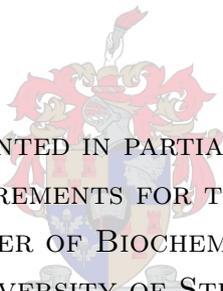


Integration of kinetic models with data  
from  $^{13}\text{C}$ -Metabolic Flux experiments

Willem Petrus Du Toit Schabort



THESIS PRESENTED IN PARTIAL FULFILMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF BIOCHEMISTRY  
AT THE UNIVERSITY OF STELLENBOSCH

Supervisor: Prof. Jacky L. Snoep  
Co-supervisor: Prof. Johann M. Rohwer  
December 2007

# Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature: .....

Date: .....

Copyright (c) 2007 Stellenbosch University  
All rights reserved

# Abstract

A detailed mathematical description of all the processes in a cell could be an informative tool for investigating biological function. Detailed kinetic models could be built either by obtaining enzyme kinetic parameters *in vitro*, or by obtaining them from time series analyses of metabolite data from rapid pulse experiments. A genome scale *in vitro* enzyme kinetic assay project would be prohibitively laborious with the current technologies. Further, there are still uncertainties about the importance of *in vivo* effects such as metabolite channelling, spatial effects and molecular crowding which could make *in vitro* determined parameters invalid. Accordingly, there is much interest in *in vivo* experiments for kinetic modelling. *In vivo* experimental methods suffer from a number of technical and even fundamental problems. Technical problems are being solved by more sensitive metabolomics tools and rapid sampling technologies. However, the large number of effectors of each enzyme reaction makes it impossible to obtain models at the level of detail possible with the *in vitro* method. Ultimately, the solution to building a genome scale Silicon Cell is to make use of both strategies. As metabolomics technologies are rapidly improving, it would thus make sense to follow the parts-based *in vitro* kinetics methodology, and carry out a detailed accuracy assessment of the model with *in vivo* experiments. To address the problem of the fundamental limit of information from concentration time-series, other *in vivo* experiments will have to be carried out as well.  $^{13}\text{C}$ -metabolic flux analysis has recently undergone vast improvements with the use of better experimental protocols and powerful algorithms for flux calculation. Incorporation of this type of experiment in the validation protocol is the aim of this

thesis, which represents an intermediary step towards using the genome-scale stoichiometric models as platforms for building genome-scale kinetic models. It is illustrated here how kinetic models can be combined with metabolic flux data in a special way which allows correct modelling of boundary conditions and validation using novel concepts. We used  $^{13}\text{C}$ -metabolic flux analysis and gas chromatography-mass-spectrometry to measure metabolic fluxes through the central metabolic pathways of the yeast *Saccharomyces cerevisiae*. This data was integrated with a previously constructed detailed kinetic model of fermentative glycolysis in the yeast to illustrate our approach. Various implications for such data integration with kinetic models were identified and a software program was designed for this purpose.

# Opsomming

'n Gedetailleerde wiskundige beskrywing van al die prosesse in 'n sel kan 'n kragtige middel wees vir die bestudering van biologiese funksies. Gedetailleerde kinetiese modelle kan gekonstrueer word vanaf ensiem kinetiese parameters soos bepaal *in vitro*, of deur dit te verky met tydreëks analise van metaboliet data vanaf puls eksperimente. 'n Genoom-skaal *in vitro* ensiemkinetika projek sou te werksintensief wees met die huidige tegnologieë. Verder is daar nog onsekerhede oor die belangrikheid van *in vivo* faktore soos metaboliet kanallering, ruimtelike faktore en molekulêre verdringing wat *in vitro* parameters ongeldig sou kon maak. Gevolglik word daar baie belang gestel in *in vivo* eksperimente vir kinetiese modellering. *In vivo* eksperimentele metodes ly onder 'n aantal tegniese en selfs fundamentele probleme. Tegniese probleme word oorbrug deur ontwikkeling van meer sensitiewe metaboloom tegnieke en vinnige monsterring. Heelaas maak die groot aantal affektore van elke ensiemreaksie dit onmoontlik om modelle met dieselfde vlak van detail te verkry as wat moontlik is met die *in vitro* metode. Gevolglik is die oplossing to die konstruksie van 'n genoom-skaal Silikon Sel dat beide strategieë gevolg moet word. Aangesien metaboloom tegnologieë tans vining verbeter, maak dit dus sin dat die deelwyse *in vitro* metode te gebruik, gevolg deur gedetailleerde akkuraatheids assessering van die model met *in vivo* eksperimente. Om die fundamentele limiet van informasie vanaf konsentrasie tydreëks data te oorbrug moet ander *in vivo* eksperimente ook uitgevoer word. <sup>13</sup>C-metaboliese fluksie analise het onlangs groot verbeteringe ondergaan met die gebruik van beter eksperimentele protokol en kragtige algoritmes vir fluksie berekeninge. Inkorporering van

hierdie tipe eksperiment in die valideringsprotokol is die doelwit van die tesis, wat 'n intermedieêre stap voor die gebruik van genoom-skaal stoichiometriese modelle as platforms vir die bou van genoom-skaal kinetiese modelle voorstel. Dit word hier gewys hoe kinetiese modelle met metaboliese fluks data gekombineer kan word in 'n spesiale manier wat modellering van rant kondisies toelaat, asook validering op grond van nuwe beginsels. Ons het  $^{13}\text{C}$ -metaboliese fluksie analise an gas chromatografie-massa-spektrometrie gebruik om fluksies te meet deur die sentrale metaboliese wieë van die gis *Saccharomyces cerevisiae*. Hierdie data was gekombineer met 'n voorheen gekonstrueerde kinetiese model van fermentatiewe glikolise in die gis om die tegniek te illustreer. Daar is klem gelê op verskeie implikasies vir sulke data integrering met kinetiese modelle en 'n sagteware program was ontwerp vir hierdie doel.

# Acknowledgements

I would like to give special thanks to a number of people:

My supervisors Professors Jacky Snoep and Johann Rohwer. Thank you for the privilege to study in an academic environment of world class. I enjoyed working in a light-hearted, friendly atmosphere.

To Dr. Nicola Zamboni and Professor Uwe Sauer. Thank you for the kind gift of the flux software and for the inspiring work you are doing.

Dr. Stefan Louw at the GC-MS. I appreciate the expertise that you brought into the project. Thank you for being a friend.

Mr. Eric Ward. Thank you for your enthusiastic participation in the design of the miniature equipment.

Dr. Christie Malherbe. Your advice on fermentations has helped me tremendously during this time. Your enthusiasm for the subject has been an inspiration for us all.

To my parents Johannes and Lulu and my brother Charl. I am very lucky to have such a loving family. Thank you for standing by me when things got tough.

My friends, thank you for the support, you mean a lot to me.

# Acronyms

**<sup>13</sup>C-MFA** <sup>13</sup>C-metabolic flux analysis

**CE** capillary electrophoresis

**BFL** biosynthetic fractional labelling

**DMFA** dimethylformamide

**EI** electron impact ionization

**EMP** Embden-Meyerhoff-Parnass pathway

**FRET** fluorescent resonance energy transfer

**FTICR-MS** fourier transform ion cyclotron resonance mass spectrometer

**GC** gas chromatography

**HMP** hexose monophosphate pathway

**HPLC** high-performance liquid chromatography

**IMFA** instationary metabolic flux analysis

**LC** liquid chromatography

**MCA** metabolic control analysis

**MFA** metabolic flux analysis

**MS** mass spectrometry

**MTBSTFA** N-(tert-butyldimethylsilyl)-N-methyltrifluoroacetamide

**NMR** nuclear magnetic resonance

**ODE** ordinary differential equation

**ORF** open reading frame

**PCA** perchloric acid

**PES** polyethersulfone

**RI** refractive index

**TOF** time-of-flight

**UPLC** ultra-pressure liquid chromatography

**UV** ultraviolet

# Contents

<b>1</b>	<b>Systems Biology, Omics, and the Silicon Cell</b>	<b>1</b>
1.1	Metabolite balancing . . . . .	2
1.2	Flux balance analysis: from observatory flux calculation to predictive network analysis . . . . .	5
1.3	Genome-scale biochemical network models . . . . .	7
1.4	Kinetic modelling and Metabolic Control Analysis . . . . .	9
1.5	Omics meets systems biology and bioinformatics . . . . .	12
1.6	Steps towards building the Silicon Cell . . . . .	14
<b>2</b>	<b>Theoretical aspects of <math>^{13}\text{C}</math>-metabolic flux analysis</b>	<b>18</b>
2.1	Analysis of data from isotope labelling experiments . . . . .	18
2.2	Detection of labelling patterns by nuclear magnetic resonance spectroscopy and mass spectrometry . . . . .	31
<b>3</b>	<b>Flux-constrained kinetic modelling: dealing with parts of a Silicon Cell</b>	<b>35</b>

3.1	Boundary conditions in kinetic models . . . . .	36
3.1.1	Concentration boundaries . . . . .	37
3.1.2	Enzyme boundaries . . . . .	38
3.1.3	Branch reactions as boundaries . . . . .	39
3.2	Simulation for accuracy assessment . . . . .	40
3.2.1	Setting of steady state pathway fluxes for validation . . . . .	40
3.2.2	Interesting properties of isotope labelling experiments . . . . .	44
3.3	Technical considerations in flux-constrained simulation . . . . .	48
3.3.1	Considerations for constraining pathway fluxes . . . . .	49
3.3.2	Implications of conserved moieties when using boundary reactions . . . . .	53
3.4	Dynamic pulse experiments: potential and limitations . . . . .	56
3.5	Simulation for prediction: Metabolic Control Analysis . . . . .	60
<b>4</b>	<b>Kinomics</b> . . . . .	<b>63</b>
4.1	Main features and uses . . . . .	63
4.2	The internal structure of Kinomics . . . . .	66
<b>5</b>	<b><sup>13</sup>C-metabolic flux experiments</b> . . . . .	<b>79</b>
5.1	Experimental procedures . . . . .	79
5.1.1	Growth conditions . . . . .	81
5.1.2	Determination of cellular growth rate . . . . .	82
5.1.3	Cellular dry weight determination . . . . .	83
5.1.4	Analysis of external metabolites with HPLC . . . . .	83

5.1.5	Analysis of labelled proteinogenic amino acids with GC-MS . . . . .	84
5.1.6	Determination of production and consumption rates of external metabolites . . . . .	85
5.1.7	Metabolic flux ratio analysis and net flux analysis . . . . .	86
5.2	Experimental results and discussion . . . . .	87
5.2.1	Physiological characterization of <i>Saccharomyces cerevisiae</i> growth on glucose . . . . .	87
5.2.2	GC-MS analysis of labelling patterns in proteinogenic amino acids. . . . .	91
5.2.3	Net flux analysis with FiatFlux . . . . .	98
<b>6</b>	<b>Flux-constrained simulation</b>	<b>103</b>
6.1	Details of the simulation . . . . .	103
6.2	Investigating the cause of model behaviour . . . . .	105
<b>7</b>	<b>Discussion and conclusion</b>	<b>109</b>
<b>A</b>	<b>Rate equations</b>	<b>114</b>
<b>B</b>	<b><sup>13</sup>C-metabolic model</b>	<b>116</b>
<b>C</b>	<b>Fermentation data plots</b>	<b>118</b>

"Progress in science depends on new techniques, new discoveries  
and new ideas, probably in that order."

Sydney Brenner, *Nature*, 5 June 1980

# Preface

In our age it has become very important to understand biochemical processes in terms of mathematical models that can capture their nonlinear and complex behaviour. Various approaches have been used to describe parts of a living cell - enzyme kinetic models, stoichiometric models, fuzzy logic and cybernetic models. Eventually, we would like to build large models that can describe a complete cell, incorporating all levels of regulation. Some have pointed out that the marriage between enzyme kinetics, stoichiometry and genetic regulation can be a detailed mechanistic one only in the high-information future, while we now have to use cybernetics and fuzzy-logic to reach practical applications [1, 2]. We agree that detailed kinetic modelling of the parts is difficult and that it is very ambitious to model a complete organism at this stage. However, the recent advances in analytical technology allow a more data rich process. It is argued that improved measurement techniques, especially those for estimation of metabolic fluxes, can now assist in a detailed kinetic modelling process of the cell.

The Silicon Cell Initiative is our long-term goal, which is aimed at a detailed kinetic model of a cell. Instead of trying to create a genome-scale model by starting with cybernetic or fuzzy modelling and working towards more detail, our starting point is a detailed bottom-up approach, while the system boundaries are well-characterized for the reference conditions with experimental data. This approach involves combining of kinetic models with larger, but static stoichiometric models. In essence, the larger stoichiometric model is used as a data model for calculating metabolic fluxes, and it is connected to the kinetic model only through a set of measured bound-

ary reactions. We focused on experiments using  $^{13}\text{C}$ -metabolic flux analysis ( $^{13}\text{C}$ -MFA) for it provides very detailed flux data. The aims of this thesis are:

1. To learn the techniques in  $^{13}\text{C}$ -MFA.
2. To investigate how the boundary reactions should be incorporated for steady state kinetic modelling.

Before our detailed discussions from Chapter 3 onwards, we start with an overview of the most relevant principles in systems biology in Chapter 1. These are set in the perspective of a rapidly changing scientific world characterized by large-scale data acquisition technologies (the 'omics') and information databases, and how *some* fields of systems biology are striving towards harnessing the information content. It is argued that strong elements of omics and bioinformatics should be incorporated into the modelling process of a Silicon Cell. Without the practicality that these elements provide, I suspect that large-scale detailed kinetic modelling will merely remain a goal for very long. Further, these elements will ensure that success in the outcome - a satisfactory description of a model organism - will not only be applicable to the model organism. These elements will help to develop a generic protocol for building genome-scale kinetic models, and allow application to many real-world problems.

While Chapter 1 is an introduction to some of the relevant themes in current systems biology and discusses some principles that are not directly used in our study, such as flux balance analysis, Chapter 2 is a technical introduction to the main principles of  $^{13}\text{C}$ -MFA. In our flux-constrained kinetic modelling approach,  $^{13}\text{C}$ -MFA and kinetic modelling (discussed in Chapter 1) are the main principles used. Chapters 1 and 2 thus together form the bulk of our literature study.

In Chapter 3 we discuss how incorporation of boundary fluxes is in principle necessary to obtain a realistic assessment of the accuracy in the modelled kinetics. It is also explained that  $^{13}\text{C}$ -MFA experiments can yield interesting

information that can be applied in a more detailed assessment of model accuracy, to be used in combination with dynamic pulse experiments. Based on these findings we describe a software program called *Kinomics* that we created for this flux-constrained modelling regime (Chapter 4).

To illustrate our approach we measured metabolic fluxes in the yeast *Saccharomyces cerevisiae* at a quasi steady state with  $^{13}\text{C}$ -constrained flux balancing, using gas chromatography-mass-spectrometry (GC-MS) and the analysis program FiatFlux (Chapter 5). Our program was then used to constrain a detailed kinetic model of yeast glycolysis with the flux data from our experiments (Chapter 6).

Apart from the specialized method development for the combination of kinetic modelling and flux analysis, this work can be seen as a starting point to further multidisciplinary studies in the context of the Silicon Cell.

# Chapter 1

## Systems Biology, Omics, and the Silicon Cell

In the current era of 'functional genomics', the meaning of 'genomics' has changed somewhat from the study of DNA and the genes that it contains, to the wider meaning of studying biological components at the genome-scale. What has caused this shift in meaning, is the strong emphasis on large-scale data acquisition which drives the continued technological breakthroughs in the omics fields. The result is that complete genomes can rapidly be sequenced and a vast array of high-throughput technologies has become available for observing thousands of cellular components at once. Of these, microarray technologies for observing the RNAs present are perhaps the best known of all. Usually, these omics are used in a fashion where the profile of a certain experimental state of importance, like a disease state, is compared to that of the normal state. Statistical methods are then used to identify components of which their presence or abundance is in some way correlated to the characteristic behaviour. Unfortunately, exploration of data in a purely statistical manner does not give deep insight into the workings of a biological system. The nonlinear and complex behaviour of biochemical systems necessitates an understanding in terms of mechanistic models that capture the nature of these systems.

One characteristic aspect of the omics is the remarkably effective techniques

for identifying, and in some instances quantifying components, but an inability to provide the kinetic features of reaction networks. That is, *how fast* are components interconverted (metabolism), how fast do components activate or inhibit one another (signal transduction) and how fast are they synthesised *de novo* and broken down (all macromolecules). These kinetic aspects hold useful information to explain the observed component patterns of the omics. This need for kinetic data types is becoming increasingly evident and significant effort has gone into the calculation of metabolic fluxes. Metabolic Flux Analysis (MFA) is now seen as an important form of data for integrated functional genomics to complement the proteomics and transcriptomics [3].

MFA uses models which are based on stoichiometry. The stoichiometry of a metabolic network is thus the first type of model that we can use to relate data to a deeper understanding of the phenotype. Stoichiometric models can be used in a number of related frameworks, of which metabolite balancing, flux balance analysis (FBA) and genome-scale network models are discussed in the first few sections of this chapter, while stable isotope based flux analysis is discussed in more detail in Chapter 2.

## 1.1 Metabolite balancing

Metabolite balancing is based on the assumption that mass is conserved in a metabolic system at a steady state. The steady state is the condition where there is mass flow through a metabolic system but the metabolite concentrations do not change over time. A set of metabolite balance equations can be constructed from stoichiometric coefficients of enzymatic reactions which describe the time differential changes in concentrations as a function of reaction rates. The balance equations can then be parameterized by measurements of clearance of extracellular substrates and increases in product concentrations in order to calculate the intracellular fluxes. Figure 1.1 shows a small metabolic network which can be described in terms of the metabolite

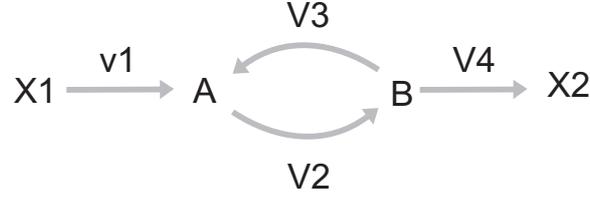


Figure 1.1: Metabolite balancing. See text.

balance equations in equation set 1.1

$$\begin{pmatrix} \mathbf{X}_1'(t) \\ \mathbf{X}_2'(t) \\ \mathbf{A}'(t) \\ \mathbf{B}'(t) \end{pmatrix} = \begin{pmatrix} -1.v_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.v_4 \\ 1.v_1 & -1.v_2 & 1.v_3 & 0 \\ 0 & 1.v_2 & -1.v_3 & -1.v_4 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'(t) \\ \mathbf{X}_2'(t) \\ 0 \\ 0 \end{pmatrix} \quad (1.1)$$

$$\begin{pmatrix} \mathbf{X}_1'(t) \\ \mathbf{X}_2'(t) \\ \mathbf{A}'(t) \\ \mathbf{B}'(t) \end{pmatrix} = \begin{pmatrix} -1.v_1 & 0 & 0 \\ 0 & 0 & 1.v_4 \\ 1.v_1 & -1.v_{23} & 0 \\ 0 & 1.v_{23} & -1.v_4 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'(t) \\ \mathbf{X}_2'(t) \\ 0 \\ 0 \end{pmatrix} \quad (1.2)$$

$$\begin{pmatrix} \mathbf{X}_1'(t) \\ \mathbf{X}_2'(t) \\ \mathbf{A}'(t) \\ \mathbf{B}'(t) \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 1 \\ 0 & -1 & 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_4 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'(t) \\ \mathbf{X}_2'(t) \\ 0 \\ 0 \end{pmatrix} \quad (1.3)$$

If the intracellular metabolites A and B are assumed to be at a steady state, their time differentials would equal zero. The occurrence of a single entry in the top left corner of equation set 1.1 indicates that if one would measure the time differential of extracellular metabolite  $X_1$ , reaction  $v_1$  could be uniquely determined, and so could  $v_4$  be uniquely determined if  $X_2$  was measured. The time differential of an external metabolite can be measured in a chemostat bioreactor as the concentration of the metabolite multiplied by a flow constant which is always known, or calculated from the clearance or production rate in case of a batch fermentation. The latter case would

be referred to as a quasi steady state. However, the network structure implicates that  $v_2$  and  $v_3$  cannot be calculated either by measurements of any of  $X_1$  or  $X_2$ , or by measuring both. This is because in both rows of  $A'(t)$  and  $B'(t)$ , the same two reaction variables  $v_2$  and  $v_3$  occur, which is referred to as rank deficiency. There are thus too many variables to calculate from metabolite balances. We can combine columns 2 and 3 into a single column which describes the net action of reactions  $v_2$  and  $v_3$  as  $v_2-v_3$  (equation set 1.2). From the known  $v_1$  we can thus calculate  $v_{23}$  which is the only other variable in the row for  $A'(t)$  and from that we can calculate  $v_4$  in  $B'(t)$ . This means that we did not actually have to measure  $B'(t)$  as the metabolite balance constraints allowed  $v_4$  to be calculated from  $A'(t)$ .

For this type of calculation, formal mathematical procedures have been defined. The stoichiometric matrix  $N$  is separated into two parts. In the first part  $N_b$ , all the fluxes are directly known from measured consumption and production fluxes. In the second part  $N_n$ , all the fluxes have to be calculated from the network constraints together with the measured part. The corresponding reaction vectors are  $r_b$  and  $r_n$  (see [4] and equation set 1.3). The stoichiometric matrix equation

$$0 = N.r \quad (1.4)$$

thus becomes

$$0 = N_b.r_b + N_n.r_n \quad (1.5)$$

which can be rewritten as

$$r_n = -N_n^{-1}.N_b.r_b \quad (1.6)$$

$N_n^{-1}$  is the precise inverse of  $N_n$ , but is actually not used very frequently, for it is only valid if  $N_n$  is square. Therefore, the relation

$$r_n = -N_n^\# .N_b.r_b \quad (1.7)$$

is used instead, where  $N_n^\#$  is the Moore-Penrose pseudo inverse and is valid for any matrix [4, 5]. It allows the calculation of  $r_n$  by a least-squares solution.

Some network structures unfortunately make the internal fluxes in metabolic systems impossible to calculate from the measured external fluxes, as shown in the example above. Mathematical algorithms have been designed to make the best use of the available data and calculate as many calculable fluxes as possible [4, 6–11]. In the example above, measurements of both  $A'(t)$  and  $B'(t)$ , which will inevitably differ due to measurement errors, can thus be used maximally because they are *balanceable*.

However, metabolite balance constraints are generally insufficient for calculation of realistic metabolic networks. A more powerful approach is to use substrates labelled with stable isotopes such as  $^{13}\text{C}$ , and then to analyse the isotope labelling patterns of metabolites. By using various statistical or algebraic methods, the labelling patterns can then be used to provide the necessary extra constraints to calculate the internal flux distributions. Theoretical frameworks and experiments for  $^{13}\text{C}$ -Metabolic Flux Analysis ( $^{13}\text{C}$ -MFA) have recently been improved to become powerful techniques for flux calculation [3, 12, 13], which is the topic of the next chapter. It is important that we extend metabolic flux experiments to the omics level to form the functional complement of the existing data technologies that have been developed to that scale. A major goal is to be able to use the recent genome-scale metabolic reconstructions [14] as the basis for flux calculations.

## 1.2 Flux balance analysis: from observatory flux calculation to predictive network analysis

Although it is very useful to know the flux pattern in a cell, it is still only a snapshot of cellular function that is obtained. The use of stoichiometric models for flux measurements thus has fundamental limits for making predictions about behaviour in different environments or genetic modifications. Also, as experimental methods do not yet allow us to calculate fluxes at a genome-scale, there is a strong emphasis on *predictions* by the genome-scale stoichiometric model, as opposed to its use as a *measurement* model. Flux balance analysis (FBA) aims to exploit our knowledge of network structure

at the genome-scale to predict cellular behaviour at a steady state.

A stoichiometric model can be used to predict the range of possible flux patterns that are allowable to operate in steady state, but does not predict which of the infinitely many possibilities will be used. It is thus not very constringent in its predictions. Accordingly, extra hypotheses are used in combination with the steady state formalism to make more constringent predictions possible. In this constraint-based modelling approach, [14,15], the hypothesis usually used is that metabolic networks are globally optimized for sustaining maximal growth rates. The experimentally determined biomass composition is thus included and an optimization routine is applied that maximizes biomass fluxes from the precursors in the determined ratios. The system is also constrained at enzyme maximal activities such as substrate transporters to set the limits for the optimization. The flux capability is frequently investigated in the dimensions of substrate uptake to study limitation properties of the network in phenotype phase plane analysis [16,17], or by defining limit potentials [18].

A question is whether these optimization criteria are valid, and whether they exist at all in nature. Certainly, not all cell types are optimized for maximizing growth rate, for then some pathways would not be used by organisms that also have more effective methods of energy production. Inference and testing of optimality criteria on actual data is at the heart of constraint-based modelling. Significant effort is being invested in the identification of optimality criteria and programs have been designed for this purpose and can apply more than one criterion [19]. An important step was the introduction of the idea that mutants usually respond to a gene deletion by striving to return to the wild-type phenotype as well as it possibly can. The MOMA approach (minimization of metabolic adjustment) [20] can be used to predict the viability of mutants. MOMA has lately been applied using phenotypic data in the form of 30 quantified metabolic fluxes from  $^{13}\text{C}$ -MFA experiments from four different wild-type conditions. This data served as realistic optimization criteria for the large-scale stoichiometric model iLL672 [21]. Gene knockout strategies for lactate overproduction in *Escherichia coli* based on two optimality criteria using OptKnock [22]

have resulted in lactate overproduction of 25-73%, proving the reliability of constraint-based FBA predictions.

FBA can also be readily extended to incorporate diverse data types and constraints. The laws of thermodynamics are applied to the analysis of metabolomics datasets in Energy Balance Analysis (EBA) [23–25]. FBA has also been extended to predict dynamic flux responses and genetic regulation [26–28] and the optimality criteria have been improved accordingly to more complex formulations [29]. Although it is not yet possible to quantify all the metabolites in the cell accurately due to difficulties with metabolomics, large-scale stoichiometric techniques thus provide a way to make use of some of this data [15]. FBA is thus becoming an efficient framework for capturing data at the genome scale to complement the omics.

### 1.3 Genome-scale biochemical network models

A range of genome-scale metabolic pathway representations have been constructed during the past few years [14]. Of note are the models of *Haemophilus influenzae* [30], *Escherichia coli* [31], *Saccharomyces cerevisiae* [32, 33] and for the important human pathogen *Mycobacterium tuberculosis* [17, 34] constructed by the groups of Palsson and others. The procedure of reconstruction starts with an annotated genome [32]. Additionally, online databases, textbooks and literature are also used. This research is a critical link to the complete genome as reference and forms a broad picture of the metabolic reactions in cells. The yeast metabolic pathway model iND750 for instance, consists of 1147 reactions and 1061 metabolic species and is fully compartmentalized into organelles [33].

Like metabolic networks, reconstructions of genome-scale protein interaction networks are also being made but are more difficult to assemble than metabolic networks [35]. The principles used are more experimentally orientated. The yeast-two-hybrid experimental method to infer protein interactions [36] can provide essential large-scale information, but is difficult to interpret for there are many false positives. Direct co-purification or in-

direct affinity tag purification can also be used efficiently to decipher the localization of proteins into organelles and complexes [35]. Much emphasis has recently been placed on integration of diverse types of genomic data to predict protein interaction networks by combining bioinformatics and experimental approaches [37]. Protein interaction networks can be incredibly large and interconnected, like the recently reconstructed human interactome map with more than 7500 mapped interactions [38]. Apart from helping to elucidate the functions of many uncharacterized proteins, interactomes are seen as a suitable framework for cellular modelling [39] as they form functional networks.

Signalling networks are also of high importance for genome-scale reconstruction [40], but are very difficult to construct. Interesting is that signalling models can be analyzed in terms of many of the same principles used for metabolism, such as elementary modes [41], and a method has been developed by which stoichiometric models and signalling models can be analyzed in a combined way [42, 43].

FBA, like other approaches that are based on constraints of network structures, is limited by the fact that the nonlinear nature of mechanisms - the enzyme mechanisms, protein interactions and signal transduction - is not fully captured in a rigid network model. These methods thus have fundamental limits in predicting the more complicated emerging properties of biochemical networks that can only be predicted by taking the concentrations of the biochemical species into account. FBA, for instance, may be used efficiently for studying the possible effects of gross manipulations such as the knockouts of enzymes, but the knowledge of more subtle changes in the biochemical macromolecules that affect their functioning cannot be used in a sensible manner. On the contrary, deep theoretical analyses of biological systems have been done for decades, but without the availability of large omics datasets, genome-scale reconstructions and bioinformatics. Advanced mathematical methods in kinetic modelling have been formulated for analysis of biochemical systems. It now makes sense that these methods should be applied in combination with bioinformatics and the omics technologies. The omics and bioinformatics can help us to create and validate kinetic

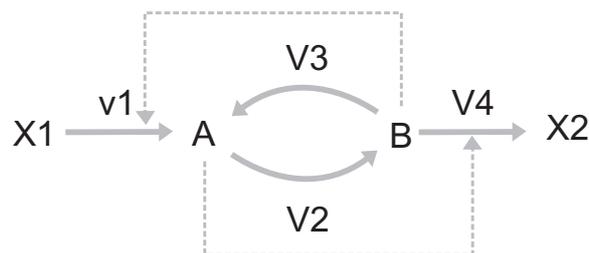


Figure 1.2: The structure of a simple kinetic model incorporating non-stoichiometric effectors. See text.

models, while the kinetic models can help us to make sense of the omics datasets and guide further experimentation.

## 1.4 Kinetic modelling and Metabolic Control Analysis

For kinetic models of metabolism, a stoichiometric model also forms the network structure as in FBA, but the reaction rates respond to the concentrations of modelled metabolites and a steady state is thus not assumed. Simulation is usually based on a numerical integration of ordinary differential equations, and sometimes partial differential equations. The method could thus be used to study a biochemical system functioning at the steady state [44, 45] or to study non-steady state behaviour such as the oscillatory behaviour found in metabolism [46, 47] and in the cell cycle [48–50].

For a good introduction to the formulation of a kinetic model, see [51]. The metabolic network in figure 1.2 with feedback and feedforward loops can be described in terms of the stoichiometric matrix equation

$$\begin{pmatrix} \mathbf{X}_1'(t) \\ \mathbf{X}_2'(t) \\ \mathbf{A}'(t) \\ \mathbf{B}'(t) \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & -1 & 1 & 0 \\ 0 & 1 & -1 & -1 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} \quad (1.8)$$

where each of the reactions in the reaction vector is now a rate equation which describes the reaction rate as a function of the metabolite concentra-

tions (M), and various enzyme kinetic parameters (P).

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} = \begin{pmatrix} v_1 \left( M[X1, A, B], P[V_{max1}, K_{eq1}, P_{1X1}, P_{1A}, P_{1B}] \right) \\ v_2 \left( M[A, B], P[V_{max2}, K_{eq2}, P_{2A}, P_{2B}] \right) \\ v_3 \left( M[A, B], P[V_{max3}, K_{eq3}, P_{3A}, P_{3B}] \right) \\ v_4 \left( M[A, B, X2], P[V_{max4}, K_{eq4}, P_{4X1}, P_{4A}, P_{4B}] \right) \end{pmatrix} \quad (1.9)$$

One could make distinctions between the different types of kinetic models based on the way they are constructed. Of note, it is important to distinguish between models constructed from rate equations and parameters obtained *in vitro*, and models of which the parameters are fitted to concentration time series data. Teusink *et. al.* characterized the individual enzymes *in vitro* for their kinetic parameters and constructed a detailed model of yeast glycolysis [44]. Rate equations in this model reflect the real cellular machinery, for example the reversible Michaelis-Menten rate equation used for the hexose isomerase

$$v = V_{max} f \frac{\frac{[S]}{K_s} - \frac{[P]}{K_s K_{eq}}}{1 + \frac{[S]}{K_s} + \frac{[P]}{K_p}} \quad (1.10)$$

Some mechanistic rate equations are very complex, for example the Monod-Wyman-Changeux equation which can describe the phosphofructokinase with its many metabolic effectors and binding cooperativity (see Appendix A). Many other detailed equations have been derived which reflect the different mechanisms by which enzymes catalyze their reactions and the order in which the implicit enzymatic steps are catalyzed. By contrast, Rizzi *et. al.* fitted kinetic parameters to a set of semi-realistic enzyme kinetic equations during an *in vivo* substrate pulse experiment [52]. In the first example, a model was constructed from the reductionistic knowledge of the parts (enzymes), whereas in the second example, time-dependent data was used to infer parameters. An important concept for us is context dependence. The realistic *in silico* enzymes in the Teusink model could function under many conditions (contexts), for it can readily be claimed that the equations describe real enzyme properties. On the other hand, a model with fitted parameters might not be used realistically in conditions that are much different

from those used for model construction. Context-independence arises with increasing accurately modelled detail that describes the mechanism of the process. It is important to note that if the same detailed equations could be used for fitting, the same level of context-independence could be achieved. However, during fitting, an integrated system of equations with many parameters and variables has to be fitted. Although separation of a model into smaller dynamic subsystems [52] provides a more direct link between dynamic data and enzyme parameters, the *current* accuracy and scope of metabolomics data does not allow such detailed equations to be used with good confidence [15]. However, quantitative metabolomics technology will improve in the future and enable fitting of detailed model parameters which will approach the detail obtainable with the *in vitro* approach. In any case, it is our objective to use such dynamic experiments mainly for validation purposes, with model construction based on characterization *in vitro*.

Whatever the method of model construction and validation, advanced mathematical frameworks have been developed for making predictions from a model, of which the most notable is Metabolic Control Analysis (MCA) [53]. Since its initial formulation [54] it has been extended several times to result in a very powerful analytical framework [55–59]. MCA provides metabolic control coefficients and metabolic response coefficients. In its original formulation, a flux control coefficient  $C_v^J$  can be defined as the response of any steady state flux  $J$  to an infinitely small change in a reaction rate  $v$ , and a concentration control coefficient  $C_v^S$  as the response of any steady state concentration  $S$  towards a small change in a reaction rate  $v$ . Flux response coefficients  $R_p^J$  and concentration response coefficients  $R_p^S$  are those sensitivities towards small changes in any of the parameters in the model. Apart from the conventional steady state analysis of biochemical systems, MCA has also been developed for oscillating biochemical systems [60]. The most recent development is a generic approach based on perturbations to characterize many other time-dependent aspects that cannot be addressed with other approaches [61]. MCA is thus a powerful method for determining the response of a biochemical system towards perturbations.

How informative the deductions from MCA may be [62], the main factor

that has been hampering the large-scale use of MCA, and kinetic modelling in general, is the uncertainty of how realistic such models are. The effects of molecular crowding, metabolite channelling and spatial effects in the cell are concerns. An important recent development in MCA has been the introduction of a framework to incorporate the uncertainties associated with kinetic parameters, metabolite concentrations and metabolic flux measurements [63]. Monte Carlo Simulation has been used to model the uncertainties, which are then projected via the (log)linear MCA formulation to the control coefficients. Incorporation of probability theory is an exciting new development that allows us to interpret model predictions for what they are. It is interesting to note that also in  $^{13}\text{C}$ -MFA [64] and FBA [65] there is great interest in Monte Carlo Simulation for its ability to handle uncertainties as well as large systems.

## 1.5 Omics meets systems biology and bioinformatics

Many of the kinetic modelling endeavours in the past have aimed to explore *possible* biological behaviours by varying parameters and equations. However, in the post-genomic era it is of crucial importance to also model *what actually happens* [66]. The large datasets obtained from omics present complex profiles and systems biology is needed to understand it better. Such models must accordingly be made at a scope and level of detail that is similar to that of the omics datasets. Both constraint-based FBA and the more complex kinetic modelling will have to be streamlined towards such a regime. Constraint-based modelling is already at the level of the omics in terms of scale, but lacks in detailed predictions. Kinetic modelling with MCA is very detailed and can now incorporate uncertainties in parameters, but lacks in scale and thus practicality. The detailed understanding of biochemical systems that large-scale kinetic modelling can provide is attractive, therefore our long-term goal is the creation of an *in silico* replica of a cell, represented by the Silicon Cell Initiative [66]. Metabolism is our starting point, but the genome-scale description will eventually take all levels of regulation into

account. We want to test the Silicon Cell under many different conditions and manipulations, therefore the rate equations must have context independence [66], the same way that an organism uses the same cellular machinery at different expression levels under many different conditions. From this viewpoint it thus makes sense that detailed enzyme kinetics (in the case of metabolic reactions) is preferable to phenomenological rate equations which do not represent any cellular mechanism.

The need for parts-based information also implies a number of other considerations. It is true that in this way the information of the parts collected during the past century can be used in an appropriate manner. However, more important is to consider that the face of biological science is changing rapidly. The recent past has shown us how biological information has been made accessible in a structured way in the form of genomic databases. Information became linked to individual genes in genomics databases as information hubs. The scientific community realized that the concept of a gene is not as constant as initially thought, and that the gene might be far from the level of functional phenotype, as a result of the hierarchical levels of regulation. Functional genomics has accordingly shifted its frame of reference somewhat towards proteins, where protein databases are now the format for linking biological information. In terms of building a parts-based Silicon Cell, this is very convenient for us. It is remarkable to note how the UniProt database has arisen as the authoritative resource for protein information [67–70]. Via UniProt, many diverse types of information from many resources are made accessible and structured around individual proteins. UniProt is continuously updated and more types of information are made available, while the process is curated by a large team of biologists. The diversity of biological databases has for long made multifaceted studies very complicated. Accordingly, a project that needed diverse data types from many resources had to be both very computer informatics intensive and required scientific knowledge of various fields, and thus only a large team could execute a multifaceted project successfully. The aim of the UniProt Consortium is to enable the individual researcher the luxury of diverse data types, structured around proteins. It is likely that all of

the most important protein-related information in databases will soon be available through UniProt. Such an initiative means that one does not need to be a bioinformatics expert to acquire diverse parts-based knowledge.

One of the recent additions to UniProt is enzyme kinetic information, which could be used directly in kinetic modelling, as is also found in the Brenda [71], CCDB [72] and SABIO-RK databases [73]. Other diverse annotations in UniProt include, for instance, associations with diseases. At first glance, the latter type of information may not be very useful from a kinetic modelling perspective, for it is not a recognizable parameter in kinetic models. Neither could diseases be completely described in terms of the function of a single protein. However, working with such information is important to keep a strong link with reality so that we can address real-world problems. We should thus make it an aim in the Silicon Cell Initiative to implement elements from bioinformatics that will help us to make systems biology practical and address real-world problems. The first genome-scale Silicon Cell should thus not only be a model to understand how the organism *Saccharomyces cerevisiae* or *Escherichia coli* works. It has to result in a generic framework for interpreting omics datasets and parts-based knowledge, which will help us to address a multitude of biological problems.

## 1.6 Steps towards building the Silicon Cell

For the Silicon Cell Initiative, a *modular approach* has been suggested [66, 74]. In this approach, the total cellular description can ultimately be created from a number of modules which are small kinetic models built by individual researchers or research teams, specializing in those parts of the cellular puzzle. Kinetic modules that have been built so far are stored in the JWS Online model database (<http://jjj.biochem.sun.ac.za>) which is a repository for parts of a Silicon Cell. General outlines for building context-independent kinetic modules have been given elsewhere [66].

Kinetic modules should be non-overlapping entities in a global cellular network which have a well-defined connectivity structure. One would therefore

think that we first need a *roadmap* in the form of a metabolic pathway network, the likes of which have been created recently for a number of organisms, including yeast [33]. We could thus build a network of connected kinetic modules on top of these representations to guide the process. This reference network should then also be our link to the information in gene, protein and metabolite databases, forming the information link between systems biology and the principles of bioinformatics.

Different from a top-down approach that would aim to first make less-detailed descriptions of the cell and then replace parts with detailed descriptions of cellular machinery [1], our approach is to start with detailed parts-based descriptions and work towards combining them. This implies that we have to obtain detailed data of a large scope of metabolism that will function: 1) as validation data for the kinetic parts, and 2) as boundary constraints for correct simulation of kinetic parts. Figure 1.3 illustrates this flux-constrained approach.

We focus here on metabolic flux data at the steady state, where steady state metabolomics and dynamic pulse experiments can follow in the future. The need for detailed flux data implies that we use stable isotope based methods, as will be explained in Chapter 2. The use of a flux model, which is inevitably a simplification of the genome-scale representation, implies that the kinetic modules are not built directly on the genome-scale network, but are linked via the *abstraction* of the flux model (see figure 1.4). As metabolic flux technology improves (which will be enabled by advances in metabolomics), the flux models will increasingly become similar to the genome-scale stoichiometric model, but flux data is always linked via the abstraction of the quantifiable flux model. We thus have to assume that the reactions which do not feature in the flux model proceed so slowly that they can be ignored in flux calculation. Chapter 3 is a theoretical treatment of the linkage between a kinetic module and metabolic flux data.

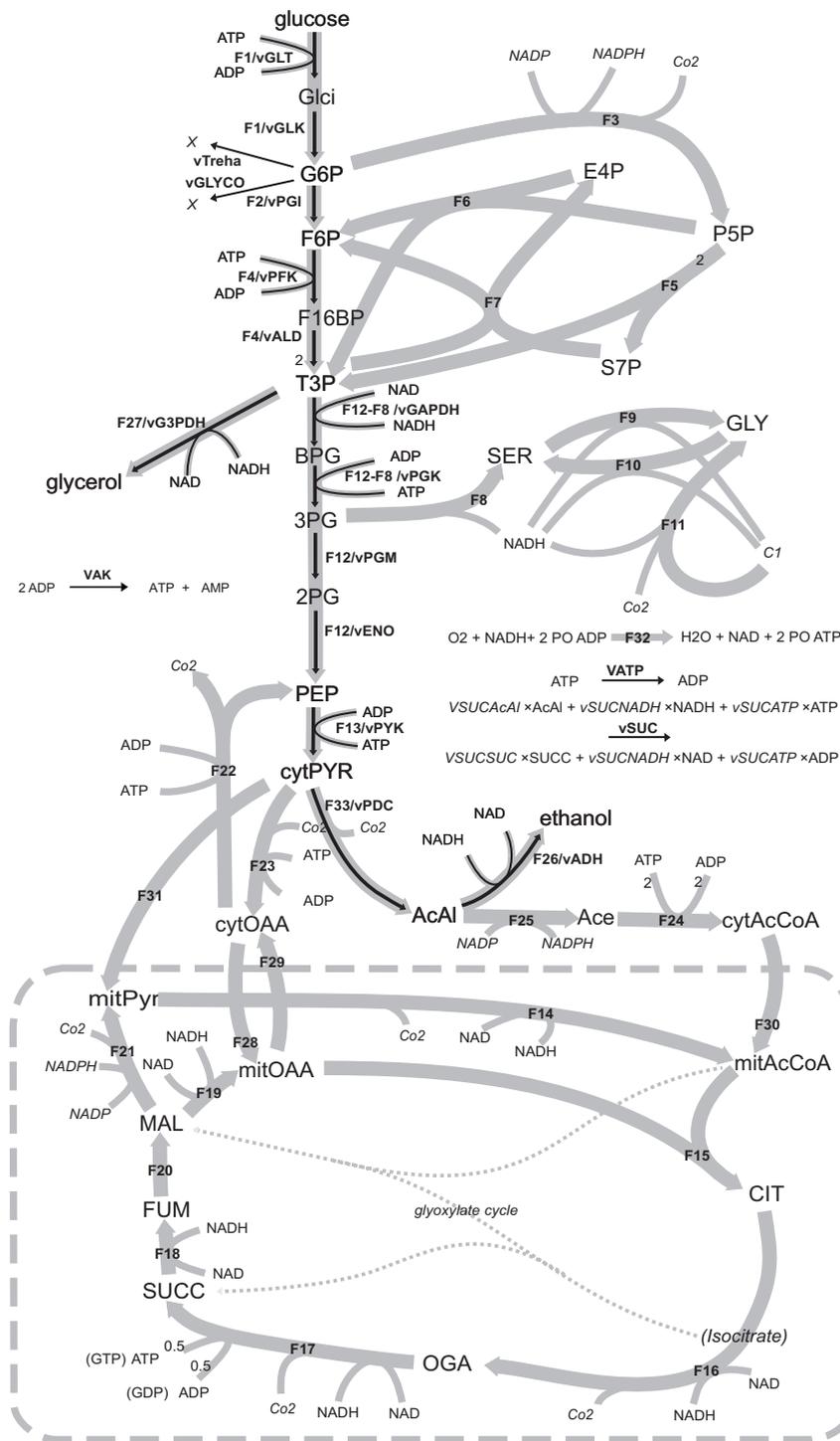


Figure 1.3: Combining a flux data model with a kinetic model. Grey arrows are measured fluxes in the flux model. Black arrows are reactions in a kinetic model. Data values are replaced with kinetics. The boundary reactions have to be incorporated correctly.

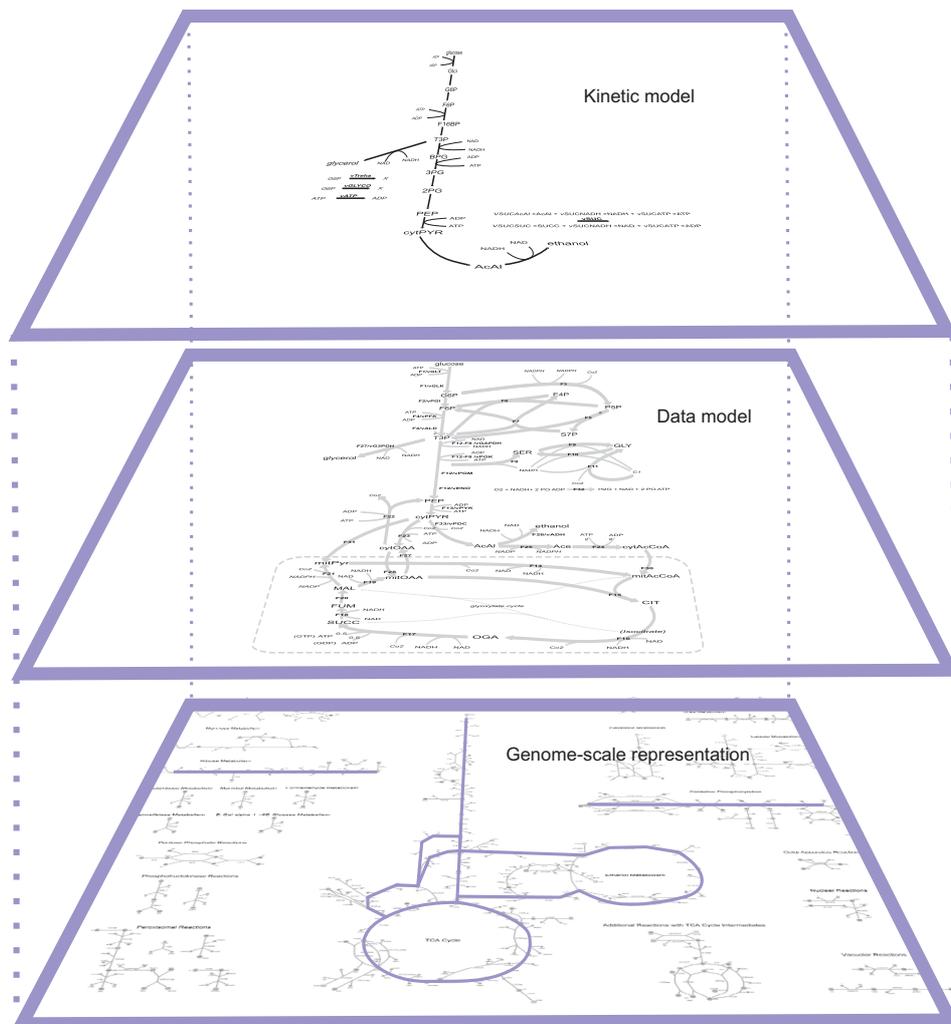


Figure 1.4: Kinetic modules are connected to the genome-scale pathway maps through data models. In this case it is a  $^{13}\text{C}$ -MFA model of central carbon metabolism which serves as boundary constraints and as basis for accuracy assessment.

## Chapter 2

# Theoretical aspects of $^{13}\text{C}$ -metabolic flux analysis

Excellent reviews on the history, experiments, mathematics and statistics behind  $^{13}\text{C}$ -MFA are given by Fernie et al. [3], Wiechert [12] and Sauer [13]. Specific mathematical concepts are described in detail in [64, 75–90], and the basis for experimental techniques are described, for nuclear magnetic resonance spectroscopy (NMR) [85, 91], for GC-MS [87, 92] and for matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) [93, 94]. Only the basic principles are explained in the next two sections.

### 2.1 Analysis of data from isotope labelling experiments

In  $^{13}\text{C}$ -MFA, cells are provided with substrates that are labelled ( $^{13}\text{C}$  as opposed to the unlabelled,  $^{12}\text{C}$  form) at certain positions and in certain percentages. The isotopic tracer atoms are then distributed in the metabolic system and by analyzing the labelling patterns of metabolic intermediates, the reaction rates through various metabolic pathways can be calculated. What determines whether a certain metabolic flux, or rather a ratio between fluxes can be calculated, is the specific labelling pattern of the substrate,

and the atom transitions in the metabolic network. Figure 2.1 gives an indication of the ways that isotope labelling information is distributed through a metabolic network.

In A, pathways J1 and J2 can be distinguished by supplying a substrate labelled in the second position (indicated by a black dot). The labelled atom is cleaved off in pathway J2, therefore it will not be present in the product. Pathway J1 cleaves off the atom in the first position, therefore its activity leads to the labelled end product. The ratio between the two pathways can thus easily be calculated as the ratio between the two isotopic forms of the end product.

In B, pathways J3 and J4 do not cleave off any of the atoms, but the atoms are rearranged only in J3. The result is that the pathways produce the same number of labelled atoms as unlabelled atoms in the end product (they produce equal *fractional labelling*). However, the labelling patterns are different. The different types of labelling patterns in a molecule are referred to as isotope isomers or isotopomers, for they are the same chemical structures, but differ only by their isotopic pattern. The flux ratio between the two pathways could thus be calculated as the ratio of these two isotopomer fractions. Note that the concentrations of these metabolites are not important, only the isotopomer fractions. Note also that for this pathway structure, we do not need to measure the isotopomer fractions in the two intermediate metabolites, for we know from the atom transitions what their isotopomer fractions will look like (if our knowledge of the pathways and carbon transitions is correct).

In C, pathways J5 and J6 neither cleave off atoms nor catalyze different carbon transitions. They result in exactly the same isotopomer fractions in the end product, which presents a problem. The only way how such a ratio could be calculated is if we would carry out an experiment in which the labelling pattern changes over time, in order to calculate a turnover rate. We would thus have to measure labelling patterns in the intermediate metabolites during a time-series in which the concentrations of the metabolites are constant (metabolic steady state) but the isotopomer fractions change over

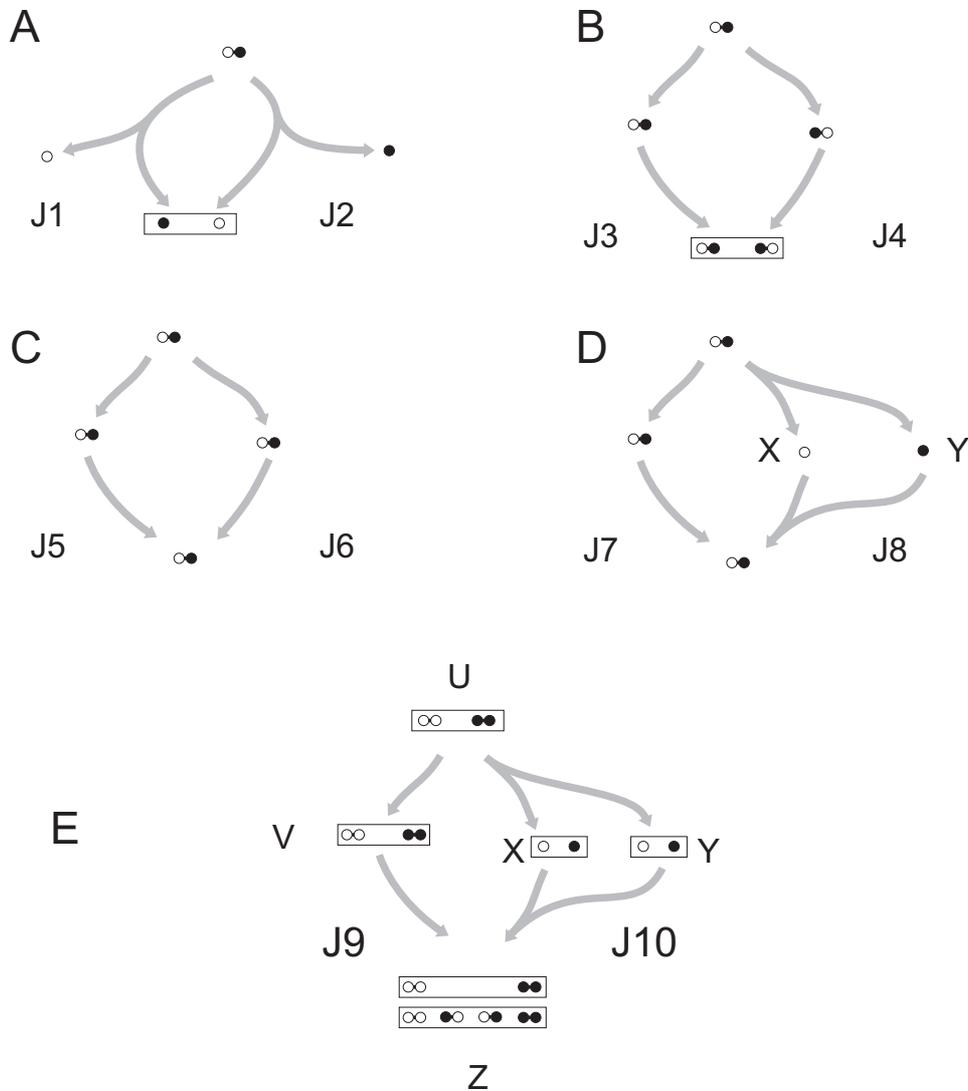


Figure 2.1: Various ways in which labelling information is generated through metabolic systems. Boxes indicate pools of a metabolite with various patterns of labelling. A. Pathways split molecules in alternative ways. B. Pathways use alternative rearrangements of atoms in molecules. C. No splitting of molecules or rearrangement of atoms. D. No splitting of molecules or rearrangement of atoms occurs, but one pathway catalyzes sequential splitting of a molecule and condensation of products. E. The scrambling effect in a condensation reaction generates information of its activity to the condensation product in a probabilistic manner.

time. Such an isotopic transient can be induced by growing the organism on the natural ( $^{12}\text{C}$ ) substrate in a chemostat, and waiting for a metabolic steady state. The substrate inflow is then switched to the labelled ( $^{13}\text{C}$ ) substrate, and the labelling patterns in the intermediate metabolites will show a labelling transient. If the concentration of a metabolite is known, as well as its labelling transient dynamics, the flux through such a pathway could be directly calculated [3, 84].

In D, pathways J7 and J8 also do not result in distinct carbon transitions which present the same problem as in C. However, this case is different due to the fact that pathway J8 splits the substrate and combines it again. At first glance, this is not an important fact, for the end product has the same labelling patterns via both pathways. However, the second part of J8 catalyzes a condensation reaction, preceded by a randomized scrambling of all the molecules in each of the pools X and Y. In this example, all substrate molecules are 100%  $^{12}\text{C}$  in the first position and 100%  $^{13}\text{C}$  in the second position, therefore all X are  $^{12}\text{C}$  and all Y are  $^{13}\text{C}$ . Chances are thus 100% that an unlabelled X and a labelled Y will combine to form 100% of  $^{12}\text{C}$ - $^{13}\text{C}$  Z molecules.

However, if the substrate would be a combination of fully unlabelled and fully labelled molecules as in E, then X and Y would each contain both  $^{12}\text{C}$  and  $^{13}\text{C}$  molecules. When they are combined again in the bottom part of J10, combination by chance will result in isotopomer fractions of Z corresponding to  $^{12}\text{C}$ - $^{12}\text{C}$ ,  $^{13}\text{C}$ - $^{12}\text{C}$ ,  $^{12}\text{C}$ - $^{13}\text{C}$  and  $^{13}\text{C}$ - $^{13}\text{C}$  patterns. This is opposed to the  $^{12}\text{C}$ - $^{12}\text{C}$  and  $^{13}\text{C}$ - $^{13}\text{C}$  isotopomer fractions that are produced by pathway J9 that does not split and recombine the molecules. Such information can be handled conveniently in terms of isotopomer distribution vectors (IDV's) and probability theory. An IDV is the vector of isotopomer fractions of a metabolite pool. The elements can have values between 0 and 1 and they add up to 1. The number of isotopomer fractions  $n$  in an IDV can be calculated from the number of atoms  $a$  as

$$n = 2^a \tag{2.1}$$

In a condensation reaction with two substrates and a product, there is a probabilistic relationship between the isotopomer fractions of the substrates and the product given by

$$IDV_Z = IDV_X \otimes IDV_Y \quad (2.2)$$

where IDV is the isotopomer distribution vector of the product ( $IDV_Z$ ) and substrates ( $IDV_X$ ,  $IDV_Y$ ) and ( $\otimes$ ) is an elementwise multiplication.  $IDV_Z$  thus has the number of

$$n_Z = n_X \times n_Y \quad (2.3)$$

elements. In this way, the IDV of any condensation product can be calculated from measured or calculated IDV's. Similarly, if the IDV's of products are known, the IDV's of their substrates can be calculated by least-squares fitting. These two calculations, along with techniques for correcting for natural abundances of isotopes [86, 90] are the basis for the derivation of the labelling patterns of metabolites via other metabolites. For example, in the retrobiosynthetic or biosynthetic fractional labelling (BFL) approach [85], the labelling patterns of metabolic intermediates are derived via labelling patterns in the amino acids, which are synthesized from these precursors and are much more abundant and stable.

In a direct probabilistic approach to  $^{13}\text{C}$ -MFA, flux ratios can be calculated based on measured or derived isotopomer fractions. A very effective method is Flux Ratio Analysis or METAFoR analysis [85, 86, 88]. For scheme E, the IDV of the product  $IDV_Z$  could be synthesized firstly through pathway J9 from pathway intermediate V. From the atom transitions it is known that  $IDV_V$  will have the same labelling pattern as the growth substrate with known  $IDV_U$ . Alternatively, the product could be synthesized through pathway J10. We can calculate the labelling pattern in its condensation product (call it  $IDV_{XY}$ ) from the split and recombined substrate  $IDV_U$ . X and Y are fragments of the substrate U isotopomers, therefore  $IDV_X$  and  $IDV_Y$  can easily be determined.  $IDV_X$  and  $IDV_Y$  can then be used to obtain  $IDV_{XY}$  by equation 2.2. The contributions of the two pathways can thus be calculated as follows:  $IDV_Z$  is produced from  $IDV_U$  and  $IDV_{XY}$  via the

relationship

$$IDV_Z = f \times IDV_{XY} + (1 - f) \times IDV_U \quad (2.4)$$

where  $f$  is the ratio of J10/(J10+J9), which can be rearranged to yield

$$f = \frac{IDV_Z - IDV_U}{IDV_{XY} - IDV_U} \quad (2.5)$$

In practice, the methods by which isotopomers are observed, sometimes do not allow the full isotopomer distributions to be calculated, but derivations from it. For instance, MS allows assignment of mass distribution vectors (MDV's). In an MDV, all the isotopomer fractions with the same number of labelled atoms are grouped in the same element, as they are observed in the same signal in MS (see figure 2.2). For a molecule with  $n$  atoms, there exists  $n+1$  mass isotopomers in the MDV. There is thus a reduction in information potential from  $2^n$  isotopomers to  $n+1$  mass isotopomers. Fortunately, we can regain some of the hidden information by using fragmentation methods in MS. For each fragment an MDV can thus be assigned which provides further constraints. For mass spectrometry data, the IDV's in all of the equations above are thus replaced by the MDV's as explained in [86]. An MDV can be assigned for each metabolite fragment in an MS and ratios are thus based on MDV's of fragments using equation 2.5.

Mass isotopomer signals can be seen for two metabolites in a GC-MS, operated in electron impact ionization mode in figure 2.3. In figure 2.4, the fragmentation patterns of the same metabolites become visible, where each fragment cluster consists of different mass isotopomer signals. Figure 2.5 gives a global view of the elution pattern during the first few minutes of elution showing the complex fragmentation patterns.

For each ratio calculated, a standard deviation can be assigned. This is based on an estimation of experimental error of the mass spectra, by comparing the amino acid MDV's with identical carbon skeletons. This error is used for calculating the standard deviation of the ratio by the law of error propagation. For a discussion of error estimation of ratios, see [86].

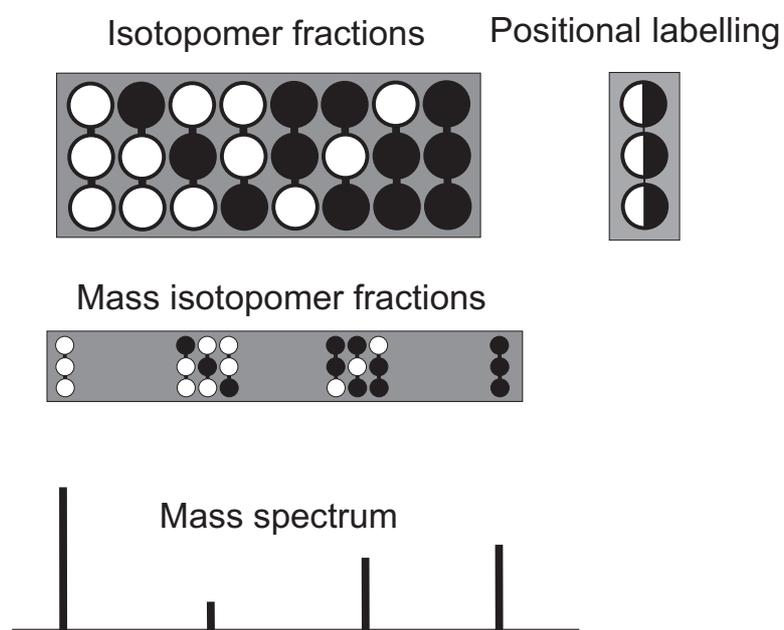


Figure 2.2: The labelling pattern of a metabolite pool can be described in terms of isotopomer fractions. Mass spectrometry data provides information of mass isotopomers, which are combinations of isotopomers that have the same mass. Positional labelling is the fraction of all atoms at a certain position in a molecule that are labelled.

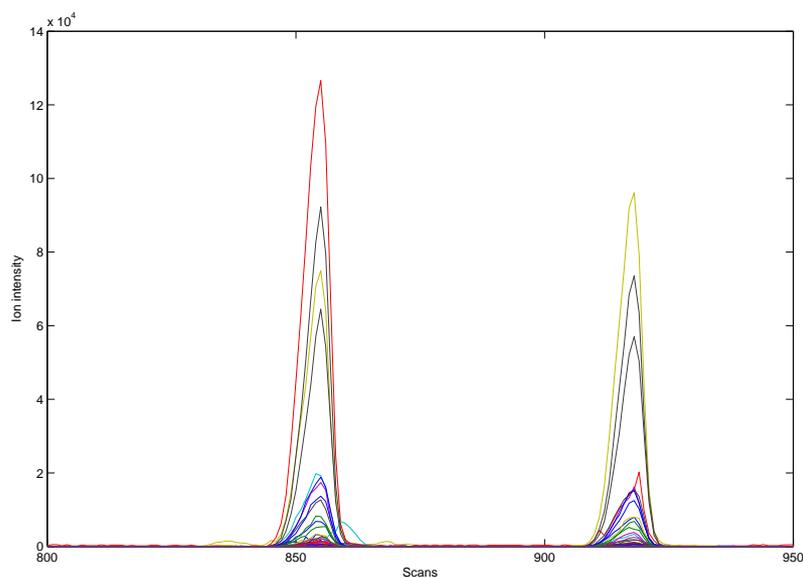


Figure 2.3: 2-Dimensional view of mass isotopomer signals in a GC-MS with a quadrupole mass spectrometer operated in electron impact ionization mode. The different lines in one peak (metabolite) are different mass isotopomer ions that elute from the GC over time. The two metabolites are alanine (left) and glycine (right). Data of hydrolyzed  $^{13}\text{C}$  glucose grown yeast biomass was obtained in this study. Details of the GC-MS procedures are given in section 5.1.5.

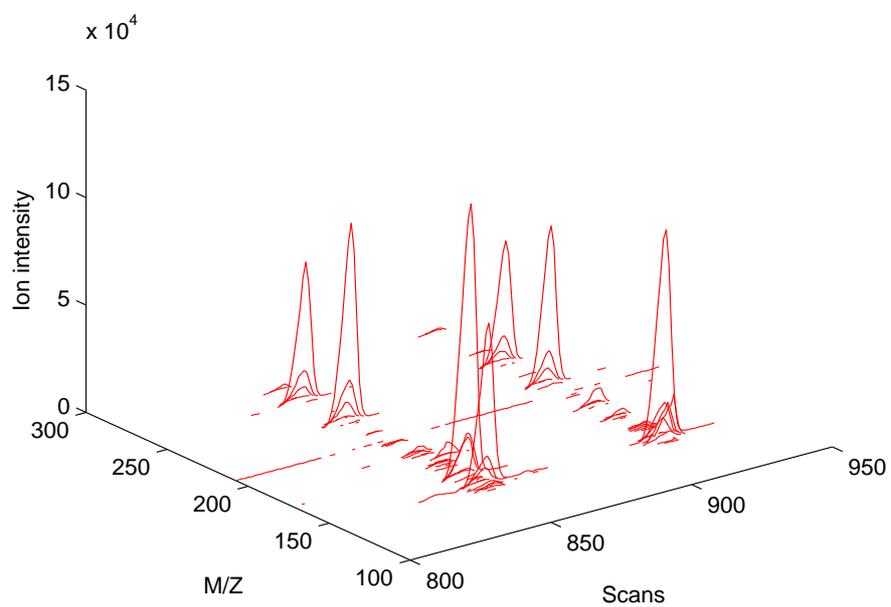


Figure 2.4: 3-Dimensional view of mass isotopomer signals in GC-MS of the hydrolyzed  $^{13}\text{C}$  glucose grown yeast biomass. Fragments of alanine (left) and glycine (right) become visible as ion clusters, each consisting of different mass isotopomers.

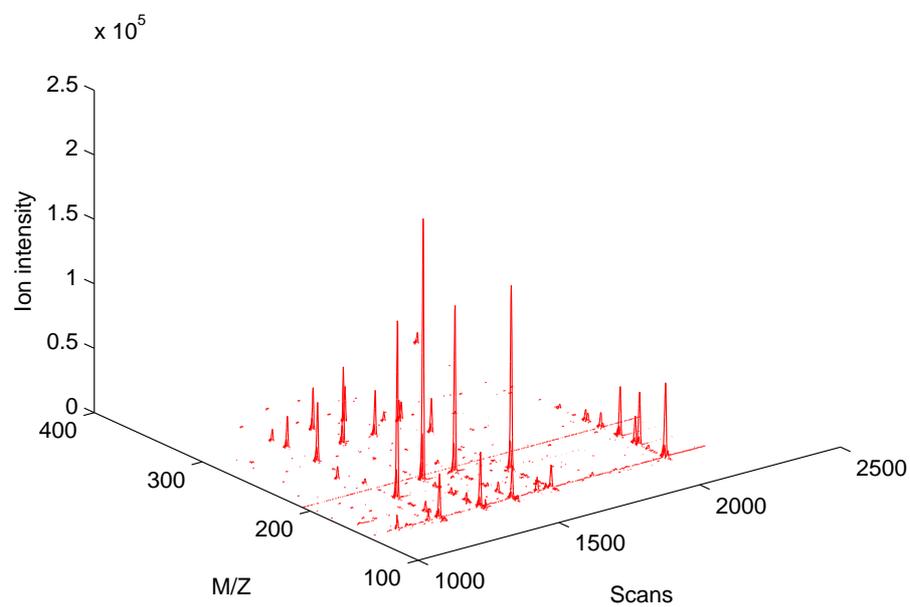


Figure 2.5: Mass isotopomer signals in GC-MS during the first few minutes of elution of the hydrolyzed  $^{13}\text{C}$  glucose grown yeast biomass. Many fragment ions are seen for each metabolite.

Flux ratio  $f$

$$f = \frac{J_{10}}{J_{10} + J_9} \quad (2.6)$$

is a constraint between fluxes at the steady state, and can be incorporated together with the stoichiometry based metabolite balance constraints of flux analysis. This equation can be rewritten as

$$0 = (1 - f)J_{10} - (f)J_9 \quad (2.7)$$

The latter equation has the same form as the rows of metabolite balances in stoichiometric flux analysis. However, for some ratios, only upper and lower limits can be assigned arising from the fact that labelling information of some metabolites is not available, and assumptions have to be made. Such ratio limits are of the forms

$$l \leq \frac{J_a}{J_a + J_b} \quad (2.8)$$

and

$$u \geq \frac{J_a}{J_a + J_b} \quad (2.9)$$

where  $l$  is a lower limit and  $u$  is an upper limit. These inequalities cannot be used in the conventional matrix based calculations of metabolite balancing, and numerical search routines have to be used. In  $^{13}\text{C}$ -constrained flux balancing [87], the equality constraint

$$0 = (1 - f)J_{10} - (f)J_9 \quad (2.10)$$

is implemented in the search routine as

$$R = (1 - f)J_{10} - (f)J_9 \quad (2.11)$$

together with metabolite balances, measured extracellular fluxes and inequality constraints.  $R$  is the residual of the constraint. The routine seeks to minimize the sum of the squared residuals, each divided by a measurement error. These measurement errors are in the form of variances (squared standard deviations) from the ratios and the extracellular metabolite production and consumption rates. For a detailed explanation of the routine and the statistical methods, see [87]. In cases where the reaction structure around

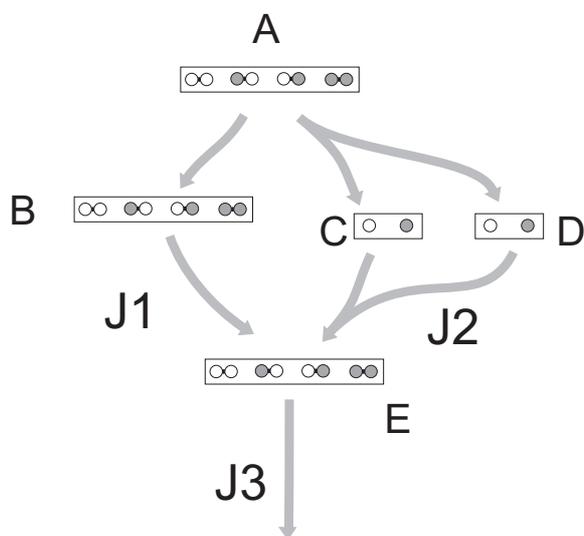


Figure 2.6: An isotopomer labelling system. For each isotopomer in the system a mass balance equation is written that describes its production rates minus its consumption rates.

a certain metabolite is very certain, the user specifies that the metabolite should be balanced. The algorithm incorporates its residual value  $R$  and seeks to minimize the scaled, squared sum of  $R$  values. Most of the metabolites will be balanced in this manner. The smaller these residuals at the end of minimization, the more certain we can be about the network model used. However, the reaction structures around cofactors such as NADH, NADPH and ATP are less certain than most other metabolites, and it is thus attempted to leave their residuals as variables, representing their net production or consumption fluxes. The algorithm does not incorporate these balances in the minimization. In our study, the FiatFlux software program [95] was used which is based on METAFoR analysis with GC-MS data and  $^{13}\text{C}$ -constrained flux balancing for net fluxes.

Another approach to calculation of metabolic fluxes is to create an isotopomer labelling system (ILS), which is a model of all the isotopomers in a metabolic system. In this approach, each isotopomer is balanced by its production and consumption fluxes. A detailed explanation of ILS's has been given in [79] and [80]. For the case of our example in figure 2.6, the mass balance equations for the isotopomers of only metabolite E, in an ILS

will look as follows:

$$\frac{d e_{00}}{dt}E = (b_{00})J1 + (c_0 \cdot d_0)J2 - (e_{00})J3 \quad (2.12)$$

$$\frac{d e_{10}}{dt}E = (b_{01})J1 + (c_0 \cdot d_1)J2 - (e_{10})J3 \quad (2.13)$$

$$\frac{d e_{01}}{dt}E = (b_{10})J1 + (c_1 \cdot d_0)J2 - (e_{01})J3 \quad (2.14)$$

$$\frac{d e_{11}}{dt}E = (b_{11})J1 + (c_1 \cdot d_1)J2 - (e_{11})J3 \quad (2.15)$$

The upper case name indicates the metabolite concentration of E and the lowercase of the same name with the binary digit subscripts indicates its isotopomer fractions. The digits represent the carbon atoms in the molecule. A 1 in the subscript indicates a  $^{13}\text{C}$  and a 0 indicates a  $^{12}\text{C}$ . For metabolic and isotopic steady state, the fractions of the isotopomers  $e_{00}$ ,  $e_{10}$ ,  $e_{01}$ ,  $e_{11}$  will not change over time. The balances for all the isotopomers of E are thus equal to zero. The assumption of isotopic steady state thus makes the concentration of E irrelevant. Further, because isotopomer fractions (and mass isotopomers) sum up to 1, the number of constraints (equations) is reduced by 1, for the last fraction can always be calculated from the other fractions. It can be seen that the probabilities of formation of isotopomer fractions in E is modelled as multiplication terms of the relevant fractions of  $c$  and  $d$ . These multiplication terms give ILS's their nonlinear characteristics, as opposed to the linear constraints of metabolite balances. Atom transitions, or rather isotopomer transitions are mapped between molecules with isotopomer mapping matrices or isotopomer transition matrices (see [79]). By now it is also evident that the number of atoms considered drastically increases the number of possible isotopomers of a molecule. There are  $2^n$  isotopomers for each molecule with  $n$  atoms which means that for a molecule of 10 carbon atoms to consider, 1024 balance equations have to be included in the model. Flux calculation with ILS's thus presents a strong need for super computing as the simulation procedure is computationally expensive, and is especially dependent on computer memory. A bigger problem is that large simultaneous forward and reverse fluxes (exchange fluxes) cause convergence problems with most of the algorithms designed thus far. The solution came with the introduction of cumomer labelling systems (CLS's) and conversion

of cumomer fractions to isotopomer fractions. A cumomer fraction is the fraction of a part of a metabolite that is labelled in a specific way. That is, regardless of the labelling of the rest of the metabolite (see [79] for an explanation of cumomers). This treatment allows matrix-based solution of isotopomer fractions as a function of flux inputs. The numerical search routine seeks to minimize the discrepancy between the simulated isotopomer fractions and measured isotopomer fractions. The flux inputs that yield closest isotopomer fraction comparisons are then taken as the best values.

Apart from the representation as isotopomers, there is also the positional representation, which is actually the more conventional formulation of a labelling system (see [76–78]). Positional labelling can be explained as the percentage of all atoms at a specific position in the molecule that are labelled and values can range from 0 to 1 (see figure 2.2). There are thus  $n$  elements in the positional labelling vector of a metabolite with  $n$  atoms and  $n$  balance equations. Accordingly, the size of the labelling system does not expand exponentially with the number of atoms in each molecule, which makes positional labelling systems computationally less expensive. However, large exchange fluxes cause the same conversion problems as with ILS's.

## 2.2 Detection of labelling patterns by nuclear magnetic resonance spectroscopy and mass spectrometry

The isotopomer distribution is the most detailed form of measurement that one can make of the labelling state and is more informative than positional labelling. Isotopomer fractions can be observed by two methods, namely MS and NMR.

NMR became particularly useful in flux analysis in two-dimensional heteronuclear single-quantum coherence [ $^{13}\text{C}$ ,  $^1\text{H}$ ] correlation spectroscopy (2D- $^{13}\text{C}$ ,  $^1\text{H}$ ]-COSY) of amino acids [85, 88]. 2D- $^{13}\text{C}$ ,  $^1\text{H}$ ]-COSY detects intact  $^{13}\text{C}$ - $^{13}\text{C}$ -fragments, therefore the experiments are usually conducted with a single substrate such as glucose which is uniformly  $^{13}\text{C}$ -labelled (at all car-

bon positions) in a certain percentage, usually 10 or 20 percent. The ability of 2D- $^{13}\text{C}$ ,  $^1\text{H}$ -COSY to quantify linear combinations of isotopomer fractions, together with its sensitivity and efficient resolution of amino acids, makes it suitable to study the central carbon metabolism. Although less sensitive than MS and thus smaller in scope, NMR can give very precise information of the positions of the heavy isotopes. It does not mean however that the scope of NMR cannot be increased [3]. A problem with  $^{13}\text{C}$  NMR methods is that it can only detect the  $^{13}\text{C}$  isotope and the  $^{12}\text{C}$  isotope is not detected, thus many isotopomer fractions are not accessible. Linear combinations of isotopomers are detected that have the same  $^{13}\text{C}$ ,  $^{13}\text{C}$ - $^{13}\text{C}$  and  $^{13}\text{C}$ - $^{13}\text{C}$ - $^{13}\text{C}$  labelling pattern, but not all fractions are detected. Additionally, some  $^{13}\text{C}$  isotopomers cannot be resolved because they have the same resonance. A new approach is to use heteronuclear spin echo difference NMR spectroscopy. This proton NMR method can be used to indirectly observe both the  $^{13}\text{C}$  and  $^{12}\text{C}$  atoms to calculate even full isotopomer distributions in some cases [91].

MS has become very practical recently and its extraordinary sensitivity, efficient coupling to pre-separation techniques, rapid experimental analysis and the small amounts of sample necessary, makes it more applicable for high-throughput data acquisition than NMR. Gas-chromatography coupled to mass spectrometry (GC-MS) is a very fast and practical method for obtaining labelling patterns in microbes and experiments have been miniaturized effectively to culture volumes as small as 1.2 mL for subsequent GC-MS analysis [87,92,96,97]. As was mentioned earlier, the retro-biosynthetic approach is the standard practice for  $^{13}\text{C}$ -MFA, wherein the labelling patterns of metabolic intermediates are accessed via the labelling patterns of the monomers of biopolymers, which are synthesized from the intermediates. For GC-MS, amino acids are freed from their biopolymers by hydrolysis. In order to make the amino acids sufficiently volatile to evaporate into the GC column, the hydrolyzed cellular matter containing the amino acids is treated with a derivatization agent such as N-(tert-butyldimethylsilyl)-N-methyltrifluoroacetamide (MTBSTFA) and a suitable organic solvent such as dimethylformamide (DMF). The derivatized amino acids are separated

from other cellular components in the GC column and electron impact ionization (EI) is applied that fragments the derivatized compounds. Where information in MS is limited by the fact that mass isotopomers are observed and not isotopomers, fragmentation provides information that resembles isotopomer fractions closer. Here, different derivatization agents may also produce alternative fragmentation patterns.

MALDI-TOF MS, which is very new to  $^{13}\text{C}$ -MFA is a very rapid and sensitive analytical tool for the analysis of external metabolites or amino acids. Analyses can be carried out without derivatization, but derivatization adds to the scope of metabolites that are observed. Culture supernatant samples as small as 1  $\mu\text{L}$  have been used for detection of labelling patterns [93,94].

From a practical perspective, the local and ILS approaches each have advantages and drawbacks. While the directed approach of flux ratio calculation works very well for calculation of a few fluxes, analysing a new metabolic system can be intimidating. The modeller needs to know all the theory behind the method in order to model explicitly the ratios directly from labelling data. The ILS approach is more generic and thus more usable for many different pathway structures. It is also more suited to extension to different situations, such as dynamic flux responses in which the concentrations, fluxes and labelling patterns change over time. The ILS programs are readily made to take network structure inputs in the form of reactions. Atom transitions are specified as strings of letters in both the substrate and the product, where the letter (atom) is transferred to the substrate at a certain position in the string. The mapping matrices or transition matrices are then automatically generated from these.

It must be said in this regard that it is easier to create such a user friendly interface in the ILS approach because of its generic nature, but it is still only a feature of the software, and not of the actual analytical method. The features that make the available ILS programs user friendly could also be applied in programs for the local approach, in a modified form. This will have to feature in programs to enable genome-scale analyses.

Computational complexity also has some practical implications. Large systems, and especially large molecules can make ILS simulations prohibitively computationally expensive, whereas the local approach does not suffer from this problem. In this regard, the group of Wiechert are implementing parallelization of the flux algorithms [98].

From a fundamental perspective, there are also advantages and drawbacks, although the two methods will ultimately provide the same answers if used correctly. ILS simulation is an integrated approach which means that it is model dependent. If there are mistakes in a network model used, the complete flux distribution will be calculated wrongly, while the strictly local flux ratio approach is better suited to validate a network structure. Genome-scale flux analysis will likely require a combined approach.

## Chapter 3

# Flux-constrained kinetic modelling: dealing with parts of a Silicon Cell

Kinetic modelling, like other forms of systems biology, now needs to embrace high-information technologies for creating and validating models. A rigorous methodology for this integration has to be developed in which kinetics can be quantitatively compared to physiological data to locate dubious parameters. There is however more to this integration process than a mere comparison of model predictions with data. For every kinetic model built, a large part of the cellular network is assumed to be non-existent.

In the first section of this chapter it is explained why it is necessary to include flux measurements of the rest of metabolism into a kinetic simulation. Only then can the existing kinetic parts be expected to portray realistic behaviour. This involves a different treatment of the system boundary conditions.

The new inclusion of boundary reactions involves a number of other implications for accuracy assessment. Establishment of extensive accuracy assessment is necessary before increasing the size of a kinetic model, and before predictions can be made from it. It is discussed that the integrated nature of kinetic models makes it difficult to locate errors, especially with only partial, or no metabolite concentration data. Section 3.2.1 explains

how the formulation of kinetic models can be altered for more stringent accuracy assessment by setting of pathway fluxes to measured values. A special type of data that we investigate for this integration and accuracy assessment process is that of  $^{13}\text{C}$ -MFA experiments. In section 3.2.2 it is shown to have interesting properties which can be used as an *in vivo* method for partial accuracy assessment.

The flux-constrained methodology introduced in this chapter involves a number of technicalities to be considered. Setting of pathway fluxes and conserved moieties hold implications which are discussed in sections 3.3.1 and 3.3.2 respectively.

Steady state data is limited in its information content and thus not very stringent as a data source for accuracy assessment. Dynamic substrate pulse experiments have high potential for accuracy assessment because they link many data points to kinetic parameters. However, these experiments suffer from a number of problems to be discussed in section 3.4. It is suggested that such experiments should be complemented by a steady state accuracy assessment protocol using the principles explained in sections 3.1 to 3.3.

Predictive capability, such as MCA, is altered, but in section 3.5 it is explained how a recent MCA framework may be used to make prediction possible.

In response to these findings, a simple prototype integration software program was written in Mathematica, which is the topic of the next chapter. This program can now serve as part of a comprehensive protocol for accuracy assessment and model improvement.

### 3.1 Boundary conditions in kinetic models

The modular approach for the Silicon Cell, in which many kinetic models are ultimately combined to form a whole, was discussed in Chapter 1. A main feature of the modular approach should be the existence of a very

large reference biochemical network which is extensively characterized with reaction and concentration data from both steady state and dynamic experiments. The structures of kinetic modules must correspond to parts in the data model. It is obvious that some measured reactions will correspond to the reactions in the kinetic model and others will fall outside of the kinetic model. Those data reactions that have a complement in the kinetic model will in some way be used for comparison with the simulation outputs of the kinetic model in the validation process. As biological pathway networks are in fact integrated systems, each kinetic part must be fully integrated with the rest of the system, either through other models or through *boundary conditions*. Due to the fact that we do not have models for the whole system, boundary conditions are used which have values that have been measured *in vivo*. All entities that are not treated as variables during a simulation, which are expected to have different values at each physiological condition are boundary conditions. They have to be measured for each such condition and used as numerical inputs in kinetic models in order to simulate the system.

The boundary conditions frequently used in detailed enzyme kinetic models are: 1) metabolite concentration boundaries and 2) maximal activities of enzymes. Entities which we assume that we can extrapolate to all conditions include enzyme parameters such as saturation, inhibition, activation and cooperativity constants, which can be measured *in vitro*. Factors such as pH and temperature also alter the behaviour of enzymes, but usually these effects are minimal over the range of conditions that a model is used for. These effects usually do not feature in the rate equations and are thus implicit in the other parameters.

### 3.1.1 Concentration boundaries

Concentration boundaries are the concentrations of metabolites that affect the modelled pathways and which are not treated as variables in a simulation. These include direct products and substrates of the pathways of interest, for instance, extracellular glucose and glycerol, and are featured in the

rate equations. They are generally treated as constant over time for steady state modelling or can equally validly be incorporated as equations that describe the measured concentrations as a function of time. For instance, if a system of rate equations is tested against time-dependent data from a dynamic substrate pulse experiment, this will be a more valid procedure than to make the assumption that a boundary metabolite stays constant. Also, the rate equations can feature concentrations of other effectors influencing them which are not synthesized or produced by the local system, but are variables of other parts of the larger system. Typically, these are activators and inhibitors of the modelled system and do not feature in its stoichiometry. They too can be treated as constants for steady state modelling or as time dependent equations.

The concentration boundaries, which have their data analogue in *metabolomics*, are thus relatively simple to incorporate: the relevant concentrations are simply used as constants or time-dependent values in the rate equations. Although there are difficulties in the technique of metabolomics itself, like identification of peaks, conversion to concentrations, sampling, derivatization, preservation and statistical considerations of combining different data sets, the final answer is a concentration which should directly relate to another concentration in the kinetic model.

### 3.1.2 Enzyme boundaries

The second type of boundary conditions referred to here is the maximal activities of the enzymes. Usually they are referred to as parameters as they are almost always used as constants. However, as the Silicon Cell will ultimately include higher levels of regulation, the enzyme concentrations will vary at different conditions, thus making them variables. They can be called boundary conditions here as they have to be measured for each condition that the model is used, before higher regulation is incorporated.

### 3.1.3 Branch reactions as boundaries

In our treatment, a third type of boundary conditions is the branch reactions, although they are not classically thought of as boundary conditions in modelling terms. These are the reactions of pathways which were considered to be unimportant at the time of model construction. They are usually excluded from the modelled network because they were assumed to have negligible reaction rates. However, the genome-scale Silicon Cell must eventually have all these smaller branch pathways also included as modules. Further, biochemical systems form extensively integrated networks. Only if the branch reactions that stem from a modelled pathway are included can the possibly accurate rate equations be expected to simulate the realistic situation. Put differently, by wrongly assuming zero branch reaction rates, a kinetic model is constrained to a linear pathway flux and both the sizes of the reaction rates and metabolite concentrations cannot represent reality if the enzyme parameters were *correct*. So firstly, it is necessary to incorporate branch pathways in order for rigorous validation to be a possibility.

How should these branch reactions be modelled? For accuracy assessment, the aim is to simulate precisely the same effect that the rest of the network has on the parts corresponding to the kinetic model. For the steady state it is thus sufficient to simulate only those branch reactions that directly consume or produce metabolite variables in the kinetic model. These can thus be the constant measured flux values as obtained by  $^{13}\text{C}$ -MFA. Let us call this set of reactions the *boundary reactions*. Similarly, for time-dependent data, the time-dependent boundary reaction data (if it becomes possible to calculate them from time-dependent metabolomics data from a pulse experiment) could be included as an equation describing the data as a function of time. By simulating the correct boundary reactions, any discrepancies in simulated values to that of the data can be related only to the errors in the kinetic parts. The boundary reactions can thus neither be the *cause* of discrepancies between simulation and data, nor *compensate* for errors in kinetic parts.

It is even a possibility that insufficient measurement techniques could cause

false negative validation results, leading to the conclusion that *in vitro* enzyme measurements are erroneous, while they may be an accurate representation of the real cellular machinery. As our aim is to create a highly realistic cell *in silico*, its validation should be treated in a quantitative manner and in essence, that we adopt the view that a precise, or nearly precise description from parts *is* possible. Thus, the small size of some of the branch fluxes should not be a key to excluded them from the simulation, as it is indeed a precisely accurate simulation that is sought after. In this way, errors in parameters could possibly be made detectible and corrected through more experimentation.

As accuracy assessment of models should precede the making of predictions from them, and as the requirements are less for the purpose of validation than for making predictions such as in MCA, considerations for prediction are discussed in a later section.

## 3.2 Simulation for accuracy assessment

This section deals with matters of establishing an optimal protocol for validation based on steady state data. The first part describes a methodology to separate a model into smaller parts for local validation, and moving the validation criterion from fluxes towards metabolite concentrations. The second part is dedicated to exploitation of the special properties of  $^{13}\text{C}$ -MFA data.

### 3.2.1 Setting of steady state pathway fluxes for validation

The fact that a bottom-up approach for detailed kinetic modelling does not use parameter fitting, implies that the simulated steady state is not necessarily at the measured steady state values. Incorporation of boundary reactions can help to put the realistic constraints on the network and enable the correct simulation for location of the errors in parameters. Upon mathematical integration of the system of differential equations, one error

will cause the whole system to predict the wrong steady state, even if all the other equations and parameters were correct. The integrated nature of a system of ODE's (see figure 3.1 A) thus makes location of the errors difficult. By separation of the individual reactions and substituting variable concentrations with measurements (see figure 3.1 B), the reaction rates can be simulated individually and a closer relation exists between measured concentrations, fluxes and enzyme kinetic parameters. This local approach was also used for validation of the Teusink model [44]. The idea of separating a complex model into smaller subparts was also followed for fitting parameters to time series data in [52].

Unfortunately, metabolite concentrations are not all measurable, and the current state of metabolomics is less quantitative than we would hope for. Accordingly, for steady state validation, we have to use small integrated systems of kinetics wherein their boundary concentrations are available and of good quality. In this way, there are concentration variables in the small integrated systems. As those concentrations could not be measured as accurately as the boundary concentrations, assessment based on these concentrations is limited, but still useful. Concentrations could be estimated from metabolomics data or from literature for less stringent validation, and the predicted metabolic flux is also usable, which can likely be measured with more confidence (see figure 3.1 C).

However, due to the problem that an integrated system is still used, even a single error in a parameter will cause the fluxes and the concentration predictions to be wrong. Further, validation upon flux data involves redundancies, much more than with concentration data. The linear constraints between fluxes at the steady state implicates that, for instance, in a linear chain of reactions, there is only one flux value for comparison against many simulated values. Also, predicted fluxes can in principle only vary between the minimum forward  $V_{max}$  and minimum reverse  $V_{max}$  in any linear reaction chain. On the contrary, predicted concentrations can range from (close to) zero to infinity, which in principle makes it more likely to detect errors in parameters based on concentrations. For steady state validation, the ideal would thus be to move the validation criteria for small integrated systems

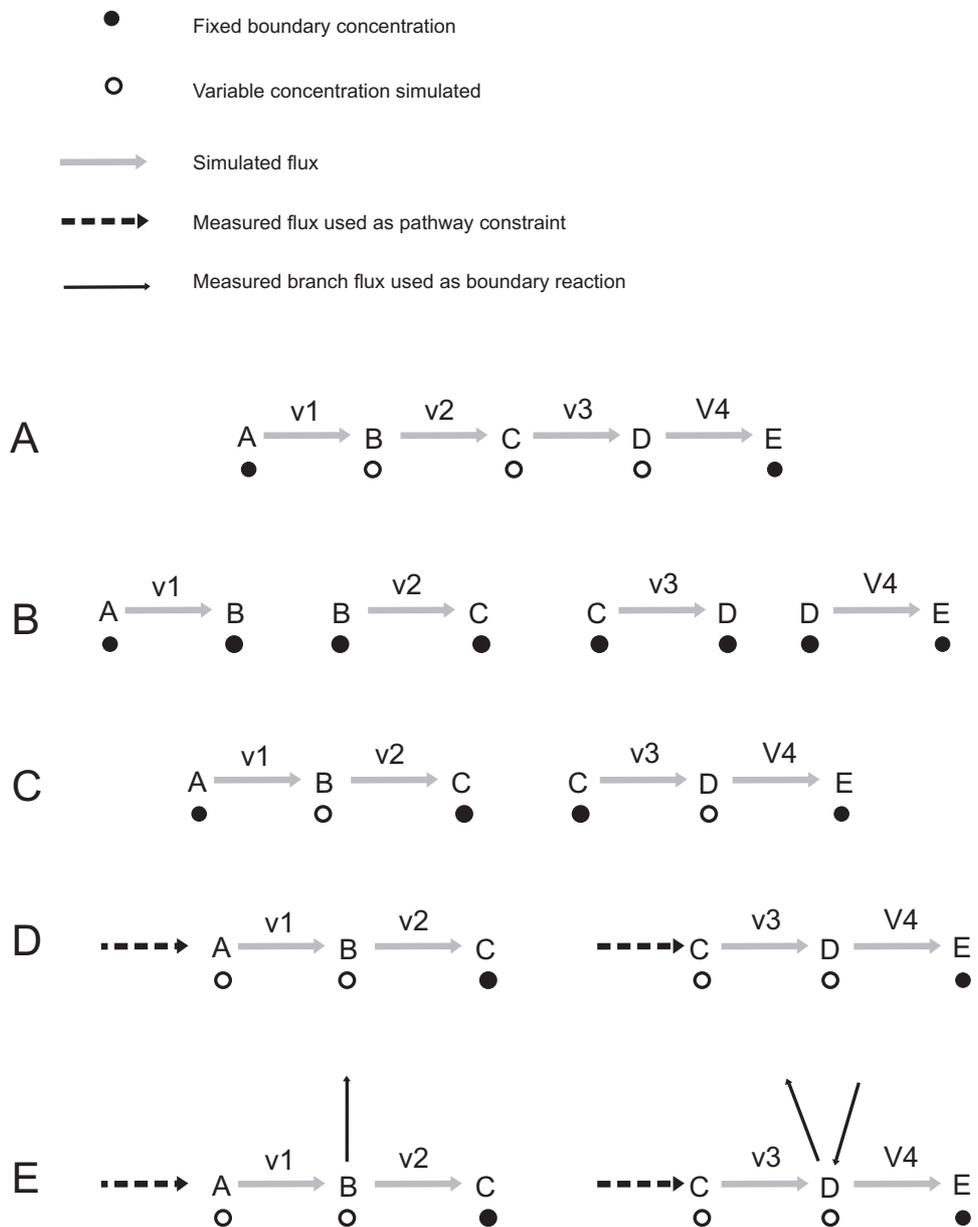


Figure 3.1: The integrated nature of a system of ODE's makes it difficult to detect errors. Small integrated systems with boundary fluxes can be simulated to make errors more obvious in a local assessment approach. See text.

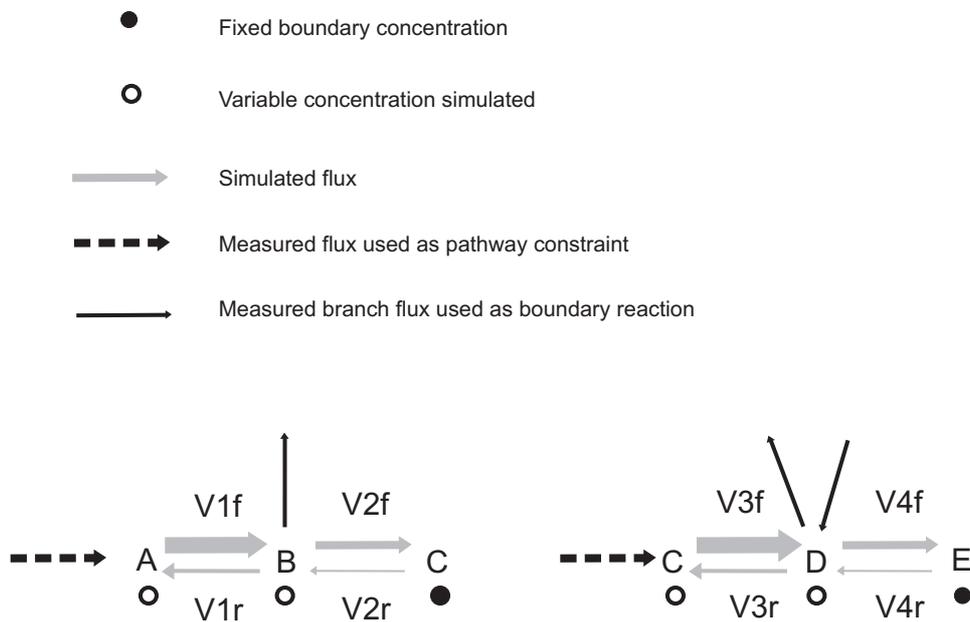


Figure 3.2: A small integrated system which allows local error detection at the steady state. The measured net flux is set, boundary fluxes are included and assessment is based on forward and reverse reactions and on some measured concentrations.

from fluxes to concentrations. This can be achieved by including a fixed reaction in the small integrated system that produces the starting metabolite and by making that metabolite a variable (see figure 3.1 D). Such a protocol would finally result in a system as is shown in figure 3.1 E, in which the boundary reactions are also included.

In summary, it is possible to simulate small kinetic systems, which allows local error detection (see figure 3.2). Constraining a pathway flux to the measured flux can be valuable in accuracy assessment as it moves the validation criterion from the interdependent flux data towards the more independent concentration domain. Accuracy assessment can be carried out on both forward and reverse reactions if the flux measurement technique allows it. Boundary reactions must be included to put realistic constraints on the system. In this methodology, partial metabolite data can be used and the presence and accuracy of the metabolite data determines the size of the simulated systems.

### 3.2.2 Interesting properties of isotope labelling experiments

Many reactions are extremely reversible. Reversibility means that reactions are simultaneously catalyzed both in the forward and reverse directions. There are accordingly an infinite number of combinations of substrate and product concentrations that can produce the same net reaction rate. This means that for one set of enzyme parameters, many combinations of substrate and product concentrations will all produce the same net reaction rate, but the sizes of their forward and reverse rates will differ. This also relates to the parameter space. For one fixed set of concentrations (which could represent the measured values) there will be many parameter sets that produce the same net reaction rate, but they will have different forward and reverse rates. The case with a parameter space can be more readily related to accuracy assessment than the case with a concentration space. As illustration of the principle that reversibility causes observation deficiency of parameters, an example is shown that illustrates the concentration space with a fixed set of parameters. The reason is that the feasible parameter space is nonlinear and difficult to visualize.

Consider the reversible Michaelis-Menten rate equation for a one-substrate, one-product reaction as given below:

$$v = \frac{V_{maxf} \frac{[S]}{K_s} - V_{maxr} \frac{[P]}{K_p}}{1 + \frac{[S]}{K_s} + \frac{[P]}{K_p}} \quad (3.1)$$

The Haldane relationship

$$K_{eq} = \frac{V_{maxf} K_p}{V_{maxr} K_s} \quad (3.2)$$

states that there is a dependency between the equilibrium constant, saturation constants and maximal forward and reverse activities. Note that the original derivation of the reversible Michaelis-Menten rate equation (equation 3.1) features maximal activities for both the forward and the reverse reactions.  $V_{maxr}$  can be replaced by the Haldane relationship as is often

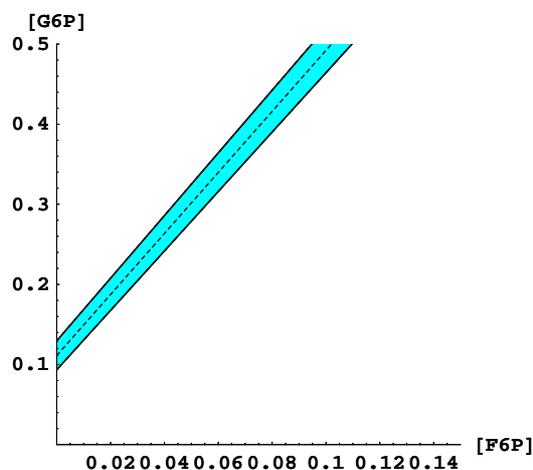


Figure 3.3: A plot of the feasible relationship between the substrate and product concentrations using the net flux as input. The reversible Michaelis-Menten rate equation was set equal to a hypothetically measured net flux of  $50 \text{ mM}\cdot\text{min}^{-1}$  with 5% error. G6P was plotted by scanning F6P. The parameters of the isomerase were obtained from the model of Teusink et al. [44] which is available on JWS Online.

done to yield equation 3.3.

$$v = V_{max} f \frac{\frac{[S]}{K_s} - \frac{[P]}{K_s K_{eq}}}{1 + \frac{[S]}{K_s} + \frac{[P]}{K_p}} \quad (3.3)$$

It is evident that there are two variables (concentrations) for one equation (reaction rate) and thus infinitely many solutions for the variables at every net reaction rate. To visualize this relationship, a plot was generated in Mathematica. The coloured space in figure 3.3 shows the feasible linear relationship that there has to be between the substrate and the product data in order for the equation and this parameter set to be accepted as correct at a specific measured net flux. The net flux ( $50 \text{ mmol}\cdot\text{min}^{-1}\cdot\text{L}^{-1}$ ) was included with a statistical error in measurements of 5 %.

There are thus many feasible concentration sets in this space. For a certain measured concentration set, the values could either fall inside or outside the error margins. If it fall inside the margins, the equation and parameter set

can be considered as accurate. If it falls outside, a different parameter set or a different equation is necessary which will move the feasible concentration space such that it corresponds to the measured values. Important to notice is that there is a large probability that the measured values will fall within the margins, as the feasible concentration space is infinite in linear combinations of substrate and product. This relates to many feasible parameter sets and large uncertainties in the parameter set.

We show here a partial solution to the problem of multi-dimensionality of rate equations in order to constrain the feasible parameter space further. Steady state isotopic tracer experiments have the potential to produce simultaneously the forward and reverse reaction rates. Wiechert [81] has also shown that it is more valid to include both the forward and reverse reactions in  $^{13}\text{C}$ -MFA calculations than just the net reactions, although both cannot always be calculated due to systemic properties of the isotope labelling system. Now, again consider equation 3.3. This equation can be rewritten in the form

$$v = v_f - v_r \quad (3.4)$$

$$v_f = V_{maxf} \frac{\frac{[S]}{K_s}}{1 + \frac{[S]}{K_s} + \frac{[P]}{K_p}} \quad (3.5)$$

$$v_r = V_{maxf} \frac{\frac{[P]}{K_s K_{eq}}}{1 + \frac{[S]}{K_s} + \frac{[P]}{K_p}} \quad (3.6)$$

which actually states that, for example the hexose-6-phosphate isomerase simultaneously catalyses a reaction in the forward direction that responds positively to its own substrate glucose-6-phosphate and a reverse reaction which reacts positively towards its own substrate, fructose-6-phosphate. In both reactions, the substrate of the other reaction acts as a competitive inhibitor, as the two rate equations share a common denominator. Whether the feature of inhibition is relevant to our approach or not, both reactions actually occur *in vivo*.

Figure 3.4 shows the effect of having both the forward and reverse reaction fluxes, as is possible with isotope labelling experiments. The forward and

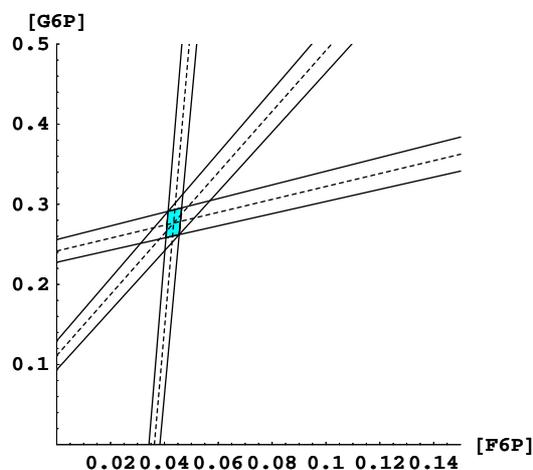


Figure 3.4: A plot of the feasible relationship between the substrate and product concentrations using the forward and reverse fluxes as inputs. The forward and reverse reactions of the reversible Michaelis-Menten rate equation was set equal to hypothetically measured forward and reverse fluxes of  $75 \text{ mM}\cdot\text{min}^{-1}$  and  $25 \text{ mM}\cdot\text{min}^{-1}$ , each with 5% error. The plot for the net flux is also included. See previous.

reverse rate equations (3.5 and 3.6) was set to  $75$  and  $25 \text{ mmol}\cdot\text{min}^{-1}\cdot\text{L}^{-1}$  to result in a net flux of  $50 \text{ mmol}\cdot\text{min}^{-1}\cdot\text{L}^{-1}$ , and each value was assigned a measurement error of 5 %. The feasible concentration space is much smaller, allowing much less uncertainty and creating a stronger case for locating dubious parameters.

Only very recently, interest has appeared in the relationship between the forward and reverse fluxes, concentrations, thermodynamics and enzyme kinetics [99]. This work suggests that the relationships could even be exploited to validate these measurements. In line with this elegant work, it is suggested in this section that the theory be used in terms of providing feasible parameter spaces for a validation criterion.

It is also important to note that this characteristic of isotope labelling experiments can only be used maximally if both the kinetic model and the isotope labelling system are modelled in a detailed form - that is, at least in terms of all metabolic intermediates in a pathway and as reversible reactions where

applicable. The current application of  $^{13}\text{C}$ -MFA however, is not informative enough to make full use of this idea, and was not used with our own experiments. The retro-biosynthetic approach [85] normally used exploits the fact that the isotopic labelling information in the metabolic intermediates is stored in the monomers of biopolymers such as proteins, solving a number of problems such as low metabolite concentrations, high turnover rates and instability of some metabolic intermediates. However, the monomers are only synthesized from a number of intermediates and thus the information potential is decreased in the retro-biosynthetic approach. Apart from the fact that only the labelling patterns of some metabolites are accessible, some fluxes, and especially those which do not alter labelling patterns, can only be resolved if time-dependent labelling of metabolites are accessed and concentrations are measured [84]. However, theorems have been derived which relate the observed apparent exchange fluxes to dependencies between individual exchange fluxes. Using these relationships in validation, feasible parameter spaces will become larger, but is still useful to obtain certain bounds. From [99, 100] it was also apparent that the specific enzyme mechanism might have an influence on the application of the theorems, therefore work still needs to be done in this regard.

### 3.3 Technical considerations in flux-constrained simulation

In the sections above, the general idea of inclusion of boundary fluxes and setting of pathway fluxes was introduced and its unique potential for accuracy assessment was emphasized. This use of a large number of measured boundary fluxes and especially the setting of pathway fluxes is a novel concept according to our knowledge, therefore a special effort has been made to identify possible artefacts to be avoided. During setting of pathway fluxes, a phenomenon was identified which was later identified as an artefact. Also, conserved moieties in kinetic models have further implications. Based on intuition and demonstration with matrix analysis, uncertainties of requirements were clarified. In the following two subsections the technicalities

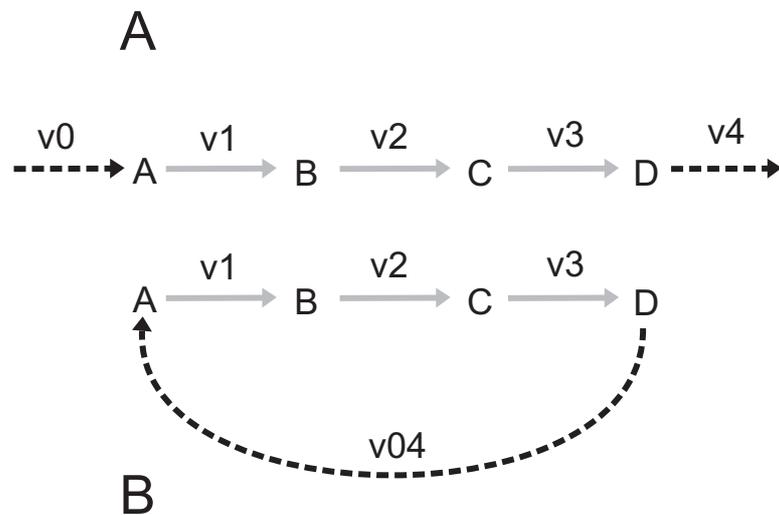


Figure 3.5: The effect of changing more than one boundary concentration in the same linear pathway to variables. See text.

involved with constraining pathway fluxes and implications of conserved moieties are demonstrated.

### 3.3.1 Considerations for constraining pathway fluxes

When pathway fluxes were set to measured values, an observation was made with regard to the number of constraints that had been included in the simulation. When both the reactions  $v_0$  and  $v_4$  were included (see figure 3.5), there was a dependency of the steady state concentration values on the starting values for the intermediate metabolites. When either  $v_0$  or  $v_4$  was included, no such dependency was seen. This dependency upon starting values is an artefact caused by the fact that the fixed  $v_0$  and  $v_4$  will always have an equal flux value, thus an unwanted conserved concentration moiety is formed. Matrices 3.7 and 3.8 show that, because  $v_0$  always equals  $v_4$ , the columns for  $v_0$  and  $v_4$  can be combined into one, giving the pathway structure of a moiety (figure 3.5 B) and linearly dependent rows in the resulting matrix 3.8. If one would include more constraints than is necessary, an unwanted artefact is created.

$$\begin{pmatrix} \mathbf{A}'(t) \\ \mathbf{B}'(t) \\ \mathbf{C}'(t) \\ \mathbf{D}'(t) \end{pmatrix} = \left( \begin{array}{c|ccc|c} 1 & -1 & 0 & 0 & 0 \\ \hline 0 & 1 & -1 & 0 & 0 \\ \hline 0 & 0 & 1 & -1 & 0 \\ \hline 0 & 0 & 0 & 1 & -1 \end{array} \right) \cdot \begin{pmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} \quad (3.7)$$

$$\begin{pmatrix} \mathbf{A}'(t) \\ \mathbf{B}'(t) \\ \mathbf{C}'(t) \\ \mathbf{D}'(t) \end{pmatrix} = \left( \begin{array}{c|ccc} 1 & -1 & 0 & 0 \\ \hline 0 & 1 & -1 & 0 \\ \hline 0 & 0 & 1 & -1 \\ \hline -1 & 0 & 0 & 1 \end{array} \right) \cdot \begin{pmatrix} v_{04} \\ v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad (3.8)$$

The degrees of freedom  $f$ , of steady state fluxes, (which is also the dimension of the  $[q \times f]$  nullspace or kernel of the matrix  $\ker(\mathbf{N})$ ) can be calculated by algebraic analysis of the  $[m \times q]$  stoichiometric matrix  $\mathbf{N}$  as

$$f = q - \text{rank}(\mathbf{N}) \quad (3.9)$$

where  $m$  is the number of variable metabolites and  $q$  is the number of reactions [4]. It is thus necessary to set only  $f$  reaction vectors to constants for constraining the whole system. This corresponds to the inclusion of  $f$  constant *flux setting reactions* in  $f$  of these paths.

There are a number of limitations to this approach. As we exploit the setting of reaction rates, it is inevitably only those boundary concentrations that *could* feature in a stoichiometric matrix that can be used. The activators and inhibitors do not feature in stoichiometry, and are thus irrelevant for this purpose. Such 'stoichiometric boundary concentrations' are generally chosen as those rows in a stoichiometric matrix which have either only one entry or all entries are of similar sign [51]. In general, any metabolite that is produced or consumed by reactions can be used, it does not have to be an extracellular metabolite. Neither does the metabolite have to be consumed in the original kinetic model but could also be produced. Constant negative boundary consumption reactions can constrain the flux in the same way that positive reactions can.

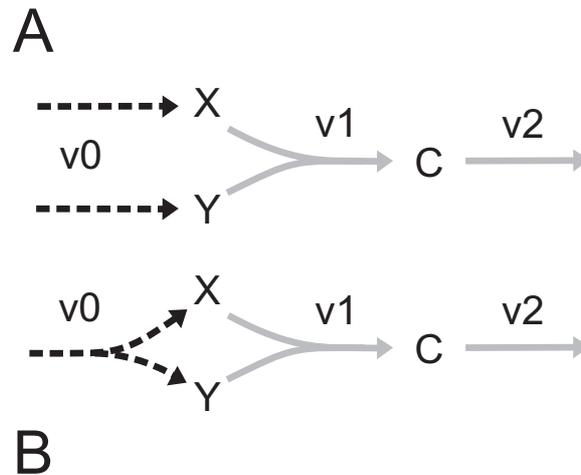


Figure 3.6: The effect of changing more than one boundary concentration in the same reaction to variables.

A further complication arises when one reaction features more than one suitable boundary concentration in its rate equation as in figure 3.6 A and B. Although it is not evident that there are bi-substrate enzyme reactions that consume extracellular metabolites, the idea is to work with small integrated systems which are parts of kinetic modules. The start and end metabolites of these kinetic systems are thus intracellular, where bi-substrate reactions are actually more abundant than single substrate reactions. The many different situations that will arise in the scope of the Silicon Cell thus deserve a generic approach.

If cases A and B represent pathways *in vivo*, they obviously function differently: In case A, metabolites X and Y are produced independently by different reactions, therefore alterations in the metabolic state will cause independent variations in concentrations of X and Y, whereas in B they are always produced simultaneously. By contrast, if cases A and B represent the simulated pathways, they seem to be different cases, like *in vivo*, but A and B are actually equal: In A, the steady state flux measurements of the two reactions in  $v_0$  must be equal due to steady state constraints, and in B,  $v_0$  produces the two substrates simultaneously. It is not necessary that both such boundary concentrations be changed to variables: One of the concentrations could be a measured value. However if they were, it must

be remembered that their time differential equations are dependent: they have the same production and consumption rates at all times, therefore their difference will at all times be equal to their difference at time zero. This dependency can be illustrated by matrix equation 3.10.

$$\begin{pmatrix} \mathbf{X}'(t) \\ \mathbf{Y}'(t) \\ \vdots \end{pmatrix} = \left( \begin{array}{c|ccc} 1 & -1 & 0 & \dots \\ 1 & -1 & 0 & \dots \\ \hline 0 & \vdots & \ddots & \\ \vdots & \vdots & \ddots & \end{array} \right) \cdot \begin{pmatrix} v_0 \\ v_1 \\ \vdots \end{pmatrix} \quad (3.10)$$

$v_0$  is the constant flux that is included in the reaction vector and rows 1 and 2 in the matrix were included as the boundary concentrations that became variables. Rows 1 and 2 only contain the stoichiometric constants for reactions  $v_0$  and  $v_1$  and zeros. Column 1 also contains only the two stoichiometric coefficients and zeros. It can be seen that rows 1 and 2 are positive linear combinations of each other, therefore the difference in their concentrations will always be the same as at time zero of simulation. This dependency of two variables means that validation potential is not maximal. Also does it imply that both these concentrations will have to be measured for use as starting values in the simulation, or at least that the difference has to be included in some way. This is actually true not only for using fixed constraining reactions, but for any such structure in a kinetic model where the rows are *positive*, linear combinations of each other. Which criterion holds the most validation information during flux constraining, using one or two variables for such a case should be analyzed statistically, taking into account the accuracy and availability of concentration data.

It will be found frequently that not all flux vectors can be set to measured values, for there might be no suitable boundary concentrations in the rate equations. Examples would be parallel pathways and loops which might not feature stoichiometric boundary concentrations in their rate equations. Also, some steps are modelled as irreversible steps and their products thus do not feature in their rate equations. A fixed consumption rate cannot be used in that case. Frequently, but not as a rule it was found that fixed consumption

rates for setting fluxes cause problems during simulation. They could be too fast for some of the reactions to respond, thus causing concentrations to have negative values. Mostly, these negative concentrations cause the solve routine to be aborted due to a division by zero in the rate equations. These occurrences can of course be partly due to errors in the kinetic equations. However, the starting values of concentrations could be the cause of these errors when fixed reactions are consuming metabolites with low starting concentrations. It is thus a non-trivial problem and one would likely have to investigate iterative simulations with large starting concentrations (while taking moieties into account), after prior analysis of maximal activities. Alternatively, we could prefer fixed production fluxes over fixed consumption fluxes.

Which one of the possibly multiple stoichiometric boundary concentrations should be used for constraining fluxes should rely on sensitivity analysis. Also, for some small integrated systems, it will be more informative not to simulate the constrained fluxes, but rather to fix the metabolite concentrations, and do comparisons on flux prediction. Sensitivity analysis will show the best criteria.

### **3.3.2 Implications of conserved moieties when using boundary reactions**

Conserved moieties are formed when enzyme reactions that have more than one substrate or product create a specific network structure that forces a group of metabolites to always be conserved in their combined concentration values [51]. This feature of conserved moieties is true regardless of whether the system is at a steady state or in a transient state. In numerical simulation, this sum will at all times be equal to the sum of the starting concentrations of those variables at time zero. This sum is thus a parameter of the system and has to be measured accurately and used as inputs. For steady state modelling it is sufficient to distribute the total sum over the initial values, but for time-simulation it is important to use the individual starting values as they were measured for the time zero. Conserved moieties

can be detected as structural moieties in a network by Gaussian elimination of the rows in a stoichiometry matrix [51, 101]. This algebraic analysis of a metabolic network can reveal important information, even outside the context of kinetic simulation as was well-illustrated for the human parasite *Trypanosoma brucei* [101]. We investigate here the implications that conserved moieties have for inclusion of boundary reactions. As discussed earlier, boundary reactions can be measured and should be integrated with the kinetic model to induce the realistic constraints. The question is now whether there are more considerations to take into account and which artefacts to avoid.

The following two considerations are of importance:

- 1) Some of the metabolic intermediates that are consumed and produced by the boundary reactions will form part of conserved moieties. As by definition there is no change in the sum of metabolites in a moiety, how can boundary fluxes be included to ensure this conservation?

- 2) Inclusion of fixed, measured boundary fluxes into conserved moieties can change the model structure in specific cases such that it seems that the conservation relations are eliminated. Does this mean that the moieties are indeed not present and also make it unnecessary to measure metabolic intermediates for starting conditions?

- 1) In the case of only one boundary reaction acting on a moiety: If the reaction has a fixed value, whether it consumes or produces metabolites from the moiety, the concentrations in the moiety will either increase to infinity or diminish to zero. If the reaction is sensitive to the concentrations in the moiety (their concentrations feature in the rate equation for the boundary reaction), the reaction will go to equilibrium. Both cases are non-realistic and should thus be eliminated. The situation can be detected as a single stoichiometry entry of boundary reactions that affects a moiety. A valid set of reaction boundary constraints to a moiety in a kinetic model thus needs to be precisely balanced to the total kinetic moiety, if fixed boundary reactions are used. For the case of sensitive boundary reactions, at least two reactions

should be present of which at least one can operate in the different direction in order to have a zero net consumption/production of the moiety.

2) It is certain that we would encounter the situation of many reactions in a genome-scale flux model that consume and produce the intermediates of a kinetic model moiety. For instance, the adenine nucleotide moiety consisting of ADP plus ATP is usually found in realistic kinetic models of central energy metabolism. The real situation *in vivo* is that there are reactions that produce those intermediates from free adenine and others that consume them, especially under growing conditions. These reactions will have to be included as branch reactions. Some of these will be reactions that *interconvert* moiety metabolites *in vivo*, thus the branch reactions will accordingly be *precisely balanced interconversion* rates (figure 3.7) which does not affect the moiety constraints. However, there will also be other reactions *in vivo* for *synthesis and breakdown* of these moiety metabolites, which is different from an interconversion. Their presence *in vivo* has the effect that the moiety constraints are released (figure 3.7). The question is: Can incorporation of fixed, measured boundary reactions of the synthesis and breakdown type also release the moiety constraints, and accordingly also eliminate the need for measurement of moiety concentrations?

As can be illustrated in figures 3.7 A and B and the corresponding stoichiometric matrices of the metabolic networks (equation 3.11 A and B), the matrices show linearly dependent rows only for the case of extra interconverting boundary fluxes (equation 3.11 B). This means that the matrices indicate moiety conservation only in the case of interconversion, as is directly obvious from the scheme (B). However, in the computer simulation, the fixed synthesis and breakdown fluxes must be precisely balanced, because their net effect on the moiety must be zero. This results in the observation that the sum constraints are *not* eliminated. It can also be seen that synthesis and breakdown reaction columns in matrix A can be combined to form exactly the same structure as matrix B with the interconversion, which therefore implies a moiety constraint.

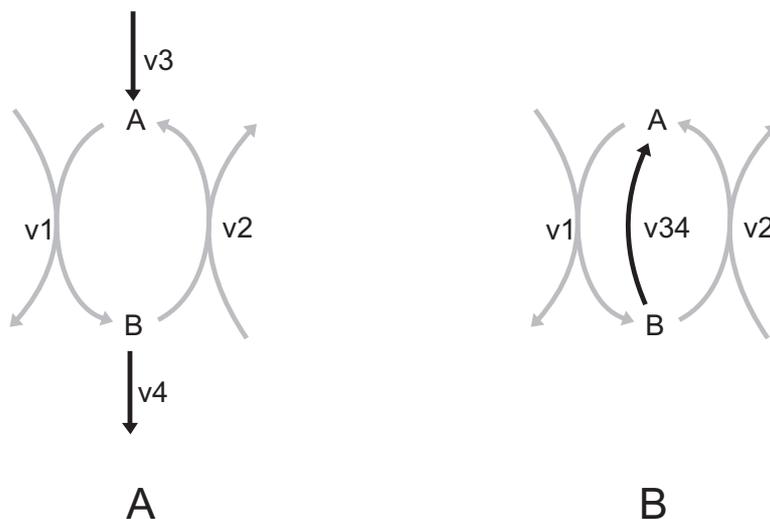


Figure 3.7: Extra synthesis and breakdown reactions included as fixed reaction rates in a conserved moiety of a kinetic model. A: as individual reactions; B: as a single interconversion reaction.

$$\begin{pmatrix} \mathbf{A}'(t) \\ \mathbf{B}'(t) \end{pmatrix} = \left( \begin{array}{cc|cc} -1 & 1 & 1 & 0 \\ 1 & -1 & 0 & -1 \end{array} \right) \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{pmatrix} \quad (3.11)$$

$$\begin{pmatrix} \mathbf{A}'(t) \\ \mathbf{B}'(t) \end{pmatrix} = \left( \begin{array}{cc|c} -1 & 1 & 1 \\ 1 & -1 & -1 \end{array} \right) \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_{34} \end{pmatrix} \quad (3.12)$$

### 3.4 Dynamic pulse experiments: potential and limitations

Steady state validation is limited in its information content. This is because only a few data points of fluxes and concentrations can be used for relating to enzyme kinetic parameters. Dynamic pulse experiments on the other hand, produce more concentration points, and if the data is of sufficient quality, more stringent relations can be drawn to parameters. As we explained

earlier, the better we could separate such a kinetic system into small systems that have sufficiently measurable boundary conditions, the better we can validate the equations and parameters. There are however a few problems with dynamic pulse experiments.

1) Metabolomics is not fully reliable yet in the determination of concentrations due to artefacts. Although very dense time-series can now be generated showing low second scale dynamics [102], the quantitative ability is not sufficient yet for large-scale kinetics.

2) The scope of metabolites that can be observed simultaneously in a given metabolomics technique is limited. The very large scope of new technologies such as GC×GC-TOF-MS [103] is also limited by the fact that GC can only carry metabolites that are volatile enough to evaporate into the GC column and still be stable. Derivatization does lower boiling points of many more metabolites effectively to allow GC, but some larger molecules in metabolism will not be observed in GC. Liquid chromatography (LC) or capillary electrophoreses (CE) should then be used in stead, each with its own problems [104]. *In vivo* NMR suffers from lower sensitivity and resolving power than MS, thus it is also limited to metabolites with larger concentrations. Observation of dynamics throughout metabolism has been identified as being very important and is still lacking [102].

3) Most metabolomics tools are invasive measurements and thus have no spatial resolution. Exceptions are *in vivo* NMR and molecular fluorescent resonance energy transfer (FRET) sensors [105]. As mentioned, NMR is unfortunately not very sensitive, thus excluding most metabolic intermediates. FRET has the potential to be very sensitive, but it has to be taken into account that a detector protein has to be encoded in the cell for each metabolite. Measuring thousands of metabolites simultaneously with FRET is therefore unlikely. Further, binding of metabolites to FRET sensors has kinetics of its own to take into account.

4) Fundamentally, the organism must have genetically expressed transporters for a substrate in order for it to be used as a pulse agent. A good example

of one is glucose, which is frequently used to perturb glycolysis. After a glucose pulse, dynamics of metabolites in the immediate pathways, such as the EMP and HMP pathways will likely be readily observed and links to parameters established with high confidence. However, further away, the dynamics will be much slower, giving time for signal transduction and gene regulation to act on the enzymes. The direct link to enzyme parameters is thus diminished. Also, for many metabolic intermediates there are no transporters and can thus not be used for the purpose of directly perturbing dynamics around them on the cellular map.

5) Sometimes the transport reaction or any reaction at the start of a reaction chain could have a maximal reaction rate significantly lower than the reactions further down the pathway. For *in vivo* validation procedures this poses a problem for there is no way to induce large reaction rates by that route using the pathway substrate.

6) Substrate pulses might not produce enough simultaneous variations in concentrations that affect one reaction, therefore insufficient perturbations are made in the concentration dimensions.

There is thus a need to improve metabolomics techniques to enable more quantitative dynamic experiments, which lies largely in the realm of analytical technology. However, even after advances in these technologies, as well as combining them, fundamental problems of dynamic pulse experiments will persist that limits the information content.

In dynamic pulse experiments, iterative simulations are carried out with a large number of parameter sets. This data not only has associated measurement errors, but also lacks in enough independent variations in substrate and product concentrations around the enzymes. Of special interest to us is that the reversibility of enzyme reactions plays a large role in making it impossible to extract the parameter values with good confidence, or to make accurate validation possible. Accordingly, large feasible parameter spaces are the result. It was explained in section 3.2.2 that the reversibility of enzyme reactions and the extra variables and parameters that it brings

makes it impossible to obtain small parameter spaces when using only concentration and net flux data. It was also explained that the data of forward and reverse fluxes from  $^{13}\text{C}$ -MFA might be used to significantly reduce the feasible parameter spaces. It is now argued that these parameter spaces from steady state flux and metabolomics experiments, especially those from  $^{13}\text{C}$ -MFA should be used in combination with the parameter spaces from pulse experiments. Logically, the intersection between the two sets will be smaller than either of the two sets. The two types of experiment might thus each produce information that the other could not produce for some reason.

It will be interesting to see the increased information from the completely general scenario - that of metabolic and isotopic non-steady state pulse experiments. All the features of both steady state MFA and pulse experiments are thus incorporated into one theory. In the context of the arguments given in this section and in section 3.2.2, it is expected that in this way, the parameter space can be reduced maximally. In this regard it is also obvious why the metabolomics tools that lend themselves to tracing isotopes - NMR and MS - have a larger information potential than other methods.

As a result of realization of this potential, the author also experimented with explicit modelling of isotopomers in simulations with detailed enzyme kinetic models. This work is still at the initial stages, therefore a description is not given here as it will form part of future investigation.

Genome-scale flux/kinetic identifiability with isotopes and pulse experiments will however prove difficult because of problems 4-6 mentioned earlier, as well as the fact that flux identifiability is diminished with an increasing number of label mixing steps. Modelling of the slower reaction networks will extensively depend on kinetics obtained *in vitro*. Hopefully, the new knowledge of *in vivo* kinetics of the fast central metabolism could teach us about the *in vivo* factors such as molecular crowding and channelling such that we can design novel high-throughput *in vitro* assay methods for modelling a Silicon Cell.

### 3.5 Simulation for prediction: Metabolic Control Analysis

If we want to make predictions from the model, like in the form of MCA, boundary reactions have to be sensitive to the intermediates they consume or produce. Such boundary reactions are actually representations of connected pathways in a global cellular network, thus incorporating only the sensitivities of the first connected enzyme reaction does not complete the picture. It might be a good approximation for describing the branch pathway if this first reaction is irreversible, as metabolic control is increased in such a step. If one would leave out the boundary reactions, a normal MCA could be done. However, the result will be at a steady state that might be far from the observed phenotype. Systemic property analysis at an unrealistic steady state simulation makes no sense at all, even though MCA may seem to be an attractive option.

A solution to this problem might be presented by a framework for MCA recently developed to incorporate uncertainties in enzyme parameters and measurements of fluxes and concentrations [63]. For the dubious parameters (or for all parameters, concentrations and flux measurements), probability distributions are generated and Monte Carlo Sampling is used to generate distributions of control properties through the matrix-based MCA formulation. The specific type of kinetic model that the authors used is the (log)linear type of model. It is important to note that the type of model the authors used is not important for the argument, and that it is not suggested here that we use the (log)linear approach which is less detailed than the *in vitro* approach. It is more important that we recognize the usefulness of a framework for incorporating uncertainties in both enzyme kinetic parameters and metabolic measurements.

(Log)linear kinetic models have simple rate equations which are fitted to metabolic data. Thus, the simulated fluxes and concentrations are those of the measurements. The MCA is thus inevitably carried out at the correct simulation results. The matrix-based formulation of MCA generates con-

trol and regulation properties only via elasticities (sensitivities of reactions towards concentrations) at a measured or assumed flux and concentration distribution. For the detailed kinetic models that we want to use in the Silicon Cell, the same matrix-based MCA formalism is valid. The elasticities at a simulated steady state can be obtained from the rate equations and used as in the (log)linear formalism. However, as we do not fit our models to data, the simulated concentrations and fluxes of the kinetic model will not be the same as the measured values, due to two reasons: 1) there are errors present in the enzyme rate equations and parameters, and 2) the absence of modelled boundary reactions that are present *in vivo*. Without some degree of parameter fitting, it is unlikely that the detailed kinetic model will precisely simulate the measured values in all dimensions. Therefore, the formalism for Monte Carlo Simulation MCA can thus not be used in exactly the same form as in [63], at least not in terms of the error distributions for metabolic measurements. A different form of the formalism will have to be used, in which inaccuracies in the simulation penalize the confidence in control and regulation properties.

Whether the model accurately simulates the correct values upon the first attempt or after many model improvement steps, for realistic MCA we will have to get to a simulated steady state that closely resembles the measured values. Inclusion of the boundary reactions in the kinetic model and inevitably also in the MCA formalism is the first step towards obtaining a more realistic simulation. Model improvement follows by more *in vitro* and dynamic substrate pulse experiments, which should result in closer resemblance of the measured fluxes. For MCA, we may need to proceed as follows:

1. Obtain concentration and flux measurements through the system.
2. Carry out the numerical simulation with the measured boundary concentrations and reactions included in the model.
3. Analyse the accuracy of the simulation. If the model does not predict satisfactory results, improve until satisfactory. Improvements can include reassessing the validity of the *in vitro* experiments, extra *in*

*vitro* measurements (especially maximal activity values), reassessment of the conversion factors used in flux and *in vitro* experiments, etc. The accuracy assessment and improvement will also largely depend on time series data from dynamic pulse experiments.

4. Create the matrix-based MCA formulation (as in [63]) with the elasticities at the steady state simulation (at which the steady state variables will not necessarily be precisely those of the measurements). For the boundary reactions, estimate realistic distributions for elasticities, considering measured fluxes, concentrations, available kinetic parameters and thermodynamics.
5. Generate control and regulation distributions.

It is likely that the branch pathways, which have smaller fluxes, will have less flux and concentration control over the central pathways which catalyze larger fluxes, as explained by Snoep et. al. [106]. Therefore it is likely that estimation of these elasticities are not as crucial as for the rest of the model. It is suggested here that the sensitivities of boundary conditions can be modelled with probability distributions to generate control features with Monte Carlo Simulation. It should be evident that this gives better understanding of the control coefficients, taking into account the calculated or assumed uncertainties.

## Chapter 4

# Kinomics

As was mentioned in Chapter 1, a modular approach was suggested for the Silicon Cell Initiative and a work plan was provided for building context independent modules [66]. Such models are currently stored in model databases such as JWS Online (<http://jjj.biochem.sun.ac.za>). The aim is now to start validating them separately on a large-scale reference physiomic dataset before they can be linked to create larger models. As was explained in Chapter 3, the integrated nature of cellular systems and special structural considerations make this integration process rather complex. Further, the available modules are full of simplifications of various forms. In response to these findings, a prototype software program was designed by the author that is useful for further progress in high-accuracy large-scale modelling and validation. In order to avoid repetitive reference to the software as 'the software', we rather call it *Kinomics*.

### 4.1 Main features and uses

In overview, Kinomics is to be used for getting kinetic modules from the JWS Online database into a form which allows:

1. numerical integration to simulate metabolic dynamics or the steady state.

2. combining of individual modules with metabolite and reaction data as boundary conditions or as validation data.
3. extraction and individual simulation of any kinetic subsystem.
4. combining of modules to form larger models.
5. visual access to the modelling process.
6. intervention at multiple stages.

*Mathematica* is a symbolic algebra system with a vast array of high-level functions. What makes this environment particularly useful in Kinomics is its satisfactory symbolic manipulation of large equations and their display. As the aim was not a complete simulation environment such as the excellent software programs PySCeS [107], Virtual Cell [108], Jarnac [109], Copasi [110], CellDesigner [111] and E-Cell [112], it should thus be developed further to work in parallel with other simulation environments to address principles which the other software programs might not have focused on. The complications encountered (described in the previous chapter) encouraged a very robust and simple design in which we can deal with the simplifications in the modules.

The modelling process is annotated throughout by the text cell annotation, describing what each cell does. A decent level of understanding of the workings of the program is at this stage necessary to use Kinomics. Automation is built in to eliminate errors that occur almost inevitably with large kinetic models.

As an example of its application, Kinomics was used to integrate a kinetic module from the JWS database into a significantly larger reaction network of central carbon metabolism, which was also our network for flux measurements, as is illustrated in figure 4.1. Kinomics makes use of a number of spreadsheets for flux and concentration input and for editing of the structure of the model. Model outputs are also written to spreadsheets. The *Mathematica* environment itself is used for a number of concise choice inputs and

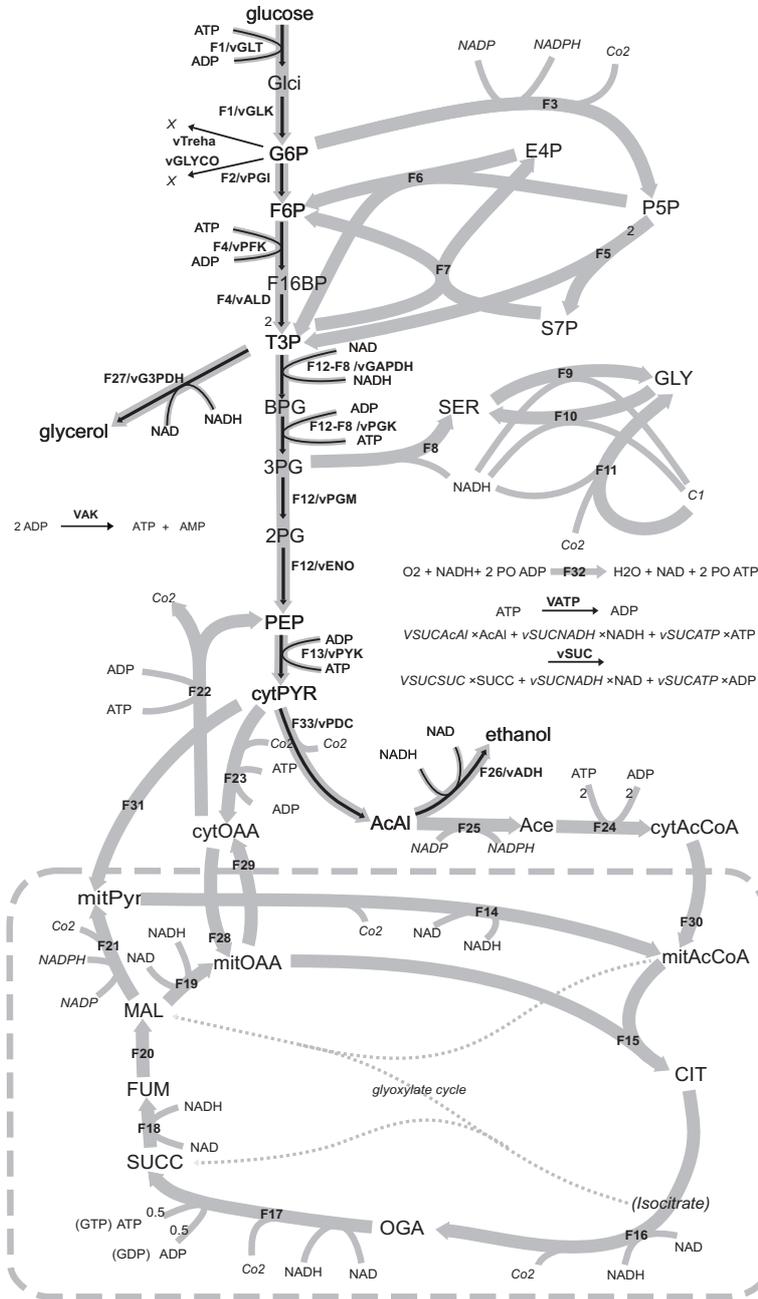


Figure 4.1: Metabolic pathways in central carbon metabolism of the yeast *Saccharomyces cerevisiae*. Grey arrows are measured fluxes in the flux model. Black arrows are reactions in a kinetic model.

for editing of rate equations. Both spreadsheets and the Mathematica environment are customizable and the simple possibility of using colours makes them both more functional for editing. We will now give a brief outline of the principles involved by using the example of the Teusink model.

## 4.2 The internal structure of Kinomics

As the software is rather simple in its design, it can readily be described in terms of a flow diagram (figure 4.2) to show the order of dependencies of structure, variables and parameters as they will be encountered when using the software. Figures 4.4 to 4.8 are amplifications of the spreadsheets in figure 4.2.

Kinetic modules are downloaded from the JWS database in PySCeS (*.psc*) format to a personal folder, as indicated by number (1) in the flow diagram. Here, further choices are made with regard to which of the modules should be used for simulation. Information is extracted from the text files and various *.xls* spreadsheets are generated for further editing. (2) is an ODE based form that contains all information with regard to the metabolic species modelled (see also figure 4.3). This format will ultimately lead to the generation of the stoichiometric network structure. Kinetic modules are added automatically onto the same spreadsheet and thus modules can be connected by *copy & paste* of spreadsheet cells in a robust manner. The default stoichiometry values can also be replaced with names in order to test variable stoichiometry as is shown here for the example of the Teusink model. For example, the stoichiometry coefficients of the succinate producing reaction, vSuc, were replaced by variable names for the individual metabolites that the reaction consumes and produces.

In (3), the names of the branch reactions that consume and produce the species in the kinetic model (the boundary reactions) are copied from output spreadsheets from the large flux measurement model (see also figure 4.4). In the current study, the measured flux outputs from the  $^{13}\text{C}$ -MFA program, FiatFlux, was processed to spreadsheets. In the FiatFlux algo-

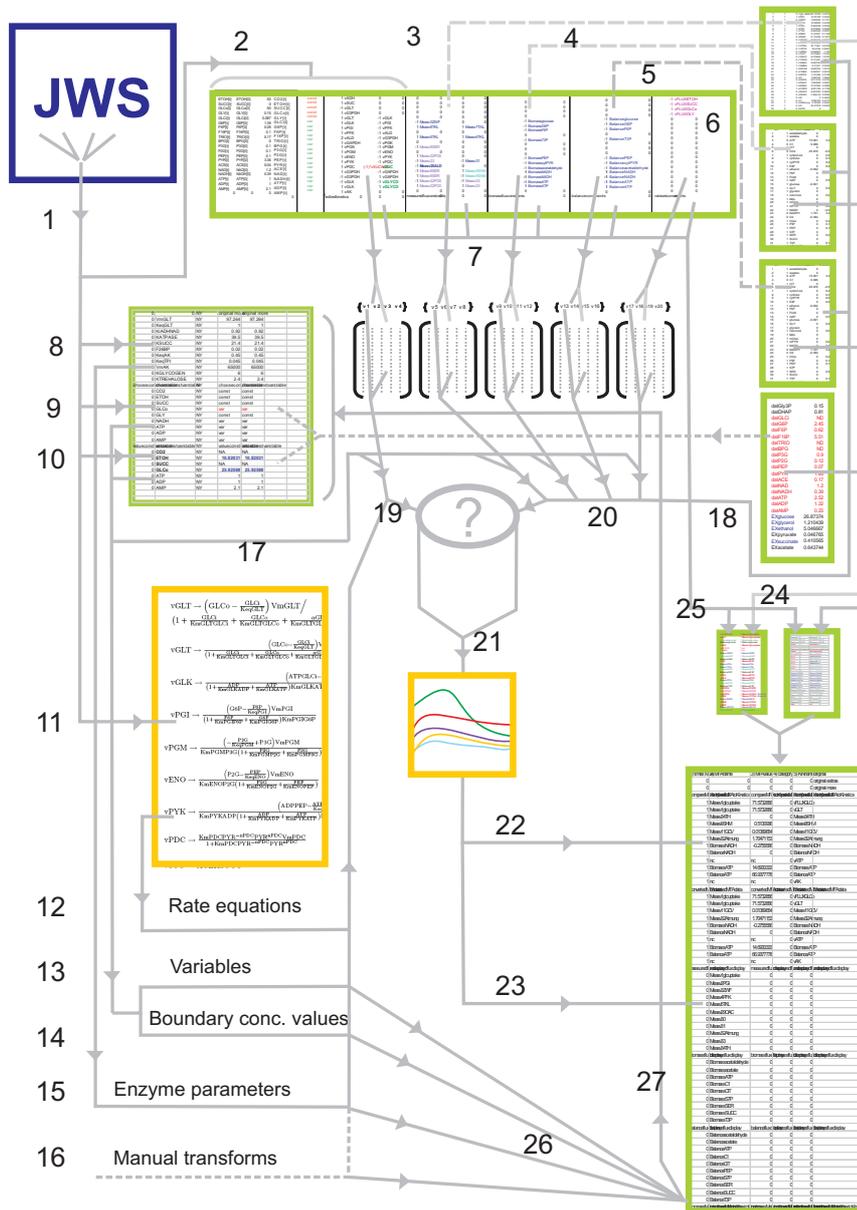


Figure 4.2: Legend on next page.

---

Figure 4.2: Flow diagram of dependencies in Kinomics. Green boxes represent spreadsheets, orange represents *Mathematica* data forms, solid grey lines represent program algorithms, and dashed grey lines represent manual input by the user. (1,2) Kinetic module files are converted to an ODE-based spreadsheet form. (3,4,5) The user copies and pastes boundary reaction names from flux output files onto the ODE-based spreadsheet. (6) Arbitrary names are given for constraining pathway fluxes. (7) Stoichiometric matrices and reaction vectors are generated. (8) A list of enzyme parameters, (9) constant/variable choices, and (10) boundary/initial concentration values are exported to a spreadsheet, to be used as basis for a series of simulations. (11) The kinetic rate equations are kept in *Mathematica* input form. Transforms are created for (12) rate equations, (13) variables, (14) boundary concentrations and (15) enzyme parameters. (16) The user can specify transforms to overrule any of the previous. (17) Constant/variable choices are used to select rows of kinetic model matrices (19) and rows in boundary matrices (20), which are parameterized by data from flux calculations. (21) After a moiety balance check, ODE's are created and numerically simulated. (22, 23) Outputs are reported in spreadsheet form which links kinetic simulation results to the larger flux model. Linkage is manually created via spreadsheets for reactions (24) and metabolomics (25). (26) All simulation inputs are included in the output form, and simulations can be repeated at a later stage (27).





```

CO2          0          0          0          0          0          0
ETOH         0          0          0          0          0          -1 vFLUXETOH
SUCC         0          0          0          0          0          -1 vFLUXSUCC
GLCo        0          0          0          0          0          1 vFLUXGLCo
GLY         0          0          0          0          0          -1 vFLUXGLY
GLCi ..... -1 Biomassglucose          -1 Balanceglucose          0          0
G6P         -1 BiomassG6P          -1 BalanceG6P          0          0
F6P         -1 BiomassF6P          -1 BalanceF6P          0          0
F16P        0          0          0          0          0          0
TRIO        -1 BiomassT3P          -1 BalanceT3P          0          0
BPG ..... 0          0          0          0          0          0
P3G         0          0          0          0          0          0
P2G         0          0          0          0          0          0
PEP         -1 BiomassPEP          -1 BalancePEP          0          0
PYR         -1 BiomasscytPYR          -1 BalancecytPYR          0          0
ACE ..... -1 Biomassacetaldehyde          -1 Balanceacetaldehyde          0          0
NAD         1 BiomassNADH          1 BalanceNADH          0          0
NADH        -1 BiomassNADH          -1 BalanceNADH          0          0
ATP         -1 BiomassATP          -1 BalanceATP          0          0
ADP         1 BiomassATP          1 BalanceATP          0          0
AMP         0          0          0          0          0          0
biomassfluxconstraints          0 balanceconstraints          0 ratesetconstraints          end

```

Figure 4.5: Module *exportfulldynamicnetwork*: *biomassfluxconstraints*, *balanceconstraints* and *ratesetconstraints*. A part of the ODE-based network format the user has to paste from net metabolic flux data, which becomes boundary reactions. The ratesetconstraints have arbitrary names and represent a measured net flux value through the pathway.

rithms, which are based on  $^{13}\text{C}$ -constrained flux balancing, constraints are based on metabolite balances, flux ratios and limits of flux ratios (see section 2.1 and [87]). To calculate net fluxes, optimization is applied to minimize metabolite balances. Accordingly, metabolite balances are not necessarily zero, indicative of either error in measurements, or in the network structure used, or as net reaction rates calculated intentionally by not balancing a metabolite. We interpreted all balances as branch fluxes that consume (negative balances) or produce (positive balances) metabolites. Accordingly, these were included in the boundary reactions (4)(figure 4.5). In the Fiat-Flux methodology, the biomass fluxes of each metabolite are also included, which are calculated from the requirements of each metabolite for biomass, multiplied by the growth rate. These biomass fluxes were also included in the boundary reactions (5)(figure 4.5).

Step (6) is included to allow the user to set pathway fluxes in the kinetic model (see figure 4.5) as was described in Chapter 3. In (7), stoichiometric matrices are generated for each of the blocks in the editing form (*edited-kinetics*, *measuredfluxconstraints*, *biomassfluxconstraints*, *balanceconstraints*

and *setrateconstraints*), each with its reaction vector. The matrices all have the same number of rows (biochemical species), which is the total number of species modelled, both the boundary concentrations and the variables.

A spreadsheet (*variationtables*) is automatically generated that contains the default enzyme parameters (8), a combined set of boundary concentrations or initial concentrations (9), and a list specifying whether each of the species should be treated as a variable or a boundary concentration (10) (figures 4.6 and 4.7). The spreadsheet form allows multiple simulations with various parameter values and variable/boundary choices. The latter feature makes it possible to simulate any subnetwork in a kinetic model.

Rate equations (see Appendix A) are obtained from the modules and kept in the Mathematica extended mathematical representation format (11) as inputs that can be edited manually in a visually understandable manner. Like the above mentioned spreadsheet formats, once this format has been created, it is no longer needed to extract information from *.psc* module files as all the information is stored.

The symbolic language of Mathematica allows the convenient functionality of transformation rules ( $/.$   $\rightarrow$ ). In (12, 13, 14, 15), transforms are created automatically that are used to replace reaction names with rate equations, and to transform rate equations with the relevant variable names, boundary concentration values, and enzyme parameter values. (16) is a step for manually creating transformation rules which can overrule any of the previous transforms, thus allowing the user to change anything with regard to individual steps, such as setting certain reactions equal to a value (for instance zero), changing rate equations, altering boundary concentration values to a time-dependent equation describing data, changing stoichiometry values, etc.

In (17), the choice of variables and boundary concentrations are used to obtain the relevant rows of each of the five stoichiometric matrices that correspond to the metabolite variables. Transforms are created automatically from FiatFlux output spreadsheets (18) that are used to parameterize the

entity	original	original.vSUCATP=4	original+MYHPLC
0	original extras	original extras	original extras
0	original more	original more	original more
VmGLT	97.264	97.264	97.264
KeqGLT	1	1	1
KmGLTGLCo	1.1918	1.1918	1.1918
KmGLTGLCi	1.1918	1.1918	1.1918
alpha	0.91	0.91	0.91
VmGLK	226.452	226.452	226.452
KeqGLK	3800	3800	3800
KmGLKGLCi	0.08	0.08	0.08
KmGLKG6P	30	30	30
KmGLKATP	0.15	0.15	0.15
KmGLKADP	0.23	0.23	0.23
VmPGI	339.677	339.677	339.677
KeqPGI	0.314	0.314	0.314
KmPGIG6P	1.4	1.4	1.4
KmPGIF6P	0.3	0.3	0.3
VmPFK	182.903	182.903	182.903
gR	5.12	5.12	5.12
LO	0.66	0.66	0.66
KmPFKF6P	0.1	0.1	0.1
CPFKF6P	0	0	0
KmPFKATP	0.71	0.71	0.71
CPFKATP	3	3	3
KPFKAMP	0.0995	0.0995	0.0995
CPFKAMP	0.0845	0.0845	0.0845
KIPFKATP	0.65	0.65	0.65
CIPFKATP	100	100	100
KPFKF26BP	0.000682	0.000682	0.000682
CPFKF26BP	0.0174	0.0174	0.0174
KPFKF16BP	0.111	0.111	0.111
CPFKF16BP	0.397	0.397	0.397
VmALD	322.258	322.258	322.258
KeqALD	0.069	0.069	0.069
KmALDF16P	0.3	0.3	0.3
KmALDGAP	2	2	2
KmALDDHAP	2.4	2.4	2.4
KmALDGAPi	10	10	10
VmGAPDHf	1184.52	1184.52	1184.52
VmGAPDhr	6549.68	6549.68	6549.68
KmGAPDHGAP	0.21	0.21	0.21
KmGAPDHBPG	0.0098	0.0098	0.0098
KmGAPDHNAD	0.09	0.09	0.09
KmGAPDHNADH	0.06	0.06	0.06
VmG3PDH	70.15	70.15	70.15
KeqG3PDH	4300	4300	4300
KmG3PDHDHAP	0.4	0.4	0.4
KmG3PDHNADH	0.023	0.023	0.023
KmG3PDHNAD	0.93	0.93	0.93
KmG3PDHGLY	1	1	1
VmPGK	1306.45	1306.45	1306.45
KeqPGK	3200	3200	3200
KmPGKBPG	0.003	0.003	0.003
KmPGKP3G	0.53	0.53	0.53
KmPGKADP	0.2	0.2	0.2
KmPGKATP	0.3	0.3	0.3
VmPGM	2525.81	2525.81	2525.81
KeqPGM	0.19	0.19	0.19
KmPGMP3G	1.2	1.2	1.2
KmPGMP2G	0.08	0.08	0.08
VmENO	365.806	365.806	365.806
KeqENO	6.7	6.7	6.7
KmENOP2G	0.04	0.04	0.04
KmENOPEP	0.5	0.5	0.5
VmPYK	1088.71	1088.71	1088.71
:	:	:	:
:	:	:	:
:	:	:	:

Figure 4.6: Module *variationtables A*. Each column in the list corresponds to a set of enzyme parameters. The list is initially obtained from the files on JWS Online. See text.

KeqPYK	6500	6500	6500
KmPYKPEP	0.14	0.14	0.14
KmPYKPYR	21	21	21
KmPYKADP	0.53	0.53	0.53
KmPYKATP	1.5	1.5	1.5
VmPDC	174.194	174.194	174.194
KmPDCPYR	4.33	4.33	4.33
nPDC	1.9	1.9	1.9
VmADH	810	810	810
KeqADH	0.000069	0.000069	0.000069
KmADHACE	1.11	1.11	1.11
KmADHETOH	17	17	17
KmADHNADH	0.11	0.11	0.11
KmADHNAD	0.17	0.17	0.17
KiADHACE	1.1	1.1	1.1
KiADHETOH	90	90	90
KiADHNADH	0.031	0.031	0.031
KiADHNAD	0.92	0.92	0.92
KATPASE	39.5	39.5	39.5
KSUCC	21.4	21.4	21.4
F26BP	0.02	0.02	0.02
KeqAK	0.45	0.45	0.45
KeqTPI	0.045	0.045	0.045
VmAK	65000	65000	65000
KGLYCOGEN	6	6	6
KTREHALOSE	2.4	2.4	2.4
CO2	const	const	const
ETOH	const	const	const
SUCC	const	const	const
GLCo	const	const	const
GLY	const	const	const
GLCi	var	var	var
G6P	var	var	var
F6P	var	var	var
F16P	var	var	var
TRIO	var	var	var
BPG	var	var	var
P3G	var	var	var
P2G	var	var	var
PEP	var	var	var
PYR	var	var	var
ACE	var	var	var
NAD	var	var	var
NADH	var	var	var
ATP	var	var	var
ADP	var	var	var
AMP	var	var	var
CO2	NA	NA	NA
ETOH	50	50	5.046666921
SUCC	NA	NA	NA
GLCo	50	50	26.87373596
GLY	0.15	0.15	1.210438787
GLCi	0.087	0.087	0.087
G6P	1.39	1.39	1.39
F6P	0.28	0.28	0.28
F16P	0.1	0.1	0.1
TRIO	5.17	5.17	5.17
BPG	0	0	0
P3G	0.1	0.1	0.1
P2G	0.1	0.1	0.1
PEP	0.1	0.1	0.1
PYR	3.36	3.36	3.36
ACE	0.04	0.04	0.04
NAD	1.2	1.2	1.2
NADH	0.39	0.39	0.39
ATP	1	1	1
ADP	1	1	1
AMP	2.1	2.1	2.1

Figure 4.7: Module *variationtables B*. Together with the enzyme parameters there are two sections: one which specifies the choice whether to make a modelled species a variable or a constant in the model, and one which specifies the value of either the starting condition or the boundary condition, based on the above choice. See text.

boundary reaction matrices corresponding to the measured fluxes, biomass fluxes and balance fluxes that consume or produce the metabolite variables. In (19), the transforms for the rate equations are used to transform the kinetic model matrix to a set of differential equations containing kinetic rate equations.

The idea behind the creation of individual matrices and vectors and the use of transforms is that it gives the user control over the process. The network is built from parts and transformations are applied stepwise, therefore it is possible to test, for instance, the effect of constraint reactions on the kinetic reaction network. An algorithm was implemented to test whether the boundary reactions have any effect on the conserved concentration moieties. The algorithm makes use of the method described in [51] by which an identity matrix is appended to the stoichiometric matrix and Gaussian elimination is applied to obtain conserved moieties. The parameterized boundary reactions (20) are then applied to all the species in the moiety and it can be established whether there is any net consumption or production to a moiety. It is for instance necessary that the production rates of moiety species must be precisely equal to their consumption rates, as was described in Chapter 3.

An algorithm was created that checks whether a certain reaction can take on a certain measured flux. This could also be done by inspection of  $V_{max}$  values, however this parameter might not feature in all rate equations. In most cases, the reverse maximal activities also do not feature explicitly, as this parameter is usually replaced by the Haldane relationship. Accordingly, an algorithm was built which makes use of a totally generic approach by which it is not necessary to specify a maximal activity in a specific direction, and thus makes automation possible. The algorithm is based on using all combinations of variables featuring in a rate equation, where the concentration variables are replaced with near-zero and near-infinity values. The reaction rates are then calculated to obtain the limits of a reaction.

If it is found that the constraints applied by the boundary reactions are feasible, all the transformations are applied and the system of differential

equations are obtained (21) which is numerically integrated in the Mathematica algorithm *NDSolve*. If a steady state is reached, the reaction rates (22) and the concentrations (23) at the end time can be written to an output spreadsheet file.

The comprehensive output file is created via two other spreadsheet forms which link reactions and biochemical species in the kinetic model to those in the larger network of reaction and concentration data. For the reaction data spreadsheet (*relateMFAtoKinetics*) and the metabolomics spreadsheet (*relateMetabolitestoKinetics*) the reaction and species vectors are obtained automatically from the kinetic model integration spreadsheet and the data spreadsheets (24, 25). By *drag & drop*, the user then lines up the kinetic entity with the data entity, as well as coefficients to relate them quantitatively (figure 4.8 A and B). Figure 4.8 A shows the list for linking the outputs of the flux-constrained kinetic model to the values of the flux model. The same names of the boundary reactions thus occur in both reaction columns. A relational coefficient is specified in the column between the lists which describes the numerical relationship between a simulated flux and a measured flux. All boundary fluxes are related with a coefficient of 1. Other coefficients could have different values, for instance if the convention of stoichiometry is different between the data model and the kinetic model. The simulated fluxes can also be a combination of measured fluxes, in the case where the pathway structures in the flux model and the kinetic model are described differently. For instance, *vGAPDH* and *vPGK* in the kinetic model are represented by the sum of *Measv12* and *Measv8* in the data model. In the original flux model, *Measv8* directly consumes T3P and not 3PG, as is shown in figure 4.9 A (reaction F8). In the original flux model, the stoichiometry convention thus describes the production of PEP as in the kinetic model, but not the consumption of T3P. The consumption of T3P to form 3PG (which is *vGAPDH* and *vPGK* in the kinetic model) does not exist in the flux model, but can be described as *Measv12 + Measv8*. The robustness in this manual linkage step thus makes it possible to handle this situation of simplification, to quantitatively link the two model types in the correct way, as is shown in figure 4.9 B.

vFLUXGLCo	1	Measv1glcuptake	GLCi	1	datGLCi
vGLT	1	Measv1glcuptake	G6P	1	datG6P
Biomassglucose	1	Biomassglucose	F6P	1	datF6P
Balanceglucose	1	Balanceglucose	F16P	1	datF16P
vGLK	1	Measv1glcuptake	TRIO	1	datTRIO
vGLYCO	1	nc	BPG	1	datBPG
vTreha	1	nc	P3G	1	datP3G
Measv3ZWF	1	Measv3ZWF	P2G	1	datP2G
BiomassG6P	1	BiomassG6P	PEP	1	datPEP
BalanceG6P	1	BalanceG6P	PYR	1	datPYR
vPGI	1	Measv2PGI	ACE	1	datACE
Measv6TKL	1	Measv6TKL	NAD	1	datNAD
Measv7TAL	1	Measv7TAL	NADH	1	datNADH
BiomassF6P	1	BiomassF6P	ATP	1	datATP
BalanceF6P	1	BalanceF6P	ADP	1	datADP
vPFK	1	Measv4PFK	AMP	1	datAMP
vALD	1	Measv4PFK	CO2	1	datCO2
Measv5TKL	1	Measv5TKL	ETOH	1	EXethanol
BiomassT3P	1	BiomassT3P	SUCC	1	EXsuccinate
BalanceT3P	1	BalanceT3P	GLCo	1	EXglucose
vG3PDH	1	Measv27GPD	GLY	1	EXglycerol
vFLUXGLY	1	Measv27GPD			
vGAPDH	1	Measv12TDH + Measv8SER			
vPGK	1	Measv12TDH + Measv8SER			
Measv8SER	1	Measv8SER			
vPGM	1	Measv12TDH			
vENO	1	Measv12TDH			
Measv22PCK	1	Measv22PCK			
BiomassPEP	1	BiomassPEP			
BalancePEP	1	BalancePEP			
vPYK	1	Measv13PYR			
Measv23	1	Measv23			
Measv31	1	Measv31			
BiomasscytPYR	1	BiomasscytPYR			
BalancecytPYR	1	BalancecytPYR			
vPDC	1	Measv33			
Biomassacetaldehyde	1	Biomassacetaldehyde			
Balanceacetaldehyde	1	Balanceacetaldehyde			
vADH	1	Measv26ADH			
vFLUXETOH	1	Measv26ADH			
vSUC	1	Measv25ALD			
vSUC	1	Measv24ACS			
vSUC	1	Measv30			
vSUC	1	Measv15CIT			
vSUC	1	Measv16ACO			
vSUC	1	Measv17KGD			
vFLUXSUCC	1	Measv17KGD			
Measv10SHM	1	Measv10SHM			
Measv14PDA	1	Measv14PDA			
Measv18SDH	1	Measv18SDH			
Measv19MDH	1	Measv19MDH			
Measv34TH	1	Measv34TH			
Measv9SHM	1	Measv9SHM			
Measv11GCV	1	Measv11GCV			
Measv32Atmung	1	Measv32Atmung			
BiomassNADH	1	BiomassNADH			
BalanceNADH	1	BalanceNADH			
vATP	1	nc			
BiomassATP	1	BiomassATP			
BalanceATP	1	BalanceATP			
vAK	1	nc			

B

A

Figure 4.8: (A) Module *relateMFAtoKinetics* links the simulated flux values of the flux-constrained kinetic model to the values of the flux data model. A relational coefficient is specified in the middle column describing the relationship between a simulated and a measured flux. The simulated fluxes can also be a combination of measured fluxes, in the case where the meaning of a certain reaction in the data model differs from that in the kinetic model. (B) Module *relateMetabolitestoKinetics*. This relates the simulated concentration values in the kinetic model with metabolomics data through a relational coefficient, or as a combination of data values.

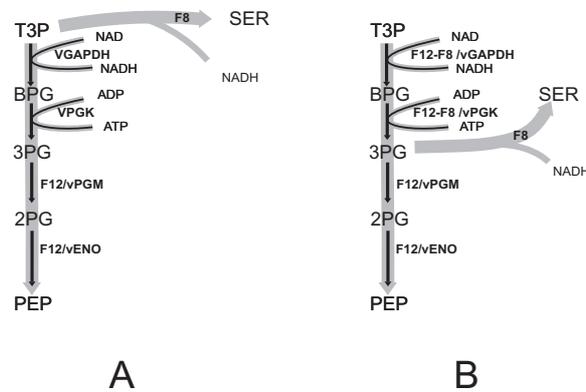


Figure 4.9: (A) The simplified pathway structure used in the flux model. The serine branch consumes T3P. (B) The serine branch was included in Kinomics as a boundary reaction consuming 3PG, as it occurs *in vivo*.

Multiple simulation outputs are reported next to one another, as well as all information necessary (26) to carry out the same simulation again. This information can automatically be reloaded (27) and used for further simulations when starting a new Mathematica session, and the results written to the same spreadsheet next to the previous results. Large output files can thus be created and the spreadsheet format allows post-simulation analysis.

Kinomics was designed by the author in its entirety and is already functional. However, it was written as a prototype program for integration and is not yet a user friendly standalone program at this stage. The useful features of the symbolic language were exploited. It seems that graphical user interfaces will be necessary to make working with large network structures and many parameters less intimidating. It could be efficient to use the J/Link extension package to Java that is available in Mathematica. Also, as bioinformatics will eventually have to be a feature in building the Silicon Cell, extensions to Python could be very useful. Python is also being used at our institute for writing the PySCeS [107] modelling software. Model building at the genome scale with thousands of reactions will definitely need extensive computer assistance and many functional links through bioinformatics.

## Chapter 5

# $^{13}\text{C}$ -metabolic flux experiments

Kinetic modules must be validated against an extensively characterized reference physiomic dataset. For genome-scale kinetic modelling, more informative experiments still have to be designed, which will include various principles such as metabolomics, isotope labelling experiments and dynamic pulse experiments. To illustrate the flux-constrained approach (see Chapter 3) with relevant experimental data,  $^{13}\text{C}$ -MFA experiments were conducted for the yeast, focussing on the central carbon metabolism. The aim here was more to learn the technique of  $^{13}\text{C}$ -MFA than it was to obtain new knowledge of yeast physiology. Experimental protocols for  $^{13}\text{C}$ -MFA had already been well-defined [113].

### 5.1 Experimental procedures

We calculated metabolic fluxes through the central metabolic pathways of yeast with the previously described flux model ([113, 114], figure 4.1) with FiatFlux [95] which is based on  $^{13}\text{C}$ -constrained flux balancing. Overall, our experimental setup of choice was to use small aerobic and anaerobic batch cultures of 20 mL in which labelling patterns were obtained with GC-MS, while external metabolites were determined with HPLC.

The choice to use both aerobic and anaerobic conditions for the yeast was due to a number of considerations with regard to the experimental conditions for which the kinetic model of glycolysis [44] was constructed. The authors used compressed yeast (Koningsgist) for determination of maximal enzyme activities and other enzyme parameters. Their model was validated against glucose run-out experiments of compressed yeast in phosphate buffer. To eliminate reactions through metabolic branches that consume intermediates from the network defined by the kinetic model (fermentative metabolism of glucose), the medium was without a nitrogen source, to eliminate growth reactions. It was also anaerobic to eliminate electron transport and oxidation of NADH.

It can be argued that the ready-grown compressed yeast, which is industrially produced, would have an enzyme activity profile characteristic of aerobic cultivation. Industrial cell biomass production is usually aerobic due to less by-product formation, which is wasteful. It is however uncertain whether the yeast might have adapted in storage to other conditions and both aerobic and anaerobic conditions were applied in our study.

More important is our choice of using not compressed yeast, but a very relevant laboratory strain CEN.PK 113-7D, and growing conditions. The idea is in fact not to keep simplifying the metabolic network by applying unusual culture conditions as is often done in kinetic modelling studies, but rather to extend the network by using rapidly growing, healthy yeast in a more complex medium. Such conditions will probably also stimulate more interest in the project as they are more realistic and applicable. One can assume that kinetic parameters in the two strains are more or less the same, but that maximal activities would have to be measured later for more accurate modelling.

In general, procedures were as described in [113] except for a few changes that are specified here.

### 5.1.1 Growth conditions

The strain *Saccharomyces cerevisiae* CEN.PK 113-7D was grown in small aerobic and anaerobic batch cultures. For aerobic growth, 20 mL cell cultures in 150 mL Erlenmeyer flasks were used on a rotary shaker at 250 rpm and kept at 30°C. For anaerobic growth, 20 mL cell cultures in 25 mL Erlenmeyer flasks were used on a rotary shaker at 150 rpm at 30°C. In the aerobic cultures, caps were made from loose-fitting aluminium foil in order to allow optimal gas exchange. In the anaerobic cultures, caps were made from tight-fitting thick silicon caps originating from 60 mL syringes through which needles were punched for flushing the headspace with nitrogen gas. The gas was bubbled through water first to reduce evaporation in the culture vessels. Although oxygen readily diffuses through silicon, it was assumed that the combination of nitrogen flushing, slower shaking, the smaller vessel, and rapid initial oxygen uptake of the yeast should have greatly reduced the dissolved oxygen concentration.

The growth medium was yeast minimal base without amino acids and ammonium sulphate (from Difco), as was also used by others [114]. This growth medium contains trace elements and vitamins and it was assumed that the uptake of the minimal components did not significantly contribute to central metabolic fluxes or the synthesis of amino acids. The yeast minimal base was supplemented with 5 g.L<sup>-1</sup> ammonium sulphate as nitrogen and sulphur source. The pH was buffered at 6.0 with 100 mM MES buffer. For growth rate determination and measurement of external metabolite concentrations, the medium was supplemented with 5 g/L glucose of 100 % naturally labelled glucose. For flux ratio analysis, the medium was supplemented with 5 g.L<sup>-1</sup> glucose of either 100 % 1-<sup>13</sup>C glucose (99.0 % purity, from Sigma Aldrich) or 20 % U-<sup>13</sup>C glucose (98.5 % purity, from Spectral Isotopes) plus 80% naturally labelled glucose. Culture vessels were autoclaved whereas media was filter-sterilized with 0.22  $\mu\text{m}$  pore size PES (polyethersulfone) filters (from Millipore).

*Note:* The Verduyn defined minimal medium most frequently used in  $^{13}\text{C}$ -MFA for yeast [113, 115] was used initially, but the yeast did not respond

well to the medium, and very low growth rates were observed. It is suspected that one of the many organic vitamin components available at the time might have deteriorated in storage. This medium is especially difficult to work with for the anaerobic cultures, which need additional lipids ergosterol and Tween 80 which the yeast cannot synthesize under anaerobic conditions [115]. The lipids are not very soluble in water and the Tween 80 crystals initially forming upon addition to water will be filtered out. The viscous Tween 80 also cannot be filter-sterilized. The best way to combine the two supplements is to dissolve the ergosterol in a small amount of ethanol and mixing it with Tween 80 which can be frozen in this way for storage. After filter-sterilization of the other components and addition to the culture vessels, the lipid mixture is heated to near-boiling point (which quickly dissolves the ergosterol crystals, re-sterilizes the content and evaporates the ethanol), and added directly to the medium. The latter procedure was used by [116] for anaerobic cultivation of the yeast in a chemostat. When using small volumes as in our study however, the lipids and water-soluble compounds need to be first sterilized separately (with filters and heat respectively), then pre-mixed and added to the vessels as such small amounts of lipid cannot be added accurately and the viscous fluid sticks to pipettes and filters. The need for these extra insoluble supplements in defined media makes anaerobic cultivation of yeast in small quantities laborious and possibly inaccurate. Miniaturized flux experiments for anaerobic cultures of yeast are also rarely seen, if ever. The use of yeast minimal base without amino acids or ammonium sulphate was very efficient and made experiments much easier.

### 5.1.2 Determination of cellular growth rate

Cell growth was followed at  $\text{OD}_{620}$  in a Titertek Multiskan PLUS MKII 96-well microtiter plate spectrophotometer and microtiter plates with  $300\ \mu\text{L}$  sample volume. Algorithms were implemented in Mathematica to fit exponential curves through values of  $\text{OD}_{620}$  to obtain maximal specific growth rates.

### 5.1.3 Cellular dry weight determination

A stock of cells was grown in batch cultures and was harvested in mid-exponential phase. For each data point, a volume corresponding to approximately 100 mg of cellular dry weight was added to a membrane filter disk with 0.22  $\mu\text{m}$  pore size in a Millipore membrane filter. Membrane filter disks were dried for 8 minutes at 330 Watt in a microwave oven and placed in a dessicator overnight. Dried filters were immediately weighed after they were taken out of the dessicator. The conversion factor was calculated as

$$c = \frac{\frac{gdw}{L}}{OD_{620}} \quad (5.1)$$

### 5.1.4 Analysis of external metabolites with HPLC

To minimize volume decrease in cell cultures due to sampling, small samples of 450  $\mu\text{L}$  were taken with a pipette of which 300  $\mu\text{L}$  was used for brief  $OD_{620}$  determination and added again to the rest afterwards. Samples were kept on ice during the process. Samples were centrifuged at 14 000 rpm at 4°C and the supernatants frozen at -80°C for later analyses. 400  $\mu\text{L}$  of sample was treated with 24.5  $\mu\text{L}$  of chilled PCA (35% m/v) and kept on ice for 10 minutes. 22  $\mu\text{L}$  of KOH (7 M) was added and kept on ice for 10 minutes. The samples were centrifuged at 14 000 rpm for 10 minutes at 4°C and the supernatants were kept for analysis with HPLC. The standards for HPLC analysis were glucose, ethanol, glycerol, succinate, acetate, and pyruvate and were treated in the same way with PCA and KOH as the samples to eliminate dilution effects. A Biorad Aminex HPX-87H ion exclusion column for organic acids (300 mm  $\times$  7.8 mm, catalogue number 125-0140) was used at 55°C on an Agilent 1100 Series HPLC system with a Waters 717 autosampler, Waters 996 phosphodiode array detector (UV) and Waters 410 differential refractometer (RI). The flow rate was 500  $\mu\text{L}\cdot\text{min}^{-1}$  and the mobile phase 5 mM  $\text{H}_2\text{SO}_4$ . Standard curves were generated for concentrations of 2, 5, 10, 20 and 50 mM and concentrations were determined as areas under the curves in Millennium software. Glucose, ethanol and glycerol were determined with RI detection while succinate, acetate and

pyruvate were determined with UV detection.

### 5.1.5 Analysis of labelled proteinogenic amino acids with GC-MS

The full 20 mL of cultures grown on labelled glucose cells were harvested in early-exponential phase at biomass concentrations between 0.1 and 0.2 g.L<sup>-1</sup> and centrifuged at 14 000 rpm at 4°C. The pellets were washed once with water and centrifuged. Pellets were frozen for later analyses. Cellular pellets were hydrolyzed using 1 mL of 6 M hydrochloric acid at 110°C in melt-sealed glass test tubes in a dedicated hydrolysis oven. 100  $\mu\text{L}$  of the hydrolysis fluid was moved to a 1 mL flat-bottomed glass vial and the acidic fluid removed by blowing a continuous stream of air through a pipette tip. Immediately before derivatization, the blow-dried pellet was dried further in an oven at 85°C for at least 20 minutes. Derivatization was achieved by addition of 50  $\mu\text{L}$  of distilled dimethylformamide (DMFA) and 50  $\mu\text{L}$  of N-(tert-butyldimethylsilyl)-N-methyltrifluoroacetamide (MTBSTFA) with a Hamilton syringe and further incubation at 85 ° C for 60 minutes.

Either 1 or 2  $\mu\text{L}$  of the derivatized sample was injected into a Hewlett Packard Agilent 6890N (GC) 5975 (MS) system. The GC-MS procedures were those already described [92] except for a few changes. An Agilent fused silica GC column with PDMS stationary phase and dimensions (60 m  $\times$  0.32 mm  $\times$  0.25 mm) was used. The eluent gas was helium and the column pressure was kept at 80 kPa. A split ratio of 1:20 was used. The initial oven temperature was 150°C and maintained for 2 min and then raised to 280°C at 3°C.min<sup>-1</sup>. The GC to MS interface temperature was 250°C, the MS ion source temperature was 200°C and electron impact ionization (EI) was used at -70 eV. MS spectra were analyzed in the range of 60-650 atom mass units at sampling intervals of 0.81 seconds per scan. A run time of 80 minutes was needed for the long column. MS data were processed using ChemStation.

*Note:* Glass Hamilton syringes should be used for handling MTBSTFA as the fumes react with polymers inside pipettes. Hydrolysis should only be carried out in dedicated hydrolysis ovens as the high temperatures and strong

acid presents a danger. Hydrolysis fluid cannot be extracted with freeze-drying as the acid fumes can damage seals on freeze-dryers.

### 5.1.6 Determination of production and consumption rates of external metabolites

Algorithms were written in Mathematica to fit exponential curves through  $OD_{620}$  data to obtain the equations

$$B_t = \left(\frac{gdw}{L}\right)_t \quad (5.2)$$

by using the conversion factor of

$$c = \frac{\frac{gdw}{L}}{OD_{620}}$$

for each  $OD_{620}$  reading for both aerobic and anaerobic conditions. Curves were also fitted through HPLC data to obtain

$$M_t = \left(\frac{\text{mmol metabolite}}{L}\right)_t \quad (5.3)$$

For concentration boundary conditions in kinetic simulations, the average concentrations during the time before sampling for  $^{13}\text{C}$  analysis in GC-MS were calculated as the integral over the time-dependent functions of concentrations, divided by the time interval as

$$\bar{M}_t = \frac{\int_{t_0}^{t_{max}} M_t dt}{dt} \quad (5.4)$$

$t_0$  was the time at which the biomass concentration was  $0.05 \text{ gdw.mL}^{-1}$  (lower accurate detection limit). At  $t_{max}$  the biomass concentration was  $0.1 \text{ gdw.mL}^{-1}$  for aerobic and  $0.2 \text{ gdw.mL}^{-1}$  for anaerobic cultures. The interval between  $t_0$  and  $t_{max}$  was used for  $dt$ .

The total derivative of each of the time-dependent concentration equations was normalized with the corresponding biomass concentration equations to obtain specific production or consumption rates as functions of time

$$M_t^{dB} = \frac{\frac{dM_t}{dt}}{B_t} \quad (5.5)$$

Further, the average production or consumption rates during the time before sampling for  $^{13}\text{C}$  analysis in GC-MS was calculated as the integral over the time-dependent functions of specific production or consumption, divided by the time interval

$$\bar{M}_t^{dB} = \frac{\int_{t_0}^{t_{max}} M_t^{dB} dt}{dt} \quad (5.6)$$

This approximates the average contribution of quasi steady state external fluxes to the labelling patterns accumulated in proteinogenic amino acids during the time interval before the sampling of labelled amino acids.

### 5.1.7 Metabolic flux ratio analysis and net flux analysis

The  $^{13}\text{C}$ -MFA software tool FiatFlux described in [95] was kindly provided by the authors and was used to analyze the MS spectra. The details of flux ratio analysis were explained in Chapter 2. The program was run in the commercial software Matlab. The necessary open source software for using network common data format (*.netCDF*, exported as *.AIA files*) in Matlab, *mexnc* and *mexcdf* was obtained and installed as described in the FiatFlux software package. A  $^{13}\text{C}$  metabolic model of central carbon metabolism that had been previously described [113,114](see figure 1.3) and also provided in the FiatFlux software package (specified as *mod\_yeast*) was used unchanged, including all specified ratios, stoichiometry and biomass requirements (see appendix B). The metabolic flux ratios determined with the *ratio.m* module was used in the *netto.m* module to obtain net metabolic fluxes. Multiple injections were performed from a number of samples to optimize the derivatization protocol. For each injection, 9 independent flux ratios were calculated. For each metabolic flux ratio, the ratio value with the smallest associated standard deviation was used as input in the *netto.m* module. The normalized, time-averaged uptake and production rates of external metabolites were also used as input in the *netto.m* module and net fluxes were calculated in the  $^{13}\text{C}$ -constrained flux balancing algorithm.

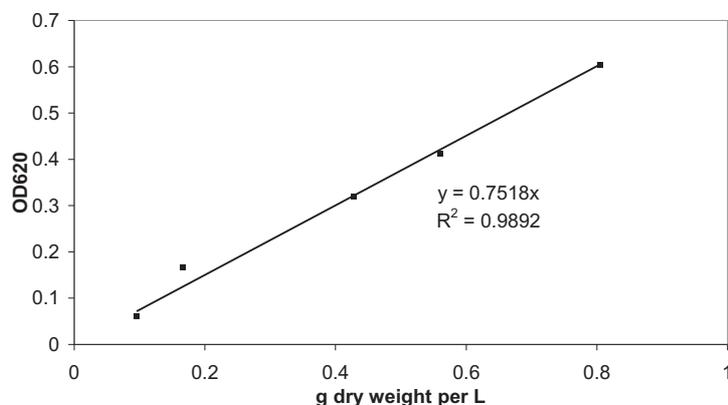


Figure 5.1: OD620 measurements plotted against cellular dry weight concentrations.

## 5.2 Experimental results and discussion

### 5.2.1 Physiological characterization of *Saccharomyces cerevisiae* growth on glucose

The use of small 20 mL cultures for economizing the experiments required samples taken for monitoring of cell growth and external metabolites to be small. The use of 620 nm instead of the conventional 600 nm was necessary due to the available optical filters for the specific microtiter plate spectrophotometer. However, the response range in the cultures at OD<sub>620</sub> conveniently allowed plate readings without any dilutions throughout the relevant exponential range. Figure 5.1 shows the linear response of OD<sub>620</sub> to values of gdw.L<sup>-1</sup>. The conversion factor  $c$  was calculated to be 1.330 gdw.L<sup>-1</sup>.OD<sub>620</sub><sup>-1</sup>.

Figures 5.2 and 5.3 show biomass concentration increase over time. The data corresponds well with exponential growth as can be seen in logarithmic scale in figures 5.4 and 5.5. Maximal specific growth rates were calculated at 0.289 .h<sup>-1</sup> for aerobic cultures and 0.343 .h<sup>-1</sup> for anaerobic cultures (See table 5.1). Anaerobic growth is usually 25 % lower than aerobic growth [117] but here the opposite was found. The value of 0.289 .h<sup>-1</sup> is significantly lower than the 0.4 .h<sup>-1</sup> found by others [113] and the reason for this is not known.

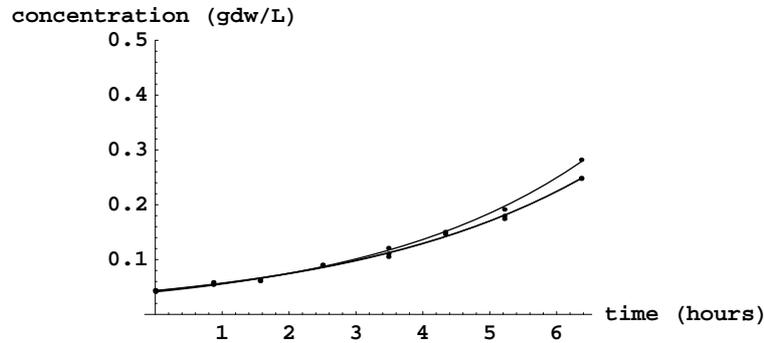


Figure 5.2: Biomass concentrations during aerobic growth. Experiments were performed in triplicate.

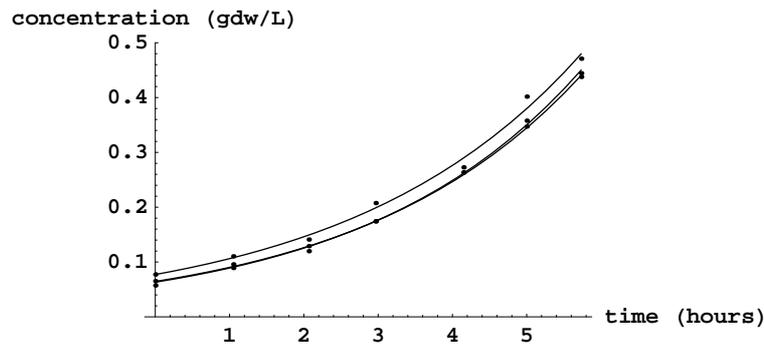


Figure 5.3: Biomass concentrations during anaerobic growth. Experiments were performed in triplicate.

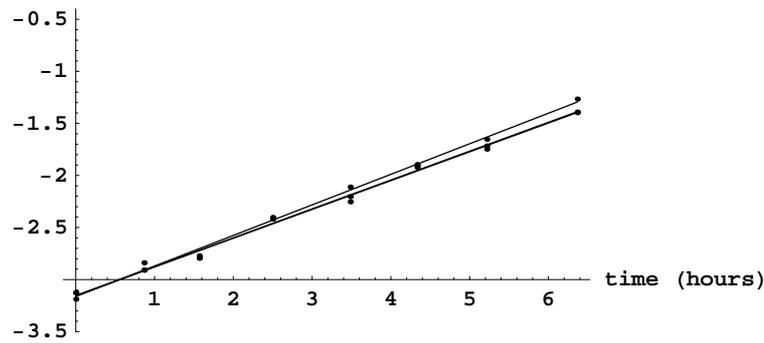


Figure 5.4: Estimation of maximal specific growth rate during aerobic growth. Experiments were performed in triplicate.

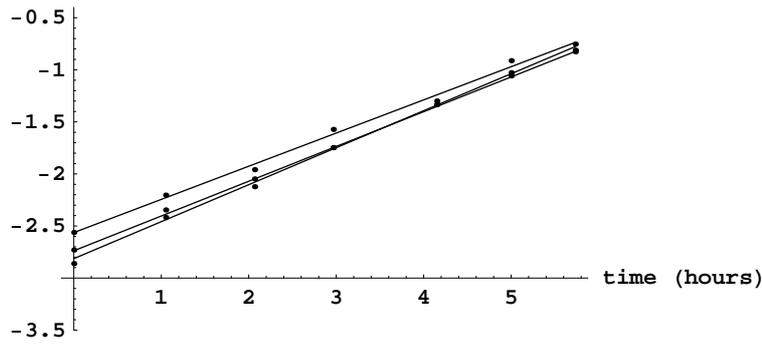


Figure 5.5: Estimation of maximal specific growth rate during anaerobic growth. Experiments were performed in triplicate.

Table 5.1: Maximal specific growth rate ( $\mu_{max}$ ) of yeast in aerobic and anaerobic cultures and biomass yield on glucose ( $Y_{Bmglc}$ ).  $Y_{Bmglc}$  was calculated as gram dry biomass produced per gram glucose consumed. Experiments were performed in triplicate.

	Aerobic growth on glucose	Anaerobic growth on glucose
$\mu_{max} (\text{h}^{-1})$	0.289	0.343
$Y_{Bmglc} (\text{gbiomass}/\text{gglucose})$	0.0979	0.0782

Table 5.2: Physiological parameters average external metabolite concentrations and average external metabolite uptake and production rates determined in aerobic and anaerobic cultures. Experiments were performed in triplicate.

	Aerobic		Anaerobic	
	Average concentration (mM)	standard deviation	Average concentration (mM)	standard deviation
glucose	26.8737	0.1584	25.9258	0.5913
ethanol	5.0466	0.0870	10.8203	0.8720
glycerol	1.2104	0.03176	1.9839	0.07974
acetate	0.6437	0.01549	0.2144	0.02924
succinate	0.4105	0.002657	N.D.	N.D.
pyruvate	0.04676	0.001159	0.2316	0.005389

	Aerobic		Anaerobic	
	Average flux ( $\text{mmol.h}^{-1}.\text{gDW}^{-1}$ )	standard deviation	Average flux ( $\text{mmol.h}^{-1}.\text{gDW}^{-1}$ )	standard deviation
glucose	-16.1024	0.4476	-23.7819	2.8424
ethanol	21.8817	0.3888	29.5872	1.1384
glycerol	4.2818	0.1108	5.3721	0.5612
acetate	1.1884	0.1844	0.7756	0.02775
succinate	0.07065	0.01458	0	N.A.
pyruvate	0.1700	0.01012	0.1571	0.02863

HPLC data of external metabolite concentrations of glucose, ethanol, glycerol, succinate, acetate and pyruvate can be seen in Appendix C, figures C.1 to C.11. Note that succinate was undetectable for anaerobic conditions and was therefore not plotted. Exponential curves could be fitted for glucose, ethanol and glycerol with excellent correspondence for both aerobic and anaerobic cultures. Fits through acetate, succinate and pyruvate were not as good which was indicative of either changes in the culture or quantification problems due to the low concentrations. Accordingly, linear equations had to be fitted through the data points of aerobic succinate production and anaerobic pyruvate production. For use in kinetic models, the averaged concentration over the initial phase before  $^{13}\text{C}$  sampling was plotted as horizontal lines over the interval, which are the values reported in table 5.2.

Time-dependent gradients of the curves were normalized with the equations for dry weight concentration, with their averaged values over the initial

time interval as seen in figures C.12 to C.22 in Appendix C. The uncertainties involved in the apparent non-exponential production rates of acetate, succinate and pyruvate can be seen more clearly in this representation. Results are summarized in table 5.2. The specific glucose uptake value of  $16.1 \text{ mmol.h}^{-1}.\text{gdw}^{-1}$  under aerobic conditions is the same as measured by [113] at the same pH of 6.0. Biomass yield on glucose was estimated at  $0.0979 \text{ g.g}^{-1}$  for aerobic conditions (See table 5.1) which is lower than the  $0.16 \text{ g.g}^{-1}$  found by [113]. Glycerol production was  $4.28 \text{ mmol.h}^{-1}.\text{gdw}^{-1}$  which is three times the value of  $1.7 \text{ mmol.h}^{-1}.\text{gdw}^{-1}$  found by [113]. For anaerobic conditions, biomass yield was  $0.0782 \text{ g.g}^{-1}$ . The slightly faster and more wasteful anaerobic metabolism of *Saccharomyces cerevisiae* is apparent from the faster glucose uptake and lower biomass yield on glucose  $\mathbf{Y}_{Bmglc}$ . Note that the carbon balance is implicit in the  $^{13}\text{C}$ -constrained MFA algorithms and that it can be accessed in the balances around each intermediate, therefore it is not done in the traditional way.

### 5.2.2 GC-MS analysis of labelling patterns in proteinogenic amino acids.

For GC-MS analysis of  $^{13}\text{C}$  labelled amino acids, cells were grown either on 100 %  $1\text{-}^{13}\text{C}$  glucose or a combination of 20 %  $\text{U-}^{13}\text{C}$  glucose and 80 % glucose of natural labelling and harvested in early-exponential phase at biomass concentrations between  $0.1$  and  $0.2 \text{ g.L}^{-1}$ . After hydrolysis and drying, the samples were derivatized with MTBSTFA and DMFA.

Figure 5.6 is a plot generated with Matlab showing the elution pattern of amino acids in GC-MS. The longer than usual 60 m GC column provided extremely efficient separation of the derivatized amino acids and simplified the analyte assignment procedure, although the number of MS spectra for each amino acid was reduced. However, this number was considered to be sufficient for observation of labelling patterns. Figures 2.3 to 2.5 used in Chapter 2 for explanatory purposes also originated from these experiments.

The *ratio.m* module of FiatFlux was used to calculate metabolic flux ratios, as was explained in Chapter 2. Due to the use of a 60 m GC column, which

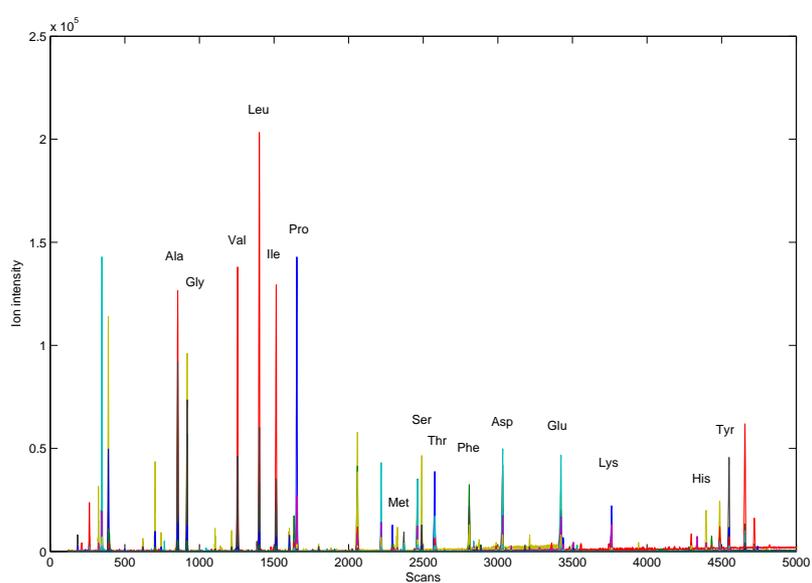


Figure 5.6: A typical GC-MS chromatogram obtained in this study. The aerobic or anaerobic yeast samples were hydrolyzed and derivatized with MTBSTFA. The 60 m column provided excellent resolving power. See section 5.1.5 for GC-MS parameters.

is twice the length normally used for analysis in FiatFlux (as in [92]), the default GC elution range for each amino acid was adapted. The sequence in which derivatized amino acids eluted from the GC column was identical as reported earlier [92]. It was found that the presence of some amino acids in the derivatized form was insufficient for assignment. Apart from a number of amino acids that are destroyed during acid hydrolysis (see [92]), the signals from especially methionine, histidine and tyrosine were weak. In order to observe all the amino acids present, we used different volumes of MTBSTFA and DMFA in the derivatization step and injection volumes of 0.5 to 2  $\mu\text{L}$ .

While overloading inevitably eliminates the use of those spectra affected by overloading, and its inherent danger to the electron source, the use of larger injection volumes (2  $\mu\text{L}$ ) and minimized volumes of the derivatization agents enabled better observation of low concentration amino acids. However, mostly the problems encountered with observation of amino acids could be ascribed to the sample preparation. It was found that all samples which showed murkiness after derivatization, indicative of unsolubilized compounds, did not contain sufficient amounts of derivatized compounds. Subsequent addition of more MTBSTFA and DMFA to these samples and incubation also did not solubilize and derivatize the compounds. It is thus indicative that these samples contained a substance that caused these results.

Water could have had this effect as it is well known to influence derivatization with MTBSTFA. Care was thus taken to use small volumes of the solution of hydrolyzed cell mass and extensive drying at 85  $^{\circ}\text{C}$  before addition of the derivatization agent. Isotope effects were also clearly observed, as was reported in [92]. In this phenomenon, the heavier isotopes move faster through the GC column and thus could result in the wrong isotope ratio calculation if all the MS spectra of a certain compound are not included in the calculation. It was assumed that it was correctly interpreted in FiatFlux.

For each GC-MS analysis, FiatFlux automatically calculated independent metabolic flux ratios from amino acid spectra. For each ratio, a standard deviation is assigned, as was explained in Chapter 2 (see [87]). Small stan-

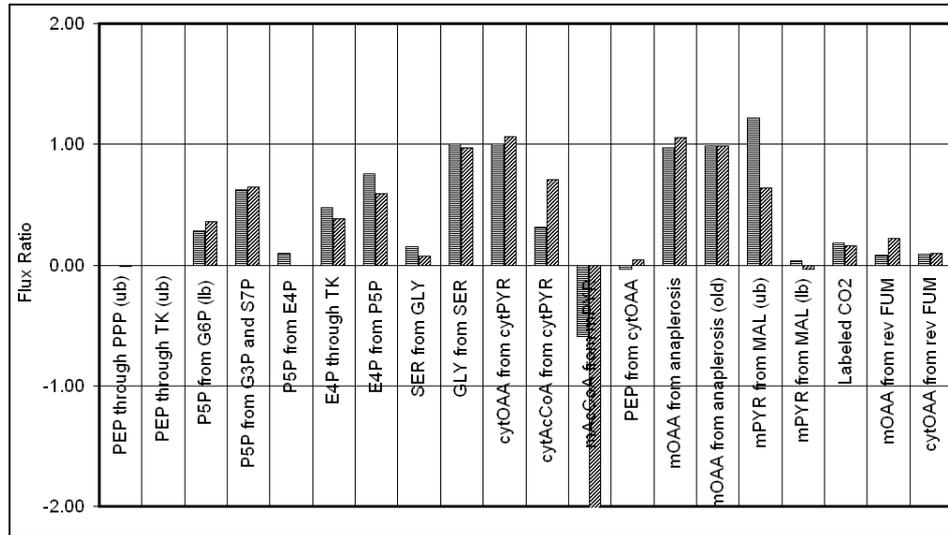


Figure 5.7: Ratios from aerobic cultivation with 20 %  $\text{U-}^{13}\text{C}$  glucose and 80 % naturally labelled glucose. 20 mL foil-capped cultures with rapid shaking were performed in duplicate. Ratio calculations were performed in FiatFlux, as explained in sections 5.1.7 and 2.1.

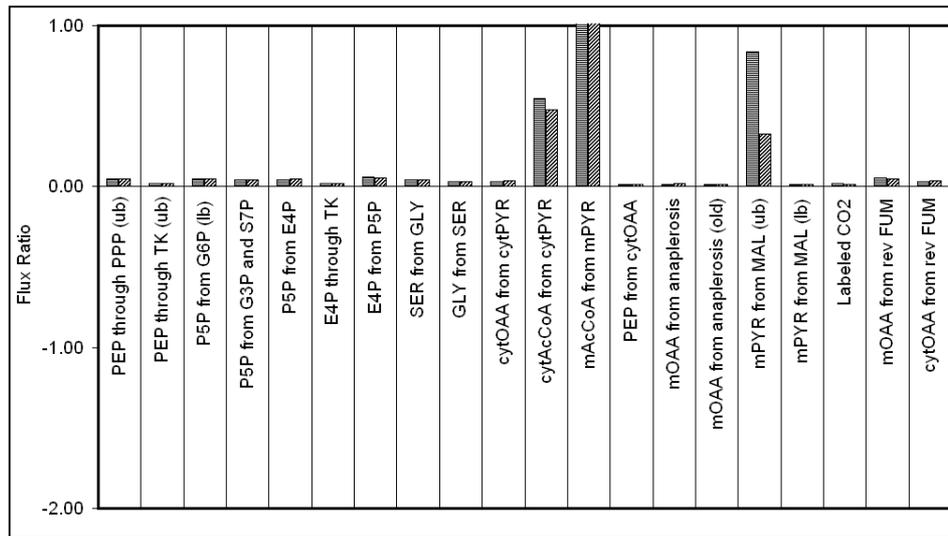


Figure 5.8: Ratio standard deviations for aerobic cultivation with 20 %  $\text{U-}^{13}\text{C}$  and 80 % naturally labelled glucose. Standard deviations correspond to ratio values in figure 5.7 and were estimated as explained in section 2.1.

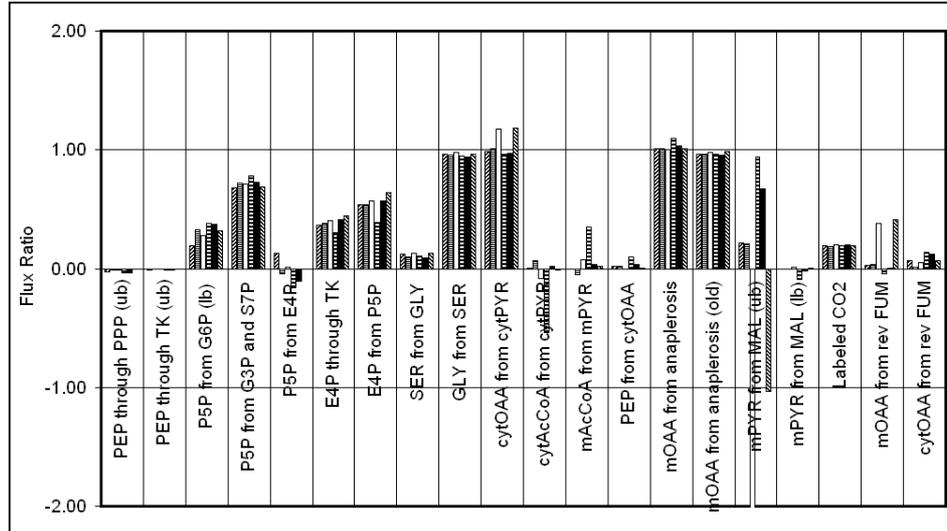


Figure 5.9: Ratios from anaerobic cultivation with 20 %  $\text{U-}^{13}\text{C}$  and 80 % naturally labelled glucose. 20 mL nitrogen flushed cultures were performed in triplicate, and two injections from each were used for optimizing the derivatization protocol. Ratio calculations were performed in FiatFlux, as explained in sections 5.1.7 and 2.1.

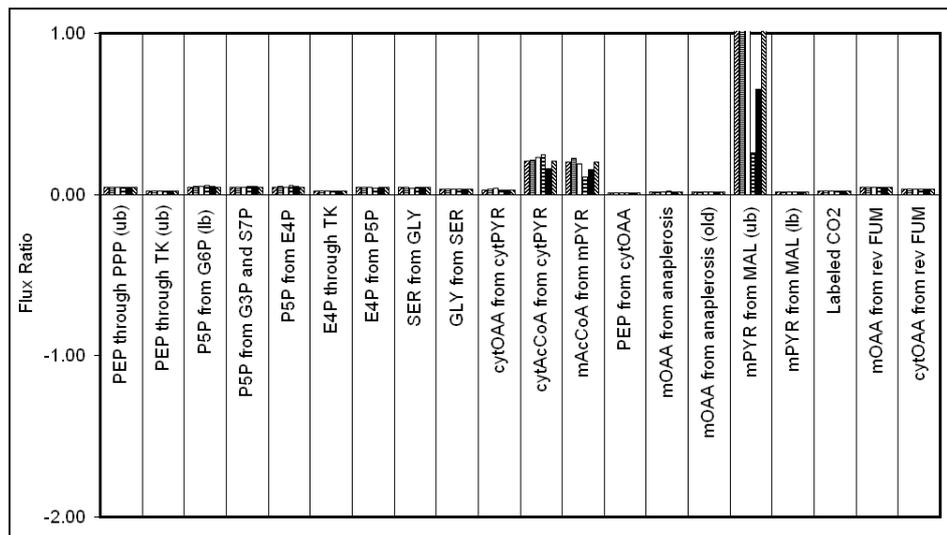


Figure 5.10: Ratio standard deviations for anaerobic cultivation with 20 %  $\text{U-}^{13}\text{C}$  and 80 % naturally labelled glucose. Standard deviations correspond to ratio values in figure 5.9 and were estimated as explained in section 2.1.

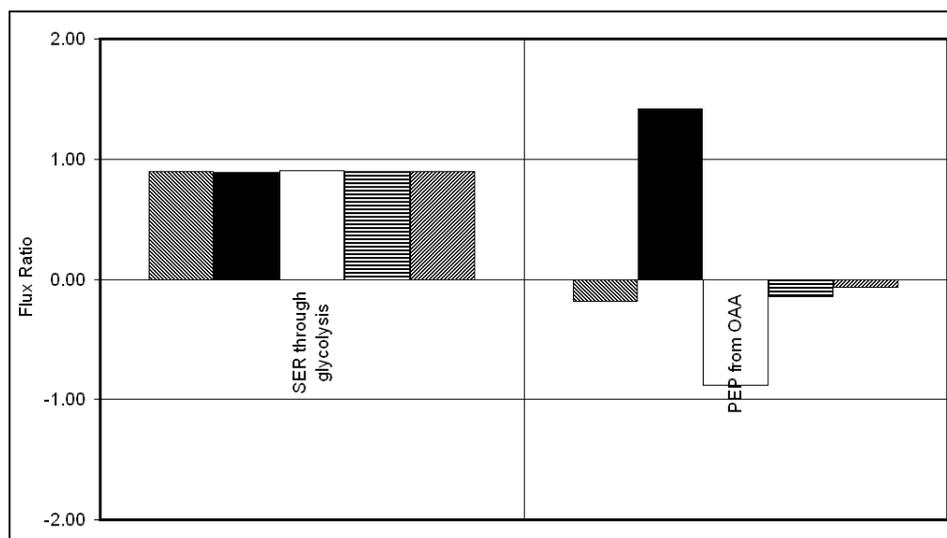


Figure 5.11: Ratios from aerobic cultivation with 100 %  $1\text{-}^{13}\text{C}$  glucose. Conditions were as for figure 5.7.

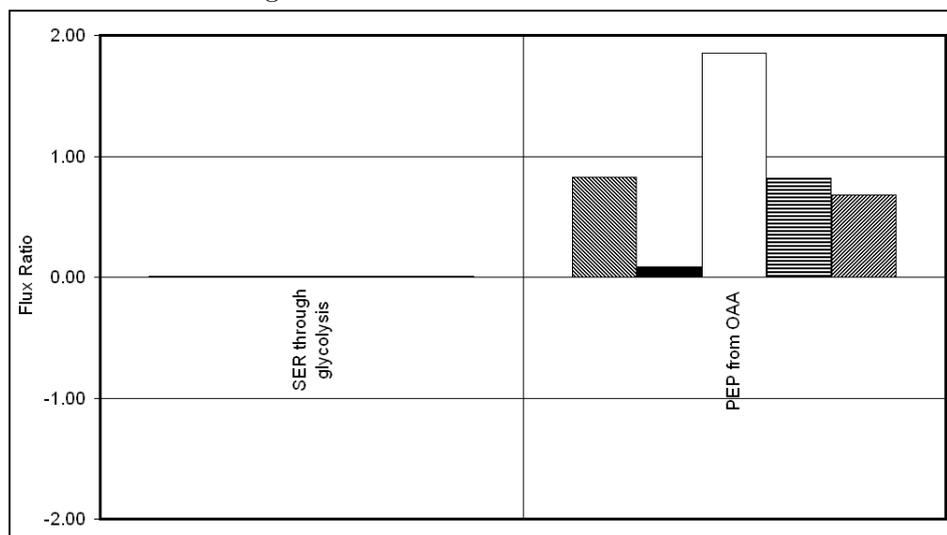


Figure 5.12: Ratio standard deviations for aerobic cultivation with 100 %  $1\text{-}^{13}\text{C}$  glucose. Standard deviations correspond to ratio values in figure 5.11.

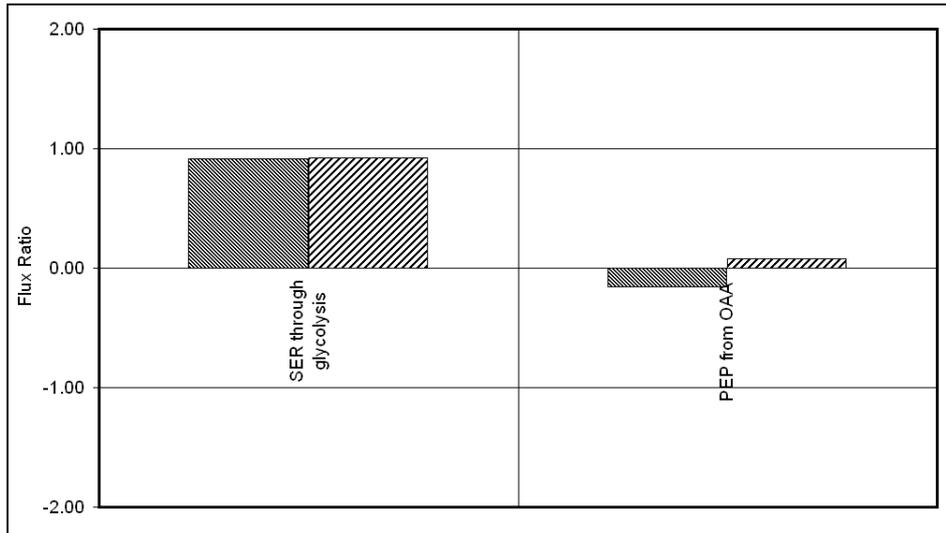


Figure 5.13: Ratios from anaerobic cultivation with 100 %  $1\text{-}^{13}\text{C}$  glucose. Conditions were as for figure 5.9.

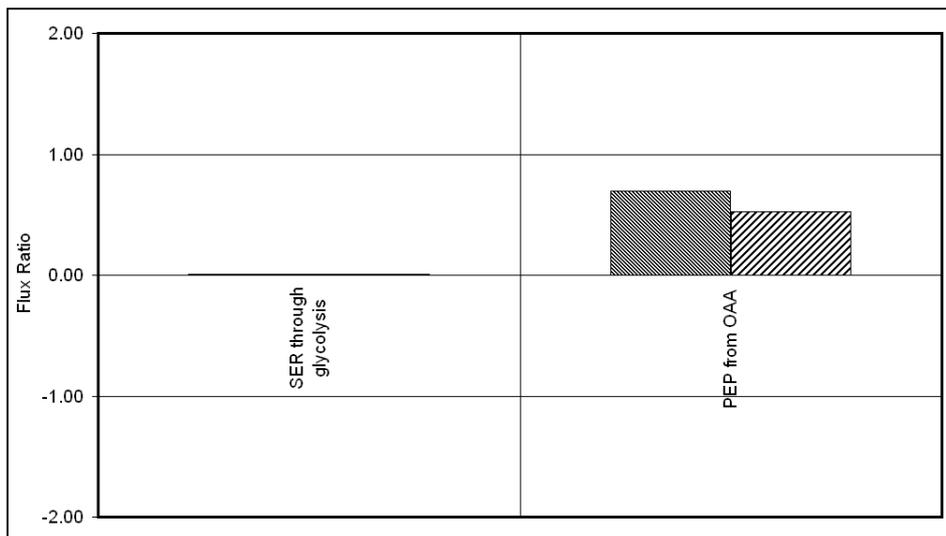


Figure 5.14: Ratio standard deviations for anaerobic cultivation with 100 %  $1\text{-}^{13}\text{C}$  glucose. Standard deviations correspond to ratio values in figure 5.13.

standard deviations correspond to high confidence in the measurements and the flux model. Results are given in figures 5.7 to 5.14.

In figures 5.9 to 5.12, the results of several injections can be seen. Due to the fact that ratios are independently calculated, for each ratio, the value was used which had the smallest error associated with it. The independence of ratios thus enabled the use of calculated ratios from different injections, where each injection had theoretically the best procedures for observing the relevant ratios, where a ratio is a function of isotope distributions of several amino acids. This practice is however only valid if it is assumed that the ratio with the lowest error is the one closest to theoretically perfect analysis and that the network assumptions are correct. The network could however be assumed as sufficient as it has been used extensively under these conditions [114]. Further, it is apparent from the results of repetitive injections that the ratio values which showed consistent values for all samples also had the smallest errors associated and the highly variable ratios had the larger errors associated. For instance, figure 5.11 and 5.12 shows this relationship clearly, thus partly validating this practice.

Using multiple injections for each ratio is however not optimal, and will likely be unnecessary, or greatly reduced if more time is spent on standardization of the sample treatment setup. Only the ratios calculated previously [113] were selected and the final values with their errors are given in table 5.3. In general, ratios correspond well with those reported elsewhere [113]. The ratio of approximately 90 % of the total carbon flux through the EMP pathway is typical and is likely a ratio of high importance for integration with a kinetic model of glycolysis. This value was found for both aerobic and anaerobic conditions.

### 5.2.3 Net flux analysis with FiatFlux

Ratio values from table 5.3 and the calculated averaged external fluxes in table 5.2 were used in the *netto.m* module of FiatFlux to estimate absolute fluxes using the *mod\_yeast* model. The  $^{13}\text{C}$ -constrained flux balancing

Table 5.3: Metabolic Flux Ratios calculated for aerobic and anaerobic cultures with standard deviations.

	Aerobic		Anaerobic	
	Ratio	SD	Ratio	SD
cytOAA from cytPYR	1.001	0.031	0.959	0.027
mOAA from anaplerosis	0.971	0.016	1.006	0.016
PEP from cytOAA	0.042	0.011	0.095	0.011
SER through glycolysis	0.899	0.011	0.921	0.011
mAcCoA from mPYR	1	1	0.346	0.106
mPYR from MAL (ub)	0.642	0.329	0.936	0.258
mPYR from MAL (lb)	0.035	0.014	0	0.015
SER from GLY	0.152	0.04	0.127	0.041
GLY from SER	0.974	0.033	0.948	0.032

algorithm repetitively converged to the solution for net fluxes as given in table 5.4 and the balance and biomass fluxes given in table 5.5. The linear optimization algorithm seeks to minimize the metabolite balances, using the ratio values as extra constraints, either as specific ratio values, or as limits for ratios. The limits are thus included not only as = constraints, but also as  $\geq$  and  $\leq$  constraints which cannot be used in matrix-based analyses (see Chapter 2, [87] and [95]). The residuals of the metabolite balances after a conversion to a local or global minimum is then reported as the individual metabolite balances as in table 5.5. These small balances were treated as the fluxes of other small reactions around these metabolites and included together with biomass reactions from the same metabolites into the simulation of the kinetic model (see Chapter 4). The optimization routine uses random starting values where the user initializes the optimization several times until a very small total residual (typically below 10) is achieved. For our experiments a total residual of below 1 was achieved. Other local minima were detected, but with much larger residuals after conversion. In particular, cycling was found between serine and glycine. This is a mathematical artefact with a closed loop of fluxes that brings no net change [21].

In order to obtain a calculable equation set, NADH had to be balanced, but neither did NADPH or ATP have to be balanced. An advantage of  $^{13}\text{C}$ -MFA is that the calculation of metabolic fluxes can be done largely independent from the assumed constraints that cofactors place on the system.

Table 5.4: Metabolic net fluxes calculated for aerobic and anaerobic cultures with absolute errors (95 % confidence interval).

Reaction	activity	reversible	Aerobic		Anaerobic	
			Flux $\frac{\text{mmol}}{\text{h}\cdot\text{gDW}}$	SD	Flux $\frac{\text{mmol}}{\text{h}\cdot\text{gDW}}$	SD
v1glc uptake	1	0	16.104	0.448	20.694	2.843
v2PGI	1	1	13.851	0.442	18.292	2.628
v3ZWF	1	0	1.700	0.176	1.743	0.314
v4PFK	1	0	14.889	0.443	19.340	2.770
v5TKL	1	1	0.561	0.059	0.574	0.105
v6TKL	1	1	0.477	0.059	0.475	0.105
v7TAL	1	1	0.561	0.059	0.574	0.105
v8SER	1	0	0.148	0.119	0.172	0.134
v9SHM	1	0	0.115	0.140	0.127	0.154
v10SHM	1	0	0.026	0.023	0.025	0.022
v11GCV	1	1	0.003	0.118	0.007	0.133
v12TDH	1	1	25.720	0.911	33.690	5.644
v13PYR	1	0	26.680	1.000	37.025	6.253
v14PDA	1	0	0.467	1.120	0.212	1.995
v15CIT	1	0	0.420	1.018	0.549	5.763
v16ACO	1	0	0.420	1.018	0.549	5.763
v17KGD	1	0	0.131	1.021	0.202	5.764
v18SDH	1	1	0.061	1.022	0.202	5.764
v19MDH	1	1	0.012	0.030	0.000	0.038
v20FUM	1	1	0.061	1.022	0.202	5.764
v21MAE	1	0	0.049	0.992	0.202	5.801
v22PCK	1	0	1.127	0.311	3.534	0.745
v23	1	0	1.921	1.068	4.544	6.396
v24ACS	1	0	0.412	0.470	0.885	3.768
v25ALD	1	1	1.600	0.507	1.661	3.768
v26ADH	1	1	21.880	0.389	29.823	1.138
v27GPD	1	1	4.282	0.111	5.169	0.561
v28OAC	1	0	0.408	0.990	0.744	6.074
v29OAC	1	0	0.000	0.059	0.195	0.304
v30	1	0	0.000	0.467	0.393	3.769
v31	1	0	1.046	0.478	0.763	3.812
v32Atmung	1	0	0.384	2.549	0.000	12.531
v33	1	0	23.480	0.639	31.484	3.715
v34TH	0	0	0.000	0.000	0.000	0.000
growth	1	1	0.289	0.006	0.343	0.020

Table 5.5: Metabolite balances and biomass fluxes from net flux analysis for aerobic and anaerobic cultures.

Metabolite	balanced	Aerobic		Anaerobic	
		balance $\frac{\text{mmol}}{\text{h}\cdot\text{gDW}}$	to biomass $\frac{\text{mmol}}{\text{h}\cdot\text{gDW}}$	balance $\frac{\text{mmol}}{\text{h}\cdot\text{gDW}}$	to biomass $\frac{\text{mmol}}{\text{h}\cdot\text{gDW}}$
AcAl	1	0	0	0	0
Ace	1	0	0	0.001	0
ATP	0	15.061	3.306	17.012	3.874
C1	0	0.086	0	0.095	0
CIT	1	0	0	0	0
CO <sub>2</sub>	0	25.976	-0.526	34	-0.616
cytAcCoA	1	0	0.412	0.003	0.483
cytOAA	1	0	0.386	0.003	0.452
cytPYR	1	0	0.063	0.001	0.074
E4P	1	0	0.084	0	0.098
ethanol	1	-0.002	0	0.236	0
F6P	1	0	0	0	0
FUM	1	0	0	0	0
G6P	1	0	0.553	0.002	0.648
glucose	1	-0.001	0	3.088	0
GLY	1	0	0.092	0	0.108
glycerol	1	0	0	-0.203	0
mitAcCoA	1	0	0.047	0	0.055
MAL	1	0	0	0	0
mitOAA	1	0	0	0	0
mitPYR	1	0	0.628	0.008	0.736
NADH	1	0	-0.062	0	-0.073
NADPH	0	1.791	3.256	1.485	3.816
O <sub>2</sub>	0	-0.384	0	0	0
OGA	1	0	0.288	0.005	0.338
P5P	1	0	0.102	0	0.119
PEP	1	0	0.168	0.001	0.196
S7P	1	0	0	0	0
SER	1	0	0.059	0	0.069
SUCC	1	0	0	0	0
T3P	1	0	0.105	0	0.123
GROWTH	1	0	-0.289	0.004	-0.339

Cofactors, especially ATP are used in multiple reactions and with unknown stoichiometry, which would introduce a lot of uncertainty to the calculation of fluxes. Also was balancing of one-carbon units (C1), carbon dioxide and oxygen not necessary. The large balances for ATP and NADPH are thus not due to small unknown reactions, but were interpreted as their calculated net consumption rates due to a multitude of reactions outside the flux model.

It was concluded that the FiatFlux software was user friendly and the automated layout of the program was especially well-designed for processing multiple datasets. A few minor problems were encountered with reports to *.xls* format in *netto.m* of FiatFlux version 1.5. Further, the *netto.m* did not detect the flux model used in the *ratio.m* module automatically. The origins were detected and the code adjusted for our own use.

## Chapter 6

# Flux-constrained simulation

### 6.1 Details of the simulation

The flux-constrained methodology was put to the test by using the results of our experiments. The experimental procedures were discussed in Chapter 5 and results were given in tables 5.4 and 5.5. The big question was whether the kinetic model of Teusink, which had been constructed for a different strain (industrial Koningsgist) could be used to describe metabolism in the strain CEN.PK 113-7D, using the same enzyme parameters. Further, maximal enzyme activities ( $V_{max}$ ) were not measured in our study.  $V_{max}$  values can be expected to vary between different strains and conditions, therefore they will inevitably have to be measured. It can be expected however, that other enzyme parameters should not change as significantly as  $V_{max}$  values. Also, concentrations of metabolites were not measured, therefore the pool sizes of conserved moieties were taken to be the same as in the original model. Very important for this study would be that parameters should not be fitted to obtain simulation results that represent the physiological data, as is frequently encountered in the literature. The essential difference is here that the mechanistic *in vitro* measurements are tested.

Our software program, Kinomics, was used to integrate the kinetic model from the JWS with the net fluxes, balance fluxes and biomass fluxes obtained for aerobic and anaerobic conditions. The measured concentrations of ex-

tracellular glucose, ethanol and glycerol from table 5.2 were used as fixed concentration boundary conditions, whereas external succinate, acetate and pyruvate concentrations do not feature in the kinetic model.

The robustness that the program provides enables one to extract the most realistic parts of a kinetic model, while less realistic parts can be replaced either by flux and concentration measurements, or more realistic models. The Teusink model features many realistic mechanistic expressions for enzyme reactions, with a few reactions of either flux data values or phenomenological expressions. The two branch reactions for glycogen and trehalose synthesis from glucose-6-phosphate had originally been modelled by the authors as constant measured boundary reactions, as is proposed for a wider application in this thesis. Also, phenomenological rate equations for succinate synthesis and ATP consumption were included as first order kinetics responding to their substrates succinate and ATP. The origin of succinate was unknown in that study, therefore it was assumed that it was derived through the glyoxylate cycle. The succinate producing rate equation was made responsive to only acetaldehyde, but acetaldehyde, NAD and ADP were consumed, while succinate, NADH and ATP was produced. A general ATPase was also included to describe all other reactions that consume ATP, such as biomass requirements and maintenance.

We can now measure metabolic fluxes more accurately by combining known biomass requirements with  $^{13}\text{C}$  isotope labelling experiments and thus elucidate the source of the succinate and more accurately calculate the ATP consumption flux. The *context dependent* rate equations for glycogen, trehalose, succinate and ATPase could thus be replaced with their accurately measured fluxes, while still using the detailed enzyme kinetics of the other reactions. The stoichiometry of the succinate reaction vSuc was thus replaced by the larger network stoichiometry, which did not feature the glyoxylate cycle, as was validated for these conditions in [114]. It is thus a case of replacing context dependent reaction blocks with measured reaction fluxes for the reference condition, while conserving detailed parts.

The ATPase in the kinetic model (vATP) was set to zero with manual

transforms, as the boundary reactions that use and produce ATP (measured fluxes, biomass fluxes and balance fluxes) account for the ATPase. The glycogen and trehalose producing reactions  $v_{\text{GLYCO}}$  and  $v_{\text{Treha}}$  were also set to zero as the biomass reactions that consume G6P account for the storage metabolites produced from G6P. Likewise,  $v_{\text{Suc}}$  was also set to zero.

The first assessment of the performance of the kinetic model was to see whether a steady state was reached upon numerical simulation. The algorithms described in Chapter 4 were used to obtain the conserved moieties in the kinetic model. Conserved concentration moieties existed for the adenines

$$\textit{Adenines} = \textit{ATP} + \textit{ADP} + \textit{AMP} \quad (6.1)$$

and the nicotinamides

$$\textit{Nicotinamides} = \textit{NADH} + \textit{NAD}^+ \quad (6.2)$$

To fulfil the concentration moiety constraints, the same values for the constituents were used as initial values as was found for the model in the JWS. The moiety balance checking algorithm was used to make sure that the net constraints on the moieties were exactly zero.

First, simulations were carried out without setting of kinetic model pathway fluxes with *setfluxes* (see Chapter 4). No steady state was reached as there was a continued increase in the F16BP concentration.

## 6.2 Investigating the cause of model behaviour

It was investigated whether the values used to convert the values from  $^{13}\text{C}$ -constrained flux analysis ( $\text{mmol}\cdot\text{h}^{-1}\cdot\text{gdw}^{-1}$ ) to the units used for kinetic modelling ( $\text{mmol}\cdot\text{min}^{-1}\cdot\text{L}^{-1}$ ) could have been the cause of an overly constrained system of rate equations. Here  $\text{gdw}$  is for the number of grams total cellular dry weight in an amount of cells, and  $\text{gprot}$  is the number of grams of total cellular protein in an amount of cells.  $\text{L}$  is the number of litres of intracellular water in an amount of cells. The conversion was calculated as

follows:

$$\frac{mmol}{min L} = \left(\frac{mmol}{h\ gdw}\right)\left(\frac{h}{min}\right)\left(\frac{gdw}{gprot}\right)\left(\frac{gprot}{L}\right) \quad (6.3)$$

where the value

$$\frac{L}{gprot} = 0.00375$$

was also used in the original model for calculations [44]. The value

$$\frac{gprot}{gdw} = 0.5$$

is typically used for yeast, but values could vary from 0.243 to 0.497 for aerobic cultivation and from 0.254 to 0.559 for anaerobic cultivation. These estimations were carried out in chemostats under various conditions of carbon to nitrogen limitation, with or without addition of benzoic acid [118]. The values of 0.5 gprot.gdw<sup>-1</sup> and 0.00375 L.gprot<sup>-1</sup> leads to a conversion factor of

$$\left(\frac{1}{60}\right)\left(\frac{1}{0.5}\right)\left(\frac{1}{0.00375}\right) = 8.89$$

and thus a glucose uptake flux in aerobic cultures of

$$(16.104)(8.89) = 143.17 \frac{mmol}{min L}$$

and

$$(20.694)(8.89) = 183.97 \frac{mmol}{min L}$$

for the anaerobic cultures. These values are significantly higher than the 108 mmol.min<sup>-1</sup>.L<sup>-1</sup> observed in the glucose runout experiments for initial validation of the kinetic model [44]. This finding, but more so, the uncertainties involved with the values for L.gprot<sup>-1</sup> and gprot.gdw<sup>-1</sup> indicated that we should focus more on these parameters. Thus, the conversion factor was varied over a wide range to test whether a steady state was reachable. Only a decrease by a factor of ten allowed conversion to a steady state. This observation indicated that a steady state was possible, and that the integrated structure is not erroneous from a *structural* perspective. A ten-fold error in the conversion factor is rather unlikely though, and other possible sources of error were investigated.

The comprehensive comparative output *.xls* spreadsheets were used to anal-

Table 6.1: Maximal activity values compared to metabolic net fluxes in aerobic and anaerobic cultures.

Vmax		MFA	Aerobic		Anaerobic	
		reaction	flux	$\frac{flux}{V_{max}}$	flux	$\frac{flux}{V_{max}}$
VmGLT	97.26	Measv1glcuptake	143.16	1.47	183.97	1.89
VmGLK	226.45	Measv1glcuptake	143.16	0.63	183.97	0.81
VmPGI	339.68	Measv2PGI	123.14	0.36	162.62	0.48
VmPFK	182.90	Measv4PFK	132.36	0.72	171.93	0.94
VmALD	322.26	Measv4PFK	132.36	0.41	171.93	0.53
VmG3PDH	70.15	Measv27GPD	38.07	0.54	45.95	0.66
VmPGM	2525.81	Measv12TDH	228.65	0.09	299.50	0.12
VmENO	365.81	Measv12TDH	228.65	0.63	299.50	0.82
VmPYK	1088.71	Measv13PYR	237.19	0.22	329.15	0.30
VmPDC	174.19	Measv33	208.74	1.20	279.89	1.61
VmADH	810.00	Measv26ADH	194.51	0.24	265.13	0.33
VmGAPDHf	1184.52	Measv12TDH + Measv8SER	229.96	0.19	301.02	0.25
VmPGK	1306.45	Measv12TDH + Measv8SER	229.96	0.18	301.02	0.23

use firstly whether the maximal activity values in the model allowed a steady state comparable to the measured fluxes. Indeed, the maximal activities of the glucose transporter vGLT and pyruvate decarboxylase vPDC were too low to sustain the measured fluxes (table 6.1). Further, a number of other  $V_{max}$  values were very close to the measured fluxes, indicating that even with very low concentrations of their products, the forward reactions could hardly produce the measured fluxes and it is likely that their substrates would build up to high levels.

The next step would be to either measure maximal activities of the enzymes for this strain and conditions or obtain more realistic values from the literature. Further analysis without this research would be futile in principle. We thus chose not to fit maximal activity values because then we would lose the context independence of the modelling process. The conversion factors also need more attention. This would likely involve measurement of cellular protein concentrations, while the amount of cell water per unit of protein is more difficult to measure and would likely be taken from the literature. Further, as flux experiments have a limited capacity for model accuracy assessment, a technology driven approach of metabolomics is needed to work in combination with flux measurements. Just like flux measurements, concentration measurements are essential as boundary values and for validation,

but also for inputs as moiety pool sizes. Only with this in place can we proceed with the protocol of localized validation with isotope labelling and metabolomics, as was introduced in Chapter 3.

## Chapter 7

# Discussion and conclusion

Genome-scale kinetic modelling is seen as a powerful method for making sense of biochemical phenomena. These models can give deep insight into the regulation of cellular processes. Unfortunately, the parameters for detailed kinetic models are difficult to obtain, which limits both their size and the confidence in their predictions. Alternatively, stoichiometric models can be created at the genome scale. Frameworks for their analyses such as FBA [14] and advanced optimization programs [22] have been developed. Analysis of stoichiometric models with FBA lacks in the fact that it does not incorporate the nonlinear behaviour of biochemical systems. Nonlinear behaviour can only be captured with kinetic models, which incorporate the concentrations of metabolic species as variables. It might seem an obvious solution for the near future to combine large stoichiometric models with kinetic models in some way.

The approach taken here was to use a large structural model, not for increasing the scope of predictions by including a wider metabolic network, but as a flux data model that facilitates validation of kinetic parts. The work described here is a part of the Silicon Cell Initiative, aimed at a parts-based genome-scale kinetic description of the yeast. The idea is to eventually create the whole from kinetic parts, built 'on top of' a genome-scale pathway map. Extensive accuracy assessments of these detailed parts are necessary before they can be linked. Apart from the general idea of combining structural

and kinetic models, two more direct aims were identified for this project, namely:

1. To learn the techniques in  $^{13}\text{C}$ -MFA.
2. To investigate how the boundary reactions should be incorporated for steady state kinetic modelling.

Chapter 3 explained why it is important to include measured boundary flux data for validation. It is both necessary for putting the correct constraints on the kinetic parts (from there the name *flux-constrained simulation*) and can be used for direct comparison. A generic theoretical approach was followed for incorporating boundary fluxes and a number of implications were discussed in Chapter 3.

It was explained that boundary fluxes alter MCA predictions, but that they have to be included in an MCA algorithm, as they are necessary to obtain accurate steady state simulations. Suggestions were made to design an MCA algorithm that incorporates uncertainties of kinetics at the boundary reactions, which would resemble an existing formalism [63] to an extent.

The many variables in a kinetic model and its integrated nature make it difficult to locate dubious parameters. It was argued that flux relations create a large degree of redundancy in the data, decreasing the validation potential of flux data. Also, rate equations have maximal activity values which put further constraints to the simulated values, and make falsely accurate simulations more likely. Rather, the validation criteria should rely more on metabolomics data, where the measured fluxes should be used as inputs in some way. It was shown how this can be done by setting pathway fluxes to measured values. A certain number of fluxes can be set, whereas additional settings will cause an artefact. There are limitations to his approach, as was pointed out. Not all pathways can be set to measured fluxes. Further, the integrated nature of a kinetic system makes it likely that the dubious parameters cannot be located. Others have separated kinetic systems into smaller subsystems to fit parameters in a more local fashion on time series

data [52]. Our objective was to develop a similar idea for local accuracy assessment on flux data. Fluxes through small kinetic subsystems can always be set to measured values. The idea is flexible with a number of criteria with regard to the availability and accuracy of metabolomics data.

Moiety constraints have implications for the use of boundary fluxes. The summed boundary fluxes acting on one moiety must equal zero, otherwise a steady state cannot be reached. They also cannot be left out, for they have an effect on the interconversion in moieties. Others have described how moieties are correctly modelled [51, 101], where the total moiety concentration has to be measured, but they have not discussed whether the inclusion of measured fluxes alter this need for data. From structural analysis it might seem obvious that these boundary fluxes eliminate the moiety. However, the fact that these boundary flux values must always cancel out around the moiety, implicates that the moiety is still a feature of the kinetic simulation. It inevitably means that the total moiety concentration has to be measured.

Validation on steady state flux and concentration data is limited due to the fact that only a few data points are used to link measured values to enzyme parameters. Dynamic pulse experiments which produce time series data will be a main feature in the validation procedure of the Silicon Cell. These experiments have their own technical challenges at this stage, but they are being solved by advances in analytical technologies. Of more importance for this discussion is that a number of fundamental problems with pulse experiments limit the information that they can provide. Even with the best analytical technologies these problems will be present. Of note is that the reversibility of enzyme reactions results in too many enzyme parameters to observe for the number of variables measured. With regard to this, it was pointed out that  $^{13}\text{C}$ -MFA has the potential to simultaneously provide forward and reverse fluxes. It can thus limit the feasible parameters further than net fluxes can. It was proposed that isotope labelling in both steady state metabolomics and dynamics pulse experiments will be very important for detailed kinetic modelling in the future.

To summarize, this chapter gave and elaborate discussion on the inclusion

of boundary fluxes into kinetic simulations. The second of the two aims were thus met. A venture was taken into the validation potential of flux data. It was explained that fluxes have fundamental limits in its validation potential, but that it still has high value, especially in the view that it can complement dynamic pulse experiments.

Based on our findings in Chapter 3, a simulation program was designed which allows for robust flux-constrained simulation. The program, called Kinomics, was described in Chapter 4. It has proven very useful thus far, although it is still rather rudimentary. It was proposed how the program will be extended for easier use and to incorporate more functionality related with bioinformatics. Although there are a number of excellent standalone simulation programs in the literature [107–112], this program seems to be unique in allowing robust connection between kinetic models and stoichiometric models.

To illustrate our approach we measured metabolic fluxes in the yeast at a quasi steady state with  $^{13}\text{C}$ -constrained flux balancing, using GC-MS and the flux analysis program FiatFlux. The experimental procedures have been well-defined by others such as the group of Sauer [113], which made the work manageable in the time scale of the project. Results were given in Chapter 5. Valuable experience has been gained with the techniques and theory behind  $^{13}\text{C}$ -MFA, which are new to our research group. The knowledge obtained which was briefly outlined in Chapter 2, together with the good quality of the data suggest that the first of our two aims has also been achieved.

Our program was then used to constrain a detailed kinetic model of yeast glycolysis [44] with the flux data from our experiments (Chapter 6). The less detailed and context dependent reactions in the kinetic model were replaced by accurate flux measurements to assess the accuracy of simulation results. The question was whether the *in vitro* kinetics of the Teusink model, which was constructed for a different strain and conditions, could be used to describe the flux data in our experiments. Early on it was found that no steady state was reached in the simulations. A closer investigation showed that in principle, steady states of the flux data could not be reached as

several maximal activity values were too low to support the fluxes. This was not seen as a defeat, as our approach already showed the first errors in the model - the maximal activity values need measurement at these conditions. By not fitting enzyme parameters to the data, the context independent mechanistic approach was followed strictly, which is the basis for further work.

## Appendix A

### Rate equations

$$v_{\text{GLT}} \rightarrow \frac{\left(\text{GLCo} - \frac{\text{GLCi}}{K_{\text{eqGLT}}}\right) V_{\text{mGLT}}}{\left(1 + \frac{\text{GLCi}}{K_{\text{mGLTGLCi}}} + \frac{\text{GLCo}}{K_{\text{mGLTGLCo}}} + \frac{\alpha \text{GLCiGLCo}}{K_{\text{mGLTGLCi}K_{\text{mGLTGLCo}}}\right) K_{\text{mGLTGLCo}}}$$

$$v_{\text{GLT}} \rightarrow \frac{\left(\text{GLCo} - \frac{\text{GLCi}}{K_{\text{eqGLT}}}\right) V_{\text{mGLT}}}{\left(1 + \frac{\text{GLCi}}{K_{\text{mGLTGLCi}}} + \frac{\text{GLCo}}{K_{\text{mGLTGLCo}}} + \frac{\alpha \text{GLCiGLCo}}{K_{\text{mGLTGLCi}K_{\text{mGLTGLCo}}}\right) K_{\text{mGLTGLCo}}}$$

$$v_{\text{GLK}} \rightarrow \frac{\left(\text{ATPGLCi} - \frac{\text{ADPG6P}}{K_{\text{eqGLK}}}\right) V_{\text{mGLK}}}{\left(1 + \frac{\text{ADP}}{K_{\text{mGLKADP}}} + \frac{\text{ATP}}{K_{\text{mGLKATP}}}\right) K_{\text{mGLKATP}} \left(1 + \frac{\text{G6P}}{K_{\text{mGLKG6P}}} + \frac{\text{GLCi}}{K_{\text{mGLKGLCi}}}\right) K_{\text{mGLKGLCi}}}$$

$$v_{\text{PGI}} \rightarrow \frac{\left(\text{G6P} - \frac{\text{F6P}}{K_{\text{eqPGI}}}\right) V_{\text{mPGI}}}{\left(1 + \frac{\text{F6P}}{K_{\text{mPGIF6P}}} + \frac{\text{G6P}}{K_{\text{mPGIG6P}}}\right) K_{\text{mPGIG6P}}}$$

$$v_{\text{GLYCO}} \rightarrow \text{KGLYCOGEN}$$

$$v_{\text{Treha}} \rightarrow \text{KTREHALOSE}$$

$$v_{\text{PFK}} \rightarrow \text{ATPF6PgR} \left(1 + \frac{\text{ATP}}{K_{\text{mPFKATP}}} + \frac{\text{F6P}}{K_{\text{mPFKF6P}}} + \frac{\text{ATPF6PgR}}{K_{\text{mPFKATPKmPFKF6P}}}\right) V_{\text{mPFK}} / \left( K_{\text{mPFKATPKmPFKF6P}} \left(1 + \frac{\text{ATP}}{K_{\text{mPFKATP}}} + \frac{\text{F6P}}{K_{\text{mPFKF6P}}} + \frac{\text{ATPF6PgR}}{K_{\text{mPFKATPKmPFKF6P}}}\right)^2 + \frac{\left(1 + \frac{\text{ATPCiPFKATP}}{K_{\text{iPFKATP}}}\right)^2 \left(1 + \frac{\text{ATPCPFKATP}}{K_{\text{mPFKATP}}}\right)^2 \left(1 + \frac{\text{ADP}^2 \text{CPFKAMPK}_{\text{eqAK}}}{\text{ATPKPFKAMP}}\right)^2 \left(1 + \frac{\text{CPFKF16BPF16P} + \text{CPFKF26BPF26BP}}{\text{KPFKF16BP} + \text{KPFKF26BP}}\right)^2 L_0}{\left(1 + \frac{\text{ATP}}{K_{\text{iPFKATP}}}\right)^2 \left(1 + \frac{\text{ADP}^2 \text{K}_{\text{eqAK}}}{\text{ATPKPFKAMP}}\right)^2 \left(1 + \frac{\text{F16P}}{\text{KPFKF16BP}} + \frac{\text{F26BP}}{\text{KPFKF26BP}}\right)^2}$$

$$v_{\text{ALD}} \rightarrow \left(\text{F16P} - \frac{\text{K}_{\text{eqTPITRIO}}}{\text{K}_{\text{eqALD}}(1 + \text{K}_{\text{eqTPI}})}\right) V_{\text{mALD}} / \left( K_{\text{mALDF16P}} \left(1 + \frac{\text{F16P}}{K_{\text{mALDF16P}}} + \frac{\text{TRIO}}{(1 + \text{K}_{\text{eqTPI}}) K_{\text{mALDDHAP}}} + \frac{\text{K}_{\text{eqTPITRIO}}}{(1 + \text{K}_{\text{eqTPI}}) K_{\text{mALDGAP}}} + \frac{\text{F16PK}_{\text{eqTPITRIO}}}{(1 + \text{K}_{\text{eqTPI}}) K_{\text{mALDF16PKmALDGAP}}} + \frac{\text{K}_{\text{eqTPITRIO}}^2}{(1 + \text{K}_{\text{eqTPI}})^2 K_{\text{mALDDHAPKmALDGAP}}}\right)$$

$$\begin{aligned}
v_{G3PDH} &\rightarrow \left( -\frac{GLYNAD}{K_{eqG3PDH}} + \frac{NADHTRIO}{1+K_{eqTPI}} \right) V_{mG3PDH} / \\
&\left( K_{mG3PDHDHAP} K_{mG3PDHNADH} \left( 1 + \frac{NAD}{K_{mG3PDHNAD}} + \frac{NADH}{K_{mG3PDHNADH}} \right) \right. \\
&\left. \left( 1 + \frac{GLY}{K_{mG3PDHGLY}} + \frac{TRIO}{(1+K_{eqTPI})K_{mG3PDHDHAP}} \right) \right) \\
v_{GAPDH} &\rightarrow \left( \frac{K_{eqTPINADTRIO} V_{mGAPDHf}}{(1+K_{eqTPI})K_{mGAPDHGAP} K_{mGAPDHNAD}} - \frac{BPGNADH V_{mGAPDHR}}{K_{mGAPDHBPG} K_{mGAPDHNADH}} \right) / \\
&\left( \left( 1 + \frac{NAD}{K_{mGAPDHNAD}} + \frac{NADH}{K_{mGAPDHNADH}} \right) \left( 1 + \frac{BPG}{K_{mGAPDHBPG}} + \frac{K_{eqTPITRIO}}{(1+K_{eqTPI})K_{mGAPDHGAP}} \right) \right) \\
v_{PGK} &\rightarrow \frac{(ADPBPGK_{eqPGK} - ATPP3G) V_{mPGK}}{\left( 1 + \frac{ADP}{K_{mPGKADP}} + \frac{ATP}{K_{mPGKATP}} \right) K_{mPGKATP} K_{mPGKP3G} \left( 1 + \frac{BPG}{K_{mPGKBPG}} + \frac{P3G}{K_{mPGKP3G}} \right)} \\
v_{PGM} &\rightarrow \frac{\left( -\frac{P2G}{K_{eqPGM}} + P3G \right) V_{mPGM}}{K_{mPGMP3G} \left( 1 + \frac{P2G}{K_{mPGMP2G}} + \frac{P3G}{K_{mPGMP3G}} \right)} \\
v_{ENO} &\rightarrow \frac{(P2G - \frac{PEP}{K_{eqENG}}) V_{mENO}}{K_{mENOP2G} \left( 1 + \frac{P2G}{K_{mENOP2G}} + \frac{PEP}{K_{mENOPEP}} \right)} \\
v_{PYK} &\rightarrow \frac{(ADPPEP - \frac{ATPPYR}{K_{eqPYK}}) V_{mPYK}}{K_{mPYKADP} \left( 1 + \frac{ADP}{K_{mPYKADP}} + \frac{ATP}{K_{mPYKATP}} \right) K_{mPYKPEP} \left( 1 + \frac{PEP}{K_{mPYKPEP}} + \frac{PYR}{K_{mPYKPYR}} \right)} \\
v_{PDC} &\rightarrow \frac{K_{mPDCPYR}^{-n_{PDC}} PYR^{n_{PDC}} V_{mPDC}}{1 + K_{mPDCPYR}^{-n_{PDC}} PYR^{n_{PDC}}} \\
v_{SUC} &\rightarrow ACEKSUCC \\
v_{ADH} &\rightarrow - \left( ETOHNAD - \frac{ACENADH}{K_{eqADH}} \right) V_{mADH} / \\
&\left( K_{iADHNAD} K_{mADHETOH} \left( 1 + \frac{ETOHK_{mADHNAD}}{K_{iADHNAD} K_{mADHETOH}} + \frac{ACEK_{mADHNADH}}{K_{iADHNADH} K_{mADHACE}} + \right. \right. \\
&\frac{NAD}{K_{iADHNAD}} + \frac{ETOHNAD}{K_{iADHNAD} K_{mADHETOH}} + \frac{ACEETOHNAD}{K_{iADHACE} K_{iADHNAD} K_{mADHETOH}} + \\
&\frac{K_{iADHNAD} K_{iADHNADH} K_{mADHACE}}{K_{iADHNAD} K_{iADHNADH} K_{mADHACE}} + \frac{K_{iADHNADH}}{K_{iADHNADH}} + \frac{ACENADH}{K_{iADHNADH} K_{mADHACE}} + \\
&\left. \left. \frac{ACEETOHNADH}{K_{iADHETOH} K_{iADHNADH} K_{mADHACE}} + \frac{ETOHK_{mADHNADH}}{K_{iADHNAD} K_{iADHNADH} K_{mADHETOH}} \right) \right) \\
v_{ATP} &\rightarrow ATPKATPASE \\
v_{AK} &\rightarrow \left( ADP^2 - \frac{AMPATP}{K_{eqAK}} \right) V_{mAK}
\end{aligned}$$

## Appendix B

# <sup>13</sup>C-metabolic model

reactions

v1glc uptake: glucose + ATP > G6P  
v2PGI: G6P = F6P  
v3ZWF: G6P > P5P + 2 \* NADPH + CO2  
v4PFK: F6P + ATP > 2 \* T3P  
v5TKL: 2\*P5P = S7P + T3P  
v6TKL: P5P + E4P = F6P + T3P  
v7TAL: S7P + T3P = E4P + F6P  
v8SER: T3P > SER + NADH  
v9SHM: SER + NADH > GLY + C1  
v10SHM: GLY + C1 > SER + NADH  
v11GCV: C1 + CO2 + NADH = GLY  
v12TDH: T3P = PEP + ATP + NADH  
v13PYR: PEP > cytPYR + ATP  
v14PDA: mPYR > mAcCoA + CO2 + NADH  
v15CIT: mOAA + mAcCoA > CIT  
v16ACO: CIT > OGA + CO2 + NADH  
v17KGD: OGA > SUCC + CO2 + 0.5\*ATP + NADH  
v18SDH: SUCC = FUM + NADH  
v19MDH: MAL = mOAA + NADH  
v20FUM: FUM = MAL  
v21MAE: MAL > mPYR + CO2 + NADPH  
v22PCK: cytOAA + ATP > PEP + CO2  
v23: cytPYR + CO2 + ATP > cytOAA  
v24ACS: acetate + 2 \* ATP > cytAcCoA  
v25ALD: acetaldehyde = acetate + NADPH  
v26ADH: acetaldehyde + NADH = ethanol  
v27GPD: T3P + NADH = glycerol  
v28OAC: cytOAA > mOAA  
v29OAC: mOAA > cytOAA  
v30: cytAcCoA > mAcCoA

v31: cytPYR > mPYR  
v32Atmung:O2 + 2\*NADH > 2\*PO\*ATP  
v33: cytPYR > acetaldehyde + CO2  
v34TH: NADPH > NADH

ratios

cytOAA from cytPYR =  $[v23]/([v29OAC]+[v23])$   
mOAA from anaplerosis =  $[v28OAC]/([v19MDH]+[v28OAC])$   
PEP from cytOAA =  $[v22PCK]/([v22PCK]+[v12TDH])$   
SER from glycolysis =  $(2*[v4PFK]-2*[v6TKL]-2*[v7TAL])/$   
 $(2*[v4PFK]+[v5TKL]+[v6TKL])$   
mAcCoA from mPYR =  $[v14PDA]/([v14PDA]+[v30])$   
mPYR from MAL (ub)  $\geq [v21MAE]/([v31]+[v21MAE])$   
mPYR from MAL (lb)  $\leq [v21MAE]/([v31]+[v21MAE])$   
SER from GLY =  $[v10SHM]/([v10SHM]+[v8SER])$   
GLY from SER =  $[v9SHM]/([v11GCV]+[v9SHM])$

biomass

$(0.438)*(237*[P5P]+140*[T3P]+660*[E4P]+1321*[PEP]+$   
 $2546*[cytOAA]+463*[SER]+726*[GLY]+371*[cytAcCoA]+$   
 $2085*[OGA]+4952*[mPYR]+371*[mAcCoA]-3599*[CO2]+$   
 $17546*[NADPH]-1393*[NADH]+10293.2*[ATP])$

$1*(234*[P5P]+142*[T3P]+101*[cytOAA]+142*[CO2]+$   
 $449*[NADPH]+528*[NADH]+2527*[ATP])$

$1*(13.04*[P5P]+6.52*[T3P]+6.52*[cytOAA]+6.52*[CO2]+$   
 $23.28*[NADPH]-42.84*[NADH]+151.868*[ATP])$

$1*(153.44*[T3P]+217.35*[cytPYR]+109.27*[cytOAA]+$   
 $1260.49*[cytAcCoA]-388.64*[CO2]+3007.9*[NADPH]-$   
 $90.44*[NADH]+1776.81*[ATP])$

$1*(1911*[G6P]+2123*[ATP])$

$1*(83.3*[OGA]+83.3*[NADPH]+333*[ATP])$

# Appendix C

## Fermentation data plots

*Note:* All experiments were performed in triplicate. Procedures were as given in section 5.2.1.

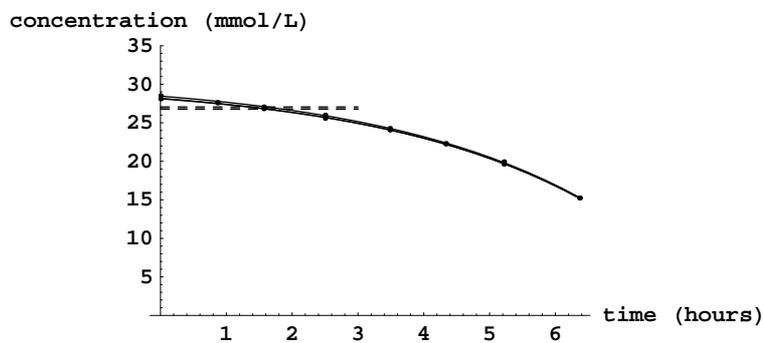


Figure C.1: Glucose concentrations during aerobic growth.

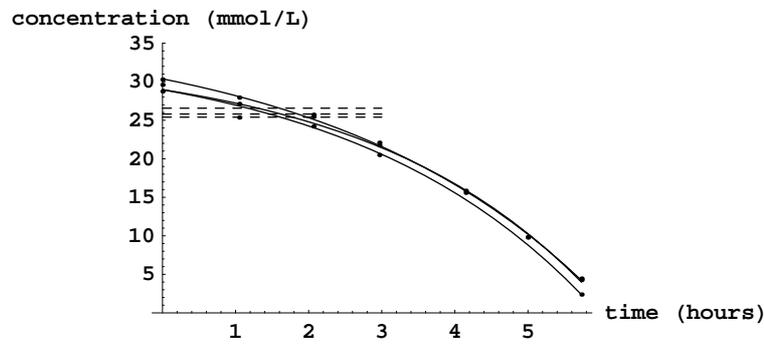


Figure C.2: Glucose concentrations during anaerobic growth.

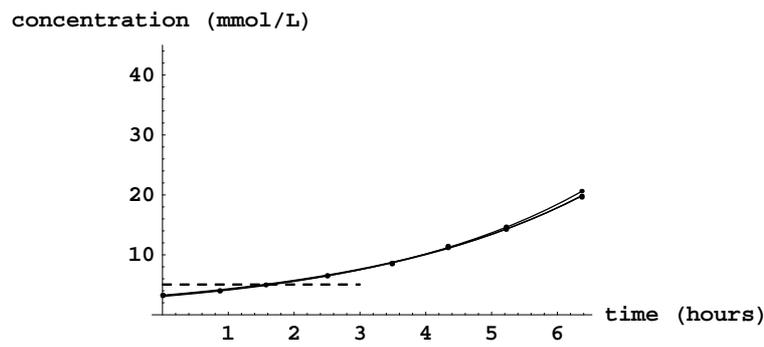


Figure C.3: Ethanol concentrations during aerobic growth.

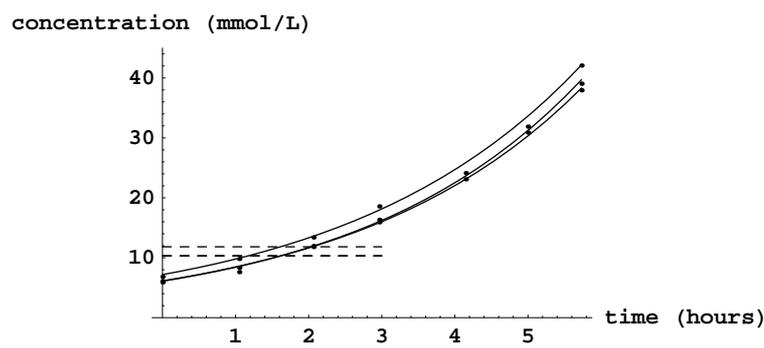


Figure C.4: Ethanol concentrations during anaerobic growth.

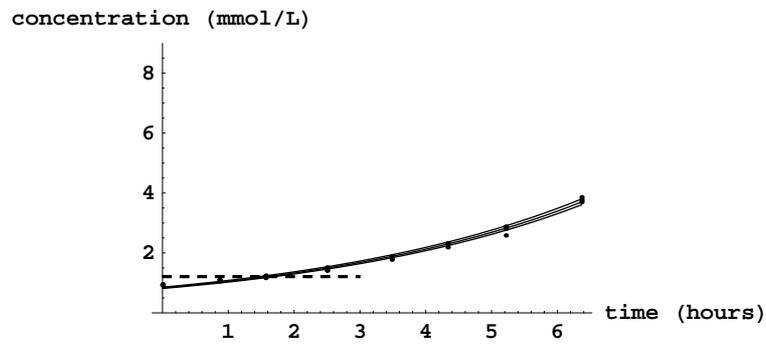


Figure C.5: Glycerol concentrations during aerobic growth.

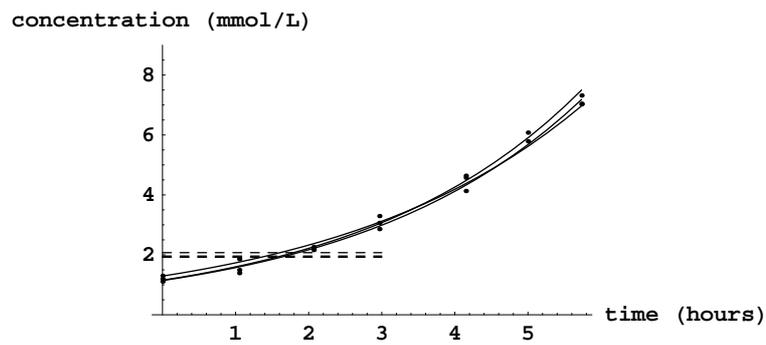


Figure C.6: Glycerol concentrations during anaerobic growth.

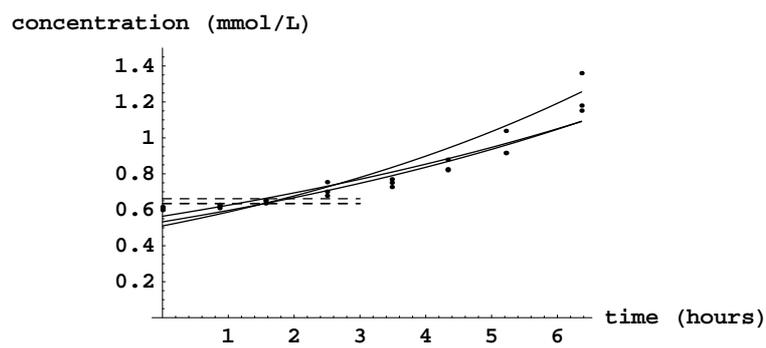


Figure C.7: Acetate concentrations during aerobic growth.

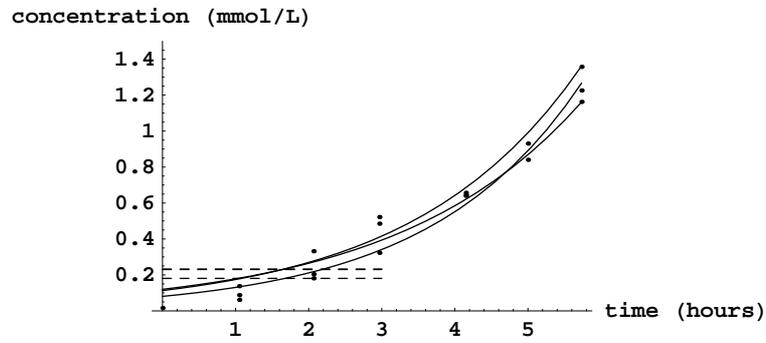


Figure C.8: Acetate concentrations during anaerobic growth.

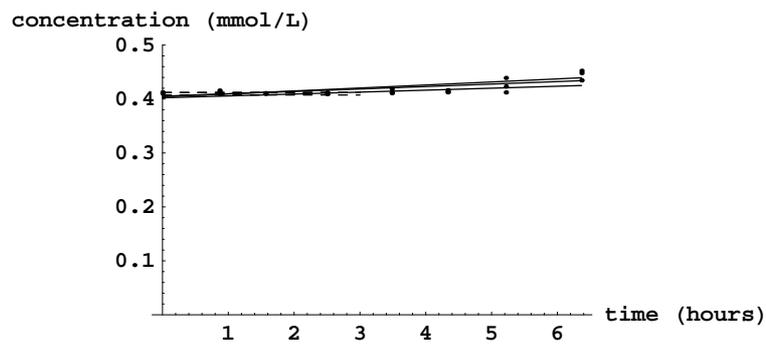


Figure C.9: Succinate concentrations during aerobic growth.

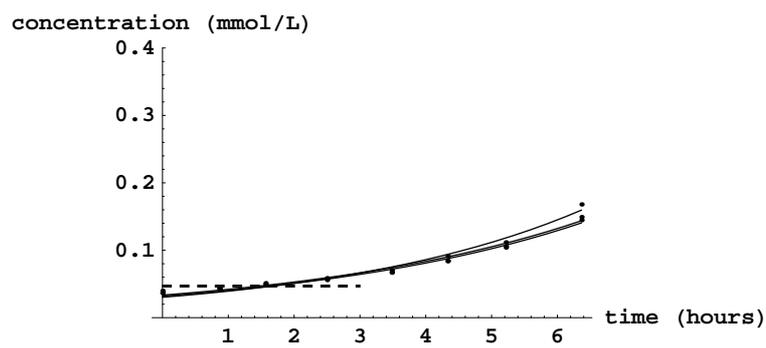


Figure C.10: Pyruvate concentrations during aerobic growth.

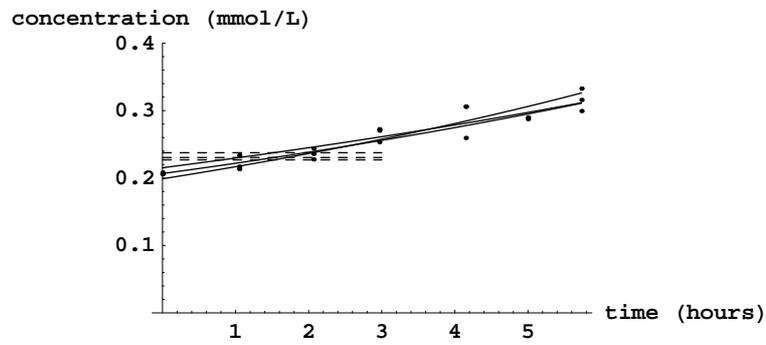


Figure C.11: Pyruvate concentrations during anaerobic growth.

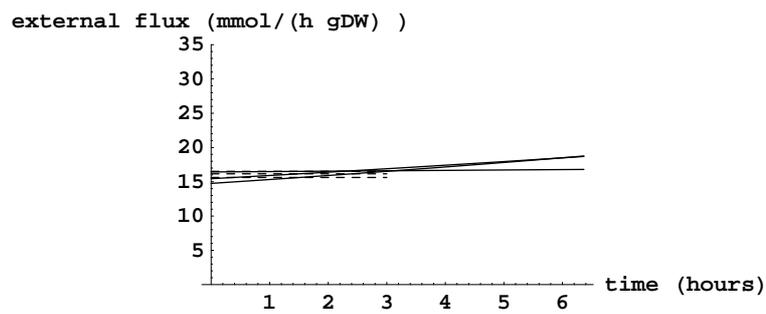


Figure C.12: Glucose consumption during aerobic growth.

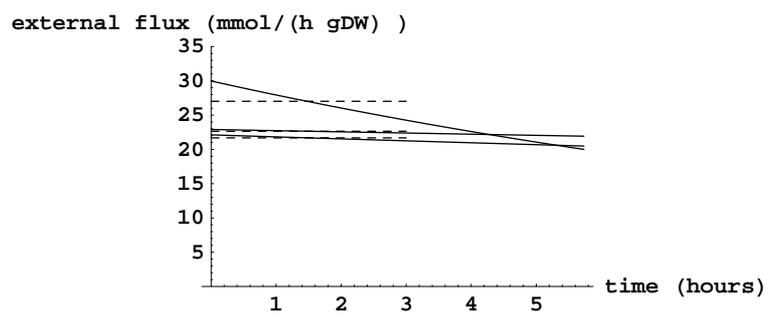


Figure C.13: Glucose consumption during anaerobic growth.

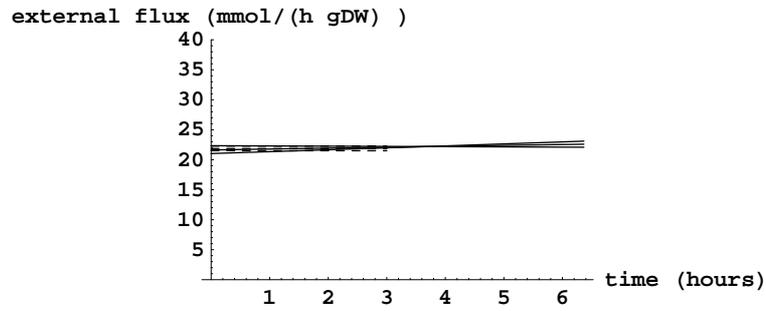


Figure C.14: Ethanol production during aerobic growth.

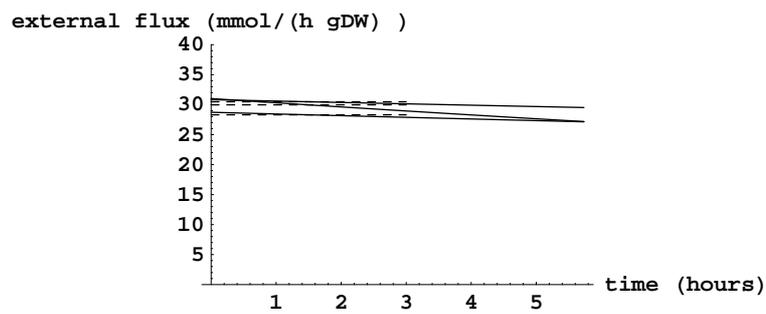


Figure C.15: Ethanol production during anaerobic growth.

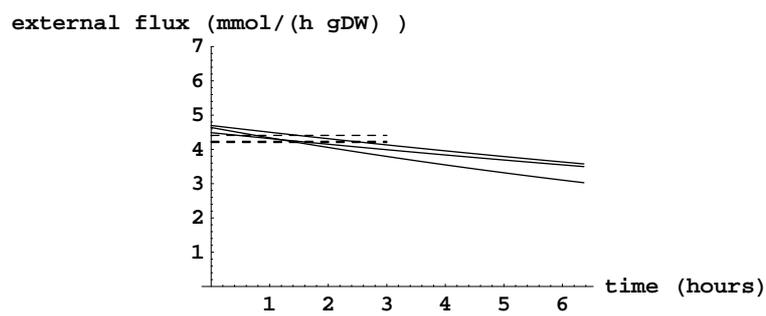


Figure C.16: Glycerol production during aerobic growth.

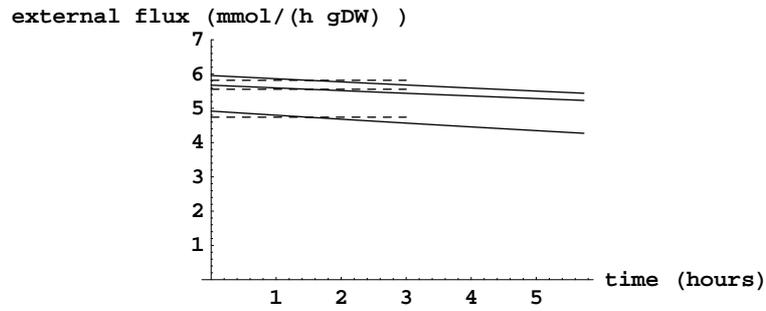


Figure C.17: Glycerol production during anaerobic growth.

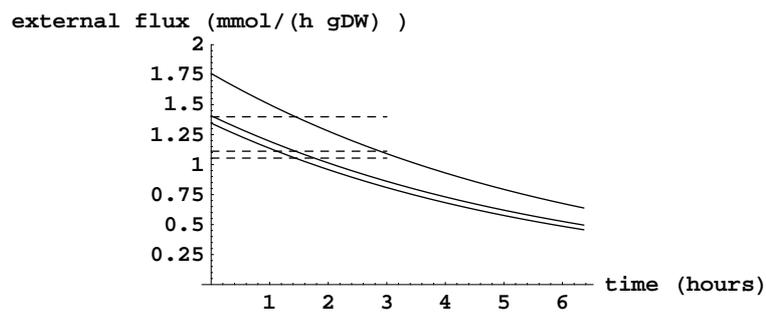


Figure C.18: Acetate production during aerobic growth.

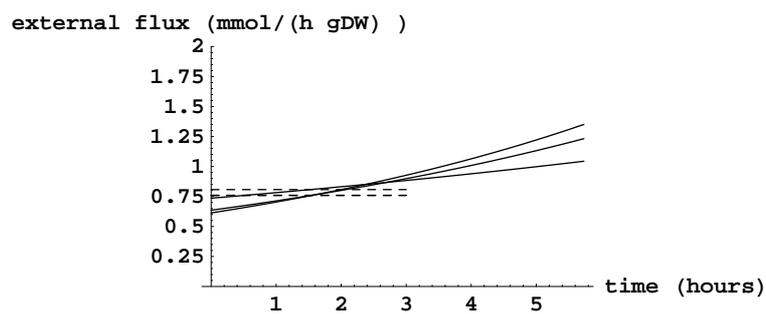


Figure C.19: Acetate production during anaerobic growth.

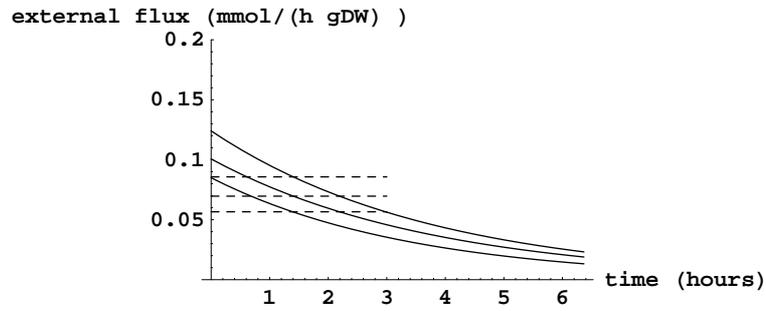


Figure C.20: Succinate production during aerobic growth.

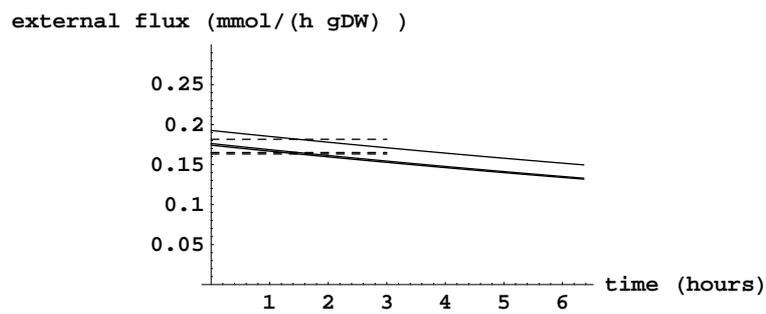


Figure C.21: Pyruvate production during aerobic growth.

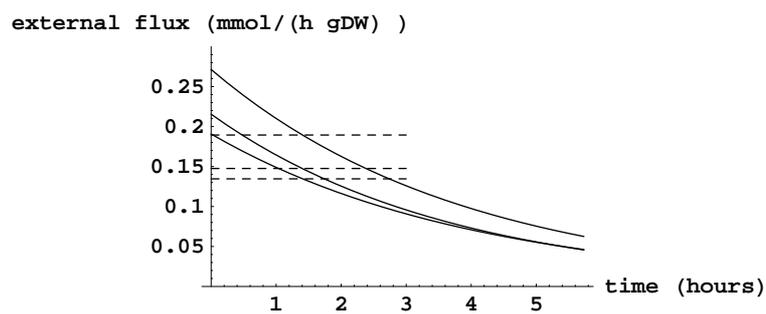


Figure C.22: Pyruvate production during anaerobic growth.

# Bibliography

- [1] J. Varner and D. Ramkrishna. Mathematical models of metabolic pathways. *Current opinion in Biotechnology*, 10:146–150, 1999.
- [2] S. Feyo de Azedevo, B. Dahm, and F.R. Oliveira. Hybrid modelling of biochemical processes: A comparison with the conventional approach. *Computers Chem. Engng*, 21:751–756, 1997.
- [3] A.R. Fernie, P. Geigenberger, and M. Stitt. Flux and important, but neglected, component of functional genomics. *Current Opinion in Plant Biology*, 8:174–182, 2005.
- [4] S. Klamt, S. Schuster, and E.D. Gilles. Calculability analysis in underdetermined metabolic networks illustrated by a model of central metabolism in purple nonsulfur bacteria. *Biotechnology and Bioengineering*, 77(7), 3 2002.
- [5] A. Ben-Israel and T.N.E. Greville. *Generalized inverses: Theory and applications*. John Wiley & Sons, New York, 1974.
- [6] W.H. Holms. The central metabolic pathways of *Escherichia coli*. Relationship between flux and control at a branchpoint, efficiency of conversion to biomass, and excretion of acetate. *Curr. Top. Cell. Regul.*, 28:69–105, 1986.
- [7] J.M. Savinell and B.O. Palsson. Optimal selection of metabolic fluxes for *in vivo* measurement. I. development of mathematical methods. *J. Theor. Biol.*, 115:201–214, 1992.
- [8] J.H. Noorman, B. Romein, K.C.A.M. Luyben, and J.J. Heijnen. Classification, error detection, and reconciliation of process information in complex biochemical systems. *Biotechnol. Bioeng.*, 49:364–367, 1996.
- [9] H.P.J. Bonarius. Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. *Trends. Biotechnol.*, 15:308–314, 1997.
- [10] R.T.J.M. van der Heijden, J.J. Heijnen, C. Hellinga, B. Romein, and K.C.A.M. Luyben. Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates. *Biotechnol. Bioeng.*, 43:3–10, 1994.
- [11] R.T.J.M. van der Heijden, B. Romein, J.J. Heijnen, C. Hellinga, and K.C.A.M. Luyben. Linear constraint relations in biochemical reaction

- systems: I. Diagnosis and estimation of gross errors. *Biotechnol. Bioeng.*, 43:11–20, 1994.
- [12] W. Wiechert.  $^{13}\text{C}$  metabolic flux analysis. *Metabolic engineering*, 3:159–206, 2001.
- [13] U. Sauer. High-throughput phenomics: experimental methods for mapping fluxomes. *Current opinion in Biotechnology*, 15:58–63, 2004.
- [14] N.D. Price, J.A. Papin, C.H. Schilling, and B.O. Palsson. Genome-scale microbial *in silico* models: the constraints-based approach. *TRENDS in Biotechnology*, 21(4), 2003.
- [15] J.M. Lee, E.P. Gianchandani, and J.A. Papin. Flux balance analysis in the era of metabolomics. *Briefings in Bioinformatics*, 7(4):140–150, 2006.
- [16] S.L. Bell and B.O. Palsson. Phenotype phase plane analysis using interior point methods. *Computers and chemical engineering*, 2004.
- [17] N. Jamshidi and B.O. Palsson. Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the *in silico* strain iNJ661 and proposing alternative drug targets. *BMC Systems Biology*, 1(26), 2007.
- [18] S.A. Wahl, R. Takors, and W. Wiechert. Interpretation of metabolic flux maps by limitation potentials and constrained limitation sensitivities. *Biotechnology and Bioengineering*, 2005.
- [19] A.P. Burgard and C.D. Maranas. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol. Bioeng.*, 82:670–677, 2003.
- [20] D. Segre, D. Vitkup, and G.M. Church. Analysis of optimality of natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci.*, 99:15112–15117, 2002.
- [21] L. Kuepfer, U. Sauer, and L.M. Blank. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.*, 15:1421–1430, 2005.
- [22] A.P. Burgard, P. Pharkaya, and C.D. Maranas. Optknock: a bilevel programming framework for identifying gene-knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.*, 84:647–657, 2003.
- [23] D.A. Beard, S.D. Liang, and H. Qian. Energy balance for analysis of complex metabolic networks. *Biophys. J.*, 83:79–86, 2002.
- [24] D.A. Beard, E. Babson, E. Curtis, and H. Qian. Thermodynamic constraints for biochemical networks. *J. Theor. Biol.*, 228:327–333, 2004.
- [25] T. Maskow and U. von Stockar. How reliable are thermodynamic feasibility statements of biochemical pathways? *Biotechnol. Bioeng.*, 92:223–230, 2005.
- [26] A. Varma and B.O. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ microbiol*, 60:3724–3731, 1994.

- [27] M.W. Covert and B.O. Palsson. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.*, 277:28058–28064, 2002.
- [28] M.W. Covert, C.H. Schilling, and B.O. Palsson. Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.*, 213:73–88, 2001.
- [29] R. Mahadevan, J.S. Edwards, and F.J. Doyle. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.*, 83:1331–1340, 2002.
- [30] C.H. Schilling and B.O. Palsson. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J. theor. Biol.*, 203:249–283, 2000.
- [31] J.S. Edwards and B.O. Palsson. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. U.S.A.*, 97(4), 2000.
- [32] J. Forster, I. Famili, P. Fu, B.O. Palsson, and J. Nielsen. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Research*, 13:244–253, 2003.
- [33] N.C. Duarte, M.J. Herrgard, and B.O. Palsson. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research*, 2004.
- [34] D.J.V. Beste, T. Hooper, G. Stewart, B. Bonde, C. Avignone-Rossa, M. Bushell, P. Wheeler, S. Klamt, A.M. Kierzek, and J. McFadden. GSMN-TB: a web-based genome scale network model of *Mycobacterium tuberculosis* metabolism. *Genome Biology*, 8(89), 2007.
- [35] M.E. Cusick, N. Klitgord, M. Vidal, and D.E. Hill. Interactome: gateway into systems biology. *Human molecular genetics*, 14(2):171–181, 2005.
- [36] T. Ito, T. Chiba, R. Ozawa, R. Yoshida, M. Hattori, and Y. Sakaki. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci.*, 98:4569–4574, 2001.
- [37] L.J. Lu, Y. Xia, A. Paccanaro, H. Yu, and M. Gerstein. Assessing the limits of genomic data integration for predicting protein networks. *Genome Research*, 15:945–953, 2005.
- [38] J.F. Rual, K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, N. Li, G.F. Berriz, F.D. Gibbons, M. Dreze, and M. Ayivi-Guedehouso. Towards a proteome-scale map of the human interactome network. *Nature*, 2005.
- [39] T. Ideker and D. Lauffenburger. Building with a scaffold: emerging strategies for high- to low-level cellular modelling. *Trends. Biotechnol.*, 2003.
- [40] A.G. Gilman. Overview of the alliance for cellular signaling. *Nature*, 420:703–706, 2002.
- [41] S. Schuster, T. Dandekar, and D.A. Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *TIBTECH*, 17, 1999.

- [42] M.J. Herrgard, M.W. Covert, and B.O. Palsson. Reconstruction of microbial transcriptional regulatory networks. *Current Opinion in Biotechnology*, 15:70–77, 2004.
- [43] S. Klamt, J. Saez-Rodrigues, J.A. Lindquist, L. Simeoni, and E.D. Gilles. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, 2006.
- [44] B. Teusink and. Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *Eur. J. Biochemistry*, 268:1–18, 2000.
- [45] M.H.N. Hoefnagel, M.J.C. Starrenburg, D.E. Martens, J. Hugenholtz, M. Kleerenbezem, I.I. van Swam, R. Bongers, H.V. Westerhoff, and J.L. Snoep. Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. *Microbiology*, 148:1003–1013, 2002.
- [46] P.O. Westermark and A. Lansner. A model of phosphofructokinase and glycolytic oscillations in the pancreatic  $\beta$ -cell. *Biophys. J.*, 85:126–139, 2003.
- [47] K. Reijenga, Y.M.G.A. van Megen, B.W. Kooi, B.M. Bakker, J.L. Snoep, H.W. van Verseveld, and H.V. Westerhoff. Yeast glycolytic oscillations that are not controlled by a single oscillophore: a new definition of oscillophore strength. *J. Theor. Biol.*, 232:385–398, 2005.
- [48] K.C. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak, and J.J. Tyson. Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell*, 11:369–391, 2000.
- [49] K.C. Chen, L. Calzone, A. Csikasz-Nagy, F.R. Cross, B. Novak, and J.J. Tyson. Integrative analysis of cell cycle control in budding yeast. *Mol. Biol. Cell*, 15:3841–3862, 2004.
- [50] J.J. Tyson and B. Novak. Regulation of the eukaryotic cell cycle: Molecular antagonism, hysteresis, and irreversible transitions. *J. theor. Biol.*, 210:249–263, 2001.
- [51] J.H.S. Hofmeyr. Steady-state modelling of metabolic pathways: A guide for the prospective simulator. *CABIOS REVIEW*, 2(1):5–11, 1986.
- [52] M. Rizzi, M. Baltes, U. Theobald, and M. Reuss. *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. mathematical model. *Biotechnology and Bioengineering*, 55(4), August 1997.
- [53] H. Kacser and J.A. Burns. The control of flux. *Symp. Soc. Expr. Biol.*, 27:65–104, 1973.
- [54] R. Heinrich and T.A. Rapoport. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *Eur. J. Biochem.*, 42:89–95, 1974.
- [55] D.A. Fell and H.M. Sauro. Metabolic control and its analysis: additional relationships between elasticities and control coefficients. *Eur. J. Biochem.*, 148:555–561, 1985.

- [56] C. Reder. Metabolic control theory: a structural approach. *J. Theor. Biol.*, 135:175–201, 1988.
- [57] B.N. Kholodenko and H.V. Westerhoff. Metabolic channelling and control of the flux. *FEBS Lett.*, 320:71–74, 1993.
- [58] V. Hatzmanikatis and J.E. Bailey. MCA has more to say. *J. Theor. Biol.*, 182:233–242, 1996.
- [59] V. Hatzmanikatis and J.E. Bailey. Effects of spatiotemporal variations on metabolic control: approximate analysis using (log)linear kinetic models. *Biotechnol. Bioeng.*, 54:91–104, 1997.
- [60] K.A. Reijenga, H.V. Westerhoff, B.N. Kholodenko, and J.L. Snoep. Control analysis for autonomously oscillating biochemical networks. *Biophys. J.*, 82:99–108, 2002.
- [61] R. Conradie, H.V. Westerhoff, J.M. Rohwer, J.H. Hofmeyr, and J.L. Snoep. Summation theorems for flux and concentration control coefficients of dynamic systems. *Systems Biology*, 153(5):314–317, 2006.
- [62] D. Kell and P. Mendez. Snapshots of systems: Metabolic control analysis and biotechnology in the postgenomic era. In A.J. Cornish-Bowden and M.L. Cardenas, editors, *Technical and medical implications of Metabolic Control Analysis*, pages 2–25. Amsterdam: Kluwer, 2000.
- [63] L. Wang, I. Birol, and V. Hatzmanikatis. Metabolic control analysis under uncertainty. *Biophys. J.*, 87:3750–3763, 2004.
- [64] V. Kadiramadathan, Y. Yang S.A. Billings, and P.C. Wright. Markov Chain Monte Carlo algorithm based metabolic flux distribution analysis on *Corynebacterium glutamicum*. *Bioinformatics*, 22(21):2681–2687, 2006.
- [65] S.J. Wiback, I. Famili, H.J. Greenberg, and B.O. Palsson. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J. Theor. Biol.*, 2004.
- [66] J. L. Snoep and H. V. Westerhoff. The Silicon Cell Initiative. *Current Genomics*, 5, 2004.
- [67] R. Leinonen, F.G. Diez, D. Binns, W. Fleishmann, R. Lopez, and R. Apweiler. Uniprot archive. *Bioinformatics*, 20(17):3236–3237, 2004.
- [68] M. Schneider, A. Bairoch, C.H. Wu, and R. Apweiler. Plant protein annotation in the Uniprot Knowledgebase. *Bioinformatics*, 138:59–66, 2005.
- [69] C.H. Wu, R. Apweiler, A. Bairoch, D.A. Natale, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M.J. Martin, R. Mazumder, C. O'Donovan, N. Redaschi, and B. Suzek. The universal protein resource (UniProt): an expanding universe of protein information. *Nucleic Acids Research*, 34:187–191, 2006.
- [70] The UniProt Consortium. The universal protein resource (Uniprot). *Nucleic Acids Research*, 35:193–197, 2007.
- [71] I. Schomburg, A. Chang, O. Hofmann, C. Ebeling, F. Ehrentreich, and D. Schomburg. Brenda: a resource for enzyme data and metabolic information. *Trends. Biochem. Sci.*, 27:54–56, 2002.

- [72] S. Sundararaj, A. Guo, B. Habibi-Nazhad, M. Rouani, P. Stothard, M. Ellison, and D.S. Wishart. The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate *in silico* modeling of *Escherichia coli*. *Nucleic Acids Res.*, 32:293–295, 2004.
- [73] U. Wittig, M. Golebiewski, R. Kania, O. Krebs, S. Mir, A. Weidemann, S. Anstein, J. Saric, and I. Rojas. SABIO-RK: Integration and curation of reaction kinetics data. In *Data integration in the life sciences*. Springer Berlin/Heidelberg, 2006.
- [74] H. V. Westerhoff, F. Bruggeman, J.S. Hofmeyr, and J. L. Snoep. Attractive models: how to make the silicon cell relevant and dynamic. *Comp. funct. genom.*, 4:155–158, 2003.
- [75] C. Zupke, R. Tompkins, D. Yarmush, and M. Yarmush. Numerical isotopomer analysis: Estimation of metabolic activity. *Analytical Biochemistry*, 247:287–293, 1997.
- [76] W. Wiechert and A.A. de Graaf. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. *Biotechnology and Bioengineering*, 55(1), 7 1997.
- [77] W. Wiechert, C. Siefke, A.A. de Graaf, and A. Marx. Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis. *Biotechnology and Bioengineering*, 55(1), 7 1997.
- [78] M.J. Arauzo-Bravo and K. Shimizu. Estimation of bidirectional metabolic fluxes from MS and NMR data using positional representations. *Genome Informatics*, 12:63–72, 2001.
- [79] W. Wiechert, M. Mollney, N. Isermann, M. Wurzel, and A.A. de Graaf. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnology and Bioengineering*, 66(2), 7 1999.
- [80] M. Mollney, W. Wiechert, D. Kownatzki, and A.A. de Graaf. Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments. *Biotechnology and Bioengineering*, 66(2), 7 1999.
- [81] W. Wiechert and M. Wurzel. Metabolic isotopomer labeling systems part I: global dynamic behaviour. *Mathematical Biosciences*, 169:173–205, 2001.
- [82] W. Wiechert and M. Wurzel. Metabolic isotopomer labeling systems part II: structural flux identifiability analysis. *Mathematical Biosciences*, 183:175–214, 2003.
- [83] W.A. van Winden, J.J. Heijnen, and P.J.T. Verheijen. Cumulative bondomers: A new concept in flux analysis from 2D [ $^{13}\text{C}$ ,  $^1\text{H}$ ] COSY NMR data. *Biotechnology and Bioengineering*, 80(70), 12 2002.
- [84] T. Grotkjaer, M. Akesson, B. Christensen, A.K. Gombert, and J. Nielsen. Impact of transamination reactions and protein turnover on labeling dynamics in  $^{13}\text{C}$ -labeling experiments. *Biotechnology and Bioengineering*, 86(2), 4 2000.

- [85] T. Szyperski. Biosynthetically directed fractional  $^{13}\text{C}$ -labelling of proteinogenic amino acids. An efficient tool to investigate intermediary metabolism. *Eur. J. Biochem*, 232:433–448, 1995.
- [86] E. Fischer and U. Sauer. Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *Eur. J. Biochem.*, 270:880–891, 2003.
- [87] E. Fischer, N. Zamboni, and U. Sauer. High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived  $^{13}\text{C}$  constraints. *Analytical Biochemistry*, 325:308–316, 2004.
- [88] U. Sauer, J.E. Bailey, and K. Wuthrich. Bioreaction network topology and metabolic flux ratio analysis by biosynthetic fractional  $^{13}\text{C}$  labeling and two-dimensional NMR spectroscopy. *Metabolic engineering*, 1:189–197, 1999.
- [89] N. Zamboni and U. Sauer. Model-independent fluxome profiling from  $^2\text{H}$  and  $^{13}\text{C}$  experiments for metabolic variant discrimination. *Genome Biology*, 5, 2004.
- [90] A.S. Wahl, M. Dauner, and W. Wiechert. New tools for mass isotopomer data evaluation in  $^{13}\text{C}$  Flux Analysis: Mass isotope correction, data consistency checking and precursor relationships. *Biotechnology and Bioengineering*, 85(3), 2 2004.
- [91] A.A. de Graaf, M. Mahle, M. Mollney, W. Wiechert, P. Stahmann, and H. Sahm. Determination of full  $^{13}\text{C}$  isotopomer distributions for metabolic flux analysis using heteronuclear spin echo difference NMR spectroscopy. *Journal of Biotechnology*, 77:25–35, 2000.
- [92] M. Dauner and U. Sauer. GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing. *Biotechnol. Prog.*, 15:642–649, 2000.
- [93] C. Wittman and A. Heinzle. Application of MALDI-TOF MS to lysine-producing *Corynebacterium glutamicum*. *Eur. J. Biochem.*, 268:2441–2455, 2001.
- [94] A. Tholey, C. Wittman, M-J Kang, D. Bungert, K. Hollemeyer, and E. Heinzle. Derivatization of small biomolecules for optimized matrix-assisted laser desorption/ionization mass spectrometry. *J. Mass Spectrom.*, 37:963–973, 2002.
- [95] N. Zamboni, E. Fischer, and U. Sauer. FiatFlux - a software for metabolic flux analysis from  $^{13}\text{C}$  glucose experiments. *BMC Bioinformatics*, 6, 2005.
- [96] L.M. Blank, F. Lehmbech, and U. Sauer. Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. *FEMS Yeast Research*, 5:545–558, 2005.
- [97] L.M. Blank, L. Kuepfer, and U. Sauer. Large-scale  $^{13}\text{C}$ -flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biology*, 6(6), 2005.
- [98] K. Noh and W. Wiechert. Parallel solution of cascaded ODE systems applied to  $^{13}\text{C}$ -labelling experiments. In *Computational Science-ICCS 2004*. Springer Berlin/Heidelberg, 2004.

- [99] W. Wiechert. The thermodynamic meaning of exchange fluxes. *Biophysical Journal*, 93:2255–2264, 2007.
- [100] D.A. Beard and H. Qian. Relationship between thermodynamic driving force and one-way fluxes in reversible processes. *PLOS ONE*, 1, 2007.
- [101] A. Cornish-Bowden and J.H.S. Hofmeyr. The role of stoichiometric analysis in studies of metabolism. *J. Theor. Biology*, 216:179–191, 2002.
- [102] H.C. Lange, M. Eman, G. van Zuijlen, D. Visser, J.C. van Dam, J. Frank, M.J. Teixeira de Mattos, and J.J. Heijnen. Improved rapid sampling for in vivo kinetics of intracellular metabolites in *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering*, 75(4), November 2001.
- [103] W. Welthagen, R.A. Shellie, J. Spranger, M. Ristow, R. Zimmermann, and O. Fiehn. Comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry (GC×GC-TOF) for high resolution metabolomics: biomarker discovery on spleen tissue extracts of obese NZO compared to lean C57BL/6 mice. *Metabolomics*, 1(1):65–73, 2005.
- [104] R. Breitling, A.R. Pitt, and M.P. Barrett. Precision mapping of the metabolome, 2006. doi:10.1016/j.tibtech.2006.10.006.
- [105] L.L. Looger, S. Lalonde, and W.B. Frommer. Genetically encoded FRET sensors for visualizing metabolites with subcellular resolution in living cells. *Plant Physiology*, 138:555–557, 2005.
- [106] J.L. Snoep, M.H.N. Hoefnagel, and H.V. Westerhoff. Metabolic engineering of branched systems: redirecting the main pathway flux. In B.N. Kholodenko and H.V. Westerhoff, editors, *Metabolic Engineering in the Post-Genomic Era*. Horizon Bioscience, Wymondham, UK, 2004.
- [107] B.G. Olivier and J.M. Rohwer. Modelling cellular systems with PySCeS. *Bioinformatics*, 21:560–561, 2005.
- [108] L.M. Loew and J.C. Schaff. Virtual Cell: a software environment for computational cell biology. *Trends. Biotechnol*, 19:401–406, 2001.
- [109] H. Sauro. Jarnac: a system for interactive metabolic analysis. In J.H.S. Hofmeyr, J.M. Rohwer, and J.L. Snoep, editors, *BioThermoKinetics 2000*, pages 221–229. Stellenbosch University Press, 2000.
- [110] S. Hoops, S. Sahle, R. Gauges, Christine Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendez, and U. Kummer. COPASI – a complex pathway simulator. *System biology*, 22(24):3067–3074, 2006.
- [111] A. Funahashi and H. Kitano. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *BioSilico*, 1:159–162, 2003.
- [112] M. Tomita, K. Hashimoto, K. Takahashi, T.S. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J.C. Venter, and K. Hutchison. E-CELL: Software environment for whole-cell simulation. *Bioinformatics*, 15:72–84, 1999.
- [113] L.M. Blank and U. Sauer. TCA cycle activity in *Saccharomyces cerevisiae* is a function of the environmentally determined specific growth and glucose uptake rates. *Microbiology*, 150:1085–1093, 2004.

- [114] H. Maaheimo, J. Fiaux, Z.P. Cakar, J.E. Bailey, U. Sauer, and T. Szyperski. Central carbon metabolism of *Saccharomyces cerevisiae* explored by biosynthetic fractional  $^{13}\text{C}$  labeling of common amino acids. *Eur. J. Biochem*, 268:2468–2479, 2001.
- [115] C. Verduyn, E. Postma, W.A. Scheffers, and J.P. van Dijken. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, 8:501–517, 1992.
- [116] M. Kroukamp. Supply-demand analysis of free-energy metabolism in *Saccharomyces cerevisiae*, 2003. Ph.D Thesis – Stellenbosch University.
- [117] P. van Hoek, J.P. van Dijken, and J.T. Pronk. Regulation of fermentative capacity and levels of glycolytic enzymes in chemostat cultures of *Saccharomyces cerevisiae*. *Enzyme Microbial Technology*, 26:724–736, 2000.
- [118] C. Larsson, A. Nilsson, A. Blomberg, and L. Gustafsson. Glycolytic flux is conditionally correlated with ATP concentration in *Saccharomyces cerevisiae*: a chemostat study under carbon or nitrogen-limiting conditions. *Journal of Bacteriology*, 179(23):7243–7250, 1997.