Stina K. Lien

Mass spectrometry based metabolomics for *Pseudomonas fluorescens*

Investigating effects of anti-sigma factor MucA and carbon sources on metabolism

Thesis for the degree of Philosophiae Doctor

Trondheim, November 2013

Norwegian University of Science and Technology
Faculty of Natural Sciences and Technology
Department of Biotechnology

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Natural Sciences and Technology
Department of Biotechnology

ISBN 978-82-471-4837-2 (printed ver.)
ISBN 978-82-471-4838-9 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2013:347

Printed by NTNU-trykk

Acknowledgments

The work presented in this Doctoral thesis has been carried out at the Department of Biotechnology (IBT), Norwegian University of Science and Technology (NTNU). It was financially supported by an internal grant at NTNU and by the ERA-NET Systems Biology of Microorganisms (SysMO) program.

The work has been supervised by Associate Professor Per Bruheim and I would like to thank him for his guidance, enthusiasm and support. This work has greatly benefitted from his knowledge in the research field and his availability and commitment as a supervisor. I would also like to thank the remaining members of the Mass Spec Metabolomics research group at IBT, Hans Fredrik Nyvold Kvitvang, Kåre Andre Kristiansen and past and present Masters Students, for contributing to a socially and academically rewarding work environment. And I am also grateful to past and present colleagues at IBT for help and collaboration with small and large issues throughout my work. I would also like to extend my gratitude to all external collaborators. At SINTEF Materials and Chemistry, Department of Biotechnology especially the Mass Spectrometry analytic team and the great support from Håvard Sletta and Randi Aune when running cultivations has been invaluable. In addition I would also like to thank Sebastian Niedenführ and Katharina Nöh at Forschungszentrum Jülich for performing the required experimental design in the software 13CFLUX2 and for their help when running simulations in the software.

Last but not least I would also like to thank my friends and family, and especially my husband Emanuel Ion Roman for being so patient and supporting through the final stages of this work, and my good friend Maria Førde Møll for all the motivational morning coffees and lunch session we have had.

Trondheim, September 2013
Stina Katrine Lien

Abstract

Pseudomonas is a bacterial genus containing species of both industrial and medical relevance due to their metabolic diversity and ability to colonize a wide variety of ecological niches including soil, water, insects, plants and animals. The biosynthetic diversity of *Pseudomonas* strains includes production of the polysaccharide alginate. Alginate has many commercial applications, but it is a complication to the human host during *P. aeruginosa* infections. In *P. aeruginosa* alginate production is often initiated by inactivation of the pleotropic anti-sigma factor MucA. The non-pathogenic strain *P. fluorescens* SBW25 used in this study does not produce alginate, but alginate production can also in this strain be initiated by MucA inactivation.

In a metabolome study the effects of inactivation of anti-sigma factor MucA was investigated, both in the presence and absence of alginate production, and when using different carbon sources. The investigation was conducted using nitrogen-limited fructose and glycerol chemostat cultivations. The strains used were *P. fluorescens* SBW25 wild type, an alginate producing *mucA*- strain and two alginate non-producing *mucA*- strains (a *mucA*- Δ*algC* strain for fructose cultivations and a *mucA*- TT*algD* strain for glycerol cultivations). Cultivation data from these chemostats showed that all *mucA*- mutants had an about 40% decreased fructose uptake rate, and an about 20% decreased glycerol uptake rate compared to the wild type (correcting for the carbon source fraction shuttled to alginate synthesis for the alginate producing *mucA*- strain). The metabolome of the various strains on the two carbon sources were characterized by preparing metabolite extracts, and analyzing them by gas chromatography – mass spectrometry (GC-MS) and liquid chromatography – tandem mass spectrometry (LC-MS/MS). The metabolome datasets showed that the different strains had distinct metabolite compositions depending on the carbon source utilized. However, amino acids and organic acids had relatively similar concentrations for all strains. Similar amino acid pools are not unexpected as all strains were cultivated at the same growth rate in nitrogen-limited chemostats. The metabolome study had two striking findings. The first finding was that for both carbon sources, the guanine nucleotide pools for the *mucA*- strains (the *mucA*- strain, the *mucA*- Δ*algC* strain and the *mucA*- TT*algD* strain)

differed from the wild type in an alginate production dependent manner. The alginate production dependent change in guanine nucleotides for the *mucA-* strains was not surprising, as GTP is utilized in alginate biosynthesis. However, what causes the specific differences in pool sizes for the alginate producing and non-producing *mucA-* strains is not clear. The second finding was that for both carbon sources, the adenine nucleotide pools for the *mucA-* strains (the *mucA-* strain, the *mucA-* Δ*algC* strain and the *mucA-* TT*algD* strain), differed from the wild type in an alginate production non-dependent manner. This resulted in a decreased energy charge (EC) for all *mucA-* strains compared to the wild type, an indication of the pleiotropic effects of MucA inactivation.

A fluxome study was performed to complement the metabolome study, and to further elucidate the effect of MucA inactivation on the metabolism of *P. fluorescens*. In this study ^{13}C-labeled fructose was used in nitrogen-limited chemostat cultivations of *P. fluorescens* SBW25 and the non-alginate producing *mucA-* Δ*algC* strain. Metabolite extract from the cultivations were analyzed by GC-MS/MS and LC-MS/MS producing metabolite mass isotopomer datasets. These datasets were then used in the simulation software 13CFLUX2 to determine intracellular fluxes in central carbon metabolism for *P. fluorescens* in the presence and absence of an active MucA. The flux distribution results revealed that the net fluxes of central carbon metabolism proceed in the same direction for both wild type and the *mucA-* Δ*algC* strain, and that they both utilize the same main route for fructose uptake. The study also revealed that there are distinct differences between the two strains at important branch points in primary metabolism, and that such differences often coincide with changes in metabolite concentrations at, or close to, the specific branch points. The fluxome study also had two striking findings. The first finding was that compared to the wild type, the *mucA- ΔalgC* strain has an increased flux through the pentose phosphate pathway (PPP), and a decreased flux through the Entner – Doudoroff pathway (EDP). The second finding was that the *mucA- ΔalgC* strain utilizes the glyoxylate shunt, at the expense of the tricarboxylic acid cycle (TCA), whilst the wild type does not. This overall flux distribution of the *mucA-* Δ*algC* strain causes it to produce less NADH than the wild type, whilst ATP and NADPH production is similar. The reduced NADH production for the *mucA-* Δ*algC* strain is

probably a factor contributing to its reduced EC, as less NADH is available for oxidative phosphorylation in this strain. The results from the metabolome study and the fluxome study have revealed new aspect of *P. fluorescens* metabolism in connection to MucA inactivation and the results can also act as a basis for future studies.

An integral part of this doctoral work was development of sample preparation protocols for metabolite extracts, and development of MS-methods for sample analysis. For the metabolome study, a targeted and a non-targeted GC-MS method for alkylated metabolites was developed, and a targeted reversed phase ion-pairing LC-MS/MS method was implemented. For the fluxome study a targeted GC-MS/MS method for alkylated metabolites was developed for detection of mass isotopomers, whilst the LC-MS/MS method from the metabolome study was expanded and quality tested for detection of mass isotopomers. In addition an isotope coded derivatization (ICD) GC-MS/MS method that improves precision for analyses of samples of silylated metabolites was developed. The silylation ICD GC-MS/MS method can complement more established normalization techniques such as isotope dilution mass spectrometry (IDMS) in the field of mass spectrometry based metabolomics.

This PhD project has contributed to the development of mass spectrometry methodology for metabolomics and fluxomics, and has employed these methods to elucidate the metabolic consequences of MucA inactivation in *P. fluorescens*.

List of papers and appendices

Original research papers

Paper I

Stina K. Lien, Hans Fredrik Nyvold Kvitvang, and Per Bruheim. (2012). "Utilization of a deuterated derivatization agent to synthesize internal standards for gas chromatography-tandem mass spectrometry quantification of silylated metabolites." Journal of Chromatography A. 1247: 118-124.

Paper II

Stina K. Lien, Håvard Sletta, Trond E. Ellingsen, Svein Valla, Elon Correa, Royston Goodacre, Kai Vernstad, Sven Even Finborud Borgos, and Per Bruheim. (2013). "Ivestigating alginate production and carbon utilization in *Pseudomonas fluorescens* SBW25 using mass spectrometry-based metabolic profiling". Metabolomics. 9: 403-417.

Paper III

Stina K. Lien, Sebastian Niedenführ, Håvard Sletta, Katharina Nöh, and Per Bruheim. (*In prepp*.). "Fluxome study reveals that inactivation of anti-sigma factor MucA cause downregulation of Entner-Doudoroff pathway and upregulation of glyoxylate shunt in *Pseudomonas fluorescens*".

Book Chapters

Appendix IV

Hans Fredrik Nyvold Kvitvang, Kåre A. Kristiansen, Stina K. Lien, and Per Bruheim. (*In pres*s). Quantitative analysis of amino and organic acids by methyl chloroformate derivatization and GC-MS/MS analysis. Springer Molecular Book series: Mass Spectrometry Methods in Metabolomics. Daniel Raftery. Humana Press.

Appendix V

Per Bruheim, Trygve Andreassen, and Stina K. Lien. (*In press*). Analytical techniques from the perspective of microbial metabolomics. Microbial Metabolomics: Advances in Molecular and Cellular Microbiology. Silas Villas-Bôas, and Katya Ruggerio. CABI Publishing.

List of abbreviations

1,3BPG	1,3-bisphosphoglycerate
2K6PGn	2-keto-6-phosphogluconate
2KGn	2-ketogluconate
2PG	2-phosphoglycerate
3PG	3-phosphoglycerate
6PGn	6-phosphogluconate
6PGnL	6-phosphogluconolactone
Ala	Alanine
Alg	Alginate biosynthesis
AMDIS	Automated Mass Deconvolution and Identification System
Ana	Anaplerotic reactions
Arg	Arginine
Asn	Asparagine
Asp	Aspartate
CE	Capillary Electrophoresis
CF	Cystic Fibrosis
Chr	Chorismate
CI	Chemical Ionization
Cit	Citrate
CLE	Carbon Labeling Experiment
CTFR	Cystic Fibrosis Transmembrane Conductance Regulator
C-THF	one-carbon group carried by tetrahydrofolate (all oxidation states)
Cys	Cysteine
DHAP	Dihydroxyacetone phosphate
DOF	Degrees of Freedom
DRS	Deconvolution Reporting Software
E4P	Erythrose 4-phosphate
EC	Energy Charge
EDP	Entner-Doudoroff Pathway
EI	Electron Ionization
EMP	Embden-Meyerhof-Parnas Pathway
EMU	Elementary Metabolite Units
ESI	Electrospray
F1P	Fructose 1-phosphate
F6P	Fructose 6-phosphate
FBP	Fructose 1,6-bisphosphate
FT-ICR	Fourier Transform Ion Cyclotron Resonance
Fum	Fumarate
G	Guluronat
G6P	Glucose-6-phosphate
GA	Glyceraldehyde
GAP	Glyceraldehyde 3-phosphate
Gat	Glycerate
GC	Gas Chromatography
GDP-M	GDP-mannose
GDP-Mu	GDP-mannuronate
Gln	Glutamine
Glu	Glutamate
Glx	Glyoxylate
Gly	Glycine
Gn	Gluconate
Gol	Glycerol
Gol3P	Glycerol 3-phosphate
His	Histidine
HomoCys	Homocysteine
ICD	Isotope Coded Derivatization
ICit	Isocitrate
IDMS	Isotope Dilution Mass Spectrometry
Ile	Isoleucine
KDPG	2-keto-3-deoxy-6-phophogluconate
K_{eq}	Thermodynamic Equilibrium Constant
K_m	Michaels constant
LC	Liquid Chromatography
Leu	Leucine
LIT	Linear Iquadrupole ion trap
Lys	Lysine
M	Mannuronat
m/z	mass to charge ratio
M1P	Mannose 1-phosphate
M6P	Mannose 6-phosphate
Mal	Malate
MCF	Methyl Chloroformate
Met	Methionine
MFA	Metabolic Flux Analysis
MIRACLE	Mass Isotopomer Ratio Analysis of U-13C-Labelled Extracts
mM	Millimolar
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
MS/MS	tandem MS
MSTFA	N-methyl-N-trimethylsilyl trifluoroacetamide
MTBSTFA	N-methyl-N-(t-butyldimethylsilyl) trifluoroacetamide
NMR	Nucelar Magnetic Reseonance
OAA	Oxaloacetate
OGA	2-oxoglutarate
Orn	Ornithine
PCA	Principal Component Analysis
PCI	Positive Chemical Ionization
PEP	Phosphoenolpyruvate
Phe	Phenylalanine
PLS-R	Patial least square resgression
PPP	Pentose Phosphate Pathway
Pro	Proline
PRPP	Phosphoribosyl-1-pyrophosphate
PYR	Pyruvate
Q	Quadrupole
R5P	Ribose-5-phosphate
Ru5P	Ribulose-5-phosphate
S7P	Sedoheptulose 7-phosphate
SAH	S-adenosyl homocysteine
SAM	S-adenosyl methionine
Ser	Serine
Suc	Succinate
Suc-CoA	Succinyl-CoA
SVD	Singular Value Decomposition
TCA	Tricarboxylic acid cycle
Thr	Threonine
ToF	Time of Flight
Trp	Thryptophane
Tyr	Tyrosine
Val	Valine
X5P	Xylulose 5-phosphate

Table of contents

1 Introduction

The topic of this doctoral work is microbial mass spectrometry (MS) based metabolomics. The focus has been an investigation of the central carbon metabolism of the bacterium *Pseudomonas fluorescens* SBW25, a body of work that has led to two publications (Paper II and Paper III). An integrated part of the work has been developing and implementing methods for preparation and analysis of metabolite extracts, and a publication on some of this work is also a part of this thesis (Paper I). The present chapter is composed of three parts giving an introduction to (*i*) the biological system (*P. fluorescens*), and to the fields of (*ii*) metabolomics and (*iii*) fluxomics. These introductions will be given in chapters 1.1, 1.2 and 1.3 respectively. Subchapter 1.3.1 gives a somewhat detailed introduction to the simulation software 13CFLUX2 used for fluxomics in Paper III. The text is an advantage, but not a prerequisite for evaluating results and biological interpretations in Paper III.

1.1 Introduction to Pseudomonas

Pseudomonas is a bacterial genus of agricultural, industrial and medical importance. For this reason the metabolism of the strain *P. fluorescens* SBW25 has been studied in this doctoral work. The present chapter will describe the central carbon metabolism of *P. fluorescens* and go into the relevance of *Pseudomonas* species for humans.

1.1.1 Nutrient uptake, metabolism and alginate production

Pseudomonas species are gram negative straight or slightly curved rods 0.5 μm – 1.0 μm in diameter by 1.5 μm to 5.0 μm in length motile by one or several polar flagella. Most species are obligate aerobes that utilize oxygen as the terminal electron acceptor, but some can also utilize nitrate (Palleroni and Moore 2004). *Pseudomonas* species are versatile bacteria capable of utilizing a large number of carbon sources and colonizing a wide variety of ecological niches including soil, water and the surface of plants and animals. *P. fluorescens* SBW25, the strain studied in this thesis, was isolated in 1989 from the leaf surface of a sugar beet plant at the University Farm, Wytham, Oxford, UK (Rainey and Bailey 1996) and its genome was sequenced in 2009 (Silby, Cerdeno-Tarraga et al. 2009). The species name originates from the production of a soluble,

greenish fluorescent pigment, produced especially under conditions of low iron availability.

The central carbon metabolism (Embden – Meyerhof – Parnas pathway / glycolysis (EMP), Entner - Doudoroff pathway (EDP), pentose phosphate pathway (PPP), tricarboxylic acid cycle including the glyoxylate shunt (TCA) and the anaplerotic reactions (Ana)), the biosynthetic route for alginate production (Alg), and routes for glucose, fructose and glycerol uptake in *P. fluorescens* is shown in Figure 1, in addition to amino acids connected to their precursor metabolites (only species transferring carbon are shown). *P. fluorescens*, as other *Pseudomonas* species, does not utilize glycolysis as they lack the phosphofructokinase activity required for complete glycolytic oxidation of glucose. Instead they use the EDP as the main catabolic route (Lessie and Phibbs 1984). Glucose is an often used carbon source in laboratory cultivations. Glucose uptake for *P. fluorescens*, as uptake of most carbohydrates, proceeds via active transport through an ATP-binding cassette (ABC) uptake system. Intracellular glucose is then phosphorylated to glucose 6-phosphate (G6P), which is glucose's entry point into central carbon metabolism. Alternatively, extracellular glucose can be oxidized to gluconate (Gn) and further to 2-ketogluconate (2KGn) in the periplasmic space, before both Gn and 2KGn are transported into the cell. Once inside the cell, both Gn and 2KGn are converted to 6-phosphogluconate (6PGn) by phosphorylation or phosphorylation and reduction, respectively (del Castillo, Ramos et al. 2007). In contrast to glucose, fructose, one of the two carbon sources utilized in this doctoral work, is imported by a phosphoenolpyruvate (PEP) – dependent phosphotransferase, leading to the formation of fructose 1-phosphate (F1P). F1P is subsequently phosphorylated to fructose 1,6-bisphosphate (FBP) (Durham and Phibbs 1982; Lessie and Phibbs 1984), or alternatively cleaved into glyceraldehyde (GA) and dihydroxyacetone phosphate (DHAP). Glycerol, the other of the two carbon sources utilized in this doctoral work, is taken up through facilitated diffusion. Once internalized, glycerol is phosphorylated to glycerol 3-phosphate (Gol3P) (Schweizer, Jump et al. 1997), which is then further oxidized to DHAP.

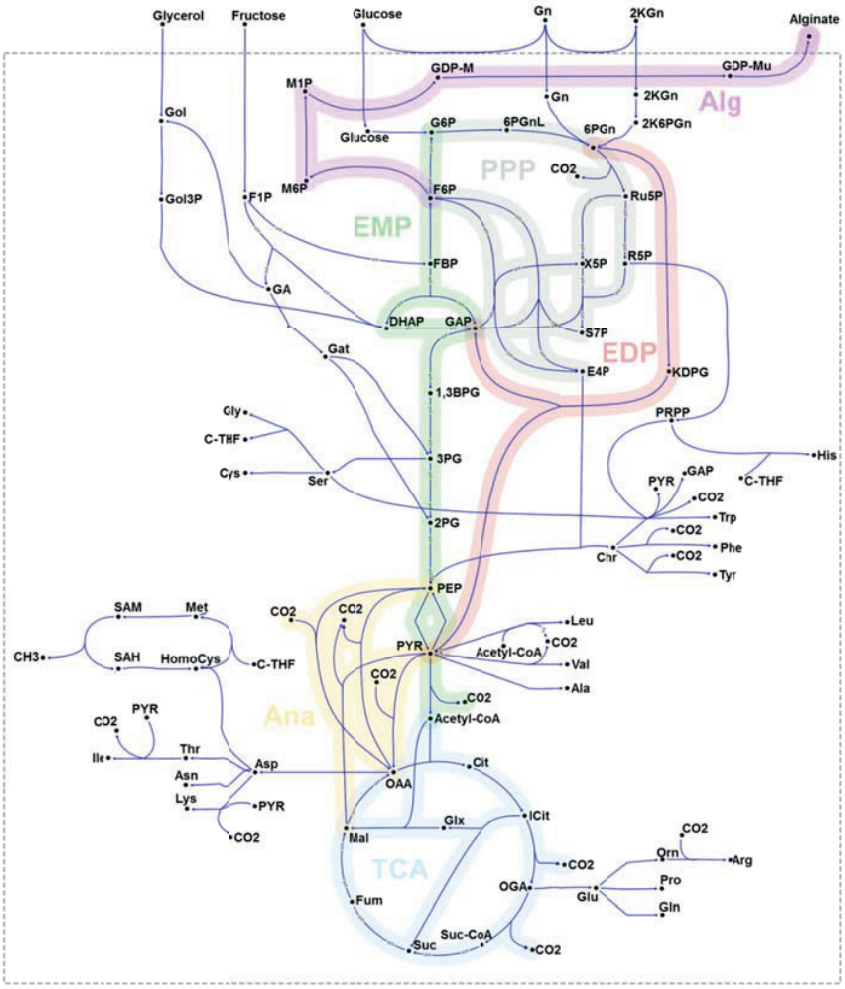


Figure 1: Central carbon metabolism of *P. fluorescens* portraying the Embden – Meyerhof - Parnas pathway / glycolysis (EMP), the Entner – Doudoroff pathway (EDP), the pentose phosphate pathway (PPP), the tricarboxylic acid cycle, the anaplerotic reactions (Ana) and alginate biosynthesis (Alg) in addition to routes of glucose, fructose and glycerol uptake. Amino acids are also shown connected to their precursor metabolites. Only carbons transferring reactions are shown. **Metabolite abbreviations: 1,3BPG: 1,3-bisphosphoglycerate; 2K6PGn: 2-keto-6-phosphogluconate; 2KGn: 2-ketogluconate; 2PG: 2-phosphoglycerate; 3PG: 3-phosphoglycerate; 6PGn: 6-phosphogluconate; 6PGnL: 6-phosphogluconolactone; Ala: Alanine; Arg: Arginine; Asn: Asparagine; Asp: Aspartate; Chr: Chorismate; Cit: Citrate; C-THF: one-carbon group carried by tetrahydrofolate (all oxidation states); Cys: Cysteine; DHAP: Dihydroxyacetonephosphate; E4P: Erythrose 4-phosphate; F1P: Fructose 1-phosphate; F6P: Fructose 6-phosphate; FBP: Fructose 1:6-bisphosphate; Fum: Fumarate; G6P: Glucose-6-phosphate; GA: Glyceraldehyde; GAP: Glyceraldehyde 3-phosphate; Gat: Glycerate; GDP-M: GDP-mannose; GDP-Mu: GDP-mannuronate; Gln: Glutamine; Glu: Glutamate; Glx: Glyoxylate; Gly: Glycine; Gn: Gluconate; Gol: Glycerol; Gol3P: Glycerol 3-phosphate; His: Histidine; HomoCys: Homocysteine; ICit: Isocitrate; Ile: Isoleucine; KDPG: 2-keto-3-deoxy-6-phophogluconate; Leu: Leucine; Lys: Lysine; M1P: Mannose 1-phosphate; M6P: Mannose 6-phosphate; Mal: Malate; Met: Methionine; OAA: Oxaloacetate; OGA: 2-oxoglutarate; Orn: Ornithine; PEP: Phosphoenolpyruvate; Phe: Phenylalanine; Pro: Proline; PRPP: Phosphoribosyl-1-pyrophosphate; PYR: Pyruvate; R5P: Ribose-5-phosphate; Ru5P: Ribulose-5-phosphate; S7P: Sedoheptulose 7-phosphate; SAH: S-adenosyl homocysteine; SAM: S-adenosyl methionine; Ser: Serine; Suc: Succinate; Suc-CoA: Succinyl-CoA; Thr: Threonine; Trp: Thryptophane; Tyr: Tyrosine; Val: Valine; X5P: Xylulose 5-phosphate.**

Although not expressed by most strains in their natural habitat, many *Pseudomonas* species harbor the biosynthetic genes necessary for alginate synthesis. The medical and industrial relevance of the polysaccharide alginate will be returned to later in the chapter, but its biosynthesis will be covered here. A review of alginate biosynthesis by *Pseudomonads* is given by Jain and Ohman (2004a). The precursor for bacterial alginate synthesis in the central carbon metabolism is fructose 6-phosphate (F6P). Alginate synthesis starts with its conversion to mannose 6-phosphate (M6P) by AlgA. M6P is then converted to mannose 1-phoshate (M1P) by AlgC, before the bifunctional enzyme AlaA converts it into GDP-mannose (GDP-M). GDP-M is then oxidized by AlgD to form GDP-mannuronic acid (GDP-Mu) utilizing two molecules of NAD^+. GDP-Mu is the monomer used for polymerization and it is simultaneously polymerized to mannuronan through β-1,4 linkages and secreted across the inner membrane, presumably by Alg8. In the periplasm AlgG is responsible for epimerization of mannuronic acid (M) residues at C5 to form guluronic acid (G) residues. Unlike epimerases of another alginate producer, *A. vinelandii*, AlgG is not able to produce consecutive G residues (so called G-blocks). In addition to epimerization, the M residues can be acetylated in the periplasm at the O-2 or O-3 position through the actions of AlgF, AlgI and AlgJ. The last step in alginate biosynthesis is its secretion through the outer membrane of the bacterial cell through a porin like protein called AlgE. The periplasmic proteins AlgX, Alg44 and AlgK have been found to be necessary for efficient alginate production, but their functional role has not been firmly established. In addition the periplasmic alginate lyase AlgL has also been found necessary for efficient alginate production, and its role is most likely to control the length of the alginate chain, to detach adherent bacteria from the surface and to clear alginate remains in the periplasm. A simplified portray of alginate synthesis is given at the top of Figure 1. With the exception of *algC*, all of the thirteen genes for the enzymes described above are contained in a single operon (the *alg* operon) as shown in Figure 2. The *alg* operon is under the control of sigma factor AlgU.

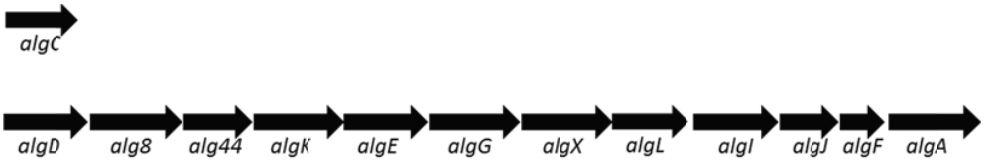


Figure 2: *algC* and the *alg* operon for alginate biosynthesis shown separately. The orientation and relative order of the genes in the *alg* operon are the same for all strains that have been investigated.

1.1.2 Pseudomonas and interactions with its surroundings

As stated previously, *Pseudomonas* species are capable of colonizing a variety of ecological niches. *P. fluorescens* and *P. putida* strains have been shown to be directly beneficial to plants by promoting plant growth and health, or indirectly beneficial by inhibiting or competing with pathogens, parasites and plant competitors. These properties enable use of *Pseudomonas* strains as biocontrol agents and as plant commensals (Harris and Stahlman 1996; Ellis, Timms-Wilson et al. 2000; Patten and Glick 2002; Haas and Keel 2003). The biosynthetic and metabolic diversity of *Pseudomonads* also opens for applications in bioremediation as they can remove or detoxify unwanted chemicals in the environment, e. g. in connection to oils spills (Desai and Banat 1997; Khan and Ahmad 2006).

As a contrast to these beneficial roles of *Pseudomonas* species, *P. syringae* is a plant pathogen (Volko, Boller et al. 1998) and *P. aeruginosa* is an opportunistic plant, insect and mammal pathogen (Elrod and Braun 1941; Elrod and Braun 1942; Bucher and Stephens 1957). Human *P. aeruginosa* infections are typically associated with burn and trauma wounds, immunocompromising diseases such as AIDS, chemotherapeutic treatments that negatively affect leukocyte number and function, and chronic lung infections in cystic fibrosis (CF) patients. CF is an autosomal recessive genetic disorder caused by a malfunctional cystic fibrosis transmembrane conductance regulator (CFTR). Although studies on mice have showed that CFTR expression in the airway epithelium is crucial for functional innate immunity to *P. aeruginosa* lung infections, there is currently no consensus concerning the molecular and cellular basis for pathogenesis and host resistance (Pier and Goldberg 2004). Early colonization of the CF

lung by *P. aeruginosa* can be cleared by aggressive antibiotic therapy, but after conversion to a mucoid phenotype, associated with overproduction of the polysaccharide alginate, the organism can no longer be eliminated (Frederiksen, Koch et al. 1997).

Alginate is an important factor in colonization by *P. aeruginosa* in the CF lung as about 80 % of *P. aeruginosa* isolates from CF patients are alginate overproducers, whilst only 1 % of other clinical *P .aeruginosa* isolates have this phenotype. In pulmonary infections alginate confers increased resistance to phagocytosis and is disadvantageous to various aspects of the host immune response. Alginate also seems to be part of an adherence mechanism for the bacterium, and it contributes to the extremely viscous nutrient rich environment in the CF lung. In addition, the mucoid phenotype of *P. aeruginosa* correlates with a decrease in the production of various other virulence factors, thereby indirectly decreasing the host immune response. This down-regulation indicate a multisystem involvement in alginate regulation (Jain and Ohman 2004b). Alginate is frequently a component of *P. aeruginosa* biofilms, but is not found in *P. putida* and *P. fluorescens* biofilms both seemingly composed more of a bacterial cellulose-like exopolymer (Tolker-Nielsen and Molin 2004).

1.1.3 The sigma factor AlgU and its anti-sigma factor MucA

As stated previously, *P. fluorescens* wild type does not produce alginate, but does as *P. aeruginosa* harbor all biosynthetic genes necessary for alginate production. All of these genes, apart from *algC* are clustered in the *alg* operon as shown in Figure 2 (Darzins, Wang et al. 1985; Morea, Mathee et al. 2001). The sigma factor AlgU (also known as AlgT or σ^{22}) is essential for transcription of *algC* and the *alg* operon, but AlgU is normally bound to the cytoplasmic domain of MucA. MucA is an inner membrane protein with one transmembrane domain, and binding of AlgU by MucA prevents transcription of the alginate biosynthetic genes. MucA thus act as an anti-sigma factor. The environmental signals that potentially release AlgU from MucA are not known. In *P. aeruginosa* conversion to a mucoid phenotype is frequently caused by a mutation in the *mucA* gene (Schurr, Yu et al. 1996; Mathee, McPherson et al. 1997), and alginate

biosynthesis has also been established in *P. fluorescens* by inactivation of MucA (Borgos, Bordel et al. 2013). Other regulatory mechanisms of alginate production include the negative regulator MucB and the positive regulators AlgR, AlgZ and AlgB (Baynham and Wozniak 1996; Ma, Selvaraj et al. 1998).

The *algU* gene is part of an operon with *mucA* and three other genes (*mucB-mucC-mucD*, in earlier litterature referred to as *algN-algM-algY*) (Ohman, Mathee et al. 1996). The sigma factor is part of the extra cytoplasmic function (ECF) family that respond to membrane stresses (Missiakas and Raina 1998). In *A. vinelandii* this cluster is, in addition to its involvement in alginate production, associated with encystment in a stress response elicited by unfavorable environments such as dry soil (Martinez-Salazar, Moreno et al. 1996). The equivalent gene cluster in *Escherichia coli* is missing the *mucD* equivalent, but the *algU* equivalent σ^E codes for an extreme heat shock sigma factor (Erickson and Gross 1989) and it seems to have a broad role as indicated by its response to envelope folding defects (Pogliano, Lynch et al. 1997).

AlgU controls several genes other than those required for alginate biosynthesis in *Pseudomonas*. For *P. aeruginosa* an unregulated AlgU has been shown to lead to elevated transcription of a disulphide bond isomerase (*dsbA*). The disulphide bond isomerase was necessary for maintaining periplasmic alkaline phosphatase activity, necessary to secure secretion of stable elastase and necessary for efficient twitching motility, but not necessary for alginate production (Malhotra, Silo-Suh et al. 2000). The mucoid phenotype of *P. aeruginosa* has also been shown to have numerous other genes up-regulated (Firoved and Deretic 2003) and *mucA*- inactivation in *P. fluorescens* has been shown to effect numerous genes including genes involved in energy generation and genes encoding ribosomal and other translation related-proteins (Borgos, Bordel et al. 2013). This clearly shows that inactivation of MucA has pleiotropic effects, not yet fully understood.

1.1.4 Properties and applications of alginates

The properties and application of alginates have been thoroughly reviewed by Skjåk-Bræk and Draget (2012), and a summary is given below. Alginates are a family of linear binary copolymers consisting of (1→4)-linked β-D-mannuronic acid (M) and α-L-guluronic acid (G) residues. They are block polymers composed of a combination of consecutive M-residues (M-blocks), consecutive G-residues (G-blocks) or blocks rich in MG dimers (MG-blocks). For any specific alginate the relative fractions of these blocks determine many of its physical and chemical properties. Commercially available alginate is currently prepared from marine brown seaweeds (*Phaeophyceae*), where it exists as a gel in the intracellular matrix conferring elasticity and physical strength. In nature alginate is also synthesized by species of bacteria in the genera *Pseudomonas* and *Azotobacter*. Harvesting alginate from algal sources is cost-efficient because natural resources can be utilized, but it has the drawback of geographical and seasonal variations. Alternatively consistent and tailor-made production could be achieved through a bacterial fermentation process using genetically modified organisms and/or *in vitro* enzymatic modification of bacterial alginate. The main difference in alginates from algal and bacterial sources is that bacterial alginate can contain O-acetyl groups on C-2 and C-3 of M-residues, and the main difference between alginates from *Pseudomonas* and *Azotobacter* is that *Pseudomonas* strains are incapable of producing alginate containing G-blocks.

According to Skjåk-Bræk and Draget (2012) alginate has numerous technical applications. Being water soluble and inert to most dyes it is an important thickener in textile printing and its film-forming properties are used in sizing/coating of paper. Alginate is also used to produce high-quality homogenous welding rods with improved performance. Traditional medical applications include use as drug delivery systems by e.g. protecting fragile colon-targeted drugs from gastric acid or to enable controlled delayed release. Alginate has also been used in wound dressings, as dental impression material and in formulations treating heartburn and gastric reflux.

In addition there are many new areas of application currently under investigation. Because alginate is capable of forming gels in the presence of bivalent cations, it is possible to create gel sphere with entrapped living cells. This can be done by mixing a cell suspension with an alginate solution and then dripping this solution into water containing bivalent cations (typically Ca^{2+}). Such immobilization of living cells has numerous applications in industry, agriculture and medicine. It can be used in ethanol production by yeast, in mass production of artificial seeds by entrapment of plant embryos and in monoclonal antibody production by entrapped hybridoma cells. In medicine, transplantation of alginate-encapsulated cells could be used in the treatment of several diseases, e.g. encapsulated adrenal chromaffin cells could be used in the treatment of Parkinson's disease, encapsulated hepatocytes could be used in the treatment of liver failure, encapsulated parathyroid cells could be used in the treatment of hypocalcemia, encapsulated Langerhans islets could be used to treat Type I diabetes and genetically altered cells could be used in the treatment of various cancers. The main purpose of the alginate capsule in such transplantations would be to protect the transplanted cells from the patient's immune response. Such protection would minimize the need to administer immunosuppressives, which would be beneficial to the patients' health. Yet another potential use for alginate is as scaffolds in tissue engineering where they could function as space-filling agents, as delivery agents for bioactive molecules or to aid organization of cells to direct tissue formation. All of these new areas of applications require optimization of alginate for the specific purpose conferring appropriate mechanical and chemical stability and managing swelling, level of contaminants, pore size and pore size distribution.

The industrial and medical relevance of *Pseudomonas* and alginate warrants the quest for a more thorough understanding of the regulation and function of the metabolism of these bacteria. For this reason *P. fluorescens* SBW25 was chosen for investigation through MS based microbial metabolomics in this doctoral work.

1.2 Microbial metabolomics – quantifying and qualifying metabolite composition of microorganisms.

Microbial metabolomics is the comprehensive study of the presence and concentrations of metabolites in the metabolic network of a microorganism. Research in this field can be seen as being comprised of three successive steps: (*i*) sample preparation, (*ii*) data acquisition through sample analysis and (*iii*) data processing and interpretation (Figure 3). These three steps will be covered separately in in sections 1.2.1, 1.2.2 and 1.2.3, respectively. The last two steps, sample analysis and data processing and interpretation are not necessarily different for microbial systems and plant or mammalian systems, but the first step, sample preparation, will be fundamentally different depending on the type of organism or tissue being investigated (Mashego, Rumbold et al. 2007). Since the biological system in this thesis is the bacterium *P. fluorescens*, the following will focus on microbial metabolomics for bacterial microorganisms.

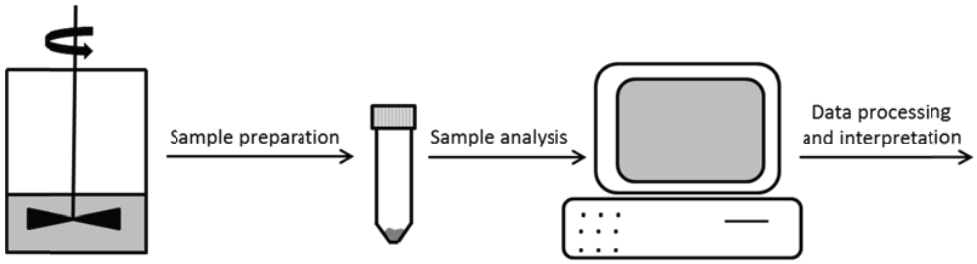


Figure 3: Overview of sequential workflow in microbial metabolomics research going from sample preparation to data acquisition and finally to data processing and interpretation.

1.2.1 Sample preparation

The process of sample preparation for bacterial metabolomics consist of removing bacteria from their growth environment and quenching the bacterial metabolism, extracting metabolites from the bacterial cells and usually some further processing such as removal of cell debris and drying of the extract (Figure 4).

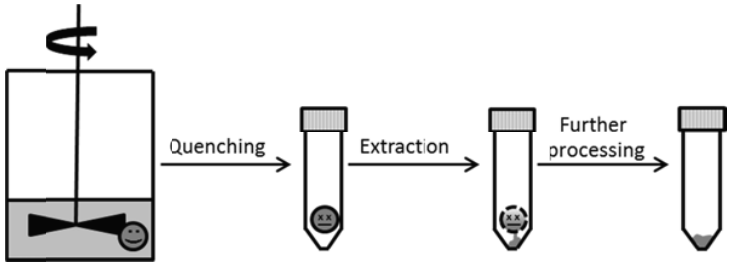


Figure 4: Overview of sample preparation of bacterial extracts going from quenching and extraction to purification.

Quenching of bacterial metabolism

Ideally the metabolism of the bacterium should be frozen (i.e. inactivated) in time at the exact moment the bacterium is removed from its growth environment (i.e. the shake flask or the fermenter). This involves arresting all enzymatic reactions and avoiding all spontaneous reactions. Typically this is done by subjecting the bacteria to temperature extremes (typically below -40 °C and above 60 °C) and/or pH extremes (typically below 2.0 and above 10.0). Using pH extremes to quench metabolism is usually avoided if mass spectrometric (MS) sample analysis is planned as the presence of ions and salts deteriorates the quality of MS analysis. The most prevalent quenching approach used for a variety of bacterial strains is cold methanol quenching where the culture sample is transferred directly into a cold methanol-water solution (Jensen, Jokumsen et al. 1999; Schaefer, Boos et al. 1999; Wittmann, Kromer et al. 2004; Winder, Dunn et al. 2008). Typically 60 % (v/v) methanol in water at -40 °C is used with a 1:3 ratio of sample and quenching solution. Other quenching solutions in use include perchloric acid at ambient and cold temperatures, hot sodium hydroxide, cold ethanol, liquid nitrogen and cold sodium chloride (Cook, Urban et al. 1976; Weuster-Botz 1997; Letisse and Lindley 2000; Buziol, Bashir et al. 2002; Wittmann, Kromer et al. 2004), but some of these are, as mentioned above, not compatible with subsequent MS analysis.

Numerous studies have shown that subjecting bacteria to any quenching solution inevitably cause metabolite leakage from the bacterial cell (Wittmann, Kromer et al. 2004; Winder, Dunn et al. 2008; Taymaz-Nikerel, de Mey et al. 2009). Because of this, measurements of intracellular metabolite concentrations from the biomass of a sample

directly transferred into the quenching solution do not accurately reflect the metabolite concentrations of cells in the growing culture. One solution to this problem is using a differential method were two separate metabolite extracts are prepared: one from the entire culture and one from the extracellular medium. These two samples are then used to calculate intracellular metabolite concentration (Taymaz-Nikerel, de Mey et al. 2009). Unfortunately such calculations for certain metabolites suffer from large standard deviations as total culture and extracellular concentrations are much larger than intracellular concentrations (per cell dry weight).

An alternative to the differential method is to separate the bacteria and the medium prior to quenching through centrifugation or quick vacuum filtration. Centrifugation is often not applicable as the g-force required to sediment cells will produce metabolite leakage and the length of time required will lead to changes in metabolism. Time is also an issue when it comes to filtration as filtration time can be a problem for metabolites with quick turnover times. Another important issue for both centrifugation and filtration is that the cell pellet or the cells on the filter need to be washed to remove remains of medium from the cells before transfer into a quenching solution. Washing can be achieved by using a sodium chloride solution. Because both hypoosmotic and cold solutions induce metabolite leakage, a solution slightly hyperosmotic to the culture medium and at the same temperature as the medium should be used (Britten and McClure 1962; Smeaton and Elliott 1967; Wittmann, Kromer et al. 2004; Bolten, Kiefer et al. 2007). Although washing cells on the filter with a sodium chloride solution can result in some leakage of metabolites, the leakage is significantly lower than the leakage produced when washing quenched cells with the same solution (Shin, Lee et al. 2010).

Extraction of metabolites from the bacterial cell

Once metabolism in quenched, metabolites need to be extracted from the bacterial cells and the best method for doing this depends on the metabolites of interest: whether the focus of the study is the entire metabolome or a specific class of metabolites. For polar and mid-polar metabolites, cold methanol extraction is well suited, whilst methanol-water-chloroform extraction is more appropriate if nonpolar metabolites are also of

interest. To extract polar metabolites boiling ethanol can be used, but one has to be aware of the possibility of degradation of thermo labile metabolites and oxidation of reduced metabolites. Acidic or alkaline extraction is also a possibility for polar metabolites, but these methods often suffer from poor metabolite recovery, can cause hydrolysis of proteins and other polymers and in addition are not compatible with MS-analysis (Villas-Bôas 2007).

Because quenching solutions inevitably cause metabolite leakage, quenching of metabolism and metabolite extraction is often performed in the same solution. When this is done using cold methanol, multiple freeze-thaw cycles are often employed to disrupt cell integrity and to aid complete extraction of metabolites (Winder, Dunn et al. 2008; Wentzel, Sletta et al. 2012). Other ways to achieve complete extraction include the use of microwave for thermo stable metabolites and the use of a French press if the number of samples to be processed is not to large (Villas-Bôas 2007).

Further processing

After extraction of metabolites from the bacteria into the surrounding solution, further processing often involves centrifugation of the sample to remove denaturized proteins and cell debris. Subsequently, for both analytical and storage purposes, it is common to concentrate or dry the metabolite extract. To avoid degradation of fragile metabolites freeze-drying is often employed, or alternatively rotary evaporation under vacuum is used. Sometimes solid phase extraction is used in combination with rotary evaporation to preserve thermo labile metabolites (Letisse and Lindley 2000).

Choosing a sample preparation protocol

When preparing bacterial metabolite extracts the best choice of quenching and extraction solution and other parameters of sample preparation clearly depend on the specific bacterium being investigated. Bacterial structure, especially cell wall and membrane composition, will greatly affect the effectiveness of any protocol. Optimized sample preparation procedures exist for a number of bacteria including gram-negative *Escherichia coli* (Winder, Dunn et al. 2008; Taymaz-Nikerel, de Mey et al. 2009),

Saccharophagus degradans (Shin, Lee et al. 2010) and *Xanthomonas campestris* (Letisse and Lindley 2000) and gram-positive *Staphylococcus aureus* (Meyer, Liebeke et al. 2010), *Lactobacillus plantarum* (Faijes, Mars et al. 2007) and *Corynebacterium glutamicum* (Wittmann, Kromer et al. 2004).

Figure 5 gives an overview of the experimental setup used by (A) Winder, Dunn et al. (2008) and (B) Taymaz-Nikerel, de Mey et al. (2009) to arrive at optimized sample preparation protocols for *E. coli*. The optimized protocols are shown as dark grey boxes, whilst light grey boxes show control samples and alternative protocols tested. In optimized protocol A a metabolite extract is prepared from the culture biomass by cold 60% methanol quenching of the culture with subsequent centrifugation to collect the cell pellet. Extraction of metabolites is then done in cold 100% methanol and the resulting intracellular concentrations are corrected for leakage during quenching using whole culture and culture medium control samples. Optimized protocol B proposes a differential method for *E. coli* where one whole culture extract and one culture medium extract is prepared through cold 60% methanol quenching and hot 75% ethanol extraction. For protocol B one would then obtain intracellular concentrations by calculating the difference between the whole cell extract and the culture medium extract. As can be seen by studying Figure 5, no experimental path during optimization of protocol A overlaps with an experimental path during optimization of protocol B. The same holds true for the experimental paths undertaken arriving at the optimized protocols in Figure 6 for (A) *S. degradans*, (B) *X. campestris*, (C) *S. aureus*, (D) *L. plantarum* and (E) *C. glutamicum*. This lack of overlap in method development strategies exemplifies a challenge in microbial metabolomics: lack of standardized protocols makes it difficult to collectively evaluate the results from different research groups. This complicates the process of choosing one out of several established protocols for a bacterium or drawing on published results for various bacteria when establishing a protocol from a new bacterium. Currently there is some research published comparing protocols for several bacteria (Bolten, Kiefer et al. 2007; Spura, Christian Reimer et al. 2009), but these are as of yet inconclusive when it comes to specifying an optimal protocol for one specific bacterium. Hopefully standardization

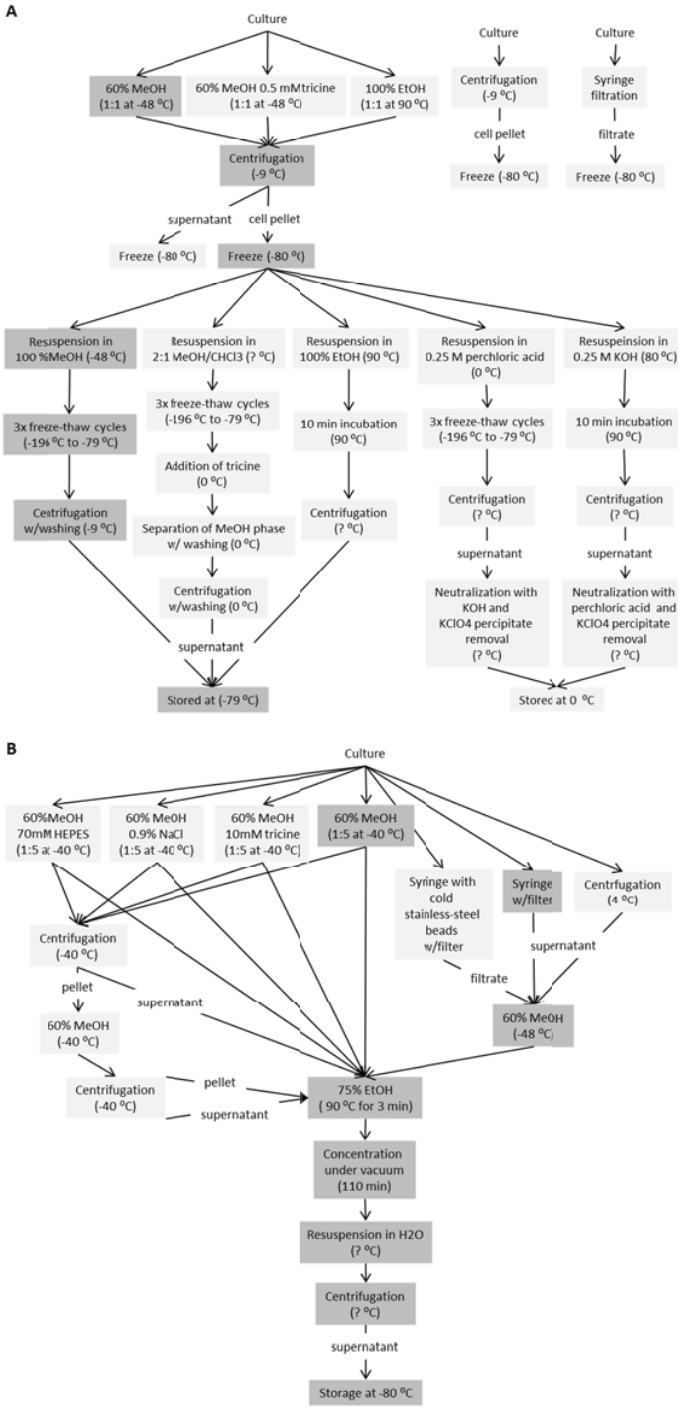


Figure 5: Overview of experimental setup for sample preparation optimization for *E. coli* by (A) Winder, Dunn et al. (2008) and (B) Taymaz-Nikerel, de Mey et al. (2009). Dark grey boxes: optimized protocol; light grey boxes: alternatives tested and control samples.

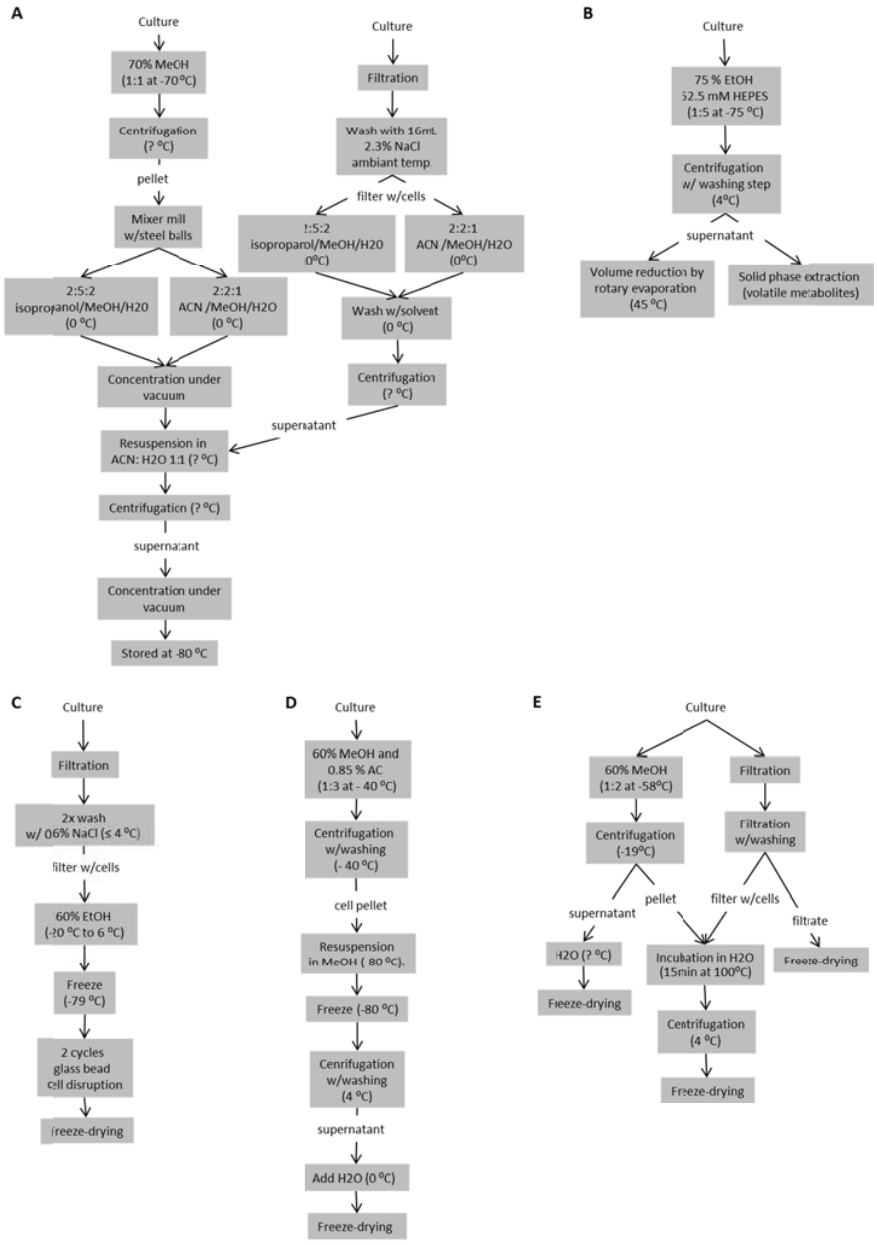


Figure 6: Optimized sample preparation protocol for gram-negative (A) *S. degradans* and (B) *X. campestris* and gram-positive (C) *S. aureus*, (D) *L. plantarum* and (E) *C. glutamicum*.

initiatives such as The Metabolomics Standards Initiative (MSI) (Fiehn, Kristal et al. 2006) will provide fruitful solutions to problems such as these in the future.

Another important issue when choosing a sample preparation protocol is to adapt it to the planned sample analysis and data processing. This involves incorporation of appropriate internal standards and appropriate control samples. These issues will be covered in chapters 1.2.2 and 1.2.3.

1.2.2 Data acquisition through sample analysis

Sample analysis approaches for metabolite extracts can be classified into two categories: non-targeted and targeted approaches. In a non-targeted approach as many metabolites as possible are detected, only limited by the analytical technique employed. Typically metabolites are not identified in the raw dataset, but at least a subset of the detected entities are usually identified during subsequent data processing and interpretation. In contrast to this, a targeted approach aims at quantitation of known metabolites. These metabolites are either defined by specific parameters of the analytical instrument or by settings of the software used in immediate processing of the raw data. Often absolute quantitative concentrations of metabolites are calculated from data in targeted approaches using calibration curves constructed from analyzed standards.

The two most commonly applied analytical techniques used for analysis of metabolite extracts are nuclear magnetic resonance (NMR) and mass spectrometry (MS) (Mashego, Rumbold et al. 2007). The main advantage of NMR over MS is its non-invasive nature, as in NMR-analysis as opposed to MS-analysis samples are not inherently consumed. The main advantage of MS-analysis over NMR-analysis is higher sensitivity, a feature often crucial to detect metabolites of low abundance. MS-techniques are currently dominating microbial metabolomics (Bruheim, Kvitvang et al. 2013), and as mass spectrometry based microbial metabolomics is the topic of this thesis, the following will focus on aspects of MS instrumentation.

Mass Spectrometry

A MS instrument needs to perform three distinctive tasks: volatile ions need to be created from the analytes in the sample, the created ions need to be separated based on their mass to charge ratio (m/z) and finally the created ions need to be detected. These three tasks are performed by the ion source, the mass analyzer and the detector of a MS instrument, respectively. Several different ion sources, mass analyzers and detectors have been developed, each with their own set of strengths and weaknesses (Gross 2004). Typically the choice of ion source and mass analyzer depends on the analytical task at hand, whilst the choice of detector is dictated by the type of mass analyzer.

The volatile ions created in the ion source are called molecular or quasimolecular ions if the analyte remains intact, but ionization can also lead to fragmentation of the analyte into a fragment ion and a neutral. Common ion sources in use include electron ionization (EI), chemical ionization (CI) and electrospray ionization (ESI) ion sources. The most distinctive feature of these ion sources is the degree to which they fragment the analyte upon ionization. CI and ESI are so-called soft ionization techniques that to a large degree leave the molecular ion intact, whilst EI, a so-called hard ionization technique to a larger extent produce fragments of the molecular ion. In non-targeted metabolomics an intact molecular ion can be used for identification purposes, if mass accuracy of the measurement is good enough. Identification can be done by using the accurate mass measurement of an unknown molecular ion to estimate its molecular formula: e.g. carbon dioxide and ethene both have a nominal mass of 28 u, but the monoisotopic mass with four significant figures is 27.99 u for CO and 28.03 u for C_2H_4, thus these two and other compounds with the same nominal mass can be separated by accurate mass measurements. EI ion sources also have an application in non-targeted metabolomics. A specific compound fragments in a distinct way depending on its chemical structure, producing a specific mass spectrum of relative intensities versus m/z for the specific compound. Because mass spectra produced by EI are very reproducible, large searchable databases of EI spectra have been constructed (e.g. publicly available NIST/EPA/NIH Mass Spectral Library which contains EI mass spectra of approximately 213,000 compounds). These databases can be used as a starting point for

identification of unknown compounds. Choice of ion source is also important in targeted metabolomics. Soft ionization techniques can be better for avoiding false positives in detection of compounds, as similar compounds will produce many similar fragment ions, although hard ionization techniques are sometimes necessary to achieve adequate sensitivity, as degree of ionization (the portion of sample ionized) is often higher for these techniques.

The main component of a mass spectrometer is the mass analyzer. In Table 1 the mass analyzers currently dominating in mass spectrometry of biological samples are listed, along with key performance characteristics. Although somewhat dated as analyzers are continuously being improved, the values in Table 1 still reflect the relative strengths and weaknesses of various mass analyzers more precisely than isolated values provided by specific vendors. In Table 1 linear dynamic range is the m/z range for which the analyzer can yield a linear response, mass accuracy is the difference between true and measured m/z divided by true m/z, resolution is the full width at half-high of a single well resolved peak divided by the mass associated with that peak, abundance sensitivity is a measurement of the signal to noise ratio and precision is the reproducibility of a single measurement (McLuckey and Wells 2001).

Table 1: Selection of mass analyzers with key performance characteristics indicated. Values obtained from McLuckey and Wells (2001) and Hu, Noll et al. (2005)[1)]

Type of mass analyzer	Linear dynamic range	Mass accuracy	Resolution	Abundance sensitivity	Precision
Sector field	10^9	1-5 ppm	10^2-10^5	10^6-10^9	0.01-1%
Quadrupole (Q)	10^7	100 ppm	10^2-10^4	10^4-10^6	0.1-5%
Time-of-Flight (ToF)	10^2-10^6	5-50 ppm	10^3-10^4	10^6	0.1-1%
Fourier Transform Ion Cyclotron Resonance (FT-ICR)	10^2-10^5	1-5 ppm	10^4-10^6	10^2-10^5	0.3-5%
Quadropole Ion Trap	10^2-10^5	50-100 ppm	10^3-10^4	10^3	0.2-5%
Orbitrap[1)]	10^2-10^3	2-5 ppm	10^4-10^6	-	-

The mass analyzers intended use will determine the relative importance of the various performance characteristics in Table 1. In non-targeted metabolite analysis good accuracy and high resolution is paramount to resolve complex samples, even with good

chromatographic separation prior to MS analysis. Because of this the mass analyzer of choice is often the Time-of-Flight (ToF) analyzer, and more recently the Orbitrap. Although both sector field instruments and Fourier Transform Ion Cyclotron Resonance (FT-ICR) instruments have better mass accuracy and higher resolution, they are often not employed because of their large space requirements and high cost. In targeted metabolite analysis, often requiring accurate quantitative data, the quadrupole analyzer (Q) is often chosen because of its high linear dynamic range and its good sensitivity. Sector field instruments, although they have an excellent dynamic range, again lose ground because of their large space requirements and their high cost.

In addition to instruments with a single mass analyzer, instruments with two or more mass analyzers in series are also in widespread use. In addition to drawing on the strengths of several mass analyzer, such tandem MS or MS/MS instruments often offer higher sensitivity and better selectivity as noise is filtered out in the additional mass selection step and ions more specific for a certain analyte can be chosen from two consecutive fragmentation steps than from a single one. Using a MS/MS instrument a compound is typically detected by recording a specific precursor-to-fragment transition (also known by the vendor specific term multiple reaction monitoring (MRM) – transition). During the last decade ion mobility analyzers, separating ions based on ion cross section rather than their m/z ratio, have also become more common, both in stand-alone instruments and as an integral part of regular MS-instruments (Kanu, Dwivedi et al. 2008).

Liquid chromatography and gas chromatography

One important issue has not yet been addressed: due to the complexity of metabolome samples MS detection in metabolomics is often preceded by a sample separation step. Liquid chromatography (LC) and gas chromatography (GC) are in most widespread use, but capillary electrophoreses (CE) is also employed. For targeted approaches a separation step is usually employed, whilst for non-targeted approaches this step is sometimes omitted. As a specific LC, GC or CE method will be most suited for specific classes of compounds, whilst other compound classes will not be detected, an analytical

method will not be truly non-targeted if a separation step is employed. Also because of the specificity of separation methods, several distinct analytical methods are often employed when using a targeted approach to cover a broad range of metabolite classes. Comparing LC and GC the main practical difference between them is that GC requires a derivatization step for most compounds to render them thermo stabile and volatile, whilst this is not required for LC. A host of different derivatization protocols exist suited for specific classes of metabolites. A thorough introduction to LC, GC and derivatization for GC in connection to microbial metabolomics can be found in Appendix V.

1.2.3 Data processing and interpretation

Data processing and interpretation is the last part of a microbial metabolome study. It is impossible to present a general outline of how to do this because it will depend on the topic being investigated and the type of data that has been collected, but generally if the number of metabolites detected is large (known and unknowns), data processing is likely to rely more on multivariate data analysis (Esbensen 2000) than on univariate statistics. In Table 2 a list of prominent papers are given sorted by whether a non-targeted or a targeted approach was taken in the investigation. The table exemplifies many of the different areas in which metabolomics studies on microbes have been utilized. To illustrate examples of data processing strategies, along with choices made for data acquisition, the following text will present six papers of Table 2 in more detail, starting with three papers utilizing a non-targeted approach and then reviewing three papers utilizing a targeted approach.

Before proceeding two general aspects of data processing are worth mentioning: Specific for non-targeted approaches is the identification of putative metabolites from accurate mass measurements using extensive metabolite databases. Approaches for doing such assignments is an area in its own right (Rogers, Scheltema et al. 2009), and will not be covered in this text. When it comes to targeted approaches, accurate and precise determination of absolute concentrations is of particular importance. As absolute quantitation is also the topic of Paper I of this Doctoral thesis, a separate section will be devoted to this at the end this subchapter.

Table 2: Examples of recent prominent papers using non-targeted and targeted approaches investigating various topics by microbial metabolomics.

Approach	Topic investigated	Organism/System	Data acqusition	Reference
Non-targeted	Metabolite profiles assosiated with different growth phases	*E. coli*	FT-ICR-MS	Takahashi, Kai et al. 2008
	Salt stress response	*S. coelicolor*	LC-LIT-Orbitrap-MS	Kol, Merlo et al. 2010
	Identification of the function of genes in silent mutations	*S. cerevisiae*	1H-NMR	Raamsdonk, Teusink et al. 2001
	Cyclic cell cycle dependent changes in metabolic state (non-targeted and targeted)	*S. cerevisiae*	LC-QqQ-LIT-MS and GCxGC-ToF-MS	Tu, Mohler et al. 2007
	Impact of infection on host hormone metabolism	Host - *Salmonella*	Qq-FT-ICR-MS	Antunes, Arena et al. 2011
	Metabolic changes in host caused by infection	Host - *M. tuberculosis*	1H-NMR	Shin, Yang 2010 et al. 2011
	Metabolic differences between drug-sensitive and drug-resistant clinical isolates	Host - *Leishmania donovani*	LC-LIT-Orbitrap-MS	t'Kindt, Scheltema et al. 2010
Targeted	Ammonia assimilation and its control	*E. coli*	LC-QqQ-MS	Yuan, Doucette et al. 2010
	Stress response to different perturbations (cold, heat, oxidative stress, lactose diauxie and stationary phase)	*E. coli*	GC-ToF-ToF-MS	Jozefczuk, Klie et al. 2010
	Absolute metabolite concentrations and the enzyme active site occpancies they imply	*E. coli*	LC-QqQ-MS	Bennett, Kimball et al. 2009
	Increasing production of L-phenylalanine	*E. coli*	Enzymatic assay	Gerigk, Bujnicki et al. 2002
	Adaptation to different carbons sources	*P. putida*	GC-Q-MS LC-ion trap-MS	van der Werf, Overkamp et al. 2008
	Increasing production of L-Valine	*C. glutamicum*	Enzymatic assay	Radmacher, Vaitsikova et al. 2002; Ternbach, Bollmann et al. 2005
	Ivestigation of metabobolic response to starvation	*S. cerevisiae* *E. coli*	LC-QqQ-MS	Brauer, Yuan et al. 2006
	Improving tolerence to acetic conditions under ethanol production	*S. cerevisiae*	GC-ToF-MS CE-ToF-MS	Hasunuma, Sanda et al. 2011

Non-targeted approaches

There are two papers in Table 2 investigating prokaryotic microorganisms by a non-targeted approach. The first of these is Takahashi, Kai et al. (2008), in which direct infusion FT-ICR-MS was used to analyze time series of *E. coli* metabolite extracts and to show that the specific metabolite composition of a sample depends on culture growth phase at time of sampling. The times series were composed of samples from exponential and stationary phase of batch fermentations and in the FT-ICR-MS raw data they were able to detect 220 independent ion groups belonging to 174 potential metabolites of which they were able to assign putative identities to 72 metabolites based on accurate mass measurements using public natural compound databases (KNApSAcK and KEGG). They analyzed their data using the multivariate data analysis techniques principal component analysis (PCA), partial least squares regression (PLS-R) and cluster analysis. In PCA the time series metabolite extracts separated into two distinct groups, one for samples taken in exponential growth and one for samples taken in stationary growth, showing the presence of growth phase dependent metabolite concentrations in the dataset. Further supporting this using PLS they were able to deduce a linear model estimating optical density values based on metabolite abundances. The cluster analysis produced 11 isolated clusters for 148 of the 220 detected ions, whilst the remaining 72 ions showed no significant correlation. For the five largest clusters they presented five separate time series graphs of average ion abundances for each cluster over time. These five graphs all had shapes distinct from each other: the first with high concentration in early exponential phase, the second with high concentration in mid-stationary phase, the third with high concentrations throughout stationary phase, the forth with high concentrations throughout stationary phase, and the fifth with high concentrations in early stationary phase. The authors also offered a biological interpretation of the clusters based on the ions in the clusters that had been assign putative identities: e.g. for the cluster that showed a metabolite abundance time profile with high concentrations in early exponential phase, one of the ions in the cluster was identified as oleic acid, a precursor in fatty acid synthesis with one double bond. This lead to the proposition that ions in this cluster belonged to

compounds related to fatty acid biosynthesis and that fatty acid biosynthesis occurs only early in the exponential growth phase of *E. coli*.

The second paper of Table 2 studying a prokaryotic organism is Kol, Merlo et al. (2010). Here the salt stress response in *Streptomyces coelicolor* is investigated and separation is used prior to MS analysis employing a LC-Orbitrap-MS with a linear quadrupole ion trap (LIT) interface. The used LC-technique, although aiding data analysis through fractionation of the sample, selects for mid-polar compounds, and thus the entire metabolome is not analyzed. Metabolite extracts from the wild type strain and five mutants derived thereof were prepared for a time series after addition of salt to cultures grown in shake flasks. Three mutants had disrupted biosynthesis of a known osmoprotectant (ectoine) and two mutants had disrupted stress regulatory enzymes. It was possible to detect 1247 independent ion groups belonging to 363 potential metabolites. The detection of more potential metabolites in this study than in the previously presented study of growth phases of *E. coli*, 363 versus 174 respectively, reflects the expected higher sensitivity attained by employing LC separation prior to MS analysis. Out of the 363 potential metabolites detected, 229 were assigned putative identities based on accurate mass searches using the metabolite specific databases ScoCyc and LIPID MAPS and a contaminant database before unidentified ions were identified with KEGG and then METLIN and finally the Human Metabolome Database. Cluster analysis was used on the time series data and identified two coherent clusters where metabolite abundances either increased or decreased for salt-stressed cultures (wild type and mutants). Similar responses to salt stress for the wild type strain and for the mutants with disrupted biosynthesis of a known osmoprotectant indicated novel responses to salt stress. Similar responses to salt stress for the wild type strain and the mutants with disruption of enzymes involved in the stress response indicated robustness of the metabolic response. Comparing time profiles of putatively identified metabolites, 15 metabolites showed an increase in abundance for all strains, possible acting as osmoprotectants, whilst three metabolites showed a decrease in abundance for all strains. Most of the metabolites with increasing responses were amino acids and di and tri-peptides, some of which have been proposed to be osmoprotectants in previous studies.

The last paper to be presented utilizing an non-targeted approach is Raamsdonk, Teusink et al. (2001). They presents a conceptual and experimental framework for revealing the phenotype of mutations that show no overt phenotype in terms of growth rate or other online measurements, i.e. so called silent mutations. Revealing such phenotypes for silent mutations can aid annotation of genes encoding proteins of unknown function. In this proof-of-principle study metabolite extracts of a steady-state grown *Saccharomyces cerevisiae* wild type strain and metabolite extracts of steady-state grown single deletion mutants of this strain were prepared. The mutants comprised of two deletion mutants having one out of two isoenzymes for 6-phosphofructokinase deleted, two deletion mutants having one of two different enzymes known to produce qualitatively similar but quantitatively different effects deleted (one that organize the assembly of the cytochrome oxidase complex and one that encodes a protein subunit of the same complex) in addition to control strains. The metabolite extracts were analyzed by 1H-NMR, and the produced spectra were, after normalization of responses and pruning of the range inspected, subjected to PCA. Another multivariate data analysis technique, discriminant function analysis (DFA) was then performed on a selection of the PC projections. DFA is a supervised technique and as such replicate metabolite extracts of different strains were defined to belong to different groups and a DFA plot minimizing within-group variance and maximizing between-group variance could be produced. In this plot the two mutants with either one of the two isoenzymes deleted clustered together, whilst the two other deletion mutants produced two separate clusters. This showed that single deletion mutants for enzymes with different function produce separate clusters, and that single deletion mutant for enzymes with the same function produce a single cluster. Thus based on a mutant's metabolome composition, it is possible to find the function of its deleted gene when analyzing its metabolome together with the metabolome of other mutants that have had genes of known function deleted.

Targeted approaches

To illustrate targeted approaches the three most cited articles of Table 2 (per year having been published) were selected. The first of these is Yuan, Doucette et al. (2009) where regulation of ammonia assimilation in *E. coli* was investigated. Using two LC-

QqQ-MS methods, one hydrophilic interaction chromatography method and one ion-pairing reversed phase chromatography method, time series of metabolite extracts from *E. coli* strains grown on filters on top of an agarose-media mixture were analyzed. A wild type and two mutants were used, the mutants having one enzyme in one of the two routes for ammonia assimilation deleted (either glutamate synthase of the glutamine synthetase/glutamate synthase cycle or glutamate dehydrogenase). The time series was composed of samples taken when ammonium concentration was limiting and of samples taken after the cells on the filter were moved to plates with high ammonium concentration. The two different LC-QqQ-MS methods were capable of detecting approximately 250 metabolites enabling collection of time series data for 59 metabolites. Out of the 59 metabolites, absolute concentrations were determined for glutamate and glutamine, whilst relative concentrations between strains were used to evaluate the remaining metabolite concentrations. The metabolome time series data for the wild type strain was visualized by a heat map showing the ratio of it to a control culture in exponential growth for all detected metabolites. Through an unbiased statistical analysis (singular value decomposition (SVD)) they were able to identify that there were two dominating characteristic response patterns in the data (accounting for 63% and 23% of the overall information in the data respectively), the first pattern showing a strong accumulation during ammonium limitation and a decrease to a normal level after alleviating ammonium limitation, the second pattern showing depletion during limitation and an elevation over normal (control) after alleviation of limitation before adjusting down to a normal level. The compound most strongly contributing to the first characteristic pattern was α-ketoglutarate and the compound most strongly contributing to the second characteristic pattern was glutamine, which could be visually verified by inspecting the heat map. This analysis showed that the compounds most strongly influences by ammonium limitation were themselves the central players, whilst homeostasis was maintained throughout much of the core metabolism. A model of the five central metabolic and regulatory enzymatic reactions in ammonium assimilation was made based on ordinary differential equations. Fitting this model, that included biochemical parameters from literature, to the experimental time series data of the wild type and mutant strains produced the first concrete example of how competition for enzyme active sites of saturated enzymes is crucial in regulation of enzyme activity

(competition of aspartate, glutamate, and glutamine for active sites of glutamate synthase and competition of glutamate, α-ketoglutarate, oxaloacetate and aspartate for the active sites of aspartate aminotransferase). This result is in contrast to the view that saturated enzymes are insensitive to substrate concentration. The model was verified by conducting control experiments which produced experimental results predicted by the model not just qualitatively but also quantitatively.

The second of the most cited articles with a targeted approach in Table 2 is Bennett, Kimball et al. (2009). In this work two LC-QqQ-MS methods, the same methods as in the previous article described, were used to determine the absolute concentrations of approximately 100 metabolites for exponentially growing *E. coli* with glucose, glycerol or acetate as the carbon source. The aim of the study was to characterize the metabolome of *E. coli* and to use this knowledge to derive enzyme active site occupancy. As in the preceding paper growth on filters on top of an agarose-media mixture was used. It was determined that the total measured intracellular metabolite pool was approximately 300 mM, with the 10 most abundant metabolites summing up to 77% of the total intracellular metabolite pool and the less abundant half of metabolites only summing up to 1.3% of the total pool. For all growth conditions glutamate, being the major nitrogen donor in a cell and the major intracellular counter ion to potassium, was the most abundant metabolite. Comparing carbon sources the majority of metabolites had different concentrations comparing glucose to glycerol and acetate (81% and 67% of metabolites with significant changes respectively), whilst the difference in metabolite concentrations were less when comparing glycerol and acetate (51% of metabolites with significant changes). The dataset was validated using thermodynamics-based metabolic flux analysis assessing feasibility of flux directionality based on calculation of Gibb's free energy. The validated dataset for growth on glucose was then used to assess enzyme saturation for substrate-enzyme pairs comparing metabolite concentrations to the enzymes Michaelis constants K_m. It was found that substrate concentration was larger than K_m, that is the enzyme was more than half saturated, for 83% of the substrate-enzymes pairs, of which 59% had a concentration more than tenfold higher than that of K_m. Having a high proportion of enzymes saturated is consistent with minimizing the high cost of protein biosynthesis

through maximizing flux per enzyme. Many of the reactions where substrate concentrations were less than K_m, were degradation reactions, whilst intermediate cases where substrate concentration and K_m was similar often involved enzymes belonging to central carbon metabolic reactions. This result is reasonable as utilization of degradation pathways to a large extent is unnecessary for exponentially growing cells on minimal media, and as reactions of central carbon metabolism are often bidirectional.

The last paper with a targeted approach to be presented is one by Hasunuma, Sanda et al. (2011). They investigated the inhibitory effect of acetic and formic acid on ethanol production by *S. cerevisiae* genetically modified to ferment xylose. Xylose is an important component of lignocellulose (plant dry matter) and its entry point into central carbon metabolism for the genetically modified strains was xylulose 5-phosphate (X5P) in PPP. Pretreatment of lignocellulose involves solubilization and hydrolysis inevitably releasing weak acids such as acetic acid and formic acid, so the inhibitory effect of these acids needs to be overcome if lignocellulose is to be used as raw material for bio-ethanol production by *S. cerevisiae*. CE-ToF-MS was used to detect 24 sugar phosphates, organic acids, nucleotides and coenzymes involved in central carbon metabolism and GC-ToF-MS was used to detect intracellular sugar and sugar alcohols xylose, xylitol and glycerol. Time series of metabolite extracts from anaerobic *S. cerevisiae* fermentations in the presence and absence of acetic acid at various concentrations (30mM and 60mM) were analyzed. Using xylose or xylose and glucose as carbon sources it was demonstrated by inspecting the time series data that the presence of acetic acid impeded xylose utilization, but not glucose utilization, and that the presence of acetic acid impeded ethanol production when glucose was not present as an alternative carbon source. These effects were stronger the higher the acetic acid concentration was. Analysis of the metabolite extracts from the fermentations using xylose as the sole carbon source showed that the decrease in xylose utilization and ethanol production coincided with an increase in the concentration of certain metabolites in the non-oxidative part of PPP (ribulose 5-phoohate (R5P), ribose 5-phosphate (R5P), seudoheptulose 7-phosphate (S7P) and erythrose 4-phosphate (E4P)) and an increase of certain glycolytic intermediates (glucose 6-phosphate (G6P), fructose 6-phosphate (F6P), fructose 1,6-bisphoshate (FBP) and dihydroxyacetone phosphate

(DHAP)). Further research was focused on attempting to alleviate the accumulation of PPP intermediates. To do this mutants with overexpression of the enzymes responsible for the rate limiting steps in the non-oxidative part of PPP, conversion of R5P and X5P to glyceraldehyde 3-phosphate (GAP) and S7P (by a transketolase) and further conversion of GAP and S7P to F6P and E4P (by a transaldolase), was constructed (refer to Figure 1 for an overview of primary metabolism). Mutants with one of the two or with both enzymes overexpressed were constructed and it was demonstrated that only overexpressing one enzyme at the time was beneficial to xylose utilization and ethanol production. Sole overexpression of the transketolase was the most beneficial giving an ethanol production of 8.33 g/L after 48h in the presence of the highest acetic acid concentration compared to 0.33 g/L for the control. The beneficial effect of transketolase overexpression was subsequently also demonstrated in the presence of formic acid.

Data correction strategies for quantitative analysis

As already stated accurate and precise absolute quantitative measurements are the aim of a targeted approach. Measurement accuracy and precision can be affected by random and systematic errors through the entire length of sample preparation and all through sample analysis, and this should be corrected for when possible. When processing datasets from analysis of metabolite extracts by a particular GC-MS, LC-MS or CE-MS method capable of detecting specific metabolite groups, it has previously not been uncommon to use a small selection of isotopically labeled metabolites, one for each metabolite group detected, to correct the responses of all metabolites in the sample. This is a flawed approach as even metabolites belonging to the same metabolite group (e.g. amino acids), are influenced differently by the physical and chemical conditions they are subjected to during quenching of metabolism, extraction of metabolites from bacterial cells, further sample preparation including derivatization if applicable and MS-analysis. Certain derivatization protocols, such as derivatization using N-methyl-N-trimethylsilylfluoroacetamide (MSTFA), produce metabolite derivatives that are inherently instable and have differing reaction kinetics. As a simple alternative solution to this problem instead of using a small set of isotopically labeled standards, it has been

proposed to analyze individual samples in a sample set when the same amount of time has passed since their derivatization (Villas-Bôas, Smart et al. 2011), however this approach might become very time consuming if the derivatization procedure in not automated.

The best correction strategy is to have one internal standard for each metabolite detected in a sample. In isotope dilution mass spectrometry (IDMS) this is achieved through using uniformly (U)-^{13}C-labeled metabolite extract prepared by running fermentations with a U-^{13}C-labeled compound as the sole carbon source prior to preparing metabolite extracts of the bacterium being investigated. Before extracting metabolites from the bacterium studied, labeled metabolite extract is added to the sample, in the original protocol in a ratio sample to labeled extract 4:1. The labeled extract is also added to the calibration standard samples that are used to construct a calibration curve. When analyzing the metabolite extracts and the calibration standards together, this enables construction of a liner calibration curve plotting area ratio (naturally labeled metabolite in calibration standard) /(U-^{13}C-labeled metabolite from labeled extract) versus known concentration in the calibration standards. This approach corrects for changes in metabolite concentration during sample preparation and effects of the analytical method because it assures that the physical and chemical behavior of the labeled and unlabeled compound is the same. The one thing that cannot be corrected by this approach is incomplete extraction of metabolites from the bacterial cell. IDMS was first reported as a LC-ESI-MS/MS method showing improved accuracy through improved linearity of the calibration curve and improved precision through reduction of the standard deviation (Wu, Mashego et al. 2005). Adaptation of IDMS has also proven successful for GC-MS (Cipollina, ten Pierick et al. 2009; Vielhauer, Zakhartsev et al. 2011).

An approach similar to IDMS, Mass Isotopomer Ratio Analysis of U-^{13}C-Labeled Extracts (MIRACLE) is proposed in (Mashego, Wu et al. 2004). In MIRACLE a solution containing cells from a fermentation using a U-^{13}C-labeled carbon source is used as a quenching solution for a second fermentation using a naturally labeled carbon

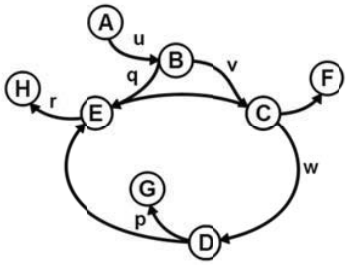
source, the difference between IDMS and MIRACLE is thus whether or not l3C-labeled metabolites have been extracted from the cells. MIRACLE is thus convenient for the relative comparison of two fermentations as metabolites can be extracted from cells at the same time, and if the two cell types are the same, incomplete extraction will not hamper interpretation of results. It is also possible to slightly alter the experimental setup and co-quench the unlabeled and naturally labeled fermentations adding the two culture samples to the quenching solution at the same time, further reducing the possibility of introducing differences between the two fermentations during sample preparation. In comparison IDMS cannot correct for effects of quenching and effects of extraction other than those caused by the chemical and physical environment, but it has the advantage over MIRACLE that it enables absolute quantification, as the labeled cells with non-extracted metabolites used in MIRACLE cannot be directly added to calibration standards.

An important issue when using labeled metabolite extracts as internal standards is that the extracts might not contain all metabolites of interest or the concentration of certain metabolites might be too low to be useful. In GC-MS analysis of metabolite extracts, the samples need to be derivatized to infer the volatility and thermal stability needed for analysis. This introduces the possibility of creating labeled internal standards for all metabolites of interest through isotope coded derivatizing (ICD) reagents. Kvitvang, Andreassen et al. (2011) presents a GC-CI-MS/MS method using this approach for derivatization with methylchloroformate (MCF), which is applicable for amino acids and carboxylic acids. During this doctoral work an equivalent method derivatizing samples with MSTFA was developed (Paper I). MSTFA is applicable to a number of metabolite groups including sugars, amino acids and carboxylic acids, but there are reports that MCF is more appropriate for the two latter groups (Villas-Bôas, Smart et al. 2011). Utilization of ICD reagents to synthesize internal standards for all metabolites detected by a specific analytical method is attractive because it is less expensive and less time consuming than attaining them commercially or producing them through individual synthesis. Although derivatization usually is unnecessary for LC-MS analysis, it has been used to increase sensitivity, selectivity and chromatographic

performance in several applications. This opens for utilization of ICD reagents to create internal standards also for LC-MS. For an extensive review on the use of ICD reagents for GC-MS and LC-MS applications in metabolic profiling see Bruheim, Kvitvang et al. (2013). The disadvantage of using ICD reagents for absolute quantitation in MS-analysis (ICD-MS) compared to IDMS is that ICD-MS does not account for events that influence metabolite concentrations in a sample prior to derivatization.

1.3 Fluxomics – quantifying reaction rates

Metabolomics deals with metabolite concentrations, but metabolite concentrations are not the only entities that describe the metabolic network of an organism. The rates of the reactions of the metabolic network are another important parameter. The collection of these rates for a network constitutes its fluxome (Sauer, Lasko et al. 1999), and the study of these rates is called fluxomics. Using classical measurements of metabolite concentrations as described in chapter 1.2, the direction of a metabolic flux can be determined if all metabolites involved in the reaction are measured and if the thermodynamic equilibrium constant (K_{eq}) is known (van der Werf, Overkamp et al. 2008; Bennett, Kimball et al. 2009). Determining the magnitude of fluxes is somewhat more complex. One of the earliest approaches was to use stoichiometric metabolic flux analysis (MFA) (Varma and Palsson 1994). In stoichiometric MFA the metabolic network is described by material balances over all network metabolites. Figure 7 displays a simple network with four intracellular metabolites B, C, D and E and three extracellular metabolites A, F and G. The metabolites are connected by fluxes u, v, w, q, p and r. To the right in the figure material balances for the intracellular metabolites at metabolic steady state are given.



B: $u - q - v = 0$
C: $v - w = 0$
D: $w - p = 0$
E: $p + q - v - r = 0$

Figure 7: A simple metabolic network with intracellular metabolites B, C, D and E and extracellular metabolites A, F, G and H connected by fluxes u, v, w, p, q and r. Material balances over the internal metabolites are given to the right.

The material balances of Figure 7 can be written using matrix notation:

$$\begin{pmatrix} u-q-v \\ v-w \\ w-p \\ p+q-v-r \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Updownarrow$$

$$\begin{pmatrix} 1 & -1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} u \\ v \\ w \\ p \\ q \\ r \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Analogues to this, the material balance for any metabolic network can be given as

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{b} \qquad (1)$$

where $\mathbf{S}$ is the stochiometric matrix, $\mathbf{v}$ is the fluxes of the network and $\mathbf{b}$ is the net metabolite uptake by the system.

The material balances for the network given in Figure 7 consist of four equations with six unknown fluxes, giving the system two degrees of freedom (DOF). This means that given a value for two distinct fluxes, e.g. u and q, the remaining four fluxes, in this case

v, w, p and r, can be calculated. In this connection u and q are termed independent fluxes or the set of free flues and v, w, p and r are termed dependent fluxes. In any metabolic network there are typically far fewer metabolites than there are fluxes relating them. This means that even if online measurements such as respiration, biomass production and product formation are incorporated, there are still several DOF. Therefore equation (1) will typically have a solution space, rather than a unique solution. If it fits the application an optimization criterion, such as maximal growth, can be introduced to give a unique solution. Stochiometric MFA has had many applications including examination of basic metabolic physiology (Majewski and Domach 1990), interpretation of experimental data (Varma and Palsson 1994), identification of targets for metabolic engineering (Varma, Boesch et al. 1993), growth medium optimization (Xie and Wang 1993) and process design optimization (Varma and Palsson 1994).

However, stochiometric MFA has some limitations. Using this approach it is not possible to determine distinct fluxes for parallel pathways unless flux measurements directly connected to the separate pathways exists. Metabolic cycles are also unattainable, as are bidirectional reactions steps and split pathways if cofactors are not balanced (Wiechert 2001). These issues can be resolved by an approach utilizing measurements of metabolite isotopomer distribution from carbon labeling experiments (CLE). In CLEs the organism being studied is fed an isotopically labeled substrate. The incorporation of labeled material cause all metabolites to exist as a collection of different isotopomers, that is a collection of isomers that have a varying number of ^{13}C-carbons at specific positions, and the different mass isotopomers (isotopomers with the same mass) of these are detectable by MS and NMR analysis. The same MS-methods as those utilized for metabolomics can be used, and the same limitations apply, but the methods need to be expanded to account for the different mass isotopomers of a metabolite (Choi and Antoniewicz 2011).

There are several different softwares available to do metabolic flux analysis based on CLE, an approach termed ^{13}C based MFA. FiatFlux, 13CFLUX2 and Metran are the

most renowned software for ^{13}C based MFA available for academic work. Other software for metabolic flux analysis based on CLE include C13 contained in The BioMet Toolbox (Cvijovic, Olivares-Hernandez et al. 2010), Flux-P (Ebert, Lamprecht et al. 2012), OpenFlux (Quek, Wittmann et al. 2009), FIA (Fluxomer Iterative Algorithm) (Srour, Young et al. 2011), NMR2FLUX (Sriram, Fulton et al. 2004), and most recently Influx_s (Sokol, Millard et al. 2012).

FiatFlux was introduced in 2005 (Zamboni, Fischer et al. 2005) and is aimed at users not necessarily familiar with CLE and numerical methods. FiatFlux is run in Matlab, has a graphical interface and is supplied with models for the central carbon metabolism of selected bacterial species. The software is predefined to take in GC-MS data (netCDF-files) of TBDMSTFA (N-(*tert*-butyldimethelsylil)-N-methyl-trifluoroacet-amide) derivatized protein-bound amino acids from CLE using 1-labeled glucose and uniformly labeled glucose. Through global material balances, intracellular fluxes are estimated from measured extracellular fluxes using flux ratios from the isotopomer measurements as constraints. Although the software can be configured to handle other organisms, other ^{13}C-substrates, other derivatization protocols and other separation techniques, the user friendliness of FiatFlux comes at the cost of versatility. In Fuhrer, Fischer et al. (2005) FiatFlux is used to determine and compare the flux map of seven bacterial species growing on glucose.

13CFLUX2 is the successor of the software 13C-FLUX introduced at the turn of the century (Wiechert, Möllney et al. 2001). It is s Linux based modularly constructed software whose programs are executed on command line. In essence global isotopomer balances and iterative least square fitting is used to determines the flux distribution of a metabolic network that produces the metabolite isotopomer distribution closest to an experimentally obtained metabolite isotopomer distribution. 13CFLUX2 has the advantage that it can incorporate all types of metabolite isotopomer measurements (MS, MS/MS and NMR) and it also contains features to aid experimental design and to do statistical analysis (Weitzel, Nöh et al. 2013). The working principles of the 13CFLUX2

software will be explained in more detail in section 1.3.1. For simplicity the biochemical network editor and data visualization software Omix (Droste, Miebach et al. 2011), can be used as a front end graphical interface for model development and measurement specifications, as well as for visualization of results. The implementation of Omix in this connection is relatively recent and its utilization is ever expanding. 13CFLUX2's predecessor has been used to characterize the anaplerotic reactions of *Corynebacterium glutamicum* identifying a futile cycle phosphoenolpyruvate – pyruvate – oxaloacetate – phosphoenolpyruvate with a flux 3-fold of the anaplerotic flux required for biosynthesis (Petersen, de Graaf et al. 2000). In another study of *C. glutamicum* 13C-MFA with 13CFLUX2 was used in addition to transcriptome and metabolome analysis to search for strategies to improve L-lysine production. It was shown that for mutants with decreased citrate synthase activity, L-lysine production increased, although the flux from acetyl-CoA and oxaloacetate to citrate remained relatively constant. It was proposed that possibly a unchanged flux would still produce an increased oxaloacetate derived L-lysine production because citrate synthase in this situation would be working at close to maximum of its capacity, leading to a high concentration of its substrates (van Ooyen, Noack et al. 2012). 13C-MFA at isotopic and metabolic steady state has recently been compared to the evolving method of 13C-MFA at isotopic non-steady and metabolic steady state using consistent dataset. In this study the current limitations of both methods were demonstrated and possible pitfalls were concluded (Noack, Nöh et al. 2011).

Metran is a software introduced in 2007 that utilize global isotopomer balancing to estimate fluxes in a fashion similar to 13CFLUX2 (Yoo, Antoniewicz et al. 2008). The concept of elementary metabolite units (EMU) (Antoniewicz, Kelleher et al. 2007) for reduction of variable complexity is utilized in Metran in the same way that the concept of cumomers (*cum*ulative isotopo*mers*, more of which in the next section) are used as the default choice for 13CFLUX2. The features of the software are reported to include statistical analysis and procedures to determine optimal experimental design, although it is difficult to give further details as, to the best of my knowledge, these procedures have yet not been published. Yoo, Antoniewicz et al. (2008) gives an example of the use of

Metran elucidation the utilization of glutamate for fatty acid synthesis in a brown adipocyte cell line.

In the following subchapter a more thorough introduction to 13CFLUX2 will be given. Knowledge of the details given here are not required when reviewing the fluxomics based paper of this thesis (Paper III), but will give a more in depth understanding of how the software works and therefore enable evaluation of the data processing done in the paper. The section is also intended as a light introduction to 13CFLUX2, only requiring basic skills in statistics and linear algebra, for anyone who might be interested in learning to use the software.

1.3.1 The software 13CFLUX2

13CFLUX2 is a powerful tool to determine the fluxes of a metabolic network. As intracellular fluxes cannot be measured directly, but have to be determined indirectly from measurements of isotopomer distributions, some basic knowledge on the mathematical principles behind 13CFLUX2 and on how the software works is a prerequisite when it comes to evaluating experimental methods and results. In this connection this section is intended to give a quick introduction to the topic.

The flux distribution of a network fed labeled substrate yields a specific isotopomer distribution

The network displayed in Figure 8 is the carbon transition network corresponding to the metabolic network in Figure 7 with q set to zero. Assuming u = 1, the system is fully determined with v = w = p = u = 1 and r = 0. In the figure each metabolite has its carbons indicated with a circle representing ^{12}C and a dot representing ^{13}C. Using this simple example it will be demonstrated how the flux distribution of a network fed a labeled substrate yields a specific isotopomer distribution for that network.

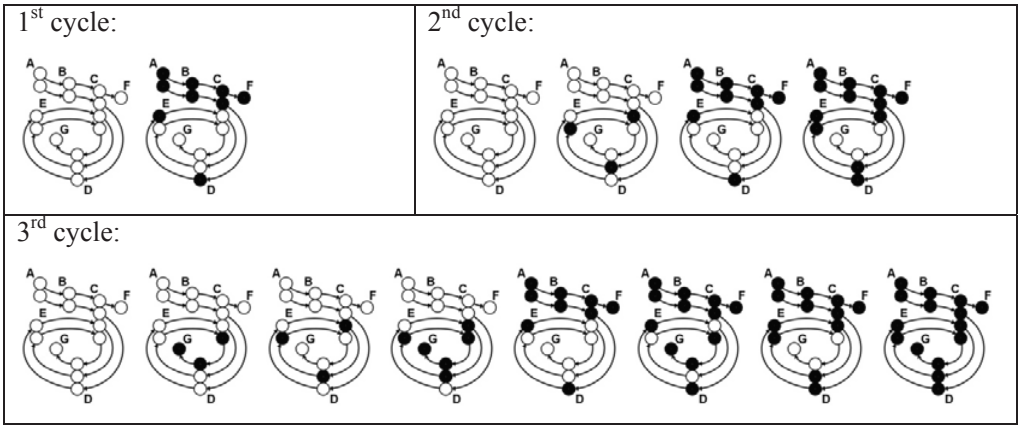


Figure 8: The carbon transitions occurring in a simple metabolic network at metabolic steady state reaching isotopic steady state when substrate A is changed from 100% unlabeled to 50% unlabeled and 50% uniformly labeled.

Imagine starting with the system at metabolic steady state, changing the feed from 100% unlabeled A to 50% unlabeled A and 50% uniformly labeled A. The system will first go through a period of isotopic unsteady state where metabolite isotopomer distributions change. First, only unlabeled E is available leading to two possible 1st cycles which produce 50% unlabeled E and 50% 1-labeled E. In the 2nd cycle with the two different isotopomers of E available, four different isotopomers of metabolite C can be produced, yielding 25% of each isotopomer of E. In the 3rd cycle, with all isotopomers of E available, eight different isotopomers of C can be produced. As for the 2nd cycle, the 3rd cycle yields 25% of each isotopomer of E, and thus isotopic steady state is reached as subsequent reactions will not lead to changes in the isotopomer distribution of the system. The isotopomer distribution of the network when the feed is 50% unlabeled A and 50% uniformly labeled A is given under condition 1 in Table 3. By convention a specific isotopomer of a metabolite is denoted by its name and a hash followed by 0's and 1's indicating ^{12}C and ^{13}C respectively. For the example network the order of carbon atoms is given from the outside of the spiral and inwards. The fraction of a specific isotopomer is indicated by a lower case letter with the 0's and 1's as subscripts (an upper case letter would indicates the amount of that specific isotopomer). Changing q from zero to 0.3 and again assuming u = 1.0, the remaining fluxes will be v = 0.7, w = p = v = u and r = 0.3. The isotopomer distribution for this situation is given under condition 2 in Table 3. Comparing condition 1 and condition 2

it is easy to see that the isotopomer distribution of a network depend on the network fluxes, and this is the basis of why measurements of metabolite isotopomer distributions can be used to calculate fluxes of a network.

Table 3: Isotopomer distribution of the spiral network with q=0 and substrate composition a_{00} = 0.50 and a_{11} = 0.50 (condition 1), with q=0.3 and substrate composition a_{00} = 0.50 and a_{11} = 0.50 (condition 2), with q=0 and substrate composition a_{00} = 0.25 and a_{11} = 0.75 (condition 3) and with q=0 and substrate composition a_{00} = 0.20, a_{10} = 0.20 and a_{11} = 0.60 (condition 4). 0 and 1 in the name of isotopomers (left column) indicate 12C and 13C, respectively.

	Condition 1	Condition 2	Condition 3	Condition 4
B#00	0.50	0.50	0.25	0.20
B#10	-	-	-	0.20
B#11	0.50	0.50	0.75	0.60
C#0000	0.125	0.1625	0.015625	0.032
C#1000	-	-	-	0.032
C#1100	0.125	0.1625	0.046875	0.096
C#0010	0.125	0.0875	0.046875	0.048
C#1010	-	-	-	0.048
C#1110	0.125	0.0875	0.140625	0.144
C#0001	0.125	0.0875	0.046875	0.048
C#1001	-	-	-	0.048
C#1101	0.125	0.0875	0.140625	0.144
C#0011	0.125	0.1625	0.140625	0.072
C#1011	-	-	-	0.072
C#1111	0.125	0.1625	0.421875	0.216
D#000	0.125	0.1625	0.015625	0.064
D#100	0.125	0.1625	0.046875	0.096
D#010	0.125	0.0875	0.046875	0.096
D#110	0.125	0.0875	0.140625	0.144
D#001	0.125	0.0875	0.046875	0.096
D#101	0.125	0.0875	0.140625	0.144
D#011	0.125	0.1625	0.140625	0.144
D#111	0.125	0.1625	0.421875	0.216
E#00	0.25	0.325	0.0625	0.16
E#10	0.25	0.175	0.1875	0.24
E#01	0.25	0.175	0.1875	0.24
E#11	0.25	0.325	0.5625	0.36
F#0	0.50	0.50	0.25	0.20
F#1	0.50	0.50	0.75	0.80
G#0	0.50	0.50	0.25	0.40
G#1	0.50	0.50	0.75	0.60
H#00	-	0.325	-	-
H#10	-	0.175	-	-
H#01	-	0.175	-	-
H#11	-	0.325	-	-

It is also worth noting that the isotopomer distribution of a network will change if the composition of the feed is changed. Comparing condition 1 to condition 3 and condition 4 in Table 3 shows how this applies to the example network. In condition 3 the feed for the example network with q = 0 is set to $a_{00} = 0.25$ and $a_{11} = 0.75$. This does not lead to new isotopomers formed compared to condition 1, but the fractions of the various isotopomers change. In condition 4 a new isotopomer is introduced in the feed as it is set to $a_{00} = 0.20$, $a_{10} = 0.20$ and $a_{11} = 0.60$. Now not only do the isotopomer fractions change, but isotopomers not previously present are also formed. As a result of how the isotopomeric state of the network depends on substrate isotope composition, determining what substrate composition to use is an essential part of experimental design (Moellney, Wiechert et al. 1999). The inevitable uncertainty in experimental isotopomer measurements leads to an experimental dataset with a collection of uncertainties. This again correlates to a flux distribution with specific confidence intervals, so it is important to choose a substrate isotopomer composition that gives experimentally determined fluxes that can shed light on the issues being investigated (i.e. if the aim of the study is to investigate if glycolysis is more active in the organism than the pentose phosphate pathway, determining that the flux from glucose 6-phosphate to fructose-6 phosphate is 3 ± 0.1 and that the flux from glucose 6- phosphate to phosphogluconolactone is 4 ± 0.3 is preferable to determining that the two fluxes are 3 ± 2 and 4 ± 3 respectively). To help determine the optimal substrate composition, 13CFLUX2 contains features that enable a quantitative comparison of the suitability of all combinations of a set of substrate isotopomers for a specific experimental objective.

The isotopomeric state of any network can be found explicitly (the mathematical framework)

For simple systems, such as the spiral network in Figures 7 and 8 when q = 0, it is easy to calculate the isotopomer distribution of the network using simple combinatorics and probability calculations as illustrated in the previous section. As the network becomes more complicated, the calculations quickly become more complex, and a structured mathematical framework is needed. In 13CFLUX2 the isotopomer distribution of the network is calculated from its fluxes by formulating material balances over all

metabolite isotopomers of the system. General isotopomer balance equations for the intracellular metabolites of the spiral network at metabolic and isotopic steady state are given by formulas (2), (3), (4) and (5):

$$\frac{d}{dt}(B \cdot b_{ij}) = u \cdot a_{ij} - (v+q)b_{ij} = 0 \qquad i,j \in \{0,1\} \qquad (2)$$

$$\frac{d}{dt}(C \cdot c_{ijkl}) = v \cdot b_{ij} \cdot e_{kl} - w \cdot c_{ijkl} = 0 \qquad i,j,k,l \in \{0,1\} \qquad (3)$$

$$\frac{d}{dt}(D \cdot d_{ijk}) = w(c_{0ijk} + c_{1ijk}) - p \cdot d_{ijk} = 0 \qquad i,j,k \in \{0,1\} \qquad (4)$$

$$\frac{d}{dt}(E \cdot e_{ij}) = p(d_{ij0} + d_{ij1}) + q \cdot b_{ij} - (r+v)e_{ij} = 0 \qquad i,j \in \{0,1\} \qquad (5)$$

This set of 32 equations for 32 unknown isotopomer is mathematically demanding to solve, and the nonlinear term in equation (3) seems to make an analytical solution impossible. However a computer aided solution of the set of equations can be achieved through introduction of the concept of *cum*ulative isotop*omer* fractions or *cumomer* fractions, meaning a specific sum of the isotopomer fractions of a metabolite (Wiechert, Mollney et al. 1999). For example the 0-cumomer fraction of metabolite B is the sum of all isotopomer fractions of B labeled 0 times or more, and is denoted b_{xx}, where x represents 0 or 1. The 1-cumomer fractions of metabolite B is the sum of all isotopomer fractions of B labeled 1 time or more at a specific position, denoted b_{1x} and b_{x1}. And last the 2-cumomer fraction of B is simply the fully labeled fraction b_{11}. In an analogous fashion 0 to n-cumomers can be formulated for any metabolite with n carbon atoms. Imagining the fictive compounds cumomers B#xx, B#1x, B#x1 and B#11, and analogue cumomers for the remaining intracellular metabolites of the spiral network, the general isotopomer balance equations (2) through (5) can be transformed into general cumomer balance equations. A key concept when doing such transformations is the weight of an isotopomer or a cumomer. The weight of an isotopomer is its' number of labeled carbons, whilst the weight of an n-cumomer is defined to be n. The transformation of isotopomer balance equations to cumomer balance equations is done by substituting 0's by x's and by eliminating terms that would violate a demand for weight conservation of all terms in an equation (weight of a quadratic term is defined to be the sum of its factor weights). The correctness of these transformation rules have been proven by Wurzel

(1997). Following these rules the general isotopomer equations (2) through (5) becomes the general cumomer balance equations (6) through (9) (for a more detailed explanation of the process of isotopomer balance formulation and their transformation into cumomer balances see Wiechert, Mollney et al. (1999)).

$$\frac{d}{dt}(B \cdot b_{ij}) = u \cdot a_{ij} - (v+q) b_{ij} = 0 \qquad i,j \in \{x,1\} \qquad (6)$$

$$\frac{d}{dt}(C \cdot c_{ijkl}) = v \cdot b_{ij} \cdot e_{kl} - w \cdot c_{ijkl} = 0 \qquad i,j,k,l \in \{x,1\} \qquad (7)$$

$$\frac{d}{dt}(D \cdot d_{ijk}) = w \cdot c_{xijk} - p \cdot d_{ijk} = 0 \qquad i,j,k \in \{x,1\} \qquad (8)$$

$$\frac{d}{dt}(E \cdot e_{ij}) = p \cdot d_{ijx} + q \cdot b_{ij} - (r+v) e_{ij} = 0 \qquad i,j \in \{x,1\} \qquad (9)$$

Because of the elimination of non-weight preserving terms, the cumomer balance equations are somewhat simpler than the isotopomer balance equations (see first terms of equations (8) and (9) compared to first term of equations (4) and (5)). This simplification has dramatic consequences: the balance equation for an n-cumomer only contains n- or lower cumomers. Because of this a cascade of linear balance equations can be constructed computing successively the 0- through n-cumomers given the values for a set of free fluxes and the input cumomer fractions a_{ij}. This cascade is given in equations (10) through (14) below.

0-cumomer balance equations:

$$\begin{aligned}
&\text{B\#xx} && : (v+q) b_{xx} = u \cdot a_{xx} \\
&\text{C\#xxxx} && : v \cdot b_{xx} \cdot e_{xx} - w \cdot c_{xxxx} = 0 \\
&\text{D\#xxx} && : w \cdot c_{xxxx} - p \cdot d_{xxx} = 0 \\
&\text{E\#xx} && : p \cdot d_{xxx} + q \cdot b_{xx} - (r+v) e_{xx} = 0
\end{aligned} \qquad (10)$$

1-cumomer balance equations:

$$
\begin{aligned}
&\text{B\#1x} && : (v+q)b_{1x} = u \cdot a_{1x} \\
&\text{B\#x1} && : (v+q)b_{x1} = u \cdot a_{x1} \\
&\text{C\#1xxx} && : v \cdot b_{1x} \cdot e_{xx} - w \cdot c_{1xxx} = 0 \\
&\text{C\#x1xx} && : v \cdot b_{x1} \cdot e_{xx} - w \cdot c_{x1xx} = 0 \\
&\text{C\#xx1x} && : v \cdot b_{xx} \cdot e_{1x} - w \cdot c_{xx1x} = 0 \\
&\text{C\#xxx1} && : v \cdot b_{xx} \cdot e_{x1} - w \cdot c_{xxx1} = 0 \\
&\text{D\#1xx} && : w \cdot c_{x1xx} - p \cdot d_{1xx} = 0 \\
&\text{D\#x1x} && : w \cdot c_{xx1x} - p \cdot d_{x1x} = 0 \\
&\text{D\#xx1} && : w \cdot c_{xxx1} - p \cdot d_{xx1} = 0 \\
&\text{E\#1x} && : p \cdot d_{1xx} + q \cdot b_{1x} - (r+v)e_{1x} = 0 \\
&\text{E\#x1} && : p \cdot d_{x1x} + q \cdot b_{x1} - (r+v)e_{x1} = 0
\end{aligned}
\qquad (11)
$$

2-cumomer balance equations:

$$
\begin{aligned}
&\text{B\#11} && : (v+q)b_{11} = u \cdot a_{11} \\
&\text{C\#11xx} && : v \cdot b_{11} \cdot e_{xx} - w \cdot c_{11xx} = 0 \\
&\text{C\#1x1x} && : v \cdot b_{1x} \cdot e_{1x} - w \cdot c_{1x1x} = 0 \\
&\text{C\#1xx1} && : v \cdot b_{1x} \cdot e_{x1} - w \cdot c_{1xx1} = 0 \\
&\text{C\#x11x} && : v \cdot b_{x1} \cdot e_{1x} - w \cdot c_{x11x} = 0 \\
&\text{C\#x1x1} && : v \cdot b_{x1} \cdot e_{x1} - w \cdot c_{x1x1} = 0 \\
&\text{C\#xx11} && : v \cdot b_{xx} \cdot e_{11} - w \cdot c_{xx11} = 0 \\
&\text{D\#11x} && : w \cdot c_{x11x} - p \cdot d_{11x} = 0 \\
&\text{D\#1x1} && : w \cdot c_{x1x1} - p \cdot d_{1x1} = 0 \\
&\text{D\#x11} && : w \cdot c_{xx11} - p \cdot d_{x11} = 0 \\
&\text{E\#11} && : p \cdot d_{11x} + q \cdot b_{11} - (r+v)e_{11} = 0
\end{aligned}
\qquad (12)
$$

3-cumomer balance equations:

$$\begin{aligned}
&\text{C\#111x} : v \cdot b_{11} \cdot e_{1x} - w \cdot c_{111x} = 0 \\
&\text{C\#11x1} : v \cdot b_{11} \cdot e_{x1} - w \cdot c_{11x1} = 0 \\
&\text{C\#1x11} : v \cdot b_{1x} \cdot e_{11} - w \cdot c_{1x11} = 0 \\
&\text{C\#x111} : v \cdot b_{x1} \cdot e_{11} - w \cdot c_{x111} = 0 \\
&\text{D\#111} \quad : w \cdot c_{x111} - p \cdot d_{111} = 0
\end{aligned} \qquad (13)$$

4-cumomer balance equation:

$$\text{C\#1111} : v \cdot b_{11} \cdot e_{1x} - w \cdot c_{1111} = 0 \qquad (14)$$

From the 0-cumomer balance equations (10) all fluxes can easily be calculated from the free fluxes because 0-cumomers per definition equals 1 (they are the sum of all isotopomers). Given all fluxes the values of 1-cumomer fractions can be calculated from equation (11), making it possible to calculate the values of 2-cumomer fractions from equations (12) and so on until the entire set of cumomer fractions is known. Knowing the cumomer fractions the isotopomer fractions can be computed because cumomer fractions and isotopomer fractions always have a one-to-one correspondence.

However cumomer balances in the form of equation (6) through (9) are not suited for automated computations. To achieve this it is convenient to utilize matrix notation. Using matrix notation a general cumomer balance equation for any network of unimolecular and bimolecular reactions is given by equation (15).

$$\frac{1}{2}\mathbf{x}^T \cdot \sum_i (v_i^{\rightarrow} \cdot \mathbf{Q}_i^{\rightarrow} + v_i^{\leftarrow} \cdot \mathbf{Q}_i^{\rightarrow}) \cdot \mathbf{x} + \sum_i (v_i^{\rightarrow} \cdot \mathbf{P}_i^{\rightarrow} + v_i^{\leftarrow} \cdot \mathbf{P}_i^{\leftarrow}) \cdot \mathbf{x} + \sum_i (v_i^{\rightarrow} \cdot \mathbf{P}_i^{inp}) \cdot \mathbf{x}^{inp} = 0 \qquad (15)$$

In equation (15) the state vector $\mathbf{x}$ contains the intracellular metabolite cumomer fractions arranged fist by metabolite, then by their subscript interpreted as a binary number (i.e. in binary order) and the state vector $\mathbf{x}^{inp}$ contains the substrate cumomer

fractions arranged in binary order. The cumomer fractions in the state vectors are numbered consecutively starting at 1. For the example network $\mathbf{x}$ and $\mathbf{x}^{inp}$ are given as:

$$\mathbf{x} = (b_{xx}, b_{x1}, b_{1x}, b_{11}, c_{xxxx,}, c_{xxx1,}, c_{xx1x,}, c_{xx11,}, c_{x1xx,}, c_{x1x1,}, c_{x11x,}, c_{x111,}, c_{1xxx,}, c_{1xx1,}, c_{1x1x,}, c_{1x11,}, c_{11xx,}, c_{11x1,}, c_{111x,}, c_{1111,}$$
$$d_{xxx}, d_{xx1}, d_{x1x}, d_{x11}, d_{1xx}, d_{1x1}, d_{11x}, d_{111}, e_{xx}, e_{x1}, e_{1x}, e_{11})$$
$$\mathbf{x}^{inp} = (a_{xx}, a_{x1}, a_{1x}, a_{11})$$

$v_i^{\rightarrow}$ in equation (15) denotes all the forward fluxes in the network and $v_i^{\leftarrow}$ denotes all the backward fluxes in the network. $\mathbf{Q}_i^{\rightarrow}$ is the three-dimensional bimolecular transition matrices given by

$$\mathbf{Q}_i^{\rightarrow} = \begin{pmatrix} \mathbf{Q}_{i,1}^{\rightarrow} \\ \vdots \\ \mathbf{Q}_{i,\dim \mathbf{x}}^{\rightarrow} \end{pmatrix}$$

where $\mathbf{Q}_{i,j}^{\rightarrow}$ is a $\dim \mathbf{x}$ square matrix whose elements are given by

$$(\mathbf{Q}_{i,j}^{\rightarrow})_{k,l} = \begin{cases} 1 & \text{if forward reaction } i \text{ combines cumomer number k} \\ & \text{and cumomer number } l \text{ to cumomer number } j \text{ and} \\ & \text{weight (k) + weight(l) = weight(j)} \\ 0 & \text{else} \end{cases}$$

$\mathbf{P}_i^{\rightarrow}$ in equation (15) is the unimolecular transition matrix for the intracellular metabolites (Wiechert and deGraaf 1997) where the elements are given by

$$(\mathbf{P}_i^{\rightarrow})_{j,k} = \begin{cases} 1 & \text{if forward reaction } i \text{ carries cumomer number } k \text{ over to cumomer number } j \text{ and weight(k) = weight(l)} \\ -1 & \text{if } j = k \text{ and forward reaction } i \text{ carries cumomers away from cumomer number } j \text{and weight(k) = weight(l)} \\ 0 & \text{else} \end{cases}$$

The other matrices $\mathbf{Q}_i^{\leftarrow}$, $\mathbf{P}_i^{\leftarrow}$ and $\mathbf{P}_i^{inp}$ are defined analogously. The procedure for explicit solution of equation (15) can be found in the appendix of (Wiechert, Mollney et al. 1999). Analogous to equation (15) for cumomer balances, a general equation for isotopomer balances for any network of unimolecular and bimolecular reactions can be formulated:

$$\frac{1}{2}\bar{\mathbf{x}}^{T} \cdot \sum_i (v_i^{\rightarrow} \cdot \bar{\mathbf{Q}}_i^{\rightarrow} + v_i^{\leftarrow} \cdot \bar{\mathbf{Q}}_i^{\rightarrow}) \cdot \bar{\mathbf{x}} + \sum_i (v_i^{\rightarrow} \cdot \bar{\mathbf{P}}_i^{\rightarrow} + v_i^{\leftarrow} \cdot \bar{\mathbf{P}}_i^{\leftarrow}) \cdot \bar{\mathbf{x}} + \sum_i (v_i^{\rightarrow} \cdot \bar{\mathbf{P}}_i^{inp}) \cdot \bar{\mathbf{x}}^{inp} = 0 \quad (16)$$

In equation (16) $\bar{\mathbf{x}}$ and $\bar{\mathbf{x}}^{inp}$ are the state vectors for intracellular metabolite isotopomer fractions and substrate isotopomer fractions respectably. Matrices $\bar{\mathbf{Q}}_i^{\rightarrow}$, $\bar{\mathbf{Q}}_i^{\leftarrow}$, $\bar{\mathbf{P}}_i^{\rightarrow}$, $\bar{\mathbf{P}}_i^{\leftarrow}$ and $\bar{\mathbf{P}}_i^{inp}$ are defined as for equation (15) with the exception that weight requirements for matrix elements are excluded.

From isotopomer measurements to flux distribution

13CFLUX2 determines the flux distribution of a metabolic network that produces the metabolite isotopomer distribution closest to an experimentally obtained metabolite isotopomer distribution. This is done by feeding the software specifications for the network being investigated, the measurements obtained and constraints given by the biomass equation. This information is contained in an fml-file, the structure of which is given in Figure 9. The main features of the fml-file can be constructed automatically using the software Omix as a front-end tool.

```
<?xml version="1.0" encoding="utf-8"?>
<fluxml xmlns="http://www.13cflux.net/fluxml">
        <info>
        </info>
        <reactionnetwork>
                <metabolitepools>
                        <pool atoms="2" id="A"/>
                </metabolitepools>
                <reaction bidirectional="false" id="w">*
                        <reduct cfg="C#1@1 C#2@1 C#3@1" id="C"/>
                        <rproduct cfg="C#2@1 C#3@1 C#4@1" id="D"/>
                        <rproduct cfg="C#1@1" id="F"/>
                </reaction>
        </reactionnetwork>
        <constraints>
                <net>*
                        <textual>
                        u=v;
                        v=1/2*w
                        </textual>
                </net>
                <xch>*
                </xch>
        </constraints>
        <configuration>
                <input pool="GLCU" type="isotopomer">*
                        <label cfg="110111">0.011</label>
                        <label cfg="011111">0.011</label>
                        <label cfg="101111">0.011</label>
                        <label cfg="111101">0.011</label>
                        <label cfg="111111">0.934</label>
                        <label cfg="111110">0.011</label>
                        <label cfg="111011">0.011</label>
                </input>
                <measurment>
                        <model>
                                <fluxmeasurement>
                                        <netflux id="fm_1">*
                                        <textual>upt</textual>
                                        <netflux/>
                                <fluxmeasurement/>
                                <labelingmeasurment>
                                        <group id="ms_group_1" scal="auto">*
                                        <textual>Asp[1,2,3,4]#M0,1,2,3,4<textual>
                                        </group>
                                <labelingmeasurment/>
                        </model>
                        <data>
                                <datum id="fm_1" stddev="0.2">2.3</datum>*
                                <datum id="ms_group_1" stddev="0.020" weight="0">0.513</datum>*
                        </data>
                </measurment>
                <simulation method="auto" type="auto">
                        <variables>
                                <fluxvalue flux="Anal" type="net">37.30<fluxvalue>*
                        </variables>
                </simulation>
        </configuration>
</fluxml>
```

Figure 9: Structure of a fml-file.

Based on the fml-file the programs of 13CFLUX2 can be used to determine a set of free fluxes, which enables calculation of all fluxes of the network as described in the previous section. From the calculated fluxes and the substrate isotopomer composition a computed isotopomer distribution can be found. This computed isotopomer distribution

is then compared to the experimentally obtained isotopomer measurements and using a least squared approach, the discrepancy between the two is iteratively reduced to find the flux distribution that best fit the experimental measurements. This approach is illustrated by equation (17)

$$\min_{\alpha} \| \mathbf{v}^{meas} - \mathbf{M}_{\bar{v}} \cdot \mathbf{K} \cdot \boldsymbol{\alpha} \|^2 + \| \mathbf{x}^{meas} - \mathbf{M}_{\bar{x}} \cdot f(\mathbf{x}^{inp}, \boldsymbol{\alpha}) \|^2 \quad \text{such that } \mathbf{S} \cdot \mathbf{K} \cdot \boldsymbol{\alpha} = \vec{\mathbf{0}} \tag{17}$$

In equation (17) $\boldsymbol{\alpha}$ is the set of free fluxes such that the dependent fluxes are given by $\mathbf{v} = \mathbf{K} \cdot \boldsymbol{\alpha}$. All isotopomer fractions are given by $\mathbf{x} = f(\mathbf{x}^{inp}, \boldsymbol{\alpha})$. Measured fluxes and measured isotopomer fractions are given by $\mathbf{v}^{meas}$ and $\mathbf{x}^{meas}$ respectively, whilst $\mathbf{M}_{\bar{v}}$ and $\mathbf{M}_{\bar{x}}$ are the measurement matrices for fluxes and isotopomer fractions respectively. A simplified example of a typical workflow in 13CFLUX2 is given in Figure 10. Notice that the workflow starts by operations on the fml-file. A more detailed description of the programs of 13CFLUX2 can be found in the 13CFLUX2 Reference Manual supplied with Weitzel, Nöh et al. (2013). Using 13CFLUX2 it was possible to further elucidate the metabolism of *P. fluorescens* in the work of this thesis.

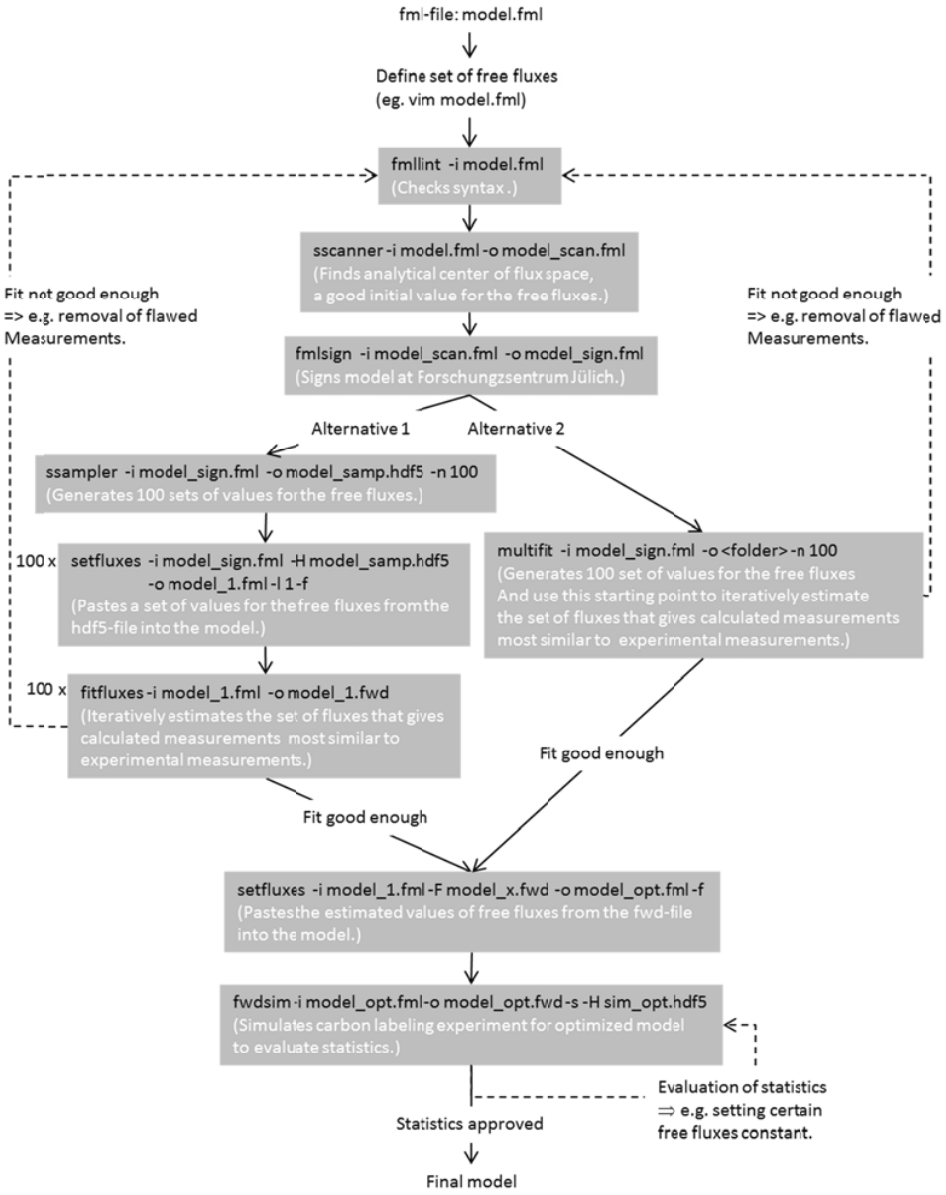


Figure 10: Simplified overview of typical workflow in 13CFLUX2. Programs of 13CFLUX2 are shown in black in grey boxes, whilst brief explanations of the programs are given in white.

2 Aims of this study

The motivation for this study was to enhance the understanding of *Pseudomonas fluorescens* metabolism in light of the *Pseudomonas* genus's importance as biocontrol agents, as cell factories and as human and plant pathogens. The aim of this work was to investigate effects of inactivation of anti-sigma factor MucA on the bacterium's metabolism through mass spectrometry based metabolomics and fluxomics. Effects of MucA inactivation was investigated both in the presence and absence of alginate biosynthesis, and when using different carbon sources. An integral part of achieving this aim, was to optimize sampling protocols for preparation of metabolite extracts and to optimize analytical MS-methods.

3 Summary of results and discussion

A metabolome study was undertaken to investigate the effects of MucA inactivation on *P. fluorescens* metabolism (Paper II). Metabolite extracts from fructose and glycerol nitrogen-limited chemostats cultivations were prepared. The strains used were *P. fluorescens* SBW25, an alginate producing *mucA*- strain and two alginate non-producing *mucA*- strains (*mucA*- Δ*algC* on fructose and *mucA*- TT*algD* on glycerol), in addition to a control strain (Δ*algC*). The metabolite extracts were analyzed by a non-targeted and a targeted GC-MS method for alkylated metabolites and a targeted reverse phase ion-pairing LC-MS/MS method. The results from the metabolome study motivated a smaller fluxome study (Paper III), aimed at complementing the results from the metabolome study and to aid in their interpretation. In the fluxome study ^{13}C-labeled fructose was used as the carbon source in nitrogen-limited chemostat cultivations of *P. fluorescens* SBW25 and the *mucA*- Δ*algC* strain. Metabolite extracts from these cultivations were analyzed by a targeted GC-MS/MS method for mass isotopomers of alkylated metabolites and the LC-MS/MS method of the metabolome study expanded to detect metabolite mass isotopomers.

A prerequisite for doing the investigation of *P. fluorescens* metabolism was developing and optimizing sample preparation protocols and analytical MS-methods. In addition to the MS-methods developed for the metabolome and the fluxome study, an isotope coded derivatization (ICD) GC-MS/MS method to diminish the effects of the inherent instability of silylated derivates was developed (Paper I).

This chapter will summarize the results from Paper I, Paper II and Paper III of this doctoral work. First, Paper I will be presented, describing the development of a quantitative GC-MS/MS method akin to the quantitative GC-MS/MS method presented in Appendix IV. Second, the overall reproducibility and suitability of the methods chosen for the metabolome study (Paper II) and the fluxome study (Paper III) of *P. fluorescens* will be presented. Third and last, the final subchapter will report the biological findings of the two studies.

3.1 Development of an isotope coded derivatizing (ICD) reagent GC-EI-MS/MS method for silylated metabolites (Paper I)

Analysis of metabolite extracts by GC-MS requires a pre-analysis derivatization step to produce volatile metabolite derivates stable enough for analysis. The two most commonly applied derivatization approaches for metabolite extracts are silylation and alkylation (Appendix V). Silylation has the advantage of being applicable to a large number of metabolite classes including alcohols, sugars, amines, acyl monophosphates and amino and non-amino organic acids (Fiehn 2008), but a complication when using this method is varying derivatizing kinetics and derivate stability for different metabolites (e.g. Paper I, Table 1). Alkylation, although only suited for amino and non-amino organic acids has the advantage of producing more stable derivates, and because the derivatization can be performed in an aqueous environment, the hydrophobic derivates can be extracted in to an organic solvent (e.g. chloroform). Extraction of the derivates to another solvent diminishes the problem of unreacted derivatization reagent in the final sample injected into the GC-MS. In a study comparing silylation and alkylation for amino and non-amino organic acids analysis of samples using alkylation was proved to be significantly more reproducible than samples using silylation (Villas-Bôas, Smart et al. 2011).

To address the reproducibility issue connected to silylation, Paper I describes the development of an ICD GC-EI-MS/MS method for silylated metabolites analogues to the published ICD GC-PCI-MS/MS method for alkylated metabolites (Appendix IV; Kvitvang, Andreassen et al. 2011). The ICD GC-EI-MS/MS method was developed using a mixture of amino acids (alanine, valine, glutamate, lysine, cysteine and tyrosine), organic acids (pyruvate, succinate, citrate and 2-oxoglutarate) and sugars (glucose, xylose, trehalose and mannitol) as a test solution. Initially both an EI ion source and a CI ion source operated in positive mode (PCI) were considered for the method, and ICD GC-MS/MS methods were developed for the test solution for both ion sources. Selectivity was found to be equally good for the two developed methods (Paper I, Supplementary Table S1 and S2), but it is possible that a PCI ion source is more appropriate for a quantitative method designed for a considerably larger number of metabolites. Because PCI is a softer ionization technique than EI, mass spectra

produced by PCI more frequently retain the quasimolecular ion and other fragment ion the high m/z range (Paper I, Figure 1). For this reason it is often easier to establish unique MRM-transitions when a PCI ion source is used, often making it more appropriate than EI for methods containing many metabolites. The reason that EI was chosen over PCI for the current ICD GC-MS/MS method was that using an EI ion source made the method significantly more sensitive. For the test solution it produced higher responses for almost 80 % of the compounds and over 10-fold higher responses for 50 % of the compounds (Paper I, Figure 2). Thus, because selectivity was equally good for the two ion sources, but because the EI ion source had superior sensitivity, it became the source of choice for the silylation ICD GC-MS/MS method.

The purpose of an ICD reagent method is having an individual labeled internal standard for each metabolite analyzed. To achieve this, the test solution was derivatized using deuterated N-methyl-N-trimethylsilyltrifluoroacetamide (d9-MSTFA), which was then spiked into a sample of the solution derivatized using ordinary MSTFA. Reanalyzing such a spiked sample nine times over an about 26 hour time period starting immediately after derivatization with subsequent spiking, produced an unexpected result: for the three earliest time points the unlabeled compounds showed a general increase in response and the labeled compounds showed a general decrease in response. Then for the six later time points the changes in responses for the unlabeled and labeled compounds followed the same trend, either both increasing or decreasing (Paper I, Figure 3 and Supplementary Figure S1). It is known that hydrolyzation and scrambling reactions occur more frequently using MSTFA as a silylation agent than using e.g. N-methyl-N-(t-butyldimethylsilyl)trifluoroacetamide MTBSTFA (Huang and Regnier 2008). Unfortunately MTBSTFA was not an option for the current method as sugars are not silylated when using MTBSTFA. It was not possible to determine the exact mechanism behind the initial inverse behavior of unlabeled and labeled compounds, but it was possible to avoid this inverse behavior by prolonging the incubation time prior to spiking the MSTFA derivatized sample with the d9-MSTFA derivatized sample (Paper

I, Figure 4). This shows that it is important to optimize the derivatization procedure for the specific samples to be analyzed if the developed ICD GC-MS/MS method is to be used.

In spiked samples where the inverse behavior of unlabeled and labeled compounds was not occurring, the effect of individual correction and group correction was compared to assess if the developed method constituted an improvement. Individual correction denotes correcting a compound's response by the response of its labeled counterpart, whilst group correction denotes correcting the response of one compound belonging to a specific metabolite group (e.g. organic acids, amino acids or sugars) by the response of one specific labeled compound belonging to that group. The two correction strategies were tested by comparing average relative standard deviations for the metabolite groups in the test solution for six consecutive runs of a spiked sample. This was done both in the presence and absence of biological matrices (serum and urine) in the sample. The results showed that although group correction did improve precision by reducing the average standard deviation for sugars, it did not always do so for organic acids. For amino acids the results showed that group correction produced no average improvement. Individual correction on the other hand worked as well as group correction for sugars, reproducibly improved precision for organic acids and in addition it also worked well for amino acids (Paper I, Figure 5). Individual correction enabled by using the developed ICD GC-EI-MS/MS thus gives significantly improved precision compared to group correction producing standard deviations of about 10 % or lower for all compounds in the test solution when reanalyzed six times over an about 15 h time period.

3.2 Overall reproducibility and suitability of sample preparation and data acquisition for the metabolome study and the fluxome study of P. fluorescens

Before initiating the metabolic and flux investigations of *P. fluorescens* SBW25 in this doctoral work, methods for sample preparation and sample analysis found in literature had to optimized for the biological system at hand. Adapting and building on the approaches found in literature, datasets from analysis of metabolite extracts of *P.*

fluorescens SBW25 could be generated. The current subchapter will briefly present the sample analysis approaches used for the metabolome study and the fluxome study, and results from principal component analysis (PCA) (Esbensen 2000) of the generated datasets will be used to illustrate the overall reproducibility and suitability of the developed methods.

3.2.1 Approach of metabolome study (Paper II)

In the metabolome study (Paper II) effects of different carbon sources, effects of alginate production and effect of inactivation of anti-sigma factor MucA on the metabolome of *P. fluorescens* SBW25 was studied. To this end several cultivations of the strains in Table 4 were performed.

Table 4: The different cultivation in the metabolome study.

Carbon source	Strain	MucA inactivation [√]	Alginate production [√]	Number of cultivations
Fructose	*P. fluorescens* SBW25			3
	P. fluorescens SBW25 Δ*algC*			2
	P. fluorescens SBW25 *mucA*-	√	√	3
	P. fluorescens SBW25 *mucA*- Δ*algC*	√		2
Glycerol	*P. fluorescens* SBW25			2
	P. fluorescens SBW25 *mucA*-	√	√	2
	P. fluorescens SBW25 *mucA*- TT*algD*	√		2

The published work focuses on metabolic profiling results from analysis of metabolite extracts using a reverse phase tributylamine ion paring LC-MS/MS method (Luo, Groenke et al. 2007) and a GC-MS method where samples were derivatized using methyl chloroformate (MCF) (Villas-Bôas, Delicado et al. 2003). These two methods are suited for phosphorylated metabolites (LC-MS/MS-method) and amino and non-amino organic acids (GC-MS method). However these two methods were not the only analytical techniques tested for the study as initially samples from preliminary test fermentations were also derivatized using MSTFA. Unfortunately GC-MS analysis in scan mode for the MSTFA derivatized samples showed that the acquired chromatograms were not very informative as they were heavily dominated by a single peak of coeluting and probably also comaximizing monosaccharaides, with other compound groups often being under the detection limit (unpublished results). Because

of this, and because sugars were not of particular interest to the study, the silylation GC-MS method was abandoned, and this is also the reason why the developed ICD GC-EI-MS/MS method (Paper I) was not expanded to more metabolites and included for use in this study.

The MCF GC-MS method used for detection of 25 amino acids and organic acids was operated in scan mode, and in addition to amino acids and organic acids 123 peaks of unknown origin were also quantified in the acquired mass spectrometric data. These unknowns were selected for quantitation by analyzing samples from the preliminary test fermentations using the freeware AMDIS (Automated Mass Deconvolution and Identification System) (Halket, Przyborowska et al. 1999). An advantage of AMDIS compared to the software supplied by the vendor (Chemstation by Agilent) is that it is able to separate signals originating from coeluting compounds in a chromatogram, a process called deconvolution. The 123 unknowns were selected by analyzing metabolite extracts from each of the four different strains grown on fructose with an AMDIS library composed of the 25 known amino acids and organic acids. For each strain the 100 most abundant non-identified deconvoluted peaks were selected and added to the library, and after eliminating several contaminant peaks and peaks that where not reproducibly detected, this resulted in 123 different unknowns compounds. These unknowns where then added to an Agilent Chemstation quantitative library along with the 25 known compounds, and joint indentification by AMDIS and Chemstation using Agilent Deconvolution Reporting Software (DRS) was used as a requirement for quantification.

The quantitative MCF GC-MS data from analysis of metabolite extracts for the cultivations of the metabolome study were normalized using samples from a reference cultivation included in each sequence. Therefore, although the responses for the unknowns could not be converted to concentrations, they were all normalized to the same reference state, eliminating instrument drift when interpretation the data. Ultimately PCA was performed on the normalized data for the 123 unknowns to look for a limited set of unknowns that were especially affected by changing carbon sources, by alginate synthesis or by MucA inactivation. The PCA scores plot showed some

clustering of cultivations depending on carbon source, alginate synthesis and MucA activity, but unfortunately the loadings plot indicated that no limited set of unknowns stood out as the cause of this separation (Paper II, Supplementary Figure S6). Because of this, no further work was undertaken to identify specific unknowns.

The LC-MS/MS and the MCF GC-MS methods used in the metabolic profiling part of the metabolome study were capable of detecting 42 common phosphorous containing metabolites and 25 amino acids and organic acids respectively. PCA was performed on the LC-MS/MS and the MCF GC-MS data for all 16 cultivations. The scores plot from this analysis showed four distinct groupings: the alginate producing strain on fructose, the non-alginate producing strains of fructose, the alginate producing strain on glycerol and the non-alginate producing strains on glycerol (Paper II, Figure 1a). Out of these four groups, the two non-alginate producing groups were closest together, and a separate PCA was performed on these two groups eliminating the variability introduced by alginate production. This analysis also produced four distinct groupings: the strains with an active MucA on fructose, the strain with an inactive MucA on fructose, the strain with an active MucA on glycerol and the strain with an inactive MucA on glycerol (Paper II, Figure 1b). The within group proximity in these PCAs shows the good reproducibility of the cultivation, sample preparation, sample analysis and data processing choices made in this study, whilst the between group distance show the appropriateness of the experimental design in answering the question of how does carbon source, alginate production and MucA activity affect the metabolome of *P. fluorescens*.

3.2.2 Approach of fluxome study (Paper III)

The question investigated in the fluxome study (Paper III) was how does MucA inactivation effect *P. fluorescens* central carbon metabolism apart from causing alginate biosynthesis. To address this issue an experimental design carbon labeling experiment (CLE) of the wild type growing on fructose was first performed to find the optimal fructose isotopomer composition for the main CLE. In the main CLE two cultivations, one of the wild type and one of the *mucA*- Δ*algC* strain, were conducted. The

metabolite extracts generated were analyzed with a MCF GC-MS/MS method developed for the purpose, and an adaptation of the LC-MS/MS methods used in the previous metabolome study. For both methods the detected compounds were limited to those present in the model of central carbon metabolism used, and the number of MRM-transitions were expanded to account for all possible metabolite mass isotopomers (Choi and Antoniewicz 2011). PCA was performed on the LC-MS/MS data and the MCF GC-MS/MS data of the nine samples of the three CLE cultivations. As for the metabolome study, the results from PCA indicated good reproducibility for the sample preparation, sample analysis and data processing choices made, and the between group distance showed the appropriateness of the experimental design in answering the biological question addressed (Paper III, Supplementary Figure S2).

3.3 Biological findings of the metabolome study and the fluxome study of P. fluorescens

3.3.1 Findings of metabolome study (Paper II)

Nitrogen-limited chemostat cultivations were performed for the strains listed in Table 4 using either fructose or glycerol as the carbon source. Metabolite extracts were prepared from samples taken from these cultivations at steady state. The aim of the metabolome study was to elucidate how MucA inactivation affects the metabolome of *P. fluorescens*, both in the presence and absence of alginate biosynthesis, and to see how the bacterium adopts to different carbon sources. For cultivations on fructose the wild type strain, the alginate producing *mucA*- strain, the alginate non-producing *mucA*- Δ*algC* strain and a Δ*algC* control strain were used. For cultivations on glycerol the wild type strain, the alginate producing *mucA*- strain and the alginate non-producing *mucA*- TT*algD* strain were used. The difference between the two double deletion mutants is that the *mucA*- Δ*algC* strain does not produce alginate because the single gene *algC* is inactivated, whilst the *mucA*- TT*algD* strain does not produce alginate because the entire *alg* operon is inactivated. The cultivation data from these chemostats showed that the *mucA*- strains (correcting for the proportion of the carbon source that is used for alginate synthesis by the *mucA*- strain), have an about 40% decrease in fructose uptake rate and an about 20% decreased glycerol uptake rate compared to the wild type (Paper II, Table 1). The following text will present the highlights from the metabolome datasets

(concentration of metabolites detected for all strains on the two carbon sources can be found in Supplementary Tables S1 and S2 of Paper II)

Effect of MucA inactivation when growing on fructose

Comparing metabolite concentrations for the two *mucA*- strains to those of the wild type, concentrations can differ little from the wild type concentrations (situation 1), differ in the same way relative to the wild type concentrations (situation 2) or change in opposite directions relative to the wild type concentrations (situation 3). In Figure 5 of Paper II concentrations for the mutant strains relative to the wild type when growing on fructose are superimposed on a map of central carbon metabolism for *P. fluorescens*. The Δ*algC* control strain differs little from the wild type, and the two *mucA*- mutants also differ little when it comes to most amino acids and organic acids (situation 1). Little change in amino acid pools is expected as the biomass production, and therefore also the requirement for precursors for protein synthesis, is the same for all strains. There are however a few exceptions to the relative unchanged concentrations of amino acids and organic acids for the *mucA*- strains: the *mucA*- Δ*algC* strain has a two-fold increased succinate concentration, and the *mucA*- strain has a two-fold decreased tyrosine and glutamate concentration.

For the other detected metabolites, predominantly phosphorylated ones, a limited set stands out as differing in the same way for both *mucA*- strains (situation 2). This set consists of fructose 1-phosphate and glyceraldehyde 3-phosphate, whose concentrations are strongly increased, 6-phoshphogluconate, whose concentration is strongly decreased, and all the three adenine nucleotides. For the adenine nucleotides the ATP concentration is strongly decreased for the *mucA*- mutants, ADP concentration is slightly increased and AMP concentration is strongly increased. Because the result is the same for all *mucA*- strains this is clearly an effect of MucA inactivation.

In contrast to the adenine nucleotides, the guanine nucleotide concentrations stand out by differing in opposite directions for the two *mucA*- strains compared to the wild type (situation 3). For the *mucA*- Δ*algC* strain all three guanine nucleotide concentrations are decreased, whilst for the *mucA*- strain all three guanine nucleotide concentrations are

increased (albeit the increase in GTP is small). Also differing in opposite directions for the two *mucA*- strains is the concentration of GDP-mannose which is not detected for the *mucA*- Δ*algC* strain but strongly increased in concentration for *mucA*- strain relative to the wild type. The increase in guanine nucleotide concentrations and the increase in GDP-mannose concentration for the alginate producing *mucA*- strain are not surprising as GTP is utilized in the synthesis of GDP-mannose, which is an intermediate in alginate biosynthesis.

Effect of MucA inactivation when growing on glycerol

In Figure 6 of Paper II concentrations for the *mucA*- TT*algD* strain and the *mucA*- strain relative to the wild type when growing on glycerol are superimposed on a map of central carbon metabolism for *P. fluorescens*. As for growth on fructose, concentrations of amino acids and organic acids generally differ little for the two *mucA*- strains compared to the wild type (situation 1). Also in accordance with results for growth on fructose, tyrosine and glutamate are again exceptions to this general trend, with tyrosine concentration being decreased for the *mucA*- strain and glutamate concentration being decreased for both mutant strains. Little variation in concentrations of amino acids and organic acids have also been reported for *P. aeruginosa* in response to MucA inactivation and utilization of different carbon sources (Frimmersdorf, Horatzek et al. 2010). Another phenomenon that is reoccurring for growth on glycerol is a change in adenine nucleotide concentrations similar for both *mucA*- strains relative to the wild type with a decreased ATP concentration, a slightly increased ADP concentration and a strongly increased AMP concentration (situation 2). When it comes to the concentrations of guanine nucleotides the change in all three of these are, as for fructose cultivations, not similar for the two *mucA*- strains (situation 3). On glycerol GTP is strongly decreased for the *mucA*- TT*algD* strain and slightly increased for the *mucA*- strain, as it was on fructose. In contrast to growth on fructose GDP and GMP concentrations are increased for both *mucA*- strains, although more strongly for the alginate producing strain.

Also specific for growth on glycerol is a strong increase in GDP-mannose (GDP-M) and strong increase in 6-phosphogluconate for both *mucA*- strains (situation 2), and a

general lower concentration of glycolytic and pentose phosphate pathway metabolites for the *mucA*- TT*algD* strain compared to the wild type with the opposite being true for the *mucA*- strain (situation 3). The increase in GDP-M for the *mucA*- TT*algD* strain is surprising as it should not be able to produce this metabolite due to an inability to produce mannose 6-phosphate.

Effects of changing the carbon source on the metabolome of the wild type strain and the *mucA*- strain

The wild type and the *mucA*- strain are the only two strains that were grown on both carbon sources, so for these strains their metabolome on the two carbon sources can be compared directly. Concentrations of the detected metabolites for these strains on both carbon sources can be found in Table 2 of Paper II, and concentrations for the wild type and the *mucA*- strain growing on glycerol relative to growth on fructose can be found superimposed on a map of central carbon metabolism in Supplementary Figure S7.

The results show that as for different mutants relative to the wild type on one carbon source, changing the carbon source for the wild type and the *mucA*- strain has little effect on most amino acid and organic acid concentrations. An exception to this general trend is lactate, malate, succinate, leucine and glutamine which all have a decreased concentration for both strains when growing on glycerol. For other metabolites changing carbon source has an effect on metabolite concentrations and for the metabolites close to the carbon source, this effect is the same for both wild type and the *mucA*- strain: fructose 1-phosphate and fructose 1,6-bisphophate concentrations are elevated for both strains when growing on fructose, and glycerol 3-phosphate (Gol3P) and dihydroxyacetone phosphate (DHAP) concentrations are elevated for both strains when growing on glycerol. That changes in metabolite concentrations occur within a subset of the metabolome closely linked to the nutrient perturbation has also been reported in previous metabolome studies (van der Werf, Overkamp et al. 2008; Yuan, Doucette et al. 2009).

For growth on glycerol it is not only Gol3P and DHAP concentrations that are increased relative to growth on fructose, but also 3-phosphoglycerate (3PG) and

phosphoenolpyruvate (PEP), which are further removed from the carbon source. Although increased metabolite concentrations cannot be directly linked to an increased flux through a specific metabolic path, it does not seem unlikely that there is an increased glycolytic flux irrespective of strain when growing on glycerol. It is also noteworthy that for metabolites of the pentose phosphate pathway, changing the carbon source has opposite qualitative effects on the wild type and the *mucA*- strain: the concentrations of the involved metabolites are decreased for the wild type on glycerol and increased for the *mucA*- strain on glycerol. This opposite effect is also true for GDP-M.

In conclusion the two most striking findings of the metabolome study are the alginate production dependent change in the guanine nucleotide concentrations for the *mucA*- strains relative to the wild type, and the changes in adenine nucleotide concentrations for the *mucA*- strains relative to the wild type independent of alginate synthesis. From the adenine nucleotide concentrations the energy charge (EC = (ATP + 0.5ADP) / (ATP + ADP + AMP)) (Atkinson 1968), a measure of the energy state of the cells, can be calculated. For the wild type the EC is 0.58 and 0.56 on fructose and glycerol respectively, for the *mucA*- strain the EC is 0.17 and 0.13 on fructose and glycerol respectively and for the double deletion mutants the EC is 0.13 and 0.20 on fructose (*mucA*- Δ*algC*) and glycerol (*mucA*- TT*algD*) respectively. It is thus clear that the energy charge of *mucA*- strains are reduced compared to the wild type. Correcting for carbon source consumption to alginate production, all *mucA*- strains have a decreased carbon source uptake compared to the wild type (an about 40% decrease on fructose and an about 20% decrease on glycerol). Because biomass production is constant, this leads to a higher biomass yield on substrate for the *mucA*- strains (again correcting for alginate production for the producing *mucA*- strain). The high biomass yield for strains with low EC seems contradictory as one would expect cells with low EC to favor catabolism over anabolism as a response to the low EC. The lack of such a response, and the similar ECs on both carbon sources for the wild type and the *mucA*- strain indicate that the absolute concentrations that the EC is calculated from, are perhaps just as important as their relative concentrations (i.e. the EC), when it comes to assessing viability and growth. All of the EC values calculated in this study are low compared to

the belief that the EC of growing bacterial cell should be higher than 0.8 and that cells become metabolically inert if the EC drops below 0.5 (Chapman, Fall et al. 1971). It is possible that the low ECs of this study are in part a product of sample preparation and analysis as e.g. an overview of several microbial studies using the cold methanol extraction protocol used it this study produced EC varying from 0.16 to 0.92 (Bolten, Kiefer et al. 2007), but it is worth pointing out that the thresholds of 0.8 and 0.5 are also experimentally determined and as such subject to experimental errors.

3.3.2 Findings of fluxome study (Paper III)

The aim of the fluxome study was to determine changes in the intracellular metabolic flux distribution of *P. fluorescens* caused by inactivation of anti-sigma factor MucA, both to generate insight into metabolism complimentary to that gained during the metabolome study and to help explain some of the findings. To determine fluxes in central carbon metabolism of *P. fluorescens* nitrogen-limited CLE chemostat cultivations were performed for the wild type and the *mucA*- Δ*algC* strain growing on fructose, and metabolite mass isotopomer datasets were generated. The mass isotopomer datasets were then used along with known fluxes in and out of the cells to determine intracellular fluxes using the simulation software 13CFLUX2.

The *mucA*- Δ*algC* mutant has a primary metabolism flux distribution distinctly different from the wild type

The flux distributions determined for the wild type and the *mucA*- Δ*algC* strain are visualized in Figure 1 and Figure 2 of Paper III respectively. The results show that all net fluxes proceed in the same direction for the two strains, and that the same main route for fructose uptake is utilized for both strains (fructose is shuttled to fructose 1,6-bisphosphate via fructose 1-phosphate) (Figure 3a, Paper III). The results also show that at important branch points in metabolism there are distinct differences for the two strains. One such distinction occurs at 6-phosphogluconate (6PGn), where one efflux goes to the Entner – Doudoroff pathway (EDP) and the other efflux goes to pentose phosphate pathway (PPP). For the wild type the major proportion of the influx goes to EDP and the minor proportion of the influx goes to PPP, whilst for the *mucA*- Δ*algC* strain the minor proportion of the influx goes to EDP and the major proportion of the influx goes to PPP (Figure 3b, Paper III). Another important distinction occur at

isocitrate (ICit) in the tricarboxylic acid cycle (TCA) where the efflux is either to glyoxylate thereby utilizing the glyoxylate shunt or to 2-oxoglutarate (OGA) as a continuum of TCA. Here the wild type does not seem to utilize the glyoxylate shunt, whilst the *mucA*- Δ*algC* strain does (Figure 3c, Paper III). There also seems to be some strain specific activity in the anaplerotic reactions, but it is difficult to pinpoint the exact difference between the wild type and the *mucA*- Δ*algC* strain as many of the determined anaplerotic fluxes are accompanied by large uncertainties (Figure 3d, Paper III).

In Figure 4a of Paper III all reactions of *P. fluorescens* metabolism involving ATP and producing reducing power are summed separately and displayed (the individual ATP, NADH and NADPH reactions are displayed in Figure 4b, c and d respectively). The figure show that NADPH production and ATP produced by substrate level phosphorylation is similar for the two strains. In contrast NADH production is different for the two strains with the *mucA*- Δ*algC* strain producing significantly less than the wild type. The *mucA*- Δ*algC* strain produce less NADH because fructose uptake is less for this strain, an a large proportion of the assimilated fructose is shuttled through PPP. In addition the entire flux that enters TCA does not complete TCA because the Glx shunt is utilized by the *mucA*- Δ*algC* strain to a significant extent.

Comparison of fluxome data to the metabolome data

Comparing metabolite concentrations and the determined fluxes reveals that differences in fluxes through specific metabolic pathways for the wild type and the *mucA*- Δ*algC* strain often coincide with differences in concentration for some of the pathway specific metabolites. Examples of this include the elevated concentration of F1P coinciding with a decreased fructose uptake, and the decreased concentration of 6PGn coinciding with an increased flux through PPP for the *mucA*- Δ*algC* strain. The increased flux through PPP also coincided with an increase in the metabolites of PPP.

An important finding of the metabolome study was the decreased EC for all *mucA*- mutants studied. The fluxome study showed that the wild type produce similar amounts of ATP as the *mucA*- Δ*algC* strain through substrate level phosphorylation, but that the NADH production is higher. The higher NADH production in theory (assuming similar

P/O ratios for both strains) enables increased ATP production through oxidative phosphorylation which would explain the higher EC of the wild type strain. Following this, perhaps oversimplified, line of though, the *mucA*- Δ*algC* strain could increase its fructose uptake to alleviate its low EC if it was perceived as a stress to the cell. This again points towards the possibility that the low ECs detected for *mucA*- strains in this study is not perceived as stressful for the cells. The results from the fluxome study presented here, show how a fluxome study can complement and help to explain results from a metabolome study.

4 Concluding remarks

In this doctoral work a quantitative GC-EI-MS/MS method addressing the inherent instability of silylated metabolites was developed (Paper I). The method relies on utilizing an isotope coded derivatization (ICD) reagent, deuterated N-methyl-N-trimethylsilyltrifluoroacetamide (d9-MSTFA), to create individual labeled internal standards for MSTFA derivatized metabolites. Tested on a solution of standards, the method produced increased precision both in the presence and absence of biological matrices (serum and urine), compared to using one compound from a specific metabolite group as the internal standard for that metabolite group. A drawback of the method is that spiking of the MSTFA derivatized samples with d9-MSTFA derivatized standards, can lead to initial instability of the derivates if the derivatization protocol is not optimized for the specific sample type. In addition the high cost of d9-MSTFA might preclude extensive use of the method. The preparation of metabolic extracts from bacterial, plant or mammalian cell cultures require several processing steps (i.e. quenching, extraction and further processing) and systematic and random errors can be introduced in all of these steps. If possible from a cost perspective and in light of the metabolites of interest, isotope dilution mass spectrometry (IDMS) would be a better choice for such samples, whilst a ICD-GC-MS method is more suited for samples requiring less processing before analysis (e.g. biological fluids).

The main topic of this doctoral work was mass spectrometry based metabolomics for *P. fluorescens* investigating pleiotropic effects of inactivation of anti-sigma factor MucA. The effects of inactivation were investigated in the presence and absence of alginate production, and when using two different carbon sources. A large metabolome study (Paper II) using a collection of strains on two different carbon sources was undertaken, and based on the results from this study a limited fluxome study (Paper III) was performed. In the fluxome study two strains on one carbon source were investigated to generate result that would not only complement the metabolomics results, but also aid in their interpretation. The results from the metabolome and the fluxome studies can, in addition to being interesting in their own right, act as stepping stones for developing experiments to further understand *P. fluorescens* metabolism. Experiments to do this

could include metabolome and fluxome studies using carbon-limited, instead of nitrogen-limited, chemostats and also studying unlimited growth in batch cultivations. It would be interesting to see which of the characteristics from the current studies would persist, and which new characteristics would emerge. Another interesting experiment would be a fluxome study especially designed to resolve fluxes of the anaplerotic reactions, or to resolve other fluxes that were poorly determined in the current study. The design of such a study would involve optimizing the composition of ^{13}C-labeled fructose for the specific fluxes of interest. Precision and accuracy of flux estimates would also benefit from more sensitive MS-instruments or MS-methods, and from metabolite extracts where a higher concentration of analytes could be attained, with the extracts still being representative for the culture from which they were sampled.

The work presented in this Doctoral thesis constitutes a significant contribution to metabolomics research for *P. fluorescens*, which is not as extensive as research for the model organisms *E. coli* and *B. subtilis*. The published metabolome dataset is also, to the best of our knowledge, the most comprehensive dataset of metabolite concentrations for *P. fluorescens* to date.

5 References

Antoniewicz, M. R., J. K. Kelleher, et al. (2007). "Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions." Metabolic Engineering **9**(1): 68-86.

Antunes, L. C. M., E. T. Arena, et al. (2011). "Impact of Salmonella Infection on Host Hormone Metabolism Revealed by Metabolomics." Infection and Immunity **79**(4): 1759-1769.

Atkinson, D. E. (1968). "Energy charge of adenylate pool as a regulatory parameter. Interaction with feedback modifiers." Biochemistry **7**(11): 4030-4034.

Baynham, P. J. and D. J. Wozniak (1996). "Identification and characterization of AlgZ, an AlgT-dependent DNA-binding protein required for *Pseudomonas aeruginosa* algD transcription." Molecular Microbiology **22**(1): 97-108.

Bennett, B. D., E. H. Kimball, et al. (2009). "Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*." Nature Chemical Biology **5**(8): 593-599.

Bolten, C. J., P. Kiefer, et al. (2007). "Sampling for metabolome analysis of microorganisms." Analytical chemistry **79**(10): 3843-3849.

Borgos, S. E., S. Bordel, et al. (2013). "Mapping global effects of the anti-sigma factor MucA in *Pseudomonas fluorescens* SBW25 through genome-scale metabolic modeling." BMC Systems Biology **7**(19).

Brauer, M. J., J. Yuan, et al. (2006). "Conservation of the metabolomic response to starvation across two divergent microbes." Proceedings of the National Academy of Sciences of the United States of America **103**(51): 19302-19307.

Britten, R. J. and F. T. McClure (1962). "The amino acid pool in *Escherichia coli*." Bacteriological Reviews **26**(3): 292-335.

Bruheim, P., H. F. N. Kvitvang, et al. (2013). "Stable isotope coded derivatizing reagents as internal standards in metabolite profiling." Journal of chromatography A **1296**(0): 196-203.

Bucher, G. E. and J. M. Stephens (1957). "A disease of grasshoppers caused by the bacterium *Pseudomonas aeruginosa* (Schroeter) Migula." Canadian Journal of Microbiology **3**(4): 611-625.

Buziol, S., I. Bashir, et al. (2002). "New bioreactor-coupled rapid stopped-flow sampling technique for measurements of metabolite dynamics on a subsecond time scale." Biotechnology and Bioengineering **80**(6): 632-636.

Chapman, A. G., L. Fall, et al. (1971). "Adenylate energy charge in *Escherichia coli* during growth and starvation." Journal of Bacteriology **108**(3): 1072-1086.

Choi, J. and M. R. Antoniewicz (2011). "Tandem mass spectrometry: A novel approach for metabolic flux analysis." Metabolic Engineering **13**(2): 225-233.

Cipollina, C., A. ten Pierick, et al. (2009). "A comprehensive method for the quantification of the non-oxidative pentose phosphate pathway intermediates in *Saccharomyces cerevisiae* by GC-IDMS." Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences **877**(27): 3231-3236.

Cook, A. M., E. Urban, et al. (1976). "Measuring the concentration of metabolites in bacteria." Analytical biochemistry **72**(1-2): 191-201.

REFERENCES

Cvijovic, M., R. Olivares-Hernandez, et al. (2010). "BioMet Toolbox: genome-wide analysis of metabolism." Nucleic Acids Research **38**: W144-W149.

Darzins, A., S. K. Wang, et al. (1985). "Clustering of mutations affecting alginic acid biosynthesis in mucoid *Pseudomonas aeruginosa*." Journal of Bacteriology **164**(2): 516-524.

del Castillo, T., J. L. Ramos, et al. (2007). "Convergent peripheral pathways catalyze initial glucose catabolism in *Pseudomonas putida*: Genomic and flux analysis." Journal of Bacteriology **189**(14): 5142-5152.

Desai, J. D. and I. M. Banat (1997). "Microbial production of surfactants and their commercial potential." Microbiology and Molecular Biology Reviews **61**(1): 47-64.

Droste, P., S. Miebach, et al. (2011). "Visualizing multi-omics data in metabolic networks with the software Omix-A case study." Biosystems **105**(2): 154-161.

Durham, D. R. and P. V. Phibbs (1982). "Fractionation and characterization of the phosphoenolpyruvate : fructose 1-phoshotransferase system from *Pseudomonas aeruginosa*." Journal of Bacteriology **149**(2): 534-541.

Ebert, B. E., A.-L. Lamprecht, et al. (2012). "Flux-P: Automating Metabolic Flux Analysis." Metabolites **2**: 872-890.

Ellis, R. J., T. M. Timms-Wilson, et al. (2000). "Identification of conserved traits in fluorescent pseudomonads with antifungal activity." Environmental Microbiology **2**(3): 274-284.

Elrod, R. P. and A. C. Braun (1941). "A phytopathogenic bacterium fatal to laboratory animals." Science **94**(2448): 520-521.

Elrod, R. P. and A. C. Braun (1942). "*Pseudomonas aeruginosa*; its role as a plant pathogen." Journal of Bacteriology **44**(6): 633-644.

Erickson, J. W. and C. A. Gross (1989). "Identification of the sigma-E subunit of *Escherichia coli* RNA-polymerase - a 2nd alternate sigma factor involved in high-temperature gene expression." Genes & Development **3**(9): 1462-1471.

Esbensen, K. H. (2000). Multivariate data analysis in practice: An introduction to multivariate data analysis and experimental design, CAMO ASA**:** 19-97.

Faijes, M., A. E. Mars, et al. (2007). "Comparison of quenching and extraction methodologies for metabolome analysis of *Lactobacillus plantarum*." Microbial Cell Factories **6**(27).

Fiehn, O. (2008). "Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry." Trac-Trends in Analytical Chemistry **27**(3): 261-269.

Fiehn, O., B. Kristal, et al. (2006). "Establishing reporting standards for metabolomic and metabonomic studies: a call for participation." OMICS A jouranl of integrative biology **10**(2): 158-163.

Firoved, A. M. and V. Deretic (2003). "Microarray analysis of global gene expression in mucoid *Pseudomonas aeruginosa*." Journal of Bacteriology **185**(3): 1071-1081.

Frederiksen, B., C. Koch, et al. (1997). "Antibiotic treatment of initial colonization with *Pseudomonas aeruginosa* postpones chronic infection and prevents deterioration of pulmonary function in cystic fibrosis." Pediatric Pulmonology **23**(5): 330-335.

REFERENCES

Frimmersdorf, E., S. Horatzek, et al. (2010). "How *Pseudomonas aeruginosa* adapts to various environments: a metabolomic approach." Environmental Microbiology **12**(6): 1734-1747.

Fuhrer, T., E. Fischer, et al. (2005). "Experimental identification and quantification of glucose metabolism in seven bacterial species." Journal of Bacteriology **187**(5): 1581-1590.

Gerigk, M., R. Bujnicki, et al. (2002). "Process control for enhanced L-phenylalanine production using different recombinant *Escherichia coli* strains." Biotechnology and Bioengineering **80**(7): 746-754.

Gross, J. H. (2004). 1 Introduction. Mass Spectrometry. Leipzig, Germany, Springer**:** 1-10.

Haas, D. and C. Keel (2003). "Regulation of antibiotic production in root-colonizing *Pseudomonas* spp. and relevance for biological control of plant disease." Annual Review of Phytopathology **41**: 117-153.

Halket, J. M., A. Przyborowska, et al. (1999). "Deconvolution gas chromatography mass spectrometry of urinary organic acids - Potential for pattern recognition and automated identification of metabolic disorders." Rapid Communications in Mass Spectrometry **13**(4): 279-284.

Harris, P. A. and P. W. Stahlman (1996). "Soil bacteria as selective biological control agents of winter annual grass weeds in winter wheat." Applied Soil Ecology **3**(3): 275-281.

Hasunuma, T., T. Sanda, et al. (2011). "Metabolic pathway engineering based on metabolomics confers acetic and formic acid tolerance to a recombinant xylose-fermenting strain of *Saccharomyces cerevisiae*." Microbial Cell Factories **10**(2).

Hu, Q. Z., R. J. Noll, et al. (2005). "The Orbitrap: a new mass spectrometer." Journal of Mass Spectrometry **40**(4): 430-443.

Huang, X. and F. E. Regnier (2008). "Differential Metabolomics Using Stable Isotope Labeling and Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry." Analytical chemistry **80**: 107-114.

Jain, S. and D. E. Ohman (2004a). 2. Alginate biosynthesis. *Pseudomonas* biosynthesis of macormolecules and molecular metabolism. J.-L. Ramos. New York, Kluwer Academic / Plenum Publishers. **3:** 63-71.

Jain, S. and D. E. Ohman (2004b). 2. Alginate biosynthesis. *Pseudomonas* biosynthesis of macormolecules and molecular metabolism. J.-L. Ramos. New York, Kluwer Academic / Plenum Publishers. **3:** 55-56.

Jensen, N. B. S., K. V. Jokumsen, et al. (1999). "Determination of the phosphorylated sugars of the Embden-Meyerhof-Parnas pathway in *Lactococcus lactis* using a fast sampling technique and solid phase extraction." Biotechnology and Bioengineering **63**(3): 356-362.

Jozefczuk, S., S. Klie, et al. (2010). "Metabolomic and transcriptomic stress response of Escherichia coli." Molecular Systems Biology **6**(364).

Kanu, A. B., P. Dwivedi, et al. (2008). "Ion mobility-mass spectrometry." Journal of Mass Spectrometry **43**(1): 1-22.

Khan, M. W. A. and M. Ahmad (2006). "Detoxification and bioremediation potential of a *Pseudomonas fluorescens* isolate against the major Indian water pollutants." Journal of environmental science and health. Part A, Toxic/hazardous substances & environmental engineering **41**(4): 659-674.

REFERENCES

Kol, S., M. E. Merlo, et al. (2010). "Metabolomic Characterization of the Salt Stress Response in *Streptomyces coelicolor*." Applied and Environmental Microbiology **76**(8): 2574-2581.

Kvitvang, H. F. N., T. Andreassen, et al. (2011). "Highly Sensitive GC/MS/MS Method for Quantitation of Amino and Nonamino Organic Acids." Analytical Chemistry **83**(7): 2705-2711.

Lessie, T. G. and P. V. J. Phibbs (1984). "Alternative pathways of carbohydrate utilization in pseudomonads." Annual Review of Microbiology **38**: 359-387.

Letisse, F. and N. D. Lindley (2000). "An intracellular metabolite quantification technique applicable to polysaccharide-producing bacteria." Biotechnology Letters **22**(21): 1673-1677.

Luo, B., K. Groenke, et al. (2007). "Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography-mass spectrometry." Journal of chromatography A **1147**: 153-164.

Ma, S., U. Selvaraj, et al. (1998). "Phosphorylation-independent activity of the response regulators AlgB and AlgR in promoting alginate biosynthesis in mucoid *Pseudomonas aeruginosa*." Journal of Bacteriology **180**(4): 956-968.

Majewski, R. A. and M. M. Domach (1990). "Simple constrained-optimization view of acetate overflow in *E.coli*." Biotechnology and Bioengineering **35**(7): 732-738.

Malhotra, S., L. A. Silo-Suh, et al. (2000). "Proteome analysis of the effect of mucoid conversion on global protein expression in *Pseudomonas aeruginosa* strain PAO1 shows induction of the disulfide bond isomerase, DsbA." Journal of Bacteriology **182**(24): 6999-7006.

Martinez-Salazar, J. M., S. Moreno, et al. (1996). "Characterization of the genes coding for the putative sigma factor AlgU and its regulators MucA, MucB, MucC, and MucD in *Azotobacter vinelandii* and evaluation of their roles in alginate biosynthesis." Journal of Bacteriology **178**(7): 1800-1808.

Mashego, M. R., K. Rumbold, et al. (2007). "Microbial metabolimics: past present and future methodologies." Biotechnology Letters **29**:1-16.

Mashego, M. R., L. Wu, et al. (2004). "MIRACLE: Mass Isotopomer Ratio Analysis of U-13C-Labeled Extracts. A New Method for Accurate Quantification of Changes in Concentrations of Intracellular Metabolites." Biotechnology and Bioengineering **85**(6): 620-628.

Mathee, K., C. J. McPherson, et al. (1997). "Posttranslational control of the algT (algU)-encoded sigma(22) for expression of the alginate regulon in *Pseudomonas aeruginosa* and localization of its antagonist proteins MucA and MucB (AlgN)." Journal of Bacteriology **179**(11): 3711-3720.

McLuckey, S. A. and J. M. Wells (2001). "Mass analysis at the advent of the 21st century." Chemical Reviews **101**(2): 571-606.

Meyer, H., M. Liebeke, et al. (2010). "A protocol for the investigation of the intracellular *Staphylococcus aureus* metabolome." Analytical biochemistry **401**(2): 250-259.

Missiakas, D. and S. Raina (1998). "The extracytoplasmic function sigma factors: role and regulation." Molecular Microbiology **28**(6): 1059-1066.

REFERENCES

Moellney, M., W. Wiechert, et al. (1999). "Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments." Biotechnology and Bioengineering **66**(2): 86-103.

Morea, A., K. Mathee, et al. (2001). "Characterization of algG encoding C5-epimerase in the alginate biosynthetic gene cluster of *Pseudomonas fluorescens*." Gene **278**(1-2): 107-114.

Noack, S., K. Nöh, et al. (2011). "Stationary versus non-stationary C-13-MFA: A comparison using a consistent dataset." Journal of Biotechnology **154**(2-3): 179-190.

Ohman, D. E., K. Mathee, et al. (1996). Regulation of the alginate (algD) operon in *Pseudomonas aeruginosa*. Molecular Biology of Pseudomonads. T. Nakazawa, K. Furukawa, D. Haas and S. Silver, ASM Press**:** 472-483.

Palleroni, N. J. and E. R. B. Moore (2004). Taxonomy of *Pseudomonads:* Experimental approaches. *Pseudomonas* genomics, life stye and molecular architecture. J.-L. Ramos. New York, Kluwer Academic / Plenum Publishers. **1:** 3-4.

Patten, C. L. and B. R. Glick (2002). "Role of *Pseudomonas putida* indoleacetic acid in development of the host plant root system." Applied and Environmental Microbiology **68**(8): 3795-3801.

Petersen, S., A. A. de Graaf, et al. (2000). "In vivo quantification of parallel and bidirectional fluxes in the anaplerosis of *Corynebacterium glutamicum*." Journal of Biological Chemistry **275**(46): 35932-35941.

Pier, G. B. and J. B. Goldberg (2004). *Pseudomonas aeruginosa* Interaction with host cells. *Pseudomonas* genomics, life stye and molecular architecture. J.-L. Ramos. New York, Kluwer Academic / Plenum Publishers. **1:** 505-534.

Pogliano, J., A. S. Lynch, et al. (1997). "Regulation of *Escherichia coli* cell envelope proteins involved in protein folding and degradation by the Cpx two-component system." Genes & Development **11**(9): 1169-1182.

Quek, L.-E., C. Wittmann, et al. (2009). "OpenFLUX: efficient modelling software for C-13-based metabolic flux analysis." Microbial Cell Factories **8**(25).

Raamsdonk, L. M., B. Teusink, et al. (2001). "A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations." Nature Biotechnology **19**(1): 45-50.

Radmacher, E., A. Vaitsikova, et al. (2002). "Linking central metabolism with increased pathway flux: L-valine accumulation by *Corynebacterium glutamicum*." Applied and Environmental Microbiology **68**(5): 2246-2250.

Rainey, P. B. and M. J. Bailey (1996). "Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome." Molecular Microbiology **19**(3): 521-533.

Rogers, S., R. A. Scheltema, et al. (2009). "Probabilistic assignment of formulas to mass peaks in metabolomics experiments." Bioinformatics **25**(4): 512-518.

Sauer, U., D. R. Lasko, et al. (1999). "Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism." Journal of Bacteriology **181**(21): 6679-6688.

Schaefer, U., W. Boos, et al. (1999). "Automated sampling device for monitoring intracellular metabolite dynamics." Analytical biochemistry **270**(1): 88-96.

Schurr, M. J., H. Yu, et al. (1996). "Control of AlgU, a member of the sigma(E)-like family of stress sigma factors, by the negative regulators MucA and MucB and

Pseudomonas aeruginosa conversion to mucoidy in cystic fibrosis." Journal of Bacteriology **178**(16): 4997-5004.

Schweizer, H. P., R. Jump, et al. (1997). "Structure and gene-polypeptide relationships of the region encoding glycerol diffusion facilitator (glpF) and glycerol kinase (glpK) of *Pseudomonas aeruginosa*." Microbiology-Sgm **143**: 1287-1297.

Shin, M. H., D. Y. Lee, et al. (2010). "Evaluation of Sampling and Extraction Methodologies for the Global Metabolic Profiling of *Saccharophagus degradans*." Analytical chemistry **82**(15): 6660-6666.

Silby, M. W., A. M. Cerdeno-Tarraga, et al. (2009). "Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*." Genome Biology **10**(5).

Skjåk-Bræk, G. and K. I. Draget (2012). Alginates: properties and applications. Polymer science: a comprehensive reference. K. Matyjaszewski and M. Möller. Amsterdam, Elsevier BV. **10:** 213-220.

Smeaton, J. R. and W. H. Elliott (1967). "Selective release of ribonuclease-inhibitor from *Bacillus subtilis* cells by cold shock treatment." Biochemical and Biophysical Research Communications **26**(1): 75-81.

Sokol, S., P. Millard, et al. (2012). "influx_s: increasing numerical stability and precision for metabolic flux analysis in isotope labelling experiments." Bioinformatics **28**(5): 687-693.

Spura, J., L. Christian Reimer, et al. (2009). "A method for enzyme quenching in microbial metabolome analysis successfully applied to gram-positive and gram-negative bacteria and yeast." Analytical biochemistry **394**(2): 192-201.

Sriram, G., D. B. Fulton, et al. (2004). "Quantification of compartmented metabolic fluxes in developing soybean embryos by employing Biosynthetically directed fractional C-13 labeling, C-13, H-1 two-dimensional nuclear magnetic resonance, and comprehensive isotopomer balancing." Plant Physiology **136**(2): 3043-3057.

Srour, O., J. D. Young, et al. (2011). "Fluxomers: a new approach for C-13 metabolic flux analysis." BMC Systems Biology **5**.

t'Kindt, R., R. A. Scheltema, et al. (2010). "Metabolomics to Unveil and Understand Phenotypic Diversity between Pathogen Populations." Plos Neglected Tropical Diseases **4**(11).

Takahashi, H., K. Kai, et al. (2008). "Metabolomics approach for determining growth-specific metabolites based on Fourier transform ion cyclotron resonance mass spectrometry." Analytical and bioanalytical chemistry **391**(8): 2769-2782.

Taymaz-Nikerel, H., M. de Mey, et al. (2009). "Development and application of a differential method for reliable metabolome analysis in *Escherichia coli*." Analytical biochemistry **386**(1): 9-19.

Ternbach, M. B., C. Bollman, et al. (2005). "Application of model discriminating experimental design for modeling and development of a fermentative fed-batch L-valine production process." Biotechnology and Bioengineering **91**(3): 356-368.

Tolker-Nielsen, T. and S. Molin (2004). The biofim lifestyre of *Pseudomonads*. *Pseudomonas* genomics, life style and molecular architecture. J.-L. Ramos. New York, Kluwer Academic / Plenum Publishers. **3:** 550-556.

REFERENCES

Tu, B. P., R. E. Mohler, et al. (2007). "Cyclic changes in metabolic state during the life of a yeast cell." Proceedings of the National Academy of Sciences of the United States of America **104**(43): 16886-16891.

van der Werf, M. J., K. M. Overkamp, et al. (2008). "Comprehensive analysis of the metabolome of *Pseudomonas putida* S12 grown on different carbon sources." Molecular BioSystems **4**(4): 315-327.

van Ooyen, J., S. Noack, et al. (2012). "Improved L-lysine production with *Corynebacterium glutamicum* and systemic insight into citrate synthase flux and activity." Biotechnology and Bioengineering **109**(8): 2070-2081.

Varma, A., B. W. Boesch, et al. (1993). "Biochemical production capabilities of *Escherichia Coli*." Biotechnology and Bioengineering **42**(1): 59-73.

Varma, A. and B. O. Palsson (1994). "Metabolic flux balanicing - Basic concepts, scientific and practical use." Nature Biotechnology **12**(10): 994-998.

Varma, A. and B. O. Palsson (1994). "Predictions for oxygen-supply control to enhance population stability of engineered production strains." Biotechnology and Bioengineering **43**(4): 275-285.

Varma, A. and B. O. Palsson (1994). "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110." Applied and Environmental Microbiology **60**(10): 3724-3731.

Vielhauer, O., M. Zakhartsev, et al. (2011). "Simplified absolute metabolite quantification by gas chromatography-isotope dilution mass spectrometry on the basis of commercially available source material." Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences **879**(32): 3859-3870.

Villas-Bôas, S. G. (2007). 3.3.3 Cell Disruption Methods. Metabolome analysis: an introduction, John Wiley & Sons**:** 58-71.

Villas-Bôas, S. G., D. G. Delicado, et al. (2003). "Simultaneous analysis of amino and nonamino organic acids as methyl chloroformate derivatives using gas chromatography-mass spectrometry." Analytical Biochemistry **322**: 134-138.

Villas-Bôas, S. G., K. F. Smart, et al. (2011). "Alkylation or silylation for analysis of amino and non-amino organic acids by GC-MS." Metabolites **1**: 3-20.

Volko, S. M., T. Boller, et al. (1998). "Isolation of new *Arabidopsis* mutants with enhanced disease susceptibility to *Pseudomonas syringae* by direct screening." Genetics **149**(2): 537-548.

Weitzel, M., K. Nöh, et al. (2013). "13CFLUX2 - high-performance software suite for 13C-metabolic flux analysis." Bioinformatics (Oxford, England) **29**(1): 143-145.

Wentzel, A., H. Sletta, et al. (2012). "Intracellular Metabolite Pool Changes in Response to Nutrient Depletion Induced Metabolic Switching in *Streptomyces coelicolor*." Metabolites **2**(1): 178-194.

Weuster-Botz, D. (1997). "Sampling tube device for monitoring intracellular metabolite dynamics." Analytical biochemistry **246**(2): 225-233.

Wiechert, W. (2001). "13C metabolic flux analysis." Metabolic Engineering **3**(3): 195-206.

Wiechert, W. and A. A. deGraaf (1997). "Bidirectional reaction steps in metabolic networks .1. Modeling and simulation of carbon isotope labeling experiments." Biotechnology and Bioengineering **55**(1): 101-117.

REFERENCES

Wiechert, W., M. Mollney, et al. (1999). "Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems." Biotechnology and Bioengineering **66**(2): 69-85.

Wiechert, W., M. Möllney, et al. (2001). "A Universal Framework for 13C Metabolic Flux Analysis." Metabolic Engineering **3**(3): 265-283.

Winder, C. L., W. B. Dunn, et al. (2008). "Global metabolic profiling of *Escherichia coli* cultures: An evaluation of methods for quenching and extraction of intracellular metabolites." Analytical chemistry **80**(8): 2939-2948.

Wittmann, C., J. O. Kromer, et al. (2004). "Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria." Analytical biochemistry **327**(1): 135-139.

Wu, L., M. R. Mashego, et al. (2005). "Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly 13C-labeled cell extracts as internal standards." Analytical biochemistry **336**: 164-171.

Wurzel, M. (1997). Stabilität und eindeutige Lösbarkeit von IsotopomerenBilanzgleichungssystemen, University of Bonn.

Xie, L. and D. I. C. Wang (1993). "Fed-batch cultivation of animal cells using different medium design concepts and feeding strategies." Biotechnology and Bioengineering **95**(2): 270-284.

Yoo, H., M. R. Antoniewicz, et al. (2008). "Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line." Journal of Biological Chemistry **283**(30): 20621-20627.

Yuan, J., C. D. Doucette, et al. (2009). "Metabolomics-driven quantitative analysis of ammonia assimilation in E. coli." Molecular Systems Biology **5**(309).

Zamboni, N., E. Fischer, et al. (2005). "FiatFlux - a software for metabolic flux analysis from C-13-glucose experiments." BMC Bioinformatics **6**(209).

Paper I

Journal of Chromatography A, 1247 (2012) 118–124

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Utilization of a deuterated derivatization agent to synthesize internal standards for gas chromatography–tandem mass spectrometry quantification of silylated metabolites

Stina K. Lien, Hans Fredrik Nyvold Kvitvang, Per Bruheim*

Department of Biotechnology, Norwegian University of Science and Technology, Sem Sælands vei 6/8, N-7491 Trondheim, Norway

ARTICLE INFO

Article history:
Received 18 January 2012
Received in revised form 13 May 2012
Accepted 14 May 2012
Available online 28 May 2012

Keywords:
Metabolomics
Quantitative metabolic profiling
GC–MS/MS
Silylation
Isotopically labelled derivatization agent
Data normalization

ABSTRACT

GC–MS analysis of silylated metabolites is a sensitive method that covers important metabolite groups such as sugars, amino acids and non-amino organic acids, and it has become one of the most important analytical methods for exploring the metabolome. Absolute quantitative GC–MS analysis of silylated metabolites poses a challenge as different metabolites have different derivatization kinetics and as their silyl-derivates have varying stability. This report describes the development of a targeted GC–MS/MS method for quantification of metabolites. Internal standards for each individual metabolite were obtained by derivatization of a mixture of standards with deuterated N-methyl-N-trimethylsilyltrifluoroacetamide (d9-MSTFA), and spiking this solution into MSTFA derivatized samples prior to GC–MS/MS analysis. The derivatization and spiking protocol needed optimization to ensure that the behaviour of labelled compound responses in the spiked sample correctly reflected the behaviour of unlabelled compound responses. Using labelled and unlabelled MSTFA in this way enabled normalization of metabolite responses by the response of their deuterated counterpart (i.e. individual correction). Such individual correction of metabolite responses reproducibly resulted in significantly higher precision than traditional data correction strategies when tested on samples both with and without serum and urine matrices. The developed method is thus a valuable contribution to the field of absolute quantitative metabolomics.

1. Introduction

Mass spectrometry (MS) based metabolite profiling has become an important tool in biological research as it aims to provide a comprehensive picture of the metabolite pools in the biological system under study [1]. The high sensitivity of MS detection and the ability to resolve complex mixtures when MS detection is combined with a chromatographic separation step (either gas chromatography (GC) or liquid chromatography (LC)) has made MS based analysis the method of choice for many applications [2]. Despite technological advancements and methodological development, many analytical challenges remain to be resolved in the field of metabolomics. Important issues are the preservation of metabolome composition during inactivation of cell metabolism, during the subsequent metabolite extraction and during the MS analysis [3]. GC and LC separation of metabolites are partly overlapping but also complimentary; they have different strengths and weaknesses when it comes to chromatographic separation of the different metabolite classes [4,5]. Utilization of GC requires a derivatization step to render the metabolites stable and volatile enough for analysis, whilst the main challenge in LC–MS based metabolite analysis is to establish robust and reproducible chromatographic conditions and to develop protocols that retain the many highly charged metabolites with low molecular weight which are not easily retained using standard reverse phase chromatographic conditions. The latter issue has been solved either by the use of ion pair reagents added to the mobile phase or with the use of hydrophilic interaction liquid chromatography (HILIC) stationary phases [6].

Of the three most commonly used GC-derivatization methods (silylation, alkylation and acylation), silylation has been the one in most widespread use in GC–MS based metabolic profiling. This is mainly because of the broad specificity of silylation which enables detection of many metabolite groups such as alcohols, sugars, amino and non-amino organic acids, amines and acyl monophosphates [7]. Electron ionization (EI) has been the standard ionization method for GC–MS as it generates very reproducible fragmentation patterns making the construction of mass spectral searchable metabolite libraries possible [8]. A major drawback of derivatization is the metabolite dependent and metabolite concentration

* Corresponding author. Tel.: +47 735933321; fax: +47 73591283.
E-mail address: Per.Bruheim@biotech.ntnu.no (P. Bruheim).

0021-9673/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.chroma.2012.05.053

dependent time needed for the derivatization to reach completeness, the possible instability of the resulting derivates and the fact that the derivatization procedure itself can introduce artefacts that can lead to misinterpretation of results in downstream data analysis. It is well known that several amino acids yield unstable derivates upon silylation, especially in the presence of water [9] and artefacts might be introduced in crude extracts since they contain salts, acids and bases that are not present in pure standard samples [10]. In addition un-reacted derivatization agent not removed from the sample solution and therefore injected into the GC–MS instrument can lead to problems such as inlet contamination and column degradation. All these factors have a negative impact on the performance of the GC–MS analysis and, subsequently, it complicates data interpretation. Attempts to standardize GC–MS methods and to develop strategies to correct or account for biases during data analyses and data processing steps have been presented [11,12]. One solution has been to do the GC–MS analysis when the same amount of time has elapsed since derivatization for all samples [13]. Another alternative, sometimes used in combination with the first one, is to include labelled internal standards for correction of sampling and instrumental biases. There are several ways to include labelled internal standards: e.g. commercially available labelled compounds for certain metabolites or metabolite groups can be included as internal standards or ^{13}C-labelled extracts can be used [14,15]. The advantage of these two approaches is that the internal standards can be added to the sample prior to sample processing steps and thereby potential metabolite losses during these steps can be monitored. However, when using representative labelled compounds to monitor a metabolite group, the labelled compound is not truly representative for all metabolites in that group, and when using ^{13}C-labelled extracts large differences in metabolite concentration and metabolite instability caused by the sample matrix can be an issue. In addition the ^{13}C-labelled extract might not contain all metabolites relevant for the study to be performed. For GC–MS analysis of metabolites, there is an additional internal standard alternative: a labelled derivatization agent can be used. This is an attractive approach as an individual internal standard can be produced for all metabolites. In addition, this approach can be used to obtain structural information on unknown peaks in the chromatogram [16–18].

In this report, we take advantage of the strong selectivity of the triple quadrupole mass analyzer and combine it with the use of a deuterated derivatization agent. The aim was to establish an quantitative GC–MS/MS method for silylated metabolites using the same approach as previously reported for methyl chloroformate (MCF) derivatization [16].

2. Materials and methods

The purpose of the following work was to develop a method for quantification of a broad range of metabolites. To this end, candidates from a range of compound groups commonly encountered in primary metabolism where chosen. The method was developed using a test solution containing six amino acids (alanine, valine, glutamate, lysine, cysteine and tyrosine), four organic acids (pyruvate, succinate, citrate and α-ketoglutarate), three carbohydrates (glucose, xylose and trehalose) and a sugar alcohol (mannitol). A 0.71 mM solution of these in deionized water was prepared from standards (Sigma–Aldrich). Matrix samples were prepared by adding 400 μL of methanol to 100 μL of blood serum or urine, allowing protein precipitation to occur at 0 °C for 30 min before centrifugation for 10 min at 13,400 rpm (Eppendorf MiniSpin). 250 μL of the supernatant was removed and dried on a speedvac concentrator (Speedvac-Savant SPD20110) before the dried material was dissolved in 50 μL freshly prepared 4%, w/v methoxyamine hydrochloride in pyridine (Sigma–Aldrich).

2.1. Derivatization (oximation and silylation)

Prior to GC–MS/MS analysis, samples where derivatized using either N-methyl-N-trimethylsilyltrifluoroacetamide containing 1% trimethylchlorosilane (MSTFA + 1% TMCS, Pierce), MSTFA (Pierce) or N-methyl-N-(trimethyl-d9-silyl)trifluoroacetamide (d9-MSTFA, Fluka). The procedure used was a modified version of the one described in [8]. Samples were prepared by drying 40 μL or 80 μL of the 0.71 mM test solution on a speedvac concentrator (Speedvac-Savant SPD20110). The dried substance was then dissolved in 20 μL freshly prepared 4%, w/v methoxyamine hydrochloride in pyridine (Sigma–Aldrich) and gently shaken at 30 °C for 90 min. 20 μL of the serum or urine containing 4%, w/v methoxyamine hydrochloride in pyridine was used instead of 20 μL of pure 4%, w/v methoxyamine hydrochloride in pyridine when preparing matrix containing samples. After incubation, 20 μL of MSTFA + 1% TMCS, MSTFA, or d9-MSTFA was added to the sample before additional incubation at 37 °C for 30 min. Samples prepared using MSTFA + 1% TMCS are referred to as trimethylsilylated-standard solutions (TMS-standard solutions) and samples prepared using d9-MSTFA are referred to as trimethyl-d9-silylated-standard solutions (d9-TMS-standard solutions). Samples with a labelled internal standard per compound were prepared by spiking the TMS-standard solution with an equal volume of the d9-TMS-standard solution. All samples analyzed on the GC–MS/MS were prepared to give a final concentration of 1.43 mM. If not otherwise indicated samples were placed on the GC–MS/MS autosampler tray holding 5 °C immediately after preparation.

2.2. GC–MS/MS analysis

Samples were run on a GC-QqQ-MS (Agilent 7890A GC-7000 MS Triple quad) equipped with an autosampler (GC PAL, CTC Analytics AG). 1 μL of the sample was injected in split mode (split ratio 10:1). The GC was operated at a constant flow of 1.2 mL/min and equipped with an Agilent DB-5MS + DG column (30 m with 10 m duragard, inner diameter 0.250 mm and film thickness 0.25 μm). The oven was kept at 60 °C for 1 min after injection before a temperature gradient of 10 °C/min was employed until reaching 325 °C. The oven was then kept at 325 °C for 10 min.

Two GC–MS/MS methods were developed: one utilizing positive chemical ionization (PCI) with methane as the reagent gas and one utilizing electron ionization (EI). The collision gas used was nitrogen and the methods multiple reaction monitoring (MRM)-transitions were unique to the compound in question within ±0.2 min of the compounds' retention time. Product ion abundances were maximized by optimal collision energy voltage.

3. Results and discussion

3.1. Positive chemical ionization (PCI) generates heavier fragment ions more suitable for MRM-transitions than electron ionization (EI), but electron ionization shows stronger sensitivity

A targeted GC–MS/MS method that employs MRM cannot make use of the comprehensive collections of silyl-derivate EI mass spectral libraries. It therefore became of interest to test the appropriateness of using positive chemical ionization (PCI) for analysis of silylated metabolites. PCI is a softer ionization technique than EI, and leads to less fragmentation of the molecular ion upon ionization. Less fragmentation of the molecular ion can be advantageous when creating a MS/MS method, as larger fragment ions are

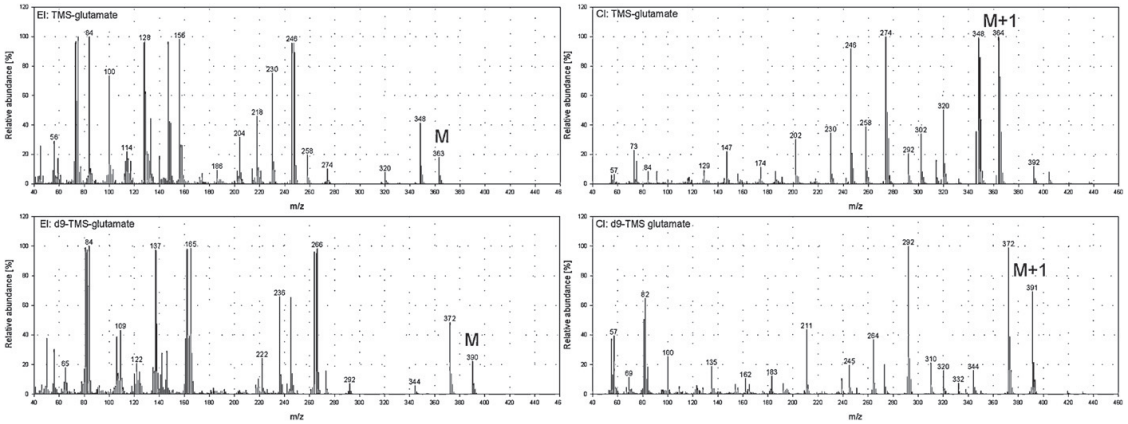


Fig. 1. EI (left column) and PCI (right column) mass spectra of TMS-glutamate (top row) and d9-TMS-glutamate (bottom row). The molecular ions [M] for EI, *m*/*z* 363 and *m*/*z* 390 for TMS-glutamate and d9-TMS-glutamate, respectively and the quasimolecular ions [M+1] for PCI, *m*/*z* 364 and *m*/*z* 391 for TMS-glutamate and d9-TMS-glutamate respectively, are indicated in the figure.

more likely to be suitable precursor ions for creating unique MRM-transitions. PCI was indeed found to have higher selectivity and sensitivity than EI in a recently developed MCF GC–MS/MS method for carboxyl- and amino-group containing metabolites [16]. Two GC–MS/MS methods were therefore developed for the 14 compounds selected as representatives for the sugar, amino acid and non-amino organic acid metabolite groups: one method utilizing EI and the other method utilizing PCI. Both methods contain two MRM-transitions per metabolite, one for the metabolite derivatized with MSTFA and one for the same metabolite derivatized with d9-MSTFA, giving a total of 28 MRM-transitions per method.

Fig. 1 displays full scan mass spectra for TMS-glutamate (top row) and d9-TMS-glutamate (bottom row) for EI (left column) and PCI (right column). MSTFA derivatized glutamate contains three trimethylsilyl groups, leading to molecular ions [M] in the EI mass spectra at *m*/*z* 363 for TMS-glutamate and at *m*/*z* 390 for d9-TMS-glutamate. For both EI spectra, [M] is of low relative abundance whilst more abundant fragment ions are found at the mid to low end of the *m*/*z* scale. In contrast to EI, the quasimolecular ions [M+1] for PCI are of high abundance: [M+1] for TMS-glutamate (*m*/*z* 364) is the base peak in the spectrum and [M+1] for d9-TMS-glutamate (*m*/*z* 391) is the third most abundant peak. Both PCI spectra contain methane-CI spectra specific adduct ions at [M+29] and [M+41]. These adduct ions are formed by electrophilic addition of $C_2H_5^+$ and $C_3H_5^+$ which are components of the methane reagent gas plasma. All of the more abundant fragment ions in the two PCI spectra (e.g. *m*/*z* 246, *m*/*z* 274 and *m*/*z* 348 for TMS-glutamate and *m*/*z* 292 and *m*/*z* 372 for d9-TMS-glutamate) reside at the mid to high end of the *m*/*z* scale. This is not surprising as PCI is a softer ionization technique than EI, leading to less fragmentation of molecules upon ionization.

The full scan mass spectra of TMS-glutamate and d9-TMS-glutamate were compared and used to find suitable precursor ions for unique MRM-transitions. In this context a unique MRM-transition is defined as a transition for an analyte that cannot be produced by its TMS or d9-TMS counterpart, or by any other compounds eluting within ±0.2 min of its retention time. For suitable precursor ions, product ion scans were performed to find unique product ions, before collision energy voltages were optimized for highest sensitivity of the MRM-transition. In an analogous fashion, MRM-transitions for all 14 metabolites were established for the GC–EI-MS/MS and GC–PCI-MS/MS methods. It is worth noting that when creating the MS/MS method for the trimethylsilyl-derivates, some high abundant precursor and product ions had to be abandoned as candidates for MRM-transitions because the ions were also produced from compounds eluting close to the compound in question. Much of the reoccurrence of MRM-transitions for coeluting compounds was caused by ions belonging to isotopic patters from multiple silicon atoms.

The selectivity and sensitivity of the GC–EI-MS/MS method and the GC–PCI-MS/MS method were compared to identify the ionization method best suited for TMS GC–MS/MS analysis. The selectivity of the EI and PCI GC–MS/MS methods were compared by running two standard solutions, one derivatized with MSTFA + 1% TMCS (TMS-standard solution) and the other derivatized with d9-MSTFA (d9-TMS-standard solution). Both samples had responses for both TMS MRM-transitions and d9-TMS MRM-transitions recorded. A selective method should have low to no response for d9-TMS MRM-transitions when analyzing the TMS-standard solution and it should have low to no response for TMS MRM-transitions when analyzing the d9-TMS-standard solution. For the two samples, only three to four weak signals approaching or in the noise region (S/N between 1 and 9) were detected (see supplementary Tables S1 and S2 for EI and PCI responses, respectively). This shows that MRM-transitions of both the GC–EI-MS/MS method and the GC–PCI–MS/MS method have good selectivity.

To evaluate the sensitivity of the two ionization methods, the responses of the TMS-standard solution and the responses of the d9-TMS-standard solutions were compared. EI produced higher responses than PCI for 12 out of 14 compounds in the TMS-standard solution and for 10 out of 14 compounds in the d9-TMS-standard solution (Fig. 2). Over half of these EI responses were over 10-fold of that of the PCI responses. EI is clearly more sensitive for TMS derivates than PCI, and the GC–EI-MS/MS method was chosen for further quantitative TMS GC–MS/MS method development.

3.2. MS responses of TMS-derivates vary significantly with storage time and storage temperature

Completeness of the derivatization reaction and stability of the resulting TMS-derivatives are major concerns when it comes to silylation. To gain some insight into these important issues for the current protocol, the stability of derivate responses over time and at different temperatures were tested. Table 1 shows the results for an experiment where TMS-standard solutions were stored at three different temperatures (5 °C, 25 °C and 37 °C) and analyzed with the

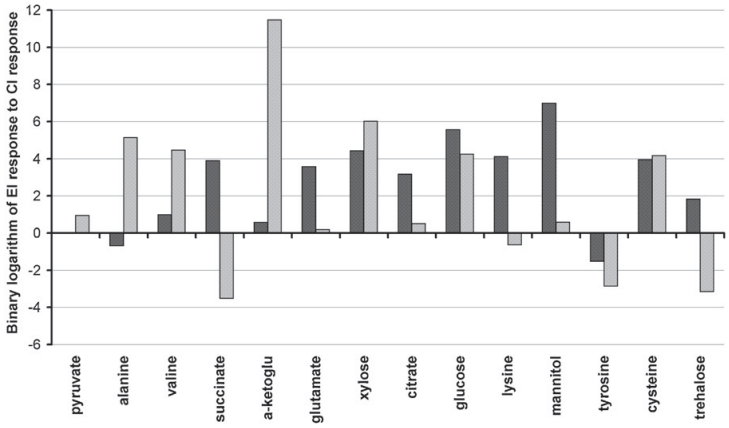


Fig. 2. EI responses relative to PCI responses plotted on a binary logarithmic scale for compounds in the TMS-standard solution (black) and the d9-TMS-standard solution (grey). (The value for TMS-pyruvate is 0.01 and is too low to be visible in the figure.)

Table 1
Relative responses for TMS-standard solution samples, TMS-standard solution samples containing serum and TMS-standard solution samples containing urine. Data for three different time points (24 h, 48 h and 72 h after derivatization) and three different storage temperatures (5 °C, 25 °C and 37 °C) are given. The values are binary logarithms of the ratio "responses at time point" over "response immediately after derivatization (0 h)".

Storage temperature	Metabolite	TMS-standard solution			TMS-standard solution containing serum			TMS-standard solution containing urine		
		24 h	48 h	72 h	24 h	48 h	72 h	24 h	48 h	72 h
5 °C	Pyruvate	0.15	0.11	0.03	0.06	−0.23	−0.23	0.19	0.00	−0.04
	Succinate	0.21	0.24	0.23	−0.18	−0.14	−0.04	−0.22	0.11	0.10
	a-Ketoglu	−0.01	−0.84	−0.01	0.06	−0.24	−0.19	0.22	0.19	0.14
	Citrate	−0.58	−0.53	−0.48	−0.29	−0.58	−0.47	0.04	−0.10	−0.07
	Alanine	−0.70	−0.49	−0.58	−0.24	−0.62	−0.62	−0.04	−0.15	−0.30
	Valine	−0.33	−0.02	0.12	−0.50	−0.43	−0.25	0.06	−0.08	−0.05
	Glutamate	−0.77	−1.45	−1.14	−0.47	−0.64	−0.69	0.03	−0.19	−0.19
	Lysine	−1.03	−1.03	−1.06	−0.63	−1.50	−0.79	−0.50	−0.49	−0.67
	Tyrosine	0.04	0.04	0.01	0.03	−0.65	−0.13	−0.05	−0.07	−0.25
	Cysteine	−2.53	−2.32	−2.25	−1.48	−2.70	−1.78	−0.71	−0.71	−1.00
	Xylose	0.23	−0.12	0.17	0.08	−0.18	−0.15	0.14	0.05	0.06
	Glucose	0.19	0.19	0.23	0.09	−0.17	−0.07	0.17	0.00	0.12
	Trehalose	0.20	0.26	0.27	0.07	−0.98	0.02	0.09	0.09	0.03
	Mannitol	0.12	0.14	0.18	0.03	−0.65	−0.07	0.14	0.05	0.09
25 °C	Pyruvate	0.47	−0.05	0.40	0.08	0.08	0.09	0.06	−0.13	−0.02
	Succinate	0.53	0.43	−0.07	0.13	0.20	−0.29	0.08	−0.45	0.20
	a-Ketoglu	0.50	0.29	0.47	0.16	0.23	0.37	0.08	0.06	0.24
	Citrate	−0.09	−0.21	0.03	−0.21	−0.14	0.04	0.04	−0.07	0.14
	Alanine	0.02	−0.53	−0.11	−0.20	−0.23	−0.28	−0.43	−0.09	−0.18
	Valine	0.33	0.01	−0.21	−0.33	0.03	0.29	0.07	0.02	0.19
	Glutamate	−0.39	−0.75	−0.58	−0.25	−0.10	0.23	−0.05	−0.30	−0.20
	Lysine	−0.82	−0.66	−0.50	−0.53	−1.01	−0.34	−0.28	−0.41	−0.36
	Tyrosine	0.43	0.30	0.46	0.08	−0.39	0.29	−0.01	−0.09	0.01
	Cysteine	−1.62	−1.83	−1.61	−1.00	−0.91	−0.74	−0.54	−0.54	−0.32
	Xylose	0.55	0.44	0.61	0.12	0.19	0.29	0.10	0.04	0.14
	Glucose	0.21	0.37	0.55	0.06	0.13	0.26	0.00	−0.07	0.16
	Trehalose	0.51	0.54	0.69	0.15	0.26	0.36	0.03	0.11	0.20
	Mannitol	0.33	0.29	0.48	0.03	−0.28	0.26	0.04	−0.01	0.15
37 °C	Pyruvate	0.35	0.21	0.24	0.21	0.20	0.14	0.12	−0.07	0.15
	Succinate	0.35	−0.06	0.39	0.30	0.33	0.39	0.05	−0.07	0.23
	a-Ketoglu	0.64	0.52	0.70	0.29	−0.06	0.44	0.27	0.13	0.47
	Citrate	0.37	0.27	0.43	0.23	0.24	0.31	0.06	−0.03	0.26
	Alanine	0.29	0.07	0.05	0.05	−0.04	−0.08	−0.10	−0.52	−0.25
	Valine	0.51	0.39	0.47	0.20	0.13	0.25	0.03	−0.21	0.21
	Glutamate	0.93	0.98	1.06	0.43	−0.16	0.61	−0.07	−0.35	−0.14
	Lysine	0.21	0.16	0.21	−0.17	−0.18	−0.15	−0.39	−0.51	−0.34
	Tyrosine	0.84	0.80	0.86	0.40	0.35	0.46	−0.03	−0.12	0.08
	Cysteine	0.26	−0.05	0.19	−0.76	−0.84	−0.85	−0.38	−0.49	−0.24
	Xylose	0.78	0.65	0.79	0.36	−0.15	0.38	0.12	−0.02	0.24
	Glucose	0.78	0.70	0.80	0.32	0.32	0.39	0.03	−0.01	0.22
	Trehalose	0.68	0.64	0.79	0.41	0.45	0.61	0.13	0.14	0.26
	Mannitol	0.39	0.32	0.44	0.24	0.27	0.33	0.04	−0.01	0.23

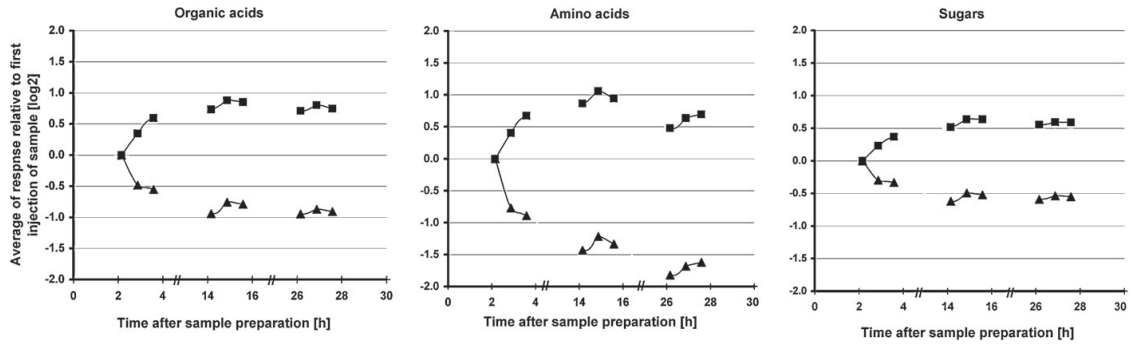


Fig. 3. Average responses of unlabelled (boxes) and labelled (triangles) organic acids, amino acids and sugars in nine runs of a TMS-standard solution spiked with d9-TMS standard solution relative to the first run.

GC–EI-MS/MS method at four different time points (0 h, 24 h, 48 h and 72 h after derivatization). For each storage temperature three different samples were prepared: one ordinary TMS-standard solution and two TMS-standard solutions containing matrices (serum and urine). The table gives the binary logarithm of responses at the three later time points relative to the first time point. For storage at 5 °C there is a general response decrease, for storage at 25 °C there is a fairly even distribution between decrease and increase in responses, and for storage of samples at 37 °C there is a general response increase (see supplementary Table S3 for a colour-coded version of Table 1 for a visualization of these general tendencies). The response instability is present also for sugars which are anticipated to be quite stable as TMS-derivatives and there is a significant change in responses even in the shortest time frame of 24 h. Surprisingly, storage at 5 °C does not seem to produce more stable responses than storage at higher temperatures and for some reason less variation between time points are observed for the TMS-standard solution containing urine. Silylation is clearly vulnerable to analytical biases, and there is a need for new analytical strategies to correct for these, especially for absolute quantification purposes.

3.3. Spiking of TMS-derivatized samples with d9-TMS-derivatized standard solution introduce instability that can be avoided by prolonged derivatization times

MSTFA derivatized samples were spiked with d9-MSTFA derivatized standard solution to introduce a true internal standard for each individual metabolite. Fig. 3 shows response averages for organic acids, amino acids and sugars for nine subsequent runs of a TMS-standard solution spiked with d9-TMS standard solution relative to the first run of the sample. Surprisingly the responses of unlabelled compounds (boxes) initially show a general increase whilst the responses of labelled compounds (triangles) initially show a general decrease (three first time points). The initial instability where unlabelled compound responses increase and labelled compound responses decrease is present for organic acids, amino acids and sugars. The opposite behaviour of unlabelled and labelled compound responses does not persist: after approximately 14 h (six later time points) unlabelled compound responses increase when labelled compound responses increase, and vice versa they decrease when labelled compound responses decrease. The initial inverse behaviour of unlabelled and labelled compound responses and the subsequent co-varying behaviour of the unlabelled and labelled compound responses is reproducible, occurs when TMCS is omitted in the derivatizing solution and also occurs in the presence of various matrices (serum and urine) (supplementary Fig. S1). The exact reason for the initial instability where unlabelled and labelled responses do not co-vary is unknown. Huang and Regnier used the approach of a labelled derivatization agent for differential GC–TOF-MS analysis but chose to use N-Methyl-N-(t-butyldimethylsilyl) trifluoroacetamide (MTBSTFA) since it is less likely to hydrolyze and because scrambling effect are more likely to occur with MSTFA [17]. As sugars are not silylated by MTBSTFA this reagent was unfortunately not an option for the present work. The characteristics of the initial instability phenomenon were not observed if TMS-standard and d9-TMS-standard solutions were run separately. Nor was it observed if equal amounts of TMS-standard solution and a TMS-blank solution were combined, or if equal amounts of d9-TMS-standard solution and a d9-TMS-blank solution were combined (data not shown). The observation of the phenomenon points to one major challenge of MSTFA derivatization: the requirement of an optimal derivatization protocol that balance completeness of derivatization reaction and derivate stability. Whilst early derivatization protocols used high temperatures (usually 60–70 °C), the standard protocol, as used in this study, operates at 30 °C for 90 min and then 37 °C for 30 min, for oximation and silylation, respectively. One explanation for the trends in Fig. 3 is that un-reacted MSTFA is able to compete with and exchange the d9-TMS groups of labelled compounds in the TMS-standards solution spiked with d9-TMS-standard solution, and that this exchange stops when all MSTFA is consumed. In accordance with this view equilibrating the TMS-standard solution sample and the d9-TMS standard solutions sample individually for a prolonged time period prior to spiking of the TMS-standard solution with the d9-TMS-standard solution should ensure that the derivatization reaction has reached completeness for all metabolites and that preferentially any un-reacted derivatization reagent has hydrolyzed and become un-reactive. Indeed when the TMS-standard solution sample and the d9-TMS-standard solutions sample was incubated at 5 °C for 20 h prior to mixing and subsequent analysis, the strong inverse behaviour of unlabelled and labelled compound responses was avoided (Fig. 4). Thus prolonging the derivatization before mixing for a time appropriate for your sample is important to ensure that the initial inverse behaviour of unlabelled and labelled compound responses does not occur.

3.4. Individual correction improves precision of silyl-derivate data more than group correction

Trimethylsilyl derivates are known to be unstable and therefore within and between sequence variability has to be addressed

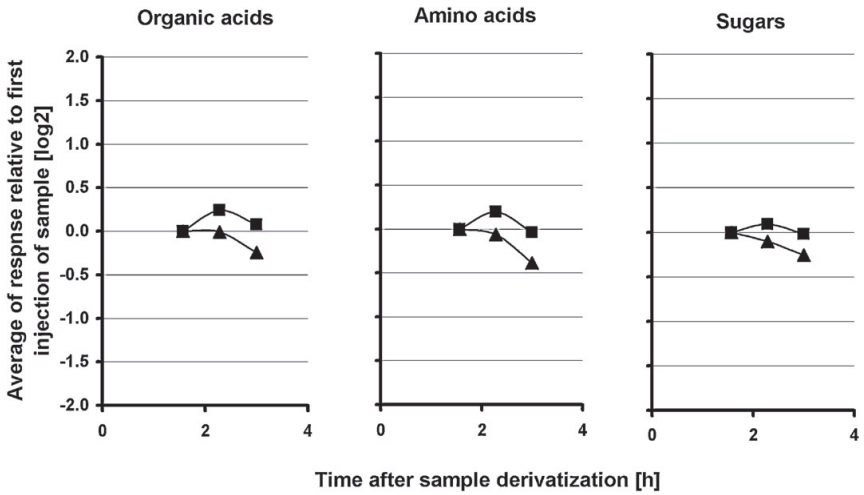


Fig. 4. Average responses of unlabelled (boxes) and labelled (triangles) organic acids, amino acids and sugars for three runs relative to the first run for a TMS-standard solution spiked with d9-TMS standard solution when the TMS-standard sample and d9-TMS standard sample were incubated for 20 h at 5 °C prior to spiking.

when evaluating data obtained by GC–MS analysis. Here we correct the responses of metabolites in the TMS-standard solution spiked with d9-TMS standards by normalizing metabolite responses by the response of their d9-TMS counterparts. This procedure has been termed d9-TMS-individual correction. The conventional procedure of normalizing responses of a compound group by using the response of one specific labelled compound belonging to that group is termed group correction. Usually this is done by adding one or more isotopically labelled compounds to the sample prior to derivatization, and then normalizing the responses of a metabolite group by the response of an isotopically labelled compound belonging to that group. For convenience, we will in the following compare d9-TMS-individual correction with what we have termed d9-TMS-group correction, instead of comparing it to the conventional way of doing group correction. For d9-TMS group correction organic acid responses in the TMS-standard solution spiked with d9-TMS-standards have been corrected using the response of d9-TMS-citrate, amino acid responses have been corrected using the response of d9-TMS-valine and sugar responses have been corrected using the response of d9-TMS-xylose. Initially d9-TMS-group correction for three subsequent runs of a TMS-standard solution spiked with d9-TMS standards and conventional group correction for three subsequent runs of a TMS-standard solution were compared. The two correction protocols provided similar average relative standard deviations (STD) of responses for the metabolite groups under consideration (organic acids, amino acids and sugars) (supplementary Fig. S2), showing that conclusions drawn when comparing d9-TMS-individual correction with d9-TMS-group correction are valid for conventional group correction as well.

As described in Section 3.3, we observed that spiking of TMS-standard solution samples with d9-TMS standard solution could lead to a phenomenon where the concentration of unlabelled compounds initially show a general increase, whilst the concentration of labelled compounds show a general decrease. Fig. 5A displays average relative STD for six runs of a TMS-standard solution spiked with d9-TMS standards analyzed whilst the concentration of labelled compounds were increasing and the concentration of unlabelled compounds were decreasing. Organic acids, amino acids and sugars are shown separately and average relative STD is displayed for uncorrected responses (black), d9-TMS-group corrected responses (grey) and d9-TMS-individually corrected responses (white). Because of the inverse change in concentrations of unlabelled and labelled compounds occurring during analysis of this sample, un-corrected responses have the highest precision (low average relative STD), whilst average relative STD is higher for both d9-TMS-group correction and d9-TMS-individual correction. However, as shown in Fig. 4, initial increase in unlabelled compound concentration and initial decrease in labelled compound concentration can be avoided by e.g. prolonged derivatization times. Fig. 5B–D shows average relative STD for six runs of three TMS-standard solutions spiked with d9-TMS standards when inverse behaviour of unlabelled and labelled compound concentrations was not occurring. One ordinary TMS-standard solution spiked with d9-TMS standards (B) and two TMS-standard solutions spiked with d9-TMS standards containing matrices (C and D) are displayed. For these samples d9-TMS-group correction produce lower average relative STD than no correction for organic acids and sugars in most instances, but does not improve average relative STD for amino acids. d9-TMS-individual correction improves average relative STD for organic acids and sugars even more than d9-TMS-group correction in most instances, and in contrast to d9-TMS-group correction, d9-TMS-individual correction improves average relative STD for amino acids as well. Using d9-TMS-individual correction thus improves relative STD, and thereby the precision of the data for all compound groups and it leads to an average relative STD of about 10% or lower. The benefits of d9-TMS individual correction were reproducible and also occurred when TMCS was omitted in the derivatization solution (data not shown). The results for d9-TMS-group correction proved to be unpredictable in the sense that often group correction decreased the average relative STD and sometimes d9-TMS group correction did not decrease the average relative STD. TMS-group correction was especially non-effective for amino acids. This is not surprising as metabolites within the groups organic acids and sugars are more similar to each other in structure and chemical properties than amino acids are. Neither is it unexpected that d9-TMS individual correction gives the most precise results. Silyl-derivates of metabolites within a metabolite group do not necessarily have the same stability and their derivatization does not necessarily follow the same kinetics, as Table 1 clearly demonstrates.

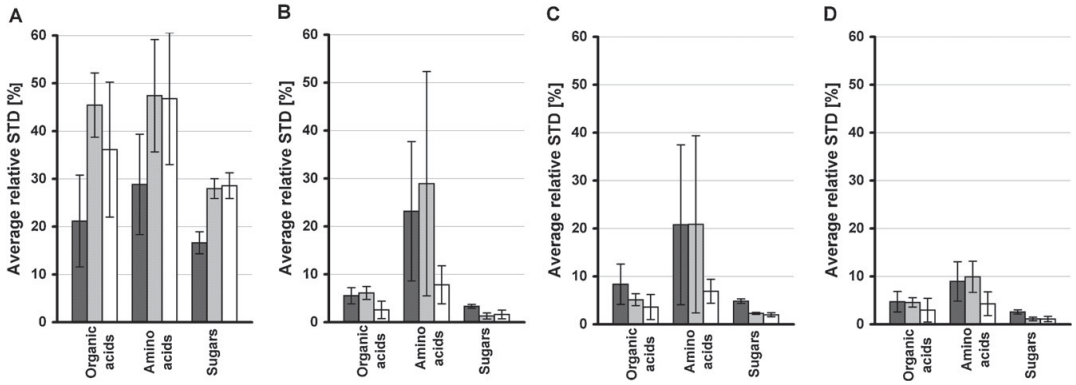


Fig. 5. Average relative standard deviation (STD) of six runs done over a 14.15 h time period for organic acids, amino acids and sugars of 4 TMS-standard solutions spiked with d9-TMS standard solution displayed with standard deviation. Black bars show relative STD for uncorrected responses, grey bars show relative STD for d9-TMS group corrected responses and white bars show relative STD for d9-TMS individually corrected responses. (A and B) Pure TMS-standard solution spiked with d9-TMS standard solution; (C) serum-containing TMS-standard solution spiked with d9-TMS standard solution; (D) urine-containing TMS-standard solution spiked with d9-TMS standard solution. Sample A was displaying inverse behaviour of unlabelled and labelled compound concentrations when run, whilst samples B, C and D were not.

3.5. Conclusion

The GC–MS/MS method presented here is a selective and sensitive method for quantification of silylated metabolites. By using the labelled derivatization reagent d9-MSTFA, deuterated counterparts of all analyzed metabolites can be made enabling normalization of all metabolite responses by individual internal standards. This report has shown that individual normalization improve data precision more than normalization of metabolite responses group by group using one labelled compound per metabolite group. The developed GC–MS/MS method is thus a valuable contribution for use in quantitative metabolomics.

Acknowledgements

This study was financed by two internal grants at NTNU.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chroma.2012.05.053.

References

[1] M. Oldiges, S. Lutz, S. Pflug, K. Schroer, N. Stein, C.S. Wiendahl, Appl. Microbiol. Biotechnol. 76 (2007) 495.
[2] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass Spectrom. Rev. 26 (2007) 51.
[3] S.G. Villas-Boas, P. Bruheim, Anal. Biochem. 370 (2007) 87.
[4] D.E. Garcia, E.E. Baidoo, P.I. Benke, F. Pingitore, Y.J. Tang, S. Villa, J.D. Keasling, Curr. Opin. Microbiol. 11 (2008) 233.
[5] J.M. Buscher, D. Czernik, J.C. Ewald, U. Sauer, N. Zamboni, Anal. Chem. 81 (2009) 2135.
[6] S.U. Bajad, W.Y. Lu, E.H. Kimball, J. Yuan, C. Peterson, J.D. Rabinowitz, J. Chromatogr. A 1125 (2006) 76.
[7] O. Fiehn, Trac-Trend Anal. Chem. 27 (2008) 261.
[8] T. Kind, G. Wohlgemuth, D.Y. Lee, Y. Lu, M. Palazoglu, S. Shahbaz, O. Fiehn, Anal. Chem. 81 (2009) 10038.
[9] E. Kaal, H.G. Janssen, J. Chromatogr. A 1184 (2008) 43.
[10] J.L. Little, J. Chromatogr. A 844 (1999) 1.
[11] H.H. Kanani, M.I. Klapa, Metab. Eng. 9 (2007) 39.
[12] H. Kanani, P.K. Chrysanthopoulos, M.I. Klapa, J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 871 (2008) 191.
[13] S.G. Villas-Boas, K.F. Smart, S. Sivakumaran, G.A. Lane, Metabolites 1 (2011) 3.
[14] M.M. Koek, B. Muilwijk, M.J. van der Werf, Anal. Chem. 78 (2006) 3839.
[15] L. Wu, M.R. Mashego, J.C. van Dam, A.M. Proell, J.L. Vinke, C. Ras, W.A. van Winden, W.M. van Gulik, J.J. Heijnen, Anal. Biochem. 336 (2005) 164.
[16] H.F.N. Kvitvang, T. Andreassen, T. Adam, S.G. Villas-Boas, P. Bruheim, Anal. Chem. 83 (2011) 2705.
[17] X.D. Huang, F.E. Regnier, Anal. Chem. 80 (2008) 107.
[18] D. Herebian, B. Hanisch, F.J. Marner, Metabolomics 1 (2005) 317.

Supplementary information
”Utilization of a deuterated derivatization agent to synthesize internal standards for gas chromatography – tandem mass spectrometry quantification of silylated metabolites”

Supplementary Table S1: Responses for the 28 MRM-transitions of the GC-EI-MS/MS method in a TMS-standard solution sample and in a d9-TMS-standard solution sample.

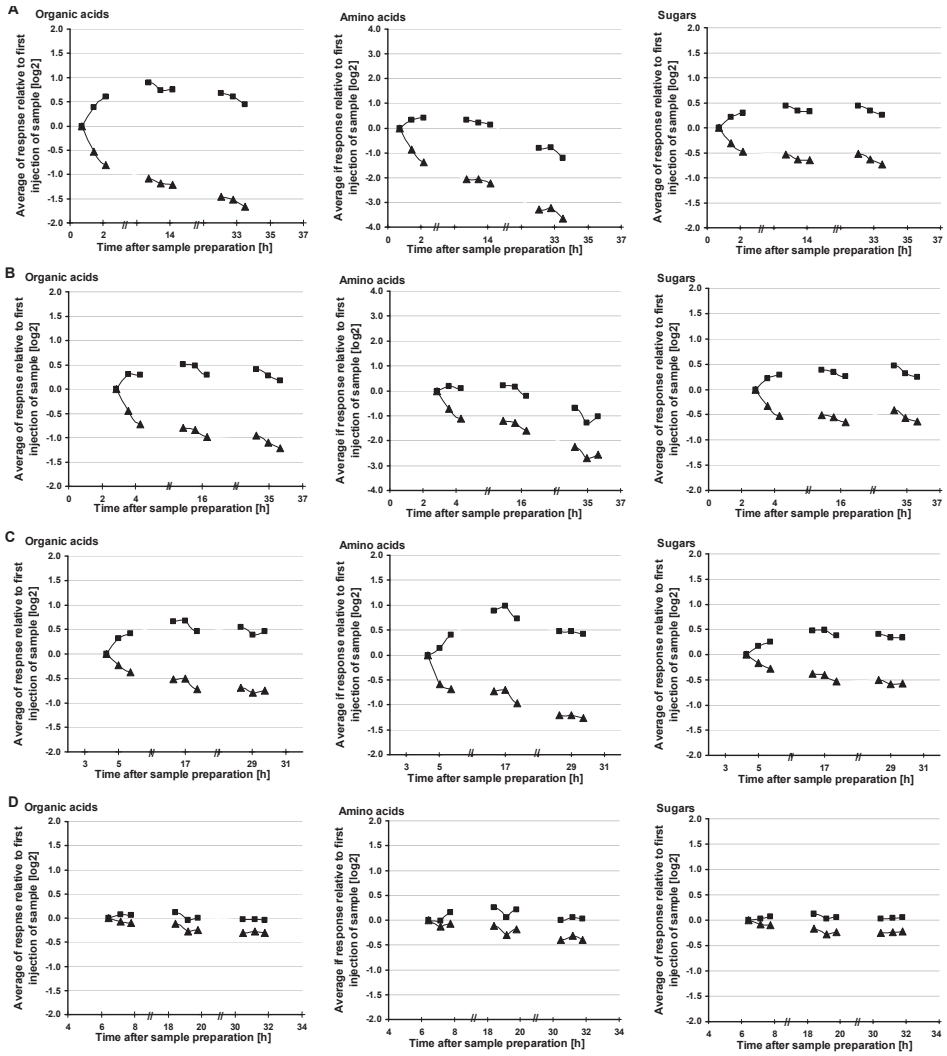
	MRM-transitions for	Response		MRM-transitions for	Response	
	TMS standard	TMS-standard	d9-TMS-standard	d9-TMS-standard	d9-TMS-standard	TMS-standard
Metabolite	solution	solution sample	solution sample	solution	solution sample	solution sample
pyruvate	175.0 -> 74.9	451336	458	179.8 -> 79.9	4477632	2060
alanine	190.0 -> 131.0	60783	0	126.2 -> 83.0	5011370	0
valine	217.9 -> 146.8	517931	0	236.1 -> 81.9	5120500	0
succinate	248.1 -> 148.0	2461704	0	236.2 -> 81.0	388007	0
a-ketoglu	304.0 -> 186.1	78830	0	211.1 -> 82.0	1186216	98
glutamate	348.1 -> 230.0	109999	0	266.2 -> 139.1	99547	0
xylose	309.1 -> 75.0	696659	0	335.4 -> 81.9	4088901	0
citrate	465.1 -> 256.9	397985	0	498.2 -> 272.0	440056	0
glucose	364.1 -> 159.8	879672	0	348.4 -> 164.0	319896	0
lysine	434.2 -> 155.9	795815	0	470.5 -> 165.0	646012	0
mannitol	345.1 -> 255.0	905106	0	457.5 -> 178.0	104764	0
tyrosine	382.1 -> 179.0	81414	0	406.2 -> 307.0	79316	0
cysteine	411.1 -> 145.9	452984	0	438.2 -> 155.0	541842	0
trehalose	330.9 -> 169.0	3860603	18	391.3 -> 178.9	208258	0

Supplementary Table S2: Responses for the 28 MRM-transitions of the GC-PCI-MS/MS method in a TMS-standard solution sample and a d9-TMS-standard solution sample.

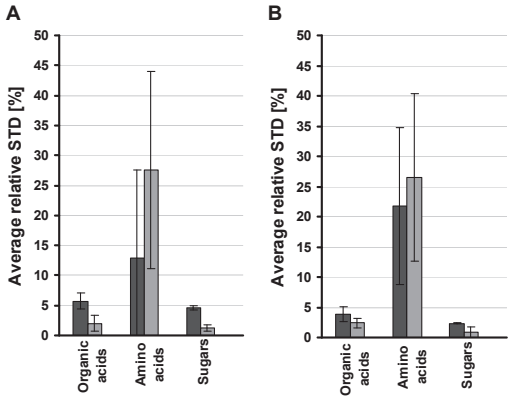
	MRM-transitions for	Response		MRM-transitions for	Response	
	TMS standard	TMS-standard	d9-TMS-standard	d9-TMS-standard	d9-TMS-standard	TMS-standard
Metabolite	solution	solution sample	solution sample	solution	solution sample	solution sample
pyruvate	173.8 -> 74.0	449611	334	199.1 -> 72.0	2335191	63
alanine	233.9 -> 116.1	97405	0	252.2 -> 125.1	142217	0
valine	262.0 -> 144.2	261920	0	280.2 -> 153.2	233479	0
succinate	291.1 -> 45.0	165234	0	182.3 -> 50.1	4439845	0
a-ketoglu	303.9 -> 75.0	53353	0	319.1 -> 80.9	417	0
glutamate	364.0 -> 274.2	9287	0	272.3 -> 245.3	87585	0
xylose	452.2 -> 330.3	32551	0	485.4 -> 274.2	63602	0
citrate	466.2 -> 258.1	44396	0	498.4 -> 272.2	310036	0
glucose	554.3 -> 188.9	18625	0	596.6 -> 204.3	16907	0
lysine	435.0 -> 156.2	46000	0	471.5 -> 193.4	999741	0
mannitol	525.3 -> 254.9	7171	0	372.3 -> 273.3	69723	0
tyrosine	398.0 -> 280.3	233449	0	406.3 -> 378.3	573512	0
cysteine	529.2 -> 264.2	29398	0	565.4 -> 282.3	30327	0
trehalose	361.1 -> 103.1	1092381	54	388.1 -> 178.2	1852266	0

Supplementary Table S3: Relative responses for TMS-standard solution samples, TMS-standard solution samples containing serum and TMS-standard solution samples containing urine. Data for three different time points (24 hours, 48 hours and 72 hours after derivatization) and three different storage temperatures (5 °C, 25 °C and 37 °C) are given. The values are binary logarithms of the ratio "responses at time point" over "response immediately after derivatization (0 hours)". Green: value is below -0.25; Black: value is between -0.25 and 0.25; Red: value is above 0.25.

Storage temperature	Metabolite	TMS-standard solution			TMS-standard solution containing serum			TMS-standard solution containing urine		
		24h	48h	72h	24h	48h	72h	24h	48h	72h
5 °C	pyruvate	0.15	0.11	0.03	0.06	-0.23	-0.23	0.19	0.00	-0.04
	succinate	0.21	0.24	0.23	-0.18	-0.14	-0.04	-0.22	0.11	0.10
	a-ketoglu	-0.01	-0.84	-0.01	0.06	-0.24	-0.19	0.22	0.19	0.14
	citrate	-0.58	-0.53	-0.48	-0.29	-0.58	-0.47	0.04	-0.10	-0.07
	alanine	-0.70	-0.49	-0.58	-0.24	-0.62	-0.62	-0.04	-0.15	-0.30
	valine	-0.33	-0.02	0.12	-0.50	-0.43	-0.25	0.06	-0.08	-0.05
	glutamate	-0.77	-1.45	-1.14	-0.47	-0.64	-0.69	0.03	-0.19	-0.19
	lysine	-1.03	-1.03	-1.06	-0.63	-1.50	-0.79	-0.50	-0.49	-0.67
	tyrosine	0.04	0.04	0.01	0.03	-0.65	-0.13	-0.05	-0.07	-0.25
	cysteine	-2.53	-2.32	-2.25	-1.48	-2.70	-1.78	-0.71	-0.71	-1.00
	xylose	0.23	-0.12	0.17	0.08	-0.18	-0.15	0.14	0.05	0.06
	glucose	0.19	0.19	0.23	0.09	-0.17	-0.07	0.17	0.00	0.12
	trehalose	0.20	0.26	0.27	0.07	-0.98	0.02	0.09	0.09	0.03
	mannitol	0.12	0.14	0.18	0.03	-0.65	-0.07	0.14	0.05	0.09
25 °C	pyruvate	0.47	-0.05	0.40	0.08	0.08	0.09	0.06	-0.13	-0.02
	succinate	0.53	0.43	-0.07	0.13	0.20	-0.29	0.08	-0.45	0.20
	a-ketoglu	0.50	0.29	0.47	0.16	0.23	0.37	0.08	0.06	0.24
	citrate	-0.09	-0.21	0.03	-0.21	-0.14	0.04	0.04	-0.07	0.14
	alanine	0.02	-0.53	-0.11	-0.20	-0.23	-0.28	-0.43	-0.09	-0.18
	valine	0.33	0.01	-0.21	-0.33	0.03	0.29	0.07	0.02	0.19
	glutamate	-0.39	-0.75	-0.58	-0.25	-0.10	0.23	-0.05	-0.30	-0.20
	lysine	-0.82	-0.66	-0.50	-0.53	-1.01	-0.34	-0.28	-0.41	-0.36
	tyrosine	0.43	0.30	0.46	0.08	-0.39	0.29	-0.01	-0.09	0.01
	cysteine	-1.62	-1.83	-1.61	-1.00	-0.91	-0.74	-0.54	-0.54	-0.32
	xylose	0.55	0.44	0.61	0.12	0.19	0.29	0.10	0.04	0.14
	glucose	0.21	0.37	0.55	0.06	0.13	0.26	0.00	-0.07	0.16
	trehalose	0.51	0.54	0.69	0.15	0.26	0.36	0.03	0.11	0.20
	mannitol	0.33	0.29	0.48	0.03	-0.28	0.26	0.04	-0.01	0.15
37 °C	pyruvate	0.35	0.21	0.24	0.21	0.20	0.14	0.12	-0.07	0.15
	succinate	0.35	-0.06	0.39	0.30	0.33	0.39	0.05	-0.07	0.23
	a-ketoglu	0.64	0.52	0.70	0.29	-0.06	0.44	0.27	0.13	0.47
	citrate	0.37	0.27	0.43	0.23	0.24	0.31	0.06	-0.03	0.26
	alanine	0.29	0.07	0.05	0.05	-0.04	-0.08	-0.10	-0.52	-0.25
	valine	0.51	0.39	0.47	0.20	0.13	0.25	0.03	-0.21	0.21
	glutamate	0.93	0.98	1.06	0.43	-0.16	0.61	-0.07	-0.35	-0.14
	lysine	0.21	0.16	0.21	-0.17	-0.18	-0.15	-0.39	-0.51	-0.34
	tyrosine	0.84	0.80	0.86	0.40	0.35	0.46	-0.03	-0.12	0.08
	cysteine	0.26	-0.05	0.19	-0.76	-0.84	-0.85	-0.38	-0.49	-0.24
	xylose	0.78	0.65	0.79	0.36	-0.15	0.38	0.12	-0.02	0.24
	glucose	0.78	0.70	0.80	0.32	0.32	0.39	0.03	-0.01	0.22
	trehalose	0.68	0.64	0.79	0.41	0.45	0.61	0.13	0.14	0.26
	mannitol	0.39	0.32	0.44	0.24	0.27	0.33	0.04	-0.01	0.23



Supplementary Figure S1: Response averages of unlabelled (boxes) and labelled (triangles) organic acids, amino acids and sugars for nine runs of a sample relative to the first run. A: a TMS-standard solution spiked with d9-TMS standard solution; B: a TMS-standard solution derivatized with MSTFA instead of MSTFA+ 1%TMCS spiked with d9-TMS standard solution; C: a serum-containing TMS-standard solution spiked with d9-TMS standard solution; D: a urine-containing TMS-standard solution spiked with d9-TMS standard solution. (For (D) the urine-containing TMS-standard solution spiked with d9-TMS standard solution, the initial inverse change in concentration of unlabeled and labelled compounds seems to be less pronounced than for the other samples. This is an experimental design effect as the samples were run alternatingly in a sequence with three subsequent runs at a time for each sample. More time had passed after derivatization before the urine-containing sample was run and the sample was therefore approaching the time where responses of unlabelled and labelled compounds co-vary).



Supplementary Figure S2: Average relative standard deviations (STD) for organic acids, amino acids and sugars in three subsequent runs of (A) a TMS-standard solution spiked with d9-TMS standard solution and (B) a TMS-standard solution. The standard deviation of the average relative STD for the compound groups is indicated by error bars. Black bars: uncorrected responses; Grey bars in A: d9-TMS-group corrected responses; Grey bars in B: conventional group corrected responses using compounds present in the sample prior to derivatization (No inverse behaviour of unlabelled compound and labelled compound concentrations was observed when the samples were run).

Paper II

Is not included due to copyright

Paper III

Is not included due to copyright

Appendix IV

Quantitative analysis of amino and organic acids by methyl chloroformate derivatization and GC-MS/MS analysis

Hans Fredrik Nyvold Kvitvang[1], Kåre A. Kristiansen[1], Stina K. Lien[1] and Per Bruheim[1]*

[1]NTNU Department of Biotechnology, Norwegian University of Science and Technology, Trondheim, Norway

Summary

Alkyl chloroformates are known for their ability to produce mixed anhydrides, and they have found use as versatile derivatization reagents for gas chromatographic (GC) separation of amino- and organic acids. Triple quadrupole mass spectrometers are excellent detectors for high sensitive and selective analysis. Here, we describe a methyl chloroformate (MCF) GC-MS/MS method for the quantitative analysis of metabolites containing amino- and/ or carboxylic groups. The method covers over sixty metabolites with quantitation limits down to low picomole range injected on column, and any metabolite with amino- and/ or carboxylic acid functional groups that yield a stable and volatile MCF-derivative can be included in the method. Absolute quantitation can be achieved by including stable isotope coded derivatization agent (d_3-MCF) and deuterated alcohol solvent (e.g. d_4-Methanol). As the carboxylic and amino groups are differently labeled, the former from the solvent methanol while the latter from MCF, this can also be used to identify number of amino- and carboxylic groups in unknown analytes in an extract.

Key words: Metabolite profiling, GC-MS, Chloroformate derivatization, Quantitative analysis, Stable isotope coded derivatization reagent

Corresponding author: Per.Bruheim@ntnu.no

1. Introduction

Amino- and organic acids are well known for their central role in metabolism. Thus, analysis, and especially the quantitative determination, of these important metabolite groups is of high interest, not only for increased understanding of function and properties of biological systems, but also for applied perspectives such as in food science and clinical diagnosis, e.g. many diseases with changed amino- and organic acid metabolism are known [1].

Amino- and organic acids can be quantitated by a variety of methods of which enzymatic assays are simplest. However, the throughput is low as this is single analyte assays. For more comprehensive analysis of the complete amino acid pool dedicated amino acid analysers using ion exchange chromatography principles and post column derivatization for UV detection have been available for decades [2]. More recently, gas chromatography (GC)/ liquid chromatography (LC)/ capillary electrophoresis (CE) separation coupled to mass spectrometry detection have become popular alternatives [3,4]. The various separation technologies come with different advantages and disadvantages, hence a general recommendation cannot be stated; this is in regard to sensitivity, selectivity, robustness, throughput (i.e. run time per analysis) etc. Mass spectrometric detection, especially the MS/MS mode of triple quadrupole (QqQ) MS instruments, is attractive for high sensitive and selective detection. They can yield quantitative data, although sometimes challenging, by use of internal and external standards, and, importantly, the analysis does not have to be limited to the metabolic end-product (amino- and organic acids), but also intermediates in the biosynthetic pathways can often be analyzed with the same experimental set-up. Thus, the MS based methods have been considered to be Metabolite Profiling methods, and such methods easily comprise fifty to hundred metabolites of various metabolite groups. This is challenging from a quantitative perspective as true internal standard, i.e deuterated or ^{13}C-labeled analogs, for each analyte is not always available or only at an unreasonable high cost. Alternatively, internal standards can be introduced through isotope coded derivatization (ICD) reagents if a derivatization step is included during the sample preparation steps [5]. Derivatization of amino and organic acids to convert the metabolites to volatile and less polar derivatives is required for GC separation. Thus, there is a low threshold to convert a (semi-) quantitative GC-MS method to an absolute quantitative method for a large number of analytes.

There are mainly two derivatization techniques, alkylation and silylation, that have been used for the GC-MS analysis of amino and organic acids. In general, silylation has been

the most frequently used derivatization reaction in the Metabolite Profiling field. In contrary to alkylation, silylation permit detection of sugars and sugar alcohols, but for amino and organic acids has the silylation technique been associated with artifact formation and less stable silyl-derivates that significantly hamper and challenge the precision of a quantitative analysis [6-8]. In a recent study it was also shown that alkylation yield higher reproducibility than silylation [9]. Alkylation is the replacement of active hydrogen in carboxylic, thiol- and amino groups with an alkyl group or, sometimes aryl group. Chloroformates have been frequently used as alkylation reagents [10], and derivatization protocols of all methyl-, ethyl-, and propyl-chloroformate have been developed and used [11-14]. Here, we describe a methyl chloroformate (MCF) GC-MS/MS method for the quantitative analysis of amino and carboxylic group containing metabolites [15]. The method takes the advantage of the selective derivatization of amino and carboxylic functional groups, and combines it with the sensitive and selective detection of triple quadrupole mass spectrometers. Additionally, quantitative precision is improved by isotope coded derivatization reagent (ICD) strategy enabling individual internal standard for all targeted analytes.

2. Materials

2.1. Derivatization reagents and equipment

The following reagents are needed for MCF derivatization (all analytical grade reagents are bought from Sigma-Aldrich). All solutions used for MCF derivatization should be prepared using distilled de-ionized (IF) water > 18 MΩ at 25°C and analytical grade reagents. 1 and 2 solutions need to be prepared in IF-water prior to derivatization:

1. 1.0 M NaOH: Weigh 4.0 g NaOH and transfer to a 100 mL volumetric flask and adjust with IF-water to 100 mL. For 4.0 M NaOH solution use 16.0 g NaOH.
2. 50 mM $NaHCO_3$: Weigh 0.42 g $NaHCO_3$ and transfer to a 100 mL volumetric flask and add IF-water up to 100 mL.
3. Pyridine
4. Methyl chloroformate (MCF)
5. Methanol
6. Chloroform
7. Na_2SO_4: (dried overnight at 500°C)

In addition, for ICD-protocol the following reagents are substituted for unlabeled methanol and MCF.

8. d_4-Methanol: #DLM-24-10x1, Cambridge Isotope Laboratories.
9. d_3-Methyl chloroformate (see Note 1).

Equipment needed for derivatization: 3.5 mL polypropylene (PP) tubes for single use alternatively silicone treated glass tubes, whirl mixer, spatula, automatic pipette 10-100 µL and 100-1000 µL range + pipette tips with high recovery, pasteur glass pipettes, GC-MS vials with inserts and RedRubber/PTFE thread seals.

2.2. GC-MS/MS instrumentation

1. For this method we used an Agilent 7890A series GC system coupled with an Agilent 7000B triple quadrupole MS, equipped with an autosampler (GC PAL, CTC Analytics AG), a DB-5MS + DG column (#122-5532G, J&W Scientific) 30 m long with 10 m guard column, 0.25 mm inner diameter, 0.25 µm film thickness, and Gooseneck Splitless liner (#20799-214.5, Restek) 4x6.5x78.5 mm for Agilent GCs. The GC should be operated in constant pressure mode with helium 6.0 as carrier gas.

3. Methods

3.1. Preparation of analytical standards for calibration curve and for internal standard use

1. Prepare 100 mM stock solutions according to manufacturer's specifications with respect to solubility and stability for each analyte. NB! methanol and ethanol should be avoided as solvents since they take part in the derivatization reaction. Acetonitrile is often a good substitute. All analytical standards should be stored at -80°C.
2. A 1 mM standard mixture (STD-mix) is prepared by transferring 100 µL of each analytical standard (100 mM) and add up to a final volume of 10 ml and distribute in appropriate aliquots (e.g. 500 µL) before freezing.

3.2. Methyl chloroformate derivatization protocol

1. Bring the STD-mix to room temperature together with the freeze-dried samples (see Note 2 for sample preparation of biological matrixes).
2. Preparation of STD-mix serial dilutions:
   - Add 300 µL of 1 mM STD-mix to a 3.5 mL PP tube (1:1 dilution)
   - Add 100 µL of 1 mM STD-mix and 300 µL IF water to another 3.5 mL PP tube (1:4 dilution), transfer 100 µL to a third PP tube and add 300 µL IF water (1:16 dilution), repeat the procedure for preparation of 1:64 and 1:256 dilutions (remember to remove 100 µL in the last dilution.)
3. Add 90 µL 4 M NaOH, 333 µL methanol, and 67 µL pyridine to the five different calibration solutions (see Note 3 and 4). In addition, add 10 µL 1 mM d_5-glutamate (see Note 5).
4. Biological samples: Dissolve the dried metabolites completely in 390 µL 1 M NaOH and transfer the samples into 3.5 mL PP. Add 10 µL 1 mM d_5-glutamate, 333 µL methanol, and 67 µL pyridine.
5. Mix the tubes on a whirl mixer for 5 s, and add 80 µL MCF and vortex for 60 s (see Note 6).
6. Add 400 µL cold chloroform, whirl mix for 10 s (see Note 7).
7. Add 400 µL 50 mM $NaHCO_3$, whirl mix for 10 s. Wait 1 minute to ensure good phase separation (see Note 8).
8. Use Pasteur pipettes and transfer the chloroform-phase (lower phase) to a new 3.5 mL PP tube ensuring no droplets of water is transferred together with the chloroform. Add 1-2

spatula of dry $NaSO_4$ (the solution should be transparent, if not add more $NaSO_4$), whirl mix for 5 s.

9. Transfer the water-free chloroform phase containing the MCF derivatized metabolites in to a GC-MS vial with insert.
10. A well-plate derivatization protocols has also been developed (see Note 9).
11. For the ICD-protocol: Transfer 30 µL of the d_3-methanol/d_3-MCF derivatized ISTD-mix into a GC-MS vial with insert and add 170 µL of the methanol/MCF derivatized samples (from 3.2.8). Mix carefully with a pipette.

3.3. GC-MS/MS analysis

1. The GC is operated in constant pressure mode with 1 bar operating pressure (see Note 10).
2. GC inlet temperature set to 290°C.
3. Sample injection (1 or 2 µL) is performed in pulsed split-less mode.
4. GC temperature gradient: 0-2 min: 45°C, thereafter a linear 10°C /min gradient to 300°C and finally kept at 300°C for 7.5 min resulting in a 35 min total run time (see Note 11).
5. MS transfer line temperature is set to 300°C.
6. The CI ion source is set to 300°C and operated in positive chemical ionization (PCI) mode with methane (2,25mL/min flow) as reagent gas (see Note 12).
7. The triple quadrupole MS was operated in multiple reaction monitoring (MRM) mode using nitrogen as collision gas (gas flow was set to 1.50 mL/min) (see Notes 13-19).

4. Notes

1. The d_3-MCF was synthesized in our own laboratory in the original publication (see [15] for detailed protocol). US laboratories can purchase d_3-MCF from Cambridge Isotope Laboratories (www.isotope.com). However, overseas shipment restrictions prevented us from buying from this company. We have later used a local chemical laboratory (www.chiron.no) for synthesis of d_3-MCF.

2. The type of biological sample determines which pretreatment should be used prior to derivatization. Proteins, if not already been removed during sample processing steps, should be removed prior to derivatization as they compete for derivatization reagent and can precipitate/ interfere with the chloroform/ water-methanol phase separation. This protocol is frequently used for protein precipitation: Add 400 µL acetonitrile to a 100 µL sample in an 1.5 mL eppendorf tube, mix and incubate on ice for 30 minutes, centrifuge, and transfer 250 µL (equal to 50 µL sample matrix) to a 3.5 mL PPT tube and freeze the samples in liquid nitrogen. Freeze-dry the samples until dry.

3. Remember to saturate the pipette-tip with liquid prior to pipetting when solutions with low viscosity are used (e.g. methanol, MCF and chloroform)

4. The volumes can be scaled as long as the ratios are kept.

5. d_5-glutamate is used as technical internal standard and also for retention time locking (RTL).

6. Original procedure uses 2 times addition of MCF and 30 seconds mixing in between, but one time MCF addition and 60 seconds mixing yield the same result.

7. Chloroform should be kept sealed (minimize presence of oxygen) at 4°C temperature in order to prevent chemical degradation.

8. Lipids might impair the phase separation and a flocculation emulsion layer will form in-between the water- (upper) and chloroform (lower) phases. A centrifugation step can ease phase separation if it is difficult to isolate and selectively pipette the chloroform phase.

9. A robotic protocol using 2 mL 96-deepwell plate (#780271, Greiner Bio-One) has been developed. The working volumes are half of those used in the standard 3.5 mL PP tube protocol and the same calibration STD-mix dilution series are used. Polystyrene flat bottom 96-well plates from Greiner Bio-One (#6551163) are used for distribution of the 1M NaOH, methanol and 50 mM $NaHCO_3$ solutions, while u-shaped bottom polypropylene 96-well-plates from Greiner Bio-One (#650261) was chosen for chloroform, pyridine and methyl chloroformate. A robotic Beckman Coulter Biomek NX^P liquid handling station was used for transferring liquids to the reaction plate containing wither standards or biological extract, whirl mixing, loading new 2 mL 96-deepwell plates for drying of chloroform phase, transferring into GC-MS vials with inserts.

10. We run the instrument in RTL-mode typically acquiring RTL data at 1 bar pressure ± 10 and 20 % when the column is new. The Agilent Mass Hunter software calculates new operating pressure to maintain the same retention times after instrument maintenance with GC-column cutting. This ease the data analysis since retention time information can be kept unchanged in the quantitative data analysis program.

11. A 30 ºC/ min gradient was used in the original method [15]. However, we recommend the longer 10 ºC/ min version if there is no limitation in instrument access. The throughput becomes lower but the method is easier to maintain since our instrument only permit time segments and not dynamic MRM with individual time windows on each MRM transition.

12. The more traditional electron impact (EI) ionization source can also be used. The EI and PCI sources were compared during development of this method, and we found it easier to establish unique MRM transitions using the softer PCI source as this retained more high molecular fragments. The PCI method also turned out to be more sensitive than the EI method for the tested metabolites.

13. The GC-MS/MS sequence starts with a solvent (chloroform) run, followed with a blank run, standard quality control mixture run and the standard mixture serial dilution samples before the real samples. We usually don't analyse more than twelve samples before a second standard quality control mixture sample is ending the sequence.

14. As a general rule to maintain selective and sensitivity are both quadrupole 1 and 3 set to unit mass resolution and dwell-time is >10 ms for all transitions. Electron multiplier voltage may be increased when running in MRM mode in order to improve sensitivity. However, gain settings will be instrument dependent and must be seen in direct relation to the total electron multiplier voltage (EMV). Therefore no general value can be recommended, but in our experiments a value between 10 and 50 is applied when running in MRM mode.

15. The MRM settings (ion pair transitions and collision energy) can be found in [15] for the MCF/d_3-MCF version of the method. As the restriction of MCF shipment may prevent laboratories to get d_3-MCF, we are currently developing alternatives, comprising use of labeled alcohol only and use of ethyl chloroformate, to the original method. These alternatives clearly are poorer than the original MCF/d_3-MCF method (less sensitive as low intensity MRM transitions must replace original high intensity MRM transitions to maintain selectivity), but the analytical precision is higher than when using external standards only (unpublished data).

16. Comprehensive Metabolite Profiling MS methods are a compromise between selectivity, sensitivity, and throughput. Qualifier MRM transitions could be included to increase the reliability, but this must be evaluated with regard to instrument performance (max number of transitions per second) and number of high intensity MRM transitions available for the individual metabolite.

17. Miscellaneous artifacts can be introduced during chemical derivatization of analytes, e.g. for aldehydes lacking hydrogen in the alpha position the basic conditions during MCF derivatization will lead to two products: reduction to the corresponding alcohol and oxidation to the corresponding carboxylic acid (i.e. the Cannizzaro reaction). The latter can be derivatized by MCF and, thus, detected by the GC-MS/MS analysis, e.g. when methylglyoxal is dissolved in a strong base as 1 M NaOH it will form lactate and its corresponding alcohol (1-hydroxypropan-2-one). Additionally, many analytes yields more than one product during derivatization, especially prominent in GC-MS analysis of silylated sugars, but chloroformate derivatization may also yield several derivatives. This is more of a challenge for MS scan Metabolite Profiling methods, but it is also very important to evaluate the behavior of all metabolites included in targeted MRM methods.

18. Data analysis: The standard curves should in general not be extrapolated. They can be forced through zero but only after visual inspection of individual curve. Results should not be reported if the highest standard concentration is exceeded.

19. Comprehensive Metabolite Profiling methods are important in the Metabolomics field that aims to cover (and preferentially quantify) as many metabolites as possible. But, clearly compromises must be taken during development, and it cannot be expected that the analytical precision holds the same level as dedicated single-/few-analyte methods.

References

1. Janeckova H, Hron K, Wojtowicz P, Hlidkova E, Baresova A, Friedecky D, Zidkova L, Hornik P, Behulova D, Prochazkova D, Vinohradska H, Peskova K, Bruheim P, Smolka V, St'astna S, Adam T (2012) Targeted metabolomic analysis of plasma samples for the diagnosis of inherited metabolic disorders. Journal of Chromatography A 1226:11-17. doi:10.1016/j.chroma.2011.09.074

2. Kaspar H, Dettmer K, Chan Q, Daniels S, Nimkar S, Daviglus ML, Stamler J, Elliott P, Oefner PJ (2009) Urinary amino acid analysis: A comparison of iTRAQ (R)-LC-MS/MS, GC-MS, and amino acid analyzer. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 877 (20-21):1838-1846. doi:10.1016/j.jchromb.2009.05.019

3. Kaspar H, Dettmer K, Gronwald W, Oefner PJ (2009) Advances in amino acid analysis. Analytical and Bioanalytical Chemistry 393 (2):445-452. doi:10.1007/s00216-008-2421-1

4. Ramautar R, Somsen GW, de Jong GJ (2013) CE-MS for metabolomics: Developments and applications in the period 2010-2012. Electrophoresis 34 (1):86-98. doi:10.1002/elps.201200390

5. Bruheim P, Kvitvang HFN, Villas-Boas SG (2013) Stable Isotope coded Derivatization reagents as Internal Standards in Metabolite Profiling. Journal of Chromatography A doi:10.1016/j.chroma.2013.03.072

6. Kanani HH, Klapa MI (2007) Data correction strategy for metabolomics analysis using gas chromatography-mass spectrometry. Metab Eng 9 (1):39-51. doi:10.1016/j.ymben.2006.08.001

7. Lien SK, Kvitvang HFN, Bruheim P (2012) Utilization of a deuterated derivatization agent to synthesize internal standards for gas chromatography-tandem mass spectrometry quantification of silylated metabolites. Journal of Chromatography A 1247:118-124. doi:10.1016/j.chroma.2012.05.053

8. Little JL (1999) Artifacts in trimethylsilyl derivatization reactions and ways to avoid them. Journal of Chromatography A 844 (1-2):1-22

9. Villas-Boas SG, Smart KF, Sivakumaran S, Lane GA (2011) Alkylation of Silylation for Analysis of Amino and Non-Amino Organic Acids by GC-MS? Metabolites 1:3-20. doi:10.3390/metabo1010003

10. Husek P (1998) Chloroformates in gas chromatography as general purpose derivatizing agents. J Chromatogr B 717 (1-2):57-91. doi:10.1016/s0378-4347(98)00136-4

11. Esterhuizen-Londt M, Downing S, Downing TG (2011) Improved sensitivity using liquid chromatography mass spectrometry (LC-MS) for detection of propyl chloroformate derivatised beta-N-methylamino-L-alanine (BMAA) in cyanobacteria. Water Sa 37 (2):133-138
12. Guo T, Geis S, Hedman C, Arndt M, Krick W, Sonzogni W (2007) Characterization of ethyl chloroformate derivative of beta-methylamino-L-alanine. J Am Soc Mass Spectrom 18 (5):817-825. doi:10.1016/j.jasms.2007.01.006

13. Kaspar H, Dettmer K, Gronwald W, Oefner PJ (2008) Automated GC-MS analysis of free amino acids in biological fluids. Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences 870 (2):222-232. doi:10.1016/j.jchromb.2008.06.018

14. Villas-Boas SG, Delicado DG, Akesson M, Nielsen J (2003) Simultaneous analysis of amino and nonamino organic acids as methyl chloroformate derivatives using gas chromatography-mass spectrometry. Analytical Biochemistry 322 (1):134-138. doi:10.1016/j.ab.2003.07.018

15. Kvitvang HFN, Andreassen T, Adam T, Villas-Boas SG, Bruheim P (2011) Highly Sensitive GC/MS/MS Method for Quantitation of Amino and Nonamino Organic Acids. Analytical Chemistry 83 (7):2705-2711. doi:10.1021/ac103245b

Appendix V

5. Analytical techniques from the perspective of microbial metabolomics

Per Bruheim[1],Trygve Andreassen[1,2], Stina K. Lien[1],

[1] Department of Biotechnology, Norwegian University of Science and Technology

[2] MR Core Facility, Department of Circulation and Medical Imaging, Norwegian University of Science and Technology

The biological sciences have undergone an 'omics revolution during the last two decades. Due to technological advances it is now possible not only to sequence any genome at a reasonable price and within a short time frame, but also to synthesise whole genomes, although only at the microbial level for the time being. In parallel with the development of genome sequencing and gene expression technologies, significant technological and methodological advances have been introduced for the study of protein and metabolite pools. However, the different 'omics fields still face challenges and are continuously evolving. Metabolomics faces particular analytical challenges due to the diverse physico-chemical properties, from highly charged to hydrophobic species, and wide dynamic range of abundances of the different metabolites within a sample, ranging from less than a picomolar up to millimolar intracellular concentrations. While many protocols for sampling, sample processing and analyses have been developed, these still have room for improvement; and there is much scope for new protocols to be developed so that biological questions can be efficiently answered.

This chapter focuses on the analytical techniques used to explore the metabolite composition in a sample. Predominantly two analytical platforms are used for the analysis of metabolites: *mass spectrometry* (MS) and *nuclear magnetic resonance* (NMR), and only these will be described in this chapter. The prerequisite for MS detection is that the metabolite is charged or can become charged during the ionization process, since the universal principle of MS is based on the detection of a molecule's mass per charge. MS is popular due to its sensitivity and high resolving power when combined with a separation step. NMR is less discriminatory than MS in the sense that all metabolites present at high concentrations can be detected, while metabolite detection on a MS instrument is dependent upon instrument settings, e.g negative ionization is used for the detection of negatively charged metabolites or metabolites that can be made negative during the ionization process, and likewise for positive ionization. One strength of NMR is that the analysis requires less processing of the sample than for MS-analysis, as whole cells and tissues can be directly analysed without any extraction step.

Mass spectrometry based metabolomics

Different methodological approaches apply for MS Metabolome studies. The simplest is the MS Fingerprinting. A biological extract can be injected directly into a MS instrument either with *direct infusion* (DI) or *flow-injection analysis* (FIA). The acquired MS fingerprint (i.e. a mass spectrum) can be used for the classification of samples, but these kinds of analyses are very sensitive to matrix-ionization suppression/competitive ionization and can lead to high variation even between repeated injections. Therefore, in-depth studies of the metabolome requires a separation step to resolve the metabolite species in time before they enter the MS instrument. Also, there are many examples of metabolites with same molecular formula (e.g. hexose-phosphates); thus, the MS analyser is not able to discriminate between the individual metabolite species, and therefore a separation stage is required.

The choice of separation technology depends on which metabolite group(s) that is under investigation (Figure 1). Traditionally, this has been separated in two classes: the metabolites that can be separated by *gas chromatography* (GC) and the metabolites that can be separated by *liquid chromatography* (LC). Usually, the division follows the line of which metabolites are volatile or can form thermo stabile volatile derivatives and those that can't. The former metabolites are amenable to GC separation, while the latter metabolites are separated with LC. Higher molecular weight metabolites usually must be separated on LC. GC and LC have been, and still are, the two dominant separation technologies for MS based Metabolomics. There is more overlap in application area than indicated in Figure 1, e.g. amino and fatty acids can also be separated by LC. However GC and LC are the preferred choice of separation technology for the metabolite groups in the left box and right box in Figure 1, respectively. A third method, *capillary electrophoresis* (CE), is not frequently present in MS Metabolomics laboratories, but clearly have its' niche with excellent separation efficiency for certain metabolites groups. These three technologies are described more in detail later in this chapter.

Additionally, two other separation technologies are indicated in Figure 1; Ion chromatography (IC) and convergence chromatography (UPC^2). Ion (exchange) chromatography has a unique selectivity in separation of ionic analytes, e.g. there are applications showing baseline separation of the analytically challenging hexose-phosphate metabolites. However, the high salt mobile phases used in IC are not compatible with MS-detection, but the manufacturer has developed a suppressor cell for converting the non-volatile mobile phase to water thus compatible with MS detection. And a recent downscaling from analytical scale to capillary scale (ICC) makes this separation technology even more promising for Metabolomics research since it was shown that the sensitivity increased several order of magnitudes compared to the analytical IC scale. UPC^2 stands for

UltraPerformance Convergence Chromatograhy™ and is based on supercritical fluid extraction (SFE) principles. There are also other manufacturers that offer technical solutions based on SFE principles. The primary mobile phase is compressed carbon dioxide but organic solvents such as methanol can also be mixed into the flow. This refined SFE technology entered the marked just recently, and there are still few applications published. It shows potential for separation of medium volatile and medium polar metabolites. Thus, it overlaps with both GC and LC application areas. It also shows promising properties for chiral separations.

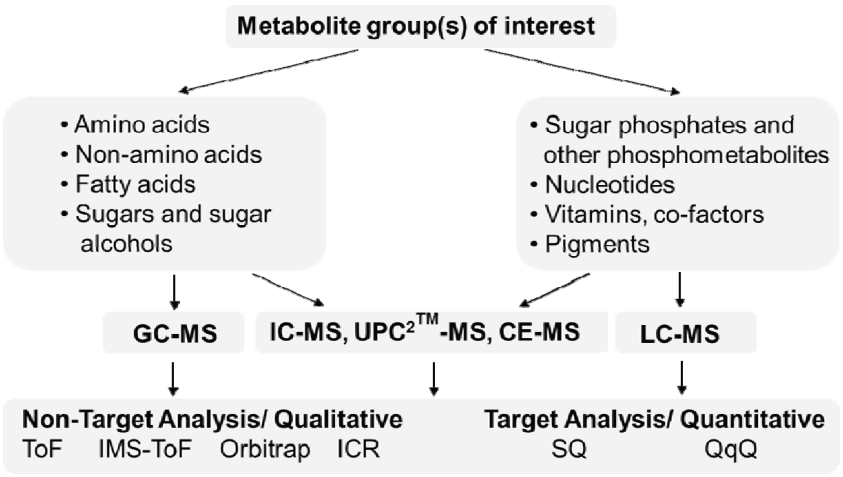


Figure 1. Outline of MS based Metabolomics. Metabolite groups are preferentially separated with different separation technologies dependent on the physic-chemical properties of the metabolites. Metabolite pool composition of biological extract can be analysed with different approaches on different MS instruments.

MS analysis can either be performed in a non-target or target way. The non-target analysis often combines MS acquisition in scan mode with multivariate and statistical analysis of the data to identify and rank the most important masses. The results are given in abundances. Thus, a non-targeted analysis is qualitative and not quantitative since the data is not calibrated to known metabolites at known concentrations. Contrary, target analysis aims at quantification of known metabolites. There are a number of different MS instrument types with quite different properties, some types better for target analysis and some types better for non-target analysis (Figure 1, lower part). The quadrupole MS instruments, for example, are well recognized for their excellent properties in quantitative analysis. The

single quadrupole (SQ) is the simplest and cheapest MS instrument available and can be operated both in scan and in *single ion monitoring* (SIM) mode, the latter being used for quantitative analysis of known compounds with the MS only analysing pre-programmed masses. The *triple quadrupole* (QqQ) instrument has both higher sensitivity and selectivity than the SQ as it can make use of unique fragmentation patterns arising in the collision cell in the MS. In *multiple monitoring mode* (MRM) the metabolites are first selected by the mass filter of Q1 (the first quadrupole) and then fragmented in Q2 (i.e. the collision cell), while Q3 works as a mass filter for selected unique fragment ions created from the precursor ion. One significant advantage with QqQ is that it can easily be combined with isotope dilution strategies for absolute quantification using ^{13}C-labelled standards, (Uehara et al., 2009; Wu et al., 2005). There are also very sensitivity quadrupole–linear trap MS instruments with MS^3 functionality on the market. The *ion trap* MS instrument is used both for quantitative and qualitative purposes. Its uniqueness lies in the MS^n fragmentation functionality which is a powerful tool in structural elucidation. While the ion trap itself has unit mass accuracy detection, it can be connected and eject ions into high accuracy mass spectrometers as *Time-of-flight* (ToF), *Orbitrap* and *Fourier transform ion cyclotron resonance* (FT-ICR). FT-ICR, for example, has a mass accuracy of better than 1 ppm and resolution over to 2 million FWHM (Junot et al., 2010). ToF MS instruments benefit from high scan rates (30 to 100 Hz) while the FT-MS types (i.e. Orbitrap and ICR) are slower working types, e.g. the Orbitrap MS must be operated in ToF resolution range (15,000 to 30,000 FWHM) to be able to acquire 2 to 3 scans per second. However, they are attractive in metabolomics applications due to their stability in accurate mass measurements. High mass accuracy MS instruments are the preferred choice for non-target analysis. More recently, the introduction of an *ion mobility*-MS instrument (IMS) has made it possible to separate and discriminate between structural isomers and other pairs of molecules with the same mass that are difficult to separate by GC and LC (Dwivedi et al., 2008); in principle it should be possible to separate the previously mentioned and analytically challenging hexose-phosphates. IMS separates ions based on their differential mobility in gas phase under the influence of a uniform weak electric field.

Clearly, equipping a laboratory for all kinds of MS-based Metabolomics applications requires several instrumental platforms. Büscher and co-workers (2009) tested different instrumental platforms for 91 metabolites (covering glycolysis, pentose phosphate pathway, tricarboxylic acid pathway, redox metabolism, amino acids and nucleotides). They concluded that for analysis on a single platform, LC provided the best coverage. However, GC is the platform of choice for smaller and more stable molecules due to its high chromatographic resolution. For a comprehensive metabolome analysis, several metabolite profiling methods should be applied. For example, van der Werf and co-workers (2007) used six different analytical MS methods for comprehensive coverage of the *E. coli* metabolome , and three

methods for a comprehensive analysis of the *Pseudomonas putida* S12 metabolome grown on different carbon sources (van der Werf et al., 2008). These studies included both LC-MS and GC-MS methods. If only a few metabolites are of interest then a metabolite target method is enough, and metabolite fingerprinting is, as already discussed, just used for classification of all large number of samples. LC-MS is preferable for the latter two approaches since *high throughput* protocols with run times down to 1–3 minutes can usually easily be developed. This enables the analysis of a high number of samples (over 500 per day per instrument). The most challenging step is to determine whether GC or LC should be used as the separation technique. The pros and cons of GC and LC will be discussed in more detail in the following paragraphs.

GC-MS

GC-MS instruments have been available for a much longer period than LC-MS instruments, as it has been technically simpler to interface the GC with a MS detector. It is a robust technology for the analysis of volatile compounds and also for semi-volatile compounds that can be derivatized to increase their volatility. The GC has better separation efficiency than LC, and GC-MS has traditionally been regarded as the prime instrumentation for metabolite profiling analysis (Fiehn, 2008). However, this prominent role of GC-MS is challenged by developments in LC-MS technology and the recent introduction of UPC^2 convergence chromatography.

As the MS can only detect charged molecules, generation of ions is the function of the ion source connecting the GC with the MS. The traditional ion source for GC-MS is EI. This is a rather strong ionization (70eV) which results in fragmentation of the molecules, and the molecular ion is for most analytes not observed in the resulting mass spectra. However, the fragmentation pattern comprised of all the fragments generated during the ionization is often unique for each analyte. EI is a reproducible ionization mode and different GC-MS instruments produce comparable fragmentation patterns, even for instruments made by different manufacturers. This has made it possible to build large databases containing fragmentation patterns for many compounds (eg. NIST/EPA/ NIH library containing over 170,000 entries and Wiley mass spectral database with 600,000 entries). These databases may not have a comprehensive coverage of metabolites as the traditional focus has been on other chemicals. However, with the recent focus on metabolome research, metabolite-specific databases, both open internet searchable and commercial databases have been constructed (e.g. the Agilent Fiehn database containing fragmentation spectra of over 700 metabolites derivatized with MSTFA).

GC can be interfaced with most types of MS detectors and information from instrument manufacturers indicates that new high-end instrument solutions will soon enter

the market; which emphasize that GC-MS will maintain a significant role as analytical technique in Metabolomics. Standard GC-MS instruments can also be equipped with chemical ionization sources and operated in both negative and positive mode. There is also possibility to connect GC instruments via an atmospheric pressure chemical ionization (APCI) source to MS instruments build for LC-MS application, e.g a custom made GC-Orbitrap MS has been built (Peterson et al., 2010) and most manufacturers offers commercial solutions. While the traditional MS detector for GC has been the quadrupole, recently GC-ToF MS has become popular as this configuration enables either higher mass accuracy determination or higher scan speed rates. The latter is often combined with fast GC and the advanced GCxGC mode, resulting in extremely high chromatographic resolution. In GCxGC mode there are two GC columns with different stationary phases serially connected to increase the resolving power. The first column, usually a long non-polar column, is connected to a short polar column with either a cryogenic modulator or a flow modulator. There are already many examples in the literature of studies of microbial metabolomes using GCxGC-ToF MS, e.g. identification of metabolite changes in yeast cultures exhibiting oscillatory behaviour and study of the yeast metabolome under respiratory and fermentative conditions (Mohler et al., 2006; Mohler et al., 2008).

The prerequisite for GC-MS analysis is that the analytes are thermally stabile and volatile since the elution occurs in gas phase, usually with an increasing temperature gradient. Most metabolites are not volatile (at least in the temperature range for operation of a GC) and must therefore be made volatile through derivatization. The most common derivatization reactions are silylation, alkylation and acylation. Derivatization agents are commercially available for each of these and most derivatization reactions are simple one- or two-step procedures. However, it is important to be aware that formation of by-products and analyte conversions (e.g. arginine to ornithine) can occur during derivatization (Garcia et al., 2008; Halket et al., 2005).

The choice of derivatization agent is based on the functional group of the metabolite of interest. Silylation has been the most popular derivatization method for analysis of metabolites since it is a versatile derivatization reaction which covers many important metabolite classes such as sugars, alcohols, amines, amino acids, non-amino (phospho-) organic acids and aromatic compounds. Derivatization with silylation is often performed in two steps, where the first step is an oximation of carbonyl groups to oximes that stabilizes the reducing sugars in the open-ring formation and also prevents decarboxylation of α-ketoacids. Several metabolite profiling TMS GC-MS studies have reported identification of over 100 metabolites, underlining the strength and potential of this methodology (Koek et al., 2006; Strelkov et al., 2004). In addition, the deconvolution processing software detected several hundred other unidentified compounds that were not contained in the databases.

One particular challenge with silylation is that many silylated metabolites are sensitive to the presence of water, especially when derivatized with MSTFA. Silylation reactions must be carried out in anhydrous conditions and the samples must be kept dry to minimize degradation of TMS-derivatives. It is shown that different metabolites have different derivatization rates. Thus, it is important to optimize the derivatization protocol. Kanani (2007; 2008) and co-workers have developed analytical strategies to face this challenge, but the in-stability of TMS-derivatives makes quantitative analysis extra challenging. Use of single, or a few, internal standards does not provide a satisfactory solution either, since their instability will not be representative for all metabolites in the sample. A potential solution is to introduce internal standards for each targeted metabolite by running separate derivatization of a standard mixture with deuterated derivatization agent. But this has been shown difficult both from an analytical point of view and deuterated MSTFA is also expensive (Lien et al., 2012a). Instrumental solution to this has been to use advanced auto-samplers with the option to run derivatization reactions on-site. This makes it possible to run many samples in a sequence with the constant time interval from derivatization to analysis. t-Butyldimethylsilylation (MTBSTFA) is less sensitive to moisture than MSTFA and reported to yield more stable derivatives, but TBS derivatives are bulkier than TMS derivatives and do not enable analysis of sugars. The choice of silylation agent is, therefore, dependent on the metabolite classes to be analysed.

Alkylation is another popular derivatization method for GC-MS analysis. Fatty acids are most often analysed on GC-MS by derivatization to their respective fatty methyl esters with BF_3 and methanol. Methyl chloroformate (MCF) derivatization has proven to be an excellent alternative to silylation for amino acids and non-amino organic acids (Husek, 1998; Villas-Boas et al., 2003). The advantage of the MCF method is that it is a rapid one-step method that can be carried out in aqueous solution using relatively inexpensive reagents and the derivatized metabolites are extracted to a chloroform phase resulting in less chemical contamination of the GC-MS. It has been shown that the MCF method is superior to the TMS method for amino acids and organic acids (Villas-Boas et al., 2011). Recently, the MCF method was further developed from a single quadrupole scan method to a triple quadrupole MRM method (Kvitvang et al., 2011). The EI source was replaced with a CI source and operated in positive CI mode. CI is a softer ionization method which preserves molecular ions and other larger fragments at a higher frequency than EI, and this was exploited to establish unique MRM transitions to over seventy amino acids and non-amino organic acids. Another feature was also included in this method: absolute quantification through the use of custom made deuterated MCF for introduction of an internal standard for all metabolites in the method.

The GC is a robust instrument with excellent chromatographic resolution, but it must be carefully operated and maintained optimal performance. The GC column must often be cut during maintenance and this shortening of column length results in earlier elution of analytes. One option to compensate for earlier elution and to avoid re-calibration of standard tables is to operate the GC in retention time locking (RTL) mode which is a software feature offered by some GC producers. RTL works best in constant pressure mode with a constant linear temperature gradient. The principle is to run an initial calibration on a new column at 0, ±10% and ±20% of running pressure and determine how the elution time of a RTL-standard varies with pressure. A correlation curve can be calculated based on this information which is used at each column shortening to estimate the lower operating pressure for keeping elution times the same as when the column was new and uncut. In our laboratory, our experience has been that this also eliminates problems with column cutting and eventually replacement.

LC-MS

Liquid-chromatography interfaced to a mass spectrometer detector has become an important and versatile analytical tool for the study of biomolecules; not only low-molecular metabolites but also biopolymers such as proteins, carbohydrates and nucleic acids can be analysed with LC-MS. Hence, the molecular weight range is much broader for LC-MS than for GC-MS. The invention of the soft ionization mode, or *electrospray ionization* (ESI), by John Fenn who was awarded the Nobel Prize in Chemistry in 2002, paved the way for analysis of biological molecules with LC-MS. Subsequent technological developments have resulted in more sensitive, robust and accurate instruments. Currently, the most sensitive instruments can detect down to attomole quantities.

There have also been significant developments in LC technology. U(H)PLC (ultra-high pressure, also called rapid resolution) operates at significantly higher pressures (>1000 bar vs. 400 bar for standard HPLC) and uses columns packed with smaller particles (sub 2μm). This greatly increases the chromatographic resolution and efficiency (Want et al., 2010). The higher chromatographic resolution is advantageous for both non-target and for target metabolite profiling analyses as it significantly shortens the analysis times and thereby increases throughput. Nano-LC has been particularly popular for proteomics applications as it significantly increases sensitivity. While nanoLC and nanoESI sources in their infancy suffered from low stability, due to technological advances these technologies are now much more robust. The run time is usually longer than traditional analytical LC, but if increased sensitivity is required then nanoLC-MS is an excellent alternative, especially for non-target metabolite profiling using accurate mass and high resolution mass spectrometers. However,

there has been a marked improvement in the ionization sources for analytical scale separations (0.05-2 ml mobile phase/ min); thus approaching nano-scale separations in sensitivity. But, still there is large room for improvement of ionization efficiencies (Wilm, 2011).

LC-MS analysis of metabolites can roughly be divided in two large groups: those that can be separated using reverse phase (RP) chromatography and those that cannot. RP stationary phases (usually C_8 or C_{18}) are robust and reproducible phases that, together with MS compatible volatile mobile phases such as water, methanol, acetonitrile and dichloromethane, work excellently for non-charged and slightly to very hydrophobic analytes. Usually the pH of the mobile phase is adjusted to ease ionization (low pH for positive ionization and high pH for negative ionization) with formic and acetic acid and ammonium hydroxide. For the most hydrophobic metabolites, e.g carotenoids, *atmospheric pressure chemical ionization* (APCI) works better than the ESI source. Even β-carotene without any oxygen atom can be ionized by APCI. However, the general guideline is that if either an oxygen or a nitrogen containing functional group is present in the molecule, the likelihood of ionization and subsequent detection by a MS using either an ESI or an APCI source is increased.

Many metabolites, especially primary metabolites in central metabolic pathways, are highly polar and charged metabolites. Thus, they are not easily retained by standard RP elution conditions. Mainly there have been two solutions to this challenge. The first has been to modifiers to the mobile phases modifiers, so-called *ion pair reagents*, such as tributylamine, dibutylammonium, hexylamine and octylammonium (Coulier et al., 2006; Luo et al., 2007; Seifar et al., 2009; Wamelink et al., 2005). These positively charged molecules mask the negative charge of the analytes, thereby improving the interaction between the hydrophobic RP stationary phase and the analytes. However, this approach has several disadvantages since it impairs ionization efficiency, resulting in lower sensitivity of detection. A practical disadvantage is that the ion pair reagents attach strongly to all surfaces in the instrument so that extensive cleaning procedures must be applied when changing to a new analytical protocol. Special precautions for not damaging the MS detector must be taken when switching from negative to positive ionization as the ion pair reagents are detected with high sensitivity in positive mode. Therefore, it is preferable to have dedicated LC-MS instruments for ion pair chromatography. Another disadvantage is that the LC columns have shorter lifespans for ion pair reagent mobile phases. Metal chelating agents have also been tested for polar anionic metabolome analyses, with promising results (Myint et al., 2009) but without becoming a frequently used method. Therefore, there has been a strong focus on testing out alternative column and mobile phase technologies. *Hydrophilic interaction liquid chromatography* (HILIC) has received the most attention and has become increasingly

popular for separation of polar analytes (Cubbon et al., 2010; Jandera, 2008). HILIC uses LC columns with silica surfaces that might or might not be slightly modified but still retain their normal phase (NP) behaviour. Common modifications include chemical binding to the silica gel with amino-, amido-, cyano-, carbamate-, diol- or polyol- ligands. These column systems are operated with a high percentage of organic solvent at the start and then increasing the percentage aqueous phase during the gradient elution. While the traditional mobile phases for NP chromatography have been only organic solvents, the HILIC columns can be used with water. The usual mobile phase combination has been acetonitrile and water with ammonium and acetic acid to adjust the pH. Development of HILIC columns has significantly advanced, and, while these chromatographic systems earlier suffered from poor stability and reproducibility, more and more reports are being published with the use of HILIC technology showing that the chromatographic performance of HILIC columns have improved significantly. Bajad and co-workers (2006) tested seven different HILIC LC columns and could reliably detect over 140 metabolites from the column with best performance, while Schiesel and co-workers (2010) developed a quantitative zwitterionic ZIC-HILIC LC-QqQ MRM covering both central primary metabolites and secondary metabolites of β-lactam antibiotics synthesis, including their precursors and extracellular degradation products. Hinterwirth and co-workers (2010) used mixed-mode hydrophilic/weak anion exchange chromatography to separate phosphorylated carbohydrate isomers. This is an interesting method for analysis of the important metabolite groups of sugar phosphates. However, they had to add a mobile phase modifier (tributyl amine) to obtain the required separation efficiency. Therefore, the method must be further developed as ion modifiers are not preferably used with MS detection. In any case, HILIC will probably be very important chromatographic choice in future LC-MS based Metabolomics, as well as the earlier mentioned ICC and UPC^2 separation technologies.

The Metabolomics community would benefit if agreement on a few methods could be reached. However, before that can be done more and different LC-column technologies must be tested, as there is no unique technology that stands out with superiority today, e.g. Yang and co-workers (2010) recently tested a pentafluorophenylpropyl (PFPP) modified silica based stationary phase and reported better performance than the optimized system by Bajad and co-workers(2006). Although one advantage LC-MS has over GC-MS is that it doesn't require derivatization, derivatization can be an alternative for LC-MS analysis of metabolite groups for which good chromatography conditions are difficult to establish (Yang et al., 2007). Derivatization also enables the inclusion of internal standards for all analytes through the use of labelled derivatization agents, as earlier discussed for the MCF GC-MS/MS method. Yang et al. (2008) used this strategy to analyse aniline derivatized glycolytic, pentose phosphate pathway and tricarboxylic acid metabolites . Clearly, different

choices are available, all with pros and cons, and the focus must not only be on which method provides the best chromatographic performance but also detection limits since the different elution conditions can have a significant effect on ionization efficiency. It is important, therefore, that new protocols not only be developed and tested on pure standards but also on real biological extracts as various matrix effects and ion suppression processes can have a strong negative effect on sensitivity and quantitative reproducibility (Remane et al., 2010; Stahnke et al., 2009).

The two common ion sources (ESI and APCI) used in LC-MS are both considered soft ionization as the molecular ion (either protonated or deprotonated) is preserved during the ionization stage. This is advantageous for identification of unknown metabolites that are scored as interesting during the analysis. The unknown metabolite is represented as a mass in the mass spectrum. If high mass accuracy MS instruments were applied then the experimentally measured mass can be used to either guess the molecular formula or directly search metabolite databases. However, it can be challenging to unambiguously identify an unknown metabolite. The metabolite might not be identified even by correctly guessing the molecular formula because different metabolites can have the same formula, e.g. $C_5H_8NO_4$ is the formula for glutamic acid but, according to the mass list made by Arita (http://www.metabolome.jp/), over ten other metabolites have the same formula. True identification, therefore, requires comparison of the unknown metabolite to standards with respect to identical retention times and MS/MS fragmentation patterns. While for GC-MS there are comprehensive databases of EI fragmentation spectra, it has not been possible so far to repeat this strategy for LC-MS. First, the LC-MS ion sources are softer and in most cases preserve the metabolite, implying that a unique fragmentation pattern is not generated. The possibility of generating fragmentation in a LC-MS therefore is limited to the MS instruments with MS/MS functionality. However, the arising fragmentation pattern varies depending on type of instrument (3D and linear trap, QqQ, QToF) which makes these fragmentation patterns less amenable to database building and searching as the GC-MS EI databases. Therefore, most of the databases built on LC-MS are only valuable at one location and on only one instrument as there can be significant differences between same kind of instrument made by the same manufactureres (i.e. it is well known that collision energies must be optimized when transferring methods between presumably identical instruments, even those made by the same manufacturer). However, there have been attempts to counter instrumentation differences by tuning of fragmentation conditions and summing several spectra acquired at different collision energies (Josephs and Sanders, 2004) and several LC-MS MS/MS libraries have become available (see http://metlin.scripps.edu/ and http://www.nist.gov/).

CE-MS

CE-MS has excellent separation efficiency for certain groups of biomolecules, such as proteins, nucleotides, nucleic acids, amino acids and (oligo-/poly-)saccharides (Klampfl, 2006). While GC and LC are pressure driven separation techniques, CE is electro-driven and based on the mobility of ions in a capillary under high voltage conditions. As with GC-MS and LC-MS, there have also been significant technical advancements in CE instrumentation and coupling to MS detectors (Klampfl, 2004; Klampfl and Buchberger, 2010). One research group in particular has been at the forefront in the use of CE-MS for metabolomics applications, and several central papers from this group should be accessed for learning about the potential of CE-MS for the analysis of several important metabolite classes as amino acids, organic acids, sugar phosphates and nucleotides (Ohashi et al., 2008; Soga et al., 2007; Soga et al., 2003). In the highly cited paper from 2003, Soga and co-workers identified over 150 metabolites using three different CE-MS methods: one for each of anionic and cationic metabolites and one for nucleotides and co-enzymes. Analysis of positively charged ions is simplest as pure fused silica capillary can be used together with an acidic electrolyte (e.g. 1 M formic acid). Negatively charged molecules are more difficult to analysis as the electroosmotic flow (EOF) must be reversed. Anions can also be adsorbed by the silica surface and different masking and pressure gradient techniques have been developed to counter this (Harada et al., 2006; Soga et al., 2007). Even sugar phosphate isomers have successfully been separated on CE-MS (Hui et al., 2007). Probably the main reason why CE-MS does not have the same outstanding role as GC-MS and LC-MS in metabolomics applications is its significantly lower sensitivity. While a μl sample is injected into a GC and LC, only a nl sample is analysed on a CE. In addition, CE connected to MS is technically challenging, and current fluctuations can often be observed. The resulting migrating time shift complicates the subsequent data analysis, especially for non-targeted data analysis where migrating time is used as an additional identifier of the mass/charge measurement.

The importance of robust cultivation, quenching, extraction and concentration protocols in MS Microbial Metabolomics

One advantage with microbial systems is that most cultivation can be scaled up so that enough biomass can be harvested. Usually, extraction of 1−5 mg dry biomass gives high intensities for the most abundant metabolites in both GC-MS and LC-MS analysis. This represents a sampling volume of 1−5 ml if the culture is harvested at a cell concentration of 1 g DW/l or approximately 3−5 OD units. Microbial metabolome (and of course other ‘ome) analyses have the advantage that the cultivations can be performed in a fermenter. This is

the most robust cultivation set-up enabling both process monitoring and control, including constant pH and dissolved oxygen. These are critical parameters that can significantly influence the metabolic state of the cell. For extensive time-series sampling, a fermenter with an operation volume of 1-2 litres would represent a culture reservoir for extensive sampling.

Quenching of metabolism and the subsequent extraction of microbial cells were thoroughly covered earlier in Chapter 2 and 3 of the book. These are critical steps for the quality of the subsequent analysis and the choice of analytical method should be decided prior to optimization of sample processing conditions. Extraction of the metabolites into a liquid is necessary for MS analysis. Quenching must happen within the time frame of turnover rates for the metabolites of interest. Extraction must be complete and performed under conditions where the metabolites are not broken down. This is technically challenging and the perfect protocol still remains to be developed. Alkaline and acidic extraction should be avoided as this introduces high levels of salt, especially if concentration of the sample is necessary. Salt may lead to less reproducibility of GC-MS analyses since the derivatization step might be sensitive to the presence of salts. Here, the MCF method has its advantage as the derivatization reaction is done in 1 M sodium hydroxide. LC-MS analysis is also sensitive to salt, but this can to a certain degree be counteracted by establishing chromatographic conditions where the target metabolites elute after most of the salts have been washed out in the void fraction. However, care should be taken and the mass spectra should be inspected for adduct ions, especially Na-adduct ions. As a significant part of the metabolite can be detected as adduct ion, this should be taken into consideration in quantitative analysis.

Some cultivation must be performed as batch-cultivations, e.g. *Streptomyces* bioprocesses where a nutrient exhaustion is triggering the secondary metabolism and antibiotic biosynthesis, e g. Wentzel et al (2012a) developed a submerged batch fermentation strategy for systems-scale multi-omics studies, and used two MS methods to follow intracellular metabolite pool changes during the metabolic switch transition phase (Wentzel et al., 2012b). However, continuous cultivations (i.e. chemostat) is the optimal and preferred fermenter mode of operation for physiological studies. The culture is independent of time at steady-state conditions in a chemostat, and thus this is the most reproducible physiological state. Importantly, the growth rate can be directly controlled by the dilution rate. This enables studies of different media, different mutants, etc. under conditions where the growth rate can be eliminated as a variable. The growth rate itself might have a significant effect on the metabolite pool composition; thus comparison of different strains, mutants, cultivation conditions etc. should, preferentially for easiest interpretation, be performed at the same growth rate. Lien et al (2012b) studied *Pseudomonas fluorescens* wild type and mutants derived thereof at the same growth rate in nitrogen-limited chemostats. They used

a GC-MS method and a LC-MS/MS method to cover primary metabolites and the nucleotide pool, and found significant changes, especially in abundance of some nucleotides, in alginate producing mutants vs. the non-producing control strains. Boer et al (2010) analysed the metabolome of steady-state grown *Saccharomyces cerevisiae* cultures at five different growth rates and in each of five different limiting nutrients and reported a high sensitivity of metabolite concentrations to the limiting nutrient's identity. The same growth rate dependency goes for gene expression studies where it also has been shown that over 50% of genes are expressed in a growth rate-dependent manner (Regenberg et al., 2006).

Most microorganisms can be cultivated under relatively simple nutrient conditions compared to mammalian cells. This is advantageous in physiological experimentation as only one carbon source, one nitrogen source, etc. are needed, and thus simplifying the interpretation of the results. Additionally, the naturally labelled substrates can be replaced with heavy isotope labelled substrates which is beneficial for internal standard usage, and also for metabolic flux estimation based on the information of fractional enrichment of the ^{13}C-isotope in metabolites in the different pathways within the cell (Tang et al., 2009; Wittmann, 2007).

There is a strong focus on target quantitative analysis in Metabolomics as these results are not only easier to compare between studies in different laboratories but also because absolute concentrations are highly desirable for interpretation of the results (i.e. in correlation with enzyme kinetics). Quantitative MS Metabolomics rely on internal standards. Not all metabolites are readily available with heavy labelling (^{2}H (deuterium) or ^{13}C) or are available at too high a cost to be used in most laboratories on a routine basis. The traditional approach has been to use a limited set of heavy labelled metabolites as internal standards for whole metabolite classes (e.g. d3-alanine can be used for amino acids, d4-succinate for non-amino organic acids). This method has severe limitations as different metabolites in the same class can have quite different stabilities during quenching, extraction and analysis. The more advanced approach is to use extracts from ^{13}C labelled cultivations as internal standards. Even microbial extracts have been used for internal standardization during metabolome analysis of mammalian cells (Uehara et al., 2009; Wu et al., 2005). The latter is challenged by the fact that different cell types have different composition of the metabolite pool, implying that not all metabolites will be quantified by its respective heavy labelled internal standard. Bennett et al. (2009) used the opposite approach, using unlabelled compounds as internal standards while the cultivations were carried out using ^{13}C-glucose. They performed the cultivations on solid media in microwells, hence the cost of using labelled substrates was reasonable, and they were able to report absolute concentrations of over 100 metabolites using a HILIC LC-MS/MS method for positively charged metabolites and an ion pair reagent RP-C_{18} LC-MS/MS method for the negatively charged metabolites.

However, this is an expensive approach for true physiological studies carried out in chemostat which requires a minimum operating volume of 100 ml to ensure reproducible cultivation conditions and proper amount of sample volume.

NMR based Metabolomics

NMR-based metabolomics has become increasingly important during the last 15 years, much thanks to the pioneering work by Nicholson and co-workers (Lindon et al., 2001; Nicholson et al., 1999). The possibility of investigating the metabolic profile of biofluids, intact tissue or even live organisms is about to make NMR an invaluable tool within medicine in areas such as diagnostics, prognostics and treatment-response.In the area of microbiology, NMR-based metabolomics has not yet been used to a great extent. However, a number of recent examples indicate that this is about to change (Behrends et al., 2012; Sonkar et al., 2012; Ye et al., 2012). There has been a steady growth in technical improvement in NMR since its introduction, and its use has branched out towards a great number of applications. The areas of most relevance to metabolomics are liquid-state, solid-state (in particular high-resolution magic angle spinning (HR-MAS)) and localized *in vivo* NMR.

Non-invasive *in vivo* NMR spectroscopy is especially attractive in clinical medicine, since metabolite composition at specific locations can be measured without surgery. Together with different imaging techniques this can help medical doctors evaluate tumors and choosing the best course of action. Naturally, this imaging derived technique is of less importance in microbiology, since microbes may be analysed directly in an NMR tube. Many experimental designs have been used to study live microorganisms in NMR magnets. Unfortunately, they all share short-comings like low resolution and low sensitivity due to the inhomogeneity and low concentration of metabolites in the sample. Although not suitable for global metabolomics analysis, such *in vivo* studies are very useful for kinetic studies and flux determination by the aid of isotope labelled compounds.

Problems with inhomogeneous samples have been dealt with by using magic angle spinning (MAS). The sample is spun, usually at thousands of hertz, with the rotational axis oriented at 54.7^{o} relative to the static magnetic field. Spinning at this so-called magic angle removes the effect of interactions responsible for broadening of spectra of heterogeneous samples. High resolution MAS (HR-MAS) is now routinely performed on soft matter like organs and tissues and the resolution of these spectra is comparable to the resolution in solution samples. Even live cells may be studied with HR-MAS, as exemplified by Hanoulle and co-workers (2005) where a novel metabolite was detected in intact mycobacteria. In this particular study, the cell culture was concentrated to a paste by centrifugation in order to

increase metabolite concentrations. Others report that magic angle spinning may affect the viability of living samples (Aime et al., 2005; Weybright et al., 1998). Another concern is that MAS rotors may shatter at high spin rates, which is especially undesirable when working with pathogenic bacteria. This has caused incentive to develop methods using much lower spinning rates (for review see Wind and Hu (2006)). Although the idea of directly monitoring metabolic composition in a live cell culture is very attractive, the concentration of most metabolites is too low to be observed. Also, the dynamic nature of living cells might cause variation of the metabolic composition during the acquisition of spectra, especially for more time-demanding multidimensional experiments. This potential problem is avoided with liquid-NMR of extracts after proper quenching and extraction, as covered in earlier chapters. The extract may also be concentrated or freeze-dried and dissolved in another solvent, making it more amenable for NMR investigation. For a comprehensive review on the use of NMR in microbiology, see Grivet and Delort (2009).

Unlike MS based techniques, NMR spectra may be obtained with minimal sample preparation, reducing the chances of introducing bias. Being able to detect all magnetically active nuclei of small molecules in a sample, NMR is as close as we get to a universal detector. The main shortcoming of NMR is its sensitivity, and a large portion of the metabolome cannot be detected. Advances in instrumentation during the last decade has lowered the detection limits significantly, especially through probe-technology. The single largest jump in sensitivity increase came with the introduction of cryoprobes which reduce thermic noise by extreme cooling of the probe electronics. Another jump came with the microprobes, providing both better sensitivity and requiring a lower sample volume. Combined, a cryogenic microprobe offers a mass sensitivity gain of about a factor of 20 compared to conventional 5 mm probes (Brey et al., 2006).

The intrinsic low sensitivity of NMR stems from the low energy difference between magnetic spin states at the currently obtainable magnetic field strengths. Following the Boltzman distribution, a low energy difference equals a small difference in population of the two states, allowing detection of only a small fraction of the magnetic dipole moments present in the sample. Stronger magnets provide a bigger energy difference and allow a larger number of nuclei to be detected. Although field strengths have increased steadily since the introduction of NMR, we are now approaching a limit for which further development in magnet technology will come at a great cost. This has led to investment in other areas to obtain better signal-to-noise ratios.

In addition to probe-technology, ways of increasing the population difference between spin states have been addressed. Polarization transfer has been used for many years to increase sensitivity of insensitive nuclei, especially by bringing the more favourable spin population difference found in ^{1}H to more insensitive ^{13}C or ^{15}N. This is the foundation of

all indirect heteronuclear 2D NMR methods used today. More drastic enhancement of sensitivity may be achieved by coupling the magnetic spin to a source of much higher polarization. This has been done by chemically induced dynamic nuclear polarization (CIDNP) (Goez, 1997), para-hydrogen induced polarization (PHIP) (Duckett and Sleigh, 1999), magnetisation transfer from polarized noble gases (Cherubini and Bifone, 2003) and microwave driven dynamic nuclear polarization (DNP). In DNP, the increased polarization of the nuclei comes from unpaired electrons, improving sensitivity by a factor of ~10^2. Griffin and co-workers (2010) have made DNP a more general way of increasing sensitivity, useful in both liquid and solid-state NMR as well as imaging (for a recent overview see Griffin and Prisner (2010)).

Most atomic nuclei have magnetic properties, enabling them to be observed by NMR. Nuclei with spin = ½ are especially suited for NMR analysis and among these ^{1}H, or the proton, stands out as the most useful.Not only is hydrogen found in almost every organic molecule, usually in many different positions and at high natural abundance, but hydrogen is amongst the most sensitive nuclei for NMR use due to its high gyromagnetic ratio. For these reasons, the proton NMR experiment has been the obvious choice for NMR based metabolomics. The experiment is fast and, when an automated sample changer is used, a large number of spectra may be recorded in a short amount of time. Other isotopes with spin = ½ that are useful within metabolomics include ^{13}C, ^{15}N, ^{19}F and ^{31}P.

Most work in NMR-metabolomics has been done on different biofluids, requiring little sample handling. A buffer is usually added to the sample for pH adjustment, plus D_2O (usually 10%) to obtain a magnetic field lock signal. The remaining ~90% of water gives a strong signal that can be removed by different water suppression techniques. When working with extracts from biological samples, the presence of large molecules may introduce inhomogeneity in the sample tube. This can cause problems in liquid-NMR since large undissolved particles cause variation in the magnetic susceptibility of the sample, resulting in peak broadening. The problem can be solved by removing those particles by centrifugation and/or filtration. Large molecules, such as large proteins or biopolymers, give broad signals due to fast relaxation of the nuclei, even when properly dissolved. These signals may be filtered out using a Carr-Purcell-Meiboom-Gill CPMG experiment that may retain the signals from slower relaxing nuclei in small molecules (Meiboom and Gill, 1958). Alternatively, signals from large molecules may be separated from the remaining small molecules due to differences in the molecular diffusion properties. This is done with diffusion-ordered spectroscopy (DOSY) using pulse field gradients (Barjat et al., 1995). Even small differences in the diffusion constant can be distinguished, helping in the assignment of peaks in complex mixtures. Even though water usually is the solvent of choice in metabolomics, a large selection of deuterated NMR solvents are available in solution-NMR. Changing to an organic

NMR solvent will allow higher concentrations of less polar metabolites, possibly enabling their detection by NMR. However, chemical shifts are highly solvent-dependent, so for comparison with reported values or spectral libraries to be possible, the spectra must be recorded from the same solvent.

For ^{1}H NMR, overlap is inevitable for complex mixtures and quantification and identification of metabolites is often challenging. This is of less concern in metabolic fingerprinting, where accurate identification and quantification of metabolites is not the main goal and classification is performed by multivariate analysis of the non-interpreted spectroscopic data. For other metabolomic studies, including metabolic profiling and target analysis, different post-acquisition strategies have been employed to deal with spectral overlap, which will be covered in later chapters. Two dimensional (2D) NMR provides greater chances for signal dispersion, and also elucidates connectivities between spins. The arsenal of different 2D techniques for the identification of novel metabolites is invaluable, but they are usually applied after isolation and purification of the target molecule. A review of the different experiments available is outside the scope of this text, but some techniques have proven useful in metabolomics and require special attention.

A large number of different 2D NMR techniques are available and they are all based on series of 1D experiments varying only by an incremented delay in the pulsesequence. As a result, 2D experiments are time consuming and have traditionally been regarded unpractical for high throughput investigations. One exception is the 2D J-resolved (JRES) experiment, in which the coupling pattern of the proton signals are dispersed in a second dimension. Because of the greater dispersion, 2D JRES has been used directly in multivariate analysis with great success. Interestingly, a skyline projection of the tilted 2D JRES spectrum gives a seemingly broad band decoupled ^{1}H NMR spectrum, called pJRES. This may be interpreted as a normal ^{1}H NMR spectrum but without interfering coupling patterns, allowing easier quantification of metabolites. The use of 2D JRES and pJRES in metabolomics has recently been reviewed (Ludwig and Viant, 2010).

Most ^{1}H signals can be found within a spectral width of approximately 10 ppm, or 4000 Hz for a 400 MHz instrument. In contrast, signals from ^{13}C nuclei are usually found within 200 ppm, or 20000 Hz at the same field strength. This reduces chances for overlap and gives ^{13}C NMR a higher resolving power. Unfortunately, the carbon-13 isotope is only found in 1.1% abundance in nature and, additionally, is about four times less sensitive than ^{1}H. Direct observation of ^{13}C is for these reasons unpractical in metabolomics where detection of many weak signals is required. Carbon-13 labeling can partly solve this problem and has been extensively used in NMR to investigate metabolic pathways. However, isotopic labeling is expensive, making it less useful for high throughput metabolomics work. Fortunately, the informative ^{13}C chemical shifts may be observed indirectly through the ^{1}H

nuclei, using different 2D techniques. One of these is Heteronuclear Single Quantum Coherence (HSQC), which shows connectivity between bonded ^{1}H and ^{13}C nuclei. Not only does this give a very well-resolved spectrum, but the indirect observation provides significant improvement in sensitivity. In addition, recent advances in pulse sequence development, such as “single scan” NMR (Frydman et al., 2002) and band-selective optimized flip-angle short-transient (SOFAST) HMQC (Schanda and Brutscher, 2005), has drastically shortened experiment time for these kind of analyses, making them much more useful for high throughput investigations.

For metabolite identification, it can be beneficial to include a separation step prior to NMR analysis by using high-performance liquid chromatography (HPLC). Hyphenated techniques like HPLC-NMR-MS are now commercially available, where the eluting HPLC peak is split before detection (Lindon et al., 2000). The eluted peaks can also be trapped on a solid matrix using solid phase extraction (SPE) before analysis. This allows higher concentration and better chances for positive identification (for review see Sandvoss et al., (2005)).

References

Aime, S., Bruno, E., Cabella, C., Colombatto, S., Digilio, G., and Mainero, V. (2005). HR-MAS of cells: A "cellular water shift" due to water-protein interactions? Magnetic Resonance in Medicine *54*, 1547-1552.

Bajad, S.U., Lu, W.Y., Kimball, E.H., Yuan, J., Peterson, C., and Rabinowitz, J.D. (2006). Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. J Chromatogr A *1125*, 76-88.

Barjat, H., Morris, G.A., Smart, S., Swanson, A.G., and Williams, S.C.R. (1995). High-resolution diffusion-ordered 2D spectroscopy (HR-DOSY) - A new tool for the analysis of complex mixtures. J Magn Reson Ser B *108*, 170-172.

Behrends, V., Williams, K.J., Jenkins, V.A., Robertson, B.D., and Bundy, J.G. (2012). Free Glucosylglycerate Is a Novel Marker of Nitrogen Stress in Mycobacterium smegmatis. J Proteome Res *11*, 3888-3896.

Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., and Rabinowitz, J.D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat Chem Biol *5*, 593-599.

Boer, V.M., Crutchfield, C.A., Bradley, P.H., Botstein, D., and Rabinowitz, J.D. (2010). Growth-limiting Intracellular Metabolites in Yeast Growing under Diverse Nutrient Limitations. Mol Biol Cell *21*, 198-211.

Brey, W.W., Edison, A.S., Nast, R.E., Rocca, J.R., Saha, S., and Withers, R.S. (2006). Design, construction, and validation of a 1-mm triple-resonance high-temperature-superconducting probe for NMR. J Magn Reson *179*, 290-293.

Buscher, J.M., Czernik, D., Ewald, J.C., Sauer, U., and Zamboni, N. (2009). Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism. Anal Chem *81*, 2135-2143.

Cherubini, A., and Bifone, A. (2003). Hyperpolarised xenon in biology. Progress in Nuclear Magnetic Resonance Spectroscopy *42*, 1-30.

Coulier, L., Bas, R., Jespersen, S., Verheij, E., van der Werf, M.J., and Hankemeier, T. (2006). Simultaneous quantitative analysis of metabolites using ion-pair liquid chromatography - Electrospray ionization mass spectrometry. Anal Chem *78*, 6573-6582.
Cubbon, S., Antonio, C., Wilson, J., and Thomas-Oates, J. (2010). Metabolomic applications of HILIC-LC-MS. Mass Spectrom Rev *29*, 671-684.
Duckett, S.B., and Sleigh, C.J. (1999). Applications of the parahydrogen phenomenon: A chemical perspective. Progress in Nuclear Magnetic Resonance Spectroscopy *34*, 71-92.
Dwivedi, P., Wu, P., Klopsch, S.J., Puzon, G.J., Xun, L., and Hill, H.H. (2008). Metabolic profiling by ion mobility mass spectrometry (IMMS). Metabolomics *4*, 63-80.
Fiehn, O. (2008). Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. Trac-Trends Anal Chem *27*, 261-269.
Frydman, L., Scherf, T., and Lupulescu, A. (2002). The acquisition of multidimensional NMR spectra within a single scan. Proceedings of the National Academy of Sciences of the United States of America *99*, 15858-15862.
Garcia, D.E., Baidoo, E.E., Benke, P.I., Pingitore, F., Tang, Y.J., Villa, S., and Keasling, J.D. (2008). Separation and mass spectrometry in microbial metabolomics. Curr Opin Microbiol *11*, 233-239.
Goez, M. (1997). Photochemically Induced Dynamic Nuclear Polarization. Advances in Photochemistry *23*, 63-163.
Griffin, R.G., and Prisner, T.F. (2010). High field dynamic nuclear polarization-the renaissance. Physical Chemistry Chemical Physics *12*, 5737-5740.
Grivet, J.P., and Delort, A.M. (2009). NMR for microbiology: In vivo and in situ applications. Progress in Nuclear Magnetic Resonance Spectroscopy *54*, 1-53.
Halket, J.M., Waterman, D., Przyborowska, A.M., Patel, R.K.P., Fraser, P.D., and Bramley, P.M. (2005). Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS. J Exp Bot *56*, 219-243.
Hanoulle, X., Wieruszeski, J.M., Rousselot-Pailley, P., Landrieu, I., Baulard, A.R., and Lippens, G. (2005). Monitoring of the ethionamide pro-drug activation in mycobacteria by H-1 high resolution magic angle spinning NMR. Biochemical and Biophysical Research Communications *331*, 452-458.
Harada, K., Fukusaki, E., and Kobayashi, A. (2006). Pressure-assisted capillary electrophoresis mass spectrometry using combination of polarity reversion and electroosmotic flow for metabolomics anion analysis. J Biosci Bioeng *101*, 403-409.
Hinterwirth, H., Lammerhofer, M., Preinerstorfer, B., Gargano, A., Reischl, R., Bicker, W., Trapp, O., Brecker, L., and Lindner, W. (2010). Selectivity issues in targeted metabolomics: Separation of phosphorylated carbohydrate isomers by mixed-mode hydrophilic interaction/weak anion exchange chromatography. J Sep Sci *33*, 3273-3282.
Hui, J.P.M., Yang, J., Thorson, J.S., and Soo, E.C. (2007). Selective detection of sugar phosphates by capillary electrophoresis/mass spectrometry and its application to an engineered E-coli host. ChemBioChem *8*, 1180-1188.
Husek, P. (1998). Chloroformates in gas chromatography as general purpose derivatizing agents. J Chromatogr B *717*, 57-91.
Jandera, P. (2008). Stationary phases for hydrophilic interaction chromatography, their characterization and implementation into multidimensional chromatography concepts. J Sep Sci *31*, 1421-1437.
Josephs, J.L., and Sanders, M. (2004). Creation and comparison of MS/MS spectral libraries using quadrupole ion trap and triple-quadruople mass spectrometers. Rapid Commun Mass Spectrom *18*, 743-759.
Junot, C., Madalinski, G., Tabet, J.C., and Ezan, E. (2010). Fourier transform mass spectrometry for metabolome analysis. Analyst *135*, 2203-2219.
Kanani, H., Chrysanthopoulos, P.K., and Klapa, M.I. (2008). Standardizing GC-MS metabolomics. J Chromatogr B *871*, 191-201.
Kanani, H.H., and Klapa, M.I. (2007). Data correction strategy for metabolomics analysis using gas chromatography-mass spectrometry. Metab Eng *9*, 39-51.

Klampfl, C.W. (2004). Review coupling of capillary electrochromatography to mass spectrometry. J Chromatogr A *1044*, 131-144.
Klampfl, C.W. (2006). Recent advances in the application of capillary electrophoresis with mass spectrometric detection. Electrophoresis *27*, 3-34.
Klampfl, C.W., and Buchberger, W. (2010). Recent Advances in the Use of Capillary Electrophoresis Coupled to High-Resolution Mass Spectrometry for the Analysis of Small Molecules. Curr Anal Chem *6*, 118-125.
Koek, M.M., Muilwijk, B., and van der Werf, M.J. (2006). Microbial metabolomics with gas chromatography/mass spectrometry (vol 78, pg 1272, 2006). Anal Chem *78*, 3839-3839.
Kvitvang, H.F.N., Andreassen, T., Adam, T., Villas-Boas, S.G., and Bruheim, P. (2011). Highly Sensitive GC/MS/MS Method for Quantitation of Amino and Nonamino Organic Acids. Anal Chem *83*, 2705-2711.
Lien, S.K., Kvitvang, H.F.N., and Bruheim, P. (2012a). Utilization of a deuterated derivatization agent to synthesize internal standards for gas chromatography-tandem mass spectrometry quantification of silylated metabolites. J Chromatogr A *1247*, 118-124.
Lien, S.K., Sletta, H., Ellingsen, T.E., Valla, S., Correa, E., Goodacre, R., Vernstad, K., Borgos, S.E.F., and Bruheim, P. (2012b). Investigating alginate production and carbon utilization in *Pseudomonas fluorescens* SBW25 using mass spectrometry-based metabolic profiling. Metabolomics *In press*.
Lindon, J.C., Holmes, E., and Nicholson, J.K. (2001). Pattern recognition methods and applications in biomedical magnetic resonance. Progress in Nuclear Magnetic Resonance Spectroscopy *39*, 1-40.
Lindon, J.C., Nicholson, J.K., and Wilson, I.D. (2000). Directly coupled HPLC-NMR and HPLC-NMR-MS in pharmaceutical research and development. J Chromatogr B *748*, 233-258.
Ludwig, C., and Viant, M.R. (2010). Two-dimensional J-resolved NMR Spectroscopy: Review of a Key Methodology in the Metabolomics Toolbox. Phytochem Anal *21*, 22-32.
Luo, B., Groenke, K., Takors, R., Wandrey, C., and Oldiges, M. (2007). Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography-mass spectrometry. J Chromatogr A *1147*, 153-164.
Meiboom, S., and Gill, D. (1958). Modified spin-echo method for measuring nuclear relaxation times. Review of Scientific Instruments *29*, 688-691.
Mohler, R.E., Dombek, K.M., Hoggard, J.C., Young, E.T., and Synovec, R.E. (2006). Comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry analysis of metabolites in fermenting and respiring yeast cells. Anal Chem *78*, 2700-2709.
Mohler, R.E., Tu, B.P., Dombek, K.M., Hoggard, J.C., Young, E.T., and Synovec, R.E. (2008). Identification and evaluation of cycling yeast metabolites in two-dimensional comprehensive gas chromatography-time-of-flight-mass spectrometry data. J Chromatogr A *1186*, 401-411.
Myint, K.T., Uehara, T., Aoshima, K., and Oda, Y. (2009). Polar Anionic Metabolome Analysis by Nano-LC/MS with a Metal Chelating Agent. Anal Chem *81*, 7766-7772.
Nicholson, J.K., Lindon, J.C., and Holmes, E. (1999). 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica *29*, 1181-1189.
Ohashi, Y., Hirayama, A., Ishikawa, T., Nakamura, S., Shimizu, K., Ueno, Y., Tomita, M., and Soga, T. (2008). Depiction of metabolome changes in histidine-starved Escherichia coli by CE-TOFMS. Mol Biosyst *4*, 135-147.
Peterson, A.C., McAlister, G.C., Quarmby, S.T., Griep-Raming, J., and Coon, J.J. (2010). Development and Characterization of a GC-Enabled QLT-Orbitrap for High-Resolution and High-Mass Accuracy GC/MS. Anal Chem *82*, 8618-8628.
Regenberg, B., Grotkjaer, T., Winther, O., Fausboll, A., Akesson, M., Bro, C., Hansen, L.K., Brunak, S., and Nielsen, J. (2006). Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae. Genome Biol *7*.

Remane, D., Meyer, M.R., Wissenbach, D.K., and Maurer, H.H. (2010). Ion suppression and enhancement effects of co-eluting analytes in multi-analyte approaches: systematic investigation using ultra-high-performance liquid chromatography/mass spectrometry with atmosphericpressure chemical ionization or electrospray ionization. Rapid Commun Mass Spectrom *24*, 3103-3108.
Sandvoss, M., Bardsley, B., Beck, T.L., Lee-Smith, E., North, S.E., Moore, P.J., Edwards, A.J., and Smith, R.J. (2005). HPLC-SPE-NMR in pharmaceutical development: capabilities and applications. Magnetic Resonance in Chemistry *43*, 762-770.
Schanda, P., and Brutscher, B. (2005). Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. Journal of the American Chemical Society *127*, 8014-8015.
Schiesel, S., Lammerhofer, M., and Lindner, W. (2010). Multitarget quantitative metabolic profiling of hydrophilic metabolites in fermentation broths of beta-lactam antibiotics production by HILIC-ESI-MS/MS. Anal Bioanal Chem *396*, 1655-1679.
Seifar, R.M., Ras, C., van Dam, J.C., van Gulik, W.M., Heijnen, J.J., and van Winden, W.A. (2009). Simultaneous quantification of free nucleotides in complex biological samples using ion pair reversed phase liquid chromatography isotope dilution tandem mass spectrometry. Anal Biochem *388*, 213-219.
Soga, T., Ishikawa, T., Igarashi, S., Sugawara, K., Kakazu, Y., and Tomita, M. (2007). Analysis of nucleotides by pressure-assisted capillary electrophoresis-mass spectrometry using silanol mask technique. J Chromatogr A *1159*, 125-133.
Soga, T., Ohashi, Y., Ueno, Y., Naraoka, H., Tomita, M., and Nishioka, T. (2003). Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. J Proteome Res *2*, 488-494.
Sonkar, K., Purusottam, R.N., and Sinha, N. (2012). Metabonomic Study of Host-Phage Interaction by Nuclear Magnetic Resonance- and Statistical Total Correlation Spectroscopy-Based Analysis. Anal Chem *84*, 4063-4070.
Stahnke, H., Reemtsma, T., and Alder, L. (2009). Compensation of Matrix Effects by Postcolumn Infusion of a Monitor Substance in Multiresidue Analysis with LC-MS/MS. Anal Chem *81*, 2185-2192.
Strelkov, S., von Elstermann, M., and Schomburg, D. (2004). Comprehensive analysis of metabolites in Corynebacterium glutamicum by gas chromatography/mass spectrometry. Biol Chem *385*, 853-861.
Tang, Y.J., Martin, H.G., Myers, S., Rodriguez, S., Baidoo, E.E.K., and Keasling, J.D. (2009). Advances in analysis of microbial metabolic fluxes via C-13 isotope labeling. Mass Spectrom Rev *28*, 362-375.
Uehara, T., Yokoi, A., Aoshima, K., Tanaka, S., Kadowaki, T., Tanaka, M., and Oda, Y. (2009). Quantitative Phosphorus Metabolomics Using Nanoflow Liquid Chromatography-Tandem Mass Spectrometry and Culture-Derived Comprehensive Global Internal Standards. Anal Chem *81*, 3836-3842.
van der Werf, M.J., Overkamp, K.M., Muilwijk, B., Coulier, L., and Hankemeier, T. (2007). Microbial metabolomics: Toward a platform with full metabolome coverage. Anal Biochem *370*, 17-25.
van der Werf, M.J., Overkamp, K.M., Muilwijk, B., Koek, M.M., van der Vat, B.J.C., Jellema, R.H., Coulier, L., and Hankemeier, T. (2008). Comprehensive analysis of the metabolome of Pseudomonas putida S12 grown on different carbon sources. Mol Biosyst *4*, 315-327.
Villas-Boas, S.G., Delicado, D.G., Akesson, M., and Nielsen, J. (2003). Simultaneous analysis of amino and nonamino organic acids as methyl chloroformate derivatives using gas chromatography-mass spectrometry. Anal Biochem *322*, 134-138.
Villas-Boas, S.G., Smart, K.F., Sivakumaran, S., and Lane, G.A. (2011). Alkylation of Silylation for Analysis of Amino and Non-Amino Organic Acids by GC-MS? Metabolites *1*, 3-20.
Wamelink, M.M.C., Struys, E.A., Huck, J.H.J., Roos, B., van der Knaap, M.S., Jakobs, C., and Verhoeven, N.M. (2005). Quantification of sugar phosphate intermediates of the pentose

phosphate pathway by LC-MS/MS: application to two new inherited defects of metabolism. J Chromatogr B *823*, 18-25.
Want, E.J., Wilson, I.D., Gika, H., Theodoridis, G., Plumb, R.S., Shockcor, J., Holmes, E., and Nicholson, J.K. (2010). Global metabolic profiling procedures for urine using UPLC-MS. Nat Protoc *5*, 1005-1018.
Wentzel, A., Bruheim, P., Øverby, A., Jakobsen, Ø.M., Sletta, H., Omara, W.A.M., Hodgson, D.A., and Ellingsen, T.E. (2012a). Optimized submerged batch fermentation strategy for systems scale studies of metabolic switching in *Streptomyces coelicolor* A3(2). BMC Systems Biology *6*.
Wentzel, A., Sletta, H., Consortium., S., Ellingsen, T.E., and Bruheim, P. (2012b). Intracellular Metabolite Pool Changes in Response to Nutrient Depletion Induced Metabolic Switching in *Streptomyces coelicolor*. Metabolites *2*, 178-194.
Weybright, P., Millis, K., Campbell, N., Cory, D.G., and Singer, S. (1998). Gradient, high-resolution, magic angle spinning H-1 nuclear magnetic resonance spectroscopy of intact cells. Magnetic Resonance in Medicine *39*, 337-345.
Wilm, M. (2011). Principles of Electrospray Ionization. Molecular & Cellular Proteomics *10*.
Wind, R.A., and Hu, J.Z. (2006). In vivo and ex vivo high-resolution H-1 NMR in biological systems using low-speed magic angle spinning. Progress in Nuclear Magnetic Resonance Spectroscopy *49*, 207-259.
Wittmann, C. (2007). Fluxome analysis using GC-MS. Microb Cell Fact *6*.
Wu, L., Mashego, M.R., van Dam, J.C., Proell, A.M., Vinke, J.L., Ras, C., van Winden, W.A., van Gulik, W.M., and Heijnen, J.J. (2005). Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly C-13-labeled cell extracts as internal standards. Anal Biochem *336*, 164-171.
Yang, S., Sadilek, M., and Lidstrom, M.E. (2010). Streamlined pentafluorophenylpropyl column liquid chromatography-tandem quadrupole mass spectrometry and global C-13-labeled internal standards improve performance for quantitative metabolomics in bacteria. J Chromatogr A *1217*, 7401-7410.
Yang, W.C., Adamec, J., and Regnier, F.E. (2007). Enhancement of the LC/MS analysis of fatty acids through derivatization and stable isotope coding. Anal Chem *79*, 5150-5157.
Yang, W.C., Sedlak, M., Regnier, F.E., Mosier, N., Ho, N., and Adamec, J. (2008). Simultaneous Quantification of Metabolites Involved in Central Carbon and Energy Metabolism Using Reversed-Phase Liquid Chromatography-Mass Spectrometry and in Vitro C-13 Labeling. Anal Chem *80*, 9508-9516.
Ye, Y.F., Zhang, L.M., Hao, F.H., Zhang, J.T., Wang, Y.L., and Tang, H.R. (2012). Global Metabolomic Responses of Escherichia coli to Heat Stress. J Proteome Res *11*, 2559-2566.