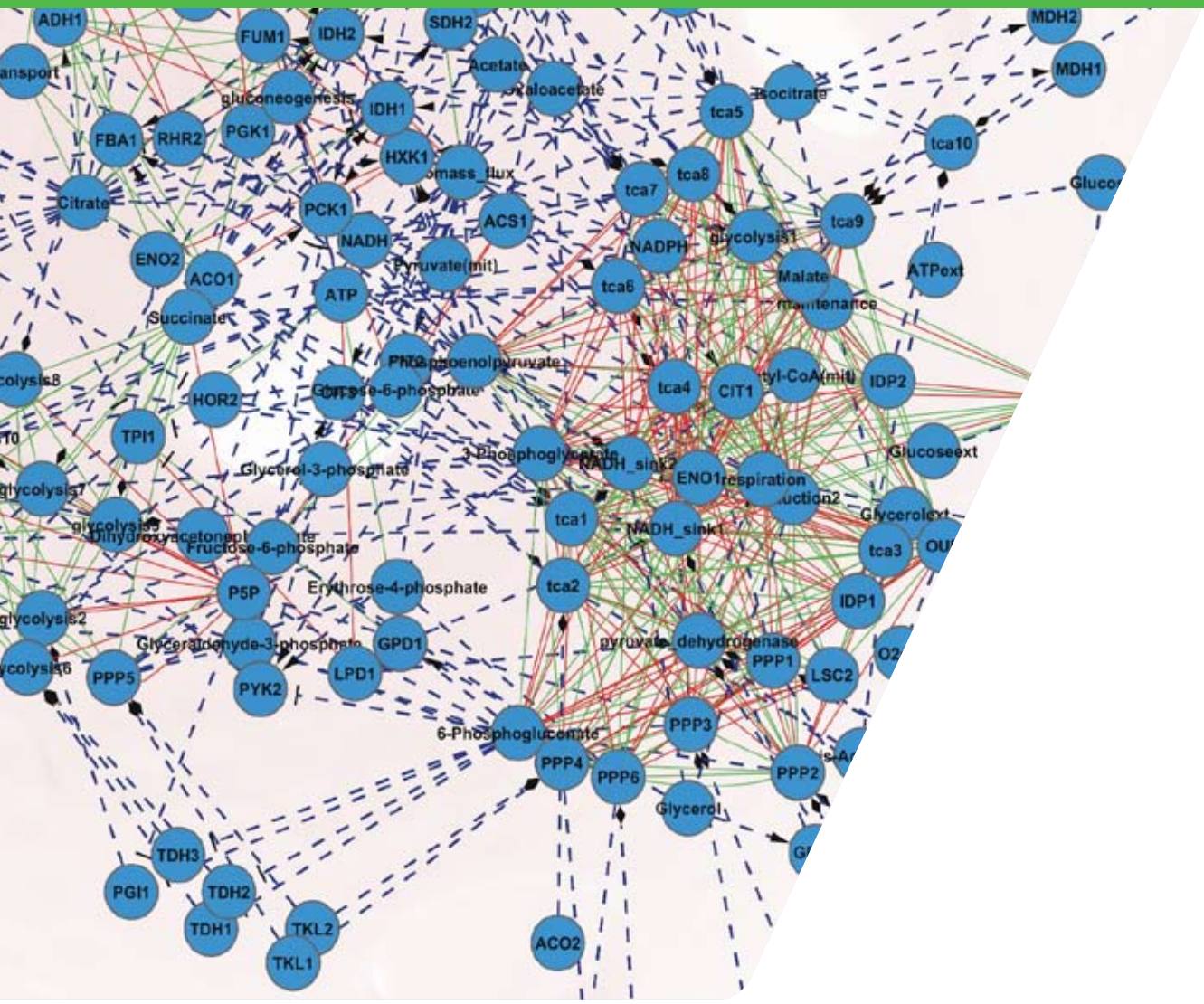

Paula Jouhten

Metabolic modelling and ^{13}C flux analysis

Application to biotechnologically important yeasts and a fungus

VTT PUBLICATIONS 724

Metabolic modelling and ^{13}C flux analysis

Application to biotechnologically important yeasts and a fungus

Paula Jouhten

Department of Biomedical Engineering and Computational Science
Faculty of Information and Natural Sciences
Helsinki University of Technology, Espoo, Finland

Graduate School in Computational Biology, Bioinformatics, and Biometry (ComBi)

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Faculty of Information and Natural Sciences, for public examination and debate in Auditorium in the building of the Department of Applied Physics (Otakaari 3) at Helsinki University of Technology (Espoo, Finland) on the 3rd of December, 2009, at 12 noon.

ISBN 978-951-38-7371-4 (soft back ed.)
ISSN 1235-0621 (soft back ed.)

ISBN 978-951-38-7372-1 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)

JULKAISIJA – UTGIVARE – PUBLISHER

VTT, Vuorimiehentie 5, PL 1000, 02044 VTT
puh. vaihde 020 722 111, faksi 020 722 4374

VTT, Bergsmansvägen 5, PB 1000, 02044 VTT
tel. växel 020 722 111, fax 020 722 4374

VTT Technical Research Centre of Finland, Vuorimiehentie 5, P.O. Box 1000, FI-02044 VTT, Finland
phone internat. +358 20 722 111, fax + 358 20 722 4374

Edita Prima Oy, Helsinki 2009

Paula Jouhten. Metabolic modelling and ^{13}C flux analysis. Application to biotechnologically important yeasts and a fungus [Aineenvaihdunnan mallinnus ja ^{13}C-vuoanalyysi. Sovellukset bioteknologisesti tärkeisiin hiivoihin ja homeeseen]. Espoo 2009. VTT Publications 724. 94 p. + app. 83 p.

Keywords metabolic modelling, metabolic flux, metabolic flux analysis (MFA), ^{13}C-labelling, ^{13}C-MFA, nuclear magnetic resonance (NMR) spectroscopy

Abstract

All bioconversions in cells derive from metabolism. Microbial metabolisms contain potential for bioconversions from simple source molecules to unlimited number of biochemicals and for degradation of even detrimental compounds. Metabolic fluxes are rates of consumption and production of compounds in metabolic reactions. Fluxes emerge as an ultimate phenotype of an organism from an integrated regulatory function of the underlying networks of complex and dynamic biochemical interactions. Since the fluxes are time-dependent, they have to be inferred from other, measurable, quantities by modelling and computational analysis. ^{13}C-labelling is crucial for quantitative analysis of fluxes through intracellular alternative pathways. Local flux ratio analysis utilises uniform ^{13}C-labelling experiments, where the carbon source contains a fraction of uniformly ^{13}C-labelled molecules. Carbon-carbon bonds are cleaved and formed in metabolic reactions depending on the *in vivo* fluxes. ^{13}C-labelling patterns of metabolites or macromolecule components can be detected by mass spectrometry (MS) or nuclear magnetic resonance (NMR) spectroscopy. Local flux ratio analysis utilises directly the ^{13}C-labelling data and metabolic network models to solve ratios of converging fluxes.

In this thesis the local flux ratio analysis has been extended and applied to analysis of phenotypes of biotechnologically important yeasts *Saccharomyces cerevisiae* and *Pichia pastoris*, and a fungus *Trichoderma reesei*. Oxygen dependence of *in vivo* net flux distribution of *S. cerevisiae* was quantified by using local flux ratios as additional constraints to the stoichiometric model of the central carbon metabolism. The distribution of fluxes in the pyruvate branching point turned out to be most responsive to different oxygen availabilities. The distribution of fluxes was observed to vary not only between the fully respiratory, respiro-fermentative and fermentative metabolic states but also between different respiro-fermentative states. The local flux ratio analysis was extended

to the case of two-carbon source of glycerol and methanol co-utilisation by *P. pastoris*. The fraction of methanol in the carbon source did not have as profound effect on the distribution of fluxes as the growth rate. The effect of carbon catabolite repression (CCR) on fluxes of *T. reesei* was studied by reconstructing amino acid biosynthetic pathways and by performing local flux ratio analysis. *T. reesei* was observed to primarily utilise respiratory metabolism also in conditions of CCR. *T. reesei* metabolism was further studied and L-threo-3-deoxy-hexulosonate was identified as L-galactonate dehydratase reaction product by using NMR spectroscopy. L-galactonate dehydratase reaction is part of the fungal pathway for D-galacturonic acid catabolism.

Paula Jouhten. Metabolic modelling and ^{13}C flux analysis. Application to biotechnologically important yeasts and a fungus [Aineenvaihdunnan mallinnus ja ^{13}C-vuoanalyysi. Sovellukset bioteknologisesti tärkeisiin hiivoihin ja homeeseen]. Espoo 2009. VTT Publications 724. 94 s. + liitt. 83 s.

Avainsanat metabolic modelling, metabolic flux, metabolic flux analysis (MFA), ^{13}C-labelling, ^{13}C-MFA, nuclear magnetic resonance (NMR) spectroscopy

Tiivistelmä

Aineenvaihdunta kattaa kaikki biomuunnokset soluissa. Mikrobiaineenvaihdunta mahdollistaa yksinkertaisten lähtöaineiden muuntamisen rajoittamattomaksi määräksi erilaisia biokemikaaleja ja jopa haitallisten aineiden hajottamisen. Aineenvaihduntavuot ovat yhdisteiden kulutus- ja tuottonopeuksia aineenvaihdunnan reaktioissa. Vuot ilmentyvät organismin todellisena fenotyyppinä, jota säätelevät yhteistoiminnallisesti solun monimutkaiset ja dynaamiset vuorovaikutusverkot. Koska vuot ovat aikariippuvaisia, ne on määritettävä mallinnuksen ja laskennallisen analyysin avulla toisista, mitattavissa olevista, suureista. ^{13}C-leimaus on välttämätöntä, jotta vuot vaihtoehtoisilla solunsisäisillä reiteillä voidaan määrittää kvantitatiivisesti. Paikallisessa vuosuhdeanalyysissa käytetään tasaista ^{13}C-leimausta, jossa hiilenlähde sisältää osuuden täydellisesti ^{13}C-leimattuja molekyylejä. *In vivo* -vuot määräävät missä suhteissa aineenvaihdunnassa katkeaa ja muodostuu uusia hiili-hiilisidoksia. Aineenvaihdunnan välituotteiden ja makromolekyylien komponenttien ^{13}C-leimauskuvioita voidaan mitata massaspektrometrialla (MS) tai ydinmagneettisella resonanssispektroskopialla (NMR). Paikallisessa vuosuhdeanalyysissa käytetään suoraan mittausinformaatiota ^{13}C-leimauskuvioista ja aineenvaihduntaverkkomalleja vuosuhteiden ratkaisemiseksi.

Väitöskirjassa paikallista vuosuhdeanalyysia laajennettiin ja sovellettiin bioteknologisesti tärkeiden hiivojen *Saccharomyces cerevisiae* ja *Pichia pastoris*, ja homeen *Trichoderma reesei* fenotyyppien analysoimiseksi. *S. cerevisiae*n *in vivo* -vuojakauman riippuvuus hapen saatavuudesta määritettiin kvantitatiivisesti käyttämällä paikallisia vuosuhteita lisärajoitteina keskeisen hiiliaineenvaihdunnan stoikiometriselle mallille. Pyruvaattiristeyksen vuojakauma osoittautui herkimmäksi eri happisaatavuuksille. Selvästi erilaiset vuojakaumat havaittiin täysin respiratiivisessa, respiro-fermentatiivisessa ja täysin fermentatiivisessa aineenvaihdunnan tilassa, mutta myös eri respiro-fermentatiivisissa tiloissa. Paikallinen vuosuhdeanalyysi laajennettiin kahden hiilenlähteen tapaukseen, jossa

P. pastoris kulutti samanaikaisesti glyserolia ja metanolia. Metanolin osuudella kokonaishiilenlähteessä ei ollut yhtä merkittävää vaikutusta vuojakaumaan kuin hiivan kasvunopeudella. Hiilikataboliittirepression (CCR) vaikutusta *T. reesei*n vuojakaumaan tutkittiin rekonstruoimalla aminohapposynteesireitit ja tekemällä paikallinen vuosuhdeanalyysi. *T. reesei*n havaittiin käyttävän pääasiassa respiratiivista aineenvaihduntaa myös repressoivissa olosuhteissa. NMR-spektroskopiaa käytettiin myös D-galakturonihapon kaboliareitin tutkimuksessa ja L-treo-3-deoksi-heksulonaatti tunnistettiin *T. reesei*n L-galaktonaattidehydrataasireaktion tuotteeksi.

Preface

This study was mainly carried out at VTT Technical Research Centre of Finland in the team of Metabolic Engineering. In addition, a part of the work was done at the Department of Computer Science, University of Helsinki. The financial support from the Finnish Funding Agency for Technology and Innovation (Tekes), Academy of Finland (Finnish Centre of Excellence programme), Graduate School in Computational Biology, Bioinformatics, and Biometry (ComBi), and Finnish Foundation for Technology Promotion (TES), is appreciatively acknowledged.

Former Vice President Juha Ahvenainen, Vice President Anu Kaukovirta-Norja, Vice President Hans Söderlund, Technology Manager Richard Fagerström, and Technology Manager Tiina Nakari-Setälä are gratefully thanked for creating excellent working facilities and possibilities. Research Professor Merja Penttilä and Team Leader Laura Ruohonen at VTT are delightfully acknowledged for the scientific guidance and their positive attitude towards this thesis work. I am also grateful to Professor Esko Ukkonen at the Department of Computer Science, University of Helsinki and Professor Jens Nielsen at Chalmers University of Technology for providing excellent facilities and high level scientific guidance during my visits.

I wish to express my deepest gratitude to my supervisor docent Dr. Hannu Maaheimo. With patience he introduced me to the field of NMR spectroscopy, ^{13}C-labelling, and metabolic fluxes and since then he has always been there to listen to me and discuss my endless flux of ideas and to support and encourage me in all situations. Hannu's enthusiasm and interest in a wide range of topics in life never stop impressing me.

My sincere thanks go to Esa Pitkänen and Dr. Ari Rantanen for the collaboration, the willingness to share knowledge, and friendship from the days when we

did not share a language to present. In our collaboration I have been especially fond of the scientific enthusiasm and special kind of humour.

I would also like to acknowledge all other co-authors of my publications: Dr. Aina Sola from the group of Professor Pau Ferrer, Dr. Tiina Pakula and Team Leader Dr. Markku Saloheimo from the group of Protein Production at VTT, Eija Rintala, Satu Hilditch, Dr. Mervi Toivari and Dr. Marilyn Wiebe from the team of Metabolic Engineering at VTT, Professor Juho Rousu from the Department of Computer Science, University of Helsinki. In addition, Dr. Marilyn Wiebe and Vytas Raulinaitis are acknowledged for fruitful discussions and revising the English language of my manuscripts.

I would like to warmly thank all my colleagues at the NMR laboratory in Viikki, at VTT in Otaniemi, and at the Department of Computer Science in Kumpula for creating friendly working atmosphere and providing assistance whenever needed. It has been a pleasure and privilege to work with all of them! I wish to express special thanks to Dr. Kai Fredriksson for encouraging words in moments of self-doubt and for bewildering discussions. My warm thanks are also due to the people in the Nielsen's Lab for Systems Biology at Chalmers University of Technology for the joyful and enthusiastic environment and all collaboration during my visit and ever since. I would also like to express my cordial thanks to Dr. Kiran Patil from DTU Technical University of Denmark for the collaboration and friendship.

I gratefully thank Professor Risto Renkonen and Professor Ulla Ruotsalainen for critically reviewing the manuscript of this thesis.

An incredible source of energy and strength for this work has been my parallel life on floorball fields. I've been living my dreams while doing research and being an athlete. I wish to thank all the floorball people in Finland and abroad, who have brought so much joy to my life. Especially, I would like thank the people that I have got to know in Tapanilan Erä, Kassu, and all the coaches who have guided my sports career. To Erik Piispa I would like to express my deepest thanks for all invaluable guidance and encouragement. Rache and Fredrik I wish to thank for the special friendship during my visit in Sweden. My warmest thanks go to Harri for the precious friendship and never-ending encouragement.

Finally, I would like to express the greatest thanks to my mother Helena for her care, support and delicate guidance in my life. Without her I would have never had all the possibilities to succeed. To my father Mikko I wish to express my warmest gratitude for the support in different venues of my life and for lighting my enthusiasm for sports.

List of publications

I Jouhten, P., Rintala, E., Huuskonen, A., Tamminen, A., Toivari, M., Wiebe, M., Ruohonen, L., Penttilä, M. & Maaheimo, H. Oxygen dependence of metabolic fluxes and energy generation of *Saccharomyces cerevisiae* CEN.PK113-1A. *BMC Systems Biology* 2008, 2:60.

II Solà, A.*, Jouhten, P.*, Maaheimo, H., Sánchez-Ferrando, F., Szyperski, T. & Ferrer, P. Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. *Microbiology* 2007, 153:281–290. *equal contributions

III Jouhten, P., Pitkänen, E., Pakula, T., Saloheimo, M., Penttilä, M. & Maaheimo, H. ^{13}C-metabolic flux ratio and novel carbon path analyses confirmed that *Trichoderma reesei* uses primarily the respirative pathway also on the preferred carbon source glucose. *BMC Systems Biology* 2009, 3:104.

IV Rantanen, A., Rousu, J., Jouhten, P., Zamboni, N., Maaheimo, H. & Ukkonen, E. An analytic and systematic framework for estimating metabolic flux ratios from ^{13}C tracer experiments. *BMC Bioinformatics* 2008, 9:266.

V Kuorelahti, S., Jouhten, P., Maaheimo, H., Penttilä, M. & Richard, P. L-galactonate dehydratase is part of the fungal path for D-galacturonic acid catabolism. *Molecular Microbiology* 2006, 61:1060–1068.

Author’s contributions

In Publication I the author of the thesis participated in the design of the study, performed chemostat cultivations, carried out the nuclear magnetic resonance (NMR) spectroscopic experiments, designed and performed the modeling and the computational work and wrote the manuscript. In Publication II the author of the thesis performed NMR spectroscopic experiments, carried out the analysis of the NMR spectral data and participated in the preparation of the manuscript. In Publication III the author of the thesis participated in the design of the study, performed the cultivations, carried out the NMR spectroscopic experiments and performed the ^{13}C-metabolic flux ratio analysis and analysed the results. The author of the thesis also together with the second author interpreted the results of the computational pathway analysis and wrote the manuscript. In Publication IV the author of the thesis constructed the models of the metabolic networks and contributed to the development of the computational methods. In Publication V the author of the thesis performed NMR spectroscopic experiments and analysis of NMR spectral data.

Contents

Appendix V of this publication is not included in the PDF version. Please order the printed version to get the complete publication (http://www.vtt.fi/publications/index.jsp).

Abbreviations

AcCoA	acetyl coenzyme A
Ala	L-Alanine
Arg	L-Arginine
Asp	L-Aspartic acid
ATP	adenosine 5-triphosphate
BDF	biosynthetically directed fractional
CCR	carbon catabolite repression
CDW	cell dry weight
CER	carbon dioxide evolution rate
CE-TOFMS	capillary electrophoresis-time-of-flight-mass spectrometry
COSY	correlation spectroscopy
D	dilution rate
DHAP	dihydroxyacetonephosphate
D_2O	deuterium oxide
EMU	elementary metabolite unit
E4P	D-erythrose 4-phosphate
FBA	flux balance analysis
FID	free induction decay
GC-MS	gas chromatography-mass spectrometry
Glu	L-Glutamic acid

Gly	Glycine
His	L-Histidine
HSQC	heteronuclear single quantum correlation
Ile	L-Isoleucine
LC-MS	liquid chromatography-mass spectrometry
LC-MS/MS	liquid chromatography-tandem mass spectrometry
Leu	L-Leucine
Lys	L-Lysine
MAE	malic enzyme
METAFoR	metabolic flux ratio
MFA	metabolic flux analysis
MOMA	minimisation of metabolic adjustment
mRNA	messenger ribonucleic acid
MS	mass spectrometry
NAD^+	nicotinamide adenine dinucleotide (oxidized)
NADH	nicotinamide adenine dinucleotide (reduced)
$NADP^+$	nicotinamide adenine dinucleotide phosphate (oxidized)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
NMR	nuclear magnetic resonance
NO	nitric oxide
Oaa	oxaloacetate
Oaa_{cyt}	cytosolic oxaloacetate
Oaa_{mit}	mitochondrial oxaloacetate
Oga	oxoglutarate
OUR	oxygen uptake rate
PDH	pyruvate dehydrogenase
Pep	phosphoenolpyruvate
PEPck	phosphoenolpyruvate carboxykinase

Phe	L-Phenylalanine
PPP	pentose phosphate pathway
Pro	L-Proline
Pyr	pyruvate
Pyr_{cyt}	cytosolic pyruvate
Pyr_{mit}	mitochondrial pyruvate
ROS	reactive oxygen species
R5P	D-ribose 5-phosphate
sd	standard deviation
SEM	standard error of the mean
Ser	L-Serine
S7P	D-sedoheptulose 7-phosphate
TCA	tricarboxylic acid
Thr	L-Threonine
TOCSY	total correlation spectroscopy
TOF	time-of-flight
Tyr	L-Tyrosine
T3P	triose 3-phosphates
YNB	yeast nitrogen base

1. Introduction – biology part

1.1 Cell factories and model organisms

Biotechnology offers possibilities for production of highly specialised biochemicals and for sustainable and economic process alternatives. It exploits cells or biocatalytes, enzymes, in bioconversions. Cells contain complete machineries for bioconversions from simple source molecules to unlimited number of biochemicals and degradation potential for breakdown of compounds even hazardous. Microorganisms are efficient cell factories whose requirements on the process conditions are modest and they often grow on inexpensive media. Even waste streams or effluents can be utilised as raw materials for bioprocesses.

Eukaryotic microorganism, yeast *Saccharomyces cerevisiae* (Figure 1), has a long history of biotechnological utilisation from conventional use as baker's yeast to production of diverse biochemicals. Concomitantly *S. cerevisiae* has been widely applied as a model organism in studies of general cell physiology. Due to the broad interest and the long history, *S. cerevisiae* is one of the most studied microorganisms with highly developed molecular biology tools and modelling frameworks [Petranovic and Vemuri, 2009; Herrgård *et al.*, 2008; Nevoigt, 2008]. Today *S. cerevisiae* is emerging as a simple eukaryotic model, a systems biology workhorse, for elucidating the mechanisms of even human diseases [Petranovic and Nielsen, 2008; Chen and Thorner, 2007]. The focus in systems biology is in understanding the function of a cell system as a whole [Lazebnik, 2002]. Since the complexity of cell systems is beyond intuitive comprehension, the core of systems biology is mathematical modelling of biological processes [Kitano, 2002]. The significant similarity of the cell function among eukaryotic cells offers promising prospects for *S. cerevisiae* models [Chen and Thorner, 2007; Petranovic and Nielsen, 2008; Botstein *et al.*, 1997]. Physiology of *S. cerevisiae* was studied in Publication I of the thesis. The organisms investi-

gated in Publications II, III and V, are fungus *Trichoderma reesei* (Figure 1) and yeast *Pichia pastoris* that are two host organisms for industrial production of natural and heterologous proteins.

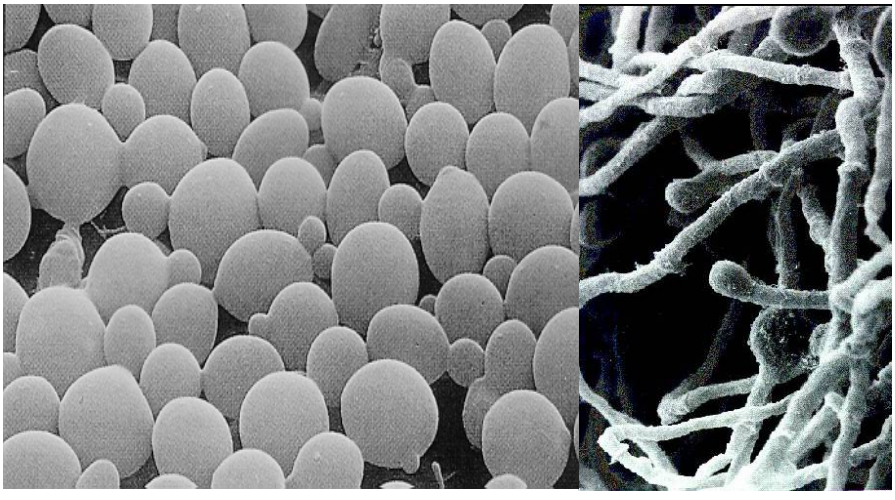

Figure 1. Budding yeast *S. cerevisiae* (on the left) and filamentous fungus *T. reesei* (on the right) are biotechnologically important production organisms.

1.2 Metabolism

All bioconversions in cells derive from metabolism. Metabolism is a set of biochemical reactions made feasible by enzymes [Stryer, 1995], which in turn are encoded by genes. Thus, the whole potential of metabolism of an organism is encoded in its genome, the complement of all genes. At present, the metabolic potential of an organism can usually be efficiently modified or engineered, with the variety of molecular biology tools available today. Metabolic engineering is, as stated by Stephanopoulos already in 1990's, "directed improvement of product formation or cellular properties through the modification of specific biochemical reaction(s) or the introduction of new one(s) with the use of recombinant DNA technology" [Stephanopoulos *et al.*, 1998].

Unicellular organisms comprise catabolism of substrates and anabolic pathways for synthesis of biomass components in a single cell. Pathways for metabolization of various carbon sources unite and a range of biosynthetic pathways initiate in central carbon metabolism, which is a common knot for catabolism and anabolism [Ma and Zeng, 2003]. Source molecules are broken down and

energy and precursors for biosynthetic pathways are produced in the central carbon metabolism.

1.3 Metabolic fluxes

Metabolic flux is a time-dependent quantity of the rate of consumption or production of compounds in a metabolic reaction [Stephanopoulos *et al.*, 1998; Nielsen *et al.*, 2003]. Metabolic fluxes are process streams of a cell factory. Therefore, a common aim of metabolic engineering is to generate changes in pathway fluxes. The essential biocatalytes, i.e. enzymes, can be amplified, deleted, and modified with versatile molecular biology tools. However, metabolic homeostasis prevails in cells that are highly balanced systems. A flux through a pathway depends on various factors in addition to the enzymes. The complement of fluxes in a cell, namely fluxome is cell's ultimate response to genetic and environmental conditions [Sauer, 2004]. The flux response emerges from an integrated function of complex and dynamic interaction networks (metabolic, signal transduction, regulatory, protein-protein interaction networks etc). Many of the components of the biochemical interaction networks such as concentrations of enzymes, other proteins, metabolites, and genome-wide gene expression levels are at present directly measurable with high-throughput systems. Since the fluxes are dependent of time, they cannot be directly measured but have to be inferred from other, measurable, quantities through a model based computational analysis.

1.4 Regulation of flux phenotype

The intertwined biochemical interaction networks of a cell form a regulatory system. The complex regulatory system enables both fine-tuned adaptive responses and robustness of the phenotype against genetic defects and fluctuations in external conditions [Kitano, 2007]. The ability to adapt to the prevailing growth conditions is essential for micro-organisms like *S. cerevisiae* that are unable to control the extracellular conditions. Furthermore, the regulatory system is capable of attenuating effects of genetic modifications on phenotype [Davies and Brindle, 1992; Schaaff *et al.*, 2004; Blank *et al.*, 2005].

The regulation of a finite change in flux can be conveniently and quantitatively divided into hierarchical and metabolic regulation [ter Kuile and Westerhoff, 2001]. Hierarchical regulation covers the steps of the central dogma of

molecular biology: gene expression, transcription, and translation. It ultimately determines the amounts of enzymes. Transcription is regulated by transcription factor proteins and other regulatory factors that may bind specifically to a gene to initiate or to speed up the formation of a messenger RNA (mRNA) [Fuda *et al.*, 2009]. The mRNA is then transferred out of the nucleus into the cytosol where it binds to ribosomes for translation. Rate of translation is dependent on various factors including ribosome density [Arava *et al.*, 2003; Brockmann *et al.*, 2007]. The degradation rates also affect the quantities of mRNAs and proteins.

Metabolic regulation of a change in flux includes everything beyond the enzyme concentrations such as activation of enzymes and kinetic control of reactions [ter Kuile and Westerhoff, 2001]. Post-translational modifications of proteins modulate their activity [Uy and Wold, 1977; Mann and Jensen, 2003]. For example phosphorylation can fully determine the activity of an enzyme [Ptacek *et al.*, 2005]. Signal transduction cascades pass phosphorylations as a response of sensing the growth conditions [Zaman *et al.*, 2008]. Reaction rates depend on the concentrations of the reactants after the particular kinetics of an enzyme [Stryer 1995]. In addition, metabolites can act as allosteric effectors and affect the reaction rates [Monod *et al.*, 1965] or even trigger regulation on the hierarchical regulatory levels [Sellick and Reece, 2003].

Metabolic homeostasis derives from dependences between flux, enzymes and metabolites and interdependences between reactions created by the metabolic network. Furthermore, hubs [Ma'ayan, 2009] such as cofactors NADH and NADPH and the energy unit ATP, are common metabolites for the whole network and create regulatory dependences even between distant pathways of the metabolic network. While there is lack of detailed knowledge on kinetic parameters and reaction mechanisms of large fraction of metabolic enzymes, thermodynamics provides insight to the dependences between fluxes and metabolites [Kummel *et al.*, 2006; Beard and Qian, 2005].

1.5 Oxygen affects flux phenotype

Oxygen conditions in nature vary between the oxygen partial pressure in air and complete anaerobiosis. Microorganisms have adapted to different ranges of oxygen availabilities depending on their natural habitats. A central role of oxygen metabolism is highly usual for any biological system [Koch and Britton, 2008]. Aerobic organisms are able to utilise oxygen as the final electron acceptor in the electron transfer chain, which is coupled to ATP synthesis. Since oxygen has

high electronegativity, the electron transfer reactions provide large transfer energy and the respiratory ATP production has an extremely high yield [Koch and Britton, 2008]. On the other hand, aerobic organisms need protection mechanisms against the deteriorating effects of oxygen [Jamieson, 1998]. Cell components are deteriorated by external oxidants and oxygen radicals generated in cell's internal oxygen utilising processes [Herrero *et al.*, 2008]. In should be noted that oxidation severely damages proteins, lipids, and nucleic acids [Jamieson, 1998; Herrero *et al.*, 2008]. Oxidative stress signalling activates repair mechanisms and degradation pathways for damaged components [Letavayová *et al.*, 2006]. Interestingly, cells counteract oxidative and reductive stresses with at least partly overlapping mechanisms [Trotter and Grant, 2002].

Aerobic environment poses a challenge also on the regulation of cell's redox balance. Cell cytosol is normally maintained reductive [López-Mirabal and Winther, 2008] and glutathione is the main buffer in the redox balancing system. In addition, it is linked to numerous cellular processes like membrane transport systems and carbon and nitrogen metabolisms [Perrone *et al.*, 2005; López-Mirabal and Winther, 2008]. Balanced redox conditions affect the metabolic homeostasis also because several metabolic reactions are redox reactions. When oxygen is not available as an electron acceptor, *S. cerevisiae* produces glycerol as a redox sink [Bakker *et al.*, 2001]. Fermentative pathway is redox neutral, but glycerol production occurs when the amount of NADH formed in biosynthesis exceeds the capacity of respiration to regenerate NADH to NAD^+. Reoxidation of NADH is prioritised under the conditions of oxygen deficiency and the carbon flux is directed to the fermentative pathway instead of to the TCA cycle [Weusthuis *et al.*, 1994; Publication I].

In *S. cerevisiae* oxygen limitation in the extracellular medium shifts the flux phenotype. The glycolytic flux is increased and ethanol production takes place [Weusthuis *et al.*, 1994; Publication I]. In the absence of ethanol production, metabolism is fully respirative. Respiro-fermentative phenotypes are observed in conditions of limited respiration. Respiratory limitation faces *S. cerevisiae* not only in lack of oxygen but also under excess glucose conditions and at high growth rate [Cortassa and Aon, 1998]. Glucose sensing and signalling network is active in conditions of excess glucose and it represses the components of the respiratory chain and the TCA cycle [Zaman *et al.*, 2009]. In contrast, the respiratory chain components have been observed to be upregulated under low oxygen conditions [Rintala *et al.*, 2009]. Aerobic alcoholic fermentation is observed in *S. cerevisiae* also at high growth rates when the glycolytic flux exceeds the

critical limit that depends on maximum respiratory rate [Vemuri *et al.*, 2007]. Exceeding the critical limit results in overflow metabolism in pyruvate branching point of central carbon metabolism and thus, to a flux to fermentative pathway [Vemuri *et al.*, 2007; Frick and Wittmann, 2005]. These observations indicate that similar metabolic states are generated by different transcriptional regulatory patterns highlighting the importance of the post-transcriptional and metabolic regulation of the phenotypes. Accordingly, the fluxes through glycolytic enzymes have previously been shown to be mainly regulated at post-transcriptional level [Daran-Lapujade *et al.*, 2007]. Furthermore, pure metabolic regulation can increase the glycolytic flux at least eight fold [van der Brink *et al.*, 2008] and the increased glycolytic flux observed under high temperature is primarily maintained by metabolic regulation [Postmus *et al.*, 2008]. On the other hand, the gluconeogenetic and glyoxylate cycle enzymes have been observed to be regulated at transcriptional level [Kolkman *et al.*, 2005].

Publication I studied the response of the metabolism of *S. cerevisiae* to the different oxygen provisions at the flux phenotypic level.

1.5.1 Oxygen responsive hierarchical regulatory mechanisms

The most well known oxygen-responsive hierarchical regulatory systems are dependent on the levels of heme and sterols [Hon *et al.*, 2003; Davies and Rine, 2006; Kwast *et al.*, 1998]. The synthesis of both of them requires molecular oxygen and thus, their levels decline in the depletion of oxygen. Hap-transcription factors respond to the levels of heme. Hap1 regulates the expression of anaerobic genes whereas Hap2/3/4/5 factors regulate expression of aerobic genes [Kwast *et al.*, 1998]. Hap1 factor has been shown to have a gentle slope in the activity profile in mild oxygen limitation but a sharp increase in severe lack of oxygen [Hon *et al.*, 2003]. Hap2/3/4/5 factors regulate genes encoding metabolic TCA cycle enzymes among others and Hap4 particularly activates catabolism of respiratory carbon sources like ethanol [Raghevendran *et al.*, 2006].

The mitochondrial production of oxidative stress mediating reactive oxygen species (ROS) and nitric oxide (NO) in low oxygen conditions have been suggested to be involved in signalling for induction of hypoxic genes [Castello *et al.*, 2006; D'Autréaux and Toledano, 2007; Woo *et al.*, 2009]. Accordingly transient oxidative stress response has been observed in *S. cerevisiae* in sudden depletion of oxygen [Dirmeier *et al.*, 2002]. Genes encoding enzymes involved in biosynthesis of fatty acids, which requires oxygen, belong to hypoxic genes.

The mitochondria are also known to signal of a respiratory defect by retrograde signalling that affects even the transcription of nuclear genes [Butow and Avadhani, 2004; Liu and Butow, 2006]. Retrograde signalling coordinates carbon and nitrogen metabolisms to respond to the requirements of the deficient state. In conditions of mitochondrial deficiency, the regulation of genes encoding TCA cycle enzymes switch from Hap-complex to retrograde regulators [Liu and Butow, 2006].

1.6 Carbon catabolite repression regulation of phenotype

Carbon catabolite repression (CCR) is a phenomenon where in presence of a preferred carbon source the pathways for utilisation of alternative carbon sources are repressed [Gancedo, 1998]. In the presence of excess glucose CCR of *S. cerevisiae* strongly represses metabolization of other carbon sources and also the respirative pathway as discussed above [Gancedo 1998; Zaman *et al.*, 2009; Westergaard *et al*., 2007]. High glucose mediates the redistribution of fluxes to respirative and fermentative pathways similarly as varying oxygen conditions and exceeding the maximum respiratory capacity [Gombert *et al.*, 2001; Nissen *et al.*, 1997; Vemuri *et al.*, 2007; Publication I]. Fermentation and high glycolytic flux enable high rate of ATP production. Glucose repression in *S. cerevisiae* is a regulatory switch that prefers high rate of ATP production instead of the high ATP yield that could be obtained from the respirative pathway.

T. reesei is naturally adapted to grow in nutrient poor environments, where it is able to use complex plant material as carbon source. *T. reesei* and number of other filamentous fungi and cellulolytic bacteria produce and secrete plant polymer hydrolyzing enzymes such as cellulases and hemicellulases to their surroundings to break down the polymers into easily metabolizable monomers [Kumar *et al.*, 2008]. The powerful machinery producing hydrolytic enzymes in *T. reesei* is under CCR when a preferred carbon source, such as glucose, is available. Small oligosaccharides or derivative parts of the polymers in the environment of the fungus act as inducers of expression of genes encoding hydrolytic enzymes. The inductive signaling is specific for particular sets of enzymes [Ilmén, 1997; Aro *et al*., 2005]. However, under high glucose concentrations, CCR overrules the inductive signals [Ilmén *et al*., 1997]. The regulatory switch of energy generation in *T. reesei* is different from the switch in *S. cerevisiae*. In *T. reesei* CCR does not cause repression of genes encoding the TCA cycle enzymes or the respiratory pathway components. Thus, CCR does not hinder the

high yield respirative energy generation in *T. reesei* [Chambergo *et al*., 2002; Gancedo, 1998]. The difference is reasonable in evolutionary sense because all the available energy in the nutritionally harsh natural habitats of *T. reesei* is valuable.

The signalling pathways for glucose repression in *S. cerevisiae* are widely studied [Zaman *et al*., 2009]. The systems biology approach has provided further understanding of the interaction of separate signalling pathways in *S. cerevisiae* in yielding specific responses to the growth conditions [Westergaard *et al*., 2007]. In *T. reesei* Cre1 is the key mediator protein of CCR [Strauss *et al.*, 1995; Ilmén *et al.*, 1996]. It is structurally highly similar to Mig1, a key protein in glucose repression in *S. cerevisiae*. Despite the sequence and structural similarity, the functional dissimilarities of Cre1 and Mig1 have led to the conclusion that glucose repression functionalities in filamentous fungi and yeasts have evolved separately [Cziferszky *et al.*, 2002; Vautard *et al*., 1999]. Pfeiffer *et al.* (2001) has also argued that the evolution from unicellular to undifferentiated multicellular organisms like *T. reesei* has been facilitated by the preference of high yield energy generation by respiration. The role of respirative metabolism in the development of multicellular organisms has recently been supported by Koch and Britton (2008).

In Publication III the distribution of intracellular metabolic fluxes in *T. reesei* were studied in different conditions of CCR.

2. Introduction – method part

2.1 Systems biology

Systems biology focuses on system level function of cells instead of the conventional approach of mere concentration on individual components [Kitano, 2002; Lazebnik, 2002]. Thus, mathematical modelling, networks of biochemical interactions, and high-throughput methods for simultaneous profiling of large numbers of cell components are essence of systems biology. Modelling is crucial for studying highly complex biological systems. Cell components and processes transfer information through interactions which enables for example adaptation mechanisms for survival and, on the contrary, phenotypic robustness against fluctuations in environmental conditions [Kitano, 2007]. An ultimate aim of systems biology is to generate predictive *in silico* models of biological systems. Modelling is an iterative process of continuous improvement of the description of the system. Models are mathematical representations of phenomena of interest and they are always simplifications of the actuality. According to the retelling of Einstein´s statement: models should be as simple as possible, but not simpler, thus the level of simplification as well as the type of the model should be designed for the purpose of the model [Klipp, 2007].

2.2 Metabolic modelling for flux analysis

The simplest models of metabolism are *black box* models in which everything else than the external fluxes of uptake and secretion is hidden in the black box. Intracellular reactions are not specified but just wrapped into the box. An overall reaction equation describes the conversion of substrates to products in the black box. Despite the obvious simplicity, the black box models can be utilised for calculation of mass balances, elemental balances and degree of reduction bal-

ances in modelling of cell factories, thus, also in calculation of process figures like yields and productivities. However, to be able to engineer the process figures, information on what occurs inside the black box is of importance.

2.2.1 Stoichiometric models

Stoichiometric models of metabolism specify the individual reactions in the system and the reaction stoichiometry relations of substrates and products. Stoichiometric models are static models and thus, do not include any reaction kinetics. While kinetic models of any medium size branching networks are still inconvenient due to the lack of knowledge and computational challenges, stoichiometric models have proven to be highly useful in metabolic studies of large networks.

2.2.1.1 Genome-wide metabolic reconstructions

The emergence of efficient sequencing and DNA techniques brought along a growing number of published fully sequenced and annotated genomes of organisms. Even the complete metabolic potential of an organism is encoded in its genome. Therefore, the availability of the genomes and the development of comparative genomics lead into reconstruction of genome wide metabolic models. The first genome-wide metabolic network reconstruction of *S. cerevisiae* was done by Förster *et al.* (2003) (Figure 2). Automatic methods exist both for full reconstruction of genome wide metabolism and for pathway searches from the given substrate to a product [Feist *et al.*, 2009 (review); Karp *et al*., 2002; Pinney *et al.*, 2005; Notebaart *et al.*, 2006]. However, after the automatic work, reliable metabolic reconstruction requires laborious manual curation including both literature checks and experimental verification of the metabolic network model [Francke *et al.*, 2005; Feist *et al.*, 2009 (review); Herrgård *et al.*, 2008 (*S. cerevisiae* consensus model); Duarte *et al.*, 2007 (human); Shinfuku *et al.*, 2009 (*Corynebacterium glutamicum*)]. The characterisation of enzymes and verification of their products as was done in Publication V in the thesis contributes to experimental validation of models. The genome wide metabolic reconstructions are stoichiometric models including static reaction descriptions and preferentially annotations of enzymes catalysing the reactions. The genome-wide metabolic models offer frameworks for investigations of the complete metabolic potential of an organism, and for data interpretation and analysis [Patil and Nielsen, 2005].

They also provide scaffolds for models of smaller and dynamic systems, and importantly they provide a link between the genome and the metabolic reactions.

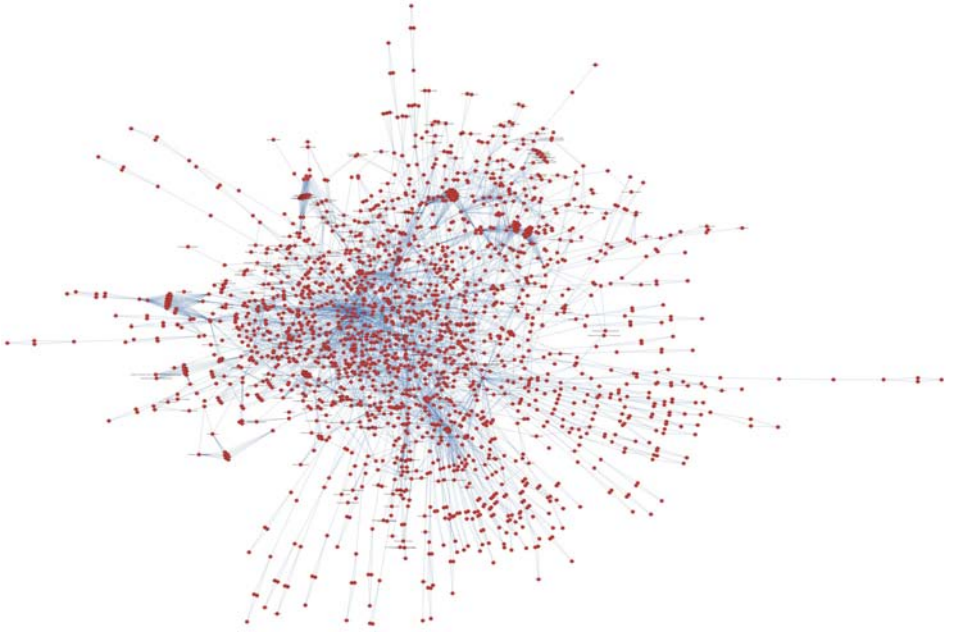

Figure 2. The genome-wide metabolic network of *S. cerevisiae* as a bipartite graph of metabolite and reaction nodes. The first genome-wide metabolic network reconstruction of *S. cerevisiae* reconstructed by Förster *et al.* (2003) included two compartments, cytosol and mitochondria, and 1175 metabolic reactions and 584 metabolites. The latest consensus model is divided into eight compartments and contains a total of 1168 metabolites and 1857 reactions and also 832 genes, 888 proteins, and 96 catalytic complexes [Herrgård *et al.*, 2008].

The reaction lists are readily converted into metabolite mass balances, functions of reaction rates, and further into stoichiometric matrices. Stoichiometric matrices can be analyzed by techniques of linear algebra to understand the metabolic potential and the structure of the metabolism of the particular organism [Palsson, 2006]. Stoichiometric reaction descriptions and a stoichiometric matrix are also the basic requirements for analysis of metabolic fluxes.

2.2.2 Kinetic models

Kinetic metabolic models include time-dependent mechanistic descriptions of reactions [Klipp, 2007]. Metabolic enzymes possess different mechanisms and

thus, the kinetic equations are enzyme specific [Stryer, 1995]. Depending on the reaction mechanism of an enzyme, the number of effectors and parameters varies. It is convenient to experimentally determine kinetics of isolated enzymes *in vitro*. However, it is likely that *in vivo* under crowded and compartmentalised conditions of a cell, the kinetics differ significantly from what is determined *in vitro*. Teusink *et al.* (2000) performed a study were a kinetic model of yeast glycolysis was set up and *in vitro* determined values for kinetic parameters were utilised for simulation. The output data from model simulations were compared to the experimental flux and metabolite data. There one finds significant discrepancies between simulated and experimental values of half of the reactions even though glycolysis is an extensively studied part of metabolism. Experimental data for computational estimation of kinetic parameters is acquired in perturbation experiments of the system of interest [Nikerel *et al.*, 2006; Vaseghi *et al.*, 1999]. However, since the number of parameters in kinetic metabolic models is usually huge, their estimation is demanding. Thus, different approximations of kinetic equations have been developed for studying the dynamic behaviour of a metabolic system [Heijnen, 2005 (review); Visser and Heijnen, 2003; Visser *et al.*, 2004; Savageau, 1970; Liebermeister and Klipp, 2006].

2.3 Metabolic flux analysis

Determination of intracellular *in vivo* fluxes is called metabolic flux analysis (MFA), which applies mass balances around metabolites according to the stoichiometric model. Thus, mass balances around metabolites are formulated as functions of the fluxes. Dilution of metabolite pools due to growth can be taken into account as a dilution term although in many cases the dilution term is negligible [Nielsen *et al.*, 2003].

$$\frac{dc_{met}}{dt} = v_{met} - \mu c_{met} \qquad (1)$$

In equation 1 c_{met} is the concentration of a particular metabolite, v_{met} the net rate of formation and consumption of the metabolite by all the fluxes in the system, and in the dilution term μ is the growth rate.

Integration of equation 1 over time for all the metabolites would yield the time dependence of concentrations $c_{met}(t)$. However, the fluxes v_{met} are often unknown functions of metabolite concentrations and unknown enzyme kinetic and other parameters as discussed above. Since it is challenging to simulate the dynamic

concentration profiles, steady state conditions are usually considered. In steady state conditions the growth rate and the metabolic fluxes are constant and there is no accumulation of intracellular metabolites but their pools also remain constant [Nielsen *et al.*, 2003]. Steady state microbial growth is reached in continuous cultures (chemostat cultures) where the rates of medium flow into and out from the bioreactor are equal. Pseudo-steady state conditions prevail in early exponential phase of a batch culture when the cells grow at maximum rate while the changes in the culture medium are still insignificant. Under assumption of steady state conditions the following form of equation 1 holds [Nielsen *et al.*, 2003]:

$$v_{met} - \mu c_{met} = 0 \tag{2}$$

Because the intracellular metabolite pools are generally very small, the dilution term is tiny compared to the fluxes producing and consuming metabolites, particularly in the central carbon metabolism that encompasses all the major fluxes [Stephanopoulos *et al.*, 1998]. When the dilution term can be assumed to be negligible, the equation 2 adapts a simple form:

$$v_{met} = 0 \tag{3}$$

which in matrix notation reads:

$$\mathbf{Nv} = \mathbf{0} \tag{4}$$

where **N** is the stoichiometric matrix with the fluxes of the system in columns and the stoichiometric coefficients of metabolites in each of the fluxes in rows and **v** is a column vector of fluxes. The stoichiometric matrix transforms the biology of metabolic reactions into mathematical framework. The matrix equation actually represents K linear mass balances for the K metabolites that contribute to J fluxes. Since there are always less metabolites K than fluxes J, the degree of freedom $F = K - J$ remains and some of the fluxes in vector v need to be determined to solve the rest of them [Nielsen *et al.*, 2003]. If the stoichiometric matrix is partitioned into two parts for measured fluxes ($\mathbf{N}_m$) and for unknown fluxes ($\mathbf{N}_u$), the equation can be rewritten in the following way:

$$\mathbf{Nv} = \mathbf{N}_m\mathbf{v}_m + \mathbf{N}_u\mathbf{v}_u = \mathbf{0} \tag{5}$$

where $\mathbf{v}_m$ is a vector of measured rates and $\mathbf{v}_u$ a vector of unknown rates. If exactly F fluxes have been measured and if $\mathbf{N}_u$ can be inverted, the unknown rates can be directly solved with matrix algebra [Stephanopoulos *et al.*, 1998]. If $\mathbf{N}_u$ has full rank ($rank(\mathbf{N}_u) = K$), it can be inverted and the unknown fluxes calcu-

lated by Gaussian elimination but if the rank of $\mathbf{N}_u$ is less than K, $\mathbf{N}_u$ is singular and the system is underdetermined. If the set of reaction stoichiometries are linearly dependent, the rank of $\mathbf{N}_u$ is less than K even though the number of measured fluxes equals degrees of freedom. Also if some of the measured rates are redundant, the rank of $\mathbf{N}_u$ is less than K.

Very seldom it is practically possible to measure enough fluxes to reach an algebraic solution to the metabolic system of linear mass balance equations. This is obviously always the case with large, genome-scale metabolic models.

2.3.1 Constraint-based analysis

The space of metabolic states has as many dimensions as reactions in the system. The stoichiometry of the reactions, equation 4, limits the space into a subspace that is a hyperplane. If the reactions are defined so that they are all positive, the plane is converted into a cone. If additionally upper bounds, maximum capacities, can be defined for the fluxes, a closed convex cone solution space is obtained. All the possible metabolic states of an organism, the feasible flux distributions, lie in that solution space. Thus, it is the space of phenotypes which an organism can express. To further shrink the solution space, additional constraints have been set up from reaction thermodynamics [Beard *et al.*, 2002; Beard *et al.*, 2004; Price *et al.*, 2004b; Price *et al.*, 2006], from experimental transcription data and from extracellular metabolome for condition-specific solution spaces [Covert and Palsson, 2002; Åkesson *et al.*, 2004; Becker and Palsson, 2008; Mo *et al.*, 2009]. The whole feasible solution space can be studied algebraically or statistically by sampling the space [Price *et al.*, 2004a; Palsson, 2006]. Randomized Monte Carlo sampling of the feasible solution space gives unbiased information on the shape and properties of the space where the true metabolic state lies [Price *et al.*, 2004a; Schellenberger and Palsson, 2009]. The null space that contains all the possible flux distribution can be studied algebraically [Palsson, 2006]. Investigation of the feasible solution space yields information on what types of solutions are possible, what parts of the metabolic network participate in the possible metabolic states, are there some limits for production of specific extracellular compounds etc. Obviously the properties of the feasible solution space contain even the properties of the true metabolic state.

Linear optimisation can be utilised to find a point solution, i.e. a single flux distribution. The approach is often called flux balance analysis (FBA) and there the optimisation requires an objective function. It is always a guess what the

organism actually optimises for in the particular conditions and it is generally not simple to set up a biologically meaningful objective function. Biomass production is an obvious choice for objective function for bacteria that grow exponentially. However, higher cells do not usually optimise for growth though uncontrolled growing cancer cells could be an exception. It is generally accepted that organisms have evolved to survive but their survival is not straightforward to define as an objective function. Different types of objective functions have been searched and suggested [Burgard and Maranas, 2003 (ObjFind); Holzhutter, 2004 (flux minimization)] and the optimised flux solutions have been tested against experimental data obtained with ^{13}C-tracer based methods [Schuetz *et al.*, 2007]. ^{13}C-tracer based methods will be presented in the next chapters. In addition to flux determination, FBA approach has been exploited for identification of optimal targets for metabolic engineering [Burgard *et al.*, 2003 (OptKnock); Pharkya *et al.*, 2004 (OptStrain)]. Furthermore, the properties of a metabolic system can be studied by defining different types of objective functions. For example production capabilities can be determined by optimising for the product formation. However, engineered organisms may not initially reach the optimal performance. Thus, FBA will not return flux phenotypes that match the reality of engineered organisms. Successful predictions of flux phenotypes of engineered organisms have been obtained with the minimization of metabolic adjustment (MOMA) to the wild type flux phenotype -principle and solved with quadratic programming [Segré *et al.*, 2002].

2.3.2 ^{13}C-metabolic flux analysis

As discussed above the determination of an objective function for FBA is often extremely difficult. In addition, the constraint-based MFA approaches, like FBA, cannot solve distributions of fluxes to parallel and alternative pathways. However, the parallel pathways usually transfer atoms in distinctive manner before they converge to a common metabolic intermediate. Thereby, utilisation of tracers has emerged. Since metabolism is all about breaking and making carbon-carbon bonds, ^{13}C, is the most common tracer in metabolic studies [Tang *et al.*, 2009]. ^{13}C is a stable carbon isotope whose natural abundance is only 1.1% [Gadian, 1982]. Other tracers are applicable for studies of specific metabolic pathways [Brosnan *et al.*, 2004 (^{15}N tracer application)].

2.3.2.1 ^{13}C-labelling and analytical methods

^{13}C-labelling for MFA is performed by introducing ^{13}C-labelled substrate to a cell culture. When the carbon source gets metabolized the tracer enrichment spreads first to the free intracellular metabolites and during extended periods of growth on ^{13}C-labelled carbon source also into the macromolecules and cell constituents. The spread of the ^{13}C-label is dependent on the metabolic fluxes and the turnover of macromolecules and storage pools. Since the alternative pathways often scramble and cleave the carbon backbones of metabolites in different ways, the ^{13}C-labelling prints information on the relative activities of the pathways into the carbon-carbon connectivities of the metabolites and into the positional fates of tracer atoms. Depending on the difference between the carbon chain modifications in the alternative pathways, different ^{13}C-labelling designs of the carbon source are optimal for resolution of the relative pathway activities [Möllney *et al.*, 1999; Araúzo-Bravo and Shimizu, 2003]. Isotopomer is a definition for isotopic isomers of a compound [Wiechert, 2001] and the isotopomers differ only in position or number of different isotopes in the molecule. Thus, molecules differing in ^{13}C-labelling patterns are isotopomers [Nielsen *et al.*, 2003]. In the following chapters two main ^{13}C-labelling approaches are presented.

Uniform labelling

In uniform ^{13}C-labelling approach the carbon source contains a fraction of molecules that are uniformly ^{13}C-labelled. Thus, they contain ^{13}C atom in all positions. Typically utilised fraction of uniformly ^{13}C-labelled molecules is around 20% [Zamboni *et al.*, 2009]. If the alternative metabolic pathways modify the carbon chain in distinctive ways, the relative activities of the pathways can be resolved with this approach. This approach can even be called bond labelling because the relative activities of the alternative pathways are actually recorded in the common product as cleaved and newly formed carbon-carbon bonds. The uniform ^{13}C-labelling approach was established in 1990's by Szyperski [Szyperski 1995] by introducing biosynthetically directed fractional (BDF) ^{13}C-labelling where a fraction of uniformly labelled carbon source was feed to microbial cells. During the steady-state growth on fractionally ^{13}C-labelled carbon source information on the relative *in vivo* activities of the pathways was recorded and significantly amplified in the ^{13}C-labelling patterns of proteinogenic amino acids. The ^{13}C-labelling patterns of proteinogenic amino acids could conveniently be

detected by two-dimensional nuclear magnetic resonance (NMR) spectroscopic experiments. Szyperski (1995) further introduced probabilistic equations for the relations between the ^{13}C-^{13}C-couplings in proteinogenic amino acids and fragmentomers, fractions of intact carbon fragments. Four fragmentomers that sum up to one of three carbon fragment of a molecule can be deduced. Fragmentomer $f^{(1)}$ represents the fraction of molecules in which the middle carbon atom and the neighboring carbons originate from different carbon source molecules, fragmentomer $f^{(2)}$ represents the fraction of molecules in which the middle carbon atom and one of the two neighboring atoms originate from the same carbon source molecule, and fragmentomer $f^{(3)}$ represents the fraction of molecules in which the middle carbon atom and both the neighboring carbons originate from the same carbon source molecule. Sometimes, if the end carbons of the three carbon fragments are in different chemical environment, even two different fragmentomers $f^{(2)}$ and $f^{(2^*)}$ can be distinguished by NMR spectroscopic methods. Fragmentomers are actually constraints for a full isotopomer distribution of a molecule. Later a GC-MS based detection method was developed to be compatible with BDF labelling [Fischer and Sauer, 2003].

Positional enrichment

The alternative ^{13}C-labelling approach is to introduce positional label(s). The positional label can be introduced in a specific position or positions of the carbon source and usually all the carbon source is equally ^{13}C-labelled. For example, the common carbon source glucose is commercially available in different compositions of ^{13}C and ^{12}C atoms. However, glucose with ^{13}C-atoms somewhere in the middle of the carbon chain is very expensive to purchase. During the growth on positionally ^{13}C-labelled carbon source, specific positions of product molecules become enriched depending on the *in vivo* activities of pathways. In positional fractional ^{13}C enrichments, the ratios of ^{13}C and ^{12}C atoms in the specific carbon positions of the product molecules provide constraints for the full isotopomer distribution of a molecule [Wiechert, 2001].

Nuclear magnetic resonance spectroscopy

Modern nuclear magnetic resonance (NMR) spectroscopy detects signals of spin possessing nuclei in a strong magnetic field after a radio frequency pulse or a sequence of pulses [Friebolin, 1991]. Nuclei that possess a spin different from

zero have different energy states in a magnetic field. Radio frequency pulses induce transitions between the energy states and create detectable macroscopic magnetisation. The magnetisation induces a current to the receiver coil. This signal is recorded and called free induction decay (FID). The FID is then Fourier transformed from time domain to frequency domain to obtain an NMR spectrum. The limited sensitivity of NMR stems from detecting only the small difference between the populations of nuclei on different energy states. The energy difference is dependent on the strength of the static magnetic field. Therefore, strong magnets are utilised. Nevertheless, NMR spectroscopy provides unlimited potential in the variety of methods that can be utilised for analysis of biological samples.

Spin possessing NMR active nuclei of main interest in analysis of biological samples are: ^{1}H, ^{13}C, ^{15}N, and ^{31}P [Gadian, 1982]. All these nuclei have spin quantum numbers of ½ and thus, have two possible energy states in a magnetic field. Proton is the most sensitive nuclei and ^{1}H NMR spectroscopy is an unbiased method because it can detect all proton containing compounds in a sample. On the other hand ^{31}P has a 100% natural abundance and therefore, it can be utilised for example for selective detection of only phosphorus containing compounds in a complex mixture.^{13}C atoms, whose natural abundance is only 1.1%, can be directly detected by NMR but not the more abundant ^{12}C atoms. Thus, ^{13}C is a suitable tracer for NMR spectroscopic studies.

NMR active nuclei give signals in an NMR spectrum on their characteristic chemical shifts [Friebolin, 1991]. The characteristic chemical shift of a nucleus depends on the nature and the chemical environment of the nucleus. Electrons in the chemical environment cause shielding of the magnetic field and thus, the magnetic field experienced by the nucleus is also dependent on its surrounding electrons. In addition, coupling to other NMR active nuclei through bonds gives rise to signal splitting. Therefore, different molecule structures have specific signal fine structures.

Complex sequences of radio frequency pulses can be designed for advanced NMR spectroscopic experiments. Magnetisation can, for example, be transferred from one type of nuclei to other types of nuclei, which targets the analysis to specific structures of interest. Higher dimensional experiments can be performed for one type nuclei (homonuclear) or for different types of nuclei (heteronuclear) [Croasmun and Carlson, 1994]. Higher dimensional experiments provide also more information about the structures of analytes because nuclei that are covalently bound together or close to each other in space can be identified.

NMR spectroscopic analyses of samples from ^{13}C-labelling experiments can provide two different types of constraints to the isotopomer distribution. If ^{13}C are detected, signal fine structures reveal fractions of couplings to adjacent ^{13}C and ^{12}C nuclei, thus, a ^{13}C-labelling status of a three carbon fragment [Szyperski, 1995]. Sometimes even longer couplings can be resolved. If protons are detected, the signal fine structure reveals the fractional enrichment of ^{13}C in the carbon coupled to the detected proton. The fraction of protons coupled to ^{13}C is observed as split satellite signals on both sides of the signal from the ^{12}C coupled protons [Friebolin, 1991].

Mass spectrometry

The advantage of mass spectrometry (MS) compared to NMR is its higher sensitivity. However, there are fundamental differences in the data that is produced by MS and NMR. MS detects molecules that have distinctive masses but it cannot distinguish between molecules having the same number of ^{13}C atoms but in different positions [Zamboni *et al.*, 2009]. Massisomer (or mass isotopomer) is the definition for isomers that differ in mass [Christensen and Nielsen, 1999]. Gas-chromatography-mass spectrometry (GC-MS) has been the most popular of MS techniques for analysis of samples from ^{13}C-labelling experiments [Wittmann, 2007]. Before the GC-MS analysis the metabolites are first derivatized to render the molecules volatile [Tang *et al.*, 2009]. Common derivatizations are silylation, acylation and alkylation [Tang *et al.*, 2009; Wittmann, 2007]. In GC-MS analytes become fragmented and derivatization agent may affect the fragmentation sites. Fragmentation yields more constraints to the full isotopomer distribution [Zamboni *et al.*, 2009]. Metabolites include natural isotopes and derivatization introduces additional atoms to the analytes. Thus, the raw data requires correction to remove them [Christensen and Nielsen, 1999; van Winden *et al.*, 2002].

Previous analyses of ^{13}C-labelling experiments, with NMR spectroscopy or GC-MS, have utilised the detection of ^{13}C-labelling patterns of proteinogenic amino acids that are abundant. GC-MS has somewhat been utilised also in analysis of free amino acids and organic acids [Wittmann *et al.*, 2002]. However, there is a delay before the label reaches proteins that are macromolecules or even some delay before the label enriches in the large free amino acid pools in cells. Therefore, the methods have not been suitable for analysis of phenomena in short time frames. In addition, the long ^{13}C-labelling experiments require lot

of expensive labelled carbon source. The application of liquid chromatography-mass spectrometry (LC-MS) in analysis of ^{13}C-labelling patterns in intracellular metabolic intermediates was demonstrated by van Winden *et al.* (2005) but the direct analysis of ^{13}C-labelling patterns of intracellular metabolites has not become a widespread method because the sensitivity severely limits the analysis as metabolic intermediates are present only in very low amounts. Later Toya *et al.* (2007) suggested CE-TOFMS (capillary electrophoresis time-of-flight mass spectrometry) for detection of ^{13}C-labelling patterns in free intracellular metabolites. CE-TOFMS is fast and the experimental set up is more flexible for analysis of various compounds than LC-MS. Kleijn *et al.* (2007) showed that the data sets from the three measurement techniques: NMR spectroscopy, LC-MS and GC-MS yielded consistent flux results in analysis of combined substrate labelling, 10% [U-^{13}C] and 90% [1-^{13}C] glucose, in glycerol over-producing *S. cerevisiae* strains. Since the flux sensitivities were found to often depend on the analysis method, a combined data set gave the most accurate flux distribution estimate. LC-MS was utilised for detection of ^{13}C in free metabolic intermediates whereas the NMR spectroscopy and GC-MS analyses were performed for ^{13}C-labelling patterns in proteinogenic amino acids and storage carbohydrates.

In order to obtain massisomer data on smaller fragments or even pure positional ^{13}C enrichment data, liquid chromatography-tandem mass spectrometry (LC-MS/MS) was introduced to ^{13}C-labelling analysis task [Iwatani *et al.*, 2007]. In LC-MS/MS the full massisomers are further fragmented and positional enrichments can be inferred from the full fragmentation data. Iwatani *et al.* (2007) applied LC-MS/MS detection to analysis of ^{13}C-labelling patterns of proteinogenic and free amino acids in *E. coli*.

2.3.2.2 ^{13}C-metabolic flux analysis – mathematical and statistical methods

The interpretation of data from ^{13}C-labelling experiments requires mathematical modelling and statistical analysis. Firstly the atom transfers in the metabolic reactions are essential to be modelled for interpretation of data form tracer experiments. Mappings of carbon atoms can be obtained from few sources [Arita, 2003; Kotera *et al.*, 2004, (KEGG rpair); Mu *et al.*, 2007 (carbon fate maps)]. Unfortunately the databases may contain errors or inconsistencies and thus, the mappings for flux analysis models require curation or at least an inspection. If the atom transfers of interest are not found in the databases, one is forced to go into mecha-

nisms of the reactions to resolve them. Furthermore, label scrambling in symmetrically reacting compounds need to be taken into account [Bernhard and Tompa, 1990]. Software such as ReMatch (http://sysdb.cs.helsinki.fi/ReMatch/) [Pitkänen *et al.*, 2008] and OpenFLUX [Quek *et al.*, 2009] aid in setting up and sharing metabolic models that include atom mappings.

The flux estimation methods that exploit ^{13}C-labelling data can be divided into two sub categories: global iterative fitting and local flux ratio analysis (possibly followed by direct flux estimation) methods. In the following chapters the features of both types of computational methods and established protocols and software are presented.

Global iterative fitting

Global iterative fitting requires modelling of label propagation in the metabolic network and set up of balance equations generally for each isotopomer. For a metabolite with n carbons there will be 2^n possible isotopomers. There will be an extremely high number of isotopomer balance equations in the system and many of them are nonlinear. Iterative fitting searches for the best fit between the observed and simulated labelling patterns [Wiechert *et al.*, 2001; Antoniewicz *et al.*, 2006]. Iteration is initiated from a guessed or a random flux distribution and ^{13}C-labelling patterns of metabolites are simulated with the model. The simulated ^{13}C-labelling patterns are compared to the observed ones and the iteration is continued until a minimum of the difference or a difference under a threshold between the simulated and the observed ^{13}C-labelling patterns is reached. The method returns a single flux distribution that gives the best global fit to all the measured data that was utilised as input.

Other methods to model ^{13}C-labelling patterns than isotopomers have been developed since the original task of simulating numerous isotopomers is computationally highly demanding. Transformation of isotopomers into cumomers enabled analytical solution as solving cascades of linear equations [Wiechert *et al.*, 1999]. Cumomers are by definition certains sums of isotopomers and cumomer fractions can incorporate both positional enrichments and isotopomer fractions. Bondomers were introduced for modelling of label propagation in uniform ^{13}C-labelling experiments [van Winden *et al.*, 2002]. Bondomers are isomer entities that differ only in numbers and positions of intact carbon-carbon bonds. Utilisation of bondomers instead of isotopomers or cumomers decreases the number of mass balance equations and similarly as isotopomers can be trans-

formed into cumomers, bondomers can be transformed into cumulative bondomers. Like cumomers cumulative bondomers enable analytical solution of sequence of linear equations. Recently an elementary metabolite units (EMUs) framework was introduced to further reduce the computational time required to simulate isotopic labelling patterns [Antoniewicz *et al.*, 2007]. EMUs are any distinct subsets of metabolites' atoms and the reaction network is decomposed into EMU reactions and a minimum amount of information required for simulations is identified. EMU framework is compatible for simulation of any type of isotopic labelling.

^{13}C-FLUX software was for a long time the only publicly available software framework for ^{13}C-MFA [Wiechert *et al.*, 2001]. It is compatible with all kinds of measurement data and provides also statistical algorithms for analysing the results. EMU framework is utilised in OpenFLUX, which is recent user-friendly software for all the steps of ^{13}C-MFA, from model building to statistical analyses [Quek *et al.*, 2009].

The drawbacks of the iterative fitting methods are that it is difficult to assure that the method reached a global minimum instead of just a local one [Ghosh *et al.*, 2005]. Moreover, if there is not enough data the method returns merely random points from the solution space but still cannot define the feasible solution space.

Local flux ratio analysis

Local flux ratio analysis utilises directly the ^{13}C-labelling data to deduce ratios of converging fluxes in the network. Thus the inaccuracies in the data or in the assumptions or errors in the network model affect the results only locally in contrast to the global methods [Zamboni *et al.*, 2009]. Algebraic equations that relate the ^{13}C-labelling pattern of a junction metabolite to the relative fluxes through the branching pathways are formulated. If it is possible to solve relative fluxes for every pair of alternative pathways in the network model i.e. as many as there are degrees of freedom in the stoichiometric model, then the determined flux ratios as additional constraints render the MFA system solvable. Approach was introduced by Fischer *et al.*, (2004) and has been implemented as software FiatFlux [Zamboni *et al.*, 2005]. The frameworks for flux ratio analysis are described in the following paragraphs.

Metabolic flux ratio analysis

Metabolic flux ratio (METAFoR) analysis was initially developed to rely on uniform ^{13}C-labelling approach by the biosynthetically directed ^{13}C-labelling of the proteinogenic amino acids and following analysis of ^{13}C-labelling patterns by two-dimensional NMR spectroscopy [Szyperski 1995; Szyperski *et al.*, 1999]. Since the carbon backbones of metabolic intermediates of central carbon metabolism are conserved in synthesis of proteinogenic amino acids and the amino acid synthesis pathways were well known for *E. coli*, Szyperski (1995) back propagated the ^{13}C-labelling patterns from the amino acids to metabolites and derived equations for ratios of converging fluxes in central carbon metabolism. Later Maaheimo *et al.* (2001) extended the method and derived flux ratio equations for compartmental metabolism of eukaryotic *S. cerevisiae*. The ^{13}C-labelling patterns of eight metabolic intermediates of central carbon metabolism of *S. cerevisiae* can be determined and utilised as parameters in the flux ratio equations. The equations derived for eukaryotic metabolism have then been utilised in analysis of metabolic states of at least the following other yeasts and a fungus *P. pastoris* [Sola *et al.*, 2004], *P. stipitis* [Fiaux *et al.*, 2003], *P. anomala* [Fredlund *et al.*, 2004] and *T. reesei* (Publication III).

The original analytical method in METAFoR analysis was ^{1}H-^{13}C HSQC (heteronuclear single quantum coherence) NMR spectroscopic experiment [Croasmun and Carlson, 1994] were the proton bound ^{13}C-nuclei and the ^{13}C-labelling status of the adjacent carbon nuclei are detected. In the experiment, signals from ^{13}C-nuclei in proteinogenic amino acids are spread into two-dimensions and found at characteristic chemical shifts in proton and carbon dimensions. According to the ^{13}C-labelling status of the adjacent carbon nuclei, different signal fine structures are formed. Coupling to an adjacent ^{13}C-nucleus splits the signal (Figure 3). All the different signal fine structures are observed on top of each other and their fractional volumes correspond quantitatively to the fractions of different three carbon isotopomers with a central ^{13}C nucleus. The fractions of different three carbon isotopomers with a central ^{13}C nucleus are obtained by iterative fitting of simulated signal fine structures on the whole multiplet signal. FCAL is a software developed for the iterative fitting and following calculation of fragmentomers from the fractions of different signal fine structures with the equations derived by Szyperski (1995) [Szyperski *et al.*, 1999]. The probabilistic equations take into account the fraction of uniformly ^{13}C-labelled substrate, the natural ^{13}C

abundance in the rest of the carbon source and the fraction of biomass synthesized during the ^{13}C-labelled feed [Szyperski *et al.*, 1995].

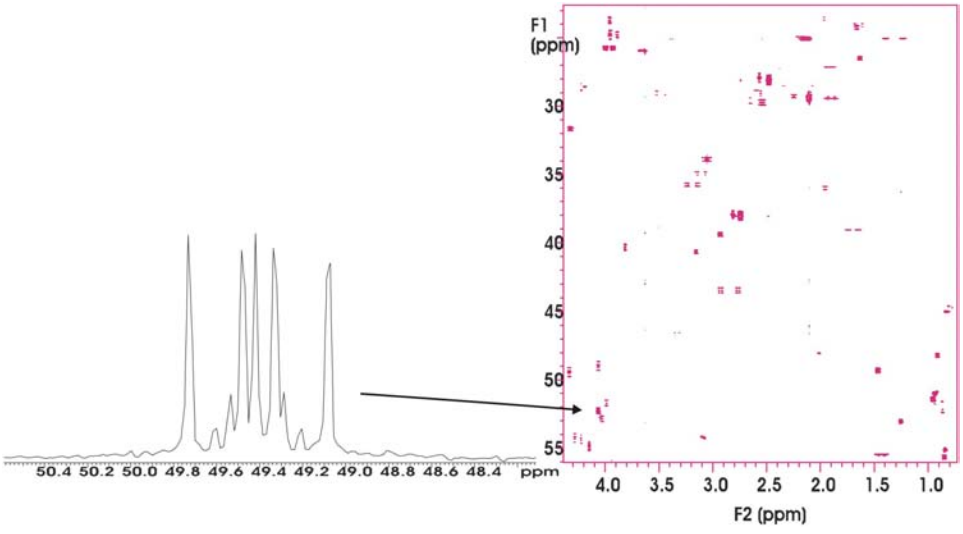

Figure 3. An example of the ^{13}C finestructure of Glu-Cα extracted from a two-dimensional ^{1}H-^{13}C HSQC NMR spectrum.

The sample preparation for METAFoR analysis is simple [Szyperski *et al.*, 1999]. The harvested biomass is just hydrolysed (6 M HCl, +110 °C). Since the information on the *in vivo* fluxes is naturally amplified in biomass, sensitivity does not limit the NMR measurement. For NMR analysis ash is removed from the hydrolysed sample and the solvent is switched to D_2O. The detection of the ^{13}C-labelling patterns of proteinogenic amino acids by 2D NMR spectroscopy can be performed without any separation steps in sample preparation.

Fischer and Sauer (2003) extended the flux ratio analysis to GC-MS as analytical method and initiated the utilisation of mixed ^{13}C-labelling, a combination of positional and fraction uniform ^{13}C-labelling. The introduction of positional label was meaningful in combination of the switch to GC-MS analysis, because even though GC-MS cannot directly quantify positional enrichments, enrichment of label in carbon fragments can be detected. Later the GC-MS analysis has been extended to analysis of ^{13}C-labelling in cell cultures in 1 ml deep-well microtiter plates enabling high throughput metabolic flux profiling [Fischer *et al.*, 2004].

Conventionally the flux ratio equations have been manually derived by experts of the metabolism of the organism under study being able to set up mean-

ingful assumptions of the fluxes for every new organism, metabolic network, ^{13}C-labelling of the carbon source and an analytical platform. Only recently a framework for systematic derivation of the equations from a given metabolic network model was introduced (Publication IV).

Local flux ratios as additional constraints in metabolic flux analysis

Local flux ratios determined from ^{13}C-labelling experiments are experimental information that can be utilised as additional constraints in a conventional MFA system [Fischer *et al.*, 2004]. If every branching point in the network model can be constrained, the system renders solvable. Fischer *et al.* (2004) applied the approach to metabolism of *E. coli*. Constraint equations were set up and the net fluxes in central carbon metabolism were solved by constrained nonlinear optimisation with Matlab function *fmincon*. Later Fredlund *et al.* (2004) set up the constraint equations for compartmentalised eukaryotic network model of *Pichia anomala*. Both studies utilised GC-MS determined ^{13}C-labelling patterns for flux ratio determination. Zamboni *et al.* (2005) implemented the approach as software FIATFLUX for net flux determination in three organisms: *E. coli, Bacillus subtilis, S. cerevisiae*. FIATFLUX contains two Matlab based modules. The first one is for determining local flux ratios of GC-MS massisomer data from ^{13}C-experiments and the second module is for estimating the net fluxes utilising the local flux ratios as additional constraints. The network models for determining the local flux ratios are fixed but the stoichiometric models for net flux determination are open for users. Constraints from ^{13}C-labelling experiments often enable leaving out the cofactors form metabolite balancing. Cofactor balancing is highly error prone since the cofactor specificities vary in isoenzymes and they are not precisely known. In Publication I net fluxes in central carbon metabolism of *S. cerevisiae* were determined under different oxygenation conditions by performing ^{13}C-labelling experiments, utilising NMR spectroscopy based METAFoR analysis instead of GC-MS analysis for determining the local flux ratios and then solving the net fluxes by optimisation having the flux ratios as additional constraints.

2.3.2.3 ^{13}C-metabolic flux analysis in large scale networks

In prolonged ^{13}C-labelling experiments all cell components and metabolites become ^{13}C-labelled but analysis of ^{13}C-labelling patterns of only limited number of metabolites is feasible with current MS and NMR techniques. However, the

measurements with both methods can be directed to specific compounds or methods can be adjusted for analysis of desired compounds. Thus, computational ^{13}C-experiment design has been investigated [Möllney *et al.*, 1999]. In addition to optimal ^{13}C-labelling designs for flux resolution [Möllney *et al.*, 1999; Araúzo-Bravo and Shimizu, 2003], sets of most informative compounds for ^{13}C-MFA can be computationally determined to target the analysis of ^{13}C-labelled samples [Rantanen *et al.*, 2006].

^{13}C-MFA becomes unfeasible in large-scale networks because of limited ^{13}C-labelling data and because of the size of the equation system. For iterative methods the number of equations grows fast as the number of additional isotopomer balances explodes when the network size increases. ^{13}C-MFA with local flux ratios as additional constraints would be computationally feasible in larger networks if it was possible to determine sufficiently many local flux ratios. Thus, the ^{13}C-labelling data limits that approach.

However, few flux analysis studies in large-scale networks have been carried out. Blank *et al.* (2005) determined the net fluxes in the central carbon metabolism of *S. cerevisiae* with local flux ratios as additional constraints in MFA and estimated the fluxes in large-scale network by minimisation of fluxes outside the central carbon metabolism. Quite recently Suthers *et al.* (2007) applied iterative flux determination approach to a large scale network of *E. coli*. Even though they included both cofactor balancing and ^{13}C-labelling data, multiple local optima that were statistically indistinguishable were identified. The ^{13}C-labelling data was insufficient for reliable flux determination in the large-scale model.

3. Aims of the research

The research included in the thesis has concentrated on studying the metabolism of mainly three organisms: yeasts *S. cerevisiae* and *P. pastoris* and a fungus *T. reesei* all of which are important production organisms in biotechnology. The focus has been on the process streams of the cell factories, the metabolic fluxes, under conditions of interest for development of production processes. The quantitative studies of the distribution of intracellular fluxes of different organisms and under different growth conditions have required development of modelling of metabolism for the analysis of fluxes (Publications I, II, III). In Publication IV and somewhat also in Publication I, computational methods for quantitative analysis of the intracellular fluxes were developed. In Publication V a novel step in the fungal metabolic pathway of catabolism of plant material compound D-galacturonic acid was identified and thus a previously unknown reaction node and interactions to reactant metabolites were set in the fungal metabolic network. The analytical tool utilised in the detection of ^{13}C-labelling in the analyses of metabolic fluxes, namely the NMR spectroscopy, was applied to verify the product of the novel enzyme.

3.1 Oxygen dependence of fluxes and underlying regulation in *S. cerevisiae*

Understanding the regulation of redox homeostasis of the organism is important for any metabolic engineering project because redox homeostasis binds together functions of even distant pathways. Redox homeostasis of an organism is naturally affected by the oxygenation of the culture that is a major parameter in industrial bioprocesses. Oxygenation is also one of the factors that most contribute to the cost of a bioprocess. The important bioproduction and model organism *S. cerevisiae* is known to exhibit various states of energy metabolism depending

on the prevailing growth conditions. However, the effect of different low oxygen conditions on the metabolic state of *S. cerevisiae*, thus the distribution of the metabolic fluxes, has not been thoroughly studied. Thus, the aim of the Publication I was to quantify the dependence of the intracellular flux distributions of *S. cerevisiae* on oxygen provision and to study the maintenance of redox homeostasis in the different oxygenation conditions. It was essential to quantify the intracellular net fluxes without including the redox cofactors in the metabolite mass balances, because the conditions were expected to strongly affect the redox balancing. Therefore, in Publication I ^{13}C-MFA was utilised to determine the net fluxes.

3.2 Two carbon source case of methanol and glycerol utilisation by *P. pastoris*

Methylotrophic yeast *P. pastoris* is a host organism for industrial production of heterologous proteins. Strong inducible promotors of the genes of methanol utilisation pathway in the peroxisomes are utilised to induce the expression of recombinant proteins. Thus, metabolism during the co-utilisation of inducer methanol and a carbon source is of interest from process development point of view. The METAFoR analysis of *P. pastoris* [Sola *et al.*, 2004] was extended to a two-carbon source co-utilization. The eukaryotic model for METAFoR analysis [Maaheimo *et al.*, 2001] was likewise extended with methanol utilisation pathway. The ^{13}C-labelling with the uniform labelling approach was performed in continuous cultures of *P. pastoris* growing on different methanol/glycerol mixtures and at two growth rates to probe the intracellular metabolic state, the ratios of intracellular metabolic fluxes, in different possible process conditions. The aim was to provide valuable information for process optimisation of recombinant protein production with *P. pastoris*.

3.3 Path identification and the effect of carbon catabolite repression on metabolic fluxes in *T. reesei*

The efficient protein expression machinery of *T. reesei* has been widely investigated but even despite the wide industrial importance the metabolism of *T. reesei* has not been largely studied and its potential is still not known. The genome of *T. reesei* has recently been published [Martinez *et al.*, 2008] but the fungus still lacks a genome-wide metabolic reconstruction.

The transcriptional response to preferred and repressive carbon source glucose and to a neutral carbon source has been studied by Chambergo *et al.* (2002). They found that excess glucose does not repress the respiratory pathway genes to the similar extent as in *S. cerevisiae* that is adapted to fast utilization of sugars by fermentation in nutrient rich environments. However, the intracellular *in vivo* fluxes of *T. reesei* were not previously studied and the effect of the different conditions of carbon catabolite repression on the intracellular fluxes has not been known.

Because of the lack of a metabolic reconstruction of *T. reesei* the biosynthetic pathways of proteinogenic amino acids in *T. reesei* were reconstructed with a recent computational pathway analysis method ReTrace [Pitkänen *et al.*, 2009] in Publication III. The reconstruction of the biosynthetic pathways of proteinogenic amino acids was essential for the application of the METAFoR analysis to quantitatively probe the intracellular flux distributions in *T. reesei*. The ^{13}C-labelling of *T. reesei* for METAFoR analysis was performed for cells growing on preferred and repressive carbon source glucose and on sorbitol. The effect of induction of cellulase gene expression on the ratios of intracellular fluxes was also studied by ^{13}C-labelling of a *T. reesei* culture growing on sorbitol induced with a small addition of inducer sophorose.

3.4 Framework for analytical determination of flux ratios

The previously established computational methods for quantitative analysis of the metabolic fluxes from ^{13}C-isotopomer measurement data relied either on manual derivation of analytic equations constraining the fluxes or on numerical solution of a highly nonlinear system of isotopomer balance equations. In the first approach, analytic equations were to be tediously derived for each organism, particular growth conditions and substrate or labelling pattern, by a domain expert while in the second approach, the global nature of an optimum solution is difficult to prove and comprehensive measurements of external fluxes to augment the ^{13}C-isotopomer data were typically required. A framework for an automatic and systematic derivation of equation systems constraining the fluxes from the model of the metabolism of an organism was developed. The framework was designed to be general for all metabolic network topologies, ^{13}C-isotopomer measurement techniques, carbon sources, and carbon source ^{13}C labelling patterns.

3.5 NMR spectroscopy as a tool in pathway identification

Fungal metabolism comprises a lot of hitherto unknown potential for biotechnology. For example a fungal pathway for D-galacturonic acid catabolism has been only partly known. D-galacturonic acid is a major component of pectin that is abundant in plant material. The first step in the fungal pathway of D-galacturonic acid catabolism was previously identified and it is an NADPH-specific D-galacturonic acid reductase generating L-galactonate [Kuorelahti *et al.*, 2005]. The next reaction in the pathway, a novel enzyme that converts L-galactonate to L-threo-3-deoxy-hexulosonate was then identified in *T. reesei*. The active enzyme was produced in the heterologous host *S. cerevisiae* and characterized. The reaction product of the enzyme L-galactonate dehydratase was analysed and identified by NMR spectroscopy. 1D and 2D NMR spectroscopic experiments were utilised for the identification of the reaction product directly in the reaction mixture.

4. Research methods

4.1 Strains

In Publication I *S. cerevisiae* CEN.PK113-1A (MATα, URA3, HIS3, LEU2, TRP1, MAL2-8c, SUC2) strain was employed. The strain was kindly provided by Dr. P. Kötter (Institut für Mikrobiologie, J.W. Goethe Universität Frankfurt, Germany) [de Jong-Gubbels *et al.*, 1998] and prior to the experiments stored in glycerol (30% v/v) at -80 °C [Wiebe *et al.*, 2008]. In Publication II a prototrophic *P. pastoris* strain expressing a heterologous protein, a *Rhizopus oryzae* lipase, under the transcriptional control of the *aox-1* promoter was employed. *P. pastoris* X-33/pPICZαA-ROL [Minning *et al.*, 2001] is the wild-type phenotype X-33 strain (Invitrogen) with the pPICZαA-derived expression vector (Invitrogen) containing the *ROL* gene, pPICZαA-ROL, integrated in its *aox-1* locus. In Publication III *T. reesei* strains QM6a (wild type) [Mandels and Reese, 1957] and QM6a with deleted *cre1* gene (unpublished) were employed. In Publication V the *S. cerevisiae* strain CEN.PK2-1D (VW-1B) was employed as the host for the heterologous expression of a *T. reesei* enzyme and was the source of the extract in the NMR spectroscopic analyses. *T. reesei* strains Rut C-30 or QM6a were employed otherwise.

4.2 Cultivations

In Publications I, II and III the organisms were cultivated either in continuous cultures in fermentors or in batch cultures in flasks. Continuous cultivations provide highly controlled culture conditions were a single parameter can be varied while everything else is kept constant. The growth rate of an organism is set by the rate of the feed and by the flow out of the reactor. Continuous cultivation operates in steady state continuous mode when all the variables have time inde-

pendent values. The steady state continuous mode is usually eventually obtained when after the initiation of the culture in a batch mode all the feed variables are kept constant. The reactor is assumed to be an ideal bioreactor where the properties of the effluent are identical to the properties of the culture in any point of the reactor. The assumption is reasonable when the mixing of the reactor is efficient. Because of the strict control, continuous cultivation provides possibilities for sampling during steady state conditions and highly reproducible cultures. The continuous cultivations utilized are described in more detail in Publications I and II and the cultivation set up in Publication I also in Wiebe *et al.* (2008).

Batch cultures in flasks are much simpler to prepare than continuous cultures. The filamentous growth of *T. reesei* complicates the bioreactor cultivations and thus in Publication II *T. reesei* was cultivated in flasks. After the initial lag-phase, during the early exponential phase in a flask culture, the growth conditions are still almost unchanged from the initial culture conditions and the organism is growing on its maximal growth rate. After passing the early exponential phase the growth conditions are not precisely known anymore. The *T. reesei* batch cultures in flasks are described in detail in Publication III.

All the cultivations for METAFoR analysis in Publications I, II, III were performed on minimal medium without amino acids. The media of the main cultures were as follows. Information on the media for inoculates etc can be found in Publications II and in Wiebe *et al.* (2008) for cultures in Publication I. In Publication I yeast was grown on defined minimal medium [Verduyn *et al.*, 1992], with 10 g glucose l^{-1} as carbon source, and supplemented with 10 mg ergosterol l^{-1} and 420 mg Tween 80 l^{-1} (a source of oleic acid). In Publication II *P. pastoris* was grown on minimal medium containing (per $1x10^{-3}$ m^3 of deionized water): Yeast Nitrogen Base (YNB), $0.17x10^{-3}$ kg; $(NH_4)_2SO_4$, $5x10^{-3}$ kg; glycerol and methanol (different ratios on w/w basis), $10x10^{-3}$ kg (total). In Publication III *T. reesei* was grown on minimal medium: $(NH_4)_2SO_4$ 7.6 g/l, KH_2PO_4 15.0 g/l, 2.4 mM $MgSO_4$, 4.1 mM $CaCl_2$, $CoCl_2$ 3.7 mg/l, $FeSO_4 \cdot 7H_2O$ 5 mg/l, $ZnSO_4 \cdot 7H_2O$ 1.4 mg/l, $MnSO_4 \cdot 7H_2O$ 1.6 mg/l, pH adjusted to 4.8 with KOH, supplemented with 2% (w/v) carbon source glucose or sorbitol.

4.3 Biosynthetically directed fractional ^{13}C-labelling

Biosynthetically directed fractional (BDF) ^{13}C-labelling was performed for *S. cerevisiae*, *P. pastoris* and *T. reesei* in studies in Publications I, II and III, respectively. *P. pastoris* and *S. cerevisiae* ^{13}C-labellings were performed in

chemostats while *T. reesei* was ^{13}C-labelled in batch cultures in flasks. After reaching a metabolic steady state in *S. cerevisiae* glucose-limited chemostats, as determined by constant physiological parameters including biomass production, carbon dioxide evolution and oxygen uptake rates (CER and OUR), alkali utilisation, and subsequently confirmed by the observation of constant extracellular and intracellular metabolites and gene transcription, 10 % of the carbon source in the medium was replaced with [U-^{13}C]glucose (Publication I). 10% [U-^{13}C]glucose was fed for 1.5 volume changes. *P. pastoris* was cultivated in chemostat on different glycerol/methanol mixtures until a metabolic steady state was reached as indicated by a constant cell density and constant oxygen and CO_2 concentrations in the bioreactor exhaust gas (Publication II). Then the culture was fed with medium containing about 10% (w/w) uniformly ^{13}C-labelled and 90% unlabelled amounts of each substrate simultaneously for one volume change. *T. reesei* ^{13}C-labelling was performed with two different set ups (Publication III). In the first set up ^{13}C-labelled carbon source was provided directly in the medium that was inoculated: in 2% (w/v) glucose minimal medium containing 10% (w/w) [U-^{13}C]glucose and in 2% (w/v) sorbitol minimal medium containing 10% (w/w) [U-^{13}C]sorbitol. In the second set up in exponential growth phase the six flask cultures were combined and then the culture broth was divided into six flasks. The final concentration of 1mM sophorose was introduced into three of the six replicate 2 l flasks to induce cellulase gene expression in *T. reesei*. An identical volume of water was added to the three control cultures. Three hours after the induction, when cellulase gene expression was expected to be at a moderate level [Ilmén *et al.*, 1997], 0.4 g of [U-^{13}C]sorbitol was added to all six cultures to initiate BDF ^{13}C-labelling. The addition of 0.4 g of [U-^{13}C]sorbitol at this time was estimated to result in a [U-^{13}C]sorbitol fraction of about 10% of the total sorbitol in the culture medium.

During steady state growth in chemostats and during quasi-steady state growth in the exponential growth phase in batch cultures the intracellular metabolic fluxes remain constant and determine the labelling patterns of carbon backbones of proteinogenic amino acids formed in biosynthesis. In the *T. reesei* cultures that were inoculated in fractionally ^{13}C-labelled medium the fraction of the initial biomass could be neglected and all the biomass is formed during ^{13}C-labelling (Publication III). Also when the ^{13}C-label was introduced after the induction of cellulase gene expression the initial biomass could be neglected. However, in chemostat cultures the fraction of biomass formed during the ^{13}C-labelling depends on the growth rate (equals dilution rate D) and the duration of the

^{13}C-labelled feed and can be estimated from the first order wash-out kinetics (Publications I and II). The *S. cerevisiae* chemostat cultures at D = 0.1 h^{-1} were fed with fractionally ^{13}C-labelled carbon source for 1.5 residence times which equals 15 h with the particular growth rate (Publication V). The *P. pastoris* chemostat cultures at D = 1.39 x 10^5 s^{-1} and 4.44 x 10^5 s^{-1} were fed with fractionally ^{13}C-labelled carbon source for 1.0 residence time (Publication II).

4.4 Sampling

The biomass samples of *S. cerevisiae* and *P. pastoris* were harvested by centrifugation (Publications I and II, respectively) and *T. reesei* mycelium by filtration (Publication III). The cell pellets and the filtrated mycelium were suspended into 10 ml of 6 M HCl and the biomass was hydrolysed in sealed glass tubes at +110 ºC for 22 h. The suspensions were dried and dissolved in H_2O for filtration through 0.2 μm filters. The filtrates were vacuum-dried and dissolved in D_2O for NMR experiments. The pH of the samples was below 1 due to residual HCl.

4.5 NMR spectroscopy

In Publication V the reaction mixture of *S. cerevisiae* extract of the strain expressing the lgd1 from *T. reesei* and 110 mM L-galactonate was analysed by NMR after different time intervals. The reaction product was identified by comparing the NMR spectrum of the reaction mixture with the NMR spectrum of pure L-galactonate. The NMR experiments were carried out at +23 °C on a Varian Inova spectrometer operating on a proton frequency of 500 MHz. The spectral widths of the 1D ^{1}H and ^{13}C spectra were 5000 Hz and 30 675 Hz, respectively. In two-dimensional homonuclear correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) experiments [Croasmun and Carlson, 1994], the spectral widths were 3400 Hz. The spinlock time for magnetization transfer along coupled nuclei in the TOCSY was 80 ms. In two-dimensional heteronuclear ^{1}H-^{13}C HSQC experiment the spectral widths in ^{1}H and ^{13}C dimensions were 1654 Hz and 10 000 Hz, respectively.

For METAFoR analysis in Publications I, II and III the ^{1}H-^{13}C HSQC NMR spectra [Croasmun and Carlson, 1994] were acquired at +40 ºC on a Varian Inova spectrometer operating at on a proton resonance frequency of 600 MHz essentially as described [Szyperski, 1995]. For each sample two spectra were acquired focusing on the aliphatic and aromatic regions. The spectral widths in

the aliphatic spectra were 6000 Hz and 5100 Hz in the ^{1}H and ^{13}C dimensions, respectively. The narrow spectral width in the ^{13}C dimension leads to back-folding of part of the signals to the empty regions of the spectrum. The spectral widths for the aromatic spectra were 6000 Hz and 2815 Hz in the ^{1}H and ^{13}C dimensions, respectively. The spectra were processed using the standard Varian spectrometer software VNMR (version 6.1, C).

4.6 Metabolic flux ratio analysis

The software FCAL (R.W. Glaser; FCAL 2.3.1) [Szyperski *et al*., 1999] was used for the integration of ^{13}C-scalar fine structures of proteinogenic amino acid carbon signals in the ^{1}H-^{13}C HSQC NMR spectra and the calculation of relative abundances of intact carbon fragments originating from a single source molecule of glucose as in Szyperski (1995).

The nomenclature utilised for the intact carbon fragments, fragmentomers, was initially described by Szyperski (1995) and has briefly been explained in Introduction. Since the carbon backbones of eight metabolic intermediates are conserved in amino acid synthesis, fragmentomer information obtained from proteinogenic amino acids can be traced back to the intermediates of central carbon metabolism [Szyperski 1995; Maaheimo *et al.*, 2001]. Mass balance equations of specific carbon fragments of the metabolic intermediates can be formulated from the propagated fragmentomer information to solve ratios of fluxes in junctions of central carbon metabolism. In Publication I the metabolic flux ratio (METAFoR) analysis relied on the compartmentalized metabolic model of *S. cerevisiae* central carbon metabolism and some of the flux ratios were calculated as formulated by Maaheimo and co-workers (2001). However, some flux ratio calculations were redefined as follows. The nomenclature of the metabolites with differentially conserved C-C connectivities that have been back propagated from the ^{13}C-labelling patterns of amino acids is <metabolite_abbreviation_> and then the following characters denote the status of the bonds in the carbon chain of the metabolite: 1 stands for an intact bond, 0 for a cleaved bond and x for either of the two. The corresponding amino acids fragmentomers are named as was explained above.

The fraction of Pep originating from phosphoenolpyruvate carboxykinase activity, denoted by X_{PEPck}, was calculated from the ratio of the fraction of Pep molecules containing an intact C1-C2 fragment and a cleaved bond between C2 and C3 (Pep_10) and the fraction of Oaa_{cyt} molecules containing the equivalent

fragments (Oaa_{cyt}_10x) (Equation 1). These fragments cannot originate from glycolysis or from the PPP [Maaheimo *et al.*, 2001]. Phe-Cα, Tyr-Cα and Asp-Cα, Thr-Cα can be traced back to the C2 of Pep and Oaa_{cyt}, respectively [Maaheimo *et al.*, 2001] (Equation 6).

$$X_{PEPck} = Pep\_10/Oaq_{cyt}\_10x = \left[f^{(2*)}\right]\{Phe,Tyr-C\alpha\}/\left[f^{(2*)}\right]\{Asp,Thr-C\alpha\} \quad (6)$$

The Oaa_{mit} molecules originating from Oga through the TCA cycle possess cleaved C2-C3 bonds. The fraction of Oaa_{mit} originating from transport over the mitochondrial membrane from Oaa_{cyt} was solved from a mass balance of intact C2-C3 fragments in Oaa_{mit}. The conserved connectivity of the C2-C3 fragment in Oaa_{mit} could be propagated back from Glu-Cα and Pro-Cα carbons that represent the C2 carbon in Oga, since the C2-C3 fragment of Oaa_{mit} is conserved in the TCA cycle as the C2-C3 fragment of Oga. The fraction of Oaa_{mit} from Oaa_{cyt}, denoted by $X_{Oaa-transport}$, was calculated as a ratio of intact C2-C3 fragments in Oga and Oaa_{cyt} (Equation 7).

$$\begin{aligned} X_{Oaa-transport} &= Oga\_x1xx/Oaa_{cyt}\_x1x \\ &= \left(\left[f^{(2)}+f^{(3)}\right]\{Glu,\mathrm{Pro}-C\alpha\}\right)/\left(\left[f^{(2)}+f^{(3)}\right]\{Asp,Thr-C\alpha,Asp-C\beta\}\right) \end{aligned} \quad (7)$$

The fraction of Oaa_{cyt} originating from Pyr_{cyt}, denoted by $X_{Oaa_{cyt}\_from\_Pyr_{cyt}}$, was solved from the mass balance of intact C2-C3 fragments (Equation 8). Since the flux from Pep to Pyr_{cyt} through phosphoenolpyruvate kinase and further through pyruvate carboxylase to Oaa_{cyt} could be assumed to be irreversible, the C2-C3 fragments of Pep were used in the mass balance equations. The conserved connectivity of the C2-C3 fragment in Pyr_{cyt} could be observed from Phe-Cα and Tyr-Cα that represent the C2 carbon of Pep (Equation 8).

$$\begin{aligned} X_{Oaa_{cyt}\_from\_Pyr_{cyt}} &= \left(Oaq_{cyt}\_x1x - Oga\_x1xx\right)/\left(Pep\_x1 - Oga\_x1xx\right) \\ &= \frac{\left[f^{(2)}+f^{(3)}\right]\{Asp,Thr-C\alpha,Asp-C\beta\}-\left[f^{(2)}+f^{(3)}\right]\{Glu,\mathrm{Pro}-C\alpha\}}{\left[f^{(2)}+f^{(3)}\right]\{Phe,Tyr-C\alpha,C\beta\}-\left[f^{(2)}+f^{(3)}\right]\{Glu,\mathrm{Pro}-C\alpha\}} \end{aligned} \quad (8)$$

The upper and lower bounds for Pyr_{mit} originating from the malic enzyme reaction, denoted by X_{MAE_ub} and X_{MAE_lb} respectively, were calculated from a mass balance of intact C2-C3 fragments of Pyr_{mit} (Equations 9 and 10). The upper and lower bounds were obtained from the assumption that the substrate fragment for malic enzyme has an equally conserved connectivity as Oga and

Oaa_{mit}. The intact fragments in Oaa_{mit} were obtained from the intact fragments in Oga since the C2-C3-C4 fragment of Oaa_{mit} is conserved in the TCA cycle in synthesis of Oga. The intact fragments in biosynthetic precursor Oga were deduced from the *f*-values of Glu and Pro carbons (Equations 9 and 10).

$$X_{MAE\_ub} = (Pep\_x1 - Pyr_{mit}\_x1)/(Pep\_x1 - Oga\_x1xx) = \frac{\left[f^{(2)} + f^{(3)}\right]\{Phe, Tyr - C\alpha, C\beta\} - \left[f^{(2)} + f^{(3)}\right]\{Ala - C\alpha, C\beta\}}{\left[f^{(2)} + f^{(3)}\right]\{Phe, Tyr - C\alpha, C\beta\} - \left[f^{(2)} + f^{(3)}\right]\{Glu, \mathrm{Pr}o - C\alpha\}} \quad (9)$$

$$X_{MAE\_lb} = (Pep\_x1 - Pyr_{mit}\_x1)/Pep\_x1 = \frac{\left[f^{(2)} + f^{(3)}\right]\{Phe, Tyr - C\alpha, C\beta\} - \left[f^{(2)} + f^{(3)}\right]\{Ala - C\alpha, C\beta\}}{\left[f^{(2)} + f^{(3)}\right]\{Phe, Tyr - C\alpha, C\beta\}} \quad (10)$$

4.7 Metabolic modelling for ^{13}C-metabolic flux analysis

In Publication I metabolic flux analysis (MFA) was used to determine intracellular net fluxes of *S. cerevisiae* under different conditions of oxygen provision (20.9%, 2.8%, 1.0%, 0.5% and 0.0% O_2 in the chemostat inlet gas), with METAFoR analysis providing additional experimental constraints to render the MFA system solvable [Fischer *et al*., 2004]. A stoichiometric model of central carbon metabolism of *S. cerevisiae* was formulated. The system boundary was set around the central carbon metabolism and the model thus included the glycolytic and the pentose phosphate pathways, the TCA cycle and the fermentative pathways, production of glycerol and anabolic fluxes from metabolic intermediates to biosynthesis. The glyoxylate cycle was omitted from the model since the METAFoR analysis data showed that the pathway was inactive (Publication I). Separate pools of Pyr, AcCoA and Oaa in the two cellular compartments, cytoplasm and mitochondria, were included in the model. Thus, they were modelled as two distinct metabolites. Mal was lumped in the same pool with Oaa_{mit}. Also the pentose phosphates formed a single pool and the triose phosphates were combined in the pools of G3P and Pep. DHAP, the precursor for glycerol synthesis, was also combined with the G3P pool. Lumping of the metabolite pools in the model is reasonable when it is meaningful to assume fast exchange between the metabolites, faster than between the metabolites and other compounds. TCA cycle metabolites were represented by the pools of citrate, Oga and Oaa_{mit}.

Scrambling of ^{13}C-labels in the symmetric molecules succinate and fumarate was taken into account [Bernhard and Tompa, 1990].

However, despite the symmetry, partial channelling of succinate and fumarate has also been seen [Sumegi *et al.*, 1993]. The transport of Pyr and Oaa across the mitochondrial membrane were included in the model but the transport of AcCoA, the final step of the cytosolic pyruvate dehydrogenase (PDH) bypass, was omitted since exogenous carnitine would be required for the carnitine shuttle to be active [Lange, 2002; Swiegers *et al.*, 2001; van Roermund *et al.*, 1999], and carnitine was not provided in the medium. In addition, carnitine acetyltransferase activity has not been detected in *S. cerevisiae* grown in anaerobic chemostats at 0.1 h^{-1} [Nissen *et al*., 1997]. However, contradictory observation has been made by Frick and Wittmann (2005) in *S. cerevisiae* strain ATCC 32167. Pyruvate by-pass via transport of AcCoA into mitocnodria was observed to be active during both respirative and fermentative growth under different dilution rates between 0.10 h^{-1} and 0.45 h^{-1}. Since acetaldehyde can freely diffuse across the mitochondrial membrane and acetaldehyde dehydrogenase (EC 1.2.1.3) and AcCoA synthetase (EC 6.2.1.1) enzymes have both been isolated in the mitochodrial proteome [Sickmann *et al*., 2003], PDH bypass could also be partially located in mitochondria and contribute directly to the formation of $AcCoA_{mit}$. In absence of fluxes inducing significantly dissimilar labelling patterns to the C2-C3 fragments of Pyr_{cyt} and Pyr_{mit} i.e. in conditions of low malic enzyme fluxes as observed in this study, ^{13}C-labelling cannot solely reveal the possible contribution of PDH bypass pathways to the carbon flux to mitochondria. However, in the cultivations performed, the expression of *ACS1* encoding the mitochondrial AcCoA synthetase, essential for the contribution of mitochondrial PDH bypass to the formation of $AcCoA_{mit}$, was negligible and the expression of *ACS2* encoding the cytosolic isoenzyme was substantially higher [Wiebe *et al.*, 2008]. Therefore, the mitochondrial PDH bypass was not included in the model.

A model of central carbon metabolism of *S. cerevisiae* with the same extent as above was formulated for development of the systematic and analytic framework for determination of flux ratios in Publication IV. In the model, some simplifications common to ^{13}C-MFA were applied by pooling metabolites whose isotopomer pools can be assumed to be fully mixed (cf. [Kleijn *et al.*, 2007]). Pooling of metabolites was carried for the pentose-phosphates in PPP, phopshotrioses between G3P and Pep in glycolysis, and Oaa and Mal in the TCA cycle. In these cases, pooling was justified by the existence of fast equilibrating, bidirectional reactions between the listed intermediates and the empirical evidence that their

isotopic labelling is not distinguishable with the current analytical tools. Cofactor metabolites were excluded from the model as cofactor specificities and activities are not accurately known for many reactions. The bulk of the carbon mappings of reactions in the metabolic network were provided by ARM project [Arita, 2003]. Carbon mappings from amino acids to their precursors conformed to [Szyperski, 1995] and [Maaheimo *et al.*, 2001]. For empirical verification of the framework it was tested by estimation of flux ratios for junction metabolites in the metabolic network of *S. cerevisiae* from the artificial data generated by the 13C-FLUX software [Wiechert *et al.*, 2001].

In Publication I the metabolic fluxes were modelled as net fluxes so that a net flux in the forward direction was assigned with a positive value and a net flux in the reverse direction was assigned with a negative value. As an exception, the transport of Oaa across the mitochondrial membrane was modelled as two one-directional transport reactions by not allowing negative net fluxes. In *S. cerevisiae* the transport of Oaa across the mitochondrial membrane can occur via mitochondrial Oaa transporter OAC1 facilitated transport [Palmieri *et al.*, 1999].

The stoichiometric model for experiments under 20.9%, 2.8% and 1.0% oxygen conditions consisted of 38 reactions coupling 34 metabolites including duplicated extracellular metabolites and uptake and production fluxes (Publication I Figure 4). The 14 fluxes across the system boundary included glucose uptake and excretion fluxes of ethanol, acetate and glycerol and the fluxes of the metabolic precursors to macromolecule synthesis for biomass production. The METAFoR analysis results were used to identify inactive reactions, to constrain the stoichiometric models for the experiments with 0.5% and 0.0% oxygen by omitting inactive fluxes to avoid numerical problems in optimization. The stoichiometric model for experiments under 0.5% oxygen consisted of 37 reactions, coupling 34 metabolites and excluding the malic enzyme activity from the first model of the network of active reactions. The compartmentalization of central carbon metabolism in anaerobic conditions is evident from the vital anabolic role of mitochondria in the absence of oxygen [Visser *et al.*, 1994]. However, in completely anaerobic conditions only the net transport of Oaa across the mitochondrial membrane is resolvable and the activities of PEPck and malic enzyme reactions cannot be quantified. Since, according to the METAFoR analysis, the PEPck reaction showed only slight activity in the other conditions studied and its activity decreased as the oxygen provided was reduced, it was omitted from the anaerobic stoichiometric model. *MAE1* has been shown to be induced in anaerobic conditions and its possible role in provision of NADPH in mitochondria in

anaerobic conditions has been discussed [Boles *et al.*, 1998]. However, the malic enzyme reaction also showed only slight activity in all the conditions where quantification was possible and had its lowest activity in 0.5% oxygen. Thus, the malic enzyme reaction was omitted from the anaerobic model. Under anaerobic conditions the stoichiometric model of the active pathways consisted of 34 reactions and 34 metabolites.

After including the measured uptake and excretion rates and the rates of metabolic precursor depletion to biomass synthesis, as determined from the composition of *S. cerevisiae* biomass previously reported [Gombert *et al.*, 2001], in the models, the linear equation systems remained underdetermined. The composition of *S. cerevisiae* biomass was assumed to be the same in all the conditions studied, since the biomass composition in the two extreme conditions, i.e. in fully aerobic and in anaerobic conditions, has been experimentally shown to be essentially the same [Gombert *et al.*, 2001; Nissen *et al.*, 1997]. Solvable systems were obtained by further constraining the MFA systems with one to six linearly independent constraints, depending on the structure of the network of active reactions from the METAFoR analysis as described by Fischer and coworkers (2004) for MS ^{13}C-labelling data. Using the constraints from the METAFoR analysis, it was not necessary to include redox cofactor mass balances in the mass balance constraints in ^{13}C MFA. Cofactor mass balances are sources of errors since the correct balancing requires detailed knowledge of the relative activities of different isoenzymes and the enzymes' redox cofactor specificities on a cell wide scale under the studied conditions. Under the conditions of different oxygen provisions, the external conditions posed different challenges on the redox homeostasis systems of the cells and their effect are not known. The mass balances of the metabolites were formulated as a linear equation system as described in [Fischer *et al.*, 2004] (Equation 11):

$$\mathbf{N}_i\mathbf{v} - \mathbf{b} = \mathbf{R}_m \tag{11}$$

where $\mathbf{N}_i$ is the stoichiometric matrix of the active network i determined from the METAFoR analysis fragmentomer data, $\mathbf{v}$ is the flux distribution vector, $\mathbf{b}$ is the vector of the measured extracellular fluxes and $\mathbf{R}_m$ is the vector of the residuals of metabolite mass balances.

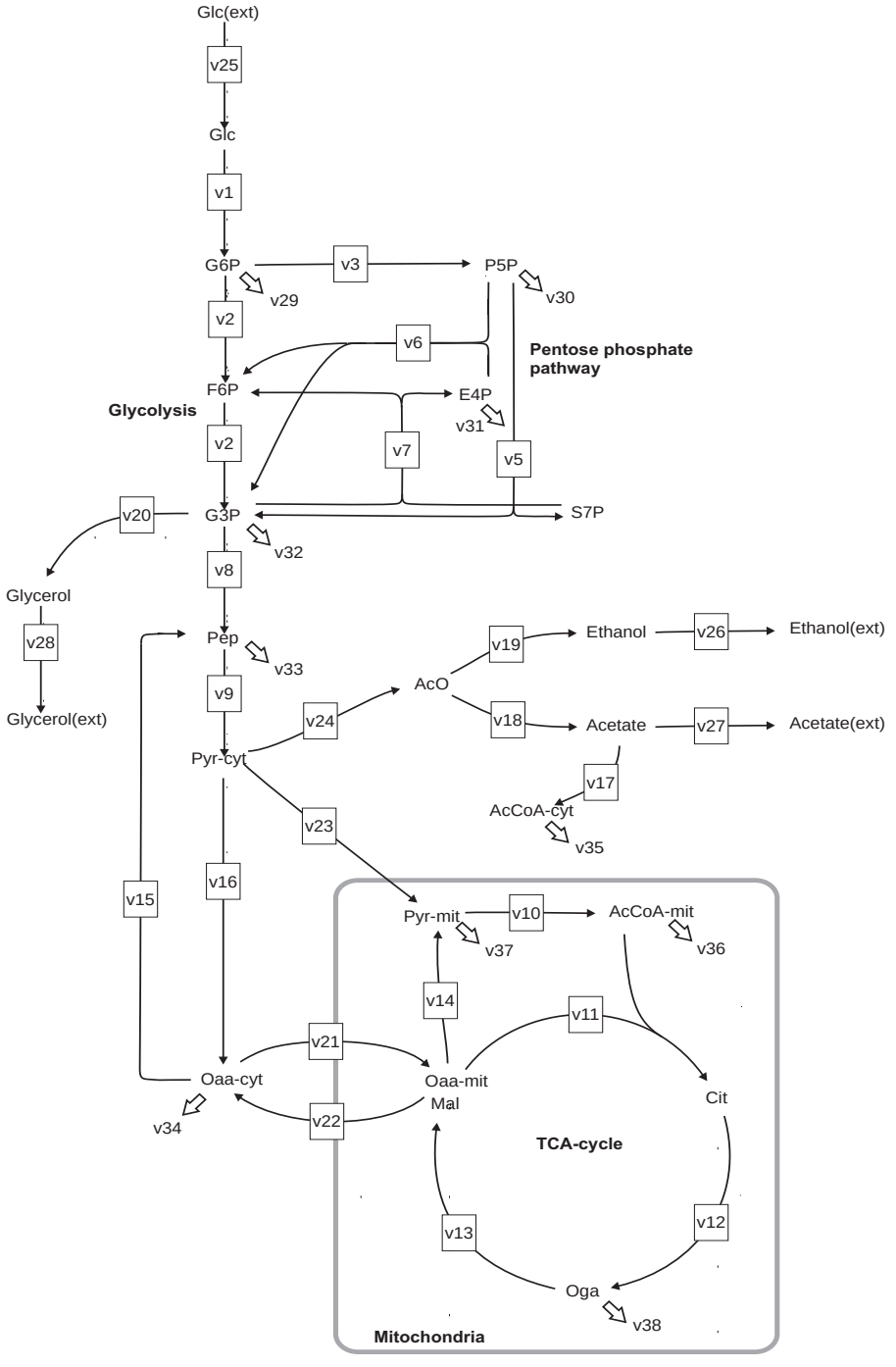

Figure 4 (Publication I). Metabolic network model of *S. cerevisiae* for net flux determination utilising flux ratios as additional constraints.

The flux ratio equations were set up according to the METAFoR analysis for the reactions in the stoichiometric models of the central carbon metabolism of *S. cerevisiae* (Equations 12 to 16, the reaction numbers are defined in Figure 4). Depending on the structure of the network of active reactions the flux ratio equations included one to six of the following (Equations 12 to 16):

the fraction of Pep from PPP assuming a maximal contribution of PPP

$$fr1 = \frac{v_5 + 3v_6 + 2v_7}{v_5 + 2v_4 + v_6} \tag{12}$$

the fraction of Pep originating from Oaa_{cyt}, X_{PEPck}:

$$fr2 = \frac{v_{15}}{v_{15} + v_8} \tag{13}$$

the fraction of Oaa_{mit} originating from Oaa_{cyt}, $X_{Oaa-transport}$:

$$fr3 = \frac{v_{21}}{v_{21} + v_{13}} \tag{14}$$

the fraction of Oaa_{cyt} originating from Pyr_{cyt}, $X_{Oaa_{cyt}\_from\_Pyr_{cyt}}$:

$$fr4 = \frac{v_{16}}{v_{16} + v_{22}} \tag{15}$$

the upper and lower bounds for Pyrmit originating from the malic enzyme reaction, X_{MAE_ub} and X_{MAE_lb}:

$$fr5 \le \frac{v_{14}}{v_{14} + v_9} \le fr6 \tag{16}$$

The following constraint equations were thus obtained from the flux ratio equations (Equations 17 to 22):

$$v_5 + 3v_6 + 2v_7 - fr1(v_5 + 2v_4 + v_6) = 0 \tag{17}$$

$$v_{15} - fr2(v_{15} + v_8) = 0 \tag{18}$$

$$v_{21} - fr3(v_{21} + v_{13}) = 0 \tag{19}$$

$$v_{16} - fr4(v_{16} + v_{22}) = 0 \quad (20)$$

$$v_{14} - fr5(v_{14} + v_9) = 0 \quad (21)$$

$$fr6(v_{14} + v_9) - v_{14} = 0 \quad (22)$$

Irreversibility was assumed for the intracellular fluxes v_3, v_4, v_8, v_9, v_{10}, x_{11}, v_{12}, v_{13}, v_{14}, v_{15}, v_{16}, v_{21}, v_{22}, v_{23}, v_{24}, for extracellular fluxes v_{25}, v_{26}, v_{27}, v_{28} and for the depletion of precursors to biosynthetic reactions and thus, only positive values were allowed for the fluxes. The minimization of the sum of the weighted square residuals of the metabolite mass balances was done using the Matlab function *fmincon* for optimization of constrained nonlinear multivariable function. The extracellular metabolite mass balances were assigned weights according to the experimental measurement error estimates. The biomass precursor metabolite mass balances were assigned ten-fold larger weights, relative to their stoichiometric coefficients in the biomass composition, since the assumption of constant biomass composition was expected to be harsh [Furukawa *et al.*, 1983]. The flux ratio constraints were included as strict constraints. The optimization was started 10000 times from random initial values to evaluate the uniqueness of the optimal flux distribution. The sensitivity of the flux distribution solutions to the noise in the flux ratio data and in the extracellular flux data was studied by Monte Carlo -simulations [Antoniewicz *et al.*, 2006]. Additive normally distributed noise with zero mean and experimentally determined variances of the flux ratios and the extracellular fluxes was sampled to the flux ratios and to the extracellular flux data, separately and simultaneously. A flux distribution was solved for each of the 100 sets of input data from 12 random initial flux distributions. Confidence intervals (95%) of the fluxes were determined.

4.8 Pathway reconstruction

In Publication III metabolic flux ratio (METAFoR) analysis was performed for *T. reesei* which lacks a genome-wide metabolic reconstruction. It was essential for the METAFoR analysis to reconstruct the biosynthetic pathways for proteinogenic amino acids of *T. reesei*. The reconstruction of amino acid biosynthetic pathways from their precursors in *T. reesei* was performed with ReTrace which is a recent computational carbon path analysis method [Pitkänen *et al.*, 2009], which can be queried to discover branching metabolic pathways in a uni-

versal metabolic database. ReTrace aims to find pathways which transfer as many atoms from source to target metabolites as possible.

The reaction database used in ReTrace analysis was KEGG LIGAND, retrieved in March 2009 [Kanehisa *et al.*, 2008]. Reaction database contained 7827 reactions, 15400 compounds. *Atom mappings*, that describe how atoms are transferred in a reaction from substrate to product metabolites, were defined for 33795 substrate-product pairs in the RPAIR database [Kotera *et al.*, 2004], which is a subdatabase of KEGG. All reactions were considered bidirectional. To compute reaction scores, a database consisting of 101136 sequences annotated with an EC number in UniProt version 9.3 [The UniProt Consortium, 2007] was queried with the 9129 protein sequences in *T. reesei* genome [Martinez *et al.*, 2008] by blastp [Altschul *et al.*, 1997] using e-value cutoff 10 to detect remote homologs. Each reaction in the KEGG database was assigned a score by taking the maximum BLAST score over all UniProt-*Trichoderma* sequence pairs, where UniProt sequence had been annotated with an EC number corresponding to the reaction. A total of 3974 reactions received a score in this procedure, while the remaining 3853 reactions were assigned a zero score. Reaction scores reflected the degree of evidence from the detection of sequence homology that there exists an enzyme catalyzing the reaction in *T. reesei*.

ReTrace operates on an atom-level graph representation of the metabolic network of all the reactions in the reaction database [Pitkänen *et al.*, 2009]. First, the metabolic network is converted into an atom graph, where nodes correspond to the atoms of metabolites and edges the atom mappings between the individual atoms. Atom mappings of carbon atoms were utilised in the reconstruction of the biosynthetic pathways of amino acids in *T. reesei*. Other atoms than carbons, such as nitrogen and sulphur, were not considered in the analysis performed in Publication III.

ReTrace [Pitkänen *et al.*, 2009] utilises a *K shortest paths* algorithm [Eppstein, 1994] to discover a number of connections between nodes in the atom graph. Given a query to find pathways from source to target metabolites (to amino acids in Publication III) ReTrace searched for shortest paths in the atom graph from any atom in source metabolites to any target metabolite atom [Pitkänen *et al.*, 2009]. Each shortest path was then processed. ReTrace traces back the target metabolite atoms along the atom mappings. Then branching points were identified as atom mappings that did not transfer the traced carbon atom towards the source [Pitkänen *et al.*, 2009]. Possible metabolites that could pro-

vide the missing carbon were determined and shortest paths from source to these metabolites were searched to determine the branches.

4.9 Localization of amino acid biosynthetic enzymes in *T. reesei*

In Publication III, TargetP, a machine learning method based on neural networks that predicts both chloroplast and mitochondrial targeting peptides and secretory signal peptides was utilized to predict the probable subcellular localization of some enzymes on the biosynthetic pathways of amino acids in *T. reesei* [Emanuelsson *et al.*, 2007; Emanuelsson *et al.*, 2000].

5. Results and discussion

In Publications I, II, III and IV ratios of intracellular fluxes were determined utilising ^{13}C-labelling experiments and the established metabolic flux ratio (METAFoR) analysis approach was extended. In Publication I flux ratios from METAFoR analysis combined with NMR spectroscopy as the analytical tool was utilized as experimentally derived additional constraints that enabled solving intracellular net fluxes under *S. cerevisiae* in different oxygenation conditions and determination of the dependence of the flux phenotype of *S. cerevisiae* on oxygen provision. In Publication II the Established METAFoR analysis was extended to a two-carbon source case to investigate the metabolism of *P. pastoris* under process conditions of recombinant protein production where methanol is used as inducer and glycerol as a carbon source. In Publication III the previously scarcely studied metabolism of filamentous fungus *T. reesei* was studied under different conditions of carbon catabolite repression. Since *T. reesei* lacks a genome-wide metabolic reconstruction the METAFoR analysis was coupled to essential reconstruction of the biosynthetic pathways of amino acids from genome level evidence. Comparison of the flux ratios of *T. reesei* to the ones observed in *S. cerevisiae* confirmed that the regulation of the central pathway fluxes is programmed in distinct ways in the two organisms. In Publication IV a systematic and analytic framework for derivation of flux ratio equations from a given model and ^{13}C-labelling data that constrains isotopomer distributions was developed. Comparison of the automatically derived flux ratios to manually by METAFoR analysis approach derived flux ratios was not straightforward because the biological information encoded in the METAFoR analysis approach by domain experts is diverse. When the model employed in the automatic derivation corresponded to the METAFoR approach, the derived flux ratios agreed well.

In Publication V NMR spectroscopy was utilized in verification of the reconstruction of a metabolic pathway in *T. reesei.* A reaction product of a step on the D-galacturonic acid catabolic pathway was identified by 1D and 2D NMR spectroscopy.

5.1 Utilization of ^{13}C-metabolic flux analysis excluded cofactor mass balances

In Publication I the dependence of intracellular flux distribution of *S. cerevisiae* on the level oxygenation was quantified by ^{13}C-MFA. *S. cerevisiae* was grown in glucose-limited chemostat cultures at a low dilution rate of 0.1 h^{-1} that ensured that the maximum respiratory rate was not reached even under fully aerobic conditions. The cultures were aerated with five different fractions of oxygen in the chemostat inlet gas: 20.9%, 2.8%, 1.0%, 0.5% and 0.0%. Wiebe *et al.* (2008) observed that at aeration of 20.9% O_2 in the chemostat inlet gas the metabolism of *S. cerevisiae* was fully respiratory and ethanol was observed in the medium under all the rest of the conditions. Duplicate cultures at each oxygenation condition were ^{13}C-labelled with uniform-labelling approach for ^{13}C-MFA analysis in Publication I. Quantitative ratios of merging fluxes in central carbon metabolism of *S. cerevisiae* were obtained by METAFoR analysis approach utilizing NMR spectroscopic detection of ^{13}C-labelling [Szyperski, 1995; Szyperski *et al.*, 1999; Maaheimo *et al.*, 2001]. The flux ratios were utilised as additional constraints to solve the mass balance equation system of the stoichiometric model of central carbon metabolism of *S. cerevisiae* as was earlier done for another eukaryote *P. anomala* with MS detected constraints by Fredlund *et al.* (2004) with the approach published by Fischer *et al.* (2004). ^{13}C-MFA enabled solving the intracellular net fluxes without including the redox cofactors in the mass balances which was essential since oxygen availability strongly affects the systems that maintain the redox homeostasis in cells.

5.2 Pyruvate branching point distribution responsive

The quantification of the net flux distributions of *S. cerevisiae* in response to different oxygenation conditions showed that the fluxes were redistributed not only between the cells grown in the fully aerobic conditions, under conditions of reduced oxygen provision and under anaerobic conditions but also between cells grown with different levels of low oxygen (2.8%, 1.0% and 0.5% O_2 in the

chemostat inlet gas) (Publication I). Although the metabolism of *S. cerevisiae* was respiro-fermentative under each of these low oxygen conditions, the actual amount of oxygen available resulted in different distribution to the respirative and fermentative pathways. The flux distribution at the pyruvate branch point, where the respirative and the fermentative pathways and the anaplerotic pathway, that operates to replenish the TCA cycle, diverge was particularly responsive to the level of reduction in oxygen provision. The respirative pathway flux decreased progressively under reduced oxygenation conditions where the availability of terminal electron acceptor limited the respiratory rate. However, the respiratory energy generation, that is highly efficient because of the high electronegativity of oxygen, provided a large fraction of ATP even under the low oxygen conditions (Table 1).

Table 1 (Publication I). Estimated fractions of respiration coupled ATP generation in *S. cerevisiae* under different oxygenation conditions.

	O_2 provided in the fermentor inlet gas				
	20.9%	2.8%	1.0%	0.5%	0.0%
ATP from respiration (%)	59	55	36	25	0
OUR (mmol $g^{-1}h^{-1}$)	2.7	2.5	1.7	1.2	0
ATP/ O[a]	0.9	1.0	1.1	1.1	-
ATP/ $2e^{-}$[b]	1.0	0.9	1.0	1.2	-

[a] Calculated from the oxygen uptake rate (OUR), [b]Calculated from the flux of electron donors to the respiratory chain.

5.3 Methanol and glycerol co-utilization extension

In Publication II a biosynthetically directed fractional ^{13}C-labelling approach was established for yeast *P. pastoris* growing on carbon substrate mixture of methanol and glycerol. Methanol is utilised as an inducer of protein production in *P. pastoris* processes. The approach allowed the quantification of the metabolic state of the TCA cycle and associated pathways under production conditions and thus was an important methodological expansion of the metabolic flux ratio (METAFoR) analysis [Szyperski, 1995; Szyperski *et al.*, 1999; Maaheimo *et al.*, 2001].

5.4 Flux distributions robust against different fractions of methanol

It was shown that co-assimilation of methanol as a carbon source does not alter the way the common amino acids are synthesized in *P. pastoris* growing on a sole multicarbon source, and that the growth on different glycerol/methanol mixtures at a given growth rate results in rather similar flux ratio profiles in the TCA cycle and related pathways as the fraction of methanol is increased (Publication II). In contrast, a clear effect of specific growth rate on the relative activity of the TCA cycle and related pathways was observed, regardless of the methanol fraction in the feed, consistent with the observation that TCA cycle activity in *S. cerevisiae* is strongly correlated with the environmentally determined specific growth rate [Blank and Sauer, 2004]. Co-assimilation of methanol as a carbon source has a clear impact with respect to the activity of the PPP, which is consistent with the increasing flux of methanol molecules towards the synthesis of central carbon metabolism intermediates (e.g. Pep), as observed when the fraction of methanol in the feed medium is increased. However, this pattern was not observed in cells growing at the higher dilution rate (where methanol is partially accumulated in the medium) suggesting that the distribution of methanol carbon into assimilatory and dissimilatory pathways may be different. Earlier ^{13}C-labelling studies of methanol metabolism of the methylotrophic yeast *H. polymorpha* [Jones and Bellion, 1991] showed that the linear methanol oxidation pathway to CO_2 only operates under extreme conditions (e.g. methanol accumulation to toxic levels), suggesting a role in detoxification.

The information from the ^{13}C-labelling and METAFoR analysis [Szyperski, 1995; Szyperski *et al.*, 1999; Maaheimo *et al.*, 2001] of *P. pastoris* on glycerol and methanol mixtures is valuable for the optimization of culture processes for the production of recombinant proteins in *P. pastoris*, where parameters such as the residual methanol concentration, specific growth rate, as well as mixed substrate culture strategies have been shown to have a dramatic impact on overall process productivity (Publication II). In addition, the information derived from the study may be relevant for the design of isotopic labelling experiments of recombinant proteins (or other cell components, e.g. cell wall glucans) for structural studies.

5.5 Metabolic flux ratio analysis of *T. reesei* necessitated reconstruction of biosynthetic pathways of amino acids

In Publication III the biosynthetic pathways of *T. reesei* were reconstructed for most of the proteinogenic amino acids with a computational carbon path analysis method ReTrace [Pitkänen *et al.*, 2009]. The method was used to search for pathways from a metabolic network consisting of all reactions found in a comprehensive metabolic reaction database, and to subsequently rank the pathways according to the degree of support from the *T. reesei*'s genome. Contiguous pathways, identical to the amino acid biosynthetic routes of *S. cerevisiae*, were found with high genome-level evidence. Origins of amino acids in *T. reesei* relevant for METAFoR analysis are shown in Figure 5.

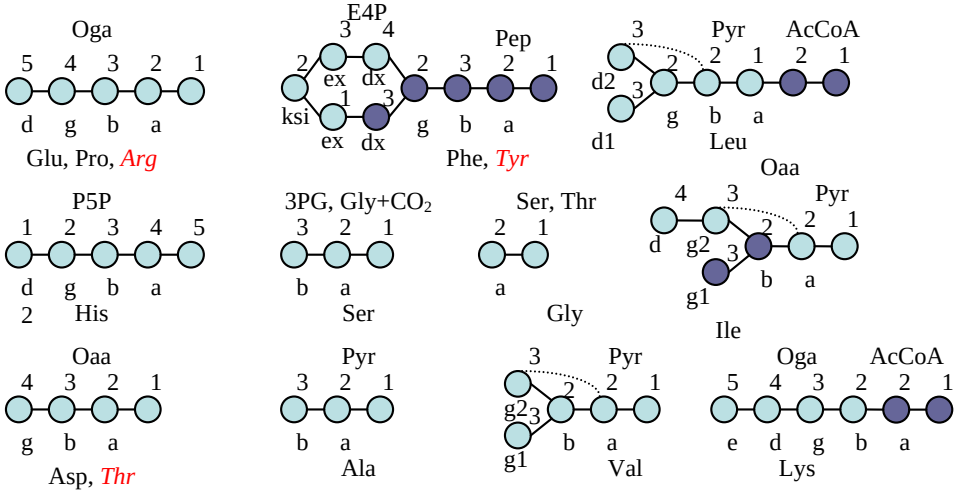

Figure 5 (Publication III). The origins of the carbon backbones of the proteinogenic amino acids in *T. reesei* that are relevant for METAFoR analysis and for which the biosynthetic pathways were reconstructed by computational pathway analysis method ReTrace [Pitkänen *et al.*, 2009]. If the biosynthetic pathway was not directly found by ReTrace, the amino acid abbreviation is denoted in red italics. The amino acid carbons are denoted in the following way: a = α, b = β, g = γ, d = δ, e = ε, ksi = ξ.

5.6 Primary respiratory metabolism

T. reesei wild type and Δ*cre1* strains were ^{13}C-labelled with uniform ^{13}C-labelling approach in batch cultures in flasks on repressive carbon source glucose and on sorbitol. The ^{13}C-labelling patterns of proteinogenic amino acids were in good

accordance with the compartmentalized model of eukaryotic central carbon metabolism, originally developed for *S. cerevisiae* [Maaheimo *et al.*, 2001]]. However, in contrast to *S. cerevisiae*, in both *T. reesei* strains Asp synthesis was observed to occur primarily from the mitochondrial pool of Oaa, under all the studied conditions.

The *T. reesei* wild type strain is known to exhibit carbon catabolite repression of hydrolytic gene expression during growth on glucose, whereas in the Δ*cre1* strain the repression is partially disturbed [Ilmén *et al.*, 1996]. The respiratory pathway does not become transcriptionally downregulated by the carbon catabolite repression in *T. reesei* as in *S. cerevisiae* [Chambergo *et al.*, 2002]. However, it is the *in vivo* fluxes that are the ultimate phenotype of an organism. The ratios of *in vivo* fluxes of *T. reesei* were solved with METAFoR analysis from the ^{13}C-labelling patterns of proteinogenic amino acids [Szyperski, 1995; Szyperski *et al.*, 1999; Maaheimo *et al.*, 2001] in the different conditions of carbon catabolite repression. This was the first time that the effect of carbon catabolite repression *T. reesei* on *in vivo* fluxes was quantitatively studied.

The relative anaplerotic flux, the flux that replenishes the TCA cycle, compared to the respiratory pathway flux was characteristic to primarily respiratory metabolism in the both *T. reesei* strains under all the studied conditions (Table 2). This indicated that *T. reesei* utilizes primarily respiratory metabolism also on preferred carbon source glucose. However, the observed relative anaplerotic fluxes suggested that the respirative activity of the TCA cycle in *T. reesei* was even slightly higher on the neutral carbon source sorbitol than on glucose. Only minor differences were observed between the *in vivo* flux distributions of the wild type and the Δ*cre1* strains. Therefore, CRE1 the key repressor of utilization of alternative carbon sources, does not mediate carbon source dependent metabolic state alterations in central carbon metabolism in *T. reesei*. The sophorose induction of cellulase gene expression did not result in significant changes in the relative requirements of proteinogenic amino acids or in the ratio of anabolic to oxidative activities of the TCA cycle.

Table 2 (Publication III). Comparison of Metabolic Flux Ratio (METAFoR) analysis of *T. reesei* to the crabtree positive and negative yeasts *S. cerevisiae* and *P. stipitis*. METAFoR analysis from fractional ^{13}C-labelling of *T. reesei* wild type and *Δcre1* strains in aerobic batch cultures on glucose and on sorbitol (Publication III), compared with the ones observed in *S. cerevisiae* and *P. stipitis* in aerobic batch cultures on glucose and in aerobic glucose-limited chemostat cultures [Maaheimo *et al.*, 2001; Fiaux *et al.*, 2003].

organism	*T. reesei*		*T. reesei*		*T. reesei*		*T. reesei*		*S. cerevisiae*		*S. cerevisiae*		*P.stipitis*		*P.stipitis*	
strain	wild type		delta *cre1*		wild type		delta *cre1*									
carbon source	glucose		glucose		sorbitol		sorbitol		glucose		glucose		glucose		glucose	
culture	batch		batch		batch		batch		batch		chemostat		batch		chemostat	
reference									Maaheimo		Fiaux		Fiaux		Fiaux	
	%	sd	%	sd	%	sd	%	sd	%	sd	%	sd	%	sd	%	sd
Pep from PPP (UB, no PEPck)	39	2	47	4	36	7	45	9	0-4		40	8	57	9	61	11
R5P from T3P and S7P	51	1	42	1	72	3	79	4	68	2	59	2	57	2	72	2
R5P from E4P	25	2	23	1	46	2	54	3	10	2	33	2	35	2	43	2
Anaplerotic flux ratio	35	1	33	2	26	3	42	5	76	4	31	2	36	2	32	2
MAE (UB)	4	0	9	1	12	2	6	5	25-30		<13		<6		<7	
MAE (LB)	2	0	6	1	9	1	4	3	nd	nd	nd	nd	nd	nd	nd	nd

UB upper bound, LB lower bound, nd not determined

5.7 Framework for analytic and systematic derivation of flux ratio equations

In Publication VI a systematic and analytic framework for ^{13}C-metabolic flux ratio analysis was introduced. Previously the utilisation of the METAFoR analysis [Szyperski, 1995; Szyperski *et al.*, 1999; Maaheimo *et al.*, 2001] has been relying on manual derivation of the equations that constrain the flux ratios. The systematic and analytic framework for ^{13}C-metabolic flux ratio analysis is a generalization and formalization of existing analytic methods for computing metabolic flux ratios [Maaheimo *et al.*, 2001; Zamboni *et al.*, 2005; Szyperski 1998] and facilitates an efficient and analytic computation of the ratios between the fluxes producing the same junction metabolite in a given metabolic network.

The model of the central carbon metabolism of *S. cerevisiae* was formulated also in a ^{13}C-FLUX [Wiechert *et al.*, 2001] format and artificial substrate labelling was employed to obtain simulated ^{13}C-labelling data for verification of the implementation of the framework for flux ratio determination. Then NMR spectroscopy data from ^{1}H-^{13}C HSQC experiments, relative intensities of fine structures that represent different combinations of ^{13}C and ^{12}C atoms coupled to a central ^{13}C atom in proteinogenic amino acids, was utilized to compare the flux ratios derived with the implemented framework with manually derived flux ratios. The experiment showed that the framework was able to provide relevant quantitative information on the distribution of metabolic fluxes, even when only constraints to the isotopomer distributions of proteinogenic amino acids are

measured. Detection of the ^{13}C-labelling status of 15 proteinogenic amino acids resulted in flux ratios for junctions in four cytosolic metabolites (Oaa, Pep, Gly and Ser) and in three mitochondrial metabolites (Oaa, AcCoA and Pyr). In addition, an upper bound for G3P molecules that had gone through a transketolase reaction could be solved when the model was first manually simplified to correspond to the model utilised in the manual derivation of the G3P flux ratio. The computed flux ratios where compared to the manually derived ones whenever the assumptions made in the manual derivation were consistent with the general model employed in the automatic derivation of the flux ratios. The automatically derived flux ratios agreed well with the manually derived ratios. Differences between the estimations could be explained by numerical instabilities and by differences in computational procedures: in manually derived ratios the estimations are based on the breakage of a single bond in different routes leading to a metabolite while in the developed framework more isotopomer information is possibly utilized in the estimation.

5.8 L-threo-3-deoxyhexulosonate is a reaction product of L-galactonate dehydratase

In Publication V the metabolism of fungus *T. reesei* was further studied and an L-galactonate dehydratase and the corresponding gene were identified from *T. reesei*. The enzyme converts L-galactonate to L-threo-3-deoxy-hexulosonate (2-keto-3-deoxy- L-galactonate) and belongs to the fungal pathway for D-galacturonic acid catabolism. L-galactonate dehydratase is the second enzyme of the pathway after the D-galacturonic acid reductase. L-galactonate dehydratase activity is present in *T. reesei* mycelia grown on D-galacturonic acid but absent when other carbon sources are used for growth. L-galactonate dehydratase is active on sugars L-galactonate and D-arabonate in which the hydroxyl groups of the C2 and the C3 in the Fischer projection are in L- and D-configuration, respectively. The enzyme was not active with sugar acids having the hydroxyl groups of C2 in D-configuration and C3 in L-configuration as in D-galactonate, D-gluconate and D-xylonate and with sugar acids having the hydroxyl groups of C2 and C3 in D-configuration as in D-gulonate.

In order to define the reaction product of L-galactonate dehydratase it was analysed by NMR spectroscopy. To generate a sufficient amount of reaction product L-galactonate was incubated in the yeast extract of the strain expressing the L-galactonate dehydratase gene. In this extract the reaction product did not

react further, which facilitated the NMR spectroscopic analysis. In the *T. reesei* mycelia extract the reaction product was degraded, making the NMR spectroscopic analysis more difficult. The NMR spectroscopic analysis showed that erythro- or threo-3-deoxy-hexulosonate was formed. Knowing the substrate of the dehydratase reaction it was concluded that the product was L-threo-3-deoxyhexulosonate. The NMR spectroscopic analysis also revealed that it was predominantly in the pyranose form. For the pyranose form two anomers are possible; the carboxyl group in R_1 and the hydroxyl group in R_2 or vice versa. The NMR spectroscopic analysis suggested that one of the anomers was predominant but it did not allow determination of which of the two anomers it was. The NMR spectroscopic analysis revealed also that the axial hydrogen at C3 is the hydrogen that was added in the reaction. However, as there are two possible chair conformations of the pyranose ring it remained unclear which of the two protons is in the axial position.

6. Conclusions and prospects

The analyses of flux ratio of different organisms revealed phenotypic differences in detail. In the studies of this thesis the differences between the regulatory principles of two eukaryotic organisms, *S. cerevisiae* and *T. reesei*, were observed on flux phenotypic level (Publications I and III). The former pursues high rate of ATP production whereas the latter seeks a high ATP yield. The distributions of fluxes to respirative and fermentative pathways were similar when *S. cerevisiae* was growing slowly with unlimited oxygenation and low glucose and *T. reesei* was growing on high glucose at maximum rate. However, the fast growing *S. cerevisiae* on high glucose diminishes the respirative pathway flux and speeds up the glycolytic flux and fermentative pathway activity [De Deken, 1966; Maaheimo *et al.* 2001; Gombert *et al.* 2001]. The differences in regulatory principles of the two organisms can be explained by the different natural habitats of the organisms and adaptive evolution of the regulatory systems. Pfeiffer *et al.* (2001) have further claimed that the preference to high ATP yield has contributed to the development of multicellular organisms.

6.1 Robust regulatory system enables stable flux phenotype

Complex and multi-level regulatory mechanisms can maintain fairly stable distribution of fluxes in altered conditions. However, attenuating changes can though be observed in the underlying levels of transcription, proteome and metabolome [Davies and Brindle, 1992; Schaaff *et al.*, 2004; Raamsdonk *et al.*, 2001]. Furthermore, an adaptation of the flux phenotype to altered conditions can occur through sequential changes in the underlying levels of cell function [van der Brink *et al.*, 2008]. There is interdependence between metabolite concentrations, enzyme concentrations and fluxes through the metabolic network,

enzyme kinetics and regulation of enzyme activity [Stryer, 1995]. Enzyme concentration is set by hierarchical regulatory system that is closely linked to the metabolic status of an organism. Nutrient sensing and feedback regulation trigger signalling cascades that affect the hierarchical regulatory system [Zaman *et al.*, 2008; Zaman *et al.*, 2009]. Coordinated responses have been observed even between metabolites and transcripts [Bradley *et al.*, 2009]. The enzyme activity cannot affect the equilibrium constant of a chemical reaction but the network can enable a shift of an effective equilibrium constant through futile cycles [Qian and Beard, 2006]. The multilevel regulatory system provides robustness that enables stability of flux phenotypes. Are the robust response mechanisms to different perturbations mechanistically similar? How has the robustness of an organism against fluctuations in its natural environment developed? The same multilevel regulatory system provides fine-tuned adaptation in some conditions. A fine piece of work was published by Bennet and co-workers (2008) where they could conclude that the regulatory system of galactose metabolism in *S. cerevisiae* functions as a low-pass filter that in dynamic conditions enables adaptation to slow changes and robustness against fast perturbations. Thus the same mechanisms can provide both robustness and adaptation. Are they different depending on the perturbation? How are the decisions between robust response and adaptation made when the cells are exposed to different perturbations? These are interesting questions that are expected to be answered by sophisticated systems biology studies when the experimental and modelling tools are reaching an adequate performance level.

It should eventually be possible to integrate data of all measurable cell constituents and integrate it with models of cellular interaction networks and compose a predictive model of cell function. There are methods for recording data on all the effector types. Quantitative metabolomics methods in particular by MS have been established [Ewald *et al.*, 2009; Buscher *et al.*, 2009]. Enzyme concentrations are determined by rates of transcription, mRNA degradation, translation and protein degradation and both protein and transcript levels can be observed in large-scale [de Groot *et al.*, 2007]. Regulatory mechanisms that affect fluxes are for example enzyme phosphorylations which can currently be monitored [Ptacek *et al.*, 2005; Huber *et al.*, 2009]. Information on allosteric interactions has been reported in literature and some is collected for example in database BRENDA (http://www.brenda-enzymes.org) [Schomburg *et al.*, 2002]. All effectors are tied together by interaction networks that are dynamic in reality. However, the static interaction network models provide a framework for the

dynamic system responses that give rise to the observed flux phenotypes. Hindrances can still be pinpointed both in experimental methods and in modelling tools. Not all the levels of cell function can be monitored in genome-wide scale and modelling tools need to be able to handle that missing data and uncertainties as well as to find suitable means to model different types of interactions and information transfer through them. Ishii *et al.* (2007) performed a pioneering data integration study in a limited system were flux phenotypic data and data on underlying regulatory levels was simultaneously recorded from central carbon metabolism of single gene deletion mutant strains of *E. coli* under different growth conditions. However, since the scale of the system was limited to the central carbon metabolism, integrative visualisation of the metabolic network was adequate for interpretation of the data. Figure 6 shows an example of integrative data analysis of the switch from fully respiratory to respiro-fermentative phenotype of *S. cerevisiae* in context of interaction network of limited extent (unpublished results). The systemic co-responses are almost too complex to visually comprehend even though the system is limited. Multiomics data integration requires tools that generate hypotheses of systemic response mechanisms.

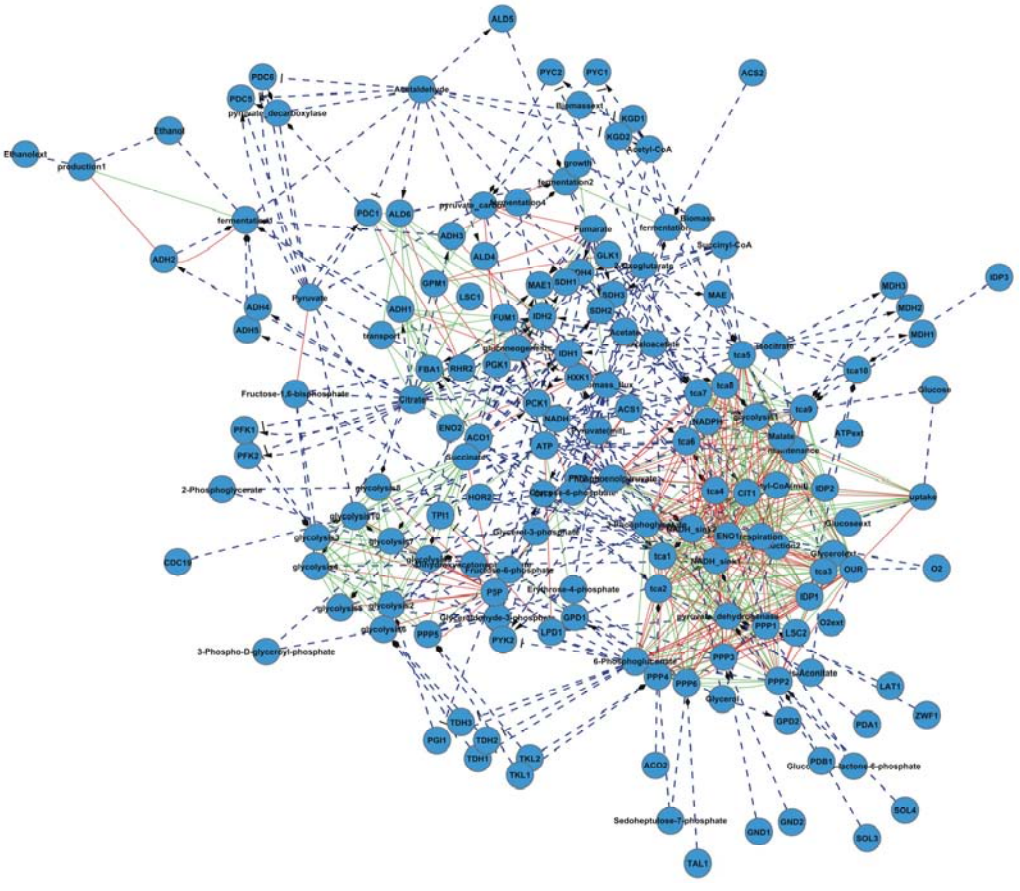

Figure 6. Co-response interactions in *S. cerevisiae* between components of central carbon metabolism in switch from fully respirative to respiro-fermentative metabolism in limited oxygen provision (unpublished results). Co-responses are calculated according to Raamsdonk *et al.* (2001) as ratios of the log fold changes in the quantities of components between the two conditions [Publication I; Rintala *et al.*, 2009]. Co-response interactions are drawn for the components having strong positive (in green) or negative (in red) co-responses.

6.2 Determinants of energy generation processes

The biological features studied in the Publications included in the thesis were effects of oxygen and different carbon sources on flux distributions of three eukaryotes *S. cerevisiae*, *T. reesei* and *P. pastoris*. Carbon sources among other nutrients [Rohde *et al.*, 2008] are known to trigger regulatory cascades that can ultimately determine the active energy production processes as in case of high

glucose sensed by *S. cerevisiae* [Zaman *et al.*, 2009]. Oxygen on the other hand is a major determinant of active energy processes in aerobic organisms. Regulation of energy processes in bioproduction hosts is naturally of importance for developed bioprocess control. Many of the regulatory mechanisms are conserved among eukaryotes and thus, the studies of simple eukaryotes such as yeasts and fungi provide a basis also for understanding the function of higher cells. Understanding of the regulation of energy generation processes is central in investigations of for example human metabolic diseases. One of the metabolic diseases, type II diabetes, is at present increasing in population of western countries [Wild *et al.*, 2004]. Energy generation processes are central also in sports. Oxygen uptake rate and carbohydrate refuelling are central parameters of sports performance [Hulston and Jeukendrup, 2009]. Oxygen provides not only an electron acceptor for efficient energy generation but also a threat of cell component damage as a strong oxidant. Both features of oxygen are related to cell ageing and thus, of major interest [Koc *et al.*, 2004; Oliveira *et al.*, 2008; Finkel and Holbrook, 2000; Lin *et al.*, 2002].

6.3 Prospects of local flux ratio analysis

Local flux ratio analysis is an efficient approach for quantitative profiling of the flux phenotype. The conventional method is extendable from inferring ^{13}C-labelling patterns of proteinogenic amino acids to detection of free metabolites and from established network structures to novel or engineered organisms by systematic derivation of constraint equations (Publication IV). The strength and advantage of the local flux ratio analysis is the independence of external fluxes and therefore, also an independence of the definition of biomass effluxes [Zamboni *et al.*, 2009]. Biomass effluxes are usually derived from growth rate and biomass composition and the experimental determination of the detailed macromolecular composition of biomass is laborious [Lange 2002; Lange and Heijnen, 2001]. Thus, flux ratios provide independent quantitative measures of distribution of fluxes also when the effluxes from the system to macromolecule synthesis is not exactly known. Analysis can be targeted to metabolic junctions of interest or as many junctions can be analysed as possible to gain constraints for net flux determination by ^{13}C-MFA.

6.4 Large-scale flux analysis

Global iterative fitting becomes computationally unfeasible when the number of isotopomer balance equations explodes. System extension requires also additional measurements of ^{13}C-labelling patterns to gain constraints for isotopomer distributions of added metabolites. Without enough constraints the flux distribution cannot be accurately solved. Combinations of ^{13}C-labelling based constraints and an objective function have been proposed for example by Blank *et al.* (2005). Nevertheless, the large-scale models can always be utilised as scaffolds where the active networks can be identified and more detailed models built [Suthers *et al.*, 2007].

Flux balance analysis (FBA) is feasible in well-defined genome scale metabolic networks. The difficulty of definition of a relevant objective function limits the utilisation of FBA in flux determination. In determination of metabolic capabilities of organisms it is highly efficient. Automatic means to infer or to identify objective functions have been proposed [Knorr *et al.* 2007; Gianchandani *et al.* 2008]. As interest in systems biology is also on higher cells, intelligent definitions of objective functions have been proposed [Heuett and Qian, 2006] and are under development.

6.5 Flux analysis in dynamic conditions

Flux analysis is turning dynamic since stationary metabolic flux analysis is not convenient for those time-dependent processes that are of biotechnological interest or compatible with analysis of fluxes in higher cells. If the specific growth rate is low, prolonged labelling time is required to reach an isotopic pseudo steady-state in macromolecule components. During prolonged labelling time the steady state may alter. Furthermore, the non-growing cells are completely incompatible with stationary biosynthetically directed ^{13}C-labelling. Direct measurement of the labelling patterns of metabolic intermediates was thought to shorten the required labelling times but the investigations showed that the labelling patterns of intermediates stabilize much later than expected. Large pools of storage compounds, protein turnover and compartmentalized pools increase the stabilisation times of the labelling patterns of metabolic intermediates to the timescale of label stabilisation in macromolecule components [Aboka *et al.*, 2009; Grotkjaer *et al.*, 2004; den Hollander *et al.*, 1981].

Dynamic ^{13}C-MFA that significantly shortens the required labelling times emerged first to metabolic stationary but isotopically non-stationary conditions [Nöh *et al.*, 2007; Selivanov *et al.*, 2006; Schaub *et al.*, 2008; Hoffmann *et al.*, 2008]. Due to the shorter labelling times, the dynamic ^{13}C-flux analysis has broader applicability to mammalian cell cultures than the conventional stationary flux analysis. Mammalian cells grow usually slower than microbial cells and steady states are hard to sustain. Recently Munger *et al.* (2008) utilised dynamic ^{13}C-labelling of cultured human fibroblasts to determine metabolic targets for antiviral therapy. Extension of ^{13}C-MFA to metabolic non-stationary states was shown *in silico* by Wahl *et al.* (2008) and applied already to *E. coli* cultures by Schaub *et al.* (2008), and to hepatic cells in a two-part study [Hoffmann *et al.*, 2008; Maier *et al.*, 2008]. ^{13}C-MFA in transient state takes advantage of a kinetic model of the metabolic system, measurements of metabolic pool sizes and time-dependent labelling patterns. Integration of data yields different things depending on the approach. The kinetic parameters can be determined more reliably, non-measured metabolite pools can be solved and sensitivities of flux distribution solutions can be decreased by integrating stationary and non-stationary data [Nöh *et al.*, 2007; Selivanov *et al.*, 2006; Schaub *et al.*, 2008]. Dynamic ^{13}C-flux analysis is also suitable for stimulus-response experiments were the systemic time-dependent responses to perturbations are investigated [Wahl *et al.*, 2008]. ^{13}C-labelling experiments in dynamic conditions enable direct probing of robustness and control of the metabolic system and the data also reveals compartmentalization of metabolite pools [Schryer *et al.*, 2009]. Furthermore, the short time-scale ^{13}C-labelling reduces the cost of experiments and thus, is compatible with high-throughput experiments.

References

Aboka FO, Heijnen JJ, van Winden WA: Dynamic ^{13}C-tracer study of storage carbohydrate pools in aerobic glucose-limited *Saccharomyces cerevisiae* confirms a rapid steady-state turnover and fast mobilization during a modest stepup in the glucose uptake rate. *FEMS Yeast Res.* 2009, 9:191–201.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997, 25:3389–3402.

Antoniewicz MR, Kelleher JK, Stephanopoulos G: Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metab Eng.* 2006, 8:324–337.

Antoniewicz MR, Kelleher JK, Stephanopoulos G: Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions. *Metab Eng.* 2007, 9:68–86.

Araúzo-Bravo M, Shimizu K: An improved method for statistical analysis of metabolic flux analysis using isotomomer mapping matrices with analytical expressions. *J Biotechnol.* 2003, 105:117–133.

Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, Herschlaq D: Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *PNAS*, 2003, 100:3889–3894.

Arita M: In Silico Atomic Tracing by Substrate-Product Relationships in *Escherichia coli* Intermediary Metabolism. *Genome Res.* 2003, 13:2455–2466.

Aro N, Pakula T, Penttilä M: Transcriptional regulation of plant cell wall degradation by filamentous fungi. *FEMS Microbiol Rev.* 2005, 29:719–735.

Bakker BM, Overkamp KM, van Maris AJA, Kötter P, Luttik MAH, van Dijken JP, Pronk JT: Stoichiometry and compartmentation of NADH metabolism in *Saccharomyces cerevisiae*. *FEMS Microbiol Rev.* 2001, 25:15–37.

Beard DA, Babson E, Curtis E, Qian H: Thermodynamic constraints for biochemical networks. *J Theor Biol*. 2004, 228:327–333.

Beard DA, Liang SD, Qian H: Energy balance for analysis of complex metabolic networks. *Biophys J.* 2002, **83**:79–86.

Beard DA, Qian H: Thermodynamics-based computational profiling of cellular regulatory control in hepatocyte metabolism. *Am J Physiol Endocrino Metab.* 2005, 288:E633–E644.

Becker SA, Palsson BO: Context-specific metabolic networks are consistent with experiments. *PLOS Comput Biol.* 2008, 4: e1000082.

Bennet MR, Lee Pang W, Ostroff NA, Baumgartner BL, Nayak S, Tsimring LS, Hasty J: Metabolic gene regulation in a dynamically changing environment. *Nature* 2008, 454:1119–1122.

Bernhard SA, Tompa P: The mechanism of succinate or fumarate transfer in the tricarboxylic acid cycle allows molecular rotation of the intermediate. *Arch Biochem Biophys*. 1990, 276:191–198.

Blank LM, Kuepfer L, Sauer U: Large-scale ^{13}C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol.* 2005, 6:R49.

Blank LM, Sauer U: TCA cycle activity in *Saccharomyces cerevisiae* is a function of the environmentally determined specific growth and glucose uptake rates. *Microbiology* 2004, 150:1085–1093.

Boles E, de Jong-Gubbels P, Pronk JT: Identification and Characterization of *MAE1*, the *Saccharomyces cerevisiae* Structural Gene Encoding Mitochondrial Malic Enzyme. *J Bacteriol.* 1998, 180:2875–2882.

Botstein D, Chervitz SA, Cherry JM: Yeast as a model organism. *Science* 1997, 277:1259–1260.

Bradley PH, Brauer MJ, Rabinowitz JD, Troyanskaya OG: Coordinated concentration changes of transcripts and metabolites in *Saccharomyces cerevisiae*. *PLOS Comput Biol*. 2009, 5:e1000270.

Brockmann R, Beyer A, Heinisch JJ, Wilhelm T: Posttranscriptional expression regulation: what determines translation rates? *PLoS Comput Biol.* 2007, 3:e57.

Brosnan JT, Brosnan ME, Nissim I: A window into cellular metabolism: hepatic metabolism of ^{15}N-labelled substrates. *Metab Eng.* 2004, 6:6–11.

Burgard AP, Maranas CD: Optimization-Based Framework for Inferring and Testing Hypothesized Metabolic Objective Functions. *Biotechnol Bioeng*. 2003, 82:670–677.

Burgard AP, Pharkya P, Maranas CD: OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng*. 2003, 85:1–11.

Buscher JM, Czernik D, Ewald JC, Sauer U, Zamboni N: Cross-platform comparison of methods for quantitative metabolomics of primary metabolism. *Anal Chem.* 2009, 81:2135–2143.

Butow RA, Avadhani NG: Mitochondrial signalling: the retrograde response. *Mol Cell* 2004, 14:1–15.

Castello PR, David PS, McClure T, Crook Z, Poyton RO: Mitochondrial cytochrome oxidase produces nitric oxide under hypoxic conditions: Implications for oxygen sensing and hypoxic signalling in eukaryotes. *Cell Metab*. 2006, 3:277–287.

Chambergo FS, Bonaccorsi ED, Ferreira AJS, Ramos ASP, Ribamar Ferreira Júnior J, Abrahao-Neto J, Farah JPS, El-Dorry H: Elucidation of the Metabolic Fate of Glucose in the Filamentous Fungus *Trichoderma reesei* Using Expressed Sequence Tag (EST) Analysis and cDNA Microarrays. *J Biol Chem.* 2002, 277:13983–13988.

Chen RE, Thorner J: Function and regulation in MAPK signaling pathways: Lessons learned from the yeast *Saccharomyces cerevisiae*. *Biochim Biophys Acta* 2007, 1773:1311–1340.

Christensen B, Nielsen J: Isotopomer analysis using GC-MS. *Metab Eng.* 1999, 1:282–290.

Cortassa S, Aon MA: The onset of fermentative metabolism in continuous cultures depends on the catabolite repression properties of *Saccharomyces cerevisiae. Enzyme Microb Technol*. 1998, 22:705–712.

Covert MW and Palsson BO: Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J Biol Chem* 2002, 277:28058–28064.

Croasmun WR, Carlson RMK: *Two-dimensional NMR spectroscopy: Applications for Chemists and Biochemists*. 2nd edition, VHC Publishers, Inc., New York, 1994, 958 p.

Cziferszky A, Mach RL, Kubicek CP: Phosphorylation Positively Regulates DNA Binding of the Carbon Catabolite Repressor Cre1 of *Hypocrea jecorina (Trichoderma reesei)*. *J Biol Chem.* 2002, 277:14688–14694.

Daran-Lapujade P, Rossell S, van Gulik WM, Luttik MAH, de Groot MJL, Slijper M, Heck AJR, Daran J-M, de Winde JH, Westerhoff HV, Pronk JT, Bakker BM: The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. *PNAS* 2007, 104:15753–15758.

D'Autréaux B, Toledano MB: ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis. *Nat rev Mol Cell Biol*. 2007, 8:813–824.

Davies BS, Rine J: A role for sterol levels in oxygen sensing in *Saccharomyces cerevisiae*. *Genetics* 2006, 174:191–201.

Davies SE, Brindle KM: Effects of overexpression of phosphofructokinase on glycolysis in the yeast *Saccharomyces cerevisiae*. *Biochemistry* 1992, 19:4729-35.

De Deken RH: The Crabtree effect: a regulatory system in yeast. *J Gen Microbiol.* 1966, 44:149–156.

de Groot MJL, Daran-Lapujade P, van Breukelen B, Knijnenburg TA, de Hulster EAF, Reinders MJT, Pronk JT, Heck AJR, Slijper M: Quantitative proteomics and transcriptomics of anaerobic and aerobic yeast cultures reveals post-transcriptional regulation of key cellular processes. *Microbiology* 2007, 153:3864–3878.

de Jong-Gubbels P, Bauer J, Niederberger P, Stückrath I, Kötter P, van Dijken JP, Pronk JT: Physiological characterisation of a pyruvate-carboxylase-negative *Saccharomyces cerevisiae* mutant in batch and chemostat. *Antonie Van Leeuwenhoek* 1998, 74:253–263.

Den Hollander JA, Behar KI, Shulman RG: ^{13}C NMR study of transamination during acetate utilization by *Saccharomyces cerevisiae*. *PNAS* 1981, 78:2693–2697.

Dirmeier R, O'Brien KM, Engle M, Dodd A, Spears E, Poyton RO: Exposure of yeast cells to anoxia induces transient oxidative stress: Implications for the induction of hypoxic genes. *J Biol Chem.* 2002, 277:34773–34784.

Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO: Global reconstruction of the human metabolic network based on genomic and bibliomic data. *PNAS* 2007, 104, 1777–1782.

Emanuelsson O, Brunak S, von Heijnen G, Nielsen H: Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc.* 2007, 2:953–971.

Emanuelsson O, Nielsen H, Brunak B, von Heijnen G: Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol.* 2000, 300:1005–1016.

Eppstein D: Finding the k shortest paths. *35th IEEE Symp. Foundations of Comp. Sci*., Santa Fe, 1994, pp. 154–165.

Ewald JC, Heux S, Zamboni N: High-Throughput Quantitative Metabolomics: Workflow for Cultivation, Quenching, and Analysis of Yeast in a Multiwell Format. *Anal Chem.* 2009, 81:3623–3629.

Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BO: Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*. 2009, 7:129–143.

Fiaux J, Cakar ZP, Sonderegger M, Wutrich K, Szyperski T, Sauer U: Metabolic-flux profiling of the yeasts *Saccharomyces cerevisiae* and *Pichia stipitis*. *Eukaryot Cell* 2003, 2:170–180.

Finkel T, Holbrook NJ: Oxidants, oxidative stress and the biology of ageing. *Nature* 2000, 408:239–247.

Fischer E, Sauer U: Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *Eur J Biochem.* 2003, 270:880–891.

Fischer E, Zamboni N, Sauer U: High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived ^{13}C constraints. *Anal Biochem*. 2004, 325:308–316.

Francke C, Siezen RJ, Teusink B: Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol.* 2005, 13:550–558.

Fredlund E, Blank LM, Schnurer J, Sauer U, Passoth V: Oxygen- and glucose-dependent regulation of central carbon metabolism in *Pichia anomala*. *Appl Env Microbiol.* 2004, 70:5905–5911.

Frick O, Wittmann C: Characterization of the metabolic shift between oxidative and fermentative growth in *Saccharomyces cerevisiae* by comparative ^{13}C flux analysis. *Microb Cell Fact*. 2005, 4:30.

Friebolin H: *Basic One- and Two-dimensional NMR Spectroscopy*. VCH Verlagsgesellschaft mbH, D-6940 Weinheim 1991, 344p.

Fuda NJ, Ardehali MB, Lis JT: Defining mechanisms that regulate RNA polymerase II transcription *in vivo*. *Nature*, 2009, 461:186–192.

Furukawa K, Heinzle E, Dunn IJ: Influence of Oxygen on the Growth of *Saccharomyces cerevisiae* in Continuous Culture. *Biotechnol Bioeng.* 1983, 25:2293–2317.

Förster J, Famili I, Fu P, Palsson BO, Nielsen J: Genome-scale reconstruction of the *Saccahromyces cerevisiae* metabolic network. *Genome Res.* 2003, 13:244–253.

Gadian DG: *Nuclear magnetic resonance and its applications to living systems*. Oxford University Press, Oxford 1982, 195p.

Gancedo JM: Yeast carbon catabolite repression. *Microbiol Mol Biol Rev.* 1998, 62:334–361.

Ghosh S, Zhu T, Grossmann I, Ataai M, Domach M: Closing the loop between feasible flux scenario identification for construct evaluation and resolution of realized fluxes via NMR. *Computers and Chemical Engineering* 2005, 29:459–466.

Gianchandani EP, Oberhardt MA, Burgard AP, Maranas CD, Papin JA: Predicting biological system objectives de novo from internal state measurements. *BMC Bioinformatics* 2008, 9:43.

Gombert AK, dos Santos MM, Christensen B, Nielsen J: Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. *J Bacteriol.* 2001, 183:1441–1451.

Grotkjaer T, Akesson M, Christensen B, Gombert AK, Nielsen J: Impact of transamination reactions and protein turnover on labeling dynamics in (13)C-labeling experiments. *Biotechnol Bioeng.* 2004, 86:209–216.

Heijnen JJ: Approximative kinetic formats used in metabolic network modelling. *Biotechnol Bioeng*. 2005, 91:534–545.

Herrero E, Ros J, Bellí G, Cabiscol E: Redox control and oxidative stress in yeast cells. *Biochim Biophys Acta* 2008, 1780:1217–1235.

Herrgård MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Blüthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novère N, Li P, Liebermeister W, Mo ML, Oliveira AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K, Spasić I, Weichart D, Brent R, Broomhead DS, Westerhoff HV, Kirdar B, Penttilä M, Klipp E, Palsson BØ, Sauer U, Oliver SG, Mendes P, Nielsen J, Kell DB: A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol.* 2008, 26:1155–1160.

Heuett WJ, Qian H: Combining flux and energy balance analysis to model large-scale biochemical networks. *J Bioinform Comput Biol*. 2006, 4:1227–1243.

Hoffmann U, Maier K, Niebel A, Vacun G, Reuss M, Mauch K: Identification of metabolic fluxes in hepatic cells from transient ^{13}C-labelling experiments: Part I. Experimental observations. *Biotechnol Bioeng.* 2008, 100:344–354.

Holzhutter H-G: The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur J Biochem.* 2004, 271:2905–2922.

Hon T, Dodd A, Dirmeier R, Gorman N, Sinclair PR, Zhang L, Poyton RO: A mechanism of oxygen sensing in yeast: multiple oxygen-responsive steps in the heme biosynthetic pathway affect Hap1 activity. *JBC* 2003, 278:50771–50780.

Huber A, Bodenmiller B, Uotila A, Stahl M, Wanka S, Gerrits B, Aebersold R, Loewith R: Characterization of the rapamycinsensitive phosphoproteome reveals that Sch9 is a central coordinator of protein synthesis. *Genes Dev.* 2009, 23:1929–1943.

Hulston CJ, Jeukendrup AE: No placebo effect from carbohydrate intake during prolonged exercise. *Int J Sport Nutr Exerc Metab.* 2009, 19:275–284.

Ilmén M: *Molecular mechanisms of glucose repression in the filamentous fungus Trichoderma reesei*. PhD thesis. VTT Publications 315, VTT Espoo; 1997. 86 p. + app. 50 p.

Ilmén M, Saloheimo A, Onnela M-L, Penttilä ME: Regulation of Cellulase Gene Expression in the Filamentous Fungus *Trichoderma reesei*. *Appl Env Microbiol.* 1997, 63:1298–1306.

Ilmén M, Thrane C, Penttilä M: The glucose repressor gene *cre1* of *Trichoderma*: Isolation and expression of a full length and truncated mutant form. *Mol Gen Genet.* 1996, 251:451–460.

Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A, Yee Ho P, Kakazu Y, Sugawara K, Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T, Hasegawa M, Takai Y, Yugi K, Arakawa K, Iwata N, Toya Y, Nakayama Y, Nishioka T, Shimizu K, Mori H, Tomita M: Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* 2007, 316:593–597.

Iwatani S, Van Dien S, Shimbo K, Kubota K, Kageyama N, Iwahata D, Miyano H, Hirayama K, Usuda Y, Shimizu K, Matsui K: Determination of metabolic flux changes during fed-batch cultivation from measurements of intracellular amino acids by LC-MS/MS. *J Biotechnol*. 2007, 128:93–111.

Jamieson DJ: Oxidative stress responses of the yeast *Saccharomyces cerevisiae*. *Yeast* 1998, 14:1511–1527.

Jones JG, Bellion E: Methanol oxidation and assimilation in *Hansenula polymorpha*. An analysis by ^{13}C n.m.r. *in vivo*. *Biochem J.* 1991, 280,475–481.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y: KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008, 36:D480–D484.

Karp PD, Paley S, Romero P: The Pathway Tools software. *Bioinformatics* 2002, 18:S225–S232.

Kitano H: Systems biology: a brief overview. *Science* 2002, 295:1662–1664.

Kitano H: Towards a theory of biological robustness. *Mol Syst Biol*. 2007, 3:137.

Kleijn RJ, Geertman J-MA, Nfor BK, Ras C, Schipper D, Pronk JT, Heijnen JJ, van Maris AJA, van Winden WA: Metabolic flux analysis of a glycerol-overproducing *Saccharomyces cerevisiae* strain based on GC-MS, LC-MS and NMR-derived ^{13}C-labelling data. *FEMS Yeast Res.* 2007, 7:216–231.

Klipp E: Modelling dynamic processes in yeast. *Yeast* 2007, 24:943–959.

Knorr AL, Jain R, Srivastava R: Bayesian-based selection of metabolic objective functions. *Bioinformatics* 2007, 23:351–357.

Koc A, Gasch AP, Rutherford JC, Kim H-Y, Gladyshev VN: Methionine sulfoxide reductase regulation of yeast lifespan reveals reactive oxygen specis-dependent and -independent components of aging. *PNAS* 2004, 101:7999–8004.

Koch LG, Britton SL: Aerobic metabolism underlies complexity and capacity. *J Physiol.* 2008, 586:83–95.

Kolkman A, Olsthoorn MMA, Heeremans CEM, Heck AJR, Slijper M: Comparative proteome analysis of *Saccharomyces cerevisiae* grown in chemostat cultures limited for glucose or ethanol. *Mol Cell Proteomics* 2005, 4:1–11.

Kotera M, Hattori M, Oh M-A, Yamamoto R, Komeno T, Yabuzaki J, Tonomura K, Goto S, Kanehisa M: RPAIR: a reactant-pair database representing chemical changes in enzymatic reactions. *Genome Inf.* 2004, 15:P062.

Kumar R, Singh S, Singh OV: Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *J Ind Microbiol Biotechnol.* 2008, 35:377–391.

Kummel A, Panke S, Heinemann M: Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Molecular Systems Biology*, 2006, 2:2006.0034.

Kuorelahti S, Kalkkinen N, Penttilä M, Londesborough J, Richard P: Identification in the mold *Hypocrea jecorina* of the first fungal D-galacturonic acid reductase. *Biochemistry* 2005, 44:11234–11240.

Kwast KE, Burke PV, Poyton RO: Oxygen sensing and the transcriptional regulation of oxygen-responsive genes in yeast. *J Exp Biol.* 1998, 201:1177–1195.

Lange HC: *Quantitative Physiology of S. cerevisiae using Metabolic Network Analysis*. PhD thesis. Technical University Delft, Delft, The Netherlands; 2002.

Lange HC, Heijnen JJ: Statistical reconciliation of the elemental and molecular biomass composition of *Saccharomyces cerevisiae*. *Biotechnol Bioeng.* 2001, 75:334–44.

Lazebnik Y. Can a biologist fix a radio? Or, what I learned while studying apoptosis. *Cancer Cell* 2002, 3:179–182.

Letavayová L, Marková E, Hermanská K, Vlčková V, Vlasáková D, Chovanec M, Brozmanová J: Relative contribution of homologous recombination and nonhomologous end-joining to DNA double-strand break repair after oxidative stress in *Saccharomyces cerevisiae*. *DNA Repair* 2006, 5:602–610.

Liebermeister W, Klipp E: Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor Biol Med Model*. 2006, 3:41.

Lin S-J, Kaeberlein M, Andalis AA, Sturtz LA, Defossez P-A, Culotta VC, Fink GR, Guarente L: Calorie restriction extends *Saccharomyces cerevisiae* lifespan by increasing respiration. *Nature* 2002, 418:344–348.

Liu Z, Butow RA: Mitochondrial retrograde signalling. *Annu Rev Genet.*. 2006, 40:159–185.

López-Mirabal HR, Winther JR: Redox characteristics of the eukaryotic cytosol. *Biochim Biophysi Acta* 2008, 1783:629–640.

Maaheimo H, Fiaux J, Çakar PZ, Bailey JE, Sauer U, Szyperski T: Central carbon metabolism of *Saccharomyces cerevisiae* explored by biosynthetic fractional ^{13}C labelling of common amino acids. *Eur J Biochem.* 2001, 268:2464–2479.

Ma'ayan A: Insights into the organization of biochemical regulatory networks using graph theory analyses. *JBC* 2009, 284:5451–5455.

Ma H-W, Zeng A-P: The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 2003, 19:1423–1430.

Maier K, Hofmann U, Reuss M, Mauch K: Identification of metabolic fluxes in hepatic cells from transient ^{13}C-labeling experiments: Part II. Flux estimation. *Biotechnol Bioeng.* 2008 100:355–370.

Mandels M, Reese ET: Induction of cellulase in *Trichoderma viride* as influenced by carbon sources and metals. *J Bacteriol.* 1957, 73:269–278.

Mann M, Jensen ON: Proteomic analysis of post-translational modifications. *Nat Biotechnol*. 2003, 21:255–261.

Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, Chapman J, Chertkov O, Coutinho PM, Cullen D, Danchin EGJ, Grigoriev IV, Harris P, Jackson M, Kubicek CP, Han CS, Ho I, Larrondo LF, de Leon AL, Magnuson JK, Merino S, Misra M, Nelson B, Putnam N, Robbertse B, Salamov AA, Schmoll M, Terry A, Thayer N, Westerholm-Parviainen A, Schoch CL, Yao J, Barbote R, Nelson MA, Detter C, Bruce D, Kuske CR, Xie G, Richardson P, Rokhsar DS, Lucas SM, Rubin EM, Dunn-Coleman N, Ward M, Brettin TS: Genome sequencing and analysis of the biomass-degrading *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotechnol.* 2008, 26:553–560.

Minning S, Serrano A, Ferrer P, Solà C, Schmid RD, Valero F: Optimisation of the high-level production of *Rhizopus oryzae* lipase in *Pichia pastoris*. *J Biotechnol.*,2001, 86:59–70.

Mo ML, Palsson BO, Herrgård MJ: Connecting extracellular metabolomic measurements to intracellualr flux states in yeast. *BMC Syst Biol.* 2009, 3:37.

Monod J, Wyman J, Changeux J-P: On the nature of allosteric transitions: A plausible model. *J Mol Biol.* 1965, 12:88–118.

Mu F, Williams RF, Unkefer CJ, Unkefer PJ, Faeder JR, Hlavacek WS: Carbon-fate maps for metabolic reactions. *Bioinformatics* 2007, 23:3193–3199.

Munger J, Bennett BD, Parikh A, Feng X-J, McArdle J, Rabitz HA, Shenk T, Rabinowitz: Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. *Nat Biotechnol*. 2008, 26:1179–1186.

Möllney M, Wiechert W, Kownatzki D, de Graaf A: Bidirectional reaction steps in metabolic networks IV: Optimal design of isotopomer labelling systems. *Biotechnol Bioeng.* 1999, 66:86–103.

Nevoigt E: Progress in metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev*. 2008, 72:379–412.

Nielsen J, Villadsen J, Lidén G: *Bioreaction engineering principles*. Kluyver Academic, New York, 2003, p. 528.

Nikerel IE, van Winden WA, van Gulik WM, Heijnen JJ: A method for estimation of elasticities in metabolic networks using steady state and dynamic metabolomics data and linlog kinetics. *BMC Bioinformatics* 2006, 7:540.

Nissen TL, Schulze U, Nielsen J, Villadsen J: Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*. *Microbiology* 1997, 143:203–218.

Notebaart RA, van Enckevort FHJ, Francke C, Siezen RJ, Teusink B: Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics* 2006, 7:296.

Nöh K, Grönke K, Luo B, Takors R, Oldiges M, Wiechert W: Metabolic flux analysis at ultra short time scale: Isotopically non-stationary ^{13}C labelling experiments. *J Biotechnol*. 2007, 129:249–267.

Oliveira GA, Tahara EB, Gombert AK, Barros MH, Kowaltowski: Increased aerobic metabolism is essential for the beneficial effects of caloric restriction on yeast life span. *J Bioenerg Biomembr.* 2008, 40:381–388.

Palmieri L, Vozza A, Agrimi G, De Marco V, Runswick J, Palmieri F, Walker JE: Identification of the Yeast Mitochondrial Transporter for Oxaloacetate and Sulfate. *JBC.*1999, 274:22184–22190.

Palsson, B. Ø. *Systems Biology: Properties of Reconstructed Networks*, 2006, Cambridge University Press, New York.

Patil KR, Nielsen J: Uncovering transcriptional regulation of metabolism by using metabolic network topology. *PNAS* 2005, 102:2685–2689.

Perrone GG, Grant CM, Dawes IW: Genetic and environmental factors influencing glutathione homeostasis in *Saccharomyces cerevisiae*. *Mol Biol Cell* 2005, 16:218–230.

Petranovic D, Nielsen J: Can yeast systems biology contribute to the understanding of human disease? *Trends Biotechnol.* 2008, 11:584-90.

Petranovic D, Vemuri GN: Impact of yeast systems biology on industrial biotechnology. *J Biotechnol.* 2009, doi:10.1016/j.jbiotec.2009.07.005, in Press.

Pfeiffer T, Schuster S, Bonhoeffer S: Cooperation and Competition in the Evolution of ATP-producing Pathways. *Science* 2001, 292:504–507.

Pharkya P, Burgard AP, Maranas CD: OptStrain: A computational framework for redesign of microbial prodcution systems. *Genome Res*. 2004, 14:2367–2376.

Pinney JW, Shirley MW, McConkey GA, Westhead DR: metaSHARK: software for automatic metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*. *Nucleic Acids Res.* 2005, 33:1399–1409.

Pitkänen E, Jouhten P, Rousu J: Inferring branching pathways in genome-scale metabolic networks with ReTrace. *BMC Syst. Biol.* 2009, 3:103.

Pitkänen E, Åkerlund A, Rantanen A, Jouhten P, Ukkonen E: ReMatch: a web-based tool to construct, store and share stoichiometric metabolic models with carbon maps for metabolic flux analysis. *J Integr Bioinform.* 2008, 5:102.

Postmus J, Canelas AB, Bouwman J, Bakker BM, van Gulik W, de Mattos MJT, Brul S, Smits GJ: Quantitative analysis of the high temperature-induced glycolytic flux increase in *Saccharomyces cerevisiae* reveals dominant metabolic regulation. *JBC* 2008, 283:23524–23532.

Price ND, Schellenberger J, Palsson BO: Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. *Biophys J*. 2004, 87:2172–2186.

Price ND, Reed JL, Palsson BO: Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol.* 2004, 2:886–897.

Price ND, Thiele I, Palsson BO: Candidate States of *Helicobacter pylori*'s Genome-Scale Metabolic Network upon Application of "Loop Law" Thermodynamic Constraints. *Biophys J*. 2006, 90:3919–3928.

Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, Fasolo J, Guo H, Jona G, Breitkreutz A, Sopko R, McCartney RR, Schmidt MC, Rachidi N, Lee S-J, Mah AS, Meng L, Stark MJR, Stern DF, De Virgilio C, Tyers M, Andrews B, Gerstein M, Schweiter B, Predki PF, Snyder M: Global analysis of protein phosphorylation in yeast. *Nature* 2005, 438:679–684.

Qian H, Beard DA: Metabolic futile cycles and their functions: a systems analysis of energy and control. *IEE Proceedings Systems Biology* 2006, 153:192–200.

Quek L-E, Wittmann C, Nielsen LK, Krömer JO: OpenFLUX: efficient modelling software for ^{13}C-based metabolic flux analysis. *Microb Cell Fact.* 2009, 8:25.

Raamsdonk LM., Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ, Westerhoff HV, van Dam K, Oliver SG: A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol.* 2001, 19:45–50.

Raghevendran V, Patil KR, Olsson L, Nielsen J: Hap4 is not essential for activation of respiration at low specific growth rates in *Saccharomyces cerevisiae*. *JBC* 2006, 281:12308–12314.

Rantanen A, Mielikäinen T, Rousu J, Maaheimo H, Ukkonen E: Planning optimal measurements of isotopomer distributions for estimation of metabolic fluxes. *Bioinformatics* 2006, 22:1198–1206.

Rintala E, Toivari M, Pitkänen J-P, Wiebe MG, Ruohonen L, Penttilä M: Low oxygen levels as a trigger for enhancement of respiratory metabolism in *Saccharomyces cerevisiae*. *BMC Genomics* 2009, 10:461.

Rohde JR, Bastidas R, Puria R, Cardenas ME: Nutritional control via Tor signaling in *Saccharomyces cerevisiae*. *Curr Opin Microbiol*. 2008, 11:153–160.

Sauer U: High-Throughput phenomics: experimental methods for mapping fluxomes. *Curr Opin Biotechnol*. 2004, 15:58–63.

Savageau M: Biochemical systems analysis: III. Dynamic solutions using power-law approximation. *J Theor Biol*. 1970, 26:215–226.

Schaaff I, Heinisch J, Zimmermann FK: Overproduction of glycolytic enzymes in yeast. *Yeast* 2004, 5:285–290.

Schaub J, Mauch K, Reuss M: Metabolic flux analysis in *Escherichia coli* by integrating isotopic dynamic and isotopic stationary ^{13}C labelling data. *Biotechnol Bioeng*. 2008, 99:1170–1185.

Schellenberger J, Palsson BO: Use of randomized sampling for analysis of metabolic networks. *JBC* 2009, 284:5457–5461.

Schomburg I, Chang A, Schomburg D: BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* 2002, 30:47–49.

Schryer DW, Peterson P, Paalme T, Vendelin M: Bidirectionality and compartmentation of metabolic fluxes are revealed in the dynamics of isotopomer networks. *Int J Mol Sci.* 2009, 10:1697–1718.

Schuetz R, Kuepfer L, Sauer U: Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol.* 2007, 3:119.

Segrè D, Vitkup D, Church GM: Analysis of optimality in natural and perturbed metabolic networks. *PNAS* 2002, 99:15112–15117.

Selivanov VA, Marin S, Lee PWN, Cascante M: Software for dynamic analysis of tracer-based metabolomic data: estimation of metabolic fluxes and their statistical analysis. *Bioinformatics* 2006, 22:2806–2812.

Sellick CA, Reece RJ: Modulation of transcription factor function by an amino acid: activation of Put3p by proline. *EMBO J*. 2003, 22:5147–5153.

Shimizu K: Metabolic flux analysis based on ^{13}C-labeling experiments and integration of the information with gene and protein expression patterns. *Adv Biochem Eng Biotechnol.* 2004, 91:1–49.

Shinfuku Y, Sorpitiporn N, Sono M, Furusawa C, Hirasawa T, Shimizu H: Development and experimental verification of a genome-scale metabolic model for *Corynebacterium glutamicum*. *Microb Cell Fact*. 2009, 8:43.

Sickmann A, Reinders J, Wagner Y, Joppich C, Zahedi R, Meyer HE, Schönfisch B, Perschil I, Chacinska A, Guiard B, Rehling P, Pfanner N, Meisinger C: The proteome of *Saccharomyces cerevisiae* mitochondria. *PNAS* 2003, 100:13207–13212.

Sola A, Maaheimo H, Ylonen K, Ferrer P, Szyperski T: Amino acid biosynthesis and metabolic flux profiling of *Pichia pastoris*. *Eur J Biochem.* 2004, 271:2462–2470.

Strauss J, Mach RL, Zeilinger S, Hartler G, Stöffler G, Wolschek M, Kubicek CP: Cre1, the carbon catabolite repressor protein from *Trichoderma reesei*. *FEBS Lett.* 1995, 376:103–107.

Stephanopoulos GN, Aristidou AA, Nielsen J: *Metabolic Engineering – Principles and Methodologies*, Academic Press, San Diego 1998, 725 p.

Stryer L: *Biochemistry*, 4th edition, W.H. Freeman and Company, New York 1995, 1064 p.

Sumegi B, Sherry AD, Malloy CR, Srere PA: Evidence for orientation-conserved transfer in the TCA cycle in *Saccharomyces cerevisiae*: ^{13}C NMR studies. *Biochemistry* 1993, 32:12725–12729.

Suthers PF, Burgard AP, Dasika MS, Nowroozi F, Van Dien S, Keasling J, Maranas CD: Metabolic flux elucidation for large-scale models using ^{13}C labelled isotopes. *Metab Eng*. 2007, 9:387–405.

Swiegers J, Dippenaar N, Pretorius I, Bauer F: Carnitine-dependent metabolic activities in *Saccharomyces cerevisiae*: three carnitine acetyltranferases are essential in a carnitine-dependent strain.*Yeast* 2001, 18:585–595.

Szyperski T: Biosynthetically directed fractional 13C-labelling of proteinogenic amino acids. An efficient tool to investigate intermediary metabolism. *Eur J Biochem.* 1995, 232:433–448.

Szyperski T: ^{13}C-NMR, MS and metabolic flux balancing in biotechnology research. *Q Rev Biophys*. 1998, 31:41–106.

Szyperski T, Glaser RW, Hochuli M, Fiaux J, Sauer U, Bailey JE, Wüthrich K: Bioreaction network topology and metabolic flux ratio analysis by biosynthetic fractional ^{13}C labelling and two-dimensional NMR spectroscopy. *Metab Eng.* 1999, 1:189–197.

Tang YJ, Garcia Martin H, Myers S, Rodriquez S, Baidoo EEK, Keasling JD: Advances in analysis of microbial metabolic fluxes via (13)C isotopic labeling. *Mass Spectrom Rev*. 2009, 28:362–375.

ter Kuile BH, Westerhoff HV: Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Letters* 2001, 500:169–171.

The UniProt Consortium: The Universal Protein Resource (UniProt). *Nucleic Acids Res.* 2007, 35:D193–197.

Teusink B, Passarge J, Reijenga CA, Esgalhado E, van der Weijden CC, Schepper M, Walsh MC, Bakker BM, van Dam K, Westerhoff HV, Snoep JL: Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *Eur J Biochem.* 2000, 267:5313–5329.

Toya Y, Ishii N, Hirasawa T, Naba M, Hirai K, Sugawara K, Igarashi S, Shimizu K, Tomita M, Soga T: Direct measurement of isotopomer of intracellular metabolites using capillary electrophoresis time-of-flight mass spectrometry for efficient metabolic flux analysis. *J Chromatogr* A. 2007, 1159:134–141.

Trotter E, Grant CM: Thioredoxins are required for protection against a reductive stress in the yeast *Saccharomyces cerevisiae*. *Mol Microbiol.* 2002, 46:869–878.

Uy R, Wold F: Posttranslational covalent modification of proteins. *Science* 1977, 198:890–896.

van der Brink J, Canelas AB, van Gulik WM, Pronk JT, Heijnen JJ, de Winde JH, Daran-Lapujade P: Dynamics of glycolytic regulation during adaptation of *Saccharomyces cerevisiae* to fermentative metabolism. *Appl Env Microbiol.* 2008, 74:5710–5723.

van Roermund C, Hettema E, Berg M, Tabak H, Wanders R: Molecular characterization of carnitine-dependent transport of acetyl-CoA from peroxisomes to mitochondria in *Saccharomyces cerevisiae* and identification of a plasma membrane carnitine transporter, Agp2p. *EMBO J.* 1999, 18:5843–5852.

van Winden W, Heijnen JJ, Verheijen PJT: Cumulative Bondomers: A New Concept in Flux Analysis from 2D [^{13}C,1H] COSY NMR Data. *Biotechnol Bioeng.* 2002, 80:731–745.

van Winden WA, van Dam JC, Ras C, Kleijn RJ, Vinke JL, van Gulik WM, Heijnen JJ: Metabolic-flux analysis of *Saccharomyces cerevisiae* CEN.PK113-7D based on

mass isotopomer measurements of ^{13}C-labeled primary metabolites. *FEMS Yeast Res.* 2005, 5:559–568.

van Winden WA, Wittmann C, Heinzle E, Heijnen JJ: Correcting mass isotopomer distributions for naturally occurring isotopes. *Biotechnol Bioeng.* 2002, 80:477–479.

Vaseghi S, Baumeister A, Rizzi M, Reuss M: *In vivo* dynamics of the pentose phosphate pathway in *Saccharomyces cerevisiae*. *Metab Eng.* 1999, 1:128–140.

Vautard G, Cotton P, Fèvre M: The glucose repressor CRE1 from *Sclerotina sclerotiorum* is functionally related to CREA from *Aspergillus nidulans* but not to the Mig proteins from *Saccharomyces cerevisiae*. *FEBS Letters* 1999, 453:54–58.

Vemuri GN, Eiteman MA, McEwen JE, Olsson L, Nielsen J: Increasing NADH oxidation reduces overflow metabolism in *Saccharomyces cerevisiae*. *PNAS* 2007, 104:2402–2407.

Verduyn C, Postma, E Scheffers WA, van Dijken JP: Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 1992, 8:501–517.

Visser D, Heijnen J: Dynamic simulation and metabolic redesign of a branched pathway using linlog kinetics. *Metab Eng.* 2003, 5:164–176.

Visser D, Schmid J, Mauch K, Reuss M, Heijnen J: Optimal re-design of primary metabolism in *Escherichia coli* using linlog kinetics. *Metab Eng.* 2004, 6:378–390.

Visser W, van der Baan AA, Batenburg-van der Vegte W, Scheffers A, Krämer R, van Dijken JP: Involvement of mitochondria in the assimilatory metabolism of anaerobic *Saccharomyces cerevisiae* cultures. *Microbiology* 1994, 140:3039–3046.

Wahl SA, Nöh K, Wiechert W: ^{13}C labelling experiments at metabolic nonstationary conditions: An exploratory study. *BMC Bioinformatics* 2008, 9:152.

Westergaard SL, Oliveira AP, Bro C, Olsson L, Nielsen J: A systems biology approach to study glucose repression in the yeast *Saccharomyces cerevisiae*. *Biotechnol Bioeng.*2007, 96:134–145.

Weusthuis RA, Visser W, Pronk JT, Scheffers WA, van Dijken JP: Effects of oxygen limitation on sugar metabolism in yeasts – a continuous-culture study of the Kluyver effect. *Microbiology* 1994, 140:703–715.

Wiebe MG, Rintala E, Tamminen A, Simolin H, Salusjärvi L, Toivari M, Kokkonen JT, Kiuru J, Ketola RA, Jouhten P, Huuskonen A, Maaheimo H, Ruohonen L, Penttilä M: Central carbon metabolism of *Saccharomyces cerevisiae* in anaerobic,

oxygen-limited and fully aerobic steady-state conditions and following a shift to anaerobic conditions. *FEMS Yeast Res.* 2008, 8:140–154.

Wiechert W: ^{13}C metabolic flux analysis. *Metab Eng.* 2001, 3:195–206.

Wiechert W, Möllney M, Isermann N, Wurzel M, de Graaf AA: Bidirectional Reaction Steps in Metabolic Networks: III. Explicit Solution and Analysis of Isotopomer Labeling Systems. *Biotechnol Bioeng.* 1999, 66:69–85.

Wiechert W, Möllney M, Petersen S, de Graaf AA: A Universal Framework for ^{13}C Metabolic Flux Analysis. *Metab Eng.* 2001, 3:265–283.

Wild S, Roglic G, Green A, Sicree R, King H: Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* 2004, 27:1047–1053.

Wittmann C: Fluxome analysis using GC-MS. *Microb Cell Fact*. 2007, 6:6.

Wittmann C, Hans M, Heinzle E: *In vivo* analysis of intracellular amino acids labelings by GC/MS. *Anal Biochem.* 2002, 307:379–382.

Woo DK, Phang TL, Trawick JD, Poyton RO: Multiple pathways of mitochondrial-nuclear communication in yeast: Intergenomic signalling involves ABF1 and affects a different set of genes than retrograde regulation. *Biochim Biophys Acta* 2009, 1789:135–145.

Zaman S, Lippman SI, Schneper L, Slonim N, Broach JR: Glucose regulates transcription in yeast through a network of signaling pathways. *Mol Syst Biol*. 2009, 5:245.

Zaman S, Lippman SI, Zhao X, Broach JR: How *Saccharomyces* responds to nutrients. *Annu Rev Genet.*. 2008, 42:2.1–2.55.

Zamboni N, Fendt S-M, Ruhl M, Sauer U: ^{13}C-based metabolic flux analysis. *Nat Protoc.* 2009, 4:878–892.

Zamboni N, Fischer E, Sauer U: FiatFlux – a software for metabolic flux analysis from ^{13}C-glucose experiments. *BMC Bioinformatics* 2005, 6:209.

Åkesson M, Forster J and Nielsen J: Integration of gene expression data into genome-scale metabolic models. *Metab Eng.* 2004, 6:285–293.

Appendix V of this publication is not included in the PDF version. Please order the printed version to get the complete publication (http://www.vtt.fi/publications/index.jsp).

Series title, number and report code of publication

VTT Publications 724
VTT-PUBS-724

Author(s)
Paula Jouhten

Title

Metabolic modelling and ^{13}C flux analysis.
Application to biotechnologically important yeasts and a fungus

Abstract

All bioconversions in cells derive from metabolism. Microbial metabolisms contain potential for bioconversions from simple source molecules to unlimited number of biochemicals and for degradation of even detrimental compounds. Metabolic fluxes are rates of consumption and production of compounds in metabolic reactions. Fluxes emerge as an ultimate phenotype of an organism from an integrated regulatory function of the underlying networks of complex and dynamic biochemical interactions. Since the fluxes are time-dependent, they have to be inferred from other, measurable, quantities by modelling and computational analysis. ^{13}C-labelling is crucial for quantitative analysis of fluxes through intracellular alternative pathways. Local flux ratio analysis utilises uniform ^{13}C-labelling experiments, where the carbon source contains a fraction of uniformly ^{13}C-labelled molecules. Carbon-carbon bonds are cleaved and formed in metabolic reactions depending on the *in vivo* fluxes. ^{13}C-labelling patterns of metabolites or macromolecule components can be detected by mass spectrometry (MS) or nuclear magnetic resonance (NMR) spectroscopy. Local flux ratio analysis utilises directly the ^{13}C-labelling data and metabolic network models to solve ratios of converging fluxes.

In this thesis the local flux ratio analysis has been extended and applied to analysis of phenotypes of biotechnologically important yeasts *Saccharomyces cerevisiae* and *Pichia pastoris,* and a fungus *Trichoderma reesei*. Oxygen dependence of *in vivo* net flux distribution of *S. cerevisiae* was quantified by using local flux ratios as additional constraints to the stoichiometric model of the central carbon metabolism. The distribution of fluxes in the pyruvate branching point turned out to be most responsive to different oxygen availabilities. The distribution of fluxes was observed to vary not only between the fully respiratory, respiro-fermentative and fermentative metabolic states but also between different respiro-fermentative states. The local flux ratio analysis was extended to the case of two-carbon source of glycerol and methanol co-utilisation by *P. pastoris*. The fraction of methanol in the carbon source did not have as profound effect on the distribution of fluxes as the growth rate. The effect of carbon catabolite repression (CCR) on fluxes of *T. reesei* was studied by reconstructing amino acid biosynthetic pathways and by performing local flux ratio analysis. *T. reesei* was observed to primarily utilise respiratory metabolism also in conditions of CCR. *T. reesei* metabolism was further studied and L-threo-3-deoxy-hexulosonate was identified as L-galactonate dehydratase reaction product by using NMR spectroscopy. L-galactonate dehydratase reaction is part of the fungal pathway for D-galacturonic acid catabolism.

ISBN
978-951-38-7371-4 (soft back ed.)
978-951-38-7372-1 (URL: http://www.vtt.fi/publications/index.jsp)

Series title and ISSN	Project number
VTT Publications 1235-0621 (soft back ed.) 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)	

Date	Language	Pages
November 2009	English, Finnish abstr.	94 p. + app. 83 p.

Name of project	Commissioned by

Keywords	Publisher
Metabolic modelling, metabolic flux, metabolic flux analysis (MFA), ^{13}C-labelling, ^{13}C-MFA, nuclear magnetic resonance (NMR) spectroscopy	VTT Technical Research Centre of Finland P.O.Box 1000, FI-02044 VTT, Finland Phone internat. +358 20 722 4520 Fax +358 20 722 4374

Julkaisun sarja, numero ja raporttikoodi

VTT Publications 724
VTT-PUBS-724

Tekijä(t)
Paula Jouhten

Nimeke

Aineenvaihdunnan mallinnus ja ^{13}C-vuoanalyysi
Sovellukset bioteknologisesti tärkeisiin hiivoihin ja homeeseen

Tiivistelmä

Aineenvaihdunta kattaa kaikki biomuunnokset soluissa. Mikrobiaineenvaihdunta mahdollistaa yksinkertaisten lähtöaineiden muuntamisen rajoittamattomaksi määräksi erilaisia biokemikaaleja ja jopa haitallisten aineiden hajottamisen. Aineenvaihduntavuot ovat yhdisteiden kulutus- ja tuottonopeuksia aineenvaihdunnan reaktioissa. Vuot ilmentyvät organismin todellisena fenotyyppinä, jota säätelevät yhteistoiminnallisesti solun monimutkaiset ja dynaamiset vuorovaikutusverkot. Koska vuot ovat aikariippuvaisia, ne on määritettävä mallinnuksen ja laskennallisen analyysin avulla toisista, mitattavissa olevista, suureista. ^{13}C-leimaus on välttämätöntä, jotta vuot vaihtoehtoisilla solunsisäisillä reiteillä voidaan määrittää kvantitatiivisesti. Paikallisessa vuosuhdeanalyysissa käytetään tasaista ^{13}C-leimausta, jossa hiilenlähde sisältää osuuden täydellisesti ^{13}C-leimattuja molekyylejä. *In vivo* -vuot määräävät missä suhteissa aineenvaihdunnassa katkeaa ja muodostuu uusia hiili-hiilisidoksia. Aineenvaihdunnan välituotteiden ja makromolekyylien komponenttien ^{13}C-leimauskuvioita voidaan mitata massaspektrometrialla (MS) tai ydinmagneettisella resonanssispektroskopialla (NMR). Paikallisessa vuosuhdeanalyysissa käytetään suoraan mittausinformaatiota ^{13}C-leimauskuvioista ja aineenvaihduntaverkkomalleja vuosuhteiden ratkaisemiseksi.

Väitöskirjassa paikallista vuosuhdeanalyysia laajennettiin ja sovellettiin bioteknologisesti tärkeiden hiivojen *Saccharomyces cerevisiae* ja *Pichia pastoris,* ja homeen *Trichoderma reesei* fenotyyppien analysoimiseksi. *S. cerevisiae*n *in vivo* -vuojakauman riippuvuus hapen saatavuudesta määritettiin kvantitatiivisesti käyttämällä paikallisia vuosuhteita lisärajoitteina keskeisen hiiliaineenvaihdunnan stoikiometriselle mallille. Pyruvaattiristeyksen vuojakauma osoittautui herkimmäksi eri happisaatavuuksille. Selvästi erilaiset vuojakaumat havaittiin täysin respiratiivisessa, respiro-fermentatiivisessa ja täysin fermentatiivisessa aineenvaihdunnan tilassa, mutta myös eri respiro-fermentatiivisissa tiloissa. Paikallinen vuosuhdeanalyysi laajennettiin kahden hiilenlähteen tapaukseen, jossa *P. pastoris* kulutti samanaikaisesti glyserolia ja metanolia. Metanolin osuudella kokonaishiilenlähteessä ei ollut yhtä merkittävää vaikutusta vuojakaumaan kuin hiivan kasvunopeudella. Hiilikataboliittirepression (CCR) vaikutusta *T. reesei*n vuojakaumaan tutkittiin rekonstruoimalla aminohapposynteesireitit ja tekemällä paikallinen vuosuhdeanalyysi. *T. reesei*n havaittiin käyttävän pääasiassa respiratiivista aineenvaihduntaa myös repressoivissa olosuhteissa. NMR-spektroskopiaa käytettiin myös D-galakturonihapon kaboliareitin tutkimuksessa ja L-treo-3-deoksi-heksulonaatti tunnistettiin *T. reesei*n L-galaktonaattidehydrataasireaktion tuotteeksi.

ISBN
978-951-38-7371-4 (nid.)
978-951-38-7372-1 (URL: http://www.vtt.fi/publications/index.jsp)

Avainnimeke ja ISSN	Projektinumero
VTT Publications 1235-0621 (nid.) 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)	

Julkaisuaika	Kieli	Sivuja
Marraskuu 2009	Englanti, suom. tiiv.	94 s. + liitt. 83 s.

Projektin nimi	Toimeksiantaja(t)

Avainsanat	Julkaisija
Metabolic modelling, metabolic flux, metabolic flux analysis (MFA), ^{13}C-labelling, ^{13}C-MFA, nuclear magnetic resonance (NMR) spectroscopy	VTT PL 1000, 02044 VTT Puh. 020 722 4520 Faksi 020 722 4374

Metabolisms of microorganisms contain possibilities for conversions of simple source molecules to unlimited number of biochemicals and for degradation of even hazardous compounds. Rates of metabolic reactions are called fluxes. They are in sense process streams of a cell factory in case of biotechnologically important organisms. Since the fluxes are time-dependent, they cannot be directly measured but have to be inferred from other, measurable, quantities by modelling and computational analysis. ^{13}C-labelling is crucial for quantitative analysis of fluxes through alternative pathways inside the cells. Fluxes emerge as an ultimate phenotype of an organism from an integrated regulatory function of the underlying networks of complex and dynamic biochemical interactions. Inferring fluxes and their regulation in simple model organisms aids in understanding for example metabolic disorders in human. The dissertation considers modelling of metabolism and ^{13}C-labelling for quantitative analysis of metabolic fluxes in yeast *Saccharomyces cerevisiae* that is an important biotechnological production and model organism, and in yeast *Pichia pastoris* and in fungus *Trichoderma reesei* that serve as efficient hosts for protein production.

ISBN 978-951-38-7371-4 (soft back ed.)
ISSN 1235-0621 (soft back ed.)

ISBN 978-951-38-7372-1 (URL: http://www.vtt.fi/publications/index.jsp)
ISSN 1455-0849 (URL: http://www.vtt.fi/publications/index.jsp)