	1	13253
L23	Synpcc7942_2230	1	11148
L24	Synpcc7942_2221	1	12449
L27	Synpcc7942_1220	1	9227
L28	Synpcc7942_0042	1	9119
L29	Synpcc7942_2224	1	7650
L31	Synpcc7942_2204	1	8799
L32	Synpcc7942_0997	1	6534
L33	Synpcc7942_1122	1	7372
L34	Synpcc7942_1614	1	5230
L35	Synpcc7942_1278	1	7840
L36	Synpcc7942_2212	1	4364
Ribosomal RNA			
16S	Synpcc7942_R0004	1	
5S	Synpcc7942_R0006	1	
23S	Synpcc7942_R0005	1	

Table A2: Gene composition of the photosystems I and II monomers of *S. elongatus* 7942.

Enzyme	Gene	Gene name	Stoichiometry
Photosystem I monomer	Synpcc7942_2049	PsaA	1
	Synpcc7942_2048	PsaB	1
	Synpcc7942_0535	PsaC	1
	Synpcc7942_1002	PsaD	1
	Synpcc7942_1322	PsaE	1
	Synpcc7942_1250	PsaF	1
	Synpcc7942_2343	PsaI	1
	Synpcc7942_1249	PsaJ	1
	Synpcc7942_0407	PsaK1 (PsaX)	1
	Synpcc7942_0920	(PsaK2) (PsaX)	1
	Synpcc7942_2342	PsaL	1
	Synpcc7942_1912a	PsaM	1
	C05306_cyt	Chlorophyll a	96

Continued on next page

Table A2 – continued from previous page				
Enzyme	Gene	Gene name	Stoichiometry	
	C02094_cyt	$\mu$ -carotene	22	
	C02059_cyt	Phylloquinone	2	
Photosystem II monomer	Synpcc7942_0424 (A)		1	
	Synpcc7942_0893 (A)		1	
	Synpcc7942_1389 (A)		1	
	Synpcc7942_0655 (D1)	PsbA (D1)	1	
	Synpcc7942_1637 (D2)	PsbD (D2)	1	
	Synpcc7942_0697	PsbB	1	
	Synpcc7942_0656	PsbC	1	
	Synpcc7942_1177	PsbE	1	
	Synpcc7942_1176	PsbF	1	
	Synpcc7942_0225	PsbH	1	
	Synpcc7942_1705	PsbI	1	
	Synpcc7942_1174	PsbJ	1	
	Synpcc7942_0456	PsbK	1	
	Synpcc7942_1175	PsbL	1	
	Synpcc7942_0699	PsbM	1	
	Synpcc7942_0224	PsbN	1	
	Synpcc7942_0294	PsbO	1	
	Synpcc7942_1038	PspP	1	
	Synpcc7942_0696	PsbT	1	
	Synpcc7942_1882	PsbU	1	
	Synpcc7942_2010	PsbV	1	
	Synpcc7942_2016	PsbX	1	
	Synpcc7942_1692	PsbY	1	
	Synpcc7942_2245	PsbZ	1	
	Synpcc7942_0343	Psb27	1	
	Synpcc7942_1679	Psb28 (W)	1	
	Synpcc7942_2478	Psb28-2 (W)	1	
		C05306_cyt	Chlorophyll a	35
		C02094_cyt	$\mu$ -carotene	11

Table A3: Gene composition of phycobilisomes in *S. elongatus* 7942. The following abbreviations are used:  $\alpha$  PC - Phycocyanin alpha subunit,  $\mu$  PC - Phycocyanin beta subunit,  $\alpha$  AP - Allophycocyanin alpha subunit,  $\mu$  AP - Allophycocyanin beta subunit, Lcr - core-rod linker protein, Lr - rod-rod linker protein, Lc - core-core linker protein, Lcm - core-thylakoid membrane linker protein.

Subunit	Description	Gene or metabolite	Gene name	Stoichiometry
Core		Phycocyanobilin		48
	24 · ( $\alpha$ AP + $\mu$ AP)	Synpcc7942_0327	ApcA	24
		Synpcc7942_0326	ApcB	24
		6 · Lc	Synpcc7942_0325	ApcC
	2 · Lcm	Synpcc7942_0328	ApcE	2
Rods (length 1)		Phycocyanobilin		+108
	+36 · ( $\alpha$ PC + $\mu$ PC)	Synpcc7942_1048	CpcA	+36
		Synpcc7942_1047	CpcB	+36
	+6 · Lcr	Synpcc7942_2030	CpcG	+6
Rods (length 2)		Phycocyanobilin		+108
	+36 · ( $\alpha$ PC + $\mu$ PC)	Synpcc7942_1048	CpcA	+36
		Synpcc7942_1047	CpcB	+36
	+6 · Lr	Synpcc7942_1049	CpcC	+6
Rods (length 3)		Phycocyanobilin		+108
	+36 · ( $\alpha$ PC + $\mu$ PC)	Synpcc7942_1048	CpcA	+36
		Synpcc7942_1047	CpcB	+36
	+6 · Lr	Synpcc7942_1049	CpcC	+6

Table A4: Gene composition of NDH I, Cytochrome b6f, ATPase, Cytochrome c oxidase, and succinate dehydrogenase.

Enzyme	Gene	Gene name	Stoichiometry
NADPH dehydrogenase	<b>Core (State M):</b>		
	Synpcc7942_1343	NdhA	1
	Synpcc7942_1415	NdhB	1
	Synpcc7942_1180	NdhC	1
	Synpcc7942_1346	NdhE	1
	Synpcc7942_1345	NdhG	1
	Synpcc7942_1743	NdhH	1
	Synpcc7942_1344	NdhI	1
	Synpcc7942_1182	NdhJ	1
	Synpcc7942_1181	NdhK	1
	Synpcc7942_0413	NdhL	1
	Synpcc7942_1982	NdhM	1

Continued on next page

Table A4 – continued from previous page

Enzyme	Gene	Gene name	Stoichiometry
	Synpcc7942_2234	NdhN	1
	Synpcc7942_2177	NdhO	1
	<b>Variable:</b>		
		State L:	
	Synpcc7942_1976	NdhD1	1
	Synpcc7942_1977	NdhF1	1
		State L':	
	Synpcc7942_1439	NdhD2	1
	Synpcc7942_1977	NdhF1	1
		State MS:	
	Synpcc7942_2092	NdhD3	1
	Synpcc7942_2091	NdhF3	1
	Synpcc7942_2093	CupA	1
	Synpcc7942_2094	CupS	1
		State MS':	
	Synpcc7942_0609	NdhD4	1
	Synpcc7942_0309	NdhF4	1
	Synpcc7942_0308	CupB	1
Cytochrome b6f	Synpcc7942_1232	PetC	2
	Synpcc7942_2426	PetM	2
	Synpcc7942_0978	PetH (FNR)	2
	Synpcc7942_1231	PetA	2
	Synpcc7942_2331	PetB	2
	Synpcc7942_2332	PetD	2
	Synpcc7942_0475	PetN	2
	Synpcc7942_1479	PetG	2
	Synpcc7942_0113	PetL	2
	Cytochrome:		
	Synpcc7942_0239	cyt f	1
	Synpcc7942_1630	cyt A (Pet J)	1
	Synpcc7942_2542	Cyt c6-2	1
ATPase	Synpcc7942_0331	A	1
	Synpcc7942_0332	C	10-15
	Synpcc7942_0333	B'	1-2
	Synpcc7942_0334	B	1-2
	Synpcc7942_0335	delta	1
	Synpcc7942_0336	alpha	3
	Synpcc7942_0337	gamma	1
	Synpcc7942_2315	beta	3
	Synpcc7942_2316	epsilon	1

Continued on next page

Table A4 – continued from previous page

Enzyme	Gene	Gene name	Stoichiometry
Cytochrome c oxidase	Synpcc7942_2603	sub I	1
	Synpcc7942_2602	subII	1
	Synpcc7942_2604	sub III	1
Succinate dehydrogenase	Synpcc7942_0314	SdhC	1
	Synpcc7942_0641	SdhA	1
	Synpcc7942_1533	SdhB	1

Table A5: Fractions of the dry weight of a *S. elongatus* cell that correspond to the quota metabolites as well as absolute amounts of the quota metabolites in one cell, assuming a dry weight of 1.5 pg/cell.

quota metabolite	fraction (g/gDW)	Amount (pg/cell)
Noncatalytic proteins	0.357	0.5712
DNA	0.031	0.0496
RNA	0.17	0.272
Cell wall	0.059	0.0944
Lipids	0.12	0.192
Cofactors and vitamins	0.029	0.0464
Ions	0.01	0.016
Pigments	0.0244	0.03904

## D *Saccharomyces cerevisiae* model

Table A6: Verduyn minimal medium composition and corresponding Yeast 6 transporters.

Compound	Transporters
D-glucose	r_1714
D-galactose	r_1710
oxygen	r_1992
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	r_1654, r_2060, r_2061
KH <sub>2</sub> PO <sub>4</sub>	r_2020, r_2005
MgSO <sub>4</sub> · 7H <sub>2</sub> O	r_2060, r_2061
D-biotin	r_1671
Ca-D-pantothenate	r_1548
Nicotonic acid	r_1967
Myo-inositol	r_1947
Thiamine hydrochloride	r_2067
Pyridoxal hydrochloride	r_2028
p-aminobenzoic acid	r_1604
Na <sub>2</sub>	r_2049
ZnSO <sub>4</sub> · 7H <sub>2</sub> O	r_2060, r_2061
MnCl <sub>2</sub> · 2H <sub>2</sub> O	-
CoCl <sub>2</sub> · 6H <sub>2</sub> O	-
CuSO <sub>4</sub> · 5H <sub>2</sub> O	r_2060, r_2061
Na <sub>2</sub> MoO <sub>4</sub> · 2H <sub>2</sub> O	-
CaCl <sub>2</sub> · 2H <sub>2</sub> O	-
FeSO <sub>4</sub> · 7H <sub>2</sub> O	r_1861
H <sub>3</sub> BO <sub>3</sub>	-
KI	-

Table A7: Additional medium components needed for anaerobic growth, and their corresponding Yeast 6 transporters.

Compound	Transporters
episterol	r_1753
ergosterol	r_1757
fecosterol	r_1788
lanosterol	r_1915
zymosterol	r_2106
14-demethyl lanosterol	r_2134
ergosta-5,7,22,24(28)-tetraen-3beta-ol	r_2137
oleate	r_2189

Table A8: Ribosome composition of *S. cerevisiae*. All gene stoichiometries are 1.

Ribosomal protein	Gene name	Gene
Small ribosomal subunit		
SA	RPS0A	YGR214W
	RPS0B	YLR048W
S2	RPS2	YGL123W
S3	RPS3	YNL178W
S3A	RPS1A	YLR441C
	RPS1B	YML063W
S4	RPS4A	YJR145C
	RPS4B	YHR203C
S5	RPS5	YJR123W
S6	RPS6A	YPL090C
	RPS6B	YBR181C
S7	RPS7A	YOR096W
	RPS7B	YNL096C
S8	RPS8A	YBL072C
	RPS8B	YER102W
S9	RPS9A	YPL081W
	RPS9B	YBR189W
S10	RPS10A	YOR293W
	RPS10B	YMR230W
S11	RPS11A	YDR025W
	RPS11B	YBR048W
S12	RPS12	YOR369C
S13	RPS13	YDR064W
S14	RPS14A	YCR031C
	RPS14B	YJL191W
S15	RPS15	YOL040C
S15A	RPS22A	YJL190C
	RPS22B	YLR367W
S16	RPS16A	YMR143W
	RPS16B	YDL083C
S17	RPS17A	YML024W
	RPS17B	YDR447C
S18	RPS18A	YDR450W
	RPS18B	YML026C
S19	RPS19A	YOL121C
	RPS19B	YNL302C
S20	RPS20	YHL015W
S21	RPS21A	YKR057W

Continued on next page

Table A8 – continued from previous page

Ribosomal protein	Gene name	Gene
S21	RPS21B	YJL136C
S23	RPS23A	YGR118W
	RPS23B	YPR132W
S24	RPS24A	YER074W
	RPS24B	YIL069C
S25	RPS25A	YGR027C
	RPS25B	YLR333C
S26	RPS26A	YGL189C
	RPS26B	YER131W
S27	RPS27A	YKL156W
	RPS27B	YHR021C
S27A	RPS31	YLR167W
	RPS28A	YOR167C
S28	RPS28B	YLR264W
	RPS29A	YLR388W
S29	RPS29B	YDL061C
	RPS30A	YLR287C-A
S30	RPS30B	YOR182C
	Large ribosomal subunit	
LP0	RPP0	YLR340W
LP1	RPP1A	YDL081C
	RPP1B	YDL130W
LP2	RPP2A	YOL039W
L3	RPL3	YOR063W
L4	RPL4A	YBR031W
	RPL4B	YDR012W
L5	RPL5	YPL131W
L6	RPL6A	YML073C
	RPL6B	YLR448W
L7	RPL7A	YGL076C
	RPL7B	YPL198W
L7A	RPL8A	YHL033C
	RPL8B	YLL045C
L8	RPL2A	YFR031C-A
	RPL2B	YIL018W
L9	RPL9A	YGL147C
	RPL9B	YNL067W
L10	RPL10	YLR075W
L10A	RPL1A	YPL220W
	RPL1B	YGL135W
Continued on next page		



Table A8 – continued from previous page

Ribosomal protein	Gene name	Gene
L11	RPL11A	YPR102C
	RPL11B	YGR085C
L12	RPL12A	YEL054C
	RPL12B	YDR418W
L13	RPL13A	YDL082W
	RPL13B	YMR142C
L13A	RPL16A	YIL133C
	RPL16B	YNL069C
L14	RPL14A	YKL006W
	RPL14B	YHL001W
L15	RPL15A	YLR029C
	RPL15B	YMR121C
L17	RPL17A	YKL180W
	RPL17B	YJL177W
L18	RPL18A	YOL120C
	RPL18B	YNL301C
L18A	RPL20A	YMR242C
	RPL20B	YOR312C
L19	RPL19A	YBR084C-A
	RPL19B	YBL027W
L21	RPL21A	YBR191W
	RPL21B	YPL079W
L22	RPL22A	YLR061W
	RPL22B	YFL034C-A
L23	RPL23A	YBL087C
	RPL23B	YER117W
L23A	RPL25	YOL127W
	RPL24A	YGL031C
L24	RPL24B	YGR148C
	RPL26A	YLR344W
L26	RPL26B	YGR034W
	RPL27A	YHR010W
L27	RPL27B	YDR471W
	RPL28	YGL103W
L27A	RPL28	YGL103W
L29	RPL29	YFR032C-A
L30	RPL30	YGL030W
L31	RPL31A	YDL075W
	RPL31B	YLR406C
L32	RPL32	YBL092W
L34	RPL34A	YER056C-A

Continued on next page

Table A8 – continued from previous page

Ribosomal protein	Gene name	Gene
L34	RPL34B	YIL052C
L35	RPL35A	YDL191W
	RPL35B	YDL136W
L35A	RPL33A	YPL143W
	RPL33B	YOR234C
L36	RPL36A	YMR194W
	RPL36B	YPL249C-A
L36A	RPL42A	YNL162W
	RPL42B	YHR141C
L37	RPL37A	YLR185W
	RPL37B	YDR500C
L37A	RPL43A	YPR043W
	RPL43B	YJR094W-A
L38	RPL38	YLR325C
L39	RPL39	YJL189W
L40	RPL40A	YIL148W
	RPL40B	YKR094C
L41	RPL41A	YDL184C
	RPL41B	YDL133C-A
Ribosomal RNA		
25S	RDN25-1	
5S	RDN5-1	
5.8S	RDN58-1	
18S	RDN18-1	

Table A9: Quota metabolite production reactions for *S. cerevisiae*. The reactions and energy requirements have been constructed following the information in the supplement of (Förster et al., 2003a).

Reactants	Stoichiometry	Products	Stoichiometry
Nocatalytic proteins			
Ala-tRNA(Ala)	0.9839201541	tRNA(Ala)	0.9839201541
Arg-tRNA(Arg)	0.3446294001	tRNA(Arg)	0.3446294001
Asn-tRNA(Asn)	0.2181008711	tRNA(Asn)	0.2181008711
Asp-tRNA(Asp)	0.6380040232	tRNA(Asp)	0.6380040232
Cys-tRNA(Cys)	0.0141540388	tRNA(Cys)	0.0141540388
Gln-tRNA(Gln)	0.2260357111	tRNA(Gln)	0.2260357111
Glu-tRNA(Glu)	0.6472255939	tRNA(Glu)	0.6472255939
Gly-tRNA(Gly)	0.6227777087	tRNA(Gly)	0.6227777087
His-tRNA(His)	0.1421837537	tRNA(His)	0.1421837537
Ile-tRNA(Ile)	0.4132550429	tRNA(Ile)	0.4132550429
Leu-tRNA(Leu)	0.6356450167	tRNA(Leu)	0.6356450167
Lys-tRNA(Lys)	0.6137705931	tRNA(Lys)	0.6137705931
Met-tRNA(Met)	0.1087287529	tRNA(Met)	0.1087287529
Phe-tRNA(Phe)	0.2871554242	tRNA(Phe)	0.2871554242
Pro-tRNA(Pro)	0.3532076054	tRNA(Pro)	0.3532076054
Ser-tRNA(Ser)	0.3975998181	tRNA(Ser)	0.3975998181
Thr-tRNA(Thr)	0.4104671262	tRNA(Thr)	0.4104671262
Trp-tRNA(Trp)	0.060905258	tRNA(Trp)	0.060905258
Tyr-tRNA(Tyr)	0.2187442365	tRNA(Tyr)	0.2187442365
Val-tRNA(Val)	0.5674482841	tRNA(Val)	0.5674482841

Continued on next page

Table A9 – continued from previous page

Reactants	Stoichiometry	Products	Stoichiometry
ATP	36.3823134562	ADP	36.3823134562
H <sub>2</sub> O	36.3823134562	phosphate	36.3823134562
		H <sup>+</sup>	36.3823134562
		protein quota	1
RNA			
AMP	0.691255	RNA quota	1
CMP	0.671719	ADP	24.614677
GMP	0.691255	H <sup>+</sup>	24.614677
UMP	0.900134	phosphate	24.614677
ATP	24.614677		
H <sub>2</sub> O	24.614677		
DNA			
dAMP	0.919380	DNA quota	1
dCMP	0.612920	ADP	26.559855
dGMP	0.612920	H <sup>+</sup>	26.559855
dTMP	0.919380	phosphate	26.559855
ATP	26.559855		
H <sub>2</sub> O	26.559855		
Cell wall			
(1 → 3)-beta-D-glucan	3.605775	cell wall quota	1
mannan	2.567065	ADP	16.553263
ATP	16.553263	H <sup>+</sup>	16.553263
H <sub>2</sub> O	16.553263	phosphate	16.553263

Continued on next page

Table A9 – continued from previous page

Reactants	Stoichiometry	Products	Stoichiometry
Lipids			
1-phosphatidyl-1D-myo-inositol	0.185284	lipid quota	1
14-demethyllanosterol	0.006782		
complex sphingolipid	0.050499		
episterol	0.011626		
ergosta-5,7,22,24(28)-tetraen-3beta-ol	0.015138		
ergosterol	0.678165		
ergosterol ester	0.098334		
fatty acid	0.024947		
fecosterol	0.013806		
lanosterol	0.003875		
phosphatidyl-L-serine	0.045171		
phosphatidylcholine	0.348771		
phosphatidylethanolamine	0.084407		
triglyceride	0.094580		
zymosterol	0.001817		
Small molecules			
riboflavin	0.431599	small molecules quota	1
sulphate	8.719170		

---

Table A10: Fractions of the dry weight of a *S. cerevisiae* cell that correspond to the quota metabolites.

quota metabolite	fraction (g/gDW)
noncatalytic proteins	0.27731
RNA	0.066546
DNA	0.0039157
cell wall	0.31472
lipids	0.0082576
small molecules	0.0022938

## E Regulation inference using deFBA time courses

Table A11: Transpose of the least squares regression matrix  $M_{LS}$  learned using deFBA time courses for the toy model in figure 8.1.

	Carb1 <sub>ext</sub>	Carb2 <sub>ext</sub>	D <sub>ext</sub>	E <sub>ext</sub>	F <sub>ext</sub>	H <sub>ext</sub>
ER1	0.00019381	0.00015021	0.0013658	-2.2705	0	0
ER2	0.00021147	0.00018655	0.00060077	-2.8392	0	0
ER3	-3.0439e-05	-5.0733e-05	0.00092267	-0.31313	0	0
ER4	-5.193e-05	-7.3483e-05	0.0036452	4.0171	0	0
ER5	-0.00016764	-0.00020332	0.0034676	5.6247	0	0
ER6	1.9984e-05	-3.89e-05	0.00079994	-0.238	0	0
ER7	-3.3693e-08	1.5392e-08	1.9897e-07	0.0037495	0	0
ER8	9.6483e-05	9.1794e-05	0.0004666	-1.8497	0	0
ERres	-7.6346e-05	-0.00014287	0.0067211	1.9126	0	0
ETC1	0.00021299	-1.8809e-06	0.00027752	-1.0494	0	0
ETC2	-0.00025606	0.00016953	0.0016208	-2.6385	0	0
ETH	0	0	0	0	0	0
ETF	0	0	0	0	0	0
S	-0.003285	-0.0059422	0.079068	120.01	0	0
RIB	0.00018009	0.00016234	0.0011605	-3.7149	0	0

Table A12: Transpose of the Lasso regression matrix  $M_{Lasso}$  (objective Lagrange multiplier  $\lambda = 0.115$ ) learned using deFBA time courses for the toy model in figure 8.1.

	Carb1 <sub>ext</sub>	Carb2 <sub>ext</sub>	D <sub>ext</sub>	E <sub>ext</sub>	F <sub>ext</sub>	H <sub>ext</sub>
ER1	0.00017539	0.00016227	0.0011829	-0	0	0
ER2	0.00019353	0.0001987	0.00041413	-0	0	0
ER3	-0	-0	0.00067714	0	0	0
ER4	-8.2981e-06	-2.4977e-06	0.0033939	0	0	0
ER5	-0.00012543	-0.0001326	0.003227	0	0	0
ER6	0	0	0.00059914	0	0	0
ER7	0	0	0	0	0	0
ER8	7.7699e-05	0.00010379	0.00028652	-0	0	0
ERres	-3.0924e-05	-7.156e-05	0.0064559	0	0	0
ETC1	0.00019352	9.9853e-06	0.00010272	-0	0	0
ETC2	-0.00018036	0.00015524	0.0014088	-0	0	0
ETH	0	0	0	0	0	0
ETF	0	0	0	0	0	0
S	-0.0033411	-0.005891	0.079585	0	0	0
RIB	0.00016291	0.00017464	0.00096803	-0	0	0

Table A13: Transpose of the Ridge regression matrix  $M_{Ridge}$  (objective Lagrange multiplier  $\lambda = 10^5$ ) learned using deFBA time courses for the toy model in figure 8.1.

	Carb1 <sub>ext</sub>	Carb2 <sub>ext</sub>	D <sub>ext</sub>	E <sub>ext</sub>	F <sub>ext</sub>	H <sub>ext</sub>
ER1	0.00020323	0.0001894	0.0011256	-1.4931e-08	0	0
ER2	0.00021314	0.00020067	0.00049013	-2.4703e-08	0	0
ER3	-2.004e-05	-2.0394e-05	0.0007607	1.9977e-09	0	0
ER4	-1.6188e-05	4.2613e-05	0.0030387	5.9358e-08	0	0
ER5	-0.00013163	-8.9192e-05	0.0028951	7.4231e-08	0	0
ER6	2.7094e-05	-1.2239e-05	0.00066033	2.0254e-09	0	0
ER7	-3.2894e-08	2.0771e-08	1.8546e-07	3.7875e-11	0	0
ER8	9.9724e-05	0.00010398	0.00038026	-1.5652e-08	0	0
ERres	-7.1847e-06	7.1492e-05	0.0055727	5.5565e-08	0	0
ETC1	0.0002077	9.3213e-06	0.00022628	-8.9323e-09	0	0
ETC2	-0.00022469	0.00021064	0.0013308	-1.6804e-08	0	0
ETH	0	0	0	0	0	0
ETF	0	0	0	0	0	0
S	-0.0024955	-0.0032847	0.065934	1.6113e-06	0	0
RIB	0.00018921	0.00019445	0.00094788	-3.0205e-08	0	0

Table A14: Transpose of the elastic net regression matrix  $M_{EN}$  (objective Lagrange multiplier  $\lambda = 0.2$ ) learned using deFBA time courses for the toy model in figure 8.1.

	Carb1 <sub>ext</sub>	Carb2 <sub>ext</sub>	D <sub>ext</sub>	E <sub>ext</sub>	F <sub>ext</sub>	H <sub>ext</sub>
ER1	0.00017805	0.00016078	0.0012046	-0	0	0
ER2	0.00019619	0.00019719	0.00043595	-0	0	0
ER3	-0	-0	0.00069879	0	0	0
ER4	-1.4297e-05	-1.1814e-05	0.0034295	0	0	0
ER5	-0.00013152	-0.00014183	0.0032626	0	0	0
ER6	0	0	0.00062081	0	0	0
ER7	0	0	0	0	0	0
ER8	8.0358e-05	0.00010227	0.00030835	-0	0	0
ERres	-3.699e-05	-8.0672e-05	0.0064909	0	0	0
ETC1	0.00019617	8.4662e-06	0.0001246	-0	0	0
ETC2	-0.00018991	0.00015721	0.0014339	-0	0	0
ETH	0	0	0	0	0	0
ETF	0	0	0	0	0	0
S	-0.0033462	-0.0058973	0.079606	0	0	0
RIB	0.00016557	0.00017315	0.00098974	-0	0	0



# Bibliography

- R. Aebersold and M. Mann. Mass spectrometry-based proteomics. *Nature*, 422 (6928):198–207, 2003.
- T. Aho, H. Almusá, J. Matilainen, A. Larjo, P. Ruusuvuori, K.-L. Aho, T. Wilhelm, H. Lähdesmäki, A. Beyer, M. Harju, et al. Reconstruction and validation of RefRec: a global model for the yeast molecular interaction network. *PLoS One*, 5(5):e10662, 2010.
- E. Albers, C. Larsson, G. Lidén, C. Niklasson, and L. Gustafsson. Influence of the nitrogen source on *Saccharomyces cerevisiae* anaerobic growth and product formation. *Applied and Environmental Microbiology*, 62(9):3187–3195, 1996.
- J. Almquist, M. Cvijovic, V. Hatzimanikatis, J. Nielsen, and M. Jirstrand. Kinetic models in industrial biotechnology – improving cell factory performance. *Metabolic Engineering*, 24:38–60, 2014.
- K. B. Andersen and K. von Meyenburg. Are growth rates of *Escherichia coli* in batch cultures limited by respiration? *Journal of Bacteriology*, 144(1):114–123, 1980.
- S. Angermayr, P. van Alphen, D. Hasdemir, G. Kramer, M. Iqbal, W. van Grondelle, H. Hoefsloot, Y. Choi, and K. Hellingwerf. Culturing *Synechocystis sp.* strain PCC 6803 with N<sub>2</sub> and CO<sub>2</sub> in a diel regime reveals multiphase glycogen dynamics with low maintenance costs. *Applied Environmental Microbiology*, 82(14):4180–4189, 2016.
- R. Ansell, K. Granath, S. Hohmann, J. M. Thevelein, and L. Adler. The two isoenzymes for yeast NAD<sup>+</sup>-dependent glycerol 3-phosphate dehydrogenase encoded by GPD1 and GPD2 have distinct roles in osmoadaptation and redox regulation. *The EMBO Journal*, 16(9):2179–2187, 1997.
- M. Ataman, D. F. Hernandez Gardiol, G. Fengos, and V. Hatzimanikatis. redGEM: Systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models. *PLoS Computational Biology*, 13(7):1–22, 07 2017.

## Bibliography

---

- H. W. Aung, S. A. Henry, and L. P. Walker. Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. *Industrial Biotechnology*, 9(4):215–228, 2013.
- B. M. Bakker, K. M. Overkamp, A. J. van Maris, P. Kötter, M. A. Luttik, J. P. van Dijken, and J. T. Pronk. Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*. *FEMS Microbiology Reviews*, 25(1):15–37, 2001.
- M. Bartl, P. Li, and S. Schuster. Modelling the optimal timing in metabolic pathway activation—use of pontryagin’s maximum principle and role of the golden section. *Biosystems*, 101(1):67–77, 2010.
- D. A. Beard, S. dan Liang, and H. Qian. Energy balance for analysis of complex metabolic networks. *Biophysical Journal*, 83:79–86, 2002.
- D. A. Beard, E. Babson, E. Curtis, and H. Qian. Thermodynamic constraints for biochemical networks. *Journal of Theoretical Biology*, 228(3):327–333, 2004.
- C. Beck, S. Hertel, A. Rediger, R. Lehmann, A. Wiegard, A. Kolsch, B. Heilmann, J. Georg, W. R. Hess, and I. M. Axmann. Daily expression pattern of protein-encoding genes and small noncoding RNAs in *Synechocystis sp.* strain PCC 6803. *Applied Environmental Microbiology*, 80(17):5195–5206, 2014.
- 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*, 104(31):12663–12668, 2007.
- J. Behre and S. Schuster. Modeling signal transduction in enzyme cascades with the concept of elementary flux modes. *Journal of Computational Biology*, 16(6):829–844, 2009.
- M. J. Behrenfeld, K. H. Halsey, and A. J. Milligan. Evolved physiological responses of phytoplankton to their integrated growth environment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1504):2687–2703, 2008.
- B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chemical Biology*, 5(8):593–599, 2009.
- D. A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers. Genbank. *Nucleic Acids Research*, 41:D36–D42, 2013.
- H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 2000. doi: 10.1093/nar/28.1.235.

- J. A. Bernstein, A. B. Khodursky, P.-H. Lin, S. Lin-Chao, and S. N. Cohen. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences*, 99(15):9697–9702, 2002.
- R. P. Bhalerao, T. Gillbro, and P. Gustafsson. Functional phycobilisome core structures in a phycocyanin-less mutant of cyanobacterium *Synechococcus sp.* PCC 7942. *Photosynthesis Research*, 45(1):61–70, 1995.
- E. Boles, P. de Jong-Gubbels, and J. T. Pronk. Identification and characterization of MAE1, the *Saccharomyces cerevisiae* structural gene encoding mitochondrial malic enzyme. *Journal of Bacteriology*, 180(11):2875–2882, 1998.
- A. Bordbar, J. M. Monk, Z. A. King, and B. Ø. Palsson. Constraint-based models predict metabolic and associated cellular functions. *Nature Reviews Genetics*, 15(2):107–120, 2014.
- H. Bremer and P. Dennis. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 3(1), 2008.
- G. C. Brown. Total cell protein concentration as an evolutionary constraint on the metabolic control distribution in cells. *Journal of theoretical biology*, 153(2):195–203, 1991.
- F. Büchel, N. Rodriguez, N. Swainston, C. Wrzodek, T. Czauderna, R. Keller, F. Mittag, M. Schubert, M. Glont, M. Golebiewski, M. van Iersel, S. Keating, M. Rall, M. Wybrow, H. Hermjakob, M. Hucka, D. B. Kell, W. Müller, P. Mendes, A. Zell, C. Chaouiya, J. Saez-Rodriguez, F. Schreiber, C. Laibe, A. Dräger, and N. Le Novère. Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Systems Biology*, 7(1):116, 2013.
- A. P. Burgard, P. Pharkya, and C. D. Maranas. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6):647–657, 2003.
- A. P. Burgard, E. V. Nikolaev, C. H. Schilling, and C. D. Maranas. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome research*, 14(2):301–312, 2004.
- R. L. Burnap. Systems and photosystems: cellular limits of autotrophic productivity in cyanobacteria. *Frontiers in Bioengineering and Biotechnology*, 3:1, 2015.
- D. Campbell, V. Hurry, A. K. Clarke, P. Gustafsson, and G. Öquist. Chlorophyll fluorescence analysis of cyanobacterial photosynthesis and acclimation. *Microbiology and Molecular Biology Reviews*, 62(3):667–683, 1998.

## Bibliography

---

- A. B. Canelas, W. M. van Gulik, and J. J. Heijnen. Determination of the cytosolic free NAD/NADH ratio in *Saccharomyces cerevisiae* under steady-state and highly dynamic conditions. *Biotechnology and Bioengineering*, 100(4):734–743, 2008.
- M. Celton, A. Goelzer, C. Camarasa, V. Fromion, and S. Dequin. A constraint-based model analysis of the metabolic consequences of increased NADPH oxidation in *Saccharomyces cerevisiae*. *Metabolic Engineering*, 14(4):366–379, 2012.
- V. Chelliah, N. Juty, I. Ajmera, R. Ali, M. Dumousseau, M. Glont, M. Hucka, G. Jalowicki, S. Keating, V. Knight-Schrijver, A. Lloret-Villas, K. Nath Natarajan, J.-B. Pettit, N. Rodriguez, M. Schubert, S. M. Wimalaratne, Y. Zhao, H. Hermjakob, N. Le Novère, and C. Laibe. BioModels: ten-year anniversary. *Nucleic Acids Research*, 43(D1):D542–D548, 2015.
- J. M. Cherry, E. L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley, E. T. Chan, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. R. Engel, D. G. Fisk, J. E. Hirschman, B. C. Hitz, K. Karra, C. J. Krieger, S. R. Miyasato, R. S. Nash, J. Park, M. S. Skrzypek, M. Simison, S. Weng, and E. D. Wong. *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Research*, 30(D1):D700–D705, 2012.
- C. Y. M. Cheung, M. G. Poolman, D. A. Fell, R. G. Ratcliffe, and L. J. Sweetlove. A diel flux balance model captures interactions between light and dark metabolism during day-night cycles in C3 and crassulacean acid metabolism leaves. *Plant Physiology*, 165(2):917–929, 2014.
- J. W. Cooley and W. F. Vermaas. Succinate dehydrogenase and other respiratory pathways in thylakoid membranes of *Synechocystis* sp. strain PCC 6803: capacity comparisons and physiological function. *Journal of Bacteriology*, 183(14):4251–4258, 2001.
- R. Costenoble, P. Picotti, L. Reiter, R. Stallmach, M. Heinemann, U. Sauer, and R. Aebersold. Comprehensive quantitative analysis of central carbon and amino-acid metabolism in *Saccharomyces cerevisiae* under multiple conditions by targeted proteomics. *Molecular Systems Biology*, 7(1):464, 2011.
- M. W. Covert, C. H. Schilling, and B. Ø. Palsson. Regulation of gene expression in flux balance models of metabolism. *Journal of Theoretical Biology*, 213(1):73–88, 2001.
- D. Davidi, E. Noor, W. Liebermeister, A. Bar-Even, A. Flamholz, K. Tummler, U. Barenholz, M. Goldenfeld, T. Shlomi, and R. Milo. Global characterization of *in vivo* enzyme catalytic rates and their correspondence to *in vitro* kcat measurements. *Proceedings of the National Academy of Sciences*, 113(12):3401–3406, 2016.

- L. M. De Godoy, J. V. Olsen, J. Cox, M. L. Nielsen, N. C. Hubner, F. Fröhlich, T. C. Walther, and M. Mann. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*, 455(7217):1251–1254, 2008.
- G. M. de Hijas-Liste, E. Klipp, E. Balsa-Canto, and J. R. Banga. Global dynamic optimization approach to predict activation in metabolic pathways. *BMC systems biology*, 8(1):1, 2014.
- P. Deuflhard and F. Bornemann. *Gewöhnliche Differentialgleichungen*. Walter de Gruyter, 2013.
- D. Deutscher, I. Meilijson, M. Kupiec, and E. Ruppin. Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nature Genetics*, 38(9): 993–998, 2006.
- J. Deutscher. The mechanisms of carbon catabolite repression in bacteria. *Current Opinion in Microbiology*, 11(2):87–93, 2008.
- S. Devoid, R. Overbeek, M. DeJongh, V. Vonstein, A. A. Best, and C. Henry. Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. *Systems Metabolic Engineering: Methods and Protocols*, pages 17–45, 2013.
- S. Diamond, D. Jun, B. E. Rubin, and S. S. Golden. The circadian oscillator in *Synechococcus elongatus* controls metabolite partitioning during diurnal growth. *Proceedings of the National Academy of Sciences*, 112(15):E1916–E1925, 2015.
- P. D. Dobson, K. Smallbone, D. Jameson, E. Simeonidis, K. Lanthaler, P. Pir, C. Lu, N. Swainston, W. B. Dunn, P. Fisher, et al. Further developments towards a genome-scale metabolic model of yeast. *BMC Systems Biology*, 4(1):145, 2010.
- M. M. dos Santos, V. Raghevendran, P. Kötter, L. Olsson, and J. Nielsen. Manipulation of malic enzyme in *Saccharomyces cerevisiae* for increasing NADPH production capacity aerobically in different cellular compartments. *Metabolic Engineering*, 6(4):352–363, 2004.
- N. C. Duarte, M. J. Herrgård, and B. Ø. Palsson. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research*, 14(7):1298–1309, 2004.
- J. S. Edwards, R. U. Ibarra, and B. Ø. Palsson. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology*, 19(2):125–130, 2001.
- M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000.

## Bibliography

---

- P. Erdrich, H. Knoop, R. Steuer, and S. Klamt. Cyanobacterial biofuels: new insights and strain design strategies revealed by computational modeling. *Microbial Cell Factories*, 13(1):128, 2014.
- P. Erdrich, R. Steuer, and S. Klamt. An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. *BMC Systems Biology*, 9(1):48, 2015.
- A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. Ø. Palsson. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3(1), 2007.
- D. Fell. *Understanding the control of metabolism*. Portland Press London, 1997.
- S.-M. Fendt and U. Sauer. Transcriptional regulation of respiration in yeast metabolizing differently repressive carbon substrates. *BMC Systems Biology*, 4(1):12, 2010.
- A. Finka and P. Goloubinoff. Proteomic data from human cell cultures refine mechanisms of chaperone-mediated protein homeostasis. *Cell Stress and Chaperones*, 18(5):591–605, 2013.
- E. Fischer and U. Sauer. Large-scale *in vivo* flux analysis shows rigidity and sub-optimal performance of *Bacillus subtilis* metabolism. *Nature Genetics*, 37(6):636–640, 2005.
- A. Flamholz, E. Noor, A. Bar-Even, and R. Milo. eQuilibrator—the biochemical thermodynamics calculator. *Nucleic Acids Research*, 40(D1):D770–D775, 2012.
- A. Flamholz, E. Noor, A. Bar-Even, W. Liebermeister, and R. Milo. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences*, 110(24):10039–10044, 2013.
- J. Förster, I. Famili, P. Fu, B. Ø. Palsson, and J. Nielsen. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Research*, 13(2):244–253, 2003a.
- J. Förster, I. Famili, B. Ø. Palsson, and J. Nielsen. Large-scale evaluation of *in silico* gene deletions in *Saccharomyces cerevisiae*. *OMICS A Journal of Integrative Biology*, 7(2):193–202, 2003b.
- C. Francke, R. J. Siezen, and B. Teusink. Reconstructing the metabolic network of a bacterium from its genome. *Trends in Microbiology*, 13(11):550–558, 2005.
- J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1 of *Springer Series in Statistics*. Springer, Berlin, 2001.

- A. M. Gleixner, D. E. Steffy, and K. Wolter. Improving the Accuracy of Linear Programming Solvers with Iterative Refinement. In *Proceedings of the 37th International Symposium on Symbolic and Algebraic Computation, ISSAC '12*, pages 187–194, 2012.
- A. M. Gleixner, D. E. Steffy, and K. Wolter. Iterative refinement for linear programming. *INFORMS Journal on Computing*, 28(3):449–464, 2016.
- A. Goelzer and V. Fromion. Resource allocation in living organisms. *Biochemical Society Transactions*, page BST20160436, 2017.
- A. Goelzer, V. Fromion, and G. Scorletti. Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6):1210–1218, 2011.
- A. Goelzer, J. Muntel, V. Chubukov, M. Jules, E. Prestel, R. Nölker, M. Mariadassou, S. Aymerich, M. Hecker, P. Noirot, et al. Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic Engineering*, 32:232–243, 2015.
- A. Goldbeter. *Biochemical Oscillations and Cellular Rhythms*. Cambridge University Press, 1997.
- W. Gottstein, S. Müller, H. Herzel, and R. Steuer. Elucidating the adaptation and temporal coordination of metabolic pathways using *in silico* evolution. *Biosystems*, 117:68–76, 2014.
- W. Gottstein, B. G. Olivier, F. J. Bruggeman, and B. Teusink. Constraint-based stoichiometric modelling from single organisms to microbial communities. *Journal of The Royal Society Interface*, 13(124):20160627, 2016.
- E. Gruchattka, O. Hädicke, S. Klamt, V. Schütz, and O. Kayser. *In silico* profiling of *Escherichia coli* and *Saccharomyces cerevisiae* as terpenoid factories. *Microbial Cell Factories*, 12(1):84, 2013.
- A. C. Guerreiro, M. Benevento, R. Lehmann, B. van Breukelen, H. Post, P. Gianisanti, A. M. Altelaar, I. M. Axmann, and A. J. Heck. Daily rhythms in the cyanobacterium *Synechococcus elongatus* probed by high-resolution mass spectrometry-based proteomics reveals a small defined set of cyclic proteins. *Molecular & Cellular Proteomics*, 13(8):2042–2055, 2014.
- O. Hädicke and S. Klamt. Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metabolic Engineering*, 13(2):204–213, 2011.
- P. Hamel, Y. Saint-Georges, B. De Pinto, N. Lachacinski, N. Altamura, and G. Dujardin. Redundancy in the function of mitochondrial phosphate transport in *Saccharomyces cerevisiae* and *Arabidopsis thaliana*. *Molecular Microbiology*, 51(2):307–317, 2004.

## Bibliography

---

- J. J. Hamilton, V. Dwivedi, and J. L. Reed. Quantitative assessment of thermodynamic constraints on the solution space of genome-scale metabolic models. *Biophysical journal*, 105(2):512–522, 2013.
- W. R. Harcombe, N. F. Delaney, N. Leiby, N. Klitgord, and C. J. Marx. The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. *PLoS Computational Biology*, 9(6):e1003091, 2013.
- B. D. Heavner and N. D. Price. Comparative analysis of yeast metabolic network models highlights progress, opportunities for metabolic reconstruction. *PLoS Computational Biology*, 11(11):e1004530, 2015.
- B. D. Heavner, K. Smallbone, B. Barker, P. Mendes, and L. P. Walker. Yeast 5—an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network. *BMC Systems Biology*, 6(1):55, 2012.
- B. D. Heavner, K. Smallbone, N. D. Price, and L. P. Walker. Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database*, 2013:bat059, 2013.
- R. Heinrich and S. Schuster. *The regulation of cellular systems*. Chapman and Hall, 1996.
- R. Heinrich and S. Schuster. The modelling of metabolic systems. Structure, control and optimality. *Biosystems*, 47(1):61–77, 1998.
- S. Henikoff and J. G. Henikoff. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences*, 89(22):10915–10919, 1992.
- M. A. Henson. Genome-scale modelling of microbial metabolism with temporal and spatial resolution. *Biochemical Society Transactions*, 43(6):1164–1171, 2015.
- M. J. Herrgård, B.-S. Lee, V. Portnoy, and B. Ø. Palsson. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Research*, 16(5):627–635, 2006.
- M. J. Herrgård, N. Swainston, P. Dobson, W. B. Dunn, K. Y. Arga, M. Arvas, N. Blüthgen, S. Borger, R. Costenoble, M. Heinemann, M. Hucka, N. Le Novère, P. Li, W. Liebermeister, M. L. Mo, A. P. Oliveira, D. Petranovic, S. Pettifer, E. Simeonidis, K. Smallbone, I. Spasić, D. Weichart, R. Brent, D. S. Broomhead, H. V. Westerhoff, B. Kürdar, M. Penttilä, E. Klipp, B. Ø. Palsson, U. Sauer, S. G. Oliver, P. Mendes, J. Nielsen, and D. B. Kell. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature Biotechnology*, 26(10):1155–1160, 2008.



- J. L. Hjersted and M. A. Henson. Optimization of fed-batch *Saccharomyces cerevisiae* fermentation using dynamic flux balance models. *Biotechnology Progress*, 22(5):1239–1248, 2006.
- H.-G. Holzhütter. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *European Journal of Biochemistry*, 271(14):2905–2922, 2004.
- C. A. Howitt and W. F. Vermaas. Quinol and cytochrome oxidases in the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochemistry*, 37(51):17944–17951, 1998.
- S. Hui, J. M. Silverman, S. S. Chen, D. W. Erickson, M. Basan, J. Wang, T. Hwa, and J. R. Williamson. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Molecular Systems Biology*, 11(2), 2015.
- R. U. Ibarra, J. S. Edwards, and B. Ø. Palsson. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature*, 420(6912):186–189, 2002.
- I. Ishtar Snoek and H. Yde Steensma. Factors involved in anaerobic growth of *Saccharomyces cerevisiae*. *Yeast*, 24(1):1–10, 2007.
- H. Ito, M. Mutsuda, Y. Murayama, J. Tomita, N. Hosokawa, K. Terauchi, C. Sugita, M. Sugita, T. Kondo, and H. Iwasaki. Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in *Synechococcus elongatus*. *Proceedings of the National Academy of Sciences*, 106(33):14168–14173, 2009.
- R. Jain and R. Srivastava. Metabolic investigation of host/pathogen interaction using MS2-infected *Escherichia coli*. *BMC Systems Biology*, 3(1):121, 2009.
- S. Jamshidi. *Comparing discrete, continuous and hybrid modelling approaches of gene regulatory networks*. PhD thesis, Freie Universität Berlin, 2012.
- J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193, 1906.
- L. Jerby and E. Ruppin. Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. *Clinical Cancer Research*, 18(20):5572–5584, 2012.
- C. Jungreuthmayer and J. Zanghellini. Designing optimal cell factories: integer programming couples elementary mode analysis with regulation. *BMC Systems Biology*, 6(1):103, 2012.
- N. Juty, R. Ali, M. Glont, S. Keating, N. Rodriguez, M. J. Swat, S. M. Wimalaratne, H. Hermjakob, N. Le Novère, C. Laibe, and V. Chelliah. BioModels: Content,

## Bibliography

---

- Features, Functionality and Use. *CPT: Pharmacometrics & Systems Pharmacology*, 2015.
- M. Kanehisa and S. Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.
- P. D. Karp, M. Latendresse, S. M. Paley, M. Krummenacker, Q. D. Ong, R. Billington, A. Kothari, D. Weaver, T. Lee, P. Subhraveti, et al. Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics*, 17(5):877–890, 2016.
- T. V. Karpinets, D. J. Greenwood, C. E. Sams, and J. T. Ammons. RNA: protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis. *BMC Biology*, 4(1):1, 2006.
- S. M. Kelk, B. G. Olivier, L. Stougie, and F. J. Bruggeman. Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. *Scientific Reports*, 2:580, 2012.
- H. Khalil. *Nonlinear Systems*. Prentice Hall, 2002.
- F. Kiefer, K. Arnold, M. Künzli, L. Bordoli, and T. Schwede. The SWISS-MODEL Repository and associated resources. *Nucleic acids research*, 37(suppl\_1): D387–D392, 2008.
- J. Kim, J. L. Reed, and C. T. Maravelias. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. *PLoS One*, 6(9):e24162, 2011.
- Z. A. King, J. Lu, A. Dräger, P. Miller, S. Federowicz, J. A. Lerman, A. Ebrahim, B. Ø. Palsson, and N. E. Lewis. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, 44(D1): D515–D522, 2016.
- H. Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, 2002.
- S. Klamt and R. Mahadevan. On the feasibility of growth-coupled product synthesis in microbial strains. *Metabolic Engineering*, 30:166–178, 2015.
- M. Klingenberg. The ADP-ATP translocation in mitochondria, a membrane potential controlled transport. *Journal of Membrane Biology*, 56(2):97–105, 1980.
- E. Klipp. Timing matters. *FEBS letters*, 583(24):4013–4018, 2009. Example 3.
- E. Klipp and R. Heinrich. Competition for enzymes in metabolic pathways: Implications for optimal distributions of enzyme concentrations and for the distribution of flux control. *Biosystems*, 54(1):1–14, 1999.

- E. Klipp, R. Heinrich, and H.-G. Holzhütter. Prediction of temporal gene expression. *European journal of biochemistry*, 269(22):5406–5413, 2002.
- E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach. *Systems biology in practice: concepts, implementation and application*. John Wiley & Sons, 2008.
- E. Klotz. Identification, assessment, and correction of ill-conditioning and numerical instability in linear and integer programs. In *Bridging Data and Decisions*, pages 54–108. INFORMS, 2014.
- S. Klumpp, Z. Zhang, and T. Hwa. Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139(7):1366–1375, 2009.
- D. Knies, P. Wittmüß, S. Appel, O. Sawodny, M. Ederer, and R. Feuer. Modeling and simulation of optimal resource management during the diurnal cycle in *Emiliana huxleyi* by genome-scale reconstruction and an extended flux balance analysis approach. *Metabolites*, 5(4):659–676, 2015.
- B. Knoke, M. Marhl, M. Perc, and S. Schuster. Equality of average and steady-state levels in some nonlinear models of biological oscillations. *Theory in Biosciences*, 127(1):1–14, 2008.
- B. Knoke, C. Bodenstern, M. Marhl, M. Perc, and S. Schuster. Jensen’s inequality as a tool for explaining the effect of oscillations on the average cytosolic calcium concentration. *Theory in Biosciences*, 129(1):25–38, 2010.
- H. Knoop, M. Gründel, Y. Zilliges, R. Lehmann, S. Hoffmann, W. Lockau, and R. Steuer. Flux balance analysis of cyanobacterial metabolism: the metabolic network of *Synechocystis sp.* PCC 6803. *PLoS Computational Biology*, 9(6):e1003081, 2013.
- K. Kochanowski, B. Volkmer, L. Gerosa, B. R. H. van Rijsewijk, A. Schmidt, and M. Heinemann. Functioning of a metabolic flux sensor in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 110(3):1130–1135, 2013.
- C. J. Krieger, P. Zhang, L. A. Mueller, A. Wang, S. Paley, M. Arnaud, J. Pick, S. Y. Rhee, and P. D. Karp. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Research*, 32(suppl 1):D438–D442, 2004.
- D. Kuan, S. Duff, D. Posarac, and X. Bi. Growth optimization of *Synechococcus elongatus* PCC 7942 in lab flasks and a 2-D photobioreactor. *The Canadian Journal of Chemical Engineering*, 93(4):640–647, 2015.
- H. Kubitschek. Cell volume increase in *Escherichia coli* after shifts to richer media. *Journal of Bacteriology*, 172(1):94–101, 1990.
- L. Kuepfer, U. Sauer, and L. M. Blank. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Research*, 15(10):1421–1430, 2005.

## Bibliography

---

- Y. Kumei, T. Nakajima, A. Sato, N. Kamata, and S. Enomoto. Reduction of G1 phase duration and enhancement of c-myc gene expression in HeLa cells at hypergravity. *Journal of Cell Science*, 93(2):221–226, 1989.
- G. Lambert, J. Chew, and M. J. Rust. Costs of clock-environment misalignment in individual cyanobacterial cells. *Biophysical Journal*, 111:883–891, 2016.
- N. Le Novère, B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, H. Dharuri, L. Li, H. Sauro, M. Schilstra, B. Shapiro, J. L. Snoep, and M. Hucka. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Research*, 34: D689–D691, 2006.
- R. Lehmann, R. Machné, J. Georg, M. Benary, I. M. Axmann, and R. Steuer. How cyanobacteria pose new problems to old methods: challenges in microarray time series analysis. *BMC Bioinformatics*, 14(1):133, 2013.
- J. A. Lerman, D. R. Hyduke, H. Latif, V. A. Portnoy, et al. *In silico* method for modelling metabolism and gene product expression at genome scale. *Nature Communications*, 3, 2012.
- C. Li, M. Donizelli, N. Rodriguez, H. Dharuri, L. Endler, V. Chelliah, L. Li, E. He, A. Henry, M. I. Stefan, J. L. Snoep, M. Hucka, N. Le Novère, and C. Laibe. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology*, 4:92, 2010.
- G.-W. Li, D. Burkhardt, C. Gross, and J. S. Weissman. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3):624–635, 2014.
- W. Liebermeister, E. Noor, A. Flamholz, D. Davidi, J. Bernhardt, and R. Milo. Visual account of protein investment in cellular functions. *Proceedings of the National Academy of Sciences*, 111(23):8488–8493, 2014.
- H. Lindhorst, S. Lucia, R. Findeisen, and S. Waldherr. Modeling metabolic networks including gene expression and uncertainties. *arXiv preprint arXiv:1609.08961*, 2016.
- H. Liu, H. Zhang, D. M. Niedzwiedzki, M. Prado, G. He, M. L. Gross, and R. E. Blankenship. Phycobilisomes supply excitations to both photosystems in a megacomplex in cyanobacteria. *Science*, 342(6162):1104–1107, 2013.
- M. Loferer-Krößbacher, J. Klima, and R. Psenner. Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis. *Applied and Environmental Microbiology*, 64(2):688–694, 1998.

- A. M. Luciani, A. Rosi, P. Matarrese, G. Arancia, L. Guidoni, and V. Viti. Changes in cell volume and internal sodium concentration in HeLa cells during exponential growth and following lonidamine treatment. *European Journal of Cell Biology*, 80(2):187–195, 2001.
- W. Ma and T. Ogawa. Oxygenic photosynthesis-specific subunits of cyanobacterial NADPH dehydrogenases. *IUBMB Life*, 67(1):3–8, 2015.
- T. D. B. Mackenzie, R. A. Burns, and D. A. Campbell. Carbon status constrains light acclimation in the cyanobacterium *Synechococcus elongatus*. *Plant Physiology*, 136(2):3301–3312, 2004.
- R. MacLean. Pleiotropy and GAL pathway degeneration in yeast. *Journal of Evolutionary Biology*, 20(4):1333–1338, 2007.
- B. Magasanik. Catabolite repression. In *Cold Spring Harbor symposia on quantitative biology*, volume 26, pages 249–256. Cold Spring Harbor Laboratory Press, 1961.
- R. Mahadevan and C. Schilling. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*, 5(4):264–276, 2003.
- J. S. Markson, J. R. Piechura, A. M. Puszynska, and E. K. O’Shea. Circadian control of global gene expression by the cyanobacterial master regulator RpaA. *Cell*, 155(6):1396–1408, 2013.
- C. M. T. Marobbio, M. A. Di Noia, and F. Palmieri. Identification of a mitochondrial transporter for pyrimidine nucleotides in *Saccharomyces cerevisiae*: bacterial expression, reconstitution and functional characterization. *Biochemical Journal*, 393(2):441–446, 2006.
- F. Matsuda, C. Furusawa, T. Kondo, J. Ishii, H. Shimizu, and A. Kondo. Engineering strategy of yeast metabolism for higher alcohol production. *Microbial Cell Factories*, 10(1):70, 2011.
- L. Michaelis and M. L. Menten. Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift*, 49(333-369):352, 1913.
- R. Milo and R. Phillips. *Cell biology by the numbers*. Garland Science, 2015.
- J. Mitchison. The growth of single cells: II. *Saccharomyces cerevisiae*. *Experimental Cell Research*, 15(1):214–221, 1958.
- H. M. Miziorko. Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Archives of Biochemistry and Biophysics*, 505(2):131–143, 2011.

## Bibliography

---

- M. L. Mo, B. Ø. Palsson, and M. J. Herrgård. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Systems Biology*, 3(1):37, 2009.
- M. Mojena, L. Bosca, and L. Hue. Effect of glutamine on fructose 2, 6-bisphosphate and on glucose metabolism in HeLa cells and in chick-embryo fibroblasts. *Biochemical Journal*, 232:521–527, 1985.
- D. Molenaar, R. van Berlo, D. de Ridder, and B. Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular Systems Biology*, 5(1), 2009.
- M. Moreira dos Santos, M. Åkesson, and J. Nielsen. Metabolic flux analysis. In B. Kholodenko and H. Westerhoff, editors, *Metabolic Engineering in the Post Genomic Era*. Horizon Bioscience, 2004.
- R. D. Morin, M. Bainbridge, A. Fejes, M. Hirst, M. Krzywinski, T. J. Pugh, H. McDonald, R. Varhol, S. J. Jones, and M. A. Marra. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques*, 45(1):81, 2008.
- A. C. Müller and A. Bockmayr. Flux modules in metabolic networks. *Journal of Mathematical Biology*, 69(5):1151–1179, 2014.
- A. Nakao, M. Yoshihama, and N. Kenmochi. RPG: the ribosomal protein gene database. *Nucleic Acids Research*, 32(suppl. 1):D168–D170, 2004.
- M. Nakao, S. Okamoto, M. Kohara, T. Fujishiro, T. Fujisawa, S. Sato, S. Tabata, T. Kaneko, and Y. Nakamura. Cyanobase: the cyanobacteria genome database update 2010. *Nucleic Acids Research*, 38(suppl 1):D379–D381, 2010.
- S. B. Needleman and C. D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48(3):443–453, 1970.
- D. L. Nelson, A. L. Lehninger, and M. M. Cox. *Lehninger Principles of Biochemistry*. Macmillan, 2008.
- A. M. New, B. Cerulus, S. K. Govers, G. Perez-Samper, B. Zhu, S. Boogmans, J. B. Xavier, and K. J. Verstrepen. Different levels of catabolite repression optimize growth in stable and variable environments. *PLoS Biology*, 12(1):e1001764, 2014.
- P. C. Ng and E. F. Kirkness. Whole genome sequencing. In M. R. Barnes and G. Breen, editors, *Genetic Variation*, pages 215–226. Springer, 2010.
- W. H. Nitschmann and G. A. Peschek. Oxidative phosphorylation and energy buffering in cyanobacteria. *Journal of Bacteriology*, 168(3):1205–1211, 1986.

- J. Nogales, S. Gudmundsson, E. M. Knight, B. Ø. Palsson, and I. Thiele. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proceedings of the National Academy of Sciences*, 109(7):2678–2683, 2012.
- I. Nookaew, M. C. Jewett, A. Meechai, C. Thammarongtham, K. Laoteng, S. Cheevadhanarak, J. Nielsen, and S. Bhumiratana. The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Systems Biology*, 2(1):71, 2008.
- R. A. Notebaart, F. H. Van Enkevort, C. Francke, R. J. Siezen, and B. Teusink. Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics*, 7(1):296, 2006.
- E. J. O’Brien, J. A. Lerman, R. L. Chang, D. R. Hyduke, and B. Ø. Palsson. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, 9(1):693, 2013.
- B. G. Olivier and F. T. Bergmann. Flux balance constraints, version 2 release 1. Available from COMBINE. <http://identifiers.org/combine.specifications/sbml.level-3.version-1.fbc.version-2.release-1>, 2015.
- J. D. Orth, I. Thiele, and B. Ø. Palsson. What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248, 2010.
- T. Österlund, I. Nookaew, S. Bordel, and J. Nielsen. Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Systems Biology*, 7(1):36, 2013.
- D. A. Oyarzún, B. P. Ingalls, R. H. Middleton, and D. Kalamatianos. Sequential activation of metabolic pathways: a dynamic optimization approach. *Bulletin of mathematical biology*, 71(8):1851–1872, 2009.
- A.-K. Pålman, K. Granath, R. Ansell, S. Hohmann, and L. Adler. The yeast glycerol 3-phosphatases Gpp1p and Gpp2p are required for glycerol biosynthesis and differentially involved in the cellular responses to osmotic, anaerobic, and oxidative stress. *Journal of Biological Chemistry*, 276(5):3555–3563, 2001.
- L. Palmieri, G. Agrimi, M. J. Runswick, I. M. Fearnley, F. Palmieri, and J. E. Walker. Identification in *Saccharomyces cerevisiae* of two isoforms of a novel mitochondrial transporter for 2-oxoadipate and 2-oxoglutarate. *Journal of Biological Chemistry*, 276(3):1916–1922, 2001.
- B. Ø. Palsson. *Systems biology: constraint-based reconstruction and analysis*. Cambridge University Press, 2015.

## Bibliography

---

- A. Papagiannakis, B. Niebel, E. C. Wit, and M. Heinemann. Autonomous metabolic oscillations robustly gate the early and late cell cycle. *Molecular Cell*, 65(2):285–295, 2017.
- G. K. Pattanayak, C. Phong, and M. J. Rust. Rhythms in energy storage control the ability of the cyanobacterial circadian clock to reset. *Current Biology*, 24(16):1934–1938, 2014.
- R. Pereira, J. Nielsen, and I. Rocha. Improving the flux distributions simulated with genome-scale metabolic models of *Saccharomyces cerevisiae*. *Metabolic Engineering Communications*, 3:153–163, 2016.
- P. Portela, S. Moreno, and S. Rossi. Characterization of yeast pyruvate kinase 1 as a protein kinase a substrate, and specificity of the phosphorylation site sequence in the whole protein. *Biochemical Journal*, 396(1):117–126, 2006.
- J. Postmus, I. Tuzun, M. Bekker, W. H. Müller, M. J. T. de Mattos, S. Brul, and G. J. Smits. Dynamic regulation of mitochondrial respiratory chain efficiency in *Saccharomyces cerevisiae*. *Microbiology*, 157(12):3500–3511, 2011.
- M. N. Price, K. M. Wetmore, A. M. Deutschbauer, and A. P. Arkin. A comparison of the costs and benefits of bacterial gene expression. *PLoS One*, 11(10):e0164314, 2016.
- N. D. Price, J. L. Reed, and B. Ø. Palsson. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology*, 2(11):886–897, 2004.
- A. C. Reimers. *Metabolic Networks, Thermodynamic Constraints, and Matroid Theory*. PhD thesis, Freie Universität Berlin, 2014.
- A. C. Reimers, F. J. Bruggeman, B. G. Olivier, and L. Stougie. Fast flux module detection using matroid theory. *Journal of Computational Biology*, 22(5):414–424, 2015.
- A.-M. Reimers and A. C. Reimers. The steady-state assumption in oscillating and growing systems. *Journal of Theoretical Biology*, 406:176–186, 2016.
- A.-M. Reimers, H. Knoop, A. Bockmayr, and R. Steuer. Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth. *Proceedings of the National Academy of Sciences*, 114(31):E6457–E6465, 2017a.
- A.-M. Reimers, H. Lindhorst, and S. Waldherr. A protocol for generating and exchanging (genome-scale) metabolic resource allocation models. *Metabolites*, 7(3):47, 2017b.
- A. Röhl and A. Bockmayr. A mixed-integer linear programming approach to the reduction of genome-scale metabolic networks. *BMC Bioinformatics*, 18(1):2, 2017.



- W. Rudin. *Principles of Mathematical Analysis (International Series in Pure & Applied Mathematics)*. McGraw-Hill Publishing Co., 1976.
- M. Rügen, A. Bockmayr, and R. Steuer. Elucidating temporal resource allocation and diurnal dynamics in phototrophic metabolism using conditional FBA. *Scientific Reports*, 5:15247, 2015.
- C. Runge. Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten. *Zeitschrift für Mathematik und Physik*, 46(224-243):20, 1901.
- R. Saha, D. Liu, A. Hoynes-O'Connor, M. Liberton, J. Yu, M. Bhattacharyya-Pakrasi, A. Balassy, F. Zhang, T. S. Moon, C. D. Maranas, et al. Diurnal regulation of cellular processes in the cyanobacterium *Synechocystis sp.* strain PCC 6803: Insights from transcriptomic, fluxomic, and physiological analyses. *mBio*, 7(3):e00464–16, 2016.
- C. Schaab, T. Geiger, G. Stoehr, J. Cox, and M. Mann. Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Molecular & Cellular Proteomics*, 11(3):M111–014068, 2012.
- J. Schellenberger, N. E. Lewis, and B. Ø. Palsson. Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophysical Journal*, 100(3):544–553, 2011.
- C. H. Schilling, S. Schuster, B. Ø. Palsson, and R. Heinrich. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnology Progress*, 15(3):296–303, 1999.
- I. Schomburg, A. Chang, S. Placzek, C. Söhngen, M. Rother, M. Lang, C. Munaretto, S. Ulas, M. Stelzer, A. Grote, et al. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Research*, 41(D1):D764–D772, 2013.
- R. Schuetz, L. Kuepfer, and U. Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*, 3(1):119, 2007.
- R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer. Multidimensional optimality of microbial metabolism. *Science*, 336(6081):601–604, 2012.
- S. Schuster and D. Fell. *Bioinformatics: from genomes to therapies*, volume 2. Wiley Online Library, 2007.
- S. Schuster and C. Hilgetag. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 2(02):165–182, 1994.

## Bibliography

---

- S. Schuster, T. Pfeiffer, and D. A. Fell. Is maximization of molar yield in metabolic networks favoured by evolution? *Journal of Theoretical Biology*, 252(3):497–504, 2008.
- M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330(6007):1099–1102, 2010.
- M. Scott, S. Klumpp, E. M. Mateescu, and T. Hwa. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Molecular Systems Biology*, 10:747, 2014.
- T. Shlomi, T. Benyamini, E. Gottlieb, R. Sharan, and E. Ruppin. Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the Warburg effect. *PLoS Computational Biology*, 7(3):e1002018, 2011.
- R. K. Shultzaberger, J. S. Boyd, S. Diamond, R. J. Greenspan, and S. S. Golden. Giving time purpose: The *Synechococcus elongatus* clock in a broader network context. *Annual Review of Genetics*, 49(1):485–505, 2015.
- M. L. Siegal. Shifting sugars and shifting paradigms. *PLoS Biology*, 13(2):e1002068, 2015.
- M. E. Skinner, A. V. Uzilov, L. D. Stein, C. J. Mungall, and I. H. Holmes. Jbrowse: a next-generation genome browser. *Genome research*, 19(9):1630–1638, 2009.
- A. L. Sonenshein. Control of key metabolic intersections in *Bacillus subtilis*. *Nature Reviews Microbiology*, 5(12):917–927, 2007.
- H.-S. Song and D. Ramkrishna. When is the quasi-steady-state approximation admissible in metabolic modeling? When admissible, what models are desirable? *Industrial & Engineering Chemistry Research*, 48(17):7976–7985, 2009.
- S. W. Sowa, M. Baldea, and L. M. Contreras. Optimizing metabolite production using periodic oscillations. *PLoS Computational Biology*, 10(6):e1003658, 2014.
- G. Stephanopoulos, A. A. Aristidou, and J. Nielsen. *Metabolic engineering: principles and methodologies*. Academic press, 1998.
- R. Steuer and B. H. Junker. Computational models of metabolism: stability and regulation in metabolic networks. *Advances in Chemical Physics*, 142:105, 2009.
- B. J. Stewart, A. Navid, K. W. Turteltaub, and G. Bench. Yeast dynamic metabolic flux measurement in nutrient-rich media by Hplc and accelerator mass spectrometry. *Analytical Chemistry*, 82(23):9812–9817, 2010.

- N. Swainston, K. Smallbone, P. Mendes, D. B. Kell, and N. W. Paton. The subliminal toolbox: automating steps in the reconstruction of metabolic networks. *Journal of Integrative Bioinformatics (JIB)*, 8(2):187–203, 2011.
- B. Szappanos, K. Kovács, B. Szamecz, F. Honti, M. Costanzo, A. Baryshnikova, G. Gelius-Dietrich, M. J. Lercher, M. Jelasity, C. L. Myers, et al. An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nature Genetics*, 43(7):656–662, 2011.
- H. B. Teicher and H. V. Scheller. The NAD(P)H dehydrogenase in barley thylakoids is photoactivatable and uses NADPH as well as NADH. *Plant Physiology*, 117(2):525–532, 1998.
- M. C. Teixeira, P. T. Monteiro, J. F. Guerreiro, J. P. Goncalves, N. P. Mira, S. C. dos Santos, T. R. Cabrito, M. Palma, C. Costa, A. P. Francisco, et al. The YEAS-TRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 42(D1):D161–D166, 2014.
- I. Thiele and B. Ø. Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5(1):93–121, 2010.
- I. Thiele, N. Swainston, R. M. Fleming, A. Hoppe, S. Sahoo, M. K. Aurich, H. Haraldsdottir, M. L. Mo, O. Rolfsson, M. D. Stobbe, et al. A community-driven global reconstruction of human metabolism. *Nature Biotechnology*, 31(5):419–425, 2013.
- B. A. Thomas, T. M. Bricker, and A. V. Klotz. Post-translational methylation of phycobilisomes and oxygen evolution efficiency in cyanobacteria. *Biochimica et Biophysica Acta - Bioenergetics*, 1143(1):104–108, 1993.
- C. T. Trinh. Elucidating and reprogramming *Escherichia coli* metabolisms for obligate anaerobic n-butanol and isobutanol production. *Applied Microbiology and Biotechnology*, 95(4):1083–1094, 2012.
- B. Trumpower. The protonmotive Q cycle. Energy transduction by coupling of proton translocation to electron transfer by the cytochrome bc<sub>1</sub> complex. *Journal of Biological Chemistry*, 265(20):11409–11412, 1990.
- C. B. Tyson, P. G. Lord, and A. E. Wheals. Dependency of size of *Saccharomyces cerevisiae* cells on growth rate. *Journal of Bacteriology*, 138(1):92–98, 1979.
- UniProt Consortium. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Research*, 42(D1):D191–D198, 2014.
- UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 45(D1):D158–D169, 2017.

## Bibliography

---

- J. P. van Dijken, R. A. Weusthuis, and J. T. Pronk. Kinetics of growth and sugar consumption in yeasts. *Antonie Van Leeuwenhoek*, 63(3-4):343–352, 1993.
- J. H. van Heerden, M. T. Wortel, F. J. Bruggeman, J. J. Heijnen, Y. J. Bollen, R. Planqué, J. Hulshof, T. G. O’Toole, S. A. Wahl, and B. Teusink. Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science*, 343(6174):1245114, 2014.
- A. Varma and B. Ø. Palsson. Metabolic flux balancing: basic concepts, scientific and practical use. *Nature Biotechnology*, 12(10):994–998, 1994.
- N. Vázquez-Laslop and A. S. Mankin. Protein accounting in the cellular economy. *Cell*, 157(3):529–531, 2014.
- I. Velázquez and J. P. Pardo. Kinetic characterization of the rotenone-insensitive internal NADH: ubiquinone oxidoreductase of mitochondria from *Saccharomyces cerevisiae*. *Archives of Biochemistry and Biophysics*, 389(1):7–14, 2001.
- W. F. Vermaas. Photosynthesis and respiration in cyanobacteria. *Encyclopedia of Life Sciences*, pages 1–7, 2001.
- A. von Kamp and S. Klamt. Enumeration of smallest intervention strategies in genome-scale metabolic networks. *PLoS Computational Biology*, 10(1):e1003378, 2014.
- K. Voss, M. Heiner, and I. Koch. Steady state analysis of metabolic pathways using petri nets. *In silico biology*, 3(3):367–387, 2003.
- A. Vozza, E. Blanco, L. Palmieri, and F. Palmieri. Identification of the mitochondrial GTP/GDP transporter in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 279(20):20850–20857, 2004.
- J. Waldbauer, S. Rodrigue, M. Coleman, and S. Chisholm. Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PLoS One*, 7(8):e43432, 2012.
- S. Waldherr and H. Lindhorst. Optimality in cellular storage via the pontryagin maximum principle. *arXiv preprint arXiv:1704.02189*, 2017.
- S. Waldherr, D. A. Oyarzún, and A. Bockmayr. Dynamic optimization of metabolic networks coupled with gene expression. *Journal of Theoretical Biology*, 365:469–485, 2015.
- C. Waldron and F. Lacroute. Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. *Journal of Bacteriology*, 122(3):855–865, 1975.
- O. Warburg. On the origin of cancer cells. *Science*, 123(3191):309–314, 1956.

- A. Y. Weiße, D. A. Oyarzún, V. Danos, and P. S. Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9):E1038–E1047, 2015.
- S. Westermark and R. Steuer. Toward multiscale models of cyanobacterial growth: A modular approach. *Frontiers in Bioengineering and Biotechnology*, 4:95, 2016.
- U. Wittig, R. Kania, M. Golebiewski, M. Rey, L. Shi, L. Jong, E. Alga, A. Weidemann, H. Sauer-Danzwith, S. Mir, et al. SABIO-RK – database for biochemical reaction kinetics. *Nucleic Acids Research*, 40(D1):D790–D796, 2012.
- B. Wu, B. Zhang, X. Feng, J. R. Rubens, R. Huang, L. M. Hicks, H. B. Pakrasi, and Y. J. Tang. Alternative isoleucine synthesis pathway in cyanobacterial species. *Microbiology*, 156:596–602, 2010.
- R. Wunderling. *Paralleler und objektorientierter Simplex-Algorithmus*. PhD thesis, Technische Universität Berlin, 1996. <http://www.zib.de/Publications/abstracts/TR-96-09/>.
- L. Yang, J. T. Yurkovich, C. J. Lloyd, A. Ebrahim, M. A. Saunders, and B. Ø. Pals-son. Principles of proteome allocation are revealed using proteomic data and genome-scale models. *Scientific Reports*, 6:36734, 2016.
- L. Ye, J. A. Berden, K. van Dam, and A. L. Kruckeberg. Expression and activity of the Hxt7 high-affinity hexose transporter of *Saccharomyces cerevisiae*. *Yeast*, 18(13):1257–1267, 2001.
- R. Young and H. Bremer. Polypeptide-chain-elongation rate in *Escherichia coli* B/r as a function of growth rate. *Biochemical Journal*, 160:185–194, 1976.
- J. Yu, M. Liberton, P. F. Cliften, R. D. Head, J. M. Jacobs, R. D. Smith, D. W. Koppe-naal, J. J. Brand, and H. B. Pakrasi. *Synechococcus elongatus* UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO<sub>2</sub>. *Scientific Reports*, 5:8132, 2015.
- T. M. Zabriskie and M. D. Jackson. Lysine biosynthesis and metabolism in fungi. *Natural Product Reports*, 17(1):85–97, 2000.
- A. Zaslaver, A. E. Mayo, R. Rosenberg, P. Bashkin, H. Sberro, M. Tsalyuk, M. G. Surette, and U. Alon. Just-in-time transcription program in metabolic pathways. *Nature genetics*, 36(5):486–491, 2004.
- S. Zhang and D. A. Bryant. The tricarboxylic acid cycle in cyanobacteria. *Science*, 334(6062):1551–1553, 2011.
- A. R. Zomorodi and C. D. Maranas. Improving the iMM904 *S. cerevisiae* metabolic model using essentiality and synthetic lethality data. *BMC Systems Biology*, 4(1):178, 2010.



# Notation

Here we summarize all the notation used in this thesis:

- $\mathbb{R}$  denotes the set of real numbers
- $\mathbb{R}_{\geq 0}$  denotes the set of real numbers that are greater or equal to zero
- $\mathbf{v}_i$  denotes the  $i$ -th element of the vector  $\mathbf{v}$
- $\mathbf{v}_A$  denotes the subvector of  $\mathbf{v}$  containing only the elements of the index set  $A$
- $S_{ij}$  denotes the element at row  $i$  and column  $j$  of the matrix  $S$
- $S_{i*}$  denotes the  $i$ -th row of the matrix  $S$
- $S_{*i}$  denotes the  $i$ -th column of the matrix  $S$
- $(\mathbf{a}, \mathbf{b})^\top$  denotes the transpose of the vector  $(\mathbf{a}, \mathbf{b})$ , i.e.,  $(\mathbf{a}, \mathbf{b})^\top := \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}$
- $\mathcal{R}$  denotes the set of reactions
- $\mathcal{M}$  denotes the set of internal metabolites
- $\mathcal{X}$  denotes the set of extracellular metabolites present in limited amounts
- $\mathcal{E}$  denotes the set of enzymes
- $\mathcal{Q}$  denotes the set of quota compounds
- $Irr$  denotes the set of irreversible reactions
- $G$  and  $TH$  denote glycogen and trehalose respectively
- $S$  denotes the stoichiometric matrix
- $\mathbf{v}$  denotes the steady-state flux vector
- $\mathbf{v}(t)$  denotes the flux vector at time point  $t$

- 
- $\tilde{\mathbf{v}}(T)$  denotes the average flux vector over the time period  $[0, T]$
  - $\bar{\mathbf{v}}$  denotes the average flux  $\tilde{\mathbf{v}}(T)$  for  $T \rightarrow \infty$
  - $\hat{\mathbf{v}}_{\mathcal{R}^E}$  denotes the enzyme production fluxes inferred using a machine learning model
  - $\mathbf{c}$  denotes the steady-state metabolite concentration vector
  - $\mathbf{c}(t)$  denotes the metabolite concentration vector at time point  $t$
  - $\bar{\mathbf{c}}(T)$  denotes the average metabolite concentration vector over the time period  $[0, T]$
  - $\bar{\mathbf{c}}$  denotes the average metabolite concentration vector  $\bar{\mathbf{c}}(T)$  for  $T \rightarrow \infty$
  - $\mathbf{e}$  denotes the steady-state enzyme concentration vector
  - $\mathbf{e}(t)$  denotes the enzyme concentration vector at time point  $t$
  - $\bar{\mathbf{e}}(T)$  denotes the average enzyme concentration vector over the time period  $[0, T]$
  - $\bar{\mathbf{e}}$  denotes the average enzyme concentration vector  $\bar{\mathbf{e}}(T)$  for  $T \rightarrow \infty$
  - $\mathbf{n}$  denotes the steady-state vector of molar amounts of both metabolites and enzymes, i.e.,  $\mathbf{n} := (\mathbf{n}_{\mathcal{M}}, \mathbf{n}_{\mathcal{E}})^{\top}$
  - $\mathbf{n}(t)$  denotes the vector of molar amounts of both metabolites and enzymes at time point  $t$
  - $\mathbf{l}, \mathbf{u}$  denote the constant lower and upper flux bounds vectors respectively
  - $\mathbf{l}(t), \mathbf{u}(t)$  denote the lower and upper flux bounds vectors at time point  $t$  respectively
  - $f(\cdot, \cdot)$  denotes deterministic kinetic rate laws
  - $K_M$  denotes the Michaelis constant
  - $k_{cat}$  denotes the turnover rate, with the variations  $k_{cat}^+$  and  $k_{cat}^-$  for the forward and reverse directions respectively
  - $V(t)$  denotes the volume of the cell population at time point  $t$
  - $\mu$  denotes the constant steady-state growth rate of a cell population
  - $\mu(t)$  denotes the growth rate at time point  $t$
  - $T_d$  denotes the division time
  - $R$  denotes the ribosome



# Glossary

**diauxic shift** Also known as diauxie, describes the growth phases of a microorganism in batch culture as it metabolizes a mixture of two sugars. Rather than metabolizing the two available sugars simultaneously, microbial cells commonly consume them in a sequential pattern, resulting in two separate growth phases.

**diurnal** Property of plant or animal behavior characterized by activity during the day, with a period of sleeping, or other inactivity, at night.

**doubling time** The period of time required for a population to double in size.

**EC number** Enzyme Commission number, a numerical classification scheme for enzymes, based on the chemical reactions they catalyze.

**enzyme** Biological macromolecule that acts as catalyst and thus increases the rate of a reaction.

**exchange reaction** Reaction that represents the transport of a metabolite from the environment to the cell or vice versa.

**external metabolite** Metabolite present in the environment of the cell.

**gene-reaction mapping** Mapping between reactions and the genes whose products are involved in the catalysis of those reactions.

**glycogen** Multibranched polysaccharide of glucose that serves as a form of energy storage.

**growth rate** The rate at which the number of individuals in a population increases.

**heterotrophic organism** Organism that cannot manufacture its own food, but relies on organic substances in the environment for growth and survival.

**internal metabolite** Metabolite present inside the cell.

## Glossary

---

**internal reaction** Reaction happening inside the cell.

**kinetic rate law** Equation that links the reaction rate with the concentrations of reactants, catalysts, and constant kinetic parameters.

**macromolecule** Very large molecule, such as protein, commonly created by polymerization of smaller subunits (monomers).

**metabolic network** Directed hypergraph that describes the biochemical reactions happening inside a cell.

**metabolite concentration** The molar amount of a metabolite divided by the volume of the cell.

**nonlimiting extracellular metabolite** Metabolite considered to be present at saturation in the medium and whose amount is not modeled explicitly.

**overflow metabolism** Seemingly wasteful strategy in which cells incompletely oxidize their growth substrate instead of using the more energetically-efficient respiratory pathway, even in the presence of oxygen.

**photoautotrophic growth** Growth of an organism using only sunlight as source of energy to synthesize organic compounds.

**quota metabolite** Metabolite or macromolecule that has no immediate catalytic role, but whose production should nevertheless be enforced.

**reaction rate** Also known as flux, it is the speed at which the reactants are transformed into products.

**respiro-fermentation** Strategy for metabolizing sugars under aerobic conditions that uses both the respiratory and the fermentative pathway.

**ribosome** Complex molecular machine, found within all living cells, that serves as the site of biological protein synthesis.

**SBML** Systems Biology Markup Language, a representation format based on XML for communicating and storing computational models of biological processes.

**spontaneous reaction** A reaction that does not need an enzyme for catalysis, but occurs without being driven by any outside force..

**steady-state assumption** Assumption that production and consumption rates for internal metabolites balance, and thus the level of internal metabolites is constant.

**stoichiometric matrix** Matrix representation of a metabolic network, where the rows correspond to internal metabolites, the columns to reactions, and the entry at row  $i$ , column  $j$  gives the stoichiometry of metabolite  $i$  in reaction  $j$ .



# Index

- Bacillus subtilis*, 18, 32, 70  
*Escherichia coli*, 17, 20, 49, 55, 70, 94, 106  
*Homo sapiens* (HeLa), 49, 50, 55, 94, 95  
*Saccharomyces cerevisiae*, 18, 21, 41, 49, 55, 61, 69, 70, 94, 98, 135–158  
*Synechococcus elongatus*, 21, 41, 72, 98, 103–133, 192  
*Synechocystis* sp., 106, 128, 129
- Gurobi, 74
- Arne Reimers, 43  
ATP maintenance, 72, 79, 114, 142, 143  
autocatalytic system, 19, 26, 64, 68, 104, 118  
average concentration, 47, 52, 53, 55, 93  
average flux, 46–48, 50, 51, 54, 55, 93  
average growth rate, 92
- balanced growth, 24, 25, 85, 112  
basis matrix, 97, 98  
BiGG, 59  
binary search, 32, 34, 40, 99, 115  
biological objective, 16, 159, 167  
biomass, 16, 38, 60, 65, 75, 82, 83, 85, 94, 137, 139, 143–145, 148, 152, 159, 162  
biomass reaction, 16, 50, 65–67, 69, 118  
biomass yield, 17, 54, 55  
BioModels, 59, 60  
BLOSUM62, 70  
bound on total proteome, 27, 31  
BRENDA, 59, 62, 69, 110, 140
- Calvin-Benson cycle, 126, 128  
carbon cycling, 117  
cellular economy, 19, 104  
cellular resource allocation, 18–21, 23–41, 43, 51, 57, 60, 81, 103–133  
circadian clock, 128, 129  
condition number, 96–98  
conditional flux balance analysis, 20, 50, 51, 55, 65, 81, 85, 87, 89–96, 99–100, 103–133  
constraint-based modeling, 15–17, 37  
convexity, 99–100  
CPLEX, 74, 98, 143  
cross section, 113, 116  
cyanobacteria, 65, 69, 85, 100, 103–133  
Cyanobase, 61  
cyclic objective, 87  
cytochrome  $b_6f$ , 107
- day-night environment, 40, 103–133  
debugging, 74–75, 96, 101

diauxie, 39, 146–147, 157, 158, 161  
 dilution, 29, 33, 34, 51, 89–96, 99, 115  
 discretization, 39, 40, 86–89, 114, 115, 142  
 division time, 116  
 doubling time, 33, 119, 147  
 dynamic enzyme-cost flux balance analysis, 20, 36–40, 43, 50, 57, 62, 81, 82, 100, 135–161, 164, 165, 167, 170  
 dynamic optimization, 39, 40  
 EC number, 61, 62  
 electron transport chain, 106, 107, 110  
 end time dependency, 82, 84, 89  
 enzyme, 14  
 enzyme capacity constraint, 37, 54, 55, 69, 88, 100, 110, 118, 140  
 enzyme concentration, 14, 16, 18, 31, 43, 44, 47  
 enzyme degradation, 35  
 enzyme production, 18, 19, 23, 26, 30, 36, 60, 61, 81, 104, 106, 109, 111, 128, 129, 140, 149, 153, 161, 163–166  
 explicit Euler, 86  
 exponential growth, 24, 25, 34, 41, 68, 82, 85, 91  
 fermentation, 18, 137, 145, 152, 153  
 flux, 14, 17, 18, 23, 26, 29, 32, 35, 37, 44–47, 71, 81, 123  
 flux balance analysis, 16–19, 27, 38, 43, 45, 47, 50, 54, 55, 117  
 Flux Balance Constraints, 57  
 flux bounds, 13, 14, 140, 142  
 flux modules, 19  
 flux variability analysis, 19  
 Genbank, 59  
 gene product, 78  
 gene-reaction mapping, 60, 106  
 genome-scale metabolic network, 12, 20, 21, 60, 136  
 glycogen, 21, 65, 69, 105, 108, 110, 126, 128–133  
 growth rate, 14, 16–18, 23–25, 28, 29, 32, 33, 39, 45, 47, 68, 71, 89, 91–93, 96, 99, 100, 104, 105, 116, 117, 119, 123  
 growth yield, 17–19, 23, 47  
 Gurobi, 98  
 Henning Knoop, 103, 104  
 Henning Lindhorst, 57  
 ill-conditioning, 98  
 implementation, 100–102  
 implicit midpoint rule, 41, 86, 114, 142  
 infeasibility diagnosis, 74  
 initial quota, 65, 67, 111, 141, 142  
 instantaneous growth rate, 71, 126, 145  
 irreducible inconsistent subsystem, 74  
 isoenzymes, 60, 61  
 Johan van Heerden, 135  
 KEGG, 59, 63, 106  
 kinetic model, 14, 19, 44, 50, 92, 95  
 kinetic parameter, 15, 16, 37, 52, 118  
 kinetic rate law, 15, 16, 26, 47, 51, 53  
 light availability, 113, 116  
 linear feasibility problem, 32, 115  
 linear growth, 82, 83, 85  
 linear program, 17, 34, 39, 40, 60, 62, 74, 85, 87, 96–98, 165, 166  
 mass-action kinetics, 52  
 MaxQuant, 59  
 ME model, 20, 32–35  
 ME models, 99

membrane, 24, 26, 64, 65  
 messenger RNA, 33, 34  
 metabolic network, 12, 13, 16, 23,  
     43–45, 52, 57, 59, 65, 92,  
     104, 106, 110, 140, 141  
 metabolite  
     external, 13, 24, 75, 82, 140  
     internal, 13, 24, 43, 75, 109, 140  
 metabolite concentration, 14, 30,  
     31, 44–47  
 Michaelis-Menten kinetics, 15, 26,  
     30, 31  
 minimal network finder, 60, 137  
 mixed integer linear programming,  
     137  
 ModelSEED, 59  
 molar amount, 14  
 molecular crowding, 18, 27, 30  
 Monod growth, 119  
 multiobjective modeling, 17  
 multiplication factor, 40, 87, 89, 99,  
     114, 116, 117, 119  
  
 NADPH dehydrogenase, 107  
 Needleman-Wunsch algorithm, 70  
 noncatalytic protein quota, 64, 67,  
     112, 119  
 nonlinear optimization problem, 28  
 notation, 12, 14, 231  
 numerical inaccuracy, 97  
 numerical instability, 74, 86  
  
 objective, 17, 38, 81, 83, 108, 114,  
     142, 159, 164, 167  
 optimal solution, 17, 19, 143, 156  
 ordinary differential equation, 14  
 overflow metabolism, 23, 28, 138,  
     139, 144–146, 157  
 oxidative pentose phosphate  
     pathway, 126, 128  
 oxygen evolution rate, 118  
 oxygen-evolving complex, 106  
  
 Pareto front, 17  
 Pathway Tools, 59  
  
 periodic environment, 85, 103–133  
 photodamage, 117  
 photon, 113  
 photosystem I, 106, 107, 113, 128  
 photosystem II, 106–108, 113, 128  
 phototrophic growth, 40  
 phycobilisome, 106, 108, 113  
 pigment, 112, 128  
 prediction horizon, 85, 143, 147,  
     148  
 Proteomaps, 59  
  
 quasi-steady-state, 15, 36, 37,  
     44–46, 96, 140, 145  
 quota compound, 64, 67, 69, 74, 76  
 quota metabolite, 26, 27, 29, 38, 40,  
     110, 111, 141  
  
 Radau collocation, 86  
 Ralf Steuer, 103  
 reaction, 78  
     exchange, 13  
     internal, 13  
     irreversible, 13, 78, 111, 140  
     reversible, 13, 78, 111, 140  
     spontaneous, 111, 140  
 reaction rate, 14  
 regulatory structure, 62, 159–171  
 resource balance analysis, 19,  
     29–33, 37, 43, 51, 55, 74, 91,  
     94, 95, 99, 159  
 respiration, 18, 108, 126, 145, 152,  
     153  
 Ribosomal Protein Gene Database,  
     59, 139  
 ribosome, 19, 23, 24, 26, 28, 30, 37,  
     59, 63, 68, 76, 104, 106,  
     108, 111, 139, 140, 148  
 ribosome translation rate, 34, 64,  
     106, 139  
 rounding error, 97, 98  
 RuBisCO, 123  
  
 SABIO-RK, 59, 69

Saccharomyces Genome Database, 61, 139  
SBML, 57, 75–80, 139  
scaling, 96–98  
self-replicator, 23–25  
sensitivity analysis, 119  
shift experiment, 102, 136, 142, 143, 147–157  
short term deFBA, 85, 102, 143, 147, 149, 159  
SoPlex, 74, 98, 116  
starvation, 69, 84  
steady-state, 15, 20, 25, 29, 39, 43–55, 90–96, 99, 109, 135, 140, 145, 157, 160, 166, 177  
stoichiometric coefficient, 14, 36, 65, 67  
stoichiometric matrix, 13, 16, 26, 29, 43, 44, 92  
storage, 21, 65, 69, 75, 76, 129, 139, 143  
substrate saturation, 30  
subunit stoichiometry, 59, 61, 62, 139  
trajectory, 52  
transporter, 24, 26, 69, 76, 141, 142, 155  
turnover rate, 15, 26, 29, 31, 32, 35, 37, 53, 59, 62, 69–70, 75, 79, 81, 87, 90, 95, 100, 110, 119, 139, 140  
UniProt, 59, 61, 63, 139  
unique solution, 19  
units, 72–73  
Warburg effect, 18  
Yeast 6, 65–67, 69, 137, 138



# Zusammenfassung

Diese Doktorarbeit ist ein Beitrag zur Systembiologie, wo große biologische Systeme wie der Metabolismus durch mathematische und rechnergestützte Modelle untersucht werden. Diese Modelle sind vereinfachte Darstellungen der biologischen Systeme und haben verschiedene Stufen von Granularität und Abstraktion, abhängig von Formalismus und Modellgröße.

Viele der größten untersuchten Netzwerke in der Systembiologie beschreiben den Metabolismus, welcher alle biochemischen Reaktionen innerhalb einer Zelle umfasst. Bis vor kurzem wurden solche großen metabolischen Netzwerke vor allem nur isoliert und unter stationären Bedingungen untersucht, ohne die Umgebungsdynamik oder die enzymatischen Ressourcen zu berücksichtigen, die notwendig sind, um alle biochemischen Reaktionen zu katalysieren. Dies geschah vor allem durch constraintbasierte Methoden und Optimierung. Während diese Methodik sich bei der Vorhersage des zellulären Verhaltens als sehr erfolgreich erwiesen hat, eignet sich dieser Ansatz nicht für Mikroorganismen, die in Umgebungen leben, die ständig Veränderungen ausgesetzt sind. Ein Beispiel sind Cyanobakterien, deren Stoffwechsel an die täglichen Veränderungen der Sonnenlichtverfügbarkeit angepasst ist. Ein weiteres Beispiel sind Hefen, die in großen Bioreaktoren leben und sich in einem dynamisch verändernden Umfeld befinden, welches durch lokale Heterogenitäten bestimmt ist.

Diese Doktorarbeit basiert auf neuen Formalismen, die die dynamische Ressourcenallokation des Stoffwechsels modellieren und dabei die Methoden der dynamischen Optimierung und Optimalsteuerung verwenden. Wir konzentrieren uns auf die Modellierung und das Verständnis der Ressourcenallokation in großen (manchmal genomweiten) metabolischen Modellen.

Nach einer Übersicht über die vorhandenen Methoden für das Studium der metabolischen Ressourcenallokation präsentiert die Arbeit eine neue mathematische Ableitung der weit verbreiteten Flussgleichgewichtsannahme für metabolische Netzwerke und zeigt, wie dies dazu genutzt werden kann, um Obergrenzen für Lösungen der dynamischen Ressourcenallokation zu bestimmen. In Vorbereitung auf die Fallstudien stellen wir eine Methodik zur Erstellung eines dynamischen Ressourcenallokationsmodells unter Verwendung von Informationen aus Online-Datenbanken dar. Darüber hinaus werden Leitlinien und nützliche Problemtransformationen zur Lösung dynamischer Ressourcenallokationsprobleme vorgestellt. Diese Theorie wird dann in zwei Fallstudien angewendet. Eine Studie betrachtet das Cyanobakterium *Synechococcus elongatus* PCC 7942 und stellt die erste genomweite dynamische Ressourcenallokationsstudie dar. Diese gibt Einblick in die zeitliche Organisation der Enzymsyntheseprozesse, die der Lichtverfügbarkeit folgen und zeigt, dass das lineare Muster der Glykogenakkumulation während der Tageszeit ein optimales Verhalten ist, welches als Kompromiss zwischen mehreren widersprüchlichen Ressourcenallokationszielen entsteht. Die zweite Fallstudie betrifft Hefe. Wir wollen verstehen, welche Mechanismen es ermöglichen, dass einige der Zellen Umgebungsübergänge überleben und andere nicht. Wir zeigen, dass Überlaufstoffwechsel und Diauxie, Phänomene die weit verbreitet in der Natur sind, optimale Verhaltensweisen aus einer Ressourcenallokationsperspektive sind. Darüber hinaus untersuchen wir, wie man Ressourcenallokationsmodelle verwenden kann, um zu verstehen, wie sich Hefe an Sauerstoff- und Nährstoffverfügbarkeitsänderungen anpasst. Wir enden mit einem Perspektivenkapitel, in dem wir einige vorläufige Ergebnisse für die Verwendung von Zeitreihen aus dynamischen Ressourcenallokationsmodellen präsentieren, um auf die regulatorischen Strukturen zu schließen, die diese optimalen Verhaltensweisen implementieren.



# Curriculum vitae

Name	<b>Alexandra-Mirela Reimers (born Grigore)</b>
Current Position	<b>Research Assistant</b> ERASysAPP European Project: ROBUSTYEAST <a href="https://www.erasysapp.eu/calls/2nd-call/robustyeast">https://www.erasysapp.eu/calls/2nd-call/robustyeast</a> Freie Universität Berlin, Berlin, Germany
University Education	<b>2014-2017 PhD student in Bioinformatics</b> Freie Universität Berlin, Berlin, Germany Thesis: Understanding metabolic regulation and cellular resource allocation through optimization Supervisor: Prof. Dr. Alexander Bockmayr <b>2012-2014 Master of Science, Bioinformatics</b> Freie Universität Berlin, Berlin, Germany Thesis: Understanding metabolic regulation through optimisation Supervisor: Prof. Dr. Alexander Bockmayr <b>2009-2012 Bachelor of Science, Bioinformatics and Computational Biology</b> Jacobs University Bremen, Bremen, Germany Thesis: Flux balance analysis on random metabolic networks Supervisor: Prof. Dr. Marc-Thorsten Hütt
Publications	<b>The steady-state assumption in oscillating and growing systems</b> Alexandra-M. Reimers, Arne C. Reimers Journal of Theoretical Biology, 406, 176-186, 2016. <b>Minimal equivalent subgraphs containing a given set of arcs</b> Arne C. Reimers, Alexandra-M. Reimers, Yaron Goldstein Theoretical Computer Science, 675, 56-63, 2017. <b>Evaluating cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth</b> Alexandra-M. Reimers, Henning Knoop, Alexander Bockmayr, Ralf Steuer Proceedings of the National Academy of Sciences, 114 (31), E6457-E6465, 2017. <b>A Protocol for Generating and Exchanging (Genome-Scale) Metabolic Resource Allocation Models</b> Alexandra-M. Reimers, Henning Lindhorst, Steffen Waldherr Metabolites, 7 (3), 47, 2017.

Workshops & Conferences	<p>Oct. 2016 Foundations of Systems Biology in Engineering Magdeburg, Germany</p> <p>Oct. 2016 International Study Group for Systems Biology Jena, Germany</p> <p>Oct. 2015 Constraint-based Reconstruction and Analysis Heidelberg, Germany</p> <p>Jun. 2015 Metabolic Pathway Analysis Braga, Portugal</p> <p>Jun. 2014 Workshop on the Economy of a Cell: Resource Allocation, Trade-Offs and Efficiency in Living Systems, Trieste, Italy</p> <p>Dec. 2013 Workshop on Integrative Omics, Pucon, Chile</p> <p>Sep. 2013 Metabolic Pathway Analysis, Oxford, UK</p> <p>Jun. 2013 Systems Medicine in Cancer, Berlin, Germany</p> <p>Aug.-Sep. 2012 Otto Warburg International Summer School on genes, metabolism and systems modelling, Shanghai, China</p>
Teaching Experience	<p><b>Oct. 2014-Mar. 2015 Teaching Assistant</b> Freie Universität Berlin, Berlin, Germany Statistics for Bioinformaticians</p> <p><b>Feb. 2012-Jun. 2012 Teaching Assistant</b> Jacobs University Bremen, Bremen, Germany Natural Science Laboratory Computer Science II - Advanced C/Python</p> <p><b>Sep. 2011-Dec. 2011 Teaching Assistant</b> Jacobs University Bremen, Bremen, Germany Engineering and Science Mathematics 1A</p>
Honours & Awards	<p><b>2013-2015 PhD Scholarship</b>, International Max Planck Research School for Computational Biology and Scientific Computing</p> <p><b>2012-2013 Predoctoral Scholarship</b>, International Max Planck Research School for Computational Biology and Scientific Computing</p> <p><b>Sep. 2015 Travel Grant within the ERA-Net on Applied Systems Biology (ERASysAPP) Mobility Program</b></p>
Languages	<p>Romanian - native or bilingual proficiency</p> <p>English - native or bilingual proficiency</p> <p>German - working proficiency</p> <p>French - limited working proficiency</p>

Berlin, 17. Oktober 2017

Alexandra-M. Reimers

# Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich alle Hilfsmittel und Hilfen angegeben habe und versichere, auf dieser Grundlage die Arbeit selbstständig verfasst zu haben. Die Arbeit wurde nicht schon einmal in einem früheren Promotionsverfahren eingereicht.

Berlin, August 2017

Alexandra-Mirela Reimers



# Acknowledgements

None of the work in this thesis would have been possible without the many people that have helped me through it. So these last two pages are about them.

My very first thanks go to my PhD supervisor, Alexander Bockmayr. Now, when I'm so close to the end, I realize that this PhD has not been about the destination, but much more about the journey and about how much I learned during this journey. And, of all the people that have shaped this path, I learned most from him. Thank you, Alexander, for the great supervision, for the many things you taught me both during my master and during my PhD. Thanks for your support and guidance, and above all for your patience, thoroughness, and objectivity. To me, the German word "Doktorvater" best describes your role in all this.

I would also like to thank Edda Klipp for agreeing to be the second reviewer of this thesis. I know it's long and it's a lot of work. Thanks, Edda, also for putting together the CSB graduate school, with which I interacted a lot at the retreats and seminars during this PhD.

This PhD would not have been possible without the International Max Planck Research School (IMPRS) for Computational Biology and Scientific Computing. I would like to thank the IMPRS for funding both my predoctoral and part of my PhD studies. Within the IMPRS, I would like to give some special thanks to Knut Reinert and Kirsten Thobe, who have agreed to be in my PhD committee. Some more special thanks go to Heike Siebert from whom I have also learned a lot both in my master and in my PhD time. A lot of support came also from Kirsten Kelleher and Fabian Feutlinske, IMPRS coordinators and the people who helped me a great deal with administration and funding. And, of course, thanks to all my IMPRS fellow students!

I would like to also acknowledge funding from the ERA-Net for Systems Biology Applications, project ROBUSTYEAST. And talking about this project, I would like to thank Henning Lindhorst for being such an awesome collaborator.

On nearly every day of work during my master and PhD time I was surrounded by the members of two research groups: Mathematics in Life Sciences and Discrete Biomathematics. And I would like to thank all their present and former members for creating such a friendly and supportive atmosphere. Thanks to Aljoscha, Annika, Arne, Hannes, Katinka, Laszlo, Lin, Ling, Marco, Neveen, Shahrads, Steffen, Therese, Yaron, Adam, Firdevs, Kirsten, Melania, Robert, and Katja! Within this big crowd, some special thanks go to all my wonderful office-mates: Annika, Arne, Lin, Steffen, and Yaron.

A big thank you to Alexander, Annika, and Arne for having proofread this thesis!

I would also like to thank my family and friends for their support and for always being there for me when I needed them.

Finally, I would like to thank Arne for being always at my side. You already know that none of this would have been possible without you. Thank you for all your love, for your patience, for being there in the good and especially in the bad moments. And most of all thanks for convincing me to continue in all the moments when I wanted to give up this PhD. No words can give credit to your help and this is why this thesis is dedicated to you.





