

Jo Maertens

CFD study of the metabolic response to fast changing
substrate concentrations due to spatial heterogeneity in
industrial fermentors

Thesis submitted in fulfilment of the requirements for the degree of
Doctor (Ph.D.) in Applied Biological Sciences.

Dutch translation of the title:

CFD-studie van de metabolische respons op snel wijzigende substraatconcentraties als gevolg van ruimtelijke heterogeniteit in industriële fermentoren

Please refer to this work as follows: Jo Maertens. CFD study of the metabolic response to fast changing substrate concentrations due to spatial heterogeneity in industrial fermentors. Ph.D. thesis, Ghent University, Belgium, 2008.

ISBN: 978-90-5989-251-4

The author and the promotors give the authorisation to consult and to copy parts of this work for personal use only. Every other use is subject to the copyright laws. Permission to reproduce any material contained in this work should be obtained from the author.

Examination Committee

Prof. dr. B. De Baets (Ghent University)
Prof. dr. ir. J. J. Heijnen (Delft University of Technology)
Prof. dr. ir. J. Pieters (Ghent University, secretary)
Prof. dr. ir. D. Ramkrishna (Purdue University)
Prof. dr. ir. W. Soetaert (Ghent University)
Prof. dr. ir. W. Verstraete (Ghent University, chair)

Supervisors

Prof. dr. ir. E. J. Vandamme (Ghent University)
Prof. dr. ir. P. A. Vanrolleghem (Université Laval)

Dean

Prof. dr. ir. H. Van Langenhove

Rector

Prof. dr. P. Van Cauwenberge

'Nu heb ik, ach, de filosofie,
geneeskunde en rechten en, o spijt,
daarnaast nog de theologie
lang bestudeerd, met noeste vlijt.
Hier sta ik nu, ik arme dwaas,
niets wijzer dan 'k al was, helaas.
'k ben doctor, ben professor bovendien,
en houd nu al zo'n jaar of tien
bij hoog en laag, van vroeg tot laat
al mijn studenten aan de praat,
beseffend niets te kunnen weten;
dat heeft zich in mijn hart gevreten.
Wel ben ik wijzer dan al die apen
van hooggeleerden, schrijvers en papen,
'k word niet gekweld door vrome twijfel,
ben ook niet bang voor duivel of hel-
maar toch, mijn vreugde is gevlogen:
geen kennis waar ik op kan bogen,
geen mens die ik iets heb te leren
of tot iets hoger kan bekeren.
Ook heb ik nergens geld of goed,
niemand die mij met eerbied groet.
Geen hond die zo zou willen leven!
Dat heeft mij tot de magie gedreven:
wie weet, als ik naar geesten luister
komt eindelijk meer licht in 't duister.
Dan moet ik niet meer, klam van 't zweet,
verkondigen wat ik niet weet,
maar krijg te zien welk krachtenspel
ten grondslag ligt aan dit bestel,
'k doorgrond de zaden en het rijpen
en hoef niet steeds naar 't woord te grijpen.

O volle maan, zag jij me maar

voor 't laatst achter mijn lessenaar,
waar ik je vaak om middernacht
met pijn in 't hart heb opgewacht,
dan, boven boeken en papier,
mijn bleke vriend, verscheen je hier!
Kon ik maar door 't gebergte dwalen
in 't zachte schijnsel van je stralen,
geesten opzoeken in hun holen
langs schemerende weitjes dolen
en niet geplaagd door muizenissen
me heilzaam in je dauw verfrissen

God weet hoelang ik mij al kwel
in mijn vervloekte, mufte cel
waar 't hemellicht niet langer straalt
maar in het glas-in-loot verschaalt!
met boekenzerk als struikelblok,
leesstof voor made, luis en spint;
een steil gewelf, tot in de nok
met kladpapiertjes volgepind;
een lorrenboedel, eeuwenoud,
met kolven, vaten, waar ik kijk,
en instrumenten volgestouwd:
dat is je wereld, dat is je rijk!

En vraag jij nog wat het kan zijn
dat jou vanbinnen zo beklemmt,
door welke mysterieuze pijn
je levenslust zo gestremd?
Terwijl het rondom klopt en bruist
in Gods natuur, grijnzen je hier,
in walm en kelderger behuisd,
de schedels toe van mens en dier.
Vlucht! Zoek de oneindige natuur!' [65]

Dankjewel,

Gino Baart, Leen Baert, Joeri Beauprez, Marianne Becquevort, Gisèle Bertens, familie Bertens, Aditya Bhagwat, Gertrude Boddin, Arnaud Boland, Andrea Casadio, Daniel Charlier, Petra Claeys, Lies Clarysse, Rony Cocquyt, Lluís Corominas, Raymond Cunin, Bernard De Baets, Bert De Boeck, Dominique Delmeire, Sofie De Maeseneire, Marjan De Mey, Dirk De Pauw, Ellen Deschepper, Jenny Devriese, Brecht Donckels, Margriet Drouillon, Joel Ducoste, Wout Duthoo, Sem Flamez, FWO-Vlaanderen, Annie Gevaert, Veerle Gevaert, Sef Heijnen, Dominique Lebbe, Gaspard Lequeux, Jin Il Kim, Dominick Maes, Frédéric Mestdagh, Eric Maertens, Evy Maertens, Nick Maertens, familie Maertens, Maria Foulquie Moreno, Ellen Mortier, Nandkishor Nere, Jan Pieters, Peter Pipelers, de prof, Cedric Raemdonck, Doraiswami Ramkrishna, Sophie Roelants, Karl Rumbold, Vivianne Rys, Bram Slabbinck, Wim Soetaert, Katrijn Standaert, Kathy Steppe, Anke Story, Hilal Taymaz, Jan Vanbiervliet, Roos Van Cauwenberghe, Sam Van Den Broeck, Philip Van Driessche, Ann van Griensven, Walter van Gulik, Ellen Van Horen, Katja Van Nieuland, Peter Vanrolleghem, Maarten Veevaete, Elisabeth Vercammen, Pieter Vergult, Timpe Vogelaers, Eveline Volcke, Hendrik Waegeman, Heidi Wouters,

voor je weet wel wat.

JM

List of abbreviations

| | |
|--------|---|
| ADP | Adenosine diphosphate |
| ATP | Adenosine triphosphate |
| cAMP | Cyclic adenosine monophosphate |
| CFD | Computational fluid dynamics |
| DNA | Deoxyribonucleic acid |
| EFM | Elementary flux mode |
| ENO | Enolase |
| EP | Extreme pathway |
| EW-TLS | Element-wise weighted total least squares |
| F6P | Fructose-6-phosphate |
| FIM | Fisher information matrix |
| G6P | Glucose-6-phosphate |
| GLC | Glucose |
| HM | Harmonic mean |
| HMM | Hidden Markov model |
| LHS | Left hand side |
| MCA | Metabolic control analysis |
| MFA | Metabolic flux analysis |
| MOMA | Minimisation of metabolic adjustment |
| mRNA | Messenger ribonucleic acid |
| MRF | Moving reference frame |
| PCA | Principal component analysis |
| PEP | Phosphoenolpyruvate |
| PGM | Phosphoglycerate mutase |
| PLS | Partial least squares |
| PRESS | Predictive residual error sum of squares |
| PTS | Phosphotransferase system |
| RANS | Reynolds-averaged Navier-Stokes |
| RHS | Right hand side |
| RPM | Revolutions per minute |
| SM | Sliding mesh |

Contents

| | | |
|---|--|-----|
| 1 | General introduction | 1 |
| 2 | Modelling with a view to target identification in metabolic engineering | 7 |
| 3 | Model-based optimisation of succinate production by <i>E. coli</i> | 29 |
| 4 | Cybernetics: some issues on the method | 41 |
| 5 | Identification and evaluation of approximative kinetic model structures | 55 |
| 6 | A modus operandi of the Bioscope to study oscillating microbial systems | 79 |
| 7 | Eulerian-Lagrangian description of a large-scale bioreactor: an averaging out approach | 89 |
| 8 | Design of a scaled-down reactor using computational fluid dynamics | 111 |
| 9 | Conclusions and perspectives | 127 |
| | Bibliography | 133 |
| | Summary | |
| | Samenvatting | |
| | Curriculum vitae | |

Chapter 1

General introduction

1.1 Introduction

The growing environmental concerns and the awareness that the world's oil supplies are limited, are factors prompting the chemical and biotechnological industries to explore nature's richness in search of methods to replace petroleum-based synthetics for the development of a biobased economy [58].

An entire branch of biotechnology, known as industrial biotechnology, is devoted to this. It uses living cells and enzymes to synthesise a wide range of products (Table 1.1) that are easily degradable, require less energy and create less waste during their production [58]. However, obliging such living cells to produce the compound of interest generally requires some modification of their metabolism. To more effectively adjust metabolism both experimental and mathematical tools have been developed to gather data and to extract information from these data with a view to modifying the cell's genetics. Such an optimisation is an iterative process of strain evaluation and modification that typically takes place under highly reproducible laboratory conditions, *i.e.*, in ideally pH, temperature, and dissolved oxygen controlled and ideally mixed fermentors with a hydraulic volume of a few litres.

However to produce the compound of interest in sufficient quantities to meet the commercial demand, the developed process has to be scaled-up. Then additional problems arise, *i.e.*,

Table 1.1: Microbially produced products and the producing organism [32]

| Producing organism | Product |
|------------------------------------|---|
| <i>Klebsiella pneumoniae</i> | 1,3-propanediol |
| <i>Aspergillus niger</i> | citric acid |
| <i>Aspergillus terreus</i> | itaconic acid |
| <i>Gluconobacter oxydans</i> | gluconic acid |
| <i>Actinobacillus succinogenes</i> | succinic acid |
| <i>Saccharomyces cerevisiae</i> | lactic acid |
| <i>Acetobacter suboxydans</i> | ascorbic acid |
| <i>Xanthomonas campestris</i> | xanthan |
| <i>Saccharomyces cerevisiae</i> | ethanol |
| <i>Corynebacterium glutamicum</i> | glutamic acid |
| <i>Candida flareri</i> | riboflavin (vitamin B ₂) |
| <i>Pseudomonas denitrificans</i> | cyanocobalamin (vitamin B ₁₂) |
| <i>Penicillium chrysogenum</i> | penicillin G |
| <i>Streptomyces orientalis</i> | vancomycin |
| <i>Streptomyces aureofaciens</i> | tetracycline |
| <i>Bacillus licheniformes</i> | α -amylase |

- biological factors, *e.g.*, the number of generations associated with the inoculum development and production phases, mutation probability, contamination vulnerability, pellet formation, cell-density, and selection pressure,
- chemical factors, *e.g.*, pH control agents, medium quality and water quality, and substrate concentrations, and
- physical factors, *e.g.*, mixing, aeration, agitation, and hydrostatic pressure,

are affected when scaling-up, all significantly influencing the overall process yield and productivity, most often in a negative way [70, 79, 183].

1.2 Aims

To gain insight in the factors leading to the suboptimally performing large-scale cultures in comparison with laboratory-scale cultures, a study of biological, chemical, and physical processes is mandatory. Uncoupling the underlying processes of different nature is difficult as some of the time constants are of the same order of magnitude. Indeed, transport phenomena influence the local conditions, which in turn influence microbial

metabolism, which in turn influence local process conditions.

Thus far, the attempts to really tackle this problem, though of major interest for the optimisation of a microbial production process *sensu largo*, have been little. Amongst others the tools to tackle this problem are not readily available. Hence, the aim of this dissertation was to develop and apply some of the tools that will be useful to investigate the widely observed reduction in process performance of large-scale cultures.

Therefore, this dissertation focusses on tools to describe/model the microbial metabolism in detail. The cellular response to the rapidly changing environmental conditions encountered in such large-scale bioreactors is indeed thought to be the main cause of the observed reduction in process performance. The models to be developed should consider the internal composition and structure of the micro-organisms, enzymatic kinetics, and the regulatory network.

For such models the gathering of experimental data to identify the model structure and its parameters and, equally important, to validate the model is a prerequisite. Hence, experimental set-ups need to be developed that mimic the large-scale conditions. These set-ups can then be used to collect the necessary intracellular metabolic data.

Finally, tools need to be developed that render the description of both biological, chemical, and physical processes that take place in large-scale bioreactors, using computational fluid dynamics models, feasible.

1.3 Outline

This dissertation consists of three parts:

In the first part tools to describe metabolism are discussed. To gain insight in the microbial metabolism, modelling can be a useful tool. Metabolic models *sensu largo* are already widely used for metabolic engineering purposes. Therefore in Chapter 2, a concise overview is given of the state of the art. Though perhaps not directly useful for the study of large-scale bioreactors, some of the reviewed methods have been applied as well: partial least squares regression has been used to identify genetic targets for the metabolic

engineering of succinate biosynthesis in *E. coli* (Chapter 3) and a method is presented to assess the uncertainty on the calculated flux control coefficients of a biochemical pathway, described by approximative metabolic models (Chapter 5).

In large-scale bioreactors zones exist with ample substrate, in general in the surroundings of the inlet of the concentrated influent, with substrate depletion and with oxygen depletion or excess in other zones. When an individual micro-organism circulates through a large-scale reactor it is sequentially exposed to these different local conditions [139, 221].

To study such phenomena, a detailed description of the biophase is mandatory. As the cellular response to the encountered rapidly changing environmental conditions in large-scale bioreactors is thought to be the main cause of the observed yield reduction. As a consequence of these variations in process conditions, a micro-organism will develop a characteristic metabolomic and proteomic make-up [44, 53, 76, 131], which will allow maximisation of its growth under those conditions, *e.g.*, mixed acid fermentation and overflow metabolism. In view of the latter, attention is devoted to the cybernetic framework in Chapter 4, especially with a view to a more detailed description of the biophase. Several rival control laws for enzyme activity have been proposed and evaluated. The rationale of the cybernetic school of thought is that a micro-organism tries to optimise its behaviour, *e.g.*, with respect to growth or substrate uptake. This is achieved by allocating the limited resources a micro-organism disposes of to these competing enzymes yielding the optimal performance, by means of a controller [138, 205, 206].

A model-based approach has thus been chosen since models can be useful tools as they are a special kind of ontology. To build models three sources of information are used: experimental data, prior knowledge gathered from the literature and databases, and its intended purpose. Modelling is however approximating and consequently the choice of the proper model structure is in general subject to individual judgement and preference. Finding the proper balance between the intended aim, prior knowledge, and the available data is however an assiduous task.

In view of a model's intended purpose, over-abstracting or oversimplifying reality can result in a model that is hard to interpret or a model that does not take into account processes of prime importance with a view to the model's intended purpose. For instance, this is probably the case for the model-based optimisation presented by Sin *et al.* (2004)

[165] which led to erroneous model predictions [166]. In contrast, complicating the model can result in a model that is poorly identifiable, *i.e.*, many different parameter sets will give almost identical fits to the calibration data (the equifinality problem) as 'they often can dance to the tune of the calibration data' [19], and one may run into the same danger. This may again be pernicious for its predictive validity [19] as these parameter sets can yield dramatically different predictions of how the system will behave as conditions change.

The need for reliable and informative data is then obvious. Therefore, tools are developed in the second part of this dissertation which may help to gather the necessary data to experimentally study microbial metabolism and to gather the necessary data with a view to parameter identification and model structure identification. To this end, a modus operandi of the Bioscope is proposed in Chapter 6 to study microbial oscillating systems.

A strategy to design a scaled-down reactor is outlined in Chapter 8. Scaled-down reactors allow to mimic on a laboratory-scale, the large-scale conditions in an attempt to anticipate the outcome on a large-scale. However, whereas the state of the art scaled-down reactors typically focus on macroscopic variables, such as circulation time and mixing time, the presented approach attempts, using computational fluid dynamics simulations, to more accurately mimic the substrate concentration dynamics observed by micro-organisms in large-scale bioreactors, as those macroscopic variables are far from ideal to be correlated with degrees of conversion.

In the last and third part of this dissertation attention is devoted to computational fluid dynamics. Computational fluid dynamics models find acceptance both in industry and academia to study the impact of spatiotemporal heterogeneity, *i.e.*, imperfect mixing, on overall process performance. The description of the biophase in a Lagrangian way, *i.e.*, following the cell's path through the reactor, is obvious since the behaviour of a micro-organism is determined by both the reigning environmental conditions and its intracellular make-up. All this is determined by what it has observed over time. Due to the stochastic nature of particle transport and the spatial heterogeneity in large-scale bioreactors, this intracellular make-up will not be identical for all micro-organisms which makes that a large number of cells must be followed to generate a view on the overall bioreactor behaviour. A method to render such calculations more feasible is therefore proposed in Chapter 7.

This dissertation ends with an overview of the main conclusions and perspectives for further research.

Chapter 2

Modelling with a view to target identification in metabolic engineering

2.1 Introduction

The well-established chemical synthesis routes face, although the era of the oil-based society has not come to an end yet, more and more competition from industrial biotechnological alternatives for the production of an increasing number of compounds, due to, *e.g.*, environmental concerns and the increasing scarcity of oil. Whereas in the past micro-organisms were typically used for the production of stereochemical [191] and complex molecules, *e.g.*, antibiotics [25, 177], nowadays they even become an interesting alternative for many bulk chemicals. In order to develop an industrial biotechnological process that can compete with the more mature chemical synthesis routes, there are 4 critical development phases:

1. The choice of the favourite micro-organism
2. Metabolic engineering
3. Scaling-up
4. Downstream processing

The second phase in the development of an economically viable industrial biotechnological process is the optimisation of the micro-organism itself using a wide range of both experimental and mathematical techniques.

To this end, due to the complexity of microbial metabolism, more and more metabolomic, proteomic, transcriptomic, and genomic data are collected [38, 78, 87, 132], which appear to be valuable to steer the process of genetic engineering with a view to the overproduction of a target compound. Indeed, these data help to elucidate the flux distribution, determine the flux controlling reactions, and yield insight in the regulation of metabolism.

In addition to these experimental techniques, mathematical methods are developed and commonly applied to interpret and to extract information from this pile of data and to identify genetic targets for the overproduction of a target compound (Table 2.1). In this context steady-state [198] and dynamic metabolic modelling [149], multivariate statistics [39, 84, 197], graph theory [136], and neural networks are used to unravel the microbial behaviour.

Finally, the development of genetic toolboxes consisting of promoter libraries [39, 69, 84] and strategies for gene knock-outs, knock-ins, knock-downs, and knock-ups [36], and the advent of functional genomics [61, 77] have allowed the directed improvement of cellular properties based on these findings in view of optimising the production host. This hotchpotch of techniques results after some iterative rounds of genetic modification and host evaluation into the development of a host with improved performance.

Such a systematic approach is obvious as the vast variety of biochemical pathways microorganisms dispose of, in order to fulfil their growth and reproduction requirements under a wide range of environmental conditions, renders them hard to fathom. A thorough understanding of the regulation of microbial processes is however a *conditio sine qua non* for the rational design of bioprocesses, as a disturbance in one part of metabolism can trigger a series of reactions on all levels of regulatory control and in all parts of metabolism. Indeed, in complex metabolic networks it is often a futile avocation to *ad hoc* predict the impact, both qualitatively and quantitatively, of a genetic intervention [12]. Hence, the popularity of models for metabolic engineering purposes. A concise overview of the use of models in this development phase will be given below (Figure 2.1).

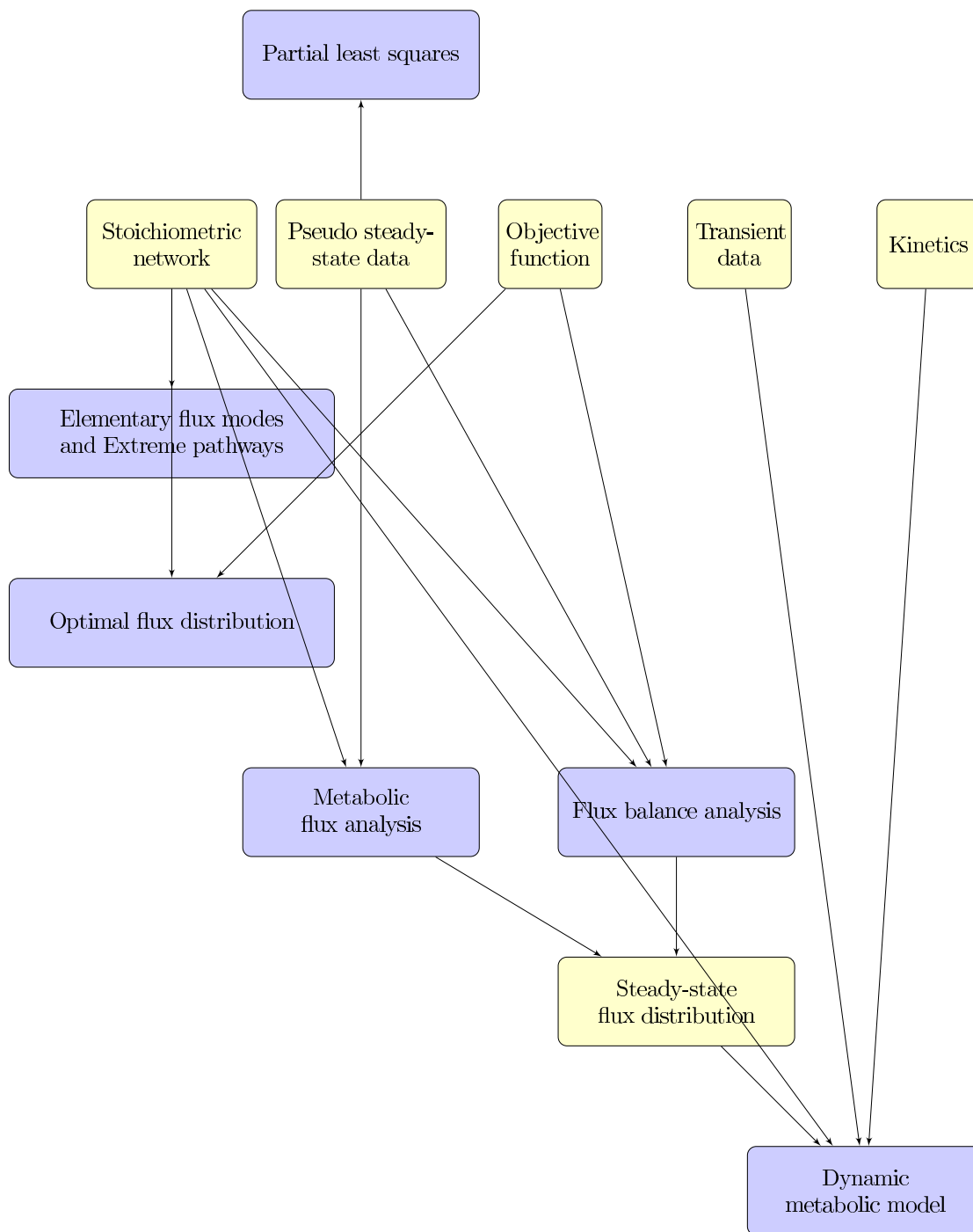


Figure 2.1: Modelling with a view to target identification in metabolic engineering. Blue blocks represent the methods, yellow blocks represent inputs.

Table 2.1: Target identification relying on metabolic modelling

| model-based optimisation method | Production host | Target compound |
|---------------------------------|------------------------|----------------------------|
| elementary flux modes | <i>E. coli</i> | L-methionine [93] |
| | <i>C. glutamicum</i> | L-methonine [93] |
| optimal flux distribution | <i>E. coli</i> | succinic acid [this study] |
| | <i>E. coli</i> | succinic acid [99] |
| flux balance analysis | <i>M. tuberculosis</i> | mycolic acid [144] |
| | <i>S. cerevisiae</i> | succinic acid [137] |
| | <i>S. cerevisiae</i> | glycerol [137] |
| | <i>S. cerevisiae</i> | vanillin [137] |
| | <i>E. coli</i> | lycopene [5] |
| | <i>E. coli</i> | L-threonine [98] |
| | <i>E. coli</i> | L-valine [135] |
| partial least squares | <i>E. coli</i> | succinic acid [34] |
| | <i>E. coli</i> | phenylalanine [200] |
| | <i>Trichoderma sp.</i> | cellulase [196] |
| dynamic metabolic modelling | <i>E. coli</i> | carnitine [27, 163] |

2.2 Stoichiometric network analysis

Ab initio, stoichiometric network models have been used to facilitate the choice of where to intervene genetically. The metabolic network comprises the metabolites and the reactions they are involved in, including formation, degradation, transport, and cellular utilisation gathered from databases [88, 128, 158] and the literature [148]. For every metabolite a mass balance can be derived:

$$\frac{dx_i}{dt} = \sum_j s_{ij}r_j - b_i \quad (2.1)$$

where s_{ij} is the stoichiometric coefficient associated with flux r_j and b_i the net transport flux of metabolite x_i . Under pseudo steady-state conditions Eq. 2.1 will reduce to:

$$0 \cong \sum_j s_{ij}r_j - b_i \quad (2.2)$$

Eq. 2.2 can be rewritten in matrix notation:

$$\begin{bmatrix} 0 \\ b \end{bmatrix} \cong S \times R \quad (2.3)$$

where S is the stoichiometric matrix, R is the vector of metabolic fluxes, and b is the vector representing m transport fluxes.

Despite success stories of metabolic model use to identify targets for modification, there have also been many false positive targets identified by these models. It is still unclear whether the well-established technique of stoichiometric modelling is fully apt to steer the process of metabolic engineering, since the kinetics and the regulation of the enzymatic reactions are not accounted for [5, 171].

Especially for the optimisation of the production of metabolites in primary metabolism that are subject to severe (redox) constraints, stoichiometric modelling is useful. It is less so for the optimisation of minor routes [195].

Once the metabolic network model is built one can resort to stoichiometric network analysis, in the absence of data. Network analysis provides for the identification of elementary flux modes, extreme pathways, and the optimal flux distribution as will be discussed below.

2.2.1 Elementary flux modes and extreme pathways

Network-based pathway analysis, *e.g.*, identification of elementary flux modes (EFMs) and extreme pathways (EPs) facilitates the assessment of network properties. Both of these methods use convex analysis, a branch of mathematics that enables the analysis of inequalities and systems of linear equations to generate a convex set of vectors that can be used to characterise all of the steady-state flux distributions of a biochemical network [134]. Both have the following properties [134]:

1. There is a unique set of elementary modes/extreme pathways for a given network.
2. Each elementary mode/extreme pathway consists of the minimum number of reactions that is required to exist as a functional unit. If any reaction in an elementary mode/extreme pathway would be removed, the whole elementary mode/extreme pathway could not operate anymore as a functional unit. This property has been called genetic independence and non-decomposability.

However, whereas elementary modes are the set of all routes through a metabolic network consistent with the latter property, extreme pathways are the systemically independent subset of elementary modes (Figure 2.2); that is, no extreme pathway can be represented as a non-negative linear combination of any other extreme pathways [134].

Both have been used to calculate product yields, to evaluate pathway redundancy, to determine correlated reaction sets, and to assess the effect of gene deletions [134]. Carlson *et al.* (2002) [28] and Kromer *et al.* (2006) [93] used elementary flux modes for rational design purposes and Carlson and Sscienc (2003) [29], Nookaew *et al.* (2007) [126], and Schwarts *et al.* (2007) [160] used the concept of elementary flux modes in combination with experimental data for network analysis.

The physiological interpretation of the results, see also Figure 2.3, and their computation for genome scale models remain however challenging [216].

2.2.2 Optimal flux distribution

The calculation of the optimal flux distribution, *e.g.*, [99], is another popular information source to steer the process of metabolic engineering. The linear programming problem can be written as:

$$\max J = b_i \tag{2.4}$$

subject to:

$$0 \cong S \times R - b \tag{2.5}$$

$$b_i = \alpha_i \tag{2.6}$$

where J is the objective function, typically the net transport flux of the compound of interest, and α_i the constraint on the net transport flux values of certain substrates i . By solving Eqs. 2.4-2.6, the maximal theoretical yield can be calculated.

Pros and cons

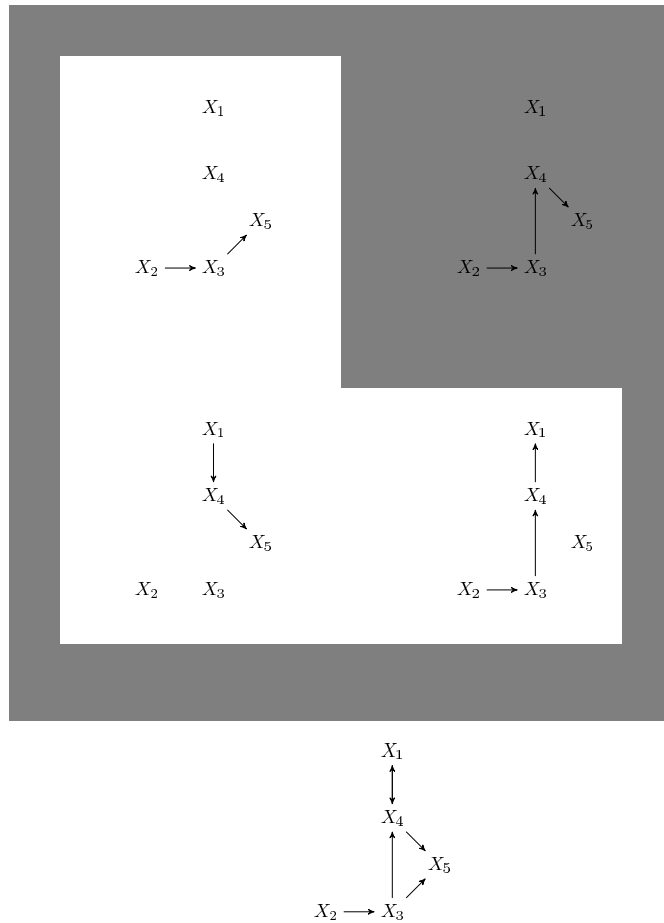


Figure 2.2: The 3 extreme pathways (white background) and 4 elementary flux modes (grey background) of the stoichiometric network. Note that the EFM (top, right) is a non-negative linear combination of 2 extreme pathways (down, right) and (down, left)

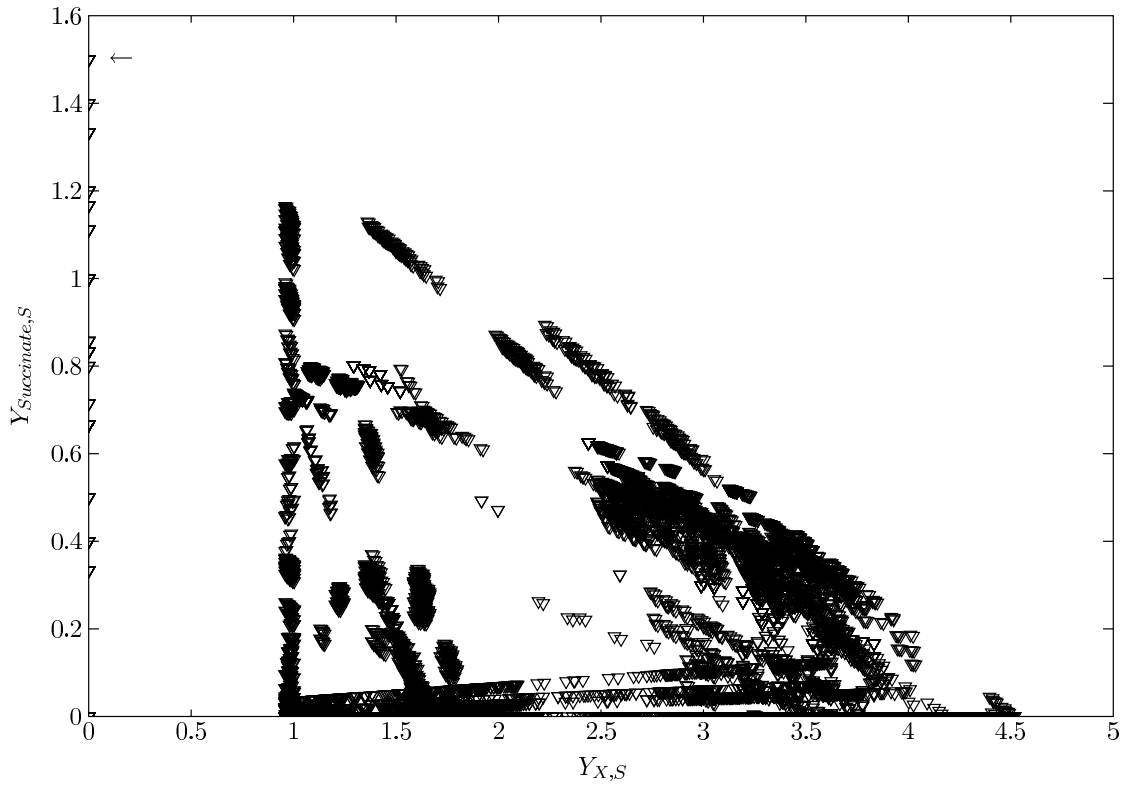


Figure 2.3: The 17528 elementary flux modes of the stoichiometric *E. coli* model of Lequeux *et al.* (2006) [100] represented as ∇ s, calculated by using Metatool 5.0 [214], and presented in the $Y_{X,S}$, $Y_{succinate,S}$ space, with $Y_{X,S}$ and $Y_{succinate,S}$ the biomass $\left[\frac{c-mole}{mole}\right]$ and succinate $\left[\frac{mole}{mole}\right]$ yield on glucose, respectively. The EFMs characterised by the optimal flux distribution, here with respect to maximal $Y_{succinate,S}$ can readily be identified (\leftarrow).

Obviously, the optimal flux distribution can be calculated, but how to achieve this optimal flux distribution *in vivo* remains unresolved since it depends considerably on the kinetics and the regulation of the enzymatic reactions, which are not accounted for [5, 171]. In addition, it focusses completely on yields whereas in reality productivity, *i.e.*, the rate at which the product is produced, is equally important.

2.3 Steady-state modelling

In the presence of data, one can resort to steady-state modelling, *e.g.*, metabolic flux analysis and flux balance analysis. Eq. 2.3 can then be rewritten as:

$$0 \cong \begin{bmatrix} S_{in} & 0 & 0 \\ S_{ex}^c & -I_{ex}^c & 0 \\ S_{ex}^m & 0 & -I_{ex}^m \end{bmatrix} \begin{bmatrix} r_{in} \\ b_{ex}^c \\ b_{ex}^m \end{bmatrix} \quad (2.7)$$

where r_{in} represents the intracellular fluxes, b_{ex}^c and b_{ex}^m the net transport fluxes to be calculated and measured, respectively. S_{in} , S_{ex}^c , and S_{ex}^m are the corresponding stoichiometric matrices and I represents a unity matrix. This equation can be rewritten as:

$$0 \cong \underbrace{\begin{bmatrix} S_{in} & 0 \\ S_{ex}^c & -I_{ex}^c \\ S_{ex}^m & 0 \end{bmatrix}}_{W_c} \underbrace{\begin{bmatrix} r_{in} \\ b_{ex}^c \end{bmatrix}}_{a_c} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ -I_{ex}^m \end{bmatrix}}_{W_m} \underbrace{\begin{bmatrix} b_{ex}^m \end{bmatrix}}_{a_m} \quad (2.8)$$

the solution of which is:

$$a_c \cong -W_c^\# W_m a_m + \text{null space}(W_c) f \quad (2.9)$$

with $W_c^\#$ the pseudo inverse of matrix W_c , with the null space defined as the set of linear independent basis vectors R_n that fulfil the equation:

$$W_c R_n = 0 \quad (2.10)$$

and f a vector with as many elements as there are columns in the null space of W_c . The

number of independent null space vectors is equal to:

$$n - \text{rank}(W_c) \quad (2.11)$$

with n the number of fluxes to be calculated.

2.3.1 Metabolic flux analysis

If Eq. 2.11 = 0 the system is determined and has one unique solution:

$$a_c \cong -W_c^\# W_m a_m \quad (2.12)$$

Substituting Eq. 2.9 in Eq. 2.8 now yields:

$$W_m a_m + W_c \left(-W_c^\# W_m a_m \right) = 0 \quad (2.13)$$

Or rewritten, since Eq. 2.10:

$$\left(W_m - W_c^\# W_c W_m \right) a_m = 0 \quad (2.14)$$

When the system is (partially) overdetermined, the extra measurements, which are specified by the so-called redundancy matrix: $W_m - W_c^\# W_c W_m$ in Eq. 2.14, can be used for statistical testing and error analysis. van der Heijden *et al.* (1994) [194] introduced a method for error detection and analysis. If the error is statistically zero, the model is consistent. If this is not the case, this error analysis method can be used to identify erroneous measurements. Removing these erroneous measurements from the data set improves the chance of a consistent result. An overview is given by Lequeux *et al.* (2006) [100].

Pros and cons

Though metabolic flux analysis (MFA) merely yields a snapshot of the metabolic state in a particular condition, it might be of some significance to steer the process of metabolic engineering as principal nodes can be identified. These principal nodes, which are characterised by significant changes in flux partitioning under different conditions, should be

regarded as potential bottlenecks [198].

It should be clear that due to the large variety of metabolic pathways, *e.g.*, parallel pathways, reversible reactions, and cycles the system is in general underdetermined. For example, genome scale models have been constructed, typically useful for the design of minimal media, *e.g.*, for *Escherichia coli* (931 reactions) [148], *Saccharomyces cerevisiae* (1175 reactions) [57], *Helicobacter pylori* (388 reactions) [157], and *Neisseria meningitidis* (496 reactions) [10]. Such genome scale models contain all known reactions, formation, degradation, transport, and cellular utilisation gathered from databases and the literature.

However, then the modeller can opt/has to reduce the metabolic network in order to get a system of feasible size, using for example an objective function or by incorporating as much knowledge, *e.g.*, prior knowledge about the flux size, and data as possible. For instance, unlabelled metabolomic data, as these data yield thermodynamic information $\Delta_r G'^o$ and consequently information about the reversibility and irreversibility of certain reactions [75, 95], labelled metabolomic data as these data yield information on split ratios [37, 201, 217], and transcriptomic data [4, 24, 100, 164], through the incorporation of additional constraints for the metabolic network, *e.g.*, presence of an enzymatic conversion, to reduce the uncertainty about the obtained flux distribution.

2.3.2 Flux balance analysis

If Eq. 2.11 > 0 the system is underdetermined. Then no unique solution exists. The question then is which of the feasible metabolic states is manifested under that condition. Flux balance analysis (FBA) [51, 142] postulates that a metabolic system exhibits a metabolic state that is optimal under some criteria. This objective is expressed as a linear combination of the fluxes contained in R . The model can then be formulated as a linear programming problem as follows:

$$\max J = \sum_j c_j r_j \quad (2.15)$$

subject to:

$$0 \cong S \times R - b \quad (2.16)$$

$$\alpha_i \leq r_i \leq \beta_i \quad (2.17)$$

where J is the objective function, c is a vector of weights, costs or benefits, linked to the fluxes, and the boundaries α_i and β_i represent known constraints on the minimum and maximum flux values.

Pros and cons

Though many objective functions have been used, the optimisation of ATP production and the optimisation of growth comply best with experimental observations [159] in many micro-organisms. The applications of FBA have been many and the *in silico* metabolic constraints predictions can be used to optimise the behaviour of interesting mutants. However, it is not because a micro-organism has the genetic potential that it will *ad hoc* perform optimally, *i.e.*, mutants created artificially are generally not subject to the same evolutionary pressure that shaped the wild type [5, 56, 162]. The method of minimisation of metabolic adjustment (MOMA) attempts to deal with this issue. Instead of maximising biomass production the mutant, *KO*, is believed to remain initially as close as possible to the wild type optimum, *WT*, in terms of flux values [162]. The objective function then becomes:

$$\min D(R_{WT}, R_{KO}) \quad (2.18)$$

with

$$D(R_{WT}, R_{KO}) = \sqrt{\sum_i^n (r_{WT} - r_{KO})^2} \quad (2.19)$$

This method heavily relies on prior knowledge (through the constraints introduced in Eq. 2.17), but at present the knowledge on the regulatory mechanisms is still lacking and fragmentary [66, 89]. In addition, in some cases no unique optimum exists and consequently many metabolic states may result in the same optimal behaviour [56].

2.4 Dynamic metabolic modelling

2.4.1 Mechanistic - approximative models

Due to the above-mentioned limitations of stoichiometric modelling, kinetic equations have been introduced in metabolic models. The general form of the mass balances of extracellular and intracellular metabolites is now given by Eq. 2.20 and Eq. 2.21, respectively:

$$\frac{dx_{S_i}}{dt} = D(x_{S_i}^0 - x_{S_i}) - \frac{x_X}{\rho_X} \sum_j s_{S_i j} r_j \quad (2.20)$$

$$\frac{dx_{M_i}}{dt} = \sum_j s_{M_i j} r_j - \mu x_{M_i} \quad (2.21)$$

with x_{M_i} and x_{S_i} the concentration of an intracellular metabolite M_i and an extracellular metabolite S_i , respectively, $s_{M_i j}$ is the stoichiometric coefficient of metabolite M_i in reaction j , r_j the rate of reaction j , ρ_X the specific weight of biomass, x_X the biomass concentration, D the dilution rate, x_S^0 the concentration of an extracellular metabolite S in the feed, and μ the specific growth rate. Note that x_S is expressed per reactor volume whereas x_M is expressed per cell volume. The term μx_M in the mass balances of the intracellular metabolites represents the dilution effect due to growth.

In mechanistic dynamic metabolic modelling, one can resort to complex *in vitro* determined mechanistic equations to describe the rate equations r_j in Eqs. 2.20-2.21 [30, 42, 149, 207].

In approximative modelling, one can resort to linear non-mechanistic kinetics to describe the rate equations r_j in Eqs. 2.20-2.21, *e.g.*, the loglinear approximation [72, 73], the GMA type power law approximation Eq. 2.22 [154], the thermokinetic approximation Eq. 2.23 [215], and the linlog approximation Eq. 2.24 [74, 209].

$$\ln\left(\frac{r_j}{J^0}\right) = \ln\left(\frac{x_E}{x_E^0}\right) + \sum_{i=1}^n \varepsilon_{M_i}^0 \ln\left(\frac{x_{M_i}}{x_{M_i}^0}\right) \quad (2.22)$$

$$\frac{r_j}{J^0} - 1 = \ln\left(\frac{x_E}{x_E^0}\right) + \sum_{i=1}^n \varepsilon_{M_i}^0 \ln\left(\frac{x_{M_i}}{x_{M_i}^0}\right) \quad (2.23)$$

$$\frac{r_j}{J^0} = \left[\frac{x_E}{x_E^0}\right] \left(1 + \sum_{i=1}^n \varepsilon_{M_i}^0 \ln\left(\frac{x_{M_i}}{x_{M_i}^0}\right)\right) \quad (2.24)$$

where the superscript ⁰ stands for the steady-state condition and with x_E the enzyme concentration, $\varepsilon_{M_i}^0$ an elasticity coefficient, and J^0 the steady-state flux. The applied equations are not as complex as mechanistic rate equations and contain less parameters to approximate the true kinetics. The rationale behind this is that metabolic redesign does not require detailed mechanistic models because of the concept of homeostasis, which implies that the micro-organism keeps its intracellular metabolite levels approximately constant [54, 170, 184]. In other words, the extrapolation range of the kinetic metabolic model does not need to be very large, as far as metabolite levels are concerned. This reasoning suggests that one can safely apply approximative kinetic equations instead of the detailed mechanistic ones that are valid over a wide range of concentration levels.

Pros and cons

The enormous variety of well regulated metabolic pathways impedes a thorough understanding of the regulation of microbial processes on the metabolomic, proteomic, transcriptomic, and genomic level in a qualitative and quantitative way. Such understanding would be beneficial for the rational design of bioprocesses, as a genetic or environmental disturbance in one part of metabolism can trigger a series of reactions on all levels of regulatory control and in all parts of metabolism [12]. Hence, in many applications, *e.g.*, metabolic engineering, 'whole cell modelling' is probably the way to go [186, 187].

However, since the knowledge about the transcriptional and translational regulation is still fragmentary, the state of the art dynamic metabolic models typically focus on the metabolome, assuming constant proteomic levels. In view of the extrapolation capacity of these models this is a drawback. Hence, in order not to violate this assumption of steady-state proteome, data for parameter identification have to be collected during a relatively short period after perturbation, this is typically within 0.2-180 s [181, 182, 207, 211].

In addition, dynamic metabolic models typically zoom in on a limited part of the micro-

bial metabolism. The resulting model typically contains a number of fluxes towards parts of the metabolism which one is not primarily interested in. When the model contains two or more of those fluxes this will create some uncertainty about the dynamic flux distribution (a steady-state model yields the steady-state flux distribution, but there is only indirect, secondary information about the dynamic evolution of the flux distribution after a perturbation of the metabolism). Only having the information of the evolution of metabolite concentrations is insufficient for these aims. Thus, in contrast to steady-state modelling, where mass balances are essential to verify the accuracy of the calculated fluxes, this check is not performed in most dynamic metabolic models [30, 60, 149, 207], as the size of the out flux of the model is not known. It should however be clear that modelling the whole metabolism would be a daunting task as well: when a perturbation passes through the metabolic network it broadens and dampens out and the information content of such data collected further on in the network is limited.

In order to reduce this uncertainty the cofactors might be used as 'closure terms', *e.g.*, the generation of NADPH, might be a good indicator for the flux through the pentose phosphate cycle. However, it should be clear that these closure terms are weak as cofactors intervene in many reactions, which are also perturbed during a pulse experiment. In addition, modelling these cofactors dynamically is not easy at all because this approach is hampered, for instance, by the inability to explain the short-term reduction in the pool size of the adenine nucleotides (AXP) after a glucose pulse [30, 207]. At present, it is still unclear what is/are the cause(s) of this reduction (adaptation would only be responsible for 15% of this gap [91], formation of adenylated compounds, *e.g.*, ADP-glucose, excretion of cAMP, ...).

Therefore, some researchers opt to describe the evolution of the cofactors as time dependent functions [30], which results in a model that is no longer useful for extrapolation. Not taking these cofactors into account 'mechanistically' thus results in a limited usefulness of the resulting model. Then, also assumptions have to be made about the evolution of the flux distributions during the transient but it is questionable whether these hold.

In order to reduce the uncertainty, one could gather a lot of data both under steady-state and dynamic conditions, *e.g.*, by perturbing the microbial cells with different substrates. However, such efforts have thus far been limited [114, 211].

The use of dynamic labelling data [213] allows as well to reduce the degrees of freedom related to the metabolic fluxes, also under dynamic conditions. However the huge variety of biochemical pathways will render such an exercise tricky, as the chosen metabolic network will influence the calculated flux distribution [111].

In addition, one should be aware that a lot of challenges still remain in the field of analytical methods, since the accurate determination of the intracellular metabolites is a considerable task as well, due to, *e.g.*, leakage and their low concentrations [49, 113, 130, 208]. For example, whereas the expected (equilibrium) ratio of the concentrations of glucose-6-phosphate [G6P] and fructose-6-phosphate [F6P], *i.e.*, $\frac{[F6P]}{[G6P]} \cong 0.25$ [18, 30, 114, 180], Bucholz *et al.* (2001) [21] find for this ratio $\frac{[F6P]}{[G6P]} \cong 0.88$.

Another issue is that the state of the art dynamic metabolic models either rely on *in vitro* determined kinetic equations or are based on approximative kinetics [30, 180, 209] and the consequences of a potentially erroneous model structure are not well known. With respect to the *in vitro* determined kinetic equations it is doubtful whether the kinetics are valid under *in vivo* conditions, as these kinetics are obtained using purified enzymes studied out of context [156, 180].

The variety of well regulated metabolic pathways also impedes a thorough understanding of the regulation of microbial processes, *e.g.*, the relative importance of the flux through pyruvate oxidase compared to the flux through pyruvate dehydrogenase is not that clear [2, 103]. Another example is the jumble of reactions around the PEP-pyruvate-oxaloacetate node. Their regulation and importance under one or the other condition is still not that well studied [26, 101, 153]. The inability to properly describe the dynamics of phosphoenolpyruvate (PEP) during the observation window of a perturbation experiment [30, 149], though a key metabolite in the primary metabolism, is the perfect illustration that setting up a metabolic model in a proper way will be demanding both for modellers and for experimentalists.

Models, whether they are approximative or mechanistic, can be useful to identify bottlenecks [54, 72, 85, 115] in metabolism and consequently could steer the process of metabolic engineering. However, since enzyme levels are not taken into account nor the influence of a genetic intervention on the metabolism, it should be clear that the extrapolation power of such models remains limited.

2.4.2 Cybernetic models

At present, one can not see the wood for the trees as the knowledge on the regulatory mechanisms is lacking and fragmentary [66, 89]. To partially circumvent this knowledge gap, the cybernetic framework can be used, since microbial species, that is, those that have undergone the process of evolution, strive to regulate their metabolism in an optimal manner [56, 112]. This reasoning is the rationale of the cybernetic school of thought: a micro-organism tries to optimise its behaviour, *e.g.*, with respect to growth or substrate uptake. This is achieved by allocating the limited resources a micro-organism disposes of to these competing enzymes yielding the optimal performance [138, 205, 206]. To this end, the cybernetic variables u and v are introduced into a kinetic model Eqs. 2.20-2.21 with the aim of substituting the unknown mechanistic details of the cell's regulatory architecture by an objective function by supposing that the metabolism of a micro-organism operates with a specific overall goal, such as the optimisation of growth.

Initially, the value of the cybernetic approach was demonstrated using relatively simple examples, typically situated in the domain of bioprocess control. In these cases, some lumped pathways competed with each other for the available resources, *e.g.*, simultaneous and sequential substrate utilisation [90] and single-substrate growth [13, 14, 190]. Then the cybernetics units could readily be identified. Enzymes belonging to the same cluster compete with each other for the same pool of resources.

Over time more challenging 'proofs of principle' were chosen, *e.g.*, in view of metabolic engineering of a production host [204, 206, 223], and the model's complexity increased. More complex networks, without lumping were considered [143, 172], but then a jumble of cybernetic units could be identified and the corresponding cybernetic variables had to be derived from the control laws. As a result, the choice of the cybernetic units became less straightforward, even quite arbitrary, and the library of cybernetic units had to be extended (convergent, divergent, linear, and cycles) [205, 206].

To overcome this, a more general framework was developed, based on the principles of optimal control theory [223]. Optimal control theory is a mathematical optimisation method for deriving control policies. It aims to find a control law for a given system such that a certain optimality criterion is achieved. In general, such a control problem

includes a gain function and a cost function relating state and control variables. An optimal controller is a set of differential equations describing the paths of the control variables that maximise the performance function. Rephrasing this in the context of a micro-organism, the cost becomes, *e.g.*, the pool of amino acids a micro-organism needs to invest for the production of a particular enzyme, and the cell’s gain could be merely growth. Young (2005) [223] opted for EFMs as cybernetic units. As elementary modes appear to be useful to understand cellular objectives for the overall metabolic network [169], the choice for the EFMs as local control level seems quite obvious. However, the choice of the associated objective function is less so. Young (2005) [223] opted for the optimisation of a harmonic mean flux J :

$$J = \frac{\sum_{i=1}^n \xi_i}{\sum_{i=1}^n \frac{\xi_i}{r_i v_i}} \quad (2.25)$$

with n the number of reactions involved in the elementary flux mode, r_i the rate of reaction i , v_i the cybernetic variable controlling enzyme activity, and ξ_i the flux through reaction i in the elementary flux mode. This objective function aims at a steady throughput through the EFM, and consequently accumulation or depletion of certain metabolites is avoided. However, its biological foundation seems difficult to grasp.

Pros and cons

Cybernetic models consider both metabolome and proteome. They apply principles of control theory with the aim of substituting the unknown mechanistic details of the cell’s regulatory architecture by an objective function by supposing that the metabolism of a micro-organism operates with a specific overall goal. Such models are therefore thought to have more extrapolation power. Although the approach thus seems appealing, given the present lack of knowledge and detailed experimental omics data and the aforementioned problems linked to mechanistic modelling, there still remain some issues unresolved: i) it is still unclear to what extent unknown regulatory mechanisms can be captured by the framework, ii) the robustness of the approach is unclear, *e.g.*, although cybernetic models are said to be able to properly describe steady-state multiplicity [122, 123], real experimental evidence to support such a claim is lacking, iii) though the cybernetic approach is a minimalistic approach, contrary to mechanistic models containing complex kinetics with a large number of (unidentifiable) parameters [42], the incorporation of enzymes and the parameters for enzyme synthesis and degradation results in many parameters

that are difficult to estimate, and iv) for even relatively small networks the number of EFMs is huge, *e.g.*, for the metabolic network of [100] 17528 EFMs are found, which use glucose as carbon source. Which EFMs to choose, remains a question hard to answer.

2.5 Multivariate statistics

Finally, multivariate statistics, principal component analysis (PCA) and partial least squares (PLS) [219, 220], are more and more used in the field of metabolism studies [39, 84, 105, 197] to interpret and to extract information from the pile of metabolomic, transcriptomic, and genomic data. By applying these methods, targets can be identified in view of further improving production hosts. Especially the use of partial least squares seems promising. The objective in PLS modelling is to find a few 'new' variables, X-scores, in such a way that the information in the dependent variables Y can be predicted as well as possible.

In fact, this projection method decomposes variables of high collinearity into one-dimensional variables, *i.e.*, an input score vector t and an output score vector u , which allows PLS to handle many and correlated predictor variables [220]. The vectors t_1 and u_1 are defined as [104]:

$$t_1 = E_0 w_1 \tag{2.26}$$

$$u_1 = F_0 c_1 \tag{2.27}$$

where E_0 is the standardised data matrix from X and F_0 is the standardised data matrix from Y [193]. The aim of this data pretreatment is to focus on the (relevant) biological information by emphasising different aspects in the data, for instance, the value of a variable relative to its average value and to reduce the influence of disturbing factors, *e.g.*, measurement noise [193]. Hence, the regression formulae for components t_1 and u_1 are given by:

$$E_0 = t_1 p_1^T + E_1 \tag{2.28}$$

$$F_0 = u_1 q_1^T + F_1 \tag{2.29}$$

where p_1 and q_1 are the loading vectors, and E_1 and F_1 are residual matrices. The linear relationship between t_1 and u_1 is calculated by:

$$u_1 = b_1 t_1 + r_1 \quad (2.30)$$

where b_1 is the regression coefficient and r_1 is the residual vector. If t_1 and u_1 cannot explain the model within a specified precision or do not contain enough information, E_0 and F_0 will be replaced by the residual matrices E_1 and F_1 . Consequently, the next latent variable vectors t_2 and u_2 are calculated by:

$$t_2 = E_1 w_2 \quad (2.31)$$

$$u_2 = F_1 c_2 \quad (2.32)$$

The regressions for components t_2 and u_2 are therefore calculated by:

$$E_1 = t_2 p_2^T + E_2 \quad (2.33)$$

$$F_1 = u_2 q_2^T + F_2 \quad (2.34)$$

This iterative procedure is repeated by using the regression residual terms obtained at the previous iteration on both the inputs and outputs at each step. The decomposition of E_0 and F_0 by score vectors is defined by:

$$E_0 = \sum_{h=1}^m t_h p_h^T + E \quad (2.35)$$

$$F_0 = \sum_{h=1}^m u_h q_h^T + F \quad (2.36)$$

Where p and q are loading vectors, E and F are residuals. For the m choice, a cross-validation method can be applied or the threshold variance of E can be used as stopping criterion [219].

In PLS one can calculate a similar kind of regression coefficients as one does in multiple linear regression. These regression coefficients relate matrix X directly to Y :

$$Y = XB + \varepsilon \tag{2.37}$$

Both regression coefficients and loading weights can be used to study the system. Note that these regression coefficients are not independent unless the number of partial least squares regression components equals the number of X -variables. By studying the loading weights, one can see how important the variable is in each latent variable. A large positive or negative weight value indicates that the corresponding X variable is highly correlated with the values in the score matrix U and hence with matrix Y . Correlation between variables can be verified by looking at the loading weights [179, 219].

van der Werf *et al.* (2005) [197] applied PLS regression to link metabolite levels to the microbial phenotype, *i.e.*, by ordering the importance of the metabolites by virtue of the weight factors, metabolites that contributed most to the phenotype of interest could be identified.

Pros and cons

van der Werf *et al.* (2005) [195, 197] successfully applied this method to select targets in view of optimisation. Such models are however completely data driven and consequently do not use the state of the art knowledge. In addition, relationships can be found between, for instance, metabolite pool sizes and a process parameter, but how to modify the cell with a view to improving the process performance remains unclear.

2.6 Conclusions

It should be clear that despite the vast lack of knowledge about the cell's regulatory architecture, the application of both experimental techniques and mathematical methods steadily yields valuable information about microbial metabolism [30, 39, 84, 136, 197, 198]. In the future, this may unambiguously guide the process of metabolic engineering.

The contemporary lack of knowledge about the functioning of the cell is however limiting the use and usefulness of many of those techniques to steer the process of metabolic engineering, *e.g.*, at present a lot of mathematically relevant questions remain unanswered, *e.g.*, which network and objective function to choose? Consequently, the era of typi-

cal data mining techniques which are useful to help unravel the complex regulation of microbial metabolism has not come to an end yet.

Chapter 3

Model-based optimisation of succinate production by *E. coli*

3.1 Introduction

Micro-organisms are already widely used for producing antibiotics, therapeutic proteins, food and feed ingredients, fuels, and vitamins. Nowadays, due to the environmental concerns and the increasing scarcity of oil, industrial biotechnological processes become an alternative for the production of an increasing number of compounds, that are typically produced using well-established chemical synthesis routes [58]. Speeding up and reducing the cost of the development of such processes is crucial to be competitive against the petroleum-based alternatives. A systematic approach, using metabolic modelling, is thought to contribute to speed up and reduce the cost of the development of commercially viable industrial biotechnological processes.

Nevertheless, expert knowledge, educated guesses, and gut feeling are still often directing the process of metabolic engineering in view of enhancing the microbial production of the target compound [152, 197], despite numerous examples where the construction of a producer strain did not turn out to be as straightforward as was initially presumed. Indeed, in complex metabolic networks, it is often a difficult task to predict the impact, both qualitatively and quantitatively, of a genetic intervention [12]. This complexity is also reflected in the metabolic models (Table 3.1). The state of the art models, genome scale models, typically consider all known reactions, formation, degradation, transport, and cellular utilisation gathered from databases and the literature.

Table 3.1: Genome-scale models

| Micro-organism | number of reactions considered | |
|---------------------------------|--------------------------------|-------|
| <i>Escherichia coli</i> | 931 | [148] |
| <i>Saccharomyces cerevisiae</i> | 1175 | [57] |
| <i>Helicobacter pylori</i> | 388 | [157] |
| <i>Neisseria meningitidis</i> | 496 | [10] |

Metabolic models can help to identify genetic targets for metabolic engineering. For instance, elementary flux modes have been used for the optimisation of L-methionine biosynthesis by *E. coli* and *C. glutamicum* [93]. Flux balance analysis has been applied for the optimisation of lycopene [5], L-threonine [98], L-valine [135], and succinic acid [34] biosynthesis by *E. coli* and of glycerol and vanillin biosynthesis by *S. cerevisiae* [137]. Partial least squares has successfully been used for the optimisation of phenylalanine biosynthesis by *E. coli* [200] and of cellulase biosynthesis by *Trichoderma sp.* [196]. Finally, dynamic metabolic modelling has been used for the optimisation of carnitine biosynthesis by *E. coli* [27].

In addition, to gain insight into the microbial metabolism, metabolomic, transcriptomic, and genomic data are typically gathered. To interpret and to extract information from the vast amount of metabolomic, transcriptomic, and genomic data, multivariate statistics, principal component analysis (PCA) and partial least squares regression (PLS) [219, 220], are more and more used in the field of metabolism related studies [39, 84, 105, 197], since these methods can handle numerous and highly correlated data. Also the elementary flux mode (see Chapter 2) data, gathered during stoichiometric network analysis can be analysed with these techniques.

Hence, the aim of this study was to develop a model-based approach for directing metabolic engineering, of which the application should result in speeding up and reducing the cost of the development of a viable industrial biotechnological process. This approach uses partial least squares regression to analyse elementary flux mode data, which are hard to interpret physiologically, and it allows to rapidly identify potential targets for metabolic engineering. This approach was illustrated by applying it to optimise succinate biosynthesis by *E. coli*.

3.2 Materials and methods

3.2.1 Metabolic model

The metabolic network model of Lequeux *et al.* (2006) [100] was used in this study. This metabolic model considers the glycolysis, with glucose transport by the PTS system, the pentose phosphate pathway, the Krebs cycle, and overflow metabolism. For each amino acid and nucleotide the anabolic reactions were included. Biosynthesis of LPS, lipid A, peptidoglycane, and the lipid bilayer are incorporated as well. The P/O ratio was set to 1.33 [108, 203]. The reactions and metabolites considered in the model are depicted in Figures 3.1 and 3.2, respectively.

3.2.2 Partial least squares

Partial least squares (PLS) regression has been performed in the software package R [140]. For a concise description of PLS the reader is referred to Chapter 2. This generalisation of multiple linear regression is able to analyse data with strongly collinear and numerous independent variables as is the case for the elementary flux modes under study. Partial least squares regression is a statistical method that links a matrix of independent variables X with a matrix of dependent variables Y . Therefore, the multivariate spaces of X and Y are transformed to new matrices of lower dimensionality that are correlated to each other. This reduction of dimensionality is accomplished by principal component analysis like decompositions that are slightly tilted to achieve maximum correlation between the latent variables of X and Y [219].

3.2.3 Elementary flux modes

The elementary flux modes of the stoichiometric *E. coli* model of Lequeux *et al.* (2006) [100] were calculated by using Metatool 5.0 [214]. For a concise description of elementary flux modes the reader is referred to Chapter 2.

3.3 Results and discussion

In this work, PLS regression was used to i) analyse the results of the elementary mode analysis and ii) to establish a relationship between the ratio of the flux through a reac-

tion to the glucose influx of an EFM and its succinate yield. To this end, for each of the 17528 EFMs of the *E. coli* model of Lequeux *et al.* (2006) [100] (Figure 3.3) this flux ratio was encoded in the matrix X (Table 3.2). The corresponding Y -variable is the succinate yield of that EFM.

Table 3.2: Construction of the matrix X , with $\xi_{i,j}$ the ratio of the flux through reaction i in EFM j to the glucose influx in EFM j

| Reaction | <i>PTS</i> | <i>PGM</i> | <i>ENO</i> | ... |
|-------------------------|---------------|---------------|---------------|-----|
| <i>EFM</i> ₁ | $\xi_{1,PTS}$ | $\xi_{1,PGM}$ | $\xi_{1,ENO}$ | |
| <i>EFM</i> ₂ | $\xi_{2,PTS}$ | $\xi_{2,PGM}$ | $\xi_{1,ENO}$ | |
| <i>EFM</i> ₃ | $\xi_{3,PTS}$ | $\xi_{3,PGM}$ | $\xi_{1,ENO}$ | |
| <i>EFM</i> ₄ | $\xi_{4,PTS}$ | $\xi_{4,PGM}$ | $\xi_{1,ENO}$ | |

Prior to data analysis, the data were appropriately pretreated. Several pretreatment methods, *i.e.*, mean centering ($x - \mu_x$) and auto-scaling $\frac{(x - \mu_x)}{\sigma_x}$, have been used [193]. Auto-scaling was finally retained as pretreatment method (Figure 3.4), since it relates best the differences in flux ratio's with succinate yield.

A PLS model was then built. First, to avoid overfitting, as this would result in a model not able to generalise to new data, cross-validation was applied to determine the appropriate number of latent variables. In cross-validation the data are split into k blocks and a one latent variable model is built from $(k-1)$ blocks of data. Based on this model, the excluded block is used for testing and an individual predictive residual error sum of squares, PRESS, is calculated. This procedure is repeated excluding each block once, and the total PRESS is calculated for the model. This procedure is then repeated for 2, 3, ..., $\min(m,n)$ latent variables, with n the sample size and m the number of variables. A series of PRESS values is obtained [102]. Wold's R criterion, given as $R = \frac{PRESS(i+1)}{PRESS(i)} \leq 1.1$, is then applied to determine the number of latent variables to be used in the final model. An additional latent variable is retained only when R is smaller than 1.1 [218]. Using this procedure 9 latent variables were retained in the PLS model.

The results, loadings and scores, are depicted in Figures 3.4-3.5. Both regression coefficients and loading weights have been used to study the system. By studying the loading weights (Figure 3.5), one can see how important the variable is in each latent variable. A large positive or negative weight value indicates that the corresponding X

variable is highly correlated with the values in the score matrix U and hence with matrix Y [179, 219]. As cellular metabolism is strongly interlinked [128, 193] it is obvious that many reactions are simultaneously affected by the different elementary flux modes. Therefore, the loadings are expected to show contributions of many different reactions (Figure 3.5).

Some of the most important reactions for succinate production are listed in Table 3.3. The targets identified by the PLS model for the genetic modification of *Escherichia coli* for succinate overproduction are in agreement with data reported in the literature [34, 81, 106, 116, 151]. This illustrates the value of this model-based approach for the identification of genetic targets. Modification of the expression of the identified genetic targets, by overexpressing or knocking out the identified genes, resulted in an enhanced production of succinate.

The proposed method yields many targets for modification. The flux through these reactions is linked in a positive or a negative way with succinate production, through mass and energy conservation laws. Further evaluation of the identified targets using, for instance, flux balance analysis and/or prior knowledge of microbial metabolism, will be useful to determine their importance under a specific condition, *e.g.*, aerobic versus anaerobic environments.

The proposed method helps to significantly reduce the computational effort to optimise microbial metabolism. For instance, since the number of possible combinations of 5 reaction-deletions in a model of 250 reactions is more than $7.8 \cdot 10^9$, and existing genome scale stoichiometric models contain a significantly higher number of reactions, genetic algorithms were applied to search for beneficial knock-out combinations [137]. A first screening of the reactions in a metabolic network, by the proposed method, is useful to render such optimisation problems more feasible.

Contrary to many other methods that are typically focussing on the identification of gene knock-out targets [5, 137], the proposed method yields the correlation (negative and positive) of the flux through each reaction with the yield of the target compound. The question of how to achieve this increased flux remains however unanswered, since the flux distribution depends considerably on the kinetics and the regulation of the enzymatic reactions, which are not accounted for [5, 171]. Still, since succinate is a primary metabo-

Table 3.3: Some of the most important reactions, identified by PLS

| Reaction | Reaction | Coefficient sign |
|----------|---|------------------|
| 155 | $FAD + Suc \longleftrightarrow FADH_2 + Fum$ | - |
| 55 | $iCit \longrightarrow Suc + Glyox$ | + |
| 105 | <i>N.N.</i> | + |
| 71 | $PEP + CO_2 + H_2O \longrightarrow OAA + PiOH$ | + |
| 87 | $Mal + NAD \longrightarrow Pyr + CO_2 + NADH + H$ | - |
| 128 | $Fum + H_2O \longrightarrow Mal$ | - |
| 136 | $Pyr + NADH + H \longrightarrow Lac + NAD$ | - |
| 127 | $FADH_2 + NAD \longrightarrow FAD + NADH + H$ | - |
| 86 | $PEP + ADP \longrightarrow Pyr + ATP$ | - |
| 126 | $AcCoA + 2NADH + H \longrightarrow EtOH + 2NAD + CoA$ | - |

lite, whose production is subject to severe (redox) constraints, stoichiometric modelling is useful.

3.4 Conclusions

A model-based genetic target identification strategy for designing a microbial strain for the production of a target compound, has been outlined. By applying partial least squares regression to the elementary flux mode data, potential targets for metabolic engineering of succinate biosynthesis in *E. coli* were identified. The targets identified by the PLS model for genetic modification of *E. coli* for succinate overproduction are in agreement with data reported in the literature.

| | | | |
|---------------------------------------|---|--------------------------------------|--|
| <i>HK:</i> | ATP + GLC \rightarrow ADP + G6P | <i>ThrSYLR:</i> | ATP + H ₂ O + HSer \rightarrow ADP + PiOH + Thr |
| <i>PGI:</i> | G6P \leftrightarrow F6P | <i>MDAPSYLR:</i> | NADPH + H + Pyr + SucCoA + Glu + AspSA \rightarrow NADP + CoA + aKGA + Suc + MDAP |
| <i>PFK:</i> | ATP + F6P \rightarrow ADP + FBP | | |
| <i>ALD:</i> | FBP \leftrightarrow G3P + DHAP | <i>LysSY:</i> | MDAP \rightarrow CO ₂ + Lys |
| <i>TPI:</i> | DHAP \leftrightarrow G3P | <i>MetSYLR:</i> | H ₂ O + SucCoA + Cys + MTHF + HSer \rightarrow Pyr + CoA + Suc + NH ₃ + Met + THF |
| <i>G3PDH:</i> | PiOH + NAD + G3P \leftrightarrow NADH + H + BPG | <i>AspSASY:</i> | ATP + NADPH + H + Asp \rightarrow ADP + PiOH + NADP + AspSA |
| <i>PGK:</i> | ADP + BPG \leftrightarrow ATP + 3PG | | |
| <i>PGM:</i> | 3PG \leftrightarrow 2PG | <i>HSerDH:</i> | NADPH + H + AspSA \leftrightarrow NADP + HSer |
| <i>ENO:</i> | 2PG \leftrightarrow H ₂ O + PEP | <i>CarPSY:</i> | 2 ATP + H ₂ O + H ₂ CO ₃ + Gln \rightarrow 2 ADP + PiOH + Glu + CarP |
| <i>PyrK:</i> | ADP + PEP \rightarrow ATP + Pyr | | |
| <i>PyrD:</i> | NAD + Pyr + CoA \rightarrow NADH + H + AcCoA + CO ₂ | <i>OrnSYLR:</i> | ATP + NADPH + H + H ₂ O + AcCoA + 2Glu \rightarrow ADP + PiOH + NADP + CoA + aKGA + Orn + Ac |
| <i>CisSY:</i> | H ₂ O + AcCoA + OAA \rightarrow CoA + Cit | | |
| <i>ACO:</i> | Cit \leftrightarrow iCit | <i>ArgSYLR:</i> | ATP + Asp + Orn + CarP \rightarrow AMP + PPiOH + PiOH + Fum + Arg |
| <i>CaDH:</i> | NAD + iCit \leftrightarrow NADH + H + CO ₂ + aKGA | | |
| <i>AKGDH:</i> | NAD + CoA + aKGA \rightarrow NADH + H + CO ₂ + SucCoA | <i>ThioRedD:</i> | NADPH + ThioRed + H \leftrightarrow NADP + ThioRedH ₂ |
| <i>SucCoASY:</i> | ADP + PiOH + SucCoA \leftrightarrow ATP + CoA + Suc | <i>H₂O₂ox:</i> | 2H ₂ O ₂ \rightarrow 2H ₂ O + O ₂ |
| <i>SucDH:</i> | FAD + Suc \rightarrow FADH ₂ + Fum | <i>FAD₂NAD:</i> | NAD + FADH ₂ \leftrightarrow NADH + FAD + H |
| <i>FumHY:</i> | H ₂ O + Fum \rightarrow Mal | <i>CoQ₂NAD:</i> | NADH + CoQ + H \leftrightarrow NAD + CoQH ₂ |
| <i>MalDH:</i> | NAD + Mal \leftrightarrow NADH + H + OAA | <i>NADH₂NADPH:</i> | NADH + NADP \leftrightarrow NAD + NADPH |
| <i>iCitL:</i> | iCit \rightarrow Suc + Glyox | <i>AICARSYLR:</i> | 6 ATP + 3H ₂ O + CO ₂ + Asp + 2Gln + Gly + FA + PRPP \rightarrow 6 ADP + PPiOH + 6PiOH + Fum + 2Glu + AICAR |
| <i>MalSY:</i> | H ₂ O + AcCoA + Glyox \rightarrow CoA + Mal | | |
| <i>PEPCB:</i> | H ₂ O + PEP + CO ₂ \rightarrow PiOH + OAA | <i>IMPSYLR:</i> | FTHF + AICAR \rightarrow H ₂ O + THF + IMP |
| <i>PEPCBKN:</i> | ATP + OAA \rightarrow ADP + PEP + CO ₂ | <i>AMP₂SYLR:</i> | Asp + GTP + IMP \rightarrow AMP + PiOH + Fum + GDP |
| <i>PyrMalCB:</i> | NAD + Mal \rightarrow NADH + H + Pyr + CO ₂ | <i>AdKN:</i> | ATP + AMP \leftrightarrow 2ADP |
| <i>LacDH:</i> | NADH + H + Pyr \leftrightarrow NAD + Lac | <i>ADPRD:</i> | ADP + ThioRedH ₂ \rightarrow ThioRed + H ₂ O + dADP |
| <i>PFLY:</i> | Pyr + CoA \rightarrow AcCoA + FA | <i>dADPKN:</i> | ATP + dADP \rightarrow ADP + dATP |
| <i>EhdDHLR:</i> | 2NADH + 2H + AcCoA \leftrightarrow 2NAD + CoA + Eth | <i>dADPPT:</i> | H ₂ O + dADP \rightarrow PiOH + dAMP |
| <i>AcKNLR:</i> | ADP + PiOH + AcCoA \leftrightarrow ATP + CoA + Ac | <i>IMPDH:</i> | NAD + H ₂ O + IMP \rightarrow NADH + H + XMP |
| <i>ActSY:</i> | Pyr + Acdh \rightarrow CO ₂ + Act | <i>GMP₂SY:</i> | ATP + H ₂ O + Gln + XMP \rightarrow AMP + PPiOH + Glu + GMP |
| <i>AcidDH:</i> | NADH + H + AcCoA \leftrightarrow NAD + CoA + Acdh | <i>GuKN:</i> | ATP + GMP \rightarrow ADP + GDP |
| <i>EhdDH:</i> | NADH + H + Acdh \leftrightarrow NAD + Eth | <i>GDPKN:</i> | ATP + GDP \rightarrow ADP + GTP |
| <i>Resp:</i> | 1.33 ADP + 1.33 PiOH + NADH + H + 0.5 O ₂ \rightarrow 1.33 ATP + NAD + 2.33 H ₂ O | <i>GDP₂RD:</i> | ThioRedH ₂ + GDP \rightarrow ThioRed + H ₂ O + dGDP |
| <i>H₂CO₃SY:</i> | H ₂ O + CO ₂ \leftrightarrow H ₂ CO ₃ | <i>dGDPKN:</i> | ATP + dGDP \rightarrow ADP + dGTP |
| <i>G6PDH:</i> | NADP + G6P \rightarrow NADPH + H + 6PGL | <i>dGDPPT:</i> | H ₂ O + dGDP \rightarrow PiOH + dGMP |
| <i>LAS:</i> | H ₂ O + 6PGL \rightarrow 6PG | <i>UMPSYLR:</i> | O ₂ + Asp + PRPP + CarP \rightarrow PPiOH + PiOH + H ₂ O + CO ₂ + UMP + H ₂ O ₂ |
| <i>PGDH:</i> | NADP + 6PG \rightarrow NADPH + H + CO ₂ + R5P | | |
| <i>RPI:</i> | R5P \leftrightarrow R5P | <i>UtrKN:</i> | ATP + UMP \rightarrow ADP + UDP |
| <i>PPE:</i> | R5P \leftrightarrow Xu5P | <i>UDPKN:</i> | ATP + UDP \rightarrow ADP + UTP |
| <i>TK1:</i> | R5P + Xu5P \rightarrow G3P + S7P | <i>CTPSY:</i> | ATP + H ₂ O + Gln + UTP \rightarrow ADP + PiOH + Glu + CTP |
| <i>TA:</i> | G3P + S7P \leftrightarrow F6P + E4P | <i>CDPKN:</i> | ATP + CDP \leftrightarrow ADP + CTP |
| <i>TK2:</i> | Xu5P + E4P \leftrightarrow F6P + G3P | <i>CDPPT:</i> | H ₂ O + CDP \rightarrow PiOH + CMP |
| <i>FBPAS:</i> | H ₂ O + FBP \rightarrow PiOH + F6P | <i>CMPKN:</i> | ATP + CMP \rightarrow ADP + CDP |
| <i>R5P2R1P:</i> | R5P \leftrightarrow R1P | <i>CDPRD:</i> | ThioRedH ₂ + CDP \rightarrow ThioRed + H ₂ O + dCDP |
| <i>PTS:</i> | GLC + PEP \rightarrow G6P + Pyr | <i>dCDPKN:</i> | ATP + dCDP \rightarrow ADP + dCTP |
| <i>PPiOH₂SY:</i> | PPiOH + H ₂ O \rightarrow 2PiOH | <i>dCDPPT:</i> | H ₂ O + dCDP \rightarrow PiOH + dCMP |
| <i>GluDH:</i> | NADPH + H + aKGA + NH ₃ \leftrightarrow NADP + H ₂ O + Glu | <i>dCTPDA:</i> | H ₂ O + dCTP \rightarrow NH ₃ + dUTP |
| <i>GluL:</i> | ATP + NH ₃ + Glu \rightarrow ADP + PiOH + Glu | <i>UDPRD:</i> | ThioRedH ₂ + UDP \rightarrow ThioRed + H ₂ O + dUDP |
| <i>GluSY:</i> | NADPH + H + aKGA + Glu \rightarrow NADP + 2Glu | <i>dUDPKN:</i> | ATP + dUDP \rightarrow ADP + dUTP |
| <i>AspSY:</i> | ATP + H ₂ O + Asp + Gln \rightarrow AMP + PPiOH + Asn + Glu | <i>dUTPPAS:</i> | H ₂ O + dUTP \rightarrow PPiOH + dUMP |
| <i>AspTA:</i> | OAA + Glu \leftrightarrow aKGA + Asp | <i>dTMP₂SY:</i> | MeTHF + dUMP \rightarrow DHF + dTMP |
| <i>AspL:</i> | ATP + NH ₃ + Asp \rightarrow AMP + PPiOH + Asn | <i>dTMPKN:</i> | ATP + dTMP \rightarrow ADP + dTDP |
| <i>AlaTA:</i> | Pyr + Glu \leftrightarrow aKGA + Ala | <i>dTDPKN:</i> | ATP + dTDP \rightarrow ADP + dTTP |
| <i>ValPyrAT:</i> | Pyr + Val \leftrightarrow aKIV + Ala | <i>dTDPPT:</i> | H ₂ O + dTDP \rightarrow PiOH + dTMP |
| <i>ValAT:</i> | aKIV + Glu \leftrightarrow aKGA + Val | <i>DHF₂RD:</i> | NADPH + H + DHF \rightarrow NADP + THF |
| <i>LeuSYLR:</i> | NAD + H ₂ O + AcCoA + aKIV + Glu \rightarrow NADH + H + CoA + CO ₂ + aKGA + Leu | <i>FTHFSYLR:</i> | NADP + H ₂ O + MeTHF \rightarrow NADPH + H + FTHF |
| <i>aKIVSYLR:</i> | NADPH + H + 2Pyr \rightarrow NADP + H ₂ O + CO ₂ + aKIV | <i>GlyCA:</i> | NAD + Gly + THF \leftrightarrow NADH + H + CO ₂ + NH ₃ + MeTHF |
| <i>leSYLR:</i> | NADPH + H + Pyr + Glu + Thr \rightarrow NADP + H ₂ O + CO ₂ + aKGA + NH ₃ + Ile | <i>AcCoA₂FRD:</i> | NADH + H + MeTHF \rightarrow NAD + MTHF |
| <i>ProSYLR:</i> | ATP + 2NADPH + 2H + Glu \rightarrow ADP + PiOH + 2NADP + H ₂ O + Pro | <i>FTHFDF:</i> | H ₂ O + FTHF \rightarrow THF + FA |
| <i>SerLR:</i> | NAD + H ₂ O + 3PG + Glu \rightarrow PiOH + NADH + H + aKGA + Ser | <i>AcCoACB:</i> | ATP + H ₂ O + AcCoA + CO ₂ \leftrightarrow ADP + PiOH + MalCoA |
| <i>SerTHM:</i> | Ser + THF \rightarrow H ₂ O + Gly + MeTHF | <i>MalCoATA:</i> | MalCoA + ACP \leftrightarrow CoA + MalACP |
| <i>H₂SSYLR:</i> | 2 ATP + 3NADPH + ThioRedH ₂ + 3H + H ₂ SO ₄ \rightarrow ADP + PPiOH + 3NADP + ThioRed + 3H ₂ O + H ₂ S + PAP | <i>AcACPSY:</i> | MalACP \rightarrow CO ₂ + AcACP |
| <i>PAPNAS:</i> | H ₂ O + PAP \rightarrow AMP + PiOH | <i>AcCoATA:</i> | CoA + AcACP \leftrightarrow AcCoA + ACP |
| <i>CysSYLR:</i> | H ₂ S + AcCoA + Ser \rightarrow CoA + Cys + Ac | <i>C120SY:</i> | 10NADPH + 10H + AcACP + 5MalACP \rightarrow 10NADP + 5H ₂ O + 5CO ₂ + C120ACP + 5ACP |
| <i>PrppSY:</i> | ATP + R5P \rightarrow AMP + PRPP | <i>C140SY:</i> | 12NADPH + 12H + AcACP + 6MalACP \rightarrow 12NADP + 6H ₂ O + 6CO ₂ + C140ACP + 6ACP |
| <i>HisSYLR:</i> | ATP + 2NAD + 3H ₂ O + Gln + PRPP \rightarrow 2PPiOH + PiOH + 2NADH + 2H + aKGA + His + AICAR | <i>C141SY:</i> | 11NADPH + 11H + AcACP + 6MalACP \rightarrow 11NADP + 6H ₂ O + 6CO ₂ + C141ACP + 6ACP |
| <i>PheSYLR:</i> | Glu + Chor \rightarrow H ₂ O + CO ₂ + aKGA + Phe | <i>C160SY:</i> | 14NADPH + 14H + AcACP + 7MalACP \rightarrow 14NADP + 7H ₂ O + 7CO ₂ + C160ACP + 7ACP |
| <i>TyrSYLR:</i> | NAD + Glu + Chor \rightarrow NADH + H + CO ₂ + aKGA + Tyr | <i>C161SY:</i> | 13NADPH + 13H + AcACP + 7MalACP \rightarrow 13NADP + 7H ₂ O + 7CO ₂ + C161ACP + 7ACP |
| <i>TrpSYLR:</i> | Gln + Ser + Chor + PRPP \rightarrow PPiOH + 2H ₂ O + G3P + Pyr + CO ₂ + Glu + Trp | <i>C181SY:</i> | 15NADPH + 15H + AcACP + 8MalACP \rightarrow 15NADP + 8H ₂ O + 8CO ₂ + C181ACP + 8ACP |
| <i>DhDoHPepAD:</i> | H ₂ O + PEP + E4P \rightarrow PiOH + Dalp | | |
| <i>DhqSY:</i> | Dalp \rightarrow PiOH + Dhq | <i>AcylTF:</i> | C160ACP + C181ACP + G63P \rightarrow 2ACP + PA |
| <i>DhsSYLR:</i> | Dhq \leftrightarrow H ₂ O + Dis | <i>Go3PDH:</i> | NADPH + H + DHAP \leftrightarrow NADP + G63P |
| <i>ShiSY:</i> | NADPH + H + Dhs \leftrightarrow NADP + Shi | <i>DGoKN:</i> | ATP + DGo \rightarrow ADP + PA |
| <i>ShiKN:</i> | ATP + Shi \rightarrow ADP + Shi3P | <i>CDPDGoSY:</i> | CTP + PA \leftrightarrow PPiOH + CDPDGo |
| <i>DhqDH:</i> | NADPH + H + Dhq \rightarrow NADP + Qa | <i>PSerSY:</i> | Ser + CDPDGo \rightarrow CMP + PSer |
| <i>ChorSYLR:</i> | PEP + Shi3P \rightarrow 2PiOH + Chor | <i>PSerDC:</i> | PSer \rightarrow CO ₂ + PEthAn |
| <i>DhsDH:</i> | Dhs \rightarrow H ₂ O + ProtoCat | <i>GlnF6P₂TA:</i> | F6P + Gln \rightarrow Glu + GA6P |
| <i>ProtoCatDC:</i> | ProtoCat \rightarrow CO ₂ + Cat | <i>GlcAnMU:</i> | GA6P \leftrightarrow GA1P |
| <i>BkaSYLR:</i> | H ₂ O + O ₂ + Cat \rightarrow Bka | <i>NAGUrTF:</i> | AcCoA + UTP + GA1P \rightarrow PPiOH + CoA + UDPNAG |
| <i>GallicSY:</i> | NAD + Dhs \rightarrow NADH + H + Gallic | <i>LipaSYLR:</i> | ATP + 2CMPKDO + 2UDPNAG + C120ACP + 5C140ACP \rightarrow ADP + 2CMP + UMP + UDP + 6ACP + Lipa + 2Ac |

Figure 3.1: Metabolic network of Lequeux *et al.* (2006) [100]: Reactions

| | | | | | |
|----------|---|--|-----------|---|---|
| 2PG | C ₃ H ₇ O ₇ P | 2-phosphoglycerate | GA6P | C ₆ H ₁₄ O ₈ NP | D-glucosamine-6-phosphate |
| 3PG | C ₃ H ₇ O ₇ P | 3-phosphoglycerate | Gallic | C ₇ H ₆ O ₅ | Gallic acid |
| 6PG | C ₆ H ₁₃ O ₁₀ P | 6-phosphogluconate | GDP | C ₁₀ H ₁₅ O ₁₁ N ₅ P ₂ | Guanosine diphosphate |
| 6PGL | C ₆ H ₁₁ O ₈ P | 6-phosphogluconolacton | GLC | C ₆ H ₁₂ O ₆ | Glucose |
| Ac | C ₂ H ₄ O ₂ | Acetate | Gleg | C ₆ H ₁₀ O ₅ | Glycogen |
| AcACP | C ₂ H ₃ O ₂ Pept | Acetyl ACP | Gln | C ₅ H ₁₀ O ₂ N ₂ | Glutamine |
| AcCoA | C ₂₃ H ₃₁ O ₁₇ N ₇ P ₃ S | Acetyl CoA | Glu | C ₅ H ₉ O ₄ N | Glutamate |
| AcOh | C ₂ H ₄ O | Acetaldehyde | Gly | C ₂ H ₅ O ₂ N | Glycine |
| ACP | HPept | Acyl carrier protein | Glyox | C ₂ H ₂ O ₃ | Glyoxylate |
| Act | C ₄ H ₈ O ₂ | Acetoine | GMP | C ₁₀ H ₁₄ O ₈ N ₅ P | Guanosine monophosphate |
| ADP | C ₁₀ H ₁₅ O ₁₀ N ₅ P ₂ | Adenosine diphosphate | Go3P | C ₃ H ₉ O ₆ P | Glycerol-3-phosphate |
| ADPHEP | C ₁₇ H ₂₇ O ₁₆ N ₅ P ₃ | ADP-Mannoseheptose | GTP | C ₁₀ H ₁₆ O ₁₄ N ₅ P ₃ | Guanosine triphosphate |
| AICAR | C ₉ H ₁₅ O ₈ N ₄ P | Amino imidazole carboxamide ribonucleotide | H | H ⁺ | Hydrogene |
| aKGA | C ₅ H ₆ O ₅ | Alpha keto glutaric acid | H2CO3 | CH ₂ O ₃ | Bicarbonate |
| aKIV | C ₅ H ₆ O ₃ | Alpha-keto-isovalerate | H2O | H ₂ O | Water |
| Ala | C ₃ H ₇ O ₂ N | Alanine | H2O2 | H ₂ O ₂ | |
| AMP | C ₁₀ H ₁₄ O ₇ N ₅ P | Adenosine monophosphate | H2S | H ₂ S | Hydrogene sulfide |
| Ar5P | C ₅ H ₁₁ O ₈ P | Arabinose-5-phosphate | H2SO4 | H ₂ O ₄ S | Sulfuric acid |
| Arg | C ₆ H ₁₄ O ₂ N ₄ | Arginine | His | C ₆ H ₉ O ₂ N ₃ | Histidine |
| Asn | C ₄ H ₈ O ₃ N ₂ | Aspartate | HSer | C ₄ H ₉ O ₃ N | Homoserine |
| Asp | C ₄ H ₇ O ₄ N | Asparagine | iCit | C ₆ H ₈ O ₇ | isocitrat |
| AspSA | C ₄ H ₇ O ₃ N | Aspartate semialdehyde | Ile | C ₆ H ₁₃ O ₂ N | Isoleucine |
| ATP | C ₁₀ H ₁₆ O ₁₃ N ₅ P ₃ | Adenosine triphosphate | IMP | C ₁₀ H ₁₃ O ₈ N ₄ P | Inosine monophosphate |
| BGalAse | C _{4,98} H _{7,58} O _{1,5} N _{1,41} S _{0,607} | Beta-galactosidase | Lac | C ₃ H ₆ O ₃ | Lactate |
| Biom | CH _{1,63} O _{0,392} N _{0,214} P _{0,021} S _{0,055} | Biomass | Leu | C ₆ H ₁₃ O ₂ N | Leucine |
| Bka | C ₃ H ₄ O ₅ | Beta ketoacidipate | Lipa | C ₁₁₀ H ₁₉₆ O ₂₈₂ N ₂ P ₂ | Lipid A |
| BPG | C ₃ H ₅ O ₁₀ P ₂ | 1,3-biphosphoglycerate | Lipid | C _{40,2} H _{77,6} O _{8,41} N _{0,771} P _{1,03} | Lipid composition |
| C120ACP | C ₁₂ H ₂₃ O ₁₆ Pept | | Lps | C ₁₇₁ H ₂₈₈ O ₆₁ N ₄ P ₂ | Lipo Poly saccharide |
| C140ACP | C ₁₄ H ₂₇ O ₁₆ Pept | | Lys | C ₆ H ₁₄ O ₂ N ₂ | Lysine |
| C141ACP | C ₁₄ H ₂₅ O ₁₆ Pept | | Mal | C ₄ H ₆ O ₅ | Malate |
| C160ACP | C ₁₆ H ₃₁ O ₁₆ Pept | | MalACP | C ₃ H ₃ O ₃ Pept | Malonyl ACP |
| C161ACP | C ₁₆ H ₂₉ O ₁₆ Pept | | MalCoA | C ₂₄ H ₃₄ O ₁₆ N ₇ P ₃ S | Malonyl CoA |
| C181ACP | C ₁₈ H ₃₃ O ₁₆ Pept | | MDAP | C ₇ H ₁₄ O ₄ N ₂ | Meso-diaminopimelate |
| CarP | CH ₃ O ₅ NP | Carbamoyl phosphate | Met | C ₅ H ₁₁ O ₂ NS | Methionine |
| Cat | C ₆ H ₆ O ₂ | Catechol | MeTHF | C ₂₀ H ₂₃ O ₆ N ₇ | Methyltetrahydrofolate |
| CDP | C ₉ H ₁₅ O ₁₁ N ₃ P ₂ | Citidine diphosphate | MTHF | C ₂₀ H ₂₃ O ₆ N ₇ | Methyl tetrahydrofolate |
| CDPDGo | C ₄₆ H ₈₃ O ₁₅ N ₃ P ₂ | CDP-diacylglycerol | NAD | C ₂₁ H ₂₈ O ₁₄ N ₇ P ₂ ⁺ | Nicotinamide adenine dinucleotide |
| CDPEthAn | C ₁₁ H ₂₁ O ₁₁ N ₄ P ₂ | CDP-ethanolamine | NADH | C ₂₁ H ₂₉ O ₁₄ N ₇ P ₂ ⁺ | |
| Chor | C ₁₀ H ₁₀ O ₆ | Chorismate | NADP | C ₂₁ H ₂₈ O ₁₇ N ₇ P ₃ ⁺ | Nicotinamide adenine dinucleotide phosphate |
| Cit | C ₆ H ₈ O ₇ | cisaconitate | NADPH | C ₂₁ H ₂₉ O ₁₇ N ₇ P ₃ | |
| CL | C ₇₇ H ₁₄₄ O ₁₆ P ₂ | Cardiolipin | NH3 | H ₃ N | Ammonia |
| CMP | C ₉ H ₁₄ O ₈ N ₃ P | Citidine monophosphate | O2 | O ₂ | Oxygen |
| CMPKDO | C ₁₇ H ₂₆ O ₁₅ N ₃ P | CMP-2-keto-3-deoxyoctanoate | OAA | C ₄ H ₄ O ₅ | Oxaloacetate |
| CO2 | CO ₂ | Carbon dioxide | Om | C ₅ H ₁₂ O ₂ N ₂ | Omithine |
| CoA | C ₂₁ H ₃₂ O ₁₆ N ₇ P ₃ S | Coenzyme A | PA | C ₃₇ H ₇₁ O ₈ P | Phosphatidyl acid |
| CoQ | C ₁₄ H ₁₈ O ₄ | Coenzyme Q. Ubiquinone (C5H8)n omitted | PAP | C ₁₀ H ₁₅ O ₁₀ N ₅ P ₂ | Phospho adenosine phosphate |
| CoQH2 | C ₁₄ H ₂₀ O ₄ | Ubiquinol | PEP | C ₃ H ₅ O ₆ P | Phosphoenolpyruvate |
| CTP | C ₉ H ₁₆ O ₁₄ N ₃ P ₃ | Citidine triphosphate | Peptido | C ₃₅ H ₅₃ O ₁₆ N ₇ | Peptidoglycane |
| Cys | C ₃ H ₇ O ₂ NS | Cysteine | PEthAn | C ₃₀ H ₇₆ O ₈ NP | Phosphatidyl ethanolamine |
| dADP | C ₁₀ H ₁₅ O ₉ N ₅ P ₂ | deoxy ADP | PG | C ₄₆ H ₇₅ O ₉ P | Phosphatidyl glycerol |
| Dalhp | C ₇ H ₁₃ O ₁₀ P | Deoxy arabinose heptulosonate | Phe | C ₉ H ₉ O ₂ N | Phenylalanine |
| dAMP | C ₁₀ H ₁₄ O ₈ N ₅ P | deoxy AMP | PROH | H ₃ O ₄ P | Phosphate |
| dATP | C ₁₀ H ₁₆ O ₁₂ N ₅ P ₃ | deoxy ATP | PPHOH | H ₄ O ₇ P ₂ | Pyrophosphate |
| dCDP | C ₉ H ₁₅ O ₁₀ N ₃ P ₂ | deoxy CDP | Pro | C ₅ H ₉ O ₂ N | Proline |
| dCMP | C ₉ H ₁₄ O ₇ N ₃ P | deoxy CMP | Prot | C _{14,8} H _{7,67} O _{1,4} N _{1,37} S _{0,046} | Protein composition |
| dCTP | C ₉ H ₁₆ O ₁₃ N ₃ P ₃ | deoxy CTP | ProtoCat | C ₇ H ₆ O ₄ | Protocatechol |
| dGDP | C ₁₀ H ₁₅ O ₁₀ N ₅ P ₂ | deoxy GDP | PRPP | C ₅ H ₁₃ O ₁₄ P ₃ | 5-phospho-alpha-D-ribose-1-pyrophosphate |
| dGMP | C ₁₀ H ₁₄ O ₇ N ₅ P | deoxy GMP | PSer | C ₄₀ H ₇₆ O ₁₀ NP | Phosphatidyl Serine |
| DGo | C ₃₇ H ₇₀ O ₅ | Diacyl glycerol | Pyr | C ₃ H ₄ O ₃ | Pyruvate |
| dGTP | C ₁₀ H ₁₆ O ₁₃ N ₅ P ₃ | deoxy GTP | Qa | C ₇ H ₁₂ O ₆ | Quinate |
| DHAP | C ₃ H ₇ O ₆ P | Dihydroxyacetone phosphate | RIP | C ₅ H ₁₁ O ₃ P | Ribose-1-phosphate |
| DHF | C ₁₃ H ₂₁ O ₈ N ₇ | Dihydrofolate | R5P | C ₅ H ₁₁ O ₈ P | Ribose-5-phosphate |
| Dhq | C ₇ H ₁₀ O ₆ | Dehydroquininate | R5P | C ₅ H ₁₁ O ₈ P | Ribulose-5-phosphate |
| Dhs | C ₇ H ₈ O ₅ | Dehydroshikimate | RNA | C _{3,23} H _{1,38} O _{7,26} N _{3,36} P | RNA composition |
| DNA | C _{3,75} H _{1,42} O ₇ N _{3,75} P | DNA composition | S7P | C ₇ H ₁₅ O ₁₀ P | Sedoheptulose-7-phosphate |
| dTDP | C ₁₀ H ₁₆ O ₁₁ N ₂ P ₂ | deoxy TDP | Ser | C ₃ H ₇ O ₃ N | Serine |
| dTMP | C ₁₀ H ₁₅ O ₈ N ₂ P | deoxy TMP | Shi | C ₇ H ₁₀ O ₅ | Shikimate |
| dTTP | C ₁₀ H ₁₇ O ₁₄ N ₂ P ₃ | deoxy TTP | Shi3P | C ₇ H ₁₁ O ₈ P | Shikimate-3-phosphate |
| dUDP | C ₉ H ₁₄ O ₁₁ N ₂ P ₂ | deoxy UDP | Suc | C ₄ H ₆ O ₄ | Succinate |
| dUMP | C ₉ H ₁₃ O ₈ N ₂ P | deoxy UMP | SucCoA | C ₂₃ H ₃₆ O ₁₉ N ₇ P ₃ S | Succinyl CoA |
| dUTP | C ₉ H ₁₅ O ₁₄ N ₂ P ₃ | deoxy UTP | THF | C ₁₉ H ₂₃ O ₉ N ₇ | Tetrahydrofolate |
| E4P | C ₄ H ₆ O ₇ P | Erythrose-4-phosphate | Thiorex | Pept | Thioredoxin |
| Eth | C ₂ H ₆ O | Ethanol | ThiorexH2 | H ₂ Pept | Reduced thioredoxin |
| F6P | C ₆ H ₁₂ O ₆ P | Fructose-6-phosphate | Thr | C ₄ H ₉ O ₂ N | Threonine |
| FA | CH ₂ O ₂ | Formic Acid | Trp | C ₁₁ H ₁₂ O ₂ N ₂ | Tryptophan |
| FAD | C ₂₇ H ₃₃ O ₁₅ N ₉ P ₂ | Flavine adenine dinucleotide | Tyr | C ₉ H ₁₁ O ₃ N | Tyrosine |
| FADH2 | C ₂₇ H ₃₅ O ₁₅ N ₉ P ₂ | | UDP | C ₉ H ₁₄ O ₁₂ N ₂ P ₂ | Uridine diphosphate |
| FBP | C ₆ H ₁₄ O ₁₂ P ₂ | Fructose-1,6-biphosphate | UDPGlc | C ₁₅ H ₂₄ O ₁₇ N ₂ P ₂ | UDP glucose |
| FTHF | C ₂₀ H ₂₃ O ₇ N ₇ | Formyl tetrahydrofolate | UDPNAg | C ₁₇ H ₂₇ O ₁₇ N ₃ P ₂ | UDP N-acetyl glucosamine |
| Fum | C ₄ H ₄ O ₄ | Fumarate | UMP | C ₉ H ₁₃ O ₈ N ₂ P | Uridine monophosphate |
| G1P | C ₆ H ₁₂ O ₆ P | Glucose-1-phosphate | UTP | C ₉ H ₁₅ O ₁₅ N ₂ P ₃ | Uridine triphosphate |
| G3P | C ₃ H ₇ O ₆ P | Glyceraldehyde-3-phosphate | Val | C ₅ H ₁₁ O ₂ N | Valine |
| G6P | C ₆ H ₁₂ O ₆ P | Glucose-6-phosphate | XMP | C ₁₀ H ₁₃ O ₉ N ₄ P | Xanthosine-5-phosphate |
| GA1P | C ₆ H ₁₄ O ₈ NP | D-glucosamine-6-phosphate | Xu5P | C ₅ H ₁₁ O ₈ P | Xylulose-5-phosphate |

Figure 3.2: Metabolic network of Lequeux *et al.* (2006) [100]: Metabolites

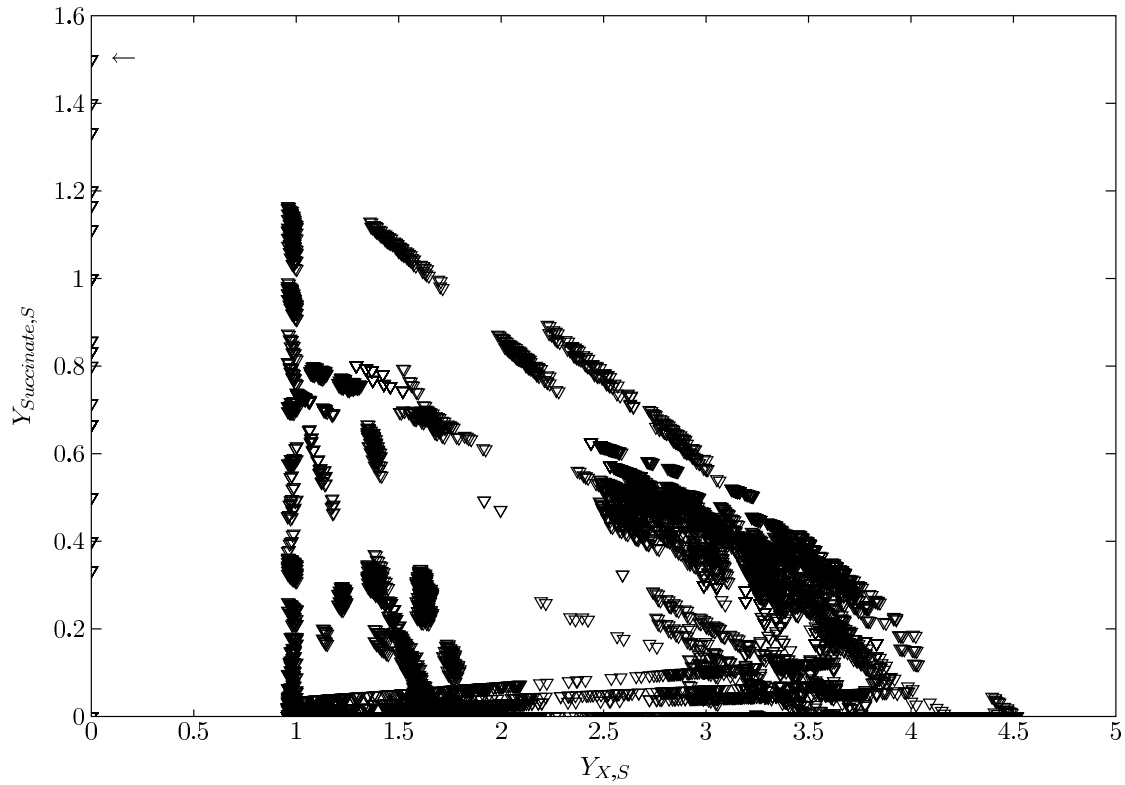


Figure 3.3: The 17528 elementary flux modes of the stoichiometric *E. coli* model of Lequeux *et al.* (2006) [100], represented as ∇ s, calculated by using Metatool 5.0 [214], presented in the $Y_{X,S}$, $Y_{succinate,S}$ space, with $Y_{X,S}$ and $Y_{succinate,S}$ the biomass $\left[\frac{c-mole}{mole}\right]$ and succinate $\left[\frac{mole}{mole}\right]$ yield on glucose, respectively. The EFMs characterised by the optimal flux distribution, here with respect to maximal $Y_{succinate,S}$, can readily be identified (\leftarrow).

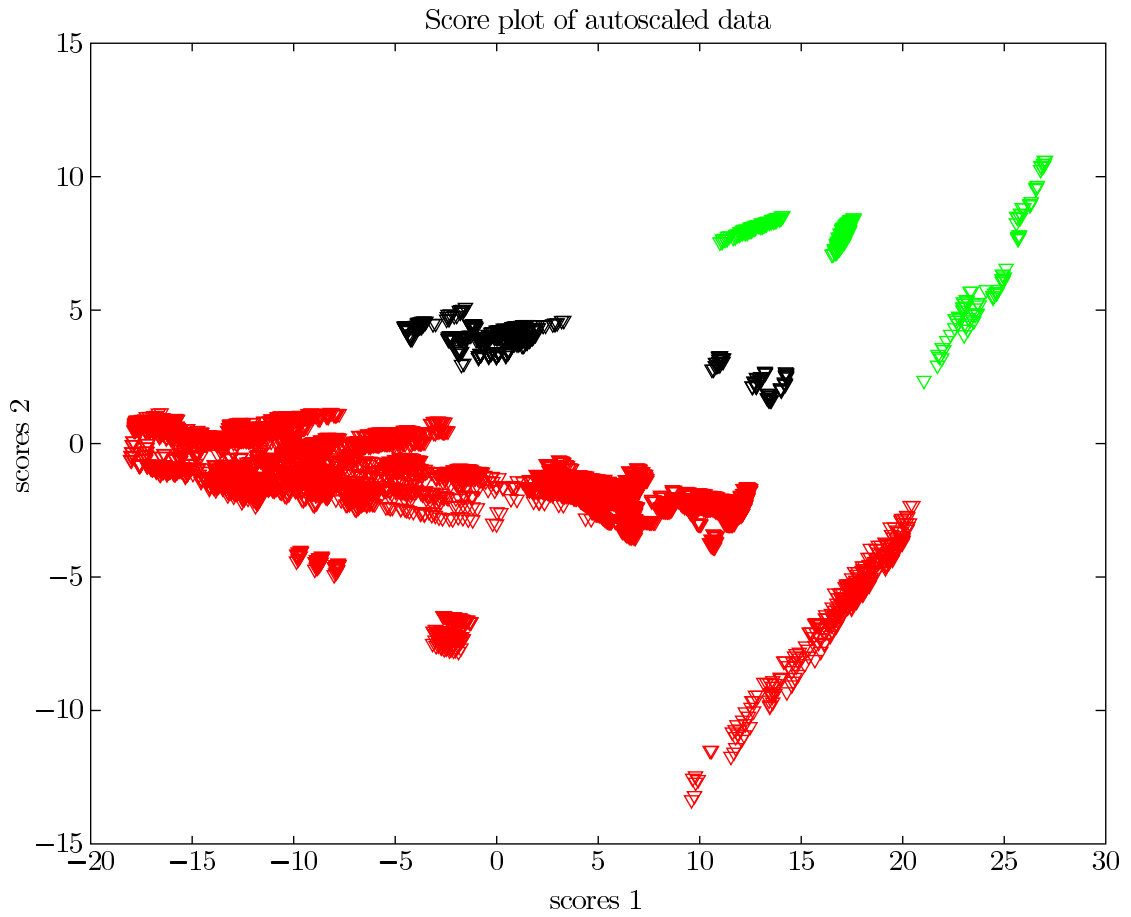


Figure 3.4: The auto-scaled data represented in a score plot. \blacktriangledown represent EFMs characterised by $Y_{succinate,S} \geq 1$, \blacktriangledown represent EFMs characterised by $0.6 \geq Y_{succinate,S} \geq 0.5$, and \blacktriangledown represent EFMs characterised by $Y_{succinate,S} \leq 0.1$.

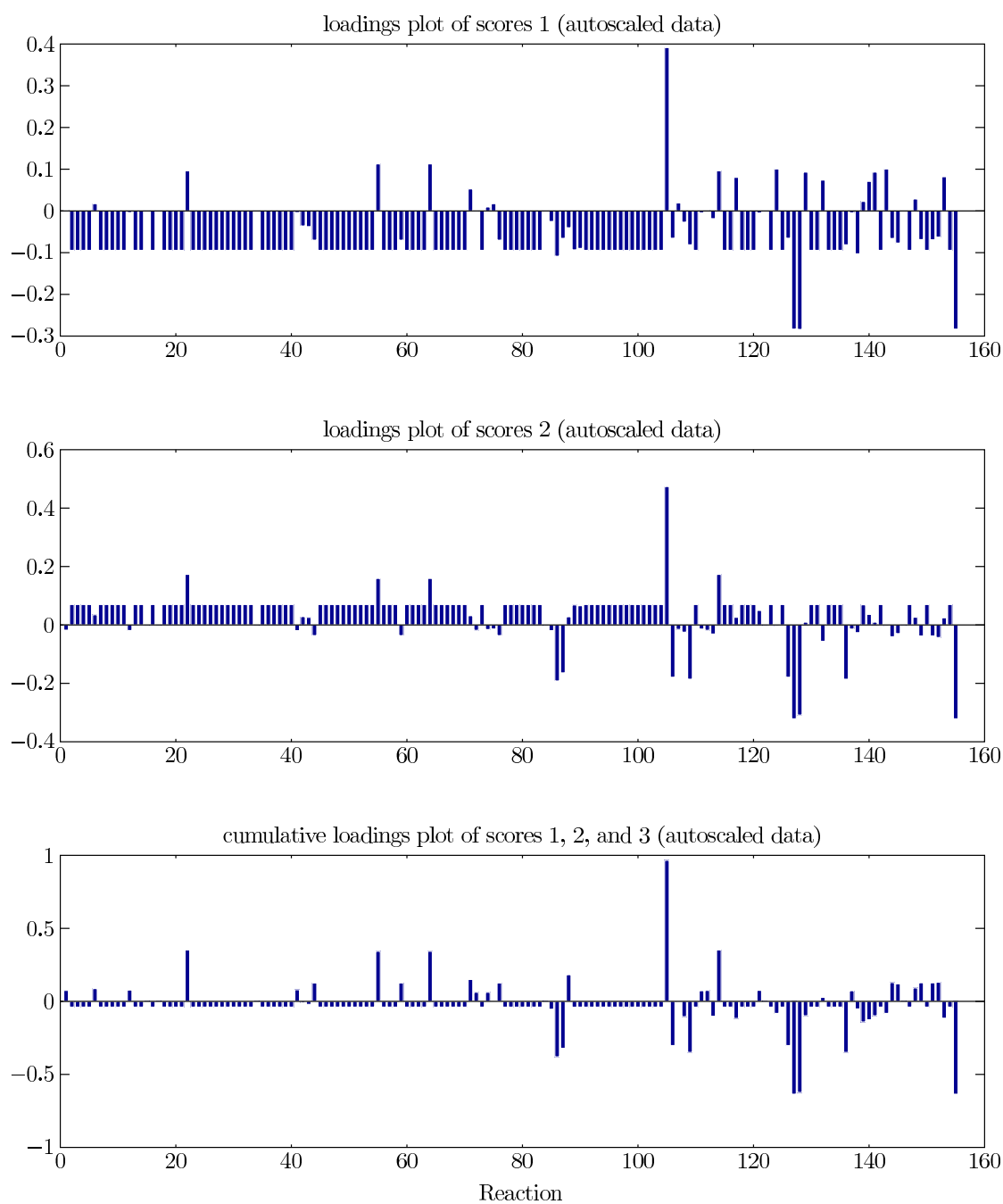


Figure 3.5: The loadings of scores 1 and 2 and the cumulative loadings of scores 1, 2, and 3. This cumulative contribution is a measure for the importance of the reaction. Positive values reflect a positive correlation between the flux and the succinate yield, negative values represent for irreversible reactions a negative correlation, for reversible reactions (R) that the direction of the flux should be the opposite as the one indicated in the model.

Chapter 4

Cybernetics: some issues on the method

4.1 Introduction

The lack of knowledge about the mechanisms micro-organisms dispose of to regulate their metabolism severely hampers the use and limits the usefulness of mechanistic modelling, especially when a detailed description of the microbial behaviour is necessary [66, 89]. In an attempt to partially circumvent this problem, cybernetic modelling introduces cybernetic variables [90] in order to accommodate for the microbial control of enzyme synthesis and activity. These variables, whose value is determined by a controller, embody the allocation of the limited resources a cell disposes of to these competing enzymes, yielding the optimal performance [138, 205, 206]. This reasoning seems acceptable as in general, it is believed that a micro-organism tries to optimise its behaviour, *e.g.*, with respect to growth or substrate uptake.

Recently, Young (2005) [223] rethought the framework and more tangibly introduced the principles of optimal control theory. Optimal control theory is a mathematical optimisation method for deriving control policies. It aims to find a control law for a given system such that a certain optimality criterion is achieved. In general, such a control problem includes a gain function and a cost function, relating state and control variables. An optimal controller is a set of differential equations describing the paths of the control variables that maximise the performance function. Rephrasing this in the context of a

micro-organism, the cost becomes, *e.g.*, the pool of amino acids a micro-organism needs to invest for the production of a particular enzyme, and the cell’s gain could be merely growth. This framework will be discussed in Section 4.2.

A key concept in cybernetic modelling is the cybernetic unit. This is a cluster of enzymes that compete with each other for the same pool of limited resources. For simple cybernetic models these cybernetic units could readily be distinguished, as they coincide with the cybernetic basic components (Figure 4.1) [13, 14, 90, 190]. The increasing model complexity, *i.e.*, over time more complex networks with less lumping were considered [143, 172], rendered the identification of these cybernetic units less straightforward and even quite arbitrary [204, 206, 223] (Figure 4.1).

To rationalise the framework, Young (2005) [223] opted for elementary flux modes (EFMs) as cybernetic units. However, whereas the choice for the EFMs as local control unit seems quite obvious, the choice of the associated objective function is not. Young (2005) [223] opted for the optimisation of a harmonic mean flux, J :

$$J = \frac{\sum_{i=1}^n \xi_i}{\sum_{i=1}^n \frac{\xi_i}{r_i v_i}} \quad (4.1)$$

with n the number of reactions, r_i the rate of reaction i , v_i the cybernetic variable controlling enzyme activity, and ξ_i the flux through reaction i in the elementary flux mode. This objective function aims amongst others at a steady throughput through the EFM in an attempt to avoid accumulation or depletion of certain metabolites [223]. However, its biological foundation is unclear.

The application of the approach in the domain of metabolic engineering requires a robust and generic framework. The cybernetic control law for enzyme activity can also cause controversy. As enzyme synthesis, enzyme activity is meticulously controlled through, for instance, allosteric control, regulation by phosphorylation/dephosphorylation and other types of covalent modifications [55, 161]. The knowledge on these regulatory mechanisms to fully model these processes mechanistically is still insufficient. Therefore, in cybernetic modelling, cybernetic variables are introduced into a kinetic model with the aim of substituting the unknown mechanistic details of the cell’s regulatory architecture. However, whereas the pool of limited resources necessary for enzyme synthesis is a concept

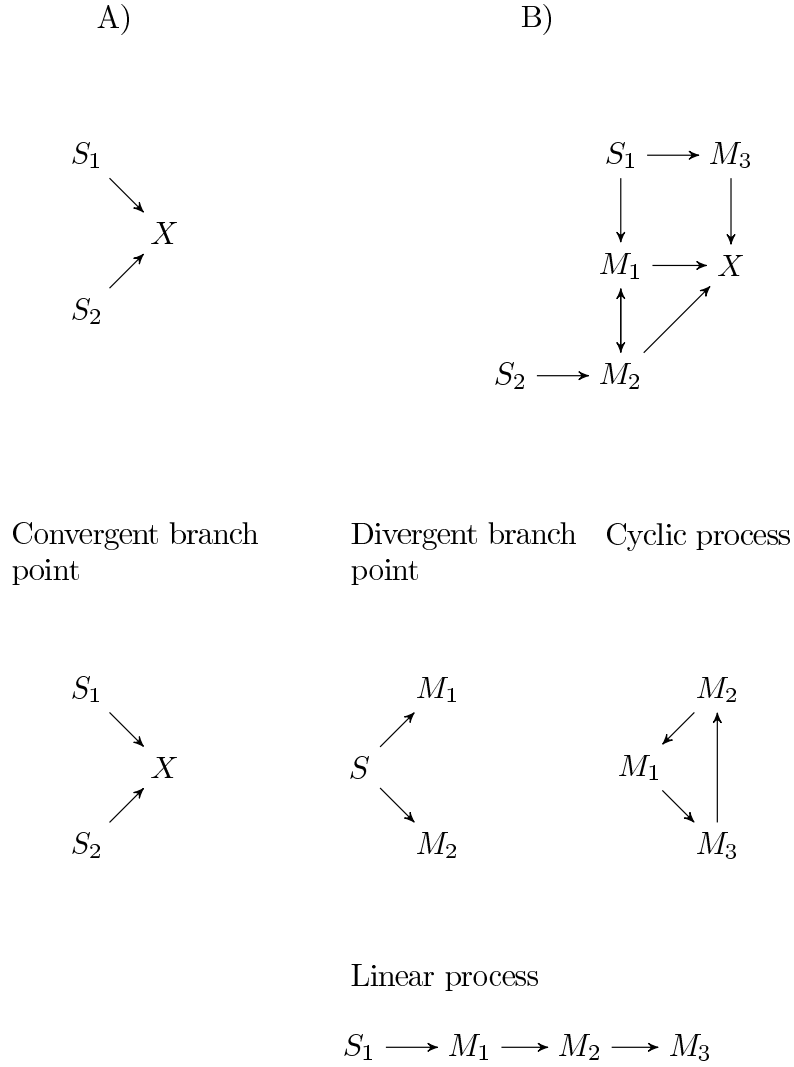


Figure 4.1: Cybernetic models and their increasing complexity, from A) the model of Kompala *et al.* (1984) [90] to B) the model of Guardia *et al.* (2000) [82] and the basic components of the cybernetic units (convergent and divergent branch points, linear and cyclic processes)

easy to grasp (as this pool could be the amino acids/ATP, necessary for the synthesis of enzymes), the concept of such a pool of limited resources for enzyme activity is far more abstract.

Thus, deriving a control law for enzyme activity seems less obvious than the derivation of the control law ruling enzyme synthesis. Kompala *et al.* (1984) [90] derived the matching law to control enzyme activity. With a view to the application of cybernetic modelling for metabolic engineering purposes, as suggested by, *e.g.*, Varner and Ramkrishna (1999) [205, 206] and Young (2005) [223] the need for a generic control law for enzyme synthesis becomes apparent. However, comprehensive arguments why the matching law would be generally valid, were not found. A typical example seemingly not coinciding with the latter, would be the regulation of glutamine synthase [8], since both the activation and inactivation of this enzyme seem to have a cost. Switzer (1977) [175] reports mechanisms for *in vivo* enzyme inactivation which also seem to have a cost.

Consequently, one may wonder what is that cost? What is the no-cost activity, *i.e.*, $v_{no-cost} = \chi$? Is there a cost for up- or down-regulation of enzyme activity? Therefore, some alternatives for this control law were derived and evaluated.

4.2 Materials and methods

4.2.1 Cybernetic framework

The derivation of the cybernetic framework is taken from Young (2005) [223]. A microbial system can be represented by a set of differential equations:

$$\dot{x} = f(x) \tag{4.2}$$

This system is subject to regulatory control both at the level of enzyme synthesis and enzyme activity. These inputs are accounted for by introducing the control vectors u for enzyme synthesis and v for enzyme activity, which specify how the resources are allocated among the various alternatives:

$$\dot{x} = f(x, u, v) \tag{4.3}$$

For clarity, the cybernetic variable controlling enzyme activity, v , will be discarded for now. It is assumed that the cell allocates its resources in such a way that the performance function J is maximised. This can be described by optimal control theory:

$$\begin{aligned} & \max J \\ & \text{subject to : } \dot{x} = f(x, u) \end{aligned} \quad (4.4)$$

Computing the optimal control is numerically quite demanding. Assuming however that regulatory decisions are made at each instant based on the projected system response over a short time interval Δt , the system can be approximated by linearisation:

$$\Delta \dot{x} = A\Delta x + B_u\Delta u + f(x(t), u^0) \quad (4.5)$$

$$A = \left. \frac{\partial f(x, u)}{\partial x} \right|_{x, u^0} \quad (4.6)$$

$$B_u = \left. \frac{\partial f(x, u)}{\partial u} \right|_{x, u^0} \quad (4.7)$$

The change in model performance ΔJ over the system's planning window Δt then becomes:

$$\Delta J = q\Delta x(t + \Delta t) - \frac{1}{2} \int_t^{t+\Delta t} (u^T \sigma_u u) d\tau \quad (4.8)$$

$$q = \frac{\partial \phi(x(t))}{\partial x} \quad \Delta J = J(t + \Delta t) - J(t) \quad (4.9)$$

in which the function $\phi(x(t))$ represents the metabolic objective function of the system and σ_u a parameter that scales the cost associated with resource investment. The solution of this optimal control problem can be derived. The Hamiltonian now becomes [224]:

$$H_1(x, u, \lambda) = -\frac{\sigma_u}{2} u^2 + \lambda^T [A\Delta x + B_u\Delta u + f(x(t), u^0)] \quad (4.10)$$

with λ the costate vector.

The state equation is given by:

$$\dot{x} = \frac{\partial H_1}{\partial \lambda} = A\Delta x + B_u\Delta u + f(x(t), u^0) \quad (4.11)$$

The stationary condition is given by:

$$0 = \frac{\partial H_1}{\partial u} = -\sigma_u u + B_u^T \lambda \quad (4.12)$$

So, one finds for u :

$$u = \frac{B_u^T \lambda}{\sigma_u} \quad (4.13)$$

in which the costate is given by:

$$-\dot{\lambda} = \frac{\partial H_1}{\partial x} = A^T \lambda \quad (4.14)$$

The boundary condition for this equation is:

$$\lambda(t + \Delta t) = q \quad (4.15)$$

Applying the boundary conditions and solving this equation gives:

$$-\int_t^{t+\Delta t} \frac{1}{\lambda} d\lambda = \int_t^{t+\Delta t} A^T dt \quad (4.16)$$

$$\frac{\lambda(t)}{\lambda(t + \Delta t)} = e^{(A^T(\Delta t))} \quad (4.17)$$

So, one finds for λ :

$$\lambda(t) = e^{(A^T(\Delta t))} q \quad (4.18)$$

Substituting Eq. 4.18 in Eq. 4.13, yields:

$$u(t) = \frac{1}{\sigma_u} B_u^T e^{(A^T(\Delta t))} q \quad (4.19)$$

$$p_i(t) = q^T e^{(A(\Delta t))} b_{ui} \quad (4.20)$$

Since the resources are limited however, this is a constrained optimisation problem. The appropriate constraints have to be added to the Hamiltonian (Eq. 4.10), yielding Eq. 4.21. The Karash-Kuhn-Tucker conditions can be derived both for the cybernetic variable controlling enzyme activity and synthesis.

$$H_2(x, u, \lambda, \eta, \nu_i) = H_1(x, u, \lambda) + \eta \left(1 - \sum_{i=1}^n u_i - \omega^2 \right) + \sum_{i=1}^n \nu_i (u_i - \kappa_i^2) \quad (4.21)$$

where ν_i, η are Lagrangian multipliers associated with the i th non-negativity constraint $u_i \geq 0$ and with the total resource constraint $\sum_{i=1}^n u_i \leq 1$, respectively, and n is the number of competing reactions. The stationary condition now becomes:

$$\frac{\partial H_2}{\partial u_i} = 0 \Rightarrow -\sigma u_i + b_{ui} \lambda = \eta - \nu_i \quad (4.22)$$

and the Karash-Kuhn-Tucker conditions are given by:

$$\frac{\partial H_2}{\partial \kappa_i} = 0 \Rightarrow 2\nu_i \kappa_i = 0 \quad (4.23)$$

$$\frac{\partial H_2}{\partial \nu_i} = 0 \Rightarrow u_i - \kappa_i^2 = 0 \quad (4.24)$$

$$\frac{\partial H_2}{\partial \omega} = 0 \Rightarrow 2\eta\omega = 0 \quad (4.25)$$

$$\frac{\partial H_2}{\partial \eta} = 0 \Rightarrow \left(1 - \sum_{i=1}^n u_i - \omega^2 \right) = 0 \quad (4.26)$$

$$u_i = \frac{b_{ui} \lambda + \eta - \nu_i}{\sigma_u} \quad (4.27)$$

The solution that simultaneously satisfies these constraints is:

$$u_i = \max\left(\frac{p_i + \eta}{\sigma_u}, 0\right) \quad (4.28)$$

Since now $u_i \geq 0$. Choosing $\sigma_u = \sum_{i=1}^n \max(p_i, 0)$ and taking into account Eq. 4.25:

$$\eta = 0 \vee \omega = 0 \quad (4.29)$$

one finally finds:

$$u_i = \frac{\max(p_i, 0)}{\sum_{i=1}^n \max(p_i, 0)} \quad (4.30)$$

Akin to the derivation of the control law ruling enzyme synthesis, u , the control law ruling enzyme activity can be derived, with $\sigma_v = \max_n(p_n)$, yielding:

$$v_i = \frac{\max(p_i, 0)}{\max_n(p_n)} \quad (4.31)$$

4.2.2 Cybernetic model

All simulations have been performed with the model of Kompala *et al.* (1984) [90], which describes the bacterial growth on mixtures of substitutable substrates, especially under conditions that give rise to diauxic growth. Two substitutable substrates are considered S_1 and S_2 , which are converted by the enzymes E_1 and E_2 , respectively, to form biomass (see also Figure 4.1). The stoichiometric and kinetic parameters used, are given in Table 4.1. The objective function is given by:

$$J = \frac{r_1 v_1}{Y_1} + \frac{r_2 v_2}{Y_2} \quad (4.32)$$

with $r_i = \frac{x_{E_i}}{x_{E_i^0}} k_i \frac{x_{S_i}}{x_{S_i} + K_{x_{S_i}}}$, $i = \{1, 2\}$. Only the control law for enzyme activity was modified during this simulation study.

Table 4.1: Stoichiometric and kinetic parameters used to simulate diauxic growth [90]

| Sugar | i | k_i (h^{-1}) | K_i (g/L) | Y_i (gB/gS_i) | α_i (h^{-1}) | β_i (h^{-1}) |
|---------|-----|--------------------|-----------------|---------------------|-------------------------|------------------------|
| Glucose | 1 | 1.08 | 0.01 | 0.52 | 1.13 | 0.05 |
| xylose | 2 | 0.82 | 0.2 | 0.5 | 0.87 | 0.05 |

4.3 Results and discussion

4.3.1 Derivation of the control law for enzyme activity

With respect to the derivation of the control law for enzyme synthesis no alternative exists. This is not the case for the control law ruling enzyme activity. Indeed, one may wonder what the no-cost activity is. Is there a cost for up- or down-regulation of enzyme activity? Therefore, some alternatives for this control law were derived and evaluated.

We replaced the cost term for enzyme activity, $\frac{\sigma_v}{2} (v)^2$, in the model performance function as presented by Young (2005) [223], with $\frac{\sigma_v}{2} (v - \chi)^2$. Analogously to Section 4.2, the solution for the optimal control problem has been derived:

$$v_i = \chi + \frac{p_i + \eta_i - \nu_i}{\sigma_v} \quad (4.33)$$

The solution that simultaneously satisfies the Karash-Kuhn-Tucker constraints, for $\chi = 0$ is:

$$v_i = \frac{\max(p_i, 0)}{\max_n(|p_i|)} \quad (4.34)$$

The control law will here be derived for the case $\chi \neq 0$. One finds for $\chi = 0.5$, and choosing $\sigma_v = \frac{\max(|p|)}{\chi}$,

$$v_i = \max \left(\frac{1}{2} \left(1 + \frac{p_i}{\max_n (|p_i|)} \right), 0 \right) \quad (4.35)$$

and for $\chi = 1$, and choosing $\sigma_v = \max_n (|p|)$,

$$v_i = \min \left(1 + \frac{p_i}{\max_n (|p_i|)}, 1 \right) \quad (4.36)$$

The enzymatic conversion capacity will be fully used, unless the return is negative. In this case down-regulation has a cost.

4.3.2 Case study

To stress the importance of the choice of the control law ruling enzyme activity, the performance of the derived control laws was evaluated for the case of sequential substrate utilisation.

The evolution of the substrates S_1 and S_2 and of the biomass X is depicted in Figure 4.2 and the cybernetic variables are depicted in Figure 4.3. Obviously, as is depicted in Figure 4.3, both enzymes will be more active using Eqs. 4.35 and 4.36, in comparison with the matching law, since the return p_i for both lumped pathways is always positive in this case. Hence, the cybernetic variables v_1 and v_2 given by Eq. 4.36 will be 1 throughout the simulation (Figure 4.3). Consequently, the enzymes are fully active, since down-regulation of enzyme activity would cost. Akin reasonings can be elaborated for $\chi = 0.5$.

The aim of the research presented here was not to come up with the 'true' control law, because this is impossible at this stage due to, *e.g.*, the limited knowledge, the lack of appropriate data, and the potential dependency on the model structure. Rather, the aim was to emphasise its importance for the model itself, as the chosen cybernetic control law will have an impact on the optimal values of the parameters to be estimated and

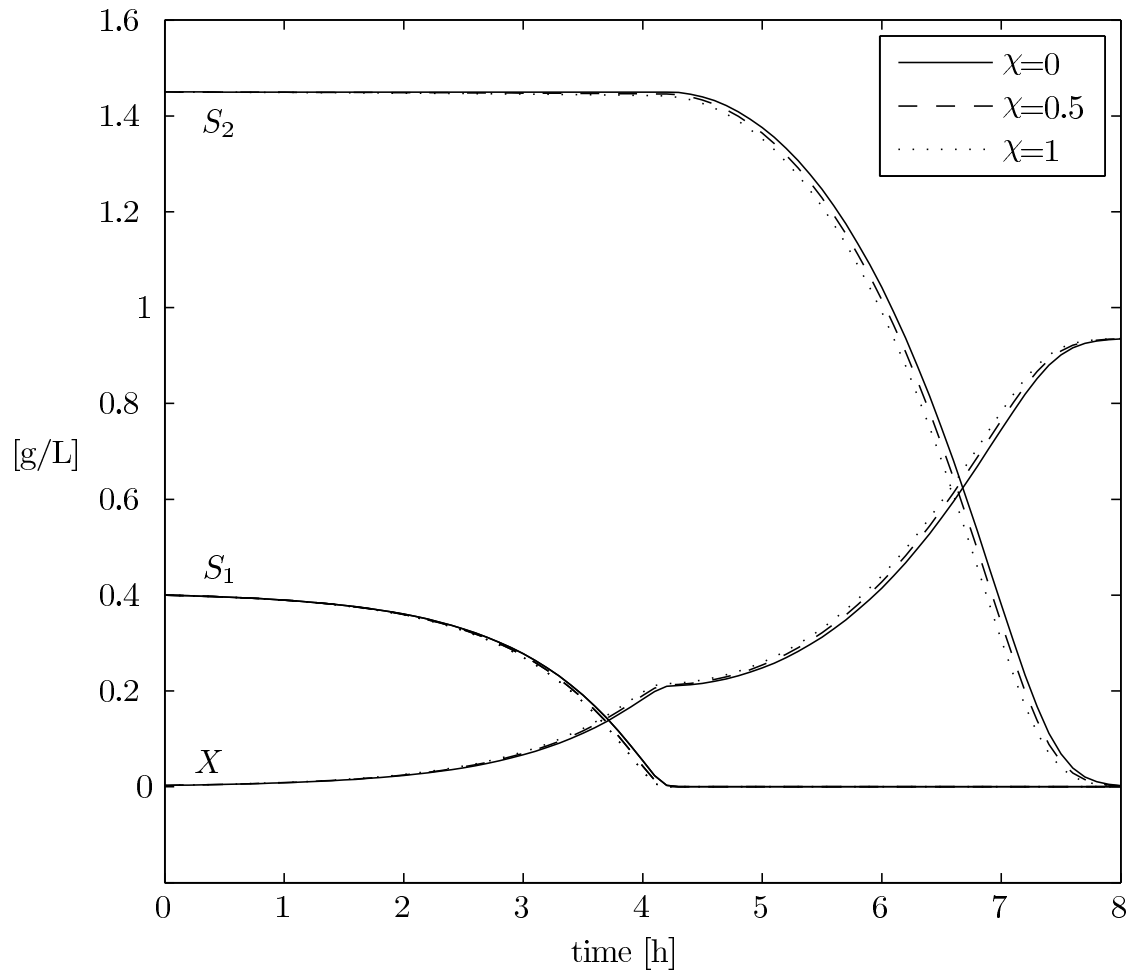


Figure 4.2: Effect of the control law derived for enzyme activity (Eqs. 4.34-4.36) on the evolution of the concentration [g/L] of glucose (S_1), xylose (S_2), and biomass (X) simulated with the model of Kompala *et al.* (1984) [90]

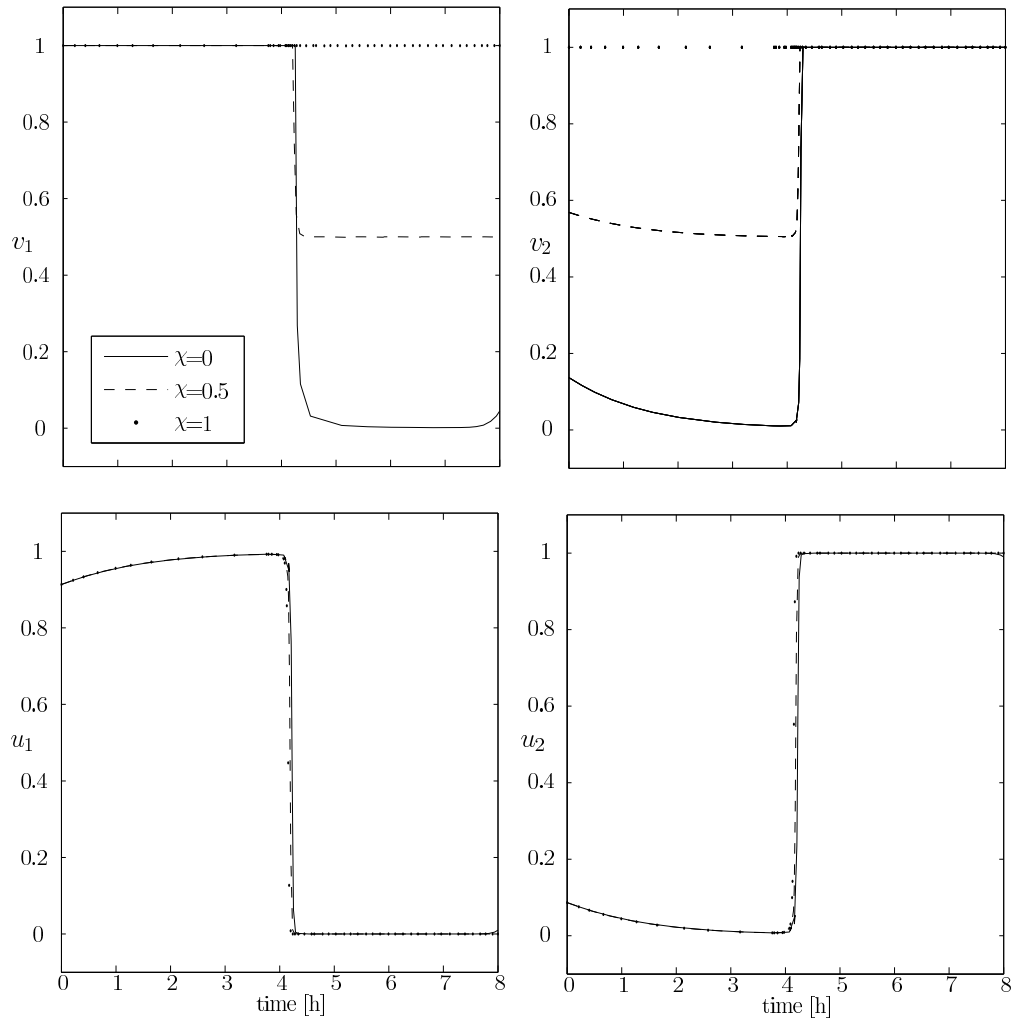


Figure 4.3: Effect of the control law derived for enzyme activity (Eqs. 4.34-4.36) on the evolution of the cybernetic variables v_1 , v_2 , u_1 , and u_2 simulated with the model of Kompala *et al.* (1984) [90]

may have an influence on the model predictions.

Based on the experimental data of Monod (1947) [117], Kompala *et al.* (1984) [90] derived the matching law (Eq. 4.34), predominantly based on the fact that no lag phase could be observed for a pregrown culture. Indeed, in this situation none of the proposed alternatives (Eqs. 4.35 and 4.36) would perform as well as the matching law. However, what was the cause for the observed behaviour? Was it indeed resource investment linked to enzyme activity? Or did the metabolic regulation of enzymes of the lumped pathways play a determining role? Indeed, substrate (use) might have a cost, which would be more or less in agreement with the matching law, but what would be the cost of an abundantly present (extracellular) substrate of an enzymatic conversion and do all enzymes compete then for the same substrate (since only one pool of limited resources to control enzyme activity is considered in the framework presented by Young (2005) [223])?

4.4 Conclusions

Ab initio, cybernetic models have typically been used in the domain of bioprocess control. Recently, the original framework was reworked by Young (2005) [223] with a view to applying this method in the domain of metabolic engineering, in order to cope with problems related to the increased model complexity. Since the application of the cybernetic approach in this domain requires a generic framework.

In view of the latter, different alternatives for the matching law have been derived and evaluated. Obviously, the choice of the control law for enzyme activity is important. However, due to the limited knowledge, issues linked to the model structure, and the lack of appropriate data it was not possible to distinguish between the rival control laws.

Although the approach seems appealing, given the present lack of knowledge, detailed experimental omics data, and some of the problems linked to 'conventional' dynamic metabolic modelling, there still remain some issues unresolved.

Chapter 5

Identification and evaluation of approximative kinetic model structures ¹

5.1 Introduction

In the past, the genetic potential *sensu largo* of a microbial strain was improved by the iterative process of random mutagenesis and screening. The advent of recombinant DNA techniques and functional genomics made it possible to apply a goal-oriented approach for genetic modification (metabolic engineering).

However, in most cases the construction of a producer strain did not turn out to be as straightforward as was initially presumed [12, 89, 195]. Indeed, in complex metabolic networks it is often a futile pursuit to *ad hoc* predict the impact, both qualitatively and quantitatively, of a genetic intervention [12]. Moreover, as the focus in metabolic engineering is shifting from the massive overexpression and inactivation of genes towards fine tuning of gene expression [39, 69, 84], the need for a reliable, quantitative predictor, *i.e.*, a model, that incorporates enzyme kinetics, regulatory mechanisms (which are in general designed to prevent overproduction), compartmentalisation, and the interactions

¹Parts of this chapter have been submitted as: J. Maertens and P. A. Vanrolleghem. Identification and evaluation of approximative kinetic model structures. BMC Bioinformatics, submitted for publication, 2008.

between distinct parts of the cellular metabolic network is growing rapidly.

Initially, stoichiometric models were applied to facilitate the choice of where to intervene genetically. However, it is still unclear whether the well established techniques of metabolic flux analysis [3, 100, 198] and flux balance analysis [5, 34, 142] are fully apt for such aims since the prediction of the optimal flux distribution depends considerably on the kinetics and the regulation of the enzymatic reactions which are not accounted for [171].

The quest for a quantitative approach also led to the development of metabolic control analysis, MCA [85]. MCA aims at eliciting the sensitivity of the metabolic flux distribution to changes of enzyme levels and thus identifies the rate controlling enzyme(s) in the pathway. The applicability of MCA is however limited due to, *e.g.*, its limited extrapolation range around the reference point, which is in general much smaller than the flux shift one aims at in metabolic engineering. It is further based on a steady-state assumption [64] and depends on (unknown) enzyme levels [54].

Although a mechanistic dynamic metabolic model is not suffering from the aforementioned shortcomings, it is no *deus ex machina* either because such a model is complex, overparameterised [42], and the parameter identification is not evident either, because of the highly nonlinear rate equations and the large number of parameters to be estimated.

To deal with the latter problem different approximative, linear non-mechanistic kinetics were suggested. The most popular approximative kinetics are i) the 'log-linear in metabolite and enzyme levels kinetics' [72, 73], ii) the so-called 'linlog kinetics' Eq. 5.1 [74, 209], iii) 'the linear in metabolite and enzyme levels kinetics', iv) the 'linear in metabolite levels kinetics' Eq. 5.2, and v) the 'GMA type power law kinetics' Eq. 5.3 [154].

$$\frac{r_k}{J^0} = 1 + \sum_{l=1}^n \varepsilon_{x_{M_l}, k}^0 \ln \left(\frac{x_{M_l}}{x_{M_l}^0} \right) \quad (5.1)$$

$$\frac{r_k}{J^0} = 1 + \sum_{l=1}^n \varepsilon_{x_{M_l}, k}^0 \left(\frac{x_{M_l}}{x_{M_l}^0} - 1 \right) \quad (5.2)$$

$$\frac{r_k}{J^0} = \prod_{l=1}^n \left(\frac{x_{M_l}}{x_{M_l}^0} \right)^{\varepsilon_{x_{M_l},k}^0} \quad (5.3)$$

with r_k the reaction rate of reaction k , J^0 the steady-state flux, x_{M_l} and $x_{M_l}^0$ the concentration of metabolite l under dynamic conditions and at steady-state, respectively, and ε^0 an elasticity coefficient.

The final aim of such a model-based approach is thus target identification for optimising a production host. These targets are those reactions that control the flux through a reaction network, which can be assessed by calculating the flux control coefficients.

Due to the complexity of metabolic networks and the limited available data for identifying the parameters of a metabolic network model, such models are in general overparameterised [42]. The resulting poorly identifiable parameters can lead to uncertain model predictions. Several approaches have thus far been presented in the field of metabolic engineering in order to deal with or assess the latter:

Nikerel *et al.* (2006) [125] simply removed the terms that contained unidentifiable model parameters. However, amongst others Degenring *et al.* (2004) [42] have observed the potential detrimental impact on the model's performance of such action. These authors reduced their overparameterised model by eliminating parameters based on a (local) sensitivity analysis. However, the importance of a parameter cannot merely be assessed using a local sensitivity analysis. Indeed, the model output can be insensitive to a parameter, but due to strong interaction effects with other parameters it can become overall important [150]. This effect can be determined by the extended FAST method [150] or the method proposed by Sobol (1993) [167].

Kresnowati *et al.* (2005) [92] made use of multiple *in silico* generated data sets to assess this uncertainty; considering the typical lack of experimental data such an approach seems far from realistic.

Hence, the aim of this study was to properly assess the uncertainty on the calculated flux control coefficients with a view to target identification in metabolic engineering. This

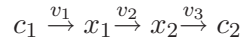
uncertainty may be the result of both an uncertain model structure and of uncertain parameter estimates. To this end, several rival approximative kinetics were used to describe an illustrative pathway. Since, the enzyme levels will be assumed constant in this study, the log-linear in metabolite and enzyme levels kinetics (i) and the linlog kinetics (ii), and the linear in metabolite and enzyme levels kinetics (iii) and the linear in metabolite levels kinetics (iv), two by two coincide, as they only differ with changing enzyme levels.

For this reason, only three approximative kinetic structures (the linlog kinetics, the GMA type power law kinetics, and the linear in metabolite levels kinetics) were retained for further analysis. In order to properly assess the uncertainty on the calculated flux control coefficients the linear kinetic parameters for each of these rival approximative kinetics were identified using a two step parameter identification procedure and the adequacy of the approximative kinetics to describe the system was evaluated.

5.2 Materials and Methods

5.2.1 Linear pathway

The pathway considered in this study was taken from Delgado and Liao (1992) [43], slightly modified by Kresnowati *et al.* (2005) [92], and is presented below. The pathway consists of four metabolites c_1 , x_1 , x_2 , and c_2 and three reactions v_1 , v_2 , and v_3 .



The complete nonlinear kinetic equations are given in Eqs. 5.4-5.6:

$$v_1 = \frac{0.2}{1 + x_1} \quad (5.4)$$

$$v_2 = \frac{1.5x_1}{0.5 + x_1} (1 + L)^{-1}, \quad L = \frac{\left(1 + \frac{x_2}{0.1}\right)^4}{\left(1 + \frac{x_2}{0.5}\right)^4} \quad (5.5)$$

$$v_3 = \frac{x_2}{1 + \frac{x_2}{2} + \frac{x_1}{40}} \quad (5.6)$$

The steady-state flux and steady-state concentrations in Table 5.1 were obtained by solving the mass balance equations for the steady-state condition.

Table 5.1: The Steady-state flux and the steady-state metabolite concentrations and the initial conditions of the perturbation (in arbitrary units)

| | c_1 | x_1 | x_2 | c_2 | J^0 |
|--|-------|-------|-------|-------|-------|
| Steady-state conditions | 2.0 | 0.411 | 0.154 | 0.0 | 0.142 |
| Initial conditions of the perturbation | 2.0 | 1.0 | 1.0 | 0.0 | |

Transient data were obtained by perturbing the steady-state (Figure 5.1). To reflect typical measurement data, normally independently distributed noise $\epsilon(0, \sigma^2)$ was superimposed on these simulated metabolite concentrations. 11 sample points were uniformly distributed over the time interval $[0, 5]$. This is realistic as sampling frequencies up to $4\text{-}5 \text{ s}^{-1}$ are reported in the literature [129].

5.2.2 Derivation of the control coefficients

The derivation of the flux control coefficients is taken from Mauch *et al.* (1997) [115]. A microbial system can be represented by a set of differential equations:

$$\dot{x} = f(x, p) \quad (5.7)$$

with x a vector that contains the state variables and p a vector that contains the parameters. C_{ij}^M is defined as the concentration control coefficient:

$$C_{ij}^M = \frac{\partial x_i}{\partial p_j} \frac{p_j}{x_i} = \frac{\partial \ln x_i}{\partial \ln p_j} \quad (5.8)$$

and C_{ij}^F is defined as the flux control coefficient:

$$C_{ij}^F = \frac{\partial v_i}{\partial p_j} \frac{p_j}{v_i} = \frac{\partial \ln v_i}{\partial \ln p_j} \quad (5.9)$$

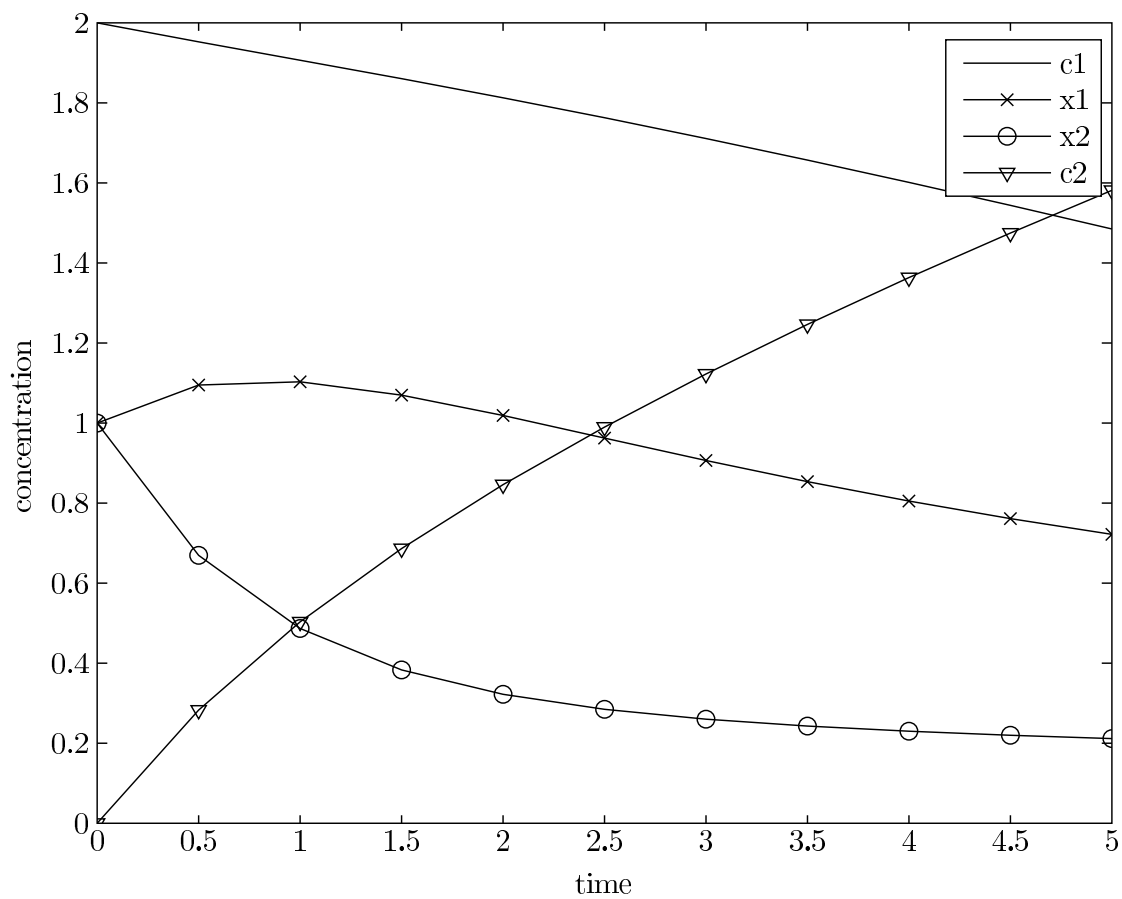


Figure 5.1: Evolution of metabolite concentrations (in arbitrary units) of c_1 , x_1 , x_2 , and c_2 after the perturbation

with v_i the rate of reaction i . The time-derivative of the first-order sensitivity of concentration x_i with respect to parameter p_j is given according to [188] by:

$$\frac{d}{dt} \left(\frac{dx_i}{dp_j} \right) = \sum_{k=1}^m \frac{\partial f_i}{\partial x_k} \frac{\partial x_k}{\partial p_j} + \frac{\partial f_i}{\partial p_j} \quad (5.10)$$

Eq. 5.10 can be written as:

$$\frac{dC_{i,j}^{M^*}}{dt} = \sum_{k=1}^m J_{ik} C_{kj}^{M^*} + \psi_{ij} \quad (5.11)$$

where $C_{i,j}^{M^*}$ denotes the non-normalised sensitivity of concentration x_i with respect to parameter p_j . $J_{i,k}$ describes the derivative of the i^{th} element of vector f with respect to the k^{th} element of state vector x . ψ_{ij} is the derivative of f with regard to the j^{th} parameter of p_s . In matrix notation Eq. 5.11 becomes:

$$\frac{dC^{M^*}}{dt} = JC^{M^*} + \psi \quad (5.12)$$

where $C^{M^*} [m, s]$ is the non-normalised concentration control matrix. $J [m, n]$ is the well-known Jacobian matrix, whilst $\psi [m, s]$ denotes the matrix containing the sensitivities of the right-hand side of Eq. 5.7 with respect to parameter vector P_s . For steady-state conditions one finds:

$$\frac{dC_0^{M^*}}{dt} = 0 \quad (5.13)$$

If J is invertible, one can write:

$$C_0^{M^*} = -J_0^{-1} \psi_0 \quad (5.14)$$

Eq. 5.14 can be transformed into the normalised concentration control matrix at steady-state, C_0^M , with the diagonal matrices $X_0 [m, m]$ and $P_S [s, s]$, which contain the steady-

state concentrations and parameter values, respectively:

$$C_0^M = -X_0^{-1} J_0^{-1} \psi_0 P_s \quad (5.15)$$

Time dependent concentration control coefficients can be obtained through integration of Eq. 5.12 using Eq. 5.14 with subsequent normalisation:

$$C^M(t) = X^{-1} \left(\int_0^t (J C^{M*} + \psi) dt + C_0^{M*} \right) P_s \quad (5.16)$$

In Eq. 5.16, $X [m, m]$ represents a matrix whose components are the time-dependent concentrations on the diagonal, and zero otherwise. Note that the time traces of state vector x can be obtained by solving Eq. 5.7.

Analogously time-dependent flux control coefficients can be derived. The differentiation of rate r_i with respect to p_j leads to:

$$\frac{dv_i}{dp_j} = \sum_{k=1}^m \frac{\partial v_i}{\partial x_k} \frac{\partial x_k}{\partial p_j} + \frac{\partial v_i}{\partial p_j} \quad (5.17)$$

and rewritten in a dimensionless form:

$$C_{ij}^F = \sum_{k=1}^m \varepsilon_{ik} C_{kj}^M + \pi_{ij} \quad (5.18)$$

in which the elasticity coefficient, ε_{ik} , is defined as:

$$\varepsilon_{ik} = \frac{\partial v_i}{\partial x_k} \frac{x_k}{v_i} \quad (5.19)$$

This elasticity coefficient describes the fractional change of the local reaction rate r_i to an infinitesimal small perturbation of concentration x_k . In other words, $\varepsilon_{i,k}$ is a measure

of the order of the local reaction rate with respect to concentration x_k . The π -elasticity coefficient, $\pi_{i,j}$, introduced in Eq. 5.18, is defined as:

$$\pi_{ij} = \frac{\partial v_i}{\partial p_j} \frac{p_j}{v_i} \quad (5.20)$$

and is used to represent the relative change of the local reaction rate r_i to a relative, infinitesimal small change of parameter p_j . In matrix notation, Eq. 5.18 becomes:

$$C^F = \varepsilon C^M + \pi \quad (5.21)$$

C^M is provided by Eq. 5.16. For steady-state conditions Eq. 5.21 becomes:

$$C_0^F = \varepsilon_0 C_0^M + \pi_0 \quad (5.22)$$

$$C_0^F = -\varepsilon_0 X_0^{-1} J_0^{-1} \psi_0 P_s + \pi_0 \quad (5.23)$$

Knowing the time traces of the concentration control coefficients, the course of the flux control coefficients can be described as:

$$C^F(t) = \varepsilon X^{-1} \left(\int_0^t (J C^{M^*} + \psi) dt + C_0^{M^*} \right) P_s + \pi \quad (5.24)$$

A flux control coefficient is a measure of how a change in the concentration of an enzyme affects the steady-state flux through that particular pathway. Hence, it is a measure of the degree of control exerted by this enzyme on the steady-state flux [222].

5.2.3 Identification procedure

A two step identification procedure has been applied. In a first step the parameters are estimated using an element-wise weighted total least squares estimator (EW-TLS). In a

second step a Bayesian approach is followed to determine the posterior distribution of the parameter estimates, using the prior distribution obtained in the first step.

Calculation of the derivatives

To obtain the parameters of the set of ordinary differential equations in a linear form the time derivatives of the metabolite concentrations have to be determined:

$$\underbrace{\frac{dc}{dt}}_{b_1 + \Delta b_1 + \epsilon} = \sum_{j=1}^m s_{c,j} J_j^0 \overbrace{\left(1 + \sum_{l=1}^n \underbrace{\left(\frac{c_{M_l}}{c_{M_l}^0} - 1 \right)}_{(A_1 + \Delta A_1)} \underbrace{\varepsilon_{c_{M_l};j}^0}_x \right)}^{r_j} \quad (5.25)$$

with m the number of reactions involved in the formation and utilisation of metabolite c . With b_1 the dependent variables, Δb_1 the errors in the dependent variables, A_1 the independent variables, ΔA_1 the errors in the independent variables, and x the linear parameters. To calculate the derivative of the concentration data with respect to time, a smoother is needed since the derivation of noisy data is an ill-posed problem. A wide variety of smoothers exist [35, 52, 155], in this work a penalised least squares smoother [52] has been used.

Element-wise weighted total least squares

A wide variety of linear estimators exists to solve Eq. 5.26:

$$(A_1 + \Delta A_1) x = b_1 + \Delta b_1 + \epsilon \quad (5.26)$$

Only when the errors in the independent variables, Δb_1 , are negligible compared to those in the dependent variables, ΔA_1 , [46], *i. e.*, $\Delta A_1 \cong 0$, the ordinary least squares estimator Eq. 5.27 yields an unbiased estimate.

$$\hat{x}_{OLS} = \min_{x \in \mathbb{R}^m} \|A_1 x - (b_1 + \Delta b_1)\|_2 \quad (5.27)$$

Unfortunately this condition does not hold here as the metabolite concentration data are inaccurately known. An estimator which does yield unbiased estimates asymptotically is known as element-wise weighted total least squares [94, 199], as it takes the errors on both dependent and independent variables into account. Firstly, the parameters were estimated with an element-wise weighted total least squares estimator, using as initial conditions the parameters obtained using an ordinary least squares estimator. Subsequently, a local nonlinear parameter estimation (simplex algorithm) was performed to obtain more accurate estimates of the parameters, which typically consists of minimising the weighted sum of squared error functional J (assuming independently and normally distributed noise), by optimally selecting the parameter values:

$$J = \sum_{i=1}^n \sum_{j=1}^m \left(\frac{b_{j,E}(i) - b_{j,IS}(i)}{\sigma_j} \right)^2 \quad (5.28)$$

with $b_{j,E}(i)$ and $b_{j,IS}(i)$ the experimentally determined and *in silico* calculated value of state variable j at time i , respectively.

Calculation of the variance-covariance matrix

The obtained parameter values are then used to calculate the Fisher information matrix, FIM , Eq. 5.29:

$$FIM = \sum_{i=1}^n \left[\left(\frac{\partial b_i(t, x)}{\partial x} \right)^T Q_k^{-1} \left(\frac{\partial b_i(t, x)}{\partial x} \right) \right] \quad (5.29)$$

with n the number of sampled time points and Q_k the measurement error variance covariance matrix given by, assuming independently and normally distributed noise:

$$Q_k = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & \sigma_j^2 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & \sigma_m^2 \end{pmatrix} \quad (5.30)$$

with m the number of measured variables and σ_j the corresponding measurement error standard deviation.

For each (state variable, parameter) combination the optimal perturbation factor θ has been determined to numerically calculate the sensitivity functions in Eq. 5.29, $\frac{\partial b}{\partial x}$, using a finite difference method, $\frac{b(t,x+\theta x)-b(t,x)}{x+\theta x-x}$. To this end, the perturbation factor out of $\{10^{-12}, 10^{-11}, \dots, 10^{-6}, \dots, 10^{-2}\}$ yielding the minimum sum of absolute relative errors (SRE) was retained, Eq. 5.31 [40, 41]:

$$SRE = \left| 1 - \frac{\frac{\partial b}{\partial x_-}}{\frac{\partial b}{\partial x_+}} \right| \quad (5.31)$$

where the subscripts, $-$ and $+$, in Eq. 5.31 stand for the sign of the perturbation factor.

According to the Cramer-Rao inequality [107] the Fisher information matrix is related to the lower bound of the parameter estimation error covariance matrix C , under some conditions, *i.e.*, the noise should be uncorrelated and normally distributed $(0, \sigma_j^2)$ [107]:

$$C \geq FIM^{-1} \quad (5.32)$$

Sampling the prior distribution of the parameters

The parameter space will then be sampled n times according to the parameter estimation variance covariance matrix, C . Then, the prior probability $P(x_i)$ for each of these sampled discrete parameter sets x_i is $\frac{1}{n}$. For every set of sampled parameter values the

model is solved. The likelihood of this set, *i.e.*, the probability of observing the data b given the parameter set x_i can then easily be calculated, assuming independently and normally distributed measurement noise [118]:

$$P(b | x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \frac{1}{\prod_{j=1}^m \sigma_j} \exp\left(-\sum_{j=1}^m \frac{(b_{j,E}(i) - b_{j,IS}(i))^2}{2\sigma_j^2}\right) \quad (5.33)$$

with $b_{j,E}(i)$ and $b_{j,IS}(i)$ the experimentally determined and *in silico* calculated value of state variable j at time i , respectively.

Calculation of the posterior distribution of the parameters

The second step of this 2 step Bayesian parameter identification procedure consists of calculating the posterior distribution on the basis of the prior distribution $P(x_i)$ [17]. The posterior distribution is given by [118]:

$$P(x_i | b) = \frac{P(b | x_i) P(x_i)}{P(b) = \int P(b | x_i) P(x_i) dx \cong \sum_{i=1}^n P(b | x_i) P(x_i)} \quad (5.34)$$

with $P(b)$ the probability of observing the data and $P(b | x_i)$ the probability of observing the data b given the parameter set x_i , Eq. 5.33.

5.3 Results and discussion

For the linear in metabolite levels, the GMA type power law, and the linlog approximative kinetic formats the parameters were estimated using sequentially the EW-TLS and a nonlinear parameter estimator. The results are depicted in Figure 5.2. A model adequacy test, χ^2 -test as described by [31], has been used to evaluate the adequacy of the different approximative kinetics to describe the collected data. All studied approximative kinetic formats performed adequate for the case presented here.

This is somewhat contradictory to Heijnen (2005) [74] who pinpoints out the advantages of the linlog kinetic format. According to Heijnen (2005) [74], allows the curvature of the linlog kinetic format to capture the true kinetics over a much larger metabolite range in comparison with the linear in metabolite levels format. Note that the original kinetic equations used in this study are highly nonlinear and one can not pretend the metabolite range observed during the transient is small.

To properly assess the uncertainty on the calculated elasticity coefficients and the flux control coefficients a Bayesian method has been applied. A first step in this Bayesian approach is the determination of the prior distribution. To this end, the by the inverse of the Fisher information matrix linearly approximated parameter estimation error covariance matrix was used. This requires the sensitivity functions of the state variables to the elasticity coefficients, which are depicted in Figure 5.3. Based on these results the parameter estimation error covariance matrix has been calculated.

The influence of the chosen prior distribution on the posterior distribution has been examined for the linlog kinetic format. Both a non-informative and an informative prior were evaluated. As non-informative prior a uniform distribution has been chosen. It is non-informative as all possible values (here) in the 95 % confidence interval (from a frequentist point of view) of the parameter estimates are a priori equally likely. As informative prior the distribution obtained through the parameter estimation error covariance matrix has been used.

From the resulting informative and non-informative prior distributions 10^5 parameter sets have been sampled (Figure 5.4). For each set the flux control coefficients and its likelihood have been calculated. Finally, the posterior probability distribution was calculated. Increasing the number of samples even more, did not alter the posterior distribution (results not shown). The resulting distribution on the flux control coefficients is given in Figure 5.5. The posterior distribution did not seem to be influenced by the used prior (data not shown).

The approach presented in this study thus attempts to take the uncertainty on the flux control coefficients into account for the purpose of identifying potential bottlenecks in the metabolic network. Even for large models [109, 147] such an approach becomes feasible, *e.g.*, by means of distributing computing [9] .

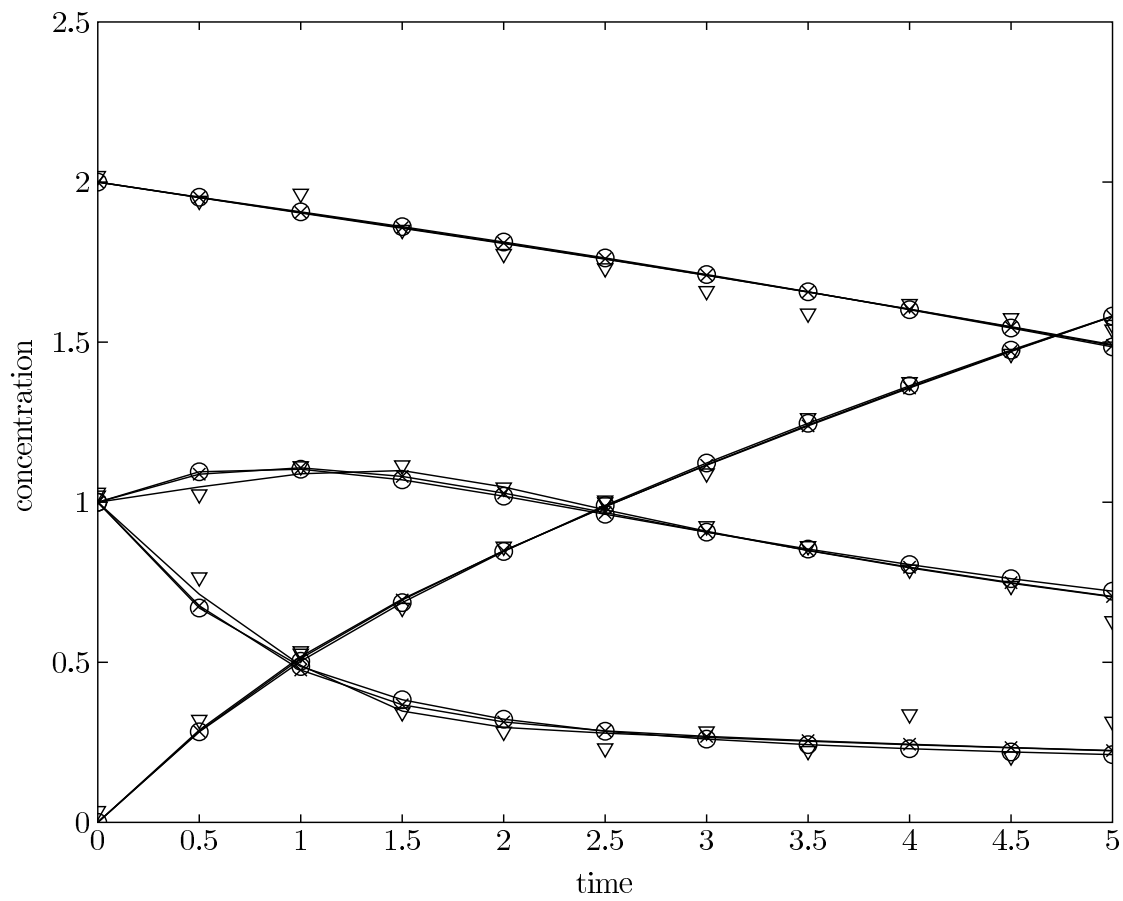


Figure 5.2: Evolution of metabolite concentrations of c_1 , x_1 , x_2 , and c_2 after the perturbation, measured (∇) and simulated by the linlog model (-o), the GMA type power law model (-), and the linear in metabolite levels model (-x).

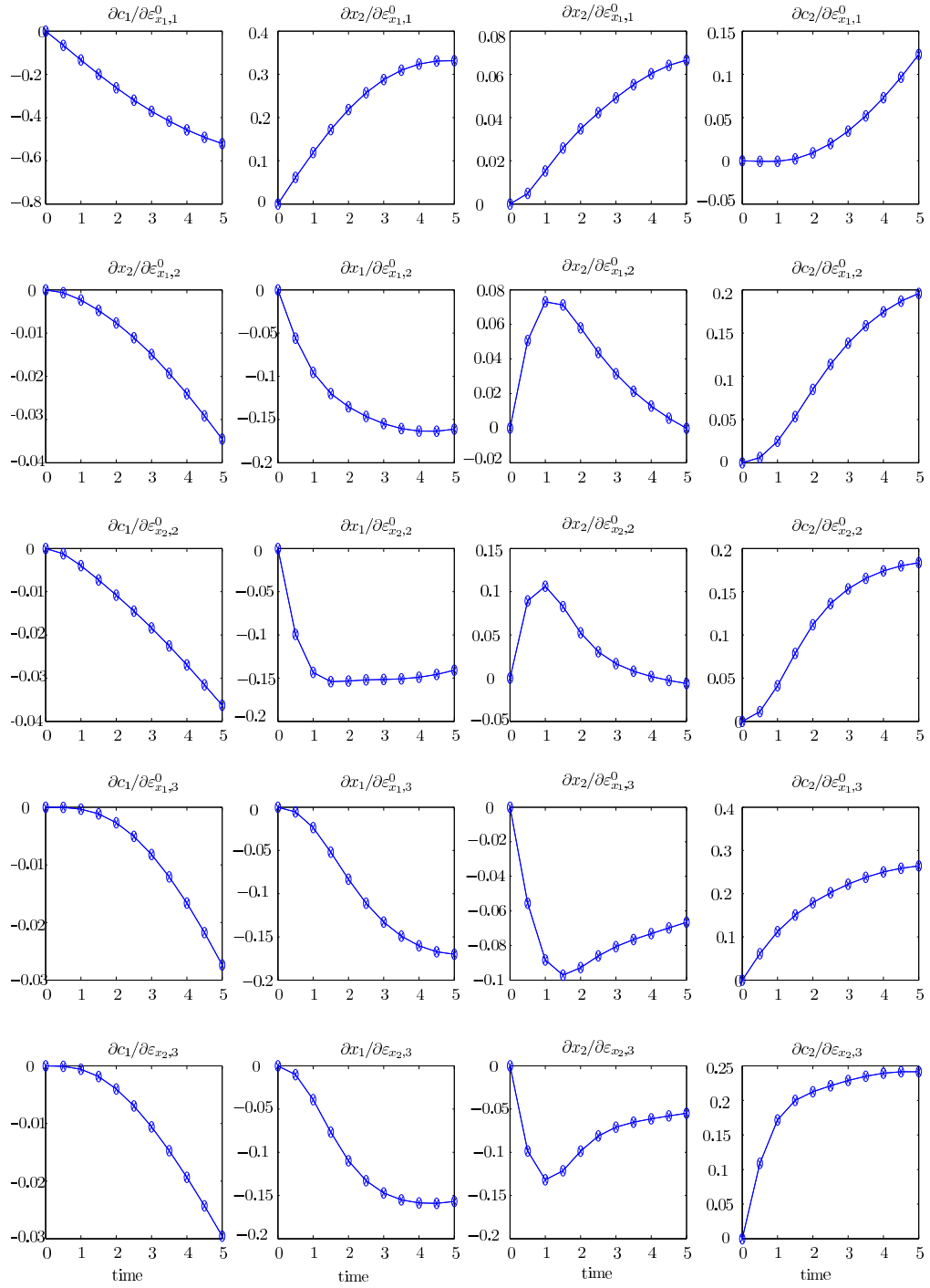


Figure 5.3: Sensitivity functions of the state variables, c_1 , x_1 , x_2 , and c_2 , to the elasticity coefficients $\varepsilon_{x_1,1}^0$, $\varepsilon_{x_1,2}^0$, $\varepsilon_{x_2,2}^0$, $\varepsilon_{x_1,3}^0$, and $\varepsilon_{x_2,3}^0$ of the linlog model after the perturbation

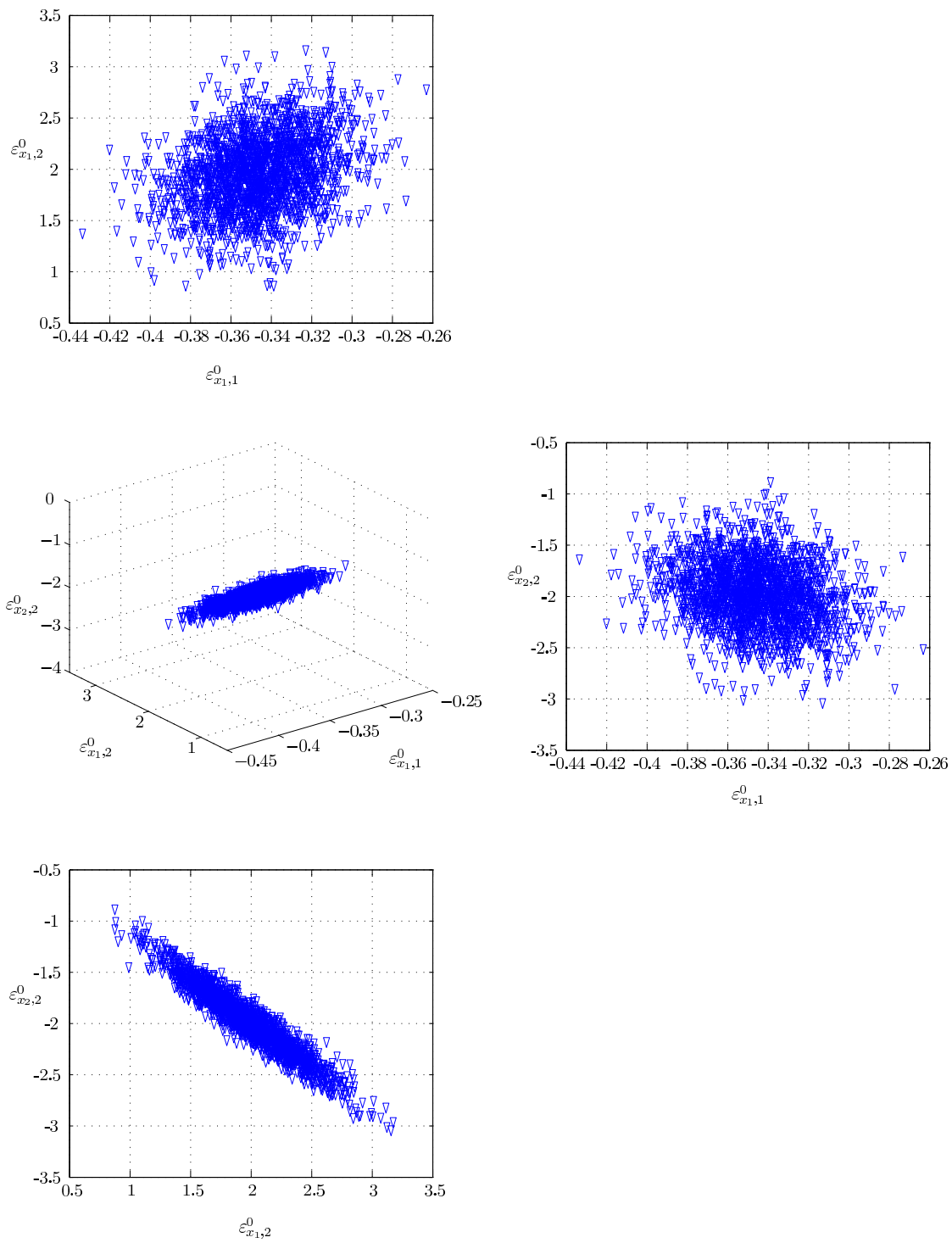


Figure 5.4: Sampling from the informative prior distribution for the linlog model, represented in the parameter space

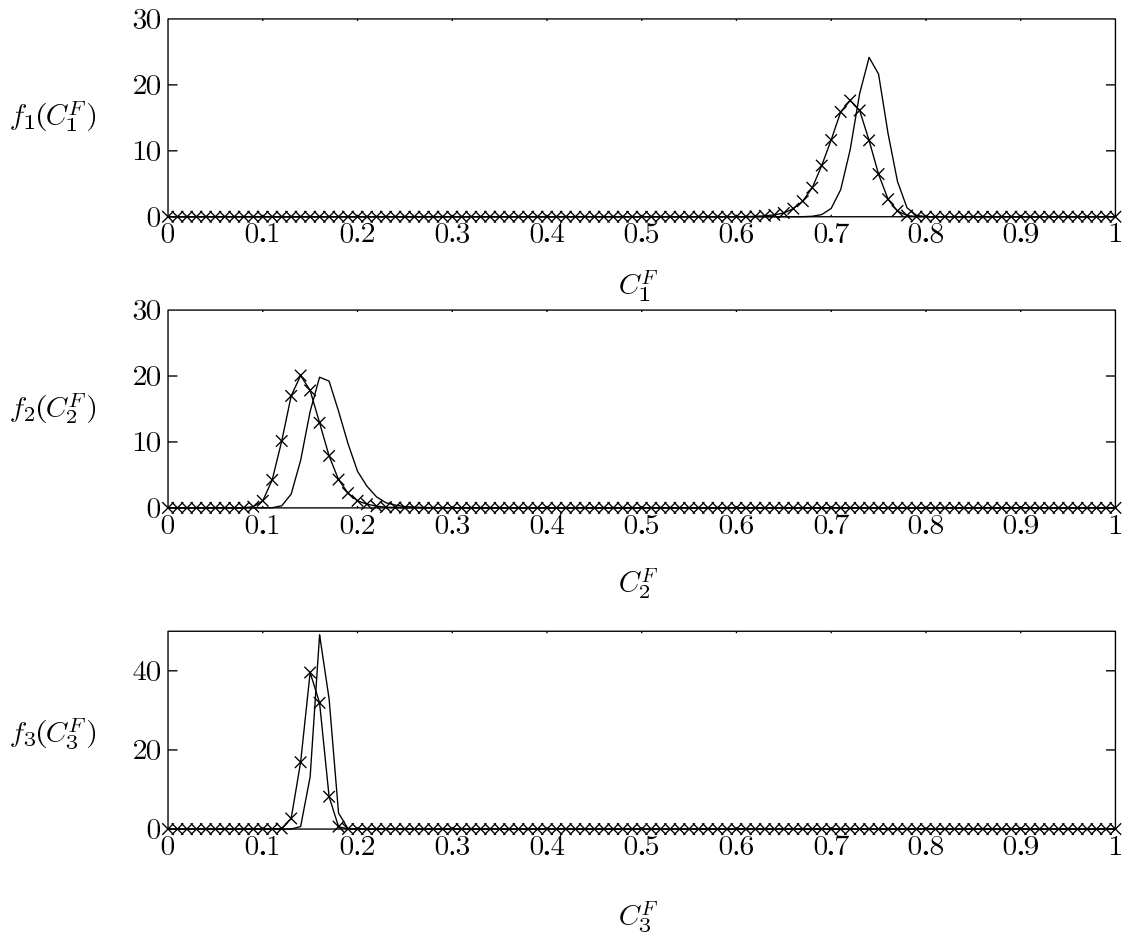


Figure 5.5: The informative prior (-x) and the posterior probability density functions (-) of the flux control coefficients of the linlog model

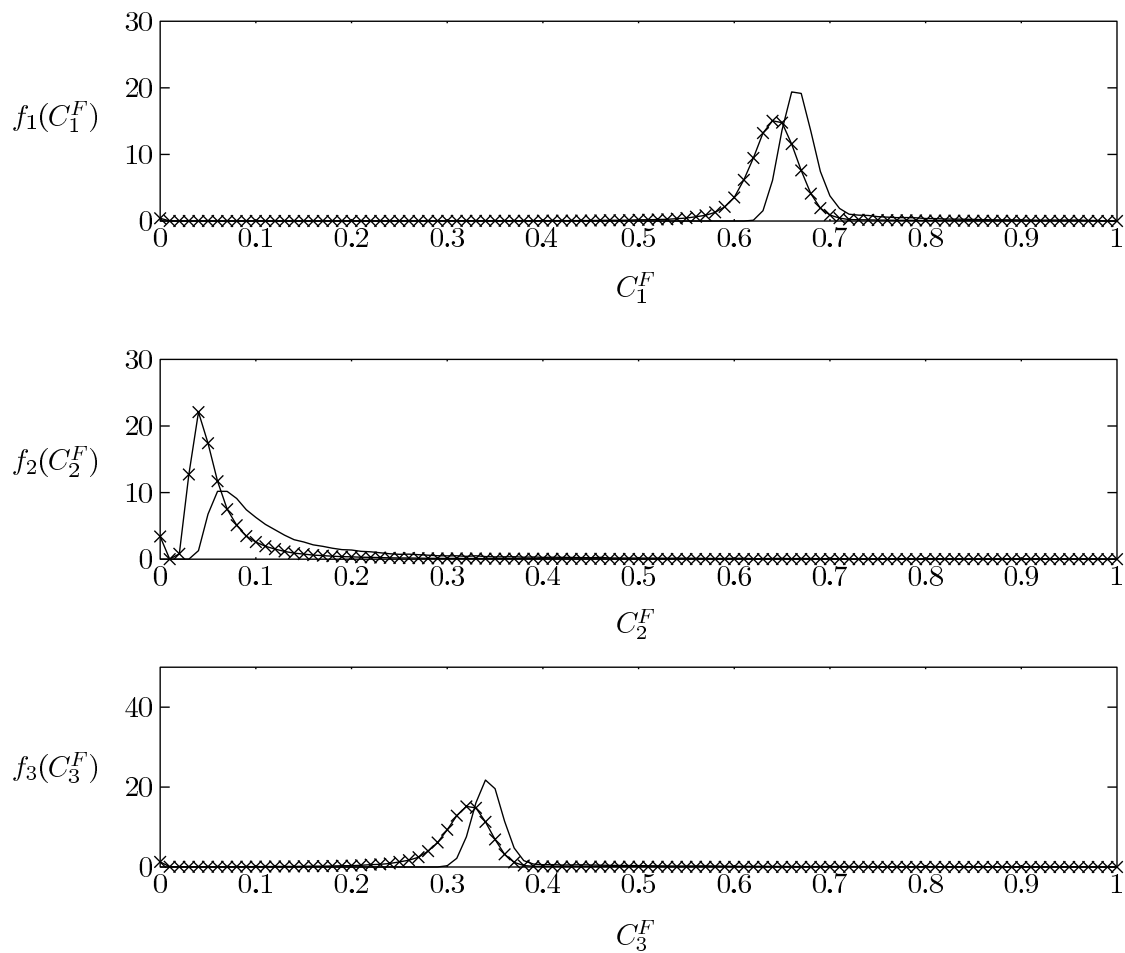


Figure 5.6: The informative prior (-x) and the posterior probability density functions (-) of the flux control coefficients of the GMA type power law model

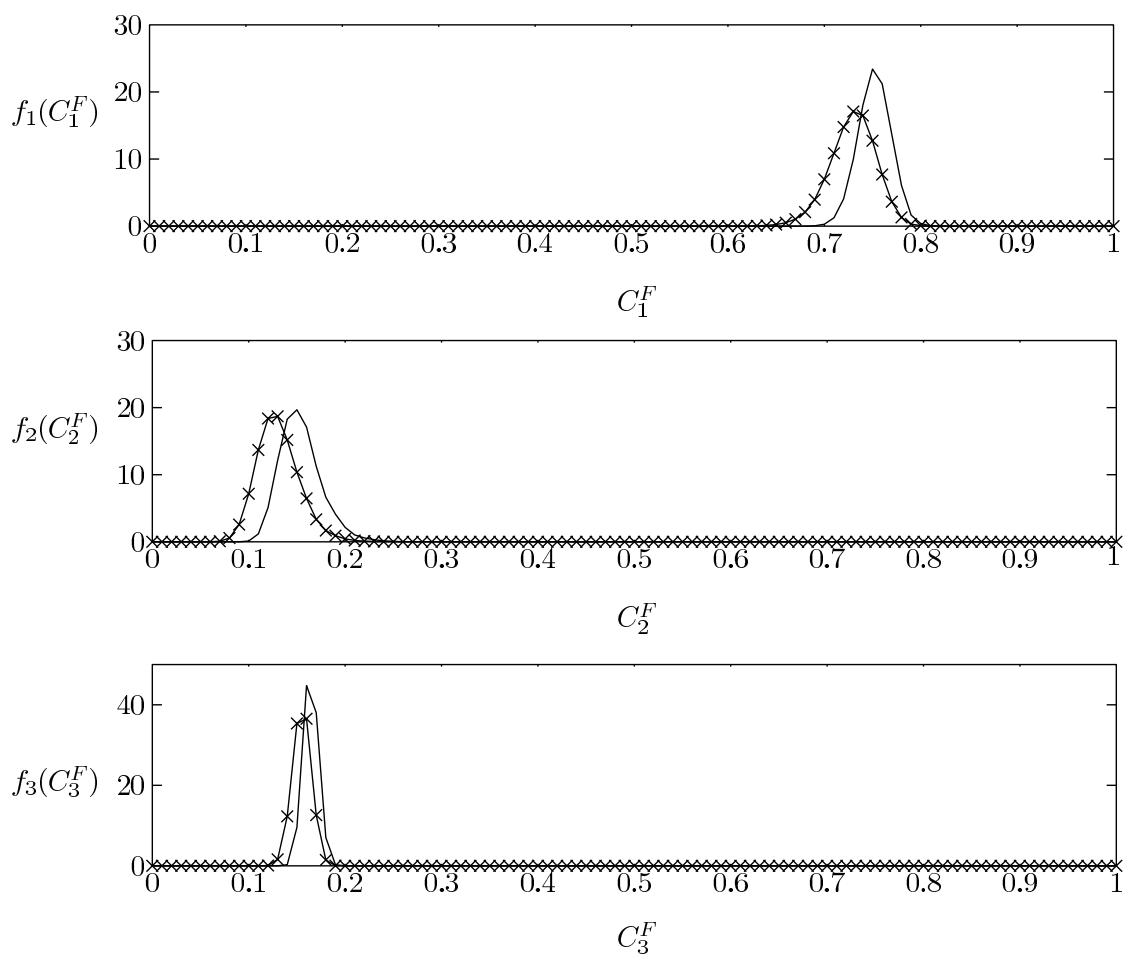


Figure 5.7: The informative prior (-x) and the posterior probability density functions (-) of the flux control coefficients of the linear in metabolite levels model

Though the order of magnitude of the calculated flux control coefficients seems in reasonable agreement with the true values (Figures 5.5-5.7), the true flux control coefficients ($C_1^F = 0.69$, $C_2^F = 0.1$, and $C_3^F = 0.21$) do not always seem to lay in the 95 % credible intervals, which could potentially be due to chance or to the model structure of the approximative kinetics. However, the applied method can be considered a more reliable way to assess the uncertainty on the calculated values of the flux control coefficients.

All approximative kinetics were judged (equally) adequate. Hence, none of the three model structures could be discarded from this analysis and all approximative kinetics were considered simultaneously in an attempt to take the uncertainty of the model structure on the calculated flux control coefficients into account. As described above, the prior and posterior distributions have been calculated, which are depicted in Figure 5.8. Note, that the obtained prior is nothing more than the rescaled superposition of the prior probability density functions of the individual approximative kinetics, as all model structures were considered equally likely. The resulting posterior distribution is however weighted, with the likelihood of each observation. As can be seen in Figures 5.5-5.8 especially the GMA type power law model is determining the resulting posterior distribution for all approximative kinetic formats.

Only a small network has been investigated and issues linked to the increased model complexity of larger models have thus not been encountered, *e.g.*, error accumulation. Whereas in the case presented here approximative kinetics appear to give fair estimates of the flux control coefficients, this approach seems to perform less well for larger networks. For example, Visser *et al.* (2004) [210] successfully re-designed primary metabolism in *E. coli* using the theoretically derived linlog elasticity coefficients and the linlog kinetic format on the basis of the model of Chassagnole *et al.* (2002) [30]. However, when one attempts to simulate the glucose perturbation experiment, that originally was used to identify the mechanistic parameters of the model, using the theoretically derived linlog elasticity coefficients and the linlog kinetic format, the predicted evolution of the metabolite concentrations did not make sense at all (data not shown).

It seems that the usefulness of such approximative kinetic formats decreases with increasing model size and complexity [92, 125, 210]. In order to collect informative data from one or more perturbation experiments for parameter identification purposes, it might be

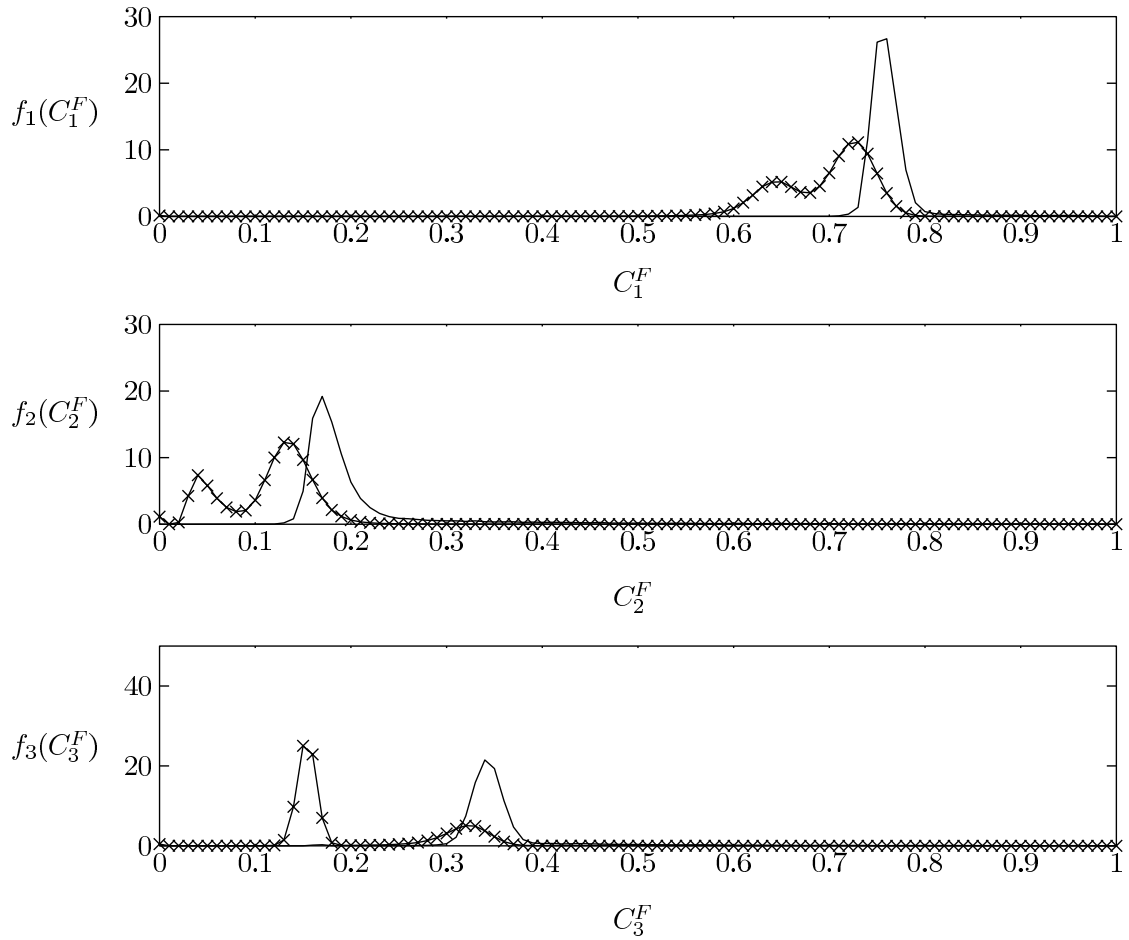


Figure 5.8: The informative prior (-x) and the posterior probability density functions (-) of the flux control coefficients of the the linlog model, the GMA type power law model, and the linear in metabolite levels model

necessary to radically perturb the cell. Probably, way beyond the envisaged metabolites' pool sizes shifts as a result of metabolic redesign and likely way beyond the metabolite range for which approximative kinetic formats yield an adequate description of the true kinetics, because a perturbation broadens and dampens out when it passes through a network [6, 7].

5.4 Conclusions

Due to the complexity of metabolic networks and the limited available data for the identification of the parameters of a metabolic network model, such models are in general overparameterised [42]. This leads to poorly identifiable parameters resulting in uncertain model predictions.

A Bayesian method is proposed to properly assess the uncertainty on the calculated flux control coefficients in view of increasing the trustworthiness of the identified metabolic engineering targets. Though the order of magnitude of the calculated flux control coefficients seems in reasonable agreement with the true values, the true flux control coefficients did not always seem to lay in the 95 % credible intervals.

All of the state of the art approximative kinetic formats: the linlog kinetics, the GMA type power law kinetics, and the linear in metabolite levels kinetics adequately described the data, even though the original kinetic equations used here are highly nonlinear and the metabolite range observed during the transient is not small.

It is shown that to a large extent the uncertainty on the calculated flux control coefficients is due to an uncertain model structure and consequently it is worth the effort to increase the trustworthiness of the identified metabolic engineering targets by means of experimental design for model discrimination.

Chapter 6

A modus operandi of the Bioscope to study oscillating microbial systems

6.1 Introduction

Whether microbial fermentation processes can be an attractive alternative for the production of many chemicals for the well-established chemical synthesis routes depends predominantly on the overall process performance. Consequently, the optimisation of microbial processes is a must, certainly compared to those more mature chemical synthesis routes. This is especially true, as in general the cell's objective function [159], *e.g.*, optimisation of growth, considerably differs from that by which the fermentation process is judged.

A systematic approach for this optimisation, by means of a genetic intervention (metabolic engineering) or by optimising the external conditions, finds more and more acceptance as it is quite difficult to predict *ad hoc* the global impact of a genetic intervention and of varying environmental conditions, respectively [12].

A systematic approach may use metabolic modelling as tool to fully understand the mechanisms, *e.g.*, allosteric control (feed-back, feed-forward control), control of protein induction, ..., which yield these altered process performances. Such metabolic models,

both steady-state [124, 198] and dynamic [30, 149] ones, are increasingly applied for the purpose of identifying the bottlenecks in the metabolic network and the elucidation of regulatory mechanisms.

However, also metabolic network models are subject to the ancient saying 'garbage in is garbage out' and consequently a properly validated model is *a conditio sine qua non* to rely on the model's outcome for process optimisation [166]. In this respect the design of experiments dedicated to model building can become an additional bottleneck. Indeed, whilst striving for a proper identification of the parameters of the metabolic network model many researchers have thus far been confronted by the limits of the available experimental data they have gathered [42].

This is due to the limited information richness of a single perturbation experiment aimed at identifying the metabolic network model's parameters and structure and deciphering regulatory mechanisms in microbial organisms. Thus, multiple experiments have to be performed starting with a culture characterised by a well defined metabolomic and proteomic state [114, 211]. Obtaining such a well defined state typically takes a lot of time and, consequently, being able to eliminate the perturbation of this state would be very welcome.

Recently, experimental set-ups have been designed in order not to perturb this well defined state and still being able to perform multiple perturbation experiments, *e.g.*, a Bioscope connected in series with a chemostat [212] (Figure 6.1, scenario I). The most important feature of a chemostat is that all culture conditions, *e.g.*, dissolved oxygen concentration, pH, cell density, ... remain constant. The effluent of the chemostat is the influent of the Bioscope. Hence, the properties of the incoming flow in the Bioscope are constant. A Bioscope is a plug flow reactor which is continuously fed by broth from the bioreactor [212]. Instead of perturbing the bioreactor itself and consequently all its biomass, the continuous flow from the bioreactor into the Bioscope is perturbed just after entering the Bioscope. Importantly, the chemostat itself is not affected by these perturbations [212]. Perturbing the reactor itself to collect a data set describing the response of the cells to the perturbation would lead to a long waiting time before another perturbation experiment can be performed because the culture must be allowed to regain its steady-state [212]. Distributed over its length the Bioscope has a number of sample ports. Because of the plug flow characteristics of the Bioscope every sample

port is linked to a sample time after perturbation. Obviously, these sample times after perturbation are determined both by the distance from the inlet of the Bioscope and the flow of broth and perturbing agent through the Bioscope. In the traditional steady-state operation of the bioreactor the sample ports of the Bioscope are opened one by one for a given amount of time in order to collect sufficient sample for the analyses to be performed.

Although the emphasis of process optimisation is nowadays shifting towards the genetic modification of microbial strains, dynamically operated cultures can be industrially interesting as well, as, for instance, the interplay of the fluctuating metabolome, *e.g.*, the ATP-paradox [168, 182], and eventually even the proteome can result in an altered process performance. In many fermentation processes, such dynamically operated cultures result in an altered yield, as the interplay of the fluctuating extracellular conditions and the altered metabolite levels (and enzymatic armamentum) results in an adapted cell with a superior [185, 189] or a deteriorated performance [16, 23, 202].

Metabolic models can also be useful tools for the optimisation of such dynamically operated cultures. Hence, the objective of this contribution is to propose a new modus operandi of the Bioscope so that this equipment can also be used to perform multiple perturbation experiments with microbial systems that are subject to a periodic operation (Figure 6.1, scenario II). This will allow to rapidly collect the necessary data in view of identifying the model's parameters and structure.

6.2 Model

The dynamic model of Chassagnole *et al.* (2002)[30] was used as data generating model. This metabolic network model describes the dynamic behaviour of the central carbon metabolism of *Escherichia coli*, *i.e.*, of 25 metabolites that are involved in 30 reactions of the glycolysis and the pentose phosphate pathway, after perturbation of a carbon limited continuous culture by a glucose pulse. The general form of the mass balances of the extracellular and intracellular metabolites is given by Eq. 6.1 and Eq. 6.2, respectively:

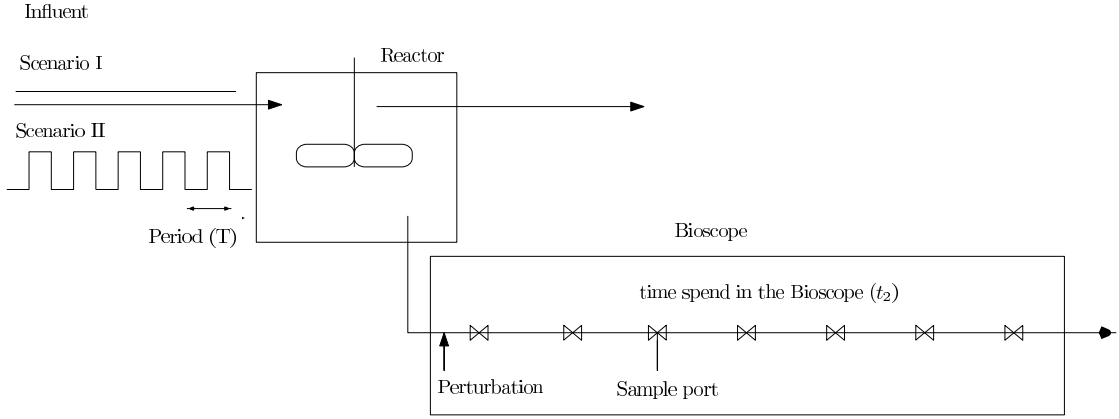


Figure 6.1: Schematic representation of the reactor configuration. Scenario I: the Bioscope is connected in series with a chemostat. Scenario II : the Bioscope is connected in series with a periodically operated reactor (period T). In order to collect cells with the same initial intracellular make-up prior to the perturbation a sample port should be opened at $[t + nT + t_2]$ s (with $n = 0,1,\dots$). Then, cells are collected that entered the Bioscope at $[t + nT]$ s, with t the time instant in the period, and that have spent t_2 s in the Bioscope.

$$\frac{dx_S}{dt} = D(x_S^0 - x_S) + f_{pulse} - \frac{x_X}{\rho_X} r_k \quad (6.1)$$

$$\frac{1}{x_x} \frac{dx_X x_M}{dt} = \frac{dx_M}{dt} + \frac{x_M}{x_X} \frac{dx_X}{dt} = \sum_k s_{jk} r_k \quad (6.2)$$

with x_M and x_S the concentration of an intracellular metabolite M and an extracellular metabolite S , respectively, s_{jk} the stoichiometric coefficient of metabolite j in reaction k , r_k the rate of reaction k , ρ_X the specific weight of biomass and x_X the biomass concentration, D the dilution rate, x_S^0 the concentration of an extracellular metabolite S in the feed, f_{pulse} a pulse of an extracellular metabolite S in the reactor, and μ the specific growth rate. Note that x_S is expressed in reactor volume whereas x_M is expressed in cell volume. The term μx_M in the mass balances of the intracellular metabolites represents the dilution effect due to growth.

6.3 Results

6.3.1 Experimental set-up

The proposed reactor configuration consists of a Bioscope connected in series with a periodically operated completely mixed bioreactor, with period T (Figure 6.1, scenario II). Such a periodic operation mode leads to a microbial system that shows limit cycle behaviour, which means that the cell's internal state periodically returns to be the same state. The new modus operandi of the Bioscope allows to selectively evaluate the perturbation behaviour of cells, that possess a particular internal state prior to the perturbation, which permits the sequential perturbation of cells with the same initial condition taken from a dynamically operated culture. Again the Bioscope principle is applied, *i.e.*, the culture from which the cells are taken is not perturbed.

In the proposed periodic operation of the bioreactor the varying conditions in the bioreactor generate a non-constant broth composition with respect to the concentrations of both extracellular and intracellular metabolites. If the sample ports would now sequentially, continuously be opened one by one, cells would be collected with a different intracellular make-up prior to the perturbation. Hence, the proposed modus operandi of the Bioscope must aim at collecting only those cells that are characterised by the same intracellular make-up prior to the perturbation. For this a control scheme to open and close the sample ports has to be applied. This has been developed below on the basis of a simulation study.

6.3.2 Simulation study

To illustrate the principle and possibilities of the proposed reactor configuration a simulation study has been performed. As mentioned above, during the periodic operation of the bioreactor, *i.e.*, when the transient behaviour has faded out, the metabolite trajectories enter a limit cycle, as illustrated in Figure 6.2 for the metabolites G6P, PEP and GLC_e in a 3D space. After every period the trajectory ends up in the same point, characterised by its intracellular metabolomic and enzymatic make-up.

The proposed method is thus able to yield, every period, the same well defined metabolomic and proteomic state. However, given the dynamic operation of the bioreactor, this state is different from the steady-state mode operation. This periodic 'initial condition'

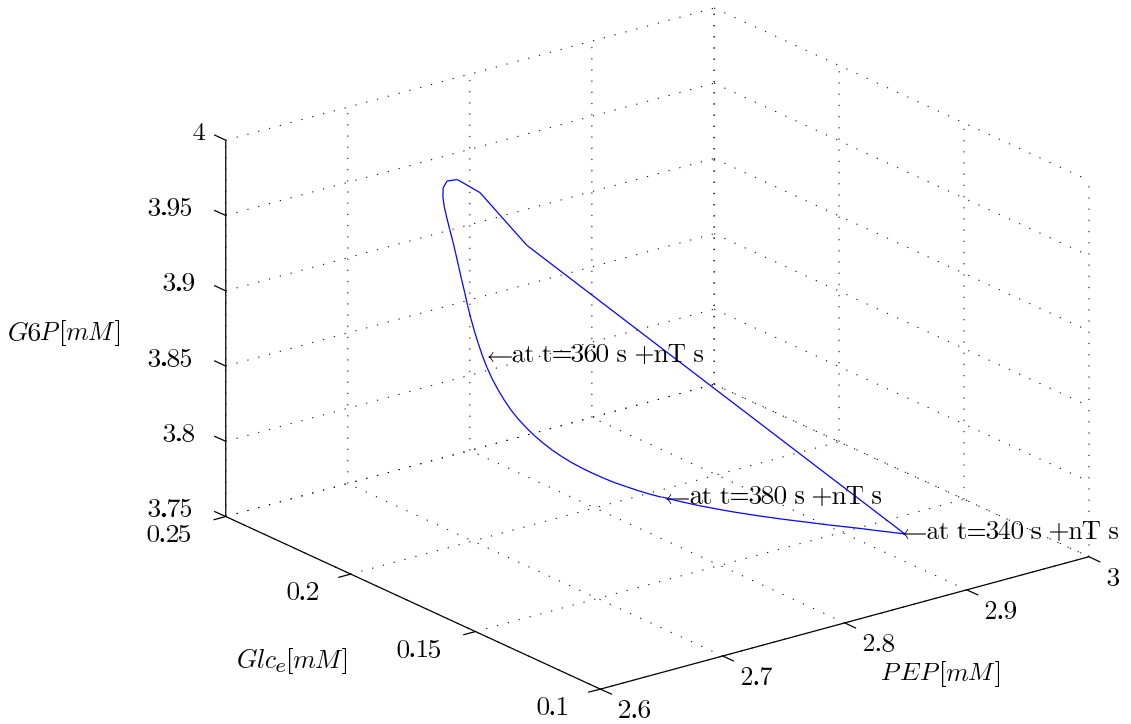


Figure 6.2: the limit cycle during the periodic operation in the reactor represented in the phase space of phosphoenolpyruvate (PEP), glucose-6-phosphate (G6P), and extracellular glucose (Glc_e).

at the inlet of the Bioscope, together with the constant perturbation applied to it, leads to a periodic variation of the extracellular and intracellular metabolite concentrations as the broth moves along the Bioscope's plug flow (Figure 6.3). Figure 6.3 depicts a schematic overview of the behaviour of the intracellular and extracellular metabolites both in the periodically operated bioreactor and in the Bioscope. In the perfectly mixed reactor, a substrate pulse is given every 60 s resulting in a periodic system.

The sample ports of the Bioscope should be controlled in such a way that only cells that had the same intracellular make-up prior to the perturbation, are collected, *i.e.*, the sample port should instantaneously open and close at $[t + nT + t_2]$ s with $n=0,1,\dots$ until enough sample is collected for the analyses to be performed. Then, cells are collected that entered the Bioscope at $[t + nT]$ s, with t the time instant in the period, and that have spent t_2 s in the Bioscope.

Figure 6.4 qualitatively depicts the trajectories in the phase space when the initial culture is perturbed at different instants during the period of the limit cycle in the bioreactor. Thus, by performing perturbation experiments at several time instants of the limit cycle, different responses can be gathered. The sample time in the Bioscope and the perturbing agent are additional degrees of freedom for this experimental set-up.

6.4 Discussion

The proposed configuration allows the execution of multiple perturbation experiments, even when the initial culture is subject to periodic conditions, via the controlled collection of samples. This operation implies that the initial reason to be of the Bioscope is preserved. Such an equipment is thus practical to study, through perturbation experiments, the complex metabolomic and proteomic interactions in periodically operated cultures for elucidating the mechanisms underlying the altered yields. In addition, only small quantities of the perturbing agent have to be used, which is a major advantage especially when the use of labelled substrates is imperative [212].

Both the sampling time in the Bioscope and the initial state of the cells to be collected prior to the perturbation are additional experimental degrees of freedom for the proposed configuration. Collecting samples in the Bioscope, with the same initial state prior to the perturbation, at distinct points in time thus allows the preservation of the culture and, consequently, should allow performing multiple experiments in a relatively short time, *e.g.*, perturbing a culture characterised by the same initial intracellular make-up, with different perturbing agents or perturbing a culture with the same perturbing agent starting with the same initial state.

6.5 Conclusions

An experimental set-up has been proposed with a view to gathering data to build and validate a dynamic metabolic model of periodically operated cultures. Such models can be useful for the optimisation of periodically operated cultures as they help to gain further insight in the complex metabolic interaction and they can predict the effect of altered conditions. This set-up allows performing multiple perturbation experiments without

perturbing the periodically operated culture itself. The perturbing agent, the sample time and the initial state, prior to the perturbation, are powerful degrees of freedom.

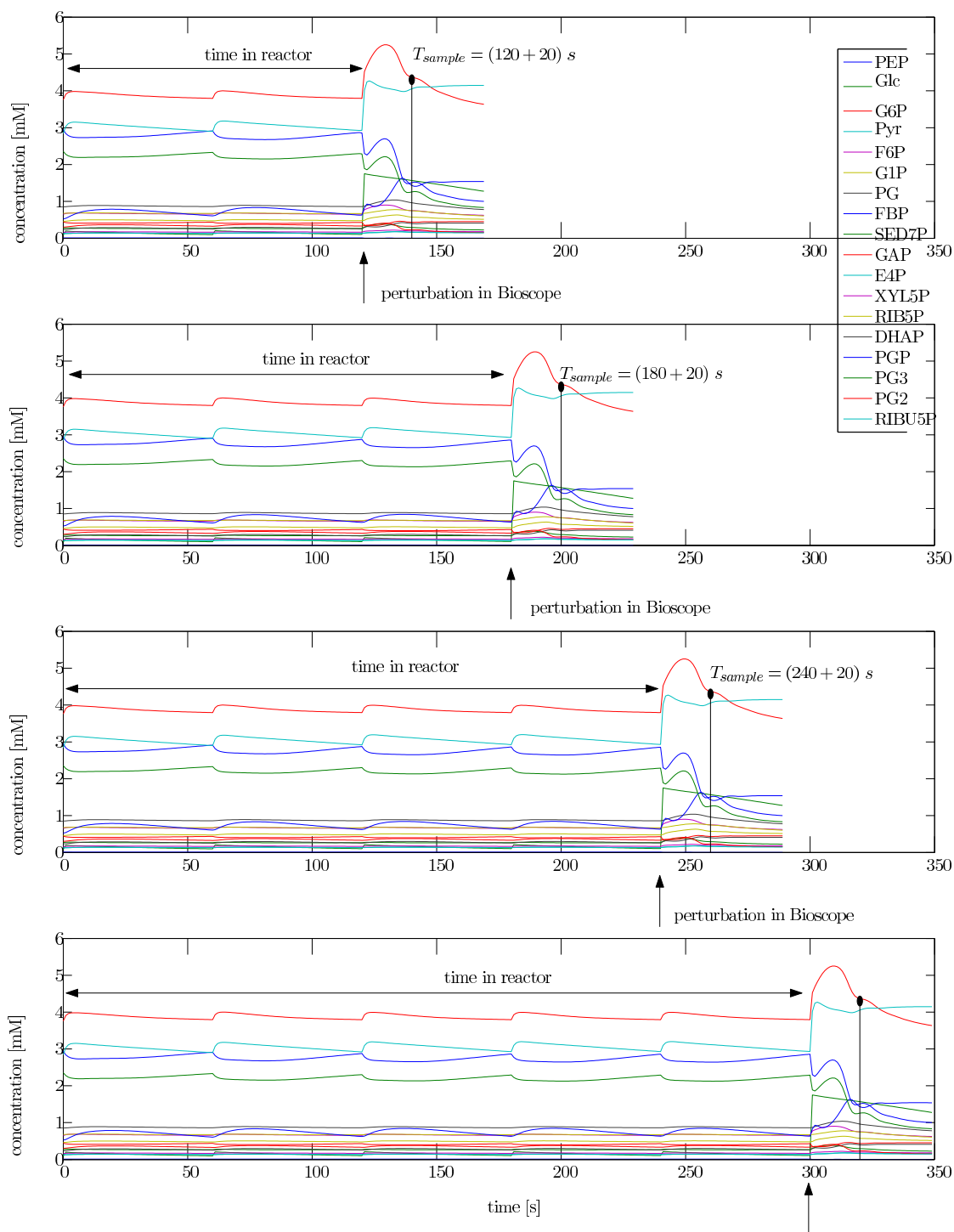


Figure 6.3: Evolution of the intracellular and extracellular metabolite concentrations of a parcel of broth in the periodically operated reactor and in the Bioscope for a parcel of broth that enters the Bioscope after 120 s (upper figure), 180 s, 240 s, and 300 s (lower figure) in function of time.

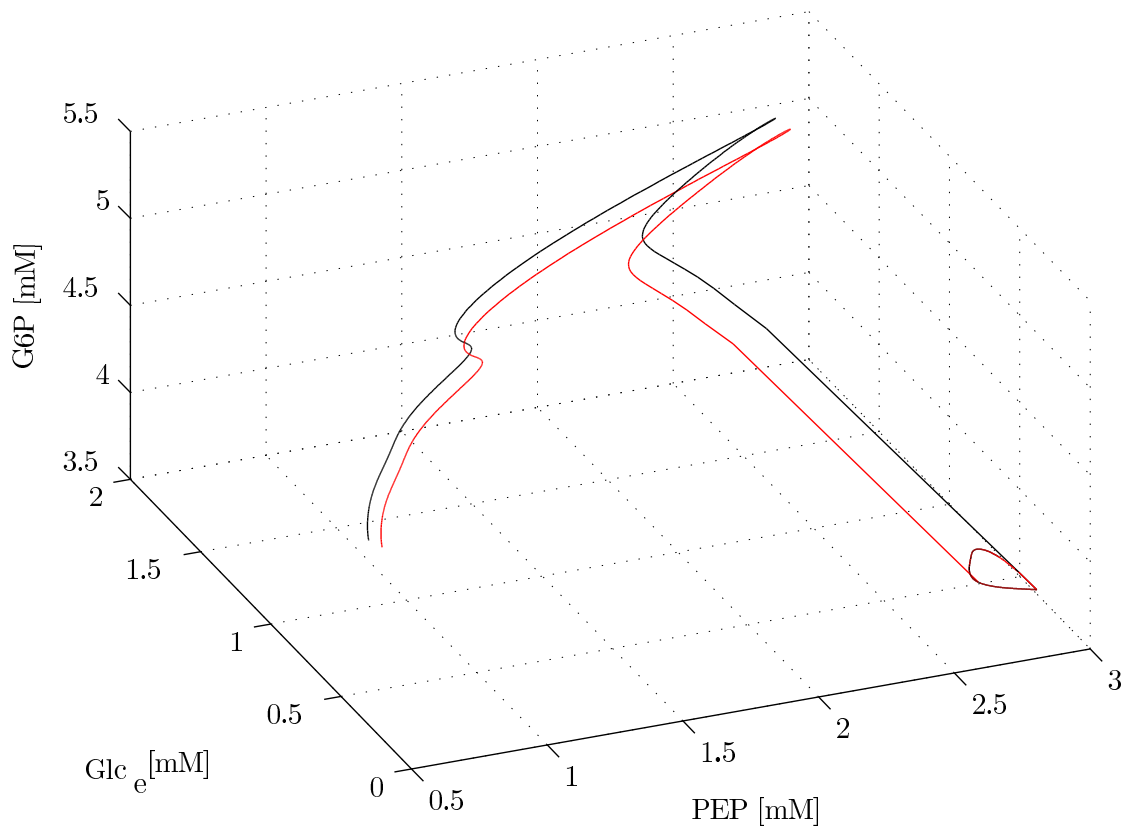


Figure 6.4: Response of the cells monitored in the Bioscope after perturbing cells originating from 2 (red and black) different instants in the limit cycle of the reactor

Chapter 7

Eulerian-Lagrangian description of a large-scale bioreactor: an averaging out approach

7.1 Introduction

The well-established chemical synthesis routes face for the production of many bulk and fine chemical more and more competition from industrial biotechnological alternatives. The development of such an industrially viable microbial process typically consists of 3 phases.

In a first phase a micro-organism is optimised itself, under laboratory conditions. Such an optimisation typically makes use of recombinant DNA techniques, functional genomics, as well as analytical [38, 181, 212] and mathematical techniques [30, 197, 198]. This mishmash of techniques has allowed a goal-oriented approach for genetic modification [11]. However, once the optimal producer has been constructed it has to be put to the test under large-scale conditions. This is due to the importance of mixing on both chemical and biological conversions [67, 173]. As a rule, scaling-up of fermentation processes from laboratory-scale to large-scale results in a significant reduction of biomass and product yields [16, 23, 202]. Finally, the product has to be purified and recovered from the fermentation broth (Downstream processing), using a wide range of physico-chemical techniques, *e.g.*, filtration, centrifugation, precipitation, ... [68]

The reduction of product yields in such a large-scale bioreactor has been attributed to imperfect mixing [139, 221]: zones exist in such large-scale bioreactors with ample substrate, in general in the surround of the inlet of the concentrated influent, with substrate depletion and with oxygen depletion or excess (laboratory-scale reactors of several litres on the contrary are typically considered as perfectly mixed). When an individual microorganism circulates through a large-scale reactor of various m^3 it is sequentially exposed to these different local conditions. The cellular response to these fast changing environmental conditions is thought to be an important cause of the observed yield reduction. In the past, many researchers have attempted to come up with a plausible explanation for the microbial response to these fast changing external conditions:

1. Hewitt *et al.* (2000) [76] postulate that the alternating production and reassimilation of organic acids like acetate, lactate, and formate due to overflow metabolism and mixed acid fermentation results in an ATP flux from biomass production towards the repetitive synthesis and degradation of certain organic acids,
2. Enfors *et al.* (2001) [53] put forward the intermittent transcriptional induction of genes, as a consequence of the rapidly changing environment in large-scale bioreactors. However as the synthesis of proteins, including folding, takes up to one hour, the rapidly changing induction and relaxation does not result in a net synthesis of proteins, and
3. Onyeaka *et al.* (2003) [131] point out pH fluctuations as a possible cause, as this will influence the proton motive force and consequently the generation of ATP.

Hence, a thorough understanding of the microbial response to the large-scale conditions would be useful for the optimisation of such processes [174]. Consequently, methods to more detailedly describe both the biophase and the physico-chemical processes in such large-scale bioreactors have been developed, *e.g.*, computational fluid dynamics models find acceptance both in industry and academia [15, 20, 62, 127] to study in many application domains the impact of spatiotemporal heterogeneity, *i.e.*, imperfect mixing, on the overall process performance.

In comparison with chemical applications, the system under study gets even more complex for biological applications since the behaviour of an individual micro-organism is also determined by its intracellular make-up, which is determined by what the micro-organism has observed over time [67, 96, 97]. Hence, a Lagrangian description, *i.e.*, following the cell's path through the reactor, is essential to take this history effect, a key element in unravelling the causes of the observed yield reduction, into account.

However, describing the biophase in a Lagrangian way is computationally demanding since a set of differential equations is linked to every micro-organism [96]. Solving this highly nonlinear system, a result of the set of intracellular balance equations and the exchange terms accounting for the transport of metabolites in and out of the microbial cell, is not trivial. To deal with the latter, Lapin *et al.* (2006) [97] opted to describe the system using an Euler-Lagrange formulation combined with a fractional-step method to allow for a stable, accurate, and numerically efficient solution of the underlying equations. This method requires however that for their three-dimensional simulation of a stirred-tank bioreactor 10^5 cells had to be tracked. Since, each of the finite volumes had to be populated with a sufficient number of microbial cells to minimise the effect of statistical error on the accuracy of the solution.

Considering however that the overall picture is merely the result of all individual micro-organisms, it may thus be conceivable that only a limited number of particles has to be tracked in order to obtain a good idea of the consumption and production of substrates and products throughout the large-scale bioreactor. Then the dynamics of the overall system can be captured by locally averaging out the behaviour of this limited number of particles over the whole population. The present contribution therefore focusses on the methods to implement such an approach taking into account the spatiotemporal heterogeneity which is characteristic for such large-scale bioreactors. Special attention will be devoted to the pitfalls using such a technique and how these were dealt with.

7.2 Materials and methods

7.2.1 Computational fluid dynamics

Reactor specifications and numerical techniques

Simulations of the flow of fluids and of the microbial conversions in a 30 m³ fermentor, stirred by four impellers, have been performed. The configuration of the fermentor is given in Table 7.1 and it is schematically depicted in Figure 7.1.

Governing flow equations

The derivation of the governing flow equations is taken from Fluent (2003) [1] and Tannehil and Anderson (1997) [178]. The flow of fluids can be described by the Navier-Stokes equations. Applying the conservation law for mass to a fluid passing through an infinitesimal control volume, yields following partial differential equation:

$$\frac{\partial}{\partial t}(\rho) + \frac{\partial}{\partial x_i}(\rho v_i) = 0 \quad (7.1)$$

with ρ the physical density of the fluid and x_i and v_i the position and velocity vector, respectively. The first term on the LHS of this equation represents the rate of increase in density in the control volume and the second term represents the rate of mass flux passing through the control surface per unit volume.

Applying the conservation law for momentum to an arbitrary control volume, yields following partial differential equation:

$$\frac{\partial}{\partial t}(\rho v_j) + \frac{\partial}{\partial x_i}(\rho v_i v_j) = \rho g + \frac{\partial}{\partial x_i}(\sigma_{ij}) \quad (7.2)$$

The two terms on the LHS of Eq. 7.2 represent the rate of increase of momentum per unit volume in the control volume and the net momentum flux in the control volume, respectively. The two terms on the RHS of Eq. 7.2 represent the gravitational force per unit of volume and the surface force per unit of volume, respectively. The components of the total stress tensor σ_{ij} are external stresses and shear stresses, which are represented by:

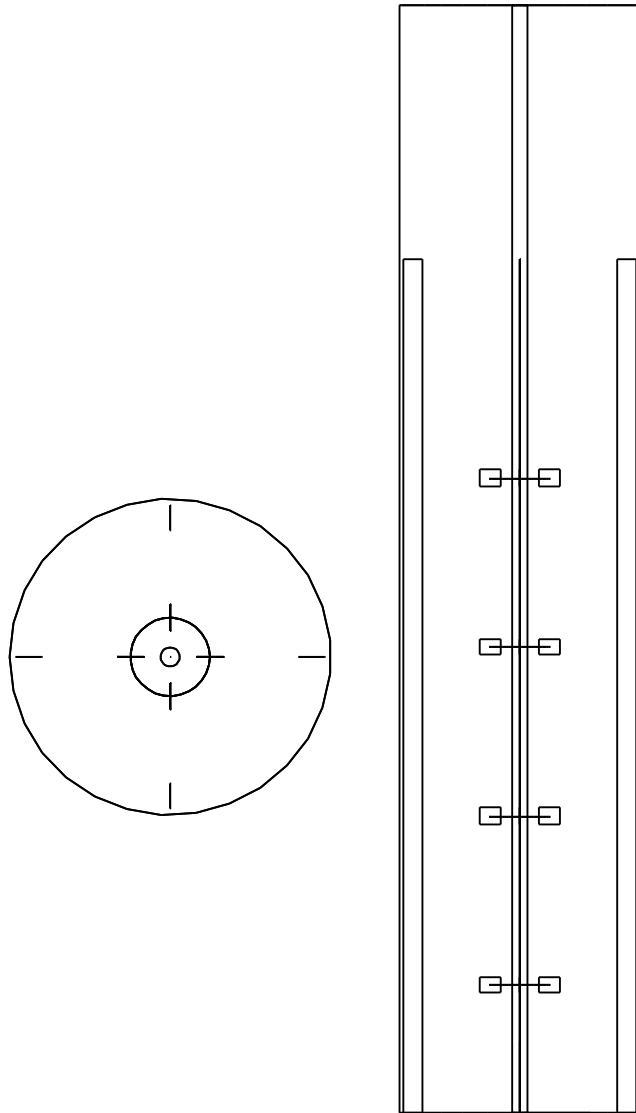


Figure 7.1: Top view and side view of the studied fermentor

Table 7.1: Fermentor configuration

| | Relative size [D] | Size [m] |
|-----------------------------|-------------------|----------|
| Fermentor diameter | 1 | 2.09 |
| Fermentor height | 4.59 | 9.60 |
| Liquid height | 2.99 | 6.25 |
| Baffle width | 0.08 | 0.17 |
| Spacing baffle-wall | 0.02 | 0.03 |
| Baffle height | 3.54 | 7.40 |
| Sparger diameter | 0.43 | 0.90 |
| Shaft diameter | 0.06 | 0.125 |
| Impeller diameter | 0.33 | 0.70 |
| Impeller blade height | 0.08 | 0.17 |
| Impeller spacing | 0.70 | 1.46 |
| Lower impeller spacing | 0.53 | 1.12 |
| Distance sparger - impeller | 0.82 | 0.58 |

$$\sigma_{ij} = -p\delta_{ij} + \tau_{ij} \quad (7.3)$$

where the pressure force, the first term on the RHS of Eq. 7.3, acts only normal to the surface of the control volume. The Kronecker-delta is defined as:

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (7.4)$$

which gives the pressure power a normal component. The viscous stress tensor is given by:

$$\tau_{ij} = \mu \left(\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) - \underbrace{\frac{2}{3}\delta_{ij} \frac{\partial v_k}{\partial x_k}}_{=0} \right) \quad (7.5)$$

with μ the molecular viscosity. The first two terms on the RHS of Eq. 7.5 represent the strain rate and the third represents dilatation. For incompressible fluids this term is equal to zero. This equation only holds for Newtonian fluids, characterised by a constant viscosity.

To resolve a turbulent flow by direct numerical simulation it is required that all relevant length scales are resolved from the smallest eddies to scales of the order of physical dimension of the problem domain, that three-dimensional computations are performed, and that the time steps must be small enough so that the small-scale motion can be resolved in a time-accurate manner. Such computations are infeasible nowadays for most applications. Time-averaged Navier-Stokes equations are used instead.

The Reynolds-averaged Navier-Stokes (RANS) equations are obtained by decomposing the dependent variables f in the conservation equations into a time-mean \bar{f} and a time-fluctuating component f' :

$$\bar{f} = \frac{1}{\Delta t} \int_{t_0}^{t_0+\Delta t} f dt \quad (7.6)$$

The time interval Δt should be chosen in such a way that its large with respect to the time constant of random fluctuations, associated with turbulence, and small in comparison with slow variations in the flow field associated with ordinary unsteady flows. The state variables in the Navier-Stokes equations are now decomposed in:

$$\begin{aligned} v_i &= \bar{v}_i + v'_i \\ u_i &= \bar{u}_i + u'_i \end{aligned} \quad (7.7)$$

Substitution in the Navier-Stokes equations and time-averaging the entire equations yields the RANS equations (note that $\overline{f'_i} = 0$). For a concise overview of the derivation, we refer to [178]. The RANS equations for mass, momentum, and chemical species for multiphase flows are given by [178]:

$$\frac{\partial}{\partial t} (\alpha_q \rho_q) + \frac{\partial}{\partial x_i} (\alpha_q \rho_q v_{i,q}) = 0 \quad (7.8)$$

$$\frac{\partial}{\partial t} (\alpha_q \rho_q v_{j,q}) + \frac{\partial}{\partial x_i} (\alpha_q \rho_q v_{i,q} v_{j,q}) = -\alpha_q \frac{\partial p}{\partial x_j} + \frac{\partial}{\partial x_i} (\alpha_q \tau_{ij,q}) + \alpha_q \rho_q g \quad (7.9)$$

$$\frac{\partial}{\partial t} (\alpha_q \rho_q \varphi_q^m) + \frac{\partial}{\partial x_i} \left(\alpha_q \rho_q v_{i,q} \varphi_q^m - \alpha_q \Gamma_q^m \frac{\partial \varphi_q^m}{\partial x_i} \right) = +S_q^m \quad (7.10)$$

For convenience, the bars indicating the time-averaged values were and will be omitted. With α_q the volume fraction, ρ_q the physical density, φ_q^m the concentration of scalar m ,

$v_{i,q}$ the velocity and $v'_{i,q}$ the fluctuations about the average velocity of phase q . The subscript q refers to the phase, *i.e.*, gas (g) or liquid (l), and the superscript m refers to the chemical species under consideration, *e.g.*, glucose. τ_q is the stress tensor of phase q , p is the total pressure, and g is the gravitational acceleration. Γ_q^m and S_q^m are the diffusion coefficient and the source term of chemical species m in phase q , respectively:

$$\Gamma_q^m = - \left(\mu + \frac{\mu_t}{\sigma_\varphi} \right) \quad (7.11)$$

$$S_q^m = f(\varphi_q^m) \quad (7.12)$$

with μ_t the turbulent viscosity and σ_φ the Schmidt number for chemical species in water assumed to be constant and equal to 0.7 [1, 59]. The formulae to calculate S_q^m will be discussed in Section 7.3. The global mass conservation is given by:

$$\sum_{q=1}^n \alpha_q = 1 \quad (7.13)$$

The viscous stress tensor $\tau_{ij,q}$ is now given by:

$$\tau_{ij,q} = \mu \left(\left(\frac{\partial v_{i,q}}{\partial x_j} + \frac{\partial v_{j,q}}{\partial x_i} \right) - \frac{2}{3} \delta_{ij} \underbrace{\frac{\partial v_{k,q}}{\partial x_{k,q}}}_{=0} \right) - \overline{\rho_q v'_{i,q} v'_{j,q}} \quad (7.14)$$

The additional term is called the Reynolds tension and represents the apparent surface gradients that are a consequence of turbulent motion. Boussinesq proposed to relate the average velocity gradient with the Reynolds tension using the average turbulent viscosity or eddy viscosity μ_t :

$$\overline{\rho_q v'_{i,q} v'_{j,q}} = \mu_t \left(\left(\frac{\partial v_{i,q}}{\partial x_j} + \frac{\partial v_{j,q}}{\partial x_i} \right) \right) - \frac{2}{3} \delta_{ij} \rho_q k \quad (7.15)$$

with k the turbulent kinetic energy.

Turbulence model

To determine μ_t , the standard $k - \epsilon$ turbulence model was used for simulating turbulent flows in the present study. The governing equations for the turbulent kinetic energy k and the turbulent kinetic energy dissipation rate ϵ were solved for both phases. Standard values for the parameters of the $k - \epsilon$ model were used in the present study (Table 7.2).

$$\begin{aligned} \frac{\partial}{\partial t} (\alpha_q \rho_q k_q) + \frac{\partial}{\partial x_i} (\alpha_q \rho_q v_{i,q} k_q) &= \frac{\partial}{\partial x_j} \left(\alpha_q \left(\mu + \frac{\mu_t}{\sigma_k} \right) \frac{\partial k_q}{\partial x_j} \right) \\ &\quad - \underbrace{\alpha_q \rho_q v'_{i,q} v'_{j,q} \frac{\partial v_{j,q}}{\partial x_i}}_{G_{k,q}} + \underbrace{\alpha_q \beta g \frac{\mu_t}{\sigma_s} \frac{\partial \phi}{\partial x_i}}_{G_{b,q}} - \alpha_q \rho_q \epsilon_q \end{aligned} \quad (7.16)$$

$$\begin{aligned} \frac{\partial}{\partial t} (\alpha_q \rho_q \epsilon_q) + \frac{\partial}{\partial x_i} (\alpha_q \rho_q v_{i,q} \epsilon_q) &= \frac{\partial}{\partial x_j} \left(\alpha_q \left(\mu + \frac{\mu_t}{\sigma_{\epsilon_q}} \right) \frac{\partial \epsilon_q}{\partial x_j} \right) \\ &\quad + \alpha_q \frac{\epsilon}{k_q} (C_{1\epsilon_q} G_{k,q} - C_{2\epsilon_q} \rho_q \epsilon_q) - \rho_k c_{2\epsilon} \frac{\epsilon_q^2}{k_q} \end{aligned} \quad (7.17)$$

in these equations G_k represents the generation of turbulence kinetic energy due to the mean velocity gradients and G_b is the generation of turbulence kinetic energy due to buoyancy, with ϕ the solid mass fraction. The turbulent viscosity μ_t is computed by combining k_q and ϵ_q as follows [178]:

$$\mu_t = \rho C_\mu \frac{k_q^2}{\epsilon_q} \quad (7.18)$$

where σ_ϵ and σ_k express the turbulent diffusive transport of the scalars k and ϵ .

Grid

The resulting set of nonlinear partial differential equations that describes the system can, in general, not be solved analytically. Numerical solution, using the finite volume method,

Table 7.2: Parameter values of the $k - \epsilon$ turbulence model [1]

| c_μ | $c_{1\epsilon}$ | $c_{2\epsilon}$ | $c_{3\epsilon}$ | σ_ϵ | σ_k |
|---------|-----------------|-----------------|-----------------|-------------------|------------|
| 0.09 | 1.14 | 1.93 | 0.80 | 1.30 | 1.00 |

requires the discretisation of this set of nonlinear partial differential equations in space and time. For this reason, the solution domain is subdivided into a finite number of small control volumes (cells) by a grid. This discretisation results in a set of coupled algebraic equations. The governing equations to be solved are strongly coupled and nonlinear and therefore they must be solved by an iterative method. The used computational grid consisted of 138144 control volumes. The solution independence on the grid size has been verified. To this end, the grid was refined with a gradient adaptation approach, which refined the grid in regions with high gradients, to 236982 control volumes.

Trajectory

Micro-organisms were modelled using a Lagrangian approach. The trajectories of the particles are determined based on a force balance [1]:

$$\rho_p \frac{du_p}{dt} = F_D (u_i - u_p) + g (\rho - \rho_p) \quad (7.19)$$

with u_i and u_p the fluid and particle velocity, respectively, ρ and ρ_p the fluid and particle density, respectively. $F_D = \frac{18\mu C_D Re}{24d_p^2}$ is the drag force, d_p is the particle diameter, $Re = \rho d_p \frac{|u_i - u_p|}{\mu}$ is the particle's Reynolds number. $C_D = a_1 + \frac{a_2}{Re} + \frac{a_3}{Re^2}$ is the drag coefficient is calculated according to [119], with a_1 , a_2 , and a_3 numerical constants for smooth spherical particles.

In the applied stochastic tracking approach, the turbulent dispersion is taken into account by integrating the trajectories using the instantaneous fluid velocity: $u = \bar{u} + u'$, with $u' = \zeta \sqrt{2k/3}$ and ζ a normally distributed random number [1]. The discrete random walk model assumes that a particle interacts with a succession of discrete fluid phase turbulent eddies, which are characterised by velocity fluctuations [1]. The approach presented by Lapin *et al.* (2004) [96] is however preferable, since the discrete random walk model shows a tendency for particles to concentrate in low-turbulence regions [1, 96].

Impeller

The earliest attempts to numerically simulate the flow field in mechanically agitated reactors applied impeller boundary conditions to model the impeller [71, 145, 146]. In this approach, the impeller is not physically modelled but represented either in terms of

boundary conditions at the surface of the volume swept by it or in terms of source terms distributed throughout its volume.

Over time several general approaches have been reported in the literature on explicit simulation of the flow field in agitated reactors. The main generalised approaches are the multiple reference frame (MRF) approach [22] and the sliding mesh (SM) approach [120, 176]. The first approach involves steady-state computations and produces a time-averaged flow field. The second approach involves transient computations to produce a time-accurate flow field. Both these approaches subdivide the computational domain into two non-overlapping regions, one region surrounding the impeller and the other representing the rest of the vessel (Figure 7.2).

The MRF approach first simulates the flow field for the inner domain surrounding the impeller in a reference frame rotating with the impeller. The resulting flow field on the interface separating the inner and outer regions then serves as boundary condition to simulate the flow field in the outer domain in an inertial frame of reference (laboratory frame of reference). This results in improved boundary conditions, which are sequentially to be applied for the simulation of the flow field in the inner domain. The procedure is repeated until a suitable numerical convergence criterion is achieved. The procedure involves steady-state approximation of essential periodic flow, correction for the relative motion and azimuthal averaging are required before using the flow field at the interface as boundary condition for solution of the flow field in the outer domain [22].

The SM approach involves transient computations to produce a time-accurate flow field. The flow equations in the inner domain are now written in a laboratory reference frame whilst it is the grid in this domain that is allowed to rotate. The rotation of the grid results however in acceleration terms which are completely equivalent to the body forces arising in non-inertial frames. The grid in the outer domain is stationary. The two regions are implicitly coupled at the interface via a SM algorithm which takes the relative motion between the two domains and performs the required interpolation into account [120].

However, whereas the MRF approach has as undesirable feature that species in the inner domain are transported relative to the impeller motion, the sliding mesh approach is, due its transient computation, computationally demanding. Therefore, it was opted to apply boundary conditions to model the impeller. The momentum source distributed

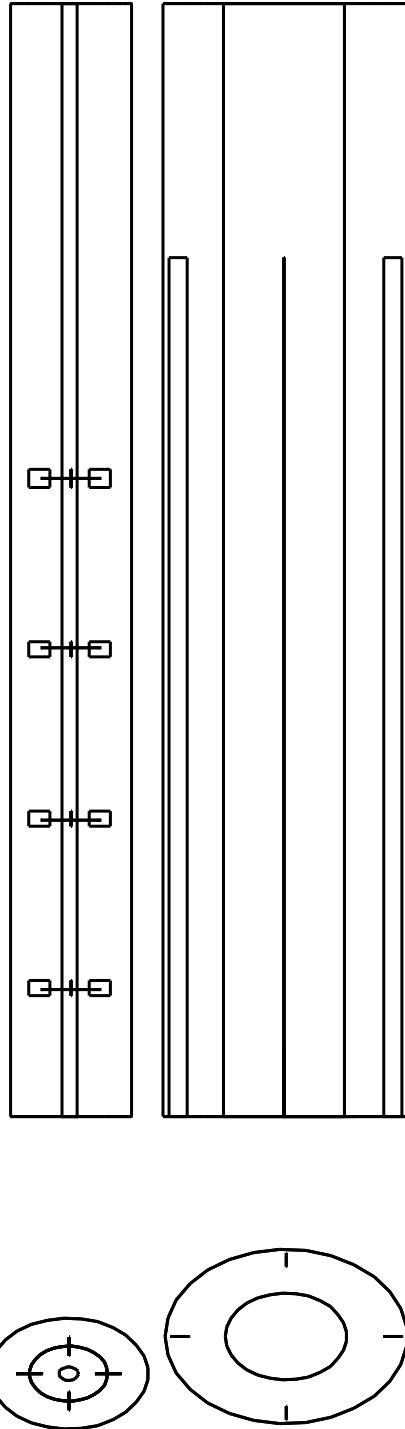


Figure 7.2: Side view and top view of an inner (left figure) and outer region (right figure)

throughout the impeller is given by:

$$M_X = \rho \frac{X}{R} \left(1 - \left(\frac{Y - Y_0}{D} \right)^2 \right) V_R \quad (7.20)$$

with R the impeller radius, X the x-distance from the center of the rotation axis, $Y - Y_0$ the y-distance from the center of the impeller, V_R the impeller tip speed, D the impeller blade thickness, and ρ the physical density. The impeller rotational speed was 115 RPM.

7.3 Results and discussion

7.3.1 Framework

The developed approach to structuredly and segregatedly describe the biophase is depicted in Figure 7.3. Herein represent the blue blocks operations that are performed by Fluent[®] (ANSYS[®], USA) typically linked to processes related to the hydrodynamics and the transport of particles and chemical species in the large-scale bioreactor and the yellow blocks represent operations that are performed by Matlab[®] (The Mathworks, USA) related to the calculation of the microbial systems and the calculation of the substrate and or space dependent functions for the chemical species' source term, S_q^m .

The rationale of the method presented here is that information on the transport of metabolites in and out the microbial cells of microbial cells in the neighbourhood of each other, both in terms substrate concentration data and position data may be used to calculate the average/overall transport of metabolites in and out of the microbial cells at a position, in this way only a limited number of particles has to be tracked in order to obtain a good idea of the transport of metabolites in and out of the microbial cells throughout the large-scale bioreactor, since the dynamics of the overall system can be captured by locally averaging out the behaviour of this limited number of particles over the population.

In this way it is possible to separate the solving of processes of completely different nature, *i.e.*, processes related to the biological system and those related to transport of fluids, particles, and substrates, which typically require a different solver. Operations related to the hydrodynamics and the transport of particles and chemical species in the

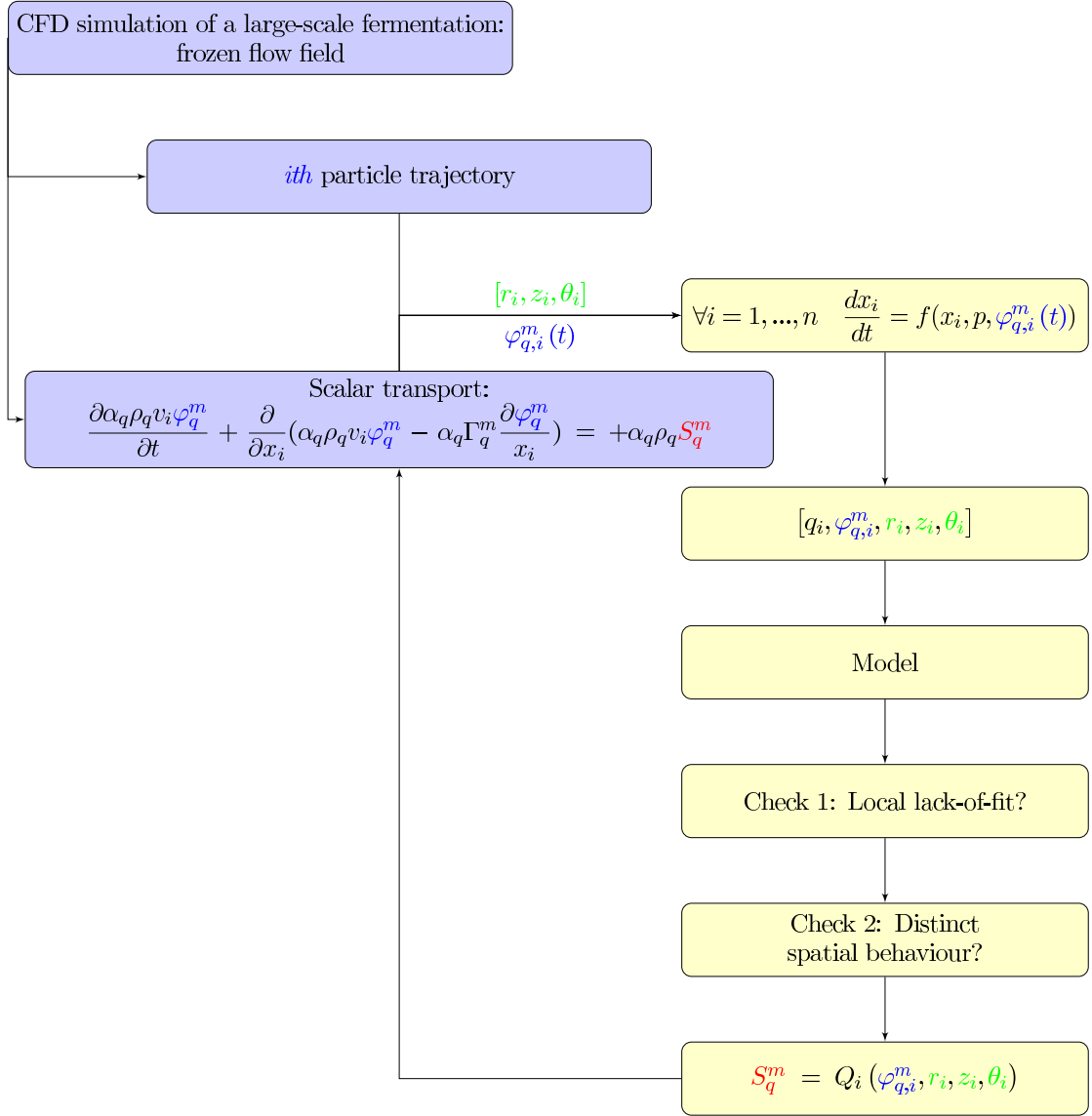


Figure 7.3: Schematic overview of the applied approach in order to structuredly and segregatedly describe the biophase. The blue blocks represent operations that are performed by Fluent[®] (ANSYS[®], USA) and the yellow blocks represent operations that are performed by Matlab[®] (The Mathworks, USA).

large-scale bioreactor were calculated by Fluent[®] (ANSYS[®], USA) and operations related to the calculation of the microbial systems and the calculation of the substrate and/or space dependent functions for the chemical species' source term S_q^m , were performed by Matlab[®] (The Mathworks).

Another advantage of the proposed method is that the number of particles that has to be tracked can be significantly reduced, since every finite volume in the calculation grid does not have to be populated with a number of particles in order to avoid statistical error on the solution. Even for a three-dimensional simulation of a stirred bioreactor, instead of the two-dimensional simulation presented here, the to be tracked number of particles is expected to be much smaller than the 10^5 particles tracked by Lapin *et al.* (2006) [97].

Every time step of the unsteady simulation, the position of all particles, calculated by Eq. 7.19, and the there reigning environmental conditions, calculated by the scalar transport equation Eq. 7.10, were recorded. Every n time steps these data were used to calculate the microbial conversions in the bioreactor and to establish a correlation between these conversion rates and the substrate concentrations. A concise overview of the individual steps will be given below.

Microbial model

For every micro-organism i in the large-scale bioreactor the production and consumption rates of metabolite m , q_i^m , have been calculated by the cybernetic model of *Saccharomyces cerevisiae* of Jones and Kompala (1999) [83] (Eq. 7.21, Figure 7.4), using the extracellular concentration data $\varphi_{q,i}^m(t)$ collected along this particle's trajectory in the large-scale bioreactor, see also Figure 7.3:

$$\dot{x}_i = f(x_i, p, \varphi_{q,i}^m(t), u_i, v_i) \rightarrow q_i^m \quad (7.21)$$

With u and v the cybernetic variables that control enzyme synthesis and activity, respectively, p the model's parameters, and x_i the intracellular state variables linked to this particle, *e.g.*, the intracellular metabolites and enzymes. In order not to unnecessarily complicate things, only the processes glucose fermentation and glucose oxidation were

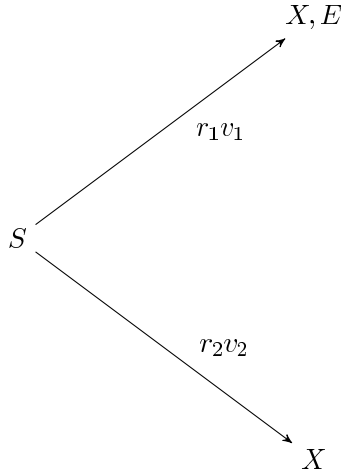


Figure 7.4: Schematic overview of the cybernetic model

considered, the rates of these process are given by r_1v_1 and r_2v_2 .

Additional checks have been implemented to verify whether the particles are still present in the calculation grid and to verify whether the integration of the microbial system has been successful, if this is not the case the data from this particle were discarded.

7.3.2 Approximate model

Subsequently, a correlation was established between the specific production and consumption rates q of micro-organisms and the substrate concentrations observed by micro-organisms:

$$q = f_1(\varphi_q^m) + \varepsilon \quad (7.22)$$

First, due to the typical nature of production and consumption rates in function of substrate concentration concentration data, *e.g.*, saturation for substrate concentration values $\varphi_q^m \gg K_{\varphi_q^m}$, with $K_{\varphi_q^m}$ the affinity constant, and the rapid changes in production and consumption rates in function of substrate concentration data for substrate concentration values $\varphi_q^m \ll K_{\varphi_q^m}$ the substrate concentration data were transformed using Eq. 7.23:

$$x_i = \log_{10} (\varphi_{q,i}^m + 1e - 4) + 4 \quad (7.23)$$

The substrate concentration range was then subdivided into a number of regions. For each of these regions the parameters b_1 , b_2 , x_0 , and a of the nonlinear function in Eq. 7.24 were identified, using a local nonlinear optimisation algorithm. As initial estimates for this nonlinear optimisation the parameter values that were calculated during the previous iteration were used or when no parameter values were available, these initial values were immediately identified from the data.

$$q = b_1 - \frac{b_2}{1 + \exp\left(-\frac{x-x_0}{a}\right)} \quad (7.24)$$

7.3.3 Quality check

The ability of these nonlinear functions to describe the conversion data as a function of substrate concentration has been verified. Special attention has been devoted to the pitfalls using the proposed technique. Therefore, it was verified whether local lack-of-fit occurred or distinct spatial behaviour was averaged out.

Local lack-of-fit

In an attempt to avoid local lack-of-fit, *i.e.*, substrate region for which the nonlinear approximation does not yield an adequate description of the data, the substrate space was subdivided into a number of subspaces. For each of these subspaces i , it was verified whether the mean and the variance of the residuals ε , with ε the difference between the calculated conversion rate, Eq. 7.21, and the conversion rate calculated by the approximate model, Eq. 7.24, were significantly different than the residuals of the rest of the population of particles j , Eqs. 7.25 and 7.26, respectively.

$$t = \frac{\bar{\varepsilon}_i + \bar{\varepsilon}_j}{\sqrt{\frac{\sigma_i^2}{n_i} + \frac{\sigma_j^2}{n_j}}} \leq t_{\frac{\alpha}{2}, df_i, df_j} \quad (7.25)$$

$$f_{\frac{\alpha}{2}, df_i, df_j} \leq f = \frac{\sigma_{\varepsilon_i}^2}{\sigma_{\varepsilon_j}^2} \leq f_{1-\frac{\alpha}{2}, df_i, df_j} \quad (7.26)$$

If so, this data set was treated separately, *i.e.*, separate functions were established to link the consumption and production rates to the environmental conditions.

$$q = f_2(\varphi_q^m) + \varepsilon \text{ if } \varphi_q^m \in SS_\varphi^i \quad (7.27)$$

Undesired averaging out distinct spatial behaviour

In an attempt to avoid averaging out distinct behaviour in function of space coordinates, for each of the substrate subspaces the data were clustered according to space coordinates, using a k-means clustering method [86]. It was verified whether the mean and the variance of the residuals of these clusters were significantly different than the residuals of the rest of the population of particles. If so, this data set was treated separately.

The appropriate approximate model is then used as source term, S_q^m , in the scalar transport equation Eq. 7.10, to calculate the substrate field in the large-scale bioreactor.

7.3.4 Case study

To illustrate the approach presented above a large-scale bioreactor was simulated two-dimensionally, where the biophase was described by the cybernetic model of Jones and Kompala (1999) [83]. Only glucose was considered as carbon-source and oxygen was assumed to be abundantly present. Glucose was continuously and constantly added near the air-liquid interface of the large-scale bioreactor. The initial biomass concentration was about 15 g/L. The resulting substrate field of glucose in the large-scale bioreactors is depicted in Figure 7.5, when the final biomass concentration was about 22 g/L.

About 700 particles were introduced in the bioreactor. Increasing this number further did not contribute to a more adequate description of the system. For each of these particles the transport of metabolites in and out of the microbial cell was calculated, and multiple nonlinear functions were used to relate the substrate concentration data to the microbial conversion data. Typical results of this procedure are depicted in Figure 7.6. Due to the much slower dynamics of enzyme synthesis and degradation in comparison with the

observed environmental variations, the cellular enzyme levels are approximately constant in the population of tracked micro-organisms, which explains the very similar behaviour of the individual microbial cells when exposed to the same substrate concentration.

Additional attention has to be devoted to concentration data that were not represented during the averaging out procedure when calculating the average/overall transport of metabolites in and out of the microbial cells. Such problems are however typically expected to occur near the concentrated influent inlet location, where due to the large substrate gradients an underrepresentation of micro-organisms is found. However, considering the typical nature of production and consumption rates in function of substrate concentration data, *e.g.*, saturation for substrate concentrations $\varphi_{q,i}^m \gg K_{\varphi_q^m}$ this problem can partially be avoided by using approximate functions which typically reflect such saturation phenomena.

The microbial model used in this study is not that complex, certainly not compared to the microbial model Lapin *et al.* (2006) [97] used in their study. However, based on an extensive literature search for dynamic metabolic models, none of them appeared to be really suitable to properly describe the biophase in large-scale bioreactors. Next to the use of a microbial model that adequately describes the biophase in large-scale bioreactors, the incorporation of the third dimension would be beneficial to obtain a more realistic description of the studied large-scale bioreactor.

It should be clear that even the proposed method to describe the biophase in a Lagrangian way is computationally demanding. Therefore, to speed up calculations two strategies have been followed. Firstly, the simulation jobs were distributed over a computer cluster, consisting of Intel Pentium 4s (CPU 3GHz, 1GB ram), *i.e.*, the tracked population was subdivided into 6 groups containing nearly an equal number of particles. These 6 sub-experiments, solving the microbial model for every particle of that group, were allocated to nodes in the calculation grid. Secondly, to minimise the time for the nonlinear optimisations the parameter values, calculated during the previous iteration, were used as initial estimates or if no such values were available these initial values were estimated immediately from the data. Both measures reduced the calculation time. It is expected that further subdividing of the tasks to be performed can further reduce the time needed for the calculations.

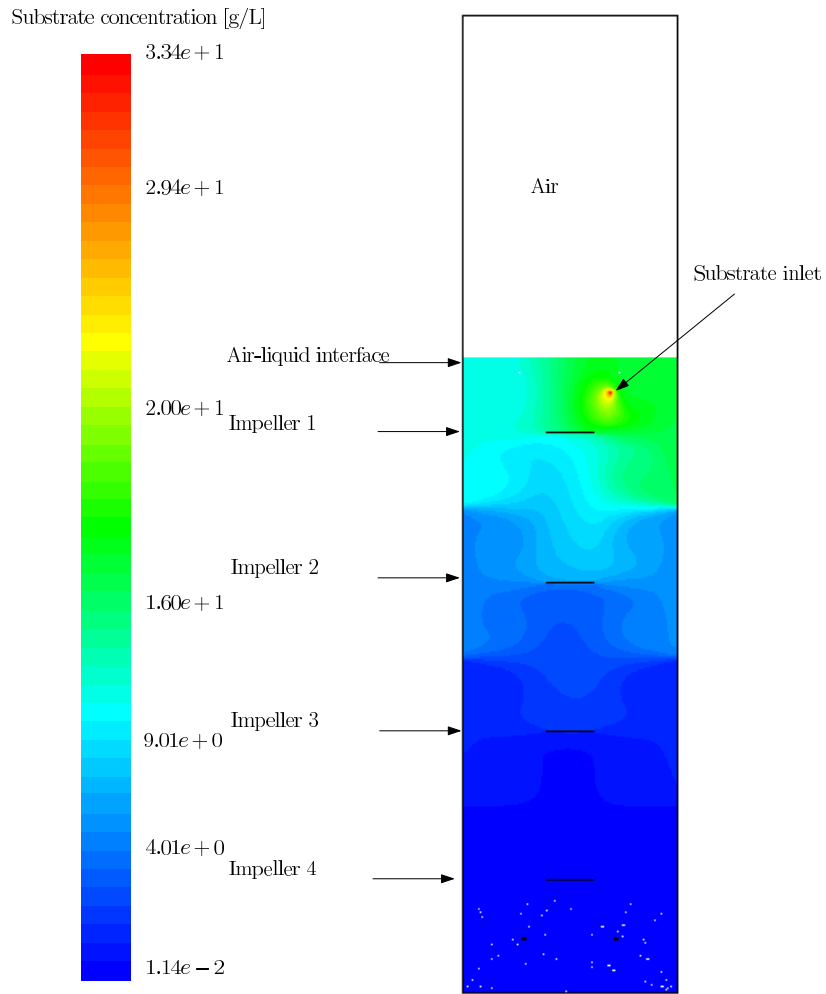


Figure 7.5: Substrate field in the large-scale bioreactor (g glucose/L)

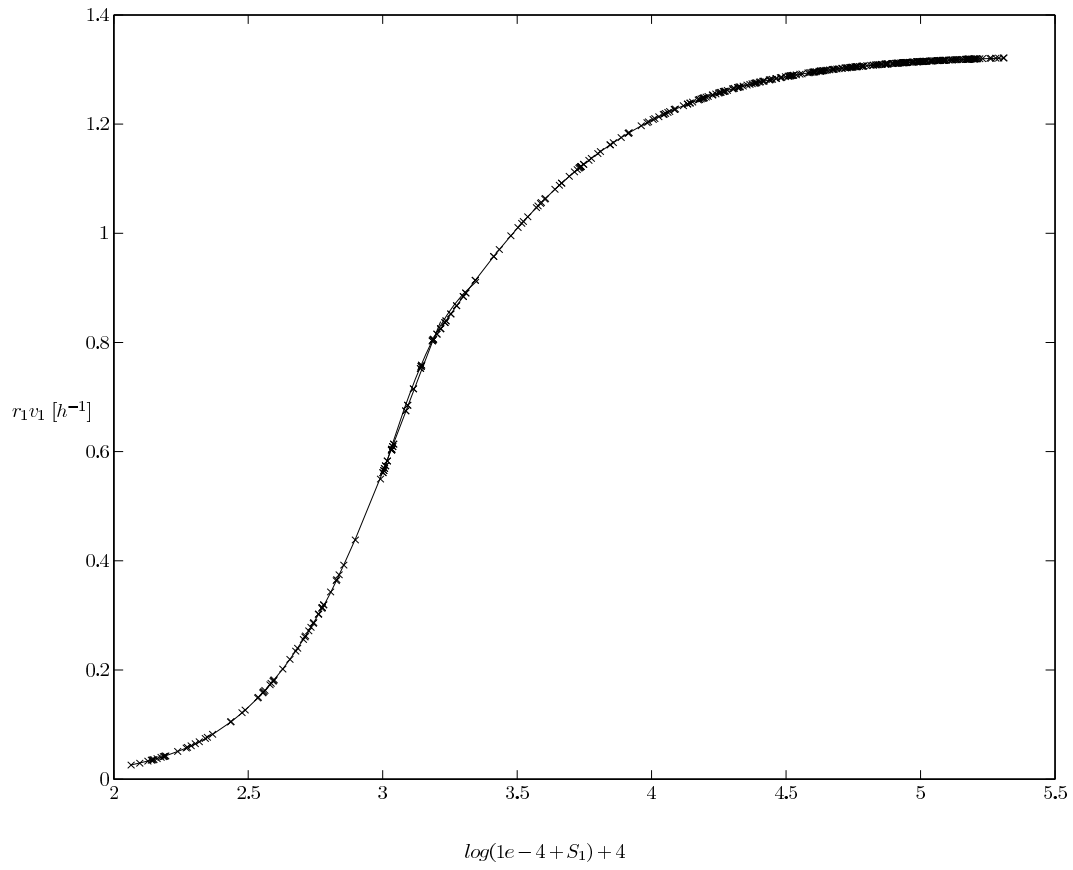


Figure 7.6: The calculated rate for $r_1 v_1$ by the cybernetic model for all tracked particles (x) in function of transformed S_1 concentration (g glucose/L) and the approximate model

7.4 Conclusions

An averaging out approach has been developed to describe a large-scale bioreactor in an Eulerian-Lagrangian way. A necessity, as the stochastic nature of particle transport in combination with the fast metabolic response to the observed fast changing environmental conditions will result in a heterogeneous population of cells.

However, solving the resulting highly nonlinear system, a result of the set of intracellular balance equations and the exchange terms accounting for the transport of metabolites in and out of the cell, is not trivial. To deal with the latter problem an averaging out approach has been developed. By averaging out the behaviour of a limited number of cells over the whole population it is already possible to get a good idea of conversions throughout the large-scale bioreactor.

Chapter 8

Design of a scaled-down reactor using computational fluid dynamics

8.1 Introduction

In view of optimising the performance of a large-scale culture, *i.e.*, scaling-up, a thorough understanding of the mechanisms responsible for the deteriorated performance of large-scale cultures in comparison with laboratory-scale cultures and of their relative importance is useful. Therefore scaling-down is a useful approach [174]: by mimicking on a laboratory-scale the large-scale conditions, this approach attempts to anticipate the outcome on a large-scale. For instance, in such laboratory-scale simulation the spatiotemporal heterogeneity, which is characteristic for large-scale reactors [139, 221], is mimicked. This can be done in a single reactor [33, 121] or by constructing a loop of two completely mixed reactors or of a completely mixed reactor and a plug flow reactor [63, 76, 131] or by means of a tubular loop reactor [133]. Mann *et al.* (1995) [110] introduced the network of zones reactor, which comprises a large number of interconnected compartments. Whether such a set-up will be popular in practise is doubtful, due to its relative complexity. Finally, Delvigne *et al.* (2005, 2006) [44, 45] came up with a still different approach, using mixing models and circulation models, but though their set-up is able to mimic macroscopic variables as mixing time and circulation time, the environmental conditions encountered by micro-organisms in their scaled-down reactor significantly differed from those found in the studied large-scale bioreactor.

It is thus still unclear how representative the currently available scaled-down reactors are

for the large-scale reactors. Therefore, this study aims at designing a more representative scaled-down reactor, which better mimics the characteristic conditions in a large-scale bioreactor, by making use of substrate concentration data observed by micro-organisms during their journey through a large-scale bioreactor. These data were obtained during a computational fluid dynamics simulation of a large-scale bioreactor.

In addition, the proposed scaled-down reactor attempts to compromise between minimising the scaled-down reactor's complexity and obtaining a realistic imitation of the large-scale conditions. Therefore, it was opted to study a controlled system consisting of two continuous stirred-tank reactors in a loop, as such a set-up still allows to exploit the naturally occurring phenomenon of blending distinct streams in large-scale bioreactors. The approach is schematically depicted in Figure 8.1 and consists of 3 steps:

1. An *in silico* large-scale fermentation is performed using a computational fluid dynamics model. For each particle, whilst it circulates throughout the large-scale reactor, its position $[r, z, \theta]$ and the substrate concentration reigning at that position φ_q^m are recorded.
2. These time series data $[\varphi_q^m, r, z, \theta]$ of stochastic nature, are subsequently used for the identification of a hidden Markov model (HMM) that captures the typical substrate concentration dynamics. This model will be used to steer the scaled-down reactor, composed of a dynamically operated reactor system consisting of two continuous stirred-tank reactors.
3. A proper controller is designed to impose the substrate concentration time series calculated by the HMM on the two-reactor system. Finally, the concentration time series data observed by a particle in the scaled-down reactor are compared with these collected in the large-scale reactor.

A concise overview of these 3 steps will be given below.

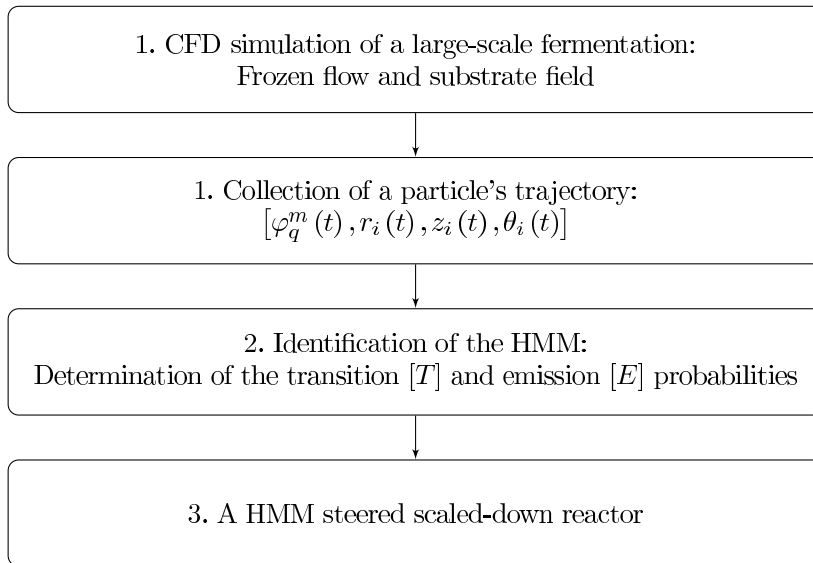


Figure 8.1: Schematic overview of the proposed approach to design a scaled-down reactor

8.2 Materials and methods

8.2.1 Computational fluid dynamics

The reader is referred to Subsection 7.2.1 for a description of the studied large-scale bioreactor and for details on the computational fluid dynamics model.

8.2.2 Hidden Markov model

The second step is the identification of a hidden Markov model (HMM) (Figure 8.1) that captures the typical concentration dynamics observed by micro-organisms in the large-scale bioreactor. HMMs are especially known for their application in temporal pattern recognition. Such as speech, musical score following, and bioinformatics [50, 141] because of their ability to capture information from series of data. For this reason a HMM will also be used in this study.

A hidden Markov model is schematically depicted in Figure 8.2. Such a model typically consists of a finite set of states $\{A, B\}$. Transitions between these states are governed by a set of probabilities called transition probabilities (Eq. 8.1). With t_{ij} the probability of going to state j from state i , here with $i, j \in \{A, B\}$. In a particular state an outcome or

observation is generated according to the associated probability distribution (Eq. 8.2). With e_{ij} the probability of emitting observation j from state i with $i \in \{A, B\}$ and $j \in \{1, 2, 3, 4\}$. If only the outcomes are visible to an observer, the states are 'hidden' to the outside. Hence, the name hidden Markov model.

$$T = \begin{pmatrix} t_{AA} & t_{AB} \\ t_{BA} & t_{BB} \end{pmatrix} \quad (8.1)$$

$$E = \begin{pmatrix} e_{A1} & e_{A2} & e_{A3} & e_{A4} \\ e_{B1} & e_{B2} & e_{B3} & e_{B4} \end{pmatrix} \quad (8.2)$$

When state i is visited, an observation token is emitted from the state's emission probability density distribution. Then according to the state's transition probability density distribution one goes to the next state. The model thus generates two series of information. For example, the following series have been generated by the HMM depicted in Figure 8.2:

$$\begin{array}{cccccccccccc} B & \rightarrow & A & \rightarrow & B & \rightarrow & B & \rightarrow & A & \rightarrow & A & \rightarrow & A & \rightarrow & A & \rightarrow & A & \rightarrow & A \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ 2 & & 3 & & 4 & & 2 & & 1 & & 2 & & 2 & & 1 & & 1 & & 3 \end{array}$$

One series, *BABBAAAAAA*, is the underlying state path, as transiting from one state to another. The other, *2342122113*, is the observed sequence, each observation being emitted from one state in the state path [192].

8.2.3 Controller scheme

This HMM will then be used to steer the conditions in the scaled-down reactor to mimic those of a large-scale reactor (Figure 8.1). To impose the dynamic behaviour of the response variables, as determined by the hidden Markov model, to the scaled-down reactor system (schematically depicted in Figure 8.3) a control scheme has been applied to the system. To this end, a PID controller in combination with an adaptive state feedback controller was used.

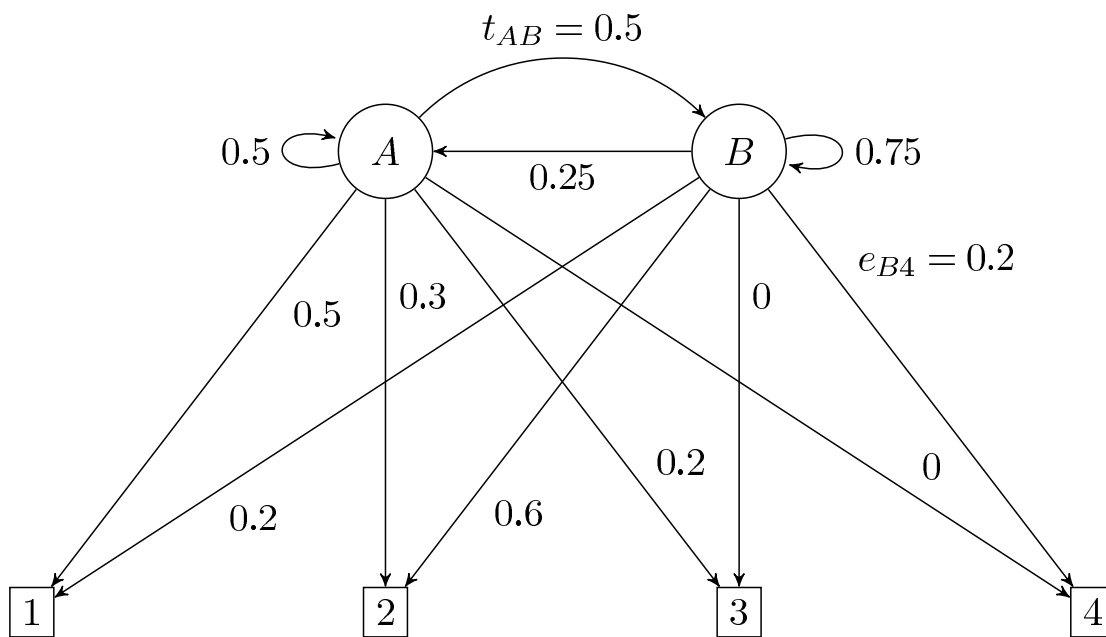


Figure 8.2: A schematic view of a HMM with states A and B and observations 1, 2, 3, and 4. The transition probabilities are represented by arrows between the states, *e.g.*, t_{AB} , and the emission probabilities are represented by arrows between the states and the observations, *e.g.*, e_{B4} .

The values of the response variables α_{ij} (the fraction of broth to be transferred from reactor i to reactor j), S_i (the concentration of the substrate S in reactor i), and V_{TOT} (the cumulative volume of reactor 1 and 2) will be controlled by adjusting the values of the manipulated variables Q_{ij} (the flow from reactor i to reactor j), Q_W (the total waste flow, which is the sum of $Q_{W1} = Q_W \frac{V_1}{V_1+V_2}$ and $Q_{W2} = Q_W \frac{V_2}{V_1+V_2}$), Q_{D_i} (a flow without any substrate to dilute reactor i), and $pulse_i$ (a substrate pulse in reactor i).

PID controllers attempt to correct the error ε between the value of a (measured) response variable x and its desired set point x^{SP} :

$$\varepsilon = \begin{bmatrix} \max(S_1 - S_1^{SP}, 0) \\ \max(S_2 - S_2^{SP}, 0) \\ \max(S_1^{SP} - S_1, 0) \\ \max(S_2^{SP} - S_2, 0) \\ V_{TOT}^{SP} - V_{TOT} \\ \alpha_{12}^{SP} - \alpha_{12} \\ \alpha_{21}^{SP} - \alpha_{21} \end{bmatrix} \quad (8.3)$$

by taking the appropriate control actions, v :

$$v = \begin{bmatrix} v_{Q_{D1}} \\ v_{Q_{D2}} \\ v_{pulse1} \\ v_{pulse2} \\ v_{Q_w} \\ v_{Q_{12}} \\ v_{Q_{21}} \end{bmatrix} \quad (8.4)$$

with,

$$v_i(t) = K_p \left(\underbrace{\varepsilon_i(t)}_P + \underbrace{\int \frac{\varepsilon_i(t)}{\tau_I} dt}_I + \underbrace{\tau_D \frac{d\varepsilon_i(t)}{dt}}_D \right) \quad (8.5)$$

with α_{ij}^{SP} in Eq. 8.3 given by:

$$\alpha_{ij} = \frac{t_{kl}}{t_{kl} + t_{km}} \quad (8.6)$$

with k the state at the previous time step in reactor i , and l and m the states at the present time step in reactor j and i , respectively.

A Proportional-Integral-Derivative (PID) controller is a generic classical feedback controller widely used in industrial control systems. The PID controller calculation (algorithm) involves three tuning parameters: the proportional, the integral, and derivative values (Eq. 8.5) [47, 48]. However, due to the general nature of PID control, it does not guarantee optimal control of the system.

A full state feedback controller is on the contrary a modern controller. This controller is, *e.g.*, employed in feedback control system theory to place the closed-loop poles of a system in predetermined locations in the s-plane [48]. Placing poles is desirable because the location of the poles corresponds directly to the eigenvalues of the system, which control the characteristics of the response of the system. Under certain conditions (if the closed-loop input-output transfer function can be represented by a state space equation) it is possible to assign a value to the system's eigenvalues, which allows to design the dynamics of the system. To this end the nonlinear system was linearised around the work point. A full state feedback controller is a typical optimal controller in which both the deviation from the set point $x - x^{SP}$ and the control action u , necessary to achieve this set point, can be penalised in the objective function J through the matrices W_1 and W_2 , respectively:

$$J = \int \left((x - x^{SP})^T W_1 (x - x^{SP}) + u^T W_2 u \right) dt \quad (8.7)$$

The control action is then given by Eq. 8.8. The controller scheme is depicted in Figure 8.4.

$$u = v + K (x - x^{SP}) \quad (8.8)$$

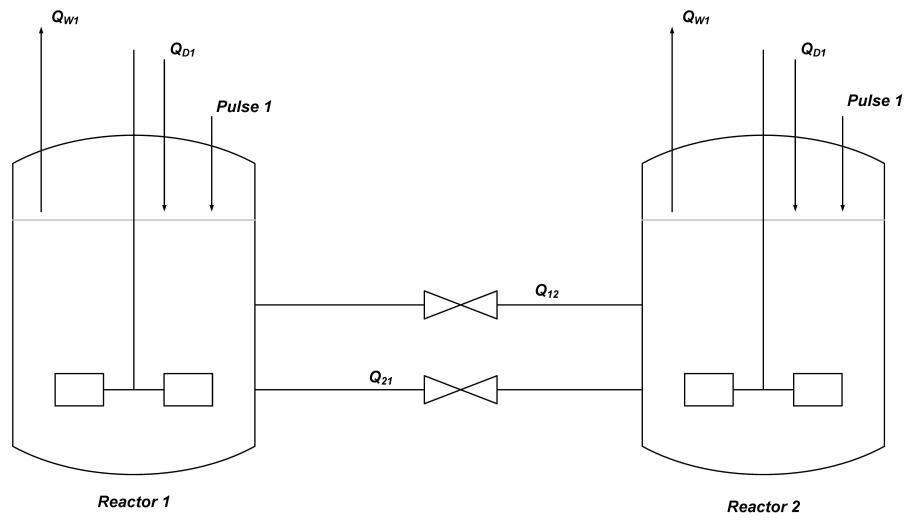


Figure 8.3: Overview of the scaled-down reactor set-up

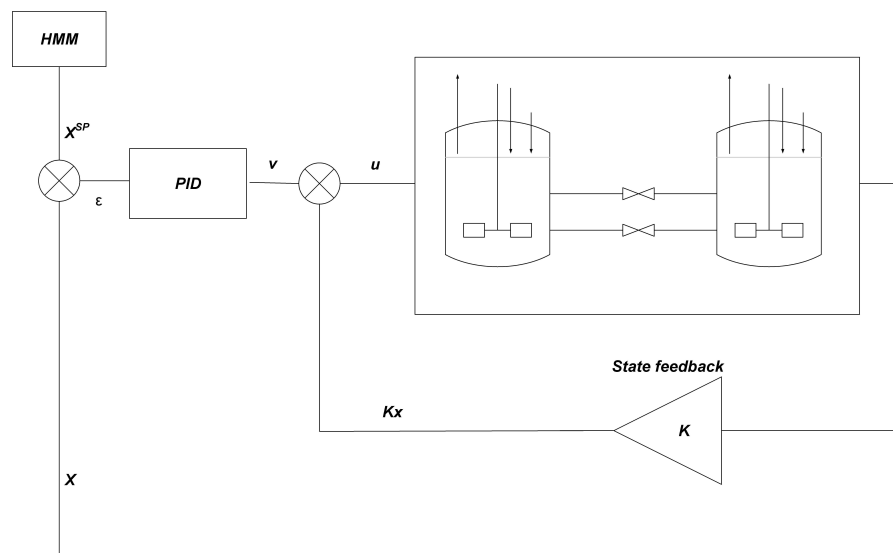


Figure 8.4: Overview of the overall control scheme, consisting of a PID controller and a state feedback control system

8.3 Results and discussion

8.3.1 Particle's trajectory

In order to get an idea about the environmental conditions micro-organisms are exposed to when travelling through a large-scale bioreactor, the substrate concentration data were collected for multiple micro-organisms along their path through the large-scale bioreactor described in Chapter 7. A micro-organism's path during its journey in the reactor is depicted in Figure 8.5. The substrate concentration data the particle encounters as a function of time are depicted as well.

8.3.2 Identification of a HMM

A typical concentration sequence encountered by a micro-organism in a non-ideally mixed bioreactor is depicted in Figure 8.5. Due to the stochastic nature of this sequence of concentration data, HMMs are typically suitable to describe such data [141]. Applying the terminology of HMMs to the case presented here, the observations are the substrate concentration data and the 'hidden' states are linked to zones in the large-scale bioreactor (Subsection 8.2.2).

As mentioned before, the scaling-down set-up presented here is a controlled system consisting of two continuous stirred-tank reactors in a loop. Such a set-up allows to exploit the naturally occurring phenomenon of blending distinct streams in large-scale bioreactors.

In the context of the scaled-down reactor presented here, these 'streams' are the broth remaining in reactor i and the broth to be transferred from reactor j to reactor i , when proceeding to the next time step. Δt is the time between two transitions. A direct consequence of the use of two completely mixed reactors is that no distinction can be made anymore between the micro-organisms contained in those 'streams' from the moment on those 'streams' have been blended, thus between micro-organisms coming from reactor j and micro-organisms that were already present in reactor i . Consequently the memory of the micro-organisms stays restricted to the preceding state. The choice for a first order hidden Markov model is then obvious, as for such a model the probability of a certain

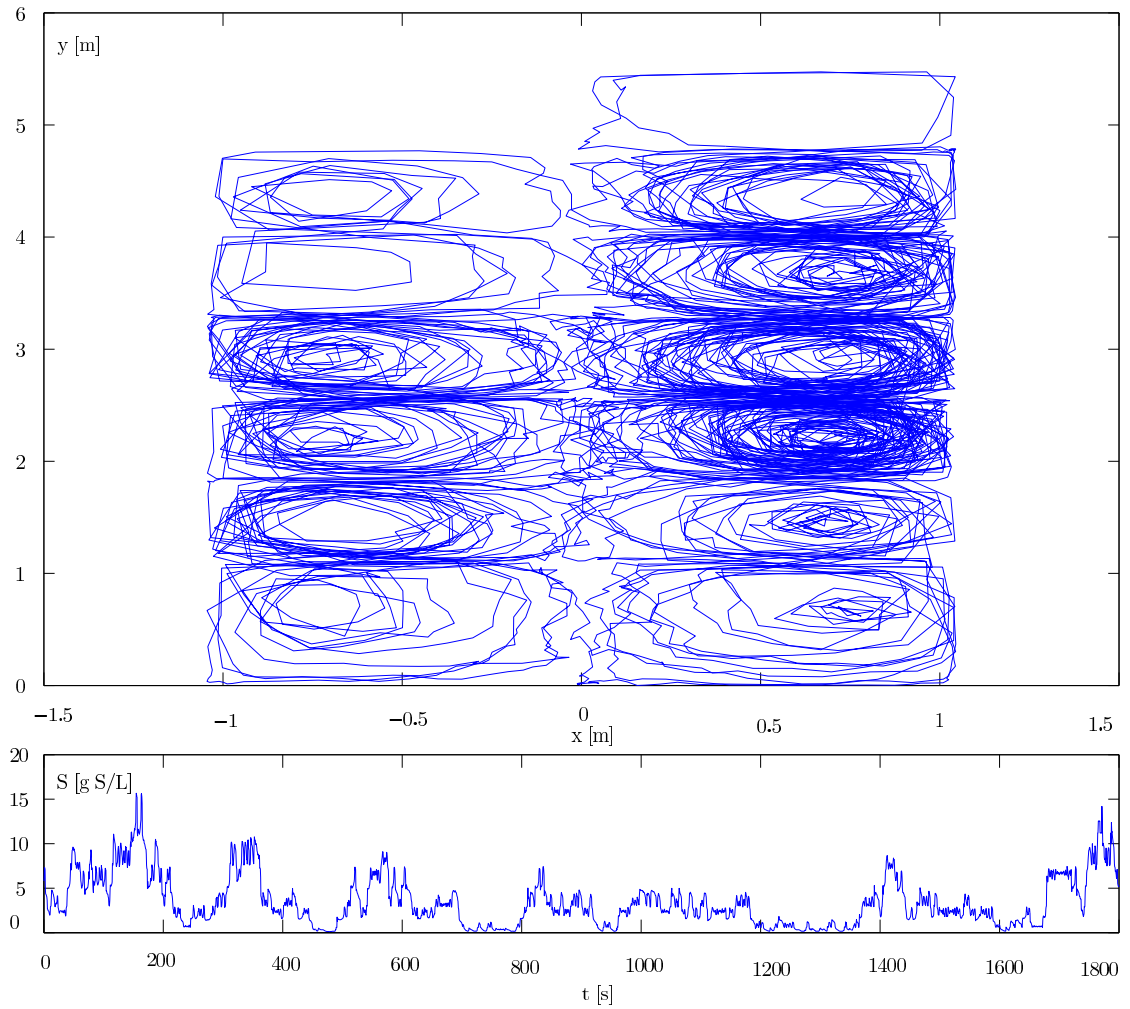


Figure 8.5: A micro-organism's trajectory in the large-scale bioreactor. Upper figure: trajectory of a micro-organism through the reactor. Lower figure: substrate concentration data the micro-organism encountered as a function of time.

state at time step $n + 1$ should only depend on the state of the previous time step n . However, then one has to be careful that all assumptions are satisfied,

$$P(\text{zone } i \mid \text{zone } j \mid \text{zone } k) = P(\text{zone } i \mid \text{zone } j \mid \text{zone } l) \quad (8.9)$$

i.e., the probability of transiting from zone k to zone j , and subsequently from zone j to zone i should be equal to the probability of transiting from zone l to zone j , and subsequently from zone j to zone i .

The determination of the states and observations of this first order hidden Markov model will be discussed below. The reactor space has to be subdivided into a number of zones, *e.g.*, Figure 8.7. These zones have been obtained by k-means clustering [86] according to both space coordinates and substrate concentrations, *i.e.*, locations in the neighbourhood of each other with similar substrate concentrations belong to a particular zone of the large-scale reactor.

Subsequently, for every micro-organism, the two data series (one series with positional and one with substrate concentration data), collected during the computational fluid dynamics simulation of the large-scale bioreactor (Figure 8.5) were transformed to the corresponding sequences in terms of states and observations. For instance, if a micro-organism is located in zone j at time $n\Delta t$, then position n in the state sequence becomes j ; if this micro-organism is located in zone k at time $(n + 1)\Delta t$, then position $n + 1$ in the state sequence becomes k . Akin, the observation sequence was built. For each zone 10 discrete substrate concentrations have been chosen, uniformly distributed over the zone's substrate range. Subsequently, each of the observed concentrations in the observation sequence was replaced by the number representing the most representative of these 10 substrate concentrations.

Based on the data collected along the micro-organism's trajectory through the large-scale bioreactor, *i.e.*, states/zones and observations/discrete concentrations, the transition and emission probability density distributions were determined.

The choice of the number of zones and the time between two transitions is nontrivial. Therefore, for every $\{C_i, \Delta t_j\}$ combination, with C_i the number of zones $\in \{10, 20, \dots, 100\}$ and Δt the time step in s between two transitions $\in \{1, 5, 10, 20\}$, a hidden Markov model

has been identified and the trajectories generated by these HMMs have been compared with the trajectories collected in the large-scale bioreactor.

The HMM that has been used to steer the scaled-down reactor was selected using following criteria: the average time of exposure to identified concentration ranges expressed in terms of percentage of total time (these concentration ranges were quite arbitrarily chosen to be $[0, \frac{g}{L}, 1, \frac{g}{L}]$, $[1, \frac{g}{L}, 2, \frac{g}{L}]$, ...), the average number of time steps a micro-organism is sequentially exposed to that concentration range, and for some of the candidate HMMs the trajectories themselves. The higher the similarity in terms of these criteria the more adequate the HMM is judged to mimic the substrate concentration data observed by micro-organisms in the large-scale bioreactor.

Finally, 70 zones were selected, which are depicted in Figure 8.7, and a time step of 10 s was retained. The results for the selected HMM are depicted in Figure 8.6. This figure shows that there is a reasonable agreement between the trajectory generated by this HMM and the trajectories truly observed in the large-scale bioreactor in terms of the aforementioned criteria. The selected HMM is thus considered to be able to describe the substrate concentration dynamics observed by micro-organisms in the large-scale bioreactor and can consequently be used to steer the scaled-down reactor.

8.3.3 A HMM driven scaled-down reactor

The selected HMM was used to steer the scaled-down reactor. Every Δt s the HMM generates for each bioreactor a new state and a new observation. The new elements of the state sequence and observation sequence, have to be transformed in terms of the response variables. For example, the fraction of broth to be transferred between the reactors is given by Eq. 8.6 and the substrate concentration set point in reactor i is given by:

$$S_i^{SP} = (e - 1) \frac{(S_{\max, state l} - S_{\min, state l})}{9} + S_{\min, state l} \quad (8.10)$$

with e the observation $e \in \{1, 2, \dots, 10\}$ generated by the HMM and $S_{\max, state l}$ and $S_{\min, state l}$ the maximal and minimal substrate concentration in zone/state l .

This hidden Markov model sets the desired set points for the response variables α_{ij} , S_i , and V_{TOT} . The evolution of the response variables S_1 , S_2 , α_{12} , and α_{21} and their set

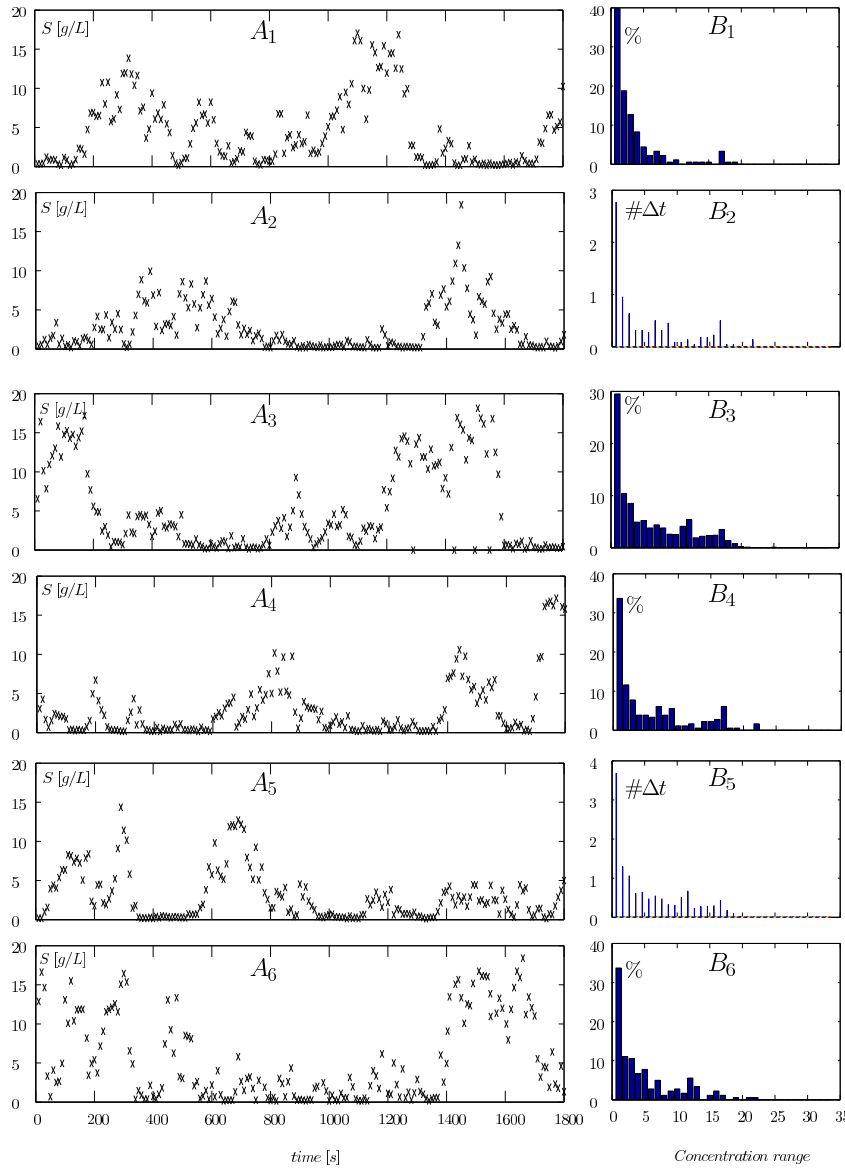


Figure 8.6: Comparison of the series of discrete substrate concentration data 'observed' by micro-organisms in the large-scale bioreactor and generated by the selected HMM: A₁ to A₅: Series of discrete substrate concentration data (×) observed by 5 micro-organisms in the large-scale bioreactor; A₆: Series of discrete substrate concentration data (×) generated by the selected HMM; B₁ and B₆: the time of exposure (in terms of percentage total time) of an organism to the identified substrate concentration ranges for the trajectory in subfigure A₁ and for the trajectory in subfigure A₆, respectively; B₃ and B₄: the average time of exposure (in terms of percentage total time) of an organism to the identified concentration ranges for series of discrete substrate concentration data in the large-scale bioreactor and generated by the selected HMM, respectively; B₂ and B₅: the average number of time steps -1 a micro-organism is sequentially exposed to the identified concentration ranges for series of discrete substrate concentration data in the large-scale bioreactor and generated by the HMM

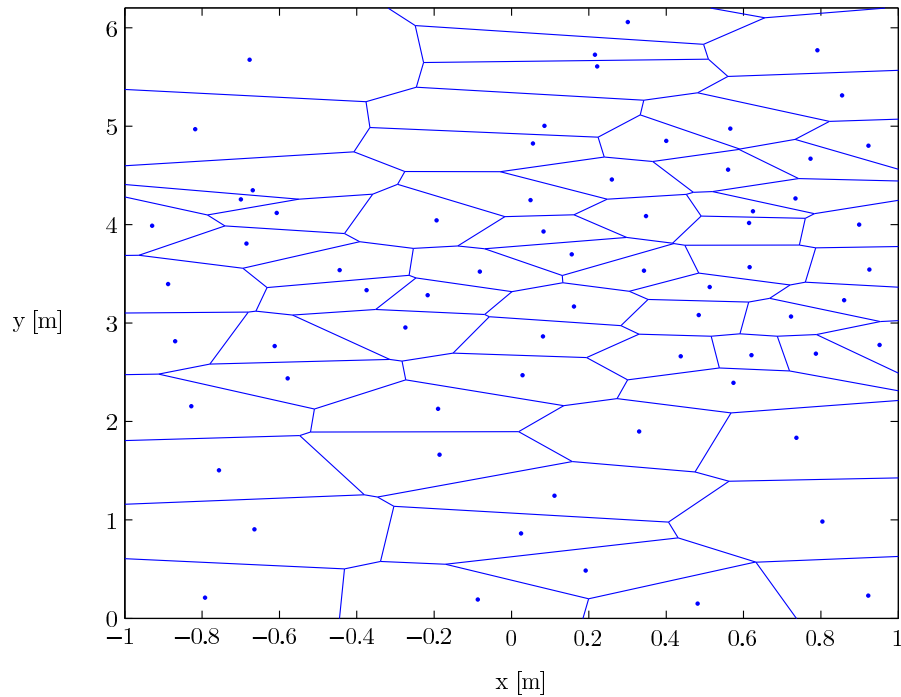


Figure 8.7: The identified zones in the large-scale bioreactor

points is given in Figure 8.8. To reflect more or less realistic conditions, constraints on the manipulated variables have been incorporated as well (maximal flow rate $Q_D 200 \frac{mL}{s}$). As shown, the response variables and their set points agree reasonably well. However, to properly assess the representativeness of the developed scaled-down reactor, it seems more correct to compare the trajectory from a micro-organism in the large-scale bioreactor (Figure 8.6) with one from a micro-organism in the scaled-down reactor (Figure 8.9). From these figures, one can conclude that the substrate concentrations observed by a micro-organism in the scaled-down reactor resemble those of a micro-organism in the *in silico* large-scale bioreactor. This is certainly the case when one compares the performance to other scaled-down reactors, *e.g.*, [45]. A more quantitative evaluation of this resemblance would be desirable. However, since a micro-organism is a highly nonlinear system it will be necessary to evaluate the response of this highly nonlinear system to the substrate concentration trajectories observed by micro-organisms in the large-scale bioreactor with those observed by micro-organisms in the scaled-down reactor. Other measures seem not fit for this aim.

The main advantage of the developed scaled-down reactor is thus that it tries to compro-

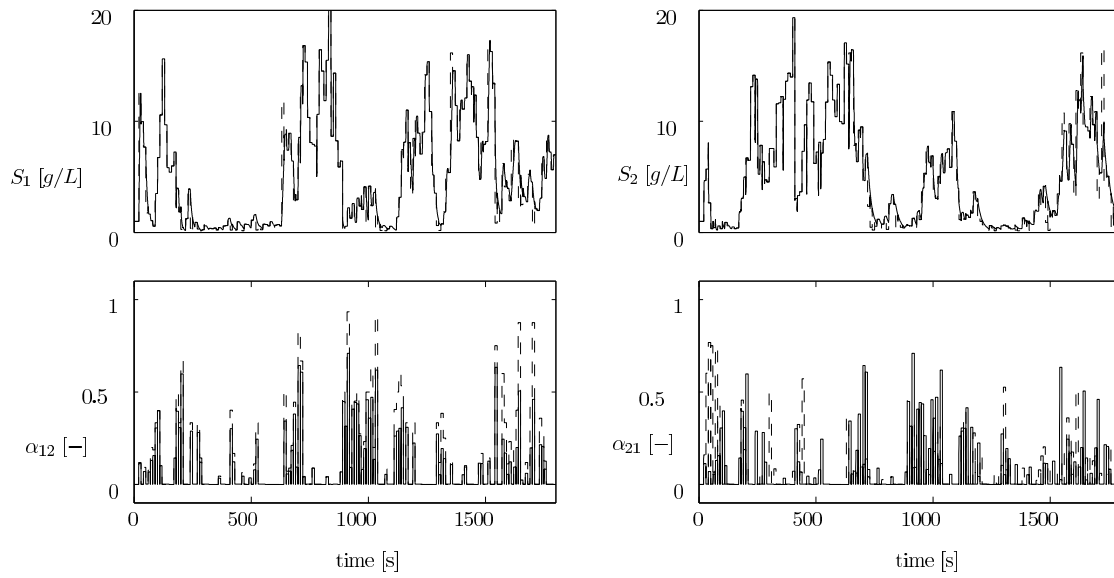


Figure 8.8: The scaled-down reactor: evolution of response variables (-) and their set points (- -).

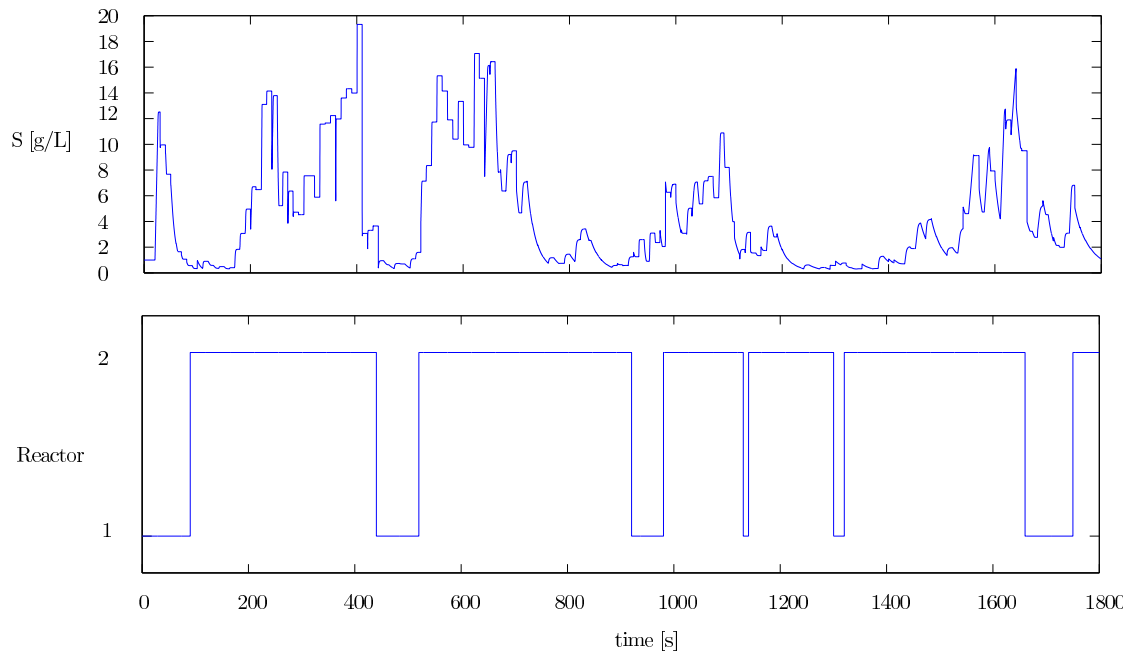


Figure 8.9: A sequence of concentration data observed by a micro-organism in the scaled-down reactor and its location in the scaled-down reactor, consisting of two continuous stirred-tank reactors

mise between the scaled-down reactor's complexity and obtaining a realistic imitation of the conditions met in large-scale bioreactors and does not focus on macroscopic variables as mixing time which in general yield an incorrect assessment of conversion efficiencies.

To impose the dynamic behaviour of the response variables, the manipulated variables were stringently controlled. The use of a maximal value for the substrate concentration set points is recommendable, since the specific conversion rates typically reach a plateau at substrate concentrations that are much larger than the affinity constant. This has not been done. Obviously, such a modification would significantly reduce the control efforts, both in terms of the waste flow Q_W , of the flows to dilute the broth in both the reactors Q_{D_1} and Q_{D_2} , and of the substrate pulses $pulse_1$ and $pulse_2$. Since then no surplus substrate would have to be added or removed. Partly due to these surplus control actions, the biomass flushes out from the present set-up and biomass retention would be needed, *e.g.*, by means of a membrane.

8.4 Conclusions

A method has been proposed to design a scaled-down reactor system on the basis of concentration data collected along a micro-organism's path in a large-scale bioreactor, rather than on the basis of macroscopic variables as mixing time and circulation time, which are far from ideal to describe improper mixing and conversion efficiencies. These data were obtained during a computational fluid dynamics simulation of a large-scale bioreactor.

The proposed set-up allows to imitate similar conditions in terms of substrate concentrations as those occurring in the large-scale bioreactor. However, due to the stringent control actions, it will be necessary to ensure biomass retention in the scaled-down reactor in order to avoid the flushing out of biomass.

The pursuit to rigorously mimic the large-scale conditions is a Moloch. Since for example a cut-off value for the substrate concentration set points could be used, since the specific conversion rates typically reach a plateau at substrate concentrations that are much larger than the affinity constant. Such a modification would significantly reduce the control efforts.

Chapter 9

Conclusions and perspectives

Meticulously optimised micro-organisms for the production of a variety of target compounds, optimised under highly reproducible and perfectly controlled laboratory-scale conditions, perform suboptimally when the process is scaled-up. This is due to biological, chemical, and physical processes which all are affected when scaling-up. The close interaction of these processes of various nature renders the study of large-scale bioreactors complex, as it is impossible to really uncouple these processes, since the time constants of those diverse processes are of the same order of magnitude: transport phenomena influence the local conditions which in turn influence microbial metabolism, which in turn influence local process conditions. In view of the latter, methods have been developed and applied in this study to investigate the biologically, chemically, and physically relevant processes that take place in large-scale bioreactors with a view to increasing insight in those processes and evaluating their importance for the widely observed yield reduction.

A detailed description of the biophase in such large-scale bioreactors seems essential. To this end the state of the art tools for modelling metabolism, typically used in the domain of metabolic engineering, were reviewed in Chapter 2, *i.e.*, stoichiometric network analysis (elementary flux modes, extreme pathways, and optimal flux distribution), steady-state metabolic modelling (metabolic flux analysis and flux balance analysis), dynamic metabolic modelling, and multivariate statistics. In the context of metabolic engineering, one should be aware that the usefulness of those tools to optimise microbial metabolism for overproducing a target compound depends predominantly on the characteristic properties of that compound. Due to their shortcomings not all meth-

ods are suitable for every kind of optimisation; issues like the dependence of the target compound's synthesis on severe (redox) constraints, the characteristics of its formation pathway, and the achievable/desired flux towards the target compound should play a role when choosing the optimisation strategy.

The vast variety of biochemical pathways micro-organisms dispose of, in order to fulfil their growth and reproduction requirements under a wide range of environmental conditions, renders them hard to fathom. Next to this tremendous amount of pathways, the lack of extensive (accurate) metabolomic, proteomic, and transcriptomic data sets also hampers the use and limits the usefulness of those mathematical methods.

For example, dynamic metabolic models might be useful tools to optimise microbial metabolism, as these models do incorporate kinetics and the regulation of enzymatic reactions. However, the drawbacks of this approach are still numerous. Models relying on *in vitro* derived mechanistic equations are overparameterised for the available data, nowadays typically collected during only one perturbation experiment. The alternative, approximative modelling is no *deus ex machina* either as in order to collect informative data for parameter identification it might be necessary to radically perturb the cell, probably way beyond the metabolite range for which approximative kinetic formats yield an adequate description of the true kinetics. In addition, these dynamic metabolic models, both mechanistic and approximative ones, zoom in on a limited part of the metabolism, which impedes mass balance checks during transient conditions. Moreover, the behaviour of cofactors is not yet modelled in a mechanistic manner, since, for instance, the pool size of the adenine nucleotides inexplicably changes during the transition from a glucose-limited to a glucose-abundant culture. Despite the rise of exchange tools like the systems biology markup language (SBML) [80], one thus should be aware that the 'plug and play' character of such model(s) (structures) remains limited.

The final aim of a dynamic model-based approach is thus target identification for optimising a production host. These targets are those reactions that control the flux through a reaction network, which can be assessed by calculating the flux control coefficients. Hence, assessing the uncertainty of the calculated flux control coefficients for the purpose of decision making/target identification in metabolic engineering is useful. This uncertainty may be the result of both an uncertain model structure and of uncertain parameter estimates. A Bayesian approach has been applied to properly assess this un-

certainty (Chapter 5).

Multiple approximative kinetic formats have been used to identify the flux control coefficients of the studied small network model. The tested approximative kinetic formats, the linlog kinetics, the linear in metabolite and enzyme levels kinetics, and the GMA type power law kinetics adequately described the data, which is somewhat contradictory to Heijnen (2005) [74] who points out the clear advantages of the linlog kinetics over the other ones. As shown, the model structure has a non negligible effect on the probability density function of the flux control coefficients and consequently it is worth the effort to search for the true model structure, *e.g.*, by means of genetic programming and optimal experimental design for model discrimination.

The usefulness of partial least squares regression as a tool to optimise microbial metabolism has been demonstrated using elementary flux mode data in Chapter 3. This approach allowed to rapidly pinpoint, without the need for experimental data, potential gene targets for succinate biosynthesis in *Escherichia coli*. The identified targets are in agreement with literature data, where modification of the expression of these genes proved to be beneficial to increase succinate yield. This approach has therefore passed a first validation round. Further evaluation of the method is however needed.

Cybernetic models were finally retained in this study, since it seemed the best method to describe the biophase in large-scale bioreactors. Indeed, a micro-organism in a large-scale bioreactor will develop a characteristic metabolomic and proteomic make-up, which will allow maximisation of its growth under those conditions, *e.g.*, mixed acid fermentation and overflow metabolism. This agrees well with the whole idea of cybernetic models that assume that a micro-organism tries to optimise its behaviour, *e.g.*, with respect to growth or substrate uptake. By allocating the resources a micro-organism disposes of to these enzymes yielding the optimal performance. Special attention has been devoted to the cybernetic control law ruling enzyme activity (Chapter 4). Several alternatives have been derived and evaluated for the conventionally used matching law. However, due to the limited knowledge, issues linked to the model structure, and the lack of appropriate data it was not possible to distinguish between the rival control laws.

Although the approach seems appealing, given the present lack of knowledge, detailed experimental omics data, and some of the aforementioned problems linked to 'conven-

tional' dynamic metabolic modelling, there still remain some issues unresolved, which will require further research.

Tools have been developed in this Ph.D. study which facilitate the gathering of data. A modus operandi of the Bioscope has been proposed in Chapter 6 for gathering data to build and validate a dynamic metabolic model of periodically operated cultures. Such models can be useful for the optimisation of periodically operated cultures as they help to gain further insight in the complex metabolic interactions and they can predict the effect of altered conditions. This set-up allows performing multiple perturbation experiments without perturbing the periodically operated culture itself, by controlling the opening and closing of the sample ports of the Bioscope. The perturbing agent, the sample time and the initial state prior to the perturbation are powerful degrees of freedom to maximise the information content of the collected data.

A method has been proposed in Chapter 8 to design a scaled-down reactor on the basis of simulated concentration data collected along a particle's path in a large-scale bioreactor, rather than using macroscopic variables as mixing time and circulation time, which are far from ideal to be linked with degrees of conversion. The proposed controlled set-up consisting of two continuous stirred-tank reactors allows to imitate similar conditions as those that occur in large-scale bioreactors. It will however be necessary to ensure biomass retention, *e.g.*, by a membrane, in the scaled-down reactor in order to avoid the flushing out of the biomass. The pursuit to rigorously mimic the large-scale conditions is a Moloch, since for example a maximal value for the substrate concentration set points could be used, as the cellular response to substrate concentrations much larger than the affinity constant reaches a plateau. The application of such a maximal value for the substrate concentration set points would also significantly reduce the control efforts.

Finally, a method has been proposed in Chapter 7 to describe the biophase in large-scale bioreactors by means of computational fluid dynamics using segregated models, in which micro-organisms are not considered identical and in which the cell is considered structured, *i.e.*, the internal composition and structure of the micro-organisms is considered. Due to the stochastic nature of particle transport and the fast metabolic response to the observed fast changing environmental conditions, this intracellular make-up is expected not to be identical for all micro-organisms. Describing the biophase in a Lagrangian way, *i.e.*, following the cell's path through the reactor, is computationally quite demanding

because a set of differential equations is linked to every micro-organism. However, by considering that the overall picture is merely the result of all individual micro-organisms it is only needed to track a limited number of particles in order to obtain a good idea of the fluxes in and out of the cells throughout the large-scale bioreactor. Indeed, the dynamics of the overall system can be captured by locally averaging out the behaviour of this limited number of particles over the whole population. Two-dimensional simulations have been performed of the large-scale bioreactor under study by means of computational fluid dynamics. Obviously, the incorporation of the third dimension would be beneficial to obtain a more realistic description of the large-scale bioreactor. However, this simplification does not derogate from the proposed method.

With respect to the elucidation of the mechanisms underlying the observed yield reduction in large-scale bioreactors the gathering of intracellular data seems essential. The gathering of (^{13}C , ^{32}P , and/or 2H) dynamic labelling data under large-scale fermentation conditions, possibly mimicked by a scaled-down reactor will be useful to study the cellular response to the observed fast changing environmental conditions. these data will also help to further investigate the hypotheses that attempt to explain the mechanisms responsible for the widely observed yield reduction in large-scale bioreactors. Some of the methods presented in this work will be useful in such investigations.

Bibliography

- [1] *Fluent 6.1 User's Guide*. Fluent Inc., Lebanon, 2003.
- [2] A. M. Abdel-Hamid, M. M. Attwood, and J. R. Guest. Pyruvate oxidase contributes to the aerobic growth efficiency of *Escherichia coli*. *Microbiology*, 147:1483–1498, 2001.
- [3] S. Aiba and M. Matsuoka. Identification of metabolic model: citrate production from glucose by *Candida lipolytica*. *Biotechnology and Bioengineering*, 21:1373–1386, 1979.
- [4] M. Akesson, J. Forster, and J. Nielsen. Integration of gene expression data into genome-scale metabolic models. *Metabolic Engineering*, 6:285–293, 2004.
- [5] H. Alper, Y.S. Jin, J. F. Moxley, and G. Stephanopoulos. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metabolic Engineering*, 7:155–164, 2005.
- [6] A. Arkin and J. Ross. Statistical construction of chemical reaction mechanisms from measured time-series. *The Journal of Physical Chemistry*, 99:970–979, 1995.
- [7] A. Arkin, P. Shen, and J. Ross. A test case of correlation metric construction of a reaction pathway from measurements. *Science*, 277:1275–1279, 1997.
- [8] M. R. Atkinson and A. J. Ninfa. Role of the GlnK signal transduction protein in the regulation of nitrogen assimilation in *Escherichia coli*. *Molecular Microbiology*, 29:431–447, 1998.
- [9] H. Attiya and J. Welch. *Distributed Computing: Fundamentals, Simulations and Advanced Topics*. John Wiley & Sons, 2004.

- [10] G. J. E. Baart, B. Zomer, A. de Haan, L. A. van der Pol, C. Beuvery, J. Tramper, and D. E. Martens. Modeling *Neisseria meningitidis* metabolism: from genome to metabolic fluxes. *Genome Biology*, 8:R316, 2007.
- [11] J. E. Bailey. Towards a science of metabolic engineering. *Science*, 252:1668–1675, 1991.
- [12] J. E. Bailey, A. R. Sburlati, V. Hatzimanikatis, K. Lee, W. A. Renner, and P. A. Tsai. Inverse metabolic engineering: a strategy for directed genetic engineering of useful phenotypes. *Biotechnology and Bioengineering*, 52:109–121, 1996.
- [13] S. Baloo and D. Ramkrishna. Metabolic regulation in bacterial cultures: I. *Biotechnology and Bioengineering*, 38:1337–1352, 1991.
- [14] S. Baloo and D. Ramkrishna. Metabolic regulation in bacterial cultures: II. *Biotechnology and Bioengineering*, 38:1353–1362, 1991.
- [15] H. J. Bart. Reactive extraction in stirred columns - a review. *Chemical Engineering and Technology*, 26:723–731, 2003.
- [16] T. Bayer, W. Zhou, K. Holzhauer, and K. Schugerl. Investigations of cephalosporin C production in an airlift tower loop reactor. *Applied Microbiology and Biotechnology*, 30:26–33, 1989.
- [17] T. Bedford and R. Cooke. *Probabilistic Risk Analysis: Foundation and Methods*. Cambridge University Press, Cambridge, 2004.
- [18] S. V. Benevolensky, D. Clifton, and D. G. Fraenkel. The effect of increased PGI on glucose metabolism in *Saccharomyces cerevisiae*. *In situ* study of the glycolytic pathway in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 269:4878–4882, 1994.
- [19] K. Beven. A manifesto for the equifinality thesis. *Journal of Hydrology*, 320:18–36, 2006.
- [20] F. Bezzo, S. Macchietto, and C. C. Pantelides. Computational issues in hybrid multizonal/computational fluid dynamics models. *AIChE Journal*, 51:1169–1177, 2005.
- [21] A. Bucholz, R. Takors, and C. Wandrey. Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatography-electrospray ionization tandem mass spectrometric techniques. *Analytical Biochemistry*, 295:129–137, 2001.

- [22] W. Bujalski, Z. Zaworski, and A. W. Nienow. CFD study of homogenization with dual Rushton turbines - comparison with experimental results. Part II: the multiple reference frame. *Transactions of the Institution of Chemical Engineers*, 80:97, 2002.
- [23] F. Bylund, E. Collet, S. O. Enfors, and G. Larsson. Substrate gradient formation in large scale bioreactors lowers cell yield and increases by-product formation. *Journal of Bioprocess Engineering*, 18:171–180, 1998.
- [24] T. Cakir, B. Kirdar, Z. I. Onsan, K. O. Ulgen, and J. Nielsen. Effect of carbon source perturbations on transcriptional regulation of metabolic fluxes in *Saccharomyces cerevisiae*. *BMC Systems Biology*, 1:18, 2007.
- [25] D. C. Cameron and C. L. Cooney. A novel fermentation: the production of R(-)-1,2-propanediol and acetol by *Clostridium thermosaccharolyticum*. *Bio/Technology*, 4:651–654, 1986.
- [26] F. Canonaco, T. A. Hess, S. Heri, T. Wang, T. Szyperski, and U. Sauer. Metabolic flux response to phosphoglucose isomerase knock-out in *Escherichia coli* and impact of overexpression of the soluble transhydrogenase UdhA. *FEMS Microbiology Letters*, 104:247–252, 2001.
- [27] M. Cánovas, V. Bernal, A. Sevilla, and J. L. Iborra. Role of wet experiment design in data generation: from *in vivo* to *in silico* and back. *In Silico Biology*, 7:S3–S16, 2007.
- [28] R. Carlson, D. Fell, and F. Sclienc. Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnology and Bioengineering*, 79:121–134, 2002.
- [29] R. Carlson and F. Sclienc. Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: creation of overall fluxes. *Biotechnology and Bioengineering*, 86:149–162, 2003.
- [30] C. Chassagnole, N. Noisommit-Rizzi, J. W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnology and Bioengineering*, 79:53–73, 2002.
- [31] B. H. Chen and S. P. Asprey. On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations. *Industrial & Engineering Chemistry Research*, 42:1379–1390, 2003.

- [32] G. K. Chotani, T. C. Dodge, A. L. Gaertner, and M. V. Arbige. Industrial biotechnology: discovery to delivery. In J. A. Kent, editor, *Kent and Riegel's Handbook of Industrial Chemistry and Biotechnology*, chapter 30, pages 1–64. Springer, 2008.
- [33] N. Cleland and S. O. Enfors. A biological system for studies on mixing in bioreactors. *Bioprocess Engineering*, 2:115–120, 1987.
- [34] S. J. Cox, L. S. Shalel, A. M. Sanchez, H. Lin, B. Peercy, G. N. Bennet, and K. Y. San. Development of a metabolic network design and optimization framework incorporating implement constraints: a succinate production case study. *Metabolic Engineering*, 8:46–57, 2006.
- [35] S. Dash, M. R. Maury, and V. Venkatasubramanian. A novel interval-halving framework for automated identification of process trends. *AIChE Journal*, 50:149–162, 2004.
- [36] K. A. Datsenko and B. L. Wanner. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, 97:6640–6645, 2000.
- [37] M. Dauner, J. E. Bailey, and U. Sauer. Metabolic flux analysis with a comprehensive isotopomer model in *Bacillus subtilis*. *Biotechnology and Bioengineering*, 76:144–156, 2001.
- [38] M. De Mey, G. Lequeux, J. Maertens, S. De Maeseneire, W. Soetaert, and E. Vandamme. Comparison of DNA and RNA quantification methods suitable for parameter estimation in metabolic modeling of microorganisms. *Analytical Biochemistry*, 352:198–203, 2006.
- [39] M. De Mey, J. Maertens, G. J. Lequeux, W. K. Soetaert, and E. J. Vandamme. Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnology*, 7:34, 2007.
- [40] D. De Pauw. *Optimal Experimental Design for Calibration of Bioprocess Models: a Validated Software Toolbox*. PhD thesis, Ghent University, Belgium, 2005.
- [41] D. J. W. De Pauw and P. A. Vanrolleghem. Practical aspects of sensitivity function approximation for dynamic models. *Mathematical and Computer Modelling of Dynamical Systems*, 12:395–414, 2006.

- [42] D. Degenring, C. Froemel, G. Dikta, and R. Takors. Sensitivity analysis for the reduction of complex metabolism models. *Journal of Process Control*, 14:729–745, 2004.
- [43] J. Delgado and J. C. Liao. Metabolic control analysis using transient metabolite concentrations. Determination of metabolite concentration control coefficients. *Biochemical Journal*, 282:965–972, 1992.
- [44] F. Delvigne, J. Destain, and P. Thomart. Bioreactor hydrodynamic effect in *Escherichia coli* physiology: experimental results and stochastic simulations. *Bioprocess Biosystems Engineering*, 28:131–137, 2005.
- [45] F. Delvigne, J. Destain, and P. Thomart. A methodology for the design of down-scale bioreactors by the use of mixing and circulation models. *Biochemical Engineering Journal*, 28:256–268, 2006.
- [46] D. Dochain and P. A. Vanrolleghem. *Dynamical Modelling and Estimation in Waste Water Treatment Processes*. IWA Publishing, London, 2001.
- [47] R. C. Dorf. *Modern Control Systems*. Addison-Wesley, 1989.
- [48] R. C. Dorf and R. H. Bishop. *Modern Control Systems*. Addison-Wesley, 1998.
- [49] W. B. Dunn, N. J. C. Bailey, and H. E. Johnson. Measuring the metabolome: current analytical technologies. *Analyst*, 130:606–625, 2005.
- [50] S. R. Eddy. What is a hidden Markov model. *Nature Biotechnology*, 22:1315–1316, 2004.
- [51] J. S. Edwards and B. Palsson. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America*, 97:5528–5533, 2000.
- [52] P. H. C. Eilers. A perfect smoother. *Analytical Chemistry*, 75:3631–3636, 2003.
- [53] S. O. Enfors, M. Jahic, A. Rozkov, B. Xu, M. Hecker, B. Jurgen, E. Kruger, T. Schweder, G. Hamer, D. O’Beirne, N. Noisommit-Rizzi, M. Reuss, L. Boone, C. Hewitt, C. Mcfarlane, A. Nienow, T. Kovacs, C. Tragardh, L. Fuchs, J. Revstedt, P. C. Friberg, B. Hjertjager, G. Blomsten, H. Skogman, S. Hjort, F. Hoeks, H. Y. Lin, P. Neubauer, R. Van der Lans, K. Luyben, P. Vrabel, and A. Manelius. Physiological responses to mixing in large bioreactors. *Journal of Biotechnology*, 85:175–185, 2001.

- [54] D. Fell. Metabolic control analysis: a survey of its theoretical and experimental development. *Biochemical Journal*, 286:313–330, 1992.
- [55] A. Fersht. *Structure and Mechanism in Protein Science. A Guide to Enzyme Catalysis and Protein Folding*. W. H. Freeman and Company, New York, 1999.
- [56] S. S. Fong, A. R. Joyce, and B. O. Palsson. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Research*, 15:1365–1372, 2005.
- [57] 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.
- [58] G. Frazzetto. White biotechnology. *EMBO Reports*, 4:835–837, 2003.
- [59] P. Ganesan and H. P. Rani. On diffusion of chemically reactive species in convective flow along a vertical cylinder. *Chemical Engineering and Processing*, 39:93–105, 2000.
- [60] J. Gao, V. M. Gorenflo, J. M. Scharer, and H. M. Budman. Dynamic metabolic modeling for a MAB bioprocess. *Biotechnology Progress*, 23:168–181, 2007.
- [61] H. Garoff. Using recombinant DNA techniques to study protein targeting in the eucaryotic cell. *Annual Review of Cell Biology*, 1:403–445, 1985.
- [62] S. C. Generalis and G. M. C. Glover. Modelling a biochemical reaction with computational fluid dynamics. *International Journal of Chemical Reactor Engineering*, 3:A50, 2005.
- [63] S. George, G. Larsson, and S. O. Enfors. A scale-down two-compartment reactor with controlled substrate oscillations: metabolic response of *Saccharomyces cerevisiae*. *Bioprocess Engineering*, 9:249–257, 1993.
- [64] C. Giersch. Concerning the measurement of flux control coefficients by enzyme titration. Steady states, quasi steady states and the role of time in control analytical experiments. *European Journal of Biochemistry*, 231:587–592, 1995.
- [65] J. W. Goethe. *Faust Oerversie*. Atheneum-Polak & Van Gennepe, 2005.
- [66] A. K. Gombert and J. Nielsen. Mathematical modelling of metabolism. *Current Opinion in Biotechnology*, 11:180–186, 2000.

- [67] W. Gujer. Microscopic versus macroscopic biomass models in activated sludge systems. *Water Science and Technology*, 45(6):1–11, 2002.
- [68] R. Gupta, Q. K. Beg, S. Khan, and B. Chauhan. An overview on fermentation, downstream processing and properties of microbial alkaline proteases. *Applied Microbiology and Biotechnology*, 60:381–395, 2002.
- [69] K. Hammer, I. Mijakovic, and P. R. Jensen. Synthetic promoter libraries - tuning of gene expression. *Trends in Biotechnology*, 24:53–55, 2006.
- [70] L. Han, S. O. Enfors, and L. Haggstrom. Changes in intracellular metabolite pools, and acetate formation in *Escherichia coli* are associated with a cell-density-dependent metabolic switch. *Biotechnology Letters*, 24:483–488, 2002.
- [71] P. S. Harvey and M. Greaves. Turbulent flow in agitated vessels. *Transactions of the Institution of Chemical Engineers*, 60:195–210, 1982.
- [72] V. Hatzimanikatis and J. E. Bailey. MCA has more to say. *Journal of Theoretical Biology*, 182:233–242, 1996.
- [73] V. Hatzimanikatis and J. E. Bailey. Effects of spatiotemporal variations on metabolic control: approximative analysis using (log) linear kinetic models. *Biotechnology and Bioengineering*, 54:91–104, 1997.
- [74] J. J. Heijnen. Approximative kinetic formats used in metabolic network modeling. *Biotechnology and Bioengineering*, 91:534–545, 2005.
- [75] C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis. Thermodynamics-based metabolic flux analysis. *Biophysical Journal*, 92:1792–1805, 2007.
- [76] C. J. Hewitt, J. Nebe-Von Caron, B. Axelsson, C. M. McFarlane, and A. W. Nienow. Studies related to the scale-up of high cell density *Escherichia coli* fed-batch using multiparameter flow cytometry: effect of a changing microenvironment with respect to glucose and dissolved oxygen concentration. *Biotechnology and Bioengineering*, 71:381–389, 2000.
- [77] P. Hieter and M. Boguski. Functional genomics: it’s all how you read it. *Science*, 278:601–602, 1997.
- [78] M. Y. Hirai, M. Yano, D. B. Goodenowe, S. Kanaya, T. Kimura, M. Awazuhara, M. Arita, T. Fujiwara, and K. Saito. Integration of transcriptomics and

- metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America*, 101:10205–12012, 2004.
- [79] M. Hoare, M. S. Levy, D. G. Bracewell, S. D. Doig, S. Kong, N. Titchener-Hooker, J. M. Ward, and P. Dunnill. Bioprocess engineering issues that would be faced in producing a DNA vaccine at up to 100 m³ fermentation scale for an influenza pandemic. *Biotechnology Progress*, 21:1577–1592, 2005.
- [80] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novere, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19:524–531, 2003.
- [81] K. Jantama, M. J. Haupt, S. A. Svoronos, X. Zhang, J. C. Moore, K. T. Shanmugam, and L. O. Ingram. Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* C that produce succinate and malate. *Biotechnology and Bioengineering*, 99:1140–1153, 2007.
- [82] M. Jesus Guardia, A. Gambhir, A. F. Europa, D. Ramkrishna, and W. H. Hu. Cybernetic modeling and regulation of metabolic pathways in multiple steady states of hybridoma cells. *Biotechnology Progress*, 16:847–853, 2000.
- [83] K. D. Jones and D. S. Kompala. Cybernetic model of the growth dynamics of *Saccharomyces cerevisiae* in batch and continuous cultures. *Journal of Biotechnology*, 71:105–131, 1999.
- [84] J. Jonsson, T. Norberg, L. Carlsson, C. Gustafsson, and S. Wold. Quantitative sequence-activity models (QSAM)-tools for sequence design. *Nucleic Acids Research*, 21:733–739, 1993.
- [85] H. Kacser and J. A. Burns. The control of flux. *Symposia of the Society for Experimental Biology*, 27:65–104, 1973.

- [86] L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data: an Introduction to Cluster Analysis*. John Wiley & Sons, 1990.
- [87] D. B. Kell. Metabolomics and systems biology: making sense of the soup. *Current Opinion in Microbiology*, 7:296–307, 2004.
- [88] I. M. Keseler, J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I. T. Paulsen, M. Peralta-Gil, and P. D. Karp. EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Research*, 33:D334–D337, 2005.
- [89] M. Koffas and S. del Cardayre. Evolutionary metabolic engineering. *Metabolic Engineering*, 7:1–3, 2005.
- [90] D. S. Kompala, D. Ramkrishna, and G. T. Tsao. Cybernetic modeling of microbial growth on multiple substrates. *Biotechnology and Bioengineering*, 26:1272–1281, 1984.
- [91] M. T. A. P. Kresnowati, W. A. van Winden, M. J. H. Almering, A. ten Pierick, C. Ras, T. A. Knijnenburg, P. Daran-Lapujade, J. T. Pronk, J. J. Heijnen, and J. M. Daran. When transcriptome meets metabolome: fast cellular responses of yeast to sudden relief of glucose limitation. *Molecular Systems Biology*, 2:49, 2006.
- [92] M. T. A. P. Kresnowati, W. A. van Winden, and J. J. Heijnen. Determination of elasticities, concentration and flux control coefficients from transient metabolic data using linlog kinetics. *Metabolic Engineering*, 7:142–153, 2005.
- [93] J. O. Kromer, C. Wittman, H. Schroder, and E. Heinzle. Metabolic pathway analysis for rational design of L-methionine production by *Escherichia coli* and *Corynebacterium glutamicum*. *Metabolic Engineering*, 8:353–369, 2006.
- [94] A. Kukush and S. Van Huffel. Consistency of elementwise-weighted total least squares estimator in a multivariate errors-in-variables model $AX = B$. *Metrika*, 59:75–97, 2004.
- [95] A. Kummel, S. Panke, and M. Heineman. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Molecular Systems Biology*, 2:2006–2034, 2006.
- [96] A. Lapin, D. Muller, and M. Reuss. Dynamic behavior of microbial populations in stirred bioreactor simulated with Euler-Lagrange methods: traveling along the

- lifeline of single cells. *Industrial & Engineering Chemistry Research*, 43:4647–4656, 2004.
- [97] A. Lapin, J. Schmid, and M. Reuss. Modeling the dynamics of *E. coli* populations in the three-dimensional turbulent fields of a stirred-tank bioreactor - a structured-segregated approach. *Chemical Engineering Science*, 61:4783–4797, 2006.
- [98] K. H. Lee, J. H. Park, T. Y. Kim, H. U. Kim, and S. Y. Lee. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Molecular Systems Biology*, 3:149, 2007.
- [99] S. Y. Lee, S. H. Homg, and S. Y. Moon. *In silico* metabolic pathway analysis and design: succinic acid production by metabolically engineered *Escherichia coli* as an example. *Genome Informatics*, 13:214–223, 2002.
- [100] G. Lequeux, L. Johansson, J. Maertens, P. A. Vanrolleghem, and G. Lidén. MFA for overdetermined systems reviewed and compared with RNA expression data to elucidate the difference in shikimate yield between carbon- and phosphate-limited continuous cultures of *E. coli* W3110. shik1. *Biotechnology Progress*, 22:1056–1070, 2006.
- [101] G. Lerondel, T. Doan, N. Zamboni, U. Sauer, and S. Ayermich. YtsJ has the major physiological role of the four paralogous malic enzyme isoforms in *Bacillus subtilis*. *Journal of Bacteriology*, 188:4727–4736, 2006.
- [102] B. Li, M. Sjostrom, and L. Eriksson. Model selection for partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 64:79–89, 2002.
- [103] M. Li, S. Yao, and K. Shimizu. Effect of *poxB* gene knockout on metabolism in *Escherichia coli* based on growth characteristics and enzyme activities. *World Journal of Microbiology and Biotechnology*, 23:573–580, 2007.
- [104] R. Li, G. Meng, N. Gao, and H. Xie. Combined use of partial least-squares regression and neural network for residual life estimation of large generator stator insulation. *Measurement Science and Technology*, 18:2074–2082, 2007.
- [105] J. C. Liao, R. Boscolo, Y. L. Yang, L. M. Tran, C. Sabatti, and V. P. Roychowdhury. Network component analysis: reconstruction of regulatory signals in biological systems. *Proceedings of the National Academy of Sciences of the United States of America*, 100:15522–15527, 2003.

- [106] H. Lin, G. N. Bennet, and K. Y. San. Fed-batch of a metabolically engineered *Escherichia coli* strain designed for high-level succinate production and yield under aerobic conditions. *Biotechnology and Bioengineering*, 90:775–779, 2005.
- [107] L. Ljung. *System Identification: Theory for the User*. Prentice Hall, New Jersey, 1999.
- [108] R. A. Majewski and M. M. Domach. Simple constrained-optimization view of acetate overflow in *E. coli*. *Biotechnology and Bioengineering*, 35:732–738, 1990.
- [109] O. Malve, M. Laine, H. Haario, T. Kirkkila, and J. Sarvala. Bayesian modelling of algal mass occurrences - using adaptive MCMC with a lake water quality model. *Environmental modelling & software*, 22:966–977, 2007.
- [110] R. Mann, S. K. Pillai, A. M. El-Hamouz, P. Ying, A. Togatorop, and R. B. Edwards. Computational fluid mixing for stirred vessels: progress from seeing to believing. *Chemical Engineering Journal*, 59:39–50, 1995.
- [111] C.D. Maranas. Computational analysis and redesign of biological networks. In *IECA 2006*, 2006.
- [112] M. R. Mashego, M. L. A. Jansen, J. L. Vinke, W. M. van Gulik, and J. J. Heijnen. Changes in the metabolome of *Saccharomyces cerevisiae* associated with evolution in aerobic glucose-limited chemostats. *FEMS Yeast Research*, 5:419–430, 2005.
- [113] M. R. Mashego, K. Rumbold, M. De Mey, E. Vandamme, W. Soetaert, and J. J. Heijnen. Microbial metabolomics: past, present and future methodologies. *Biotechnology Letters*, 29:1–11, 2007.
- [114] M. R. Mashego, W. M. van Gulik, and J. J. Heijnen. Metabolome dynamic responses of *Saccharomyces cerevisiae* to simultaneous rapid perturbations in external electron acceptor and electron donor. *FEMS Yeast Research*, 7:48–66, 2007.
- [115] K. Mauch, S. Arnold, and M. Reuss. Dynamic sensitivity analysis for metabolic systems. *Chemical Engineering Science*, 52:2589–2598, 1997.
- [116] C. S. Millard, Y. P. Chao, J. C. Liao, and M. I. Donnely. Enhanced production of succinic acid by overexpression of phosphoenolpyruvate carboxylase in *Escherichia coli*. *Applied and Environmental Microbiology*, 62:1808–1810, 1996.
- [117] J. Monod. The phenomenon of enzymatic adaptation and its bearing on problems of genetics and cellular differentiation. *Growth Symposium*, 11:223–289, 1947.

- [118] M. G. Morgan and M. Henron. *Uncertainty*. Cambridge University Press, Cambridge, 2004.
- [119] S. A. Morsi and A. J. Alexander. An investigation of particle trajectories in two-phase flow systems. *Journal of Fluid Mechanics*, 55:193–208, 1972.
- [120] J. Y. Murthy, S. R. Mathur, and D. Choudhary. CFD simulation of flows in stirred tank reactors using a sliding mesh technique. *Institution of Chemical Engineers Symposium Series*, 136:341–348, 1994.
- [121] P. K. Namdev, P. K. Yegneswaran, B. G. Thompson, and M. R. Gray. Experimental simulation of large-scale bioreactor environments using a monte-carlo method. *Canadian Journal of Chemical Engineering*, 69:513–519, 1991.
- [122] A. Namjoshi, A. Kienle, and D. Ramkrishna. Steady-state multiplicity in bioreactors: bifurcation analysis of cybernetic models. *Chemical Engineering Science*, 58:793–800, 2003.
- [123] A. Narang, A. Konopka, and D. Ramkrishna. New patterns of mixed-substrate utilization during batch growth of *Escherichia coli* K12. *Biotechnology and Bioengineering*, 55:747–757, 1997.
- [124] J. Nielsen. Metabolic engineering: techniques for analysis of targets for genetic manipulations. *Biotechnology and Bioengineering*, 58:125–132, 1998.
- [125] I. E. Nikerel, W. A. van Winden, W. M. van Gulik, and J. J. Heijnen. A method for estimation of elasticities in metabolic networks using steady state and dynamic metabolomics data and linlog kinetics. *BMC Bioinformatics*, 7:540, 2006.
- [126] I. Nookaew, A. Meechai, C. Thammarongtham, K. Laoteng, V. Ruanglek, S. Cheevadhanarak, J. Nielsen, and S. Bhumiratana. Identification of flux regulation coefficients from elementary flux modes: a systems biology tool for analysis of metabolic networks. *Biotechnology and Bioengineering*, 97:1535–1549, 2007.
- [127] T. Norton and D. W. Sun. Computational fluid dynamics (CFD) - an effective and efficient design and analysis tool for the food industry: a review. *Trends in Food Science and Technology*, 17:600–620, 2006.
- [128] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 27:29–34, 1999.

- [129] M. Oldiges, M. Kunze, D. Degenring, G. A. Sprenger, and R. Takors. Simulation, monitoring and analysis of pathway dynamics by metabolic profiling in the aromatic amino acid pathway. *Biotechnology progress*, 20:1623–1633, 2004.
- [130] M. Oldiges, S. Lutz, S. Pflug, K. Schroer, N. Stein, and C. Wiendahl. Metabolomics: current state and evolving methodologies and tools. *Applied Microbiology and Biotechnology*, 76:495–511, 2007.
- [131] H. Onyeaka, A. W. Nienow, and C. J. Hewitt. Further studies related to the scale-up of high cell density *Escherichia coli* fed-batch fermentations: the additional effect on changing microenvironment when using aqueous ammonia to control pH. *Biotechnology and Bioengineering*, 84:474–484, 2003.
- [132] B. O. Palsson. *In silico* biology through "omics". *Nature Biotechnology*, 20:649–650, 2002.
- [133] M. Papagianni, M. Mattey, and B. Kristiansen. Design of a tubular loop bioreactor for scale-up and scale-down of fermentation processes. *Biotechnology Progress*, 19:1498–1504, 2003.
- [134] J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson. Comparison of network-based pathway analysis methods. *Trends in Biotechnology*, 22:400–405, 2004.
- [135] J. H. Park, K. H. Lee, T. Y. Kim, and S. Y. Lee. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome and *in silico* gene knockout simulation. *Proceedings of the National Academy of Sciences of the United States of America*, 104:7797–7802, 2007.
- [136] K. R. Patil and J. Nielsen. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proceedings of the National Academy of Sciences of the United States of America*, 102:2685–2689, 2005.
- [137] K. R. Patil, I. Rocha, J. Forster, and J. Nielsen. Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics*, 6:308, 2005.
- [138] P. R. Patnaik. Are microbes intelligent beings?: an assessment of cybernetic modeling. *Biotechnology Advances*, 18:267–288, 2000.

- [139] D. J. Pollard, A. P. Isin, P. A. Shamlou, and M. D. Lilly. Reactor heterogeneity with *Saccharopolyspora erythraea* airlift fermentations. *Biotechnology and Bioengineering*, 58:453–463, 1998.
- [140] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006.
- [141] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77:257–286, 1989.
- [142] R. Ramakrishna, J. S. Edwards, A. McCulloch, and B. O. Palsson. Flux-balance analysis of mitochondrial energy metabolism: consequences of system stoichiometric constraints. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, 280:695–704, 2001.
- [143] R. Ramakrishna, D. Ramkrishna, and A. E. Konopka. Cybernetic modeling of growth in mixed, substitutable substrate environments: preferential and simultaneous utilization. *Biotechnology and Bioengineering*, 52:141–151, 1996.
- [144] K. Raman, P. Rajagopalan, and N. Chandra. Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs. *PLoS Computational Biology*, 1:e46, 2005.
- [145] V. V. Ranade and J. B. Joshi. Flow generated by a disc turbine: part II. Mathematical modeling and comparison with experimental results. *Transactions of the Institution of Chemical Engineers*, 68:34–50, 1990.
- [146] V. V. Ranade, J. B. Joshi, and A. G. Marathe. Flow generated by pitched blade turbines–II: simulation using k- ϵ model. *Chemical Engineering Communications*, 81:225–248, 1989.
- [147] J. Ranta and R. Maijala. A probabilistic transmission model of *Salmonella* in the primary broiler production chain. *Risk Analysis*, 22:47–58, 2002.
- [148] J. L. Reed, T. D. Vo, C. H. Schilling, and B. O. Palsson. An expanded genome-scale model of *Escherichia coli* K-12. *Genome Biology*, 4:54, 2003.
- [149] M. Rizzi, M. Baltes, U. Theobald, and M. Reuss. *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. Mathematical model. *Biotechnology and Bioengineering*, 55:592–608, 1997.

- [150] A. Saltelli, S. Tarantola, and K. P. S. Chan. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41:39–56, 1999.
- [151] A. M. Sanchez, G. N. Bennett, and K. Y. San. Efficient succinic acid production from glucose through overexpression of pyruvate carboxylase in an *Escherichia coli* alcohol dehydrogenase and lactate dehydrogenase mutant. *Biotechnology Progress*, 21:358–365, 2005.
- [152] C. N. S. Santos and G. Stephanopoulos. Combinatorial engineering of microbes for optimizing cellular phenotypes. *Current Opinion in Chemical Biology*, 12:168–176, 2008.
- [153] U. Sauer and B. J. Eikmanns. The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiology Reviews*, 29:765–794, 2005.
- [154] M. A. Savageau. *Biochemical System Analysis*. Addison-Wesley Publishing Company, Reading, 1st edition, 1976.
- [155] A. Savitzky and M. J. E. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36:1627–1639, 1964.
- [156] U. Schaefer, W. Boos, R. Takors, and D. Weuster-Botz. Automated sampling device for monitoring intracellular metabolite dynamics. *Analytical Biochemistry*, 270:88–96, 1999.
- [157] C. H. Schilling, M. W. Covert, I. Famili, G. M. Church, J. S. Edwards, and B. O. Palsson. Genome-scale metabolic model of *Helicobacter pylori* 26695. *Journal of Bacteriology*, 184:4582–4593, 2002.
- [158] I. Schomburg, O. Hofmann, C. Bansch, A. Chang, and D. Schomburg. Enzyme data and metabolic information: BRENDA, a resource for research in biology, biochemistry, and medicine. *Gene Function and Disease*, 3-4:109–118, 2000.
- [159] R. Schuetz, L. Kuepfer, and U. Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*, 3:119, 2007.
- [160] J. M. Schwartz, C. Gauguin, J. Nacher, A. de Daruvar, and M. Kanehisa. Observing metabolic functions at the genome scale. *Genome Biology*, 8:R123, 2007.

- [161] I. H. Segel. *Enzyme Kinetics. Behavior and Analysis of Rapid Equilibrium and Steady State Enzyme Systems*. John Wiley & Sons, New York, 1993.
- [162] D. Segre, D. Vitkup, and G. M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99:15112–15117, 2002.
- [163] A. Sevilla, J. W. Schmid, K. Mauch, J. L. Iborra, M. Reuss, and Cánovas. Model of central and trimethylammonium metabolism for optimizing L-carnitine production by *E. coli*. *Metabolic Engineering*, 7:401–425, 2005.
- [164] T. Shlomi, Y. Eisenberg, R. Sharan, and E. Ruppin. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Molecular Systems Biology*, 3:101, 2007.
- [165] G. Sin, G. Insel, D. S. Lee, and P. A. Vanrolleghem. Optimal but robust N and P removal in SBRs: a model-based systematic study of operation scenarios. *Water Science and Technology*, 50(10):97–105, 2004.
- [166] G. Sin, K. Villez, and P. A. Vanrolleghem. Application of a model-based optimisation methodology for nutrient removing SBRs leads to falsification of the model. *Water Science and Technology*, 53(4-5):95–103, 2006.
- [167] I. M. Sobol. Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment*, 1:407–414, 1993.
- [168] O. J. G. Somsen, M. A. Hoeben, E. Esgahado, J. L. Snoep, D. Visser, R. T. J. M. Van der Heijden, J. J. Heijnen, and H. V. Westerhoff. Glucose and the ATP paradox in yeast. *Biochemical Journal*, 352:593–599, 2000.
- [169] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E. D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420:190–193, 2002.
- [170] G. Stephanopoulos and T. Simpson. Flux amplification in complex metabolic networks. *Chemical Engineering Science*, 52:2607–2627, 1997.
- [171] G. N. Stephanopoulos, A. A. Aristidou, and J. Nielsen. *Metabolic Engineering*. Academic Press, 1998.

- [172] J. V. Straight and D. Ramkrishna. Cybernetic modeling and regulation of metabolic pathways. growth on complementary nutrients. *Biotechnology Progress*, 10:574–587, 1994.
- [173] A. P. J. Sweere. *Response of Baker's Yeast to Transient Environmental Conditions Relevant to Large-Scale Fermentation Processes*. PhD thesis, T. U. Delft, The Netherlands, 1988.
- [174] A. P. J. Sweere, K. C. A. M. Luyben, and N. W. F. Kossen. Regime analysis and scale-down: tools to investigate the performance of bioreactors. *Enzyme and Microbial Technology*, 9:386–398, 1987.
- [175] R. L. Switzer. The inactivation of microbial enzymes *in vivo*. *Annual Review of Microbiology*, 31:135–154, 1977.
- [176] G. Tabor, A. D. Gosman, and R. Issa. Numerical simulation of flow in a mixing vessel stirred by a Rushton turbine. *Institution of Chemical Engineers Symposium Series*, 140:25–34, 1996.
- [177] Y. Takiguchi, H. Mishima, M. Okuda, M. Terao, A. Aoki, and R. Fukuda. Milbemycins, a new family of macrolide antibiotics: fermentation, isolation and physico-chemical properties. *Journal of Antibiotics*, 33:1120–1127, 1980.
- [178] J. C. Tannehil, D. A. Anderson, and R. H. Pletcher. *Computational Fluid Mechanics and Heat Transfer*. Taylor and Francis, 1997.
- [179] S. Teppola, S. P. Mujunen, and P. Minkkinen. Partial least squares modeling of an activated sludge plant: a case study. *Chemometrics and Intelligent Laboratory Systems*, 38:197–208, 1997.
- [180] B. Teusink, J. Passarge, C. A. Reijenga, E. Esgalhado, C. C. van der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. van Dam, H. V. Westerhoff, and J. L. Snoep. Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *European Journal of Biochemistry*, 267:5313–5329, 2000.
- [181] U. Theobald, W. Mailinger, M. Baltes, M. Rizzi, and M. Reuss. *In vivo* analysis of metabolic dynamics in *Saccharomyces cerevisiae*: I. Experimental observations. *Biotechnology and Bioengineering*, 55:305–316, 1997.

- [182] U. Theobald, W. Mailinger, M. Reuss, and M. Rizzi. *In vivo* analysis of glucose-induced fast changes in yeast adenine nucleotide pool applying a rapid sampling technique. *Analytical Biochemistry*, 214:31–37, 1993.
- [183] M. Thiry and D. Cingolani. Optimizing scale-up fermentation processes. *Trends in Biotechnology*, 20:103–105, 2002.
- [184] S. Thomas and D. A. Fell. The role of multiple enzyme activation in metabolic flux control. *Advances in Enzyme Regulations*, 38:65–85, 1998.
- [185] M. Tohyama, T. Patarinska, Z. Qiang, and K. Shimizu. Modeling of the mixed culture and periodic control for PHB production. *Biochemical Engineering Journal*, 10:157–173, 2002.
- [186] M. Tomita. Whole-cell simulation: a grand challenge of the 21st century. *Trends in Biotechnology*, 19:205–210, 2001.
- [187] M. Tomita, K. Takahashi, Y. Matsuzaki, K. Saito, S. Tanida, K. Yugi, J. C. Venter, and C. A. Hutchison 3rd. E-cell: software environment for whole-cell simulation. *Bioinformatics*, 15:72–84, 1999.
- [188] R. Tomovich and M. Vukobratovich. *General Sensitivity Theory*. American Elsevier, 1972.
- [189] M. A. Trujillo-Roldan, C. Pena, O. T. Ramirez, and E. Galdino. Effect of dissolved oxygen tension on the production of alginate by *Azotobacter vinelandii*. *Biotechnology Progress*, 17:1042–1048, 2001.
- [190] B. G. Turner, D. Ramkrishna, and N. B. Jansen. Cybernetic modeling of bacterial cultures at low growth rates: single-substrate systems. *Biotechnology and Bioengineering*, 1989:252–161, 1989.
- [191] S. Ui, A. Mimura, M. Ohkuma, and T. Kudo. Formation of a chiral acetoinic compound from diacetyl by *Escherichia coli* expressing meso-2,3-butanediol dehydrogenase. *Letters in Applied Microbiology*, 28:457–460, 1999.
- [192] C. P. Unsworth, M. R. Cowper, S. McLaughin, and B. Mulgrew. A new method to detect nonlinearity in a time-series: synthesizing surrogate data using a Kolmogorov-Smirnoff tested, hidden Markov model. *Physica D*, 155:51–68, 2001.

- [193] R. A. van den Berg, H. C. J. Hoefsloot, J. A. Westerhuis, A. K. Smilde, and M. J. van der Werf. Centering, scaling and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, 7:142, 2006.
- [194] 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: II. Diagnosis and estimation of gross errors. *Biotechnology and Bioengineering*, 43:11–20, 1994.
- [195] M. J. van der Werf. Towards replacing closed with open target selection strategies. *Trends in Biotechnology*, 23:11–16, 2005.
- [196] M. J. van der Werf. TNO: microbial biotechnology. www.tno.nl/downloads/%5CMicrobial%20Production%20Proceses1.pdf, 2008.
- [197] M. J. van der Werf, R. H. Jellema, and T. Hankemeier. Microbial metabolomics: replacing trial-and-error by the unbiased selection and ranking of targets. *Journal of Industrial Microbiology and Biotechnology*, 32:234–252, 2005.
- [198] W. M. van Gulik, W. T. de Laat, J. L. Vinke, and J. J. Heijnen. Application of metabolic flux analysis for the identification of metabolic bottlenecks in the biosynthesis of penicillin-G. *Biotechnology and Bioengineering*, 68:602–618, 2000.
- [199] S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem Computational Aspects and Analysis*. SIAM, Philadelphia, 1991.
- [200] N. van Luijk, K. Rochat, B. Muilwijk, B. van der Werff-van der Vat, S. Bijlsma, L. Coulier, R. Jellema, T. Hankemeier, and M. J. van der Werf. Strain improvement with seven-league boots: application of metabolomics to improve phenylalanine production by *Escherichia coli*. www.tno.nl/downloads/KvL_FBI_Poster_democoli_metabolomics.pdf, 2008.
- [201] W. A. van Winden, J. C. van Dam, C. Ras, R. J. Kleijn, J. L. Vinke, W. M. van Gulik, and J. J. Heijnen. Metabolic-flux analysis of *Saccharomyces cerevisiae* CEN. PK113-7D based on mass isotopomer measurements of ¹³C-labeled primary metabolites. *FEMS Yeast Research*, 5:559–568, 2005.
- [202] F. Varder and M. D. Lilly. Effect of cycling dissolved oxygen concentration on product formation in penicillin fermentations. *European Journal of Applied Microbiology*, 14:203–211, 1982.

- [203] A. Varma and B. O. Palsson. Metabolic capabilities of *Escherichia coli*: II. Optimal growth patterns. *Journal of Theoretical Biology*, 165:503–522, 1993.
- [204] J. Varner and D. Ramkrishna. Application of cybernetic models to metabolic engineering: investigation of storage pathways. *Biotechnology and Bioengineering*, 58:282–291, 1998.
- [205] J. Varner and D. Ramkrishna. Metabolic engineering from a cybernetic perspective. 1. Theoretical preliminaries. *Biotechnology Progress*, 15:407–425, 1999.
- [206] J. Varner and D. Ramkrishna. Metabolic engineering from a cybernetic perspective: aspartate family of amino acids. *Metabolic Engineering*, 1:88–116, 1999.
- [207] S. Vaseghi, A. Baumeister, M. Rizzi, and M. Reuss. *In vivo* dynamics of the pentose phosphate pathway in *Saccharomyces cerevisiae*. *Metabolic Engineering*, 1:128–140, 1999.
- [208] S. G. Villas-Boas and P. Bruheim. Cold glycerol-saline: the promising quenching solution for accurate intracellular metabolite analysis of microbial cells. *Analytical Biochemistry*, 370:87–97, 2007.
- [209] D. Visser and J. J. Heijnen. Dynamic simulation and metabolic re-design of a branched pathway using lin-log kinetics. *Metabolic Engineering*, 5:164–176, 2003.
- [210] D. Visser, J. W. Schmid, K. Mauch, M. Reuss, and J. J. Heijnen. Optimal re-design of primary metabolism in *Escherichia coli* using linlog kinetics. *Metabolic Engineering*, 6:378–390, 2004.
- [211] D. Visser, G. A. van Zuylen, J. C. van Dam, M. R. Eman, A. Proll, C. Ras, L. Wu, W. M. van Gulik, and J. J. Heijnen. Analysis of *in vivo* kinetics of glycolysis in aerobic *Saccharomyces cerevisiae* by application of glucose and ethanol pulses. *Biotechnology and Bioengineering*, 88:157–167, 2004.
- [212] D. Visser, G. A. van Zuylen, J. C. van Dam, A. Oudshoorn, M. R. Eman, C. Ras, W. M. van Gulik, J. Frank, G. W. van Dedem, and J. J. Heijnen. Rapid sampling for analysis of *in vivo* kinetics using the bioscope: a system for continuous-pulse experiments. *Biotechnology and Bioengineering*, 79:674–681, 2002.
- [213] E. O. Voit, F. Alvarez-Vasquez, and K. J. Sims. Analysis of dynamic labeling data. *Mathematical Biosciences*, 191:83–99, 2004.

- [214] A. von Kamp and S. Schuster. Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*, 22:1930–1931, 2006.
- [215] H. V. Westerhoff and K. Van Dam. *Thermodynamics and Control of Biological Free Energy Transduction*. Elsevier, Amsterdam, 1987.
- [216] S. J. Wiback and B. O. Palsson. Extreme pathway analysis of human red blood cell metabolism. *Biophysical Journal*, 83:808–818, 2002.
- [217] 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*, 66:69–85, 1999.
- [218] S. Wold. Cross-validation estimation of the number of components in factor and principal component analysis. *Technometrics*, 24:397–405, 1978.
- [219] S. Wold, M. Sjostrom, and L. Eriksson. PLS-regression: a basic tool for chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58:109–130, 2001.
- [220] S. Wold, J. Trygg, A. Berglund, and H. Antti. Some recent developments in PLS modeling. *Chemometrics and Intelligent Laboratory Systems*, 58:131–150, 2001.
- [221] B. Xu, M. Jahic, G. Blomstein, and S. O. Enfors. Glucose overflow metabolism and mixed acid fermentation in aerobic large-scale fed-batch processes with *Escherichia coli*. *Applied Microbiology and Biotechnology*, 51:564–571, 1999.
- [222] Y. T. Yang, G. N. Bennet, and K. Y. San. Genetic and metabolic control. *Electronic Journal of Biotechnology*, 1:3, 1998.
- [223] J. D. Young. *A System-Level Mathematical Description of Metabolic Regulation Combining Aspects of Elementary Flux Mode Analysis with Cybernetic Control Laws*. PhD thesis, Purdue University, USA, 2005.
- [224] S. H. Zak. *Systems and Control*. Oxford University Press, New York, 2003.

Summary

Scaling-up fermentation processes from laboratory-scale conditions to large-scale conditions generally results in a reduction of the overall process yield and productivity. This is due to the interplay of biological, chemical, and physical factors. In this work, different tools have been developed and applied which may help to elucidate the mechanisms causing this generally observed yield reduction.

Then, tools to describe micro-organisms in detail are necessary. Hence, the state of the art approaches for metabolic modelling, typically used in the domain of metabolic engineering, were reviewed. The strategy to be followed for optimising a production host for overproducing a target compound should predominantly depend on its characteristic properties. In this respect, issues like the dependence of the target compound's synthesis on severe (redox) constraints, the characteristics of its formation pathway, and the achievable/desired flux towards the target compound should play a role when choosing the optimisation strategy. Still, due to the vast variety of biochemical pathways and the lack of extensive data sets the usefulness of these mathematical techniques remains limited. In this Ph.D. study some of the reviewed methods have been applied, such as partial least squares, approximative metabolic modelling, and cybernetic modelling.

The usefulness of partial least squares regression has been demonstrated using elementary flux mode data. It was possible to rapidly pinpoint potential targets for modification of the microbial production of succinate by *Escherichia coli*, without the need for experimental data. The identified targets are in agreement with the literature data (modification of the expression of these genes proved to be beneficial to increase succinate yield). This approach has therefore passed a first validation round. Further evaluation is however needed.

Conversely, a dynamic model-based approach focusses on the identification of the flux controlling reactions, which are targets for genetic modifications. In view of decision-making in metabolic engineering, it is important to assess the uncertainty on the calculated flux control coefficients. Both an uncertain model structure and uncertain parameter estimates can be the cause for the overall prediction uncertainty. For an illustrative pathway this uncertainty has been properly assessed. Multiple approximative kinetic formats have been used to identify the flux control coefficients of the small network model studied. It has been shown that the applied model structure significantly influences the distribution of the flux control coefficients.

Micro-organisms in large-scale bioreactors are characterised by a particular metabolomic and proteomic make-up, which allows maximisation of their growth under those conditions, *e.g.*, mixed acid fermentation and overflow metabolism. Since this complies well with the idea behind cybernetic modelling, cybernetic models were finally retained to describe the biophase in large-scale bioreactors. The rationale of the cybernetic school of thought is that micro-organisms are believed to optimise their behaviour, *e.g.*, with respect to growth or substrate uptake. This is achieved by allocating, by means of a controller, the limited resources a micro-organism disposes of to these enzymes yielding the optimal performance. In spite of recent efforts to increase the robustness of the approach, *e.g.*, by introducing elementary flux modes as intermediate level of control, there still remain some issues unresolved. For instance, several rival control laws for enzyme activity have been derived. These rival control laws had a different no-cost activity and are based on the fact that mechanisms have been reported in the literature for both the activation and inactivation of enzymes, which may have a cost. However, due the lack of appropriate data it was not possible to distinguish between those rival control laws.

Subsequently, set-ups are discussed which may help to gather the necessary data to experimentally study microbial metabolism and to gather the necessary data with a view to parameter identification and model structure identification. To this end, a modus operandi of the Bioscope is proposed to study microbial oscillating systems. A strategy has been proposed to control the opening and closing of the sample ports, so that this equipment can also be used to collect the samples from multiple perturbation experiments, without perturbing the microbial oscillating culture from which the cells are taken.

A strategy to design a scaled-down reactor is outlined as well. The innovative aspect

of the presented approach is that it attempts to mimic the environmental conditions observed by the micro-organisms, by making use of computational fluid dynamics simulation results, rather than to focus on macroscopic variables, such as circulation time and mixing time, as those macroscopic variables are far from ideal to be correlated with degrees of conversion. Such scaled-down reactors allow to mimic on a laboratory-scale, the large-scale conditions in an attempt to anticipate the outcome on a large-scale. The proposed controlled set-up, a controlled system consisting of two continuous stirred-tank reactors in a loop, allows to imitate similar conditions as those that occur in large-scale bioreactors. To reduce the control efforts one could use a maximal value for the substrate concentration set points, since the cellular response to environmental concentrations much larger than the affinity constant becomes saturated.

Finally, a method has been proposed to use segregated models, in which micro-organisms are not considered identical, and in which the cells are structured, *i.e.*, the internal composition and structure of the micro-organisms is considered, to describe the biophase in large-scale bioreactors using computational fluid dynamics. The description of the biophase in a Lagrangian way, *i.e.*, following the cell's path through the reactor, is an obvious choice since the behaviour of a micro-organism is determined both by the reigning environmental conditions and its intracellular make-up. This intracellular make-up is expected not to be identical for all micro-organisms, due to the stochastic nature of particle transport and the fast metabolic response to the observed fast changing environmental conditions. Such an approach is computationally quite demanding because every micro-organism is linked to a set of differential equations. However, by considering that the overall picture is merely the result of all individual micro-organisms it is only needed to track a limited number of particles in order to obtain a good idea of the consumption and production of metabolites throughout the large-scale bioreactor. Indeed, the dynamics of the overall system can be captured by averaging out the behaviour of this limited number of particles over the whole population, hereby making use of prior knowledge about the microbial behaviour.

Samenvatting

Het opschalen van microbiële culturen van laboratoriumschaal naar productieschaal leidt in de regel tot een reductie van de procesopbrengst. Dit is te wijten aan een samenspel van biologische, chemische en fysische factoren. In dit werk werden verschillende methodes ontwikkeld die kunnen helpen bij het ontrafelen van de mechanismen die aan de oorsprong liggen van deze reductie.

Voor zo'n studie zijn methodes die toelaten het microbiële metabolisme in detail te beschrijven, belangrijk. Daarom werd een stand van zaken opgemaakt van het metabolisch modelleren. Metabolische modellen worden typisch toegepast in het domein van metabolische engineering voor de optimalisatie van productiestammen met het oog op de overproductie van een doelmolecule. De afhankelijkheid van zijn synthese van redoxbeperkingen, de karakteristieke eigenschappen van zijn syntheseroute en de bereikbare flux richting de doelmolecule zouden een rol moeten spelen bij de keuze van de optimalisatiestrategie.

Het nut van partiële kleinste-kwadraten regressie werd geïllustreerd met behulp van elementaire flux mode data. Zo konden mogelijke doelwitten voor genetische modificatie geïdentificeerd worden met het oog op de microbiële productie van succinaat door *E. coli*. De geïdentificeerde doelwitten zijn in overeenstemming met de literatuur; in dewelke aangetoond werd dat modificatie van de expressie van deze genen leidt tot een verhoogde succinaatopbrengst. Deze aanpak heeft daarom een eerste validatieronde doorstaan. Verdere evaluatie is evenwel nodig.

Een dynamisch metabolisch model focust op het identificeren van de snelheidsbepalende stappen in een reactienetwerk, wat typisch doelwitten zijn voor genetische modificatie. Met het oog op het nemen van beslissingen in metabolische engineering is het belangrijk

om de onzekerheid omtrent de berekende snelheidsbepalende stappen adequaat te kunnen inschatten. Deze onzekerheid kan zowel het gevolg zijn van een onzekere modelstructuur als van onnauwkeurig gekende parameterwaarden. Voor een illustratief reactienetwerk werd deze onzekerheid nagegaan. Meerdere approximatieve kinetieken werden gebruikt om de snelheidsbepalende stappen van het bestudeerde reactienetwerk te identificeren. Hieruit bleek dat de modelstructuur een significante invloed heeft op de distributies van de snelheidsbepalende stappen.

Micro-organismen in een productieschaalreactor worden gekenmerkt door specifieke metabolische en proteomische niveaus die onder die condities toelaten groei te maximaliseren, bijvoorbeeld door overflow metabolisme. Aangezien dit gedrag overeenstemt met het concept dat aan de basis ligt van het cybernetisch modelleren, werd dit type modellen gebruikt om de biofase in dergelijke reactoren te beschrijven. De rationale achter het cybernetisch modelleren is dat micro-organismen hun gedrag optimaliseren met betrekking tot groei door de beperkte middelen waar de cel over beschikt te investeren in die enzymen die een optimaal gedrag verzekeren. Ondanks recente pogingen om de robuustheid van de methode te vergroten, bijvoorbeeld door het introduceren van elementaire flux modes als intermediair regelniveau, blijven toch nog een aantal zaken onopgelost. Een aantal controlewetten voor enzymactiviteit werden afgeleid en geëvalueerd. Deze rivaliserende wetten worden gekenmerkt door een verschillende 'geen kost' activiteit en steunen op het feit dat in de literatuur verschillende mechanismen voor activatie en inactivatie van enzymen werden beschreven, die een kost hebben. Door het gebrek aan geschikte data was het evenwel niet mogelijk om tussen deze wetten een onderscheid te maken.

Vervolgens werden experimentele opstellingen ontworpen die kunnen helpen bij het vergaren van de nodige data om het microbiële metabolisme te bestuderen en om de benodigde data te verzamelen met het oog op het schatten van parameters en het identificeren van een geschikte modelstructuur. Hiertoe werd een *modus operandi* voor de Bioscope voorgesteld, zodat dit apparaat ook kan worden aangewend om oscillerende microbiële systemen te bestuderen. Door het openen en sluiten van de staalnamepoorten zo te regelen dat enkel cellen worden gecollecteerd met een zelfde geschiedenis, kunnen in de Bioscope meerdere perturbatie experimenten worden uitgevoerd, zonder de oscillerende microbiële cultuur waarvan de biomassa afkomstig is, te verstoren. Dit versnelt uiteraard het vergaren van data voor metabolische modellering.

Een strategie werd voorgesteld om een scaled-down reactor te ontwerpen. Het innovatieve aspect van de voorgestelde aanpak is dat deze poogt de door micro-organismen waargenomen omgevingsomstandigheden in productieschaalreactoren na te bootsen, eerder dan te focussen op macroscopische variabelen als mengtijd en circulatietijd. Deze macroscopische variabelen zijn immers verre van ideaal om gecorreleerd te worden aan omzettingsgraden, waar het uiteindelijk om gaat. Een dergelijke scaled-down reactor laat toe om op laboratoriumschaal de omstandigheden na te bootsen in een productieschaal reactor, zodat reeds geanticipeerd kan worden op het resultaat op productieschaal. De voorgestelde experimentele opstelling, die uit twee geregelde reactoren in een kring bestaat, laat toe om gelijkaardige omstandigheden na te bootsen als die in productieschaalreactoren. Om de regelacties te reduceren zou men een maximale waarde voor de substraatconcentratie wenswaarde kunnen gebruiken, aangezien de cellulaire respons op substraatconcentraties veel groter dan de affiniteitsconstante een plateau bereikt.

Tot slot werd een methode voorgesteld die toelaat om zowel de interne structuur en samenstelling van micro-organismen als de heterogeniteit van de microbiële populatie in een productieschaalreactor via stromingsdynamica modellen te beschrijven. Bij de Lagrangiaanse aanpak wordt het pad van een micro-organisme doorheen de reactor gevolgd. Daar het gedrag van een micro-organisme zowel wordt bepaald door de heersende omgevingsomstandigheden als door zijn intracellulaire toestand, ligt het voor de hand om deze aanpak te gebruiken voor het beschrijven van de biofase in zo'n reactor. Door de stochastische aard van partikeltransport en de snelle metabolische respons op de snel variërende omgevingsomstandigheden, wordt deze toestand niet geacht identiek te zijn voor alle micro-organismen. Een dergelijke aanpak is evenwel computationeel veeleisend, aangezien elk micro-organisme gelinkt is met een stelsel differentiaalvergelijkingen. Door te beschouwen dat het gedrag van de totale populatie niet meer is dan de resultante van de individuele micro-organismen is het enkel nodig om een beperkt aantal cellen te volgen om een goed idee te krijgen van de consumptie- en productiesnelheden doorheen de reactor. De dynamiek van het systeem kan immers gevat worden door het gedrag van dit beperkt aantal cellen lokaal uit te middelen over de gehele populatie. Hierbij werd gebruik gemaakt van voorkennis omtrent het microbiële gedrag, bijvoorbeeld het feit dat saturatie van de substraatopnamesnelheid optreedt bij substraatconcentraties die vele malen groter zijn dan de affiniteitsconstante.

Curriculum vitae

Jo Maertens

Studies

2007 - Present: Scientific collaborator at the Department of Applied Mathematics, Biometrics and Process Control (Ghent University)

2003 - 2007: Research assistant of the Fund for Scientific Research Flanders at the Department of Applied Mathematics, Biometrics and Process Control (Ghent University)

1998 - 2003: Bio-Engineer in Chemistry (Magna cum laude)

Faculty of Agricultural and Applied Biological Sciences, Ghent University

- Additional courses

2001-2002: French (Ghent University, Talencentrum)

2003-2004: Deterministic Models in Operational Research (Ghent University, Prof. Vanmaele)

2003-2004: Metabolic Engineering and Functional Genomics (TU Denmark, Prof. Nielsen)

2004-2007: Spanish (PCVO)

Publications

- S. W. H. Van Hulle, S. Van Den Broeck, J. Maertens, K. Villez, B. M. R. Donckels, G. Schelstraete, E. I. P. Volcke, and P. A. Vanrolleghem. Construction, start-up and operation of a continuously aerated laboratory-scale SHARON reactor in view of coupling with an Anammox reactor. *Water SA*, 31: 327 – 334, 2005. (IF = 0.445, TC = 7)

- M. De Mey, G. Lequeux, J. Maertens, S. De Maeseneire, W. Soetaert, and E. Vandamme. Comparison of DNA and RNA quantification methods suitable for parameter estimation in metabolic modeling of microorganisms. *Analytical Biochemistry*, 353, 198 – 203, 2006. (IF = 2.67, TC = 4)
- G. Lequeux, L. Johansson, J. Maertens, P. A. Vanrolleghem, and G. Lidén. MFA for overdetermined systems reviewed and compared with RNA expression data to elucidate the difference in shikimate yield between carbon- and phosphate-limited continuous cultures of *E. coli* W3110.shik1. *Biotechnology Progress*, 22, 1056 – 1070, 2006. (IF = 1.985, TC = 2)
- M. De Mey, J. Maertens, G. J. Lequeux, W. K. Soetaert, and E. J. Vandamme. Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnology*, 7, 34, 2007. (IF = 2.74, TC = 3)
- M. De Mey, G. J. Lequeux, J. J. Beauprez, J. Maertens, E. Van Horen, W. K. Soetaert, P. A. Vanrolleghem, and E. J. Vandamme. Comparison of different strategies to reduce acetate formation in *Escherichia coli*. *Biotechnology Progress*, 23, 1053 – 1063, 2007. (IF = 2.102, TC = 1)
- M. De Mey, G. J. Lequeux, J. Maertens, C. I. De Muynck, W. K. Soetaert, and E. J. Vandamme. Comparison of protein quantification and extraction methods suitable for *E. coli* cultures. *Biologicals*, 36, 198–202, 2008. (IF = 0.921, TC = 0)
- F. Mestdagh, J. Maertens, T. Cucu, K. Delporte, C. Van Peteghem, and B. De Meulenaer. Impact of additives to lower the formation of acrylamide in a potato model system through pH reduction and other mechanisms. *Food Chemistry*, 107, 26 – 31, 2008. (IF = 2.433, TC = 2)
- B. M. R. Donckels, D. J. W. De Pauw, B. De Baets, J. Maertens, and P. A. Vanrolleghem. An anticipatory approach to optimal experimental design for model discrimination. *Chemometrics and Intelligent Laboratory Systems*, submitted for publication, 2008.
- J. Maertens and P. A. Vanrolleghem. Identification and evaluation of approximative kinetic model structures. *BMC Bioinformatics*, submitted for publication, 2008.

Conferences and research stays

- 2006, October 30 - November 3: Conference '3rd International *E. coli* Alliance Conference on Systems Biology' (Jeju Island, Republic of Korea)
- 2006, May 1 - 31: Research stay (West Laffayette, Indiana, USA)
Purdue University
School of Chemical Engineering (Prof. D. Ramkrishna)
- 2005, June 19 - 24: Conference: 'Computational Fluid Dynamics in Chemical Reaction Engineering' (Barga, Italy)
- 2004, September 19 - 23: Conference: 'Metabolic Engineering V: Genome to Product' (Lake Tahoe, California, USA)
- 2004, June 1 - 4: Research stay (Lund, Sweden)
Lund University
Department of Chemical Engineering (Prof. G. Lidén)
- 2004, May 22 - 28: Ph.D. course: 'Metabolic Engineering and Functional Genomics' (Lyngby, Denmark)
Technical University of Denmark
Center for Microbial Biotechnology (Prof. J. Nielsen)
- 2003, February 9 - 22: Intensive course: 'Agriculture: Source of Raw Material for Industry' (Vienna, Austria)
Universität für Bodenkultur (Prof. R. Verhe)
- 2002, September 22 - 27: Research stay (Nijmegen, The Netherlands)
Katholieke Universiteit Nijmegen
Department of Microbiology (Prof. M.S.M. Jetten, Dr. M. Strous)
- 2002, February - June: Erasmus (Montpellier, France)
l' Institut des Sciences de l' Ingénieur de Montpellier (Prof. P. Charlier)
l' Ecole Nationale Supérieure de Chimie de Montpellier