Technische Universität München
Systembiotechnologie

MODEL BASED ANALYSIS OF CENTRAL METABOLIC PATHWAYS OF *HALOMONAS ELONGATA*

Christiana Verena Sehr

Vollständiger Abdruck der von der Fakultät für Maschinenwesen
der Technischen Universität München zur Erlangung des akademischen
Grades eines Doktor-Ingenieurs
genehmigten Dissertation.

Vorsitzender:	Prof. Dr. S. Berensmeier
Prüfer:	1. Prof. Dr. A. Kremling
	2. Prof. Dr. E. Voit
	3. TUM Junior Fellow Dr. A. Marin-Sanguino

Die Dissertation wurde am 30.06.2016 bei der Technischen Universität
München eingereicht und durch die Fakultät für Maschinenwesen
am 22.09.2016 angenommen.

ABSTRACT

Halophiles, like *Halomonas elongata*, show great biotechnological potential, especially concerning the production of compatible solutes. Since the metabolic details, which lead to its production, remain unclear, a thorough analysis is required. The scope of the thesis is to elucidate certain aspects of the metabolism of *H. elongata* applying mathematical modeling.

Due to the absence of large data sets, which leaves a great number of degrees of freedom, the analysis is characterized by the sequential application of constraint-based techniques, which reduce the dimensionality of the parameter space. Since the genome of *H. elongata* is known, the workflow is initiated by metabolic reconstruction, followed by a stoichiometric analysis of the resulting metabolic network via Flux Balance Analysis, which indicates the major paths through the network, leading to ectoine production. The analysis if followed by a thermodynamic feasibility analysis of the resulting flux distributions, which moreover returns information concerning feasible metabolite concentrations. The framework is finally completed by the reduction of the model via thermodynamic shortening and the utilization of the gained insights in building a stable dynamical model.

The application of the workflow has produced a series of models that were able to identify optimal pathways, as well as thermodynamic bottlenecks. These models are furthermore comparable with fermentation data. The results show that a sequential application of constraint-based techniques already provides great insights into the metabolic network of an organism, even without the availability of large data sets.

The challenge in mathematical modeling:
... not to produce the most comprehensive descriptive model but to produce the simplest possible model that incorporates the major features of the phenomenon of interest.

— Howard Emmons [73]

ACKNOWLEDGEMENTS

This thesis represents the end of my PhD. The past four years have been very eventful with several ups and downs. I am very glad that I got the offer to do a PhD at the Specialty Devision for Systems Biotechnology at the 'Technische Universität München'. This opportunity did not only complement my studies in mathematics but also broadened my mind with respect to coding and awoke my passion for it.

In this respect, I want to thank my supervisors Professor Andreas Kremling and Dr. Alberto Marin-Sanguino for this chance and their guidance throughout my PhD. Moreover, I thank Professor Eberhard Voit, who is kind enough to read and evaluate my thesis and to fly all the way to Munich to participate in my defense. Professor Michael Savageau deserves a lot of credit as well, for allowing me to stay at his lab in Davis and for guiding me through the Design Space Analysis. In this respect, I also thank Dr. Jason Lomnitz for his patient assistance with the Design Space Toolbox and for making it so much fun.

Moreover, I want to thank my family, especially my mom, who supported me in every possible way. Without their steady encouragement I would not have succeeded.

Alfonso Chacón Roldán deserves a lot of credit for always being there, when I needed help with my work or moral support. I am so lucky to have you!

Last but not least, I give thanks to my uncle Anton Güth, who was so kind to read the thesis and correct my spelling, while actually being on vacation.

CONTENTS

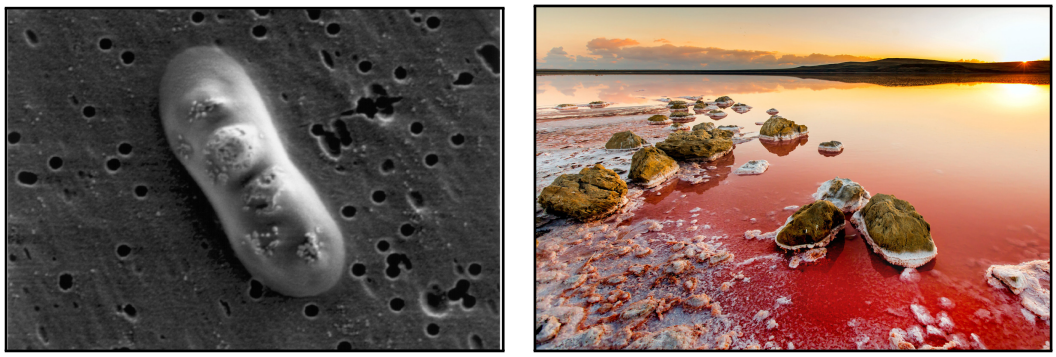
LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

The present thesis deals with the organism *Halomonas elongata* and intends to contribute to a quantitative understanding of its metabolism with the help of mathematical modeling. The aim is to formulate a model of the metabolism, concerning the product ectoine, which could furthermore be used to guide the design of improved strains able to achieve higher yields and also produce ectoine efficiently from a wider spectrum of raw materials.

The organism in question is *H. elongata*, a halophilic γ-proteobacterium, see Figure 1 (a), which is commonly found in hypersaline environments with salt-concentrations exceeding 0.5 mol/l, like the salt lake in Figure 1 (b) [58, 153]. In order to survive, the organism is bound to establish an osmotic equilibrium by synthesis and/or import of the compatible solute ectoine [48]. Since compatible solutes protect proteins, membranes and even whole cells against denaturation, inactivation and inhibition by heat or hyperosmotic stress, they turn out to be very useful for biotechnological applications, especially with respect to cosmetics. Ectoine is superior in its protective abilities, in particular against UV radiation. For this reason it is produced annually on a scale of tons with its producer strain *H. elongata* [153]. Due to its industrial importance, it is necessary to gain a deeper understanding of the mechanisms in *H. elongata*, which lead to ectoine production. The development of a mathematical model, which is able to further elucidate the understanding of the metabolism of *H. elongata*, with a view to guide the way to experimental improvements of the organism with respect to product yield, is the focus of the present thesis.



(a) (b)

Figure 1: *H. elongata* (a) [97] and its habitat (b) [166].

H. elongata is still a relatively unknown organism. Therefore, the amount of information available is manageable. Due to Schwibbert et al. [153], the complete genome of the bacterium has been sequenced and the ectoine metabolism

identified and characterized. Furthermore, data concerning fermentation experiments and enzyme assays are available, although not yet published. For this reason, the resulting mathematical model should be able to be combined with the fermentation data and based on its analysis able to propose further experiments.

Since the amount of available data is marginal, many different models can fit all available data. Therefore, instead of inferring one such model, a workflow is established, which connects different modeling techniques in a sequential way to gain further insights and reduce the degrees of freedom before analyzing any data. The framework includes the following tools:

STOICHIOMETRY In a first step, the metabolic network is analyzed with respect to mass conservation via Flux Balance Analysis (FBA), since this technique requires little information concerning enzyme kinetics or metabolite concentrations. Consequently, one obtains an idea of the flux distributions within the network, the maximal production yields and the maximal possible P:O ratio.

THERMODYNAMICS Subsequently, the flux distributions are analyzed from an energetic perspective, which adds further restrictions to the analysis and verifies, if the flux distributions are feasible. The tools used in this step are the Network-Embedded Thermodynamic (NET) analysis and the Max-min Driving Force (MDF) algorithm, complemented by the Specific Ion interaction Theory (SIT), which allows the thermodynamic analysis at elevated salt concentrations.

DYNAMICS The analysis is concluded by a consideration of regulation within the network, which follows a significant model-reduction with help of the newly developed method, *Thermodynamic shortening*. The tool used in the definition of the dynamical model is the Biochemical Systems Theory (BST). Another possible method would be the well-known *parameter estimation* [13]. Since this method depends on the available data sets and only mirrors the conditions under which the data have been generated, it has been discarded as too specific.

The mentioned tools yield qualitatively different information, which are able to reduce the degrees of freedom significantly, if taken together. This smooths the way for a subsequent application of the available experimental data. The challenge of this workflow approach lies in the initial selection and subsequent integration of the methods described above. The framework has been published in 2015 in [155]. Around the same time, the group of Hatzimanikatis published a paper, which is concerned with the same problem and in which they choose a similiar approach [10]. The fact that other groups work on the same problem emphasizes how relevant the topic is.

The thesis is composed of three parts: The first part (Material and Methods) reviews existing analytic methods, starting out with metabolic reconstruction, followed by the application of tools from thermodynamics and finally methods, which allow the analysis of a dynamical model. In order to combine them in a functional workflow, a certain missing tool had to be developed in the course of this work. It is further described at the beginning of the second part (Results). The chapters discuss the application of the newly developed workflow with respect to the metabolism of *H. elongata* to shed light on distinctive features of the network, like the absence of the fructose-1,6-biphosphate phosphatase or the presence of a sodium-translocating version of complex I in the respiratory chain, in addition to the classical proton-dependent one.

MATERIAL AND METHODS

2 METABOLIC RECONSTRUCTION AND STOICHIOMETRIC CONSTRAINTS

In order to build a mathematical model of the metabolism of *H. elongata*, the genome, as presented by Schwibbert et al. [153], has to be examined. This examination is needed to gain insight into the different gene products and their relation to each other. The process, called *metabolic reconstruction*, allows the modeler to obtain a first impression of the metabolic structure. A general protocol for this process has been written by Thiele and Palsson [169]. Several scientists work on software to find automated, or at least semi-automated, ways to cope with this task but so far they have not yet come up with a flawless tool. Thus, the community is left with manual evaluations of such an automated reconstruction [50, 169]. For this reason, the initial task of this project consists of the manual evaluation of key genes of the genome of *H. elongata* and the subsequent formulation and analysis of the obtained metabolic structure.

2.1 ANNOTATION

The process of metabolic network reconstruction generally begins with the automatic annotation of relevant genes, which identifies all open reading frames (ORFs) of the genome [137]. These are the genes, which encode a protein or functional RNA. Schwibbert et al. [153] already used automatic annotation strategies to find the ORFs of *H. elongata* and to identify their gene products. This step is followed by a manual evaluation.

The most reliable evaluation approach is a database search for homologs [84]. Essentially this means the identification of similar sequences in other better understood genomes in publicly available databases. Since mutations in the genome have likely appeared, ORFs do not have to be completely identical. For sequences to be called homologous and biological relevant, they need to fulfill a certain criteria: The similarity score of the alignment shall not be improvable by the addition or dismissal of any letters, representing nucleotides or aminoacids. Such an alignment is called *high-scoring segment pair (HSP)* and is the aim of a homology search.

In order to find homologs, one can either consider the DNA or the protein sequence, but the latter has several distinct advantages and therefore might lead to a greater sensitivity. Two of the most common advantages of protein sequences are: First, that statistical significance can be accomplished with much shorter sequences, because there are more amino acids than nucleotides, and therefore fewer repetitions of certain patterns. Second, it is less likely for an amino acid to be replaced during evolution, than for a nucleotide.

The manual annotation of the genome of *H. elongata* is therefore done on the protein level, with help of the database UniProt, which is an useful tool to find homologs to known ORFs. UniProt [31] is the largest protein-database and combines the data from Swiss-Prot, TrEMBL (computer-annotated supplement to Swiss-Prot) and PIR (Protein Information Resource). It is updated regularly and moreover makes use of the Basic Local Alignment Search Tool (BLAST), which is the most widely used and fastest sequence-alignment tool so far [84]. Although BLAST uses a solid statistical foundation, it still produces false positives, as well as false negatives. One could try to improve the search result by only considering hits with very high statistical significance, to eliminate the false positives but this would also result in missing numerous homologs. Therefore, it seems more reasonable to check the outcome of the search manually and decide on the most reasonable homolog. This causes the genome annotation to be the rate limiting step in most genome projects.

Once a homologous sequence is found, the corresponding encoded protein can be identified, thus detecting the homologs of the ORFs and at the same time their potential function in the genome as well.

Despite significant improvements, genome annotation is far from perfect and may contain multiple incorrect or misleading outcomes, due to automatic annotation. Moreover, UniProt records include functional predictions, which have been neither published in a scientific journal nor experimentally confirmed. Therefore, all findings need to be reviewed case by case to increase reliability when identifying the best match. For purposes of the present work, the focus of the manual evaluation is on the ORFs for the central metabolism of *H. elongata*, for which a descriptive mathematical model is developed.

As an illustration of the complexity consider an actual example. It is commonly accepted that *E. coli* and *H. elongata* are close relatives. The annotation for ORF *Helo_3010* of *H. elongata* using BLAST in Uniprot results in a 54 % match with the phosphoenolpyruvate carboxylase of *E. coli* (strain K12). There are better matches with other organisms available (for example *H. salina* at 90.6 %) but given the close relationship, *E. coli* has the advantage to be experimentally well understood. Thus, there is a significant likelihood that the enzyme of the homolog is already functionally characterized. This is indeed borne out, the enzyme of the homolog has been characterized in the laboratory and is therefore the best annotation match for *Helo_3010* [63], supporting the conclusion that the gene *Helo_3010* probably encodes the phosphoenolpyruvate carboxylase in *H. elongata* as well.

2.2 STOICHIOMETRIC CONSTRAINTS

The identification of the gene products and their relations goes hand in hand with their conversion into a mathematical representation to be useable for computation of network and/or physiological capabilities [50]. This is the basis for a genome-scale model. Schwibbert et al. reconstructed the basic structure of the metabolism of *H. elongata* for a core-model in [153]. This existing model has

been extended and modified in the course of this project as described in the results section (Chapter 6).

Following the metabolic reconstruction, the next important step is the deployment of a large range of computational tools to analyze functional network properties [137]. In a perfect world one could utilize a dynamic mathematical model to reflect the time dependency of the available data. Unfortunately, the necessary mechanistic details and kinetic parameters for this approach are rarely available for large-scale networks, as stated by Klamt and Stelling [81] and Beard and Qian [19]. Therefore, one may focus on the evaluation of systemic properties instead, for instance with a methodology called *constraint-based modeling*, which deals, among others, with the stoichiometry of the metabolic network [50, 137]. This approach does not require detailed kinetic parameters or a-priori knowledge of concentrations or fluxes but is able to include them in form of constraints, if available [19]. It has proven to be an effective strategy for predicting the physiological response of microorganisms [98] and metabolic capabilities [90]. Constraint-based modeling takes advantage of physiochemical constraints such as mass balance, energy balance, and flux limitations to describe the potential behavior of an organism [78]. In this respect, one assumes that the organism will reach a steady state that satisfies the physiochemical constraints under any given environmental condition. To identify such a steady state, an optimization is carried out, which has to satisfy the given constraints [154]. One form of constraints, which can be derived directly from the metabolic reconstruction, are stoichiometric constraints, since the reconstruction gives information with respect to the interplay of metabolites and reactions. This information is collected in the *stoichiometric matrix*, as will be explained in the following.

2.2.1 *The stoichiometric matrix*

The stoichiometric coefficients of the metabolites in question are typically represented by a stoichiometric matrix (nxm), whose n rows correspond to the metabolites and the m columns reflect the reactions of the network and therefore the chemical transformations caused by the enzymes [30, 119, 137]. In this way, the stoichiometric matrix describes the structural relationships between the network components. This matrix is quite important because it represents the translation of biological knowledge into mathematical terms [96]. A metabolite, which takes part in the j^{th}-reaction as a substrate, has a negative entry and a metabolite, which takes part as a product, has a positive entry in the j^{th}-column. As an example, one may consider the following network with five metabolites and four reactions:

$$X_0 \xrightarrow{v_1} X_1 \xrightarrow{v_2} X_2 \xrightarrow{v_3} X_3 \xrightarrow{v_4} X_4$$

The stoichiometric matrix in this case would be:

$$\mathbf{S} = \begin{array}{c} \\ X_0 \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & v_2 & v_3 & v_4 \end{array} \\ \begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

The metabolites in a stoichiometric matrix can be divided into internal and external species [81]. The external species are thought to be sinks or sources, which might physically lie outside of the cell and take only part in a single reaction. They are the reason, why the stoichiometric coefficients of a reaction may only have positive or only negative signs, when external species are ignored by removing their rows. In the example above, the external species are X_1 and X_4. The rest are internal species.

Once the stoichiometric matrix has been determined, mass balances involving internal species can be represented as follows (vectors are marked by arrows and matrices are written in bold) [87, 96]:

$$\frac{d\vec{x}}{dt} = \mathbf{S}\,\vec{v}(\vec{x}) - \mu\,\vec{x} \tag{1}$$

$\vec{x}$ (nx1) is a vector of intracellular metabolite concentrations, while $\vec{v}(\vec{x})$ (mx1) is a flux vector depending on named metabolites and μ represents the specific growth rate of the cell. Because the dilution term $\mu\,\vec{x}$ is much smaller, than $\mathbf{S}\,\vec{v}$, in the context of this work, it can be disregarded. Equation 1 is the *dynamic mass balance equation* and it describes the evolution of the concentration of each metabolite, x_i, over time. The intracellular reaction mechanisms of metabolic networks are still unknown, thus one usually reverts to the assumption that the internal species are in a pseudo-steady state, because the time scales inside the cell are much faster, than outside. Therefore, the following formula (the *law of mass conservation*) applies for the internal species [96]:

$$\mathbf{S}\,\vec{v} = 0 \tag{2}$$

Equation 2 describes the space, which includes every feasible flux distribution. This space can be further restricted by, e.g., the definition of lower bounds for irreversible reactions, like $\vec{v} \geqslant 0$ or additional upper bounds concerning the capacity of fluxes, as $\vec{v} \leqslant v^{max}$. As displayed in Figure 2, the possible solution space decreases with every addition.

The stoichiometric matrix, without external species, can be characterized by four fundamental subspaces: the (right) null space, the left null space, the row subspace and the column subspace [87]:

- The *row subspace* indicates temporal changes of variables and therefore enables the formation of pool variables. Its dimension is r (the rank of the stoichiometric matrix), which is usually smaller than n, because $\mathbf{S}$ typically does not have full rank.

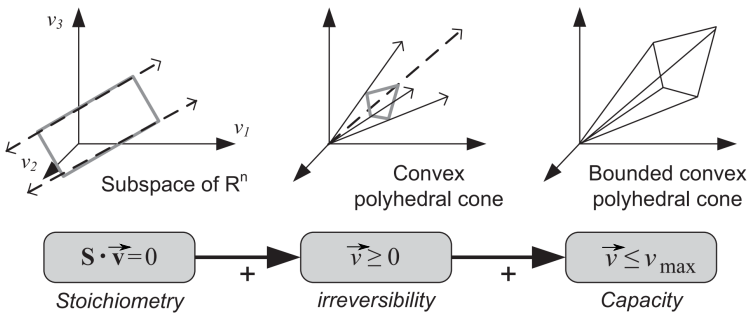


Figure 2: In addition to Equation 2 one can add further restrictions, like the irreversibility of certain reactions and/or maximum capacity constraints, derived from enzyme or transport capacities, resulting in a decreased solution space [96].

- The *column space* indicates the weight of each rate in the velocity vector and can be used for model reduction. Its dimension is also r.
- The *null space* indicates the stationary flux maps, the ones, which fulfill Equation 2. Its dimension is $m - r$.
- The *left nullspace* describes the conservation conditions and therefore the time invariants. A conservation equation is the weighted sum of a metabolite concentration, which stays constant. Its dimension is $n - r$.

2.2.2 *Flux Balance Analysis (FBA)*

There are several constraint-based approaches with respect to stoichiometric modeling, like *convex analysis* or *metabolic flux analysis*. One of the most well known examples is the *Flux Balance Analysis (FBA)* [96, 172, 173]. In this case, the constraints are predefined through the underlying biochemical network. FBA makes it possible to calculate the flow of metabolites through a metabolic network and thereby allows for the prediction of the growth rate of an organism or the rate of production of other metabolites of interest [119], because it is assumed that cells have evolved to achieve an optimal behavior owing to evolutionary pressure [96]. That is, cells tend to regulate their fluxes towards an optimal flux distribution. A central aspect of FBA is to find such optimal flux distributions via linear programming.

In order to apply FBA, one has to take the list of stoichiometrically balanced biochemical reactions from the metabolic reconstruction and convert them into a mathematical model by forming the stoichiometric matrix, as is typical for a constraint-based approach (see Figure 3 for an overview). Growth can be incorporated through a single reaction, which represents the consumption of the precursors for biomass production (yellow column in Figure 3). The green columns represent the exchange reactions, which stand for the flow of certain metabolites in and out of the cell. FBA, as a constraint-based method, follows the law of mass conservation. This homogeneous system of linear equations

demands that the production and consumption of an internal metabolite must cancel each other out.

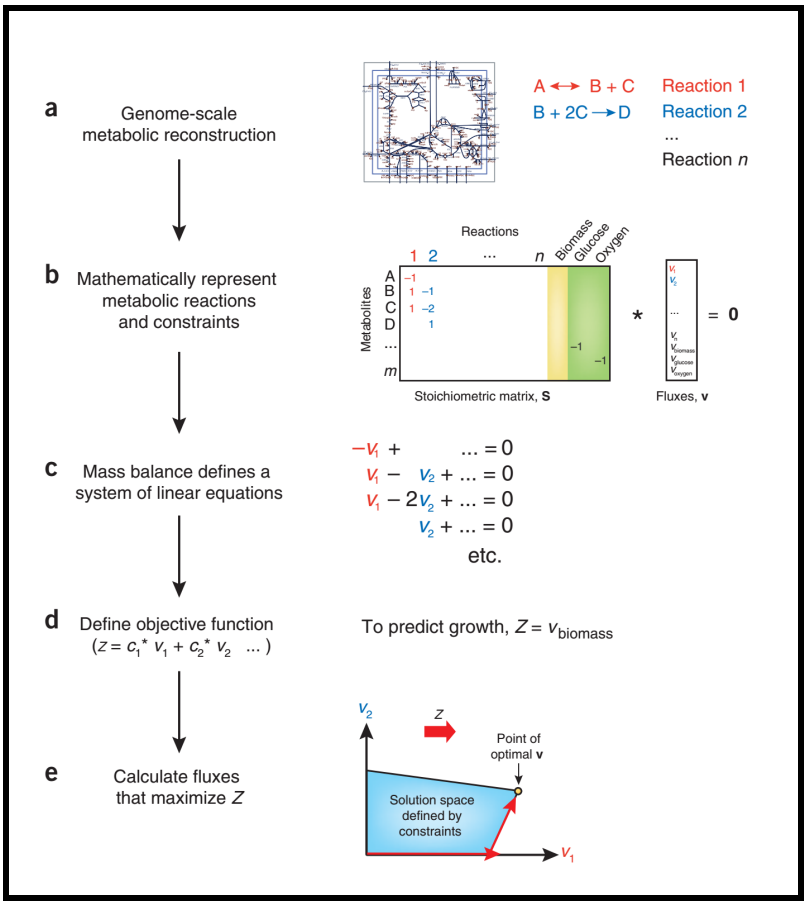


Figure 3: (a) Metabolic reconstruction provides insights into the network structure. (b) This reconstruction is converted into a mathematical model by forming the stoichiometric matrix. Growth is incorporated with an additional reaction (yellow column). Exchange reactions (green columns) represent the flow in and out of the cell. (c) At steady state, the flux through each reaction is given by Equation 2. (d) In order to build the optimization problem, an objective function needs to be defined, like one for growth-maximization. (e) Linear programming is used to maximize or minimize the objective function to identify an optimal flux distribution [119].

As a metabolic network usually has more reactions, than metabolites, it is underdetermined and therefore there is more than one solution, an issue that has been addressed by Kelk et al. [79]. One has to choose an objective, a metabolite/product, which is supposed to be optimized, such as biomass. The objective function Z can be any linear combination of fluxes: $Z = \vec{c}^{\,T} \cdot \vec{v}$, where $\vec{c}$ is a vector of weights. With the use of linear programming, it is possible to identify a flux distribution, which is a minimization or maximization of the objective function, as long as the problem is feasible. In order to get a biologically feasible solution, one can constrain the reaction directions of the known irreversible reactions. In case that a reaction is irreversible forward, the flux has to be positive ($v \geqslant 0$), if it is irreversible backward, the flux is negative ($v \leqslant 0$). The reversible reactions stay unconstrained and if a reaction should not take

part in the optimization, its upper and lower bound are set to zero. Despite all the constraints, the space of possible flux distributions can be quite large, as outlined exemplarily through the blue region in Figure 3 (e) [119] but the single optimal flux distribution in this simple example is highlighted through a little dot in the upper right corner.

A typical formulation of such an optimization problem is the following example. It describes the optimization of biomass, where the upper and lower bounds of the fluxes are defined by $\vec{v_L}$ and $\vec{v_U}$:

$$\begin{aligned} &\max Z = v_{bio} \\ &\text{s.t:} \\ &\quad \mathbf{S}\,\vec{v} = \vec{0} \\ &\quad \vec{v_L} \leqslant \vec{v} \leqslant \vec{v_U} \end{aligned} \tag{3}$$

In general, one should also define an upper bound for the uptake of some limiting nutrient (v_{in}), like glucose, to avoid the solution to be infinite growth. Then, all the fluxes have to be seen as yields, because their values depend on the limiting nutrient, which is the input of the system and the vector of solutions becomes the vector of yields. The k^{th} entry would be: $Y_{k,in} = \frac{v_k}{v_{in}}$. There exists a disagreement between the usage of optimal yields or optimal rates, as stated by Schuster, Pfeiffer, and Fell [149], e.g. rate maximization matches experimental results for the energy production pathway in *S. cerevisiae* better, while yield maximization reproduces the experimental observations in *E. coli*. Throughout this work, the flux distributions will be seen as vectors of yields.

Both the uptake of some input or the output can be constrained. Therefore, one can set a certain minimal constraint for the production of, for example, biomass and optimize the production of another metabolite or even the uptake of some limiting nutrient. This way, it is possible to ensure, for instance, the same biomass production as in Equation 3, if its maximal growth value (v_{bio}^{max}) is used as constraint in Equation 4 but with minimal uptake of glucose (v_{in}). This reformulation of a linear problem is often used in trade-off analysis [34] and it is also a well established method in multi-criteria programming [165].

$$\begin{aligned} &\min Z = v_{in} \\ &\text{s.t:} \\ &\quad \mathbf{S}\,\vec{v} = \vec{0} \\ &\quad \vec{v_L} \leqslant \vec{v} \leqslant \vec{v_U} \\ &\quad v_{bio}^{max} \leqslant v_{bio} \end{aligned} \tag{4}$$

A solution to the problem, as formulated in Equation 4, can be interpreted as the flux distribution that is able to achieve a pre-established growth rate, while making the most efficient use of the input [155].

Another relevant element of FBA is the determination of the importance of single reactions for the overall system performance by studying knock-out mutations [81]. Here, certain reactions are constrained to have a zero-flux to gain

insight into the new direction of the system-fluxes and to determine how this affects the objective: Is the optimal value diminished or not affected at all? Edwards and Palsson [43], for instance, studied knock-out mutations to identify optimal growth phenotypes of the metabolism of *E. coli* for different growth conditions.

Therefore, using FBA on metabolic networks rather aims at constructing a model, which is capable of predicting the phenotype, which will be expressed under certain circumstances, than to determine the flux distribution, which fits a certain data set [96].

Despite its utility, FBA does have limitations: If a model includes closed cycles, FBA solutions exhibiting a net flux around such a cycle can contradict Kirchhoff's second law for electric circuits, because it is a thermodynamically infeasible loop. These cycles are sets of reactions, which carry a flux around and therefore are 'perpetual motion machines' [40]. Luckily, this problem can be prevented by formulating a new optimization problem, which keeps the optimal value of the actual objective fixed, while minimizing the sum of the absolute values of the fluxes ($-\delta \leqslant \overrightarrow{v} \leqslant \delta,\ \delta > 0$). This formulation makes use of the L1-norm. The method actually aims to approximate a flux distribution that minimizes cellular investment into enzymes. As a side effect, the infeasible loops vanish, as stated by Desouki et al. [40, 70, 148]. The loopless COBRA algorithm by Schellenberger, Lewis, and Palsson [147], developed in 2011 involving mixed-integer programming, is another alternative to address inherent FBA limitations.

The usage of the FBA method enables great insights into the topology of a network. Nevertheless, it can only provide a limited validity in its predictions, because it does not usually take regulation into account (except rFBA – regulatory Flux Balance Analysis [51]) and moreover dismisses reactant concentrations and any form of gradients. Furthermore, because of the underdetermined nature of the stoichiometric matrix, a FBA solution is only one of many, so-called *degenerated*, solutions [161, 181]. There can be many more solutions, which satisfy the conditions but are difficult to predetermine. For this reason, further constraints are needed to deal with the general lack of knowledge on kinetic parameters and reaction mechanisms [90] and to restrict the potential solution space. A first approach would be the incorporation of the first law of thermodynamics, termed *energy conservation*. This law states that the overall driving force through internal cycles must be zero [68]. Furthermore, the second law of thermodynamics can be included as well [16, 132, 182]. In this way, it is possible to determine feasible reaction directions, and facilitate the introduction of metabolite concentrations [19], as demonstrated in the following papers: [15, 16, 69, 90, 133]. The procedure will be explained in the subsequent chapter.

3 THERMODYNAMICS

Most interactions in biological networks, such as diffusion, mass transport, protein-protein interactions and free energy of biotransformations, are governed or driven by thermodynamics, which reveals that its consideration in mathematical modeling is critical [162]. During the last 16 years, the application of thermodynamics with respect to entire metabolic networks has increased significantly. This is due to the fact that during this period the number of annotated genome sequences has grown tremendously, which made their examination possible in the first place [12]. Several approaches have been developed during that time. The most common are the network-embedded thermodynamic analysis (NET-analysis) [90, 184], the energy balance analysis (EBA) [16, 17] and the thermodynamics-based flux analysis (TFA) [68]. Each one will be explained in a short discourse, followed by a discussion concerning the method-choice with respect to this work.

The EBA approach by Beard and Qian, controls the directionality and the bounds of the fluxes due to the use of the second law of thermodynamics, which is stating that each internal reaction with non-zero flux must dissipate energy [15]. It moreover utilizes constraints for metabolite concentrations [12, 17]. In contrast, NET and TFA analysis, while constraining the fluxes through the use of the second law of thermodynamics as well, use the value of the Gibbs energy as a linear function of the logarithms of the metabolite concentrations (or activities) instead of a constant or continuous variable. They moreover value the necessity of adjusting the Gibbs energies to physiological conditions, like variabilities in pH or ionic strength, because not all metabolites appear in the same cell compartment and thus may be exposed to different conditions. In this respect, Maskow and Stockar [100] elaborate that feasible pathways could falsely be labeled as infeasible or vice versa without careful consideration of physiochemical conditions. In contrast to EBA, TFA and NET-analysis both consider the transmembrane ion transport, which makes them more comprehensive [74, 184]. However, the approaches are not completely equivalent: First, NET uses linear programming, while TFA uses mixed-integer linear programming (MILP); second, the NET approach requires a predetermination of the directionality of the fluxes, while TFA considers all fluxes as bidirectional. For this reason, Ataman and Hatzimanikatis [12] prefer the TFA approach and consider EBA and NET as special cases of TFA. Nevertheless, the NET approach has been the method of choice in the present work. Henry *et al* [68] and Ataman *et al* [12] named several reasons, why TFA should be preferred over NET but none of them applies here: Neither is the core-model incomplete, nor does it bother that the flux directions have to be predefined. On the contrary, it is in fact convenient, because the flux distributions of the FBA are to be tested for their thermodynamic feasibility, which involves their directionalities. Furthermore,

MILP is not needed, considering the size of the core-model. It can be analyzed easily via linear programming. Finally, because the possibilities of the network should be analyzed to the fullest, it seems reasonable to determine the feasible ranges of the Gibbs energies, as it is characteristic for the NET approach [90].

3.1 NETWORK-EMBEDDED THERMODYNAMIC ANALYSIS (NET-ANALYSIS)

As already mentioned, flux distributions FBA returns, will be tested for thermodynamic feasibility to be able to distinguish between biologically feasible and non-feasible FBA solutions and to gain insight into the control mechanisms of the cell. For this reason, it is very important to examine the network in its entirety, because the energies of the reactions are coupled.

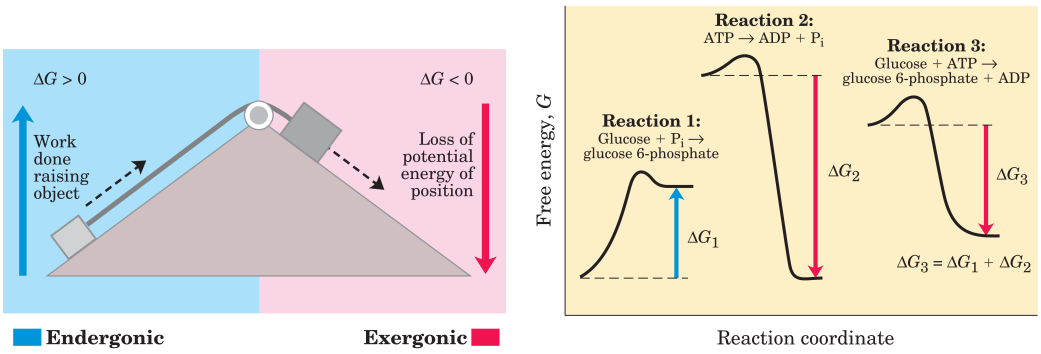


Figure 4: (a) displays how the downward motion of the bigger box releases potential energy that can do mechanical work. The potential energy, made available by spontaneous downward motion, is an exergonic process (pink) and can be coupled to the endergonic upward movement of a smaller box (blue). (b) shows the same principals with respect to chemical reactions: In reaction 1, the formation of glucose 6-phosphate yields a product of higher energy, than the two substrates. This is therefore an endergonic reaction and cannot react spontaneously. In reaction 2, the breakdown of ATP is highly exergonic and can therefore drive an endergonic reaction, when the two reactions are coupled, as can be seen in Reaction 3 [110].

To illustrate this point, consider the mechanic example in Figure 4 (a). At the top of the plane sits a box, which has a certain amount of potential energy, because of its elevation. It tends to slide downwards, which offers the possibility to couple it to a smaller box, which is needed to move upwards but cannot do it on its own. The energy available to carry out this task is the *Gibbs energy*, $\Delta_r G$. This concept can be applied to chemical reactions as well: Each reactant contains a certain amount of potential energy, related to the nature and number of its bounds. If a reaction occurs spontaneously, it releases Gibbs energy, if the substrates have a higher free energy as the products. Such a reaction is called *exergonic* and the decline in Gibbs energy from substrate to product has a negative value ($\Delta_r G < 0$). The alternative are *endergonic* reactions, which require an energy input to proceed ($\Delta_r G > 0$). As shown in Figure 4 (b), it is possible to couple endergonic and exergonic reactions to drive otherwise unfavorable re-

actions [110] Therefore, it is critical to consider the network as a whole, instead of single reactions, to make energy coupling possible [90].

Before the algorithm can be explained in its complexity, some basics have to be established. As mentioned above, a prerequisite for the NET-analysis is the knowledge of the directions of the fluxes. These directions can be gained from FBA with application of the signum function (in short *sign*) on the flux distribution vector, resulting in the so-called *signvector*. The signum function translates positive values to 1 and negative values to -1. Zero stays the same [19], see Figure 5 for an example.

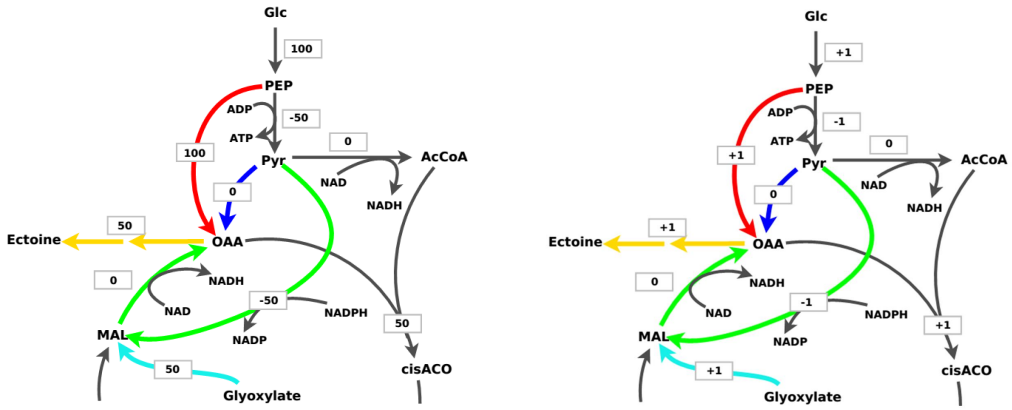


Figure 5: The signum function extracts the sign of any real number. Forward fluxes of (a) are translated to 1 and backward fluxes to -1 in (b).

Moreover, the compartment conditions need to be defined, that is the pH, ionic strength (abbreviated by I) and membrane potential (ΔV) in the different cell compartments, which will be needed to correct the formation Gibbs energies (see Section 3.1.1) later on.

The *Gibbs energy of a reaction j* can be calculated from the *Gibbs energies of formation* ($\Delta_f G$) of the species involved in the reaction:

$$\Delta_r G_j = \sum_i s_{ij} \Delta_f G_i \tag{5}$$

whereas s_{ij} are the stoichiometric coefficients of $\mathbf{S}$.

3.1.1 *Gibbs energy of formation*

Assuming that the metabolite concentration is spatially invariant and that temperature and pressure are constant, it is possible to define a single Gibbs energy of formation ($\Delta_f G$) for each metabolite species, which is valid throughout a single cell compartment [53].

The *Gibbs energy of formation of a species i* is defined as

$$\Delta_f G_i = \Delta_f G_i^\circ + R\,T\ln(a_i) \tag{6}$$

$\Delta_f G_i^\circ$ is called the *standard Gibbs energy of formation*. It is the Gibbs energy change, when a mole of the species in its standard state (in the ideal aqueous solution at 1 mol/l), is formed from its elements in their reference state. A superscript circle is used to designate a standard condition [1, 53]. R is the *universal gas constant* (in $J/mol\,K^{-1}$) and T is the *absolute temperature* (in K). a_i is the *activity of a species i*, and it's given by $a_i = \gamma_i x_i$, where γ_i is the *activity coefficient* and x_i the concentration. The activity coefficient accounts for the deviations from ideal behavior of the system [28]. Because equilibrium concentrations are easier to measure, than activities, a possible reformulation of Equation 6 is [100]:

$$\Delta_f G_i = \underbrace{\Delta_f G_i^\circ + R\,T\ln(x_i)}_{\text{ideal }\Delta_f G} + R\,T\ln(\gamma_i) \tag{7}$$

The first two terms form the Gibbs energy of formation under ideal conditions (ideal $\Delta_f G$), meaning in distilled water, that is $\gamma_i = 1$, $I = 0$. The last term, $R\,T\ln(\gamma_i)$, is the correction term of the Gibbs energy of formation for ionic strength. In general, one uses the extended Debye-Hückle term (s. Section 3.4) for γ_i, in combination with the common logarithm (base 10), but this term only applies for relatively low values of ionic strength. This issue will be explained in greater detail in Section 3.4. Equation 7 can be rewritten in a way that the standard Gibbs energy of formation ($\Delta_f G^\circ$) and the activity coefficient term ($R\,T\ln(\gamma)$) are combined in one term to $\Delta_f G_i^0$. The superscript zero accentuates that the standard Gibbs energy of formation is now a function of ionic strength [53]:

$$\Delta_f G_i = \Delta_f G_i^0 + R\,T\ln(x_i) \tag{8}$$

$\Delta_f G_i$ not only needs to be corrected for ionic strength but for the pH of the compartment as well, which demands a transformation of $\Delta_f G_i^0$. Since protons are then incorporated through the pH, they are not considered as reactants [74]. Using the extended Debye-Hückle equation (as described in see Section 3.4), the conversion factor $\ln 10$ ($\log \rightarrow \ln$) and a legendre transformation, one gains the *standard transformed Gibbs energy of formation* (the transformation is indicated by an additional accent) [1]:

$$\Delta_f G_i'^0 = \Delta_f G_i^0 + R\,T\,nH_i\,\log(10)\,pH - \frac{2.91482\,(z_i^2 - nH_i)\,\sqrt{I}}{1 + 1.6\,\sqrt{I}} \tag{9}$$

With nH_i being the number of protons and z_i being the charge of the species. $\Delta_f G'^0$ replaces $\Delta_f G_i^0$ in Equation 8 and one gains the formula for the *transformed Gibbs energy of formation* ($\Delta_f G'$).

So far, the term *species* has been used, instead of *reactant* or *metabolite*, because reactants are considered to be made up of sums of particular species, the so-called *pseudo isomers*. For example, ATP exists as several distinct species, which are assumed to be rapidly interconverted [1]:

$$ADP^{3-} \Leftrightarrow HADP^{2-} \Leftrightarrow H_2ADP^{-} \Leftrightarrow H_3ADP \tag{10}$$

Each one of this different species has its own $\Delta_f G_i^0$. Alberty measured several species but not quite enough for the core-model of *H. elongata* [72]. Certain reactions can be combined to eliminate some of the unknown reactants albeit not all. The standard Gibbs energies of formation have to be collected from somewhere else. The only other possibility is the use of estimated values, which have been provided by the *Group contribution* (GC) and *Component Contribution (CC) Method*, whereas the second method is an enhancement of the first [72, 114]. The GC Method deals with the molecular structures of the different species. It decomposes the species into a set of smaller molecular substructures, whose Gibbs energies of formation can be estimated. Afterwards, the species is composed again and the Gibbs energies of formation of the species is the sum over the Gibbs energies of formation of its substructures [72]:

$$\Delta_f G_{est}^0 = \sum_{i=1}^{N_{gr}} n_i \Delta_{gr} G_i^0 \tag{11}$$

$\Delta_f G_{est}^0$ is the estimated $\Delta_f G^0$, n_i is the number of instances of group i in the molecular structure and N_{gr} is the number of different groups for which $\Delta_{gr} G_i^0$ is known. Noor et al. [114] developed the CC Method after realizing that the error rate of the GC is too high. He therefore combined it with a second method, which he terms the *Reactant Contribution (RC) Method*. This method was developed 1957 by Burton [86]. It is based on experimental data and manages to calculate unknown energies out of known ones applying the first law of thermodynamics [114]. The CC Method splits each reaction into two independent reactions. One of these sub-reactions is calculated with RC, if enough data are available, the other is calculated using GC. They are combined subsequently giving the value of the RC Method more weight.

At this stage, all Gibbs energies of formation of the different species of a reactant are known and corrected for pH and ionic strength (see Equation 9), and one can finally calculate the *standard transformed Gibbs energy of formation of the reactant* k, respective its pseudo isomer group [1]:

$$\Delta_f G_k'^0 = -RT \log \left(\sum_{i=1}^{N_{iso}} \exp \left(-\frac{\Delta_f G_i'^0}{RT} \right) \right) \tag{12}$$

whereas N_{iso} is the number of different species of reactant k. This aggregation is feasible, because, when pseudo isomers are in equilibrium, they have the same chemical potential [1].

Only one more element is missing to be able to understand the algorithm, the consideration of the ion transport through the membrane. By default, metabolites that are present in different cell compartments cannot exchange unless a respective transport process is defined in the model [184]. In order to include the Gibbs energy of transport processes, two forces of the membrane have to be considered: the *proton motive force*, which accounts for the effect of the pH gradient of the membrane and the *membrane potential*, which accounts for the transfer of charged molecules from the inside of the cell to the outside ($\Rightarrow \Delta V < 0$). Where protons travel through the membrane, the Gibbs energy has to be corrected for the proton motive force, which releases energy, as well as the membrane potential, which costs energy. The transport of other charged molecules, like sodium, is determined by the membrane potential.

$$\begin{aligned} \text{Transport protons } (\Delta_{pH}G): &\quad s\,(R\,T\,\ln 10\,\Delta pH - z\,F\,\Delta V) \\ \text{Transport sodium } (\Delta_{V}G): &\quad -\,s\,z\,F\,\Delta V \end{aligned}$$

$$\Delta_r G'_{Transport} = \Delta_V G + \Delta_{pH} G \tag{13}$$

s represents the stoichiometric coefficient of the cation, while F is the Faraday constant, z stands for the charge, ΔpH is the difference in pH ($pH_{intracellular}$ - $pH_{extracellular}$) and ΔV ($V_{intracellular}$ - $V_{extracellular}$) represents the difference in electric potential. Corresponding to Henry, Broadbelt, and Hatzimanikatis [68], the ΔV of *E. coli* can be estimated as follows:

$$\Delta V = 33.33 \cdot \Delta pH - 143.33 \tag{14}$$

whereas the term 143.33 represents the effect of the different ions.

These terms have to be added to the Gibbs energy of a reaction, if transport-processes are taking place, as in the case of the ATPase.

3.1.2 *The algorithm*

Now, the actual algorithm of the NET-analysis will be considered. It is displayed in Equation 15. As already mentioned, the objective is to gain an idea of the ranges the Gibbs energy of each reaction can exhibit. To arrive at the boundary conditions, the Gibbs energy is minimized and maximized to get the feasible interval. As one can see in the second line, the transport is added to the (transformed) Gibbs energy. If no transport takes places, the value of $\Delta_r G'_{Transport}$ is zero. Because a certain reaction direction has to be tested for feasibility, the Gibbs energy has to be multiplied with the sign. The second law of thermodynamics demands for this product to be negative and $-\Delta_r G'$ is termed the *Driving force* of a reaction [102]. The third line corresponds to the transformed Gibbs energy of formation, which has been discussed in detail already. The last constraints are the boundaries for the concentrations of all

metabolites, which usually are fixed between 0.1 µmol/l and 10 mmol/l, with the exception of metabolites like H_2O [14, 68, 111].

$$\begin{aligned} \min/\max\ & \Delta_r G'_k \\ & (\Delta_r G'_j + \Delta_r G'_{Transport,j}) \cdot \mathrm{sign}(v_j) < 0 \\ & \Delta_r G'_j = \sum_i s_{ij} \Delta_f G'_i \\ \text{s.t:}\ & \\ & \Delta_f G'_i = \Delta_f G'^{0}_i + R\,T \ln(x_i) \\ & x_{min} \leqslant x_i \leqslant x_{max} \end{aligned} \tag{15}$$

These minimizations and maximizations of the Gibbs energies are done for each reaction. At first, they are done without enforcing any directions on the reactions (all signs are zero), which returns one set of intervals for the Gibbs energies of all reactions. The process is repeated with directions, as defined in the signvector, which returns a second set of intervals for the Gibbs energies of all reactions. The procedure results in two sets of intervals for each reaction. The intervals without enforced direction are naturally wider, because the system is less restricted.

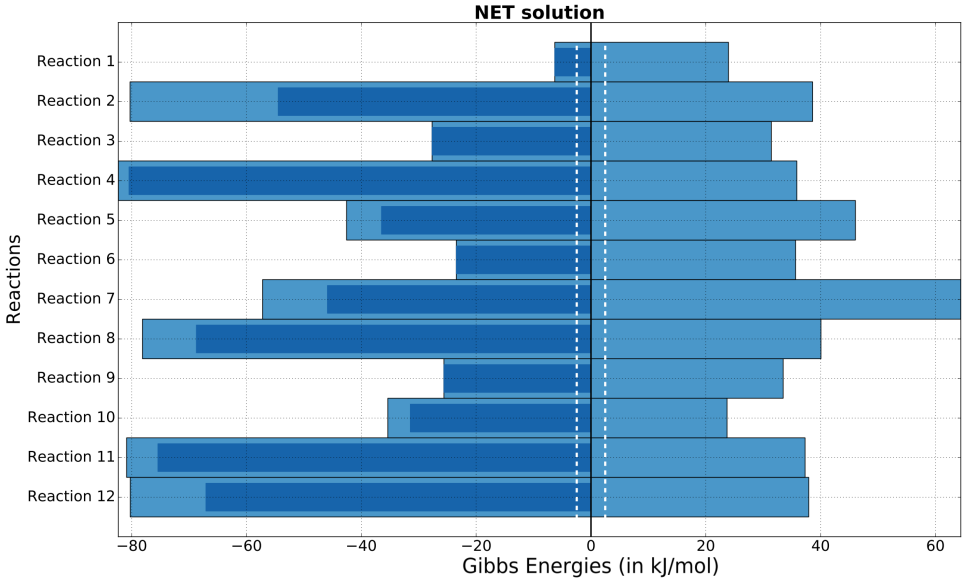


Figure 6: Gibbs energy intervals from the NET method, applied to the reactions of glycolysis, ordered from the import of glucose to lactate production. The light blue intervals represent the intervals in case that no reaction directions are given. The dark blue ones represent the intervals, which result, if all reactions are enforced to proceed in forward direction. The dashed white lines display the interval, which indicates closeness to thermodynamic equilibrium.

As an exemplary output, consider Figure 6. The light blue bars indicate the possible Gibbs energy intervals due to the provided concentration ranges. The darker blue bars display the thermodynamically feasible ranges after introducing reaction directions. It is obvious that all reactions were enforced to only

move in a forward direction (negative Gibbs energy). The dashed white lines indicate the closeness to thermodynamic equilibrium. The interval they describe is [-2.5; 2.5] kJ/mol [18].

If a reaction displays a Gibbs energy, which is not close to thermodynamic equilibrium ($|\Delta_r G'| > 2.5$ kJ mol^{-1}) and therefore has a significant driving force, it can be assumed that an active regulation of the corresponding enzyme has a significant effect on the pathway [68, 184].

In Figure 7 one can see the effect of reaction directions on the possible Gibbs energies of formation of the metabolites A, B, C and D. Because the energy through the network only flows from a higher to a lower level, the ranges are quite restricted [90].

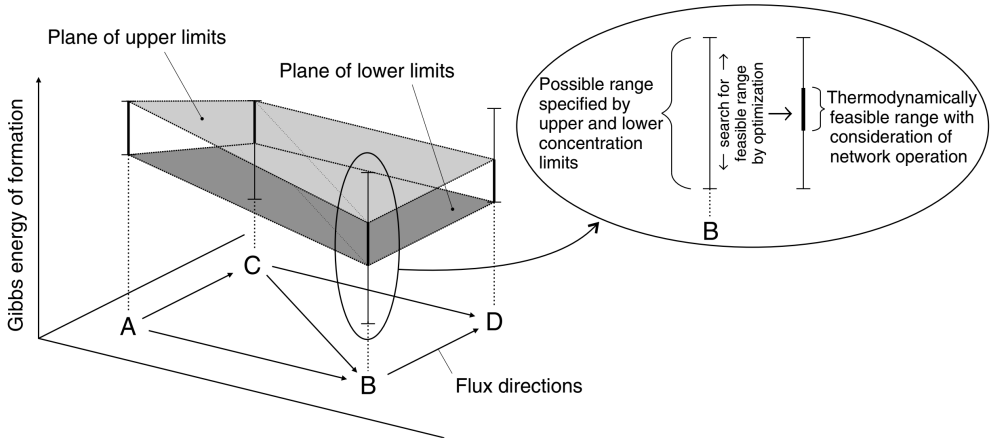


Figure 7: Illustration of the mutual thermodynamic interdependencies of reactions in a network. Possible ranges for the reactant's Gibbs energies of formation, taking into account only the concentration ranges of the metabolites (A-D), are shown with confined vertical bars. Owing to the provided flux directions and the cooperative action of the reactions in the network, however, the thermodynamically feasible ranges are smaller, which is highlighted by the bold parts of the bars. A flux can only flow from a higher to a lower level of Gibbs energy of formation. The space between the displayed planes consists of the thermodynamically feasible Gibbs energies of formation and, thus, describes the feasible concentration space [90].

The algorithm, as displayed in Equation 15, can also be used for the optimization of the logarithms of the metabolite concentrations. As before, two optimizations are executed, a minimization and a maximization but this time of each metabolite instead of each reaction. Therefore, a plot, as Figure 6, can also be drawn for the logarithms of the metabolite concentrations. It is even possible to let the optimization of the concentrations follow the optimization of the reactions, to be able to use the Gibbs energy intervals as additional constraints. In this way, one can couple both optimizations directly [90].

3.1.3 *Standard transformed Gibbs energy*

The *standard transformed Gibbs energy*, which is the Gibbs energy, when all concentrations equal 1 mol/l ($\Rightarrow \ln x = 0$), is given by [1]:

$$\begin{aligned}\Delta_r G_j'^0 &= \sum_i s_{ij} \Delta_f G_i'^0 + \Delta_r G'_{Transport,j} \\ &= -R\,T\,\ln(K'_{eq})\end{aligned} \tag{16}$$

It follows:

$$K'_{eq} = e^{-\frac{\Delta_r G_j'^0}{RT}} \tag{17}$$

K'_{eq} is the equilibrium constant, which represents the ratio between the concentrations of products and substrates in equilibrium [2]. Thus, in equilibrium, it is also defined as:

$$\prod_j |x_j|_0^{s_{i,j}} = K'_{eq,i} \tag{18}$$

3.2 MAX-MIN DRIVING FORCE (MDF)

The insight into the ranges of the Gibbs intervals and the metabolite concentrations is already an improvement of the analysis but an interesting question is, if a point estimate for a single solution of a given thermodynamic setting is feasible. Would it be possible to sample individual values out of the intervals for each reaction and each metabolite? It can be argued that this approach would neither be feasible, nor an optimal choice. Noor et al. [111] already provided a solution for this, although their motivation was a different one. They determined that, if a reaction reaches thermodynamic equilibrium ($|\Delta_r G'| \leqslant 2.5$ kJ mol^{-1} [18]), exponentially more enzyme counterproductively catalyzes the reverse reaction, since it applies that the Gibbs energy is proportional to the logarithm of the ratio of forward (v_f) and reverse flux (v_b) [18]:

$$\Delta G = -R\,T\,\log\frac{v_f}{v_b} \tag{19}$$

Therefore, being in thermodynamic equilibrium reduces the net rate at which the reaction proceeds. This results in the need for a higher enzyme level to keep the net flux up, which in turn is quite costly for the cell, since the production of additional enzymes costs ATP. Noor et al. therefore developed an algorithm, which shifts the Gibbs energies of the reactions as far away from thermodynamic equilibrium as possible by maximizing their driving forces. This method allows the ranking of different pathways for their cost efficiency and to determine the rate limiting reactions of a pathway. Moreover, it results in a thermodynamically feasible solution-vector and is therefore added to the analysis framework. This algorithm, which is termed *Max-min Driving Force*, will be explained in detail in the following.

In order to maximize the driving forces of all reactions, one uses the minimum over all reaction driving forces (namely B) as an optimization goal and maximizes it, within the predefined concentration limits, which shifts it as far away from equilibrium as possible. The algorithm is also based on linear programming, see Equation 20.

$$
\begin{aligned}
&\max_{x,B} B \\
&\quad \text{s.t:} \\
&\qquad -\operatorname{sign}(\overrightarrow{v}) \cdot (\overrightarrow{\Delta_r G}'^{0} + R\,T \cdot \mathbf{S}^T \cdot \ln \overrightarrow{x}) \geqslant B \\
&\qquad x_{min} \leqslant x_i \leqslant x_{max} \\
&\qquad B \text{ free}
\end{aligned}
\tag{20}
$$

The NET (Equation 15) and the MDF algorithm are directly related, since it applies that (see Equation 16): $\overrightarrow{\Delta_r G}'^{0} + R\,T \cdot \mathbf{S}^T \cdot \ln \overrightarrow{x} = \overrightarrow{\Delta_r G}' + \overrightarrow{\Delta_r G}'_{Transport}$

Therefore, all solutions of the MDF algorithm are embedded in the intervals calculated by NET-analysis, as can be seen in Figure 8.

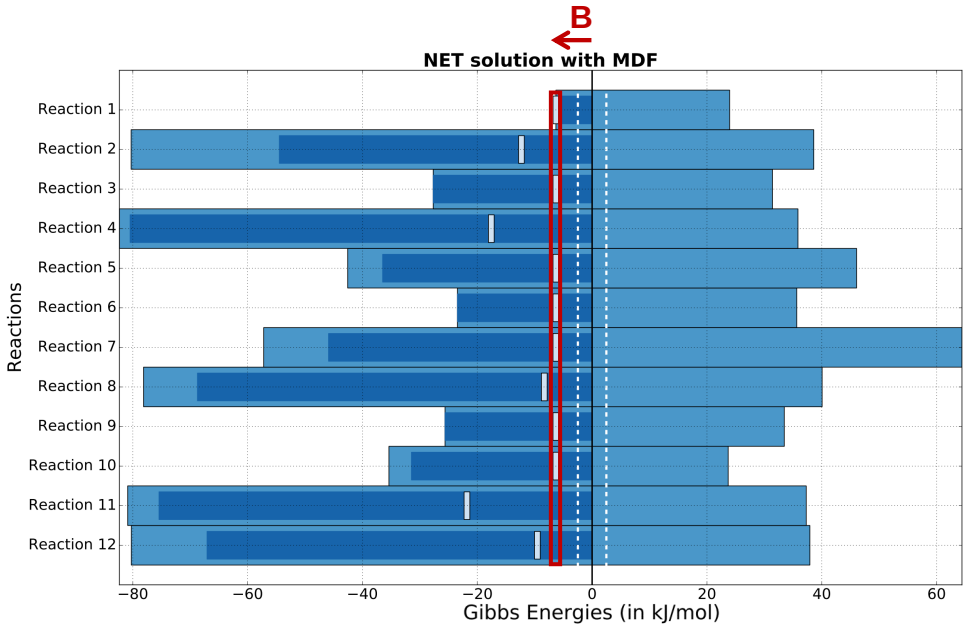


Figure 8: The figure displays a solution of the NET (as before in Figure 6) and the MDF algorithm. The small white bars are the MDF solution. The smallest distance to the equilibrium is the optimal minimal driving force (B), indicated by the red bar.

In case of the MDF algorithm only one optimization is done, instead of several for each reaction and each metabolite. It results in a solution vector, which consists of Gibbs energy values for all reactions and concentrations for all metabolites. A solution for the Gibbs energies with the MDF algorithm can be found in Figure 8. The small white bars are the Gibbs energies, which result from the maximization of B. The value of B represents the smallest distance from the thermodynamic equilibrium, illustrated through the red bar.

Another way to display B can be seen in Figure 9 (a). B is the *minimal driving force* of the pathway and therefore the smallest step in the energy-flow diagram, it identifies the bottleneck-reactions of a network. The enzymes, which catalyze these reactions, will constrain the activity of the pathway, unless they are present either in high concentrations or especially fast catalysts [111].

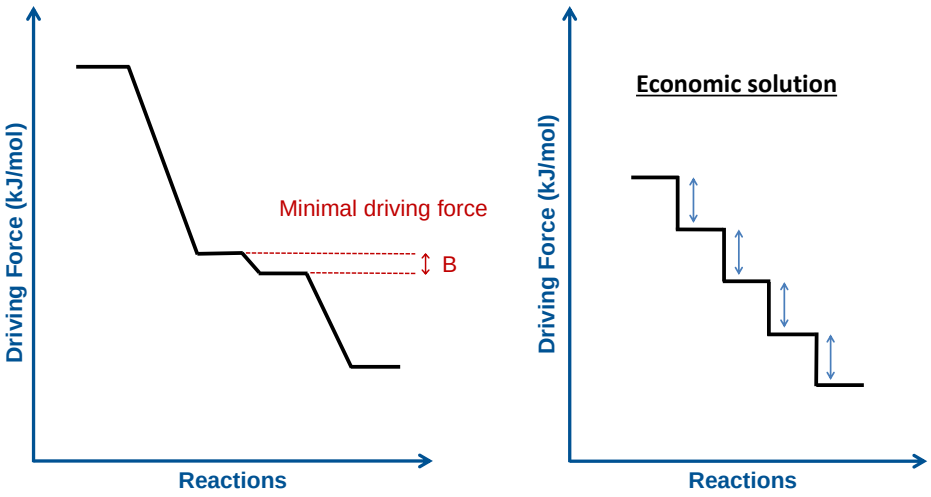


Figure 9: (a) shows that B represents the bottleneck of a pathway, because it identifies the reaction with the lowest driving force. (b) displays a MDF solution with wide boundaries concerning concentrations. The differences in driving force are more or less the same in each step. This solution is therefore weighted as the most economic, because of the lowest possible enzyme investment.

Another finding of the group Elad Noor's was that, if the concentration limits are reasonable wide, the solution of the MDF algorithm results in Gibbs energies, which are quite similar to one another (see the supplementary information of [111]), as displayed in Figure 9 (b). If the limits are narrow, one gets a solution as in (a), where some bottlenecks exist with low Gibbs energies and some reactions are highly exergonic.

Figure 10 displays an example of a small linear pathway. $\Delta_r G'^0$ is assumed to be zero for both reactions and the reactions are irreversible forward for the sake of simplicity. The application of the MDF algorithm, under the assumption that the concentration intervals are [1; 10] kJ/mol, produces two inequalities, which need to be solved. In order to maximize B, X_3 needs to have the lowest concentration possible, therefore $X_3 = 1$ kJ/mol and X_1 needs to be as high as possible and therefore 10 kJ/mol. The optimal concentration for B lies beneath both curves in Figure 10 and is maximal, where both curves intersect. Therefore, the Gibbs energies of both reactions are equal.

The finding of the uniform distribution of $\Delta_r G$ can be associated with the design principle of *metabolic simplicity* [103, 104], which has been confirmed to play a role in metabolic pathway architecture [112]. This principle indicates that a metabolic conversion will be carried out by using the shortest pathway possi-

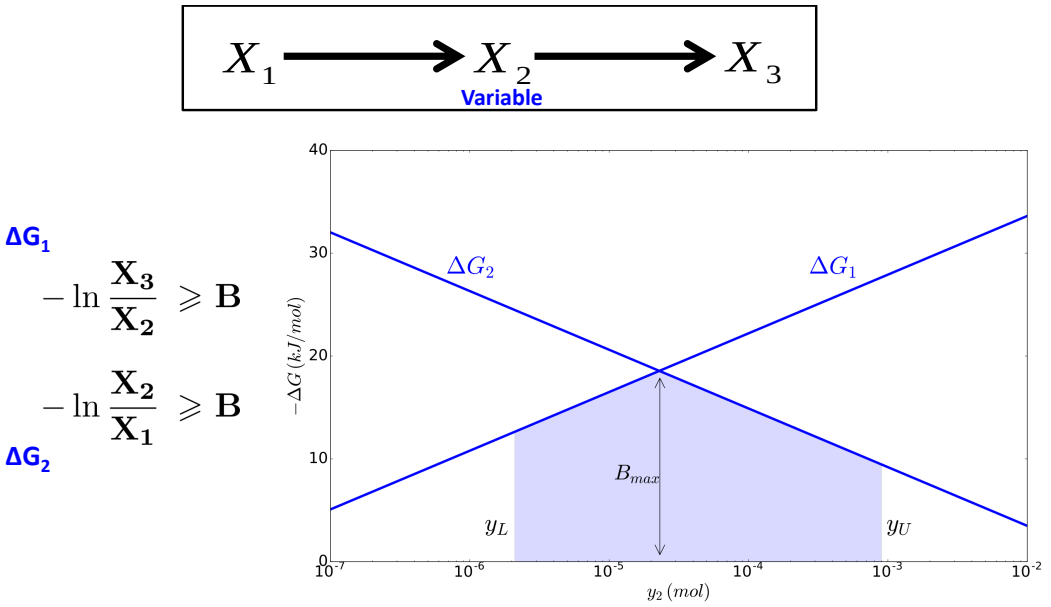


Figure 10: The figure displays the application of the MDF method on a linear pathway with only two reactions. This results in the equalization of two inequalities in order to calculate the maximal B and therefore the Gibbs energies. The calculation shows that the optimal solution can be found, if both Gibbs energies are exactly the same, as long as the concentration limits of X_2 are not violated (indicated through the blue area). y_2 is the logarithm of X_2.

ble. The justification is that two pathways, which perform the same conversion, exhibit the same overall Gibbs energy. The shorter pathway will have higher driving forces in each step and therefore needs less enzyme to keep up a high flux, which is more energy efficient.

The MDF principle will now be applied to a real life example, the ammonia assimilation.

Ammonia assimilation

Ammonia assimilation in bacteria is known to be mediated mainly by two mechanisms: First, the reductive amination of α-ketoglutaric acid to yield glutamate, catalyzed by glutamate dehydrogenase (GDH) and, second, formation of glutamine from glutamate and ammonium, catalyzed by glutamine synthetase (GS), followed by the transfer of the amidegroup to α-ketoglutaric acid, catalysed by glutamate synthase (GOGAT), which results in the net-synthesis of one glutamate molecule. The sum of the GS/GOGAT system corresponds to the GDH equation with the addition of the hydolysis of one ATP, see Figure 11.

Some bacteria seem to possess only the second mechanism, others, like *E. coli*, exhibit both [23, 65]. In order to find out which pathway is superior, one could apply FBA and would get that it is the GDH-pathway, because it saves ATP for other processes. The metabolic simplicity principle would lead in the same direction but it has been shown that *E. coli*, e.g., prefers the GS/GOGAT system.

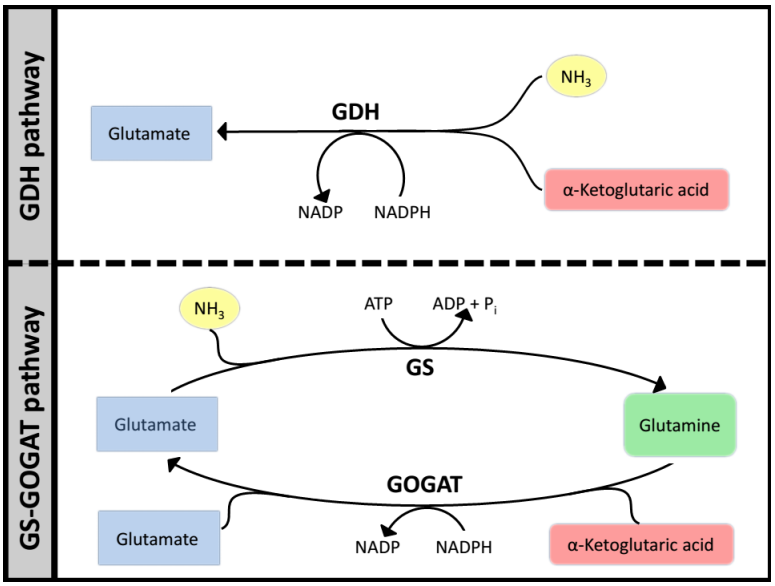


Figure 11: The two typical systems for ammonia assimilation. The GDH pathway needs only one enzyme, the GS/GOGAT system needs two and moreover hydrolyzes one ATP.

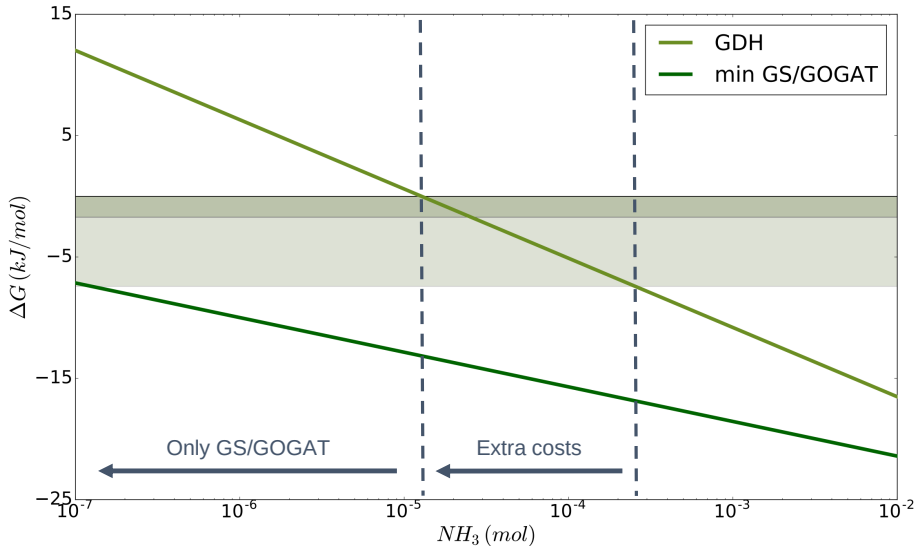


Figure 12: The figure can be divided in three parts. In the first part, GDH is thermodynamically unable to carry any flux in forward direction in contrast to the GS/GOGAT system. In the second part, GDH can carry flux but with very high costs, at least as long as it is in the dark green part, which basically means the reaction is close to equilibrium. The lighter green part allows for it to carry flux with less cost but still is close to equilibrium and therefore still the minor choice. Eventually both pathways converge and are both good choices for ammonia assimilation.

The results of the application of the MDF method are displayed in Figure 12. Detailed calculations can be found in [155].

The figure shows the Gibbs energies of the two pathways in dependency on the ammonia concentration (in mol/l). The darker green line is the sum of both Gibbs energies of the GS/GOGAT system, the other line stands for the Gibbs energy of the GDH pathway. The figure can be divided in three parts. In

the first part, for low ammonia concentrations, GDH exhibits a positive Gibbs energy and is therefore not able to execute the forward direction, whereas the GS/GOGAT system performs the conversion easily. In the second part of the figure, GDH is able to proceed in forward direction but the reaction is still quite close to thermodynamic equilibrium, which means high enzyme costs. Therefore, it is still better for the cell to use the GS/GOGAT system. Both curves converge at high ammonia concentrations, which allows for the GDH to be used. Therefore, the GS/GOGAT system is a clear adaption to environments, which are poor in ammonia.

3.3 DRIVING FORCE VS. CATALYTIC EFFICIENCY OF ENZYMES

In order to clarify the role of thermodynamics and the importance of driving forces, the example of a reversible Michaelis Menten reaction will be analyzed, showing how the driving force can drastically change the efficiency of enzyme catalysis.

There are different ways of describing the rate law ($\nu(\overrightarrow{x})$) of a reaction, like the Michaelis Menten description of irreversible enzymatic reactions, which has been introduced by Victor Henri [67] and rationalized by, among others, Michaelis and Menten in 1913 [105]. The irreversible rate law has been further developed by Haldane to describe reversible reactions [33]: *reversible Michaelis Menten kinetics* [113]. Haldane already noticed an inherent dependency between the kinetic parameters of a reaction and its thermodynamics, as stated by Noor et al. In [113] they use this insight to rewrite the reversible Michaelis Menten kinetics in a way, which clearly shows this dependency.

The original reversible Michaelis Menten kinetics equation describes the following mechanism:

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_4}{\overset{k_3}{\rightleftharpoons}} EP \underset{k_6}{\overset{k_5}{\rightleftharpoons}} E + P \tag{21}$$

E is the amount of free enzyme, S and P represent substrate and product concentration and ES and EP are the enzymes, which carry substrate or product.

By assuming steady state of intermediate complexes, one can calculate the appropriate rate law:

$$v = E\frac{k_{cat}^{+}\frac{S}{K_S} - k_{cat}^{-}\frac{P}{K_P}}{1 + \frac{S}{K_S} + \frac{P}{K_P}} = E\,k_{cat}^{+}\frac{\frac{S}{K_S}}{1 + \frac{S}{K_S} + \frac{P}{K_P}}\left(1 - \frac{P}{S}\frac{k_{cat}^{-}/K_P}{k_{cat}^{+}/K_S}\right) \tag{22}$$

whereas the k_{cat} values are the maximal forward and backward rates per unit of enzyme and K_S and K_P are the Michaelis constants [113]. Under the assumption that a reaction has reached equilibrium, the relation $\frac{k_{cat}^{+}/K_S}{k_{cat}^{-}/K_P}$ is called *Haldane relationship*, which states:

$$\frac{k_{cat}^{+}/K_S}{k_{cat}^{-}/K_P} = K'_{eq} \tag{23}$$

K'_{eq} is a thermodynamic constant, called the *equilibrium constant*, which has already been defined. Because of the definition of the Gibbs energy and Equation 16, it applies that [113]:

$$\begin{aligned} v &= E\, k^+_{cat} \cdot \frac{\frac{S}{K_S}}{1 + \frac{S}{K_S} + \frac{P}{K_P}} \cdot \underbrace{\left(1 - e^{\frac{\Delta_r G'}{RT}}\right)}_{1-\theta} \\ &= \underbrace{k^+_{cat}\, E\,(1-\theta)}_{\text{effective enzyme}} \cdot \frac{\frac{S}{K_S}}{1 + \frac{S}{K_S} + \frac{P}{K_P}} \end{aligned} \tag{24}$$

θ denotes the *thermodynamic term*, which describes the distance to equilibrium ($\theta = e^{\Delta_r G'^0/RT+\ln(P/S)}$). For a reaction ν_i with j reactants it therefore applies [27, 122]:

$$\prod_j |x_j|_0^{s_{i,j}} \;=\; \theta_i\, K'_{eq,i} \tag{25}$$

which is a reformulation of Equation 18. Equation 18 can be seen as the special case in which equilibrium is reached and θ equals one. $(1-\theta)$ is bounded between 0 and 1. The rate law is written in the direction $S \rightarrow P$, but can easily be written in the reverse direction.

Even without knowledge of the concrete mechanisms of the reaction or the numerical values of the parameters, much can already be understood from Equation 24. The product $k^+_{cat}\, E\,(1-\theta)$ describes the fraction of the enzyme, which actually catalyzes productively. It therefore establishes the minimal amount of enzyme that can sustain a certain flux [52]. The effective enzyme ($k^+_{cat}\, E\,(1-\theta)$) increases monotonically but not linearly with distance to equilibrium, compare Figure 13. The notation ΔG_x indicates the Gibbs energy at which the enzyme can operate up to x% of its V_{max}.

It can clearly be seen that a driving force of more or less 8 kJ/mol is already enough for an enzyme to operate close to 100% of its V_{max}, which signifies that a minimal driving force of 8 kJ/mol with the MDF algorithm is sufficient to ensure an efficient pathway [155].

Considering Equation 24 once more, it is obvious that the reversible Michaelis Menten equation is not independent of the product concentration, in contrast to the irreversible one. However, it can become independent under certain conditions: Assuming the driving force of a reaction is very high and the backward flux therefore negligible, results in $\theta \approx 0$, but the rate law still depends on the product concentration in the denominator of the κ-term. Only if $P \ll K_p$ holds, are the reversible and irreversible Michaelis Menten kinetics equivalent [113].

3.4 IONIC STRENGTH (SIT)

As already mentioned before, the Gibbs energies of formation of each species have to be corrected for ionic strength to make sure the Gibbs energy for the

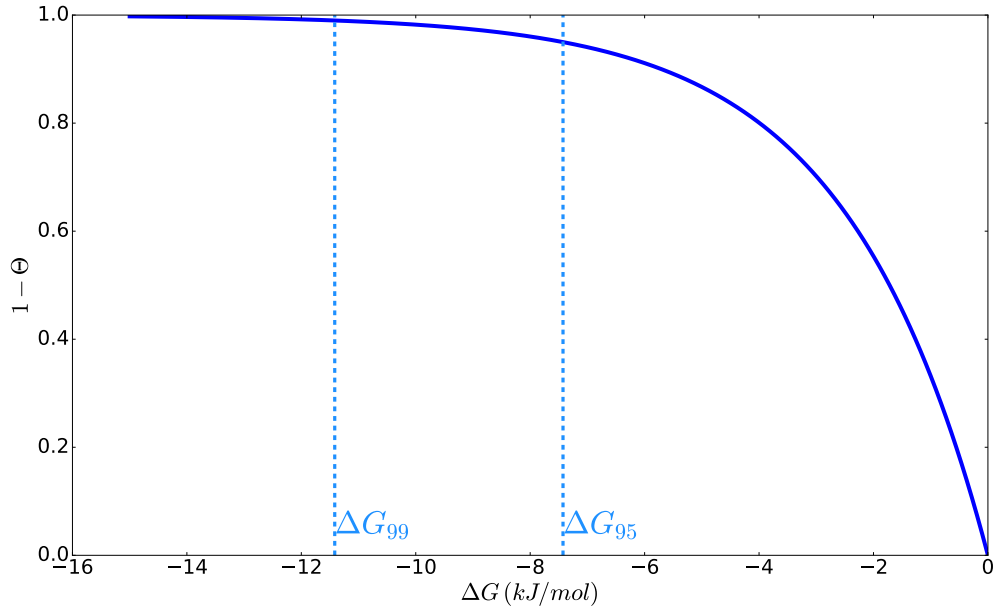


Figure 13: When ΔG is similar enough to RT (typically assumed to be 2.5 kJ/mol for biochemical reactions), $1-\theta$ decreases rapidly, since $\theta = exp(\frac{\Delta_r G'}{RT})$. $\Delta G_{95\%}$ and $\Delta G_{99\%}$ are the affinities at which the enzyme can operate at 95% and 99% of its V_{max} respectively.

overall reaction is corrected as well, because the interactions of the molecules and ions in a solution have to be accounted for. This correction can have many different appearances. Debye and Hückle in 1923 were the first to define such a correction term, which suggested that the forces, which cause deviations from ideality of solutions are due to long range interionic attractions between ions of different charge [28, 36–38]. Ions coordinate themselves in solution and build clouds of molecules of contrary charge around themselves, which brings more stability. Debye and Hückle therefore developed an equation, which relates the activity coefficient of a salt with the ionic strength and ionic charge (z), although it can only be applied for very dilute solutions ($I < 0.001$ mol/l [64]) and is independent of the chemical nature of the ions and their size. The correction term, in dependence on the ionic strength (I), is defined as:

$$-\log\gamma = A\,z^2\sqrt{I} \tag{26}$$

A is a function of temperature and equals 0.509 ($\sqrt{\frac{kg}{mol}}$) at 25°C. In most processes and procedures much higher concentrations of electrolytes are required. Therefore, the Debye-Hückle equation was refined and became the *extended Debye-Hückle equation*, which includes an ion size parameter:

$$-\log\gamma = \frac{A\,z^2\sqrt{I}}{1+B\,a\sqrt{I}} \tag{27}$$

B is another constant and a is the sum of the radii of the two ions of the electrolyte, which has to be treated as an empirical parameter [28]. If the solution contains more than one electrolyte it is thermodynamically inconsistent to

use this equation for calculating activity coefficients [62]. Besides, it can only be applied, if the ionic strength of the solution lies in-between 0.05 and 0.25 mol/l. Higher ionic strengths, as presented in the habitat of *H. elongata*, require a more complex approach.

Higher ionic strengths are covered with more sophisticated equations, like the one developed by Pitzer [123–130, 159, 160], whose motivation was to be able to predict activity coefficients in multicomponent electrolyte solutions. The basis was the extended Debye-Hückle equation, which was completed by terms, which represent short range interactions, as well as triple ion interactions [3, 28]. Still, it cannot be used to describe the behavior of systems involving nonionic substances, because it does not take into account interactions between electrons of the ions leading to chemical bonding between them [28].

Alternatively, one can use the *Specific Ion Interaction Theory (SIT)* [3, 59], which is equivalent to Pitzer [59, 60] but differs in the form of the denominator in the Debye-Hückle term.

An extensive comparison of the SIT and Pitzer formulations of the ion interaction approach can be found in references [45, 60], as well as in chapter 9 of [3]. The key problem, when using Pitzer for ion-complexes, is to determine the concentration dependence of the coefficient for short range forces. This requires experimental data at low ionic strength.

The SIT approach has clear advantages, as outlined in [**allard1997av**]:

- Robustness in analyzing experimental data from diverse sources
- Capacity to provide good estimates of activity coefficients
- Simplicity of use
- Possibility, by using charge/ion size correlations, to estimate unknown values for its parameters, the so-called ion interaction coefficients.

The different approaches for the calculation of the activity coefficient of NaCl can be found in Figure 14.

The formula for the SIT approach is the following:

$$\log(\gamma_j) \;=\; -\frac{A\, z_j^2\, \sqrt{I}}{1 + 1.5\,\sqrt{I}} + \sum_k \epsilon(j,k,I)\, m_k \tag{28}$$

In the first approximation of the virial expansion, only one interaction term is used to describe the non-electrostatic interactions between two species; the interaction parameters (ϵ) are considered to be independent of the ionic strength and for a given pair of ions to be independent of the presence of other ions. This is the basis for the Specific Ion Interaction Theory developed by Bronsted, Guggenheim and Scatchard. The specific ion interaction parameters are empirical, and have to be determined experimentally, either from activity coefficient data or from equilibrium constants [59]. m_k are the molalities. The value 1.5 $kg^{1/2}{\cdot}mol^{-1/2}$ was proposed by Scatchard [146]. It was found that this value

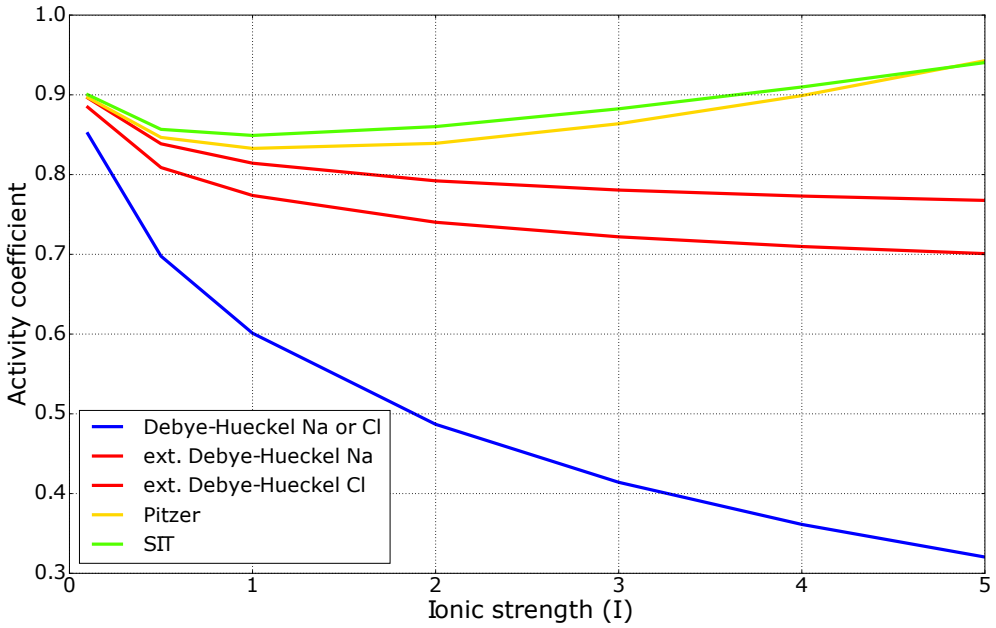


Figure 14: A comparison of Debye-Hückel, extended Debye-Hückel, SIT and Pitzer approaches concerning NaCl with varying ionic strength (in mol/l). One can see the similarity of SIT and Pitzer, as mentioned in [59, 60].

minimizes the ionic strength dependence of ϵ between I = 0.5 and 3.5 mol/l. It is an empirical parameter, which is correlated to ϵ.

Two main assumptions form the basis of SIT [**allard1997av**]: 1. The ionic medium ions will make the main contribution to the value of $\log \gamma_j$ for the reactions ions. This often makes it possible to simplify the summation, so that only ion interaction coefficients between the substrates/products in the chemical reactions and the ionic medium ions are included.
2. According to the Bronsted principle of specific interactions, the ion interaction coefficients ϵ are expected to be small for ions of the same charge sign. Similarly, they should be small for interactions between uncharged species and electro-neutral combinations of ions [24]. The rationale behind this is that ions of the same charge sign are usually separated from each other due to electrostatic repulsion and therefore ϵ would be small.

For the sake of simplicity, it is convenient to assume the interaction coefficients for uncharged species to be zero [61, 71], which generally fits (CO_2 is an exception [3]).

Another important variable is the *activity of water*: Equilibria involving H_2O as a reactant or product require a correction for the activity of water, a_{H_2O}:

$$\log a_{H_2O} = -\frac{\phi M_{H_2O} \sum_k m_k}{\ln 10} \tag{29}$$

ϕ is the osmotic coefficient, whose values for single electrolytes have been compiled by various authors, like [136]. In mixed electrolytes with several components at high concentrations, it may be necessary to use Pitzer's equation. A

table of the different activities of water, depending on the concentration, can be found in [59]. In seawater a correction is most likely and a sufficiently good approximation may be obtained by considering NaCl as the dominant electrolyte.

4 DYNAMICS

The constraint-based analysis tools presented so far (FBA, thermodynamics) have a number of significant advantages. First, they are linear, which makes them easy to handle even in case of large systems. Second, the amount of information needed is reasonably small [155]. Still, they draw an incomplete picture of the metabolism, because they omit regulation, which is a key feature in metabolic networks. The tool kit has to be completed by the inclusion of an analysis tool for dynamics, which allows the introduction of kinetics into the model. This is not trivial, the components of biochemical networks often cooperate in ill-characterized ways, almost always containing nonlinearities [176]. Since it is not possible to find an exact mathematical description of biochemical processes, approximations have to suffice [175]. The idea is to approximate the rate law. One could try a linear approximation but this solution would not be valid for a wide range of concentration values. A better choice are nonlinear approximations, which are in fact valid over a much wider range. In order to be useful, they must be based on the essential nonlinearity of the rate law but still be simple enough to allow mathematical treatment [144]. Common nonlinear approximations of the rate law, which are used frequently, are the Michaelis Menten and the Hill equation. Since they are quite complex, they are inpracticable with respect to larger reaction systems. The information necessary to describe a mathematical model using this type of equations is often unavailable or incomplete. It would be desirable to use a method, which helps to identify key regulatory interactions, without being tied to a certain choice of function shapes and parameter values [94]. For instance, for Michaelis Menten-like reactions it is not enough to know which metabolite or effector inhibits a particular reaction but also what type of inhibition is present [175]. A framework, which is easy to apply, is the Biochemical systems theory (BST), which has been proposed by Michael Savageau in 1969 [140–142]. This framework gets by with qualitative knowledge about variables and the reactions they influence. Its basics will be explained in the following section.

4.1 BIOCHEMICAL SYSTEMS THEORY

BST is based on nonlinear approximations of the rate law, which can be achieved by the linear approximation of reaction kinetics in logarithmic coordinates. Reaction kinetics describe an instantaneous temporal change in concentration of substrate or product [175]. Taking the logarithm of the rate law allows for an approximation, which is easy to treat mathematically, because it generates straight

lines and therefore retains the nonlinear properties [87, 144]. The following Taylor approximation around an operating point x_0 results in:

$$\begin{aligned} \log(v) &= \log(v_0) + \left.\frac{\partial \log(v)}{\partial \log(x)}\right|_0 \cdot (\log(x) - \log(x_0)) \\ &= f \cdot \log(x) + \log(\gamma) \end{aligned} \tag{30}$$

whereas f describes the slope of the rate (v) vs. the corresponding variable (x) in a log-log plot and γ represents the intercept. For an illustration, see Figure 15.

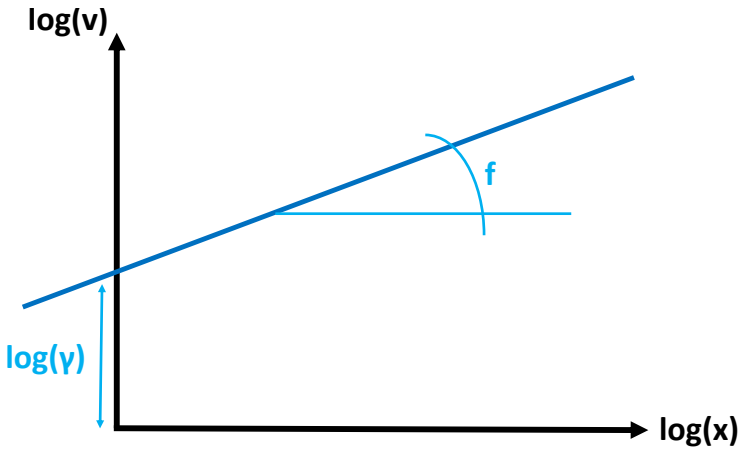


Figure 15: Illustration of the relationship of γ and f in a log-log plot

From Equation 30 follows:

$$\log(\gamma) = \log(v_0) - f \cdot \log(x_0) \tag{31}$$

In rectangular coordinates Equation 30 represents a power law, approximating the rate law at the operating value of the relevant concentration [144]:

$$v = \gamma \cdot x^f \tag{32}$$

γ is the apparent *rate constant* for the reaction, which can be positive or negative, depending on the flux: If it is a synthesis reaction, $\gamma_{i,j}$ would be positive, if it is a degradation reaction, $\gamma_{i,j}$ is negative. f is the apparent *kinetic order* with respect to x. In general, the rate law depends on several concentration variables, which can still be approximated and described by a power law:

$$\begin{aligned} v &= \gamma\, x_1^{f_1}\, x_2^{f_2} \cdots x_n^{f_n} \\ &= \gamma \prod_{i=1}^{n} x_i^{f_i} \end{aligned} \tag{33}$$

Savageau [144] stated that the curvilinear nature of a power law function makes it a perfect candidate for nonlinear approximations, because it conforms to the actual rate law over a much wider range of concentration values.

As illustrated by Equation 1 in Chapter 2, the change in concentration can be described as a linear combination of fluxes that either produce or consume

each component. These fluxes can be replaced by the power law approximation of the rate laws, which yields:

$$\begin{aligned}\dot{x}_i &= \gamma_{i,1}\prod_{j=1}^{n} x_j^{f_{i,j,1}} + \gamma_{i,2}\prod_{j=1}^{n} x_j^{f_{i,j,2}} + \ldots + \gamma_{i,k}\prod_{j=1}^{n} x_j^{f_{i,j,k}}, \qquad \text{for } i = 1,\ldots,m \\ &= \sum_{k=1}^{P_i} \alpha_{i,k}\prod_{j=1}^{n} x_j^{g_{i,j,k}} - \sum_{k=1}^{N_i} \beta_{i,k}\prod_{j=1}^{n} x_j^{h_{i,j,k}}, \qquad \text{for } i = 1,\ldots,m \end{aligned} \tag{34}$$

A system of this form is called a *generalized mass action (GMA-) system*. All GMA-system equations have the same mathematical form but differ in their parameters. Thus, Equation 1 is a GMA-system as well.

Because the fluxes can be divided into production and degradation processes, one can use $\alpha_{i,k}$ and $\beta_{i,k}$ as the production and depletion rate constants for such a classification, instead of the $\gamma_{i,j}$. The kinetic orders $f_{i,j,k}$ are also renamed into $g_{i,j,k}$ and $h_{i,j,k}$, consistent with the classification. Therefore, the parameters $\alpha_{i,k}$ and $g_{i,j}$ always refer to production or synthesis, whereas $\beta_{i,k}$ and $h_{i,j}$ always refer to degradation or loss. The first subscript (i) represents the affected variable, the second (j) stands for the effecting variable. If a kinetic order is positive, it indicates an activating effect of x_j on x_i, if it is negative, it represents an inhibiting effect. If it is zero, x_j does not have any effect [175]. The number of influxes into x_i is P_i, the number of effluxes if N_i. Each one of the n metabolites of Equation 1 can be written in such a way.

Another kind of system, which uses power laws as well, but only one degradation and one depletion term per metabolite, is called an *S-system*. The S stands for synergism:

$$\dot{x}_i = \alpha_i \prod_{j=1}^{n} x_j^{g_{i,j}} - \beta_i \prod_{j=1}^{n} x_j^{h_{i,j}}, \qquad i = 1,2,\ldots,m \tag{35}$$

The positive term is an aggregation of all synthesis fluxes, while the negative one aggregates all degradation fluxes [87]. In case of a GMA model, the fluxes are not aggregated together but each flux is approximates separately [176]. Although the functional representations at branch points are different in the GMA and S-system forms, the dynamics of the resulting models are often rather similar [176] and they are exactly the same at the operating point [175].

In biochemical system theory, one distinguishes three main entities: The *dependent variables*, representing components, which are affected by the system and whose values can change (e.g. intermediates or products); furthermore, *independent variables*, which represent unaffected components, such as enzymes. They are typically assumed constant during an experiment; and last but not least, *parameters*, such as kinetic orders and rate constants, which reflect the structure of a system and are constant in general. The distinction between dependent and independent variables always depends on the focus of the analysis. The term independent variable normally has a different meaning in mathematics but since the only independent variable in mathematical sense is time, there

is no ambiguity in reusing the term for all the variables, which are not affected by the system, that is, which are independent of it [175].

A pathway composed of many metabolites is represented by a collection of differential equations for the dependent variables. The independent variables stay constant. Still, they take part in the equation system, because they may affect the dependent variables. An example system, which will play a role in the subsequent chapter, is displayed in Figure 16. It consists of three dependent and two independent metabolites, four reactions and one inhibition (represented by $x_3^{g_{1,3}}$). In this case the S- and GMA-system representation are exactly the same, because each dependent metabolite has only one in- and one efflux, as represented in Equation 36.

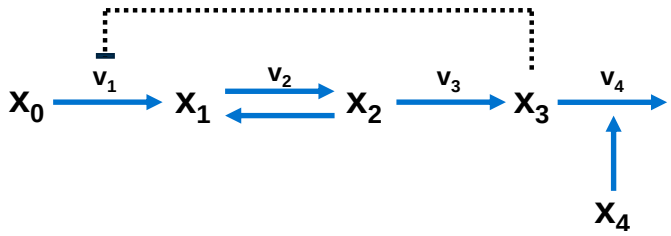


Figure 16: BST example with five metabolites, four reactions and one inhibition.

$$\begin{aligned} \dot{x}_1 &= v_1 - v_2 &&= \alpha_1 x_0^{g_{1,0}} x_3^{g_{1,3}} - \beta_2 x_1^{h_{2,1}} x_2^{h_{2,2}} \\ \dot{x}_2 &= v_2 - v_3 &&= \alpha_2 x_1^{g_{2,1}} x_2^{g_{2,2}} - \beta_3 x_2^{h_{3,2}} \\ \dot{x}_3 &= v_3 - v_4 &&= \alpha_3 x_2^{g_{3,2}} - \beta_4 x_3^{h_{3,4}} x_4^{h_{4,4}} \end{aligned} \tag{36}$$

Studying the map of the pathway shows that the degradation of x_1 constitutes the same process as the synthesis of x_2, although the terms $\beta_2 x_1^{h_{2,1}} x_2^{h_{2,2}}$ and $\alpha_2 x_1^{g_{2,1}} x_2^{g_{2,2}}$ appear to be quite different. The same applies for the conversion from x_2 to x_3. The *precursor-product relationships*, which describe the equivalence of such terms, impose constraints on the choice of the parameter values. It therefore has to apply that $\alpha_2 = \beta_2$, $\alpha_3 = \beta_3$, $h_{2,1} = g_{2,1}$, $h_{2,2} = g_{2,2}$ and $h_{3,2} = g_{3,2}$ [175].

Therefore Equation 36 can be rewritten to:

$$\begin{aligned} \dot{x}_1 &= \alpha_1 x_0^{g_{1,0}} x_3^{g_{1,3}} - \beta_2 x_1^{h_{2,1}} x_2^{h_{2,2}} \\ \dot{x}_2 &= \beta_2 x_1^{h_{2,1}} x_2^{h_{2,2}} - \beta_3 x_2^{h_{3,2}} \\ \dot{x}_3 &= \beta_3 x_2^{h_{3,2}} - \beta_4 x_3^{h_{3,4}} x_4^{h_{4,4}} \end{aligned} \tag{37}$$

Connection between BST and thermodynamics

As mentioned above, there are different ways of describing the rate law of a reaction. In addition to the power law, other rate equations exist, e.g. the Michaelis Menten description of irreversible enzymatic reactions, as well as the reversible Michaelis Menten kinetics, both mentioned in Section 3.3:

$$v = E\,k_{cat}^{+} \cdot \frac{\frac{S}{K_S}}{1 + \frac{S}{K_S} + \frac{P}{K_P}} \cdot \left(1 - e^{\frac{\Delta_r G'}{RT}}\right) = V^{+} \cdot \kappa \cdot (1 - \theta) \tag{38}$$

V^{+} denotes the *capacity term*, κ is a *saturation fraction* and θ is the *thermodynamic term*, which has been introduced already. κ and $(1-\theta)$ are both bounded between 0 and 1.

Kinetic orders $\left(\frac{\partial \ln(v)}{\partial \ln(x)}\right)$ are very often referred to as elasticities in the context of *Metabolic Control Analysis (MCA)* [66, 75], which is a sister discipline to BST. These two can easily be translated into one another. Noor et al. [113] calculated the kinetic order of the rate law, as defined in Equation 38 with respect to the substrate concentration and gained the following insight:

$$g_S^v = \frac{\partial \ln(v)}{\partial \ln(S)} = \kappa + \frac{\theta}{1-\theta} \tag{39}$$

This can be interpreted as follows: At low driving forces, the kinetic order is dominated by the *thermodynamic contribution*, $\frac{\theta}{1-\theta}$. The kinetic order can therefore be determined without knowing any of the kinetic constants. Compare Figure 17, which shows the change in $\frac{\theta}{1-\theta}$ at decreasing driving force.

The kinetic order $g_{i,j}$ can therefore be defined as

$$g_{i,j} = g_{i,j}^{k} - s_{i,j}\,\frac{\theta}{1-\theta} \tag{40}$$

$s_{i,j}$ will be negative for substrates (compare Equation 39) and positive for products. $g_{i,j}^{k}$ is the kinetic contribution and matches κ. It is bounded between zero and one for Michaelis Menten-type kinetics and between zero and the Hill coefficient for allosteric rate laws [155, 163].

Therefore, for reactions far away from equilibrium ($|\Delta_r G| \gg 2.5\,kJ/mol$), the influence of thermodynamics on the kinetic orders can be ignored, because the thermodynamic contribution tends to zero.

4.2 STEADY STATE ANALYSIS

The dynamic behavior of a model can only be analyzed, when the system of differential equations is solved and evaluated. Still, many important aspects at or close to a steady state can be analyzed with much simpler methods. These are actually of great importance, because most metabolic networks operate close to a steady state [175]. For this reason, the present section will be concerned with the calculation of the steady state of a given system.

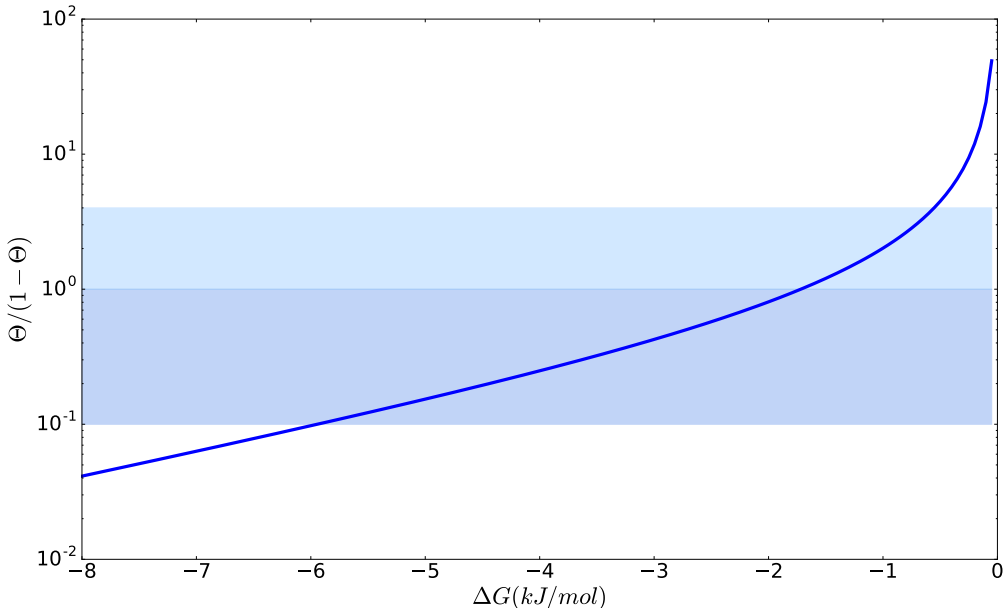


Figure 17: Change in the thermodynamic contribution, $\frac{\theta}{1-\theta}$, in logarithmic coordinates, with varying Gibbs energy. The dark blue area illustrates the part in which the thermodynamic and kinetic contribution are comparable, if one assumes Michaelis Menten-type kinetics, because then they both lie in between 0 and 1. Allosteric enzymes have kinetic contributions of magnitudes between 0 and 4, which is highlighted through the light blue area for an thermodynamic contribution between 1 and 4 [155, 163].

The steady state describes the situation, where the temporal change in concentration of the metabolites equals zero. Therefore, $\dot{x} = 0$, compare Equation 2. This does not mean that nothing is happening in the pathway. Material is still flowing through the systems but there are no temporal net-changes. The fluxes are in balance [175].

Considering, for instance, an S-system, it applies that:

$$0 = \alpha_i \prod_{j=1}^{n} x_j^{g_{i,j}} - \beta_i \prod_{j=1}^{n} x_j^{h_{i,j}}, \qquad i = 1, 2, ..., m \tag{41}$$

This equation can be rearranged with the help of the logarithm ($\ln(x) = y$):

$$\alpha_i \prod_{j=1}^{n} x_j^{g_{i,j}} = \beta_i \prod_{j=1}^{n} x_j^{h_{i,j}} \tag{42}$$

$$\ln(\alpha_i) + \sum_{j=1}^{n} g_{i,j} \ln(x_j) = \ln(\beta_i) + \prod_{j=1}^{n} h_{i,j} \ln(x_j) \tag{43}$$

$$\sum_{j=1}^{n} (g_{i,j} - h_{i,j})\, y_j = \ln\left(\frac{\beta_i}{\alpha_i}\right) \tag{44}$$

Some variables can be merged in order to simplify the system: $a_{i,j} = g_{i,j} - h_{i,j}$ and $b_i = \ln\left(\frac{\beta_i}{\alpha_i}\right)$. It follows:

$$\sum_{j=1}^{n} a_{i,j}\, y_j \; = b_i, \qquad i \; = \; 1, 2, ..., m \tag{45}$$

In matrix notation: $\mathbf{A}\,\vec{y} \; = \; \vec{b}$

Therefore, in spite of being a fully nonlinear system, the steady state of a S-system is the solution of a linear equation. The coefficients of dependent and independent variables can be partitioned in two matrices, $\mathbf{A_D}$ and $\mathbf{A_I}$, respectively and it applies:

$$\mathbf{A_D}\,\vec{y_D} \; + \; \mathbf{A_I}\,\vec{y_I} \; = \; \vec{b} \tag{46}$$

Simple linear algebra can then be used to solve $\vec{y_D}$ as a function of the independent variables and the parameters of the system. There exists a unique steady state, if the determinant of the quadratic matrix $\mathbf{A_D}$ is non-zero, otherwise there can be infinite steady states [87].

4.3 SENSITIVITY ANALYSIS

Sensitivity analysis is a framework, which allows the examination of a system's response to different influences. Typically, two types of influences can be present [175]: First, persistent changes in the numerical values of independent variables, which can be caused by a change in environment. The system may therefore assume a new steady state. The relative change of a system property due to such a scenario is called *logarithmic gain* and represents a measure for the robustness of a model. Second, permanent changes in system parameters (kinetic orders or rate constants), which is the consequence of a structural change in the system due to a mutation or disease and results in altered enzyme activity. The change of a system property due to structural change is called *sensitivity*.

There exist, among others, the following four different types of logarithmic gains and sensitivities, which allow testing the quality of a model in steady state:

- Logarithmic gain of a steady state concentration with respect to a change in an independent variable
- Logarithmic gain of a flux with respect to a change in an independent variable
- Sensitivity of a steady state concentration with respect to a change in a parameter
- Sensitivity of a flux with respect to a change in a parameter

The calculation of these logarithmic gains and sensitivities will be explained below.

LOGARITHMIC GAIN OF A METABOLITE The relative change in a variable corresponds to an absolute change in its logarithm [175]:

$$\frac{\dot{x}(t)}{x(t)} = \frac{d}{dt}\ln(\dot{x}(t)) = \dot{y}(t) \tag{47}$$

whereas $y(t) = \ln(x(t))$.

The logarithmic gain of the steady state of a dependent variable x_i with respect to a change in an independent variable x_j, $L(x_i, x_j)$, can therefore be defined as:

$$L(x_i, x_j) = \frac{\partial \ln(x_i)}{\partial \ln(x_j)} = \frac{\partial y_i}{\partial y_j} \tag{48}$$

Because a dependent variable is usually influenced by several independent variables, it is convenient to summarize the different logarithmic gains in a matrix ($\mathbf{L}(x_D, x_I)$). The line i contains all logarithmic gains of a dependent variable x_i. For a system with p dependent and q independent variables, the matrix would exhibit the dimension $p \times q$.

LOGARITHMIC GAIN OF A FLUX The flux in power law representation has been defined in Equation 33. A change in an independent variable x_j has two effects on a flux v_i. The first one is a direct, the second an indirect one, caused by the changes in the metabolites as a response to the change in x_j [175]. This results in:

$$L(v_i, x_j) = \frac{\partial \ln(v_i)}{\partial \ln(x_j)} = f_{i,j} + \sum_{k=1}^{n} f_{i,k} L(x_k, x_j) \tag{49}$$

These logarithmic gains, $L(v, x)$, can also be summarized in a matrix, which would have n rows and q columns in case of n fluxes and q independent variables in the system.

SENSITIVITY OF A METABOLITE The changes in parameters, which are analyzed via the sensitivities, are not persistent ones, as for the logarithmic gains, but permanent ones. Such a change might be the availability of a new enzyme, which increases the speed of certain reactions and therefore alters the rate constants. As in case of the logarithmic gains, the calculation of the sensitivities starts out with the calculation of the steady states, see Section 4.2, and moreover uses partial derivatives. The calculation of the sensitivities of a steady state concentration with respect to change in rate or kinetic constants differs from each other. Thus, they have to be considered separately [175].

Sensitivity with respect to rate constants For the sensitivity of the dependent metabolite x_i with respect to the rate constant α_j, it applies that:

$$S(x_i, \alpha_j) = \frac{\partial \ln(x_i)}{\partial \ln(\alpha_j)} = -\frac{\partial \ln(x_i)}{\partial \ln(\beta_j)} \tag{50}$$

The matrix, which includes all sensitivities with respect to the rate constants is $\mathbf{S}(x, \alpha)$, which equals $-\mathbf{S}(x, \beta)$.

Sensitivity with respect to kinetic orders The sensitivity of the dependent metabolite x_i with respect to kinetic orders ($g_{j,k}$ or $h_{j,k}$) can be calculated as follows:

$$S(x_i, g_{j,k}) = \frac{\partial \ln(x_i)}{\partial \ln(g_{j,k})} = \frac{\partial x_i}{\partial g_{j,k}} \frac{g_{j,k}}{x_i} \tag{51}$$

SENSITIVITY OF A FLUX The sensitivities of a flux with respect to different groups of parameters are considered independently as well.

Sensitivity with respect to rate constants The sensitivity of a flux v_i with respect to changes in rate constants (α_p or β_p) is given by:

$$\begin{aligned} S(v_i, \alpha_p) &= \frac{\partial \ln(v_i)}{\partial \ln(\alpha_p)} = \frac{\partial v_i}{\partial \alpha_p} \frac{\alpha_p}{v_i} \\ &= \begin{cases} 1 + \sum_{j=1}^{n} f_{i,j} S(x_j, \alpha_p), & \text{if } p = i \\ \sum_{j=1}^{n} f_{i,j} S(x_j, \alpha_p), & \text{if } p \neq i \end{cases} \end{aligned} \tag{52}$$

Sensitivity with respect to kinetic orders The sensitivity with respect to changes in the kinetic orders ($g_{j,k}$ or $h_{j,k}$) is:

$$S(v_i, g_{j,k}) = \frac{\partial \ln(v_i)}{\partial \ln(g_{j,k})} = \frac{\partial v_i}{\partial g_{j,k}} \frac{g_{j,k}}{v_i} \tag{53}$$

4.4 MATHEMATICALLY CONTROLLED COMPARISON

The concepts introduced in the previous section are very helpful for the examination of the quality of a system but their potential is not completely exploited. They can also be used to compare different kinds of systems to figure out which is superior with respect to functional effectiveness measures. Michael Savageau, who developed the biochemical systems theory, analyzed different kinds of design principles with respect to unbranched pathways to uncover, if variations in design happened by chance or if they might have been selected for functional reasons. A design principle, as defined by Savageau, is a rule that characterizes a feature of a class of systems, such that the discovery of the rule allows one not only to understand known instances but also to predict new instances within the class [145, 150]. Comparing different classes is not an easy task, since two objects of the same class can already exhibit differences in their quality. One therefore needs a systematic approach.

A first step would be the definition of the classes, which are about to be compared. The alternatives should differ from the reference in only a single process

that therefore becomes the focus of the analysis. The reference class could be an unbranched pathway with only irreversible reactions, e.g., the alternative could be designed accordingly, except for a feedback inhibition of the first reaction by the end product [144]. The definition is easily done with power law models, because information about the kind of inhibition, e.g., is not necessary. Furthermore, a protocol has to be established, which allows the comparison of this kind of classes, such as the *Mathematically Controlled Comparison (MCC)* [5–9, 144]. According to MCC, the classes must be equivalent on two levels: internal and external.

INTERNAL EQUIVALENCE Since a process is modified, while building the alternative from the reference class, the parameters (kinetic parameters, rate constant) of this single reaction may differ as well. All other parameters should stay the same, which is defined as *internal equivalence*. The notation will be that a parameter (p) of the alternative system will be indicated by a prime: p'. Thus, internal equivalence is expressed in the condition $p = p'$, which has to apply for all parameters which are not taking part in the process of interest.

EXTERNAL EQUIVALENCE With internal equivalence, most parameters are equal between the reference and alternative classes. The remaining parameters represent degrees of freedom, which need to be eliminated. This can be done by ensuring that the systemic behaviors of both classes are identical to an external observer [150]. This is called *external equivalence*. Typically, the steady state values, fluxes and selected logarithmic gains are forced to be equal, which allows the determination of parameters of the alternative class as a function of the parameters in the reference class. Such constraints are imposed until all free parameters of the alternative class are determined, which eliminates all differences that are not inherent to the class but only characteristic of some particular cases.

Subsequently, measures of functional effectiveness, relevant to the biological context of these designs are determined and used to compare the classes and determine their remaining differences [150]. These measures of functional effectiveness are clarified in the following subsection. MCC therefore provides qualitative statements about the superiority of one design over the other with respect to certain characteristics but it cannot provide a quantitative one. If one wishes to know how much better one design is, the analysis has to be executed with specific values for the parameters involved. Given that these often are not available, Alves and Savageau sampled them from distributions representing prior knowledge about the likely ranges of their values and afterwards analyzed them through statistical approaches [5–9, 150].

They examined, among others, the performance of an unbranched pathway with feedback inhibition of the first reaction by the end-product and interchanged the position of a single irreversible reaction within the pathway, which otherwise only exhibits reversible reactions [9]. Applying different criteria of functional effectiveness (see the following subsection), their analysis revealed

that the best position for an irreversible step within this pathway, would be the first. Moreover, they concluded that a fully reversible pathway is more stable and robust, than a fully irreversible one, although it is possible to optimize its performance with a single irreversible step in the beginning of the pathway.

Criteria of functional effectiveness

The comparison of different pathway designs has been clarified but on what grounds? What are the criteria, which should be fulfilled by a pathway to be considered superior? Why is end-product inhibition the most prevalent pattern of control in biochemical networks [7, 144]? Savageau listed and discussed different kinds of quantitative criteria, which can be considered in unbranched metabolic pathways, to answer such questions. Important ones are the following [6, 9, 144]:

- A biochemical pathway should be robust, which means insensitive to fluctuations in the parameter values, which concerns the sensitivities of the steady states and the fluxes (see Section 4.3).
- The flux through the pathway should be responsive to changes in demand for the final product, which ensures that the amount of material flowing is directly coupled to the metabolic needs. This concerns the logarithmic gain $L(v, x_{n+1})$.
- The system should respond quickly to environmental changes, requiring short transition times.
- The steady state of the system should be dynamically stable following small perturbations in the concentration variables or would otherwise be dysfunctional.

The stability of a steady state is an especially interesting criteria, which will be considered further. Four factors exhibit a great influence, according to Savageau: The strength of feedback inhibition, the number of reactions in the pathway, the degree of equality of the kinetic orders and the alternative patterns of feedback inhibition.

STRENGTH OF FEEDBACK INHIBITION AND PATHWAY LENGTH Savageau compared unbranched pathways of different length and with different strength of end-product inhibition.

As it turned out, a shorter pathway, i.e. one of only three steps, allows for higher feedback signals, than one with more steps, without losing stability. This can be seen in Figure 18. On the left, the systems stay stable until the strength of the feedback signal turns 4 or higher. In case of a pathway with more steps, like one with 10 (panel on the right), the system already starts oscillating with a feedback signal of about 0.8. Therefore, in order to ensure stability, the pathway should either be quite short or it should not allow for high feedback signals, because it will not be able to actively regulate the system any longer.

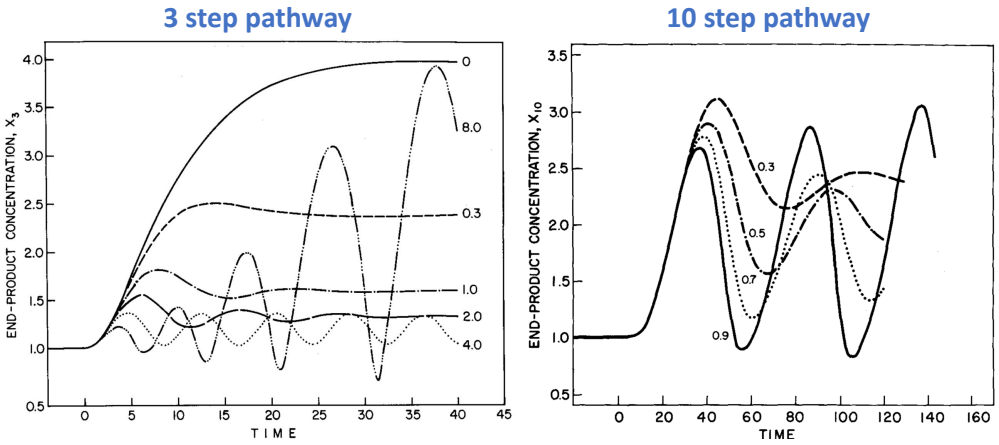


Figure 18: Strength of feedback inhibition vs. stability. The longer the pathway, the easier does a feedback signal cause instability [143].

EQUALIZATION OF KINETIC PARAMETERS Another possibility to stabilize a system is the control of the kinetic parameters. If these parameters are very different in value for each reaction, they promote stability. This finding actually makes great sense, because it means that some of the reactions occur at much faster rates. That is, these reactions are irrelevant for the system's temporal behavior and thus the pathway may be considered to have fewer reactions, than it actually does, from a kinetic point of view. This effect is termed *kinetic shortening*. The different time scales will be especially important in the following section.

ALTERNATIVE PATTERNS OF CONTROL As already mentioned, the most prevalent pattern of control in biochemical networks is the feedback inhibition. Savageau [6, 144] examined the different possibilities of feedback control in an unbranched pathway and concluded that the end-product feedback inhibition of the first reaction is the one closest to the boundary of instability. Savageau concluded that stability might be an important criteria for selection but it does not necessarily have to be maximized. Stabilization beyond the degree necessary for survival might even be a disadvantage, because it makes the system's temporal response to change very sluggish.

4.5 MODEL REDUCTION VIA TIME HIERARCHIES

As one can imagine, methods, like MCC, might become extremely complicated as soon as more complex chemical reaction systems are analyzed. Biological systems often involve tens or even hundreds of reactants participating in a multitude of reactions and occur on a broad spectrum of time scales (denoted as hierarchies) [183]. Analysis of such detailed models are hindered by different factors: First, they include a great number of mostly unknown kinetic parameters; second, these models are often inherently stiff due to widely different orders of magnitude of the values of the kinetic parameters; finally, it is a sig-

nificant computational challenge to simulate such a model [57, 177]. To facilitate this task, a reduction of the model, which still captures the time evolution of species composition with sufficient accuracy is required. Different approaches exist: One is the reduction via time-scale analysis, useful for models with fast and slow reactions and can be applied to linear as well as non linear models. Vora and Daoutidis developed a method, which is based on singular perturbation analysis and which allows the systematic identification of independent quasi-steady-state constraints for the fast reactions of a system. All fast reactions are thus assumed to be in quasi-steady-state. In the end one obtains a model of the slow dynamics, which is independent of the fast reaction rate expressions and moreover non-stiff. Naturally, the approach leads to a change of variables, which results in linear combinations of metabolite concentrations, whose dynamics evolve only slowly.

The system of equations can be divided into *slow* (s) and *fast* (f) variables, as follows:

$$\dot{x} = \mathbf{S}_s \vec{v_s} + \mathbf{S}_f \vec{v_f} \tag{54}$$

with p fast and $q - p$ slow reactions [87]. A reaction can be described by $v_i = \gamma_i \cdot h^i(x_j)$, where γ_i is the rate constant for the i-th reaction (compare Equation 32) and $h^i(x_j)$ is a non-linear function describing the dependence of the reaction rate on the concentration of the j-th reactant. According to singular perturbation theory, the characteristic time constant of a reaction i is $t_i = 1/\gamma_i$. A small time constant indicates that the reaction proceeds rapidly, a big one the contrary [57]. A reaction can be classified into fast or slow on the basis of this time constant, which can be calculated for systems close to steady state, if their time scale is not already known from literature. To accomplish this, the eigenvalues of the linearized model ($\dot{x} = J x$, J is the Jacobi matrix of the system) have to be calculated and it applies that $t_i = 1/|\mathrm{Re}(\lambda_i)|$ [87].

Because the rate vector consists of a large (γ) and a small term ($h_f(x_j)$), the system is stiff, which can be overcome by scaling v_f with a single parameter (ϵ), which then contains the stiffness. With singular perturbation arguments, the quasi-steady-state conditions for the fast reactions can then be determined and the fast modes can be eliminated from Equation 54 by a change of variables: a vector of fast reactions (z_f) in the slow time scales replaces v_f [57].

$$\dot{x} = \mathbf{S}_s \vec{v_s} + \mathbf{S}_f \vec{z_f} \tag{55}$$

In a next step, the system is reduced via a coordinate change: $\vec{P} = \mathbf{C}\vec{x}$. It applies:

$$\begin{aligned} \dot{\vec{P}} &= \mathbf{C}\dot{\vec{x}} \\ &= \mathbf{C}\mathbf{S}_s \vec{v_s} + \mathbf{C}\mathbf{S}_f \vec{z_f} \end{aligned} \tag{56}$$

Because the variables $\vec{P}$ are not yet true slow variables, as long as z_f is present, $\mathbf{C}$ has to be taken from the left null space (see Chapter 2) of $\mathbf{S}_f$. This decouples the slow dynamics from the fast reaction rates z_f.

The paper of Gerdtzen, Daoutidis, and Hu includes the application of the method. One example concerns the central carbon metabolism in erythrocytes. The original model contains 20 metabolites and 25 reactions. Seven of the reactions are identified as fast, which includes reversible, as well as irreversible reactions. By applying the reduction method, five reactions were eliminated and only 13 pools ($\vec{P}$) of metabolites were left.

It has been pointed out that constraint-based methods, as FBA and thermodynamics, can be combined and further used to gain insight to build and analyze a dynamical model. The combination of these constraint-based methods did not present any difficulties, as has been shown with the ammonia assimilation example in Chapter 3. Still, kinetic parameters are usually not known, which complicates the transition from stoichiometry to dynamics. Therefore, an additional method will be introduced in Chapter 5, which will help to bridge the gap. It has been published in [155].

RESULTS

5 THERMODYNAMIC SHORTENING

5.1 INTRODUCTION

To derive a dynamic model, which is able to mirror experimental results correctly, is still a difficult task in case of biological systems. This is caused by (among others) the lack of techniques to measure kinetic parameters on a large scale. Therefore, it seems reasonable to constrain the possible values of the components to eliminate any parameter combination, which is not feasible. Subsequently, one can concentrate on identifying likely evolutionary strategies and possible ways to achieve them, which will reduce the amount of possible states even further. This approach makes sense, because physics and chemistry constrain all theoretically possible phenotypes, which are available to evolution and only 'successful' strategies prevail on the long term [157]. The workflow is illustrated in Figure 19.

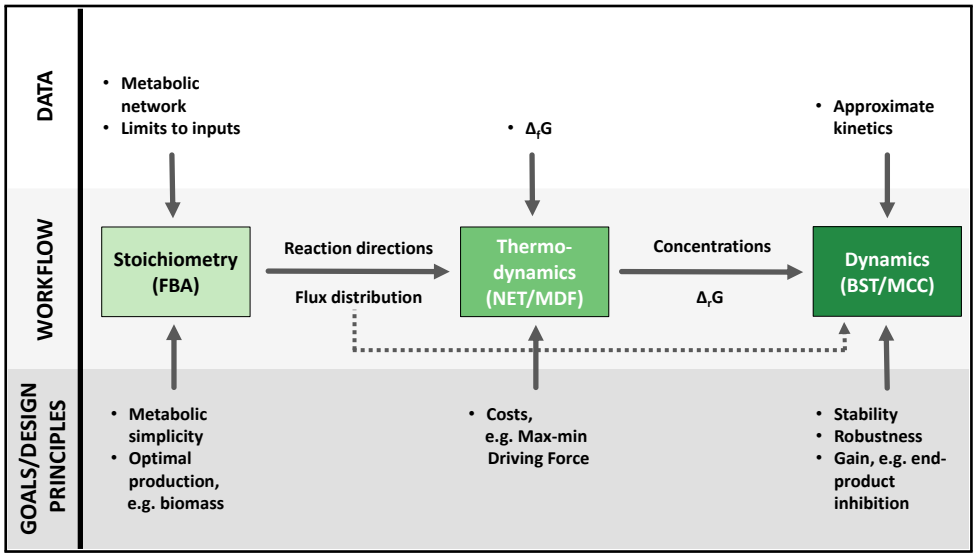


Figure 19: A sequential workflow from metabolic reconstructions to dynamic models. Each of the three techniques featured here incorporates different kinds of information into the process. This information can be of two kinds: different types of data, shown in the upper part of the figure, and evolutionary considerations at different levels, shown in the lower part of the figure. Furthermore, the results of each step are fed to the next. The flux distributions provided by FBA can be used as an input to thermodynamics and the results of both approaches can be translated to parameters for a dynamic model, which can be formulated and analyzed according to BST and MCC.

As has been mentioned, being close to or far from equilibrium has a great effect on a reaction. If a reaction is close to thermodynamic equilibrium, the forward and backward direction of a reaction are nearly equal, which elimi-

nates the net-flux and results in higher costs with respect to enzyme investment. The opposite case is a reaction, which is far away from equilibrium. Assuming its Gibbs energy is negative, it follows that the backward flux is almost non-existent and the reaction is considered as irreversible forward. These two are extreme cases, which can be considered on the pathway level as well: An unbranched pathway, whose reactions are close to thermodynamic equilibrium, versus an unbranched pathway, whose reactions are far away from equilibrium. The dynamics of the first version have already been elaborated by Michael Savageau, as described in the previous chapter, while the second is less well understood. The system is clearly more complicated. It is already known that the kinetic orders are dominated by the thermodynamic contribution, which tends to infinity, if thermodynamic equilibrium is approximated, and therefore counters possible perturbations of substrate and product, see Figure 17. Since the ratio of product and substrate tends to the equilibrium constant, it is possible to establish a method, which simplifies the pathway in a way, which is similar to the one described in Section 4.5, utilizing thermodynamic data.

5.2 THERMODYNAMIC SHORTENING

The topic *model reduction* has already been mentioned in Section 4.5. Although the presented method has its advantages, it cannot be denied that it also has some disadvantages. First, the fast reactions have to be determined, which requires information from literature or the calculation of the eigenvalues and eigenvectors from, e.g., a parameter estimation analysis. Second, one introduces extra algebraic conditions via the quasi-steady-state assumption, which constrain the system. An alternative reduction, which is based on the idea of time hierarchies, is *thermodynamic shortening*. It is introduced by the means of the unbranched pathway example from before (Chapter 4), see Figure 20, which is followed by a short discussion about the differences of the two model reduction methods.

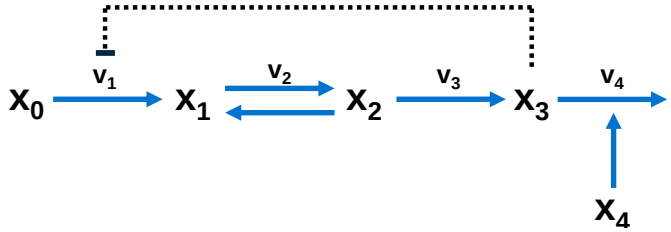


Figure 20: Unbranched pathway with feedback inhibition and a single reversible step.

Thermodynamic shortening uses a classification of the reactions similar to Gerdtzen but instead of fast and slow reactions, it considers reactions, which are far from or close to thermodynamic equilibrium. It is possible that this division matches the one based on time hierarchies, but not necessarily. The advantage of thermodynamic shortening lies in the possibility to use the results of the Max-min Driving Force algorithm, concerning the distance to equilibrium. Moreover, it enables the reformulation of the model in a reduced form, which

will be consistent with the behavior of the full network, as was the case for the time hierarchy reduction.

x_0 and x_4 are independent variables, while all the other state variables are dependent.

Defining Equation 1 for the dependent variables results in:

$$\begin{aligned} \dot{x_1} &= v_1 - v_2 \\ \dot{x_2} &= v_2 - v_3 \\ \dot{x_3} &= v_3 - v_4 \end{aligned} \tag{57}$$

Assuming that reaction ν_2 is in thermodynamic equilibrium, one can reorder the reactions, as has been done for the time hierarchies:

$$\dot{\vec{x}} = \mathbf{S}_{irr}\, \vec{v_{irr}}(\vec{x}) + \mathbf{S}_{eq}\, \vec{v_{eq}}(\vec{x}) \tag{58}$$

irr means *irreversible*, while eq stands for *reversible*, respective *in equilibrium*. It should be noted that $\vec{x}$ includes all variables, dependent as well as independent. The dependent variables can be further differentiated into free ($\vec{x_F}$) and bound variables ($\vec{x_B}$), which are connected via the equilibrium constant: If a reaction is in thermodynamic equilibrium, the equilibrium constant, K'_{eq} ($\theta = 1$), can be calculated and used to express the bound variable of the reaction through the free variable (as well as the independent variables, if present in the reaction): $K'_{eq} = \frac{x_2}{x_1}$.

Therefore the concentration of reactant x_2 can be calculated from the product of the substrate concentration x_1 and the equilibrium constant, since no independent variables are taking part in this example reaction. That is, x_2 is then defined as a bound variable, which depends on the free variable x_1. The other way around is equally possible. One therefore might define $\vec{x_B}$ as $[x_2]$, which can be calculated with help of the equilibrium constant, which in turn follows from the standard Gibbs energy (see Equation 17) and the concentrations of the free $[x_1, x_3]$ and independent metabolites $[x_0, x_4]$. This can be defined as follows, in logarithmic coordinates:

$$\log(\vec{K'_{eq}}) = -\mathbf{S}^{\mathbf{T}}_{\mathbf{eq,F}} \log(\vec{x_F}) + \mathbf{S}^{\mathbf{T}}_{\mathbf{eq,B}} \log(\vec{x_B}) + \mathbf{S}^{\mathbf{T}}_{\mathbf{eq,I}} \log(\vec{x_I}) \tag{59}$$

It follows:

$$\log(\vec{x_B}) = -(\mathbf{S}^{\mathbf{T}}_{\mathbf{eq,B}})^{-1} \cdot \left[\mathbf{S}^{\mathbf{T}}_{\mathbf{eq,F}} \log(\vec{x_F}) + \mathbf{S}^{\mathbf{T}}_{\mathbf{eq,I}} \log(\vec{x_I}) - \log(\vec{K'_{eq}})\right] \tag{60}$$

Because Equation 60 is a power law, it can easily be used to eliminate x_B and yet preserve the structure of the system.

As has been done for the reduction of the time hierarchies, one defines a new variable P, which is a linear combination of the former variables and therefore called *pool*. It applies that:

$$\dot{\vec{P}} = \mathbf{C}\,\mathbf{S}_{\mathbf{irr}}\, \vec{v_{irr}}(\vec{x}) + \mathbf{C}\,\mathbf{S}_{\mathbf{eq}}\, \vec{v_{eq}}(\vec{x}) \tag{61}$$

The pools must be defined, such that the fluxes $\vec{v_{eq}}$ vanish, since they become internal fluxes. This is important because the net-fluxes of $\vec{v_{eq}}$ are zero in equilibrium, which forces the in- and effluxes to be zero as well, in line with mass balance, if $\vec{v_{eq}}$ remains. In order to achieve that, $\mathbf{C}$ has to be chosen from the left null space of $\mathbf{S_{eq}}$. It remains:

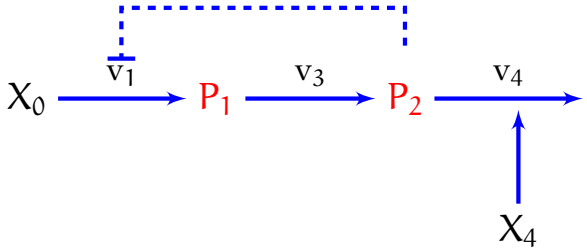
$$\dot{\vec{P}} = \mathbf{C}\,\mathbf{S_{irr}}\,\vec{v_{irr}}(\vec{x}) \tag{62}$$

Applying the method on the example yields the following results:

$$\begin{aligned} P_1 &= x_1 + x_2 \\ &= (1 + K'_{eq,2})\,x_1 \\ P_2 &= x_3 \end{aligned} \tag{63}$$

$$\begin{aligned} \dot{P_1} &= v_1 - v_3 \\ \dot{P_2} &= v_3 - v_4 \end{aligned} \tag{64}$$

As one can see in Equation 63, the pools are only dependent on the free metabolites. The bound ones are replaced, whereas one has to keep in mind that the fluxes still depend on the dependent, as well as the independent metabolites in Equation 64. Only two differential equations are left. The pool diagram, which displays all interconnections, looks as follows:



As can be seen in Equation 34 and 35, the differential equations of a dynamic model are supposed to depend on themselves and the parameters. Because the fluxes of the equation system 64 are still dependent on the old variables, it has to be further transformed. The aim is to gain a system, which is dependent on the pools themselves.

So far, the fluxes are dependent on the dependent and independent metabolites, as has been seen in Equation 33:

$$v_i = \gamma_i \prod_j x_{Bj}^{f^B_{i,j}} \prod_k x_{Fk}^{f^F_{i,k}} \prod_l x_{Il}^{f^I_{i,l}} \tag{65}$$

The parameters f can be combined in a single matrix $\mathbf{F_x}$, which describes the effect of the metabolites on the fluxes. In this example $\mathbf{F_x}$ looks as follows:

$$\mathbf{F_x} = \begin{array}{c} \\ v_1 \\ v_3 \\ v_4 \end{array} \begin{array}{c} \begin{array}{ccccc} x_2 & x_1 & x_3 & x_0 & x_4 \end{array} \\ \begin{pmatrix} 0 & -f^F_{1,1} & -f^F_{1,3} & f^I_{1,0} & 0 \\ f^B_{3,2} & 0 & -f^F_{3,3} & 0 & 0 \\ 0 & 0 & f^F_{4,3} & 0 & f^I_{4,4} \end{pmatrix} \end{array} \tag{66}$$

The rows of the matrix correspond to the remaining fluxes (v_2 is not present anymore) and the columns correspond to the bound, the free and the independent metabolites. In this case, it is assumed that all products have a negative effect ($-f$) on their synthesis reaction, while all substrates are expected to have a positive effect ($+f$). The inhibiting effect of x_3 on v_1 is expressed through $-f^F_{1,3}$. If more known regulatory effects exist they can easily be added to the matrix.

Since $\mathbf{F_x}$ consists of all the derivatives of the logarithms of the fluxes with respect to the metabolites, it can be written as follows:

$$\mathbf{F_x} = \frac{\partial \log v}{\partial \log x} = \underline{\overline{V}}^{-1} \frac{\partial \overrightarrow{v}}{\partial \overrightarrow{x}} \underline{\overline{X}} \tag{67}$$

$\underline{\overline{V}}^{-1}$ and $\underline{\overline{X}}$ are diagonal matrices with the irreversible fluxes, respectively the metabolites as diagonal elements.

As has been indicated, the fluxes should not be dependent on the dependent metabolites anymore, but on the pools and the independent metabolites:

$$v_i = \tilde{\gamma}_i \prod_j p_j^{f^p_{i,j}} \prod_l x_{Il}^{f^I_{i,l}} \tag{68}$$

and for the $f^p_{i,j}$ it shall apply that:

$$\mathbf{F'_P} = \underline{\overline{V}}^{-1} \frac{\partial \overrightarrow{v}}{\partial \overrightarrow{P}} \underline{\overline{P}} \tag{69}$$

$\underline{\overline{P}}$ is a diagonal matrix with the pools as diagonal elements.

In order to be able to calculate Equation 68, $\frac{\partial \overrightarrow{v}}{\partial \overrightarrow{P}}$ needs to be calculated ($\vec{P}$ represents a vector of the pools):

$$\begin{aligned}
\frac{\partial \overrightarrow{v}}{\partial \overrightarrow{P}} &= \frac{\partial \overrightarrow{v}}{\partial \overrightarrow{x_F}} \frac{\partial \overrightarrow{x_F}}{\partial \overrightarrow{P}} + \frac{\partial \overrightarrow{v}}{\partial \overrightarrow{x_B}} \frac{\partial \overrightarrow{x_B}}{\partial \overrightarrow{x_F}} \frac{\partial \overrightarrow{x_F}}{\partial \overrightarrow{P}} \\
&= \left[\frac{\partial \overrightarrow{v}}{\partial \overrightarrow{x_F}} + \frac{\partial \overrightarrow{v}}{\partial \overrightarrow{x_B}} \underline{\overline{X}}_B \frac{\partial \log \overrightarrow{x_B}}{\partial \log \overrightarrow{x_F}} \underline{\overline{X}}_F{}^{-1} \right] \frac{\partial \overrightarrow{x_F}}{\partial \overrightarrow{P}} \\
&= \left[\frac{\partial \overrightarrow{v}}{\partial \overrightarrow{x_F}} + \frac{\partial \overrightarrow{v}}{\partial \overrightarrow{x_B}} \underline{\overline{X}}_B \frac{\partial \log \overrightarrow{x_B}}{\partial \log \overrightarrow{x_F}} \underline{\overline{X}}_F{}^{-1} \right] \left(\frac{\partial \overrightarrow{P}}{\partial \overrightarrow{x_F}} \right)^{-1}
\end{aligned} \tag{70}$$

The change in-between the last two steps of Equation 70 is valid: It applies that $d\overrightarrow{P} = \frac{\partial \overrightarrow{P}}{\partial \overrightarrow{x_f}} d\overrightarrow{x_f}$ and therefore $d\overrightarrow{x_f} = \left[\frac{\partial \overrightarrow{P}}{\partial \overrightarrow{x_f}} \right]^{-1} d\overrightarrow{P}$. Since it also applies that $d\overrightarrow{x_f} = \frac{\partial \overrightarrow{x_F}}{\partial \overrightarrow{P}} d\overrightarrow{P}$, one can conclude that: $\left[\frac{\partial \overrightarrow{P}}{\partial \overrightarrow{x_f}} \right]^{-1} = \frac{\partial \overrightarrow{x_F}}{\partial \overrightarrow{P}}$.

In order to complete the calculation, the single components of Equation 70 need to be determined:

$$\frac{\partial \overrightarrow{v}}{\partial \overrightarrow{x_F}} = \underline{\overline{V}} \mathbf{F}_{\mathbf{x_F}} \underline{\overline{X}}_F{}^{-1} \tag{71}$$

$$\frac{\partial \overrightarrow{v}}{\partial \overrightarrow{x_B}} = \overline{V}\,\mathbf{F_{x_B}}\,\overline{X}_B^{-1} \tag{72}$$

whereas $\overline{X}_F^{-1}$ and $\overline{X}_B^{-1}$ are diagonal matrices, which contain the reciprocals of the free and bound metabolites as diagonal elements.

From Equation 60 it follows that:

$$\frac{\partial \log \overrightarrow{x_B}}{\partial \log \overrightarrow{x_F}} = -(\mathbf{S^T_{eq,B}})^{-1}\,\mathbf{S^T_{eq,F}} \tag{73}$$

The vector of pools can be differentiated as follows:

$$\begin{aligned}\frac{\partial \overrightarrow{P}}{\partial \overrightarrow{x_F}} &= \mathbf{C_F} + \mathbf{C_B}\frac{\partial \overrightarrow{x_B}}{\partial \overrightarrow{x_F}} \\ &= \mathbf{C_F} + \mathbf{C_B}\overline{X}_B\,\frac{\partial \log \overrightarrow{x_B}}{\partial \log \overrightarrow{x_F}}\,\overline{X}_F^{-1}\end{aligned} \tag{74}$$

Therefore, $\mathbf{F'_P}$ can be calculated as:

$$\mathbf{F'_P} = [\mathbf{F_{x_F}}, \mathbf{F_{x_B}}] \cdot \underbrace{\left[\begin{bmatrix}\mathbf{I_F} \\ \frac{\partial \overrightarrow{P}}{\partial \overrightarrow{x_F}}\end{bmatrix} \overline{X}_F^{-1} \cdot \left[[\mathbf{C_F}, \mathbf{C_B}] \begin{bmatrix}\mathbf{I_F} \\ \overline{X}_B\,\frac{\partial \log \overrightarrow{x_B}}{\partial \log \overrightarrow{x_F}}\,\overline{X}_F^{-1}\end{bmatrix}\right]^{-1} \cdot \overline{P}\right]}_{\mathbf{U}} \tag{75}$$

$$= \begin{pmatrix} f^p_{1,1} & f^p_{1,2} \\ f^p_{3,1} & f^p_{3,2} \\ 0 & f^p_{4,2} \end{pmatrix} \tag{76}$$

Because the fluxes of the pools still depend on the independent variables, it applies that:

$$\mathbf{F_P} = \begin{array}{c} \\ v_1 \\ v_3 \\ v_4 \end{array} \begin{array}{c} \begin{array}{cccc} P_1 & P_2 & x_0 & x_4 \end{array} \\ \begin{pmatrix} f^p_{1,1} & f^p_{1,2} & f^I_{1,0} & 0 \\ f^p_{3,1} & f^p_{3,2} & 0 & 0 \\ 0 & f^p_{4,2} & 0 & f^I_{4,4} \end{pmatrix} \end{array} \tag{77}$$

Now, that the fluxes are only dependent on the pools and the independent variables, the differential equation system 64 can be rewritten as follows:

$$\begin{aligned}\dot{P_1} &= \tilde{\gamma_1}\,P_1^{f^p_{1,1}}\,P_2^{f^p_{1,2}}\,x_0^{f^I_{1,0}} - \tilde{\gamma_3}\,P_1^{f^p_{3,1}}\,P_2^{f^p_{3,2}} \\ \dot{P_2} &= \tilde{\gamma_3}\,P_1^{f^p_{3,1}}\,P_2^{f^p_{3,2}} - \tilde{\gamma_4}\,P_2^{f^p_{4,2}}\,x_4^{f^p_{4,4}}\end{aligned} \tag{78}$$

The $\tilde{\gamma_i}$ can be calculated based on Equation 31, that is, they are dependent on the independent metabolites, if v_i depends on them. Thus, the terms for the

independent metabolites in the differential equations cancel out. $x_4^{f_{4,4}^p}$ in $\dot{P}_2$, for instance, cancels out, since the same term appears in the denominator of $\tilde{\gamma}_4$.

The following subsection deals with a direct comparison of both presented model reduction approaches on the basis of a simple example. This comparison shall illustrate the similarities and discrepancies between their resulting classifications.

Comparison of thermodynamic shortening and model reduction via time hierarchies

In this section the two approaches are compared with respect to their categorization on the basis of a simple example, which can easily be calculated and analyzed. Consider the unbranched pathway:

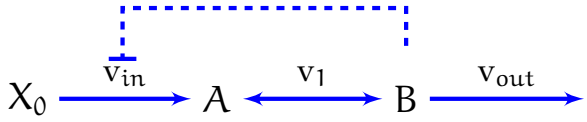


Table 1 lists the different cases, which are about to be examined and sums up the results. V_{max} refers to the velocity of equation v_1, while θ refers to its distance to equilibrium.

The rate laws are defined by power laws:

$$\begin{aligned} v_{in} &= k_{in} \cdot B^{-g} \\ v_{out} &= k_{out} \cdot B \\ v_1 &= k_1^f A - k_1^r B \end{aligned} \tag{79}$$

The steady states can be calculated as follows:

$$\begin{aligned} \dot{A} &= v_{in} - v_1 \\ &= k_{in} \cdot B^{-g} - k_1^f A + k_1^r B \\ \dot{B} &= v_1 - v_{out} \\ &= k_1^f A - k_1^r B - k_{out} \cdot B \\ &= k_1^f A - (k_1^r + k_{out}) B \end{aligned}$$

which results in:

$$|A|_0 = \frac{k_{in} \left(\frac{k_{in}}{k_{out}}\right)^{\frac{-g}{g+1}} + k_1^r \left(\frac{k_{in}}{k_{out}}\right)^{\frac{1}{g+1}}}{k_1^f}$$

$$|B|_0 = \left(\frac{k_{in}}{k_{out}}\right)^{\frac{1}{g+1}}$$

			Thermodynamic shortening applicable
		$\theta \ll 1$	$\theta \approx 1$
Time hierarchies applicable	$V_{max} \uparrow$	conflict with assumption	$v_{eq} = v_f = v_1$
	$V_{max} \downarrow$	$v_{eq} = v_f = \emptyset$	$v_{eq} \neq v_f$

Table 1: Comparison of both reduction methods, based on different assumptions concerning reaction ν_1. $V_{max} \uparrow$ means that $V_{max} \geqslant v_{in}, v_{out}$, while $V_{max} \downarrow$ stands for $V_{max} \leqslant v_{in}, v_{out}$. Since $\theta \approx 1$ means that ν_1 is close to thermodynamic equilibrium, the reduction via thermodynamic shortening is only applicable, if this holds. Furthermore, the reduction via time hierarchies can only be applied, if $V_{max} \uparrow$.

It therefore applies that:

$$\frac{|B|_0}{|A|_0} = \frac{\left(\frac{k_{in}}{k_{out}}\right)^{\frac{1}{g+1}} k_1^f}{k_{in}\left(\frac{k_{in}}{k_{out}}\right)^{\frac{-g}{g+1}} + k_1^r \left(\frac{k_{in}}{k_{out}}\right)^{\frac{1}{g+1}}} \tag{80}$$

$$= \frac{k_1^f}{k_{out} + k_1^r} \tag{81}$$

In the following, four different scenarios are discussed: The scenarios, in which it is assumed that ν_1 is a fast reaction ($V_{max} \geqslant v_{in}, v_{out}$) and a reduction via time hierarchies therefore possible, but at the same time the reaction is assumed to be either close ($\theta \approx 1$) or far from equilibrium ($\theta \ll 1$). As well as the scenarios that ν_1 is assumed to be slow but also close or far from equilibrium. This discussion sheds light on the question, if the subsequent classifications of the two reduction methods are the same and under which circumstances. The fact that the system is able to be reduced via thermodynamic shortening, even if the reduction via time hierarchies would not apply ($V_{max} \downarrow$, $\theta \approx 1$), points out that the newly developed reduction method is a valuable addition to the existing reduction methods.

- $V_{max} \geqslant v_{in}, v_{out}$ and $\theta \ll 1$:
  It applies that $k_1^f \gg k_{in}$ and $k_1^f \gg k_{out}$ and therefore:

  $$\frac{|B|_0}{|A|_0} = \frac{1}{\frac{k_{out}}{k_1^f} + \frac{k_1^r}{k_1^f}}$$

  $$\approx \frac{k_1^f}{k_1^r} = Keq$$

  It follows that the results contradict the assumption, since ν_1 is not supposed to be in equilibrium. Therefore, it can be followed that the fast reaction has to be in equilibrium.

- $V_{max} \leqslant v_{in}, v_{out}$ and $\theta \ll 1$:
  It applies that $k_1^f \ll k_{in}$ and $k_1^f \ll k_{out}$ and therefore:

$$\begin{aligned}\frac{|B|_0}{|A|_0} &= \frac{k_1^f}{k_{out}} \\ &\approx 0\end{aligned}$$

  Thus, ν_1 is considered as a slow reaction, which is also not close to equilibrium. Both approaches would result in the same classification and neither of them would be able to reduce the system any further.

- $V_{max} \geqslant v_{in}, v_{out}$ and $\theta \approx 1$:
  If ν_1 is in equilibrium, it applies that $\frac{k_1^f}{k_1^r} = K_{eq}$ and therefore:

$$\begin{aligned}\frac{|B|_0}{|A|_0} &= \frac{k_1^f}{k_{out} + k_1^r} \\ &= \frac{1}{\frac{k_{out}}{k_1^f} + \frac{1}{K_{eq}}}\end{aligned}$$

  If $V_{max} \geqslant v_{in}, v_{out}$ applies, then $k_1^f \to \text{inf}$ and therefore $\frac{|B|_0}{|A|_0} \to K_{eq}$
  Thus, ν_1 is considered a fast reaction, which is also close to equilibrium. Both approaches return to the same classification and are both able to reduce the system.

- $V_{max} \leqslant v_{in}, v_{out}$ and $\theta \approx 1$:
  In this case, ν_1 is in equilibrium but it does not apply that $k_1^f \to \infty$. Still, it is nevertheless possible to gain $\frac{|B|_0}{|A|_0} \to K_{eq}$, if $k_{out} \to 0$:

$$\begin{aligned}\frac{|B|_0}{|A|_0} &= \frac{k_1^f}{k_{out} + k_1^r} \\ &\to K_{eq}\end{aligned}$$

  Thus, ν_1 is considered a slow reaction, which is also close to equilibrium. The approaches gain different classifications with respect to the reactions at hand. Only thermodynamic shortening is able to reduce the system.

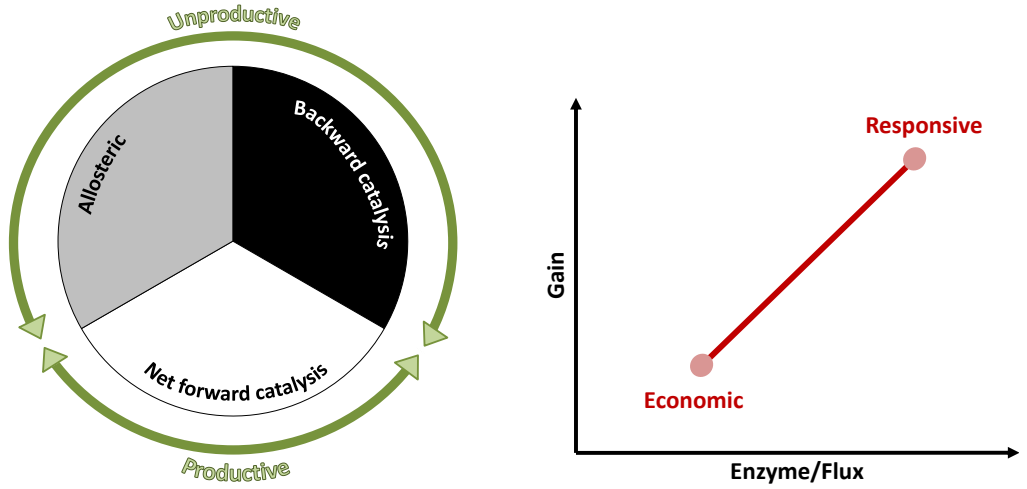
In conclusion, it has been seen that both reduction methods may result in the same flux-classification but they as well may not, depending on the conditions. In any case, although it has been concluded in the example system that the fast reaction needs to be in thermodynamic equilibrium, this does not apply universally. Irreversible reactions might as well be fast, as can be seen in the example in [57].

The inclusion of thermodynamics into the process of dynamical model-building has been shown to be a powerful tool to reduce the number of equations.

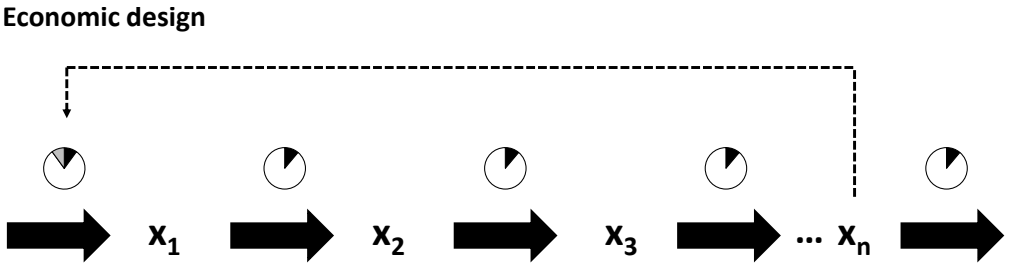
5.3 ECONOMIC VS. RESPONSIVE DESIGN

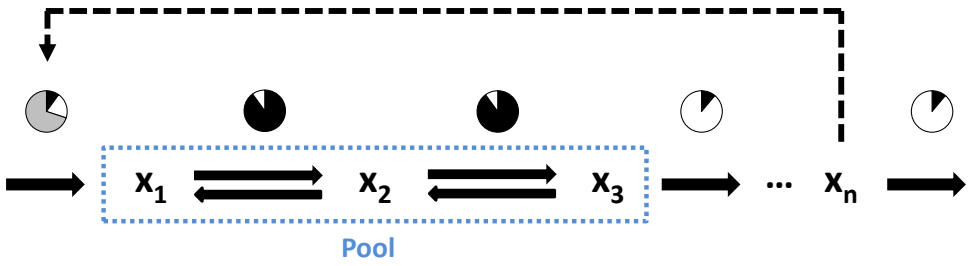
The two extreme cases of a given pathway, the fully irreversible and fully reversible pathway, have two biological significant phenotypes: The responsive and the economic design. They are both based on reasonable ideas but still very opposed in their properties. The economic design aims to be as cost efficient as possible by investing the least amount of enzyme, following the principle of the Max-min Driving Force, with enzymes working at saturation and few inhibitors, see Figure 21 (c). The alternative pathway design aims at the best performance, fulfilling the criteria of functional effectiveness (see Chapter 4). As has been stated before, the best performance can be achieved, when an unbranched pathway exhibits a first irreversible step, followed by steps close to thermodynamic equilibrium, which desensitizes the pathway with respect to changes in certain enzymes [9]. The robustness and stability can be further increased by thermodynamic shortening of the pathway, since a shorter pathway is in general more robust [144]. Furthermore, the design allows for a higher feedback signal, which reduces the amount of 'working' enzyme even further. The two contrary designs are considerable as two extremes in a trade-off (see figure Figure 21 (b)) and it might very well be that pathways alternate between both, depending on the circumstances.

As long as the demand is low, the flux stays low as well and therefore can be kept with a reasonable amount of enzymes, even if they operate at a fraction of their V_{max} (compare Equation 24), which corresponds to the responsive design. However, as soon as the demand increases, the cell needs to be prepared to respond quickly with an increase in flux to raise production. This scenario shifts the reactions further away from equilibrium and therefore closer to the economic design.



(a) (b)





(c)

Figure 21: The circles in (a),(c) schematically show the reduction of enzyme activity, due to different causes. In white, the fraction of the enzyme that is actually catalyzing the reaction forward. In grey the fraction that is inactive due to allosteric inhibition (or more exactly the fraction in which the total activity is reduced). In black, the fraction of activity that is lost due to the flux of the reverse reaction or due to insaturation of the enzyme; (b) shows the two alternative pathways in performance space and the trade-off between them; (c) displays a more detailed depiction of the two modes of operation: The economic variant carries a high flux and all its enzymes are operating close to their V_{max} due to weak inhibition and distance to equilibrium. In the responsive variant, the first enzyme is inhibited and two of the reactions are close to equilibrium. The efficiency of the enzymes is much lower but so is the flux they have to carry.

The remaining chapters will focus on the application of the presented methods on the model of *H. elongata*. To start out, a stoichiometric model of the organism is needed and will be discussed in Chapter 6.

6 THE METABOLIC RECONSTRUCTION OF *H. ELONGATA*

The model used in this thesis is based on a core-model, published by Schwibbert et al. in 2011 [153]. The following sections will introduce the organism *H. elongata* and its product ectoine and will furthermore outline the model of its metabolism, which has evolved since 2011.

6.1 *Halomonas elongata*

As mentioned, *Halomonas elongata* (type strain DSM 2581^T) is a halophilic γ-proteobacterium, commonly found in environments too inhospitable for most organisms [153]. The bacterium is closely related to another halophilic proteobacterium, *Chromohalobacter salexigens* [11], of the family Halomonadaceae (69% sequence identity), such that it was assumed to be one and the same until recently.

Halophilic organisms live in saline environments, such as salt lakes, coastal lagoons and man-made salterns with more than 0.5 mol/l NaCl concentration [58]. Some can even survive entrapment in salt rock [179]. Given the preferences, they are challenged by two stress factors, low water potential and high inorganic ion concentrations. The higher water potential inside the cell causes water efflux to abolish the water gradient. This would cause most cells to shrink and die. *H. elongate*, however, uses one of two known osmoadaption mechanisms to cope with fluctuations in the osmolarity of the growth medium: The organic-osmolyte mechanism.

6.1.1 *Salt-in cytoplasm vs. organic-osmolyte mechanism*

The two known mechanisms for osmoadaption are the salt-in-mechanism and the organic osmolyte mechanism [46, 58, 85, 153].

SALT-IN CYTOPLASM MECHANISM The salt-in mechanism was discovered in *Halobacteria* [91] and seems to be the typical archaeal strategy of osmoadaption [39, 80, 92, 168]. Although the environment of halophilc organisms contains great amounts of NaCl, sodium (Na^+) is not accumulated in the cell. The organisms instead allow potassium (K^+) to float in until the molar concentrations are balanced with its counter ion, chlorid (Cl^-) [44, 93]. There are, however, some anaerobic halophilic bacteria known to accumulate either K^+ or Na^+, depending on the growth phase [117, 134]. The saline cytoplasm requires that most of the enzymes in the cells are enriched in acidic amino acids and, at the same time, they are strictly dependent on K^+ and or Na^+ for activity [46].

ORGANIC-OSMOLYTE MECHANISM The organic-osmolyte mechanism is widespread among bacteria and eukarya and also present in some methanogenic archaea [91]. These organisms keep the cytoplasm to a large extend free of KCl and the design of the cells interior remains basically unchanged [101, 138, 174], although they allow for a short-term influx of K^+ after a sudden increase in salinity [85]. This cation serves as a cellular messenger and activates the accumulation of highly water-soluble organic compounds (like sugars, polyols, amino acids and/or amino acid derivatives) by either *de novo* synthesis or by uptake from the environment. These enable the maintenance of the osmotic equilibrium with the environment [135]. Because the organic compounds do not disturb the cell's metabolism, they are called 'compatible solutes' [25]. Compatible solutes effectively stabilize proteins and even whole cells by mitigating detrimental effects of freezing, drying and high temperatures [22, 95]. The most common compounds are the amino acid derivatives glycine-betaine and ectoine [116, 135] and more procaryotes are able to synthesize the latter, than the former, e.g. *H. elongata* [56, 135, 156]. Some bacteria can also use ectoine as an energy source [171] or to stabilize whole cells against stresses, such as UV radiation or cytotoxins [26, 55, 77, 83]. This property makes ectoine a valuable ingredient for health and skin care products.

Concluding, halophiles, which use the organic-osmolyte mechanism, are more flexible, given the wide range of salt concentrations in which they can prosper (3 to 30%) [174].

The ability of halophilic organisms to withstand extreme stress factors attracts interest from the scientific community and establishes significant utility for processes and applications in biotechnology. Organic osmolytes, such as ectoine, establish osmotic equilibrium and stabilize biological structures. Already in the early 1990's novel biotechnological processes were successfully developed using halophilic bacteria, such as *H. elongata*, as producing strains for organic osmolytes [91].

6.1.2 *Ectoine*

Ectoine (1,4,5,6-tetrahydro-2-methyl-4-pyrimidinecarboxylic acid) (Figure 22) is a compatible solute of high interest for the pharmaceutical and cosmetic industry. Examples include protection of organisms, organs, tissues, cells or the organic building blocks thereof from chemical radicals and oxidatively active compounds [152]; prevention and treating of gastrointestinal and amyloidal diseases [151]; and stabilization of skin moisture content [107].

Ectoine was first discovered in the extremely halophilic phototrophic bacterium *Ectothiorhodaspira halochloris* [56] and has been characterized by ^{13}C-NMR spectroscopy, mass spectrometry and infrared spectroscopy (IR). Because it dissolves in water quite easily (up to 6 mol/l water at 4°C), it possesses a non-ionic character at physiological pH-values and hence meets the requirements for an organic osmolyte [121].

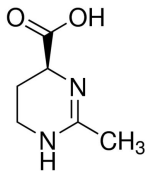


Figure 22: Ectoine ($C_6H_{10}N_2O_2$) [158]

The production of ectoine through the organic-osmolyte mechanism prevents the cell from being shrunk by osmosis. This allows halophilic organisms to live in environments with very high salt concentrations (10 times higher than sea water), without disturbing the cell's metabolism, even at high cytoplasmic concentrations. For instance, *H. elongata* can tolerate salt concentrations well above 10% (1.7 mol/l) NaCl [153].

ECTOINE SYNTHESIS Ectoine is the main compatible solute of *H. elongata* [156]. It is synthesized from L-aspartic-4-semialdehyde, which is an important synthesis-intermediate in the aspartate family of amino acids [115, 121, 153].

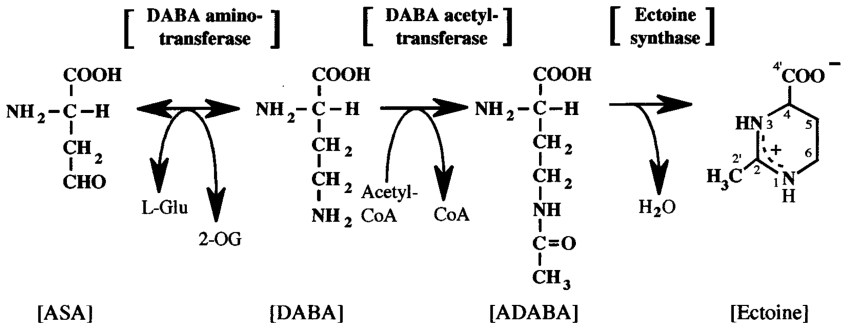


Figure 23: Ectoine synthesis [115]

The ectoine synthesis consists of three steps, as displayed in Figure 23: L-aspartic-4-semialdehyde (ASA) is converted into L-2,4-diaminobutyrate (DABA) via transamination and DABA is in turn converted to N-γ-acetyl-diaminobutyrate (ADABA) by acetylation with acetyl coenzyme A (AcCoA), which then yields ectoine by circularization. The three enzymes involved in this are DABA amino-transferase, DABA acetyltransferase, and ectoine synthase in order of reaction [115]. Ectoine can also be utilized as both a carbon and a nitrogen source by *H. elongata*. The doeABCD (**d**egradation **o**f **e**ctoine ABCD) genes are responsible for its degradation [153].

The following sections deal with the metabolic network of *H. elongata*.

6.2 BASIC MODEL

The corner stones of the model for *H. elongata* have been laid by the automatic and manual annotation of Schwibbert et al., with the help of the KEGG and BRENDA database. They built a so-called core-model, which reflects the conversion of glucose to ectoine, without the incorporation of growth, because

they aimed to explore the different metabolic options to produce ectoine via FBA. The core-model is the product of the careful reduction of a richer model with several reactions, which are not relevant for ectoine production from a sugar-source. Similar models exist for organisms which are highly related to *H. elongata*, for example *E. coli*. However, in order to explore some newly identified metabolic differences, the model has been adjusted. A map of the metabolic pathways is displayed in Figure 24.

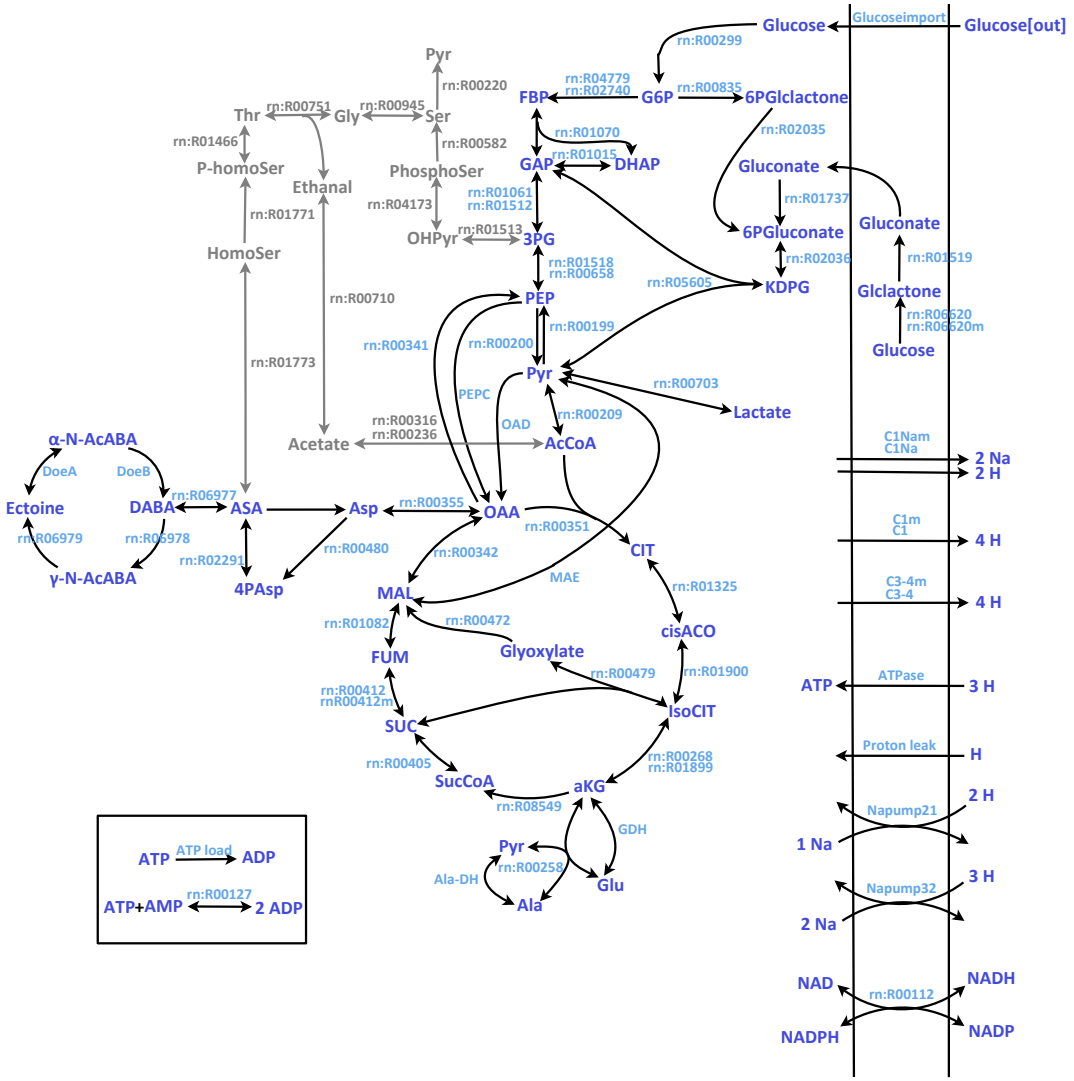


Figure 24: Pathway map of the core-model. The grey reactions and metabolites are included in the model but are only active in case that growth is considered. They therefore play a minor role. All other reactions are displayed in light blue, while the metabolites are displayed in dark blue. Most reactions are named like their KEGG identifiers. The long versions of the metabolite names, as well as the model equations can be found in Appendix B.

The subsequent sections deal with important aspects of the core-model of *H. elongata*.

6.3 ANAPLEROTIC REACTIONS

Since the product, ectoine, could not be produced without the use of at least one of the anaplerotic reactions, they play a major role in the metabolism of *H. elongata*. Moreover, the importance has already been pointed out in [153]. Therefore, their role will be examined in this analysis.

The anaplerotic reactions of *H. elongata* are the *oxaloacetate decarboxylase (OAD)* (EC 4.1.1.3, rn:R00217), the *PEP carboxylase (PEPC)* (EC 4.1.1.31, rn:R00345) and the *malic enzyme (MAE)* (EC 1.1.1.38, rn:R00214). They count as anaplerotic reactions, because they appear to replenish carbon in the TCA-cycle, compare Figure 24. The single reactions are explored below.

6.3.1 *Oxaloacetate Decarboxylase (OAD)*

The oxaloacetate decarboxylase is a membrane-bound pump, which imports sodium into the cell:

$$\text{Pyruvate} + 2\,\text{Na}_{\text{out}} + \text{CO}_2 \quad \Longrightarrow \quad \text{Oxaloacetate} + 2\,\text{Na}_{\text{in}} + \text{H}_2\text{O}$$

This fact is quite important when it comes to FBA, because the sodium has to be exported again, which requires energy. However, because it follows the conversion of PEP to pyruvate, two ATPs per glucose molecule can be gained initially. This issue will be explained in more detail in the following chapter.

In contrast to *H. elongata*, *E. coli* does not exhibit any form of OAD, which is an important difference between the two organisms.

6.3.2 *PEP Carboxylase (PEPC)*

The model includes two enzymes, which catalyze the direct interconversion of PEP and oxaloacetate: The *PEP carboxylase* and the *PEP carboxykinase* (EC 4.1.1.49, rn:R00341). The PEP carboxylase transforms pyruvate to oxaloacetate and therefore participates in the production of ectoine from sugar:

$$\text{PEP} + \text{CO}_2 \quad \Longrightarrow \quad \text{Oxaloacetate} + \text{Phosphate}$$

The PEP carboxykinase is assumed to catalyze the reverse direction, since the forward direction is kinetically unfavorable for this enzyme [153, 185].

6.3.3 *Malic Enzyme (MAE)*

The malic enzyme connects pyruvate with malate, which is then processed to oxaloacetate via the malate dehydrogenase (EC 1.1.1.37, rn:R00342). As in the case of OAD, two ATP units are gained with the interconversion from PEP to pyruvate beforehand. This enzyme is usually known to catalyze the reverse direction but the FBA done by Schwibbert et al. [153] assumed that the forward direction would be energetically preferable. To test, if this hypothesis is actually thermodynamically feasible, is one of the objectives of this work. The reaction

is defined as follows:

$$NADPH + CO_2 + Pyruvate + H \quad \Longrightarrow \quad NADP + Malate + H_2O$$

Not only the direction of the reaction differs to the enzyme used by *E. coli* but also the cofactors. Instead of NAD and NADH it is, based on personal communications from Viktoria Kindzierski, assumed that NADP and NADPH are used.

6.3.4 *Glyoxylate Shunt (GLX)*

There exists a fourth option to replenish the TCA-cycle: The glyoxylate shunt (rn:R00479, rn:R00472). These enzymes represent a short-cut through the TCA-cycle and because some of the carbon is lost before it reaches oxaloacetate, the GLX is not considered a real anaplerotic reaction but rather like an anaplerotic strategy, which is nevertheless important.

6.4 PHOSPHOFRUCTOKINASE (PFK VS. PPI-PFK)

Glucose can be metabolized in different ways. Common in bacteria is the Embden-Meyerhof-Parnas pathway. Other organisms prefer the Entner-Doudoroff pathway, which is similar with respect to the overall scheme, see figure Figure 25.

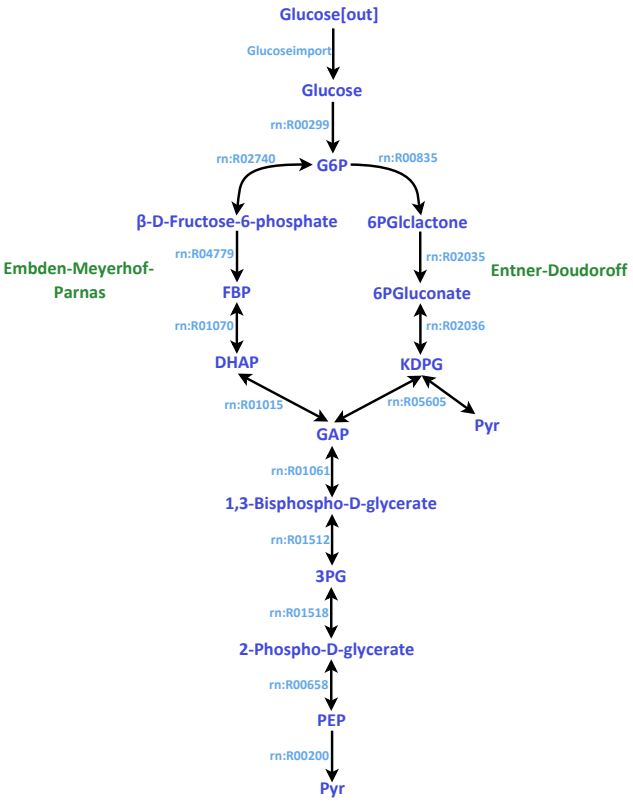


Figure 25: Embden-Meyerhof-Parnas and Entner-Doudoroff pathway.

The Entner-Doudoroff pathway is versatile. Different organisms exhibit different expressions: Besides the constitutive pathway (e.g. in *Z. mobilis*), there exist the linear pathway (e.g. in *E. coli*) and the cyclic pathway (e.g. in *Pseudomonas*) [32]. The annotation raised the question, if *H. elongata* uses the linear or the cyclic pathway. They differ in the way glycolysis is executed. In the case of the linear pathway, the Entner-Doudoroff pathway is only used for metabolization of certain carbohydrates, as gluconate, while the ordinary carbohydrate metabolism proceeds via the Embden-Meyerhof-Parnas pathway. In the case of the cyclic pathway, the core of the central carbohydrate metabolism is formed by the Entner-Doudoroff pathway and the Embden-Meyerhof-Parnas pathway is used in the reverse direction. The basic model was initially built with the linear pathway but experimental evidence has been generated that indicates the presence of the cyclic version. The difference lies in the phosphofructokinase. The 'normal' one is an ATP-dependent enzyme (PFK) (EC 2.7.1.11, rn:R04779). An alternative is represented by the pyrophosphate-dependent phosphofructokinase (PPi-PFK) (EC 2.7.1.90, rn:R00764). One hint that indicates it could be the second enzyme has been given by the closest relative of *H. elongata*: *Chromohalobacter salexigens*. This bacterium exhibits the PPi-PFK and utilizes therefore a cyclic pathway. Another hint has been given by the absence of the D-fructose-1,6-bisphosphate phosphatase (FBPP, F6P + Pi <===> FBP + H2O) (EC 3.1.3.11, rn:R00762), which usually accompanies PFK. This enzyme is needed to catalyze the reverse direction, since the PFK is highly unlikely to progress backwards. The two PFK options are defined by the following reactions:

PFK: F6P + ATP $\Longrightarrow$ FBP + ADP

PPi-PFK: F6P + PPi $\Longleftrightarrow$ FBP + Pi

If the ATP-dependent PFK is missing in the metabolism, it indicates that the Embden-Meyerhof-Parnas pathway cannot proceed in a forward direction, because the PPi-PFK catalyzes the reverse direction. The Embden-Meyerhof-Parnas pathway is therefore not an option to metabolize the sugar-source, which only leaves the Entner Doudoroff. Thus, certain precursors, which are needed for biomass production are not build. This is overcome by the backward-catalysis of the Embden-Meyerhof-parnas pathway. For that reason, this version of the pathway is called cyclic [32].

6.5 RESPIRATORY CHAIN

The respiratory chain of *H. elongata* matches the one of *E. coli* in most parts. However, *H. elongata* has one additional channel, the sodium-motive NADH:ubiquinone oxidoreductase (EC 1.6.5.8, $C1^{Na}$) [21]. This channel is similar to NDH-1-type NADH:ubiquinone oxidoreductase (EC 1.6.5.3, complex I or C1), which is also present in *E. coli*. They both translocate protons across the inner membrane per molecule of oxidized NADH, helping to build the elec-

trochemical potential difference used to produce ATP. $C1^{Na}$ exports two protons and two sodium molecules instead of four protons, as C1 does, compare Figure 26. It is for instance needed to export the sodium, which is imported through OAD. The same applies for the two antiporters (the orange channels in named figure). Complex III and IV have been combined to *C3-4*.

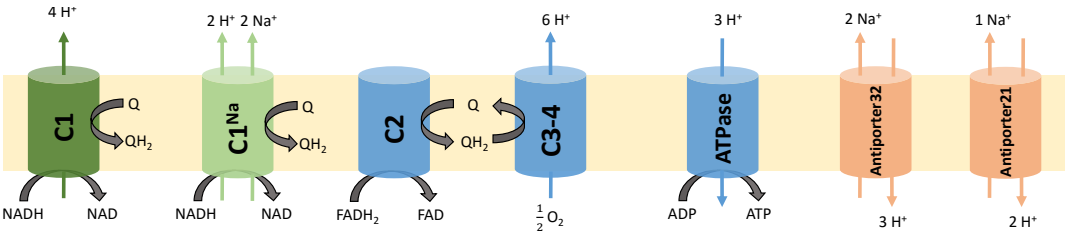


Figure 26: The respiratory chain. The blue channels are the complexes II-IV of the respiratory chain, as well as the ATPase. Complex III and IV have been combined to one: C3-4. The dark green channel is complex I (C1), the light green channel is the additional complex I channel in *H. elongata*, which exports sodium out of the cell. The antiporters, responsible for the exchange of protons and sodium are marked in orange. The inputs to this module are $FADH_2$ and NADH, while FAD and NAD serve as outputs.

6.5.1 *P:O ratio*

The P:O ratio of an organism describes the number of ATP-molecules, which can be obtained by oxidative phosphorylation per molecule of oxygen. This parameter cannot be derived from the genome annotation.

The current model does not fix the P:O ratio but allows it to manifest itself through the flux distribution in the FBA. Its value can be extracted from the FBA by means of the ratio of the fluxes *ATP load* and *C3-4*. To prevent the P:O ratio from having an unreasonably high value it can be controlled by either the reaction *proton leak*, which simulates a proton leak into the cell, or by the antiporters.

6.5.2 *Glucoseimport*

The original model assumed that glucose, as well as fructose, are imported through the PTS-system. New annotation results, however, show that the PTS-system is only used for fructose. Glucose is imported directly and afterwards phosphorylated to α-D-glucose-6-phosphate through the glycokinase-reaction (EC 2.7.1.2, rn:R00299).

6.5.3 *Quinones*

The original basic model did not differentiate between the different kinds of quinones and always used ubiquinone/ubiquinol. However, it is now known that, e.g. *E. coli*, uses menaquinone/menaquinol as well [4, 170]. Its presence

in *H. elongata* should therefore be considered and analyzed. In general, aerobic gram-positive bacteria contain menaquinone, most gram-negative types contain ubiquinone, but a few gram-negatives, like *E. coli*, contain roughly equal amounts of both quinones [120]. The difference between both quinones is the redox potential: Menaquinone has a very low redox potential (-74 mV) compared to that of ubiquinone (100 mV) [54]. They both mediate the electron transport reactions in the respiratory chains of bacteria. Because a preference for either quinone is not yet clear, all reactions, which use ubiquinone so far (which includes the reactions of the respiratory chain) have been defined again with menaquinone/menaquinol. The glucose dehydrogenase, which is a membrane bound enzyme concerning the Entner Doudoroff, accepts electrons from the pyrroloquinoline quinone (PQQ), which in turn interchanges electrons with menaquinone or ubiquinone, dependending on the growth conditions [88].

Moreover, the succinate dehydrogenase reaction (EC 1.3.99.1, rn:R00412) had to be changed because so far it used FAD/$FADH_2$ as substrate/product, which is not possible because they are covalently bound to the enzyme. Ubiquinone/ubiquinol or menaquinone/menaquinol are actually substrate and product and therefore the reaction changes to [110]:

$$\text{Succinate} + \text{Q} \quad \Longrightarrow \quad \text{Fumarate} + \text{QH}_2$$

$$\text{Succinate} + \text{MQ} \quad \Longrightarrow \quad \text{Fumarate} + \text{MQH}_2$$

All reactions, which are newly defined with menaquinone/menaquinol, instead of ubiquione/ubiquinol are marked by $\mathfrak{m}$, like $\mathfrak{rn} : \mathtt{R00412}\mathfrak{m}$.

6.5.4 *Additional reactions without gene assignment*

Some processes, like the consumption of ATP for the maintenance of the cell, are not modeled. Nevertheless, they are accounted for through the addition of extra reactions. One such reaction, which describes the *ATP-hydrolysis* is the *ATP load*. Another one is called *Proton leak*. This reaction is a proxy for all reactions, which cause a proton influx, while skipping the respiratory chain, like uncoupling agents.

Moreover, a *transhydrogenase* reaction has been added to account for the interconversion of NADH and NADPH. *E. coli*, e.g., has two transhydrogenase versions, the UdhA and the PntAB. One is soluble and the other is membrane bound and therefore proton-translocating [139]. Recent annotation results showed that *H. elongata* only owns the PntAB transhydrogenase, which is therefore incorporated into the model with its KEGG-code rn:R00112:

$$\text{NADPH} + \text{NAD}^+ + \text{H}_{\text{in}} \quad \Longleftrightarrow \quad \text{NADP}^+ + \text{NADH} + \text{H}_{\text{out}}$$

NADH and NADPH are quite different function wise, because NADH serves energy-production as the main respiratory cofactor and can be recovered through the TCA-cycle, while NADPH attends anabolic reduction reactions (i.e. biosynthesis). NADPH can be recovered via the Pentose-Phosphate-Pathway, the isocitrate dehydrogenase and the transhydrogenase reaction.

6.6 INCORPORATION OF BIOMASS PRODUCTION INTO THE MODEL

The core-model of *H. elongata* consists of the main elements for ectoine production from sugar but it does not take into account the possibility of cell growth, which is a typical objective of every cell. Ectoine production and cell growth are competitive and cannot be satisfied equally at the same time. Therefore the comparison of both goals becomes an interesting point to pursue. In order to do so, the reactions for the precursors of biomass and the growth reaction itself have to be introduced. Several precursors for biomass production are build in the glycolysis but not all. Because the exact composition of the biomass of *H. elongata* is not known so far, this analysis will be mostly based on *E. coli*, which is also gram-negative and thus a well-studied representative. Where information on the composition of certain biomass precursors is available, they are modified to reflect *H. elongata*.

The biomass of a bacterial cell is primarily build up from seven components [109]:

- Protein (55%)
- DNA (3.1%)
- RNA (20.5%)
- Lipids (9.1%)
- Carbohydrates (2.5%))
- Peptidoglycan (2.5%)
- Lipopolysaccharides (3.4%)

These components have to be broken down into their precursors, until all interconnections between the core-model and the new components are identified and linked. This is essential for the subsequent use of FBA.

The procedure has been roughly pictured by Feist et al. in Figure 27. As one can see, the units change from mmol/g Dw to g/g Dw as soon as the macromolecular level is reached (Dw stands for dry weight). This is due to the fact that biomass cannot be measured in mmol/g Dw. The units therefore have to be adjusted.

In the following, the different precursors of biomass will be considered, the exact calculations and reactions can be found in the appendix. The components, which add up to biomass as depicted in Figure 28.

6.6.1 *Protein*

The proteins consist of amino acids, connected to the core-model through metabolites as oxaloacetate, which is the precursor of the aspartate family. Figure 29 summarizes the different connections.

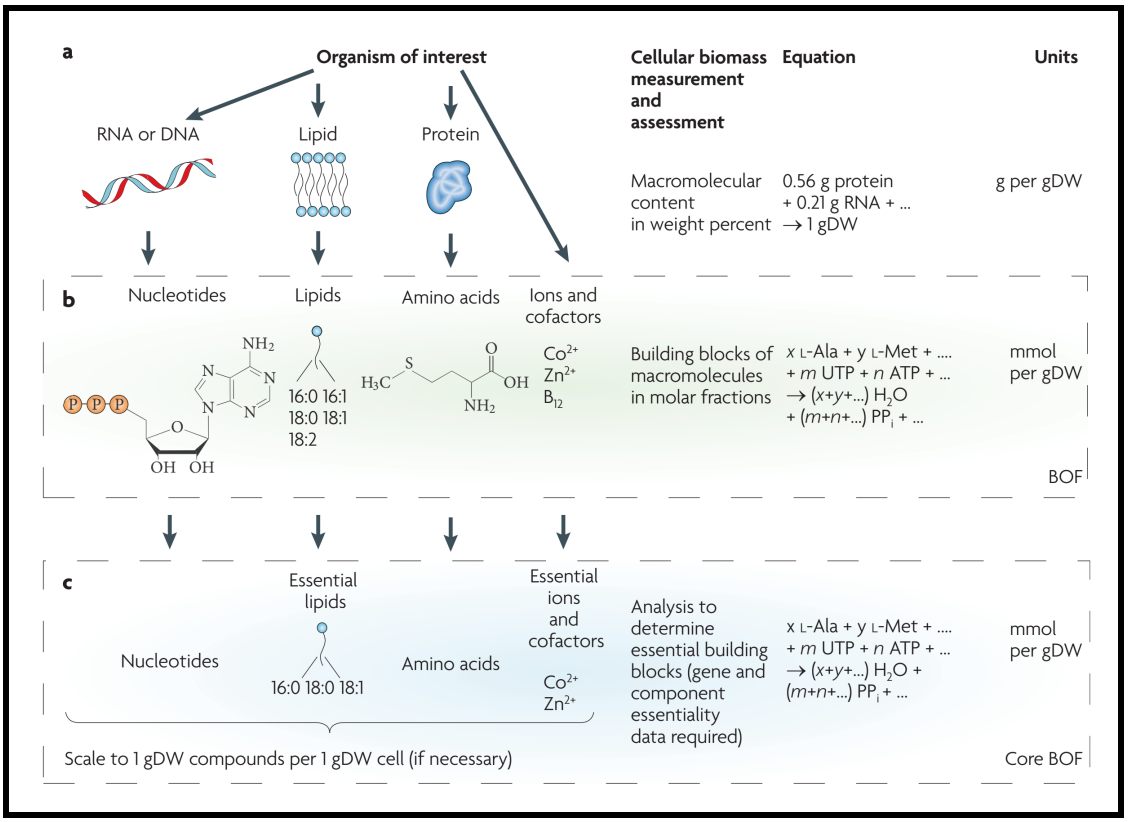


Figure 27: Procedure to gain the biomass composition according to Feist et al. [50]. In a first step, (a), the macromolecular weight percent contribution for each component needs to be determined. The data with respect to *E. coli* can be found in [109]. Each macro molecule is then broken down into its building blocks (b) (BOF stands for biomass objective function). These building blocks in turn are also broken down to establish a connection to the existing core-model (c).

The synthesis reactions of the amino acids have been looked up in the KEGG database and been compared with the stoichiometries in other *E. coli* models [131, 167]. It was thereby implied that the two organisms would not be distinguishable with respect to their amino acids. Neidhardt published a table, which depicts the content of each amino acids in a general protein (in mol %) [109]. The reaction, which has been introduced to reflect this relation is *BmAminoacidreaction*. Stephanopoulos provided the formula to calculate 1 g of protein per g Dw [164], which is based on the molecular weights of the components:

39.1 mmol ATP + 9.1 mmol Aminoacids $\Longrightarrow$ 1 g Protein + 39.1 mmol ADP + 39.1 mmol P_i

6.6.2 *DNA and RNA*

DNA, as well as RNA, are composed of nucleotides. Again, the synthesis reactions have been looked up in KEGG but also in Lehninger. The formulas are therefore once more based on *E. coli*, given the close relationship to *H. elon-*

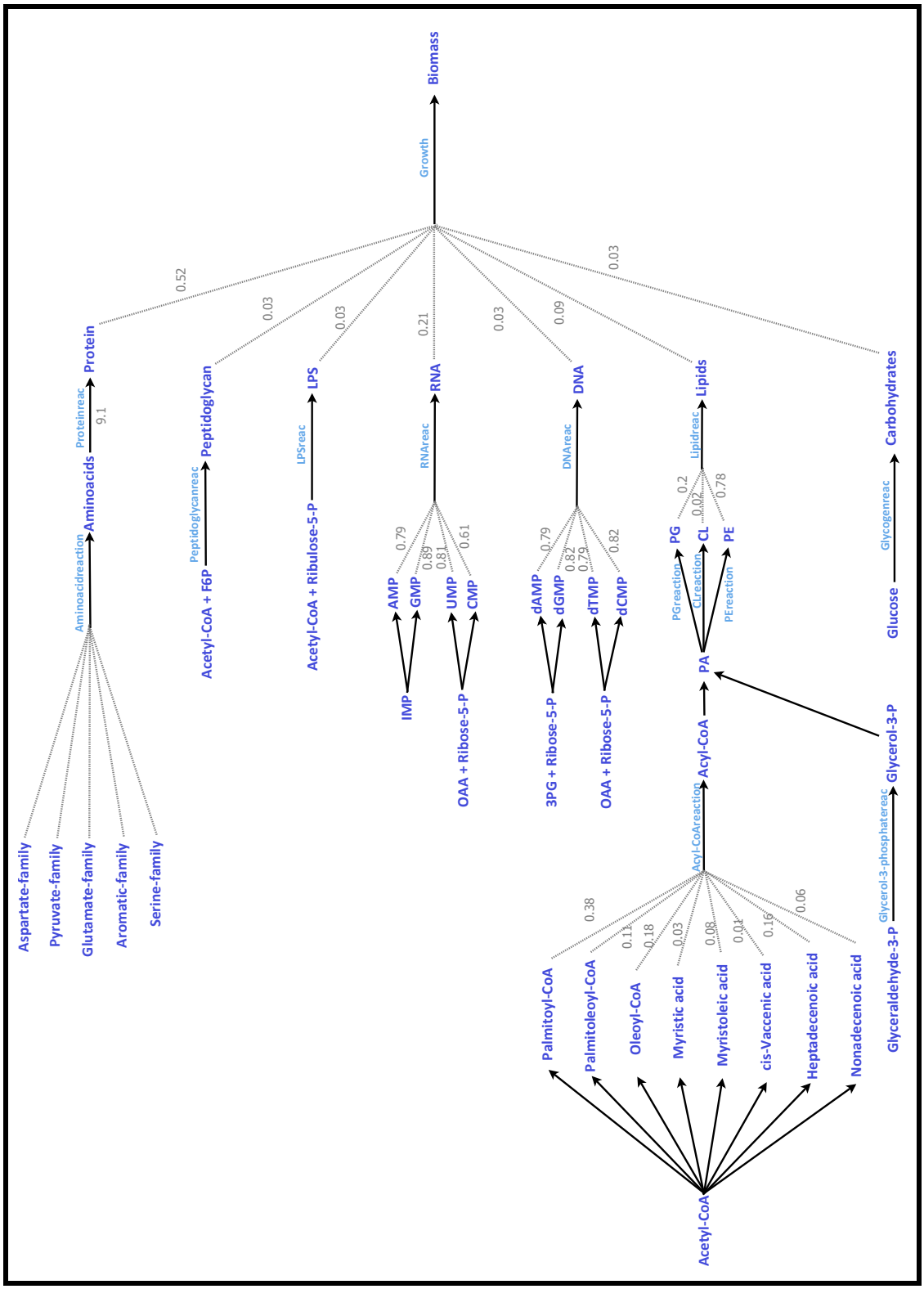


Figure 28: Pathway map for biomass synthesis. Metabolites are once again displayed in dark blue, while reactions are displayed in light blue.

gata. It can be adjusted easily as soon as new data indicate differences. Again,

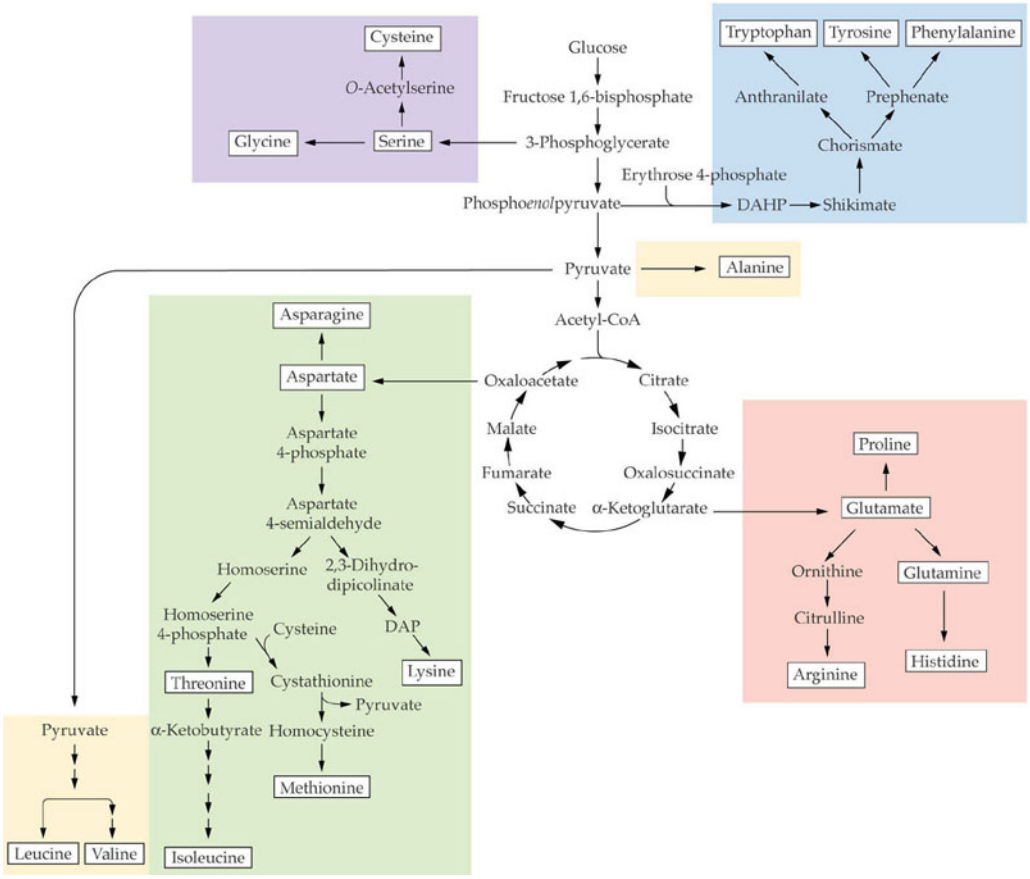


Figure 29: The amino acid families and their connection to the core-model of *H. elongata* [35]. Green: Aspartate-family; Yellow: Pyruvate-family; Orange: Glutamate-family; Blue: Aromatic-family; Purple: Serine-family.

Stephanopoulos provided formulas to calculate 1 g of DNA and RNA per g Dw. The new reaction formulas are *BmDNAreac* and *BmRNAreac* (see appendix).

The following subcomponents of biomass provided a greater challenge.

6.6.3 *Lipids*

The basis of most lipids is acetyl-CoA because it is needed to assemble fatty acids. Depending on the length of the chain and the kind of carbon bounds, one distinguishes different kinds of fatty acids. The active form is called *acyl-CoA*, which is a generic term, like amino acids. The proportion of each fatty acid in acyl-CoA had been collected in [131]. All fatty acids have been incorporated in one single reaction, *BmAcyl-CoAreaction*. To finish the process of lipid biosynthesis, the fatty acids have to be provided with their head groups. The final phospholipids are *phosphatidylethanolamine (PE)*, *phosphatidylglycerol (PG)* and *cardiolipin (CL)*. Pramanik and Keasling provided a phospholipid composition as well but in this case, data from *H. elongata* exist [178]. Because the synthesis reactions of PE, PG and CL are already calculated in g/g Dw, no further conversions are necessary.

6.6.4 *Carbohydrates*

The carbohydrate, which is considered, is glycogen. Glycogen consists of glucose. The exact amount used for glycogen production can be calculated: Because glucose has a molecular weight of 180 g/mol, 1 g of glycogen consists of 5.55 mmol glucose. The synthesis reaction therefore is:

5.55 mmol (Glucose + ATP - ADP - Pi) $\Longrightarrow$ 1 g Glycogen

6.6.5 *Peptidoglycan*

Peptidoglycan is a polymer, which forms a layer outside the plasma membrane of most bacteria. It usually consists of N-acetylglucosamine (NAG), N-acetylmuramic acid (NAM), Meso-diaminopimelate and different amino acids [99]. The composition of the amino acids differs from organism to organism. The one of *H. elongata* can also be found in [178]. The amino acids present in the peptidoglycan of *H. elongata* are leucine, glycine, glutamate and alanine. The calculation of the synthesis reaction can be found in the appendix.

6.6.6 *Lipipolysaccharides*

Lipopolysaccharides are a major component of the membrane of gram-negative bacteria. Its structur varries among different kinds of bacteria. Because there is no information available on its composition within *H. elongata*, *E. coli* is used as a reference [167]. The derivation of the molecular mass and the synthesis equation is displayed in the appendix.

6.6.7 *Biomass synthesis*

With all precursors available in g/g DW it is possible to define the synthesis equation of biomass, which is called *BmGrowth*. The formula can be derived from [109]:

0.031 DNA + 0.205 RNA + 0.55 protein + 0.091 lipids + 0.025 carbohydrates +

0.025 peptidoglycan + 0.034 lipopolysaccharides $\Longrightarrow$ 1 g biomass

Without the incorporation of biomass, the core-model consists of 79 reactions and 80 metabolites. The inclusion of biomass increases the number of reactions to 154 and the number of metabolites to 160.

7

FLUX BALANCE ANALYSIS (FBA)

Flux Balance Analysis is a suitable way to start the mathematical analysis of a biochemical system, because it provides insight into the processes within a network, without demanding too much information in advance. It allows the modeler to gain an idea of the capabilities of a system, e.g. the theoretical production limits, based on the model. The model of *H. elongata*, with biomass production included, is analyzed under different perspectives: First, with respect to ectoine production; second, with respect to growth to get an idea of how much carbon of the sugar-source could theoretically be used for biomass production. Third, ectoine is considered as a part of biomass, which shifts the focus to its influence on biomass production. Finally, a variable, proton leak, is examined, which allows the comparison with experimental data. The comparison of the two different glycolytic pathways is an objective throughout the whole FBA analysis. The anaplerotic reactions play an important role with respect to growth and ectoine production. Therefore, their role is examined as well.

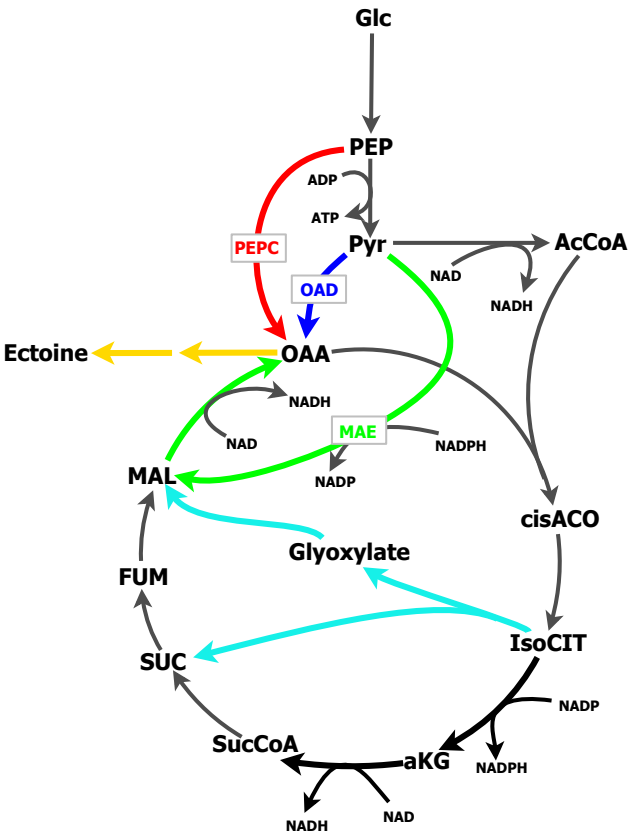


Figure 30: The figure displays a shortened pathway version (EMP pathway used), which lies the focus on the anaplerotic reactions and energy production.

Figure 30 displays all the reactions of interest in different colors. These colors will keep their meaning throughout the chapter. The glycolytic pathway variant displayed is the Emden-Meyerhof-Parnas pathway. The Entner-Doudoroff pathway variant looks similar but the flux, coming from glucose, would end up in phosphoenolpyruvate (PEP), following glyceraldehyde-3-phosphate (GAP), as well as pyruvate, compare Figure 24. The reactions of interest stay the same.

7.1 MAXIMIZATION OF ECTOINE PRODUCTION

In the first analysis, the focus is on ectoine maximization alone and growth is not considered. As has been explained in the materials and methods part, it is possible to define certain constraints in FBA to see how the system behaves under these conditions. The objective (ectoine maximization) stays the same throughout this section. The value of the input, glucose, has to be fixed to allow for a certain flux through the network. For purpose of this analysis, the input-flux is set to 100 mmol/(g Dw h), which stays the same throughout the whole thesis and will serve as a reference for the yields. It is known that a cell needs a certain amount of energy to maintain its status quo. This requirement is represented by the ATP load reaction as discussed in the previous chapter. Limiting this reaction restricts, among others, the ATP demand of the cell. In this way, the optimization problem can be further constrained, which offers the possibility to gain an idea of the influence of the ATP load on the choice of direction through the whole network. In each optimization the ATP load reaction is given a value between zero and 800 mmol ATP/100 mmol Glc, although it is known that a cell usually does not need more than 10 mmol ATP/100 mmol Glc to stay alive [118]. For each fixed ATP load value (ν_{ATP}^{fix}) an optimization is executed and the values of the fluxes of interest are saved and afterwards depicted in a stacked area plot, which shows how the carbon flows through the pathway.

The optimization problem is formulated as follows:

$$
\begin{aligned}
&\max Z = v_{ectoine} \\
&\text{s.t:} \\
&\quad \mathbf{S}\,\vec{v} = \vec{0} \\
&\quad \vec{v_L} \leqslant \vec{v} \leqslant \vec{v_U} \\
&\quad v_{ATP}^{fix} \leqslant v_{ATP}
\end{aligned}
\tag{82}
$$

An exemplary stacked area plot of such an optimization is depicted in Figure 31 (a). The x-Axis displays the varying ATP load (in mmol ATP/100 mmol Glc). The y-Axis represents the fluxes the carbon takes on its way through the network. It can either end up in ectoine production through one of the anaplerotic reactions or is burned to satisfy the ATP load, which is represented by the black arrows in Figure 30 or the black area in (a). Note for clarification: The respiratory chain, which satisfies the ATP load, is activated through NADH, which in turn is gained through the reactions marked in black in Fig-

ure 30. These reactions also use up carbon for CO_2-production. Therefore, they represent the usage of carbon for energy production.

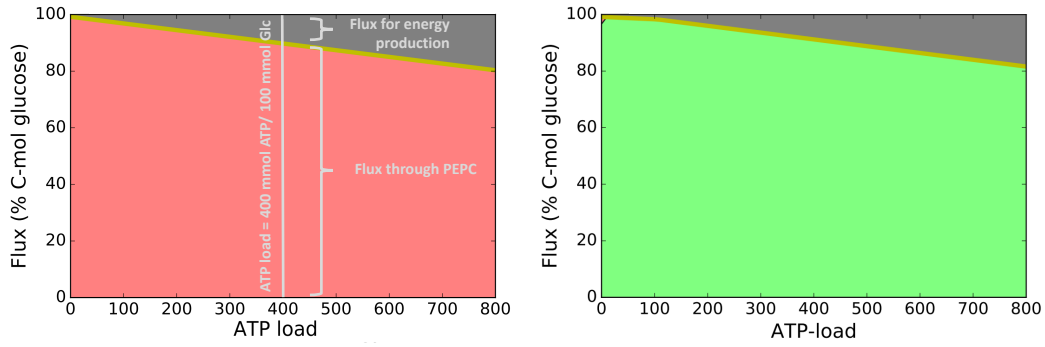


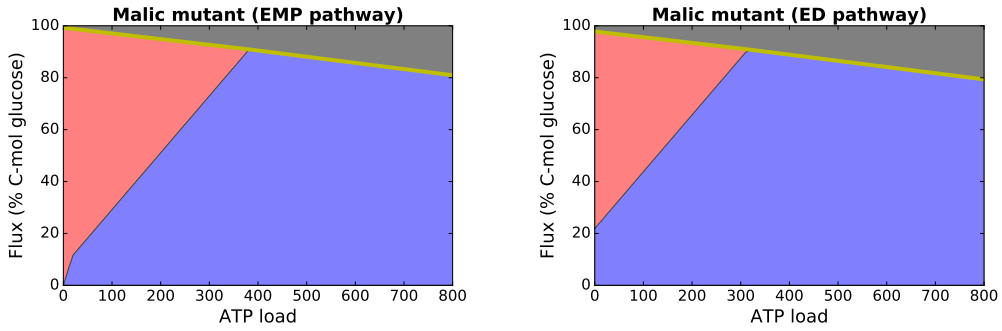
Figure 31: (a) is an exemplary stacked area plot, which explains the concept for ectoine optimization for different ATP load values (in mmol ATP/100 mmol Glc); (b) displays the flux distributions, which result, if malic enzyme is able to perform the transition from pyruvate to malate. For each ATP load value (per 100 mmol Glc) MAE is the superior flux-choice, if the ectoine-yield is maximized.

As can be seen in Figure 30, the anaplerotic reactions all flow into oxaloacetate, which has only one effluxes, one which ends up in ectoine. That is, the fluxes of the green, blue, red and cyan arrows add up to the value of the yellow one, which represents the amount of the produced ectoine. Because the carbon has to flow through at least one of the colored reactions to account for ectoine maximization, the stacked area plot does not exhibit white areas. In this way, each vertical line, which can be drawn in such a plot, accounts for an optimization with fixed ATP load, the value is represented by the x-coordinate and reflects how much of the carbon ends up in ATP production (in carbon-mol %) and how much flows through either of the anaplerotic reactions and ends up in ectoine production. The line drawn in Figure 31 (a) can be interpreted as follows: 90% of the carbon influx flows through PEPC and ends up in ectoine production, 10% have to be burned to satisfy the ATP load of 400 mmol ATP/100 mmol Glc. The glyoxylate shunt is something special, because it serves energy as well as ectoine production. Therefore, part of its flux will add up to the yellow line, while part of it will be above, serving energy production.

Optimizing the EMP pathway for glycolysis results in the stacked area plot displayed in Figure 31 (b). It shows the flux distributions without any extra constraints. As can be seen , independent of the ATP load, malic enzyme seems to be the superior anaplerotic reaction for ectoine production. As long as the ATP load stays below 100 mmol ATP/100 mmol Glc, almost all the glucose is catabolized to ectoine, which is represented through the yellow line. As soon as this threshold is crossed, the carbon is further needed to satisfy the other ATP consuming reactions of the cell, represented through the increasing black area. Alternative solutions are possible, for instance is PEPC an acceptable alternative to MAE, as long as the ATP load reactions is fixed on a relatively low value.

Solutions, which are equally optimal with respect to their objective, are called *degenerated*. They are difficult to find, because FBA only returns one solution in each optimization, while there usually exist thousands of equally optimal solutions. It is hard to tell with FBA alone which solution is more reasonable and closer to reality.

As has been shown, malic enzyme appears to be the superior option when it comes to ectoine production, according to FBA. Still, it is known that the other anaplerotic reactions are important and need to be analyzed in more detail as well. The FBA results indicate that MAE is the preferred anaplerotic reaction, which does not allow for the examination of the other anaplerotic reactions. To accommodate the analysis, mutants are defined, which lack the ability to use malic enzyme. One such mutant is the malic mutant. To analyze the differences between OAD and PEPC as well, two double mutants are defined: the malic OAD mutant and the malic PEPC mutant.



(a) (b)

Figure 32: Figure displaying stacked area plots for the pathways with inactivated malic enzyme. (a) shows the stacked area plot, which has been created with the model, which uses the EMP pathway; (b) displays the stacked area plot, which has been achieved with the model, which uses the ED pathway.

Figure 32 displays the flux distributions of the malic mutant in case that either EMP or ED is active. It can be seen that OAD seems superior to PEPC, because it is the only remaining anaplerotic reaction at high ATP load values. This is caused by the ATP, which is gained through the PEP to pyruvate reaction. One might wonder, why OAD and malic enzyme are not equivalent, since both follow this reaction. The reason is sodium, which gains access to the cell, when OAD is active. As mentioned already, these cations have to be exported again, but the process is coupled to an import of protons. This proton-influx reduces the amount of protons, which are available to produce ATP with the ATPase. Thus, the export of the sodium indirectly costs energy. The exact amount lost depends on the membrane protein used to export the sodium. The different options are the sodium dependent complex I ($C1^{Na}$) of the respiratory chain and two antiporters, which exchange sodium with protons: the 1:2 Na^+/H^+ antiporter (Napump21) and the 2:3 Na^+/H^+ antiporter (Napump32), see Figure 33.

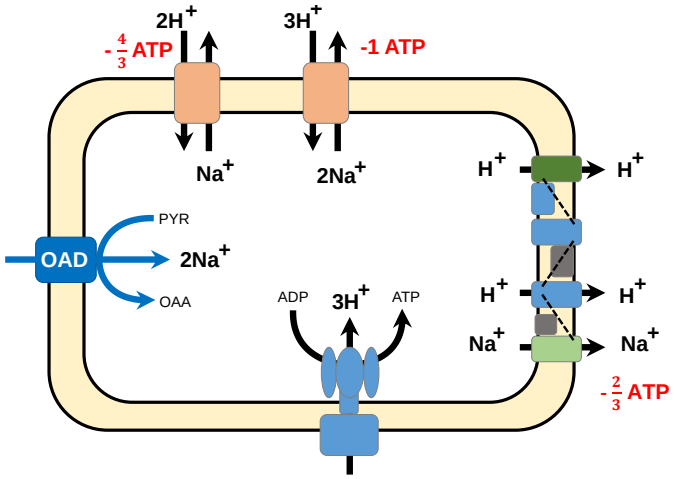


Figure 33: Sodium economy: The use of OAD is coupled to an import of two sodium cations, which have to be exported again. Three different options exist, which entail different costs (in red) with respect to energy. The respiratory chain, on the right, is the most energy-efficient option to export sodium (the colors correspond to the ones from Figure 26). The two antiporters (in orange) import protons and therefore are not very energy-efficient. The ATPase needs to import three protons to produce one ATP molecule.

The more efficient option is the respiratory chain, which uses $C1^{Na}$ instead of C1. Because the proton-export is divided in half, if $C1^{Na}$ is used instead of C1, the ATPase misses out on two of three protons per sodium-pair. Thus, the energy deficit equals $\frac{2}{3}$ ATP. The two antiporters exhibit a worse balance: The antiporter Napump32 exchanges the sodium-pair with three protons and therefore costs one ATP. The antiporter Napump21 makes the same exchange with four protons and therefore exhibits the highest opportunity costs, because the deficit is more than one ATP molecule. For this reason, the best case scenario is the use of $C1^{Na}$ and although the step from PEP to pyruvate results in the gain of one ATP, the export of sodium costs $\frac{2}{3}$ of it, which leaves an overall gain of $\frac{1}{3}$ ATP. The use of malic enzyme does not cost any ATP units and is therefore superior.

In the course of the analysis, a ranking could be established: Malic enzyme, catalyzing the reaction from pyruvate to malate, is the best option to produce ectoine, because of the extra ATP gained in the step from PEP to pyruvate. The next best option is the use of OAD, which still gains $\frac{1}{3}$ of an ATP unit. OAD is followed by PEPC, which lacks the step from PEP to pyruvate and therefore the gain of any ATP. The worst option is the glyoxylate shunt, because it does not offer any extra energy, but rather costs carbon in the TCA cycle, which cannot be used for ectoine production. Nevertheless, the glyoxylate shunt is a good choice for energy production, if only little carbon needs to be invested to satisfy the ATP load, as can be seen in Figure 34.

In this case, malic enzyme and PEPC are inactivated, which means that PEPC is no possible choice for a low ATP load anymore, as has been the case in Figure 32. Thus, OAD has to be used but this goes along with the use of the

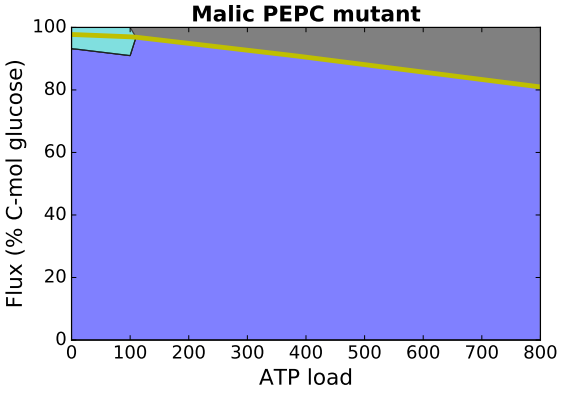


Figure 34: Malic PEPC mutant with active EMP pathway. Only OAD is left as an anaplerotic reaction. To use this enzyme, the respiratory chain has to be activated, in order to export sodium in the most efficient way. Therefore, some carbon needs to be burned, which can be achieved in a cheaper way by the use of the TCA shortcut: the glyoxylate shunt. As soon has the ATP load increases, more carbon needs to be burned and the shortcut is insufficient to satisfy the demand.

respiratory chain to export sodium. It lowers the ectoine-yield from the start. Instead of using the reactions marked in black, which would cost more carbon, the shortcut through the glyoxylate shunt is taken. Therefore, only 2 instead of 6 carbons are consumed in each turn. This is only valid, as long as the amount of energy needed is small. As soon as a certain threshold is reached, the energy-demand and therefore the NADH-demand is too high and more carbon has to be consumed.

As has already been indicated, the Embden-Meyerhof-Parnas and the Entner-Doudoroff pathway are compared as well. As can be seen in Figure 32, the flux distributions are slightly different, although the ranking stays the same. Still, the Entner-Doudoroff does not gain as much energy as the Embden-Meyerhof-Parnas. The energy supply is only half [52]. Therefore, the ectoine-yield is 1-2% lower throughout all considered pathway versions, which indicates that the ED pathway is inferior. This finding raises the question why *H. elongata* would prefer the ED over the EMP pathway. This question will be followed up in the thermodynamics.

7.2 MAXIMIZATION OF BIOMASS PRODUCTION

The next focus of the FBA is the biomass maximization. In order to be comparable to the former results, the same reactions as before are considered, although one has to keep in mind that their flux might be lower. Moreover, not the whole carbon flux will be depicted, resulting in white areas, because a major part ends up in biomass without the use of the anaplerotic reactions. The constraints for ATP load and glucose input stay the same as before.

As can be seen in Figure 35, the flux through the anaplerotic reactions is much lower, than before. As mentioned, a significant amount of carbon takes

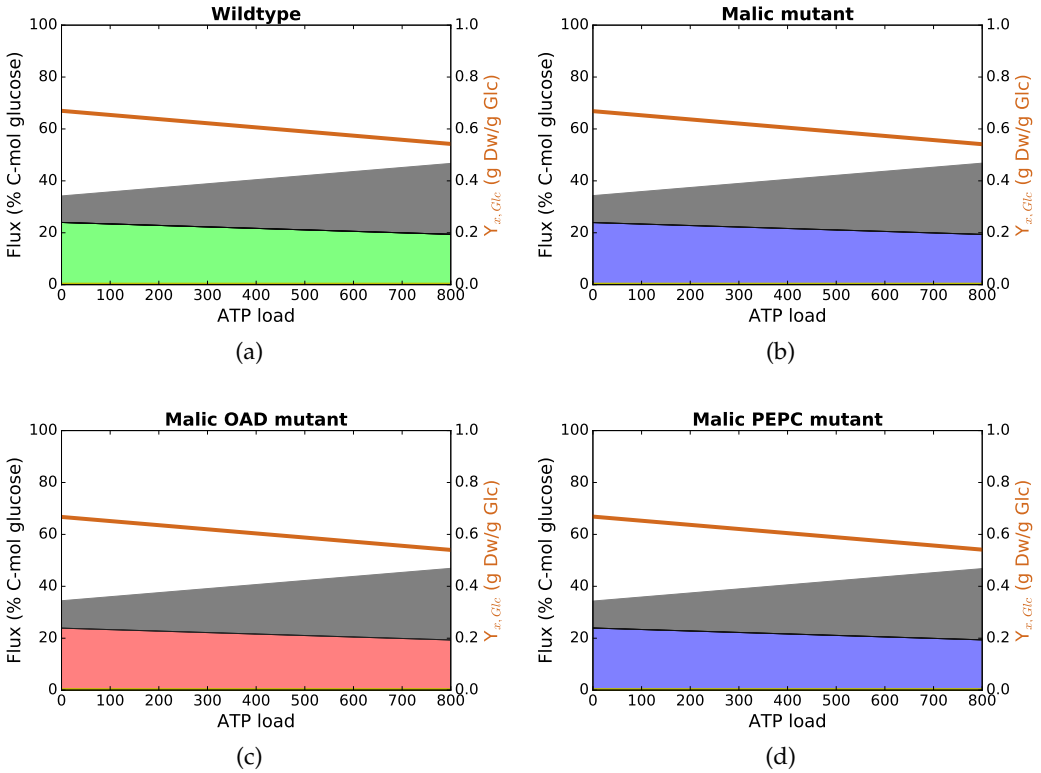
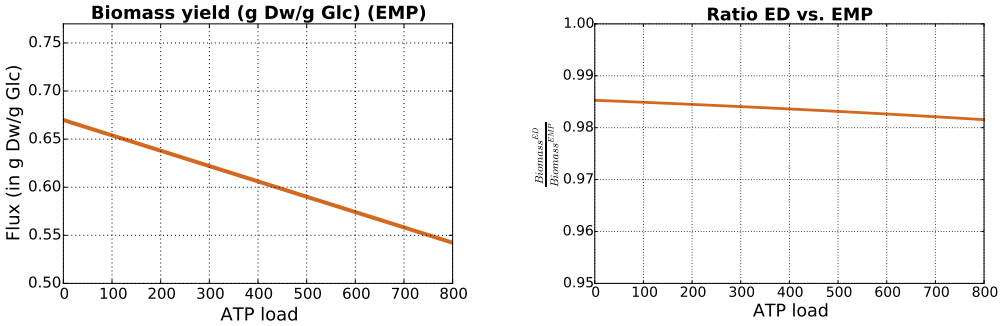


Figure 35: Optimization of biomass with an active EMP pathway. (a) wildtype; (b) malic mutant; (c) malic OAD mutant; (d) malic PEPC mutant. The brown line represents the biomass-yield with respect to the amount of glucose fed (100 mmol/g Dw, respective 18 g/g Dw). The biomass production cannot be depicted in mmol Dw/mmol Glc, because it cannot be measured that way. It is measured in g Dw/g Glc, which results in a yield considering gram. The maximal yield is 0.67 g Dw/g Glc, according to the analysis. The change in yield is depicted in brown.

other routes through the network to end up in biomass. Because there are so many, they cannot be displayed all together, resulting in a large unfilled white space. The yellow line, which showed how much ectoine has been produced, is replaced by a brown one, which represents the biomass-yield with increasing ATP load. The maximal value is 0.67 g Dw/g Glc. At an ATP load of 800 mmol ATP/100 mmol Glc, the yield is decreased by 20% to 0.54 g Dw/g Glc. As can be seen in the different figures, the mutants have no effect on the biomass-yield, the decrease is always the same. Moreover, because more precursors are needed for biomass production, more energy must be invested, which results in an increased CO_2 production, as represented by the black area. The ranking stays the same, although the glyoxylate shunt does not serve as an alternative flux anymore. Interestingly, the malic mutant ceases the use of PEPC. Previously PEPC was used as long as the energy demand was low and it was possible to prevent the use of the respiratory chain, which costs carbon. In this case, the respiratory chain is already needed for biomass production. Thus, it does not cost more carbon to use OAD at low ATP load.

The same analysis has been repeated with the Entner-Doudoroff pathway, instead of Embden-Meyerhof-Parnas and again, a 2% lower yield resulted. Figure 36 (a) represents the change in biomass in g Dw/g Glc, scaled by the molecular weight of glucose, with the use of the EMP pathway, while (b) displays the ratio of the optimal values, which results from the use of the ED or the EMP pathway. As one can see, the value is close to 1, because the difference is low, but it increases a bit with increasing ATP load. The flux profile stays the same, independent of the pathway choice. Therefore, the stacked area plots of the ED-version are not displayed.



(a) (b)

Figure 36: (a) displays the production of biomass in g Dw/g Glc, if the EMP pathway is active, while (b) illustrates the ratio between the production with the ED and EMP pathway.

7.3 OPTIMIZATION OF BIOMASS PRODUCTION WITH ECTOINE CONTENT INCORPORATED

The previous analysis considered ectoine and biomass production as independent of each other. In fact, ectoine is part of biomass, although its exact content is not known, because it varies with the salt concentration of the medium. Literature supports an assumption of up to 20% [42]. In order to examine the whole spectrum, the fraction parameter, α, is introduced as a variable. In the following, the ATP load reaction is set to 10 mmol ATP/100 mmol Glc, which appears to be a reasonable estimation, since similar values have been measured for *E. coli* [118]. Still, ATP is produced with the ATPase for synthesis reaction, etc.. The formula, which represents the incorporation of ectoine (in g), is the following [42]:

$$(1-\alpha)\,\text{biomass} + \alpha\,\text{ectoine} \quad = \quad \text{biomass with ectoine}$$

α is varied between zero and one and therefore between the two possible extremes of the formula: Biomass contains either no ectoine or consists only of ectoine. The flux distributions from before for an ATP load of 10 mmol ATP/100 mmol Glc can be found at the boundaries of the new stacked area plots. The principle stays the same, the carbon flux through the anaplerotic reactions and the energy production are displayed in C-mol % glucose, as well

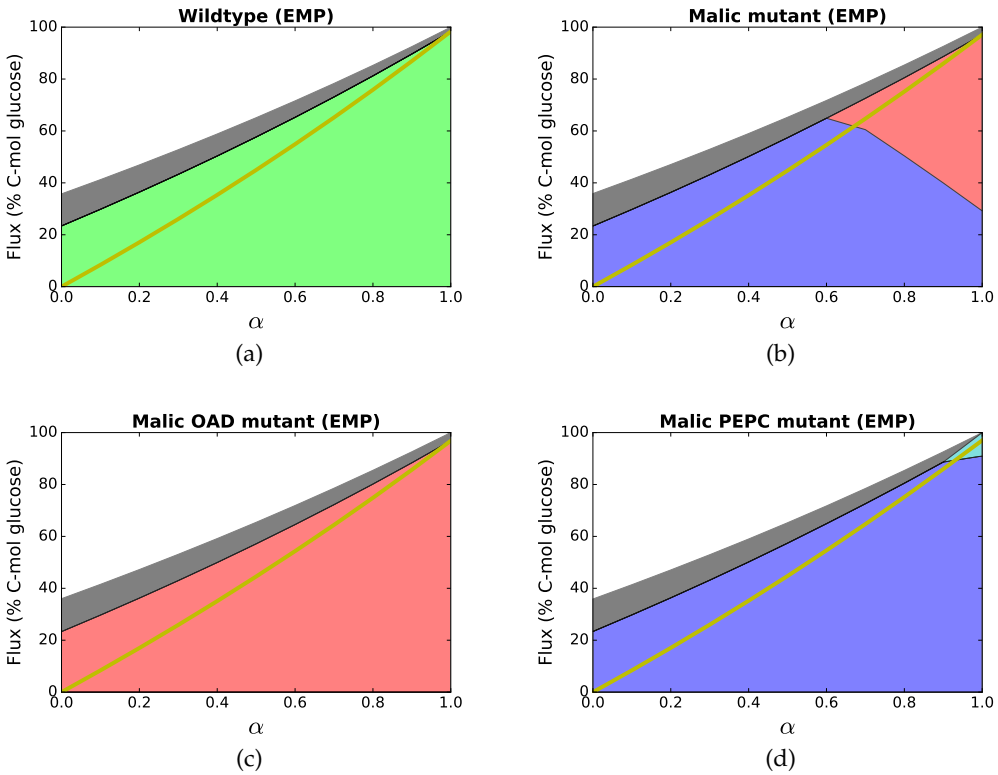
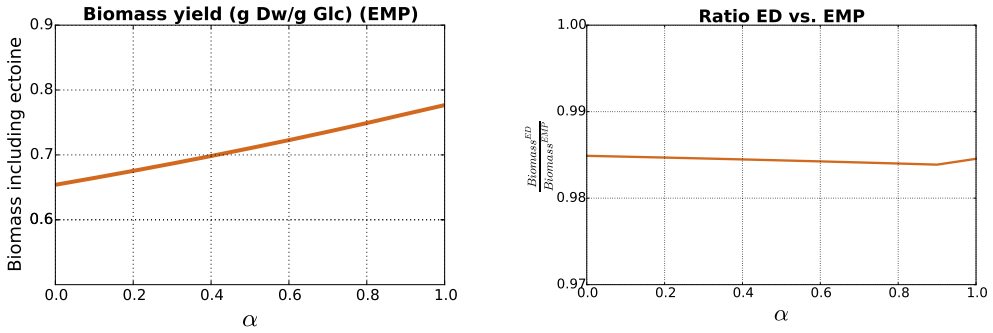


Figure 37: Stacked area plots for biomass optimization (including ectoine) with the EMP pathway active. (a) wildtype; (b) malic mutant; (c) malic OAD mutant; (d) malic PEPC mutant. The variable (α) is the ectoine-content fraction. The boundaries display the extreme cases of no ectoine or only ectoine in biomass. The results with the ED pathway are quite similar.

as the ectoine production (yellow line). The results for the EMP pathway are displayed in Figure 37. If compared with the ED pathway, one would see that the energy investment of the ED pathway is always higher, than for the EMP pathway, while the ectoine-yield stays lower. Moreover, as could already be seen in Figure 32, OAD is even more important in the case of the ED pathway, than in case of the EMP. All the flux of the anaplerotic reactions, which is above the yellow line, is the one, which flows into biomass, everything below flows into ectoine.

Another direct comparison of the EMP and ED pathway can be found in Figure 38. Again, α is the variable, which means the left boundary in (a) represents the optimal biomass production without the inclusion of ectoine, while at the right boundary the optimal value for ectoine production (in g Dw/g Glc) is reached. The overall yield of ectoine is 0.78 g ectoine/g Glc and the overall yield of biomass without ectoine content has been 0.67 g Dw/g Glc. (b) displays the ratio between the optimal values with the ED and EMP pathway and one can see that this ratio stays below 1 but is not significantly affected by the change in ectoine content.



(a) (b)

Figure 38: (a) displays the production of biomass with ectoine content in g Dw/g Glc, if the EMP pathway is active, while (b) illustrates the ratio between the production with the ED and EMP pathway. As one can see in (a), the higher the ectoine content, the more product in g Dw/g Glc. The fraction of ectoine does not have a great effect on the difference between EMP and ED pathway. The EMP pathway still ensures more product but this does not change depending on ectoine.

So far, the resulting yields have been quite high, since they are the maximal theoretical yields. The following section incorporates the regulation via the proton leak to link the theoretical yields with experimental results of fermentation experiments.

7.4 PROTON LEAK AND COMPARISON WITH EXPERIMENTAL DATA

The P:O ratio of the model can be regulated via the proton leak (in mmol H^+/100 mmol Glc), allowing the adjustment of the FBA by lowering the optimal output to gain results, which are closer to experimental results. Recent experiments (not yet published) have shown that the yield of biomass including ectoine is 0.5 g Dw/g Glc at a salt concentration of 1 mol/l.

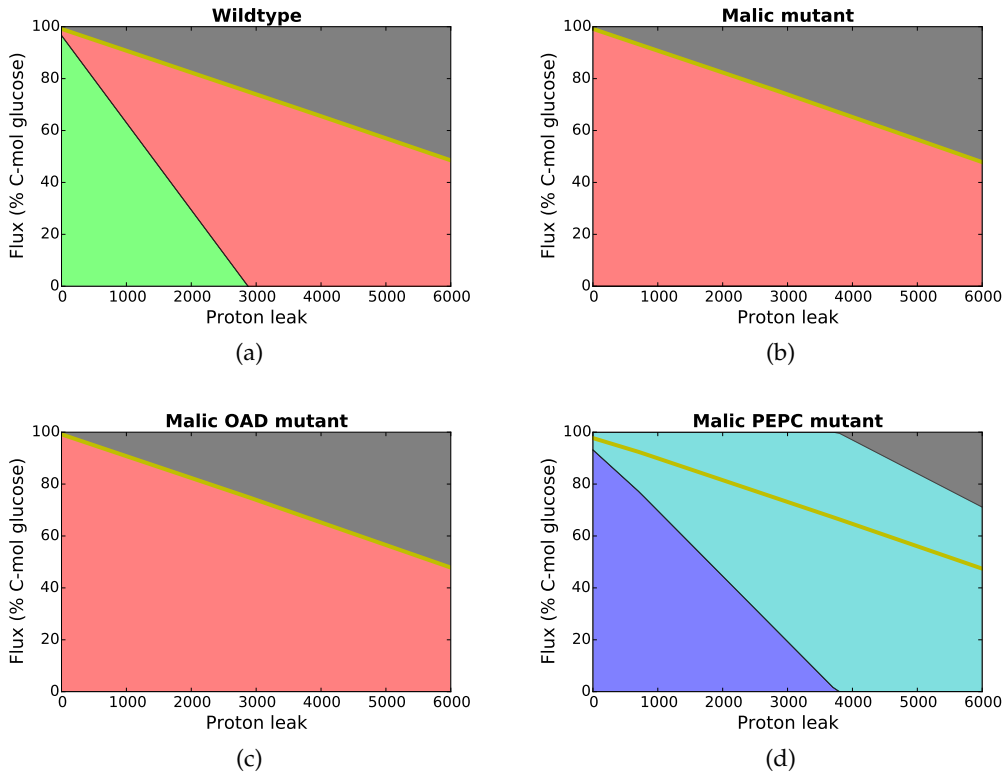


Figure 39: Optimization of ectoine production in dependence on a proton leak (in mmol H^+/100 mmol Glc) into the cell (EMP pathway). As depicted, the proton leak has a significant effect on the ectoine optimization and the use of the anaplerotic reactions.

Figure 39 shows the effect of the proton leak on ectoine optimization with an active EMP pathway. The results for the ED pathway are similar and not displayed. The malic enzyme looses its superior position with increasing proton leak, when optimizing for ectoine. It actually changes direction from malate to pyruvate, possibly to help the transhydrogenase exporting protons with the supply of NADPH. PEPC becomes more important, than before and is also superior with respect to OAD, because OAD lowers the export of protons with the use of the sodium-translocation C1. OAD is only considered as an option, if malic enzyme and PEPC are not available and even then, the glyoxylate shunt seems to be the superior choice.

What about biomass including ectoine? The paper from Doetsch et al. includes a figure, which displays the ectoine content in biomass for different salt concentrations. At a salt concentration of 1 mol/l, the ectoine content is about

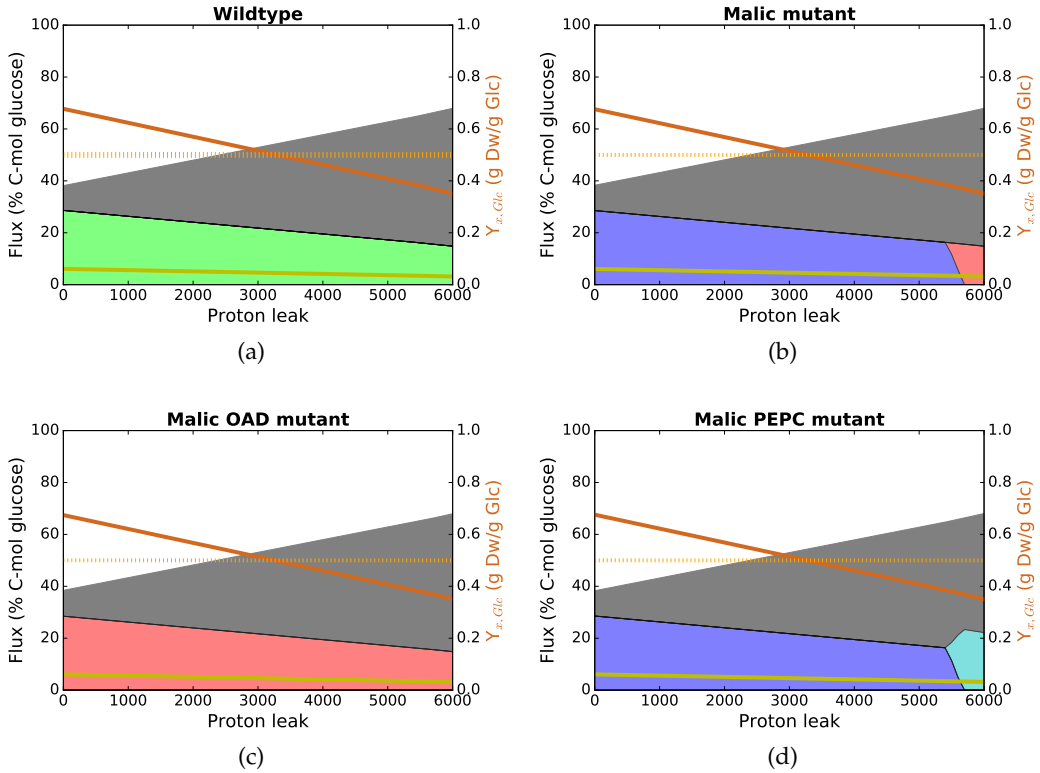
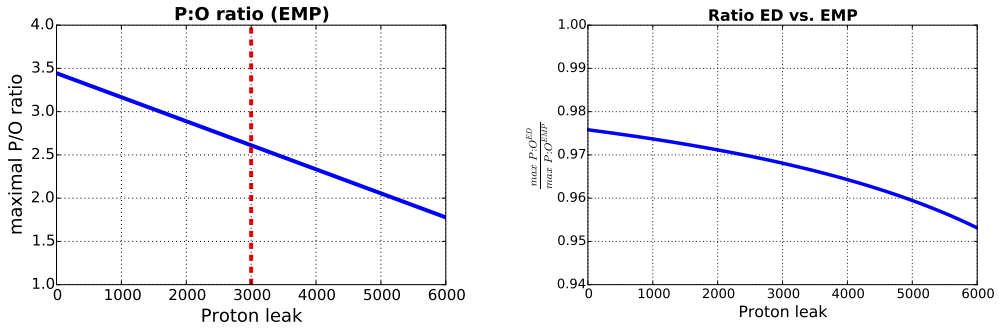


Figure 40: Optimization of biomass containing 7.1% ectoine in dependence on the proton leak (EMP pathwy).

7.1%. Therefore, the α from before is set to 0.071 and biomass, including ectoine, is optimized again. This time the proton leak is used as a variable to see, if it is possible to lower the theoretical biomass yield from 0.67 g Dw/g Glc to 0.5 g Dw/g Glc (marked by an orange dotted line). Figure 40 shows that a proton leak of about 3000 mmol H^+/100 mmol Glc seems to be sufficient. The preferences with respect to the anaplerotic reactions stay the same as in the other sections and the production of CO_2 is much higher. The similar results for the ED pathway are not displayed.

For a thorough examination, the maximal P:O ratio is calculated, dependent on the proton leak. Therefore, the ATP load reaction, which represents the need for ATP of the cell, is maximized for every proton leak and its value is set in proportion to the usage of the oxygen, represented by the flux through C3-4. The resulting ratio is illustrated in Figure 41 (a) where the proton leak of 3000 mmol H^+/100 mmol Glc, which ensures a theoretical biomass-yield of 0.5 g Dw/g Glc, is marked by a red dotted line. The resulting maximal P:O ratio is 2.6. Plotting the results of the EMP and ED pathway against each other, results in (b), showing that the proton leak indeed has an effect on the ratio of EMP and ED pathway. It drifts further apart.



(a) (b)

Figure 41: P:O ratio. (a) displays the P:O ratio for different proton leaks, while (b) depicts the difference between the results with EMP and ED pathway.

As mentioned previously, the FBA serves as an easy start in the mathematical modeling process of the metabolism of *H. elongata*. It predicts that the EMP pathway is overall better, than the ED pathway, although the difference is minor. MAE, if processing from pyruvate to malate, is the superior option to produce ectoine, independent of the glycolytic pathway. The next best option is OAD, then comes PEPC. One can theoretically gain 0.78 g ectoine/g Glc or 0.67 g Dw/g Glc but it has to be kept in mind that these are very optimistic assumptions. Constraining the P:O ratio via the proton leak reaction, lowers the biomass yield to values, which are compareable with experimental data.

The question remains, why *H. elongata* would exhibit the PPi-PFK instead of a normal PFK and if malic enzyme is actually able to catalyze the reaction from pyruvate to malate. Moreover, the flux distributions in general should be checked for their thermodynamic practicability. It has been explained in the previous section that ubiquinone and menaquinone are both included in the model, because it was not clear which one might be present or if both are present, as it is the case in *E. coli*. This question is not answerable with FBA, because FBA does not differentiate between them.

These matters will be discussed in the following chapter. The thermodynamic analysis will not be able to consider the whole core-model with biomass included, because the formation energies for the macro molecules are not available so far. Therefore, mostly the basic model, without growth, will be considered in the following. However, the last section will deal with a growth related flux distribution, whose active fluxes, which have already been matched with formation energies, will be analyzed thermodynamically.

The sign vectors of the FBA have to be updated with respect to the transhydrogenase in order to be analyzed thermodynamically. This reaction should not be restricted in the thermodynamics, because it would not allow for certain ratios of $\frac{\text{NAD}}{\text{NADH}}$ or $\frac{\text{NADP}}{\text{NADPH}}$. Moreover, it sometimes causes unrealistic solutions in the FBA, which would invalidate the thermodynamics.

8

THERMODYNAMIC RESULTS

The metabolism of *H. elongata* is largely unknown, which provokes several questions concerning the functionality. These will be examined in greater detail from an energetic point of view in the course of this chapter. Some have already been looked into from a stoichiometric angle but not all. For instance, the differences between the goals of ectoine and energy production have not been considered yet.

The following analysis will initially concentrate on differences in certain modules of the network, like the advantages of a sodium-translocating complex I ($C1^{Na}$) in the respiratory chain. From there it will evolve to questions concerning the overall metabolism.

8.1 RESPIRATORY CHAIN: COMPLEX I AND QUINONES

The respiratory chain is a very important module in *H. elongata*, because it provides energy, which is needed for synthesis reactions, like the synthesis of proteins, etc.. Thus, gaining insight into the processes and composition of this module is fundamental for the understanding of the whole metabolism. The respiratory chain of *H. elongata* contains two versions of complex I: One, which is only proton-translocating (C1) and one, which translocates sodium as well ($C1^{Na}$). These two channels are compared with respect to their thermodynamic properties. The analysis is coupled with a comparison of the ubi- and menaquinone, since quinones play an important part as electron shuttles in the respiratory chain. At first, the two complexes are compared via FBA. This analysis is followed by a thermodynamic analysis, which uses the flux distributions as signvectors.

In order to illustrate the differences between the two complexes, the analysis focuses on a subnetwork, the respiratory chain, as displayed in Figure 26. This module includes the ATPase, the antiporters and the different complexes. In order to analyze the differences between the two complex I versions, the module is extended by two reactions: First, the ATP load reaction, which serves as a measure of ATP production and which has been explained before; second, a new reaction (Na^+-intake), similar to the proton leak reaction, which mirrors all processes, which cause a sodium-intake of the cell, e.g. through OAD (see Figure 33). This reaction will serve as a variable in FBA. Instead of glucose, the inputs to the subnetwork are 100 mmol/(g Dw h) $FADH_2$ and 100 mmol/(g Dw h) NADH, while FAD and NAD serve as outputs, compare Figure 26.

A first comparison of the two complexes is done via FBA. The objective used is the maximization of the ATP production, which is represented by the ATP load reaction, to gain insight into how much energy could be produced. As has

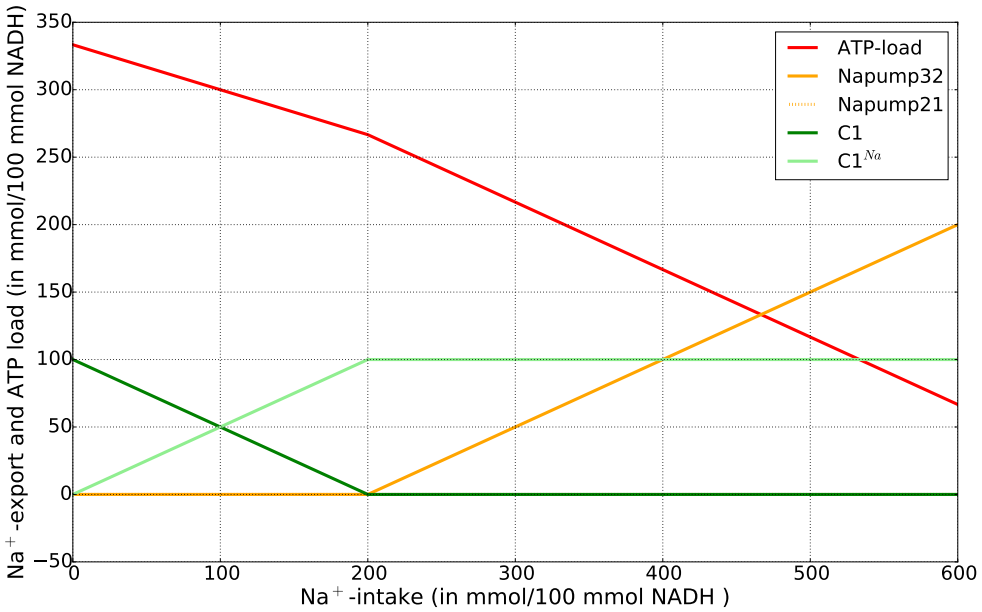


Figure 42: Flux distribution of the respiratory chain with varying sodium-intake. At low Na^+-intake, C1 is the superior channel, since it ensures a higher proton export for the ATP production. With increasing intake, $C1^{Na}$ gains importance, since the sodium needs to be exported again.

been done in the previous chapter, the objective function is optimized several times, in dependency on a given variable: The sodium-intake flux (in mmol Na^+/100 mmol NADH). The sodium-intake is chosen as a variable, since *H. elongata* thrives in high salt environments and certain substances, like compatible solutes, can only be imported, if coupled to sodium. Moreover, this accommodates processes, which import sodium into the cell, like the reaction catalyzed by OAD. Both complex I versions are active in the analysis. The fluxes in Figure 42 display the yields in dependence on 100 mmol NADH/(g Dw h). Figure 42 shows the development of the fluxes, as the proton-motive-force and the sodium-motive-force are built. As one can see, the classical version of complex I is superior, if sodium is not entering the cell, since it ensures a higher ATP production due to the greater proton export. As soon as the cell starts to take sodium in, the sodium-translocating complex I is used. Its flux increases with increasing intake till it replaces the classical complex I completely at an intake of 200 mmol sodium/100 mmol NADH. In general, the ATP production decreases with increasing sodium-intake, since sodium needs to be exported in addition to protons. The slope becomes even steeper as soon as the antiporters are needed as well, because the amount of electrons available to serve the respiratory chain does not stay above the electron-influx through the sodium-uptake anymore. Because the usage of 1 mmol NADH/(g Dw h) pushes two electrons through the membrane, this threshold lies at 200 mmol sodium/100 mmol NADH. The usage of the antiporters has higher opportunity costs, as has been explained in Figure 33 and therefore the decrease in ATP production is more significant. Since the Napump32 is still the superior antiporter, Napump21 stays inactive. The more the uptake increases, the more flux needs

to be channeled through the mentioned antiporter, since the respiratory chain would need more electrons to increase the flux through $C1^{Na}$.

An interesting fact, which is not displayed here, is that as soon as the sodium-intake is higher, than 120 mmol Na^+/100 mmol NADH, the ATP production of the respiratory chain with only $C1^{Na}$ would outperform the one with only C1. The reason is the ATP-balance: C1 might export two protons more, than $C1^{Na}$ in each turn, but since C1 would need the help of an antiporter to export any sodium and since the use of the antiporter causes an uptake of 3 protons, the balance is negative. The subnetwork with $C1^{Na}$ does not rely on the antiporter as long as enough electrons are available (see above). Therefore, the module with C1 is more productive at low sodium-intake, while the module with $C1^{Na}$ outperforms C1 at higher intake.

To further examine the differences between both channels, a thermodynamic analysis is executed. It is expected that $C1^{Na}$ exhibits other advantages besides being more effective at higher sodium intake. Moreover, the question arises, why *H. elongata* should have two antiporters, if one does not seem to be needed. This might have thermodynamic reasons as well. A possible explanation would be that the superior antiporter looses its ability to perform the exchange reaction in forward direction under certain conditions. The analysis of the two subnetwork-versions, one with only C1 and one with only $C1^{Na}$, is separately executed with ubiquinone and menaquinone, which might be equivalent concerning FBA but due to their different redox potentials are expected to exhibit differences with respect to thermodynamics. The module using ubiquinone will be examined first.

8.1.1 *Quinones*

The proton- and sodium-translocating modules are analyzed with respect to their feasibility under varying conditions. The conditions, which would be controllable by a experimentalist, are the pH outside of the cell, the salt concentration in the medium and the membrane potential, since it depends on the pH-difference of the cell (see Equation 14). The reaction directions used for this analysis follow from FBA. They are applied in the NET algorithm, which returns intervals for the Gibbs energies and metabolite concentrations. These intervals are of minor interest in the present analysis. It suffices to know which combination of the mentioned conditions result in feasible solutions. In order to examine the behavior under the predefined conditions, intervals are defined for all of them: The pH outside of the cell is varied between [5; 8], the salt concentration takes on values between [0.1; 2] mol/l and the membrane potential varies between [-200; -50] mV, dependent on the ΔpH. The internal pH is fixed at a value of 7.6, which is reasonable for *H. elongata*. The ionic strength changes with a varying NaCl. The internal ionic strength is considered to be constant 0.25 mol/l [174]. Each combination of the varying conditions is then tested with the NET algorithm and saved, if it turns out to be a feasible one.

The feasible combinations for a salt concentration of 0.1 mol/l or 2 mol/l, if only ubiquinone shuttles the electrons, are displayed in Figure 43.

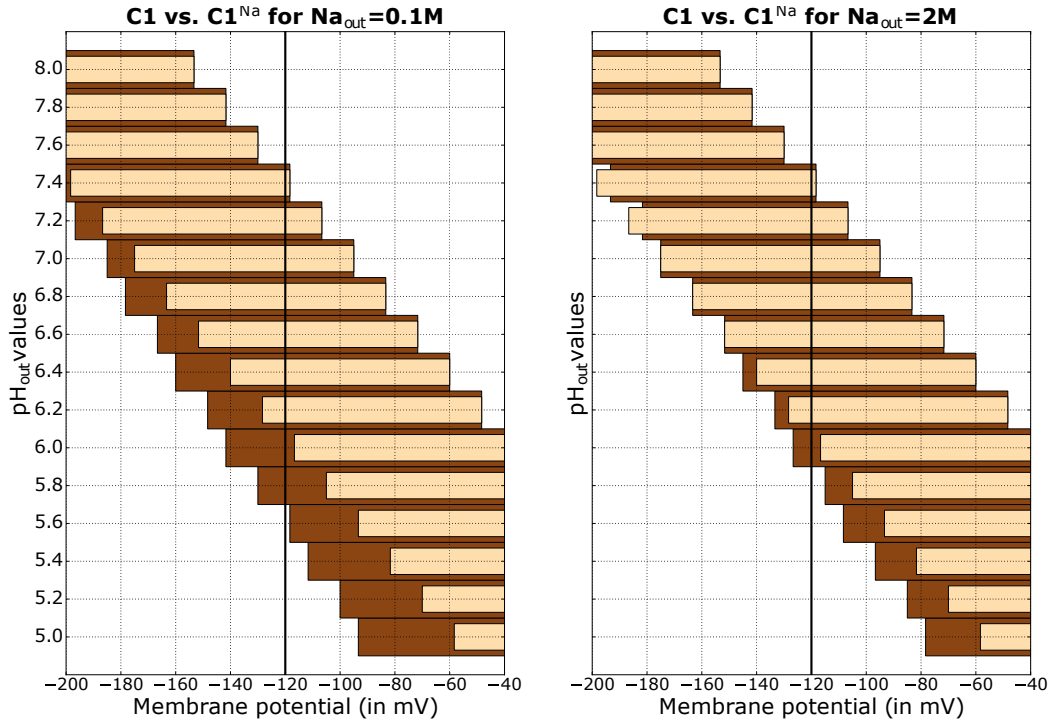


Figure 43: Comparison of the adaptability of the two enzymes (C1 and C1^{Na}) of the respiratory chain to different environmental conditions (salt concentration of 0.1 mol/l vs. 2 mol/l). The dark brown intervals belong to the module, which translocates protons, as well as sodium (C1^{Na}) and the lighter brown ones belong to the module, which only builds a proton-motive-force (C1).

The left panel shows the results for a salt content of 0.1 mol/l (= 0.1 M), while the right panel shows the results for 2 mol/l (= 2 M) salt. Both panels display the feasible combinations of pH_{out} and membrane potential for the two module-versions, the one, which uses C1 (light brown intervals) and the one, which uses C1^{Na} (dark brown intervals). Each external pH can be matched with an interval concerning the potential, since the term, which accounts for the influence of ions in Equation 14 has been varied between [35; 290] mV, to account for a membrane potential range of [-200; -50] mV in case of each pH. As one can see, the intervals with respect to the membrane potential are in general larger, if the sodium-dependent channel is present, at least for an external pH below 7. This can be explained with the proton-motive-force: The greater the pH difference, the lower the tendency for protons to leave the cell. This offers the sodium-dependent channel an advantage, since it exports fewer protons. It can therefore be concluded that this module is compatible with a stronger potential at lower pH_{out}. Comparing the two plots with each other indicates that a change in NaCl has a significantly greater effect on C1^{Na}, than C1, because the dark brown intervals become shorter with increasing salt (compare the right panel). One might expect the subnetwork with the sodium-translocating system to have an advantage at high external pH, since alkaliphiles are supposed

to rely on the sodium-motive-force [108]. In order to be true for this system, a sodium-dependent ATPase would have to be added. So far its existence in *H. elongata* has not been confirmed. Considering the figure in general, it appears that a higher pH outside of the cell is accompanied by a more negative membrane potential, which is not surprising, since this has already been established in *E. coli* [63, 186]. The vertical line at -120 mV indicates the typical membrane potential of *E. coli*, which serves as a reference value [106].

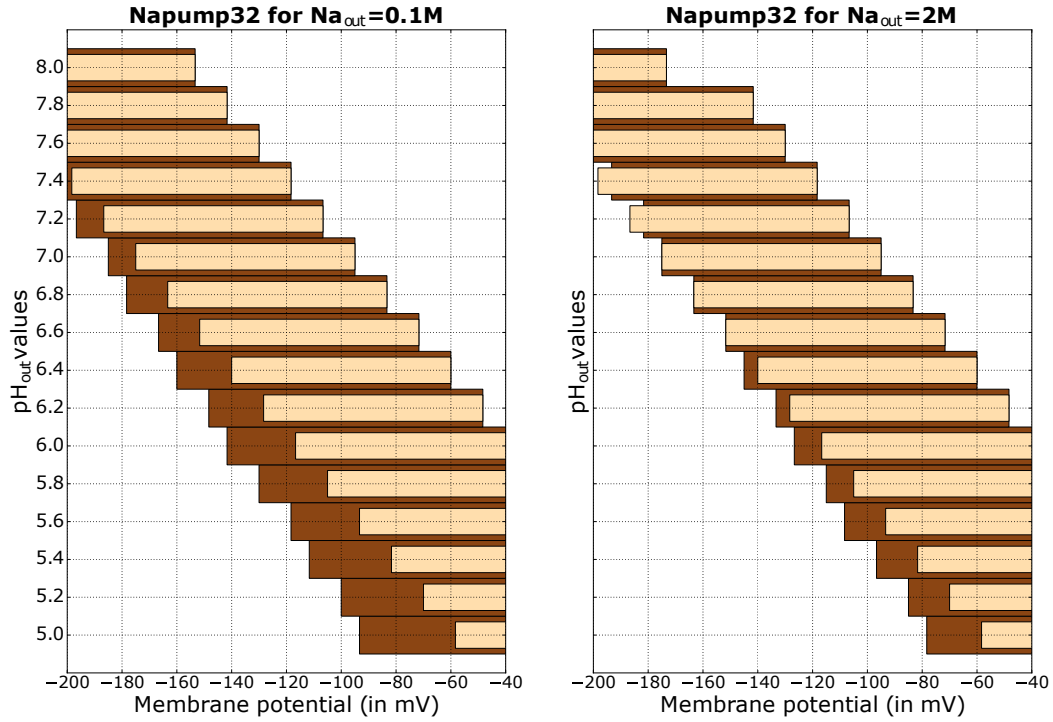


Figure 44: The adaptability of the antiporter Napump32 for ubiquinone. The intervals are almost exactly the same as before.

As mentioned, the role of the 1:2 Na^+/H^+ antiporter is unclear so far. Thus another thermodynamic analysis is executed, using the same varying conditions as before. This time signvectors are used, which exhibit active antiporters. The principle is the same as before, the different conditions are varied and the feasible combinations of pH and membrane potential displayed. Figure 44 shows that the feasible combinations are almost exactly the same as before. The only difference can be found at a pH of 8. The feasible interval turned out to be smaller. It can be concluded that the 2:3 antiporter does not loose its ability to proceed in a forward direction. The same applies for the 1:2 antiporter (result not displayed). Thus it is still unclear what role the inferior antiporter has. Considering the Gibbs intervals has at least shown that the 1:2 antiporter has a lower tendency to change direction, than the 2:3 antiporter.

The analysis is repeated. This time with mena- instead of ubiquinone. Comparing Figure 45 with Figure 43 shows that the respiratory chain with ubiquinone can operate at much wider potential ranges with respect to both versions of complex I. Especially at high salt, which is rather compatible with biological observations. Therefore, it seems reasonable to assume that it is more likely for

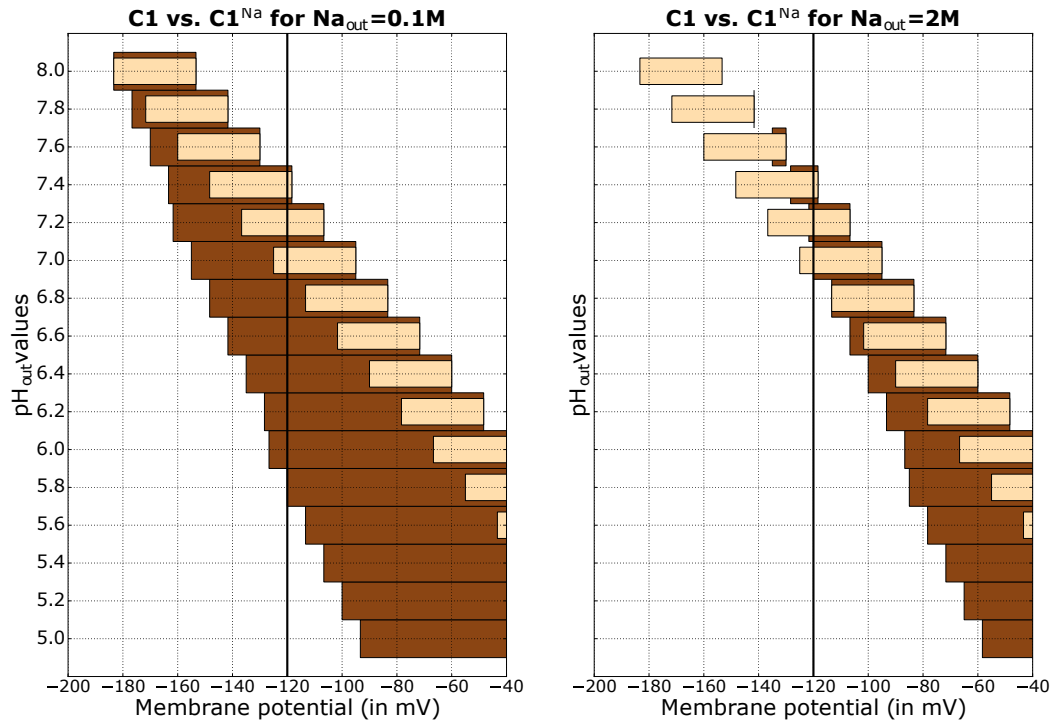


Figure 45: Comparison of the adaptability of the two enzymes to different environment conditions in case that menaquinone is used. The dark brown intervals belong to $C1^{Na}$ and the lighter brown ones to C1

ubiquinone to shuttle the electrons in the respiratory chain of *H. elongata*, than menaquinone. It is conceivable that both are present but menaquinone will be excluded in the remaining part of the analysis, since it might cause infeasibilities.

8.1.2 *Respiratory chain with MDF*

Given the importance of the respiratory chain, a further analysis from another angle is executed, the enzymatic costs of the pathway, determined by the Max-min Driving Force algorithm. This might offer new insights into the comparison of the effects of the proton-motive-force and the sodium-motive-force. This time the ratios between the major cofactors, $\frac{\text{ATP}}{\text{ADP}}$ and $\frac{\text{NAD}}{\text{NADH}}$, are also included in the analysis, which constrains the thermodynamic analysis further. These ratios are very important, since they ensure the upkeep of certain gradients. Moreover, constraining the ratios, rather than the absolute co-factor concentrations, is better, since they are more conserved in many cases [111]. In general, it seems that $\frac{\text{ATP}}{\text{ADP}}$ appears to lie in an interval of [3; 10], while $\frac{\text{NAD}}{\text{NADH}}$ should lie in an interval of [10; 20] [20, 68, 100, 111]. Since one has to choose certain values for both ratios, the values proposed by Henry et al. are assumed as fixed constrains. This might not be applicable in case of the whole core-model, but a small module, like the respiratory chain, is easier to constrain. The constraints are: $\frac{\text{ATP}}{\text{ADP}} = 3.2$ and $\frac{\text{NAD}}{\text{NADH}} = 19$.

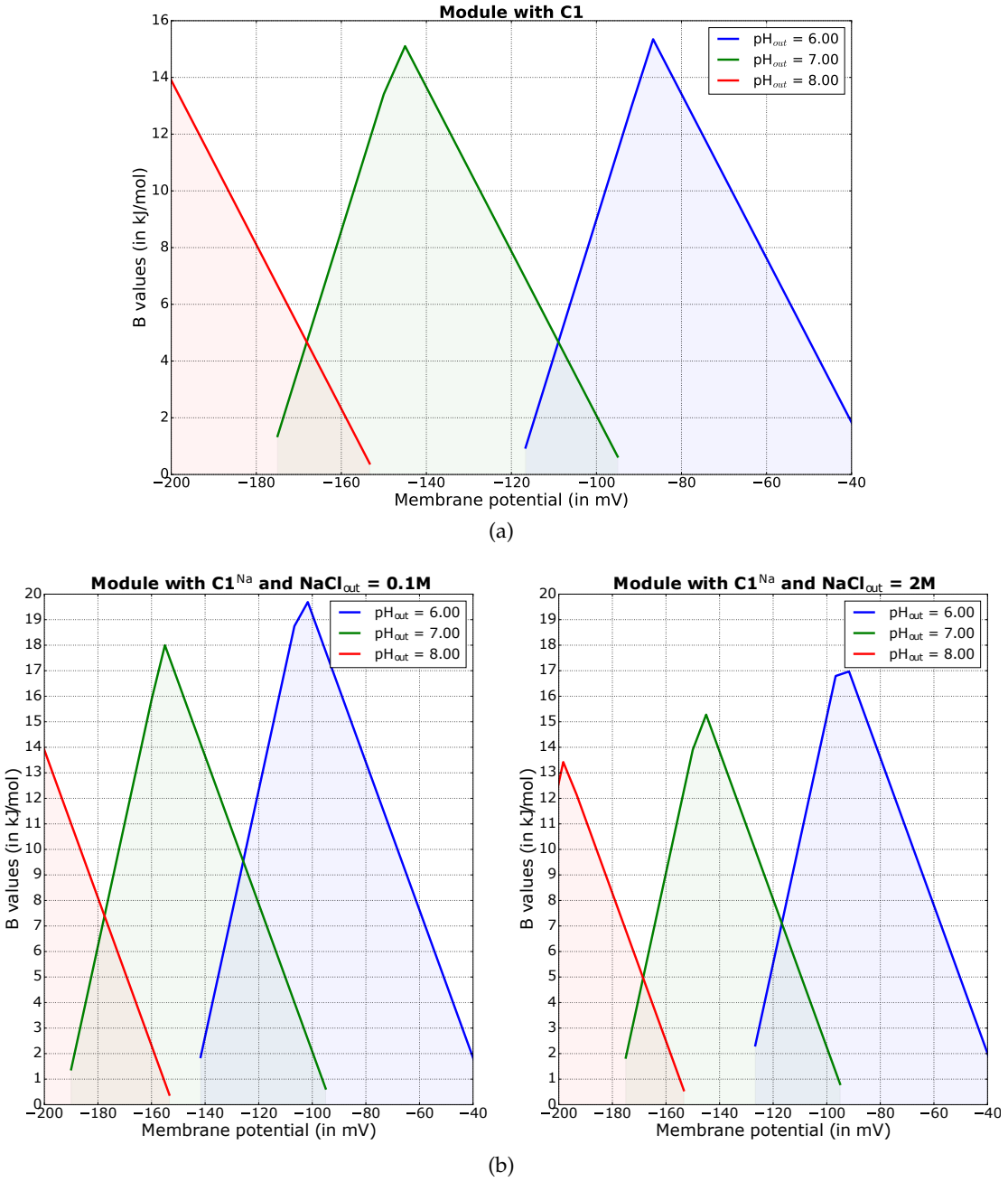


Figure 46: Comparing the respiratory chain with C1 (a) and C1^{Na} (b). Plotted are the minimal driving force solutions for varying membrane potential and pH_{out}.

As before, the different external parameters are varied but this time the minimal driving force values (B) are collected and plotted in Figure 46. (a) belongs to the module, which builds the proton-motive-force, (b) represents the subnetwork, which builds the sodium-motive-force as well. The x-Axis is the membrane potential (in mV), while the y-Axis displays the minimal driving force values for each external pH (indicated through the colors red, green and blue). The effect of the salt concentrations can be seen by comparison of the plots within the same row. Obviously, the complex I, which only translocates protons, is not affected by sodium and therefore is not affected by an increase in

salt concentration. Thus, only one figure is displayed in (a). This does not apply for the system with $C1^{Na}$ (Figure 46 (b)). In general it seems that the higher the external salt concentration, the lower the minimal driving force, since the maximal B values become smaller and therefore closer to equilibrium. Only at a $pH_{out} = 8$ does B stay the same, more or less independent of the salt concentration. Moreover, a lower external pH allows for higher minimal driving force values, which makes the system pathway cheaper. In general, the minimal driving force values seem to be higher in case of the respiratory chain, which builds the sodium-motive-force. Therefore, it turns out to be the cheaper pathway variant. Interestingly, one can once more see the strong correlation between membrane potential and external pH.

Comparing the plots concerning the system with C1 with respect to the different pH values, is seems that a pH_{out} of 7 is the best fit, since it allows for a wide interval concerning the membrane potential and includes the reference potential -120 mV. An optimal value for optimal operation seems to be the combination of $pH_{out} = 7$ and $\Delta V = -155$ mV. This potential is higher in magnitude, than the reference value but as has been stated before, the membrane of *H. elongata* is different to the one of *E. coli* with respect to the lipid composition and the peptidoglycan. The ions cannot go through as easily, which might allow for a potential of -155 mV.

In case of the respiratory chain with $C1^{Na}$, the conditions are different: a pH of 6 ensures a higher minimal driving force, than a pH of 7, independent of the sodium concentration. This also means a lower magnitude with respect to the membrane potential, which might not be reasonable in nature, due to the fact stated above. If one chooses a pH_{out} of 7 instead, the minimal driving force still exhibits higher or comparable values to the one in (a). In general, the possible bottleneck values are quite high for both systems, since the subnetwork is small. Figure 13 displayed the dependence of the productivity of an enzyme on the driving force of the reaction it catalyzes. It has been illustrated that a driving force of more or less 8 kJ/mol is already enough for an enzyme to operate close to 100% of its V_{max}. Therefore, it is not necessary to limit the optimal feasible ranges of the membrane potential to just the peaks of each curve. A wider interval can be allowed.

Concluding, one can say that $C1^{Na}$ in general offers higher magnitudes with respect to Gibbs energies but is also counteracted by higher external sodium concentrations.

8.2 PFK VS. PPI-PFK

As has been explained before, *H. elongata* exhibits the PPi-PFK (rn:R00764) instead of the ATP-dependent PFK (rn:R04779), which has a major consequence concerning the kind of glycolysis used. In this section the differences between both PFKs with respect to reversibility are examined via consideration of another subnetwork: The glycolysis from α-D-glucose-6-phosphate to lactate

The analysis is initiated by a comparison via FBA. As in Chapter 7, the ATP load is used as a variable, ectoine is maximized and the optimized values are collected for both pathway versions and then displayed in Figure 47. One version uses the PFK, the other the PPi-PFK, which is assumed to be reversible.

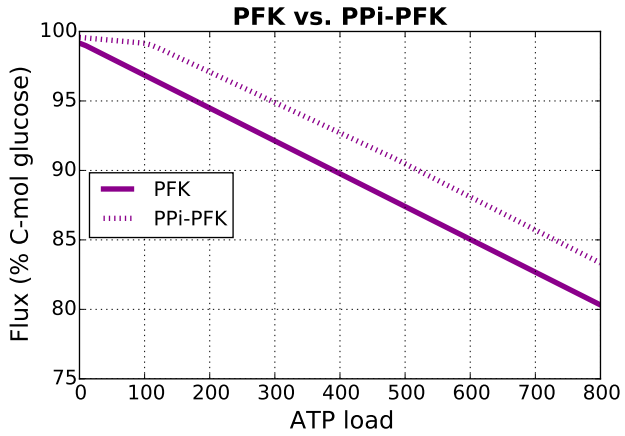
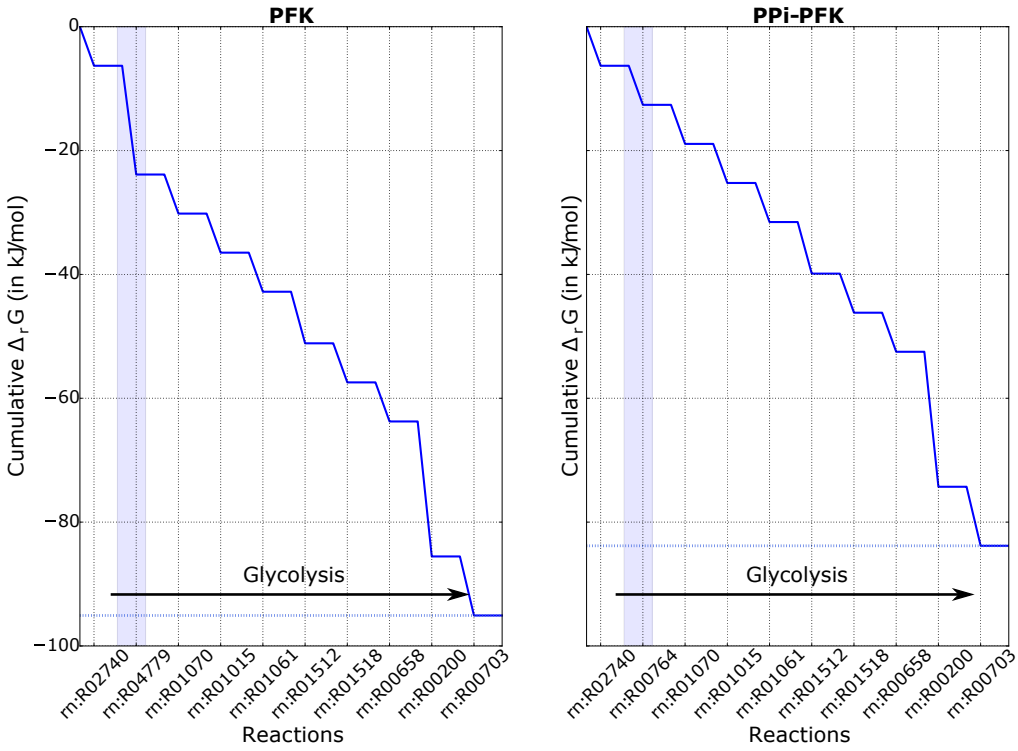


Figure 47: Ectoine production with PFK or PPi-PFK active for varying ATP load. As can be seen, the yield is higher in case of the PPi-PFK.

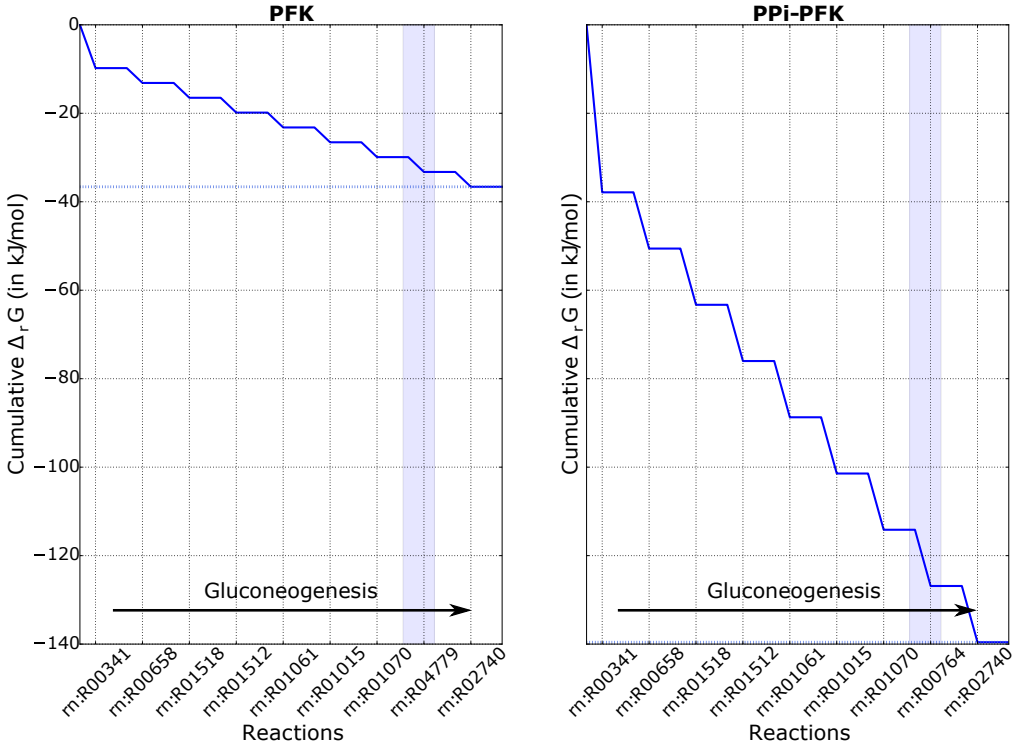
As it turns out, the pathway variant, which uses the PPi-PFK allows for higher product-yields, than the PFK. However, this only applies, if the enzyme actually is able to catalyze the forward direction, as the PFK does. This will be checked in the following.

The thermodynamic analysis is executed also using the MDF algorithm but without changing the external conditions of the cell: The membrane potential is fixed near the reference value of *E. coli* (-123 mV), since it is calculated via Equation 14 with an internal pH of 7.6 and an external pH of 7. The external salt concentration is set to 1 mol/l.

An initial analysis compares the thermodynamic profile of the pathway with either PFK, catalyzing the forward direction, see Figure 48 (a). The Gibbs energies of all reactions, as calculated by the MDF algorithm, are summed up and the change in driving force from one reaction to the next is displayed as steps. One can easily see that the PFK is energetically more favorable in the forward direction, than the PPi-PFK, since its cumulative driving force (~ 95 kJ/mol) is much higher. The step size of the PFK is higher as well, when comparing the step sizes of the reactions marked in blue. Moreover, it can be seen that the MDF algorithm provides a solution in which the step size from one reaction to the other is mostly conform. This generally applies for small modules, like the



(a)



(b)

Figure 48: Energetical profiles of the general glycolysis moving in forward (a) and backward direction (b) (gluconeogenesis), in order to compare which direction is more likely for which PFK. It can be seen that the PFK has a higher tendency to catalyze the forward and the PPi-PFK the backward direction. The reactions can be found in Figure 24, the formulas are displayed in Appendix B.

glycolysis module in which the boundaries of the concentrations do not differ too much from each other.

A second analysis is concerned with the backward catalysis. In this case the pathway starts with oxaloacetate instead of lactate, because it marks the start of gluconeogenesis. All direction signs from the previous analysis are switched. Figure 48 (b) shows that the pathway with the PFK has a much lower tendency to catalyze the backward direction, than the one with the PPi-PFK (compare the blue marked areas). The cumulative driving force of the PPi-PFK is much higher (~ 140 kJ/mol), which makes sense, since the PFK is in general assumed to be irreversible forward and is therefore attended by D-fructose-1,6-bisphosphate phosphatase (FBPP), which catalyzes the backward direction instead.

Is is concluded that glycolysis with PFK present has a higher tendency to catalyze the forward direction, while glycolysis with PPi-PFK has a higher tendency to execute the backward direction. Since the gene for the PPi-PFK has been found in the genome but the gene for the phosphatase has not, it seems even more likely that the PPi-PFK is a substitute and that *H. elongata* uses the cyclic Entner-Doudoroff pathway.

The comparison of the two glycolytic variants will take place in the following section. Furthermore, the model will be considered as a whole, instead of single sub-modules.

8.3 EMP VS. ED

The experimental and homologous results so far indicate that *H. elongata* makes use of the Entner-Doudoroff pathway in forward direction to catabolize sugar and that it uses the Embden-Meyerhof-Parnas pathway in backward direction for certain synthesis reactions, compare Figure 25. As has been seen in the chapter concerning FBA, the ED pathway does not produce as much ATP as the EMP and therefore produces less product. This raises the question why an organism would actually use the ED pathway. Flamholz et al. and Klingner et al. [52, 82] followed up on this issue. Flamholz et al. proposed that the use of the ED pathway might be beneficial from a thermodynamic point of view, while Klingner et al. discovered that strains of marine bacteria, which use the ED pathway, exhibit a more robust resistance against oxidative stress, which is typical for this environment. The current section executes a thermodynamic comparison of both pathway versions.

The analysis is similar to the one in the previous section. Again, a membrane potential of -123 mV and a pH_{out} of 7 is used. The external sodium concentration is varied between 0.1 mol/l, 1 mol/l and 2 mol/l, which results in three different curves in each of the following plots. The plots display the thermodynamic profile again. In the course of this analysis the whole core-model (without biomass) is examined via FBA and MDF algorithm but only the reactions of interest, meaning the ones from EMP and ED pathway are displayed, as

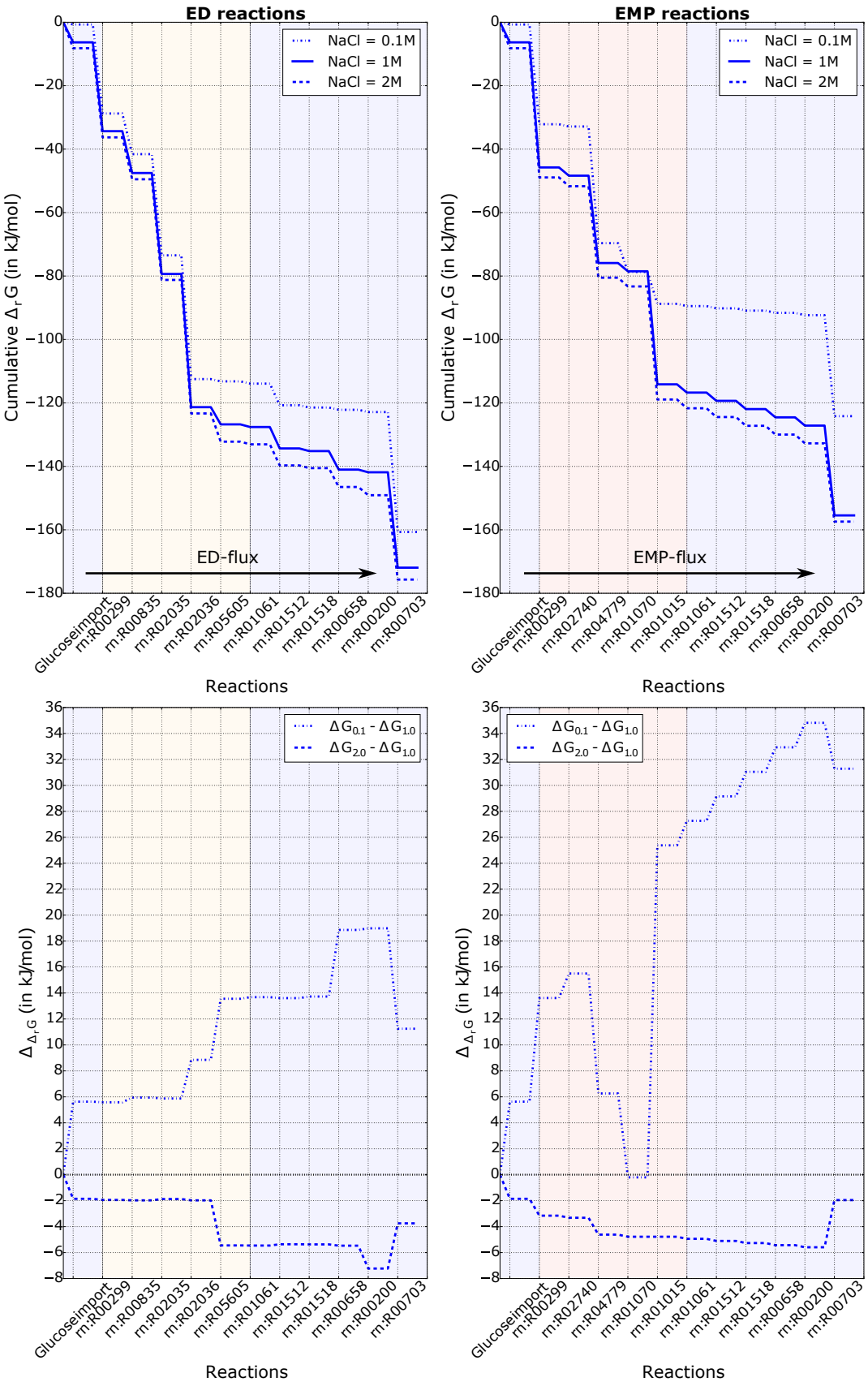
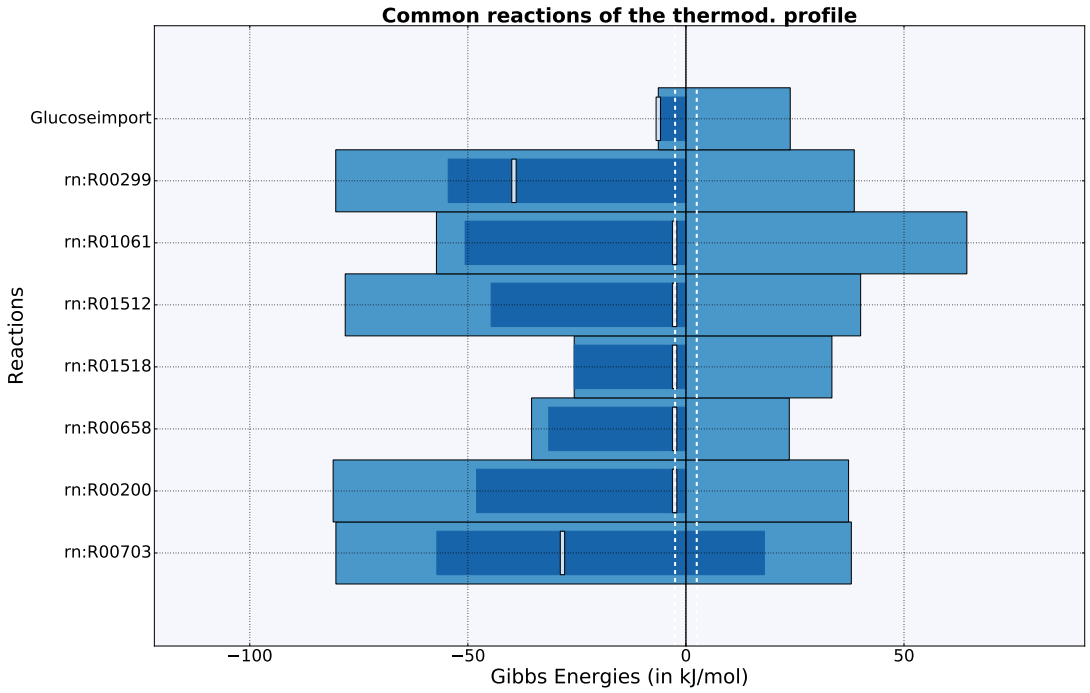


Figure 49: Displaying the comparison of EMP and ED pathway, influenced by the different external sodium concentrations. The second row displays the difference between the Gibbs energies of the reactions to identify those, which are mostly influenced and how by the change of the sodium concentration. The reactions can be found in Figure 24, the formulas are displayed in Appendix B.

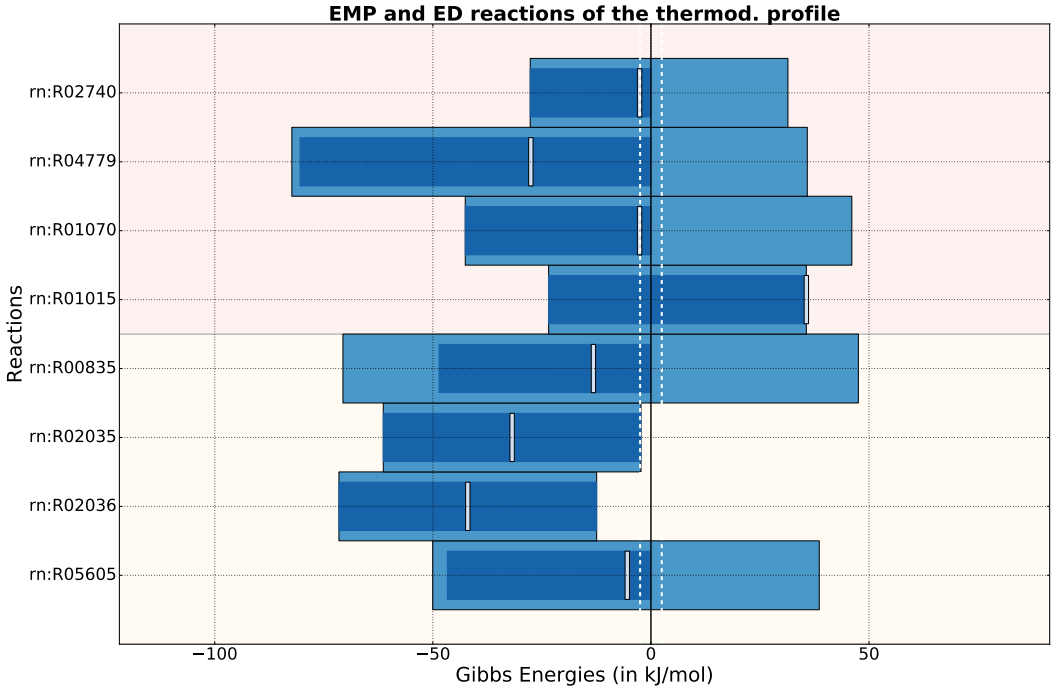
shown in Figure 49. This time the two versions of complex I are not examined, since the EMP pathway cannot be optimized without C1. The reactions, which overlap in both pathways are highlighted in blue, while the reactions, which belong either to the EMP or the ED pathway are highlighted in red or yellow. The results for the varying salt concentrations in the first row are indicated through different line styles, as can be seen in the legend. The first row of Figure 49 displays the thermodynamic profile, while the second row displays how the Gibbs energies of the reactions (represented by ΔG_x) are affected by the difference in salt concentrations (as indicated by the lowercase x). The dotted line, for instance, represents the difference in Gibbs energy at 0.1 mol/l and 1 mol/l salt, compare the legend.

As can be seen in the first row, the ED pathway is indeed thermodynamically more favorable, because the cumulative driving force is significantly higher (170 kJ/mol in case of 1 mol/l salt). It can therefore be assumed that ED is the cheaper pathway with respect to enzyme costs. Moreover, the increase in salt seems to have a positive effect on the thermodynamics in both versions, although the effects seems to be higher in case of the EMP pathway. This is caused by an improvement of the Gibbs energy of the *glucoseimport* reaction, as well as the reactions of the respiratory chain in case of the EMP pathway and the OAD, due to an increase of the ionic strength outside of the cell, which facilitates the import of sodium. Moreover, the difference between $\Delta G_{0.1}$ and $\Delta G_{1.0}$ is higher, than between $\Delta G_{1.0}$ and $\Delta G_{2.0}$, since the change in salt-content is also 10-fold, as can be seen in the plots in the second row: The driving force change in case of the dotted line is much greater, than the dashed one. The difference in-between the height of the steps of neighboring reactions indicates how much they are affected by the change in salt. With respect to the ED pathway, the *glucoseimport* reaction, as well as reactions rn:R05605 (2-dehydro-3-deoxy-phosphogluconate aldolase) and rn:R00658 (phosphopyruvate hydratase) are mostly affected. One can also see that only rn:R00703 (lactic acid dehydrogenase) is negatively influenced by the shift in concentration. For EMP, reaction rn:R01015 (triose-phosphate-isomerase) is affected significantly. Still, this result has to be considered carefully because the Max-min Driving Force exhibits degenerated solutions as well, since its based on a linear optimization, as FBA, and the bottleneck reactions can lie in some other part of the pathway.

In Figure 50 one can find the Gibbs intervals from the NET algorithm and the Max-min Driving Force solution from the thermodynamic profiles. The first plot contains the reactions, which are common in both plots and marked already as such in Figure 49 by the blue background. The second plot contains the remaining reactions of each pathway. The dashed white lines indicate the area, in which a reaction is assumed close to equilibrium ($|\Delta_r G| \leqslant 2.5$ kJ/mol). As one can see, the Max-min Driving Force solutions of the EMP reactions are clustering close to zero, which causes the huge difference between ED and EMP. The ED pathway has four big drops, the EMP only two.



(a)



(b)

Figure 50: Comparison of the Gibbs energy intervals of EMP and ED pathway. The upper plot contains all the common reactions of both pathways. The lower plot contains the reactions, which are specific for each pathway. The ones with the red background belong to the EMP pathway, while the ones with the yellow background belong to the ED pathway. The white lines mark the closeness to thermodynamic equilibrium, as has been explained before.

Going forward, all analysis attempts will use flux distributions, which use the cyclic ED instead of the EMP pathway.

The following section will deal with the question, if malic enzyme is actually able to catalyze the forward direction, as indicated by the FBA, or if this solution is not thermodynamically feasible after all.

8.4 MALIC ENZYME

As discussed in Chapter 7, if malic enzyme (MAE) actually belongs to the anaplerotic reactions and catalyzes the reaction from pyruvate to malate, it ensures the highest ectoine-yield. So far, malic enzyme is only known as catalyzing the backward direction. Therefore, the flux distribution gained by FBA will be checked for thermodynamic feasibility. Although the whole basic core-model (without biomass) will take part in the FBA and NET-analysis, only the reactions belonging to the anaplerotic reactions and the TCA-cycle will be displayed. The results can be found in Figure 51.

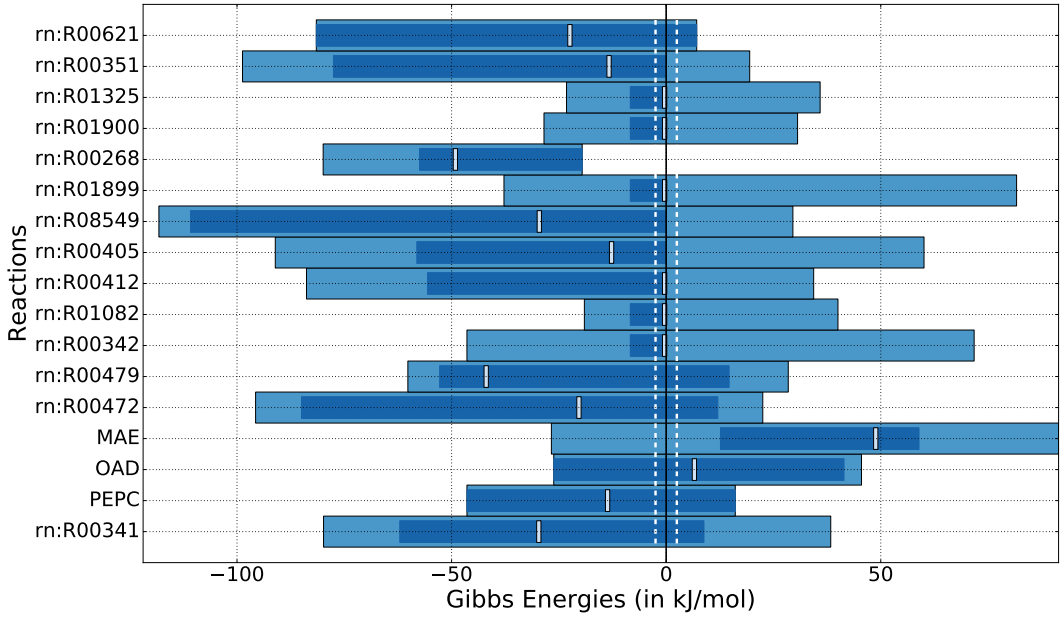


Figure 51: Solution, where the malate dehydrogenase (rn:R00342) is forced to go forward and MAE is left free. One can see that it is nevertheless only catalyzing the opposite direction. The reaction formulas are displayed in Appendix B.

The thermodynamic analysis exhibits feasibility problems with respect to the malate dehydrogenase (rn:R00342) and MAE. Both reactions are supposed to catalyze the forward direction, as determined via FBA. The malate dehydrogenase exhibits problems to move in forward direction as can be seen in Figure 51. The darker blue interval is very small and the MDF solution is in thermodynamic equilibrium. This reaction is a thermodynamic bottleneck in general [111]. It turns out that it is thermodynamically not possible to couple malic enzyme and malate dehydrogenase in forward direction. The NET algorithm only returns a feasible solution, if the reaction direction of malic enzyme is

not restricted. The result of this analysis is displayed in Figure 51 and one can see that malic enzyme has a clear tendency to go backwards. This is true for an external salt concentration of 1 mol/l and 2 mol/l. The external pH was chosen to be 7, which resulted in a membrane potential of -123 mV. For as yet unexplained reasons, the NET algorithm cannot find a feasible solution at an external salt concentration of 0.1 mol/l, at least not for the chosen pH_{out}. The analysis has been repeated for different combinations of pH_{out}, membrane potential and salt concentration again, but in all cases, the malic enzyme has to catalyze the backward direction. Furthermore, the activity of malic enzyme has been tested in the laboratory. The enzyme showed low activity in case of catalysis from pyruvate to malate, which serves as further evidence.

Therefore, the malic enzyme has the same purpose in *H. elongata*, as it does in other organisms: To be a workaround for the malate dehydrogenase, in case it proceeds in backward direction. Based on this insight, the following sections will not consider malic enzyme as an anaplerotic reaction anymore. The flux distributions of interest with respect to ectoine optimization are now the ones using OAD, PEPC or the glyoxylate shunt.

The subsequent chapter will deal with different goals of the cell and how they affect the thermodynamics.

8.5 ECTOINE- VS. ENERGYOPTIMIZATION (WITH AMMONIA ASSIMILATION)

Two different goals of the cell will be compared with respect to thermodynamics: The aim to produce as much energy as possible for all the different energy-dependent processes and the aim to produce ectoine in order to protect the cell from hyperosmotic shock. In a first step, two different flux distributions are obtained via FBA. The flux distribution, which serves ectoine maximization, uses the Entner-Doudoroff pathway to process glucose and the OAD-enzyme as an anaplerotic reaction, if available. The flux distribution, which serves energy maximization is obtained with the objective *max ATP load* and does not use the anaplerotic reactions at all, instead it takes advantage of the TCA cycle. Certain reactions of the model are not active in either one of the two flux distributions obtained. Therefore, a reduced version of Figure 24 has been created, as can be found in Figure 52. This figure will be the focus throughout this section.

Both flux distributions obtained are translated to signvectors and analyzed thermodynamically. Again, the co-factor ratios, as proposed by Henry et al., are used ($\frac{ATP}{ADP}$ = 3.2, $\frac{NAD}{NADH}$ = 19 and $\frac{NADP}{NADPH}$ = 1.2) and the external conditions varied, while using the MDF algorithm to find feasible solutions.

Figure 53, displays the flux distribution of ectoine (a) and energy (b) optimization for an ATP load of 10 mmol ATP/100 mmol Glc (a good estimate supported by the literature [118]). The differences between the proton translocating (C1) and the sodium-dependent complex I ($C1^{Na}$) are still of interest. Therefore, they are compared as well. The optimizations, which are displayed in Figure 53, e.g., exhibit an inactivated $C1^{Na}$ but active C1. An interesting flux

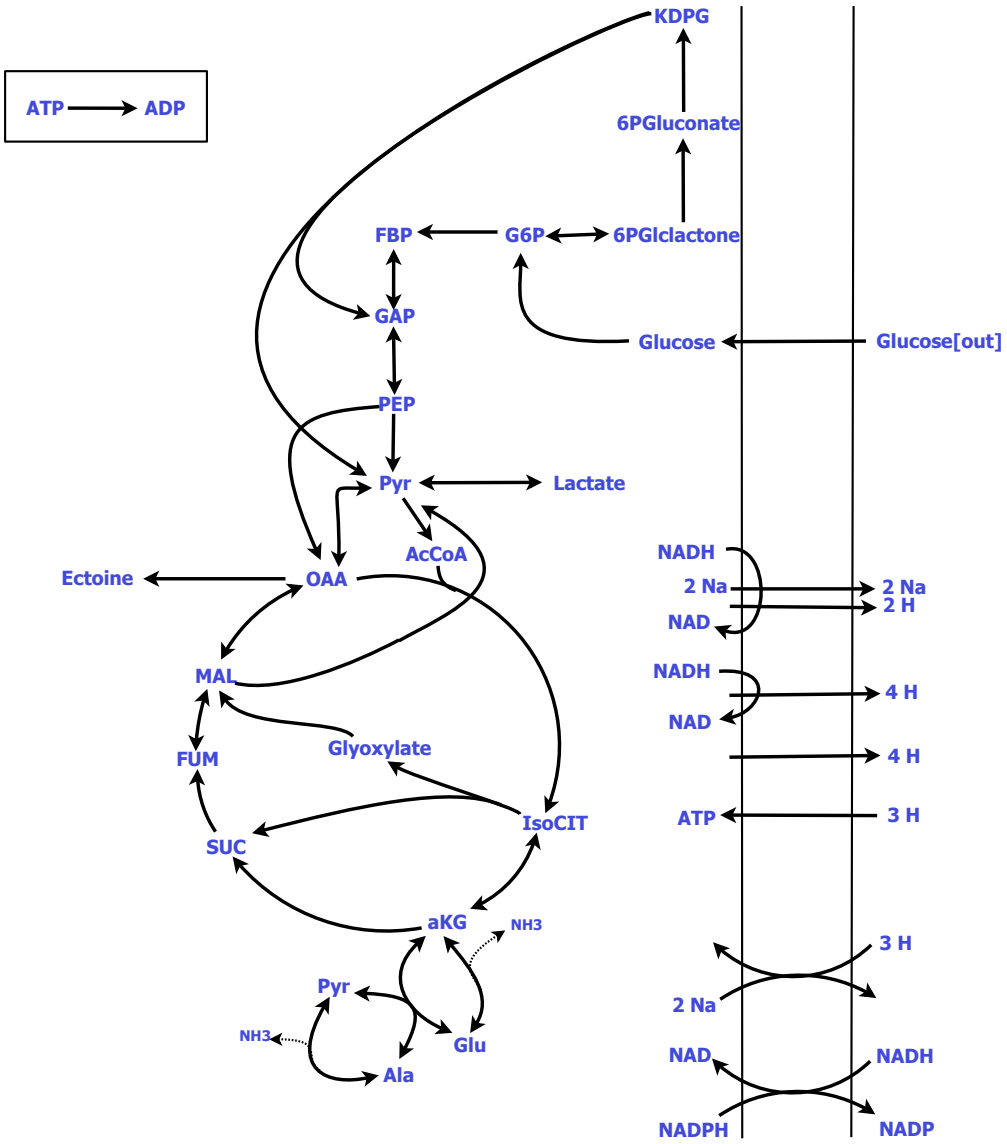
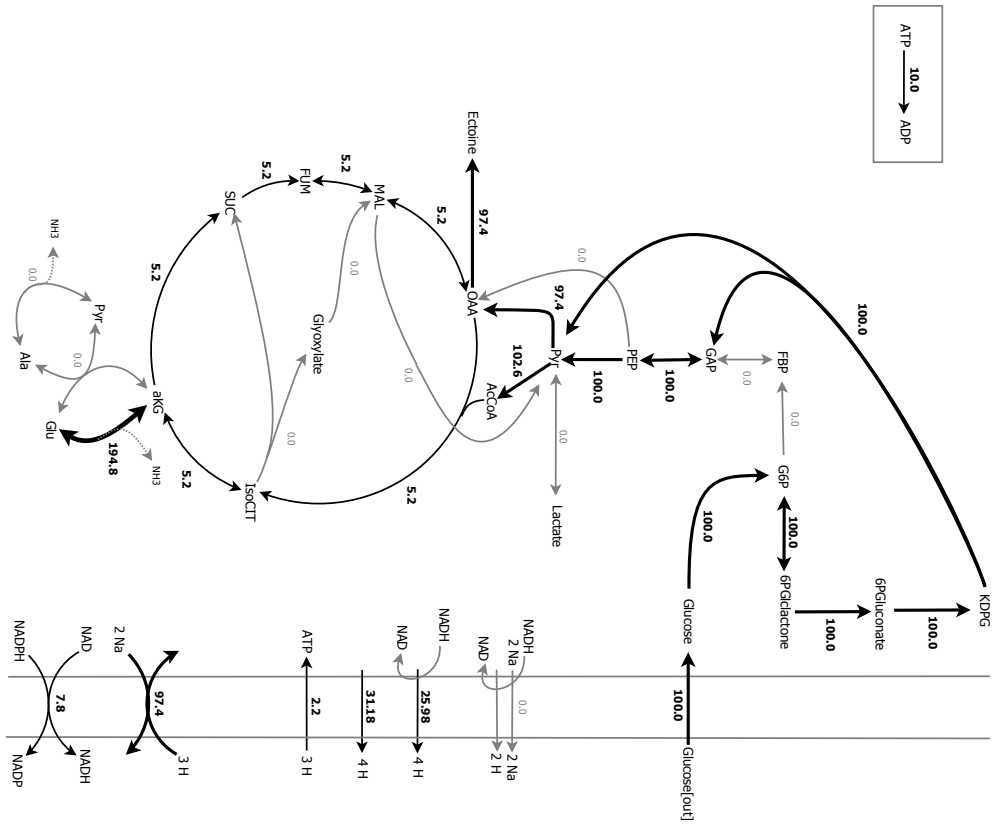


Figure 52: Reduced version of Figure 24. Only the reactions, which are relevant for ectoine or energy optimization are displayed.

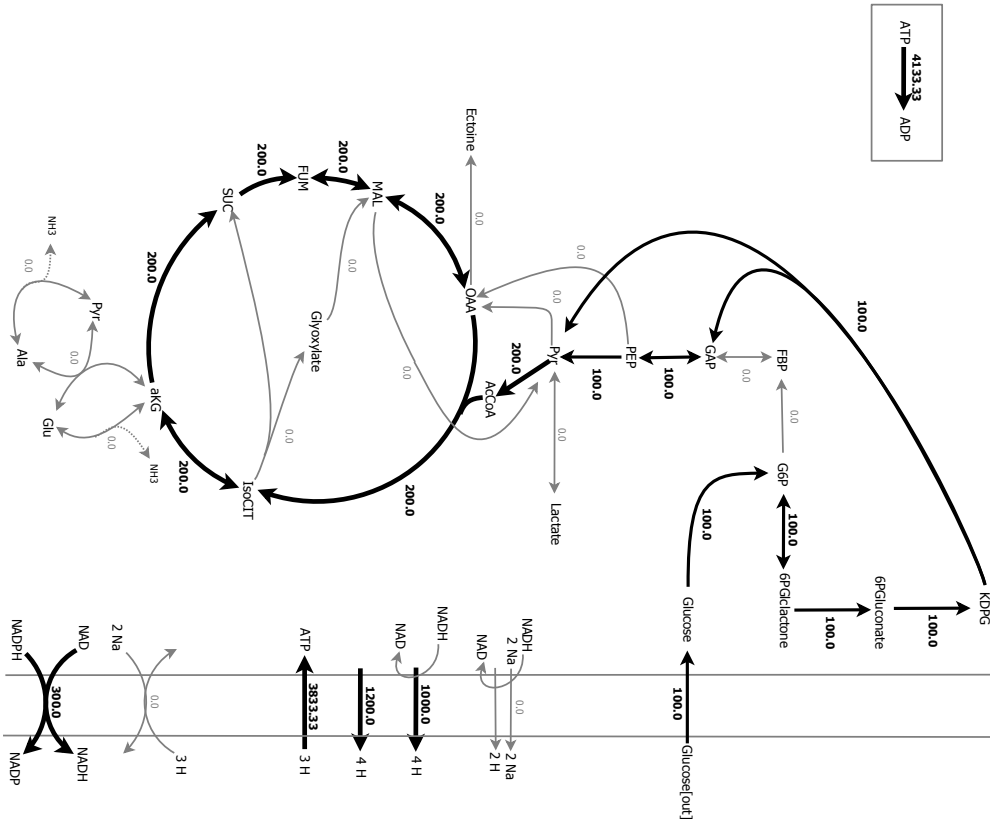
with respect to ectoine optimization is the flux, which allows the cell to import ammonia. The present system has two of those, the glutamate (GDH) and the alanine dehydrogenase (Ala-DH). There is no difference between those two concerning FBA, that is the resulting flux distribution always exhibits a flux through the same enzyme, the GDH. In order to compare the thermodynamic results with the ones obtained with the alanine dehydrogenase, the glutamate dehydrogenase will be inactivated in certain cases.

8.5.1 *Only proton-translocating channel active ($C1^{Na}$ inactive)*

As can be seen in Figure 53, all reactions, which exhibit a flux in the FBA are black and the stronger the flux, the thicker the arrow. It is obvious that those



(a)



(b)

Figure 53: Comparison of the ectoine- (a) and energy-fluxes (b) for an ATP load of 10 mmol ATP/100 mmol Glc and with C1 active in the FBA. The TCA cycle is obviously more important for energy production.

two cell goals result in quite different flux distributions. The ectoine opimization makes the least use of the TCA cycle possible, because its use would burn up carbon, which would in turn not be available for ectoine production. The energy optimization favors the burning of carbon, because it results in energy production. Moreover, the ammonia assimilation takes no part in this process. Now, what does the thermodynamic analysis reveal? Do the feasible conditions of both objectives overlap?

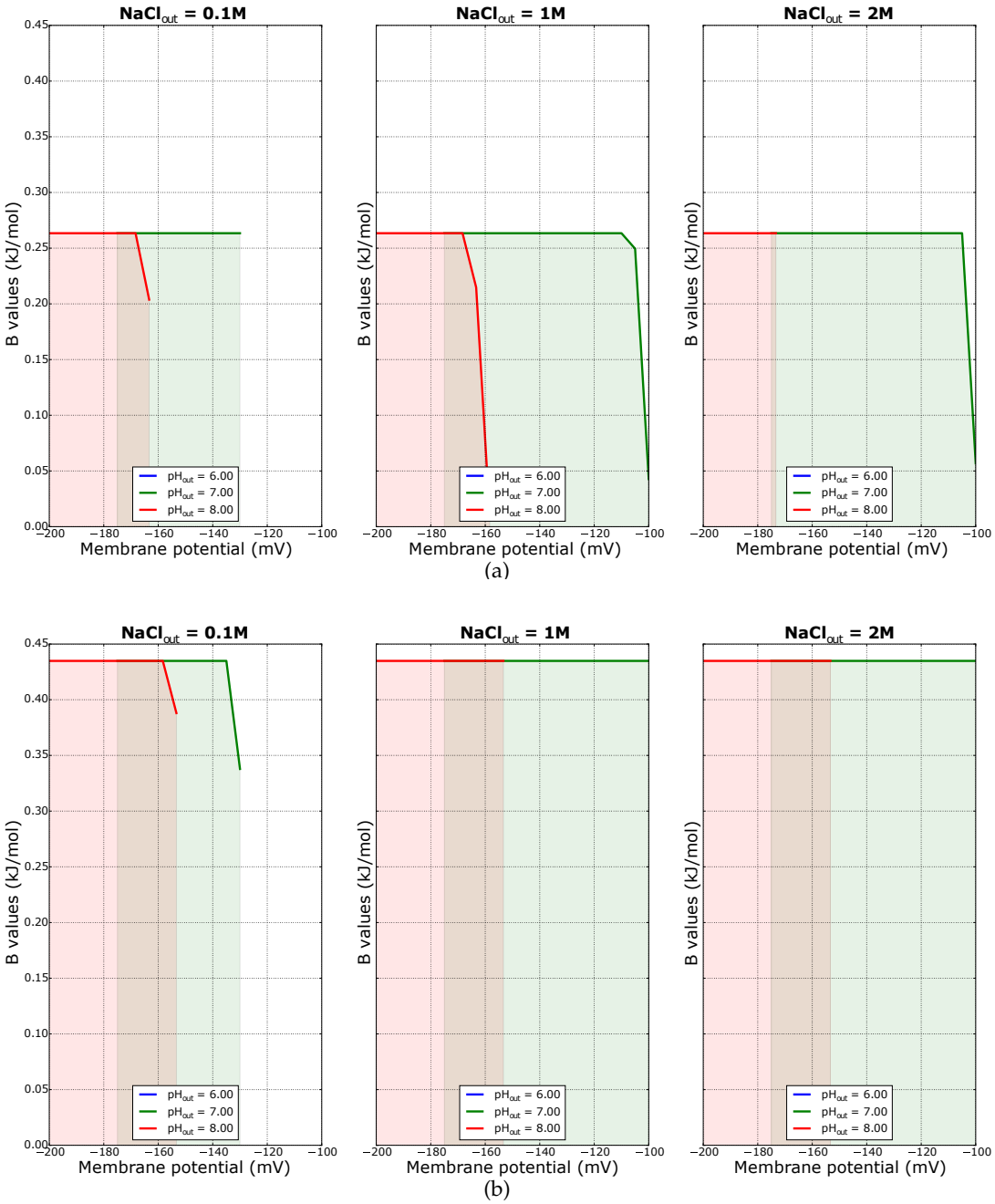


Figure 54: Comparison of the MDF solutions for ectoine (a) and energy (b) optimization with only C1 active in the flux distributions.

In Figure 54 one can find the feasible thermodynamic boundary conditions for both flux distributions. (a) displays the conditions for the ectoine optimization for different salt concentrations. (b) shows the same for the energy optimization. As has been done before for Figure 46, the different external conditions have been varied and if a feasible solution of the MDF algorithm existed, it has been saved and displayed. The external pH has been varied between 6, 7 and 8. As can be seen in the legend: A pH_{out} of 6 is not feasible in case of the whole network, at least not for a membrane potential between -100 mV and -200 mV, and therefore not displayed. The reason is the import of glucose. The formation energies for glucose in- and outside of the cell are corrected for the respective pH. This, in combination with the concentration intervals chosen (see Section B.4), results in a positive Gibbs energy and therefore in a backward catalysis. The problems with non-alkaline pH have been discovered before by Noor's group [111]. They found out that it also has a negative effect on the TCA-cycle.

As has been seen before, a higher pH_{out} of 8 goes along with higher magnitudes of membrane potential. Therefore, the feasible intervals of pH_{out} equal 7 or 8 are almost completely different. This was already obvious in Figure 46 and is confirmed here for the whole model and moreover with respect to two different cell goals. The first thing, which catches the eye is the maximal possible B-value. The energy optimization offers values, which are almost twice the size of the ones, offered by ectoine optimization. Moreover, the maximal value is not reached for only one membrane potential value but for a whole range, in contrast to the peaks in Figure 46. The values of B are also much smaller, than before. Independent of the cell goal, the bottlenecks of the pathway operate close to thermodynamic equilibrium. They are mainly residing in the TCA cycle and furthermore in its in- and effluxes in case of the ectoine optimization. A higher salt concentration seems to have a positive effect on both cell goals, if the external pH is 7. The feasible range concerning the membrane potential in combination with an external pH of 7 seems to be more or less the same for both goals. In case of an external pH of 8, they differ: The energy optimization seems to be unaffected by the different salt concentrations at a pH_{out} of 8, whereas the feasible membrane potential decreases with increasing salt concentration in case of the ectoine optimization and the same pH.

8.5.2 *Only the proton- and sodium-translocating channel active (C1 inactive)*

To be able to compare the gained results with the ones, which would come out of the system with $C1^{Na}$, the analysis is repeated. Most of the figures of the present and the following section are displayed in Appendix C but referred to in the main text. The resulting conclusions are still presented in detail. As can be seen in Figure 70, the ectoine production has actually a higher yield, if $C1^{Na}$ is active, instead of C1 but at the same time, the energy production displays a lower yield, than before. Therefore, in the FBA, $C1^{Na}$ is preferred for ectoine production, since OAD is active and imports sodium, which has to be exported again (the ATP load is still fixed at 10 mmol ATP/100 mmol Glc),

while the proton-translocating channel is preferred for energy production, since it ensures a higher export of protons.

Figure 71 shows the thermodynamic feasibility ranges. As before and in all the following figures, an external pH value of 6 is not feasible for a membrane potential in-between -100 mV and -200 mV. The maximal B-values stay the same as before, meaning that the bottlenecks of the system optimizing energy are still close to equilibrium, but not as close as the ones in case of the system with ectoine production. The intervals are quite similar to before but they are wider with respect to energy optimization at a pH_{out} of 7 at low salt concentrations. This is interesting, considering that in this case the FBA would result in a flux distribution with active C1 rather, than an active $C1^{Na}$. Considering the pathway serving ectoine optimization, the MDF algorithm indicates that $C1^{Na}$ has certain advantages at low salt concentrations with respect to an external pH of 7, which backs up the FBA. The results concerning an external pH of 8 stay exactly the same, independent of the optimization-goal of the cell.

8.5.3 *Both complex I versions active*

So far the two complex I versions have only been considered independent of each other. In a next step both enzymes are kept active during the FBA but as expected, the algorithm chooses the best of the two options: In case of ectoine optimization and an ATP load of 10 mmol ATP/100 mmol Glc, only $C1^{Na}$ is active, while the algorithm results in a non-zero C1-flux in case of energy optimization, because more protons are exported for energy production. Therefore, the results are the same as displayed in Figure 70 (a) and in Figure 53 (b), which leaves the thermodynamic results unchanged as well. Still, the question remains, if the FBA algorithm would result in a different flux distribution, if the ATP load would be set higher. This is the case with ectoine optimization, as can be seen in the stacked area plot 55, which is similar to the ones in Chapter 7.

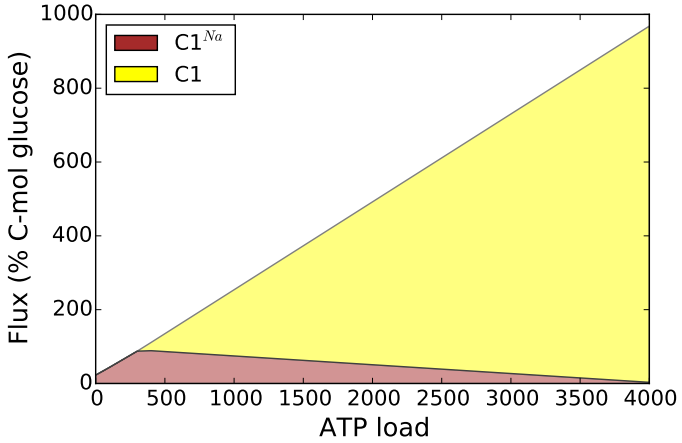


Figure 55: The fluxes of C1 and $C1^{Na}$ in case of ectoine optimization and under varying ATP load (in mmol ATP/100 mmol Glc). As can be seen, the importance of C1 (yellow area) increases with increasing ATP load.

With increasing ATP load, the cell cannot cover the ATP-shortage alone with $C1^{Na}$ but starts to use both enzymes simultaneously and increases the use of C1 with increasing ATP load, because it yields more ATP, as long as the net-influx of sodium is not constrained at the same time (compare Figure 42). Again, two flux distributions are compared, the one with an ATP load of 10 mmol ATP/100 mmol Glc, which has been analyzed already and one with an ATP load of 600 mmol ATP/100 mmol Glc, which uses C1 as well as $C1^{Na}$, see Figure 72 in the appendix for a comparison of the flux distributions.

The resulting signvectors are analyzed thermodynamically as well, which is displayed in Figure 73 in the appendix. The thermodynamic analysis shows that the addition of C1 has an effect at low and high salt: In low salt situations, the feasible range of the membrane potential at a pH_{out} of 7 decreases, while in case of high salt environments, the feasible range at a pH_{out} of 8 increases marginally.

8.5.4 *Ammonia assimilation*

The network exhibits two ways to import ammonia. Therefore, a comparison of both possibilities might offer insights into the preferences of the organism. It has already been shown that from a stoichiometric view they appear the same, which always results in flux distributions with active glutamate dehydrogenase (GDH) flux. Therefore, in order to examine the effect, the GDH is inactivated and the resulting flux distributions are analyzed with the MDF algorithm. This is only done for the objective ectoine optimization, because ammonia assimilation does not play a role in energy optimization. Again, both versions of complex I are compared. Figure 74 in the appendix displays both flux distributions, with GDH inactivated. The FBA is otherwise not altered (compare Figure 53 (a) and Figure 70 (a)).

Figure 56 shows that the inactivation of the glutamate dehydrogenase affects the thermodynamic feasibility significantly. In case of low salt, low pH_{out} values are not feasible anymore and the feasible interval at a pH_{out}=8 is much smaller, than before. With higher salt, the conditions get better but it is still worse, than with the GDH. First, the improvement is caused by the use of OAD, which imports sodium. The higher the sodium concentration outside of the cell, the easier the import of sodium into the cell, which has a positive effect on the overall driving force. Second, the improvement is caused by the change in Gibbs energy of *Glucoseimport*, which has a higher tendency to catalyze the forward direction. It could be assumed that glucose is not imported alone but might undertake a symport with sodium. This possibility has been tested but a symport with sodium does not improve the glucose import enough to ensure a forward catalysis at low salt and a pH_{out} of 7. The usage of ATP or the import of protons does not improve the Gibbs energy either.

One can still see that the variant with C1 is a little better at high salt concentrations, at least for a pH_{out} of 7. The feasible ranges at an external pH of 8, are not affected at all by the choice between the two complex I versions. Inter-

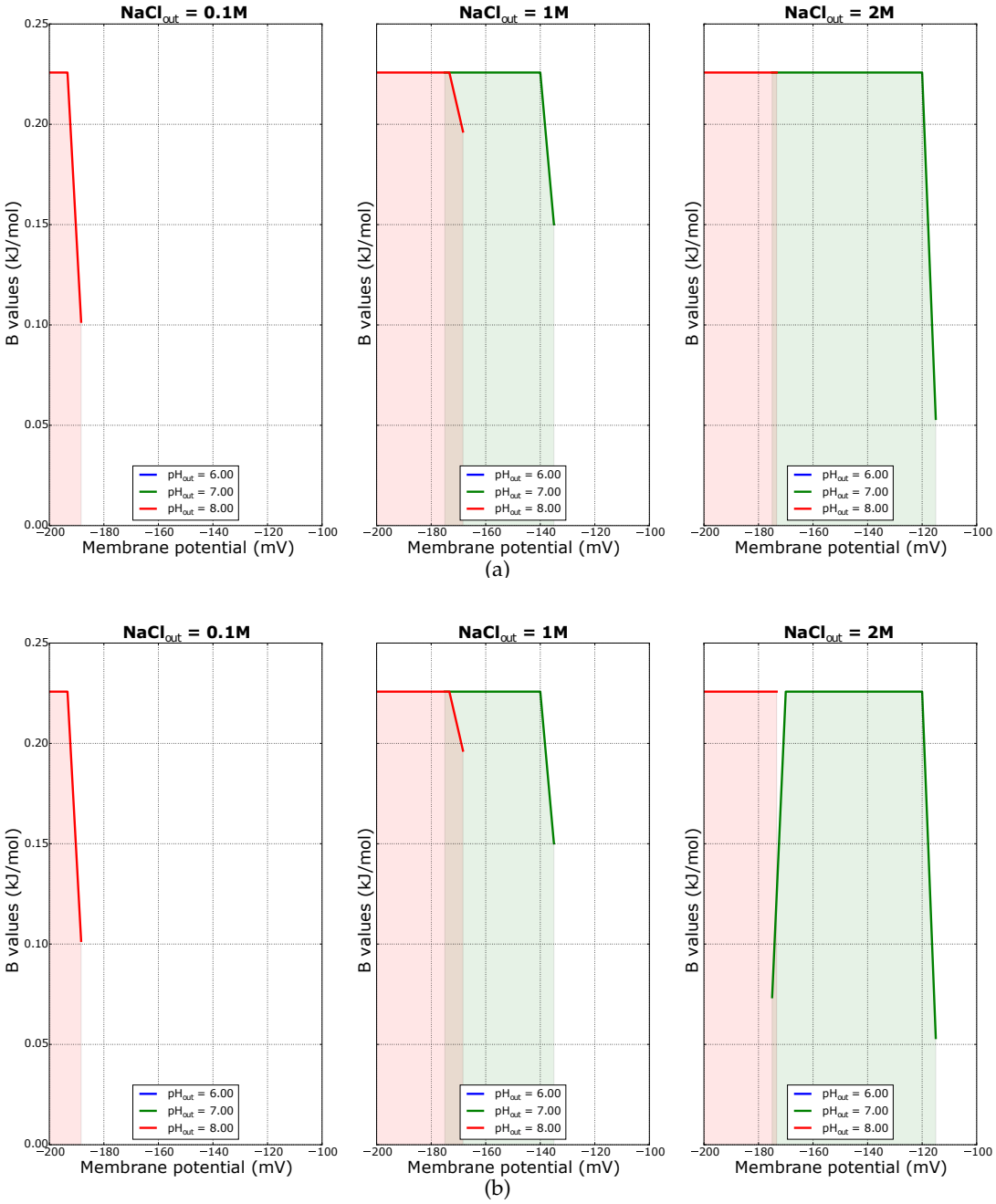


Figure 56: Comparison of the MDF solutions of the flux distributions with only the proton-translocating or sodium-translocating channel active and with the glutamate dehydrogenase inactive.

estingly, enzyme assays done in the lab showed that Ala-DH has a very high activity, which indicates its importance. Maybe it is an enzyme, which is used especially at high salt concentrations.

In a next step, the flux distributions at an ATP load of 10 mmol and 600 mmol ATP/100 mmol Glc are compared, to examine if the results differ, when Ala-DH is used. The flux distributions can be seen in Figure 75, which are quite similar to Figure 72. It follows the thermodynamic analysis. The first row

in Figure 76 corresponds to Figure 56 (b), because in both cases only $C1^{Na}$ is active. One can see that the addition of C1 has no effect on the feasibility range in case of an external pH of 7 but does have a positive effect for an external pH of 8 and at higher salt concentrations, which is the same effect as before with active glutamate dehydrogenase (compare Figure 73). Still, the feasibility ranges are much smaller, than in case of an active GDH.

The following can be concluded:

- A pH_{out}=6 is not feasible in case of the basic core-model for neither of the two optimization-goals of the cell, since it does not allow the forward catalysis of the *glucoseimport*.
- The minimial driving force is higher with respect to energy optimization but both flux distributions exhibit reactions close to thermodynamic equilibrium, due to bottlenecks in the TCA cycle.
  - Ectoine optimization:
    * Active GDH:
      · The FBA-algorithm results in a flux distribution, which exhibits an active flux of the sodium translocating channel, as long as the ATP load of the cell is low.
      · At low salt, the flux distribution, with active $C1^{Na}$ exhibits a wider membrane potential range at a pH_{out} of 7, than the one, which uses the proton-translocating channel (compare Figure 54 (a) with Figure 71 (a) for low salt).
      · Simultaneously, at high salt, the membrane potential ranges are quite the same for a pH_{out} of 7. Thus, the choice of complex I does not affect the thermodynamics.
      · With increasing salt, the feasible membrane potential range for pH_{out}=8 becomes smaller, independent of the choice of complex I version.
      · The simultaneous use of the proton-translocating and the proton- and sodium-translocating channel has a positive effect on the membrane potential at high salt but a negative one at low salt (compare Figure 73 (a) with (b)).
    * Active Ala-DH:
      · The thermodynamic feasibility is negatively affected, especially at low salt
      · The differences between C1 and $C1^{Na}$ vanish almost completely
  - Energy optimization:
    * At low salt, the flux distribution, with an active sodium-dependent channel exhibits a wider membrane potential range in combination with pH_{out}=7, than the one with C1 (compare Figure 54 (b) and 71 (b)).

* At high salt, the membrane potential ranges are exactly the same with either C1 or $C1^{Na}$.
* The thermodynamically feasible range of the membrane potential is not affected by the salt concentration in case of an external pH of 8.

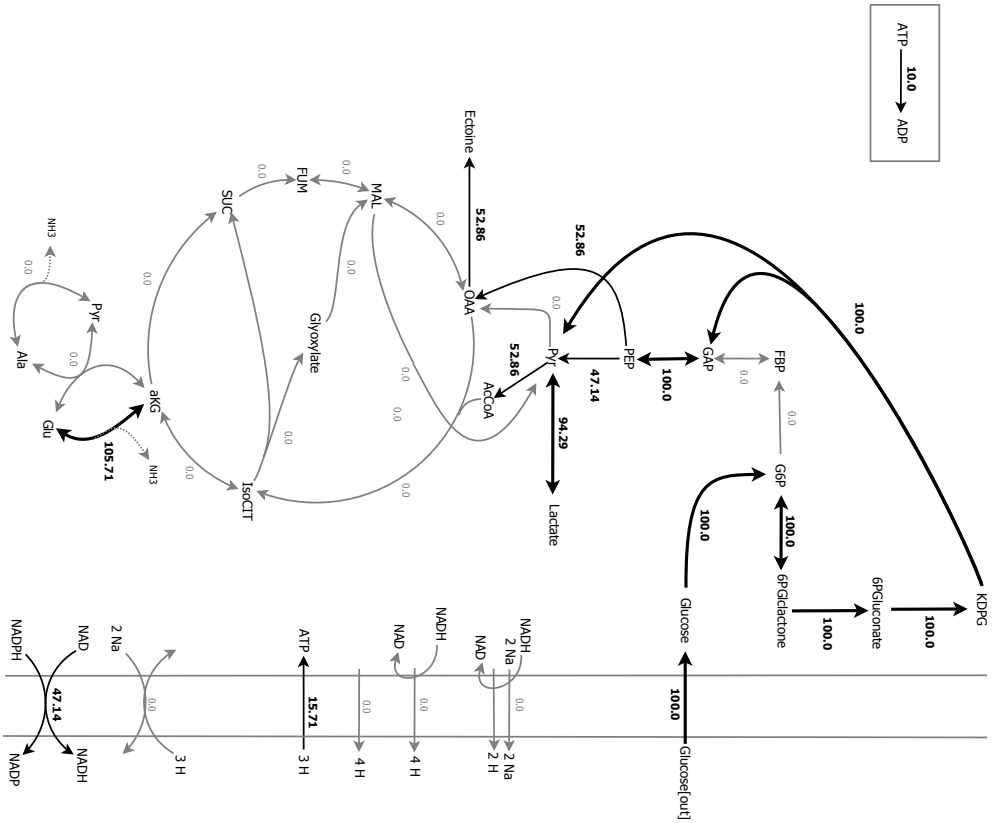
In a next step, the anaplerotic reactions (besides malic enzyme) are compared with each other to see, if the thermodynamic analysis agrees with the ranking of the FBA or if it draws a different picture.

8.6 ANAPLEROTIC REACTIONS (WITH AMMONIA ASSIMILATION)

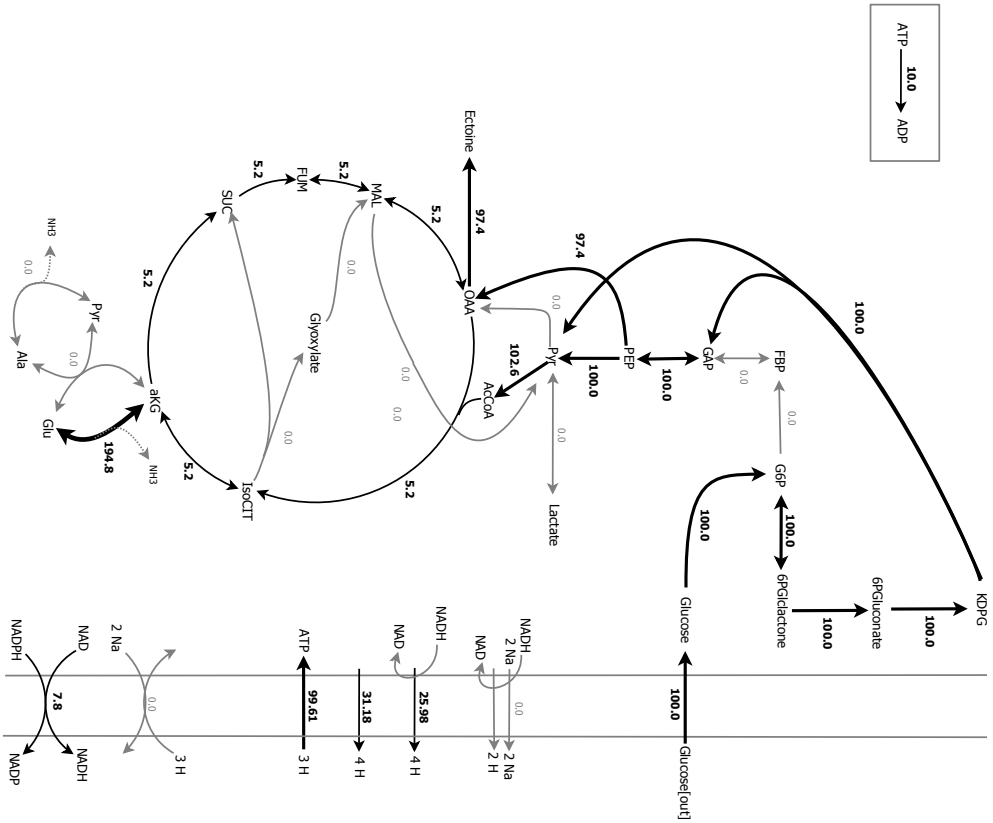
This present section compares the different in silico mutants, which have been defined in the FBA. The wildtype, as presented in the FBA, will not play a role anymore, because malic enzyme is now defined as catalyzing the backward reaction. Therefore, this new defined wildtype replaces the malic mutant. Both produce the same flux distributions in the FBA. One might notice that the previous section already discussed the new wildtype. It remains the malic OAD mutant, which is now called the OAD mutant. The malic PEPC mutant corresponds to the new defined wildtype, since OAD is superior with respect to stoichiometry. To further analyze the glyoxylate shunt, another mutant is defined, the OAD PEPC mutant, which misses OAD as well as PEPC. The methods from the previous section are repeated with the objective of *ectoine optimization* and the results compared.

8.6.1 *OAD mutant*

The OAD mutant will make a start, which means that OAD is inactivated. As in the last section, the effects of proton-dependent and proton- as well as sodium-translocating complex are compared with each other. As one can see in Figure 57, the difference in yield between the flux distributions with C1 and $C1^{Na}$ is significantly higher, than in case of the wildtype, respective the PEPC mutant. If the proton-dependent channel is inactivated, the system skips the whole respiratory chain, besides the ATPase and makes use of anaerobic respiration, because the sodium, which would be pumped out through $C1^{Na}$, would have to be pumped in again, which costs energy. The yield of ectoine is much lower, than before (52.06 mmol/100 mmol Glc) and an ATP load above 120 mmol ATP/100 mmol Glc cannot be satisfied. If C1 is active, the yield of ectoine stays high, which shows that it is quite important for the cell to have the proton-dependent channel, if OAD is not available for some reason. Now, which kind of insights can be gained through a thermodynamic analysis? As expected, the anaerobic solution is more exergonic, because the system exhibits less fluxes and is therefore less restricted. Interestingly, the thermodynamic feasibility clearly improves with increasing salt concentration. Still, it is not an optimal solution, because of the low ectoine-yield. The version with the proton-dependent channel is quite similar to the result of the wildtype (see Figure 54



(a)



(b)

Figure 57: (a) shows the flux distribution of the OAD mutant for an ATP load of 10 mmol ATP/100 mmol Glc in case that $C1^{Na}$ is active but C1 not. Still, $C1^{Na}$ exhibits no flux and the flux distribution reveals that the system uses anaerobic respiration. (b) displays the flux distribution in case that C1 is active.

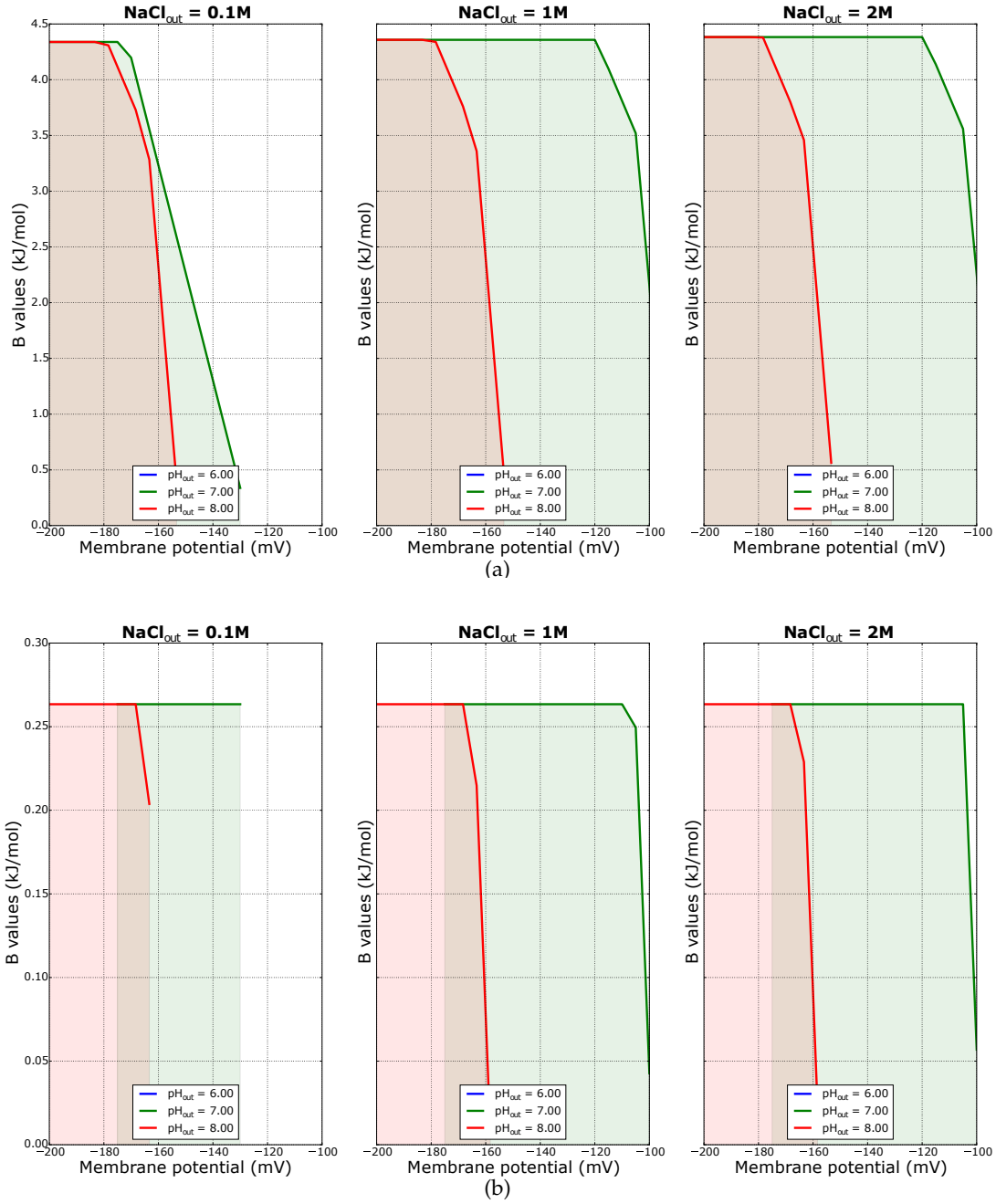


Figure 58: MDF solutions for the OAD mutant, which is forced to dismiss either C1 (a) or $C1^{Na}$ (b). As can be seen in (a), the lower the yield, the higher the B value.

(a)), although the OAD mutant is a little better with respect to high salt and an external pH of 8.

In case that both, C1 and $C1^{Na}$, are active in the FBA, it is not surprising that the resulting flux distribution only uses the proton-translocating complex, independent of the ATP load, which corresponds to the results in Figure 57 (b). The thermodynamic analysis stays the same as well (see Figure 58 (b)).

8.6.2 *OAD PEPC mutant*

When OAD and PEPC are inactivated simultaneously, the glyoxylate shunt is used instead. Again, the two versions of complex I are compared. First, the system without C1 and second the system, which exhibits C1 and $C1^{Na}$. In the second case, the FBA channels no flux through the channel, which builds the sodium-motive force, because OAD is not available. The results show that the case with only $C1^{Na}$ is the worst so far. The additional inactivation of PEPC therefore has a great effect on the efficiency of the system. No product is produced, see Figure 77. In case that the proton-motive force building channel is active, product can still be produced, although the product yield is much lower, than before. This is caused by the fact that carbon has to be burned in order to use the glyoxylate shunt.

Looking at the thermodynamics, it is obvious that there exists a trade-off between ectoine-yield and exergonism: B_{max} is much higher, than before and the feasible ranges are quite wide, see Figure 78. Still, the solution is not acceptable, since the yield is so low. Considering (b), one recognized the same solution as with the OAD mutant and C1 active (compare with Figure 58 (b)). Therefore, the knockout of PEPC in addition to OAD does not have any effect on the thermodynamics, if C1 is active, although it has an effect on the FBA: A lower yield of ectoine. Therefore, the glyoxylate shunt is still only the last resort, when it comes to ectoine production.

In a next step the effect of the ammonia assimilation with Ala-DH is analyzed.

8.6.3 *Ammonia assimilation*

The GDH is again inactivated in the FBA to gain a better understanding of the importance of the Ala-DH. As it turns out, the use of Ala-DH is not advantageous, confirming earlier results.

At first the OAD mutant is considered. Because the OAD mutant generally prefers the proton-translocating channel over the proton- and sodium-translocating channel, as has been seen in Section 8.6.1, the effect of using $C1^{Na}$ is compared with allowing both enzymes to be active at once. In case the proton-motive force building complex is not available, the system settles for anaerobic fermentation again, as displayed in the appendix in Figure 79 (a) (compare with Figure 57). The ammonia assimilation is done via Ala-DH. The thermodynamic analysis shows that the minimal driving force value is smaller, than with the use of GDH. Even if C1 is active in the FBA (resulting in Figure 79 (b)), the value of B decreases a little, from 0.26 kJ/mol to 0.23 kJ/mol (compare Figure 58 with Figure 80 (b)).

In case of the PEPC OAD mutant, the inactivation of C1 results in the same flux distribution, as in Figure 77, because ammonia assimilation is not taking place. The thermodynamic analysis is therefore the same as before. However,

if $C1^{Na}$ is inactivated, the FBA settles for a flux distribution similar to the one in Figure 77 (b), displayed in Figure 81 (a). The thermodynamic results, as displayed in Figure 81 (b), are comparable with Figure 78 (b) but the minimal driving force is again smaller.

The following can be concluded:

- Active GDH:
  - OAD mutant
    - The FBA returns a flux distribution with anaerobic respiration in case that only $C1^{Na}$ is active and the ectoine-yield is only 0.5 mmol/mmol Glc. If C1 is active, the yield is still high.
    - The B-values in case of the $C1^{Na}$ version are higher, since the system is less restricted. However, it is obvious that the OAD mutant depends on C1, due to the FBA results.
  - OAD PEPC mutant
    - Since OAD is still inactivated, C1 is the first complex I choice. The additional inactivation of PEPC further accentuates this point.
    - The additional inactivation of PEPC has no effect on the thermodynamics, if C1 is active.
  - The thermodynamic analysis does not reveal which anaplerotic reaction is preferable, although the FBA does.
- Active Ala-DH:
  - The thermodynamic feasibility is again negatively affected

8.7 BIOMASS

The available thermodynamic data does not cover the whole network of the core-model containing biomass but it is still possible to take a flux distribution, which ensures biomass maximization and analyze the thermodynamic feasibility of those reactions, which take part in the basic core-model (without biomass included). These results can then be compared with the ones gained with ectoine optimization, for instance with respect to the EMP and ED pathway.

As can be seen in Figure 59, the thermodynamic profile with respect to the ED pathway stays almost completely the same as in Figure 49. Only the step sizes of rn:R05605 and rn:R00703 are affected. The driving force of rn:R05605 becomes higher, while the driving force of rn:R00703 becomes smaller. The overall driving force is a bit higher. The step size in case of the EMP pathway changes as well. The driving force in general is smaller but the step sizes from rn:R01015 on are not that balanced anymore. The positive effect of sodium stays the same, although its not as significant as before.

Another possible analysis is the comparison of the proton- and sodium-translocating channels in the respiratory chain. As has been done before for

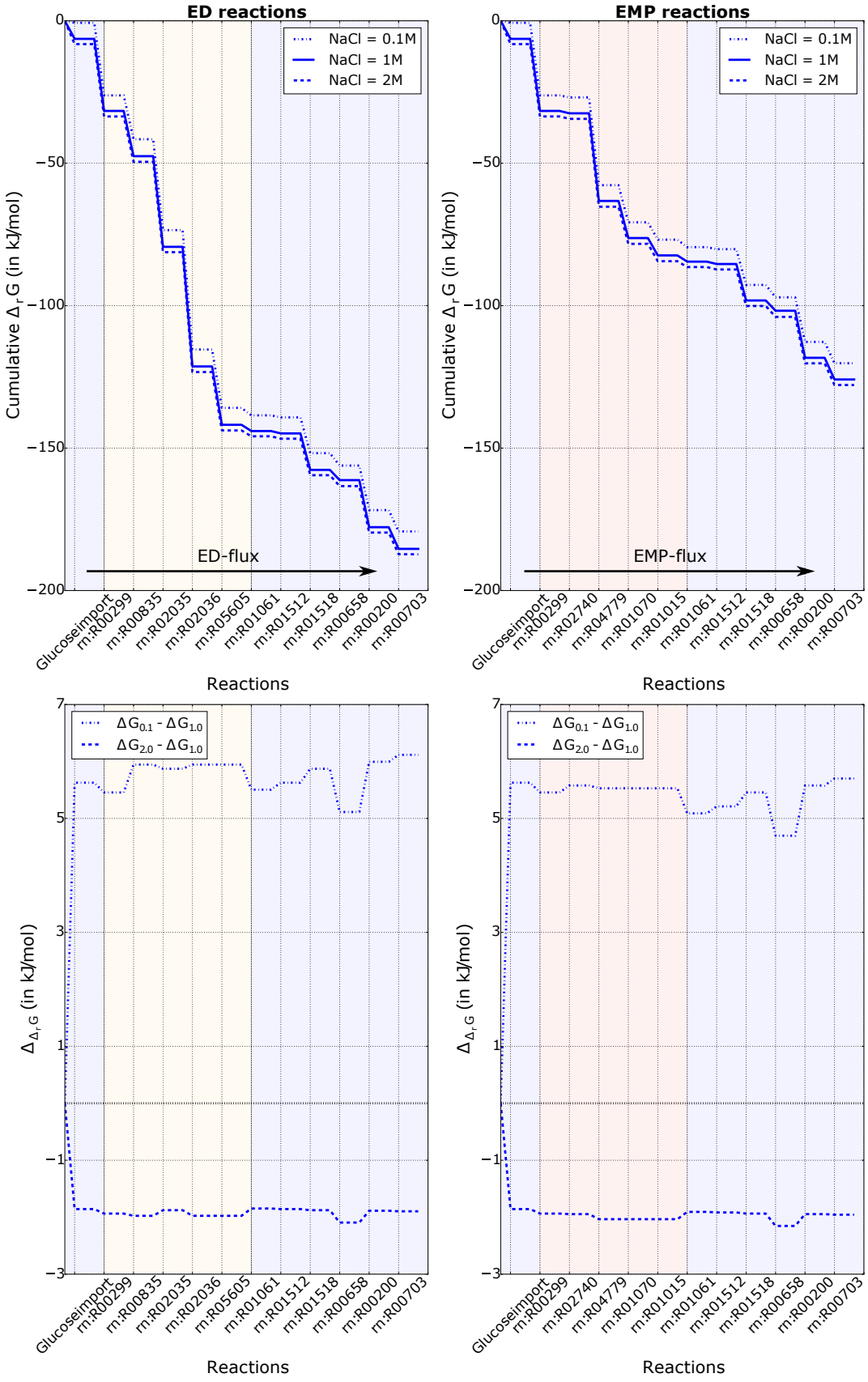
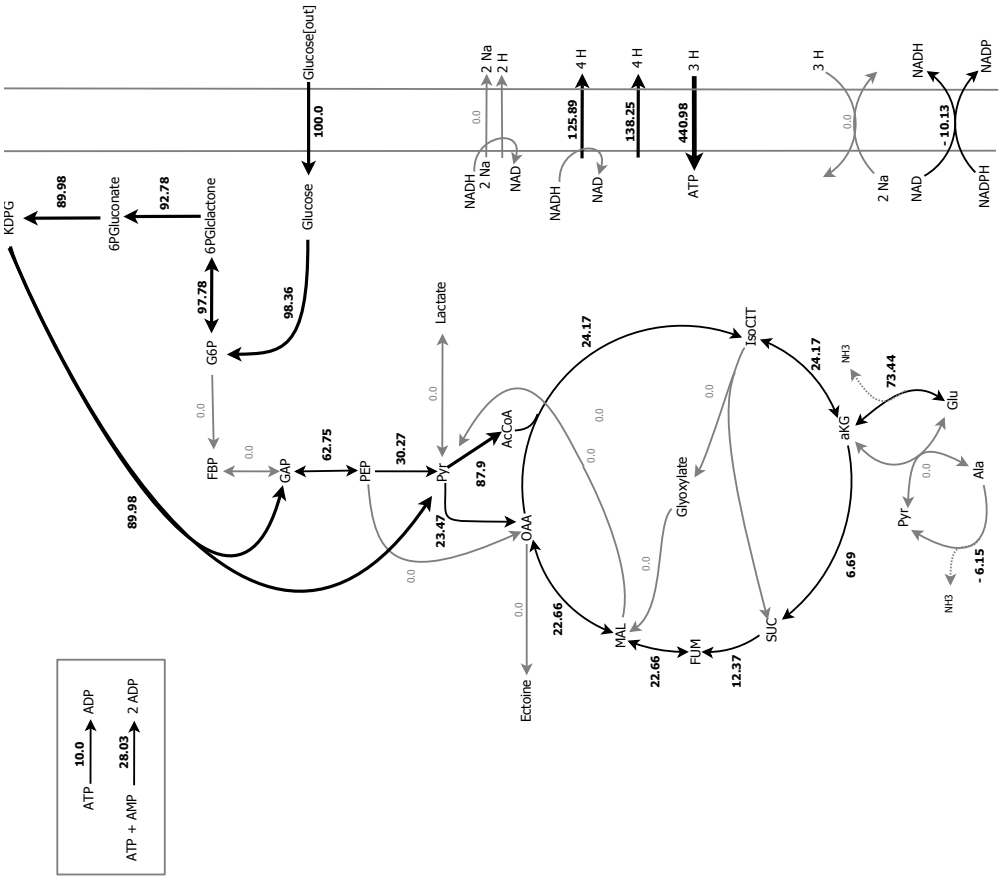
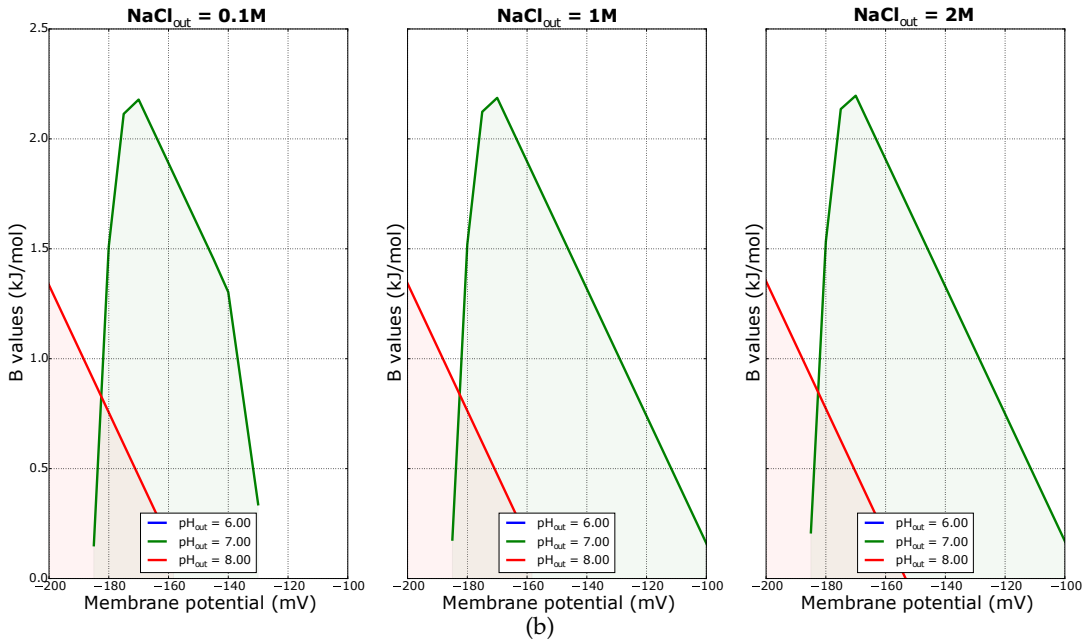


Figure 59: Glycolysis-comparison of the core-model with biomass and proton-translocating complex I, compare Figure 49 for further information.



(a)



(b)

Figure 60: (a) Fluxes of biomass optimization with C1 active. During biomass optimization, more fluxes are active, than actually displayed but in order to be comparable with Figure 53, etc., the reactions are left out. (b) shows the corresponding MDF solution at different conditions.

ectoine optimization, two FBAs have been executed. In both cases, glucose is set to 100 mmol/g Dw and the ATP load is assumed to be 10 mmol ATP/100 mmol Glc. In one case only the proton-translocating channel is active, in the other the proton- and sodium-translocating one. The flux distribution in case of an active $C1^{Na}$ channel behaves odd, due to a great import of protons through the transhydrogenase, and therefore its analysis is excluded. The flux distribution with active C1 can be seen in Figure 60 (a). It is analyzed with the MDF algorithm, while membrane potential, pH_{out} and salt content in the medium are varied, as has been done before for ectoine and energy optimization. The resulting feasible membrane potential intervals for the different pH values and salt contents can be found in (b). As can be seen, the optimal value for B is much higher, than for the ectoine or energy optimizations. The feasible membrane potential intervals from before are included in the ones for biomass optimization. Moreover, the salt content does not seem to have a great effect on the thermodynamic analysis. Therefore, it seems that the biomass-extension makes the system thermodynamically advantageous.

In conclusion, it has been shown that the Entner-Doudoroff pathway has its advantages with respect to thermodynamics, which answers the question, why *H. elongata* uses it. Moreover, the malic enzyme could be excluded as a possible anaplerotic reaction. Several analysis have been performed with respect to the comparison of C1 and $C1^{Na}$. As it turns out, the sodium-translocating channel has its advantages, since it allows more flexibility with respect to different membrane potential. However, the question of the thermodynamically superior anaplerotic reaction could not be answered so far. The analysis of a growth-optimizing flux distribution revealed that the inclusion of biomass in the model improves the thermodynamics.

The upcoming chapter describes the application of the reduction via thermodynamic shortening for the core-model without biomass and analyzes the resulting dynamic model with respect to stability.

9 DYNAMICS

The processes described in this chapter concern the application of thermodynamic shortening with respect to the model of *H. elongata*. The procedure results in a dynamic model, which is based on the steady state fluxes from FBA and the concentration values of the MDF algorithm. The model is subsequently analyzed with respect to its stability behavior and the dependencies of the matrices $\mathbf{F_x}$ and $\mathbf{F_P}$.

9.1 FROM CONSTRAINT-BASED TO DYNAMIC MODELING

Since the main scope of the thesis is the improved understanding of ectoine metabolism, the dynamical analysis will be restricted to the basic core-model, which lacks biomass production. Such a model can, for instance, be used to reflect the situation of an osmotic upshock, which results in the termination of growth and the initiation of ectoine production. Because the core-model not only includes glycolysis, TCA-cycle and ectoine production, but also the respiratory chain, as well as the synthesis of certain amino acids, such as serine and threonine, the total number of reactions reaches 79, accompanied by 69 dependent and 11 independent metabolites. The reduction of such a system to a size, which can be analyzed dynamically, is a challenge, since dynamically relevant information must not be lost. Two possible reduction approaches have already been discussed: the reduction based on time hierarchies (see Chapter 4) and thermodynamic shortening, which has been developed in the context of this work (see Chapter 5). It represents the preferred option since no extra algebraic constraints are necessary and the results of the thermodynamic analysis come into use.

The procedure begins with the application of FBA, which uses ectoine maximization as the objective and considers the ATP load to be 10 mmol ATP/100 mmol Glc. The resulting flux distribution is translated into a signvector and analyzed via the MDF algorithm. The external conditions used are the following: The NaCl concentration is chosen to be 1 mol/l, the pH_{out} is 7 and the membrane potential is set to -123 mV. Certain ratios for the cofactors are defined, the values are taken from Henry et al. It is known, that most reactions of glycolysis and TCA-cycle are generally in thermodynamic equilibrium [110, 180]. The Gibbs energies of mentioned reactions have therefore been bound between [-2.5; 2.5] kJ/mol [18]. The remaining irreversible reactions of glycolysis and TCA-cycle are the glucokinase (rn:R00299), the PFK (rn:R04779/rn:R00764), the pyruvate kinase (rn:R00200), the citrate synthase (rn:R00351), the oxoglutarate dehydrogenase (rn:R08549), the malate dehydrogenase (rn:R00342) and the isocitrate dehydrogenase reaction (rn:R00236 + rn:R01899).

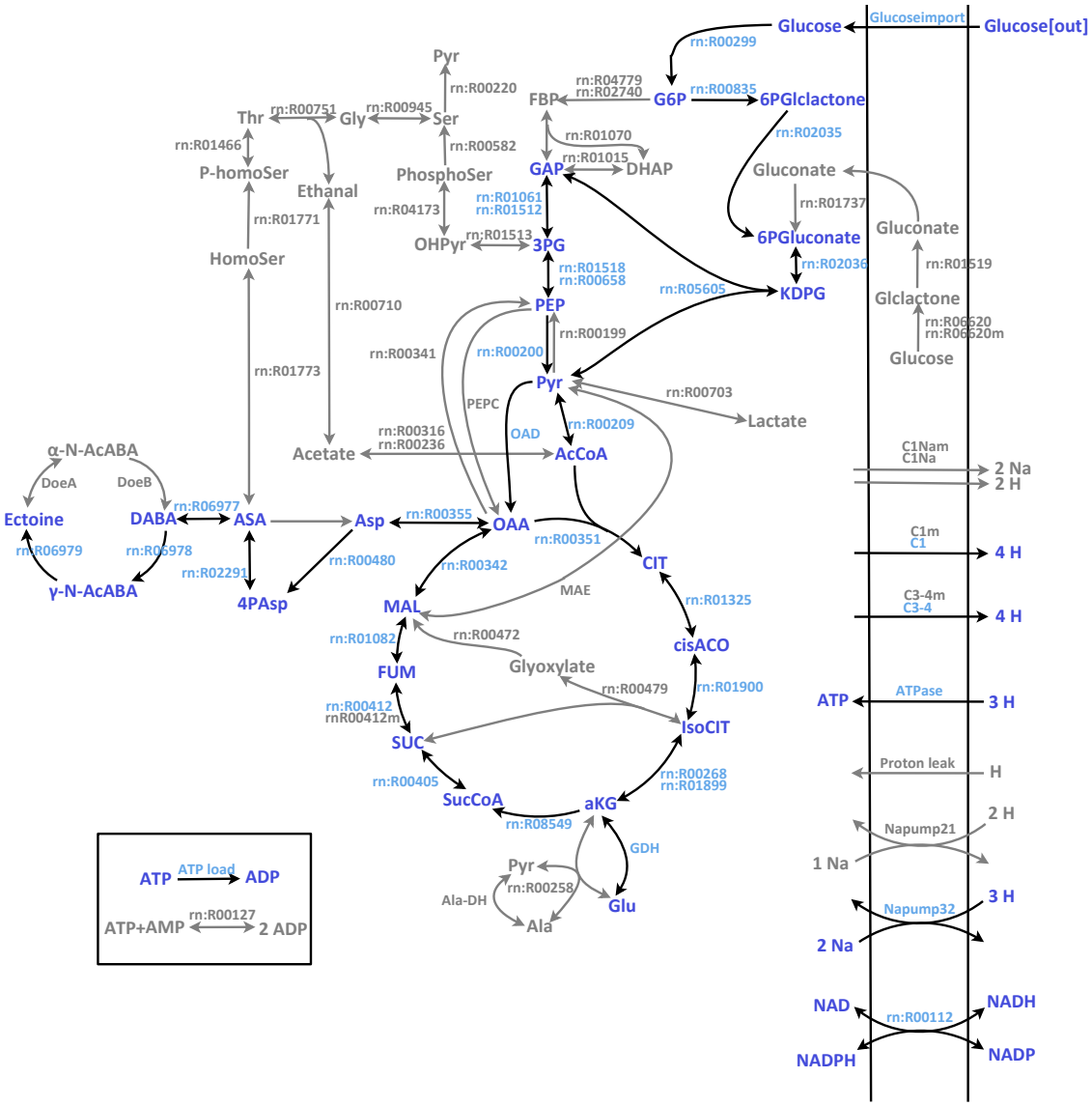


Figure 61: The pathway map is similar to Figure 24. However, all reactions, which are inactive in the flux distribution following ectoine optimization are displayed in grey. The metabolites, which are grey, are defined as independent metabolites. The active reactions and metabolites are displayed in blue.

The results are subsequently evaluated: Only the active fluxes are relevant, the reactions, which show no flux in the flux distribution, are therefore discarded and displayed in grey in Figure 61. This reduces the number of reactions to 36. Metabolites, which took only part in such reactions are then defined as independent and displayed in grey as well. They have to stay in the model, because they might have a regulatory effect on the remaining reactions, although they do not take part as substrate or product anymore. Furthermore, metabolites, like co-factors or in- and outputs are defined as independent metabolites as well, since they are constants. This procedure results in 29 dependent and 51 independent metabolites. The thermodynamic analysis provides a classifica-

tion, with respect to the distance to equilibrium, which can be used in thermodynamic shortening: The reactions, whose Gibbs energies lie in-between [-2.5; 2.5] kJ/mol are defined as reversible, all others are classified as irreversible. The 36 remaining reactions can therefore be divided into 16 reversible and 20 irreversible reactions. Values for the metabolite concentrations and equilibrium constants are also obtained. The K'_{eq} will be needed to calculate the bound metabolites (compare Equation 60).

As has been explained before, certain metabolites can be pooled together, see Section 4.5 and 5.2. The procedure results in 13 pools, which are going to be the new variables of the dynamic model:

- P_0 = α-D-Glucose-6-phosphate
- P_1 = L-2,4-Diaminobutyrate
- P_2 = Succinyl-CoA + CoA
- P_3 = Oxalosuccinate + cis-Aconitate + Isocitrate + Citrate
- P_4 = N-γ-Acetyldiaminobutyrate
- P_5 = Acetyl-CoA
- P_6 = 2-Oxoglutarate + Glutamate
- P_7 = Pyruvate + Succinyl-CoA + Succinate + L-Aspartic-4-semialdehyde + 4-Phospho-L-aspartate + Oxaloacetate + Fumarate + Malate + L-Aspartate
- P_8 = 3-Phospho-glycerate + 2-Phospho-D-glycerate + PEP + Glyceraldehyde-3-phosphate + 1,3-Bisphospho-D-glycerate
- P_9 = 6-phospho-D-gluconate
- P_{10} = KDPG
- P_{11} = 6-phospho-D-gluconolactone
- P_{12} = Glucose

An illustration of such a pool can be found in Figure 62. It displays the pool P_3, which concatenates a part of the TCA-cycle. Dependent metabolites, which are not connected to others via a reversible reaction, cannot be pooled together and therefore define a single pool, as can for instance be seen in case of pool P_0.

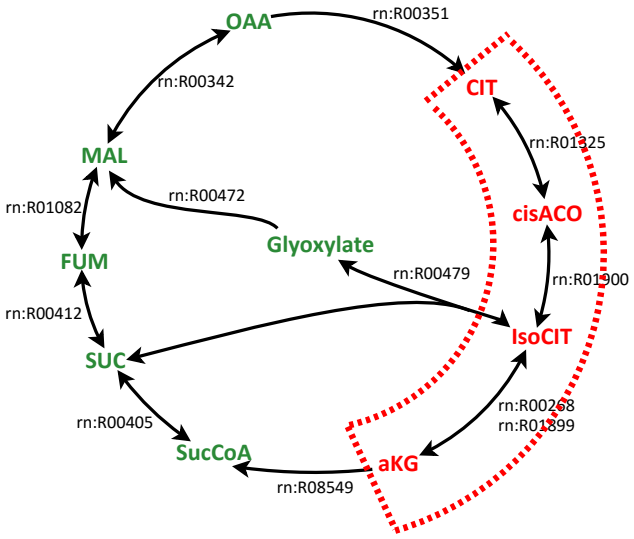


Figure 62: Visualization of pool 3.

Since 16 reactions are close to thermodynamic equilibrium, the dimension of $\overrightarrow{x}_B$ is 16, which leaves 13 free metabolites. The bound metabolites can be calculated with the help of the equilibrium constants, see Equation 60. The dimension of $\overrightarrow{x}_I$ is 51. The matrix $\mathbf{F_x}$ can be built as has been explained in Section 5.2. Its dimension is 20x80. Each substrate is expected to have a positive effect on its reaction, while each product is expected to have a negative effect, which makes the pathway reversible. It might initially seem odd that a product of an irreversible reaction influences the reaction, but as has been discussed in Section 3.3, the product only loses influences, if $P \ll K_p$ holds. Furthermore, it is known from Kadir et al. [76] that NADH has a negative effect on the citrate synthase. This fact is therefore included as an additional constraint. The exact values of the kinetic parameters are unknown, but the range is established [163]: It is assumed that most kinetic parameters are in a range of [-1; 1]. This fact enables the use of ensemble modeling, that is to sample the values of the kinetic parameters randomly [89]. The sign of the parameter depends on its effect. An activating effect goes along with a positive kinetic order, while an inhibiting effect is represented by a negative kinetic order. Allosteric regulations, as the effect of NADH on the citrate synthase, are expected to take on values between [-2; 2]. The part concerning the dependent metabolites ($[\mathbf{F_{x_B}}, \mathbf{F_{x_F}}]$) has been drawn as a color plot in Figure 63. One can see the different metabolite sections and identify the influence of each metabolite group: A positive influence is indicated in brown, a negative one in orange. The allosteric inhibition would be colored in orange but since NADH is an independent metabolite, it is not displayed.

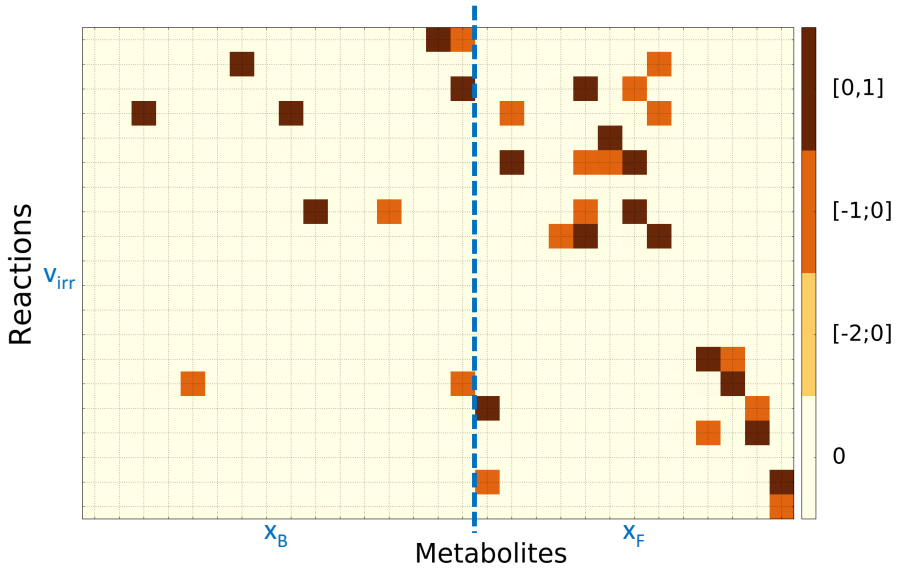


Figure 63: Matrix $[\mathbf{F}_{x_B}, \mathbf{F}_{x_F}]$ as a color plot.

Thousands of $\mathbf{F}_x$ matrices can be sampled, its values are uniformly distributed. For each of this matrices, the matrix $\mathbf{F}_P$ can be calculated, which in turn can be used to calculate the system of differential equations of the pools (the reaction names of the model are replaced by short forms to simplify the reading):

$$
\begin{aligned}
\dot{P}_0 &= -v_{15} + v_{18} \\
\dot{P}_1 &= v_3 - v_5 \\
\dot{P}_2 &= -v_2 + v_5 + v_7 \\
\dot{P}_3 &= -v_1 + v_7 \\
\dot{P}_4 &= -v_4 + v_5 \\
\dot{P}_5 &= v_2 - v_5 - v_7 \\
\dot{P}_6 &= v_1 - v_8 \\
\dot{P}_7 &= v_0 - v_2 - v_3 - v_7 + v_8 + v_{14} \\
\dot{P}_8 &= -v_0 + v_{14} \\
\dot{P}_9 &= -v_{13} + v_{16} \\
\dot{P}_{10} &= v_{13} - v_{14} \\
\dot{P}_{11} &= v_{15} - v_{16} \\
\dot{P}_{12} &= -v_{18} + v_{19}
\end{aligned}
\tag{83}
$$

As has been done before, a pool diagram can be drawn, which displays all interconnections of the different pools, see Figure 64.

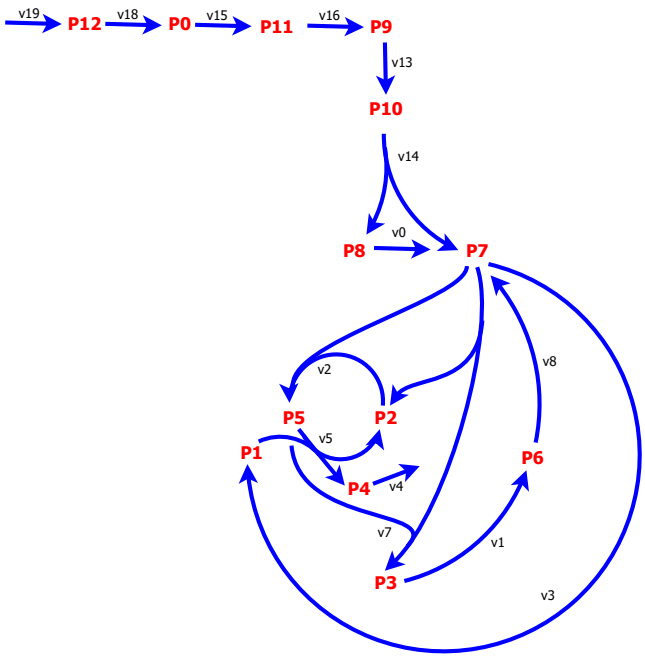


Figure 64: Pool diagram for the core-model. The arrows indicating regulation are excluded due to the complexity of the diagram.

Replacing the rate laws by power laws, depending on the pools, the differential equation system becomes:

$$
\begin{aligned}
\dot{P}_0 &= -\tilde{\gamma}_{15}\, P_0^{f_{15,0}}\, P_{11}^{f_{15,11}} + \tilde{\gamma}_{18}\, P_0^{f_{18,0}}\, P_{12}^{f_{18,12}} \\
\dot{P}_1 &= \tilde{\gamma}_3\, P_1^{f_{3,1}}\, P_2^{f_{3,2}}\, P_6^{f_{3,6}}\, P_7^{f_{3,7}} - \tilde{\gamma}_5\, P_1^{f_{5,1}}\, P_2^{f_{5,2}}\, P_4^{f_{5,4}}\, P_5^{f_{5,5}}\, P_7^{f_{5,7}} \\
\dot{P}_2 &= -\tilde{\gamma}_2\, P_2^{f_{2,2}}\, P_5^{f_{2,5}}\, P_7^{f_{2,7}} + \tilde{\gamma}_5\, P_1^{f_{5,1}}\, P_2^{f_{5,2}}\, P_4^{f_{5,4}}\, P_5^{f_{5,5}}\, P_7^{f_{5,7}} + \tilde{\gamma}_7\, P_2^{f_{7,2}}\, P_3^{f_{7,3}}\, P_5^{f_{7,5}}\, P_7^{f_{7,7}} \\
\dot{P}_3 &= -\tilde{\gamma}_1\, P_3^{f_{1,3}}\, P_6^{f_{1,6}} + \tilde{\gamma}_7\, P_2^{f_{7,2}}\, P_3^{f_{7,3}}\, P_5^{f_{7,5}}\, P_7^{f_{7,7}} \\
\dot{P}_4 &= -\tilde{\gamma}_4\, P_4^{f_{4,4}} + \tilde{\gamma}_5\, P_1^{f_{5,1}}\, P_2^{f_{5,2}}\, P_4^{f_{5,4}}\, P_5^{f_{5,5}}\, P_7^{f_{5,7}} \\
\dot{P}_5 &= \tilde{\gamma}_2\, P_2^{f_{2,2}}\, P_5^{f_{2,5}}\, P_7^{f_{2,7}} - \tilde{\gamma}_5\, P_1^{f_{5,1}}\, P_2^{f_{5,2}}\, P_4^{f_{5,4}}\, P_5^{f_{5,5}}\, P_7^{f_{5,7}} - \tilde{\gamma}_7\, P_2^{f_{7,2}}\, P_3^{f_{7,3}}\, P_5^{f_{7,5}}\, P_7^{f_{7,7}} \\
\dot{P}_6 &= \tilde{\gamma}_1\, P_3^{f_{1,3}}\, P_6^{f_{1,6}} - \tilde{\gamma}_8\, P_2^{f_{8,2}}\, P_6^{f_{8,6}}\, P_7^{f_{8,7}} \\
\dot{P}_7 &= \tilde{\gamma}_0\, P_2^{f_{0,2}}\, P_7^{f_{0,7}}\, P_8^{f_{0,8}} - \tilde{\gamma}_2\, P_2^{f_{2,2}}\, P_5^{f_{2,5}}\, P_7^{f_{2,7}} - \tilde{\gamma}_3\, P_1^{f_{3,1}}\, P_2^{f_{3,2}}\, P_6^{f_{3,6}}\, P_7^{f_{3,7}} \\
&\quad -\tilde{\gamma}_7\, P_2^{f_{7,2}}\, P_3^{f_{7,3}}\, P_5^{f_{7,5}}\, P_7^{f_{7,7}} + \tilde{\gamma}_8\, P_2^{f_{8,2}}\, P_6^{f_{8,6}}\, P_7^{f_{8,7}} + \tilde{\gamma}_{14}\, P_2^{f_{14,2}}\, P_7^{f_{14,7}}\, P_8^{f_{14,8}}\, P_{10}^{f_{14,10}} \\
\dot{P}_8 &= -\tilde{\gamma}_0\, P_2^{f_{0,2}}\, P_7^{f_{0,7}}\, P_8^{f_{0,8}} + \tilde{\gamma}_{14}\, P_2^{f_{14,2}}\, P_7^{f_{14,7}}\, P_8^{f_{14,8}}\, P_{10}^{f_{14,10}} \\
\dot{P}_9 &= -\tilde{\gamma}_{13}\, P_9^{f_{13,9}}\, P_{10}^{f_{13,10}} + \tilde{\gamma}_{16}\, P_9^{f_{16,9}}\, P_{11}^{f_{16,11}} \\
\dot{P}_{10} &= \tilde{\gamma}_{13}\, P_9^{f_{13,9}}\, P_{10}^{f_{13,10}} - \tilde{\gamma}_{14}\, P_2^{f_{14,2}}\, P_7^{f_{14,7}}\, P_8^{f_{14,8}}\, P_{10}^{f_{14,10}} \\
\dot{P}_{11} &= \tilde{\gamma}_{15}\, P_0^{f_{15,0}}\, P_{11}^{f_{15,11}} - \tilde{\gamma}_{16}\, P_9^{f_{16,9}}\, P_{11}^{f_{16,11}} \\
\dot{P}_{12} &= -\tilde{\gamma}_{18}\, P_0^{f_{18,0}}\, P_{12}^{f_{18,12}} + \tilde{\gamma}_{19}\, P_{12}^{f_{19,12}}
\end{aligned}
\tag{84}
$$

The $\tilde{\gamma}_i$ are calculated via Equation 31 with the help of the steady state fluxes and the steady state concentrations. Their denominators do not exhibit independent metabolites, since these canceled out, as has been explained in Chapter 5. The kinetic parameters are collected in matrix $\mathbf{F_P}$.

The interdependence of the matrices $\mathbf{F'_x}$ and $\mathbf{F_P}$ will be examined in the following section.

9.2 INTERDEPENDENCE OF THE PARAMETER MATRICES

It has been shown in Equation 75 that $\mathbf{F_P}$ can be directly calculated from $\mathbf{F'_x}$ (=[$\mathbf{F_{x_F}}$, $\mathbf{F_{x_B}}$]). Both matrices are linked by matrix $\mathbf{U}$, which is a (29x13)-matrix with full rank. Therefore, the distributions of the parameters of $\mathbf{F'_x}$ and $\mathbf{F_P}$ are related as well. The question is how. The parameters of the original $\mathbf{F_x}$ are uniformly distributed and mostly in the interval [-1; 1], but how does this affect $\mathbf{F_P}$?

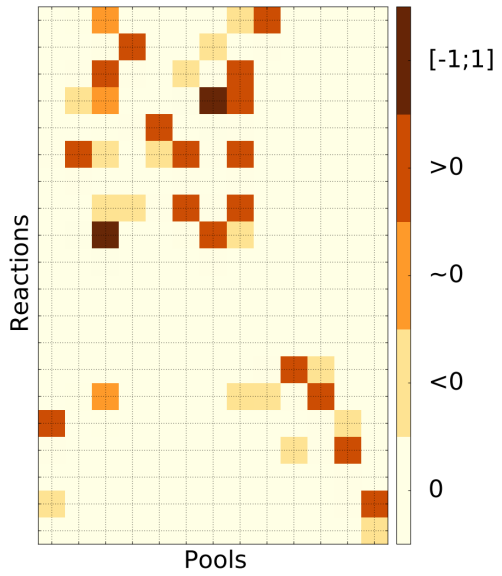
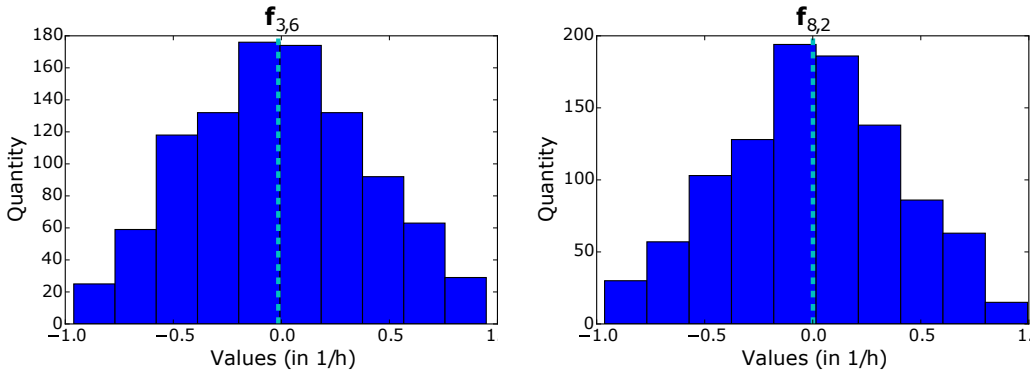


Figure 65: Matrix $\mathbf{F'_P}$ as a color plot. Most parameters show clear tendencies with respect to their sign.

1000 sets of parameters of $\mathbf{F_P}$ are calculated with Equation 75 and afterwards examined. As it turns out, most parameters of $\mathbf{F_P}$ are uniformly distributed as well but they tend to have a certain sign and stay mostly either in the interval [-1; 0] or [0; 1], see Figure 65. Two parameters are not uniformly but triangular distributed: $f^P_{3,6}$ and $f^P_{8,2}$. They both lie in the interval [-1; 1], see the histograms in Figure 66.



(a) (b)

Figure 66: Histograms of two $\mathbf{F_P}$ parameters: $f^P_{3,6}$ and $f^P_{8,2}$.

These results allow to sample the values for $\mathbf{F_P}$ directly, without the need for $\mathbf{F_x}$ first.

The subsequent section will analyze the dynamics of two different pathway designs in unreduced and reduced form with respect to their stability behavior.

9.3 STABILITY ANALYSIS

Chapter 5 listed some of the criteria, which should be fulfilled by a biochemical system. One of those is stability, because it is essential for survival. To understand the interplay between kinetics and thermodynamics, one considers the simplest possible system able to sustain the fluxes, provided by FBA. For this purpose, all allosteric modulations are set to zero and only the action of substrates and products are considered as influences on reactions. As seen before, there are two alternative strategies. On the one hand, the efficiency of the enzymes can be maximized by minimizing the inhibition that the reactions undergo by its products (economic design). On the other hand, a pathway can response better to changes in supply or demand, when reactions respond to both, substrates and products (responsive design). In order to compare these strategies, 1000 models of each type have been sampled and the eigenvalues of their Jacobian matrices analyzed for stability (absence of eigenvalues with positive real parts). The results are displayed in the subsequent figures.

In the left panel of Figure 67, a scatter plot of all eigenvalues is displayed, normed by the maximal eigenvalue of each system. It can be seen that the systems exhibit a great amount of eigenvalues with positive real part (displayed in white). Practically each system exhibits at least one eigenvalue with positive real part, considering the histogram in the right panel. It turns out that these systems are almost never stable.

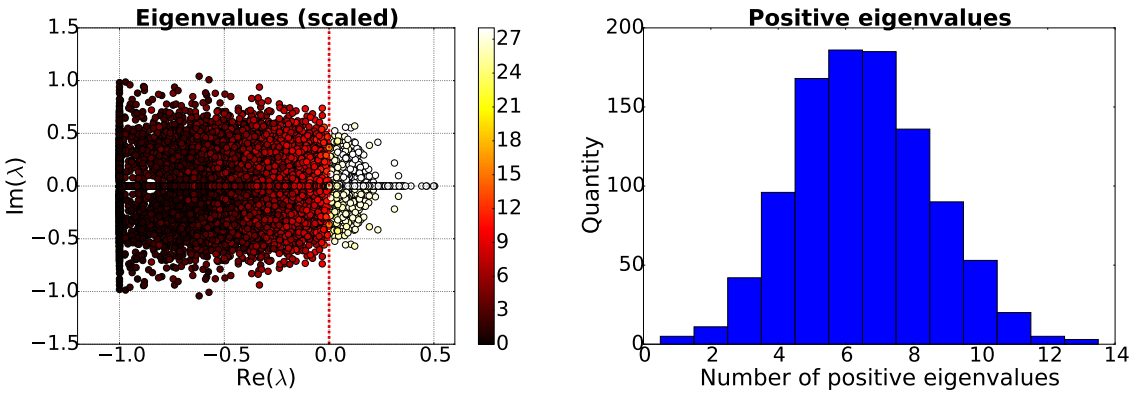


Figure 67: The left panel displays a scatter plot of the eigenvalues of each reversible system. The eigenvalues of each system are scaled and furthermore sorted in increasing order and color coded accordingly. The right panel reveals the distribution of the number of positive eigenvalues in each system. There are usually about six or seven positive eigenvalues.

A similar analysis for the irreversible case shows a clear tendency towards more stable systems. The results are displayed in Figure 68. The left panel exhibits the normed eigenvalues of the systems, which seem to be mostly negative. The right panel confirms this assumption: The systems tend to exhibit less eigenvalues with positive real part, than before. More or less 40% of the irreversible systems turn out to be stable.

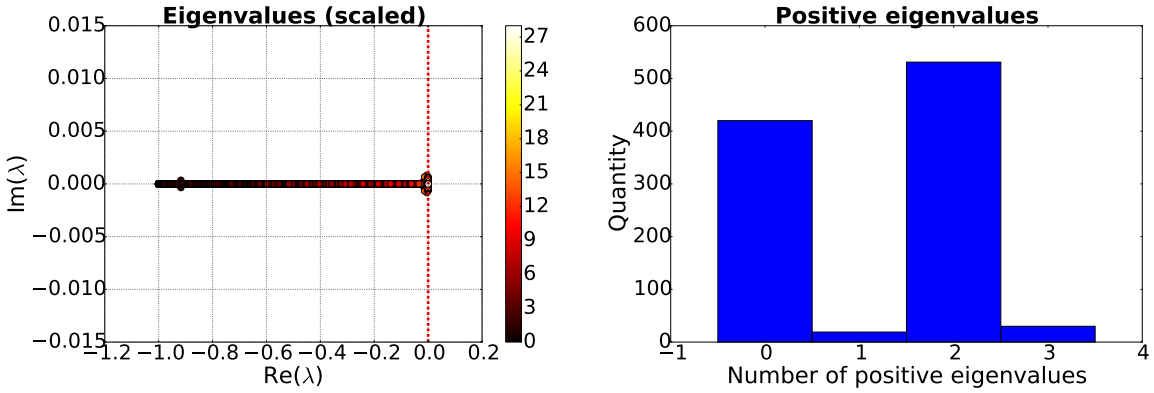


Figure 68: The left panel displays a scatter plot of the eigenvalues of each irreversible system. The eigenvalues of each system are again scaled and furthermore sorted in increasing order and color coded accordingly. The right panel reveals the distribution of the number of positive eigenvalues in each system. As can be seen, the number of positive eigenvalues is low.

Although it may be counter intuitive, it has been shown that a fully irreversible step at the beginning of a pathway increases its stability [9]. The rational for this being that the product inhibition on a sequence of consecutive reactions is dynamically equivalent to a feedback inhibition, which is known to cause instability in long pathways [144]. Of course, the extrapolation from simple pathways to full networks is not straight forward, which explains the need

to generalize concepts, such as thermodynamic shortening to such complex cases.

As it turns out, applying thermodynamic shortening on either of the two different system designs, improves their stability significantly. 10% of the reduced reversible and 99.9% of the reduced irreversible systems turn out to be stable.

Figure 69 depicts the dynamics of such a stable shortened system in response to a bolus of glucose. The pools have been normalized by their steady state values, while the initial point of the simulation has been the steady state solution of the pools, with one exception, pool 12, which represents glucose. Its normalized steady state value has been changed to five, to cause a disturbance of the system. As can be seen, the yellow line, which represents pool 12, decreases to its steady state, while all other pools are shortly disturbed but return to their steady state as well. The simulation serves as a proof of concept, leaving additional degrees of freedom for selection, such as logarithmic gains or realistic time scales.

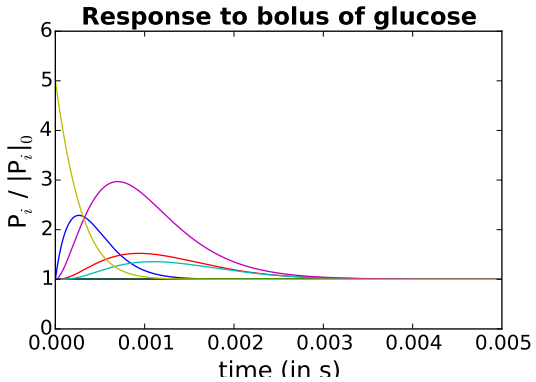


Figure 69: Simulation of a stable reduced system in response to a bolus of glucose. The pools are normalized by their steady state values.

In conclusion, it has been shown that the workflow is functional and able to return a stable dynamic model. The explained method shows great potential as a starting point to dynamic models. The inclusion of experimental data and realistic time scales would cause further improvements of the method. Moreover, the stochastic approach concerning the distributions can be combined with the one from Michael Savageau, concerning MCC, to apply further criteria of functional effectiveness, besides stability [5–9, 150].

10 CONCLUSIONS AND OUTLOOK

The aim of the present work has been to improve the understanding of the metabolism of *H. elongata* with the help of mathematical modeling. The idea was to reduce the uncertainty of the system and to discard unsound hypothesis a priori, before applying any experimentally based information. Thus, the need for a framework emerged, which combines different kinds of modeling approaches in a sequential manner and hence reduces the degrees of freedom. As can be seen in the results section, the workflow has successfully been established and applied with respect to *H. elongata*. A short outline of the gained insights into the metabolism will be summarized below:

STOICHIOMETRY The stoichiometric analysis confirmed the great influence of the anaplerotic reactions, especially on ectoine production. Moreover, a ranking of these reactions could be established. It has been shown that the variation of the proton leak reaction lowers the product yields to values, which are comparable with fermentation data and therefore biologically more realistic.

THERMODYNAMICS The thermodynamic analysis helped clarifying the evolutionary success of the Entner-Doudoroff pathway in so many bacteria [82], as well as the physiological role of the PPi-PFK. Furthermore, it has been revealed that MAE is not viable as an anaplerotic reaction, pinpointing how limitations of FBA can be overcome by a subsequent thermodynamic analysis. It has also been shown that a sodium-dependent complex I enables *H. elongata* to tap the sodium gradient to provide exergonism, where is needed, enhancing the metabolism. The comparison of the flux distributions of ectoine- and energy-optimization revealed that both exhibit thermodynamic bottlenecks. In addition, the Ala-DH reaction seems to improve with increasing salt-concentration.

DYNAMICS The model has been reduced with thermodynamic shortening and the information from FBA and MDF, which successfully bridges the gap between constraint-based and dynamic modeling. In addition, it has been possible to obtain stable dynamic models, especially with the reduced system. The results demonstrate the success of the established framework, which already limits the possible dynamics of the system. Moreover, the interdependencies of the matrices $\mathbf{F_x}$ and $\mathbf{F_P}$ have been revealed.

The current dynamic analysis considers only single steady states. However, the cell switches between different states, like energy- and ectoine-production. Instead of analyzing them independently, as has been done so far, they should be linked and the transition in-between those states should be analyzed with the application of an additional approach: The Design Space Analysis [49].

Moreover, the statistical MCC-version of Michael Savageau [5–9, 150] can be applied to the present dynamical model to examine the regulation more closely.

APPENDIX

A

APPENDIX

A.1 BIOMASS COMPOSITION

The section contains most of the calculations done with respect to biomass composition. The part about the proteins is missing, because the calculations are easily done with the proportions already given ([109, 164]). The glycogen synthesis is left out as well, because it has already been calculated in the main text.

A.1.1 *DNA*

The formulas for DNA and RNA synthesis have been provided by Stephanopoulos et al. [164] but as a test, the calculations are repeated.

The amount percent (in µmol/g DW) have been given by Neidhardt. The molecular mass (in g/mol) for each nucleotide is known as well, see Table 2.

This way, the molecular mass of DNA can be calculated:

$$\begin{aligned} 0.2465\,\mathrm{dAMP} + 0.2465\,\mathrm{dTMP} + 0.253\,\mathrm{dCMP} + 0.253\,\mathrm{dGMP} &= 326.63\,\mathrm{g/mol} \\ &= 0.3266\,\mathrm{g/mmol} \end{aligned}$$

Because the DNA synthesis involves the investment of energy, the overall reaction can be calculated as follows:

$$3.4\,\mathrm{ATP} + 0.25\,\mathrm{dAMP} + 0.25\,\mathrm{dGMP} + 0.25\,\mathrm{dTMP} + 0.25\,\mathrm{dCMP} \Longrightarrow 3.4\,\mathrm{ADP} + 3.4\,\mathrm{Pi} + \mathbf{1\,mmol}\,\mathrm{DNA}$$

$$3.4\,\mathrm{ATP} + 0.25\,\mathrm{dAMP} + 0.25\,\mathrm{dGMP} + 0.25\,\mathrm{dTMP} + 0.25\,\mathrm{dCMP} \Longrightarrow 3.4\,\mathrm{ADP} + 3.4\,\mathrm{Pi} + \mathbf{0.33\,g}\,\mathrm{DNA}$$

$$\tfrac{1}{0.33}(3.4\,\mathrm{ATP} + 0.25\,\mathrm{dAMP} + 0.25\,\mathrm{dGMP} + 0.25\,\mathrm{dTMP} + 0.25\,\mathrm{dCMP} - 3.4\,\mathrm{ADP} - 3.4\,\mathrm{Pi}) \Longrightarrow \mathbf{1\,g}\,\mathrm{DNA}$$

The results are similar enough that there exists no conflict in taking the given formulas ([109, 164]).

Metabolite	**Amount percent (µmol/g DW)**	**proportion**	**Molecular mass (in g/mol)**
dAMP	24.7	0.2465	331.2
dCMP	25.4	0.253	307.2
dGMP	25.4	0.253	347.2
dTMP	24.7	0.2465	322.2

Table 2: Molecular mass of DNA components

Metabolite	Amount percent (μmol/g DW)	proportion	Molecular mass (in g/mol)
AMP	165	0.262	347.2
CMP	126	0.2	323.2
GMP	203	0.322	363.2
UMP	136	0.216	324.2

Table 3: Molecular mass of RNA components

A.1.2 *RNA*

The procedure is the same as with the DNA. The sources stay the same as well, see Table 3:

The molecular mass of RNA:

$$\begin{aligned} 0.26\,\mathrm{AMP} + 0.2\,\mathrm{CMP} + 0.32\,\mathrm{GMP} + 0.22\,\mathrm{UMP} &= 342.584\ \mathrm{g/mol} \\ &= 0.3426\ \mathrm{g/mmol} \end{aligned}$$

Because the RNA synthesis involves the investment of energy as well, the overall reaction can be calculated as follows:

$$2.4\,\mathrm{ATP} + 0.26\,\mathrm{AMP} + 0.2\,\mathrm{CMP} + 0.32\,\mathrm{GMP} + 0.22\,\mathrm{UTP} \Longrightarrow 2.4\,\mathrm{ADP} + 2.4\,\mathrm{Pi} + \mathbf{1\,mmol}\,\mathrm{RNA}$$

$$2.4\,\mathrm{ATP} + 0.26\,\mathrm{AMP} + 0.2\,\mathrm{CMP} + 0.32\,\mathrm{GMP} + 0.22\,\mathrm{UTP} \Longrightarrow 2.4\,\mathrm{ADP} + 2.4\,\mathrm{Pi} + \mathbf{0.34\,g}\,\mathrm{RNA}$$

$$\tfrac{1}{0.34}(2.4\,\mathrm{ATP} + 0.26\,\mathrm{AMP} + 0.2\,\mathrm{CMP} + 0.32\,\mathrm{GMP} + 0.22\,\mathrm{UTP} - 2.4\,\mathrm{ADP} - 2.4\,\mathrm{Pi}) \Longrightarrow \mathbf{1\,g}\,\mathrm{RNA}$$

A.1.3 *Lipids*

A.1.3.1 *Acyl-CoA*

In order to calculate the mass of the Acyl-CoA, its synthesis equation is needed. For that the proportions from Pramanik [131] are used. But these are in g% not mol% and they therefore have to be recalculated (it is known that they are g%, because the reference of the paper included unidentified lipids, which cannot be given in mol).

The molecular mass is calculated as follows (compare Table 4):

$$\begin{aligned} \frac{0.38}{256} + \frac{0.11}{254} + \frac{0.18}{282} + \frac{0.03}{228} + \frac{0.08}{226} + \frac{0.16}{270} + \frac{0.01}{284} + \frac{0.06}{296} &= 3.8 \cdot 10^{-3}\ \mathrm{g/mol} \\ &= 3.8\ \mathrm{g/mmol} \end{aligned}$$

And with this, the synthesis equation can be calculated in mmol:

Metabolite	Molecular mass (in g/mol)	Total fatty acid (in g%)
Palmitoyl	256	38.23
Palmotoleoyl	254	10.74
Oleoyl	282	17.91
Myristoyl	228	2.68
Myristoleoyl	226	7.7
Heptadecanoyl	270	16.11
cis-Vaccenoyl	284	0.9
Nonadecenoyl	296	5.73

Table 4: Molecular mass and composition of fatty acids

Metabolite	Molecular mass (in g/mol)
Acyl	22.97
Serine	105.09
Glycerol-3-phosphate	172.07

Table 5: Molecular mass of PE components

$$\frac{0.38}{256}\,\text{Palmitoyl-CoA} + \frac{0.11}{254}\,\text{Palmitoleoyl-CoA} + \frac{0.18}{282}\,\text{Oleoyl-CoA} + \frac{0.03}{228}\,\text{Myristoyl-CoA} + \frac{0.08}{226}\,\text{Myristoleoyl-CoA} + \frac{0.16}{270}\,\text{Heptadecanoyl-CoA} + \frac{0.01}{284}\,\text{cis-Vaccenoyl-CoA} + \frac{0.06}{296}\,\text{Nonadecenoyl-CoA} \implies \mathbf{1\ g}\ \text{Acyl-CoA}$$

$$\frac{0.38}{256}\,\text{Palmitoyl-CoA} + \frac{0.11}{254}\,\text{Palmitoleoyl-CoA} + \frac{0.18}{282}\,\text{Oleoyl-CoA} + \frac{0.03}{228}\,\text{Myristoyl-CoA} + \frac{0.08}{226}\,\text{Myristoleoyl-CoA} + \frac{0.16}{270}\,\text{Heptadecanoyl-CoA} + \frac{0.01}{284}\,\text{cis-Vaccenoyl-CoA} + \frac{0.06}{296}\,\text{Nonadecenoyl-CoA} \implies \mathbf{3.8\ mmol}\ \text{Acyl-CoA}$$

$$\frac{0.38}{256\cdot 3.8}\,\text{Palmitoyl-CoA} + \frac{0.11}{254\cdot 3.8}\,\text{Palmitoleoyl-CoA} + \frac{0.18}{282\cdot 3.8}\,\text{Oleoyl-CoA} + \frac{0.03}{228\cdot 3.8}\,\text{Myristoyl-CoA} + \frac{0.08}{226\cdot 3.8}\,\text{Myristoleoyl-CoA} + \frac{0.16}{270\cdot 3.8}\,\text{Heptadecanoyl-CoA} + \frac{0.01}{284\cdot 3.8}\,\text{cis-Vaccenoyl-CoA} + \frac{0.06}{296\cdot 3.8}\,\text{Nonadecenoyl-CoA} \implies \mathbf{1\ mmol}\ \text{Acyl-CoA}$$

A.1.3.2 *Phosphatidylethanolamine (PE)*

The synthesis equation of PE is [110] (see Table 5):

$$\begin{aligned} 2\,\text{Acyl} + \text{Serine} + \text{Glycerol-3-phosphate} &= 323\,\text{g/mol} \\ &= 0.32\,\text{g/mmol} \end{aligned}$$

Thus, its molecular mass in g is:

$$2\,\text{Acyl-CoA} + \text{Serine} + \text{Glycerol-3-phosphate} + 3\,\text{ATP} \Longrightarrow 2\,\text{CoA} + 3\,\text{AMP} + 3\,\text{PPi} + CO_2 + \mathbf{1\,mmol}\,\text{PE}$$
$$2\,\text{Acyl-CoA} + \text{Serine} + \text{Glycerol-3-phosphate} + 3\,\text{ATP} \Longrightarrow 2\,\text{CoA} + 3\,\text{AMP} + 3\,\text{PPi} + CO_2 + \mathbf{0.32\,g}\,\text{PE}$$

$$\tfrac{1}{0.32}(2\,\text{Acyl-CoA} + \text{Serine} + \text{Glycerol-3-phosphate} + 3\,\text{ATP} - 2\,\text{CoA} - 3\,\text{AMP} - 3\,\text{PPi} - CO_2) \Longrightarrow \mathbf{1\,g}\,\text{PE}$$

A.1.3.3 *Phosphatidylglycerol (PG)*

The synthesis equation of PG is [110] (see Table 5):

$$\begin{aligned} 2\,\text{Acyl} + 2\,\text{Glycerol-3-phosphate} &= 390.088\,g/mol \\ &= 0.39\,g/mmol \end{aligned}$$

Thus:

$$2\,\text{Acyl-CoA} + 2\,\text{Glycerol-3-phosphate} + 3\,\text{ATP} + H_2O \Longrightarrow 2\,\text{CoA} + 3\,\text{AMP} + 3\,\text{PPi} + CO_2 + \mathbf{1\,mmol}\,\text{PG}$$
$$2\,\text{Acyl-CoA} + 2\,\text{Glycerol-3-phosphate} + 3\,\text{ATP} + H_2O \Longrightarrow 2\,\text{CoA} + 3\,\text{AMP} + 3\,\text{PPi} + CO_2 + \mathbf{0.39\,g}\,\text{PG}$$

$$\tfrac{1}{0.323}(2\,\text{Acyl-CoA} + 2\,\text{Glycerol-3-phosphate} + 3\,\text{ATP} + H_2O - 2\,\text{CoA} - 3\,\text{AMP} - 3\,\text{PPi} - CO_2) \Longrightarrow \mathbf{1\,g}\,\text{PG}$$

A.1.3.4 *Cardiolypin (CL)*

The synthesis equation of CL originates from Lehninger ([110]):
CL: 2 PG $\Longrightarrow$ *CL* + Glycerol

$$4\,\text{Acyl-CoA} + 4\,\text{Glycerol-3-phosphate} + 6\,\text{ATP} + 2\,H_2O \Longrightarrow \text{CL} + \text{Glycerol} + 4\,\text{CoA} + 6\,\text{AMP} + 6\,\text{PPi}$$

The molecular mass in g/mmol is (see Table 5):

$$\begin{aligned} 4\,\text{Acyl} + 4\,\text{Glycerol-3-phosphate} &= 780.176\,g/mol \\ &= 0.78\,g/mmol \end{aligned}$$

Thus:

$$4\,\text{Acyl-CoA} + 4\,\text{Glycerol-3-phosphate} + 6\,\text{ATP} + 2\,H_2O \Longrightarrow 4\,\text{CoA} + 6\,\text{AMP} + 6\,\text{PPi} + \text{Glycerol} + \mathbf{1\,mmol}\,\text{CL}$$
$$4\,\text{Acyl-CoA} + 4\,\text{Glycerol-3-phosphate} + 6\,\text{ATP} + 2\,H_2O \Longrightarrow 4\,\text{CoA} + 6\,\text{AMP} + 6\,\text{PPi} + \text{Glycerol} + \mathbf{0.78\,g}\,\text{CL}$$

$$\frac{1}{0.78}(4\,\text{Acyl-CoA} + 4\,\text{Glycerol-3-phosphate} + 6\,\text{ATP} + 2\,H_2O - 4\,\text{CoA} - 6\,\text{AMP} - 6\,\text{PPi} - \text{Glycerol}) \Longrightarrow \mathbf{1\,g}\,\text{CL}$$

The formula for the lipid composition is taken from the Vreeland-paper [178]:

0.074 CL + 0.42 PG + 0.383 PE + unknown $\Longrightarrow$ **1 g** *Lipids*

Because the paper has another metabolite, the unknown lipids, the proportions of PE, PG and CL do not sum 1 g, thus the proportions have to be recalculated, keeping the relations to one another. Therefore:

0.08 CL + 0.48 PG + 0.44 PE $\Longrightarrow$ **1 g** *Lipids*

A.1.4 *Peptidoglycan*

In order to calculate the mass of the peptidoglycan, its synthesis equation is needed, as well as the synthesis equations of the presursors. These can be achieved from Pramanik [131]:

UDPGA (UDP-N-Acetylglucosamine):
β-D-Fructose-6-phosphate + AcetylCoA + Glutamine + UTP $\Longrightarrow$ CoA + *UDPGA* + PPi + Glutamate

UDPAM (UDP-Acetylmuramate):
PEP + NADH + UDPGA + H $\Longrightarrow$ NAD + Pi + *UDPAM*

ASPSA (Aspartate-β-semialdehyde):
ATP + NADPH + H + L-Aspartate $\Longrightarrow$ ADP + NADP + Pi + *ASPSA*

DAP (L,L-2,6-Diaminopimelate):
Succinyl-CoA + Pyruvate + NADPH + ASPSA + H + Glutamate $\Longrightarrow$ Succinate + CoA + α-Ketoglutarate + NADP + *DAP*

The synthesis equation for peptidoglycan cannot be taken as well, because the composition in *H. elongata* is quite different from the one in *E. coli* [178]. The mentioned paper offers the precursors of peptidoglycan, as well as their molar ratios. In this case the ones for a NaCl-concentration of 3.4 mol/l have been taken, because these seem to be most convenient and it is mentioned that the ratios don't really change with salt. Subsequently, they have been recalculated in a way that makes everything dependent on UDPGAM instead of DAP. Subsequently, the molecular mass can be calculated:

Peptidoglycan:
UDPGAM + UDPGA + $\frac{2}{3}$ DAP + $\frac{3}{5}$ Leucine + $\frac{4}{15}$ Glycine + $\frac{5}{3}$ Alanine + Gluta-

Metabolite	Molecular mass (in g/mol)
GA	221.21
GAM	293.3
Alanine	89.1
Glutamate	147.13
Leucine	131.18
Glycine	75.067
DAP	190.296

Table 6: Molecular mass of peptidoglycan components

mate + 5 ATP $\Longrightarrow$ 5 ADP + 5 Pi + 2 UDP + *Peptidoglycan*

Thus:

$$\begin{aligned}\text{GAM} + \text{GA} + \tfrac{2}{3}\,\text{DAP} + \tfrac{3}{5}\,\text{Leucine} + \tfrac{4}{15}\,\text{Glycine} + \tfrac{5}{3}\,\text{Alanine} + \text{Glutamate} &= 1035.726\,\text{g/mol} \\ &= 1.036\,\text{g/mmol}\end{aligned}$$

So 1.04 g of peptidoglycan corresponds to 1 mmol of peptidoglycan.

Therefore:

$$\text{UDPGAM} + \text{UDPGA} + \tfrac{2}{3}\,\text{DAP} + \tfrac{3}{5}\,\text{Leucine} + \tfrac{4}{15}\,\text{Glycine} + \tfrac{5}{3}\,\text{Alanine} + \text{Glutamate} + 5\,\text{ATP} \implies 5\,\text{ADP} + 5\,\text{Pi} + 2\,\text{UDP} + \mathbf{1\,mmol}\,\mathtt{peptidoglycan}$$

$$\text{UDPGAM} + \text{UDPGA} + \tfrac{2}{3}\,\text{DAP} + \tfrac{3}{5}\,\text{Leucine} + \tfrac{4}{15}\,\text{Glycine} + \tfrac{5}{3}\,\text{Alanine} + \text{Glutamate} + 5\,\text{ATP} \implies 5\,\text{ADP} + 5\,\text{Pi} + 2\,\text{UDP} + \mathbf{1.04\,g}\,\mathtt{peptidoglycan}$$

$$\tfrac{1}{1.04}\,\text{UDPGAM} + \tfrac{1}{1.04}\,\text{UDPGA} + \tfrac{2}{3\cdot 1.04}\,\text{DAP} + \tfrac{3}{5\cdot 1.04}\,\text{Leucine} + \tfrac{4}{15\cdot 1.04}\,\text{Glycine} + \tfrac{5}{3\cdot 1.04}\,\text{Alanine} + \tfrac{1}{1.04}\,\text{Glutamate} + \tfrac{5}{1.04}\,\text{ATP} - \tfrac{5}{1.04}\,\text{ADP} - \tfrac{5}{1.04}\,\text{Pi} - \tfrac{2}{1.04}\,\text{UDP} \implies \mathbf{1\,g}\,\mathtt{peptidoglycan}$$

A.1.5 *Lipopolysaccharides*

In order to calculate the mass of the lipopolysaccharides, its synthesis equation, as well as the synthesis equations of the presursors are needed. These can be derived from Taymaz-Nikerel [167], the numbers in brakets are the equation numbers of the source:

Malonyl-CoA (214,222):
Acetyl-CoA + ATP + HCO_3 (=CO_2) $\Longleftrightarrow$ ADP + H + *Malonyl-CoA* + Pi

Dodecanoyl-CoA (215,217):
5 Malonyl-CoA + Acetyl-CoA + 15 H + 10 NADPH $\Longrightarrow$ 5 CO_2 + 5 CoA + 5 H_2O + 10 NADP + *Dodecanoyl-CoA*

3hmrs-CoA (R-3-hydroxy-myristoyl-CoA) (213):
Dodecanoyl-CoA + Malonyl-CoA + 2 H + NADPH $\Longrightarrow$ CO_2 + CoA + NADP + *3hmrs-CoA*

CMPKDO (CMP-3-deoxy-D-manno-octulosonate) (42,160,164,163):
PEP + Ribulose-5-phosphate + 2 H_2O + CTP $\Longrightarrow$ PPi + 2 Pi + *CMPKDO*

u23ga (UDP-2,3-bis(3-hydroxytetradecanoyl) glucosamine) (199,196,195):
UACGAM + 2 3hmrs-CoA + H_2O $\Longrightarrow$ Acetate + 2 CoA + H + *u23ga*

ADPHEP-L,D (ADP-L-glycero-D-manno-heptose) (29,193,174,173,172,170):
β-D-Fructose-6-phosphate + 2 ATP + Erythrose-4-P + H_2O $\Longrightarrow$ ADP + Glyceraldehyde-3-phosphate + PPi + Phosphate + *ADPHEP-L,D*

UDP-Glucose (208):
α-D-Glucose-6-phosphate + H + UTP $\Longrightarrow$ PPi + *UDPGlucose*

CDPETN (CDP-Ethanolamine) (rn:R02055,rn:R07376,rn:R01468,rn:R02038):
ATP + Serine + CTP $\Longleftrightarrow$ ADP + PPi + CO_2 + *CDPETN*

Lipopolysaccharides (206,184,194,161,176,177,185):
2 u23ga + ATP + H_2O + 4 CMPKDO + Dodecanoyl-CoA + Myristoyl-CoA + 3 ADPHEP-L,D + 2 CDPETN + 2 UDP-Glucose $\Longrightarrow$ UMP + 15 H + 3 UDP + 4 ADP + 4 CMP + 2 CDP + 2 CoA + *Lipopolysaccharides*

ACP is often used in the paper but this acyl-carrier-protein is just an intermediate step and therefore discarded. Instead the CoA-versions are used. This explains little discrepancies in the equations and the source.

The molecular weight of the lipopolysaccharides can be calculated as follows:

$$\begin{aligned} 2\,\text{23ga} + 4\,\text{KDO} + \text{Dodecanoyl} + \text{Myristoyl} + 3\,\text{ADPHEP-L,D} + \text{Ethanolamine} + 2\,\text{Glucose} &= 4931.74\ \text{g/mol} \\ &= 4.932\ \text{g/mmol} \end{aligned}$$

So 4.9 g of lipopolysaccharides corresponds to 1 mmol of lipopolysaccharides.

Therefore:

Metabolite	Molecular mass (in g/mol)
Glucosamine + Myristic + Dodecanoyl (=23ga)	607.86
KDO	238.19
Dodecanoyl	200.32
Myristoyl	228.38
ADPHEP-L,D	617.36
Ethanolamine	61.08
Glucose	180.16

Table 7: Molecular mass of lipopolysaccharides components

$$2\ \text{u23ga} + \text{ATP} + \text{H}_2\text{O} + 4\,\text{CMPKDO} + \text{Dodecanoyl-CoA} + \text{Myristoyl-CoA} + 3\,\text{ADPHEP-L,D} + 2\,\text{CDPETN} + 2\,\text{UDP-Glucose} \implies \text{UMP} + 15\,\text{H} + 3\,\text{UDP} + 34\,\text{ADP} + 4\,\text{CMP} + 2\,\text{CDP} + 2\,\text{CoA} + \mathbf{1\ mmol}\ \textit{lipopolysaccharides}$$

$$2\ \text{u23ga} + \text{ATP} + \text{H}_2\text{O} + 4\,\text{CMPKDO} + \text{Dodecanoyl-CoA} + \text{Myristoyl-CoA} + 3\,\text{ADPHEP-L,D} + 2\,\text{CDPETN} + 2\,\text{UDP-Glucose} \implies \text{UMP} + 15\,\text{H} + 3\,\text{UDP} + 34\,\text{ADP} + 4\,\text{CMP} + 2\,\text{CDP} + 2\,\text{CoA} + \mathbf{4.9\ g}\ \textit{lipopolysaccharides}$$

$$\tfrac{2}{4.9}\ \text{u23ga} + \tfrac{1}{4.9}\,\text{ATP} + \tfrac{1}{4.9}\,\text{H}_2\text{O} + \tfrac{4}{4.9}\,\text{CMPKDO} + \tfrac{1}{4.9}\,\text{Dodecanoyl-CoA} + \tfrac{1}{4.9}\,\text{Myristoyl-CoA} + \tfrac{3}{4.9}\,\text{ADPHEP-L,D} + \tfrac{2}{4.9}\,\text{CDPETN} + \tfrac{2}{4.9}\,\text{UDP-Glucose} - \tfrac{1}{4.9}\,\text{UMP} - \tfrac{15}{4.9}\,\text{H} - \tfrac{3}{4.9}\,\text{UDP} - \tfrac{34}{4.9}\,\text{ADP} - \tfrac{4}{4.9}\,\text{CMP} - \tfrac{2}{4.9}\,\text{CDP} - \tfrac{2}{4.9}\,\text{CoA} \implies \mathbf{1g}\ \textit{lipopolysaccharides}$$

B

MODEL EQUATIONS AND CONCENTRATION INTERVALS

B.1 BASIC MODEL

Glycolysis and Gluconeogenesis

Name	Formula
Glucoseimport	$Glucose_{out}$ ===> Glucose
rn:R00299	ATP + Glucose ===> α-D-Glucose-6-phosphate + ADP
rn:R04779	β-D-Fructose-6-phosphate + ATP ===> β-D-Fructose-1,6-bisphosphate + ADP
rn:R00764	β-D-Fructose-6-phosphate + PPi <===> β-D-Fructose-1,6-bisphosphate + Phosphate
rn:R02740	α-D-Glucose-6-phosphate <===> β-D-Fructose-6-phosphate
rn:R01070	β-D-Fructose-1,6-bisphosphate <===> Glyceraldehyde-3-phosphate + Glycerone-phosphate
rn:R01015	Glyceraldehyde-3-phosphate <===> Glycerone-phosphate
rn:R01061	NAD + Glyceraldehyde-3-phosphate + Phosphate <===> NADH + 1,3-Bisphospho-D-glycerate + H
rn:R01512	ATP + 3-Phospho-glycerate <===> ADP + 1,3-Bisphospho-D-glycerate
rn:R01518	2-Phospho-D-glycerate <===> 3-Phospho-glycerate
rn:R00658	2-Phospho-D-glycerate <===> PEP + H_2O
rn:R06620	Glucose + PQQ ===> D-glucono-1,5-lactone + PQQH2
Quinonereacm	PQQH2 + MQ <===> PQQ + MQH2
Quinonereac	PQQH2 + Q <===> PQQ + QH2
rn:R01519	D-glucono-1,5-lactone + H_2O ===> D-gluconate
rn:R01737	ATP + D-gluconate ===> ADP + 6-phospho-D-gluconate
rn:R02036	6-phospho-D-gluconate <===> KDPG + H_2O
rn:R05605	KDPG <===> Glyceraldehyde-3-phosphate + Pyruvate
rn:R00835	α-D-Glucose-6-phosphate + NADP ===> 6-phospho-D-glucono-lactone + NADPH
rn:R02035	6-phospho-D-glucono-lactone + H_2O ===> 6-phospho-D-gluconate

Table 8: Glycolysis and Gluconeogenesis

TCA cycle	
Name	**Formula**
rn:R00351	Oxaloacetate + Acetyl-CoA + H_2O ===> CoA + Citrate
rn:R01325	Citrate <===> cis-Aconitate + H_2O
rn:R01900	cis-Aconitate + H_2O <===> Isocitrate
rn:R00268	Oxalosuccinate + H_2O <===> 2-Oxoglutarate + CO_2
rn:R01899	Isocitrate + NADP <===> Oxalosuccinate + NADPH + H
rn:R08549	NAD + CoA + 2-Oxoglutarate + H_2O ===> NADH + Succinyl-CoA + CO_2 + H
rn:R00405	ADP + Succinyl-CoA + Phosphate <===> ATP + Succinate + CoA
rn:R00412	Succinate + Q <===> Fumarate + QH_2
rn:R00412m	Succinate + MQ <===> Fumarate + MQH_2
rn:R01082	Fumarate + H_2O <===> Malate
rn:R00342	NAD + Malate <===> Oxaloacetate + NADH + H
rn:R00341	Oxaloacetate + ATP + H_2O ===> PEP + ADP + CO_2
rn:R00479	Isocitrate <===> Glyoxylate + Succinate
rn:R00472	Glyoxylate + Acetyl-CoA + H_2O ===> Malate + CoA
rn:R00621	TPP + 2-Oxoglutarate + H_2O <===> 3-Carboxy1hydroxypropyl-ThPP + CO_2
Pyruvate metabolism	
rn:R00199	ATP + Pyruvate + H_2O ===> PEP + AMP + Phosphate
rn:R00200	PEP + ADP ===> ATP + Pyruvate
rn:R00209	NAD + CoA + Pyruvate + H_2O <===> NADH + Acetyl-CoA + CO_2 + H
OAD	Pyruvate + 2 Na_{out} + CO_2 ===> Oxaloacetate + 2 Na + H_2O
PEPC	PEP + CO_2 ===> Oxaloacetate + Phosphate
MAE	NADPH + Pyruvate + CO_2 + H <===> NADP + Malate + H_2O
rn:R00710	NAD + Ethanal + H_2O <===> Acetate + NADH + H
rn:R00316	Acetate + ATP <===> Acetyladenylate + PPi
rn:R00236	CoA + Acetyladenylate <===> Acetyl-CoA + AMP
rn:R00703	NAD + L-Lactate <===> NADH + Pyruvate + H
Ammonia assimilation	
GDH	NADH + 2-Oxoglutarate + NH_3 + H <===> NAD + Glutamate + H_2O
rn:R00258	2-Oxoglutarate + L-alanine <===> Pyruvate + Glutamate
Ala-DH	NAD + L-alanine + H_2O <===> NADH + Pyruvate + NH_3

Table 9: TCA cycle, pyruvate metabolism and ammonia assimilation

Glycine, serine and threonine metabolism	
Name	**Formula**
rn:R01513	NAD + 3-Phosphoglycerate <===> 3-Phosphohydroxypyruvate + NADH + H
rn:R04173	Dexfosfoserine + 2-Oxoglutarate <===> 3-Phosphohydroxypyruvate + Glutamate
rn:R00582	Dexfosfoserine + H_2O ===> Serine + Phosphate
rn:R00220	Serine ===> Pyruvate + NH_3
rn:R00945	Glycine + 5,10-Methylene-THF + H_2O <===> Tetrahydrofolate + Serine
rn:R00751	L-Threonine <===> Glycine + Ethanal
rn:R01466	O-Phospho-L-homoserine + H_2O <===> L-Threonine + Phosphate
rn:R01771	L-Homoserine + ATP ===> ADP + O-Phospho-L-Homoserine
rn:R01773	L-Homoserine + NAD <===> NADH + L-Aspartic-4-semialdehyde + H
rn:R06977	Glutamate + L-Aspartic-4-semialdehyde <===> L-2,4-Diaminobutyrate + 2-Oxoglutarate
rn:R02291	L-Aspartic-4-semialdehyde + NADP + Phosphate <===> 4-Phospho-L-aspartate + NADPH + H
rn:R00480	L-Aspartate + ATP ===> ADP + 4-Phospho-L-aspartate
rn:R00355	L-Aspartate + 2-Oxoglutarate <===> Oxaloacetate + Glutamate

Table 10: Glycine, Serine and Threonine metabolism

Respiratory chain	
Name	**Formula**
C1	NADH + Q + 4 H ===> NAD + QH_2 + 4 H_{out}
C1m	NADH + MQ + 4 H ===> NAD + MQH_2 + 4 H_{out}
$C1^{Na}$	NADH + 2 Na + Q + 2 H ===> NAD + 2 Na_{out} + QH_2 + 2 H_{out}
$C1^{Na}$m	NADH + 2 Na + MQ + 2 H ===> NAD + 2 Na_{out} + MQH_2 + 2 H_{out}
C2	$FADH_2$ + Q ===> FAD + QH_2
C2m	$FADH_2$ + MQ ===> FAD + MQH_2
C3-4	QH_2 + 0.5 O_2 + 4 H ===> Q + 6 H_{out} + H_2O
C3-4m	MQH_2 + 0.5 O_2 + 4 H ===> MQ + 6 H_{out} + H_2O
ATPase	ADP + 3 H_{out} + Phosphate ===> ATP + H_2O + 3 H
Napump32	2 Na + 3 H_{out} ===> 2 Na_{out} + 3 H
Napump21	Na + 2 H_{out} ===> Na_{out} + 2 H

Table 11: Respiratory chain

Ectoine metabolism	
DoeA	Ectoine + H_2O <===> N-α-Acetyldiaminobutyrate
DoeB	N-α-Acetyldiaminobutyrate + H_2O ===> Acetate + L-2,4-Diaminobutyrate
rn:R06978	L-2,4-Diaminobutyrate + Acetyl-CoA ===> CoA + N-γ-Acetyl-diaminobutyrate
rn:R06979	N-γ-Acetyl-diaminobutyrate ===> Ectoine + H_2O
Additional reactions	
ATP load	ATP + H_2O ===> ADP + Phosphate
rn:R00127	ATP + AMP <===> 2 ADP
rn:R00112	NADH + H_{out} + NADP <===> NAD + NADPH + H
leak	H_{out} ===> H

Table 12: Ectoine metabolism and additional reactions

B.2 INCLUDING BIOMASS INTO THE MODEL

Additional reactions needed for the synthesis of biomass precursors	
Name	**Formula**
rn:R00014	Pyruvate + TPP <===> 2-α-Hydroxyethy-TPP + CO_2
rn:R00149	2 ATP + NH_3 + CO_2 + H_2O <===> 2 ADP + Carbamoylphosphate + Phosphate
rn:R01528	6-phospho-D-gluconate + NADP <===> NADPH + Ribulose-5-phosphate + CO_2 + H
rn:R01056	Ribulose-5-phosphate <===> Ribose-5-phosphate
rn:R01049	ATP + Ribose-5-phosphate <===> AMP + PRPP
rn:R01067	β-D-Fructose-6-phosphate + 3-Phospho-glycerate <===> Xylulose-5-phosphate + Erythrose-4-P
rn:R01529	Ribulose-5-phosphate <===> Xylulose-5-phosphate
rn:R01655	5,10-Methylene-THF + H_2O <===> 10-Formyl-THF + H
rn:R02016	NADP + Thioredoxin <===> NADPH + Thioredoxindisulfide + H
rn:R00188	AMP + Phosphate <===> PAP + H_2O
H2Ssynthesis	2 ATP + 3 NADPH + Thioredoxin + 3 H + SO_4 <===> ADP + 3 NADP + H_2S + Thioredoxindisulfide + PAP + PPi + 3 H_2O

AICAR-synthesisreac	5 ATP + PRPP + 10-Formyl-THF + CO_2 + 2 H_2O + 2 Glutamine + L-Aspartate + Glycine ===> Tetrahydrofolate + 5 ADP + Fumarate + AICAR + PPi + 5 Phosphate + 2 Glutamate
Chorismate-synthesisreac	2 PEP + ATP + NADH + Erythrose-4-P + H ===> NAD + ADP + Chorismate + 4 Phosphate
IMPreac	10-Formyl-THF + AICAR ===> Tetrahydrofolate + IMP + H_2O
SAdenosylreac	ATP + S-Adenosyl-L-homocysteine + 5-Methyl-THF + 2 H_2O <===> Tetrahydrofolate + S-Adenosyl-L-methionine + Adenosine + PPi + Phosphate
Adenosinereac	ADP + AMP <===> ATP + Adenosine
5_MethylTHFreac	5,10-Methylene-THF + NADPH + H <===> NADP + 5-Methyl-THF
Glycerolreac	ATP + NADH + Glycerol + H ===> NAD + ADP + Glycerone-phosphate
UDPGAreac	β-D-Fructose-6-phosphate + Acetyl-CoA + Glutamine + UTP ===> CoA + UDPGA + PPi + Glutamate
UDPAMreac	PEP + NADH + UDPGA + H ===> NAD + UDPAM + Phosphate
ASPSAreac	ATP + NADPH + H + L-Aspartate ===> ADP + NADP + ASPSA + Phosphate
DAPreac	Succinyl-CoA + Pyruvate + NADPH + ASPSA + H + Glutamate ===> Succinate + CoA + 2-Oxoglutarate + NADP + DAP
MalonylCoAreac	ATP + Acetyl-CoA + CO_2 ===> ADP + Malonyl-CoA + Phosphate + H
DodecanoylCoAreac	Acetyl-CoA + 10 NADPH + 5 Malonyl-CoA + 15 H ===> 5 CoA + 10 NADP + Dodecanoyl-CoA + 5 CO_2 + 5 H_2O
3hmrsCoAreac	NADPH + Malonyl-CoA + Dodecanoyl-CoA + 2 H ===> CoA + NADP + 3-hmrs-CoA + CO_2
CMPKDOreac	PEP + Ribulose-5-phosphate + 2 H_2O + CTP ===> CMPKDO + PPi + 2 Phosphate
u23gareac	UDPGA + 2 3-hmrs-CoA + H_2O ===> Acetate + 2 CoA + u23ga + H
ADPHEPreac	β-D-Fructose-6-phosphate + 2 ATP + Erythrose-4-P + H_2O ===> ADP + Glyceraldehyde-3-phosphate + ADPHEP + PPi + Phosphate
UDPglucosereac	α-D-Glucose-6-phosphate + H + UTP ===> UDP-Glucose + PPi
Ethanolaminereac	ATP + Serine + CTP <===> ADP + CDP-Ethanolamine + PPi + CO_2

Table 13: Additional reactions needed for the synthesis of biomass precursors.

Nucleotide synthesis	
Name	**Formula**
AMPreac	ATP + IMP + L-Aspartate ===> ADP + AMP + Fumarate + Phosphate
GMPreac	NAD + ATP + IMP + 2 H_2O + Glutamine ===> NADH + AMP + GMP + PPi + H + Glutamate
UMPreac	Oxaloacetate + 5 ATP + NADPH + Ribose-5-phosphate + 2 NH_3 + H ===> 5 ADP + NADP + UMP + H_2O
CMPreac	Oxaloacetate + 7 ATP + NADPH + Ribose-5-phosphate + 3 NH_3 + H ===> 7 ADP + NADP + CMP + H_2O
dAMPreac	3 NAD + 9 ATP + 5,10-Methylene-THF + 3-Phospho-glycerate + 2 NADPH + Ribose-5-phosphate + 5 NH_3 + CO_2 + H_2O ===> Tetrahydrofolate + 3 NADH + 9 ADP + 2 NADP + dAMP + H
dGMPreac	3 NAD + 11 ATP + 5,10-Methylene-THF + 3-Phospho-glycerate + NADPH + Ribose-5-phosphate + 5 NH_3 + CO_2 + 2 H_2O ===> Tetrahydrofolate + 3 NADH + 11 ADP + NADP + dGMP + 2 H
dTMPreac	Oxaloacetate + 5 ATP + 5,10-Methylene-THF + 3 NADPH + Ribose-5-phosphate + 2 NH_3 + 3 H ===> Tetrahydrofolate + 5 ADP + 3 NADP + dTMP + 3 H_2O
dCMPreac	Oxaloacetate + 7 ATP + 2 NADPH + Ribose-5-phosphate + 3 NH_3 + 2 H ===> 7 ADP + 2 NADP + dCMP + 2 H_2O
Additional amino acids	
Arginineреac	2 ATP + NADH + AcetylCoA + NADPH + Carbamoylphosphate + NH_3 + 2 H + L-Aspartate + Glutamate ===> Acetate + NAD + ADP + CoA + AMP + Fumarate + NADP + PPi + 2 Phosphate + Arginine
Asparaginereac	ATP + NH_3 + L-Aspartate ===> AMP + PPi + Asparagine
Cysteinereac	Acetyl-CoA + H_2S + Serine ===> Acetate + CoA + Cysteine
Glutaminereac	ATP + NH_3 + Glutamate ===> ADP + Phosphate + Glutamine
Histidinereac	2 NAD + ATP + PRPP + 3 H_2O + Glutamine ===> 2 NADH + 2-Oxoglutarate + AICAR + 2 PPi + Phosphate + 2 H + Histidine
Isoleucinereac	NADPH + 2-α-Hydroxyethy-TPP + H + Glutamate + L-Threonine ===> TPP + 2-Oxoglutarate + NADP + NH_3 + H_2O + Isoleucine
Leucinereac	NAD + 2 Pyruvate + Acetyl-CoA + NADPH + Glutamate ===> NADH + CoA + 2-Oxoglutarate + NADP + 2 CO_2 + Leucine
Lysinereac	ATP + NADH + Succinyl-CoA + Pyruvate + NADPH + 2 H + L-Aspartate + Glutamate ===> NAD + Succinate + ADP + CoA + 2-Oxoglutarate + NADP + CO_2 + Phosphate + Lysine

Methioninereac	ATP + NADH + 5,10-Methylene-THF + Succinyl-CoA + NADPH + H_2O + 2 H + Cysteine + L-Aspartate ===> Tetrahydrofolate + NAD + Succinate + ADP + CoA + NADP + Phosphate + Methionine + Serine
Phenylalaninereac	Chorismate + Glutamate ===> 2-Oxoglutarate + CO_2 + H_2O + Phenylalanine
Prolinereac	ATP + NADH + NADPH + 2 H + Glutamate ===> NAD + ADP + NADP + H_2O + Phosphate + Proline
Tryptophanreac	PRPP + Chorismate + Glutamine + Serine ===> Glyceraldehyde -3-phosphate + Pyruvate + PPi + CO_2 + 2 H_2O + Tryptophan + Glutamate
Tyrosinereac	NAD + Chorismate + Glutamate ===> NADH + 2-Oxoglutarate + CO_2 + H + Tyrosine
Valinereac	2 Pyruvate + NADPH + H + Glutamate ===> 2-Oxoglutarate + NADP + CO_2 + H_2O + Valine
Aminoacidreaction	0.06 Arginine + 0.05 Asparagine + 0.02 Cysteine + 0.05 Glutamine + 0.02 Histidine + 0.05 Isoleucine + 0.08 Leucine + 0.06 Lysine + 0.03 Methionine + 0.03 Phenylalanine + 0.04 Proline + 0.01 Tryptophan + 0.03 Tyrosine + 0.08 Valine + 0.1 L-alanine + 0.05 L-Aspartate + 0.05 Glycine + 0.05 Glutamate + 0.04 Serine + 0.05 L-Threonine ===> Aminoacids

Lipid synthesis

Name	Formula
AMPreac	ATP + IMP + L-Aspartate ===> ADP + AMP + Fumarate + Phosphate
Glycerol-3-phosphatereac	NADH + Glyceraldehyde-3-phosphate + H ===> NAD + Glycerol-3-phosphate
Palmitoyl-CoAreac	7 ATP + 8 Acetyl-CoA + 14 NADPH + 14 H ===> 7 ADP + 7 CoA + 14 NADP + Palmitoyl-CoA + 14 H_2O
Palmitoleoyl-CoAreac	7 ATP + 8 Acetyl-CoA + 14 NADPH + 14 H ===> 7 ADP + 7 CoA + 14 NADP + Palmitoleoyl-CoA + 14 H_2O
Oleoyl-CoAreac	NAD + 8 ATP + 9 Acetyl-CoA + 16 NADPH + 15 H ===> NADH + 8 ADP + 8 CoA + 16 NADP + Oleoyl-CoA + 15 H_2O
MyristoylCoAreac	6 ATP + 7 Acetyl-CoA + 12 NADPH + 12 H ===> 6 ADP + 6 CoA + 12 NADP + Myristoyl-CoA + 12 H_2O
MyristoleoylCoAreac	6 ATP + 7 Acetyl-CoA + 12 NADPH + 12 H ===> 6 ADP + 6 CoA + 12 NADP + Myristoleoyl-CoA + 12 H_2O
HeptadecanoylCoAreac	S-Adenosyl-L-methionine + Palmitoleoyl-CoA ===> S-Adenosyl-L-homocysteine + Heptadecanoyl-CoA

cis-VaccenoylCoAreac	8 ATP + 9 Acetyl-CoA + 16 NADPH + 16 H ===> 8 ADP + 8 CoA + 16 NADP + cis-Vaccenoyl-CoA + 16 H_2O
NonadecenoylCoAreac	SAdenosylLmethionine + Oleoyl-CoA ===> S-Adenosyl-L-homocysteine + Nonadecenoyl-CoA
Acyl-CoAreaction	0.39 Palmitoyl-CoA + 0.11 Palmitoleoyl-CoA + 0.17 Oleoyl-CoA + 0.03 Myristoyl-CoA + 0.09 Myristoleoyl-CoA + 0.16 Heptadecanoyl-CoA + 0.01 cis-Vaccenoyl-CoA + 0.01 Nonadecenoyl-CoA ===> Acyl-CoA
PEreaction	9 ATP + 3 Glycerol-3-phosphate + 6 Acyl-CoA + 3 Serine ===> 6 CoA + 9 AMP + PE + 9 PPi + 3 CO2
PGreaction	8 ATP + 5 Glycerol-3-phosphate + 5 Acyl-CoA + 3 H2O ===> CoA + 8 AMP + PG + 8 PPi
CLreaction	8 ATP + 5 Glycerol-3-phosphate + 5 Acyl-CoA + 3 H2O ===> 5 CoA + 8 AMP + 1 Glycerol + CL + 8 PPi

Biomass synthesis

Name	**Formula**
Proteinreac	39.1 ATP + 9.1 Aminoacids ===> 39.1 ADP + Protein + 39.1 Phosphate
DNAreac	11 ATP + 0.8 dAMP + 0.8 dGMP + 0.8 dTMP + 0.8 dCMP ===> 11 ADP + DNA + 11 Phosphate
RNAreac	7 ATP + 0.8 AMP + 0.9 GMP + 0.8 UMP + 0.6 CMP ===> 7 ADP + RNA + 7 Phosphate
Lipidreac	0.4 PE + 0.5 PG + 0.1 CL ===> Lipids
Glycogenreac	6 ATP + 6 Glucose ===> 6 ADP + Carbohydrates + 6 Phosphate
Peptidoglycanreac	5 ATP + 1 UDPGA + 1 UDPAM + 0.6 DAP + 0.6 Leucine + 2 L-alanine + 0.3 Gly + 1 Glutamate ===> 5 ADP + Peptidoglycan + 5 Phosphate + 2 UDP
LPSreac	0.2 ATP + 0.2 Dodecanoyl-CoA + 0.8 CMPKDO + 0.4 u23ga + 0.6 ADPHEP + 0.4 UDPGlucose + 0.4 CDPEthanolamine + 0.2 Myristoyl-CoA + 0.2 H2O ===> 0.8 ADP + 0.4 CoA + Lipopolysaccharides + 0.2 UMP + 0.8 CMP + 3 H + 0.6 UDP + 0.4 CDP
Growth	0.03 DNA + 0.2 RNA + 0.6 Protein + 0.1 Lipids + 0.03 Carbohydrates + 0.03 Peptidoglycan + 0.03 Lipopolysaccharides ===> Biomass

Table 14: Nucleotide synthesis, additional amino acid reactions, lipid biosynthesis and biomass synthesis.

B.3 ABBREVIATIONS WITH RESPECT TO METABOLITES

Abbreviations

Abbreviation	Metabolite	Abbreviation	Metabolite
3hmrs-CoA	R-3-hydroxy-myristoyl-CoA	γ-N-acABA	Nγ-Acetyl-diaminobutyrate
3PG	3-Phospho-glycerate	GAP	Glyceraldehyde-3-phosphate
4PAsp	4-Phospho-L-aspartate	Glclactone	D-glucono-1,5-lactone
5-Methyl-THF	5-Methyltetrahydrofolate	Glu	Glutamate
6PGlclactone	6-phospho-D-gluconolactone	Gluconate	D-gluconate
6PGluconate	6-phospho-D-gluconate	Gly	Glycine
AcCoA	AcetylCoA	Glycerol-3-P	Glycerol-3-phosphate
ADPHEP	ADP-mannoheptose	HomoSer	L-Homoserine
aKG	2-Oxoglutarate	IMP	Inosinic acid
Ala	L-Alanine	IsoCIT	Isocitrate
α-N-AcABA	N-α-Acetyldiaminobutyrate	KDPG	2-Keto-3-deoxy-6-phosphogluconate
ASA	L-Aspartic-4-semialdehyde	LPS	Lipposaccharid
Asp	L-Aspartate	MAL	Malate
ASPSA	L-Aspartate 4-semialdehyde	MeTHF	5,10-Methylene-THF
CDPEthanolamine	CDP-Ethanolamine	MQ	Menaquinone
cisACO	cisAconitate	MQ_2	Menaquinol
CIT	Citrate	OAA	Oxaloacetate
CL	Cardiolypin	OHPyr	3-Phosphonooxy-pyruvate
CMPKDO	CMP-2-Keto-3-deoxyoctanoate	PA	Phosphatidyl acid
DABA	L-2,4-Diaminobutyrate	PAP	Phosphoadenosine phosphate
DAP	L,L-2,6-Diaminopimelate	PE	Phosphatidyl ethanolamine
DHAP	Glycerone-phosphate	PEP	Phosphoenol-pyruvate
F6P	β-D-Fructose-6-phosphate	PG	Phosphatidyl glycerol
FBP	β-D-Fructose-1,6-bisphosphate	P-homoSer	O-Phospho-L-homoserine
FUM	Fumarate	PhosphoSer	Dexfosfoserine
G6P	α-D-glucose-6-phosphate	PQQ	Pyrroloquinoline quinone

$PQQH_2$	Reduced pyrrolo-quinoline quinone	SucCoA	Succinyl-CoA
PRPP	Phosphoribosyl pyrophosphate	THF	Tetrahydrofolate
Pyr	Pyruvate	Thr	Threonine
Q	Quinone	TPP	Thiamine diphosphate
QH_2	Quinol	u23ga	UDP-2,3-bis(3-hydroxy-tetradecanoyl) glucosmine
Ribose-5-P	Ribose-5-phosphate	UDPAM	UDP-Acetylmuramate
Ribulose-5-P	Ribulose-5-Phosphate	UDPGA	UDP-N-Acetylglucosamine
Ser	Serine	UDPGlucose	UDP-glucose
SUC	Succinate		

Table 15: Abbreviations

B.4 CONCENTRATION INTERVALS

Boundaries for metabolite concentrations		
Metabolite	**Lower bound**	**Upper bound**
CO_2 [68]	0.0001	0.0001
Glucose	10^{-7}	0.02
$Glucose_{out}$	0.03	0.03
Glutamate	10^{-7}	0.1
H_2O	1	1
L-Aspartic-4-semialdehyde	$6\ 10^{-8}$	0.015
Na [100, 174]	0.04	0.25
Na_{out}	[0.1; 1; 2]	[0.1; 1; 2]
NH_3 [68]	0.019	0.019
O_2 [41, 68]	$8.2\ 10^{-6}$	$6.5\ 10^{-4}$
Oxalosuccinate	$6\ 10^{-8}$	0.015
Phosphate [29, 47]	0.05	0.2
PPi	0.001	0.1

Table 16: Special bounds for metabolite concentrations (in mol/l). The general limits are $[10^{-7}; 0.015]$ mol/l

C

FIGURES BELONGING TO CHAPTER 8

Several figures have been produced with respect to Section 8.5 and Section 8.6, which can't be displayed in the main text due to a shortage of space but might nevertheless be interesting to the reader. These figures are all together displayed here.

C.1 ECTOINE- VS. ENERGYOPTIMIZATION

The first figures concern the ectoine and energy optimization with respect to the model, in which C1 is inactive but $C1^{Na}$ is active. The flux distributions are displayed in Figure 70. The resulting thermodynamically feasible ranges are displayed in Figure 71.

Figure 72 compares the flux distributions gained via ectoine optimization in case of an ATP load of 10 mmol or 600 mmol ATP/100 mmol Glc. The thermodynamic comparison can be found in Figure 73.

Figure 74 displays the flux distributions with either C1 or $C1^{Na}$ active and GDH inactivate in the FBA. Therefore only Ala-DH can be used for ammonia assimilation.

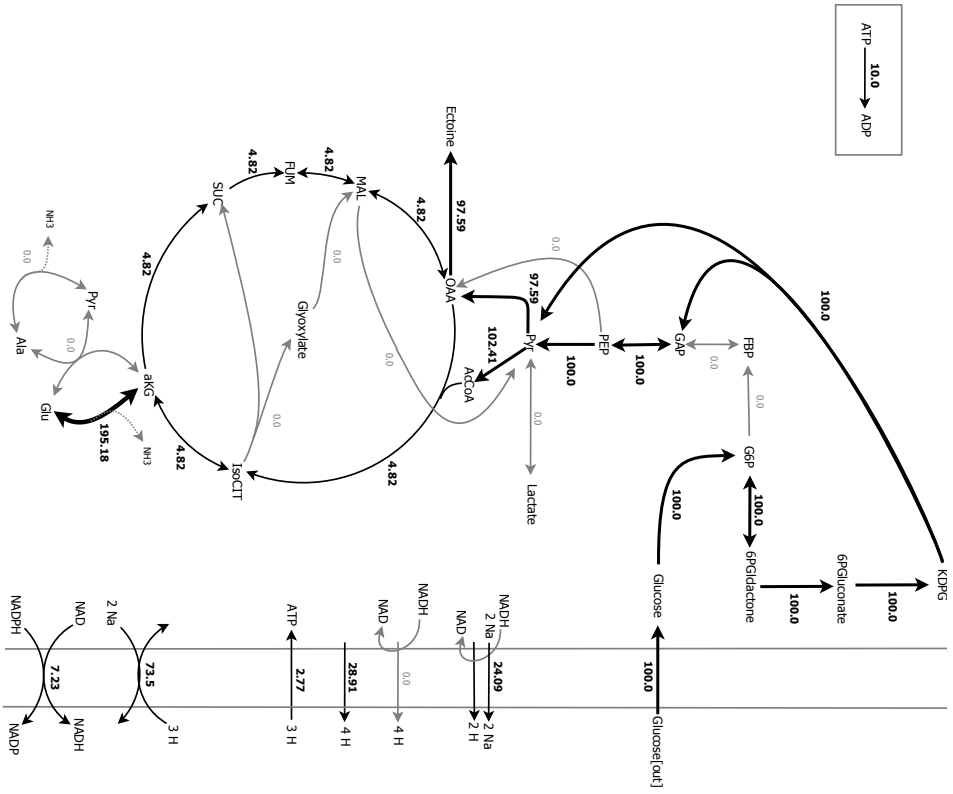
The comparison of Figure 72 has been repeated in Figure 75. This time without the use of the GDH. The thermodynamic comparison can be found in Figure 76.

C.2 ANAPLEROTIC REACTIONS

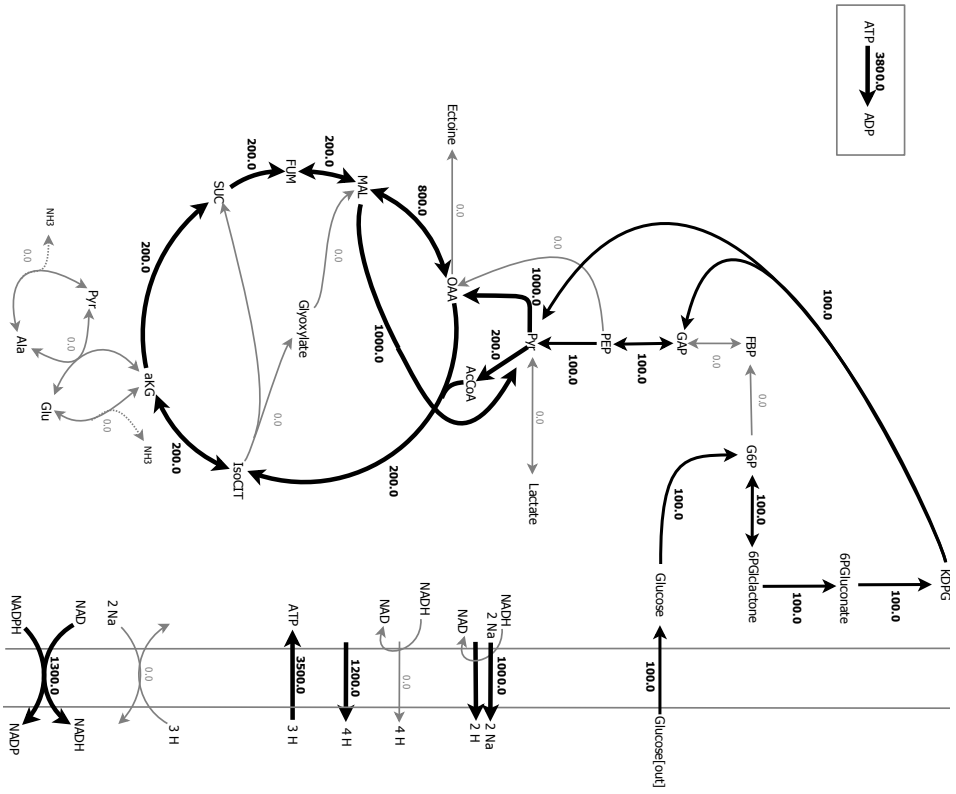
Figure 77 displays the flux distribution of the PEPC OAD mutant in case that only $C1^{Na}$ or both version of complex I are active in the FBA. The thermodynamic analysis is displayed in Figure 78.

In case that the GDH is inactive, the following flux distibutions can be gained with $C1^{Na}$ active or both active: see Figure 79. The thermodynamic analysis is displayed in Figure 80

Figure 81 displays the outcoming in case of an inactivation of OAD, PEPC and GDH.



(a)



(b)

Figure 70: Comparison of the ectoine- (a) and energy-fluxes (b) for an ATP load of 10 mmol ATP/100 mmol Glc, with only $C1^{Na}$ active in the FBA. The ectoine production is increased in comparison with Figure 53, while the energy production is decreased.

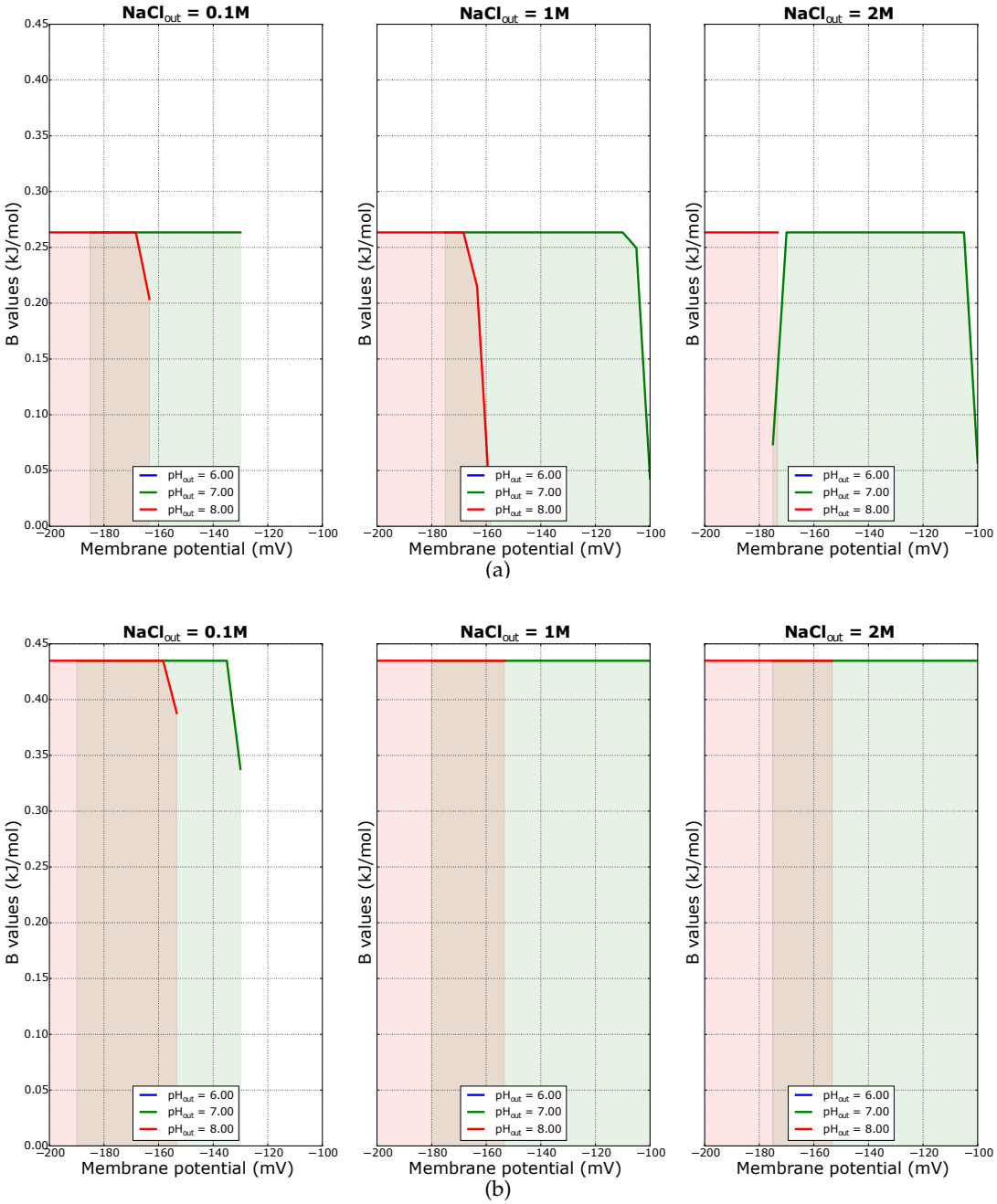
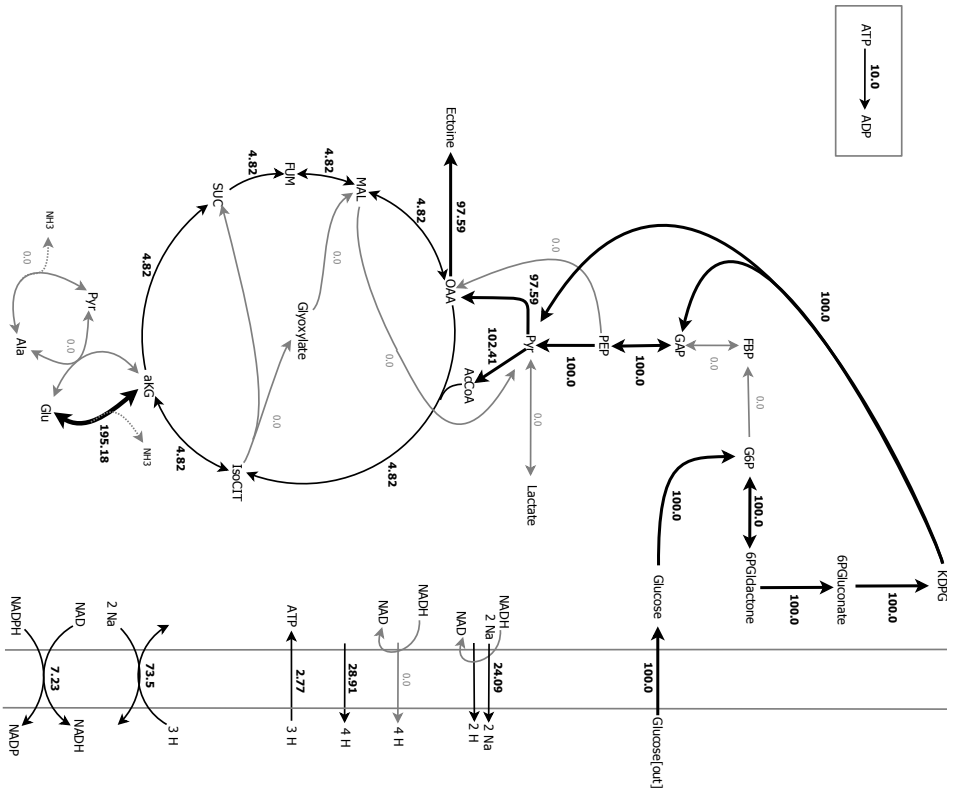


Figure 71: Comparison of the MDF solutions for ectoine (a) and energy optimization (b) with only $C1^{Na}$ active in the flux distributions.



(a)

(b)

Figure 72: (a) shows the flux distribution of the ectoine optimization for an ATP load of 10 mmol ATP/100 mmol Glc, as in Figure 70 (a), where only $C1^{Na}$ is active. (b) displays the flux distribution for an ATP load of 600 mmol ATP/100 mmol Glc, where both enzymes are active simultaneously. The ectoine yield is lower in the second case, because carbon is needed for ATP production.

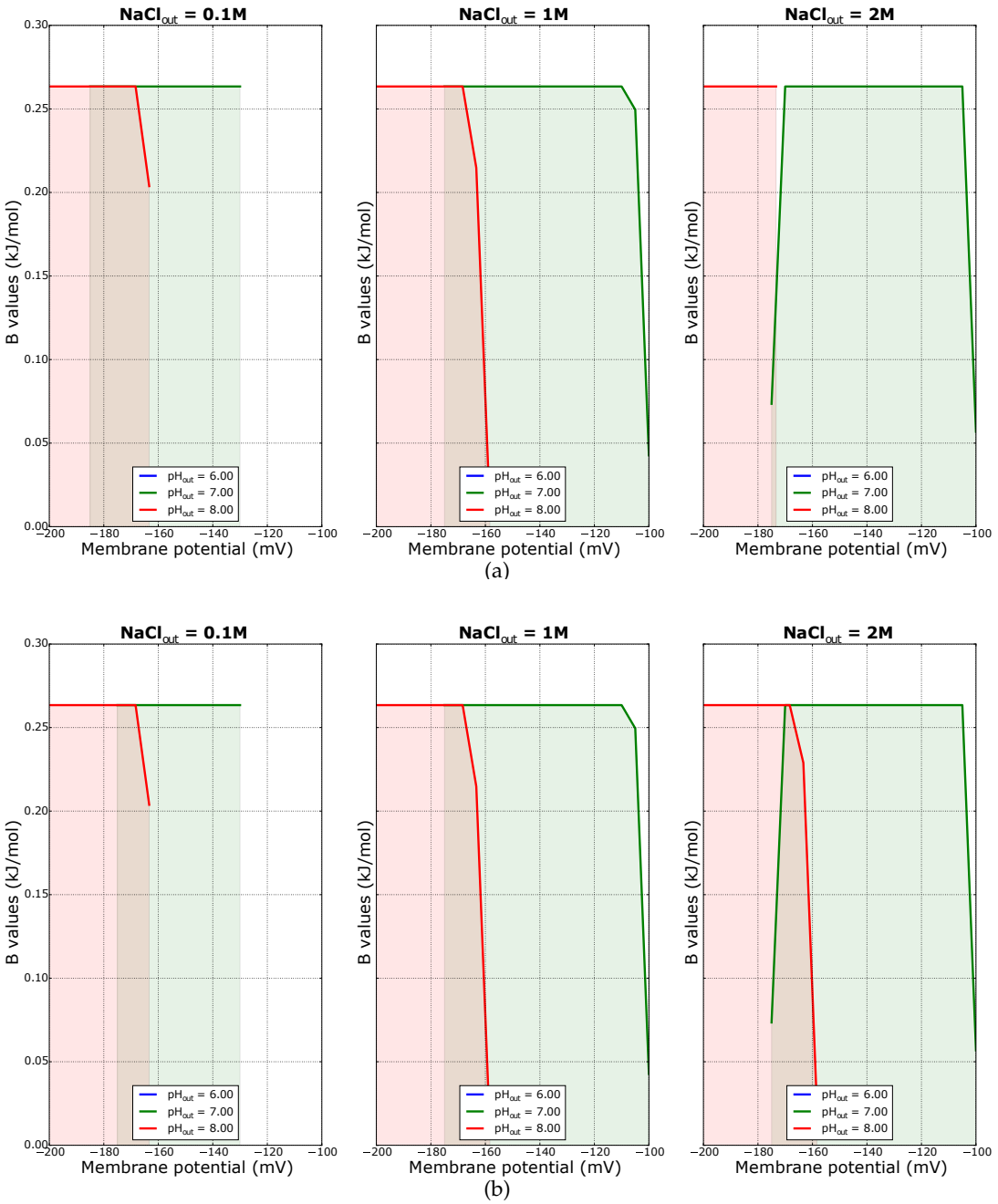
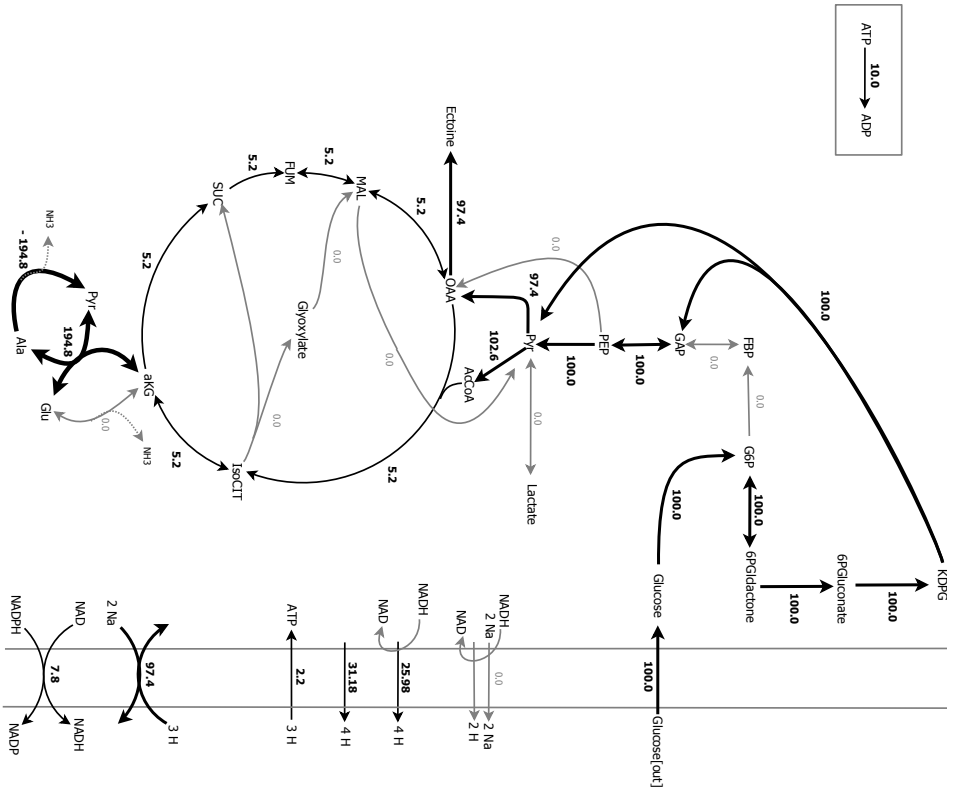
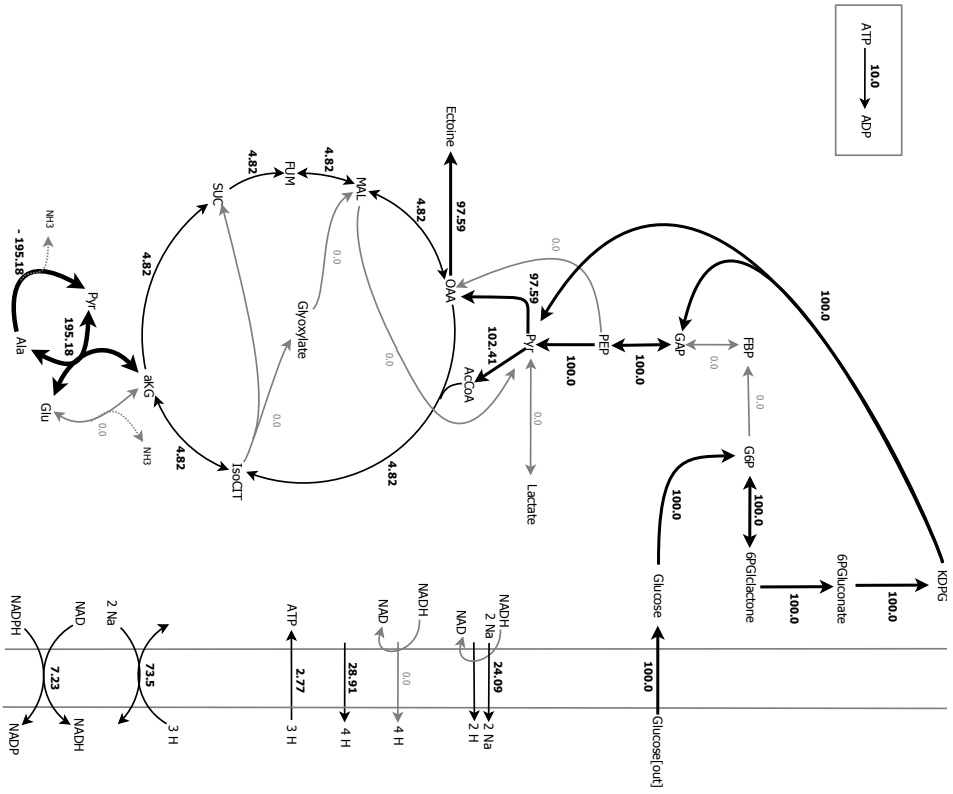


Figure 73: (a) shows the feasible ranges for the ectoine optimization under the condition that $C1^{Na}$ is used, which matches the first row of Figure 71. (b) displays the feasible ranges in case that both enzymes are active at once in the flux distribution.

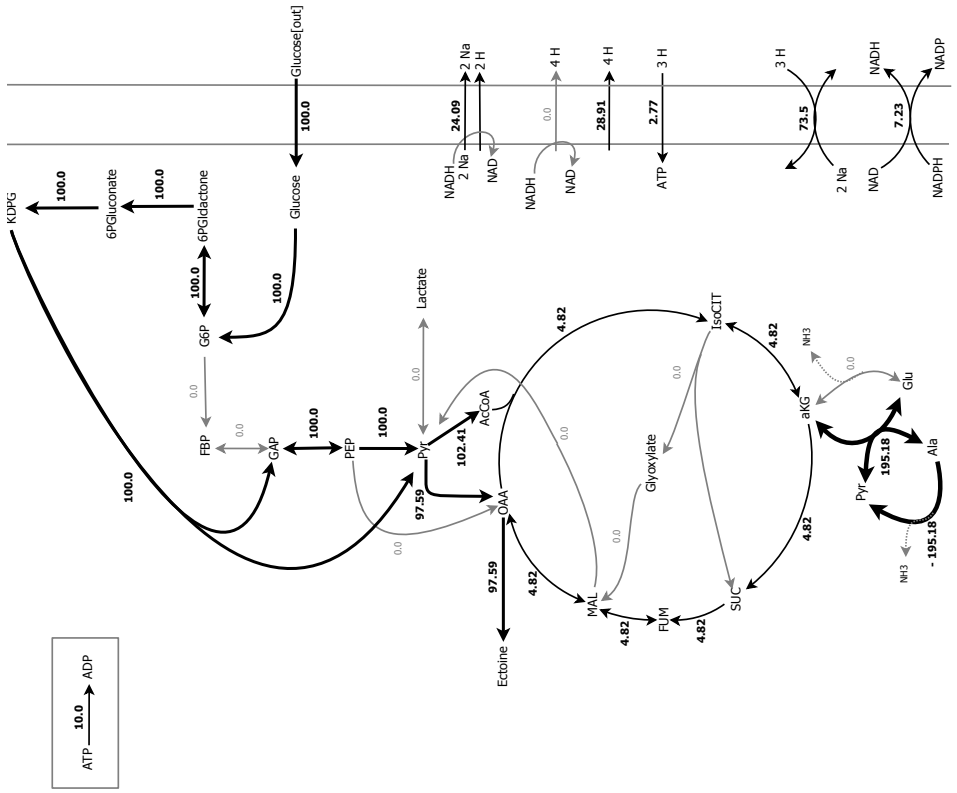


(a)

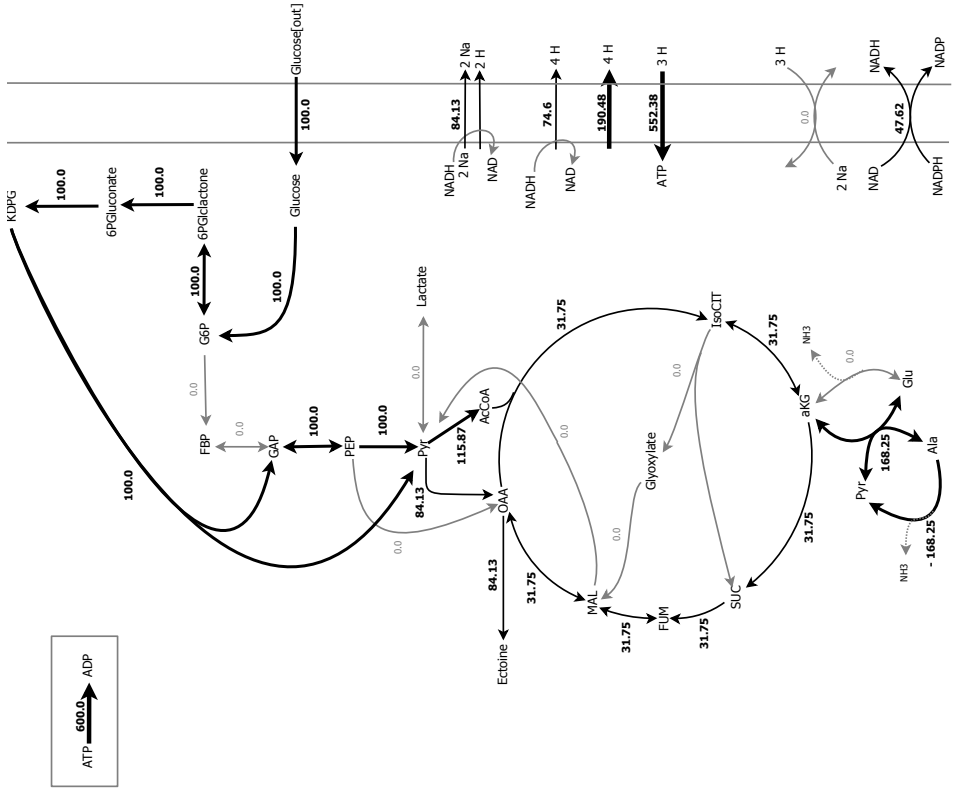


(b)

Figure 74: Comparison of the ectoine maximizations for an ATP load of 10 mmol ATP/100 mmol Glc, with either C1 (a) or C1^{Nq} (b) active and with the GDH shut off.



(a)



(b)

Figure 75: (a) shows the flux distribution of the ectoine optimization for an ATP load of 10 mmol ATP/100 mmol Glc, where only $C1^{Na}$ is active and GDH is inactive. (b) displays the flux distribution for an ATP load of 600 mmol ATP/100 mmol Glc, where both enzymes are active simultaneously and the GDH still inactive.

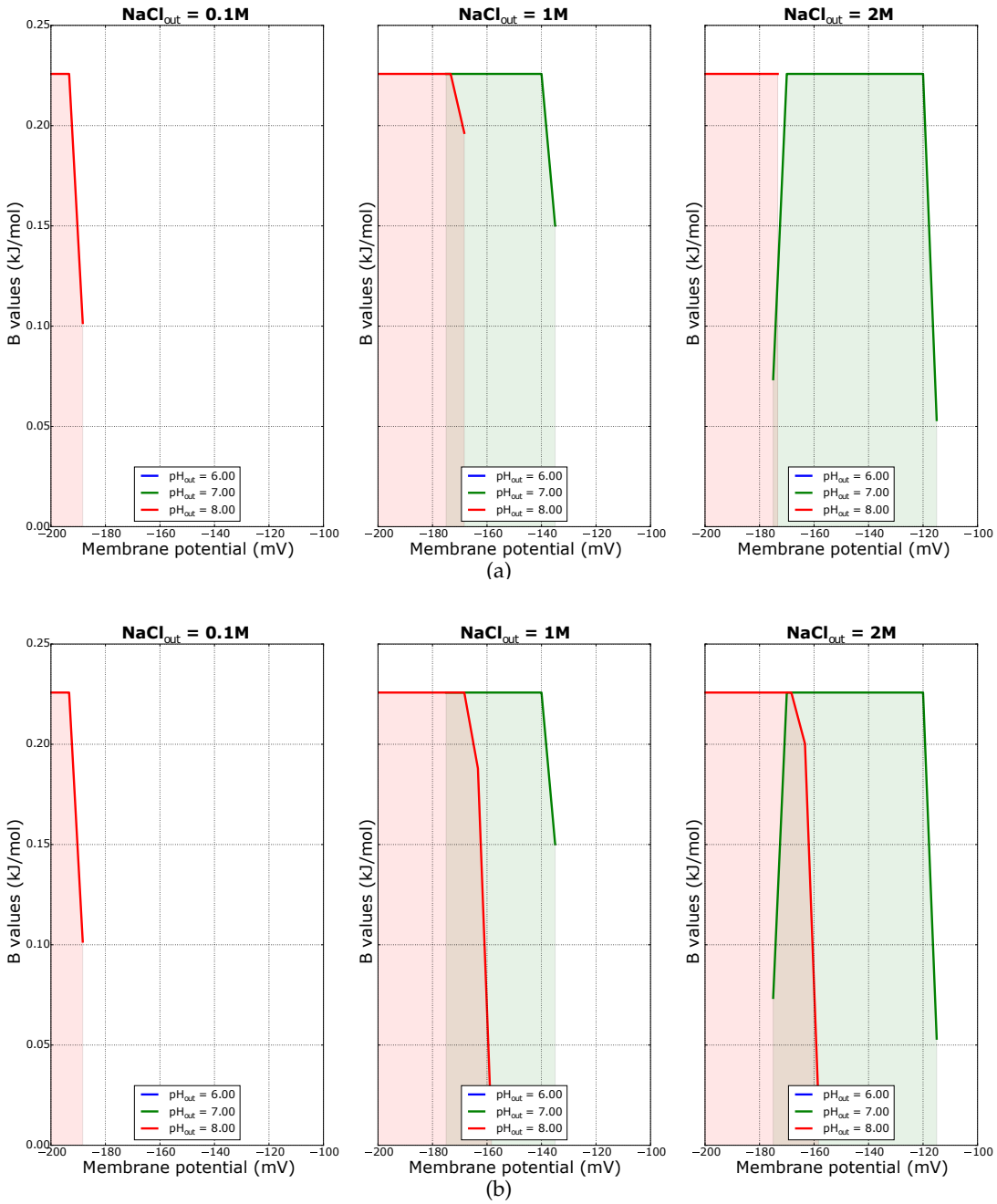
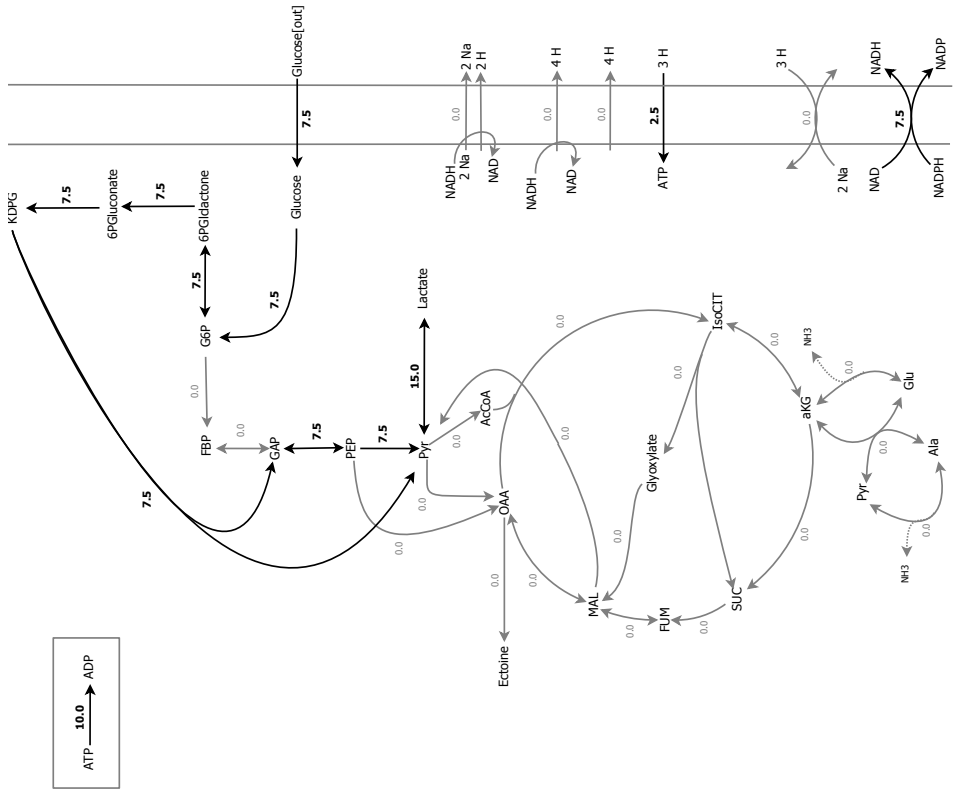
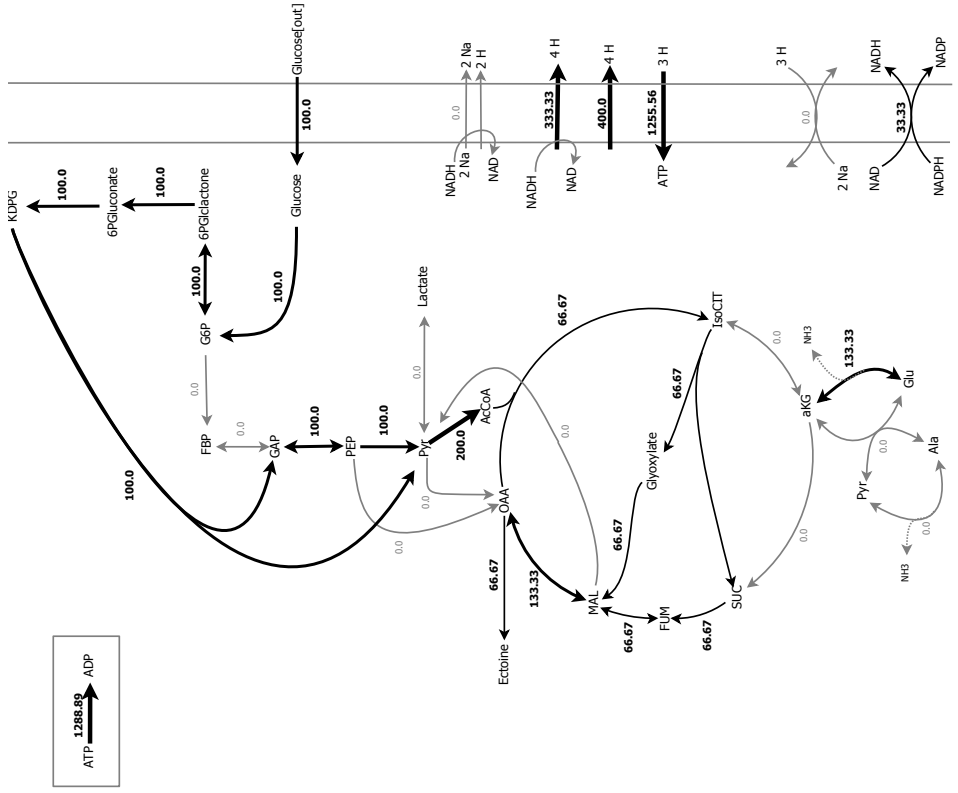


Figure 76: (a) shows the MDF solutions for the ectoine optimizing flux distribution under the condition that C1^{Na} is used and GDH inactive. (b) displays the solutions in case that both enzymes are active at once but the GDH still inactive.



(a)



(b)

Figure 77: (a) shows the flux distribution for the PEPC OAD mutant in case that $C1^{Na}$ is active in the FBA but C1 not for an ATP load of 10 mmol ATP/100 mmol Glc. However, $C1^{Na}$ does not exhibit a flux (b) displays the flux distribution in case that both complex I versions are active but only C1 exhibits a flux.

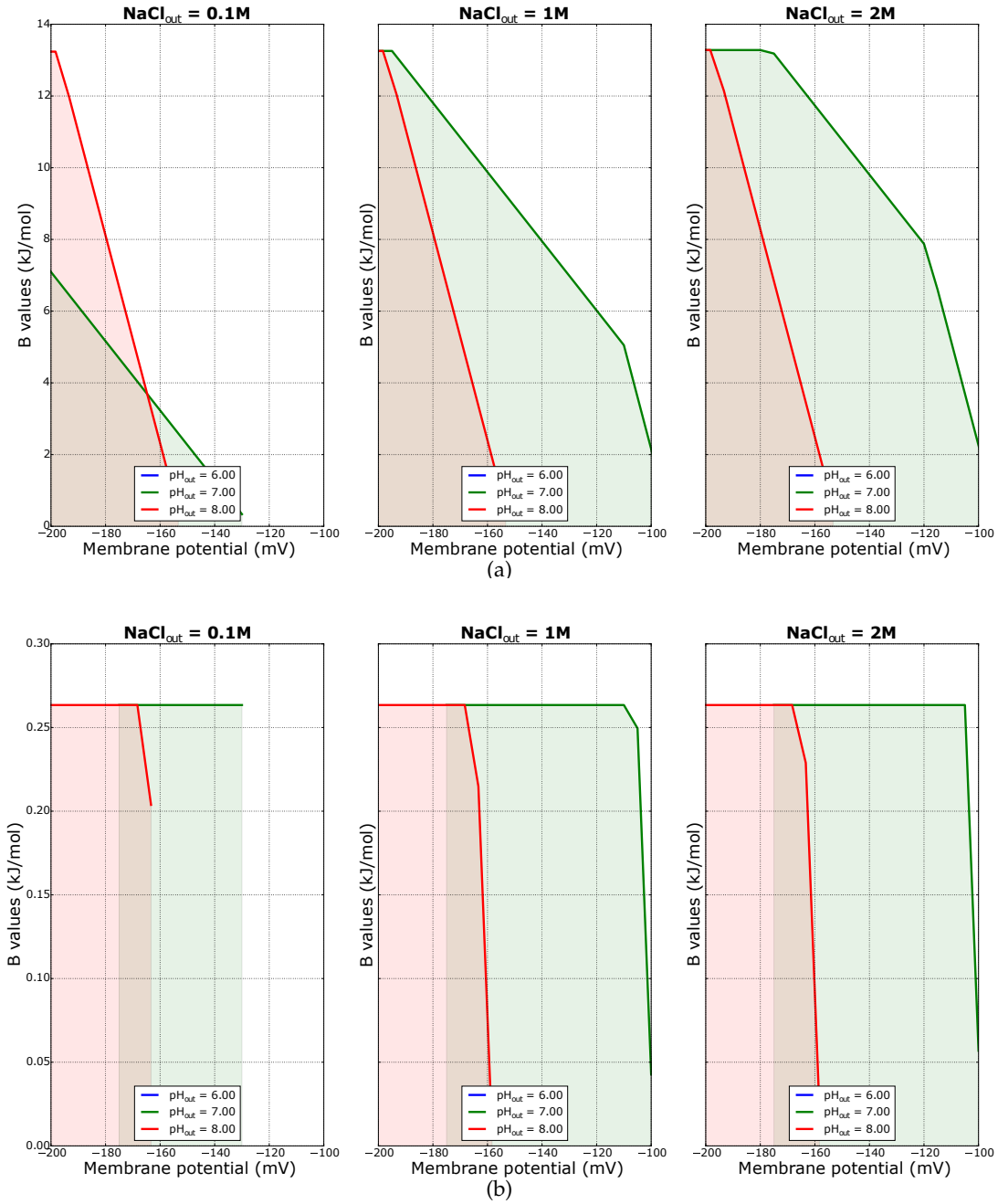
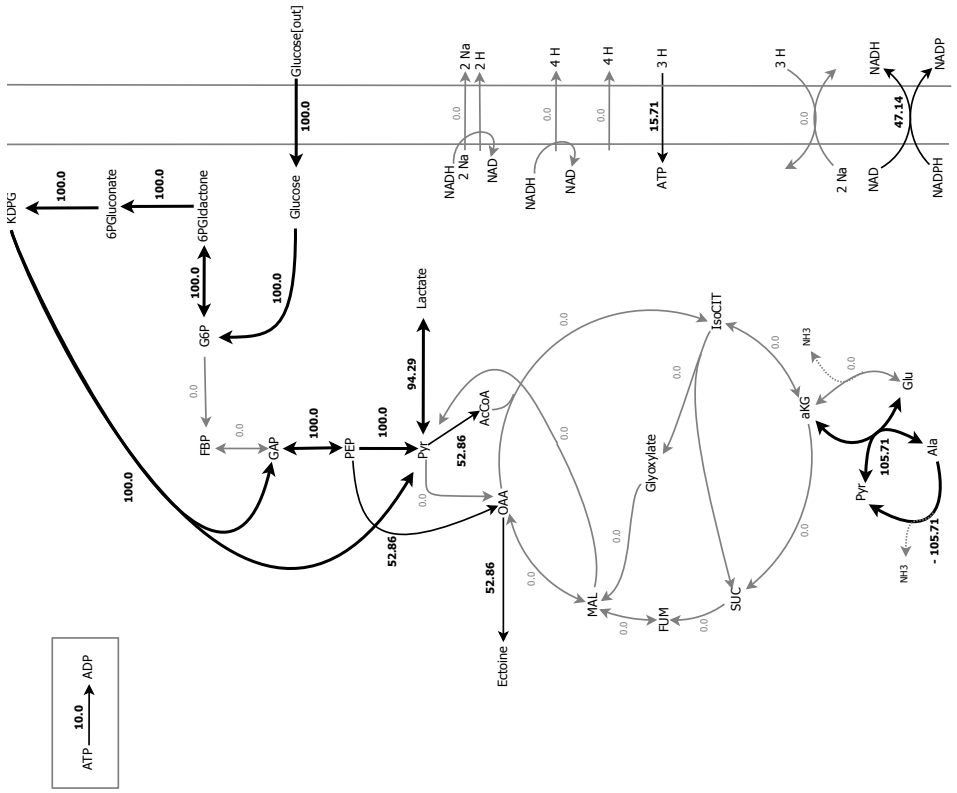
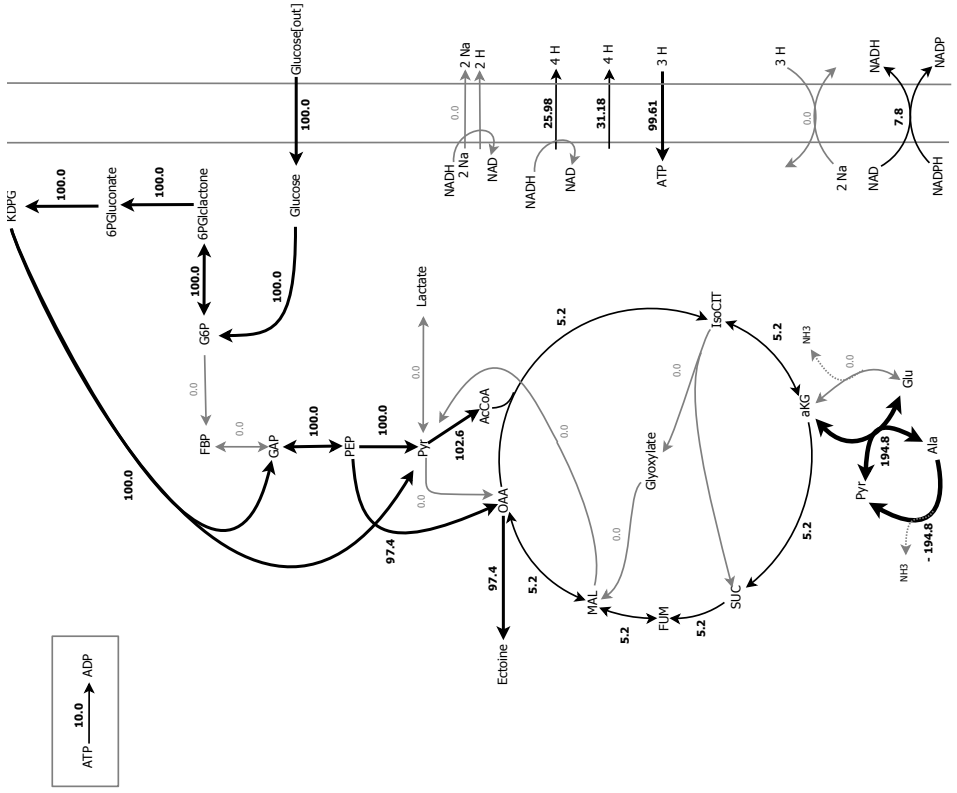


Figure 78: (a) shows the conditions in case of anaerobic respiration. (b) displays the conditions in case that both enzymes were active in the FBA, but only C1 exhibits a flux.



(a)



(b)

Figure 79: (a) shows the flux distribution of the OAD mutant for an ATP load of 10 mmol ATP/100 mmol Glc in case that $C1^{Na}$ is active but C1 not. (b) displays the case that both are active during FBA but in both cases, $C1^{Na}$ shows no flux. The glutamate dehydogenase is inactive.

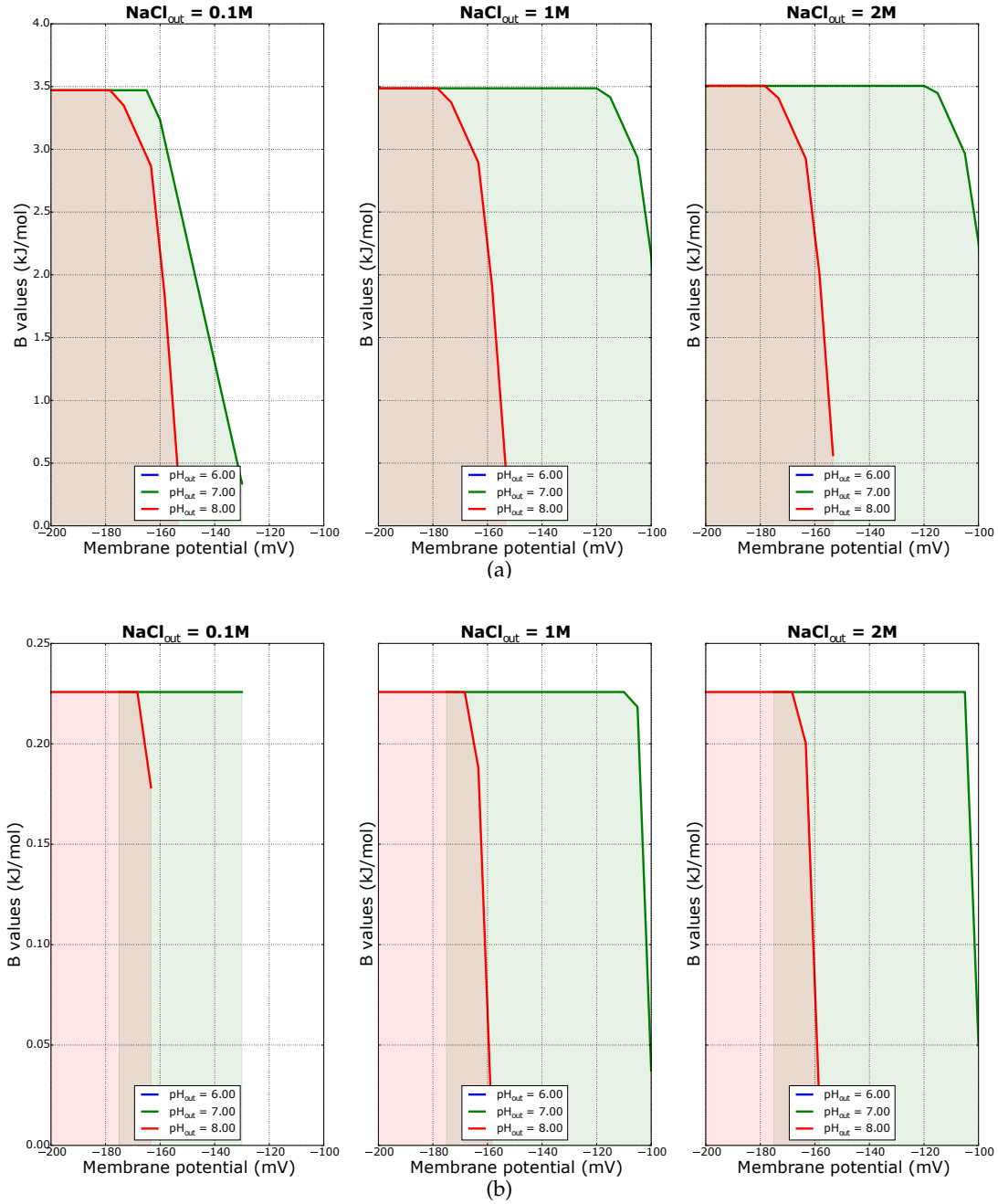
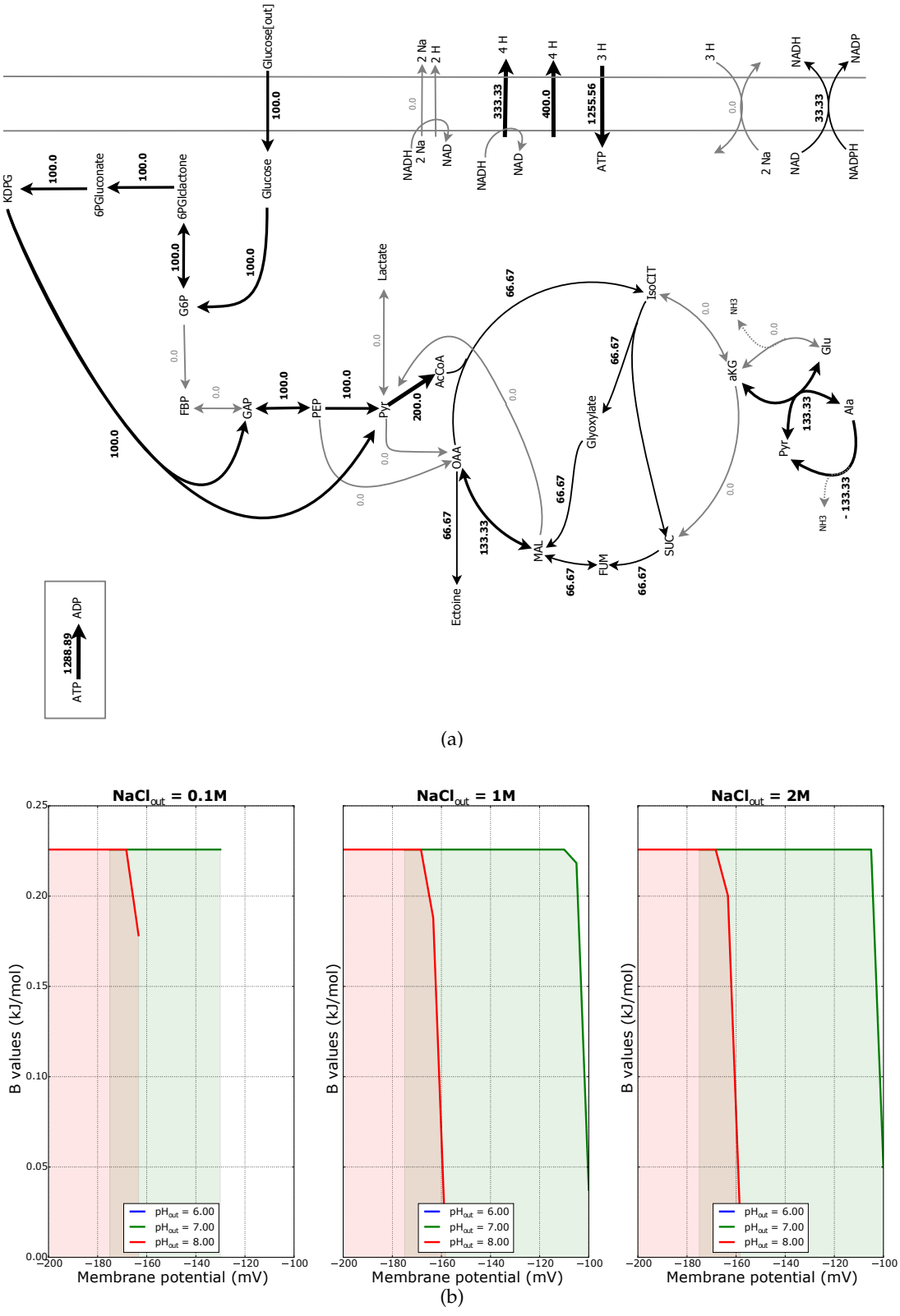


Figure 80: (a) displays the feasible ranges of the OAD mutant for an ATP load of 10 mmol ATP/100 mmol Glc in case that C1^{Na} is active but C1 not, during FBA. The second shows the opposite. In both cases, the glutamate dehydogenase is inactive. The anaerobic respiration causes a higher exergonism.



(a)

(b)

Figure 81: (a) displays the flux distribution of the PEPC OAD mutant in case that only C1 is active for an ATP load of 10 mmol ATP/100 mmol Glc, while the glutamate dehydrogenase is inactive; (b) shows the minimal driving force solution for the displayed flux distribution.

BIBLIOGRAPHY

[1] Robert A Alberty. *Thermodynamics of biochemical reactions*. John Wiley & Sons, 2005.

[2] Robert A Alberty, Athel Cornish-Bowden, Robert N Goldberg, Gordon G Hammes, Keith Tipton, and Hans V Westerhoff. "Recommendations for terminology and databases for biochemical thermodynamics." In: *Biophysical chemistry* 155.2 (2011), pp. 89–103.

[3] Bert Allard. *Modelling in aquatic chemistry*. Nuclear Energy Agency, 1997.

[4] Adrián F Alvarez, Claudia Rodriguez, and Dimitris Georgellis. "Ubiquinone and menaquinone electron carriers represent the yin and yang in the redox regulation of the ArcB sensor kinase." In: *Journal of bacteriology* 195.13 (2013), pp. 3054–3061.

[5] Rui Alves and Michael A Savageau. "Comparing systemic properties of ensembles of biological networks by graphical and statistical methods." In: *Bioinformatics* 16.6 (2000), pp. 527–533.

[6] Rui Alves and Michael A Savageau. "Effect of overall feedback inhibition in unbranched biosynthetic pathways." In: *Biophysical journal* 79.5 (2000), pp. 2290–2304.

[7] Rui Alves and Michael A Savageau. "Extending the method of mathematically controlled comparison to include numerical comparisons." In: *Bioinformatics* 16.9 (2000), pp. 786–798.

[8] Rui Alves and Michael A Savageau. "Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways." In: *Bioinformatics* 16.6 (2000), pp. 534–547.

[9] Rui Alves and Michael A Savageau. "Irreversibility in unbranched pathways: preferred positions based on regulatory considerations." In: *Biophysical journal* 80.3 (2001), pp. 1174–1185.

[10] Stefano Andreozzi, Ljubisa Miskovic, and Vassily Hatzimanikatis. "iSCHRUNK–In Silico Approach to Characterization and Reduction of Uncertainty in the Kinetic Models of Genome-scale Metabolic Networks." In: *Metabolic engineering* 33 (2016), pp. 158–168.

[11] David R Arahal, M Teresa García, Carmen Vargas, David Cánovas, Joaquin J Nieto, and Antonio Ventosa. "Chromohalobacter salexigens sp. nov., a moderately halophilic species that includes Halomonas elongata DSM 3043 and ATCC 33174." In: *International journal of systematic and evolutionary microbiology* 51.4 (2001), pp. 1457–1462.

[12] Meric Ataman and Vassily Hatzimanikatis. "Heading in the right direction: thermodynamics-based network analysis and pathway engineering." In: *Current Opinion in Biotechnology* 36 (2015), pp. 176–182.

[13] Julio R Banga and Eva Balsa-Canto. "Parameter estimation and optimal experimental design." In: *Essays in biochemistry* 45 (2008), pp. 195–210.

[14] Arren Bar-Even, Elad Noor, Yonatan Savir, Wolfram Liebermeister, Dan Davidi, Dan S Tawfik, and Ron Milo. "The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters." In: *Biochemistry* 50.21 (2011), pp. 4402–4410.

[15] Daniel A Beard, Eric Babson, Edward Curtis, and Hong Qian. "Thermodynamic constraints for biochemical networks." In: *Journal of theoretical biology* 228.3 (2004), pp. 327–333.

[16] Daniel A Beard, Shou-dan Liang, and Hong Qian. "Energy balance for analysis of complex metabolic networks." In: *Biophysical journal* 83.1 (2002), pp. 79–86.

[17] Daniel A Beard and Hong Qian. "Thermodynamic-based computational profiling of cellular regulatory control in hepatocyte metabolism." In: *American Journal of Physiology-Endocrinology and Metabolism* 288.3 (2005), E633–E644.

[18] Daniel A Beard and Hong Qian. "Relationship between thermodynamic driving force and one-way fluxes in reversible processes." In: *PloS one* 2.1 (2007), e144.

[19] Daniel A Beard and Hong Qian. *Chemical biophysics: quantitative analysis of cellular systems*. Cambridge University Press, 2008.

[20] Bryson D Bennett, Elizabeth H Kimball, Melissa Gao, Robin Osterhout, Stephen J Van Dien, and Joshua D Rabinowitz. "Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli." In: *Nature chemical biology* 5.8 (2009), pp. 593–599.

[21] Yulia V Bertsova and Alexander V Bogachev. "The origin of the sodium-dependent NADH oxidation by the respiratory chain of Klebsiella pneumoniae." In: *FEBS letters* 563.1 (2004), pp. 207–212.

[22] Nuno Borges, Ana Ramos, Neil D Raven, Richard J Sharp, and Helena Santos. "Comparative study of the thermostabilizing properties of mannosylglycerate and other compatible solutes on model enzymes." In: *Extremophiles* 6.3 (2002), pp. 209–216.

[23] ALFREDO F BRAÑA, NANCY PAIVA, and ARNOLD L DEMAIN. "Pathways and regulation of ammonium assimilation in Streptomyces clavuligerus." In: *Microbiology* 132.5 (1986), pp. 1305–1317.

[24] JN Brønsted. "Studies on solubility. IV. The principle of the specific interaction of ions." In: *Journal of the American Chemical Society* 44.5 (1922), pp. 877–898.

[25] AD Brown. "Microbial water stress." In: *Bacteriological Reviews* 40.4 (1976), p. 803.

[26] Elisabetta Buommino, Chiara Schiraldi, Adone Baroni, Iole Paoletti, Monica Lamberti, Mario De Rosa, and Maria Antonietta Tufano. "Ectoine from halophilic microorganisms induces the expression of hsp70 and hsp70B' in human keratinocytes modulating the proinflammatory response." In: *Cell stress & chaperones* 10.3 (2005), p. 197.

[27] Jonathan J Burbaum, Ronald T Raines, W John Albery, and Jeremy R Knowles. "Evolutionary optimization of the catalytic effectiveness of an enzyme." In: *Biochemistry* 28.24 (1989), pp. 9293–9305.

[28] Alfred Richard Burkin. *Chemical hydrometallurgy: Theory and principles*. Vol. 1. World Scientific, 2001.

[29] JHB Christian and Judith A Waltho. "Solute concentrations within cells of halophilic and non-halophilic bacteria." In: *Biochimica et biophysica acta* 65.3 (1962), pp. 506–508.

[30] Bruce L Clarke. "Stoichiometric network analysis." In: *Cell biophysics* 12.1 (1988), pp. 237–253.

[31] UniProt Consortium et al. "UniProt: a hub for protein information." In: *Nucleic acids research* (2014), gku989.

[32] Tyrrell Conway. "The Entner-Doudoroff pathway: history, physiology and molecular biology." In: *FEMS Microbiology Letters* 103.1 (1992), pp. 1–28.

[33] Athel Cornish-Bowden. *Fundamentals of enzyme kinetics*. John Wiley & Sons, 2013.

[34] GB Danzig. *Linear programming and extensions. 1963*.

[35] David Hildebrand. "Plant Biochemistry Lecture 24." `http://www.uky.edu/~dhild/biochem/24/lect24.html`. 2016.

[36] P. Debye and E. Hückel. "The theory of electrolytes. (1) Lowering of freezing point and related phenomena." In: *Physik. Z.* 24 (1923), pp. 185–206.

[37] P. Debye and E. Hückel. "The theory of electrolytes. (2) The limiting law of electrical conductivity." In: *Ibid.* 24 (1923), pp. 305–325.

[38] P. Debye and E. Hückel. "The theory of electrolytes. (3) Osmotic equation of state and the activity of strong electrolytes in dilute solutions." In: *Ibid.* 25 (1924), pp. 97–107.

[39] Patrick P Dennis and Lawrence C Shimmin. "Evolutionary divergence and salinity-mediated selection in halophilic archaea." In: *Microbiology and Molecular Biology Reviews* 61.1 (1997), pp. 90–104.

[40] Abdelmoneim Amer Desouki, Florian Jarre, Gabriel Gelius-Dietrich, and Martin J Lercher. "CycleFreeFlux: efficient removal of thermodynamically infeasible loops from flux distributions." In: *Bioinformatics* (2015), btv096.

[41] Pauline M Doran. *Bioprocess engineering principles*. Academic press, 1995.

[42] A. Dötsch, J. Severin, A. Wolfgang, E. A. Galinski, and J.-U. Kreft. "A mathematical model for growth and osmoregulation in halophilic bacteria." In: *Microbiology* (2008), pp. 2956–2969.

[43] Jeremy S Edwards and Bernhard O Palsson. "Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions." In: *BMC bioinformatics* 1.1 (2000), p. 1.

[44] Henryk Eisenberg and Ellen J Wachtel. "Structural studies of halophilic proteins, ribosomes, and organelles of bacteria adapted to extreme salt concentrations." In: *Annual review of biophysics and biophysical chemistry* 16.1 (1987), pp. 69–92.

[45] MP Elizalde and JL Aparicio. "Current theories in the calculation of activity coefficients—II. Specific interaction theories applied to some equilibria studies in solution chemistry." In: *Talanta* 42.3 (1995), pp. 395–400.

[46] Nuno Empadinhas and Milton da Costa. "Osmoadaptation mechanisms in prokaryotes: distribution of compatible solutes." In: *International Microbiology* 11.3 (2008), pp. 151–161.

[47] Kjell Magne Fagerbakke, Svein Norland, and Mikal Heldal. "The inorganic ion content of native aquatic bacteria." In: *Canadian journal of microbiology* 45.4 (1999), pp. 304–311.

[48] C Fallet, P Rohe, and E Franco-Lara. "Process optimization of the integrated synthesis and secretion of ectoine and hydroxyectoine under hyper/hypo-osmotic stress." In: *Biotechnology and bioengineering* 107.1 (2010), pp. 124–133.

[49] Rick A Fasani and Michael A Savageau. "Automated construction and analysis of the design space for biochemical systems." In: *Bioinformatics* 26.20 (2010), pp. 2601–2609.

[50] Adam M Feist, Markus J Herrgård, Ines Thiele, Jennie L Reed, and Bernhard Ø Palsson. "Reconstruction of biochemical networks in microorganisms." In: *Nature Reviews Microbiology* 7.2 (2009), pp. 129–143.

[51] Ciarán P Fisher, Andrzej M Kierzek, Nick J Plant, and J Bernadette Moore. "Systems biology approaches for studying the pathogenesis of non-alcoholic fatty liver disease." In: *World journal of gastroenterology* 20.41 (2014), pp. 15070–15078.

[52] Avi Flamholz, Elad Noor, Arren Bar-Even, Wolfram Liebermeister, and Ron Milo. "Glycolytic strategy as a tradeoff between energy yield and protein cost." In: *Proceedings of the National Academy of Sciences* 110.24 (2013), pp. 10039–10044.

[53] Ronan MT Fleming, Ines Thiele, and HP Nasheuer. "Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to Escherichia coli." In: *Biophysical chemistry* 145.2 (2009), pp. 47–56.

[54] Naoko Fujimoto, Tomoyuki Kosaka, and Mamoru Yamada. "Menaquinone as Well as Ubiquinone as a Crucial Component in the Escherichia coli Respiratory Chain." In: *Chem Biol* (2012), pp. 187–208.

[55] Kentaro Furusho, Toshihiro Yoshizawa, and Shinichi Shoji. "Ectoine alters subcellular localization of inclusions and reduces apoptotic cell death induced by the truncated Machado–Joseph disease gene product with an expanded polyglutamine stretch." In: *Neurobiology of disease* 20.1 (2005), pp. 170–178.

[56] Erwin A Galinski, Heinz-Peter PFEIFFER, and Hans G TRÜPER. "1, 4, 5, 6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid." In: *European Journal of Biochemistry* 149.1 (1985), pp. 135–139.

[57] Ziomara P Gerdtzen, Prodromos Daoutidis, and Wei-Shou Hu. "Non-linear reduction for kinetic models of metabolic reaction networks." In: *Metabolic engineering* 6.2 (2004), pp. 140–154.

[58] Katrin Grammann, Angela Volke, and Hans Jörg Kunte. "New type of osmoregulated solute transporter identified in halophilic members of the Bacteria domain: TRAP transporter TeaABC mediates uptake of ectoine and hydroxyectoine in Halomonas elongata DSM 2581T." In: *Journal of bacteriology* 184.11 (2002), pp. 3078–3085.

[59] Grenthe, I. and Mompean, F. and Spahiu, K. and Wanner, H. "Guidelines for the extrapolation to zero ionic strength." https://www.oecd-nea.org/dbtdb/guidelines/tdb2.pdf. 2013.

[60] Ingmar Grenthe and A Plyasunov. "On the use of semiempirical electrolyte theories for modeling of solution chemical data." In: *Pure and applied chemistry* 69.5 (1997), pp. 951–958.

[61] Ingmar Grenthe, Hans Wanner, Isabelle Forest, et al. "Chemical thermodynamics of uranium." In: (1992).

[62] Edward Armand Guggenheim. *Thermodynamics: An Advanced Treatment for Chemists and Physicist*. Vol. 2. North-Holland, 1949.

[63] FRANCETTE Hamaide, DONN J Kushner, and G DENNIS Sprott. "Proton motive force and Na+/H+ antiport in a moderate halophile." In: *Journal of bacteriology* 156.2 (1983), pp. 537–544.

[64] HS Harned and BB Owen. "The Physical Chemistry of Electrolytic Solutions, Reinhold Pub." In: *Corp., New York* 354 (1958).

[65] Wally C van Heeswijk, Hans V Westerhoff, and Fred C Boogerd. "Nitrogen assimilation in Escherichia coli: putting molecular data into a systems perspective." In: *Microbiology and Molecular Biology Reviews* 77.4 (2013), pp. 628–695.

[66] R. Heinrich and T.A. Rapoport. "A linear steady-state treatment of enzymatic chains." In: *Eur. J. Biochem.* 42 (1974), pp. 89–95.

[67] Victor Henri. *Lois générales de l'action des diastases*. Librairie Scientifique A. Hermann, 1903.

[68] Christopher S Henry, Linda J Broadbelt, and Vassily Hatzimanikatis. "Thermodynamics-based metabolic flux analysis." In: *Biophysical journal* 92.5 (2007), pp. 1792–1805.

[69] William J Heuett and Hong Qian. "Combining flux and energy balance analysis to model large-scale biochemical networks." In: *Journal of bioinformatics and computational biology* 4.06 (2006), pp. 1227–1243.

[70] Hermann-Georg Holzhütter. "The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks." In: *European Journal of Biochemistry* 271.14 (2004), pp. 2905–2922.

[71] Wolfgang Hummel, Urs Berner, Enzo Curti, FJ Pearson, and Tres Thoenen. "Nagra/PSI chemical thermodynamic data base 01/01." In: *Radiochimica Acta* 90.9-11/2002 (2002), pp. 805–813.

[72] Matthew D Jankowski, Christopher S Henry, Linda J Broadbelt, and Vassily Hatzimanikatis. "Group contribution method for thermodynamic analysis of complex metabolic networks." In: *Biophysical journal* 95.3 (2008), pp. 1487–1499.

[73] John Stockie. "What is mathematical modeling." http://people.math.sfu.ca/~stockie/research/cfdgroup/models.pdf.

[74] Stefan J Jol, Anne Kümmel, Vassily Hatzimanikatis, Daniel A Beard, and Matthias Heinemann. "Thermodynamic calculations for biochemical transport and reaction processes in metabolic networks." In: *Biophysical journal* 99.10 (2010), pp. 3139–3144.

[75] H. Kacser and J.A. Burns. "The control of flux." In: *Symp. Soc. Exp. Biol.* 27 (1973), pp. 65–104.

[76] Tuty Asmawaty Abdul Kadir, Ahmad A Mannan, Andrzej M Kierzek, Johnjoe McFadden, and Kazuyuki Shimizu. "Modeling and simulation of the main metabolism in Escherichia coli and its several single-gene knockout mutants with experimental verification." In: *Microbial Cell Factories* 9.1 (2010), p. 1.

[77] Mathumai Kanapathipillai, Georg Lentzen, Michael Sierks, and Chan Beum Park. "Ectoine and hydroxyectoine inhibit aggregation and neurotoxicity of Alzheimer's β-amyloid." In: *FEBS letters* 579.21 (2005), pp. 4775–4780.

[78] Kenneth J Kauffman, Purusharth Prakash, and Jeremy S Edwards. "Advances in flux balance analysis." In: *Current opinion in biotechnology* 14.5 (2003), pp. 491–496.

[79] Steven M Kelk, Brett G Olivier, Leen Stougie, and Frank J Bruggeman. "Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks." In: *Scientific reports* 2 (2012).

[80] Sean P Kennedy, Wailap Victor Ng, Steven L Salzberg, Leroy Hood, and Shiladitya DasSarma. "Understanding the adaptation of Halobacterium species NRC-1 to its extreme environment through computational analysis of its genome sequence." In: *Genome Research* 11.10 (2001), pp. 1641–1650.

[81] Steffen Klamt and Jörg Stelling. "Stoichiometric and constraint-based modeling." In: *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts* (2006), pp. 73–96.

[82] Arne Klingner, Annekathrin Bartsch, Marco Dogs, Irene Wagner-Döbler, Dieter Jahn, Meinhard Simon, Thorsten Brinkhoff, Judith Becker, and Christoph Wittmann. "Large-Scale 13C Flux Profiling Reveals Conservation of the Entner-Doudoroff Pathway as a Glycolytic Strategy among Marine Bacteria That Use Glucose." In: *Applied and environmental microbiology* 81.7 (2015), pp. 2408–2422.

[83] Sonja Kolp, Markus Pietsch, Erwin A Galinski, and Michael Gütschow. "Compatible solutes as protectants for zymogens against proteolysis." In: *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1764.7 (2006), pp. 1234–1242.

[84] Eugene V Koonin and Michael Y Galperin. *Sequence - Evoluion - Function*. Kluwer Academic Publishers, 2002.

[85] Annette Kraegeloh and Hans Kunte. "Novel insights into the role of potassium for osmoregulation in Halomonas elongata." In: *Extremophiles* 6.6 (2002), pp. 453–462.

[86] Hans Adolf Krebs and HL Kornberg. *Energy transformations in living matter*. Springer, 1957.

[87] Andreas Kremling. *Systems biology: mathematical modeling and model analysis*. CRC Press, 2013.

[88] A Kröger, Vladimir Dadák, Martin Klingenberg, and Friederike Diemer. "On the role of quinones in bacterial electron transport. Differential roles of ubiquinone and menaquinone in Proteus rettgeri." In: *European journal of biochemistry/FEBS* 21.3 (1971), p. 322.

[89] Lars Kuepfer, Matthias Peter, Uwe Sauer, and Jörg Stelling. "Ensemble modeling for analysis of cell signaling dynamics." In: *Nature biotechnology* 25.9 (2007), pp. 1001–1006.

[90] Anne Kümmel, Sven Panke, and Matthias Heinemann. "Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data." In: *Molecular systems biology* 2.1 (2006).

[91] Hans Jörg Kunte. "Osmoregulation in halophilic bacteria." In: *Extremophiles, Encyclopedia of Life Support (EOLSS)* 2 (2012), pp. 263–277.

[92] JANOS K Lanyi. "Salt-dependent properties of proteins from extremely halophilic bacteria." In: *Bacteriological Reviews* 38.3 (1974), p. 272.

[93] H Larsen. "The halobacteria's confusion to biology." In: *Antonie van Leeuwenhoek* 39.1 (1973), pp. 383–396.

[94] Hannes Link, Dimitris Christodoulou, and Uwe Sauer. "Advancing metabolic models with kinetic information." In: *Current opinion in biotechnology* 29 (2014), pp. 8–14.

[95] Karin Lippert and Erwin A Galinski. "Enzyme stabilization be ectoine-type compatible solutes: protection against heating, freezing and drying." In: *Applied Microbiology and Biotechnology* 37.1 (1992), pp. 61–65.

[96] Francisco Llaneras and Jesús Picó. "Stoichiometric modelling of cell metabolism." In: *Journal of Bioscience and Bioengineering* 105.1 (2008), pp. 1–11.

[97] LookForDiagnosis.com. "Halomonas (Halomonas elongata)." `https://lookfordiagnosis.com/mesh_info.php?term=Halomonas&lang=1`. 2014.

[98] Radhakrishnan Mahadevan, Daniel R Bond, Jessica E Butler, Abraham Esteve-Núñez, Maddalena V Coppi, Bernhard O Palsson, Christopher H Schilling, and DR Lovley. "Characterization of metabolism in the Fe (III)-reducing organism Geobacter sulfurreducens by constraint-based modeling." In: *Applied and environmental microbiology* 72.2 (2006), pp. 1558–1568.

[99] John M Martinko and MT Madigan. *Brock biology of microorganisms*. 2005.

[100] Thomas Maskow and Urs von Stockar. "How reliable are thermodynamic feasibility statements of biochemical pathways?" In: *Biotechnology and Bioengineering* 92.2 (2005), pp. 223–230.

[101] Alastair T Matheson, G Dennis Sprott, IJ McDonald, and Horace Tessier. "Some properties of an unidentified halophile: growth characteristics, internal salt concentration, and morphology." In: *Canadian journal of microbiology* 22.6 (1976), pp. 780–786.

[102] Alan D McNaught and Alan D McNaught. *Compendium of chemical terminology*. Vol. 1669. Blackwell Science Oxford, 1997.

[103] Enrique Meléndez-Hevia and Angel Isidoro. "The game of the pentose phosphate cycle." In: *Journal of theoretical biology* 117.2 (1985), pp. 251–263.

[104] Enrique Meléndez-Hevia, Thomas G Waddell, and Francisco Montero. "Optimization of metabolism: the evolution of metabolic pathways toward simplicity through the game of the pentose phosphate cycle." In: *Journal of theoretical biology* 166.2 (1994), pp. 201–220.

[105] Leonor Michaelis and Maud L Menten. "Die kinetik der invertinwirkung." In: *Biochem. z* 49.333-369 (1913), p. 352.

[106] Ron Milo and Rob Philips. *Cell Biology by the numbers*. Taylor & Francis Ltd., 2016.

[107] L. Motitschke, H. Driller, and E. Galinski. *Ectoin and ectoin derivatives as moisturizers in cosmetics*. US Patent 6,403,112. June 2002. URL: https://www.google.com/patents/US6403112.

[108] Armen Y Mulkidjanian, Pavel Dibrov, and Michael Y Galperin. "The past and present of sodium energetics: may the sodium-motive force be with you." In: *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1777.7 (2008), pp. 985–992.

[109] Frederick Carl Neidhardt, John L Ingraham, and Moselio Schaechter. "Physiology of the bacterial cell: a molecular approach." In: (1990).

[110] David L Nelson, Albert L Lehninger, and Michael M Cox. *Lehninger principles of biochemistry*. Macmillan, 2008.

[111] Elad Noor, Arren Bar-Even, Avi Flamholz, Ed Reznik, Wolfram Liebermeister, and Ron Milo. "Pathway thermodynamics highlights kinetic obstacles in central metabolism." In: *PLoS Comput Biol* 10.2 (2014), e1003483.

[112] Elad Noor, Eran Eden, Ron Milo, and Uri Alon. "Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy." In: *Molecular cell* 39.5 (2010), pp. 809–820.

[113] Elad Noor, Avi Flamholz, Wolfram Liebermeister, Arren Bar-Even, and Ron Milo. "A note on the kinetics of enzyme action: a decomposition that highlights thermodynamic effects." In: *FEBS letters* 587.17 (2013), pp. 2772–2777.

[114] Elad Noor, Hulda S Haraldsdóttir, Ron Milo, and Ronan MT Fleming. "Consistent estimation of Gibbs energy using component contributions." In: *PLoS Comput Biol* 9.7 (2013), e1003098.

[115] Hisayo Ono, Kazuhisa Sawada, Nonpanga Khunajakr, Tao Tao, Mihoko Yamamoto, Masayuki Hiramoto, Atsuhiko Shinmyo, Mitsuo Takano, and Yoshikatsu Murooka. "Characterization of biosynthetic enzymes for ectoine as a compatible solute in a moderately halophilic eubacterium, Halomonas elongata." In: *Journal of bacteriology* 181.1 (1999), pp. 91–99.

[116] Aharon Oren. "Microbial life at high salt concentrations: phylogenetic and metabolic diversity." In: *Saline systems* 4.2 (2008), p. 13.

[117] Aharon Oren, Mikal Heldal, and Svein Norland. "X-ray microanalysis of intracellular ions in the anaerobic halophilic eubacterium Haloanaerobium praevalens." In: *Canadian Journal of Microbiology* 43.6 (1997), pp. 588–592.

[118] Jeffrey D Orth, Tom M Conrad, Jessica Na, Joshua A Lerman, Hojung Nam, Adam M Feist, and Bernhard Ø Palsson. "A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011." In: *Molecular systems biology* 7.1 (2011), p. 535.

[119] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. "What is flux balance analysis?" In: *Nature biotechnology* 28.3 (2010), pp. 245–248.

[120] KP Pandya and HK King. "Ubiquinone and menaquinone in bacteria: A comparative study of some bacterial respiratory systems." In: *Archives of biochemistry and biophysics* 114.1 (1966), pp. 154–157.

[121] Petra Peters, EA Galinski, and HG Trüper. "The biosynthesis of ectoine." In: *FEMS microbiology letters* 71.1-2 (1990), pp. 157–162.

[122] Gösta Petterson. "Evolutionary optimization of the catalytic efficiency of enzymes." In: *European Journal of Biochemistry* 206.1 (1992), pp. 289–295.

[123] Kenneth S Pitzer. "Thermodynamics of electrolytes. I. Theoretical basis and general equations." In: *The Journal of Physical Chemistry* 77.2 (1973), pp. 268–277.

[124] Kenneth S Pitzer. "Thermodynamics of electrolytes. V. Effects of higher-order electrostatic terms." In: *Journal of Solution Chemistry* 4.3 (1975), pp. 249–265.

[125] Kenneth S Pitzer and Janice J Kim. "Thermodynamics of electrolytes. IV. Activity and osmotic coefficients for mixed electrolytes." In: *Journal of the American Chemical Society* 96.18 (1974), pp. 5701–5707.

[126] Kenneth S Pitzer and Guillermo Mayorga. "Thermodynamics of electrolytes. II. Activity and osmotic coefficients for strong electrolytes with one or both ions univalent." In: *The Journal of Physical Chemistry* 77.19 (1973), pp. 2300–2308.

[127] Kenneth S Pitzer and Guillermo Mayorga. "Thermodynamics of electrolytes. III. Activity and osmotic coefficients for 2–2 electrolytes." In: *Journal of Solution Chemistry* 3.7 (1974), pp. 539–546.

[128] Kenneth S Pitzer, John R Peterson, and Leonard F Silvester. "Thermodynamics of electrolytes. IX. Rare earth chlorides, nitrates, and perchlorates." In: *Journal of Solution Chemistry* 7.1 (1978), pp. 45–56.

[129] Kenneth S Pitzer, Rabindra N Roy, and Leonard F Silvester. "Thermodynamics of electrolytes. 7. Sulfuric acid." In: *Journal of the American Chemical Society* 99.15 (1977), pp. 4930–4936.

[130] Kenneth S Pitzer and Leonard F Silvester. "Thermodynamics of electrolytes. VI. Weak electrolytes including H3PO4." In: *Journal of Solution Chemistry* 5.4 (1976), pp. 269–278.

[131] J. Pramanik and J. D. Keasling. "Stoichiometric Model of *Escherichia coli* Metabolism: Incorporation of Growth-Rate Dependent Biomass Composition and Mechanistic Energy Requirements." In: *John Wiley & Sons* (1997), pp. 398–421.

[132] Nathan D Price, Iman Famili, Daniel A Beard, and Bernhard Ø Palsson. "Extreme pathways and Kirchhoff's second law." In: *Biophysical journal* 83.5 (2002), p. 2879.

[133] Hong Qian, Daniel A Beard, and Shou-dan Liang. "Stoichiometric network theory for nonequilibrium biochemical systems." In: *European Journal of Biochemistry* 270.3 (2003), pp. 415–421.

[134] S Rengpipat, SE Lowe, and JG Zeikus. "Effect of extreme salt concentrations on the physiology and biochemistry of Halobacteroides acetoethylicus." In: *Journal of bacteriology* 170.7 (1988), pp. 3065–3071.

[135] Mary F Roberts. "Organic compatible solutes of halotolerant and halophilic microorganisms." In: *Saline systems* 1.5 (2005), pp. 1–30.

[136] Robert Anthony Robinson and Robert Harold Stokes. *Electrolytic solutions*. Butterworth, 1959.

[137] Eytan Ruppin, Jason A Papin, Luis F De Figueiredo, and Stefan Schuster. "Metabolic reconstruction, constraint-based analysis and game theory to probe genome-scale metabolic networks." In: *Current opinion in biotechnology* 21.4 (2010), pp. 502–510.

[138] M Sadler, M McAninch, R Alico, and LI Hochstein. "The intracellular Na+ and K+ composition of the moderately halophilic bacterium, Paracoccus halodenitrificans." In: *Canadian journal of microbiology* 26.4 (1980), pp. 496–502.

[139] Uwe Sauer, Fabrizio Canonaco, Sylvia Heri, Annik Perrenoud, and Eliane Fischer. "The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of Escherichia coli." In: *Journal of Biological Chemistry* 279.8 (2004), pp. 6613–6619.

[140] Michael A Savageau. "Biochemical systems analysis: I. Some mathematical properties of the rate law for the component enzymatic reactions." In: *Journal of theoretical biology* 25.3 (1969), pp. 365–369.

[141] Michael A Savageau. "Biochemical systems analysis: II. The steady-state solutions for an n-pool system using a power-law approximation." In: *Journal of theoretical biology* 25.3 (1969), pp. 370–379.

[142] Michael A Savageau. "Biochemical systems analysis: III. Dynamic solutions using a power-law approximation." In: *Journal of theoretical biology* 26.2 (1970), pp. 215–226.

[143] Michael A Savageau. "Optimal design of feedback control by inhibition." In: *Journal of molecular evolution* 5.3 (1975), pp. 199–222.

[144] Michael A Savageau. "Biochemical systems analysis." In: (1976).

[145] Michael A Savageau. "Design principles for elementary gene circuits: Elements, methods, and examples." In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 11.1 (2001), pp. 142–159.

[146] George Scatchard. *Equilibrium in Solutions; Surface and Colloid Chemistry*. Harvard University Press, 1976.

[147] Jan Schellenberger, Nathan E Lewis, and Bernhard Ø Palsson. "Elimination of thermodynamically infeasible loops in steady-state metabolic models." In: *Biophysical journal* 100.3 (2011), pp. 544–553.

[148] Robert Schuetz, Nicola Zamboni, Mattia Zampieri, Matthias Heinemann, and Uwe Sauer. "Multidimensional optimality of microbial metabolism." In: *Science* 336.6081 (2012), pp. 601–604.

[149] Stefan Schuster, Thomas Pfeiffer, and David A Fell. "Is maximization of molar yield in metabolic networks favoured by evolution?" In: *Journal of theoretical biology* 252.3 (2008), pp. 497–504.

[150] John H Schwacke and Eberhard O Voit. "Improved methods for the mathematically controlled comparison of biochemical systems." In: *Theoretical Biology and Medical Modelling* 1.1 (2004), p. 1.

[151] T. Schwarz, G. Lentzen, and J. Krutmann. *Orally Used Compatible Solutes Containing Agents*. US Patent App. 11/885,687. Mar. 2009. URL: https://www.google.com/patents/US20090060876.

[152] Thomas Schwarz. *Use of compatible solutes as substances having free radical scavenging properties*. US Patent 20,030,147,937. Aug. 2003.

[153] Karin Schwibbert, Alberto Marin-Sanguino, Irina Bagyan, Gabriele Heidrich, Georg Lentzen, Harald Seitz, Markus Rampp, Stephan C Schuster, Hans-Peter Klenk, Friedhelm Pfeiffer, et al. "A blueprint of ectoine metabolism from the genome of the industrial producer Halomonas elongata DSM 2581T." In: *Environmental microbiology* 13.8 (2011), pp. 1973–1994.

[154] Daniel Segre, Dennis Vitkup, and George M Church. "Analysis of optimality in natural and perturbed metabolic networks." In: *Proceedings of the National Academy of Sciences* 99.23 (2002), pp. 15112–15117.

[155] Christiana Sehr, Andreas Kremling, and Alberto Marin-Sanguino. "Design Principles as a Guide for Constraint Based and Dynamic Modeling: Towards an Integrative Workflow." In: *Metabolites* 5.4 (2015), pp. 601–635.

[156] Jörg Severin, Axel Wohlfarth, and Erwin A Galinski. "The predominant role of recently discovered tetrahydropyrimidines for the osmoadaptation of halophilic eubacteria." In: *Journal of General Microbiology* 138.8 (1992), pp. 1629–1638.

[157] Oren Shoval, Hila Sheftel, Guy Shinar, Yuval Hart, Omer Ramote, Avi Mayo, Erez Dekel, Kathryn Kavanagh, and Uri Alon. "Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space." In: *Science* 336.6085 (2012), pp. 1157–1160.

[158] Sigma-Aldrich. "Ectoine." http://www.sigmaaldrich.com/catalog/product/sigma/e2271?lang=de&region=DE.

[159] Leonard F Silvester and Kenneth S Pitzer. "Thermodynamics of electrolytes. 8. High-temperature properties, including enthalpy and heat capacity, with application to sodium chloride." In: *The Journal of Physical Chemistry* 81.19 (1977), pp. 1822–1828.

[160] Leonard F Silvester and Kenneth S Pitzer. "Thermodynamics of electrolytes. X. Enthalpy and the effect of temperature on the activity coefficients." In: *Journal of Solution Chemistry* 7.5 (1978), pp. 327–337.

[161] Kieran Smallbone and Evangelos Simeonidis. "Flux balance analysis: a geometric perspective." In: *Journal of theoretical biology* 258.2 (2009), pp. 311–315.

[162] Keng Cher Soh and Vassily Hatzimanikatis. "Network thermodynamics in the post-genomic era." In: *Current opinion in microbiology* 13.3 (2010), pp. 350–357.

[163] Albert Sorribas, Benito Hernández-Bermejo, Ester Vilaprinyo, and Rui Alves. "Cooperativity and saturation in biochemical networks: a saturable formalism using Taylor series approximations." In: *Biotechnology and bioengineering* 97.5 (2007), pp. 1259–1277.

[164] G. N. Stephanopoulos, A. A. Aristidou, and Nielsen J. *Metabolic Engineering*. Academic Press, 1998.

[165] Ralph E Steuer. *Multiple criteria optimization: theory, computation, and applications*. Wiley, 1986.

[166] Suzy Strutner and Sergey Anashkevych. "Ultra-Salty Lake Koyashskoye Is So Pink, It Looks Like Another Planet." http://www.huffingtonpost.com/2015/06/09/koyashskoye-pink-lake-crimea-ukraine-photos_n_7534644.html. 2015.

[167] Hilal Taymaz-Nikerel, Amin Espah Borujeni, Peter JT Verheijen, Joseph J Heijnen, and Walter M van Gulik. "Genome-derived minimal metabolic models for Escherichia coli MG1655 with estimated in vivo respiratory ATP stoichiometry." In: *Biotechnology and bioengineering* 107.2 (2010), pp. 369–381.

[168] Andreas Tebbe, Christian Klein, Birgit Bisle, Frank Siedler, Beatrix Scheffer, Carolina Garcia-Rizo, Jan Wolfertz, Volker Hickmann, Friedhelm Pfeiffer, and Dieter Oesterhelt. "Analysis of the cytosolic proteome of Halobacterium salinarum and its implication for genome annotation." In: *Proteomics* 5.1 (2005), pp. 168–179.

[169] Ines Thiele and Bernhard Ø Palsson. "A protocol for generating a high-quality genome-scale metabolic reconstruction." In: *Nature protocols* 5.1 (2010), pp. 93–121.

[170] G Unden and J Bongaerts. "Alternative respiratory pathways of< i> Escherichia coli</i>: energetics and transcriptional regulation in response to electron acceptors." In: *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1320.3 (1997), pp. 217–234.

[171] C Vargas, Mohamed Jebbar, R Carrasco, Carlos Blanco, MI Calderón, F Iglesias-Guerra, and JJ Nieto. "Ectoines as compatible solutes and carbon and energy sources for the halophilic bacterium Chromohalobacter salexigens." In: *Journal of applied microbiology* 100.1 (2006), pp. 98–107.

[172] Amit Varma and Bernhard O Palsson. "Metabolic capabilities of Escherichia coli: I. Synthesis of biosynthetic precursors and cofactors." In: *Journal of theoretical biology* 165.4 (1993), pp. 477–502.

[173] Amit Varma and Bernhard O Palsson. "Metabolic capabilities of Escherichia coli II. Optimal growth patterns." In: *Journal of Theoretical Biology* 165.4 (1993), pp. 503–522.

[174] Antonio Ventosa, Joaquín J Nieto, and Aharon Oren. "Biology of moderately halophilic aerobic bacteria." In: *Microbiology and Molecular Biology Reviews* 62.2 (1998), pp. 504–544.

[175] Eberhard O Voit. *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists*. Cambridge University Press, 2000.

[176] Eberhard O Voit. "Biochemical systems theory: a review." In: *ISRN Biomathematics* 2013 (2013).

[177] Nishith Vora and Prodromos Daoutidis. "Nonlinear model reduction of chemical reaction systems." In: *AIChE Journal* 47.10 (2001), pp. 2320–2332.

[178] RH Vreeland, R Anderson, and RG Murray. "Cell wall and phospholipid composition and their contribution to the salt tolerance of Halomonas elongata." In: *Journal of bacteriology* 160.3 (1984), pp. 879–883.

[179] Russell H Vreeland, William D Rosenzweig, and Dennis W Powers. "Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal." In: *Nature* 407.6806 (2000), pp. 897–900.

[180] John R Williamson. "Glycolytic control mechanisms I. Inhibition of glycolysis by acetate and pyruvate in the isolated, perfused rat heart." In: *Journal of Biological Chemistry* 240.6 (1965), pp. 2308–2321.

[181] Edwin H Wintermute, Tami D Lieberman, and Pamela A Silver. "An objective function exploiting suboptimal solutions in metabolic networks." In: *BMC systems biology* 7.1 (2013), p. 98.

[182] Feng Yang, Hong Qian, and Daniel A Beard. "Ab initio prediction of thermodynamically feasible reaction directions from biochemical network stoichiometry." In: *Metabolic engineering* 7.4 (2005), pp. 251–259.

[183] Antonios Zagaris, Hans G Kaper, and Tasso J Kaper. "Fast and slow dynamics for the computational singular perturbation method." In: *Multiscale Modeling & Simulation* 2.4 (2004), pp. 613–638.

[184] Nicola Zamboni, Anne Kümmel, and Matthias Heinemann. "anNET: a tool for network-embedded thermodynamic analysis of quantitative metabolome data." In: *BMC bioinformatics* 9.1 (2008), p. 199.

[185] Nicola Zamboni, Hannu Maaheimo, Thomas Szyperski, Hans-Peter Hohmann, and Uwe Sauer. "The phosphoenolpyruvate carboxykinase also catalyzes C 3 carboxylation at the interface of glycolysis and the TCA cycle of Bacillus subtilis." In: *Metabolic engineering* 6.4 (2004), pp. 277–284.

[186] Dan Zilberstein, V Agmon, S Schuldiner, and E Padan. "Escherichia coli intracellular pH, membrane potential, and cell growth." In: *Journal of bacteriology* 158.1 (1984), pp. 246–252.

DECLARATION

I hereby declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given in the bibliography.

München, June 2016

Christiana Verena Sehr

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

`https://bitbucket.org/amiede/classicthesis/`