LUND UNIVERSITY

Development of Protocols for Metabolomics in Biomedical Research using Chemometrics

Danielsson, Anders

2010

Development of Protocols for Metabolomics in Biomedical Research using Chemometrics

Anders Danielsson

LUND UNIVERSITY

Faculty of Engineering, LTH
ANALYTICAL CHEMISTRY
2010

Akademisk avhandling som för avläggande av teknologie doktorsexamen vid tekniska fakulteten, LTH, vid Lunds universitet, offentligen skall försvaras fredagen den 5 mars 2010, klockan 13.15 i Jubileumsaulan, ingång 59, Skånes Universitetssjukhus, Malmö.
Fakultetsopponent: Docent Henrik Antti, Kemiska institutionen, Umeå universitet, Umeå, Sverige.

Doctoral thesis which, by due permission of the Faculty of Engineering, LTH, at Lund University, will be publicly defended on Friday 5th of March, 2010, at 1.15 p.m. in Jubileumsaulan, entrance 59, Skåne University Hospital, Malmö, for the degree of Doctor of Philosophy in Engineering.
Faculty opponent: Associate Professor Henrik Antti, Department of chemistry, Umeå University, Umeå, Sweden.

Organization	Document name
LUND UNIVERSITY Department of Analytical Chemistry P.O. Box 124, SE-221 00 Lund SWEDEN	DOCTORAL DISSERTATION
	Date of issue
	Sponsoring organization

Author(s)
Anders Danielsson

Title and subtitle

Development of Protocols for Metabolomics in Biomedical Research using Chemometrics

Abstract

Metabolomics is a rapidly growing research field. It aims for quantification of all the metabolites in a biological sample such as plasma or cells and has, for example, been used to find disease biomarkers and to elucidate gene function. However, analysis of the complete metabolome puts high demands on the methods used. For instance, the methods should be unbiased to accurately depict the in vivo status in the cell. Furthermore, the methods must have very high resolution and sensitivity to allow detection of all metabolites. To approach these high goals, the protocols used in metabolomics need to be thoroughly optimised.

As the amount of information contained in the metabolome is immense, efficient methods are needed both to plan experiments and to convert the data to useful information. For this task, chemometrics is an ideal approach as it allows efficient experimental planning and multivariate data analysis. The experimental planning is sometimes referred to design of experiments and aims to systematically and simultaneously vary experimental factors in a structured manner. Hence, fewer experiments are generally needed to efficiently map how the system is affected by prevailing factors. The multivariate data analysis employs powerful projection and regression methods to find patterns in data, create system models and classify data.

In this thesis two thorough developments of metabolomics protocols and three metabolomics investigations, relevant to metabolic regulation in diabetes patients and insulin-producing cells, are presented. The design of experiments approach and multivariate data analysis were applied. The developed protocols were optimised and validated for the analysis of human blood plasma and adherent cell cultures, respectively, and included optimisation from the sample preparation to the analysis with gas chromatography/mass spectrometry. The first of the metabolomics studies aimed to find biomarkers reflecting metabolic regulation during an oral glucose tolerance test in humans to aid in the diagnosis of diabetes. The second study was performed on clonal β-cells and aimed to find metabolic regulation coupled to the amplifying pathway of insulin secretion. The last study aimed to identify metabolic dysregulation in clonal β-cells growing under lipotoxic and glucotoxic conditions, respectively. In all studies, metabolomics extended and deepened the understanding of metabolic regulation in cells and patients. As such, metabolomics will help to find explanations for metabolic diseases such as diabetes

Key words: Metabolomics, chemometrics, design of experiments, gas chromatography, mass spectrometry, multivariate analysis, metabolism, diabetes

Classification system and/or index termes (if any):

Supplementary bibliographical information:		Language English
ISSN and key title:		ISBN 978-91-7422-237-1
Recipient's notes	Number of pages 210	Price
	Security classification	

Development of Protocols for Metabolomics in Biomedical Research using Chemometrics

Anders Danielsson

LUND
UNIVERSITY

Faculty of Engineering, LTH
ANALYTICAL CHEMISTRY
2010

Doctoral thesis
Printed by Media-Tryck
Lund University
ISBN 978-91-7422-237-1

To my family

Table of contents

Acknowledgements

Well, now I'm sitting here, thinking that everything has come to an end with this thesis and that I have so many people to thank. In so many ways, this book is the work of many people and I just hope that I don't forget to express my gratitude to anyone. I will now change to Swedish.

Först vill jag tacka mina handledare, **Peter Spégel**, **Hindrik Mulder** och **Lo Gorton** som likt de tre musketörerna verkligen hjälpt till i alla väder. Det är mycket riktigt en spännande konstellation, precis som det beskrevs då jag började! **Peter!** Dig har jag att tacka för så mycket, såväl på jobbet som på fritiden! Du har en makalös förmåga att alltid kunna hjälpa till på alla sätt! Jag begriper inte hur du fixar allt samtidigt! **Hindrik!** Tack för all din support och entusiasm! Det har gjort att forskningen faktiskt blivit just sådär rolig som jag alltid hoppats på! **Lo!** Utan dig hade inget av det här blivit verklighet och för det är jag dig oerhört tacksam.

Ett jättestort tack till alla mina nuvarande och tidigare kollegor i Molekylär metabolism: **Vladimir, Thomas, Isabel, Cecilia, Siri, Jelena, Laila, Kalle, Disa, Ashkan, Hedvig** och **Karin**. Ni är fantastiska hela bunten och jag känner mig lyckligt lottad att ha kommit till er!

Ett stort tack till **Thomas Moritz** som varit till mycket stor hjälp som medförfattare och outtömlig kunskapsbank!

Ett stort tack går också till tidigare och nuvarande kollegor på Teknisk analytisk kemi, Analytisk kemi och Organisk kemi. Ett speciellt stort tack går till **Kerstin**, **Maggan, Clas, Olov** och **Maria!**

Till alla mina vänner, speciellt **Heffa**, **Danne**, **Pelle**, **Jodde**, **Harald**, **Lagerås, Peter K, Lars T** och **Lars N**, vill jag ge ett stort tack och den här boken, som jag vid något tillfälle kommer att kontrollera om ni har läst ☺

Till hela min underbara familj och släkt, speciellt **Mamma**, **pappa**, **systrar med familjer** och **Duvan**: Liksom så mycket annat, hade denna bok inte blivit av om det inte vore för min underbara familj som i alla väder givit mig det stöd som behövts och mycket mer därtill. Det började med ett hemmabyggt tjuvlarm på toa och slutade med en avhandling i analytisk kemi! Jag älskar er!

Till familjen **Nielsen** vill jag också rikta ett stort tack och vill också passa på att önska mig ett skånskt-svenskt slangordslexikon och ett svenskt-svenskt/tyskt. ☺

Till min alldeles underbara **Tettan**! Nu äntligen är det här klart och nu slipper du kanske höra med om ”trippel-quaddar”, ”Fourier transform jon-cyklotron resonans mass spektrometrar”, ”TOFFAR”, principal komponenter och metaboliter! Tack för all din hjälp och fantastiska stöd! Jag älskar dig snäckan!

List of papers

The content of this thesis is based on the following papers.

I. Development of a GC/MS based Metabolomics Protocol by means of Statistical Experimental Design
Anders P. H. Danielsson, Thomas Moritz, Hindrik Mulder, Peter Spégel
Submitted

II. Development and Optimisation of a Metabolomics Method for Analysis of Adherent Cell Cultures
Anders P.H. Danielsson, Thomas Moritz, Hindrik Mulder, and Peter Spégel
Submitted

III. Metabolomic Analysis of a Human Oral Glucose Tolerance Test reveals Fatty acids as Reliable Indicators of Regulated Metabolism
Peter Spégel, Anders P.H. Danielsson, Karl Bacos, Cecilia L.F. Nagorny, Thomas Moritz, Hindrik Mulder, and Karin Filipsson
*Metabolomics, **2009, In press** (available online)*

IV. Unraveling the Tight Coupling of Metabolism and Insulin Secretion in Clonal β-cells (INS-1 832/13)
P. Spégel, A.P.H. Danielsson, V.V. Sharoyko, C.L.F. Nagorny, G. Sharp, S. Straub, and H. Mulder
Manuscript

V. Investigation of Lipotoxicity-induced Metabolic Alterations in Clonal β-cells (INS-1 832/13)
Anders P.H. Danielsson, Cecilia L.F. Nagorny, Siri Malmgren, Hindrik Mulder, and Peter Spégel
Manuscript

Author contributions to the papers

Paper I

The author contributed to a major part in planning of the experiments, performed all the experiments and a major part of the data analysis, and wrote most of the manuscript.

Paper II

The author contributed substantially to the planning of the experiments, performed all the experiments and a major part of the data analysis, and wrote a major part of the manuscript.

Paper III

The author contributed substantially to the data analysis and assisted in writing the manuscript.

Paper IV

The author contributed to a major part in the sample preparation, aided the first author in analysing the samples and assisted in interpretation of the data. The author wrote a significant part of the manuscript.

Paper V

The author performed a major part of the experiments and all of the data analysis, and wrote a substantial part of the manuscript.

Acronyms and abbreviations

3D-SUS	Three-dimensional shared and unique structure
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
APCI	Atmospheric pressure chemical ionisation
CCC	Central composite circumscribed
CCF	Central composite face-centred
CE	Capillary electrophoresis
CE/MS	Capillary electrophoresis/mass spectrometry
CI	Chemical ionisation
CoA	Coenzyme A
COST	Change one factor separately at a time
CPT1	Carnitine palmitoyltransferase 1
DAG	Diacylglycerol
DIMS	Direct infusion mass spectrometry
DOE	Design of experiments
EI	Electron impact
ESI	Electrospray ionisation
ETC	Electron transport chain
FFA	Free fatty acid
FTMS	Fourier transform mass spectrometry
FT-ICR MS	Fourier transform ion cyclotron resonance mass spectrometry
GABA	γ-aminobutyric acid
GC	Gas chromatography
GC/MS	Gas chromatography/mass spectrometry
GLUT	Glucose transporter
GSIS	Glucose-stimulated insulin secretion
HILIC	Hydrophilic interaction liquid chromatography

HMCR	Hierarchical multicurve resolution
i.d.	Inner diameter
IT	Ion trap
LC	Liquid chromatography
LC/MS	Liquid chromatography/Mass spectrometry
MALDI	Matrix-assisted laser desorption ionisation
Malonyl-CoA	Malonyl coenzym-A
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
MSTFA	*N*-methyl-*N*-trimethylsilyltrifluoroacetamide
MTBSTFA	*N*-*tert*-butyldimethylsilyl-*N*-methyltrifluoroacetamide
MVDA	Multivariate data analysis
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NIPALS	Nonlinear iterative partial least squares
NMR	Nuclear magnetic resonance
OGTT	Oral glucose tolerance test
OPLS	Orthogonal projections to latent structures
PCA	Principal component analysis
PLS	Projections to latent structures
Q-TOF	Quadrupole time of flight
RI	Retention index
RIA	Radioimmunoassay
ROS	Reactive oxygen species
T2D	Type 2 diabetes
TAG	Triacylglycerol
TBDMS	*tert*-butyldimethylsilyl
TCA	Tricarboxylic acid cycle
TIC	Total ion count
TMS	Trimethylsilyl
TOF	Time of flight
UPLC	Ultra performance liquid chromatography

Abstract

Metabolomics is a rapidly growing research field. It aims for quantification of all the metabolites in a biological sample such as plasma, saliva, cerebrospinal fluid or cells. Because the metabolite levels in a biological sample are the end result of the regulatory processes in cells, metabolomics is a very powerful approach for characterisation of phenotypes. Metabolomics has been used to find disease biomarkers, investigate influences of heavy metals on the metabolism and to elucidate gene function. However, analysis of the complete metabolome puts high demands on the methods used. For instance, the methods should be unbiased to accurately depict the *in vivo* status in the cell. Furthermore, the methods must have very high resolution and sensitivity to allow detection of all metabolites. To approach these high goals, the protocols used in metabolomics need to be thoroughly optimised.

The amount of information contained in the metabolome is immense. Consequently, the data set collected from a metabolomics study is very large. To extract the relevant information from such large sets of data, efficient methods are needed both to plan experiments and to convert the data to useful information. For this task, chemometrics is an ideal approach as it allows efficient experimental planning and multivariate data analysis. The experimental planning is sometimes referred to as statistical experimental design or design of experiments. It aims to systematically and simultaneously vary experimental factors in a structured manner. Hence, fewer experiments are generally needed to efficiently map how the system is affected by prevailing factors. The multivariate data analysis employs powerful projection and regression methods to find patterns in data, create system models and classify data. Hence, chemometrics provides a framework for efficient experimental design and an efficient approach for information retrieval.

In this thesis two thorough developments of metabolomics protocols and three metabolomics investigations, relevant to metabolic regulation in diabetes patients and insulin-producing cells, are presented. The design of experiments approach and multivariate data analysis were applied. The developed protocols were optimised and validated for the analysis of human blood plasma and adherent cell cultures, respectively, and included optimisation from the sample preparation to the analysis with gas chromatography/mass spectrometry. The first of the metabolomics studies aimed to find biomarkers reflecting metabolic regulation

during an oral glucose tolerance test in humans to aid in the diagnosis of diabetes. The second study was performed on clonal β-cells and aimed to find metabolic regulation coupled to the amplifying pathway of insulin secretion. The last study aimed to identify metabolic dysregulation in clonal β-cells growing under lipotoxic and glucotoxic conditions, respectively. In all studies, metabolomics extended and deepened the understanding of metabolic regulation in cells and patients. As such, metabolomics will help to find explanations for metabolic diseases such as diabetes

Populärvetenskaplig sammanfattning

Under senare år har forskningen kring sjukdomar, orsakade av rubbningar i ämnesomsättningen, blivit allt viktigare. Till exempel är den vanligaste dödsorsaken i västvärlden hjärt-kärlsjukdomar vars bakomliggande orsak, atheroskleros, beror på en rubbning i lipidmetabolismen. En annan stor folksjukdom är typ 2 diabetes som har undersökts i denna avhandling. Idag lider över 200 miljoner människor av diabetes och om 20 år förutspås det vara nära 400 miljoner. Diabetes kännetecknas av att kroppen inte klarar att hålla en lagom nivå av glukos, d v s socker, i blodet. Detta beror på att frisättningen av hormonet insulin från de s.k. β-cellerna är störd och att känsligheten för insulin är nedsatt. Resultatet blir en generell rubbning av metabolismen. Rubbningen yttrar sig också bl. a. genom onaturligt höga halter av blodfetter. Vad som orsakar störningen och den minskade insulinkänsligheten är ännu inte klarlagt, vilket till stor del beror på den stora komplexiteten hos metabolismen.

Metabolismen sker i ett mycket omfattande och komplext nätverk av kemiska reaktioner som har till uppgift att omvandla det vi äter och det kroppen innehåller till energi och byggstenar. De kemiska reaktionerna sker framför allt i kroppens celler mellan små molekyler, de s.k. metaboliterna som har mycket varierade egenskaper. Metaboliterna, som kollektivt kallas metabolomet, uppskattas vara ca 3000 i en människa och förekommer i mycket varierande halter.

Delvis på grund av den stora komplexiteten hos metabolomet har man traditionellt begränsat sig till att studera en eller några få metaboliter åt gången. Detta tillvägagångssätt har ett flertal nackdelar. Till exempel är det sannolikt att förändringar som observeras i en begränsad del av metabolomet även bidrar till förändringar i andra delar som man inte undersöker. En annan nackdel är att man på förhand måste ha hypotes om vad som ska hända för att veta vilken eller vilka metaboliter man vill studera.

På senare år har man, tack vare en massiv utveckling av både instrument och analysmetoder, börjat försöka undersöka hela metabolomet samtidigt. I regel kan man undersöka ca 200 metaboliter åt gången. Detta kallas metabolomik och inom detta område används traditionella analytiska tekniker såsom masspektrometri och kromatografi i kombination med multivariat data analys som är en del av

kemometri. Multivariat data analys är väl anpassad för att kunna utvärdera den enorma och komplexa mängd data som man normalt får fram i en metabolomikstudie. Metoderna bygger på att effektivt extrahera och visualisera den information som är signifikant.

En annan del av kemometri bygger på att med matematiska modeller planera sina försök, s.k. statistisk experimentell design. Detta är speciellt lämpligt då man med få experiment vill undersöka vilka faktorer som påverkar ett system. Då man kombinerar statistisk experimentell design med multivariat data analys har man ett effektivt verktyg med vilket man också kan optimera en process.

I två av studierna som ligger till grund för denna avhandling har kemometriska metoder använts för att optimera arbetsprotokoll för metabolomikstudier på human blodplasma respektive odlade β-celler med hjälp av gaskromatografi i kombination med masspektrometri. I tre andra studier har metabolomik använts för att undersöka vad som händer i metabolomet hos människor och odlade β-celler då glukos förbränns. I en av dessa studier har även påverkan av ett överflöd av fettsyror och glukos på odlade β-celler undersökts

Introduction

A key aspect in understanding a biological system is to study the metabolism as the concentrations of the metabolites in a cell are sensitive to changes in both the surrounding environment and also to genetic changes. Hence, the metabolites can be viewed as the end products of the regulatory processes in a cell (1). The study of the complete set of metabolites in a cell, i.e. the metabolome is called metabolomics and aims for an unbiased quantification of all the components in the metabolome. This will provide a snapshot of the *in vivo* metabolism. The levels of the metabolites can then be used to elucidate the effects of a perturbation of the regulatory processes in a cell. In addition, the common use of powerful multivariate data analysis (MVDA) (2-3) adds the ability of linking the perturbation to significant changes in metabolite levels, i.e. creating a model of the investigated metabolism. Hence, it is a screening approach well suited for finding metabolic patterns and biomarkers. The primary benefit of this global approach, compared to traditional approaches where a single or a few metabolites are analysed, is that there need not be any hypothesis *a priori*; instead metabolomics can be hypothesis-generating. This is highly useful when studying metabolism which basically is a complex network of coupled equilibria between several thousand metabolites.

There are many applications for metabolomics are many. To mention a few, it has been used to elucidate effects of drugs on metabolism (4-6), the influence of heavy metals on the metabolism (7) and finding biomarkers for disease (8-9). Of special importance today, in the post-genomic era, is to use metabolomics to link the genotype to the phenotype (1, 10-11). The connection between the genome and the other "omes" can be described in the continuously expanding "omics-cascade" (Fig. 1). A problem with genomics, transcriptomics and proteomics is that although several genomes have been completely mapped and their expression levels can be measured, many genes and the proteins they encode have no known function. Furthermore, the functions of many proteins are assigned solely on homology to other proteins, which is not necessarily correct (1). As metabolomics is the "ome" closest to the phenotype it may help in elucidating the function of a gene in combination with the other disciplines and functional studies.

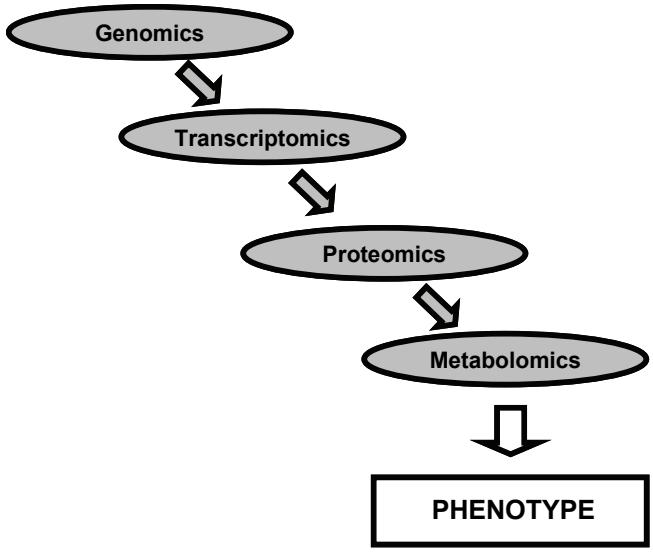

Figure 1 The "omics cascade" describing the connection between genomics, transcriptomics, proteomics, metabolomics and the phenotype.

Investigating effects on the metabolome is a highly complex task, putting extreme demands on the methods used. The size of the metabolome is dependent on the organism and has been estimated to be 200 000 metabolites in plants (1) and at least 2000 metabolites (10) in the human metabolome. Metabolites also have widely different physical and chemical properties and are present at very different concentrations. The diversity of the metabolome is also greater than that of the proteome and the genome, which both have a comparatively low number of building blocks (1). Hence, the methods used should have very high resolution in order to detect the vast amount of metabolites. Also, the method should be unbiased in order to give an accurate snapshot of the *in vivo* metabolism. If biases are introduced in the investigation, the metabolite levels will not reflect the actual situation in the sample. The vast differences in abundance, from pmol to mmol, for a single metabolite as well as between metabolites, require wide linear ranges and low detection limits. This enables the quantification of small variations also of metabolites of low abundance, which is very important as it is rarely the actual metabolite level that is important for the biological function, but rather the variation.

Due to the complexity of the metabolome, no method is currently able to measure all metabolites simultaneously. This warrants comprehensive method optimisation from the sample pre-treatment to the detection. The majority of studies employ methods based on chromatographic separation with mass spectrometric (MS) detection or nuclear magnetic resonance (NMR) spectroscopy. Out of these

methods gas chromatography (GC) with MS detection (GC/MS) is the most common and is the method used in the present work. Common samples include plasma (12), urine (13), cerebrospinal fluid (14) and cells. Among the cell types are human hepatocytes (15), yeast (16-19) and *E.Coli (20-21)*. Regardless of sample type, their pre-treatment needs to be unbiased, robust and reproducible. Therefore, development of pre-treatment protocols is essential for the accuracy of metabolomics investigations. Several pre-treatment protocols have been developed for a wide range of biofluids and cells (12, 18, 22-25) and several studies have focussed on the comparison of protocols (16, 19, 21).

Metabolomics investigations frequently generate vast amounts of highly complex data due to the metabolome complexity and the concurrent influence of several experimental variables. This requires efficient tools for experimental planning and for data analysis in order to first systematically induce perturbations and subsequently extract relevant data. For this task, design of experiments (DOE) (26-27) and MVDA (2-3) is a well suited approach. This approach allows simultaneous and systematic variation of several experimental factors, i.e. changes in glucose concentration, temperature and oxygen concentration, and uses powerful projection and regression methods for information retrieval.

Objectives

The objectives were to develop comprehensive metabolomics protocols and to elucidate metabolic regulation related to stimulus-secretion. More specifically the aims were to

- develop a metabolomics protocol for blood plasma.
- develop a metabolomics protocol for adherent cell cultures for the study of clonal β-cells.
- characterise metabolic regulation after glucose stimulation *in vivo* in humans and *in vitro* in clonal β-cells
- characterise metabolic dysregulation under lipotoxic, glucotoxic and glucolipotoxic conditions in clonal β-cells.

Metabolism - the target system for metabolomics studies

Metabolism comprises a very large interconnected set of chemical reactions divided into metabolic pathways that maintain the cellular processes (28). This is accomplished by breaking and forming chemical bonds, thereby consuming or releasing energy as well as creating new molecules. The pathways are catalysed by enzymes and controlled by the current needs of the cell, mediated by chemical signals. The energy is derived from carbohydrates, lipids and proteins, which are broken down to provide energy and building blocks for the cell. The molecular products and intermediates that participate in these processes, the metabolites, are collectively known as the metabolome. The metabolites are highly diverse with regards to their chemical properties and functionalities. This is in part due to the many different functions for which the metabolites are responsible. For instance, the energy in the body is mainly stored as triglycerides which are well suited for this purpose as these are energy-rich and do not take up as much space as carbohydrates. Another reason is that the metabolic processes are carried out in different cellular compartments characterised by differences in the chemical environment (pH, hydrophobicity, etc.).

In this section a brief overview of the metabolic phenomena studied in connection to diabetes is presented.

Glucose homeostasis and glucose-stimulated insulin secretion

An essential component of whole body metabolism is the metabolic regulation exerted by insulin. Glucose is the prime stimulus of insulin secretion, and it is frequently referred to as glucose-stimulated insulin secretion (GSIS), and occurs in the pancreatic β-cells. The β-cells, reside in the islets of Langerhans, which are present in the pancreas. They act as blood glucose sensors and regulators, constantly monitoring the glucose level in the blood and releasing insulin to maintain it at a physiological level (~5 mM). This is particularly important after a

meal, and serves to make glucose available for metabolism in peripheral tissues. Skeletal muscle is the quantitatively most important tissue in the body for insulin-stimulated glucose uptake followed by the adipose tissue. Although glucose uptake in the liver is not insulin-dependent, this organ plays a very important role in glucose homeostasis. It extracts glucose from the blood when the sugar is abundant after a meal and converts it into glycogen, the storage form of glucose. When glucose levels in the blood drop, glycogen is broken down to glucose, which is released from the liver. Hereby, blood glucose levels are maintained, which is very important for the brain, an organ dependent solely on glucose for its energy production. Whether the liver stores or releases glucose is determined by insulin released from the pancreatic β-cells.

GSIS is a highly complex process which is not fully understood (**Papers III-V)**. The process is controlled by two mechanisms; the triggering pathway and the amplifying pathway (29). The triggering pathway (Fig. 2) is initiated by the uptake of glucose into the β-cell, which is mediated by glucose transporters, GLUT1 and GLUT2, present in the plasma membrane. Following uptake, glucose is metabolised through glycolysis, producing adenosine triphosphate (ATP), reduced nicotinamide adenine dinucleotide (NADH) and pyruvate. Importantly, the first step in β-cell glycolysis is catalysed by glucokinase. This enzyme has a high K_m for glucose, ensuring that the activity of the enzyme is regulated in the physiological range of glucose concentrations. Glucokinase thus serves as a glucosensor in the β-cell. The glycolytic end product pyruvate is further metabolised in the mitochondria in the tricarboxylic acid (TCA) cycle. The resultant production of NADH drives the electron transport chain (ETC), producing more ATP which is transported to the cytosol. While the ATP level is increased the level of adenosine diphosphate (ADP) is decreased, causing an increase in the ATP/ADP-ratio. This closes the ATP-sensitive potassium channels (K_{ATP}-channels) which normally transport K^+ out of the cell to maintain a negative membrane potential. Closure of the K_{ATP}-channels causes the membrane potential (ψ) to rise until a threshold value is reached and the whole membrane is depolarised. The depolarisation opens voltage-dependent Ca^{2+} channels and the cytosolic Ca^{2+} concentration is therefore rapidly increased. This is the triggering signal for insulin secretion from intracellular stores of insulin granules.

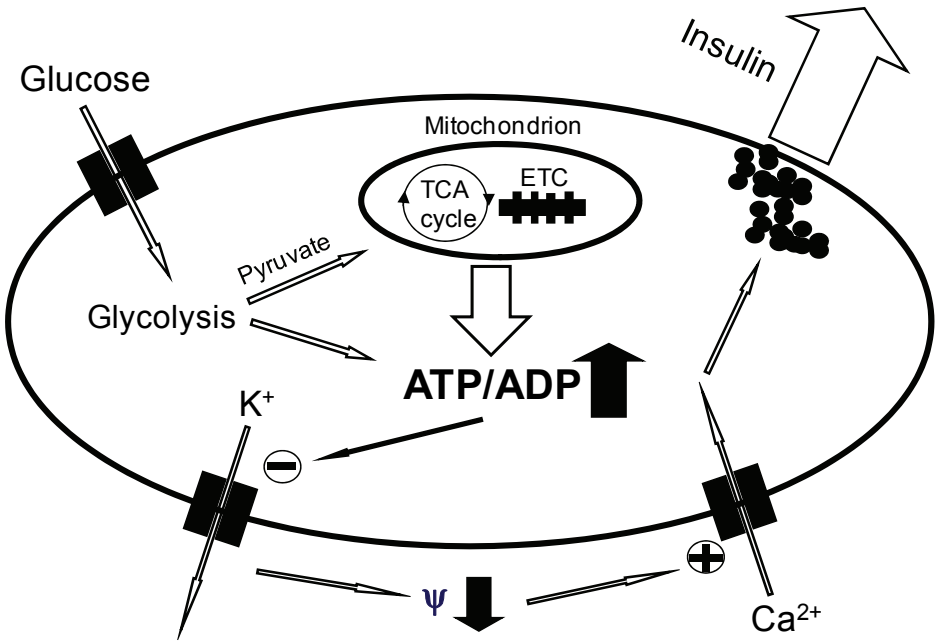

Figure 2 A schematic, describing the triggering pathway of GSIS.

In contrast to the triggering pathway, the amplifying pathway is not well understood (**Paper IV)**. It further augments and enhances the secretion of insulin but it is not known which signals that are involved in this process. Nevertheless, several metabolic intermediates have been proposed as signals stimulating this process (30). These include glutamate, malonyl coenzym-A (malonyl-CoA) and the reduced form of nicotinamide adenine dinucleotide phosphate (NADPH).

Lipotoxicity and glucotoxicity

It has become clear that β-cell function is disturbed in Type 2 diabetes (T2D) (31). A strong line of research in the field has addressed how and why excessive lipids could contribute to this process (32). Lipotoxicity is characterised by an

accumulation of lipids in tissues other than adipose tissue, causing a toxic environment for the cells. The cause of lipotoxicity are not known and it has been suggested that it is due to a malfunction in the storage function of the adipose tissue causing free fatty acids (FFA) accumulation in other tissues (33). Affected tissues are primarily the liver, the pancreatic islets and muscle. The subsequent toxic environment in these tissues adversely affects cellular functions. Lipotoxicity has been connected to many diseases such as the metabolic syndrome, cardiovascular disease and T2D. Of main interest in this thesis is T2D and hence the effects on the pancreatic β-cell metabolism and insulin secretion (**Paper V)**.

In T2D, elevated plasma levels of FFAs are regularly found and it has been shown in rodent islets that exposure to elevated FFA levels cause a decrease in GSIS (34). Similarly, in experiments on human islets, it has also been shown that elevated levels of FFAs decrease the insulin secretion, reduce insulin sensitivity and may induce apoptosis in the β-cell (35) However, the effects on the β-cells are time-dependent. While an increase in FFA levels during a brief exposure causes an increase in insulin secretion, a longer exposure causes impaired insulin secretion and possibly apoptosis (35). A key concept in many hypotheses regarding lipotoxicity is lipid metabolism and the import of FFA into the mitochondria. This transport is tightly controlled by the activity of carnitine-palmitoyl tranferase 1 (CPT1), which in turn is dependent on the level of malonyl-CoA, which inhibits the action of CPT1. The inhibition results in increased levels of triacylglycerols (TAGs), diacylglycerols (DAGs) and ceramides which are synthesised in the cytosol. Particularly the latter have been implicated in apoptosis (36).

Closely associated to lipotoxicity in T2D is the occurrence of glucotoxicity where the elevated glucose levels have been susggested to induce apoptosis of β-cells and metabolic dysfunction. The toxic effect is here assumed to be mediated by reactive oxygen species (ROS). Lipotoxicity is often observed together with glucotoxicity in T2D, suggesting a coupled action – glucolipotoxicity (37). However, it has not been established whether there is an interaction of glucose and lipids in these metabolically “toxic” situations, and to what extent the adverse effects on the β-cells are critical. Neverthless, due to the decreased insulin secretion, both the glucose level and the levels of FFAs are normally elevated in T2D. The cause and effect sequence of glucose and lipid abnormalities and cellular dysfunction in T2D has also not been established. Clearly, many questions remain and metabolomics analysis may provide some novel insight.

Methods for metabolomics

The complexity of metabolomics samples puts high demands on the methods used to study them. In fact, no single method exist with the ability to measure all metabolites simultaneously (38). Instead, combinations of several methods have to be used to cover a larger portion of the metabolome. Not only do the methods themselves differ in their characteristics and performance, but also in the type of sample pre-treatment required and which type of metabolites that can be measured. In this section, the most common methods for metabolomics are briefly described and compared. The method of choice in this thesis, gas chromatography/mass spectrometry (GC/MS) will be described in more detail (**Papers I-V)**. For a more detailed description of the below discussed methods, the reader is referred to the review by Büscher (39).

Gas chromatography/mass spectrometry

GC combined with MS is the most common approach for metabolomics and metabolic profiling analyses and has been applied to a wide array of investigations on the metabolome (12, 21-22, 40-41). Advantages with this method include a very high separation efficiency, robustness and high throughput (39). The main drawback is that it can only be used to separate volatile analytes. As most metabolites are non-volatile this adds the requirement to derivatise the metabolites (see below). However, all metabolites can not be made volatile, especially larger metabolites which form strong intermolecular interactions. Another drawback is that thermolabile metabolites may be degraded. Also this can be counteracted somewhat by the use of a suitable derivatisation method.

The retention is based on partitioning between the mobile phase, consisting of a carrier gas and a stationary phase consisting of a liquid residing on the inside wall of the capillary. The stationary phase may be of many different types but is in metabolomics approaches usually a non-polar phase with high inertness. The partitioning is in turn governed by the vapour pressure (p^0) of the analytes and their interaction with the stationary phase, described by the activity factor (γ). Their combination is often called the effective volatility. This can be described by

the retention factor (*k'*) which represents the molar fraction of an analyte in each phase at equilibrium and also has a relationship to the retention time (equation 1).

$$k' = \frac{RT\rho_s}{\gamma p^0 M_s} \cdot \frac{V_s}{V_g} = \frac{t_r - t_m}{t_m} \quad (1)$$

Here, R is the gas constant, T is the temperature, M_s is the molar mass of the stationary phase, ρ_s is the density of the stationary phase and V_s and V_g is the volume of stationary phase and mobile phase, respectively. Finally, t_r is the retention time for the analyte and t_m is the column void time. From this expression it is obvious that the retention time decreases with increasing effective volatility. Hence, volatile analytes (high p^0) with weak interactions with the stationary phase (γ>1) will be eluted first. The retention can be controlled by changing the temperature which both directly influences the *k'* (equation 1) and also influences the vapour pressure according to the Clausius-Clapeyron equation (equation 2). It should be mentioned that the temperature also influences the activity factor, but only to a minor extent and it can therefore be regarded as constant for an analyte in the temperature range used in GC.

$$\ln p^0 = \frac{-\Delta H_{vap}}{RT} + C \quad (2)$$

Here, ΔH_{vap} is the vaporisation enthalpy and C is a constant. Hence, an increased temperature increases the vapour pressure and decreases the retention and also dominates over the direct effect on the retention factor (equation 1) and the effect on the activity factor. The strong dependence of the retention time on the temperature is utilised when applying a temperature gradient. The gradient increases the temperature, thereby reducing the analysis time. However, the increase in temperature has to be thoroughly optimised as a too steep gradient may reduce the separation efficiency (**Papers I** and **II**). Identification of analytes can be performed based on their retention time but it is most commonly performed by the analyte retention index (42) (RI) which relates the retention time of the analyte to a homologous series of e.g. alkanes. RI, allows for comparison of analyte retention on several gas chromatographs. However, identification based on retention is not sufficient as many analytes coelute.

The gas chromatograph is usually coupled to a time-of-flight (TOF) mass analyser, via an electron impact (EI) ion source. Other ion sources can also be used, for instance chemical ionisation (CI). In the EI ion source, the analytes are fragmented into ions by a stream of electrons with a kinetic energy of usually 70 eV. After

ionisation, the ionic fragments are accelerated in to the mass analyser where the ionised analytes are separated according to their TOF in the so called flight tube which is often equipped with a reflectron to compensate for the energy distribution obtained in the acceleration step. As with any MS method, the separation depends on the ionic fragments mass-to-charge ratio (*m/*z). Identification of the analytes can now be performed based both on RI and by matching their mass spectra to mass spectra in databases. The spectrum of an analyte is often highly specific due to the robust fragmentation pattern.

Alternative methods for metabolomics

Mass spectrometry-based methods

There are several other methods used for metabolomics and similar methodologies out of which the most commonly used are chromatography methods coupled to MS (39). Among these liquid chromatography-mass spectrometry (LC/MS), is the most common (25). Like GC, the separation is based on partitioning, but here between a liquid flow and a stationary phase often covalently bound to particles. Hence, the metabolites do not have to be volatile and therefore do not require derivatisation prior to analysis. Like in GC, the stationary phases may be of many different types. This will determine the selectivity. Common approaches are hydrophilic interaction liquid chromatography, (HILIC) (43), and ion-pairing chromatography which are both suitable for the predominantly polar metabolome. The primary advantages with LC/MS over GC/MS is the ability to analyse the pure metabolites, the ability to analyse large polar metabolites and that it can handle thermolabile metabolites (1). These advantages result in an improved coverage of the metabolome (39). Disadvantages with LC/MS compared to GC/MS are primarily its lower reproducibility, lower robustness, longer analysis times and somewhat lower separation efficiency (39). The reproducibility problem originates from that buffers are normally freshly prepared before each study, which may generate increased day-to-day variation. The lower separation efficiency is primarily due to that the separation is performed in a packed column with a flowing liquid as mobile phase. The longer analysis time is due to the limited flow rates that can be used with the electrospray ionisation (ESI) ion sources and that the separation efficiency is adversely affected at higher flow rates. To counteract theses disadvantages, ultra performance liquid chromatography (UPLC) has been developed. It has already been applied in numerous studies (4, 15, 44-46), and uses smaller particles and narrower columns to primarily increase the separation efficiency. This has several additional advantages, including shorter

analysis time and lower solvent consumption. In addition, the flow rate limitation set by the ESI ion source can be reduced by the use of atmospheric pressure chemical ionisation (APCI), which tolerates increased flow rates.

Capillary electrophoresis/mass spectrometry, (CE/MS) has also been used for metabolomics analysis (47-48). Separation occurs in a capillary over which a high voltage is applied. The basis for separation is differences in electrophoretic mobility which relates to the size, shape and charge of the analytes. Advantages with this method are high separation efficiencies and a coverage of the metabolome comparable to LC/MS (39). The primary disadvantage is that neutral analytes or analytes with the same charge-to-size-ratio can not be separated unless they are highly different in shape. In addition, the pH is critical for the separation as it controls the charge for the analytes and also the magnitude of the endoosmotic flow. Further, CE/MS produces less robust retention times than GC and LC

Recently, so called comprehensive two-dimensional gas chromatography, GCxGC, has emerged as a valuable tool for metabolomics. It uses two GC columns with opposite polarity connected via a thermal modulator to reduce the band broadening (10, 49-50). Very high separation efficiency can be achieved. The two columns are normally coupled to a TOF-MS due to the need for high scanning speeds.

The rapid development of MS has resulted in a large variety of mass analysers, both for coupling to LC, GC or CE. Next to the TOF (with reflectron), the second most common mass analyser is the quadrupole which has lower performance than the reflectron-TOF. It is, however often combined in tandem MS arrangements like triple-quadrupoles or quadrupole-TOF (Q-TOF) with increased performance and utility for structural elucidation. A mass analyser with similar performance is the ion trap (IT) which, like the quadrupole, is often used in combination with other mass analysers as for example a quadrupole. The highest performing of all mass spectrometers are the Fourier transform MS (FTMS) (51) and orbitrap (52). They offer the best resolution, lowest detection limits and highest mass accuracy. They can also be coupled to, for example, a quadrupole to further increase the performance. For a detailed comparison and description of these mass analysers, the reader is referred to the review by Dettmer (10).

Direct injection or infusion mass spectrometry (DIMS) refers to the introduction of the sample into the ion source without prior separation (10). It has been used primarily for metabolic fingerprinting but can, with high resolution MS, be almost as efficient as hyphenated methods (53). Because this technique puts high demands on the resolution, FTMS, Q-TOF, orbitrap and triple quadrupoles are the

methods of choice. Advantages with this method are the high throughput as samples may be run in only a few minutes and a somewhat more unbiased approach due to the lack of sample preparation. Disadvantages are that isomers can not be distinguished and the quite large impact of ion suppression.

Finally matrix-assisted laser desorption ionisation quadrupole ion trap TOF (MALDI-QIT-TOF) has been used for metabolomics (54). Similar to DIMS, MALDI-TOF lacks complicated sample preparation steps. Additionally, it is highly sensitive and therefore requires only minute amounts of samples. However, like the other methods lacking a separation step it suffers from ion suppression (55) and the inability of separating isomers.

Nuclear magnetic resonance spectroscopy

In comparison to the MS-based methods nuclear magnetic resonance (NMR) spectroscopy holds the advantage of being a truly quantitative method as quantification is not performed relative to any standard (23). It has been widely used for metabolomics (6, 9, 56-58). Disadvantages with this method are a lower sensitivity and dynamic range, and perhaps most important, its lower resolution (23). The lack of sensitivity and resolution leads to that only the most abundant metabolites are detected, typically 30 components.

Sample pre-treatment for GC/MS

Although metabolomics use highly advanced analytical equipment, an Achilles heel is the sample pre-treatment which needs to be fast, efficient and unbiased. This is a complicated task as this procedure is prone to introduce biases towards the vast amounts of chemical and physical properties of the metabolites in the sample. A bias towards a metabolite or metabolite class will lead to an inaccurate measurement of the actual metabolite concentrations. In addition, the protocols must also allow for a high throughput and be gentle. Aggressive treatments are therefore not suitable as they may cause degradation of the metabolites. Several articles have described the development of sample pre-treatment protocols for different types of biological samples for example yeast, bacteria, plant cells, erythrocytes and human blood plasma (12, 16, 18-19, 21-22, 25, 59). In this section preparation of cells and blood plasma will be briefly described.

Preparation of cell samples

For a global metabolite analysis of cell samples, several steps are required and depending on the cell type and culture conditions different approaches are needed. These differences are primarily due to the culture conditions and the stability of the cell envelope. However, several steps are independent of cell type: 1) Quenching of the metabolism, 2) metabolite extraction. Both steps are sometimes performed at the same time thereby increasing the sample throughput. Most developments reported in the literature deal with chemostat and batch cultures, i.e. cells that are suspended in the culture medium (16, 18-19, 21, 59). In these studies it is important that no leakage of metabolites occur into the culture medium, which is lost when the cells are harvested. This does not necessarily apply to other types of cultures, as for example adherent cells where the cells can easily be separated from the culture medium prior to quenching. The cell envelope stability differs widely depending on the cell type. For instance, the cells may (bacteria, plant cells) or may not (mammalian cells) have a cell wall which requires tougher methods for cell lysis.

Quenching

As the cells rapidly change their metabolism, and therefore metabolite levels, in response to environmental changes, a rapid stop of the metabolism is critical for an accurate reflection of the metabolism (1, 19). Common procedures include rapid changes in pH or temperature. Rapid pH changes are often induced by addition of acids, for example perchloric acid, or bases like sodium hydroxide. These procedures may potentially hydrolyse many of the metabolites and are also not suitable for high throughput analysis since a neutralisation step generally must be introduced (1, 21). These chemicals may also cause problems in the following derivatisation step (21). Rapid changes in temperature include for example boiling and freezing of samples (16). These procedures are often performed in the presence of an organic solvent, out of which cold methanol is the most commonly used, even for chemostat and batch cultures, although several studies have shown that metabolite leakage occur (16, 18-19, 21, 59). It has also been suggested that methanol does not completely stop enzyme activities (16). Nevertheless, cold methanol quenching remains a simple and mild method. Other common methods include freezing samples in liquid nitrogen which is considerably more tedious.

Cell lysis and extraction

Following quenching the cells must be lysed and the intracellular metabolites extracted. This step is probably the most critical step because the lysis efficiency is dependent on the cell type and the extraction efficiency is highly dependent on the chemical properties of the metabolites. For example, simultaneous extraction of lipids, sugars and amino acids presents a great challenge due to their widely

differing polarities. In addition, the protocol needs to allow for high throughput and be non-destructive. Hence, developing an efficient extraction protocol for metabolomics is a complicated task and numerous protocols exist. Most of these are based on liquid-liquid extraction with violent shaking. A widely used protocol employs a two-phase system of chloroform, methanol and water (16, 21, 60). This approach normally provides good coverage of the metabolome but is more time-consuming and thus not suitable for high throughput protocols. Additionally, there is a risk for sample loss due to the difficulty in separating the two phases quantitatively. Nevertheless, it has been found to be the most suitable method for *S.Cerevisae* (yeast) (16). Other methods include potassium hydroxide, perchloric acid, hot water and boiling ethanol (16). These methods may degrade the metabolites and therefore not suitable for metabolomics sampling. An approach that is much more straight-forward and which has in several studies provided good coverage is extraction in cold pure methanol, usually accompanied with freeze-thaw cycles (19-21). Other advantages with the use of this method are that it can be directly combined with the quenching step, readily precipitates proteins, and has minimal influence on pH. The obtained samples can easily be concentrated which is usually required prior to derivatisation. A disadvantage is the use of freeze-thaw cycles which make the approach more time consuming.

Derivatisation

As only volatile analytes can be analysed using GC/MS and most metabolites are non-volatile, the primary aim of the derivatisation is to increase the volatility. There are several additional reasons to derivatise the metabolites, for instance to modulate the chromatographic behaviour to improve the separation of analytes with similar properties. In addition, improvements of the detectability and/or stability are common aims. However, despite the advantages gained with derivatisation there are several drawbacks. Obviously, it introduces at least one additional step, which may be quite time-consuming and tedious. For example, an optimised derivatisation method for silylation may take as long as 17 h to complete (22) and includes a heating step that may be invasive. Perhaps, the most serious drawback is that the yield in the derivatisation reactions also differs among the analytes and rarely approaches 100% in the sample. Therefore the measured composition determined in the sample may not accurately reflect the sample composition. Common for most derivatisation methods is that artefact compounds are produced which may complicate the GC/MS analysis. Some reagents can not be used with certain columns as they react with the stationary phase thereby causing heterogeneities. Many reagents also produce several derivatives from single metabolites. This may be prevented by an additional methoximation step. Lastly, the derivatisation is generally performed off-line and generally not

quenched prior to analysis; instead the reaction may proceed until the sample is injected on the column.

To increase the volatility, the derivatisation aims to reduce the intermolecular attractive forces between the metabolites by substituting polar functionalities in the molecule with a non-polar group. This has additional benefits in that it also improves the chromatographic properties of the metabolites by reducing the risk of adsorption of the analytes to the column wall, injector liner and to polar silanol-groups in the stationary phase. Adsorption can otherwise cause peak asymmetry, worsening the resolution and complicating the interpretation of the chromatogram.

Silylation

Several derivatisation methods have been used for metabolomics, of which the most common type is silylation with an alkyl-silane reagent (22, 39). These reagents exchanges acidic protons on hydroxyls (-OH), thiols (-SH), phosphates ($-OPO_3H_2$) and amines ($-NH_2$) with an alkyl-silane group through an S_N2 reaction. The reagents used for this purpose reacts with water and hydroxyl functionalities first. The sample components will therefore need to be completely dry prior to the reaction and no alcohols can be used as solvent. The most common silylation reaction replaces the acidic protons with a trimethylsilyl-group (TMS) (Fig. 3B). Several reagents can produce TMS derivatives but the most common silylation reagent is *N*-methyl-*N*-(trimethylsilyl)-trifluoroacetamide (MSTFA) which has been used in numerous studies (1, 12, 22). Compared to second most common TMS reagent, *bis*-trimethylsilyltrifluoracetamide, it generates more volatile analytes, has fewer side reactions and also derivatises $-NH_2$ groups more readily (41).

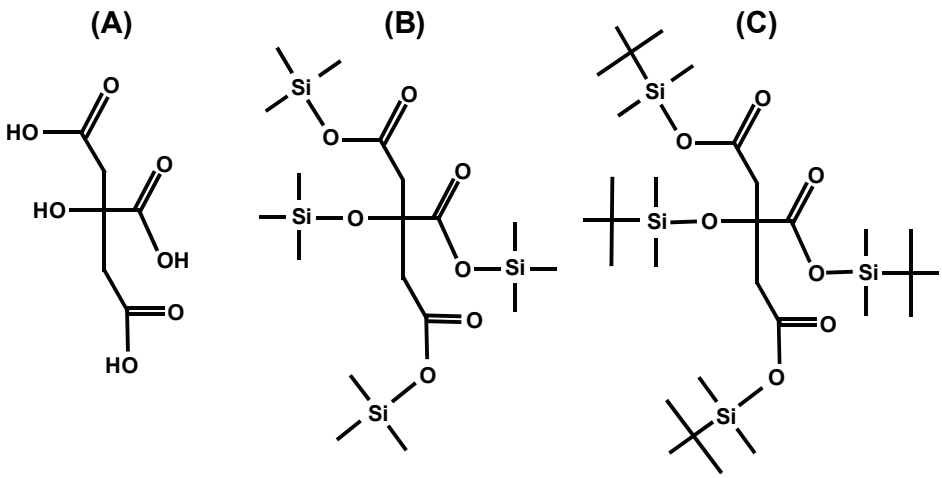

Figure 3 The structures for **(A)** citric acid, **(B)** citric acid 4TMS and **(C)** citric acid 4TBDMS

Another derivatising reagent is *N*-methyl-*N*-*tert*-butyldimethylsilyltrifluoroacetamide, MTBSTFA, which replaces the active hydrogens with a *tert*-butyldimethylsilyl (TBDMS) group (Fig 3C). This type of silylation has been used in fewer metabolomics studies (18, 39, 41, 61-62) compared to MSTFA derivatisation. It has previously also been used for structure elucidation (41). It has several advantages over MSTFA derivatisation. A significant advantage is that it generates a distinct and high intensity fragment due to cleavage of a stable *tert*-butyl radical from the derivatives (41, 63). This fragment consequently has the mass of the derivative minus the mass of the *tert*-butyl radical and is commonly referred to as $[M-57]^+$ fragment. The high intensity and known mass of this fragment has the benefit of aiding in identification (41). Other benefits assigned to the TBDMS derivatives are increased thermal and hydrolytic stability compared to the TMS derivatives (61, 64).

A special characteristic of TBDMS derivatisation compared to TMS derivatisation is related to the bulkiness of the TBDMS group. This bulkiness prevents full derivatisation of primary amines, sugars and sugar phosphates (18, 41). Therefore, quantification of compounds containing amine functionalities is simplified as there is one peak less to measure, while the quantification of sugars and sugar phosphates is no longer feasible as only a small part are volatile enough to be separated. However, the elimination of sugars and sugar phosphates is not necessarily a drawback. In blood plasma, the high concentration of glucose generally disables quantification of metabolites eluting with and close to glucose. Thus, with MTBSTFA these metabolites may be easier to quantify.

Oximation

Although silylation readily yields chromatographable compounds it is usually combined with an oximation step prior to silylation (12, 18, 22, 41). As the name implies, the product of an oximation is an oxime and it is formed by reaction between the oximation reagent, i.e. an alkoxyamine hydrochloride or a hydroxyamine hydrochloride, and carbonyl functionalities. This is primarily performed to prevent cyclisation of reducing sugars which otherwise may be detected as five peaks if TMS derivatisation is used (22). The origin of these peaks is due to that the reducing sugars naturally are present as several anomers in equilibrium. After TMS derivatisation, five tautomers are normally detected; one open-chain, two furanoses and two pyranoses (22). The multiple peaks and also the equilibrium between these forms, pose a problem in the quantification. However, after oximation of the carbonyl oxygen of the anomer carbon, only two peaks corresponding to the *syn*- and *anti*-forms are normally detected due to the reduced rotation around the C=N bond (22, 41). Additional advantages with oximation are that α-ketoacids are protected against decarboxylation (41, 65).

Chemometrics

To handle the complexity of metabolomics data, especially the vast amount of factors and responses involved in the study of the metabolome, sophisticated data handling methods are needed. Chemometrics can be viewed upon as a set of statistical tools which offer efficient experimental planning and analysis of complex systems (27). The experimental planning is often called DOE or statistical experimental design and enables systematic variation of factors that one wishes to explore (26-27, 66-69). It is particularly useful when there is a multitude of factors and responses that interact. The analysis part MVDA which allows for efficient extraction of information from highly complex data (2-3). By combining DOE and MVDA methods, the experimental results are likely to contain the sought information which can be efficiently extracted and modelled.

Design of experiments

In many scientific disciplines, it is necessary to investigate the behaviour of a system where many settings affect the outcome of the experiments. The aim is often to find the optimum settings and/or parameters in an experiment to achieve maximum performance and the best possible outcome. The settings can often be changed independently of each other. These settings are termed *factors.* The outcome of an experiment can also be observed as several types of results, and these are termed *responses.* A system where many factors are varied and several responses are measured becomes very complex and difficult to analyse, especially in the case with responses that depend on factor interactions and/or higher-order terms of several factors. Traditionally these systems are investigated by changing one factor while the others are held constant. This approach is often referred to as the COST approach (Change one factor separately at a time) and has several drawbacks (26). First, it does not take factor interactions or higher-order terms into account and, second, it will require large amounts of experiments to enable a correct description of the system.

A better approach for investigating such systems is DOE. It allows for systematic and simultaneous variation of the factors and system modelling using regression models and projection methods. In contrast to the COST approach, interactions

between factors can be discovered and also higher-order terms can be found. An additional benefit of using DOE is that it is normally highly time-efficient compared to the COST approach as the experiments are carefully selected and the factors are varied simultaneously.

DOE objectives

There are three main objectives for DOE, namely screening, optimisation and robustness testing (26). In screening, the aim is to find the most influential factors and also to efficiently map the experimental space, i.e. determine the ranges for the factors. The designs used for this purpose are therefore often intended to assess only the main factors, the interactions and their ranges. Higher-order terms are not considered at this stage, unless prior knowledge of the system is available. After the experiments have been performed and the data have been collected, the data are interpreted by means of MVDA methods and the most influential factors are determined. Led by the screening results, new experiments can be set up to find the optimum factor settings for the system. For this purpose the experimental design need to include at least interaction and quadratic factors. Therefore this objective requires more experiments to be performed. The last objective is robustness testing where the tolerance of the optimised system to changes in the factor settings is investigated.

Mathematical models

A system composed of several factors influencing several responses, can be described by a simple empirical model as described by equation 3.

$$\mathbf{Y} = f(\mathbf{X}) \tag{3}$$

Here **Y** represents a matrix of responses and **X** a matrix of factors. The function ***f(X)*** is then approximated by a polynomial that describes the relationship between the response and the factors. Depending on the objective of the study, the polynomials include different terms. For screening studies often only the main effects and the interaction effects are of interest. Therefore, the following empirical model is often applied, described in equation 4 for a single response for three factors, i.e. a third-order interaction model (69).

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 + E \tag{4}$$

Equation 4 describes a simple system with only three factors, x_1, x_2 and x_3, and their interactions, x_1x_2, x_1x_3, x_2x_3 and $x_1x_2x_3$ affecting a single response y are assessed. The magnitude of the coefficients b, for each factor is a measure of the

influence of the factor on y. The intercept is described by b_0 and represents the mean effect of all factors. As models are approximations of the underlying system, there will also be an error or residual in the model which is described by E. To optimize the above system, at least quadratic terms need to be added, resulting in a quadratic model (equation 5).

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 + b_1^2x_1^2 + b_2^2x_2^2 + b_3^2x_3^2 + E \quad (5)$$

The added terms results in that additional experiments need to be performed. The coefficients, b_i, are calculated from the experiments by the use of MVDA which will be later described in detail.

Model designs

Full factorial designs

One of the most frequently used designs is the full factorial design. This design is simple to understand and forms the basis for other more advanced designs. It is therefore useful to describe this model in detail. Going back to the above system with three factors, it can be calculated how many experiments that are required to asses all the effects if the number of factor levels are specified. In a system with k factors, 2^k experiments will be needed. Hence, for a system with three factors at two levels, the number of experiments will be $2^3 = 8$ experiments. However, a set of replicates is usually added to allow the reproducibility of the experiments to be measured. If the reproducibility can be considered being similar over the entire experimental space, the replication is preferably performed in the design centre point. Adding the replicates in the centre also enables curvature in the response to be detected in 2-level factorial designs. Such a design can easily be graphically illustrated (Fig. 4). However, it is important to understand that the reproducibility of the centre points might not in all cases accurately assess the reproducibility over the entire design. Especially as the experiments in the outer parts of the design may use more extreme settings than what is used in the model centre. Ideally, all experiments should have replicates.

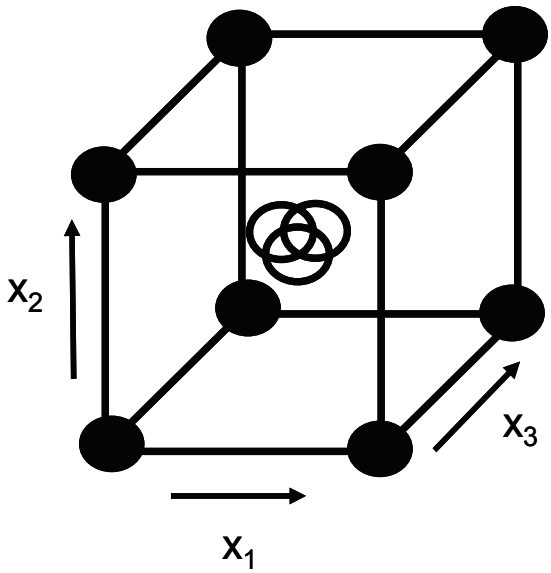

Figure 4 A full factorial design with 3 factors (x_1, x_2, x_3) and two levels (solid circles). Included are also 3 centrepoints (circles).

Central composite designs

The central composite design is a two-level full factorial design combined with replicated experiments in the centre and symmetrically positioned axial points on each side (26-27). There are two kinds of central composite designs; the face-centred (CCF) and the central composite circumscribed (CCC). These designs are useful primarily for optimisation as it investigates each factor at three (CCF) or five (CCC) levels. This allows for checking for curvature in the responses.

Mixture designs

To optimise the composition of a mixture, a mixture design can be used (26-27). These designs are different from process designs because the constituents in the mixture can not be changed independently. Instead the sum of these constituents is always 100% or 1 and each constituent is included as a fraction of the total sum. This is referred to as closure and requires a different treatment compared to process designs. Two types of factors can be incorporated in a mixture design: mixture factors or fillers. The difference is that only the mixture factors (also called formulation factors) have an effect on the response(s). The filler may for instance be an inert solvent.

D-optimal designs

The D-optimal design is a computer-generated design which is exceedingly useful in special cases of experimental design. These cases do for example include when the experimental region is irregular. Irregular design regions can occur for instance in a mixture design where part of the mixture is inapplicable. There are several more cases where D-optimal designs are useful and for these, the reader is referred to a review by de Aguiar (70) where also the properties of this design is thoroughly described. The D-optimal designs are created from a candidate set of experiments to cover an as large volume as possible of the experimental region. The candidate set contains all theoretically possible experiments. To select the experiments that cover the maximum volume, the criterion of D-optimality is used. This criterion means that when the determinant from the matrix $(X'X)^{-1}$, where X is the design matrix, is minimised, the maximum volume of the experimental region is covered.

Multivariate data analysis

After having performed the experiments, preferably using DOE, the data need to be evaluated and converted into useful information. Usually, the number of factors and the number of responses are large, yielding a large data matrix, $\boldsymbol{X}$, with N rows and K columns. Each row represents an observation (sample), run at the conditions from the experimental design; each column represents a factor (predictor), which in GC/MS based metabolomics is the measured peak area of a detected metabolite. However, the data are also contaminated with noise and the peak areas of different metabolites may differ up to a factor of several thousands (71). To extract relevant data from such a large, heterogeneous and contaminated data set, several data pre-treatment steps followed by powerful regression and projection methods are needed.

Raw data pre-treatment

Before the useful information can be extracted from the GC/MS raw data by multivariate data analysis methods, the data are pre-processed in several steps. The first steps include converting the retention times to Kovats retention indexes and calculating the peak areas of all eluted analytes. The former step is simply performed by analysing a sample of homologous n-alkanes. The latter is a highly complicated task due to the many samples and overlapping peaks that are normally found in analyses of biological samples. The peak overlaps are the result of the

large number of analytes and artefacts that are eluted and the insufficient resolution of all current separation methods. Software packages exist that are capable of accomplishing these operations. For instance ChromaTOF™ software (Leco Corp., St Joseph, MI, USA). In this thesis hierarchical multicurve resolution (HMCR), a software tool developed for MATLAB™, was used (72).

HMCR is applied on the un-processed raw data from the GC/MS analyses. These data can be viewed as a three-dimensional structure where the three dimensions contain the mass spectral information, the chromatographic information and the sample number, respectively. These data require several steps of pre-processing before they can be subjected to multivariate modelling. All these steps are taken care of by the use of semi-automated scripts developed for MATLAB™ (72-73). A detailed analysis on this subject is outside the scope of this thesis. Only a brief description will be presented here.

The data collected for each m/z – channel is first smoothed by a moving average to reduce the noise, and the baseline for each sample is corrected for. The chromatograms from each of the samples are then aligned by determining the maximum covariance. This corrects for the common occurrence of retention drift and makes it possible to divide the chromatograms into time windows. The limits for the windows are set at positions at which the intensity for all m/z-channels is low. Consequently, the window size varies along the retention time axis. The mean signal for all these low intensity points, in all chromatograms, is also used as a measure of the signal background which is then removed in each time window.

Each time window now contains a number of peaks out of which many are overlapping and will thus have to be deconvoluted in order to allow for an accurate quantification. The deconvolution is performed by alternating regression (74) which yields resolved chromatographic profiles with their corresponding mass spectra. From this information, the peak areas for each deconvoluted component can be calculated due to that the mass spectrum must be homogeneous across the whole peak. The peak areas and the retention index for the peak maxima are then conveniently put in a matrix that can be used for further processing by multivariate data analysis tools.

After the deconvolution has finished it should be checked as erroneous m/z:s are sometimes used for the integration. This occurs primarily in regions where there are many co-eluting analytes and hence overlapping peaks as it is in these regions that it will be difficult to elucidate which m/z that belongs to which peak. This is performed by comparing the pure spectra to those used for the peak area calculations and manually plot the m/z:s for these overlapping peaks. If errors are found the peaks need to be manually integrated using a single or a few m/z:s

which are not always possible to find. Hence, even though current deconvolution software approaches are powerful, they are still not a real substitute for having an optimal chromatographic resolution. For a thorough description and comparison of deconvolution software the reader is referred to the paper by Lu (75).

Finally, metabolite identification is performed using database searches based on Kovats retention index and mass spectra.

Normalisation, centring, scaling and transformation

The data table containing the peak areas (or peak heights) needs to be further treated in order to focus the data analysis on the induced variation (2-3). The metabolomics data have several properties that make it difficult to analyse without pre-treatment. The large differences in abundance of the analytes need to be compensated for, as the biological effect of a metabolite is related to the variation rather than to the absolute level. The induced variation is also mixed with other types of variation, systematic and random, emanating from biological and technical sources. This variation should be reduced, or preferably eliminated, in order to study the sought variation. The technical variation can in part be reduced by four main operations: normalisation, centring, scaling and transformation. These will briefly be discussed below.

The first step to perform is normalisation, which aims to compensate for systematic errors appearing in the metabolomics protocol. Examples where errors occur are during the extraction, sample derivatisation, chromatography, and the mass spectrometry (76). The systematic part of these errors occur because of differences in metabolite properties in conjunction with the method selectivity and result in that the calculated peak area no longer is proportional to the original concentration of the analyte. A remedy is therefore to relate the peak area of the analyte to the peak area of an internal standard which is added in a known amount and should have the same properties as the analyte. Therefore, stable isotope-labelled counterparts of the metabolites are often used as internal standards. The assumption is that the isotope-labelled standard only is affected by the systematic variation (76).

Even though most studies use isotope-labelled standards, there are several drawbacks with the use of this method. For instance, internal standards for all metabolites do not exist and are generally expensive. Therefore, many studies use only a few internal standards and normalise metabolites lacking a proper internal standard by a standard having approximately the same properties. A more innovative approach is to use a few standards, preferably covering many metabolite properties, and use principal component analysis (see below) to

describe their variation in all sample runs, and normalise all metabolites using the scores of the first principal component (77). Each sample is therefore approximately divided by the average level of the internal standards. However, this requires that the PCA model describes almost all of the variation in the first component.

A more serious drawback with normalisation based on internal standards is the requirement that the standards produce unique fragments with detectable intensities for identification and quantification. Hence, the mass shifts should also be quite large to not be confused with the isotope patterns of the unlabelled counterparts, e.g. cross-contribution. Equally important is that the chromatographic resolution must be high enough to avoid cross-contribution from co-eluting analytes (76). Careful choice of internal standards and an optimised chromatography is therefore crucial. High performance mass spectrometers like the FT-ICR MS (78-79) do alleviate the problem of unique mass fragments as they offer very high (mass) resolution and sensitivity. As the instrumental performance currently is limited, several methods have been developed to reduce the influence of cross-contribution by means of mathematical methods. These methods are however, outside the scope of this thesis and the reader is referred to a recent paper by Redestig (76). A common problem when studying cellular samples is difficulties to control the actual amount of sample. Here, normalisation to one or a group of sample components reflecting the sample size is required. In biomedical research this normalisation is usually performed using, e.g. the transcripts of a housekeeping gene or the total protein content. In GC/MS based metabolomics, the total ion count (TIC) can be used as it reflects all detectable metabolites, in analogy to the total protein normalisation.

After normalisation has been performed it is assumed that the systematic variation for each analyte is the variation induced by the experimental design. However, further complications of metabolomics samples must now be alleviated. First, the widely different abundances of the metabolites cause two problems: an offset between the mean concentrations of the analytes and a difference in the magnitude of the variance (71). This will likely cause erroneous conclusions if not corrected. To correct for the offsets, the mean for all measurements of a response is withdrawn from each measurement. This is called centring and efficiently focuses the investigation on the fluctuations in the measurements rather than on their magnitude. Thus, now only the variation for each analyte is studied. However, the magnitude of the variation is still dependent on the abundances of the analytes; consequently highly abundant analytes will be more influential in the following data analysis if the data are not scaled. Several methods for scaling are available (71, 80-81) with the common aim that the data should be directly comparable. The most frequently used scaling method and also the method applied in this thesis is

autoscaling or unit variance (UV) scaling. It converts all variances to unit variance by dividing each measurement with their standard deviation.

Finally, some measured responses may need to be transformed mainly to compensate for heteroscedasticity. Heteroscedasticity means that the variance for a randomised variable is not constant (71). This may cause a deviation, skewness, of the residual distribution due to a few measurements with extreme variance resulting in that these metabolites dominate the model. Another reason to transform a response is to make modelling of non-linear factor – response relationships possible, by linearisation. A common type of transformation is the log-transformation, which is applied in this thesis.

Projection methods

Consider the metabolomics data in the pre-treated data matrix, **X**. These data contain for the moment only a vast number of observations, *N,* and factors, *K,* with no apparent relationship. Moreover, the observations have usually been collected under different conditions. A common approach to describe a complex system is to visualise it in a coordinate system, which traditionally will require as many dimensions as there are variables. As the data from a metabolomics investigation frequently contains 100 or more variables due to all the measured responses, plotting the raw data will not provide an effective approach for interpretation. Hence, for efficient visualisation, the number of dimensions needs to be reduced while still preserving the data as the global approach will be lost if the number of variables is reduced. This is accomplished with projection methods that aim to find latent variables and structures describing the data more efficiently, i.e. in fewer dimensions. The latent variables are orthogonal and the projection methods can therefore be seen as a change of the coordinate system.

The projection methods can be divided into two main variants: unsupervised and supervised methods. The difference between these methods is that for the unsupervised methods, no prior information about the data is needed. In contrast, the supervised methods require prior information about the observations. The supervised methods include regression and classification methods. Regardless of the method used, much of the interpretation power comes from powerful graphical tools which allow elucidation of variable importance, relationships between variables and relationships between the measured responses and *a priori* information. In this section, the unsupervised method principal component analysis, PCA (82), will be described followed by two supervised regression methods, projections to latent structures, PLS (83), and orthogonal PLS, OPLS (84).

Principal component analysis, PCA

The most common multivariate projection method is PCA. It aims to find principal components, e.g. latent variables, to describe the variation in the data in fewer dimensions. The principal components are found by the use of the nonlinear iterative partial least square, (NIPALS) algorithm (85). The pre-treated data matrix, **X,** containing *N* observations and *K* factors is decomposed into a scores matrix, **T,** a loadings matrix, **P,** and a residual matrix **E,** according to equation 6.

$$\mathbf{X} = \mathbf{TP}^{'} + \mathbf{E} \tag{6}$$

The principal components consist of the product of an individual score, **t**, and a loading, **p,** while **E** describes the error in the approximation. The first principal component describes the largest proportion of the data variation and the second principal component describes the second largest and so on until all the significant latent variables in the data are described.

$$\mathbf{X} = t_1 p_1^{'} + t_2 p_2^{'} + t_3 p_3^{'} + ... + t_z p_z^{'} + \mathbf{E} \tag{7}$$

Hence, PCA seeks to maximise the explained variance in the data. The significant latent variables can be found in several ways. The most common approach is cross validation (27, 86) which will be described below but it can also be found by determining the eigenvalues for each component (27). The data that are not found significant is contained in the residual matrix, **E**, and should only contain noise, i.e. non-systematic randomised variation.

Projections to latent structures, PLS, and orthogonal PLS, OPLS

In contrast to PCA, PLS is a supervised method that seeks the linear relationships between the factors in **X** and *a priori* information, i.e. the responses, in matrix **Y** (2, 83). Hence, PLS is a regression method and can be used to find the influence of the factors in **X** has on the responses in **Y**. Alternatively, a discriminant analysis can be performed, PLS-DA. The **Y** matrix then contains information of classes as dummy variables. PLS is performed by fitting two PCA-like models for **X** and **Y** and aligning these simultaneously. Thereby the covariance between **X** and **Y** is maximised, offering maximal ability to predict **Y** from **X**. A typical example is when optimizing the factor settings for a process with the use of DOE (**Paper I** and **II**). Similarly to PCA, both the **X** and the **Y** matrix are decomposed into scores, loadings and residual matrices (equation 8 and 9)

$$\mathbf{X} = \mathbf{TP}^{'} + \mathbf{E} \tag{8}$$

$$\mathbf{Y} = \mathbf{U}\mathbf{C}' + \mathbf{G} \quad (9)$$

Here, **T** is the score matrix for **X** and **P** is the loading matrix for **X,** while **U** and **C** are the score matrix and loading matrix for **Y**, respectively. As two models are created there are also two residual matrices, **E** and **G** for the projection of **X** and **Y**, respectively. The PLS components are found by the PLS-NIPALS (85) algorithm and are analogous to PCA, described by combinations of **TP'** and **UC'**. The scores for **X**, i.e. the **T** matrix, are also good predictors for **Y**. Hence, we can also predict **Y** from the scores from **X**, according to

$$\mathbf{Y} = \mathbf{T}\mathbf{C}' + \mathbf{F} \quad (10)$$

where, **F** is a residual matrix arising from the projection of **Y** from **T**. The scores can also be written as

$$\mathbf{T} = \mathbf{X}\mathbf{W}^* = \mathbf{X}\mathbf{W}(\mathbf{P}'\mathbf{W})^{-1} \quad (11)$$

where $\mathbf{W}^*$ is the matrix with transformed weights which are the coefficients for the linear combinations with the original variables (83) and **W** is the weight matrix. Hence, the weights can be used to rewrite the expressions for **Y** (and **X**)**.**

$$\mathbf{Y} = \mathbf{X}\mathbf{W}(\mathbf{P}'\mathbf{W})^{-1}\mathbf{C}' + \mathbf{F} = \mathbf{X}\mathbf{W}^*\mathbf{C}' + \mathbf{F} = \mathbf{X}\mathbf{B} + \mathbf{F} \quad (12)$$

From equation 12 it is clear that **Y** can be estimated from the values in **X** by the use of the **B** matrix, which contains the regression coefficients.

$$\mathbf{B} = \mathbf{W}(\mathbf{P}'\mathbf{W})^{-1}\mathbf{C}' = \mathbf{W}^*\mathbf{C}' \quad (13)$$

A limitation of the PLS approach is that the scores used to predict the responses in **Y**, contain all systematic variation in the **X** matrix. Some of this systematic variation may not be linearly related **Y** and can therefore cause difficulties in model interpretation. To address this problem, orthogonal PLS (OPLS) was developed (3, 84). The term orthogonal comes from the ability of OPLS to divide the systematic variation in two parts, the predictive variation and the orthogonal variation. Additionally, a residual matrix of unexplained variation is also created. This is performed according to equation 14 and 15, where the **X** and **Y** matrices are decomposed into scores, loadings and residuals.

$$\mathbf{X} = \mathbf{T}_P\mathbf{P}_P' + \mathbf{T}_o\mathbf{P}_o' + \mathbf{H} \quad (14)$$

$$\mathbf{Y} = \mathbf{T}_P\mathbf{C}_P' + \mathbf{I} \quad (15)$$

Here, $\mathbf{T_p}$ is the score matrix for the predictive data in **X**, and $\mathbf{P_p}$ the corresponding loadings. The orthogonal variation is described by the scores in $\mathbf{T_o}$ and $\mathbf{P_o}$. The weights used to predict **Y** from **X** are contained in the $\mathbf{C_p}$ matrix and the residuals for **X** and **Y** are contained in the **H** and **I** matrices, respectively. The predictive variation is the systematic variation in **X** that is linearly related to the responses in **Y** and can be used for the prediction. In contrast, the orthogonal variation is the systematic variation that is not related to **Y** and instead explains other systematic variation in **X**. As the orthogonal variation is systematic, it may describe other phenomena of interest in the data and OPLS allows for further investigations of this so called structured noise.

Model validation

After a model has been created, it needs to be optimised and validated before it can be interpreted (69). Several tests can be made to assess the validity and performance. The total variation in the system can be described as the sum of squares, *SS*, which can be further divided into three parts according to

$$SS = SS_{reg} + SS_{resid} = SS_{reg} + SS_{pe} + SS_{lof} \tag{16}$$

where SS_{reg} accounts for the variation that is explained by the regression model and SS_{resid} accounts for the variation that could not be explained by the model, i.e. the residual variation. The residual variation can be further divided into variation caused by pure error, SS_{pe} and the variation caused by lack of fit, SS_{lof}. The pure error emanates from a lack in repeatability or reproducibility while the lack of fit comes from that the model poorly describes the variation in the model. From the variations, several statistical measures of the model quality and predictive ability can be constructed. The amount of explained variation gives information about the amount of explained variation in the model and is formulated as

$$R^2 = \frac{SS_{reg}}{SS} \tag{17}$$

Hence, a model with an explained variation close to one describes most of the data in the system and consequently has small residuals. However, R^2 is not a good measure of model quality as it is very sensitive to the degrees of freedom. Just by adding more factors to the model, the explained variation can become close to one. Another, much more rigid parameter for testing the model quality is the cross validated predictability (86), Q^2. As the name implies Q^2 relies on cross validation to evaluate the predictive power of the model. Cross validation works by leaving out one or more observations from the calculated model and predict the responses. The calculated responses are compared to the responses obtained with the

complete model and the prediction error sum of squares, PRESS, is calculated (equation 18). This is performed several times with different values being left out each time.

$$Q^2 = \frac{SS - PRESS}{SS} \tag{18}$$

In contrast to R^2, increasing the number of factors in the model does not increase Q^2 and is therefore the most important parameter to evaluate when creating a model.

It is also of interest to assess the pure error variation described by SS_{pe} and the lack of fit variation, SS_{lof}. The pure error variation should be small compared to the total variation in the model. Otherwise it is not possible to obtain a good model as the variation to a large extent comes from random variations, such as the injection in GC. This can also be interpreted as that the experimental design should induce sufficiently much variation in the data to always be larger than the pure error. To test the significance of the pure error its variance is compared to the variance of the model error by an *F*-test (69). Ideally, the variances should be of equal size. This means that the residual variation is only due to random variation. However, if the pure error is significantly smaller, the residual variation is mainly due to the model error and the model has lack of fit and does not accurately describe the data. This is often also visible as a low value for Q^2. Often additional terms or better control of the pure error increases the model quality. However, a drawback with this method is that to estimate the pure error, the experimental design must include replicates.

Results and discussion

Development of metabolomics protocols (Papers I and II)

In this thesis, two metabolomics protocols were comprehensively optimised using DOE in combination with MVDA. This approach is ideal for this type of optimisation as each step is influenced by many settings (factors) and several performance criteria (responses) are assessed in each step. Moreover, the settings often interact, either synergistically or antagonistically.

In **paper I**, a metabolomics protocol for blood plasma analysis using MTBSTFA as silylating reagent was developed and optimised. Derivatisation with MTBSTFA offers an improved derivative stability compared to the more common reagent MSTFA. The use of MTBSTFA also facilitates the data analysis, which is primarily due to the generation of high intensity fragments with an m/z of $[M-57]^+$ for each analyte (41). In addition, carbohydrates are eliminated from the gas chromatography step. This is beneficial for blood plasma analysis, where analytes with similar retention otherwise are very difficult to detect.

A critical step in metabolomics analysis is the sample pre-treatment step, since it is prone to introduce biases towards metabolite properties (16, 18, 20, 59). While many protocols have been developed, the majority of these are optimised for chemostat or batch cultures for yeast and bacteria. Hence, in **paper II**, a development and optimisation of a metabolomics protocol for adherent cell cultures, clonal β-cells (INS-1 832/13), is presented. Here, MSTFA was used as silylation reagent.

Cell disruption and extraction (Paper II)

After quenching the cells in pure methanol at -80°C, the samples were extracted at 100% (v/v) methanol or 82% (v/v) methanol using either a ball-mill, vortexing (violent shaking) or snap freezing in liquid nitrogen. An attempt to optimise a single phase system with methanol chloroform and water was performed in a pilot study and showed that the addition of chloroform did not increase the coverage (data not shown). The commonly used two-phase system, chloroform/methanol/water, was not an option due to its polarity biases and

throughput limitations (19, 21), although some reports claim better coverage of the metabolome with this procedure (16). Therefore, only a single phase system with methanol and water was optimised in this study. In total, six combinations were evaluated by use of several OPLS and OPLS-DA models. Based on these models, it was found that the highest extraction efficiency was gained with ball-milling at 82% methanol.

Sample derivatisation optimisation (Paper I)

The two-step derivatisation procedure was optimised with respect to temperature, duration and the composition of both reaction solutions. The experimental plan was generated with a D-optimal design. It was found that the oximation settings had a strong influence on the yield of all analytes, which generally decreased with increasing oximation duration and temperature. In analogy, the same was true for the silylation settings, although a larger number of responses were now oppositely affected. For both steps, modulation of the reaction solvent by additional organic modifiers showed almost no benefits. For this reason, organic modifiers were excluded. In summary, it was found that methoximation is best performed at 20°C for 4 h. The following silylation step is best performed at 100°C for 2h 50 min.

Optimisation of the injector settings (Papers I and II)

The settings for injection temperature and purge vent time were optimised using three level full factorial designs to obtain the maximum peak areas for all analytes. An injection volume of 1 µl was used in both studies to reduce liner and column contamination. A high injector purge vent time was beneficial for all analyte responses in both studies and was therefore set at the highest investigated value of 115 s. In contrast, the analytes showed heterogeneous behaviour when changing the injector temperature. While some analyte responses benefitted from an increase, other responses were decreased. An interesting finding in **paper I** was that partially derivatised metabolites, i.e. metabolites lacking silylation on some functionality, had a significantly decreased thermal stability to that of their fully silylated counterpart. The optimum injector temperature was set to 270°C in **paper I** and to 260°C in **paper II**.

Gas chromatography optimisation (Papers I and II)

In each of the studies, two columns that are commonly used for metabolomics studies were investigated: a 10 m (DB5MS, inner diameter (i.d.) 0.18 mm, film thickness 0.18 µm) and a 30 m (DB5MS, i.d. 0.25 mm, film thickness 0.25 µm). For each column, a three level face-centred CCF design was applied to find the optimum settings for high peak capacity, symmetric peaks, high sample throughput and large peak heights. The investigated settings were the injector purge vent time, the carrier gas flow rate, initial gradient temperature, initial

gradient temperature duration and temperature gradient. The injector purge vent time was included also here as it may influence the separation efficiency. To facilitate the interpretation, the two designs were merged.

In both studies it was found that the short column was sensitive to the factor settings, especially a high injector purge vent time which caused a detrimental decrease in peak capacity. This severely limits the application of this column as it already has inferior peak capacity compared to the longer column. Hence, only the longer column was further optimised. For this column the purge vent time of 115 s could be used with only a minor decrease in peak capacity. The flow rate was set to 1 mL/min due to that higher settings decreased the peak capacity, likely due to mass transport limitations between the mobile and stationary phase. The overall dominating factor for the chromatography was the temperature gradient rate which was set to 25°C/min in **paper I** and 15°C/min in **paper II**. At these settings the long purge vent time only caused a minor decrease in peak capacity. The initial temperature duration only had small effects on the peak capacity and was set to the lowest value of 2 min to increase the sample throughput. Finally, increasing the initial temperature decreased the analysis time as it decreased the retention for the last eluted peak. However, the initial temperature also had effects on, for example, the peak capacity and the peak shape. After a thorough evaluation of all significant interactions, the optimum settings were found to be 60°C and 82°C, for **paper I** and **paper II**, respectively.

Mass spectrometry (Paper I)

For the mass spectrometry, the ion source temperature and acquisition rate were investigated in order to find if an increase in the peak areas could be obtained. For this purpose, a three level full factorial design was used. It was found that an increase in ion source temperature increased the peak areas for all analytes and it was therefore set to the highest setting of 250°C. An increased acquisition rate had a generally low influence on the peak areas, but a significant decrease was observed for late eluted analytes. A possible drawback with a higher ion source temperature is that the fragmentation pattern might change, rendering spectra that are incompatible with many databases. This was investigated by calculating the ratio of the peak areas for $[M-57]^+$ and a small, unique fragment for each detected analyte It was found that this ratio was affected by the ion source temperature. Thus, the ion source temperature affects the measured mass spectra and a temperature far-off the range found in common spectra databases should therefore not be used. Most interestingly, the acquisition rate affected the ratio for early and late eluted metabolites, whereas this effect was insignificant for the metabolites with intermediate elution times. The acquisition rate also affected the ratio, which probably is related to some technical feature of the equipment used. The optimum settings were found to be 250°C and an acquisition rate of 20 Hz

Sample drift comparison (Paper I)

As a final validation of the performance, the sample drift, i.e. the change in peak area response with run order was compared for human blood plasma samples, derivatised with MTBSTFA or MSTFA. The validation was performed by creating two models, one for each sample set, relating the peak area to the run order with OPLS. It was found that the MTBSTFA sample set showed less drift. This effect was probably due to the larger TBDMS groups, which prevent the formation of disilylated amines on amino acids. Disilylated amines on amino acids can be formed when using MSTFA. This occurs quite slowly and can thus occur during the analysis as the reaction is not quenched. Indeed, both mono- and disilylated amines were found to drift in the MSTFA sample set.

Method validation (Papers I and II)

In both papers validation with real samples were performed. In **paper I,** a set of human blood plasma samples was used, while in **paper II**, clonal β-cells (INS-1 832/13) were used. The validity of the proposed methods was evaluated with regards to linear range, precision and detection limit. The linear ranges were rigorously tested using both lack-of-fit tests and by cross-validating the linear model. For the developed method in **paper I,** all metabolites exhibited linear responses up to 80% of the stock blood plasma concentration with low detection limits and high precision. The developed method in **paper II** showed linear responses ranging from a single well in a 96-well plate to a 10 cm culture dish for many metabolites. For both methods the detection limits were found on par or better than for other methods.

Applications of metabolomics to stimulus-secretion coupling (Papers III-V)

Glucose homeostasis and its tight coupling to insulin secretion are essential for whole body metabolism. This homeostasis is determined by a combination of GSIS and how well the hormone exerts its effect in peripheral organs, such as skeletal muscle and adipose tissue. If the latter, termed insulin sensitivity is impaired, more insulin will be required to control glucose homeostasis. Not much is known about the global metabolic response to glucose stimulation of the pancreatic β-cell, and its correlation to insulin secretion. It is however clear that the peripheral action of insulin affects a large variety of metabolic pathways. Thus, regulation of several classes of plasma metabolites can be expected. In order to

elucidate this complex machinery, we studied the metabolic response of glucose stimulation in plasma and pancreatic β-cells in three papers.

Paper III

Metabolomics, using GC/MS in conjunction with insulin secretion measurements were applied to investigate the metabolic response in humans, upon an oral glucose tolerance test (OGTT). The OGTT is commonly used to diagnose diabetes and is solely based on the blood glucose level. The objective of the study was to investigate whether other metabolites also could be reliable indicators of regulated metabolism. With the use of several OPLS models, the variation in the data could be divided into metabolic variation induced by the OGTT and gender, and instrumental variation. An OPLS-DA model resulted in an almost perfect separation according to gender, which indicates a pronounced uninduced biological variation. The OPLS model describing the variation over time induced by the OGTT showed clear differences according to the time after the OGTT. Pronounced drift could be found in the data, even though a comprehensive normalisation was performed. In order to find metabolites uniquely induced by the OGTT, the correlation loadings, p(corr), from this model were plotted against p(corr) for the other two models. It was clearly seen that the fatty acids and an unknown metabolite N85 were uniquely regulated by the OGTT as their p(corr) values were high for OGTT and low for the other models.

However, it is also of interest to find metabolites that correlate to the insulin secretion profile which was found to peak at 30 min. Therefore two OPLS models containing data for 0-30 min and 30-120 min after the OGTT were created. Their correlation loadings were plotted against each other. As the fatty acids had low loadings in both time intervals, they were down-regulated in both time intervals. In contrast, the positions for mannose, glucose and several unidentified metabolites (N79, N35, N67, N10, N45, and N83) indicated that these metabolites were passing a minimum or maximum at 30 min, i.e. these were correlating or inversely correlating to the insulin secretion profile. The findings from the models describing the response to the OGTT were confirmed using raw data plots. As an additional test, Pareto scaling was applied to the data for comparison to the UV-scaling that was initially used. In accordance with the UV-scaled data, fatty acids and N85 were concluded to be reliable markers for the OGTT. However, the loadings along the predictive component were very different. Pareto scaling of the data would have resulted in that some of the strongest regulations had been overlooked.

Paper IV

The global metabolic response of clonal β-cells (INS-1 832/13) to glucose stimulation was studied with the purpose to elucidate the amplifying pathway. The amplifying pathway augments insulin secretion after that it has been triggered by a rise in intracellular Ca^{2+}. It has long been assumed that this is due to the rising level of a single metabolic signal. Glutamate, NADPH and long chain acyl-CoA are among those suggested to convey this signal. In this study, a total of 195 putative metabolites were quantified by GC/MS at 3, 6, 10 and 15 min after glucose stimulation. Additionally, cell perifusion was used to determine the dynamics of the insulin secretion profile. From an initial screening, the largest alterations in the metabolome were found in the first 15 min after the glucose stimulation. Hence, they coincided with the peak of insulin secretion. From an OPLS model relating the metabolome to the time for quenching it was found that the samples clustered according to their quenching time. From the loading plot it could be concluded that this was primarily due to a decrease in the levels of long chain fatty acids and amino acids and an increase in glycolytic and TCA-cycle intermediates. To find the rates of onset and decline of the metabolic regulation, correlation loadings from two models describing the changes in the metabolome from 0 to 10 min and 10 to 15 min were plotted. These intervals border at the time of the peak in insulin secretion as measured in the perifusion experiments.

From the correlation loading plot, it was found that both glycolytic and TCA cycle intermediates generally increased in both time intervals. The positions of the TCA cycle intermediates differed in their time of onset, with the early and late intermediates having a fast onset. The position of the pentose phosphate shunt intermediate ribose 5-phosphate showed that this metabolite has a local maximum at approximately 10 min. Most amino acids were found to be down-regulated over the whole time-span, which was expected as proteolysis and protein synthesis are affected by the glucose availability. The exceptions from this trend were glutamate, pyroglutamate and alanine which instead increased over the whole time-span. Interestingly, aspartate and glutamate, involved in mitochondrial shuttling, displayed the strongest decay and onset of all amino acids, respectively. This is most likely due to that net oxidation of NADH is needed for maintained glycolysis. Moreover, glutamate dehydrogenase reversibly converts glutamate to α-ketoglutarate to fuel the TCA cycle during starvation. The fatty acids showed heterogeneous behaviour with short chain fatty acid being up-regulated while long-chain fatty acids were down-regulated. This may be due to a shift from β-oxidation during starvation to fatty acid synthesis after stimulation. Several unidentified carbohydrates and ribose 5-phosphate were also found to correlate with the insulin secretion profile, while fatty acids were found having an inverse correlation. The common denominator for these metabolites is the involvement of

NADPH, suggesting that NADPH may be the coupling factor. However, the complex interaction of all pathways as shown in the present investigation, suggests that a more likely scenario is that a metabolic pattern stemming from several pathways plays a role in stimulus-secretion coupling.

Paper V

The effect of excess palmitate and glucose on stimulus-secretion coupling in clonal β-cells (INS-1 832/13) was studied with GC/MS metabolomics, insulin secretion and viability assays to elucidate the metabolic effects of glucotoxicity, lipotoxicity and glucolipotoxicity. High levels of free fatty acids and glucose are commonly found in conjunction with T2D. This has been suggested to render a toxic environment for the cells. It has also been shown that high levels of fatty acids disturb stimulus-secretion coupling. Here, models for lipotoxicity and glucotoxicity were developed using DOE. In addition, a possible glucolipotoxic model was developed.Measuring insulin secretion, it was found that a rise from 2.8 mM to 16.7 mM caused a 40-fold increase in insulin release for cells cultured under normal conditions. In contrast, under the three toxic conditions, GSIS diminished to 5-fold. By using OPLS, a clear clustering according to growth conditions were revealed for all four conditions.

In analogy with **paper III** and **IV**, three new OPLS models were calculated and their correlation loadings were plotted to find uniquely regulated metabolites for each of the toxic conditions. In addition, a 3D-SUS plot was constructed (three-dimensional shared and unique structure plot), From these plots, γ-aminobutyric acid (GABA) was found to be uniquely regulated in the comparison of gluco-lipotoxicity and glucotoxicity and in the comparison of glucotoxicity and lipotoxicity, whereas it was shared between glucolipotoxicity and lipotoxicity. In the lipotoxic and glucolipotoxic model, GABA was down-regulated. In contrast, it was up-regulated under glucotoxic conditions, implying that glucotoxicity and lipotoxicity exert their effects on different metabolic pathways. Intermediates for both the TCA cycle and glycolysis were found to roughly reflect the level of glucose used during culture. Their levels were thus largely unaltered by lipotoxicity. However, citrate and succinate were exceptions. Citrate was shown to increase during lipotoxic conditions, suggesting that acetyl-CoA from β-oxidation supplies the TCA cycle with carbons when the glucose level is low. Succinate levels were shown to lose its glucose-responsiveness in the presence of palmitate. The levels of fumarate were glucose-responsive, which may indicate a sufficient replenishment of the TCA cycle through the anaplerotic malate-aspartate shuttle. In the presence of high glucose, the levels of most essential amino acids, urea and ornithine were down-regulated, indicating that proteolysis is reduced by the increase in glucose. Fatty acid metabolism was also affected in the different

toxicity models. Glycerol 3-phosphate increased in the glucotoxic and lipotoxic models but was unaltered in the glucolipotoxic model. The fatty acid levels increased in the lipotoxic and glucolipotoxic model, while the levels were generally decreased in the glucotoxic model.

Major conclusions

- Multivariate approaches in experimental planning and analysis, i.e. DOE in combination with PLS, constitute a highly efficient framework for developing new protocols for metabolomics.

- Metabolomics, combining high throughput analytical techniques with chemometrics, is a powerful approach to elucidate effects of perturbations on the metabolome.

- The application of metabolomics to elucidate effects of metabolic alterations is capable of being hypothesis-generating rather than being hypothesis-driven.

- Fatty acids are reliable markers of metabolic regulation in an OGTT and may thus be able to increase the precision of diagnostics for glucose tolerance and diabetes.

- It is unlikely that a single metabolic signal accounts for the amplifying pathway of insulin secretion. Instead, a pattern of metabolite levels from several metabolic pathways or a common denominator for several of these, is likely to serve as amplifiers in β-cell stimulus-secretion coupling.

- Differential regulation of GABA in glucotoxicity and lipotoxicity was found, while metabolites from glucose metabolism were less affected than expected.

References

1. Fiehn O. Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol. 2002;48(1):155-71.
2. Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C, Wold S. Multi- and Megavariate Data Analysis Part I - Basic Priciples and Applications. 2 ed. Umeå: Umetrics; 2006.
3. Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C, Wold S. Multi- and Megavariate Data Analysis Part II - Advanced Applications and Method Extensions. 2 ed. Umeå: Umetrics 2006.
4. Cai S, Huo T, Xu J, Lu X, Zheng S, Li F. Effect of mitiglinide on Streptozotocin-induced experimental type 2 diabetic rats: A urinary metabonomics study based on ultra-performance liquid chromatography-tandem mass spectrometry. J Chromatogr B. 2009;877(29):3619-24.
5. Madsen R, Lundstedt T, Trygg J. Chemometrics in metabolomics - A review in human disease diagnosis. Anal Chim Acta.In Press, Accepted Manuscript.
6. Sussulini A, Prando A, Maretto DA, Poppi RJ, Tasic L, Banzato CuEM, et al. Metabolic Profiling of Human Blood Serum from Treated Patients with Bipolar Disorder Employing 1H NMR Spectroscopy and Chemometrics. Anal Chem. 2009 10/28/;81(23):9755-63.
7. Ducruix C, Vailhen D, Werner E, Fievet JB, Bourguignon J, Tabet J-C, et al. Metabolomic investigation of the response of the model plant Arabidopsis thaliana to cadmium exposure: Evaluation of data pretreatment methods for further statistical analyses. Chemom Intell Lab Syst. 2008;91(1):67-77.
8. Issaq HJ, Blonder J. Electrophoresis and liquid chromatography/tandem mass spectrometry in disease biomarker discovery. J Chromatogr B. 2009;877(13):1222-8.
9. Zhang S, Nagana Gowda GA, Asiago V, Shanaiah N, Barbas C, Raftery D. Correlative and quantitative 1H NMR-based metabolomics reveals specific metabolic pathway disturbances in diabetic rats. Anal Biochem. 2008;383(1):76-84.
10. Dettmer K, Aronov P, A. , Hammock B, D. . Mass spectrometry-based metabolomics. Mass Spectrom Rev. 2007;26(1):51-78.
11. Oresic M. Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction. Nutrition, Metabolism and Cardiovascular Diseases. 2009;19(11):816-24.
12. A J, Trygg J, Gullberg J, Johansson AI, Jonsson P, Antti H, et al. Extraction and GC/MS Analysis of the Human Blood Plasma Metabolome. Anal Chem. 2005 11/08/;77(24):8086-94.
13. Rafii M, Elango R, House JD, Courtney-Martin G, Darling P, Fisher L, et al. Measurement of homocysteine and related metabolites in human plasma

and urine by liquid chromatography electrospray tandem mass spectrometry. J Chromatogr B. 2009;877(28):3282-91.
14. Koek MM, Bakels F, Engel W, van den Maagdenberg A, Ferrari MD, Coulier L, et al. Metabolic Profiling of Ultrasmall Sample Volumes with GC/MS: From Microliter to Nanoliter Samples. Anal Chem. 2009 11/30/;82(1):156-62.
15. Croixmarie V, Umbdenstock T, Cloarec O, Moreau Al, Pascussi J-M, Boursier-Neyret C, et al. Integrated Comparison of Drug-Related and Drug-Induced Ultra Performance Liquid Chromatography/Mass Spectrometry Metabonomic Profiles Using Human Hepatocyte Cultures. Anal Chem. 2009 07/09/;81(15):6061-9.
16. Canelas AB, ten Pierick A, Ras C, Seifar RM, van Dam JC, van Gulik WM, et al. Quantitative Evaluation of Intracellular Metabolite Extraction Techniques for Yeast Metabolomics. Anal Chem. 2009 08/04/;81(17):7379-89.
17. Cozzolino D, Flood L, Bellon J, Gishen M, Lopes MDB. Combining near infrared spectroscopy and multivariate analysis as a tool to differentiate different strains of Saccharomyces cerevisiae: a metabolomic study. Yeast. 2006;23(14-15):1089-96.
18. Ewald JC, Heux Sp, Zamboni N. High-Throughput Quantitative Metabolomics: Workflow for Cultivation, Quenching, and Analysis of Yeast in a Multiwell Format. Anal Chem. 2009 03/25/;81(9):3623-9.
19. Villas-Bôas SG, Højer-Pedersen J, Åkesson M, Smedsgaard J, Nielsen J. Global metabolite analysis of yeast: evaluation of sample preparation methods. Yeast. 2005;22(14):1155-69.
20. Prasad Maharjan R, Ferenci T. Global metabolite analysis: the influence of extraction methodology on metabolome profiles of Escherichia coli. Anal Biochem. 2003;313(1):145-54.
21. Winder CL, Dunn WB, Schuler S, Broadhurst D, Jarvis R, Stephens GM, et al. Global Metabolic Profiling of Escherichia coli Cultures: an Evaluation of Methods for Quenching and Extraction of Intracellular Metabolites. Anal Chem. 2008 03/11/;80(8):2939-48.
22. Gullberg J, Jonsson P, Nordström A, Sjöström M, Moritz T. Design of experiments: an efficient strategy to identify factors influencing extraction and derivatization of Arabidopsis thaliana samples in metabolomic studies with gas chromatography/mass spectrometry. Anal Biochem. 2004;331(2):283-95.
23. Parab GS, Rao R, Lakshminarayanan S, Bing YV, Moochhala SM, Swarup S. Data-Driven Optimization of Metabolomics Methods Using Rat Liver Samples. Anal Chem. 2009 01/13/;81(4):1315-23.
24. Rammouz RE, Létisse F, Durand S, Portais J-C, Moussa ZW, Fernandez X. Analysis of skeletal muscle metabolome: Evaluation of extraction methods for targeted metabolite quantification using liquid chromatography tandem mass spectrometry. Anal Biochem.In Press, Corrected Proof.

25. Ying Z, Jiye A, Wang G, Qing H, Bei Y, Weibin Z, et al. Organic solvent extraction and metabonomic profiling of the metabolites in erythrocytes. J Chromatogr B. 2009;877(18-19):1751-7.
26. Eriksson L, Johansson E, Kettaneh-Wold N, Wikström C, Wold S. Design of Experiments - Principles and Applications. 3 ed. Umeå: Umetrics; 2008.
27. Brereton R, G. Chemometrics: Data Analysis for the Laboratory and Chemical Plant. 1 ed: John Wiley & Sons, Ltd.; 2003.
28. Berg J, M., Tymoczko J, L., Stryer L. Biochemistry. 6 ed. New York: W.H. Freeman and Company; 2007.
29. Henquin JC. Triggering and amplifying pathways of regulation of insulin secretion by glucose. Diabetes. 2000 November 2000;49(11):1751-60.
30. Maechler P. Mitochondria as the conductor of metabolic signals for insulin exocytosis in pancreatic beta-cells. Cell Mol Life Sci. 2002;59(11):1803-18.
31. Florez JC. Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: Where are the insulin resistance genes? Diabetologia. 2008;51(7):1100-10.
32. Unger RH. Lipotoxicity in the pathogenesis of obesity-dependent NIDDM. Genetic and clinical implications. Diabetes. 1995 August 1995;44(8):863-70.
33. Robertson RP, Harmon J, Tran POT, Poitout V. β-Cell Glucose Toxicity, Lipotoxicity, and Chronic Oxidative Stress in Type 2 Diabetes. Diabetes. 2004 February 2004;53(suppl 1):S119-S24.
34. Sako Y, Grill V, E. A 48-hour Lipid Infusion in the Rat Time-Dependently Inhibits Glucose-Induced Insulin Secretion and B Cell Oxidation Through a Process Likely Coupled to Fatty Acid Oxidation. Endocrinology. 1990 October 1, 1990;127(4):1580-9.
35. Marchetti P, Del Prato S, Lupi R, Del Guerra S. The pancreatic beta-cell in human Type 2 diabetes. Nutrition, Metabolism and Cardiovascular Diseases. 2006;16(-Supplement_1):S3-S6.
36. Shimabukuro M, Higa M, Zhou Y-T, Wang M-Y, Newgard CB, Unger RH. Lipoapoptosis in Beta-cells of Obese Prediabeticfa/fa Rats. J Biol Chem. 1998 December 4, 1998;273(49):32487-90.
37. Poitout V, Robertson RP. Glucolipotoxicity: Fuel Excess and beta-Cell Dysfunction. Endocrine reviews : issued quarterly for the Endocrine Society. 2008;29(3):351-66.
38. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends Biotechnol. 2004;22(5):245-52.
39. Büscher JM, Czernik D, Ewald JC, Sauer U, Zamboni N. Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism. Anal Chem. 2009 02/23/;81(6):2135-43.
40. Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L. Metabolite profiling for plant functional genomics. Nat Biotech. 2000;18(11):1157-61.
41. Fiehn O, Kopka J, Trethewey RN, Willmitzer L. Identification of Uncommon Plant Metabolites Based on Calculation of Elemental Compositions

Using Gas Chromatography and Quadrupole Mass Spectrometry. Anal Chem. 2000 07/07/;72(15):3573-80.
42. Tanaka K, Hine D, West-Dull A, Lynn T. Gas-chromatographic method of analysis for urinary organic acids. I. Retention indices of 155 metabolically important compounds. Clin Chem. 1980 December 1, 1980;26(13):1839-46.
43. Mohamed R, Varesio E, Ivosev G, Burton L, Bonner R, Hopfgartner Gr. Comprehensive Analytical Strategy for Biomarker Identification based on Liquid Chromatography Coupled to Mass Spectrometry and New Candidate Confirmation Tools. Anal Chem. 2009 08/24/;81(18):7677-94.
44. Zelena E, Dunn WB, Broadhurst D, Francis-McIntyre S, Carroll KM, Begley P, et al. Development of a Robust and Repeatable UPLC-MS Method for the Long-Term Metabolomic Study of Human Serum. Anal Chem. 2009 01/26/;81(4):1357-64.
45. Bruce SJ, Jonsson P, Antti H, Cloarec O, Trygg J, Marklund S, L., et al. Evaluation of a protocol for metabolic profiling studies on human blood plasma by combined ultra-performance liquid chromatography/mass spectrometry: From extraction to data analysis. Anal Biochem. 2008;372(2):237-50.
46. Bruce SJ, Tavazzi I, Parisod Vr, Rezzi S, Kochhar S, Guy PA. Investigation of Human Blood Plasma Sample Preparation for Performing Metabolomics Using Ultrahigh Performance Liquid Chromatography/Mass Spectrometry. Anal Chem. 2009 03/26/;81(9):3285-96.
47. Chalcraft KR, Lee R, Mills C, Britz-McKibbin P. Virtual Quantification of Metabolites by Capillary Electrophoresis-Electrospray Ionization-Mass Spectrometry: Predicting Ionization Efficiency Without Chemical Standards. Anal Chem. 2009 03/10/;81(7):2506-15.
48. Lapainis T, Rubakhin SS, Sweedler JV. Capillary Electrophoresis with Electrospray Ionization Mass Spectrometric Detection for Single-Cell Metabolomics. Anal Chem. 2009 06/11/;81(14):5858-64.
49. Almstetter MF, Appel IJ, Gruber MA, Lottaz C, Timischl B, Spang R, et al. Integrative Normalization and Comparative Analysis for Metabolic Fingerprinting by Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry. Anal Chem. 2009 06/12/;81(14):5731-9.
50. Pasikanti KK, Ho PC, Chan ECY. Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. J Chromatogr B. 2008;871(2):202-11.
51. Brown S, C. , Kruppa G, Dasseux J-L. Metabolomics applications of FT-ICR mass spectrometry. Mass Spectrom Rev. 2005;24(2):223-31.
52. Qizhi H, Robert JN, Hongyan L, Alexander M, Mark H, Cooks RG. The Orbitrap: a new mass spectrometer. J Mass Spectrom. 2005;40(4):430-43.
53. Giavalisco P, Hummel J, Lisec J, Inostroza AC, Catchpole G, Willmitzer L. High-Resolution Direct Infusion-Based Mass Spectrometry in Combination with Whole 13C Metabolome Isotope Labeling Allows Unambiguous

Assignment of Chemical Sum Formulas. Anal Chem. 2008 11/17/;80(24):9417-25.
54. Miura D, Fujimura Y, Tachibana H, Wariishi H. Highly Sensitive Matrix-Assisted Laser Desorption Ionization-Mass Spectrometry for High-Throughput Metabolic Profiling. Anal Chem. 2009 12/16/;82(2):498-504.
55. Annesley TM. Ion Suppression in Mass Spectrometry. Clin Chem. 2003 7/1;49(7):1041-4.
56. Al Zweiri M, Sills GJ, Leach JP, Brodie MJ, Robertson C, Watson DG, et al. Response to drug treatment in newly diagnosed epilepsy: A pilot study of 1H NMR- and MS-based metabonomic analysis. Epilepsy Research.In Press, Corrected Proof.
57. Rai RK, Tripathi P, Sinha N. Quantification of Metabolites from Two-Dimensional Nuclear Magnetic Resonance Spectroscopy: Application to Human Urine Samples. Anal Chem. 2009 11/17/.
58. Zhang S, Zheng C, Lanza IR, Nair KS, Raftery D, Vitek O. Interdependence of Signal Processing and Analysis of Urine 1H NMR Spectra for Metabolic Profiling. Anal Chem. 2009 07/06/;81(15):6080-8.
59. Bolten CJ, Kiefer P, Letisse F, Portais J-C, Wittmann C. Sampling for Metabolome Analysis of Microorganisms. Anal Chem. 2007 04/06/;79(10):3843-9.
60. Koning Wd, Dam Kv. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. Anal Biochem. 1992;204(1):118-23.
61. Birkemeyer C, Kolasa A, Kopka J. Comprehensive chemical derivatization for gas chromatography-mass spectrometry-based multi-targeted profiling of the major phytohormones. J Chromatogr A. 2003;993(1-2):89-102.
62. Yuan K, Kong H, Guan Y, Yang J, Xu G. A GC-based metabonomics investigation of type 2 diabetes by organic acids metabolic profile. Journal of Chromatography, B. 2007;850(1-2):236-40.
63. Ohie T, Fu X-w, Iga M, Kimura M, Yamaguchi S. Gas chromatography-mass spectrometry with tert.-butyldimethylsilyl derivatization: use of the simplified sample preparations and the automated data system to screen for organic acidemias. J Chromatogr B. 2000;746(1):63-73.
64. Yu Z, Peldszus S, Huck PM. Optimizing gas chromatographic-mass spectrometric analysis of selected pharmaceuticals and endocrine-disrupting substances in water using factorial experimental design. J Chromatogr A. 2007;1148(1):65-77.
65. Tam YY, Normanly J. Determination of indole-3-pyruvic acid levels in Arabidopsis thaliana by gas chromatography-selected ion monitoring-mass spectrometry. J Chromatogr A. 1998;800(1):101-8.
66. Araujo PW, Brereton RG. Experimental design II. Optimization. TrAC, Trends Anal Chem. 1996;15(2):63-70.
67. Araujo PW, Brereton RG. Experimental design I. Screening. TrAC, Trends Anal Chem. 1996;15:26-31.
68. Araujo PW, Brereton RG. Experimental design III. Quantification. TrAC, Trends Anal Chem. 1996;15(3):156-63.

69. Lundstedt T, Seifert E, Abramo L, Thelin B, Nyström Å, Pettersen J, et al. Experimental design and optimization. Chemom Intell Lab Syst. 1998;42(1-2):3-40.
70. de Aguiar PF, Bourguignon B, Khots MS, Massart DL, Phan-Than-Luu R. D-optimal designs. Chemom Intell Lab Syst. 1995;30(2):199-210.
71. van den Berg R, Hoefsloot H, Westerhuis J, Smilde A, van der Werf M. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics. 2006;7(1):142.
72. Jonsson P, Johansson AI, Gullberg J, Trygg J, A J, Grung B, et al. High-Throughput Data Analysis for Detecting and Identifying Differences between Samples in GC/MS-Based Metabolomic Analyses. Anal Chem. 2005 08/04/;77(17):5635-42.
73. Jonsson P, Gullberg J, Nordstrom A, Kusano M, Kowalczyk M, Sjostrom M, et al. A Strategy for Identifying Differences in Large Series of Metabolomic Samples Analyzed by GC/MS. Anal Chem. 2004 02/11/;76(6):1738-45.
74. Karjalainen EJ. The spectrum reconstruction problem - Use of alternating regression for unexpected spectral components in two-dimensional spectroscopies. Chemom Intell Lab Syst. 1989;7(1-2):31-8.
75. Lu H, Liang Y, Dunn WB, Shen H, Kell DB. Comparative evaluation of software for deconvolution of metabolomics data based on GC-TOF-MS. TrAC, Trends Anal Chem. 2008;27(3):215-27.
76. Redestig H, Fukushima A, Stenlund H, Moritz T, Arita M, Saito K, et al. Compensation for Systematic Cross-Contribution Improves Normalization of Mass Spectrometry Based Metabolomics Data. Anal Chem. 2009 09/10/;81(19):7974-80.
77. Chorell E, Moritz T, Branth S, Antti H, Svensson MB. Predictive Metabolomics Evaluation of Nutrition-Modulated Metabolic Stress Responses in Human Blood Serum During the Early Recovery Phase of Strenuous Physical Exercise. J Proteome Res. 2009 03/25/;8(6):2966-77.
78. Marshall A, G., Hendrickson C, L. , Jackson G, S. . Fourier transform ion cyclotron resonance mass spectrometry: A primer. Mass Spectrom Rev. 1998;17(1):1-35.
79. Leon C, Rodriguez-Meizoso I, Lucio M, Garcia-Cañas V, Ibañez E, Schmitt-Kopplin P, et al. Metabolomics of transgenic maize combining Fourier transform-ion cyclotron resonance-mass spectrometry, capillary electrophoresis-mass spectrometry and pressurized liquid extraction. J Chromatogr A. 2009;1216(43):7314-23.
80. Craig A, Cloarec O, Holmes E, Nicholson JK, Lindon JC. Scaling and Normalization Effects in NMR Spectroscopic Metabonomic Data Sets. Anal Chem. 2006 02/18/;78(7):2262-7.
81. Keun HC, Ebbels TMD, Antti H, Bollard ME, Beckonert O, Holmes E, et al. Improved analysis of multivariate data by variable stability scaling: application to NMR-based metabolic profiling. Anal Chim Acta. 2003;490(1-2):265-76.

82. Pearson K. On lines and planes of closest fit to systems of points in space. Philosophical Magazine Series 6. 1901;2(11):559 - 72.
83. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst. 2001;58(2):109-30.
84. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). J Chemom. 2002;16(3):119-28.
85. Wold H. Nonlinear estimation by iterative least squares procedures. David F, editor. New York: Wiley; 1996.
86. Wold S. Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. Technometrics. 1978;20(4):397-405.