Uniting *In Silico* and *In Vivo* Systems Biology: a New Concept to Approximate Theory to Real-Life Flux Distributions

I n a u g u r a l - D i s s e r t a t i o n

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Thorsten Fallisch

aus Düsseldorf

2006

Für Sonja

Contents

List of Figures

List of Tables

Abbreviations

BM	Biomass
Cgl	*Corynebacterium glutamicum*
CUBIC	**C**ologne **U**niversity **B**io**I**nformatics **C**enter
EBA	energy balance analysis
EEB	external energy balance
FBA	flux balance analysis
GC	Gas chromatography
Glc	Glucose
NaAc	(Sodium) Acetate
LC	Liquid chromatography
MS	Mass spectrometry

Chapter 1

Introduction

It is well known in the scientific community that in the past years a steadily growing amount of complete genomes of cellular organisms has become available. Alongside, the wish to understand the metabolism of those life-forms grew. A paradigm shift, necessary to comprehend the metabolic interactions, is going through biology: research moves from reductionist approaches, which take the single units as a basis, to holistic, system-based concepts (Kauffman *et al.*, 2003).

This shift can be seen in "wet-lab" biology as well as bioinformatics. Investigations carried out in laboratories e. g. move from efforts to find out as much as possible about a single enzyme to system based methods like metabolic profiling (compare 2.5). Within bioinformatics, in recent years, the field of theoretical systems biology evolved. Prior areas are often focused on single molecules, their structures, properties, and interactions. Examples stretch from *ab initio* protein structure prediction (Osguthorpe, 1999, 2000) to homology modeling, which takes advantage of structural information available in databases (Al-Lazikani *et al.*, 2001); from protein-protein docking (Vajda & Camacho, 2004) to genome annotation (Reed *et al.*, 2006), which describes the connection between the genomic sequence and the function of the coded protein.

The methods used in theoretical systems biology vary. Petri nets (see e.g. (Pinney *et al.*, 2003; Hartmann, 2006)), for instance, are an approach that has to be classified more reductionist: modeling the network as a composition of single reactions, the latter are the simulated units. In parallel, strategies exist that take pathways rather than reactions as the basic units. Examples are shortest path analyses (e.g. (Rahman *et al.*, 2005)) or the concept of *elementary modes* (Schuster *et al.*, 1999). One level up, flux balance analysis (e.g. (Kauffman *et al.*, 2003)) simulates the metabolism as a whole, yielding information about the constituting reactions from above rather than simulating them separately. And of course hybrid methods have been developed, like the combination of flux balance analysis (*FBA*) and a subset of the elementary modes called *extreme pathways* (Schilling *et al.*, 2000, 2001).

Flux balance analysis, as a fully systemic approach, is solely based on the topology of the metabolic network. This is expressed as a stoichiometric matrix and is set into a steady state. Mathematically, this procedure yields a simple linear equation system in which the vector of the reaction rates contains the unknown variables (see chapter 2.6 for details). By imposing additional constraints this system can eventually be solved, resulting in a flux distribution for the network under investigation. Edwards *et al.* showed that *FBA* yields results that are in accordance with nature (Edwards *et al.*, 2001). Daniel Beard and his colleagues refined the method and called it *energy balance analysis* (*EBA*). They added rules derived from the fundamental laws of thermodynamics (2.1.1). Details about *EBA* can be found in chapter 2.7.

Knowledge of the current flux pattern within a cellular network provides information about the way the organism operates. Especially comparisons of flux vectors derived under different constraints enable biologists to understand the respondence of life to varying environmental circumstances.

Thereby, the powerful adaptability of bacteria can be further elucidated. We may understand how exactly some bacteria metabolize given carbon sources in excess amino acids, which are then excreted. And by predicting the adaption to perturbances or changes in the nutrimental supplies, scientists are able to gain knowledge usable to improve biotechnological production yields. Of course, there exist methods to determine flux distributions experimentally (e. g. (Wendisch *et al.*, 2000)). But these suffer from being slower and more costly than predictive approaches. Additionally, certain constellations of metabolic reactions and pathways are necessary to determine the desired flux rates. In consequence, experimental approaches usually only cover small parts of the investigated reaction networks.

Unfortunately, all current methods used to *predict* metabolic flux patterns result in intervals for each flux, not in discrete values. The smaller the intervals are, the more distinct and the more *valuable* is the information gained. This work aims at the reduction of these intervals. We will take advantage of available data and unite systemic approaches from bioinformatics with those from "wet-lab" work. Namely, a method will be developed which integrates metabolic profiling data into *EBA* analyses. To accomplish this an idea that bases on thermodynamics is employed: the same network is analyzed under different environmental conditions. Thereby, the information gathered from the corresponding profiling experiments is used to implement a "greater than / less than" relation between the same reactions under the alternative conditions. This shall narrow the gap between nature and prediction and reduce the size of the flux space that remains after an energy balance analysis.

To test and validate the developed concept, we must employ a model of a bacterial organism. There is a variety of genome based reaction networks available. The bacterium modeled in most detail is certainly *Escherichia coli* (see e. g. (Reed & Palsson, 2003)). Other examples include *Helicobacter pylori* (Thiele *et al.*, 2005; Schilling *et al.*, 2002; J.F. *et al.*, 1997) and the

eukaryotic *Saccharomyces cervisiae* (Förster *et al.*, 2003; Mewes *et al.*, 1997). A more detailed discussion can be found in chapter 5.1.

We are in need of experimental metabolic profiling data. In CUBIC[1] research laboratories an organism called *Corynebacterium glutamicum* is under referring investigation. *C. glutamicum* is a gram-positive, aerobic, and non-pathogenous soil bacterium. Discovered in 1957 at the research institute of Kyowa Hakko in Tokyo, the genome was sequenced more than 40 years later, see e. g. (Ikeda & Nakagawa, 2003) and (Kalinowski *et al.*, 2003). In the meantime the biotechnological industry learned to use *C. glutamicum* as a means to produce various amino acids and vitamins in noticeable amounts. Most important is the production of the L-amino acids glutamate and lysine. The exact quantities given in literature vary slightly. According to Silberbach *et al.* more than 10^6 t of L-glutamate and around $5.6 \cdot 10^5$ t of L-lysine are produced by this bacterium annually worldwide (Silberbach *et al.*, 2005). Hüser and coworkers give the same amount of L-glutamate but a slightly lower amount of L-lysine ($4.5 \cdot 10^5$ t / year) in (Hüser *et al.*, 2003)[2].

In the first part of this thesis a genome based model of the *Corynebacterium glutamicum* metabolism is built. It is supported by the use of DNA microarray analyses data. We verified the model by comparison to experimental flux data. In the second part, the concept to unite the metabolic profiling data and the energy balance analysis is introduced. It is shown that this combination leads to a further reduction of the acquired steady state flux space. Further on, we performed a complete single knock out mutation study of the constructed model. The metabolic network, its verification, the new concept and its application, and the knock out study are discussed in detail.

---

[1] CUBIC = **C**ologne **U**niversity **B**io**I**nformatics **C**enter

[2] The differences in the stated amounts of L-lysine produced by *C. glutamicum* may be related to the times the articles were published (2003 and 2005).

Chapter 2

Theory

This chapter will provide the necessary background to understand what is described in the methods, results, and discussion.

2.1 Thermodynamics

In the following, the very basic ideas of (bio)chemical thermodynamics will be explained. To follow the procedures taken within this thesis, one does not need detailed knowledge of thermodynamics. Thus, what is to be explained, will be introduced on a high level of abstraction. The interested reader will find more detailed information in every standard textbook of physical chemistry. We suggest the lecture of (Atkins, 1990) at this point.

2.1.1 The fundamental laws of thermodynamics

The behavior of all processes occuring in the universe is governed by three fundamental laws. The first two of them are needed to follow this work.

2.1.1.1 The first law of thermodynamics

P.W. Atkins states the first law of thermodynamics as: "The internal energy of a system is constant unless it is changed by doing work or by heating"

(Atkins, 1990, page 31). If we view the universe as the mentioned system, and express the law in a more colloquial way, it states: energy can neither disappear into nor be created out of nothing.

Written as an equation the first law of thermodynamics reads:

$$\Delta U = q + w \tag{2.1}$$

where ΔU denotes the change in internal energy of a closed system, q is the amount of heat passing through its boundary, and w is the amount of work passing through the system's boundary.

The point of view, when talking about internal energies as well as energies at all, is within the system under investigation. It follows that energy supplied to the system has a positive sign and energy leaving the system has a negative sign.

Further on, one has to know that the internal energy is a state function. This means changes in U are only dependent on the states of the starting and endpoint of a process; they are independent of the way the process takes. In other words: the change in U is the same no matter *how* energy is supplied to the system.

2.1.1.2   The second law of thermodynamics

The second law of thermodynamics concerns the direction of spontaneous processes. Originally it says that spontaneous processes are always accompanied by an increase in entropy (entropy is a measure of chaos). By traversing some transformations, which are explained in detail in (Atkins, 1990, p. 82 ff.), one arrives at the following equation:

$$\mathrm{d}G_{T,p} \leq 0 \tag{2.2}$$

$\mathrm{d}G$ is the change in the *Gibbs free energy* (also called the *free enthalpy*) during the process

2.1.2 The Gibbs free energy or free enthalpy

The free enthalpy G is dependent on the enthalpy H, which again depends on the internal energy U (see equation 2.1). The correlations are:

$$H \;=\; U + p \cdot V \tag{2.3}$$

$$G \;=\; H - T \cdot S \tag{2.4}$$

Here, p denotes the actual pressure, V the volume, T the temperature and S is the before-mentioned entropy.

The Gibbs free energy can be classified into different divisions. The ones relevant for this thesis are

1. the Gibbs free energy of formation: G_f

2. the Gibbs free energy of reaction: G_r

2.1.2.1 Gibbs free energy of formation

G_f is the energy that is needed to build a molecule from its atoms. Since it is a very extensive process to determine this for each molecule experimentally, Mavrovouniotis developed a group contribution method in (Mavrovouniotis, 1998). Basically, the free enthalpy of reaction (see below) is calculated by equation 2.5. Educts and products are split into their constituent groups. After doing this for a set of reactions, one gains a linear equation system, which can be solved by multiple regression. As a result G_f values are gained for parts of molecules. In a final step these G_f values may be summed up in order to calculate free enthalpies for new molecules.
The above cited article determines the Gibbs free energies for increments in aqueous solution at a pH of seven. As explained in 2.1.3 this is the natural environment for biochemical reactions.
The programming and calculation of the G_f values for all molecules in this

work has been accomplished by Kai Hartmann (Hartmann, 2006).

2.1.2.2 Gibbs free energy of reaction

The Gibbs free energy of reaction is simply defined as the free enthalpy of all products minus the free enthalpy of all educts

$$\Delta G_r = \sum_{products} G_i - \sum_{educts} G_j \qquad (2.5)$$

If referring to reactions in aqueous solution (like usual in biochemistry) and to standard states, the free enthalpy of a reaction can be calculated as

$$\Delta G_r = \Delta G_r{}^{\ominus} - RT \cdot lnK \qquad (2.6)$$

$\Delta G_r{}^{\ominus}, R$ and T denote the reaction free enthalpy under standard conditions, the gas constant[1] and the actual temperature, respectively.

In the above equation K states the equilibrium constant, which in detail is written as:

$$K = \frac{\prod_p [p]^{\nu_p}}{\prod_e [e]^{\nu_e}} \qquad (2.7)$$

p are the products, e the educts of the reaction; ν denotes the order of the reaction in the referring metabolite[2].

Spontaneous (bio)chemical reactions are always accompanied by a negative Gibbs free energy of reaction, which is calculated by equation 2.6 !

2.1.3 Standard states in chemistry and biology

When comparing any thermodynamical values one must adjoin information of the states these values were derived under. In order to normalize this,

---

[1]$R = 8.31451 \frac{J}{K \cdot mol}$

[2]About the order of reactions read (Atkins, 1990, p. 782 ff.)

chemists have defined standard states, which are mostly used when publishing thermodynamical data. There are two sets of data currently used as standard values. The *standard temperature and pressure* (STP) and the *standard ambient temperature and pressure* (SATP). See table 2.1.

Table: 2.1: **Standard States in Chemistry and Biology**
The standards states used in chemistry and biology. STP = standard temperature and pressure. SATP = standard ambient temperature and pressure. *BIO* = standard states used in biology. For STP/SATP compare (Atkins, 1990, page 11), for BIO see (Voet & Voet, 1992, page 51). For states used in this work also compare (Mavrovouniotis, 1998). Variables derived under SATP conditions are signed by a ($^{\ominus}$), *BIO*-based variables by a prime ($\prime$).

STP	T	=	0°C	=	273.15 K
	p	=	1 atm	=	101.325 kPa
	pH	=	0		
SATP	$T^{\ominus}$	=	25°C	=	298.15 K
	$p^{\ominus}$	=	1 bar	=	100 kPa
	$pH^{\ominus}$	=	0		
BIO	T'	=	25°C	=	298.15 K
	$p\prime$	=	1 bar	=	100 kPa
	pH'	=	7		

All values used in this work refer to the state called *BIO* in the above table.

2.2   Biochemical reaction networks

This section shall briefly describe the specialties of biochemical networks. At first the important terms are explained. Then a short comparison between biochemical networks and other networks will be given. At last a discussion of thermodynamic aspects follows.

```
GATTACACTAATGT
CTAATGTGATTACA
```

Figure 2.1: **Complementary DNA strands**
DNA molecules consist of two strands that are complementary to each other. In four available bases there exist two complementary pairs, which are A / T and G / C. (A) adenine, (T) thymine, (G) guanine, (C) cytosine.

2.2.1 Genes - Enzymes - Metabolites

The examination of biochemical networks usually starts at genome level. The DNA, as the carrier of the genetic information, states the genome. This consists of genes, each of which encodes a protein. Looking at the DNA as a molecule, it basically consists of four different nucleotides (often also called bases): adenine (A) and thymine (T), guanine (G) and cytosine (C). There exist some more, but due to their minor appearances they shall be neglected now. Accordingly, a strand of DNA can be represented by a line of the four letters ATGC. However, DNA consists of two complementary strands that connect to each other. Thereby A and T are complementary and G and C are. Representing the DNA as a line of letters, an example is given in figure 2.1.

When a gene gets *expressed*[3] - meaning a protein is built out of the included information - it happens in two steps. During *transcription* the DNA double strand gets temporarily split and the complementary strand of **one** of the DNA strands is built. Thereby only a part of the DNA (e. g. one gene) is transcribed. The resulting molecule is one-stranded and termed RNA. During *translation* this RNA molecule is used to construct a protein. The coherences are summarized in the central dogma of molecular biology, which

---

[3]The term *expression* may either refer to the combination of transcription and translation, in which case the term *protein-expression* would be more appropriate, or it may refer to *gene-expression*, which describes the transcription of one gene, in contrast to the whole genome.
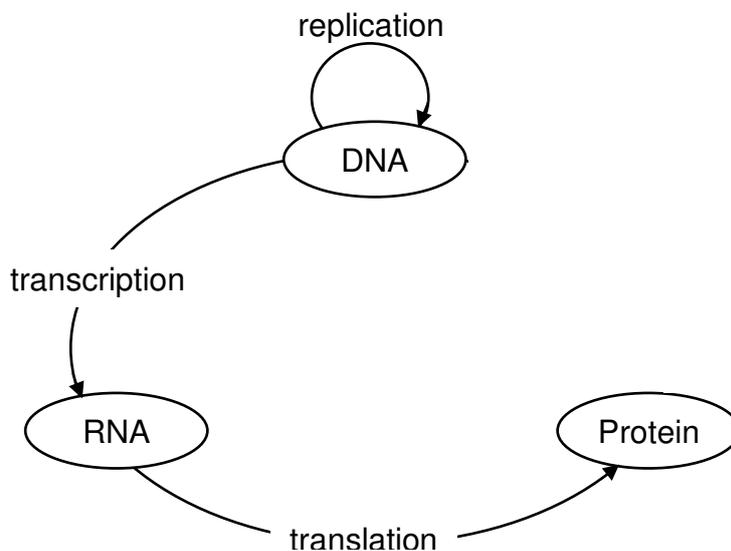
Figure 2.2: **Central dogma of molecular biology**
The arrows show the transfer of genetic information. There is no transfer of such from proteins to DNA, RNA or Proteins. In other words: proteins can only be *products* of genetic information.

is shown in figure 2.2.

Proteins are macromolecules which consist of 300 amino acids in average. These in turn are built of 19 - 29 atoms. Proteins are the workers of a cell. They carry out functions like membrane-crossing transport, conversion of molecules, or construction of other proteins. A special subgroup of the proteins is called *enzymes* which catalyze biochemical reactions. The molecules participating in those reactions are called *metabolites*.

When analyzing metabolic networks[4], the focus of interest may be laid on the metabolites or the enzymes. Within the thesis at hand we strongly concentrate on the metabolites. This induces some special aspects, that have

---

[4]The terms *metabolic network*, *reaction network* or *biochemical network* are interchangeable in this context.

to be taken care of, during the inspection of such networks. The following sections explain those aspects.

2.2.2 Biochemical networks - bipartite graphs

Networks can be expressed as graphs: the DNS-servers within the internet, the wastewater canal system under a city, or the atoms of a molecule that are connected via chemical bonds. There are lots of other examples.
A graph consists of nodes and edges. In the examples mentioned, the nodes would be the servers, the canal junctions, or the atoms, and the edges would be the phone lines, the canals, and the bonds, respectively. In all these networks the following facts are given:

- All nodes are on the same hierarchic level. In other words: there exists only one kind of nodes.

- Edges do connect exactly two nodes (which may be identical).

Regarding metabolic networks, the nodes are the metabolites and the edges refer to the reactions. But by using this analogy, problems arise. Most biochemical reactions do **not** convert <u>one</u> molecule into <u>one</u> other. Usually two or more metabolites are transformed into two or more other metabolites. This would mean that edges need to connect more than two nodes. This is indeed the usual way chemical reactions are written in textbooks as is depicted in figure 2.3

Additionally, stoichiometric coefficients complicate the matter. In a reaction like $[2A + B \rightarrow C]$ <u>two</u> metabolites of type A participate. Since there only may exist <u>one</u> node of type A, we have to attach a weight to an edge. Using the way shown in figure 2.3 (B), the weight of "two" has to be attached to the edge leading from A to *enzyme*.

When analyzing metabolic networks, one should keep the characteristics discussed above in mind.
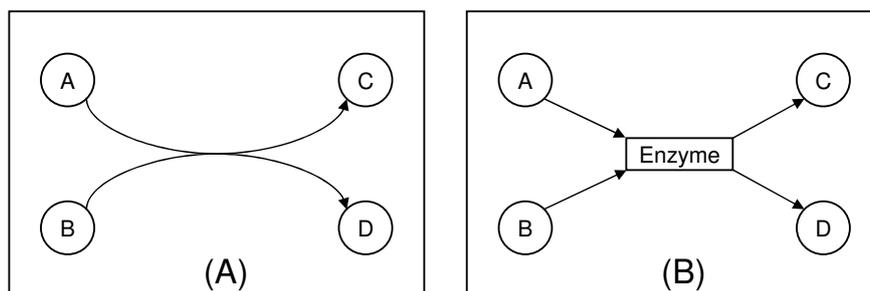
Figure 2.3: **Biochemical networks are bipartite graphs**
A) shows the way biochemical reactions are usually depicted. B) shows
the underlying bipartite graph. This knows two kinds of nodes. In case of
metabolic networks these refer to metabolites and reactions.

2.2.3 Thermodynamic aspects

In section 2.1.1.2 it was said that ΔG has to be negative, if a reaction is to
take place spontaneously. It is possible to reverse spontaneous reactions, e. g.
by heating. In this thesis we assume the model to exist under the standards
usual in biology (25°C, 1 atm, pH 7; compare table 2.1). All reactions are
to occur spontaneously ($\Delta G < 0$) and additionally we presume that there
exist no other ways of energy transport within the cell than by metabolites.
Zooming out to the whole network, it follows that also the energy balance of
the network itself has to be negative. Thereby, "energy balance" only refers
to the energy of compounds entering or leaving the cell. Energy transported
into the cell by heat, light, or other means is excluded at this point.

The aspects above are included in the thesis at hand. Energy balance analysis
(2.7) is used when referring to single reactions. The external energy balance
(3.5) depicts the energy proportions flowing into and out of the system as a
whole.

2.3 Enzyme and reaction kinetics

It was already stated on page 11 that enzymes are molecules which catalyze (bio)chemical reactions. Put in simple words, they lower energetic barriers that have to be overcome in order for a reaction to take place. Thereby, enzymes facilitate reactions which would otherwise be blocked by these barriers. Details about the method of operation and underlying thermodynamical issues can be found in (Voet & Voet, 1992). Usually, enzyme-catalyzed reactions do not follow simple chemical reaction kinetics (see below). E.g., Michaelis and Menten state the kinetic law for an enzyme-catalyzed reaction with one substrate, in absence of a reverse reaction, (see e. g. (Voet & Voet, 1992)) as:

$$v_0 = \frac{v_{max}[\text{S}]}{K_M + [\text{S}]} \tag{2.8}$$

v_0: initial velocity, v_{max}: maximal velocity, [S]: concentration of substrate, K_M: Michaelis constant.

Athel Cornish-Bowden describes the more complicated laws for multi-substrate reactions, which may include reverse reactions, in (Cornish-Bowden, 1995). Within the scope of this thesis we will assume linear reaction kinetics (i.e. chemical kinetics) for enzyme-catalyzed reactions. Under certain constraints (compare chapters 3.6.2.2 and 5.4.1 for assumed constraints and discussion) this is reasonable.

Ruled by linear reaction kinetics, a reaction velocity becomes dependent on its substrates *and* products in the following way

$$v \ = \ v_1 - v_2 \tag{2.9}$$

with

$$v_1 \ = \ k_1 * \prod_i [i]^{\nu_i} \tag{2.10}$$

$$v_2 \ = \ k_2 * \prod_j [j]^{\nu_j} \tag{2.11}$$

v states the overall reaction velocity, v_1 is the reaction rate in forward direction, k_1 the referring kinetic constant, v_2 and k_2 refer to backward direction. $[i]$ is the concentration of substrate i, $[j]$ the one of product j and ν_i and ν_j are the orders of the reactions in metabolites i and j respectively (compare (Atkins, 1990, p. 782 ff.)).

Additionally, some words have to be said concerning enzyme catalysis. Most enzymes are able to catalyze one chemical transformation at a time. There are exceptions, but they are rare. Thus, one can imagine a situation where lots of substrates need to be transformed by one enzyme molecule, which would not be able to accomplish that workload. This situation would be referred to as *saturation*. In general, saturation means that an increase in the substrate concentration does not result in an increase of the reaction rate.
Enzymes are themselves complex biological molecules for which regulatory mechanisms exist, e.g. feedback mechanisms. There are two general types of feedback: *activation* and *inhibition*. Both may be triggered by chemical compounds interacting with the enzyme.
Since enzymes are biological molecules, they may be digested by other enzymes. Also disintegration by external factors (e.g. by heat) may play a role. This in turn could change the overall flux through the catalyzed reaction. If disintegration occurs faster than expression, we will move closer to saturation.
More details on all of the above topics can be found in any standard biochemistry textbook (e.g. (Voet & Voet, 1992)).

2.4 DNA microarrays

DNA microarray analyses give a crude quantitative determination of the expression (2.2.1) of certain genes.

Individual DNA samples representing the genes of a genome are arrayed on membranes or glass plates. The RNA from a specific organism is taken, fluorescently labeled and brought onto the array. If a complementary gene fragment exists on the plate, the RNA will bind to it. Then all free RNA molecules left are washed away. By usage of a laser the marked RNA can be visualized. Since it is known which spot refers to which gene, the expressed genes can be identified. The relative intensity of the fluorescence refers to the amount of RNA molecules existent within the examined tissue.

So all genes that are transcribed from the genome at hand can be identified and relatively quantified. Of course "all" only includes those genes which are represented on the microarray. Interested persons are referred to (Hüser *et al.*, 2003) for further reading.

2.5 Metabolic profiling

Metabolic profiling experiments enable scientist to characterize the phenotypes of organisms on an experimental basis. A *metabolic profile* can be established along two dimensions. The term either represents the measured concentration of **one** metabolite versus time or the concentrations of **all** metabolites at one discrete point of time. Within the work at hand we usually refer to the latter variant.

2.5.1 Metabolic profiling - experiments

The metabolic profiling data implemented in this thesis was derived by a method that has been established by Sergey Strelkov *et al.* in (Strelkov

et al., 2004). Strelkov used a combination of gas chromatography and mass spectrometry (GC/MS) to quantify the metabolites in bacterial probes.

The chain of experiments starts with raising *C. glutamicum* under defined conditions. Thereby, fermenter technology is given the preference over shaking flasks, due to the possibility to control the pH value and to keep the oxygen saturation constant. Nutrition, however, is only supplied once at the beginning of the experiment. This method is referred to as *batch cultivation*, contrary to fed-batch (feed at the beginning and afterwards whenever nutrient limitation arises) or continuous supply of feeding materials. After starting the growth by initial feeding, in given time intervals probes are taken from the culture. The bacteria therein are lysed and the metabolic content is analyzed.

In a first step, the mixture is separated with the help of gas chromatography. The GC peak areas are used to quantify the individual components. In a second step, mass spectrometric analyses shall help to identify the metabolites. It must be known that the identification is usually carried out by comparing the retention time from the GC as well as the mass spectrum to the data retrieved by usage of commercially available standards. Databases may help at this point (Schauer *et al.*, 2005).

If neither standards are purchasable nor database entries exist, identification becomes complicated.

2.5.2 Metabolic profiling - interpretability

In principle, it is possible to perform quantitative analyses that yield discrete amounts. One needs to measure calibration curves for all components in advance. But, since this is a rather time consuming task, it has not been done for the data used in the work at hand. Thus, only a relative interpretation of the experimental data is allowed. Additionally, this relativity is only valid in one dimension. If the concentration of a specific metabolite changes between different measurements (e. g. between cultivation on glucose or acetate), it

can be determined that it doubles, bisects etc. Differences between diverse components, however, cannot be interpreted. E.g., there may be the result that ATP concentration doubles if the organism is raised on excess nitrogen, contrary to nitrogen starvation. But there will be no result as: ATP concentration is half of NADH concentration.

2.6   Flux balance analysis

FBA - an acronym of flux balance analysis - starts with the mass balance for a single metabolite x:

$$\frac{\mathrm{d}[x]}{\mathrm{d}t} = \vec{\mathbf{s}} \cdot \vec{\mathbf{v}} \qquad (2.12)$$

$[x]$ denotes the concentration of the metabolite; t is the time; $\vec{\mathbf{s}}$ is called the stoichiometric vector, which contains the stoichiometric coefficients of x in all reactions of the network; $\vec{\mathbf{v}}$ is the flux vector containing the flux values of all the network's reactions.

By zooming out to the whole network, $[x]$ becomes the concentration vector $\vec{\mathbf{c}}$ and $\vec{\mathbf{s}}$ becomes the stoichiometric matrix $\underline{\underline{\mathbf{S}}}$. Within $\underline{\underline{\mathbf{S}}}$, each row refers to one metabolite and each column to one reaction.

$$\frac{\mathrm{d}\vec{\mathbf{c}}}{\mathrm{d}t} = \underline{\underline{\mathbf{S}}} \cdot \vec{\mathbf{v}} \qquad (2.13)$$

During a flux balance analysis only a very short time slice $\mathrm{d}t$ is considered. This leads to the assumption of the metabolic network existing in a steady state. The effect is that the concentrations of all metabolites do not change over time. Expressed mathematically:

$$\vec{\mathbf{0}} = \underline{\underline{\mathbf{S}}} \cdot \vec{\mathbf{v}} \qquad (2.14)$$

The result is an under-determined, homogeneous, linear system of equations,

$$
\begin{array}{c} \\ A \\ B \\ C \\ D \end{array}
\begin{pmatrix}
-1 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & -1 & 0 & -1 & 0 \\
1 & 0 & 0 & -1 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & -1
\end{pmatrix}
*
\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \end{pmatrix}
= \vec{0}
$$

(A)   **(B)**
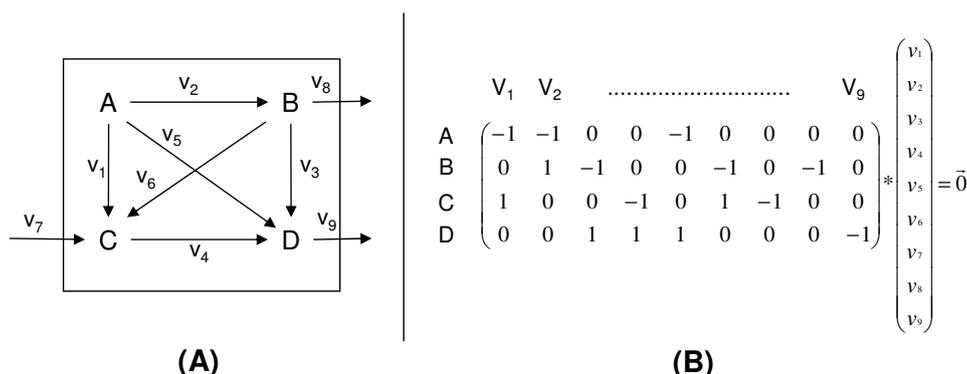
Figure 2.4: **FBA basics**
A) shows an example network with four metabolites, six internal and three membrane-crossing reactions. B) gives the referring basic equation of flux balance analysis (compare eq. 2.14). Note that membrane-crossing reactions (also called *external*) are defined to lead from the cytoplasm to a no deeper defined environment, hence their stoichiometric coefficients are always -1. Differentiation between *influxes* and *effluxes* is accomplished by imposition of constraints onto the *fluxes* (contrary to the *reactions*).

since natural metabolic networks normally own more reactions than metabolites (see also (Kauffman *et al.*, 2003; Schilling *et al.*, 2000, 2001)). Compare also figure 2.4.

As a consequence we are not able to obtain a discrete solution, but will get a solution space. Due to its biochemical background and its shape, this solution space is called the *steady state flux cone*. The cone is convex and, if no further constraints are imposed on the system, open at the top (see figure 2.5).

In general, there are side constraints made that influence those fluxes which refer to membrane transport reactions; e.g. the carbon source influx should be limited. According to the biochemical and environmental conditions, which are to be represented by the model, additional constraints may be imposed. For instance, the nitrogen influx could be limited to simulate the behavior under referring starvation. To analyze fluxes in an anaerobic environment, oxygen uptake can be prohibited. Any side constraints can be introduced as inequalities:
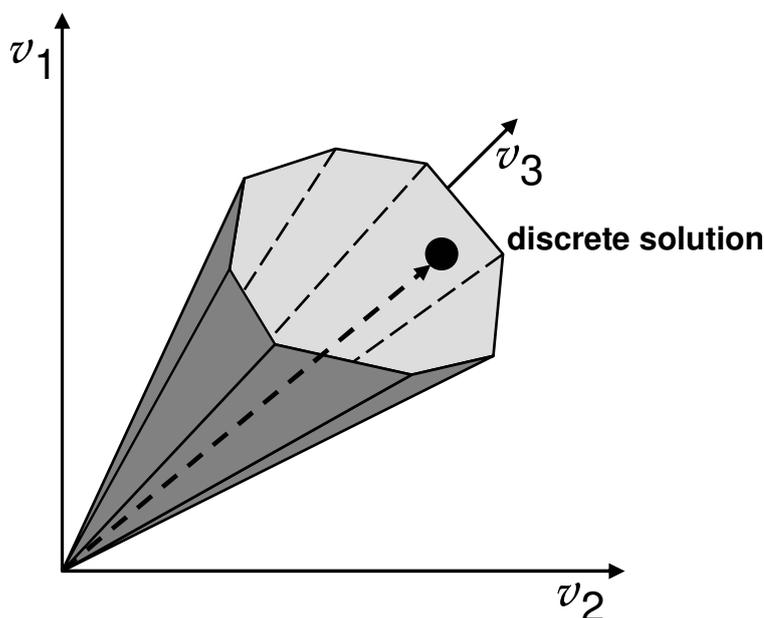
Figure 2.5: **Steady state flux cone**
The result of equation 2.14 is a polyhedral cone emanating from the origin and stretching into the metabolic flux space. Due to its biochemical background it is called the *steady state flux cone*. Figure cloned from (Schilling *et al.*, 2000).

$$\alpha_i \leq v_i \leq \beta_i \qquad (2.15)$$

with α_i and β_i being arbitrary numbers.

In order to obtain interpretable results, at least as many constraints have to be imposed as are necessary to close the steady state flux cone.

All possible flux distributions (i.e.: vectors $\vec{v}$) the system can achieve lie within the cone. Thus, its analysis presents the metabolic phenotypes the system is able to achieve.

To pick a discrete solution out of the flux cone, a sound objective function is stated and a linear optimization applied. Scientists agreed in regarding

the production of biomass as a reasonable objective function, at least for unicellular organisms (Segrè *et al.*, 2002).

$$
\begin{aligned}
\text{Maximize} \quad & \sum_{j} c_j \cdot v_j \\
\text{subject to} \quad & \sum_{ij} S_{ij} \cdot v_j = 0 \quad | \quad \alpha \leq v_j \leq \beta \quad\quad (2.16)
\end{aligned}
$$

i: metabolites, j: reactions; c_i: coefficients in objective; S_{ij}: stoichiometric coefficients; v_j: fluxes.

Mahadevan and Schilling explain in (Mahadevan & Schilling, 2003) that even after performing a linear optimization, in most cases, there is no single discrete solution. On the contrary, there are more than one possible flux distributions which yield the exact same biomass production rate. Figure 2.6 shows how variable fluxes constitute in general.

If only one of the depicted elements exists in the analyzed model, after optimization not one, but infinite solutions exist, which lead to the same optimum.

2.7 Energy balance analysis

The research group of Daniel Beard noted, that, by application of *FBA* only, flux distributions may occur which lack thermodynamic feasibility (Beard *et al.*, 2002). The first law of thermodynamics (2.1.1.1) states the conservation of energy. Applied to metabolic networks, it can be deduced that the sum of the reaction free enthalpies over a closed loop is zero. The second law (2.1.1.2) says that spontaneous chemical reactions are accompanied by a negative reaction free enthalpy (2.1.2.2).
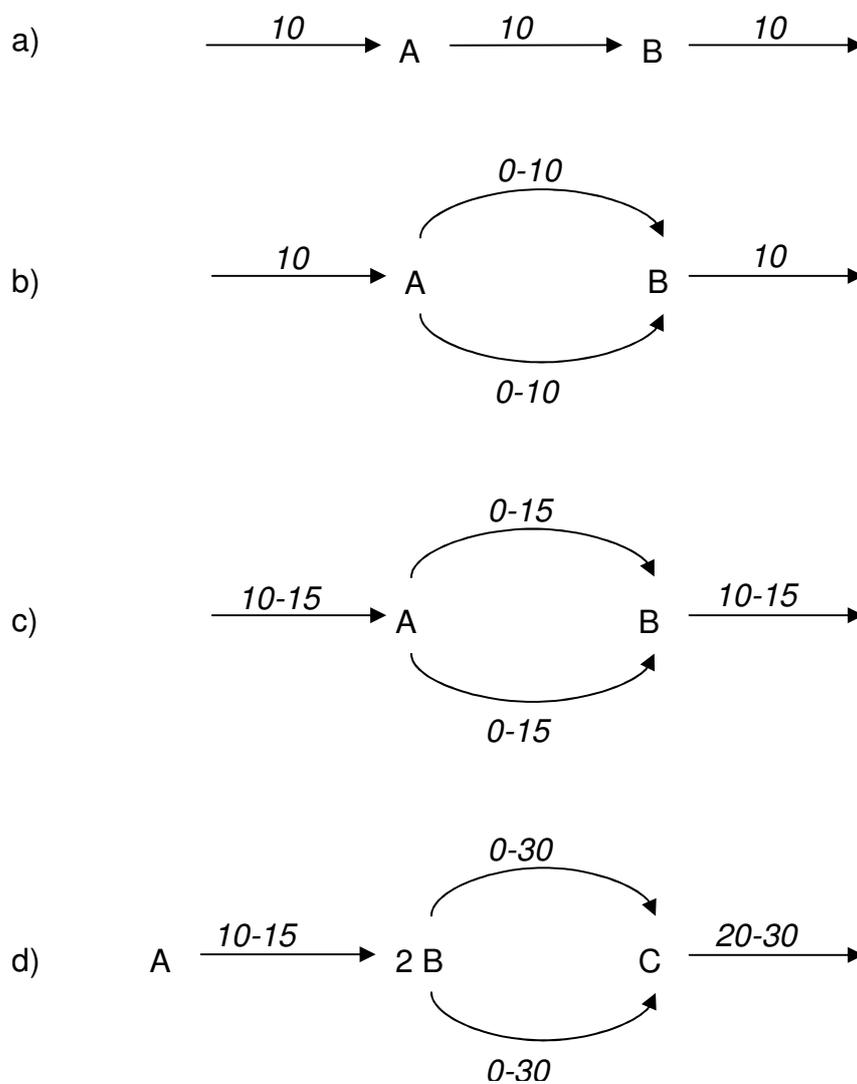
Figure 2.6: **How do variable fluxes occur ?**
a) The flux towards A is fixed, so all others are fixed, too. b) The flux towards A is fixed. Then it can divert, with the sum of both possibilities being 10, always. c) If the flux towards A is variable itself, this has to be added. d) Also stoichiometric coefficients have to be taken into account (A → 2·B).

$$\vec{\mathbf{0}} = \underline{\underline{\mathbf{C}}} \cdot \Delta\vec{\mathbf{G}}_{\mathbf{r}} \qquad \text{(first law)} \qquad (2.17)$$

$$0 \le v_i \cdot \Delta G_i \, , \; \forall i \qquad \text{(second law)} \qquad (2.18)$$

$\underline{\underline{\mathbf{C}}}$: full cycle basis of the system; $\Delta\vec{\mathbf{G}}_{\mathbf{r}}$: vector of Gibbs energies of reactions; v_i: flux of reaction i.

By combining these statements, it follows that no cyclic fluxes are allowed, because the equal sign in eq. 2.18 only holds true if $v_i = 0$. On his website (Beard, 2005) Beard offers programs for download that accomplish the inclusion of equations 2.17 and 2.18. Matlab (The MathWorks, 2005) is employed as programming environment. Due to the thermodynamic / energetic background of the theory, Beard *et al.* termed it *energy balance analysis* (*EBA*).

The step from *FBA* to *EBA* further constraints the feasible solution space of equation 2.14. But it also necessitates the employment of nonlinear programming, because the additional constraints cannot be stated in a linear way. As a consequence, the typical problems attached to nonlinear programs enter the analysis. In a nutshell: the optimization routines may get stuck at local optima, rather than finding the global optimum. Additionally, it can be hard to find out whether the denoted result in fact is a local or the global optimum. A possibility to overcome this obstacle is explained in the following chapter

2.7.1 *Ab initio* prediction of reaction directions

In (Yang *et al.*, 2005), Feng Yang introduced the possibility to calculate thermodynamically feasible reaction directions *ab initio*. This is done by fixing some necessary external fluxes (e. g. carbon source influx and biomass production rate) and, afterwards, successively minimizing and maximizing

each flux of the network under the constraints of *EBA*. If thereby a valid positive value for a flux is found, it is obvious that the forward direction is feasible. The opposite holds true for minimization and negative flux values. The point is: the optimum found does not have to be the global optimum. If only one positive value is found, it is enough to state that the forward direction is allowed. After determination of all feasible reaction directions, *linear* optimization is applied once more to find the *global* optimum of the systemic objective. The above delineated procedure was used in this work, as explained in 3.4.

2.8   Databanks

2.8.1   BRENDA

The *Braunschweiger Enzym Datenbank* (BRENDA) is the main collection of enzyme functional data available (Schomburg *et al.*, 2004). Started in 1987, BRENDA now covers around 4200 EC numbers representing more than 83000 different enzyme molecules. The information supplied covers enzyme nomenclature, interaction to ligands and information regarding those, functional parameters like K_M values and turnover numbers, organism related information, enzyme structure and molecular properties, as well as links to referring literature (compare (Schomburg, 1987)).

2.8.2   KEGG

The *Kyoto Encyclopedia of Genes and Genomes* (KEGG) is the second biological database used throughout the thesis at hand (Kanehisha & Goto, 2000). Contrary to BRENDA, it emphasizes the systemic point of view, while only covering basic information about enzymes and the reactions catalyzed by them. KEGG contains graphical depictions of metabolic pathways (Kanehisha, 2005), similar to the well known wall charts by Gerhard Michal (Gerhard, 1999). Additionally, it provides an unambiguous coding of all re-

actions and metabolites, which proves very helpful when processing the data with a computer. Compounds which can be addressed by multiple names in different reactions (which again may be addressed by varying names) are always addressed by the same identifier.

Chapter 3

Methods

All computational calculations described in this work have been carried out on a standard personal computer, Intel Pentium 4 CPU 2.4 GHz, 1 GB RAM.

As operating systems we used mainly the *mandriva* (formerly mandrake) linux distribution (Mandriva, 2003). Additionally, Microsoft Windows XP (Microsoft Corporation, 2003) was used when employing the company's office package.

Programming languages used throughout this work are C/C++ (Stroustrup, 1997), *Matlab* scripting (The MathWorks, 2005), and *Perl* (Wall *et al.*, 2000).

3.1 CMP

CMP is an acronym of *Cubic Metabolome Project*. This is a C/C++ programming library, which has been developed to solve problems related to biochemical networks. All programming work accomplished in C/C++ in this thesis is based on the CMP library. Interested persons can download the CMP source code under the following web address:

`http://www.biotool.uni-koeln.de`

3.1.1 CMP details

As shown in figure 3.1, the structure of CMP is hierarchic. Data is stored within CMP in a metabolic network. This network consists of reactions. Each reaction owns educts and products, which are stored as sets of molecules. Obviously, these consist of molecules, which are build of atoms. Each level is designed to store the appropriate data, as well as to provide functions necessary to handle this data. E.g., along with a reaction, information about its reversibility is stored and can be changed easily at any time. For detailed information see the above-mentioned website.
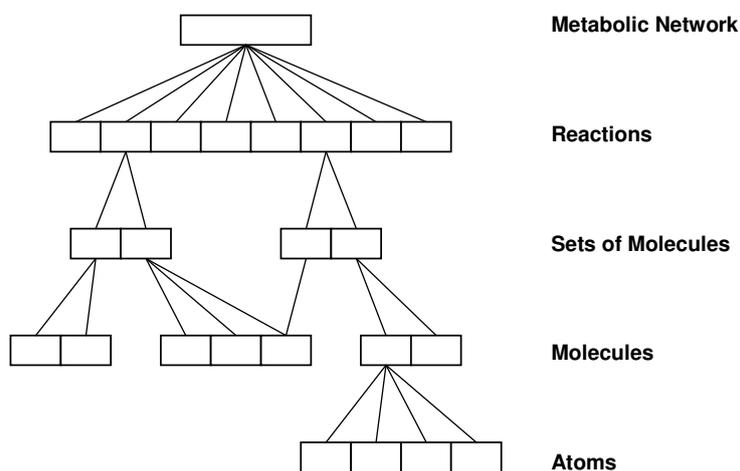
Figure 3.1: **CMP Structure**
The comprehensive structure of the CMP programming library. Each depicted level is implemented as a class of its own, supplied with appropriate functions and data-storage solutions.

3.2 The *C. glutamicum* model

In the following, the process of modeling the metabolism of *C. glutamicum* is explained. At first we will describe which assumptions and abstractions have to be made. Next, the actual procedure will be depicted, including the usage of DNA microarray data and the reasons for implementing reactions with non-annotated genes. At last some additional comments will be made.

3.2.1 Underlying assumptions

Basic genome-based modeling can be described very briefly as annotating the genome and looking up all possibly occurring reactions in databases like KEGG (Kanehisha, 1997) or BRENDA (Schomburg *et al.*, 2004) (compare 2.8). Therefrom, directly a metabolic model can be established. Doing so, one must be aware of some generalizations hidden in this procedure:

- if one is not using microarray data (compare 3.2.3), it is assumed that each coding gene is expressed

- and each gene expression product (i.e. protein) is available in sufficient amount, so enzyme saturation is not an issue (see also 3.6.2.2).

- also, each enzyme is available everywhere, at any time

- the same holds true for all metabolites

- usually, no enzyme specific properties are modeled (e. g. activation or inhibition; also compare 3.6.2.2)

Most of the time, compartmentation is neglected. This holds true for our model, since *C. glutamicum* is a prokaryotic bacterium. Another point, which shall be mentioned explicitly though being obvious, is: processes not modeled are considered to have no influence on the modeled part of the organism.

3.2.2 The modeling process

The *C. glutamicum* reaction network designed in this thesis is based on the bacterial genome. Though the annotation of the genome has been published by now (compare (Ikeda & Nakagawa, 2003) and (Kalinowski *et al.*, 2003)), we used an in-house annotation, performed between 2002 and 2004 by Dr. Urte Wendt. This is due to the fact that the publishing date of the annotations just cited lies after the date the model-building process took place. Later on, we included the published annotations during refinement processes.

The modeling process is depicted in the flowchart diagram 3.2 on page 31. We started with the genome annotation and connected the resulting list of EC-numbers to the KEGG database (Kanehisha, 1997; Kanehisha & Goto, 2000; Kanehisha, 2005). In constant alignment with literature (Takaç *et al.*, 1998) a list of reactions was achieved. Since the KEGG database is sometimes erroneous, a manual check for correctness of all reactions used in our network has been performed. This check included the examination of the reaction's mass-, redox-, and charge-balance.

At a later stage of the work, the set of reactions has been adapted to an assumed intracellular pH of 7. This has been performed by Kai Hartmann.

The refinement process, following the initial reaction network, included closing gaps in pathways as well as removing loose ends. To achieve the latter, either a reaction crossing the cellular membrane has been added or the reaction-sequence leading to the dead end has been removed. In this context, we tried to allow only metabolites to cross the membrane which are known to do so (i.e. sources for carbon, phosphate, nitrogen, or sulfur and oxygen or carbon dioxide). Additionally, biomass components are modeled to be able to leave the cell, but in fact they do not pass the cellular membrane but constitute it.

It has to be emphasized that the procedure of refinement and expansion has been an ongoing process throughout the whole work. At some stages,

only the flux analysis (see 2.6) of the system at hand revealed existing deficiencies. Even though the model construction chronologically has to come before the flux analysis, this work cannot be split into two parts, but has to be understood as consisting of two tasks, which are highly interconnected.

To learn which irreversibility information has initially been used, the reader is referred to section 3.4.2.

3.2.3 Using DNA microarray analyses data

For background theory to DNA microarray analysis see 2.4.

Within the research group of Prof. Dr. Reinhard Krämer at the Cologne University, Dr. Maike Silberbach *et al.* carried out DNA microarray analyses of the *Corynebacterium glutamicum*. Details about the methods applied are published in (Silberbach *et al.*, 2005; Silberbach, 2005). In a nutshell, Silberbach *et al.* used reverse transcription to transform RNA molecules into cDNA, which has subsequently been quantified by microarray analyses.
In the thesis at hand, we used this data to further improve and justify the *C. glutamicum* model constructed. Therefore, we calculated the *R-value*, that is defined in (Hüser *et al.*, 2003) as

$$R = \frac{\text{signal mean} - \text{background mean}}{\text{background standard deviation}} \tag{3.1}$$

and is suitable to discriminate between significant and nonsignificant transcription. Then we connected the expressed genes to enzymes, if possible. This was accomplished by using our annotation (compare 3.2.2, page 29) as well as the genome annotation published by Kalinowski *et al.* (Kalinowski *et al.*, 2003).
The reader should know, that Hüser *et al.* claimed *R-values* ≥ 2 to state a significant transcription.
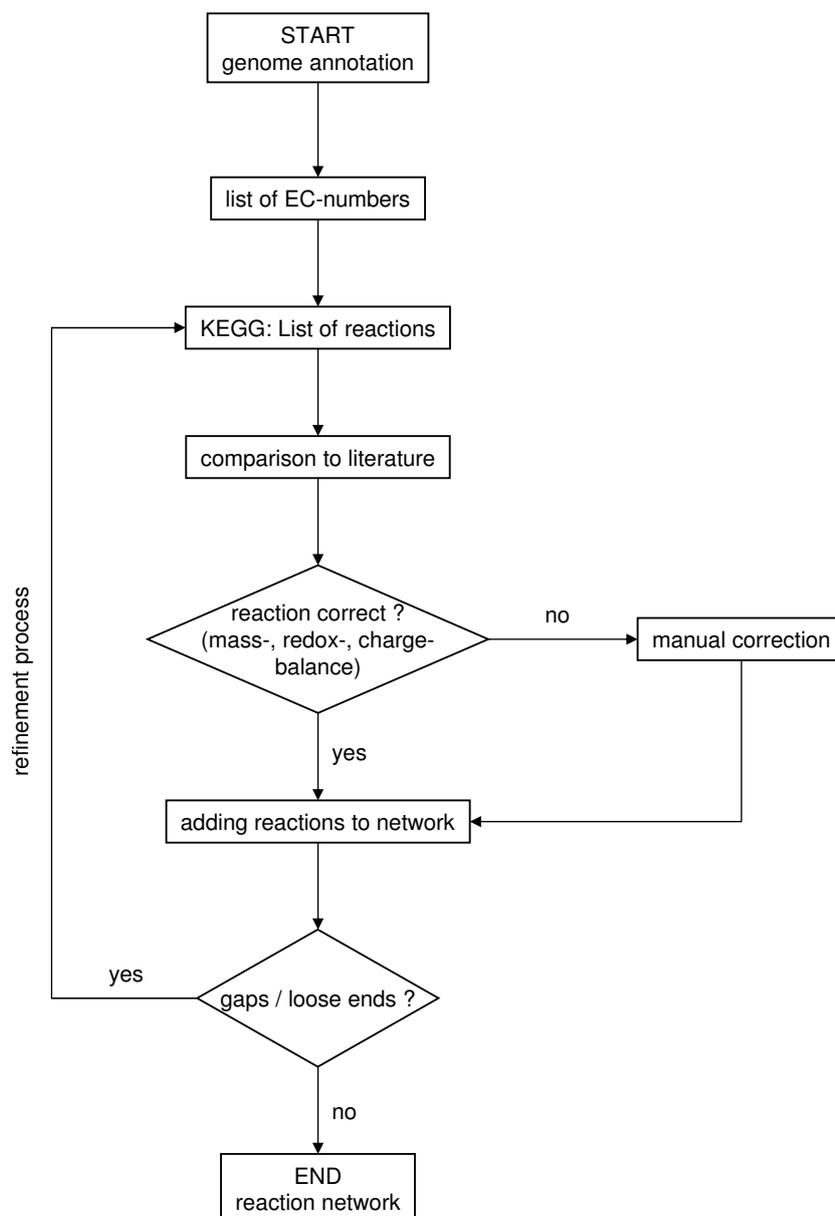
Figure 3.2: **Process of building the model**
Flowchart of the model building process. Genome annotation performed by Dr. Urte Wendt, (Ikeda & Nakagawa, 2003), and (Kalinowski *et al.*, 2003). All reactions have been manually checked for correctness in terms of mass-, redox-, and charge-balance. All reactions have later on been adapted to an intracellular pH of 7 by Kai Hartmann.
See the following references for KEGG: (Kanehisha, 1997; Kanehisha & Goto, 2000; Kanehisha, 2005) , Literature used in comparison: (Takaç *et al.*, 1998).

During the process of establishing the connections between a gene and an enzyme, the following redundancies could be observed:

1. there may be more than one gene coding the same enzyme.

2. there may be more than one enzyme annotated to one gene.

Since we only needed to know *if* a gene is expressed, and were not interested in the exact amount, for us it was sufficient to identify at least one gene that is expressed in a significant amount. Thus, the first-mentioned redundancy states no obstacle.

Concerning the second one, we had no means to rank the enzymes annotated to one gene. In order to act scientifically, it was not possible to decide manually which enzyme is coded by the referring gene. In consequence, we could either neglect the genes coding more than one possible enzyme or take all enzymes as expressed, if the gene is transcribed. We assumed all genes coded to be expressed, otherwise too many genes would have been omitted. But only those enzymes were added to the network that fitted into its structure. E.g. enzymes coding reactions which are not connected to the model were neglected.

3.2.4 Non-annotated reactions

During the process of model building, it was sometimes necessary to add reactions which are either not annotated or which additionally do not exist within KEGG (Kanehisha, 2005). Thereby, the non-annotated reactions can be divided in two types:

1. reactions which are catalyzed by an enzyme, but this enzyme is not annotated within the *C. glutamicum* genome.

2. reactions which are not attached to a catalyzing enzyme.

The reason to include reactions like those were

- to close gaps in obvious pathways

- to cross the cellular membrane

Most of the gap-closing reactions are isomeric transformations[1]. We only introduced transformations between conformational isomers or configurational isomers, if the latter can be achieved at low costs and with the help of water only. In these cases we assumed that no help of an enzyme is needed to accomplish the reaction. An example, very popular in this work, is the transformation between the open chain and closed ring form of sugars.

3.2.5   Biomass constitution

Basically, we used the biomass as defined in (Takaç *et al.*, 1998). But to satisfy the genome annotation (see 3.2.2, page 29 ff.) we had to neglect certain molecules. There was no hint, that pathways producing them exist in *Corynebacterium glutamicum*. Table 4.6 in the results section on page 71 shows the detailed composition of biomass as used in this work.

3.2.6   State of the metabolites

It has been mentioned before (3.2.2, page 29) that all molecules have been adapted to an intracellular pH of 7. The according process was accomplished manually by Kai Hartmann (Hartmann, 2006).

3.2.7   Storage of the *C. glutamicum* model

The model derived is stored within two text-files.
The main reaction file contains information about the enzymes catalyzing a reaction (2.3), the pathways the reaction belongs to, and the reaction equation itself. This file has the KEGG reaction file format (Goto *et al.*, 2004, section 2.1). We extended this format, in a consistent way, by an identifier

---

[1]Isomers are molecules which consist of the same atoms being arranged differently in space. For more details about the different kinds of isomeries, see e. g. (Hollemann & Wiberg, 1995, p. 322 ff.)

called METABOLISM, which is followed by the part of the metabolism the reaction belongs to in our model.

The second file used to store the network has the format of the KEGG *reaction_ main.lst*-file (Goto *et al.*, 2004, section 5.10). It only contains information about the reversibilities of the reactions used (see 3.4.2, too).

3.3   Flux Balance Analysis (*FBA*)

For details about the theory of flux balance analysis see chapter 2.6.

The implementation of the *FBA*, as performed in this work, is based on the *CMP* library (see 3.1).

To carry out the linear optimization, we used a C/C++ library from IBM called OSL, which is short for *Optimization Subroutine Library*. OSL was freely available for academic purposes and is a subroutine library able to solve linear optimization problems in an acceptable time (i.e. below one second for the current model (see 4.1)). Unfortunately, IBM has ended support for this product during the time of our work, but interested persons can have a look at the manual online at (IBM, 1995).

The data, necessary to carry out a flux balance analysis, are the ones given in the theory chapter (2.6). In a nutshell these are:

- the stoichiometric reaction equations

- information on the reactions' (ir)reversibilities

- information about bounds on fluxes. This includes external fluxes as well as internal ones (see 2.6, especially fig. 2.4).

The software package programmed to perform the flux balance analysis has to be given the information stated in the last item of the above list via an additional file. The format of the output files, containing the optimal flux

distributions, is defined by the optimization subroutine library.

Scripts in the *Perl* programming language (Wall *et al.*, 2000) have been written to transform these into various formats, e.g. to be read by Microsoft Excel (Microsoft Corporation, 2003) or the Open Office package (SUN Microsystems, 2004).

Most of the literature concerning flux balance analysis (e.g. (Schilling *et al.*, 2002; Edwards & Palsson, 1999, 2000)) describes the objective function as the production of biomass. In case of *E. coli*, it has also been shown that predictions based on this assumption are in accordance with experimental data (Edwards *et al.*, 2001). Using the biomass constitution described in 3.2.5, we will show that optimizing biomass allows us to reproduce experimental results in theory (compare results in chapter 4.2).

In this work, the simplex algorithm (Lawler, 2001, p. 43 ff.) was used to compute all linear optimizations.

3.4 Energy Balance Analysis (*EBA*)

For details about the theory of energy balance analysis see chapter 2.7.

The program to perform the *EBA* was written in the *Matlab* programming language (The MathWorks, 2005). We used and adapted the programs placed at disposal by Daniel Beard under (Beard, 2005).

As explained in 2.7 and 2.7.1, there are two ways of carrying out the energy balance analysis. One is to perform a nonlinear analysis which includes the necessary constraints (Beard *et al.*, 2002). The other is to calculate the feasible reaction directions in advance as explained in (Yang *et al.*, 2005).

We tried the first-mentioned method, but encountered severe problems in finding the optimum within the nonlinear optimization problem at hand (see 2.7). Thus, we switched to the *ab initio* prediction of thermodynam-

ically feasible reaction directions. Trying to use the whole matrix while calculating the complete cycle basis (see eq. 2.17 & 2.18, page 23) made the task uneconomic, because the calculation time easily exceeded days, without the computing getting finished. Since the directions of all outer reactions are known in advance, a shift to the prediction of only the internal reaction directions followed.

Calculation time was reduced to hours rather than days by this step.

This is what was done:

- extract the part of the stoichiometric matrix containing only the internal reactions

- calculate the full cycle basis (compare equation 2.17,p. 23) using the matlab *cycle* function supplied by Daniel Beard (Beard, 2005)

- use the optimization method described in (Yang *et al.*, 2005). We changed this method as described below (3.4.1).

- achieve a prediction of all reaction directions for thermodynamically feasible flux distributions

As a result of this procedure we obtained a new file, supplying constraints on all internal reactions, so the fluxes within the system were only allowed to distribute according to thermodynamic feasibility. As Feng Yang explains in (Yang *et al.*, 2005) and Daniel Beard also states on his website (Beard, 2005), one has to check whether there are any possible cycles left after the prediction of the flux directions.

In all analyses presented in this work under the label of *EBA*, no cycles were left after the initial prediction of the reaction directions (at this point see also 3.4.2, concerning the initially given reaction irreversibilities). Thus, thermodynamic feasibility was indeed ensured for all flux distributions gained under the constraints of *EBA*.

As for all flux balance analyses (see 3.3), we used biomass (compare 3.2.5) production as the objective function and the simplex algorithm (Lawler, 2001, p. 43 ff.) to compute all linear optimization problems. To compute the nonlinear optimizations the *Matlab* (The MathWorks, 2005) function `fmincon` was used.

3.4.1 Applied changes to the optimization method

Yang *et al.* suggest to impose all necessary constraints on a network and then to minimize and maximize each flux within, in order to achieve information about the thermodynamically reasonable reaction directions (compare (Yang *et al.*, 2005)). As explained in 2.7.1, this procedure does not need to find the global optimum of the nonlinear optimization problem at hand, but it is sufficient to reach a point other than zero to state that a correspondingly signed flux exists.

After implementing this method, we found the predicted reaction directions differing slightly between varying calculations. This is caused by the optimizer sometimes getting stuck on a negative value, regarding it as the optimum even though a positive optimum exists. Since we tried random initial guesses (starting points) for the optimizer until he found *any* optimum, sometimes a negative (local) optimum was found, sometimes a positive was.

Our changes were not only to analyze the reaction currently optimized, but to do so for the whole achieved flux vector. It might happen, while optimizing reaction x, a positive value was found for reaction y. Since the whole flux vector has to be reasonable in a thermodynamic sense, obviously there are solutions containing a positive flux through reaction y. But, while maximizing v_y explicitly, a negative optimum was found, suggesting no positive flux exists for reaction y. We figured this to be an effect of the optimizer getting stuck in a local (negative) optimum. Thus, we considered the whole flux vector at each optimization. Additionally, sometimes a flux is predicted

to equal zero. This happens if minimization as well as maximization result in zero as the particular optimal value. In these cases, we asked the result to be confirmed 20 times, meaning we tried 20 different, randomly chosen, starting points for the very same optimization. That brute approach showed to be necessary after a flux was predicted to be zero which was lying within the central carbon metabolism.

The whole process is included in the flowchart 3.4 on page 41.

3.4.2 Identification of initial irreversibilities

As shown in figure 3.4, we started with a network consisting only of reversible reactions. It was tried next to predict all feasible reaction directions from scratch. As expected, this did not result in a network with totally feasible reaction directions. The check whether there are still cycles left (compare 3.4) turned out positive. Now, the reaction which was most common in the existing cycles was identified. A search in the KEGG database (Kanehisha, 2005) was performed, retrieving eventually existing information about this reaction's irreversibility. If such information was found, it was implemented and a new *ab initio* prediction of the feasible reaction directions was computed. If no such information was found, the next most common reaction was identified and used. This procedure was repeated until there were no more cycles left after the flux direction prediction.

This way, we ensured to use as few initial information as possible in the most effective way possible. There have been a few exceptions to the method described. Details about that can be found in section 3.4.3.

3.4.3 Special cases

The reactions listed in table 3.1 have been added to the network as a final step of the model building process. Integrating them made the calculation of the full cycle basis (see eq. 2.17, p. 23) too costly, concerning calculation time. Including additional irreversibility information did not help here,

Table: 3.1: **Reactions not included in irreversibility prediction**
The following reactions have not been included during the prediction of the thermodynamically feasible reaction directions. They have been added afterwards with the irreversibilities shown. See text (3.4.3) for details.

reaction 134	ATP + Glycerol => ADP + sn-Glycerol 3-phosphate
reaction 135	Glycerol + NADP+ <=> D-Glyceraldehyde + NADPH + H+
reaction 137	D-Glyceraldehyde + NAD+ + H2O <=> D-Glycerate + NADH + 2 H+
reaction 136	ATP + D-Glycerate <=> ADP + 3-Phospho-D-glycerate

because the algorithm used (Beard, 2005) calculates the cycle basis from the internal stoichiometric matrix (compare 2.6, especially fig. 2.4,p. 19) without taking the irreversibility information into account. Irreversibilities are added later, while performing the optimization method explained in (Yang *et al.*, 2005).

As a consequence, we followed the chain of arguments stated next and integrated the reactions without any further test if the energy balance analysis constraints are fulfilled:

1. before adding the above mentioned reactions, sn-Glycerol 3-phosphate could only be produced and was directly converted to biomass afterwards.

2. if the reactions 134, 135, 137 and 136 are added, including irreversibility information which only allows another way of *producing* sn-Glycerol 3-phosphate, no additional cycles can occur.

3. KEGG database (Kanehisha, 2005) says reaction 134 is irreversible in the according way.

Figure 3.3 depicts the situation.
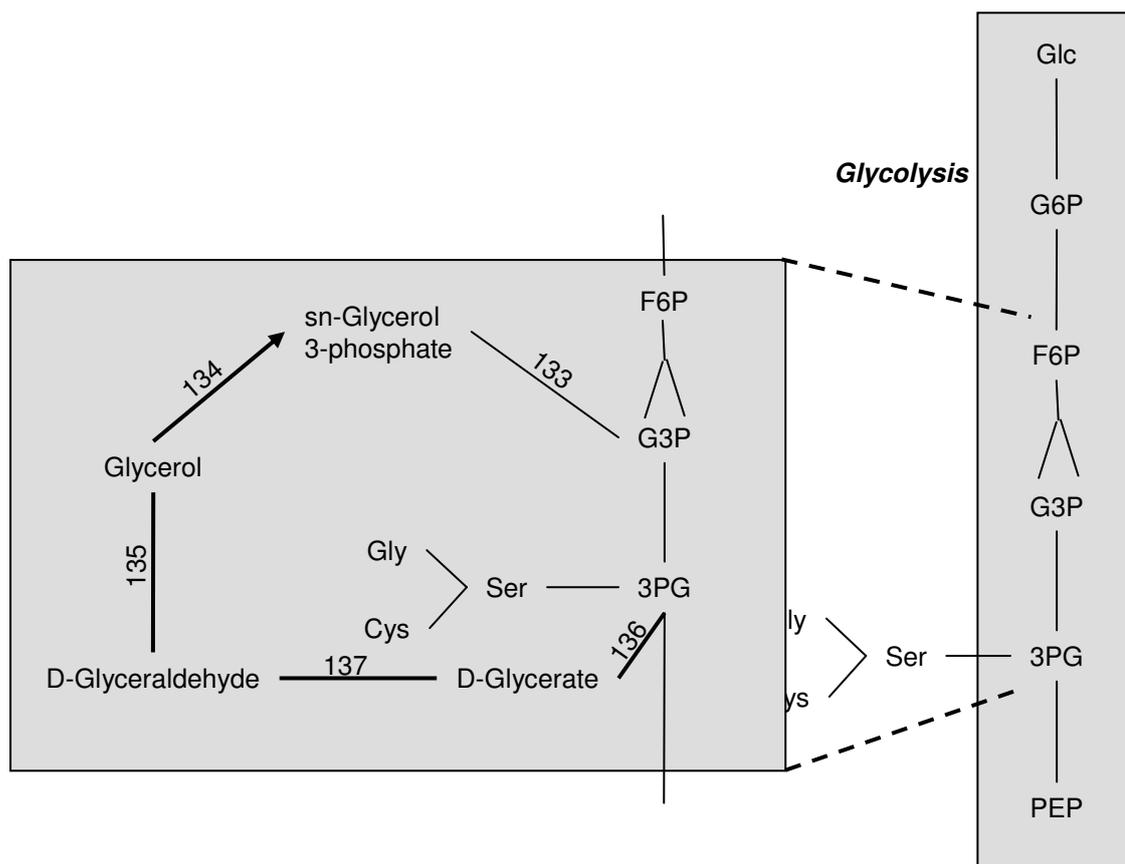
Figure 3.3: **Reactions not included in irreversibility prediction**
The *EBA* predicted reaction 133 to be irreversible towards sn-Glycerol 3-phosphate. The reactions 136,137,135 and 134 (bold lines or arrows) have not been included during the prediction of thermodynamically feasible reaction directions (see table 3.1). They have been added afterwards with irreversibilities (arrows) shown.
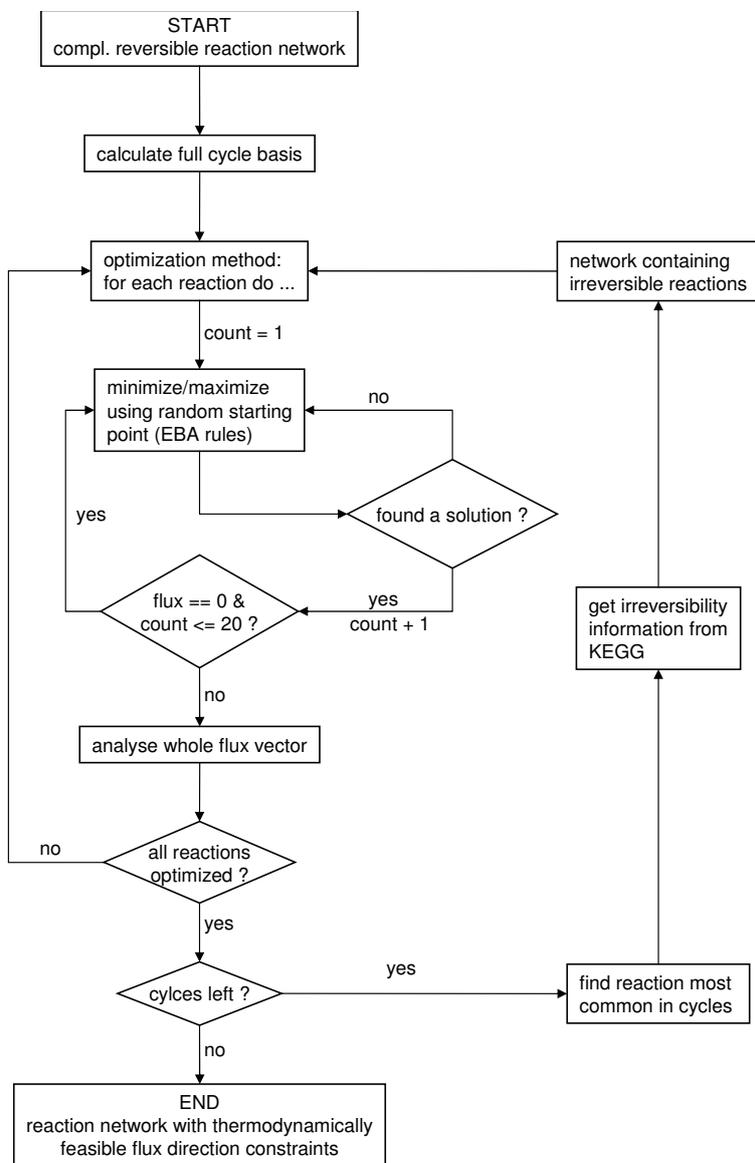
Figure 3.4: **EBA flowchart**
A flowchart diagram of the energy balance analysis process. For the theory of *EBA* see 2.7. The procedure shown includes matlab code as well as C++ code based on CMP (3.1)

3.5 External Energy Balance (*EEB*)

The idea behind developing the external energy balance is the following: besides a thermodynamic feasible arrangement of the fluxes within the simulated network, one has to analyze thoroughly the energy balance spanning the cellular system as a whole.

This is a simple consequence of the first law of thermodynamics (see 2.1.1.2), which states that energy can neither vanish nor appear spontaneously. But since a living organism dissipates energy as heat, and we are not able to measure its amount, the condition that has to be fulfilled is:
the amount of energy *leaving* the system as metabolites must be smaller than or equal to the amount of energy *entering* the system in metabolic form.

$$\sum_{i=in} \Delta G_i * |v_i| \geq \sum_{j=out} \Delta G_j * |v_j| \tag{3.2}$$

v_i and v_j are the reaction velocities of the incoming and outgoing reactions, respectively.

The *greater*-part of the 'greater or equal' sign refers to the energy dissipated by the system. The value of the free enthalpy of formation ΔG_f is defined as the free enthalpy of a reaction building a molecule from the elements in standard state (compare table 2.1). The ΔG_f values have been calculated as described in the theory chapter 2.1.2.1. The software accomplishing this task has been developed by Kai Hartmann.
After imposing the above equation to the system, we introduced the next step -"*EEB* - Analysis"- which is explained in the following section.

3.5.1 *EEB*- Analysis

A living system is not able to metabolize all entering molecules into other molecules, which are leaving, without loss of energy. For a reaction to

actually occur, there has to be a drop in energy from the educts to the
products. As a consequence of this, in combination with the second law of
thermodynamics (section 2.1.1.2), a living system dissipates energy as heat.
Since we can calculate the energy entering or leaving the system, by summing
up each side of equation 3.2, we know the amount of energy consumed by
the metabolism. The task at hand now is to relate this amount to the model
analyzed. Stated different: how much energy could the bacterium gain from
the sources it consumes ?

The reader should be aware, that this is not the total amount contained
within the incoming chemical molecules. This would mean, that the organ-
ism digests the molecules down to elemental level, and this does not happen.
But what is the maximum energy "hidden" within the compounds used ?

We suggest, that the total conversion of all carbon entering the cell into
carbon dioxide leaving it frees the maximum possible amount of energy.

Thus, in order to attach a biologically relevant meaning to the amount of
heat *actually* dissipated, one has to relate it to the amount of heat dissi-
pated if the organism completely transfers the available carbon into carbon
dioxide. The latter has been computed by simply subtracting the left side
of equation 3.2 from it's right side, after optimizing the system on carbon
dioxide production. It is obvious, that one has to use the same amount of
carbon flowing into the cell in both cases.

In a nutshell:

1. optimize the model on carbon dioxide production

2. calculate the amount of heat dissipated (HD) as:

$$HD_{CO_2} = \sum_{j=out} \Delta G_j * |v_j| - \sum_{i=in} \Delta G_i * |v_i|$$

3. optimize the model on another objective (usually biomass production (compare 2.6))

4. again calculate the amount of heat dissipated (HD) as:

$$HD_{Biomass} = \sum_{j=out} \Delta G_j * |v_j| - \sum_{i=in} \Delta G_i * |v_i|$$

5. relate $HD_{Biomass}$ to HD_{CO_2}

3.6 Metabolic profiling

The procedure of metabolic profiling experiments has been explained in 2.5. The following sections show the core idea of this thesis. We will explain the developed method of implementing the profiling-data into the *FBA/ EBA*-analyses. This contains the underlying idea, the method of implementation, and the procedure to choose the correct metabolic profiles to use.

3.6.1 Metabolic profiles and reaction fluxes

The following sentence states the basic idea of this work:

The higher the concentration of a reaction's substrate, the faster the reaction.

Readers educated in chemistry will note two things:

1. due to greatly varying kinetic constants between different reactions (compare eqs. 2.10 & 2.11) this only holds true if comparing the same reaction under different circumstances

2. the reaction's velocity depends on the amounts of educts as well as of products (2.3)

Because of this and of some arguments founded in biochemistry we had to make the assumptions outlined in the following section.

3.6.2 Prerequisites and assumptions

3.6.2.1 About metabolic concentrations

Up to the time this thesis has been written, the determination of the metabolic profiles (see 2.5) has not been complete. There have been some metabolites our profiling team was able to measure, but not yet to identify. Other metabolites could not be measured for various reasons (2.5.1).

Thus, we made the following assumption:

All metabolites whose concentration could not be determined are assumed to have the same concentration under all circumstances taken into account.

The results (chapter 4) show this assumption to be reasonable.

3.6.2.2 About (enzyme-) kinetics

We assume the network's reactions to occur at conditions that approximate chemical rather than enzyme kinetics. In other words: *enzyme kinetics are linear with respect to each metabolite.* This means :

- the order of the reactions is one in each metabolite (i.e. the stoichiometric coefficient, if that metabolite appears more than once in the reaction)

- enzyme expression is no limiting factor and is the same in all networks considered.

- enzyme kinetics is far from saturation. E.g. for Michaelis-Menten kinetics this means $[S] \ll K_M \Rightarrow v \approx \frac{V_{max}[S]}{K_M}$ (see 2.3)

- we do not consider any effects changing the reactions' kinetics, such as inhibition, activation, enzyme-disintegration, or others

If the reaction rate is linearly dependent on the substrate concentrations

$$v = k * \prod_i [i]^{\nu_i} \quad | \quad \nu_i = 1 \quad \forall \, i \tag{3.3}$$

k: kinetic constant, $[i]$: concentration of metabolite i, ν_i: order of reaction in i

it is easy to understand that the (forward) flux grows with growing substrate amount. But, as said in 3.6.1, considering this is not sufficient. The flux of the backward reaction cannot be neglected, because its change may exceed the change of the forward reaction's flux.

An example may further illuminate the matter: suppose a reaction the forward flux of grows from 10 to 15 $\frac{\text{mmol}}{\text{g}(DW)\cdot\text{h}}$ and the flux in backward direction from 5 to 12 $\frac{\text{mmol}}{\text{g}(DW)\cdot\text{h}}$; then the overall reaction velocity (see eq. 2.9) *decreases* even though the forward flux increases.

Compare the same arbitrary reaction '$A \rightleftharpoons B$' under two different circumstances and call it R and R', respectively. It shall be, that the concentration of A is higher for R than for R': $[A] > [A']$. Then obviously the flux v for reaction R is greater than v' for R'. For this to hold true, while additionally considering the backward flux, we must compare three cases:

1. $[B] < [B']$

2. $[B] = [B']$

3. $[B] > [B']$

Case 1: the overall reaction flux increases from R' to R, because the forward flux increases while the backward flux decreases

Case 2: the overall flux increases, too. The forward flux increases while the backward flux stays the same. This is the case e. g. for all reactions of which we do not know the product concentrations. (In 3.6.2.1 we said all unknown concentrations are assumed to be constant.)

Case 3: this is the more complicated case, because the increase of the

backward flux may outbalance the increase of the forward flux. In order for the overall velocity to increase, the following condition must hold true:

$$\frac{[A]}{[A']} > \frac{[B]}{[B']} \qquad (3.4)$$

For *any* reaction this transforms into

$$\frac{\prod_i [i]^{\nu_i}}{\prod_{i'} [i']^{\nu_{i'}}} > \frac{\prod_j [j]^{\nu_j}}{\prod_{j'} [j']^{\nu_{j'}}} \qquad (3.5)$$

with i denoting all educts and j all products of the reaction (ν are the orders of the reactions in the referring metabolites).

The interested reader can find the mathematical proof, that equation 3.5 is sufficient to ensure a growing overall flux, in appendix C.

3.6.3 Method of implementation

To be able to neglect kinetic constants, the same reaction under alternate circumstances needs to be compared (see 3.6.1). In CUBIC laboratories this is realized by raising *Corynebacterium glutamicum* using different carbon sources. These feeding conditions have then been reproduced in the work at hand. Namely, we simulate, analyze, and compare the *C. glutamicum* metabolism while being raised on

- acetate (*NaAc*) and

- glucose (*Glc*)

In the following, we will add the shortcuts *NaAc* and *Glc* as indices to e.g. metabolic concentrations to indicate the conditions used during the

determination of the referring values.

If the concentration of a reaction's substrate $[\text{S}]_{Glc}$ is greater than $[\text{S}]_{NaAc}$, with all assumptions stated in 3.6.2 being fulfilled, we know from section 3.6.1 that the flux through the reaction following the substrate is greater for Glc than for $NaAc$ conditions: $v_{Glc} > v_{NaAc}$.

In order to improve the results available from a flux balance analysis or energy balance analysis, we introduce the above condition into the optimization process. This is done by optimizing both systems (under Glc or $NaAc$ conditions) at the same time and simultaneously imposing the new constraint.

The task of optimizing the systems in parallel was accomplished by extending the stoichiometric matrix (see fig. 2.4, p. 19) in a way shown below.

Let S_{Glc} be the matrix for the glucose-fed organism and S_{NaAc} the one for acetate conditions. They are combined to S_c, as:

$$S_c = \begin{pmatrix} S_{NaAc} & \cdots & 0 & \cdots \\ \vdots & \ddots & \vdots & \ddots \\ 0 & \cdots & S_{Glc} & \cdots \\ \vdots & \ddots & \vdots & \ddots \end{pmatrix} \tag{3.6}$$

To add new constraints of the type $v_{NaAc} > v_{Glc}$ (v being the same reaction in both cases) we appended additional rows to S_c. These rows only carried entries in the columns referring to v and stated a "greater than" equation.

In terms of reactions and metabolites, we introduced an artificial metabolite to the system, which built the link between the two parts of it (refer to chapter 2.6 about the set-up of the stoichiometric matrix).

It has to be noted, that the range of the method outlined above does not stop with the combination of two cases (for us: *NaAc* and *Glc*). The same procedure can be extended to implement data from three or more different data sets. Also, it is not necessary that the partial matrices look exactly the same. The only prerequisite that must be fulfilled is, that the reactions which are used to add new constraints are identical. This is because only then the kinetic constants can be reduced, as said in section 3.6.1.

3.6.3.1   Notes about the method of optimization

While optimizing two matrices, which are not connected, at the same time, the combined objective function (see 2.6) can simply be the sum of the two partial ones. But, if the matrices are connected, we may run into problems. The newly integrated, combinatorial constraint can make the solution space become concave (compare chapter 2.6, especially figure 2.5 on page 20). In that case, the optimization will reach a point where the algorithm would have to make a decision: point d in figure 3.5. The optimal solution in the concave space lies either at b or c. Since both are totally equal regarding the summed objectives, merely chance determines which result will be shown.

Biologically, there is no reason that both parts of the system must own a maximized objective simultaneously. It suffices that one flux distribution is optimized while the other is merely feasible and the combinatorial constraints are satisfied. Thus, we optimized one part of the matrix after the other, including the new constraints in all runs.

3.6.4   Selecting the metabolic profiles

The reader should be aware of the types of data obtained from metabolic profiling experiments before proceeding this chapter. Information on this can be found in section 2.5.
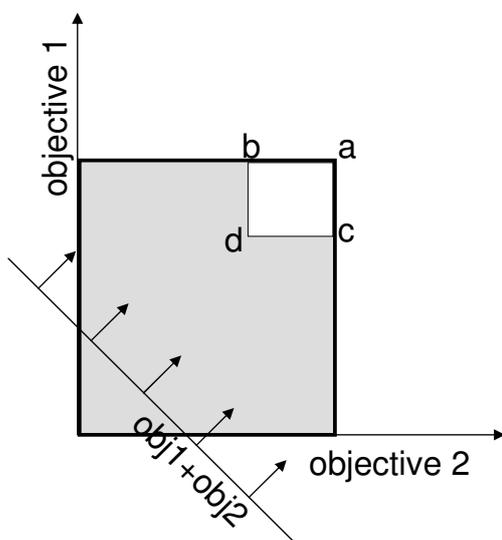
Figure 3.5: **Convex and concave space**
Solid line = convex, gray area = concave space. The task be, to optimize *objective 1 + objective 2*. Within the convex space, the optimum lies at *a*. Within the concave space the optimum can either be at *b* or *c*.

In chapter 3.3 (page 35), it was said that the *C. glutamicum* model is optimized on biomass production. To connect this to experimental results, we must identify data that has been received from organisms actually producing the maximum possible biomass.

For bacteria grown in a fermenter, like described in 2.5.1, this refers to the period of exponential growth. Thus, the first assignment was to identify the phases of exponential growth. After this, the data had to be normalized. Last but not least, we needed to identify the metabolic profiles that showed the highest correlation to each other (*NaAc* vs. *Glc*).

3.6.4.1   Identification of the period of exponential growth

The first step during the identification of the period of exponential growth
was to set inner and outer bounds on the possible range. To explain this, we
will anticipate a figure from the results section (4.3.1).
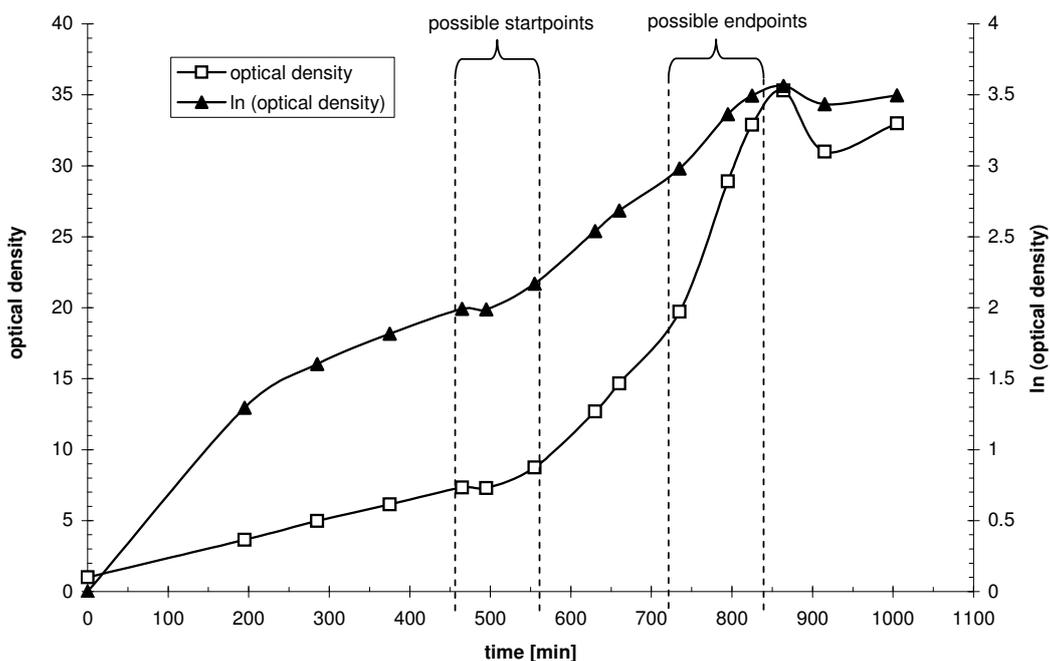
Figure 3.6: **Growth curve for glucose fed *C. glutamicum* (1)**
Boxed line = optical density, triangulared line = natural logarithm of OD.
Dashed areas show start and end ranges of the period of exponential growth.

In the above figure, the dashed lines enclose possible ranges for the start-
and endpoint of the period of exponential growth. Within these ranges, for
all possible combinations of data points:

1. from the curve ln(optical density) over time take one point from the
   start- and end-range each. Thereby, take care that there is at least one
   data point between the start- and endpoint.

51

2. calculate the square of the correlation coefficient as defined by Pearson[2].

The combination of start- and endpoint that gave the highest squared Pearson Coefficient was chosen.

We aimed for the identification of the phase of *exponential* growth. This means, the natural logarithm of the chosen data should have a linear dependence on time during this period. Thus, we tried to find the particular start- and endpoint, that allow the best possible linear regression on the data in between.

It showed to be necessary to define possible start- and end-ranges manually, because sometimes the linear regression improved by adding points which definitely lie beyond the period of exponential growth (like e. g. (31; 915 min) in figure 3.6).

Obviously, the linear regression of a set of two neighboring data points would be perfect. In consequence, the determined phase of exponential growth must at least include three data points.

3.6.4.2 Normalizations

What may be called a *data set* in the following, refers to a metabolic profile of all metabolites at one point in time.

The normalization included two steps, which will be explained in detail next:

- adjust all data sets to include the same data only

- normalize data sets to the same total amount of metabolites

Adjust data sets to include the same data
Of course, this is not to be understood as: "the data sets should not differ". The point is, not all metabolites could be detected in all experiments

---

[2] The squared Pearson correlation coefficient r_p^2 (compare eq. 3.8, p. 53) is the usual means to quantify the quality of a linear regression (the closer to one, the better). See e.g. (Koehler *et al.*, 2002) for further details.

(see 2.5.1). What we needed at the end of the day was a relation between the concentrations of the same metabolite for different environmental conditions. Thus, this metabolite must occur in all data sets. Also, the normalization explained next would not make sense if we had considered metabolites only occurring in some profiles.

Normalize data sets to the same total amount of metabolites

$$\sum_x [x]^i = \sum_x [x]^j = \sum_x [x]^k = \ldots \tag{3.7}$$

x = metabolites, i,j,k,... = profiles

The biological idea behind this is that a cell only has a limited space available. Thus, it is a reasonable assumption that the total amount of metabolites is the same for all cells.

3.6.4.3   Identifying the highest correlating profiles

The means used to identify the highest correlating profiles has been the *Pearson correlation coefficient*:

$$r_p = \frac{\sum(x_i - \overline{x}) \sum(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \cdot \sum(y_i - \overline{y})^2}} \tag{3.8}$$

We calculated the coefficient for all *NaAc*-profiles against all *Glc*-profiles. The combination showing the maximal correlation has been used further on. In terms of systems biology, it is reasonable to compare only those systems which are most similar. Only then, a meaning can be attached to the differences between those systems. For systems being highly diverse from the very beginning, the single difference has less impact.

3.6.5 Selecting the metabolites

After having defined the metabolic profiles that shall be used, the distinct metabolites need to be selected. If one aims for the deduction of new constraints from the concentration ratio of one metabolite in different systems, the ratio has to have a significant size. What exactly *significant* means, has to be defined by the scientist. In case of this thesis, a ratio of

$$\frac{[S]_{NaAc}}{[S]_{Glc}} \geq 5$$

$$\text{or} \tag{3.9}$$

$$\frac{[S]_{NaAc}}{[S]_{Glc}} \leq 0.2$$

was considered to be significant.

Additionally, two other conditions had to be fulfilled. One was stated in equation 3.5 on page 47 and concerned the ratio of the reaction's products and thus the backward reaction. Secondly, the reaction using S as a substrate definitely needs to have a major flux in forward direction. If this is not given, we do not know if the molecule S, we are looking at, is in fact a substrate or a product. Since the result of the whole process is a constraint of the type $v_{NaAc} \geq v_{Glc}$, we cannot deduce such a statement, if we do not know that both reactions (under NaAc or Glc feeding) show the same direction. This can be verified by performing the robustness analysis explained in the following section (3.7). Only if network topology and thermodynamics constricted the flux through the referring reaction to be forward, the possible combinatorial constraint was imposed on the system.

The whole process of choosing the profiles, normalization, and metabolite-selection is shown in figure 3.7.
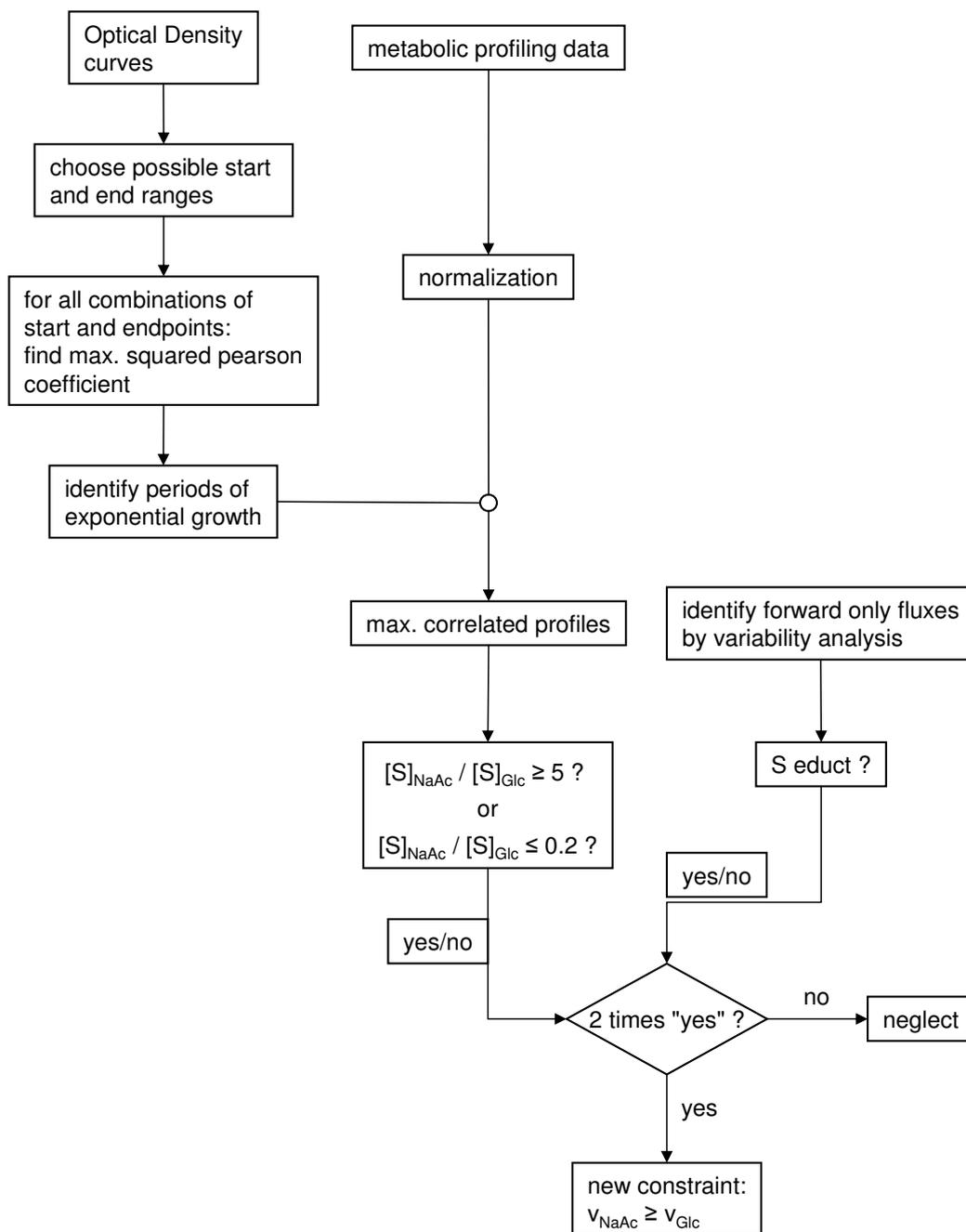
Figure 3.7: **Process of identification and implementation of new combinatorial constraints**
The above figure shows the process of identifying and implementing new combinatorial constraints. The procedure is explained in detail throughout section 3.6.

3.7   Robustness Analysis

In chapter 2.6 it was explained that a flux balance analysis probably never reaches one distinct solution. When analyzing a model of a living organism, we will always end up with a solution space. But, how to estimate the size and shape of this solution space?

In this work, we used a robustness analysis, which is the opposite of a sensitivity analysis. The procedure is simple:

- optimize the system on biomass production

- introduce a new constraint: biomass production be at least $90\,\%$ of its maximal possible rate

- while keeping the carbon influx constant, minimize and maximize each flux successively

The outcome are the intervals each flux can vary in while biomass production levels at its maximum. Thus, this data can be interpreted as showing the robustness of the metabolic network against changes in the single fluxes.

3.8   Imposed constraints

3.8.1   Biomass / CO_2 ratio

Preliminary analyses during the process of model-building showed that it is not sufficient to constrain the carbon influx and maximize the biomass production rate to achieve biologically reasonable results. This leads to a nearly complete transformation of all carbon into biomass. The carbon dioxide production, which was found to be approximately one third of the incoming carbon in experiments, nearly drops to zero in the simulation.

As a consequence, we adopted the ratio between the biomass and CO_2 efflux from literature (Wendisch *et al.*, 2000). Thereby, this ratio was assumed to stay the same, independent of the actual flux distribution. Also, slightly

different ratios for glucose respectively acetate fed organisms were used. See above cited article and tables 4.7, 4.9, and 4.11 on pages 72, 80, and 83 for details.

3.8.2 Carbon related constraints

3.8.2.1 The biomass production rate

In (Schlegel, 1985, p. 195) the definition of the growth rate μ is given as

$$\mu = \frac{\ln \text{OD} - \ln \text{OD}_0}{t - t_0} \left[\frac{1}{\text{h}}\right] \tag{3.10}$$

To obtain μ from experiments, the period of exponential growth is determined as described in section 3.6.4.1. Following, the growth rate is calculated, using the start- and endpoint data in the above definition.
Turning to the simulation, the efflux of biomass can be calculated from equation 3.10:

$$v_{BM} = \mu \cdot \frac{1}{\text{M}(BM)} \left[\frac{1}{\text{h}\frac{\text{g}(DryWeight)}{\text{mol}(BM)}}\right] \tag{3.11}$$

Thereby, $\text{M}(BM)$ denotes the molar mass of biomass, which can be derived as the sum of all biomass components times their respective stoichiometric factors.

$$\text{M}(BM) = \sum_i \nu_i \, \text{M}(i) \tag{3.12}$$

i: biomass components, ν: stoichiometric factor

(see reaction 145 in appendix A)

3.8.2.2 How to calculate carbon related constraints

The procedure to determine the carbon related constraints was the same for all analyses, though the exact values differed. To define all constraints connected to carbon fluxes, three steps have been undertaken:

1. calculate the (experimental) biomass production rate and define the biomass efflux accordingly (see 3.8.2.1)

2. by using the biomass / CO_2 ratio, mentioned in section 3.8.1, calculate the CO_2 efflux

3. since biomass and carbon dioxide are the only ways for carbon to flow out of the system, calculate the necessary influx of the carbon source as

$$v_C^{in} = x_{BM} \frac{\text{mmol}(C)}{\text{mmol}(BM)} \cdot v_{BM}^{out} \frac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}}$$
$$+ x_{CO_2} \frac{\text{mmol}(C)}{\text{mmol}(CO_2)} \cdot v_{CO_2}^{out} \frac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}} \quad (3.13)$$

v denotes a flux and x the content of carbon in either biomass (BM) or CO_2.

To calculate the influx of the actual carbon source, the reciprocal of its carbon content has to be multiplied with v_C^{in}. As said before (3.7), during all analyses the carbon source influx was kept constant, while the biomass efflux had to reach at least 90 % of its maximum possible value. Also the CO_2 efflux has not been fixed, but was given a lower bound.

3.8.3 Non-carbon related constraints

The non-carbon metabolites which were allowed to cross the cellular membrane are listed in table 3.2. The bounds imposed on the referring transmembrane reactions are given, too.

Table: 3.2: **Non-carbon transmembrane fluxes**
The bounds shown refer to *outgoing* reaction directions.

Metabolite	lower bound	upper bound
Water	$-\infty$	∞
Proton (H^+)	$-\infty$	∞
Ammonia	$-\infty$	0
Oxygen	$-\infty$	0
Phosphate	$-\infty$	0

3.9   Model validation

3.9.1   Preceding adaptions

The model, build in this thesis, was validated by reproduction of experimentally derived flux distributions (Wendisch *et al.*, 2000). To be able to do this, we had to perform some calculations in order to adapt the flux units. Based on (Takaç *et al.*, 1998) we used

$$\left[\frac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}} \right] \tag{3.14}$$

as flux units. In contrast, the flux units given in (Wendisch *et al.*, 2000) are:

$$\left[\frac{\text{nmol}}{\text{mg}(Protein)\cdot\text{min}} \right] \tag{3.15}$$

Additionally, Wendisch *et al.* used a factor of

$$0.5\,\frac{\text{g}(Protein)}{\text{g}(DryWeight)} \tag{3.16}$$

for their calculations (compare (Marx *et al.*, 1996)).

The figures 4.3 and 4.4 in section 4.2 (p. 72 ff.) show the fluxes measured by Wendisch and his colleagues after adaption to our flux units (equation 3.14).

3.9.2 Method and required accuracy

It was impractical to try to come as close to the experimental values as possible, because problems, similar to those shown in figure 3.5 on page 50, arose. In a nutshell, due to joint minimization there exist several minima, one of which would need to be chosen. While minimizing the difference between the predicted fluxes and the published ones, eventually one reaches a point where a deviation-reduction of one flux entails a deviation-increase of another flux. To decide which way to take, a weight would have to be attached to these fluxes. But, there is no biochemical reason to rank one reaction more important than any other. Thus, we set the same limit for all measured fluxes: the reproduction of the reaction rates determined by Wendisch *et al.* had to be accomplished with an accuracy of less than 30 % deviation. The method to test this shall be described next:

1. maximize biomass production, while disregarding experimental values

2. impose constraints on all fluxes that have been measured by Wendisch and coworkers:

   - lower bound = $v_i(measured) \cdot 70\%$
   - upper bound = $v_i(measured) \cdot 130\%$

3. maximize biomass production again and see if its value changed

We considered the ability to produce the same amount of biomass under the new constraints (item 2) to be sufficient to call our model validated.

3.10 Special analyses methods

3.10.1 Estimation of the flux space size

Flux balance analysis does not leave us with a single solution, but with a solution space (compare chapter 2.6). When performing a robustness analysis, as described in 3.7, we get a first glance at the solution space of the problem at hand. Another, less detailed view of the solution space is explained in the following.

If we look at two fluxes, which can vary like shown in figure 3.8, we do know nothing about the connection between those fluxes. It may be, that '$v_1 = v_2$' always has to be satisfied. That would leave the part of line a that lies inside the dotted box as the solution space. It may also be, that the two fluxes v_1 and v_2 are totally independent. In this case, the solution space would be $\Delta v_1 \cdot \Delta v_2$.
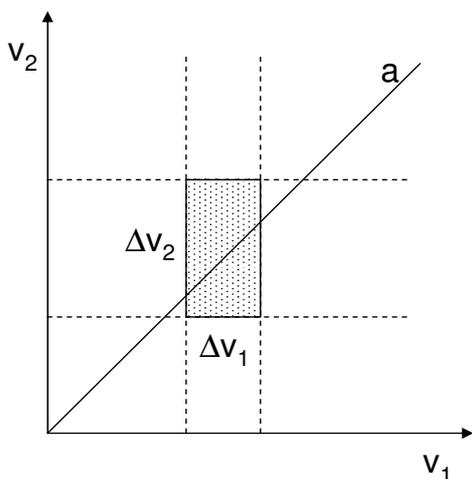
Figure 3.8: **Estimation of the flux space size**
See text (3.10.1) for details.

In any case, the area of the flux space, that can be covered by those two fluxes, will be smaller than or equal to the dotted rectangle.

$$\text{Flux Space} \leq \Delta v_1 \cdot \Delta v_2 \tag{3.17}$$

If we zoom to more than two dimensions, the structure we build is a cuboid for three dimensions and a hypercuboid for more than three dimensions. The volume of that hypercuboid can be taken as an upper estimation of the solution space's size:

$$\text{Flux Space} = \prod_i v_i \qquad \forall\; v_i \neq 0 \tag{3.18}$$

$$v_i \text{ refer to reaction rates}$$

In order to get meaningful results, we must only take those fluxes into account, that are different from zero. Also, to be able to compare results of different analyses, only those fluxes can be used that are non-zero in all analyses.

3.10.1.1 Flux space development graphs

Equation 3.18 gives an upper estimate of the size of the flux space. To make the development visible, we may sort all flux variabilities, then build their product consecutively, and plot the development over the reactions. In order to scale down the usually large interval covered, we use half logarithmic plotting.

The idea is, to visualize the way the flux variabilities work together to build the solution space. An outstanding, low or highly variable flux would cause a jump in this graph. A smooth curve, on the contrary, hints to a sound distribution of flux variabilities.

One must be aware that the final value of this analysis is an *upper* estimation of the size of the solution space. Even if this estimate is high, it may be, that the real size is much smaller, due to the coherences outlined in figure 3.8.

3.11 Knock out study

We calculated all possible single knock out mutants of the modeled organism. Thereby, we used the energy balance analysis as described in section 3.4. To be able to compare the results, the initial irreversibilities (see 3.4.2) determined for the wild-type model were used. The analysis of the knock out mutants focused on the biomass production rate.

For all reactions that are not catalyzed by an enzyme, the knock out mutants were constructed as if each of those reactions were catalyzed by a single enzyme. We included these reactions, because in some cases we could not assume the reactions to take place without enzymatic help, but no enzyme had yet been attached to them. A more detailed inspection can be found in the discussion (5.5). Reactions which are catalyzed by more than one possible enzyme have not been included in this study, since only *single* knock out mutants were computed.

Chapter 4

Results

4.1 The *Corynebacterium glutamicum* metabolism model

Building a genome based model of the investigated organism, like described in section 3.2.2, was one of the main parts of this work.

The metabolic network of the investigated bacterium, as used throughout this thesis, consists of the central carbon metabolism as well as all production pathways necessary to synthesize amino acids, fatty acids, and all other metabolites needed to build biomass as defined in (Takaç *et al.*, 1998). A detailed depiction of the model would either be too large or too confusing if pressed onto one or two pages. Thus, we only provide a schematic overview in figure 4.2 and a complete list of all reactions and metabolites is located within appendix A. The abbreviations used in figure 4.2 are given in table 4.1. The representing stoichiometric matrix (see 2.6) owns the dimensions $225 \cdot 244$, compare also figure 4.1.

For a list of the initially given reversibilities see table 4.3.3.2 on page 83.

Number of reactions		244
	thereof internal	206
	thereof external	38
Number of metabolites		225
	thereof internal	206
	thereof external	19

Figure 4.1: **Stoichiometrix Matrix Overview**
The stoichiometric matrix of the *C. glutamicum* model.

4.1.1 DNA microarray analyses data

About background theory concerning DNA transcription and microarrays the reader is referred to sections 2.2.1 and 2.4. Special information about how DNA array data is used within this work is given in 3.2.3, page 30.

Table 4.2 shows a statistical overview of the data derived from the microarray experiments and the annotation. As shown in this table, there are 11 enzymes for which no gene is expressed with an $R\text{-}value \geq 2$ (compare equation 3.1, page 30). Table 4.3 lists those enzymes and the reactions catalyzed by them, explicitly. For details about the change of the systemic responses when neglecting those enzymes see the knock out section 4.5.

The interested reader can find an additional table (B.1), listing all enzymes separated into the classes *expressed*, *weakly expressed*, and *non-annotated*, in appendix B.
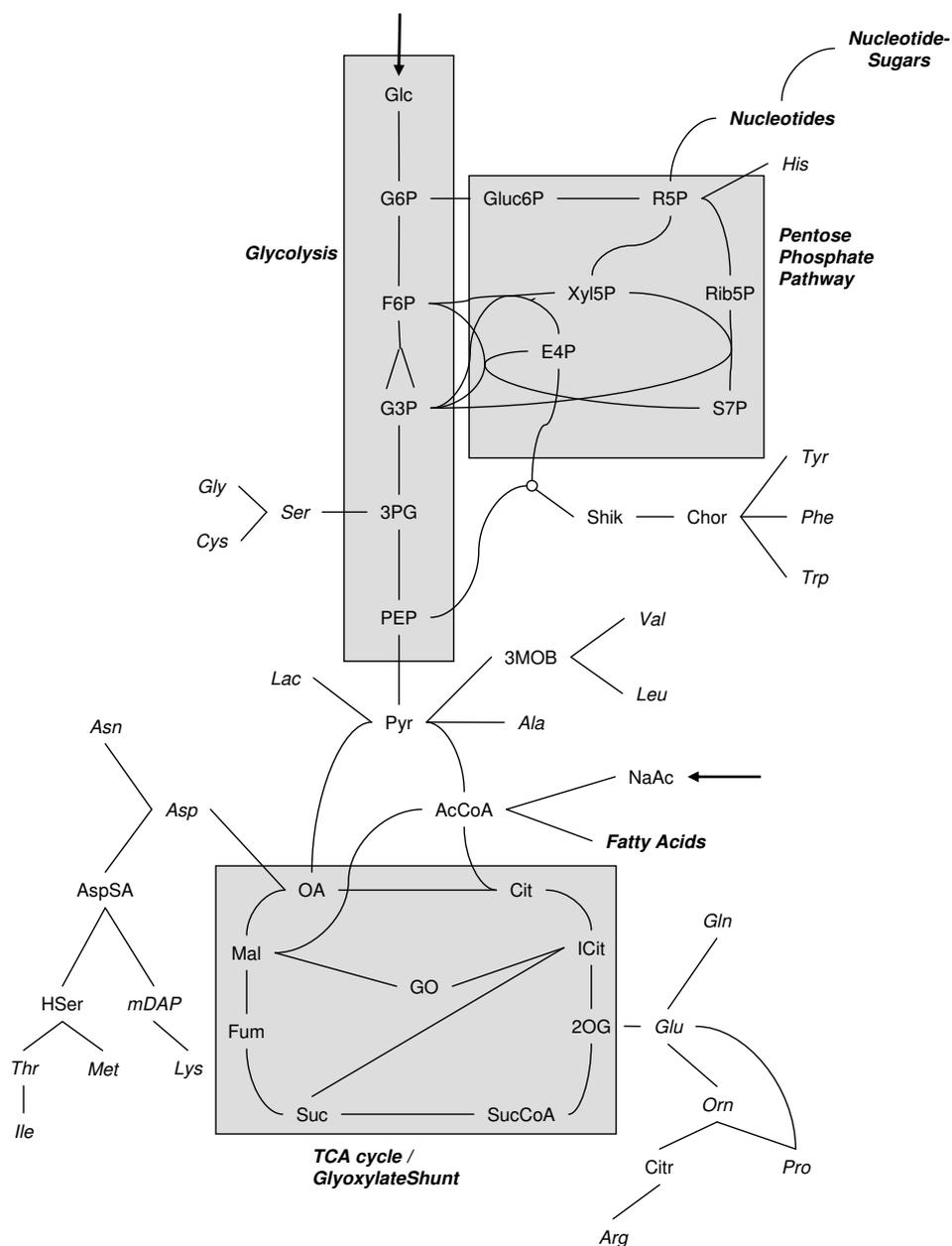
Figure 4.2: **Overview of the *C. glutamicum* model**
A schematic overview of the model used throughout this work. **Bold** arrows depict entries for the carbon sources glucose and acetate. All *italic metabolites* are part of the biomass. ***Bold and italic*** are metabolic pools that flow into biomass, or pathway names. The abbreviations used are given in table 4.1. A complete list of all reactions and metabolites can be found in appendix A.

Table: 4.1: **Abbreviations used in figure 4.2**
The abbreviations that are used in figure 4.2, arranged in alphabetical order.

Abbr.	Name	Abbr.	Name
2OG	2-Oxoglutarate	Ile	L-Isoleucine
3MOB	3-methyl-2-oxobutanoate	Lac	(S)-Lactate
3PG	3-Phospho-D-glycerate	Leu	L-Leucine
AcCoa	Acetyl-CoA	Lys	L-Lysine
Ala	L-Alanine	Mal	(S)-Malate
Arg	L-Arginine	mDAP	meso-2,6-Diaminoheptanedioate
Asn	L-Asparagine	Met	L-Methionine
Asp	L-Aspartate	NaAc	Sodium Acetate
AspSA	L-Aspartate 4-semialdehyde	OA	Oxaloacetate
Chor	Chorismate	Orn	L-Ornithin
Cit	Citrate	PEP	Phosphoenolpyruvate
Citr	L-Citrulline	Phe	L-Phenylalanine
Cys	L-Cysteine	Pro	L-Proline
E4P	D-Erythrose 4-phosphate	Pyr	Pyruvate
F6P	D-Fructose 6-phosphate	R5P	D-Ribulose 5-phosphate
Fum	Fumarate	Rib5P	D-Ribose 5-phosphate
G3P	Glyceraldehyde 3-phosphate	S7P	Sedoheptulose 7-phosphate
G6P	D-Glucose 6-phosphate	Ser	L-Serine
Glc	D-Glucose	Shik	Shikimate
Gln	L-Glutamine	Suc	Succinate
Glu	L-Glutamate	SucCoA	Succinyl-Co
Gluc6P	6-Phospho-D-gluconate	Thr	L-Threonine
Gly	Glycine	Trp	L-Tryptophan
GO	Glyoxylate	Tyr	L-Tyrosine
His	L-Histidine	Val	L-Valine
HSer	L-Homoserine	Xyl5P	D-Xylulose 5-phosphat
ICit	Isocitrate		

Table: 4.2: **DNA microarray analyses and annotation data**
Statistical analysis of the data derived from DNA microarray experiments. The *Model* column shows data referring only to the *C. glutamicum* model used in this work. The *Annotations* column refers to data derived from the combination of our inhouse annotation (see 3.2.2) and the one performed by Kalinowski and coworkers (Kalinowski *et al.*, 2003).

Description	Model	Annotations
Number of genes	302	3002
Number of enzyme coding genes	302	953
Number of enzymes (incl. 23 non-annotated in model)	186	610
Number of genes expressed with an *R-value* < 2	44	520
Number of enzymes with no encoding gene expressed with an *R-value* ≥ 2	11	69
Above as percentage of all enzymes	5.9 %	11.3 %
Number of genes not measured on microarrays	20	207
Above as percentage of all genes	6.6 %	6.9 %

Table: 4.3: **Weakly expressed genes**
All enzymes used within the scope of this work's *C. glutamicum* model that have no encoding gene expressed with an *R-value* ≥ 2 (see 3.2.3, p. 30).

Enzyme	Reaction(s)	reaction(s) also catalyzed by[*]	Enzyme	Reaction(s)	reaction(s) also catalyzed by[*]
1.1.1.1	reaction 135	1.1.1.2	2.7.1.71	reaction 107	—
1.2.1.11	reaction 85	—	2.7.4.9	reaction 187	—
1.2.1.59	reaction 11	1.2.1.12		reaction 188	—
2.1.2.3	reaction 175	—	3.1.3.12	reaction 151	—
2.1.3.2	reaction 182	—	4.2.3.5	reaction 104	—
2.3.1.12	reaction 38	1.8.1.4 & 1.2.4.1	5.4.99.16	reaction 146	—

[*]the referring genes are expressed with an *R-value* ≥ 2

4.1.2 Non-annotated reactions

For the reasons stated in 3.2.4 there was a number of reactions that were added to the model without being connected to an enzyme that is known to exists in *C. glutamicum*. These can be separated into two species:

1. reactions which are catalyzed by an enzyme, according to KEGG (Kanehisha & Goto, 2000), but the enzyme has not been annotated (see table 4.4).

2. reactions which are not connected to a catalyzing enzyme. These reactions are marked by '0.0.0.0' as a catalyzing enzyme in table A.1 and are explicitly listed in table 4.5

Table: 4.4: **Non-annotated reactions**
All non-annotated enzymes used in this work's *C. glutamicum* model and the reactions catalyzed by them.

Enzyme Number	Reaction	Enzyme Number	Reaction
1.3.5.1	part of reaction 198	2.6.1.2	reaction 45
1.3.99.11	reaction 183	2.7.1.3	reaction 7
1.6.1.1	reaction 200	2.7.2.12	reaction 42
	part of reaction 196	2.7.7.38	reaction 139
1.6.1.2 & 1.8.1.9 & 1.17.4.2	reaction 140	3.1.3.15	reaction 61
	reaction 141	3.1.3.45	reaction 138
	reaction 142	3.5.4.9	reaction 194
	reaction 189	3.6.1.5	part of reaction 197
1.8.2.1	part of reaction 196		part of reaction 198
2.3.1.1	reaction 96	3.6.1.8	reaction 158
2.3.1.157	reaction 157	4.1.1.3	reaction 39
2.5.1.55 & 5.3.1.13	reaction 143	5.4.99.5	reaction 105
		6.3.4.16	reaction 46

Table: 4.5: **Non-catalyzed reactions** All reactions present in the *C. glutamicum* model that are not attached to a catalyzing enzyme. In table A.1, appendix A these reactions are marked by '0.0.0.0' as a catalyzing enzyme. Membrane transport is modeled without cost. *Source / Sink* reactions are artificial reactions that do not represent any real reaction.

Reaction		**Comment**
77	:	reaction marked *spontaneous* in KEGG
115	:	reaction marked *non enzymatic* in KEGG
144	:	reaction taken from (Takaç *et al.*, 1998), lumps 4 steps together, which do not have complete EC numbers attached to them
145	:	biomass production (generic reaction)
Isomeric conversions		
Reactions	:	203, 204, 205, 206, 207
Membrane transport		
Reactions	:	202, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225
Source / Sink		
Reactions	:	226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 240, 239, 241, 242, 243, 244

4.1.3 Biomass constitution

As noted in 3.2.5 we adapted the biomass constitution given in (Takaç *et al.*, 1998). Table 4.6 shows how biomass is constituted in this work.

Table: 4.6: **Biomass constitution**
The metabolites and their corresponding stoichiometric factors building up biomass are shown. This corresponds directly to reaction 244 given in appendix A, table A.1.

Stoichiometric factor	Metabolite	Stoichiometric factor	Metabolite
0.0235	ADP-D-glycero-D-manno-heptose	0.27	L-Isoleucine
		0.42	L-Leucine
0.185	ATP	0.32	L-Lysine
0.0235	CMP-3-deoxy-D-manno-octulosonate	0.14	L-Methionine
		0.0593	L-Ornithine
0.12	CTP	0.17	L-Phenylalanine
0.02	dATP	0.2	L-Proline
0.02	dCTP	0.377	L-Serine
0.02	dGTP	0.24	L-Threonine
0.02	dTTP	0.05	L-Tryptophan
0.258	Fatty Acid	0.13	L-Tyrosine
0.58	Glycine	0.4	L-Valine
0.2	GTP	0.0276	meso-2,6-Diaminoheptanedioate
0.5352	L-Alanine		
0.28	L-Arginine	0.0235	Myristic Acid
0.22	L-Asparagine	0.0235	Myristoleic Acid
0.22	L-Aspartate	0.129	sn-Glycerol 3-phosphate
0.09	L-Cysteine	0.1697	UDPglucose
0.2776	L-Glutamate	0.0433	UDP-N-acetyl-D-glucosamine
0.25	L-Glutamine	0.0276	UDP-N-acetylmuramate
0.09	L-Histidine	0.13	UTP

4.2 Model validation

The method used to validate the established model has been described in section 3.9 (p. 58 ff.). Figures 4.3 and 4.4 show the fluxes measured by Wendisch and coworkers in (Wendisch *et al.*, 2000). As said in section 3.9.2, we were able to achieve these fluxes with an accuracy of $\pm 30\,\%$ by using the same carbon input and the given biomass $/\,CO_2$ ratio (see 3.8.1), only. Table 4.7 details the imposed constraints.

The results of a robustness analysis, as described in 3.7, are shown in figure 4.5.

Table: 4.7: **Imposed constraints for model validation**
These are the constraints imposed to reproduce the flux distribution published in (Wendisch *et al.*, 2000) and shown in figures 4.3 and 4.4.

Methods used			
EBA (see 3.4)			
EEB (see 3.5)			
Constraints imposed			
Carbon source	Influx $\left[\frac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}}\right]$	CO_2 efflux $\left[\frac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}}\right]$	Biomass efflux $\geq 90\,\%$ of $\left[\frac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}}\right]$
Acetate	16.2	≥ 23	0.2236333
Glucose	4.44	≥ 15.72	0.2597953
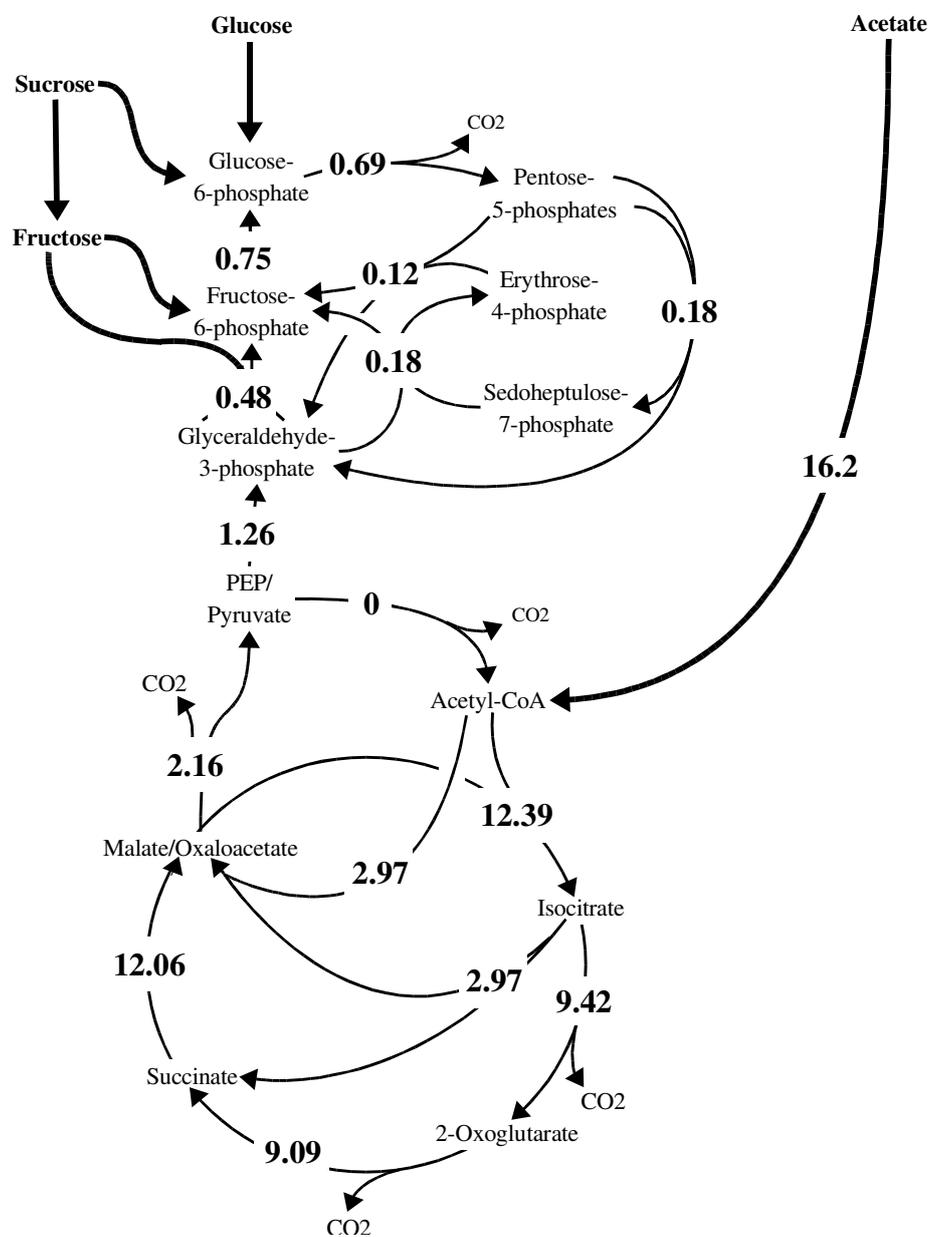
Figure 4.3: **Metabolic fluxes in *C. glutamicum* during growth on acetate**
See table 4.7 for constraints used to reproduce displayed reaction rates. Fluxes given in $\frac{\text{mmol}}{\text{g}(DW)\cdot\text{h}}$. All fluxes have been measured and previously published by Wendisch and coworkers in (Wendisch *et al.*, 2000). Figure cloned from above reference.
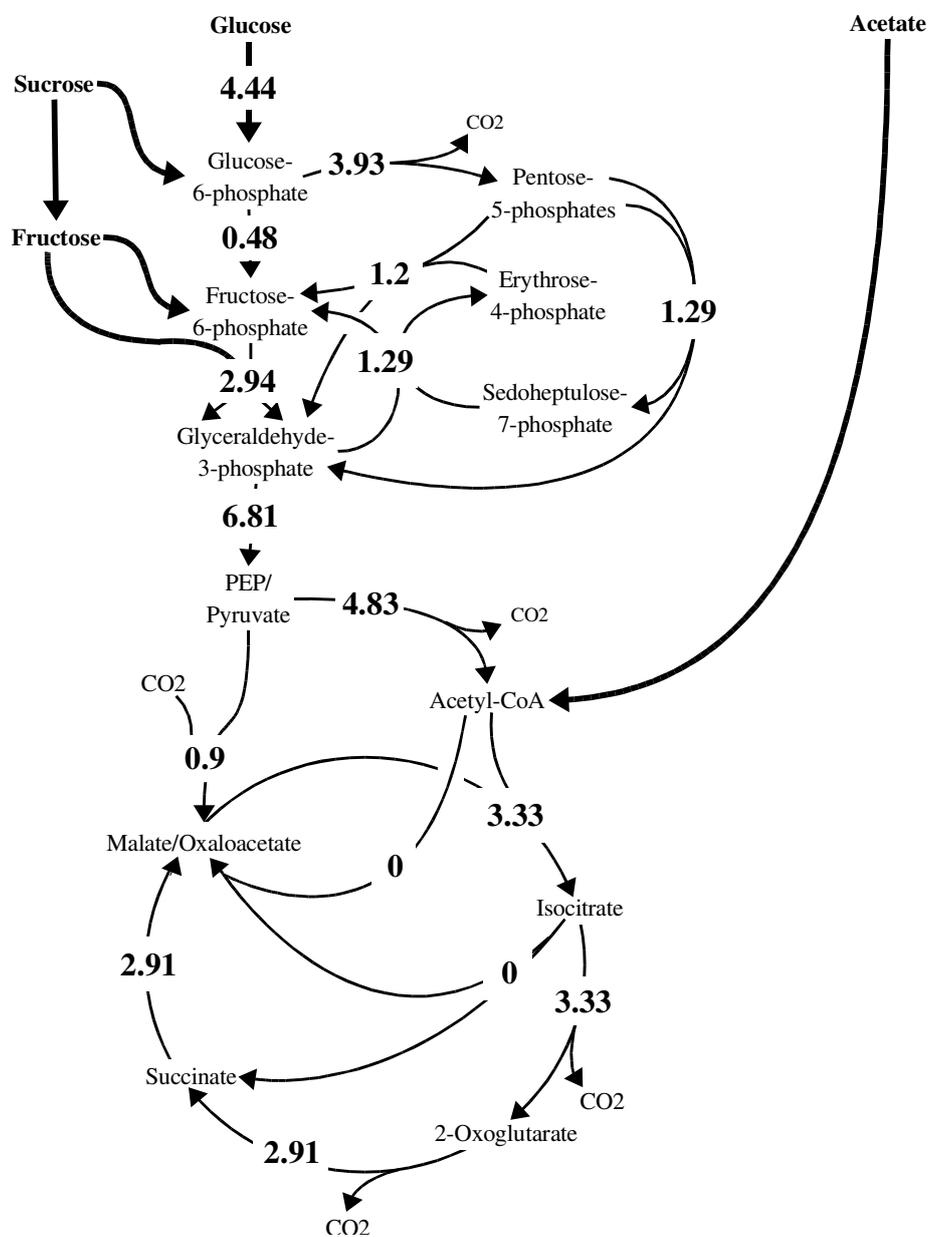
Figure 4.4: **Metabolic fluxes in *C. glutamicum* during growth on glucose**

See table 4.7 for constraints used to reproduce displayed reaction rates. Fluxes given in $\frac{\text{mmol}}{\text{g}(DW)\cdot\text{h}}$. All fluxes have been measured and previously published by Wendisch and coworkers in (Wendisch *et al.*, 2000). Figure cloned from above reference.
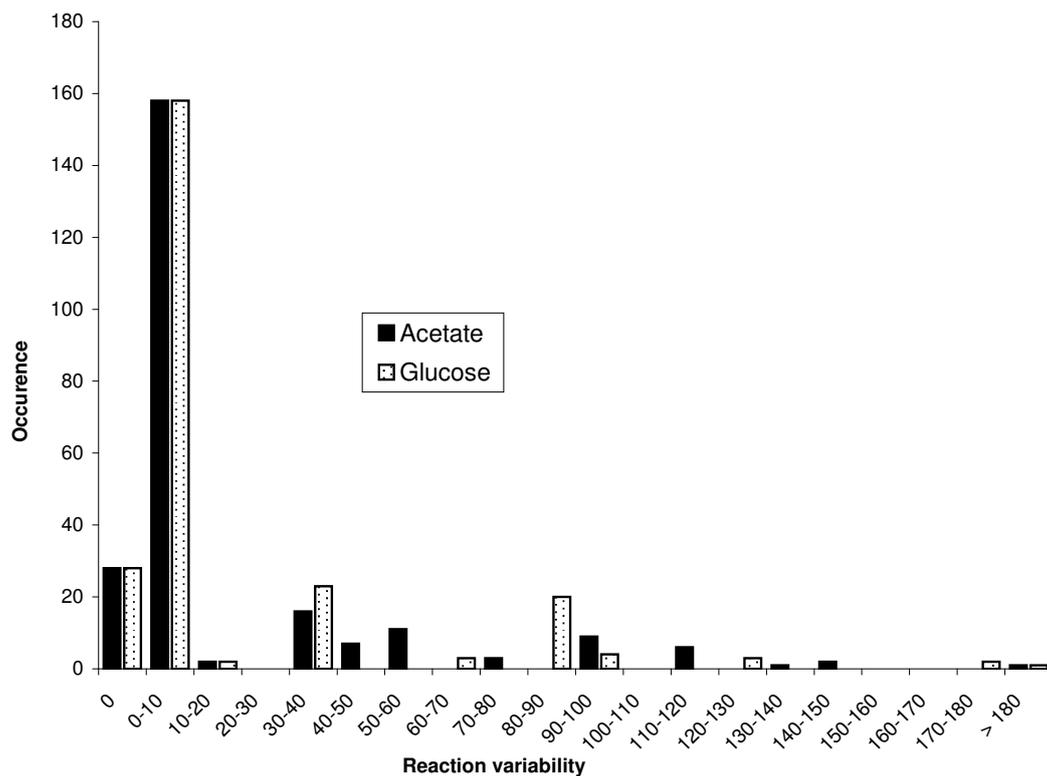
Figure 4.5: **Robustness analysis of the *C. glutamicum* metabolism during model validation**

Analysis performed with constraints given in table 4.7. Figure shows the predicted reaction variabilities for constraints imposed to reproduce the flux distribution published in (Wendisch *et al.*, 2000) and shown in figures 4.3 and 4.4.

4.3 Using laboratory conditions

In the metabolic profiling experiments, carried out in CUBIC laboratories, we achieved growth-rates different from those found by Wendisch *et al.* (Wendisch *et al.*, 2000). Since no carbon dioxide production rates have been determined, in all following calculations we reproduced the laboratory conditions by using the growth rates found and thereby assumed, the CO_2 / Biomass ratio equals the one given in (Wendisch *et al.*, 2000). The determination of the growth-rates is described in detail in section 3.6.4.1, page 51 ff.

The figures 4.6 to 4.9 in the next section show the OD-curves measured. Table 4.8 provides an overview of the determined growth rates.

The following chapters show the results obtained by using different sets of constraints.

All metabolic profiling experiments using glucose as carbon source have been carried out by Dr. Sebastian Buchinger. All metabolic profiling experiments using acetate as carbon source have been carried out by Eliane Hornemann. The latter data has also been published in the diploma thesis of Ms. Hornemann (Hornemann, 2005).

4.3.1 Identification of periods of exponential growth

The method referring to the figures shown below is described in section 3.6.4.1, page 51 ff. Table 4.8 gathers all important data from these figures.
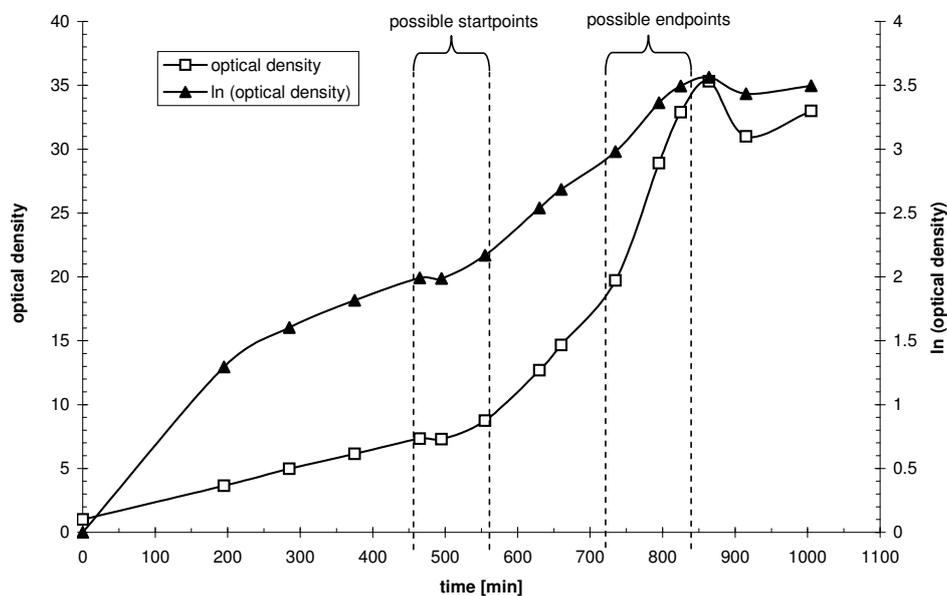
Figure 4.6: **Growth curve for glucose fed *C. glutamicum* (1)**
Boxed line = optical density, triangulared line = natural logarithm of OD.
Dashed areas show start and end ranges of the period of exponential growth.
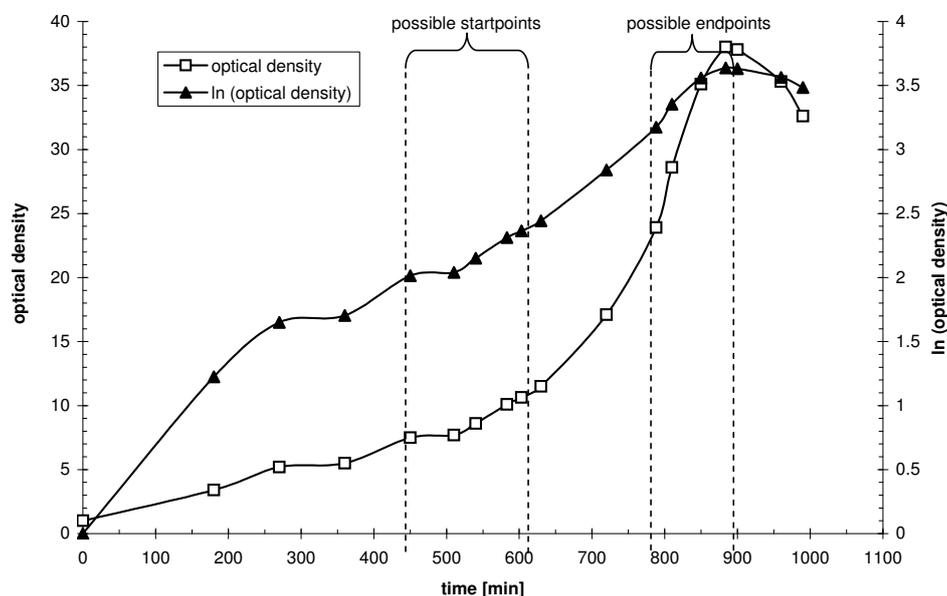
Figure 4.7: **Growth curve for glucose fed *C. glutamicum* (2)**
Boxed line = optical density, triangulared line = natural logarithm of OD.
Dashed areas show start and end ranges of the period of exponential growth.
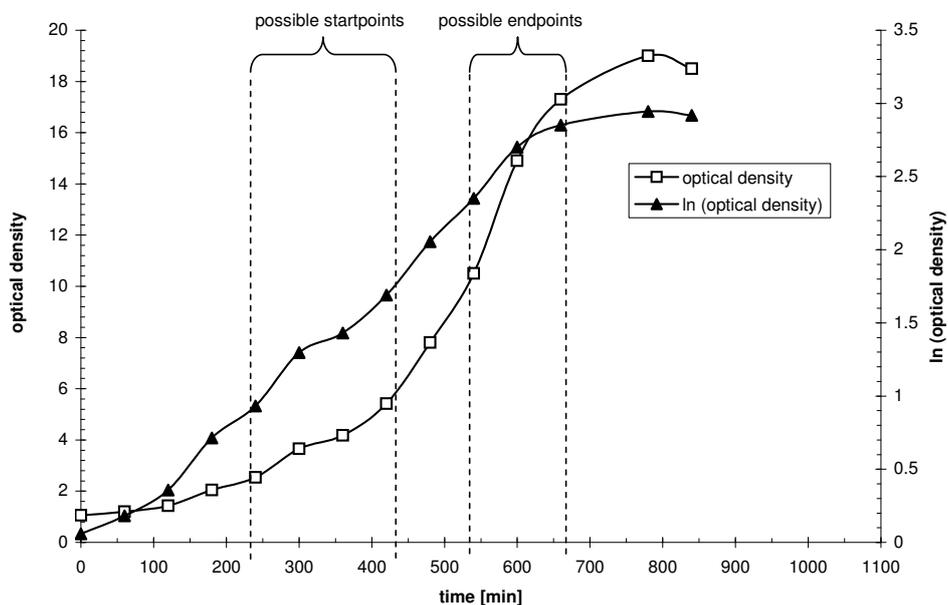
Figure 4.8: **Growth curve for acetate fed *C. glutamicum* (1)**
Boxed line = optical density, triangulared line = natural logarithm of OD.
Dashed areas show start and end ranges of the period of exponential growth.
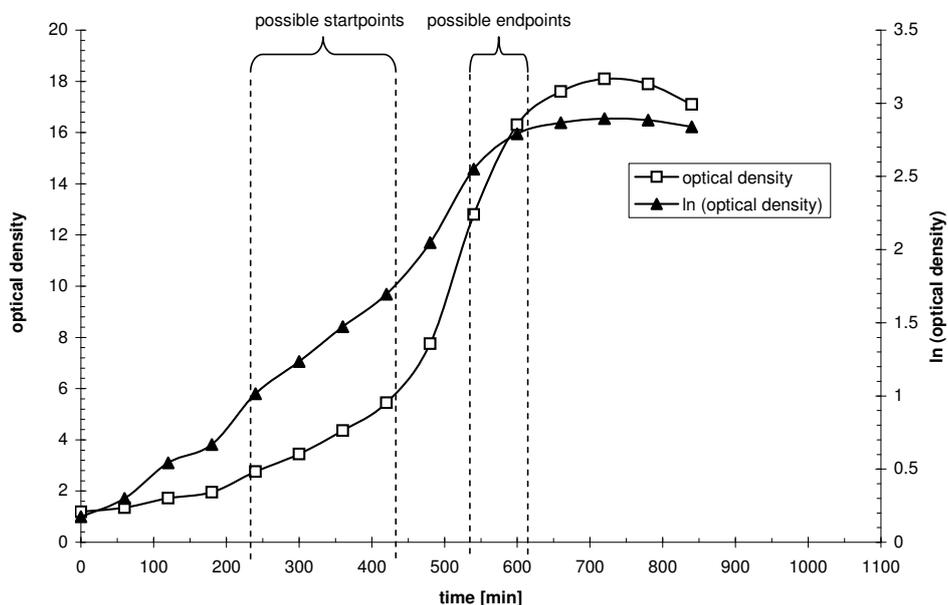
Figure 4.9: **Growth curve for acetate fed *C. glutamicum* (2)**
Boxed line = optical density, triangulared line = natural logarithm of OD.
Dashed areas show start and end ranges of the period of exponential growth.

78

Table: 4.8: **Growth rate data**

Overview of the growth rates of the periods of exponential growth, derived as described in section 3.6.4.1, page 51 ff.

Carbon source	Figure	Growth rate [/h]	Mean [/h]	Ratio (Glc/NaAc)
Glucose (1)	4.6	0.293	0.279 ± 0.014	
Glucose (2)	4.7	0.265		0.734
Acetate (1)	4.8	0.333	0.380 ± 0.047	
Acetate (2)	4.9	0.427		

4.3.2 Flux Balance Analysis

The results shown next have been obtained by a flux balance analysis, as outlined in section 3.3.

Table 4.9 shows the constraints imposed during the flux balance analysis of the *Corynebacterium glutamicum* model. Figure 4.10 shows the referring robustness analysis. Particular attention shall be directed to the two bars at the very right of the diagram. They show that a rather large fraction of all fluxes levels at their upper bound[1], meaning they would be unbound if the constraint was not placed. All reactions are considered reversible during a flux balance analysis.

---

[1]The upper and lower bounds are arbitrarily set, for the only reason that the optimizer would fail, if they were infinity.

Table: 4.9: **Constraints imposed on *FBA***

These are the constraints imposed during a flux balance analysis of the *C. glutamicum* metabolism. The biomass production rates translate into the mean growth rates shown in table 4.8, if a biomass constitution as defined in table 4.6 on page 71 is used.

Methods used			
FBA (see 3.3)			
Constraints imposed			
Carbon source	Influx $\left[\dfrac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}}\right]$	CO_2 efflux $\left[\dfrac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}}\right]$	Biomass efflux $\geq 90\%$ of $\left[\dfrac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}}\right]$
Acetate	21.27	≥ 30.2	0.2935782
Glucose	3.68	≥ 13.04	0.2150686

Figure 4.10: **Robustness analysis under *FBA* conditions**

Analysis performed under constraints given in table 4.9. The figure shows the predicted reaction variabilities for constraints imposed to reproduce growth rates measured in CUBIC laboratories (compare 4.3.1). All reactions are considered reversible.

4.3.3 Energy Balance Analysis

The reader is referred to section 3.4 for details about energy balance analyses. On page 37 it was explained, that we used an adapted optimization method (3.4.1) to determine the feasible reaction directions in terms of *EBA*. Also, in section 3.4.2 we explained how the initial irreversibilities were calculated. After predicting the allowed reaction directions, we identified all remaining cyclic fluxes and successively introduced initial irreversibilities for the reactions most common in the cycles until all cycles were eliminated.

The following sections show the results concerning the initially used irreversibilities and the results of an energy balance analysis.

4.3.3.1 EBA - initial irreversibilities

To be able to calculate thermodynamically feasible constraints on all reaction directions, it was necessary to give initial irreversibilities for 24 reactions (see above or section 3.4.2). Otherwise cyclic fluxes could occur, undermining thermodynamic feasibility. Table 4.10 lists those reactions in detail.

Table: 4.10: **Initial irreversibilities**
All reactions for which an initial direction information had to be given (see text and compare 3.4.2, page 38).

No.	Reaction	Reaction details
1.	reaction 4	D-Fructose 1,6-bisphosphate + H2O => D-Fructose 6-phosphate + Orthophosphate
2.	reaction 7	ATP + D-Fructose => ADP + D-Fructose 1-phosphate + H+
3.	reaction 8	ATP + D-Fructose => ADP + beta-D-Fructose 6-phosphate + H+
4.	reaction 27	Citrate + CoA + H+ <= Acetyl-CoA + H2O + Oxaloacetate

Table: 4.10 . . . Continued from previous page

5.	reaction 33	2-Oxoglutarate + Lipoamide + H+ => S-Succinyldihydrolipoamide + CO2
6.	reaction 36	Isocitrate => Succinate + Glyoxylate
7.	reaction 38	Pyruvate + CoA + NAD+ => Acetyl-CoA + CO2 + NADH
8.	reaction 39	Oxaloacetate + H+ => Pyruvate + CO2
9.	reaction 54	2 L-Glutamate + NADP+ <= L-Glutamine + 2-Oxoglutarate + NADPH + H+
10.	reaction 56	ATP + L-Glutamate + NH3 => ADP + Orthophosphate + L-Glutamine + H+
11.	reaction 57	L-Glutamine + H2O => L-Glutamate + NH3
12.	reaction 97	N2-Acetyl-L-ornithine + H2O => Acetate + L-Ornithine
13.	reaction 112	ATP + L-Glutamate => ADP + L-Glutamyl 5-phosphate
14.	reaction 133	sn-Glycerol 3-phosphate + NAD+ <= Glycerone phosphate + NADH + H+
15.	reaction 134	ATP + Glycerol => ADP + sn-Glycerol 3-phosphate
16.	reaction 148	(1,4-alpha-D-Glucosyl)n + alpha-D-Glucose <= Maltose
17.	reaction 149	UTP + D-Glucose 1-phosphate + H+ => Pyrophosphate + UDPglucose
18.	reaction 151	alpha,alpha'-Trehalose 6-phosphate + H2O => alpha,alpha-Trehalose + Orthophosphate
19.	reaction 158	ATP + H2O => AMP + Pyrophosphate + H+
20.	reaction 159	ATP + NAD+ => ADP + NADP+ + H+
21.	reaction 197	2 NADH + 4 ADP + 4 Orthophosphate + Oxygen + 6 H+ => 2 NAD+ + 4 ATP + 6 H2O
22.	reaction 198	2 FADH2 + 2 ADP + 2 Orthophosphate + Oxygen + 2 H+ => 2 FAD + 2 ATP + 4 H2O

Table: 4.10 . . . Continued from previous page

23.	reaction 199	Pyrophosphate + H2O => 2 Orthophosphate + H+
24.	reaction 201	NADP+ + H2O => Orthophosphate + NAD+

4.3.3.2 EBA - robustness analyses

Table 4.11 shows the constraints imposed on the energy balance analysis. The corresponding robustness analysis is shown in figure 4.11. The initial irreversibilities shown in the previous section have been used herein.

Table: 4.11: **Constraints imposed on *EBA***
Constraints imposed to perform an energy balance analysis of the *C. glutamicum* metabolism. External energy balance related constraints are imposed, too. The biomass production rates translate into the mean growth rates shown in table 4.8, if a biomass constitution as defined in table 4.6 on page 71 is used.

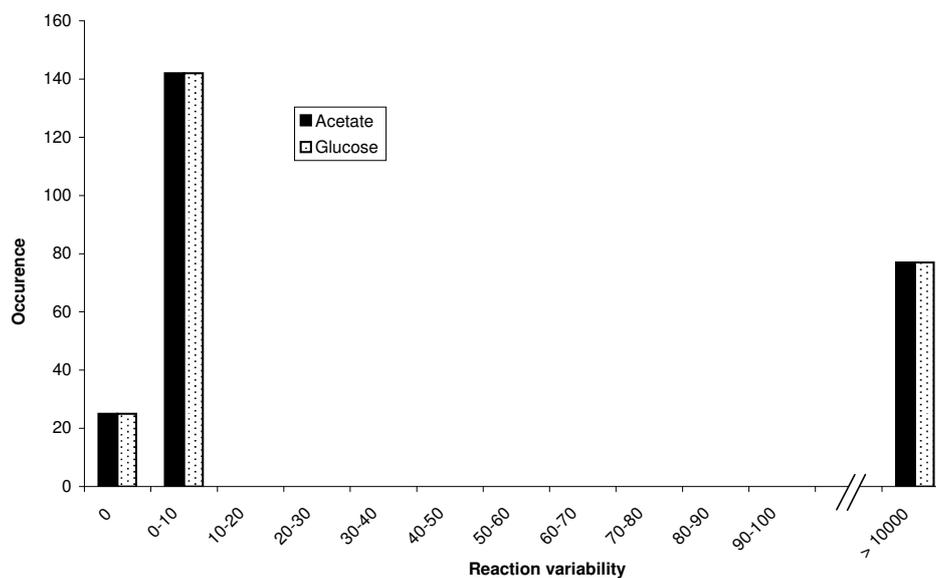| Methods used			
EBA (see 3.4)			
EEB (see 3.5)			
Constraints imposed			
Carbon source	Influx $\left[\frac{\text{mmol}}{\text{g}(\textit{DryWeight})\cdot\text{h}}\right]$	CO_2 efflux $\left[\frac{\text{mmol}}{\text{g}(\textit{DryWeight})\cdot\text{h}}\right]$	Biomass efflux $\geq 90\%$ of $\left[\frac{\text{mmol}}{\text{g}(\textit{DryWeight})\cdot\text{h}}\right]$
Acetate	21.27	≥ 30.2	0.2935782
Glucose	3.68	≥ 13.04	0.2150686

Figure 4.11: **Robustness analysis under *EBA* conditions**
Analysis performed under constraints given in table 4.11. The figure shows
the predicted reaction variabilities for constraints imposed to reproduce
growth rates measured in CUBIC laboratories (compare 4.3.1). For reac-
tion irreversibilities see 4.3.3.2.

4.3.4 Comparison of *FBA* and *EBA*

The details about *FBA* and *EBA* are described in sections 3.3 and 3.4.

Figure 4.12 shows a comparison of the results of the two methods. We set the variability of each flux of the *FBA* to 100 % and afterwards calculated the *EBA*-variabilities as percentages with respect to them. In figure 4.13 a flux space development graph (see 3.10.1.1) for both, *FBA* and *EBA*, is shown.
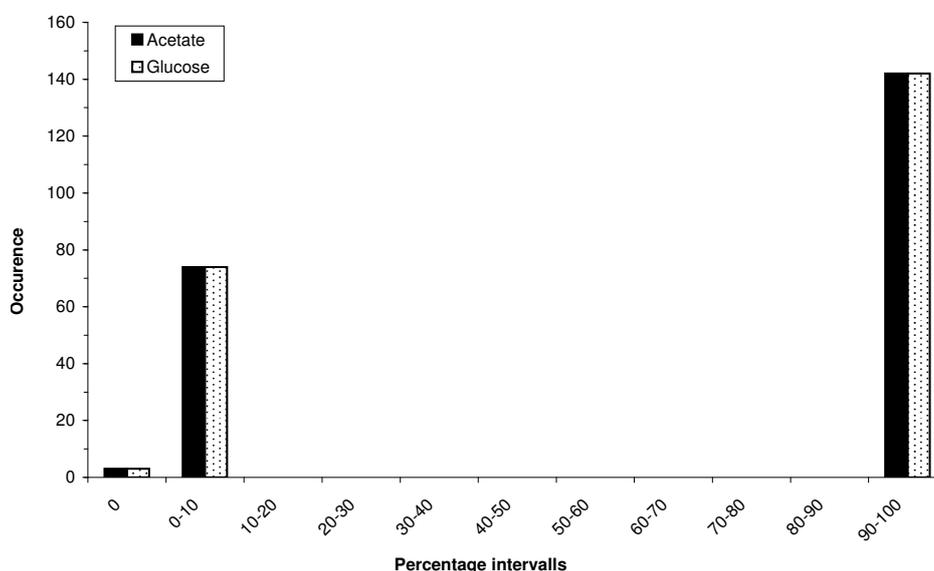
Figure 4.12: **Comparison of robustness analyses under *FBA* and *EBA* conditions**
The variabilities of the robustness analysis in figure 4.11 (*EBA*) are calculated as percentage values with respect to the the data shown in figure 4.10 (*FBA*). The referring constraints are given in table 4.11 and table 4.9, respectively.
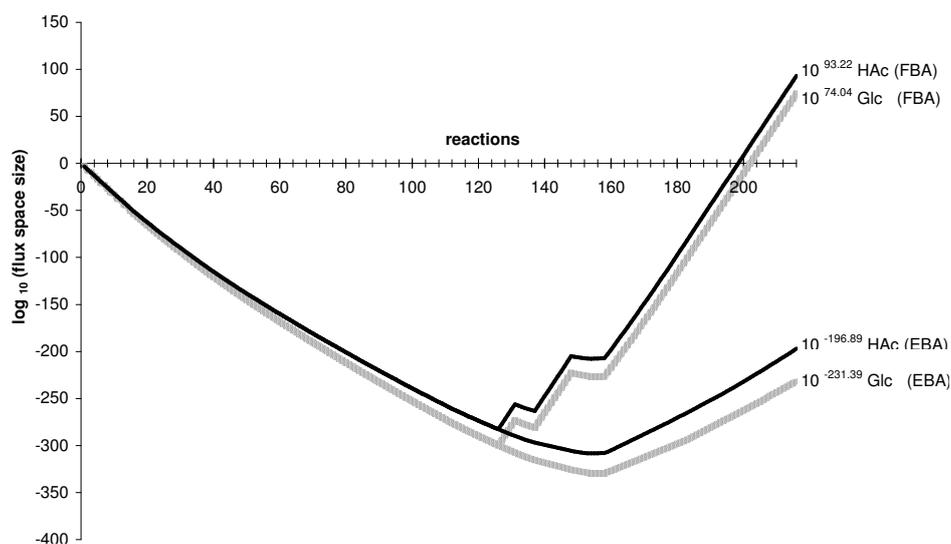
Figure 4.13: **Flux space development**
A comparison of the flux space development (compare 3.10.1.1) of both carbon sources (NaAc = acetate, Glc = glucose) under *FBA* and *EBA* conditions. The whole data set has been sorted for increasing variabilitites of acetate feeding under *FBA* constraints. Numbers given to the right are absolute values of upper estimates of flux space sizes in units of $\left[\frac{\mathrm{mmol}}{\mathrm{g}(DryWeight)\cdot\mathrm{h}}\right]^{216}$. For details about the constraints imposed, see tables 4.9 and 4.11.

Table: 4.12: **EEB-Analysis**
External energy balance analysis of acetate and glucose as carbon sources, under constraints given in table 4.11 on page 83. Numbers shown are ΔG^0 values (compare 2.1.2), calculated as explained in section 3.5.

	Acetate as carbon source	Glucose as carbon source	
ΔG^0_{BM}	-11334.24	-5581.49	$\left[\frac{J}{g(DryWeight)\cdot h}\right]$
$\Delta G^0_{CO_2}$	-16442.56	-9323.72	$\left[\frac{J}{g(DryWeight)\cdot h}\right]$
$\Delta\Delta G^0 = \Delta G^0_{CO_2} - \Delta G^0_{BM}$	-5108.32	-3742.23	$\left[\frac{J}{g(DryWeight)\cdot h}\right]$
$\Delta\Delta G^0$ in % of $\Delta G^0_{CO_2}$	31.07 %	40.14 %	

4.3.5   External energy balance analysis

Section 3.5.1 explains the method of an external energy balance analysis. In a nutshell: we suppose the energy the system can gain from metabolizing all inflowing carbon into CO_2 states the energy maximally available to the system. Thus, the difference between this energy and the drop in free enthalpy (compare 2.1.2) related to biomass production shows the amount of Gibbs free energy disposed to biomass.

Table 4.12 shows the energies dissipated for both carbon sources (acetate and glucose) under the constraints of table 4.11 on page 83. Special focus shall be laid on the bottom line, which shows the percentage of the energy available to the system that is transferred into biomass. The values of the free enthalpies used for the calculation of the external energy balance are given in table B.2, appendix B. A detailed explanation and discussion of the *EEB* results can be found in section 5.3.4.

4.4 Metabolic profiling

Chapter 3.6 gives a detailed description of the methods to implement metabolic profiling data that have been developed in this work.

In the following, we will at first present the metabolic profiling data used. Then, we will show the results obtained by using this data and the differences to the results obtained without the metabolic profiling data.

4.4.1 Metabolic profiling data

Table 4.13 shows the metabolic profiling data used in the work at hand. This data has been obtained after

- the identification of the periods of exponential growth. For method see 3.6.4.1, results are displayed in section 4.3.1.

- all normalizations described in section 3.6.4.2.

- identification of the highest correlating profiles as described in 3.6.4.3.

- selecting the metabolites as described in 3.6.5.

The reactions for which we introduced new constraints (compare 3.6.3), based on the data shown in table 4.13, are listed in table 4.14.

Table: 4.13: **Metabolic profiling data**
The metabolic profiling data implemented in this work. Only relevant metabolites are shown (compare 3.6.5). Acetate fed profile refers to datapoint at 600 minutes in figure 4.9; glucose fed profile refers to datapoint at 630 minutes in figure 4.6. These profiles correlated with a Pearson coefficient (see eq. 3.8, page 53) of 0.9102. As said in section 2.5.2 the measured values are pseudo-concentrations. Thus, no units can be given. For each datapoint 3 probes have been measured. Abbreviations: CS = carbon source, SD = standard-deviation, NaAc = acetate, Glc = glucose.

Metabolite	Concentration (CS = NaAc)	SD [%]	Concentration (CS = Glc)	SD [%]	Ratio NaAc / Glc
Glycerone phosphate	1125.51	14.63	12926.32	36.87	0.09
L-Valine	520318.08	19.01	3803855.93	16.63	0.14
Alanine	152335.93	1.57	974248.18	16.20	0.16
Glucose	36018.17	21.52	178445.68	30.52	0.20
Shikimate	6368.78	5.70	1167.76	54.76	5.45
Pyruvate	52213.84	33.21	9412.97	24.01	5.55
L-Proline/D-Proline	4312791.00	5.23	660870.64	10.66	6.53
Glucose 6-phosphate	63758.09	13.28	9598.33	51.70	6.64
Phenylalanine	52098.86	18.07	6694.11	40.59	7.78
Ribose 5-phosphate	57243.76	7.08	6512.81	50.23	8.79
N-Acetyl-L-glutamate	395504.81	2.83	29016.78	27.66	13.63
D-Fructose 6-phosphate	90374.43	3.07	6047.76	59.80	14.94
Fumarate	146132.19	26.36	8427.08	18.21	17.34
Fructose	12775.36	37.80	276.25	69.03	46.25
3-Phospho-D-glycerate	32709.30	12.48	475.32	49.10	68.82

Table: 4.14: **Reactions affected by metabolic profiling data**
The reactions affected by the data shown in table 4.13. Reactions' educts
must at least have a ratio of 5 between acetate fed and glucose fed profile
(equation 3.9, page 54). Also, the condition stated in equation 3.5 on page 47
is fulfilled for all reactions below.

Metabolite	Reaction	Direction
Glycerone phosphate	reaction 133	backward
L-Valine	none	—
Alanine	none	—
Glucose	reaction 1	forward
	reaction 204	backward
Shikimate	reaction 107	forward
Pyruvate	reaction 73	backward
	reaction 38	forward
	reaction 39	backward
	reaction 45	backward
	reaction 86	forward
	reaction 72	forward
L-Proline / D-Proline	none	—
Glucose 6-phosphate	reaction 150	forward
	reaction 206	forward
	reaction 9	backward
Phenylalanine	none	—
Ribose 5-phosphate	reaction 17	forward
N-Acetyl-L-glutamate	reaction 100	forward
D-Fructose 6-phosphate	reaction 154	forward
Fumarate	reaction 30	backward
Fructose	none	—
3-Phospho-D-glycerate	reaction 117	forward
	reaction 136	backward

4.4.2 Energy balance analysis

The results in this section are based on the theory of *EBA* (3.4) as well as on the developed method of implementation for metabolic profiling data (3.6).

Figure 4.14 shows a robustness analysis under the constraints shown in table 4.15. The reader should note, that the data for the model fed on glucose looks the same as in figure 4.11. The data for the acetate fed model is different, though.

Table: 4.15: **Constraints imposed on *EBA* including metabolic profiling data** Constraints imposed to perform an energy balance analysis of the *C. glutamicum* metabolism. External energy balance related constraints are imposed, too. Constraints deduced from metabolic profiling data (see table 4.13) are implemented. The biomass production rates translate into the mean growth rates shown in table 4.8, if a biomass constitution as defined in table 4.6 on page 71 is used.

Methods used			
EBA (see 3.4)			
EEB (see 3.5)			
Constraints imposed			
Carbon source	Influx $\left[\dfrac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}}\right]$	CO$_2$ efflux $\left[\dfrac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}}\right]$	Biomass efflux $\geq 90\,\%$ of $\left[\dfrac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}}\right]$
Acetate	21.27	≥ 30.2	0.2935782
Glucose	3.68	≥ 13.04	0.2150686
Constraints from metabolic profiling data			
Combinatorial constraint	**Reactions**		
Acetate $\leq$ Glucose	1, 9, 30, 39, 45, 73, 136		
Glucose $\leq$ Acetate	17, 38, 72, 86, 100, 107, 117, 133, 150, 154, 204, 206		
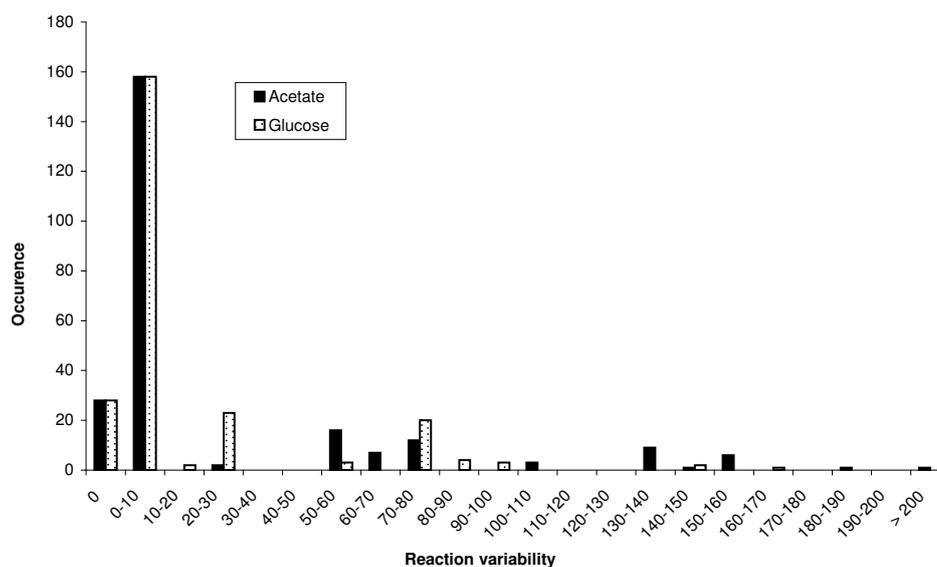
Figure 4.14: **Robustness analysis under *EBA* and metabolic profiling deduced conditions**
Analysis performed with constraints given in table 4.15. The figure shows the predicted reaction variabilities for constraints imposed to reproduce growth rates measured in CUBIC laboratories (compare 4.3.1) and constraints deduced from metabolic profiles (see 4.4). For reaction irreversibilities see 4.3.3.2.

4.4.3 Comparison of *EBA* excluding and including combinatorial constraints

Figure 4.15 gives a representation of the changes between the simulations based only on the rules of *EBA* (3.4) and *EEB* (3.5), on the one hand, and the simulations including the rules deduced from the metabolic profiling data (tab. 4.13), on the other hand.
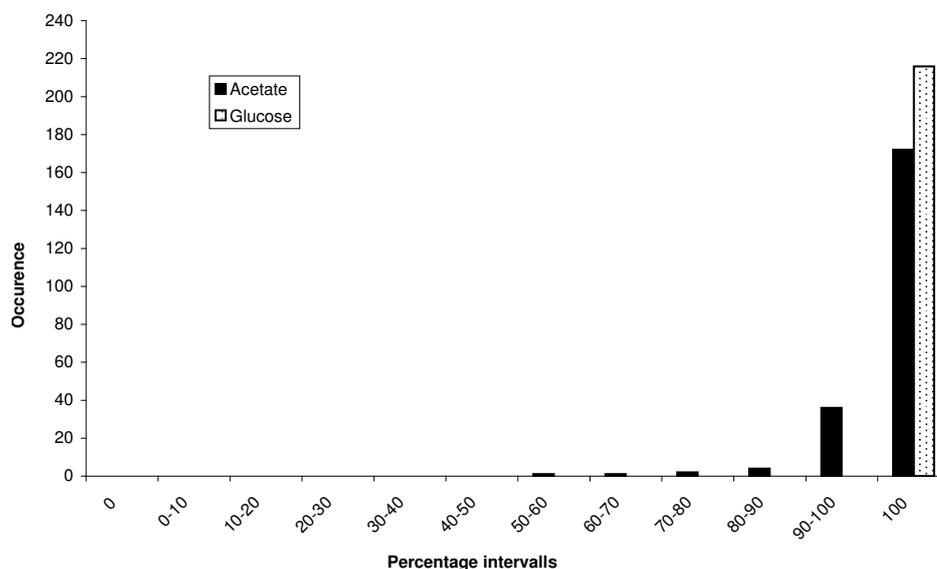
Figure 4.15: **Comparison of robustness analyses under *EBA* conditions alone and *EBA* including metabolic profiling deduced constraints**
The changes in the solution space, based on the additional constraints derived from metabolic profiling experiments (compare table 4.15). The compared analyses are shown individually in figures 4.11, p. 84 (*EBA*) and 4.14, p. 92 (*EBA*, metabolic profiling).

Only the solution space of the acetate fed model is reduced compared to its shape outlined in figure 4.11. The solution space based on glucose as carbon source stays the same. Figure 4.16 shows the development of the flux space, as introduced in section 3.10.1.1. The change of the upper estimate of its size, after implementing the combinatorial constraints, is only marginal here.
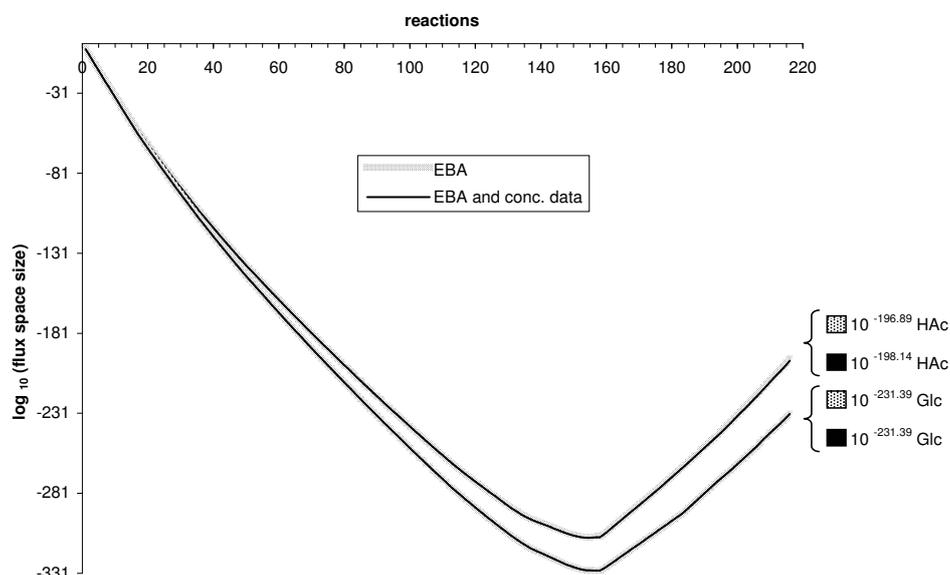
Figure 4.16: **Comparison of the flux space developments under *EBA* conditions alone and *EBA* including metabolic profiling deduced constraints**

A comparison of the flux space development (see 3.10.1.1) of both carbon sources (NaAc = acetate, Glc = glucose) under *EBA/EEB* conditions excluding and including combinatorial constraints. The whole data set has been sorted for increasing variabilitites of acetate feeding under *EBA/EEB* constraints. Numbers given to the right are absolute values of upper estimates of flux space sizes in units of $\left[\dfrac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}} \right]^{216}$. For details about the constraints imposed, see table 4.11 and 4.15.

4.4.4 External energy balance analysis

The results for the external energy balance analysis do not differ from the ones shown in table 4.12 on page 87. In other words: by imposing combinatorial constraints, derived from metabolic profiling data, the external fluxes do not change at all.

Table: 4.16: **Knock out study results**

Overview of the results of a knock out study using single knock out mutations.

No. of		mutations that ...
65	...	do not affect the ability to produce biomass
3	...	only make biomass production impossible for *glucose* feeding conditions, but still allow a thermodynamically feasible flux distribution[*]
123	...	make biomass production impossible for *glucose* and *acetate* feeding conditions, but still allow a thermodynamically feasible flux distribution[*]
17	...	do not allow any feasible flux distribution (classified as *lethal*)

[*]the model is unable to produce biomass, but can still metabolize carbon into CO_2

4.5   Knock out study

We performed an energy balance analysis (see 3.4) for all possible single knock out mutants of the modeled organism, as described in section 3.11. Table 4.16 shows an overview of the mutations' effects. A detailed listing of the mutations that effect those enzymes whose referring genes are weakly expressed (compare results in table 4.3 on page 68) is given in the following section.

The constraints imposed during this study are the ones listed in table 4.11 on page 83. For comparability with the results of the wild-type analyses, during the *EBA*, the initial irreversibilities determined for the *C. glutamicum* wild type (table 4.3.3.2, p. 83) were used.

4.5.1   Knock out study details

Table 4.17 shows the effects of the knock outs of all genes that are only weakly expressed within *C. glutamicum* (compare table 4.3, page 68).

For all knock outs listed in table 4.17, there still is a possibility to metabolize inflowing carbon into CO_2, although the ability to produce biomass (i.e. to

Table: 4.17: **Knock out study details - weakly expressed genes**
Effects of knock out mutations affecting genes which are only weakly expressed, according to DNA microarray experiments (see table 4.3 on page 68).

Enzyme	Reaction(s)	Effect on biomass production
1.1.1.1	reaction 135	no effect*
1.2.1.11	reaction 85	growth impossible†
1.2.1.59	reaction 11	no effect*
2.1.2.3	reaction 175	growth impossible†
2.1.3.2	reaction 182	growth impossible†
2.3.1.12	reaction 38	no effect*
2.7.1.71	reaction 107	growth impossible†
2.7.4.9	reaction 187 reaction 188	growth impossible†
3.1.3.12	reaction 151	no effect
4.2.3.5	reaction 104	growth impossible†
5.4.99.16	reaction 146	no effect

*there are other -stronger expressed- enzymes attached to the reaction
†the model is unable to produce biomass, but can still metabolize carbon into CO_2

grow) is destroyed in some cases.

The effects of theoretically knocking out those enzymes which are not annotated (see tab. 4.4, p. 69) are shown in table 4.18.

For most of the reactions not attached to a catalyzing enzyme (compare table 4.5 on page 70) there is no evidence these reactions are in need of an enzyme. The only reaction that might require enzymatic help is reaction 144:

$$\text{Sedoheptulose 7-phosphate} + \text{ATP} + \text{H}+ \quad \leftrightarrow \quad \text{ADP-D-glycero-D-manno-heptose} + \text{Pyrophosphate}$$

The knock out of the potential enzyme catalyzing the above reaction would lead to the incapability to grow, but the ability to metabolize carbon into

CO_2 would remain.

Of further interest are the 17 knock outs, that are classified as *lethal* to the organism, i.e. that prevent any valid flux distribution. They are listed in table 4.19. As can be seen there, it is vital that most of the metabolites which are modeled to cross the membrane also are allowed to do so. Closing the passage for H_2O, O_2, CO_2 and H^+ will be as lethal as preventing the carbon sources from entering. Inhibition of the other common external fluxes (P_i, SO_4^{2-}, NH_3) leads to the inability to grow while still carbon consumption in combination with carbon dioxide production is possible.

All other knock out mutations with deadly effects are affecting internal reactions. They will be discussed on page 123 ff.

Table: 4.18: **Knock out study details - non-annotated enzymes**
Effects of theoretical knock out mutations affecting genes which have not been annotated (compare tab. 4.4, p. 69). *Growth impossible* refers to the incapability to produce biomass, but a remaining ability to metabolize inflowing carbon into CO_2. *Lethal* represents the complete lack of the capability to metabolize inflowing carbon.

Enzyme Number	Reaction(s)	Effect on biomass production
1.3.5.1	part of reaction 198	lethal
1.3.99.11	reaction 183	growth impossible
1.6.1.1	reaction 200 part of reaction 196	growth impossible
1.6.1.2 & 1.8.1.9 & 1.17.4.2	reaction 140 reaction 141 reaction 142 reaction 189	growth impossible
1.8.2.1	part of reaction 196	growth impossible
2.3.1.1	reaction 96	no effect
2.3.1.157	reaction 157	growth impossible

Continued on next page ...

Table: 4.18 . . . Continued from previous page

2.5.1.55 & 5.3.1.13	reaction 143	growth impossible
2.6.1.2	reaction 45	growth impossible
2.7.1.3	reaction 7	no effect
2.7.2.12	reaction 42	no effect
2.7.7.38	reaction 139	growth impossible
3.1.3.15	reaction 61	growth impossible
3.1.3.45	reaction 138	growth impossible
3.5.4.9	reaction 194	growth impossible
3.6.1.5	part of reaction 197 part of reaction 198	lethal
3.6.1.8	reaction 158	no effect
4.1.1.3	reaction 39	no effect
5.4.99.5	reaction 105	growth impossible
6.3.4.16	reaction 46	growth impossible

Table: 4.19: **Knock out study details - lethal knock outs**
All knock out mutations that prevent *any* valid flux distribution.

Enzyme Number	Reaction(s)
Membrane transport reactions	
-.-.-.-*	reaction 226, 227, 229, 232, 233 & 238
Internal reactions	
1.1.1.37	reaction 26
1.3.5.1	reaction 198
1.3.99.1	reaction 29 & 198
1.9.3.1	reaction 197 & 198
1.10.2.2	reaction 196, 197 & 198
2.3.1.8	reaction 40
2.3.3.1	reaction 27
2.7.1.2	reaction 1
3.6.1.5	reaction 197 & 198
4.2.1.2	reaction 30
4.2.1.3	reaction 31

*Membrane transport has been modeled without enzymatic help

Chapter 5

Discussion

5.1 The *Corynebacterium glutamicum* metabolism model

The model we constructed and used afterward contains all reactions that constitute the central carbon metabolism. Also all pathways necessary to metabolize supplied carbon sources into biomass are present; so the model is able to grow (see results section 4.1). Comparing this capability with the data shown in table 4.2 on page 68, we have to say that the model, on the one hand, contains all important reactions. Since there are only 302 out of 2489 annotated genes used within this model, on the other hand, we may not call this a genome scale model.

A closer look at the model shows a detailed modeling of the central carbon metabolism. Namely, this includes the glycolysis, the pentose phosphate pathway, and the citric acid cycle. Examining the pathways that produce amino acids, fatty acids, and other biomass components, the level of detail declines. The network supplies the ability to produce biomass, but lacks the provision of variances to do so.

Nevertheless, to our knowledge this is the largest model of the *Corynebacterium glutamicum* that has been constructed so far. Examples of *C. glu-*

tamicum reaction networks built before have been published.

Wendisch and his coworkers measured fluxes in the central carbon metabolism in (Wendisch *et al.*, 2000) and thereby used a model established in the same research group before (Marx *et al.*, 1996). This model only describes the central carbon metabolism and contains 16 reactions.
Kiefer, Wittmann and Zelder published experimentally determined flux distributions during growth on glucose or fructose together with Elmar Heinzle in (Kiefer *et al.*, 2004) . Later in 2004 they published analyses using a similar model during growth on sucrose (Wittmann *et al.*, 2004). In the first cited article they describe the model as containing 42 reactions in total. The largest model constructed by now has been published by Takaç *et al.* in 1998 in (Takaç *et al.*, 1998). It consists of 115 reactions. But one has to be aware, that this model has been designed far in advance of the publishing of the *C. glutamicum* genome, which was in 2003 (Ikeda & Nakagawa, 2003; Kalinowski *et al.*, 2003). It was based on the *E. coli* metabolic network and described by the authors as "a comprehensive metabolic network [...] proposed for glutamic acid bacteria and used in a stoichiometrically based flux balance model for L-glutamate production" (Takaç *et al.*, 1998, Abstract, 1st sentence).

In comparison to other organisms modeled, we might have a look at *Escherichia coli* first. The *E. coli* genome (Blattner *et al.*, 1997) has a similar size, compared to *C. glutamicum*: 4288 protein coding genes against 3002. But, since more than "Thirteen Years of Building Constraint-Based In Silico Models of *Escherichia coli*" (Reed & Palsson, 2003) have passed, the corresponding *in silico* models grew up to a network of 931 reactions (Reed *et al.*, 2003), whereof 720 have already been published by Edwards and Palsson in 2000 (Edwards & Palsson, 2000).
Other organisms, genome based metabolic networks have been modeled for, are *Haemophilus influenzae* Rd. (Edwards & Palsson, 1999; Fleischmann

Table: 5.1: **Examples of genome based models**
An overview of the sizes of genomes and *in silico* metabolic networks of some organisms. Except yeast, these all are prokayotic bacteria. Annotations: 1) Kalinowski *et al.* (2003), 2) Blattner *et al.* (1997), 3) Fleischmann *et al.* (1995), 4) Tatusov *et al.* (1996), 5) J.F. *et al.* (1997), 6) Mewes *et al.* (1997).

Organism	Base pairs	protein-coding genes	annotated genes	Enzymes	Reactions in model	
C. glutamicum	3282708	3002	2489	197	234	[1]
E. coli	4639221	4288	2659	904	931	[2]
H. influenzae	1830137	1749	1610	293	488	[3,4]
H. pylori	1667867	1590	1091	341	554	[5]
S. cervisiae	$\geq$ 12 Mill.	6000	2598	708	1175	[6]

et al., 1995), *Helicobacter pylori* (Thiele *et al.*, 2005; Schilling *et al.*, 2002; J.F. *et al.*, 1997) and also *Saccharomyces cervisiae* (Förster *et al.*, 2003; Mewes *et al.*, 1997) as an example of an eukaryotic organism. Table 5.1 gives a comparative overview of the genome and model sizes.

5.1.1 DNA microarray analyses data

We used the data derived from DNA microarray analyses by Silberbach *et al.* (Silberbach *et al.*, 2005; Silberbach, 2005) to justify and improve the *C. glutamicum* network. But there are some issues which have to be considered when supporting the model of a metabolic network by microarray experiments.

- A researcher can make a good confirmation of gene-expression (see 2.2.1) by finding values that definitely rise above the experimental noise. But a value inside the noise does not necessarily represent a non-existent expression. In other words: an apparently non-existent transcription does not proof a non-existence of the coded protein.

- Transcription $\neq$ translation. A gene transcribed into mRNA and a piece of mRNA translated into a protein are separated steps, which

occur independent of each other. Thus, measuring the transcription only, provides *hints* which enzymes might exist inside the cell, but is *no proof* of it. In other words: An apparently present transcription does not proof the availability of the referring protein.

Despite the items stated above, DNA transcription levels may very well be used as a clue to whether an enzyme exists or does not.

First, although there are some very stable RNA molecules, usually RNA is metabolized rather fast. Therefore, a very low transcription rate may indeed be interpreted as a low protein expression, because nearly non-existing RNA obviously cannot be translated into protein. Second, a high occurrence of RNA hints to an ongoing translation. In consequence this may be interpreted as the protein being expressed.

It shall be emphasized that we did *not* use the DNA microarray analyses as a proof of *protein* expression. We considered them to be hints which should be followed further. Concerning the data pointing to weakly expressed genes, we analyzed the necessity of the referring enzymes in the knock out study (see section 3.11 for methods and 4.5 for results).

5.1.2   Non-annotated reactions

The reasons for implementing non-annotated reactions have been discussed in section 3.2.4. Either we needed to close gaps in obvious pathways or membrane transport had to be modeled. We distinguished two types of non-annotated reactions. On the one hand, reactions exist which are occuring spontaneously and are not catalyzed by an enzyme at all. On the other hand, we have reactions being enzyme dependent, but the referring enzyme is not annotated from the *C. glutamicum* genome.

Concerning the first, there is no reason not to use these reactions in our model, because they may occur as soon as the transformed metabolites are present. The latter ones, on the other hand, have to be examined more

closely. This has been done in context of the knock out studies; the referring discussion starts on page 120.

5.1.3 Biomass constitution

The reader should notice, that the components that create biomass and are listed in table 4.6 (p. 71) are considered to be precursors, not the biomass itself. Takaç and colleagues defined a biomass composition that they used for the analysis of glutamic acid producing bacteria (Takaç *et al.*, 1998). We adapted this slightly, according to the knowledge we have, based on the bacterial genome (Ikeda & Nakagawa, 2003; Kalinowski *et al.*, 2003). Within the biomass used in the work at hand, there exist precursors to build nucleic acids, proteins, the bacterial plasma-membrane, and the cell wall.

The substrates used to produce nucleic acids can be separated into these leading to DNA: *dATP, dCTP, dGTP, dTTP* and those leading to RNA: *ATP, CTP, GTP, UTP*. All amino acids listed in table 4.6 are included as protein constituents. Further on, the bacterial membrane can be build from fatty acids and lipopolysaccharides. The fatty acids are represented by *myristic* and *myristoleic acid* as well as by the generic "fatty acid", which is a statistical mean of saturated and non-saturated fatty acids. *ADP-D-glycero-D-manno-heptose* and *CMP-3-deoxy-D-manno-octulosonate* are employed to synthesize lipopolysaccharides. The outermost layer of the hull of gram-positive bacteria is the cell wall. The molecules *meso-2,6-Diaminoheptanedioate, UDPglucose, UDP-N-acetyl-D-glucosamine* and *UDP-N-acetylmuramate* are listed as the necessary precursors.

The biomass published in (Takaç *et al.*, 1998) is based on the composition determined for *E. coli*. As already mentioned, we adapted this based on the genome. Additionally, even though the DNA and the protein network of the two prokaryotic bacteria are different, the statistical distribution of the biomass precursor molecules can be assumed to be the same.

It shall be mentioned at this point, that a more detailed biomass composition

exists for *E. coli* and can be downloaded from the website of the research group of Bernhard Palsson (Palsson, 2005). It is part of the sophisticated model published in (Reed *et al.*, 2003).

Studies concerning the robustness of the network against changes in biomass constitution would go beyond the scope of this work. But Varma and Palsson published corresponding studies in 1995 for the *E. coli* bacterium (Varma & Palsson, 1995). They remarked a robustness of the flux balance analysis results against minor changes in biomass constitution. We may also claim, that we were able to reproduce experimental results with our model, using the biomass constitution given in table 4.6 on page 71.

5.2   Model validation

The method to validate the constructed reaction network has been described in section 3.9 on page 58 ff. In table 4.7, page 72, it is shown, that we used the *EBA* method (chapter 3.4) while affirming the model. Obviously, it has to be able to reproduce real-life situations with thermodynamically necessary constraints being imposed.

By using restrictions derived from the work of Wendisch and colleagues (Wendisch *et al.*, 2000), after preceding adaptions[1] (see 3.9.1), we have been able to reproduce the measured flux distributions within a range of $\pm 30\%$. This value includes experimental errors as well as shortcomings in the comparison of two models not being identical. Volker Wendisch and his coworkers deduced metabolic fluxes from the enrichment of labeled carbon atoms (^{13}C) in various metabolites. Thereby they used a model of the central metabolism of the *Corynebacterium glutamicum*, which contains 16 reactions. Experimental errors can be found during the determination of growth rates, biomass

---

[1] The adaptions we made result in a simple factor, which is the same for all fluxes. Thus, they do not change the relative flux distributions found.

yields, carbon consumption rates and ^{13}C enrichment. Details are given in (Wendisch *et al.*, 2000).

Taking the given arguments into account, we consider an accuracy of $\pm 30\%$ close enough to experimentally determined values to claim our model validated.

5.3 Using laboratory conditions

Ideally, one would know the growth rates as well as the amount of carbon that is metabolized into CO_2 by the organism. We only had the growth rates at disposal. Thus, we had to use a published ratio between biomass and carbon dioxide production. It seemed reasonable to adapt that ratio from the same literature we used before to verify our model (Wendisch *et al.*, 2000). The assumption here is, that a decrease or increase of the growth rate is founded in a referring change of the carbon consumption. We do not think, that major changes of the biomass production rate may be following a greater change of the ratio between CO_2 and biomass creation, along with a constant carbon influx.

The determination of the experimental growth rates shall be discussed next.

5.3.1 Identification of periods of exponential growth

The bacterial cultures raised on acetate grow faster than the ones raised on glucose, during their periods of exponential growth, as can be seen from table 4.8 on page 79. The difference between the mean values is approximately 30 %. Even though only two values are available per nutrient, we consider this enough to be taken as a fact, not a coincidence that may change in a greater data set.

In case of glucose as a carbon provider, a smaller deviation between the values can be observed, than in case of acetate. This, on the contrary, may very well be an artifact, which probably vanishes in a greater data set.

It needs to be mentioned, that the identification of the periods of exponential growth is more substantiated in case of glucose. The reader is referred to figures 4.6 to 4.9 on page 77 ff. It can be seen that especially the start of the interesting phase is hard to identify in those graphs where acetate is used as energy source. In consequence, we used a wider range to identify the starting points, trusting the method described in section 3.6.4.1 (page 51).

5.3.2 Flux Balance Analysis

In figure 4.10 (page 80) we observe around 80 fluxes that touch the upper limit imposed on all fluxes. This behavior indicates them being unbounded[2]. It stems from the possibility of biochemical information - basically the type and arrangement of chemical bonds - flowing in circles.

To rearrange chemical bonds needs free enthalpy (ΔG), if this process is to occur spontaneously (compare 2.1.1.2). Thus, a cyclic flux has to be connected to a loss of ΔG, which is not considered during a flux balance analysis. Otherwise cyclic fluxes would not be possible.

As a consequence, using a flux balance analysis approach in combination with a completely reversible reaction system yields results which are not realistic. We observed exactly what Daniel Beard, Shou-dan Liang and Hong Qian discussed in (Beard *et al.*, 2002). *FBA* alone is not sufficient to obtain thermodynamically feasible results. Accordingly, we implemented the energy balance analysis, which is described in the above cited article among others (Beard *et al.*, 2004; Qian *et al.*, 2003; Yang *et al.*, 2005; Beard & Qian, 2005).

5.3.3 Energy Balance Analysis

Overviewing the energy balance analysis of the *Corynebacterium glutamicum* metabolism we can claim to be able to predict a meaningful flux vector; "meaningful" in the sense of thermodynamics. What initial information had

---

[2]The mentioned limits are arbitrarily set, for the only reason that the optimizer would fail, if they were infinity.

to be given to achieve this and the comparison to the results of the flux balance analysis will be discussed next.

5.3.3.1 *EBA*- initial irreversibilities

Initial irreversibilities have to be given for 24 reactions. Otherwise, the energy balance analysis would not yield a thermodynamically feasible flux distribution, but would still allow cyclic fluxes. From a systemic point of view, only one possible cyclic flux would corrupt the whole flux vector. A possible constriction of some reaction directions may lead to a reduction of the amount of unconstrained reactions, followed by a reduction of the flux space size. But, if we are looking at the whole cell at a time, either the flux distribution is thermodynamically sound or it is not. And if there are fluxes left which are unbounded, then it is not a feasible solution.

We started to identify the reactions to be irreversible from the beginning by calculating the matrix of all internal cycles (2.7). A subsequent prediction of feasible reaction directions revealed which cycles were still left. The reaction most common in these cycles was identified and its irreversibility information retrieved from the KEGG database (Kanehisha & Goto, 2000). Again, a prediction of feasible reaction directions followed. The procedure was applied iteratively until cyclic fluxes were completely prevented. A more detailed description can be found in section 3.4.2.
Thus, the choice of reactions, we used the reversibility information of, is only dependent on the procedure, which ensures the usage of as few initial information as possible while providing the necessary effect. There is no biological meaning whatsoever attached to it.

5.3.3.2 *EBA*- robustness analysis and comparison to *FBA*

The constraints on the external fluxes during the energy balance analysis (table 4.11, page 83) have to be the same as those being imposed during the flux balance analysis (table 4.9, page 80). By inspection of figure 4.11,

page 84, a much more detailed distribution of flux variances than in the refer-ring *FBA* figure 4.10 is recognizable. A very important fact is the complete lack of unconstrained fluxes. The one flux which is able to vary in an inter-val greater 200 has its limits at -273.7 and +38.7. This is the flux through a phosphor interchange-reaction between NADH and NADPH (reaction 200).

Figure 4.12 on page 85 shows a direct comparison of the variabilities of all fluxes, *FBA* versus *EBA*. The comparing method is explained briefly on page 85. It can be observed that about 70-80 reactions are much more constrained after taking thermodynamic feasibility into account.   When analyzing that figure, one has to be aware of the following:
The variability of the highly variable *FBA* fluxes is only dependent on the upper and lower bounds imposed. Remembering that these are arbitrary (compare 4.3.2), the amount of reduction from *FBA* to *EBA* is insignificant. Significant is only the fact *that* a lot of fluxes are further constrained and the *amount of them.*

The last figure comparing the flux and energy balance analyses is figure 4.13 on page 86, which shows the flux space development as described in 3.10.1.1. It illustrates that energy balance analysis indeed scales down the flux cone in a large amount. The estimated size drops from $10^{93.22}$ for acetate feeding and $10^{74.04}$ for glucose feeding to $10^{-196.89}$ and $10^{-231.39}$, respectively.
For both types of analysis, the reactions are sorted the same way. Thus, we may deduce from the graph that those reactions which are already bound under *FBA* conditions have the same variabilities under *EBA* conditions. This is because the lines exactly overlap in the left part of the figure. Also, it is obvious that the reactions which seem to be unbounded under *FBA* (right part of the graph) are much more constrained, when thermodynamical aspects are considered (*EBA*).
Finally, for both methods one can compare the behavior of the metabolism when it is fed with different carbon sources. In case of *FBA* as well as in case

of *EBA* the total size of the solution space is smaller if the bacterium grows on glucose. The difference grows from *FBA* to *EBA*. This is understandable, since all *FBA* reaction variabilities on the right side of figure 4.12 have the same maximal value (for acetate and glucose conditions)[3]. Thus, the development of the acetate related solution space parallels the one related to glucose, when looking at these unbounded reactions. Contrary, the *EBA* variabilities develop differently, allowing the gap between the size of the acetate and the glucose flux cone to increase.

Summed up, we definitely observe a large confinement, when applying constraints which meet the thermodynamic requirements of an energy balance analysis against those referring to *FBA* alone.

5.3.3.2.1 A closer look Contrary to flux balance analysis, during an energy balance analysis cyclic fluxes are prevented (compare 2.7). If matter is allowed to flow in closed loops, the flux through these loops may rise up to infinity without breaking any of the *FBA* constraints. But even when prohibiting cyclic fluxes, high variabilities still may occur.

As said above, the most variable flux belongs to reaction 200, which is a shuttle between NADH and NADPH:

$$\mathrm{NADPH + NAD^+ \leftrightarrow NADP^+ + NADH}$$
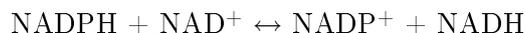
According to table 4.4, the *NADH transhydrogenase* (1.6.1.1), which catalyzes this reaction, is not annotated from the genome. But a knock out of it leads to the incapability to grow (compare table 4.18). Thus, it was necessary to include the enzyme into the model.
The reason for the high variability of reaction 200 will be another possibility to perform its transformation, hidden within the network. This will probably

---

[3]As mentioned before, this maximal value is arbitrarily set to prevent programming related problems.

be a chain of reactions that cycles carbon while shifting the phosphor. Thermodynamically, this is not breaking the laws constituting the energy balance analysis, since there is in fact a reaction attached to that "cycle". Another thing worth considering is: the system might direct a higher carbon flow through the pentose phosphate pathway. That will oxidize one carbon to CO_2 while producing two molecules of NADPH. The alternative is, to direct that carbon through the glycolysis and produce NADH, instead. Later on, the carbon would be oxidized during the citric acid cycle. Depending on the pathway taken, the NADH/NADPH shuttle will work the desired direction.

The next most variable reaction lies in the heart of the glycolysis and describes the phosphorylation of fructose-6-phosphate to fructose-1,6-bisphosphate by use of ATP (reaction 3). But, there is a reaction (no. 4) that can restore fructose-6-phosphate from fructose-1,6-bisphosphate while setting free one of the phosphates. Summed up, those two reactions can be used to split ATP into ADP and phosphate (compare figure 5.1).
The same can be achieved by summing up:

reaction 158	ATP	+	H_2O	$\leftrightarrow$	AMP	+	PP_i
reaction 160	ATP	+	AMP	$\leftrightarrow$	ADP	+	ADP
reaction 199	PP_i	+	H_2O	$\leftrightarrow$	P_i	+	P_i
$\sum/2$	ATP	+	H_2O	$\leftrightarrow$	ADP	+	P_i
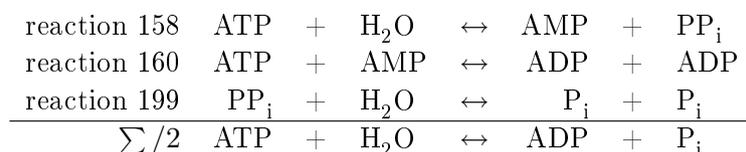
Thus, there exist two alternate pathways to decompose ATP. Hence reaction 3 is highly variable. If the situation depicted in figure 5.1 stated the only possibility to dephosphorylate ATP, a certain flux through the futile cycle[4] would always be necessary. In consequence the variability of reaction 3 would be lower.
The flux through reaction 3 & 4's substitutes may rise up to nearly half the value of reaction 3. This is reasonable, because the latter one needs to run two times to achieve the same as its alternative pathway. That this

---

[4]The example shown in figure 5.1 is called a futile cycle. That is a cyclic flux that serves the decomposition of some metabolites, attached to the cycle, which are not needed by the cell.
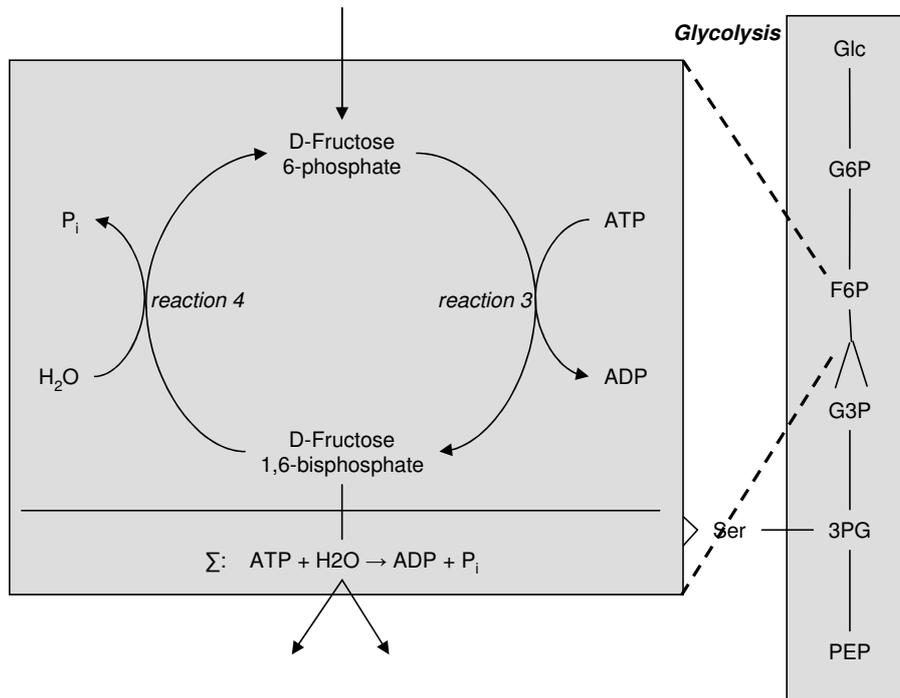
Figure 5.1: **Phosphorylation of fructose 6-phosphate**
If the reactions 3 & 4 run in the directions shown, their sum will be the hydrolysis of ATP. A cyclic flux that serves the decomposition of some metabolites, attached to the cycle, which are not needed by the cell is called a *futile cycle*.

alternative way does not reach up to exactly half the value tells us there is a flux that *has to* go through reaction 3; very probably, this is the carbon flowing through the glycolysis.
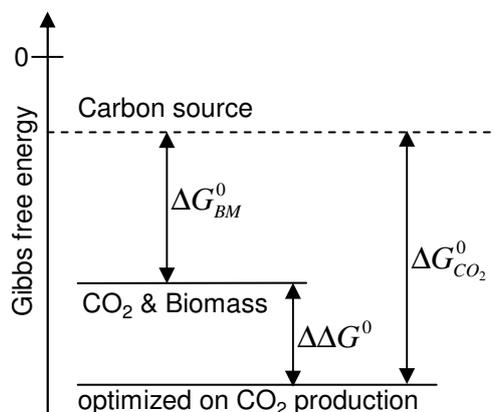
Figure 5.2: **Values used in an *EEB***
$\Delta G^0_{CO_2}$ is considered to be the maximal energy the organism can gain from the carbon source; it is connected to the optimization of the model on CO_2 production. ΔG^0_{BM} is the energy dissipated if the organism produces biomass (BM) and CO_2. Accordingly, $\Delta\Delta G^0$ is the amount of energy the bacterium transfers into biomass.

5.3.4 External energy balance analysis

The results, this part of the discussion refers to, are shown in table 4.12 on page 87. As a reminder (compare section 3.5.1): we regard the amount of heat that is set free during the total conversion of inflowing carbon into CO_2 as the maximal amount of heat that is hidden in the metabolites entering the system. The bottom line of table 4.12 shows how much of this energy is converted into biomass for the two carbon sources analyzed throughout this work. Figure 5.2 demonstrates the coherences.

Interestingly, even though the *Corynebacterium glutamicum* grows faster if raised on acetate than on glucose (see table 4.8, p. 79), it is able to make better use of the latter. Ca. 40 % of the available energy are transferred into biomass, contrary to about 30 % in case of acetate. Examined from another angle, raised on glucose, the bacterium dissipates about 60 % of the available energy, whereas approximately 70 % are dissipated if raised on acetate.

5.3.5   General considerations

If we constrain all external fluxes to specific values, they obviously do not generate variable internal fluxes. In this case, variability generally appears if there is more than one possible pathway able to accomplish a certain task. In case of the studies at hand, we did fix the amount of carbon flowing into the cell. But we allowed a certain tolerance to the way this carbon may leave the system (compare tables 4.9 and 4.11). This adds up to the possible variabilities already present. Taking the stoichiometric factors (which may even be fraction numbers) into account will eventually end up with the variabilities summarized in figure 4.11. The correlations are also explained in chapter 2.6 and are shown in figure 2.6 on page 22.

In (Mahadevan & Schilling, 2003), Mahadevan and Schilling performed an analysis, similar to our robustness analysis (3.7), for the *E. coli* metabolic network published by Edwards and Palsson in (Edwards & Palsson, 2000); concerning the *E. coli* model also compare 5.1. Contrary to the analyses throughout this work, they fixed the carbon influx and biomass outflux to certain values. Next, they minimized and maximized each flux throughout the network. Doing so, 29 reactions showing a higher variability than $10^{-5}\left[\frac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}}\right]$ for *E. coli* being raised on glucose were found. Using acetate and lactate as carbon sources, they found 19 and 28 variable reactions. When comparing these results to the ones shown in sections 4.3.2 and 4.3.3, one must be aware of the following:

- As said above, Mahadevan and Schilling fixed the value of the biomass formation rate, while we demanded it to be $\geq 90\,\%$ of its maximum.

- The metabolic network constructed by Palsson and his colleagues uses irreversibility information based on biological databases. Additionally, Mahadevan and Schilling eliminated some reactions (or set irreversibility constraints) to remove futile cycles (see fig. 5.1, p. 111).

Due to the first item, we will get greater flux intervals for all fluxes, as explained above. We required a $90\,\%$ biomass production to allow a certain flexibility between biomass and CO_2 production. Since we adapted the biomass / CO_2 ratio from literature (compare 3.8.1), this takes experimental errors into account (see also section 5.2). Due to the second item, Mahadevan and Schilling probably received a smaller solution space. The more constraints one imposes on a constraint-based analysis, the smaller the solution space becomes. Contrarily, we used the minimum amount of irreversibility constraints needed, by applying energy balance analysis. First, this makes the results shown in the thesis at hand *ab initio* results, with the exception of the initially given irreversibilities (compare 4.3.3.2). Second, elimination of futile cycles, as performed by Mahadevan and his colleague, is not enough to ensure thermodynamic feasibility, as Beard, Liang, and Qian explain in (Beard *et al.*, 2002). Centralized, we may have slightly higher variabilities than expected when comparing to related publications, but we only use as few information as necessary to predict thermodynamically feasible flux distributions.

5.4 Metabolic profiling

The main idea of the work at hand is the development of a method to implement data from metabolic profiling experiments (compare 3.6) in order to further constrain the metabolic flux space. Thus, we will at first discuss the assumptions made in advance of the implementation. The profiling data and the results obtained after implementing them will be considered next. Thereby, it will become clear that we were indeed able to invent a sound method which leads to the expected reduction of the metabolic flux space.

5.4.1 Prerequisites and Assumptions

This section refers to the assumptions that have been laid out in 3.6.2, page 45. First we regarded the concentrations of all metabolites whose profiles could not be measured as constant under all circumstances (p. 45). Since we used the profiles (*NaAc* and *Glc*) showing the highest correlation (compare section 3.6.4.3), this is a reasonable hypothesis. Ongoing research in the field of experimental metabolic profiling should decrease the amount of unknown concentrations in the near future. Second, we assumed linear enzyme kinetics (3.6.2.2). Concerning the four points stated on page 45 the following can be said:

- the order of the reaction in each metabolite equals the corresponding stoichiometric coefficient:
  we consider the influence of all molecules on the reaction rate to be the same. Even though a connection between the reaction's mechanism and its order in the individual educts is given, since we do have no information about the mechanism, we assign the same importance to each molecule.

- enzyme expression is the same in all networks (here: *Glc* & *NaAc*) considered:
  this is one of the assumptions that underlie the whole modeling process

116

(see 3.2.1). Thus, it is a logical consequence to assume this for all networks considered.

- enzyme kinetics is far from saturation:
  suppose we had a saturated enzyme in our network. Then its substrates would have accumulated. This is not conform to the assumption of a steady state flux distribution, which is the basic idea of *FBA* (see 2.6).

- there are no effects changing the reaction kinetics:
  flux balance analysis as well as its enhancement energy balance analysis is a systemic method that is not dynamic. In consequence, effects like inhibition, enzyme-disintegration, and others are supposed to be covered by the analysis itself. These are properties that cannot be modeled on a local scale while using flux balance methods.

In closing it can be said, the assumptions made are as well reasonable as necessary for the analyses.

5.4.2 Metabolic profiling data

The metabolites which could be used to further constrain the flux space (compare 3.6.5), as well as their measured concentrations, are shown in table 4.13 on page 89 . They gather from various parts of the metabolism and no focus on any special pathway can be assessed. Thus, the effects of the new conditions also affect the whole model and results will be seen globally, rather than locally.
Table 4.14, page 90, shows the reactions that are affected by the data. The first thing to be noticed is, that there may be more than one reaction affected by each metabolite. On the other hand, the constraints that could be deduced from the metabolic concentration data of *valine, alanine, proline, phenylalanine* and *fructose* did not lead to any change in the possible flux patterns.

The reason is simple: if the new, combinatorial requirement is

$$v_{NaAc} \leq v_{Glc}$$

and the bounds on v_{NaAc} and v_{Glc} are

$$a \leq v_{NaAc} \leq b$$
$$\text{and}$$
$$a \leq v_{Glc} \leq c$$
$$\text{with}$$
$$b \leq c$$

before adding the new constraint, then no effect can be observed. The original bounds a, b and c hold true even after imposing the new restriction. While interpreting this, we must be aware that - using the robustness analysis (3.7) and the flux space development graphs (3.10.1.1) - no effect can be *observed*. But there *is* an effect hidden, nevertheless.
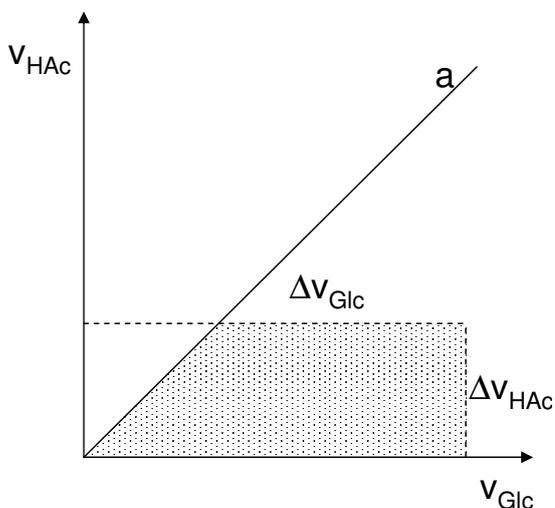
Figure 5.3: **Hidden effect of combinatorial constraints**
If the constraint is $v_{NaAc} \leq v_{Glc}$, then Δv_{NaAc} and Δv_{Glc} remain. But the *real* possible solution space left, is only the dotted area, in contrast to the dashed rectangle, that constituted it before.

Figure 5.3 clarifies the issue. The possible ranges Δv_{NaAc} and Δv_{Glc} can still be used to full capacity. Thus, the results of the analyses mentioned above

show no difference. But the flux space is in fact restricted. The reason why these hidden effects are not analyzed further is simply a matter of calculation time. We would have to fix one of the fluxes and test the other's variability against different fixed values. And since there may be more than two fluxes influencing each other, we would have to do this analogously for 3, 4, 5, ... all fluxes.

The combinatorial constraints, derived from the metabolic profiling data, and their effects are discussed within the context of the energy balance analysis in the following section.

5.4.3 Energy balance analysis excluding and including combinatorial constraints

In figure 4.14 on page 92 the results of an energy balance analysis under the constraints of table 4.15 (page 91, incl. combinatorial constraints) are shown. A closer comparison of this figure to the *EBA* figure derived without the additional conditions (fig. 4.11, p. 84) reveals that the flux variabilities referring to the acetate fed metabolism are further constricted, while the ones connected to glucose as a carbon source remain unchanged. These differences are explicitly expressed in figure 4.15 on page 93.

We see around 25 % of all reaction variabilities shrink to a different degree when the new constraints are imposed on the system. This means, by considering further thermodynamical issues (compare section 3.6.1), we are able to reduce the possible flux space.

This reduction is also visible in the flux space development graph 4.16 on page 94. The values at the right show a reduction of the upper estimate of the flux space size (see 3.10.1) by a factor of ca. 100. More precisely, the solution space is diminished to 5.25 % of its original value. As said in the previous section, one also has to take into account that there is a further

reduction hidden within the constraints imposed (see figure 5.3).

In conclusion it can be said that we achieved our goal of reducing the steady state flux cone by implementing metabolic profiling data (compare introduction, page 3). To accomplish that, we developed a new concept which can be combined with any other enhancement of the flux balance analysis (e.g. *EBA*). Linear programming only is needed as a means to carry out the computing, circumventing the problems usually attached to nonlinear programming problems (compare 2.7).

5.4.4   External energy balance analysis

The fact that external fluxes do not change when applying additional constraints (see 4.4.4, page 94) very probably is related to the limitedness of the model. Since the only carbon-related reactions crossing the cellular membrane are transporting the carbon source, CO_2, or biomass, we do not expect any change therein as long as we do not change the carbon influx.
This behavior will very probably change when opening more systemic gates that allow carbon to cross the cellular membrane.

5.5   Knock out study

From table 4.16 (page 95) we learn, that the *C. glutamicum* metabolism is robust against a number of knock out mutations, referring to the ability to grow. Considering our model, 65 of the 302 genes may be knocked out and it will still be able to produce biomass. On the other hand, there are 123 knock out mutations, that bar the system from growth, while still valid flux distributions exist. These remaining possibilities to metabolize inflowing nutrients channel all carbon into the production of carbon dioxide. Further on, there exist 17 knock outs we classify as *lethal*, meaning they circumvent *any* valid flux distribution.
The mutations that only leave the ability to produce CO_2 obviously prevent

the growth of the model. In terms of biology, this is lethal. But, the metabolism keeps the ability to gain energy from inflowing nutrients (it is able to degrade them to CO_2). It "just" is not able to produce biomass out of them. Thus, a more detailed model or a slightly altered biomass composition might restore the power to grow. Knock out mutations classified lethal inhibit any flux distributions. Their damage is severe and - probably - cannot be easily overcome.

Nevertheless, generally speaking it may be presumed that a more detailed model will result in more alternate pathways, which are able to fulfill the same systemic task. As a consequence, the robustness against genes becoming unavailable should rise. Also, right now we are only observing three types of consequences after gene knock outs. Either the capability to grow remains unchanged or the system completely lacks it. In the latter case either the model still provides active pathways, which are able to metabolize inflowing carbon into CO_2, or the mutations are lethal. A more detailed model might soften this hard separation. In real life organisms a mutation which makes a gene unusable may *reduce* the growth rate, but does not have to be either lethal or without effect. Imaginable is even a biomass production *in*creasing effect of a knock out mutation if it entails the disappearance of an inhibitor. But this behavior cannot be simulated using the model at hand.

5.5.1 Knock out study details

Some of the genes used to construct the *C. glutamicum* model shall be discussed in detail. On the one hand, those genes which are only weakly expressed (4.1.1) need to be examined. On the other hand, we have to take a closer look at the enzymes which have no annotated gene attached (4.1.2). It also seems reasonable to explore the reactions that have no catalyzing enzyme assigned. Of further interest are finally those knock outs that are lethal to the organism.

Weak gene-expression

Table 4.17 on page 96 shows 11 enzymes whose genes are only weakly transcribed according to the DNA microarray analyses, performed in chapter 4.1.1 (compare also table 4.3, page 68). Five of them can be knocked out without affecting the ability to grow in any way. It should be remarked, that three of those five genes have substitutes available. That means a knock out of those genes may only result in a lower amount of the coded enzyme, but will not remove it completely from the organism. Since we assume all enzymes, which are available in any amount, to be available in sufficient amount (compare page 28), no effect on the system can be observed.

The remaining six mutations lead to an incapability to produce biomass, while still metabolic pathways leading to carbon dioxide are active. As said above, this represents a model that is able to gain energy from supplied nutrients, but lacks the ability to make use of the available matter.

Remarkably, none of the genes which are only weakly expressed is vital to the organism (i.e. none of their knock outs is classified *lethal*). Thus it seems reasonable to assume that a genome scale model will even show less impact of the six mutations last-mentioned.

Non-annotated enzymes

Table 4.18 (p. 97) displays the results of removing those enzyme, which could not be annotated from the given genome. Most of the reactions listed cannot be neglected without consequence. The five reactions marked *no effect* in the table mentioned may in fact be neglected, but have been included to close obvious gaps.

This behavior - the necessity of most of the non-annotated reactions - very probably roots in the incompleteness of our model. As said before (page 100), though the model used is the largest *C. glutamicum* model constructed so far, it is far from being complete when referring to the genome. A more detailed examination and - consecutively - delineation of possible pathways may very well lead to alternate chains of reactions from one node

(metabolite) to another. In consequence, the reactions now being vital may become one of several possibilities and thereby may become negligible.

Non-catalyzed reactions

Only one of the reactions that do not have an enzyme attached might be in need of a catalyst: reaction 144, which is the only one producing the biomass component *ADP-D-glycero-D-manno-heptose* (compare table 4.6, page 71). Its removal entails a limited metabolism (no biomass production), but does not have to be classified lethal (see above).

The metabolite is included in the biomass as a precursor for the lipopolysaccharid synthesis. Since there is only *CMP-3-deoxy-D-manno-octulosonate* available as a substitute, we figured *ADP-D-glycero-D-manno-heptose* to be essential and included reaction 144 in the model. Additionally, the KEGG database does not provide full EC numbers of the reactions leading to the above heptose-derivative. Thus we could not determine if the relevant enzymes are annotated from the *C. glutamicum* genome.

Lethal knock outs

We explained on page 97 that there are only three common membrane transports which can be closed without lethal effects. Namely, the external fluxes for P_i, SO_4^{2-} or NH_3 may be constricted to zero while the metabolism remains active but incapable to grow.

Looking at the deadly internal knock outs, 5 of the 11 enzymes are needed within the respiratory chain: 1.10.2.2, 1.3.5.1, 1.3.99.1, 1.9.3.1, and 3.6.1.5. Taking away the ability to breath will naturally kill the organism.

The enzyme 1.1.1.37 is responsible for the conversion of malate to oxaloacetate within the citric acid cycle. It seems reasonable that it is not possible to destroy such a central gene without consequences. The same holds true for the following enzymes, which all catalyze parts of the same pathway: 2.3.3.1 (acetyl-CoA + oxaloacetate → citrate), 4.2.1.3 (citrate → isocitrate), and 4.2.1.2 (fumarate → malate).

The remaining enzymes are 2.3.1.8 and 2.7.1.2, the latter of which is responsible for the conversion of glucose to glucose-6-phosphate by use of ATP. The protein carrying the EC number 2.3.1.8 is used by *Corynebacterium glutamicum* to build acetyl-CoA from acetate. Both of them are understandably important reactions for the bacterium, considering the fundamentality of their products.

Reflection

We observe 3 general kinds of effects following single knock out mutations:

- no effect at all

- incapability to grow, but remaining activity of other metabolic pathways, which transform inflowing carbon into CO_2.

- lethality

There exist only a few lethal knock outs, whereas there are a lot of knock outs preventing growth. But, these data have to be considered with precaution. The case of reaction 144 described above is not uncommon (see non-catalyzed reactions). A knock out necessarily inhibits the production of biomass, if the referring reaction is the only one producing a metabolite which is part of the biomass. This may be different, if there are more reactions added to the network, which build this specific compound.

Also it should be taken into consideration to "soften" the biomass constitution. The components of biomass listed on page 71 are considered as the precursors (see 5.1.3) that are needed to build the cell. It may be possible to replace some of them by other metabolites; or to use ranges rather than discrete stoichiometric factors for some of the molecules. In consequence the model should react more robust against minor changes affecting the biomass production and/or constitution.

5.6 Conclusions

The established model has been discussed and classified within section 5.1. It consists of the main biochemical pathways and all additional reactions necessary to produce biomass. The model is able to grow on varying carbon sources, two of which have been considered within the work at hand (acetate and glucose).

Flux balance analyses have been carried out for a variety of organisms, some are listed in table 5.1 on page 102. Along publishing the largest model for glutamate producing bacteria up to now (Takaç *et al.*, 1998), Takaç and colleagues also performed *FBA* analyses under varying constraints. The *C. glutamicum* model described in section 4.1 is able to reproduce those results. Since we used a larger and more detailed model, naturally we achieved more in-depth results. Also - contrary to the workgroup from Ankara University - we employed a systemic approach to calculate possible reaction directions (3.4). To our knowledge, we are the first group besides D. Beard *et al.*, who are the authors of the energy balance analysis, which used this framework on a metabolic network model.

A research group also very active on the field of flux balance analysis is the *Systems Biology Research Group* from the University of California, San Diego (Palsson, 2005), led by Prof. Bernhard Palsson. Their *FBA* related research directs into developing new methods for result interpretation. One of the achievements was the publication of the so-called *extreme pathways* (Schilling *et al.*, 2000, 2001). They are a unique subset of the *elementary flux modes* introduced by Schuster *et al.* in (Schuster *et al.*, 1999, 2000). An elementary flux mode is a minimal combination of reactions, that can provide active flux through the network. It is found by consecutively eliminating reactions from the network while still valid flux distributions are possible. Eventually, one reaches a point where the elimination of one more reaction would be lethal, meaning no valid flux arrangement can be found any more.

The set of all possible pathways that are just one step away from break-down is called the set of elementary modes. The before mentioned extreme pathways are a proper subset of those, which is unique by the algorithm calculating them (Schilling *et al.*, 2000). A more detailed comparison of the two concepts can be found elsewhere (Jason *et al.*, 2004).

Another method, which has to be mentioned here, is the *phenotype phase plane analysis*. Phenotype phase planes are areas demarcated by the projection of the extreme pathways onto relevant subspaces. An example would be the carbon source uptake rate against the oxygen uptake rate. By phenotype phase plane analysis different regions, each of which describes another phenotype, can be established on this coordinate system. The theory is described briefly in (Schilling *et al.*, 2001) and in more detail in (Edwards *et al.*, 2002).

Contrary to focusing on the interpretation, we focused on the analyzing methods. First, we considered it necessary to implement the energy balance analysis (3.4) in order to achieve a thermodynamically sound flux distribution. On this basis only, we have been able to introduce new constraints, also based on thermodynamics. Metabolic profiling data was employed to further smallen the solution space. Compare sections 3.6, 4.4, and 5.4 for method, results, and discussion, respectively. As far as we know, there haven't been any other attempts of implementing metabolic profiling data into the systemic simulation of metabolic networks. Also, other approaches than *EBA* to improve the results of *FBA* are unknown to us.

Flux balance analysis does not deliver a unique solution, even after optimizing an objective function (Mahadevan & Schilling, 2003). Energy balance analysis is able to reduce the possible solution space significantly (compare figure 4.12 on page 85), but nevertheless it does not yield a discrete value. We aimed at a further reduction of the remaining flux space. Even by considering thermodynamical aspects imposed onto the model by the metabolic profiling data, we did not reduce the valid solutions to one vector. But we

did further reduce the available solution space.

As said in the introduction, our work aimed at closing the gap between prediction and nature and at the reduction of the valid flux space. Therefore, we planned to build a genome based model of the *Corynebacterium glutamicum* metabolism and to validate that model by experimental data. In section 4.1 we introduced the model. It has been validated as planned in section 4.2. Using a newly developed method to unite metabolic profiling with energy balance analysis (chapter 3.6) we also have indeed been able to further reduce the available solution space. The referring results and discussion can be found in sections 4.4.3 and 5.4.3.

Along the way, we analyzed the external energy balance (see 3.5, 4.3.5 and 5.3.4) and performed a single knock out mutation study (see 3.11, 4.5 and 5.5) to further elucidate the modeled organism.

Centralized, it can be said that we have created a new concept to combine knowledge gained from systemic laboratory approaches with an established method to analyze and simulate metabolic networks. This concept has been demonstrated on the example of the *C. glutamicum* metabolism and we thereby showed that we accomplished our goals.

Chapter 6

Outlook

The concept presented in this work brings together data from metabolite profiling experiments (2.5) and a theoretical systems biology approach (*EBA*, see 2.7). One drawback during the development was the incompleteness of the data from the metabolic profiling (compare 3.6.2.1). Even though we discussed the amount of unknown concentrations to be declining on page 116, there will always be metabolites left that cannot be measured. To quantify metabolites like described in 2.5.2, either the compound needs to be purchasable or available in databases. This is not given for around 30 % of the molecules listed in table A.2 (appendix A). If we consider, that the reactions used in the constructed model mainly constitute the central carbon metabolism, this percentage will probably be higher when zooming out to the whole cellular network. Also, larger molecules cannot be measured, since it is not possible to evaporate them during the gas chromatography (compare 2.5.1). Examples include such important substances as ATP or NADH. Currently, research at CUBIC starts to develop LC-MS[1] methods to be able to expand the range of detectable molecules. But very probably, we will never be able to quantify all of them that exist within the *C. glutamicum* metabolism. And we definitely won't be within the next few years.

---

[1]**L**iquid **C**hromatography - **M**ass **S**pectrometry

The same problems will be faced when analyzing a different organism. Thus, it is reasonable to start generating concepts to predict unknown metabolic concentrations. These, in turn, can be used to further reduce the solution space of the energy balance analysis. Daniel Beard and Hong Qian published a study on the hepatocyte metabolism in (Beard & Qian, 2005) in which they determined unknown metabolic concentrations by minimizing the enzyme activity differences between two operational modes of the cell. This could be translated to the *Corynebacterium glutamicum* network. E.g., it is possible to parallel the work at hand and interpret the external conditions (the feeding with different nutrients) as "operational modes". A minimization of the difference of the total amount of metabolic concentrations may be the first step towards a more detailed objective function.

There have been approaches developed, which base on flux balance analysis, but move to different objective functions. In (Holzhütter, 2004) Herman-Georg Holzhütter introduces the idea of a minimization of the sum of all fluxes in a metabolic network, while keeping some "target fluxes" constant or within small intervals. Since already Segrè noted in 2002 (Segrè *et al.*, 2002) that optimal biomass production is not necessarily a sound target for all models, it is definitely worth considering objective functions different from that. Also, Holzhütter stated that optimal biomass production can be regarded as a special case of flux minimization: there is a relation between trying to achieve a certain biomass production at minimal costs and achieving maximal biomass production at given costs. A successive application of both objectives seems promising, too. First, the biomass production will be optimized. Second, it is tried to achieve this at minimal costs.

Another idea is called MOMA and was published in (Segrè *et al.*, 2002). MOMA is an acronym of *minimization of metabolic adjustments* and is focused on the analysis of perturbed metabolic networks; the most common perturbation being a knock out mutation. The underlying thesis of MOMA

is that random knock outs - like performed in laboratories - did not occur under the pressure of evolution. Thus, there is no reason to assume these mutants to be optimized on biomass production anymore. Instead, the authors suggest the optimal flux distribution of perturbed networks being the one that is as close to the wild type results as possible. In consequence, for knock out mutants, they do not optimize the biomass production rate, but minimize the distance between the wild type and the mutant flux distribution:

$$D(\vec{\mathbf{w}}, \vec{\mathbf{v}}) = \sqrt{\sum_i (w_i - v_i)^2} \qquad (6.1)$$

D: distance in flux space; $\vec{\mathbf{w}}$: wild type flux vector; $\vec{\mathbf{v}}$: mutant flux vector

So the task is to find the vector $\vec{\mathbf{v}}$ such that $D(\vec{\mathbf{w}}, \vec{\mathbf{v}})$ is minimized. In the above cited article, the authors claim to be able to reach results better than *FBA* alone, while analyzing mutants. The drawback is: the flux distribution of the wild type organism must be known in order to find the most similar one from the mutant's flux space. While this method does not help in determining the flux vector of the *C. glutamicum* wild type, it may be helpful when analyzing the knock out organisms that are raised in CUBIC laboratories. It should be considered as well to adapt the approach to map experimental results (e. g. (Wendisch *et al.*, 2000)) as best as possible. This may help in identifying the flux rates of reactions not covered in experiments.

In order to gain more biological knowledge, focus should also be laid on the interpretation of the results. At first, an application of enhanced methods that already exist should be performed. Expressing the flux distribution in terms of extreme pathways (Schilling *et al.*, 2000, 2001) or analyzing different phenotype phase planes (Schilling *et al.*, 2001; Edwards *et al.*, 2002) might lead to new insights. After that, it should be considered to take advantage of the available concentration data, in combination with predictions of the missing data, if available. One approach is currently under development by Asad Rahman from CUBIC. His idea is to rank cellular pathways by the

amount of Gibbs free energy (2.1.2) that is lost or gained while traveling them. Since different molecules are involved in different pathways, the ΔG varies, dependent on the particular way that is taken. The ΔG values can easily be refined by using the available concentration data. Further improvement can be gained by employing predicted concentration data to fill empty variables.

At this point, a last suggestion shall be the combination of pathway analyses with flux prediction. One idea is to rank pathways by the flux through them. Another concept will be briefly outlined next.

A flux gives the velocity of a reaction. The units usually are $\left[\frac{\text{mmol}}{\text{g}(DryWeight)\cdot\text{h}}\right]$. The reciprocal gives information about the time it takes to convert one mmol of a compound. If this is united with a shortest path analysis, the result could be called a *fastest path analysis*. It could answer questions like: which is the fastest way to convert one mol of **A** into one mol of **B**?

A closing remark: there is a really wide spectrum of possible applications and concepts that can be developed within systems biology. It will definitely go beyond the scope of this work to discuss them all. The thesis at hand is a first step into uniting metabolic profiling experiments with the framework of flux prediction methods. And we think it is definitely worth it, to further follow this path.

Appendix A

Reactions and Metabolites

Table: A.1: **Classification of the Networks Reactions**
The classification of the networks reactions into different pathways is not
unambiguous. It is meant to be a hint.

	Glycolysis/Glyconeogenesis	
1	ATP + D-Glucose <=> ADP + D-Glucose 6-phosphate + H+	2.7.1.2
2	2-Phospho-D-glycerate <=> Phosphoenolpyruvate + H2O	4.2.1.11
3	ATP + D-Fructose 6-phosphate <=> ADP + D-Fructose 1,6-bisphosphate + H+	2.7.1.11
4	D-Fructose 1,6-bisphosphate + H2O <=> D-Fructose 6-phosphate + Orthophosphate	3.1.3.11
5	D-Glucose 6-phosphate <=> D-Fructose 6-phosphate	5.3.1.9
6	Sucrose + H2O <=> D-Fructose + D-Glucose	3.2.1.20
7	ATP + D-Fructose <=> ADP + D-Fructose 1-phosphate + H+	2.7.1.3
8	ATP + D-Fructose <=> ADP + beta-D-Fructose 6-phosphate + H+	2.7.1.4
9	D-Glucose 1-phosphate <=> alpha-D-Glucose 6-phosphate	5.4.2.2
10	(2R)-2-Hydroxy-3-(phosphonooxy)-propanal <=> Glycerone phosphate	5.3.1.1
11	(2R)-2-Hydroxy-3-(phosphonooxy)-propanal + Orthophosphate + NAD+ <=> 3-Phospho-D-glyceroyl phosphate + NADH + H+	1.2.1.12 1.2.1.59
12	beta-D-Fructose 1,6-bisphosphate <=> Glycerone phosphate + (2R)-2-Hydroxy-3-(phosphonooxy)-propanal	4.1.2.13

Table: A.1 ... Continued from previous page

13	ATP + 3-Phospho-D-glycerate <=> ADP + 3-Phospho-D-glyceroyl phosphate	2.7.2.3
14	2-Phospho-D-glycerate <=> 3-Phospho-D-glycerate	5.4.2.1
15	ATP + D-Fructose 1-phosphate <=> ADP + beta-D-Fructose 1,6-bisphosphate + H+	2.7.1.56
Pentose Phosphate Pathway		
16	D-Glucose 6-phosphate + NADP+ <=> D-Glucono-1,5-lactone 6-phosphate + NADPH + H+	1.1.1.49
17	ATP + D-Ribose 5-phosphate <=> AMP + 5-Phospho-alpha-D-ribose 1-diphosphate + H+	2.7.6.1
18	D-Ribose 5-phosphate <=> D-Ribulose 5-phosphate	5.3.1.6
19	6-Phospho-D-gluconate + NADP+ <=> D-Ribulose 5-phosphate + CO2 + NADPH	1.1.1.44
20	D-Ribulose 5-phosphate <=> D-Xylulose 5-phosphate	5.1.3.1
21	Sedoheptulose 7-phosphate + (2R)-2-Hydroxy-3-(phosphonooxy)-propanal <=> D-Ribose 5-phosphate + D-Xylulose 5-phosphate	2.2.1.1
22	Sedoheptulose 7-phosphate + (2R)-2-Hydroxy-3-(phosphonooxy)-propanal <=> D-Erythrose 4-phosphate + D-Fructose 6-phosphate	2.2.1.2
23	beta-D-Fructose 6-phosphate + (2R)-2-Hydroxy-3-(phosphonooxy)-propanal <=> D-Erythrose 4-phosphate + D-Xylulose 5-phosphate	2.2.1.1
24	D-Glucono-1,5-lactone 6-phosphate + H2O <=> 6-Phospho-D-gluconate + H+	3.1.1.31
Tricarboxylic Acid Cycle		
25	Isocitrate + NADP+ <=> 2-Oxoglutarate + CO2 + NADPH	1.1.1.42
26	(S)-Malate + NAD+ <=> Oxaloacetate + NADH + H+	1.1.1.37
27	Citrate + CoA + H+ <=> Acetyl-CoA + H2O + Oxaloacetate	2.3.3.1
28	ATP + Succinate + CoA <=> ADP + Orthophosphate + Succinyl-CoA	6.2.1.5
29	Succinate + FAD <=> FADH2 + Fumarate	1.3.99.1
30	(S)-Malate <=> Fumarate + H2O	4.2.1.2
31	Citrate <=> Isocitrate	4.2.1.3
32	Dihydrolipoamide + NAD+ <=> Lipoamide + NADH + H+	1.8.1.4

Continued on next page ...

Table: A.1 . . . Continued from previous page

33	2-Oxoglutarate + Lipoamide + H+ <=> S-Succinyldihydrolipoamide + CO2	1.2.4.2
34	Succinyl-CoA + Dihydrolipoamide <=> CoA + S-Succinyldihydrolipoamide	2.3.1.61
Glyoxylate Shunt		
35	(S)-Malate + CoA + H+ <=> Acetyl-CoA + H2O + Glyoxylate	2.3.3.9
36	Isocitrate <=> Succinate + Glyoxylate	4.1.3.1
Pyruvate Metabolism		
37	ATP + Pyruvate <=> ADP + Phosphoenolpyruvate + H+	2.7.1.40
38	Pyruvate + CoA + NAD+ <=> Acetyl-CoA + CO2 + NADH	2.3.1.12 1.8.1.4 1.2.4.1
39	Oxaloacetate + H+ <=> Pyruvate + CO2	4.1.1.3
40	Acetyl-CoA + Orthophosphate <=> CoA + Acetyl phosphate	2.3.1.8
41	ATP + Acetate <=> ADP + Acetyl phosphate	2.7.2.1
42	Pyrophosphate + Acetate <=> Orthophosphate + Acetyl phosphate	2.7.2.12
43	Orthophosphate + Oxaloacetate + H+ <=> H2O + Phosphoenolpyruvate + CO2	4.1.1.31
44	(S)-Lactate + NAD+ <=> Pyruvate + NADH + H+	1.1.1.27
Alanine Biosynthesis		
45	L-Alanine + 2-Oxoglutarate <=> Pyruvate + L-Glutamate	2.6.1.2
Arginine Biosynthesis		
46	2 ATP + NH3 + CO2 + H2O <=> 2 ADP + Orthophosphate + Carbamoyl phosphate + 3 H+	6.3.4.16
47	N-(L-Arginino)succinate <=> Fumarate + L-Arginine	4.3.2.1
48	Carbamoyl phosphate + L-Ornithine <=> Orthophosphate + L-Citrulline + H+	2.1.3.3
49	ATP + L-Citrulline + L-Aspartate <=> AMP + Pyrophosphate + N-(L-Arginino)succinate + H+	6.3.4.5
Asparagine/Aspartate Metabolism		
50	L-Aspartate + 2-Oxoglutarate <=> Oxaloacetate + L-Glutamate	2.6.1.1

Table: A.1 ... Continued from previous page

51	ATP + L-Aspartate + L-Glutamine + H2O <=> AMP + Pyrophosphate + L-Asparagine + L-Glutamate + H+	6.3.5.4
Cysteine Biosynthesis		
52	L-Serine + Acetyl-CoA <=> O-Acetyl-L-serine + CoA	2.3.1.30
53	O-Acetyl-L-serine + Hydrogen sulfide <=> L-Cysteine + Acetate + H+	2.5.1.47
Glutamine/Glutamate Metabolism		
54	2 L-Glutamate + NADP+ <=> L-Glutamine + 2-Oxoglutarate + NADPH + H+	1.4.1.13
55	L-Glutamate + NADP+ + H2O <=> 2-Oxoglutarate + NH3 + NADPH + H+	1.4.1.4
56	ATP + L-Glutamate + NH3 <=> ADP + Orthophosphate + L-Glutamine + H+	6.3.1.2
57	L-Glutamine + H2O <=> L-Glutamate + NH3	3.5.1.2
Glycine Biosynthesis		
58	Glycine + Tetrahydrofolate + NAD+ <=> 5,10-Methylenetetrahydrofolate + NH3 + CO2 + NADH	2.1.2.10
Histidine Biosynthesis		
59	Phosphoribosyl-ATP + Pyrophosphate <=> ATP + 5-Phospho-alpha-D-ribose 1-diphosphate	2.4.2.17
60	L-Histidinol + 2 NAD+ + H2O <=> L-Histidine + 2 NADH + 3 H+	1.1.1.23
61	L-Histidinol phosphate + H2O <=> L-Histidinol + Orthophosphate	3.1.3.15
62	L-Histidinol phosphate + 2-Oxoglutarate <=> 3-(Imidazol-4-yl)-2-oxopropyl phosphate + L-Glutamate	2.6.1.9
63	D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate <=> 3-(Imidazol-4-yl)-2-oxopropyl phosphate + H2O	4.2.1.19
64	Phosphoribosyl-ATP + H2O <=> Phosphoribosyl-AMP + Pyrophosphate + H+	3.6.1.31
65	Phosphoribosyl-AMP + H2O + H+ <=> 5-(5-Phospho-D-ribosylaminoformimino)-1-(5-phosphoribosyl)-imidazole-4-carboxamide	3.5.4.19

Continued on next page ...

Table: A.1 . . . Continued from previous page

66	1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide + L-Glutamate + D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate + H+ <=> N-(5'-Phospho-D-1'-ribulosylformimino)-5-amino-1-(5"-phospho-D-ribosyl)-4-imidazolecarboxamide + L-Glutamine	2.4.2.-
67	5-(5-Phospho-D-ribosylaminoformimino)-1-(5-phosphoribosyl)-imidazole-4-carboxamide <=> N-(5'-Phospho-D-1'-ribulosylformimino)-5-amino-1-(5"-phospho-D-ribosyl)-4-imidazolecarboxamide	5.3.1.16
Isoleucine Biosynthesis		
68	L-Isoleucine + 2-Oxoglutarate <=> (R)-2-Oxo-3-methylpentanoate + L-Glutamate	2.6.1.42
69	(R)-2,3-Dihydroxy-3-methylpentanoate + NADP+ <=> (S)-2-Hydroxy-3-methyl-3-oxopentanoate + NADPH + H+	1.1.1.86
70	(S)-2-Aceto-2-hydroxybutanoate <=> (S)-2-Hydroxy-3-methyl-3-oxopentanoate	1.1.1.86
71	(R)-2,3-Dihydroxy-3-methylpentanoate <=> (R)-2-Oxo-3-methylpentanoate + H2O	4.2.1.9
72	2-Oxobutanoate + Pyruvate + H+ <=> (S)-2-Aceto-2-hydroxybutanoate + CO2	2.2.1.6
Leucine Biosynthesis		
73	2-Acetolactate + CO2 <=> 2 Pyruvate + H+	2.2.1.6
74	L-Leucine + 2-Oxoglutarate <=> 4-Methyl-2-oxopentanoate + L-Glutamate	2.6.1.42
75	2,3-Dihydroxy-3-methylbutanoate <=> 3-Methyl-2-oxobutanoate + H2O	4.2.1.9
76	2-Isopropylmalate + CoA + H+ <=> Acetyl-CoA + 3-Methyl-2-oxobutanoate + H2O	2.3.3.13
77	4-Methyl-2-oxopentanoate + CO2 <=> 3-Carboxy-4-methyl-2-oxopentanoate + H+	0.0.0.0
78	2-Acetolactate + NADPH + H+ <=> 2,3-Dihydroxy-3-methylbutanoate + NADP+	1.1.1.86

Table: A.1 . . . Continued from previous page

79	2-Isopropylmalate <=> 2-Isopropylmaleate + H2O	4.2.1.33
80	3-Isopropylmalate <=> 2-Isopropylmaleate + H2O	4.2.1.33
81	3-Isopropylmalate + NAD+ <=> 3-Carboxy-4-methyl-2-oxopentanoate + NADH + H+	1.1.1.85
Lysine Biosynthesis		
82	meso-2,6-Diaminoheptanedioate + H+ <=> L-Lysine + CO2	4.1.1.20
83	ATP + L-Aspartate <=> ADP + 4-Phospho-L-aspartate	2.7.2.4
84	L-Homoserine + NADP+ <=> L-Aspartate 4-semialdehyde + NADPH + H+	1.1.1.3
85	L-Aspartate 4-semialdehyde + Orthophosphate + NADP+ <=> 4-Phospho-L-aspartate + NADPH + H+	1.2.1.11
86	L-Aspartate 4-semialdehyde + Pyruvate <=> 2,3-Dihydrodipicolinate + 2 H2O + H+	4.2.1.52
87	N-Succinyl-LL-2,6-diaminoheptanedioate + H2O <=> Succinate + LL-2,6-Diaminoheptanedioate	3.5.1.18
88	LL-2,6-Diaminoheptanedioate <=> meso-2,6-Diaminoheptanedioate	5.1.1.7
89	2,3,4,5-Tetrahydrodipicolinate + NADP+ <=> 2,3-Dihydrodipicolinate + NADPH + H+	1.3.1.26
90	Succinyl-CoA + 2,3,4,5-Tetrahydrodipicolinate + H2O <=> CoA + N-Succinyl-2-L-amino-6-oxoheptanedioate	2.3.1.117
91	N-Succinyl-LL-2,6-diaminoheptanedioate + 2-Oxoglutarate <=> N-Succinyl-2-L-amino-6-oxoheptanedioate + L-Glutamate	2.6.1.17
Methionine Biosynthesis		
92	5-Methyltetrahydrofolate + L-Homocysteine <=> Tetrahydrofolate + L-Methionine	2.1.1.13
93	L-Cystathionine + H2O <=> L-Homocysteine + NH3 + Pyruvate	4.4.1.8
94	Acetyl-CoA + L-Homoserine <=> CoA + O-Acetyl-L-homoserine	2.3.1.31
95	O-Acetyl-L-homoserine + L-Cysteine <=> L-Cystathionine + Acetate + H+	2.5.1.49
Ornithine Biosynthesis		
96	Acetyl-CoA + L-Glutamate <=> CoA + N-Acetyl-L-glutamate + H+	2.3.1.1

Table: A.1 ... Continued from previous page

97	N2-Acetyl-L-ornithine + H2O <=> Acetate + L-Ornithine	3.5.1.14
98	L-Ornithine <=> L-Proline + NH3	4.3.1.12
99	N2-Acetyl-L-ornithine + 2-Oxoglutarate <=> N-Acetyl-L-glutamate 5-semialdehyde + L-Glutamate	2.6.1.11
100	ATP + N-Acetyl-L-glutamate <=> ADP + N-Acetyl-L-glutamate 5-phosphate	2.7.2.8
101	N-Acetyl-L-glutamate 5-semialdehyde + Orthophosphate + NADP+ <=> N-Acetyl-L-glutamate 5-phosphate + NADPH + H+	1.2.1.38
Phenylalanine Biosynthesis		
102	L-Phenylalanine + 2-Oxoglutarate <=> Phenylpyruvate + L-Glutamate	2.6.1.9 2.6.1.1
103	Prephenate + H+ <=> Phenylpyruvate + H2O + CO2	4.2.1.51
104	5-O-(1-Carboxyvinyl)-3-phosphoshikimate <=> Chorismate + Orthophosphate	4.2.3.5
105	Chorismate <=> Prephenate	5.4.99.5
106	2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate + Orthophosphate <=> Phosphoenolpyruvate + D-Erythrose 4-phosphate + H2O	2.5.1.54
107	ATP + Shikimate <=> ADP + Shikimate 3-phosphate + H+	2.7.1.71
108	Shikimate + NADP+ <=> 3-Dehydroshikimate + NADPH + H+	1.1.1.25
109	2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate <=> 3-Dehydroquinate + Orthophosphate	4.2.3.4
110	3-Dehydroquinate <=> 3-Dehydroshikimate + H2O	4.2.1.10 4.2.1.11
111	Phosphoenolpyruvate + Shikimate 3-phosphate <=> Orthophosphate + 5-O-(1-Carboxyvinyl)-3-phosphoshikimate	2.5.1.19
Proline Biosynthesis		
112	ATP + L-Glutamate <=> ADP + L-Glutamyl 5-phosphate	2.7.2.11
113	L-Proline + NADP+ <=> (S)-1-Pyrroline-5-carboxylate + NADPH + H+	1.5.1.2
114	L-Glutamate 5-semialdehyde + Orthophosphate + NADP+ <=> L-Glutamyl 5-phosphate + NADPH + H+	1.2.1.41

Continued on next page ...

Table: A.1 . . . Continued from previous page

115	L-Glutamate 5-semialdehyde <=> (S)-1-Pyrroline-5-carboxylate + H2O	0.0.0.0
Serine Biosynthesis		
116	O-Phospho-L-serine + H2O <=> L-Serine + Orthophosphate	3.1.3.3
117	3-Phospho-D-glycerate + NAD+ <=> 3-Phosphonooxypyruvate + NADH + H+	1.1.1.95
118	O-Phospho-L-serine + 2-Oxoglutarate <=> 3-Phosphonooxypyruvate + L-Glutamate	2.6.1.52
Threonine Biosynthesis		
119	L-Threonine <=> 2-Oxobutanoate + NH3	4.3.1.19
120	O-Phospho-L-homoserine + H2O <=> L-Threonine + Orthophosphate	4.2.3.1
121	ATP + L-Homoserine <=> ADP + O-Phospho-L-homoserine + H+	2.7.1.39
Tryptophane Biosynthesis		
122	Chorismate + L-Glutamine <=> Anthranilate + Pyruvate + L-Glutamate + H+	4.1.3.27
123	N-(5-Phospho-D-ribosyl)anthranilate + Pyrophosphate <=> Anthranilate + 5-Phospho-alpha-D-ribose 1-diphosphate	2.4.2.18
124	L-Serine + Indoleglycerol phosphate <=> L-Tryptophan + Glyceraldehyde 3-phosphate + H2O	4.2.1.20
125	1-(2-Carboxyphenylamino)-1'-deoxy-D-ribulose 5'-phosphate + H+ <=> Indoleglycerol phosphate + CO2 + H2O	4.1.1.48
126	N-(5-Phospho-D-ribosyl)anthranilate <=> 1-(2-Carboxyphenylamino)-1'-deoxy-D-ribulose 5'-phosphate	5.3.1.24
Tyrosine Biosynthesis		
127	L-Tyrosine + 2-Oxoglutarate <=> 3-(4-Hydroxyphenyl)pyruvate + L-Glutamate	2.6.1.9 / 2.6.1.1
128	Prephenate + NAD+ <=> 3-(4-Hydroxyphenyl)pyruvate + CO2 + NADH	1.3.1.12
Valine Biosynthesis		
129	L-Valine + 2-Oxoglutarate <=> 3-Methyl-2-oxobutanoate + L-Glutamate	2.6.1.42

Table: A.1 ... Continued from previous page

	Fatty Acid Production	
130	7 Acetyl-CoA + 12 NADPH + 6 ATP + 5 H+ + H2O <=> 7 CoA + Myristic Acid + 12 NADP+ + 6 ADP + 6 Orthophosphate	2.3.1.85
131	7 Acetyl-CoA + 11 NADPH + 6 ATP + 4 H+ + H2O <=> 7 CoA + Myristoleic Acid + 11 NADP+ + 6 ADP + 6 Orthophosphate	2.3.1.85
132	8.2 Acetyl-CoA + 14 NADPH + 7.2 ATP + 6 H+ + H2O <=> 8.2 CoA + Fatty Acid + 14 NADP+ + 7.2 ADP + 7.2 Orthophosphate	2.3.1.85
	Other Biomass Components	
133	sn-Glycerol 3-phosphate + NAD+ <=> Glycerone phosphate + NADH + H+	1.1.1.94
134	ATP + Glycerol <=> ADP + sn-Glycerol 3-phosphate	2.7.1.30
135	Glycerol + NADP+ <=> D-Glyceraldehyde + NADPH + H+	1.1.1.1 1.1.1.2
136	ATP + D-Glycerate <=> ADP + 3-Phospho-D-glycerate	2.7.1.31
137	D-Glyceraldehyde + NAD+ + H2O <=> D-Glycerate + NADH + 2 H+	1.2.1.3
138	3-Deoxy-D-manno-octulosonate 8-phosphate + H2O <=> 3-Deoxy-D-manno-octulosonate + Orthophosphate	3.1.3.45
139	CTP + 3-Deoxy-D-manno-octulosonate <=> Pyrophosphate + CMP-3-deoxy-D-manno-octulosonate	2.7.7.38
140	ATP + NADH + H+ <=> dATP + NAD+ + H2O	1.6.1.2 1.17.4.2 1.8.1.9
141	GTP + NADPH + H+ <=> dGTP + NADP+ + H2O	1.6.1.2 1.17.4.2 1.8.1.9
142	CTP + NADPH + H+ <=> dCTP + NADP+ + H2O	1.6.1.2 1.17.4.2 1.8.1.9
143	3-Deoxy-D-manno-octulosonate 8-phosphate + Orthophosphate <=> Phosphoenolpyruvate + D-Ribulose 5-phosphate + H2O	5.3.1.13 2.5.1.55

Table: A.1 ... Continued from previous page

144	Sedoheptulose 7-phosphate + ATP + H+ <=> ADP-D-glycero-D-manno-heptose + Pyrophosphate	0.0.0.0
Biomass Production		
145	41.139 H2O + 41.139 ATP + 0.58 Glycine + 0.25 L-Glutamate + 0.25 L-Glutamine + 0.17 L-Phenylalanine + 0.13 L-Tyrosine + 0.2 L-Serine + 0.05 L-Tryptophan + 0.22 L-Aspartate + 0.22 L-Asparagine + 0.32 L-Lysine + 0.09 L-Cysteine + 0.14 L-Methionine + 0.24 L-Threonine + 0.27 L-Isoleucine + 0.2 L-Proline + 0.28 L-Arginine + 0.48 L-Alanine + 0.4 L-Valine + 0.42 L-Leucine + 0.09 L-Histidine + 0.185 ATP + 0.2 GTP + 0.13 UTP + 0.12 CTP + 0.02 dATP + 0.02 dGTP + 0.02 dCTP + 0.02 dTTP + 0.129 sn-Glycerol 3-phosphate + 0.129 L-Serine + 0.258 Fatty Acid + 0.0157 UDPglucose + 0.0235 Myristic Acid + 0.0235 Myristoleic Acid + 0.0235 CMP-3-deoxy-D-manno-octulosonate + 0.0235 ADP-D-glycero-D-manno-heptose + 0.0157 UDP-N-acetyl-D-glucosamine + 0.0276 UDP-N-acetyl-D-glucosamine + 0.0276 UDP-N-acetylmuramate + 0.0552 L-Alanine + 0.0276 meso-2,6-Diaminoheptanedioate + 0.0276 L-Glutamate + 0.154 UDPglucose + 0.048 L-Serine + 0.0593 L-Ornithine <=> 1.0 Biomass + 41.139 ADP + 41.139 Orthophosphate + 41.139 H+	0.0.0.0
Sugars Metabolism		
146	Maltose <=> alpha,alpha-Trehalose	5.4.99.16
147	(1,4-alpha-D-Glucosyl)n + Orthophosphate <=> D-Glucose 1-phosphate	2.4.1.1
148	(1,4-alpha-D-Glucosyl)n + alpha-D-Glucose <=> Maltose	2.4.1.25
Trehalose Biosynthesis		
149	UTP + D-Glucose 1-phosphate + H+ <=> Pyrophosphate + UDPglucose	2.7.7.9
150	UDPglucose + D-Glucose 6-phosphate <=> UDP + alpha,alpha'-Trehalose 6-phosphate + H+	2.4.1.15

Table: A.1 ... Continued from previous page

151	alpha,alpha'-Trehalose 6-phosphate + H2O <=> alpha,alpha-Trehalose + Orthophosphate	3.1.3.12
Aminosugars Metabolism		
152	UTP + N-Acetyl-D-glucosamine 1-phosphate + H+ <=> Pyrophosphate + UDP-N-acetyl-D-glucosamine	2.7.7.23
153	Phosphoenolpyruvate + UDP-N-acetyl-D-glucosamine <=> UDP-N-acetyl-3-(1-carboxyvinyl)-D-glucosamine + Orthophosphate	2.5.1.7
154	L-Glutamine + D-Fructose 6-phosphate <=> L-Glutamate + D-Glucosamine 6-phosphate	2.6.1.16
155	D-Glucosamine 1-phosphate <=> D-Glucosamine 6-phosphate	5.4.2.2
156	UDP-N-acetylmuramate + NADP+ <=> UDP-N-acetyl-3-(1-carboxyvinyl)-D-glucosamine + NADPH + H+	1.1.1.158
157	Acetyl-CoA + D-Glucosamine 1-phosphate <=> CoA + N-Acetyl-D-glucosamine 1-phosphate + H+	2.3.1.157
Purine Metabolism		
158	ATP + H2O <=> AMP + Pyrophosphate + H+	3.6.1.8
159	ATP + NAD+ <=> ADP + NADP+ + H+	2.7.1.23
160	ATP + AMP <=> ADP + ADP	2.7.4.3
161	ATP + GDP <=> ADP + GTP	2.7.4.6
162	ATP + GMP <=> ADP + GDP	2.7.4.8
163	5-Phosphoribosylamine + Pyrophosphate + L-Glutamate <=> L-Glutamine + 5-Phospho-alpha-D-ribose 1-diphosphate + H2O	2.4.2.14
164	N6-(1,2-Dicarboxyethyl)-AMP <=> Fumarate + AMP	4.3.2.2
165	IMP + H2O <=> 1-(5'-Phosphoribosyl)-5-formamido-4-imidazolecarboxamide	3.5.4.10
166	IMP + NAD+ + H2O <=> Xanthosine 5'-phosphate + NADH + H+	1.1.1.205
167	GTP + IMP + L-Aspartate <=> GDP + Orthophosphate + N6-(1,2-Dicarboxyethyl)-AMP + 2 H+	6.3.4.4
168	ATP + Xanthosine 5'-phosphate + L-Glutamine + H2O <=> AMP + Pyrophosphate + GMP + L-Glutamate + 2 H+	6.3.5.2

Table: A.1 . . . Continued from previous page

169	ATP + 5-Phosphoribosylamine + Glycine <=> ADP + Orthophosphate + 5'-Phosphoribosylglycinamide + H+	6.3.4.13
170	ATP + 2-(Formamido)-N1-(5'-phosphoribosyl)acetamidine <=> ADP + Orthophosphate + Aminoimidazole ribotide + 2 H+	6.3.3.1
171	1-(5-Phospho-D-ribosyl)-5-amino-4-imidazolecarboxylate + H+ <=> Aminoimidazole ribotide + CO2	4.1.1.21
172	10-Formyltetrahydrofolate + 5'-Phosphoribosylglycinamide <=> Tetrahydrofolate + 5'-Phosphoribosyl-N-formylglycinamide + H+	2.1.2.2
173	ATP + 5'-Phosphoribosyl-N-formylglycinamide + L-Glutamine + H2O <=> ADP + Orthophosphate + 2-(Formamido)-N1-(5'-phosphoribosyl)acetamidine + L-Glutamate + H+	6.3.5.3
174	1-(5'-Phosphoribosyl)-5-amino-4-(N-succinocarboxamide)-imidazole <=> Fumarate + 1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide	4.3.2.2
175	10-Formyltetrahydrofolate + 1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide <=> Tetrahydrofolate + 1-(5'-Phosphoribosyl)-5-formamido-4-imidazolecarboxamide	2.1.2.3
176	ATP + 1-(5-Phospho-D-ribosyl)-5-amino-4-imidazolecarboxylate + L-Aspartate <=> ADP + Orthophosphate + 1-(5'-Phosphoribosyl)-5-amino-4-(N-succinocarboxamide)-imidazole + H+	6.3.2.6
Pyrimidine Metabolism		
177	ATP + UDP <=> ADP + UTP	2.7.4.6
178	ATP + UMP <=> ADP + UDP	2.7.4.14
179	ATP + CDP <=> ADP + CTP	2.7.4.6
180	ATP + UTP + L-Glutamine + H2O <=> ADP + Orthophosphate + CTP + L-Glutamate + 2 H+	6.3.4.2
181	Orotidine 5'-phosphate + H+ <=> UMP + CO2	4.1.1.23
182	Carbamoyl phosphate + L-Aspartate <=> Orthophosphate + N-Carbamoyl-L-aspartate + H+	2.1.3.2
183	(S)-Dihydroorotate + NAD+ <=> Orotate + H+ + NADH	1.3.99.11

Table: A.1 ... Continued from previous page

184	Orotidine 5'-phosphate + Pyrophosphate <=> Orotate + 5-Phospho-alpha-D-ribose 1-diphosphate	2.4.2.10
185	(S)-Dihydroorotate + H2O <=> N-Carbamoyl-L-aspartate + H+	3.5.2.3
186	ATP + dTDP <=> ADP + dTTP	2.7.4.6
187	ATP + dTMP <=> ADP + dTDP	2.7.4.9
188	ATP + dUMP <=> ADP + dUDP	2.7.4.9
189	UDP + NADPH + H+ <=> dUDP + NADP+ + H2O	1.6.1.2 1.17.4.2 1.8.1.9
Tetrahydrofolate Metabolism		
190	Tetrahydrofolate + NADP+ <=> Dihydrofolate + NADPH + H+	1.5.1.3
191	5,10-Methylenetetrahydrofolate + Glycine + H2O <=> Tetrahydrofolate + L-Serine	2.1.2.1
192	5,10-Methylenetetrahydrofolate + NADP+ <=> 5,10-Methenyltetrahydrofolate + NADPH	1.5.1.5
193	5-Methyltetrahydrofolate + NAD+ <=> 5,10-Methylenetetrahydrofolate + NADH + H+	1.7.99.5
194	5,10-Methenyltetrahydrofolate + H2O <=> 10-Formyltetrahydrofolate + H+	3.5.4.9
195	dUMP + 5,10-Methylenetetrahydrofolate <=> Dihydrofolate + dTMP	2.1.1.45
Sulfur metabolism		
196	Sulfate + 4 NADPH + 6 H+ <=> Hydrogen sulfide + 4 NADP+ + 4 H2O	1.8.1.2 1.8.2.1 1.10.2.2 1.6.5.3 1.6.1.1
Respiratory Reactions		
197	2 NADH + 4 ADP + 4 Orthophosphate + Oxygen + 6 H+ <=> 2 NAD+ + 4 ATP + 6 H2O	1.6.5.3 1.10.2.2 1.9.3.1 3.6.1.5

Table: A.1 ... Continued from previous page

198	2 FADH2 + 2 ADP + 2 Orthophosphate + Oxygen + 2 H+ <=> 2 FAD + 2 ATP + 4 H2O	1.3.5.1
		1.3.99.1
		1.10.2.2
		1.9.3.1
		3.6.1.5
Other Reactions		
199	Pyrophosphate + H2O <=> 2 Orthophosphate + H+	3.6.1.1
200	NADPH + NAD+ <=> NADP+ + NADH	1.6.1.1
201	NADP+ + H2O <=> Orthophosphate + NAD+	3.1.3.-
Isomeric Conversions		
202	Maltose <=> Maltose (external)	0.0.0.0
203	(2R)-2-Hydroxy-3-(phosphonooxy)-propanal <=> Glyceraldehyde 3-phosphate	0.0.0.0
204	D-Glucose <=> alpha-D-Glucose	0.0.0.0
205	D-Fructose 6-phosphate <=> beta-D-Fructose 6-phosphate	0.0.0.0
206	D-Glucose 6-phosphate <=> alpha-D-Glucose 6-phosphate	0.0.0.0
207	beta-D-Fructose 1,6-bisphosphate <=> D-Fructose 1,6-bisphosphate	0.0.0.0
Membrane Transport Reactions		
208	glutamate <=> glutamate (external)	0.0.0.0
209	lactate <=> lactate (external)	0.0.0.0
210	trehalose <=> trehalose (external)	0.0.0.0
211	lysine <=> lysine (external)	0.0.0.0
212	glucose <=> glucose (external)	0.0.0.0
213	phosphate <=> phosphate (external)	0.0.0.0
214	CO2 <=> CO2 (external)	0.0.0.0
215	H2O <=> H2O (external)	0.0.0.0
216	O2 <=> O2 (external)	0.0.0.0
217	L-Phenylalanine <=> L-Phenylalanin (external)	0.0.0.0
218	D-Fructose <=> D-Fructose (external)	0.0.0.0
219	H+ <=> H+ (external)	0.0.0.0
220	Saccharose <=> Saccharose (external)	0.0.0.0

Table: A.1 . . . Continued from previous page

221	NH3 <=> NH3 (external)	0.0.0.0
222	Biomass <=> Biomass (external)	0.0.0.0
223	Sulfate <=> Sulfate (external)	0.0.0.0
224	Acetic acid <=> Acetic acid (external)	0.0.0.0
225	L-Gln <=> L-Gln (external)	0.0.0.0
226	external H2O <=>	0.0.0.0
227	external Oxygen <=>	0.0.0.0
228	external phosphate <=>	0.0.0.0
229	external CO2 <=>	0.0.0.0
230	external NH3 <=>	0.0.0.0
231	external glutamate <=>	0.0.0.0
232	external glucose <=>	0.0.0.0
233	external acetic acid <=>	0.0.0.0
234	external lysine <=>	0.0.0.0
235	external sulfate <=>	0.0.0.0
236	external glutamine <=>	0.0.0.0
237	external L-Phenylalanine <=>	0.0.0.0
238	external H+ <=>	0.0.0.0
239	external sucrose <=>	0.0.0.0
240	external fructose <=>	0.0.0.0
241	external lactate <=>	0.0.0.0
242	external maltose <=>	0.0.0.0
243	external trehalose <=>	0.0.0.0
244	external biomass <=>	0.0.0.0

Table: A.2: **Classification of the Networks Metabolites**
The classification of the networks metabolites into different pathways is not unambiguous. It is meant to be a hint.

	Glycolysis/Glyconeogenesis
1	(2R)-2-Hydroxy-3-(phosphonooxy)-propanal
2	2-Phospho-D-glycerate
3	3-Phospho-D-glycerate
4	3-Phospho-D-glyceroyl phosphate
5	D-Fructose
6	D-Fructose 1,6-bisphosphate
7	D-Fructose 1-phosphate
8	D-Fructose 6-phosphate
9	D-Glucose
10	D-Glucose 1-phosphate
11	D-Glucose 6-phosphate
12	Glycerone phosphate
13	Sucrose
14	alpha-D-Glucose 6-phosphate
15	beta-D-Fructose 1,6-bisphosphate
16	beta-D-Fructose 6-phosphate
	Pentose Phosphate Pathway
17	5-Phospho-alpha-D-ribose 1-diphosphate
18	6-Phospho-D-gluconate
19	D-Erythrose 4-phosphate
20	D-Glucono-1,5-lactone 6-phosphate
21	D-Ribose 5-phosphate
22	D-Ribulose 5-phosphate
23	D-Xylulose 5-phosphate
24	Sedoheptulose 7-phosphate

Table: A.2 ... Continued from previous page

	Tricarboxylic Acid Cycle
25	(S)-Malate
26	2-Oxoglutarate
27	Citrate
28	CoA
29	Dihydrolipoamide
30	FAD
31	FADH2
32	Fumarate
33	Isocitrate
34	Lipoamide
35	Oxaloacetate
36	S-Succinyldihydrolipoamide
37	Succinate
38	Succinyl-CoA
	Glyoxylate Shunt
39	Glyoxylate
	Pyruvate Metabolism
40	(S)-Lactate
41	Acetate
42	Acetyl phosphate
43	CO2
44	Phosphoenolpyruvate
45	Pyruvate
	Alanine Biosynthesis
46	L-Alanine
	Arginine Biosynthesis
47	Carbamoyl phosphate
48	L-Arginine
49	L-Citrulline
50	N-(L-Arginino)succinate

Table: A.2 . . . Continued from previous page

Asparagine/Aspartate Metabolism	
51	L-Asparagine
52	L-Aspartate
Cysteine Biosynthesis	
53	Hydrogen sulfide
54	L-Cysteine
55	L-Serine
56	O-Acetyl-L-serine
Glutamine/Glutamate Metabolism	
57	L-Glutamate
58	L-Glutamine
59	NH3
Glycine Biosynthesis	
60	Glycine
Histidine Biosynthesis	
61	3-(Imidazol-4-yl)-2-oxopropyl phosphate
62	5-(5-Phospho-D-ribosylaminoformimino)-1-(5-phosphoribosyl)-imidazole-4-carboxamide
63	D-erythro-1-(Imidazol-4-yl)glycerol 3-phosphate
64	L-Histidine
65	L-Histidinol
66	L-Histidinol phosphate
67	N-(5'-Phospho-D-1'-ribulosylformimino)-5-amino-1-(5''-phospho-D-ribosyl)-4-imidazolecarboxamide
68	Phosphoribosyl-AMP
69	Phosphoribosyl-ATP
Isoleucine Biosynthesis	
70	(R)-2,3-Dihydroxy-3-methylpentanoate
71	(R)-2-Oxo-3-methylpentanoate
72	(S)-2-Aceto-2-hydroxybutanoate
73	(S)-2-Hydroxy-3-methyl-3-oxopentanoate

Table: A.2 . . . Continued from previous page

74	2-Oxobutanoate
75	L-Isoleucine
Leucine Biosynthesis	
76	2,3-Dihydroxy-3-methylbutanoate
77	2-Acetolactate
78	2-Isopropylmalate
79	2-Isopropylmaleate
80	3-Carboxy-4-methyl-2-oxopentanoate
81	3-Isopropylmalate
82	3-Methyl-2-oxobutanoate
83	4-Methyl-2-oxopentanoate
84	L-Leucine
Lysine Biosynthesis	
85	2,3,4,5-Tetrahydrodipicolinate
86	2,3-Dihydrodipicolinate
87	4-Phospho-L-aspartate
88	L-Aspartate 4-semialdehyde
89	L-Homoserine
90	L-Lysine
91	LL-2,6-Diaminoheptanedioate
92	N-Succinyl-2-L-amino-6-oxoheptanedioate
93	N-Succinyl-LL-2,6-diaminoheptanedioate
94	NADP+
95	NADPH
96	meso-2,6-Diaminoheptanedioate
Methionine Biosynthesis	
97	5-Methyltetrahydrofolate
98	L-Cystathionine
99	L-Homocysteine
100	L-Methionine
101	O-Acetyl-L-homoserine

Table: A.2 . . . Continued from previous page

Ornithine Biosynthesis	
102	L-Ornithine
103	L-Proline
104	N-Acetyl-L-glutamate
105	N-Acetyl-L-glutamate 5-phosphate
106	N-Acetyl-L-glutamate 5-semialdehyde
107	N2-Acetyl-L-ornithine
Phenylalanine Biosynthesis	
108	2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate
109	3-Dehydroquinate
110	3-Dehydroshikimate
111	5-O-(1-Carboxyvinyl)-3-phosphoshikimate
112	Chorismate
113	L-Phenylalanine
114	Phenylpyruvate
115	Prephenate
116	Shikimate
117	Shikimate 3-phosphate
Proline Biosynthesis	
118	(S)-1-Pyrroline-5-carboxylate
119	L-Glutamate 5-semialdehyde
120	L-Glutamyl 5-phosphate
Serine Biosynthesis	
121	3-Phosphonooxypyruvate
122	O-Phospho-L-serine
Threonine Biosynthesis	
123	L-Threonine
124	O-Phospho-L-homoserine
Tryptophane Biosynthesis	
125	1-(2-Carboxyphenylamino)-1'-deoxy-D-ribulose 5'-phosphate
126	Anthranilate

Table: A.2 . . . Continued from previous page

127	Glyceraldehyde 3-phosphate
128	Indoleglycerol phosphate
129	L-Tryptophan
130	N-(5-Phospho-D-ribosyl)anthranilate
Tyrosine Biosynthesis	
131	3-(4-Hydroxyphenyl)pyruvate
132	L-Tyrosine
Valine Biosynthesis	
133	L-Valine
Fatty Acid Production	
134	Acetyl-CoA
135	Fatty Acid
136	Myristic Acid
137	Myristoleic Acid
Other Biomass Components	
138	3-Deoxy-D-manno-octulosonate
139	3-Deoxy-D-manno-octulosonate 8-phosphate
140	ADP-D-glycero-D-manno-heptose
141	CMP-3-deoxy-D-manno-octulosonate
142	CTP
143	D-Glyceraldehyde
144	D-Glycerate
145	Glycerol
146	H2O
147	NAD+
148	NADH
149	dATP
150	dCTP
151	dGTP
152	sn-Glycerol 3-phosphate

Table: A.2 . . . Continued from previous page

Biomass Production	
153	Biomass
Sugars Metabolism	
154	(1,4-alpha-D-Glucosyl)n
155	Maltose
156	alpha,alpha-Trehalose
157	alpha-D-Glucose
Trehalose Biosynthesis	
158	UDPglucose
159	alpha,alpha'-Trehalose 6-phosphate
Aminosugars Metabolism	
160	D-Glucosamine 1-phosphate
161	D-Glucosamine 6-phosphate
162	N-Acetyl-D-glucosamine 1-phosphate
163	UDP-N-acetyl-3-(1-carboxyvinyl)-D-glucosamine
164	UDP-N-acetyl-D-glucosamine
165	UDP-N-acetylmuramate
Purine Metabolism	
166	1-(5'-Phosphoribosyl)-5-amino-4-(N-succinocarboxamide)-imidazole
167	1-(5'-Phosphoribosyl)-5-amino-4-imidazolecarboxamide
168	1-(5'-Phosphoribosyl)-5-formamido-4-imidazolecarboxamide
169	1-(5-Phospho-D-ribosyl)-5-amino-4-imidazolecarboxylate
170	10-Formyltetrahydrofolate
171	2-(Formamido)-N1-(5'-phosphoribosyl)acetamidine
172	5'-Phosphoribosyl-N-formylglycinamide
173	5'-Phosphoribosylglycinamide
174	5-Phosphoribosylamine
175	ADP
176	AMP
177	ATP
178	Aminoimidazole ribotide

Table: A.2 ... Continued from previous page

179	GDP
180	GMP
181	GTP
182	H+
183	IMP
184	N6-(1,2-Dicarboxyethyl)-AMP
185	Orthophosphate
186	Pyrophosphate
187	Tetrahydrofolate
188	Xanthosine 5'-phosphate
Pyrimidine Metabolism	
189	(S)-Dihydroorotate
190	CDP
191	N-Carbamoyl-L-aspartate
192	Orotate
193	Orotidine 5'-phosphate
194	UDP
195	UMP
196	UTP
197	dTDP
198	dTMP
199	dTTP
200	dUDP
201	dUMP
Tetrahydrofolate Metabolism	
202	5,10-Methenyltetrahydrofolate
203	5,10-Methylenetetrahydrofolate
204	Dihydrofolate
Sulfur metabolism	
205	Sulfate

Table: A.2 ... Continued from previous page

Respiratory Reactions	
206	Oxygen
Membrane Transport Reactions	
207	external CO2
208	external H+
209	external H2O
210	external L-Phenylalanine
211	external NH3
212	external Oxygen
213	external acetic acid
214	external biomass
215	external fructose
216	external glucose
217	external glutamate
218	external glutamine
219	external lactate
220	external lysine
221	external maltose
222	external phosphate
223	external sucrose
224	external sulfate
225	external trehalose

Appendix B

Tables

Table: B.1: **Enzyme data**

All enzymes used in the *Corynebacterium glutamicum* model, separated into the categories: expressed, weakly expresses, and non-annotated.

Enzymes expressed ($R \geq 2$)
1.1.1.158, 1.1.1.2, 1.1.1.205, 1.1.1.23, 1.1.1.25, 1.1.1.27, 1.1.1.3, 1.1.1.37, 1.1.1.42, 1.1.1.44, 1.1.1.49, 1.1.1.85, 1.1.1.86, 1.1.1.94, 1.1.1.95, 1.10.2.2, 1.2.1.12, 1.2.1.3, 1.2.1.38, 1.2.1.41, 1.2.4.1, 1.2.4.2, 1.3.1.12, 1.3.1.26, 1.3.99.1, 1.4.1.13, 1.4.1.4, 1.5.1.2, 1.5.1.3, 1.5.1.5, 1.6.5.3, 1.7.99.5, 1.8.1.2, 1.8.1.4, 1.9.3.1, 2.1.1.13, 2.1.1.45, 2.1.2.1, 2.1.2.10, 2.1.2.2, 2.1.3.3, 2.2.1.1, 2.2.1.2, 2.2.1.6, 2.3.1.117, 2.3.1.30, 2.3.1.31, 2.3.1.61, 2.3.1.8, 2.3.1.85, 2.3.3.1, 2.3.3.13, 2.3.3.9, 2.4.1.1, 2.4.1.15, 2.4.1.25, 2.4.2.-, 2.4.2.10, 2.4.2.14, 2.4.2.17, 2.4.2.18, 2.5.1.19, 2.5.1.47, 2.5.1.49, 2.5.1.54, 2.5.1.7, 2.6.1.1, 2.6.1.11, 2.6.1.16, 2.6.1.17, 2.6.1.42, 2.6.1.52, 2.6.1.9, 2.7.1.11, 2.7.1.2, 2.7.1.23, 2.7.1.30, 2.7.1.31, 2.7.1.39, 2.7.1.4, 2.7.1.40, 2.7.1.56, 2.7.2.1, 2.7.2.11, 2.7.2.3, 2.7.2.4, 2.7.2.8, 2.7.4.14, 2.7.4.3, 2.7.4.6, 2.7.4.8, 2.7.6.1, 2.7.7.23, 2.7.7.9, 3.1.1.31, 3.1.3.-, 3.1.3.11, 3.1.3.3, 3.2.1.20, 3.5.1.14, 3.5.1.18, 3.5.1.2, 3.5.2.3, 3.5.4.10, 3.5.4.19, 3.6.1.1, 3.6.1.31, 4.1.1.20, 4.1.1.21, 4.1.1.23, 4.1.1.31, 4.1.1.48, 4.1.2.13, 4.1.3.1, 4.1.3.27, 4.2.1.10, 4.2.1.11, 4.2.1.19, 4.2.1.2, 4.2.1.20, 4.2.1.3, 4.2.1.33, 4.2.1.51, 4.2.1.52, 4.2.1.9, 4.2.3.1, 4.2.3.4, 4.3.1.12, 4.3.1.19, 4.3.2.1, 4.3.2.2, 4.4.1.8, 5.1.1.7, 5.1.3.1, 5.3.1.1, 5.3.1.16, 5.3.1.24, 5.3.1.6, 5.3.1.9, 5.4.2.1, 5.4.2.2, 6.2.1.5, 6.3.1.2, 6.3.2.6, 6.3.3.1, 6.3.4.13, 6.3.4.2, 6.3.4.4, 6.3.4.5, 6.3.5.2, 6.3.5.3, 6.3.5.4
Enzymes weakly expressed ($R \leq 2$)
1.1.1.1, 1.2.1.11, 1.2.1.59, 2.1.2.3, 2.1.3.2, 2.3.1.12, 2.7.1.71, 2.7.4.9, 3.1.3.12, 4.2.3.5, 5.4.99.16
Enzymes non-annotated
1.3.5.1, 1.3.99.11, 1.6.1.1, 1.6.1.2, 1.8.1.9, 1.17.4.2, 1.8.2.1, 2.3.1.1, 2.3.1.157, 2.5.1.55, 5.3.1.13, 2.6.1.2, 2.7.1.3, 2.7.2.12, 2.7.7.38, 3.1.3.15, 3.1.3.45, 3.5.4.9, 3.6.1.5, 3.6.1.8, 4.1.1.3, 5.4.99.5, 6.3.4.16

Table: B.2: **Gibbs energies of external molecules**
ΔG_f^0 values of all metabolites able to cross the cellular membrane; calculated by Kai Hartmann as described in 2.1.2.1. Since all amino acids are assumed to form proteins, a correcting summand (18.8 KJ/mol) is added to their ΔG_f^0 values, taking the resulting peptid-bonds into account. *Water* is assumed to have an extracellular concentration of 55.5 mol/L.

Metabolite	$\Delta G_f^0 \left[\frac{\text{KJ}}{\text{mol}}\right]$	Metabolite	$\Delta G_f^0 \left[\frac{\text{KJ}}{\text{mol}}\right]$
Biomass molecules			
ADP-D-glycero-D-manno-heptose	-1051.9	L-Isoleucine	-330.1
		L-Leucine	-330.1
ATP	-193.3	L-Lysine	-336.8
CMP-3-deoxy-D-manno-octulosonate	-1606.7	L-Methionine	-296.3
		L-Ornithine	-362.8
CTP	-573.2	L-Phenylalanine	-194.2
dATP	-24.7	L-Proline	-285.0
dCTP	-404.6	L-Serine	-498.8
dGTP	-202.1	L-Threonine	-504.2
dTTP	-573.2	L-Tryptophan	-95.8
Fatty Acid	-366.9	L-Tyrosine	-356.1
Glycine	-356.1	L-Valine	-337.3
GTP	-370.7	meso-2,6-Diaminoheptanedioate	-684.1
L-Alanine	-350.2		
L-Arginine	-364.5	Myristic Acid	-281.6
L-Asparagine	-490.4	Myristoleic Acid	-202.9
L-Aspartate	-677.4	sn-Glycerol 3-phosphate	-484.5
L-Cysteine	-320.1	UDPglucose	-1473.6
L-Glutamate	-670.3	UDP-N-acetyl-D-glucosamine	-1432.2
L-Glutamine	-483.3	UDP-N-acetylmuramate	-1680.7
L-Histidine	-210.5	UTP	-769.0
Other molecules			
Acetate	-366.9	Oxygen	+16.4
Ammonia	-79.3	Phosphate	-166.7
Carbon dioxide	-326.4	Sulfate	-744.53
Glucose	-886.2	Water	-227.17
H^+	± 0		

Appendix C

Proof: required concentration-ratios

Given the assumptions stated in chapter 3.6.2.2, reaction rates follow the general law:

$$v_1 = k_1 * \prod_i [i]^{\nu_i} \tag{C.1}$$

where v_1 is the actual reaction velocity of the forward reaction, k is the kinetic constant, $[i]$ are the concentrations of all metabolites i and ν_i is the order of the reaction in the referring metabolite i (Atkins, 1990, p. 782 ff.). Accordingly

$$v_2 = k_2 * \prod_j [j]^{\nu_j} \tag{C.2}$$

states the rate-law for the backward direction.

In the following we will replace the parts

$$\prod_i [i]^{\nu_i} \quad \text{by} \quad A$$

$$\text{and} \tag{C.3}$$

$$\prod_j [j]^{\nu_j} \quad \text{by} \quad B$$

From 2.3 we know that the total reaction rate is given as

$$v = v_1 - v_2 \tag{C.4}$$

The only reactions taken into account are those which definitely have a forward flux attached to them:

$$v > 0 \Leftrightarrow v_1 > v_2 \tag{C.5}$$

Now, we want to examine the ratio of the rates of the same reaction under different circumstances, the second of which shall be marked by a prime ($'$): v and v'.

From equation C.1 and C.3 we directly can see that $v_1 > v_1'$ if $A > A'$.

As said in section 3.6.2.2, there is no problem if $B \leq B'$. Otherwise, the backward reaction rate increases due to an increasing B. In that case, it may happen that the increase in v_2 outruns the increase in v_1 and the overall reaction velocity *decreases* despite an increase in A against A'.

Our statement concerning this problem is:

Statement

The following conditions

$$A \quad > \quad A'$$

$$\textit{and} \tag{C.6}$$

$$\frac{A}{A'} \quad > \quad \frac{B}{B'}$$

are sufficient to ensure a ratio

$$\frac{v}{v'} > 1 \tag{C.7}$$

even if $B > B'$.

Given

$$\frac{A}{A'} > 1 \tag{C.8}$$

$$\frac{B}{B'} > 1 \tag{C.9}$$

$$v_1 > v_2 \quad \Leftrightarrow \quad k_1 A > k_2 B \tag{C.10}$$

$$v_1' > v_2' \quad \Leftrightarrow \quad k_1 A' > k_2 B' \tag{C.11}$$

From basic chemistry it is known that rate constants are always positive. A and B are replacement characters for products of metabolic concentrations, which are positive by definition.

$$k_1, k_2, k_1', k_2' \quad > \quad 0 \tag{C.12}$$

$$A, B, A', B' \quad > \quad 0 \tag{C.13}$$

Proof

Inserting equation C.3 in C.1 and C.2, we get, with the help of equation C.4, for the ratio v/v':

$$\frac{v}{v'} = \frac{k_1 A - k_2 B}{k_1 A' - k_2 B'} \tag{C.14}$$

reduce by $k_1 A' + k_2 B'$

$$\begin{aligned}
\frac{v}{v'} &= \frac{(k_1 A - k_2 B) * (k_1 A' + k_2 B')}{(k_1 A' - k_2 B') * (k_1 A' + k_2 B')} \\
&= \frac{k_1^2 AA' - k_2^2 BB' + k_1 k_2 (AB' - A'B)}{(k_1^2 A'^2 - k_2^2 B'^2)} \\
&= \frac{k_1^2 AA' - k_2^2 BB'}{(k_1^2 A'^2 - k_2^2 B'^2)} + \frac{k_1 k_2 (AB' - A'B)}{(k_1^2 A'^2 - k_2^2 B'^2)} \qquad \text{(C.15)}
\end{aligned}$$

To achieve equation C.7 it is sufficient, that the first part of C.15 is greater one, while the second part is positive. Thus, it must be that:

$$\frac{k_1^2 AA' - k_2^2 BB'}{(k_1^2 A'^2 - k_2^2 B'^2)} > 1 \qquad \text{(C.16)}$$

$$\Leftrightarrow \quad k_1^2 AA' - k_2^2 BB' > k_1^2 A'^2 - k_2^2 B'^2 \qquad \text{(C.17)}$$

$$\Leftrightarrow \quad \frac{k_1^2 AA' A'}{A'} - k_1^2 A'^2 > \frac{k_2^2 BB' B'}{B'} - k_2^2 B'^2 \qquad \text{(C.18)}$$

$$\Leftrightarrow \quad k_1^2 A'^2 * \left(\frac{A}{A'} - 1\right) > k_2^2 B'^2 * \left(\frac{B}{B'} - 1\right) \qquad \text{(C.19)}$$

Using squared equation C.11, it is sufficient that the following equation is fulfilled for C.16 to hold true:

$$\begin{aligned}
\left(\frac{A}{A'} - 1\right) &> \left(\frac{B}{B'} - 1\right) \\
\Leftrightarrow \quad \frac{A}{A'} &> \frac{B}{B'} \qquad \text{(C.20)}
\end{aligned}$$

As said above, additionally to C.16, the second part of equation C.15 must be positive to ensure C.7:

$$\frac{k_1 k_2 (AB' - A'B)}{k_1^2 A'^2 - k_2^2 B'^2} > 0 \tag{C.21}$$

The square of equation C.11 tells us, that the denominator of C.21 is always positive. Then, obviously the condition of the numerator to be positive is ample to finish the proof.

$$k_1 k_2 (AB' - A'B) \quad > \quad 0 \tag{C.22}$$

$$\Leftrightarrow \quad AB' \quad > \quad A'B \tag{C.23}$$

$$\Leftrightarrow \quad \frac{A}{A'} \quad > \quad \frac{B}{B'} \tag{C.24}$$

The last transformation is possible, because $A', B' > 0$ (eq. C.13).

Equation C.24 is exactly the same as C.20 which again equals the statement in C.6.

Thus, if equation C.6 is fulfilled, then C.7 is ensured.

q.e.d.

Bibliography

Al-Lazikani, B., Jung, J., Xiang, Z. & Honig, B. (2001). Protein structure prediction. *Current Opinion in Structural Biology*, **5**, 51–56.

Atkins, P.W. (1990). *Physical Chemistry*. Oxford University Press, Oxford, 4th edition.

Beard, D.A. (2005). Computational tools download website: `http://bbc.mcw.edu/Computation/projects/computationaltools.htm`, last visited july 2005.

Beard, D.A., Babson, E., Curtis, E. & Qian, H. (2004). Thermodynamic constraints for biochemical networks. *Journal of Theoretical Biology*, **228**, 327–333.

Beard, D.A., Liang, S. & Qian, H. (2002). Energy Balance for Analysis of Complex Metabolic Networks. *Biophysical Journal*, **83**, 79–86.

Beard, D.A. & Qian, H. (2005). Thermodynamic-based computational profiling of cellular regulatory control in hepatocyte metabolism. *American Journal of Physiology - Endocrinology and Metabolism*, **288**, E633–E644.

Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Wayne Davis, N., Kikpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. & Shao, Y. (1997). The Complete Genome Sequence od *Escherichia coli* K-12. *Science*, **277**, 1453–1462.

Cornish-Bowden, Athel (1995). *Fundamentals of Enzyme Kinetics*. Portland Press Ltd., London.

Edwards, J.S., Ibarra, R.U. & Palsson, B.Ø. (2001). *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology*, **19**(2), 125–130.

Edwards, J.S. & Palsson, B.Ø. (1999). Systems Properties of the *Haemophilus influenzae* Rd Metabolic Genotype. *Journal of Biological Chemistry*, **274**(25), 17410–17416.

Edwards, J.S. & Palsson, B.Ø. (2000). The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the USA*, **97**(10), 5528–5533.

Edwards, J.S., Ramakrishna, R. & Palsson, B.Ø. (2002). Characterizing the metabolic phenotype: A Phenotype Phase Plane Analysis. *Biotechnology and Bioengineering*, **77**(1), 27–36.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L., Glodek, A., Kelley, J.M., Weidman, J.F., Phillipps, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O. & Venter, J.C. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269**, 496–512.

Förster, J., Famili, I., Fu, P., Palsson, B.Ø. & Nielsen, J. (2003). Genome-Scale Reconstruction of the *Saccharomyces cervisiae* Metabolic Network. *Genome Research*, **13**(2), 244–253.

Gerhard, M. (1999). *Biochemical Pathways.* Spektrum. Akad. Verl., Heidelberg, Berlin.

Goto, S., Hattori, M., Nishioka, T. & Kanehisa, M. (2004). LIGAND - Database of Chemical Compounds and Reactions in Biological Pathways, Manual `ftp://ftp.genome.ad.jp/pub/kegg/ligand/ligand.doc`, last visited july 2005.

Hartmann, K. (2006). *Title to be defined.* PhD thesis, *Cologne University Bioinformatics Center*. In preparation.

Hollemann, A.F. & Wiberg, E. (1995). *Lehrbuch der Anorganischen Chemie.* de Gruyter, Berlin, New York, 101st edition.

Holzhütter, H.G. (2004). The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *European Journal of Biochemistry*, **271**, 2905–2922.

Hornemann, Eliane (2005). Metabolomanalyse von *Corynebacterium glutamicum* nach Kultivierung auf verschiedenen Kohlenstoff- und Stickstoffquellen.

Hüser, A.T., Becker, A., Brune, I., Dondrup, M., Kalinowski, J., Plassmeier, J., Pühler, A., Wiegräbe, I. & Tauch, A. (2003). Development of a *Corynebacterium glutamicum* DNA microarray and validation by genome-wide expression profiling during growth with propionate as carbon source. *Journal of Biotechnology*, **106**, 269–286.

IBM (1995). OSL V1R2.1 Guide and Reference, website: `http://publibfp.boulder.ibm.com/cgi-bin/bookmgr/BOOKS/ekkosl01/CCONTENTS`, last visited july 2005.

Ikeda, M. & Nakagawa, S. (2003). The *Corynebacterium glutamicum* genome: features and impacts in biotechnological processes. *Applied Microbiology and Biotechnology*, **62**(2-3), 99–109.

Jason, A.P., Stelling, J., Price, N.D., Klamt, S., Schuster, S. & Palsson, B.Ø. (2004). Comparison of network-based pathway analysis methods. *Trends in Biotechnology*, **22**(8), 400–405.

J.F., Tomb, O., White, A.R., Kerlavage, R.A., Clayton, G.G., Sutton, R.D., Fleischmann, K.A., Ketchum, H.P., Klenk, S., Gill, B.A., Dougherty, K., Nelson, J., Quackenbush, L.X., Zhou, E.F., Kirkness, S., Peterson, B., Loftus, D., Richardson, R., Dodson, H.G., Khalak, A., Glodek, K., McKenney, L.M., Fitzgerald, N., Lee, M.D., Adams, E.K., Hickey, D.E., Berg, J.D., Gocayne, T.R., Utterback, J.D., Peterson, J.M., Kelley, M.D., Cotton, J.M., Weidman, C., Fujii, C., Bowman, L., Watthey, E., Wallin, W.S., Hayes, M., Borodovsky, P.D., Karp, H.O., Smith, C.M., Fraser & J.C., Venter (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori* . *Nature*, **388**, 539–547.

Kalinowski, J., Bathe, B., Bartels, D., Bischoff, N., Bott, M., Burkovski, A., Dusch, N., Eggeling, L., Eikmanns, BJ., Gaigalat, L., Goesmann, A., Hartmann, M., Huthmacher, K., Kramer, R., Linke, B., McHardy, AC., Meyer, F., Mockel, B., Pfefferle, W., Puhler, A., Rey, D.A., Ruckert, C., Rupp, O., Sahm, H., Wendisch, V.F., Wiegrabe, I. & Tauch, A. (2003). The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *Journal of Biotechnology*, **104**(1-3), 5–25.

Kanehisha, M. (1997). A database for post-genome analysis. *Trends in Genetics*, **13**, 375–376.

Kanehisha, M. (2005). Kyoto Encyclopedia of Genes and Genomes website: `http://www.kegg.com`, last visited july 2005.

Kanehisha, M. & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acid Research*, **28**, 27–30.

Kauffman, K.J., Prakash, P. & Edwards, J.S. (2003). Advances in flux balance analysis. *Current Opinion in Biotechnology*, **24**, 491–496.

Kiefer, P., Heinzle, E., Zelder, O. & Wittmann, C. (2004). Comparative Metabolic Flux Analysis of Lysine-Producing *Corynebacterium glutamicum* Cultured on Glucose or Fructose. *Applied and Environmental Microbiology*, **70**(1), 229–239.

Koehler, W., Schachtel, G. & Voleske, P. (2002). *Biostatistik*. Springer, Berlin Heidelberg New York, 3rd edition.

Lawler, E. (2001). *Combinatorial Optimization, Networks and Matroids*. Dover Publications, Inc., Mineola, New York.

Mahadevan, R. & Schilling, C.H. (2003). The effects of alternate optimal solutions in constrained-based genome-scale metabolic models.

Mandriva (2003). Mandriva linux website: `http://www.mandriva.com/`, last visited july 2005.

Marx, A., de Graaf, A.A., Wiechert, W., L., Eggeling & Sahm, H. (1996). Determination of the fluxes in the central carbon metabolism of *Corynebacterium glutamicum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing. *Biotechnology & Bioengineering*, **49**, 111–129.

Mavrovouniotis, M.L. (1998). Estimation of Standard Gibbs Energy Changes of Biotransformations. *Journal of biological Chemistry*, **266**(22), 14440–14445.

Mewes, H.W., Albermann, K., Bähr, M., Frishman, D., Gleissner, A., Hani, J., Heumann, K., Kleine, K., Maierl, A., Oliver, S.G., Pfeiffer, F. & Zollner, A. (1997). Overview of the yeast genome. *Nature*, **387**, 7–8.

Microsoft Corporation (2003). Microsoft website: `http://www.microsoft.com`, last visited july 2005.

Osguthorpe, D.J. (1999). Improved *ab initio* predictions with a simplified, flexible geometry model.

Osguthorpe, D.J. (2000). *Ab initio* protein folding.

Palsson, B.O. (2005). Systems Biology Research Group from the University of California, San Diego: `http://gcrg.ucsd.edu`, last visited february 2006.

Pinney, J.W., Westhead, D.R. & McConkey, G.A. (2003). Petri Net representations in systems biology. *Biochemical Society Transactions*, **31**, 1513–1515.

Qian, H., Beard, D.A. & Liang, S. (2003). Stoichiometric network theory for nonequilibrium biochemical systems. *European Journal of Biochemistry*, **270**, 415–421.

Rahman, S.A., Advani, P., Schunk, R. & Schomburg, D. (2005). Metabolic pathway analysis web service (Pathway Hunter Tool ant CUBIC). *Bioinformatics*, **21**(7), 1189–1193.

Reed, J.L., Famili, I., Thiele, I. & Palsson, B.Ø. (2006). Towards multidimensional genome annotation. *Nature Reviews Genetics*, **7**, 130–141.

Reed, J.L. & Palsson, B.Ø. (2003). Thirteen years of Building Constraint-Based In Silico Models of *Escherichia coli* . *Journal of Bacteriology*, **185**(9), 2692–2699.

Reed, J.L., Vo, T.D., Schilling, C.H. & Palsson, B.Ø. (2003). An expanded genome-scale model of *Escherichia coli* K-12(*i*JR904 GSM/GPR). *Genome Biology*, **4**(9).

Schauer, N., Steinhauser, D., Strelkov, S., Schomburg, D., Allison, G., Moritz, T., Lundgren, K., Roessner-Tunali, U., Forbes, M.G., Willmitzer, L., Fernie, A.R. & Kopka, J. (2005). GC-MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS letters*, **579**, 1332–1337.

Schilling, C.H., Covert, M.W., Iman, F., Church, G.M., Edwards, J.S. & Palsson, B.Ø. (2002). Genome-Scale Metabolic Model of *Helicobacter pylori* 26695. *Journal of Bacteriology*, **184**(16), 4582–4593.

Schilling, C.H., Edwards, S.J., Letscher, D. & Palsson, B.Ø. (2001). Combining Pathway Analysis with Flux Balance Analysis for the Comprehensive Study of Metabolic Systems. *Biotechnolgy and Bioengineering*, **71**(4), 286–306.

Schilling, C.H., Letscher, D. & Palsson, B.Ø. (2000). Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective. *Journal of theoretical Biology*, **203**, 229–248.

Schlegel, H.G. (1985). *Allgemeine Mikrobiologie*. Georg Thieme Verlag, Stuttgart, New York, 6th edition.

Schomburg, D. (1987). BRENDA: `http://www.brenda.uni-koeln.de`, last visited november 2005.

Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G. & Schomburg, D. (2004). BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Research*, **32** Database issue, D431–D433.

Schuster, S., Dandekar, T. & Fell, D.A. (1999). Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering. *Trends in Biotechnology*, **17**(2), 53–60.

Schuster, S., Fell, D.A. & Dandekar, T. (2000). A general definition of metabolic pathways for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, **18**(3), 326–332.

Segrè, D., Vitkup, D. & Church, G.M. (2002). Analysis of optimality in natural and pertubed metabolic networks. *PNAS*, **99**(23), 15112–15117.

Silberbach, M. (2005). *Biochemische, biotechnologische und molekularbiologische Analyse eines Signaltransduktionsweges in Corynebacterium glutamicum.* PhD thesis, *Cologne University Bioinformatics Center.*

Silberbach, M., Hüser, A., Kalinowski, J., Pühler, A., Walter, B., Krämer, R. & Burkovski, A. (2005). DNA microarray analysis of the nitrogen starvation response of *Corynebacterium glutamicum.* *Journal of Biotechnology*, **119**(4), 357–367.

Strelkov, S., von Elstermann, M. & Schomburg, D. (2004). Comprehensive analysis of metabolites in *Corynebacterium glutamicum* by gas chromatography/mass spectrometry. *Biological Chemistry*, **385**, 853–861.

Stroustrup, B. (1997). *The C++ programming language.* Addison-Wesley Publishing Company, Massachusetts, USA, 3rd edition.

SUN Microsystems (2004). Open Office package website: `http://www.openoffice.org`, last visited july 2005.

Takaç, S., Güzide, C., Mavutina, F. & Dervakos, G. (1998). Metabolic flux distribution for the optimized production of L-glutamate. *Enzyme and Microbial Technology*, **23**, 286–300.

Tatusov, R.L., Mushegian, A.R., Bork, P., Brown, N.P., Hayes, W.S., Borodovsky, M., Rudd, K.E. & Koonin, E.V. (1996). Metabolism and evolution of Haemophilus influenzae deduced from a whole-genome comparison with *Escherichia coli* . *Current Biology*, 6(3), 279–91.

The MathWorks (2005). Matlab website: `http://www.mathworks.com/`, last visited july 2005.

Thiele, I., Vo, T.D., Price, N.D. & Palsson, B.Ø. (2005). Expanded Metabolic Reconstruction of *Helicobacter pylori* (*i*IT341 GSM/GPR): an In Silico Genome-Scale Characterization of Single- and Double-Deletion Mutants. *Journal of Bacteriology*, **187**(16), 5818–5830.

171

Vajda, S. & Camacho, C.J. (2004). Protein-protein docking: is the glass half-ful or half-empty? *Trends in Biotechnology*, **22**, 110–116.

Varma, A. & Palsson, B.Ø. (1995). Parametric Sensitivity of Stoichiometric Flux Balance Models Applied to Wild-Type *Escherichia coli* Metabolism. *Biotechnology and Bioengineering*, **45**, 69–79.

Voet, D. & Voet, J.G. (1992). *Biochemie.* VCH Verlagsgesellschaft mbH, Weinheim, Germany, 1st edition.

Wall, L., Cristiansen, T. & Orwant, J. (2000). *Programming Perl.* O'Reilly & Assiciates, Inc., Sebastopol, USA, 3rd edition.

Wendisch, V.F., De Graaf, A.A., Sahm, H. & Eikmanns, B.J. (2000). Quantitative Determination of Metabolic Fluxes during Coutilization of Two Carbon Sources: Comparative Analyses with *Corynebacterium glutamicum* during Growth on Acetate and/or Glucose. *Journal if Bacteriology*, **182**(11), 3088–3096.

Wittmann, C., Kiefer, P. & Zelder, O. (2004). Metabolic Fluxes in *Corynebacterium glutamicum* during Lysine Production with Sucrose as Carbon Source. *Applied and Environmental Microbiology*, **70**(12), 7277–7287.

Yang, F., Qian, H & Beard, D.A. (2005). *Ab Initio* Prediction of Thermodynamically Feasible Reaction Directions from Biochemical Network Stoichiometry. *Metabolic Engineering*, in Press.

Kurzzusammenfassung

In den letzten Jahren expandierte das Feld der Bioinformatik von seinem bisherigen Fokus der proteinverwandten Themen zur Untersuchung ganzer, zellulärer Organismen. Mittlerweile wurde eine ganze Zahl von Bakterien modelliert, wobei das am detailliertesten untersuchte mit Sicherheit *Escherichia coli* ist. Als systemische Untersuchungsmethode hat sich in der wissenschaftlichen Gemeinschaft die sogenannte *Flux Balance Analysis* (*FBA*) etabliert. Darin wird das zu untersuchende, metabolische Netzwerk als Matrix dargestellt, die "stöchiometrische Matrix". Unter der Annahme dass System befände sich in einem (temporären) Fließgleichgewicht erhält man ein homogenes lineares Gleichungssystem, welches typischerweise unterbestimmt ist. Eine Zielfunktion wird ausgewählt und das entstehende lineare Program gelöst. Als Ergebnis kann so eine diskrete Lösung aus dem Lösungsraum ausgewählt werden. In der vorliegenden Arbeit wurde ein genombasiertes Modell des *Corynebacterium glutamicum* erstellt und mit Hilfe der *FBA* analysiert. Wir haben eine Verbesserung der *Flux Balance Analysis* implementiert, welche auch thermodynamische Aspekte in Betracht zieht, die *Energy Balance Analysis* (*EBA*). Weiterhin verwendeten wir Daten aus Metabolic-Profiling Experimenten, um den Analysen nach *EBA* zusätzliche Beschränkungen aufzuerlegen. Der Vergleich des Organismus unter verschiedenen Umweltbedingungen ermöglicht es, unbekannte kinetische Konstanten zu eliminieren und neue Anforderungen während der *Energy Balance Analysis* zu stellen. Die notwendigen Daten wurden durch Aufzucht des Bakteriums auf Glukose oder Acetat gewonnen. Die Methodik führt zu einer weiteren Verkleinerung des Lösungsraumes und hilft auf diese Weise die Lücke zwischen Theorie und echtem Leben zu schließen. Die übergeordnete Natur der entwickelten Technik ermöglicht deren Anwendung mit jedem Modell und deren Kombination mit allen weiteren, möglichen Verbesserungen der *Flux Balance Analysis*.

Abstract

Within the last years, bioinformatics expanded from its focus on protein related research to the investigation of whole organisms. Up to date, a variety of bacteria has been modeled, in most detail *Escherichia coli.* As a systemic approach, flux balance analysis (*FBA*) has established itself in the scientific community to analyze steady state flux distributions. Within *FBA* the metabolic network is expressed in terms of a matrix, called the stoichiometric matrix. The assumption of the system to exist in a (temporary) steady state leads to a homogeneous linear system of equations, which is typically underdetermined. By application of an objective function and computation of the linear program that unfolds, one can select one discrete solution out of the existing solution space. In this work, we built a genome based model of the *Corynebacterium glutamicum* and analyzed it in terms of flux balance analysis. We implemented an enhancement of *FBA*, called energy balance analysis, that considers thermodynamical issues. We further used metabolic profiling data to impose more constraints on the analyses. By comparing the organism under different environmental conditions, we were able to neglect unknown kinetic constants and to establish new requirements during the energy balance analysis. Namely, we used data derived by raising the *C. glutamicum* on acetate or glucose. This procedure leads to a further reduction of the solution space and thereby helps to close the gap between predictions and real-life flux distributions. The comprehensive nature of the technique enables it to be applied to any model and to be combined with any other enhancement of the flux balance analysis.

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. D. Schomburg betreut worden.

- Keine Teilpublikationen -

———————————————————

Lebenslauf

Thorsten Fallisch
geboren am 9. Oktober 1974
verheiratet
deutsch

Lehrtätigkeit

11/2001 - 05/2006	Universität zu Köln,
	Dozent, Wissenschaftlicher Mitarbeiter
	Aufgaben: Konzeptionieren und Durchführen einer
	englischsprachigen Vorlesung (40 Std./Jahr) für
	internationale Stipendiaten
	Thema: Matrizencodierung chemischer Reaktionen,
	Anforderungen und Grenzen der Softwareentwicklung in
	der Bioinformatik, Analyse und Simulation genombasierter
	Stoffwechselmodelle - verschiedene Methoden im Vergleich
10/2003-04/2004	Supervision der Praktikumsarbeit einer Studentin
	Thema der Arbeit: *Analyse der Berechnungsdauer aller*
	"Elementary Modes" eines zellulären Netzwerkes in
	Abhängigkeit von Netzwerkgröße und -struktur

Studium

11/2001 - 05/2006	Universität zu Köln,
	Promotionsstudium Biochemie
	<u>Thema</u>: *Aufbau und Simulation eines genombasierten*
	metabolischen Modells, Entwicklung von Konzepten zur
	Integration metabolischer Profildaten in die Simulation
10.10.2001	Diplom Chemie
08/2001 - 10/2001	Schriftliche Ausarbeitung der Diplomarbeit
	<u>Titel</u>: *Vorhersage der SCOP-Klassifizierung aus*
	Alignmentdaten unter Verwendung neuronaler Netzwerke
02/2001 - 08/2001	Auslandsaufenthalt am *Stockholm Bioinformatics Center*,
	Universität Stockholm zur Durchführung der Diplomarbeit
02.06.1998	Vordiplom Chemie
04/1996	Universität zu Köln, Beginn des Chemiestudiums
10/1995 - 04/1996	Universität Münster, Studium der Rechtswissenschaften

Zivildienst / Ausbildung

07/1994 - 09/1995	Rettungssanitäter bei der Johanniter-Unfall-Hilfe e.V.
07/1994 - 12/1994	Ausbildung zum Rettungssanitäter

Schulausbildung

26.05.1994	Abitur

Köln, 31. Januar 2006

_____