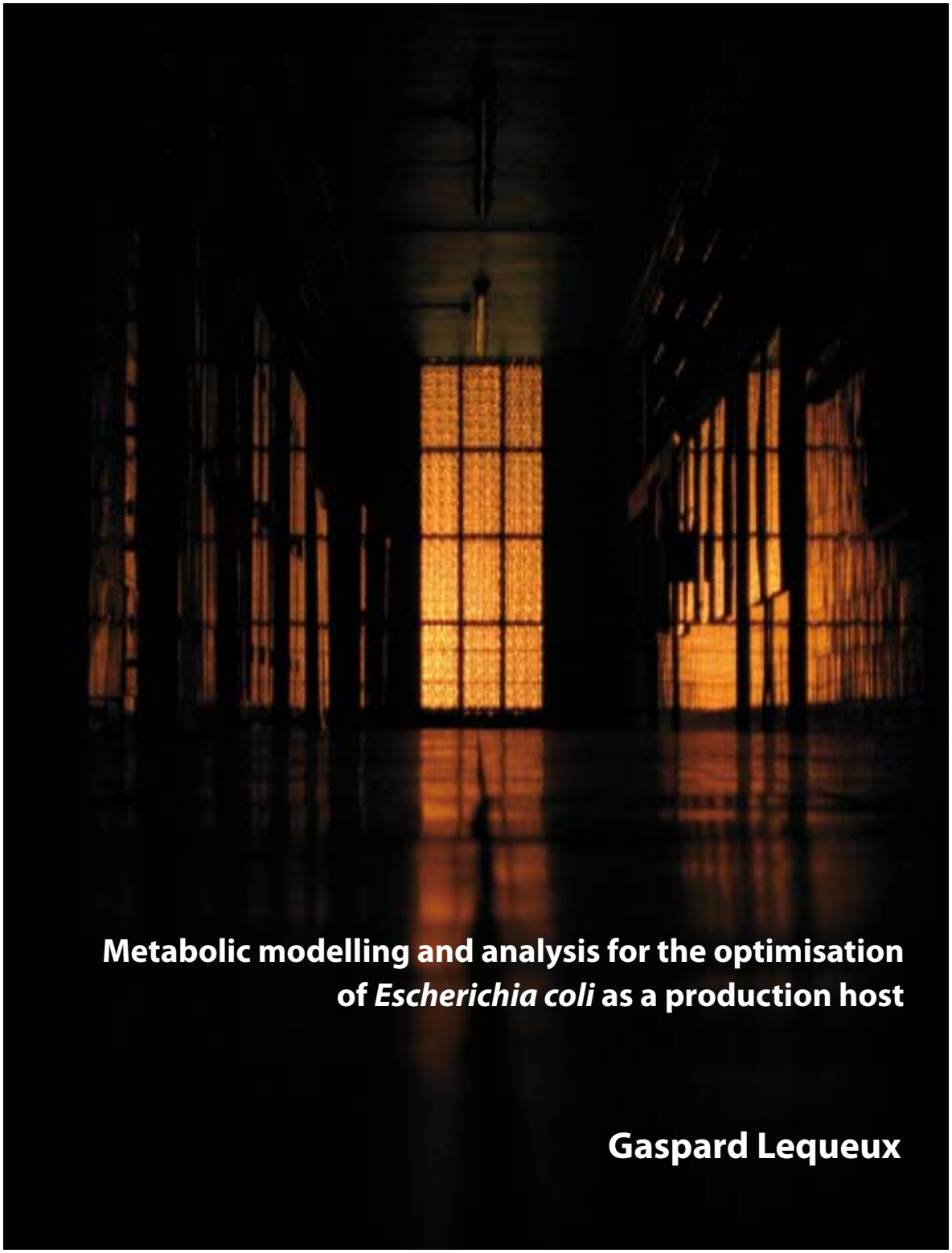


Metabolic modelling and analysis for the optimisation of *Escherichia coli* as a production host

Gaspard Lequeux

UNIVERSITEIT GENT

FACULTEIT BIO-INGENIEURSWETENSCHAPPEN

Examination Committee

Prof. dr. Jean-Pierre Ottoy (Ghent University)
Em. Prof. dr. Raymond Cunin (Vrije Universiteit Brussel)
Prof. dr. ir. Gunnar Lidén (Lund University, Sweden)
dr. ir. Rene van der Heijden (SPAD it, The Netherlands)
Prof. dr. ir. Wim Soetaert (Ghent University)
Prof. dr. ir. Erick Vandamme (Ghent University)
Prof. dr. Patrick Van Oostveldt (Ghent University)

Supervisor

Prof. dr. ir. Peter Vanrolleghem (Department of Applied Mathematics, Biometrics and Process Control, Ghent University)

Dean

Prof. dr. ir. Herman Van Langenhove

Rector

Prof. dr. Paul Van Cauwenberge

Gaspard Lequeux

Metabolic modelling and analysis for the optimisation of *Escherichia coli* as a production host

Thesis submitted in fulfilment of the requirements for the degree of
Doctor (Ph.D.) in Applied Biological Sciences, Cell and Gene Biotechnology.

Dutch translation of the title:
Metabolische modellering en analyse voor de optimalisatie van Escherichia coli
als productiestam

Please refer to this work as follows: Gaspard Lequeux, 2008. Metabolic modelling and analysis for the optimisation of *Escherichia coli* as a production host. Ph.D. thesis, Ghent University, Ghent, Belgium.

ISBN: 978-90-5989-219-4

Woord vooraf

Een doctoraat schrijven gebeurt niet alleen. Een heel systeem van bureau- en laboratten was hiervoor nodig. Vooreerst mijn promotor, *let's move to Canada* Peter, die mij de kans gaf om te doctoreren, en later nooit het vertrouwen verloor om de verschillende experimenten tot een goed einde te brengen. Gunnar, his fine humor and subtle way of making scientific research publishable. "De Prof" Erick Vandamme en *laten we een watergevecht houden* Wim, voor de mooie bruggen tussen Limab en Biomath.

Het begon met de ingenieursthesis. Peter moest met een plan B op de proppen komen. Iets dat zeker zou lukken. Metabolische modellering dan maar. De eerste artikels, waren niet zo interessant, niet technisch genoeg. Maar dan kwamen die papers van Noorman en dat ongepubliceerd werk... Ik was verkocht. En Peter blijkbaar ook, want ik mocht blijven en mee aan de wieg staan van de metabolische groep binnen Biomath. En iets insturen naar een conferentie. Maar ja, dat ongepubliceerd werk waar zoveel op gebaseerd was, hoe verwijs je daarnaar? Gelukkig was Rene een goede vriend van Peter, en voila, mijn eerste internationale samenwerking. Pas 4.5 jaar later hebben we elkaar in levende lijve ontmoet.

Samen met Sam als top van Biomath (dixit Peter) de metabolische modelleringsgroep uit de grond proberen stampen. Allebei bevestigen we uiteindelijk de trend die we waarnamen onder assistenten... Later kwam daar Jo bij —de fratsen die we samen uitgehaald hebben zijn niet meer te tellen—, Aditya, *de voorzienige* en Brecht *de wanhopige, maar zijn collegae zorgen wel voor hem.* For a short time we had Joel; without him, the rest of Biomath would never know what we were doing. Gino, erbij gekomen anno 2007 en op het laatste nippertje nog heel het doctoraat gelezen. Tinne, voor de duizenden stikstofanalyses. En dan de ontelbare bureaugenoten en de rest van de vakgroep, voor de gezellige sfeer, de sappige roddels, de *laten we elkaar in de plassen duwen* lentewandelingen (die traditioneel plaatsvinden na 21 juni)...

Natuurlijk geen modelleren zonder data. Daarvoor was de samenwerking met Limab onontbeerlijk. Initieel het project schrijven met Marjan en Sofie Vandedrinck, later de kleurrijke samenwerking met Marjan. Uiteindelijk is het ons toch gelukt om modellen en experimenten samen te brengen. De *wij zijn beter dan Sartorius*[1] *en gaan het lab ombouwen* samenwerking met Joeri, met het legendarische hoogtepunt toen we een bug vonden in het stuurprogramma

[1] Welliswaar dankzij de hulp van Ralf Bruene van Sartorius, Duitsland

van de nieuwe reactor. Katja die zo graag de experimenten in vijfvoud herhaalde. Ellen, de relativerende noot en RNA extractie experte waar iedereen om vecht. *Laten we het chocolade popje van mijn thesisbegeleider opeten* Hendrik. Karl, we weten nu nog altijd niet waar je hond begraven is. Evelien, hebben we het labo niet eens onder water gezet? De memorabele memo-avonden. En natuurlijk de rest van de Limab, voor de gezellige sfeer en het gemak waarmee jullie op de kasten te jagen zijn.

Dankzij Peter, een Zweedse excursie. The welcome feeling of *some things you don't want to know and let's celebrate the breakthrough result today because tomorrow we will find a reason why it is not that revolutionary* Louise and Thomas, their fantastic wedding party. The Christmas eve modelling sessions with Andreas. *Swedish tradition* Tobias, sorry to introduce you to Vim and the regular expression world; you probably lost much time with those interesting toys. Anneli, present everywhere in Sweden and Denmark. Klaus, after months we realised that we both enjoyed reading Russian literature (well, we used to, before our phd-time). All the people working at the Chemical Engineering Department, for the nice atmosphere, the much appreciated coffee breaks (well, for me, start-of-the-day event), the legendary Christmas and Crayfish parties and so much more.

Walter van TUDelft, voor de elementenanalyse die je er tussendoor geduwd hebt.

David, voor de opvang tussen de twee Zweedse periodes, voor de leuke trektocht samen, de goeie babbels (over wat zullen we maar in het midden laten :)).

Mijn ouders, zussen en broer, die ervoor zorgden dat ik mij nooit thuis voelde in Gent: te gehecht aan het ouderlijke thuis. De wekelijkse voedselbedeling, de back-up wanneer de treinen weer eens staakten, de geanimeerde reality-soap maaltijden: zal broer een motorfiets kopen of niet? De geïnteresseerde vragen over het doctoraat waarop het standaard antwoord was: hmmm.

Veerle, dé stimulans om dit doctoraat op tijd af te werken (*E. coli* experimenten, computers en zelfs Linux zijn onbelangrijk geworden; en MS Windows problemen werden plezant om op te lossen, weliswaar ligt dat niet aan Microsoft). Een dag Gentse feesten werd systematisch begonnen en beëindigd met het controleren van de *E. coli* culturen. Elke morgen kwam je mij op tijd halen (ik ben nooit zo regelmatig toegekomen op den bureau); op het einde had je door dat ik volledig geen uren meer had en liet je mij de vrijheid om mijn 25 uren per dag dagen te (over)leven. Je hebt me overgehaald om naar de voetbal te gaan kijken, emh, mee te gaan als fotograaf. Je was de enige aan wie ik mijn problemen kwijt wou, tijdens de legendarische wandelingen... Dank ook aan je ouders en zus voor het tweede thuis.

Gaspard Lequeux
Gent, 17 januari 2008

Samenvatting

Metabolische modellering werd gebruikt voor het onderzoeken van verschillende strategieën ter optimalisatie van *Escherichia coli* als productiestam. In het proefschrift wordt eerst een overzicht gegeven van Metabolische Flux Analyse (MFA). Ten eerste worden verschillende technieken uitgelegd om de kwaliteit van een metabolisch model te testen (elementaire samenstellingstest, zoeken naar doodlopende reactieroutes, parallelle reactieroutes detecteren). Ten tweede wordt uitgelegd hoe een metabolisch model algemeen kan opgelost worden waarbij de fluxen gebalanceerd worden. Verder wordt besproken hoe de oorzaak van inconsistente meetingen gevonden kan worden via de vectorvergelijkingstest.

Een nadeel bij het gebruik van MFA is dat het alleen toepasbaar is op culturen in evenwichtstoestand terwijl vele interessante processen niet in evenwicht zijn. Daarom werd een nieuwe techniek voorgesteld, dynamische MFA, waarin tijdreeksen van transiënte experimenten worden omgezet naar een vorm bruikbaar voor MFA. In dynamische MFA wordt de afgeleide genomen van de concentratieprofielen op elk tijdspunt onder studie. Deze afgeleide wordt gebruikt voor het berekenen van de netto uitwisselingsfluxen die dan gebruikt worden in de MFA analyse. Maar het berekenen van afgeleiden leidt tot een vergroting van de ruis aanwezig in de data. Daarom moeten geschikte vereffeningnsalgoritmes toegepast worden. De voorgestelde polynoomvereffening vertoont enkele gebreken en vereist vele manuele afstellingen. Desondanks werden uiteindelijk toch bruikbare resultaten verkregen.

De hierboven beschreven wiskundige technieken werden toegepast op een aantal interessante biologische en industrieel relevante gevallenstudies.

Een *E. coli* stam, genetisch gemodificeerd om meer shikiminezuur te produceren (shikiminezuur is een basismolecule voor het maken van Tamiflu®, een antiviraal geneesmiddel), heeft een hogere shikiminezuur opbrengst onder koolstofrijke condities dan onder koolstofgelimiteerde condities. Met behulp van MFA werd aangetoond dat dit niet veroorzaakt wordt door een lagere koolstofflux naar de aromatische aminozuur syntheseroute (shikiminezuur is een nevenproduct van die route), maar door een hogere excretie van andere nevenproducten in die route. Door het gebruik van MFA kon de intracellulaire ATP hydrolyse flux berekend worden voor verschillende groeisnelheden van koolstof- en fosforgelimiteerde culturen zodat

de onderhoudscoëfficiënten voor beide soorten culturen konden bepaald worden. Er werd aangetoond dat *E. coli* cellen onder koolstofrijke condities geen groeigeassocieerde onderhoudsbehoeften hebben, maar dat ze wel de maximaal mogelijke hoeveelheid ATP verbruiken zodat hun niet-groeigeassocieerde onderhoudsbehoefte ongeveer even hoog is als de totale onderhoudsbehoefte van koolstofgelimiteerde culturen gekweekt bij groeisnelheden bijna even hoog als de uitspoelsnelheid.

Een typisch probleem bij *E. coli* culturen is het acetaat overloop metabolisme. Verschillende genetische strategieën werden onderzocht om deze ongewenste acetaatproductie tegen te gaan. Het uitschakelen van de acetaatproductieroute zorgde voor een lagere acetaatproductie, maar er werd meer lactaat in de cultuurvloeistof teruggevonden. De strategie waarbij het fundamentele probleem dat er niet genoeg koolstof van de glycolyse naar de citroenzuurcyclus gaat, werd verholpen door overexpressie van *ppc*, coderende voor PEP carboxylase, en gaf duidelijk betere resultaten. Uiteraard gaf de combinatie van de twee methoden ook goede resultaten.

Vervolgens werd onderzocht of de *ppc* overexpressiemutant een betere stam was voor het produceren van recombinant eiwit. Het voorbeeldeiwit dat hiervoor gebruikt werd, was β-galactosidase. Een plasmide met daarop *LacZ* (coderende voorβ-galactosidase) tezamen met een IPTG induceerbare promotor, werden ingebracht in *wild-type* cellen en in *ppc* overexpresserende mutanten. Van beide stammen werden chemostaatculturen gekweekt. Bij het bereiken van de evenwichtstoestand werd de productie van recombinant eiwit gestart door het toevoegen van IPTG aan de cultuurvloeistof. Een twaalf keer hogere eiwitproductie werd waargenomen in de *ppc* overexpresserende stam.

Het gedrag van koolstofgelimiteerde continue culturen die werden omgeschakeld naar NH_3-limitatie en *vice versa*, werd onderzocht met behulp van dynamische MFA. Wanneer omgeschakeld werd van stikstoflimitatie naar stikstofovervloed (koolstoflimitatie), werd een lagfase van meerdere uren waargenomen waarin de cellen bijna ophielden met groeien. Hiervoor kon geen goede verklaring gegeven worden, maar de lengte van de lagfase (5 uren) suggereert dat de oorzaak waarschijnlijk genetisch is. In de literatuur werd dergelijke lagfase totnogtoe enkel waargenomen bij *glnE knock-out* mutanten. Het gen *glnE* encodeert ATase, het eiwit verantwoordelijk voor de regulatie van glutamine synthetase. Wanneer onder stikstofrijke condities ATase niet meer werkt, ontstaat er toxiciteit van NH_3 te wijten aan accumulatie van glutamine/glutamaat. Deze lagfase werd niet waargenomen bij het omschakelen van koolstofgelimiteerde culturen naar stikstoflimitatie.

Het concept van de chemostaat werd uitgebreid door aan continue culturen oscillerende omgevingscondities op te leggen. Deze oscillaties bleven in dit doctoraatsonderzoek beperkt tot wisselingen tussen aerobiosis en anaerobiosis. Twee perioden van oscillatie werden uitgetest: 4 minuten en 30 minuten. In beide gevallen was de eerste helft van de periode aeroob en de

tweede helft anaeroob. Met behulp van Affymetrix® microarrays werden genexpressieprofielen gegenereerd op verschillende tijdstippen in de oscillatie. Deze profielen werden vergeleken met die van volledig aerobe en volledig anaerobe culturen. Uit de 30 minuten experimenten bleek dat cellen zich sneller aanpasten aan aerobiosis dan aan anaerobiosis. In de experimenten met een oscillatieperiode van 4 minuten, was het verschil tussen de genexpressieprofielen in de aerobe en anaerobe fase minimaal en de expressieprofielen van beide fasen leken meer op die van de volledig aerobe culturen dan op die van de volledig anaerobe culturen.

Dit werk geeft een overzicht van MFA en breidde er de toepasbaarheid van uit naar dynamische experimenten. Het nut van de voorgestelde methodologieën werd geïllustreerd aan de hand van biologisch interessante en industrieel relevante voorbeelden. Daarnaast werd een nieuwe experimentele opstelling uitgeprobeerd: het gecontroleerd laten oscilleren van omgevingscondities bij chemostaatculturen. Deze opstelling gaf veelbelovende resultaten.

Summary

Different strategies for the optimisation of *Escherichia coli* as production organism were investigated using metabolic modelling tools. An overview of the state of the art of metabolic flux analysis (MFA) is given, starting with different metabolic model check methods (elemental composition check, dead-end detection, parallel pathway identification) and continuing with the general technique for solving overdetermined metabolic models and balancing the measured fluxes or eventually, in the case of flux data being inconsistent between each other, finding the wrong measurement with the vector comparison test.

A drawback of MFA is that it is only applicable to cultures in pseudo steady state. Therefore a novel technique, called dynamic MFA, is presented. Dynamic MFA transforms time series of transient experimental data to a form suitable for MFA. Dynamic MFA is based on taking the derivative of the concentration profile at each time instant of interest and using this derivative to calculate net fluxes, that are subsequently used in the MFA framework. However, taking the derivative increases the noise on the data, thus suitable data smoothing algorithms have to be used. Polynomial smoothing is proposed, but it has some difficulties and much manual tuning is needed. Still, acceptable results were obtained.

These mathematical techniques were then applied on different biologically interesting and industrially relevant cases.

E. coli strains genetically engineered to produce more shikimate (an interesting starting compound for a number of chemicals, *e.g.* oseltamivir phosphate, the active compound of Tamiflu®, a drug against influenza), has a higher shikimate yield when grown under carbon-abundant conditions (phosphor-limited cultivations) than under carbon-limiting conditions. Using MFA, it is shown that this is not due to a lower flux towards the aromatic pathway (shikimate is an intermediary compound of this pathway), but is caused by the increased excretion of other metabolites involved in the aromatic pathway. Also, having access to internal flux values for ATP hydrolysis at different growth rates for cultures under carbon- and phosphor-limitation, the maintenance parameters are calculated for both types of cultivations. It is suggested that *E. coli* cells under carbon-abundance have no growth-associated maintenance but are utilising the maximal possible amount of ATP and the non-growth-

associated maintenance is as high as the total maintenance requirements of cells grown under carbon-limiting conditions near the wash-out rate.

A typical problem of *E. coli* cultivations is the acetate overflow metabolism. Different genetic strategies were assessed to reduce this waste acetate production. When disabling the acetate pathway, less acetate was indeed produced, but more lactate was found in the reaction broth. The approach in which the more fundamental problem of insufficient carbon going from the glycolysis to the Krebs cycle is remedied by overexpressing *ppc*, encoding PEP carboxylase, gave better results. Obviously, the third approach in which both genetic strategies were combined, also performed well.

Next, it was assessed whether a *ppc* overexpressing mutant had improved capabilities for producing recombinant proteins. The model protein used was β-galactosidase. A plasmid containing *LacZ* (encoding β-galactosidase) with an IPTG inducible promoter, was introduced in wild-type cells and *ppc* overexpressing strains. For both, chemostats were run, and when steady state was attained, IPTG was added to the reactor broth, starting the recombinant protein production. A twelve times higher protein production flux was observed in the *ppc* mutant.

The intracellular flux distributions of carbon-limited continuous cultures switched to NH_3-limitation and *vice versa*, was investigated using dynamic MFA. When switching from nitrogen-limitation to nitrogen-abundance (carbon-limitation), a lag phase of several hours in which the growth of cells almost stopped, was observed. No clear reason could be given for this. However, the extended period of the lag phase (5 hours) suggests that the cause is probably genetic. So far, such lag phase has only been described in literature for *glnE* knock-out mutants. Gene *glnE* encodes ATase, the protein responsible for the regulation of glutamine synthetase. When, under nitrogen excess, ATase is not functioning anymore, the toxicity of NH_3 is due to accumulation of glutamine/glutamate. Such a lag phase was not observed in the case of carbon-limited cultures switched to nitrogen-limitation.

Extending the concept of chemostats, continuous cultures with oscillating environmental conditions were run. The oscillations were between aerobiosis and anaerobiosis. Two periods were chosen for the oscillations: 4 minutes and 30 minutes and in each case half of the period was aerobic, the other half being anaerobic. Gene expression profiles, using Affymetrix® microarrays, were generated at different time points in the oscillations and compared to the gene expression profiles of fully aerobic and anaerobic cultures. In the 30 minutes oscillating experiment, it was observed that cells adapted faster to aerobiosis than to anaerobiosis. In the experiment where the oscillation period was 4 minutes, the difference in gene expression profile between the aerobic phase and the anaerobic phase was minimal and the expression profile of both phases were more similar with the fully aerobic culture than the fully anaerobic one.

This work reviewed and extended the applicability of MFA to dynamic experiments. The usefulness of the methodologies presented were illustrated on different biologically interesting and industrially relevant case studies. Furthermore a novel kind of experiments was conducted: applying controlled environmental oscillation on chemostat cultures. This setup generated promising results.

Contents

Chapter 1

General introduction

Micro-organisms are more and more used for the production of chemicals, referred to as white biotechnology. Indeed micro-organisms have many advantages compared to the classical oil-based chemistry: they utilise renewable resources (thus are carbon neutral), do not produce highly toxic waste products, are operated under gentle conditions (no high pressures or high temperatures needed) and can produce many different compounds. For fine chemistry and food chemistry they have the advantage that molecules with the right chirality are produced.

Although micro-organisms have the potential to produce almost every imaginable organic chemical, a major disadvantage is the sometimes assiduous way in which strains have to be engineered to make the production of the wanted chemical economically viable. However, this is not much different from the classical chemistry 100 years ago, when the concept of atoms was barely understood. The same now happens with biochemistry. Fifty years ago, the basic structure of DNA was unveiled. Now the exponential growth phase of biochemistry is happening: whole genomes are sequenced, production can be improved by direct mutations instead of random mutations with selection pressure, and even completely synthetic biology is emerging (Benner & Sismour, 2005).

To successfully use *E. coli* as a production host, knowledge of metabolism and genetics is essential. The quest for knowledge can be directly motivated by a practical problem (*e.g.* why is the shikimate yield different for different limiting conditions?), but can also be steered by curiosity (*e.g.* what happens during transient conditions? How do the cells react to oscillating environmental inputs?). Both types of questions are addressed in this work.

Cultivation experiments generate much data and it is not always easy to make sense of them. Therefore, mathematical techniques are needed to analyse and interpret those data. The mathematical tools should be able to integrate experimental data with (biochemical) knowledge already available, generating additional insight in the functioning of the microbial cell.

One such technique that combines experimentally measured extracellular fluxes with metabolic knowledge, is metabolic flux analysis (MFA). MFA allows calculating intracellular fluxes when only extracellular metabolite concentration measurements are available. Knowledge of intracellular fluxes allows one to better understand where and how all the nutrients entering the cell, are used.

A key drawback of metabolic flux analysis is that it can only be applied on pseudo steady state cultures while interesting phenomena can be observed during transient environmental conditions. Therefore, a mathematical transformation of the metabolite concentration measurement time series was proposed that allows to use dynamic data in MFA calculations. Dynamic MFA opens a whole new field of experimentation involving flux calculations. The power of observing internal fluxes with only extracellular concentration measurements is not limited anymore to chemostat cultures, but is now also available for dynamic cultivations.

With more and more laboratories having access to transcriptomics, proteomics and even metabolomics, reproducible (microbial) cell cultures are increasingly used (Hoskisson & Hobbs, 2005). In the 1960s, chemostats were essentially used to characterise the kinetics of cellular growth. However, it appears that no unique value for kinetic parameters can be obtained experimentally and this is not due to experimental limitations, but is inherent to bacteria, as they are experts at adapting to changing growth conditions (Ferenci, 1999). The inventors of the chemostat (Novick & Szilard, 1950), were aware from the beginning that evolutionary pressure limited the number of generations that a strain could be cultivated and behave similarly. In this work chemostats were used to elucidate a number of metabolic questions regarding the production of interesting compounds.

Outline of the thesis

In the whole thesis it is tried to understand the inner workings of the cellular mechanisms, mostly on metabolic level, but also on genetic level.

For the metabolic understanding, the powerful technique of metabolic flux analysis is used. A key stone of metabolic flux analysis is the metabolic model. The correctness of this model is crucial to be able to trust the results obtained in the metabolic flux analysis. To increase the confidence one can have in the metabolic model and the subsequent results obtained with it, and to facilitate the development of such a metabolic model, chapter 2 reviews different tools and checks. The second part of this chapter reviews the methods for generating intracellular flux data based on the extracellular measurement data. Different statistical tests are discussed. The chapter ends with an overview of how all those methods were implemented in order to automate the applications of those tools and checks as much as possible. The software toolbox is versatile enough for use in a multitude of cases where metabolic modelling can help in understanding the cultivation processes.

Shikimate is an interesting starting compound for a number of chemicals (*e.g.* oseltamivir phosphate, the active compound of Tamiflu®, a drug against influenza). When cultivating *E. coli* strains, genetically engineered to produce shikimate, more shikimate is obtained under carbon-abundant conditions (phosphorus limited) than under carbon-limited conditions. To elucidate the biochemical reason for this, at first sight, strange behaviour, a number of chemostats at different dilution rates and for the two different limitations were run. The data were analysed with MFA and interesting observations were made (chapter 3).

One of the major drawbacks when using *E. coli* cultivations is the unwanted production of acetate. In chapter 4 different genetic strategies for reducing acetate were proposed and tested. The first strategy disabled the acetate production pathway, the second strategy channeled more carbon to the citric acid cycle and the third mutant was a combination of both. Metabolic flux analysis was used to access the possible differences in flux distributions for those three mutants. A strain is proposed that excretes far less fermentative metabolites and that is usable as generic production host.

This strain was tested for its suitability as protein production host in chapter 6. A *LacZ* (encoding β-galactosidase, chosen as model protein) plasmid with an IPTG inducible promoter was introduced in the *E. coli* mutant. After the cells were in steady state, IPTG was added and the reactor was intensively sampled. The experiment was performed with the mutated strain and with the wild-type. Dynamic MFA was used to combine transient concentration measurements with biological network topology.

Chapter 5 introduces dynamic MFA and applies it to study the intracellular effects of transients between two different steady states. Nitrogen-limited chemostats were switched to carbon-limitation and *vice versa*. Some interesting phenomena were observed and are discussed.

Succinate is an interesting industrial compound, with applications in food and pharmaceutical products, surfactants and detergents, green solvents and biodegradable plastics (Zeikus *et al.*, 2004). Under anaerobic conditions, valuable amounts of succinate are produced, however, the process productivity is low (Lin *et al.*, 2005). It was hypothesised that oscillating between aerobic and anaerobic conditions, could increase the succinate yield. On a more fundamental level, those oscillations should show some interesting phenomena going on in the transcriptome. Transcriptome analysis during transition from aerobic to anaerobic conditions and *vice versa* has been done (Partridge *et al.*, 2006, 2007), but the effects of oscillations was never investigated. Therefore, to study the expression profiles of different genes of central carbon metabolism, microarrays were hybridised with cDNA made from RNA extracts from broth samples taken at different time points in the period of the oscillation. Besides fully aerobic and anaerobic continuous cultures, two periods for oscillations were selected. One

experiment was performed in which aerobiosis and anaerobiosis were alternated with a period of 4 minutes and one experiment in which this period was 30 minutes (chapter 7).

This dissertation ends with an overview of the main conclusions drawn from the methodological, modelling and experimental work. Perspectives for future research are proposed as well (chapter 8).

Part I

MFA

Chapter 2

Building and solving steady state metabolic flux models[1]

Metabolic flux analysis (MFA) can be used to get more knowledge about the flux distribution of intracellular reactions. MFA is frequently applied on underdetermined systems, *i.e.* the solution of the model can only be calculated under certain assumptions *e.g.* maximal biomass production (Savinell & Palsson, 1992). But metabolic models can also be used when the system is overdetermined, *i.e.* when there are more measurements available than degrees of freedom (van Gulik *et al.*, 2000). The statistical methods discussed in this chapter were initially developed for black box models (Wang & Stephanopoulos, 1983; van der Heijden *et al.*, 1994b; van der Heijden & Heijnen, 1995), but can easily be applied to metabolic models as well (Romein, 2001).

After an introduction about the stoichiometric matrix and the nullspace, different model checks are reviewed. Then the most general method for solving metabolic models is presented. It will be clarified that one can still calculate some fluxes and perform statistical tests to check whether the measurement dataset is consistent even if parts of the model are underdetermined (Romein, 2001). The presented methodologies here are more general than what can be found in Stephanopoulos *et al.* (1998) and Nielsen *et al.* (2003).

[1]Parts of this chapter were published as G. Lequeux, R. van der Heijden, S. Van Den Broeck & P. A. Vanrolleghem (2004). Computational methods to determine conserved moieties and parallel pathways in metabolic network models. In *Proceedings 9th IFAC Conference on Computer Applications in Biotechnology CAB9*. Nancy, France, March 28–31 and in G. Lequeux, L. Johansson, J. Maertens, P. A. Vanrolleghem & G. Lidén (2006). 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

2.1 Mathematical representation of metabolic models

Different ways of summarising biochemical network information can be found in literature. Two main tendencies can be discerned: the one with the stoichiometric matrix *sensu stricto*, S (Stephanopoulos *et al.*, 1998; Nielsen *et al.*, 2003), and the more general representation with the extended stoichiometric matrix, W (Noorman *et al.*, 1991). Both representations are useful: the S matrix is used for searching parallel pathways, while the W matrix is used for solving the metabolic model.

2.1.1 The stoichiometric matrix S

A set of reaction equations can be described by the following matrix equation:

$$\frac{\mathrm{d}\boldsymbol{c}}{\mathrm{d}t} = \mathsf{S}\,\boldsymbol{v} - \boldsymbol{r} \tag{2.1}$$

Herein matrix S is the stoichiometric matrix (constructed as explained in equation 2.2), vector $\boldsymbol{v}$ contains the reaction rates and vector $\boldsymbol{r}$ contains the exchange rates (*i.e.* the net production or consumption rates) of the different metabolites. Vector $\boldsymbol{c}$ represents the intracellular concentrations of the metabolites.

When pseudo steady state is assumed, there is no difference between the net production and consumption rates so the left side of equation 2.1 can be set to zero and be rewritten more explicitly as:

$$\begin{bmatrix} S_{1,1} & S_{1,2} & \dots & S_{1,m} \\ S_{2,1} & S_{2,2} & \dots & S_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p,1} & S_{p,2} & \dots & S_{p,m} \\ S_{p+1,1} & S_{p+1,2} & \dots & S_{p+1,m} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,1} & S_{1,2} & \dots & S_{n,m} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \\ r_{p+1} \\ \vdots \\ r_n \end{bmatrix} \tag{2.2}$$

The m different columns of the stoichiometric matrix S represent the m different reactions with corresponding rates v_1 to v_m. Of the n metabolites involved in the network, there are p metabolites that are considered to be exchangeable with the environment (the first p rows of the stoichiometric matrix). The other $n - p$ metabolites occur only inside the cell and are not exchanged with the environment. Hence, the last $n - p$ exchange rates (r_{p+1} to r_n) are equal to zero.

Equation 2.2 can be split into two parts. One part that involves the metabolites that are exchangeable with the environment and another part that involves the metabolites of which

there is no net production or consumption (because of the pseudo steady state assumption):

$$\begin{cases} \mathsf{S}_{\mathsf{acc}}\ \boldsymbol{v} = \boldsymbol{r_{acc}} \\ \mathsf{S}_{\mathsf{pss}}\ \boldsymbol{v} = \boldsymbol{r_{pss}} = \mathbf{0} \end{cases} \tag{2.3}$$

$\mathsf{S}_{\mathsf{acc}}$ contains the rows of S that correspond to the exchangeable metabolites (*i.e.* the first p rows of the stoichiometric matrix in equation 2.2). The corresponding exchange rates are summarised in $\boldsymbol{r_{acc}}$. Thus $\boldsymbol{r_{acc}}$ contains the p first r's of equation 2.2, *i.e.* all exchange rates that are not zero. $\mathsf{S}_{\mathsf{pss}}$ is constituted of those rows that agree with the non accumulating metabolites (*i.e.* those that are in pseudo steady state). Hence, the corresponding rates, summarised in $\boldsymbol{r_{pss}}$ are equal to zero.

2.1.2 The extended stoichiometric matrix W

A second way (compared to equation 2.1) of representing a biochemical reaction network is:

$$\frac{\mathrm{d}\boldsymbol{c}}{\mathrm{d}t} = \mathsf{W}\ \boldsymbol{a}(\boldsymbol{c}) \tag{2.4}$$

where $\boldsymbol{c}$ is a vector containing the concentration of the different metabolites inside the cell, matrix W is the extended stoichiometric matrix and vector $\boldsymbol{a}$ is the vector with all the reaction rates. If steady state is assumed, the concentration of the metabolites inside the cell do not change, the reaction rates are independent of the intracellular metabolite concentrations and equation 2.4 can be simplified to

$$\mathsf{W}\ \boldsymbol{a} = \mathbf{0} \tag{2.5}$$

W is called the extended stoichiometric matrix because it contains more than pure enzymatic conversions alone. The exchange of metabolites with the environment is added as reactions taking from, or removing matter to a buffer (Schilling *et al.*, 1999). Thus input and output from the system is modelled as (*e.g.* glucose):

$$\longrightarrow \mathrm{GLC} \tag{2.6}$$

The corresponding rate in the flux vector $\boldsymbol{a}$ is the input rate for glucose.

2.1.3 Relationship between S and W

Equation 2.2 can be rewritten as (using the fact that: $r_{p+1}, \ldots, r_n$ are equal to zero):

$$\begin{bmatrix} S_{1,1} & \ldots & S_{1,m} & -1 & 0 & \ldots & 0 \\ S_{2,1} & \ldots & S_{2,m} & 0 & -1 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{p,1} & \ldots & S_{p,m} & 0 & 0 & \ldots & -1 \\ S_{p+1,1} & \ldots & S_{p+1,m} & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{n,1} & \ldots & S_{n,m} & 0 & 0 & \ldots & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_m \\ r_1 \\ \vdots \\ r_p \end{bmatrix} = \mathbf{0} \tag{2.7}$$

This equation can be summarised to

$$\begin{bmatrix} \mathsf{S}_{\mathsf{acc}} & -\mathsf{I} \\ \mathsf{S}_{\mathsf{pss}} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{r} \end{bmatrix} = \mathbf{0} \tag{2.8}$$

which matches equation 2.5.

2.2 The pseudo inverse of a matrix

The solution of a linear system:

$$\mathsf{A}\,\boldsymbol{x} = \boldsymbol{b} \tag{2.9}$$

can be found by calculating the inverse of matrix A:

$$\boldsymbol{x} = \mathsf{A}^{-1}\,\boldsymbol{b} \tag{2.10}$$

This can only be done if A is invertible, which means that the system of equations should be neither overdetermined nor underdetermined. In MFA those conditions are rarely fulfilled.

In the case of an overdetermined system a 'least square' solution can be calculated by use of the (left) pseudo inverse:

$$\boldsymbol{x} = \left(\mathsf{A}^{\mathsf{T}}\,\mathsf{A}\right)^{-1}\,\mathsf{A}^{\mathsf{T}}\,\boldsymbol{b} \tag{2.11}$$

The left pseudo inverse can thus be defined as $\left(\mathsf{A}^{\mathsf{T}}\,\mathsf{A}\right)^{-1}\,\mathsf{A}^{\mathsf{T}}$ and to be calculable, $\mathsf{A}^{\mathsf{T}}\,\mathsf{A}$ must be nonsingular. This means that the number of rows of A should be less or equal to the number of columns. Furthermore, those rows should be linearly independent.

A more general method to calculate the pseudo inverse, applicable on every matrix, is based on Singular Value Decomposition (Golub & Van Loan, 1996) (a good explanation of SVD, the pseudo inverse and its application in MFA can be found in van der Heijden, 1991). Given the matrix A with dimensions $\langle m \times n \rangle$, there exist orthonormal matrices:

$$\mathsf{U}_{\langle m \times m \rangle} \text{ and } \mathsf{V}_{\langle n \times n \rangle} \tag{2.12}$$

so that

$$\mathsf{A} = \mathsf{U}\,\mathsf{S}\,\mathsf{V}^{\mathsf{T}} \tag{2.13}$$

where S is an $\langle m \times n \rangle$ dimensional non square diagonal matrix. The elements on the diagonal are the singular values of A. The number of singular values that are not zero is equal to the rank of A. When A has full rank and $m = n$, the inverse of A is (the inverse of an orthonormal matrix is equal to its transposed)

$$\mathsf{A}^{-1} = \mathsf{V}\,\mathsf{S}^{-1}\,\mathsf{U}^{\mathsf{T}} \tag{2.14}$$

The inverse of a diagonal matrix is calculated by inverting the elements on the diagonal one by one. If A is not fully ranked, or A is not square, a partial inverse can be calculated by

inverting only singular values in S that are not zero. This way the pseudo inverse of every matrix can be calculated:

$$\mathsf{A}^{\#} = \mathsf{V}\ \mathsf{S}^{\#}\ \mathsf{U}^{\mathsf{T}} \tag{2.15}$$

This allows to calculate a solution not only when the system is fully- or overdetermined, but also when parts of the system are underdetermined. The next section explains how one can make a distinction, by use of the nullspace, between those underdetermined parts (whose obtained solution is obviously not unique) and the determined ones.

2.3 Nullspace of a matrix

If the system as given in equation 2.9 is underdetermined, the solution given with the pseudo inverse:

$$\boldsymbol{x} = \mathsf{A}^{\#}\ \boldsymbol{b} \tag{2.16}$$

is only one of the infinite solutions. To determine which elements of $\boldsymbol{x}$ have a single solution and to determine a relationship between the infinity of solutions for the other elements of $\boldsymbol{x}$, the nullspace can be calculated.

The nullspace is defined as the set of linear independent vectors, $\boldsymbol{x_n}$, that fulfil the equation:

$$\mathsf{A}\ \boldsymbol{x_n} = \boldsymbol{0} \tag{2.17}$$

The number of independent nullspace vectors is equal to the number of columns in A minus the rank of A. From the definition it is clear that each nullspace vector can be added an arbitrary number of times to the base solution given in equation 2.16. Thus the complete solution is:

$$\boldsymbol{x} = \mathsf{A}^{\#}\ \boldsymbol{b} + \text{nullspace}\,(A)\ \ \boldsymbol{f} \tag{2.18}$$

where $\boldsymbol{f}$ is a vector with as many elements as there are vectors (columns) in the nullspace of A. For all possible values of $\boldsymbol{f}$ the solution remains valid.

Parallel pathways in the metabolic model typically yield nullspace vectors if there are no fluxes measured from one of those pathways. Elements (rows) of $\boldsymbol{x}$ that have only zeros in the nullspace of A are fully determined. Thus, a system of equations can have some unknowns for which a unique solution can be found and some for which this is not possible.

2.4 Model check techniques

According to Edwards *et al.* (2002), biological discovery will not be limited by the availability of data, but by the lack of available tools to analyse and interpret these data. One of these tools is metabolic modelling. Information about reactions is put into a network model. Several

mathematical tools can then be used to extract useful information from the model about the micro-organism under study (Noorman *et al.*, 1991; Schilling *et al.*, 1999, 2000; Klamt *et al.*, 2002).

In this section three such techniques will be discussed. First, the simple but very useful elemental consistency test will be explained. Secondly the detection of metabolic dead-ends will be discussed, and finally the concept of null cycles will be explained. The first two techniques are useful when collecting the set of relevant biochemical reactions and constructing the metabolic model. They check whether the set of reactions makes sense. The last technique, detection of parallel pathways by evaluating the presence of null cycles, allows to find parts of the metabolic model that will be underdetermined if no intracellular fluxes are provided as input data when solving the model. This knowledge can be used during the model building phase, where it can be decided to not include some reactions (*e.g.* the salvage pathways for nucleotides).

2.4.1 Elemental consistency check

Because the stoichiometric matrix S is the foundation of all techniques used in metabolic modelling, it is important to make sure that this matrix is correct *i.e.* that each reaction obeys the chemical law of conservation. Therefore a second matrix is constructed, the elemental composition matrix E.

The atomic composition of each metabolite that occurs in the stoichiometric matrix is put in a column of the elemental composition matrix. The order in which the metabolites occur in the elemental composition matrix should be the same as in the stoichiometric matrix. So if a metabolite occurs in row i of the stoichiometric matrix, it should also appear in column i of the elemental composition matrix.

The different rows of the elemental composition matrix represent the different atoms that are used to build up the metabolites. When electrical charge is also accounted for, an extra row is added to the elemental composition matrix.

This gives an elemental composition matrix with dimensions $\langle q \times n \rangle$, with q the number of atoms considered to be relevant for the used metabolites and n the number of metabolites considered in the model, as seen in equation 2.2.

If the stoichiometric matrix is correct, the following equation should hold:

$$\mathsf{E}\,\mathsf{S} = 0 \tag{2.19}$$

This equation should give the zero matrix with dimension $\langle q \times m \rangle$. If, for example, element (i, j) of that matrix is not zero, it can be concluded that the balance of atom i in reaction

j is not correct. Before doing further work with the metabolic model, elemental consistency errors should be corrected.

2.4.2 Dead-ends

Not every set of individual reactions forms an acceptable and meaningful metabolic reaction system. Sometimes a set of reactions cannot function because one reaction is absent. For example, if one forgets to include the first reaction of the glycolysis, the other reactions have a steady state flux of zero. Those sets of reactions that surely have a flux equal to zero, are called dead-ends. Dead-ends can also be caused by accumulation instead of depletion, for example when a transport reaction for an intracellular metabolite to the environment is omitted.

Furthermore, it is desirable to know whether the metabolites that are assumed to be exchangeable, are really exchangeable. If not, they will be detected with the same methodology as the one with which dead-ends in reaction sets are detected.

To find dead-ends, the test proposed by van der Heijden & Heijnen (1995) can be used. Initially the test was developed to detect dead-ends in reaction fluxes, but it can easily be extended for the detection of metabolites that are initially set as exchangeable, but are not. This test makes use of the extended stoichiometric matrix W.

The pseudo steady state assumption implies that the formation rate of each non-accumulating metabolite (corresponding to the rows $p+1$ to n in matrix W) must equal its consumption rate. If a metabolite is produced by a certain reaction and not consumed (by some other reactions), then that reaction will have a flux equal to zero. On the other hand, a reaction that consumes a metabolite that is not produced in any other reaction, will also have a zero flux.

For the exchangeable metabolites (rows 1 to p from matrix W) the same is true: there must be a reaction that produces or consumes a net amount of the exchangeable metabolite. Otherwise the exchange rate is zero. To detect the zero fluxes or exchange rates, equation 2.5 is solved (Golub & Van Loan, 1996):

$$\boldsymbol{a} = \mathsf{W}^{\#}\,\mathbf{0} + \text{nullspace}\,(\mathsf{W})\;\;\boldsymbol{F} = \text{nullspace}\,(\mathsf{W})\;\;\boldsymbol{F} \tag{2.20}$$

$\mathsf{W}^{\#}$ is the pseudo inverse of the matrix W and $\boldsymbol{F}$ is a column vector with as many elements as there are dimensions in the nullspace of W. Vector $\boldsymbol{F}$ represents the freedom of the system *i.e.* how much fluxes (v_i's) or exchange rates (r_i's) may be chosen before there is an unique solution. If a reaction or exchange rate is part of a dead-end, the flux is zero, whatever the elements of $\boldsymbol{F}$ are. This means that the fluxes v_i or exchange rates r_i corresponding to rows of the nullspace of W containing only zero's, are dead-ends. Dead-ends do not contribute to the

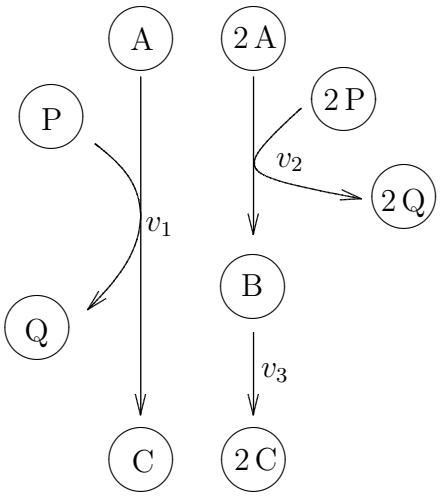


Figure 2.1: Metabolic network with one nullcycle.

final solution and can be removed from the model. However it should be investigated whether they are not caused by the omission of a reaction. In the latter case the model should be corrected.

2.4.3 Nullcycles and parallel pathways

The number of parallel pathways in a metabolic network is a measure for its robustness. The net production of two parallel routes should be strictly the same. Thus, futile cycles are not parallel pathways as they have ATP consumption as a net effect. Under thermodynamic considerations, each pair of parallel pathways should go in the same direction. To mathematically detect them, they are considered to run in opposite direction such that the net production is zero: the second parallel pathway of the pair consumes what the first pathway produces and vice versa. This gives for equation 2.2:

$$\mathsf{S}\,\boldsymbol{v} = \mathbf{0} \tag{2.21}$$

It represents a homogeneous system of equations. To find the non-trivial solutions, the nullspace of S has to be taken. By definition (Golub & Van Loan, 1996), each vector $\boldsymbol{k}$ (column) of the nullspace of a matrix M obeys the equation: $\mathsf{M}\,\boldsymbol{k} = \mathbf{0}$.

The dimension (number of columns) of the nullspace of S is equal to the number of pairs of parallel pathways. Each pair of parallel pathways defines a set of reactions that, if run in the stoichiometric quantities dictated by the coefficients of the nullspace vector, produces nothing. Therefore, those pairs of parallel pathways are called nullcycles.

In figure 2.1 an example of a metabolic network with one nullcycle is given: $\{2v_1, -v_2, -v_3\}$. Running v_2 and v_3 in opposite direction cancels the effect of two times reaction v_1.

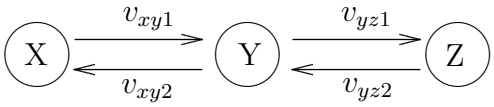


Figure 2.2: Network with two independent nullcycles but it seems as if there are three.

Several couples of parallel pathways can be extracted from a nullcycle, depending on where the cycle is broken. In the nullcycle of figure 2.1 ($\{2v_1, -v_2, -v_3\}$), a first couple of parallel pathways can be $\{-v_2, 2v_1\}$ and $\{v_3\}$ (breaking the nullcycle at metabolite B and C). Another couple of parallel pathways can be constructed if the nullcycle is broken at metabolite A and C: $\{2v_1\}$ and $\{v_2, v_3\}$.

It is desirable to choose parallel pathways that give insight into the metabolic model. One property of the chosen parallel pathways should be that they are thermodynamically feasible. But even then, there can be more than one possible couple of parallel pathways in a nullcycle if some reactions are considered reversible. Therefore, the extra condition that all reactions are irreversible, can be used to unequivocally extract the parallel pathways from a nullcycle. Reactions with positive coefficients in the nullcycle make up the first parallel pathway and reactions with negative coefficients form the second parallel pathway. In the example of figure 2.1, the first parallel pathway is $\{2v_1\}$ and the second is $\{v_2, v_3\}$.

To know what the net reaction of the parallel pathways of a nullcycle is, the reactions of one of the parallel pathways are multiplied with the absolute value of their coefficients of the nullspace and the resulting net reaction is calculated. With the above example this can be summarised in the following matrix:

$$\begin{bmatrix} & v_2 & v_3 & Net\ reaction \\ \text{A} & -2 & 0 & -2 \\ \text{B} & 1 & -1 & 0 \\ \text{C} & 0 & 2 & 2 \\ \text{P} & -2 & 0 & -2 \\ \text{Q} & 2 & 0 & 2 \end{bmatrix} \tag{2.22}$$

It can be seen that the net reaction is the same as the reaction of the first parallel pathway $\{2v_1\}$.

Finding nullcycles relies on finding a base for a nullspace. However, different bases can be constructed for the same nullspace. This corresponds to the fact that different nullcycles can be detected but not all are independent of each other. In figure 2.2 three nullcycles can be seen: $\{v_{xy1}, v_{xy2}\}$, $\{v_{yz1}, v_{yz2}\}$ and $\{v_{xy1}, v_{yz1}, v_{xy2}, v_{yz2}\}$. However, the last nullcycle is the sum of the two first or the second is equal to the last minus the first. In fact, there really

are only two nullcycles. All the others can be constructed from the two nullcycles that are selected.

The question remains which nullcycle to choose. In the example of figure 2.2 it seems logical to take the first two, $\{v_{xy1}, v_{xy2}\}$ and $\{v_{yz1}, v_{yz2}\}$, as representation of the possible nullcycles. But it may be that it is easier to recognise the process of nullcycle three ($\{v_{xy1}, v_{yz1}, v_{xy2}, v_{yz2}\}$) than that of nullcycle two. Furthermore, a nullcycle with few reactions can generate a complicated net reaction of the corresponding pair of parallel pathways. Thus, even if all possible nullcycles (all possible vectors of all possible bases) could be determined, finding criteria to select the simplest nullcycles is not a trivial task.

Although the nullcycles cannot be determined unequivocally, the technique is useful to measure how much redundancy exists in the stoichiometric model. Redundancy is defined by Edwards & Palsson (1998) as the capability of the cell to redistribute its metabolic fluxes when faced with the loss of one or multiple enzymes; more nullcycles means more alternative pathways when one pathway is blocked. The selected nullcycles can be used to determine which internal reaction rates (the v_i's of equation 2.2) should be measured to solve equation 2.2, because net exchange rates (the p first r_i's of equation 2.2) do not give information about the partition workload between the different parallel pathways. To solve equation 2.2, at least one reaction rate of each nullcycle should be measured.

As can been seen from figure 2.2, if both a reaction and the reverse reaction are included in the stoichiometric matrix, it will give an extra nullcycle.

2.5 Solving an overdetermined metabolic model

A metabolic network model is basically a system of linear equations. There are two methods to solve such a system. The first one works directly with nullspace calculations (van der Heijden & Heijnen, 1995). It is an elegant method but has a major drawback: it is not possible to use internal flux measurements to reduce the number of degrees of freedom in the system. Only exchange rates can be used. The second method (well described in Noorman *et al.* (1996)) is more general and is the one that is used in this work.

2.5.1 Solution of the metabolic model

Equation 2.5 is split up in two parts: one for the rates that are measured and thus known (subscript m), and one for the rates to be calculated (subscript c):

$$\mathsf{W_c}\,\boldsymbol{a_c} + \mathsf{W_m}\,\boldsymbol{a_m} = \mathbf{0} \tag{2.23}$$

This equation can be solved to $\boldsymbol{a_c}$:

$$\boldsymbol{a_c} = -\mathsf{W}_\mathsf{c}^{\#}\ \mathsf{W}_\mathsf{m}\ \boldsymbol{a_m} + \text{nullspace}\,(\mathsf{W}_\mathsf{c})\ \ \boldsymbol{f} \tag{2.24}$$

2.5.2 Redundant measurements

The solution obtained in equation 2.24 can be substituted into equation 2.23:

$$\mathsf{W}_\mathsf{m}\ \boldsymbol{a_m} - \mathsf{W}_\mathsf{c}\ \left(\mathsf{W}_\mathsf{c}^{\#}\ \mathsf{W}_\mathsf{m}\ \boldsymbol{a_m} + \text{nullspace}\,(\mathsf{W}_\mathsf{c})\ \ \boldsymbol{f}\right) = \mathbf{0} \tag{2.25}$$

As a matrix multiplied with its nullspace is equal to zero, this equation can be rewritten as:

$$\left(\mathsf{W}_\mathsf{m} - \mathsf{W}_\mathsf{c}\ \mathsf{W}_\mathsf{c}^{\#}\ \mathsf{W}_\mathsf{m}\right)\ \boldsymbol{a_m} = \mathbf{0} \tag{2.26}$$

When the system is overdetermined, *i.e.* when there are more measurements than degrees of freedom (and assuming the measurements do not agree perfectly), the terms between brackets in equation 2.26 form a non-zero matrix. All independent rows of that matrix can be combined in the redundancy matrix R (van der Heijden *et al.*, 1994a,b):

$$\mathsf{R}\ \boldsymbol{a_m} = \mathbf{0} \tag{2.27}$$

It should be noted that it is not because there is redundancy in the measurements that the nullspace of the extended stoichiometric matrix will be void. It is perfectly possible (and usual) to have parts of the system of equations that are overdetermined and parts that are underdetermined.

Redundant measurements can be used to enhance the confidence in the measurements. This is explained in Wang & Stephanopoulos (1983) for black box models, but the method also applies to stoichiometric models (Romein, 2001). The real value of a measurement $\boldsymbol{a_m}$ is equal to the measured value $\widetilde{\boldsymbol{a_m}}$ minus some random noise $\boldsymbol{\delta}$:

$$\boldsymbol{a_m} = \widetilde{\boldsymbol{a_m}} - \boldsymbol{\delta} \tag{2.28}$$

Equation 2.27 gives the general formula. Applied in a statistical context, distinction should be made between a real value of a variable (that can never be known), the measurement of that variable (indicated with a tilde), and the estimate of that variable (indicated with a hat). Combining equation 2.28 and 2.27 yields:

$$\mathsf{R}\ \widetilde{\boldsymbol{a_m}} = \mathsf{R}\ \ (\boldsymbol{a_m} + \boldsymbol{\delta}) = \mathsf{R}\ \boldsymbol{\delta} = \boldsymbol{\epsilon} \tag{2.29}$$

$\boldsymbol{\epsilon}$ being the vector of residuals. It was proven that to minimise the error $\boldsymbol{\delta}$ on the measurements, the following objective function J has to be minimised (Madron *et al.*, 1977):

$$J = \boldsymbol{\delta^T}\ \mathsf{P}_\epsilon^{-1}\ \boldsymbol{\delta} \tag{2.30}$$

P_ϵ is the variance covariance matrix of the vector of residuals and equals:

$$\mathsf{P}_\epsilon = \mathsf{R}\,\mathsf{P}_\delta\,\mathsf{R}^\mathsf{T} \tag{2.31}$$

The variance covariance matrix of $\boldsymbol{\delta}$ (P_δ) is equal to the variance covariance matrix of the measurements $\mathsf{P}_{\widetilde{\mathsf{a_m}}}$.

The solution of the minimisation problem of equation 2.30 gives an estimate for $\boldsymbol{\delta}$:

$$\widehat{\boldsymbol{\delta}} = \mathsf{P}_\delta\,\mathsf{R}^\mathsf{T}\,\mathsf{P}_\epsilon^{-1}\,\boldsymbol{\epsilon} \tag{2.32}$$

which in turn gives an estimate for the measurements (I is the identity matrix):

$$\widehat{\boldsymbol{a_m}} = \left(\mathsf{I} - \mathsf{P}_{\widetilde{\mathsf{a_m}}}\,\mathsf{R}^\mathsf{T}\,\left(\mathsf{R}\mathsf{P}_{\widetilde{\mathsf{a_m}}}\,\mathsf{R}^\mathsf{T}\right)^{-1}\,\mathsf{R}\right)\,\widetilde{\boldsymbol{a_m}} \tag{2.33}$$

That these estimated values are better than the measured ones, is obvious from the variance covariance matrix of the estimated values, since the second term of $\mathsf{P}_{\widehat{\mathsf{a_m}}}$ is always positive:

$$\mathsf{P}_{\widehat{\mathsf{a_m}}} = \left(\mathsf{I} - \mathsf{P}_{\widetilde{\mathsf{a_m}}}\,\mathsf{R}^\mathsf{T}\,\left(\mathsf{R}\mathsf{P}_{\widetilde{\mathsf{a_m}}}\,\mathsf{R}^\mathsf{T}\right)^{-1}\,\mathsf{R}\right)\,\mathsf{P}_{\widetilde{\mathsf{a_m}}} \tag{2.34}$$

Making an estimate of the measured fluxes thus reduces their uncertainty.

The estimated measurements should be used to calculate the unknown fluxes. Equation 2.24 becomes:

$$\widehat{\boldsymbol{a_c}} = -\mathsf{W}_\mathsf{c}^\# \,\mathsf{W}_\mathsf{m}\,\widehat{\boldsymbol{a_m}} + \text{nullspace}\,(\mathsf{W}_\mathsf{c})\;\boldsymbol{f} \tag{2.35}$$

and the variance covariance matrix of the unknown fluxes equals (van der Heijden, 1991):

$$\mathsf{P}_{\widehat{\mathsf{a_c}}} = \mathsf{W}_\mathsf{c}^\# \,\mathsf{W}_\mathsf{m}\,\mathsf{P}_{\widehat{\mathsf{a_m}}}\,\mathsf{W}_\mathsf{m}^\mathsf{T}\,\mathsf{W}_\mathsf{c}^{\#T} \tag{2.36}$$

Only the rows and columns of $\mathsf{P}_{\widehat{\mathsf{a_c}}}$ corresponding to elements of $\widehat{\boldsymbol{a_c}}$ that have no freedom left in the nullspace of W_c (*i.e.* for which the corresponding rows in the nullspace of W_c contain only zeros), are relevant.

2.5.3 Statistical test of the quality of the measurements

In the previous section it was explained how to increase the reliability of the measured fluxes. It is also possible to check whether the measurements are consistent. If so, the residual vector ϵ should be equal to zero. To investigate whether this is the case, the following test statistic h is used (Reilly & Carpani, 1963):

$$h = \boldsymbol{\epsilon}^{\boldsymbol{T}}\,\mathsf{P}_\epsilon^{-1}\,\boldsymbol{\epsilon} \tag{2.37}$$

The H_0 hypothesis of this statistic is that measurements are consistent and that h is equal to zero.

This test statistic is equivalent to the objective function J of equation 2.30 (Wang & Stephanopoulos, 1983). Thus, if P_{δ}^{-1} is a diagonal matrix, h can be rewritten as:

$$h = J = \widehat{\boldsymbol{\delta}}^{\boldsymbol{T}} \, \mathsf{P}_{\delta}^{-1} \, \widehat{\boldsymbol{\delta}} = \sum_{j=1}^{n} \frac{\widehat{\delta}_j^2}{\sigma_{\delta_{j,j}}^2} \tag{2.38}$$

with n being the number of elements in the vector $\boldsymbol{a_m}$, *i.e.* the number of measured fluxes. As the error δ on the measurement is assumed to be normally distributed, the terms of the sum in the equation above are normally distributed with variance equal to one. Thus the test statistic h follows, by definition, a χ^2 distribution.

It was proven that the number of degrees of freedom of that χ^2 distribution is equal to n minus the rank of R (van der Heijden *et al.*, 1994b).

2.5.4 Vector comparison test

If the quality of the measurements test rejects the H_0 hypothesis (stating that the errors are equal to zero), one can detect which measurement(s) is erroneous. A simple approach would be to try to remove measurements one by one, and check if the statistical test passes (the serial elimination method of Wang & Stephanopoulos (1983)). But statistically this is not good practice, as there are some issues with independence of the tests (van der Heijden *et al.*, 1994b). A better approach is to use the vector comparison test (van der Heijden *et al.*, 1994b).

Equation 2.27 can be written more explicitly as:

$$\boldsymbol{R_1} \, a_{m_1} + \boldsymbol{R_2} \, a_{m_2} + \cdots + \boldsymbol{R_i} \, a_{m_i} + \cdots + \boldsymbol{R_n} \, a_{m_n} = \boldsymbol{\epsilon} \tag{2.39}$$

with $\boldsymbol{R_i}$ representing the different columns of R and a_{m_i} the different measured fluxes. If one of the measurements is really wrong by an amount τ, the expected value (E) of ϵ is:

$$\begin{aligned} \mathrm{E}[\boldsymbol{\epsilon}] &= \mathrm{E}[\boldsymbol{R_1} \, a_{m_1} + \boldsymbol{R_2} \, a_{m_2} + \ldots \; + \boldsymbol{R_i} \; (a_{m_i} + \tau) + \cdots + \boldsymbol{R_n} \, a_{m_n}] \\ &= \mathrm{E}[\boldsymbol{R_1} \, a_{m_1} + \boldsymbol{R_2} \, a_{m_2} + \cdots + \; \boldsymbol{R_i} \, a_{m_i} + \cdots + \boldsymbol{R_n} \, a_{m_n}] + \boldsymbol{R_i} \, \tau \\ &= \boldsymbol{R_i} \, \tau \end{aligned}$$

The same reasoning can be applied when more measured fluxes are wrong. Then the residual vector $\boldsymbol{\epsilon}$ will be a linear combination of those erroneous measurements. For each combination of subvectors $\mathsf{R_s}$ of the redundancy matrix R, a statistical test is run with H_0 hypothesis that the residual vector ϵ is a linear combination of $\mathsf{R_s}$. The statistic used for this test:

$$h_{\Delta} = \boldsymbol{\epsilon}^{\boldsymbol{T}} \, \mathsf{P}_{\epsilon}^{-1} \, \boldsymbol{\epsilon} - \boldsymbol{\epsilon}^{\boldsymbol{T}} \, \mathsf{P}_{\epsilon}^{-1} \, \mathsf{R_s} \, \left(\mathsf{R_s^T} \, \mathsf{P}_{\epsilon}^{-1} \, \mathsf{R_s} \right)^{-1} \, \mathsf{R_s^T} \, \mathsf{P}_{\epsilon}^{-1} \, \boldsymbol{\epsilon} \tag{2.40}$$

follows a χ^2 distribution with $\mathrm{rank}(\mathsf{R}) - \mathrm{rank}(\mathsf{R_s})$ degrees of freedom (van der Heijden *et al.*, 1994b). Running the test for each possible combination of vectors of R is time consuming.

But it makes no sense to take combinations with more than $\mathrm{rank}(\mathsf{R}) - 1$ vectors, as the degree of freedom of the corresponding statistic is then zero or less. This is also intuitive: a vector of n elements is a linear combination of every independent set of n vectors of length n. As such, the problem of too many combinations becomes worse the more fluxes are measured. A luxury problem.

2.6 Implementation

Three main parts can be discerned in the generation of metabolic flux maps (figure 2.3). First, the metabolic model has to be constructed. Then the data have to be collected and preprocessed in a form suitable for metabolic modelling. And finally the data and the model are combined.

2.6.1 Metabolic model generation and checks

Collecting the different reactions of the metabolic model together with the elemental composition matrix, is done in a spreadsheet program (OpenOffice.org). Each reaction and each metabolite is named. This name is carried on in all subsequent calculations allowing easy debugging and interpretation of the obtained results.

The list of reactions, exported as a text file, is converted by a Perl (`www.perl.org`) program (mmm.pl in figure 2.3) to a stoichiometric matrix S. Besides that, other vectors are defined containing the names of the exchangeable and measured metabolites. Also, the elemental matrix is generated.

Those matrices and vectors are read by a Maple program (Heck, 1993) by which an elemental composition check and dead-end test are performed. If both tests pass, another Maple routine searches for parallel pathways (modelanalysis in figure 2.3). Based on those tests, the model is adapted until errors are no longer encountered. The result is a mathematically correct metabolic model.

2.6.2 Preprocessing the measurement data

When performing a cultivation, different measurement data are collected and have to be combined in a format suitable for automatic post processing. The data obtained online via the MFCS/win computer system (Sartorius BBI, Germany) are semi manually exported and the relevant parts are selected and saved to the raw data file. The program written for this in SciLab (`www.scilab.org`, fda.sci in figure 2.3) allows to easily visualise selected data and extract time intervals for further processing. Furthermore a basic filter is available to remove outliers (*e.g.* due to malfunctioning equipment). The lab analysis data obtained after

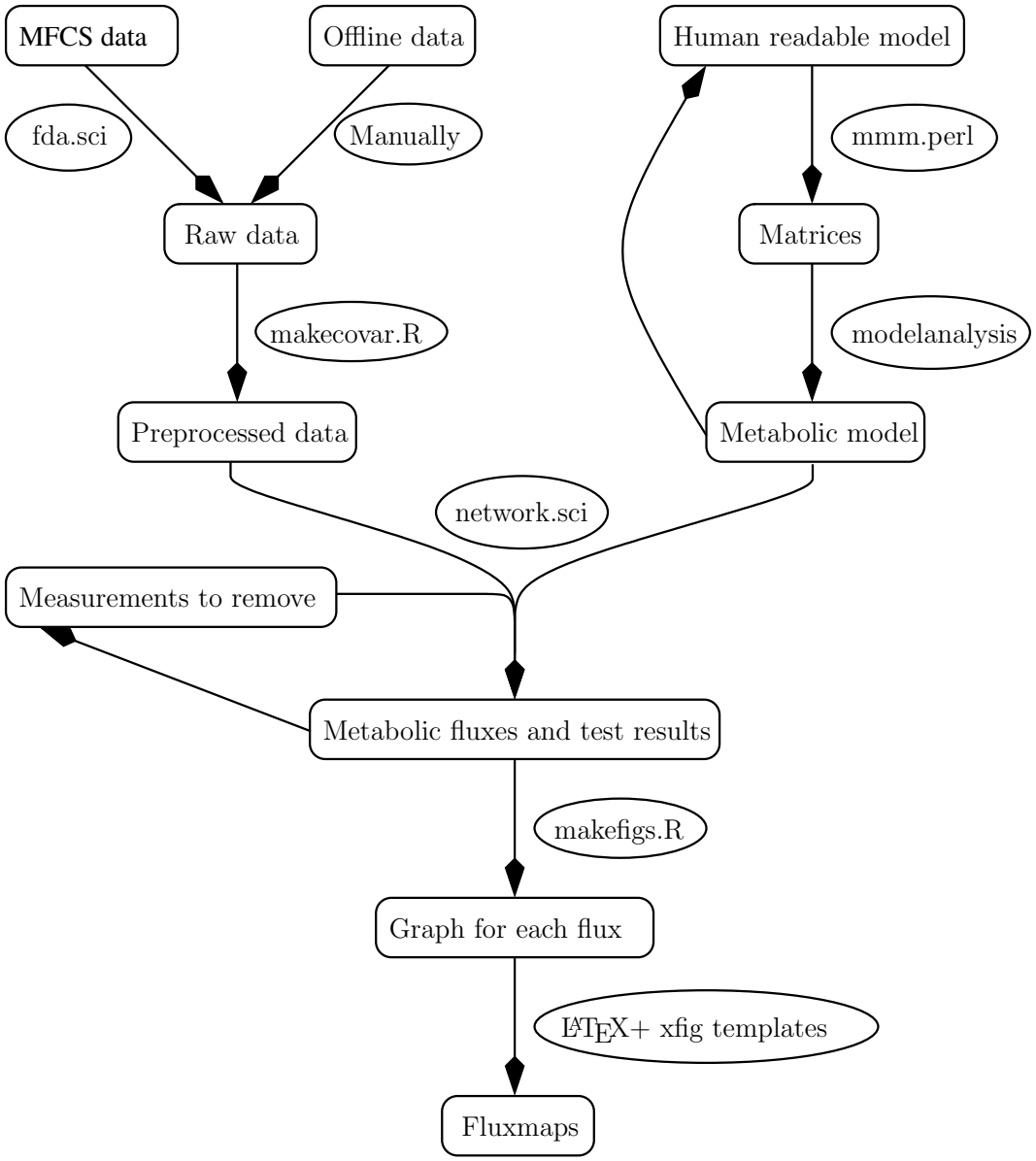


Figure 2.3: Combining measurement data with metabolic models through the different programs, finally gives flux maps.

the culture is finished (cell dry weight, HPLC, nitrogen and phosphorus analysis) must be manually added to the raw data file.

Converting the different measurement data to units suitable for input in the metabolic model (in this work mol/l/h was chosen) and generating the variance covariance matrix from the different repeats of the measurements, is done in R (makecovar.R in figure 2.3; Ihaka & Gentleman, 1996; R Development Core Team, 2006). The covariance matrix is calculated from the raw data and refined on theoretical bases. First, the residuals are calculated by subtract-

ing the raw measurement from the mean value. This is done for each measurement in each experiment. The residuals from all experiments are collected and on those residuals, the covariance matrix is calculated, using the Pearson method with pairwise complete observations, as explained in the help pages of R. Then, all covariances of measurements that are assumed to have no correlations, are set to zero. Actually, all measurements but the HPLC ones, were assumed to be independent of each other. This covariance matrix is subsequently used when transforming the raw measurement data expressed in grams/liter (products in the reactor broth) or fractions (gas exhaust) to moles/liter/hour. The covariance matrix is transformed accordingly using first order linearisation the details of which can be found in appendix A.

2.6.3 Combining the metabolic model with the data

Solving the metabolic model, balancing the measurements and performing the statistical tests is done in SciLab (network.sci in figure 2.3). The fluxes for the different experiments, together with their variances, are saved to a single file, to facilitate the generation of combined graphs for the different experiments.

For each metabolite and each flux, a graph is automatically generated (makegraphs.R in figure 2.3). The different graphs are combined to a flux map of which the template was created in Xfig (`www.xfig.org`). Xfig allows to embed LaTeXcommands in the figures. In this case, the commands for inclusion of the different graphs were given in those fluxmap skeletons and when running the figure through LaTeX(`www.ctan.org`), the actual graphs are shown.

The whole process was completely automated with make scripts (`www.gnu.org/software/make/`).

The software was applied in the metabolic modelling studies presented in chapter 3, to elucidate the difference in shikimate yield between carbon- and phosphate-limited cultures and in chapter 4, to compare the metabolic implications of different strategies for reducing acetate formation. The methodology for metabolic modelling was extended to be applicable under transient conditions also (chapter 5 and 6). For this, a program was written that extracts fluxes from the data obtained from a transient experiment. The details of this can be found in chapter 5.

Chapter 3

MFA compared with RNA expression data to elucidate the difference in shikimate yield between carbon- and phosphate-limited continuous cultures of *E. coli* W3110.shik1[1]

Shikimic acid is an interesting starting material for production of many chemical compounds, with as major example Tamiflu®, used for treatment of influenza (De Clercq, 2002). As the extraction of shikimate from the plant *Illicium* is expensive, production strategies via fermentation are developed. One of the strains that has been genetically modified to produce more shikimate is *E. coli* W3110.shik1 (Johansson *et al.*, 2005). Shikimate production however is very low under carbon-limiting conditions.

In this chapter, the metabolic flux analysis for overdetermined systems, as explained in chapter 2, is applied to a set of experiments aiming at finding differences in flux distribution between carbon-limited and carbon-abundant (phosphate-limited) cultures of the *E. coli* strain W3110.shik1 (Johansson *et al.*, 2005). This strain was genetically modified to produce more shikimate (an intermediate of the aromatic amino acid pathway). Under carbon-limited conditions, more byproducts, such as dehydroshikimate and dehydroquinate, and less shikimate

[1]Part of this chapter was published in G. Lequeux, L. Johansson, J. Maertens, P. A. Vanrolleghem & G. Lidén (2006). 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

can be found (Johansson *et al.*, 2005; Knop *et al.*, 2001) than under carbon-abundance. It is shown that this lower shikimate yield under carbon-limited conditions is not due to a lower flux going to the aromatic pathway, on the contrary, the flux is higher.

The second part compares fluxes obtained from flux balancing with RNA expression level measurements. Although it is sometimes tried to solve the problem of parallel pathways with the aid of RNA expression levels (Huang *et al.*, 2003), it is not done here because it is expected to be impossible (Åkesson *et al.*, 2004). A high expression level not necessarily correlates with a high flux. There is still translational and metabolic regulation that has influence on the metabolism. However it is interesting to see how for some fluxes the variations over different culture conditions correlates with the RNA expression levels of the corresponding genes. Some fluxes are negatively correlated with expression levels. But most of the fluxes are not correlated with gene expression levels, suggesting that transcriptional regulation only happens at metabolic nodes.

3.1 Materials and methods

3.1.1 Experimental setup

Strains and cultivations

W3110 was obtained from the American Type Culture Collection (ATCC 27325) and was used as host for the shikimic acid producing strain (referred to as W3110.shik1 in the text). This strain had the following genetic modifications (Johansson *et al.*, 2005):Δ*aroL*, tryptophan and phenylalanine feedback resistant *aroG*, tryptophan feedback resistant *trpE* and *tnaA*. Furthermore, W3110.shik1 was cloned with plasmid pSGs26 (derived from pBR322) containing tyrosine and phenylalanine feedback resistant *aroF* and two antibiotic resistance markers amp^R (ampicillin) and $tetc^R$ (tetracycline).

In the first set of experiments W3110.shik1 six phosphate-limited and four carbon-limited continuous cultures with W3110.shik1 were run at dilution rates ranging from $0.05\,h^{-1}$ to $0.3\,h^{-1}$.

A second set of experiments consisted of phosphate and carbon-limited chemostats of the wild-type W3110 and the shikimate producing strain W3110.shik1 at a dilution rate of $0.2\,h^{-1}$. Each combination was performed in duplicate. Besides input and output fluxes, RNA expression levels were also determined.

Inoculum preparation

The medium used for inoculum preparation, 100 ml, was the same for carbon-limited and phosphate-limited experiments. For W3110 (wild-type), the medium contained per liter: 20 g

glucose (VWR International, Stockholm, Sweden), 7.5 g K_2HPO_4, 0.3 g NH_4Fe(III)-citrate, 2.1 g citric acid monohydrate, 0.1 ml $MgSO_4 \cdot 7\,H_2O$ (1 M), 0.0037 g $(NH_4)_6(Mo_7O_{24}) \cdot 4\,H_2O$, 0.0029 g $ZnSO_4 \cdot 7\,H_2O$, 0.0247 g H_3BO_3, 0.0025 g $CuSO_4 \cdot 5\,H_2O$, 0.0158 g $MnCl_2 \cdot 4\,H_2O$ and 0.055 mg thiamine (Sigma-Aldrich, Steinheim, Germany), 1.2 ml H_2SO_4 (96 %) (VWR International). For W3110.shik1, the medium additionally contained: 0.010 g p-hydroxybenzoic acid, 0.010 gram potassium p-amino benzoate, 0.010 g 2,3- dihydroxybenzoic acid (vitamins), and 0.015 g tetracycline hydrochloride (antibiotic) (Sigma-Aldrich). Inoculum cultures were grown in baffled E-flasks of 250 ml at 37 °C at a shaker speed of 200 rpm up to an OD-value of 1.5–2.5, to allow exponential growth. The reactor was inoculated with 40 ml of the inoculum W3110.shik1 or 20 ml of the W3110 culture.

Culture media

The glucose and mineral solutions were sterilised separately at 121 °C for 20 min and mixed afterwards. Antibiotic solution, vitamins, trace metals and $MgSO_4$-solution were added by sterile filtration through a 0.2 mm Minisart cellulose acetate filter (Sartorius AG, Goettingen, Germany). All solutions were prepared using deionised water. 1.5 l of medium was prepared for the batch cultivation. The working volume of the chemostat was around 1.3 l, and pH was adjusted to 7.0 by addition of NH_4OH (25 %) before and after sterilisation. Level control kept the volume of the cultivation broth constant.

Carbon-limited cultivations

The composition of the C-limitation medium for the batch phase and the chemostat phase of the W3110 was the same as for the preculture medium, except for some minor changes. The medium contained 0.13 g/l antifoam 286 (Sigma- Aldrich). In addition, the glucose concentration was 10 and 25 g/l in batch and chemostat phase, respectively. The composition of the medium for growth of W3110.shik1 was the same as that for W3110, except that it also contained 0.015 g/l tetracycline hydrochloride.

Phosphate-limited cultivations

The composition of the P-limited medium in batch phase was per liter: 20 gram glucose, 92 ml H_3PO_4 (85 %), 5.39 g NH_4SO_4, 3.32 g NaOH, 1.66 gram KOH, 0.52 g $MgSO_4 \cdot 7\,H_2O$, 0.133 g antifoam 286, 3.73 ml H_2SO_4 (96 %); trace metals (1): 0.093 g $FeSO_4 \cdot 7\,H_2O$, 0.079 g citric acid monohydrate; trace metals (2): 0.0073 g $CoCl_2 \cdot 6\,H_2O$, 0.00207 g $MnCl_2 \cdot 4\,H_2O$, 0.00103 g ZnCl (Sigma-Aldrich). The medium of W3110.shik1 also contained 0.015 g/l tetracycline hydrochloride. In chemostat operation, the phosphate solution and the solution of the remaining medium components were fed separately to the reactor. The phosphate feed contained 0.9 ml/l H_3PO_4 (85 %), whereas the second feed solution contained all other components. Since the phosphate solution contributed extra volume and thereby diluted the second

feed solution the latter was up concentrated 1.3 times, *e.g.* the glucose concentration was 32 g/l giving a concentration of 25 g/l in the reactor.

Culture methods

All cultures were carried out using a Biostat CT culture vessel (Sartorius BBI Systems GmbH, Melsungen, Germany) with a maximum working volume of 3.5 l. Temperature (37 °C), pH (7.0), stirring rate (750 rpm) and airflow rate (0.75 slpm) were controlled by the program MFCS/win shell 2.0 (Sartorius BBI) via the control unit DCU Biostat C (Sartorius BBI). pO_2 was measured with an InPro 6000 (Mettler Toledo Gmbh, Giessen, Germany). The pH was measured with a pH-meter of type 405-DPAS-SC-K8S/120 (Mettler Toledo). For maintaining the pH at 7.0, H_2SO_4 (2 M) and NH_4OH (25 %) were used in the C-limited case and H_2SO_4 (1 M) and NH_4OH (12.5 %) in the P-limited case. O_2 and CO_2 content in the off gas were measured by an Innova 1311 (INNOVA Air Tech Instruments, Ballerup, Denmark). Two balances and pumps were also coupled to the equipment allowing precise measurement of the feed rate during chemostat operation. Data from all equipment except for one of the pumps were routinely logged by the data acquisition software.

In the C-limited case, chemostat operation was started when glucose was depleted (indicated by a rapid reduction of the CO_2 in the off gas). In the P-limited case, the chemostat was started when the phosphate was consumed (which coincided with a small peak in the pO_2). Establishment of steady state was confirmed from measurement of the off-gas composition. At least five residence times were allowed to pass before each steady state. Sampling for OD and for metabolite analyses were carried out at fermentation start, end of batch, and at steady state during chemostat operation. Samples for dry weight were taken at the end of the batch cultivation and at steady state during chemostat cultivation.

3.1.2 Analytical methods

The analytical methods for determining the amount of metabolites and cell dry weights are described in Johansson *et al.* (2005). 3.5 % of the dry weight was considered as ash (Battley, 1995).

For the elemental composition of biomass, 50 ml of fermentation broth was taken and centrifuged for 20 minutes at 2000 g and 4 °C. The pellet was washed once with 40 ml ice cold 0.9 % NaCl and again centrifuged and decanted. The pellets were then frozen in liquid nitrogen, vacuum dried and analysed with a CHNS-O analyser (Model EA1108, Carlo Erba Instruments, Italy) (van Gulik *et al.*, 2000).

Sampling for transcriptome analysis was performed at steady state. To minimise mRNA degradation, 50 ml falcon tubes containing about 25 ml of ice were filled with reactor broth

as fast as possible. The tubes were centrifuged for 1 min at 2000 g and the pellet was frozen in liquid nitrogen. This method was done according to the recommendations from Brigitte Regenberg from DTU (Denmark Technical University). Those recommendations were followed for the C-limited experiments, as they were performed first. But later, when it was decided to do P-limited experiments, it was realised that DTU had only worked with yeast, of which the RNA is more stable that that from *E. coli*. Therefore, in the phosphate limited experiments, the mRNA sampling method was improved by adding 1.25 ml 'stop-solution', containing about 95 % ethanol and 5 % phenol, to the tubes prior to sampling. Even then a stress response can not be excluded. It was observed that some heat-chock genes were activated. They should however be activated in all cases.

Total RNA was extracted by using the Fastprep® system including the FastRNA®Pro Blue kit (Qbiogene, Montréal, Canada). Present DNA was degraded by addition of DNAse (VWR International) to a concentration of $0.2\,\mathrm{U}\mu^{-1}\mathrm{l}^{-1}$. The samples were then held at 37 °C for 20 min. The reaction was interrupted by addition of 0.5 M EDTA (Sigma-Aldrich) to a final concentration of 10 mM. The RNA samples were further cleaned by using a RNeasy microelute Cleanup kit (Qiagen, Venlo, The Netherlands). The quality of RNA was checked by running the samples on a 1 % TBE gel. The RNA samples were then sent to SWEGENE Microarray Resource Centre (Lund, Sweden) where the quality of RNA was further checked by using Nanodrop ND1000 (Nanodrop Technologies. Wilmington. USA), which gives very accurate concentration and 260/280 ratio figures. In addition quality testing of the RNA was carried out using the Agilent Bioanalyser 2100 (Agilent Technologies, Palo Alto, USA). cDNA synthesis, cDNA fragmentation and preparation of the hybridisation mixture was carried out according to the recommendations of the manufacturer of the microarrays (Affymetrix, Santa Clara, USA). Hybridisation, washing, staining and scanning of the microarrays (Affymetrix *E. coli* antisense genome arrays AS v2) were performed by using GeneChip Hybridisation Oven 640, GeneChip Fluidic Station 450 and GeneChip Scanner 2500 (all from Affymetrix). Data acquisition and gene expression data analysis was carried out using MAS 5.0 (Affymetrix). A chip to chip normalisation was performed in this program by scaling to a median intensity of 100. For statistic analysis the Bayesian test was carried out using a web-interface version of Cyber-T found at `http://cybert.microarray.ics.uci.edu/` (original url was: `http://visitor.ics.uci.edu/genex/cybert/` Long *et al.* (2001)).

The original data together with the analysis and discussion, can be found in Johansson & Lidén (2006). For this study, only selected genes were used. For those genes, the expression levels together with the associated variances were taken. Thus all genes shown here, are not necessarily significantly different from each other.

3.2 The metabolic model

3.2.1 Biomass composition

To investigate whether biomass composition varied for different dilution rates, the elemental composition of the biomass was determined (the elements C, H, O, N and S were measured; three repeats were performed). A significant difference was only found for oxygen and to a lesser extent nitrogen (figure 3.1). The oxygen content increases with increasing growth rate, which is the reason why the total molecular mass (expressed in C-moles) is rising slightly with increasing growth rate (figure 3.2).

Measurements of the protein, DNA and RNA content were performed to see whether the change in oxygen content was reflected in the biomass composition. However, no significant differences were found. The data were highly variable with protein content varying from 50 % to 80 % of the biomass and there was no correlation with the growth rate. Furthermore, the influence of the biomass composition on the calculated fluxes was very low. Thus a different biomass composition for each dilution rate would not be more correct than using the same biomass composition for each dilution rate even when neglecting the varying oxygen content of the actual biomass (figure 3.1). Therefore, a constant biomass composition was used for every dilution rate (and for every strain): 70 % protein, 12 % RNA, 3 % DNA and 15 % Other (expressed in g/gDW). Those values were mainly based on literature (Carlson & Srienc, 2004b). When some freedom was left in the literature data (confidence values), it was tried to strike a balance with the data shown in figure 3.1 and in a lesser extend, as there was more noise on those, the protein, DNA and RNA measurements.

The latter fraction (the 15 % Other of above) is subdivided into four other components: lipopolysaccharide (LPS), lipids, peptidoglycan and glycogen. According to Pramanik & Keasling (1997), the relative occurrence on weight basis of those four compounds for the average biomass composition of *E. coli* at a growth rate of $1\,h^{-1}$, is 0.194 LPS, 0.520 lipids, 0.143 peptidoglycane and 0.143 glycogen.

The composition of the lipid fraction can be found for different *E. coli* strains at different growth rates in Pramanik & Keasling (1997). The variation seems rather small, so constant composition was taken: 75 % phosphatidylethanolamine, 20 % phosphatidylglycerol and 5 % cardiolipin.

For DNA each nucleotide was assumed to be equally represented in molar units (Pramanik & Keasling, 1997). For RNA the distribution was 26.2 % ATP, 32.2 % GTP, 20.0 % CTP and 21.6 % UTP in molar units (Pramanik & Keasling, 1997).

The biomass used in the metabolic model had a molecular mass of 26.22 and the following elemental composition: $CH_{1.91}O_{0.506}N_{0.252}P_{0.015}S_{0.007}$. The molecular mass is somewhat higher

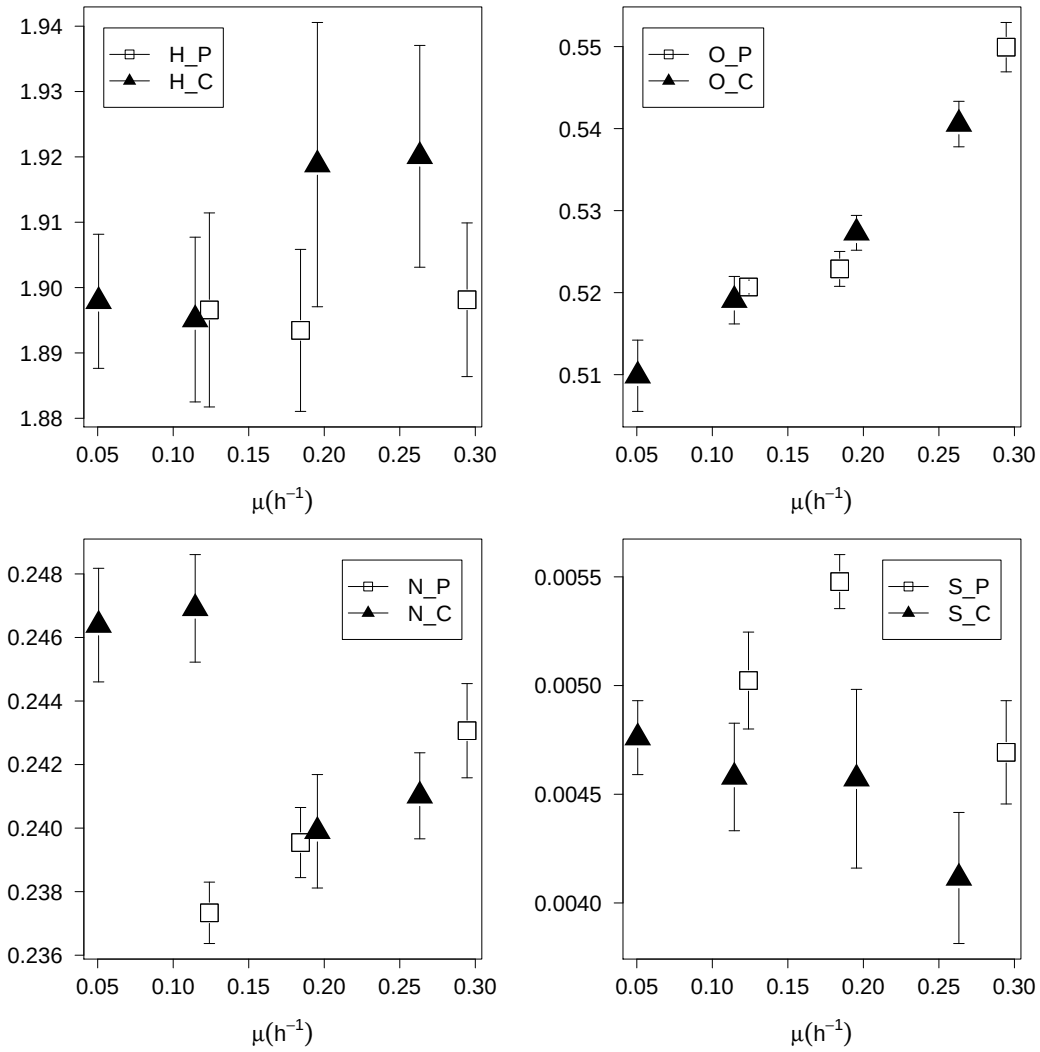


Figure 3.1: The elemental composition of biomass (expressed in C-moles) at different dilution rates. H (upper left), O (upper right), N (lower left) and S (lower right). The triangles represent the values for C-limited cultures and the squares those for the P-limited ones. Error bars are for the standard deviation.

as what would be expected from figure 3.2, because the biomass composition was mainly based on data from Carlson & Srienc (2004b), as it is difficult to extract the biomass composition from the elemental composition data (figure 3.1).

3.2.2 The different reactions

The metabolic model included the glycolysis, with glucose transported by the PTS system (Chen *et al.*, 1997), the pentose phosphate pathway, the Krebs cycle, ethanol, acetate and formate formation.

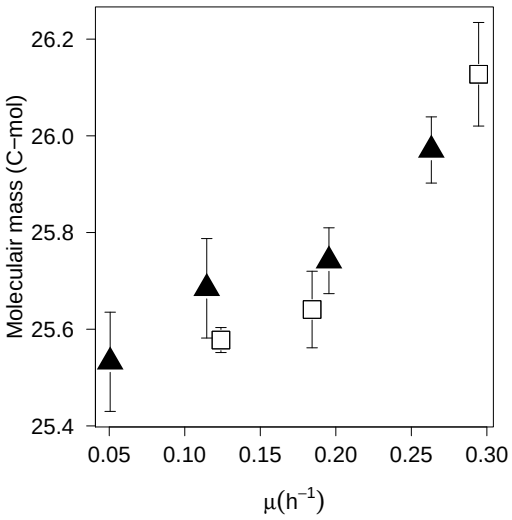


Figure 3.2: The molar mass of the biomass (in C-moles) as function of the dilution rate. Notation as in figure 3.1.

It is generally assumed that the glyoxylate pathway is not active in the *E. coli* K12 family (from which the strain used in this study is a member) if glucose is provided as a carbon source (Noronha *et al.*, 2000; Phue & Shiloach, 2004; Cozzone, 1998). Instead PEP carboxylase was used as regenerating reaction for the Krebs cycle metabolites (Sauer *et al.*, 1999; Wick *et al.*, 2001). The difference between both pathways is that PEP carboxylase uses one more ATP (Wick *et al.*, 2001).

For each amino acid and nucleotide the anabolic reactions were included. In order to avoid parallel pathways (and thus parts of the model that can not be solved), no 'salvage' pathways were used. Biosynthesis of LPS, lipid A, peptidoglycane and the lipid bilayer were also added. As explained above, a simple biomass composition was used for every growth rate.

Shikimic acid is an intermediate metabolite in the aromatic acid pathway. Not only shikimic acid is excreted but also dehydroquinic acid (precursor of shikimic acid), quinic acid, dehydroshikimate, protocatechuate and gallic acid (Draths & Frost, 1991). Those last two were not detected in the fermentation broth and thus were not included in the model.

Sources for the reactions were mainly the ecocyc database (`http://www.ecocyc.org/`)(Keseler *et al.*, 2005), the database provided by the university of California (`http://systemsbiology.ucsd.edu/organisms/ecoli.html`)(Reed *et al.*, 2003) and the KEGG database (`http://www.genome.ad.jp/kegg/`)(Kanehisa & Goto, 2000). The P/O ratio was set to 1.33 (Varma & Palsson, 1993).

Thus the constructed model contained 137 reactions (figure 3.3) and 151 metabolites of which 16 were exchangeable: Dhq, H_2SO_4, Ac, Cit, Shi, H_2O, O_2, PiOH, GLC, CO_2, Eth, Qa, Biom,

Dhs, NH_3 and FA. All parallel pathways were removed. There were no dead ending reactions and the elemental consistency test was passed (as described in section 2.4).

There are 143 independent equations in the model and $137 + 16$ unknowns. Thus, at least 10 measurements must be performed. Twelve exchange fluxes were measured: O_2, GLC, CO_2, Cit, FA, Shi, Dhs, Dhq, Qa, Ac, Eth and Biom. The consistency of the measurements could thus be assessed and if the consistency test failed, the vector comparison test could be used to search for the dissonant measurement.

3.3 Results and discussion

3.3.1 Flux analysis

The data

Ten chemostat experiments with the W3110.shik1 strain, were performed in which 12 exchange rates were measured. As explained above two measurements were redundant. The statistical test to evaluate whether the measurements agree could be run, and even when the measurements did not agree the vector comparison test could point to the wrong measurement. However, this is two extra measurements is rather low for the vector comparison test and the results should be interpreted cautiously.

Although no significant difference between the carbon in and outflux could be detected (figure 3.4), in two out of 10 experiments, the measurements were not consistent. Running the vector comparison test revealed in one case that the consumed amount of oxygen was too high. Therefore, this measurement was removed.

In a second experiment, the ethanol measurement was detected as most probable cause for inconsistency. But removing this measurement gave a high ethanol flux prediction, which was never detected in reality for any of the dilution rates. However, in this dataset the oxygen measurement was identified as another possible cause for the non-fit of the model. Removing the oxygen measurement and keeping the ethanol measurement, yielded a consistent dataset and realistic model predictions.

The different fluxes

Figure 3.5 shows the relative fluxes (in all the figures 3.5, 3.8 and 3.9 each molar flux is divided by the biomass flux) for the glycolysis and the pentose phosphate pathway. Figure 3.10 gives the most important fluxes in C-moles/C-moles of biomass. It can be seen that for the carbon-limited cultures, each mole of glucose gives two moles of biomass (the graph for the PTS reaction shows that the relative flux of glucose going into the cell is 0.5, thus to produce one mole of biomass, 0.5 mol of glucose is consumed), or put in another way, three

PGI: G6P ⟷ F6P
PFK: ATP + F6P ⟶ ADP + FBP
ALD: FBP ⟷ G3P + DHAP
TPI: DHAP ⟷ G3P
G3PDH: PiOH + NAD + G3P ⟷ NADH + H + BPG
PGK: ADP + BPG ⟷ ATP + 3PG
PGM: 3PG ⟷ 2PG
ENO: 2PG ⟷ H2O + PEP
PyrK: ADP + PEP ⟶ ATP + Pyr
PyrD: NAD + Pyr + CoA ⟶ NADH + H + AcCoA + CO2
CitSY: H2O + AcCoA + OAA ⟶ CoA + Cit
ACO: Cit ⟷ iCit
CitDH: NAD + iCit ⟷ NADH + H + CO2 + aKGA
AKGDH: NAD + CoA + aKGA ⟶ NADH + H + CO2 + SucCoA
SucCoASY: ADP + PiOH + SucCoA ⟷ ATP + CoA + Suc
SucDH: FAD + Suc ⟶ FADH2 + Fum
FumHY: H2O + Fum ⟷ Mal
MalDH: NAD + Mal ⟷ NADH + H + OAA
PEPCB: H2O + PEP + CO2 ⟶ PiOH + OAA
PFLY: Pyr + CoA ⟶ AcCoA + FA
EthDHLR: 2 NADH + 2 H + AcCoA ⟷ 2 NAD + CoA + Eth
AcKNLR: ADP + PiOH + AcCoA ⟷ ATP + CoA + Ac
Resp: 1.33 ADP + 1.33 PiOH + NADH + H + 0.5 O2 ⟶ 1.33 ATP + NAD + 2.33 H2O
H2CO3SY: H2O + CO2 ⟷ H2CO3
G6PDH: NADP + G6P ⟶ NADPH + H + 6PGL
LAS: H2O + 6PGL ⟶ 6PG
PGDH: NADP + 6PG ⟶ NADPH + H + CO2 + Rl5P
PPI: Rl5P ⟷ R5P
PPE: Rl5P ⟷ Xu5P
TK1: R5P + Xu5P ⟷ G3P + S7P
TA: G3P + S7P ⟷ F6P + E4P
TK2: Xu5P + E4P ⟷ F6P + G3P
PTS: GLC + PEP ⟶ G6P + Pyr
PPiOHHY: PPiOH + H2O ⟶ 2 PiOH
GluDH: NADPH + H + aKGA + NH3 ⟷ NADP + H2O + Glu
GluLI: ATP + NH3 + Glu ⟶ ADP + PiOH + Gln
AspSY: ATP + H2O + Asp + Gln ⟶ AMP + PPiOH + Asn + Glu
AspTA: OAA + Glu ⟷ aKGA + Asp
AlaTA: Pyr + Glu ⟷ aKGA + Ala
ValAT: aKIV + Glu ⟷ aKGA + Val
LeuSYLR: NAD + H2O + AcCoA + aKIV + Glu ⟶ NADH + H + CoA + CO2 + aKGA + Leu
aKIVSYLR: NADPH + H + 2 Pyr ⟶ NADP + H2O + CO2 + aKIV
IleSYLR: NADPH + H + Pyr + Glu + Thr ⟶ NADP + H2O + CO2 + aKGA + NH3 + Ile
ProSYLR: ATP + 2 NADPH + 2 H + Glu ⟶ ADP + PiOH + 2 NADP + H2O + Pro
SerLR: NAD + H2O + 3PG + Glu ⟶ PiOH + NADH + H + aKGA + Ser
SerTHM: Ser + THF ⟶ H2O + Gly + MeTHF
H2SSYLR: 2 ATP + 3 NADPH + ThioredH2 + 3 H + H2SO4 ⟶ ADP + PPiOH + 3 NADP + Thiored + 3 H2O + H2S + PAP
PAPNAS: H2O + PAP ⟶ AMP + PiOH
CysSYLR: H2S + AcCoA + Ser ⟶ CoA + Cys + Ac
PrppSY: ATP + R5P ⟶ AMP + PRPP
HisSYLR: ATP + 2 NAD + 3 H2O + Gln + PRPP ⟶ 2 PPiOH + PiOH + 2 NADH + 2 H + aKGA + His + AICAR
PheSYLR: Glu + Chor ⟶ H2O + CO2 + aKGA + Phe
TyrSYLR: NAD + Glu + Chor ⟶ NADH + H + CO2 + aKGA + Tyr
TrpSYLR: Gln + Ser + Chor + PRPP ⟶ PPiOH + 2 H2O + G3P + Pyr + CO2 + Glu + Trp
DhDoPHepAD: H2O + PEP + E4P ⟶ PiOH + Dahp
DhqSY: Dahp ⟶ PiOH + Dhq
DhsSYLR: Dhq ⟷ H2O + Dhs
ShiSY: NADPH + H + Dhs ⟷ NADP + Shi
ShiKN: ATP + Shi ⟶ ADP + Shi3P
DhqDH: NADPH + H + Dhq ⟶ NADP + Qa
ChorSYLR: PEP + Shi3P ⟶ 2 PiOH + Chor
ThrSYLR: ATP + H2O + HSer ⟶ ADP + PiOH + Thr
MDAPSYLR: NADPH + H + Pyr + SucCoA + Glu + AspSA ⟶ NADP + CoA + aKGA + Suc + MDAP
LysSY: MDAP ⟶ CO2 + Lys
MetSYLR: H2O + SucCoA + Cys + MTHF + HSer ⟶ Pyr + CoA + Suc + NH3 + Met + THF
AspSASY: ATP + NADPH + H + Asp ⟶ ADP + PiOH + NADP + AspSA
HSerDH: NADPH + H + AspSA ⟷ NADP + HSer
CarPSY: 2 ATP + H2O + H2CO3 + Gln ⟶ 2 ADP + PiOH + Glu + CarP
OrnSYLR: ATP + NADPH + H + H2O + AcCoA + 2 Glu ⟶ ADP + PiOH + NADP + CoA + aKGA + Orn + Ac
ArgSYLR: ATP + Asp + Orn + CarP ⟶ AMP + PPiOH + PiOH + Fum + Arg
ThioredRD: NADPH + Thiored + H ⟷ NADP + ThioredH2
H2O2ox: 2 H2O2 ⟶ 2 H2O + O2
FAD2NAD: NAD + FADH2 ⟷ NADH + FAD + H
AICARSYLR: 6 ATP + 3 H2O + CO2 + Asp + 2 Gln + Gly + FA + PRPP ⟶ 6 ADP + PPiOH + 6 PiOH + Fum + 2 Glu + AICAR
IMPSYLR: FTHF + AICAR ⟶ H2O + THF + IMP
AMPSYLR: Asp + GTP + IMP ⟶ AMP + PiOH + Fum + GDP
AdKN: ATP + AMP ⟷ 2 ADP
ADPRD: ADP + ThioredH2 ⟶ Thiored + H2O + dADP
dADPKN: ATP + dADP ⟶ ADP + dATP
IMPDH: NAD + H2O + IMP ⟶ NADH + H + XMP
GMPSY: ATP + H2O + Gln + XMP ⟶ AMP + PPiOH + Glu + GMP
GuKN: ATP + GMP ⟶ ADP + GDP
GDPKN: ATP + GDP ⟶ ADP + GTP
GDPRD: ThioredH2 + GDP ⟶ Thiored + H2O + dGDP
dGDPKN: ATP + dGDP ⟶ ADP + dGTP
UMPSYLR: O2 + Asp + PRPP + CarP ⟶ PPiOH + PiOH + H2O + CO2 + UMP + H2O2
UrKN: ATP + UMP ⟶ ADP + UDP
UDPKN: ATP + UDP ⟶ ADP + UTP
CTPSY: ATP + H2O + Gln + UTP ⟶ ADP + PiOH + Glu + CTP
CDPKN: ATP + CDP ⟷ ADP + CTP
CMPKN: ATP + CMP ⟶ ADP + CDP
CDPRD: ThioredH2 + CDP ⟶ Thiored + H2O + dCDP
dCDPKN: ATP + dCDP ⟶ ADP + dCTP
UDPRD: ThioredH2 + UDP ⟶ Thiored + H2O + dUDP
dUDPKN: ATP + dUDP ⟶ ADP + dUTP
dUTPPPAS: H2O + dUTP ⟶ PPiOH + dUMP
dTMPSY: MeTHF + dUMP ⟶ DHF + dTMP
dTMPKN: ATP + dTMP ⟶ ADP + dTDP
dTDPKN: ATP + dTDP ⟶ ADP + dTTP
DHFRD: NADPH + H + DHF ⟶ NADP + THF
FTHFSYLR: NADP + H2O + MeTHF ⟶ NADPH + H + FTHF
GlyCA: NAD + Gly + THF ⟷ NADH + H + CO2 + NH3 + MeTHF
MeTHFRD: NADH + H + MeTHF ⟶ NAD + MTHF
AcCoACB: ATP + H2O + AcCoA + CO2 ⟷ ADP + PiOH + MalCoA
MalCoATA: MalCoA + ACP ⟷ CoA + MalACP
AcACPSY: MalACP ⟶ CO2 + AcACP
C120SY: 10 NADPH + 10 H + AcACP + 5 MalACP ⟶ 10 NADP + 5 H2O + 5 CO2 + C120ACP + 5 ACP
C140SY: 12 NADPH + 12 H + AcACP + 6 MalACP ⟶ 12 NADP + 6 H2O + 6 CO2 + C140ACP + 6 ACP
C160SY: 14 NADPH + 14 H + AcACP + 7 MalACP ⟶ 14 NADP + 7 H2O + 7 CO2 + C160ACP + 7 ACP
C181SY: 15 NADPH + 15 H + AcACP + 8 MalACP ⟶ 15 NADP + 8 H2O + 8 CO2 + C181ACP + 8 ACP
AcylTF: C160ACP + C181ACP + Go3P ⟶ 2 ACP + PA
Go3PDH: NADPH + H + DHAP ⟷ NADP + Go3P
DGoKN: ATP + DGo ⟶ ADP + PA
CDPDGoSY: CTP + PA ⟷ PPiOH + CDPDGo
PSerSY: Ser + CDPDGo ⟶ CMP + PSer
PSerDC: PSer ⟶ CO2 + PEthAn
GlnF6PTA: F6P + Gln ⟶ Glu + GA6P
GlcAnMU: GA6P ⟷ GA1P
NAGUrTF: AcCoA + UTP + GA1P ⟶ PPiOH + CoA + UDPNAG
LipaSYLR: ATP + 2 CMPKDO + 2 UDPNAG + C120ACP + 5 C140ACP ⟶ ADP + 2 CMP + UMP + UDP + 6 ACP + Lipa + 2 Ac
A5PIR: Rl5P ⟷ Ar5P
PGLCMT: G6P ⟷ G1P
CMPKDOSYLR: 2 H2O + PEP + Ar5P + CTP ⟶ PPiOH + 2 PiOH + CMPKDO
ADPHEPSY: ATP + S7P ⟶ PPiOH + ADPHEP
UDPGlcSY: G1P + UTP ⟶ PPiOH + UDPGlc
EthANPT: CMP + PEthAn ⟷ CDPEthAn + DGo
LpsSYLR: 3 ADPHEP + 3 CMPKDO + 2 UDPGlc + Lipa + 2 CDPEthAn ⟶ 3 ADP + 3 CMP + 2 CDP + 2 UDP + Lps
PGSYLR: Go3P + CDPDGo ⟶ PiOH + CMP + PG
CLSY: PG + CDPDGo ⟶ CMP + CL
PeptidoSYLR: 5 ATP + NADPH + H + PEP + 3 Ala + MDAP + 2 UDPNAG ⟶ 5 ADP + 7 PiOH + NADP + UMP + UDP + Peptido
GlcgSY: ATP + G1P ⟶ ADP + PPiOH + Glcg
ATPHY: ATP + H2O ⟶ ADP + PiOH
DNASYLR: 2 H2O + 0.246 dATP + 0.254 dGTP + 0.254 dCTP + 0.246 dTTP ⟶ 2 PiOH + DNA
RNASYLR: 0.262 ATP + 2 H2O + 0.322 GTP + 0.2 CTP + 0.216 UTP ⟶ 2 PiOH + RNA
ProtnSYLR: 2 ATP + 4 H2O + 0.0961 Ala + 0.05506 Arg + 0.04505 Asn + 0.04505 Asp + 0.01702 Cys + 0.04905 Gln + 0.04905 Glu + 0.1151 Gly + 0.01802 His + 0.05405 Ile + 0.08408 Leu + 0.06406 Lys + 0.02903 Met + 0.03504 Phe + 0.04104 Pro + 0.04004 Ser + 0.04705 Thr + 0.01101 Trp + 0.02603 Tyr + 0.07908 Val + 2 GTP ⟶ 2 ADP + 4 PiOH + 2 GDP + Protn
LipidSYLR: 0.0266 CL + 0.202 PG + 0.7714 PEthAn ⟶ Lipid
BiomSYLR: 0.003472 Glcg + 0.0002027 Lps + 0.0006801 Peptido + 0.002408 DNA + 0.009191 RNA + 0.1454 Protn + 0.002774 Lipid ⟶ Biomn

Figure 3.3: List of reactions used in the model. A description of the metabolites can be found in appendix B.1.

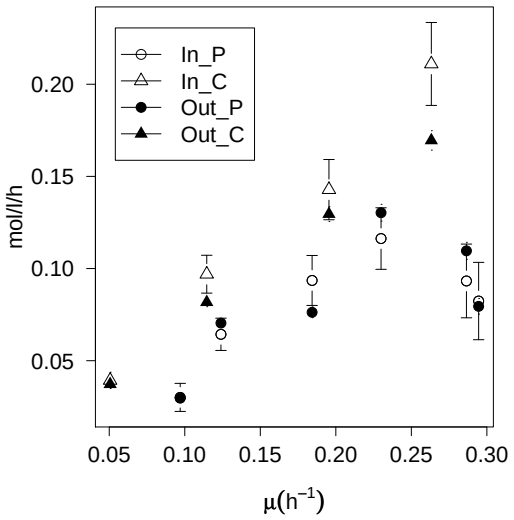


Figure 3.4: Moles of carbon going in the reactor versus moles of carbon leaving the reactor. Open symbols are for influxes, closed symbols for the carbon leaving the reactor. Triangles are for C-limited cultures and bullets for P-limited ones. Error bars represent the standard deviation.

of the six carbons of glucose are used for biomass construction (this can also nicely be seen in figure 3.10 where three C-moles of per C-mole of biomass is taken into the cell). The same graph also shows that under P-limited conditions, at low dilution rates, the biomass yield is lower; but strangely at a dilution rate of $0.25\,h^{-1}$ it tends to become better than under carbon-limitation (less glucose is consumed per mole of biomass produced than under carbon-limitation). Looking at the ATP hydrolysis (figure 3.6) and the respiration (figure 3.7) there is no difference between the carbon-limited and phosphate-limited cultures at higher dilution rates. Thus, the better biomass yield of P-limited cultures at high dilution rates is not due to less CO_2 production or less ATP hydrolysis. The reason for this should be found in the aromatic pathway.

In the first reaction of the aromatic pathway, one can see that for high dilution rates the flux through, the synthesis reaction of Dahp (figure 3.9), is higher for carbon-limited cultures than for phosphate-limited ones. Considering that the fluxes through the chorismate synthesis reaction are all equal (because a fixed biomass composition was used) the increased flux through the Dahp formation reaction necessarily means an increased excretion of products upstream of chorismate. Unfortunately, those byproducts were found not to be shikimic acid, the target molecule, but dehydroquinate and dehydroshikimate. Thus, carbon-limited cultures have a larger flux through the shikimate pathway but do not produce more shikimate because the extra flux is wasted on other byproducts. If this byproduct excretion could be eliminated, carbon-limited cultures would clearly be better for shikimate production.

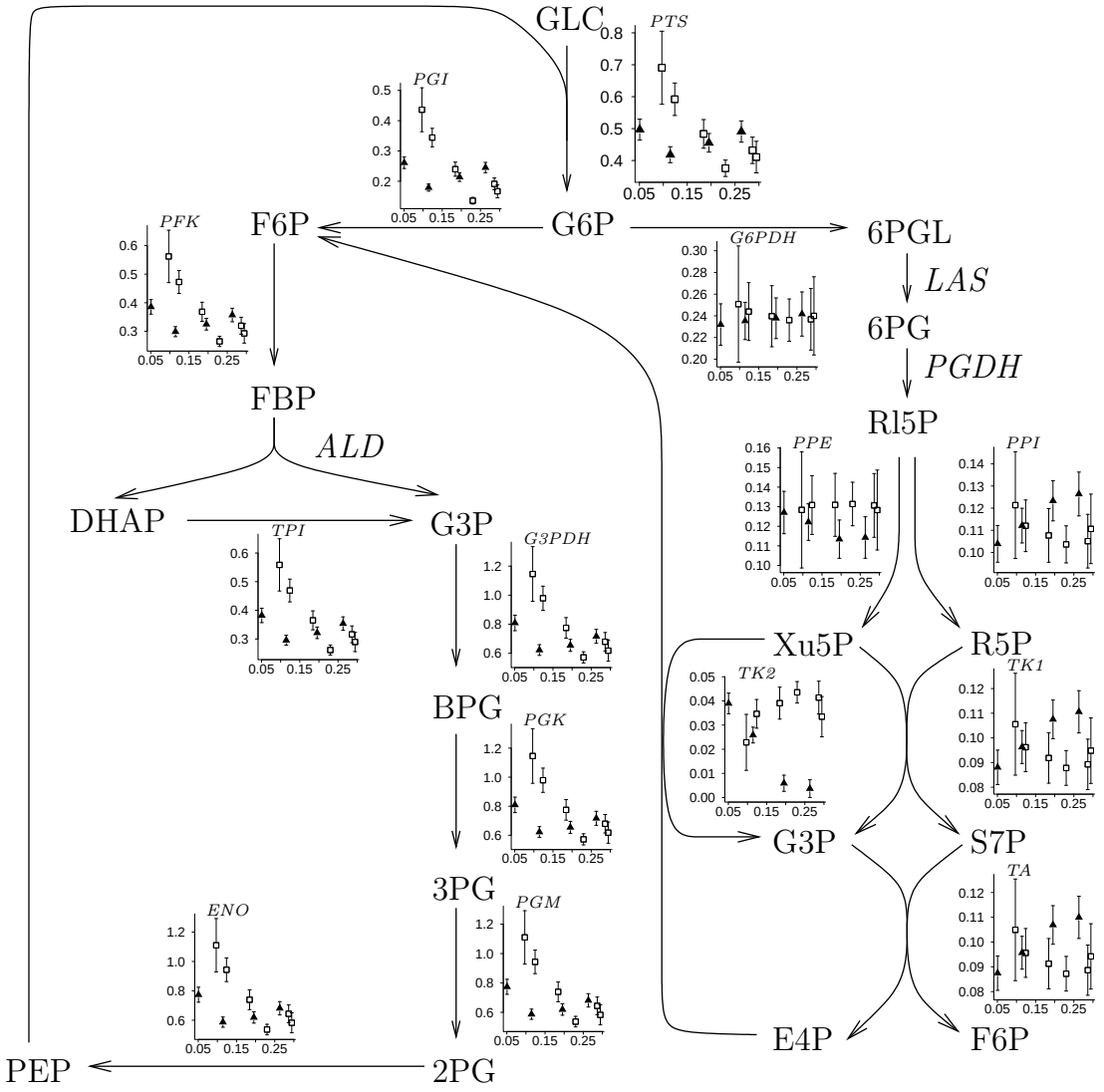


Figure 3.5: Relative fluxes (mol/mol Biom , each flux is normalized against the biomass flux) in the glycolysis and pentose phosphate pathway for different growth rates. Squares are used for the phosphate-limited cultures and triangles represent the data from the carbon-limited cultures. Metabolites are typeset in upright font, names of reactions in italic. Abscissa: growth rate, ordinate: relative fluxes. The error bars represent the standard deviation.

The larger flux through the shikimate route for carbon-limited cultures gives no significant rise in the G6P fluxes that enter the pentose phosphate pathway. The differentiation can be found in the amount of carbon that goes from the PPP to the glycolysis pathway via F6P and G3P. A split occurs at Rl5P. More E4P is needed to sustain the production of Dahp under carbon-limitation, therefore the reaction catalysed by TK2 is lower or even zero at higher dilution rates for carbon-limited cultures This makes more Xu5P available, which in turn allows the TK1 flux to increase. This allows the TA flux to increase, leading to a higher

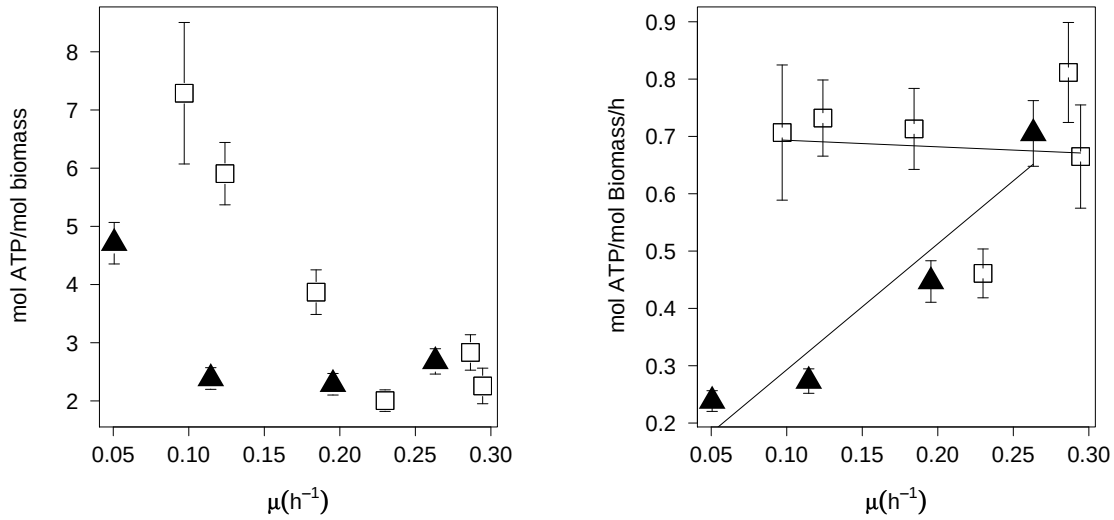


Figure 3.6: Left: mole ATP hydrolysed per mole of biomass formed. Right: mole ATP per mole of biomass per hour that is hydrolysed. From this data, the maintenance coefficients are calculated, assuming a constant P/O ratio of 1.33. Notation as in figure 3.5.

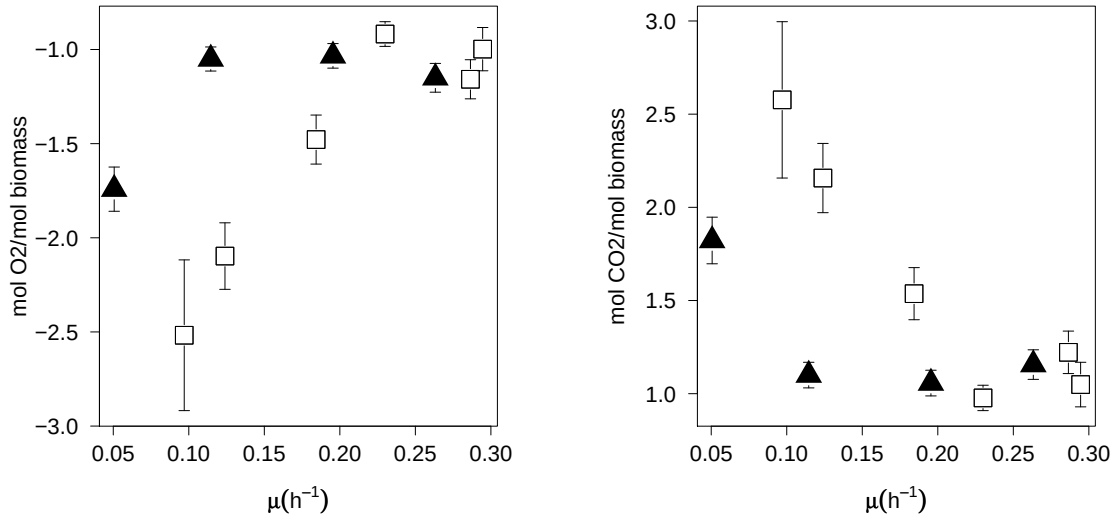


Figure 3.7: Moles of oxygen consumed (left) and moles of carbon dioxide produced (right) per mole of biomass formed. Notation as in figure 3.5.

E4P production. As, under carbon-limitation, the flux through TK2 is lower, the one through PPE can also be lower while the one through PPI is higher. This is in line with the studies where an increased flux to the shikimate pathway was achieved by overexpression of Tk1. An amplification of this enzyme in combination with an overexpressed DAHP-synthase has shown to double the flow into the pathway in comparison to the case where only DAHP-synthase was overexpressed (Draths *et al.*, 1992).

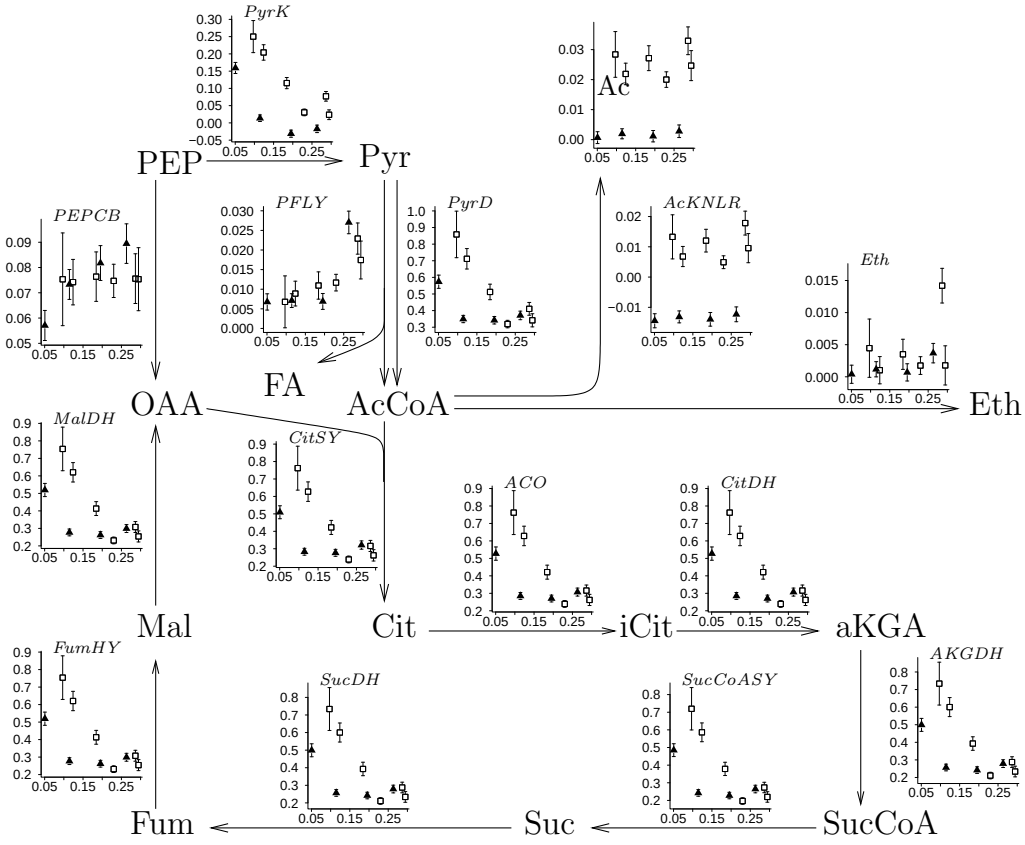


Figure 3.8: Fluxes in the Krebs cycle and the fermentative pathways. Notation as in figure 3.5.

The fluxes through glycolysis (figure 3.5) follow the same pattern as those from ATP hydrolysis (figure 3.6) and respiration (figure 3.7). This pathway is mainly used to fuel the citric acid cycle (figure 3.8) for generating ATP and biomass precursors. Interesting to note is the negative flux in carbon-limited cultures going through AcKNLR (figure 3.8). Actually, no acetate was supplied to the medium. The acetate consumed is provided by the cysteine, ornithine and lipid A synthesis reactions. Acetate consumption in the presence of glucose, although uncommon, was reported in literature (Varma & Palsson, 1994a).

The flux through PEP carboxylase (figure 3.8) is completely dependent on the biomass formation reaction (as is the one through the chorismate synthesis reaction, figure 3.9). This can nicely be seen as the constant flux for every phosphate-limited culture. For the carbon-limited cultures this is not true, as there was some citric acid consumed in the experiment with the lowest growth rate while it was produced in the other ones. The amounts of citric acid consumed or produced were however very low: there is no significant difference between the carbon-limited and phosphate-limited cultures in the flux through PEP carboxylase.

The calculated ATP hydrolysis flux allows to investigate the maintenance requirement for

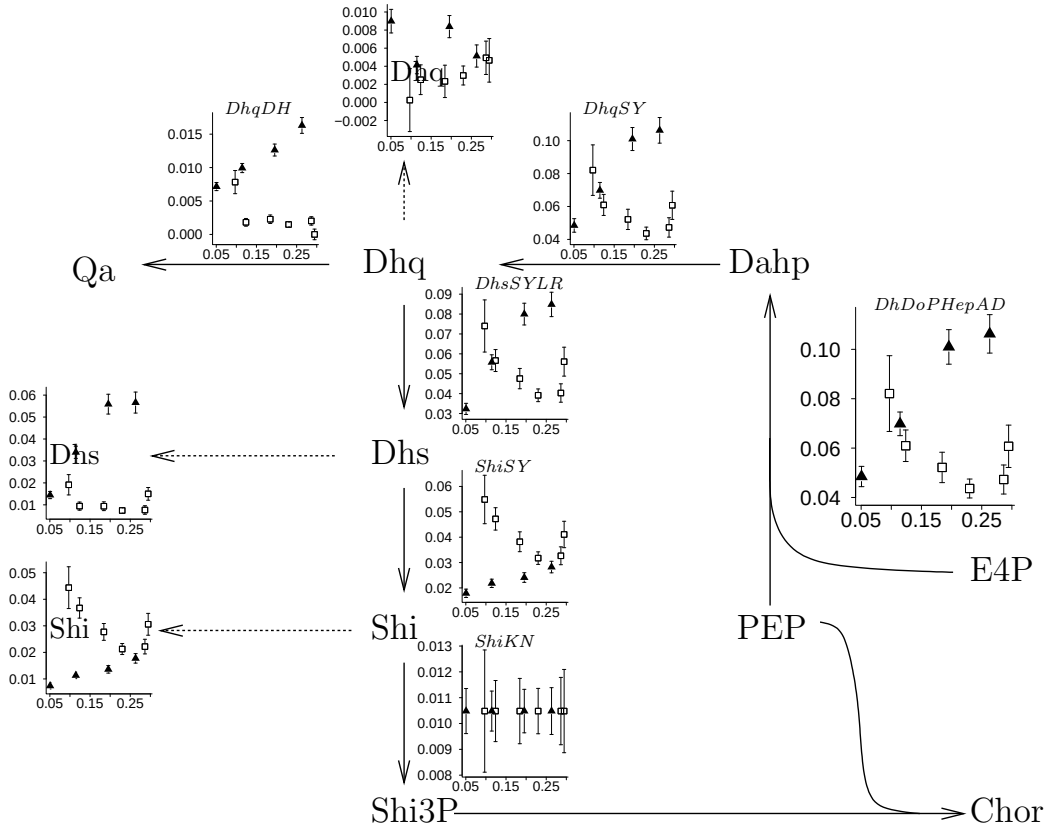


Figure 3.9: Fluxes in the shikimate pathway. Notation as in figure 3.5.

different growth rates and limiting conditions (figure 3.6). Phosphate-limited cultures have a higher non-growth-associated maintenance compared to carbon-limited ones, as under carbon-abundance there is no need for an efficient carbon utilisation. The non-growth-associated maintenance for phosphate-limited cultures is 0.71 mole ATP per mole of biomass per hour. For carbon-limited cultures it is 0.073 mol/mol BM/h. This is conform the values reported in literature: 0.20 mol/mol BM/h (Varma & Palsson, 1994a), 0.073 mol/mol BM/h (Kayser *et al.*, 2005) and 0.12 mol/mol BM/h (Carlson & Srienc, 2004a).

The growth-associated maintenance for phosphate-limited cultures was around zero. For carbon-limitation it was found to be 2.19 mol/mol BM. In the literature both low values, 0.34 mol/mol BM (Varma & Palsson, 1994a), and high values, 2.6 mol/mol BM (Carlson & Srienc, 2004a) can be found.

Apparently under carbon-abundant conditions, the cells, even at the lowest growth rate, produce as much ATP as possible and this production does not significantly increase (can not increase due to P-limitation?) with higher growth rates. Carbon-limited cultures want

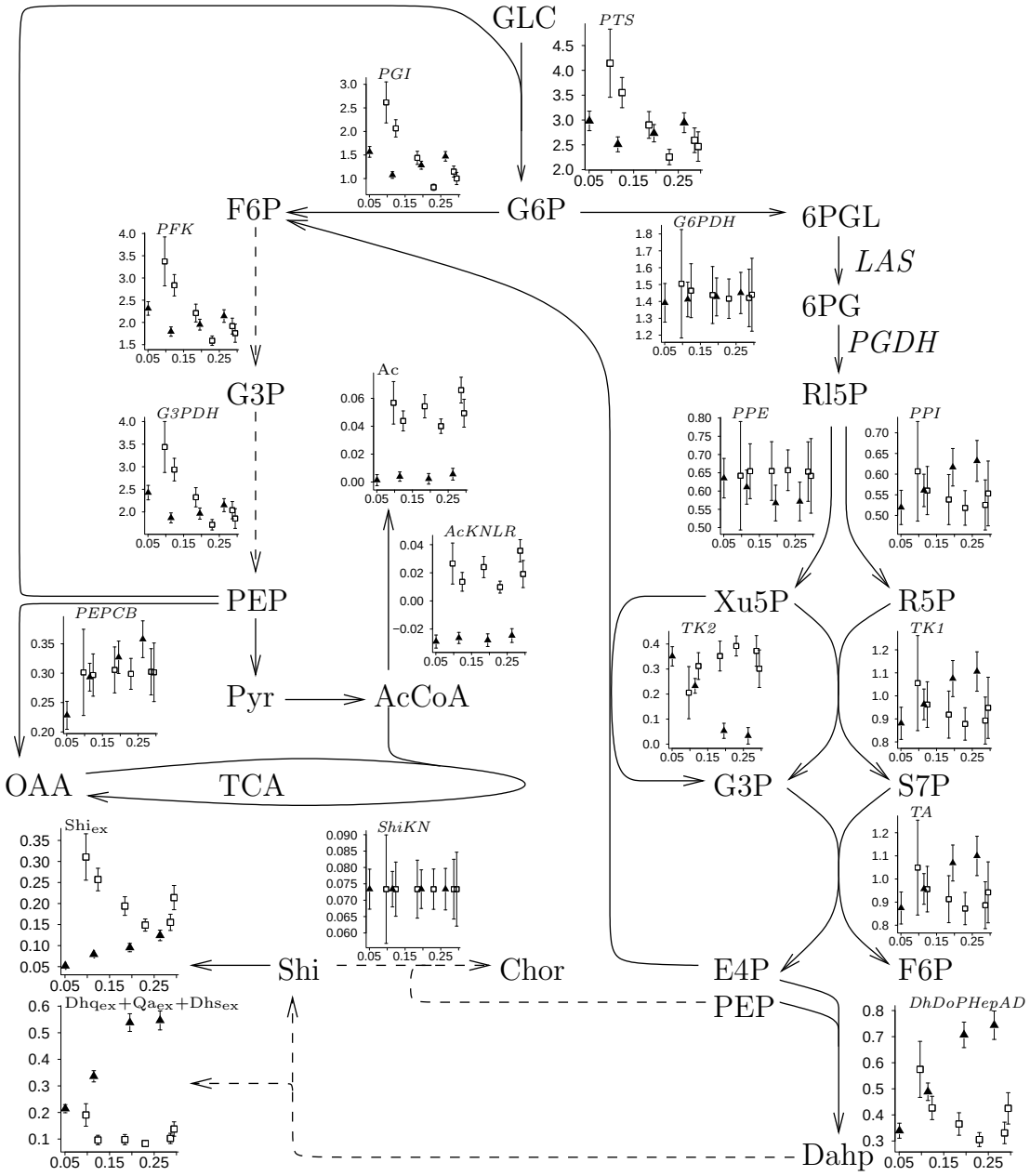


Figure 3.10: Summary of figures 3.5, 3.8 and 3.9 in which the relative fluxes (as compared to the biomass flux) are expressed in C-moles. Broken lines are for lumped reactions. TCA: Tricarboxylic acid cycle. Notations as in figure 3.5.

to optimise the utilisation of the available glucose and try to minimise the maintenance cost resulting in a high growth dependent maintenance. But the total maintenance is always lower than in carbon-abundant cultures (in the right part of figure 3.6, the curve of the P-limited cultures, squares, lies above the curve of the C-limited ones, triangles). At a growth rate of $0.28\,\mathrm{h}^{-1}$ both curves meet each other. This is also around the point where the cells start to

wash out. This suggests that the maintenance level of the phosphate-limited cultures is really the maximum ATP production that the cell can sustain.

3.3.2 RNA expression compared with metabolic fluxes

To investigate whether there exists a correlation between RNA expression levels and metabolic fluxes, four different kinds of cultures were performed, in which flux data and RNA expression levels were determined. It should be noted that those experiments were completely different from the experiments described above and are fully discussed in Johansson & Lidén (2006). Two different experiments were done with the wild-type strain (*E. coli* W3110), one carbon-limited and one phosphate-limited and two experiments were done with the modified strain (*E. coli* W3110.shik1), also one carbon- and one phosphate-limited. Each experiment was performed twice. The continuous cultures were all run at a dilution rate of $0.2\,\mathrm{h}^{-1}$.

In general RNA expression levels are compared by taking the $\log_2$ of the fraction of expression levels under modified conditions against expression levels in a reference state. In this study, the mean of the values found at the four different conditions was used as reference state and not the wild-type fermentation under carbon- or phosphate-limited conditions. This because as the RNA expression levels have to be compared with the fluxes, the same mathematical treatment should be applied to the flux values (relative molar fluxes were used: each flux expressed in $\mathrm{mol\,l}^{-1}\,\mathrm{h}^{-1}$ was divided by the biomass flux). However, for the wild-type, some fluxes in the shikimate producing pathway are zero (more precisely, there is no quinic acid production). For the modified strain, some other fluxes are zero (no formic acid formation). Thus, for those fluxes, no good reference state exists. Therefore, the mean of the four values was used as reference state, each value was divided by it and the $\log_2$ was taken. This was done for both the RNA expression levels and the fluxes that were not zero (figures 3.11, 3.12 and 3.13).

The flux through a reaction is not always controlled by the enzyme level, as metabolic control analysis shows, but can also be limited by the availability of reactants. It is not expected that there is a direct correlation between RNA expression levels and enzyme levels, even less between RNA expression levels and metabolic fluxes. Only when RNA expression is absent, one can say for sure that the corresponding flux is zero. And the absence of expression of a gene can only be disproved, not proven (Åkesson *et al.*, 2004). However, in some cases there can be qualitative correlations between RNA expression levels and metabolic fluxes (Oh & Liao, 2000).

For some (key) reactions an increase in flux correlates with an increase or decrease in RNA expression levels (figures 3.11, 3.12 and 3.13). Some correlations were detected in the glycolysis and the aromatic pathway while in the PPP and the citric acid cycle the reactions seem to be driven by supply and demand and not by RNA expression of their genes; the differences

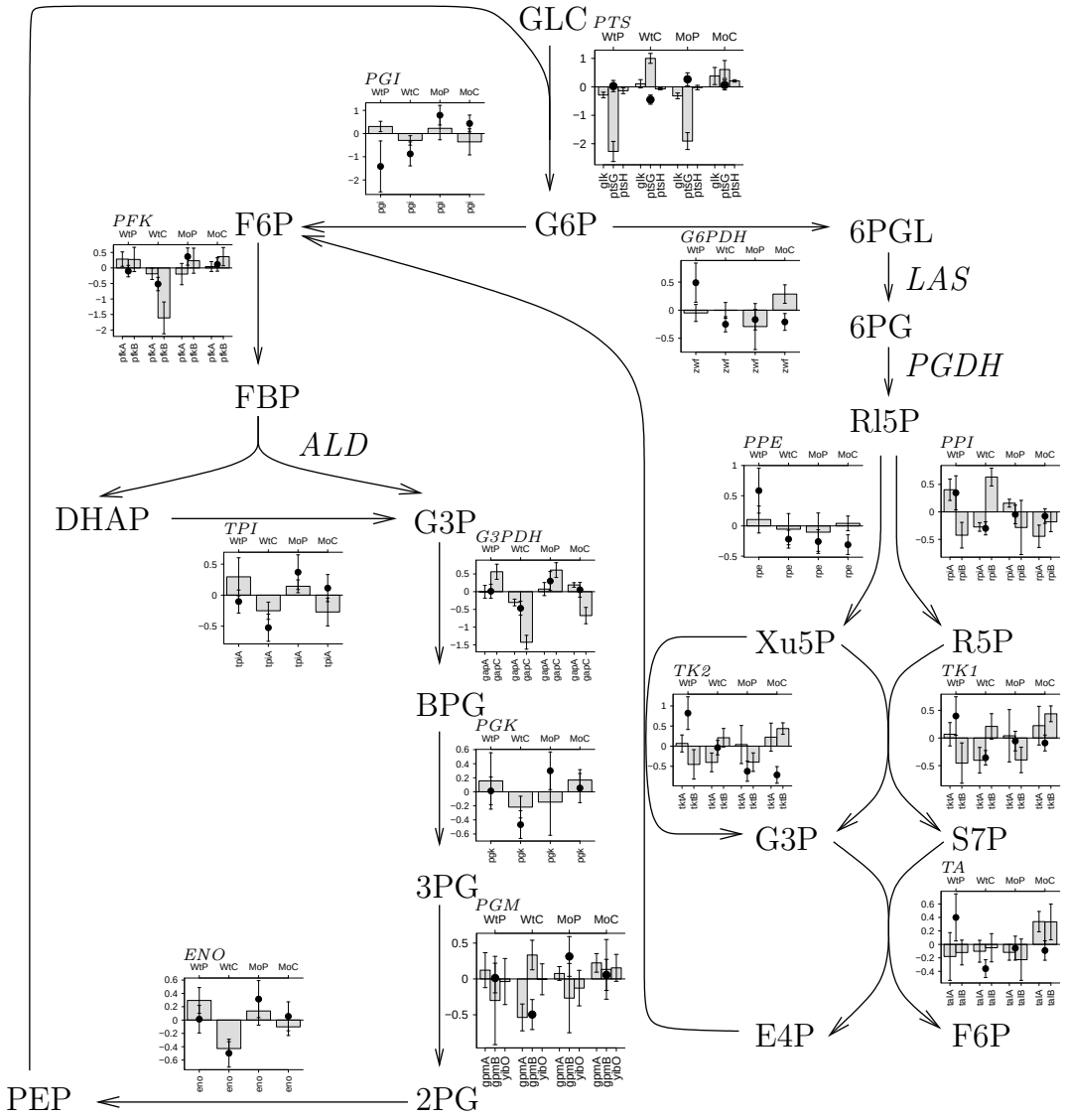


Figure 3.11: Fluxes compared with the RNA expression data for the central carbon metabolism. Abscissa: growth rate, ordinate: relative fluxes. The error bars represent the standard deviation. The dots represent the fluxes and the bars the gene expression levels. Both are log_2 transformed. The gene names can be found in the lower part of each graph while the upper part labels the different experiments: WtP for wild type strain, P-limited culture, WtC for wild type strain, C-limited culture, MoP for modified strain, P-limited culture and MoC for modified strain, C-limited culture. Reaction names are typeset in italic, metabolite names in upright fonts.

smaller than $\log_2$ (such as *MalSY* in figure 3.12) are not significant.

In the following paragraphs examples are given of these three cases: positive correlation, negative correlation and no correlation.

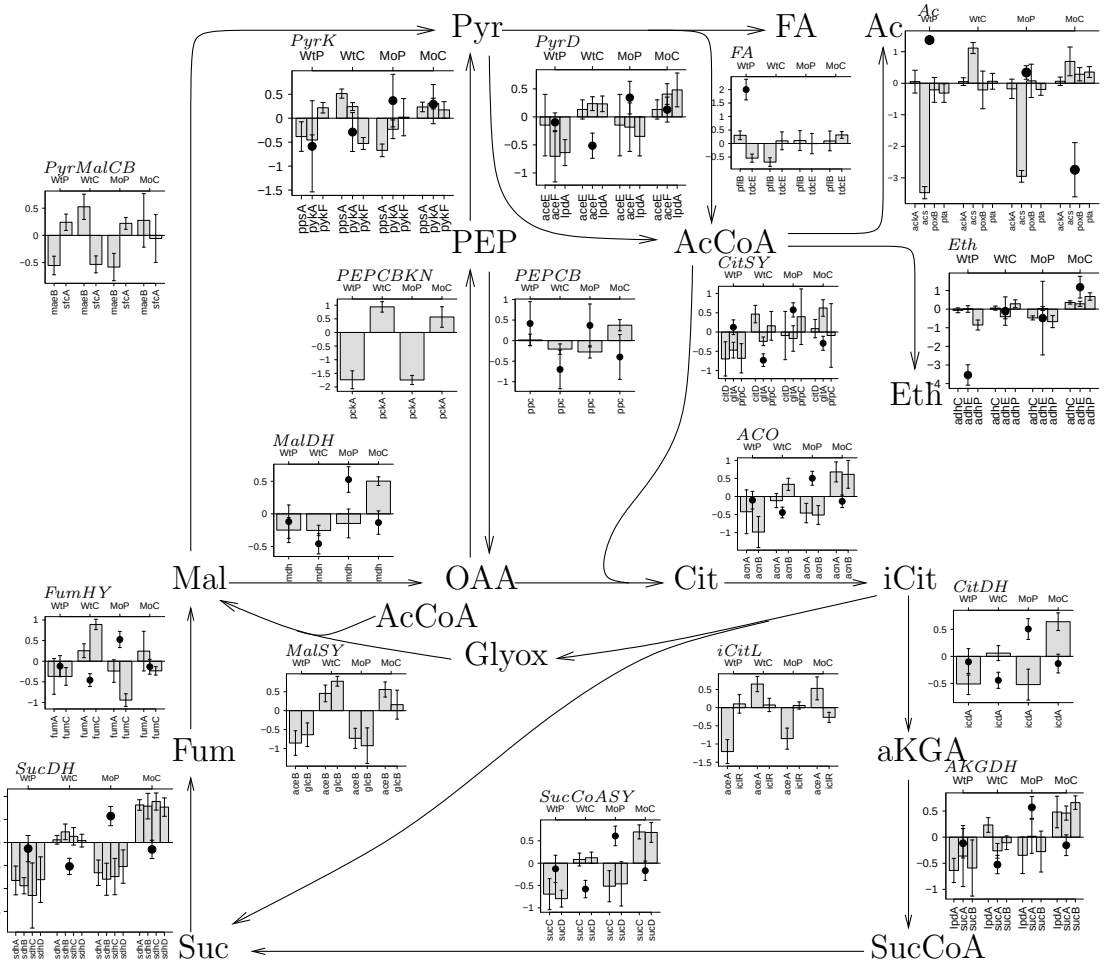


Figure 3.12: Fluxes compared with the RNA expression data for the Krebs cycle. Notation as in figure 3.11.

Positive correlation

The modified strain has the feedback inhibition acting on *aroF* (from tyrosine and phenylalanine) and *aroG* (from trypthophane and phenylalanine) removed. This can be seen both in the fluxes and in the RNA expression levels (figure 3.13).

The expression of *ydiB* does not seem to be correlated with the ShiSY reaction (figure 3.13), but it does correlate with the production of quinate. Carbon-limitation in the modified strain gives high intracellular shikimate concentrations (Johansson & Lidén, 2006) causing a higher expression of *ydiB* which gives a high flux through DhqDH. No effect is seen on *aroE*. As, *ydiB*, *aroF* codes for ShiSY but the gene product of *ydiB* also catalyses the conversion of dehydroquinate to quinate. However, this upregulation of *ydiB* is only seen in the modified strain, in which feedback inhibition of *aroF* and *aroG* is removed. Probably that in the wild-type, no accumulation of shikimate occurs due to proper regulation of the initial reaction of

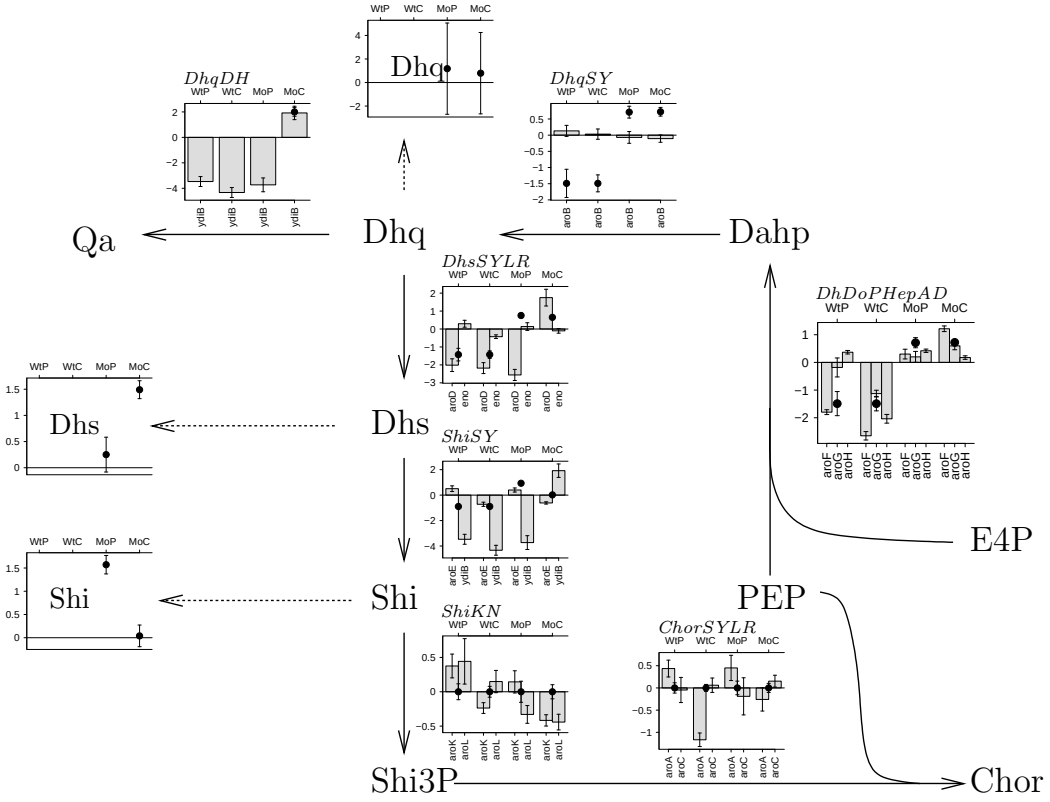


Figure 3.13: Fluxes compared with the RNA expression data in the aromatic amino acid pathway. Notation as in figure 3.11.

the aromatic pathway.

Negative correlation

Whereas in C-limited cultures *ptsG* is more expressed than in carbon-abundant ones, the specific uptake of glucose is lower (figure 3.11). An explanation might be that under such conditions the cell tries to optimise the uptake of carbon by expressing more *ptsG* (Hua *et al.*, 2004). It is known that glucokinase (coded by *glk*, figure 3.11) does not vary much under different conditions, as the main entrance for glucose into the *E. coli* cell is via PTS (Meyer *et al.*, 1997). It can be seen that the *glk* expression is almost constant. However, the gene product of *ptsH*, Hpr (heat stable, histidyl phosphorylatable protein) is also part of the PTS system. Hpr is the subunit responsible for transferring the phosphate group to the sugar imported in the cell while the gene product of *ptsG* is the subunit that imports the glucose molecule and releases it phosphatated in the cell (Postma *et al.*, 1993). Both have their own operon and promoters. It makes sense that to scavenge glucose, only the part of the enzyme responsible for attracting glucose is expressed.

Negative correlation with the flux is also found for *acs*, converting acetate to acetyl-CoA (figure 3.12): this gene codes for acetyl coenzyme A synthetase and is even expressed when no acetate is available (Shin *et al.*, 1997). This seems to fit the case here: under carbon-limiting conditions, the expression of this gene is stronger: more need for carbon? Indeed *poxB*, *ackA* and *pta* are not as influenced by the carbon needs of the cell: they code for enzymes involved in acetate production, not consumption (even if some of them are reversible).

The data suggest that carbon-abundant cultures have a low expression of genes coding for pyruvate dehydrogenase (*PyrD* in figure 3.12, *acdE*, *aceF* and *lpdA* are all members of the same operon and code for subunits of the pyruvate dehydrogenase complex) and a high flux through that reaction, while carbon-limited cultures show the opposite behaviour. This confirms the high non-growth and low growth-associated maintenance found for the phosphate-limited cultures: the flux through the citric acid cycle (mainly used for generating ATP via respiration) is saturated. This can also be observed in the Krebs cycle itself: the flux through it is higher in P- than in C-limitation while the genes are downregulated. Furthermore, the flux in the modified strain is generally higher than the one in the wild-type strain.

No correlation

In general the fluxes in the PPP pathway do not follow the RNA expression patterns (figure 3.11). Fluxes seem to be driven by the need to build blocks for the amino acid synthesis reactions, *in casu* the aromatic pathway. The lack of correlation of the fluxes with the RNA expression data could also be due to the high variance that all measurements have.

3.4 Conclusions

With the aid of MFA, it was shown that in the *E. coli* W3110.shik1 strain, the difference in shikimic acid yield between carbon-rich and carbon-limited cultures is not due to a lower flux in the aromatic amino acid pathway, but due to a larger excretion of dehydroshikimic acid and dehydroquinic acid. The flux entering the shikimate pathway was even higher in carbon-limited cultures and this due to, or caused by less carbon going from the pentose phosphate pathway to the glycolysis.

Under carbon abundance, the ATP production rate is maximal. It was shown that growth-associated maintenance, under P-limitation, is almost zero, but non-growth-associated maintenance is as high as the total maintenance under carbon-limitation at the wash out point.

Flux data were compared with RNA expression data. Most of the fluxes were not driven by the amount of expression of the corresponding gene. However, some fluxes were correlated, especially those from the altered genes: *aroF* and *aroG* feedback inhibition, giving a higher flux through the aromatic pathway, and *ydiB* overexpression, causing a high quinate production.

Finally, a number of important fluxes are negatively correlated with the RNA expression level: this to maximise the flux through that reaction, for example *ptsG* under carbon-limitation.

Chapter 4

Comparing different strategies to reduce acetate formation in *E. coli*[1]

E. coli cells produce acetate as an extracellular byproduct of aerobic cultivations under excess-glucose conditions. This phenomenon is referred to as overflow metabolism. Acetate is undesirable, because it retards growth even at concentrations as low as 0.5 g/l (Nakano *et al.*, 1997) and inhibits protein formation (Ko *et al.*, 1994). Most process designs or genetic modifications to overcome acetate formation ultimately aim to balance growth rate and oxygen consumption (Eiteman & Altman, 2006).

Process improvement approaches involve designing growth media or conditions that eliminate or reduce acetate formation. In contrast, genetic approaches involve altering the genetic profile of the strain itself, to restrict the biochemical synthesis of acetate. Process approaches generally focus on reducing the carbon flux entering in the cells and thus diminish the specific growth rate and efficiency of the production process.

For these reasons, this research focuses on genetic approaches to overcome acetate formation. Several interrelated methods have been used to reduce acetate formation genetically. The three partially overlapping strategies are (1) approaches that directly reduce glucose consumption, (2) approaches that directly reduce carbon flow to acetate, and (3) approaches that address underlying metabolic and regulatory mechanisms leading to acetate formation.

Because the first strategy is a genetic equivalent of bioprocess approaches, this research focuses on the two other strategies based on metabolic modelling and engineering. Three different strains were constructed: one in which the acetate production pathway is disabled, one in which more carbon is channeled to the citric acid cycle, and one that combines both

[1]This chapter was published in M. De Mey, G. J. Lequeux, J. J. Beauprez, J. Maertens, E. Van Horen, W. K. Soetaert, P. A. Vanrolleghem & E. J. Vandamme (2007b). Comparison of different strategies to reduce acetate formation in *Escherichia coli*. *Biotechnology Progress*, 23:1053–1063

approaches. In the first part of the study, the strains were grown in batch cultures. In the second part, steady state experiments were performed and metabolic flux analysis was applied to calculate the intracellular fluxes.

4.1 Materials and methods

4.1.1 Bacterial strain

Escherichia coli MG1655 [λ^-, F^-, *rph*–1, (*fnr*$^-$ 267)del] (labeled WT in the figures) was obtained from the Netherlands Culture Collection of Bacteria (NCCB, Utrecht, The Netherlands). *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB* [λ^-, F^-, *rph*–1, (*fnr*$^-$ 267)del, Δ*ackA–pta*, Δ*poxB*] (labeled 3KO in the figures), *E. coli* MG1655 Δp*ppc ppc–p37* [λ^-, F^-, *rph*–1, (*fnr*$^-$ 267)del, Δp*ppc ppc–p37*] (labeled PPC in the figures), and *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* [λ^-, F^-, *rph*–1, *rfb*–50, (*fnr*$^-$ 267)del, Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37*] (labeled 3KO–PPC in the figures) were constructed in the Laboratory of Genetics and Microbiology (MICR, VUB, Belgium), using the method described by Datsenko & Wanner (2000).

4.1.2 Culture conditions

Media

The Luria Broth (LB) medium consisted of 1 % triptone peptone (Difco, Erembodegem, Belgium), 0.5 % yeast extract (Difco) and 0.5 % sodium chloride (VWR, Leuven, Belgium).

The minimal medium consisted of 2 g/l NH_4Cl, 5 g/l $(NH_4)_2SO_4$, 2 g/l KH_2PO_4, 0.5 g/l NaCl, 0.5 g/l $MgSO_4 \cdot 7\,H_2O$, 16.33 g/l glucose $\cdot\, H_2O$, 1 ml/l vitamin solution and 100 µl/l molybdate solution. Vitamine solution consisted of 3.6 g/l $FeCl_2 \cdot 4\,H_2O$, 5 g/l $CaCl_2 \cdot 2\,H_2O$, 1.3 g/l $MnCl_2 \cdot 2\,H_2O$, 0.38 g/l $CuCl_2 \cdot 2\,H_2O$, 0.5 g/l $CoCl_2 \cdot 6\,H_2O$, 0.94 g/l $ZnCl_2$, 0.0311 g/l H_3BO_4, 0.4 g/l $Na_2EDTA \cdot 2\,H_2O$ and 1.01 g/l thiamine $\cdot$ HCl. The molybdate solution contained 0.967 g/l $Na_2MoO_4 \cdot 2\,H_2O$.

For flask culture medium, glucose and $MgSO_4$ was separately autoclaved (121 °C, 20 minutes) from the salts and combined after cooling down together with the filter-sterilised (pore size 0.2 µm, Millipore bottletopfilter, Millipore, Antwerpen, Belgium) vitamine and molybdate solutions. Before autoclaving, the pH of the salts solution was set at 7 with 1 M K_2HPO_4.

All components for the medium of the continuous culture were dissolved and filter-sterilised (pore size 0.22 µm, Sartobran, Sartorius, Belgium). The pH of the medium was left at approximately 5.4.

Cultivations

A preculture from a single colony on a LB-plate was started in 5 ml LB medium during 8 hours at 37 °C on an orbital shaker at 200 rpm. From this culture, 2 ml was transferred to 100 ml minimal medium in a 500 ml shake flask, and incubated for 16 hours at 37 °C on an orbital shaker at 200 rpm. Fifty milliliter of this culture was used to inoculate the reactor, a 1.2 l Biostat B culture vessel with 1 l working volume (Sartorius-BBI systems, Melsungen, Germany). The culture conditions were: 37 °C, stirring at 800 rpm, airflow rate of 1 l/min. The pH was maintained at 7 with 0.5 M H_2SO_4 and 4 M KOH. The exhaust gas was cooled down to 4 °C by an exhaust cooler (Frigomix 1000, Sartorius-BBI, Melsungen, Germany). A continuous feed of 4 ml/h 10 % solution of silicone antifoaming agent (BDH 331512K, VWR Int Ltd., Poole, England) was added to the culture vessel. The volume of the cultivation broth was kept constant with level control.

4.1.3 Measurements

Online measurements

Dissolved oxygen, pH, stirrer rate, temperature and airflow were continuously logged to the computer running MFCS/win 2.1 (Sartorius-BBI, Melsungen, Germany). Two balances under the influent and effluent barrel were coupled to the system to allow precise measurement of the dilution rate. Oxygen and carbon dioxide content of the off gas were measured by an URAS 10E off gas analyser (Hartmann & Braun, Germany), connected to the MFCS/win software.

Sampling of the reactor

The bioreactor contains in its interior a harvest pipe (HPLC tubing, Bio-Rad, Eke, Belgium) connected to a reactor port, linked outside to a Masterflex 14 tubing (Cole-Parmer, Antwerpen, Belgium) followed by a harvest port with a septum for sampling. The other side of this Masterflex 16 tubing is connected back to the reactor vessel. This system is referred to as the rapid sampling loop. During sampling, reactor broth is pumped around in the sampling loop. It has been estimated that, at a flow rate of 150 ml/min, the reactor broth needs 0.04 s to reach the harvest port and 3.2 s to re-enter the reactor. At a pO_2 level of 50 %, there is around 3 mg/l of oxygen in the liquid. The pO_2 level should never go below 20 %. Thus 1.8 mg/l of oxygen may be consumed during transit through the harvesting loop. Assuming an oxygen uptake rate of 0.4 g oxygent/g biomass/h (the maximal oxygen uptake rate found in the experiments), this gives for 5 g/l biomass, an oxygen uptake rate of 2 g/l/h or 0.56 mg/l/s, which multiplied by 3.2 s (residence time in the loop) gives 1.8 mg/l oxygen consumption.

In order to stop the metabolism of cells during the sampling, reactor broth was sucked through the harvest port in a syringe filled with 65 g stainless steel beads cooled down to −20 °C

(Mashego *et al.*, 2003), immediately followed by cold centrifugation (15000 g, 5 min, 4 °C).

In the batch experiments, a sample for OD_{600} and extracellular measurements was taken each hour using the rapid sampling loop and the cold stainless bead sampling method. When exponential growth was reached, the sampling frequency was increased to every 30 minutes.

Chemostat cultures were followed by measuring OD_{600}, extracellular metabolites and cell dry weight. Samples for MFA were taken after the cells attained steady state, which required at least five residence times, without significant perturbations. Six different samples were taken for extracellular metabolites measurements and five for cell dry weight determination.

Cell density measurements

Cell density of the culture was frequently monitored by measuring optical density at 600 nm (Uvikom 922 spectrophotometer, BRS, Brussel, Belgium). Cell dry weight was obtained by centrifugation (15 min, 5000 g, GSA rotor, Sorwall RC–5B, Du Pont Instruments, Goffin Meyvis, Kapellen, Belgium) of 20 g reactor broth in pre-dried and weighted falcons. The pellets were subsequently washed once with 20 ml physiological solution (9 g/l NaCl) and dried at 70 °C to a constant weight. To be able to convert OD measurements to biomass concentrations, a correlation curve of the OD to the biomass concentration was made.

HPLC analysis of glucose and organic acids

The concentrations of glucose and organic acids were determined on a Varian Prostar HPLC system (Varian, Sint-Katelijne-Waver, Belgium), using an Aminex HPX-87H column (Bio-Rad, Eke, Belgium) heated at 65 °C, equipped with a 1 cm reversed phase precolumn, using 5 mM H_2SO_4 (0.6 ml/min) as mobile phase. Detection was done by a dual-wave UV-VIS (210 nm and 265 nm) detector (Varian Prostar 325) and a differential refractive index detector (Merck LaChrom L-7490, Merck, Leuven, Belgium).

Phosphorous determination

The phosphate determination is based on the formation of a phosphomolybdate complex which absorbs at 820 nm (Gawronski & Benson, 2004; Ames & Dubin, 1960; Lowry & Lopez, 1946). First, the sample is deproteinised by adding 100 µl 0.6 N perchloric acid to 100 µl sample and diluted by adding 800 µl 0.1 N sodium acetate. 150 µl diluted and deproteinised sample was added to 350 µl acorbate/molybdate solution (one part 10 % ascorbate and 6 parts 0.42 % ammonium molybdate in 1 N H_2SO_4). The reaction mixture was incubated for 20 minutes at 60 °C. To stop the reaction, 500 µl stop solution (2 % sodiumcitrate tribasic dehydrate, 2 % acetic acid) was added to 500 µl reaction mixture. The absorbance at 820 nm was measured in a microplate reader (680 XR microplate reader, Bio-Rad, Eke, Belgium). For each batch of measurements, a calibration curve was made using a stock solution of 1 M KH_2PO_4. This

stock solution was diluted to a concentration of 0.002 M KH_2PO_4, and standard series from 0 M to 0.002 M were made for calibration.

Nitrogen determination

Quantification of nitrogen in the culture medium was performed using the kit LCK238 from Hach Lange Gmbh (Mechelen, Belgium) according to the manufacturers protocol.

4.1.4 Metabolic model

The theory of metabolic flux analysis as explained in chapter 2 was used. All fluxes were expressed in mol/l/h and model calculations were performed in these units.

The model described in section 3.2 was used with some minor modifications (different metabolites that are excreted) and corrections (in the biomass)(figure 4.1).

A constant biomass composition was used for every dilution rate: 55 % protein, 20.5 % RNA, 3.1 % DNA, 11.1 % lipids, 4.2 % LPS, 3.1 % peptidoglycane and 3.1 % glycogene (Pramanik & Keasling, 1997). This resulted in a biomass composition of $CH_{1.63}O_{0.392}N_{0.244}P_{0.021}S_{0.006}$, with molecular mass of 24.16 g/mol.

The model contains 137 reactions and 151 metabolites of which 11 were considered exchangeable with the environment: NH_3, PiOH, Biom, GLC, Lac, OAA, Suc, O_2, CO_2, H_2O and H_2SO_4. All parallel pathways were removed. There were no dead-end reactions, and the elemental consistency test was successfully passed (see section 2.4).

Eight metabolites were analysed for: GLC, NH_3, PiOH, Biom, O_2, CO_2, Lac, and Suc. The model contains 143 independent equations and 148 (137 internal fluxes and 11 exchangeable metabolites) unknown fluxes. Thus, at least five measurements should be performed to solve the model. Hence, three extra measurements were available to run the vector comparison test to detect and remove erroneous measurements (as described in section 2.5).

4.2 Results and discussion

Even under aerobic conditions, when glucose is excessively present, *E. coli* cells produce acetate as an extracellular byproduct. Acetate is unwanted because it slows down growth and it has a negative effect on recombinant protein production. Figure 4.2 depicts the central metabolism of *E. coli*. A first genetic approach to minimise acetate formation is to cut off the flow of carbon to acetate. To achieve this, an *ackA-ptA*, *poxB* knock-out strain (3KO) was created.

PGI: G6P ⟷ F6P
PFK: ATP + F6P ⟶ ADP + FBP
ALD: FBP ⟷ G3P + DHAP
TPI: DHAP ⟷ G3P
G3PDH: PiOH + NAD + G3P ⟷ NADH + H + BPG
PGK: ADP + BPG ⟷ ATP + 3PG
PGM: 3PG ⟷ 2PG
ENO: 2PG ⟷ H2O + PEP
PyrK: ADP + PEP ⟶ ATP + Pyr
PyrD: NAD + Pyr + CoA ⟶ NADH + H + AcCoA + CO2
CitSY: H2O + AcCoA + OAA ⟶ CoA + Cit
ACO: Cit ⟷ iCit
CitDH: NAD + iCit ⟷ NADH + H + CO2 + aKGA
AKGDH: NAD + CoA + aKGA ⟶ NADH + H + CO2 + SucCoA
SucCoASY: ADP + PiOH + SucCoA ⟷ ATP + CoA + Suc
SucDH: FAD + Suc ⟶ FADH2 + Fum
FumHY: H2O + Fum ⟷ Mal
MalDH: NAD + Mal ⟷ NADH + H + OAA
PEPCB: H2O + PEP + CO2 ⟶ PiOH + OAA
LacDH: NADH + H + Pyr ⟷ NAD + Lac
PFLY: Pyr + CoA ⟶ AcCoA + FA
AcKNLR: ADP + PiOH + AcCoA ⟷ ATP + CoA + Ac
Resp: 1.33 ADP + 1.33 PiOH + NADH + H + 0.5 O2 ⟶ 1.33 ATP + NAD + 2.33 H2O
H2CO3SY: H2O + CO2 ⟷ H2CO3
G6PDH: NADP + G6P ⟶ NADPH + H + 6PGL
LAS: H2O + 6PGL ⟶ 6PG
PGDH: NADP + 6PG ⟶ NADPH + H + CO2 + Rl5P
PPI: Rl5P ⟷ R5P
PPE: Rl5P ⟷ Xu5P
TK1: R5P + Xu5P ⟷ G3P + S7P
TA: G3P + S7P ⟷ F6P + E4P
TK2: Xu5P + E4P ⟷ F6P + G3P
PTS: GLC + PEP ⟶ G6P + Pyr
PPiOHHY: PPiOH + H2O ⟶ 2 PiOH
GluDH: NADPH + H + aKGA + NH3 ⟷ NADP + H2O + Glu
GluLI: ATP + NH3 + Glu ⟶ ADP + PiOH + Gln
AspSY: ATP + H2O + Asp + Gln ⟶ AMP + PPiOH + Asn + Glu
AspTA: OAA + Glu ⟷ aKGA + Asp
AlaTA: Pyr + Glu ⟷ aKGA + Ala
ValAT: aKIV + Glu ⟷ aKGA + Val
LeuSYLR: NAD + H2O + AcCoA + aKIV + Glu ⟶ NADH + H + CoA + CO2 + aKGA + Leu
aKIVSYLR: NADPH + H + 2 Pyr ⟶ NADP + H2O + CO2 + aKIV
IleSYLR: NADPH + H + Pyr + Glu + Thr ⟶ NADP + H2O + CO2 + aKGA + NH3 + Ile
ProSYLR: ATP + 2 NADPH + 2 H + Glu ⟶ ADP + PiOH + 2 NADP + H2O + Pro
SerLR: NAD + H2O + 3PG + Glu ⟶ PiOH + NADH + H + aKGA + Ser
SerTHM: Ser + THF ⟶ H2O + Gly + MeTHF
H2SSYLR: 2 ATP + 3 NADPH + ThioredH2 + 3 H + H2SO4 ⟶ ADP + PPiOH + 3 NADP + Thiored + 3 H2O + H2S + PAP
PAPNAS: H2O + PAP ⟶ AMP + PiOH
CysSYLR: H2S + AcCoA + Ser ⟶ CoA + Cys + Ac
PrppSY: ATP + R5P ⟶ AMP + PRPP
HisSYLR: ATP + 2 NAD + 3 H2O + Gln + PRPP ⟶ 2 PPiOH + PiOH + 2 NADH + 2 H + aKGA + His + AICAR
PheSYLR: Glu + Chor ⟶ H2O + CO2 + aKGA + Phe
TyrSYLR: NAD + Glu + Chor ⟶ NADH + H + CO2 + aKGA + Tyr
TrpSYLR: Gln + Ser + Chor + PRPP ⟶ PPiOH + 2 H2O + G3P + Pyr + CO2 + Glu + Trp
DhDoPHepAD: H2O + PEP + E4P ⟶ PiOH + Dahp
DhqSY: Dahp ⟶ PiOH + Dhq
DhsSYLR: Dhq ⟷ H2O + Dhs
ShiSY: NADPH + H + Dhs ⟷ NADP + Shi
ShiKN: ATP + Shi ⟶ ADP + Shi3P
ChorSYLR: PEP + Shi3P ⟶ 2 PiOH + Chor
ThrSYLR: ATP + H2O + HSer ⟶ ADP + PiOH + Thr
MDAPSYLR: NADPH + H + Pyr + SucCoA + Glu + AspSA ⟶ NADP + CoA + aKGA + Suc + MDAP
LysSY: MDAP ⟶ CO2 + Lys
MetSYLR: H2O + SucCoA + Cys + MTHF + HSer ⟶ Pyr + CoA + Suc + NH3 + Met + THF
AspSASY: ATP + NADPH + H + Asp ⟶ ADP + PiOH + NADP + AspSA
HSerDH: NADPH + H + AspSA ⟷ NADP + HSer
CarPSY: 2 ATP + H2O + H2CO3 + Gln ⟶ 2 ADP + PiOH + Glu + CarP
OrnSYLR: ATP + NADPH + H + H2O + AcCoA + 2 Glu ⟶ ADP + PiOH + NADP + CoA + aKGA + Orn + Ac
ArgSYLR: ATP + Asp + Orn + CarP ⟶ AMP + PPiOH + PiOH + Fum + Arg
ThioredRD: NADPH + Thiored + H ⟷ NADP + ThioredH2
H2O2ox: 2 H2O2 ⟶ 2 H2O + O2
FAD2NAD: NAD + FADH2 ⟷ NADH + FAD + H
AICARSYLR: 6 ATP + 3 H2O + CO2 + Asp + 2 Gln + Gly + FA + PRPP ⟶ 6 ADP + PPiOH + 6 PiOH + Fum + 2 Glu + AICAR
IMPSYLR: FTHF + AICAR ⟶ H2O + THF + IMP
AMPSYLR: Asp + GTP + IMP ⟶ AMP + PiOH + Fum + GDP
AdKN: ATP + AMP ⟷ 2 ADP
ADPRD: ADP + ThioredH2 ⟶ Thiored + H2O + dADP
dADPKN: ATP + dADP ⟶ ADP + dATP
IMPDH: NAD + H2O + IMP ⟶ NADH + H + XMP
GMPSY: ATP + H2O + Gln + XMP ⟶ AMP + PPiOH + Glu + GMP
GuKN: ATP + GMP ⟶ ADP + GDP
GDPKN: ATP + GDP ⟶ ADP + GTP
GDPRD: ThioredH2 + GDP ⟶ Thiored + H2O + dGDP
dGDPKN: ATP + dGDP ⟶ ADP + dGTP
UMPSYLR: O2 + Asp + PRPP + CarP ⟶ PPiOH + PiOH + H2O + CO2 + UMP + H2O2
UrKN: ATP + UMP ⟶ ADP + UDP
UDPKN: ATP + UDP ⟶ ADP + UTP
CTPSY: ATP + H2O + Gln + UTP ⟶ ADP + PiOH + Glu + CTP
CDPKN: ATP + CDP ⟷ ADP + CTP
CMPKN: ATP + CMP ⟶ ADP + CDP
CDPRD: ThioredH2 + CDP ⟶ Thiored + H2O + dCDP
dCDPKN: ATP + dCDP ⟶ ADP + dCTP
UDPRD: ThioredH2 + UDP ⟶ Thiored + H2O + dUDP
dUDPKN: ATP + dUDP ⟶ ADP + dUTP
dUTPPPAS: H2O + dUTP ⟶ PPiOH + dUMP
dTMPSY: MeTHF + dUMP ⟶ DHF + dTMP
dTMPKN: ATP + dTMP ⟶ ADP + dTDP
dTDPKN: ATP + dTDP ⟶ ADP + dTTP
DHFRD: NADPH + H + DHF ⟶ NADP + THF
FTHFSYLR: NADP + H2O + MeTHF ⟶ NADPH + H + FTHF
GlyCA: NAD + Gly + THF ⟷ NADH + H + CO2 + NH3 + MeTHF
MeTHFRD: NADH + H + MeTHF ⟶ NAD + MTHF
AcCoACB: ATP + H2O + AcCoA + CO2 ⟷ ADP + PiOH + MalCoA
MalCoATA: MalCoA + ACP ⟷ CoA + MalACP
AcACPSY: MalACP ⟶ CO2 + AcACP
C120SY: 10 NADPH + 10 H + AcACP + 5 MalACP ⟶ 10 NADP + 5 H2O + 5 CO2 + C120ACP + 5 ACP
C140SY: 12 NADPH + 12 H + AcACP + 6 MalACP ⟶ 12 NADP + 6 H2O + 6 CO2 + C140ACP + 6 ACP
C160SY: 14 NADPH + 14 H + AcACP + 7 MalACP ⟶ 14 NADP + 7 H2O + 7 CO2 + C160ACP + 7 ACP
C181SY: 15 NADPH + 15 H + AcACP + 8 MalACP ⟶ 15 NADP + 8 H2O + 8 CO2 + C181ACP + 8 ACP
AcylTF: C160ACP + C181ACP + Go3P ⟶ 2 ACP + PA
Go3PDH: NADPH + H + DHAP ⟷ NADP + Go3P
DGoKN: ATP + DGo ⟶ ADP + PA
CDPDGoSY: CTP + PA ⟷ PPiOH + CDPDGo
PSerSY: Ser + CDPDGo ⟶ CMP + PSer
PSerDC: PSer ⟶ CO2 + PEthAn
GlnF6PTA: F6P + Gln ⟶ Glu + GA6P
GlcAnMU: GA6P ⟷ GA1P
NAGUrTF: AcCoA + UTP + GA1P ⟶ PPiOH + CoA + UDPNAG
LipaSYLR: ATP + 2 CMPKDO + 2 UDPNAG + C120ACP + 5 C140ACP ⟶ ADP + 2 CMP + UMP + UDP + 6 ACP + Lipa + 2 Ac
A5PIR: Rl5P ⟷ Ar5P
PGLCMT: G6P ⟷ G1P
CMPKDOSYLR: 2 H2O + PEP + Ar5P + CTP ⟶ PPiOH + 2 PiOH + CMPKDO
ADPHEPSY: ATP + S7P ⟶ PPiOH + ADPHEP
UDPGlcSY: G1P + UTP ⟶ PPiOH + UDPGlc
EthANPT: CMP + PEthAn ⟷ CDPEthAn + DGo
LpsSYLR: 3 ADPHEP + 3 CMPKDO + 2 UDPGlc + Lipa + 2 CDPEthAn ⟶ 3 ADP + 3 CMP + 2 CDP + 2 UDP + Lps
PGSYLR: Go3P + CDPDGo ⟶ PiOH + CMP + PG
CLSY: PG + CDPDGo ⟶ CMP + CL
PeptidoSYLR: 5 ATP + NADPH + H + PEP + 3 Ala + MDAP + 2 UDPNAG ⟶ 5 ADP + 7 PiOH + NADP + UMP + UDP + Peptido
GlcgSY: ATP + G1P ⟶ ADP + PPiOH + Glcg
ATPHY: ATP + H2O ⟶ ADP + PiOH
DNASYLR: 2 H2O + 0.246 dATP + 0.254 dGTP + 0.254 dCTP + 0.246 dTTP ⟶ 2 PiOH + DNA
RNASYLR: 0.262 ATP + 2 H2O + 0.322 GTP + 0.2 CTP + 0.216 UTP ⟶ 2 PiOH + RNA
ProtSYLR: 2 ATP + 3 H2O + 0.0961 Ala + 0.05506 Arg + 0.04505 Asn + 0.04505 Asp + 0.01702 Cys + 0.04905 Gln + 0.04905 Glu + 0.1151 Gly + 0.01802 His + 0.05405 Ile + 0.08408 Leu + 0.06406 Lys + 0.02903 Met + 0.03504 Phe + 0.04104 Pro + 0.04004 Ser + 0.04705 Thr + 0.01101 Trp + 0.02603 Tyr + 0.07908 Val + 2 GTP ⟶ 2 ADP + 4 PiOH + 2 GDP + Prot
LipidSYLR: 0.0266 CL + 0.202 PG + 0.7714 PEthAn ⟶ Lipid
BiomSYLR: 0.004561 Glcg + 0.0002663 Lps + 0.0008933 Peptido + 0.002291 DNA + 0.01446 RNA + 0.1227 Prot + 0.003642 Lipid ⟶ Biom

Figure 4.1: List of reactions used in the model. A description of the metabolites can be found in appendix B.1.

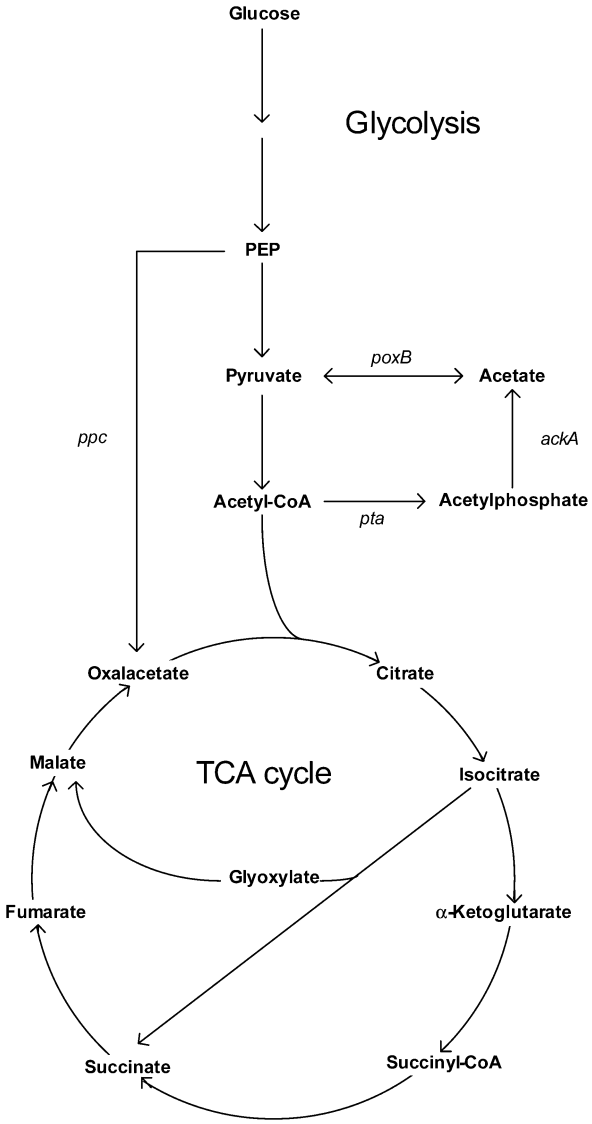


Figure 4.2: The central metabolism of *E. coli*. PEP, phosphoenolpyruvate; *ackA*, acetate kinase; *poxB*, pyruvate oxidase; *ppc*, phosphoenolpyruvate carboxylase; *pta*, acetylphosphotransferase.

A second strategy to reduce acetate formation, involves addressing the underlying biochemical mechanisms that lead to acetate. Recent results have potentially elucidated the metabolic and regulatory causes of acetate formation and the links between acetate formation and recombinant protein production. Of the amino acids that combine to form the protein product, ten are biochemically derived from TCA cycle metabolites, eleven from glycolysis metabolites and 4 are derived from the pentose phosphate pathway. An important fact is that when *E. coli* is grown in a medium with glucose as the sole carbon source, nearly all the carbon used to synthesise the ten amino acids derived from TCA metabolites must flow through the

anaplerotic pathway mediated by the enzyme phosphoenolpyruvate carboxylase (Eiteman & Altman, 2006). In addition, glycolysis and TCA cycle both generate the reduced cofactor $NADH + H^+$. In an aerobic environment, $NADH + H^+$ can be reoxidised to NAD^+, which is necessary to drive the glycolysis and the TCA cycle. The rate at which $NADH + H^+$ is formed proportionally increases with the rate of glucose consumption. However, the specific oxygen uptake rate will reach a plateau when the maximum rate at which *E. coli* can consume oxygen is achieved, despite its availability. At this point, cellular respiration is not able to regenerate sufficient NAD^+, resulting in the accumulation of $NADH + H^+$, which plays an important role in acetate formation. First, $NADH + H^+$ is a strong allosteric inhibitor of the enzyme citrate synthase, which mediates the first step of the TCA cycle. Thus, accumulation of $NADH + H^+$ will inhibit the carbon flow to the TCA cycle and consequently decrease the formation of the ten TCA cycle derived amino acids. Secondly, the redox ratio $NADH + H^+/NAD^+$ is involved in the regulation of the ArcA regulatory system in *E. coli*, which represses the expression of several genes in the TCA cycle. *E. coli* will respond to increased $NADH + H^+$ levels by redirecting the carbon flow to acetate production that generates less $NADH + H^+$ (Eiteman & Altman, 2006).

A strategy to divert the carbon flow from acetate to the TCA cycle and to produce less $NADH + H^+$ is to overexpress PEP carboxylase encoded by the gene *ppc*. Previous research already indicated that overexpressing ppc reduces acetate formation (De Maeseneire *et al.*, 2006; Farmer & Liao, 1997). Therefore, a *ppc* overexpressing mutant of *E. coli* was created in which the natural *ppc* promoter was replaced with a strong constitutive artificial promoter (PPC) (De Mey *et al.*, 2007c). Finally, the two approaches were combined by replacing the natural *ppc* promoter with p37 in *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB* (3KO-PPC).

4.2.1 Batch experiments

Different *E. coli* strains were studied, representing different approaches to minimise acetate formation; initially, batch cultures were performed on glucose-limiting minimal medium under aerobic conditions with the following strains: *E. coli* MG1655 (WT), *E. coli* MG1655 Δp*ppc ppc–p37* (PPC), *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB* (3KO), and *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* (3KO-PPC).

During the culture, the growth progress (OD_{600}) and extracellular metabolite concentrations were determined. The results are depicted in figure 4.3.

From these data, the growth rate (μ) and the relevant metabolite yields (Y metabolite) were determined for the different strains. The growth rate was calculated by determining the slope of the curve, obtained by plotting $\ln(OD_{600})$ in function of the time using the software Sigma-plot 10.0. The metabolite yields were calculated by determining the slope of the curves, obtained by plotting the metabolite concentration in function of the glucose consumption, also

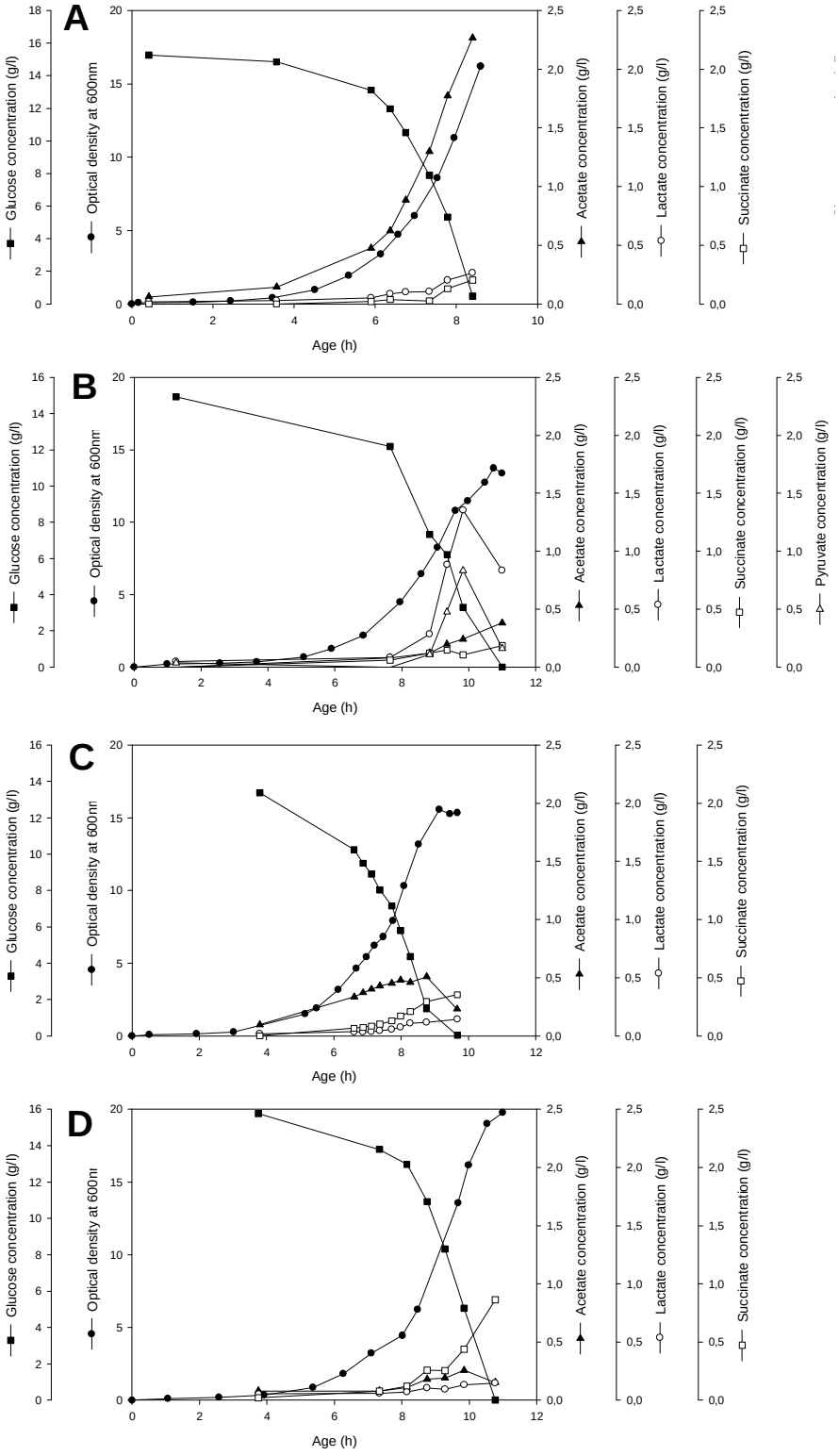


Figure 4.3: Evolution of biomass and extracellular metabolites during an aerobic batch culture of (A) WT, (B) 3KO, (C) PPC and (D) 3KO-PPC.

by use of Sigma-plot 10.0. The results are given in figures 4.4 and 4.5.

Cutting off the carbon flow towards acetate results in a decreased growth rate to 80 %, as compared to the wild-type. This is caused by the decreased availability of energy as the acetate pathway generates one ATP extra and consumes one more $NADH + H^+$ compared to the lactate pathway (which is used when the acetate pathway is knocked out). The same effect is observed in the *ppc* overexpressing strain. By diverting the carbon flow from phophoenolpyruvate (PEP) to oxaloacetate instead of pyruvate, less PEP is available for the PTS to transport glucose into the cell. In addition, combination of cutting off the carbon flow towards acetate and diverting the carbon flow from PEP toward oxaloacetate, has no additional effect.

However, a decrease of 20 % in maximal growth rate (as compared to the wild-type) still corresponds to a growth rate of $1.82\,h^{-1}$ and a generation time of 0.38 h, which indicates that these mutant strains remain fast growing and thus applicable for large industrial processes.

Figure 4.5 illustrates that the acetate yield in the wild-type is quite high. Cutting off the carbon flow towards acetate reduces the acetate yield significantly (7.6 times). However, the lactate yield increases substantially (13.5 times). This indicates that the genetic approach of interrupting the carbon flow toward acetate results in minimising acetate. However, other byproducts are formed, making this strategy unfavorable. This confirms data found in literature (Contiero *et al.*, 2000; El-Mansi & Holms, 1989; Yang *et al.*, 1999; Dittrich *et al.*, 2005). All of them report a strong reduction in acetate production when *ackA* and *pta* are eliminated. This is at the expense of the growth rate and is accompanied by an increase in the production of other fermentation products such as lactate and formate.

The genetic approach of redirecting the carbon flow from PEP towards oxaloacetate seems more promising. The acetate yield decreases 11.5 times. The succinate yield on the other hand increases significantly (1.8 times), whereas the lactate yield decreases 1.6 times, indicating an increased flux towards the TCA cycle and thus towards the amino acid precursors. In the literature, overexpression of *ppc* from a plasmid is described (Farmer & Liao, 1997) and a reduced acetate production is obtained. In this contribution, in contrast to what is described in Farmer & Liao (1997), *ppc* is constitutively overexpressed by replacing the natural promoter in the *E. coli* genome. Therefore, this approach introduces minimal changes in the host, is easily applicable to industrial strains, and there is no need for a selection marker (*e.g.* antibiotic) and an inducer (*e.g.* IPTG) to maintain the plasmid and to express the gene, respectively.

Combination of the two genetic approaches results in similar culture behaviour as replacing the natural *ppc* promoter by a strong promoter, except for the succinate yield that increases 2.5 times, as compared to the wild-type.

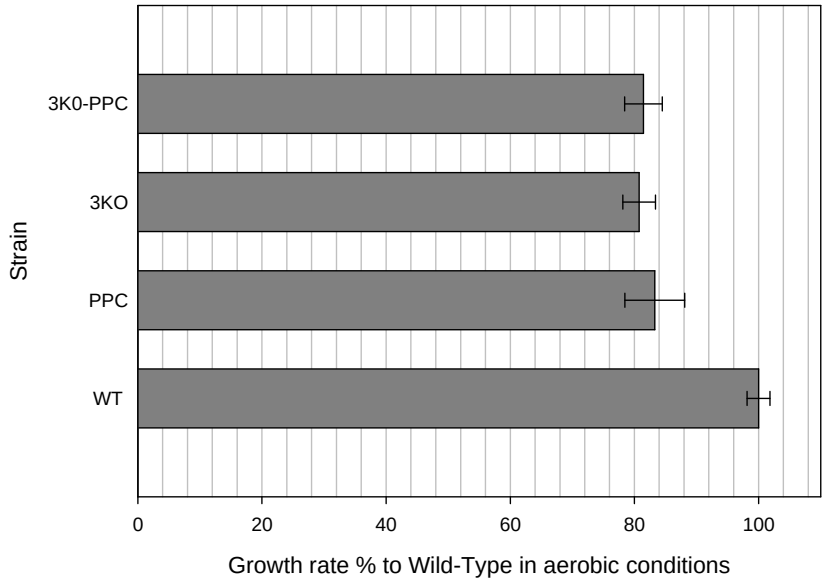


Figure 4.4: Growth rate (% to *E. coli* MG1655) in aerobic conditions on glucose-limiting minimal medium for different strains. Error bars represent the standard deviation.

Considering the results of the batch experiments, it was decided to perform chemostat experiments to elucidate the influence of these genetic manipulations on the central metabolism under C-limitation in a chemostat.

4.2.2 Chemostat experiments and MFA

To collect data for the metabolic steady state model of *E. coli* MG1655, chemostat experiments were performed with different mutant strains and at minimally three different dilution rates per strain.

For each experiment, it was verified whether the culture was C-limited by measuring the residual glucose concentration in the supernatant.

Metabolic flux analysis was performed for further characterisation of the different *E. coli* strains.

Data quality assessment

To assess the quality of the data, the carbon balance of the different chemostat experiments for the different strains was evaluated (figure 4.6). The carbon balance is calculated from the measured fluxes. Figure 4.6 shows that five experiments (3KO-PPC at D = $0.113\,h^{-1}$, WT at D = $0.143\,h^{-1}$, 3KO-PPC at D = $0.145\,h^{-1}$, PPC at D = $0.163\,h^{-1}$ and 3KO-PPC at D = $0.35\,h^{-1}$) have a carbon balance that significantly does not close. However, all the

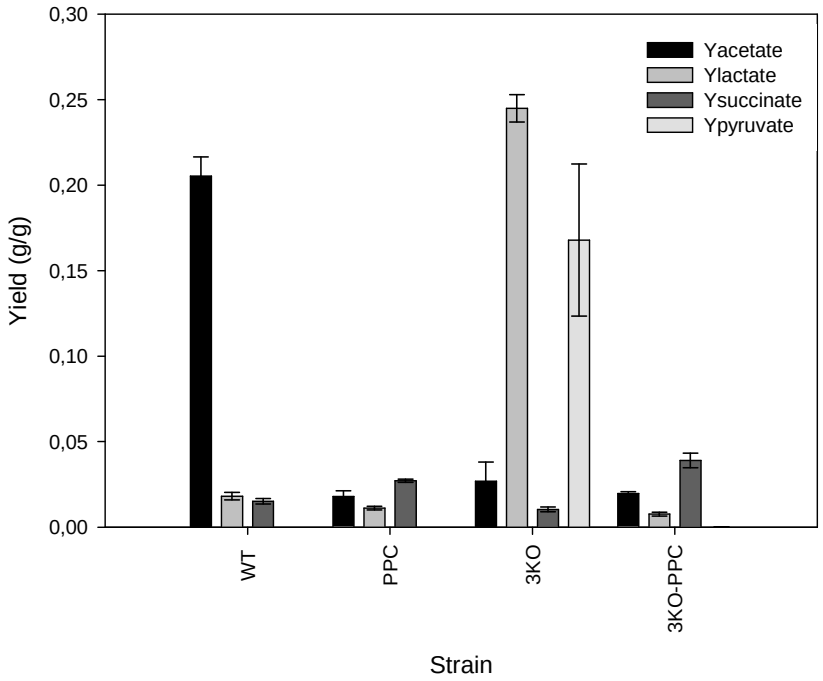


Figure 4.5: Metabolite yields (g/g) of acetate, lactate, succinate, and pyruvate for different *E. coli* strains during a batch experiment in aerobic conditions on glucose-limiting minimal medium. Error bars represent the standard deviation.

experiments but 3KO-PPC at D = $0.143\,h^{-1}$ were kept in the data set, as the calculated fluxes were in accordance with those at other dilution rates. Experiment 3KO-PPC at D = $0.145\,h^{-1}$ was not used in the subsequent calculations because the error on the carbon balance is much larger and the statistical test performed on the model residuals rejected this experiment.

The ratio between the carbon dioxide excretion rate and the oxygen uptake rate is called respiratory quotient (RQ). From figure 4.7 it can be seen that the RQs do not significantly differ from one.

For each chemostat, the metabolic model was solved and the statistical test to assess the quality of fit of the model to the data was run. The H_0 hypothesis of this test is that the residuals (defined as the measured fluxes minus the model predicted fluxes) are zero (see section 2.5.3 for a detailed discussion of this test). When this test was not passed successfully, the vector comparison test was used to find (and ultimately remove) the wrong measurement (see section 2.5.4). For four models, the residuals were significantly different from zero: experiment WT at D = $0.227\,h^{-1}$, 3KO at D = $0.093\,h^{-1}$, PPC at D = $0.257\,h^{-1}$ and 3KO-PPC at D = $0.145\,h^{-1}$. Subsequently, erroneous measurements were removed.

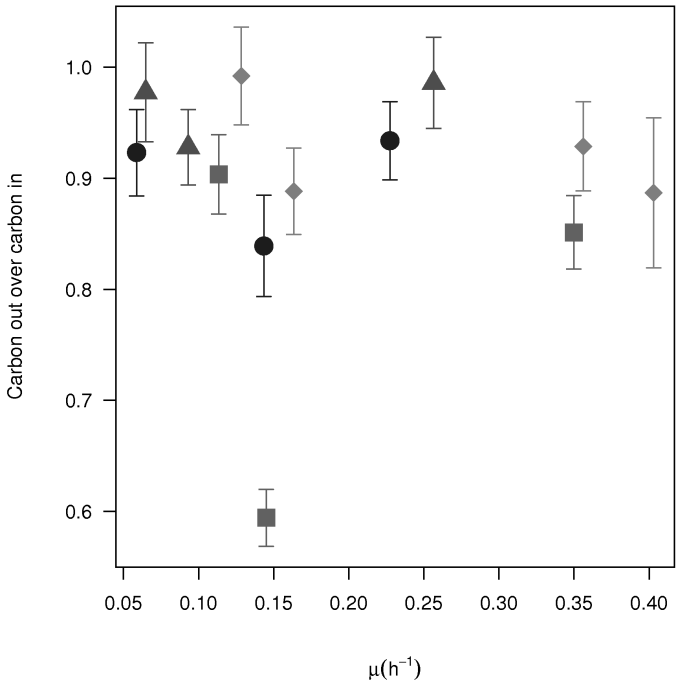


Figure 4.6: Ratio of carbon out over carbon in. Error bars represent the standard deviation. WT = dot; 3KO = triangle; PPC = diamond; 3KO-PPC = square.

For experiment WT at D = $0.227\,h^{-1}$, the NH_3 measurement was removed as suggested by the vector comparison test.

In the case of experiment 3KO at D = $0.093\,h^{-1}$, the vector comparison model was fully accepted when the O_2 measurement was removed.

In the case of experiment PPC at D = $0.257\,h^{-1}$, it was not clear from the vector comparison test which measurements were causing the misfitting of the model. Most suggested vectors contained PiOH and when removing this measurement, the model was fully accepted.

For experiment 3KO-PPC at D = $0.145\,h^{-1}$, the model was not accepted and the vector comparison test suggested to remove the NH_3 and the biomass measurements. However, after removing those measurements, the P-value of the χ-statistical test for acceptance of the model was still only 0.06. Because of this low P-value and because of the ratio carbon out over carbon in was equal to 0.6, it was decided not to take these measurements into account for further analysis.

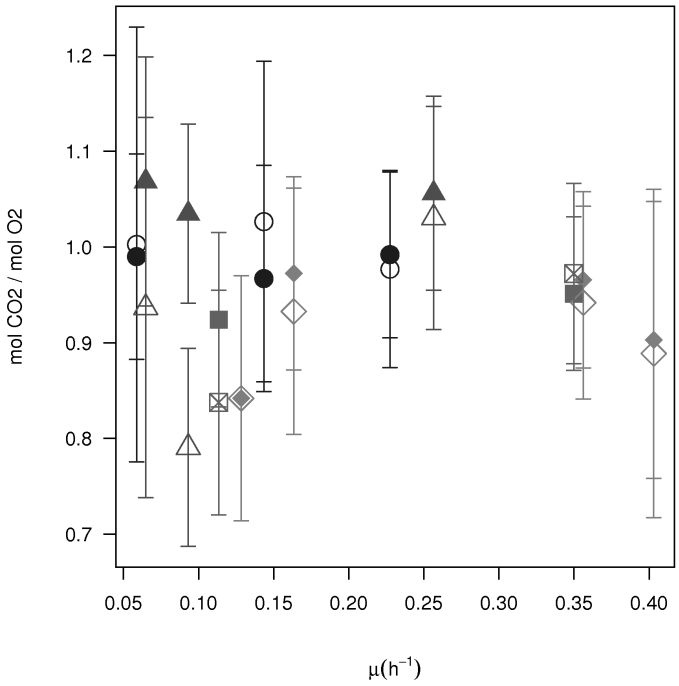


Figure 4.7: Respiration coefficients. Open symbols stand for the measured values; solid symbols stand for the corrected values by the model. Error bars represent the standard deviation. WT = dot; 3KO = triangle; PPC = diamond; 3KO-PPC = square.

MFA results

Figure 4.8 depicts the fluxes in the glycolysis, the pentose-phosphate pathway, the TCA cycle and the fermentation pathways. The perfect linear relationship for the different dilution rates in the pentose-phosphate pathway originates from the fact that this pathway is, under the assumptions of this model, solely used for generating biomass precursors.

Although some slight changes in some fluxes appear, figure 4.8 shows that in general no significant difference in the flux map for the different strains can be observed, indicating no negative influence of the genetic modifications on primary metabolism at the dilution rates studied.

The biomass yield on glucose is around 0.45 C-mol biomass per C-mol glucose. MFA can quantitatively show where in the central carbon metabolism the carbon is taken out for biomass precursors. At the G6P node, more than half of the carbon flows toward the pentose-phosphate pathway and not to the glycolysis. However, most of this flows back into the glycolysis as F6P and finally only 15 % (taking into account the CO_2 production in the beginning of the pentose phosphate pathway) of the glucose is channeled via the pentose-phosphate cycle to biomass.

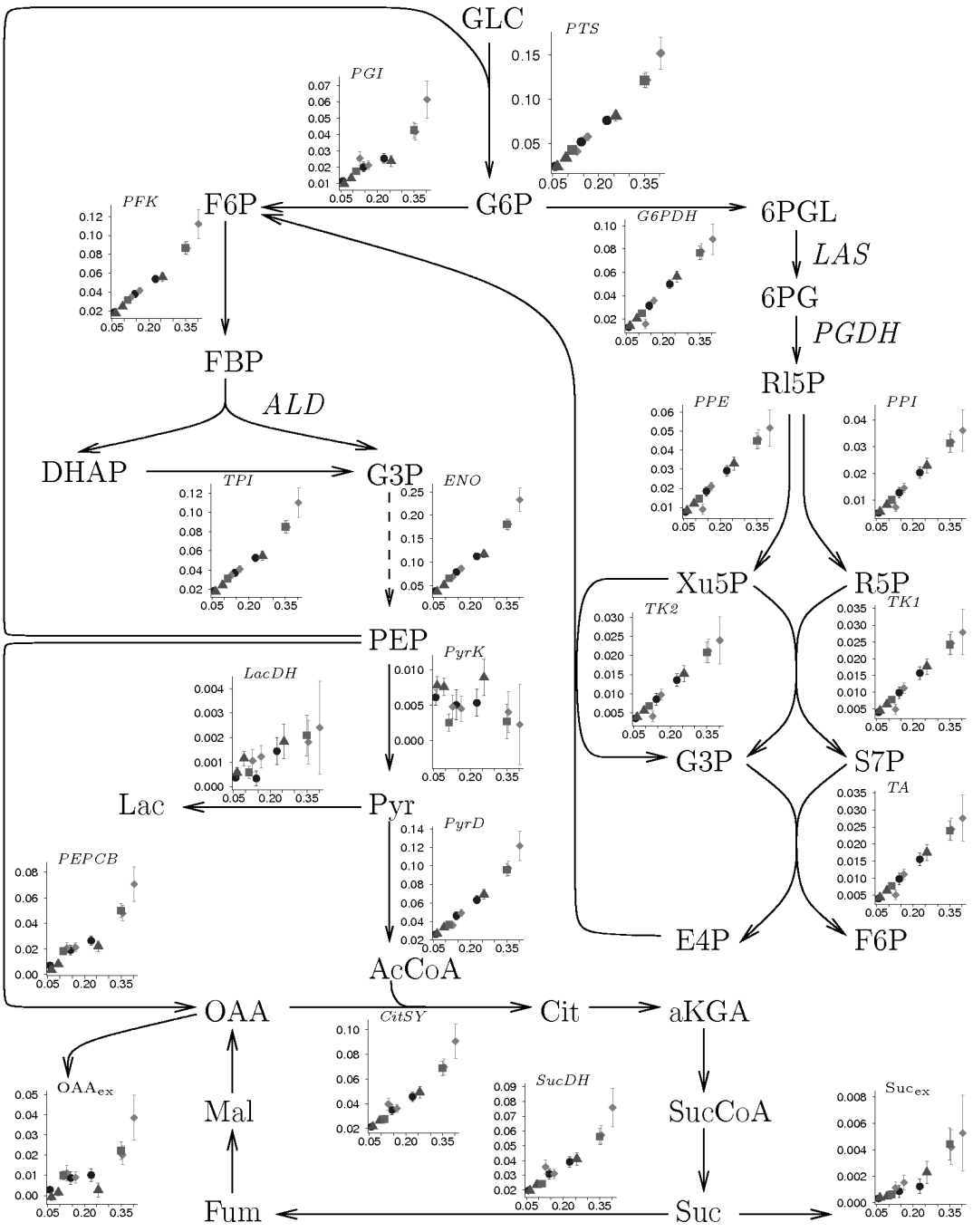


Figure 4.8: Relative fluxes to biomass (mol/mol biomass/h) in the glycolysis, the pentose-phosphate pathway, the TCA cycle, and the fermentation pathway. Abscissa: growth rate, ordinate: relative fluxes. Error bars represent the standard deviation. WT = dot; 3KO = triangle; PPC = diamond; 3KO-PPC = square.

The flux of PyrK, converting PEP to pyruvate, is essentially zero for all dilution rates. This is in agreement with the results found in Carlson & Srienc (2004b). They generated potentially efficient pathways for biomass production using elementary mode analysis. All those pathways have a very low or zero PyrK flux, which means that all pyruvate originates from the PTS reaction.

PyrD only consumes 75 % of the available pyruvate, the remaining amount is used for biomass precursor synthesis. Thus, around 13 % of the glucose carbon is converted to biomass via pyruvate.

For the fatty acid formation, AcCoA is used. This accounts for around 5 % of the glucose carbon going to biomass. Finally, around 10 % glucose carbon is going to biomass precursors via citric acid cycle metabolites (replenished via the PEPCB reaction).

Maintenance

The calculated ATP hydrolysis flux allows to investigate the maintenance requirements for the strains. Figure 4.9 shows the energy requirements for different dilution rates. As with the other fluxes, the modified strains have really no different maintenance requirements, indicating that there is no negative influence of the genetic modifications, on the energy metabolism at the dilution rates studied.

The growth-associated maintenance is the slope of the line shown in figure 4.9, while the non-growth-associated maintenance is found by taking the intercept of this line. For non-growth-associated maintenance, a value of 0.088 mol/mol BM/h (BM = BioMass) was obtained. This conforms with the values reported in the literature for the wild-type: 0.20 mol/mol BM/h (Varma & Palsson, 1994a) 0.073 mol/mol BM/h (Kayser *et al.*, 2005), and 0.12 mol/mol BM/h (Carlson & Srienc, 2004b). The growth-associated maintenance is 1.70 mol/mol BM. In literature, both low values, 0.34 mol/mol BM (Varma & Palsson, 1994a), and high values, 2.6 mol/mol BM (Carlson & Srienc, 2004b), can be found.

4.3 Conclusions

Three different approaches to minimise acetate formation were compared: (1) one approach that directly reduces carbon flow to acetate (3KO); (2) one approach that addresses underlying metabolic and regulatory mechanisms that lead to acetate formation (PPC); and (3) one approach that combines the two strategies above (3KO-PPC).

Batch cultures under aerobic conditions were performed to compare these mutants to the wild-type. All approaches resulted in a decrease of about 20 % of the growth rate. However,

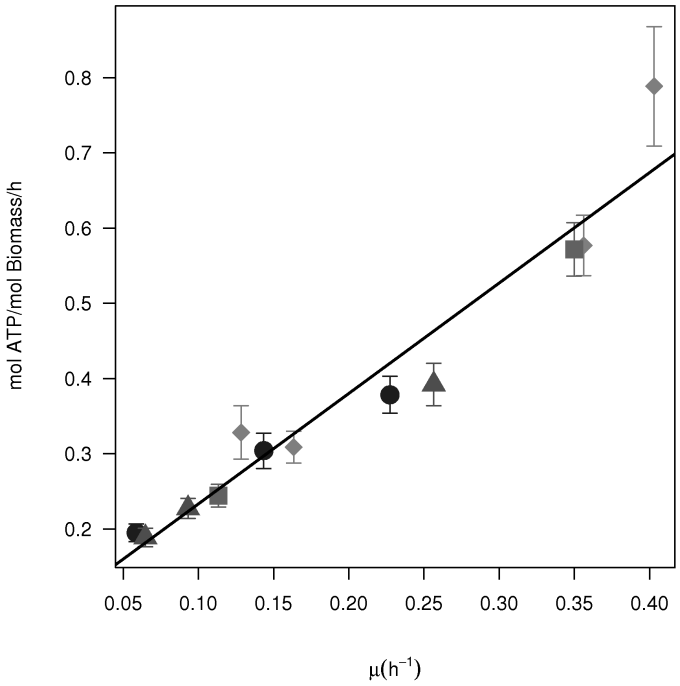


Figure 4.9: Moles of ATP per mol of biomass per hour that are hydrolysed for the different strains at different growth rates. Error bars represent the standard deviation. WT = dot; 3KO = triangle; PPC = diamond; 3KO-PPC = square.

these mutant strains remain fast growing and, thus, are suitable for application in large scale bioreactor processes.

Cutting off the carbon flow towards acetate (3KO) reduces the acetate yield significantly. On the other hand, the lactate yield increases substantially. The genetic approach of interrupting the carbon flow towards acetate results in minimising acetate, but other byproducts are formed, making this strategy unfavorable. This confirms data found in the literature (Contiero *et al.*, 2000; El-Mansi & Holms, 1989; Yang *et al.*, 1999; Dittrich *et al.*, 2005). The genetic approach of redirecting the carbon flow from PEP towards oxaloacetate (PPC) is more promising. The acetate and lactate yields decrease, whereas the succinate yield increases significantly. A combination of the two genetic approaches results in a similar behaviour as replacing the natural *ppc* promoter by a strong promoter, except for the succinate yield that increases more.

For further characterisation of the strains, chemostat experiments were conducted. Subsequently, the data of the chemostat experiments were used to perform metabolic flux analysis. The flux maps for the different strains showed no significant difference, indicating no negative influence of the genetic modifications on primary metabolism at the dilution rates studied.

Also for the cellular energy metabolism (*i.e.* maintenance), no significant differences could be observed.

Taking into account the results of these batch cultures and MFA, the strain PPC can be used as a possible host for recombinant protein production. The genetic alterations in this strain show no unfavorable impact on primary metabolism and cellular energy metabolism. In addition, during batch cultures more succinate production is observed, indicating an increased flux toward the TCA cycle and thus towards the amino acid precursors.

Part II

Dynamic metabolic flux analysis

Chapter 5

Dynamic metabolic flux analysis: what happens inside the cell when changing the limiting substrate from C to N and *vice versa*?

Metabolic flux analysis (MFA) permits to calculate intracellular fluxes by measuring only exchange rates with the environment. The main requirement for this methodology is that the cell is in a pseudo steady state, thus that there is no accumulation or depletion of intracellular metabolites.

Flux balance analysis (FBA) uses linear optimisation for solving the metabolic model and does not directly use measurements. The same requirements as for MFA are applicable to FBA: the reactor should be in steady state. However, some examples of FBA applied in not stationary cases can be found. In Varma & Palsson (1994b), flux balance analysis is used for describing fed batches by iteratively solving the model for maximal biomass production. At each time point, the available substrate is calculated from the results of the FBA in the previous step. The time profiles of cell density, glucose and by-products were quantitatively predicted. This concept of dynamic FBA (DFBA) is further developed by Mahadevan *et al.* (2002). They formalise the methodology of Varma & Palsson (1994b) and name it static optimisation-based DFBA. They also introduce a new methodology, called dynamic optimisation-based DFBA, in which the optimisation is done over the entire time period of interest to obtain time profiles of fluxes. The methodology of dynamic optimisation-based DFBA is combined with the concept of minimisation of metabolic adjustment (MOMA) to model myocardial energy metabolism (Luo *et al.*, 2006). Minimisation of metabolic adjustment (MOMA), is an alternative goal function to better fit FBA predictions to mutated strains (Segrè *et al.*, 2002). These results show that the cellular metabolism does not always remain optimal during

transient perturbations. In the comparative study of Schuetz *et al.* (2007) of different goal functions, is was shown that only ATP or biomass maximisation has the best predicting power compared to 13 C analysis. However, they only calculated one point for the whole batch and did not use the methodology for batch modelling of Luo *et al.* (2006).

Traditionally many chemostats are performed with one limiting substrate (mostly carbon-limitation). The use of two limiting substrates has been studied less extensively (Lee *et al.*, 1984; Egli, 1991; Zinn *et al.*, 2004). Lee *et al.* (1984) experimentally showed that one can change the limiting substrate (ammonium or carbon) by changing the dilution rate. This was generalised by Egli (1991) who showed that the zone of dual substrate limitation is dependent on the growth rate.

Outline of the work

The first part of this chapter extends MFA so that it can be applied to systems that are not in steady state. The mathematical methodology needed to apply dynamic MFA is explained.

In the second part of this chapter, dynamic MFA is applied on experiments where steady state cultures were switched from one limitation to another: from glucose limitation to ammonia limitation and *vice versa*. The nitrogen-limiting medium contained 30 g/l glucose and 0.65 g/l ammonia, the carbon-limiting medium 15 g/l glucose and 1.3 g/l ammonia. The media were designed as such that for both conditions the same biomass concentration was achieved in the reactor broth. To assess the intracellular fluxes, dynamic metabolic flux analysis is applied. At each time point, the exchange fluxes of metabolites are determined based on their concentration in the reactor broth. Those fluxes are used to solve an overdetermined metabolic model, resulting in the determination of the intracellular flux values.

5.1 Materials and methods

5.1.1 Bacterial strain

Escherichia coli MG1655 [λ^-, F^-, *rph*–1, (*fnr*$^-$ 267)del] was obtained from the Netherlands Culture Collection of Bacteria (NCCB, Utrecht, The Netherlands).

5.1.2 Culture conditions

Media

Luria Broth (LB) medium and medium used for the shake flask cultivations is the same as described in section 4.1.2.

The minimal medium consisted of 2.5 g/l (N-limited medium) or 5 g/l (C-limited medium)

$(NH_4)_2SO_4$, 2 g/l KH_2PO_4, 0.5 g/l NaCl, 0.5 g/l $MgSO_4 \cdot 7\,H_2O$, 33 g/l (N-limited medium) or 16.5 g/l (C-limited medium) Glucose $\cdot\, H_2O$, 1 ml/l vitamin solution and 100 µl/l molybdate solution (as described in section 4.1.2). All components for the medium of the continuous culture were dissolved and filter-sterilised (pore size 0.22 µm, Sartobran, Sartorius, Belgium). The pH was left at approximately 5.4.

Cultivations

Culture conditions and inoculation are described in section 4.1.2.

The experiment in which the limiting substrate was changed from glucose to ammonia was conducted at a dilution rate of $0.155\,h^{-1}$. The experiment in which the limitation was changed from nitrogen to carbon, was conducted at a dilution rate of $0.142\,h^{-1}$.

For each experiment, steady state samples were taken: optical density (OD), cell dry weight (CDW) and HPLC analysis. These samples were taken after waiting at least five residence times after the batch phase. To ensure that the perturbations caused by sampling did not influence steady state, another fiver residence times were waited before switching the limiting nutrient. The transition between the two limitations was frequently sampled for OD and metabolite data (HPLC). Subsequently again five residence times were waited before sampling the second steady state condition.

5.1.3 Measurements

All measurement techniques are the same as those described in section 4.1.3.

5.1.4 Data Analysis

Due to the amount of data that had to be combined for each experiment, a custom program was written in Python using the scipy scientific library (Jones *et al.*, 2001–2007; Pérez & Granger, 2007) to process the data and apply the algorithm as explained in section 5.2 (polynomial fitting, extracting the derivatives and calculating the fluxes). Data were reorganised to a format suitable for the programs written for metabolic flux analysis as explained in chapter 2 of this dissertation.

Each point in the time series of the transient experiments is only measured once. This is deemed sufficient, as the error propagation in the calculations is not done when approximating the data with polynomes (as described in section 5.2.2). In the MFA calculations afterwards however, a variance covariance matrix is needed. In this, it is not the absolute values of the variances that are important, but the relative ones. They represent the correctness of a measurement and thus the weight it will have in balancing the measurements in the overdetermined system. Therefore the variance covariance matrix calculated in chapter 4

was used for the MFA part of this analysis, as the covariance matrix obtained in chapter 4, is from samples processed the same way as those from this chapter.

The MFA modelling was performed as described in chapter I. The metabolic model used for this study contains 136 reactions (figure 5.1) and 150 metabolites of which 12 are exchangeable with the environment. There are 142 independent equations and $136 + 12 = 148$ unknowns. Thus at least 6 measurements have to be performed in order to fully solve the model. Actually 10 exchange metabolites were measured (CO_2, O_2, NH_3, PiOH, acetate, lactate, pyruvate, succinate, glucose and biomass) giving 4 redundant measurements for flux balancing.

5.2 Dynamic metabolic flux analysis

5.2.1 DMFA: the concept

In classical metabolic flux analysis, the measured fluxes are obtained by quantification of the metabolite concentrations in the reactor broth of a chemostat culture. But what should be done under transient conditions? Measurements of external metabolites can not directly be used for fluxes. In Provost & Bastin (2004) MFA is applied to the batch phase of the culture of Chinese Hamster Ovary (CHO) cells. The time series of the measured metabolite concentrations are approximated with a line. The slope of those lines is a direct measure of the fluxes entering and leaving the cells and is used to calculate internal reaction rates via MFA.

This approach of Provost & Bastin (2004) can be formalised and extended. Instead of limiting oneself to cases where linear regression is sufficient to adequately capture the dynamics of the measured metabolites during the transients, one can take the derivative in each point, which is, by definition, a flux.

The differential equation that governs the change of a component in the reactor broth can be written as:

$$\frac{\mathrm{d}C}{\mathrm{d}t} = D\ (C_{in} - C) + r_p \tag{5.1}$$

Where C is the concentration of the component, C_{in} the concentration of the component in the influent, r_p the production rate and D the dilution rate.

From equation 5.1 one may calculate the net reaction rate r_p, using the following approximation for the derivative in a point n of the time series:

$$\left.\frac{\mathrm{d}C}{\mathrm{d}t}\right|_n = \frac{C_n - C_{n-1}}{t_n - t_{n-1}} \tag{5.2}$$

Eventually, more points around the point of interest could be used to perform linear regression. However, because of the noisiness of the data, this way of working does not give satisfying

PGI: G6P ⟷ F6P
PFK: ATP + F6P ⟶ ADP + FBP
ALD: FBP ⟷ G3P + DHAP
TPI: DHAP ⟷ G3P
G3PDH: PiOH + NAD + G3P ⟷ NADH + H + BPG
PGK: ADP + BPG ⟷ ATP + 3PG
PGM: 3PG ⟷ 2PG
ENO: 2PG ⟷ H2O + PEP
PyrK: ADP + PEP ⟶ ATP + Pyr
PyrD: NAD + Pyr + CoA ⟶ NADH + H + AcCoA + CO2
CitSY: H2O + AcCoA + OAA ⟶ CoA + Cit
ACO: Cit ⟷ iCit
CitDH: NAD + iCit ⟷ NADH + H + CO2 + aKGA
AKGDH: NAD + CoA + aKGA ⟶ NADH + H + CO2 + SucCoA
SucCoASY: ADP + PiOH + SucCoA ⟷ ATP + CoA + Suc
SucDH: FAD + Suc ⟶ FADH2 + Fum
FumHY: H2O + Fum ⟷ Mal
MalDH: NAD + Mal ⟷ NADH + H + OAA
PEPCB: H2O + PEP + CO2 ⟶ PiOH + OAA
LacDH: NADH + H + Pyr ⟷ NAD + Lac
PFLY: Pyr + CoA ⟶ AcCoA + FA
AcKNLR: ADP + PiOH + AcCoA ⟷ ATP + CoA + Ac
Resp: 1.33 ADP + 1.33 PiOH + NADH + H + 0.5 O2 ⟶ 1.33 ATP + NAD + 2.33 H2O
H2CO3SY: H2O + CO2 ⟷ H2CO3
G6PDH: NADP + G6P ⟶ NADPH + H + 6PGL
LAS: H2O + 6PGL ⟶ 6PG
PGDH: NADP + 6PG ⟶ NADPH + H + CO2 + Rl5P
PPI: Rl5P ⟷ R5P
PPE: Rl5P ⟷ Xu5P
TK1: R5P + Xu5P ⟷ G3P + S7P
TA: G3P + S7P ⟷ F6P + E4P
TK2: Xu5P + E4P ⟷ F6P + G3P
PTS: GLC + PEP ⟶ G6P + Pyr
PPiOHHY: PPiOH + H2O ⟶ 2 PiOH
GluDH: NADPH + H + aKGA + NH3 ⟷ NADP + H2O + Glu
GluLI: ATP + NH3 + Glu ⟶ ADP + PiOH + Gln
AspSY: ATP + H2O + Asp + Gln ⟶ AMP + PPiOH + Asn + Glu
AspTA: OAA + Glu ⟷ aKGA + Asp
AlaTA: Pyr + Glu ⟷ aKGA + Ala
ValAT: aKIV + Glu ⟷ aKGA + Val
LeuSYLR: NAD + H2O + AcCoA + aKIV + Glu ⟶ NADH + H + CoA + CO2 + aKGA + Leu
aKIVSYLR: NADPH + H + 2 Pyr ⟶ NADP + H2O + CO2 + aKIV
IleSYLR: NADPH + H + Pyr + Glu + Thr ⟶ NADP + H2O + CO2 + aKGA + NH3 + Ile
ProSYLR: ATP + 2 NADPH + 2 H + Glu ⟶ ADP + PiOH + 2 NADP + H2O + Pro
SerLR: NAD + H2O + 3PG + Glu ⟶ PiOH + NADH + H + aKGA + Ser
SerTHM: Ser + THF ⟶ H2O + Gly + MeTHF
H2SSYLR: 2 ATP + 3 NADPH + ThioredH2 + 3 H + H2SO4 ⟶ ADP + PPiOH + 3 NADP + Thiored + 3 H2O + H2S + PAP
PAPNAS: H2O + PAP ⟶ AMP + PiOH
CysSYLR: H2S + AcCoA + Ser ⟶ CoA + Cys + Ac
PrppSY: ATP + R5P ⟶ AMP + PRPP
HisSYLR: ATP + 2 NAD + 3 H2O + Gln + PRPP ⟶ 2 PPiOH + PiOH + 2 NADH + 2 H + aKGA + His + AICAR
PheSYLR: Glu + Chor ⟶ H2O + CO2 + aKGA + Phe
TyrSYLR: NAD + Glu + Chor ⟶ NADH + H + CO2 + aKGA + Tyr
TrpSYLR: Gln + Ser + Chor + PRPP ⟶ PPiOH + 2 H2O + G3P + Pyr + CO2 + Glu + Trp
DhDoPHepAD: H2O + PEP + E4P ⟶ PiOH + Dahp
DhqSY: Dahp ⟶ PiOH + Dhq
DhsSYLR: Dhq ⟷ H2O + Dhs
ShiSY: NADPH + H + Dhs ⟷ NADP + Shi
ShiKN: ATP + Shi ⟶ ADP + Shi3P
ChorSYLR: PEP + Shi3P ⟶ 2 PiOH + Chor
ThrSYLR: ATP + H2O + HSer ⟶ ADP + PiOH + Thr
MDAPSYLR: NADPH + H + Pyr + SucCoA + Glu + AspSA ⟶ NADP + CoA + aKGA + Suc + MDAP
LysSY: MDAP ⟶ CO2 + Lys
MetSYLR: H2O + SucCoA + Cys + MTHF + HSer ⟶ Pyr + CoA + Suc + NH3 + Met + THF
AspSASY: ATP + NADPH + H + Asp ⟶ ADP + PiOH + NADP + AspSA
HSerDH: NADPH + H + AspSA ⟷ NADP + HSer
CarPSY: 2 ATP + H2O + H2CO3 + Gln ⟶ 2 ADP + PiOH + Glu + CarP
OrnSYLR: ATP + NADPH + H + H2O + AcCoA + 2 Glu ⟶ ADP + PiOH + NADP + CoA + aKGA + Orn + Ac
ArgSYLR: ATP + Asp + Orn + CarP ⟶ AMP + PPiOH + PiOH + Fum + Arg
ThioredRD: NADPH + Thiored + H ⟷ NADP + ThioredH2
H2O2ox: 2 H2O2 ⟶ 2 H2O + O2
FAD2NAD: NAD + FADH2 ⟷ NADH + FAD + H
AICARSYLR: 6 ATP + 3 H2O + CO2 + Asp + 2 Gln + Gly + FA + PRPP ⟶ 6 ADP + PPiOH + 6 PiOH + Fum + 2 Glu + AICAR
IMPSYLR: FTHF + AICAR ⟶ H2O + THF + IMP
AMPSYLR: Asp + GTP + IMP ⟶ AMP + PiOH + Fum + GDP
AdKN: ATP + AMP ⟷ 2 ADP
ADPRD: ADP + ThioredH2 ⟶ Thiored + H2O + dADP
dADPKN: ATP + dADP ⟶ ADP + dATP
IMPDH: NAD + H2O + IMP ⟶ NADH + H + XMP
GMPSY: ATP + H2O + Gln + XMP ⟶ AMP + PPiOH + Glu + GMP
GuKN: ATP + GMP ⟶ ADP + GDP
GDPKN: ATP + GDP ⟶ ADP + GTP
GDPRD: ThioredH2 + GDP ⟶ Thiored + H2O + dGDP
dGDPKN: ATP + dGDP ⟶ ADP + dGTP
UMPSYLR: O2 + Asp + PRPP + CarP ⟶ PPiOH + PiOH + H2O + CO2 + UMP + H2O2
UrKN: ATP + UMP ⟶ ADP + UDP
UDPKN: ATP + UDP ⟶ ADP + UTP
CTPSY: ATP + H2O + Gln + UTP ⟶ ADP + PiOH + Glu + CTP
CDPKN: ATP + CDP ⟷ ADP + CTP
CMPKN: ATP + CMP ⟶ ADP + CDP
CDPRD: ThioredH2 + CDP ⟶ Thiored + H2O + dCDP
dCDPKN: ATP + dCDP ⟶ ADP + dCTP
UDPRD: ThioredH2 + UDP ⟶ Thiored + H2O + dUDP
dUDPKN: ATP + dUDP ⟶ ADP + dUTP
dUTPPPAS: H2O + dUTP ⟶ PPiOH + dUMP
dTMPSY: MeTHF + dUMP ⟶ DHF + dTMP
dTMPKN: ATP + dTMP ⟶ ADP + dTDP
dTDPKN: ATP + dTDP ⟶ ADP + dTTP
DHFRD: NADPH + H + DHF ⟶ NADP + THF
FTHFSYLR: NADP + H2O + MeTHF ⟶ NADPH + H + FTHF
GlyCA: NAD + Gly + THF ⟷ NADH + H + CO2 + NH3 + MeTHF
MeTHFRD: NADH + H + MeTHF ⟶ NAD + MTHF
AcCoACB: ATP + H2O + AcCoA + CO2 ⟷ ADP + PiOH + MalCoA
MalCoATA: MalCoA + ACP ⟷ CoA + MalACP
AcACPSY: MalACP ⟶ CO2 + AcACP
C120SY: 10 NADPH + 10 H + AcACP + 5 MalACP ⟶ 10 NADP + 5 H2O + 5 CO2 + C120ACP + 5 ACP
C140SY: 12 NADPH + 12 H + AcACP + 6 MalACP ⟶ 12 NADP + 6 H2O + 6 CO2 + C140ACP + 6 ACP
C160SY: 14 NADPH + 14 H + AcACP + 7 MalACP ⟶ 14 NADP + 7 H2O + 7 CO2 + C160ACP + 7 ACP
C181SY: 15 NADPH + 15 H + AcACP + 8 MalACP ⟶ 15 NADP + 8 H2O + 8 CO2 + C181ACP + 8 ACP
AcylTF: C160ACP + C181ACP + Go3P ⟶ 2 ACP + PA
Go3PDH: NADPH + H + DHAP ⟷ NADP + Go3P
DGoKN: ATP + DGo ⟶ ADP + PA
CDPDGoSY: CTP + PA ⟷ PPiOH + CDPDGo
PSerSY: Ser + CDPDGo ⟶ CMP + PSer
PSerDC: PSer ⟶ CO2 + PEthAn
GlnF6PTA: F6P + Gln ⟶ Glu + GA6P
GlcAnMU: GA6P ⟷ GA1P
NAGUrTF: AcCoA + UTP + GA1P ⟶ PPiOH + CoA + UDPNAG
LipaSYLR: ATP + 2 CMPKDO + 2 UDPNAG + C120ACP + 5 C140ACP ⟶ ADP + 2 CMP + UMP + UDP + 6 ACP + Lipa + 2 Ac
A5PIR: Rl5P ⟷ Ar5P
PGLCMT: G6P ⟷ G1P
CMPKDOSYLR: 2 H2O + PEP + Ar5P + CTP ⟶ PPiOH + 2 PiOH + CMPKDO
ADPHEPSY: ATP + S7P ⟶ PPiOH + ADPHEP
UDPGlcSY: G1P + UTP ⟶ PPiOH + UDPGlc
EthANPT: CMP + PEthAn ⟷ CDPEthAn + DGo
LpsSYLR: 3 ADPHEP + 3 CMPKDO + 2 UDPGlc + Lipa + 2 CDPEthAn ⟶ 3 ADP + 3 CMP + 2 CDP + 2 UDP + Lps
PGSYLR: Go3P + CDPDGo ⟶ PiOH + CMP + PG
CLSY: PG + CDPDGo ⟶ CMP + CL
PeptidoSYLR: 5 ATP + NADPH + H + PEP + 3 Ala + MDAP + 2 UDPNAG ⟶ 5 ADP + 7 PiOH + NADP + UMP + UDP + Peptido
GlcgSY: ATP + G1P ⟶ ADP + PPiOH + Glcg
ATPHY: ATP + H2O ⟶ ADP + PiOH
DNASYLR: 2 H2O + 0.246 dATP + 0.254 dGTP + 0.254 dCTP + 0.246 dTTP ⟶ 2 PiOH + DNA
RNASYLR: 0.262 ATP + 2 H2O + 0.322 GTP + 0.2 CTP + 0.216 UTP ⟶ 2 PiOH + RNA
ProtSYLR: 2 ATP + 3 H2O + 0.0961 Ala + 0.05506 Arg + 0.04505 Asn + 0.04505 Asp + 0.01702 Cys + 0.04905 Gln + 0.04905 Glu + 0.1151 Gly + 0.01802 His + 0.05405 Ile + 0.08408 Leu + 0.06406 Lys + 0.02903 Met + 0.03504 Phe + 0.04104 Pro + 0.04004 Ser + 0.04705 Thr + 0.01101 Trp + 0.02603 Tyr + 0.07908 Val + 2 GTP ⟶ 2 ADP + 4 PiOH + 2 GDP + Prot
LipidSYLR: 0.0266 CL + 0.202 PG + 0.7714 PEthAn ⟶ Lipid
BiomSYLR: 0.004561 Glcg + 0.0002663 Lps + 0.0008933 Peptido + 0.002291 DNA + 0.01446 RNA + 0.1227 Prot + 0.003642 Lipid ⟶ Biom

Figure 5.1: List of reactions used in the model. A description of the metabolites can be found in appendix B.1.

results. Therefore approximation techniques have to be used. In this work, polynomial approximation was chosen.

5.2.2 Polynomial fitting

Savitzky & Golay (1964) developed a method to generate a smooth curve (with derivatives) from data that have uniformly spaced abscissa: a polynomial through $2m + 1$ data points is fitted and the central point of the set is approximated with the value given by that polynomial. This is repeated for each point of the dataset. They prove that this equals to use a convolution function, for which the values can be calculated once and then subsequently used in the data smoothing algorithm, resulting in a substantial decrease of computation time. Steinier *et al.* (1972) made some corrections to the initial paper and introduced matrix notations.

With use of the polynomial fitting algorithms found in Jones *et al.* (2001–2007) a software tool was developed to approximate the data by using different polynomials. In this tool, equally spaced abscissa were not needed. To cover the whole time series, a moving window approach was used (figure 5.2). The points inside the outer window, W_1, are used for fitting the polynomial. The derivative of the polynomial in the points of interest falling in the inner window W_2 are used for dynamic MFA. Making the inner window smaller than the outer window ensures a smooth transition between the different polynomials. However, if the inner window is too small compared to the outer window, the noise on the data is not filtered sufficiently, yielding non realistic fluctuations in the derivatives (the derivatives represent the change in concentration).

The windows are moved appropriately to cover the whole data range. For each time series, optimal values for W_1 and W_2 have to be sought (*i.e.* the experimenter tries different values and visually evaluates the quality of the fit). Sometimes one polynomial suffices to nicely interpolate the whole data series, but for most metabolite time series, multiple polynomials have to be chained together.

Some of the available data series were impossible to smooth with polynomials. They behaved like a typical logistic curve. Hence, the logistic curve was added to the set of functions that could be used to approximate the data:

$$L(t) = p_0 \, \frac{1 + p_1 \, e^{p_3 \, t}}{1 + p_2 \, e^{p_3 \, t}} \tag{5.3}$$

The polynomial or logistic functions give a smoothed concentration time profile and enable to calculate the derivative of the concentrations over time. This can subsequently be used in equation 5.1 to calculate the cellular production (or consumption) fluxes:

$$r_p = \frac{\mathrm{d}C}{\mathrm{d}t} - D \; (C_{in} - C) \tag{5.4}$$

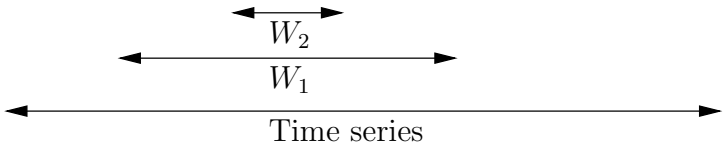


Figure 5.2: A moving window is run through a time series of data. W_1 is the polynomial fitting window, W_2 is the interpolation window.

Having r_p, classical MFA techniques (as explained in section 2.5) can be applied at each time point. The results of those models, evaluated at the different time points, can then be combined to obtain a time profile of internal fluxes.

5.2.3 Growth rate calculation

During the transient period, the assumption that the growth rate is equal to the dilution rate does not hold anymore. However, equation 5.4 can explicitly be written for biomass:

$$r_{p,Biom} = \frac{\mathrm{d}C_{Biom}}{\mathrm{d}t} + D\, C_{Biom} \tag{5.5}$$

from which the growth rate can be calculated as:

$$\mu = \frac{r_{p,Biom}}{C_{Biom}} = \frac{\mathrm{d}C_{Biom}}{\mathrm{d}t}\ \frac{1}{C_{Biom}} + D \tag{5.6}$$

5.3 Results

5.3.1 Polynomial fitting

The polynomials fitted through the data (figure 5.3 and 5.4) are used to obtain the concentration, C, and the derivative of the concentration to the time, $\mathrm{d}C/\mathrm{d}t$ in each point. From those values, the conversion rate can be calculated (equation 5.4). Biomass concentrations were calculated from OD values, after performing a calibration curve with cell dry weight (CDW).

Much tuning is needed of the parameters W_1, W_2 (see figure 5.2 on page 71) and the maximal order of the polynomials, to get adequate fits. Even then, taking the derivative will not always yield satisfying results. To avoid the steep derivatives that polynomials sometimes give, the derivative was calculated by taking the polynomial value of the previous point and calculating the slope of this point to the point of interest. One can observe this in the figures 5.3 and 5.4 as the interpolated points are connected with straight lines and not with the original polynomials.

At time zero, the media were simply switched. No cells were extracted and resuspended in the other medium, as done in Atkinson *et al.* (2002). Thus at each switch of limiting compound,

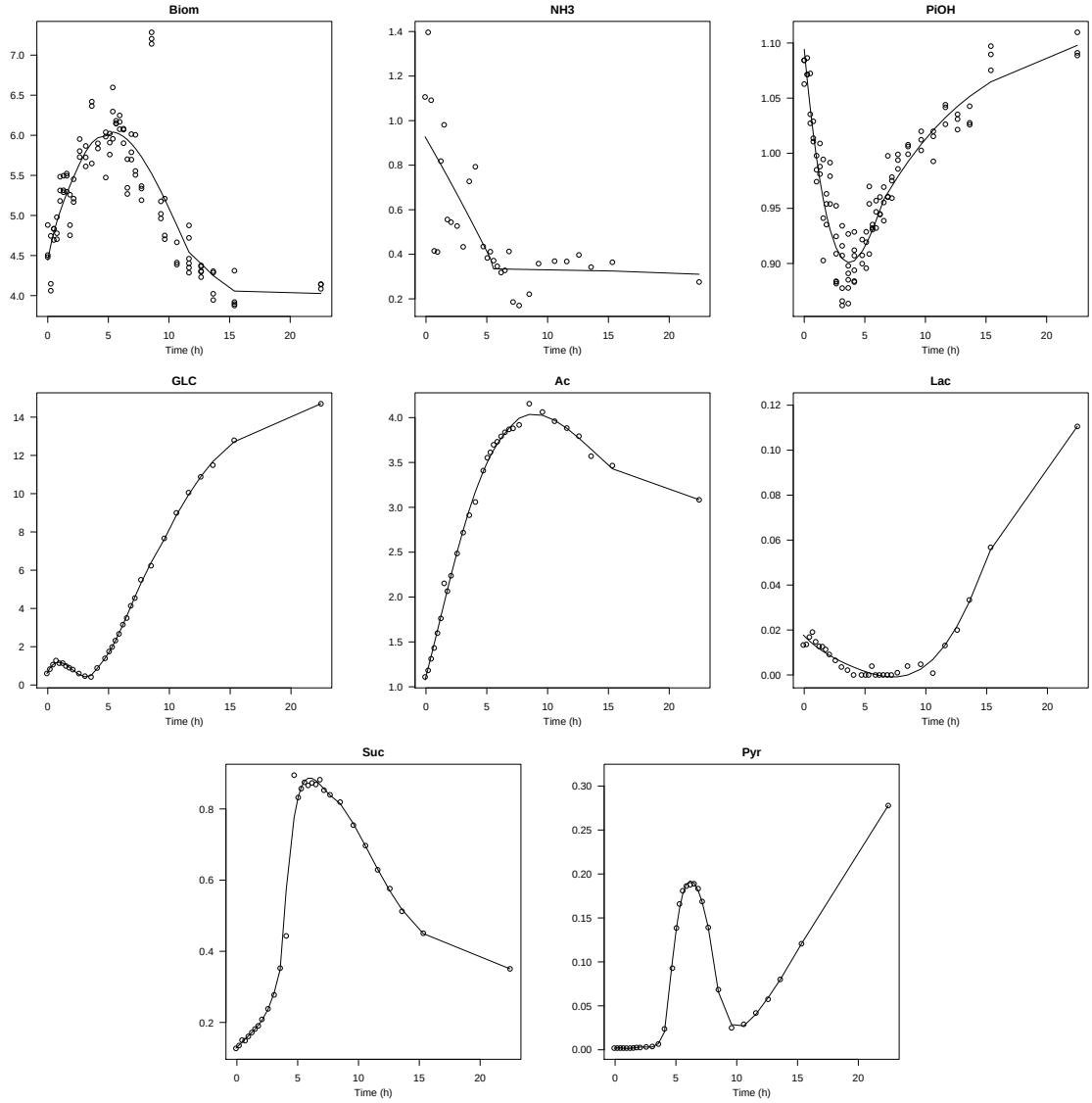


Figure 5.3: Polynomial fit of some metabolites (expressed in g/l) for the experiment where carbon-limiting medium was replaced with nitrogen-limiting medium. The switch is made at time zero, when the cells are in carbon-limited steady state. Steady state values of five residence times before the switch and five residence times after the switch are not shown.

there is an abundance of both, as the medium limited in carbon is saturated with nitrogen while the medium limited in nitrogen is saturated with glucose. In the N to C limitation case, nitrogen abundance is reached before all the glucose is depleted (figure 5.4). For the C to N limitation experiment it seems as if N limitation is reached not much later than when carbon abundance starts (figure 5.3). However, it should be noted that the steady state concentration of NH_3 in the C to N limitation experiment is around 0.2 g/l (data not shown on figure 5.3, as this value falls out the time frame shown) while it is 0.1 g/l in the N to C limitation case. Probably that even this is too high and that for low concentrations of nitrogen, the LCK238

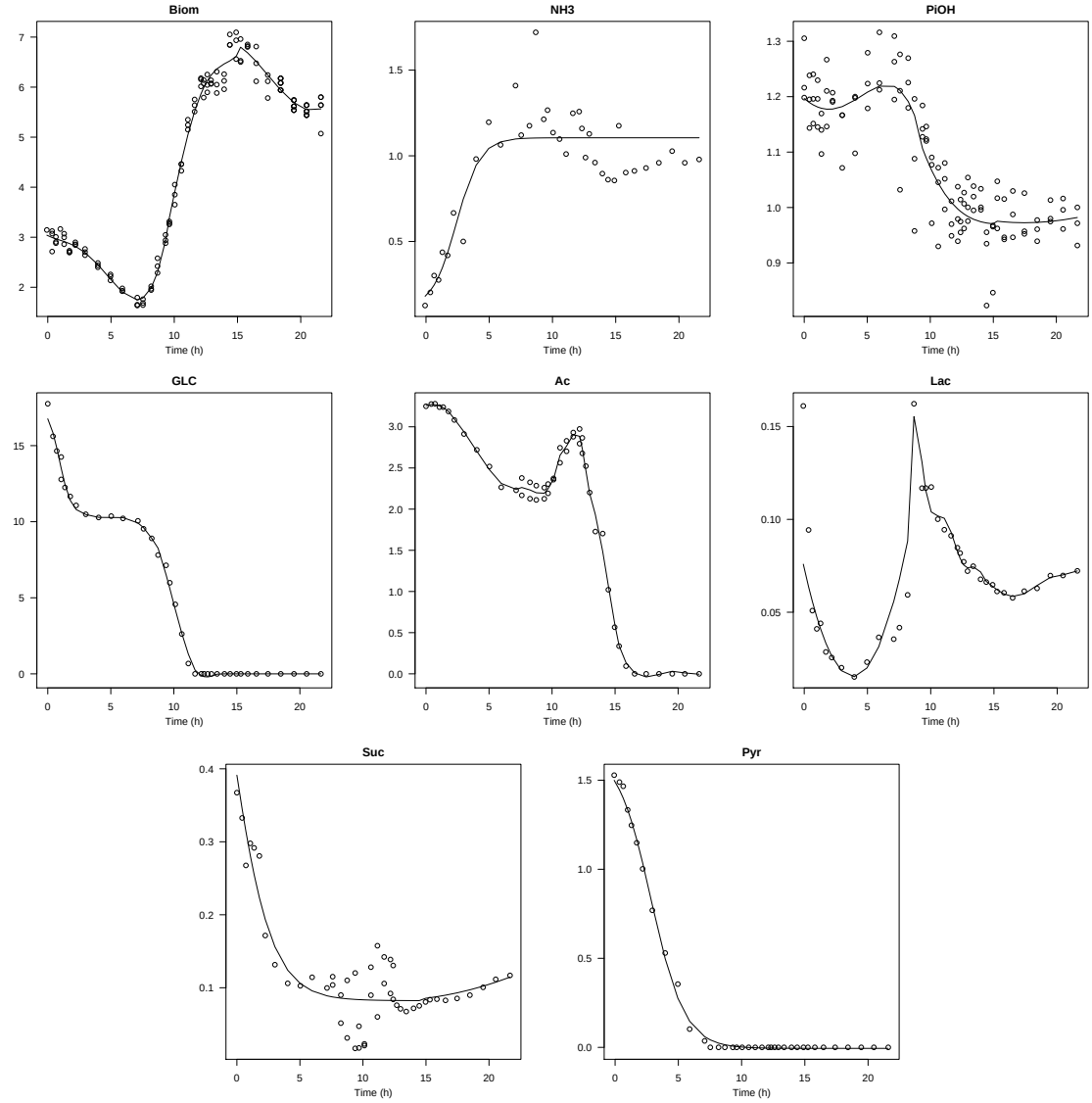


Figure 5.4: Polynomial fit of some metabolites (expressed in g/l) for the experiment where nitrogen-limiting medium was replaced with carbon-limiting medium. The switch is made at time zero, when the cells are in nitrogen-limited steady state. Steady state values of five residence times before the switch and five residence times after the switch are not shown.

nitrogen measurement kit from Hach Lange is not suitable.

5.3.2 Carbon balance

As can be seen in figure 5.5, the carbon balances are in most cases not significantly different from one. In the N to C substrate limitation experiment, the errors on the carbon balance are larger. This could be due to some accumulation effects in the first hours after the switch, where the biomass concentration decreases. The excess carbon out around 12 hours after the

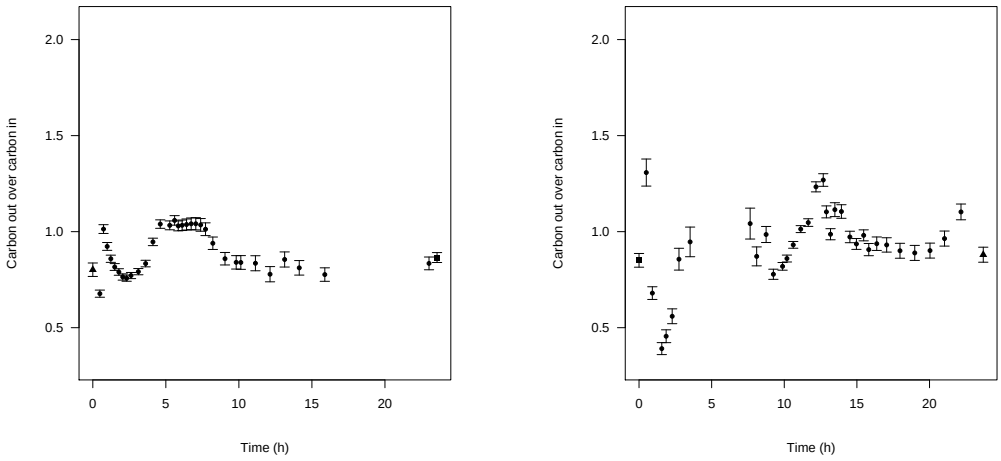


Figure 5.5: Carbon balance. Left: C-limitation to N-limitation; right: N-limitation to C-limitation. Error bars represent the standard deviations.

switch coincides with the acetate peak and the start of the glucose limitation (right part of figure 5.4). However, these data points are not eliminated and the analysis is continued with these points included.

5.3.3 Metabolic flux analysis

A long lag phase was observed in the experiment in which the nitrogen-limited culture was switched to carbon-limitation (figure 5.4). To try to find a possible explanation for this, the nitrogen metabolism is briefly reviewed.

Nitrogen metabolism regulation

The central molecule in the signal transduction mechanisms responsible for the regulation of nitrogen metabolism in *E. coli* is PII (Ninfa & Atkinson, 2000). To be active, PII must be uridylylated by a bifunctional uridylyltransferase/uridylylremoving enzyme (UTase/UR, *glnD* gene product). High glutamine concentrations in the cell (indicating ammonium sufficiency) activate the UR activity: PII is not uridylylated and is inactive. Low glutamine concentrations (indicating an ammonium limitation) result in UTase activity and PII is activated to PII-UMP. It was shown *in vitro* (Jiang & Ninfa, 1999; Ninfa & Jiang, 2005) that very high concentrations of alpha-ketoglutarate bind with PII. In this state, PII would not be able to bind with NtrB which finally would result in NtrC being phosphorylated and thus activated (NtrB and NtrC are nitrogen metabolism regulators). But this hypothesis could not be proven experimentally (Maheswaran & Forchhammer, 2003).

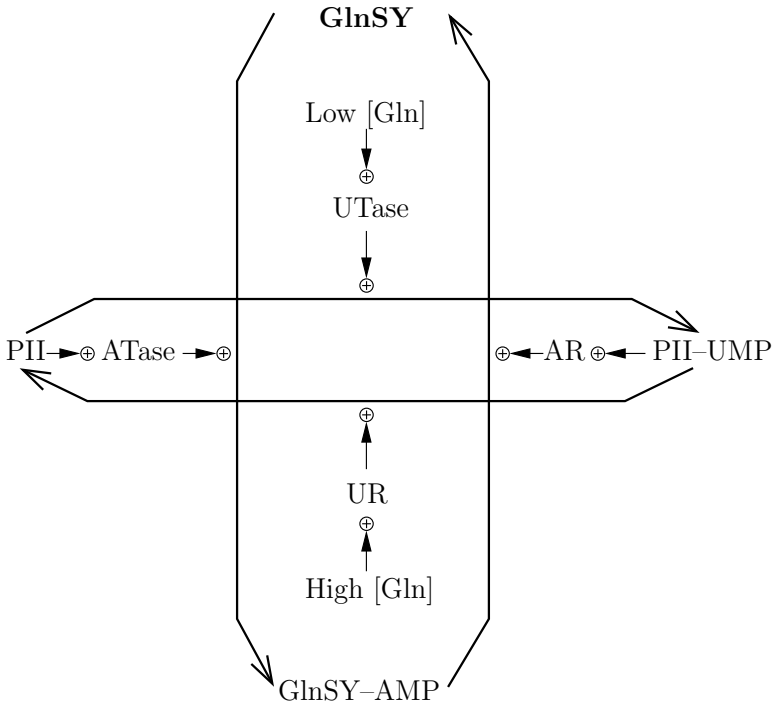


Figure 5.6: Metabolic regulation of GlnSY (Glutamine synthetase). UTase: uridylyltransferase, UR: uridylyl removing function of the UTase; ATase: adenylyltransferase, AR adenylyl removing function of ATase.

The three reactions for incorporation of ammonia in *E. coli* are:

$$\begin{aligned}
\text{GluDH}: &\quad \text{aKGA} + \text{NH}_3 + \text{NADPH} \rightarrow \text{Glutamate} + \text{NADP}^+ \\
\text{GlnSY}: &\quad \text{Glutamate} + \text{NH}_3 + \text{ATP} \rightarrow \text{Glutamine} + \text{ADP} \\
\text{GluSY}: &\quad \text{Glutamine} + \text{aKGA} + \text{NADPH} \rightarrow 2\,\text{Glutamate} + \text{NADP}^+
\end{aligned}$$

Under nitrogen rich conditions, NH_3 is incorporated into the cellular metabolism by glutamate dehydrogenase (GluDH). At concentrations below 1 mM, the affinity of GluDH to NH_3 is too low and the reaction tends to occur in reverse direction, liberating ammonia (Burkovski, 2003). Then glutamine synthetase (GlnSY, gene product of *glnA*) and glutamate synthase (GluSY) take over. When growing under energy rich conditions, 15 % of the ATP requirement of the cells is due to the GlnSY/GluSY system (Reitzer, 2003), thus GlnSY has to be down-regulated under nitrogen rich conditions (figure 5.6).

The regulation of GlnSY is achieved by adenylylation, catalysed by adenylyltransferase (ATase, gene product of *glnE*). PII-UMP (nitrogen poor conditions) stimulates the deadenylylation function of ATase and GlnSY-AMP is converted to active GlnSY, while PII (nitrogen rich conditions) activates the adenylylation of GlnSY (Bueno *et al.*, 1985).

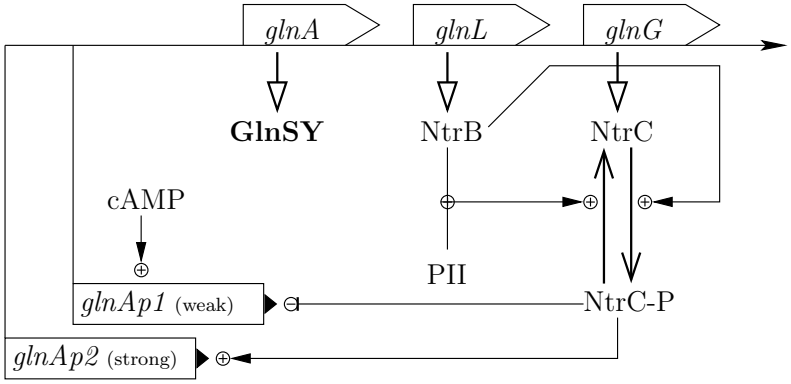


Figure 5.7: Genetic regulation of *glnALG* by the NtrB/NtrC two component system.

The transcription of *glnALG* is regulated by the NtrB/NtrC two component system (figure 5.7). Phosphorylation of NtrC (gene product of *glnG*) is catalysed by NtrB (gene product of *glnL*). But NtrB combined with PII (thus under nitrogen sufficient conditions) has phosphatase activity and dephosphorylates NtrC (Keener & Kustu, 1988). The *glnALG* operon is transcribed from tandem promoters. Expression from the weak promoter *glnAp1* (using σ^{70}-factor, Mao *et al.*, 2007) is repressed by NtrC-P and activated by high cAMP levels (typically found in carbon-limiting conditions) while expression from the strong promoter *glnAp2* (using σ^{54}-factor, Mao *et al.*, 2007) is activated by NtrC-P (Reitzer & Magasanik, 1985). Thus basal levels of glutamine synthetase, and its gene expression regulators NtrB and NtrC are assured. The genes *glnB* (PII), *glnD* (UTase) and *glnE* (ATase) are constitutively expressed and thus not under influence of the nitrogen status of the cells (van Heeswijk *et al.*, 1993).

Actually, *E. coli* contains two PII like proteins: the gene product from *glnB*, that is constitutively expressed (Mao *et al.*, 2007) and the gene product of *glnK*. The expression of *glnK* depends on the presence of UTase, NtrC and the absence of ammonia (van Heeswijk *et al.*, 1996).

The operon containing *glnK* also has the genetic code for *amtB*. AmtB facilitates diffusion of NH_3 and is not an ammonium transporter (Soupene *et al.*, 2002; Javelle *et al.*, 2005). However, NH_4^+ rather than NH_3 is present in the medium at neutral pH. Structural determination of AmtB showed that there is an extracellular high affinity ammonium ion binding site for scavenging NH_4^+ at very low concentrations (also to discriminate against water, which otherwise might compete too effectively with ammonia). As no functional group for splitting the H^+ ion off is present near this site, small basic molecules (*e.g.* phosphates, carbonates) must be present in the periplasmic region (Winkler, 2006).

Prokaryotes in which the gene pair *glnK amtB* is not present, are pathogenic bacteria, whose genomes have undergone reductive evolution and who get their nitrogen from the host organ-

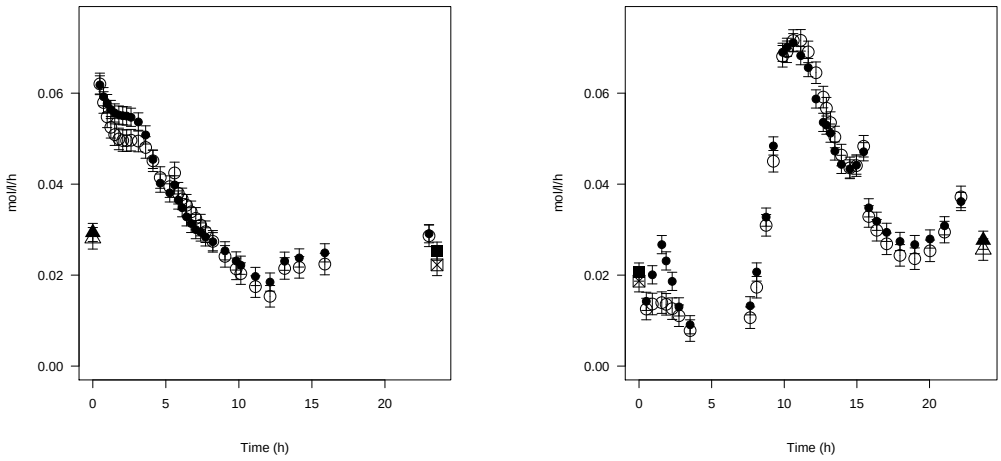


Figure 5.8: Fluxes in mol/l/h of the biomass production. Open symbols are the values as derived from the polynomials with formula 5.4, closed symbols are values obtained after flux balancing. Left: C-limitation to N-limitation; right: N-limitation to C-limitation. The first point left of each figure is the steady state value before the medium switch. The last point right on each figure is the steady state value after at least 50 hours. Error bars represent the standard deviations.

ism (Thomas *et al.*, 2000). Interestingly it was shown that *E. coli* lacking *amtB* grows faster on the poor nitrogen source arginine than the wild-type, leading to the hypothesis that NH_3 may diffuse outward through AmtB when the internal concentration is much higher than the external one (Soupene *et al.*, 1998). Under ammonium abundance, GlnK-UMP (signaling nitrogen rich conditions) binds to AmtB and inactivates it (Couts *et al.*, 2002; Javelle *et al.*, 2004).

Lag phase when switching from N-limitation to N-abundance

In the beginning of the switch from nitrogen-limiting medium to glucose-limiting medium, the glucose-limiting medium adds nitrogen to the culture broth while there is still an abundance of glucose. This can be seen in figure 5.4, in the glucose profile. The glucose concentration is only zero after 12 hours while the nitrogen concentration already reaches steady state after about 5 hours. At the medium switch, as there is extra nitrogen, the biomass growth could increase, but apparently *E. coli* needs some time to adapt to the new conditions. The biomass flux even decreases as there is nitrogen added and becomes almost zero (figure 5.8). Only seven hours after the medium switch, the growth rate increases sharply to attain almost $0.6\,h^{-1}$ (see the right part of figure 5.9). Actually this increased growth rate coincides with the uptake of all glucose left in the broth (figure 5.4).

The ATP hydrolysis reaction (figure 5.10) lumps all the maintenance requirements and futile

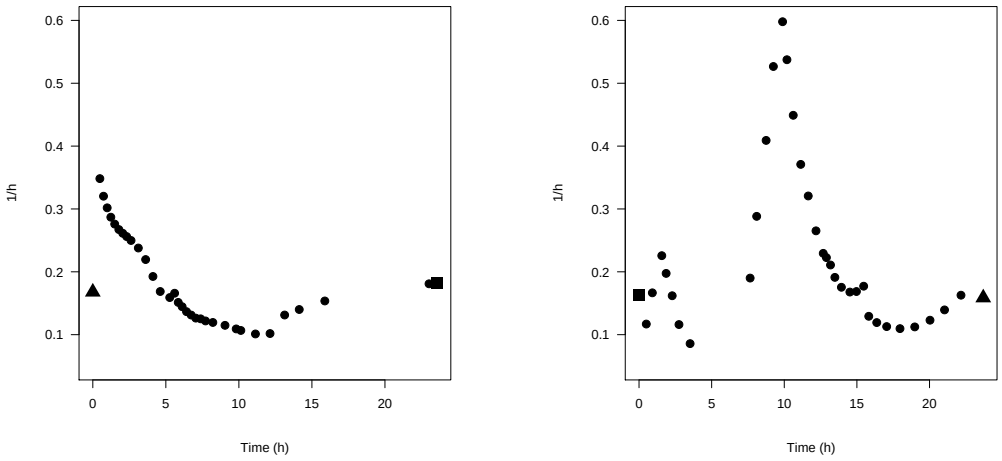


Figure 5.9: The growth rate of the cells during the transients. Left: C-limitation to N-limitation; right: N-limitation to C-limitation. The first point left of each figure is the steady state value before the medium switch. The last point right on each figure is the steady state value after at least 50 hours. Error bars represent the standard deviations.

cycles (as including all the futile cycles in the metabolic model would yield parallel pathways and the system of equations would not be solvable without measuring intracellular fluxes). It can be seen that during the lag phase of the experiment switching nitrogen to carbon-limitation (right part of figure 5.10), the ATP maintenance requirement is around 1 molATP/molBiom/h, which is high compared to the values found in chapter 3, figure 3.6. It seems even more unusually high because there is minimal growth during this period. This high ATP requirement in the beginning of the lag phase correlates with a high uptake of glucose (PTS reaction in figure 5.12) and an increased flux in the TCA cycle (5.12).

Normally bacterial cells are highly resistant to ammonium and the negative effects of high NH_4^+ concentrations are due to an enhanced osmolarity or increased ionic strength of the medium and not caused by ammonium itself (Müller *et al.*, 2006). But when changing the medium from nitrogen-limitation to carbon-limitation, the environment of the bacteria switches from low NH_4^+ concentrations to high ones, and it could be that *E. coli* needs a certain time to adapt. The scavenging active during N-limitation could still be active in the beginning of the transition, resulting in a high influx of ammonia. To counter this, active efflux is needed, draining ATP and explaining the high energetic cost in the beginning of the addition of ammonium (right part of figure 5.10).

Such a mechanism has been observed in plants, where NH_4^+ toxicity is due to the inability of root cells to limit the influx of ammonium. The high cytosolic NH_4^+ concentration activates high-capacity, energy demanding, ammonium efflux systems. This ammonium efflux can constitute as much as 80 % of primary influx, resulting in a futile cycle of nitrogen across

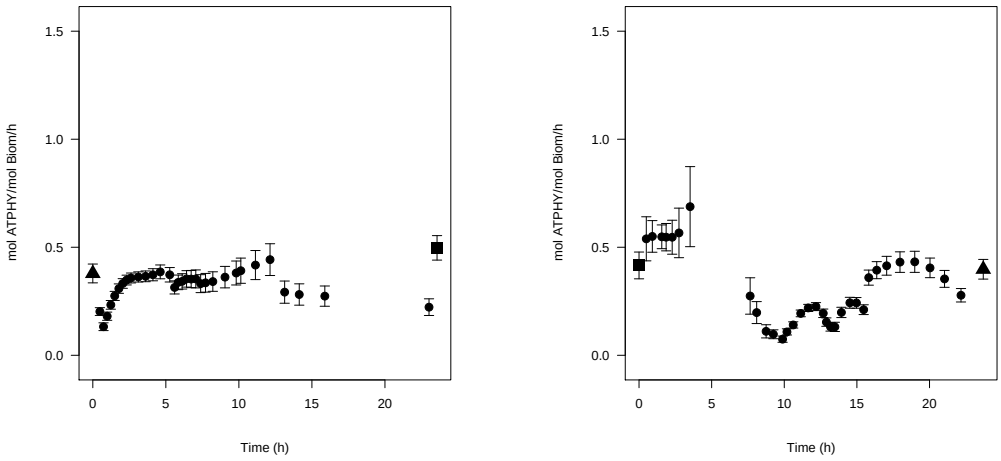


Figure 5.10: Fluxes through the ATP hydrolysis reaction (ATPHY reaction); it can be considered as a measure for the maintenance requirement of the cells. Left: C-limitation to N-limitation; right: N-limitation to C-limitation. The first point left of each figure is the steady state value before the medium switch. The last point right on each figure is the steady state value after at least 50 hours. Error bars represent the standard deviations.

the plasma membrane of root cells. This futile cycle carries a high energetic cost that is independent of N metabolism and is accompanied by a decline in growth. Plants that are resistant to high ammonium concentrations (*e.g. Oriza sativa*), limit the influx of ammonium by lowering the polarisation of the cellular membrane. This lowering of membrane polarisation is not observed in *Hordeum vulgare*, sensitive to ammonium (Britto *et al.*, 2001).

However, the lack of growth resulting from adding ammonium to cells adapted to ammonium limitation has previously only be observed in cells lacking *glnE* (giving the gene product ATase, responsible for activating/inactivating glutamine synthetase) while wild-type *E. coli* did not show an impaired growth when shifted from ammonium limitation to excess (Kustu *et al.*, 1984; Atkinson & Ninfa, 1998). Furthermore, the lack of growth in *glnE* mutants could be alleviated in cells constitutively expressing only low levels of glutamine synthetase (Kustu *et al.*, 1984), suggesting that the toxicity of ammonium in this case is not due to NH_4^+ itself, but maybe to accumulation of glutamine and/or glutamate. Normally, when shifting from ammonium limitation to carbon-limitation, the amount of active glutamine synthetase decreases instantly (Friedrich & Magasanik, 1977). Possibly due to the lack of ATase, the enzymatic glutamine synthetase regulation is lost, and only genetic regulation is present (figure 5.6). If this would be the case in the strain used in this study, it could explain the long lag phase after switching: the cells have to wait until enough glutamine synthetase is denatured before being able to grow again.

Thus it seems like *E. coli* MG1655, the strain used in this study, has some anomalies in its

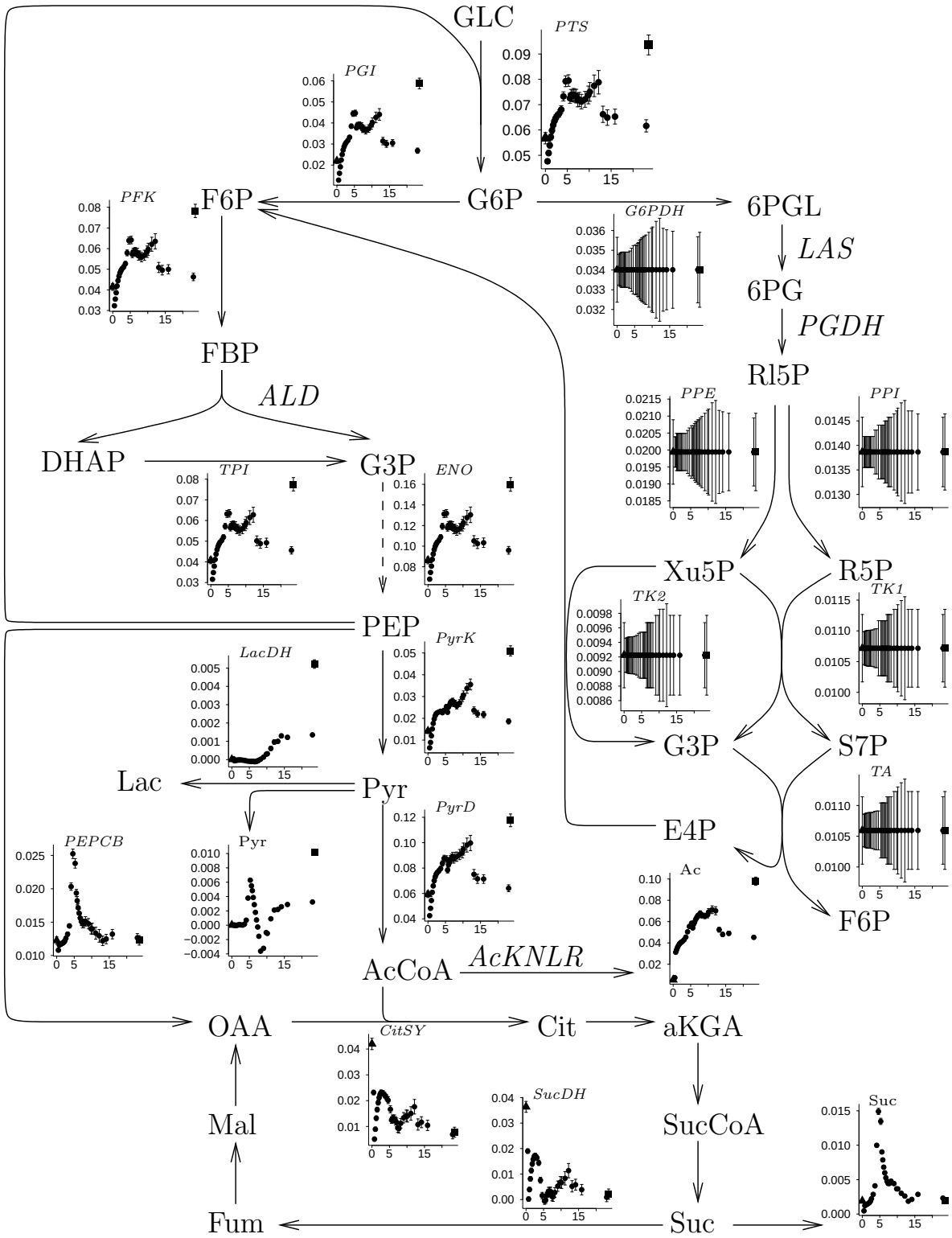


Figure 5.11: Fluxmap of the glycolysis, penthose phosphate pathway and citric acid cycle for the experiment where carbon-limiting medium is changed to nitrogen-limiting medium at time zero. The ordinate on each graph represents the flux expressed in mol/mol BM/h while the abscissa represents time. The first point left of each figure is the steady state value before the medium switch. The last point right on each figure is the steady state value after at least 50 hours. Error bars represent the standard deviations.

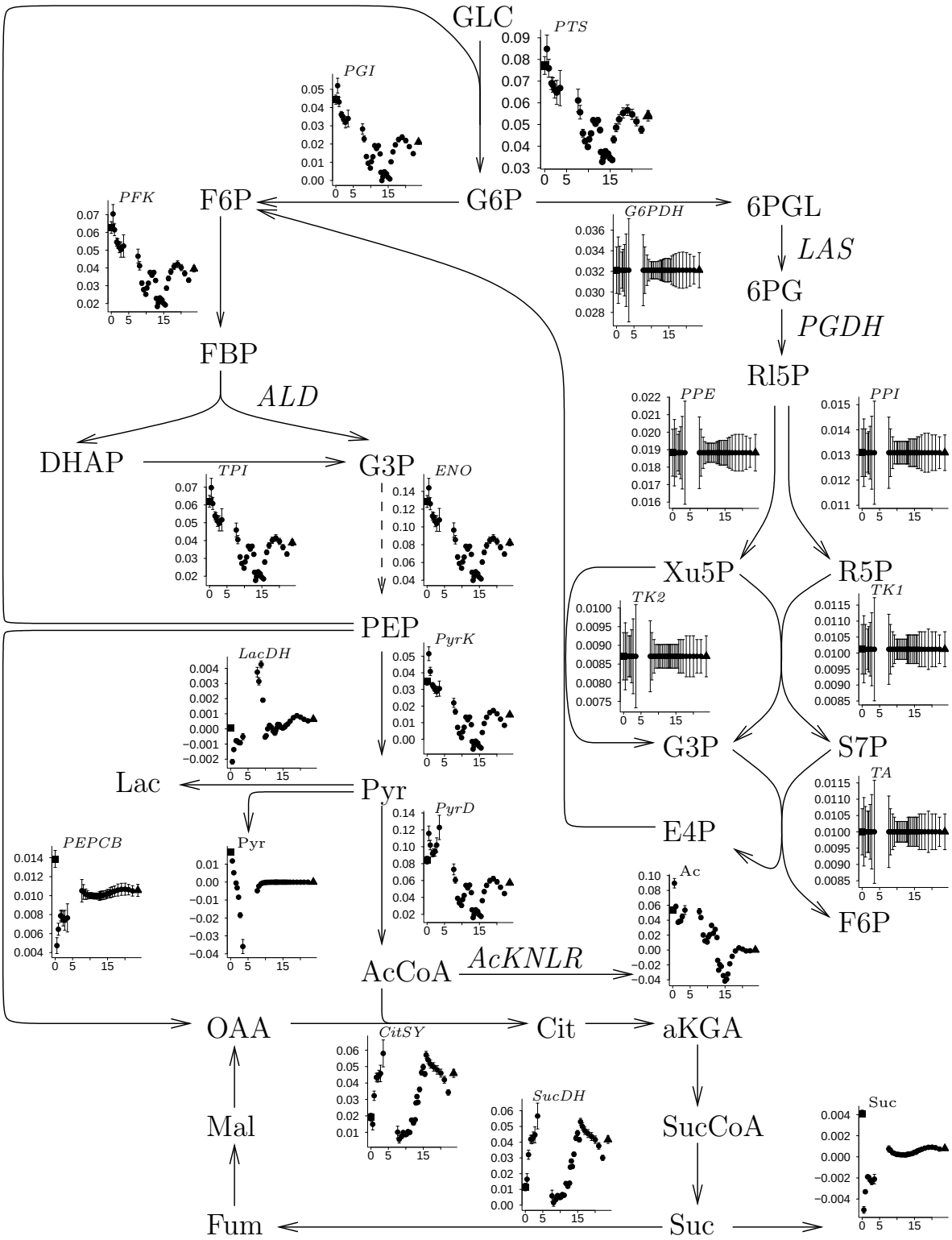


Figure 5.12: Fluxmap of the glycolysis, penthose phosphate pathway and citric acid cycle for the experiment where nitrogen-limiting medium is changed to carbon-limiting medium at time zero. The ordinate on each graph represents the flux expressed in mol/mol BM/h while the abscissa represents time. The first point left of each figure is the steady state value before the medium switch. The last point right on each figure is the steady state value after at least 50 hours. Error bars represent the standard deviations.

nitrogen metabolism. It is known that multiple strains are denominated under this name (Soupene *et al.*, 2003). The strain used in this work originated from the Coli Genetic Stock Center (CGSC) (personal communications from the Netherlands Culture Collection of Bacteria). The NCCB acquired this strain from CGSC prior to 1991. At that time, the strain was sent out as being *fnr*$^-$. Later it was found that the strain was actually a mixture of *fnr*$^+$ and *fnr*$^-$ strains (personal communication from Mary Berlyn). For the experiments in this chapter, it turned out that the *fnr*- strain was used. More research is needed to find out why this strain does not grow well when NH_4^+ is added to a nitrogen-limiting environment.

The lag phase found in the N-limitation to C-limitation experiment stands in sharp contrast with the experiment in which the carbon-limiting medium is replaced with the nitrogen-limiting one and thus carbon is added in the beginning of the switch while nitrogen is not yet limiting. The growth rate increases instantly (figure 5.8). It is known that during carbon-limitation, sugar regulons are upregulated, by internally synthesised sugars (Death & Ferenci, 1994). This was observed in chapter 3, where it can be seen in figure 3.11, that under glucose limitation *ptsG* is upregulated. The results of the experiments described here show that this upregulation permits the cells to almost instantly increase their growth rate when carbon is added to a carbon-limited culture. The cells grow as fast as possible, until all nitrogen is consumed and thus the glucose concentration in the reactor broth increases around the same time as the nitrogen decreases (figure 5.3).

The peak of succinate (and the smaller peak of pyruvate) occurring around 5 hours after the medium switch (figure 5.11) is probably due to the depletion of ammonium (figure 5.3). The high flux through the glycolysis can not be stopped sufficiently quickly and the carbon is excreted as pyruvate or diverted to PEPCB to be excreted as succinate.

Growth rates

Figure 5.9 depicts the different growth rates during the transient conditions occurring when switching the limiting substrate. The same trends as in the biomass flux (figure 5.8) can be observed. For the C- to N-limitation case, a sharp increase in growth rate is observed in the beginning, while for the N- to C-limitation the growth rate decreases to almost zero before increasing to $0.6\,h^{-1}$, not far from the maximal growth rate of $0.7\,h^{-1}$. In the carbon to nitrogen-limitation case, the maximal possible growth rate is never achieved and after a sharp increase in the beginning to $0.35\,h^{-1}$ the growth rate decreases. Acetate inhibition (Luli & Strohl, 1990) could be a possible cause. Externally this is not observable in the first hour (figure 5.3), but internally it can be seen that together with the growth rate, the acetate flux increases dramatically (figure 5.11) suggesting an high intracellular acetate concentration responsible for inhibition.

The sharp increase in growth rate of the cells at the beginning of the medium switch in the

C- to N-limitation experiment (left part of figure 5.9), is not accompanied by an increased flux through the Krebs cycle (figure 5.11, the first hour after the steady state point, labeled with a triangle). An increase in biomass flux does not imply that intracellular fluxes relative to the biomass flux should also increase. On the contrary, in the beginning the cells are more energy-efficient than during the carbon-limited steady state (left subfigure of figure 5.10). However, very quickly the flux through the glycolysis increases.

5.4 Conclusions and perspectives

The concept of dynamic metabolic flux analysis was introduced. It was shown how to transform extracellular measurement data from dynamic experiments to flux values. It is not always clear in transient experiments whether the decrease in extracellular metabolites is due to the cells stopping production and the remaining product diluting out or whether the cells are actively taking up the metabolites. Transforming the extracellular time series of concentrations to flux values can give the answer to that question. Furthermore, those flux values can help to get insight into the intracellular reactions.

The transformation of time series of concentration measurements to flux values is based on differentiation (in the mathematical sense: finding the derivative) of those time series. Differentiation typically amplifies the noise on the data, therefore a noise-reducing step is needed prior to the differentiation. In this work, this was done by polynomial filtering. Extensive manual tuning of the parameters W_1 and W_2 (windows of the polynomial interpolation, see figure 5.2) was needed to get acceptable filtering. Manually drawing approximating (bezier) curves to the measurement data would be a better approach because a researcher would be able to incorporate his expert knowledge of the cellular processes and culture methods. For example, in the beginning and at the end of the time series, the curves should have a first derivative being zero, as the transient experiments start with cells in steady state and (another) steady state is reached at the end. Also, the information seen in one metabolite could be used in fitting the curve of other metabolites. For example, in figure 5.4, there is a small dip at the beginning of the fitted curve of PiOH. This is probably just caused by random noise, as at the same time the biomass goes down and one would expect an increase in phosphate concentration of the broth. When drawing the curves manually, it would also be easier to ignore differences in variance in a time series (*e.g.* Suc in figure 5.4).

Of course, manually drawing curves is not very reproducible and there will be (small) differences between different individuals. Automated methods that are able to cope with non-uniform variance, that can do the fit for the whole dataset at once (and thus use the time series from one metabolite for the fitting of other ones) and in which it is possible to define beginning and ending conditions to the curves (derivatives being zero) are preferred, but they will probably also require a lot of fine-tuning of the fitting parameters and may still not give

satisfying results.

A lag phase was observed when cells adapted to a nitrogen-limiting environment, were supplied with nitrogen. During this lag phase of several hours, ATP demand was high. No clear physiological reason could be given. The only similar case described in literature for *E. coli* is a *glnE* knock out, in which case the toxicity was not due to ammonium itself, but to the accumulation of glutamine/glutamate because the downregulation of glutamine synthetase is not operative (glutamine synthetase is downregulated by the gene product of *glnE*). A possible explanation for the long lag phase observed in the N-to-C-limitation experiment could be that sufficient glutamine synthetase has to denature before the accumulation of glutamine/glutamate does not impair growth anymore.

Chapter 6

Protein overproduction strain: influence on metabolism[1]

Escherichia coli is widely used for recombinant protein production, largely because it is well characterised, fast and inexpensive to grow, and relatively easy to alter genetically. During the biochemical synthesis of proteins, production of nonessential metabolites can waste carbon and energy that might otherwise be directed towards protein production. A prominent example of a metabolite that accumulates during aerobic growth of *E. coli* on glucose is acetate. Enzymatically synthesised from acetyl-coenzyme A (acetyl-CoA) in two steps —phosphotransacetylase (encoded by *pta*) converts acetyl-CoA to the intermediate acetyl phosphate, which is subsequently converted to acetate with the generation of ATP by acetate kinase (*ack*)— acetate is also widely considered inhibitory to growth and protein production (Aristidou *et al.*, 1994; Chou *et al.*, 1994; Jensen & Carlsen, 1990; Luli & Strohl, 1990).

Approaches for increasing cell density or protein yield in *E. coli* often focus on the reduction of acetate formation, and a variety of methods has been studied (De Mey *et al.*, 2007a). The production of acetate can be blocked altogether, for example, by mutations in the *pta* and/or *ack* genes (Dedhia *et al.*, 1994; Diaz Ricci *et al.*, 1991; LeVine *et al.*, 1980; San *et al.*, 1994). Alternatively, acetate accumulation can be reduced by redirecting it or its precursors to other biochemicals. For example, pyruvate can be converted to acetoin by acetolactate synthase (Aristidou *et al.*, 1994, 1995). Other strategies include various glucose feeding regimes (Kleman *et al.*, 57, 1991, 1996; Kleman & Strohl, 1992, 1994; Lee, 1996; Riesenberg & Guthke, 1996), limiting growth rate by carbon-limited fed-batch schemes (Akesson *et al.*, 2001; Kim *et al.*, 2004; Konstantinov *et al.*, 1990; Lee *et al.*, 1999; Lee, 1996; Lin *et al.*, 2001) and utilisation of alternative feeds such as glycerol (Hahm *et al.*, 1994; Lee, 1996), mannose (Andersen &

[1]Parts of this chapter will shortly be submitted: M. De Mey, G. J. Lequeux, J. J. Beauprez, J. Maertens, E. Van Horen, W. Soetaert, P. A. Vanrolleghem & E. J. Vandamme (2008). Transient metabolic modeling of *Escherichia coli* MG1655 and MG1655 $\Delta ackA - pta$, $poxB$ Δp$pc - p37$ for recombinant β-galactosidase production. In preparation.

von Meyenburg, 1980) or fructose (Aristidou *et al.*, 1999). Also, supplementing the medium (*e.g.* amino acids) has proven to be positive on reducing acetate (Han *et al.*, 1992). Another approach to keep the growth rate below the threshold for acetate production is the pH-stat, where a nutrient feed is activated when the pH increases and variants, where the culture is dosed with more nutrients than necessary. Unfortunately, many of these strategies reduce the glucose uptake rate, which can simultaneously reduce the rate of protein production.

Recombinant protein production is believed to diminish the flow in the TCA cycle through the withdrawal of the intermediates that serve as biochemical precursors. Ten amino acids are biochemically derived from TCA cycle intermediates: aspartate, asparagine, methionine, threonine, isoleucine, and lysine are derived from oxaloacetate, while glutamate, arginine, proline, and glutamine are derived from α-ketoglutarate. The additional metabolic burden resulting from recombinant protein production would likely further diminish the availability of oxaloacetate, which could lead to additional acetate formation from acetyl-CoA. If withdrawal of TCA cycle intermediates limits cell growth and protein production and consequently increases acetate accumulation, then providing cells with improved metabolic means to replenish these TCA cycle intermediates should increase protein production.

Anaplerotic biochemical pathways are the enzymatic reactions that replenish TCA cycle intermediates. In *E. coli*, the principal anaplerotic pathway during growth on glucose is the formation of oxaloacetate from phosphoenolpyruvate (PEP) by PEP carboxylase, and this single pathway must supply carbon for the amino acids and other cellular building blocks derived from TCA cycle intermediates. Previous research indicated that a combination of directly reducing the carbon flow to acetate and anticipating on the underlying metabolic and regulatory mechanism that lead to acetate, is the most promising approach to overcome acetate formation and improve recombinant protein production (De Mey *et al.*, 2007b). Also, expressing the heterologous anaplerotic enzyme pyruvate carboxylase of *Rhizobium etli* in *E. coli* allows replenishing the TCA cycle intermediates and seemed to have a positive influence on recombinant protein production (March *et al.*, 2002).

Aim of work

The aim of this work was to improve our understanding of recombinant protein synthesis in *E. coli*. To achieve this, we analysed the production of a model recombinant protein, β-galactosidase, in response to the constitutive overexpression of the anaplerotic PEP carboxylase pathway. Therefore, a *ppc* overexpressing mutant of *E. coli* MG1655 $\Delta ackA$–pta, $\Delta poxB$ was created in which the natural *ppc* promoter was replaced with a strong constitutive artificial promoter (De Mey *et al.*, 2007c).

Metabolic flux analysis permits to calculate intracellular reaction fluxes by measuring only exchange rates of the cell with the environment. Thus with concentration measurements of

the metabolites in the reactor broth, MFA permits to know the fluxes in the cells. But this requires that the reactor is in steady state. However, in this work, β-galactosidase production is started after the cultures reach steady state. Therefore, dynamic MFA (from which the principles are explained in section 5.2) is utilised and MFA is applied to an overdetermined system that is not in steady state. At each time point, the fluxes are calculated and used as input for the metabolic flux model. In this way the intracellular fluxes of the transient phase during which recombinant protein production is induced, can be revealed.

6.1 Material and methods

6.1.1 Bacterial strain and plasmids

Escherichia coli MG1655 [λ^-, F^-, *rph*–1, (*fnr*$^-$ 267)del] (labeled WT in the figures) was obtained from the Netherlands Culture Collection of Bacteria (NCCB, Utrecht, The Netherlands). *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* [λ^-, F^-, *rph*–1, *rfb*–50, (*fnr*$^-$ 267)del, Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37*] was constructed in the Laboratory of Genetics and Microbiology (MICR) using the method of Datsenko & Wanner (2000). Plasmid pTrcHisTopoLacZ was obtained from Invitrogen (Merelbeke, Belgium). This expression plasmid contains the model recombinant protein β-galactosidase coded by the gene *LacZ*.

The expression plasmid pTrcTopoHisLacZ was transformed in competent cells of *E. coli* MG1655 and *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* using the simplified procedure of Hanahan *et al.* (1991).

6.1.2 Culture conditions, sampling and sample analysis

Carbon-limited chemostat experiments of *E. coli* were performed in two liter Biostat B culture vessels (Sartorius-BBH Systems, Melsungen, Germany) with 1.5 l working volume as described in section 4.1.2. Sampling for CDW, extracellular measurements and metabolites analysis was done as described in section 4.1.3.

6.1.3 Step experiment

Step experiments were performed after the cells attained steady state. Then, during 7 hours a 1 mM IPTG solution (isopropyl-β-D-thiogalactopyranoside) was supplied to the vessel to induce the *ptrc* promoter on the pTrcHisTopoLacZ plasmid.

During the step experiment, every 15 min samples for OD_{600}, extracellular metabolite measurements and β-galactosidase activity were taken using the rapid sampling loop and the cold stainless bead sampling method, immediately followed by centrifugation (section 4.1.3).

6.1.4 Data analysis

Glucose and organic acids were determined by High Performance Liquid Chromatography (HPLC) on a Varian Prostar HPLC system (Varian, Sint-Katelijne-Waver, Belgium). Phosphate determination was done as described in section 4.1.3. Quantification of total nitrogen in the culture medium was performed using the LCK238 kit from HACH Lange GmbH (Mechelen, Belgium) according to the manufacturers protocol.

6.1.5 β-galactosidase assay

Cell lysis using the EasyLyse-kit (Epicentre Biotechnologies, BIOzymTC, Landgraaf, The Netherlands) was done as recommended by the manufacturer in the EasyLyse Bacterial Protein Extraction Solution manual. β-galactosidase activity was assayed as described by Miller (1972) with some modifications: 50 µl sample was added to 200 µl 16 mM ortho-nitrophenyl-galactopyranoside (ONPG, in phosphate buffer: 100 mM phosphate, 1 mM $MgCl_2$, pH 7.4). The absorbance at 415 nm was measured in a microplate reader (680 XR microplate reader, Bio-Rad, Eke, Belgium) during 10 minutes. For each batch of measurements, a calibration curve was made using a stock solution of 1 unit/µl commercial β-galactosidase of *Escherichia coli* (Fluka, Bornem, Belgium). This stock solution was used to make a standard series from 0 units/µl to 0.003 units/µl.

6.1.6 Dynamic metabolic flux analysis

Dynamic MFA was done as described in section 5.2.

6.2 Results and discussion

The capacity for recombinant protein production was compared for two strains: MG1655 and MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37*. Therefore, the commercially available expression plasmid pTrcHisTopoLacZ was transformed in competent cells of MG1655 and MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37*, respectively. Insertion of the plasmid was checked by antibiotic resistance and control restriction digest on purified plasmids with restriction enzymes NcoI and AlwnI (New England Biolabs, The Netherlands) in NEB Buffer 4 (New England Biolabs) at 37 °C. The resulting DNA fragment pattern was checked on a 1.2 % agarosegel in TAE-buffer (50x stock solution, Invitrogen, Merelbeke, Belgium)

6.2.1 Step experiment

To compare the production of recombinant β-galactosidase of *E. coli* MG1655 with and without constitutive phosphoenolpyruvate carboxylase activity and cutting off the acetate pathway, a chemostat experiment at a dilution rate of 0.1 h^{-1} was done for both strains. When

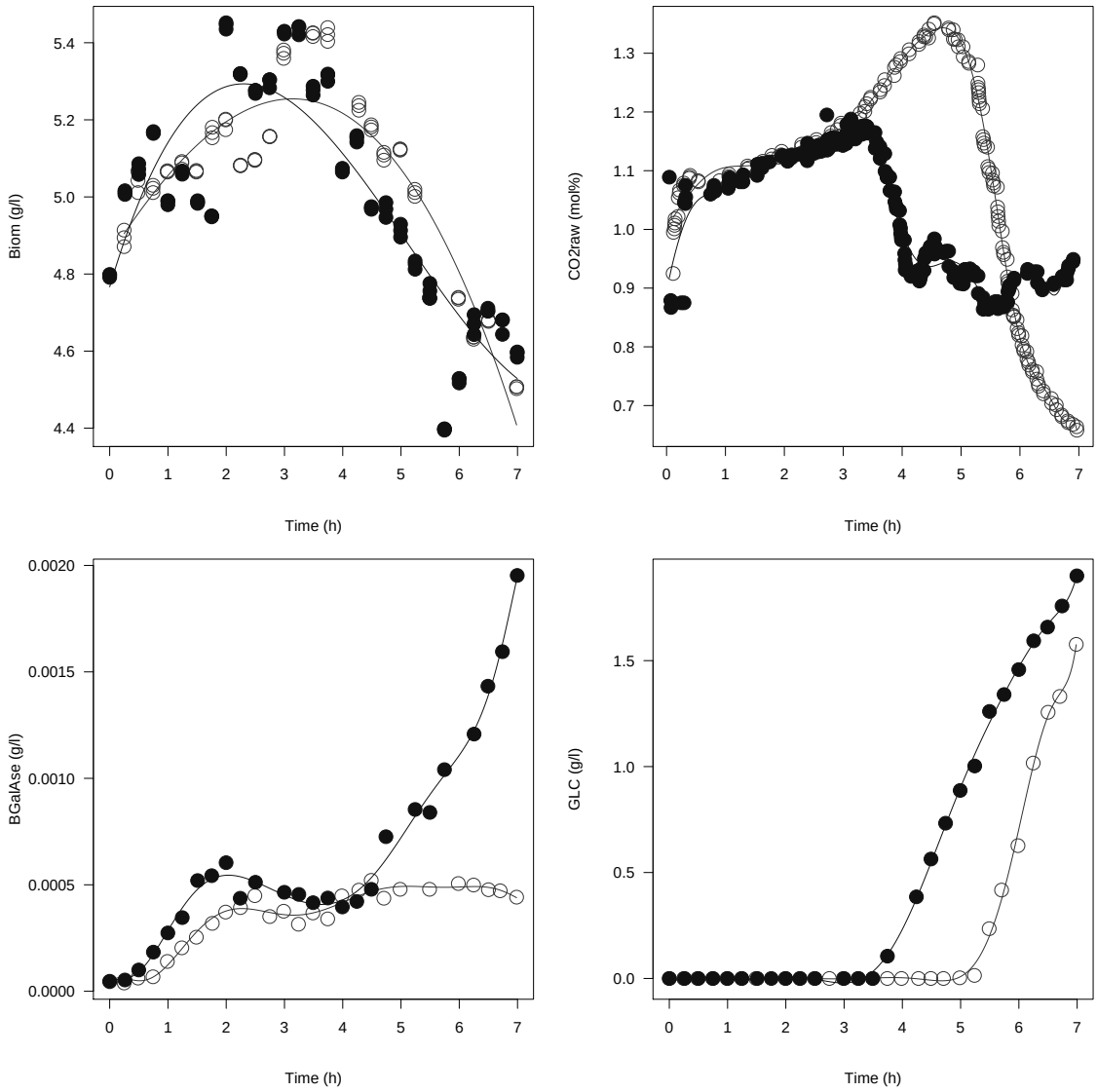


Figure 6.1: Time series for the biomass concentration (based on the OD measurements), % CO_2 in the exhaust gas, β-galactosidase concentration (based on activity measurements) and residual glucose for seven hours during the induction of pTrcHisTopoLacZ in MG1655 (open symbols) and MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* (closed symbols). The lines through the points represent the polynomial fits.

adding IPTG in the feed, to obtain continuous induction, excessive foaming was observed. This was caused by a too high protein production. For this reason, no IPTG was added in the feed and a step experiment was conducted when the steady state phase was reached. During seven hours, the *ptrc* promoter of the pTrcHisTopoLacZ plasmid was induced with IPTG and every 15 min, a sample was taken. In figure 6.1 the evolution of the biomass, % CO_2, intracellular β-galactosidase activity and residual glucose for MG1655 and MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* during the pulse experiment is shown. The extracellular

β-galactosidase activity was also measured in the supernatant of the broth, but no activity was detected.

During the step experiment, the biomass (figure 6.1) follows the same pattern for the two different strains. First, there is a slight increase in biomass concentration followed by a stagnation. Subsequently, there is a second increase followed by a decrease around 4 h after the start of induction. Both strains have a sharp increase in CO_2 production during the first 30 minutes of the induction followed by a gradual increase. After 3.5 h, a difference between the two strains can be noticed in the CO_2 data. While for MG1655 the CO_2 keeps increasing until 5 h followed by an exponential drop to 0.65 %, the CO_2 of MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* drops immediately but not as low as for MG1655 and then fluctuates around 0.9 %

During the first two hours of the step experiment, the cells produce more β-galactosidase reflected by more enzyme activity (figure 6.1). The increase in β-galactosidase is faster and higher in MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* which indicates that this strain is a better recombinant protein producer. Next there is a slight decline and stagnation in the β-galactosidase production. After 3.5 hours, the β-galactosidase production in MG1655 improves somewhat and finally stagnates. β-galactosidase production in MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37*, on the other hand, keeps increasing. At the end, after 7 hours of induction, the β-galactosidase production in MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* is approximately 5 times higher compared to the β-galactosidase production in MG1655. Also a difference in glucose consumption (figure 6.1) for the two different strains can be observed. After 4 h of induction, the glucose consumption of MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* slows down whereas the glucose consumption of MG1655 decreases only after 5 h.

The difference in recombinant protein production between the two strains may have the following explanation. At the start of the step experiment the cells are in optimal condition for recombinant protein production and the production increases. This increase is higher in MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37*, suggesting that the recombinant protein production in this strain is more efficient (higher pool of precursors in the Krebs cycle due to the constitutive *ppc* expression) compared to MG1655. After 2 hours, the optimum production capacity of the cells is reached; the production capacity thereafter declines somewhat and then stagnates. Because MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* grows slower than MG1655 (likely due to the *ppc* overexpression redirecting PEP to oxaloacetate instead of to pyruvate necessary for the PTS) and is more efficient in protein production, this strain more easily adapts to the stress situation of IPTG induction, resulting in a steady increase in β-galactosidase production whereas the β-galactosidase production of MG1655 stagnates. This evolution was confirmed by the Western blot of the intracellular crude cell extract (data not

shown).

6.2.2 Dynamic metabolic flux analysis

To assess if and how metabolic fluxes are affected by the constitutive overexpression of *ppc* and by cutting off the acetate pathway, metabolic modelling can be used. However, MFA is normally only usable in chemostat experiments, *i.e.* when intracellular fluxes do not change over time. In order to compare the dynamics of the fluxes over time during the step experiment, dynamic metabolic flux analysis was used.

The metabolic model used was the same as depicted in figure 4.1 with the β-galactosidase synthesis reaction added (a description of the metabolites can be found in appendix B.1):

BGalAseSYLR: 1.994 ATP + 2.992 H2O + 0.0507 Ala + 0.06197 Arg + 0.05352 Asn + 0.07042 Asp + 0.0169 Cys + 0.04225 Gln + 0.06197 Glu + 0.09296 Gly + 0.04507 His + 0.03662 Ile + 0.09296 Leu + 0.01972 Lys + 0.0338 Met + 0.04225 Phe + 0.04507 Pro + 0.05634 Ser + 0.05352 Thr + 0.02254 Trp + 0.03099 Tyr + 0.07042 Val + 1.994 GTP $\longrightarrow$ 1.994 ADP + 3.989 PiOH + 1.994 GDP + BGalAse

The model contains 138 reactions and 152 metabolites of which 12 were considered exchangeable with the environment: NH_3, PiOH, Biom, GLC, Lac, OAA, Suc, O_2, CO_2, H_2O, H_2SO_4 and BGalAse. Seven metabolites were measured: Biom, GLC, Lac, Suc, O_2, CO_2 and BGalAse. There are 144 independent equations in the model and 138 + 12 = 150 unknowns. Thus at least 6 measurements should be performed to solve the model. In this case there were 7, thus measurements could be balanced (see section 2.5). Only one polynomial through the whole range of data points was fitted, since this was sufficient to capture the dynamics in the measurements. Figure 6.1 shows the polynomials fitted to some of the data.

Figure 6.2 draws the dynamics in the relative fluxes of the central metabolism, the fermentation pathways and the recombinant production pathway during the step experiment. As expected, the initial flux through PTS was smaller in *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37*. However, one hour after induction, the PTS flux was higher in *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37*, except for two hours between $4 - 6$ h.

The flux towards fermentation products (acetate and lactate) was initially smaller in *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37*, but followed roughly the same dynamics in both strains until 4 h of induction. From that point on, the flux towards fermentation products was maintained in *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* whereas it first increased drastically in the wild-type *E. coli* and then seemed to decrease again.

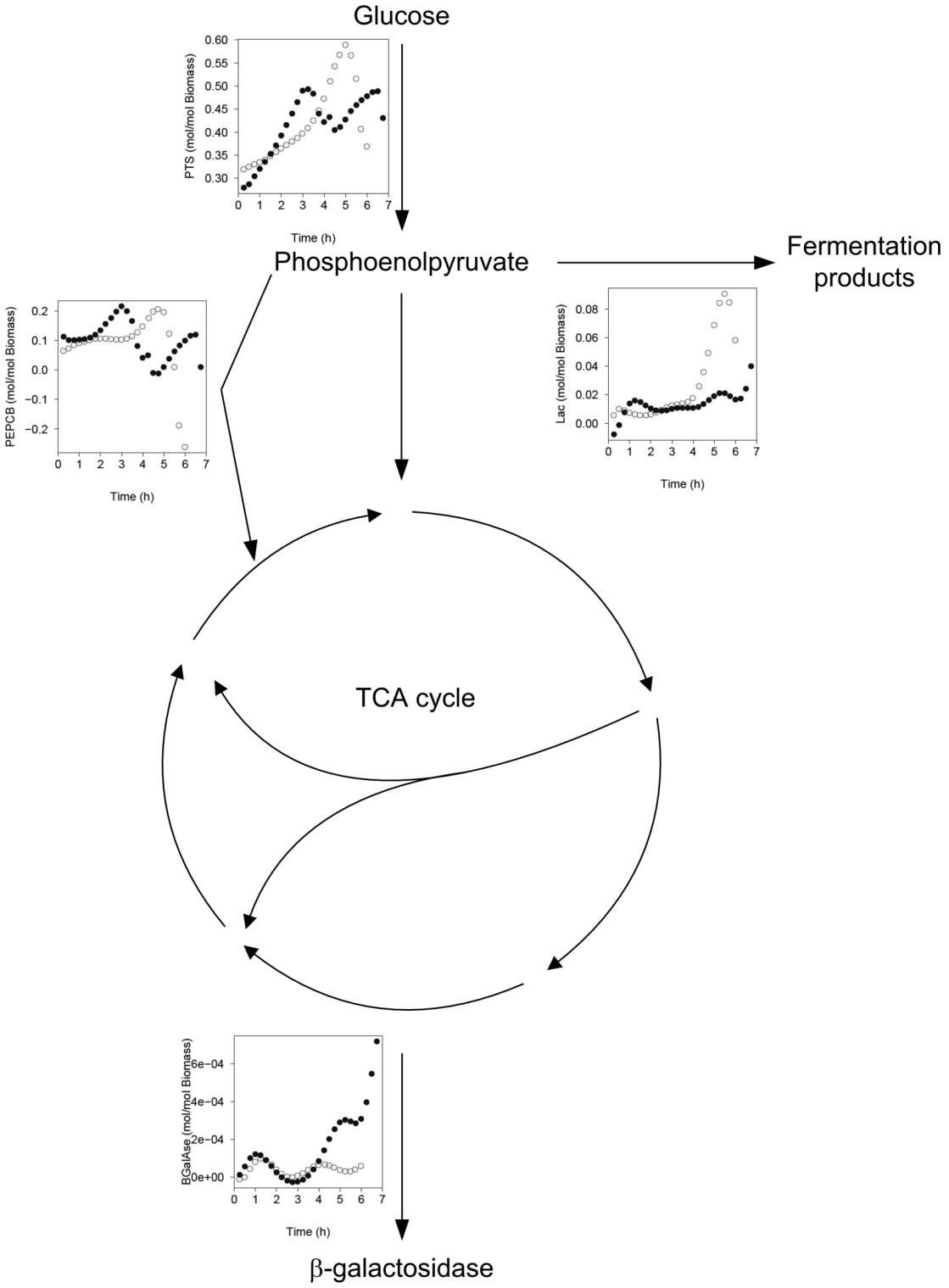


Figure 6.2: The dynamics in relative fluxes (each flux is divided by the biomass flux) of the central carbon metabolism, the fermentative pathways and the recombinant protein production pathway. The fluxes for MG1655 are represented by open symbols) and those for MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* by closed symbols.

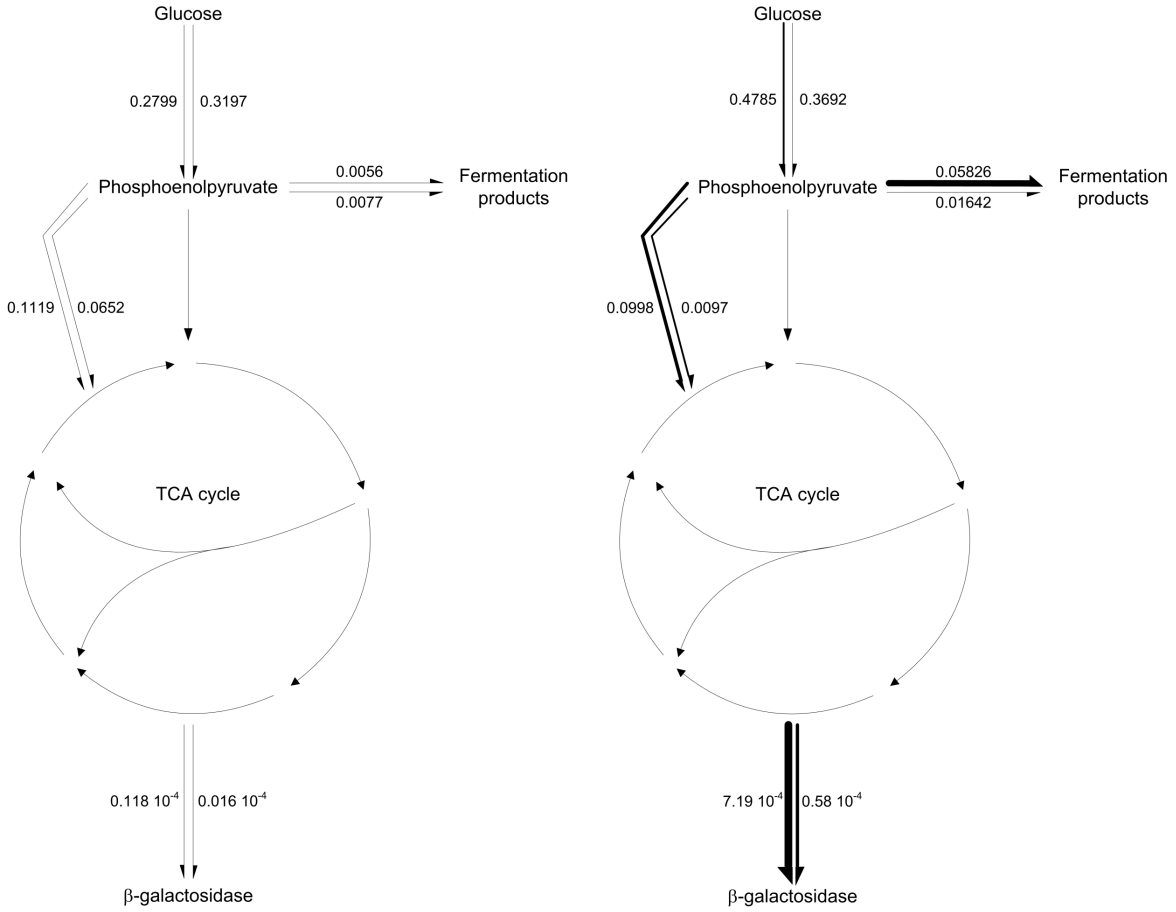


Figure 6.3: Relative flux map of the central carbon metabolism, fermentative pathways and the recombinant protein production pathway for *E. coli* MG1655 (right/top fluxes) and *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* (left/bottom) fluxes at the beginning of the step experiment (left figure) and at the end of the 7 hours induction (right figure).

Overall, the flux through PEP carboxylase is higher in *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* than in the wild-type. The flux towards the *β*-galactosidase followed the same dynamics for both strains during the first four hours of induction. From that point on, the flux towards *β*-galactosidase was maintained in the wild-type, whereas it increased quickly in *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37*. This indicates that the overexpression of an anaplerotic pathway (PEP carboxylase) has a positive effect on recombinant protein production.

Figure 6.3 depicts the relative fluxes in the central metabolism at the start of the step experiment and at the end of induction. The relative flux towards fermentation products is higher (3.6 times) in the wild-type E. coli than in MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* after circa 7 hours of induction. In MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37*, on the other hand, the flux through PEP carboxylase towards the Krebs cycle is significantly higher,

supplying the precursor pools for protein production. Consequently, the flux towards β-galactosidase is 12 times larger in the MG1655 $\Delta ackA$–pta, $\Delta poxB$, Δpppc ppc–$p37$ mutant than in the wild-type.

6.3 Conclusions

To improve our understanding of recombinant protein synthesis in *E. coli*, we analysed the production of a model recombinant protein, β-galactosidase, in response to the constitutive overexpression of an anaplerotic reaction (PEP carboxylase). To compare the production of recombinant β-galactosidase of *E. coli* MG1655 with and without constitutive phosphoenolpyruvate carboxylase activity and cutting off the acetate pathway, a chemostat experiment at $D = 0.1\,h^{-1}$ was performed for both strains prior to a step experiment. During 7 hours, the *ptrc* promoter of the pTrcHisTopoLacZ plasmid was induced with IPTG. In order to compare the dynamics of the fluxes over time during the step experiment, dynamic metabolic flux analysis was applied. This revealed that the relative flux towards fermentation products was higher in the wild-type *E. coli* than in the MG1655 $\Delta ackA$–pta, $\Delta poxB$, Δpppc ppc–$p37$ mutant. Furthermore, in the latter one, the flux towards β-galactosidase was significantly higher, resulting in five times more protein activity.

These results confirm our premise that by improving the metabolic capacity to replenish the TCA intermediates, recombinant protein production could be improved. Similar results were obtained by March *et al.* (2002): to replenish the TCA cycle intermediates, the authors expressed the anaplerotic enzyme pyruvate carboxylase of *Rhizobium etli*.

Part III

Oscillations – mRNA arrays

Chapter 7

Oscillating environmental oxygen conditions

7.1 Influence of oxygen on metabolism

The advent of oxygen is relatively recent in the evolution of life (Webster, 2003). The first 2 billion years of evolution happened in an anaerobic environment. Then, two billion years ago, the increase of photosynthesis by the cyanobacteria gradually changed the atmosphere with methane, nitrogen and ammonia as major components, to what we have now. Thus oxygen utilisation enzymes were relatively recently introduced in bacteria, hence their ability to grow, or at least survive, in anoxic environments. Eukaryotes originated when oxygen was present in the atmosphere and for them oxygen became an essential molecule to survive. Even yeast needs some essential nutrients (*e.g.* methionine) when grown anaerobically, because of oxygen-dependent steps in the biosynthesis of those metabolites.

Procaryotes conserved the ability to function without oxygen. They have a complex regulation mechanism on the genetic and metabolic level that switches on aerobic metabolism whenever O_2 is present and goes back to the ancestral pathways of anaerobic growth when oxygen is absent. In this work, the effect of oscillating between aerobic and anaerobic conditions was investigated at the level of the transcriptome and the metabolome. The model organism was *E. coli*. Two different periods for the oscillations were chosen: one of 4 minutes and one of 30 minutes. Half of the time, air was blown in the culture, the other half a mixture of nitrogen and carbon dioxide. The fluctuations of the concentrations of external metabolites (fermentation products) over one such aerobic/anaerobic sequence, was measured. The changes in transcription levels were investigated using Affymetrix microarrays.

7.1.1 Electron acceptors in *E. coli*

E. coli, being a procaryote, can use several electron acceptors (reviewed in Bunn & Poyton, 1996). When growing under fully aerobic conditions, cytochrome o oxidase (encoded by *cydABCDE*) is active. When oxygen starts to be limiting, cytochrome d oxidase (encoded by *cyoAB*) is used. Cytochrome o has a high V_{max}, but a low affinity for oxygen while cytochrome d has a low V_{max} but a high affinity for oxygen, a classic specialisation in *E. coli*. Aerobically grown cells contain around 300 molecules cytochrome o and 200 molecules cytochrome d. During anaerobic or microaerobic growth, a 140-fold repression of expression of cytochrome o is observed, while the expression of cytochrome d increases 3-fold (Cotter *et al.*, 1990). Thus, in the presence of oxygen both cytochromes are available in the cells. However, cytochrome d does not contribute to the H^+ gradient (actually, high oxygen content inhibits cytochrome d), but is probably involved in the protections of the cells against reactive oxygen species (Bunn & Poyton, 1996).

In the absence of oxygen, *E. coli* can produce up to five alternative oxidoreductases (Bunn & Poyton, 1996): fumarate reductase (encoded by *frdABCD*), two nitrate reductases (encoded by *narGHJI* and *narZYWV*), dimethyl sulfoxide (DMSO) and trimethylamine-N-oxide (TMAO) reductase (encoded by *dmsABC*) and TMAO reductase (encoded by *torA*). TMAO reductase has a much more narrow substrate specificity than DMSO/TMAO reductase.

Only *cyoABCDE* is fully expressed in the presence of oxygen. Operon *cydAB* is fully expressed under microaerobic conditions while the expression of *narGHJI*, *frdBCD* and *dmsABC* increases under anaerobic conditions even in the absence of an alternative electron acceptor. Nitrate increases the expression of *narGHJI* even more and decreases the expression of *frdABCD* and *dmsABC* (Bunn & Poyton, 1996). Three regulatory networks are responsible for the transcriptional regulation of those electron acceptors (Gunsalus, 1992). The NarX, NarL, and NarQ regulator network controls respiratory gene expressions in response to nitrate availability. Aerobic-anaerobic control is done by the Fnr and the ArcAB regulatory system, which will be discussed in more detail below.

7.1.2 Regulation of the respiration metabolism

The main regulators in the oxygen response in the *E. coli* metabolism are the two component anoxic redox control (Arc) system, and the fumarate and nitrate reduction (Fnr) protein (figure 7.1). The two components of the Arc system are the membrane-bound redox sensor ArcB, whose activity is inhibited by oxidised quinones (Georgellis *et al.*, 2001), and ArcA, responsible for the signal transduction to the genetic regulation (Green & Paget, 2004). When oxygen starts to be scarce and thus also oxidised quinones, ArcB autophosphorylates and then transphosphorylates ArcA (Matsubara & Mizuno, 2000). Phosphorylated ArcA represses expression of many genes involved in the respiratory metabolism. For example, the Krebs cycle

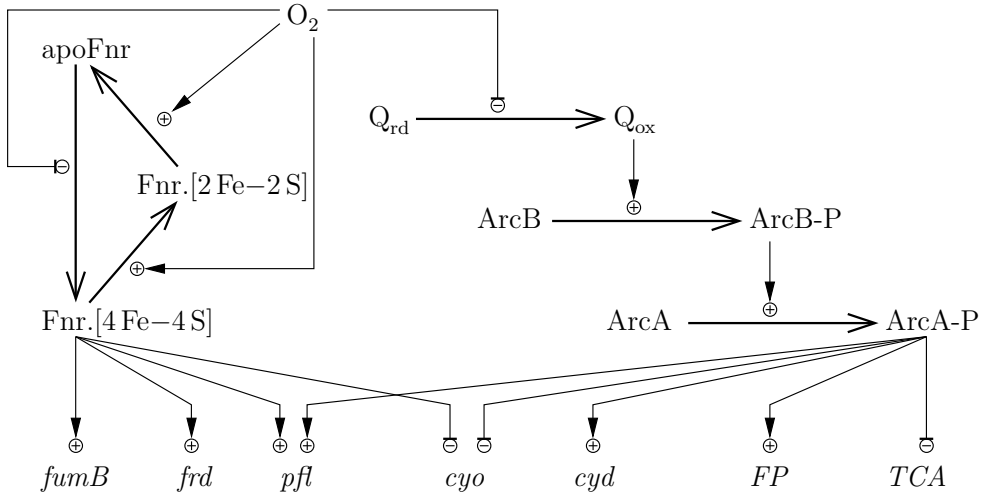


Figure 7.1: The influence of ArcA and Fnr on the transcription of different genes from the central carbon pathways. Q_{rd}: reduced quinone; Q_{ox}: oxidised quinone; *fumB*: fumarate hydratase; *frd* fumarate reductase; *pfl*: pyruvate formate lyase; *cyo*: cytochrome o; *cyd*: cytochrome d; *FP*: genes from the fermentative pathways; *TCA*: genes from the Krebs cycle.

is interrupted (Shalel-Levanon *et al.*, 2005b), the glyoxylate shunt is inactivated (Georgellis *et al.*, 2001) and *cyoABCDE* (coding for cytochrome o) is not expressed anymore (Bunn & Poyton, 1996). At the same time, ArcA also activates expression of other genes: fermentative pathways (Georgellis *et al.*, 2001), *cydAB* coding for cytochrome d and *pfl* coding for pyruvate formate lyase. ArcA is also involved in conjugal fertility functions (Spiro & Guest, 1991).

Unlike ArcB, Fnr is a direct O_2 sensor (Khoroshilova *et al.*, 1997). It is similar to the cyclic AMP receptor protein (CRP) but has no cAMP binding domain (Spiro, 1994). Fnr consists of two domains, a carboxy-terminal DNA-binding region for binding to specific targets and an amino-terminal sensory domain containing four cysteines binding the Fe-S clusters that function as direct oxygen sensors (Green & Paget, 2004). The active form of the Fnr regulator is a homodimer with each Fnr molecule containing a [4 Fe−4 S] cluster. In the presence of oxygen, the [4 Fe−4 S] cluster is converted to a [2 Fe−2 S] cluster (probably via a [3 Fe−4 S] cluster, releasing H_2O_2, Green *et al.*, 1996; Crack *et al.*, 2004), resulting in inactivation of Fnr (Khoroshilova *et al.*, 1997). Further incubation with oxygen results in the conversion of [2 Fe−2 S].Fnr to apoFnr (Achebach *et al.*, 2005; Sutton *et al.*, 2004b). The total amount of Fnr in a cell does not vary much (Sutton *et al.*, 2004a). Under aerobic conditions, Fnr occurs in the cell as a monomer in the apo form, with cysteine disulfide bridges. When O_2 is removed, the apoFnr is converted to a [4 Fe−4 S].Fnr dimer (Popescu *et al.*, 1998; Bates *et al.*, 1995). However, there is a lag phase due to the cysteine disulfides that have to be reduced for incorporation of the [4 Fe−4 S] cluster. This lag phase varies from less than 100 minutes (Tran *et al.*, 2000) to more than 200 minutes (Achebach *et al.*, 2004). Addition of

glutaredoxins greatly decreases the lag phase to less than one hour (Achebach *et al.*, 2004).

Many kinetic studies exist that track the time needed for inactivation of [4 Fe−4 S].Fnr. An overview of the different values reported in literature can be found in Achebach *et al.* (2005) who demonstrated that *in vitro* conversion of [4 Fe−4 S].Fnr to apoFNR requires around 5 minutes. This is rather similar to the 4.5 minutes needed for inactivation of [4 Fe−4 S].Fnr found by Sutton *et al.* (2004a) for *in vivo* conditions while *In vitro*, 3.4 minutes were needed to convert [4 Fe−4 S].Fnr to [2 Fe−2 S].Fnr.

The expression of one third of the genes of *E. coli* are altered when changing from aerobic to anaerobic conditions. The expression levels of most of those genes are directly or indirectly influenced by ArcA and 50 % of those genes are regulated by Fnr (Salmon *et al.*, 2005, 2003). There seems to be common functions between Fnr and Lrp (leucine responsive regulatory protein): 25 % of the genes regulated by Fnr are also regulated by Lrp. Furthermore, *lrp* expression increases five-fold in fnr^- strains (Salmon *et al.*, 2003; Hung *et al.*, 2002).

When switching an aerobic culture to microaerobic conditions, a peak in Fnr is observed after five minutes, after which ArcA starts to increase and reaches steady state 10 minutes later (Partridge *et al.*, 2007). This is consistent with the fact that Fnr activates the transcription of *arcA* (Compan & Touati, 1994) and that ArcA represses *fnr* expression under microaerobic conditions (Shalel-Levanon *et al.*, 2005a). At the same time, Pdhc (pyruvate dehydrogenase complex) is partially replaced by Pfl (pyruvate phosphate lyase). Pfl is inactivated by O_2, therefore the expression of *yfiD* also increases. YfiD repairs oxygen inactivated Pfl under microaerobic conditions (Zhu *et al.*, 2007). Furthermore, the increased expression of cytochrome d (the cytochrome having a high affinity for O_2) scavenges oxygen that would otherwise damage Pfl (Alexeeva *et al.*, 2002). Simultaneous activity of Pfl and Pdhc was also observed by Alexeeva *et al.* (2000).

To split the Krebs cycle, *sucCD* is repressed so that succinyl-CoA is no longer converted to succinate. Hence, two branches are created. One branch uses fumarate as electron acceptor and produces succinate, while the other branch is just active enough to produce alfaketoglutaric acid, needed as precursor for amino acid production (Cox *et al.*, 2006). Indeed, ArcA represses *gltA* (citrate synthase), *icd* (isocitrate dehydrogenase), and *sdhC* (converting succinate to fumarate) (Shalel-Levanon *et al.*, 2005b). Transcription of fumarate reductase (*frd*) is activated, hence fumarate can function as electron acceptor and succinate is produced.

Adding oxygen to anaerobically cultivated cells, results within 5 minutes in decreased transcription of genes associated with anaerobic metabolism and increases transcription of genes related to aerobic metabolism (Partridge *et al.*, 2006). Genes for the pyruvate dehydrogenase complex were transcribed 5 minutes after the switch, while expression of *sucCD* only occurs after 10 minutes. Also expressed within 5 minutes after the switch, is *dctA*, coding for an acid

symporter, *lctPRD*, coding for a lactate transporter, and *kdgT*, coding for an oxoglutarate transporter. This is probably meant for use in the Krebs cycle the first ten minutes, as this cycle is not yet completely functional then.

Until recently, it was believed that ArcA did not have an influence on fully aerobic and fully anaerobic cultures; ArcA was the regulator of microaerobic conditions, while Fnr played its role under more anaerobic conditions (Shalel-Levanon *et al.*, 2005c; Alexeeva *et al.*, 2003, 2000). This is however in contrast with the results of Partridge *et al.* (2007) who showed an Fnr peak activity before an increase in ArcA, when switching an aerobic culture to microaerobic conditions. Furthermore, Perrenoud & Sauer (2005) demonstrated that ArcA also influences fully aerobic and fully anaerobic cultures: *arcA* knock-out mutants had a 60 % increased TCA flux, compared to the wild-type.

7.2 Experimental setup

To investigate the effects of oscillating oxygen input concentrations on the cells, four kinds of experiments were run: two steady state experiments, one aerobic and one anaerobic, and two experiments in which the gas flow oscillated between a nitrogen oxygen mixture (air) and a nitrogen/carbon dioxide mixture (nitrogen with 3 % CO_2). One oscillation was run with a period of 4 minutes and one with a period of 30 minutes. For each, the period was divided in two equal parts. In the first part, the reactor was aerated with air, and in the second part the nitrogen CO_2 mixture was blown through it. The experiments were run at a dilution rate of $0.13\,h^{-1}$ (with a standard deviation of $0.02\,h^{-1}$).

Sampling for mRNA, extracellular metabolites and OD was done at least five residence times after the reactor was set in steady state operation or oscillating mode. For the experiments with oscillating gas concentrations, the reactor was sampled at different moments in the same period, during two consecutive periods. Table 7.1 contains the detailed mRNA sampling times.

7.3 Materials and methods

7.3.1 Bacterial strain

Escherichia coli MG1655 [λ^-, F^-, *rph*–1] was obtained from the Coli Genetic Stock Center (CGSC). It was explicitly checked to not have the *fnr* deletion, as some strains with this name have it (Soupene *et al.*, 2003).

Table 7.1: Properties of the different mRNA samples. The Ss stands for steady state; P04, P30: oscillation experiment with a period of 4 respectively 30 minutes; Ae: aerobic phase; An: anaerboib phase; the number at the end of the SampleID indicates when in the aerobic/anaerobic phase the sample was taken; The ArrayID is the same as the SampleID, except that the replicate number is added.

ArrayID	SampleID	ExpID[1]	Period	STiP[2]	Phase	Replicate
SsAe_1	SsAe	SsAe	Ss	-	Aerobic	1
SsAe_2	SsAe	SsAe	Ss	-	Aerobic	2
SsAn_1	SsAn	SsAn	Ss	-	Anaerobic	1
SsAn_2	SsAn	SsAn	Ss	-	Anaerobic	2
P04Ae02_1	P04Ae02	P04	4	2	Aerobic	1
P04An02_1	P04An02	P04	4	4	Anaerobic	1
P04Ae02_2	P04Ae02	P04	4	2	Aerobic	2
P04An02_2	P04An02	P04	4	4	Anaerobic	2
P30Ae02_1	P30Ae02	P30	30	2	Aerobic	1
P30Ae04_1	P30Ae04	P30	30	4	Aerobic	1
P30Ae08_1	P30Ae08	P30	30	8	Aerobic	1
P30Ae15_1	P30Ae15	P30	30	15	Aerobic	1
P30An02_1	P30An02	P30	30	17	Anaerobic	1
P30An04_1	P30An04	P30	30	19	Anaerobic	1
P30An08_1	P30An08	P30	30	23	Anaerobic	1
P30An15_1	P30An15	P30	30	30	Anaerobic	1
P30Ae02_2	P30Ae02	P30	30	2	Aerobic	2
P30Ae04_2	P30Ae04	P30	30	4	Aerobic	2
P30Ae08_2	P30Ae08	P30	30	8	Aerobic	2
P30Ae15_2	P30Ae15	P30	30	15	Aerobic	2
P30An02_2	P30An02	P30	30	17	Anaerobic	2
P30An04_2	P30An04	P30	30	19	Anaerobic	2
P30An08_2	P30An08	P30	30	23	Anaerobic	2
P30An15_2	P30An15	P30	30	30	Anaerobic	2

[1]ID of the experiment.

[2]Sampling time in the period. Expressed in minutes.

7.3.2 Culture conditions

Media

The Luria Broth (LB) medium consisted of 1 % triptone peptone (Difco, Erembodegem, Belgium), 0.5 % yeast extract (Difco) and 0.5 % sodium chloride (VWR, Leuven, Belgium).

The minimal medium contained 6.75 g/l NH_4Cl, 1.25 g/l $(NH_4)_2SO_4$, 1.15 g/l KH_2PO_4, 0.5 g/l NaCl, 0.5 g/l $MgSO_4 \cdot 7\,H_2O$, 16.33 g/l glucose $\cdot\, H_2O$, 1 ml/l vitamin solution, 100 µl/l molybdate solution and 100 µl/l selenium solution. Vitamine solution contained 3.6 g/l $FeCl_2 \cdot 4\,H_2O$, 5 g/l $CaCl_2 \cdot 2\,H_2O$, 1.3 g/l $MnCl_2 \cdot 2\,H_2O$, 0.38 g/l $CuCl_2 \cdot 2\,H_2O$, 0.5 g/l $CoCl_2 \cdot 6\,H_2O$, 0.94 g/l $ZnCl_2$, 0.0311 g/l H_3BO_4, 0.4 g/l $Na_2EDTA \cdot 2\,H_2O$ and 1.01 g/l thiamine $\cdot$ HCl. The molybdate solution contained 0.967 g/l $Na_2MoO_4 \cdot 2\,H_2O$. The selenium solution contained 42 g/l SeO_2. Anaerobically *E. coli* did not grow without selenium in the medium. Selenium is needed for anaerobic growth, probably as cofactor for pyruvate formate lyase (Enoch & Lester, 1972; Lester & Demoss, 1971).

For flask culture medium, glucose and $MgSO_4$ was separately autoclaved (121 °C, 20 minutes) from the salts and combined with the filter-sterilised (pore size 0.2 µm, Millipore bottletopfilter, Millipore, Antwerpen, Belgium) vitamine, molybdate and selenium solutions, after cooling down. Before autoclaving, the pH of the salts solution was set at 7 with 1 M K_2HPO_4.

All components for the medium of the continuous culture were dissolved and filter-sterilised (pore size 0.22 µm, Sartobran, Sartorius, Belgium). The pH was left at approximately 5.4.

Cultivations

A preculture from a single colony on a LB-plate was started in 5 ml LB medium during 8 hours at 37 °C on an orbital shaker at 200 rpm. From this culture, 2 ml was transferred to 100 ml minimal medium in a 500 ml shake flask, and incubated for 16 hours at 37 °C on an orbital shaker at 200 rpm. Sixty milliliter of this culture was used to inoculate the reactor, a 2 l Biostat B culture vessel with 1.5 l working volume (Sartorius-BBI systems, Melsungen, Germany). The culture conditions were: 37 °C, stirring at 800 rpm, gas flow rate of 1.5 l/min. The pH was maintained at 7 with 0.5 M H_2SO_4 and 4 M KOH. The exhaust gas was cooled down to 4 °C by an exhaust cooler (Frigomix 1000, Sartorius-BBI, Melsungen, Germany). A continuous feed of 4 ml/h 10 % solution of silicone antifoaming agent (BDH 331512K, VWR Int Ltd., Poole, England) was added to the culture vessel.

Gas composition during the experiments in which the culture conditions were oscillated between anaerobic and aerobic conditions, was controlled by the MFCS/win software (Sartorius-BBI systems, Melsungen, Germany). To maintain the same mixing conditions between the two phases of the oscillation, the gas flow had to be constant. Therefore during the anaerobic phase, a mixture of nitrogen gas (97.7 %) and CO_2 (2.3 %), to counter the stripping action of the flushing gas, was blown through the reactor. During the aerobic phase, air was used as aeration gas.

From figure 7.2 it can be seen that some time is needed (around 10 seconds) to get the reactor broth to aerobic or anaerobic conditions when the gas mixture is changed. Two

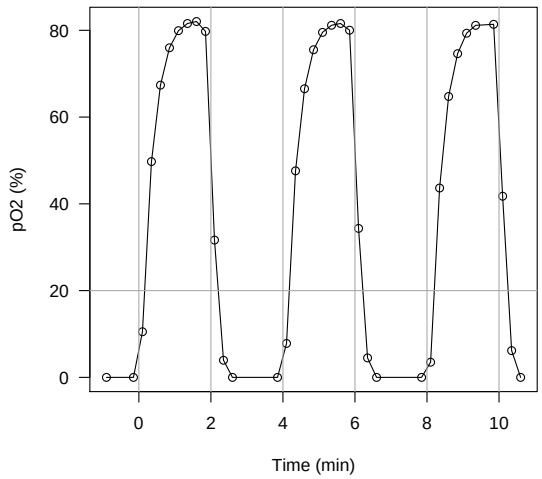


Figure 7.2: Illustration of the responsiveness of the pO_2 probe to changes in the gas composition blown through the reactor. At 0, 4 and 8 minutes, the MFCS/win controller switches to air. At 2, 6 and 10 minutes, the anaerobic phase is started by blowing a mixture of nitrogen and CO_2 through the vessel.

minutes is probably the shortest time that can be taken between each switch. Already now, the cells are less time in fully anaerobic conditions than in aerobic and micro-aerobic ones. As biomass continues to consume oxygen when stopping the oxygen supply, one would expect that anaerobiosis is faster attained that aerobiosis when starting the oxygen flow after the anaerobic phase. However, figure 7.2 shows that the 20 % level is a little bit faster attained when going from anaerobiosis to aerobiosis than the other way around. But getting the high oxygen concentration levels when entering the aerobic phase, takes more time than loosing them when going into the anaerobic phase.

7.3.3 Analytical methods

Online measurements, the sampling method of the reactor, the cell density measurement method and HPLC analysis, were done as described in section 4.1.3.

RNA samples were taken with the stainless steel cold beads method as discribed before (section 4.1.3). These samples were centrifuged in a Heraeus biofuge stratos (Thermo Electron LED GmbH, Langenselbold, Germany) at 15.000 rpm in eppendorf tubes at 4 °C. The supernatant was discarded and the pellet was washed with PBS buffer (Merck, Overijse, Belgium) and centrifuged at the same conditions as the first time. Finally the pellets were resuspended in 100 µl of PBS buffer and 500 µl RNAlater reagent (Qiagen, Venlo, The Netherlands).

An RNeasy midi kit (Qiagen) was used for RNA extraction. Purity was then confirmed with a Nanodrop (Nanodrop technologies, France). The ratio of absorbance at 260/230 and 260/280 nm was measured to ensure the absence of contaminating DNA, proteins and salts. The

extraction can be assumed successful when both ratios lay between 1.8 and 2. The purity of the RNA was further controlled on a FA-agarose gel, as advised by Qiagen.

Microarray analyses were performed by the VIB MicroArrays Facility (`www.microarrays.be`). Affymetrix GeneChip E. coli Genome 2.0 arrays were used. Total RNA was checked for its integrity and purity using Agilent Bioanalyzer and Nanodrop spectrophotometer, respectively. Probes were prepared from 10 µg total RNA, showing no signs of degradation or impurities (260/230 and 260/280 absorbance ratios > 1.8), according to Affymetrix's guidelines. Briefly, total RNA was supplemented with Poly-A RNA controls (P/N 900433, Affymetrix, UK) and reverse transcribed to cDNA using a random primer. After RNA degradation with NaOH, cDNA was purified (MinElute PCR Purification Kit, Qiagen, P/N 28004), and analysed again for yield ($30 - 120$ µg) and purity (260/230 and 260/280 absorbance ratio > 1.8). The cDNA was then fragmented using Dnase-I (P/N 27-0514-01, Amersham Biosciences) and terminally labeled with biotinylated GeneChip DNA Labeling Reagent (P/N 900542, Affymetrix, UK). A minimum of 1 µg fragmented probe was resuspended in 80 µl hybridisation buffer containing 3 nM control oligo B2 (P/N 900301, Affymetrix, UK) and hybridised in a rotisseri oven at 45 °C for 16 hours at 60 rpm. The genechips were washed and stained in the GeneChip Fluidics Station 400 (Affymetrix, UK) using Mini_prok2v1 protocol, and subsequently scanned with the GeneChip Scanner 3000 (Affymetrix, UK). Image analysis was performed in GCOS.

The data analysis of the microarray results was done in R (R Development Core Team, 2006) using toolboxes of the bioconductor project (Gentleman *et al.*, 2004). The affy package (Irizarry *et al.*, 2003a; Gautier *et al.*, 2004) was used to import the CEL files into R. CEL files are the result of the conversion of the raw images obtained from the array scanner, to probe intensity values (Bolstad *et al.*, 2005).

7.4 mRNA arrays: Methods and results

Affymetrix GeneChips are photochemically synthesised (Fodor *et al.*, 1991) arrays of short nucleotide probes of around 20 bases (Lockhart *et al.*, 1996). Each gene is represented by 10 to 20 probe pairs per gene (Nguyen *et al.*, 2002). A probe pair consists of one oligonucleotide that matches perfectly (PM) and one that has a single mismatch for the central base in the oligonucleotide (MM). Different arrays for different organisms exist. In this study, the *E. coli* array v2 from Affymetrix was used. Biotin labeled cRNAs are hybridised to the array. The degree of matching is assessed using laser confocal fluorescence scanning (Fodor *et al.*, 1993). The obtained image is then converted to intensity values for each PM and MM probe. Different steps are needed for converting those probe intensity values to expression values for the different genes. These steps are explained in the next section.

7.4.1 Extracting expression levels from probe intensities

Each GeneChip array has only one sample hybridised to it. Thus, to compare different experimental setups, the intensities of the different arrays have to be scaled and normalised before (relative) gene expressions can be computed. Four steps can be discerned in turning probe intensity data into expression measures (Gautier *et al.*, 2004):

1. Background correction. This is done for each chip separately and corrects for the gradients and differences of hybridisation that can occur in the different chips. In the MAS software (provided by Affymetrix), the array is split in a number of rectangular zones (typically 16). In each zone, the lowest 2 % of the probes is chosen as the background for that zone. Some smoothing is done with the background in the other zones. This corrected value is subtracted from each probe intensity value if the remaining is still not smaller than zero (Affymetrix, 2002). The Robust Multiarray Average algorithm (RMA, Irizarry *et al.*, 2003b) subtracts a background value based on modelling of the PM signal intensities.

   As the sequence of the probes has influence on the hybridisation, background modelling techniques that incorporate those probe sequences were developed (Wu *et al.*, 2004; Zhang *et al.*, 2003).

2. Normalisation. To be able to compare different GeneChips, normalisation between chips is needed. It is one of the most important steps to get meaningful results from the GeneChips (Schadt *et al.*, 2000). Normalisation is done for all probes on an array. PM and MM probes are not treated separately. Different methods for normalisation can be used.

   Initially, the method used by Affymetrix for scaling different arrays, assumes that the difference between two arrays is linear and has an intercept of zero (Schadt *et al.*, 2000). This assumption is not always valid. The distribution in the low-intensity signals behaves differently than the distribution of the high-intensity signals (Schadt *et al.*, 2000). As dividing the arrays into two intensity blocks, where a linear regression can be performed in each block, does not keep the normalisation curve smooth, Schadt *et al.* (2000) propose smoothing splines. However, Li & Wong (2001b) had no problem running piecewise median lines for normalisation.

   Cyclic loess and contrast based methods make use of the M versus A plots (Bolstad *et al.*, 2003). In two colour channel experiments, M is the difference in log expression values and A is the average of the log expression values (Dudoit *et al.*, 2002). For one channel arrays, instead of the two colours, two arrays can be used. For normalisation of multiple arrays, all pairwise combinations are looked at.

Quantile normalisation transforms the data such that the distribution of probe intensities for each array are the same (Bolstad *et al.*, 2003). This is the method used by the RMA algorithm (Irizarry *et al.*, 2003b).

The vsn methodology combines background correction and normalisation in one single step (Huber *et al.*, 2002, 2003). Typically the (log-)ratio, is used to access the difference in expression between two experimental setups. However, weakly expressed genes can have high fold changes that are actually not significant because the variance for the expression values for lowly expressed genes is much higher than for highly expressed genes. The vsn methodology makes use of the generalised log-ratio which coincides with the usual log-ratio when the expression values are large, but shrinks towards zero when both expression values are small. This transformation is used for normalisation, and normalised expression values are obtained with a constant variance across the whole range of expression levels (hence Variance Stabilising normalisation, Huber *et al.*, 2003).

Bolstad *et al.* (2003) demonstrated that the quantile normalisation is the preferred method as it gives the best results in terms of speed and for variance and bias criteria. They also concluded that methods using a baseline array should be avoided. However, that study focused on normalisation and did not include the vsn methodology.

3. PM correction with MM. Mismatch probes are included on Affymetrix arrays to quantify non-specific and cross-hybridisation. In the first version of the Affymetrix software, MM signals were simply subtracted from the PM values. However, sometimes the MM signals are higher than the PM ones giving complications when taking the logarithm. MM signals being higher than PM signals occurs mostly in lowly expressed genes, the reason being that MM probes only lower the binding affinity for PM and not for random sequences. Hence, in lower expressed genes most of the signal in PM probes is due to background noise, making not much difference with the MM probes (Zhang *et al.*, 2003). Therefore, in the new version of the Affymetrix software, idealised MM values are used (Gautier *et al.*, 2004). Anyway, the usefulness of background subtraction is questioned (Li & Wong, 2001b) and in many cases this step is omitted (Gautier *et al.*, 2004) as background subtraction indeed increases accuracy but comes with a high cost in precision (Huber *et al.*, 2005).

4. Computation of expression values from probe intensities. Each gene on the Affymetrix array is represented by 11-20 probes. Many methods exist to summarise probe intensities to gene expression values. Simply taking the mean value of the different probes does not give satisfying results (Li & Wong, 2001a; Irizarry *et al.*, 2003b). MAS 5 (the software from Affymetrix to analyse microarrays), is based on Tukey's biweight (Gautier *et al.*, 2004). The RMA algorithm (Irizarry *et al.*, 2003b) fits a linear model to each probe set

using a robust procedure, typically median polish (Holder *et al.*, 2001). Other model-based approaches are described in Li & Wong (2001a) and Sásik *et al.* (2002).

The RMA methodology (Robust Multiarray Average, Irizarry *et al.*, 2003b) is the most used combination of techniques to generate expression values out of probe intensities. It consists of the following steps: background correction based on the modelling of the PM signals, normalisation of the arrays using quantile normalisation (introduced by Bolstad *et al.*, 2003), no PM correction, and the so-called median polish summarisation method. However, in a comparative study using qRT-PCR (reverse-transcription polymerase chain reaction), Qin *et al.* (2006) concluded that RMA was not the most optimal methodology. They advise to use gcRMA (which uses sequence-based background adjustment, Wu *et al.*, 2004) or if that is not available, MAS 5 (the algorithm used by Affymetrix software) or the dChip mismatch model (Li & Wong, 2001a). This concords with the findings of Choe *et al.* (2005): MAS 5 for background correction and PM adjustment, RMA (median polish) as expression summary method and the choice of the normalisation algorithm does not matter. Note that Choe *et al.* (2005) did not include the gcRMA methodology in the set of algorithms they compared.

MA plots are used to summarise responses. In two colour channels M is the difference in log expression values while A is the average of those values (Dudoit *et al.*, 2002):

$$M = \log_2(R) - \log_2(G) \qquad A = \frac{1}{2}\left(\log_2(R) + \log_2(G)\right) \tag{7.1}$$

where R and G stand for the red and green colour in the two colour arrays. Note that logarithms base 2 are used, as intensities are typically stored as 16 bit values. For single channel arrays, the G values of the equation above are replaced by a synthetic mean value of the array set. However, in this study the mean value of the arrays of the two aerobic cultures, SsAe_1 and SsAe_2, was used. Each MA plot (for example figures 7.3 and 7.4) is also accompanied by its median M value and the interquartile range (the range around the median in which 50 % of the values are found) of the M values. The median should be around zero and the IQR as small as possible. In each MA plot, a smooth line (local polynomial fitting) is fitted to the data (white line in figures 7.3 and 7.4). This line should be as close as possible to the $M = 0$ horizontal axis. Ideal experiments have MA plots where most of the dots are uniformly distributed over the A range and close around the $M = 0$ line. The dots outside this region should be clearly outside, so that the M values being different from zero are clearly due to differential expression and not some random noise. A more in-depth description of MA plots can be found in Colantuoni *et al.* (2003).

Different methodologies for extracting expression levels from probe intensities were tried. The quality of those was accessed with MA plots. All calculations were performed starting from the affy package (Irizarry *et al.*, 2003a; Gautier *et al.*, 2004) of the bioconductor framework

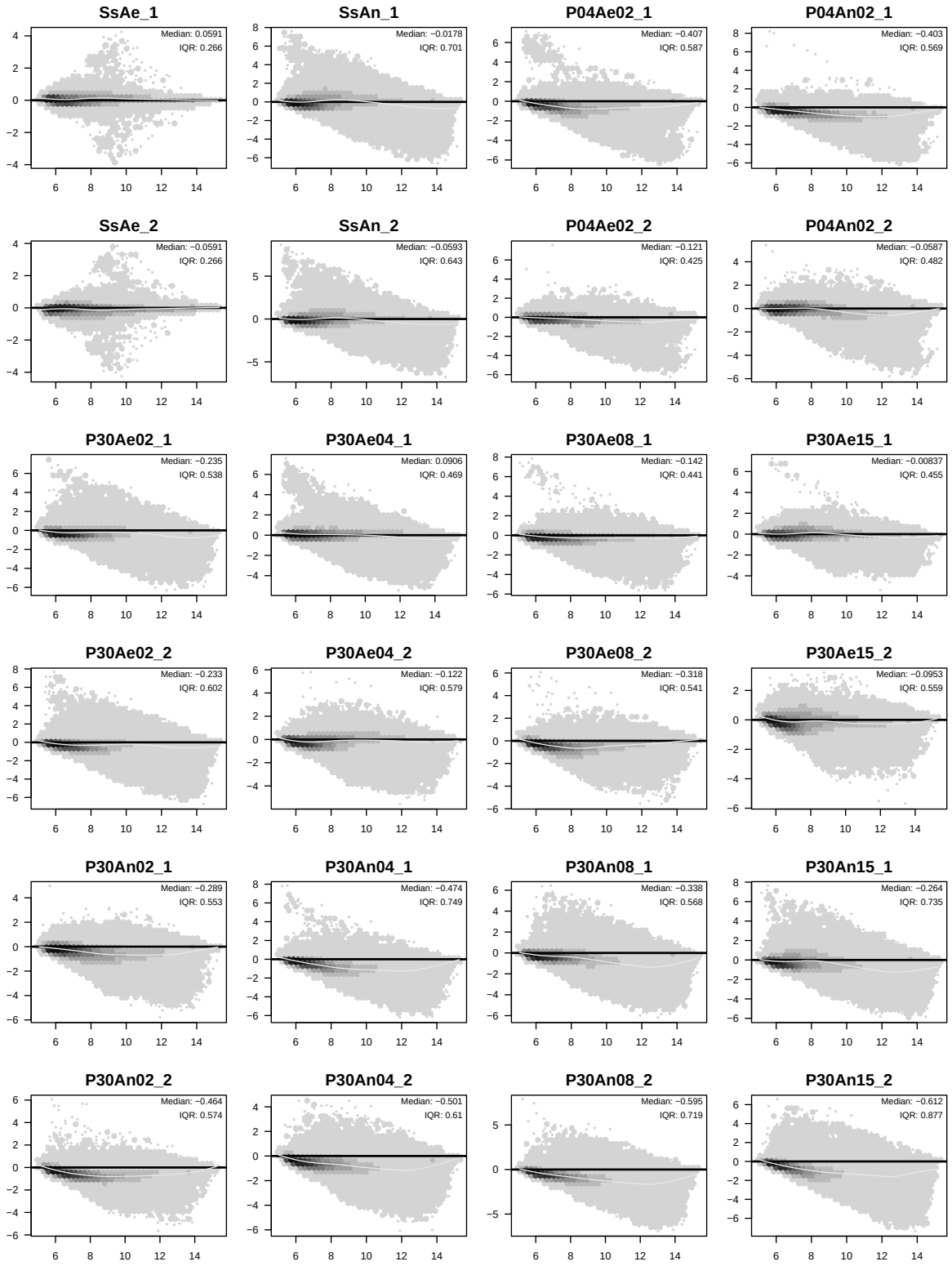


Figure 7.3: MA plots of the raw data. Plot names are explained in table 7.1. The x-axis contains the sum of $\log_2$ expression values of the individual arrays and the synthetic array made of the combination of SsAe_1 and SsAe_2 while the y-axis contains the $\log_2$ differences.

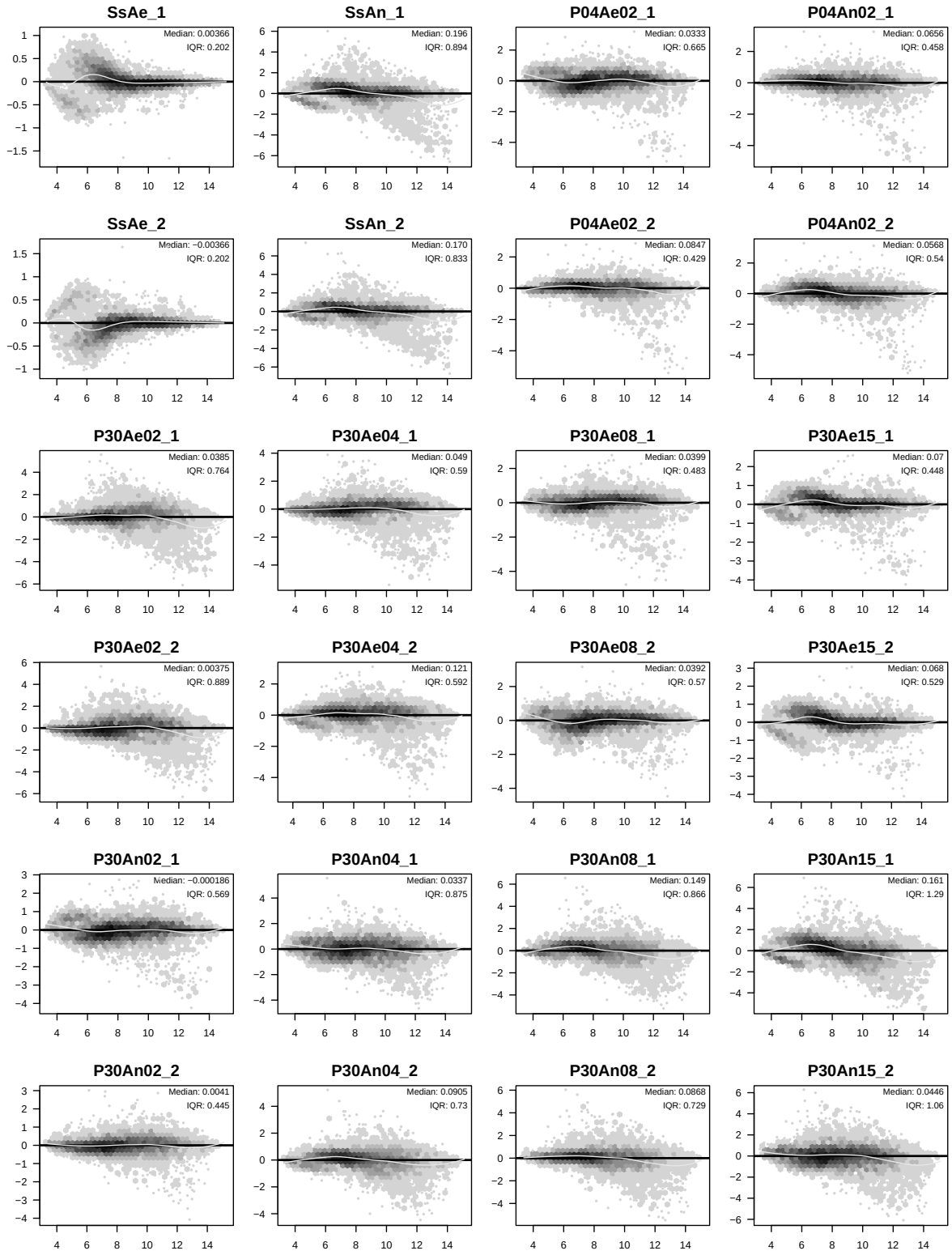


Figure 7.4: MA plots of the RMA normalised data after removing genes that were not differentially expressed in the two repeats of at least one experimental condition. Notation as in figure 7.3.

(Gentleman *et al.*, 2004). The most straight-forward method for transforming the raw data obtained from the CEL files (figure 7.3) is the RMA method (figure 7.4, Irizarry *et al.*, 2003b). One can see that indeed, RMA normalisation lowers the variance (the IQR's in figure 7.4 are lower than those in figure 7.3). The vsn normalisation algorithm looks attractive as it explicitly aims at stabilising the variance (Huber *et al.*, 2003). However, using vsn for normalisation and keeping the other steps as in RMA, did not give satisfying results (data not shown). They were worse than the results obtained with RMA normalisation (figure 7.4), something also observed by Qin *et al.* (2006) in a comparative study of different methodologies.

Using a background correction method based on the sequence of the probes is given as the best performing method in Qin *et al.* (2006). However, in this case, when using gcRMA (a background correction method based on probe sequences, Wu *et al.*, 2004) for background correction, and keeping the other steps as in RMA, the resulting MA plots were worse than when using RMA normalisation (data not shown).

Thus, the best methodology for summarising the probe intensities seems to be RMA.

In the subsequent analysis genes that were not expressed, were removed from the normalisation. Presence/Absence was calculated with the MAS5 method. From the 10206 genes, 7776 remained present.

7.4.2 Statistical tests

Once gene expression data are obtained from the raw arrays, differences in expression levels have to be searched for. To do this, the R library limma was used (Smyth, 2005). Limma makes use of linear models to analyse array data. The estimation of the variance on the expression data is done using an empirical Bayes method. An excellent introduction of linear models for analysing microarray experiments, using the Bayesian approach, can be found in Smyth (2004).

Linear models for microarrays

Assuming n different conditions for which microarrays were analysed, the expression of a certain gene y over the different microarrays can be summarised in a vector $\boldsymbol{y}$:

$$\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T \tag{7.2}$$

The expected values of the gene expressions in the different conditions is:

$$\mathrm{E}[\boldsymbol{y}] = \mathsf{X}\,\boldsymbol{\alpha} \tag{7.3}$$

In this equation, X is the design matrix and $\boldsymbol{\alpha}$ is the coefficient vector.

In the simple case of only analysing the data for experiment SsAe (aerobic conditions), SsAn (anaerobic conditions), a possible design matrix is (each experiment is repeated 2 times):

$$\begin{array}{c} \text{SsAe\_1} \\ \text{SsAe\_2} \\ \text{SsAn\_1} \\ \text{SsAn\_2} \end{array} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \tag{7.4}$$

Substituting this design matrix in equation 7.3, yields:

$$\mathrm{E}\begin{bmatrix} y_{SsAe} \\ y_{SsAn} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{m,1} \\ \alpha_{m,2} \end{bmatrix} \tag{7.5}$$

where $\alpha_{m,1}$ and $\alpha_{m,2}$ are elements of the vector $\boldsymbol{\alpha}$. The first element of $\boldsymbol{\alpha_m}$ represents the mean expression value for SsAe while $\alpha_{m,2}$ is the expression value for the anaerobic experiment. The difference in expression between the two experimental conditions is then $\alpha_{m,1} - \alpha_{m,2}$. As expression levels of the different treatments can easily be extracted, this arrangement of the design matrix is called the group-means parameterisation (Smyth, 2005).

Another possible arrangement of the design matrix is:

$$\begin{array}{c} \text{SsAe\_1} \\ \text{SsAe\_2} \\ \text{SsAn\_1} \\ \text{SsAn\_2} \end{array} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{7.6}$$

Which yields, after substitution in equation 7.3:

$$\mathrm{E}\begin{bmatrix} y_{SsAe} \\ y_{SsAn} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{c,1} \\ \alpha_{c,2} \end{bmatrix} \tag{7.7}$$

The first element of $\boldsymbol{\alpha_c}$ again represents the expression value of SsAe, but the second element, $\alpha_{c,2}$ stands for the difference in expression between SsAe and SsAn. Thus, for this arrangement of the design matrix X, contrasts between different treatments can easily be extracted. Hence it is called the treatment-contrasts parameterisation (Smyth, 2005).

Major requirements of the design matrix are that there are just as many columns as there are factors in the experiment (in the example above, this is 2, in the experimental setup of this chapter, it is 12) and that all columns are linearly independent. However, the contrasts that ones want to analyse are not always linearly independent and the number of interesting contrasts is not always equal to the number of different experimental conditions. One can

manually search how to combine the different elements of $\boldsymbol{\alpha}$, as in equation 7.7, but a more systematical way consists of by defining a contrast matrix C. A contrast matrix transforms $\boldsymbol{\alpha}$ to a new vector $\boldsymbol{\beta}$, the elements of which have a biological meaning:

$$\beta = \mathsf{C}^{\mathsf{T}}\ \alpha \tag{7.8}$$

In the first examples above (the design matrix given in equation 7.4), assuming only the difference in expression is of interest, the contrast vector $\boldsymbol{\beta_m}$ has only one column and a possible contrast matrix is:

$$\mathsf{C_m} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \Longrightarrow \quad \boldsymbol{\beta_m} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \alpha_{m,1} \\ \alpha_{m,2} \end{bmatrix} = \mathrm{E}[y_{SsAn}] - \mathrm{E}[y_{SsAe}] \tag{7.9}$$

A possible contrast matrix for the second design matrix (equation 7.6) is (again assuming only the difference in expression is of interest):

$$\mathsf{C_c} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \Longrightarrow \quad \boldsymbol{\beta_c} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\mathsf{T}} \begin{bmatrix} \alpha_{c,1} \\ \alpha_{c,2} \end{bmatrix} = \mathrm{E}[y_{SsAn}] - \mathrm{E}[y_{SsAe}] \tag{7.10}$$

For each gene, the linear model of equation 7.3 is fitted and the coefficients, $\boldsymbol{\alpha}$**,** and contrasts, $\boldsymbol{\beta}$, are estimated. Associated with each estimate is a variance. For a certain gene y, the variance on the expression value, is σ^2, estimated as s^2:

$$\mathrm{var}(\widehat{y}) = s^2 \tag{7.11}$$

A positive definite matrix V is estimated with each $\boldsymbol{\alpha}$ such that:

$$\mathrm{var}\left(\widehat{\boldsymbol{\alpha}}\right) = \mathsf{V}\ s^2 \tag{7.12}$$

resulting in the covariance matrix of the contrast vector, $\boldsymbol{\beta}$, being:

$$\mathrm{var}\left(\widehat{\boldsymbol{\beta}}\right) = \mathsf{C}^{\mathsf{T}}\ \mathsf{V}\ \mathsf{C}\ s^2 \tag{7.13}$$

The square roots of the diagonals of the matrix $\mathsf{C}^{\mathsf{T}}\ \mathsf{V}\ \mathsf{C}$, $\sqrt{u}$, are called the unscaled standard deviations. Assuming each element of $\boldsymbol{\beta}$ is normally distributed as $N\left(0, u_i\, s^2\right)$, the ordinary t-statistics for a contrast β_i can be calculated:

$$t_i = \frac{\widehat{\beta}_i}{s\ \sqrt{u_i}} \tag{7.14}$$

which tests the assumption that β_i is zero. However, as the number of genes tested is very large and the number of repeats for microarrays are typically very low, there will always be some β_i with very small variance, giving a high t-value even if β_i is small, resulting in the rejection of the null hypothesis. Different *ad hoc* solutions have been proposed. They all increase s. The difference lies in the justification of the argument added to the standard deviation (Lönnstedt

& Speed, 2002). In limma, a Bayesian approach is used for this justification. Information on the variance of the other genes is included in the estimation of the variance of the gene expression values, resulting in an improved standard deviation, $\tilde{s}$. Based on this new $\tilde{s}$, a modified t-statistic can be calculated (Lönnstedt & Speed, 2002; Smyth, 2004).

In this work, the group-means parameterisation was used. Selected contrasts are assessed and the expression over the different experiments for genes from the central carbon metabolism are discussed. To assess differences in expression, a p-value of 0.01 was used. If not stated otherwise, plots of gene expressions are relative to the continuous aerobic cultures.

7.5 Results and discussion

7.5.1 Aerobic versus anaerobic conditions

The difference in expression profile between the aerobic and anaerobic phase for the three different setups (being continuous cultures *sensu stricto*, oscillating aeration with a period of 4 minutes and oscillating aeration with a period of 30 minutes) was assessed. The samples compared were those for which the cells were as long as possible in the aerobic or in the anaerobic phase. Thus, the following contrasts were generated: SsAe–SsAn (the differences in expression between aerobic cultures and anaerobic cultures), P04Ae02–P04An02 and P30Ae15–P30An15 (difference in expression at the end of the aerobic phase and the end of the anaerobic phase for the 4 minutes and the 30 minutes period experiments, respectively).

When selecting genes that are differentially expressed, two parameters are important. The p-value, which was kept at 0.01, and the minimal fold change (the amount with which a gene is upregulated compared to the reference condition). The correlation between p-values and fold changes can be visualised in so-called volcano plots (figure 7.5), where the $-\log_2$ of the p-values are plotted against the $\log_2$ fold changes. Ideally, genes with a low p-value should have a large fold change. However, for this data set this was not the case. Figure 7.5 shows that many of the p-values lower than 0.01 are for log fold changes less than one. The significance of those differential expressions is questionable, as in the MA plots of the replicates of the experiments (figure 7.6) it is observed that the differences in log expression values in replicated experiments are commonly larger than two. Thus, log fold changes of 2 or less can be due to experimental variation and are not necessarily biologically significant.

Figure 7.7 shows the Venn diagrams for the three contrasts described above for different minimal fold changes. When setting no minimal fold change (figure 7.7a), the number of differentially expressed genes between SsAe and SsAn and between P30Ae15 and P30An15 roughly corresponds to the number of differentially expressed genes reported by Salmon *et al.* (2003).

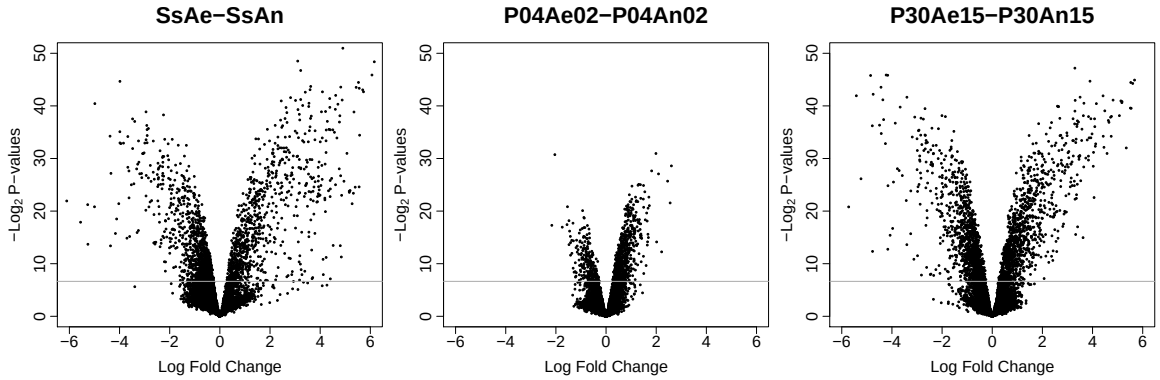


Figure 7.5: Volcano plots. The horizontal line stands for the p-value of 0.01.

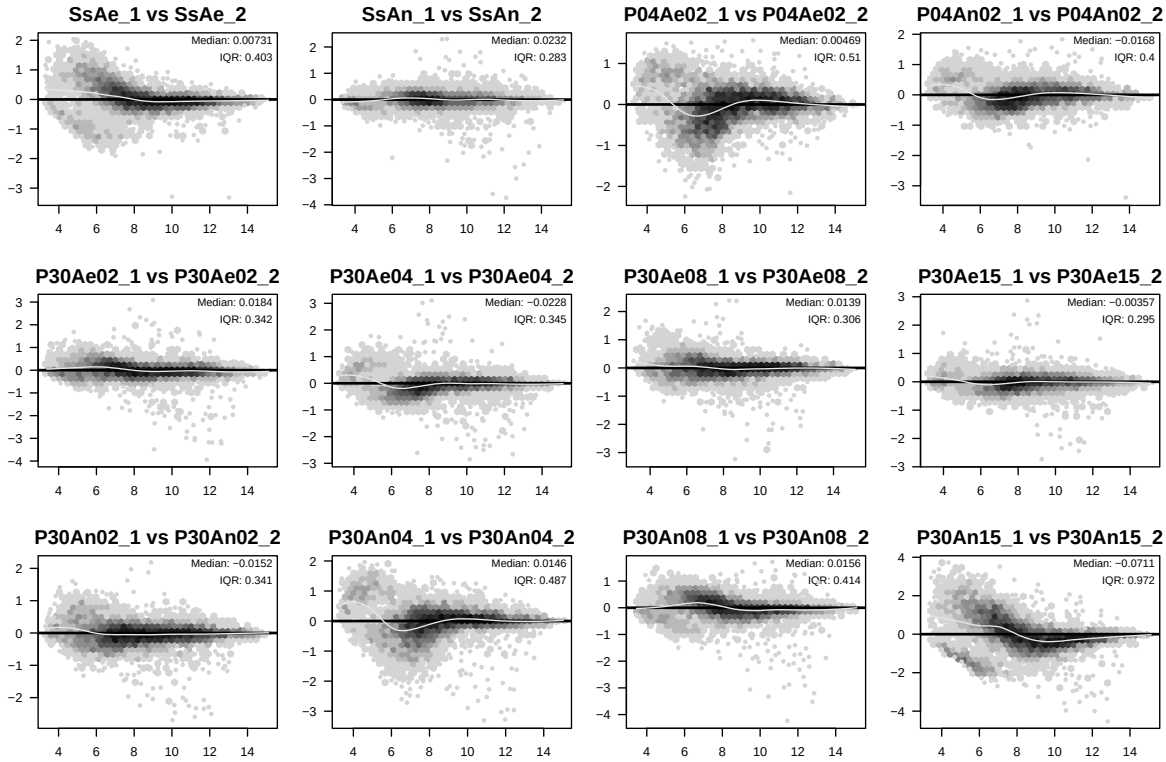
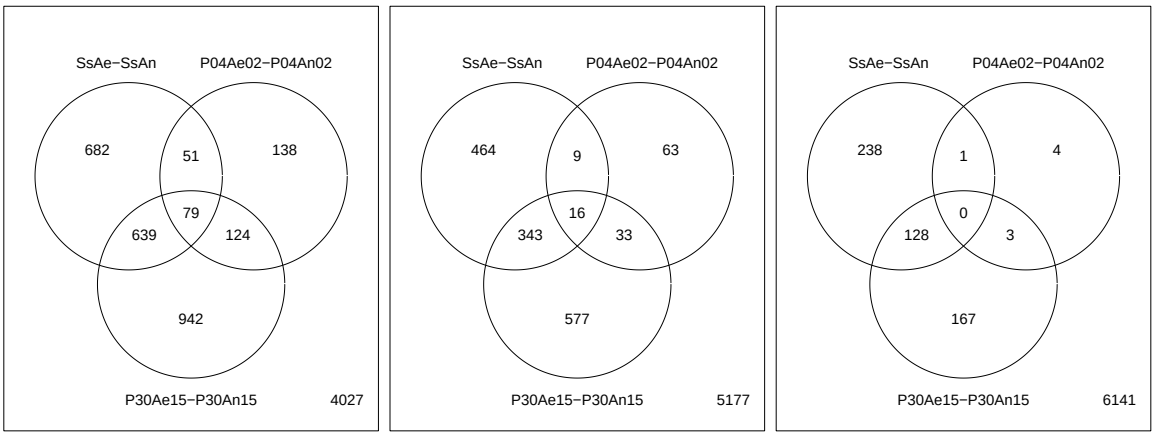


Figure 7.6: MA plots of repeats of the RMA normalised data. The x-axis contains the log sum of the expression values of the two repeats and the y-axis the log difference.

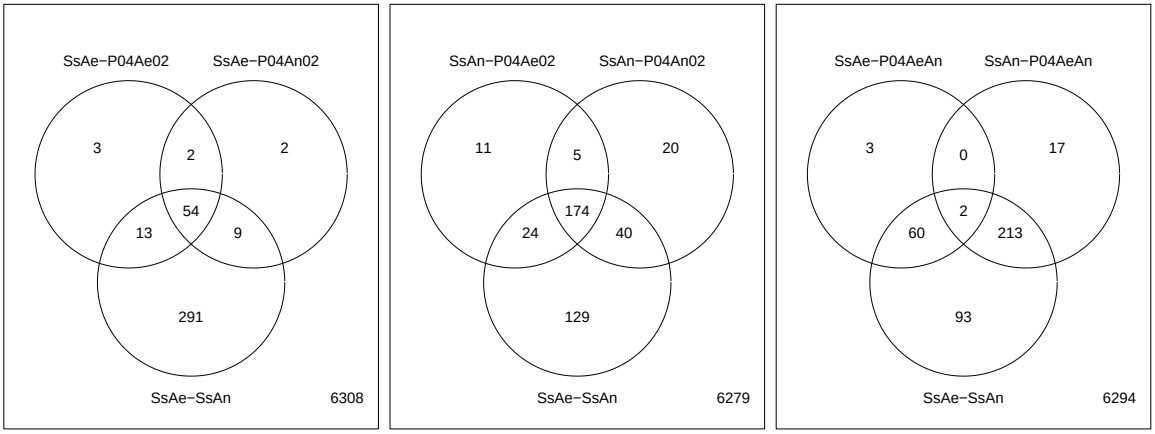
Figure 7.7 further shows that many genes differentially expressed between fully aerobic cultures and fully anaerobic cultures are not the same as those differentially expressed in the different phases of the oscillating experiment with a period of 30 minutes and to a lesser extent, that of 4 minutes.

As already observed in the volcano plots (figure 7.5), the number of differentially expressed



(a) Log fold change > 0 **(b)** Log fold change > 1 **(c)** Log fold change > 2

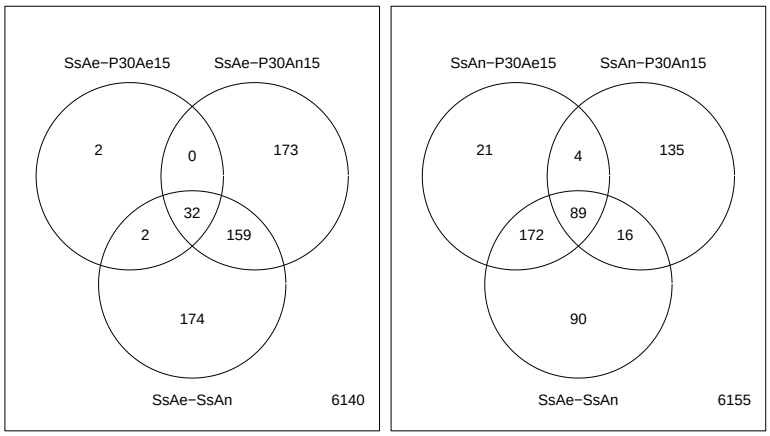
Figure 7.7: Venn diagrams of the genes differently expressed between aerobic and anaerobic conditions. The 3 diagrams compare the same contrasts, but the minimal fold change for selecting genes as differentially expressed, is increased.



(a) **(b)** **(c)**

Figure 7.8: (a) Comparison of the P04 phases with the aerobic continuous culture. (b) Comparison of P04 phases with the anaerobic continuous culture. (c) Experiments P04Ae02 and P04An02 are treated as replicates of one experiment, P04AeAn, and are compared with the aerobic and anaerobic cultures. The selected genes all have a log fold change of minimally two.

genes lowers drastically when increasing the minimal fold change (figure 7.7). The decrease is spectacular for the P04 experiment: most of the genes differentially expressed have a log fold change of less than 2. Thus it seems that there is not much difference in genetic expression between the aerobic and anaerobic phase when oscillating the oxygen supply with a period of 4 minutes. Furthermore, the P04 experiments have more genes differentially expressed with anaerobically grown cells (figure 7.8b) than with aerobic cultures (figure 7.8a), suggesting



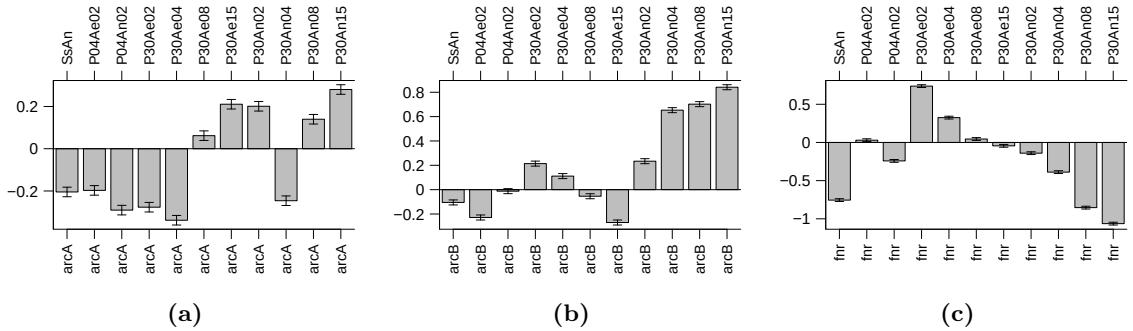
(a) **(b)**

Figure 7.9: Comparison of the 30 minutes oscillation experiment at the end of the aerobic and anaerobic phase with (a) aerobic and (b) anaerobic continuous cultures. The selected genes all have a log fold change of minimally two.

that the 4 minutes oscillating cultures have more in common with aerobically grown cells than with anaerobic cultures.

Most of the genes from P04 that are differentially expressed between the continuous cultures, are common to the aerobic phase of P04 and the anaerobic phase of P04 (54 and 174, figures 7.8a and 7.8b), suggesting again that there is not much difference between the two phases. Therefore the arrays for experiments P04Ae02 and P04An02 were combined in P04AeAn and compared with the continuous cultures (figure 7.8c). It can be seen that most of the genes differentially expressed between P04AeAn and SsAe or SsAn are also differentially expressed between SsAe and SsAn. Thus in the P04 cultures, some genes (62, in figure 7.8c) are regulated as if under anaerobic conditions, but a large part of the genes playing a role in the anaerobic versus aerobic metabolism, are expressed as under aerobic conditions (215 of the 268 genes differentially regulated between SsAe and SsAn are also up or down regulated when comparing SsAn with P04AeAn, figure 7.8c).

The picture for the experiment in which the oxygen input was alternated with a period of 30 minutes, is completely different (figure 7.9). At the end of the aerobic phase, the gene expression profile is almost the same as under fully aerobic conditions (SsAe-P30Ae15, figure 7.9a) while at the end of the anaerobic phase, many genes are still differentially expressed compared to the anaerobic continuous culture (SsAn-P30An15, figure 7.9b). Thus it seems as if adaptation to aerobiosis is much faster than to anaerobiosis, at least at the gene level. Also many genes differentially expressed between SsAn and P30S15 are not differentially expressed between SsAe and SsAn (figure 7.9b).



(a) **(b)** **(c)**

Figure 7.10: Relative gene expressions of the global regulators *arcAB* (a,b) and *fnr* (c). The y-axis contains the relative $\log_2$ gene expressions relative to the aerobic continuous culture (SsAe). Error bars are for the Bayesian corrected standard deviation.

7.5.2 The oxygen regulators ArcAB and Fnr

The expression profile of *arcA* is rather constant during the experiments (figure 7.10a). However, the one from *arcB* and *fnr* does change over the course of the P30 oscillating experiments (figure 7.10b,c). Normally, it is expected that the total content of Fnr (active and inactive) in the cells is constant (Sutton *et al.*, 2004a). In the experiments described in this work, *fnr* expression is systematically lower under anaerobic conditions. The difference in expression of *fnr* between the aerobic and the anaerobic phase of the P04 experiments is less pronounced than between the aerobic and anaerobic continuous cultures.

Apparently some time is needed to lower the *fnr* expression when switching to anaerobic conditions. This can be observed from the gradual decrease in *fnr* expression in the anaerobic phase of the P30 experiments (figure 7.10c). However, when switching back to aerobic conditions, the increase in *fnr* gene expression is immediate and overshoots during the first 10 minutes of the aerobic phase (P30Ae02 and P30Ae04 in figure 7.10c). A possible explanation could be that under aerobic conditions Fnr is more easily degraded and that to keep the protein concentration constant, more *fnr* has to be expressed.

Fnr normally has a positive effect on the expression of *arcA* (Compan & Touati, 1994). This seems not to be the case in these experiments: there is not really a correlation between the expression profile of *fnr* and *arcA* (figure 7.10). However, *fnr* gene expression seems to be negatively correlated with *arcB* in the P30 experiments.

7.5.3 Central carbon metabolism

The effect of oxygen on central carbon metabolism is mostly sensed on the pyruvate node and the Krebs cycle. Under anaerobic conditions, the pyruvate dehydrogenase complex is substituted by pyruvate formate lyase and the citric acid cycle is branched at the alfaketoglutarate–

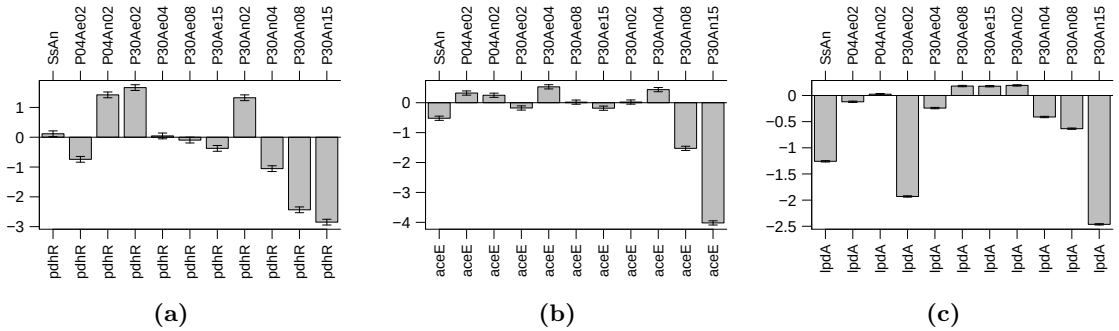


Figure 7.11: Relative gene expressions for genes involved in the functioning of the pyruvate dehydrogenase complex: (a) *pdhR*, encoding a regulator of PDHc, (b) *aceE* and (c) *lpdA*. Notation as in figure 7.10.

succinate node. This is discussed below.

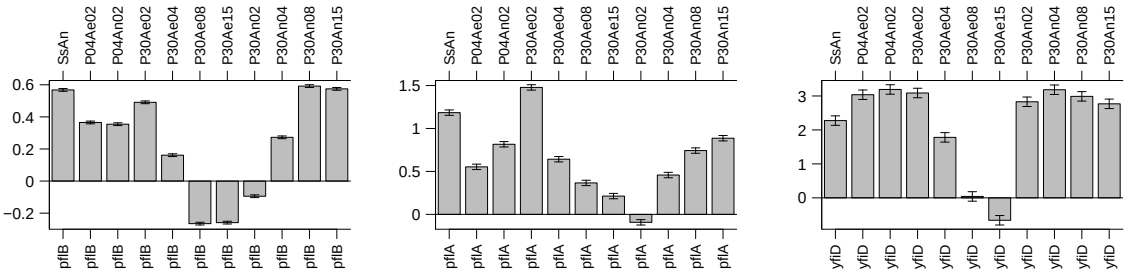
Pyruvate node

The pyruvate dehydrogenase complex (PDHc), is the gene product of the *pdhR-aceEF-lpd* operon (figure 7.11). However, *lpd* also provides subunits for the alfaketoglutarate dehydrogenase complex, catalysing the conversion of alfaketoglutarate to succinyl-CoA in the Krebs cycle. Therefore, the *pdhR-aceEF-lpd* operon contains different promoters to produce at least three different transcripts: *aceEF-lpd*, *aceEF* and *lpd*. PdhR negatively regulates the synthesis of PDHc (Quail *et al.*, 1994). The slower response of *lpd* correlates with the expression profile of *sucABCD* (figure 7.5.3) encoding the other subunits of alfaketoglutarate dehydrogenase (*sucAB*) and succinate dehydrogenase (*sucCD*).

The genes involved in pyruvate formate lyase, converting pyruvate to acetyl-CoA and formate under anaerobic conditions, are immediately upregulated when entering anaerobiosis (figure 7.12). The *pfl* operon is upregulated by Fnr (Cox *et al.*, 2005) and Fnr reacts fast to changing environmental oxygen conditions (Partridge *et al.*, 2007). The differential expression of *pflB*, coding for Pfl, is less pronounced than *pflA*, an activator of Pfl and *yfiD*, a protein involved in the reparation of Pfl, when damaged by oxygen (Zhu *et al.*, 2007). Pfl is rapidly inactivated in the presence of small amounts of oxygen.

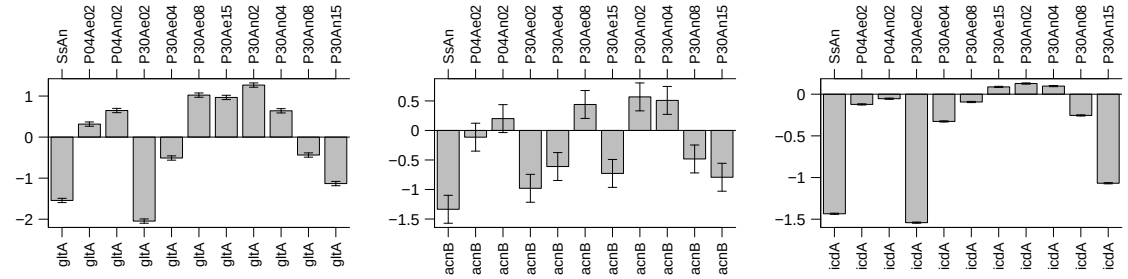
Krebs cycle

Under anaerobic conditions, the Krebs cycle is split in an oxidative and a reductive branch. The oxidative branch is only used for biomass precursors and is not very active. This is reflected in the gene expression profiles of *gltA*, *acnB* and *icd* coding for citrate synthase, aconitase B and isocitrate dehydrogenase (figure 7.13). The oxidative part of the Krebs cycle ends at alfaketoglutarate, hence the stronger downregulation of *sucABCD* coding for the



(a) **(b)** **(c)**

Figure 7.12: Relative gene expressions of (a) *pflB*, coding for pyruvate formate lyase, (b) *pflA*, coding for an activator of Pfl, and (c) *yfiD*, whose gene product reactivates Pfl after oxygen damage. Notation as in figure 7.10.



(a) **(b)** **(c)**

Figure 7.13: Expression profiles of the first three genes from the Krebs cycle: (a) *gltA* (citrate synthase) (b) *acnB* (aconitase B) and (c) *icd* (isocitrate dehydrogenase). Notation as in figure 7.10.

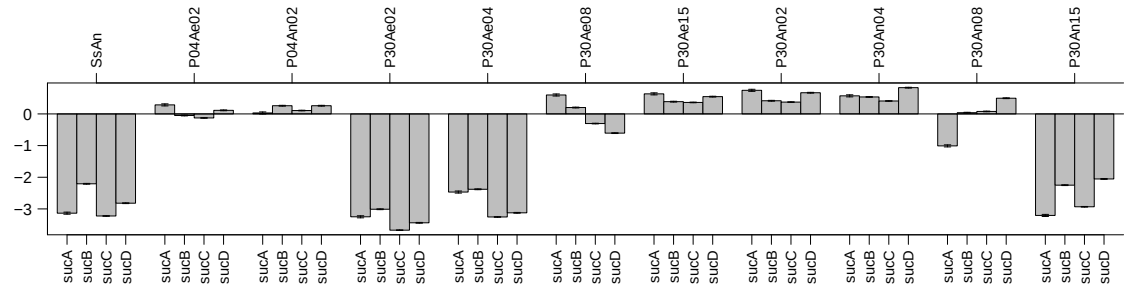


Figure 7.14: Relative gene expressions of the transcripts for alfaketoglutarate dehydrogenase (*sucAB*) and succinyl-CoA synthetase (*sucCD*). Notation as in figure 7.10.

enzymes transforming alfaketoglutarate to succinate: unlike the first three reactions of the Krebs cycle (figure 7.13), alfaketoglutarate dehydrogenase has to be completely inactivated under anaerobic conditions (figure 7.14). However, the genetic down- or upregulation of *sucABCD* is not fast: only after 8 minutes the expression was lowered/increased, something

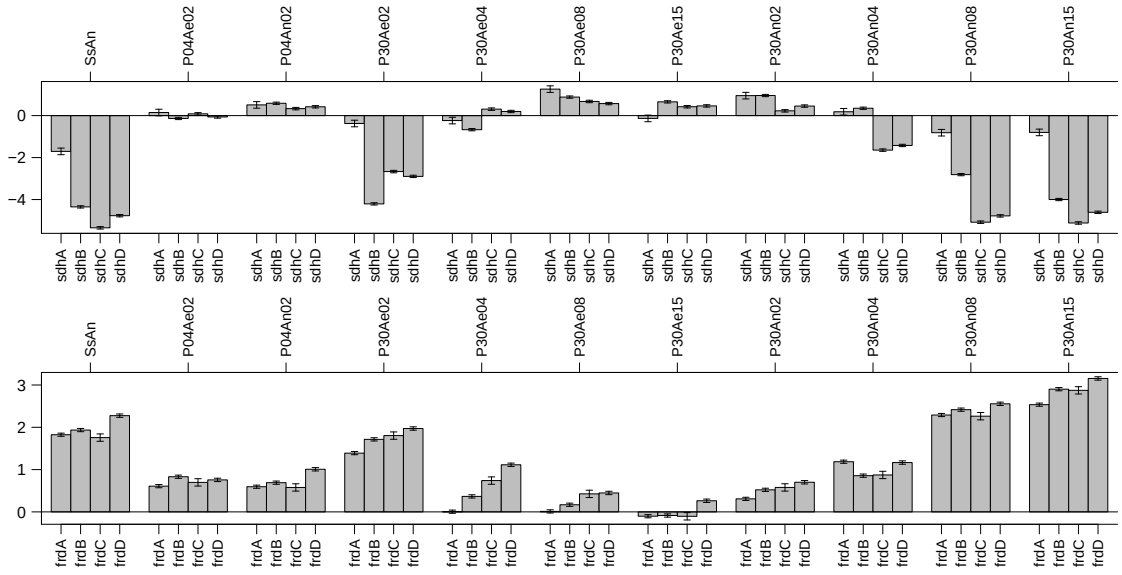
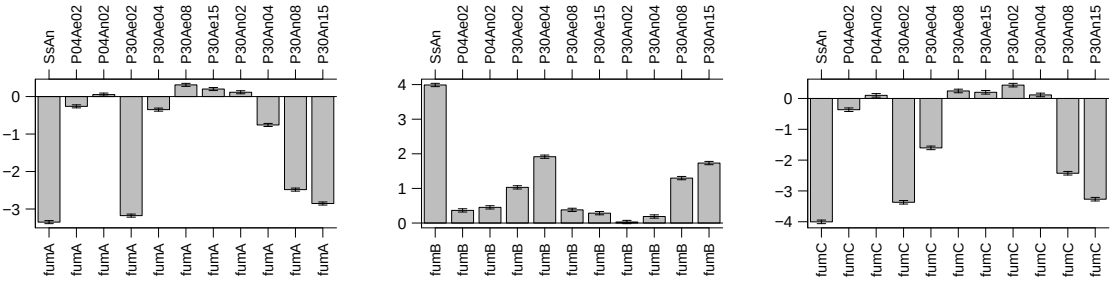


Figure 7.15: Relative gene expressions of succinate dehydrogenase (*sdhABCD*) and fumarate reductase (*frdABCD*). Notation as in figure 7.10.

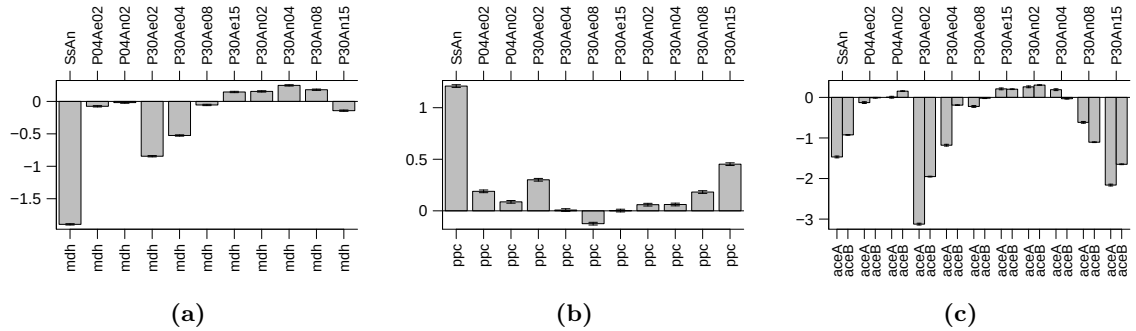


(a) **(b)** **(c)**

Figure 7.16: Expression profile of the genes involved in the conversion of fumarate to malate. (a) FumA is active under microaerobic conditions, (b) FumB converts malate to fumarate and is important in the anaerobic respiration, and (c) FumC is highly active under aerobic conditions. Notation as in figure 7.10.

also observed when switching cell cultures from aerobic to anaerobic conditions and *vice versa* Partridge *et al.* (2007, 2006).

The adaptation to changing oxygen levels is faster for the expression of the genes encoding the enzymes of the reductive branch of the Krebs cycle (figures 7.15 and 7.16). Indeed *sdhABCD*, encoding succinate dehydrogenase is downregulated by Fnr, while frdABCD, the fumarate reductase operon, reducing fumarate to succinate, and *fumB*, whose gene product converts malate to fumarate, are both upregulated by Fnr. The peak activity of Fnr occurs around 5 minutes after the oxygen switch (Partridge *et al.*, 2007). The operon for malate dehydrogenase



(a) **(b)** **(c)**

Figure 7.17: Relative expression profiles for *mdh*, encoding malate dehydrogenase, *ppc*, encoding PEP carboxylase and *aceAB*, encoding the enzymes of the glyoxylate pathway. Notation as in figure 7.10.

(*mdh*, figure 7.17a) is regulated by ArcA (Cox *et al.*, 2005). Activation of ArcA is slower than that of Fnr (Partridge *et al.*, 2007), hence the slow response of *mdh* to changing oxygen concentrations. The carbon needed for reducing fumarate to succinate should come from PEP, via PEP carboxylase (encoded by *ppc*), the enzyme that combines CO_2 with PEP to form oxaloacetate. However, compared to the fully anaerobic cultures, not much upregulation is observed for *ppc* during the anaerobic phases of the oscillating experiments (figure 7.17b). The glyoxylate pathway, downregulated by *ArcA* (Cox *et al.*, 2005), is clearly downregulated during anaerobic conditions (figure 7.17c). Note how *mdh* and *aceAB*, both regulated by ArcA, are still downregulated during the two first minutes of aerobiosis in the 30 minutes oscillation experiments.

Fermentative pathways

The expression of *poxB* is not altered during the oscillations (figure 7.18c) while the other pathway for producing acetate, via phosphate acyltransferase (encoded by *pta*) and acetate kinase (encoded by *ackA*), is upregulated in the anaerobic and P30 cultures. Acetyl-CoA synthetase, encoded by *acs*, serves to metabolise acetate by converting it to acetyl-CoA. The reaction is completely irreversible *in vivo* because ATP is hydrolysed to AMP and pyrophosphate and the pyrophosphate is quickly removed by the pyrophosphatases present in the cells (Kumari *et al.*, 2000). Regulation of this gene is complex and involves the cyclic AMP receptor protein (CRP), Fnr, the glyoxylate shunt repressor IclR and its activator FadR (Kumari *et al.*, 2000). In the P30 experiment, the downregulation of *acs* nicely correlates with the upregulation of *ackA-pta* (figure 7.18a and 7.18b) suggesting that the acetate produced during the anaerobic phase is again assimilated in the aerobic phase. Indeed, no significant amounts of acetate were observed in the oscillating cultures.

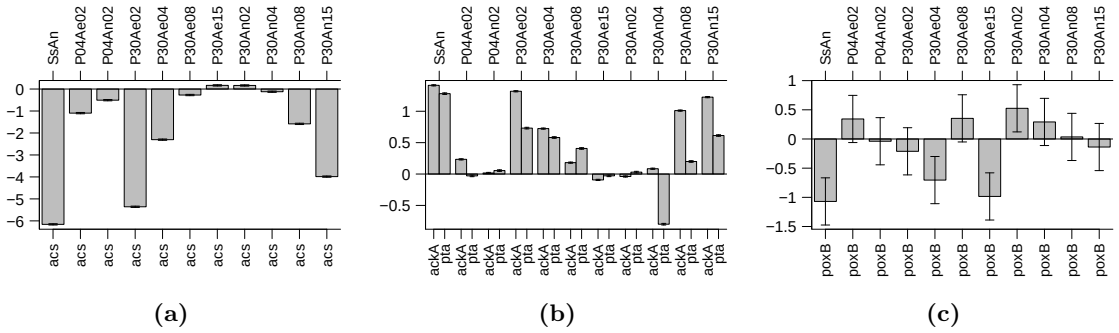


Figure 7.18: Relative expression profile of (a) *acs* (conversion of acetate to acetyl-CoA), (b) *ackA* and *ptaA* (conversion of acetyl-CoA to acetate via acetylphosphate), and (c) *poxB* (coding for pyruvate oxydase which transforms pyruvate to acetate). Notation as in figure 7.10.

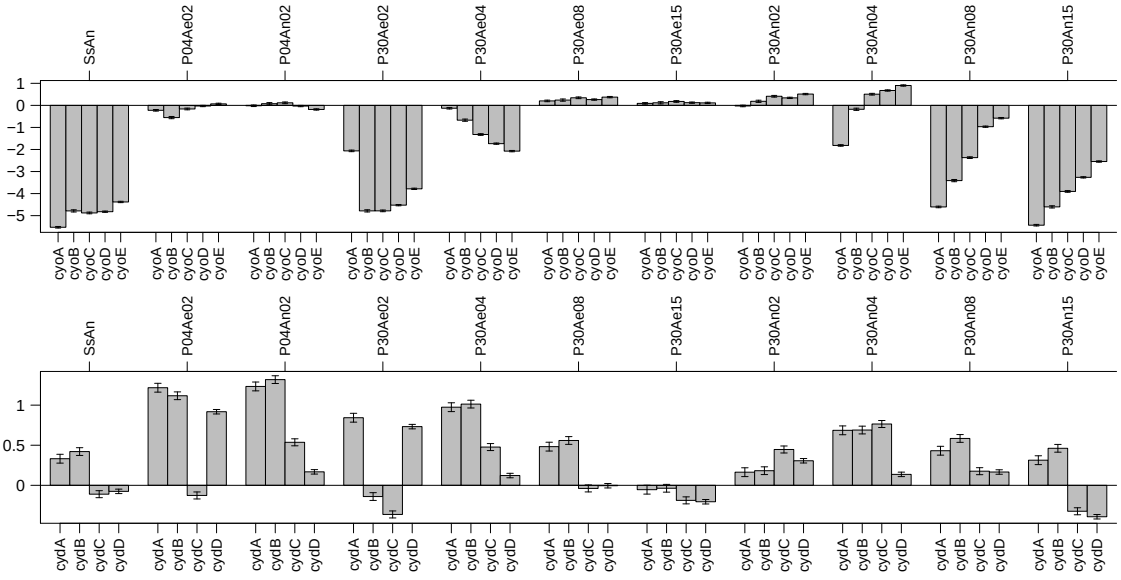


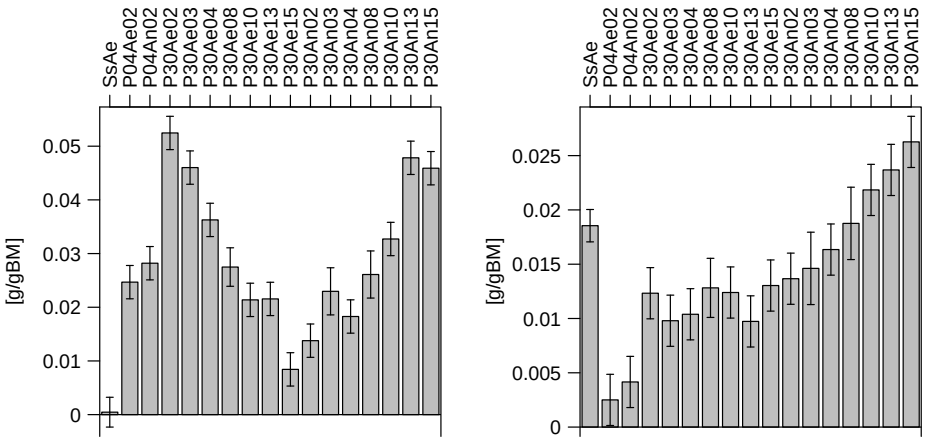
Figure 7.19: Relative gene expressions of the operon of (a) cytochrome o (*cyo*, top) and (b) cytochrome d (*cyd*, bottom). Notation as in figure 7.10.

Electron acceptors

Expression of *cyd*, encoding cytochrome d, active under microaerobic conditions, is activated by ArcA and repressed by Fnr while *cyo*, encoding cytochrome o, active under oxygen abundant conditions, is repressed by both ArcA and Fnr (Cox *et al.*, 2005). This explains the strong repression of the *cyo* operon and the not as strong repression of *cyd* (figure 7.19).

Effects of the oscillations

The amount of fermentative products excreted during oscillating operations, was significantly lower than during anaerobic cultivations. Indeed, using the classical methods for calculating



(a) Acetate **(b)** Succinate

Figure 7.20: Fluxes for (a) acetate and (b) succinate. In the P30 experiment, extra samples were taken for HPLC analysis as compared to the mRNA samples, at 3, 10 and 13 minutes of the aerobic and anaerobic phase. Data for the anaerobic experiments are not shown and were for acetate 0.39 ± 0.01 g/g BM/h and for succinate 0.280 ± 0.007 g/g BM/h. Error bars represent the standard deviation.

mean and standard deviation, the values found were not significantly different from zero. However, the metabolite data are mathematically not different from gene expression data and the same Bayesian corrections can be applied, resulting in much smaller variances (figure 7.20). It can be seen how, in the P30 experiments, the acetate concentrations (figure 7.20a) increases when the *acs* expression is low (figure 7.18a) and the expression of *ackA-pta* is high (figure 7.18b). The pattern for succinate was different. The concentration gradually increased during the anaerobic phase of the P30 experiment but when the aerobic phase started, succinate quickly reentered the cellular metabolism (figure 7.20b). Correlation with the gene expression patterns is less clear (figure 7.15).

The biomass production rate of the P04 experiment was reduced by 25 % compared to the aerobic cultures and that of P30 by 34 %. Thus, the increase in succinate production that was hoped for, did not occur. A possible cause for this is the introduction of futile cycles by activating aerobic and anaerobic pathways at the same time.

There is not much difference in expression of the Krebs cycle genes for the two phases of the P04 experiment. The gene expression levels are rather similar to those of the fully aerobic cultures (the reference in all the figures) with the exception of *frd* and *cyd*. On can thus reasonably assume that the Krebs cycle remains active, even in the anaerobic phase, but that fumarate reductase and cytochrome d are added to the metabolic system, resulting in an extra inefficiency: cytochrome d does not contribute to the H^+ gradient (Puustinen *et al.*,

1991), needed for ATP generation. Furthermore, *acs* is not downregulated, thus, all acetate that could be produced during the anaerobic phase is quickly re-assimilated by acetyl-CoA synthetase, at the cost of 2 ATPs. And succinate dehydrogenase and fumarate reductase are countering each other. This could explain the lower biomass yields and the lack of succinate excretion.

The same can be said for the P30 experiment. However, the duration of both phases is long enough so that the metabolism can partially adapt. Probably more acetate is produced during the anaerobic phase than in the P04 experiments (*acs* is also downregulated during the anaerobic phase, figure 7.18a) but it is reasimilated in the aerobic phase. This oscillating behaviour of acetate was observed in the HPLC measurements (figure 7.20a). Probably this futile cycle over acetate is responsible for the extra 9 % of decrease in biomass yield compared to the decrease found for the P04 experiment.

7.6 Conclusions

The influence of controlled oscillations in the oxygen supply on the gene expression levels was investigated. Four types of experiments were performed: (a) fully aerobic continuous cultures, (b) fully anaerobic continuous cultures, (c) cultures in which the oxygen input was oscillated with a period of 4 minutes and (d) cultures in which the period of oscillation was 30 minutes.

The gene expression levels of the 4 minutes oscillation experiments were more similar with those of the fully aerobic continuous cultures than with the anaerobic ones (figure 7.8). In the P30 experiments, the similarity between the expression profile after 15 minutes of aerobiosis and the fully aerobic cultures, was larger than the similarity between the gene expression profile after 15 minutes of anaerobiosis and the fully anaerobic experiments (figure 7.9). This suggests that the adaptation to aerobiosis is faster than to anaerobiosis, *i.e.* during the anaerobic phase many aerobic pathways are still active, while the anaerobic pathways are disabled as soon as oxygen is blown through the reactor.

In the P04 experiment, this overlapping is complete: *frd*, *cyd* are both active at the same time. Frd (fumarate reductase) and cytochrome d are normally active only under anaerobic and microaerobic conditions. Cytochrome d oxidises NADH but does not contribute to the H^+ gradient needed for ATP generation and the action of fumarate reductase is countered by succinate dehydrogenase. In the P30 experiment, an extra futile cycle is added: *acs*, encoding acetyl-CoA synthetase, responsible under aerobic conditions for the uptake of acetate, is upregulated in the aerobic phase of the oscillation. This way, the acetate produced during the anaerobic phase is remetabolised, at the cost of two ATPs.

One can conclude that short interruptions of oxygen supply do not trigger much acetate production and the loss of productivity is then caused by cytochrome d, wasting NADH. Longer interruptions of oxygen supply trigger acetate formation. When reentering aerobic conditions, this acetate triggers the expression of *acs*, creating a futile cycle.

The expected increase in succinate was not obtained. On the other hand, the increase in *ppc* was not as large as expected to produce more succinate. Combined with the downregulation of the glyoxylate pathway during the anaerobic phase, not much more carbon is entering the Krebs cycle under oscillating conditions than when grown fully aerobically.

Chapter 8

General conclusions and perspectives

Different experimental conditions, genetic and environmental, were combined with different mathematical approaches to understand the metabolism of *Escherichia coli* with the goal of its optimisation for the production of interesting compounds.

8.1 Mathematical techniques

The central piece in the mathematical techniques used in this work, is the metabolic model. Conceptually, this is not a very complex model: a matrix in which the rows represent the metabolites and the columns refer to the reactions. However, due to the size of this matrix, errors very easily leap in. This was countered in two ways. First, software was written to convert a list of reactions to a stoichiometric matrix. A list of reactions is simpler to manage and also clearer to the user's thinking than a large matrix of mostly zeros. Secondly, different model checking techniques were implemented: the elemental consistency test checks whether the stoichiometry of the metabolic model is correct. The dead-end test checks whether the set of reactions makes sense. Parallel pathways are reported so that potential underdetermined parts of the model can be found. Care was taken to make the software modular and the results easily interpretable: reaction names and metabolite names are reported instead of row or column numbers so that the use of coding schemes for reaction and metabolites is unnecessary.

Metabolic flux analysis (MFA) allows calculating intracellular fluxes when only exchange fluxes are measured. Furthermore, with MFA it is possible to balance *a priori* contradicting measurements and with the vector comparison test it can be investigated which measurements are really wrong. Again in the software implementation, reaction and metabolite names are reported and not matrix row or column numbers, making interpretation of the results easier.

MFA was successfully applied (1) to elucidate the difference in shikimate fluxes between carbon and phosphorus limited cultures and (2) to investigate any negative effect of mutation introduced into *E. coli* strains to reduce acetate production.

A major limitation of MFA is that it is only applicable on chemostat cultures. However, interesting phenomena can be observed under transient conditions. Therefore, MFA was extended to dynamic MFA. Fluxes were calculated based on the derivative of time series of concentration measurements. Those fluxes could subsequently be used for MFA, again generating time series of intracellular fluxes. The technique was successfully applied to transient experiments where the limiting substrate was changed from glucose to ammonia and *vice versa* and on step experiments where the comparison was made in recombinant protein production capabilities between a wild-type strain and a *ppc* overexpressing one. However, it had to be concluded that the explaining power of (dynamic) MFA on secondary metabolism is not as high as on primary metabolites. The fluxes to secondary metabolites are very small compared to the ones in primary metabolism and thus do not influence the primary metabolic routes much.

One bottleneck of dynamic MFA is that time series of measurements are noisy and should be filtered before taking the derivative for calculating fluxes. This filtering was found to be very tricky: not much data is available (each measurement is very expensive in both money and time), and thus the classical filtering techniques are not really suitable. In this work, polynomial filtering was used, using a moving window approach. However, the fine tuning of the parameters (degree of polynomial, size of moving window) was time consuming and the results were not always satisfying. More research is needed to find a suitable method for smoothing the time series of metabolite concentration measurements. Probably the best approach will be to manually draw bezier curves.

Another extension of the smoothing that should be pursued is the error propagation from the raw measurement data to the MFA calculated fluxes. This error propagation was rigorously implemented for pure MFA but was abandoned in the polynomial smoothing because of its complexity. MFA and especially flux reconciliation is heavily based on the covariance matrix. In the transient experiments performed in this work, the covariance matrix of steady state experiments whose fluxes were measured with the same equipment, was used as an approximation. However it would be better to derive this matrix from the polynomial smoothing itself. Confidence intervals could then be calculated from which the variance of an interpolated point could be derived.

8.2 *E. coli* as production host

Different cultivation techniques were used to increase understanding of *E. coli* metabolism and find routes for increased production. In a first set of experiments chemostats were used

to assess the difference in metabolism between (1) different environmental conditions and (2) different mutations. The second set of experiments consisted of switching the environmental conditions from one state to another state and looking at their effect on metabolism. In the third cultivation technique, the switching was periodic and *E. coli* cell cultures were subjected to oscillations in oxygen supply.

8.2.1 Steady state experiments

E. coli can be used to produce many different compounds. In this work, different conditions, genetic and environmental, were used to try to influence its metabolism. In chapter 3 it was shown, by using MFA, that the lower shikimate yield in carbon-limited cultures compared to carbon-abundant ones was not due to a lower carbon flux going into the aromatic amino acid production route but because more excretion of shikimate precursors. The flux in carbon-limited cultures, to the aromatic pathway was even higher than in phosphate-limited ones. Furthermore, with the aid of metabolic flux analysis, it was possible to determine the maintenance requirements under the different conditions. It was concluded that *E. coli*, when cultivated under carbon-abundant conditions, utilises the maximally possible amount of ATP. The non-growth-associated maintenance of P-limited cultures was as large as the total maintenance requirements of C-limited cells at their maximal growth rate.

When using *E. coli* as production host, acetate is an unwanted by-product. Different genetic strategies to reduce acetate production, were investigated (chapter 4). Three strains were assessed: one in which the acetate pathway was knocked out, one in which the *ppc* gene (encoding PEP carboxylase, that converts PEP and CO_2 to oxaloacetate) was overexpressed and one in which both mutations were combined. Simply blocking the acetate pathway did nothing to the fundamental cause of the overflow metabolism but increased lactate production. More carbon had to be diverted to the citric acid cycle. Indeed the PPC strain was found to be a favorable production strain: less overflow metabolites were excreted. Further characterisation of the strains in chemostat experiments showed no difference in flux distribution between the different mutants, indicating that the mutations caused no negative impact on primary metabolism.

8.2.2 Transient experiments

Production of acetate could be reduced by overexpressing *ppc*, increasing the flux towards the Krebs cycle. In a second experiment, it was demonstrated that this increased flux towards the Krebs cycle is also advantageous for protein production. Two strains were compared: one wild-type and one in which *ppc* was overexpressed. β-galactosidase was used as model protein and in both strains a *LacZ* plasmid was introduced. After attaining steady state, expression of *LacZ* was induced by supplying IPTG to the cultures. The resulting production profile was

monitored and the experimental data were analysed with dynamic MFA. It was shown that the *ppc* overexpression mutant, had a significant increase in protein production capabilities.

The influence of switching NH_3-limited cultures to glucose limitation and *vice versa* was investigated using dynamic MFA (chapter 5). When switching a N-limited culture to N-abundant conditions, a lag phase was observed in which the growth almost completely stopped and ATP demand was high. No clear reason could be given for this. Such a lag phase was only observed before in knock-outs of *glnE*, whose gene product, ATAse, regulates the activity of glutamine synthetase. Toxicity in that case was not due to ammonia itself, but to the accumulation of glutamine/glutamate as glutamine synthetase was not downregulated anymore.

The extended period of the lag phase (5 hours) suggests that the cause is probably genetic. With microarrays, the possible genes involved could be tracked in future studies. However, it is quite certain that many genes will be influenced by the switch from nitrogen-limiting conditions to nitrogen-abundant ones. It will thus still be difficult to find the mechanism behind this lag phase. Another possibility would be to create mutants and asses their behaviour. However, it is not clear at this stage which mutations would possibly reveal the cause of this lag phase. Furthermore, the strain is behaving normally under other conditions, and thus the only way of knowing the influence of a mutation on the lag phase is redoing that transient experiment. Not really a fast screening method...

8.2.3 Oscillating environmental conditions

The influence of oscillating *E. coli* cultures between aerobiosis and anaerobiosis was investigated using microarrays. It was found that cells adapt faster to aerobic conditions than to anaerobic ones: the experiment in which the period of oscillation was four minutes, had a gene expression profile more similar to fully aerobic cultures than to fully anaerobic ones. However, cytochrome d, responsible for NADH oxidation under microaerobic conditions, and fumarate reductase were found to be more expressed. Cytochrome d oxidises NADH but does not contribute to the H^+ gradient needed for ATP production. This could explain the lower biomass yield for the P04 experiments compared to fully aerobic cultures.

In the experiment with a 30 minutes period, the similarity of the gene expression profile between cells after 15 minutes aerobiosis and a fully aerobic culture was higher than between cells after 15 minutes of anaerobiosis and a fully anaerobic culture. However, unlike in the P04 experiments, different genes were expressed in the aerobic and anaerobic phase. Furthermore, besides the upregulation of cytochrome d, a second inefficiency was detected: acetate produced in the anaerobic phase activates the expression of *acs*, whose gene product is responsible for the uptake of acetate at the cost of two ATPs. This could explain the even lower biomass yield obtained in the P30 experiments compared to the P04 ones.

The expected increase in succinate production combined with a relatively fast growth was not observed when oscillating the *E. coli* cells between aerobic and anaerobic conditions. Different strategies could be tried to try to increase the succinate production:

- As adaptation to aerobic conditions is faster than to anaerobic conditions, the aerobic phase could be shortened.
- A major futile cycle is introduced by the reasimilation of acetate via acetyl-coA synthase. Knocking out *acs* will only increase acetate production. If succinate has to be produced, more carbon should be diverted from acetate towards the Krebs cycle, and the same strategy as for enhancing recombinant protein production can be used. Hence the *E. coli* MG1655 Δ*ackA–pta*, Δ*poxB*, Δp*ppc ppc–p37* strain of chapter 4 would be an ideal candidate to redo this oxygen oscillating experiment with.
- The *acs* geneproduct typically scavenges carbon. It could be expected that it is not as highly expressed under carbon-abundant conditions as under carbon-limiting ones, like *ptsG* in phosphate-limited cultures is not as highly expressed as in carbon-limited cultures. Hence, instead of working with carbon-limited cultures, P- or N-limited ones could be envisaged.

Appendix A

Variance covariance calculations

All the variance covariance calculations are done based on linearisation of the functions. Given the variables $x_1 \dots x_n$ with their variance covariance matrix:

$$S_x = \begin{bmatrix} \sigma_{1,1}^2 & \cdots & \sigma_{1,n} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1} & \cdots & \sigma_{n,n}^2 \end{bmatrix} \tag{A.1}$$

and the transformation to other variables y_i, with the functions $y_i = f_i(x_1 \dots x_n)$ for $i = 1 \dots m$, the jacobian matrix is defined as:

$$J = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_n} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{A.2}$$

and the new variance covariance matrix is then equal to:

$$S_y = J \; S_x \; J^{\mathsf{T}} \tag{A.3}$$

This general formula is illustrated in the examples below.

A.1 Sum

Given a with σ_a^2 and b with σ_b^2 and assuming there is no correlation between a and b, the variance of $a + b$ is calculated as:

$$\sigma_{a+b}^2 = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \sigma_a^2 + \sigma_b^2 \tag{A.4}$$

A.2 Substraction

Given a with σ_a^2 and b with σ_b^2 and assuming there is no correlation between a and b, the variance of $a - b$ is calculated as:

$$\sigma_{a-b}^2 = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \sigma_a^2 + \sigma_b^2 \tag{A.5}$$

A.3 Multiplication

Given a with σ_a^2 and b with σ_b^2 and assuming there is no correlation between a and b, the variance of $a * b$ is calculated as:

$$\sigma_{a\,b}^2 = \begin{bmatrix} b & a \end{bmatrix} \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \sigma_a^2 b^2 + \sigma_b^2 a^2 \tag{A.6}$$

A.4 Division

Given a with σ_a^2 and b with σ_b^2 and assuming there is no correlation between a and b, the variance of a/b is calculated as:

$$\sigma_{a/b}^2 = \begin{bmatrix} \frac{1}{b} & -\frac{a}{b^2} \end{bmatrix} \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix} \begin{bmatrix} \frac{1}{b} \\ -\frac{a}{b^2} \end{bmatrix} = \frac{\sigma_a^2}{b^2} + \sigma_b^2 \frac{a^2}{b^4} \tag{A.7}$$

Appendix B

List of abbreviations

B.1 List of metabolites

2PG	$C_3H_7O_7P$	2-phophoglycerate
3PG	$C_3H_7O_7P$	3-phophoglycerate
6PG	$C_6H_{13}O_{10}P$	6-phosphogluconate
6PGL	$C_6H_{11}O_9P$	6-phosphogluconolacton
Ac	$C_2H_4O_2$	Acetate
AcACP	C_2H_3OPept	Acetyl ACP
AcCoA	$C_{23}H_{34}O_{17}N_7P_3S$	Acetyl CoA
ACP	HPept	Acyl carier protein
ADP	$C_{10}H_{15}O_{10}N_5P_2$	Adenosine diphosphate
ADPHEP	$C_{17}H_{27}O_{16}N_5P_2$	ADP-Mannoheptose
AICAR	$C_9H_{15}O_8N_4P$	Amino imidazole carboxamide ribonucleotide
aKGA	$C_5H_6O_5$	Alpha keto glutaric acid
aKIV	$C_5H_8O_3$	Alpha-keto-isovalerate
Ala	$C_3H_7O_2N$	Alanine
AMP	$C_{10}H_{14}O_7N_5P$	Adenosine monophosphate
Ar5P	$C_5H_{11}O_8P$	Arabinose-5-phosphate
Arg	$C_6H_{14}O_2N_4$	Arginine
Asn	$C_4H_8O_3N_2$	Asparagine
Asp	$C_4H_7O_4N$	Aspartate
AspSA	$C_4H_7O_3N$	Aspartate semialdehyde
ATP	$C_{10}H_{16}O_{13}N_5P_3$	Adenosine triphosphate
BGalAse	$C_{4.98}H_{7.58}O_{1.5}N_{1.41}S_{0.0507}$	Beta-galactosidase
Biom	$CH_{1.63}O_{0.392}N_{0.244}P_{0.021}S_{0.00565}$	Biomass
Biomn	$CH_{1.91}O_{0.506}N_{0.252}P_{0.0149}S_{0.0067}$	Biomass
BPG	$C_3H_8O_{10}P_2$	1-3-biphosphoglycerate
C120ACP	$C_{12}H_{23}OPept$	
C140ACP	$C_{14}H_{27}OPept$	
C160ACP	$C_{16}H_{31}OPept$	
C181ACP	$C_{18}H_{33}OPept$	
CarP	CH_4O_5NP	Carbamoyl phosphate
CDP	$C_9H_{15}O_{11}N_3P_2$	Citidine diphosphate
CDPDGo	$C_{46}H_{83}O_{15}N_3P_2$	CDP-diacylglycerol
CDPEthAn	$C_{11}H_{20}O_{11}N_4P_2$	CDP-ethanolamine
Chor	$C_{10}H_{10}O_6$	Chorismate
Cit	$C_6H_8O_7$	cisaconitate
CL	$C_{77}H_{144}O_{16}P_2$	Cardiolipin
CMP	$C_9H_{14}O_8N_3P$	Citidine monophosphate
CMPKDO	$C_{17}H_{26}O_{15}N_3P$	CMP-2-keto-3-deoxyoctanoate
CO2	CO_2	Carbondioxide
CoA	$C_{21}H_{32}O_{16}N_7P_3S$	Coenzyme A
CTP	$C_9H_{16}O_{14}N_3P_3$	Citidine triphosphate
Cys	$C_3H_7O_2NS$	Cysteine
dADP	$C_{10}H_{15}O_9N_5P_2$	deoxy ADP
Dahp	$C_7H_{13}O_{10}P$	Deoxy arabino heptulosonate
dATP	$C_{10}H_{16}O_{12}N_5P_3$	deoxy ATP
dCDP	$C_9H_{15}O_{10}N_3P_2$	deoxy CDP
dCTP	$C_9H_{16}O_{13}N_3P_3$	deoxy CTP
dGDP	$C_{10}H_{15}O_{10}N_5P_2$	deoxy GDP
DGo	$C_{37}H_{70}O_5$	Diacyl glycerol
dGTP	$C_{10}H_{16}O_{13}N_5P_3$	deoxy GTP
DHAP	$C_3H_7O_6P$	Dihydroxyaceton phosphate
DHF	$C_{19}H_{21}O_6N_7$	Dihydrofolate
Dhq	$C_7H_{10}O_6$	Dehydroquinate
Dhs	$C_7H_8O_5$	Dehydroshikimate
DNA	$C_{9.75}H_{14.2}O_7N_{3.75}P$	DNA composition
dTDP	$C_{10}H_{16}O_{11}N_2P_2$	deoxy TDP
dTMP	$C_{10}H_{15}O_8N_2P$	deoxy TMP
dTTP	$C_{10}H_{17}O_{14}N_2P_3$	deoxy TTP
dUDP	$C_9H_{14}O_{11}N_2P_2$	deoxy UDP
dUMP	$C_9H_{13}O_8N_2P$	deoxy UMP
dUTP	$C_9H_{15}O_{14}N_2P_3$	deoxy UTP
E4P	$C_4H_9O_7P$	Erythrose-4-phosphate
Eth	C_2H_6O	Ethanol
F6P	$C_6H_{13}O_9P$	Fructose-6-phosphate
FA	CH_2O_2	Formic Acid
FAD	$C_{27}H_{33}O_{15}N_9P_2$	Flavine adeninen dinucleotide
FADH2	$C_{27}H_{35}O_{15}N_9P_2$	
FBP	$C_6H_{14}O_{12}P_2$	Fructose-1-6-biphosphate
FTHF	$C_{20}H_{23}O_7N_7$	Formyl tetrahydrofolate
Fum	$C_4H_4O_4$	Fumarate
G1P	$C_6H_{13}O_9P$	Glucose-1-phosphate
G3P	$C_3H_7O_6P$	Glyceraldehyde-3-phosphate
G6P	$C_6H_{13}O_9P$	Glucose-6-phosphate

GA1P	$C_6H_{14}O_8NP$	D-glucosamine-6-phosphate
GA6P	$C_6H_{14}O_8NP$	D-glucosamine-6-phosphate
GDP	$C_{10}H_{15}O_{11}N_5P_2$	Guanosine diphosphate
GLC	$C_6H_{12}O_6$	Glucose
Glcg	$C_6H_{10}O_5$	Glycogen
Gln	$C_5H_{10}O_3N_2$	Glutamine
Glu	$C_5H_9O_4N$	Glutamate
Gly	$C_2H_5O_2N$	Glycine
GMP	$C_{10}H_{14}O_8N_5P$	Guanosine monophosphate
Go3P	$C_3H_9O_6P$	Glycerol-3-phosphate
GTP	$C_{10}H_{16}O_{14}N_5P_3$	Guanosine triphosphate
H2CO3	CH_2O_3	Bicarbonate
H2O	H_2O	Water
H2O2	H_2O_2	
H2S	H_2S	Hydrogene sulfide
H2SO4	H_2O_4S	Sulfuric acid
His	$C_6H_9O_2N_3$	Histidine
H	H^+	Hydrogene
HSer	$C_4H_9O_3N$	Homoserine
iCit	$C_6H_8O_7$	isocitraat
Ile	$C_6H_{13}O_2N$	Isoleucine
IMP	$C_{10}H_{13}O_8N_4P$	Inosine monophosphate
Lac	$C_3H_6O_3$	Lactate
Leu	$C_6H_{13}O_2N$	Leucine
Lipa	$C_{110}H_{196}O_{32}N_2P_2$	Lipid A
Lipid	$C_{40.2}H_{77.6}O_{8.41}N_{0.771}P_{1.03}$	Lipid composition
Lps	$C_{171}H_{298}O_{81}N_4P_2$	Lipo Poly sacharide
Lys	$C_6H_{14}O_2N_2$	Lysine
Mal	$C_4H_6O_5$	Malate
MalACP	$C_3H_3O_3Pept$	Malonyl ACP
MalCoA	$C_{24}H_{34}O_{19}N_7P_3S$	Malonyl CoA
MDAP	$C_7H_{14}O_4N_2$	Meso-diaminopimelate
Met	$C_5H_{11}O_2NS$	Methionine
MeTHF	$C_{20}H_{23}O_6N_7$	Methyleen tetrahydro folate
MTHF	$C_{20}H_{25}O_6N_7$	Methyl tetrahydrofolate
NAD	$C_{21}H_{28}O_{14}N_7P_2^+$	Nicotinamide adenine dinucleotide
NADH	$C_{21}H_{29}O_{14}N_7P_2$	
NADP	$C_{21}H_{28}O_{17}N_7P_3^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	$C_{21}H_{29}O_{17}N_7P_3$	
NH3	H_3N	Ammonia
O2	O_2	Oxygen
OAA	$C_4H_4O_5$	Oxaloacetate
Orn	$C_5H_{12}O_2N_2$	Ornithine
PA	$C_{37}H_{71}O_8P$	Phosphatidyl acid
PAP	$C_{10}H_{15}O_{10}N_5P_2$	Phospho adenosine phosphate
PEP	$C_3H_5O_6P$	Phosphoenolpyruvate
Peptido	$C_{35}H_{53}O_{16}N_7$	Peptidoglycane
PEthAn	$C_{39}H_{76}O_8NP$	Phosphatidyl ethanolamine
PG	$C_{40}H_{75}O_9P$	Phosphatidyl glycerol
Phe	$C_9H_{11}O_2N$	Phenylalanine
PiOH	H_3O_4P	Phosphate
PPiOH	$H_4O_7P_2$	Pyrophosphate
Pro	$C_5H_9O_2N$	Proline
Prot	$C_{4.8}H_{7.67}O_{1.4}N_{1.37}S_{0.046}$	Protein composition
Protn	$C_{4.8}H_{9.67}O_{2.4}N_{1.37}S_{0.046}$	Protein composition
PRPP	$C_5H_{13}O_{14}P_3$	5-phospho-alpha-D-ribosyl-1-pyrophosphate
PSer	$C_{40}H_{76}O_{10}NP$	Phosphatidyl Serine
Pyr	$C_3H_4O_3$	Pyruvate
Qa	$C_7H_{12}O_6$	Quinate
R5P	$C_5H_{11}O_8P$	Ribose-5-phosphate
Rl5P	$C_5H_{11}O_8P$	Ribulose-5-phosphate
RNA	$C_{9.58}H_{13.8}O_{7.95}N_{3.95}P$	RNA composition
S7P	$C_7H_{15}O_{10}P$	Sedoheptulose-7-phosphate
Ser	$C_3H_7O_3N$	Serine
Shi	$C_7H_{10}O_5$	Shikimate
Shi3P	$C_7H_{11}O_8P$	Shikimate-3-phosphate
Suc	$C_4H_6O_4$	Succinate
SucCoA	$C_{25}H_{36}O_{19}N_7P_3S$	Succinyl CoA
THF	$C_{19}H_{23}O_6N_7$	Tetrahydrofolate
Thiored	Pept	Thioredoxin
ThioredH2	H_2Pept	Reduced thioredoxin
Thr	$C_4H_9O_3N$	Threonine
Trp	$C_{11}H_{12}O_2N_2$	Tryptophan
Tyr	$C_9H_{11}O_3N$	Tyrosine
UDP	$C_9H_{14}O_{12}N_2P_2$	Uridine diphosphate
UDPGlc	$C_{15}H_{24}O_{17}N_2P_2$	UDP glucose
UDPNAG	$C_{17}H_{27}O_{17}N_3P_2$	UDP N-acetyl glucosamine
UMP	$C_9H_{13}O_9N_2P$	Uridine monophosphate
UTP	$C_9H_{15}O_{15}N_2P_3$	Uridine triphosphate
Val	$C_5H_{11}O_2N$	Valine
XMP	$C_{10}H_{13}O_9N_4P$	Xanthosine-5-phosphate
Xu5P	$C_5H_{11}O_8P$	Xylulose-5-phosphate

B.2 Full name of the reactions

A5PIR	Arabinose 5 phosphate isomerase
AcACPSY	beta-ketoacyl-ACP synthase
AcCoACB	AcCoA carboxylase
AcCoATA	AcCoA-ACP transacetylase
AcdhDH	Acetaldehyde dehydrogenase
AcKNLR	Phosphate acetyl transferase + acetate kinase
ACO	Aconitase
ActSY	Acetoin synthesis
AcylTF	Acyltransferase
AdKN	Adenylaat kinase
ADPHEPSY	ADP-L-glycero-D-mannoheptose-6-epimerase
ADPRD	ADP reductase
AICARSYLR	AICAR synthesis (lumped reaction)
AKGDH	alfa-ketoglutate dehydrogenase
aKIVSYLR	alfa-ketoIsoValerate synthesis (lumped reaction)
AlaTA	Alanine transaminase
ALD	Aldolase
AMPSYLR	AMP synthesis (lumped reaction)
ArgSYLR	Arginine synthesis (lumped reaction)
AspLI	Aspartate Amonia ligase
AspSASY	Aspartate Semi aldehyde synthese (lumped reaction)
AspSY	Asparagine synthetase
AspTA	Aspartate transaminase
ATPHY	ATP hydrolysis
BGalAseSYLR	BGalAse synthesis (lumped reaction)
BiomSYLR	Biomass synthesis
BkaSYLR	Beta ketoadipate synthesis (lumped reaction)
C120SY	C12.0-ACP synthesis
C140SY	C14.0-ACP synthesis
C141SY	C14.1-ACP synthesis
C160SY	C16.0-ACP synthesis
C161SY	C16.1-ACP synthesis
C181SY	C18.1-ACP synthesis
CarPSY	Carbamoyl phosphate synthase
CDPDGoSY	CDP-Diacylglycerol synthetase
CDPKN	CDP kinase

CDPPT	CDP phosphatase
CDPRD	CDP reductase
ChorSYLR	Chorismate synthesis (lumped reaction)
CitDH	Isocitrate dehydrogenase
CitSY	Citrate synthase
CLSY	Cardiolipin synthase
CMPKDOSYLR	CMPKDO synthesis (lumped reaction)
CMPKN	CMP kinase
CoQ2NAD	Quinone reductase
CTPSY	CTP synthase
CysSYLR	Cysteine synthesis (lumped reaction)
dADPKN	dADP kinase
dADPPT	dADP phosphatase
dCDPKN	dCDP kinase
dCDPPT	dCDP phosphatase
dCTPDA	dCTP deaminase
dGDPKN	dGDP kinase
dGDPPT	dGDP phosphatase
DGoKN	Diacylglycerol kinase
DhDoPHepAD	Dehydro deoxyphosphoheptonate aldolase
DHFRD	DHF reductase
DhqDH	Dehydro quinate dehydrogenase
DhqSY	Dhq synthase
DhsDH	Dhs dehydratase
DhsSYLR	Dhs synthesis
DNASYLR	DNA formation (lumped reaction)
dTDPKN	dTDP kinase
dTDPPT	dTDP phosphatase
dTMPKN	dTMP kinase
dTMPSY	Thymidilate synthase
dUDPKN	dUDP kinase
dUTPPPAS	dUTP pyrophosphatase
ENO	Enolase
EthANPT	Ethanolamine phosphotransferase
EthDH	Ethanol dehydrogenase
EthDHLR	Ethanol dehydrogenase (lumped reaction)
FAD2NAD	FAD2NAD
FBPAS	Fructose-1,6-bisphosphatase
FTHFDF	FTHF deformylase
FTHFLY	Formate THF ligase
FTHFSYLR	FTHF synthesis (lumped reaction)
FumHY	Fumarate hydratase
G3PDH	Glyceraldehyde-3-phosphate dehydrogenase
G6PDH	Glucose-6-phosphate dehydrogenase
GallicSY	Gallic acid formation
GDPKN	GDP kinase
GDPRD	GDP reductase
GlcAnMU	Phosphoglucosamine mutase
GlcgSY	Glycogen synthase
GlnF6PTA	Glutamine fructose-6-phosphate transaminase
GluDH	Glutamate dehydrogenase
GluLI	Glutamate-ammonia ligase
GluSY	Glutamate synthase
GlyCA	gcv system
GMPSY	GMP Synthase
Go3PDH	Glycerol-3-phosphate dehydrogenase
GuKN	Guanilate kinase
H2CO3SY	CO2 - HCO3 equilibration reaction
H2O2ox	hydroperoxidase
H2SSYLR	H2S synthesis (lumped reaction)
HisSYLR	Histidine biosynthesis (lumped reaction)
HK	Hexokinase
HSerDH	Homoserine dehydrogenase
iCitL	Isocitrate lyase
IleSYLR	Isoleucine synthesis (lumped reaction)
IMPDH	IMP dehydrogenase
IMPSYLR	IMP synthesis (lumped reaction)
LacDH	Lactate dehydrogenase
LAS	Lactonase
LeuSYLR	Leucine synthesis (lumped reaction)
LipaSYLR	Lipid A synthesis (lumped reaction)
LipidSYLR	Lipid formation (lumped reaction)
LysSY	Diaminopimelate decarboxylase
MalCoATA	MalCoA ACP transacetylase
MalDH	Malate dehydrogenase
MalSY	Malate synthase
MDAPSYLR	MDAP synthesis (lumped reaction)
MeTHFRD	MeTHF reductase
MetSYLR	Methionine synthesis (lumped reaction)
NADH2NADPH	NADH2NADPH
NAGUrTF	N-Acetylglucosamine-1-phosphate-uridyltransferase
OrnSYLR	Ornithine synthesis (lumped reaction)
PAPNAS	3' - 5' Bisphosphate nucleotidase (PAP degradation)
PEPCB	PEP-carboxylase
PEPCBKN	PEP-carboxykinase
PFK	Phosphofructokinase
PFLY	Pyruvate formate lyase
PGDH	6-phosphogluconaat dehydrogenase
PGI	Phophoglucoisomerase
PGK	Phosphoglycerate kinase
PGLCMT	Phosphoglucomutase
PGM	Phosphoglycerate mutase
PGSYLR	PG synthesis (lumped reaction)
PheSYLR	Phenylalanine biosynthesis (lumped reaction)
PPE	Phospho pentose epimerase
PPI	Phospho pentose isomerase
PPiOHHY	PPiOH hydrolase
ProSYLR	Prolyne biosynthesis (lumped reaction)
ProtoCatDC	ProtoCat decarboxylase
ProtSYLR	Protein formation (lumped reaction)
PrppSY	Phosphoribosyl pyrophosphate synthase
PSerDC	Phosphatidylserine decarboxylase
PSerSY	Phosphatidylserine synthase
PTS	Phosphoenoltransferase
PyrD	Pyruvate dehydrogenase
PyrK	Pyruvate kinase
PyrMalCB	Pyruvic-malic carboxylase
R5P2R1P	Ribose-5-phosphate conversion to ribose-1-phosphate
Resp	Respiration
RestSYLR	Rest
RNASYLR	RNA formation (lumped reaction)
SerLR	Serine biosynthesis
SerTHM	Serine transhydroxy methylase
ShiKN	Shikimate kinase
ShiSY	Shikimate synthesis
SucCoASY	SuccinylCoA synthetase
SucDH	Succinate dehydrogenase
TA	Transaldolase
ThioredRD	Thioredoxine reductase
ThrSYLR	Threonine synthesis (lumped reaction)
TK1	Transketolase1
TK2	Transketolase2
TPI	Triose phosphate isomerase
TrpSYLR	Tryptophan biosynthesis
TyrSYLR	Tyrosine biosynthesis (lumped reaction)
UDPGlcSY	UDP glucose-1-phosphate uridylyltransferase
UDPKN	UDP kinase
UDPRD	UDP reductase
UMPSYLR	UMP synthesis (lumped reaction)
UrKN	Uridilate kinase
ValAT	Branched chain AA aminotransferase
ValPyrAT	Valine Pyruvate aminotransferase

B.3 List of genes

Gene functions were taken from the EcoCyc database (Keseler *et al.*, 2005).

aceA	Isocitratelyase
aceB	Malate synthase
aceF	Lipoate acetyltransferase / dihydrolipoamide acetyltransferase, subunit of pyruvate dehydrogenase multienzyme complex
ackA	Acetate kinase
acn	Aconitase
acs	AcetylCo synthase
adhCEP	Ethanol dehydrogenase
adhE	Acetyldehydrogenase
aldH	Aldehyde dehydrogenase
amtB	AmtB ammonium Amt transporter
arcA	ArcA transcriptional dual regulator
arcB	ArcB sensory histidine kinase
aroF	Subunit of 2-dehydro-3-deoxyphosphoheptonate aldolase
aroG	Subunit of 2-dehydro-3-deoxyphosphoheptonate aldolase
aroL	shikimate kinase II
cyd	Cytochrome bd terminal oxidase
cyo	Cytochrome bo terminal oxidase
dctA	DctA dicarboxylate DAACS transporter
dms	Dimethyl sulfoxide reductase
eno	Enolase
fba	Fructosebisphosphate aldolase
fnr	FNR transcriptional dual regulator
frd	Fumarate reductase
fumA	Fumarase
fumB	Fumarase B
fumC	Fumarase C
gapA	Glyceraldehyde-3-phosphate dehydrogenase
glk	Glucokinase
glnA	Adenylyl-glutamine synthetase
glnB	PII
glnD	Uridylyltransferase / uridylyl-removing enzyme
glnE	Glutamine synthetase adenylyltransferase
glnG	Nitrogen regulatory protein NtrC
glnK	Nitrogen regulatory protein GlnK
glnL	Nitrogen regulatory protein NtrB
gltA	Citrate synthase
icd	Isocitrate dehydrogenase
kdgT	2-dehydro-3-deoxy-D-gluconate transporter
lacZ	Betagalactosidase
lct	LldP lactate transporter
ldhA	Lactate dehydrogenase
lpd	Subunit of lipoamide dehydrogenase, 2-oxoglutarate dehydrogenase complex, gcv system and pyruvate dehydrogenase multienzyme complex
lrp	Lrp transcriptional dual regulator
maeAB	Malic enzyme
mdh	Malate dehydrogenase
mgsA	Methylglyoxal synthase
narGHJI	Nitrate reductase A
narZYWV	Nitrate reductase Z
pck	Phosphoenolpyruvatecarboxykinase
pdh	Pyruvate dehydrogenase
pdhR	PdhR transcriptional dual regulator
pfkAB	6-phophofructokinase
pfl	Pyruvateformatelyase
pgi	Phosphogluco-isomerase
pgk	Phophoglyceratekinase
poxB	Pyruvate oxidase
ppc	Phosphoenolpyruvatecarboxylase
pta	Acetylphosphotransferase
ptsG	PTS Enzyme IIB and IIC domains
ptsH	HPr (heat stable, histidyl phosphorylatable protein)
pyk	Pyruvate kinase
rph	RNase PH monomer
sdhABCD	Succinate dehydrogenase
sucAB,lpd	Alfaketoglutarate dehydrogenase
sucCD	Succinate thiokinase
tnaA	L-cysteine desulfhydrase / tryptophanase
torA	Subunit of trimethylamine N-oxide reductase I
trpE	Anthranilate synthase component I
ydiB	Shikimate dehydrogenase / quinate dehydrogenase
yfiD	Stress-induced alternate pyruvate formate-lyase subunit

Bibliography

S. Achebach, T. Selmer & G. Unden (2005). Properties and significance of apoFNR as a second form of air-inactivated [4Fe-4S].FNR of *Escherichia coli*. *FEBS Journal*, 272:4260–4269.

S. Achebach, Q. H. Tran, A. Vlamis-Gardikas, M. Müllner, A. Holmgren & G. Unden (2004). Stimulation of Fe-S cluster and insertion into apoFNR by *Escherichia coli* glutaredoxins 1, 2 and 3 in vitro. *FEBS Letters*, 565:203–206.

Affymetrix (2002). *Statistical Algorithms Description Document*. `http://www.affymetrix.com/support/technical/whitepapers/sadd_whitepaper.pdf`.

M. Åkesson, J. Förster & J. Nielsen (2004). Integration of gene expression data into genome-scale metabolic models. *Metabolic Engineering*, 6:285–293.

M. Akesson, P. Hagander & J. P. Axelsson (2001). Avoiding acetate accumulation in *Escherichia coli* cultures using feedback control of glucose feeding. *Biotechnology and Bioengineering*, 73:223–230.

S. Alexeeva, B. De Kort, G. Sawers, K. J. Hellingwerf & M. J. Teixeira De Mattos (2000). Effects of limited aeration and of the ArcAB system on intermediary pyruvate catabolism in *Escherichia coli*. *Journal of Biotechnology*, 182:4934–4940.

S. Alexeeva, K. J. Hellingwerf & M. J. Teixeira de Mattos (2002). Quantitative assessment of oxygen availability: Perceived aerobiosis and its effect on flux distribution in the respiratory chain of *Escherichia coli*. *Journal of Bacteriology*, 184:1402–1406.

S. Alexeeva, K. J. Hellingwerf & M. J. Teixeira de Mattos (2003). Requirement of ArcA for redox regulation in *Escherichia coli* under microaerobic but not anaerobic or aerobic conditions. *Journal of Bacteriology*, 185:204–209.

B. N. Ames & D. T. Dubin (1960). The role of polyamines in the neutralization of bacteriophage deoxyribonucleic acid. *The Journal of Biological Chemistry*, 253:769–775.

K. B. Andersen & K. von Meyenburg (1980). Are growth rates of *Escherichia coli* in batch cultures limited by respiration? *Journal of Bacteriology*, 144:114–123.

A. A. Aristidou, K.-Y. San & G. N. Bennett (1994). Modification of central metabolic pathway in *Escherichia coli* to reduce acetate accumulation by heterologous expression of *bacillus subtilis* acetolactate synthase gene. *Biotechnology and Bioengineering*, 44:944–951.

A. A. Aristidou, K.-Y. San & G. N. Bennett (1995). Metabolic engineering of *Escherichia coli* to enhance recombinant protein production through acetate reduction. *Biotechnology Progress*, 11:475–478.

A. A. Aristidou, K.-Y. San & G. N. Bennett (1999). Improvement of biomass yield and recombinant gene expression in *Escherichia coli* by using fructose as the primary carbon source. *Biotechnology Progress*, 15:140–145.

M. R. Atkinson, T. A. Blauwkamp, V. Bondarenko, V. Studitsky & A. J. Ninfa (2002). Activation of the *glnA*, *glnK* and *nac* promoters as *Escherichia coli* undergoes the transition from nitrogen excess growth to nitrogen starvation. *Journal of Bacteriology*, 184:5358–5363.

M. R. Atkinson & A. J. Ninfa (1998). Role of the GlnK signal transduction protein in the regulation of nitrogen assimilation in *Escherichia coli*. *Molecular Microbiology*, 29:431–447.

D. M. Bates, B. A. Lazazzera & P. Kiley (1995). Characterisation of FNR* mutant proteins indicates two distinct mechanisms for altering oxygen regulation of the *Escherichia coli* transcription factor FNR. *Journal of Bacteriology*, 177:3972–3978.

E. H. Battley (1995). An apparent anomaly in the calculation of ash-free dry weights for the determination of cellular yields. *Applied and Environmental Microbiology*, 61:1655–1657.

S. A. Benner & M. Sismour (2005). Synthetic biology. *Nature Reviews Genetics*, 6:533–543.

B. M. Bolstad, R. A. Irizarry, M. Åstrand & T. P. Speed (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19:185–193.

B. M. Bolstad, R. A. Irizarry, L. Gautier & Z. Wu (2005). Preprocessing high-density oligonucleotide arrays. In Gentleman *et al.* (2005), chapter 2, pp. 13–32.

D. T. Britto, M. Y. Siddiqi, A. D. Glass & H. J. Kronzucker (2001). Futile transmembrane NH_4^+ cycling: A cellular hypothesis to explain ammonium toxicity in plants. *Proceedings of the National Academy of Sciences of the United States of America*, 98:4255–4258.

R. Bueno, G. Pahel & B. Magasanik (1985). Role of *glnB* and *glnD* gene products in regulation of the *glnALG* operon of *Escherichia coli*. *Journal of Bacteriology*, 164:816–822.

H. F. Bunn & R. Poyton (1996). Oxygen sensing and molecular adaptation to hypoxia. *Physiological Reviews*, 76:839–885.

A. Burkovski (2003). I do it my way: Regulation of ammonium uptake and ammonium assimilation in *Corynebacterium glutamicum*. *Archives of Microbiology*, 179:83–88.

R. Carlson & F. Srienc (2004a). Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: Creation of overall flux states. *Biotechnology and Bioengineering*, 86:149–162.

R. Carlson & F. Srienc (2004b). Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: Identification of reactions. *Biotechnology and Bioengineering*, 85:1–19.

R. Chen, W. M. G. J. Yap, P. W. Postma & J. E. Bailey (1997). Comparative studies of *Escherichia coli* strains using different glucose uptake systems: Metabolism and energetics. *Biotechnology and Bioengineering*, 56:583–590.

S. E. Choe, M. Boutros, A. M. Michelson, G. M. Church & M. S. Halfon (2005). Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. *Genome Biology*, 6:R16.

C.-H. Chou, G. N. Bennett & K.-Y. San (1994). Effect of modulated glucose uptake on high-level recombinant protein production in a dense *Escherichia coli* culture. *Biotechnology Progress*, 10:644–647.

C. Colantuoni, G. Henry, C. M. L. S. Bouton, S. L. Zeger & J. Pevsner (2003). SNOMAD: Biologist-friendly web tools for the standardization and normalization of microarray data. In Parmigiani *et al.* (2003), chapter 9, pp. 210–228.

I. Compan & D. Touati (1994). Anaerobic activation of *arcA* transcription in *Escherichia coli*: roles of Fnr and ArcA. *Molecular Microbiology*, 11:955–964.

J. Contiero, C. M. Beatty, S. Kumari, C. L. DeSanti, W. R. Strohl & A. J. Wolfe (2000). Effects of mutations in acetate metabolism on high-cell-density growth of *Escherichia coli*. *Journal of Industrial Microbiology*, 24:421–430.

P. A. Cotter, V. Chepuri, R. B. Gennis & R. P. Gunsalus (1990). Cytochrome o (*cyoABCDE*) and d (*cydAB*) oxidase gene expression in *Escherichia coli* is regulated by oxygen, pH, and the *fnr* gene product. *Journal of Bacteriology*, 172:6333–6338.

G. Couts, G. Thomas, D. Blakey & M. Merrick (2002). Membrane sequestration of the signal transduction protein GlnK by the ammonium transporter AmtB. *The European Molecular Biology Organization Journal*, 21:536–545.

S. J. Cox, S. Shalel-Levanon, G. N. Bennett & K.-Y. San (2005). Genetically constrained metabolic flux analysis. *Metabolic Engineering*, 7:445–456.

S. J. Cox, S. Shalel-Levanon, A. Sanchez, H. Lin, B. Peercy, G. N. Bennett & K.-Y. San (2006). Development of a metabolic network design and optimization framework incorporating implementation constraints: A succinate production case study. *Metabolic Engineering*, 8:46–57.

A. J. Cozzone (1998). Regulation of acetate metabolism by protein phosphorylation in enteric bacteria. *Annual Review of Microbiology*, 52:127–164.

J. Crack, J. Green & A. Thomson (2004). Mechanism of oxygen-sensing by the bacterial transcription factor fumarate-nitrate reduction (FNR). *The Journal of Biological Chemistry*, 279:9278–9286.

K. A. Datsenko & B. L. Wanner (2000). 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.

E. De Clercq (2002). Strategies in the design of antiviral drugs. *Nature Reviews/Drug Discovery*, 1:13–25.

S. L. De Maeseneire, M. De Mey, S. Vandedrinck & E. J. Vandamme (2006). Metabolic characterisation of *E. coli* citrate synthase and phosphoenolpyruvate carboxylase mutants in aerobic cultures. *Biotechnology Letters*, 28:1945–1953.

M. De Mey, S. L. De Maeseneire, W. K. Soetaert & E. J. Vandamme (2007a). Minimizing acetate formation in *E. coli* fermentations. *Journal of Industrial Microbiology and Biotechnology*, 34. 689–700.

M. De Mey, G. J. Lequeux, J. J. Beauprez, J. Maertens, E. Van Horen, W. Soetaert, P. A. Vanrolleghem & E. J. Vandamme (2008). Transient metabolic modeling of *Escherichia coli* MG1655 and MG1655 $\Delta ackA - pta$, $poxB$ Δp$ppc - p37$ for recombinant β-galactosidase production. In preparation.

M. De Mey, G. J. Lequeux, J. J. Beauprez, J. Maertens, E. Van Horen, W. K. Soetaert, P. A. Vanrolleghem & E. J. Vandamme (2007b). Comparison of different strategies to reduce acetate formation in *Escherichia coli*. *Biotechnology Progress*, 23:1053–1063.

M. De Mey, J. Maertens, G. J. Lequeux, W. K. Soetaert & E. J. Vandamme (2007c). Construction and model-based analysis of a promoter library for *E. coli* : An indispensable tool for metabolic engineering. *BMC Biotechnology*, 7:34.

A. Death & T. Ferenci (1994). Between feast and famine: Endogenous inducer synthesis in the adaptation of *Escherichia coli* to growth with limiting carbohydrates. *Journal of Bacteriology*, 176:5101–5107.

N. N. Dedhia, T. Hottiger & J. E. Bailey (1994). Overproduction of glycogen in *Escherichia coli* blocked in the acetate pathway improved cell growth. *Biotechnology and Bioengineering*, 44:132–139.

J. C. Diaz Ricci, L. Regan & J. E. Bailey (1991). Effect of alteration of the acetic acid synthesis pathway on the fermentation pattern of *Escherichia coli*. *Biotechnology and Bioengineering*, 38:1318–1324.

C. R. Dittrich, R. V. Vadali, G. N. Bennett & K.-Y. San (2005). Redistribution of metabolic fluxes in the central aerobic metabolic pathway of *E. coli* mutant strains with deletion of the *ackA-pta* and *poxB* pathways for the synthesis of isoamyl acetate. *Biotechnology Progress*, 21:627–631.

K. Draths & J. Frost (1991). Conversion of D-glucose into catechol: The not-so-common pathway of aromatic biosynthesis. *Journal of the American Chemical Society*, 113:9361–9363.

K. Draths, D. Pompliano, D. Conley, J. Frost, A. Berry, G. Disbrow, R. Staversky & J. Lievense (1992). Biocatalytic synthesis of aromatics from D-glucose: The role of transketolase. *Journal of the American Chemical Society*, 114:3956–3962.

S. Dudoit, Y. H. Yang, M. J. Callow & T. P. Speed (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica*, 12:111–139.

J. Edwards, M. Covert & B. Palsson (2002). Metabolic modelling of microbes: the flux-balance approach. *Environmental Microbiology*, 4:133–140.

J. Edwards & B. Palsson (1998). How will bioinformatics influence metabolic engineering? *Biotechnology and Bioengineering*, 58:162–169.

T. Egli (1991). On multiple-nutrient-limited growth of microorganisms, with special reference to dual limitation by carbon and nitrogen substrates. *Antonie van Leeuwenhoek*, 60:225–234.

M. A. Eiteman & E. Altman (2006). Overcoming acetate in *Escherichia coli* recombinant protein fermentations. *Trends in Biotechnology*, 24:530–533.

E. M. T. El-Mansi & W. H. Holms (1989). Control of carbon flux to acetate excretion during growth of *Escherichia coli* in batch and continuous cultures. *Journal of General Microbiology*, 135:2875–2884.

H. G. Enoch & R. L. Lester (1972). Effects of molybdate, tungstate, and selenium compounds on formate dehydrogenase and other enzyme systems in *Escherichia coli*. *Journal of Bacteriology*, 110:1032–1040.

W. R. Farmer & J. C. Liao (1997). Reduction of aerobic acetate production by *Escherichia coli*. *Applied and Environmental Microbiology*, 63:3205–3210.

T. Ferenci (1999). "Growth of bacterial cultures" 50 years on: Towards an uncertainty principle instead of constants in bacterial growth kinetics. *Research in Microbiology*, 150:431–438.

S. P. A. Fodor, R. P. Rava, X. C. Huang, A. C. Pease, C. P. Holmes & C. L. Adams (1993). Multiplexed biochemical assays with biological chips. *Nature*, 364:555–556.

S. P. A. Fodor, J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu & D. Solas (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science*, 251:767–773.

B. Friedrich & B. Magasanik (1977). Urease of *Klebsiella aerogenes*: Control of its synthesis by glutamine synthetase. *Journal of Bacteriology*, 131:446–452.

L. Gautier, L. Cope, B. M. Bolstad & R. A. Irizarry (2004). affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*, 20:307–315.

J. D. Gawronski & D. R. Benson (2004). Microtiter assay for glutamine synthetase biosynthetic activity using inorganic phosphate detection. *Analytical Biochemistry*, 327:114–118.

R. Gentleman, V. Carey, W. Huber, R. A. Irizarry & S. Dudoit, editors (2005). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Springer.

R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang & J. Zhang (2004). Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, 5:R80.

D. Georgellis, O. Kwon & C. C. Lin (2001). Quinones as the redox signal for the Arc two-component system of bacteria. *Science*, 292:2314–2316.

G. Golub & C. Van Loan (1996). *Matrix Computations*. Johns Hopkins University Press, Baltimore and London, third edition.

J. Green, B. Bennett, P. Jordan, E. Ralph, A. J. Thomson & J. R. Guest (1996). Reconstitution of the [4Fe-4S] cluster in FNR and demonstration of the aerobic–anaerobic transcription switch *in vitro*. *Biochemical Journal*, 316:887–892.

J. Green & M. S. Paget (2004). Bacterial redox sensors. *Nature Reviews Microbiology*, 2:954–967.

R. P. Gunsalus (1992). Control of electron flow in *escherichia coli*: coordinated transcription of respiratory pathway genes. *Journal of Bacteriology*, 174:7069–7074.

D. H. Hahm, J. Pan & J. S. Rhee (1994). Characterization and evaluation of a *pta* (phosphotransacetylase) negative mutant of *Escherichia coli* HB101 as production host of foreign lipase. *Applied Microbiology and Biotechnology*, 42:100–107.

K. Han, H. C. Lim & J. Hong (1992). Acetic acid formation in *Escherichia coli* fermentation. *Biotechnology and Bioengineering*, 39:663–671.

D. Hanahan, J. Jessee & F. R. Bloom (1991). Plasmid transformation of *Escherichia coli* and other bacteria. *Methods in Enzymology*, 204:63–113.

A. Heck (1993). *Introduction to Maple*. Springer-Verlag.

D. Holder, R. F. Rauberras, V. B. Pikounis, V. Svetnik & K. Shoper (2001). Statistical analysis of high density oligonucleotide arrays: A SAFER approach. *Proceedings of the ASA Annual Meeting Atlanta, GA*.

P. A. Hoskisson & G. Hobbs (2005). Continuous culture – making a comeback? *Microbiology*, 151:3153–3159.

Q. Hua, C. Yang, T. Oshima, H. Mori & K. Shimizu (2004). Analysis of gene expression in *Escherichia coli* in response to changes of growth-limiting nutrient in chemostat cultures. *Applied and Environmental Microbiology*, 70:2354–2366.

G. S. Huang, M.-Y. Hong & Y.-C. Liu (2003). Incorporation of DNA chip technology to the simulation and validation of flux analysis in yeast diauxic growth. *Life Sciences*, 72:2525–2531.

W. Huber, Irizarry & R. Gentleman (2005). Preprocessing overview. In Gentleman *et al.* (2005), chapter 1, pp. 3–12.

W. Huber, A. von Heydebreck, H. Sültmann, A. Poustka & M. Vingron (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18 Suppl. 1:S96–S104.

W. Huber, A. von Heydebreck, H. Sültmann, A. Poustka & M. Vingron (2003). Parameter estimation for the calibration and variance stabilization of microarray data. *Statistical Applications in Genetics and Molecular Biology*, 2:Article 3.

S.-P. Hung, P. Baldi & G. W. Hatfield (2002). Global gene expression profiling in *Escherichia coli* K12. The effects of leucine-responsive regulatory protein. *The Journal of Biological Chemistry*, 277:40309–40323.

R. Ihaka & R. Gentleman (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5:299–314.

R. A. Irizarry, L. Gautier & L. M. Cope (2003a). An R package for analysis of Affymetrix oligonucleotide arrays. In Parmigiani *et al.* (2003), chapter 4, pp. 102–119.

R. A. Irizarry, B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf & T. P. Speed (2003b). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4:249–264.

A. Javelle, E. Severi, J. Thornton & M. Merrick (2004). Ammonium sensing in *Escherichia coli*. *The Journal of Biological Chemistry*, 279:8530–8538.

A. Javelle, G. Thomas, A.-M. Marini, R. Krämer & M. Merrick (2005). *In vivo* functional characterization of the *Escherichia coli* ammonium channel AmtB: Evidence for metabolic coupling of AmtB to glutamine synthetase. *Biochemical Journal*, 390:215–222.

E. B. Jensen & S. Carlsen (1990). Production of recombinant human growth hormone in *escherichia coli*: expression of different precursors and physiological effects of glucose, acetate and salts. *Biotechnology and Bioengineering*, 36:1–11.

P. Jiang & A. J. Ninfa (1999). Regulation of autophosphorylation of *Escherichia coli* nitrogen regulator II by the PII signal transduction protein. *Journal of Bacteriology*, 181:1906–1911.

L. Johansson & G. Lidén (2006). Transcriptome analysis of a shikimic acid producing strain of *Escherichia coli* W3110 grown under carbon- and phosphate-limited conditions. *Journal of Biotechnology*, 126:528–545.

L. Johansson, A. Lindskog, G. Silfversparre, C. Cimander, K. F. Nielsen & G. Lidén (2005). Shikimic acid production by a modified strain of *E. coli* (W3110.shik1) under phosphate-limited and carbon-limited conditions. *Biotechnology and Bioengineering*, 92:541–552.

E. Jones, T. Oliphant, P. Peterson *et al.* (2001–2007). SciPy: Open source scientific tools for Python. `http://www.scipy.org/`.

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

A. Kayser, J. Weber, V. Hecht & U. Rinas (2005). Metabolic flux analysis of *Escherichia coli* in glucose-limited continuous culture. I. Growth-rate-dependent metabolic efficiency at steady state. *Microbiology*, 151:693–706.

J. Keener & S. Kustu (1988). Protein kinase and phosphoprotein phosphatase activities of nitrogen regulatory proteins NtrB and NtrC of enteric bacteria: Roles of the conserved amino-terminal domain of NtrC. *Proceedings of the National Academy of Sciences of the United States of America*, 85:4976–4980.

I. M. Keseler, J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I. T. Paulsen, M. Peralta-Gil & P. D. Karp (2005). EcoCyc: A comprehensive database resource for *Escherichia coli*. *Nucleic Acids Research*, 33:D334–D337.

N. Khoroshilova, C. Popescu, E. Münck, H. Beinert & P. J. Kiley (1997). Iron-sulfur cluster disassembly in the FNR protein of *Escherichia coli* by O_2: [4Fe-4S] to [2Fe-2S] conversion with loss of biological activity. *Proceedings of the National Academy of Sciences of the United States of America*, 94:6087–6092.

B. S. Kim, S. C. Lee, S. Y. Lee, Y. K. Chang & H. N. Chang (2004). High cell density fed-batch cultivation of *Escherichia coli* using exponential feeding combined with pH-stat. *Bioprocess and Biosystems Engineering*, 26:147–150.

S. Klamt, S. Schuster & E. Gilles (2002). Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. *Biotechnology and Bioengineering*, 77:734–751.

G. Kleman & W. Strohl (1994). Acetate metabolism by *Escherichia coli* in high-cell-density fermentation. *Applied and Environmental Microbiology*, 60:3952–3958.

G. L. Kleman, J. J. Chalmers, G. W. Luli & W. R. Strohl (1991). Glucose-stat, a glucose-controlled continuous culture. *Applied and Environmental Microbiology*, 57:918–923.

G. L. Kleman, J. J. Chalmers, G. W. Luli & W. R. Strohl (57). A predictive and feedback control algorithm maintains a constant glucose concentration in fed-batch fermentations. *Applied and Environmental Microbiology*, 1991:910–917.

G. L. Kleman, K. M. Horken, F. R. Tabita & W. R. Strohl (1996). Overproduction of recombinant ribulose 1,5-bisphosphate carboxylase/oxygenase from *synechococcus sp.* strain pcc6301 in glucose-controlled high-cell-density fermentations by *Escherichia coli* K-12. *Applied and Environmental Microbiology*, 62:3502–3507.

G. L. Kleman & W. R. Strohl (1992). High cell density and high-productivity microbial fermentation. *Current Opinion in Biotechnology*, 3:93–98.

D. R. Knop, K. M. Draths, S. S. Chandran, J. L. Barker, R. von Daeniken, W. Weber & J. W. Frost (2001). Hydroaromatic equilibration during biosynthesis of shikimic acid. *Journal of the American Chemical Society*, 123:10173–10182.

Y.-F. Ko, W. E. Bentley & W. A. Weigand (1994). A metabolic model of cellular energetics and carbon flux during aerobic *Escherichia coli* fermentation. *Biotechnology and Bioengineering*, 43:847–855.

K. Konstantinov, M. Kishimoto, T. Seki & T. Yoshida (1990). A balanced DO-stat and its application to the control of acetic acid excretion by recombinant *Escherichia coli*. *Biotechnology and Bioengineering*, 36:750–758.

S. Kumari, C. M. Beatty, S. J. W. Browning, Douglas F Busby, E. J. Simel, G. Hovel-Miner & A. J. Wolfe1 (2000). Regulation of Acetyl Coenzyme A synthetase in *Escherichia coli*. *Journal of Bacteriology*, 182:4173–4179.

S. Kustu, J. Hirschman, D. Burton, J. Jelesko & J. C. Meeks (1984). Covalent modification of bacterial glutamine synthetase: Physiological significance. *Molecular and General Genetics*, 197:309–317.

A. L. Lee, M. M. Ataai & M. L. Shuler (1984). Double-substrate-limited growth of *Escherichia coli*. *Biotechnology and Bioengineering*, 26:1398–1401.

J. Lee, S. Y. Lee, S. Park & A. P. J. Middelberg (1999). Control of fed-batch fermentations. *Biotechnology Advances*, 17:29–48.

S. Y. Lee (1996). High cell-density culture of *Escherichia coli*. *Trends in Biotechnology*, 14:98–105.

G. Lequeux, L. Johansson, J. Maertens, P. A. Vanrolleghem & G. Lidén (2006). 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.

G. Lequeux, R. van der Heijden, S. Van Den Broeck & P. A. Vanrolleghem (2004). Computational methods to determine conserved moieties and parallel pathways in metabolic network models. In *Proceedings 9th IFAC Conference on Computer Applications in Biotechnology CAB9*. Nancy, France, March 28–31.

R. L. Lester & J. A. Demoss (1971). Effects of molybdate and selenite on formate and nitrate metabolism in *Escherichia coli*. *Journal of Bacteriology*, 105:1006–1014.

S. M. LeVine, F. Ardeshier & G. F.-L. Ames (1980). Isolation and characterization of acetate kinase and phosphotransacetylase mutants of *Escherichia coli* and *salmonella typhimurium*. *Journal of Bacteriology*, 143:1081–1085.

C. Li & W. H. Wong (2001a). Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proceedings of the National Academy of Sciences of the United States of America*, 98:31–36.

C. Li & W. H. Wong (2001b). Model-based analysis of oligonucleotide arrays: Model validation, design issues and standard error application. *Genome Biology*, 2:research0032.

H. Lin, G. N. Bennett & K.-Y. San (2005). Metabolic engineering of aerobic succinate production systems in *Escherichia coli* to improve process productivity and achieve the maximum theoretical succinate yield. *Metabolic Engineering*, 7:116–127.

H. Y. Lin, B. Mathiszik, B. Xu, S. O. Enfors & P. Neubauer (2001). Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivation of *Escherichia coli*. *Biotechnology and Bioengineering*, 73:347–357.

D. J. Lockhart, H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton & E. L. Brown (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology*, 14:1675–1680.

A. Long, H. Mangalam, B. Chan, L. Tolleri, G. Hatfield & P. Baldi (2001). Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework: Analysis of global gene expression in *Escherichia coli* K12. *The Journal of Biological Chemistry*, 276:19937–19944.

I. Lönnstedt & T. Speed (2002). Replicated microarray data. *Statistica Sinica*, 12:31–46.

O. H. Lowry & J. A. Lopez (1946). The determination of inorganic phosphate in the presence of labile phosphate esters. *The Journal of Biological Chemistry*, 162:412–428.

G. W. Luli & W. R. Strohl (1990). Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Applied and Environmental Microbiology*, 56:1004–1011.

R.-Y. Luo, S. Liao, G.-Y. Tao, Y.-Y. Li, S. Zeng, Y.-X. Li & Q. Luo (2006). Dynamic analysis of optimality in myocardial energy metabolism under normal and ischemic conditions. *Molecular Systems Biology*, 2:71.

F. Madron, V. Veverka & V. Vanecek (1977). Statistical analysis of material balance of a chemical reactor. *American Institute of Chemical Engineers Journal*, 23:482–486.

R. Mahadevan, J. S. Edwards & F. J. Doyle (2002). Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophysical Journal*, 83:1331–1340.

M. Maheswaran & K. Forchhammer (2003). Carbon-source-dependent nitrogen regulation in *Escherichia coli* is mediated through glutamine-dependent GlnB signalling. *Microbiology*, 149:2163–2172.

X.-J. Mao, Y.-X. Huo, M. Buck, A. Kolb & Y.-P. Wang (2007). Interplay between CRP-cAMP and PII-Ntr systems forms novel regulatory network between carbon metabolism and nitrogen assimilation in *Escherichia coli*. *Nucleic Acids Research*, 35:1432–1440.

J. C. March, M. A. Eiteman & E. Altman (2002). Expression of an anaplerotic enzyme, pyruvate carboxylase, improves recombinant protein production in *Escherichia coli*. *Applied and Environmental Microbiology*, 68:5620–5624.

M. R. Mashego, W. M. van Gulik, J. L. Vinke & J. J. Heijnen (2003). Critical evaluation of sampling techniques for residual glucose determination in carbon-limited chemostat culture of Saccharomyces cerevisiae. *Biotechnology and Bioengineering*, 83:395–399.

M. Matsubara & T. Mizuno (2000). The SixA phospho-histidine phosphatase modulates the ArcB phosphorelay signal transduction in *Escherichia coli*. *FEBS Letters*, 470:118–124.

D. Meyer, C. Shneider-Fresenius, R. Horlacher, R. Peist & W. Boos (1997). Molecular characterization of glucokinase from *Escherichia coli* K-12. *Journal of Bacteriology*, 179:1298–1306.

J. H. Miller (1972). *Experiments in molecular genetics*. Cold Spring Harbor Press.

T. Müller, B. Walter, A. Wirtz & A. Burkovski (2006). Ammonium toxicity in bacteria. *Current Microbiology*, 52:400–406.

K. Nakano, M. Rischke, S. Sato & H. Maerkl (1997). Influence of acetic acid on the growth of *Escherichia coli* K12 during high-cell-density cultivation in a dialysis reactor. *Applied Microbiology and Biotechnology*, 48:597–601.

D. V. Nguyen, A. B. Arpat, N. Wang & R. J. Carroll (2002). DNA microarray experiments: Biological and technological aspects. *Biometrics*, 58:701–717.

J. Nielsen, J. Villadsen & G. Lidén (2003). *Bioreaction Engineering Principles*. Kluwer Academic, Plenum Publishers, New York, second edition.

A. J. Ninfa & M. R. Atkinson (2000). PII signal transduction proteins. *Trends in Microbiology*, 8:172–179.

A. J. Ninfa & P. Jiang (2005). PII signal transduction proteins: sensors of α-ketoglutarate that regulate nitrogen metabolism. *Current Opinion in Microbiology*, 8:168–173.

H. Noorman, J. Heijnen & K. Luyben (1991). Linear relations in microbial reaction systems: A general overview of their origin, form and use. *Biotechnology and Bioengineering*, 38:603–618.

H. Noorman, K. Luyben & J. Heijnen (1996). Classification, error detection, and reconciliation of process information in complex biochemical systems. *Biotechnology and Bioengineering*, 49:364–376.

S. B. Noronha, H. J. C. Yeh, T. F. Spande & J. Shiloach (2000). Investigation of the TCA cycle and the glyoxylate shunt in *Escherichia coli* BL21 and JM109 using 13C-NMR/MS. *Biotechnology and Bioengineering*, 68:316–327.

A. Novick & L. Szilard (1950). Description of the chemostat. *Science*, 112:715–71715–7166.

M.-K. Oh & J. C. Liao (2000). Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. *Biotechnology Progress*, 16:278–286.

G. Parmigiani, E. S. Garret, R. A. Irizarry & S. L. Zeger, editors (2003). *The Analysis of Gene Expression Data. Methods and Software*. Springer.

J. D. Partridge, G. Sanguinetti, D. P. Dibden, R. E. Roberts, R. K. Poole & G. Jeffrey (2007). Transition of *Escherichia coli* from aerobic to micro-aerobic conditions involves fast and slow reacting regulatory components. *The Journal of Biological Chemistry*, 282:11230–11237.

J. D. Partridge, C. Scott, Y. Tang, R. K. Poole & G. Jeffrey (2006). *Escherichia coli* transcriptome dynamics during the transition from anaerobic to aerobic conditions. *The Journal of Biological Chemistry*, 281:27806–27815.

F. Pérez & B. E. Granger (2007). IPython: A system for interactive scientific computing. *Computing in Science and Engineering*, 9:21–29.

A. Perrenoud & U. Sauer (2005). Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in *Escherichia coli*. *Journal of Bacteriology*, 187:3171–3179.

J.-N. Phue & J. Shiloach (2004). Transcription levels of key metabolic genes are the cause for different glucose utilization pathways in *Escherichia coli* B (BL21) and *Escherichia coli* K (JM109). *Journal of Biotechnology*, 109:21–30.

C. V. Popescu, D. M. Bates, H. Beinert, E. Münck & P. J. Kiley (1998). Mössbauer spectroscopy as a tool for the study of activation/inactivation of the transcription regulator FNR in whole cells of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 95:13431–13435.

P. W. Postma, J. W. Lengeler & G. R. Jacobson (1993). Phosphoenolpyruvate: Carbohydrate phosphotransferase systems of bacteria. *Microbiological Reviews*, 57:543–594.

J. Pramanik & J. Keasling (1997). Stoichiometric model of *Escherichia coli* metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnology and Bioengineering*, 56:398–421.

A. Provost & G. Bastin (2004). Dynamic metabolic modelling under the balanced growth condition. *Journal of Process Control*, 14:717–728.

A. Puustinen, M. Finel, T. Haltia, R. B. Gennis & M. Wikström (1991). Properties of the two terminal oxidases of *Escherichia coli*. *Biochemistry*, 30:3936–3942.

L.-X. Qin, R. P. Beyer, F. Hudson, N. J. Linford, D. E. Morris & K. F. Kerr (2006). Evaluation of methods for oligonucleotide array data via quantitative PCR. *BMC Bioinformatics*, 7:23.

M. A. Quail, D. J. Haydon & J. R. Guest (1994). The *pdhR-aceEF-lpd* operon of *Escherichia coli* expresses the pyruvate dehydrogenase complex. *Molecular Microbiology*, 12:95–104.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. `http://www.R-project.org`.

J. L. Reed, T. D. Vo, C. H. Schilling & B. Ø. Palsson (2003). An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biology*, 4:R54.

P. Reilly & R. Carpani (1963). Application of statistical theory of adjustment to material balances. In *13th Chemical Engineering Conference*. Montreal, Canada.

L. Reitzer (2003). Nitrogen assimilation and global regulation in *Escherichia coli*. *Annual Review of Microbiology*, 57:155–176.

L. Reitzer & B. Magasanik (1985). Expression of *glnA* in *Escherichia coli* is regulated at tandem promoters. *Proceedings of the National Academy of Sciences of the United States of America*, 82:1979–1983.

D. Riesenberg & R. Guthke (1996). High-cell-density cultivation of microorganisms. *Applied Microbiology and Biotechnology*, 51:422–430.

B. Romein (2001). *Mathematical Modelling of Mammalian Cells in Suspension Culture*. Phd, Technische Universiteit Delft, The Netherlands.

K. Salmon, S.-P. Hung, K. Mekjian, P. Baldi, G. W. Hatfield & R. P. Gunsalus (2003). Global gene expression profiling in *Escherichia coli* K12. The effects of oxygen availability and FNR. *The Journal of Biological Chemistry*, 278:29837–29855.

K. Salmon, S.-P. Hung, N. Steffen, R. Krupp, P. Baldi, G. W. Hatfield & R. P. Gunsalus (2005). Global gene expression profiling in *Escherichia coli* K12. Effects of oxygen availability and ArcA. *The Journal of Biological Chemistry*, 280:15084–15096.

K.-Y. San, G. N. Bennett, A. A. Aristidou & C. H. Chou (1994). Strategies in high-level expression of recombinant protein in *Escherichia coli*. *Annals of the New York Academy of Sciences*, 721:257–267.

R. Sásik, E. Calvo & J. Corbeil (2002). Statistical analysis of high-density oligonucleotide arrays: A multiplicative noise model. *Bioinformatics*, 18:1633–1640.

U. Sauer, D. R. Lasko, J. Fiaux, M. Hochuli, R. Glaser, T. Szyperski, K. Wüthrich & J. E. Bailey (1999). Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. *Journal of Biotechnology*, 181:6679–6688.

J. Savinell & B. Palsson (1992). Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *Journal of Theoretical Biology*, 154:421–454.

A. Savitzky & M. J. E. Golay (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36:1627–1639.

E. E. Schadt, C. Li, C. Su & W. H. Wong (2000). Analyzing high-density oligonucleotide gene expression array data. *Journal of Cellular Biochemistry*, 80:192–202.

C. Schilling, D. Letscher & B. Palsson (2000). Theory for systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology*, 203:229–248.

C. Schilling, S. Schuster, B. Palsson & R. Heinrich (1999). Metabolic pathway analysis: Basic concepts and scientific applications in the post-genomic era. *Biotechnology Progress*, 15:296–303.

R. Schuetz, L. Kuepfer & U. Sauer (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*, 3:119.

D. Segrè, D. Vitkup & G. M. Church (2002). 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.

S. Shalel-Levanon, K.-Y. San & G. N. Bennett (2005a). Effect of ArcA and FNR on the expression of genes related to the oxygen regulation and the glycolysis pathway in *Escherichia coli* under microaerobic growth conditions. *Biotechnology and Bioengineering*, 92:147–159.

S. Shalel-Levanon, K.-Y. San & G. N. Bennett (2005b). Effect of oxygen, and ArcA and FNR regulators on the expression of genes related to the electron transfer chain and the TCA cycle in *Escherichia coli*. *Metabolic Engineering*, 7:364–374.

S. Shalel-Levanon, K.-Y. San & G. N. Bennett (2005c). Effect of oxygen on the *Escherichia coli* ArcA and FNR regulation systems and metabolic responses. *Biotechnology and Bioengineering*, 89:556–564.

S. Shin, S. G. Song, D. S. Lee, J. G. Pan & C. Park (1997). Involvement of *iclR* and *rpoS* in the induction of *acs*, the gene for acetyl coenzyme A synthetase of *Escherichia coli* K-12. *FEMS Microbiology Letters*, 146:103–108.

G. K. Smyth (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 3:Article 3.

G. K. Smyth (2005). limma: Linear models for microarray data. In Gentleman *et al.* (2005), chapter 23, pp. 397–420.

E. Soupene, L. He, D. Yan & S. Kustu (1998). Ammonia acquisition in enteric bacteria: Physiological role of the ammonium/methylammonium transport B (AmtB) protein. *Proceedings of the National Academy of Sciences of the United States of America*, 95:7030–7034.

E. Soupene, H. Lee & S. Kustu (2002). Ammonium/methylammonium transport (Amt) proteins facilitate diffusion of NH_3 bidirectionally. *Proceedings of the National Academy of Sciences of the United States of America*, 99:3926–3931.

E. Soupene, W. C. van Heeswijk, J. Plumbridge, V. Stewart, D. Bertenthal, H. Lee, G. Prasad, O. Paliy, P. Charernnoppakul & S. Kustu (2003). Physiological studies of *Escherichia coli* strain MG1655: Growth defects and apparent cross-regulation of gene expression. *Journal of Bacteriology*, 185:5611–5626.

S. Spiro (1994). The FNR family of transcriptional regulators. *Antonie van Leeuwenhoek*, 66:23–36.

S. Spiro & J. R. Guest (1991). Adaptive responses to oxygen limitation in *Escherichia coli*. *Trends in Biochemical Sciences*, 16(310–314).

J. Steinier, Y. Termonia & J. Deltour (1972). Comments on smoothing and differentiation of data by simplified least square procedure. *Analytical Chemistry*, 44:1906–1909.

G. Stephanopoulos, A. Aristidou & J. Nielsen (1998). *Metabolic Engineering. Principles and Methodologies*. Academic Press.

V. R. Sutton, E. L. Mettert, H. Beinert & P. J. Kiley (2004a). Kinetic analysis of the oxidative conversion of the $[4Fe\text{-}4S]^{2+}$ cluster of FNR to a $[2Fe\text{-}2S]^{2+}$ cluster. *Journal of Bacteriology*, 186:8018–8025.

V. R. Sutton, A. Stubna, T. Patschkowski, E. Münck, H. Beinert & P. J. Kiley (2004b). Superoxide Destroys the $[2Fe\text{-}2S]^{2+}$ Cluster of FNR from *Escherichia coli*. *Biochemistry*, 43:791–798.

G. Thomas, G. Coutts & M. Merrick (2000). The *glnKamtB* operon, a conserved gene pair in prokaryotes. *Trends in Genetics*, 16:11–14.

Q. H. Tran, T. Arras, S. Becker, G. Holighaus, G. Ohlberger & G. Unden (2000). Role of glutathione in the formation of the active form of the oxygen sensor FNR ([4Fe-4S].FNR) and in the control of FNR function. *European Journal of Biochemistry*, 267:4817–4824.

R. van der Heijden (1991). *State Estimation and Error Diagnosis for Biotechnological Processes*. Phd thesis, Technische Universiteit Delft, Nederland.

R. van der Heijden & J. Heijnen (1995). A comprehensive approach for structural analysis of metabolic networks. Unpublished.

R. van der Heijden, J. Heijnen, C. Hellinga, B. Romein & K. Luyben (1994a). Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates. *Biotechnology and Bioengineering*, 43:3–10.

R. van der Heijden, B. Romein, J. Heijnen, C. Hellinga & K. Luyben (1994b). Linear constraint relations in biochemical reaction systems: II. Diagnosis and estimation of gross errors. *Biotechnology and Bioengineering*, 43:11–20.

W. M. van Gulik, W. T. A. M. de Laat, J. L. Vinke & J. J. Heijnen (2000). Application of metabolic flux analysis for the identification of metabolic bottlenecks in the biosynthesis of Penicillin-G. *Biotechnology and Bioengineering*, 68:602–618.

W. C. van Heeswijk, S. Hoving, D. Molenaar, B. Stegeman, D. Kahn & H. V. Westerhoff (1996). An alternative PII protein in the regulation of glutamine synthetase in *Escherichia coli*. *Molecular Microbiology*, 21:133–146.

W. C. van Heeswijk, M. Rabenberg, H. V. Westerhoff & D. Kahn (1993). The genes of the glutamine synthetase adenylylation cascade are not regulated by nitrogen in *Escherichia coli*. *Molecular Microbiology*, 9:443–457.

A. Varma & B. Ø. Palsson (1993). Metabolic capabilities of *Escherichia coli* I. Synthesis of biosynthetic precursors and cofactors. *Journal of Theoretical Biology*, 165:477–502.

A. Varma & B. Ø. Palsson (1994a). Metabolic flux balancing: Basic concepts, scientific and practical use. *Biotechnology*, 12:994–998.

A. Varma & B. Ø. Palsson (1994b). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology*, 60:3724–3731.

N. Wang & G. Stephanopoulos (1983). Application of macroscopic balances to the identification of gross measurement errors. *Biotechnology and Bioengineering*, 25:2177–2208.

K. A. Webster (2003). Evolution of the coordinate regulation of glycolytic enzyme genes by hypoxia. *The Journal of Experimental Biology*, 206:2911–2922.

L. M. Wick, M. Quadroni & T. Egli (2001). Short- and long-term changes in proteome composition and kinetic properties in a culture of *Escherichia coli* during transition from glucose-excess to glucose-limited growth conditions in continuous culture and *vice versa*. *Environmental Microbiology*, 3:588–599.

F. K. Winkler (2006). Amt/MEP/Rh proteins conduct ammonia. *European Journal of Physiology*, 451:701–707.

Z. Wu, R. A. Irizarry, R. Gentleman, F. Martinez-Murillo & F. Spencer (2004). A model-based background adjustment for oligonucleotide expression arrays. *Journal of the American Statistical Association*, 99(468):909–917.

Y.-T. Yang, G. N. Bennett & K.-Y. San (1999). Effect of inactivation of *nuo* and *ackA-pta* on redistribution of metabolic fluxes in *Escherichia coli*. *Biotechnology and Bioengineering*, 65:291–297.

J. G. Zeikus, M. K. Jain & P. Elankovan (2004). Biotechnology of succinic acid production and markets for derived industrial products. *Applied Microbiology and Biotechnology*, 51:545–552.

L. Zhang, M. F. Miles & K. D. Aldape (2003). A model of molecular interactions on short oligonucleotide microarrays. *Nature Biotechnology*, 21:818–821.

J. Zhu, S. Shalel-Levanon, G. Bennett & K.-Y. San (2007). The YfiD protein contributes to the pyruvate formate-lyase flux in an *Escherichia coli arcA* mutant strain. *Biotechnology and Bioengineering*, 97:138–143.

M. Zinn, B. Witholt & T. Egli (2004). Dual nutrient limited growth: models, experimental observations, and applications. *Journal of Biotechnology*, 113:263–279.

ISBN 978-90-5989-219-4