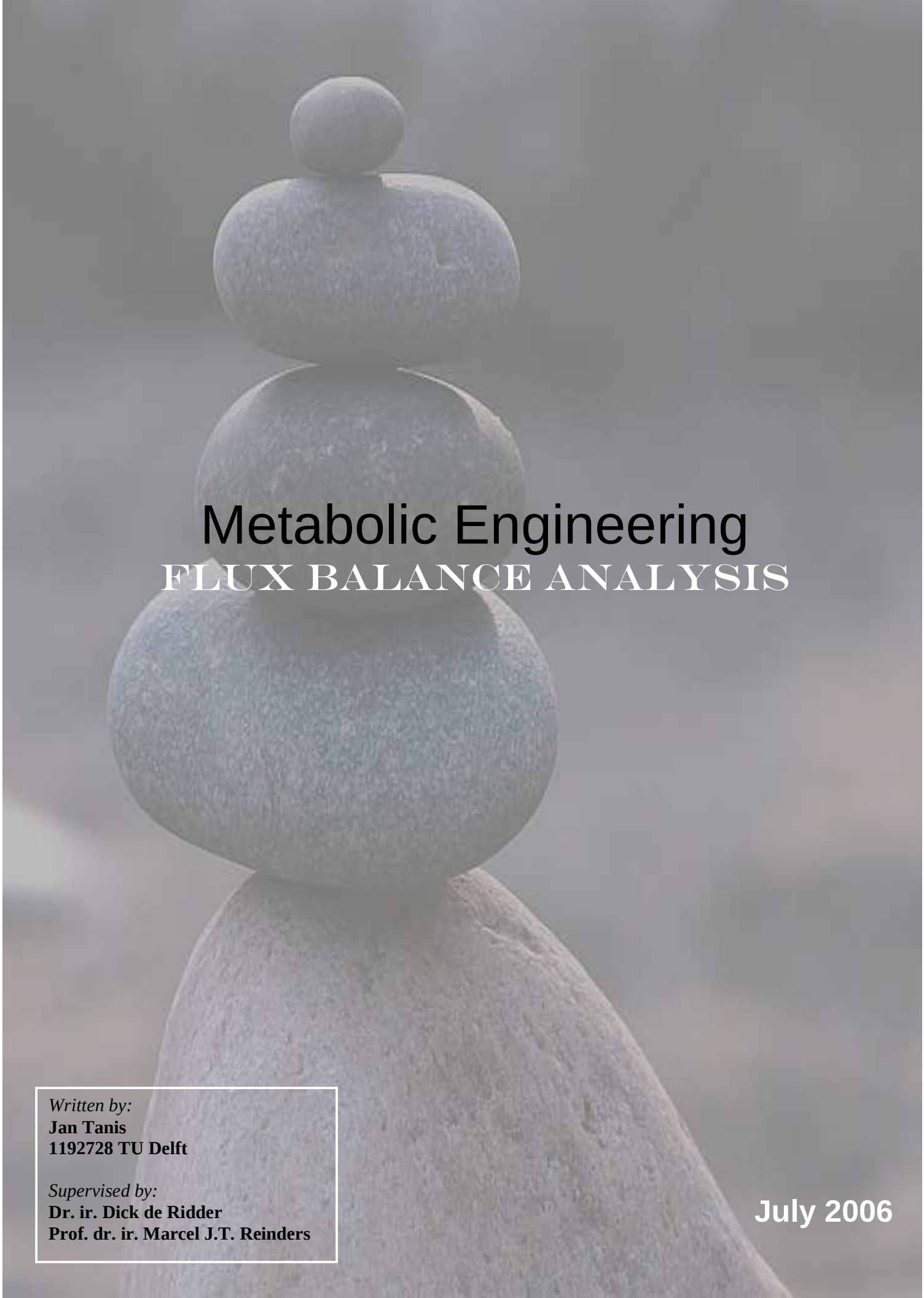

Metabolic Engineering
FLUX BALANCE ANALYSIS

Written by:
Jan Tanis
1192728 TU Delft

Supervised by:
Dr. ir. Dick de Ridder
Prof. dr. ir. Marcel J.T. Reinders

July 2006

Abstract

Biologists are interested in the capabilities of metabolic networks. A number of mathematical and biological models were developed to predict these capabilities based on stoichiometric and kinetic information. The predicted outcomes can help to create new hypotheses which can be tested and used within metabolic engineering to optimize certain pathways in an organism.

Three models will be explained and discussed in this paper. *Metabolic control analysis* will predict single solutions, but requires lots of information about the parameters of the enzymes in the organism. This will form a drawback of the method, because a lot of this kinetic information is not yet available. The second approach, *carbon flux analysis* is a combined mathematical/biological approach based on modified carbon molecules which are fed to the organism. Based on measuring where these molecules are located, predictions can be made about the fluxes within the model. The third model, *flux balance analysis*, will be discussed in more detail.

Flux balance analysis is a constraints-based approach. This model will not create a single prediction, but instead it will create a solution space based on stoichiometric information. The solution space can then be reduced by biochemical and environmental constraints. Based on an objective function, optimalization can be applied to find the flux distribution which optimizes this objective function.

Adding gene regulation to the model will further reduce the solution space and can even help predict outcomes under different conditions. This second generation of flux balance models can open new doors by adding more additional information about the organism and further reducing the solution space.

Table of contents

1 Introduction 4

2 Metabolic Networks 5
2.1 Metabolic Engineering 5
2.2 Metabolic reconstruction 5
2.3 Models 6
2.3.1 Flux Balance Analysis 6
2.3.2 Carbon Flux Analysis 7
2.3.3 Metabolic Control Analysis 9
2.4 Validation 10

3 Flux Balance Analysis 11
3.1 Model construction 11
3.1.1 Mass balance 12
3.1.2 Steady State 13
3.1.3 Elementary modes & Extreme pathways 15
3.2 Constraints 17
3.2.1 Energy balance analysis 18
3.3 Geometrical representation 19
3.4 Optimization 20
3.4.1 Linear Programming 20
3.4.2 Multi Integer Linear Programming 20

4 Gene regulated Flux Analysis 21
4.1 Regulation of gene expression 21
4.2 Reduction of the steady-state solution space 24

5 Discussion & Conclusion 27

Bibliography 28
Appendix A 30
A1. Linear Programming in Matlab 30
A2. Singular Value Decomposition (SVD) ..... **Error! Bookmark not defined.**

1 Introduction

Motivation

Cells are using their metabolism to transform substrates into a wide range of products through chemical reactions. These products will be used for growth, maintenance and proliferation of the cell. Enzymes with catalytic properties are the main players in this process and are responsible for speeding up the chemical reactions. To make an overview of the chemical reactions used in metabolism, metabolic networks are used to display the paths which substrates need to follow to form new products.

Industrial companies are exploiting the metabolism of microorganisms to create products for their own needs. The substrates used by the cells metabolism are only partly used for biomass and product synthesis, the remainder is being used for energy supply. One goal in metabolic engineering is to optimize metabolic networks to reach a maximum product synthesis and/or minimal energy use.
To reach this goal, quantification of the metabolic network is necessary.

To quantify metabolic networks means to be able to measure the substrate-fluxes in the metabolic network. Metabolic networks contain pathways which are built up from chemical reactions. In many cases the route and direction of the pathways are known, the quantity of the flux trough the pathways however is far more difficult to determine.

Overview

Different techniques have been developed to get a better insight of metabolic fluxes. This paper will describe known techniques used in metabolic engineering with one in particular, namely flux balance analysis (FBA). The basic idea behind this approach is not look for the precise flux solution, but to determine the limitations of a metabolic system. This will form a solution space with all possible flux solutions and can be used to create new hypotheses for further research.

The basic principles of metabolic engineering are given in section 2. In this section the models will be discussed that are able to determine the fluxes of a metabolic network. Because this paper will be based on the flux balance model, section 3 will discuss this model in more detail such as the constraints and optimization methods used for this model. Section 4 will be used to discuss gene regulation as an extra constraint in a flux balance model.

The remaining open questions about these modeling approaches will be discussed in the final section. Finally, some conclusions drawn during this research study about flux balance analysis will be presented.

2 Metabolic Networks

Metabolism is the biochemical modification of chemical elements in living organisms and cells. This process can be divided into two categories: anabolism and catabolism. Anabolism is the process where complex molecules are formed by smaller molecules, whereas catabolism is the breakdown of complex molecules. To sustain life and therefore build molecules, energy is needed from the breakdown of nutrients such as glucose and fatty acids. Metabolism is therefore a precise balance between the construction and breakdown of molecules needed for growth, maintenance and the proliferation of cells.

Cells are using their metabolism in principle for their own needs, but the metabolism of micro-organisms can also be exploited for other purposes. A well known example is the use of baker's yeast as leaving agent in bread or alcohol production with sugar as substrate. The bakers yeast's metabolism uses a substrate carbohydrate to produce the by-product carbon dioxide, which causes the dough to rise. This is one of the many examples where metabolism is being used to convert one product into another. Figure 1 shows a graphical representation of the yeast fermentation of alcohol with sugar as substrate.

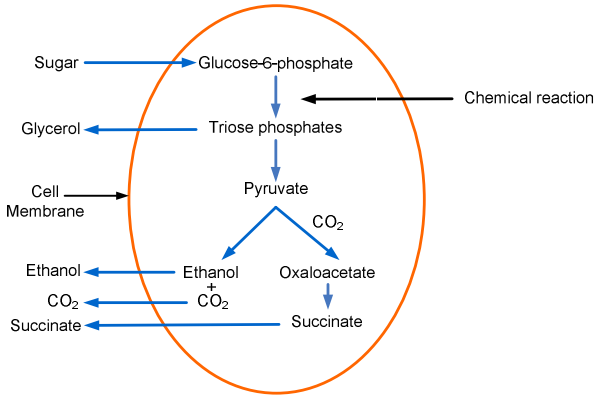

Figure 1 - Yeast fermentation

2.1 Metabolic Engineering

Metabolic engineering is an area of research pursued by scientific, pharmaceutical and industrial institutes. Each of these is trying to get a better insight into the metabolism of organisms. Better insight into the human metabolism, for example, gives pharmaceutical companies opportunities in drug development. Industrial companies are especially interested in finding an optimal metabolism for micro-organisms which will lead, for example, to less substrate uptake or higher biomass production of the micro-organisms.

The metabolism of organisms is often represented as networks of chemical reactions which modify molecules in the cell, as shown in Figure 2. Feasible solutions of such networks are reached with a certain flux distribution which describes the rate of each flux within the network. Optimal results or optimal pathways are routes through the metabolic network which optimize a certain objective as higher biomass for example. Genetic engineering, specific environmental conditions and the proper substrates are tools to find and utilize these optimal pathways.

2.2 Metabolic reconstruction

Metabolic networks are being reconstructed with the use of the sequenced genome of the organism. Additional information such as annotated sequences, biochemical information, strain-specific information is often used, for example, to reconstruct the metabolic network *Escherichia coli (E. coli)* MG1655, as showed in [4].

.

2.3 Models

The metabolic network of an organism describes which products are constructed from which substrates. However, the genomic data used to construct the metabolic network is not sufficient to describe the flux distributions, in other words the rate at which the products are made. Feasible flux distributions of a metabolic network are often represented by a vector which describes the rate of all fluxes within the network. Figure 2 shows a simplified representation of a metabolic network. Boxes **A**, **B** and **C** represent the substrate concentrations. Fluxes b_1, b_2 and b_3 represent the exchange substrate-fluxes which transfer substrate/products in and out of the cell via the cell membrane. The internal substrate-fluxes are represented by v_1, v_2 and v_3.

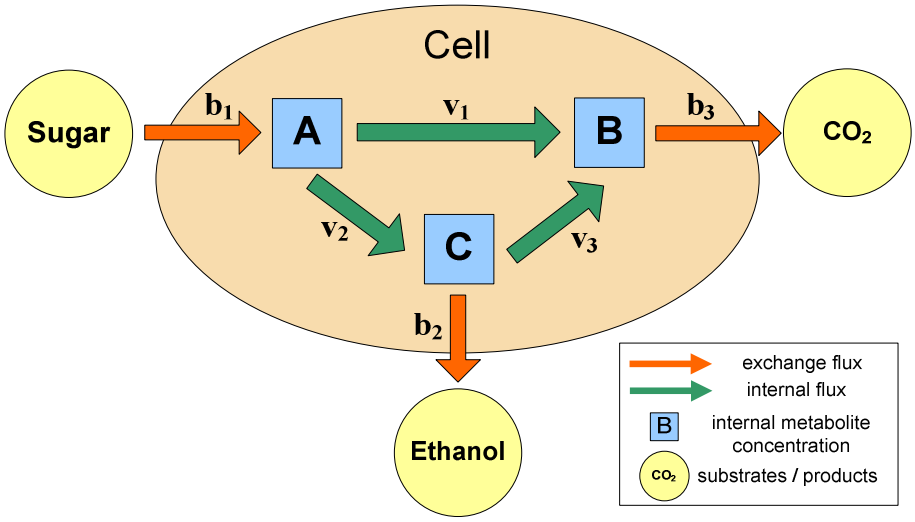

Figure 2 – Global representation of a metabolic network

Exchange fluxes are relative easy to determine. They represent for example oxygen and sugar uptake from outside the cell or secretion of ethanol and carbon dioxide. The hardest problem is to determine the internal fluxes. These are hidden inside the cell and are hard to measure without interrupting the cell's biological process. Predicting and interpreting metabolic flux distributions through a metabolic network requires the use of mathematical modeling and computational simulation.

Most of the existing methods requires detailed kinetic and concentration information about enzymes and various cofactors that influence the enzymes activity. However, even though the biological information is growing rapidly, there is still not enough information to describe the complete cellular metabolism in such mathematical detail for a single cell. Various models have been developed to overcome this shortage of information. Three of them will be discussed in this paper, with the emphasis on flux balance analysis.

2.3.1 Flux Balance Analysis

Flux balance analysis (FBA) is a modeling approach based on the constraints of a metabolic network. The main source of constraints is the stoichiometric matrix which describes precisely which substrates at which quantity are necessary to create a new product. Because this literature study puts the emphasis on this subject, chapter 3 will be used to discuss this modeling approach in more detail.

2.3.2 Carbon Flux Analysis

Carbon Flux Analysis is based on a combination of biological and mathematical methods to analyze metabolic networks ([12],[13]). This approach has been developed due the shortcomings of stoichiometry-based approaches, which will be discussed in detail in chapter 3. Stoichiometric-based approaches are not able to resolve bidirectional steps and cannot observe cyclic fluxes. This shortcoming can easily be seen in the following figures. When the fluxes leaving branch point A in Figure 3 are unknown and cannot to be measured, the fluxes through the separate metabolic routes cannot to be determined. A similar example is shown in Figure 4 where a circular metabolic pathway is shown. When the branch point fluxes between A-B and A-C are unknown, it will be impossible to determine the split fluxes with only stoichiometry information. Bidirectional reaction steps as shown in Figure 5 are very common in the cell. Stoichiometry-based approaches are unable to determine the two fluxes and will only be able to determine the *net flux*, which is the sum of the two fluxes.

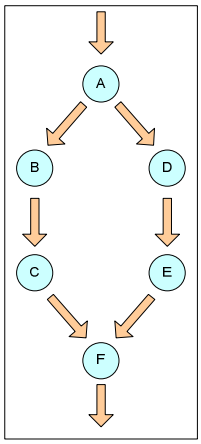

Figure 3 - Parallel pathway

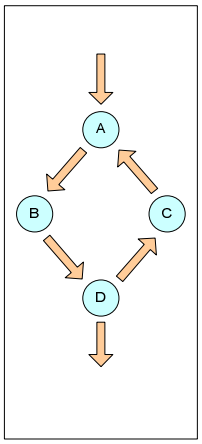

Figure 4 - Circular pathway

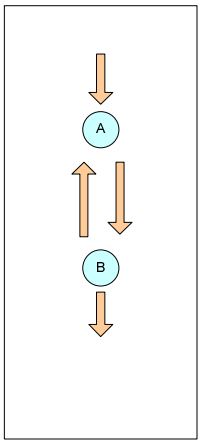

Figure 5 - Bidirectional pathway

To overcome this shortcoming, it became clear that additional information is required to complement the extracellular flux data. A carbon labeling experiment (CLE) has been carried out by feeding a specifically ^{13}C-labeled substrate in a metabolic stationary state. This modified carbon molecule had been modified and can be distinguished from the other carbon molecules in the metabolic network.

^{13}C-labeled substrates are fed to the cell. The assumption is made that enzymes within the cell cannot sense the difference between the normal and the modified carbon molecules. The labeled carbon atoms are then distributed all over the metabolic network. The isotopic enrichment in the intracellular metabolite pools can then be measured by mass spectrometry (MS) and nuclear magnetic resonance (NMR) instruments. The resulting data measured from the system provides a large amount of additional information to quantitate the intracellular fluxes.

Algorithm

The algorithm described by Wiechert in [12], is widely used to evaluate a carbon labeling experiment and proceeds as follows:

1. Guess a flux distribution over the metabolic network which fulfills the stoichiometric balance equations. The sum of the outgoing fluxes of a branch point must be equal to the incoming flux.
2. Simulate a carbon labeling experiment on this flux distribution and the known labeling state of the input substrate.
3. Compute from the outgoing isotopomer distribution the measured values that would come out if the guessed fluxes were present in system
4. Compute the difference between the measurements predicted by the simulation and these measurements which were actually obtained. The discrepancy is usually measured by a sum of squares where each single residual value is weighted by the corresponding measurement standard deviation.

Based on the computed discrepancy a systematic variation of the guessed fluxes is performed by applying an optimization algorithm. This implies an iteration of the steps 2-4.

The evaluation of a CLE is based on a rather complicated mathematical model that describes how the labeled material distributes over the metabolic network. The centerpiece of data evaluation is a simulation of the CLE. In this simulation it is assumed that the intracellular fluxes are already known. Based on guessed flux values and the known input substrate composition the stationary distribution of 13C labeled molecules over the network can be computed.

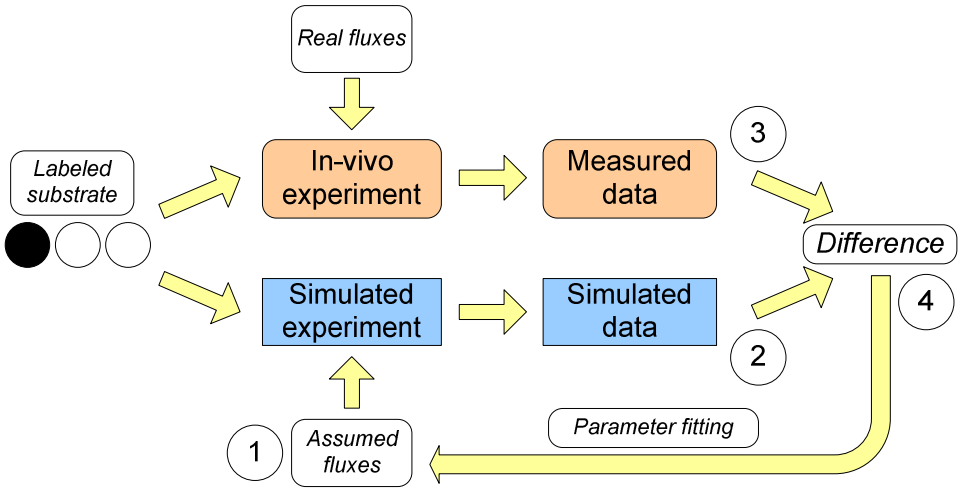

Figure 6 – graphical representation of Wiechert's algorithm

2.3.3 Metabolic Control Analysis

Metabolic control analysis is based on the analysis of the regulatory structures in cells that lead to certain fluxes in a metabolic network. As described in [13], enzymes are in this case the major players responsible for the fluxes. The benefit of this approach is that this model will not only describe the internal fluxes, it will also explain why these fluxes occur. Analysis methods using only the stoichiometric matrix can only study the interactions between different pathways in the metabolic network, but they are not able to determine how the fluxes are controlled.

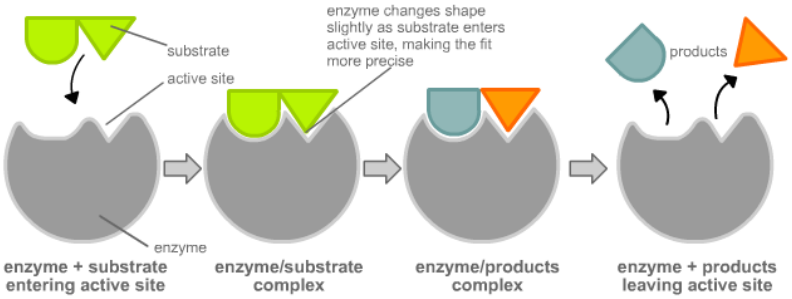

Figure 7– product catabolism performed by enzymes (***source: www.wikipedia.org***)

In this modeling approach, a metabolic network is formed by single enzyme-catalyzed reaction steps. These reaction steps are responsible for modifying substrates in the cell and creating new products, as illustrated in Figure 7. Mathematical functions are used to describe these reaction steps by modeling them based on the kinetic properties of the enzymes. Complex reaction steps will easily have ten or more kinetic parameters like inhibitors and other co-factors that have influence on the enzymes. This makes it difficult to define the kinetic formulas, because a lot of data is required to estimate these parameters.

Data

Published data is often measured in different organisms or different strains, under different conditions. The estimated kinetic parameters can therefore differ in magnitude which makes them hard to use in the modeling process. The active enzyme concentrations are stored in a vector $\mathbf{e}$ and correspond with the rate of product formation of the enzyme. This is extremely difficult to measure because it depends on the expression level of the enzyme. Micro-arrays can be used to measure the expression level of the DNA, but the stability of RNA and translation factors can influence the final enzyme activities. The vector $\mathbf{e}$ can also be estimated with experiments that measure the concentration of the enzyme from cell extracts, but results can be influenced by the cell disruption due to the measurement.

When kinetic parameters, enzyme concentrations, stoichiometric information and substrate concentrations are combined, a model can be created as shown in (1). In stationary state, the system is studied under the condition that the concentrations of the metabolites are constant.

$$\mathbf{N} \cdot \mathbf{v}(\boldsymbol{\alpha}, \mathbf{s}, \mathbf{e}; \mathbf{x}) = 0 \qquad \textbf{(1)}$$

Where $\mathbf{N}$ stand for the stoichiometric matrix, $\boldsymbol{\alpha}$ denotes the vector of all enzyme kinetic parameters, $\mathbf{s}$ is the vector of extracellular substrate concentrations, $\mathbf{e}$ is the vector of enzyme concentrations and $\mathbf{x}$ denotes the vector of intracellular metabolite concentrations.

2.4 Validation

The hardest part of the modeling process is testing the validity of the model. In the common definition of validity for models, it needs to predict the outcomes of all experiments in the development scope. This means that the model needs to resist all falsification attempts. The irony behinds this definition is that model validity can never be proven because there are always an unlimited number of experiments to perform on the model.

Statistical procedures are used for testing validity hypotheses. The tested model is assumed to be valid and fitted on the experimental data. Although it is hard to ensure the validity of the outcomes of mathematical model build up from biological information, it can provide new information for new experiments.

3 Flux Balance Analysis

Most mathematical models built for modeling metabolic networks are based on calculating a single solution which describes all fluxes in a metabolic network. Flux balance analysis (FBA) is based on the idea to look at the limitations of a metabolic network. This approach narrows the range of possible states that a metabolic network can adopt based on different constraints which cells must obey.

This constraint-based approach provides a basis for understanding the structure and function of biochemical reaction networks. Additionally, each of these constraints can be described mathematically, offering a geometric interpretation of the effect that each successive constraint places on the metabolic function. Each metabolic flux of the network constitutes a dimension of the solution space and can therefore be represented by an axis in a graph. The graph combined with constraints of the cell will produce a hyperplane which represents a solution space for a specific metabolic network with all possible states that the network can take.

3.1 Model construction

Flux balance models require the definition of all the metabolic reactions and metabolites used in the model. As described in section 2.2, the genome sequence is a good starting point. The product of gene is annotated by homology searches as to identify all the metabolic enzymes. These enzymes can be used to describe all chemical reactions they catalyze. The result of this research is given in Figure 8. As can be seen in the figure, flux 3 and 4 form a reversible chemical reaction. This means that a reactant (**B** or **D**) can be transformed back and forward into its original chemical structure. In many papers, this reaction is split into two separated fluxes, which results in consequently positive fluxes instead a mixture of negative and positive fluxes.

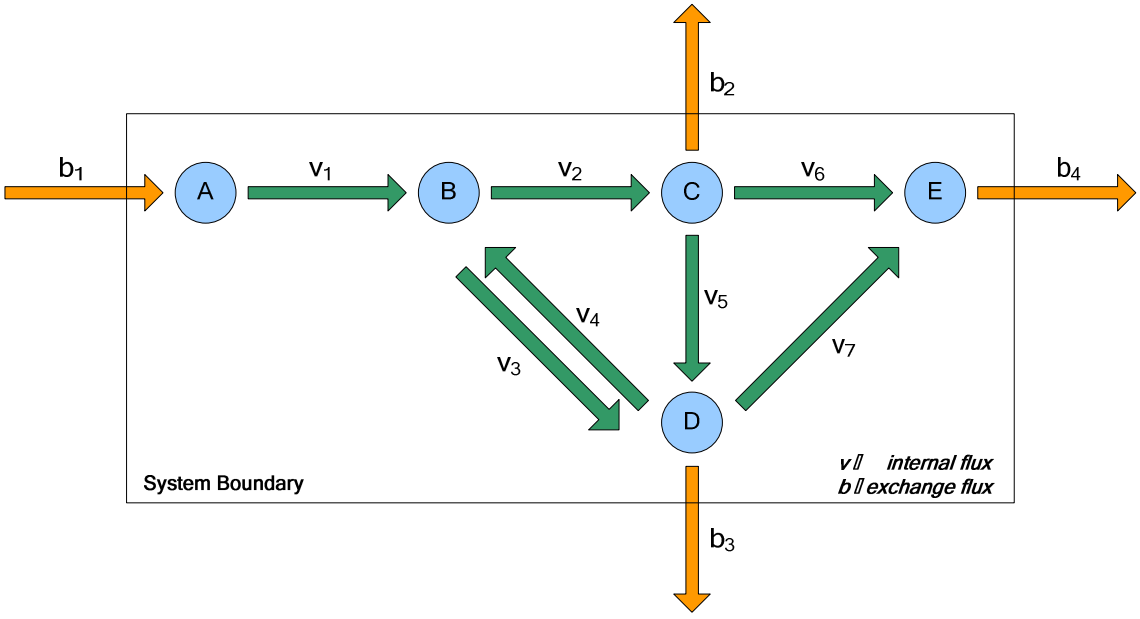

Figure 8 – Chemical reaction scheme

3.1.1 Mass balance

The steady-state of metabolic networks can be described in terms of mass balance equations. These equations describe the change of concentration over time based on the fluxes through the metabolic network. The change of concentration is the difference between the rates at which the metabolite is produced and consumed. An example of mass balance equations extracted from Figure 8 is given by (2). It can be seen that the sum of all variables in the equations equals the external fluxes and fulfills the definition of the conservation of mass principle (no mass is lost during the experiment).

$$\begin{aligned}
\frac{dA}{dt} &= -v_1 + b_1 \\
\frac{dB}{dt} &= v_1 + v_4 - v_2 - v_3 \\
\frac{dC}{dt} &= v_2 - v_5 - v_6 - b_2 \\
\frac{dD}{dt} &= v_3 + v_5 - v_4 - v_7 - b_3 \\
\frac{dE}{dt} &= v_6 + v_7 - b_4
\end{aligned} \tag{2}$$

The mass balance equations can be described by a stoichiometric matrix $\mathbf{S}$ and a flux vector $\mathbf{v}$. A stoichiometric matrix $\mathbf{S}$ is a $m \times n$ matrix where **m** corresponds to the number of metabolites and **n** is the number of chemical reactions or fluxes taking place within the metabolic network. The matrix describes the relationship between the metabolites and the products. It describes which metabolites at which quantity are necessary to create a specific product. The vector $\mathbf{v}$ corresponds to the quantity of the fluxes within the metabolic network. It describes the rate at which a metabolite is chemically modified into another metabolite. The relationship between the concentrations of the metabolites and the stoichiometric matrix is shown in (3).

$$\begin{bmatrix} \frac{dA}{dt} \\ \vdots \\ \frac{dE}{dt} \end{bmatrix} = \mathbf{S} \cdot \mathbf{v} \tag{3}$$

If (2) is represented in terms of a stoichiometric matrix $\mathbf{S}$ and a flux vector $\mathbf{v}$ as in (3), equation (4) is derived.

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \\ \frac{dD}{dt} \\ \frac{dE}{dt} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 & -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \mathbf{v_1} \\ \vdots \\ \mathbf{v_7} \\ \mathbf{b_1} \\ \vdots \\ \mathbf{b_4} \end{bmatrix} \tag{4}$$

It must be noted that other values than zeros and ones can be used within stoichiometric matrices. Stoichiometric matrices are used to describe balanced chemical reactions; more reactions can be necessary to create one single product. A well known example is the formation of liquid water from the diatomic gasses hydrogen and oxygen. Using only one hydrogen molecule and one oxygen molecule, as shown in (5), will not create a balanced chemical reaction.

$$H_2 + O_2 \rightarrow H_2O + energy \quad \textbf{(5)}$$

Therefore two molecules of hydrogen are used to create 2 molecules of liquid water. This example is shown in (6) and forms a balanced chemical reaction. In a stoichiometric matrix, the chemical reaction of liquid water will be represented as shown in (7).

$$2H_2 + O_2 \rightarrow 2H_2O \quad \textbf{(6)}$$

	H_2	O_2	H_2O
Reaction A	**-2**	**-1**	**2**

(7)

3.1.2 Steady State

The most important assumption made FBA is that the model will be in steady state. This means that all concentrations of the metabolites in the model are constant, i.e. change over time will be zero. This assumption is made in FBA because the metabolic transients are more rapid than both cellular growth rates and the dynamic changes in the organism's environment [11]. Metabolism typically has transients that are extremely fast which results in a steady state for the complete system in matter of seconds. Changes in the metabolism made by gene regulation will become noticeable only after minutes or even hours. This assumption is also known as the quasi steady-state approximation. With this approximation, we assume that the change of concentrations over time will be approximately zero and can rewrite (3) into (8). This opens new opportunities to calculate the internal fluxes.

$$\mathbf{S} \cdot \mathbf{v} = 0 \quad \textbf{(8)}$$

The mass balance equations presented in (2) are linear equations. Rewritten (8), the null space of $\mathbf{S}$ can be obtained. The null space is a set of all operands $\mathbf{v}$ which solves (8). It therefore contains all possible solutions for the metabolic network under the steady-state condition. The dimension of the null space will be determined by the number of free variables in the original set of linear equations which is referred as the rank of the stoichiometric matrix. The Rank theorem, (9), can then be used to determine the dimension of the null space.

$$\dim Null\ (\mathbf{S}) + rank\ (\mathbf{S}) = n \quad \textbf{(9)}$$

The null space itself can be determined with the singular value decomposition method. With this method, a set of basis vectors is derived which are linear independent of each other. The minimal number of vectors needed to span the null space is equal to its dimension. The stoichiometric matrix given in (4) has a rank of 5, which gives a null space with a dimension of 6. This means that 6 basis vectors $\beta = \{b_1, \ldots, b_6\}$ in $\mathbb{R}^{11}$ are needed to span the null space. The basis vectors derived with the singular value decomposition are theoretically feasible to describe all possible solutions within the null space based on (8). However, a theoretical description is not enough for describing solutions for a metabolic network. The basis vectors must also be biochemically feasible, which means that the solution they represent must not conflicting with other defined constraints as shown in (11) and (12), [10].

The basis vectors which span the null space of (8) are shown in (10). The first seven entries for each vector correspond to the internal fluxes ($\boldsymbol{v_i}$), whereas the remaining entries corresponds to the exchange fluxes ($\boldsymbol{b_i}$). Taking a closer look at the 4th and 5th basis vector reveals negative values for internal fluxes (orange squares). In biochemical terms, negative values for chemical reaction rates are impossible and therefore vectors with negative values have no biochemically meaningful information.

$$\beta = \{basis_1, \ldots, basis_6\} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \leftarrow \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \quad \textbf{(10)}$$

Because the six basis vectors are linearly independent, they can be used to create another set of six vectors which span the same null space. To create a set of basis vectors which is biochemical feasible, two constraints are imposed on the internal and external fluxes. Equation (11) ensures that each internal flux is positive and (12) imposes further restrictions on the exchange fluxes. If a reactant is known to enter the system and does not exit, then a constraint can be applied to that particular flux to be negative.

$$\mathbf{v_i} \geq 0 \quad \textbf{(11)}$$

$$\begin{aligned} \mathbf{b_i} &\geq 0 \ \text{(exit only)} \\ \mathbf{b_i} &\leq 0 \ \text{(enter only)} \end{aligned} \quad \textbf{(12)}$$

Under the above constraints, the basis vectors $\mathbf{b_i}$ can be rewritten into a new set of vectors $\mathbf{p_i}$. An algorithm described in [9] can be used to find this new set of vectors which obey above constraints.)

$$\mathbf{P} = \begin{Bmatrix} \mathbf{p_1} = basis_1 \\ \mathbf{p_2} = basis_2 \\ \mathbf{p_3} = basis_3 \\ \mathbf{p_4} = basis_4 + basis_1 \\ \mathbf{p_5} = basis_5 + basis_2 \\ \mathbf{p_6} = basis_5 + basis_6 \end{Bmatrix}^T = \begin{matrix} \mathbf{p_1} & \mathbf{p_2} & \mathbf{p_3} & \mathbf{p_4} & \mathbf{p_5} & \mathbf{p_6} \end{matrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \leftarrow \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \quad \textbf{(13)}$$

The transformation shown in (13) results in a new set of basis vectors. The new set of vectors $\mathbf{p_i}$ is a one-to-one mapping with the original basis vectors which spans the null space. This means that every flux distribution, which is a solution of the null space, can be written as a linear combination of the pathways traced by the basis vectors in $\mathbf{p_i}$. Many papers describe the pathways given by set $\mathbf{p_i}$ as *Extreme Pathways* because of their properties. These properties are explained in section 3.1.3. The extreme pathways of the metabolic network used in the example are shown in Figure 9.

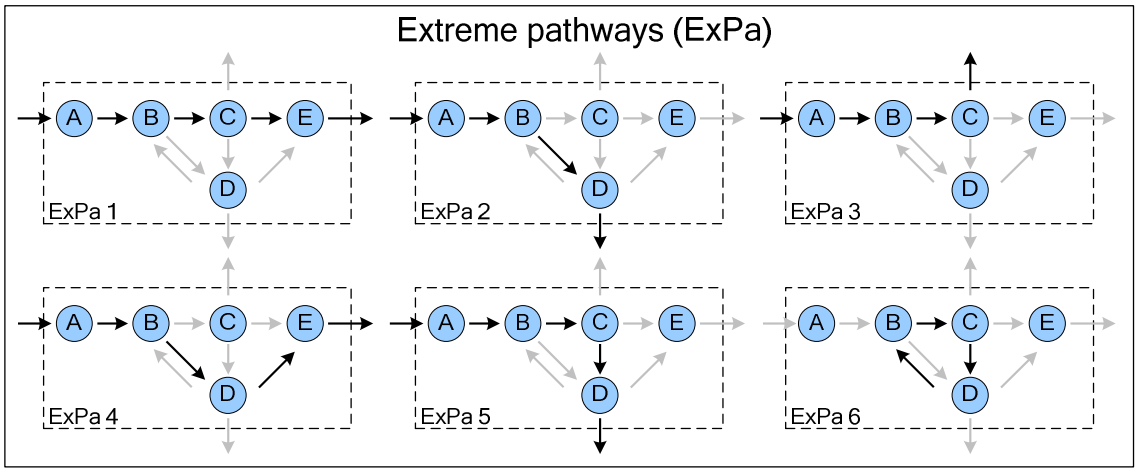

Figure 9 – Extreme pathways formed by the basis vectors as shown in (13)

In mathematical terms, every solution of the null space can written as the sum of all extreme pathways $\boldsymbol{EP_i}$ combined with a weight factor $\boldsymbol{w_i}$ for each vector as shown in (14).

$$C = \{v \mid v = \sum_{i=1}^{k} w_i EP_{i,}\ w_i \geq 0\ \forall i\} \qquad \textbf{(14)}$$

3.1.3 Elementary modes & Extreme pathways

As shown in the previous section, the null space of the stoichiometric matrix can be written in terms of extreme pathways through a metabolic network. Elementary flux modes can also be used to represent the null space [7]. Although these are close related to extreme pathways, there are differences between them. An overview of these is given in Table 1 and Figure 10.

Table 1 – comparison between elementary and extreme pathways

	Elementary flux modes	Extreme pathways
1	There is a unique set of elementary modes for a given metabolic network.	There is a unique set of extreme pathways for a given metabolic network.
2	Each elementary mode consists of the minimum number of reactions that it needs to exist as a functional unit. If any reaction in an elementary mode were removed, the whole elementary could not operate as a functional unit.	Each extreme pathway consists of the minimum number of reactions that it needs to exist as a functional unit.
3	The elementary modes are the set of all routes through a metabolic network consistent with property 2.	The extreme pathways are the systemically independent subset of elementary modes; that is, no extreme pathways can be represented as a nonnegative linear combination of any other extreme pathway.

The table with properties reveals that the two types of pathways share two properties. Both must be unique and must consist of the minimum number of reactions to become a functional pathway in the metabolic network. This means that every extreme pathway or elementary mode is the smallest sub network of the metabolic network that functions in steady-state. The difference between the two types becomes clear in Figure 10, where all pathways created from a single reaction network are the same, except for extreme pathway 4, which is not an extreme pathway.

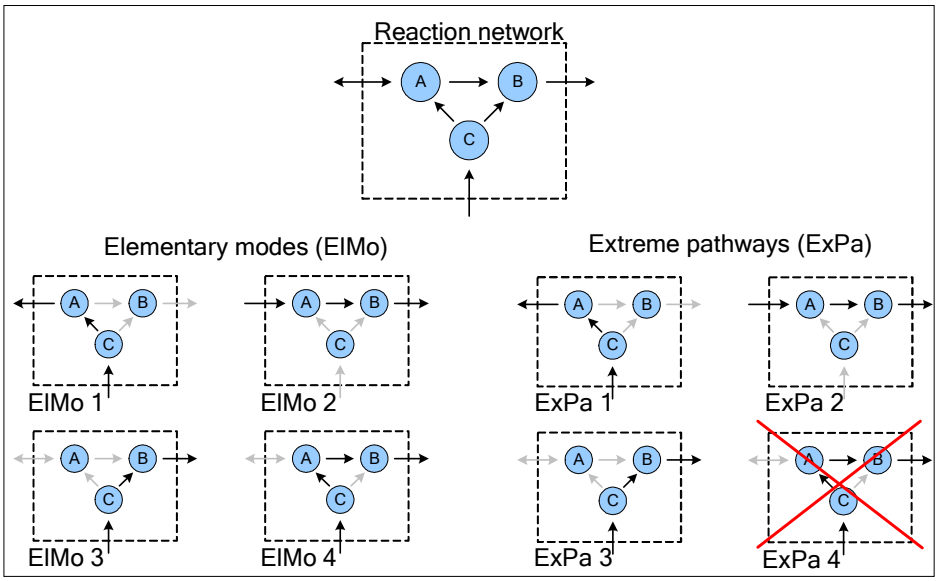

Figure 10 - Comparison Elementary modes and exptreme pathways of a metabolic network

The third property of extreme pathways explains that each pathway cannot be a linear combination of other pathways. Taking a closer look at Figure 10 reveals that ExPa4 is a linear combination of ExP1 and ExP2. Remember that the exchange fluxes can be positive and negative as shown in (12), this in contrast with the internal fluxes which are constrained to be positive as explained in section 3.1.2. This means that a linear combination of ExPa1 and ExPa2 can exclude the exchange flux (red eclipse in Figure 11) connected with metabolite **A**. This exchange flux can be excluded, because $-\mathbf{b}_1 + \mathbf{b}_1 = 0$. The combination of ExPa1 and ExPa2 will be similar to ExPa4, which is a contradiction with property 3 of the extreme pathways. In this way, elementary modes are always a superset of the extreme pathways.

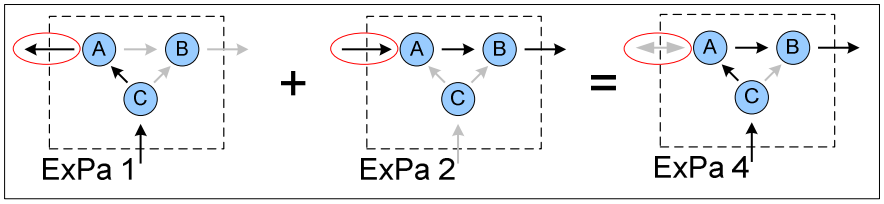

Figure 11 – Linear combination of extreme pathways

3.2 Constraints

Once a system has been defined in terms of mass balance equations, the constraints imposed by the enzyme or transporter capacities (maximum uptake or reaction rates for example) can be considered and incorporated into the model. Cells have to deal with many constraints during their lifetime. FBA will use these constraints to determine the limitations of the cells under different environmental conditions. Different kind of constraints can be used to limit possible solutions. An overview of constraints used by flux balance analysis is given [8].

- **Physico-chemical** - Physico-chemical constraints are the so called 'hard' constraints on cell functions. These constraints will not change due to environmental conditions. Examples of these hard constraints are mass, energy and momentum which is conserved in the cell, which means that during the experiment the total quantity of these variables, will be equal to there initial value. The conservation of mass is described by the mass balance equations and will form the most important constraints for this model. The same principle accounts for energy and is reviewed in a review of energy balance analysis [1].

  The chemical reaction rates are determined by the kinetic properties of enzymes and available substrate concentration. Enzymes can perform several million catalytic reactions every second. The maximum capacity of an enzyme in terms of velocity is defined by V_{max}. In this state, all active sites on the enzyme are occupied by substrates. The amount of substrate is therefore important to determine the velocity of an enzyme. An important kinetic parameter used for enzymes is the Michealis-Menten constant (K_m) which represents the amount of substrates required an enzyme to reach half its maximum velocity. The Enzyme turn-over parameter is the number of substrate molecules converted into products per unit time when the enzyme is fully saturated with the substrate. This number is generally less then 10^4 sec^{-1}. So both the substrate concentration and kinetic parameters will determine the chemical reaction rate.

- **Topobiological** - The crowding of molecules in cells leads to topobiological or three-dimensional constraints. An example is the DNA tightly packed within the nucleus because DNA stretched will be a 1000 times larger then the size of the cell. At the same time DNA has to be accessed for transcription in high quantities and fast. Another example is the ratio between the available amount of tRNA molecules and ribosomes, which are respectively the building blocks and factories for proteins.

- **Environmental** - Environmental conditions in cells are time and condition dependent. Nutrient concentrations, PH value and temperature are examples of environmental constraints. It is fundamental that with each experiment the used media and environmental conditions are well reported when they are used in models. Data used from different laboratories without the proper conditions written down are often limited used for *in silico* modeling.

Although these are useful constraints, it is not immediately clear how they can be implemented in mathematical form. Especially the topobiological constraints seem hard to implement in a flux balance model. Environmental constraints are more obvious because quantities as temperature, PH-value and nutrient conditions will have an influence on the chemical reaction rate, which in turn has an influence in the flux balance model. Constraints most discussed in papers are the mass balance constraints and energy balance constraints. These are easy to implement in FBA in terms of matrixes and vectors.

3.2.1 Energy balance analysis

Although the mass balance equations form the main constraints used in flux balance analysis, other constraints can help to further tight the solution space as explained in section 3.1. A detailed example is given by [1], which describes the use of an energy balance model in collaboration with flux balance analysis. Energy balance analysis is based on the theory and methodology for enforcing the laws of thermodynamics which is applied in metabolic flux analysis.

The chemical potential of an element is the increase of energy when one unit of this element is added to a system. According to the second law of thermodynamics, fluxes must flow from reactants of higher chemical potential to ones of lower chemical potential. Many of the feasible fluxes, determined with FBA, violate the second law of thermodynamics and are therefore infeasible.

[1] describes that the energy-balance equations are obtained from the null-space matrix obtained from an alternative matrix $\mathbf{S}'$ derived from the stoichiometric matrix. Matrix $\mathbf{S}'$ is formed by removing the columns that interact with the exchange fluxes. Matrix $\mathbf{S}'$ will then represent the set of internal loops of chemical reactions. The null-space matrix from matrix $\mathbf{S}'$ is derived with singular value decomposition (15). The diagonal of Matrix $\mathbf{\Lambda}$ (16) stores the singular values $\boldsymbol{\lambda_i}$. The remaining columns are all equal to zero. If $\mathbf{n_c}$ is the number of chemical reactions and $\mathbf{r}$ is the rank of Matrix $\mathbf{S}'$, then matrix $\mathbf{B}$ will provide a $(\mathbf{n_c} - \mathbf{r})$-dimensional null space of $\mathbf{S}'$. This null space is represented by matrix $\mathbf{K}$ as shown in (17).

$$\mathbf{S}' = \mathbf{A}\mathbf{\Lambda}\mathbf{B}^{\mathbf{T}} \tag{15}$$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_m & 0 & \dots & 0 \end{bmatrix} \tag{16}$$

$$\mathbf{K} = \begin{bmatrix} \mathbf{b}_{1,r+1} & \dots & \mathbf{b}_{n_c,r+1} \\ \vdots & & \vdots \\ \mathbf{b}_{1,n_c} & \dots & \mathbf{b}_{n_c,n_c} \end{bmatrix} = [\mathbf{k}_1, \dots, \mathbf{k}_{(\mathbf{n_c}-\mathbf{r})}] \tag{17}$$

Associated with each internal flux $\mathbf{j_i}$ is a chemical potential difference $\mathbf{\Delta\mu}$. These potential differences must satisfy a law similar to the Kirchhoff's loop law in electrical circuits shown in (18). Vector $\mathbf{\Delta\mu}$ lists the number of $\mathbf{n_c}$ chemical potential differences associated with the reaction fluxes. Because each vector $\mathbf{k_i}$ in matrix $\mathbf{K}$ provides weights for exactly balancing the chemical reaction equations, solutions to (18) balance the global potential energy of the network. An example of this balance is shown in [1].

$$\mathbf{K}\Delta\boldsymbol{\mu} = 0 \tag{18}$$

The second law of thermodynamics takes the form of an inequality constraint for each flux as shown in (19) which ensures that entropy[1] production is positive for each reaction. Equations (18) and (19) represent thermodynamic constraints that should be considered in addition to the flux balance constraints. Equation (19) provides a link between the mass balance equations and energy balance equations.

$$\mathbf{J_i} \cdot \Delta\boldsymbol{\mu} < 0 \tag{19}$$

1. specifies the amount of disorder or randomness in a system bearing energy or information.

3.3 Geometrical representation

A great benefit of the flux balance approach is the capability of the visualization of the solutions. The flux balance approach works with vectors and constraints which can be visualized in a 3D-graph as shown in Figure . The graph shows a cone generated from the extreme pathways of the previous example. Each edge of the cone represents one extreme pathway of the metabolic network. Because of the limitation, the graph can only show three fluxes of the extreme pathway at the same time.

The cone is a graphical representation of all solutions of the null space obtained from the stoichiometric matrix, which means that every point within the cone represent a flux distribution that obeys the steady-state assumption. Linear programming (section 3.4.1) can be used to find the best flux distribution based on a certain objective function. Figure shows only the basis vectors. Equation (14) shows that these can be multiplied by a weight factor. The cone shown in this graph is based on a weight factor at one for all vectors. Without the use of constraints on the fluxes, the size of the cone will be unlimited.

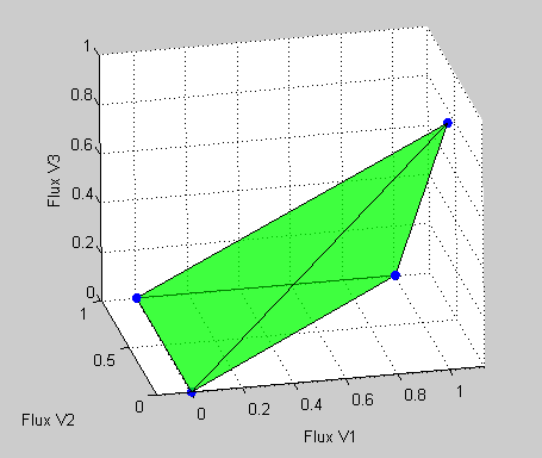

Figure 12 - 3D cone

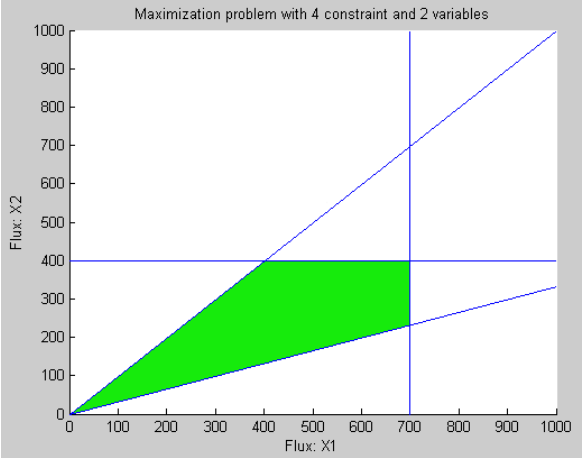

Figure 13 – Cone constrained by flux limitations

With these constraints a bounded solution space (green space in Figure) can be created wherein every possible metabolic flux distribution, or every possible metabolic phenotype of the cell, must lie. The solution of the described flux balance equations must lie in the closed solution space. Extra constraints in the form of experimental measurements and gene regulation can be added to reduce the size of the solution space. This will be discussed in more detail in chapter 4. It must be noted that these constraints do not have the same consistency as physiochemical constraints and only add information in certain conditions.

As noted before, FBA will not give an exact flux-distribution. It can be used to calculate and predict metabolic flux distributions and to analyze the capabilities of a metabolic network based on the stoichiometric, thermodynamic and reaction capacity constraints. Linear programming can be used to calculate the optimal flux-distribution based on the solution space created by the constraints. This optimal flux-distribution will be based on a predefined objective function. By calculating and examining optimal flux distributions under various conditions, it is possible to generate quantitative hypothesis *in silico* that may be tested experimentally.

3.4 Optimization

When the metabolic network is described in mathematical terms, optimization can be applied to determine an optimal flux distribution based on a certain objective function. In [11] it is argued that optimization of growth usually confers an advantage selected by evolution. The determined flux distribution can be used to describe experimental results and to predict how cells will respond on their environment.

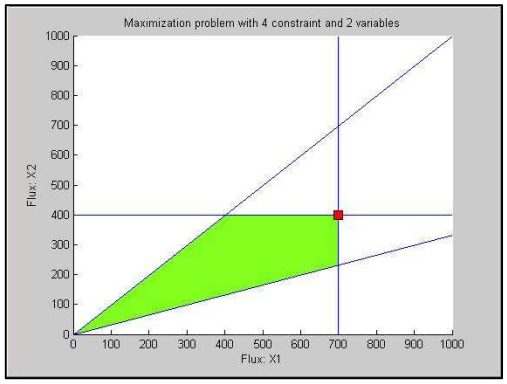

Figure 12 - optimal flux distribution base on an objective function

3.4.1 Linear Programming

Linear programming is used to solve algebraic equations subject to constraints. The objective is to optimize a certain function that is constrained by a number of inequality functions. A very common procedure used for linear programming is the simplex method. This was in fact the first method developed to solve linear programs. The first approach of the simplex method is to convert the linear programming problem into a system of linear equations. In the case of FBA, the mass balance equations and constraints are used. Appendix A shows in detail how the optimal values for algebraic systems are determined.

3.4.2 Multi Integer Linear Programming

Some algebraic systems with an objective function have more then one single optimal value. These values are called alternate optima. In [6], an algorithm called recursive mixed-integer linear programming (MILP) is proposed, which is based on the linear programming algorithm to find all those alternate optima.

4 Gene regulated Flux Analysis

In this chapter, two papers will be discussed which introduce regulation of gene expression within FBA. [2] describes how regulation can be used to create time series and discovers information about the network that would not be discovered without the use of gene regulation. The second paper, [3], also uses gene regulation to reduce the solution space created by the stoichiometric matrix.

4.1 Regulation of gene expression

An estimation of 660 metabolic genes are present in the genome of *E. coli*, [2]. These genes have some roll within the metabolism of the cell. It has been thought that 400 genes have a regulatory function. During research on the K-12 MG1655 genome, 178 genes were annotated as regulatory genes or putative regulatory genes.

The high level of transcriptional regulation in this and other organisms has a significant effect on the cell's behavior. Previous FBA models, developed before 2001, have not incorporated the regulation of gene expression and the activity of the expressed gene product, [2]. These models assumed that all gene products are available in the metabolic network to contribute to an optimal solution. The presence and activity of many metabolites strongly depends on the regulatory process in the cell and will form constraints in the metabolic network.

FBA-models of *E. coli* without gene regulatory incorporated, leads to certain incorrect predictions of cellular-level behavior [4]. To predict cell behavior on a general scale, regulatory constraint should be incorporated into the model. Unlike physico-chemical constraints as mass balance equations, regulatory constraints are self-imposed by the organism, and presumably represent the result of an optimal evolutionary process.

The transcriptional regulatory structure can be described by Boolean logic equations. The expressions of transcriptional units are then described with a value 1 when the unit is transcribed and value 0 if it is not. The same accounts for the presence of enzymes and certain conditions inside or outside the cell. Boolean logic equations are formed up of variables (transcriptional units, enzymes, etc) and operators as AND, OR and NOT, which can be used to build up the equations which describes the expression of transcription units.

An example of a simple regulation circuit is given in Figure 13 which contains one gene **G** which is transcribed by a process **E**. The enzyme which is then translated catalyses reaction **rxn** which converse substrate **A** to product **B**. Product **B** interacts with a binding site on gene **G** such that the transcription process trans is inhibited. So process **trans** will occur if gene **G** is present in the genome and product **B** is not present to bind to the DNA. In logical terms, this part of Figure 13 can be written as shown in (20).

$$\mathbf{trans} = \mathrm{IF}(\mathbf{G}) \wedge \neg(\mathbf{B}) \qquad \mathbf{20}$$

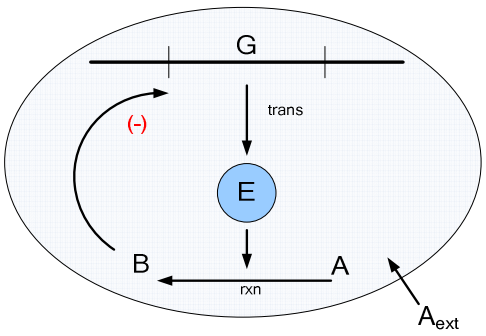

Figure 13 –regulatory circuit

Chemical reaction **rxn** can also be written in logical terms as shown in (21).

$$\mathbf{rxn} = \mathrm{IF}(\mathbf{A}) \wedge (\mathbf{E}) \quad \textbf{(21)}$$

The presence of enzymes or regulatory enzymes in the cell at a certain point in time depends both on the transcription of the enzyme in previous time and the rates of synthesis and decay of the enzyme. Under steady state conditions, the average protein synthesis and degradation times are equal.

If the presence of all regulated enzymes in the metabolic network is determined for a given time-interval ($\mathbf{t_1} \rightarrow \mathbf{t_2}$) and it is known that a certain enzyme is not present during that time-interval, then the flux through that will be catalyzed by this enzyme can be set to zero. This restriction maybe used as a temporary constraint on the metabolic network.

$$\mathbf{v_k}(\mathbf{t}) = 0, \quad \text{when } \mathbf{t_1} \le \mathbf{t} \le \mathbf{t_2} \quad \textbf{(22)}$$

If an enzyme is present during a given time-interval, the reaction which it catalyzed will be left unconstrained and will be determined with the use of FBA. Thus, known regulation of gene expression can be represented with the use of Boolean logic and incorporated within flux balance analysis.

In [2] it is shown that FBA alone would not be able to predict a certain diauxic shift within in the metabolism along a time line. The example shows the growth of the cell based on two carbon sources in the extracellular space. Both concentrations where set on an initial value of 10mM. Region **A** shown in the time line of Figure 16 shows that **carbon1** is preferred during cell growth, whereas in Region **C** **carbon2** is preferred. Region **B** shows that the cell growth stops when there is a shift between the two carbon sources. In this period, a transport process **Tc2** is being upregulated and synthesized and during this period, no carbon sources are being transported into the cell, which means that the cell growth during this period stops.

The regulation between the two carbon sources is shown in Figure 14 and Figure 15. The two bars below the two figures represent the gene regulation. It can be seen that in the second figure, **Tc2** becomes active and will be responsible for the transport of **carbon2** into the cell. Without the addition of these regulatory constraints, the system would use both carbon sources at the same time to maximize the production of biomass conditions and would fail to predict the diauxic shift in Figure 16. During the diauxic shift, the biomass stops growing for a period. More examples are given in the paper to show the importance of gene regulation within FBA.

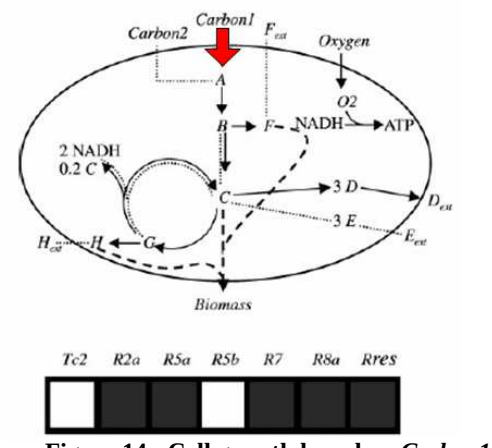

Figure 14 - Cell growth based on *Carbon1*

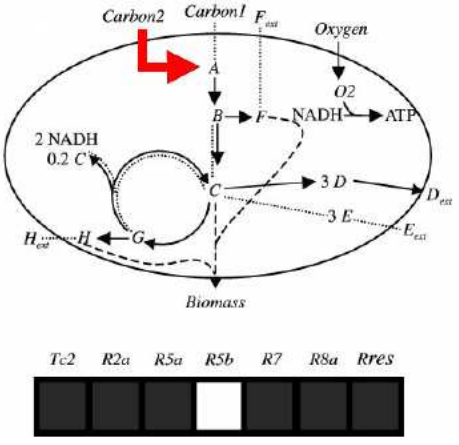

Figure 15 – Shift to C*arbon2*

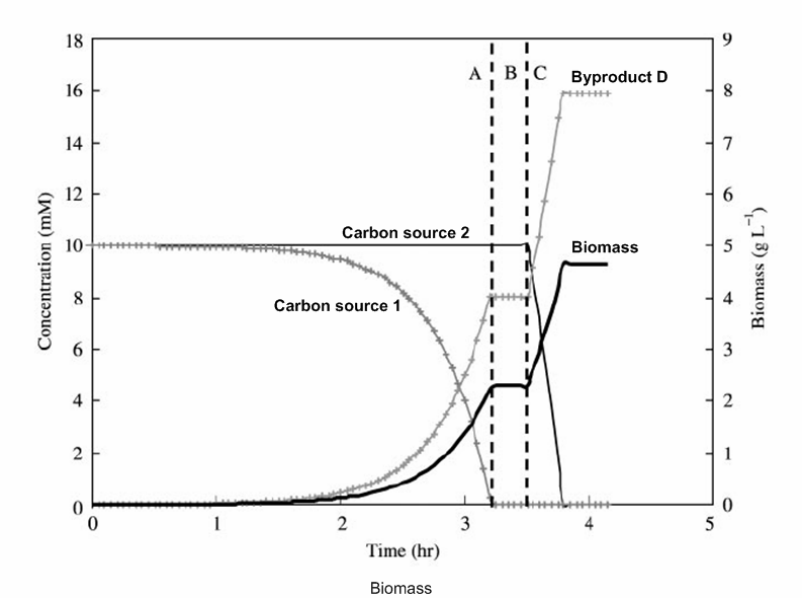

Figure 16 – A diauxic shift is predicted in time-interval B

4.2 Reduction of the steady-state solution space

The second paper of Covert *et al* [3], shows how gene regulation can be used to reduce the solution space within FBA. As explained in [2], the first generation of models based on flux balance analysis did not incorporated regulation of gene expression. In [3] it is shown that the use of these additional constraints prevents the use of a subset of the extreme pathways derived from the hard constraints (stoichiometric matrix) and thus reducing the solution space of allowable metabolic functions. Covert *et* al [3], shows how the reduction of the solution space due to regulatory constraints using extreme pathways analysis.

Transcriptional regulations are not as hard in comparison with mass balance equations. They are only temporally valid as described by the regulatory system. When these constraints are valid, they further reduce the size of the solution space and change its shape as shown in Figure 17 and Figure 18. The addition of known regulatory mechanisms forms the second generation of constraints bases models of complex biochemical reaction networks: models that combine metabolic flux-balance formalism and regulation of gene expression.

As shown in [2], Boolean logic equations are used to describe the regulatory system of an organism. Using these restrictions, some extreme pathways may be become infeasible when external nutrients, for example, are absent from the external medium or when the expression of the gene responsible for producing a metabolic flux has been repressed.

Regulatory events impose temporary, adjustable constraints on the solution as shown in Figure 18 and Figure 17. The first figure shows the solution space of a metabolic network based on the null space of the stoichiometric matrix. If the flux through a certain reaction is knocked-out due to transcriptional regulation, then one or more extreme pathway vectors, which define the boundaries of the solution space, are removed and the volume of the solution space will be reduced. This is shown in Figure 17 where extreme pathway P1 is removed due transcriptional regulation and the solution space is restricted to a smaller space with fewer solutions.

We can extend this to the example with the carbon sources in Figure 14 and Figure 15. Process **Tc2** will become inactive due the presence of **Carbon1**. This means that pathways which include **Tc2** will become infeasible when **Carbon1** is present in the external medium. So, when **Tc2** for example is present in extreme pathway 1 in Figure 17 and Figure 18, this complete pathway will become infeasible when **Carbon1** is present in the external medium.

[3] demonstrated this approach on a skeleton system which has 80 extreme pathways. As regulatory constraints are applied to the system, the number of feasible extreme pathways is reduced to a number between 2 and 26 extreme pathways. The method handled in [3] shows how regulatory mechanisms are used to constrain network functions and produce a small range of physiologically meaningful behavior from all allowable network functions.

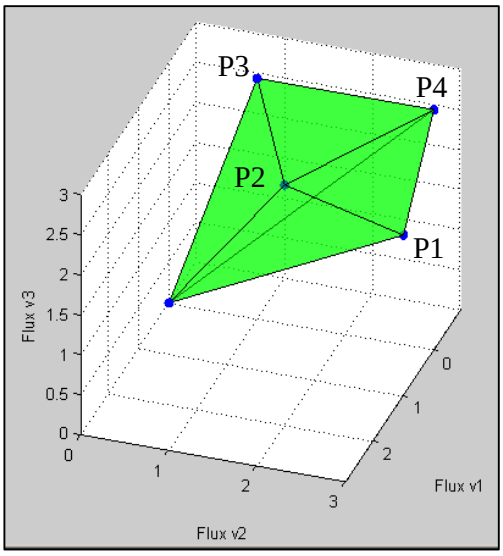

Figure 17 – Solution space without regulatory constraints

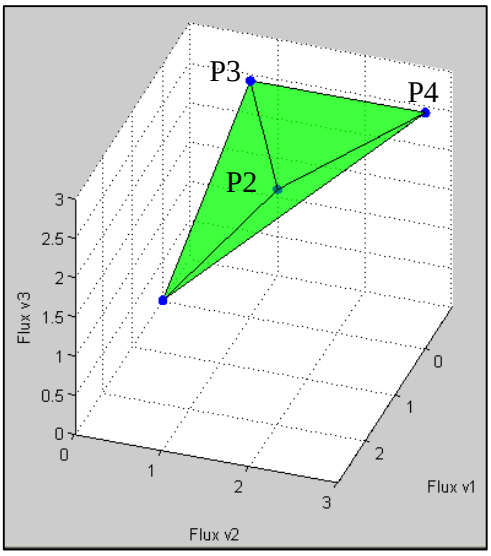

Figure 18 – Solution space with P1 not permitted due regulatory constraint

In total, 80 extreme pathways were calculated based on the metabolic network shown in Figure 14. All extreme pathways are calculated by the algorithm of Shilling, 2000, and are shown in the appendix of [3], paper. Given the total of five inputs towards the metabolic network which are represented by Boolean Logic, there are total of $2^5 = 32$ possible environments conditions which may be recognized by the cell.

The complete list of all possible environmental conditions which can be recognized by the cell are given in the original paper [3]. A small subset of this list is given in Table 2 which shows a subset of possible environmental conditions and the set of repressed enzymes under those conditions. The extreme pathways listed on the right of Table 2 remain feasible under both environmental and regulatory constraints. The complete list of feasible pathways can also be found in the original paper.

Environments					Repressed enzymes				Pathways	Pathway list
C1	C2	F	H	O2		R_{5b}	R_{8a}	T_{c2}	26	P2, P4, P5, P6, P8, P9, P10, P12, P29, P30, P31, P32, P33, P34, P35, P36, P37, P38, P45, P46, P47, P48, P49, P50, P51, P52
C1	C2	F	H		R_{5a}		R_{8a} R_{res}	T_{c2}	10	P39, P40, P41, P42, P43, P44, P49, P50, P51, P52
C1	C2	F		O2		R_{5b}		T_{c2}	8	P29, P30, P33, P34, P45, P46, P49, P50
C1	C2	F			R_{5a}		R_{res}	T_{c2}	4	P41, P42, P49, P50
C1	C2		H	O2		R_{5b}	R_{8a}	T_{c2}	14	P2, P5, P6, P9, P10, P30, P31, P34, P35, P37, P46, P47, P50, P51
C1	C2		H		R_{5a}		R_{8a} R_{res}	T_{c2}	5	P39, P42, P43, P50, P51

Table 2 – Subset of all possible environmental conditions, repressed enzymes and feasible extreme pathways

Some extreme pathways can be eliminated directly because they consist of enzymes which cannot be expressed at the same time due regulatory constraints. Table 2 also shows that according the environmental conditions, the maximum number of extreme pathways is 26 and the minimum is 2. This means a reduction of extreme pathways made by the constraints is between 67.5% and 97.5%.

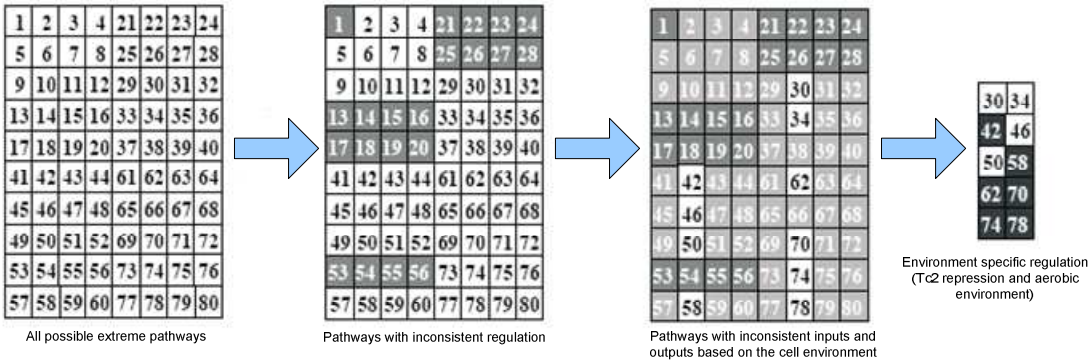

Figure 19 – Reduction of 80 extreme pathways to 4 extreme pathways with the use of constraints

The paper shows that due to regulatory constraints, 21 of the total 80 extreme pathways are always excluded. By considering only the pathways with the appropriate inputs and outputs based on the cell environments (C1, C2, O2 as inputs and D_{ext}, E_{ext} as outputs), another 49 pathways can be eliminated. The last ten remaining pathways can be subjected to more specific environmental conditions used in the above example like repression of **Tc2** process due the presence of **Carbon1**. In this situation, an additional 6 pathways can be eliminated.

The remaining 4 extreme pathways can be used to span the reduced solution space. The example used in [3] shows a 3-dimensional space in where the C1 uptake rate, oxygen uptake rate and growth rate where projected. In [5] it is shown that pathway 30 is the line of optimality (LO) as shown in Figure 21. This line is optimal in relation with the other three lines, since no carbon is lost in secretion of byproduct; pathway 34 in Figure 21 includes secretion of D_{ext} and therefore gives a lower biomass yield then pathway 30. Pathway 46 and 50 are both fermentative which means that they transform sugar into carbon dioxide and alcohol which doesn't require oxygen. So both pathways overlap on the X-axis in Figure 21.

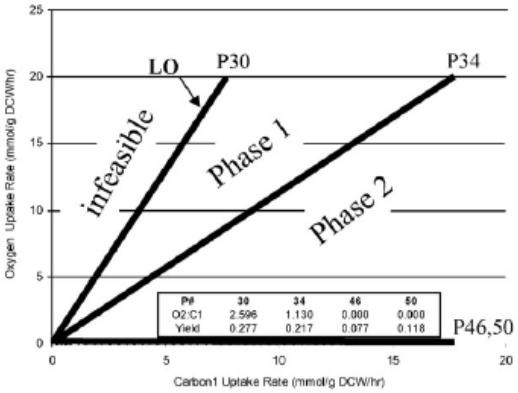

Figure 21 – 2D Phenotype Phase Plane (Carbon & Oxygen)

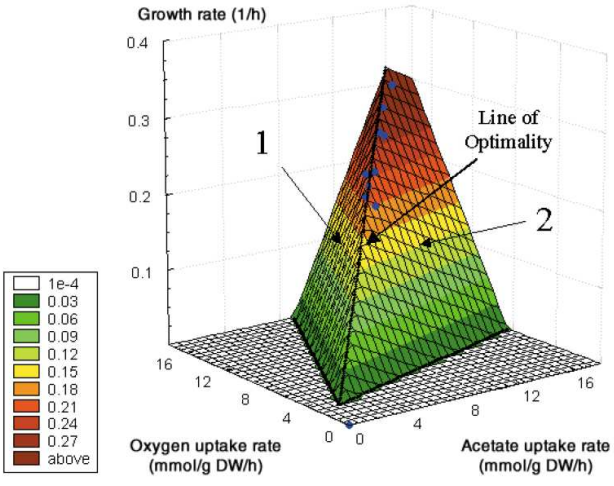

Figure 20 – 3D Phenotype Phase Plane (Carbon, Oxygen & Growth rate)

5 Discussion & Conclusion

The goal of this paper was to describe and discuss the modeling approaches which are able to analyze the fluxes within a metabolic network. This paper describes three of them with the emphasis on flux balance analysis.

Metabolic flux analysis (MCA) is a modeling approach which needs an enormous amount of information of the system to be modeled. The major drawback of this approach is just the lack of this information. A lot of biological information is already known, but many things are still unclear.

Another approach is to use carbon labeling experiment. This approach is based on using modified carbon molecules to analyze the route of these modified molecules. Although some old papers describe the price of these experiments as a drawback, some newer papers tell that these experiments are becoming more affordable.

The last approach investigated was flux balance analysis (FBA). This approach is based on the constraints of a certain metabolic network. It will not predict a precise solution, but instead it will provide a space which contains every possible solution of the network. This range of solutions can provide new ideas and hypothesis which can be investigated by biologists.

The solution space is created by the null space of the stoichiometric matrix which will describes which kind of substrate is necessary and at which quantity to create other products. The null space determined by the singular value decomposition will contain all theoretically feasible solutions, but many of them will not be biologically feasible. An algorithm is used which transform the basis vectors of the solution space into a set of extreme pathways which are biochemical feasible. These new vectors are called the extreme pathways because they represent functional pathways through the metabolic network with a minimal number of chemical reactions.

The great benefit of FBA is the opportunity to represent the extreme pathways in a graph. Each extreme pathway will represented an edge of the cone which holds all feasible solutions based on the stoichiometric matrix. The cone can be constrained in various ways. Hard physical constraints are given by the mass and internal energy balances within the biological system. Other constraints such as gene-regulation and substrate concentration are more environment-specific and are only valid under certain conditions.

Gene-regulation can be applied to investigate the model under conditions where specific genes are knocked-out. This will delete those extreme pathways which incorporate those genes. These deletions will reduce the cone and therefore the range of possible solutions will decrease.

Every modeling approach has its advantages and disadvantages. Metabolic control analysis is very precise and explains the presence of each flux, but the lack of all kinetic information forms a major drawback. Carbon flux experiments will resolve fluxes, but will be used to explore small pieces of the metabolic network and will be time consuming. FBA will not give a precise solution, but instead it will provide a complete range of solutions.

The final question remains how we can improve the search for consistent links between the genotype and phenotype. Introducing new data sources within FBA can further help reduce the range of solutions. Gene expression in combination with transcription factors might give additional information about the presences of certain enzymes. This knowledge will provide extra constraints which can be implemented in the model. FBA provides a stable basis based on the stoichiometry of the organism. Expanding the model with additional data sources will only improve the model and will supply new hypothesis to be tested *in vitro*.

Bibliography

1. Beard D. A., Liang S. D., Qian H.
"Energy balance for analysis of complex metabolic networks"
Biophysical journal, **83**(1), 79-86
(**2002**)

2. Covert M. W., Schilling C. H., Palsson B.
"Regulation of gene expression in flux balance models of metabolism"
J Theor Biol, **213**(1), 73–88
(**2001**)

3. Covert M. W., Palsson B. O.
"Constraints-based models: regulation of gene expression reduces the steady-state solution space"
J Theor Biol, **221**(3), 309-325
(**2003**)

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

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

6. Lee S., Palakornkule C., Domach M. M., Grossmann I. E.
"Recursive MILP model for finding all the alternate optima in LP models for metabolic networks"
Computers and Chemical Engineering, **24**(2), 711-716
(**2000**)

7. Papin J. A., Stelling J., Price N. D., Klamt S., Schuster S., Palsson B. O.
"Comparison of network-based pathway analysis methods"
Trends in biotechnology(Regular ed), **22**(8), 400-405
(**2004**)

8. Price N. D., Reed J. L., Palsson B. O.
"Genome-scale models of microbial cells: evaluating the consequences of constraints"
Nature Reviews Microbiology, **2**(11), 886-897
(**2004**)

9. Schilling C. H., Letscher D., Palsson B. O.
"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**(3), 229-248
(**2000**)

10. Schilling C. H., Palsson B. O.
"The underlying pathway structure of biochemical reaction networks"
Proceedings of the National Academy of Sciences, **270**(3), 415-421
(**2003**)

11. Varma A., Palsson B. O.
"Metabolic flux balancing: basic concepts, scientific and practical use"
Bio/technology(New York, N Y 1983), **12**(10), 994-998
(**1994**)

12. Wiechert W.
"13 C Metabolic Flux Analysis"
Metabolic Engineering, **3**(3), 195-206
(**2001**)

13. Wiechert W.
"Modeling and simulation: tools for metabolic engineering"
Journal of Biotechnology, **94**(1), 37-63
(**2002**)

Appendix A

A1. Linear Programming in Matlab

Linear programming problems are easily solved in the mathematical program Matlab from *The Mathworks Inc*. The difficulty lies in defining the problem in mathematical terms. To explain the use op linear programming in Matlab, we will use a minimization problem with two variables and two constraints.

- **Constraints** - Constraints of the optimization problem are normally described in algebraic equations. To use those equations in Matlab, we need to rewrite the equation in terms of a matrix and two vectors. Each row of the matrix represents one constraint and the columns represent the weights of the variables. The variables itself are describes by a $(nx1)$ vector *x* so are the constants of the equation described by a $(nx1)$ vector *b*.

$$S = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 3 & 1 \end{bmatrix}; b = \begin{pmatrix} 120 \\ 225 \\ 175 \end{pmatrix}; x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$S \cdot x = b$ will form the next algebraic equations

$$x_1 + x_2 = 120$$
$$x_1 + 2x_2 = 225$$
$$3x_1 + x_2 = 175$$

- **Objective function -** The objective function determines the variables in vector form which has to be maximized. The values in the vector define the weight for each variable used in the optimization problem.

$$f_{\min} = -0.2x_1 - 0.35x_2$$

- **Bounds** – The bounds are used to determine the hard bounds of each variable. Two vectors are used to determine respectively the lower and upper bound of each variable used in the optimization problem.

$$LB = \begin{pmatrix} 0 \\ 0 \end{pmatrix}; UB = \begin{pmatrix} 200 \\ 200 \end{pmatrix}$$

```
S    = [1 1; 1 2; 3 1];
b    = [120 225 175];
fmin = [-0.2 -0.35];

LB  = [0;0];
UB  = [200;200];

[Xmax,Valmax] = linprog(fmax,S,b,[],[],LB,UB)
```

Matlab Code

- **Graphical representation** – The best part of linear programming is the ability to show a good graphical representation of the constraints, bound and optimal solution of a linear programming problem. Each constraint can be represented by a straight line in a graph. The bounds are depending on the variable, horizontal or vertical lines which bound the possible values for the variables.

  The next two graphics shows two linear programming examples; figure A2 shows a simple maximization problem and figure A1 shows the graphical representation of the minimization example used above. The green areas show the field of possible values for the two variables x_1 en x_2. With the use of either Matlab or linear programming by hand, we can calculate the optimal point based on type of problem and the objective function. The optimal for these two problems are displayed as a red square in the graphs.

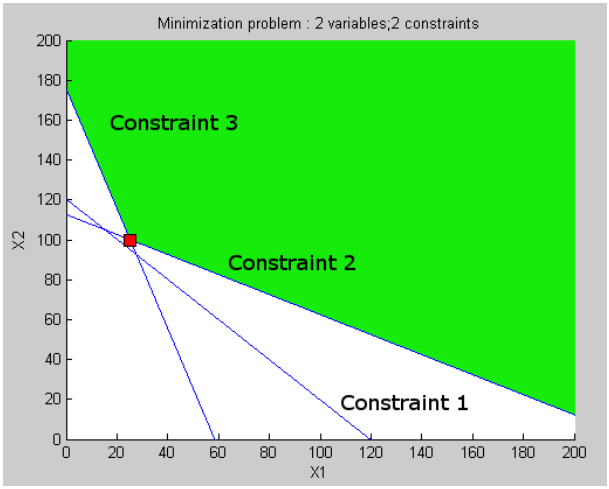

Figure A1 - Minimization problem

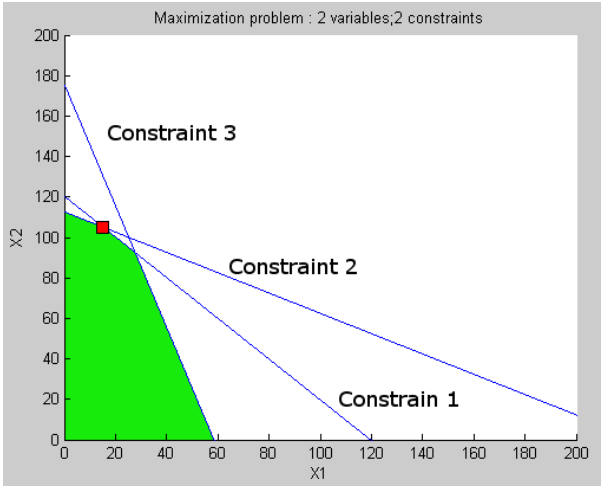

Figure A2 - Maximization problem

- **Double-sided linear constraints** - Some constraints are so-called double sided linear constraints which mean that an equation is bounded from two sides which is shown in Figure 12.