Tools and Applications for One- and Two-dimensional Gas Chromatography – Time-of-Flight Mass Spectrometry-based Metabolomics

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

an der naturwissenschaftlichen Fakultät IV

-Chemie und Pharmazie-

der Universität Regensburg

vorgelegt von

Martin F. Almstetter

aus Ingolstadt

2011

Diese Doktorarbeit entstand in der Zeit von August 2007 bis November 2011 am Institut für Funktionelle Genomik der Universität Regensburg.

Die Arbeit wurde angeleitet von Prof. Dr. Peter J. Oefner.

Promotionsgesuch eingereicht am 07. Dezember 2011.

Für meine Eltern

Danksagung

Nun ist es endlich soweit und ich freue mich sehr, an dieser Stelle einigen Leuten danken zu können, die zum Gelingen dieser Arbeit beigetragen haben.

In erster Linie möchte ich mich bei **Prof. Peter Oefner** bedanken, der es mir ermöglichte, am Institut für Funktionelle Genomik zu promovieren. Vielen Dank für die hervorragenden Voraussetzungen am Institut, die Vergabe des interessanten Themas, das stetige Interesse an meiner Arbeit und dass ich auf internationalen Tagungen präsentieren durfte.

Ein Dank gebührt auch **Prof. Frank-Michael Matysik**, der sich als Stellvertreter der Fakultät für Chemie bereiterklärt hat, das Erstgutachten zu übernehmen.

Der größte Dank gilt meiner direkten Betreuerin, der Leiterin der Metabolomics-Gruppe und Labor-Mutti, **Dr. Katja Dettmer-Wilde**. Deine ausgezeichnete fachliche Betreuung und dein Engagement waren essenziell für den Erfolg dieser Arbeit. Vielen Dank für die vielen hilfreichen Ratschläge, die Weitergabe deines ausgeprägten analytisch-chemischen Wissens und den Grundsatz, dass wissenschaftliches Arbeiten immer auch Spaß machen muss.

Ganz herzlich möchte ich natürlich auch der treuen Laborkollegin zu meiner Rechten, **Nadine Nürnberger**, danken. Deine fröhliche Art und dein auflockerndes Wesen haben so manch tristen Laboralltag erhellt. Ohne deinen ständigen Support im Labor hätte ich einige Experimente (u.a. den *Coli*-Dance) nicht bewältigen können.

Ein aufrichtiges „Mahlzeit" an **Dr. Michael (Onkel) Gruber** für deine ansteckende gute Laune, deine Motivation und nicht zuletzt für deine fachliche und finanzielle Unterstützung.

Danke **Inka Appel** für INCA, dein bioinformatisches Fachwissen und deine Bereitschaft zur Kooperation zwischen „Nerds" und „Labrats".

Dziękuje, merci und xiéxie an meine Kollegen der Metabolomics-Gruppe **Magdalena (Pietruszka) Waldhier**, **Christian Wachsmuth** und **Dr. Wentao Zhu** für die angenehme Zusammenarbeit, den Zusammenhalt, den regen Ideenaustausch und die Förderung meiner Sprachkenntnisse. Ich werde die schöne Zeit im Metabolomics-Team nicht vergessen.

1 Table of Contents

2 Abbreviations and acronyms

(v/v)	volume/volume percent
(w/v)	weight/volume percent
1D	1st dimension/one-dimensional
2D	2nd dimension/two-dimensional
5-HIAA	5-hydroxyindolacetic acid
AA	amino acid
Ala	alanine
APCI	atmospheric pressure chemical ionization
API	atmospheric pressure ionization
ASA	acetylsalicylic acid
Asn	asparagine
Asp	aspartic acid
AUC	area under the curve
biol. tripl.	biological triplicate
calib. tripl.	calibration triplicate
CD	cyclodextrin
CDW	cell dry weight
CE	capillary electrophoresis
corr.	corrected
COW	correlation optimized warping
csv	character separated value
CVal	cross-validation
Cys	cysteine
đ	mean difference
DDA	dodecanoic acid
ECD	electron capture detector
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EI	electron ionization
FAME	fatty acid methyl ester
FC	fold change
FDA	Food and Drug Administration

FDR	false discovery rate
FID	flame ionization detector
FPR	false positive rate
GABA	γ-aminobutyric acid
GC	gas chromatography
GC×GC	comprehensive two-dimensional gas chromatography
Gln	glutamine
Glu	glutamic acid
Glucose-6P	glucose-6-phosphate
Gly	glycine
GUI	graphical user interface
HCA	hierarchical cluster analysis
HepDA	heptadecanoic acid
HexDA	hexadecanoic acid
HPLC	high-performance liquid chromatography
ID	internal diameter
Ile	isoleucine
IS	internal standard
LB	Luria-Bertani
LC	liquid chromatography
Leu	leucine
LLOQ	lower limit of quantification
LMCS	longitudinal modulating cryogenic system
LOD	limit of detection
LOQ	limit of quantification
LR	linear range
Lys	lysine
m/z	mass-to-charge ratio
MCF	methyl chloroformate
Met	methionine
MS	mass spectrometry/mass spectrometer
MS/MS	tandem quadrupole mass spectrometry
MPS	multipurpose sampler
MSTFA	*N*-methyl-trimethylsilyltrifluoroacetamide

NA	nonanoic acid
NAD$^+$/NADH	nicotinamide adenine dinucleotide (oxidized and reduced form)
NADP$^+$/NADPH	nicotinamide adenine dinucleotide phosphate (oxidized and reduced form)
NDA	nonadecanoic acid
NIST	National Institute of Standards and Technology
NMR	nuclear magnetic resonance
NPD	nitrogen phosphorus detector
OD	optical density
ODA	octadecanoic acid
PARAFAC	parallel factor analysis
PCA	principal component analysis
PCDA	principal component discriminant analysis
PCF	propyl chloroformate
PDA	pentadecanoic acid
Phe	phenylalanine
PLS-DA	partial least squares - discriminant analysis
Pro	proline
PTV	programmed-temperature vaporization
QC	quality control
qMS	quadrupole mass spectrometry
R	correlation coefficient
R^2	square of the linear regression coefficient R
ROC	receiver operator characteristic
RSD	relative standard deviation
RT	retention time
S/N	signal-to-noise ratio
SC	Statistical Compare
SD	standard deviation
Ser	serine
SIM	selected ion monitoring
SPME	solid-phase microextraction
TCA	tricarboxylic acid cycle/citric acid cycle
TDA	tridecanoic acid

TE	technical error
TeDA	tetradecanoic acid
Thr	threonine
TIC	total ion chromatogram
TMS	trimethylsilyl
TOF	time-of-flight
TPR	true positive rate
Tyr	tyrosine
U	unique
U-^{13}C	uniform ^{13}C labeled
UDA	undecanoic acid
ULOQ	upper limit of quantification
Val	valine

3 Motivation

The metabolome represents the quantitative complement of low molecular-weight metabolites present in a cell or body fluid under certain physiological conditions. It reflects the cellular processes and thus directly the biochemical phenotype of a living system. Metabolomics is defined as the systematic study of metabolites and their response to environmental, nutritional, genetic, and pathophysiological influences. The ultimate objective is the quantitative analysis of the entire metabolome in a single run. Currently, not a single method, but rather a combination of analytical techniques is required to accomplish this challenging task because of the greatly differing physicochemical properties of the hundreds to thousands of metabolites present in biological systems.[1, 2]

Metabolomics data is typically tackled by employing either metabolic fingerprinting or metabolite profiling approaches. Metabolic fingerprinting assembles all the analytical information gathered from a sample and thus provides a snapshot of metabolism at a given state, while metabolite profiling focuses on the quantitative analysis of a metabolite class or metabolites associated with a selected biochemical pathway.[2]

Gas chromatography coupled to mass spectrometry (GC-MS) has become a common tool for metabolomic investigations.[3] Despite its excellent chromatographic resolution, one-dimensional (1D) GC cannot resolve the multitude of metabolites present in physiological fluids or tissue/cell extracts. Comprehensive two-dimensional gas chromatography (GC×GC) uses a thermal modulator incorporated between two columns with orthogonal separation characteristics and thus enables a multiplicative increase in peak capacity, enhanced resolution, lower limits of detection (LODs), and a structured separation space.[4] Coupled to an electron ionization (EI) fast acquisition time-of-flight mass spectrometer (TOFMS) for the identification and quantification of analytes, GC×GC is perfectly suited for metabolic fingerprinting.

Aim #1: Development of a data processing strategy for comparative metabolic fingerprinting by GC×GC-TOFMS

Setup and validation

The initial goal of this present work was the development and validation of an algorithm for the processing of GC×GC-TOFMS metabolic fingerprinting data for the

comparative characterization of multiple samples and their constituents. To this end, initial data processing functions provided by the LECO ChromaTOF software, such as baseline correction, spectrum deconvolution, and peak detection, combine, integration, and identification, were adopted to generate a peak list for every sample analyzed. For a comparative analysis, identical peaks had to be recognized across all peak lists. Due to a lack of efficient solutions for data alignment, the Integrative Normalization and Comparative Alignment (INCA) module was developed that in a first step compensated for run-to-run retention time (RT) fluctuations before aligning the entire data from all peak lists into one final data matrix according to 1st dimension (1D) and 2nd dimension (2D) retention times as well as mass spectral information. The area integral of the *m/z* 73 ion trace of the trimethylsilyl (TMS) group was employed as a quantitative measure originating from the characteristic fragmentation behavior of silylated metabolites upon EI ionization. Statistical classification and testing strategies were used to visualize the data and to determine differences in metabolite abundance, respectively. The accuracy of peak detection and alignment of 1.1- to 4-fold changes in metabolite concentration was validated by a spike-in experiment.

Application

The GC×GC-TOFMS comparative metabolic fingerprinting strategy was applied to the metabolome analysis of two *Escherichia coli* (*E. coli*) strains which vary in glycolytic flux distribution. That the signal intensity of the *m/z* 73 ion reflected true differences in metabolite abundance was verified by absolute quantification of class-distinguishing metabolites using compound-specific fragment ions and stable isotope-labeled standards in a separate metabolite profiling approach.

The same procedure was utilized in 1D mode (GC-TOFMS) to determine differences between blood serum and plasma.

Comparison

Data acquired for the spike-in experiment and the comparative metabolic fingerprinting of *E. coli* strains were used to directly compare the performance of INCA versus the Statistical Compare (SC) alignment tool, which recently became commercially available as an add-on option for ChromaTOF version 4. Advantages and limitations of the two algorithmic data processing strategies were assessed.

Aim #2: Performance evaluation and application of GC×GC-TOFMS metabolite profiling

Another objective was the comprehensive evaluation of GC×GC-TOFMS in comparison to other GC-MS techniques for metabolite analysis. For this purpose, GC×GC-TOFMS was evaluated against GC-APCI-TOFMS, GC-EI-TOFMS, GC-CI-qMS, and GC-EI-qMS in terms of reproducibility, dynamic range, and limits of detection and quantification using a mix of 43 metabolites from different chemical classes and metabolic pathways and 12 stable isotope-labeled standards.

Further, a GC×GC-TOFMS metabolite profiling approach was established for the analysis of amino acid enantiomers (AAEs) as their methanol/methyl chloroformate derivatives (MeOH/MCF). The goal was to improve enantiomer resolution and quantification compared to a previously published 1D-GC-qMS approach with single ion monitoring (SIM) [5] by using a combination of a γ-cyclodextrin (CD) chiral 1D column (Rt-γDEXsa) and two different 2D columns. The method was applied to the analysis of sera of patients suffering from liver cirrhosis.

4 Background

An abbreviated version of this chapter was published in Analytical Bioanalytical Chemistry.[6] Parts were also published in book chapters [7] and [8].

4.1 Metabolomics

Metabolomics or metabonomics, both terms are nowadays used interchangeably [9], is an intrinsic part of systems biology, which focuses on the study of molecular components (genes, transcripts, proteins and metabolites) in a biological system and the complex interactions of these constituents. The transcriptome, proteome, and metabolome all reflect the physiological status of a cell, tissue, or organism and their response differs upon both genetic and environmental (age, diet, disease, drug, lifestyle) influences (Figure 4.1).
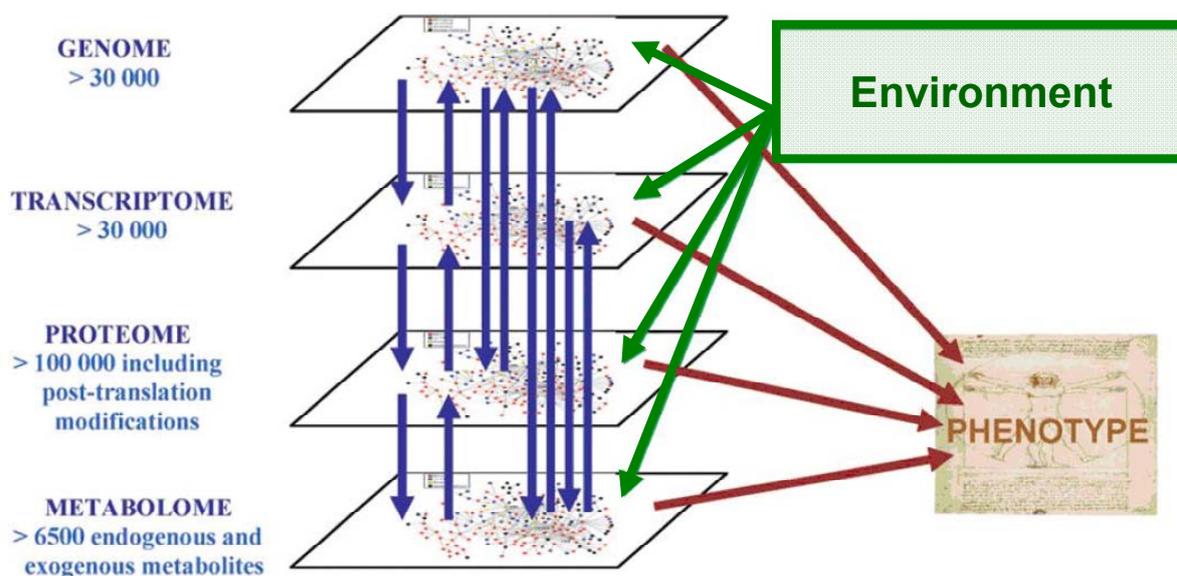
Figure 4.1. Complex interactions of constituents from all functional levels and the environment generate the phenotype. Based on its downstream position the metabolome represents the most immediate measure of the biological phenotype. Modified from [10].

Alterations at the metabolite level (concentration and flux) tend to be more pronounced than at the transcript or protein level as the metabolome is positioned farthest downstream of gene expression and closest to the biological phenotype. Additionally, primary metabolism is a highly dynamic system. Opposed to turnover in the proteome and transcriptome (minutes to hours), metabolic flux operates on time

scales of seconds. The study of the metabolome, whether applied individually or in a more integrated way, is a sensitive and rapid phenotypic measure for detecting altered physiological states.[11-14]

4.1.1 The fundamentals of metabolomics

Metabolomics is the systemic study of metabolite profiles, their composition and dynamics as influenced by genetic modifications, physiological stimuli, environmental, nutritional or other factors. Metabolites are low molecular-weight organic (or inorganic [15, 16]) compounds (< 1,500 Da) that are formed by the numerous biosynthetic and catabolic pathways within a biological system (endogenous) or originate from host-specific microbes and the consumption of food nutrients and pharmaceuticals (exogenous). The ultimate ambition in metabolomics studies is the global detection and quantification of all metabolites in a single analysis. Although the number of metabolites in most organisms is estimated to be lower than that of genes and proteins, a holistic analysis by a single analytical platform may remain technologically infeasible due to the great chemical diversity of metabolites present over a wide concentration range.[17-20]

Compared to the chemically very similar nucleotides in genomics and transcriptomics and amino acids in proteomics, the metabolome consists of various substance classes, including alcohols, ketones, carbohydrates, amino and organic acids, lipids, nucleotides, polyols, and others. Metabolites can therefore vastly differ in molecular weight, size, charge, volatility, acidity, polarity or hydrophobicity. Due to this broad chemical space, special branches, such as lipidomics and glycomics, have developed within the field of metabolomics.[21-23]

The number of metabolites is organism and sample type dependent. While prokaryotes, like *E. coli*, hold approximately 750 metabolites [24], the number for eukaryotic systems can range from around 1100 in yeast [25] to many thousands in humans [26, 27] and up to tens or hundreds of thousands in plants [28, 29] and fungi [30]. In higher organisms multiple metabolomes (specific cell, tissue, biofluid, and microbiomes of the gastrointestinal tract) coexist that vary distinctly in number and types of metabolites present. To date metabolic databases only contain a fraction of all metabolites, lacking many lipids, xenobiotics, their metabolites, and metabolites emerging from interacting metabolomes.[31-33]

4.1.2 Strategies in metabolomics studies

Albeit the terminology differs among related publications, commonly two experimental strategies are applied to metabolomic research, targeted (metabolite profiling and metabolite target analysis) and untargeted studies (metabolic fingerprinting and metabolic footprinting). Examples are shown in Figure 4.2.
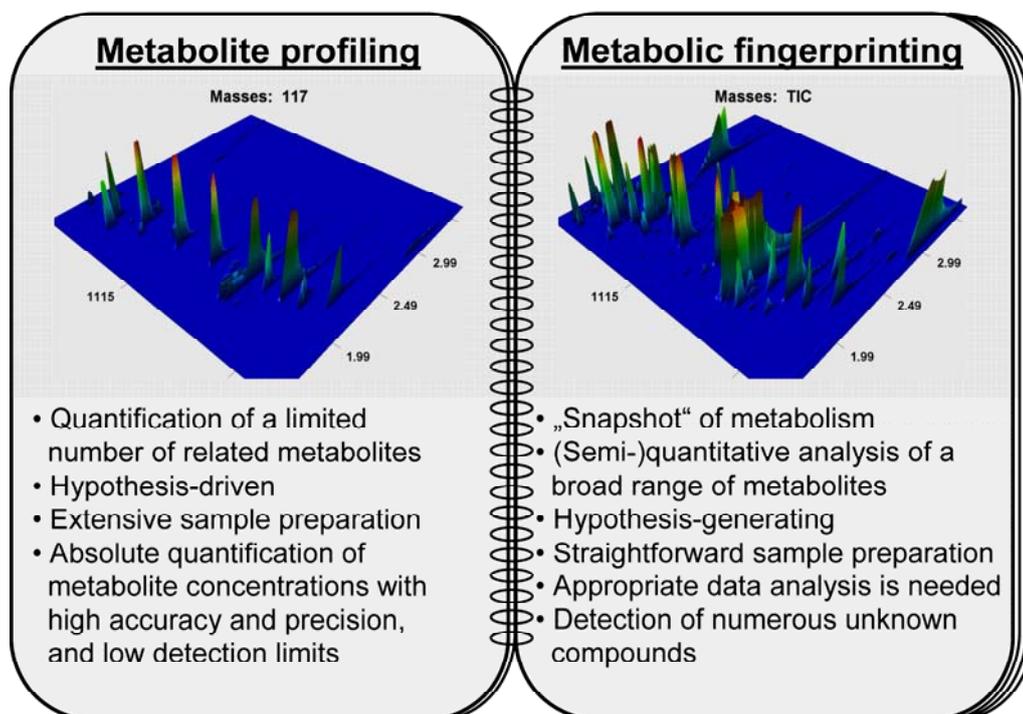
Figure 4.2. The most common strategies in metabolomics research. Reproduced from [6].

Metabolite profiling focuses on the analysis of a limited number of (generally known) metabolites associated with a selected biochemical pathway (e.g. TCA cycle) or a specific class of compounds (e.g. amino acids, fatty acids).[34] Metabolite target analysis is a more specific approach, which concentrates on selected analytes only, e.g. biomarkers of disease. Both strategies are classically hypothesis-driven and mostly provide absolute quantification of metabolite concentrations with high accuracy and precision, and low detection limits. On the other hand, metabolic fingerprinting is a non-biased hypothesis-generating approach with the experimental goal to gather analytical information on an extensive range of intra-cellular metabolites directly yielding a snapshot of metabolism at a given time.[35, 36] The metabolic fingerprints are screened comprehensively for signals that distinguish populations. Such global studies are typically carried out on analytical platforms with fast acquisition rates. Due to extremely large datasets chemometric or comparative

visualization techniques are often required. The equivalent to metabolic fingerprinting at the extracellular level is called metabolic footprinting.[37, 38] In contrast to targeted analyses the simultaneous detection of hundreds to thousands of metabolites in a single analysis is usually accompanied by limited metabolite information (lots of unknowns are detected) and a lack of absolute quantitative data. A targeted follow-up study with a directed sample preparation is recommended to verify interesting identifications.

4.2 Experimental preparation

A metabolomics study is divided into sample collection and preparation, data acquisition and processing, bioinformatic analysis, and interpretation. In addition, method optimization and validation, and an appropriate quality control play an essential role. Various parameters determine a well-thought-out experimental design. Depending on the biological system that is being studied an adequate number of samples per group need to be obtained. The more complex the metabolome the more samples are recommended. However, sample sizes can be reduced if certain factors (age, gender, etc.) are controlled or a time-series experiment on the same individuals is performed. During sample preparation artificially introduced variability has to be kept at a minimum to avoid compositional changes of the metabolome. That is usually facilitated by standardized quenching, extraction, and derivatization protocols.

4.2.1 Quenching and extraction

Quenching, the process of decreasing or completely inhibiting enzymatic activity, can be accomplished by freezing in liquid nitrogen, acid treatment or by using cold-buffered methanol solutions.[39, 40] The subsequent extraction of metabolites can vary according to the experimental strategy. More extensive extraction techniques are typically performed for targeted analyses, whereas extraction in untargeted studies is optimized towards detecting as many metabolites as possible.

Solvent and solid-phase extractions are established extraction techniques, but some sample types demand additional procedures. Tissue specimens (brain, kidney, etc.) require homogenization and mechanical or chemical lysation of cell walls followed by liquid–liquid extraction to release and extract intracellular metabolites.[41, 42] The high

protein content of serum and plasma specimens requires the precipitation of proteins during the extraction process.[43, 44] Urine may require urease treatment as high concentrations of urea are unfavorable especially for GC-MS analyses.[45] Variable fluid intake affects the comparison of urinary metabolite patterns. Adjusting sample volumes, e.g. based on the creatinine concentration, is advisable.

4.2.2 Derivatization

Only few metabolites are naturally volatile (e.g. alcohols, esters, monoterpenes). The majority lacks sufficient volatility and, therefore, requires chemical derivatization to expand the capacity of metabolites amenable to GC-based methods. There are many derivatization techniques that have been tailored for particular metabolite classes, such as alkyl chloroformates for amino/organic acids [46, 47] or esterification for fatty acids [48]. A frequently-used derivatization method is silylation [49] as it modifies a diversity of functional groups (hydroxyl, carboxylic acid, amine, amide, imine, thiol, phosphate) and, hence, is perfectly suited for untargeted approaches. The derivatives generally become less polar, more volatile, and thermally more stable. Two-step derivatization, which involves (m)ethoximation forming stable derivatives with carbonyl moieties, followed by trimethylsilylation, is a highly versatile procedure [50] that is routinely used in GC×GC based metabolomics studies (Figure 4.3).

1. **Methoximation**
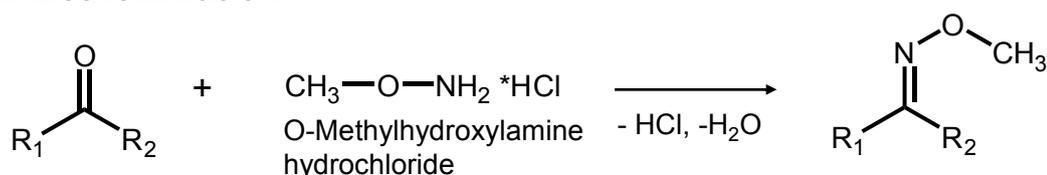
2. **Silylation**
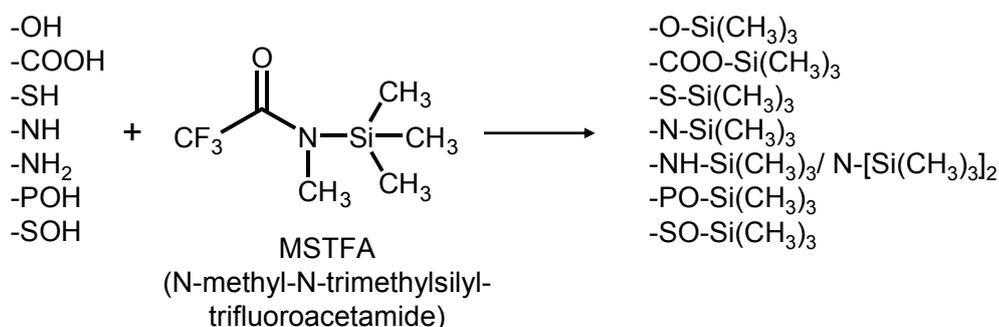
Figure 4.3. Example of a two-step derivatization procedure (primarily used in this work). Reproduced from [7].

After derivatization a single metabolite can be represented by more than one product (feature) and vice versa. While E-Z isomerism of methoxyamines can form two signals, partial silylation can form numerous ones for certain metabolites, e.g. amino acids. In addition, steric hindrance, degradation or rearrangement reactions can take place.[49, 51]

4.3 Analytical techniques for metabolome analysis

An assortment of analytical tools has been used to cope with the complex information space in metabolomics research. Nuclear Magnetic Resonance (NMR) spectroscopy [52, 53], GC-MS [54] and liquid chromatography (LC) – MS [55] are the major techniques in use today, but many others have been applied including capillary electrophoresis (CE) – MS [56] and infrared/Raman spectroscopy.[57] All of these methods can be applied to broad metabolite classes, but each one has its distinct advantages and disadvantages regarding detection limits, throughput, sensitivity, interferences, and the kind of information that is provided. In the end, maximum coverage of the metabolome can only be obtained by a combination of analytical platforms (and sample preparation techniques).[58, 59]

This chapter highlights the analytical techniques that were mainly employed in the course of this thesis.

4.3.1 Comprehensive two-dimensional gas chromatography – time-of-flight mass spectrometry

Gas chromatography is an excellent technique for the qualitative and quantitative analysis of volatile compounds or analytes that can be derivatized to obtain a sufficient vapor pressure. The one-dimensional separation process provides adequate resolution for various analytical tasks. However, conventional 1D-GC does not suffice to globally resolve the numerous components present in extracts of physiological fluids and tissues resulting in many co-elutions.

There are multiple ways to cope with these co-elutions. Selective sample preparation or pre-fractionation is rather suited for the targeted analysis of (also low abundant) metabolites than for metabolic fingerprinting, where sample preparation should be kept to a minimum to afford coverage of as many metabolites as possible.

Incorporating selective detectors, such as a nitrogen/phosphorus (NPD) or electron capture detector (ECD), is another alternative, but again is only advisable if selected components are analyzed. A mass spectrometer is a rather universal detector for both compound identification and quantification (for more details see chapter 4.4.1). Capillary GC in combination with MS has thus become a widely used tool for metabolomic investigations.[50] However, as current mass spectral deconvolution algorithms are only effective to a certain extent and are not helpful in case of overlapping isomers, improvement of chromatographic resolution is still desirable. Another strategy is to change selectivity, e.g. stationary phase, or to improve column efficiency, using for instance a longer column. Since resolution increases only with the square root of the column length, changing the selectivity is often more appropriate. On the other hand, a different stationary phase may result in baseline separation of a given peak pair but can cause co-elutions in other parts of the chromatogram. The concept of multidimensional GC (MDGC, GC-GC), which has been presented several decades ago [60], serially couples two columns with different selectivity. However, in the heart-cut technique at best only few chromatographic bands (unresolved peaks) can be selected for further separation on the second column rendering the technique unsuited for across-the-board screening of metabolites. The introduction of comprehensive two-dimensional gas chromatography overcame this limitation.

4.3.1.1 Fundamentals

In 1991, Phillips and coworker introduced comprehensive two-dimensional gas chromatography, which enables the separation of the complete sample in both dimensions.[61] Two columns with orthogonal separation characteristics are connected in series by a modulator resulting in enhanced resolution, a multiplicative increase in peak capacity, and a structured separation space.[62] The modulator is located between the columns to accumulate/focus the effluent from the first column and periodically transfer it to the second column in small concentrated segments.[63] In case of thermal modulation (see chapter 4.3.1.2) the focusing effect carries the additional benefit of creating narrow second dimension peaks and, thereby, increasing peak heights that in turn enhance detection sensitivity.[64] Due to its superior chromatographic resolving power, GC×GC is particularly suitable for the

separation of low molecular-weight analytes in complex samples.[61, 65] A picture and scheme of the GC×GC setup utilized in this work is shown in Figure 4.4.
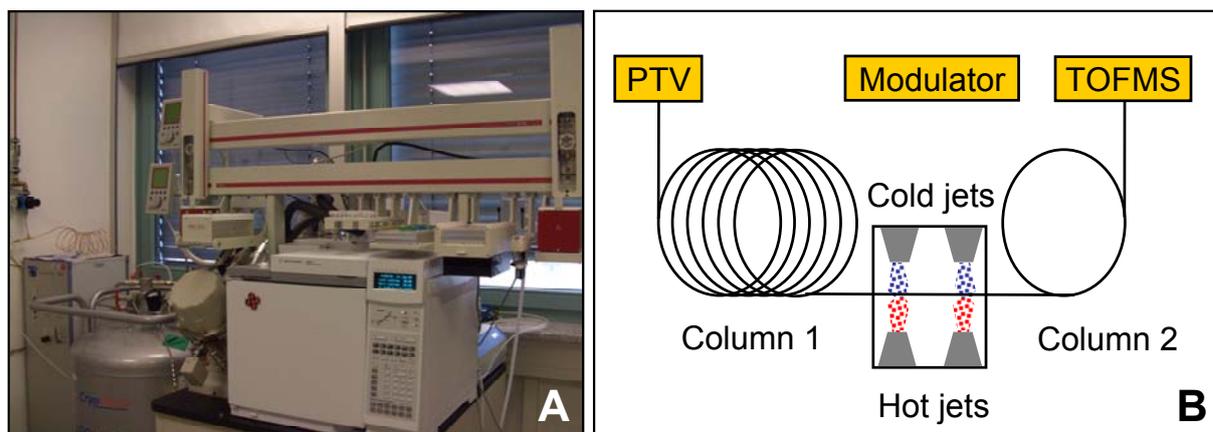
Figure 4.4. (A) Actual and (B) schematic setup of a GC×GC-EI-TOFMS. A thermal modulator is used to alternately cool and heat the incoming analytes and to release them periodically in packages onto the 2nd dimension column. PTV, Programmed Temperature Vaporizer. Reproduced in part from [7].

4.3.1.2 Modulators

For a true two-dimensional separation it is necessary to maintain the chromatographic resolution achieved in the first dimension. Therefore, at least three modulations per peak are required.[64] While one fraction of an analyte is separated on the second dimension column the next fraction is already sampled in the modulator. With peak widths of 6 to 25 seconds in the first column and 3 to 4 required modulations the separation in the second column can only last 2 to 8 seconds. Consequently, short narrow columns are used for fast analysis in the 2D, because the separation on the second column must be finished before the next fraction is transferred. Otherwise so-called wrap-around effects can occur. Due to an analysis time of 2 – 8 seconds, separations in the 2D are basically isothermal.

A number of modulators are available.[66] They can be divided into flow modulators and thermal modulators. Although improvements have been made regarding flow modulation [67], thermal modulators have become more prevalent. They use a localized temperature difference to create the required retention/release.[68, 69] The earliest modulators, like the sweeper, have applied heat.[61, 70] It consists of a modulator capillary that connects the columns and a rotating slotted heater, which periodically rotates over the modulator capillary to desorb, spatially compress and

release fractions. Simultaneously, modulators were developed that alternately cool and heat one or two segments of the GC capillary (or transfer line) and thus were able to achieve a wider volatility range. The first cryo-modulator was the longitudinal modulating cryogenic system (LMCS, Figure 4.5).[71, 72] The LMCS uses expanding liquid carbon dioxide for analyte trapping. It needs no additional heat source as the trap after a fixed time interval is moved mechanically to an upstream position in order to expose the focused fraction to the GC oven air temperature.
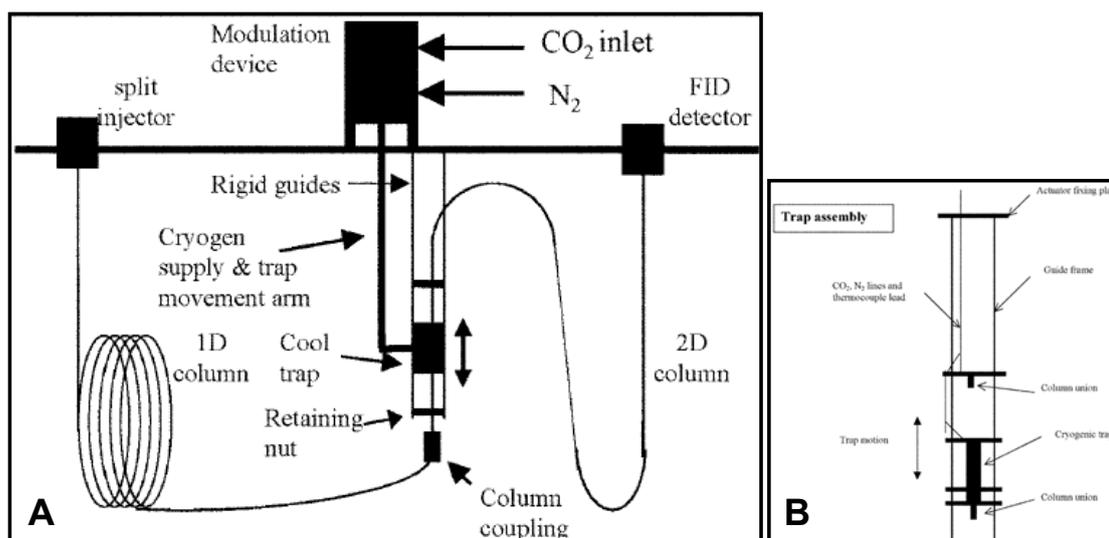
Figure 4.5. Schematic diagram of (A) a GC×GC setup using a LMCS and (B) the corresponding trap assembly. Modulation is accomplished by pneumatic movement of the trap. Liquid CO_2 is supplied through a restrictor allowing the liquid to expand. A secondary flow of N_2 prevents ice build-up at the ends. Taken from [73].

Today jet-based modulators with either carbon dioxide or (liquid) nitrogen for cooling are applied. The main benefit here is the absence of moving parts. In our studies a dual-stage four jet modulator is used that consists of two cold jets operating with cold nitrogen and two hot jets using hot air (Figure 4.6). The jets are used in alternating fashion. The first stage of modulation starts with cooling the eluate. The focused fraction is released by a stream of hot air and then trapped again in the second stage (cold jet on). Finally, the cold jet in the second stage is turned off and the hot jet turned on to transfer the fraction onto the second column. While the fraction is released the cold jet in the first stage is turned on again to trap the next fraction. With this dual-stage setup one stage is always cold to prevent analyte breakthrough during the heating phase. The focusing effect achieved during modulation creates very narrow peaks in the range of 50 – 200 ms in the 2nd dimension. Consequently, fast detectors are needed to appropriately describe the peaks.
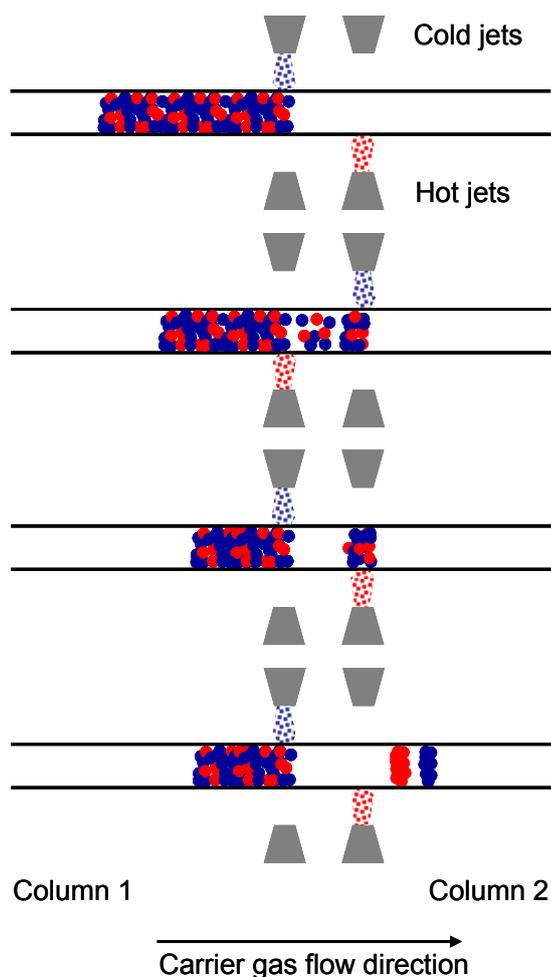
Figure 4.6. Simplified representation of a modulation process using a dual-stage four-jet modulator. Reproduced from [8].

4.3.1.3 Detectors

Acquisition of at least 15 data points across a typical 2D GC peak 100 ms in width requires a data acquisition rate of 150 Hz, which can be delivered for example by modern generation flame ionization detectors (FIDs) [74] or ECDs [75]. However, the detection method of choice for metabolomics studies is MS. A mass spectrometer provides structural information, which can be used for elucidation of unknowns, and high selectivity throughout the chromatogram. While rapid scanning quadrupole mass spectrometers suitable for GC×GC are available, most metabolomics studies to date have used fast acquisition time-of-flight mass spectrometers with nominal mass resolution. GC×GC coupled to an EI-TOFMS is predestined for the characterization of metabolic fingerprints taking the entire available (also non-targeted) information from all experiments into account for subsequent statistical analysis. TOFMS instruments like LECO´s (Corporation, St. Joseph, MI, USA) Pegasus IV are able to

acquire up to 500 mass spectra per second (at unit-mass resolution), which is highly suitable for proper reconstruction of the very narrow 2D peaks and reliable deconvolution of overlapping peaks.

4.3.1.4 Data visualization

GC×GC data are commonly visualized as contour plots. However, the detector records a continuous chromatogram. The raw chromatogram is cut into a series of individual chromatograms based on the modulation time. These short 2D chromatograms are then stacked to each other, creating a three dimensional plane, with 1D retention time as x-axis, 2D retention time as y-axis, and peak intensity as z-axis. The peaks are represented as contour plots or in a 3D view with color coding as abundance measure (see Figure 4.7).
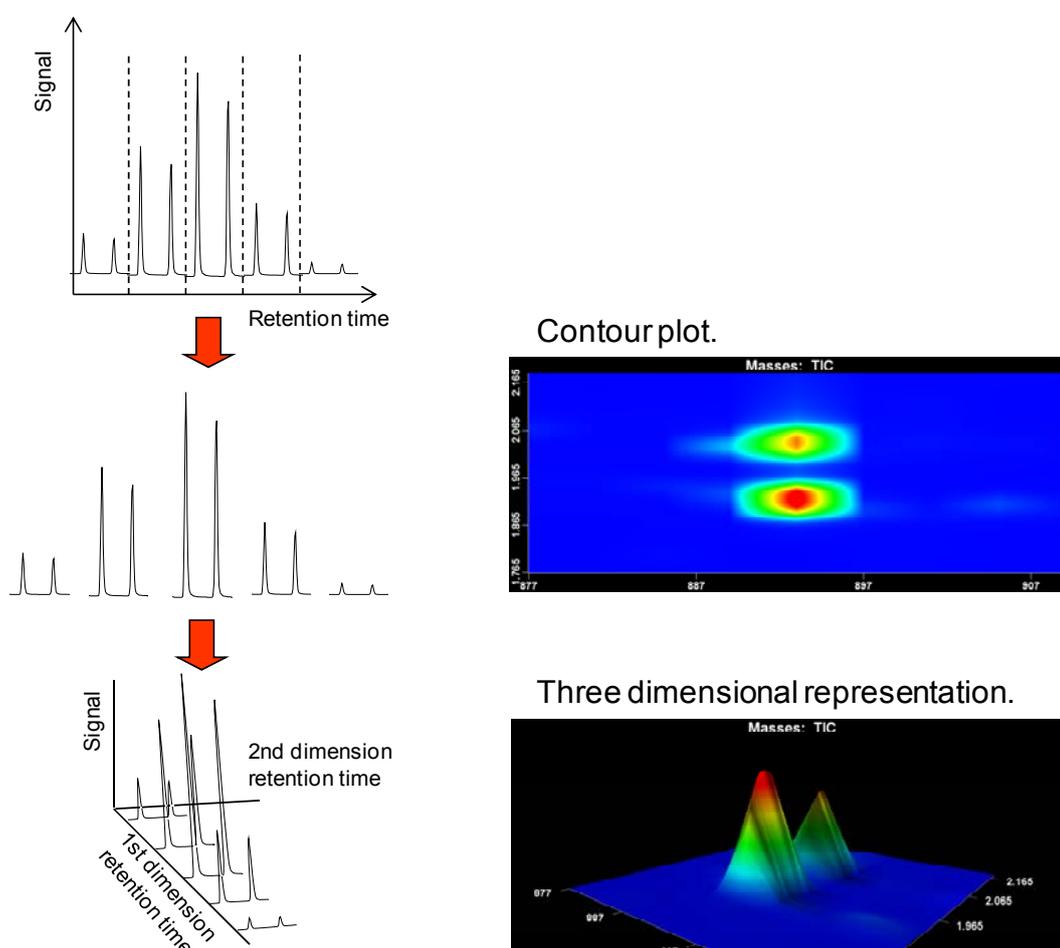
Figure 4.7. Visualization of GC×GC data. The raw GC chromatogram is cut into a series of 2nd dimension chromatograms whose duration is defined by the modulation time. The individual chromatograms are than stacked to each other creating a three dimensional plane. Reproduced from [8].

4.3.2 Gas chromatography – atmospheric pressure chemical ionization – time-of-flight mass spectrometry

In 1973, Horning et al. introduced atmospheric pressure chemical ionization (APCI) as a soft ionization technique for coupling GC to MS.[76, 77] Over the years the technique hardly attracted attention until recently, when both Schiewek et al. and McEwen and McKay modified an atmospheric pressure ionization (API) source for simultaneous operation in LC and GC mode.[78, 79] These APCI sources were recently made commercially available. They enable hyphenation of GC to high-resolution mass spectrometers.

APCI is widely employed for ionization of small, semi-polar to polar metabolites. The soft-ionization technique mainly generates quasi-molecular ions. Fragmentation is thus substantially reduced in comparison to hard ionization methods, like EI. A corona discharge needle provides electron ionization of the surrounding nitrogen gas. The subsequent plasma formation leads to an ion/charge transfer to water vapour. The hydronium ion-water clusters then produce the protonated molecular ions.

Overall, GC-APCI-TOFMS offers high chromatographic resolution and peak capacity of volatile compounds in addition to a promising ionization sensitivity and mass accuracy for the quasi-molecular ions. The high-resolution accurate mass measurement can be utilized to generate a sum formula. Considering in addition the isotopic pattern of a feature [80] as well as chemical and heuristic rules [81] will further facilitate the determination of unknown analytes. GC-APCI-TOFMS is particularly suited for the analysis of complex mixtures, but has only been applied to date in a few targeted studies for pharmaceutical research [82], analysis of foodstuff [83], and metabolomics [84, 85].

4.4 From raw GC×GC data to relevant biological information

A prerequisite for statistical data analysis is an unbiased and reproducible data processing. There are two ways of handling the acquired analytical raw data. The raw data are either converted to a specific format (typically NetCDF, network common data format) and exported for external work-up or processed in situ using the vendor software of the instrument manufacturer. In GC×GC-TOFMS based metabolomics the main data processing methods applied to raw data include

background correction, deconvolution, peak picking, peak integration, and peak merging of modulations belonging to one compound. The performance of each processing step has an influence on the quality of the final data and, ultimately, on the value of biological information extracted from the data. When using vendor software typically a peak table is obtained for each sample that typically comprises hundreds to thousands of features and their characteristics (1D and 2D retention time, mass spectrum, peak area, etc.). A comparative analysis of entire profiles requires reliable and automated data alignment to recognize identical metabolites in every sample. Data evaluation is achieved by using univariate or multivariate statistics. This chapter discusses method development for data processing, alignment and analysis and their implementation in metabolomics studies. A general overview is given in Figure 4.8.
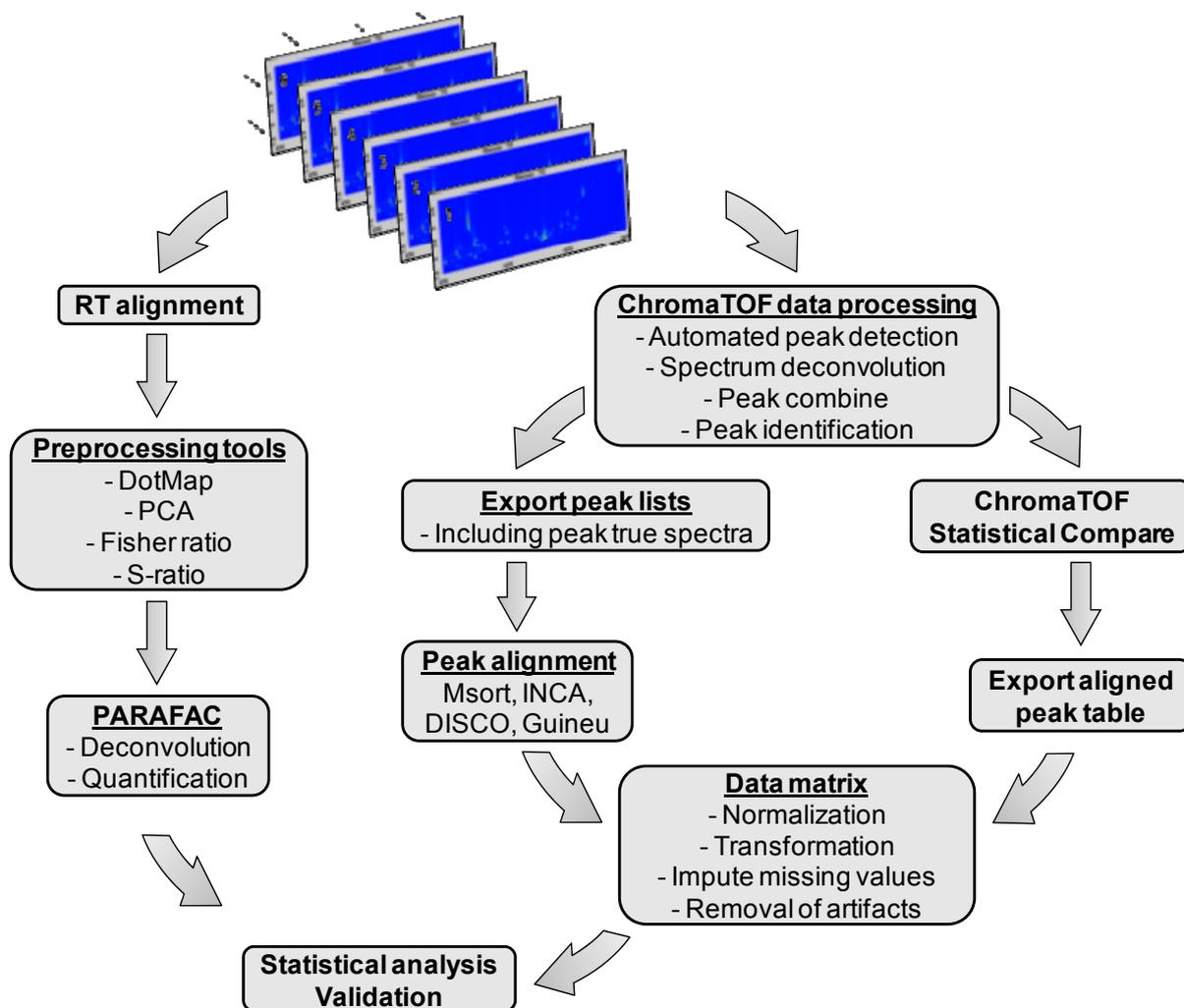
Figure 4.8. Possible routes that have been proposed in GC×GC based metabolomics studies to handle raw data. Reproduced from [6].

4.4.1 Data processing

Deconvolution is a mathematical operation that exploits differences in mass spectral information to separate overlapping peaks thereby enhancing the analytical resolution even further. It is particularly suited for the complex and large-scale datasets generated by the GC×GC-TOFMS instrument. It also yields a "pure" mass spectrum for each feature for mass spectral matching, compound identification and quantification. Two software packages are commonly used for deconvolution of GC×GC-TOFMS raw data, the commercial ChromaTOF software by LECO Corporation (St. Joseph, MI, USA) and parallel factor analysis (PARAFAC).[86] A third software solution that merits mentioning is GC Image.[87, 88] However, so far it has only been employed in GC×GC-based metabolomics for the processing and construction of the GC×GC contour plots of data from FAMEs of various bacteria.[89]

4.4.1.1 Deconvolution using ChromaTOF

ChromaTOF, like other commercial software packages, was originally intended for data acquisition and processing of 1D data, but was modified accordingly to handle data from 2D separations. ChromaTOF´s most important features besides data acquisition certainly are true signal deconvolution®, peak find, second dimension peak combine via spectral matching, peak integration and identification (based on mass spectral similarity to library spectra), classification, peak table compilation, quantification (semiquantitative or based on calibrations), data visualization (3D landscape or 2D color plots), and export. Koek et al. (2010) quantitatively evaluated the efficiency of the ChromaTOF software in terms of non-targeted GC×GC-TOFMS data processing.[90] A set of mouse liver specimens and pooled quality controls thereof were measured by GC×GC-TOFMS (Figure 4.9) and GC-qMS, and data processing results were compared. Targeted GC-qMS data processing involved inspection and, if necessary, manual correction of the integration of all quantified metabolites. For GC×GC-TOFMS, a target table was constructed manually and missing values across the samples were filled by lowering the mass spectral match threshold in a separate processing run; the rest (deconvolution, peak find/integration/identification, 2D peak combine) was fully automated. Individual peak list entries were quantified based on unique masses determined by the ChromaTOF software.

RSDs of the MS response of target compounds found by both GC methods were compared. For 70 % of these analytes accurate peak areas could be obtained with ChromaTOF. The inaccuracy of the remaining peak areas originated mostly from poor deconvolution, which caused errors in combining peaks from the same metabolite in the second dimension.
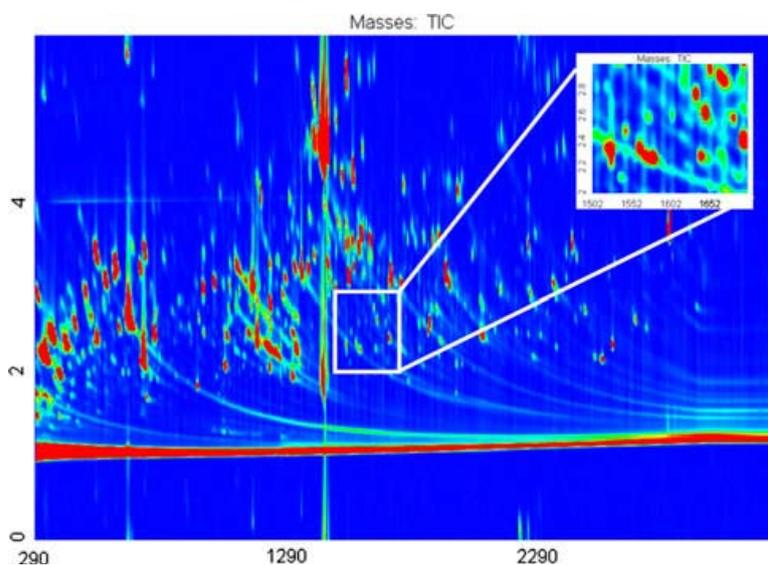
Figure 4.9. Contour plot of a GC×GC-TOFMS total ion chromatogram (TIC) of a pooled mouse liver sample. Taken from [90].

Although the data processing strategy pursued in this paper was very time-consuming and recommended for studies of 30-50 samples only, the work clearly illustrated the need for improvement of existing deconvolution algorithms implemented in commercial software (here ChromaTOF).

4.4.1.2 PARAFAC

The evolution of PARAFAC

The PARAFAC algorithm resolves and quantifies target analytes in higher order (3 and greater) data arrays. More dimensions can be obtained by adding analytical dimensions and/or combining multiple samples. For third order instrumentation, like GC×GC-TOFMS, only a single sample profile is needed due to the extra dimension in chemical selectivity. Sinha et al. (2004) were the first to apply PARAFAC to GC×GC-TOFMS based metabolomics data. PARAFAC initiated by trilinear decomposition (TLD) was used to deconvolute isomers on three overlapping species of possibly isomeric monosaccharide derivatives in a chromatographic subregion of a Huilmo grass metabolite extract.[91]

Initially, PARAFAC algorithms always required input on the specific number of factors in order to create the deconvolution model, which made it a semi-automated rather than a completely automated method. A solution was presented by Hoggard and Synovec (2007), who complemented TLD-initiated PARAFAC with an alternating least square (ALS) method.[92] This improved PARAFAC version automatically selected the adequate number of factors by starting off with a one-factor model for each sample and maximizing the number of factors (multi-factor model) until overfitting (i.e. "splitting" of the analyte signal) occurred. The mass spectral loadings of first and second column profiles of each model were matched against the mass spectrum of the target analyte. Overfitting was indicated by more than one factor in the same model showing match values above a user-defined mass spectral threshold. To avoid excessive computation times, the number of factors tested had to be limited.

In 2008, PARAFAC was successfully applied to a respective subsection of different GC×GC-TOFMS chromatograms (amongst others a yeast cell extract, Figure 4.10) in an automated way without requiring assumptions about analyte identities.[93] This non-targeted method was stated to be applicable to an entire chromatogram by individually analyzing all its subsections. However, the time needed to process a single sample was estimated to take tens of hours. The samples were spiked with compounds yielding known signals to qualitatively validate the performance of the method, but no quantitative data was shown.
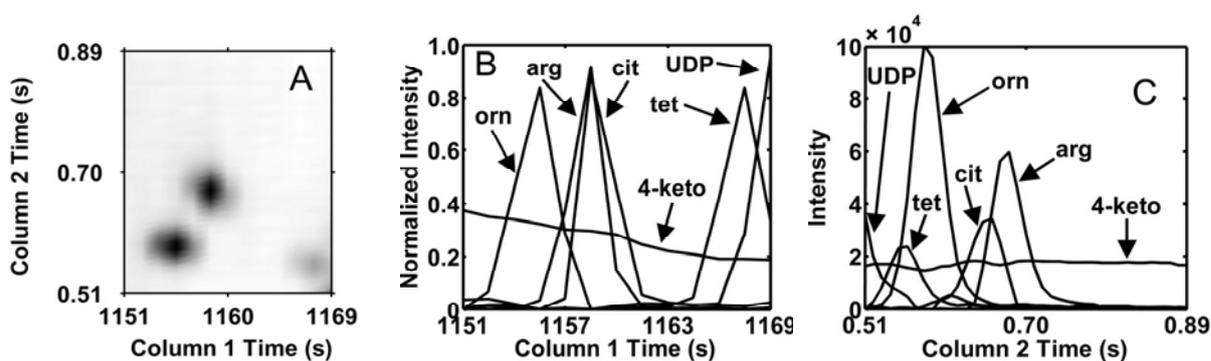
Figure 4.10. (A) Grayscale TIC 2D separation of the derivatized metabolite raw data subsection, loadings on the first (B) and second (C) separation dimension of the selected factors; orn, ornithine; cit, citric acid; 4-keto, 4-ketoglucose; arg, arginine; tet, tetradecanoic acid; UDP, UDP glucose fragment. Adapted from [93].

In order to find metabolites of interest in the raw data and particularly those that differ between samples or sample classes, multivariate analysis tools were employed. Deconvolution was then applied only to the interesting regions, which demanded considerably less computation time than conventional deconvolution of an entire chromatogram.

PARAFAC preprocessing tools

In 2004, the DotMap algorithm was developed to quickly locate candidate analytes among other sample components by using a weighted mass spectral similarity metric to scan all observed mass spectra.[94] This dot-product mass spectral matching algorithm was applied to find derivatized target metabolites in time segments of raw GC×GC-TOFMS data from human infant urine. In case of interference by other components respective regions were deconvoluted by TLD-initiated PARAFAC to provide pure profiles of the respective target metabolite.

In 2006, Pierce et al. introduced two approaches to determine natural chemical differences between classes of complex samples without a priori knowledge of particular target regions. First, a PCA was used prior to PARAFAC deconvolution to quickly capture differences between chromatograms (based on the two selective mass channels m/z 73 and 217) of organic acid extracts of plant samples.[95] Raw chromatographic data were translated into lower-dimensional principal components (PCs), which were retained and further evaluated when covering a relevant portion of the total variance (here PC1 and PC2). Mohler et al. (2006) proceeded similarly to identify 26 class-differentiating metabolite peak locations between yeast cell extracts grown on different carbon sources. Data from m/z 73, 205, and 387 were subjected to PCA, normalized to the summed TIC, and mean centered. Most variable metabolites were quantified using an in-house developed PARAFAC graphical user interface (GUI).[96] In the second approach by Pierce et al. (2006) an automated Fisher ratio method was incorporated as a front end tool.[97] Contrary to PCA, the Fisher ratio technique was supervised and thus more robust against within-class variation in discriminating samples. The Fisher ratio method was applied to entire 4D datasets (all mass channels were scanned providing another dimension) by using a point-by-point indexing scheme (considers each point in the separation space independently) for feature selection to discover potential regions with biological significance. The method was evaluated by means of a spike-in experiment and

applied to urine specimens from pregnant and non-pregnant women to detect unknown differences in organic acid metabolite composition.

Mohler et al. (2007) [98] took advantage of the progress made in the development of chemometric software by applying the Fisher ratio algorithm [97] with the automated PARAFAC GUI [92] and a Students´ t-test to the previously analyzed yeast metabolite data from cells grown under fermenting and respiring conditions (Figure 4.11). The number of group-distinguishing metabolite peaks could be nearly tripled employing a Fisher ratio threshold just above the noise level. In addition, Mohler et al. (2008) established the signal ratio (S-ratio) method to analyze sample differentiating locations of GC×GC-TOFMS data in 24 time interval measurements (injected in triplicate) from yeast cells grown under continuous, nutrient-limited conditions.[99] The S-ratio method was specifically developed to cope with the oscillatory behavior of the yeast metabolome as PCA and Fisher analysis were ill-suited for such a large number of sample classes and similar metabolite concentrations of contiguous time intervals. The phase delay of the cycling metabolites alternated widely in relation to the oxygen consumption cycling pattern. The S-ratio captured the amplitude (depth-of-modulation) of the cycles at each m/z based on the strongest (S_{max}) and the weakest (S_{min}) signal intensity for each metabolite. Finally, the most selective m/z ions for a given metabolite were retained to find locations of those metabolites that showed periodic patterns.
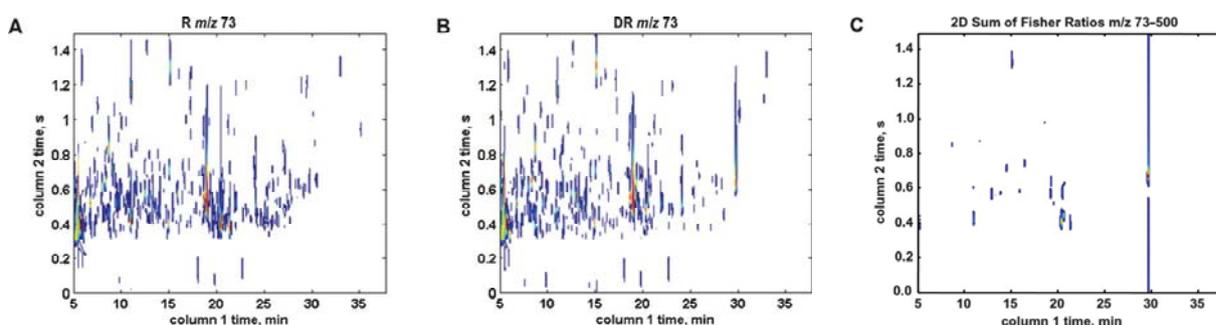
Figure 4.11. GC×GC-TOFMS contour plots at m/z 73 of (A) fermenting and (B) respiring yeast cells. (C) The sum of Fisher ratios plot for m/z 73-500. The signals in the plot provide the locations of chromatographic peaks that differentiate the classes (A) and (B). The streaking in (C) originates from trehalose, which is present in excess in (B). Taken from [98].

For all of the between-species metabolite profiling approaches presented in this section relative abundances based on TIC chromatographic peak volumes were given for the highly significant component peaks. A prerequisite for proper

implementation of the PCA, Fisher ratio, and S-ratio preprocessing tools, besides a sufficient sampling of the first dimension slices to reconstruct a column 1 peak profile, was a precise retention time alignment between samples to eliminate retention time variations. The next chapter takes a closer look at different alignment algorithms for both raw and already processed data.

4.4.2 Alignment

Accurate and preferably automated retention time alignment is essential for comparative characterization of multiple samples or sample classes. Identical metabolites have to be recognized over many GC×GC-TOFMS runs. Alignment of metabolic fingerprinting analyses poses a great challenge. The entire available information from all experiments has to be taken into account in order to maximize the possibility of detecting novel biomarkers. Numerous solutions have been presented in recent years for the alignment of GC×GC-TOFMS data.

In 2005, the first application of GC×GC-TOFMS to a metabolomics study was described analyzing tissue extracts from spleens of obese and lean mice.[100] In a subsequent publication, the same authors applied several univariate analysis strategies to the same dataset.[101] Chromatogram subtraction, averaging routines, weighting factors, Student's t-test, and automated peak comparison using the Compare function implemented in the LECO ChromaTOF software were employed to directly compare GC×GC metabolite fingerprints against a reference chromatogram. Because not all signals were found in all chromatograms, each sample had to serve as reference leading to computing-intensive comparisons.

In the PARAFAC community, algorithms have been proposed to compensate for run-to-run variations [102, 103], but they are only applicable to small segments of the 2D chromatogram. A retention time correction of the entire chromatogram in both separation dimensions was accomplished by Pierce et al. [104] (2005) using an indexing scheme with a piecewise RT alignment algorithm and by Zhang et al. [105] (2008) employing a correlation optimized warping (COW) algorithm. Whereas the piecewise alignment algorithm only accepts simple scalar shifts, the COW algorithm interpolatively stretches and compresses local regions to maximize the correlation between the warped and the reference chromatogram. The COW alignment

mechanism was demonstrated on chromatograms of standard mixtures (Figure 4.12) and two serum replicates that had been recorded 5 months apart.

Homogeneous samples (of the same origin or replicates) could be directly aligned based on their TIC chromatograms, while for heterogeneous samples selected ion count (SIC) chromatographic profiles were required to determine adequate warping parameters. Most approaches presented so far use raw instrument data without considering any global data processing functions (automated baseline correction, mass spectral deconvolution, peak picking, integration, library search, and signal/noise filtering). Therefore, data alignment was primarily based on 1D and 2D retention times ignoring the additive information given by a metabolite´s characteristic fragment ion mass spectrum.
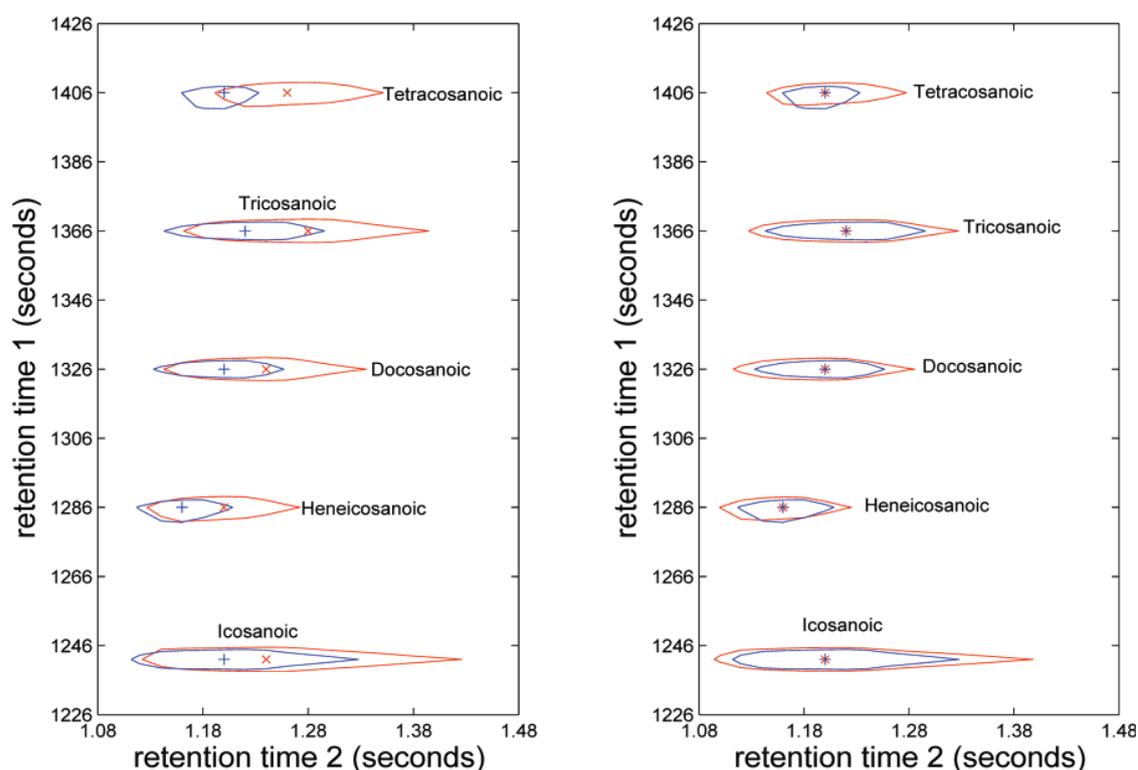
Figure 4.12. Alignment of two TIC chromatograms of a standard mix using the COW algorithm. Taken from [105].

Oh et al. (2008) were the first to use the data processing options provided by the ChromaTOF software to create peak tables for subsequent alignment by their own developed MSort® software.[106] MSort® works with 1D and 2D RTs, the linear correlation (Pearson coefficient) of fragment mass spectra, and, optionally, the peak name as assigned by ChromaTOF for peak merging and sorting. Ultimately, a new peak table representing all peaks in all chromatograms was generated. The software

was tested on standard metabolite mixtures and human serum specimens spiked with the standard mixture. A disadvantage of the software is the utilization of a user-defined RT window, whose fixed size in both RT dimensions affects the software´s reliability and efficiency. Furthermore, the algorithm can only handle small and linear RT distortions and needs considerable computational power for larger datasets.

In 2009, we developed a retention time correction and data alignment tool called INCA for comparative metabolic fingerprinting based on peak lists generated by ChromaTOF as a starting point.[107] The accuracy of processing, alignment, and detection of fold changes in metabolite concentration was validated by a spike-in experiment. In a proof-of-principle study we applied the metabolic fingerprinting strategy to quantitatively discriminate metabolomes of mutant and wild type *E. coli* strains. Further details are presented in chapter 6.

In a follow-up study, the peak lists generated from the spike-in and *E. coli* experiments were utilized to evaluate and directly compare the performance of INCA and the Statistical Compare (SC) alignment module, which had been made commercially available in 2010 as an add-on to ChromaTOF software version 4.[108] Enhancements and limitations of the two algorithms are discussed in chapter 7.

Wang et al. (2010) proposed a distance and spectrum correlation optimization (DISCO) algorithm to manage large nonlinear RT shifts.[109] DISCO used peak lists generated by ChromaTOF as input. After z-score transformation of metabolite RTs, DISCO selected a set of so-called landmark peaks from a specific number of samples based on Euclidean distances of two-dimensional retention times and mass spectrum similarity via Pearson correlation coefficient. RT variation of the landmark peaks was assumed to reflect those of all other metabolite peaks in the same sample. The two-dimensional RT space was divided into numerous subsections containing landmark peaks. A local partial linear fitting function in each RT dimension was applied to adjust RT shifts followed by metabolite peak alignment of all samples using a progressive RT map searching method based on the optimized Euclidean distance and mass spectrum similarity. The algorithm enabled alignment of data acquired under various experimental conditions (shown in Figure 4.13).

The method was claimed to be suitable for metabolomics; but instead of assessing its performance on metabolites, a mixture of saturated n-alkanes and the 8270 MegaMix from Restek containing mostly cyclic and polycyclic aromatic hydrocarbons

were used. Hence, the claim has been supported insufficiently to date. Also, the performance of DISCO was evaluated on five out of (only) six deuterated semivolatile internal standards spiked into derivatized extracts from rat plasma and no quantitative data was shown.
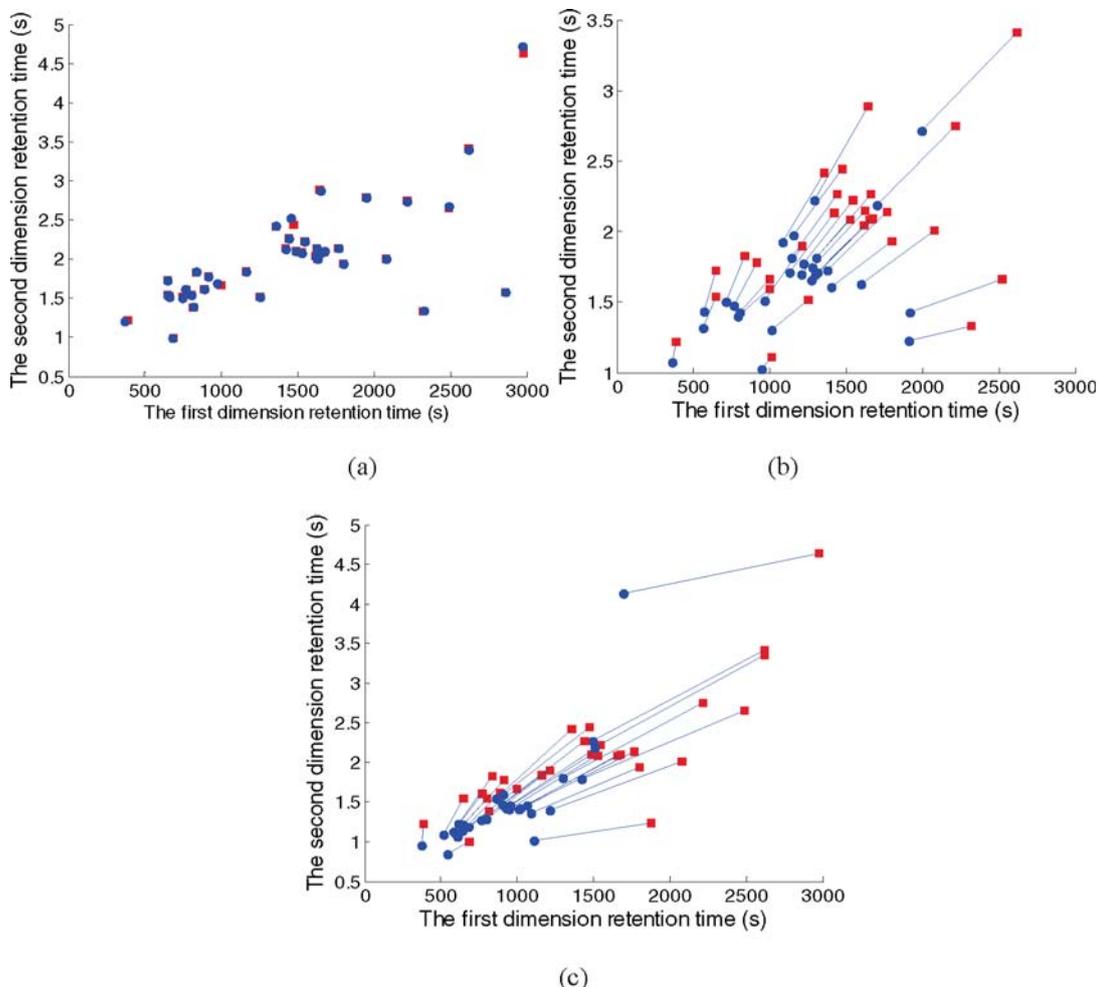
Figure 4.13. Performance of the DISCO algorithm in detecting landmark peaks from data acquired under different experimental conditions. Temperature programs were ramped at (a) 5, (b) 7, and (c) 10°C/min, respectively. Blue circles are landmark peaks in sample S^5_1 and red squares are corresponding landmark peaks in sample S^5_2 (a), S^7_1 (b), and S^{10}_1 (c). Taken from [109].

Recently, Castillo et al. (2011) presented a data analysis platform named Guineu for large GC×GC-TOF-MS datasets.[110] Like previously described algorithms, Guineu uses peak lists generated by vendor software (e.g. ChromaTOF) as input and aligns data based on retention times and mass spectral information (eventually identification). The freely available open source software package offers in addition data normalization, numerous filtering tools, the use of retention indices, and a tool

for group-type identification of compounds. Mainly peak areas from total ion chromatograms were used for quantification purposes. Concentrations of selected compounds were separately determined using ChromaTOF and characteristic *m/z* ions. In 440 serum specimens acquired over a period of 20 days the average RSD was under 10 % for the internal standards. Equally good was the day-to-day repeatability for the quantitative analysis of amino and carboxylic acids in control serum samples over a period of 3 months.

4.4.3 Data analysis and validation

Prior to statistical analysis, data are often pretreated, e.g. normalized, transformed, missing values are imputed, and artifacts are removed. The strategy for statistical analysis, e.g. univariate or multivariate, regression and classification techniques, should be selected according to experimental setup and the hypothesis under testing. The main statistical methods used in each study are given in the last column of Table S1. Unsupervised approaches, such as principal component analysis (PCA), and supervised approaches, such as partial least squares - discriminant analysis (PLS-DA) and Fisher ratio analysis were among the most popular multivariate methods, but others, including principal component discriminant analysis (PCDA) or hierarchical cluster analysis (HCA) can be also found.

In most cases, metabolomics datasets contain more variables than samples. Simultaneous testing of thousands of variables may yield small p-values by chance resulting in false positives. Thus, p-values must be corrected for multiple testing. The family-wise type I error rate can be controlled by conservative correction algorithms such as Bonferroni [111] or Westfall and Young's step-down adjusted p-values [112]. Another way of correcting for multiple comparisons is the estimation of the false discovery rate (FDR).[113] The FDR of a list of features is the expected relative frequency of false positives in it. These approaches have been introduced to the metabolomics field in 2006 [114] and are now commonly used.

In terms of validation, figures of merits including LOD, linear range, lower limit of quantification (LLOQ), accuracy, and precision are determined for metabolite profiling. For metabolic fingerprinting, class prediction and cross-validation (CVal) methods have been adopted to reduce the probability of obtaining correlations by chance.

4.5 GC×GC based metabolome analysis

This chapter focuses on targeted and non-targeted approaches in GC×GC based metabolomics studies. The GC×GC based applications are categorized based on biological matrices and briefly summarized in Table S1 in the appendix.

4.5.1 Biofluids

Biofluids can be obtained non-invasively by excretion (urine, sweat), secretion (breast milk, bile), or invasively by needle (blood, cerebrospinal fluid, cyst fluid). To date GC×GC-based studies have been focused almost exclusively on the metabolomes of blood and urine.

The analysis of whole blood is not common; rather serum or plasma is used. Both are isolated from whole blood in separate preparation steps. Serum is obtained as a supernatant after centrifugation of clotted blood. Plasma represents the liquid phase of blood after treatment with anticoagulants like citrate, EDTA or heparin. Serum and plasma are among the most complex biofluids requiring at least the removal of proteins as described in chapter 4.2.1. Urine may be used directly, but adjustment for urine concentration differences and removal of urea are recommended.

4.5.1.1 The blood metabolome

By way of comparison, more studies employing GC×GC-TOFMS have been published on the analysis of the serum/plasma than the urine metabolome.

Suomalainen and colleagues applied GC×GC-TOFMS to the measurement of amino acids in serum [115] and skeletal muscle specimens [116] collected from a mouse model of progressive late-onset mitochondrial myopathy. In comparison to wild type animals, mice carrying mutant mtDNA helicase Twinkle had increased serum levels of most amino acids.[115] In skeletal muscle, in contrast, only the levels of serine and alanine were significantly increased in affected mice.[116] Feeding a ketogenic, high-fat diet normalized the aberrant amino acid levels.

Asiago et al. (2010) studied 257 retrospective serum specimens obtained from 56 breast cancer patients using both GC×GC-TOFMS and NMR.[117] An aliquot of 200 µL serum was subjected to protein precipitation by the addition of 400 µL methanol and centrifugation, the protein pellet was further washed with chloroform, the combined

extracts were dried and derivatized using MTBSTFA (N-methyl-N-(tert-butyldimethylsilyl)trifluoroacetamide). Four metabolite markers, namely glutamic acid, N-acetyl-glycine, 3-hydroxy-2-methyl-butanoic acid, and nonanedioic acid were selected from the GC×GC-TOFMS data and used together with NMR data to build a predictive model for early detection of recurrent breast cancer.

Lankinen et al. (2011) analyzed plasma samples to study the effects of rye bread consumption.[118] In total 231 metabolites, including organic acids, sterols, and alcohols, were identified in the plasma samples. Rye bread consumption over eight weeks resulted in increased levels of ribitol, indoleacetic acid, and ribonic acid.

Beckstrom et al. (2011) studied perinatal asphyxia in a non-human primate model by analyzing heparinized blood specimens.[119] The abundance of ten metabolites was significantly changed in comparison to control animals. While the known biomarkers lactate and creatinine could be confirmed, new metabolites were identified as potential markers including succinate, malate, and arachidonic acid. Li et al. (2009) performed a biomarker discovery study by analyzing plasma (the anticoagulant used was not reported) specimens from 48 type 2 diabetes mellitus patients and 31 healthy controls.[120] The study yielded five potential biomarkers including glucose (to be expected), 2-hydroxyisobutyric acid, linoleic acid, palmitic acid, and phosphate.

GC×GC-TOFMS and lipidomics were used in a longitudinal metabolome study in children to detect changes in the serum metabolome preceding the onset of diabetes mellitus type 1. The GC×GC-TOFMS metabolomics dataset comprised 75 metabolites that had been measured in 419 samples.[121] Reduced serum concentrations of succinic acid and altered lipid profiles were observed at birth for children who progressed to diabetes. In 2011, Oresic et al. used the same analytical platforms to study the serum metabolome associated with schizophrenia and other psychotic disorders.[122]

Velagapudi et al. (2009) compared the serum metabolome and the lipidomes of serum, adipose tissue, and liver, respectively, of conventionally raised and germ-free mice to investigate the effect of gut microbiota on host energy and lipid metabolism.[123] Analysis of the serum metabolome was performed by GC×GC-TOFMS resulting in the identification of 185 metabolites. A clear segregation of the metabolic fingerprints based on colonization status of the gut microflora was observed. Increased levels of microbially derived metabolites, including 3-

hydroxyphenylpropionic acid, hydrocinnamic acid, and rhamnose, were found in addition to changes in serum levels of endogenous sugars, amino acids, fatty acids, and metabolites involved in energy metabolism and phase II metabolism.

GC×GC-TOFMS has not only been used for global metabolome investigation but also for metabolite profiling of selected compound classes, such as elucidation of the plasma fatty acid profile [124] and the analysis of amino acids [116, 125]. Waldhier et al. (2011) demonstrated improved enantiomer resolution and quantification of free D- an L- amino acids in serum and urine by GC×GC-TOFMS compared to 1D-GC-qMS.[126] Further details are presented in chapter 8.

A number of publications used serum or plasma as a model sample for method development and optimization. A differential metabolomics approach applying stable isotope labeling for quantification by GC×GC-TOFMS was investigated by Huang and Regnier (2008).[127] Samples were either derivatized with deuterium labeled or unlabeled derivatization agent, then mixed and analyzed. For derivatization N-methyl-N-tert-butyldimethylsilyl-trifluoroacetamide (MTBSTFA) was used, because derivatives are less prone to hydrolysis and label scrambling. Initial studies were performed using amino acid, fatty acid, and organic acid standards to test the efficacy of comparative quantification by isotope ratio analysis and the method was then applied to human serum. Compared to fatty acids, inferior average accuracies and precisons were obtained for isotope ratios of amino and organic acids. One drawback of this approach is that deuterium labeled compounds generally elute at lower retention times than their non-labeled analogs. The retention time gap increases with the number of deuterium atoms in the derivative and, consequently, with the number of derivatizeable functional groups per compound, which proved unfavorable for the quantification efficiency using this approach. Another disadvantage is the size of MTBSTFA. Derivatization is not always possible, particularly when target compounds contain sterically hindered sites.[128] O'Hagan et al. (2007) used closed-loop optimization, in which the complete process of parameter setting, GC run performance, and data analysis was automated and iterated to obtain a method that yielded the maximum number of peaks detected in serum specimens.[129] Gröger and Zimmermann (2011) employed parallel computing to accelerate data processing and analysis proceedings.[130] The setup was tested on a dataset obtained from plasma specimens of male Sprague–Dawley rats exposed to side stream cigarette smoke.

A common problem in metabolomics studies is the wide concentration range of metabolites in biological matrices. Koek et al. (2008) addressed the issue of mass loadability by using a wider bore column (0.32 mm ID) in the second dimension.[131] Mass loadability was improved 10-fold compared to a conventional narrow bore second dimension column (0.10 mm ID) at the expense of a 40 % decrease in peak capacity. The setup was used to analyze 15 fetal bovine serum specimens demonstrating better peak shapes for metabolites eluting close to high abundant compounds in case of the wider bore column combination (exemplified in Figure 4.14).
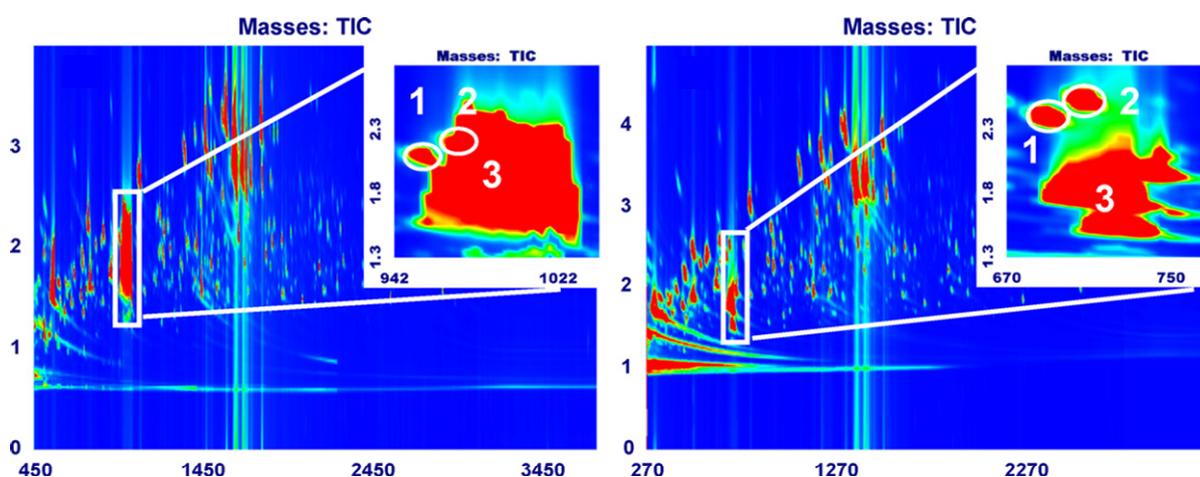
Figure 4.14. Total ion chromatograms of fetal bovine serum samples using a conventional setup (left) compared to a high loadability setup (right). 1, serine; 2, threonine; 3, urea. Reproduced from [131].

4.5.1.2 The urinary metabolome

The application of GC×GC-TOFMS to urine specimens has focused mainly on the analysis of inborn errors of metabolism. Wojtowicz et al. (2010) developed an automated data processing strategy for the detection of 21 target metabolites that are indicative of 6 defects of metabolism including thymidine phosphorylase deficiency, mevalonic aciduria, hawkinsinuria (also called 4-alpha-hydroxyphenylpyruvate hydroxylase deficiency), aromatic L-amino acid decarboxylase deficiency, propionic acidaemia, and medium-chain acyl-CoA dehydrogenase deficiency (an example is shown in Figure 4.15).[132] Sample preparation consisted of direct ethoximation in urine, followed by extraction with ethyl acetate. The organic phase was dried and further extracted using a methanol/acetone mixture. The extract was dried and trimethylsilylation was performed. The method was applied to analyze six patient

samples. Automated data processing resulted in the correct annotation of all pathological markers, which was superior over 1D-GC-MS.
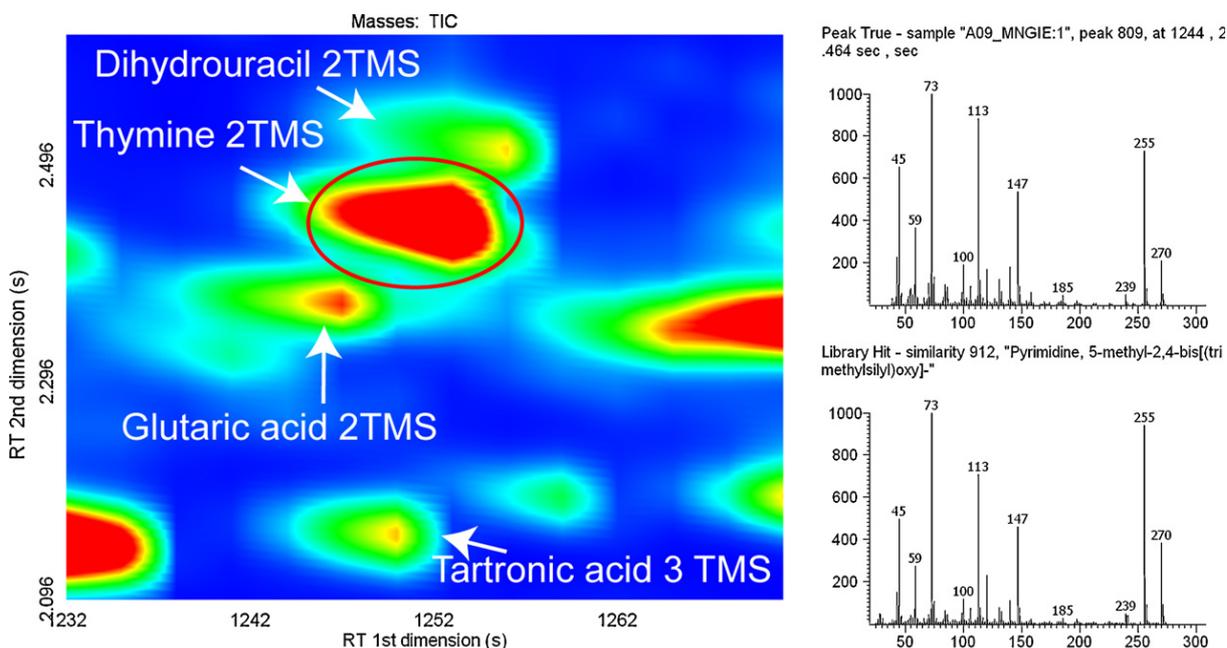
Figure 4.15. GC×GC-TOFMS contour plot zoom-in section of urine from a patient with thymidine phosphorylase deficiency. Dihydrouracil 2TMS, glutaric acid 2TMS, and tartronic acid 3TMS are resolved from thymine 2TMS. Thymine 2TMS is identified with a NIST spectral match of 912. Taken from [132].

Kouremenos et al. (2010) studied the urinary profiles of 5 patients with different inborn errors of metabolism.[133] The study included the comparison of two column combinations (nonpolar/polar versus polar/nonpolar) with the polar/nonpolar column combination making better use of the 2D separation space. The quantification capabilities of GC×GC-TOFMS using stable isotope-labeled internal standards were demonstrated for five metabolites. Furthermore, a novel urinary metabolite, crotonyl glycine, was identified that might serve as a diagnostic marker for mitochondrial 3-hydroxy-3-methylglutaryl CoA synthase deficiency. Urine specimens were diluted to a fixed creatinine concentration, followed by methoximation, extraction with ethyl acetate, and derivatization with BSTFA (bis-(trimethylsilyl)trifluoroacetamide) containing 1% TMCS (trimethylchlorosilane). Microwave assisted derivatization was performed as previously optimized by Kouremenos et al. (2010) employing a mix of standard compounds (amino acids, organic acids, sugars, sugar alcohols, and fatty acids).[134] Sinha et al. (2004) used GC×GC-TOFMS to analyze metabolites in a human infant urine specimen to demonstrate the applicability of their DotMap algorithm (see chapter 7.1.2.2).[94]

4.5.2 Cell culture and tissue extracts

Pasikanti et al. (2010) performed a metabolic footprinting study comparing the metabolite pattern in cell culture media from tumorigenic and nontumorigenic uroepithelial cells.[135] A culture media aliquot of 200 µL was used and protein precipitation with methanol was performed. The SC tool from LECO (presented and evaluated in chapter 7) was used for data analysis. Twenty metabolites, including glucose, other sugars, sugar alcohols, and malic acid, were identified that differed significantly between the two cell types.

Mucosal biopsies from the ascending colon were investigated using both lipidomics and GC×GC-TOFMS based metabolomics to study the pathophysiology of irritable bowel syndrome.[136] Tissue samples for GC×GC-TOFMS were extracted using methanol. A total of 107 metabolites were identified in the GC×GC chromatograms and used for further analysis. The most prominent change with an almost 14-fold up-regulation in patients compared to controls was observed for 2(3H)-furanone.

Kleemann et al. (2010) studied the pathogenesis of insulin resistance in male ApoE3Leiden transgenic mice that were fed a high-fat diet.[137] Liver tissue samples were freeze-dried overnight, homogenized, and extracted using 80 % (v/v) methanol in water. GC×GC-TOFMS analysis revealed significant changes in the concentrations of glucose, gluconeogenesis and Krebs cycle metabolites, and branched amino acids. The samples were used by Koek et al. (2010) to evaluate commercially available software for non-target processing of GC×GC-TOFMS data.[90]

Mervaala et al. (2010) investigated metabolic profiles in angiotensin II–induced cardiac hypertrophy by analyzing heart tissue specimens from transgenic and control rats, respectively.[138] The GC×GC-TOFMS dataset comprised 247 metabolites, of which 112 showed significant differences. Further applications of GC×GC-TOFMS include a biomarker discovery study in spleen tissue extracts of obese NZO mice [100] and the analysis of L-β-methylamino-alanine in mouse and human brain tissues [139].

4.5.3 Bacteria and yeast

GC×GC studies on the metabolomes of bacteria and yeast were mainly performed by the Lidstrom group and Synovec and colleagues, respectively, both at the University

of Washington in Seattle, using PARAFAC and associated preprocessing tools. Details about the methodologies are given in chapter 4.4.1.2.

4.5.3.1 Bacteria

The Lidstrom group focused their GC×GC-TOFMS analyses exclusively on *Methylobacterium extorquens* AM1, a facultative methylotroph capable of generating energy and assimilating carbon from reduced one-, two-, or four-carbon substrates such as methanol, ethylamine, or succinate, respectively.[140-142] It presents an attractive model system for the application of novel metabolomics-related techniques because of the in-depth understanding of its genetics and biochemistry.

In 2008, Guo and Lidstrom investigated metabolite profiles of methoximated/trimethylsilylated extracts from cells grown on either methanol or succinate.[140] Absolute quantification of amino acids and major intermediates of central C_1 and multicarbon metabolism was accomplished by spiking different amounts of standard mixtures into cell cultures. In addition, Fisher ratio analysis at *m/z* 147 determined thirty-six differentially expressed metabolites. Thirteen could be identified by matching to library mass spectra. PARAFAC GUI was used for peak deconvolution and calculation of normalized peak volumes from reconstructed 3D peaks (Figure 4.16). The completely overlapping peaks of 3-hydroxyisobutyrate and 3-hydroxybutyrate were successfully deconvoluted, which allowed for the detection of differences in abundance of both isomers between cells grown on methanol and succinate, respectively. Further, increased concentrations of methylfumaric and methylmalic acid in C_1 metabolism indicated an active glyoxylate regeneration cycle.

Yang et al. (2009) utilized the complementary nature of LC- and GC-based methods by applying liquid chromatography – tandem quadrupole mass spectrometry (LC-MS/MS) and GC×GC-TOFMS with PARAFAC data analysis to compare concentrations of 39 targeted metabolites involved in central carbon metabolism of *M. extorquens* AM1 grown on ethylamine (C_2) and succinate (C_4), respectively.[141] The abundance of twenty intermediates changed between differentially grown cells reflecting the pathways linked to C_2 and C_4 metabolism. Agreement of the two methods and quantification accuracy were verified for seven metabolites that had been detected by both methods.
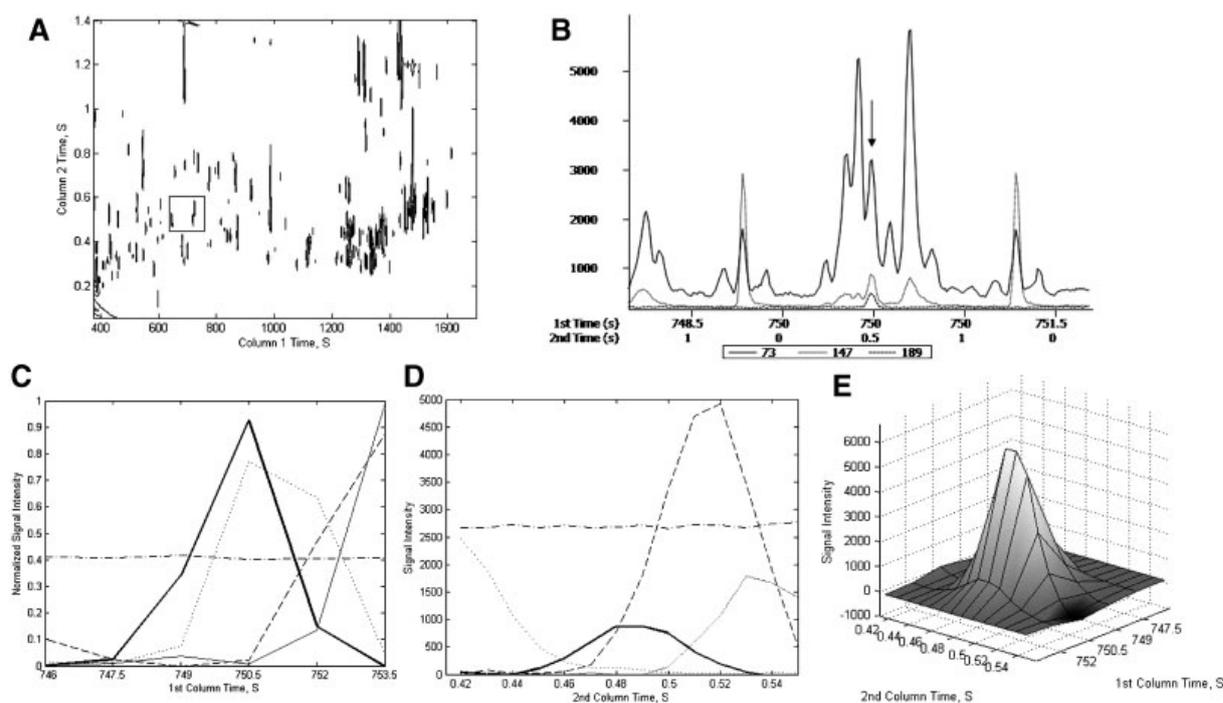
Figure 4.16. Demonstration of using PARAFAC for peak quantification. (A) TIC contour plot of a derivatized methanol chemostat culture extract analyzed by GC×GC-TOFMS. (B) 2D plot of the boxed region in (A) at *m/z* 73, 147, 189. The peak with an arrow on top is identified as glyceric acid, triTMS. (C) Glyceric acid, triTMS (solid lines) was completely separated from four overlapping components (dashed and dotted lines) on both column 1 (C) and column 2 (D). (E) The 3D reconstruction of the glyceric acid, triTMS peak. Reproduced from [140].

In 2010, Okubo et al. determined the capability of *M. extorquens* AM1 mutants to grow on C_2 compounds in the absence of malyl-CoA/β-methylmalyl-CoA lyase and malate synthase activity, respectively.[142] The extensive study included a ^{13}C labeling experiment, microarray gene expression analysis, enzyme activity assays, and metabolite determination by means of LC-MS/MS and GC×GC-TOFMS using the in-house developed signal ratio method to detect differences between mutants. One of the major findings of the study was the identification of an alternative route for glyoxylate consumption in *M. extorquens* AM1, in which glyoxylate is converted to intermediates of central metabolism via a part of the serine cycle coupled to the glycine cleavage system. This alternative pathway complements the malate synthase reaction, which is a bottleneck in *M. extorquens* AM1 cultivated on C_2 compounds.

Methyl esters of fatty acids from *Brevundimonas diminuta*, *Chryseobacterium gleum*, and *Stenotrophomonas maltophilia*, respectively, were analyzed by David et al. (2008) using the commercial Sherlock Microbial Identification System, GC×GC-FID,

and GC-MS in electron impact and positive chemical ionization mode.[89] The enhanced selectivity and group type separation of GC×GC allowed a more complete elucidation of the fatty acids in the microorganisms (Figure 4.17).
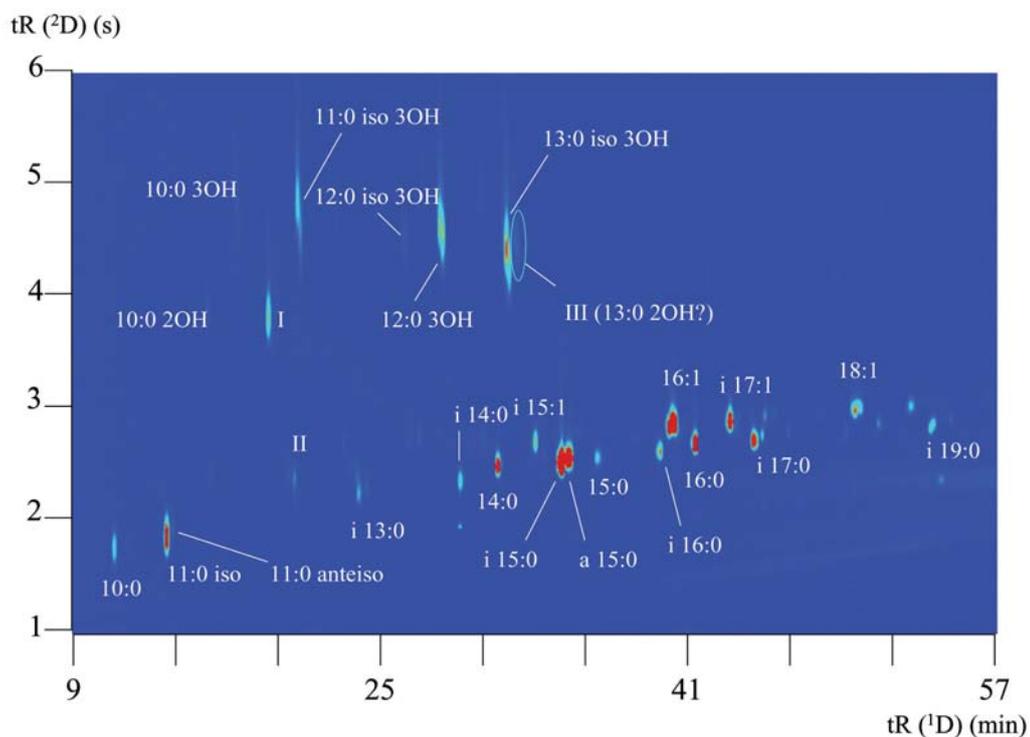
Figure 4.17. Structured separation space of FAMEs from *Stenotrophomonas maltophilia* using GC×GC-FID. Taken from [89].

4.5.3.2 Yeast

The Synovec group covered most of the GC×GC-TOFMS based metabolomic research on yeast. Mohler et al. (2006) compared metabolite extracts isolated from yeast cells grown on either glucose (fermentation) or ethanol (respiration).[96] PCA on three selective mass channels combined with PARAFAC and LECO ChromaTOF could locate and identify 26 class-distinguishing metabolites.

A Fisher ratio/PARAFAC GUI approach on the same dataset raised the number of identifications nearly three-fold.[98] A Student´s t-test with a 95 % confidence limit determined 54 metabolites to be statistically changing. In 2008, Mohler et al. developed the S-ratio/PARAFAC GUI method, another multivariate classification analysis tool geared specifically to time-interval experiments. In 24 time-point measurements of yeast cells, 44 identified unique metabolites were located that exhibited cycling with a depth-of-modulation amplitude of greater than three.[99] Humston et al. (2008) employed the same method to detect time-dependent

metabolite changes in wild type and *snf1Δ* mutant yeast cells under repressing and depressing conditions.[143] In a follow-up study, GC×GC-TOFMS and LC-MS/MS were used to investigate metabolite differences between wild type and four mutant strains (*snf1Δ*, *adr1Δ*, *cat8Δ*, and *adr1Δcat8Δ*) grown in fermentable synthetic complete medium containing 5 % glucose.[144] For this study, metabolite levels were correlated with RNA data for pathways comprising statistically significant metabolites (TCA cycle, glyoxylate cycle, and gluconeogenesis).

Cooper et al. (2010) developed a high-throughput profiling technique to quantify primary amine-containing metabolites in the *Saccharomyces cerevisiae* yeast deletion collection by means of capillary electrophoresis coupled to a laser-induced fluorescence detector (CE-LIF) of cell extracts labeled with 4-fluoro-7-nitro-2,1,3-benzoxadiazole (NBD-F).[145] Amino acid peak identities were verified by spike-in experiments and parallel quantification of trimethylsilylated amino acids by means of GC×GC-TOFMS. Comparison of data obtained for four samples showed that 12 out of 13 amino acids measured by both CE-LIF and GC×GC-TOFMS were correlated at R^2 >0.7 (6 out of 13 had R^2 >0.9). Glutamine yielded a low correlation, as it did not separate from valine by CE.

4.5.4 The plant kingdom

An array of analytical strategies has been proposed for plant metabolomics.[146] GC-MS has proven the most popular among these. In recent years GC×GC-TOFMS approaches have emerged.

In 2005, Hope et al. first optimized the GC×GC-TOFMS separation space for the analysis of TMS derivatives of amino and organic acids and then applied the method to two extracts of common lawn grass samples.[147] Pierce et al. (2006) employed PCA data mining on entire GC×GC-TOFMS chromatograms to discover differences between organic acid extracts of basil, peppermint, and the sweet herb stevia.[95] Kusano et al. (2007) combined GC-TOFMS and GC×GC-TOFMS for comprehensive non-targeted metabolite profiling of brown rice seeds from the world rice core collection (WRC), a representative set of Asian rice cultivars.[148] 1D-GC-TOFMS was used for high-throughput profiling of all rice cultivars, followed by GC×GC-TOFMS for a detailed analysis of a representative subset. Since the WRC covered most of the DNA polymorphisms, metabolite phenotyping provided an insight into the metabolite

diversity of rice natural variants and facilitated the selection of nutritionally useful rice varieties.

Following publication of the genome sequence of the green alga *Chlamydomonas reinhardtii* in 2007 [149], May et al. (2008) employed GC×GC-TOFMS based metabolomics and LC-MS based shotgun proteomics profiling technologies complemented with *in silico* genome annotation methods to characterize the molecular repertoire of this model organism under reference conditions.[150] Computational metabolic modeling revealed new putative genes, pathways, and enzymatic links. In 2009, Kempa et al. proposed a workflow for flux analysis incorporating GC×GC-TOFMS, ChromaTOF deconvolution, and an algorithm for batchwise data processing to study the ^{13}C-pattern in *C. reinhardtii*.[151] Doebbe et al. (2010) reported a differential GC×GC-TOFMS analysis of metabolite profiles between a high H_2-producing *C. reinhardtii* strain and the corresponding wild type before and during the H_2 production phase.[152] In contrast to GC-MS, the number of resolved peaks could be increased by an order of magnitude. Remarkable differences were detected in anaerobic pathways of starch degradation, fatty acid and lipid synthesis, and formate/ethanol production that provided potential targets for metabolic engineering to further enhance substrate supply for the hydrogenase(s) in the chloroplast.

As part of the META-PHOR project (METAbolomics for Plants Health and OutReach), Allwood et al. (2009) presented an extensive GC-TOFMS ring experiment based on major differential metabolite features of melon, broccoli, and rice sample extracts for short-term inter-laboratory reproducibility.[153] During the course of the study conventional GC-TOFMS was also compared to GC×GC-TOFMS.

Finally, Johanningsmeier and McFeeters (2010) described a non-targeted solid-phase microextraction (SPME) GC×GC-TOFMS approach for discriminative analysis of fermented cucumber volatiles before and after anaerobic spoilage making use of the Reference and Compare function of the ChromaTOF software.[154] A composite sample that comprised equal volumes of aliquots from each treatment served as reference. Analysis of variance (ANOVA) of analyte log peak areas detected 33 metabolites that changed significantly in concentration after spoilage.

4.5.5 Miscellaneous

This chapter takes a brief look at GC×GC studies of metabolites in milk, beverages, invertebrates, and feces.

Hyötyläinen et al. (2004) tested four column combinations (see Table S1) by GC×GC-FID to find optimal conditions for the separation of dietary milk derived fatty acids as their methyl ester derivatives.[155] Best results, in particular for C18 fatty acids, were obtained with a nonpolar 100% dimethylpolysiloxane HP-1 column coupled to a polar narrow-bore (0.05 mm ID) Carbowax column as the second dimension column. Vlaeminck et al. (2007) utilized the superior separation efficiency of GC×GC for the analysis of milk fatty acid profiles from dairy cows fed a control versus a ration diet supplemented with marine algae.[156] FAMEs were separated using both nonpolar/polar and polar/nonpolar column combinations. The well-ordered structure in the 2D contour plot facilitated identification of known and unknown compounds.

Mayadunne et al. (2005) compared two different column sets, low-polarity/polar and polar/nonpolar, for the separation of amino acids as their propyl chloroformate derivatives.[157] Standards were characterized by GC-MS, GC×GC-FID, and GC×GC-TOFMS. A selection of samples (wines, beer, and honey) was analyzed by GC×GC-FID using the favored column combination comprising a polar 30 m × 0.25 mm ID × 0.25 µm SolGel-WAX and a nonpolar 1.5 m × 0.1 mm ID × 0.1 µm BP-1. Junge et al. (2007) applied GC×GC to the separation of amino acid enantiomers as their ethyl chloroformate derivatives in samples of beer by coupling a Chirasil-L-Val column with a low-polarity column in the second dimension.[158] In addition, a chiral/polar column combination and various derivatization techniques were tested.

In further GC×GC-TOFMS studies, Ralston-Hooper et al. (2008 & 2010) employed MSort® (for details see chapter 4.4.2) to examine metabolomic changes among distinct populations of aquatic invertebrates [159, 160], while Aura et al. (2010) used an anaerobic *in vitro* colon model combined with a global metabolomics approach to identify novel metabolites synthesized by the fecal microbiota.[161]

5 Experimental section – Material, methods and instrumentation

5.1 Chemicals

Metabolite standards, Luria-Bertani (LB) broth, odd-numbered saturated straight chain fatty acids (C9-C19), methoxylamine hydrochloride, isooctane, pyridine, MCF, norvaline, all solids of L and D configured amino acids (AAs), racemates of proteinogenic AAs, [^{2}H$_7$]-trans-cinnamate, [2,2,4,4-^{2}H$_4$]citrate, [U-^{13}C]glucose, and [U-^{13}C]β-hydroxybutyrate were purchased from Sigma-Aldrich/Fluka (Taufkirchen, Germany). Glycerol was from BDH Prolabo (VWR International, Vienna, Austria), [U-^{13}C]lactose from Omicron Biochemicals (South Bend, USA), methanol (LC-MS grade) and chloroform (HPLC grade) were from Fisher Scientific GmbH (Ulm, Germany), [U-^{13}C]fumarate, [U-^{13}C]lactate, [U-^{13}C]pyruvate, [U-^{2}H]succinate, and a [U-^{13}C, U-^{15}N] cell free amino acid mix from Eur-isotop (Saint-Aubin Cedex, France), DL-[2,3,3-^{2}H$_3$]malate and [4,6,7-^{2}H$_3$]5-hydroxyindole-[^{2}H$_2$]acetate from CDN Isotopes Inc. (Quebec, Canada), and *N*-methyl-*N*-trifluoroacetamide (MSTFA) from Macherey-Nagel (Dueren, Germany). The water was purified using a PURELAB Plus system (ELGA LabWater, Celle, Germany).

5.2 Sample preparation

5.2.1 Automated methoximation/silylation

Automated methoximation/silylation was carried out in metabolic fingerprinting experiments in capters 6 and 7 and in the quantitative analysis of β-hydroxybutyrate, glucose, and intermediates of the citrate cycle in chapter 9.

After sample evaporation, the vials were closed with magnetic crimp caps and placed in a cooled tray (5 °C) for automated handling using the MPS-2 Prepstation sample robot (Gerstel, Muehlheim, Germany). Fifty microliters of 10 mg/mL methoxylamine hydrochloride in pyridine were added and incubated at 60 °C for 60 min. Prior to trimethylsilylation, 10 µL of internal standard solution containing odd-numbered, saturated straight chain fatty acids (C9–C19) at a concentration of 1 mM were added

to the sample residue. Then, 50 µL MSTFA were added and incubated at 60 °C for 60 min. Samples were vortexed continuously at 500 rpm during incubation in the agitator. For analysis, a sample volume of 1.5 µL was injected in splitless mode.

5.2.2 Derivatization with methyl chloroformate

Methanol/methyl chloroformate derivatization was used for quantitative analysis of amino acid enantiomers (chapter 8) as described by Waldhier et al.[5]

Magdalena Waldhier (Institute and Chair of Functional Genomics, University of Regensburg) optimized the derivatization steps and performed in part the sample preparation together with Nadine Nürnberger (Institute and Chair of Functional Genomics, University of Regensburg).

5.2.3 Derivatization with propyl chloroformate

Propyl chloroformate (PCF) derivatization was employed for quantitative analysis of amino acids (chapter 9) according to Kaspar et al.[162]

Briefly, AAs were derivatized directly in the aqueous biological sample, 20-50 µL of which were transferred together with 20 µL of the stabilization reagent to a 2-mL autosampler vial (Gerstel). The vial was closed with a magnetic crimp cap to allow automated handling by the MPS-2 Prepstation sample robot (Gerstel) in all the following steps. First, the sample was diluted with water up to 225 µL, followed by addition of 10 µL of a norvaline solution (200 µM) and 10 µL of internal standard mix. A mixture of uniformly ^{13}C, ^{15}N-labeled alanine, glycine, valine, leucine, isoleucine, threonine, serine, proline, asparagine, aspartate, methionine, glutamate, phenylalanine, glutamine, lysine, histidine, tyrosine, and tryptophan, as well as [2,5,5-$^{2}H_3$]α-aminoadipic acid, were used as internal standards with a concentration range from 0.0438 to 1.4175 mM. To increase the pH of the solution, 120 µL of 0.33 M sodium hydroxide solution were added, followed by 50 µL of picoline in propanol, which acts as a catalyst for the derivatization reaction (solution provided by Phenomenex Inc., Torrence, CA, USA). The vial was moved to an agitator and the solution was mixed at 750 rpm for 0.2 min at 35 °C. Fifty µL of propyl chloroformate in chloroform were added to the sample, the solution was mixed for 0.2 min (750 rpm, 35 °C), equilibrated for 1 min and again mixed for 0.2 min. To extract the derivatives, 250 µL of isooctane were added and the vial was vortexed for 0.2 min

(750 rpm, 35 °C). For analysis, an aliquot (2.5 µL) was taken from the upper organic phase and injected at a split ratio of 1:15.

5.3 Instrumentation

5.3.1 Metabolic fingerprinting/quantitative analysis of selected metabolites

A LECO (Corporation, St. Joseph, MI, USA) Pegasus 4D GC×GC-TOFMS instrument was used comprising an Agilent Technologies Model 6890 GC (Agilent, Palo Alto, CA, USA), a dual-stage, quad-jet thermal modulator, a secondary oven coupled to a fast-acquisition time-of-flight mass spectrometer providing unit mass resolution, a PTV injector (Gerstel), and a MPS-2 Prepstation sample robot (Gerstel) for automated sample derivatization and handling. The robot was equipped with two agitators for sample incubation and two syringes of different volumes. A 10-µL syringe was used for internal standard addition and sample injection, while reagents were added by means of a 250-µL syringe. Between adding steps, the syringes were washed at least 3 times with isooctane. Samples were kept in a cooled tray at 5 °C.

The GC×GC column set consisted of a Rxi®-5ms column (30 m × 0.25 mm ID × 0.25 µm film thickness) from Restek (GmbH, Bad Homburg, Germany) equipped with a 5 m × 0.25 mm ID deactivated pre-column (Agilent) as the first dimension column and a Rtx®-1701 (2 m × 0.1 mm ID × 0.1 µm film thickness, Restek) as the second dimension column. The oven temperature was initially held at 50 °C for 0.2 min, raised at 8 °C/min to 265 °C, and held for 10 min. A positive offset of 5 °C was used for the second dimension column and a 15 °C offset relative to the first dimension column for the modulator. The column flow (constant) was 1 mL He/min. The temperature of the PTV Injector was set at 50 °C for 0.5 min and ramped at 12 °C/s to 250 °C over 1 min. A chemically inert SILTEC liner (Gerstel) was used. For GC×GC analyses a 4 s modulation period was used (0.6 s hot pulse time and 1.4 s cool time). Mass spectra were acquired from *m/z* 40-600 at a rate of 200 spectra/s. For 1D-GC-TOFMS analysis the modulator was turned off and mass spectra were collected at 50 spectra/s. The solvent delay was always 8 min.

5.3.2 Quantitative analysis of amino acid enantiomers

The GC×GC-TOFMS instrument described in chapter 5.3.1 was used. Following settings were altered: an Rt®-γDEXsa column (30 m × 0.25 mm ID × 0.25 µm film thickness, Restek) protected by a 5 m × 0.25 mm ID fused silica deactivated guard column (Agilent) was emloyed as the first dimension column, while an Rtx®-1701 (2 m × 0.1 mm ID × 0.1 µm film thickness, Restek) or a ZB-AAA (2 m × 0.25 mm ID × 0.1 µm film thickness, Phenomenex) column was employed as the second dimension column. The helium flow-rate was set at 1.9 mL/min (constant flow). The solvent delay was kept at 19 min. Temperature programs are detailed in chapter 8.

5.3.3 Quantitative analysis of amino acids

Quantitative analysis of AAs was performed as previously described.[162] An Agilent Technologies Model 6890 GC equipped with a 5975 Inert XL MSD, a PTV injector (Gerstel), and a MPS-2 Prepstation sample robot (Gerstel) was used. The syringes were washed with propanol after adding aqueous solutions and with chloroform and propanol after addition of organic solutions.

The GC column was a ZB-AAA (15 m × 0.25 mm ID × 0.1 µm film thickness, Phenomenex). The oven temperature was initially held at 70 °C for 1 min, raised at 30 °C/min to 300 °C, and held for 3 min. The column flow was 1.1 mL He/min. The temperature of the PTV Injector was set at 50 °C for 0.5 min and ramped at 12 °C/s to 320 °C (5 min). A chemically inert SILTEC liner (Gerstel) was used. The transfer line to the mass spectrometer was kept at 310 °C. The MS was operated in scan (*m/z* 50-420) and SIM mode. For SIM, appropriate ion sets were selected and two characteristic mass fragments of the derivatized amino acids were used for almost all amino acids, except for the labeled amino acids.

5.3.4 Miscellaneous

Solvent evaporation was performed on a Combi Dancer Infra-Red Vortex-Evaporator (Hettich AG, Baech, Switzerland).

Shake flasks for cultivation of *E. coli* were incubated in an Innova 4430 Incubator Shaker (New Brunswick Scientific GmbH, Nuertingen, Germany).

Optical density (OD) measurements were performed on an Ultrospec 3100 pro UV/VIS Spectrophotometer (Amersham Biosciences, Buckinghamshire, UK).

For the sterilization of all materials utilized for the cultivation and preparation of bacterial samples a steam sterilizer (Varioklav, Thermo Fisher Scientific Inc., Waltham, MA, USA) was used.

5.4 Software

Analysis of the acquired raw data generated on the GC×GC-TOFMS instrument was performed using the LECO ChromaTOF software versions 3.34 to 4.32. For metabolic fingerprinting, regions of the 2D chromatogram containing excessive noise were excluded. Baseline correction, deconvolution, and peak picking were performed. Signals exceeding a predefined *S/N* were selected and combined in the second dimension separation using a spectral matching factor. The definition of a separate *S/N* for 2nd dimension subpeaks was an improvement of ChromaTOF version 4 over earlier versions. An override of the allowed 2nd dimension retention time shift was set to improve the separation of closely eluting peaks with an identical 1D retention time and a similar 2D RT. For compound identification, commercial standards were run individually and EI spectra were matched against both an in-house library (for details see Table S2 in the appendix) and the NIST 05 library. After data processing a corresponding peak table was obtained for each sample. Various characteristics per feature, such as name, retention times, area, similarity, *S/N*, quant mass, and unique mass were set. Note that unique mass refers to a metabolite-specific unique *m/z* ion trace, which exhibits the least amount of interferences compared to other *m/z* ion traces available in a compound's fragment spectrum. Quant mass represents an assigned *m/z* ion trace to calculate a compound's peak area/height.

5.5 Validation Methods

5.5.1 LOD and LOQ

The LOD was defined as the concentration producing an *S/N* of at least 3:1. The linear range was determined by the lower (LLOQ) and the upper limit of quantification

(ULOQ), which were defined as the lowest and highest point of the calibration curve with an accuracy between 80 % and 120 %, in agreement with the FDA Guide for Bioanalytical Method Validation.[163] In addition, peaks had to exceed an *S/N* of 8:1 in case of the LLOQ.

5.5.2 ROC curve

Receiver operator characteristic (ROC) curves were used to graphically present the compromise between the false positive and false negative rates. The plots show the false positive rate on the x-axis and the true positive (1-false negative) rate on the y-axis. A discrimination of different abundance levels is good if a small false positive rate and a small false negative rate are obtained. A ROC curve can be quantified by determining the area under the curve (AUC). The closer the AUC is to 1, the better is the differentiation of the various levels.

5.5.3 Bland-Altman plot

Bland-Altman plots were applied to assess the agreement between different methods.[164] The Bland-Altman analysis is a graphical model, also known as Tukey mean difference plots, which displays for each sample the concentration difference y_n between data obtained by two analytical methods or sampling techniques *a* and *b* (y = c_a - c_b) against the average x_n of the two concentrations (x_n = $(x_a + y_b)/2$). Every Bland-Altman plot includes three horizontal lines that mark the mean difference and the upper and lower limits of agreement which are defined as mean difference ± 1.96 times SD of the difference ($\bar{d} \pm 1.96$ SD). An ideal agreement between two datasets is observed, if the absolute mean difference is zero.

6 Development of an alignment tool for GC×GC-TOFMS comparative metabolic fingerprinting

6.1 Introduction

As described in chapter 4.2.2, comparative analysis of metabolic fingerprints requires a reliable and preferentially automated data alignment to recognize identical metabolites across many samples. Initial algorithms for GC×GC-TOFMS data alignment have been developed, but were either based on raw data using retention times only or limited to a fixed retention time window. Furthermore, the methods were computationally intensive and lacked quantitative information rendering a global discriminative study impossible.

This project was performed in cooperation with Inka Appel (Institute of Functional Genomics, Chair of Statistical Bioinformatics, University of Regensburg), who as part of her Ph.D. thesis focused on the bioinformatics work and, in particular, the development of an algorithm for normalizing retention times and aligning GC×GC-TOFMS data with the goal of combining multiple measurements into one data matrix for subsequent multivariate statistical analysis. The author of the present thesis, on the other hand, performed all experiments related to the validation and biomedical application of the algorithm. Their equal contributions to this body of research are documented in two joint first authorships in *Analytical Chemistry* [107] and the *Journal of Chromatrography A* [108].

Starting point were the data processing functions provided by the LECO ChromaTOF software, such as baseline correction, mass spectral deconvolution, peak picking, signal/noise filtering, peak integration, and library search. They were used to generate peak lists for each sample. Retention time variations in the complete series of measurements were compensated by fitting linear models to each retention time plane in every sample. In a subsequent step, the peak lists were aligned in one data matrix based on 1D and 2D retention times as well as spectral information (number of fragment ions and relative intensities). The final data matrix contained peak intensities of the universal *m/z* 73 ion for all features as a quantitative measure. Multivariate analysis and testing strategies were used to visualize the data and determine differences in metabolite abundance.

The accuracy of peak alignment and detection of 1.1- to 4-fold changes in metabolite concentration were validated by a spike-in experiment. The tools were further applied to the comparative metabolic fingerprinting of a wild type versus a double-mutant strain of *E. coli* lacking the transhydrogenases UdhA and PntAB, which play important roles in NADPH metabolism. Finally, the data obtained were also compared with previously published CE-TOFMS data.[165]

This chapter was published in [107].

6.2 Material and Methods

6.2.1 Sample preparation

6.2.1.1 Spike-in experiment

Hundred microliters of an aqueous-methanolic *E. coli* BL21 extract were transferred to a 2-mL glass vial with a 0.2-mL glass insert, followed by the addition of a mixture of 20 metabolites: 2-hydroxybutyrate, 3-hydroxybutyrate, 2-hydroxy-3-methylbutyrate, 3-methyl-2-oxovalerate, malonate, nicotinate, phenylacetate, dimethylsuccinate, decanoate, mandelate, adipate, erythritol, phenyllactate, triethanolamine, dodecanoate, suberate, xylitol, vannilate, mannitol, and eicosanoate.

Blank samples containing only *E. coli* extract and seven spike levels were generated with 0.25, 0.275, 0.3125, 0.375, 0.5, 0.625, and 1.0 nmol absolute of each analyte. Each sample was prepared in six replicates. Samples were evaporated and derivatized as described in chapter 5.2.1. In total, 42 spiked and six blank samples were measured in random order by GC×GC-TOFMS.

6.2.1.2 Bacterial strains and growth conditions

The *E. coli* wild type strain MG1655 and the mutant UdhA-PntAB (MG1655 Δ*Udh*-Δ*PntAB*) were cultured at 37 °C in 250-mL shake flasks containing 30 mL of LB and harvested after reaching stationary phase growth. Cell dry weight (CDW) was determined by vacuum filtration of 10 mL medium containing cells at different optical densities. The filters were dried at 120 °C and weighed empty and with dried cells. Cells were harvested by fast filtration as previously described.[165] Briefly, 2-10 mL of cell suspension (depending on OD) were filtrated by vacuum filtration using

polyethersulfonate (PESU) filters (0.45 µm pore size, 25 mm, Sartorius AG, Göttingen, Germany). The filters were then washed with an identical volume of NaCl solution (0.9 % (w/v), room temperature), before they were submersed in 3 mL of quenching/extraction solution in 50-mL Falcon tubes (Fisher Scientific, PA, USA). The extraction solution consisted of 80 % (v/v) methanol in water at -20 °C and contained the extraction standards [U-^{13}C]lactate, [U-^{13}C]pyruvate, [U-^{2}H]succinate, [2,3,3-^{2}H$_3$]malate, [U-^{13}C]glucose, [^{2}H$_7$]cinnamate, and norvaline at 1 mM each. The tube was vortexed for 60 s; the filter was removed and checked visually for complete dissolution of cells. Falcon tubes were transferred to liquid nitrogen for 3 min, thawed in an ice bath for 15 min, and briefly vortexed. This freeze-thaw cycle was repeated three times for complete cell disruption. The sample was centrifuged at 4 °C and 3,375 xg for 5 min and the supernatant transferred to a new tube. Pellets were re-extracted twice with 0.5 mL of 80 % (v/v) methanol at -20 °C and the extracts were combined. The extracts were concentrated in a vacuum evaporator, transferred into a 2-mL glass vial with a 0.2-mL glass insert, and evaporated and derivatized as described in 5.2.1. The metabolic fingerprints of the *E. coli* wild type MG1655 and the *E. coli* MG1655 double-mutant Δ*Udh*-Δ*PntAB* were compared by cultivating each strain in three separate flasks and by filtrating each culture in triplicates. In total, nine samples per strain and three medium blanks were measured in random order using GC×GC-TOFMS.

6.2.1.3 Calibration

Calibration was performed using metabolite standard solutions in methanol. One hundred µL of different calibration levels were transferred to a glass vial with glass insert and 20 µL of the surrogate solution containing [U-^{13}C]lactate, [U-^{13}C]pyruvate, [U-^{2}H]succinate, [2,3,3-^{2}H$_3$]malate, [U-^{13}C]glucose, [^{2}H$_7$]trans-cinnamate, and norvaline (1 mM each) were added. The standards were dried and derivatized as described in 5.2.1. In total, twelve calibration points were generated over a concentration range of 0.5-1000 µM.

6.2.2 Data processing

Raw data were processed with the LECO ChromaTOF software version 3.34. Regions of the 2D chromatogram containing excessive noise were excluded. Baseline correction, deconvolution, and peak picking were performed. Signals with a

predefined *S/N* of ≥500 were selected and combined in the second dimension separation using a spectral matching factor of 700. An override of the allowed 2nd dimension retention time shift was set at 0.150 s early and 0.050 s late. For compound identification, commercial standards were run individually and EI spectra were matched against both an in-house library and the NIST 05 library. Every sample was processed as described above and a corresponding peak table was generated. As a quantitative measure the peak area of the *m/z* 73 ion trace was integrated. For each chromatogram the peak list was exported in a character separated value (csv) file format containing peak names, 1D and 2D retention times, area integrals for *m/z* 73, similarity values of EI spectra to the mass spectral libraries, and the deconvoluted mass spectra containing only fragment ions with a relative intensity ≥5 % of the base peak.

6.2.3 Scaling

To account for shifts in retention times, linear models were fitted to the series of measurements. For 1D and 2D retention time two independent linear models for each measurement were learned separately. To that end, retention times for odd-numbered fatty acids included as derivatization standards in each measurement were extracted. A given linear model consists of a scaling factor and an offset, which best correlate the observed retention times for the standards in the corresponding measurement to the median across all measurements. The fatty acids used as standards typically had well separated peaks that occurred uniformly over the entire separation space except for the early part of the chromatogram, where smaller odd-numbered fatty acids could not be included as standards because of their natural occurrence in *E. coli*. This ensured that the linear models were well supported across a wide range of retention times. By applying the scaling parameters to the corresponding retention times of all peaks, possible shifts were compensated. The code was implemented in the statistical computing package R (http://www.r-project.org).[166]

6.2.4 Alignment

After retention time scaling, the peak lists were aligned in one data matrix based on 1D and 2D retention times as well as spectral information. In spite of scaling, small

shifts in retention times remained. Similarly, EI spectra varied in the number of *m/z* ions detected and relative signal intensities over time. These were taken into account by four user-defined tolerance parameters, one for each retention time, number of fragment ions with a relative intensity ≥5 % of the base peak, and relative intensities of the respective fragment ions. The generated matrix contained areas detected for *m/z* 73 and had one column per measurement and one row per feature, where a feature corresponded to a set of aligned peaks. Each feature was characterized by the average and interval of 1st and 2nd dimension retention times and its EI spectrum consisting of *m/z* ions with scaled ion abundances. Ion abundances of spectra were always scaled so that the highest abundance was 999.

To generate the alignment matrix, all the peak lists for each measurement were read into a raw peak list and this list was then sorted according to retention times. Peaks of the complete peak list were added one by one into the alignment matrix. A raw peak was aligned to a feature in the alignment matrix, if all tolerance parameters were satisfied. Otherwise a new feature containing the peak's retention times and spectrum was generated. A retention time was considered close enough, if its distance from the entry's corresponding mean of interval boundaries was smaller than the corresponding tolerance parameter. Two spectra were defined as similar if their fragment ions had a sufficient overlap and the difference in ion abundances of the overlapping ions was small. Upon successful alignment, a new union set of fragment ions and respective relative abundances was generated. Finally, the appropriate area detected for *m/z* 73 was stored into the aligned data matrix. In this matrix, entries that corresponded to undetected features were set to zero. The peak alignment module was implemented in R and could be accessed via the shell. The pseudocode can be found in Figure S1 in the appendix.

The module for retention time scaling and peak alignment was implemented by Inka Appel and therefore named INCA – Integrative Normalization and Comparative Analysis.

6.2.5 Normalization, testing and validation

During measurements, small variations in the amount of sample injected might be caused by a number of environmental effects such as humidity or temperature. To compensate for such fluctuations, we used the area integral of the *m/z* 271 trace of

tridecanoic acid, which had been added to all samples as an internal derivatization standard of known concentration (1 mM), to normalize all peak areas prior to log-transformation.

Spectral features distinguishing the wild type from the mutant *E. coli* strain were selected using a t-test with equal variances. Since thousands of features were tested simultaneously, small p-values occurred by chance and false positive features were expected. The false discovery rate of a list of features was the expected relative frequency of false positives in it. False discovery rates were estimated according to Benjamini and Hochberg [113] using the R package MULTTEST.[167]

Metabolic fingerprinting aims to decide, whether a biological sample belongs to one of two predefined classes (e.g. wild type vs. mutant), exclusively based on a vector of metabolite abundances. The problem is embedded into the realm of supervised machine learning. Hence, the nearest shrunken centroid learning algorithm was used which originally had been developed for gene expression fingerprinting by means of DNA microarrays.[168] The inputs of this method are vectors of metabolite abundances from both sample classes, the so-called metabolic fingerprints; the output is a metabolic signature. The signature is a rule converting fingerprints into class predictions. In order to evaluate fingerprinting performance objectively, a leave-one-out cross-validation scheme was applied. In turn, each sample was excluded from the data, a signature was learned from the remaining data, and the excluded sample was predicted. The resulting error rate estimated fingerprinting accuracy in practice. Testing and validation was performed by Inka Appel using the statistical computing software R.

6.3 Results

6.3.1 Method validation via spike-ins

INCA was validated by spiking twenty standard compounds into an *E. coli* wild type strain extract at seven different concentration levels, each of which was prepared in six replicates and measured in random order by GC×GC-TOFMS. The total number of features available for optimizing and testing the retention time correction and alignment tools was 800 after exclusion of two runs because of technical problems.

A series of steps were performed to compile a metabolic fingerprint from the raw data in a single peak list (Figure 6.1). Using the LECO ChromaTOF software, first the regions of the 2D chromatogram containing excessive noise were excluded, thereby cutting on average the number of features per sample in half. In a second step, the ChromaTOF software was used for baseline correction, the merging of peaks generated in the second dimension separation using spectral matching, and the exclusion of signals not exceeding a predefined S/N of ≥500.
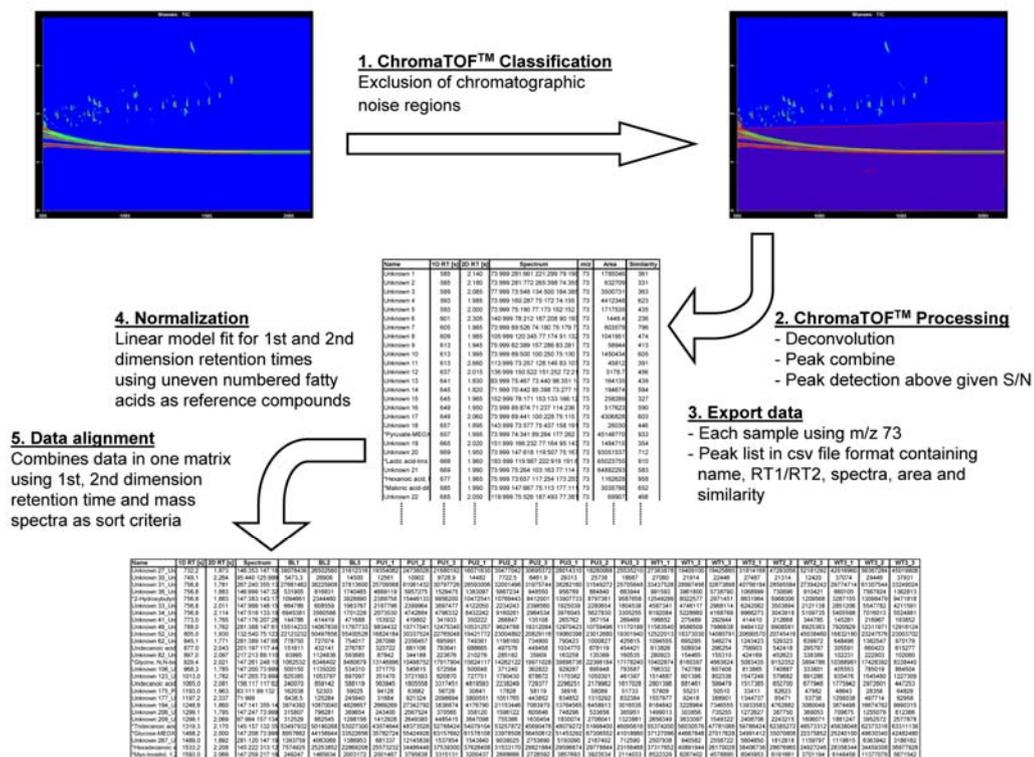
Figure 6.1. Workflow of the various processing steps. Reproduced from [107].

The EI spectra of trimethylsilyl derivatives showed a typically abundant ion with m/z 73 that corresponded to the trimethylsilyl cation $[(CH_3)_3Si]^+$, while derivatives with at least two TMS groups always yielded a strong signal at m/z 147 corresponding to the pentamethyldisiloxane cation $[C_5H_{15}Si_2O]^+$.[49] This characteristic fragmentation behavior was exploited for data analysis by using only the area integrals for m/z 73 as a quantitative measure for all peaks in the peak lists. A compound-specific unique m/z ion could not be used, as the ChromaTOF software was unfortunately unable to constantly assign the same abundant unique mass to an identical compound in the different peak lists. Deconvoluted spectra with a mass threshold of ≥5 % were used accordingly. After exporting the peak lists of each sample, the similarity measure was

used to automatically locate the odd-numbered fatty acid standards in the peak list, which always showed similarities of >95 %.

Prior to alignment, retention times were normalized by fitting linear models separately to the first and second dimension retention times using the odd-numbered fatty acids as reference points. Afterwards, peak alignment was carried out using consecutively first dimension retention time, second dimension retention time, *m/z* values, and relative ion intensities of EI spectra as sorting criteria.

In order to apply the peak alignment effectively, optimal tolerance parameters were determined by testing a total of 288 parameter combinations on the 20 standard compounds that had been spiked into an *E. coli* wild type extract (Table 6.1). Sixteen of the parameter settings were able to align the 20 standard compounds in a single row each, yielding 800 out of 800 true positives. The number of true positives depended more on the tolerance parameters chosen for spectra than for retention times.

Table 6.1. Parameters and associated tolerances tested for feature alignment. t_1st and t_2nd correspond to first and second dimension retention time, while t_ol and t_rate correspond to relative overlaps of fragment ions and their relative signal intensities for a given feature.

Parameter	Tolerance
t_1st [s]	4, 8, 12
t_2nd [s]	0.025, 0.05, 0.075, 0.1, 0.125, 0.15
t_ol	1.0, 0.9, 0.8, 0.7
t_rate	0.1, 0.2, 0.3, 0.4

The optimal parameter setting chosen for subsequent analysis of features was 8 s and 0.1 s for 1st and 2nd dimension retention time, respectively, and 90 % and 40% overlap of *m/z* values and relative ion intensities, respectively. Using this setting, a total of 4,726 features remained of the GC×GC-TOFMS spike-in data. However, some features were not detected in all samples, possibly indicating noise. Therefore, only features detected in at least 50 % of all samples were selected for further analysis. This feature reduction resulted in 517 features with a maximum of 17 % zero values. This was highly compatible with multivariate analysis according to Gika et al., who had suggested that up to 50 % zero values might be tolerated.[169] Zero values were set to the minimum area count assuming that peaks were missing

because they fell below the background noise. The peak areas of the features were then normalized with the area integrals of tridecanoic acid at m/z 271 and subjected to a log transformation for further validation.

The spike-in dataset allowed for the comprehensive evaluation of the ability of GC×GC-TOFMS together with INCA to detect fold changes. When comparing two groups of samples with different spike-in levels, only features spiked into the matrix should vary. Pairwise t-tests of each spike-in level with the biological background data were performed to detect differential features assuming equal variance in both groups. Sorting the list of features in decreasing order according to the absolute value of the t-statistics put differential features on top of the list that should ideally be represented by the spike-in metabolites. For each rank of the t-statistic, true positives and negatives, and respectively false positives and negatives could be determined by comparing the features up to the respective rank to the actual list of spike-ins. The efficiency across all thresholds is illustrated in ROC curves. Here the false positive rate (FPR, or 1-specificity) is plotted versus the true positive rate (TPR, or sensitivity) for all possible t-statistic ranks (Figure 6.2). While it proved impossible to discriminate a 1.1-fold change from the background, higher fold changes yielded increasing empirical accuracies (92 % and 96 % for the 2- and 4-fold change, respectively). This matches accuracies reported for gene expression data.[170]

Hence, genuine changes in metabolite concentration could be detected against a background of thousands of spectral signals subject to random measurement fluctuations that arise in the analysis of complex biological matrices. Next, we asked whether fold changes could be reproduced quantitatively. Linear dependencies between expected and observed fold changes were observed. However, the slope and offset of regression lines differed from metabolite to metabolite. Figure 6.3 exemplifies the spike-in data for 2-hydroxybutyrate. Thereby, all possible pairwise fold changes were plotted. Corresponding figures for the entire spike-in dataset can be found in Figure S2 in the appendix. Reproducibility of fold changes was assessed using univariate linear regression on the integrals of both the universal m/z 73 fragment ion trace (Fig. 6.3A) and the analyte-specific fragment ion trace (Fig. 6.3B) for each of the 20 different spiked analytes investigated over the fold changes.
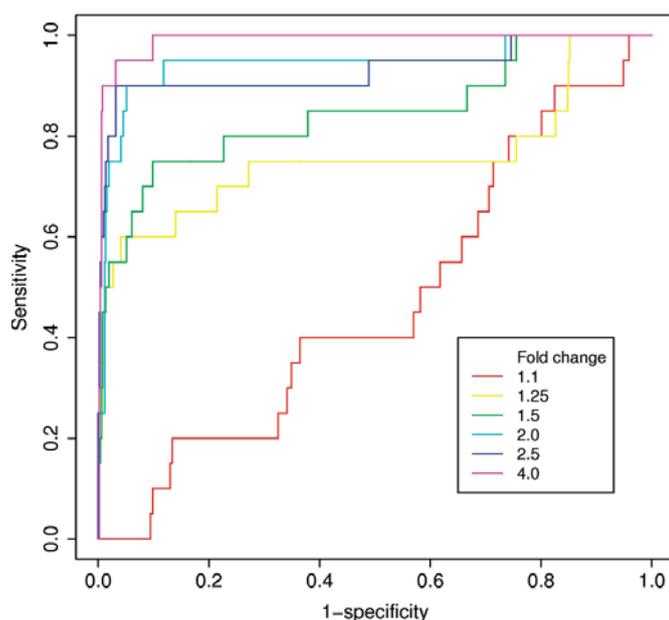
Figure 6.2. ROC curves for different fold changes. Sensitivity and 1-specificity improve with increasing fold change. Reproduced from [107].

Regression equations, regression coefficients, and relative standard deviations (RSDs) are given in the figures. The correlation between observed and spiked-in fold changes increased when integrating a unique mass trace rather than the *m/z* 73 fragment ion trace. The ranges of regression coefficients and RSDs improved from 0.630 – 0.985 and 7.8 – 52.9 % to 0.874 – 0.993 and 5.0 – 19.3 %, respectively. In summary, fold changes could be reproduced up to linear transformations.
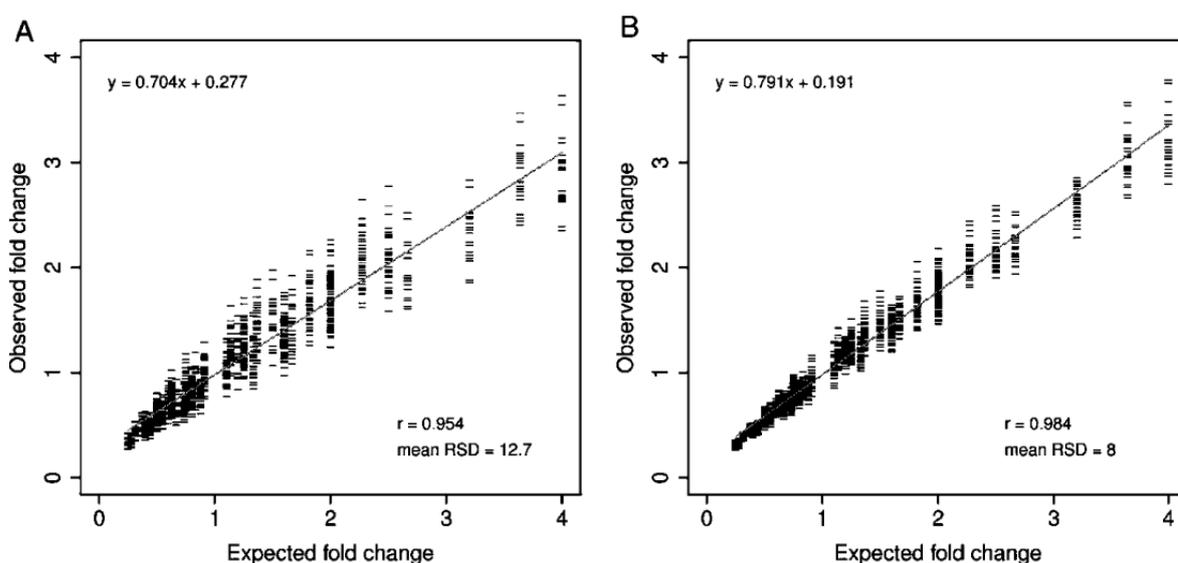
Figure 6.3. Linear dependency between expected and observed fold changes of 2-hydroxybutyrate for both (A) the universal *m/z* 73 fragment ion trace and (B) the analyte-specific fragment ion trace *m/z* 205. All possible pairwise fold changes are plotted. Reproduced from [107].

6.3.2 Comparison of *E. coli* strains

Figure 6.4 shows a comparison of a 1D total ion chromatogram of the mutant *E. coli* strain (Fig. 6.4A) with 2D total ion chromatograms of the mutant (Fig. 6.4B) and the respective wild type strain (Fig. 6.4C).
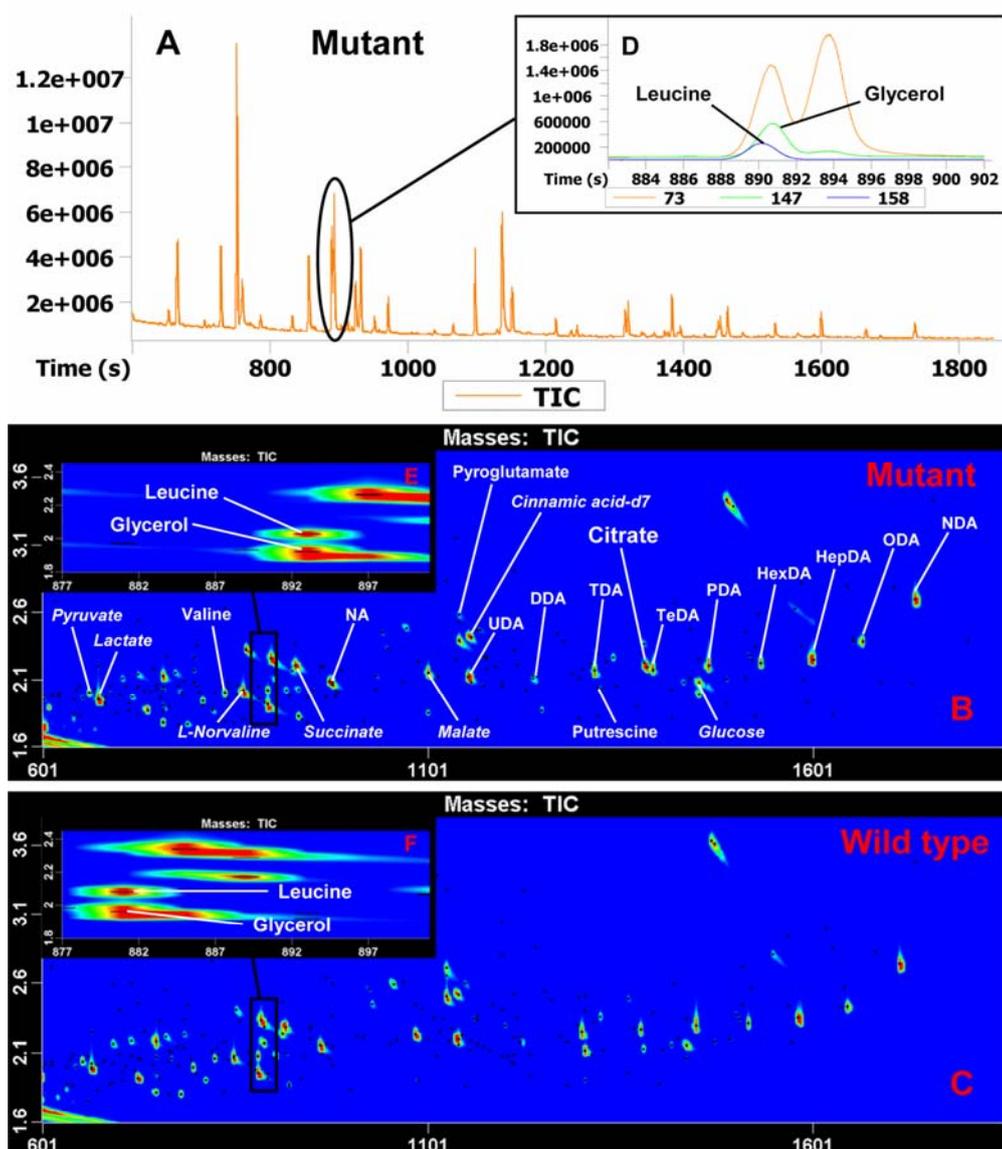
Figure 6.4. (A) TIC chromatogram of mutant Δ*UdhA*-Δ*PntAB E. coli* MG1655 strain using 1D-GC-TOFMS. (B) TIC chromatogram of the same mutant strain sample using GC×GC-TOFMS. (C) GC×GC-TOFMS separation of the wild type *E. coli* MG1655 strain. Major metabolites are labeled and the added internal standards are italicized. Selected compounds, in particular leucine and glycerol, are shown in the small windows D, E, and F. Even without data analysis, the comparison of mutant and wild type sample shows an apparent difference in the abundance of citrate (B vs. C). Enhanced separation is achieved by using GC×GC (E, F) instead of 1D-GC (D). Furthermore, retention time shifts are observed that require scaling (E vs. F). Reproduced from [107].

The advantage of GC×GC over conventional 1D-GC is exemplified for the separation of glycerol and leucine. In 1D-GC-TOFMS, the two compounds could only be resolved by MS on the basis of the characteristic EI fragment ions of m/z 147 and 158 for glycerol and leucine, respectively (Fig. 6.4D). In contrast, glycerol and leucine are baseline separated in the second dimension of GC×GC-TOFMS (Fig. 6.4E and F). Since ChromaTOF failed to deconvolute the signals of glycerol and leucine consistently in the 1D-GC-TOFMS spectra, the latter method failed, in contrast to GC×GC-TOFMS, to detect the significant difference in leucine abundance between the two strains.

Following exclusion of peaks with S/N <500 and retention time scaling, peak alignment was carried out using 1D and 2D retention times and the deconvoluted mass spectra as sorting criteria. After alignment, a total of 2,259 features remained (Table 6.2). Only features that had been detected in at least 50 % of all samples (9 out of 18) were further analyzed. This arbitrary cut-off was selected with respect to the group-size of 9 to ensure that features not present in one group but in the other were not excluded from the list. A total of 398 features with a maximum of 25 % zero values remained.

Table 6.2. GC×GC-TOFMS data for comparison of the *E. coli* strains. Reproduced from [107].

Total features	2259
Reduced features[1]	398
Significant features[2]	48
Metabolites identified that distinguished the two *E. coli* strains	pyruvate, citrate, cis-aconitate, succinate, α-ketoglutarate, fumarate, malate, glucose, glucose-6-phosphate, myo-inositol, indole, GABA, glycine, proline, lysine, leucine, isoleucine, valine, pyroglutamate, putrescine, itaconate, aminoadipate, tartrate, orotate, furoate, caproate, sucrose

[1] Features exceeding S/N ≥500 in 9 out of 18 samples.

[2] Metabolites that differed significantly (FDR <0.05) in abundance between the strains.

Zero values might result from (i) peaks that were actually not present, (ii) signals falling below predefined S/N thresholds, (iii) features not matching the alignment parameters, (iv) ineffective deconvolution of raw mass spectra, and (v) peaks that were not identified according to initial data processing settings. Especially, if values for almost a whole group, like a fold change level, are missing no inference can be

made about the mean and variance of that group. Zero values were problematic for subsequent statistical analysis and were thus set to the minimum area count assuming that peaks were missing because they fell below the background noise. The features were then normalized with the area integrals of tridecanoic acid at *m/z* 271, subjected to a log transformation and a PCA (Figure 6.5). A clear segregation of the two *E. coli* strains and the blanks was obtained. Moreover, the variability among the *E. coli* strains was notably higher than the variability among the blanks indicating that the metabolic fingerprints captured true biological variation.
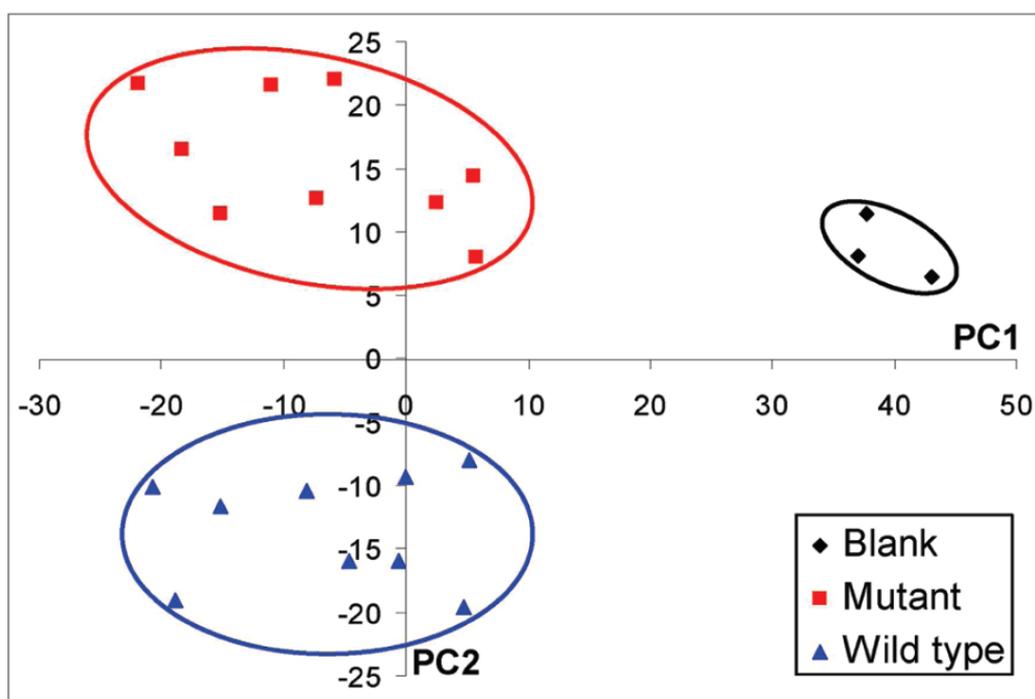
Figure 6.5. PCA of a randomly aligned dataset of *E. coli* mutant and wild type strains. Feature reduction yielded 398 features present in at least 50 % of samples within one group. PC1 accounted for 33 % of the variance, PC2 for 16 %. Reproduced from [107].

A t-test with equal variances identified a list of 48 features that differed significantly between the two strains with an estimated false discovery rate of <0.05, indicating significance also after correction for multiple testing. In other words less than 3 false positives were expected among the 48 identified features. It should be noted that a feature does not necessarily equal a metabolite. Several amino acid signals were identified in the chromatograms corresponding to a single amino acid. It is well known that partial silylation can result in more than one signal for an amino acid. Also, certain metabolites such as those containing a carbonyl group, e.g. glucose, glucose-6-phosphate, and α-ketoglutarate, form two isomers upon methoximation

that are separated by GC. In addition, steric hindrance, degradation, or rearrangement reactions can occur and cause multiple possible analyte structures of the same metabolite.[49, 51] Thus, a metabolite can be represented by more than one feature and *vice versa* a feature can represent more than one metabolite. Twenty-seven metabolites were identified accounting for 33 of the 48 features (Table 6.2).

Significant features do not imply that metabolic fingerprinting is possible. In order to verify that strains can be correctly classified into wild type vs. mutant through their metabolic fingerprints, we learned metabolite signatures using the shrunken centroid classifier by Tibshirani et al.[168] In a leave-one-out cross-validation procedure 100 % of the metabolic fingerprints were classified correctly. Most of the compounds were intermediates of the citrate cycle as expected from our previous analysis of the two *E. coli* strains employing CE-TOFMS.[165] Importantly, the latter method also confirmed, that the additional significant features discovered by GC×GC-TOFMS were true positives.

6.3.3 Comparative quantification using metabolite profiling

Selected metabolites, including pyruvate, succinate, fumarate, malate, α-ketoglutarate, cis-aconitate, citrate, myo-inositol, and glucose-6-phosphate, were quantified using their respective unique mass in a metabolite profiling approach. A calibration was performed using a custom-made metabolite standard solution with different calibration levels. The linear calibration range was 0.5 – 250 µM. Results are depicted in Figure 6.6. A significant increase of TCA cycle intermediates in the mutant compared to the wild type *E. coli* was observed, thus confirming the GC×GC-TOFMS quantitative data based on the universal *m/z* 73 fragment ions.

Analytes and standards are listed in Table 6.3. First and second dimension retention times, correlation coefficients, linear calibration ranges, calibration parameters, relative standard deviations of one calibration triplicate, and the average RSDs of biological triplicates are presented. In addition, the average RSDs of biological triplicates using *m/z* 73 is shown. The calibration triplicates of the internal standards were normalized by using the area integrals of tridecanoic acid at *m/z* 271. Utilizing the *m/z* 73, the RSDs of the analytes varied from 13.2 to 26.3 %. In contrast, using the respective unique mass, the RSDs of the analytes ranged from 7.4 to 24.5 % for

metabolite extraction and GC×GC-TOFMS analysis. The RSDs for calibration replicates varied from 7.0 to 15.2 %.

Table 6.3. Figures of merit for the metabolite profiling of significant compounds.[1] Reproduced from [107].

Compounds	1D, 2D retention time [s]	r	Linear equation	RSD calib. tripl. (%)	RSD biol. tripl. *m/z* 73 (%)	RSD biol. tripl. *m/z* U (%)
Analytes						
Pyruvate	661, 1.995	0.999	y= 0.200 x 0.008	7.5	16.7	7.4
Succinate	929, 2.225	0.992	y= 0.106 x 0.008	9.9	19.9	16.2
Fumarate	957, 2.220	0.991	y= 0.102 x 0.017	11.0	21.8	19.7
Malate	1101, 2.165	0.992	y= 0.059 x 0.010	7.0	17.8	16.5
α-Ketoglutarate[2]	1153, 2.320	0.997	y= 0.001 x -4.0e-005	10.3	19.0	14.3
α-Ketoglutarate[2]	1177, 2.270	0.998	y= 0.015x -0.001	11.6	23.5	18.6
cis-Aconitate	1317, 2.330	0.999	y= 0.020 x 0.001	9.8	19.2	18.4
Citrate	1385, 2.190	0.998	y= 0.078 x 0.012	13.0	13.2	17.5
Myo-inositol	1589, 2.085	0.997	y= 0.067 x 0.013	15.2	26.3	24.5
Glucose-6P[2]	1757, 2.770	0.999	y= 0.015 x -0.001	12.5	19.6	12.2
Glucose-6P[2]	1773, 2.870	0.999	y= 0.003 x -4.8e-005	9.5	24.2	16.6
Internal extraction standards						
[U-^{13}C]Pyruvate	661, 1.995	na[3]	na	4.3	16.7	14.4
[U-^{13}C]Lactate	673, 1.960	na	na	7.1	18.8	16.6
Norvaline, OTMS	721, 2.155	na	na	3.1	14.9	18.7
Norvaline, N-, OTMS	857, 2.015	na	na	8.2	12.6	20.1
[U-^{2}H]Succinate	925, 2.250	na	na	4.7	7.5	7.2
[2,3,3-^{2}H$_3$]Malate	1097, 2.145	na	na	7.7	14.4	11.6
[^{2}H$_7$]trans-Cinnamate	1149, 2.465	na	na	9.5	13.8	15.9
[U-^{13}C]Glucose[2]	1453, 2.025	na	na	9.8	13.5	10.5
[U-^{13}C]Glucose[2]	1469, 2.020	na	na	10.7	17.1	22.0
Internal derivatization standards						
Nonanoate	973, 2.060	na	na	2.4	15.2	8.6
Undecanoate	1153, 2.140	na	na	2.3	9.5	6.3
Tridecanoate[4]	1317, 2.185	na	na	na	na	na
Pentadecanoate	1465, 2.220	na	na	1.7	7.4	6.4
Heptadecanoate	1601, 2.275	na	na	2.6	12.4	9.0
Nonadecanoate	1737, 2.725	na	na	4.7	11.0	10.0

[1] linear calibration range was 0.5 – 250 µM.

[2] form isomers upon methoximation.

[3] na, not applicable.

[4] used to normalize internal standards.
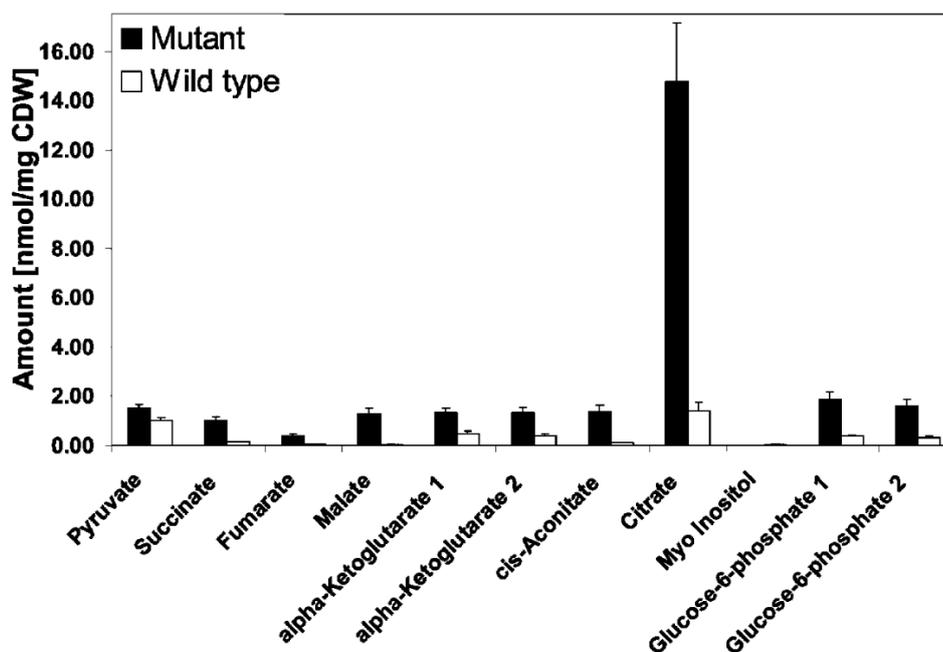
Figure 6.6. Absolute quantification of a selected set of metabolites, whose abundance differed significantly (FDR of <0.05) between *E. coli* mutant and wild type strain using their respective unique mass. Reproduced from [107].

6.3.4 Comparison of GC×GC-TOFMS and CE-TOFMS data

GC×GC-TOFMS data obtained were compared with CE-TOFMS data reported previously by Timischl et al. [165] for the same *E. coli* strains. In total, Timischl et al. made 58 high-confidence identifications of anionic metabolites in the two *E. coli* strains grown in LB medium and harvested in the stationary phase. Out of the 58 metabolites, 37 showed significant differences in abundance between the strains based on a Student´s t-test (p <0.05) without correction for multiple testing. Eighteen of the 37 metabolites were also found by GC×GC-TOFMS to differ significantly with an FDR of <0.05 (see Table S3 in the appendix). Alanine, aspartate, benzoate, glucosamine-6-phosphate, glucuronate, glycerolphosphate, octanoate, pantothenate phosphoenolpyruvate, phosphoglycerate, and glutamate (a total of 11) showed no significant difference by GC×GC-TOFMS, but had been found to differ significantly by CE-TOFMS. However, glutamate is known to partly lose H_2O and rearrange to pyroglutamate, which is also formed from glutamine through loss of NH_3.[51] This explains why pyroglutamate differed significantly between the strains in the GC×GC-TOFMS analysis. Finally, GC×GC-TOFMS failed to detect 8 metabolites, namely UMP, UDP, UDP-glucose, CMP, dTMP, AMP, glutathione, and maltose. The majority of those compounds are nucleotides that are not amenable to GC analysis. In

contrast to CE-TOFMS, GC×GC-TOFMS found α-ketoglutarate and tartrate to differ significantly between the strains. Finally, we were able to detect significant differences for glycine, indole, myo-inositol, and putrescine, which had not been detected by CE-TOFMS, and for GABA and furoate, which had not been confirmed by a commercial standard in the CE analysis. A complete comparison between GC×GC-TOFMS and CE-TOFMS is given in Table S3 in the appendix.

6.4 Discussion

6.4.1 Metabolic fingerprinting

The GC×GC-TOFMS based metabolic fingerprinting approach presented in this chapter, in combination with classification and testing algorithms, provides an effective and simple means to classify samples according to metabolite abundances, potentially making it a valuable tool in many application areas, such as medical diagnosis and prognostication. The retention time correction and peak alignment algorithm INCA facilitates the comparison of fingerprints by (i) fitting retention times of unknown features to linear models using the homologous series of odd-numbered fatty acids as reference points, and by (ii) combining peak lists into a single data matrix based on first and second dimension retention times and spectral information. The algorithm further capitalizes on the characteristic fragmentation behavior of silylated metabolites upon EI ionization yielding a true mass spectrum, which can be used in library searches for identification purposes. In addition a trimethylsilyl cation with a characteristic *m/z* 73 ion is obtained, the integral of which can be used as a quantitative measure for all features detected above a predefined *S/N*.

INCA particularly benefits from gathering the complete information from all peak lists. Thereby no data is lost and the decision about further data handling is left to the user. Furthermore, INCA is capable of automatically selecting optimal alignment tolerance parameters based on spiked-in compounds for any biological matrix, which proves to be a further user-friendly attribute.

It must be pointed out that INCA, like MSort®, or the later developed DISCO or Guineu (for details see chapter 4.4.2) use automatically processed data by vendor software as input and, therefore, rely on the quality and transparency of the provided data. Alignment as well as quantification can suffer from well-known problems like

ineffective deconvolution, peak picking, peak merging and peak integration, or signals falling below the predefined *S/N* threshold. This results in missing values which need to be handled carefully for subsequent statistical analysis.

A disadvantage of ChromaTOF, in particular, is the lack of quantification options for metabolic fingerprinting. A compound-specific unique *m/z* ion could not be exploited for external data alignment due to the inability of the ChromaTOF software to constantly assign the same abundant unique mass to an identical compound across all peak lists. However, as shown in this chapter, while use of the universal *m/z* 73 fragment ion for all compounds in a sample matrix might yield increased variance compared to the use of compound-specific unique *m/z* ions, it has proven sufficiently reliable for quantification purposes and, therefore, represents a reasonable alternative.

The *E. coli* mutant versus wild type experiment was repeated together with Christian Wachsmuth (Institute and Chair of Functional Genomics, University of Regensburg).[171] Extracts were analyzed by both GC×GC-TOFMS and GC-APCI-TOFMS. Thirteen metabolites could be confirmed significant by both methods including all TCA cycle intermediates and the related itaconate. As a group, those analytes had been regulated most significantly (for more details see Table S4). Six metabolites, namely thymine, dihydroorotate, ornithine, N-acetylaspartate, N-acetylneuraminate, and N-acetylputrescine, were assigned using the GC-APCI-TOFMS data and a *de novo* identification approach.[171] These six metabolites were not identified by the initial GC×GC-TOFMS study. The identification of derivatives of dihydroorotate, N-acetylneuraminate, and N-acetylputrescine proved difficult as their characteristic EI spectra were contained in neither the NIST 05 nor the Fiehn metabolite library. N-acetylaspartate had already been identified as significantly different using CE-TOFMS metabolic fingerprinting.[165]

6.4.2 Metabolite profiling

The performance of GC×GC-EI-TOFMS for metabolite profiling was comprehensively evaluated and within the scope of the master thesis of C. Wachsmuth compared to GC-EI-TOFMS, GC-APCI-TOFMS, GC-EI-qMS, and GC-CI-qMS in terms of reproducibility, dynamic range, limit of detection and quantification using a standard mixture of 43 metabolites from different chemical classes and metabolic pathways

and 12 stable isotope-labeled standards (for details see Table S5 in the appendix). The results were published in Wachsmuth et al.[171]

The standard mixture was prepared from separate stock solutions by C. Wachsmuth. It included each compound at a concentration of 1 mM in methanol and was serially diluted covering 31 points over a concentration range of 0.002 – 3000 µM. GC×GC-EI-TOFMS raw data were processed with the LECO ChromaTOF software v4.32 according to chapters 6.2.2 and 7.2.1, respectively.

For calibration, compounds were normalized by the corresponding stable isotope-labeled standard or by the closest eluting or structurally similar internal standard compound. A characteristic fragment ion was selected as quantifier for each compound according to Table S5 in the appendix. Calibration curves were obtained for every standard by plotting the response as a function of the absolute amount of standard in nmol. Whenever a certain analyte gave two peaks, e.g. cis-trans isomers of methoximated compounds or partly silylated amino acids, only the more intense one was used.

In Table 6.4, linear ranges and RSDs of calibration replicates are given for the 43 analytes and the five techniques. R square values were classified in three groups (R^2 >0.995, 0.990<R^2<0.995, 0.980<R^2<0.990) and the respective group is indicated for each analyte and technique. The exact R^2 values are given in Table S6 in the appendix. Linear regression analysis revealed overall excellent R^2 values exceeding 0.995 for more than 70% of analytes in case of GC×GC-EI-TOFMS, GC-APCI-TOFMS, and GC-EI-qMS, and for almost half of all compounds in case of GC-EI-TOFMS and GC-CI-qMS. R^2 values below 0.99 were only observed for Ile-2TMS employing GC×GC-EI-TOFMS, and GABA-3TMS, erythritol, and phenylacetate for GC-EI-TOFMS. Based on our results and literature data [84, 172], distinct improvements in linearity are attributed to the use of twelve stable isotope-labeled internal standards, because analytes with a corresponding stable isotope-labeled internal standard performed overall better than the other analytes.

Evaluation of quantification ranges (Table 6.4) highlighted GC×GC-EI-TOFMS as the most powerful among the tested techniques due to its ability to measure concentration levels over three orders of magnitude down to LLOQs in the sub-micromolar range. Compared to 1D-GC-EI-TOFMS, thermal refocusing of analytes eluting from the first column yielded narrow peaks and, consequently, lower LLOQs.

Table 6.4. Linear ranges and RSDs for derivatization replicates (n=5) for master mix compounds analyzed with the five different techniques. Exact R² values are given in Table S6 in the appendix. The degree of silylation is only indicated for metabolites forming more than one derivative. Reproduced from [171].

Compound	GC-APCI-TOFMS LR [µM]	RSD (%)[b]	GC×GC-EI-TOFMS LR [µM]	RSD (%)[b]	GC-EI-TOFMS LR [µM]	RSD (%)[b]	GC-CI-qMS LR [µM]	RSD (%)[c]	GC-EI-qMS LR [µM]	RSD (%)[c]
Ala-2TMS	1.95-250[2]	7.6	0.98-250[1]	6.7	7.81-500[2]	12.0	46.88-2000[2]	10.7	15.63-750[1]	6.5
Val-2TMS	0.49-125[1]	2.5	0.12-500[1]	3.3	7.81-750[2]	11.1	11.72-1500[2]	8.3	3.91-1500[1]	3.2
Leu-2TMS	0.49-250[2]	8.6	3.91-375[2]	3.8	15.63-250[2]	6.0	11.72-1000[1]	1.8	31.25-750[2]	4.0
Ile-2TMS	0.49-187.5[1]	6.5	3.91-375[3]	3.7	31.25-250[2]	7.7	23.44-750[2]	10.6	46.88-2000[1]	4.3
Pro-2TMS	7.81-1000[2]	5.4	7.81-750[1]	4.7	46.88-250[2]	5.1	62.5-1000[2]	6.8	46.88-500[2]	4.0
Gly-3TMS	0.49-93.75[1]	3.2	0.12-750[1]	7.2	23.44-750[2]	6.7	15.63-500[2]	2.4	11.72-3000[1]	3.8
Putrescine-4TMS	0.98-93.75[1]	2.5	11.72-750[2]	7.5	62.5-750[2]	7.3	15.63-500[2]	8.7	1.95-3000[1]	2.0
GABA-3TMS	1.95-62.50[2]	8.7	0.98-500[2]	3.8	23.44-250[3]	5.9	15.63-375[2]	17.3	7.81-3000[1]	4.0
Lactate[a]	1.95-750[2]	4.9	0.12-750[1]	1.2	3.91-750[1]	2.3	62.5-2500[1]	1.6	15.63-2000[1]	4.0
2-OH-butyrate	0.24-187.5[1]	6.3	0.12-250[1]	2.3	7.81-750[1]	5.0	62.5-3000[1]	8.0	15.63-3000[1]	3.2
3-OH-butyrate[a]	0.06-187.5[1]	4.9	0.06-750[1]	1.4	3.91-750[1]	2.4	15.63-3000[1]	5.1	3.91-3000[1]	3.7
Me-malonate	0.06-125[1]	3.5	0.24-750[1]	6.2	15.63-250[1]	7.7	7.81-2000[1]	0.9	7.81-2000[1]	3.4
Glycerol	3.91-46.88[2]	3.9[d]	1.95-750[1]	3.2	7.81-125[1]	2.9	7.81-500[1]	3.2	3.91-2000[1]	3.0
Succinate[a]	0.24-93.75[1]	2.5	0.12-187.5[1]	1.5	7.81-250[2]	1.8	31.25-1000[2]	2.1	7.81-3000[1]	1.9
Glycerate	0.06-62.50[1]	4.8	0.06-750[2]	4.6	7.81-750[1]	4.5	31.25-1000[1]	3.1	7.81-2000[1]	2.8
Fumarate[a]	0.49-250[1]	2.9	0.12-750[1]	2.0	0.49-750[1]	2.1	3.91-3000[1]	0.9	1.95-3000[1]	2.0
Malate[a]	3.91-62.50[1]	3.8	0.03-750[1]	3.0	3.91-750[1]	3.2	23.45-1000[1]	6.1	1.95-2000[1]	2.4
Adipate	3.91-93.75[1]	2.7	0.12-375[2]	9.3	15.63-750[1]	8.9	62.5-1500[2]	0.4	7.81-3000[1]	2.6
Erythritol	3.91-93.75[2]	5.4	0.06-93.75[1]	1.2	3.91-500[3]	8.4	46.88-1000[2]	4.0	3.91-3000[1]	2.1
cis-Aconitate	0.12-62.50[1]	1.7	0.06-375[1]	2.2	15.63-750[1]	1.9	46.88-1000[2]	10.4	3.91-1500[1]	5.2
Citrate[a]	1.95-62.50[1]	2.9	0.12-500[1]	2.1	3.91-375[2]	2.9	11.72-2000[2]	7.9	0.98-1000[1]	2.9
Pyruvate[a]	0.24-500[1]	4.7	0.06-750[1]	2.1	15.63-750[1]	3.2	23.44-3000[1]	4.9	3.91-3000[1]	3.5
2-Ketobutyrate	1.95-375[1]	2.1	0.12-250[2]	6.5	31.25-750[2]	14.4	62.5-1000[2]	6.9	7.81-3000[1]	4.2
3-Me-2-oxovalerate	0.24-125[1]	5.0	0.49-375[1]	2.1	15.63-750[1]	2.1	62.5-750[2]	2.2	62.5-1500[2]	4.4
4-Me-2-oxovalerate	3.91-187.5[1]	5.1	0.12-750[2]	3.2	3.91-250[1]	3.7	46.88-500[2]	4.4	7.81-750[1]	1.3
α-Ketoglutarate	0.06-93.75[2]	4.8	0.98-750[1]	4.0	15.63-250[2]	8.0	46.88-750[1]	5.0	7.81-1500[2]	5.3
Phenylacetate	3.91-125[1]	2.0	0.24-750[1]	2.4	23.44-375[3]	2.2	46.88-2000[2]	4.1	31.25-1000[2]	3.2
Phenylpyruvate	0.49-93.75[1]	4.2	3.91-750[1]	4.2	15.63-187.5[2]	4.6	23.44-750[1]	1.7	62.5-3000[2]	2.7
OH-phenylacetate	0.24-62.50[1]	5.4	0.06-375[1]	3.8	0.98-750[2]	11.8	11.72-1500[1]	2.0	7.81-1500[1]	3.6
Homovanillate	3.91-62.50[1]	0.4	0.98-500[2]	3.3	1.95-750[2]	10.6	23.44-2500[1]	2.0	3.91-1500[2]	6.0
Hippurate-1TMS	0.49-93.75[1]	3.3	93.75-750[1]	nd	93.75-750[2]	nd	31.25-750[1]	1.1	46.88-2000[1]	4.6
Homogentisate	0.98-46.88[1]	5.7[d]	0.24-93.75[1]	1.9	3.91-750[2]	11.2	11.72-1000[1]	2.4	7.81-1500[1]	3.8
OH-phenylpyruvate	0.49-93.75[1]	1.9	7.81-750[2]	5.3	15.63-250[2]	9.9	31.25-250[2]	10.2	7.81-250[2]	2.2
5-HIAA-2TMS[a]	0.98-187.5[1]	3.7	0.12-500[1]	1.7	23.44-250[1]	2.0	31.25-750[1]	1.9	62.5-3000[1]	7.6
Phenyllactate	1.95-93.75[1]	1.9	0.12-500[1]	3.9	1.95-750[1]	9.6	46.88-750[1]	1.9	7.81-3000[1]	2.7
Nicotinate	1.95-62.50[1]	2.3	0.98-750[1]	9.6	0.98-750[1]	10.5	3.91-2000[1]	2.0	3.91-2000[1]	2.4
Glycerol-1-P	1.95-93.75[1]	1.1	3.91-750[1]	3.2	3.91-250[2]	3.4	46.88-1500[2]	6.5	7.81-250[2]	1.3
3-P-glycerate	1.95-62.50[1]	1.4	15.63-375[2]	3.4	46.88-375[2]	3.5	46.88-2000[1]	3.9	7.81-250[2]	1.8
Glucose-6P	1.95-125[1]	5.4	23.43-750[2]	6.3	46.88-750[2]	10.7	62.5-1500[2]	8.5	3.91-500[1]	1.9
Lactose[a]	3.91-125[2]	8.2	0.03-750[1]	1.8	3.91-750[1]	2.7	7.81-3000[1]	3.1	1.95-3000[1]	2.5
Myo-Inositol	1.95-93.75[1]	1.4	0.24-93.75[1]	1.4	0.98-250[2]	0.6	23.44-2000[2]	10.8	7.81-2500[1]	3.6
Glucose[a]	0.06-250[1]	3.3	0.03-750[1]	0.7	0.98-750[1]	2.6	23.44-3000[2]	5.7	3.91-3000[1]	3.5
Fructose	0.98-93.75[1]	1.6	0.12-750[1]	1.6	7.81-750[2]	2.3	23.44-1500[2]	9.2	7.81-2000[1]	1.9

[a] Compounds with a corresponding internal standard, [b] conc. = 62.5 µM n=5, [c] conc. = 500 µM n=5,

[d] Standard conc. above ULOQ.

[1] R² >0.995, [2] 0.990<R²<0.995, [3] 0.980<R²<0.990; nd, not detected.

Overall, due to a higher ion transmission, LLOQs were lower for TOF mass analyzers. However, because of detector saturation, signals did not scale linearly up to the highest measured concentration for most metabolites, as it was the case for quadrupole mass analyzers. GC×GC-EI-TOFMS was superior to all other techniques yielding an average linear range of more than three orders of magnitude. LLOQs were in the range of 0.03 to 93.75 µM. GC-APCI-TOFMS showed an average linear range of about one order of magnitude less in concentration with LLOQs ranging from 0.06 to 7.81 µM. GC-EI-qMS and GC-CI-qMS had LLOQs in the range of 0.98 to 62.5 µM and 3.91 to 62.5 µM, respectively. In the case of hippurate and compounds with a phosphate group (glycerol-1-phosphate, 3-phosphoglycerate, and glucose-6-phosphate) the LLOQs for methods with EI were surprisingly high and soft ionization at atmospheric pressure proved favorable for those analytes.

LODs were also compared (see Table S6 in the appendix) and they were in the sub-micromolar range in case of TOF mass analyzers (except GC×GC-EI-TOFMS) and in the lower micromolar range when quadrupole mass analyzers had been used. While GC×GC-EI-TOFMS showed outstanding low LODs in the nanomolar range, GC-APCI-TOFMS performed best among the remaining techniques. Overall, trends for LODs were comparable to LLOQs.

Finally, RSDs for calibration replicates (fivefold derivatized, see Table 6.4) and technical replicates (fivefold injected, see Table S6 in the appendix) of a calibration point were compared to assess the precision of the analytical procedure. For a better method comparison a standard concentration well in the linear range of the respective method was chosen, *i.e.* 500 µM and 62.5 µM for techniques employing quadrupole and TOF mass analyzers, respectively. As can be seen from Table 6.4, GC-EI-qMS showed the best performance with RSDs below 8 % for all compounds. These values are better than those reported in literature[172], especially in case of amino acids, which is probably due to the utilization of more appropriate internal standards. RSDs were below 10 % for GC×GC-EI-TOFMS and GC-APCI-TOFMS and below 12 % for GC-EI-TOFMS and GC-CI-qMS with the exception of 2-ketobutyrate (14.4 %) in case of GC-EI-TOFMS and GABA (17.3 %) in case of GC-CI-qMS. RSDs for technical replicates were in total lower than those for calibration replicates and for more than 80 % of the compounds below 5 %. However, regarding repeatability, amino acids performed worst irrespectively of the GC-MS method employed, because silylation is not an appropriate derivatization method for this

class of compounds as previously reported [172] and more suitable methods should be used.[53, 173, 174] For other chemical groups there was no bias noticed towards high or low RSDs. In case of compounds with a corresponding stable isotope-labeled internal standard, RSDs were distinctly lower.

7 Comparison of two algorithmic GC×GC-TOFMS data processing strategies for metabolic fingerprinting

7.1 Introduction

The SC option for GC×GC-TOFMS data was introduced in 2010 by LECO as an add-on software package to ChromaTOF version 4. SC provides peak alignment, statistical information on various peak properties, and a Fisher ratio calculation module.

Raw data acquired in chapter 6.2.1 for the spike-in experiment and the comparative metabolic fingerprinting of a wild type versus a double-mutant *E. coli* strain were used to evaluate the performance of SC and to compare it to our in-house developed retention time correction and data alignment algorithm INCA. Starting with identical peak lists generated by LECO's ChromaTOF software, the accuracy of peak alignment and detection of 1.1- to 4-fold changes in metabolite concentration was assessed by means of differentially spiked *E. coli* extracts (analogous to Chapter 6.3.1). To provide the same quality input signals for both alignment routines, the universal *m/z* 73 ion trace of the trimethylsilyl group was used as a quantitative measure for all features.

Since SC operates as part of the ChromaTOF software, the feasibility of employing the signal intensities of metabolite-specific unique *m/z* ions for quantification instead of *m/z* 73 ions was studied.

Multivariate statistical analyses were performed in order to evaluate the respective enhancements and limitations of the two algorithmic data processing strategies.

This work was performed together with Inka Appel (Institute of Functional Genomics, Chair of Statistical Bioinformatics, University of Regensburg). The results in this chapter were published in [108].

7.2 Methods

7.2.1 Data processing

Raw data were processed with the ChromaTOF software version 4.32. Parameters were chosen according to chapter 6.2.2. A separate *S/N* for 2nd dimension subpeaks, which was set at ≥50, was an improvement of ChromaTOF v4 over previous versions.

7.2.2 Data alignment

7.2.2.1 Statistical Compare

In the SC feature of ChromaTOF the processed samples were added to a sample table and assigned to their respective groups. The alignment processing method provided two parameters for retention time and mass spectral matching, respectively. RT match criteria were taken into account by a maximum RT difference and a maximum number of modulation periods between peaks. For spectral matching, a mass threshold and a minimum similarity match were defined. Besides a separate *S/N* for peaks not found by the initial peak finding, thresholds for analytes to be kept for statistical evaluation (minimum number of samples or minimum percent of samples in a class that contain the analyte) could be defined. Optionally, it was possible to fill the table with zero values in case that an analyte was not found in a specific sample.

The alignment parameters for retention time and mass spectral matching were chosen in accordance with the optimal parameter setting determined for INCA (see chapter 7.3.1). The maximum RT difference was set at 4 s, the maximum number of modulation periods at 2. A minimum similarity mass spectral match of 80 % was chosen. Further, in the deconvoluted mass spectra only *m/z* ions with relative base peak intensities ≥10 % were considered for alignment. The separate *S/N* for peaks not found by the initial peak finding was set at ≥50. Maximum data output for statistical analysis was obtained by setting the minimum number of samples containing the analyte and the minimum percent of samples in a class that contain the analyte at relatively low values of 1 and 10 %, respectively. Peak alignment was

consequently carried out using 1st and 2nd dimension retention times and mass spectra as sort criteria.

Pairwise sample comparisons for all samples analyzed are made peak by peak. Peaks across samples that are within the retention time window specified in the data processing method and share the best spectral match are grouped together. The spectral match uses the NIST match algorithm, the same as employed in library searches. Following the pairwise comparisons, common peaks are linked from sample to sample to create groups of common analyte peaks. During the grouping conflicts may arise due to the parameters specified in the data processing method (RT window and mass spectral match threshold) and/or variability in peak interferences and RT shifts. The software takes various steps encoded in non-accessible source code to resolve these conflicts. Failure to resolve conflicts results in the exclusion of peaks from the final table.

Each peak is assigned a peak quality based on the peak shape of the extracted unique mass ion chromatogram and the quality of the deconvolution. The peak that represents the best quality within a group of aligned peaks is selected to conduct a library search for that group of peaks and the name of the matching library compound appears in the compound table.

Deconvolution performed during data preprocessing may yield different unique masses for identical compounds in different samples. Nonetheless, SC has to choose a consistent mass for quantification from the unique masses. This quant mass is selected according to the most common unique mass for a specific peak across all samples. The remaining alternating masses are adjusted according to this predominant mass, provided that the peak profile is good. If the peak profile is poor and thus does not allow the extraction of a unique *m/z* ion due to interferences, the ratio of the quant mass to the available unique mass will be determined to calculate peak height and peak area. Alternatively, a quant mass may be assigned by manually replacing the prevalent unique *m/z* ion trace for a feature across all peak lists by the *m/z* ion trace of choice, *e.g. m/z* 73.

The aligned peak table was exported in a csv file format. The table was transformed to obtain a suitable data matrix for multivariate statistical analysis, which had one column per measurement and one row per feature. A feature was characterized by

peak name, average of 1D and 2D retention times, area count, the unified quant mass, and the respective area integrals for each measurement.

7.2.2.2 INCA

For INCA, it was not possible to exploit a compound-specific unique *m/z* ion, as the algorithm worked outside LECO´s proprietary software and, therefore, had no access to the raw data during alignment. INCA instead relied on the characteristic EI fragmentation behavior of trimethylsilyl derivatives by using only the area integrals of the *m/z* 73 ion traces as a quantitative measure.

A peak list in csv file format was exported for each chromatogram containing peak name, 1D and 2D retention times, area count of *m/z* 73, and the deconvoluted mass spectra containing only *m/z* ions with a relative intensity ≥10 % of the base peak. The implementation of a separate *S/N* for 2nd dimension subpeaks in ChromaTOF v4 enabled more accurate peak integration compared to previous versions, but also increased the overall background noise. Hence, the intensity threshold for *m/z* ions relative to the base peak had to be raised from ≥5 to ≥10 % (compared to chapter 6.2.2). Retention time scaling and peak alignment was performed in the same way as in chapter 6.2.3 and 6.2.4 to yield the aligned data matrix.

7.2.3 Normalization and testing

See chapter 6.2.5.

7.3 Results

INCA and SC were evaluated based on the spike-in dataset (for details see chapter 6.2.1.1 and 6.3.1). To ensure direct comparability, INCA alignment was repeated with ChromaTOF software version 4.32. For SC exactly the same preprocessing steps were chosen to compile metabolite fingerprints from the raw data (schematically shown in Figure 7.1). The performances of the alignment routines SC and INCA were directly compared by manually selecting the universal *m/z* 73 ion trace of the trimethylsilyl group as the same quality input signal.

7.3.1 Statistical Compare with *m/z* 73 versus INCA

A total of 887 features were extracted from the GC×GC-TOFMS spike-in data using SC with *m/z* 73 as quant mass. However, some features were not detected in every sample causing missing values. Reasons for missing values were already addressed in chapter 6.3.2. Additionally, only features detected in at least 50 % of all samples were selected for further analysis removing primarily those features, which could not be aligned to other peaks at all. This strict feature reduction resulted in 458 features with a maximum of 14 % zero values.

1. ChromaTOF™ Classification
Exclusion of chromatographic noise regions

2. ChromaTOF™ Processing
- Deconvolution
- Peak combine
- Peak detection above given S/N
- Integration of subpeaks

3. ChromaTOF™ Statistical Compare
- Generate classes
- Add samples to classes
- Peak alignment using 1D RT, 2D RT and mass spectra as sort criteria

4. Export
- Aligned peak table in csv file format

5. Data conversion
- Feature selection
- Normalization with IS
- Log transformation

Analyte	Name	Mass	Count	1st Dimension Average	2nd Dimension Average	Area Average
71	Analyte 70	89	40	673.1	2.11113	1383396.51
1164	*Eicosanoic acid, trimethylsilyl ester	117	40	1817.5	3.00613	2994213.87
73	*Hexanoic acid, trimethylsilyl ester	173	40	674.0	1.96313	100793.990
714	*Dodecanoic acid, trimethylsilyl ester	117	40	1237	2.102	5137492.56
404	*Methionine, trimethylsilyl ester	104	40	1029.5	2.46612	9812062.74
1160	Analyte 1200	167	40	1757.9	3.37087	554767.409
921	*Fructose-MEOX-TMS, 1,3,4,5,6-pentakis-O-(trim	103	40	1429.1	1.9305	1672514.22
74	*Malonic acid-ditms	147	40	681.7	1.99663	17924822.9
80	Pyruvic acid-diTMS, 2-Propenoic acid, 2-[(trimethyl	147	40	694.1	1.97535	1490743.50
415	*Aspartate, N-(trimethylsilyl)-, bis(trimethylsilyl) es	232	40	1037.2	2.56462	1456026.95
412	*Aspartate, bis(trimethylsilyl) ester	160	40	1039.1	2.36637	56077543.7
84	*Valine, N-(trimethylsilyl)-, trimethylsilyl ester	144	40	697	2.51025	742433.520
916	Analyte 954	149	40	1421.1	2.61837	260650.536
422	Analyte 426	189	40	1045	2.01037	101195.001
87	Analyte 86	113	40	705	1.91912	2141267.40
707	Analyte 741	164	40	1233.2	2.2055	436072.152
423	*AIB, tric(trimethylsilyl) ester	174	40	1045	2.052	320424.856
91	Analyte 89	187	40	708.5	1.84488	113050.477
912	Glucose (pyranose-tms), Glucopyranose, 1,2,3,4,f	204	40	1421	1.96938	158277.064
907	*Fumaric acid- diTMS	245	40	1417.1	1.99187	2850099.21
1145	Analyte 1186	163	40	1695.3	2.32987	489662.760
92	l-Alanine, N-(trimethylsilyl)-, trimethylsilyl ester	116	40	712.6	1.95638	10406190.6
906	*Fructose-MEOX-TMS, 1,3,4,5,6-pentakis-O-(trim	307	40	1417	1.93262	84685.828
441	*Decanoic acid, trimethylsilyl ester	117	40	1065	2.07575	3912257.48
444	*Indole, 1-(trimethylsilyl)-	117	40	1065.8	2.48488	10218892.7

Peak	Mass	RT1_mean	RT2_mean	Count	Spike1a	Spike1.1a	Spike1.25a	Spike1.5a	Spike2a	Spike2.5a	Spike4a	Spike1b	Spike1.1b
Analyte 9	81	601.80	1.944	40	-3.518	-4.728	-4.557	-3.667	-3.775	-3.687	-3.837	-4.309	-4.105
Analyte 14	113	610.50	2.670	40	-5.161	-4.756	-4.182	-4.909	-4.486	-4.690	-3.404	-4.038	-3.833
Analyte 17	73	614.50	1.860	40	-1.882	-2.394	-2.168	-2.269	-2.245	-2.140	-1.621	-1.707	-1.698
Analyte 19	102	629.90	1.847	40	-3.081	-4.343	-0.994	-3.456	-2.284	-2.619	0.392	-0.398	-0.341
*Ethanolamin	174	630.70	1.802	40	-4.255	-4.185	-4.093	-4.283	-4.262	-4.192	-3.795	-4.690	-4.173
Analyte 24	131	634.30	1.784	40	-3.766	-4.112	-4.536	-4.191	-4.073	-3.816	-3.330	-4.732	-3.814
Lactic acid -d	117	637.80	1.789	40	-2.821	-2.626	-2.791	-3.075	-3.199	-3.648	-2.638	-2.467	-2.729
Analyte 37	207	641.70	1.713	40	-3.692	-3.208	-3.709	-3.780	-3.348	-3.816	-3.101	-3.423	-3.305
Analyte 43	152	642.60	1.967	40	-1.772	-1.754	-1.236	-1.836	-1.576	-1.497	0.024	-0.549	-0.790
Analyte 47	174	655.10	1.989	40	-2.881	-2.938	-2.760	-2.980	-2.918	-2.751	-2.428	-2.455	-2.556
*Malonic acid	147	658.70	1.842	40	-4.949	-5.147	-4.945	-5.081	-5.077	-4.906	-4.204	-4.732	-4.603
Analyte 55	207	659.90	1.724	40	-3.496	-3.243	-3.827	-3.724	-3.374	-3.589	-2.742	-3.237	-2.980
Lactic acid -d	117	667.00	1.935	40	-2.140	-1.856	-2.223	-2.077	-2.058	-1.987	-1.575	-2.029	-2.015
Analyte 69	178	671.10	2.148	40	-5.760	-6.601	-5.555	-5.801	-5.892	-5.563	-4.549	-5.044	-5.272
Analyte 70	89	673.10	2.211	40	-3.803	-3.836	-3.290	-3.234	-2.912	-2.561	-1.390	-3.763	-3.577
*Hexanoic ac	173	674.80	1.963	40	-5.778	-5.691	-5.759	-5.931	-6.154	-6.100	-5.488	-5.388	-5.560

Figure 7.1. Schematic workflow of the various data processing steps. Reproduced from [108].

Missing values were set to the minimum area count assuming that peaks were missing, because they had fallen below the background noise. The peak areas at *m/z* 73 of the remaining features were then normalized against the peak area at *m/z* 271 of tridecanoic acid and subjected to a log transformation for further validation.

For the INCA alignment, optimal tolerance parameters were determined on the 20 standard compounds that had been spiked into an *E. coli* wild type extract. The optimal parameter setting chosen for subsequent analysis of features was 4 s and

0.05 s for 1D and 2D retention time, respectively, and 80 % and 20 % overlap of *m/z* values and relative ion intensities, respectively. The parameters were more stringent than in our previous work due to the improved quantification of subpeaks by the new ChromaTOF version. With the use of this parameter setting, a total of 4,404 features remained. Data reduction resulted in 447 features with a maximum of 17 % zero values. The peak areas at *m/z* 73 for these features were then also normalized against the peak area of tridecanoic acid at *m/z* 271 and log transformed.

While the initial number of features extracted by INCA had been several-fold greater (n=4,404) than that of SC (n=887), data reduction yielded a comparatively similar number of features for INCA (n=447) and SC (n=458), respectively. The spike-in dataset was used to assess the ability of GC×GC-TOFMS in combination with either alignment algorithm to detect fold changes. T-statistics and ROC curves were compiled as described in chapter 6.3.1. Briefly, pairwise t-tests of each spike-in level with the biological background data were carried out to detect differential features assuming equal variance in both groups. After sorting the list of features in decreasing order according to the absolute value of the t-statistics, each rank of the t-statistic was assigned as true positive and negative, and respectively false positive and negative by comparing the features up to the respective rank to the actual list of spike-ins. The efficiency is shown in ROC curves and their respective AUC. The false positive rate is plotted versus the true positive rate for all possible t-statistic ranks. ROC curves for both INCA (A) and SC (B) are presented in Figure 7.2.

Irrespectively of the alignment algorithm used, it proved impossible to discriminate a 1.1-fold change from the background (AUC for INCA and SC were 0.58 and 0.59, respectively). SC (AUC of 0.90) performed better than INCA (AUC of 0.87) in distinguishing a 1.25-fold change from the background. For higher fold changes, the AUCs improved and became increasingly similar for the algorithms, but did not reach 1 even at the highest spike-in level. In Figure 7.2C, the AUCs of INCA vs. SC were plotted for all possible pairwise fold changes. Overall, SC performed comparable to INCA with a small advantage at the lower and a slight disadvantage at the higher fold changes.
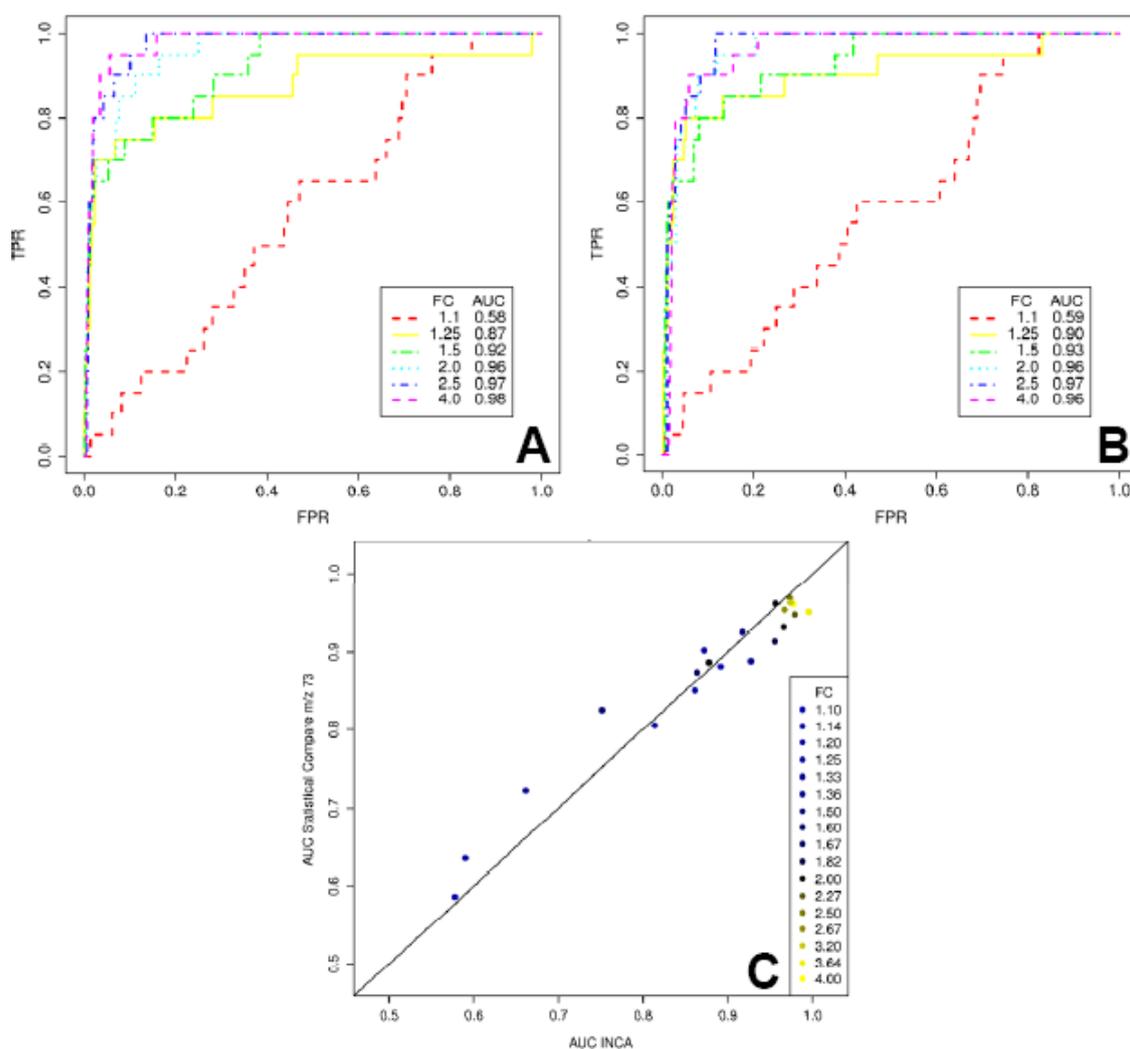
Figure 7.2. ROC curves for different fold changes using (A) INCA and (B) SC alignment, both using *m/z* 73 for quantification, and corresponding AUC values. TPR and FPR improve with increasing fold changes. (C) AUC comparison plot between INCA and SC for all pairwise fold changes. SC performed comparable to INCA in the spike-in experiment. Reproduced from [108].

7.3.2 SC with *m/z* 73 versus unique *m/z*

To fully exploit the potential of SC, the same peak lists of the GC×GC-TOFMS spike-in data were used, but the *m/z* 73 ion trace was replaced by a compound-specific unique mass. Here, SC could further capitalize on its access to chromatographic data during the alignment procedure. Quant masses could be verified and if necessary modified assuring that the same abundant quant *m/z* ion trace was assigned to an identical compound in different samples. In case the quant ion could not be extracted, a ratio of the chosen quant ion to the existent unique *m/z* ion was determined, which represented a viable approximation.

Alignment yielded 1,222 features that were reduced to 632 features with a maximum of 14 % zero values after elimination of those features that had been detected in less than 50 % of the samples. After setting missing values to the minimum area count the features´ peak areas were finally normalized with the area integrals of tridecanoic acid at *m/z* 271 and log transformed.

Representative ROC curves with their respective AUCs making use of unique *m/z* are shown in Figure 7.3A. Analogously, all pairwise fold changes were plotted with their respective AUCs to compare the unique *m/z* with the *m/z* 73 alignment (Figure 7.3B). Although it was still impossible to distinguish a 1.1-fold change from the background, a 1.25-fold change could be distinguished more effectively by the unique *m/z* alignment with an AUC of 0.92 compared to an AUC of 0.90 for the *m/z* 73 alignment. The 2.5-fold and 4-fold changes resulted in nearly ideal AUCs of 0.99. However, the use of unique *m/z* ions as quantifiers still yielded false positives even at the highest fold change tested. Overall, the use of unique *m/z* outperformed *m/z* 73 in the spike-in experiment.
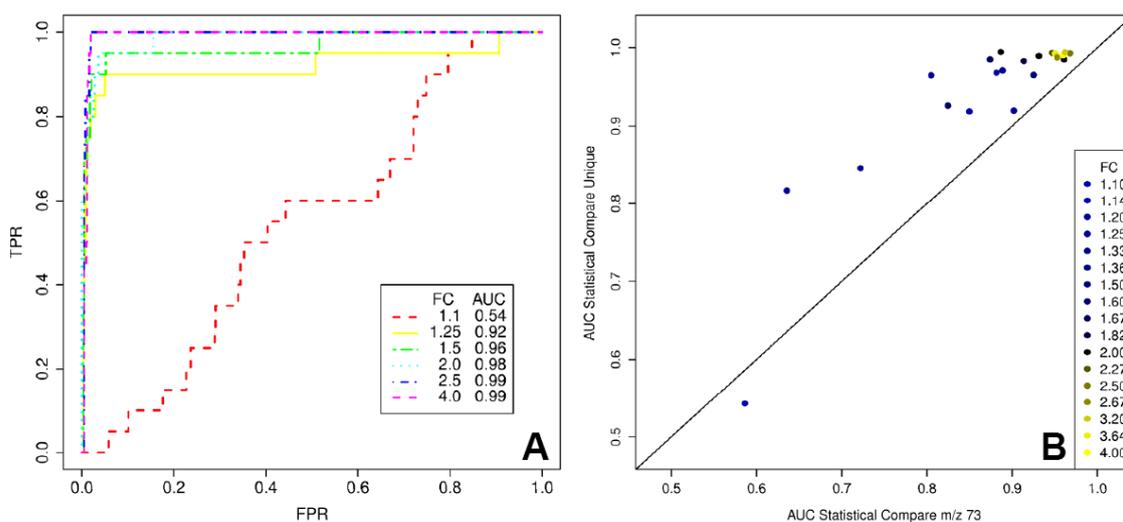
Figure 7.3. (A) ROC curves for different fold changes employing unique *m/z* ions that were extracted automatically by SC as quantifiers and corresponding AUC values. TPR and FPR improve with increasing fold changes. (B) AUC comparison plot between unique *m/z* and *m/z* 73 for all pairwise FCs. Unique *m/z* outperformed *m/z* 73 in the spike-in experiment. Reproduced from [108].

7.3.3 Evaluation of fold changes

Genuine changes in metabolite concentration have to be detected against a background of numerous features. Therefore, it was determined whether the spiked-

in fold changes could be reproduced quantitatively using expected versus observed fold change plots. Linear dependencies between expected and observed fold changes were observed as exemplified in Figure 7.4 for triethanolamine. Corresponding figures for the other spike-ins can be found in Figure S3 in the appendix. Triethanolamine quantified via the integral of the *m/z* 262 (SC) instead of the *m/z* 73 trace (INCA) showed a nearly ideal regression line with a slope of 1 and an offset of 0.

Reproducibility of fold changes was assessed using univariate linear regression on the integrals of both the universal *m/z* 73 fragment ion trace and the analyte-specific fragment ion trace extracted from SC for each of the 20 different spike-ins over the fold changes investigated. The respective regression equations, regression coefficients, and RSDs are given in the figures. The correlation between observed and expected fold changes increased when integrating the unique mass trace extracted by SC rather than the *m/z* 73 fragment ion trace used by INCA. The ranges of regression coefficients and RSDs improved from 0.827 – 0.992 and 5.8 – 20.7 % to 0.882 – 0.994 and 4.2 – 20.7 %, respectively. Interestingly, the highest RSD (20.7 %) belonged to phenylacetate, for which SC had extracted automatically the *m/z* 73 ion trace as quant mass due to the lack of alternative compound-specific *m/z* ions.
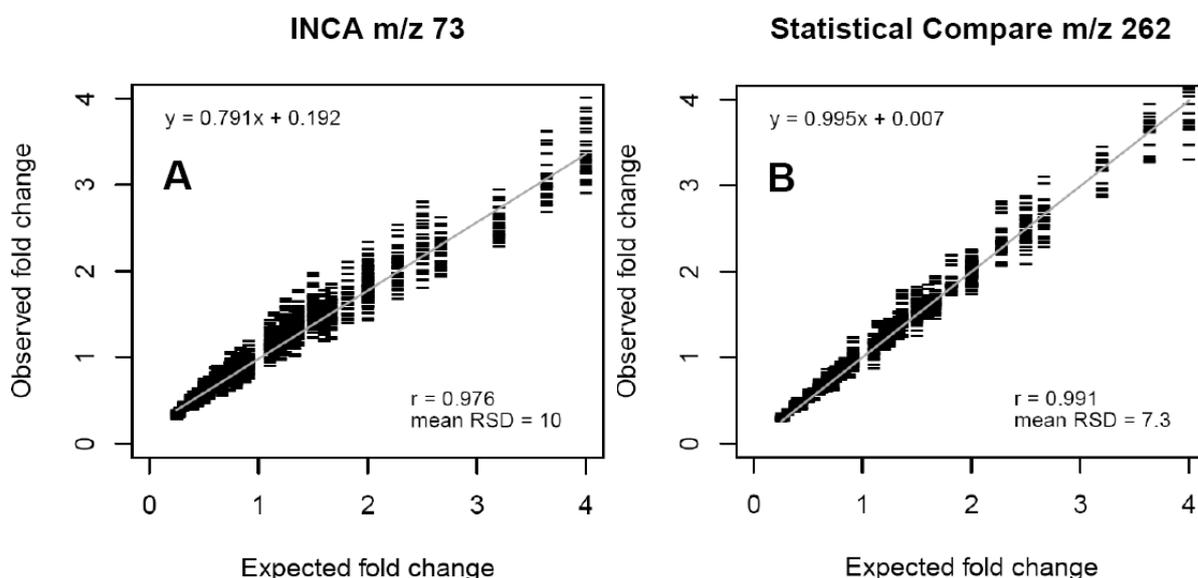
Figure 7.4. Linear dependency between expected and observed fold changes of triethanolamine for both (A) the universal *m/z* 73 fragment ion trace used by INCA and (B) the analyte-specific fragment ion trace *m/z* 262 extracted by Statistical Compare. All possible pairwise fold changes are plotted. Reproduced from [108].

A less than optimal performance of SC in the automated extraction of quant ions is illustrated by the compound dimethylsuccinate. As evidenced in Figure 7.5, the choice of *m/z* 231 rather than the automatically extracted *m/z* 147 would have yielded a superior linear dependence between expected and observed fold changes.
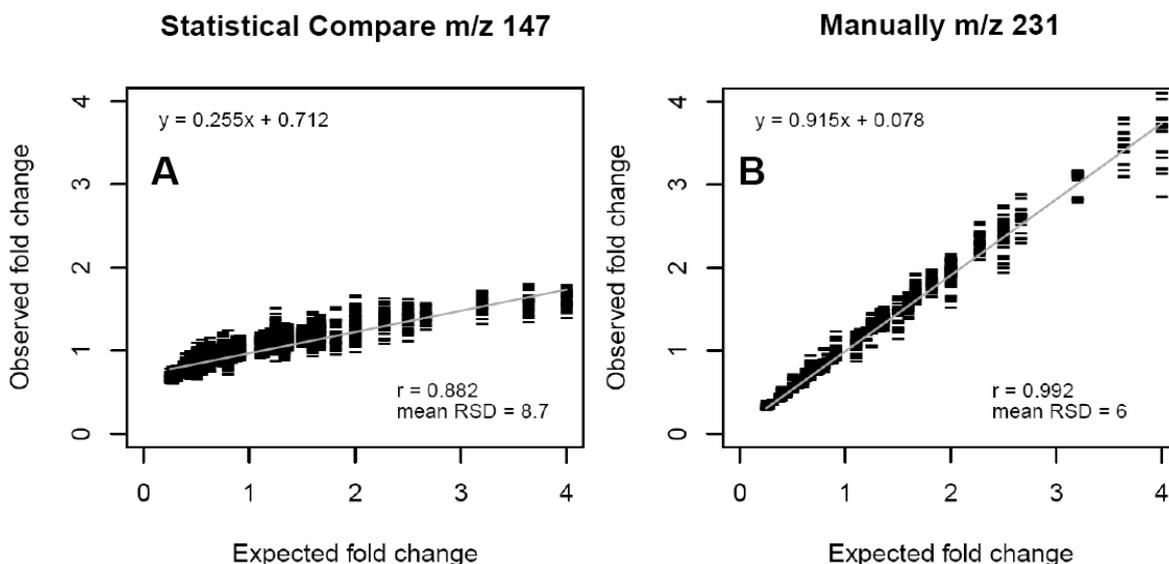
Figure 7.5. Linear dependency between expected and observed fold changes of dimethylsuccinate for (A) the analyte-specific fragment ion trace *m/z* 147 extracted automatically by SC and (B) the manually assigned analyte-specific fragment ion trace *m/z* 231. All possible pairwise fold changes are plotted. Reproduced from [108].

7.3.4 Comparison of *E. coli* strains – INCA versus SC with *m/z* U

Both INCA and SC were applied to the comparative metabolic fingerprinting of a wild type versus a double-mutant strain of *E. coli*. Following exclusion of peaks with a *S/N* <500 and a 2D subpeak *S/N* <50, peak alignment was carried out with SC and INCA according to the parameter settings described in 7.2.2.1 and 7.3.1, respectively, using 1D and 2D retention times and the mass spectra as sorting criteria.

After SC alignment, a total of 809 features remained (Table 7.1). Only features that had been detected in at least 50 % of all samples (9 out of 18) were further analyzed. This arbitrary cut-off had been selected with respect to the group size of 9 to ensure that features present in one group but not the other were not excluded from the list. A total of 497 features with a maximum of 15 % zero values remained. Missing values were set to the minimum area count; feature integrals were normalized against the *m/z* 271 integral of tridecanoic acid, subjected to log transformation and a PCA (Figure 7.6). A clear segregation of the *E. coli* strains and blanks was obtained.

Table 7.1. Comparison of the performance of SC and INCA in the extraction of features from GC×GC-TOFMS data and the identification of metabolites that distinguish a wild type from a double-mutant *E. coli* strain lacking the transhydrogenases UdhA and PntAB. Metabolites printed in bold and italic indicate those detected by one of the two alignment algorithms only. Reproduced from [108].

	SC	INCA
Total features	809	2625
Reduced features[1]	497	422
Significant features[2]	56	53
Metabolites identified that distinguished the *E. coli* strains	citrate, cis-aconitate, succinate, α-ketoglutarate, fumarate, glucose, glucose-6-phosphate, myo-inositol, indole, GABA, proline, lysine, leucine, isoleucine, valine, pyroglutamate, aminoadipate, tartrate, orotate, ***mannitol***, furoate, caproate, sucrose	***pyruvate***, citrate, cis-aconitate, succinate, α-ketoglutarate, fumarate, ***malate***, glucose, glucose-6-phosphate, myo-inositol, indole, GABA, ***glycine***, proline, lysine, leucine, isoleucine, valine, pyroglutamate, ***putrescine***, ***itaconate***, aminoadipate, tartrate, orotate, furoate, caproate, sucrose

[1] Features exceeding *S/N* ≥500 in 9 out of 18 samples.

[2] Metabolites that differed significantly (FDR <0.05) in abundance between the strains.
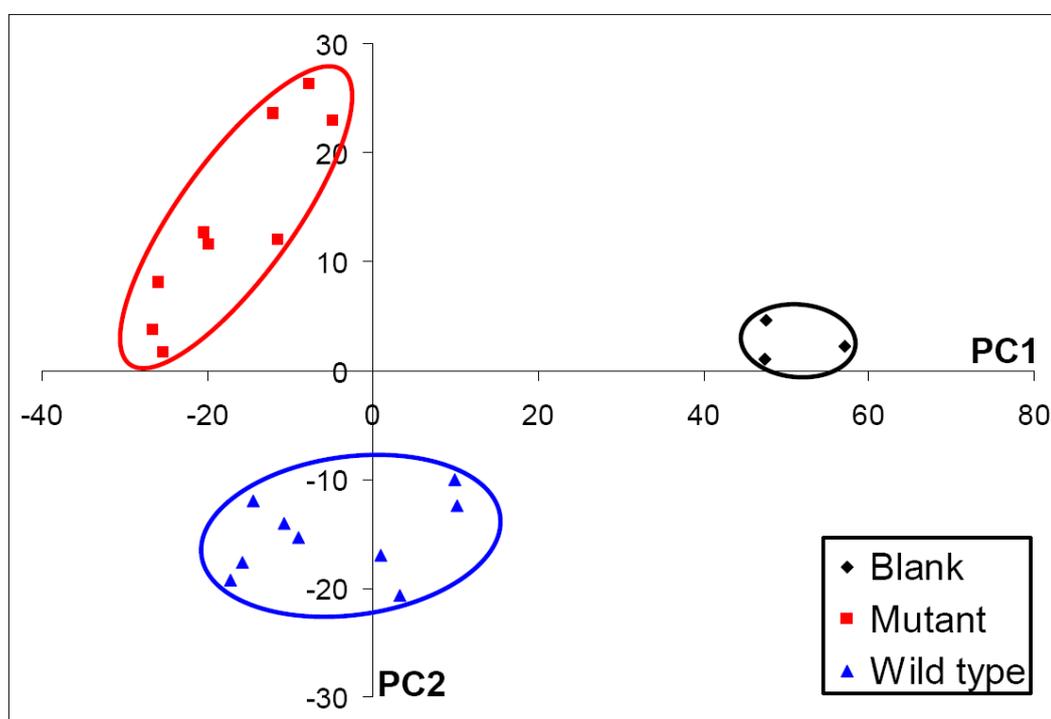
Figure 7.6. PCA of randomly aligned metabolic fingerprints, using SC, for three biological, respectively analytical replicates each of an *E. coli* mutant and wild type strain. Feature reduction yielded 497 features present in at least 50 % of all samples. PC1 accounted for 36 % of the variance, PC2 for 10 %. Reproduced from [108].

A t-test with equal variances identified a list of 56 features that differed significantly between the two strains with an estimated false discovery rate of <0.05, indicating significance also after correction for multiple testing. Out of the 56 features, 23 metabolites were identified accounting for 29 of the 56 features (Table 7.1). A feature does not necessarily equal a metabolite (also described in chapter 6.3.2) as partial silylation, steric hindrance, degradation or rearrangement reactions can occur and cause multiple possible analyte structures of the same metabolite. Further, methoximation may yield two signals for certain compounds due to E-Z isomerism of methoxyamines.[49, 51] Hence, a metabolite can be represented by more than one feature and *vice versa* a feature can represent more than one metabolite.

With INCA similar results as in chapter 6.3.2 were obtained. INCA alignment yielded 2,625 features that were reduced to 422 features with a maximum of 17 % zero values. Out of the 53 features that significantly distinguished the *E. coli* strains, 27 metabolites were identified accounting for 33 of the 53 features.

Comparing the lists of metabolites identified by the two alignment routines, mannitol represented an additional metabolite found by SC only, whereas SC failed to detect glycine, itaconate, malate, putrescine, and pyruvate, although their concentrations had been verified in chapter 6.3.2 to differ significantly between the two strains. The problem here was the failure of SC to extract useful unitized unique quantifiers with sufficient intensity across all samples. This is exemplified by putrescine, for which SC had extracted the *m/z* 170 trace as quant ion, yielding RSDs of 56 % and 120 %, respectively, for *E. coli* mutant and wild type strain, while the respective RSDs for *m/z* 73 trace were 32% and 39 %. The latter were sufficient to detect a significant difference in the intracellular levels of putrescine between the wild type and the mutant *E. coli* strain. Such shortcomings of the current version of SC may be overcome in the future by the implementation of an intensity range threshold for the automated extraction of suitable quant ions. This way masses can be avoided whose intensities are either in saturation or slightly above the noise level, as these masses typically yield increased variance.

7.4 Discussion

In contrast to INCA, SC operates within the ChromaTOF software and thus has access to the raw data during the alignment process. This enables SC to (i) select an

analyte-specific *m/z* ion rather than the universal *m/z* 73 trace of the trimethylsilyl ion as a quantitative measure for identical compounds across different samples and (ii) set a separate *S/N* for signals that fell initially below the predefined *S/N*, which reduces zero values. These attributes give SC an edge in the detection of changes in analyte concentration twofold or less.

SC fails at times to extract quant ions with adequate intensity. These ions typically yield excessive variance in abundance that makes a discriminate analysis impossible. Additionally, SC excludes features from the final table if conflicts in the peak grouping have occurred. The fact that the user is denied access to these conflicts to possibly reconcile them is a clear limitation of the tool. SC also assigns a single peak name to a whole group of aligned peaks based on the best quality mass spectrum within a group. However, the automatic peak name assignment is not always advantageous, because the user loses the ability to reconstruct the correct alignment of each metabolite across all samples. Obviously, there is need for improvement of SC in certain areas to make it a more efficient tool for comparative metabolic fingerprinting.

INCA, on the other hand, guarantees that no data is lost and the decision about further data handling is left to the user. An additional advantage is the capability of INCA to automatically select optimal alignment tolerance parameters based on spiked-in compounds for any biological matrix. A drawback of INCA certainly is the lack of alternative quantification options. The *m/z* 73 ion is often affected by increased variance due to overload or saturation of the detector. Alternatively the TIC could be utilized for quantification, but it also tends to be affected by typically abundant *m/z* 73 and 147 ions that derive from TMS derivatives.

8 GC×GC-TOFMS quantitative analysis of amino acid enantiomers in physiological fluids

8.1 Introduction

The enantioselective analysis of amino acids is attracting increasing interest due to a better understanding of the biological relevance of D-AAs.

D-AAs can be found in bacteria, mammalians, plants, and food.[175-181] In humans, D-AAs mainly derive from bacterial metabolism and intake of fermented food.[175-177, 182] D-AA elimination typically proceeds via renal excretion. In addition, enzymes that catalyze D-AA decomposition, such as D-AA oxidase and D-aspartate oxidase, have been found in brain, liver, and kidney.[183, 184]

Separation and quantification of amino acid enantiomers rely primarily on chromatographic and electrophoretic strategies.[185] Waldhier et al. (2010) developed a GC-qMS method based on pre-column methanol/methyl chloroformate derivatization and separation on a γ-cyclodextrin based column (Rt-γDEXsa) for the quantitative analysis of proteinogenic AAEs.[5] Twelve out of the twenty proteinogenic amino acids eluted from the Rt-γDEXsa column (Ala, Gly, Val, Leu, Ile, Pro, Thr, Asp, Ser, Met, Asn, and Phe). Baseline separation was accomplished for each corresponding enantiomer pair (except for Phe enantiomers). Baseline separation of D-Ile/L-Leu, L-Thr/L-Asp, and L-Ser/D-Met, respectively, could not be achieved. While the use of unique *m/z* ions enabled the quantification of L-Thr, L-Asp, and D-Met, it was not possible to quantify the structural isomers D-Ile and L-Leu due to their identical *m/z* ions. Furthermore, a double peak impeded the quantification of L-Ser.

GC×GC was employed by Junge et al. (2007) for the first time to separate AAEs as their ethyl chloroformate derivatives in samples of beer by coupling a enantioselective 25 m × 0.25mm ID × 0.16 μm Chirasil-L-Val column to a 3 m × 0.1 mm ID × 0.1 μm BP-1 column in the second dimension.[158] The study presented in this chapter widens the scope of GC×GC-based AAE analysis by applying a combination of an Rt-γDEXsa chiral 1D column with different 2D columns to the separation and also quantification of AAEs as their MeOH/MCF derivatives in

physiological fluids. The performance was compared to that of the previously published GC-qMS method.[5]

The work in this chapter was published in [126] with Magdalena Waldhier (Institute and Chair of Functional Genomics, University of Regensburg), as joint first author.

8.2 Materials and Methods

8.2.1 Sample preparation and instrumentation

Aqueous stock solutions were prepared from the L- and D-AA solids at concentrations of 11 – 60 mM. Racemic stock solutions contained enantiomers in the range of 7.0 – 46.5 mM. A master mixture was prepared based on these solutions containing Gly and all enantiomers at a concentration of 1 mM each. A [U-^{13}C, U-^{15}N] cell free amino acid mix served as internal standard. The combined D+L-concentrations of the 20 proteinogenic AAs contained in the mixture ranged between 0.37 and 2.58 mM, but no information on D-AA ratios was available.

A set of 25 serum specimens from patients with liver cirrhosis and 16 control sera were analyzed to demonstrate relevance for medical diagnostics. At the time of blood drawing, none of the patients had been treated with peptide antibiotics or other drugs that might affect D-AA serum levels. Serum specimens were stored at -80°C. Specimens were provided by collaborators at the University Hospital of Regensburg with approval from the institutional review board.

Sample and standard solution aliquots of 150 µL and 20 µL of internal standard mix were used for quantitative analysis. Derivatization with MeOH/MCF was performed as described in chapter 5.2.2 or [5]. Sample preparation was carried out by Magdalena Waldhier and Nadine Nürnberger (Institute and Chair of Functional Genomics, University of Regensburg).

For analysis, a sample volume of 1.5 µL was injected in splitless mode. GC×GC-TOFMS instrument settings were used according to chapter 5.3.2. Temperature programs (Table 8.1) were optimized by Magdalena Waldhier. Method A was used for the Rt®-γDEXsa/Rtx®-1701 column set, whereas method B was applied to the Rt®-γDEXsa/ZB-AAA column set.

8.2.2 Data processing

Raw data were processed with the LECO ChromaTOF software version 3.34. For compound identification, EI spectra were matched against an in-house generated library containing the MeOH/MCF-derivatized D- and L-AAs and compared to the EI spectra of the validated GC-qMS method.[5]

Table 8.1. Optimized GC temperature programs for the column sets RT®-γDEXsa/Rtx®-1701 (A) and Rt®-γDEXsa/ZB-AAA (B), respectively.

Method	Temperature program
A	70°C (1 min) – 2°C/min – 150°C (10 min) – 2°C/min – 180°C (25 min)
B	70°C (1 min) – 2°C/min – 130°C (12.5 min) – 8°C/min – 150°C (4 min) – 4°C/min – 190°C (6 min)

8.2.3 Quantification and method validation

Calibration was performed with the Rt®-γDEXsa/ZB-AAA column set using a 16 point-serial dilution of the master mixture. It contained all AAs, whose elution from the Rt®-γDEXsa column as MeOH/MCF-derivatives could be detected, over a concentration range of 31 nM to 1 mM each. The applied internal standard solution contained 20 uniformly ^{13}C-, ^{15}N-labeled AA types at different concentrations between 0.37 and 2.58 mM. As the labeled AAs had been obtained from algae, they were present in the mix at their naturally occurring ratios in algae. Most D-AAs were detected in the labeled mix but their abundance was too low (<3.4 % of the respective L-AA area) to be suitable as internal standards. Thus, the stable isotope-labeled L-enantiomers of Ala, Val, Leu, Ile, Pro, Thr, Asp, Ser, Met, Asn, and Phe were used to correct areas of the corresponding D- and L-AA in quantitative measurements. Correspondingly, stable isotope-labeled Gly, the only natural nonchiral AA, was used for correcting Gly area integrals. Eight aliquots of a human serum specimen were prepared and analyzed to test method reproducibility.

8.3 Results

Two different column combinations for the GC×GC separation of MeOH/MCF derivatives of AAEs were tested. In GC×GC the length of the 2D column is kept short to allow the elution of analytes within one modulation period. Length of the 1D column, in contrast, is not limited. Therefore, the chiral Rt®-γDEXsa column was

always chosen as 1D stationary phase, as the resolution of enantiomers improves with increasing length of the chiral column. Secondly, the Rt®-γDEXsa phase reacted sensitive to high temperatures, resulting in reduced enantioselectivity and accelerated column degradation as shown previously.[5] The increased temperature in the 2D separation is thus unfavorable for AAE resolution.

8.3.1 Rt®-γDEXsa/Rtx®-1701 column set

Initially, a 2 m midpolarity 14 % cyanopropylphenyl/86 % dimethyl polysiloxane Rtx®-1701 capillary column served as the 2D column. A characteristic chromatogram of the master standard obtained under the optimized temperature program (method A, Table 8.1) is shown in Fig. 8.1. The Rtx®-1701 column compensated for the lack of selectivity of the Rt®-γDEXsa for L-Thr/L-Asp and L-Ser/D-Met, respectively, and allowed their baseline resolution in the second dimension. However, L-Leu and D-Ile still coeluted. MCF-Phe enantiomers were not resolved at all by the chiral Rt®-γDEXsa column and, therefore, could not be separated in the second dimension either.
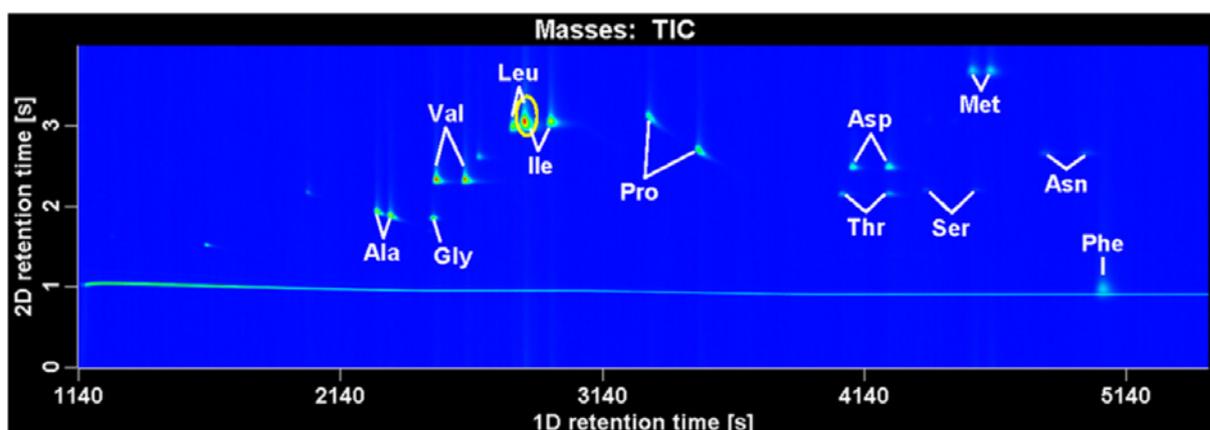
Figure 8.1. Total ion current (TIC) 2D chromatogram of the master mixture using the Rt®-γDEXsa/Rtx®-1701 column set and GC method A. D-Ile/L-Leu isomers are encircled. Reproduced from [126].

8.3.2 Rt®-γDEXsa/ZB-AAA column set

Kaspar et al. (2008) had demonstrated the successful separation of Leu and Ile as their PCF derivatives on a 15 m ZB-AAA capillary column that reportedly consists of 50 % phenyl/50 % dimethyl polysiloxane.[53, 186] Therefore, the ZB-AAA column was

tested as 2D column. A representative chromatogram of the master mixture obtained after optimization of the temperature program (method B, Table 8.1) is pictured in Fig. 8.2A.
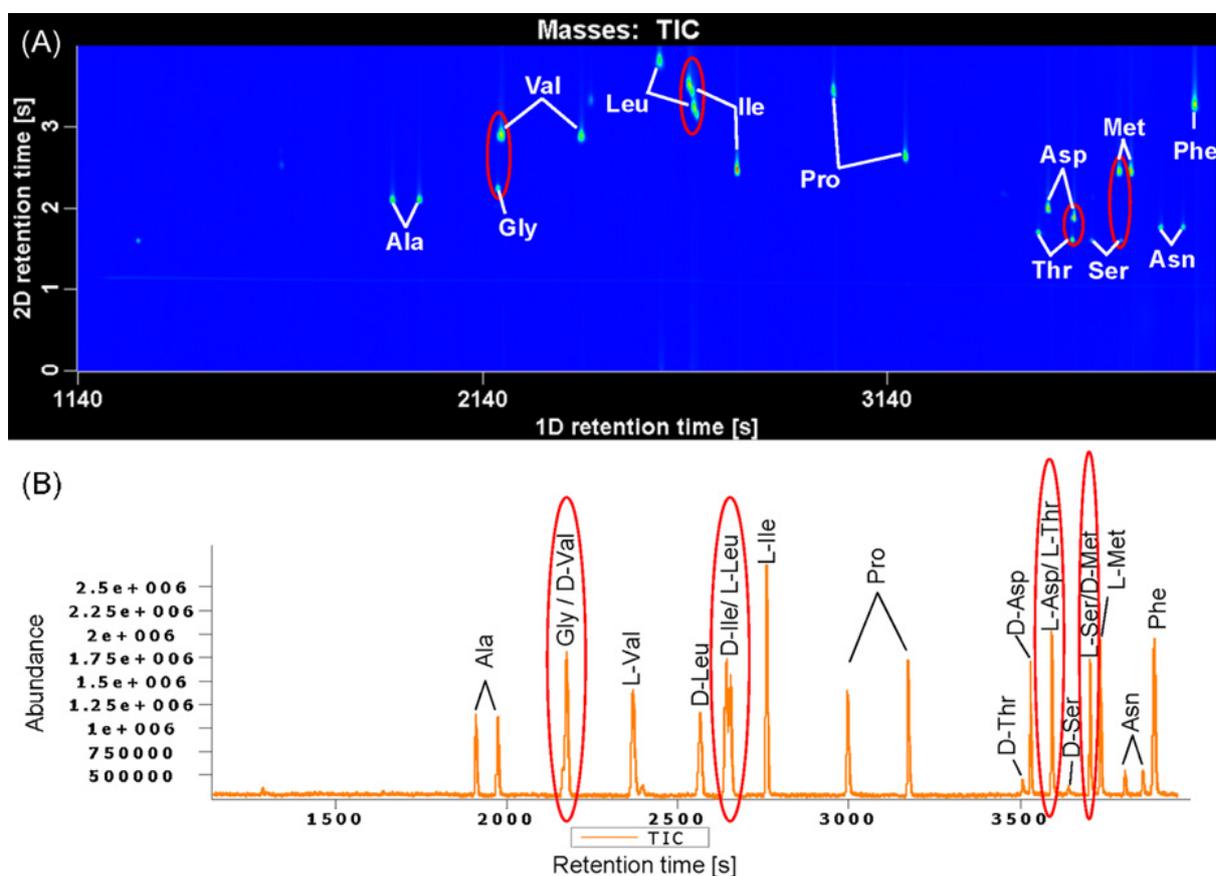
Figure 8.2. Total ion chromatograms of the master mixture recorded in (A) 2D and (B) 1D mode using the Rt®-γDEXsa/ZB-AAA column set and method B. Analytes coeluting from the Rt®-γDEXsa column are encircled in both chromatograms. Reproduced from [126].

As can be seen, baseline separation was achieved for all analytes except for the Phe enantiomers. In contrast, 1D-GC-TOFMS yielded 4 unresolved peak pairs (Figure 8.2B) reflecting the insufficient selectivity of the Rt®-γDEXsa column. Selectivity was improved by using thermal modulation and the orthogonal separation properties of the 2D ZB-AAA column. Enhanced resolution of enantiomers that were separated by a temperature program on the chiral 1D column was achieved due to the fast and therewith approximately isothermal separation on the 2D column. It allows additional separation of enantiomers even though the 2D selector provides no stereospecific retardation. This can be seen for the enantiomers of Leu, Ile, and Pro in Fig. 8.2A.

8.3.3 Quantification and method validation

Calibration curves were generated for absolute quantification of AAEs. Table 8.2 lists fragment ion masses chosen for quantification, as well as values for linear range, R^2, and reproducibility. For direct comparison, Table 8.2 also includes the values obtained for 1D-GC-qMS. The GC-qMS method is specified in [5] (Table 1, E). GC×GC-TOFMS calibration curves were linear with the square values of the sample correlation coefficient R ranging between 0.9938 and 0.9992, which is comparable, albeit inferior to the 1D method, whose R^2 values ranged from 0.9956 to 1.0000. LLOQs were in the range of 0.03 – 2 µM. Octaplicate analysis of a serum specimen yielded RSDs below 5 % except for D-Ala (12.2 %), D-Asp (16.6 %), and D-Asn (9.9 %), whereas hexaplicate 1D-GC-qMS analysis yielded in general RSDs <4 %.

Further method validation was exclusively handled by Magdalena Waldhier and can be found in her Ph.D. thesis (in preparation) or in [126].

Table 8.2. Comparison of linear ranges, R^2 values, and relative standard deviations for AAE analysis in serum by GC×GC-TOFMS and GC-qMS. Reproduced from [126].

AA	Quantifier analyte + U-^{13}C,^{15}N-labeled IS (m/z)	GC×GC-TOFMS			GC-qMS		
		LR [µM]	R^2	RSD (%) (n=8)	LR [µM]	R^2	RSD (%) (n=6)
D-Ala	102 + 105	0.031 -1000	0.9971	12.15	0.031 -500	0.9998	1.10
L-Ala	102 + 105	0.061 -1000	0.9974	2.16	0.122 -500	0.9996	0.95
Gly	88 + 90	0.031* -1000	0.9979	3.78	0.977 -500	0.9996	0.87
D-Val	130 + 135	0.244 -500	0.9971	-	0.244 -500	0.9999	-
L-Val	130 + 135	0.061* -500	0.9961	3.81	0.244 -500	0.9999	0.81
D-Leu	144 + 150	0.244 -1000	0.9974	-	0.244 -500	1.0000	-
L-Leu	144 + 150	0.061 -1000	0.9954	1.28	-	-	-
D-Ile	144 + 150	0.061 -1000	0.9970	-	-	-	-
L-Ile	144 + 150	0.061 -1000	0.9973	3.07	0.061 -500	0.9998	0.87
D-Pro	128 + 133	0.031 -500	0.9988	4.52	0.061 -500	0.9998	1.31
L-Pro	128 + 133	0.061 -500	0.9970	3.38	0.061 -500	0.9999	0.73
D-Thr	147 + 149	0.122* -500	0.9986	-	0.977 -1000	0.9996	-
L-Thr	147 + 149	0.061* -500	0.9980	3.53	1.953 -1000	0.9988	1.62
D-Asp	160 + 164	0.031* -1000	0.9938	16.64	0.122 -500	0.9998	3.87
L-Asp	160 + 164	0.061* -1000	0.9990	3.18	0.977 -500	0.9994	0.70
D-Ser	118 + 121	0.244* -1000	0.9987	-	1.953 -1000	0.9999	-
L-Ser	118 + 121	0.031* -1000	0.9991	2.37	31.25 -1000	0.9956	-
D-Met	162 + 167	1.953 -1000	0.9988	-	0.488 -1000	0.9999	-
L-Met	162 + 167	0.977 -1000	0.9992	2.56	0.244 -1000	0.9996	1.06
D-Asn	127 + 132	0.061* -1000	0.9973	9.93	0.244 -1000	0.9999	-
L-Asn	127 +132	0.031* -1000	0.9983	3.43	0.244 -1000	0.9999	1.02
DL-Phe	162 + 171	0.061 -1000	0.9978	3.41	0.061 -1000	0.9995	1.13

* At least two additional calibration points were included at the lower end of the GC×GC calibration curve compared to the GC-qMS calibration curve.

8.3.4 Application to serum specimens from patients with liver cirrhosis

An exemplary biomedical application of GC×GC-TOFMS was performed by analyzing serum specimens from n=16 healthy probands and n=25 patients with liver cirrhosis. Data analysis was performed by M. Waldhier. The clear trend of increased serum levels of D-AAs in liver cirrhosis patients is visible in Fig. 8.3A.
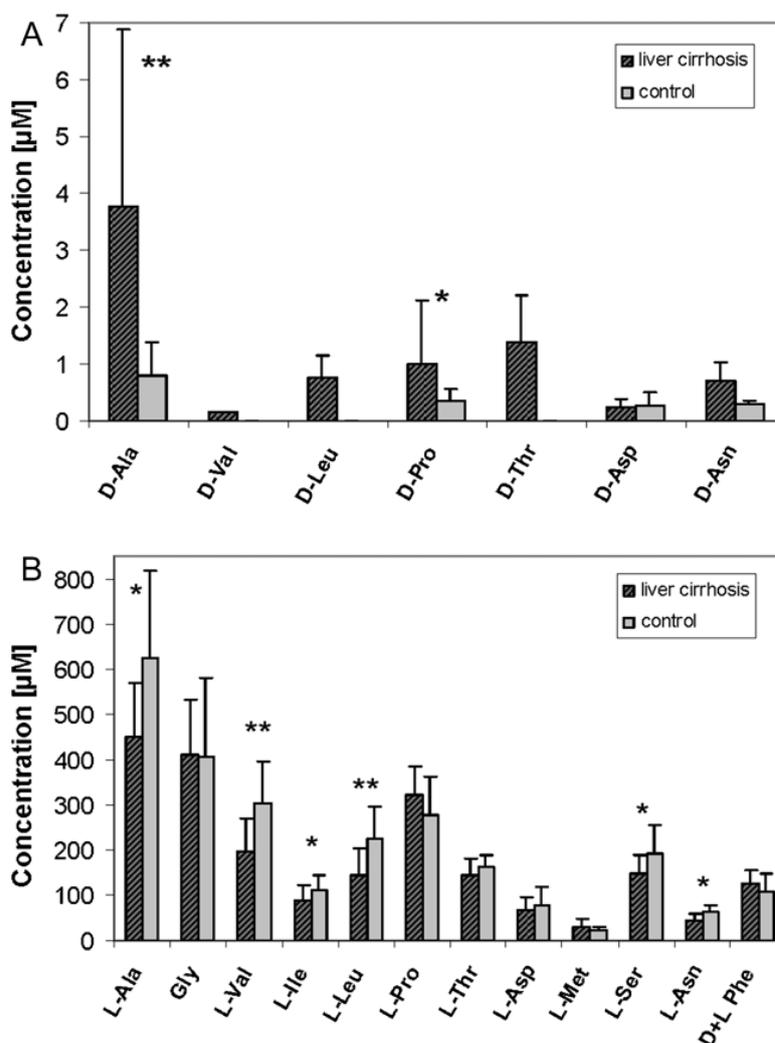
Figure 8.3. Mean concentrations and SDs of (A) D-AAs and (B) L-AAs in serum specimens of n=25 patients suffering from liver cirrhosis compared to the respective values of n=16 healthy probands (*p <0.05, **p <0.001). Reproduced from [126].

Only D-Asp concentrations were not increased compared to the respective mean concentration of the control group. D-Val, D-Leu, and D-Thr were found above the LLOQ only in some of the serum specimens of cirrhosis patients. D-Ile and D-Met were detected in none of the samples and D-Ser could not be measured due to the interference by a matrix compound. Compared to the controls, significantly elevated

serum concentrations of D-Pro (p=0.013) and D-Ala (p=0.00008) were observed as determined by a two-tailed Student's t-test. Mean D-Asn concentrations were not tested because of an insufficient number (n=2) of values above LLOQ in the control group. Determined concentrations of D-Ala, D-Pro, and D-Asn + D-Asp in control sera were in good agreement with respective serum levels reported previously for healthy volunteers.[187, 188] L-AA serum levels, on the other hand, were significantly decreased in liver cirrhosis specimens as shown in Fig. 8.3B: L-Ala (p=0.0051), L-Ile (p=0.036), L-Ser (p=0.0099), L-Asn (p=0.0011), L-Val (p=0.00028), and L-Leu (p=0.0003). The increased D-AA serum levels observed for liver cirrhosis reflect the loss of intact liver parenchyma and, hence, a reduced enzymatic capacity to catalyze D-AA decomposition.

8.3.5 Comparison of GC×GC-TOFMS and GC-qMS

In comparison to GC-qMS, GC×GC-TOFMS data analysis was considerably more time-consuming. Another problem resulted from the slightly earlier elution of internal standards compared to the respective analytes as shown in Fig. 8.4. Since the base fraction peak of the labeled AA appeared one modulation prior to the analyte maximum, the analyte-to-standard area ratio varied for each fraction. Thus, peaks from all modulations had to be integrated to determine the response.
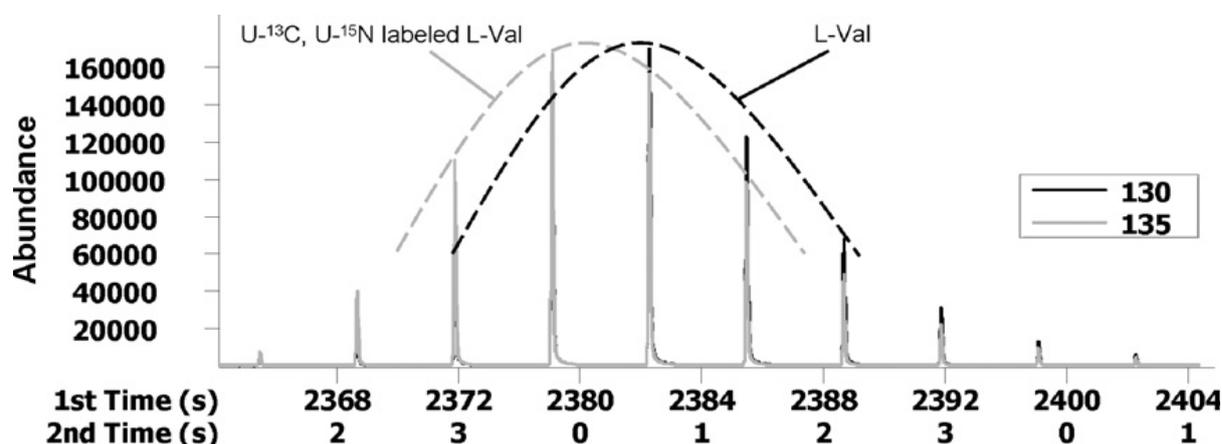
Figure 8.4. Extracted ion chromatogram of L-Val quantification masses. Peak envelopes (dashed lines) of L-Val and U-^{13}C, U-^{15}N labeled L-Val are indicated to demonstrate their position shift and the varying analyte-to-internal standard area ratio for each fraction. Reproduced from [126].

Numerous fractions had to be integrated manually and summed up, leading to a significant increase in data analysis time. These problems were caused by the

spectral match factor that was used to combine all modulated peaks. Due to the impact of noise, low abundant modulated peaks suffer from an insufficient spectral match and, therefore, are not integrated automatically. Reproducible manual inclusion of a great number of sub-fraction integrals is a challenging task and contributes to higher RSDs for GC×GC-TOFMS quantitative data as compared to the excellent repeatability of chiral GC-qMS analysis (RSDs <4 %).

As a result of the concentration discrepancy between D- and L-AAs (which may be as great as three orders of magnitude) we observed a great number of modulations and peak tailing for L-AAs and stable isotope-labeled internal standards. Optimally, each first dimension peak should be sampled three to five times, which was the case for D-AAs. The correction of the area integral of a D-AA with a stable isotope-labeled L-AA that generates ten or more modulations is suboptimal and, consequently, lowers repeatability of D-AA quantification results as reflected in comparatively great RSDs for D-Asp (16.6 %), D-Ala (12.2 %), and D-Asn (9.9 %), whereas RSDs for the other target analytes were <5 %. A set of stable isotope-labeled D-AAs for D-AA area integral correction would thus improve precision of quantification results, but is unfortunately not available at present.

The same number of AA types eluted from the Rt$^®$-γDEXsa column in GC×GC analysis as in GC-qMS analysis. Reasons for the failure to detect the proteinogenic AAs Glu, Gln, Cys, Lys, His, Arg, Tyr, and Trp remain to be elucidated, but may be due to chemical or thermal derivative instability as reported for propyl chloroformate derivatized arginine [53], or strong interactions with the CD core leading to strong retardation and broad peaks that vanish into the baseline. However, due to its higher peak capacity the presented GC×GC approach has the potential to be expanded to enantiomers of other small organic acids.

In the end, the optimized GC×GC-TOFMS method presented in this chapter for the determination of AAEs in physiological fluids is at present to the best of our knowledge the most effective technique for AAE analysis, enabling quantification of 8 D-AAs in serum or urine in a single chromatographic run.

8.4 Discussion

The combination of a γ-CD column (Rt$^®$-γDEXsa) and an AA selective (ZB-AAA) column resulted in the baseline separation of 20 AAEs as their MeOH/MCF

derivatives. GC×GC-TOFMS compared to GC-qMS with SIM yielded distinctly lower LOQs for Gly, L-Val, and both enantiomers of Thr, Asp, Ser, and Asn. Analytes eluting late from the chiral column showed an improved LLOQ in GC×GC-TOFMS, as column bleeding could be separated from all analytes in the 2D. In general, LLOQs were in the range of 0.03 – 2 µM. Reproducibility of the analysis of a serum specimen in octaplicate ranged from 1.3 to 16.6%. The method was applied to the comparison of AAE serum concentrations in patients suffering from liver cirrhosis to a control group. Significantly increased D-AA concentrations were found for the patient group, whereas L-AA levels were decreased.

Peak resolution was improved by coupling two columns with orthogonal separation characteristics. All target analytes were resolved adequately from matrix peaks except for D-Ser in serum, a distinct improvement over GC-qMS on an Rt®-γDEXsa column that had failed to resolve D-Asn and both enantiomers of Thr and Ser from matrix compounds. Separation of D-Met/L-Ser and L-Thr/L-Asp, respectively, which had coeluted from the chiral column, was accomplished with the GC×GC approach. Separation of D-Ile/L-Leu was possible but required decreased operation temperatures for both columns and, as a consequence, increased analysis time from 44 to 66 min.

9 GC-MS based metabolic fingerprinting and profiling of serum versus plasma collection

9.1 Introduction

A widely used biological fluid in mammalian studies is blood, as it reflects systemic changes in the metabolome. Blood is commonly either sampled as plasma, that still contains clotting factors, or as serum. Advantages of plasma over serum are quick processing, as there is no need to wait for blood to clot, higher yield, lower risk of haemolysis and thrombocytolysis, and virtually no interference from postcentrifugal coagulation that can occur in serum.[189] On the other hand, disadvantages of plasma relative to serum are interferences with analysis caused by anticoagulants, enzyme inhibitors, fibrinogen, and cations contained in anticoagulants. For plasma collection different anticoagulants and additives are used, including EDTA, heparin, sodium fluoride/potassium oxalate (NaF-KOx), and citrate. Anticoagulation occurs by binding calcium ions (citrate, EDTA, and oxalate) or by inhibiting thrombin activity (heparinates, hirudin). The choice of additive depends on the analysis to be performed.

If heparin is used that contains sodium metabisulfite as a preservative, the quantification of cystine and homocystine can be impaired due to the formation of sulfocysteine and homosulfocysteine.[190] Fluoride inhibits the enzyme enolase in the glycolytic pathway and thereby prevents the degradation of glucose, with KOx added as an anticoagulant. Waring et al. (2007) compared NaF-KOx and serum gel tubes in the determination of blood glucose in humans.[191] Serum was promptly separated from gel tube samples and refrigerated, whereas NaF-KOx samples were not separated until immediately before analysis. A small relative negative bias ranging from 3.2 % to 6.3 % was observed for samples collected in NaF-KOx tubes, caused by dilution and the onset of glycolysis. Consequently, serum can be used for glucose determination if special serum-separating tubes are employed and if the samples are immediately frozen after centrifugation. Mikesh et al. (2008) studied human blood collected in NaF-KOx tubes and stored at room temperature.[192] While fluoride prevents lactate production by inhibiting enolase in the glycolytic pathway, glucose is still consumed up to 3 hours after blood collection, which is attributed to the

continuing metabolism of glucose to glucose-6-phosphate, as the enzymes upstream of enolase remain active. Further, the initial decrease in glucose is the greater the higher the white blood cell, red blood cell, or platelet-count in the blood specimen under investigation. However, storing the blood samples with or without fluoride addition on ice showed only a minimal decrease in glucose and increase in lactate.[192] A drawback of NaF-KOx plasma for metabolic fingerprinting by GC-MS is the excess of oxalate that may interfere with the analysis. The same applies for using sodium citrate as anticoagulant. Moreover, citrate is an important metabolite of the TCA cycle and the use of citrate as an anticoagulant will prevent its accurate analysis. While NaF-KOx and citrate plasma are not well suited for general metabolomics applications, the question remains, which type of blood sample is preferable for metabolomics studies. Deprez et al. (2002) studied the stability of rodent plasma collected into heparinized containers for ^{1}H-NMR analysis in toxicology studies.[193] The time elapsing between the drawing of blood and centrifugation to separate plasma from blood cells was found not to be critical if less than 35 min, with no changes in small metabolites normally associated with cell leakage and haemolysis observed. No significant changes were identified in samples, which were snap-frozen and stored for up to 9 months at −80 °C. However, storing the samples at 4 °C or room temperature led to changes in metabolite concentrations, such as an increase in glycerol and choline indicating lipid hydrolysis. Teahan et al. (2006) qualitatively compared NMR metabolic profiles of human serum and heparin-plasma. For serum and plasma, four different clot contact times up to 180 min and keeping the samples during this time on ice or room temperature were investigated.[194] The differences between NMR profiles of serum and heparin-plasma were found to be minimal. Overall, the inter-individual differences were larger than differences in spectra caused by increasing clotting time. Nevertheless, changes in metabolite profile were smaller when samples had been kept on ice between blood sampling and cell separation for periods not exceeding 35 min.[194]

Further, lubricants, polymeric surfactants, plasticizers, separator gel components, and heparin have been reported as potential sources of interference in mass spectrometric analyses.[195, 196] Moreover, MS based studies have been performed that evaluated the impact of different blood collection tubes on selected analytes (aldosterone and testosterone/dihydrotestosterone, respectively).[109, 197] However, to the best of our knowledge, no comprehensive study has been performed on the

impact on the suitability of plasma versus serum for mass spectrometry based metabolic fingerprinting.

To this end, bovine serum, EDTA-plasma, and EDTA-plasma fortified with acetylsalicylic acid (ASA) as antioxidant were compared with regard to their applicability for metabolomics studies. Metabolic fingerprints were generated by means of GC-TOFMS. For data processing, the LECO ChromaTOF software in combination with the in-house developed retention time correction and data alignment tool INCA were used. To confirm that the observed signal intensities in the GC-TOFMS fingerprints reflected true metabolite abundances, 19 amino acids, glucose, and 6 organic acids were quantified using stable isotope-labeled internal standards.

The results presented in this chapter were published with Dr. Katja Dettmer (Institute and Chair of Functional Genomics, University of Regensburg) as joint first author in [198].

9.2 Material and Methods

9.2.1 Collection of blood specimens

Blood specimens were drawn by jugular venipuncture into 7-mL evacuated EDTA-coated tubes and 7-mL evacuated serum-gel tubes (Vacuette, Greiner Bio-One, Kremsmünster, Austria) from 18 lactating Brown Swiss dairy cows held at the research farm Veitshof of the Technical University Munich, Freising-Weihenstephan, Germany. The tubes were gently inverted 4 times to ensure proper mixing of the blood and additive in the specimens. Plasma was separated by centrifugation (15 min, 2000 xg, 4 °C) immediately after sampling and stored frozen at -80 °C until analysis. To prepare EDTA-ASA-plasma blood was also drawn in into 7-mL evacuated EDTA-coated tubes, followed by the immediate addition of 200 µL stabilization solution prior to sample mixing. The aqueous stabilization solution contained 0.3 M Na_2EDTA and 1 % (w/v) ASA. Blood serum samples were kept at room temperature for blood to clot and centrifuged (15 min, 2000 xg, 4 °C) within 1 hour after blood collection.

9.2.2 GC-MS analysis

9.2.2.1 GC-qMS of amino acids

Quantitative analysis of amino acids was performed in 20-µL sample aliquots using GC-qMS after derivatization with propyl chloroformate.[162] For details see chapters 5.2.3 and 5.3.3.

9.2.2.2 GC-TOFMS for metabolic fingerprinting and quantification of β-hydroxybutyrate, glucose, glycolysis and citrate cycle intermediates

Ten-µL serum/plasma aliquots were extracted with 50 µL methanol containing the extraction standards $[2,2,4,4-^2H_4]$citrate, $[U-^{13}C]$fumarate, $[U-^{13}C]$glucose, $[U-^{13}C]$β-hydroxybutyrate, $[U-^{13}C]$lactate, $[2,3,3-^2H_3]$malate, $[U-^{13}C]$pyruvate, $[U-^2H]$succinate, $[^2H_7]$trans-cinnamate, and unlabeled norvaline. Samples were vortexed and then centrifuged at 4 °C and 3,375 xg for 5 min. The supernatant was transferred into a 2-mL glass vial with a 0.2-mL glass insert. Pellets were re-extracted twice more with 50 µL of methanol. The extracts were combined, and evaporated and derivatized as described in chapter 5.2.1.

Sample preparation using 10-µL aliquots was repeated for metabolic fingerprinting without the addition of the extraction standards due to potential interferences of the stable isotope-labeled analogs with the endogenous compounds in data analysis.

Calibration was performed using metabolite standard solutions in methanol. Hundred µL of different calibration levels were transferred into a glass vial with glass insert and 10 µL of the surrogate solution containing stable isotope-labeled $[2,2,4,4-^2H_4]$citrate, $[U-^{13}C]$fumarate, $[U-^{13}C]$glucose, $[U-^{13}C]$β-hydroxybutyrate, $[U-^{13}C]$lactate, $[2,3,3-^2H_3]$malate, $[U-^{13}C]$pyruvate, $[U-^2H]$succinate, $[^2H_7]$trans-cinnamate, and unlabeled norvaline (1 mM each) were added. The standards were dried and derivatized as described in chapter 5.2.1. Twelve calibration points were generated over a concentration range of 0.5 – 500 µM.

For metabolic fingerprinting and targeted quantitative analysis of citrate, fumarate, glucose, β-hydroxybutyrate, lactate, malate, pyruvate, and succinate, the LECO Pegasus GC×GC-TOFMS instrument was used in 1D mode according to chapter 5.3.1. The modulator was turned off and mass spectra were collected at 50 spectra/s.

Sample preparation of serum, EDTA-plasma and EDTA-ASA-plasma samples was performed by Nadine Nürnberger (Institute and Chair of Functional Genomics, University of Regensburg).

9.2.2.3 Data Analysis

Raw data were processed with the LECO ChromaTOF software version 3.35 according to chapter 6.2.2. Peaks with an *S/N* ≥250 were selected. Retention time correction and data alignment was performed using INCA as described in chapter 6.2.3 and 6.2.4. Normalization and testing were performed according to chapter 6.2.5. INCA was adapted to the analysis of 1D-GC-TOFMS data by simply excluding the 2D retention time as a parameter. Quantile plots were generated for data visualization. For each feature pairwise quantiles were plotted in different colors to visualize the variance in metabolite abundances.

Quantile plots were generated by Wolfram Gronwald (Institute and Chair of Functional Genomics, University of Regensburg). Bland-Altman analyses were done by Katja Dettmer.

9.3 Results and Discussion

9.3.1 Metabolite fingerprinting

For an overall comparison of the metabolite fingerprints obtained by GC-TOFMS analysis of methanolic extracts of serum, EDTA-plasma and EDTA-ASA-plasma samples, peak lists were generated with the LECO ChromaTOF software that included all signals with a *S/N* ≥250. For each peak the area of the *m/z* 73 fragment ion trace, which is shared by all TMS derivatives, was exported as a quantitative measure. The peak lists from all samples were aligned into a single data matrix using INCA. After alignment a total of 654 features were exported from the GC-TOFMS data.

Not all features were detected in all samples. Therefore, only features present in 10 or more (>50 %) samples per group were selected for further analysis, which left a total of 128 features. The area integral of each feature was normalized by using the area integral of tridecanoic acid at *m/z* 271 to account for variability during derivatization and injection. The normalized data was then subjected to a log

transformation. To visualize the variance in metabolite abundances of the aligned dataset quantile plots were generated (Figure 9.1).

Figure 9.1 shows a 1D chromatogram of an EDTA-ASA-plasma sample. For each feature of one dataset 0 %- and 100 %- (blue), 10 %- and 90 %- (light blue), 30 %- and 70 %- (yellow), and 50 %-quantiles (red) were plotted against the features´ median of the other dataset (black). Two quantile plots of this kind are exemplified in Figure 9.1B and C.
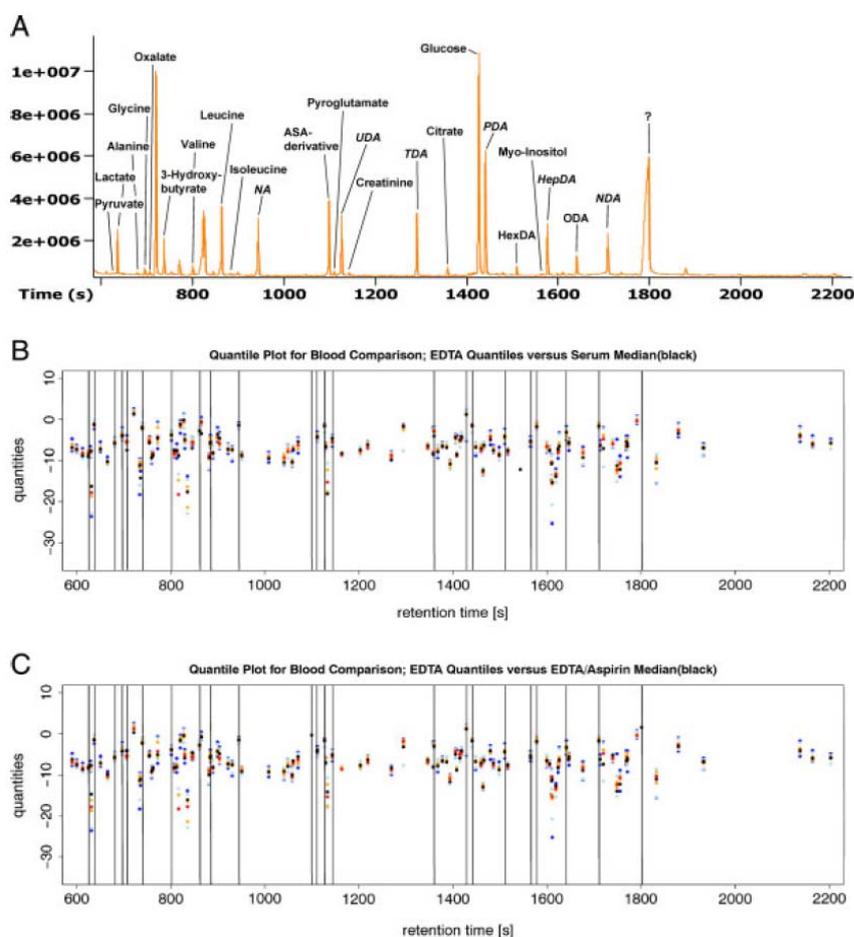
Figure 9.1. TIC chromatogram of an EDTA-ASA-plasma sample using GC-TOFMS (A). Several major components are labeled and the added internal standards are italicized. Quantile plots of aligned metabolic fingerprints of 18 EDTA-plasma samples (B, C). 0 %- and 100 %- (blue), 10 %- and 90 %- (light blue), 30 %- and 70%- (yellow), and 50 %-quantiles (red) are plotted against the features´ median of 18 bovine serum (B) and EDTA-ASA-plasma (C) samples (black). Reproduced from [198].

When comparing the quantiles of EDTA-plasma to the median of EDTA-ASA-plasma 63 % of the analytes´ medians were located within the 30 %- and 70 %-quantile and

95 % within the 10 %- and 90 %-quantiles. The comparison of EDTA-plasma quantiles to the median of serum yielded 65 % of the analytes´ medians within the 30 %- and 70 %-quantile and 87 % within the 10 %- and 90 %-quantiles.

A t-test with equal variances was performed for group-wise comparison and the p-values obtained were corrected for multiple testing according to Benjamini and Hochberg (Table 9.1).[113] A total of 6, 9, and 21 significant features with a FDR of <0.05 were identified when comparing EDTA- versus EDTA-ASA-plasma, EDTA-plasma versus serum, and EDTA-ASA-plasma versus serum, respectively. Not surprisingly, the feature discriminating EDTA-ASA-plasma mostly from EDTA-plasma and serum, respectively, was a peak representing acetylsalicylic acid (ASA derivative). The comparison of EDTA-plasma with serum yielded citrate and oxalate as significantly discriminating features. These compounds also differed significantly in concentration between EDTA-ASA-plasma and serum. In addition, a significant difference in lactate concentration was found when comparing EDTA-ASA-plasma with serum. Also, several unknowns were detected that could not be matched against both the custom-made and the NIST 05 library with a minimum similarity of 70 %.

Table 9.1. Number and identities of the significant features detected in the different sample types by means of INCA-based analysis of metabolic fingerprinting GC-TOFMS data. Reproduced from [198].

	EDTA vs. EDTA-ASA	EDTA vs. Serum	EDTA-ASA vs. Serum
Significant features (FDR <0.05)	6	9	21
Metabolites identified that distinguished the sample groups	Acetyl salicylic acid derivative, unknowns	Citrate, oxalate, unknowns	Acetyl salicylic acid derivative, citrate, lactate, oxalate, unknowns

To verify that the variance of the features was not a result of analytical imprecision, quantile plots of technical replicates were created (Figure 9.2). One serum, EDTA-plasma, and EDTA-ASA-plasma sample, respectively, was injected 5 times each. The peak lists were processed according to the steps described above. Zero and 100 %-quantiles are represented by blue, 25 %- and 75 %-quantiles by yellow, and the 50 %-quantile by red points. The quantile plots for the technical replicates show slight to no variance in metabolite abundance.
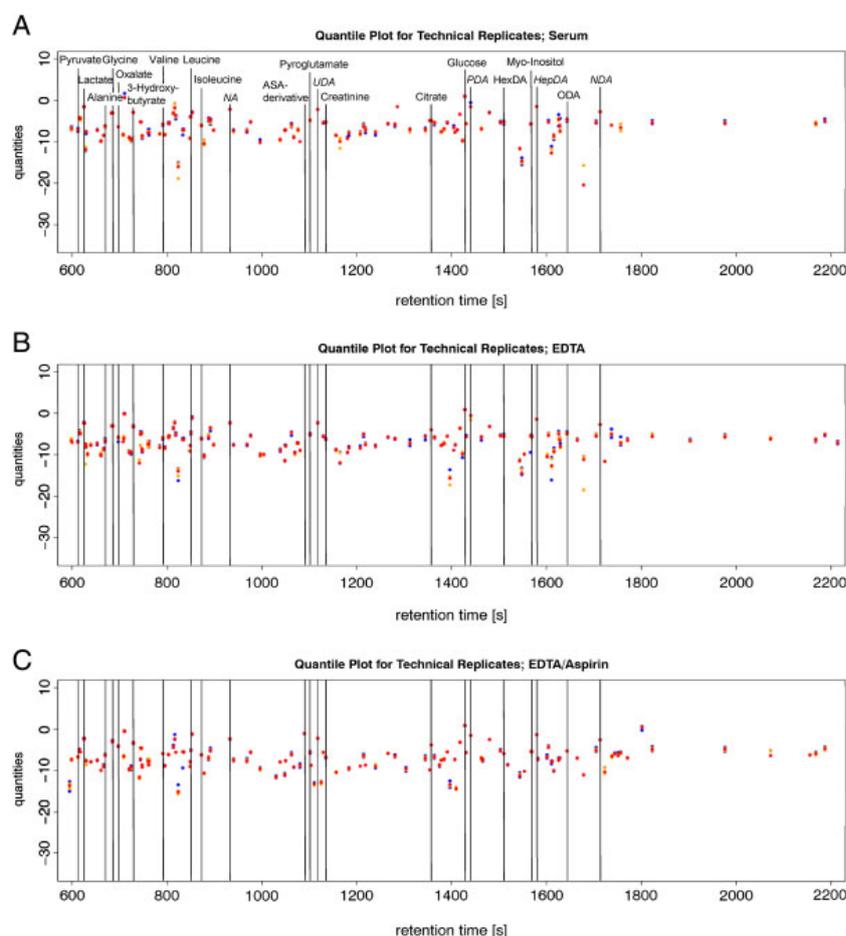
Figure 9.2. Quantile plot of technical replicates. One serum (A), one EDTA-plasma (B), and one EDTA-ASA-plasma (C) sample was injected 5 times each. Zero and 100 %-quantiles are represented by blue, 25 %- and 75 %-quantiles by yellow, and the 50 %-quantile by red points. Reproduced from [198].

9.3.2 Targeted analysis

To confirm the results of the metabolic fingerprinting approach a total of 26 metabolites, including 19 amino acids, glucose, and 7 organic acids, were quantified in serum, EDTA-plasma, and EDTA-ASA-plasma samples collected from 18 dairy cows, of which 7 were in early and 11 in late lactation. LLOQ, average concentration with standard deviation, and concentration range observed for each metabolite in the three sample types are listed in Table 9.2.

The LLOQ of the amino acids and the remaining analytes were determined for injection volumes of 20 μL and 10 μL, respectively. With the exception of aspartate, all analytes were detected above the LLOQ in all samples. Interestingly, aspartate could be detected above the LLOQ in 10 serum samples, but only in 3 and 2 of the

ETDA-plasma and EDTA-ASA-plasma samples, respectively. This indicates that overall serum yields slightly higher aspartate values compared to plasma. It should be noted that the samples were analyzed in random order to avoid any bias introduced by the instrumental analysis. Aspartate was excluded from further analysis due to the low number of data points. As expected the highest abundant metabolite among the quantified analytes was glucose with an average concentration of 2.8 mM in EDTA-plasma and 2.6 mM in EDTA-ASA-plasma and serum.

Table 9.2. LLOQ, average concentration with SD, and concentration range for the analytes quantified in EDTA-plasma, EDTA-ASA-plasma, and serum (n=18 for each). Reproduced from [198].

	LLOQ	EDTA-Plasma		EDTA-ASA-Plasma		Serum	
		Mean ± SD [µM]	Range [µM]	Mean ± SD [µM]	Range [µM]	Mean ± SD [µM]	Range [µM]
Pyruvate	4.88 [a]	54 ± 9	44 - 76	49 ± 8	28 - 62	47 ± 11	30 - 70
Lactate	2.44 [a]	511 ± 129	310 - 818	508 ± 150	254 - 906	662 ± 105	530 - 842
β-Hydroxy-butyrate	4.88 [a]	534 ± 297	304 - 1636	517 ± 330	326 - 1778	536 ± 300	360 - 1666
Succinate	9.76 [a]	36 ± 4	28 - 44	33 ± 6	24 - 48	38 ± 5	30 - 46
Malate	4.88 [a]	21 ± 3	16 - 26	19 ± 3	14 - 28	20 ± 5	8 - 34
Citrate	4.88 [a]	286 ± 61	174 - 398	262 ± 62	164 - 360	212 ± 49	136 - 278
Glucose	4.88 [a]	2801 ± 374	2020 - 3276	2623 ± 251	2076 - 3038	2661 ± 274	2050 - 2970
Alanine	0.75 [b]	211 ± 49	128 - 336	201 ± 43	127 - 295	234 ± 57	142 – 353
Glycine	7.50	404 ± 146	229 - 717	381 ± 134	213 - 639	466 ± 156	284 - 849
α-Amino-butyrate	0.75	24 ± 8	14 - 45	23 ± 7	10 – 34	25 ± 8	14 – 42
Valine	0.75	183 ± 47	89 - 281	174 ± 43	88 - 262	202 ± 58	112 - 325
Leucine	0.75	105 ± 24	53 - 158	99 ± 22	53 – 151	117 ± 31	68 – 190
Isoleucine	2.25	93 ± 20	46 - 134	88 ± 19	44 - 131	103 ± 26	56 - 154
Proline	0.75	69 ± 13	36 - 90	66 ± 13	35 – 91	79 ± 17	45 - 117
Asparagine	30.00	35 ± 8	17 - 45	33 ± 8	19 - 44	40 ± 9	23 – 59
Aspartate [c]	7.50	9 ± 1	8 - 10	8 ± 1	7.6 - 8.4	9 ± 1	8 - 12
Methionine	7.50	18 ± 4	10 - 26	18 ± 4	11 – 25	21 ± 5	14 – 32
Glutamate	7.50	44 ± 10	34 - 66	40 ± 10	25 - 68	51 ± 13	36 - 88
Phenyl-alanine	2.25	46 ± 9	32 - 62	44 ± 9	30 – 60	52 ± 10	39 – 72
Ornithine	2.25	40 ± 15	17 - 72	36 ± 14	16 - 69	45 ± 17	19 - 76
Glutamine	75.00	246 ± 46	150 - 307	235 ± 45	140 – 310	281 ± 63	178 – 403
Lysine	2.25	73 ± 16	43 - 103	66 ± 16	36 - 104	84 ± 21	51 – 127
Histidine	30.00	44 ± 17	17 - 73	42 ± 15	16 – 61	48 ± 19	19 - 87
Tyrosine	2.25	45 ± 12	22 - 63	41 ± 11	24 - 59	48 ± 14	25 – 71
Tryptophan	0.75	50 ± 11	31 - 76	46 ± 9	30 – 65	56 ± 13	35 - 82
Hippurate	4.50	101 ± 15	81 - 127	96 ± 17	66 - 132	114 ± 20	80 - 141

[a] LLOQ was calculated for a sample volume of 10 µL.
[b] LLOQ was calculated for a sample volume of 20 µL.
[c] Quantified above LLOQ in n=3 for EDTA-plasma, n=2 for EDTA-ASA-plasma, and n=10 for serum.

The three sample types were compared with each other by Bland-Altman plots.[164] Since it is not possible to show all Bland-Altman plots, the mean difference, limits of

agreement, and percentage of the mean difference from the mean concentration are listed in Table 9.3 comparing serum versus EDTA-plasma, serum versus EDTA-ASA-plasma, and EDTA-plasma versus EDTA-ASA-plasma, respectively. An ideal agreement between two datasets is observed, if the absolute mean difference is almost zero. In order to evaluate the data agreement a cutoff of 15 % was defined to be acceptable for the percentage of the mean difference (%đ) from the mean concentration.

Table 9.3. The mean difference (đ), limits of agreement (±1.96 SD of the differences), and the percentage of the difference from the mean of the Bland-Altman analyses are presented. Calculations were performed with concentrations in µM. Reproduced from [198].

	Serum vs. EDTA-plasma			Serum vs. EDTA-ASA-plasma				EDTA plasma vs. EDTA-ASA-plasma		
	đ	1.96 SD	%đ from mean	đ	1.96 SD	%đ from mean	%đ from mean corr.	đ	1.96 SD	%đ from mean
Pyruvate	-6.7	11.6	13.3	-2.4	20.9	5.1	13.6	4.2	16.2	8.2
Lactate	151.3	243.3	25.8	154.6	285.2	26.4	26.7	3.2	241.9	0.6
β-Hydroxy-butyrate	2.1	93.9	0.4	19.7	121.1	3.7	0.5	17.6	111.0	3.3
Succinate	2.1	12.4	5.7	5.1	14.5	14.5	5.7	3.0	15.3	8.8
Malate	-0.3	11.0	1.6	1.3	12.0	6.7	1.7	1.7	8.7	8.4
Citrate	-74.1	65.2	29.8	-50.9	49.9	21.5	47.8	23.2	60.0	8.5
Glucose	-140.1	430.8	5.1	38.0	479.9	1.4	4.9	178.1	507.1	6.6
Alanine	22.1	51.7	10.0	32.3	68.1	14.9	10.7	10.2	30.4	4.9
Glycine	62.1	120.3	14.3	84.8	127.0	20.0	22.3	22.7	62.9	5.8
α-Amino-butyrate	1.8	11.1	7.5	2.8	8.4	11.9	6.9	1.0	9.9	4.4
Valine	19.5	51.5	10.1	28.3	62.7	15.1	6.9	8.8	21.7	4.9
Leucine	12.5	32.1	11.2	18.0	38.4	16.6	7.7	5.6	11.8	5.4
Isoleucine	10.2	27.0	10.5	14.8	31.4	15.5	7.4	4.6	10.3	5.1
Proline	9.9	18.5	13.5	13.4	22.2	18.6	12.6	3.5	9.5	5.1
Asparagine	5.5	9.7	14.6	7.0	10.5	19.0	12.3	1.5	5.4	4.5
Methionine	2.5	5.4	12.6	2.9	6.1	15.1	11.7	0.5	5.2	2.5
Glutamate	7.3	16.3	15.4	11.0	19.1	23.9	19.1	3.7	12.3	8.6
Phenyl-alanine	5.9	13.5	12.1	7.9	15.4	16.5	10.8	2.0	5.8	4.5
Ornithine	5.2	12.0	12.2	8.9	14.6	21.8	7.1	3.7	5.7	9.6
Glutamine	37.5	97.9	14.2	45.9	95.4	17.8	15.5	10.9	40.0	4.5
Lysine	11.2	25.8	14.3	17.8	26.9	23.7	10.5	6.6	18.3	9.5
Histidine	3.8	19.1	8.3	6.4	17.7	14.2	8.1	2.5	9.9	5.9
Tyrosine	3.4	13.8	7.3	7.2	13.8	16.1	5.6	3.8	9.0	8.8
Tryptophan	6.2	13.4	11.7	9.4	16.0	18.4	9.7	3.2	6.1	6.8
Hippurate	12.9	27.9	12.0	17.7	35.1	16.8	12.4	4.8	22.1	4.8

When comparing serum to EDTA-plasma the majority of analytes show a positive value for the mean difference demonstrating that slightly higher values were

measured in serum. Only pyruvate, malate, citrate, and glucose yielded a negative mean difference. However, the %đ was less than 15 % for most analytes. Only lactate and citrate featured values of 26 % and 30 %, respectively.

The comparatively lower values for citrate in serum might be due to the formation of complexes with Ca^{2+} and Mg^{2+} ions.[53] The divalent cation-citrate complexes formed will precipitate during methanol extraction of metabolites, resulting in lower citrate values. In EDTA-plasma, the EDTA will chelate Ca^{2+} and Mg^{2+} ions yielding higher citrate values. EDTA-plasma should therefore be favored in studies that focus on the quantification of citrate and the related metabolites isocitrate and cis-aconitate, which possibly form analogous divalent cation complexes in serum. The discrepancy in lactate, on the other hand, is probably due to ongoing glycolysis in the waiting period for blood to clot prior to separation of serum from the blood cells by centrifugation, resulting in slightly lower glucose values and increased lactate concentrations.

The representative Bland-Altman plots (Figure 9.3A and B) clearly show that lactate levels are significantly (p=0.007) lower in EDTA-plasma than serum, while the reverse is true for citrate (p=0.002). Adding ASA to EDTA-plasma did not show any major effects on the analytes quantified here. The average concentrations across all analytes were approximately 5.2 % lower in EDTA-ASA-plasma in comparison to EDTA-plasma, which is attributed to the dilution of the plasma (about 3.5 mL) through addition of the aspirin solution (200 µL), which accounts for about 5.7 %. In general the data from EDTA-plasma and EDTA-ASA-plasma agreed well with %đ less than 10 %. However, data were not corrected for the dilution, because the blood volume sampled might vary slightly from sample to sample, which makes the collection of EDTA-ASA-plasma less accurate.

The comparison of serum with EDTA-ASA-plasma revealed higher mean differences with lower concentrations for most analytes in EDTA-ASA-plasma. Again this can be attributed to the dilution of plasma by the Aspirin solution. In order to test this hypothesis, the difference in concentration between EDTA-plasma and EDTA-ASA-plasma found for each analyte in the individual samples was added to the concentration determined in the EDTA-ASA-plasma to correct for the dilution and the Bland-Altman analysis comparing EDTA-ASA-plasma to serum was performed again (Table 9.3, %đ from mean corr.). The resulting %đ values were comparable to the %đ values obtained for the comparison of serum versus EDTA-plasma.
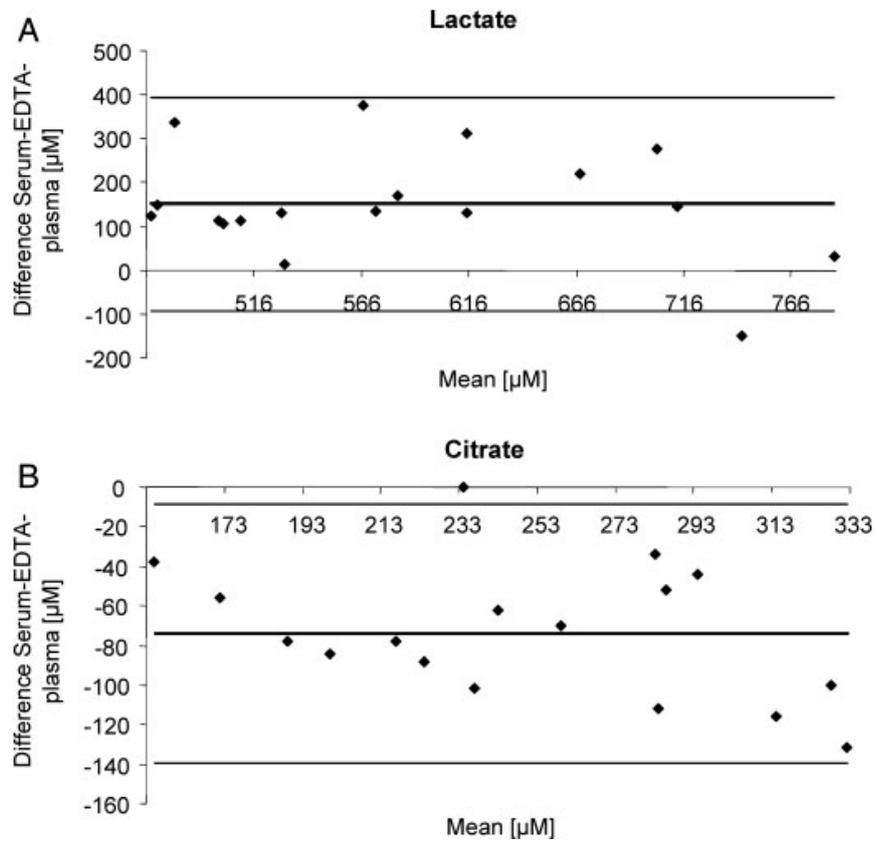
Figure 9.3. Bland-Altman plots for lactate and citrate comparing serum versus EDTA-plasma. The mean difference ($\bar{d}$) is indicated in the plots by the centerline, while the outer lines represent the limits of agreement corresponding to ±1.96 SD of the differences. Reproduced from [198].

10 Conclusions and perspectives

GC×GC particularly in combination with TOFMS has emerged as a promising and powerful analytical tool for metabolomics investigations. Compared to conventional 1D-GC, GC×GC provides a more comprehensive coverage of metabolomes due to increased chromatographic resolution, peak capacity, detection sensitivity, and broader dynamic range. In the early stages of the doctoral research work presented here, application of GC×GC-TOFMS to metabolomics was impeded by the lack of appropriate data processing, alignment, and analysis tools for the comparative analysis of metabolic fingerprints taking the entire available, also non-targeted information from all experiments into account for statistical analysis and, thereby, the identification of novel biomarkers.

In this thesis, a GC×GC-TOFMS based metabolic fingerprinting strategy is presented that was developed in an interdisciplinary collaboration between analytical chemistry and bioinformatics. Data preprocessing tools provided by vendor software were used in combination with the in-house programmed retention time correction and peak alignment module INCA to form a data matrix suitable for multivariate statistical analysis. The metabolic fingerprinting routine was validated by a spike-in experiment, compared to the later introduced commercial alignment tool Statistical Compare, and applied to biomedical studies.

GC×GC-EI-TOFMS was comprehensively evaluated over various GC-MS techniques (GC-EI-TOFMS, GC-CI-qMS, GC-EI-qMS, and GC-APCI-TOFMS) used for metabolomics studies employing a standard mixture of 43 compounds. GC×GC-TOFMS showed the best performance among the tested techniques with linear ranges of over three orders of magnitude, LLOQs in the sub-micromolar range, and LODs in the nanomolar range. RSDs for calibration replicates (fivefold derivatized) were below 10 % and technical replicates (fivefold injected) below 5 % for 80 % of all compounds.

GC×GC-TOFMS was applied to the quantitative analysis of amino acid enantiomers as their methyl chloroformate derivatives in physiological fluids. GC×GC-TOFMS method was validated by analyzing AAEs in urine and serum specimens, respectively, and by comparing the results with those of a previously established GC-qMS method using single ion monitoring. Peak resolution was enhanced by coupling

a γ-cyclodextrin (Rt$^{\circledR}$-γDEXsa) with an amino acid selective (ZB-AAA) column. Twenty AAEs including the structural isomers L-leucine/D-isoleucine could be baseline separated. Except for methionine enantiomers, improved LLOQs were obtained in contrast to the GC-qMS method. The analysis of a serum specimen in octaplicate yielded RSDs below 5 %, except for D-Ala, D-Asp, and D-Asn. AAE levels in patients suffering from liver cirrhosis showed significantly increased D-AA concentrations and slightly decreased L-AA levels compared to a control group.

Despite the progress made, there is still need for improvement in the processing of GC×GC-TOFMS metabolic fingerprinting data. Particularly the precision of deconvolution algorithms needs to be further optimized to preserve the quality of the provided data. Also, data processing strategies need to be refined to enable better use of adequate metabolite-specific unique *m/z* ions for quantification.

A lingering issue is the computationally intensive handling of GC×GC-TOFMS data. As a consequence of high sample throughput, multidimensional separation, and increased mass spectral acquisition rates, datasets become extremely large. Complete chromatographic analyses can only become more attractive if state-of-the-art architectures, like 64-bit and multi-core processing, are supported and further advances are made by implementing less computationally demanding algorithms.

Metabolomics aims to determine absolute concentration levels for all metabolites in a biological system. However, due to practical limitations, e.g. the absence of standard reference material, global GC×GC-TOFMS studies still rely on relative rather than absolute concentrations to compare metabolites in different samples. Relative quantification tends to have its flaws as metabolite recoveries can vary according to differences in matrix composition.

So far, GC×GC-TOFMS has been mostly used for smaller metabolomics studies. Its robustness to run large sample sets over a long period has yet to be determined. In this context the development of protocols to correct for batch effects will be of uttermost importance.

The metabolomics field certainly trends towards routine implementation of GC×GC-TOFMS into a global metabolomic platform that uses complementary tools including LC-MS and NMR. Also, combining GC×GC with high-resolution mass analyzers should drive future efforts in characterizing unknown components of the metabolome.

In the end, the rapid rate of methodical, hardware and software developments promises solutions to many of these problems in years to come.

11 References

(1)   Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. *Trends Biotechnol* **1998**, *16*, 373-378.
(2)   Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom Rev* **2006**, *26*, 51-78.
(3)   Dunn, W. B. *Phys Biol* **2008**, *5*, 011001.
(4)   Bertsch, W. *J. High Resol. Chromatogr.* **1999**, *22*, 647–665.
(5)   Waldhier, M. C.; Dettmer, K.; Gruber, M. A.; Oefner, P. J. *J Chromatogr B Analyt Technol Biomed Life Sci* **2010**, *878*, 1103-1112.
(6)   Almstetter, M. F.; Dettmer, K.; Oefner, P. J. *Anal Bioanal Chem* **2011**, accepted.
(7)   Almstetter, M. F.; Dettmer, K.; Oefner, P. J. In *Functional Genomics - Methods and Protocols, Second Edition*; Kaufmann, M., Klinger, C., Eds.; Humana Press: New York, USA, 2012; Vol. 815, pp 399-412.
(8)   Dettmer, K.; Almstetter, M. F.; Wachsmuth, C. J.; Oefner, P. J. In *The Handbook of Plant Metabolomics - Metabolite Profiling and Networking*; Wiley-VCH: Weinheim, Germany, 2012; in print.
(9)   Nicholson, J. K.; Lindon, J. C. *Nature* **2008**, *455*, 1054-1056.
(10)   Dunn, W. B.; Broadhurst, D. I.; Atherton, H. J.; Goodacre, R.; Griffin, J. L. *Chem Soc Rev* **2011**, *40*, 387-426.
(11)   Bruggeman, F. J.; Westerhoff, H. V. *Trends Microbiol* **2007**, *15*, 45-50.
(12)   van der Greef, J.; Martin, S.; Juhasz, P.; Adourian, A.; Plasterer, T.; Verheij, E. R.; McBurney, R. N. *J Proteome Res* **2007**, *6*, 1540-1559.
(13)   Urbanczyk-Wochniak, E.; Luedemann, A.; Kopka, J.; Selbig, J.; Roessner-Tunali, U.; Willmitzer, L.; Fernie, A. R. *EMBO Rep* **2003**, *4*, 989-993.
(14)   Dunn, W. B.; Bailey, N. J.; Johnson, H. E. *Analyst* **2005**, *130*, 606-625.
(15)   Kell, D. B. *BMC Med Genomics* **2009**, *2*, 2.
(16)   Mounicou, S.; Szpunar, J.; Lobinski, R. *Chem Soc Rev* **2009**, *38*, 1119-1138.
(17)   Dettmer, K.; Hammock, B. D. *Environ Health Perspect* **2004**, *112*, A396-397.
(18)   Forster, J.; Famili, I.; Fu, P.; Palsson, B. O.; Nielsen, J. *Genome Res* **2003**, *13*, 244-253.
(19)   Fiehn, O. *Plant Mol Biol* **2002**, *48*, 155-171.
(20)   Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181-1189.
(21)   German, J. B.; Gillies, L. A.; Smilowitz, J. T.; Zivkovic, A. M.; Watkins, S. M. *Curr Opin Lipidol* **2007**, *18*, 66-71.
(22)   Morelle, W.; Canis, K.; Chirat, F.; Faid, V.; Michalski, J. C. *Proteomics* **2006**, *6*, 3993-4015.
(23)   Bou Khalil, M.; Hou, W.; Zhou, H.; Elisma, F.; Swayne, L. A.; Blanchard, A. P.; Yao, Z.; Bennett, S. A.; Figeys, D. *Mass Spectrom Rev* **2010**, *29*, 877-929.
(24)   Nobeli, I.; Ponstingl, H.; Krissinel, E. B.; Thornton, J. M. *J Mol Biol* **2003**, *334*, 697-719.
(25)   Herrgard, M. J.; Swainston, N.; Dobson, P.; Dunn, W. B.; Arga, K. Y.; Arvas, M.; Bluthgen, N.; Borger, S.; Costenoble, R.; Heinemann, M.; Hucka, M.; Le Novere, N.; Li, P.; Liebermeister, W.; Mo, M. L.; Oliveira, A. P.; Petranovic, D.; Pettifer, S.; Simeonidis, E.; Smallbone, K.; Spasic, I.; Weichart, D.; Brent, R.; Broomhead, D. S.; Westerhoff, H. V.; Kirdar, B.; Penttila, M.; Klipp, E.; Palsson, B. O.; Sauer, U.; Oliver, S. G.; Mendes, P.; Nielsen, J.; Kell, D. B. *Nat Biotechnol* **2008**, *26*, 1155-1160.

(26)   Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; Mandal, R.; Sinelnikov, I.; Xia, J.; Jia, L.; Cruz, J. A.; Lim, E.; Sobsey, C. A.; Shrivastava, S.; Huang, P.; Liu, P.; Fang, L.; Peng, J.; Fradette, R.; Cheng, D.; Tzur, D.; Clements, M.; Lewis, A.; De Souza, A.; Zuniga, A.; Dawe, M.; Xiong, Y.; Clive, D.; Greiner, R.; Nazyrova, A.; Shaykhutdinov, R.; Li, L.; Vogel, H. J.; Forsythe, I. *Nucleic Acids Res* **2009**, *37*, D603-610.

(27)   Duarte, N. C.; Becker, S. A.; Jamshidi, N.; Thiele, I.; Mo, M. L.; Vo, T. D.; Srivas, R.; Palsson, B. O. *Proc Natl Acad Sci U S A* **2007**, *104*, 1777-1782.

(28)   Hall, R.; Beale, M.; Fiehn, O.; Hardy, N.; Sumner, L.; Bino, R. *Plant Cell* **2002**, *14*, 1437-1440.

(29)   Oksman-Caldentey, K. M.; Saito, K. *Curr Opin Biotechnol* **2005**, *16*, 174-179.

(30)   Smedsgaard, J.; Nielsen, J. *J Exp Bot* **2005**, *56*, 273-286.

(31)   Nicholson, J. K.; Holmes, E.; Lindon, J. C.; Wilson, I. D. *Nat Biotechnol* **2004**, *22*, 1268-1274.

(32)   Han, X.; Gross, R. W. *Expert Rev Proteomics* **2005**, *2*, 253-264.

(33)   Goodacre, R. *J Nutr* **2007**, *137*, 259S-266S.

(34)   Gates, S. C.; Sweeley, C. C. *Clin Chem* **1978**, *24*, 1663-1673.

(35)   Aharoni, A.; Ric de Vos, C. H.; Verhoeven, H. A.; Maliepaard, C. A.; Kruppa, G.; Bino, R.; Goodenowe, D. B. *Omics* **2002**, *6*, 217-234.

(36)   Ellis, D. I.; Dunn, W. B.; Griffin, J. L.; Allwood, J. W.; Goodacre, R. *Pharmacogenomics* **2007**, *8*, 1243-1266.

(37)   Allen, J.; Davey, H. M.; Broadhurst, D.; Heald, J. K.; Rowland, J. J.; Oliver, S. G.; Kell, D. B. *Nat Biotechnol* **2003**, *21*, 692-696.

(38)   Kaderbhai, N. N.; Broadhurst, D. I.; Ellis, D. I.; Goodacre, R.; Kell, D. B. *Comp Funct Genomics* **2003**, *4*, 376-391.

(39)   Theobald, U.; Mailinger, W.; Reuss, M.; Rizzi, M. *Anal Biochem* **1993**, *214*, 31-37.

(40)   Villas-Boas, S. G.; Hojer-Pedersen, J.; Akesson, M.; Smedsgaard, J.; Nielsen, J. *Yeast* **2005**, *22*, 1155-1169.

(41)   Wu, H.; Southam, A. D.; Hines, A.; Viant, M. R. *Anal Biochem* **2008**, *372*, 204-212.

(42)   Kohler, M.; Machill, S.; Salzer, R.; Krafft, C. *Anal Bioanal Chem* **2009**, *393*, 1513-1520.

(43)   Want, E. J.; O'Maille, G.; Smith, C. A.; Brandon, T. R.; Uritboonthai, W.; Qin, C.; Trauger, S. A.; Siuzdak, G. *Anal Chem* **2006**, *78*, 743-752.

(44)   Bruce, S. J.; Tavazzi, I.; Parisod, V.; Rezzi, S.; Kochhar, S.; Guy, P. A. *Anal Chem* **2009**, *81*, 3285-3296.

(45)   Shoemaker, J. D.; Elliot, W. H. *Journal of Chromatography B: Biomedical Sciences and Applications* **1991**, *562*, 125-138.

(46)   Husek, P. *J Chromatogr B Biomed Sci Appl* **1998**, *717*, 57-91.

(47)   Villas-Boas, S. G.; Delicado, D. G.; Akesson, M.; Nielsen, J. *Anal Biochem* **2003**, *322*, 134-138.

(48)   Eder, K. *J Chromatogr B Biomed Appl* **1995**, *671*, 113-131.

(49)   Halket, J. M.; Zaikin, V. G. *Eur J Mass Spectrom (Chichester, Eng)* **2003**, *9*, 1-21.

(50)   Roessner, U.; Wagner, C.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. *Plant J* **2000**, *23*, 131-142.

(51)   Steinhauser, D.; Kopka, J. *Exs* **2007**, *97*, 171-194.

(52)   Dieterle, F.; Riefke, B.; Schlotterbeck, G.; Ross, A.; Senn, H.; Amberg, A. *Methods Mol Biol* **2011**, *691*, 385-415.

(53) Gronwald, W.; Klein, M. S.; Kaspar, H.; Fagerer, S. R.; Nurnberger, N.; Dettmer, K.; Bertsch, T.; Oefner, P. J. *Anal Chem* **2008**, *80*, 9288-9297.

(54) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom Rev* **2007**, *26*, 51-78.

(55) Theodoridis, G.; Gika, H. G.; Wilson, I. D. *Mass Spectrom Rev* **2011**, doi: 10.1002/mas.20306, Epub ahead of print.

(56) Ramautar, R.; Mayboroda, O. A.; Somsen, G. W.; de Jong, G. J. *Electrophoresis* **2011**, *32*, 52-65.

(57) Ellis, D. I.; Goodacre, R. *Analyst* **2006**, *131*, 875-885.

(58) Biais, B.; Allwood, J. W.; Deborde, C.; Xu, Y.; Maucourt, M.; Beauvoit, B.; Dunn, W. B.; Jacob, D.; Goodacre, R.; Rolin, D.; Moing, A. *Anal Chem* **2009**, *81*, 2884-2894.

(59) t'Kindt, R.; Morreel, K.; Deforce, D.; Boerjan, W.; Van Bocxlaer, J. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009**, *877*, 3572-3580.

(60) Zakaria, M.; Gonnord, M.-F.; Guiochon, G. *J. Chromatogr.* **1983**, *271*, 127-192.

(61) Liu, Z. Y.; Philips, J. B. *J. Chromatogr. Sci.* **1991**, *29*, 227-231.

(62) Bertsch, W. *J. High Resol. Chromatogr.* **2000**, *23*, 167-181.

(63) Khummueng, W.; Harynuk, J.; Marriott, P. J. *Anal Chem* **2006**, *78*, 4578-4587.

(64) Gorecki, T.; Harynuk, J.; Panic, O. *J Sep Sci* **2004**, *27*, 359-379.

(65) Mondello, L.; Tranchida, P. Q.; Dugo, P.; Dugo, G. *Mass Spectrom Rev* **2008**, *27*, 101-124.

(66) Dalluge, J.; Beens, J.; Brinkman, U. A. *J Chromatogr A* **2003**, *1000*, 69-108.

(67) Poliak, M.; Kochman, M.; Amirav, A. *J Chromatogr A* **2008**, *1186*, 189-195.

(68) Ledford, E. B.; Billesbach, C. A. *J. High Resolut. Chromatogr.* **2000**, *23*, 202-204.

(69) Harynuk, J.; Górecki, T. *J. Sep. Sci.* **2002**, *25*, 304-310.

(70) Phillips, J. B.; Gaines, R. B.; Blomberg, J.; van der Wielen, F. W. M.; Dimandja, J.-M.; Green, V.; Granger, J.; Patterson, D.; Racovalis, L.; de Geus, H.-J.; de Boer, J.; Haglund, P.; Lipsky, J.; Sinha, V.; Ledford Jr., E. B. *J. High Resol. Chromatogr.* **1999**, *22*, 3-10.

(71) Kinghorn, R. M.; Marriott, P. J. *J. High Resol. Chromatogr.* **1999**, *22*, 235–238.

(72) Marriott, P. J.; Kinghorn, R. M. *TrAC Trends in Analytical Chemistry* **1999**, *18*, 114-125.

(73) Pursch, M.; Sun, K.; Winniford, B.; Cortes, H.; Weber, A.; McCabe, T.; Luong, J. *Anal Bioanal Chem* **2002**, *373*, 356-367.

(74) Beens, J.; Brinkman, U. A. *Analyst* **2005**, *130*, 123-127.

(75) Kristenson, E. M.; Korytar, P.; Danielsson, C.; Kallio, M.; Brandt, M.; Makela, J.; Vreuls, R. J.; Beens, J.; Brinkman, U. A. *J Chromatogr A* **2003**, *1019*, 65-77.

(76) Horning, E. C.; Horning, M. G.; Carroll, D. I.; Dzidic, I.; Stillwell, R. N. *Anal Chem* **1973**, *45*, 936-943.

(77) Horning, E. C.; Carroll, D. I.; Dzidic, I.; Haegele, K. D.; Lin, S.; Oertli, C. U.; Stillwell, R. N. *Clin Chem* **1977**, *23*, 13-21.

(78) Schiewek, R.; Lorenz, M.; Giese, R.; Brockmann, K.; Benter, T.; Gab, S.; Schmitz, O. J. *Anal Bioanal Chem* **2008**, *392*, 87-96.

(79) McEwen, C. N.; McKay, R. G. *J Am Soc Mass Spectrom* **2005**, *16*, 1730-1738.

(80) Kind, T.; Fiehn, O. *BMC Bioinformatics* **2006**, *7*, 234.

(81) Kind, T.; Fiehn, O. *BMC Bioinformatics* **2007**, *8*, 105.

(82) Bristow, T.; Harrison, M.; Sims, M. *Rapid Commun Mass Spectrom* **2010**, *24*, 1673-1681.

(83)   Garcia-Villalba, R.; Pacchiarotta, T.; Carrasco-Pancorbo, A.; Segura-Carretero, A.; Fernandez-Gutierrez, A.; Deelder, A. M.; Mayboroda, O. A. *J Chromatogr A* **2011**, *1218*, 959-971.

(84)   Carrasco-Pancorbo, A.; Nevedomskaya, E.; Arthen-Engeland, T.; Zey, T.; Zurek, G.; Baessmann, C.; Deelder, A. M.; Mayboroda, O. A. *Anal Chem* **2009**, *81*, 10071-10079.

(85)   Pacchiarotta, T.; Nevedomskaya, E.; Carrasco-Pancorbo, A.; Deelder, A. M.; Mayboroda, O. A. *J Biomol Tech* **2010**, *21*, 205-213.

(86)   Harshman, R. A. *UCLA Working Papers in Phonetics* **1970**, *16*, 1–84.

(87)   Ni, M.; Reichenbach, S. E.; Visvanathan, A.; TerMaat, J.; Ledford, E. B., Jr. *J Chromatogr A* **2005**, *1086*, 165-170.

(88)   Hollingsworth, B. V.; Reichenbach, S. E.; Tao, Q.; Visvanathan, A. *J Chromatogr A* **2006**, *1105*, 51-58.

(89)   David, F.; Tienpont, B.; Sandra, P. *J Sep Sci* **2008**, *31*, 3395-3403.

(90)   Koek, M. M.; van der Kloet, F. M.; Kleemann, R.; Kooistra, T.; Verheij, E. R.; Hankemeier, T. *Metabolomics* **2010**, *7*, 1-14.

(91)   Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Fraga, C. G.; Nilsson, E. J.; Synovec, R. E. *J Chromatogr A* **2004**, *1056*, 145-154.

(92)   Hoggard, J. C.; Synovec, R. E. *Anal Chem* **2007**, *79*, 1611-1619.

(93)   Hoggard, J. C.; Synovec, R. E. *Anal Chem* **2008**, *80*, 6677-6688.

(94)   Sinha, A. E.; Hope, J. L.; Prazen, B. J.; Nilsson, E. J.; Jack, R. M.; Synovec, R. E. *J Chromatogr A* **2004**, *1058*, 209-215.

(95)   Pierce, K. M.; Hope, J. L.; Hoggard, J. C.; Synovec, R. E. *Talanta* **2006**, *70*, 797-804.

(96)   Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Young, E. T.; Synovec, R. E. *Anal Chem* **2006**, *78*, 2700-2709.

(97)   Pierce, K. M.; Hoggard, J. C.; Hope, J. L.; Rainey, P. M.; Hoofnagle, A. N.; Jack, R. M.; Wright, B. W.; Synovec, R. E. *Anal Chem* **2006**, *78*, 5068-5075.

(98)   Mohler, R. E.; Dombek, K. M.; Hoggard, J. C.; Pierce, K. M.; Young, E. T.; Synovec, R. E. *Analyst* **2007**, *132*, 756-767.

(99)   Mohler, R. E.; Tu, B. P.; Dombek, K. M.; Hoggard, J. C.; Young, E. T.; Synovec, R. E. *J Chromatogr A* **2008**, *1186*, 401-411.

(100)   Welthagen, W.; Shellie, R. A.; Spranger, J.; Ristow, M.; Zimmermann, R.; Fiehn, O. *Metabolomics* **2005**, *1*, 65-73.

(101)   Shellie, R. A.; Welthagen, W.; Zrostlikova, J.; Spranger, J.; Ristow, M.; Fiehn, O.; Zimmermann, R. *J Chromatogr A* **2005**, *1086*, 83-90.

(102)   Fraga, C. G.; Prazen, B. J.; Synovec, R. E. *Anal Chem* **2001**, *73*, 5833-5840.

(103)   van Mispelaar, V. G.; Tas, A. C.; Smilde, A. K.; Schoenmakers, P. J.; van Asten, A. C. *J Chromatogr A* **2003**, *1019*, 15-29.

(104)   Pierce, K. M.; Wood, L. F.; Wright, B. W.; Synovec, R. E. *Anal Chem* **2005**, *77*, 7735-7743.

(105)   Zhang, D.; Huang, X.; Regnier, F. E.; Zhang, M. *Anal Chem* **2008**, *80*, 2664-2671.

(106)   Oh, C.; Huang, X.; Regnier, F. E.; Buck, C.; Zhang, X. *J Chromatogr A* **2008**, *1179*, 205-215.

(107)   Almstetter, M. F.; Appel, I. J.; Gruber, M. A.; Lottaz, C.; Timischl, B.; Spang, R.; Dettmer, K.; Oefner, P. J. *Anal Chem* **2009**, *81*, 5731-5739.

(108)   Almstetter, M. F.; Appel, I. J.; Dettmer, K.; Gruber, M. A.; Oefner, P. J. *J Chromatogr A* **2011**, *1218*, 7031-7038.

(109)   Wang, B.; Fang, A.; Heim, J.; Bogdanov, B.; Pugh, S.; Libardoni, M.; Zhang, X. *Anal Chem* **2010**, *82*, 5069-5081.

(110) Castillo, S.; Mattila, I.; Miettinen, J.; Oresic, M.; Hyotylainen, T. *Anal Chem* **2011**, *83*, 3058-3067.

(111) Abdi, H. In *Encyclopedia of Measurement and Statistics*; Salkind, N. J., Ed.; Sage Publications: Thousand Oaks, USA, 2007, pp 103-107.

(112) Westfall, P. H.; Young, S. S. *Resampling-based multiple testing: examples and methods for P-value adjustment*; John Wiley and Sons: New York, USA, 1993.

(113) Benjamini, Y.; Hochberg, Y. *Journal of the Royal Statistical Society: Series B* **1995**, *57*, 289-300.

(114) Broadhurst, D. I.; Kell, D. B. *Metabolomics* **2006**, *2*, 171-196.

(115) Ahola-Erkkila, S.; Carroll, C. J.; Peltola-Mjosund, K.; Tulkki, V.; Mattila, I.; Seppanen-Laakso, T.; Oresic, M.; Tyynismaa, H.; Suomalainen, A. *Hum Mol Genet* **2010**, *19*, 1974-1984.

(116) Tyynismaa, H.; Carroll, C. J.; Raimundo, N.; Ahola-Erkkila, S.; Wenz, T.; Ruhanen, H.; Guse, K.; Hemminki, A.; Peltola-Mjosund, K. E.; Tulkki, V.; Oresic, M.; Moraes, C. T.; Pietilainen, K.; Hovatta, I.; Suomalainen, A. *Hum Mol Genet* **2010**, *19*, 3948-3958.

(117) Asiago, V. M.; Alvarado, L. Z.; Shanaiah, N.; Gowda, G. A.; Owusu-Sarfo, K.; Ballas, R. A.; Raftery, D. *Cancer Res* **2010**, *70*, 8309-8318.

(118) Lankinen, M.; Schwab, U.; Seppanen-Laakso, T.; Mattila, I.; Juntunen, K.; Mykkanen, H.; Poutanen, K.; Gylling, H.; Oresic, M. *J Nutr* **2010**, *141*, 31-36.

(119) Beckstrom, A. C.; Humston, E. M.; Snyder, L. R.; Synovec, R. E.; Juul, S. E. *J Chromatogr A* **2011**, *1218*, 1899-1906.

(120) Li, X.; Xu, Z.; Lu, X.; Yang, X.; Yin, P.; Kong, H.; Yu, Y.; Xu, G. *Anal Chim Acta* **2009**, *633*, 257-262.

(121) Oresic, M.; Simell, S.; Sysi-Aho, M.; Nanto-Salonen, K.; Seppanen-Laakso, T.; Parikka, V.; Katajamaa, M.; Hekkala, A.; Mattila, I.; Keskinen, P.; Yetukuri, L.; Reinikainen, A.; Lahde, J.; Suortti, T.; Hakalax, J.; Simell, T.; Hyoty, H.; Veijola, R.; Ilonen, J.; Lahesmaa, R.; Knip, M.; Simell, O. *J Exp Med* **2008**, *205*, 2975-2984.

(122) Oresic, M.; Tang, J.; Seppanen-Laakso, T.; Mattila, I.; Saarni, S. E.; Saarni, S. I.; Lonnqvist, J.; Sysi-Aho, M.; Hyotylainen, T.; Perala, J.; Suvisaari, J. *Genome Med* **2011**, *3*, 19.

(123) Velagapudi, V. R.; Hezaveh, R.; Reigstad, C. S.; Gopalacharyulu, P.; Yetukuri, L.; Islam, S.; Felin, J.; Perkins, R.; Boren, J.; Oresic, M.; Backhed, F. *J Lipid Res* **2010**, *51*, 1101-1112.

(124) Tranchida, P. Q.; Costa, R.; Donato, P.; Sciarrone, D.; Ragonese, C.; Dugo, P.; Dugo, G.; Mondello, L. *J Sep Sci* **2008**, *31*, 3347-3351.

(125) McGaw, E. A.; Phinney, K. W.; Lowenthal, M. S. *J Chromatogr A* **2010**, *1217*, 5822-5831.

(126) Waldhier, M. C.; Almstetter, M. F.; Nurnberger, N.; Gruber, M. A.; Dettmer, K.; Oefner, P. J. *J Chromatogr A* **2011**, *1218*, 4537-4544.

(127) Huang, X.; Regnier, F. E. *Anal Chem* **2008**, *80*, 107-114.

(128) Schummer, C.; Delhomme, O.; Appenzeller, B. M.; Wennig, R.; Millet, M. *Talanta* **2009**, *77*, 1473-1482.

(129) O'Hagan, S.; Dunn, W. B.; Knowles, J. D.; Broadhurst, D.; Williams, R.; Ashworth, J. J.; Cameron, M.; Kell, D. B. *Anal Chem* **2007**, *79*, 464-476.

(130) Gröger, T.; Zimmermann, R. *Talanta* **2011**, *83*, 1289-1294.

(131) Koek, M. M.; Muilwijk, B.; van Stee, L. L.; Hankemeier, T. *J Chromatogr A* **2008**, *1186*, 420-429.

(132) Wojtowicz, P.; Zrostlikova, J.; Kovalczuk, T.; Schurek, J.; Adam, T. *J Chromatogr A* **2010**, *1217*, 8054-8061.

(133) Kouremenos, K. A.; Pitt, J.; Marriott, P. J. *J Chromatogr A* **2010**, *1217*, 104-111.

(134) Kouremenos, K. A.; Harynuk, J. J.; Winniford, W. L.; Morrison, P. D.; Marriott, P. J. *J Chromatogr B Analyt Technol Biomed Life Sci* **2010**, *878*, 1761-1770.

(135) Pasikanti, K. K.; Norasmara, J.; Cai, S.; Mahendran, R.; Esuvaranathan, K.; Ho, P. C.; Chan, E. C. *Anal Bioanal Chem* **2010**, *398*, 1285-1293.

(136) Kajander, K.; Myllyluoma, E.; Kyronpalo, S.; Rasmussen, M.; Sipponen, P.; Mattila, I.; Seppanen-Laakso, T.; Vapaatalo, H.; Oresic, M.; Korpela, R. *World J Gastroenterol* **2009**, *15*, 6068-6074.

(137) Kleemann, R.; van Erk, M.; Verschuren, L.; van den Hoek, A. M.; Koek, M.; Wielinga, P. Y.; Jie, A.; Pellis, L.; Bobeldijk-Pastorova, I.; Kelder, T.; Toet, K.; Wopereis, S.; Cnubben, N.; Evelo, C.; van Ommen, B.; Kooistra, T. *PLoS One* **2010**, *5*, e8817.

(138) Mervaala, E.; Biala, A.; Merasto, S.; Lempiainen, J.; Mattila, I.; Martonen, E.; Eriksson, O.; Louhelainen, M.; Finckenberg, P.; Kaheinen, P.; Muller, D. N.; Luft, F. C.; Lapatto, R.; Oresic, M. *Hypertension* **2010**, *55*, 508-515.

(139) Snyder, L. R.; Hoggard, J. C.; Montine, T. J.; Synovec, R. E. *J Chromatogr A* **2010**, *1217*, 4639-4647.

(140) Guo, X.; Lidstrom, M. E. *Biotechnol Bioeng* **2008**, *99*, 929-940.

(141) Yang, S.; Sadilek, M.; Synovec, R. E.; Lidstrom, M. E. *J Chromatogr A* **2009**, *1216*, 3280-3289.

(142) Okubo, Y.; Yang, S.; Chistoserdova, L.; Lidstrom, M. E. *J Bacteriol* **2010**, *192*, 1813-1823.

(143) Humston, E. M.; Dombek, K. M.; Hoggard, J. C.; Young, E. T.; Synovec, R. E. *Anal Chem* **2008**, *80*, 8002-8011.

(144) Humston, E. M.; Dombek, K. M.; Tu, B. P.; Young, E. T.; Synovec, R. E. *Anal Bioanal Chem* **2011**, *401*, 2387-2402.

(145) Cooper, S. J.; Finney, G. L.; Brown, S. L.; Nelson, S. K.; Hesselberth, J.; MacCoss, M. J.; Fields, S. *Genome Res* **2010**, *20*, 1288-1296.

(146) Hegeman, A. D. *Brief Funct Genomics* **2010**, *9*, 139-148.

(147) Hope, J. L.; Prazen, B. J.; Nilsson, E. J.; Lidstrom, M. E.; Synovec, R. E. *Talanta* **2005**, *65*, 380-388.

(148) Kusano, M.; Fukushima, A.; Kobayashi, M.; Hayashi, N.; Jonsson, P.; Moritz, T.; Ebana, K.; Saito, K. *J Chromatogr B Analyt Technol Biomed Life Sci* **2007**, *855*, 71-79.

(149) Merchant, S. S.; Prochnik, S. E.; Vallon, O.; Harris, E. H.; Karpowicz, S. J.; Witman, G. B.; Terry, A.; Salamov, A.; Fritz-Laylin, L. K.; Marechal-Drouard, L.; Marshall, W. F.; Qu, L. H.; Nelson, D. R.; Sanderfoot, A. A.; Spalding, M. H.; Kapitonov, V. V.; Ren, Q.; Ferris, P.; Lindquist, E.; Shapiro, H.; Lucas, S. M.; Grimwood, J.; Schmutz, J.; Cardol, P.; Cerutti, H.; Chanfreau, G.; Chen, C. L.; Cognat, V.; Croft, M. T.; Dent, R.; Dutcher, S.; Fernandez, E.; Fukuzawa, H.; Gonzalez-Ballester, D.; Gonzalez-Halphen, D.; Hallmann, A.; Hanikenne, M.; Hippler, M.; Inwood, W.; Jabbari, K.; Kalanon, M.; Kuras, R.; Lefebvre, P. A.; Lemaire, S. D.; Lobanov, A. V.; Lohr, M.; Manuell, A.; Meier, I.; Mets, L.; Mittag, M.; Mittelmeier, T.; Moroney, J. V.; Moseley, J.; Napoli, C.; Nedelcu, A. M.; Niyogi, K.; Novoselov, S. V.; Paulsen, I. T.; Pazour, G.; Purton, S.; Ral, J. P.; Riano-Pachon, D. M.; Riekhof, W.; Rymarquis, L.; Schroda, M.; Stern, D.; Umen, J.; Willows, R.; Wilson, N.; Zimmer, S. L.; Allmer, J.; Balk, J.; Bisova, K.; Chen, C. J.; Elias, M.; Gendler, K.; Hauser, C.; Lamb, M. R.; Ledford, H.;

Long, J. C.; Minagawa, J.; Page, M. D.; Pan, J.; Pootakham, W.; Roje, S.; Rose, A.; Stahlberg, E.; Terauchi, A. M.; Yang, P.; Ball, S.; Bowler, C.; Dieckmann, C. L.; Gladyshev, V. N.; Green, P.; Jorgensen, R.; Mayfield, S.; Mueller-Roeber, B.; Rajamani, S.; Sayre, R. T.; Brokstein, P.; Dubchak, I.; Goodstein, D.; Hornick, L.; Huang, Y. W.; Jhaveri, J.; Luo, Y.; Martinez, D.; Ngau, W. C.; Otillar, B.; Poliakov, A.; Porter, A.; Szajkowski, L.; Werner, G.; Zhou, K.; Grigoriev, I. V.; Rokhsar, D. S.; Grossman, A. R. *Science* **2007**, *318*, 245-250.

(150) May, P.; Wienkoop, S.; Kempa, S.; Usadel, B.; Christian, N.; Rupprecht, J.; Weiss, J.; Recuenco-Munoz, L.; Ebenhoh, O.; Weckwerth, W.; Walther, D. *Genetics* **2008**, *179*, 157-166.

(151) Kempa, S.; Hummel, J.; Schwemmer, T.; Pietzke, M.; Strehmel, N.; Wienkoop, S.; Kopka, J.; Weckwerth, W. *J Basic Microbiol* **2009**, *49*, 82-91.

(152) Doebbe, A.; Keck, M.; La Russa, M.; Mussgnug, J. H.; Hankamer, B.; Tekce, E.; Niehaus, K.; Kruse, O. *J Biol Chem* **2010**, *285*, 30247-30260.

(153) Allwood, J. W.; Erban, A.; de Koning, S.; Dunn, W. B.; Luedemann, A.; Lommen, A.; Kay, L.; Loscher, R.; Kopka, J.; Goodacre, R. *Metabolomics* **2009**, *5*, 479-496.

(154) Johanningsmeier, S. D.; McFeeters, R. F. *J Food Sci* **2010**, *76*, C168-177.

(155) Hyöyläinen, T.; Kallio, M.; Lehtonen, M.; Lintonen, S.; Perajoki, P.; Jussila, M.; Riekkola, M. L. *J Sep Sci* **2004**, *27*, 459-467.

(156) Vlaeminck, B.; Harynuk, J.; Fievez, V.; Marriott, P. J. *Eur. J. Lipid Sci. Technol.* **2007**, *109*, 757-766.

(157) Mayadunne, R.; Nguyen, T. T.; Marriott, P. J. *Anal Bioanal Chem* **2005**, *382*, 836-847.

(158) Junge, M.; Huegel, H.; Marriott, P. J. *Chirality* **2007**, *19*, 228-234.

(159) Ralston-Hooper, K.; Hopf, A.; Oh, C.; Zhang, X.; Adamec, J.; Sepulveda, M. S. *Aquat Toxicol* **2008**, *88*, 48-52.

(160) Ralston-Hooper, K. J.; Adamec, J.; Jannash, A.; Mollenhauer, R.; Ochoa-Acuna, H.; Sepulveda, M. S. *J Appl Toxicol* **2010**, Epub ahead of print.

(161) Aura, A. M.; Mattila, I.; Hyotylainen, T.; Gopalacharyulu, P.; Bounsaythip, C.; Oresic, M.; Oksman-Caldentey, K. M. *Mol Biosyst* **2010**, *7*, 437-446.

(162) Kaspar, H.; Dettmer, K.; Gronwald, W.; Oefner, P. J. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008**, *870*, 222-232.

(163) FDA. *Guidance for industry bioanalytical method validation.*; Center for Drug Evaluation and Research, U.S. Department of Health and Human Services, Rockville, USA, 2001.

(164) Bland, J. M.; Altman, D. G. *Lancet* **1986**, *1*, 307-310.

(165) Timischl, B.; Dettmer, K.; Kaspar, H.; Thieme, M.; Oefner, P. J. *Electrophoresis* **2008**, *29*, 2203-2214.

(166) R-DevelopmentCoreTeam *R: A language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, 2007.

(167) Pollard, K. S.; Dudoit, S.; van der Laan, M. J. *MULTTEST Multiple testing procedures and applications to genomics*; Division of Biostatistics, University of California: Berkeley, USA, 2004.

(168) Tibshirani, R.; Hastie, T.; Narasimhan, B.; Chu, G. *Proc Natl Acad Sci U S A* **2002**, *99*, 6567-6572.

(169) Gika, H. G.; Macpherson, E.; Theodoridis, G. A.; Wilson, I. D. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008**, *871*, 299-305.

(170) Lemon, W. J.; Liyanarachchi, S.; You, M. *Genome Biol* **2003**, *4*, R67.

(171) Wachsmuth, C. J.; Almstetter, M. F.; Waldhier, M. C.; Gruber, M. A.; Nurnberger, N.; Oefner, P. J.; Dettmer, K. *Anal Chem* **2011**, *83*, 7514-7522.

(172) Villas-Bôas, S. G.; Smart, K. F.; Sivakumaran, S.; Lane, G. A. *Metabolites* **2011**, *I*, 3-20.

(173) Kaspar, H.; Dettmer, K.; Gronwald, W.; Oefner, P. J. *Anal Bioanal Chem* **2009**, *393*, 445-452.

(174) Kaspar, H.; Dettmer, K.; Chan, Q.; Daniels, S.; Nimkar, S.; Daviglus, M. L.; Stamler, J.; Elliott, P.; Oefner, P. J. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009**, *877*, 1838-1846.

(175) Brückner, H.; Hausch, M. *Chromatographia* **1989**, *28*, 487-492.

(176) Brückner, H.; Hausch, M. In *Amino Acids: Chemistry, Biology and Medicine*; Lubec, G., Rosenthal, G. A., Eds.; ESCOM Scientific: Leiden, The Netherlands, 1990, pp 1172-1182.

(177) Brückner, H.; Hausch, M. *Milchwissenschaft* **1990**, *45*, 357-360.

(178) Friedman, M. *J Agric Food Chem* **1999**, *47*, 3457-3479.

(179) Friedman, M. *Chem Biodivers* **2010**, *7*, 1491-1530.

(180) Pätzold, R.; Schieber, A.; Brückner, H. *Biomed Chromatogr* **2005**, *19*, 466-473.

(181) Zahradnickova, H.; Husek, P.; Simek, P.; Hartvich, P.; Marsalek, B.; Holoubek, I. *Anal Bioanal Chem* **2007**, *388*, 1815-1822.

(182) Ali, H. S.; Patzold, R.; Bruckner, H. *Amino Acids* **2010**, *38*, 951-958.

(183) D'Aniello, A.; D'Onofrio, G.; Pischetola, M.; D'Aniello, G.; Vetere, A.; Petrucelli, L.; Fisher, G. H. *J Biol Chem* **1993**, *268*, 26941-26949.

(184) Pilone, M. S. *Cell Mol Life Sci* **2000**, *57*, 1732-1747.

(185) Waldhier, M. C.; Gruber, M. A.; Dettmer, K.; Oefner, P. J. *Anal Bioanal Chem* **2009**, *394*, 695-706.

(186) Cipollina, C.; ten Pierick, A.; Canelas, A. B.; Seifar, R. M.; van Maris, A. J.; van Dam, J. C.; Heijnen, J. J. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009**, *877*, 3231-3236.

(187) Brückner, H.; Schieber, A. *Biomed Chromatogr* **2001**, *15*, 166-172.

(188) Armstrong, D. W.; Gasper, M.; Lee, S. H.; Zukowski, J.; Ercal, N. *Chirality* **1993**, *5*, 375-378.

(189) Guder, W. G.; Narayanan, S.; Wisser, H.; Zawta, B. *Diagnostic samples: from the patient to the laboratory - The impact of preanalytical variables on the quality of laboratory results*, 4th ed.; Wiley-VCH: Weinheim, Germany, 2009.

(190) Parvy, P.; Bardet, J.; Rabier, D.; Bonnefont, J. P.; Kamoun, P. *Clin Chem* **1989**, *35*, 178.

(191) Waring, W. S.; Evans, L. E.; Kirkpatrick, C. T. *J Clin Pathol* **2007**, *60*, 820-823.

(192) Mikesh, L. M.; Bruns, D. E. *Clin Chem* **2008**, *54*, 930-932.

(193) Deprez, S.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N.; Waterfield, C. J. *J. Pharm. Biomed. Anal.* **2002**, *30*, 1297-1310.

(194) Teahan, O.; Gamble, S.; Holmes, E.; Waxman, J.; Nicholson, J. K.; Bevan, C.; Keun, H. C. *Anal. Chem.* **2006**, *78*, 4307-4318.

(195) Drake, S. K.; Bowen, R. A. R.; Remaley, A. T.; Hortin, G. L. *Clin. Chem.* **2004**, *50*, 2398-2401.

(196) Mei, H.; Hsieh, Y.; Nardo, C.; Xu, X.; Wang, S.; Ng, K.; Korfmacher, W. A. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 97-103.

(197) Taylor, P. J.; Cooper, D. P.; Gordon, R. D.; Stowasser, M. *Clin. Chem.* **2009**, *55*, 1155-1162.

(198) Dettmer, K.; Almstetter, M. F.; Appel, I. J.; Nurnberger, N.; Schlamberger, G.; Gronwald, W.; Meyer, H. H.; Oefner, P. J. *Electrophoresis* **2010**, *31*, 2365-2373.

12 Appendix

Figure S1. Pseudocode for the peak alignment. Reproduced from [107].

```
Initialize threshold parameter t_1st, t_2nd, t_ol, t_rate
Aligned <- matrix()  # initiate emtpy alignment matrix with columns 'Name', 'rt1st_lower_bound',
'rt1st_upper_bound', 'rt2nd_lower_bound', 'rt2nd_upper_bound', 'spectrum S', areas for files,
'mean_rt1st', 'mean_rt2nd'
Peaks <- matrix(measurements)   # read raw data with columns 'Name', 'rt1st', 'rt2nd',
'spectrum S', 'area'
for each peak in Peaks:
  if Aligned is EMPTY: generate new entry using peak; next peak
  rt1st <- which(abs(Aligned$mean_rt1st - peak$rt1st) <= t_1st)
  if any rt1st:
    rt2nd <- which(abs(Aligned$mean_rt2nd[rt1st] - peak$rt2nd) <= t_2nd)
    if any rt2nd:
      for each S in Aligned$S[rt1st[rt2nd]]:
        is <- duplicated fragment ions of S$ions and peak$ions
        smallerS <- min(len(S), len(peak$S))
        if len(is) >= smallerS*t_ol:
          h <- abs(peak$S$h[is] - S$h[is])
          if max(h) <= t_rate: # align peak to Aligned entry
            update retention time boundaries and means
            update S as union set of fragment ions and renormalize heights
            remember area at specified file position
            next peak
  else: generate new entry using peak
```

Figure S2. Linear dependencies between expected and observed fold changes of the 20 different spike-in metabolites for both the universal *m/z* 73 and the metabolite-specific *m/z* fragment ion trace investigated over all pairwise fold changes. Reproduced from [107].

Figure S2 continued.

m/z 73 Malonate

y = 0.798x + 0.195
r = 0.934
mean RSD = 17.7

m/z unique Malonate

y = 1.074x + −0.063
r = 0.993
mean RSD = 6.7

m/z 73 Phenylacetate

y = 0.977x + 0.127
r = 0.63
mean RSD = 52.9

m/z unique Phenylacetate

y = 0.621x + 0.352
r = 0.933
mean RSD = 14.9

m/z 73 Nicotinate

y = 0.586x + 0.395
r = 0.857
mean RSD = 18.8

m/z unique Nicotinate

y = 0.364x + 0.603
r = 0.874
mean RSD = 13.8

m/z 73 Dimethylsuccinate

y = 0.521x + 0.475
r = 0.797
mean RSD = 22.1

m/z unique Dimethylsuccinate

y = 0.939x + 0.07
r = 0.949
mean RSD = 16.4

Figure S2 continued.

m/z 73 Decanoate

y = 0.521x + 0.453

r = 0.908
mean RSD = 14.9

m/z unique Decanoate

y = 0.756x + 0.23

r = 0.954
mean RSD = 14.2

m/z 73 Mandelate

y = 0.503x + 0.48

r = 0.843
mean RSD = 21.3

m/z unique Mandelate

y = 0.836x + 0.161

r = 0.936
mean RSD = 18

m/z 73 Adipate

y = 0.873x + 0.124

r = 0.955
mean RSD = 15.5

m/z unique Adipate

y = 1.015x + 0.005

r = 0.915
mean RSD = 19.3

m/z 73 Erythritol

y = 0.613x + 0.357

r = 0.959
mean RSD = 11.3

m/z unique Erythritol

y = 0.849x + 0.14

r = 0.973
mean RSD = 11.7

Figure S2 continued.

m/z 73 Phenyllactate — y = 0.259x + 0.715, r = 0.782, mean RSD = 15.3

m/z unique Phenyllactate — y = 0.38x + 0.585, r = 0.903, mean RSD = 12.2

m/z 73 Triethanolamine — y = 0.881x + 0.134, r = 0.907, mean RSD = 23.3

m/z unique Triethanolamine — y = 0.986x + 0.023, r = 0.958, mean RSD = 16.2

m/z 73 Dodecanoate — y = 0.623x + 0.363, r = 0.904, mean RSD = 17.4

m/z unique Dodecanoate — y = 0.793x + 0.194, r = 0.962, mean RSD = 13.1

m/z 73 Suberate — y = 0.982x + 0.022, r = 0.976, mean RSD = 13.3

m/z unique Suberate — y = 1.083x + −0.07, r = 0.984, mean RSD = 11

Figure S2 continued.

m/z 73 Xylitol

y = 0.721x + 0.254

r = 0.985
mean RSD = 7.8

observed fold change

expected fold change

m/z unique Xylitol

y = 0.946x + 0.053

r = 0.986
mean RSD = 9.6

observed fold change

expected fold change

m/z 73 Vanillate

y = 1.115x + −0.094

r = 0.972
mean RSD = 14.9

observed fold change

expected fold change

m/z unique Vanillate

y = 1.066x + −0.052

r = 0.977
mean RSD = 12.5

observed fold change

expected fold change

m/z 73 Mannitol

y = 0.885x + 0.11

r = 0.971
mean RSD = 12.7

observed fold change

expected fold change

m/z unique Mannitol

y = 0.952x + 0.051

r = 0.971
mean RSD = 13.3

observed fold change

expected fold change

m/z 73 Eicosanoate

y = 0.838x + 0.159

r = 0.928
mean RSD = 19.2

observed fold change

expected fold change

m/z unique Eicosanoate

y = 0.863x + 0.126

r = 0.983
mean RSD = 9.5

observed fold change

expected fold change

Figure S3. Linear dependencies over all pairwise fold changes between expected and observed fold changes of the 20 different spike-in metabolites for both the universal *m/z* 73 and the metabolite-specific unique *m/z* ion trace selected by Statistical Compare. Reproduced from [108].

Figure S3 continued.

m/z 73 Decanoate

m/z 117 Decanoate

m/z 73 Mandelate

m/z 179 Mandelate

m/z 73 Adipate

m/z 111 Adipate

m/z 73 Erythritol

m/z 217 Erythritol

Figure S3 continued.

m/z 73 Phenyllactate

y = 0.311x + 0.657

r = 0.861
mean RSD = 13

m/z 193 Phenyllactate

y = 0.344x + 0.618

r = 0.942
mean RSD = 7.3

m/z 73 Triethanolamine

y = 0.791x + 0.192

r = 0.976
mean RSD = 10

m/z 262 Triethanolamine

y = 0.995x + 0.007

r = 0.991
mean RSD = 7.3

m/z 73 Dodecanoate

y = 0.357x + 0.605

r = 0.939
mean RSD = 8

m/z 117 Dodecanoate

y = 0.665x + 0.306

r = 0.972
mean RSD = 9.7

m/z 73 Suberate

y = 0.883x + 0.107

r = 0.992
mean RSD = 6.1

m/z 187 Suberate

y = 0.927x + 0.068

r = 0.988
mean RSD = 8.1

Figure S3 continued.

m/z 73 Xylitol
y = 0.745x + 0.23
r = 0.99
mean RSD = 6

m/z 217 Xylitol
y = 1.001x + 0.003
r = 0.985
mean RSD = 9.5

m/z 73 Vannilate
y = 0.819x + 0.166
r = 0.977
mean RSD = 10.4

m/z 297 Vannilate
y = 0.969x + 0.029
r = 0.991
mean RSD = 7.4

m/z 73 Mannitol
y = 0.889x + 0.102
r = 0.989
mean RSD = 7.7

m/z 319 Mannitol
y = 0.842x + 0.144
r = 0.991
mean RSD = 6.3

m/z 73 Eicosanoate
y = 0.823x + 0.161
r = 0.979
mean RSD = 9.5

m/z 117 Eicosanoate
y = 0.964x + 0.041
r = 0.973
mean RSD = 12.8

Table S1. Overview of the GC×GC based metabolomics publications presented in chapter 4.4 and 4.5. Reproduced from [6].

Authors (Year) [Reference]	Sample matrix	Pretreat-ment	Column set [m×mm ID×µm film thickness]	Detection	Data handling	Quantification	Statistical analysis
Ahola-Erkkilä et al. (2010) [115]	Serum	M, S	Rtx®-5 (10×0.18×0.2) BPX-50 (1.5×0.1×0.1)	TOFMS	ChromaTOF	relative	t-test
Allwod et al. (2009) [153]	Plant	M, S	VF-5ms (30×0.25×1.0) VF-17ms (2×0.1×0.2)	TOFMS	ChromaTOF	-	-
Almstetter et al. (2009 & 2011) [107, 108]	Bacterium	M, S	Rxi®-5ms (30×0.25×0.25) Rtx®-1701 (2×0.1×0.1)	TOFMS	ChromaTOF/ INCA	relative, absolute	PCA, t-test, FDR, ROC, CVal
Asiago et al. (2010) [117]	Serum	S	DB-5 (30×0.25×0.25) DB-17 (1×0.1×0.1)	TOFMS	ChromaTOF	relative	PLS-DA, t-test, box plots, CVal, ROC
Aura et al. (2010) [161]	Feces	M, S	Rtx®-5 (10×0.18×0.20) BPX-50 (1.5×0.1×0.1)	TOFMS	ChromaTOF	relative	PLS-DA, ANOVA, FDR, heatmap
Beckstrom et al. (2011) [119]	Plasma	M, S	Rtx®-5ms (20×0.25×0.5) Rtx®-200ms (2×0.18× 0.2)	TOFMS	Fisher ratio/ PARAFAC	relative	t-test, PCA, heatmap
Castillo et al. (2011) [110]	Plasma	M, S	Rtx®-5 (10×0.18×0.2) BPX-50 (1.5×0.1×0.1)	TOFMS	ChromaTOF/ Guineu	relative	PCA
Cooper et al. (2010) [145]	Yeast	M, S	Rtx®-5ms (20×0.25×0.5) Rtx®-200ms (2×0.18×0.2)	TOFMS	ChromaTOF	relative	t-test, heatmap
David et al. (2008) [89]	Bacterium	Me	HP-5ms (30×0.25×0.25) BPX-70 (1×0.1× 0.2)	FID	GC Image		-
Doebbe et al. (2010) [152]	Plant	M, S	Rtx®-5ms (30×0.25×0.25) BPX-50 (2×0.1×0.1)	TOFMS	ChromaTOF	relative	Fisher ratio, t-test
Gröger and Zimmermann (2011) [130]	Plasma	M, S	SolGel-1ms (30×0.25×0.25) BPX-50 (2×0.1×0.1)	TOFMS	MatLab	relative	PLS-DA, t-test, CVal
Guo and Lidstrom (2008) [140]	Bacterium	M, S	Rtx®-5 (20×0.25×0.5) Rtx®-200 (2×0.18×0.2)	TOFMS	Fisher ratio/ PARAFAC	relative	-
Hoggard and Synovec (2008) [93]	Yeast	M, S	Rtx®-5ms (20×0.25×0.5) Rtx®-200ms (2×0.18×0.2)	TOFMS	PARAFAC	-	-
Hope et al. (2005) [147]	Standards, plant	S	DB-5 (10×0.18×0.18) DB-17 (2×0.1×0.1)	TOFMS	PARAFAC	-	-
Huang and Regnier (2008) [127]	Standards, serum	S	DB-5 (10×0.18×0.18) DB-17 (1×0.1× 0.1)	TOFMS	ChromaTOF	relative	-

Authors (Year) [Reference]	Sample matrix	Pretreat-ment	Column set [m×mm ID×µm film thickness]	Detection	Data handling	Quantification	Statistical analysis
Humston et al. (2008) [143]	Yeast	M, S	Rtx®-5ms (20×0.25×0.5) Rtx®-200ms (2×0.18×0.2)	TOFMS	S-ratio/ PARAFAC	relative	PCA
Humston et al. (2011) [144]	Yeast	M, S	Rtx®-5ms (20×0.25×0.5) Rtx®-200ms (2×0.18×0.2)	TOFMS	S-ratio/ PARAFAC	relative	PCA, ANOVA, t-test, heatmap
Hyötyläinen et al. (2004) [155]	Milk	Me	Set 1/2: CP-7420 (100×0.25×0.25) HP-5ms (1.5×0.1×0.1) or HP-1 (1×0.1×0.1) Set 3/4: HP-1 (30×0.25×0.25) Cyano (1×0.1×0.1) or DB-WAX (0.35×0.05×0.1)	FID	MatLab	relative	-
Johanningsmeier and McFeeters (2010) [154]	Plant	SPME	SolGel-WAX (30×0.25×0.25) Rtx®-1701 (1×0.1×0.1)	TOFMS	ChromaTOF	relative	ANOVA, FDR
Junge et al. (2007) [158]	Standards, beer	MCF, ECF, TFA-ME	Chirasil-L-Val (25×0.25×0.16) BPX-50 (1 or 3×0.1×0.1) or BP-1 (1 or 3×0.1×0.1)	FID	-	relative	-
Kajander et al. (2009) [136]	Mucosa	M, S	Rtx®-5ms (10×0.18×0.2) BPX-50 (1.1×0.1×0.1)	TOFMS	ChromaTOF	relative	PLS-DA, Wilcoxon test, FDR, CVal
Kempa et al. (2009) [151]	Plant	M, S	VF-5ms (30×0.25×0.25) VF-17ms (1×0.1×10)	TOFMS	ChromaTOF/ MetMax	-	PCA, ICA
Kleemann et al. (2010) [137]	Liver	E, S	BPX-50 (30×0.25×0.25) BPX-5 (2×0.32×0.25)	TOFMS	ChromaTOF	relative	ANOVA
Koek et al. (2008) [131]	Standards, bovine serum	E, S	BPX-50 (30×0.25×0.25) BPX-5 (1×0.1×0.1) (2×0.25×0.25) (2×0.32×0.25) (2×0.32×0.5)	TOFMS	Chemstation/ ChromaTOF	relative	-
Koek et al. (2010) [90]	Mouse liver	E, S	BPX-50 (30×0.25×0.25) BPX-5 (2×0.32×0.25)	TOFMS	ChromaTOF	relative	PCA, PCDA, box plots, ANOVA
Kouremenos et al. (2010a) [134]	Standards	M, S	BPX-5 (30×0.25×0.25) BPX-50 (1×0.1×0.1)	TOFMS	ChromaTOF	relative	-

Authors (Year) [Reference]	Sample matrix	Pretreat-ment	Column set [m×mm ID×µm film thickness]	Detection	Data handling	Quantification	Statistical analysis
Kouremenos et al. (2010b) [133]	Urine	S	BPX-5 (30×0.25×0.25) BPX-50 (1×0.1×0.1) or BPX-50 (30×0.25×0.25) BPX-5 (1×0.1×0.1)	TOFMS	ChromaTOF	relative	-
Kusano et al. (2007) [148]	Plant	M, S	Rtx®-5Sil MS (30×0.25×0.25) Rtx®-50 (1×0.18×0.2)	TOFMS	ChromaTOF	relative	t-test
Lankinen et al. (2011) [118]	Plasma	M, S	Rtx®-5 (10×0.18×0.2) BPX-50 (1.5×0.1×0.1)	TOFMS	ChromaTOF	relative	t-test, FDR, Spear-man correlation
Li et al. (2009) [120]	Plasma	S	DB-5 (30×0.25×0.25) DB-1701 (1.4×0.1×0.1)	TOFMS	GC×GC workstation	relative	OSC-PLS-DA, t-test, CVal
May et al. (2008) [150]	Plant	M, S	VF-5ms (30×0.25×0.25) VF-17ms (1×0.1×10)	TOFMS	ChromaTOF	-	-
Mayadunne et al. (2005) [157]	Wine, beer, honey	PCF	Set 1:BPX-5 (30×0.25×0.25) BPX-50 (2×0.1×0.1) Set 2: SolGel-WAX (30×0.25×0.25) BP-1 (1.5×0.1×0.1)	FID, TOFMS	-	absolute	-
McGaw et al. (2010) [125]	Plasma	S, PCF	Rtx®-5 (38×0.18×0.2) Rxi®-17 (1×0.1×0.1)	TOFMS	ChromaTOF	relative	-
Mervaala et al. (2010) [138]	Heart	M, S	Rtx®-5 (10×0.18×0.2) BPX-50 (1.5×0.1×0.1)	TOFMS	ChromaTOF	relative	PLS-DA, ANOVA, Wilcoxon test, FDR
Mohler et al. (2006) [96, 98, 99]	Yeast	M, S	Rtx®-5ms (20×0.25×0.5) Rtx®-200ms (2×0.18×0.2)	TOFMS	PCA/ PARAFAC	relative	-
Mohler et al. (2007)	Yeast	M, S	Rtx®-5ms (20×0.25×0.5) Rtx®-200ms (2×0.18×0.2)	TOFMS	Fisher ratio/ PARAFAC	relative	t-test
Mohler et al. (2008)	Yeast	M, S	Rtx®-5ms (20×0.25×0.5) Rtx®-200ms (2×0.18×0.2)	TOFMS	S-ratio/ PARAFAC	relative	PCA
Oh et al. (2008) [106]	Standards, serum	E, S	DB-5 (10×0.18×0.18) DB-17 (1×0.1×0.1)	TOFMS	ChromaTOF/ MSort	-	ROC
O´Hagan et al. (2007) [129]	Serum	M, S	DB-1 (30×0.25×0.25) BPX-50 (1.5×0.1×0.1)	TOFMS	ChromaTOF	-	PCA

Table S1 continued.

Authors (Year) [Reference]	Sample matrix	Pretreat-ment	Column set [m×mm ID×µm film thickness]	Detection	Data handling	Quantification	Statistical analysis
Okubo et al. (2010) [142]	Bacterium	M, S	Rtx®-5 (20×0.25×0.5) Rtx®-200 (2×0.18×0.2)	TOFMS	S-ratio/ PARAFAC	relative	t-test
Oresic et al. (2008) [121]	Serum	M, S	Rtx®-5 (10×0.18×0.2) BPX-50 (1.5×0.1×0.1)	TOFMS	ChromaTOF	relative	Wilcoxon test, FDR, heatmap
Oresic et al. (2011) [122]	Serum	M, S	Rtx®-5 (10×0.18×0.2) BPX-50 (1.5×0.1×0.1)	TOFMS	ChromaTOF	relative	ANOVA, beanplot, CVal, ROC
Pasikanti et al. (2010) [135]	Cells	M, S	DB-1 (30×0.25×0.25) Rxi®-17 (1.5×0.1×0.1)	TOFMS	ChromaTOF	relative	PCA, PLS-DA, Welch t-test
Pierce et al. (2006) [95]	Plant	M, S	DB-5 (20×0.25×0.5) Rtx®-200 (2×0.18×0.2)	TOFMS	PCA/ PARAFAC	relative	-
Pierce et al. (2006) [97]	Urine	M, S	DB-1 (20×0.25×0.5) or Rtx®-5ms (20×0.25×0.5) Rtx®-200 (2×0.18×0.2)	TOFMS	Fisher ratio/ PARAFAC	relative	-
Ralston-Hooper et al. (2008) [159]	Inverte-brates	E, S	Rtx®-5 (10×0.18×0.2) DB-17 (1.1×0.18×0.18)	TOFMS	ChromaTOF/ MSort	relative	t-test, FDR, PCA
Ralston-Hooper et al. (2010) [160]	Inverte-brates	M, S	Rtx®-5 (10×0.18×0.2) DB-17 (1.1×0.18×0.18)	TOFMS	ChromaTOF/ MSort	relative	t-test, ANOVA, FDR, HCA
Shellie et al. (2005) [101]	Tissue	M, S	Set 1: PDMS (30×0.25×0.25) 50% Phenyl polysilphenylene siloxane (1.5×0.1×0.1) Set 2: 5% Phenyl PDMS (30×0.25×0.25) 50% Phenyl polysilphenylene siloxane (2×0.1×0.1)	TOFMS	ChromaTOF	relative	t-test
Sinha et al. (2004) [94]	Infant urine	O, S	DB-5 (20×0.25×0.5) Rtx®-200 (2×0.18×0.2)	TOFMS	DotMap/ PARAFAC	-	-
Sinha et al. (2004) [91]	Plant	O, S	DB-5 (10×0.18×0.18) DB-17 (2×0.1×0.1)	TOFMS	PARAFAC	-	-
Snyder et al. (2010) [139]	Brain tissue	M, S	Rtx®-5 (20×0.25×0.5) Rtx®-200 (2×0.18×0.2)	TOFMS	PARAFAC	relative	-
Tranchida et al. (2008) [124]	Plasma	Me	Equity-1 (30×0.25×0.25) Supelcowax™ 10 (0.95×0.1×0.1)	FID	-	-	-

Table S1 continued.

Authors (Year) [Reference]	Sample matrix	Pretreatment	Column set [m×mm ID×µm film thickness]	Detection	Data handling	Quantification	Statistical analysis
Tyynismaa et al. (2010) [116]	Muscle	M, S	Rtx®-5 (10×0.18×0.2) BPX-50 (1.5×0.1×0.1)	TOFMS	ChromaTOF	relative	t-test
Velagapudi et al. (2009) [123]	Serum	M, S	Rtx®-5 (10×0.18×0.2) BPX-50 (1.5×0.1×0.1)	TOFMS	ChromaTOF	relative	PLS-DA, t-test, ANOVA, FDR
Vlaeminck et al. (2007) [156]	Milk	Me	BPX-5 (30×0.25×0.25) BPX-20 (0.85×0.1×0.2)	FID	Chemstation	relative	t-test
Wachsmuth et al. (2011) [171]	Standards, bacterium	M, S	Rxi®-5ms (30×0.25×0.25) Rtx®-1701 (2×0.1×0.1)	TOFMS	ChromaTOF	relative	
Waldhier et al. (2011) [126]	Serum, urine	MCF	Rt®-γDEXsa (30×0.25×0.25) Rtx®-1701 (2×0.1×0.1) or ZB-AAA (2×0.25×0.1)	TOFMS	ChromaTOF	absolute	t-test
Wang et al. (2010) [109]	Standards, rat plasma	E, S	Rxi®-5ms (30×0.25×0.25) BPX-50 (1.2×0.1×0.1)	TOFMS	ChromaTOF/ DISCO	-	ROC
Welthagen et al. (2005) [100]	Spleen tissue	M, S	Rtx®-1ms (30×0.25×0.25) BPX-50 (1.5×0.1×0.1)	TOFMS	ChromaTOF	relative	t-test
Wojtowicz et al. (2010) [132]	Urine	S	Rxi®-5ms (30×0.25×0.25) BPX-50 (2.5×0.1×0.1)	TOFMS	ChromaTOF	relative	-
Yang et al. (2009) [141]	Bacterium	M, S	Rtx®-5 (20×0.25×0.5) Rtx®-200 (2×0.18×0.2)	TOFMS	Fisher ratio/ PARAFAC	absolute, relative	t-test, Fisher ratio
Zhang et al. (2008) [105]	Standards, serum	E, S	DB-5 (10×0.18×0.18) DB-17 (1×0.1×0.1)	TOFMS	ChromaTOF	-	-

M methoxymation, E ethoxymation, O oxymation, S silylation, Me methylation, TFA-ME *N*-trifluoroacetyl methyl ester derivatization.

Table S2. Our in-house EI mass spectral library consists of approx. 150 metabolites including sugars, alcohols, fatty acids, organic acids, amino acids, and amino acid metabolites. Some metabolites yield multiple derivatization products.

No.	Metabolite
1	N-Acetyl glucosamine, MeOx, 4TMS
2	cis-Aconitic acid, 3TMS
3	Adenosine, 4TMS
4	Adipic acid, 2TMS
5	Alanine, N-TMS, TMS
6	α-Aminoadipic acid, N-TMS, 2TMS
7	γ-Aminobutyric acid, 2TMS
8	γ-Aminobutyric acid, 3TMS
9	Aminoisobutyric acid, 2TMS
10	Aminoisobutyric acid, 3TMS
11	α-Aminopimelic acid, N-TMS, 2TMS
12	Anthranilic acid, N-TMS, TMS
13	Arachidonic acid, TMS
14	Ascorbic acid, 4TMS
15	Asparagine, N-2TMS, N-TMS, TMS
16	Asparagine, N,N-TMS, 2TMS
17	Aspartic acid, 2TMS
18	Aspartic acid, N-TMS, 2TMS
19	Benzoic acid, TMS
20	Cholesterol, TMS
21	Citric acid, 4TMS
22	Creatinine, N,N-2TMS, TMS
23	Cystathionine, N,N-2TMS, 2TMS
24	Cysteine, N,S-2TMS, TMS
25	Cystine, N,N-2TMS, 2TMS
26	Decanoic acid, TMS
27	Dimethylsuccinic acid, 2TMS
28	Docosahexaenoic acid, TMS
29	Docosanoic acid, TMS
30	Dodecanoic acid, TMS
31	Eicosanoic acid, TMS
32	Erucic acid, TMS
33	Erythritol, 4TMS
34	Ethanolamine, 3TMS
35	Fructose, MeOx, 5TMS
36	Fructose-6-phosphate, MeOx, 6TMS
37	Fumaric acid, 2TMS
38	Galactose, MeOx, 6TMS
39	Glucose, MeOx, 5TMS
40	Glucose-6-phosphate, MeOx, 6TMS
41	Glutamic acid, N-TMS, 2TMS
42	Glutamine, N-TMS, 2TMS
43	Glutaric acid, 2TMS
44	Glyceric acid, 3TMS
45	Glycerol, 3TMS
46	Glycerol-1-phosphate, 4TMS
47	Glycerol-2-phosphate, 4TMS
48	Glycine, N-2TMS, TMS
49	Glycine, N-TMS, TMS
50	Heptadecanoic acid, TMS
51	Heptanoic acid, TMS
52	Hexadecanoic acid, TMS
53	Hexanoic acid, TMS
54	Hippuric acid, N-TMS, TMS
55	Hippuric acid, TMS
56	Histamine, N,1-2TMS
57	Histidine, N,1-2TMS, TMS
58	Homogentisic acid, TMS
59	Homovanillic acid, 2TMS
60	3-Hydroxyanthranilic acid, N-TMS, 2TMS
61	Hydroxybenzyl alcohol, 2TMS
62	2-Hydroxybutyric acid, 2TMS
63	3-Hydroxybutyric acid, TMS
64	5-Hydroxyindoleacetic acid, 3TMS
65	3-Hydroxykynurenine, N,N-2TMS, 2TMS
66	2-Hydroxy-3-methylbutyric acid, 2TMS
67	Hydroxyphenylacetic acid, TMS
68	Hydroxyphenylpyruvic acid, TMS
69	Hydroxyproline, 2TMS
70	Hydroxyproline, 3TMS
71	Indole-3-pyruvate, MeOx, TMS
72	Indole, 1-TMS
73	Indole-1-acetic acid, TMS
74	Indole-3-acetic acid, 1-TMS, TMS
75	Indole-3-lactic acid, 1-TMS, 2TMS
76	Indole-3-propionic acid, 1-TMS, TMS
77	Isoleucine, N-TMS, TMS
78	Isoleucine, TMS
79	Itaconic acid, 2TMS
80	2-Ketobutyric acid, TMS
81	α-Ketoglutaric acid, MeOx-TMS
82	Ketosuccinic acid, 3TMS
83	Kynurenic acid, 2TMS
84	Kynurenine, N-TMS, TMS
85	Kynurenine, N,N-2TMS, TMS
86	Lactic acid, 2TMS
87	Lactose, 7TMS
88	Leucine, N-TMS, TMS
89	Leucine, TMS
90	Linoleic acid, TMS
91	α-Linolenic acid, TMS
92	Lysine, 3TMS
93	Lysine, N-TMS, N-2TMS, TMS
94	Lysine, N,N-2TMS, TMS
95	Malic acid, 3TMS
96	Malonic acid, 2TMS
97	Maltose, 8TMS
98	Mandelic acid, 2TMS
99	Mannitol, 6TMS
100	Mannose, MeOx, 5TMS
101	Methionine, N-TMS, TMS
102	Methionine, TMS
103	Methylhistidine, N-TMS, TMS
104	Methylmalonic acid, 2TMS
105	3-Methyl-2-oxovaleric acid, TMS
106	4-Methyl-2-oxovaleric acid, TMS
107	Myo-Inositol-phosphate, 7TMS

Table S2 continued.

No.	Metabolite
108	Myo-Inositol, 6TMS
109	Nicotinamide, TMS
110	Nicotinic acid, TMS
111	Nonadecanoic acid, TMS
112	Nonanoic acid, TMS
113	Norvaline, N-TMS, TMS
114	Octadecanoic acid, TMS
115	Octanoic acid, TMS
116	Oleic acid, TMS
117	Ornithine, N,N-2TMS, TMS
118	Ornithine, N-TMS, N-2TMS, TMS
119	Orotic acid, 3TMS
120	Oxalacetate, MeOx, 2TMS
121	Oxalic acid, 2TMS
122	Palmitelaidic acid, TMS
123	Pantothenic acid, 3TMS
124	Pentadecanoic acid, TMS
125	Pentanoic acid, TMS
126	Phenylacetic acid, TMS
127	Phenylalanine, N-TMS, TMS
128	Phenylalanine, TMS
129	Phenyllactic acid, 2TMS
130	Phenylpyruvic acid, TMS
131	Phosphoenolpyruvic acid, MeOx, 3TMS
132	3-Phosphoglyceric acid, TMS
133	O-Phosphorylethanolamine, 4TMS
134	Proline, 1-TMS, TMS
135	Proline, TMS
136	Putrescine, N-2TMS, N-2TMS
137	Pyroglutamate, 1-TMS, TMS
138	Pyruvate, MeOx, TMS
138	Quinolinic acid, 2TMS
139	Ribitol, 5TMS
140	Ribose, MeOx, 4TMS
141	Salicylic acid, TMS, acetate
142	Sarcosine, N-TMS, TMS
143	Serine, 2TMS
144	Serine, N-TMS, 2TMS
145	Serotonine, N,N-2TMS, TMS
146	Serotonine, N-2TMS, N-TMS, TMS
147	Shikimic acid, 4TMS
148	Sorbitol, 6TMS
149	Suberic acid, TMS
150	Succinic acid, 2TMS
151	Sucrose, 8TMS
152	Tartaric acid, 4TMS
153	Taurine, N-2TMS, TMS
154	Taurine, N-TMS, TMS
155	Tetradecanoic acid, TMS
156	Threonine, 2TMS
157	Threonine, N-TMS, 2TMS
158	Tridecanoic acid, TMS
159	Triethanolamine, 3TMS
160	Tryptophan, N,1-2TMS, TMS
161	Tryptophan, N-TMS, TMS
162	Tyrosine, 2TMS
163	Tyrosine, N-TMS, 2TMS
164	Undecanoic acid, TMS
165	Valine, N-TMS, TMS
166	Valine, TMS
167	Vanillic acid, 2TMS
168	Xanthurenic acid, 3TMS
169	Xylitol, 5TMS
170	Xylose, MeOx, 4TMS

Table S3. Comparison of metabolites identified in extracts of the two E. coli strains grown in LB medium and harvested in stationary phase using CE-TOFMS and GC×GC-TOFMS analysis. Reproduced from [107].

No.	Metabolite	CE-TOFMS		GC×GC-TOFMS	
		Detected	Significant (p <0.05)	Detected	Significant (FDR <0.05)
1	Aminoadipate	x	x	x	x
2	Caproate	x	x	x	x
3	cis-Aconitate	x	x	x	x
4	Citrate	x	x	x	x
5	Fumarate	x	x	x	x
6	Glucose	x	x	x	x
7	Glucose-6-phosphate	x	x	x	x
8	Isoleucine	x	x	x	x
9	Itaconate	x	x	x	x
10	Leucine	x	x	x	x
11	Lysine	x	x	x	x
12	Malate	x	x	x	x
13	Orotate	x	x	x	x
14	Proline	x	x	x	x
15	Pyruvate	x	x	x	x
16	Succinate	x	x	x	x
17	Sucrose	x	x	x	x
18	Valine	x	x	x	x
19	Alanine	x	x	x	
20	Aspartate	x	x	x	
21	Benzoate	x	x	x	
22	Glucosamine-6-phosphate	x	x	x	
23	Glucuronate	x	x	x	
24	Glutamate	x	x	x*	
25	Glycerolphosphate	x	x	x	
26	Octanoate	x	x	x	
27	Pantothenate	x	x	x	
28	Phosphoenolpyruvate	x	x	x	
29	Phosphoglycerate	x	x	x	
30	AMP	x	x		
31	CMP	x	x		
32	dTMP	x	x		
33	Glutathione	x	x		
34	Maltose	x	x		
35	UDP	x	x		
36	UDP-glucose	x	x		
37	UMP	x	x		
38	alpha-Ketogluterate	x		x	x
39	Tartrate	x		x	x
40	Furoate	x[1]	x	x	x
41	GABA	[2]	x	x	x
42	Glycine			x	x
43	Myo-inositol			x	x
44	Indole			x	x
45	Putrescine			x	x
46	Pyroglutamate			x[1]*	x
47	Caprate	x		x	
48	Glutarate	x		x	
49	Heptanoic acid	x		x	
50	Histidine	x		x	
51	Lactate	x		x	
52	Phenylalanine	x		x	
53	Glycerate	x[1]	x	x	
54	Glycerol	x[1]	x	x	

No.	Metabolite	CE-TOFMS Detected	CE-TOFMS Significant (p <0.05)	GC×GC-TOFMS Detected	GC×GC-TOFMS Significant (FDR <0.05)
55	2-Hydroxy-3-methylbutyrate			x	
56	2-Hydroxybutyrate			x	
57	3-Hydroxybutyrate			x	
58	3-Methyl-2-oxovalerate			x	
59	4-Methyl-2-oxovalerate			x	
60	Arachidate			x	
61	Decanoate			x	
62	Dodecanoate			x	
63	Ethanolamine			x	
64	Fructose			x	
65	Glucosamine			x	
66	Hexadecanoate			x	
67	Methionine			x	
68	N-Acetylglucosamine			x	
69	Nicotinamide	[3]		x	
70	Nicotinate			x	
71	Octadecanoate			x	
72	O-Phosphorylethanolamine			x	
73	Palmitoleate			x	
74	Tetradecanoate			x	
75	Threonine	[3]		x	
76	Tyrosine			x	
77	Valerate	[3]		x	
78	Cadaverine			x[1]	
79	ADP	x			
80	CDP	x			
81	NAD$^+$/NADH	x			
82	UDP-D-glucuronate	x			
83	Uridine	x			
84	NADP$^+$/NADPH	x			
85	Pentose-phosphate	x		[4]	
86	4-Hydroxybenzyl alcohol	[3]			
87	Adenine	[3]			
88	Adenosine	[3]			
89	Arginine	[3]			
90	cyclic-AMP	[3]			
91	Cytidine	[3]			
92	Hypoxanthine	[3]			
93	Uracil	[3]			
94	Glutamine	[3]		*	
95	Nonanoate	x		IS[5]	
96	*Cinnamate-d7*	-		IS	
97	*Glucose-$^{13}C_6$*	-		IS	
98	*Glutamate-d5*	IS		-	
99	*Lactate-$^{13}C_3$*	IS		IS	
100	*Malate-d3*	IS		IS	
101	*Norvaline*	-		IS	
102	*PIPES*	IS		-	
103	*Pyruvate-$^{13}C_3$*	-		IS	
104	*Succinate-d4*	IS		IS	
105	*Uneven numbered fatty acids (C9-C19)*	-		IS	

* Glutamate and glutamine rearrange to pyroglutamate, [1] not confirmed by a commercial standard,
[2] γ-aminobutyrate was detected, but not confirmed by a commercial standard, [3] Metabolite was not detected by CE-TOFMS in *E. coli* extracts from cells grown in LB medium in the stationary phase, but were found in extracts of cells grown with minimal medium, [4] not properly assignable.

Table S4. Comparison of metabolites identified in extracts of two *E. coli* strains using GC×GC-EI-TOFMS and GC-APCI-TOFMS analyses. Reproduced from [171].

Metabolite	GC×GC-EI-TOFMS (2009) 27 identified metabolites out of 48 features		GC-APCI-TOFMS (2010) 25 identified metabolites out of 45 features	
	Detected	Significant (adjusted p-value[a] <0.05)	Detected	Significant (adjusted p-value[a] <0.05)
Citrate	x	x	x	x
Fumarate	x	x	x	x
Succinate	x	x	x	x
α-Ketoglutarate	x	x	x	x
Orotate	x	x	x	x
Itaconate	x	x	x	x
cis-Aconitate	x	x	x	x
Lysine	x	x	x	x
Valine	x	x	x	x
Leucine	x	x	x	x
Myo-inositol	x	x	x	x
Malate	x	x	x	x
Isoleucine	x	x	x	x
Proline	x	x	x	
Putrescine	x	x	x	
Aminoadipate	x	x	x	
Pyroglutamate	x	x	x	
Caproate	x	x	x	
Furoate	x	x	x	
Glucose	x	x	x	
Glucose-6P	x	x	x	
Sucrose	x	x	x	
Pyruvate	x	x	x	
Indole	x	x	x	
Glycine	x	x	x	
Tartrate	x	x	x	
GABA	x	x	x	
Pantothenate	x		x	x
Nicotinate	x		x	x
Phosphoglycerate	x		x	x
Methionine	x		x	x
Phosphoenolpyruvate	x		x	x
Phenylalanine	x		x	x
Dihydroorotate			x	x
N-Acetyl-Aspartate			x	x
N-Acetyl-Neuraminate			x	x
Thymine			x	x
Acetylputrescine			x	x
Ornithine			x	x
Glutamate-5-phosphate[b]			x	x

[a] False discovery rate according to Benjamini and Hochberg.[113]

[b] Tentatively identified without commercial standard.

Table S5. Mastermix compounds, internal standards and uneven numbered fatty acids with their respective quantification traces (hard and soft ionization techniques). Modified from [171].

Compound	Retention times (1D, 2D) [s]	*m/z* quantifier EI fragment ion	*m/z* quantifier [M+H]$^+$ ion
Pyruvate	591, 1.860	174	190
Lactate	603, 1.800	219	235
2-Ketobutyrate	643, 1.905	188	204
Ala-2TMS	647, 1.835	190	234
2-Hydroxybutyrate	671, 1.835	205	249
3-Hydroxybutyrate	703, 1.885	233	249
3-Methyl-2-oxovalerate	727, 1.960	203	232
4-Methyl-2-oxovalerate	759, 1.965	200	232
Methylmalonate	759, 2.065	247	263
Val-2TMS	763, 1.890	218	262
Leu-2TMS	819, 1.890	232	276
Glycerol	827, 1.790	218	309
Ile-2TMS	843, 1.910	232	276
Nicotinate	843, 2.370	180	196
Phenylacetate	847, 2.225	193	209
Pro-2TMS	851, 1.950	216	260
Gly-3TMS	859, 1.915	276	292
Succinate	859, 2.145	172	263
Glycerate	879, 1.910	292	323
Fumarate	891, 2.125	245	261
Malate	1035, 2.010	233	351
Adipate	1039, 2.190	111	291
Erythritol	1055, 1.790	217	411
GABA-3TMS	1071, 1.990	304	320
α-Ketoglutarate	1107, 2.195	198	320
Phenyllactate	1119, 2.190	193	311
Phenylpyruvate	1123, 2.260	265	266
Hydroxyphenylacetate	1163, 2.235	179	297
Putrescine-4TMS	1251, 1.910	214	377
cis-Aconitate	1251, 2.220	229	391
Glycerol-1-phosphate	1271, 2.160	357	461
Homovanillate	1271, 2.325	326	327
Phosphoglycerate	1307, 2.280	357	475
Citrate	1315, 2.140	273	481
Homogentisate	1323, 2.215	384	385
Hippurate-1TMS	1331, 3.005	236	252
Fructose	1371, 1.835	307	570
Hydroxyphenylpyruvate	1375, 2.255	190	354
Glucose	1399, 1.870	319	570
Myo-Inositol	1523, 1.955	305	613

Table S5 continued.

Compound	Retention times (1D, 2D) [s]	m/z quantifier EI fragment ion	m/z quantifier [M+H]$^+$ ion
5-Hydroxyindoleacetate-2TMS	1591, 2.975	335	336
Glucose-6-phosphate	1691, 2.385	387	722
Lactose	2043, 3.500	361	948
Surrogates			
[U-^{13}C]pyruvate	591, 1.860	177	193
[U-^{13}C]lactate	603, 1.800	222	238
[U-^{13}C]3-hydroxybutyrate	703, 1.885	237	253
Norvaline-2TMS	787, 1.885	144	262
[U-^{2}H]succinate	855, 2.125	176	267
[U-^{13}C]fumarate	891, 2.125	249	265
[2,3,3-^{2}H$_3$]malate	1031, 2.045	236	354
[^{2}H$_7$]trans-cinnamate	1079, 2.370	212	228
[2,2,4,4-^{2}H$_4$]citrate	1315, 2.090	276	485
[U-^{13}C]glucose	1399, 1.870	323	576
[4,6,7-^{2}H$_3$]5-hydroxyindole-[^{2}H$_2$]acetate	1587, 2.970	340	341
[U-^{13}C]lactose	2043, 3.500	367	960
Derivatization standard – uneven numbered fatty acids			
Nonanoic acid	903, 1.980	215	231
Undecanoic acid	1083, 2.025	243	259
Tridecanoic acid	1247, 2.040	271	287
Pentadecanoic acid	1395, 2.075	299	315
Heptadecanoic acid	1531, 2.160	327	343
Nonadecanoic acid	1659, 2.285	355	371

Table S6. R^2 values, LODs and RSDs for technical replicates for evaluated GC-MS approaches. Reproduced from [171].

Compound	GC-APCI-TOFMS			GC×GC-EI-TOFMS			GC-EI-TOFMS			GC-CI-qMS			GC-EI-qMS		
	R^2	LOD [µM]	RSD[b] (%)	R^2	LOD [µM]	RSD[b] (%)	R^2	LOD [µM]	RSD[b] (%)	R^2	LOD [µM]	RSD[c] (%)	R^2	LOD [µM]	RSD[c] (%)
Ala-2TMS	0.9907	0.98	2.5	0.9979	0.49	4.5	0.9908	1.95	8.3	0.9913	7.81	4.6	0.9955	7.81	2.4
Val-2TMS	0.9960	0.24	3.8	0.9985	0.03	3.2	0.9944	0.24	8.8	0.9919	1.95	4.5	0.9975	1.95	2.1
Leu-2TMS	0.9934	0.49	8.1	0.9930	0.98	5.9	0.9932	11.72	8.2	0.9979	3.91	6.7	0.9910	7.81	2.0
Ile-2TMS	0.9972	0.49	7.0	0.9843	0.98	3.2	0.9907	1.95	5.9	0.9901	3.91	2.3	0.9954	7.81	2.4
Pro-2TMS	0.9926	1.95	2.7	0.9974	3.91	7.1	0.9906	23.45	7.3	0.9935	7.81	3.8	0.9913	11.72	2.5
Gly-3TMS	0.9969	0.24	5.0	0.9981	0.03	4.6	0.9905	0.98	7.6	0.9920	1.95	4.4	0.9981	3.91	2.4
Putrescine-4TMS	0.9981	0.98	3.1	0.9920	7.81	2.3	0.9933	23.45	3.0	0.9908	7.81	6.0	0.9980	0.98	1.0
GABA-3TMS	0.9910	1.95	1.5	0.9944	0.49	3.8	0.9890	1.95	5.9	0.9900	0.98	5.5	0.9965	0.98	0.4
Lactate[a]	0.9949	0.98	0.3	0.9994	0.01	0.7	0.9975	3.91	2.5	0.9958	7.81	8.1	0.9964	1.95	2.3
2-OH-butyrate	0.9966	0.12	2.5	0.9995	0.01	2.0	0.9958	0.06	1.7	0.9967	23.44	4.8	0.9978	3.91	2.6
3-OH-butyrate[a]	0.9984	0.06	2.4	0.9992	0.01	0.8	0.9971	0.98	0.2	0.9980	11.72	1.6	0.9990	1.95	3.7
Me-malonate	0.9956	0.06	3.7	0.9972	0.06	4.2	0.9953	0.24	3.3	0.9968	3.91	4.5	0.9978	7.81	0.9
Glycerol	0.9903	1.95	4.5[d]	0.9976	0.01	2.1	0.9978	0.98	3.5	0.9967	3.91	4.0	0.9972	1.95	0.5
Succinate[a]	0.9973	0.12	4.0	0.9978	0.02	2.6	0.9940	1.95	3.7	0.9935	11.72	4.8	0.9994	1.95	2.1
Glycerate	0.9968	0.06	3.3	0.9949	0.02	2.8	0.9969	0.98	4.2	0.9964	15.63	3.0	0.9980	1.95	0.5
Fumarate[a]	0.9977	0.24	1.1	0.9997	0.01	1.5	0.9962	0.24	0.6	0.9992	0.98	5.4	0.9991	0.06	0.7
Malate[a]	0.9967	0.24	2.9	0.9988	0.01	0.8	0.9955	0.98	0.6	0.9958	7.81	1.4	0.9993	0.24	1.5
Adipate	0.9973	0.24	4.4	0.9908	0.01	3.0	0.9963	3.91	3.4	0.9934	15.63	3.4	0.9988	0.98	0.8
Erythritol	0.9948	0.06	2.6	0.9960	0.01	1.7	0.9898	1.95	2.7	0.9925	11.72	1.5	0.9993	0.98	1.0
cis-Aconitate	0.9969	0.06	1.9	0.9979	0.02	1.8	0.9962	3.91	7.1	0.9929	15.63	10.1	0.9950	0.98	0.6
Citrate[a]	0.9976	0.49	1.4	0.9994	0.01	0.6	0.9947	1.95	0.3	0.9915	7.81	11.3	0.9993	0.12	1.2
Pyruvate[a]	0.9986	0.12	1.3	0.9996	0.004	1.2	0.9981	3.91	3.7	0.9974	7.81	9.1	0.9994	0.98	4.3
2-Ketobutyrate	0.9969	0.24	4.0	0.9901	0.02	5.1	0.9900	11.72	3.2	0.9913	15.63	6.2	0.9989	3.91	3.0
3-Me-2-oxovalerate	0.9980	0.12	2.3	0.9979	0.12	2.7	0.9963	0.24	4.1	0.9930	23.44	5.8	0.9933	11.72	6.2

Table S6 continued.

Compound	GC-APCI-TOFMS			GC×GC-EI-TOFMS			GC-EI-TOFMS			GC-CI-qMS			GC-EI-qMS		
	R^2	LOD [µM]	RSD[b] (%)	R^2	LOD [µM]	RSD[b] (%)	R^2	LOD [µM]	RSD[b] (%)	R^2	LOD [µM]	RSD[c] (%)	R^2	LOD [µM]	RSD[c] (%)
4-Me-2-oxovalerate	0.9960	0.24	6.0	0.9928	0.02	4.5	0.9956	0.24	4.3	0.9900	11.72	2.7	0.9976	3.91	1.5
α-Ketoglutarate	0.9947	0.06	3.1	0.9965	0.03	1.4	0.9901	1.95	2.8	0.9961	15.63	4.4	0.9948	1.95	1.4
Phenylacetate	0.9970	0.98	3.1	0.9981	0.03	2.9	0.9875	0.03	3.0	0.9948	11.72	3.0	0.9925	7.81	2.5
Phenylpyruvate	0.9980	0.24	1.8	0.9983	0.49	2.0	0.9901	15.63	1.4	0.9986	7.81	2.8	0.9923	31.25	1.9
Hydroxyphenylacetate	0.9969	0.24	4.8	0.9973	0.02	2.2	0.9911	0.49	4.5	0.9954	3.91	3.7	0.9984	1.95	0.8
Homovanillate	0.9969	0.49	0.5	0.9940	0.24	2.8	0.9935	0.98	3.1	0.9974	3.91	2.4	0.9926	0.98	1.3
Hippurate-1TMS	0.9977	0.49	0.8	0.9963	93.75	nd	0.9930	93.75	nd	0.9984	15.63	5.5	0.9963	31.25	3.0
Homogentisate	0.9963	0.24	4.6[d]	0.9968	0.06	2.3	0.9941	3.91	2.7	0.9967	3.91	4.4	0.9985	0.49	1.9
OH-phenylpyruvate	0.9990	0.24	3.6	0.9935	1.95	4.5	0.9943	7.81	5.3	0.9907	15.63	11.0	0.9924	1.95	2.1
5-HIAA-2TMS[a]	0.9987	0.49	2.7	0.9990	0.03	3.9	0.9956	11.72	4.9	0.9959	3.91	4.1	0.9972	11.72	0.7
Phenyllactate	0.9960	0.49	1.3	0.9986	0.06	5.6	0.9956	0.98	5.3	0.9990	7.81	8.3	0.9975	0.98	2.0
Nicotinate	0.9976	0.49	1.0	0.9967	0.01	2.1	0.9983	0.49	2.7	0.9980	3.91	1.5	0.9992	0.98	4.1
Glycerol-1-P	0.9974	0.24	1.8	0.9975	1.95	2.3	0.9931	1.95	2.5	0.9949	7.81	5.6	0.9930	3.91	1.6
3-P-glycerate	0.9971	0.24	4.0	0.9932	11.72	3.9	0.9934	31.25	3.7	0.9953	7.81	3.7	0.9920	3.91	1.4
Glucose-6P	0.9965	0.98	2.1	0.9925	7.81	4.2	0.9948	15.63	5.8	0.9904	31.25	2.6	0.9953	0.98	2.5
Lactose[a]	0.9920	0.98	6.6	0.9995	0.02	0.3	0.9972	0.49	0.3	0.9981	7.81	7.4	0.9994	1.95	0.6
Myo-Inositol	0.9980	0.98	2.4	0.9980	0.02	2.5	0.9907	0.06	2.3	0.9920	1.95	7.3	0.9986	0.12	0.9
Glucose[a]	0.9984	0.06	0.5	0.9998	0.02	0.2	0.9964	0.49	0.3	0.9949	11.72	3.9	0.9992	1.95	1.9
Fructose	0.9981	0.12	1.8	0.9985	0.03	0.5	0.9911	0.12	1.0	0.9926	3.91	3.4	0.9983	1.95	1.6

[a] Compounds with a corresponding internal standard, [b] n=5, conc. = 62.5 µM, [c] n=5, conc. = 500 µM, [d] Standard conc. above ULOQ.

nd, not detected.

13 Curriculum vitae

Personal Data

Name	Martin F. Almstetter
Date of birth	19.06.1979
Nationality	German

Education

08/2007 – present	Natural Sciences Graduate Student at the Institute of Functional Genomics, University of Regensburg, Germany (Advisor: Prof. Dr. Peter J. Oefner)
12/2006 – 07/2007	Scientific assistant at the Institute of Analytical Chemistry, Instrumental Analysis, University of Regensburg, Germany
03/2006 – 11/2006	Diploma thesis at the Institute of Analytical Chemistry, Chemo- and Biosensors, University of Regensburg, Germany, on *"Luminescent Trace Oxygen Sensors Applicable to High Temperature Conditions"* (Advisor: Prof. Dr. Otto S. Wolfbeis)
11/1999 – 11/2006	Chemistry studies at the University of Regensburg
1998 – 1999	Sportfördergruppe of the German Army
1989 – 1998	High school in Ingolstadt

Stipends and Awards

Scholarship for the International Symposium on Chromatography (ISC) 2008

Biomedical research prize 2009 by the University of Regensburg for the publication: Almstetter et al., Anal Chem 2009; 81(14): 5731-9

Poster Award, 1st Prize: Almstetter MF, Dettmer K, Oefner PJ, *Strategies for quantitative metabolite analysis with GC×GC-TOFMS*, 27th International Symposium on Chromatography, September, 21st – 25th 2008, Münster, Germany

Activities

Baseball (German National Team, 1. Bundesliga)

14 Publications and Presentations

14.1 Publications

[1] **Almstetter MF**, Oefner PJ, Dettmer K.

Comprehensive two-dimensional gas chromatography in metabolomics.

Analytical Bioanalytical Chemistry; Review [accepted].

[2] Dettmer K, **Almstetter MF**, Wachsmuth CJ, Oefner PJ.

Comprehensive two-dimensional gas chromatography for metabolomics.

Book chapter in *The Handbook of Plant Metabolomics - Metabolite Profiling and Networking;* Wiley-VCH: Weinheim, 2012; in print.

[3] **Almstetter MF**, Oefner PJ, Dettmer K.

Metabolic fingerprinting using comprehensive two-dimensional gas chromatography – time-of-flight mass spectrometry (GC×GC-TOFMS).

Book chapter in *Functional Genomics - Methods and Protocols, 2nd edition*; Humana Press: New York, 2012; 815, 399-412.

[4] Wachsmuth CJ, **Almstetter MF**, Waldhier MC, Gruber MA, Nürnberger N, Oefner PJ, Dettmer K.

Performance evaluation of gas chromatography – atmospheric pressure chemical ionization – time-of-flight mass spectrometry for metabolic fingerprinting and profiling.

Analytical Chemistry 2011; **83**(19): 7514-22.

[5] **Almstetter MF***, Appel IJ*, Dettmer K, Gruber MA, Oefner PJ.

Comparison of two algorithmic data processing strategies for metabolic fingerprinting by comprehensive two-dimensional gas chromatography – time-of-flight mass spectrometry.

Journal of Chromatography A 2011; **1218**(39): 7031-8. *Equal first authors/contributed equally.

[6] Waldhier MC*, **Almstetter MF***, Nürnberger N, Gruber MA, Dettmer K, Oefner PJ.

Improved enantiomer resolution and quantification of free D-amino acids in serum and urine by comprehensive two-dimensional gas chromatography – time-of-flight mass spectrometry.

Journal of Chromatography A 2011; **1218**(28): 4537-44. *Equal first authors/contributed equally.

[7] Dettmer K, Nürnberger N, Kaspar H, Gruber MA, **Almstetter MF**, Oefner PJ.

Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols.

Analytical Bioanalytical Chemistry 2011; **399**(3): 1127-39.

[8] Dettmer K*, **Almstetter MF***, Appel IJ, Nürnberger N, Schlamberger G, Gronwald W, Meyer HDD, Oefner PJ.

Comparison of serum versus plasma collection in gas chromatography – mass spectrometry based metabolomics.

Electrophoresis 2010; **31**(14): 2365-73. *Equal first authors/contributed equally.

[9] Klein MS*, **Almstetter MF***, Schlamberger G, Nürnberger N, Dettmer K, Oefner PJ, Meyer HHD, Wiedemann S, Gronwald W.

NMR and mass spectrometry based milk metabolomics in dairy cows during early and late lactation.

Journal of Dairy Science 2010; **93**(4): 1539-50. *Equal first authors/contributed equally.

[10] **Almstetter MF***, Appel IJ*, Gruber MA, Lottaz C, Timischl B, Spang R, Dettmer K, Oefner PJ.

Integrative normalization and comparative analysis for metabolic fingerprinting by comprehensive two-dimensional gas chromatography – time-of-flight mass spectrometry.

Analytical Chemistry 2009; **81**(14): 5731-9. *Equal first authors/contributed equally.

14.2 Oral Presentations

Data Processing Strategies for Metabolic Fingerprinting by GC×GC-TOFMS.
Invited talk, 2nd European GC×GC Symposium by LECO, Regensburg, Germany, Sep 20th, 2011.

Alignment Routines for Metabolic Fingerprinting by GC×GC-TOFMS.
Invited talk, European LECO Pegasus Users Meeting, Aachen, Germany, Nov 30th, 2010.

Comparison of Alignment Routines for Metabolic Fingerprinting by GC×GC-TOFMS.
34th International Symposium on Capillary Chromatography (ISCC) and 7th GC×GC Symposium, Riva del Garda, Italy, May 31st, 2010.

Gas Chromatography – Mass Spectrometry in Metabolomics.
Symposium of the Bavarian Genome Research Network, Regensburg, Germany, April 26th, 2010.

GC×GC-TOFMS in Metabolomics.
20th Postgraduate Seminar of the Separation Science workgroup in the German Chemical Society (GDCh), Hohenroda, Germany, Jan 11th, 2010.

Integrative Normalization and Comparative Analysis (INCA) for Metabolic Fingerprinting.
Invited talk, 1st European GC×GC Symposium by LECO, Amsterdam, Netherlands, Jun 18th, 2009.

Integrative Normalization and Comparative Analysis (INCA) for Metabolic Fingerprinting by Comprehensive Two-dimensional Gas Chromatography – Time-of-flight Mass Spectrometry.
Invited talk, LECO Practice Workshop, Mönchengladbach, Germany, Oct 28th, 2008.

GC×GC-TOFMS in Metabolomics.
32nd International Symposium on Capillary Chromatography and 5th GC×GC Symposium, Riva del Garda, Italy, May 26th, 2008.

14.3 Poster Presentations

Klein MS, **Almstetter MF**, Nürnberger N, Louis C, Sigl G, Wiedemann S, Meyer HDD, Dettmer K, Oefner PJ, Gronwald W.
Nuclear Magnetic Resonance and Mass Spectrometry Based Milk and Plasma Metabolomics in Dairy Cows during Early and Late Lactation.
Fugato Status-Seminar, Kassel, Germany (2011).

Waldhier MC, **Almstetter MF**, Gruber MA, Dettmer K, Oefner PJ.
Chiral amino acid analysis of physiological fluids by GC-MS.
34th International Symposium on Capillary Chromatography and 7th GC×GC Symposium, Riva del Garda, Italy (2010).

Klein MS, **Almstetter MF**, Schlamberger G, Nürnberger N, Dettmer K, Oefner PJ, Meyer HDD, Wiedemann S, Gronwald W.
NMR and mass spectrometry based milk metabolomics in dairy cows during early and late lactation.
31st German Chemical Society Discussion Meeting on Magnetic Resonance in Chemistry and Materials Research, Dresden, Germany (2009).

Almstetter MF, Dettmer K, Oefner PJ.
Strategies for quantitative metabolite analysis with GC×GC-TOFMS.
27th International Symposium on Chromatography, Münster, Germany (2008), awarded as best poster.

15 Summary

The study of cellular metabolite profiles and changes therein due to genetic and environmental influences is termed metabolomics. A major, yet to be realized goal is the detection and quantification of all metabolites in a single analysis.

In this work, a comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (GC×GC-TOFMS) method was designed and validated that assembles the entire available analytical information from each sample in one data matrix for subsequent statistical evaluation. For the purpose of data merging we developed and implemented the retention time correction and data alignment tool INCA (Integrative Normalization and Comparative Alignment). The INCA module capitalized on the characteristic fragmentation behavior of silylated metabolites upon EI ionization by using the integral of the *m/z* 73 ion trace of the trimethylsilyl (TMS) group as quantitative measure for all features. The method was applied to reveal differences in metabolite composition between (i) an *Escherichia coli* wild type and a double-mutant strain lacking the transhydrogenases UdhA and PntAB and (ii) serum and plasma. Subsequently, we evaluated the performance of the Statistical Compare alignment function introduced later by LECO for GC×GC-TOFMS data and compared it to INCA.

GC×GC-TOFMS was comprehensively evaluated against various 1D-GC-MS techniques using a set of 43 metabolite standards from different chemical classes and metabolic pathways. GC×GC-TOFMS proved to be the most powerful method with a linear range of more than three orders of magnitude, LLOQs in the sub-micromolar and LODs in the nanomolar range.

GC×GC-TOFMS was also employed for absolute quantification of amino acid enantiomers (AAEs) as their methyl chloroformate derivatives and results were compared to those of a previously established 1D-GC-qMS method with single ion monitoring. The coupling of a γ-cyclodextrin (Rt[®]-γDEXsa) with an amino acid selective (ZB-AAA) column resulted in enhanced peak resolution. Twenty AAEs including the critcal peak pair L-leucine/D-isoleucine, which exhibited equal fragmentation behavior upon EI ionization in 1D-GC-MS, could be baseline separated. Except for methionine enantiomers, distinctly improved LLOQs were obtained. The method was applied to the analysis of AAE serum concentrations in

patients suffering from liver cirrhosis and showed significantly increased D-AA concentrations and slightly decreased L-AA levels compared to a control group.

16 Zusammenfassung

Die Fachrichtung Metabolomik beschäftigt sich mit der umfassenden Analyse von zellulären Stoffwechselprodukten in Abhängigkeit von genetischen und umweltbedingten Einflüssen. Ein wesentliches Ziel ist die Detektion und Quantifizierung möglichst aller Metabolite in einer einzigen Messung.

In dieser Arbeit wurde auf Basis der komprehensiven zweidimensionalen Gaschromatographie gekoppelt mit der Flugzeitmassenspektrometrie (GC×GC-TOFMS) ein Verfahren zur Extraktion der gesamten zur Verfügung stehenden analytischen Informationen aus einem GC×GC-TOFMS Lauf und der Zusammenführung mehrerer Läufe in einer Datenmatrix zur nachfolgenden statistischen Auswertung des gesamten Datensatzes entwickelt. Zum Zweck der Datenzusammenführung haben wir den INCA (Integrative Normalization and Comparative Alignment) Algorithmus zur Retentionszeit-Korrektur und zum Datenabgleich entwickelt und implementiert. INCA nutzt das charakteristische Fragmentierungsverhalten von silylierten Metaboliten bei vorangehender Elektronen-Ionisierung (EI) und setzt das Integral der *m/z* 73 Ionenspur der Trimethylsilylgruppe als quantitative Größe für alle Einträge der Peak-Listen ein. Diese vergleichende Metabolic Fingerprinting Strategie wurde verwendet, um globale metabolische Veränderungen zu detektieren. Zum einen wurde mittels GC×GC-TOFMS ein *E. coli* Wildtypstamm im Vergleich mit der entsprechenden Deletionsmutante *UdhA-PntAB* untersucht, in der beide Nikotinamidnukleotid-Transhydrogenasen fehlen. Zum anderen wurden Unterschiede zwischen Serum- und Plasma-Proben analysiert. In einem separaten Verfahren wurde die später entwickelte Statistical Compare (SC) Funktion von LECO zur Datenanordnung in die Datenprozessierungs-Strategie mit eingebunden. Diverse Vor- und Nachteile von SC wurden im direkten Vergleich zu INCA ermittelt.

Das Leistungsvermögen der GC×GC-TOFMS im Vergleich zu anderen GC-MS Techniken wurde anhand von 43 Metaboliten von unterschiedlichen chemischen Klassen und metabolischen Stoffwechselwegen evaluiert. GC×GC-TOFMS erwies sich als die leistungsfähigste Methode mit linearen Bereichen von mehr als drei Größenordnungen, unteren Bestimmungsgrenzen im sub-mikromolaren und Nachweisgrenzen im nanomolaren Bereich.

Desweiteren wurde GC×GC-TOFMS zur absoluten Quantifizierung von Aminosäure Enantiomeren als deren Methylchloroformat Derivate angewendet. Die Resultate wurden mit denen einer im Vorfeld etablierten 1D-GC-qMS Methode mit Einzelionen-Nachweis verglichen. Die Kopplung einer γ-Cyclodextrin (Rt®-γDEXsa) mit einer Aminosäure-selektiven (ZB-AAA) Säule führte zu einer deutlich verbesserten Auflösung. Zwanzig Aminosäure Enantiomere konnten bis zur Grundlinie voneinander getrennt werden, einschließlich des kritischen Peakpaares L-Leucin/D-Isoleucin, das in der 1D-GC-MS zur gleichen Retentionszeit eluierte und dasselbe Fragmentierungsmuster zeigte. Es wurden, mit Ausnahme der Methionin Enantiomere, im Allgemeinen verbesserte untere Bestimmungsgrenzen erzielt. Die Methode wurde zur Bestimmung der Konzentrationen von Aminosäure Enantiomeren in Serumproben von Leberzirrhose-Patienten angewandt. Im Vergleich zu einer Kontroll-Gruppe wurden erhöhte D- und leicht verminderte L-Aminosäure Konzentrationen festgestellt.

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Regensburg, 06.12.2011

Martin Almstetter