

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/138666> holds various files of this Leiden University dissertation.

Author: Laan, T. van der

Title: High-throughput quantification and unambiguous identification for metabolomics

Issue Date: 2020-12-14

**High-throughput quantification and
unambiguous identification for
metabolomics**

The publication of this thesis was financially supported by:

Leiden Academic Centre for Drug Research

Leiden University Libraries

AB Sciex

Cover design: Anna van Duijn

Thesis lay-out: Tom van der Laan

Printing: Ede Printservice

© Copyright, Tom van der Laan, 2020

ISBN: 978-90-831117-2-8

All rights reserved. No part of this book may be reproduced in any form or by any means without permission of the author.

High-throughput quantification and unambiguous identification for metabolomics

Proefschrift

Ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op maandag 14 december 2020
klokke 15:00 uur

door

Tom van der Laan

Geboren te Amersfoort, Nederland
in 1992

PROMOTOR

Prof.dr. T. Hankemeier

CO-PROMOTOR

Dr.ir. A.C. Dubbelman

Dr. A.C. Harms

PROMOTIECOMMISSIE

Prof.dr. H. Irth (Chair)

Prof.dr. J.A. Bouwstra (secretary)

Prof.dr.ir. P.J. Schoenmakers

Prof.dr. J.P.M. van Duynhoven

Prof.dr. G. Hopfgartner

Prof.dr. E.C.M. de Lange

The research described in this thesis was performed at the division Systems Biomedicine and Pharmacology of the Leiden Academic Centre for Drug Research (LACDR), Leiden University (Leiden, The Netherlands). The research was financially supported by COAST/NWO (project no. 053.21.118).

CONTENTS

Chapter 1	General introduction and scope	1
Chapter 2	Fast LC-ESI-MS/MS analysis and influence of sampling conditions for gut metabolites in plasma and serum <i>Scientific Reports (2019)</i>	21
Chapter 3	High-throughput fractionation coupled to mass spectrometry for improved quantitation in metabolomics <i>Analytical Chemistry (2020)</i>	45
Chapter 4	Data independent acquisition for the quantification and identification of metabolites in plasma <i>Submitted for publication</i>	77
Chapter 5	Fractionation platform for target identification using off-line directed 2D-LC, MS and NMR <i>Analytica Chimica Acta (2020)</i>	101
Chapter 6	Conclusions and perspectives	137
Addendum	Nederlandse Samenvatting Curriculum Vitae Acknowledgements	153

Chapter 1

General introduction and scope

METABOLOMICS

Metabolomics is the discipline which focuses on the analysis of small molecules within a biological system. These small molecules are referred to as metabolites and comprise for example amino acids, fatty acids, sugars, hormones and vitamins. The metabolome is the collection of all metabolites in a biological system and the comprehensive profiling of this metabolome is the main goal in metabolomics. Because metabolites are the smallest molecules present in biological systems, they are often a result of several chemical processes.¹ These processes are affected by two factors: the genome and the exposome. The genome represents all genetic material of an organism and its influence on metabolite levels has been demonstrated for numerous metabolic pathways including amino acids, vitamin and cofactors, fatty acid metabolism and glucose homeostasis.² The exposome, on the contrary, represents all environmental factors to which a biological system is exposed during lifetime.³ For example, the intestinal absorption of fatty acids after food ingestion directly increases the systemic circulation of fatty acids.⁴ A more indirect effect can be exemplified by the serum metabolite levels of smokers. In comparison with non-smokers, smokers demonstrate a significantly different metabolic profile, which is reversible after the cessation of smoking.⁵ Next to food ingestion and smoking behavior, many other environmental factors, e.g. infectious agents, pollutions, gut microbiota and stress, can influence the metabolome composition as well. Because metabolite levels are influenced by both genetic and environmental factors, it offers a functional readout of a biological system. This readout can be useful for numerous applications. In this thesis, we focused on applications in the field of healthcare and food.

Metabolomics in healthcare

The metabolome is often considered as the phenotype of a biological system. This phenotype is specific to each individual and can, therefore, be a great asset in personalized healthcare.^{6,7} Personalization can be accomplished by comparing large-scale metabolic profiles of patients and healthy people (see Figure 1). These profiles can be used to identify metabolites with diagnostic or prognostic power.

A diagnostic biomarker can be used to indicate whether an individual is suffering from a certain disease whereas a prognostic biomarker can monitor the health status of an individual or predict drug response.⁸ One of the most well-known diagnostic metabolite biomarkers is phenylalanine, which is routinely screened in newborn babies to diagnose phenylketonuria (metabolic disorder).⁹

Static prognostic biomarkers can be screened before the administration of a drug. The levels of these metabolites are used to predict whether a drug will result in the desired effect. Serotonin

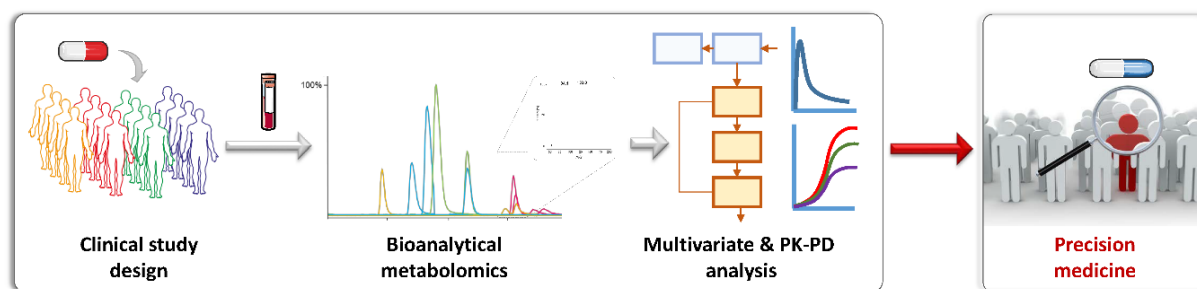


Figure 1. Precision medicine workflow using clinical metabolomics-based biomarker discovery and clinical pharmacology. Adapted from Kohler *et al.*⁸

serum levels, for example, have shown to be highly correlated with response to aspirin.¹⁰ The determination of serotonin levels in cardiovascular patients could help in predicting aspirin treatment efficacy. Patients, for whom drug efficacy is not expected, can more easily be identified preventing unnecessary side effects and facilitating the consideration of other treatments.

Dynamic prognostic biomarkers possess valuable information in determining the health status of an individual. These biomarkers can be used to monitor disease progression or drug response. Preventive screening of early biomarkers, that are indicative for the onset of a disease, may allow for dietary and lifestyle changes rather than more radical drug treatments or surgical intervention.¹¹ Wang *et al.* demonstrated that five branched-chain amino acids had a very high predictive power for the development of diabetes.¹² Preventive screening of these potential biomarkers can serve as a wakeup call for people that are in a high-risk category. Before the development of the actual disease, people can change their lifestyle increasing their life quality and decreasing the time and cost pressure on the healthcare system. In more severe disease stages, dynamic biomarkers can indicate the direction of the pathophysiology and demonstrate the efficacy of an administered drug. Huang *et al.* demonstrated seven metabolites that could potentially serve as prognostic biomarkers for endocrine therapy response in prostate cancer patients.¹³ Good responders demonstrated a similar metabolic profile in comparison with the healthy control group whereas the metabolic profile of poor responders remained abnormal after endocrine therapy.

Metabolomics in food

Metabolomics is applied to a wide variety of sample types, which includes food samples. Metabolic profiling of food samples is used in the agricultural sector and the food industry for many different purposes.¹⁴ The applications range from health regulations to the experience of taste.

Metabolomics can be used to assess food safety. Metabolic screening of food products is of high importance for human health because it can prevent food products containing toxic metabolite levels from entering the market. By the use of proper metabolomics tools, toxic levels of food

additives and pesticides can easily be identified.¹⁴ Metabolomics is also used to assess the safety of genetically modified crops by evaluating the metabolic differences of genetically modified crops and their natural counterparts.¹⁵ This can provide an extra dimension to the risk assessment of genetic modification and can help to decipher potential environmental and health hazards.

On the other hand, metabolomics is also used for the improvement of food. The link between plant genomics and metabolomics can help in understanding the underlying biochemical processes on certain trait formations in vegetables and fruits.¹⁶ This information can contribute to the development of crop species with higher agronomic and nutritional value. Taste is another difficult biochemical process that researchers try to explain by using metabolomics. To find taste-related metabolites, metabolic profiling is often combined with a taste evaluation performed by a trained sensory panel (see Figure 2).^{17,18} Metabolites that significantly correlate with the experience of a certain taste profile can be identified, which may contribute to the understanding of taste experience. A typical example of a taste-related metabolite is glutamate, which is mainly held accountable for the experience of umami flavours in food.¹⁹

METABOLOMICS ANALYSIS

Metabolomics is becoming increasingly important. The increased demand for metabolic profiles requires the use of sophisticated analytical instruments and methodologies. The most dominant techniques in the field of metabolomics are mass spectrometry (MS) and nuclear magnetic resonance (NMR). These techniques are used to address the two most important questions in metabolomics: what is the quantity of a metabolite and what is the identity of a metabolite in a

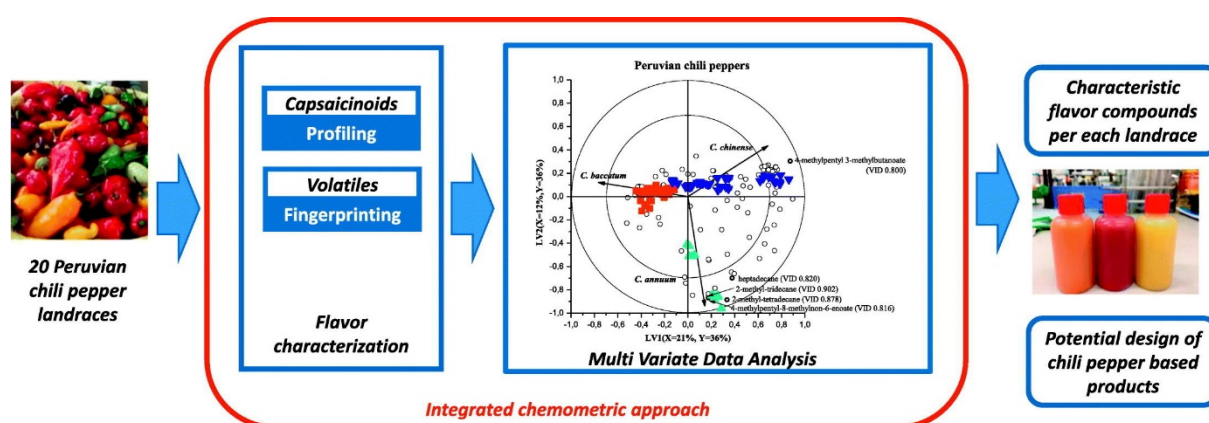


Figure 2. Pathway for the identification of taste-related features in food products. The profiling of several compound classes and flavor characterization can result in the identification of characteristic flavor compounds. Adapted from Morales-Soriano et al.¹⁸

sample. Figure 3 shows the increasing trend in metabolomics publications and the most popular analytical technologies that are used for the analysis of metabolites.

Quantification

The quantity of a metabolite is often measured in a targeted metabolomics analysis in which the metabolites of interest are already known in advance. A targeted analysis is used for the measurement of already validated biomarkers or metabolites that are expected to have a diagnostic or prognostic value.^{20,21} Because the identity of these metabolites is already known, it becomes easier to extract quantitative information from the acquired data. NMR is very powerful when it comes to the quantification of the more abundant metabolites in a complex sample because of its high reproducibility.²² The high reproducibility is caused by the fact that the NMR signal is proportional to the number of protons in case of ^1H NMR. ^1H NMR is most often used for quantitative purposes because the analysis of this nucleus results in the highest sensitivity. However, when it comes to the quantification of low abundant metabolites, MS is the technique of choice because of its low detection limits and high selectivity.¹ Triple quadrupole (QQQ) MS is the golden standard for MS-based quantification of small molecules in biological samples because of its high dynamic range, robustness and sensitivity.¹ QQQ analyses make use of MS/MS scans, in

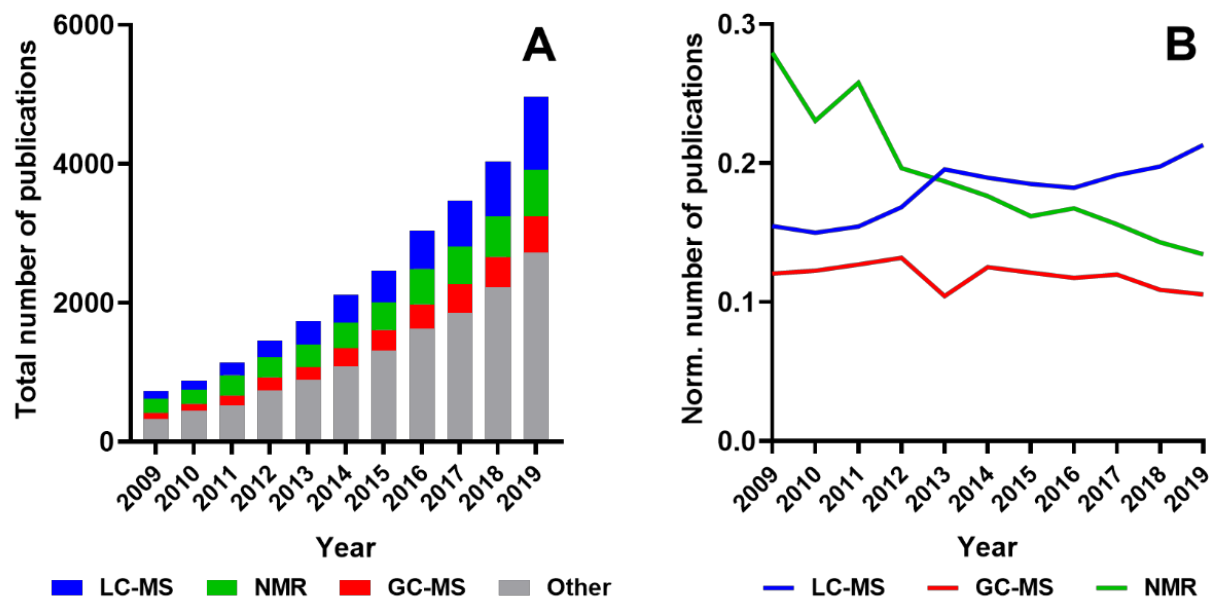


Figure 3. Overview of a literature survey on PubMed including metabolom* or metabonom* in the title or abstract in a period between 2009 and 2019. (A) The total number of publications using the most popular metabolomics techniques, i.e. NMR, LC-MS and GC-MS, are indicated with different colors. (B) Normalized number of publications using the most popular metabolomics techniques. The search terms are mentioned in Table S1 of the supplementary information.

which precursor mass of the analytes is selected and subsequently fragmented into product ions. Specific precursor and product ion transitions are optimized for each analyte and used for quantitative purposes.²³ High-resolution mass spectrometers are also used for quantitative purposes.²⁴ The benefit of high-resolution MS is that a single scan comprises all m/z values in a certain mass range. Therefore, the analysis is rather fast (only one scan needed, instead of multiple transition in tandem MS) and suitable for untargeted approaches because there is no preselection of analytes.

Identification

Structural elucidation of a metabolite is generally preceded by an untargeted metabolomics screening. In untargeted metabolomics, the metabolites of interest are not known in advance and, therefore, the used analytical methods strive to be non-selective to cover as many metabolites as possible.²⁵ The acquired data in these platforms can be used to correlate spectral features to a biological effect. Once a feature demonstrates a significant correlation, it becomes useful to know its structural formula which helps in understanding the biological meaning behind this effect.²⁶

^1H NMR is also predominantly used for metabolite identification.²⁷ Since proton signals of metabolites tend to overlap in the presence of complex biological samples, two-dimensional NMR has been used to obtain better resolved data. Homonuclear, e.g. ^1H - ^1H , and heteronuclear, e.g. ^1H - ^{13}C , two-dimensional NMR has shown to greatly improve spectral resolution and reduces peak overlap.²⁸ Many metabolite identifications are conducted for metabolites that have been identified before (non-novel metabolite identification). This makes it possible to search one- and two-dimensional NMR spectra in metabolite databases, like MMCD²⁹, BMRB³⁰, HMDB³¹ and COLMAR³², to facilitate metabolite identification.

High-resolution mass spectrometers, e.g. time-of-flight and orbitrap mass spectrometers, are mostly used for untargeted metabolomics screens.²⁵ A high-resolution scan comprises m/z values of all positive or negative ions in a certain mass range. The accuracy of these measurements is sufficient to suggest the elemental composition of a metabolite, especially when it is combined by the intensity ratios of the naturally occurring isotopes.³³ Despite its high selectivity, high-resolution scans do not provide structural information, unless it is combined with MS/MS fragmentation. The fragmentation of metabolites results in essential information for the structural annotation of an unknown metabolite. For non-novel metabolite identification, a spectrum of the produced product ions can be searched in MS/MS databases like NIST (<https://www.nist.gov/srd>), HMDB³¹, mzCloud (<https://www.mzcloud.org/>) and MassBank of North America (MoNA; <https://mona.fiehnlab.ucdavis.edu/>).

CHALLENGES IN METABOLOMICS ANALYSIS

Metabolic profiles have shown to be very useful for several disciplines. High-end analytical equipment has been used in order to obtain metabolic profiles in biological samples. However, these techniques are subjected to several challenges, which prevent their optimal use for metabolomics applications. The range of analytical techniques and methodologies used for metabolomics applications is wide, which can be explained by the specific advantages and disadvantages that are inherently coupled to the use of certain analytical methodologies. Each metabolomics application has its demands and, therefore, benefits from a specific type of analytical equipment and methodology. Here, we will introduce the main challenges of MS- and NMR-based metabolomics and possibilities on how to overcome them.

Mass spectrometry quantification

MS analysis requires analytes to be charged and in the gas phase. Electrospray ionization (ESI) is the most commonly used technique for the production of gas-phase ions before MS analysis. However, when multiple compounds enter the ESI source simultaneously, they start to enhance or suppress each other's ionization. This effect is known as matrix effect.³⁴ Especially with regards to complex samples (i.e. biological samples), matrix effect can result in a dramatically decreased

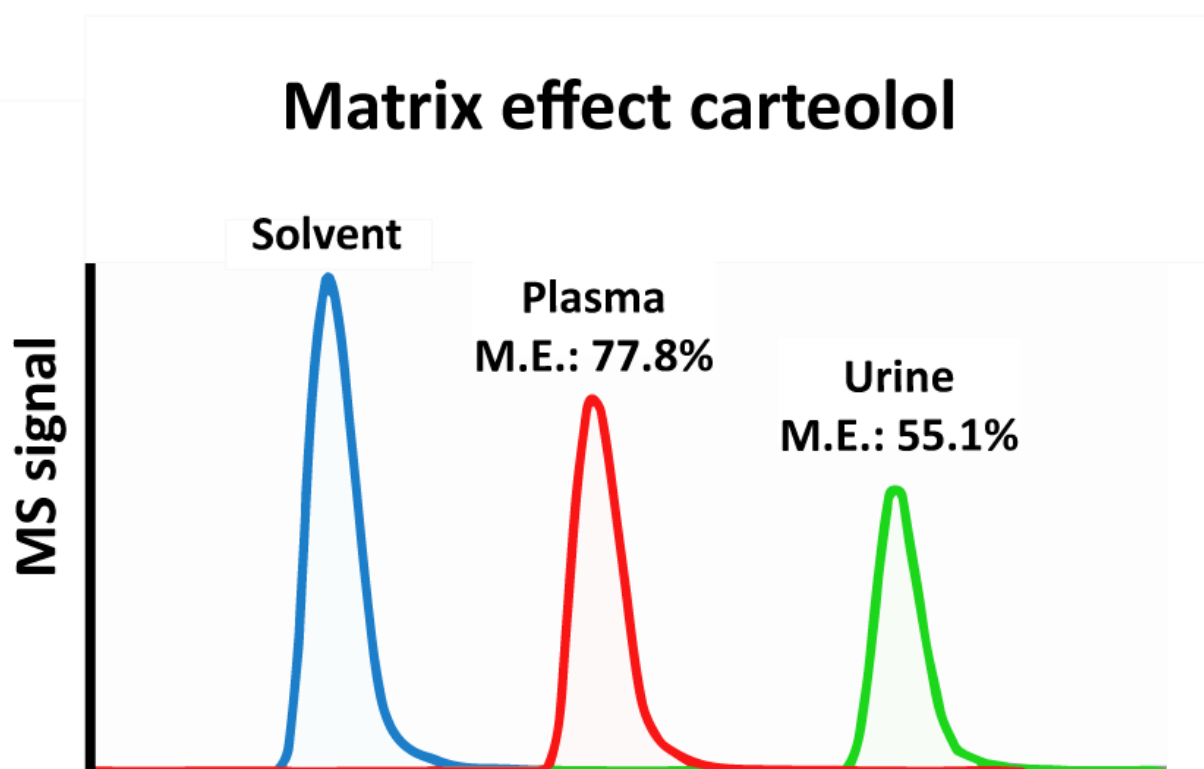


Figure 4. The MS signal of carteolol in different biological matrices. Matrix effect (M.E.) is observed in the plasma and urine analysis. The plasma and urine matrix reduced the MS signal of carteolol to 77.8% and 55.1% of its original value, respectively. Adapted from Gonzalez et al.⁷⁰

sensitivity due to ion suppression and very poor repeatability because of matrix differences (see Figure 4).³⁵ This impairs correct quantification as the signal of analytes can drop below detection limits and the quantification accuracy and precision can be severely compromised. To minimize matrix effect, MS is almost always coupled to a separation technique to reduce the sample complexity before MS analysis.³⁶ Liquid chromatography (LC) is the most commonly used separation technique, because of its high versatility.³⁷

Although the combination of LC and MS is the workhorse in MS-based metabolomics, the throughput is limited by the gradient time of the LC separation and many metabolomics applications require the quantitative analyses of thousands of samples to discover or validate a potential biomarker.³⁸ Therefore, high-throughput platforms are of utmost importance to analyze large-scale metabolomics studies. Separation-free MS-analyses have been developed to increase throughput. These platforms are mostly performed utilizing a flow-injection analysis (FIA) in which no separation takes place and a sample plug is directly introduced into the mass spectrometer.³⁹ The sample preparation of these methodologies is often performed by a dilute-and-shoot approach, which aims at minimizing the matrix effect at the ionization source by diluting the sample. For protein-rich samples, organic solvents are used for the dilutions in order to precipitate the proteins.⁴⁰ This overcomes the ion suppression caused by proteins. However, sample dilutions may result in concentrations of metabolites below detection limits and high abundant matrix compounds can, even after dilution, still cause severe ion suppression.³⁹

Previous studies have shown that salts and phospholipids are held responsible for a majority of signal suppression in blood analyses.^{41,42} Sample preparation offers great potential in removing these ion suppressors in a limited amount of time by making use of parallel executions or on-line hyphenation to MS analysis. The most commonly used sample preparation techniques are liquid-liquid extraction (LLE) and solid-phase extraction (SPE) and these techniques can be used to remove salts and phospholipids in a high-throughput fashion. LLE techniques have been used to separate polar and non-polar metabolites in biological samples. In these methods, a polar and non-polar fraction are obtained, which realizes the removal of phospholipids from polar metabolites.^{43,44,45} LLE extractions can be performed in parallel and coupled to FIA-MS to improve sample throughput. Solid-phase extractions can also be performed in parallel, but have also been coupled on-line to mass spectrometry in the RapidFire system.⁴⁶ The on-line elution of an SPE cartridge has been shown to be feasible within only 8.5 seconds of analysis time. Different SPE cartridges can be utilized to cover a wide range of metabolite classes.

Current high-throughput sample preparation techniques generally result in two fractions. Because only two fractions are obtained, the cleanup efficiency of these approaches remains rather limited. In high-throughput LLE applications, these fractions are directly measured by MS,

which does not allow for further within-fraction separation. The on-line elution of SPE cartridges also results in minimal within-fraction separation, because SPE sorbents generally have large particles. Therefore, compounds that are present in the same fraction will be simultaneously introduced into the mass spectrometer and still result in matrix effect. This problem can be overcome by the use of a more comprehensive and high-performance sample preparation module. Multiple serially coupled SPE columns can be used to achieve a wider coverage of metabolites in a single analysis. High performance columns can be implemented to establish within-fraction separation in a limited amount of time, because the smaller particle size results in a more efficient separation. The implementation of these columns is hardly at the expense of analysis time when combined with fast solvent switches. To maximize the throughput further, short high performance columns can be used allowing for shorter elution steps.

Mass spectrometry identification

Throughput is also a pressing issue in untargeted MS analyses. With regards to analysis time, feature correlation and identification are ideally performed within the same analysis. This means that the exact mass and fragmentation pattern of all analytes should be captured in one analysis. Low-resolution instruments, e.g. QQQ, as well as high-resolution instruments, e.g. Q-ToF and Orbitrap, are able to acquire fragmentation data. All these techniques make use of a low-resolution precursor ion selection followed by CID fragmentation and subsequent detection of product ions. The rate at which these MS/MS scans are acquired is limited by the scan time of the MS. Scan times as low as 30 msec have been used for quantitative and qualitative purposes for high-resolution mass spectrometry.⁴⁷ However, when a mass range of 500 m/z has to be fragmented at this scan time, the cycle time of one MS/MS scan already exceeds 15 seconds (including switching time of the Q1). A substantial amount of LC peak widths are smaller than 15 seconds, which means that a substantial amount of compounds will not be fragmented using this approach.⁴⁸

To overcome this issue, more sophisticated fragmentation techniques have been developed. These techniques comprise data dependent acquisition (DDA) and data independent acquisition (DIA) (see Figure 5). In DDA analyses, precursor ions are selected based on abundance or an inclusion list (combined with an exclusion list) and subsequently fragmented.⁴⁹ An advantage of this approach is that there is no time wasted on the fragmentation of m/z windows that do not contain any analytes. However, since fragmentation time is limited by the peak width and MS scan time, there is a high possibility that lower abundant analytes are not selected for fragmentation. Therefore, the coverage of this fragmentation is limited, while a wide coverage is key in untargeted analyses.

A more comprehensive fragmentation technique is called DIA. DIA ensures the fragmentation of all analytes within a certain mass range. One type of DIA is MS^{ALL} (also known as all-ion

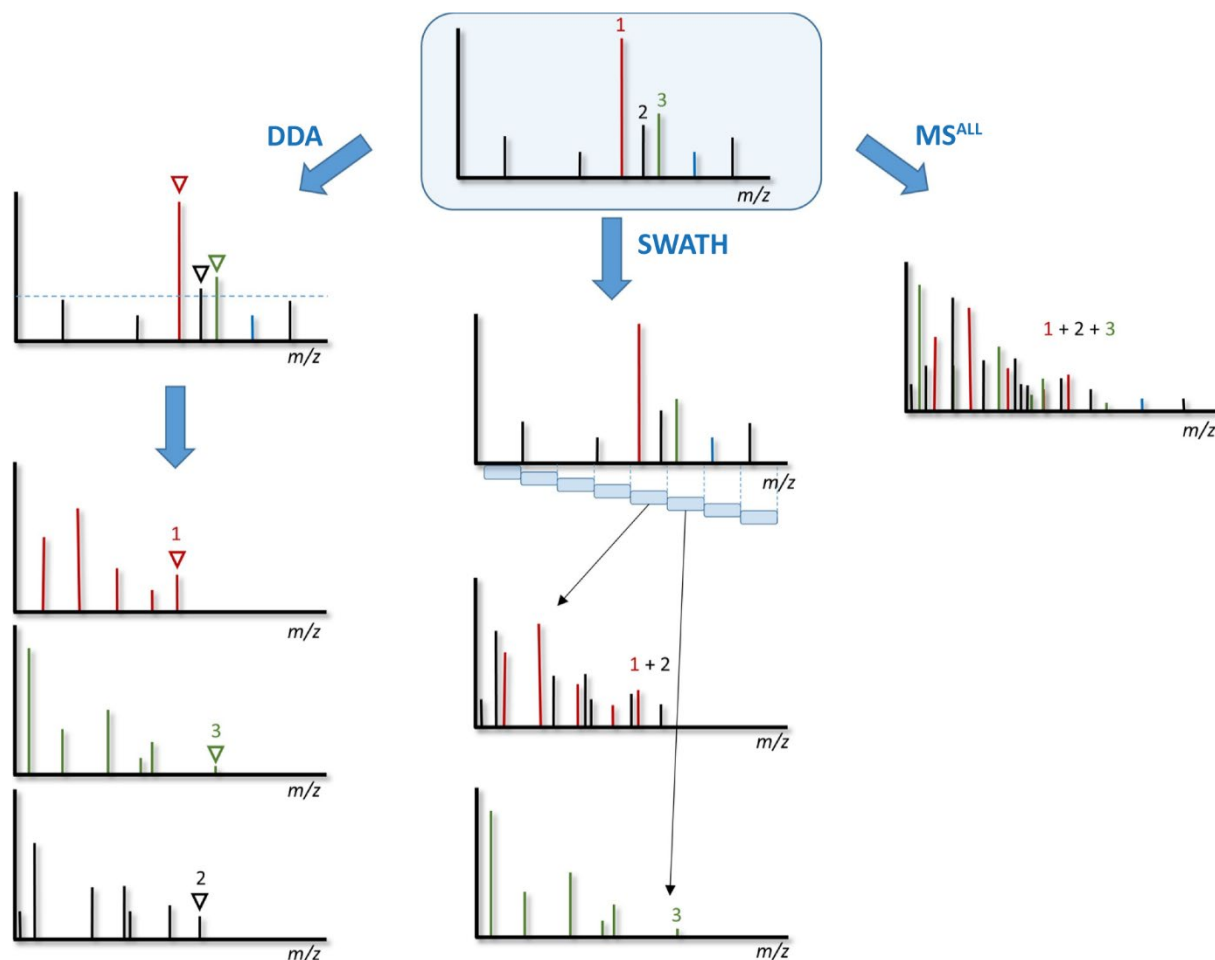


Figure 5. MS/MS procedures for data-dependent acquisition and data-independent acquisition, i.e. SWATH and MS^{ALL}. Adapted from Fenaille et al.²⁵

fragmentation or MS^E), in which a whole mass range is fragmented in one MS/MS window.^{50,51} Although all masses are fragmented, the selectivity of this technique is rather limited in complex samples, because it easily results in overlapping product ions.⁵² Moreover, the formation of product ions from multiple precursor ions may cause very complex MS/MS spectra, which limits its usefulness for compound identification.⁵³ A more selective type of DIA is called sequential window acquisition of all theoretical masses (SWATH).⁵⁴ In SWATH, the mass range is divided into several MS/MS windows (5-50 Dalton), which are fragmented consecutively. The window width is set based on the mass range, peak width and MS scan time. The SWATH windows have a fixed width or variable width that is dependent on the density of precursor ions. Because the number of precursor ions per MS/MS scan is smaller in SWATH relatively to MS^{ALL}, it is also expected that SWATH MS/MS spectra are more useful for quantification and identification. The identification power of SWATH has already been shown in several publications.^{55,56,57}

Although SWATH overcomes part of the shortcomings of MS^{ALL}, the fact remains that multiple precursor ions are fragmented simultaneously, which allows for the possibility of product ion

overlap and complex MS/MS spectra. Up until now, platforms involving SWATH identification were only used in combination with conventional liquid-chromatography separations. Therefore, the identification performance in high-throughput platforms remains unknown. Fast chromatography methods generally result in more coelution and smaller peak widths, which both increase the number of precursor ions per MS/MS window. Moreover, the quantitative performance of SWATH in both conventional and high-throughput platforms has never been addressed. Therefore, a more extensive evaluation of the versatility of SWATH in terms of metabolite identification and quantification is needed. If the performance proves to be sufficient, SWATH offers a great opportunity in combining untargeted analyses, identification and quantification in a single platform.

Nuclear magnetic resonance

The most well-known challenges in NMR analyses are the limited sensitivity and signal overlap.²² NMR analyses are generally 10 to 100 times or more less sensitive than MS analyses and allow for a substantially lower metabolic coverage in biological samples.⁵⁸ Therefore, NMR analyses are mostly performed for metabolites which are expected to be present at a concentration of 1 μM or higher. Moreover, signal overlap is an often-occurring problem in NMR-based metabolomics as many metabolites consist of similar functional groups. This can be problematic for quantification purposes as some peaks are not fully resolved which prevents proper peak integration. In addition, low abundant molecules can be overshadowed by high abundant molecules, which decreases the metabolic coverage and prevents metabolites from being identified.

Researchers have tried to overcome the challenges of sensitivity and peak overlap to make full use of the advantages of NMR analyses. The sensitivity of NMR has been improved by the use of higher field strengths (up to 1.2 GHz), cryogenically cooled probes and hyperpolarization.^{59,60,61} However, these technological advancements are at the expense of instrumental costs. A more cost-efficient solution to improve NMR sensitivity is the use of microcoils. The volume of microcoils goes down to the microliter/submicroliter range and drastically improves the mass sensitivity of NMR analyses.⁶² This allows for more sensitive analyses of pre-concentrated or mass limited samples.

Although these technological advancements improve the sensitivity, the complexity of biological samples still causes peak overlap and lower abundant molecules can still be 'hidden' behind higher abundant molecules. This is exemplified by the overlap of trimethylamine-N-oxide (TMAO), betaine and glucose signals in the NMR analysis of a blood sample (Figure 6). TMAO and betaine are potential biomarkers for cardiovascular disease.⁶³ However, at physiological pH, the peaks of these potential biomarkers overlap with each other as well as with the highly abundant glucose signals.⁶⁴ Therefore, accurate quantification of these potential biomarkers remains

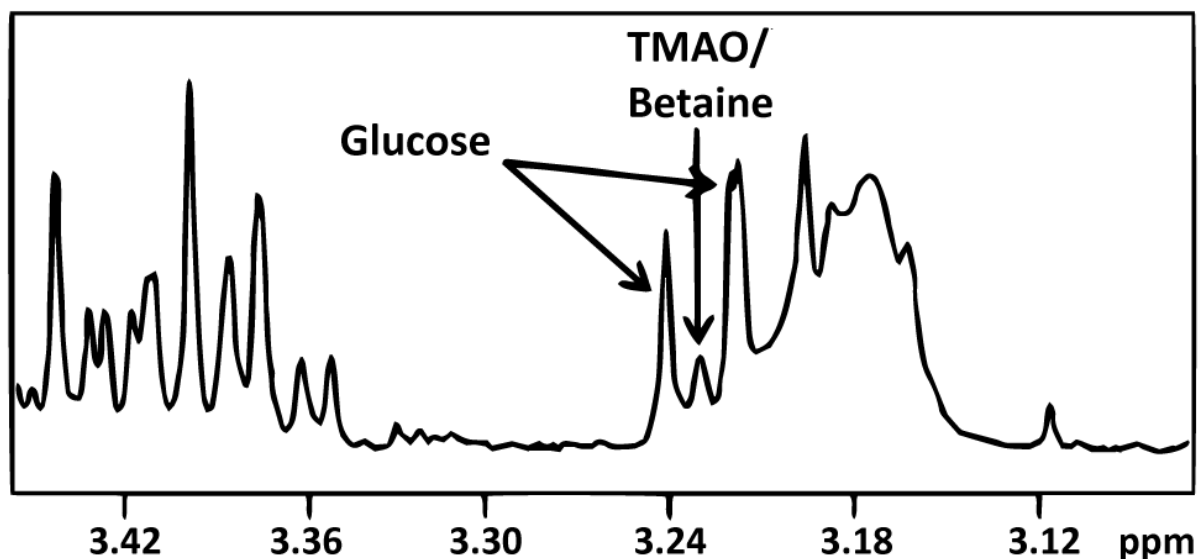


Figure 6. The analysis of a blood sample at physiological pH using a clinical ^1H NMR analyzer (400 MHz). The NMR spectra shows overlap of TMAO, betaine and glucose. Adapted from Garcia et al.⁶⁴

challenging using NMR. This is a classic example why improvements are also needed in the sample preparation methods that take place before NMR analysis. Fractionation provides a promising approach to lower the complexity of biological samples by dividing a sample into cleaner fractions. Chromatographic separations are often used to accomplish this.^{65,66} Conventionally, one dimensional chromatography is used for the cleanup of analytes before NMR analysis with collection of fractions of the eluent. More recently, also two-dimensional chromatography has been applied to improve the cleanup efficiency.^{67,68} However, its use has been rather limited and the potential has not been fully investigated.

The power of two-dimensional chromatography is highly dependent on the orthogonality between the two chromatographic separations. This orthogonality is determined by the selection of chromatography phases and by the chemical properties of the analytes.⁶⁹ Therefore, the most optimal orthogonal combination of chromatography columns is highly dependent of the sample mixture. An ideal platform should be able to adapt to differences in sample mixtures. However, a fixed combination of chromatography phases is commonly used, which offers no flexibility. This shortcoming can be overcome by the use of a directed two-dimensional chromatography, which is driven by the chemical properties of the analytes. After the first dimension, an orthogonality assessment could be made for each fraction, which can be used to select the most orthogonal second dimension. This will ensure that every analyte is fractionated by a combination of chromatography phases that is most powerful in terms of cleanup efficiency. As a result, highly purified features can be obtained, which facilitates the identification.

SCOPE OF THIS THESIS

The aim of this thesis is to accelerate metabolic profiles by enhancing the throughput of metabolite quantifications and improving the confidence of metabolite identifications. The challenge of achieving fast quantification in metabolomics is the presence of severe matrix effects during the MS analysis of complex samples. Complex samples also result in challenges during metabolite identification as complex MS/MS spectra and peak overlap in ^1H NMR complicate structure elucidation. The goal of this thesis is to tackle these challenges by the development and application of innovative fractionation approaches and state-of-the-art MS and NMR analyses. Fast and high performance fractionation strategies are used to decrease matrix effect during MS analysis. This prevents the need for extensive chromatographic separations, which is beneficial for sample throughput. A slower but more in-depth fractionation is realized by a directed two-dimensional chromatography method, which is adaptable to chemical differences of unknown compounds and complex samples. This fractionation allows for the efficient purification of unknown features, which is highly advantageous for NMR identification.

The aim of **Chapter 2** was to develop a high throughput platform for the quantification of gut metabolites in blood by MS. These metabolites have been linked to an increased risk for cardiovascular diseases and, therefore, there is an increased demand for the analysis of these potential biomarkers. A high-throughput analysis was achieved by fractionating blood samples using reversed phase chromatography. The use of fast solvent switches instead of a time-consuming gradient allowed for an analysis time of only 3 minutes. The LC column was unconventionally used to trap known ion suppressors leaving the analytes in a cleaned flow-through. It was shown that phospholipids caused more ion suppression than salts, and that the removal of phospholipids resulted in substantially lower ion suppression during the analysis of the polar gut metabolites.

The platform in **chapter 2** allowed for the quantification of five metabolites. The aim of **chapter 3** was to extend the fractionation approach in order to quantify 50 chemically diverse metabolite biomarkers using a multi-dimensional fractionation. Reversed phase and ion exchange chromatography were used for a fractionation based on polarity and charge, which allocated phospholipids and salts over different fractions minimizing their adverse effect during ESI. Matrix effect was further reduced by the use of high performance columns, which resulted in a better peak shape. This allowed for the separation of phospholipids from other metabolites in the apolar fraction. In comparison with a flow injection analysis, ion suppression decreased substantially by the fractionation approach. Although multiple columns were serially coupled, the analysis time was only 3 minutes per MS polarity because of the implementation of short columns and fast solvent switches.

The analysis of **chapter 3** allowed for a comprehensive targeting of the metabolome, but lacked isomeric separation. Isomers can be distinguished, however, by the use of fragmentation (in case of structural isomers). Data-independent acquisition is a promising methodology to acquire the exact mass and fragmentation data of all features in a single analysis. Currently, the use of DIA has been limited to the identification of unknown compounds. In **chapter 4**, the aim was to develop an untargeted analysis platform using DIA that allowed for both the identification of metabolites as well as the quantification of chromatographically unresolved isomers. The combination of conventional chromatography and variable SWATH allowed for the accurate quantification of structural isomers and the correct structure annotation of metabolites. The use of variable SWATH clearly outperformed other DIA strategies, i.e. fixed SWATH and MS^{ALL}, which often resulted in product ion overlap and complex MS/MS spectra. The combination of conventional chromatography and variable SWATH offers a great potential for combined identification and quantification platforms.

The DIA methodologies of **chapter 4** only allow for an identification based on spectral overlap of an unknown compound with a spectral MS library. When no spectral library hits can be found or when a higher identification confidence is necessary, the MS identification can be complemented by NMR. Therefore, the aim of **chapter 5** was to develop an identification workflow including MS and NMR. We have developed a directed two-dimensional chromatography fractionation in order to efficiently purify unknown compounds from a complex matrix prior to NMR analysis. The fractionation platform includes an orthogonality assessment for the second dimension, which ensures the most efficient combination of chromatography types per unknown compound and complex sample in terms of isolation efficiency. The platform resulted in the purification of five taste-related features in soy sauce. The unknown features were tentatively identified by MS and the structure was confirmed by the NMR analysis of the purified features.

In **chapter 6**, a general conclusion and perspective of the reported studies is provided. The overall performance improvements in terms of metabolite quantification and identification are discussed. The chapter is finalized with suggestions and future directions for advanced metabolomics analyses.

REFERENCES

1. Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–269 (2012).
2. Shin, S. Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).
3. Wild, C. P. The exposome: From concept to utility. *Int. J. Epidemiol.* **41**, 24–32 (2012).
4. Ramírez, M., Amate, L. & Gil, A. Absorption and distribution of dietary fatty acids from different sources. *Early Hum. Dev.* **65**, 95–101 (2001).

5. Xu, T. *et al.* Effects of smoking and smoking cessation on human serum metabolite profile: Results from the KORA cohort study. *BMC Med.* **11**, 1–14 (2013).
6. Everett, J. R. Pharmacometabonomics in humans: A new tool for personalized medicine. *Pharmacogenomics* **16**, 737–754 (2015).
7. Beger, R. D. *et al.* Metabolomics enables precision medicine: “A White Paper, Community Perspective”. *Metabolomics* **12**, (2016).
8. Kohler, I., Hankemeier, T., van der Graaf, P. H., Knibbe, C. A. J. & van Hasselt, J. G. C. Integrating clinical metabolomics-based biomarker discovery and clinical pharmacology to enable precision medicine. *Eur. J. Pharm. Sci.* **109**, S15–S21 (2017).
9. Wilcken, B. & Wiley, V. Newborn screening. *Pathology* **40**, 104–115 (2008).
10. Ellero-Simatos, S. *et al.* Pharmacometabolomics reveals that serotonin is implicated in aspirin response variability. *CPT Pharmacometrics Syst. Pharmacol.* **3**, 1–9 (2014).
11. Trivedi, D. K., Hollywood, K. A. & Goodacre, R. Metabolomics for the masses: The future of metabolomics in a personalized world. *New Horizons Transl. Med.* **3**, 294–305 (2017).
12. Wang, T. J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nat. Med.* **17**, 448–453 (2011).
13. Huang, G. *et al.* Metabolomic evaluation of the response to endocrine therapy in patients with prostate cancer. *Eur. J. Pharmacol.* **729**, 132–137 (2014).
14. Kim, S., Kim, J., Yun, E. J. & Kim, K. H. Food metabolomics: From farm to human. *Curr. Opin. Biotechnol.* **37**, 16–23 (2016).
15. Simó, C., Ibáñez, C., Valdés, A., Cifuentes, A. & García-Cañas, V. Metabolomics of genetically modified crops. *Int. J. Mol. Sci.* **15**, 18941–18966 (2014).
16. Carreno-Quintero, N., Bouwmeester, H. J. & Keurentjes, J. J. B. Genetic analysis of metabolome-phenotype interactions: From model to crop species. *Trends Genet.* **29**, 41–50 (2013).
17. Xia, J., Broadhurst, D. I., Wilson, M. & Wishart, D. S. Translational biomarker discovery in clinical metabolomics: An introductory tutorial. *Metabolomics* **9**, 280–299 (2013).
18. Morales-Soriano, E. *et al.* Flavor characterization of native Peruvian chili peppers through integrated aroma fingerprinting and pungency profiling. *Food Res. Int.* **109**, 250–259 (2018).
19. Lindemann, B. The Discovery of Umami. *Chem. Senses* **27**, 843–844 (2002).
20. Cai, Y. & Zhu, Z. J. A high-throughput targeted metabolomics workflow for the detection of 200 polar metabolites in central carbon metabolism. *Methods Mol. Biol.* **1859**, 263–274 (2019).
21. van der Laan, T. *et al.* Fast LC-ESI-MS/MS analysis and influence of sampling conditions for gut metabolites in plasma and serum. *Sci. Rep.* **9**, 12370 (2019).
22. Markley, J. L. *et al.* The future of NMR-based metabolomics. *Curr. Opin. Biotechnol.* **43**, 34–40 (2017).
23. Roberts, L. D., Souza, A. L., Gerszten, R. E. & Clish, C. B. Targeted metabolomics. *Curr Protoc Mol Biol.* 1–34 (2012). doi:10.1002/0471142727.mb3002s98.Targeted
24. Kapoore, R. V. & Vaidyanathan, S. Towards quantitative mass spectrometry-based metabolomics in microbial and mammalian systems. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **374**, (2016).
25. Fenaille, F., Barbier Saint-Hilaire, P., Rousseau, K. & Junot, C. Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand? *J. Chromatogr. A* **1526**, 1–12 (2017).
26. Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D. & McLean, J. A. Untargeted Metabolomics Strategies—Challenges and Emerging Directions. *J. Am. Soc. Mass Spectrom.* **27**, 1897–1905 (2016).
27. Holmes, E. *et al.* Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature* **453**, 396–400 (2008).
28. Bingol, K. & Brüscheiler, R. Multidimensional APPROACHES to NMR-based metabolomics. *Anal. Chem.* **86**, 47–57 (2014).
29. Cui, Q. *et al.* Metabolite identification via the Madison Metabolomics Consortium Database

- [3]. *Nat. Biotechnol.* **26**, 162–164 (2008).
30. Ulrich, E. L. *et al.* BioMagResBank. *Nucleic Acids Res.* **36**, 402–408 (2008).
31. Wishart, D. S. *et al.* HMDB: A knowledgebase for the human metabolome. *Nucleic Acids Res.* **37**, 603–610 (2009).
32. Robinette, S. L., Zhang, F., Brüscheweiler-Li, L. & Brüscheweiler, R. Web server based complex mixture analysis by NMR. *Anal. Chem.* **80**, 3606–3611 (2008).
33. Rathahao-Paris, E., Alves, S., Junot, C. & Tabet, J. C. High resolution mass spectrometry for structural identification of metabolites in metabolomics. *Metabolomics* **12**, 1–15 (2016).
34. Chin, C., Zhang, Z. P. & Karnes, H. T. A study of matrix effects on an LC/MS/MS assay for olanzapine and desmethyl olanzapine. *J. Pharm. Biomed. Anal.* **35**, 1149–1167 (2004).
35. Furey, A., Moriarty, M., Bane, V., Kinsella, B. & Lehane, M. Ion suppression; A critical review on causes, evaluation, prevention and applications. *Talanta* **115**, 104–122 (2013).
36. Gowda, G. A. N. & Djukovic, D. Overview of Mass Spectrometry-Based Metabolomics: Opportunities and Challenges. *Methods Mol Biol* **1198**, 3–12 (2014).
37. Xiao, J. F., Zhou, B. & Resson, H. W. Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *TrAC - Trends Anal. Chem.* **32**, 1–14 (2012).
38. Zampieri, M., Sekar, K., Zamboni, N. & Sauer, U. Frontiers of high-throughput metabolomics. *Curr. Opin. Chem. Biol.* **36**, 15–23 (2017).
39. Nanita, S. C. & Kaldon, L. G. Emerging flow injection mass spectrometry methods for high-throughput quantitative analysis. *Anal. Bioanal. Chem.* **408**, 23–33 (2016).
40. Souverain, S., Rudaz, S. & Veuthey, J. L. Protein precipitation for the analysis of a drug cocktail in plasma by LC-ESI-MS. *J. Pharm. Biomed. Anal.* **35**, 913–920 (2004).
41. Ghosh, C., Shinde, C. P. & Chakraborty, B. S. Influence of ionization source design on matrix effects during LC-ESI-MS/MS analysis. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **893–894**, 193–200 (2012).
42. Truffelli, H., Palma, P., Famigliani, G. & Cappiello, G. An overview of matrix effects in liquid chromatography-mass spectrometry. *Indian J. Exp. Biol.* **30**, 491–509 (2011).
43. Bligh, E.G. and Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, (1959).
44. Folch, J., Lees, M. & Sloane Stanley, G. H. A simple method for the isolation and purification of total lipides from animal tissues. *J. Biol. Chem.* **226**, 497–509 (1956).
45. Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A. & Schwudke, D. Lipid extraction by methyl-terf-butyl ether for high-throughput lipidomics. *J. Lipid Res.* **49**, 1137–1146 (2008).
46. Zhang, X. *et al.* SPE-IMS-MS: An automated platform for sub-sixty second surveillance of endogenous metabolites and xenobiotics in biofluids. *Clin. Mass Spectrom.* **2**, 1–10 (2016).
47. Dubbelman, A. C. *et al.* Mass spectrometric recommendations for Quan/Qual analysis using liquid-chromatography coupled to quadrupole time-of-flight mass spectrometry. *Anal. Chim. Acta* **1020**, 62–75 (2018).
48. Hsieh, E. J., Bereman, M. S., Durand, S., Valaskovic, G. A. & MacCoss, M. J. Effects of Column and Gradient Lengths on Peak Capacity and Peptide Identification in nanoflow LC-MS/MS of Complex Proteomic Samples. *J. Am. Soc. Mass Spectrom.* **24**, 148–153 (2013).
49. Stahl, D. C., Swiderek, K. M., Davis, M. T. & Lee, T. D. Data-controlled automation of liquid chromatography/tandem mass spectrometry analysis of peptide mixtures. *J. Am. Soc. Mass Spectrom.* **7**, 532–540 (1996).
50. Navarro-Reig, M. *et al.* Metabolomic analysis of the effects of cadmium and copper treatment in: *Oryza sativa* L. using untargeted liquid chromatography coupled to high resolution mass spectrometry and all-ion fragmentation. *Metallomics* **9**, 660–675 (2017).
51. Plumb, R. S. *et al.* UPLC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation Robert. *Rapid Commun. Mass Spectrom.* 4129–4138 (2008). doi:10.1002/rcm.2550
52. Bilbao, A. *et al.* Processing strategies and software solutions for data-independent acquisition in mass spectrometry. *Proteomics* **15**, 964–980 (2015).
53. Zhu, X., Chen, Y. & Subramanian, R. Comparison of information-dependent acquisition,

- SWATH, and MS All techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *Anal. Chem.* **86**, 1202–1209 (2014).
54. Bonner, R. & Hopfgartner, G. SWATH acquisition mode for drug metabolism and metabolomics investigations. *Bioanalysis* **8**, 1735–1750 (2016).
 55. Collins, B. C. *et al.* Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat. Commun.* **8**, 1–11 (2017).
 56. Roemmelt, A. T., Steuer, A. E., Poetzsch, M. & Kraemer, T. Liquid chromatography, in combination with a quadrupole time-of-flight instrument (LC QTOF), with sequential window acquisition of all theoretical fragment-ion spectra (SWATH) acquisition: Systematic studies on its use for screenings in clinical and foren. *Anal. Chem.* **86**, 11742–11749 (2014).
 57. Scheidweiler, K. B., Jarvis, M. J. Y. & Huestis, M. A. Nontargeted SWATH acquisition for identifying 47 synthetic cannabinoid metabolites in human urine by liquid chromatography-high-resolution tandem mass spectrometry. *Anal. Bioanal. Chem.* **407**, 883–897 (2015).
 58. Emwas, A. H. *et al.* Nmr spectroscopy for metabolomics research. *Metabolites* **9**, (2019).
 59. Moser, E., Laistler, E., Schmitt, F. & Kontaxis, G. Ultra-high field NMR and MRI-the role of magnet technology to increase sensitivity and specificity. *Front. Phys.* **5**, 1–15 (2017).
 60. Jézéquel, T. *et al.* Absolute quantification of metabolites in tomato fruit extracts by fast 2D NMR. *Metabolomics* **11**, 1231–1242 (2015).
 61. Nikolaou, P., Goodson, B. M. & Chekmenev, E. Y. NMR Hyperpolarization Techniques for Biomedicine. *Chemistry (Easton)*. **21**, 3156–3166 (2015).
 62. Grimes, J. H. & O’Connell, T. M. The application of micro-coil NMR probe technology to metabolomics of urine and serum. *J. Biomol. NMR* **49**, 297–305 (2011).
 63. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
 64. Garcia, E. *et al.* NMR quantification of trimethylamine-N-oxide in human serum and plasma in the clinical laboratory setting. *Clin. Biochem.* **50**, 947–955 (2017).
 65. Brennan, L. NMR-based metabolomics: From sample preparation to applications in nutrition research. *Prog. Nucl. Magn. Reson. Spectrosc.* **83**, 42–49 (2014).
 66. Van Duynhoven, J. P. M. & Jacobs, D. M. Assessment of dietary exposure and effect in humans: The role of NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **96**, 58–72 (2016).
 67. Jiao, L. *et al.* Preparative isolation of flavonoid glycosides from *Sphaerophysa salsula* using hydrophilic interaction solid-phase extraction coupled with two-dimensional preparative liquid chromatography. *J. Sep. Sci.* **40**, 3808–3816 (2017).
 68. Wang, W. *et al.* Efficient separation of high-purity compounds from *Oxytropis falcata* using two-dimensional preparative chromatography. *J. Sep. Sci.* **40**, 3593–3601 (2017).
 69. Bassanese, D. N. *et al.* Protocols for finding the most orthogonal dimensions for two-dimensional high performance liquid chromatography. *Talanta* **134**, 402–408 (2015).
 70. González, O. *et al.* Matrix Effect Compensation in Small-Molecule Profiling for an LC-TOF Platform Using Multicomponent Postcolumn Infusion. *Anal. Chem.* **87**, 5921–5929 (2015).

SUPPLEMENTARY INFORMATION

Table S1. The search terms that were used for the PubMed literature survey of Figure 1.

Analytical technique	Search term
Total number	(metabonom*[Title/Abstract]) OR metabolom*[Title/Abstract])
LC-MS	(metabonom*[Title/Abstract]) OR metabolom*[Title/Abstract]) AND ((*LC-MS*[Title/Abstract] OR *LC/MS*[Title/Abstract]) OR (*liquid*[Title/Abstract] AND *chromatography*[Title/Abstract] AND *mass*[Title/Abstract] AND *spectrometry*[Title/Abstract]))
GC-MS	(metabonom*[Title/Abstract]) OR metabolom*[Title/Abstract]) AND ((*GC-MS*[Title/Abstract] OR *GC/MS*[Title/Abstract]) OR (*gas*[Title/Abstract] AND *chromatography*[Title/Abstract] AND *mass*[Title/Abstract] AND *spectrometry*[Title/Abstract]))
NMR	(metabonom*[Title/Abstract]) OR metabolom*[Title/Abstract]) AND ((*NMR*[Title/Abstract]) OR (*nuclear*[Title/Abstract] AND *magnetic*[Title/Abstract] AND *resonance*[Title/Abstract]))

Chapter 2

Fast LC-ESI-MS/MS analysis and influence of sampling conditions for gut metabolites in plasma and serum

Based on:

Tom van der Laan, Tim Kloots, Marian Beekman, Alida Kindt, Anne-Charlotte Dubbelman, Amy Harms, Cornelia M van Duijn, P. Eline Slagboom, Thomas Hankemeier

Fast LC-ESI-MS/MS analysis and influence of sampling conditions for gut metabolites in plasma and serum

Scientific Reports (2019)

ABSTRACT

In the past few years, the gut microbiome has been shown to play an important role in various disorders including in particular cardiovascular diseases. Especially the metabolite trimethylamine-N-oxide (TMAO), which is produced by gut microbial metabolism, has repeatedly been associated with an increased risk for cardiovascular events. Here we report a fast liquid chromatography tandem mass spectrometry (LC-MS/MS) method that can analyze the five most important gut metabolites with regards to TMAO in three minutes. Fast liquid chromatography is unconventionally used in this method as an on-line cleanup step to remove the most important ion suppressors leaving the gut metabolites in a cleaned flow-through fraction, also known as negative chromatography. We compared different blood matrix types to recommend best sampling practices and found citrated plasma samples demonstrated lower concentrations for all analytes and choline concentrations were significantly higher in serum samples. We demonstrated the applicability of our method by investigating the effect of a standardized liquid meal (SLM) after overnight fasting of 25 healthy individuals on the gut metabolite levels. The SLM did not significantly change the levels of gut metabolites in serum.

INTRODUCTION

The gut microbiome has recently been shown to play an important role in cardiovascular disease (CVD) and various other disorders.^{1,2,3} Especially the gut metabolite trimethylamine-N-oxide (TMAO) has drawn a lot of attention, as it proved to be an important biomarker for CVD.^{1,4,5} A schematic overview of the TMAO biosynthesis is depicted in Figure 1. Wang *et al.* studied the influence of gut metabolites and the gut microbiome on CVD by comparing the metabolic profile between healthy volunteers and patients who had suffered a heart attack.⁶ Not only TMAO, but also its precursor choline was found to be significantly correlated with an increased risk for CVD.

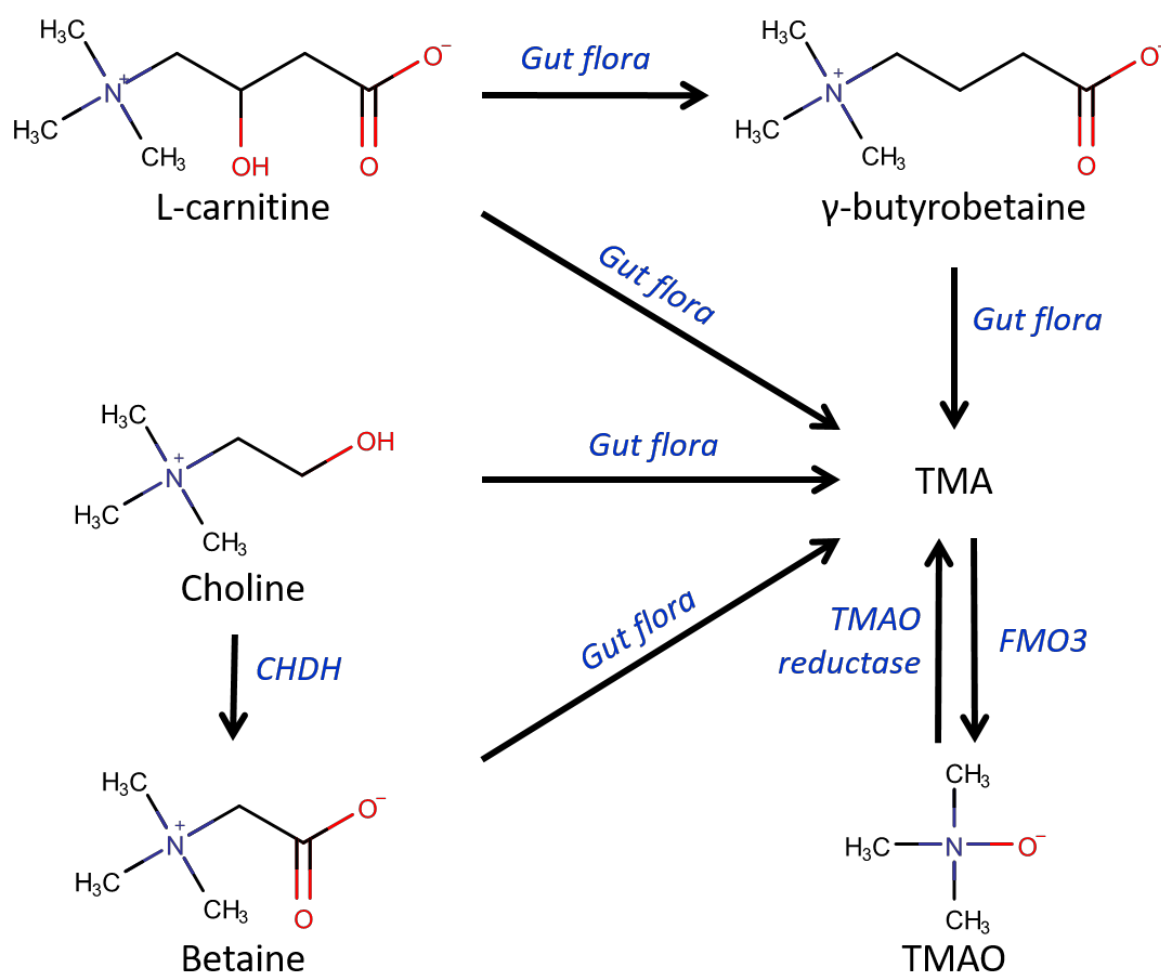


Figure 1. TMAO biosynthesis. L-carnitine and choline are derived from the diet and are endogenously synthesized. L-carnitine can be converted by the gut flora to TMA directly or with γ -butyrobetaine as an intermediate. Choline is converted to betaine by choline dehydrogenase (CHDH), with highest expression in the renal cortex. Betaine, as well as choline are a substrate for the gut flora to form TMA. TMA is converted to TMAO by flavin-containing monooxygenase 3 (FMO3) and TMAO is reduced to TMA by TMAO reductase. This figure has been adapted from the literature.^{2,6,7,9,11}

In the following years, more research focused on elucidating the role of TMAO and the gut microbiome in CVD. The clinical relevance of TMAO was emphasized by a positive correlation between plasma levels of TMAO and major cardiovascular events (death, myocardial infarction, or stroke) in 4000 patients.³ In addition to TMAO formation, choline is oxidized into betaine, which is mostly occurring in the renal cortex.⁵ Betaine has also been reported as a biomarker for CVD.⁷ L-carnitine, like choline and betaine, is a substrate for the gut microbiome to produce trimethylamine (TMA), a precursor of TMAO.^{2,8} Koeth *et al.* showed that L-carnitine can also contribute to TMAO production via a distinct route. They demonstrated that the majority of dietary L-carnitine is converted to γ -butyrobetaine (GBB) instead of TMA.⁹ GBB is converted into TMAO via TMA and accelerates atherosclerosis in mice. L-carnitine and choline are mostly derived from diet and are also endogenously synthesized.^{2,10} These metabolites are the main source of the TMAO biosynthesis. L-carnitine, GBB, choline and betaine are all substrates for the gut microbiota to synthesize the precursor of TMAO: TMA.¹¹ Therefore, targeting these metabolites with the end product TMAO in a single analytical platform, can provide important information about the TMAO biosynthesis pathway in relation to cardiovascular diseases.

Techniques that have been used to analyze gut metabolites are proton nuclear magnetic resonance (¹H NMR) and mass spectrometry. There are only a few studies that analyzed TMAO with ¹H NMR.^{12,13} This is probably due to the difficulties that are faced during ¹H NMR analysis of the TMAO metabolism, such as the inability to distinguish betaine signals from TMAO signals at pH values above 5.8¹⁴ and the overlap of glucose signals with TMAO, betaine and choline¹⁵. The latter is especially problematic for plasma since glucose is highly abundant in this matrix. Combined with the limited sensitivity of ¹H NMR, one can expect that mass spectrometry (MS) is the method of choice for the analysis of gut metabolites in biological fluids.

In order to increase throughput, flow injection analysis (FIA) and direct injection (DI) MS methods have been developed to measure gut metabolites.^{16,17} Although the analysis time is minimized in these studies, ion suppression and sensitivity are not addressed in the method validation. This makes it difficult to assess to what extent the matrix is interfering during these analyses. Moreover, these methods were only used to measure urine. Therefore, it is questionable whether these methods are suitable for analyzing gut metabolites in plasma as this is a more complex matrix due to the presence of proteins and lipids.^{18,19,20} Over the last five years, numerous analytical methods have been developed to study gut metabolites in plasma with liquid chromatography (LC).^{2,3,5,21,22,23,24,25,26,27,28,29} Both normal phase and reversed phase stationary phases have been used for the chromatographic separation. Fast hydrophilic interaction liquid chromatography (HILIC) analysis has been used to analyze TMA and TMAO in plasma in 2.5 minutes.²⁶ However, TMAO levels alone do not provide a complete picture to determine the

increased risk for CVD after choline and L-carnitine ingestion. Recently, methods have been developed that included TMAO, betaine, choline and carnitine in their target analytes.^{24,28,29} However, the throughput of the methods is compromised since the analysis time is 10 minutes or more. To our knowledge, no method has been published yet targeting the full TMAO biosynthesis pathway, as depicted in Figure 1, in a high-throughput fashion.

In this study, a fast LC-ESI-MS/MS method was developed and validated for the analysis of the gut metabolites L-carnitine, GBB, choline and betaine and TMAO in plasma. An analytical column with a C18 stationary phase was used to realize an on-line cleanup of the sample before introducing it to the MS. This stationary phase does not retain the analytes but does retain important matrix interferences, such as (phospho-)lipids. This type of chromatography is also referred to as negative chromatography.³⁰ As a result, a fast analysis of the gut metabolites in a cleaned flow-through is possible using ESI-MS/MS. This allowed us to quantify the five gut metabolites with an analysis time of only three minutes. A main evaluation point of the method is the influence of the matrix components on the ionization process and means to correct for matrix effect. This is especially important for high-throughput LC-MS since these platforms often have less separation, which generally results in a negative effect on electrospray ionization. To stress the importance of the on-line cleanup, we compared our LC-method with a FIA-MS method. The major contributors to ion suppression were identified by evaluating signal suppression of known ion suppressors from literature. We also evaluated the influence of the blood collection procedures on the gut metabolite concentrations. This should give insight into the performance of the method in different blood matrices and whether it is possible to compare different matrices with each other. The method was shown to be applicable to blood as is illustrated in a challenge test with a choline-containing beverage. Here, gut metabolite levels were analyzed in fasted serum levels and compared with serum levels taken about 30 minutes after the consumption of a standardized liquid meal.

MATERIALS & METHODS

Chemicals

Water was obtained from an arium pro UF/VF water purification system with a Sartopore 2 0.2 µm filter (Sartorius Stedim, Rotterdam, The Netherlands). High performance liquid chromatography (HPLC) and ULC-MS grade methanol (MeOH) and HPLC grade acetonitrile (ACN) were purchased from Actua-all (Oss, The Netherlands). Formic acid 99% was purchased from Acros Organics (Geel, Belgium). 1,2-dinonadecanoyl-sn-glycero-3-phosphocholine (PC 19:0/19:0) was purchased from Avanti Polar Lipids, inc (Alabaster, USA). Sodium chloride originated from Sigma-Aldrich Chemie B.V. (Zwijndrecht, The Netherlands). The suppliers of the

standards and deuterated internal standards can be found in the Supplementary information (Table S1). Pooled citrated plasma (October 2016) was used for method development and was purchased from Sanquin (Amsterdam, The Netherlands).

Sample preparation

A 10 μL aliquot of plasma was transferred into a 1.5-mL Eppendorf tube and mixed with 10 μL of deuterated internal standards in MeOH. During method validation, 10 μL of calibration standards in MeOH were added to the sample. A final volume of 100 μL was reached by adding MeOH. The Eppendorf tubes were thoroughly mixed by vortex mixing them for 30 seconds. Then, the Eppendorf tubes were centrifuged for 5 min (15700 g at 4 °C). After centrifugation, 80 μL of the supernatant was collected and transferred into an autosampler vial containing a 150 μL insert.

LC-MS/MS

Samples were measured using a UPLC Agilent Infinity II (1290 Multisampler, 1290 Multicolumn Thermostat and 1290 High Speed Pump) (Agilent Technologies, Waldbronn, Germany) coupled to an AB SCIEX quadrupole-ion trap 6500 (QTRAP) (AB Sciex, Massachusetts, USA). The ionization source was the Turbo Spray Ion Drive and was set in positive mode with a capillary spray voltage set at 2500V. Declustering potential, Entrance potential collision cell exit potential curtain gas collision gas temperature (TEM) ion source gas 1 and ion source gas 2 were set at 70V, 10V, 10V, 20 psi, medium, 350 °C, 80 psi and 70 psi, respectively. The analyte and deuterated internal standard MRM transitions and the optimized collision energies are mentioned in the Supplementary information (Table S1). We have developed an LC and FIA method. The LC analyses were carried out using an AccQ-Tag™ Ultra C18 column (2.1 \times 100 mm, 1.7 μm). The aqueous mobile phase A consisted of 0.1% formic acid in water and mobile phase B consisted of 0.1% formic acid in ACN. The gradient started at 5% B and was held at this value for 0.8 min. The gradient increased linearly to 50% B in 0.05 min and to 100% B in an extra 0.10 min. The mobile phase composition was held at 100% B for 1.25 min before it returned to 5% B in 0.02 min. Finally, the gradient was kept at 5% B for 0.78 min to re-equilibrate the column. The total analysis time was 3 min. FIA was carried out using a mobile phase consisting of formic acid/MeOH/H₂O (0.1/90/10, v/v/v). The total analysis time was 1 minute. The injection volume was set at 1 μL and the flow rate at 700 $\mu\text{L}/\text{min}$ for both LC and FIA analyses. The raw data was analyzed using MultiQuant (AB SCIEX, Version 3.0.2)

Matrix effect evaluation

The matrix effect was determined by the ratio of the peak area of the deuterated internal standards in a plasma sample to the peak area of the deuterated internal standards in a water sample.

$$(1) \text{ Matrix effect} = \frac{\text{Peak area ISTD in matrix}}{\text{Peak area ISTD in academic}} \times 100\%$$

In addition, individual matrix components were added to a water sample to determine their corresponding matrix effect on the target analytes: plasma, physiological salt, physiological phosphatidylcholine (PC) and the combination of physiological salt and PC. The physiological salt and PC solution consisted of 154 mM NaCl and 2,8 mM 1,2-dinonadecanoyl-sn-glycero-3-phosphocholine (PC 19:0/19:0), respectively.

Calibration lines

The standards and their deuterated internal standards were weighed and dissolved in MeOH to reach a 1 mg/mL stock solution. The stock solutions were stored at – 80 °C. A calibration line was constructed by several dilutions of the standard stock solutions. Eight calibration concentrations (C8-C1) in methanol were used for each standard in which C8 contained the highest concentration and the subsequent concentrations were 1:1 dilutions of the previous concentration. The ninth calibration point did not contain the standards (C0). The concentration of the deuterated internal standards was set to mimic the physiological concentrations of the corresponding analytes in plasma and was added to all nine calibration standards. The concentrations of C8 and the deuterated internal standards are mentioned in the Supplementary information (Table S2). The calibration lines were constructed by plotting the ratio between the peak area of the standards and the peak area of their corresponding deuterated internal standards on the y-axis against the concentration of the analytes on the x-axis. Calibration lines were constructed in water samples and in plasma samples.

Method validation parameters

Method validation was performed in water, plasma and serum. Repeatability (N=5) was determined by the RSD of pooled citrated plasma samples. The intermediate precision (N=26) was determined by the RSD of pooled citrated plasma samples measured on three different days. Blank effect was calculated by the ratio of the peak area in blank samples (pure methanol, N=4) to the peak area in pooled serum samples (N=14).

$$(2) \text{ Blank effect} = \frac{\text{Peak area in blank samples}}{\text{Peak area in in pooled serum samples}} \times 100\%$$

Linearity, limit of detection (LOD) and lower limit of quantification (LLOQ) were determined using the calibration curve in water. The linearity of the calibration line was calculated by the correlation coefficient (R^2). LOD (μM) and LLOQ (μM) were calculated using the standard deviation (SD) of the peak area of C1 (three replicates) and the peak area of a blank. The response factor (RF) was calculated by dividing the peak area of C1 by the [C1].

$$(3) LOD = \frac{3 \times SD (\text{peak area } C1) + \text{peak area blank}}{RF\left(\frac{\text{peak area } C1}{[C1]}\right)}$$

$$(4) LLOQ = \frac{10 \times SD (\text{peak area } C1) + \text{peak area blank}}{RF\left(\frac{\text{peak area } C1}{[C1]}\right)}$$

Different blood collection procedures and freeze-thaw cycles

Fresh blood samples of five healthy volunteers were drawn. Every blood sample of each healthy volunteer was treated in four different ways within one hour. Plasma samples were prepared by collecting whole blood in EDTA, heparin or citrate containing tubes. Subsequently, the tubes were centrifuged at 2000g at 4°C for 10 min. The supernatant was collected. Serum samples were prepared by allowing the whole blood to clot at room temperature for 30 min. Afterwards, the samples were centrifuged at 2000g at 21°C for 10 min. The supernatant was collected. For the TMAO platform, 10 µL of each sample (N=3) was aliquoted immediately after plasma and serum were obtained. The aliquots were stored at -80°C, except for the fresh sample measurements (fresh). A freeze-thaw (FT) cycle was defined by freezing the samples at -80°C and having the samples thaw on crushed ice for 1 hour. Samples were stored at -80°C for 1 day, 2 days and 3.5 hours for FT1, FT2 and FT3, respectively. After thawing, the samples were stored in the freezer or directly prepared and analyzed. The third freeze-thaw cycle was also thawed at room temperature (FT3R). The data was normalized on the average concentration of the fresh heparin treated plasma. Every compound for each healthy volunteer was normalized separately. The statistical analysis of the different blood collection procedures was performed by a one-way ANOVA analysis. For this, we used the data from the first freeze-thaw cycle, because a blood sample is commonly kept at -80 °C prior to analysis.

Effect of a standardized liquid meal

The influence of a standardized liquid meal on the serum levels of the gut metabolites was evaluated in 25 older adults (11 women and 14 men with a mean age of 64.5 ± 5.2 years) participating in the Growing Old Together (GOTO) study.⁵³ Because the GOTO study only reports on the response to the GOTO lifestyle intervention study (baseline and follow-up), the challenge test with a standardized liquid meal (SLM) is described in this section. The first blood collection was in fasting status and took place between 8 and 9 am in the hospital after at least 10 h of fasting. The second blood collection was taken on average 33 min (SD ± 1 min) after a 200 mL SLM Nutridrink™ challenge between 9 am and 12 am on the same day. Nutridrink is a liquid oral nutritional supplement (Nutricia Advanced Medical Nutrition, Zoetermeer, The Netherlands; 1.5 kcal/mL (6.25 kJ/mL), 35 En% fat, 49 En% carbohydrates, 16En% protein).⁴⁸ The SLM contained 110 mg of choline and did not contain carnitine. The blood was collected in standard

coagulation tubes of which serum was collected after coagulation of the blood. Quality control (QC) samples were prepared by pooling 5 μ L of every individual study sample. In the batch design, a QC sample was analyzed after every tenth sample. Each QC sample and every seventh sample was measured twice. All tests were performed on the log₂ transformed absolute concentrations of the gut metabolites. A paired t-test was used to assess differences between the two time points. The response to the challenge test was calculated as the difference of the log₂ transformed non-fasted serum values and the log₂ transformed absolute fasted serum values. These differences were tested for an association to gender using an unpaired, standard t-test. Further, a heat map of these differences was constructed using hierarchical clustering. A multiple testing correction (Benjamini-Hochberg, <0.05) was used to correct the p-values for the false discovery rate (FDR). The described studies were approved by the Medical Ethical Committee of the Leiden University Medical Center and comply with relevant guidelines and regulations. Volunteer inclusion was based on informed consent.

RESULTS AND DISCUSSION

Method development

In the first stage of the method development, unique precursor-product ion transmissions were optimized for all standards and deuterated internal standards. Betaine-d₉ was chosen as a deuterated internal standard for betaine because we found background interferences in usable multiple reaction monitoring (MRM) transition for betaine-d₃ in plasma. In the LC-MS analysis, the most important MS variable was the collision energy, which was optimized to reach the highest sensitivity. The LC gradient started with 95% water to maximize the retention of the matrix interferences by the LC column. Formic acid was added to facilitate the ionization of our target analytes. Since it was expected that matrix interference would mainly occur by the presence of lipids, a C₁₈ column of 100 mm was used to provide sufficient load ability and retention of the matrix components. A shorter column would have improved the overall throughput slightly but could have also compromised the matrix removal. We did not use a time-consuming gradient because we did not aim for a chromatographic separation. Therefore, the mobile phase composition was almost directly switched to 100% organic solvent after the analysis of the analytes in order to wash the bound matrix components from the column before returning to the starting composition. The chromatograms demonstrated good peak shapes (Figure 2).

Figure 2 demonstrates that the signal of the gut metabolites is in the same order of magnitude and the gut metabolites are coeluting with their deuterated internal standards, which is required for absolute quantification. An interesting finding was the appearance of an extra peak in the chromatogram of GBB. The extra peak elutes earlier than the larger peak and is baseline

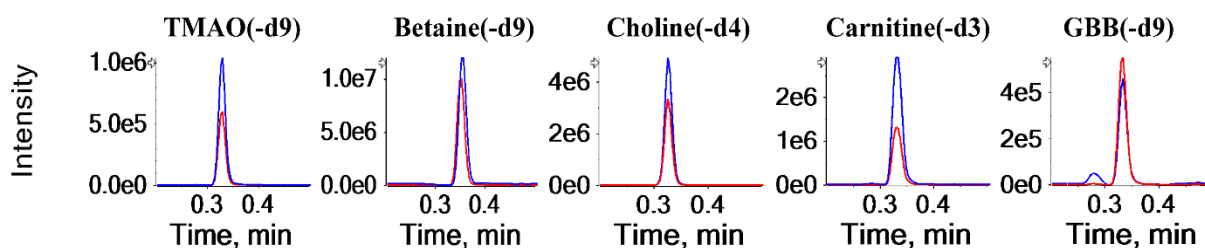


Figure 2. Extracted ion chromatograms of the LC-MS/MS analysis of the gut metabolites (blue line) and their deuterated internal standards (red line). There are a minimum of 12 data points on each peak. The data is obtained from a pooled serum sample. The chromatograms demonstrate a clear overlap in retention time and peak area of the deuterated internal standards.

separated. Therefore, it did not cause any problems during peak integration. To ensure that this peak was another compound than GBB, we spiked a plasma sample with various concentrations of GBB (see supplementary information Figure S1). The peak area of the earlier eluting peak did not change in size whilst the peak area of the later eluting peak did. This result indicates that the earlier eluting peak is indeed caused by another compound and not by the analyte GBB. The mobile phase of the FIA-MS method was optimized in order to reach to highest sensitivity. Therefore, 90% methanol and 0.1% formic acid were added to facilitate the electrospray ionization. Sensitivity is an important parameter in our method because we wanted to be able to measure a wide physiological range. This is especially important for TMAO, which is present in a wide physiological range (0.73-126 μM).²² Acetonitrile was used for the LC-MS analysis because of its stronger eluting power in comparison to methanol. This decreased the washing time of the LC column. The rest of the analysis parameters are the same for the FIA and LC analysis.

Matrix effect evaluation

Phospholipids are a major contributor to ion suppression.^{19,20} Phosphatidylcholines (PCs) are the most abundant phospholipids and this particular type is responsible for most of the ion suppression caused by phospholipids because of their high abundance.^{31,32} The physiological concentration of the sum of all PCs in human plasma mentioned in the literature is 211.3 mg/dL (157.0-327.0).^{33,34,35,36} For demonstration purposes in our study, one PC standard (211.3 mg/dL 1,2-dinonadecanoyl-sn-glycero-3-phosphocholine) was used to represent the PCs in plasma. The concentration of this standard was chosen to reflect the total concentration of PCs in plasma. Other well-known ion suppressors that are present in plasma in high concentrations are salts.^{37,38} To explore the effect of salts on the ionization efficiency we used a physiological saline solution (154 mM NaCl).

In order to examine the performance of the on-line cleanup, the matrix effect of various added matrix components was evaluated for the LC-MS and a FIA-MS analysis. The matrix effect was

determined by the ratio of the peak area of the deuterated internal standards in a matrix sample to the peak area of the deuterated internal standards in a water sample. We used deuterated internal standards because they are chemically similar to the corresponding gut metabolites and there were no background levels present in plasma for their MS/MS chromatograms. Figure 3 shows the matrix effect of the evaluated matrices on the peak area of the deuterated internal standards. The physiological PC concentration caused severe matrix effects when applying FIA-MS, reducing the signal to be 39 percent of the original signal. On the contrary, during the LC-MS analysis the PC matrix did not affect the signal. The matrix effect of the deuterated internal standards in the PC matrix was 103%, which was virtually identical to the reference values obtained in a water sample. This finding suggests that the LC column efficiently removed the PCs from the elution region of the analytes. The addition of a physiological salt concentration also caused substantial ion suppression. The FIA and LC analysis are both affected by this matrix because the salts coelute with the analytes in both methods. Betaine-d9 is considerably less affected during the LC analysis. This was expected since betaine has a slightly longer retention time on the LC column and will therefore elute further away from the unretained salts. Choline-d4 is not suppressed by the salt and is even enhanced in the FIA-MS method, which could be explained by the permanent positive net charge on the analyte. In the FIA, the PC and salt reduced the signal to 39 and 81% of the original signal, respectively, stressing the importance of PC removal before MS analysis. The matrix effect of PC and salt together already resembles the matrix

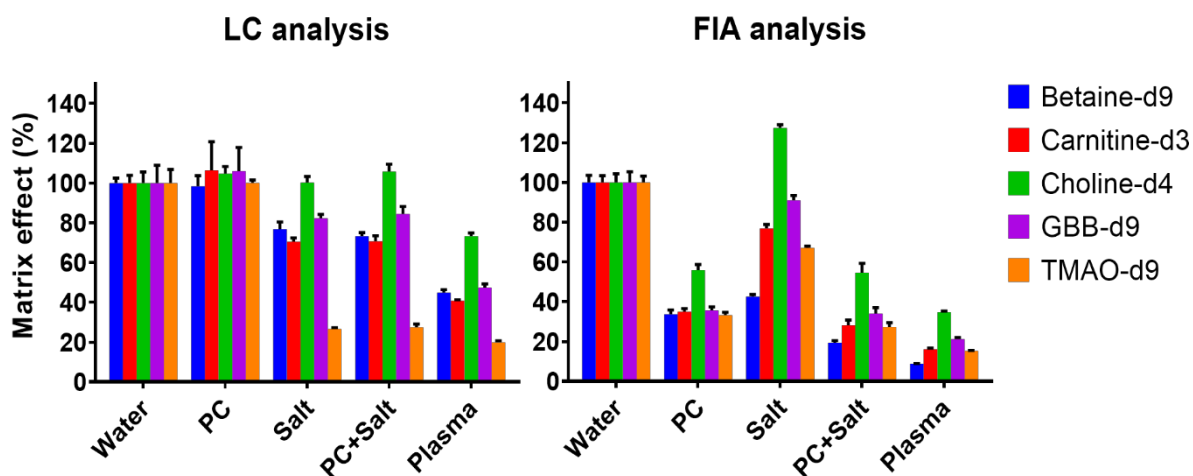


Figure 3. The matrix effect of various matrices on the peak area of the deuterated internal standards. Matrix effect was determined by the ratio of the peak area of the deuterated internal standards in a matrix sample to the peak area of the deuterated internal standards in a water sample. Physiological PC, salt, a combination of both and plasma have been evaluated for both FIA-MS and LC-MS.

effect caused by plasma, suggesting that these matrix constituents might be indeed responsible for the vast majority of matrix effect measured in plasma. This applies mainly to the PCs since these compounds are suppressing more than the salts. Although the suppression in the LC-MS analysis is considerably less, it has not been completely removed. Salts and other highly polar compounds (e.g. amino acids, creatinine, uric acid) that coelute with the gut metabolites still cause ion suppression. However, this remaining effect can be compensated by the use of deuterated internal standards. In addition to the reduced matrix effect, the elution of the retained matrix components can be switched to waste resulting in less contamination of the MS. In summary, the on-line matrix removal resulted in a cleaner analysis without suffering in terms of analysis time.

Method validation parameters

Table 1 summarizes the results of the method validation. In order to determine the LOD and LLOQ, a calibration curve was constructed in a water sample. Since we were aiming for absolute quantification using limited separation, we decided to calculate the LOD and LLOQ using the standard deviation of three replicates (see formula 3 and 4 in the methods section). This method comprises all sources of variability as well as the response. Lowest physiological concentration reported in the literature was at least six times higher than the LLOQ value, indicating a sufficient sensitivity of the developed method.^{22,39,40,41} The correlation coefficient showed an excellent linearity (≥ 0.998) for all calibration curves. The analysis of blank samples resulted in a limited blank effect ($< 1.5\%$). The repeatability and intermediate precision of pooled citrated plasma samples demonstrated an RSD consistently below 10%, indicating a high repeatability of the analysis. The used gradient resulted in a robust analysis since we were able to analyze more than 5000 samples using the same column without a decrease in performance.

Table 1. Method validation parameters for the fast LC-ESI-MS/MS method.

Analyte	Elution time (min)	LOD (μM)	LLOQ (μM)	Linearity (R^2)	Blank effect (%)	Repeatability (RSD, %)	Intermediate precision (RSD, %)
TMAO	0.33	0.04	0.1	0.999	0.5	6.1	5.2
Betaine	0.35	0.9	2.7	0.999	0.8	1.2	3.4
Choline	0.33	0.5	1.5	0.999	0.7	3.6	4.2
Carnitine	0.33	0.7	1.7	0.999	0.5	6.2	6.7
GBB	0.33	0.02	0.04	0.998	1.3	5.9	6.3

Different blood collection procedures and freeze-thaw cycles

Differences in concentrations of small metabolites between different blood collection procedures have been observed and should be taken into account.^{42,43} Therefore, we have evaluated four different blood collection procedures: ethylenediaminetetraacetic acid (EDTA) plasma, heparin plasma, citrated plasma and serum (Figure 4). The validated analytical method showed a good linearity with a mean R^2 of 0.998 (0.993-1.000) for all different collection procedures (Table 2) indicating the slopes of the calibration curves made in the different blood matrices as well as the calibration curve in a water sample were highly comparable with a mean RSD of 3.0% (1.6-3.4). The comparison to the uncorrected data (see Supplementary information Table S3), which demonstrated a mean R^2 of 0.971 (0.946-0.995) and a mean RSD of 28.9% (1.8-60.0), indicates that the deuterated internal standards are capable of extending the dynamic range, correcting the matrix effect of different blood matrices and allow for absolute quantification. We performed a one-way ANOVA and Post-Hoc analysis to investigate whether there was a statistical difference between the different collection procedures in terms of absolute concentrations. We have used freeze-thaw cycle 1 (FT1) for this comparison since most analyzed samples undergo one freeze-thaw cycle. All citrated plasma analyte concentrations demonstrated significantly lower values compared to the other blood collection procedures, except for betaine which did not reach this significance in comparison to EDTA and heparin treated blood. Lower citrate analyte concentrations can be explained by the way citrate plasma is obtained. When blood is treated with citrate, there is a dilution step involved because citrate is added as a liquid.⁴² In our experiment, the blood to citrate ratio was 9:1 and on average, citrate analyte concentrations were 11% lower in comparison to EDTA and heparin analyte concentrations. This decrease clearly reflects the dilution ratio of liquid in citrate plasma collection tubes. In contrast to the other analytes, choline concentrations were 17% higher in serum samples in comparison to EDTA and heparin samples.

Table 2. The slope and correlation coefficient of the calibration curves in a water sample and in different blood matrices using deuterated internal standard correction.

	TMAO		Betaine		Choline		Carnitine		GBB	
	Slope	R ²	Slope	R ²	Slope	R ²	Slope	R ²	Slope	R ²
Water	0.20	0.998	0.031	0.999	0.11	0.999	0.052	0.998	0.71	0.998
Heparin	0.19	0.998	0.030	0.998	0.11	1.000	0.050	0.999	0.74	0.993
EDTA	0.19	0.993	0.033	0.999	0.12	0.999	0.054	0.999	0.73	0.997
Citrate	0.19	0.999	0.032	1.000	0.11	0.998	0.049	0.999	0.72	1.000
Serum	0.21	0.998	0.030	0.997	0.11	1.000	0.053	0.999	0.72	0.999
RSD of slopes (%)	3.1		3.4		3.2		3.2		1.6	

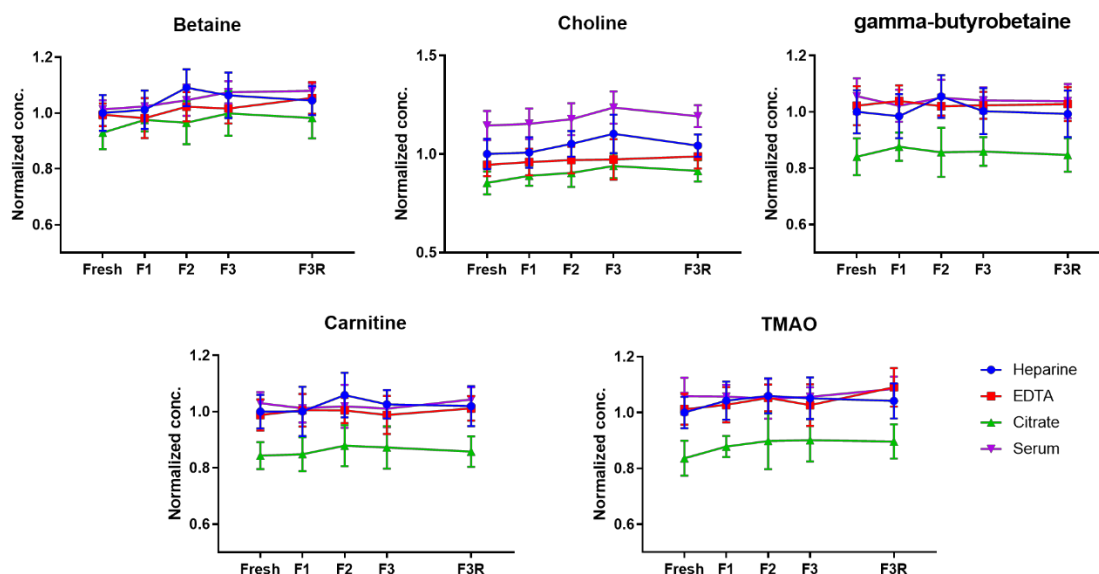


Figure 4: Different blood collection procedures of five healthy volunteers. Normalized concentrations of gut metabolites are plotted against the freeze-thaw (FT) cycles.

A higher analyte concentration in serum compared to plasma is a well-known phenomenon in metabolomics and can be explained in two ways.⁴³ First of all, platelets release metabolites into serum during the coagulation process.⁴⁴ Secondly, the coagulation process causes the clotting of fibrinogen. By removing the clot, the volume fraction of coagulation proteins is removed. The remaining analytes are left in a lower volume which makes them more concentrated.⁴⁵ Since the last explanation should have increased all analyte concentrations in serum, it seems more likely that the platelets have released choline into the serum. This is in accordance with a study of Petty and Scrutton, who demonstrated that platelets store and release choline.⁴⁶ In order to demonstrate the effect of different freeze-thaw cycles, we have compared analyte levels of fresh blood samples to samples which were exposed to one or multiple freeze thaw-cycle(s). Figure 4 demonstrates that gut metabolite levels are not shown to be affected by the number of freeze-thaw cycles. This is advantageous because study samples often experience one or more freeze-thaw cycles. On the other hand, the type of whole blood treatments can have a significant impact on the measured gut metabolites concentrations. This finding emphasizes that the treatment of whole blood is an important parameter and that it should be taken into consideration during clinical applications.

Comparison of fasting state and meal challenge

Figure 5 demonstrates the gut metabolite levels of 25 healthy volunteers during fasting (samples taken in the morning after overnight fasting) and half an hour after the consumption of a standardized liquid meal (SLM). The SLM contained 110 mg of choline – equivalent to the choline

content of approximately one egg or 140 g of fish or meat⁴⁷– and can therefore be considered as a substantial source of TMAO precursors. However, none of the targeted metabolites revealed a statistically significant difference between the fasting and SLM-challenged time point (FDR corrected p -value <0.05 , see Supplementary information Table S4) and no clear trend could be identified in the before-after plots in Figure 5, unless stratified by gender. The heat map in Figure 6 also demonstrates that there is no uniform response to the SLM. Volunteers with increased (cluster A) and decreased (cluster C) gut metabolite levels after the SLM-challenge could clearly be distinguished based on hierarchical clustering. The difference in response might be explained by the gender of the volunteers since differences in betaine, carnitine, choline, γ -butyrobetaine serum levels were significantly correlated with the gender of the volunteers (FDR corrected p -value <0.05 , see Supplementary information Table S5). In general, men and women demonstrated increased and decreased gut metabolite levels, respectively, after the challenge of SLM consumption compared to fasting. The decreasing levels in female volunteers were surprising because this trend has not been observed before after food intake. The observed trend does indicate that gender might be an important parameter in determining the gut metabolite response caused by food intake. However, when fasting and SLM-challenged serum levels were stratified by gender, only male betaine serum levels significantly changed between the two time points (FDR corrected p -value <0.05 , see Table S4 in the Supplementary information). TMAO response was not correlated with the gender of the volunteers and showed most dissimilarities (distance in the dendrogram) in comparison to the other gut metabolites. The deviant behavior of TMAO might be explained by the reversibility of the TMA conversion to TMAO. Although a gender-related trend was observed, the differences between the two time points and volunteers might have also been caused by other sources of variability like the type and quantity of bacteria in the gut microbiome, variability of expression of choline dehydrogenase in the kidney cortex and biosynthesis of carnitine and choline. For example, a decrease in choline serum levels can be caused by an

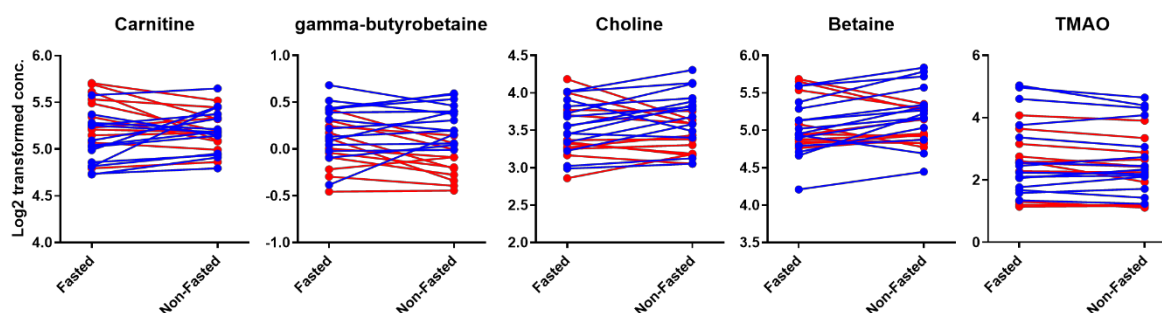


Figure 5. Before-after plots of gut metabolite serum levels of 25 volunteers during fasting and non-fasting (i.e. 33 min after a meal). The 14 male volunteers are depicted in blue and the 11 female volunteers are depicted in red.

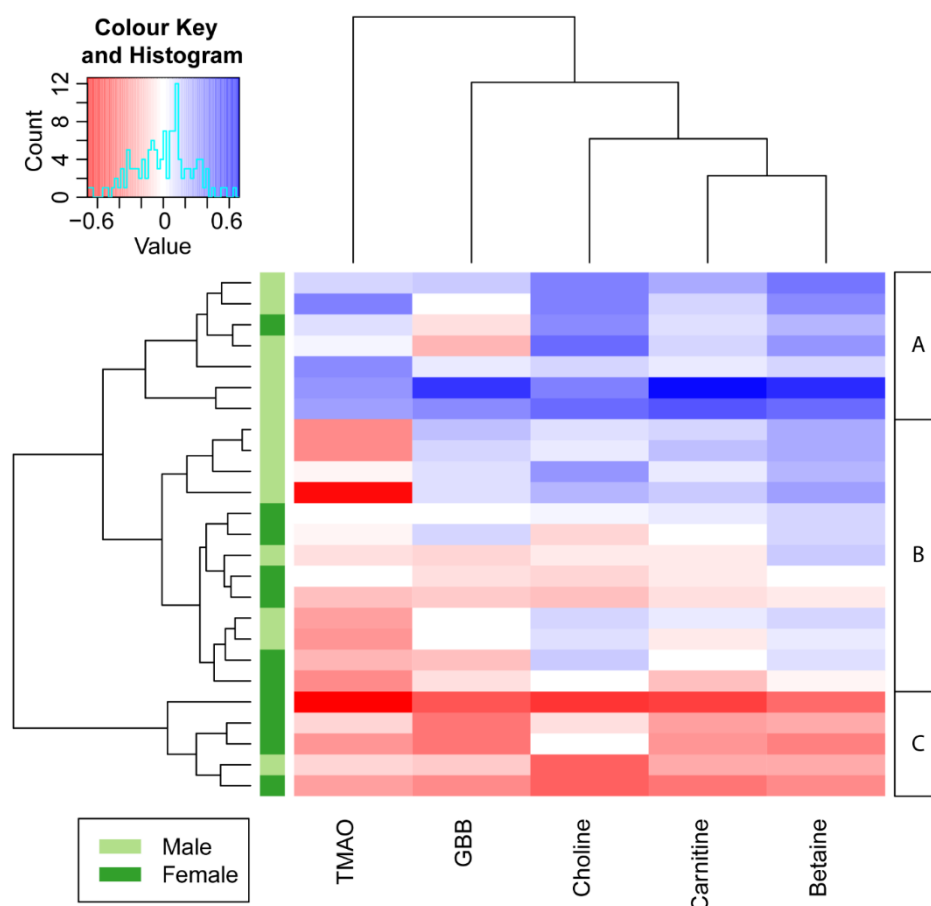


Figure 6. Heat map of the log₂ transformed serum level differences between the fasted and non-fasted time point. Blue and red cells represent increased and decreased SLM-challenged metabolite levels in serum, respectively. The volunteers and metabolites are sorted by hierarchical clustering. Dendrograms are used to show distances between samples (rows) and metabolites (columns). Based on the hierarchical clustering, three groups could be identified indicating volunteers with increased levels (A), limited differences (B) and decreased levels (C). It can be clearly seen that A and C consist of mainly male and female volunteers, respectively.

increased expression of choline dehydrogenase and an increase in carnitine and GBB could be a result of an increased biosynthesis of carnitine.

Although the gut metabolite serum levels did not change significantly between the fasted and SLM- challenged time point, an increase in glucose and branched-chain amino acids was demonstrated by Schutte *et al.* in the same sample collection and SLM-challenge as used in this study.⁴⁸ This indicates that the SLM is able to elevate certain metabolites in blood within half an hour. In addition to that, Lang *et al.* demonstrated that FMO3 catalyzes TMA *N*-oxygenation with a k_{cat} of more than thirty per minute indicating that TMAO could be formed within half an hour.⁴⁹ In contrast to our finding, Cho *et al.* found a significant increase in plasma TMAO levels after the consumption of foods containing TMAO or its precursors.⁵⁰ In that study, volunteers fasted

overnight and were fed a diet containing eggs, beef or fish. After 15 minutes there was already a significant increase in plasma TMAO levels observed for all diets. The difference might be explained by the fact that only male volunteers were enrolled in the latter study. In accordance to our results, male volunteers demonstrated increased levels of gut metabolites after the consumption of a meal. Female volunteers, however, demonstrated an opposite trend in our results. Another difference can be found in the administration of 'real' dairy products in contrast to the SLM-challenge ('artificial drink') that has been used in our study. The SLM did not contain TMAO or carnitine and could therefore not increase TMAO levels by direct absorption of TMAO or via microbial conversion of carnitine. Other studies did not find a strong association between animal food intake and plasma concentrations of TMAO, betaine and choline.^{51,52} Although clear differences in animal food intake could be identified, corresponding plasma levels of gut metabolites could not be correlated with this variation. However, the difference between fasting and food intake was not examined in these studies. Therefore, these studies do not reveal whether food consumption itself contributes to different gut metabolite levels.

CONCLUSION

We have developed a high throughput LC-MS/MS platform to analyze gut metabolites in plasma and serum. The method can quantify five gut metabolites in three minutes per sample. By using an on-line cleanup approach, the most important ion suppressors, PCs, were removed from the elution region of the gut metabolites reducing the ion suppression significantly. The blood collection procedure had a significant impact on the measured concentrations of the gut metabolites. Therefore, the type of blood collection is of great importance within a study, and can be of great importance when different studies are compared to each other because inter-study variability could be caused by the collection procedures rather than, for instance, a diet or drug treatment. Our results did not demonstrate a significant diet-related effect of the gut metabolite levels in serum, since the gut metabolites hardly changed 30 minutes after the intake of an SLM. This is in contrast to another study, in which an increase was observed. However, this difference could have been caused by the administration of different types of food. We did observe volunteers with increased and decreased levels, which could have been driven by the gender. Therefore, the influence of fasting on the gut metabolite levels remains a subject of debate and is a parameter that should be studied in more detail. In summary, gut metabolites have been reported to have great potential to become a valuable biomarker in cardiovascular diseases, but this requires validation in large studies. Our simple and fast method facilitates further research and validation studies on the role of gut metabolites in relation to cardiovascular events and should stimulate other researchers to elucidate the biology behind this role. If TMAO is fully accepted as a valuable biomarker for cardiovascular diseases, our relatively easy method can help

clinicians to make fast diagnoses of cardiovascular disease risks and can possibly contribute to prevent major cardiovascular events like death, myocardial infarction, or stroke.

ACKNOWLEDGEMENTS

We would like H. Eka D. Suchiman to be acknowledged for preparing the serum samples of the GOTO study. The GOTO study was financially supported by the Netherlands Consortium for Healthy Ageing (grant 050-060-810), in the framework of the Netherlands Genomics Initiative, Netherlands Organization for Scientific Research (NWO); by BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007) and by the Netherlands CardioVascular Research Initiative (CVON201-03).

REFERENCES

1. Wang, Z. et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472, 57–63 (2011).
2. Koeth, R. A. et al. Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* 19, 576–585 (2013).
3. Tang, W. H. W. et al. Intestinal Microbial Metabolism of Phosphatidylcholine and Cardiovascular Risk. *N. Engl. J. Med.* 368, 1575–1584 (2013).
4. Griffin, J. L., Wang, X. & Stanley, E. Does Our Gut Microbiome Predict Cardiovascular Risk? *Circ. Cardiovasc. Genet.* 8, 187–191 (2015).
5. Bennett, B. J. et al. Trimethylamine-N-Oxide, a Metabolite Associated with Atherosclerosis, Exhibits Complex Genetic and Dietary Regulation. *Cell Metab.* 17, 49–60 (2013).
6. Wang, Z. et al. Prognostic value of choline and betaine depends on intestinal microbiota-generated metabolite trimethylamine-N-oxide. *Eur. Heart J.* 35, 904–910 (2014).
7. Lever, M., George, P. M., Dellow, W. J., Scott, R. S. & Chambers, S. T. Homocysteine, glycine betaine, and N,N-dimethylglycine in patients attending a lipid clinic. *Metabolism.* 54, 1–14 (2005).
8. Zeisel, S. H. & Warriar, M. Trimethylamine N -Oxide, the Microbiome, and Heart and Kidney Disease . *Annu. Rev. Nutr.* 37, 157–181 (2017).
9. Koeth, R. A. et al. γ -butyrobetaine is a proatherogenic intermediate in gut microbial metabolism of L-carnitine to TMAO. *Cell Metab.* 20, 799–812 (2014).
10. Zeisel, S. H. & Da Costa, K. A. Choline: An essential nutrient for public health. *Nutr. Rev.* 67, 615–623 (2009).
11. Fennema, D., Phillips, I. R. & Shephard, E. A. Trimethylamine and Trimethylamine N-Oxide, a Flavin-Containing Axis Implicated in Health and Disease. *Drug Metab. Dispos.* 44, 1839–1850 (2016).
12. Wojtowicz, W. et al. Serum and urine ^1H NMR-based metabolomics in the diagnosis of selected thyroid diseases. *Sci. Rep.* 7, 1–13 (2017).
13. Garcia, E. et al. NMR quantification of trimethylamine-N-oxide in human serum and plasma in the clinical laboratory setting. *Clin. Biochem.* 50, 947–955 (2017).
14. Zuppi, C. et al. ^1H NMR spectra of normal urines: Reference ranges of the major metabolites. *Clin. Chim. Acta* 265, 85–97 (1997).
15. Nicholls, A. W., Mortishire-Smith, R. J. & Nicholson, J. K. NMR Spectroscopic-Based Metabonomic Studies of Urinary Metabolite Variation in Acclimatizing Germ-Free Rats. *Chem. Res. Toxicol.* 16, 1395–1404 (2003).

16. Hsu, W.-Y. et al. Rapid screening assay of trimethylaminuria in urine with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 21, 1915–1919 (2007).
17. Johnson, D. W. A flow injection electrospray ionization tandem mass spectrometric method for the simultaneous measurement of trimethylamine and trimethylamine N -oxide in urine. *J. Mass Spectrom.* 43, 495–499 (2008).
18. Furey, A., Moriarty, M., Bane, V., Kinsella, B. & Lehane, M. Ion suppression; A critical review on causes, evaluation, prevention and applications. *Talanta* 115, 104–122 (2013).
19. Ghosh, C., Shinde, C. P. & Chakraborty, B. S. Influence of ionization source design on matrix effects during LC-ESI-MS/MS analysis. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 893–894, 193–200 (2012).
20. Little, J. L., Wempe, M. F. & Buchanan, C. M. Liquid chromatography-mass spectrometry/mass spectrometry method development for drug metabolism studies: Examining lipid matrix ionization effects in plasma. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 833, 219–230 (2006).
21. Kuka, J. et al. Suppression of intestinal microbiota-dependent production of pro-atherogenic trimethylamine N-oxide by shifting L-carnitine microbial degradation. *Life Sci.* 117, 84–92 (2014).
22. Wang, Z. et al. Measurement of trimethylamine-N-oxide by stable isotope dilution liquid chromatography tandem mass spectrometry. *Anal. Biochem.* 455, 35–40 (2014).
23. Zhao, X., Zeisel, S. H. & Zhang, S. Rapid LC-MRM-MS assay for simultaneous quantification of choline, betaine, trimethylamine, trimethylamine N -oxide, and creatinine in human plasma and urine. *Electrophoresis* 36, 2207–2214 (2015).
24. Steuer, C., Schutz, P., Bernasconi, L. & Huber, A. R. Simultaneous determination of phosphatidylcholine-derived quaternary ammonium compounds by a LC-MS/MS method in human blood plasma, serum and urine samples. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 1008, 206–211 (2016).
25. Mi, S., Zhao, Y.-Y., Jacobs, R. L. & Curtis, J. M. Simultaneous determination of trimethylamine and trimethylamine N -oxide in mouse plasma samples by hydrophilic interaction liquid chromatography coupled to tandem mass spectrometry. *J. Sep. Sci.* 40, 688–696 (2017).
26. Heaney, L. M., Jones, D. J. L., Mbasu, R. J., Ng, L. L. & Suzuki, T. High mass accuracy assay for trimethylamine N-oxide using stable-isotope dilution with liquid chromatography coupled to orthogonal acceleration time of flight mass spectrometry with multiple reaction monitoring. *Anal. Bioanal. Chem.* 408, 797–804 (2016).
27. Ocque, A. J., Stubbs, J. R. & Nolin, T. D. Development and validation of a simple UHPLC-MS/MS method for the simultaneous determination of trimethylamine N-oxide, choline, and betaine in human plasma and urine. *J. Pharm. Biomed. Anal.* 109, 128–135 (2015).
28. Liu, J. et al. Simultaneous targeted analysis of trimethylamine-N-oxide, choline, betaine, and carnitine by high performance liquid chromatography tandem mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 1035, 42–48 (2016).
29. Kadar, H. et al. A multiplexed targeted assay for high-throughput quantitative analysis of serum methylamines by ultra performance liquid chromatography coupled to high resolution mass spectrometry. *Arch. Biochem. Biophys.* 597, 12–20 (2016).
30. Lee, M. F. X., Chan, E. S. & Tey, B. T. Negative chromatography: Progress, applications and future perspectives. *Process Biochem.* 49, 1005–1011 (2014).
31. Ismaiel, O. A., Zhang, T., Jenkins, R. G. & Karnes, H. T. Investigation of endogenous blood plasma phospholipids, cholesterol and glycerides that contribute to matrix effects in bioanalysis by liquid chromatography/mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 878, 3303–3316 (2010).
32. Myher, J. J., Kuksis, a & Pind, S. Molecular species of glycerophospholipids and sphingomyelins of human erythrocytes: improved method of analysis. *Lipids* 24, 396–407 (1989).

33. Seppänen-Laakso, T. et al. Major human plasma lipid classes determined by quantitative high-performance liquid chromatography, their variation and associations with phospholipid fatty acids. *J. Chromatogr. B Biomed. Sci. Appl.* 754, 437–445 (2001).
34. Quehenberger, O. et al. Lipidomics reveals a remarkable diversity of lipids in human plasma. *J. Lipid Res.* 51, 3299–3305 (2010).
35. Dougherty, R. M., Galli, C., Ferro-Luzzi, A. & Iacono, J. M. Lipid and phospholipid fatty acid composition of plasma, red blood cells, and platelets and how they are affected by dietary lipids: a study of normal subjects from Italy, Finland, and the USA. *Am. J. Clin. Nutr.* 45, 443–455 (1987).
36. Maldonado, E. N., Romero, J. R., Ochoa, B. & Aveldaño, M. I. Lipid and fatty acid composition of canine lipoproteins. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.* 128, 719–729 (2001).
37. Trufelli, H., Palma, P., Famiglini, G. & Cappiello, G. An overview of matrix effects in liquid chromatography-mass spectrometry. *Indian J. Exp. Biol.* 30, 491– 509 (2011).
38. Annesley, T. M. Ion suppression in mass spectrometry. *Clin. Chem.* 49, 1041–1044 (2003).
39. Bain, M. A., Faull, R., Fornasini, G., Milne, R. W. & Evans, A. M. Accumulation of trimethylamine and trimethylamine-N-oxide in end-stage renal disease patients undergoing haemodialysis. *Nephrol. Dial. Transplant.* 21, 1300–1304 (2006).
40. Psychogios, N. et al. The human serum metabolome. *PLoS One* 6, (2011).
41. Vernez, L., Wenk, M. & Krähenbühl, S. Determination of carnitine and acylcarnitines in plasma by high-performance liquid chromatography/electrospray ionization ion trap tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 18, 1233–1238 (2004).
42. Gonzalez-Covarrubias, V., Dane, A., Hankemeier, T. & Vreeken, R. J. The influence of citrate, EDTA, and heparin anticoagulants to human plasma LC-MS lipidomic profiling. *Metabolomics* 9, 337–348 (2013).
43. Yu, Z. et al. Differences between human plasma and serum metabolite profiles. *PLoS One* 6, 1–6 (2011).
44. Naoki, Y. et al. Sphingosine 1-phosphate, a bioactive sphingolipid abundantly stored in platelets, is a normal constituent of human plasma and serum. *J. Biochem.* 973, 969–973 (1997).
45. Kronenberg, F. et al. Influence of hematocrit on the measurement of lipoproteins demonstrated by the example of lipoprotein(a). *Kidney Int.* 54, 1385–1389 (1998).
46. Petty, A. C. & Scrutton, M. C. Release of Choline Metabolites from Human Platelets: Evidence for Activation of Phospholipase D and of Phosphatidylcholine-specific Phospholipase C. *Platelets* 4, 23–29 (1993).
47. Patterson, K. Y. et al. USDA Database for the Choline Content of Common Foods In collaboration with. *Environ. Heal.* 1–37 (2008). doi:10.15482/USDA.ADC/1178141
48. Schutte, B. A. M. et al. The effect of standardized food intake on the association between BMI and 1 H-NMR metabolites. *Sci. Rep.* 6, 1–6 (2016).
49. Lang, D. H. et al. Isoform specificity of trimethylamine N-oxygenation by human flavin-containing monooxygenase (FMO) and P450 enzymes Selective catalysis by fmo3. *Biochem. Pharmacol.* 56, 1005–1012 (1998).
50. Cho, C. E. et al. Trimethylamine-N-oxide (TMAO) response to animal source foods varies among healthy young men and is influenced by their gut microbiota composition: A randomized controlled trial. *Mol. Nutr. Food Res.* 61, 1–12 (2017).
51. Kühn, T. et al. Intra-individual variation of plasma trimethylamine-N-oxide (TMAO), betaine and choline over 1 year. *Clin. Chem. Lab. Med.* 55, 261–268 (2017).
52. Rohrmann, S., Linseisen, J., Allenspach, M., von Eckardstein, A. & Müller, D. Plasma Concentrations of Trimethylamine-N-oxide Are Directly Associated with Dairy Food Consumption and Low-Grade Inflammation in a German Adult Population. *J. Nutr.* 146, 283–289 (2016).
53. van de Rest, O. et al. Metabolic effects of a 13-weeks lifestyle intervention in older adults: The Growing Old Together Study. *Aging (Albany, NY)*. 8, 111–26 (2016).

SUPPLEMENTARY INFORMATION**Table S1. Supplier, collision energy and transitions of the standards and deuterated internal standards.**

Name	(Deuterated internal) standard	Supplier	Collision energy (eV)	Transition (m/z)
Betaine	Betaine hydrochloride	Sigma Aldrich	30	118.1→58.2
Betaine d-9	N-(Carboxymethyl)-trimethyl-d9 ammonium chloride	CDN isotopes	30	127.1 → 66.1
L-carnitine	L-carnitine hydrochloride	Sigma Aldrich	25	162.3→85.1
L-carnitine-d3	L-carnitine-d3 hydrochloride	CDN isotopes	25	165.3→85.1
choline	Choline chloride	Sigma Aldrich	25	104.2→60.1
choline-d4	Choline-1,1,2,2-d4 chloride	CDN isotopes	25	108.2→60.1
TMAO	Trimethylamine N-oxide dihydrate	Sigma Aldrich	25	76.1→58.1
TMAO-d9	Trimethylamine N-oxide-d9	Cambridge Isotope Laboratories	25	85.1→66.1
GBB	(3-carboxypropyl)trimethyl-ammonium chloride	Sigma Aldrich	20	146.2→87.1
GBB-d9	(3-carboxypropyl)trimethyl-d9-ammonium chloride	CDN isotopes	20	155.2→87.1

Table S2. Concentration of the C8 of the standards and deuterated internal standards.

C8 concentration of the standards		Deuterated internal standard concentration	
Name	concentration (μM)	Name	concentration (μM)
TMAO	43.2	TMAO-d9	3.56
Betaine	716.1	Betaine-d9	24.59
Choline	286.5	Choline-d4	13.92
Carnitine	556.5	Carnitine-d3	19.93
γ-butyrobetaine	17.6	γ-butyrobetaine -d9	1.05

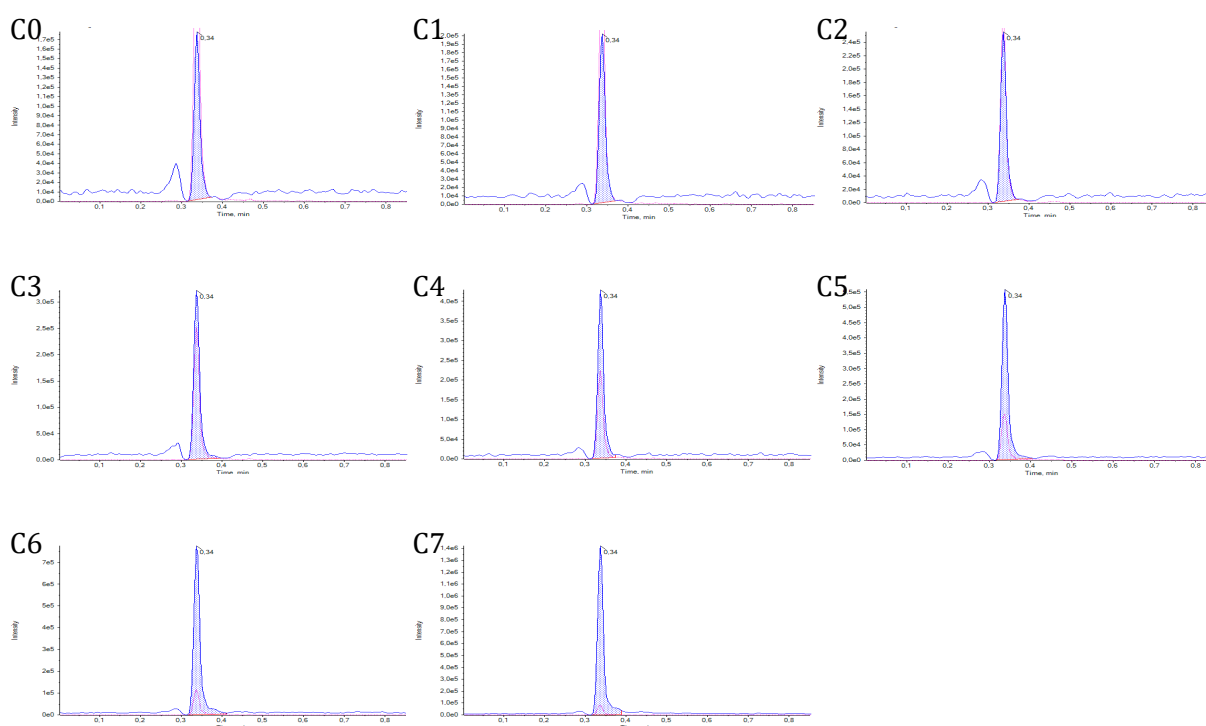


Figure S1. The extracted ion chromatogram of γ -butyrobetaine at increasing concentrations (C0-7) of the standard. The blue line represents γ -butyrobetaine signal and the red line represents the deuterated internal standard γ -butyrobetaine-d9 signal. It can be clearly seen that the small peak before γ -butyrobetaine is baseline separated and is not increasing with increasing concentrations of the standard γ -butyrobetaine.

Table S3. The slope and correlation coefficient of the calibration curves in a water sample and in different blood matrices without using deuterated internal standard correction.

	TMAO		Betaine		Choline		Carnitine		GBB	
	Slope	R ²	Slope	R ²	Slope	R ²	Slope	R ²	Slope	R ²
Water	97523	0.950	40692	0.978	77780	0.960	142427	0.973	128063	0.956
Heparin	72464	0.983	13422	0.981	79004	0.966	82676	0.966	45480	0.986
EDTA	74382	0.975	11887	0.977	80663	0.946	69215	0.946	36374	0.962
Citrate	81862	0.988	12458	0.988	81660	0.951	72620	0.951	42873	0.995
Serum	71976	0.985	14119	0.988	80952	0.970	81514	0.970	46438	0.990
RSD of slopes (%)	12.1		60.0		1.8		29.9		57.3	

Table S4. Statistical difference between the fasting and non-fasting time point for all volunteers and stratified by gender. Significant differences are indicated (* indicates p-value <0.05).

Compound name	gender	p-value	FDR corrected
Betaine	all	0.076	0.142
Carnitine	all	0.849	0.849
Choline	all	0.296	0.370
gamma-butyrobetaine	all	0.441	0.508
TMAO	all	0.056	0.120
Betaine	Male	0.001*	0.009*
Carnitine	Male	0.036*	0.091
Choline	Male	0.013*	0.065
gamma-butyrobetaine	Male	0.141	0.235
TMAO	Male	0.544	0.583
Betaine	Female	0.243	0.349
Carnitine	Female	0.031*	0.091
Choline	Female	0.256	0.349
gamma-butyrobetaine	Female	0.007*	0.053
TMAO	Female	0.020*	0.075

Table S5. Association between gender and log2 transformed fasted and non-fasted serum level differences. Significant correlations are indicated (* indicates p-value <0.05).

Compound name	p-value	FDR corrected
Betaine	0.001*	0.004*
Carnitine	0.002*	0.004*
Choline	0.013*	0.016*
gamma-butyrobetaine	0.002*	0.004*
TMAO	0.211	0.211

Chapter 3

High-throughput fractionation coupled to mass spectrometry for improved quantitation in metabolomics

Based on:

Tom van der Laan, Anne-Charlotte Dubbelman, Kevin Duisters, Alida Kindt, Amy C. Harms and Thomas Hankemeier

High-throughput fractionation coupled to mass spectrometry for improved quantitation in metabolomics

Analytical Chemistry (2020)

ABSTRACT

Metabolomics is emerging as an important field in life sciences. However, a weakness of current mass spectrometry (MS) based metabolomics platforms is the time-consuming analysis and the occurrence of severe matrix effects in complex mixtures. To overcome this problem, we have developed an automated and fast fractionation module coupled on-line to MS. The fractionation is realized by the implementation of three consecutive high performance solid-phase extraction columns consisting of a reversed phase, mixed-mode anion exchange and mixed-mode cation exchange sorbent chemistry. The different chemistries resulted in an efficient interaction with a wide range of metabolites based on polarity and charge and allocation of important matrix interferences like salts and phospholipids. The use of short columns and direct solvent switches allowed for fast screening (3 min per polarity). In total, 50 commonly-reported diagnostic or explorative biomarkers were validated with a limit of quantification that was comparable with conventional LC-MS(/MS). In comparison with a flow injection analysis without fractionation, ion suppression decreased from 89% to 25% and the sensitivity was 21 times higher. The validated method was used to investigate the effects of circadian rhythm and food intake on several metabolite classes. The significant diurnal changes that were observed stress the importance of standardized sampling times and fasting states when metabolite biomarkers are used. Our method demonstrates a fast approach for global profiling of the metabolome. This brings metabolomics one step closer to the implementation into the clinic.

INTRODUCTION

Metabolomics is increasingly important in the field of life sciences. It is used for the screening of inborn errors of metabolism¹, precision medicine² and discovery of new biomarkers for health, disease and intervention.³ To accommodate this increased interest, there is a need for fast and comprehensive screening of the metabolome.⁴ Mass spectrometry (MS) is a highly sensitive technique and MS-based methods can screen a large range of metabolites in a single run.⁵ This makes MS highly suitable for comprehensive metabolomics. The downside of MS is that it often requires extensive sample preparation and separation to reduce interferences of complex biological samples at the ionization source.⁶

Flow injection analysis coupled to mass spectrometry (FIA-MS) is an appealing approach in fast and comprehensive screening since there is no chromatography that discriminates against compound classes or decreases the throughput.⁷ The sample preparation of these methodologies is often a 'dilute-and-shoot' approach, whereby dilution is applied to decrease the interference of the sample matrix at the ionization source. However, these methods often suffer in terms of sensitivity because the analytes are also diluted or high abundant matrix interferences still cause severe ion suppression.⁸ Therefore, sample preparation remains an important aspect in fast MS analysis in order to decrease the sample complexity while maintaining a sufficient analyte concentration. Liquid-liquid extraction (LLE) has been performed in parallel and coupled to FIA-MS to improve throughput and coverage.⁹ On the other hand, solid-phase extraction (SPE) has been coupled on-line to mass spectrometry in the RapidFire system resulting in analyses times of around 8.5 seconds.¹⁰ By using LLE or different SPE sorbents in parallel, however, the cleanup efficiency remains limited. Generally, these approaches only result in two fractions (water/organic fraction in LLE and flow-through/elution fraction in SPE) and fractions are ionized at once without within-fraction separation.

In this work, we demonstrate a comprehensive and fast sample preparation method coupled on-line to MS. The method utilizes two important chemical properties of the metabolome: polarity and charge. Three consecutive high performance (particle size $\leq 5 \mu\text{m}$) SPE columns, consisting of a reversed phase, mixed-mode cation exchange and mixed-mode anion exchange sorbent chemistry, are coupled on-line to a mass spectrometer. This ensured the allocation of metabolites into different fractions (flow-through; polar/neutral, reversed phase; apolar, cation exchange; polar and positive, anion exchange; polar and negative). Moreover, it also removed known ion suppressors from different fractions minimizing their adverse effects during electrospray ionization. Phospholipids and salts are held responsible for a majority of signal suppression during electrospray ionization of plasma samples.¹¹ By using a fractionation approach based on polarity and charge, phospholipids are retained on the reversed phase column whereas positive

and negative salt ions are trapped on and eluted from the cation and anion exchange, respectively. Another benefit of serially coupled columns is the flow-through fraction, which is cleaned by three sorbent chemistries instead of one in conventional single-column methods. The advantage of on-line fractionation over off-line fractionation is that it allows for some separation between compounds within a fraction prior to electrospray ionization. Hereby, retained ion suppressors could elute at another time than retained analytes. To our knowledge, this is the first publication that reports the use of serially coupled high performance SPE columns to realize an on-line fractionation including some separation prior to MS analysis. The strength of this platform is emphasized by the use of short analytical columns which allow for fast solvent switches while still benefitting from chromatographic separation.

We have developed a targeted platform for the analysis of 50 commonly-reported diagnostic or explorative biomarkers.^{12,13,14} These compounds belong to the following compound classes: amino acids, amines, purines, sugars, acylcarnitines, organic acids and fatty acids. We present a fast on-line sample preparation method that fractionates these compound classes in plasma. Several on-line SPE columns have been evaluated for their ability to fractionate plasma prior to MS analysis. The optimized methods for both positive and negative electrospray ionization mode have been validated and applied in a study investigating the effect of circadian rhythm and food intake on several metabolite classes. This study should give insight into the diurnal variations of the studied biomarkers. These variations are important to assess because they could potentially be misinterpreted as disease or intervention related variations. This misinterpretation compromises the diagnostic and explorative power of a potential biomarker.

MATERIALS & METHODS

Chemicals

An overview of the used (internal) standards and concentrations is provided in the Supplementary information (Table S1 and S2). Water was obtained from an arium pro UF/VF water purification system with a Sartopore 2 0.2 μm filter. Methanol (Ultra-LC-MS grade) was purchased from Actu-All (Oss, The Netherlands). Ammonium hydroxide (28-30 wt% solution of ammonia in water) and formic acid (98%+) were purchased from Acros Organics (Bleiswijk, The Netherlands). Ammonium acetate ($\geq 99.0\%$) and ammonium formate ($\geq 99.995\%$) were purchased from Sigma-Aldrich (Zwijndrecht, The Netherlands).

Method development

We have used polymeric mixed-mode ion exchange columns because they provide a superior pH stability over other ion exchange sorbent types. Several ion exchange columns have been evaluated according to the retention, trapping and elution performances of representative

standards. We tested four low performance (particle size $>5\mu\text{m}$), four high performance Sepax (particle size $1.7\text{-}5\mu\text{m}$) and four high performance Zirchrom (particle size $3\mu\text{m}$) SPE columns. The low performance, Sepax and Zirchrom SPE columns were composed of four mixed-mode ion exchange types (strong cation exchange (SCX), strong anion exchange (SAX), weak cation exchange (WCX) and weak anion exchange (WAX)). Similar loading and elution buffers were used for each type of ion exchange. The evaluated ion exchange columns, loading and elution buffers explored during development can be found in the Supplementary information (Table S3). The selected ion exchange columns were coupled to a reversed phase column and ordered in a way that was most beneficial in terms of matrix effect reduction and peak shape. The reversed phase column was a ZORBAX Extend-C18, $2.1\text{ mm} \times 5\text{ mm}$, $1.8\text{ }\mu\text{m}$ guard column from Agilent Technologies Netherlands (Waldbronn, The Netherlands).

Five cationic compounds were used to represent different types of cations (leucine, glutamic acid, arginine, hypoxanthine and choline) and four anionic compounds were used to represent different types of anions (lactic acid, malic acid, citric acid and indoxyl sulfate). The amino acids consisted of cationic and anionic functional groups. Glucose functioned as a neutral marker and indicated whether ions were efficiently removed from the column flow-through.

Validation

Individual stock solutions and calibration mixtures were stored at $-80\text{ }^{\circ}\text{C}$. In each specific fraction, there was at least one internal standard present. In total seven calibration points were used (C1-7). The highest calibration concentration is referred to as C7 (Supplementary information Table S1) and the subsequent concentrations were 1:1 dilutions of the previous concentration. All calibration standards were included in the same stock solution and all calibration solutions were composed of 69% methanol in water. C0 was prepared by adding 69% MeOH without standards. Within the calibration range, C4 and the internal standard concentration were set to mimic the physiological concentration of the analyte found on the Human Metabolome Database (HMDB).¹⁵ Calibration curves were constructed by standard addition of the calibration standards to plasma samples. The repeatability of the method was determined by the relative standard deviation of three replicates of three different concentrations (C0, C2 and C4). The intermediate precision was determined by the relative standard deviation of three different concentrations (C0, C2 and C4) on three different days (N=9). The matrix effect was determined by the ratio of the peak area of the internal standard in a plasma and water sample.¹⁶ Ion suppression was determined by subtracting 100% by the matrix effect. Ion suppression of ion enhanced compounds was set at 0% when calculating the mean ion suppression.

$$(1) \text{ Matrix effect} = \frac{\text{Area ISTD in plasma}}{\text{Area ISTD in water}} \times 100\%$$

$$(2) \text{ Ion suppression} = 100\% - \text{Matrix effect}$$

The carryover was evaluated as the ratio of the peak area in a blank sample and the peak area in a pooled plasma sample that was analyzed just before the blank (N=3). Ten concentration levels of internal standards were used to determine the limit of detection (LOD) and lower limit of quantification (LLOQ). The highest concentration was C6 which was four times the physiological value of the unlabeled counterpart (Supplementary information Table S2) and the subsequent concentrations were 1:1 dilutions of the previous concentration. The LOD (formula 3) and LLOQ (formula 4) were determined by the following formula which used the peak area of a blank, the standard deviation (SD) of the lowest concentration with a S/N greater than 3 (C_{low}) and the response factor (RF), which was calculated by the ratio of the peak area and concentration of C_{low} .

$$(3) \text{ LOD} = \frac{3 \times SD_{\text{area } C_{low}} + \text{area}_{\text{Blank}}}{\left(\frac{\text{area}_{C_{low}}}{[C_{low}]}\right)}$$

$$(4) \text{ LLOQ} = \frac{10 \times SD_{\text{area } C_{low}} + \text{area}_{\text{Blank}}}{\left(\frac{\text{area}_{C_{low}}}{[C_{low}]}\right)}$$

Sample preparation

During the method validation, 30 μL EDTA plasma aliquots, 30 μL of calibration standard and 30 μL of the internal standard solution, H_2O and MeOH were mixed reaching a total volume of 195 μL and 71% MeOH. The mixture was vigorously vortexed and centrifuged (10 min, 16100 g, 4 $^\circ\text{C}$). After centrifugation, 100 μL of the supernatant was transferred into an autosampler vial containing a 150 μL insert. Study samples were prepared by mixing 15 μL EDTA, 15 μL of internal standard solution, H_2O and MeOH reaching a total volume of 97.5 μL and 71% MeOH (same ratios as during method validation). The vortex and centrifuge step remained the same and 50 μL of the supernatant was transferred into an autosampler vial containing a 150 μL insert.

The flow injection analysis (FIA) sample preparation was adapted from Carducci *et al.*¹⁷ Ten microliters of EDTA plasma and internal standard solution were mixed with methanol, water and acetic acid to reach a final solution of 80% methanol, 0.1% acetic acid and a plasma dilution ratio of 100. This dilution ratio was found to give the highest sensitivity after testing plasma dilution ratios of 10 to 500. An adjusted Bligh and Dyer LLE was also performed prior to the FIA.¹⁸ Ten microliters of EDTA plasma and internal standard solution were extracted with methanol, dichloromethane and water (v/v/v, 2/2/1.8) reaching a total volume of 1000 μL . 200 μL of the

apolar and 200 μL of the polar fraction were evaporated and separately reconstituted in 200 μL 0.1% acetic acid in 80% MeOH.

Fractionation and mass spectrometry

A Shimadzu Nexera UHPLC (Darmstadt, Germany) was connected to a Sciex X500R QToF (Darmstadt, Germany). The setup was extended by a stand-alone Agilent 1260 Infinity Isocratic Pump (Waldbronn, Germany) and two VICI six-port valves (Rotterdam, The Netherlands). Figure 1 shows a schematic overview of the setup.

The injection volume of the fractionation method was set at 1 μL and the flow rate at 800 $\mu\text{L}/\text{min}$. In positive mode, the C18, WAX and SCX columns were loaded consecutively. The mobile phases consisted of 0.2% formic acid in water for loading (gradient pump: A), 2 mM ammonium acetate in methanol for the C18 elution (gradient pump: B) and 100 mM ammonium acetate pH 10 for ion exchange elution (IEX pump). In negative mode, the C18 and WAX columns were loaded consecutively. The mobile phases consisted of 2 mM ammonium acetate in water for loading (gradient pump: A), 2 mM ammonium acetate in methanol (gradient pump: B) for the C18 elution and 100 mM ammonium formate pH 10.5 for ion exchange elution (IEX pump). When the gradient pump was selected, the IEX pump pumped the solvent back to the solvent bottle. When the IEX pump was selected, the gradient pump flow was directed to waste. By using two other six-port valves, the IEX columns could be switched in and out of the line of the LC flow. The total runtime

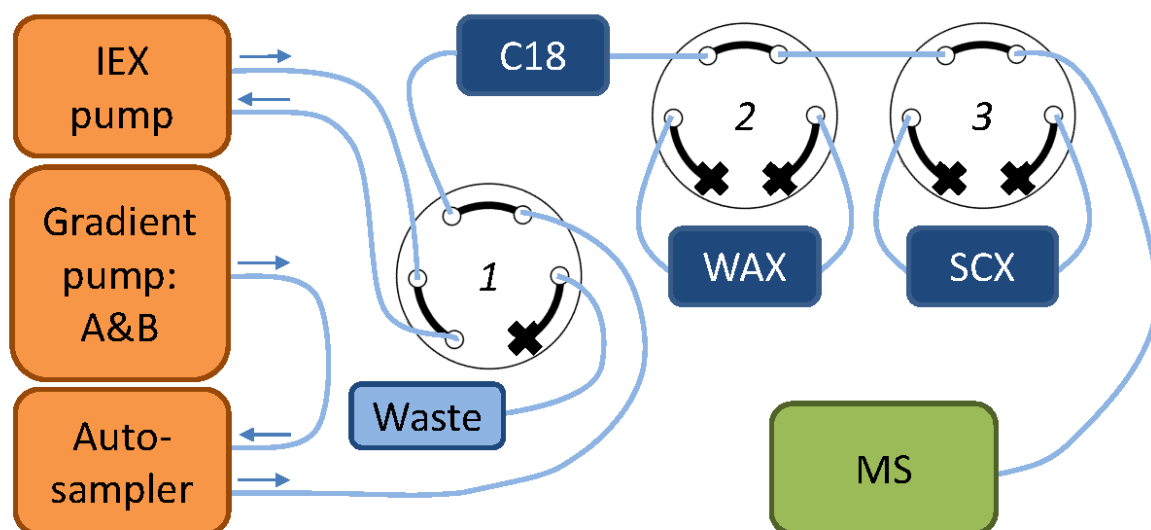


Figure 1. Schematic representation of the on-line fractionation setup. Valve 1, which was located on the mass spectrometer, was used to change between the IEX pump and the gradient pump. Valve 2 and 3, VICI valves, were used to switch the ion exchange columns in or out of line.

was 3 min and the detailed timetable of the fractionation in positive and negative mode can be found in the Supplementary information (Table S4 and S5).

The flow injection analysis (FIA) method was adapted from Carducci *et al.*¹⁷ The injection volume was set at 20 μL and the flow rate at 80 $\mu\text{L}/\text{min}$. The mobile phase consisted of 80% methanol in water. Although the mobile phase contained no additives, the sample diluent contained 0.1% acetic acid which was sufficient to promote ionization. At 0.8 minutes, the flow rate was increased to 800 $\mu\text{L}/\text{min}$ for 0.5 minutes in order to flush the system and at 1.3 minutes the flow rate returned to 80 $\mu\text{L}/\text{min}$. The total analysis time was 1.4 minutes. The MS parameters can be found in the Supplementary information (Table S6).

The data was processed in Analytics of Sciex OS 1.6. For the targeted processing, the analytes were integrated by integrating the signal of the M+H (in positive mode) and M-H (in negative mode) ion with an XIC width of 0.01 Dalton. Glucose was measured as an M+Na ion and choline was measured as an M+ ion. The untargeted data processing was performed using the 'Non-targeted Peaks' function in Analytics (see detailed information in the Supplementary information table S11C).

Effect of circadian rhythm and food intake on metabolite classes

The effect of circadian rhythm and food intake on the metabolite classes was evaluated for ten healthy male volunteers (aged 18-45 years). The clinical study was approved by the Ethical Committee of the Centre for Human Drug Research Leiden and all volunteers signed an informed consent form. The study design has previously been published.¹⁹ In short, blood samples were collected over 24 hours under uniform conditions for food intake, physical activity and night rest. At each time point, 20 mL of blood was drawn into two 10 mL BD Vacutainer® K2EDTA tubes and kept on ice. The tubes were gently inverted multiple times and centrifuged (1000 g, 15 min, 4 °C). Plasma samples were aliquoted and stored at -80 °C prior to analysis. A quality control (QC) was prepared by pooling 15 μL of every individual study sample. A QC sample was analyzed every 10 samples. Metabolites with an RSD below 15% throughout the QC samples were included in the data analysis.

Each metabolite was normalized on the first time point and subsequently log-transformed using the natural logarithm. Then, the metabolites were allocated to six different compound classes (amino acids, amines, hexose, acylcarnitines, organic acids or fatty acids). An overview of the compound classes is provided in the Supplementary information (Table S7). Within each compound class, all metabolite concentrations were averaged per time point and volunteer. A Wilcoxon Signed Rank test was used to assess the change in this mean per time point relative to

the baseline.²⁰ A multiple comparisons correction (Benjamini–Yekutieli, < 0.1) was used to adjust the p-values for multiple testing.²¹ All statistical analyses were performed in R (version 3.4.3).²²

RESULTS AND DISCUSSION

Method development

This study aims to develop an efficient and fast methodology to minimize matrix effects, focusing on salt and (phospho)lipid removal. Lipid removal was accomplished by a reversed phase column and salt removal by mixed-mode ion exchange columns. An Agilent ZORBAX Extend-C18 UPLC guard column was selected as the reversed phase column because it demonstrated superior separation and peak shape over low performance SPE columns.

Table 1 provides an overview of the performance of the evaluated ion exchangers. The grading scheme is depicted by numbers and colours indicating good (positive and green) or bad (negative and red) performances. Table 1 indicates that the WCX columns had a relatively low trapping efficiency as most of the analytes eluted at the dead time (grade 0). Most of the analytes were efficiently retained or trapped (grades 1 and 2, respectively) by the SCX columns. However, choline could not be eluted in the Hysphere column and arginine caused breakthrough (grade -1) in the Zirchrom column indicating a superior performance of the Sepax column. The right part of Table 1 shows that all SAX columns did not allow the desorption of indoxyl sulfate (grade -3) indicating that this type of anion exchanger could be exhausted over time due to the irreversible

Table 1. Evaluation of different mixed-mode cation and anion exchange columns. The grading scheme is as follows: elution at dead time: 0; retention: 1; trapped and eluted: 2; trapped and separated during elution: 3; no peak visible: -3; extreme tailing: -2; breakthrough: -1).

		Low performance		High performance			
		Hysphere	Oasis	Sepax		Zirchrom	
		Strong	Weak	Strong	Weak	Strong	Weak
Cation exchange	Leu	2	0	1	0	0	1
	Glu	2	0	1	0	2	-2
	Arg	-2	1	2	0	-1	2
	Hpx	2	0	1	0	0	0
	Choline	-3	-2	1	0	0	1
Score		1	-1	6	0	1	2
Anion exchange	Lactate	-1	1	1	1	-3	-2
	Malate	2	2	2	2	-3	-2
	Citrate	2	2	-2	2	-3	-2
	Indoxyl sulfate	-3	-3	-3	3	-3	2
	Score		0	2	-2	8	-12

binding of analytes. The Sepax WAX was suitable for all representative analytes, whereas the Oasis column was too strong (grade -3 for indoxyl sulfate) and the Zirchrom column repeatedly resulted in extreme tailing (grade -2). The Sepax SCX and WAX columns were unsurpassed in terms of retention and trapping and allowed for the analysis of all representative compounds. Therefore, we selected these columns for the trapping of the ionic species. The combination of a WAX and SCX also provided the possibility to use a similar elution buffer for both columns. The elution from a WAX column requires a high pH in order to remove the positive charge on the sorbent, whereas the high pH removes the positive charge of the analytes during the elution of an SCX column. Besides, the high pH is accomplished by the use of ammonia, which is a suitable counter-ion for an SCX column.

The silica material of the ZORBAX Extend-C18 guard column was end-capped with methyl groups which made the sorbent resistant to high pH. Therefore, this particular column could be permanently in line with the flow. In contrast, the IEX columns were switched out of the line during C18 elution because this improved retention. In negative mode, the WAX elution profile was better in the absence of the SCX column. Since the SCX column did not contribute to the reduction of ion suppression in negative mode, this column was permanently switched out of the line during the analysis in negative mode. The IEX methods were further optimized to improve retention and peak shapes and to minimize carry-over.

Fractionation characteristics

Figure 2 shows the chromatograms of a pooled plasma sample measured with the final fractionation methods in positive and negative mode. The chromatogram contains three different fractions in positive mode (flow-through: polar neutral/positive; IEX: polar positive and C18: apolar) and three fractions in negative mode (flow-through: polar neutral/negative; IEX: polar negative and C18: apolar). An overview of the fractions and charge of the analytes during loading is supplied in the Supplementary information (Table S7). The elution profile of the phospholipids in the negative fractionation method is measured in positive MS polarity (because of ionization efficiency) and shown in the Supplementary information (Figure S1). The phospholipids are separated from both the acylcarnitines and the fatty acids and therefore could not suppress their ionization. This stresses the importance of the combined on-line fractionation and separation. If these fractions were collected off-line and subsequently injected into the MS, the phospholipids would have been ionized simultaneously with the fatty acids and acylcarnitines. The salts were most likely divided over the mixed-mode ion exchangers (SCX and WAX in positive mode and WAX in negative mode) and eluted during the ion exchange elution. By allocating these known ion suppressors over different fractions, we minimized the ion suppression in a limited amount of time.

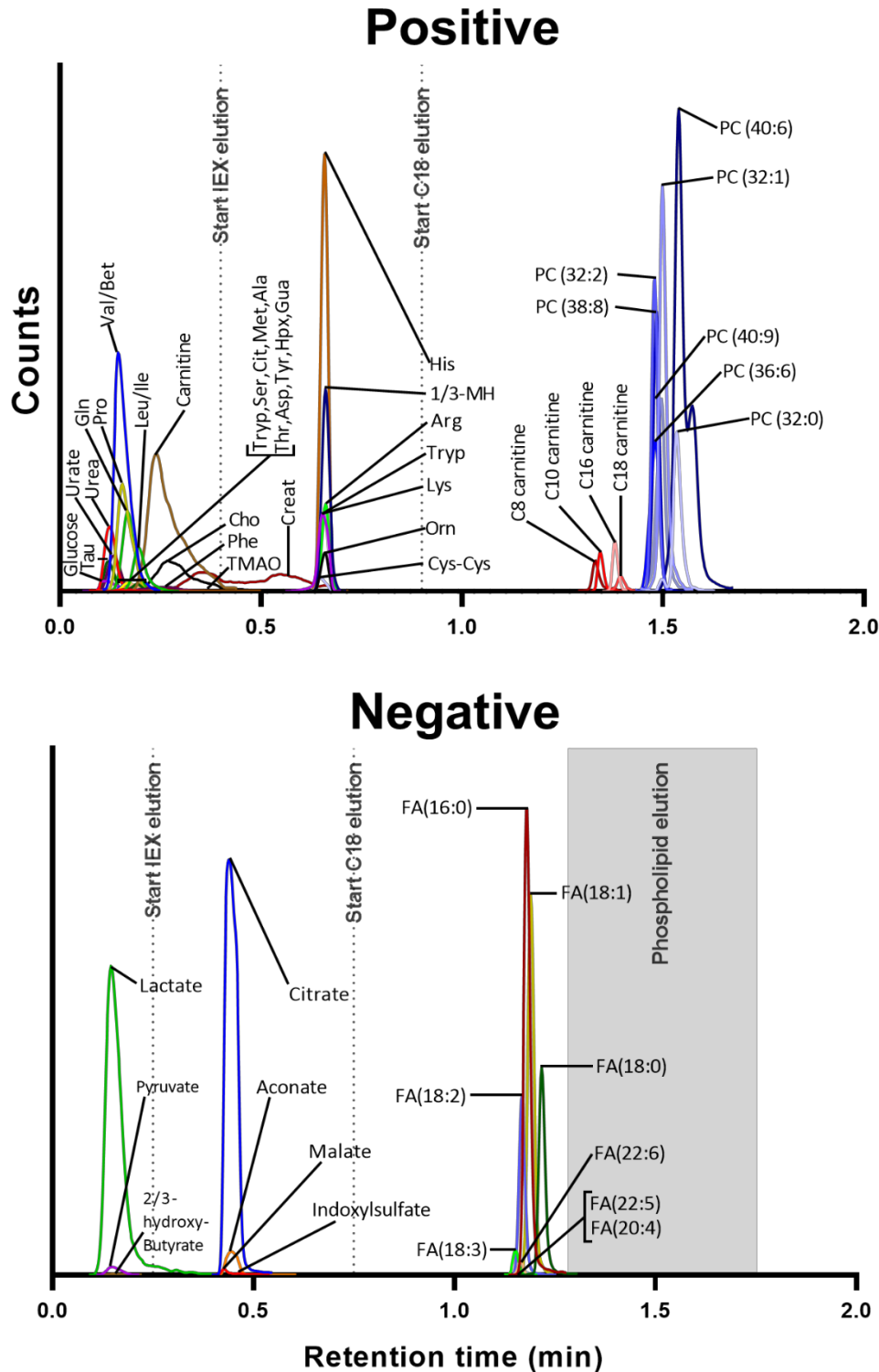


Figure 2. Extracted ion chromatogram of a pooled plasma sample measured by the fractionation method in positive and negative mode. The phospholipid elution window (phospholipid elution profile shown in Supplementary information Figure S1) in negative mode is indicated by the grey area. All the ions are measured by $M+H$ in positive mode and $M-H$ in negative mode, apart from hexose which was measured as a sodium adduct. For visualization purposes, the phospholipids and fatty acids were extracted by using the one ^{13}C m/z value.

In general, the flow-through fraction contained analytes that were polar and consisted of a zero and/or one net charge during loading. Singly-charged compounds experienced some retention in positive mode, but no retention in negative mode. The lack of retention might be explained by the counter-ion effect of the high concentrations of salts in plasma. In positive mode, a remaining negative charge on the acids might have impaired the retention of amino acids. The second fraction comprised all the components that were trapped on the ion exchange columns. A compound was efficiently trapped on the IEX column if it consisted of multiple net charges or was in equilibrium between one net charge and multiple charges at the pH during loading. The third fraction consisted of all the apolar compounds, which were efficiently trapped on and eluted from the C18 column.

Creatinine was strongly retained but not trapped on the SCX column. Creatinine had one positive net charge and two additional neutral nitrogen atoms, which could have potentially increased the interaction with the stationary phase. We did not find any other compounds that resulted in multiple peaks due to breakthrough or multiple trappings. Non-gaussian shaped peak areas were obtained by integrating the area under the curve between the two intersections with the baseline. These compounds were corrected by their corresponding internal standard because their peak shape and retention time were similar (see Supplementary information Figure S2 for the example of creatinine(-D3)). Other analytes were corrected either by their corresponding internal standard or by an internal standard that coeluted. the area under the curve between the two intersections with the baseline. These compounds were corrected by their corresponding internal standard because their peak shape and retention time were similar (see Supplementary information Figure S2 for the example of creatinine(-D3)). Other analytes were corrected either by their corresponding internal standard or by an internal standard that coeluted.

Method validation

The validation was performed by assessing the repeatability, intermediate precision, carryover, LOD, LLOQ and the matrix effect of the method. The results of the validation can be found in Supplementary information (Table S8).

The mean repeatability and intermediate precision were 6.0 and 7.1%, respectively. The relative standard deviation of 48 compounds was below 15% and two components varied more than 15%: TMAO and guanine. This was most likely caused by the low signal of these analytes due to the low physiological concentration and the low molecular weight. In total, 1071 injections were performed on the same set of columns with a sufficient repeatability as is shown in the validation (first injections) and biological application (last injections). The coefficient of determination (R^2) was on average 0.995, which indicated a good linearity of the fractionation method. The linearity of 47 compounds was higher than 0.99 and three compounds revealed a linearity lower than 0.99.

The linearity of C16- and C18-carnitine was compromised by matrix interferences since a calibration curve constructed in water demonstrated a sufficient linearity (>0.99). All the acylcarnitines were corrected by the same internal standard, i.e. octanoylcarnitine-d3. This internal standard corrected well for co-eluting analytes C8- and C10-carnitine. C16- and C18-carnitine were more strongly retained and eluted further away from the internal standard and closer to the (phospho)lipids. Therefore, the linearity of these analytes would be improved by the correction of a more apolar internal standard. The lower linearity of docosapentanoic acid was found for both plasma and water samples. The reason for this was unknown.

The LOD and LLOQ were determined by spiking several internal standards in plasma. This was done because the analytes of interest were endogenous and differences in chromatography were observed between water and plasma samples. Figure 3 demonstrates that physiological blood levels as reported in literature were higher than the calculated LLOQ indicating a sufficient sensitivity of the method. The average carryover was 0.5% when a blank sample was measured after a QC sample. In total 48 compounds demonstrated a lower carryover than 2%. There were two compounds with a higher carryover: methionine (5.3%) and decanoylcarnitine (2.4%). The carryover of methionine can be explained by the fact that sulfur sticks to stainless steel.²³ The reason for the carryover of decanoylcarnitine was unclear. Although a slight carryover has been observed, we expect no problems with respect to the quantification of study samples. The analytes of interest are endogenous compounds, which are present in every studied person. This will ensure that a small carryover will have a limited effect on the quantification values of the analytes.

Fractionation versus flow injection analysis and conventional liquid chromatography

In order to demonstrate the cleanup efficiency of the fractionation method, we measured spiked internal standards in plasma and water. Hereby, the matrix effect, ion suppression and LLOQ were determined for the fractionation and an FIA method. Figure 3 shows that the mean ion suppression of the fractionation method was 25%, whereas the mean ion suppression in the FIA method was 89%. We have previously reported the effects of salts and phospholipids on the ESI.¹¹ The fractionation method provides a fast solution to minimize ion suppression caused by these matrix interferences. The use of three orthogonal columns allocated phospholipids, negative and positive salts into three different fractions. The on-line elution into the MS and the use of high performance SPE columns allowed for the separation between analytes and matrix interferences within a fraction. An additional LLE step prior to the FIA decreased the ion suppression to 80% (see Supplementary information Table S9). This decrease in ion suppression was predominantly observed for compounds in the apolar fraction, i.e. fatty acids and acylcarnitines. However, the ion suppression of these compounds was still considerably less in our fractionation method. For

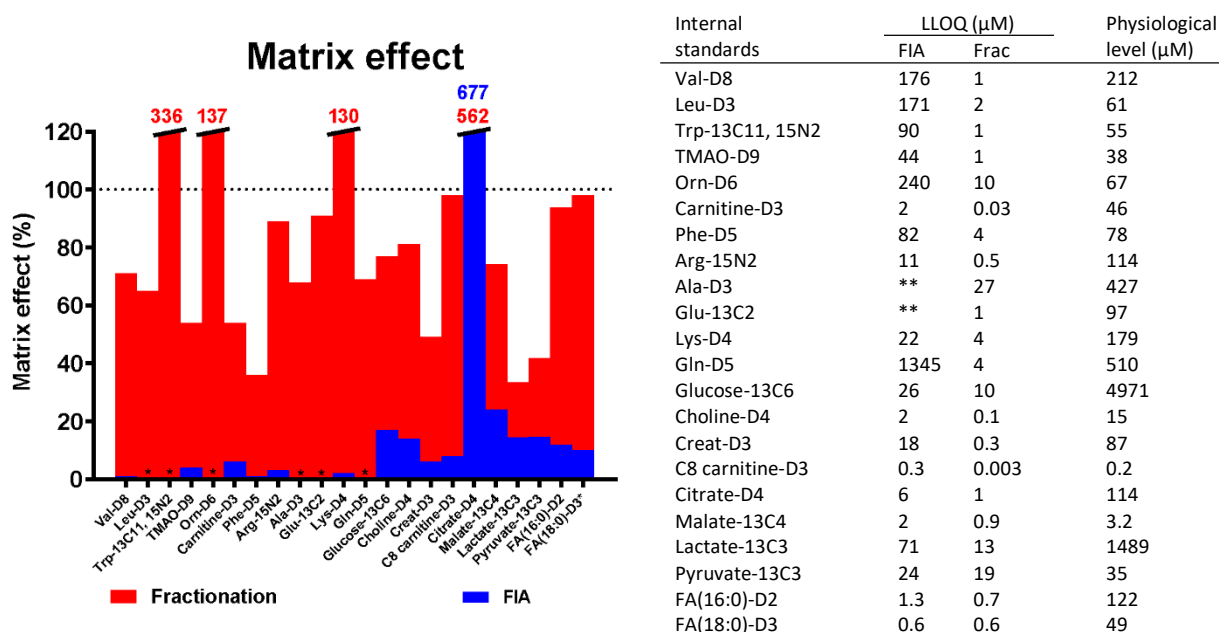


Figure 3: Performance comparison of the fractionation (Frac) method and flow injection analysis (FIA). The graph shows the matrix effect for each internal standard measured by either the fractionation method (red) or FIA (blue). Compounds with 0% matrix effect (indicated by *) were not detected at C4 levels. Compounds that experienced ion enhancement (matrix effect > 100%) were cut off at a matrix effect of 120% (values are indicated in corresponding colours). The table on the right shows the lower limit of quantification (LLOQ) of FIA and fractionation as well as the physiological plasma levels (HMDB values).²⁹ (**=not detected at C7 levels)

metabolites in the polar fraction, the ion suppression was comparable with FIA without LLE. LLE demonstrates little cleanup efficiency because samples are only fractionated based on polarity and the obtained fractions are analyzed at once without further separation.

The fractionation method demonstrated a superior sensitivity in comparison with FIA. The mean LLOQ of the fractionation method was 21 times lower which ensured a sufficient sensitivity to measure physiological levels in plasma. In contrast, 9 out of 22 analytes could not be quantified using the FIA method due to insufficient sensitivity (LLOQ higher than physiological levels). The substantial difference in ion suppression was most likely responsible for the differences in sensitivity. The performance improvement was mainly reflected in positive mode. In negative mode, the improvement in ion suppression and sensitivity was smaller. This is in accordance to other studies, in which was shown that ion suppression is less occurring in negative ionization mode.^{24,25} Although the FIA method is faster (1.4 versus 3 min), the findings in Figure 3 emphasize the necessity of on-line fractionation prior to electrospray ionization.

We have also compared the LLOQ of the ISTDs with the LLOQ of conventional LC-MS analyses reported in literature (see Supplementary information Table S10). These findings demonstrated

that the sensitivity of fractionation and LC-MS is in a similar range. This was also expected because of the limited ion suppression in the fractionation method and a comparable peak width, injection volume and flow rate with regards to general LC-MS. However, differences in, for example, LLOQ determinations, used mass spectrometer (tandem and high-resolution) and derivatization might complicate this comparison. It does indicate that we are at least in a comparable sensitivity range relative to LC-MS. This is also emphasized by the coverage of the fractionation method in comparison with conventional reversed phase (RP) and hydrophilic interaction chromatography (HILIC) separations. The number of unique retention time and m/z features was 2289, 3475 and 3529 for fractionation, RP and HILIC, respectively (the methodologies are presented in the Supplementary information Table S11). The difference in coverage is mostly explained by the additional isomeric separation that is experienced in conventional chromatography as the number of unique m/z features was practically similar (2089, 2465 and 2325 for fractionation, RP and HILIC, respectively).

Our fractionation approach enables the analysis of multiple compound classes in 3 minutes per polarity, whereas conventional LC-MS usually requires a gradient time of around 3-30 minutes per compound class (see Table S10). The analysis time of LC-MS can be reduced by the use of faster gradients. However, in order to realize a comprehensive targeting of the metabolome, multiple LC separations would be needed (e.g. HILIC and RP for polar and apolar, respectively). The inclusion of multiple chromatographic gradients drastically decreases the overall throughput of the analysis. Moreover, the equilibration and flushing time of conventional LC columns (3-15 cm) is substantially higher in comparison with short chromatographic columns (0.5-1 cm). The benefit of an integrated fractionation approach is due to the use of multiple short chromatographic columns, which allow for an efficient separation, while little time is spent on gradients and column equilibration/flushing. The challenge of using a fractionation approach instead of conventional chromatography is the lack of isomeric separation. This could be overcome by the use of ion-mobility and MS/MS experiments.

Effect of circadian rhythm and food intake on metabolite classes

It is known that there are trends in metabolite levels due to the circadian rhythm and food intake.²⁶ These fluctuations are important to take into account when metabolites are studied or used as biomarkers. Different sampling times throughout the day could cause variations in metabolite levels that are not attributable to a studied disease or intervention. For this, we profiled ten healthy volunteers on ten different time points on a time scale of 24 hours. This study should clarify the significance of these diurnal changes.

After the data acquisition, 47 compounds were included in the data analysis and three compounds were excluded. Fatty acid 16:0 and 18:0 had an RSD of more than 15% due to fluctuating background levels. C18 carnitine also had an RSD of more than 15%. The reason for this was unclear. Figure 4 shows that our validated platform allowed us to demonstrate significant changes of metabolite classes throughout the day (false discovery rate (FDR) adjusted p-values are listed in the Supplementary information Table S12). All compound classes changed significantly from the baseline, apart from the amines. The amines (quaternary amines, creatinine, urea and uric acid) did not reveal a significant difference over a period of 24 hours. This is in accordance with our prior work, in which we demonstrated that gut metabolites (quaternary amines) were not affected by the fasting state of an individual.¹¹ The amino acid levels started to rise after wake time. The levels remained high throughout the morning/afternoon and decreased again towards baseline levels just before dinner. After dinner, the amino acids increased again and subsequently returned to baseline levels during night rest. The increase in amino acids after wake time and in the afternoon/evening has also been observed in prior studies.²⁰

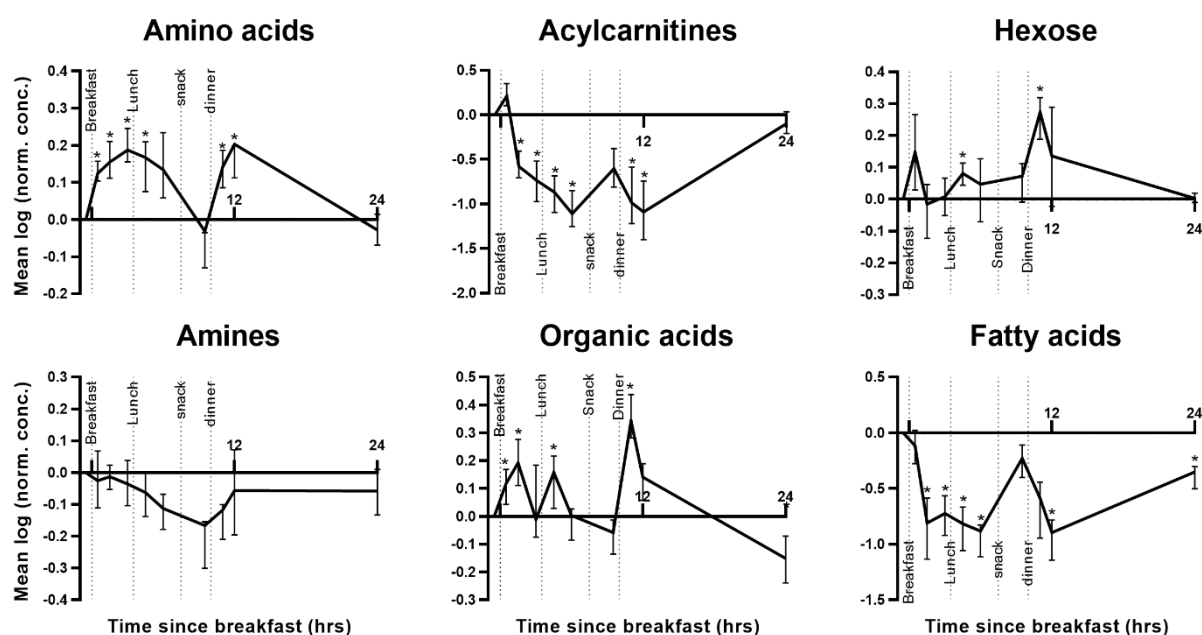


Figure 4. The mean natural logarithm of normalized metabolite concentrations over time. Within each compound class, metabolites were averaged per time point and volunteer. The mean of these curves over the 10 volunteers are depicted and the pointwise interquartile range (IQR) of the volunteers is presented in the error bars. Time points that are significantly different from the baseline are indicated (* FDR adjusted p-values < 0.1). The time frame comprises four standardized feeding times and meals and one night rest. The time is presented with respect to the breakfast time.

The hexose and organic acid levels significantly increased after the feeding times (except for hexose after breakfast which did not reach FDR corrected significance). When sugar is available, glucose is the main source of the citric acid cycle. This explains the similarities of the hexose and organic acid trends since organic acids are the main constituents in the citric acid cycle. The fatty acids concentrations decreased throughout the day and increased just before dinner and after 24 h, which has been observed before.²⁶ During (overnight) fasting, glucose is mainly depleted, switching the main energy source to fatty acids. In this state, fatty acids are released from triglycerides by lipolysis, which explains the high fatty acid levels prior to dinner and after a night rest.²⁷ In order to accommodate the increased demand for fatty acids, acylcarnitines are put in place to transport the fatty acids into the mitochondria for β -oxidation.²⁸ This explains the similarities between the fatty acid and acylcarnitine profile. Sampling time is an indispensable parameter to take into account when metabolites are used or studied as biomarkers. Food intake and circadian rhythm significantly change compound classes from baseline levels. Therefore, sampling times and fasting states should be standardized when metabolites are used for diagnosis, clinical studies or biomarker discovery. This should further strengthen the use of discovered metabolite biomarkers in personalized health care.

CONCLUSIONS

Although much progress has been made in the analysis of metabolites, fast and global profiling of the metabolome in complex matrices remains a challenging aspect. For this purpose, we demonstrated a fast and comprehensive fractionation method coupled on-line to mass spectrometry. The three serially coupled high performance SPE columns resulted in a fractionation based on polarity, charge, and removed important ion suppressors from different fractions. The on-line and orthogonal setup realized a flow-through which was cleaned by three different sorbent chemistries and a within-fraction separation of analytes and ion suppressors. The comparison with FIA emphasized the performance improvement achieved with the fractionation method. In a limited amount of time, the fractionation method drastically lowered the ion suppression as well as the detection limits. The on-line fractionation demonstrated similar quantification limits in comparison to the conventional LC-MS analyses. This proves that on-line fractionation enables the analysis of a large range of metabolites without suffering in terms of sensitivity. The developed fractionation method was able to demonstrate fluctuations of metabolite classes in blood samples from healthy volunteers on different time points throughout the day, which could be explained by underlying metabolic processes. These significant diurnal variations are important for clinicians when metabolites are used as biomarkers. Standardized sampling times and fasting states should minimize variations caused by food intake and circadian rhythm on the disease or intervention related variations. This work provides a methodology to

target multiple metabolite classes within a single analytical platform without suffering in terms of analysis time. This development in comprehensive and fast metabolite screening should encourage researchers and clinicians to make full use of the field of metabolomics and to further investigate the value of potential prognostic and diagnostic biomarker metabolites.

ACKNOWLEDGEMENTS

The authors are grateful to receive funding for this research from The Netherlands Organization for Scientific Research (NWO) in the framework of the Technology Area TA-COAST (Fund New Chemical Innovations. project no. 053.21.118). This research was also part of the Netherlands X-omics Initiative and partially funded by NWO, project 184.034.019.

REFERENCES

1. Mussap, M., Zaffanello, M. & Fanos, V. Metabolomics: a challenge for detecting and monitoring inborn errors of metabolism. **6**, (2018).
2. Balashova, E., Maslov, D. & Lokhov, P. A Metabolomics Approach to Pharmacotherapy Personalization. *J. Pers. Med.* **8**, 28 (2018).
3. Kaushik, A. K. & DeBerardinis, R. J. Applications of metabolomics to study cancer metabolism. *Biochim. Biophys. Acta - Rev. Cancer* **1870**, 2–14 (2018).
4. Miggliels, P., Wouters, B., Westen, G. J. Van, Dubbelman, A. & Hankemeier, T. Novel technologies for metabolomics: more for less. *TrAC - Trends Anal. Chem.* **2018**, 1–9 (2018).
5. Gowda, G. A. N. & Djukovic, D. Overview of Mass Spectrometry-Based Metabolomics: Opportunities and Challenges. *Methods Mol Biol* **1198**, 3–12 (2014).
6. Ismaiel, O. A., Halquist, M. S., Elmanly, M. Y., Shalaby, A. & Karnes, H. T. Monitoring phospholipids for assessment of matrix effects in a liquid chromatography-tandem mass spectrometry method for hydrocodone and pseudoephedrine in human plasma. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **859**, 84–93 (2007).
7. Enot, D. P. *et al.* Preprocessing, classification modeling and feature selection using flow injection electrospray mass spectrometry metabolite fingerprint data. *Nat. Protoc.* **3**, 446–470 (2008).
8. Nanita, S. C. & Kaldon, L. G. Emerging flow injection mass spectrometry methods for high-throughput quantitative analysis. *Anal. Bioanal. Chem.* **408**, 23–33 (2016).
9. Zhang, J. *et al.* High-throughput salting-out assisted liquid/liquid extraction with acetonitrile for the simultaneous determination of simvastatin and simvastatin acid in human plasma with liquid chromatography. *Anal. Chim. Acta* **661**, 167–172 (2010).
10. Zhang, X. *et al.* SPE-IMS-MS: An automated platform for sub-sixty second surveillance of endogenous metabolites and xenobiotics in biofluids. *Clin. Mass Spectrom.* **2**, 1–10 (2016).
11. van der Laan, T. *et al.* Fast LC-ESI-MS/MS analysis and influence of sampling conditions for gut metabolites in plasma and serum. *Sci. Rep.* **9**, 12370 (2019).
12. Trivedi, D. K., Hollywood, K. A. & Goodacre, R. Metabolomics for the masses: The future of metabolomics in a personalized world. *New Horizons Transl. Med.* **3**, 294–305 (2017).
13. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
14. Mayo Clinic, M. M. L. Rochester 2018 Interpretive Handbook. (2018). doi:10.1016/S1002-0721(13)60014-9
15. Wishart, D. S. *et al.* HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
16. Matuszewski, B. K., Constanzer, M. L. & Chavez-Eng, C. M. Strategies for the assessment of

- matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. *Anal. Chem.* **75**, 3019–30 (2003).
17. Carducci, C. *et al.* Quantitative determination of guanidinoacetate and creatine in dried blood spot by flow injection analysis-electrospray tandem mass spectrometry. *Clin. Chim. Acta* **364**, 180–187 (2006).
 18. Bligh, E.G. and Dyer, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, (1959).
 19. Duisters, K. *et al.* Intersubject and Intrasubject Variability of Potential Plasma and Urine Metabolite and Protein Biomarkers in Healthy Human Volunteers. *Clin. Pharmacol. Ther.* **0**, 1–9 (2019).
 20. Thompson, D. K. *et al.* Daily Variation of Serum Acylcarnitines and Amino Acids. *Metabolomics* **8**, 556–565 (2012).
 21. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
 22. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.r-project.org/>.
 23. Nagu, M., Abdulhadi, A., Huwaiji, A. & Alanazi, N. M. Effect of Elemental Sulfur on Pitting Corrosion of Steels Muthukumar. *Insights Anal. Electrochem.* **4**, 1–7 (2018).
 24. Ghosh, C., P. Shinde, C. & S. Chakraborty, B. Ionization Polarity as a Cause of Matrix Effects, its Removal and Estimation in ESI-LC-MS/MS Bio-analysis. *J. Anal. Bioanal. Tech.* **01**, 1–7 (2010).
 25. Oldekop, M. L., Rebane, R. & Herodes, K. Dependence of matrix effect on ionization polarity during LC-ESI-MS analysis of derivatized amino acids in some natural samples. *Eur. J. Mass Spectrom.* **23**, 245–253 (2017).
 26. Dallmann, R., Viola, A. U., Tarokh, L., Cajochen, C. & Brown, S. A. The human circadian metabolome. *Proc. Natl. Acad. Sci.* **109**, 2625–2629 (2012).
 27. Ali, A. H. *et al.* Free fatty acid storage in human visceral and subcutaneous adipose tissue: Role of adipocyte proteins. *Diabetes* **60**, 2300–2307 (2011).
 28. Kompare, M. & Rizzo, W. B. Mitochondrial Fatty-Acid Oxidation Disorders. *Semin. Pediatr. Neurol.* **15**, 140–149 (2008).
 29. Human Metabolome Database. Available at: <http://www.hmdb.ca/>.

SUPPLEMENTARY INFORMATION

Phospholipids in negative fractionation

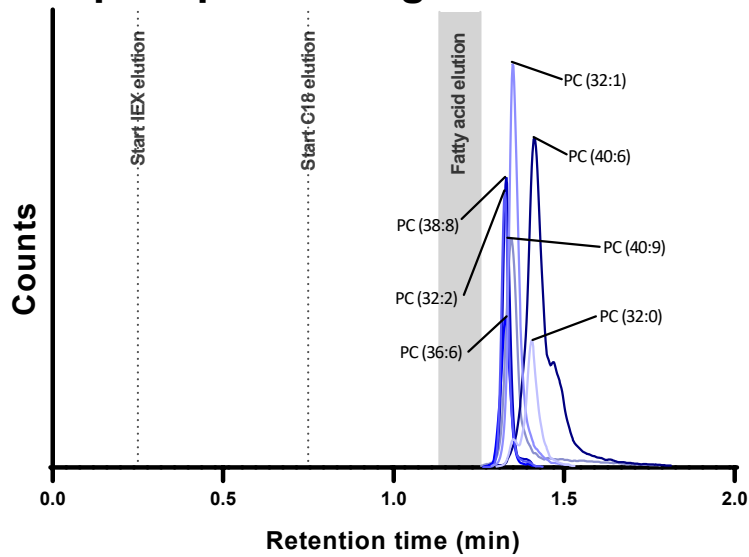


Figure S1. Extracted ion chromatogram of phospholipids in a pooled plasma sample measured by the negative fractionation method and positive MS polarity. The elution window of the fatty acids are indicated by the grey area.

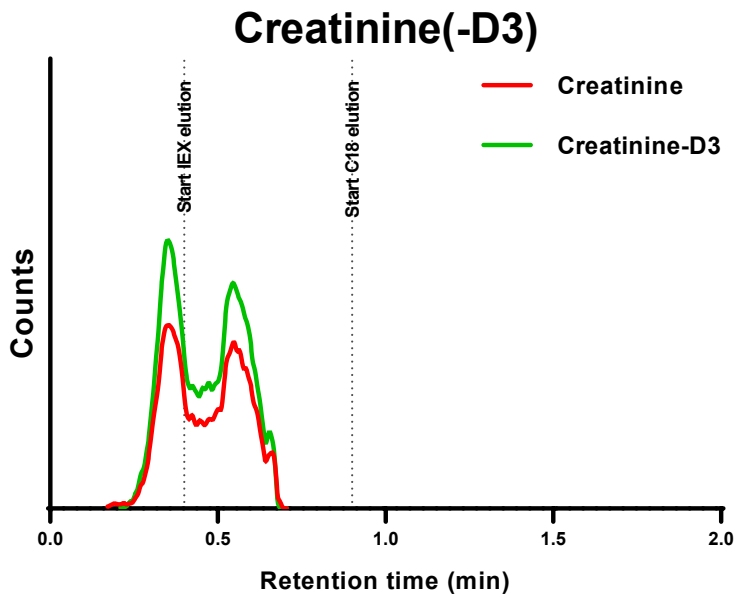


Figure S2. Extracted ion chromatogram of creatinine and creatinine-D3 measured by the fractionation method in a pooled plasma sample in positive mode. The peak shape and retention time are overlapping.

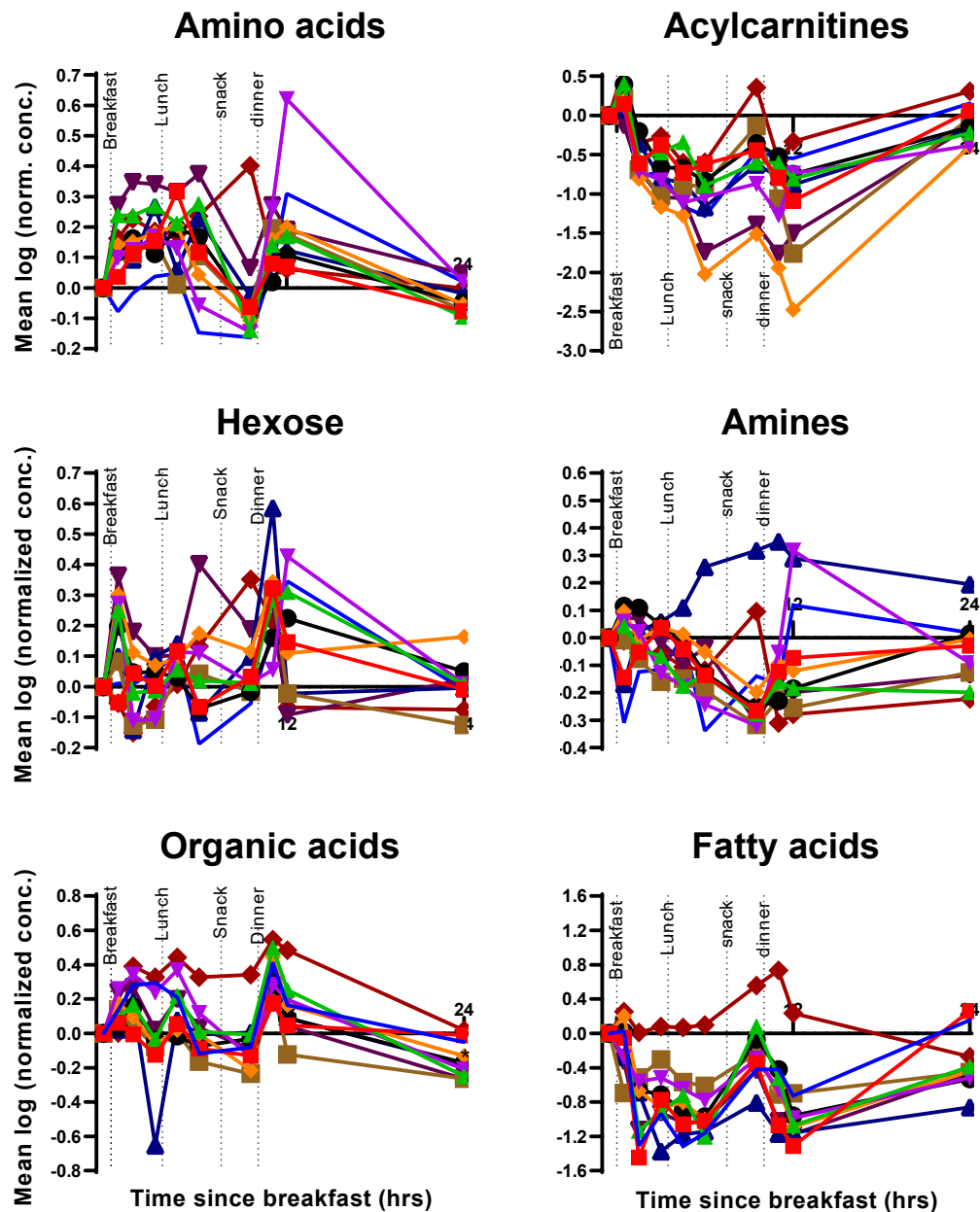


Figure S3. The individual mean natural logarithm of metabolite concentrations over time. Normalization was performed on the first time point. Within each compound class, metabolites were averaged per time point. The 10 volunteers are depicted in different colours. The time frame comprises four standardized feeding times and meals and one night rest. The time is presented with respect to the breakfast time.

Table S1. Calibration standards.

Name	CheBi ID	Standard	Supplier	Solvent	C7 (μ M)
1-methylhistidine	70958	1-methylhistidine	HMDB	Water	80
3-methylhistidine	70959	3-methylhistidine	HMDB	Water	40
Alanine	16977	L-alanine	Sigma-Aldrich	Water	1600
Arginine	16467	L-arginine hydrochloride	Sigma-Aldrich	Water	400
Betaine	17750	Betaine hydrochloride	Sigma-Aldrich	Water	300
Carnitine	16347	L-Carnitine hydrochloride	Sigma-Aldrich	Water	200
Citrulline	16349	Citrulline	HMDB	Water	100
Cystine	16283	L-cystine	Fluka	1M Hydrochloric acid	100
Glutamic acid	16015	L-glutamic acid	Sigma-Aldrich	Water	400
Glutamine	18050	L-glutamine	Fluka	Water	2000
Histidine	15971	L-histidine monohydrochloride monohydrate	Fluka	Water	600
Isoleucine	17191	L-Isoleucine	Fluka	Water	200
Leucine	15603	L-Leucine	Fluka	Water	200
Lysine	18019	L-lysine monohydrochloride	Fluka	Water	800
Methionine	16643	L-methionine	Fluka	Water	200
Ornithine	15729	L-Ornithine hydrochloride	Sigma-Aldrich	Water	400
Phenylalanine	17295	L-phenylalanine	Fluka	Water	300
Proline	17203	L-proline	Fluka	Water	800
Serine	17115	L-serine	Fluka	Water	600
Taurine	15891	Taurine	Fluka	Water	400
Threonine	16857	L-threonine	Fluka	Water	600
Tryptophan	16828	L-tryptophan	Sigma-Aldrich	0.5M hydrochloric acid	200
Tyrosine	17895	L-tyrosine	Fluka	1M hydrochloric acid	200
Valine	16414	L-valine	Fluka	Water	800
TMAO	15724	Trimethylamine-N-oxide dihydrate	Sigma-Aldrich	Water	100
Choline	15354	Choline chloride	Sigma-Aldrich	Water	100
Creatinine	16737	Creatinine	Sigma-Aldrich	Water	400
Urea	16199	Urea	Sigma-Aldrich	Water	10000
Glucose	17634	D-(+)-glucose	Sigma-Aldrich	Water	10000
Guanine	16235	Guanine	HMDB	0.1M sodium hydroxide	8
Hypoxanthine	17368	Hypoxanthine	Sigma-Aldrich	1M sodium hydroxide	200
Uric acid	17775	Uric acid	Sigma-Aldrich	1M sodium hydroxide	2000
Octanoylcarnitine	18102	Octanoyl -L-carnitine HCl	VU medical center	Methanol	1
Decanoylcarnitine	68830	Decanoyl-L-carnitine HCl	VU medical center	Methanol	1
Palmitoylcarnitine	17490	Hexadecanoyl-L-carnitine HCL	VU medical center	Methanol	0,4
Stearoylcarnitine	84644	Octadecanoyl-L-carnitine HCl	VU medical center	Methanol	0,2
Aspartic acid	22660	DL-aspartic acid	Sigma-Aldrich	1M sodium hydroxide	160
Citric acid	30769	Citric acid	Sigma-Aldrich	Water	800
Lactic acid	422	lactic acid lithium salt	Acros organics	Water	3000
Isocitric acid	30887	Isocitric acid	HMDB	Water	40
Pyruvic acid	32816	sodium pyruvate	Sigma-Aldrich	Water	400
Aconitic acid	32805	Cis-aconitic acid	Sigma-Aldrich	Water	225
Alpha-ketoglutaric acid	30915	Alpha-ketoglutaric acid disodium salt hydrate	Fluka	Water	80
2-Hydroxybutyric acid	1148	2-Hydroxybutyric acid sodium salt (97%)	Sigma-Aldrich	Water	400
3-Hydroxybutyric acid	20067	3-hydroxybutyric acid	Sigma-Aldrich	Water	800
Malic acid	6650	DL-Malic acid	Sigma-Aldrich	Water	120
Indoxyl sulfate	43355	indoxyl sulfate potassium salt	Sigma-Aldrich	Methanol	20
Alpha-linolenic acid	27432	Alpha-linolenic acid	Cayman chemical	Ethanol	10
Docosahexaenoic acid	28125	Docosahexaenoic acid	Cayman chemical	Ethanol	4
Docosapentaenoic acid	65136	Docosapentaenoic acid	Cayman chemical	Ethanol	2
Linoleic acid	17351	Linoleic acid	Sigma-Aldrich	Methanol	200
Arachidonic acid	15843	Arachidonic acid	Cayman chemical	Ethanol	30
Palmitic acid	15756	Palmitic acid	HMDB	Methanol	200
Oleic acid	16196	Oleic acid	Sigma-Aldrich	Ethanol	200
Stearic acid	28842	Stearic acid	Sigma-Aldrich	Isopropanol	100

Table S2. Internal standards. The C6 values are the concentrations (μM) which were present in the internal standard solution. During sample preparation, the same volume of internal standard solution and plasma is added.

Name	Standard	Supplier	Solvent	C6 (μM)
Phenyl alanine	DL-Phenyl-d5-alanine	CDN isotopes	Water	320
Trimethylamine-N-oxide	Trimethylamine N-oxide-D9	Cambridge Isotope	Water	160
Alanine	DL-ALANINE-2,2,3,3-d3, 99.8% D	CDN isotopes	Water	1600
Choline	Choline-D4 Chloride	CDN Isotopes	Water	60
Carnitine	L-carnitine-d3 HCl (methyl-d3)	CDN Isotopes	Water	180
Creatinine	Creatinine-D3	Cambridge Isotope	Water	360
Valine	L-VALINE (D8)	Cambridge isotopes	Water	800
Leucine	DL-Leucine-D3	CDN Isotopes	1M hydrochloric acid	240
Ornithine	L-Ornithine-D6 hydrochloride	CDN Isotopes	Water	280
Hypoxanthine	Hypoxanthine-D2	CDN Isotopes	1M sodium hydroxide	140
Glutamine	L-GLUTAMINE (2,3,3,4,4-D5)	Cambridge isotopes	Water	2000
Lysine	L-LYSINE 2 HCL (4,4,5,5-D4)	Cambridge isotopes	Water	720
Glutamic acid	L-GLUTAMIC ACID (1,2-13C2)	Cambridge isotopes	1M sodium hydroxide	400
Arginine	L-Arginine-15N2 hydrochloride	Cortecnet	Water	400
Tryptophan	L-TRYPTOPHAN (U-13C11, U15N2)	Cambridge isotopes	Water	200
Glucose	Glucose-13C6	Cortecnet	90% methanol	20000
Octanoylcarnitine	Octanoyl carnitine-D3 HCl	CDN Isotopes	Methanol	0,8
Pyruvic acid	Pyruvic acid-13C3	Omicron Biochemicals	Water	140
Lactic acid	Lactic acid-13C3	Biomedical Isotopes	Water	4000
Malic acid	Malic acid-13C4	Cambridge Isotope	Water	24
Citric acid	Citric acid-D4	Cambridge Isotope	Water	480
Palmitic acid	Palmitic acid-D2	Cambridge Isotope	Ethanol	480
Stearic acid	Stearic acid-D3	Cambridge Isotope	Ethanol	200

Table S3. Solid-phase extraction column information.

Brand	Type	Phase	Functional group	Amount of sorbent (mg)	Particle size (μm)	pH stability	Loading buffer	Elution buffer
Hysphere	SCX	PS-DVB	Sulfonate	13	10	1-14	0.1% formic acid	100 mM ammonium acetate pH 10
Hysphere	SAX	PS-DVB	Quaternary amine	13	10	1-14	5 mM ammonium acetate	1% formic acid
Oasis	WCX	Oasis HLB	Carboxylate	10.4	30	0-14	5 mM ammonium acetate	1% formic acid
Oasis	WAX	Oasis HLB	tertiary/secondary amine	10.4	30	0-14	5 mM ammonium acetate	100 mM ammonium formate pH 10
Sepax	SCX	PS-DVB	Sulfonate	Not disclosed	3	2-12	0.1% formic acid	100 mM ammonium acetate pH 10
Sepax	SAX	PS-DVB	Quaternary amine	Not disclosed	1.7	2-12	5 mM ammonium acetate	1% formic acid
Sepax	WCX	PS-DVB	Carboxylate	Not disclosed	1.7	2-12	5 mM ammonium acetate	1% formic acid
Sepax	WAX	PS-DVB	Tertiary amine	Not disclosed	5	2-12	5 mM ammonium acetate	100 mM ammonium formate pH 10
Zirchrom	SCX	Zirconia dioxide	Ethylenediamine-N,N'-tetramethylphosphonic acid	150	3	1-10	0.1% formic acid	100 mM ammonium acetate pH 10
Zirchrom	SAX	Zirconia dioxide	Polyethyleneimine	150	3	1-12	5 mM ammonium acetate	1% formic acid
Zirchrom	WCX	Zirconia dioxide	Phosphate	150	3	1-10	5 mM ammonium acetate	1% formic acid
Zirchrom	WAX	Zirconia dioxide	Polyethyleneimine	150	3	3-9	5 mM ammonium acetate	100 mM ammonium formate pH 9

Table S4. Fractionation LC and valve parameters for positive mode.

Time (min)	Flow rate (ml/min)	%A	%B	IEX mobile phase	WAX and SCX columns
0.00	0.8	100	0	Out	In
0.40	0.8	100	0	In	In
1.00	0.8	100	0	Out	In
1.01	0.8	0	100	Out	Out
2.10	0.8	0	100	Out	Out
2.11	0.8	100	0	Out	Out
2.35	0.8	100	0	Out	In
3.00	0.8	100	0	Out	In

Table S5. Fractionation LC and valve parameters for negative mode.

Time (min)	Flow rate (mL/min)	%A	%B	IEX mobile phase	WAX column
0.00	0.5	100	0	Out	In
0.25	0.5	100	0	In	In
0.30	0.5	100	0	In	In
0.31	0.8	100	0	In	In
0.75	0.8	100	0	Out	Out
0.76	0.8	20	80	Out	Out
0.95	0.8	15	85	Out	Out
0.96	0.8	0	100	Out	Out
2.00	0.8	0	100	Out	Out
2.01	0.8	100	0	Out	Out
2.25	0.8	100	0	Out	In
3.00	0.8	100	0	Out	In

Table S6. Mass spectrometry parameters.

Platform	Polarity	Gas 1 (psi)	Gas 2 (psi)	Curtain gas (psi)	Temp (°C)	Spray voltage (V)	Declustering Potential (V)	Collision energy (eV)	Mass range (m/z)
Fractionation, RP and HILIC	Positive	40	60	40	650	5500	80	5	50-880
	Negative	40	60	40	650	-4500	-60	-5	50-880
FIA	Positive	40	40	30	550	5500	80	5	50-500
	Negative	40	40	30	550	-4500	-80	-5	50-500

Table S7. Overview of the different fractions.

Compounds	pH at loading	Method	Net charge during loading	Fraction	Compound class
Valine/Betaine	2.5	Positive	+/-neutral	Flow-through	Amino acids/amines
Leucine/Isoleucine	2.5	Positive	+/-neutral	Flow-through	Amino acids
Tryptophan	2.5	Positive	+/-neutral	SCX	Amino acids
Trimethylamine-N-oxide	2.5	Positive	Neutral	Flow-through	Amines
Ornithine	2.5	Positive	+/-++	SCX	Amino acids
Carnitine	2.5	Positive	+/-neutral	Flow-through	Amines
1/3-methylhistidine	2.5	Positive	+/-++	SCX	Amino acids
Phenylalanine	2.5	Positive	+/-neutral	Flow-through	Amino acids
Serine	2.5	Positive	+/-neutral	Flow-through	Amino acids
Citrulline	2.5	Positive	+/-neutral	Flow-through	Amino acids
Methionine	2.5	Positive	+/-neutral	Flow-through	Amino acids
Arginine	2.5	Positive	+/-++	SCX	Amino acids
Alanine	2.5	Positive	+/-neutral	Flow-through	Amino acids
Cystine	2.5	Positive	+/-++/neutral	SCX	Amino acids
Aspartic acid	2.5	Positive	+/-neutral	Flow-through	Amino acids
Glutamic acid	2.5	Positive	+/-neutral	Flow-through	Amino acids
Histidine	2.5	Positive	+/-++	SCX	Amino acids
Lysine	2.5	Positive	+/-++	SCX	Amino acids
Proline	2.5	Positive	+/-neutral	Flow-through	Amino acids
Threonine	2.5	Positive	+/-neutral	Flow-through	Amino acids
Urea	2.5	Positive	Neutral	Flow-through	Amines
Glutamine	2.5	Positive	+/-neutral	Flow-through	Amino acids
Glucose	2.5	Positive	Neutral	Flow-through	Hexose
Tyrosine	2.5	Positive	+/-neutral	Flow-through	Amino acids
Choline	2.5	Positive	+	Flow-through	Amines
Hypoxanthine	2.5	Positive	+/-neutral	Flow-through	Purines
Guanine	2.5	Positive	+/-neutral	Flow-through	Purines
Uric acid	2.5	Positive	Neutral	Flow-through	Amines
Creatinine	2.5	Positive	+	Flow-through/SCX	Amines
Taurine	2.5	Positive	Neutral	Flow-through	Amino acids
Palmitoylcarnitine	2.5	Positive	+/-neutral	C18	Acylcarnitines
Stearoylcarnitine	2.5	Positive	+/-neutral	C18	Acylcarnitines
Decanoylcarnitine	2.5	Positive	+/-neutral	C18	Acylcarnitines
Octanoylcarnitine	2.5	Positive	+/-neutral	C18	Acylcarnitines
(Iso)Citric acid	7	Negative	-3	WAX	Organic acids
Malic acid	7	Negative	-2	WAX	Organic acids
Lactic acid	7	Negative	-	Flow-through	Organic acids
Pyruvic acid	7	Negative	-	Flow-through	Organic acids
Aconitic acid	7	Negative	-3	WAX	Organic acids
Alpha-ketoglutaric acid	7	Negative	-2	WAX	Organic acids
2/3-Hydroxybutyric acid	7	Negative	-	Flow-through	Organic acids
Indoxylsulfate	7	Negative	-(sulfate)	WAX	Organic acids
Palmitic acid	7	Negative	-	C18	Fatty acids
Stearic acid	7	Negative	-	C18	Fatty acids
Alpha-linolenic acid	7	Negative	-	C18	Fatty acids
Docosaehaenoic acid	7	Negative	-	C18	Fatty acids
Docosapentaenoic acid	7	Negative	-	C18	Fatty acids
Linoleic acid	7	Negative	-	C18	Fatty acids
Arachidonic acid	7	Negative	-	C18	Fatty acids
Oleic acid	7	Negative	-	C18	Fatty acids

Table S8. Method validation parameters (CO = carryover and ME = matrix effect).

Name	ISTD	Repeatability			Intermediate precision			CO (1 st)	CO (2 nd)	R ²	ME (%)	LOD (µM)	LLOQ (µM)
		C0 RSD (%)	C2 RSD (%)	C4 RSD (%)	C0 RSD (%)	C2 RSD (%)	C4 RSD (%)						
Val/Bet	Val	1.6	2.0	1.4	3.4	2.8	2.3	0.0	0.0	0.992	71	0.4	1
Leu/Ile	Leu	2.9	1.1	1.7	2.5	3.1	2.9	0.0	0.0	0.993	65	0.5	1
Trp	Tryp	5.8	0.9	2.0	4.7	1.8	2.4	0.0	0.0	0.997	336	0.3	1
TMAO	TMAO	18.1	3.6	1.0	14.8	7.5	3.1	0.0	0.0	0.999	54	0.4	1
Orn	Lys	13.9	3.4	1.4	11.1	6.5	3.8	0.0	0.0	0.994	137	3	10
Carnitine	Car	3.8	0.1	2.0	3.0	1.4	1.7	0.0	0.0	1.000	54	0.01	0.03
1/3-Mhis	Arg	10.5	0.8	1.4	8.8	1.7	2.0	0.0	0.0	0.996			
Phe	Phe	6.6	1.6	3.6	7.2	5.9	4.9	0.0	0.0	0.993	36	1	4
Ser	Val	6.3	2.4	1.9	4.9	4.6	4.2	0.0	0.0	0.994			
Cit	Val	3.9	2.1	2.6	6.0	4.7	5.8	0.0	0.0	0.997			
Met	Val	5.4	1.8	1.6	5.1	3.8	3.5	5.3	1.5	0.999			
Arg	Arg	3.3	1.2	0.8	3.6	1.8	2.3	2.0	0.0	0.999	89	0.2	0.5
Ala	Glu	4.7	1.4	3.8	5.2	2.3	3.8	0.0	0.0	0.997	68	8	27
Cys-Cys	Lys	6.0	2.2	3.7	7.9	3.3	5.0	0.0	0.0	0.991			
Asp	Val	4.8	3.7	4.3	5.9	6.3	4.2	0.0	0.0	0.995			
Glu	Leu	3.2	0.7	1.4	3.2	3.5	2.4	0.0	0.0	0.996	91	0.9	1
His	Arg	12.1	1.6	0.9	10.0	4.6	2.9	0.5	0.1	0.995			
Lys	Lys	4.3	2.7	3.3	3.6	3.0	3.1	0.0	0.0	0.995	130	1	4
Pro	Val	2.3	1.9	2.4	1.8	2.1	1.6	0.0	0.0	0.998			
Thr	glu	5.7	1.8	3.2	4.2	4.7	2.4	0.0	0.0	0.995			
Urea	Phe	4.2	3.9	3.4	4.7	3.1	2.4	0.0	0.0	0.993			
Gln	glu	3.6	0.7	2.0	2.7	1.9	1.8	0.0	0.0	0.999	69	1	4
Glucose	Glucose	7.2	1.2	7.9	5.3	2.7	5.1	0.0	0.0	0.998	77	3	10
Tyr	Glu	2.2	6.4	3.7	4.4	7.6	4.1	0.0	0.0	0.999			
Choline	Choline	3.4	1.4	1.2	2.7	1.5	1.6	0.0	0.0	0.999	81	0.04	0.1
Hyp	Phe	6.6	4.2	1.2	12.1	4.9	2.7	0.0	0.0	0.998			
Gua	Leu	18.9	8.2	6.5	15.4	8.8	6.7	0.0	0.0	0.991			
Urate	Glu	1.6	2.2	1.3	2.3	3.7	4.3	0.0	0.0	0.991			
Creat	Creat	4.2	0.4	1.5	3.4	1.3	1.3	0.0	0.0	0.999	49	0.09	0.3
Tau	Glu	3.2	2.1	2.6	3.7	4.4	4.4	0.0	0.0	0.997			
C16 carnitine	C8 carnitine	3.8	4.0	1.4	5.5	9.0	10.0	1.8	0.0	0.982			
C18 carnitine	C8 carnitine	11.6	6.0	3.3	11.2	11.7	6.9	0.0	0.0	0.987			
C10 carnitine	C8 carnitine	4.1	1.4	5.4	9.1	6.0	8.6	2.4	1.3	0.990			
C8 carnitine	C8 carnitine	4.6	3.1	5.0	5.8	3.8	4.1	0.0	0.0	0.994	98	0.002	0.003
Citrate	Citrate	1.5	1.3	3.5	2.0	1.8	3.3	0.3	0.2	0.999	562	0.7	1
Malate	Malate	10.6	3.2	0.6	14.7	5.1	3.9	0.4	0.0	0.993	75	0.4	0.9
Lactate	Lactate	1.5	0.7	4.1	1.9	3.0	2.7	0.0	0.0	0.991	34	5	13
Pyruvate	Lactate	4.0	4.8	2.9	4.9	4.9	5.0	0.0	0.0	0.993	41	9	19
Aconate	Citrate	5.8	1.9	3.2	8.5	4.7	3.6	0.9	0.0	0.998			
α-keto-glutarate	Citrate	8.5	7.3	4.1	11.4	6.8	6.6	2.0	0.1	0.996			
2/3-hydroxy-butyrate	Lactate	8.3	0.6	0.7	7.0	4.0	3.2	0.0	0.0	0.999			
Indoxyl-sulfate	Citrate	3.1	3.7	7.3	9.6	6.7	8.0	0.0	0.0	0.995			
FA(16:0)	FA(16:0)	8.9	10.1	10.9	13.5	8.6	11.0	0.8	0.0	0.996	94	0.6	0.7
FA(18:0)	FA(16:0)	9.2	6.6	5.0	14.4	8.4	9.7	0.0	0.0	0.996	98	0.2	0.6
FA(18:3)	FA(16:0)	3.6	3.7	0.9	13.6	14.3	6.9	1.9	0.8	0.996			
FA(22:6)	FA(16:0)	7.6	2.3	0.3	6.9	4.5	6.2	0.4	0.0	0.998			
FA(22:5)	FA(18:0)	6.3	12.6	2.5	7.9	10.7	6.8	1.6	0.3	0.974			
FA(18:2)	FA(18:0)	5.0	3.3	2.2	11.6	5.6	2.7	1.2	0.3	0.998			
FA(20:4)	FA(18:0)	4.9	3.0	1.9	7.9	3.8	4.5	2.0	0.6	0.998			
FA(18:1)	FA(18:0)	3.7	6.0	2.1	4.7	6.4	5.2	0.9	0.1	0.996	562		

Table S9. The matrix effect and fraction of 22 internal standards in fractionation, liquid-liquid extraction coupled to flow injection analysis (LLE-FIA) and flow injection analysis (FIA). An ion suppression of 0% indicates a compound that is either enhanced or not suppressed. An ion suppression of 100% means that the compound is not detected.

Compounds	Fractionation		LLE-FIA		FIA
	Matrix effect (%)	Fraction	Matrix effect (%)	fraction	Matrix effect (%)
Valine-d8	71	Flow-through	5	Polar	1
Leucine-d3	65	Flow-through	4	Polar	0
Tryptophan-13C11, 15N2	336	SCX	4	Polar	0
TMAO-d9	54	Flow-through	9	Polar	4
Ornithine d6	137	SCX	2	Polar	0
Carnitine-d3	54	Flow-through	14	Polar	6
Phenylalanine-d5	36	Flow-through	4	Polar	1
Arginine-15N2	89	SCX	7	Polar	3
Alanine-d3	68	Flow-through	0	Polar	0
Glutamic acid-13C2	91	Flow-through	0	Polar	0
Lysine-d4	130	SCX	5	Polar	2
Glutamine-d5	69	Flow-through	1	Polar	0
Glucose-13C6	77	Flow-through	25	Polar	17
Choline-d4	81	Flow-through/SCX	24	Polar	14
Creatinine-d3	49	Flow-through	20	Polar	6
Octanoyl carnitine-d3	98	C18	32	Apolar	8
Citric acid-D4	562	WAX	86	Polar	677
Malic acid-13C4	75	WAX	19	Polar	24
Lactic acid-13C3	34	Flow-through	14	Polar	14
Pyruvic acid-13C3	41	Flow-through	11	Polar	15
Palmitic acid-D2	94	C18	76	Apolar	12
Stearic-D3	98	C18	70	Apolar	10

Table S10. The LLOQ of literature LC-MS methods. The LLOQ determination, MS mode, derivatization, analysis time and reference are also indicated in the table.

	LLOQ (μM)	LLOQ determination	MS mode	Derivatization	Analysis time (min)	Reference
Valine	10	Unknown	Tandem	Yes	7	Trabado <i>et al.</i> ¹
Leucine	10	Unknown	Tandem	Yes	7	Trabado <i>et al.</i> ¹
Tryptophan	5	Unknown	Tandem	Yes	7	Trabado <i>et al.</i> ¹
TMAO	0.1	S/N	Tandem	No	10	Liu <i>et al.</i> ²
Ornithine	5	Unknown	Tandem	Yes	7	Trabado <i>et al.</i> ¹
Carnitine	0.06	S/N	Tandem	No	10	Liu <i>et al.</i> ²
Phenyl alanine	5	Unknown	Tandem	Yes	7	Trabado <i>et al.</i> ¹
Arginine	5	Unknown	Tandem	Yes	7	Trabado <i>et al.</i> ¹
Alanine	20	Unknown	Tandem	Yes	7	Trabado <i>et al.</i> ¹
Glutamic acid	5	Unknown	Tandem	Yes	7	Trabado <i>et al.</i> ¹
Lysine	10	Unknown	Tandem	Yes	7	Trabado <i>et al.</i> ¹
Glutamine	20	Unknown	Tandem	Yes	7	Trabado <i>et al.</i> ¹
Glucose	11	S/N	Tandem	No	4	Matsunami <i>et al.</i> ³
Choline	0.1	S/N	Tandem	No	10	Liu <i>et al.</i> ²
Creatinine	10	Unknown	Tandem	Yes	7	Trabado <i>et al.</i> ¹
C8 carnitine	0.006	S/N	Tandem	Yes	22	Giesbertz <i>et al.</i> ⁴
Citrate	0.2	S/N	Tandem	No	Unknown	Kadhi <i>et al.</i> ⁵
Malate	0.2	S/N	Tandem	No	Unknown	Kadhi <i>et al.</i> ⁵
Lactate	3	S/N	Tandem	No	Unknown	Kadhi <i>et al.</i> ⁵
Pyruvate	1	Unknown	Tandem	No	3.5	Chuang <i>et al.</i> ⁶
FA(16:0)	0.8	Calibration curve	High-resolution	No	15	Takahashi <i>et al.</i> ⁷
FA(18:0)	0.07	calibration curve	High-resolution	No	15	Takahashi <i>et al.</i> ⁷

References Table S10

1. Trabado, S. *et al.* The human plasma-metabolome: Reference values in 800 French healthy volunteers; Impact of cholesterol, gender and age. *PLoS One* **12**, 1–17 (2017).
2. Liu, J. *et al.* Simultaneous targeted analysis of trimethylamine-N-oxide, choline, betaine, and carnitine by high performance liquid chromatography tandem mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **1035**, 42–48 (2016).
3. Matsunami, R. K., Angelides, K. & Engler, D. A. Development and validation of a rapid ¹³C6-Glucose Isotope Dilution UPLC-MRM Mass Spectrometry Method for Use in determining system accuracy and performance of blood glucose monitoring devices. *J. Diabetes Sci. Technol.* **9**, 1051–1060 (2015).
4. Giesbertz, P., Ecker, J., Haag, A., Spanier, B. & Daniel, H. An LC-MS/MS method to quantify acylcarnitine species including isomeric and odd-numbered forms in plasma and tissues. *J. Lipid Res.* **56**, 2029–2039 (2015).
5. Al Kadhi, O., Melchini, A., Mithen, R. & Saha, S. Development of a LC-MS/MS Method for the Simultaneous Detection of Tricarboxylic Acid Cycle Intermediates in a Range of Biological Matrices. *J. Anal. Methods Chem.* **2017**, 1–12 (2017).
6. Chuang, C. K. *et al.* A method for lactate and pyruvate determination in filter-paper dried blood spots. *J. Chromatogr. A* **1216**, 8947–8952 (2009).
7. TAKAHASHI, H. *et al.* Long-Chain Free Fatty Acid Profiling Analysis by Liquid Chromatography–Mass Spectrometry in Mouse Treated with Peroxisome Proliferator-Activated Receptor α Agonist. *Biosci. Biotechnol. Biochem.* **77**, 2288–2293 (2013).

Table S11a. The reversed phase (RP) gradient performed on an UPLC HSS T3 (1.7 μm , 2.1 x 100 mm). Mobile phase A and B consisted of 0.1% formic acid in water and acetonitrile, respectively. The column oven was set at 30 °C. The sample preparation, injection volume and MS parameters were similar to the fractionation method.

Time (min)	Mobile phase B (%)	Flow rate (mL/min)
0.0	0	0.4
7.0	15	0.4
10	55	0.4
10.5	100	0.4
12.5	100	0.4
12.6	0	0.4
15.0	0	0.4

Table S11b. The hydrophilic interaction liquid chromatography (HILIC) gradient performed on a Sequant ZIC-HILIC (3 μm , 2.1 x 100 mm). Mobile phase A consisted of 10 mM ammonium formate and 0.075% formic acid in 90/10 (v/v) acetonitrile/water and mobile phase B of 10 mM ammonium formate and 0.075% formic acid in 10/90 (v/v) acetonitrile/water. The column oven was set at 30 °C. The sample preparation, injection volume and MS parameters were similar to the fractionation method.

Time (min)	Mobile phase B (%)	Flow rate (mL/min)
0.0	0	0.5
1.2	0	0.5
9.16	75	0.5
14.0	75	0.5
14.2	0	0.5
18.0	0	0.5

Table S11c. The number of detectable features in the fractionation, RP and HILIC method as determined by the 'Non-targeted Peaks' function in Analytics of Sciex OS 1.6. The minimum retention time was determined by the dead volume of the platform and the maximum retention time was determined by the start of the column equilibration. The peak detection sensitivity was set at exhaustive, the Area Ratio Threshold (Unknown/Control) was set at 0 and the 'Group peaks by adduct or charge' function was checked. The number of features was corrected for the presence of multiple adducts. Feature m/z values were rounded to two decimal points. Features that were identified in both positive and negative mode were indicated as one feature.

Platform	Unique retention time and m/z features	Unique m/z features
Fractionation	2342	2113
RP	3517	2478
HILIC	3534	2326

Table S12. P-values and FDR adjusted p-values of several compound classes on different time points in comparison to baseline levels.

Compound class	Time after breakfast (hours)	P-value	FDR adjusted p-value
Amino acids	0.5	0.006	0.063
	1.5	0.004	0.046
	3	0.002	0.032
	4.5	0.002	0.032
	6	0.037	0.296
	9.5	0.16	0.942
	11	0.002	0.032
	12	0.002	0.032
	24	0.131	0.829
Amines	0.5	0.846	1
	1.5	0.557	1
	3	0.232	1
	4.5	0.105	0.686
	6	0.064	0.498
	9.5	0.084	0.576
	11	0.084	0.576
	12	0.557	1
	24	0.16	0.942
Acylcarnitines	0.5	0.014	0.13
	1.5	0.002	0.032
	3	0.002	0.032
	4.5	0.002	0.032
	6	0.002	0.032
	9.5	0.006	0.063
	11	0.002	0.032
	12	0.002	0.032
	24	0.232	1
Hexose	0.5	0.02	0.166
	1.5	0.77	1
	3	0.922	1
	4.5	0.002	0.032
	6	0.557	1
	9.5	0.16	0.942
	11	0.002	0.032
	12	0.084	0.576
	24	1	1
Organic acids	0.5	0.002	0.034
	1.5	0.004	0.046
	3	0.846	1
	4.5	0.006	0.058
	6	0.846	1
	9.5	0.105	0.686
	11	0.002	0.034
	12	0.02	0.166
	24	0.006	0.058
Fatty acids	0.5	0.193	1
	1.5	0.004	0.046
	3	0.004	0.046
	4.5	0.004	0.046
	6	0.004	0.046
	9.5	0.105	0.686
	11	0.037	0.296
	12	0.004	0.046
	24	0.01	0.097

Chapter 4

Data independent acquisition for the quantification and identification of metabolites in plasma

Based on:

Tom van der Laan, Isabelle Boom, Joshua Maliepaard, Anne-Charlotte Dubbelman, Amy Harms and Thomas Hankemeier

Data independent acquisition for the quantification and identification of metabolites in plasma

Submitted for publication (2020)

ABSTRACT

Fragmentation data is of great value in mass spectrometry-based metabolomics. It is used to enhance the selectivity for the quantification of analytes or to provide structural information for the identification of features. A popular fragmentation technique for non-targeted analysis is called data independent acquisition (DIA) because it provides fragmentation data for all analytes in a specific mass range. In this work, we demonstrated the strengths and weaknesses of DIA. Two types of chromatography (fractionation/3min and HILIC/18min) and three DIA protocols (variable SWATH, fixed SWATH and MS^{ALL}) were used to evaluate the performance of DIA. Our results show that fast chromatography often results in product ion overlap and complex MS/MS spectra, which reduces the quantitative and qualitative power of these DIA protocols. Wide Q1 windows (i.e. MS^{ALL}) demonstrated a similar effect, which was observed in both chromatography types. The combination of SWATH and HILIC allowed for the correct identification of 20 metabolites using the NIST library. After SWATH window customization (i.e. variable SWATH), we were able to quantify ten structural isomers with a mean accuracy of 103% (91-113%). The robustness of the variable SWATH and HILIC method was demonstrated by the accurate quantification of these structural isomers in 10 highly diverse blood samples. Since the combination of variable SWATH and HILIC results in good quantitative and qualitative fragmentation data, it is promising for both targeted and untargeted platforms. This should decrease the number of platforms needed in metabolomics and increase the value of a single analysis.

INTRODUCTION

Fluctuations in the levels of metabolites have frequently been correlated to health, disease or response to treatment.^{1,2,3} Therefore, the identity and quantity of these metabolites can help researchers and clinicians to determine the health status of an individual and eventually to realize personalized medicine.⁴ Liquid chromatography (LC) coupled to mass spectrometry (MS) is one of the most commonly used techniques to identify and quantify metabolites in biological samples by qualitative and quantitative analyses, respectively.^{5,6} This is due to its high sensitivity, selectivity and potential for high-throughput.

With high-resolution MS, a full MS scan can be used to suggest the elemental composition of a compound based on the measured m/z value. However, the intact mass alone does not provide enough structural information to identify unknown features or resolve compounds with the same m/z value, i.e. isomers. Fragmentation patterns obtained by tandem mass spectrometry (MS/MS) provide structural information and can, therefore, aid in identifying metabolites and resolving isomers. Therefore, fragmentation data is a great asset in both qualitative and quantitative analyses.

MS/MS can be subdivided into two categories: targeted and untargeted analyses, whereby the latter can be further subdivided into data dependent (DDA) and data independent acquisition (DIA) protocols. An overview of the most commonly used MS/MS modes are listed in Table 1. In

Table 1. Overview of the most commonly used tandem mass spectrometry modes. MRM: multiple reaction monitoring, MRM^{HR}: high-resolution multiple reaction monitoring, PRM: parallel reaction monitoring, DDA: Data Dependent Acquisition, DIA: Data Independent Acquisition, SWATH: Sequential Window Acquisition of All Theoretical Mass Spectra and AIF: All-Ion Fragmentation.

Name	MRM	MRM ^{HR} /PRM	DDA	DIA	
				SWATH	MS ^{ALL} /MS ^E /AIF
Q1 size	Narrow	Narrow	Narrow	Medium	Wide
High/low resolution	Low	High	High	High	High
Precursor selection	User defined	User defined	Based on parent ion intensity*	No selection	No selection
Targeted/untargeted	Targeted	Targeted	Untargeted	Untargeted	Untargeted

*Can also be based on e.g. mass defect or isotope pattern

targeted analyses, one or multiple metabolites are selected based on the hypothesis of a biological question. Most often these analyses are performed by multiple reaction monitoring (MRM) using a triple quadrupole (QqQ) mass spectrometer in which a precursor and product ion(s) for each analyte are optimized.⁷ A more selective and comprehensive targeted approach can be realized by MRM high-resolution (MRM^{HR}) (also known as parallel reaction monitoring (PRM)). In this mass spectrometry mode, a preselected precursor is isolated and fragmented followed by a high-resolution scan of all the produced product ions. Targeted analyses only allow for the quantification and identification of preselected features. This means that it is possible to optimize analytical parameters specifically to an analyte, which can be beneficial in terms of selectivity and sensitivity. It also leads to less unnecessary data, because of the preselection of interesting targets.

Untargeted analysis aims at a comprehensive measurement of all analytes in a biological system, including chemical unknowns.^{8,9} MS/MS scans are indispensable in untargeted analysis to provide structural information. Untargeted analyses, i.e. DDA and DIA based protocols, consist of a high-resolution full scan followed by a single or multiple MS/MS experiment(s). In DDA, precursor ions are usually selected and fragmented based on intensity resulting in the fragmentation of the most abundant precursor ions.¹⁰ A major drawback of these methodologies is the limited time for fragmentation which is caused by the peak width and MS scan time. Time limits the fragmentation to the most abundant precursor ions, which can result in missing fragmentation data for low abundant ions. Although this can be partly overcome by including a dynamic exclusion list, the fact remains that this type of analysis is naturally biased towards certain analytes.¹¹ Chromatograms acquired through DDA can not be used for quantification, because often the apex of the peak is missed and/or there are not enough data points on the peak for accurate quantitation.¹²

DIA fragmentation is not triggered by the intensity of precursor ions but is predefined by the user. Here, a certain mass range is selected and all precursors within this mass range are fragmented. Since DIA methodologies do not involve a selection of precursor (apart from the selected mass range), they are often considered more suitable for untargeted analyses.^{9,8,13} One example of DIA is MS^{ALL} (also known as All-Ion Fragmentation (AIF) and MS^E) in which the whole mass range is fragmented at once.^{14,15} This technique makes use of a wide Q1 filter and allows for the detection of the fragments of all precursors in a single scan. A more selective DIA approach is called Sequential Window Acquisition of All Theoretical Mass Spectra (SWATH) in which a mass range of interest is divided into several subranges.¹³ Typically, SWATH scans are performed by Q1 window sizes ranging from 3-50 Dalton. In proteomics, this window size is typically 25 Dalton.¹⁶ Since the DIA analyses allow for the fragmentation of all precursors, the quality of DIA acquired MS/MS spectra might be compromised by the larger Q1 window size. More precursors will enter

the collision cell concurrently with increasing Q1 window sizes. As a result, complex MS/MS spectra can hamper the proper identification of chemical unknowns and overlap in product ion formation can lead to inaccurate quantification.⁸ This is especially problematic for high-throughput analyses whereby numerous analytes are coeluting. Moreover, fast analyses often go hand in hand with a narrow chromatographic peak width resulting in less time for fragmentation.

The fragmentation data of DIA methodologies can be used for two purposes: identification and quantification. This offers a great potential for combined quantitative and qualitative platforms, which are increasingly implemented in the field of metabolomics.¹⁷ For the identification of metabolites, the acquired MS/MS spectra can be compared with commercially available spectral libraries, such as NIST, mzCloud, HMDB and Metlin. These libraries are easily searched in open access web interfaces and/or by the integration into data analysis software. Metabolite identifications are scored based on the similarity in fragmentation patterns between the acquired spectra and library spectra. It is also possible to re-interrogate the data for the identification of unknown features found at a later stage, since MS/MS spectra of all precursors are acquired. The human metabolome consists of numerous metabolites, which makes its composition very complex. The Human Serum Metabolome, for example, reports the presence of 4651 metabolites.¹⁸ Because of the complexity of the human metabolome, the use of comprehensive fragmentation techniques is as important as the use of comprehensive libraries. Current spectral libraries cover a substantial part of the human metabolome (e.g. 3500 human metabolites in NIST 2017¹⁹ and 3724 endogenous metabolites in mzCloud²⁰). This coverage is essential in identifying unknown metabolites using MS/MS data.

For quantification, fragmentation can help in quantifying compounds that are not resolved by their exact mass, like isomers. Chromatographic separation is often used to resolve isomers before introducing them into the MS. However, achieving baseline separation between isomers often requires extensive method development and these analytical platforms are usually targeted to only a single isomer pair.^{21,22} Generally, DIA platforms are not optimized to resolve specific isomer pairs. The fragmentation data in DIA analyses, however, can be used to differentiate between structural isomers which have the same elemental composition but differ in elemental organization. As a result, different fragmentation patterns can be obtained within a structural isomer pair.^{23,24} Diagnostic product ions can be used to differentiate between structural isomers and can also help to quantify these compounds individually.

In this work, we evaluate the versatility of DIA methodologies for metabolomics. Both a conventional LC-MS and a high-throughput platform were evaluated based on their compatibility with DIA analyses. These platforms differ in separation efficiency and peak width and thereby reflect the demands of different types of chromatography in terms of DIA analyses. The

compatibility assessment is based on the qualitative and quantitative performances of the MS/MS scans. In order to demonstrate the quantitative performance, we have quantified five structural isomer pairs by SWATH and MS^{ALL} using diagnostic product ions. The quantified values are compared with MRM^{HR} for accuracy, since this is the most selective fragmentation mode. The qualitative performance is defined by identifying a set of known metabolites using a commercially available MS/MS library (NIST). This comprehensive comparison of different DIA methodologies using different separation mechanisms should demonstrate the usefulness and the limitations of DIA methodologies in terms of quantification and identification.

MATERIALS AND METHODS

Chemicals

Standards were purchased from HMDB (Edmonton, Alberta, Canada), Sigma-Aldrich (Zwijndrecht, The Netherlands) and Fluka (Seelze, Germany). Internal standards were purchased from Cambridge Isotopes (Tewksbury, Massachusetts, United States), Cortecnet (Voisins-Le-Bretonneux, France) and CDN Isotopes (Nieuwegein, The Netherlands). An overview of the (internal) standards and concentrations is provided in the Supplementary information (Table S1 and S2).

Acetonitrile (LC-MS grade) was purchased from Biosolve B.V. (Valkenswaard, The Netherlands) and methanol (Ultra-LC-MS grade) was purchased from Actu-All (Oss, The Netherlands). Ammonium formate ($\geq 99.995\%$) and ammonium acetate ($\geq 99.0\%$) were purchased from Sigma-Aldrich (Zwijndrecht, The Netherlands). Ammonium hydroxide (28-30 wt% solution of ammonia in water) and formic acid (98%+) were purchased from Acros Organics (Bleiswijk, The Netherlands). Pooled heparin plasma (June 2018) was used for quantitative and qualitative performance evaluation and purchased from Sanquin (Amsterdam, The Netherlands). Ten diverse EDTA plasma samples were purchased from Bio IVT (Westbury, New York, United States). The clinical variables are mentioned in the supplementary information (table S3).

Standard and internal standard solutions

Individual (internal) standards were dissolved in water. Calibration and internal standard solutions were prepared by mixing the individual standards into a final solvent composition of 75% methanol in water. The C8 calibration solution was the highest concentration and the subsequent calibration solutions (C7-C1) were 1:1 dilutions of the previous solution. The C0 calibration solution was a 75% methanol solution that did not contain the calibration standards. The internal standard and C4 calibration concentration were set to mimic the physiological concentration as found in HMDB.²⁵

Sample preparation

During sample preparation, 15 μL of heparin plasma, 15 μL of internal standard solution and 15 μL of calibration standard solution were mixed with 30 μL of methanol by vortex mixing. Non-spiked samples were prepared by replacing the calibration standard solution by 15 μL of 75% methanol. Subsequently, the samples were centrifuged at 4 $^{\circ}\text{C}$ and 16100 g for 10 min. Fifty μL of supernatant was transferred into a 1.5 mL HPLC vial containing a 150 μL insert.

LC-MS

A Sciex X500R QToF (Darmstadt, Germany) was used for the MS analysis. A Shimadzu Nexera UHPLC (Darmstadt, Germany) was extended with an Agilent 1260 infinity isocratic pump (Waldbronn, Germany) and two VICI six-port valves (Rotterdam, The Netherlands).

The high-throughput analyses were performed by an on-line fractionation method, which has been published before.²⁶ In short, a ZORBAX Extend-C18 2.1 \times 5 mm, Sepax SCX 2.1 \times 10 mm and Sepax WAX 2.1 \times 10 mm were serially coupled in order to trap apolar, positively and negatively charged analytes, respectively. Three mobile phases are used for loading (0.2% formic acid in water), ion-exchange elution (100 mM ammonium acetate pH 10 in water) and C18 elution (2mM ammonium acetate in methanol). The flow rate was 800 $\mu\text{L}/\text{min}$ for loading/C18 elution and 500 $\mu\text{L}/\text{min}$ for ion-exchange elution. The injection volume was set at 1 μL . The polarity, ion source gas 1, ion source gas 2, curtain gas, CAD gas, temperature, declustering potential and spray voltage were set at positive, 40 psi, 60 psi, 40 psi, 7 psi, 650 $^{\circ}\text{C}$, 80 V and 5500 Volt, respectively.

The conventional LC analysis was performed by a HILIC separation. The HPLC column was a SeQuant ZIC-HILIC 2.1 \times 100 mm column (Amsterdam, The Netherlands). Mobile phase A consisted of 10 mM ammoniumformate and 0.075/90/10 (v/v/v) formic acid/acetonitrile/water and mobile phase B consisted of 10 mM ammoniumformate and 0.075/10/90 (v/v/v) formic acid/acetonitrile/water. The gradient started at 0% B. After 1.2 minutes the gradient linearly increased to 75% B in 7.96 min. The gradient was kept at 75% B for 4.84 minutes in order to flush the column. Subsequently, the gradient decreased to 0% B in 0.2 minutes and was kept at this value for 3.8 minutes to equilibrate the column. The flow rate and the injection volume were set at 500 $\mu\text{L}/\text{min}$ and 3 μL , respectively. The polarity, ion source gas 1, ion source gas 2, curtain gas, CAD gas, temperature, declustering potential and spray voltage were set at positive, 40 psi, 60 psi, 35 psi, 7 psi, 575 $^{\circ}\text{C}$, 80 V and 5500 Volt, respectively.

All MS methods consisted of a full scan of 50 ms at a collision energy 5 eV followed by at least one MS/MS scan of 30 ms at a collision energy of 20 eV. The MRM^{HR} analysis was conducted in product ion scan mode using Sciex OS 1.5 (Darmstadt, Germany). MS^{ALL} was performed in SWATH mode using one MS/MS window ranging from m/z 75 to 250. The SWATH analysis in the fractionation

and HILIC method consisted of 7 and 30 SWATH windows, respectively. Variable SWATH windows were calculated using the 'SWATH Variable Window Calculator V1.1' from Sciex (Supplementary information Figure S1). The windows sizes of the variable and fixed SWATH methods are depicted in the supplementary information (Table S4 and S5).

Diagnostic product ions and quantification of structural isomers.

Structural isomer standards in water were analyzed individually by a flow injection analysis and an MRM^{HR} scan in order to find the corresponding product ions. Product ions and corresponding intensities were obtained by taking the average mass spectrum of the analyte peak. All product ions that demonstrated an intensity of at least 1 percent of the most abundant product ion were included. Overlapping product ions (similar exact mass) within an isomer pair were excluded. The remaining diagnostic product ions were used for the quantification of the structural isomers.

Diagnostic product ions were quantified by the fractionation and HILIC method by plotting the peak area ratio of a diagnostic product ion against the concentration. Nine different calibration points were used to construct the calibration curve (C0-C8). All calibration points were used for the calibration curve, unless they were outside of the linear range ($R^2 < 0.99$). Five replicates of a plasma pool were analyzed to determine the physiological concentrations. Product ions that were included for quantification demonstrated no peak overlap with other common product ions, had a linear calibration curve (≥ 0.99) and deviated less than 15% ($n=5$). Each analyte was quantified using four different MS methods: MRM^{HR}, variable SWATH, fixed SWATH and MS^{ALL}. The MRM^{HR} scan was taken as the reference value. Quantification accuracies are reported as a percentage of the MRM^{HR} scan (see formula 1).

$$(1) \text{ Quantification Accuracy} = \frac{[DIA \text{ protocol}]}{[MRM^{HR}]} \times 100\%$$

The remaining diagnostic product ions demonstrated a quantification accuracy of 85-115%. The robustness of the HILIC and variable SWATH method was assessed by determining the quantification accuracies of the remaining diagnostic product ions in blood samples of ten male donors. The subjects were selected based on their differences in clinical variables, i.e. age, race, BMI, fasting state and smoking habit (see supplementary information table S3). The diversity in the blood samples should push the performance of the HILIC and variable SWATH method to its extremes.

Identification of metabolites

The NIST 2017 library was used to evaluate the qualitative power of the DIA MS/MS spectra. The library was uploaded in Library View and accessed via Analytics in Sciex OS. The identification was performed by a candidate search, which shows the best spectral match in the selected library.

Twenty known metabolites were identified using five technical replicates. The mean library hit score and the number of correct hits were noted. The absence of a library match and an incorrect structural assignment were given a library hit score of 0%.

RESULTS AND DISCUSSION

Method development

For the assessment of the DIA performance, we have chosen two different chromatography methods to demonstrate the possibilities and limitations of DIA-based analyses. The difference between the two chromatography methods is mainly driven by time (Figure 1). In the fractionation method, the analytes of interest all elute within the first minute of the analysis. This causes very narrow peak widths and severe coelution. As a result, there will be little time for fragmentation and a high chance of complex MS/MS spectra. In contrast, the HILIC method demonstrates much more separation and also considerably broader peaks. Figure 1 demonstrates that there is even a slight separation within the structural isomer pairs using a HILIC separation. This means that in the HILIC method, there is more time for fragmentation and a lower chance of complex MS/MS spectra. The comparison of the two methods should demonstrate the influence of chromatographic separation and peak width on the selectivity of the DIA MS/MS spectra.

During SWATH method development, we aimed for the highest selectivity by including as many windows as possible while still having at least ten data point across the chromatographic peak for absolute quantitation. For this, we calculated the maximum number of SWATH windows that still ensured ten data points on the narrowest chromatographic peak.¹⁶ This resulted in a maximum of 7 windows in the fractionation method and 30 windows in the HILIC method. The fixed SWATH

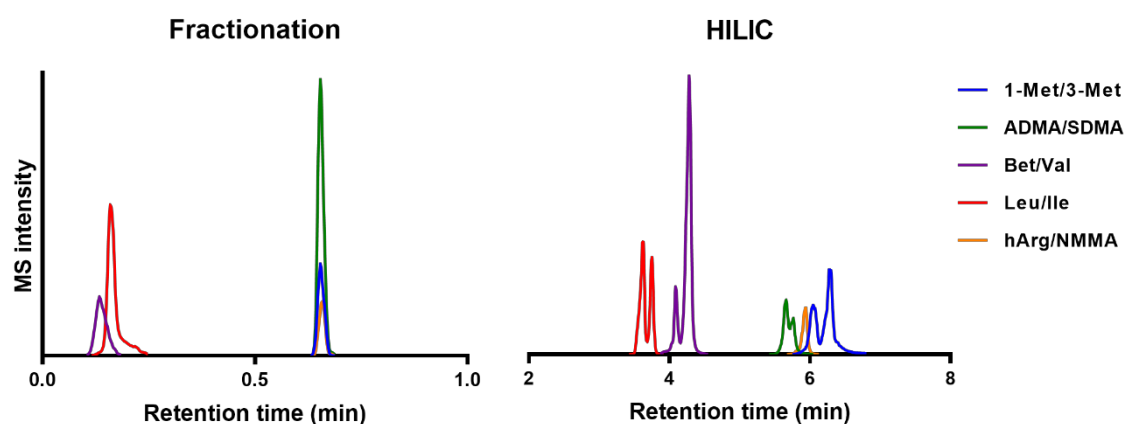


Figure 1. Extracted ion chromatograms of five structural isomer pairs. The fractionation method does not demonstrate chromatographic separation within an isomer pair. The HILIC method does demonstrate slight chromatographic separation within an isomer pair.

methods were constructed by dividing the mass range by the number of SWATH windows. A one Dalton overlap between the windows was included to ensure the inclusion of all precursors in the desired mass range. The fixed SWATH window size was 25 (+1) and six (+1) Dalton in the fractionation and HILIC method, respectively. The window sizes of the variable SWATH methods were calculated using the density of all precursor ions in the desired mass range (75-250 m/z).²⁷ The window sizes were adjusted in order to have an equal density of precursor ions in each SWATH window.

Quantitation of structural isomers

Five structural isomer pairs were quantified by analyzing diagnostic product ions. The accuracies of the best performing product ions are depicted in Table 2 and the performance of all product ions is presented in the supplementary information (Table S6). In general, more diagnostic product ions could be integrated reliably by SWATH in comparison with MS^{ALL}. The extracted ion chromatograms in MS^{ALL} were often disturbed by fluctuating or high baselines, which led in some cases to the absence of certain product ion peaks. In general, the baseline signal increased with increasing Q1 windows (as shown by Figure 2). The elevated baselines indicated that background

Table 2. The quantification accuracy of structural isomers in comparison with MRM^{HR}. Accuracies were divided into three groups: 85-115% (green), 115-120% (yellow) and >120% (red). Compounds that could not be quantified due to an insufficient linearity (<0.99), high variability (>15%) or integrations problems (peak overlap or too high baseline) are indicated by the zero values.

Analyte	Variable SWATH		Fixed SWATH (%)		MSE (%)	
	<i>Frac</i>	<i>HILIC</i>	<i>Frac</i>	<i>HILIC</i>	<i>Frac</i>	<i>HILIC</i>
ADMA	88	96	0	102	0	0
SDMA	107	101	96	110	106	0
hArg	109	99	389	102	271	0
NMMA	739	111	6383	226	1944	0
Bet	106	102	103	134	136	124
Val	98	91	93	91	99	101
1-Met	157	113	0	109	0	102
3-Met	97	99	97	89	193	89
Ile	106	112	113	99	129	90
Leu	144	110	148	117	171	168

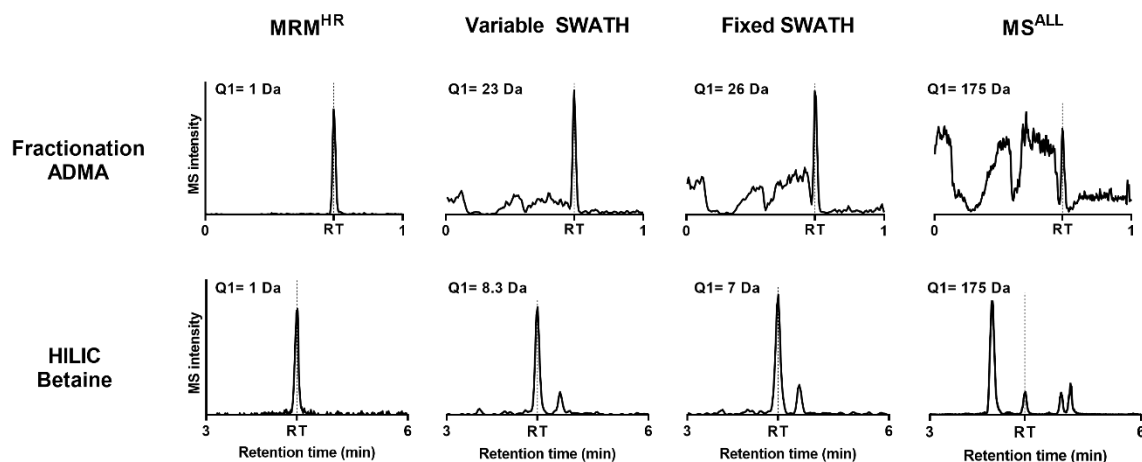


Figure 2. Extracted ion chromatograms of the different types of MS/MS scans with corresponding Q1 window sizes. The extraction width is 0.02 m/z. Fluctuating baselines can be observed in the fractionation method with increasing Q1 window sizes. Multiple peaks can be seen in HILIC with increasing Q1 window sizes.

ions produced similar products as some of the structural isomers. Besides, extracted ion chromatograms of MS^{ALL} often showed multiple peaks, indicating the presence of common product ions produced by different precursors. In some cases, coelution of these common product ions obstructed proper peak integration. Similar integration problems occurred in the SWATH analysis. However, the frequency and severity of these problems were noticeably less. MRM^{HR} resulted in the cleanest chromatograms with low baselines and no overlapping peaks. These findings are in accordance to the work of Venable *et al.*, in which was shown that the appearance of background peaks and high background noise decreases with decreasing Q1 window sizes.²⁸

In MS^{ALL}, six of the ten compounds had a quantification accuracy above 115% (129-1944%), indicating that multiple precursor ions indeed produced similar product ions. Especially the fractionation method was compromised by this particular DIA mode. The quantification values were considerably higher than MRM^{HR}, which was expected because of the high degree of coelution and large window sizes. The analysis of complex matrices, in particular, is prone to suffer from product ion overlap because of the high chance of coelution of compounds with similar fragmentation patterns.²⁷ Although the product ions were diagnostic within an isomer pair, they lacked selectivity in the presence of a biological matrix. By decreasing the size of the Q1 windows in the SWATH analysis, the accuracies drastically improved. This applied to both the fractionation and HILIC method.

The variable SWATH method improved the quantification accuracies more than the fixed SWATH method in both chromatography methods. This demonstrates that even with a conventional LC separation and relatively small SWATH windows (6+1 Dalton), the selectivity is still dependent

on the customization of the SWATH windows (fixed versus variable). The power of customized SWATH windows is emphasized by a comparison of the three MS/MS methods. Table 2 shows that 85%, 60% and 30% of the individual isomer quantifications were good (>85% and <115%) in variable SWATH, fixed SWATH and MS^{ALL}, respectively, indicating that a higher degree of Q1 window customization results in a more accurate quantification. The combination of variable SWATH and HILIC resulted in the accurate quantification of all investigated isomers. The quantitative performance of the latter platform was maintained when the five isomers pairs were quantified in blood samples of ten male donors (see Figure 3). The high correlation with MRM^{HR} demonstrated that differences in biological samples did not interfere with the selectivity of the quantification of the selected isomers. Figure 3 shows that the correlation coefficient of 9 structural isomers was higher than 0.90, with a mean quantification accuracy of 102% (96-108%) per diagnostic fragment. The strongly correlated and highly accurate quantification values demonstrate that the quantification accuracies of variable SWATH and HILIC are not affected by a large variation in blood sample matrices. Figure S2 and Table S7 in the supplementary information reveal that the other diagnostic product ions followed the same trend. The only product ion that behaved differently was the product ion of betaine. Even though the quantification values were strongly correlated between MRM^{HR} and variable SWATH ($R^2=0.86$), the mean quantification accuracy was 157%. Since the relative standard deviation of the quantification accuracies throughout the ten blood samples was low (9%) and the correlation with MRM^{HR} high, it was expected that the quantification values were structurally increased in all volunteers rather than influenced by product ions derived from the blood samples. However, the exact mechanism of this increase remained unclear.

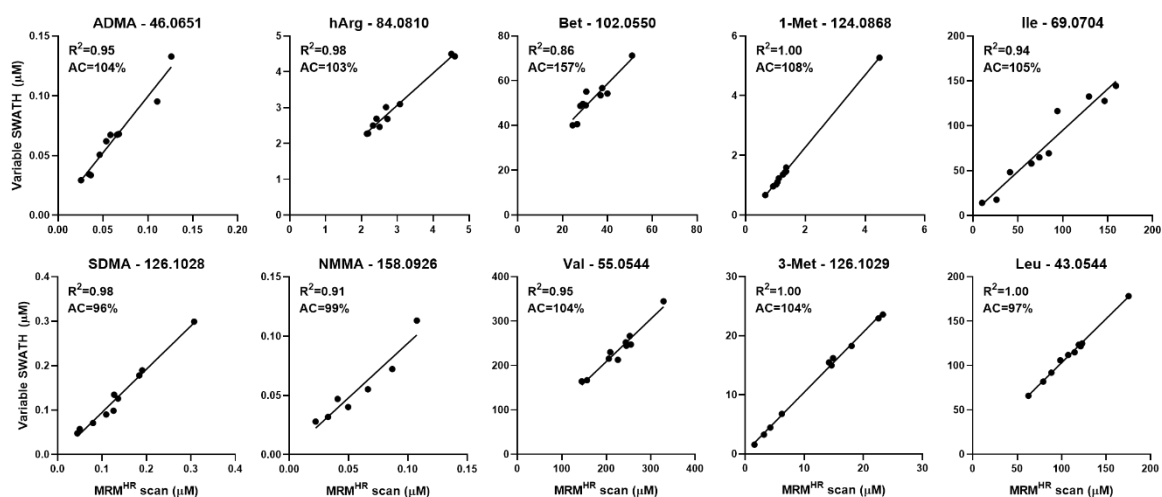


Figure 3. Correlation (R^2) and accuracy (AC) of the quantification values of 10 structural isomers in 10 volunteers. Each data point represents the analyte concentration quantified by MRM^{HR} and variable SWATH. The quantification values of the best performing product ion are depicted in the graph. NMMA was not detected in three volunteers.

In accordance to our results, Zhang *et al.* demonstrated a performance improvement by using variable SWATH windows over fixed SWATH windows. The performance improvement was found in both the identification and quantification of peptides.¹² Table 2 also demonstrates that DIA methodologies benefit from extensive chromatographic separations since DIA fragmentation in the HILIC method resulted in a more accurate quantification. Especially MS^{ALL} analyses require a thorough separation because the size of the Q1 window adds little extra selectivity.¹³ In addition to DIA type and chromatographic separation, MS/MS quantifications improve with increasing peak widths because smaller SWATH windows can be applied resulting in a more selective fragmentation.¹² This has to be taken into account when DIA protocols are implemented in high-throughput platforms in which severe coelution and small peak widths often occur. Our results show that using a HILIC separation variable SWATH, 9 out of 10 structural isomers can be quantitated accurately in a robust manner. However, it does not mean that all product ions in this methodology are selective. Table S6 in the supplementary information shows that there are still 8 out of the 28 product ions that could not be accurately quantified. Therefore, the possibility of product ion overlap should always be considered when DIA methodologies are used for quantification.

Identification of metabolites

In metabolomics, untargeted analyses can be performed to correlate unknown features with a certain physiological effect. Once a feature demonstrates a correlation with an investigated response, it is important to unravel its structure in order to understand its biochemical mechanism in the studied effect. Identification strategies are crucial to transform data (m/z values) into information (chemical structures). MS/MS data of unknown features can be used to search for potential candidates in commercially available spectral libraries, like NIST. Here, we discuss the qualitative power of DIA-acquired MS/MS spectra using a spectral library search.

The library hit scores and number of correct library hits using SWATH and MS^{ALL} are listed in Table 3. The identities of the metabolites were known in advance, which facilitated the evaluation of the correctness of the structure annotation. As shown by Table 3, practically all identifications were correct using a HILIC separation and SWATH. This confirms the qualitative power of SWATH, which has been shown numerous times before.^{29,30,31} The identification in our HILIC method and the latter referenced literature all make use of a conventional chromatographic separation which results in relatively clean MS/MS spectra. In contrast, the fractionation method, which makes use of a limited separation, demonstrated a decrease in qualitative performance as five and six of the 20 compounds could not be identified in variable and fixed SWATH, respectively. The higher degree of coelution and large window sizes resulted in more complex MS/MS spectra. Deconvolution was applied to remove several impurities from the MS/MS scan as it is known that

Table 3. Identification of twenty known metabolites using SWATH and MSALL. The experimental MS/MS spectra were compared with the NIST 2017 library. The mean score of the library hits (N=5) and the number of correct hits are listed in the table. MS/MS spectra that could not be matched or resulted in a wrong hit were given a score of 0%.

Analyte	Variable SWATH				Fixed SWATH				MS ^{ALL}			
	Frac		HILIC		Frac		HILIC		Frac		HILIC	
	Score (%)	Hits (#)	Score (%)	Score (%)	Score (%)	Hits (#)	Score (%)	Hits (#)	Score (%)	Hits (#)	Score (%)	Hits (#)
Proline	100	5/5	100	5/5	100	5/5	100	5/5	100	5/5	100	5/5
Arginine	100	5/5	100	5/5	0	0/5	99	5/5	100	5/5	99	5/5
Acetyl-carnitine	99	5/5	99	5/5	98	5/5	99	5/5	96	5/5	98	5/5
Creatinine	100	5/5	99	5/5	100	5/5	99	5/5	95	5/5	99	5/5
Ornithine	0	0/5	99	5/5	0	0/5	99	5/5	40	2/5	96	5/5
Methionine	76	5/5	99	5/5	0	0/5	99	5/5	0	0/5	99	5/5
Carnitine	98	5/5	98	5/5	96	5/5	98	5/5	78	5/5	97	5/5
Citrulline	98	5/5	97	5/5	99	5/5	94	5/5	0	0/5	50	3/5
Tyrosine	94	5/5	97	5/5	94	5/5	96	5/5	92	5/5	98	5/5
Phenylalanine	76	5/5	96	5/5	57	4/5	96	5/5	41	3/5	97	5/5
Histidine	98	5/5	95	5/5	98	5/5	95	5/5	99	5/5	3	1/5
Serine	0	0/5	93	5/5	0	0/5	97	5/5	0	0/5	19	3/5
Tryptophan	63	5/5	93	5/5	77	5/5	91	5/5	0	0/5	94	5/5
Glutamine	93	5/5	93	5/5	92	5/5	94	5/5	97	5/5	93	5/5
Cystine	88	5/5	92	5/5	92	5/5	91	5/5	0	0/5	93	5/5
Taurine	0	0/5	92	5/5	0	0/5	89	5/5	0	0/5	0	0/5
Lysine	0	0/5	91	5/5	88	5/5	89	5/5	0	0/5	93	5/5
Asparagine	0	0/5	82	5/5	0	0/5	92	5/5	0	0/5	0	0/5
Glutamic acid	90	5/5	62	4/5	94	5/5	82	5/5	97	5/5	0	0/5
γ-butyrobetaine	54	5/5	44	5/5	73	5/5	60	5/5	0	0/5	0	0/5
Average	66		91		63		93		47		66	

this decreases the complexity of the DIA-based MS/MS scans.³² However, deconvolution has a limited power when severe coelution and wide Q1 windows dissociate the link between the precursor and product ion. A separation based on fractions results often in identical retention times for several precursors, which makes it difficult to link a precursor ion to its product ions by retention time.

As shown in Figure 4, MS/MS spectra become more complex with increasing Q1 window sizes, which is clearly demonstrated by a comparison of the SWATH and MS^{ALL} spectra. This explains the fact that MS^{ALL} only allowed for the unequivocal identification of 9 and 13 of the 20 compounds in the fractionation and HILIC method, respectively. The identification of taurine in Figure 4 illustrates the problems of coelution and large window sizes because only a HILIC separation

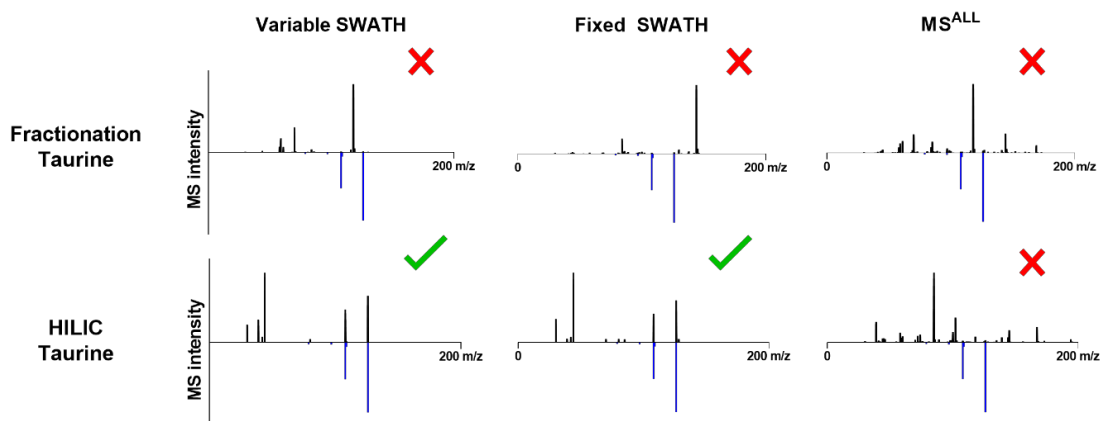


Figure 4. Experimental (black) and NIST library (blue) MS/MS spectra comparison. The HILIC separation in combination with SWATH resulted in the assignment of the correct structure.

using SWATH resulted in a correct identification. Although the two most abundant MS/MS peaks of the library spectra were present in all DIA-acquired spectra, these peaks were overshadowed by numerous other higher intensity peaks. This obstructed the correct identification of taurine.

The qualitative power of SWATH clearly outperformed MS^{ALL} , which has been shown before.³³ The identification performance did not seem to be affected by a higher degree of SWATH window customization as the number of correct hits and library hit scores were virtually similar between variable and fixed SWATH. This in contrast to a study of Zhang *et al.*, in which variable Q1 windows improved the selectivity of the SWATH analysis of metabolites in urine.¹² The investigated mass range in the latter study, however, was substantially higher than in our study (50-915 m/z versus 75-250 m/z). As a result, the difference in ion density throughout the mass range was considerably higher resulting in a larger variation in variable window sizes in order to equalize the ions density in each SWATH windows. Therefore, the fixed and variable SWATH window size differed to greater extent than in our study. This resulted in a more prominent difference in the quality and identification efficiency of the acquired SWATH spectra.

CONCLUSION

Data independent acquisition (DIA) is an attractive technique in mass spectrometry-based metabolomics. It offers the user the possibility to acquire fragmentation data of all analytes in a defined mass range without introducing a bias towards certain analytes. However, the possibility of product ion overlap and complex MS/MS spectra may impair the quantitative and qualitative performance of these MS/MS scans. Our work demonstrates that the performance of DIA methodologies is highly dependent on the type of chromatography and the organization of the Q1 filters. Fast chromatography in combination with wide Q1 filters tends to result in fragmentation

data that are often unusable for qualitative and quantitative purposes. In contrast, smaller-sized Q1 windows (i.e. SWATH) in combination with hydrophilic interaction liquid chromatography (HILIC) separation resulted in correct metabolite identifications and more often in accurate quantification values. Especially customized SWATH windows (i.e. variable SWATH) demonstrated a superior performance, since it allowed for the accurate quantification of 10 structural isomers. Apart from one compound, the quantification accuracies remained good when these isomers were quantified in ten highly diverse blood samples demonstrating the robustness of the analysis. Although a window size of 25 Dalton is common practice in proteomics, we have experienced inaccurate quantifications using fixed SWATH windows of six Daltons in our metabolomics application. Therefore, it seems that the density of precursor ions per mass unit is higher in metabolomics compared to proteomics. Since we have shown good qualitative and quantitative performances using variable SWATH windows of around six Dalton, we would recommend the use of this window size in metabolomics applications in order to maintain a proper selectivity and speed during fragmentation. The use of variable SWATH windows and a HILIC separation are promising for combined targeted and untargeted platforms since it allows for both quantification and identification. Targeted platforms involving SWATH can be re-interrogated at a later stage to identify interesting unknown features. On the other hand, untargeted platforms using SWATH can also be used to quantify critical isomer pairs. The combination of targeted and untargeted metabolomics platforms should facilitate the analytical method development workflow and increase the information provided by a single analysis.

ACKNOWLEDGEMENTS

The authors are grateful to receive funding for this research from The Netherlands Organization for Scientific Research (NWO) in the framework of the Technology Area TA-COAST (Fund New Chemical Innovations. project no. 053.21.118).

REFERENCES

1. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
2. Mayo Clinic, M. M. L. Rochester 2018 Interpretive Handbook. (2018). doi:10.1016/S1002-0721(13)60014-9
3. Trivedi, D. K., Hollywood, K. A. & Goodacre, R. Metabolomics for the masses: The future of metabolomics in a personalized world. *New Horizons Transl. Med.* **3**, 294–305 (2017).
4. Balashova, E., Maslov, D. & Lokhov, P. A Metabolomics Approach to Pharmacotherapy Personalization. *J. Pers. Med.* **8**, 28 (2018).
5. Xiao, J. F., Zhou, B. & Resson, H. W. Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *TrAC - Trends Anal. Chem.* **32**, 1–14 (2012).
6. Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–269 (2012).
7. Roberts, L. D., Souza, A. L., Gerszten, R. E. & Clish, C. B. Targeted metabolomics. *Curr Protoc*

- Mol Biol.* 1–34 (2012). doi:10.1002/0471142727.mb3002s98.Targeted
8. Fenaille, F., Barbier Saint-Hilaire, P., Rousseau, K. & Junot, C. Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand? *J. Chromatogr. A* **1526**, 1–12 (2017).
 9. Wang, R., Yin, Y. & Zhu, Z. J. Advancing untargeted metabolomics using data-independent acquisition mass spectrometry technology. *Anal. Bioanal. Chem.* 4349–4357 (2019). doi:10.1007/s00216-019-01709-1
 10. Stahl, D. C., Swiderek, K. M., Davis, M. T. & Lee, T. D. Data-controlled automation of liquid chromatography/tandem mass spectrometry analysis of peptide mixtures. *J. Am. Soc. Mass Spectrom.* **7**, 532–540 (1996).
 11. Johnson, D., Boyes, B., Fields, T., Kopkin, R. & Orlando, R. Optimization of data-dependent acquisition parameters for coupling high-speed separations with LC-MS/MS for protein identifications. *J. Biomol. Tech.* **24**, 62–72 (2013).
 12. Zhang, Y. *et al.* The Use of Variable Q1 Isolation Windows Improves Selectivity in LC-SWATH-MS Acquisition. *J. Proteome Res.* **14**, 4359–4371 (2015).
 13. Bonner, R. & Hopfgartner, G. SWATH acquisition mode for drug metabolism and metabolomics investigations. *Bioanalysis* **8**, 1735–1750 (2016).
 14. Plumb, R. S. *et al.* UPLC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation Robert. *Rapid Commun. Mass Spectrom.* 4129–4138 (2008). doi:10.1002/rcm.2550
 15. Navarro-Reig, M. *et al.* Metabolomic analysis of the effects of cadmium and copper treatment in: *Oryza sativa* L. using untargeted liquid chromatography coupled to high resolution mass spectrometry and all-ion fragmentation. *Metallomics* **9**, 660–675 (2017).
 16. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, 1–17 (2012).
 17. Dubbelman, A. C. *et al.* Mass spectrometric recommendations for Quan/Qual analysis using liquid-chromatography coupled to quadrupole time-of-flight mass spectrometry. *Anal. Chim. Acta* **1020**, 62–75 (2018).
 18. The Human Serum Metabolome. <https://serummetabolome.ca/>
 19. Yang, X. *New Features of the 2017 NIST Tandem Mass Spectral Library.* (2017).
 20. mzCloud – Statistics. Available at: <https://www.mzcloud.org/Stats>.
 21. Yan, Z., Maher, N., Torres, R., Cotto, C. & Hastings, B. Isobaric metabolite interferences and the requirement for close examination of raw data in addition to stringent chromatographic separations in liquid chromatography/tandem mass spectrometric analysis of drugs in biological matrix. *Rapid Commun. Mass Spectrom.* **22**, 2021–2028 (2008).
 22. Martens-Lobenhoffer, J., Surdacki, A. & Bode-Böger, S. M. Fast and precise quantification of l-homoarginine in human plasma by hilic-isotope dilution-MS-MS. *Chromatographia* **76**, 1755–1759 (2013).
 23. Armirotti, A., Millo, E. & Damonte, G. How to Discriminate Between Leucine and Isoleucine by Low Energy ESI-TRAP MSn. *J. Am. Soc. Mass Spectrom.* **18**, 57–63 (2007).
 24. Martens-Lobenhoffer, J. & Bode-Böger, S. M. Chromatographic-mass spectrometric methods for the quantification of l-arginine and its methylated metabolites in biological fluids. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **851**, 30–41 (2007).
 25. Human Metabolome Database. Available at: <http://www.hmdb.ca/>.
 26. Laan, T. Van Der *et al.* High-Throughput Fractionation Coupled to Mass Spectrometry for Improved Quantitation in Metabolomics. *Anal. Chem.* (2020). doi:10.1021/acs.analchem.0c01375
 27. Bilbao, A. *et al.* Processing strategies and software solutions for data-independent acquisition in mass spectrometry. *Proteomics* **15**, 964–980 (2015).
 28. Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1**, 39–45 (2004).

29. Collins, B. C. *et al.* Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. *Nat. Commun.* **8**, 1–11 (2017).
30. Roemmelt, A. T., Steuer, A. E., Poetzsch, M. & Kraemer, T. Liquid chromatography, in combination with a quadrupole time-of-flight instrument (LC QTOF), with sequential window acquisition of all theoretical fragment-ion spectra (SWATH) acquisition: Systematic studies on its use for screenings in clinical and foren. *Anal. Chem.* **86**, 11742–11749 (2014).
31. Scheidweiler, K. B., Jarvis, M. J. Y. & Huestis, M. A. Nontargeted SWATH acquisition for identifying 47 synthetic cannabinoid metabolites in human urine by liquid chromatography-high-resolution tandem mass spectrometry. *Anal. Bioanal. Chem.* **407**, 883–897 (2015).
32. Tsugawa, H. *et al.* MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **12**, 523–526 (2015).
33. Zhu, X., Chen, Y. & Subramanian, R. Comparison of information-dependent acquisition, SWATH, and MS All techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *Anal. Chem.* **86**, 1202–1209 (2014).

SUPPLEMENTARY INFORMATION

Table S1. The standards of the analytes with corresponding internal standard correction, supplier and C8 concentration.

Name	ChEbi ID	Internal standard correction		Standard	Supplier	Solvent	C8 (μM)
		<i>Fractionation</i>	<i>HILIC</i>				
1-Met	70958	1-Met-D3	1-Met-D3	1-methylhistidine	HMDB	H ₂ O	640
3-Met	70959	1-Met-D3	1-Met-D3	3-methylhistidine	HMDB	H ₂ O	320
Bet	17750	Val-D8	Bet-D9	Betaine hydrochloride	Sigma-Aldrich	H ₂ O	1200
Val	27266	Val-D8	Val-D8	L-valine	Fluka	H ₂ O	3200
Ile	17191	Leu-D3	Leu-D3	L-Isoleucine	Fluka	H ₂ O	800
Leu	15603	Leu-D3	Leu-D3	L-Leucine	Fluka	H ₂ O	800
hArg	27747	Arg- ¹⁵ N ₂	Lys-D4	L-homoarginine HCl	Sigma-Aldrich	H ₂ O	80
NMMA	28229	Arg- ¹⁵ N ₂	1-Met-D3	NG-methyl-L-Arginine Acetate Salt	Sigma-Aldrich	H ₂ O	16
ADMA	17929	Arg- ¹⁵ N ₂	1-Met-D3	NG,NG-Dimethylarginine dihydrochloride	Sigma-Aldrich	H ₂ O	16
SDMA	25682	Arg- ¹⁵ N ₂	1-Met-D3	NG,NG'-Dimethyl-L-arginine di(p-hydroxyazobenzene-p'-sulfonate)	Sigma-Aldrich	H ₂ O	16

Table S2. The internal standards with corresponding supplier and C4 concentration.

Name	Internal standard	Supplier	Solvent	Concentration (μM)
Val-D8	L-valine (D8)	Cambridge isotopes	H ₂ O	200
Leu-D3	DL-leucine-D3	CDN Isotopes	1M HCl	60
Lys-D4	L-lysine 2 HCl (4,4,5,5-D4)	Cambridge isotopes	H ₂ O	180
Arg- ¹⁵ N ₂	L-arginine- ¹⁵ N ₂ hydrochloride	Cortecnet	H ₂ O	100
Bet-D9	N-(Carboxymethyl)-N,N,N-trimethyl-d9-ammonium Chloride	CDN Isotopes	H ₂ O	75
1-Met-D3	1-methyl-d3-L-histidine	CDN Isotopes	H ₂ O	80

Table S3. Clinical variables of ten healthy male subjects.

Gender	Age	Race	BMI	Fasted	Smoker
Male	33	Hispanic	31	YES	NO
Male	56	African American	35	NO	NO
Male	61	Hispanic	20	NO	YES
Male	24	Hispanic	31	NO	YES
Male	33	Hispanic	30	NO	NO
Male	56	African American	33	NO	YES
Male	34	African American	27	NO	YES
Male	56	African American	41	YES	NO
Male	54	Hispanic	30	NO	NO
Male	32	Hispanic	31	NO	YES

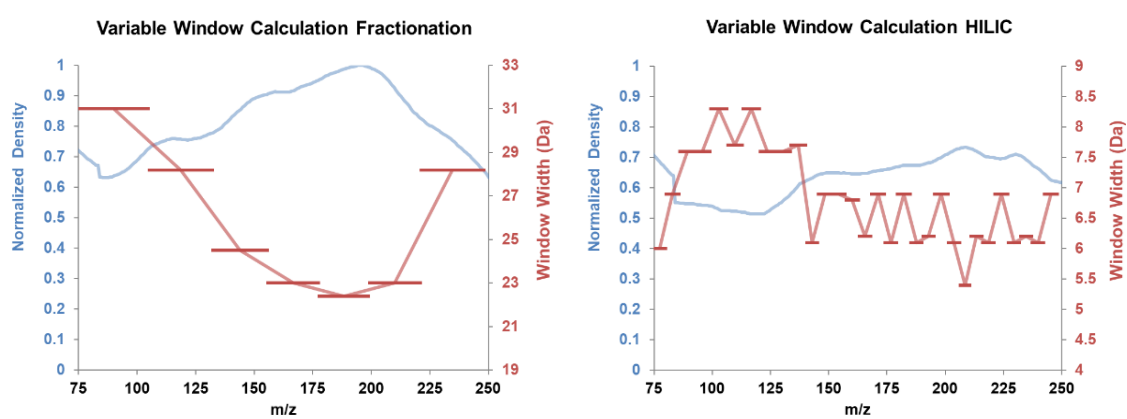


Figure S1. Variable window calculator V1.1 results. The left y-axis demonstrates the normalized density of the MS intensities of a full scan (75-250 m/z). The right y-axis demonstrates the window width of the calculated SWATH windows. The HILIC method has 30 windows and the fractionation method has 7 windows.

Table S4. SWATH window sizes for the fractionation method.

SWATH windows	Variable SWATH		Fixed SWATH	
	Start mass	Stop mass	Start mass	Stop mass
1	74.5	105.5	75	101
2	104.5	132.7	100	126
3	131.7	156.2	125	151
4	155.2	178.2	150	176
5	177.2	199.6	175	201
6	198.6	221.6	200	226
7	220.6	248.8	225	250

Table S5. SWATH window sizes for the HILIC method.

SWATH windows	Variable SWATH		Fixed SWATH	
	<i>Start mass</i>	<i>Stop mass</i>	<i>Start mass</i>	<i>Stop mass</i>
1	74.5	80.5	75	81
2	79.5	86.4	80	87
3	85.4	93	86	93
4	92	99.6	92	99
5	98.6	106.9	98	105
6	105.9	113.6	104	111
7	112.6	120.9	110	117
8	119.9	127.5	116	123
9	126.5	134.1	122	129
10	133.1	140.8	128	135
11	139.8	145.9	134	141
12	144.9	151.8	140	147
13	150.8	157.7	146	153
14	156.7	163.5	152	159
15	162.5	168.7	158	165
16	167.7	174.6	164	171
17	173.6	179.7	170	177
18	178.7	185.6	176	183
19	184.6	190.7	182	189
20	189.7	195.9	188	195
21	194.9	201.8	194	201
22	200.8	206.9	200	207
23	205.9	211.3	206	213
24	210.3	216.5	212	219
25	215.5	221.6	218	225
26	220.6	227.5	224	231
27	226.5	232.6	230	237
28	231.6	237.8	236	243
29	236.8	242.9	242	249
30	241.9	248.8	248	250

Table S6. The quantification accuracy of structural isomers in comparison with a MRM^{HR}. Values between 85-115% are indicated in green. Values between 115-120% are indicated in yellow. Values outside these ranges are indicated in red. Compounds that could not be quantified due to an insufficient linearity (<0.99), high variability (>15%) or integration problems (peak overlap or too high baseline) are indicated by the zero values.

Analyte	Product ions (m/z)	Variable SWATH (%)		Fixed SWATH (%)		MS ^{ALL} (%)	
		Fractionation	HILIC	Fractionation	HILIC	Fractionation	HILIC
ADMA	46.0651	88	96	0	102	0	0
	112.0873	0	119	0	121	0	0
	114.1028	233	111	0	119	0	0
SDMA	126.1028	0	86	0	94	183	0
	172.1081	107	101	96	110	106	0
hArg	60.0560	109	122	0	117	0	0
	84.0810	0	112	0	91	0	0
	85.0651	191	0	389	126	271	0
	86.0967	0	99	0	102	0	0
	127.0870	0	141	0	159	0	0
	130.0866	0	104	0	95	0	0
	147.1133	0	95	0	0	0	0
NMMA	155.0818	0	89	0	140	0	0
	115.0870	739	136	6383	226	1944	0
Bet	158.0926	0	111	0	0	0	0
	58.0655	106	195	91	0	150	0
	59.0733	0	226	114	0	0	0
	102.0550	0	102	0	0	0	358
Val	117.7023	131	124	103	134	136	124
	55.0544	98	112	93	115	99	104
	57.0582	170	91	167	91	184	101
1-Met	72.0811	0	291	148	258	142	0
3-Met	124.0868	157	113	0	109	0	0
	85.0765	338	109	288	103	0	0
	95.0608	0	96	2120	97	1723	0
Ile	126.1029	97	99	97	89	193	89
	69.0704	106	112	113	99	129	90
Leu	43.0544	144	110	148	117	171	168

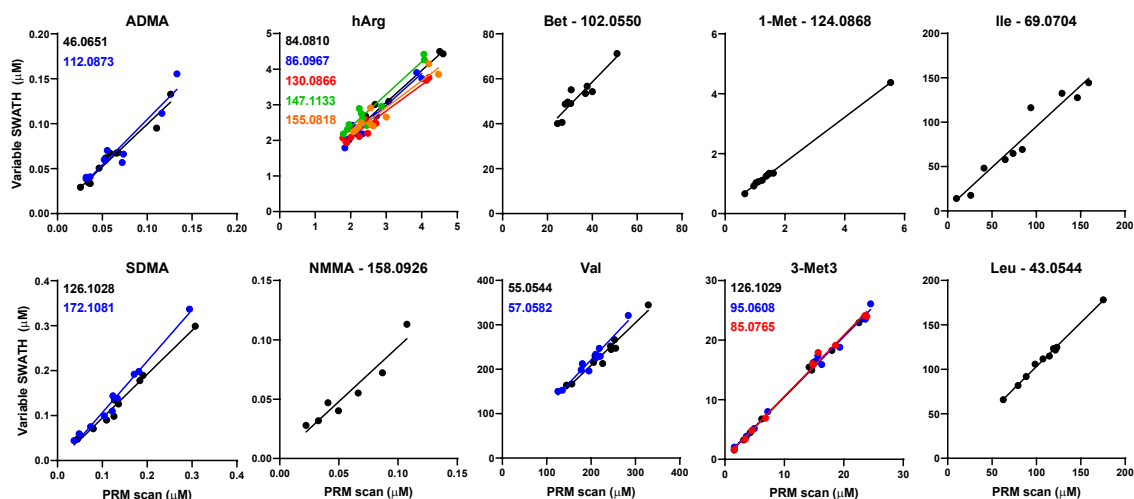


Figure S2. The quantification values of 10 structural isomers in 10 volunteers. Each data point represents the analyte concentration quantified by MRM^{HR} and variable SWATH. Different diagnostic product ions of one structural isomer are plotted in different colours.

Table S7. Correlation and accuracy of the quantification values of 10 structural isomers in 10 volunteers.

Analyte	Product ions (<i>m/z</i>)	Correlation (R^2)	Accuracy (%)
ADMA	46.0651	0.95	104
	112.0873	0.91	104
SDMA	126.1028	0.98	96
	172.1081	0.98	108
hArg	84.0810	0.98	103
	86.0967	0.95	103
	130.0866	0.94	100
	147.1133	0.91	113
	155.0818	0.90	101
NMMA	158.0926	0.91	99
Bet	102.0550	0.86	157
Val	55.0544	0.95	104
	57.0582	0.93	110
1-Met	124.0868	1.00	108
3-Met	85.0765	0.99	105
	95.0608	0.99	105
	126.1029	1.00	104
Ile	69.0704	0.94	105
Leu	43.0544	1.00	97

Chapter 5

Fractionation platform for target identification using off-line directed 2D- LC, MS and NMR

Based on:

Tom van der Laan, Hyung Elfrink, Fatemeh Azadi-Chegeni, Anne-Charlotte Dubbelman, Amy Harms, Doris Jacobs, Ulrich Braumann, Aldrik H. Velders, John van Duynhoven and Thomas Hankemeier

Fractionation platform for target identification using off-line directed two-dimensional chromatography, mass spectrometry and nuclear magnetic resonance

Analytica Chimica Acta (2020)

ABSTRACT

The unambiguous identification of unknown compounds is of utmost importance in the field of metabolomics. However, current identification workflows often suffer from error-sensitive methodologies, which may lead to incorrect structure annotations of small molecules. Therefore, we have developed a comprehensive identification workflow including two highly complementary techniques, i.e. liquid chromatography (LC) combined with mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR), and used it to identify five taste-related retention time and m/z features in soy sauce. An off-line directed two-dimensional separation was performed in order to purify the features prior to the identification. Fractions collected during the first dimension separation (reversed phase low pH) were evaluated for the presence of remaining impurities next to the features of interest. Based on the separation between the feature and impurities, the most orthogonal second dimension chromatography (hydrophilic interaction chromatography or reversed phase high pH) was selected for further purification. Unknown compounds down to tens of micromolar concentrations were tentatively annotated by MS and structurally confirmed by MS and NMR. The mass (0.4-4.2 μg) and purity of the isolated compounds were sufficient for the acquisition of one and two-dimensional NMR spectra. The use of a directed two-dimensional chromatography allowed for a fractionation that was tailored to each feature and remaining impurities. This makes the fractionation more widely applicable to different sample matrices than one-dimensional or fixed two-dimensional chromatography. Five proline-based 2,5-diketopiperazines were successfully identified in soy sauce. These cyclic dipeptides might contribute to taste by giving a bitter flavour or indirectly enhancing umami flavour.

INTRODUCTION

Metabolomics has gained a lot of interest in numerous disciplines like life sciences, environmental sciences and the food industry.^{1,2,3} Despite the tremendous developments in metabolite analysis, metabolomics still encounters several analytical challenges. Metabolite identification is considered as one of the major bottlenecks in present-day metabolomics because of its labor-intensive and error-prone methodologies.⁴ Structure elucidation of unknown compounds is essential in order to translate analytical data into useful information. In 2007, the Metabolomics Standards Initiative (MSI) defined four levels of identification confidence.⁵ Level 1 represents a positive identification of an unknown compound. Level 2 and 3 are the putatively annotation of compounds and compound classes, respectively. Unidentified and unclassified compounds that can be differentiated based on spectral data are classified as level 4. Although these guidelines are dated, they are still used as the standard for metabolite identification. Currently, the Metabolite Identification Task Group of the Metabolomics Society is revising the reporting standards. In this work, we aim for level 1 identification with regards to the MSI guidelines and level C with regards to the newly proposed reporting standards.

Generally, metabolite identification annotates metabolites that have been characterized before, which is referred to as non-novel identification. Non-novel identifications are mainly performed by co-characterization with a synthetic standard. In order to reach level 1 identification, a minimum of two physical and/or chemical properties (e.g. fragmentation pattern, chromatographic retention, NMR spectra) of the unknown metabolite and the synthetic standard have to overlap. Two commonly used analytical techniques that are used for the analysis of such properties are mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy. Combined MS and NMR identification strategies have drawn an increased interest because they are highly complementary which increases the identification confidence of unknown metabolites.^{6,7}

High-resolution mass spectrometers report an accurate mass and intensity ratio of the isotopes that can be sufficient to suggest the elemental composition of a certain m/z value.⁸ The list of possible elemental compositions can be further constrained by using MS/MS data in which a fragment ion has to contain no other elements than the precursor ion and a precursor ion has to contain all elements of the fragment ion.⁹ MS/MS scans also capture structural information of an m/z value and can be compared with spectral libraries to search for identification hits. Once a hit has been found, the metabolite can be purchased/synthesized and measured using the same MS methodology. The retention time and MS/MS spectra of the synthesized compound can be compared with an unknown metabolite and lead to level 1 identification. In case no synthesized

compound is available, confident identifications can be achieved by combining structural information obtained from MS with NMR.^{10,11}

NMR spectroscopy can also be used for level 1 identification by comparing typical spin patterns of a synthesized compound with an unknown metabolite. Main challenges for NMR-based identifications are, however, low sensitivity and signal overlap.¹² Many metabolites consist of similar molecular structural motifs which often result in similar peaks in the NMR spectra.⁶ Signal overlap is especially problematic for low abundance compounds since they are prone to be overshadowed by signals of high abundance compounds. In addition, the concentration of low abundance compounds are often below the detection limits because of the limited sensitivity of NMR.¹³ Signal overlap and insufficient sensitivity can result in missing signals of an unknown compounds and can, therefore, obstruct identification or structure confirmation. Chromatographic separation can be used to drastically improve the power of NMR by isolating target compounds from complex matrices into cleaner fractions which can be up-concentrated for improved sensitivity.^{14,15,16} The purification can even further be improved by including a second orthogonal chromatographic separation.¹⁷ In addition, 2D NMR spectroscopy can be exploited for identification purposes and complemented with 1D NMR quantification methods.

In this study, we demonstrate a platform for the identification of retention time and m/z features in complex samples. In order to unravel the structure of these features, a tentative mass spectrometry-based identification is performed. Thereafter, NMR and MS analyses are used to confirm the structure proposed by MS. Since complex samples consist of a complex matrix and the abundance of certain compounds can be low, we have developed a comprehensive fractionation approach in order to decrease the sample complexity and to increase the concentration prior to the NMR analysis. The fractionation consisted of an off-line directed two-dimensional chromatography, which was tailored to the unknown feature and sample impurities. Fractions collected during the first dimension were evaluated for the presence of impurities. The stationary phase that resulted in the best separation between the unknown feature and the impurities was selected for the second dimension fractionation. We have applied our platform to the identification of five taste-related features in soy sauce. This identification platform provides a general approach for metabolite identification and can be tailored to specific features and samples types. The complementary structure confirmation by MS and NMR ensures a feature identification with high certainty.

MATERIALS AND METHODS

Chemicals and product

Acetonitrile Ultra LC-MS was purchased from Actua-All (Oss, The Netherlands). Formic acid (98%+) was purchased from Acros Organics (Bleiswijk, The Netherlands). Ammonium formate ($\geq 99.995\%$) and cyclo(Pro-Pro) were acquired from Sigma-Aldrich (Zwijndrecht, The Netherlands). Cyclo(Pro-Phe), cyclo(Pro-Gly) and cyclo(Pro-Thr) were purchased from Bachem AG (Bubendorf, Switzerland). Cyclo(Pro-Leu) was obtained from Santa Cruz Biotechnology (Heidelberg, Germany). Deuterated water (D_2O , 99.9%) was purchased from Euriso-top (Saint-Aubin, France). All soy sauce products used in this study were commercially available. The soy sauce that was used for the identification and structure confirmation of the unknown compounds was obtained from a local supermarket in Korea.

Fractionation chromatography and analytical chromatography

The identification platform included in total three off-line fractionation chromatography methods and one analytical chromatography method. All fractionation chromatography methods used an injection volume, flow rate and column temperature of 100 μ L, 2 mL/min and 30 °C, respectively.

The reversed phase low pH (*RP low pH*) method used a C18 Atlantis T3, 4.6 x 150 mm, 3 μ m particle size column. Mobile phase A and B consisted of 0.1% formic acid in water and acetonitrile, respectively. The column oven was set at 30 °C. The gradient started at 0% B and increased linearly to 15% B in 7 minutes. Subsequently, the gradient increased to 55% B in 3 minutes and to 100% B in an additional 0.5 minutes. The gradient was kept at 100% B for 2 minutes to flush the column. The gradient decreased to 0% B in 0.1 min and was kept at this value for 2.4 minutes to equilibrate the column. The total gradient time was 15 minutes.

The reversed phase high pH (*RP high pH*) method used a C18 Kinetex EVO, 4.6 x 150 mm, 5 μ m particle size column. The gradient profile was identical to the *RP low pH* gradient, but mobile phase A and B consisted of 2 mM ammonium formate pH 9 in water and 95/5 (v/v) acetonitrile/water, respectively.

The hydrophilic interaction chromatography (*HILIC*) method used a Sequant ZIC-*HILIC*, 4.6 x 100 mm, 5 μ m particle size column. Mobile phase A consisted of 10 mM ammonium formate and 0.075/90/10 (v/v/v) formic acid/acetonitrile/water and mobile phase B of 10 mM ammonium formate and 0.075/10/90 (v/v/v) formic acid/acetonitrile/water. The gradient started at 0% B for 1.2 minutes. The gradient increased linearly to 75% B in 7.96 minutes and was kept at this value for 6.04 minutes to flush the column. The gradient decreased to 0% B in 0.2 minutes and stayed at this value for 4.8 minutes to equilibrate the column. The total gradient time was 20.2 minutes.

The analytical chromatography *RP low pH* method used a C18 Acquity UPLC HSS T3, 2.1 x 100 mm, 1.8 μm particle size column. The injection volume and flow rate were 1 μL and 0.4 mL/min, respectively. The gradient profile was identical to the fractionation *RP low pH* method.

For the fractionation, liquid chromatography (LC) separations were performed on a Waters Acquity UPLC system (Waters, Etten-Leur, The Netherlands) and fractions were collected in 0.35 min time windows using a Waters Fraction Manager (Waters, Etten-Leur, The Netherlands). The analytical chromatography and three fractionation chromatography methods that were coupled to MS were performed on a Shimadzu Nexera UHPLC (Darmstadt, Germany).

MS analysis

MS analyses were performed on a Sciex X500R QToF (Darmstadt, Germany). The fractionation chromatography methods were coupled to MS via a 1:20 flow split in order to locate the features in the acquired fractions, assess the orthogonality of the second dimension chromatography and to identify the features. The analytical *RP low pH* method was directly coupled to the MS and was used for structure confirmation and quantification by MS.

The exact mass and isotope pattern were analyzed using Sciex OS v1.5. MS/MS scans were acquired in product ion scan mode and matched to three commonly used spectral libraries: NIST 2017, the MassBank of North America (MoNA) and mzCloud. NIST 2017 was searched in Sciex OS 1.5 by a candidate search. MoNA (<https://mona.fiehnlab.ucdavis.edu>) and mzCloud (<https://www.mzcloud.org>) were searched via the web interface. The mass tolerance of the spectral search was set at 0.01 Dalton. When this mass tolerance did not result in a confident hit, the mass tolerance was increased to 0.3 Dalton. Features that did not result in a confident library hit were evaluated further by searching their elemental composition in the Dictionary of Natural products (DNP, dnp.chemnetbase.com). The standards of the proposed structures were used for structure confirmation by comparing the retention time and MS/MS spectra of the standard and feature.

The original concentration of the identified compounds in soy sauce were determined by means of standard addition. Known concentrations of the standards were spiked to soy sauce at seven different levels (C1-C7). C7 was the highest calibration concentration and contained all the standards at relevant concentrations. The subsequent calibration levels were all 1:1 dilutions of the previous calibration point. Calibration samples were prepared by mixing 10 μL soy sauce, 10 μL calibration standards and 80 μL H₂O. C0 represented the non-spiked soy sauce and was prepared by replacing the calibration standards by H₂O. In order to construct a calibration line, the peak area obtained by analysis of the calibration samples was plotted against the concentration of the calibration standards. The original concentration of the identified

compounds was determined by dividing the intercept of the calibration line by the slope of the calibration line.

NMR analysis

1D and 2D NMR spectra were recorded on 600 MHz NMR spectrometers equipped with cryoprobes suitable for sample tube diameters of 1.7 (TCI MicroCryoProbe) and 5 mm (TCI CryoProbe). In both cases samples were measured in 1.7 mm (30 μ l) sample tubes. The pulse sequences for the 2D heteronuclear single quantum coherence (HSQC), total correlation spectroscopy (TOCSY), double quantum filtered homonuclear correlation spectroscopy (DQF COSY) and heteronuclear multiple-bond correlation spectroscopy (HMBC) experiments were taken from the Bruker library. 1D ^1H NMR spectra were deconvoluted with the CHENOMX software.

The concentration of the identified compounds in the final fraction and the total proton concentration in non-fractionated soy sauce and the final fraction were determined using the Pulse Length-based Concentration determination (PULCON) methodology¹⁸ on a Bruker 600 MHz NMR spectrometer equipped with an Avance III HD console and a 5 mm TCI CryoProbe probehead. PULCON is an external method in which the absolute intensities in two different one-dimensional NMR spectra are correlated based on the principle of reciprocity.^{19,20,21} Based on this principle, the 90° pulse length is inversely proportional to the NMR signal strength for a sample in a given r.f. coil. Therefore, if the concentration of a reference (ref) sample is known and the 90° pulse is precisely calibrated, the unknown concentration (U) of each analyte can be obtained using the following formula:

$$C_U = kC_{ref} \frac{A_U T_U \theta_{90}^U n_{ref}}{A_{ref} T_{ref} \theta_{90}^{ref} n_U}$$

A indicates the integral over the resonances, T , the sample temperature in Kelvin, θ_{90} , the 90° pulse length, and n , the number of transients used for the measurement of the reference and the unknown sample. The correction factor k stands for any differences between the experiments such as incomplete relaxation or different receiver gains which results in variation in signal intensities of the reference and the analyte. The equation is valid when the experiments are performed with the same NMR probe, which is properly tuned and matched to the same amplifier. Trimethylsilylpropanoic acid (TSP) at a concentration of 11.6 mM was used as an external reference. For quantification measurements, each 1D ^1H NMR experiment was recorded with a 27-second relaxation delay and 128 number of scans at 300K. For each fraction, 90° pulse length was calibrated individually. For quantification purposes, all spectra were manually phase and baseline corrected and integral regions were set manually. The quantification of the identified

compounds in the final fraction was based on manual integration of assigned and non-overlapping signals. The proton purity of the final fractions and non-fractionated soy sauce were defined as the ratio between the proton concentration of the identified compounds and the total proton concentration excluding the water peak. The concentrations of the identified compounds in non-fractionated soy sauce were determined by mass spectrometry as described in the previous section. The other concentrations were determined by PULCON.

$$\textit{Proton purity (soy sauce/final fraction)} = \frac{[\text{H}^1]\textit{identified compound}}{[\text{H}^1]\textit{total}}$$

Identification workflow

The identification workflow is depicted in Figure 1. The workflow started with m/z and retention time features in a biological matrix that correlated with a certain effect. **(1)** The complex matrix was fractionated using the first dimension chromatography (*RP low pH*). Ten injections of 100 μL were fractionated resulting in a total fraction volume of 7 mL. **(2)** 10 μL of the first dimension fractions was analyzed using the first dimension fractionation coupled to mass spectrometry. The fraction containing the highest abundance of the feature was selected and evaluated for remaining impurities. **(3)** 50 μL of the selected first dimension fractions was analyzed using the second dimension chromatography (*HILIC* and *RP high pH*) coupled to mass spectrometry. The orthogonality of the second dimension was evaluated based on the separation between the feature and remaining impurities. The chromatography phase that resulted in the most efficient isolation of the features was selected for further purification. **(4)** The selected first dimension fractions were freeze-dried and reconstituted in 300 μL of mobile phase A (*RP high pH*) or 50/50 (v/v) acetonitrile/water (*HILIC*). Two injections 100 μL were fractionated using the selected second dimension chromatography resulting in a total fraction volume of 1.4 mL. **(5)** 10 μL of the second dimension fractions was analyzed using the second dimension fractionation coupled to mass spectrometry. The features were identified by mass spectrometry and feature-containing fractions were selected. **(6)** The proposed structure was confirmed by MS by comparing the standard with the feature in the full biological matrix using the analytical chromatography method coupled to mass spectrometry. **(7)** The selected second dimension fractions were evaporated and reconstituted in 40 μL D_2O for NMR analysis reaching a maximum concentration factor of 17. The proposed structure was confirmed by NMR by comparing the standard with the purified feature after the second chromatography dimension.

Complex food application

In an inhouse study performed at Unilever, we have found five retention time and m/z features that were strong predictors for the sensory scores of fermented soybean flavour (Figure S1 in the supplementary information). These features were measured by reversed phase LC-MS wherein in

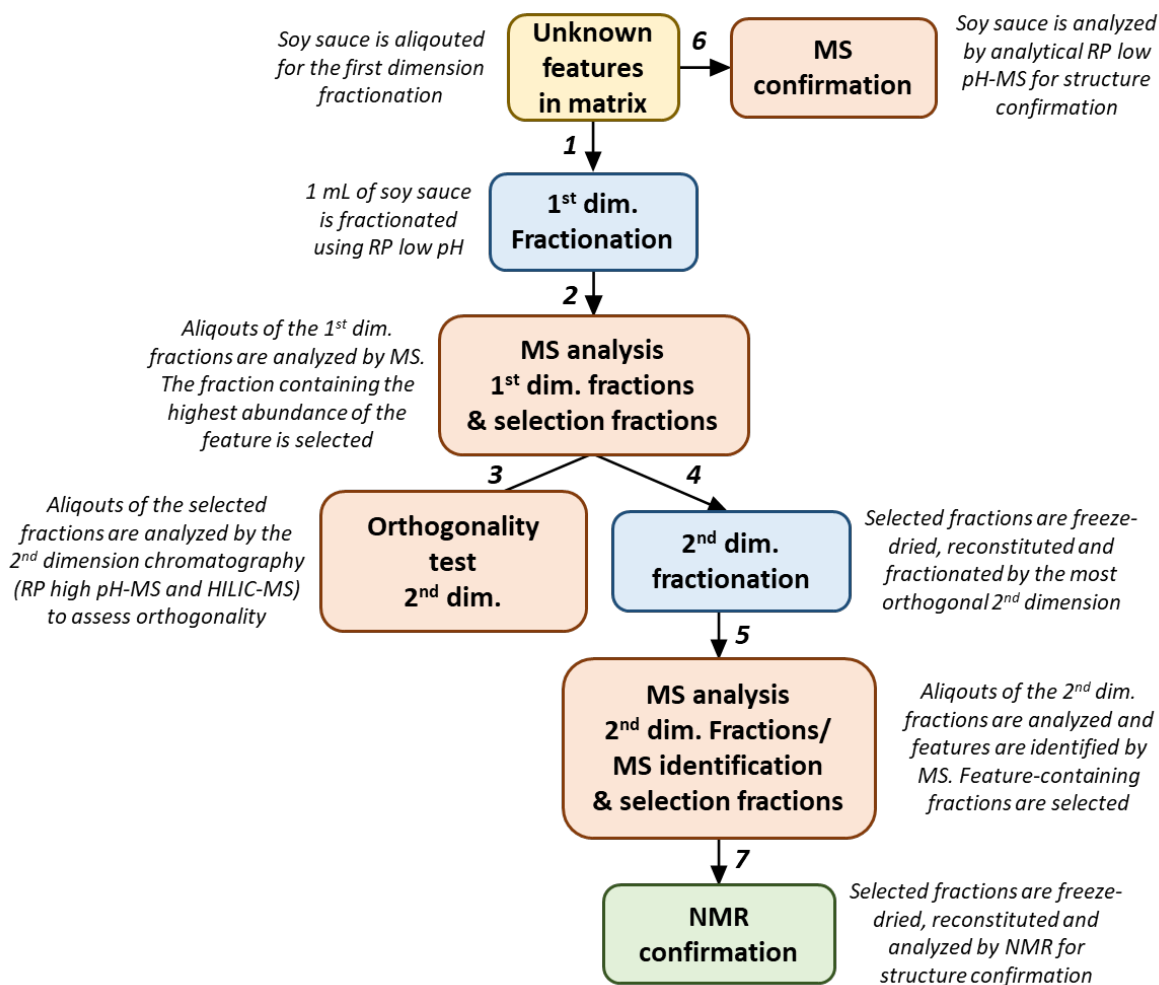


Figure 1. The MS/NMR-based identification workflow.

total 622 features were found (method can be found in Table S1 in the supplementary information). The soy sauce with the strongest fermented soybean flavour according to the partial least squares (PLS) model was selected for the identification workflow. This particular soy sauce demonstrated the highest abundance of the five taste-related features.

RESULTS AND DISCUSSION

In this study, we aimed for structure identification of retention time and m/z features followed by structure confirmation by two complementary techniques: mass spectrometry (MS) and nuclear magnetic resonance (NMR). Because of the high complexity of complex samples and the low apparent concentration of some unknown compounds, a thorough fractionation procedure was often required in order to yield clean and concentrated fractions for NMR analysis. A directed two-dimensional chromatography is used to tailor the fractionation procedure per individual feature. The contaminants that are present after the first dimension chromatography were used to select

the most orthogonal chromatography for the second dimension. This resulted in an efficient isolation procedure that is adjustable to different features and sample matrices.

Sample fractionation

We have applied the identification platform that is depicted in Figure 1 to isolate five features in soy sauce. The fractionation procedure is demonstrated by an example (feature A) in Figure 2. After the first dimension chromatography, the mass spectrum of the feature-containing fraction was compared with a blank fraction. The comparison with a blank fraction demonstrated which peaks resulted from the mobile phase and which peaks resulted from the fraction itself. m/z peaks present in the blank were attributed to the mobile phase and other m/z peaks were selected as undesired contaminants from soy sauce. Thereafter, the fractions were injected into the second dimension chromatography methods, i.e. fractionation reversed phase high pH (*RP high pH*) and fractionation *HILIC*. The m/z values of the feature of interest as well as the contaminants were extracted in order to assess the orthogonality of the second dimension. The method that resulted in the cleanest separation of the feature from the contaminants was selected for further fractionation.

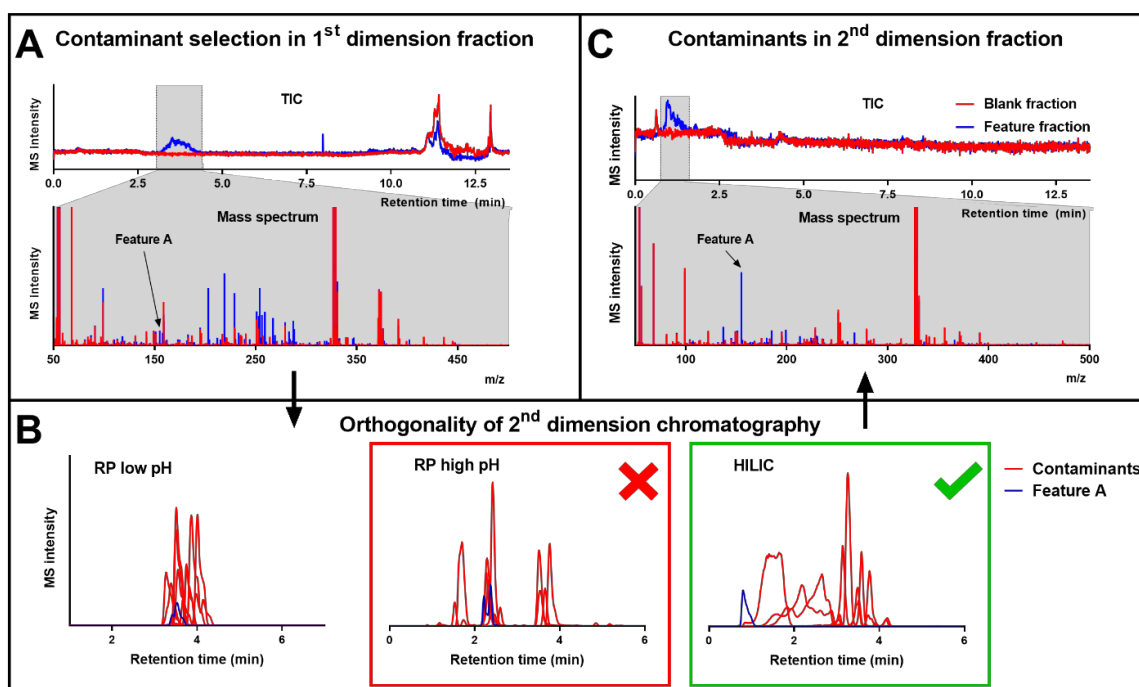


Figure 2. The fractionation workflow of feature A. (1) The mass spectrum (ranging over the grey area) of the feature-containing and blank fraction measured in the first dimension. m/z values that are present in the fraction next to the feature are selected as contaminants. (2) The orthogonality of the second dimension chromatography methods is assessed by extracting the feature and contaminant m/z values. (3) Contamination of the feature-containing fraction is assessed after the second dimension chromatography.

An overview of the identified features is provided in Table 1 and the orthogonality assessment is shown in Figure S2 of the supplementary information. As Figure S2 shows, the fractions obtained after the first dimension still contained highly abundant contaminants (represented by the red MS signals). Therefore, the purity of these fractions could still be improved by adding a second dimension chromatography step. For feature A, *HILIC* was selected for further purification as feature A clearly demonstrated the least overlap with the contaminants in this method. *RP high pH* was selected as the second dimension for feature B and E because of similar reasons. *HILIC* and *RP high pH* both demonstrated a powerful cleanup efficiency for feature C, as both methods resulted in little to no overlap with the contaminants. In this case, the selection of the second dimension was based on peak shape, which was substantially better in the *RP high pH* method. Smaller peak widths are beneficial for fractionation as it allows for a higher yield of the feature per fraction. Although fractions could have been collected over a longer time, we tried to minimize the collection time per fraction as much as possible. The chance of contaminant introduction into the fraction increases with the collection time. Even if the chromatogram looks clean around the feature of interest, undetected contaminants (neutral and/or negatively charged compounds) can still be present and possibly interfere with the NMR analysis. The second dimension fractionation of feature D demonstrated that not only contaminants of other masses but also isomeric contaminants should be taken into account. For this feature, *RP high pH* was able to separate three peaks for the feature mass whereas with *HILIC* they eluted as one. Isomeric contaminants may interfere with NMR structure confirmation to a greater extent than other contaminants because of the high chemical similarity.

The use of one-dimensional chromatography for compound fractionation is well studied and common practice in structural elucidation using NMR.^{22,23} However, as demonstrated by our results, one-dimensional chromatography fractionation still leaves highly abundant contaminants in the purified fractions. Second dimension chromatography ideally provides orthogonality to the first dimension resulting in cleaner fractions. The majority of our results demonstrated that the combination of *RP low pH* and *RP high pH* resulted in the most efficient fractionation. This is in

Table 1. Overview of the identified features.

Feature ID	A	B	C	D	E
Mass (<i>m/z</i>)	155.082	195.113	199.108	211.14	245.13
Soy conc. (mM)	0.02	0.2	0.02	0.08	0.02
2nd dimension	<i>HILIC</i>	<i>RP high pH</i>	<i>RP high pH</i>	<i>RP high pH</i>	<i>RP high pH</i>
Library	DNP	DNP	MoNA	MoNA	MoNA
Cyclo	Pro-Gly	Pro-Pro	Pro-Thr	Pro-Leu	Pro-Phe

accordance to a study of Gilar *et al.*, in which was demonstrated that the highest peak capacity was obtained by two RP chromatography methods that used significantly different pH conditions.²⁴ However, the combination of *RP low pH* and *RP high pH* resulted in very little orthogonality for feature A, which highly benefitted from the combination of *RP low pH* and *HILIC*. Although orthogonality between two chromatography sorbents types may have been proven, orthogonality is still highly dependent on the chemistry of the targeted compounds.²⁵ This is emphasized in our work because different first dimension fractions required different second dimension chromatography methods.

MS identification and structure confirmation

The MS/MS spectra of the soy features were uploaded to three commonly used spectral libraries. The confidence of the library hits was assessed by two criteria. Firstly, the elemental composition of the library precursor ion had to be similar to the elemental composition suggested by the exact mass and isotope pattern of the soy feature. Secondly, the product ion m/z values found in the library hit had to be present in the soy feature MS/MS spectra. A library hit was labeled as confident when it met the previously described criteria whereas a library hit was labeled as poor when it met one or none of the criteria. The search in the NIST 2017 library did not result in any library hit. The search in the mzCloud library resulted in poor library hits. The search in the MoNA library resulted in three confident hits, i.e. cyclo(Pro-Thr), cyclo(Pro-Leu) and cyclo(Pro-Phe) for feature C, D and E, respectively.

The experimental MS/MS spectra of two features (A and B) did either result in no library hit or poor library hits. Therefore, their elemental composition was searched in the dictionary of natural products, which resulted in a list of potential candidates. Thereafter, all experimental MS/MS spectra were compared and evaluated for the presence of known substructures, which is common practice in metabolite identification.²⁶ All experimental MS/MS spectra contained the fragment 70.07 m/z , which is the immonium ion for proline.²⁷ The three features that could be matched using the MoNA spectral library indeed contained a proline amino acid in addition to another amino acid in a 2,5-diketopiperazine structure. Since the MS/MS spectra of the other two features also contained the characteristic fragment 70.07 m/z , it was hypothesized that these features also contained a proline amino acid. The molecular formula search in the dictionary of natural products revealed proline-based structures for the two remaining features.

The five suggested dipeptides were purchased, measured and compared with the soy features. Figure S3 in the supplementary information demonstrates that the MS/MS spectra of the standards and soy features overlap. The highest abundant product ions of the standards were all found in the MS/MS spectra of the soy features. Some additional product ions that were found in the soy MS/MS spectra were most likely caused by impurities. Since soy sauce is a complex matrix,

it can be expected that some coeluting precursor ions with a similar nominal mass pass through the Q1 filter in addition to the studied feature. The structures of the identified features are depicted in Figure 3.

NMR structure confirmation and quantification of purified fractions

For all the purified soy features and dipeptide standards, two-dimensional homo- (DQCOSY, TOCSY) and heteronuclear (HSQC, HMBC) NMR spectra were recorded. An overview of the ^1H and ^{13}C NMR signal assignments is given in Table 2. The soy features were of sufficient purity and concentration to recognize the typical ^1H TOCSY NMR spin patterns of the amino acids present in the suggested dipeptides.²⁸ For several dipeptides the ^1H NMR spectral assignments could be complemented with assignments of ^{13}C NMR resonances. These assignments were all in the expected regions. The ^1H NMR signals of the standards were identical to the ones identified in the fractions (see Table S2 in the supplementary information).

The overlaid 2D ^1H - ^1H COSY spectra of feature B (blue) and cyclo(Pro-Pro) (red) are presented in Figure 4. The data illustrates that the concentration of the features in the complex mixture is sufficient for the recording of 2D spectra with a small volume MicroCryoProbe optimized for mass sensitivity. It also shows that the resonances of the soy feature and the dipeptide standard are identical, thus confirming the spectral assignments for cyclo(Pro-Pro) in Table 2. Moreover, the majority of the correlations in the spectra of the fraction were arising from cyclo(Pro-Pro) indicating the high purity of the fraction. For the other features, spectral assignments were also performed by the 2D ^1H - ^1H COSY spectra of the corresponding dipeptide (see supplementary information Figure S4).

Figure 5 presents the 1D spectra of feature B. The non-overlapping resonances of cyclo(Pro-Pro) between 2.25-2.37 ppm, 3.41-3.60 ppm and 4.40-4.48 ppm were used for quantification. The number of underlying protons was accounted for and thus a concentration of 0.6 mM was determined, as it is stated in Figure 5. In a similar manner, the other purified soy features were quantified. The results are summarized in Table 2.

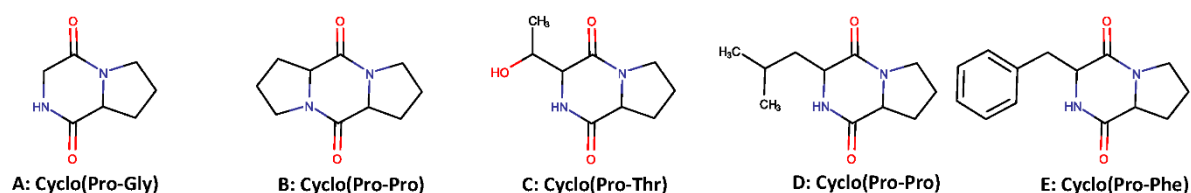


Figure 3. Structural formulas of identified features. No chirality is indicated and it is assumed that only L-amino acids are present.

Table 2. The concentration and the mass of the dipeptides in the final fractions, the ratio between the final fractions and non-fractionated soy sauce with regards to concentration (dipeptide concentration final fraction/soy sauce) and proton purity (dipeptide proton purity final fraction/soy sauce) and $^1\text{H}/^{13}\text{C}$ NMR assignments of soy features by spectral comparison with the dipeptide standards. *The signals of cyclo(Pro-Pro) appeared twice as high with respect to the other dipeptides because of the symmetry. This was taken into account during the quantification.

Cyclo		Pro-Gly		Pro-Pro		Pro-Thr		Pro-Leu		Pro-Phe	
Final fraction	Conc. (mM)	0.3		0.6*		0.05		0.3		0.1	
	Mass (μg)	1.7		4.7*		0.4		2.5		1.2	
Final fraction Soy sauce	Conc.	15		3		3		4		5	
	Proton purity (10^2)	13		5		8		1		16	
Nuclei analysis		^{13}C	^1H	^{13}C	^1H	^{13}C	^1H	^{13}C	^1H	^{13}C	^1H
Other Amino acid	C α	56.6	3.89 4.18			63	4.15	56.6	4.29	59.2	4.59
	C1					68.8	4.45	40.6	1.65 1.83	40.2	3.13 3.28
	C2										
	C3							26.8	1.81		
	C4										
	Me $_1$					21.2	1.31	23.9	0.92		
	Me $_2$							24.9	0.94		
	Φ 1									132.9	7.22
	Φ 2									131.5	7.39
	Φ 3									130.2	7.32
Proline	(C=O) C1	--	4.33	63.5	4.43	61.6	4.30	--	4.34	--	4.06
	C2	--	1.96	30	2.00	30.9	1.91	--	1.97	--	0.79
				2.34		2.31		2.35		2.33	
	C3	--	1.96	25.5	2.00	24.5	1.99	24.4	1.97	--	
			2.07		2.09		2.08	--	2.09		
NH-C4	--			3.56	48.1	3.66				--	3.34
			3.55	48.0	3.47	48.1	3.51		3.56		3.5

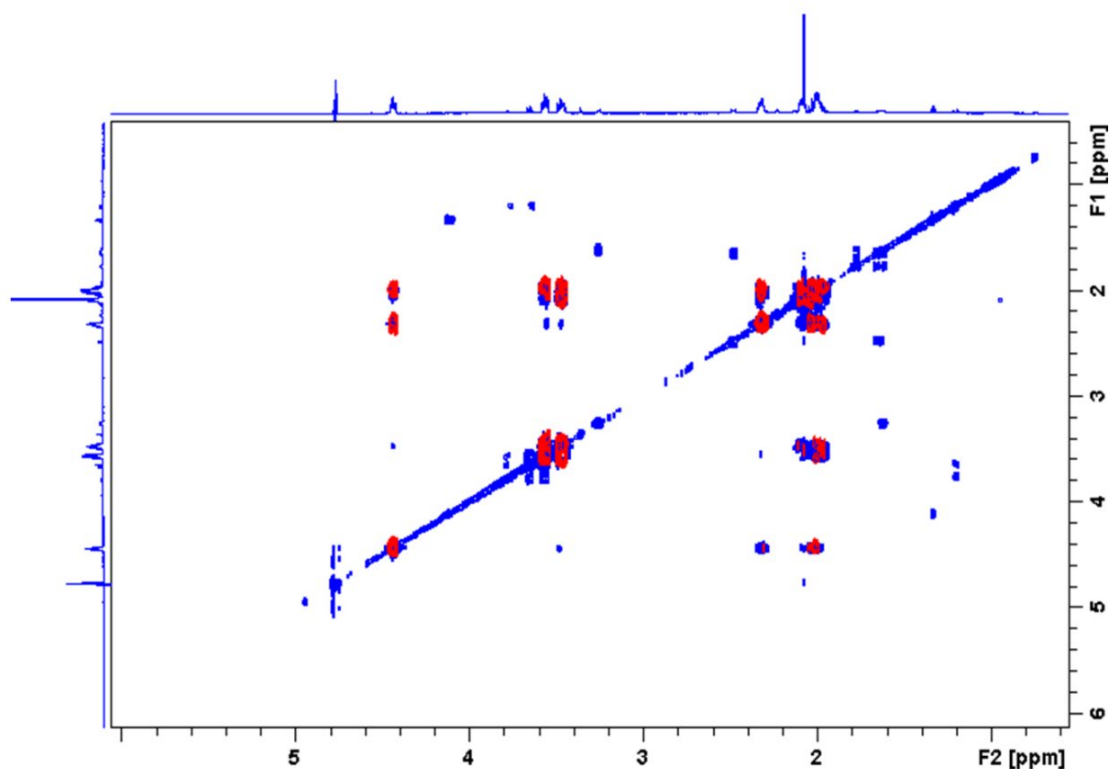


Figure 4. 2D ^1H - ^1H COSY spectra of the second dimension fraction containing cyclo(Pro-Pro) in blue overlaid with the cyclo(Pro-Pro) standard in red.

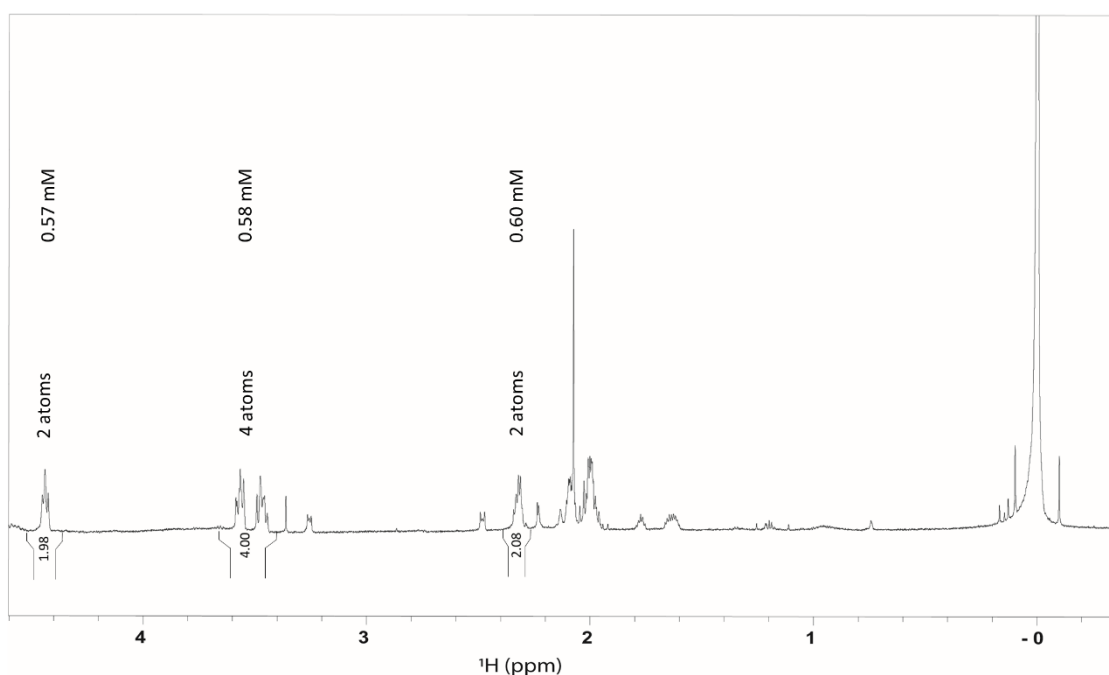


Figure 5. 1D ^1H NMR spectrum of the second dimension fraction containing cyclo(Pro-Pro). Integrated resonances, number of associated protons and calculated concentrations are indicated. The measured mass is $4.7\ \mu\text{g}$ (based on a sample volume of $40\ \mu\text{L}$ and a molar mass of $194.23\ \text{g/mol}$).

Compared with soy sauce (Table 1), the concentrations of the dipeptides in the final fractions (Table 2) increased by a factor of 6, on average. Moreover, this increase in concentration was accompanied by a substantial removal of the background signals. In comparison with soy sauce, the proton purity improved by a factor of three orders of magnitude, on average, in the final fractions. The cleanup efficiency of our platform is also emphasized by the decreased complexity of the NMR spectra of the final fractions in comparison with the NMR spectrum of non-fractionated soy sauce (see supplementary information Figure S5). This enabled the recording of high quality 2D NMR spectra in a MicroCryoProbe (1.7 mm sample tubes), optimized for mass sensitivity in small (30 μ l) volumes. When the 1.7 mm sample tubes were measured in a conventional 5 mm Cryoprobe, the sensitivity was sufficient for recording high quality 1D NMR spectra which allowed for accurate quantification.

Despite the increase in concentration, the dipeptides were not fully recovered. From the maximum concentration factor of 17, 15-88% was actually achieved. This can be explained by the fraction collection procedure in which the collection vial was switched every 0.35 minutes. When a dipeptide was eluting during the switching time, this dipeptide was collected into multiple fractions. Eventually, the fraction containing the highest abundance of the dipeptide was selected for further purification aiming for the highest analyte/impurity ratio. This resulted in the loss of the dipeptide that ended up in a non-selected fraction. The purification of Cyclo(Pro-Gly) did not suffer substantial losses as the concentration factor of 15 was close to the maximum factor. This dipeptide was mostly fractionated into one fraction instead of multiple. Although the dipeptides were not fully recovered, the fractionation still allowed for a higher concentration in the final fraction as compared to the original concentration in the investigated soy sauce. This emphasizes the power of off-line fractionation prior to NMR analysis, because it enables the up-concentration of purified fractions after the chromatographic separation. In on-line LC-NMR applications, compounds would be diluted because of the mobile phase flow and diffusion caused by the considerable amount of post-column tubing.²⁹ This problem can be overcome by the use of an SPE trapping prior to the NMR analysis. However, an LC-SPE-NMR platform often results in a tedious analytical setup and method optimization and lacks efficiency for highly polar compounds.³⁰ Besides, pure and concentrated fractions, which are quantified by NMR, can be used to construct calibration curves using LC-MS and quantify unknown features in the original sample when authentic standards are unavailable.^{31,32} This makes our identification platform suitable for de novo identification with subsequent quantification.

2,5-diketopiperazines and taste

Soy sauce is a popular seasoning that is used particularly in Chinese and Southeast Asian cuisines. It mainly delivers a salty and umami taste, but also adds a characteristic flavour. Amino acids,

sugars, organic acids and minerals are considered as main taste components of soy sauce.^{33,34} Since soy sauce is a fermented product, studies have focused on protein degradation and thus several taste-active amino acids and peptides have already been identified.³⁵ However, there are still many unknown molecules that provide the typical flavour of soy sauce.

We have identified five taste-related cyclic dipeptides, 2,5-diketopiperazines (2,5-DKPs), that derive from proline condensation reactions. 2,5-DKPs can be produced through proteolysis of microorganisms during fermentation or cyclization of linear peptides due to thermal processing.³⁶ Especially proline-based 2,5-DKPs are produced in both heated and fermented foods. Since soy sauce production involves both heating and fermentation, it can be expected that these particular cyclic dipeptides are present in soy sauce. 2,5-DKPs are often found in food and beverages and have shown to be key inducers of bitter taste.³⁷ In addition, hydrophobic peptides that contain proline have been associated with bitter sensations, because proline-peptides favour binding to the bitter taste receptor.³⁸ Although 2,5-DKPs have mostly been linked to bitter taste, these compounds have also been linked to umami taste to a lesser extent.³⁷ Moreover, Zhu *et al.* showed that bitter-tasting linear peptides were able to significantly enhance umami taste in the presence of soy sauce and monosodium glutamate.³⁹ Since cyclic dipeptides generally have a stronger bitter taste in comparison with linear peptides³⁷, it can be hypothesized that the synergistic effect on umami taste is also stronger. This might explain the correlation between the concentration of proline-based 2,5-DKPs and the taste experience of soy sauce. However, more research is needed to prove this effect. In summary, proline-based 2,5-DKPs may affect bitterness and indirectly enhance umami perception in soy sauce.

CONCLUSIONS

The unambiguous identification of unknown compounds remains one of the most challenging aspects of metabolomics. The combination of two highly complementary techniques, however, can drastically increase the confidence of structure annotations. In addition, when the spectral matching of MS data does not result in a library hit (e.g. compound has not been identified before or does not fragment), NMR is crucial for de-novo structure elucidation. Therefore, we have developed a workflow for metabolite identification combining NMR and MS. A comprehensive fractionation method was developed to purify five taste-related unknowns from soy sauce to allow for the NMR analysis of m/z and retention time features from a complex sample. The directed two-dimensional fractionation demonstrated that different second dimension chromatography types were needed in order to isolate the unknown compound from other matrix components. The use of a one-dimensional or a fixed two-dimensional fractionation would have resulted in a less efficient purification. The purified fractions were clean and concentrated enough

to allow for metabolite identification by MS and structure confirmation by MS and NMR. Although the current study presents the identification of unknown compounds in soy sauce, the developed method is not limited to this particular sample type. The evaluation of impurities and unknown compounds between the first and second dimension ensures that this methodology can account for differences in sample matrices. Each first dimension fraction will benefit from the most orthogonal second dimension chromatography in order to remove remaining impurities. If needed, it is also possible to extend the number of second dimension chromatography types to push the versatility of the fractionation even more. In our study, however, the use of only two different second dimension chromatography types was already sufficient to successfully identify five unknown compounds. All taste-related soy features were identified as proline-based 2,5-diketopiperazines (2,5-DKPs). These compounds are generally known to affect bitterness of food products. Moreover, bitter tasting peptides have been shown to enhance umami perception in soy sauce as well. More research should be conducted to prove that this is also the case for proline-based 2,5-diketopiperazines.

ACKNOWLEDGEMENTS

The authors are grateful to receive funding for this research from The Netherlands Organization for Scientific Research (NWO) in the framework of the Technology Area TA-COAST (Fund New Chemical Innovations. project no. 053.21.118).

REFERENCES

1. Beger, R. D. *et al.* Metabolomics enables precision medicine: “A White Paper, Community Perspective”. *Metabolomics* **12**, (2016).
2. Lin, C. Y., Viant, M. R. & Tjeerdema, R. S. Metabolomics: Methodologies and applications in the environmental sciences. *J. Pestic. Sci.* **31**, 245–251 (2006).
3. Kim, S., Kim, J., Yun, E. J. & Kim, K. H. Food metabolomics: From farm to human. *Curr. Opin. Biotechnol.* **37**, 16–23 (2016).
4. Dunn, W. B. *et al.* Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics* **9**, 44–66 (2013).
5. Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **3**, 211–221 (2007).
6. Pan, Z. & Raftery, D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal. Bioanal. Chem.* **387**, 525–527 (2007).
7. Castillo-Munoz, N. *et al.* Flavonol 3- O -Glycosides Series of *Vitis vinifera* Cv. *J. Agric. Food Chem.* **57**, 209–219 (2009).
8. Rathahao-Paris, E., Alves, S., Junot, C. & Tabet, J. C. High resolution mass spectrometry for structural identification of metabolites in metabolomics. *Metabolomics* **12**, 1–15 (2016).
9. Rojas-Chertó, M. *et al.* Elemental composition determination based on MS n. *Bioinformatics* **27**, 2376–2383 (2011).
10. van der Hoof, J. J. J. & Rankin, N. Metabolite Identification in Complex Mixtures Using Nuclear Magnetic Resonance Spectroscopy. in *Modern Magnetic Resonance* 1–32 (Springer International Publishing, 2016). doi:10.1007/978-3-319-28275-6_6-1

11. Boiteau, R. M. *et al.* Structure elucidation of unknown metabolites in metabolomics by combined NMR and MS/MS prediction. *Metabolites* **8**, (2018).
12. Markley, J. L. *et al.* The future of NMR-based metabolomics. *Curr. Opin. Biotechnol.* **43**, 34–40 (2017).
13. Lei, Z., Huhman, D. V. & Sumner, L. W. Mass spectrometry strategies in metabolomics. *J. Biol. Chem.* **286**, 25435–25442 (2011).
14. Brennan, L. NMR-based metabolomics: From sample preparation to applications in nutrition research. *Prog. Nucl. Magn. Reson. Spectrosc.* **83**, 42–49 (2014).
15. Bhatia, A., Sarma, S. J., Lei, Z. & Sumner, L. W. UHPLC-QTOF-MS/MS-SPE-NMR: A Solution to the Metabolomics Grand Challenge of Higher-Throughput, Confident Metabolite Identifications. in *Methods in Molecular Biology* **2037**, 113–133 (Humana Press Inc., 2019).
16. Van Duynhoven, J. P. M. & Jacobs, D. M. Assessment of dietary exposure and effect in humans: The role of NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **96**, 58–72 (2016).
17. Pirok, B. W. J., Stoll, D. R. & Schoenmakers, P. J. Recent Developments in Two-Dimensional Liquid Chromatography: Fundamental Improvements for Practical Applications. *Anal. Chem.* **91**, 240–263 (2019).
18. Wider, G. & Dreier, L. Measuring protein concentrations by NMR spectroscopy. *J. Am. Chem. Soc.* **128**, 2571–2576 (2006).
19. Hoult, D. I. & Richards, R. E. The signal-to-noise ratio of the nuclear magnetic resonance experiment. *J. Magn. Reson.* **24**, 71–85 (1976).
20. Hoult, D. I. The Principle of Reciprocity in Signal Strength Calculations. *Concepts Magn. Reson.* **12**, 173–187 (2000).
21. Van der Klink, J. J. The NMR reciprocity theorem for arbitrary probe geometry. *J. Magn. Reson.* **148**, 147–154 (2001).
22. Willmann, J., Mahlstedt, K., Leibfritz, D., Spraul, M. & Thiele, H. Characterization of sphingomyelins in lipid extracts using a HPLC-MS-offline-NMR method. *Anal. Chem.* **79**, 4188–4191 (2007).
23. Bird, S. S. *et al.* Structural characterization of plasma metabolites detected via LC-electrochemical coulometric array using LC-UV fractionation, MS, and NMR. *Anal. Chem.* **84**, 9889–9898 (2012).
24. Gilar, M., Olivova, P., Daly, A. E. & Gebler, J. C. Orthogonality of separation in two-dimensional liquid chromatography. *Anal. Chem.* **77**, 6426–6434 (2005).
25. Bassanese, D. N. *et al.* Protocols for finding the most orthogonal dimensions for two-dimensional high performance liquid chromatography. *Talanta* **134**, 402–408 (2015).
26. Lynn, K. S. *et al.* Metabolite identification for mass spectrometry-based metabolomics using multiple types of correlated ion information. *Anal. Chem.* **87**, 2143–2151 (2015).
27. Smith, L. L., Herrmann, K. A. & Wysocki, V. H. Investigation of gas phase ion structure for proline-containing b 2 ion. *J. Am. Soc. Mass Spectrom.* **17**, 20–28 (2006).
28. Wüthrich, K. NMR with Proteins and Nucleic Acids. *Europhys. News* **17**, 11–13 (1986).
29. Walker, G. S. & O’Connell, T. N. Comparison of LC-NMR and conventional NMR for structure elucidation in drug metabolism studies. *Expert Opin. Drug Metab. Toxicol.* **4**, 1295–1305 (2008).
30. Schlotterbeck, G. & Ceccarelli, S. M. LC-SPE-NMR-MS: A total analysis system for bioanalysis. *Bioanalysis* **1**, 549–559 (2009).
31. Van Duynhoven, J. *et al.* Rapid and sustained systemic circulation of conjugated gut microbial catabolites after single-dose black tea extract consumption. *J. Proteome Res.* **13**, 2668–2678 (2014).
32. Espina, R. *et al.* Nuclear magnetic resonance spectroscopy as a quantitative tool to determine the concentrations of biologically produced metabolites: Implications in metabolites in safety testing. *Chem. Res. Toxicol.* **22**, 299–310 (2009).
33. Shiga, K. *et al.* Metabolic profiling approach to explore compounds related to the umami intensity of soy sauce. *J. Agric. Food Chem.* **62**, 7317–7322 (2014).
34. Kaneko, S., Kumazawa, K. & Nishimura, O. Isolation and identification of the umami

- enhancing compounds in Japanese soy sauce. *Biosci. Biotechnol. Biochem.* **75**, 1275–1282 (2011).
35. Apriyantono, A., Setyaningsih, D., Hariyadi, P. & Nuraida, L. Sensory and peptides characteristics of soy sauce fractions obtained by ultrafiltration. *Adv. Exp. Med. Biol.* **542**, 213–226 (2004).
 36. Otsuka, Y. *et al.* Investigation of the formation mechanism of proline-containing cyclic dipeptide from the linear peptide. *Biosci. Biotechnol. Biochem.* **83**, 2355–2363 (2019).
 37. Borthwick, A. D. & Da Costa, N. C. 2,5-diketopiperazines in food and beverages: Taste and bioactivity. *Crit. Rev. Food Sci. Nutr.* **57**, 718–742 (2017).
 38. Zhao, C. J., Schieber, A. & Gänzle, M. G. Formation of taste-active amino acids, amino acid derivatives and peptides in food fermentations – A review. *Food Res. Int.* **89**, 39–47 (2016).
 39. Zhu, X., Chen, Y. & Subramanian, R. Comparison of information-dependent acquisition, SWATH, and MS All techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *Anal. Chem.* **86**, 1202–1209 (2014).

SUPPLEMENTARY INFORMATION

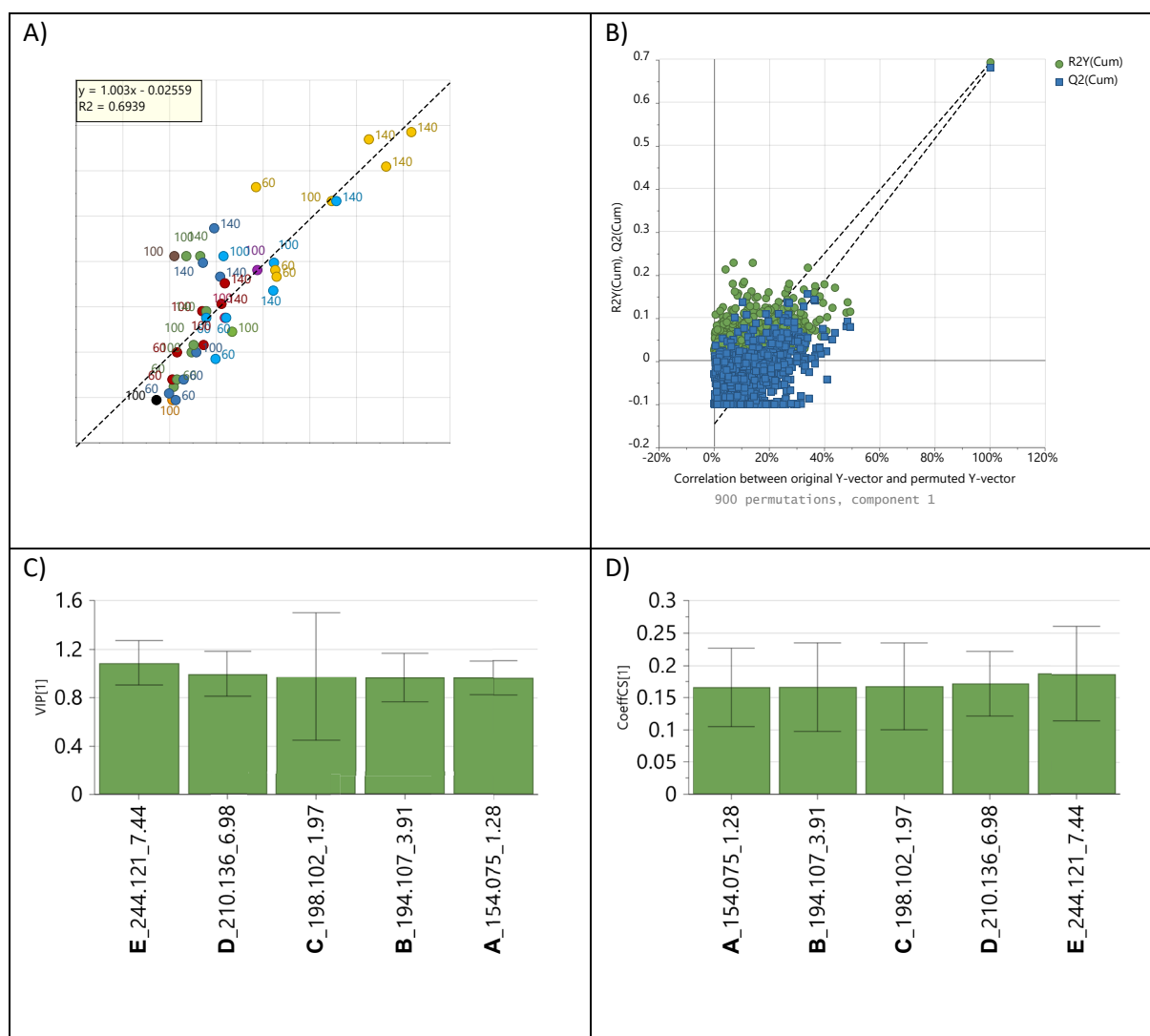
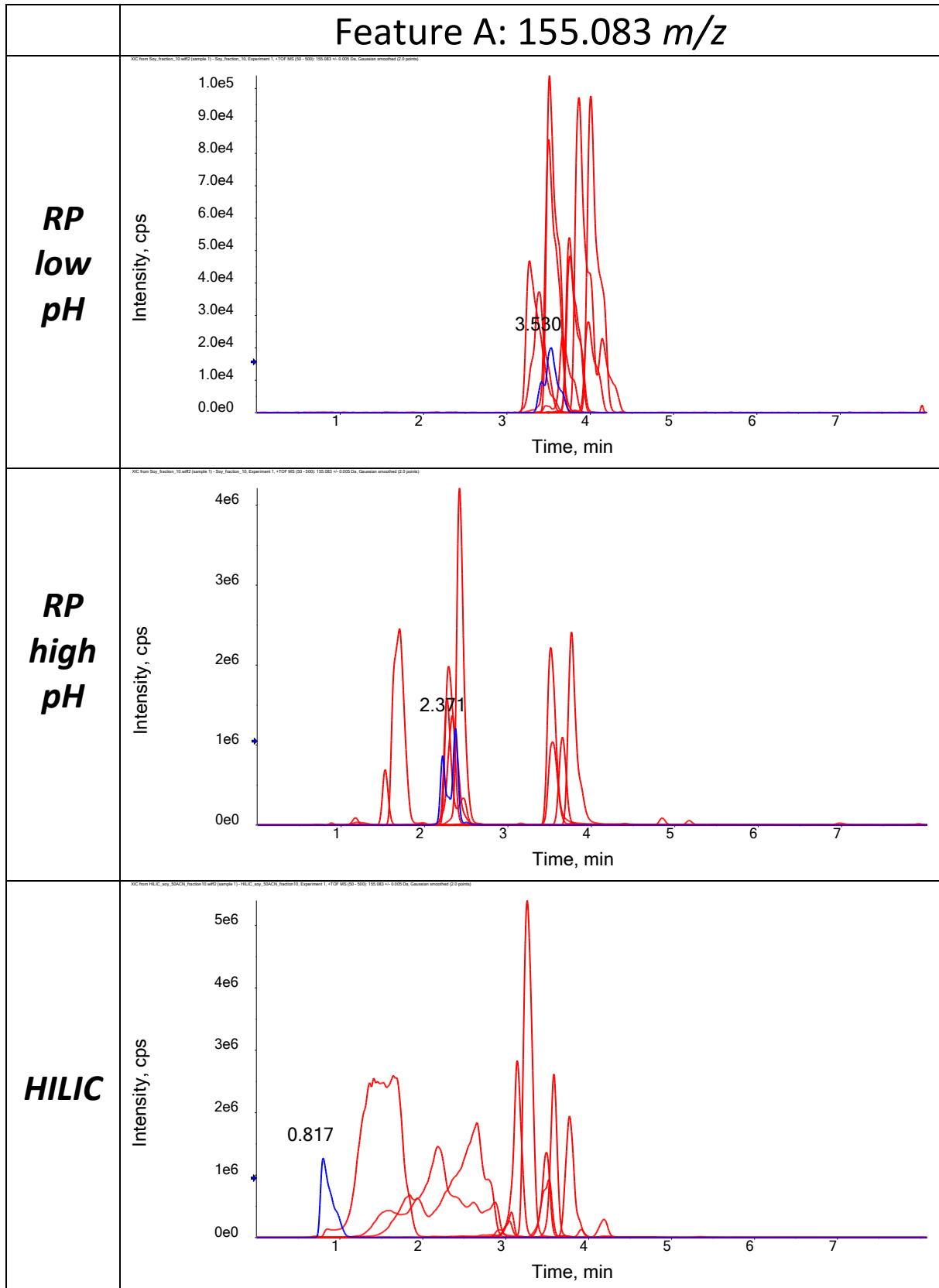


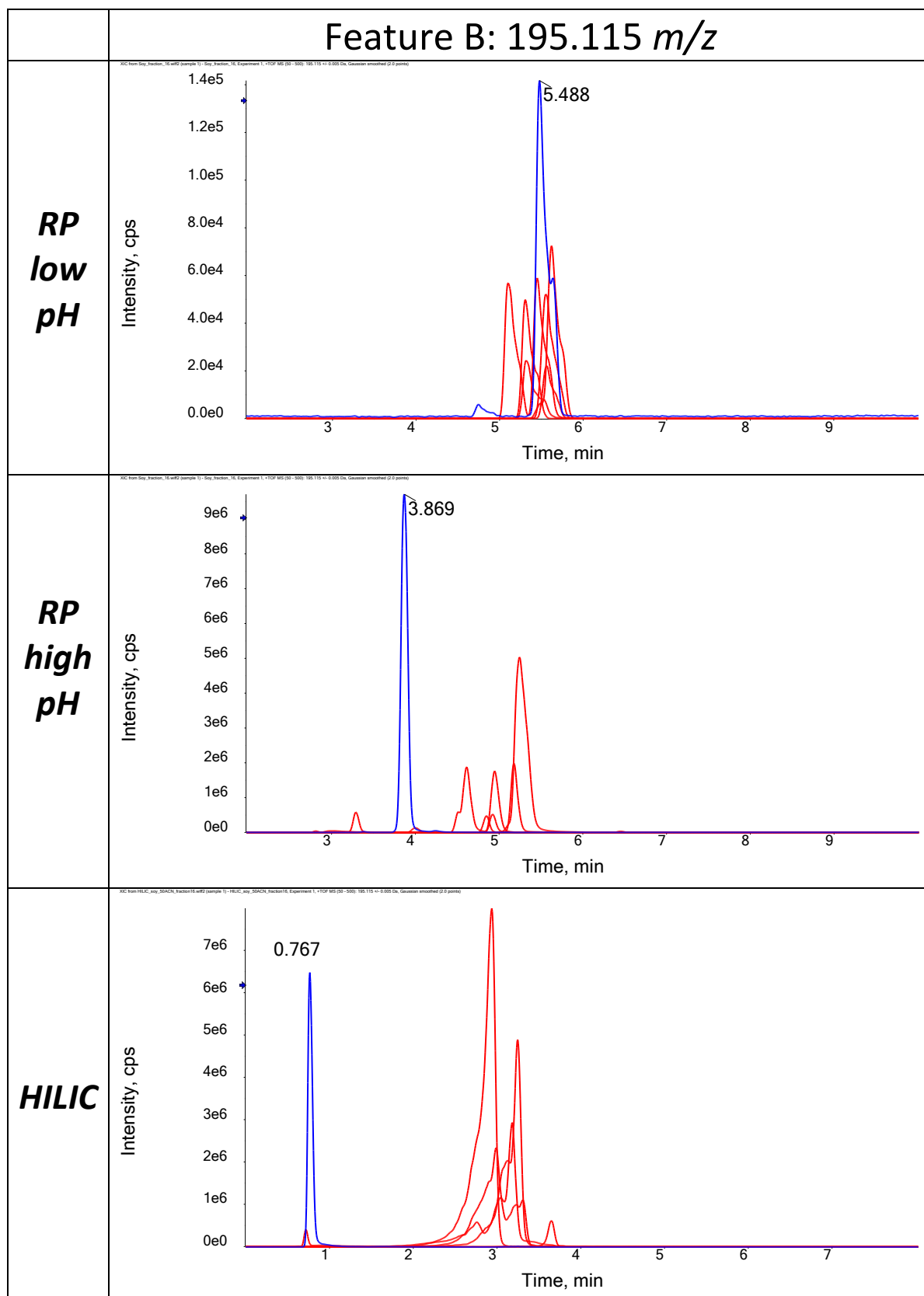
Figure S1. PLS-model predicting the average sensory scores of fermented soybean flavour from five unknown features measured amongst 617 features by LC-MS-based global profiling (method can be found in Table S1). The sensory scores were evaluated by a trained sensory panel. The model was calculated on one principal component and with 7-fold cross-validation using SIMCA (version 15.02. Sartorius Stedim Biotech, Umäa, Sweden). Panel A) shows the actual versus predicted sensory scores of fermented soybean flavour. The sensory scores were evaluated from bouillons to which 10 different commercially available soy sauces were added at three different relative concentrations. The color indicates the soy sauce; black is the reference sample to which no soy sauce was added. The labels indicate the soy sauce concentrations. Panel B) shows the predictability of the PLS-model ($Q^2=0.682$) when compared to 900 permutation tests ($Q^2 < 0.3$), demonstrating that the PLS-model is statistically significant. Panel C) is the variable of importance plot indicating the ranked contribution of the five unknown features of the model. Panel D) is the coefficient plot, showing that the five unknown features positively contribute to the prediction. The soy sauce S04 was used in this study for metabolite identification.

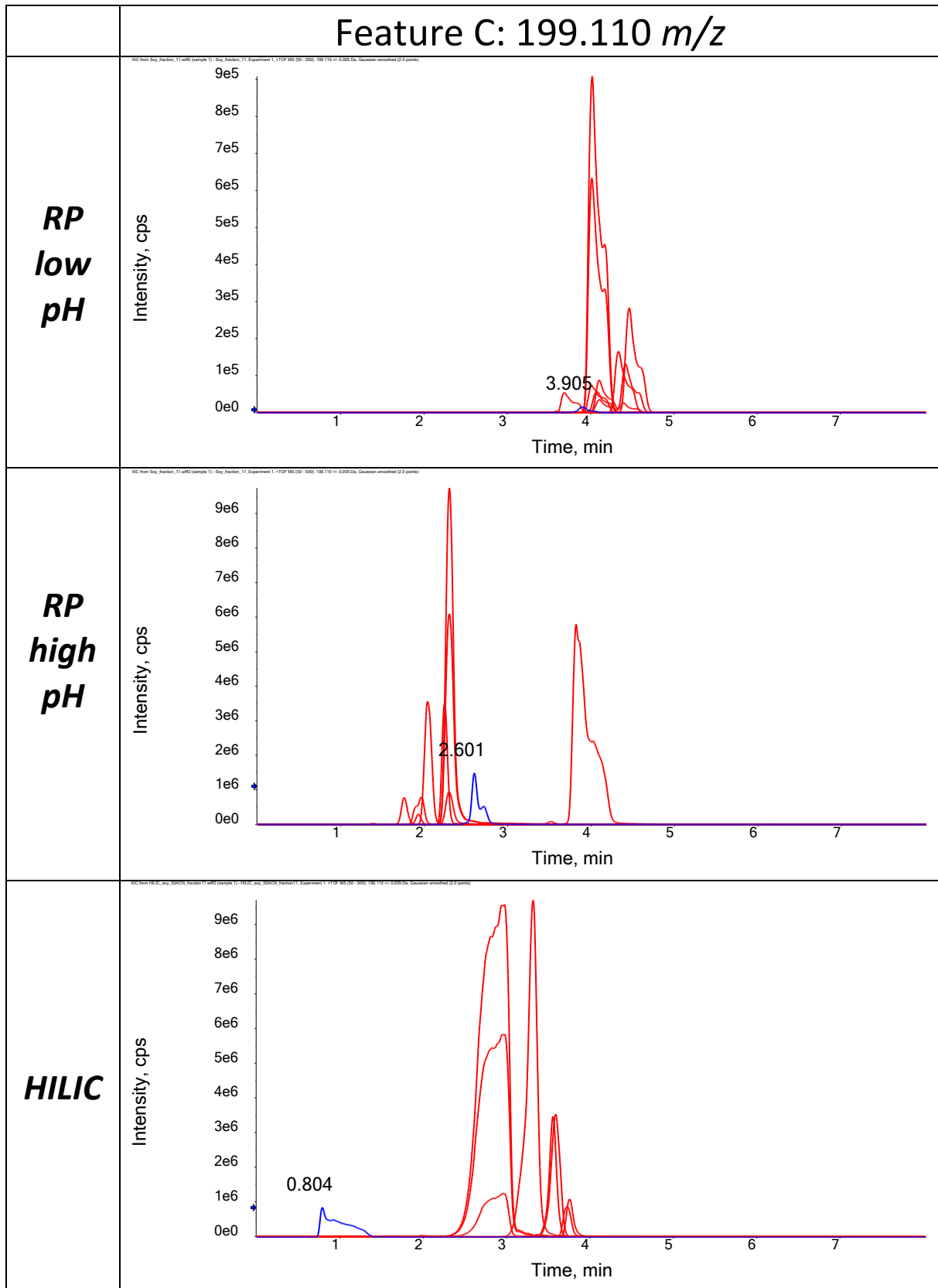
Table S1. The LC-MS-based global profiling gradient performed on an ACCQ-TAG ULTRA C18 column (1.7 μm , 2.1 x 100 mm). Mobile phase A and B consisted of 2% formic acid in water and acetonitrile, respectively, the injection volume was 5 μL and the MS was set at positive polarity.

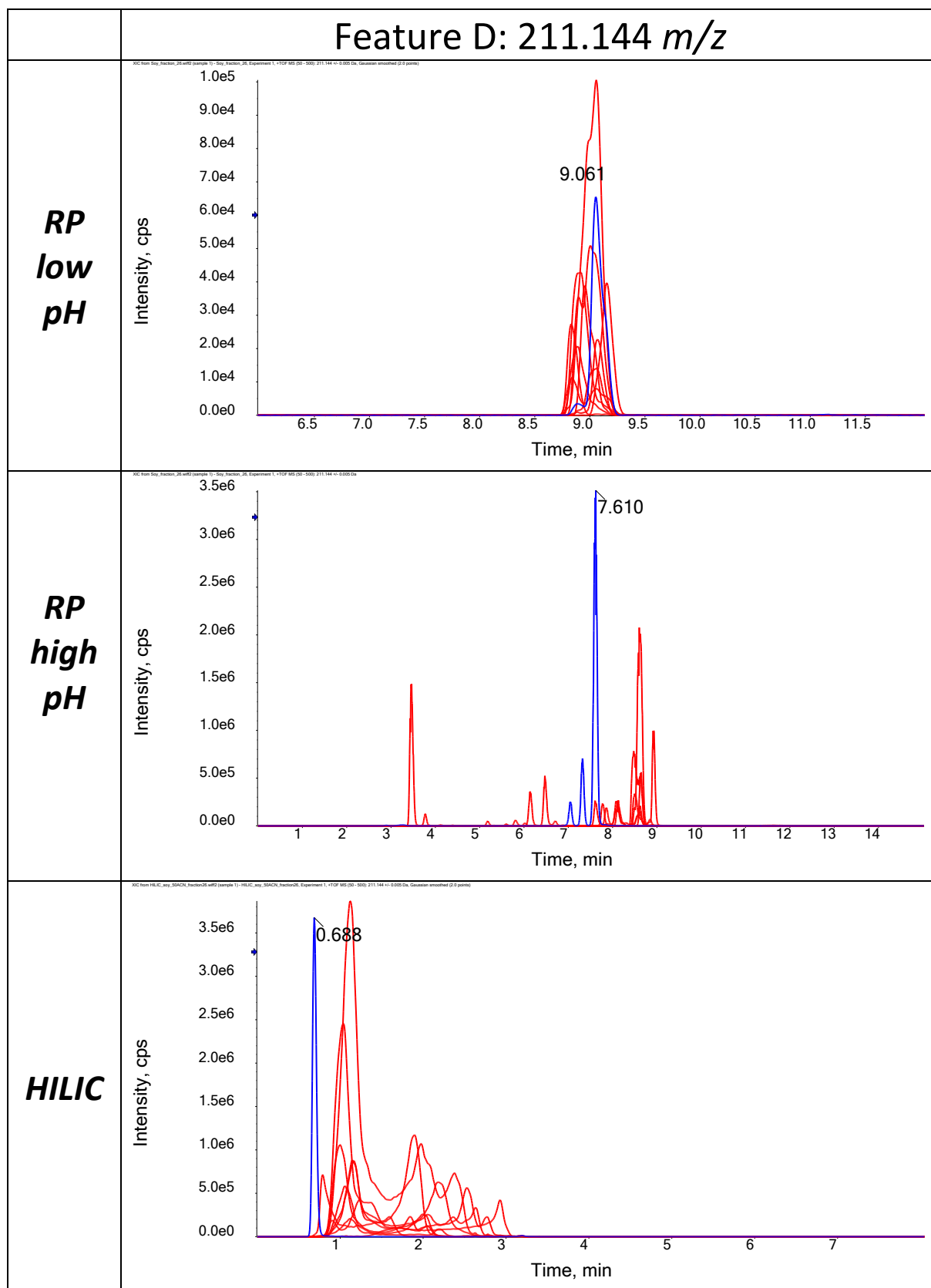
Time (min)	Mobile phase B (%)	Curve	Flow rate (mL/min)
0.0	0.2	-	0.7
7.0	20.0	7	0.7
7.5	60.0	6	0.7
7.6	95.0	6	0.7
8.5	95.0	6	0.7
8.6	0.2	6	0.7
10.5	0.2	6	0.7

Figure S2. Orthogonality of RP high pH and HILIC for features A-E. The extracted ion chromatograms of the features and the contaminants are depicted in blue and red, respectively. All separations are performed on the fractionation chromatography methods.









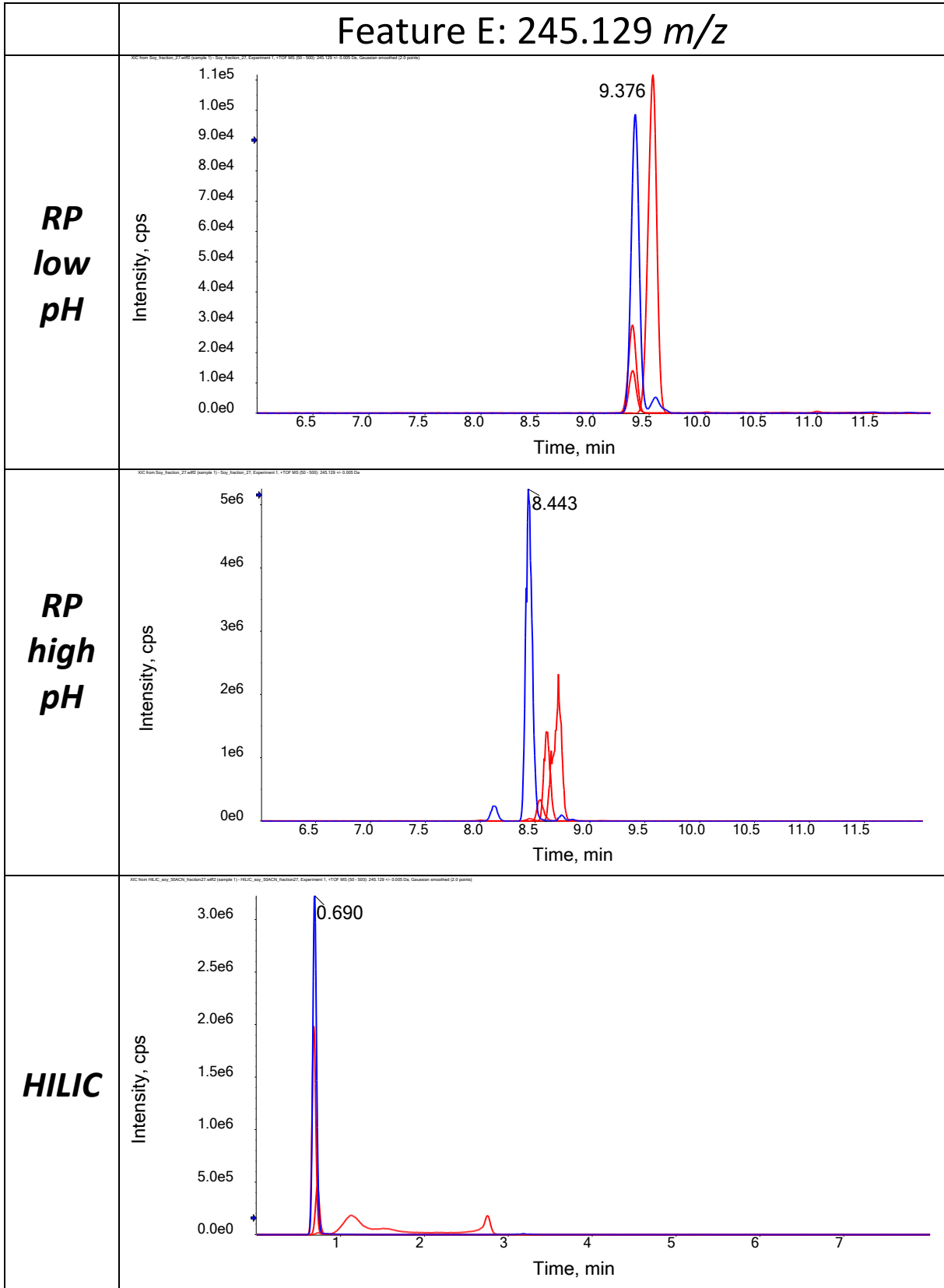
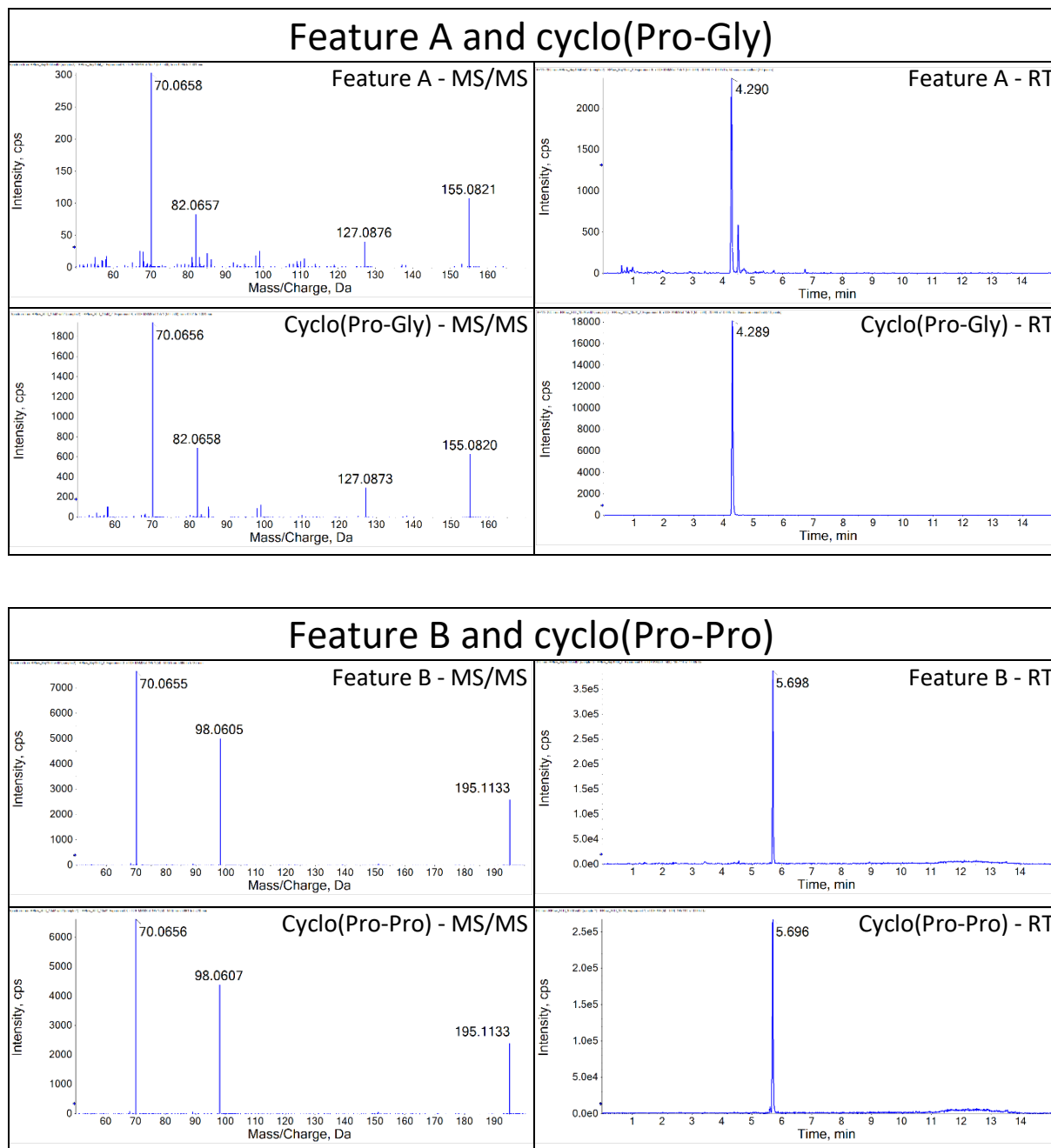
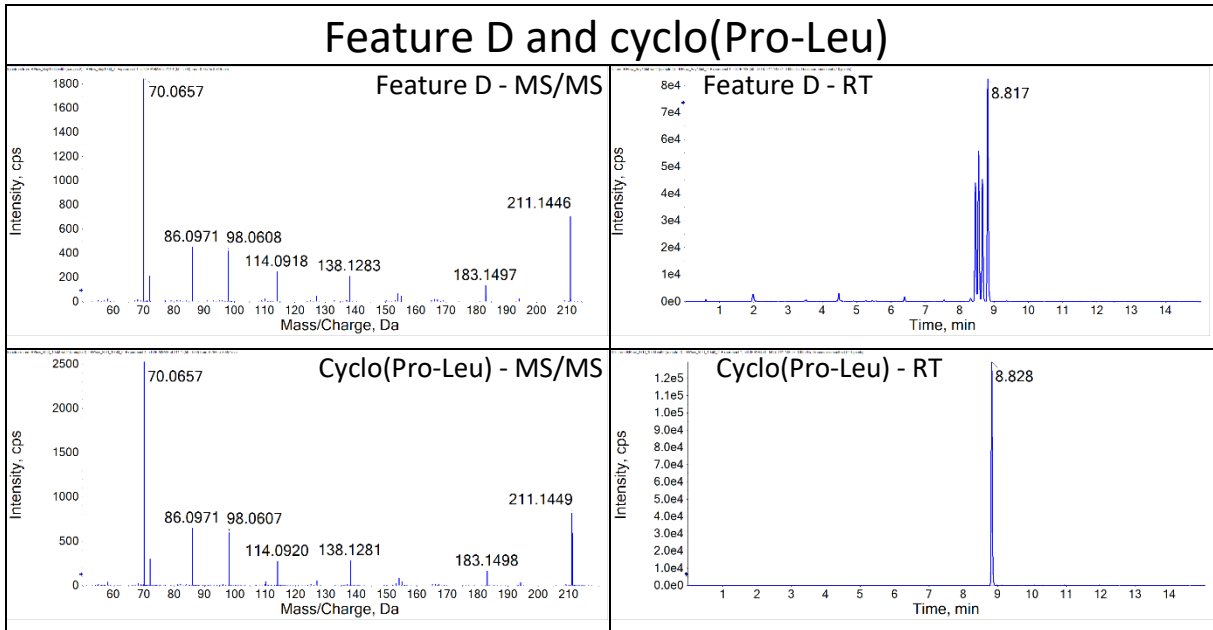
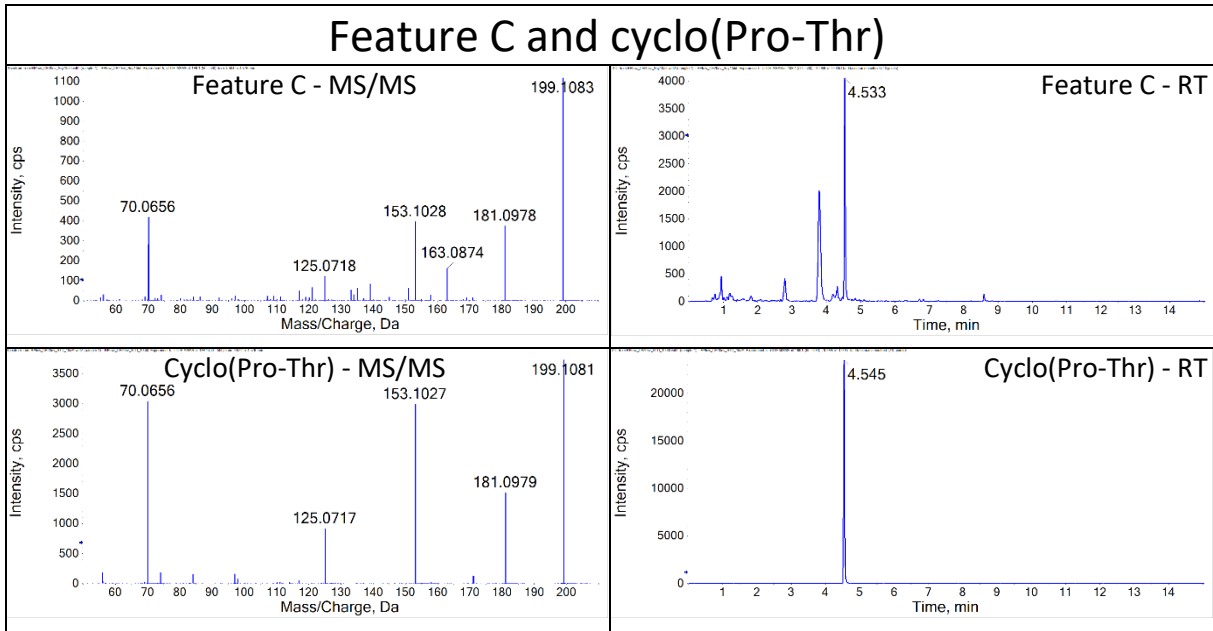


Figure S3. The MS/MS spectra and retention time (RT) of features A-E and the suggested cyclic dipeptide standard. The analytical reversed phase low pH method was used for the analysis.





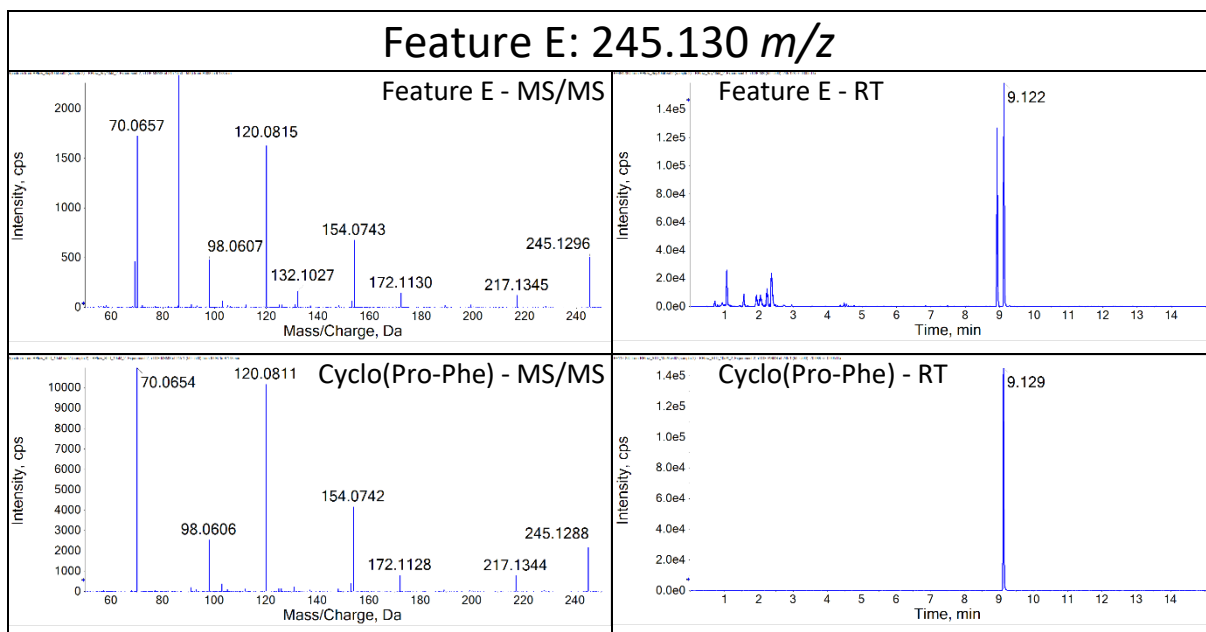
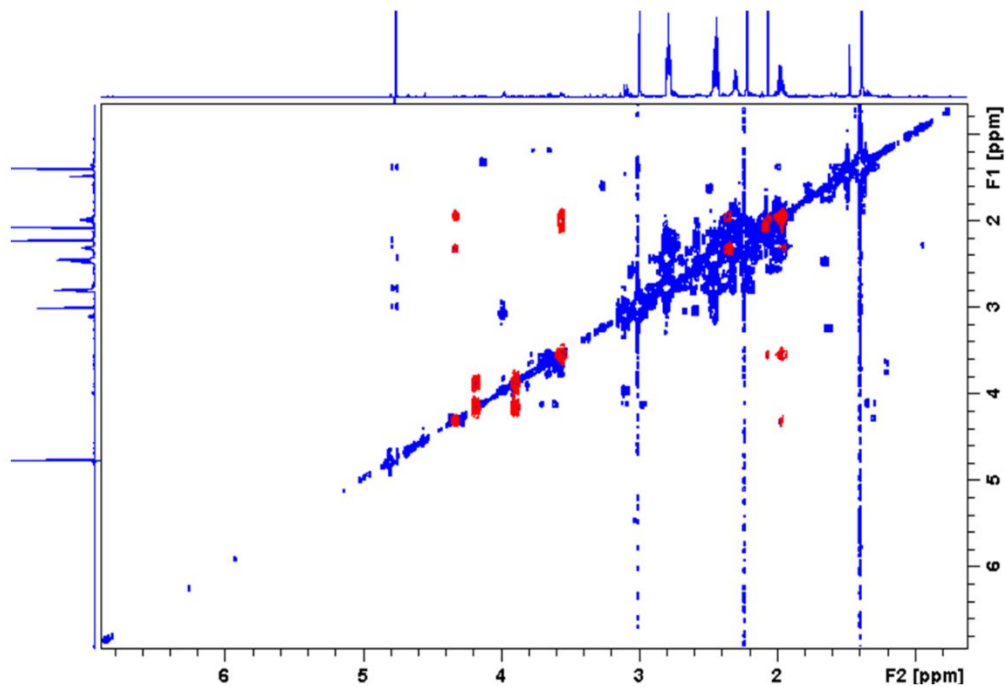
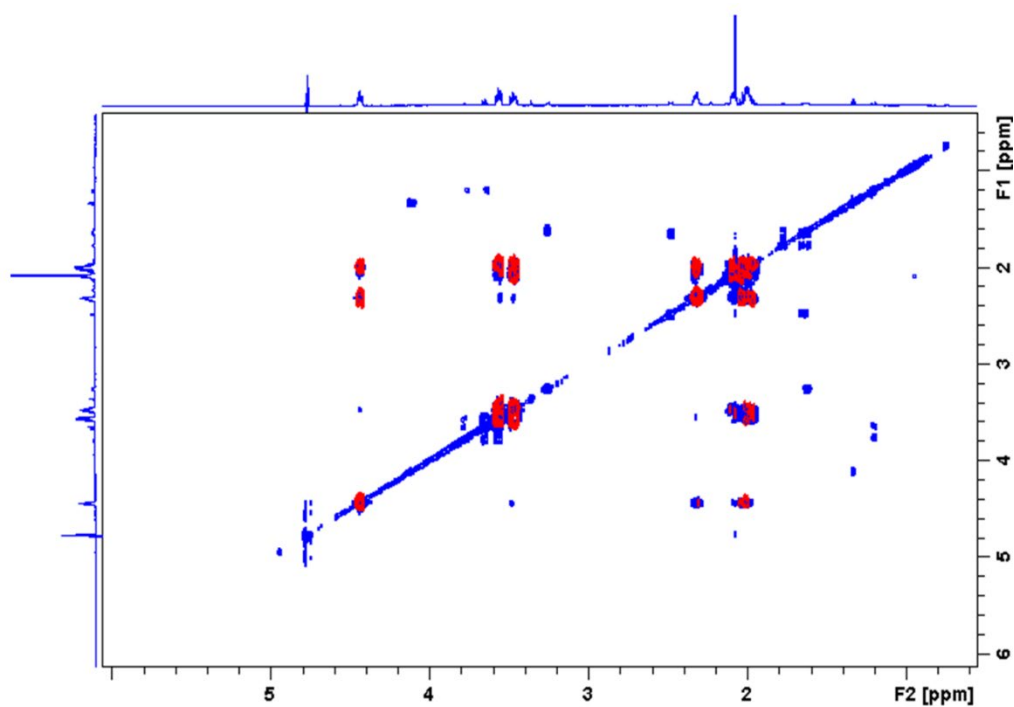


Figure S4. 2D ^1H - ^1H COSY spectra of the second dimension fractions in blue overlaid with the dipeptide standards in red.

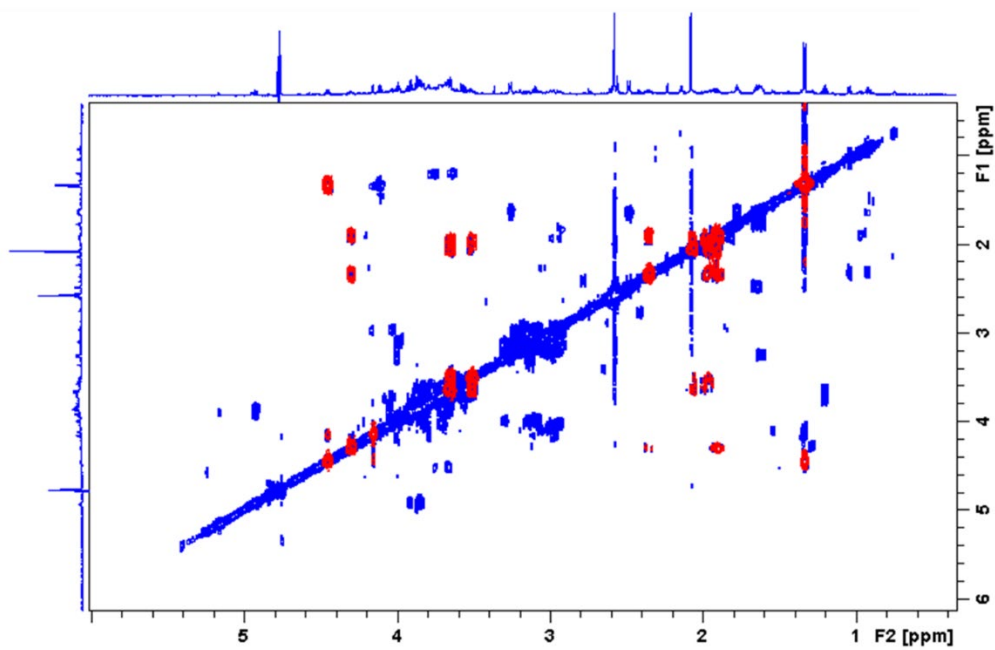
Feature A (blue) and Cyclo(Pro-Gly) (Red)



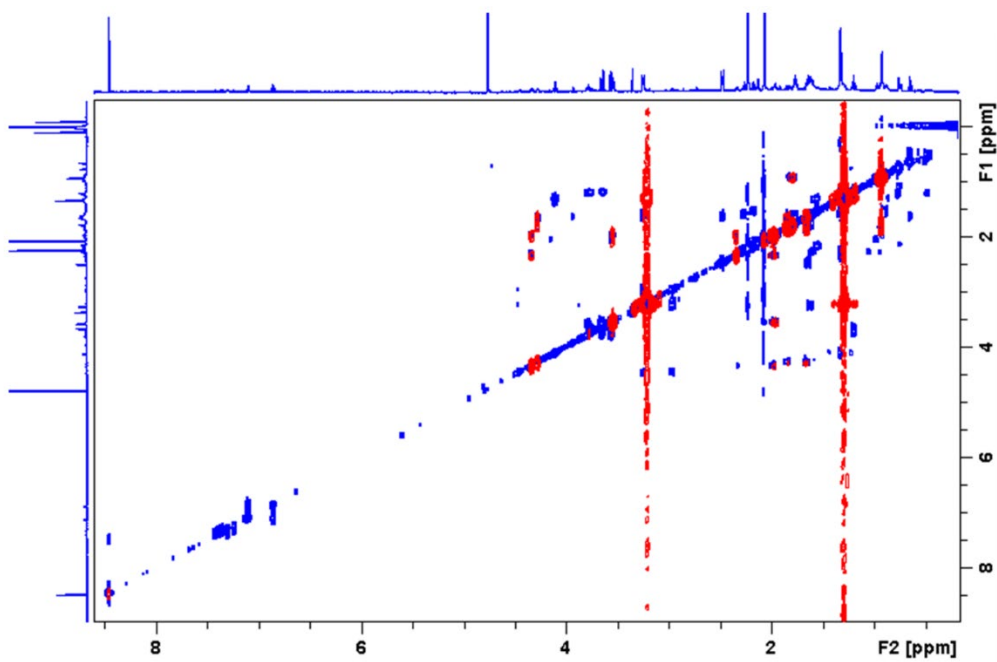
Feature B (blue) and Cyclo(Pro-Pro) (Red)



Feature C and Cylo(Pro-Thr)



Feature D and Cyclo(Pro-Leu)



Feature E and Cyclo(Pro-Phe)

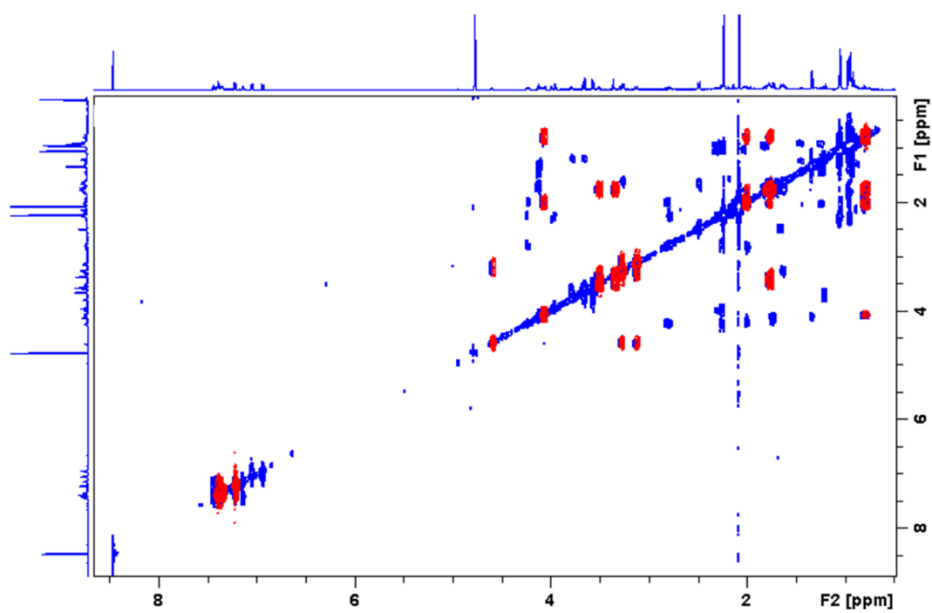
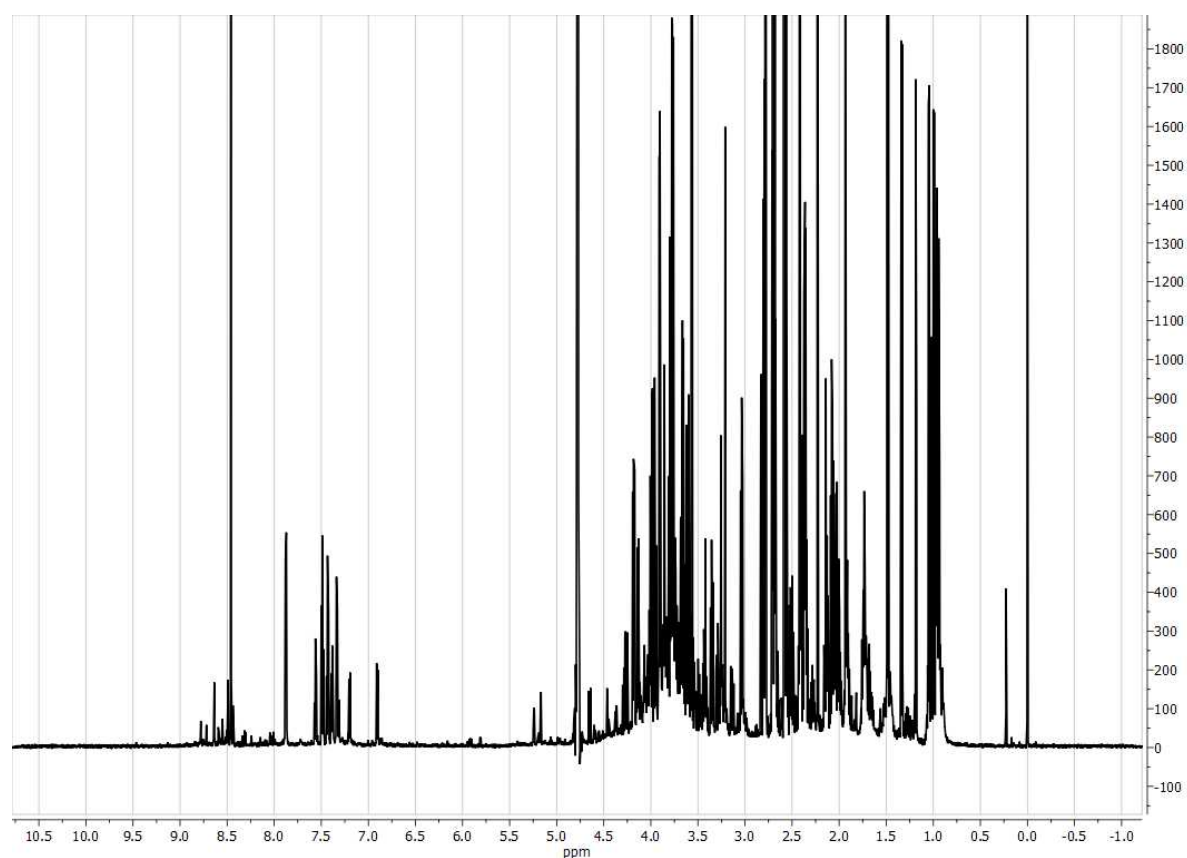


Table S2. Overview of $^1\text{H}/^{13}\text{C}$ NMR assignments of dipeptides standards.

Cyclo		Pro-Gly		Pro-Pro		Pro-Thr		Pro-Leu		Pro-Phe	
Nuclei analysis		^{13}C	^1H	^{13}C	^1H	^{13}C	^1H	^{13}C	^1H	^{13}C	^1H
Other Amino acid	C α	48.3	3.89 4.18			63	4.15	56.2	4.28	59.2	4.58
	C1					68.8	4.45	40.6	1.65	40.7	3.12 3.28
	C2										
	C3							26.8	1.81		
	C4										
	Me ₁					21.3	1.33	23.5	0.92		
	Me ₂							25.1	0.94		
	Φ 1									132.8	7.21
	Φ 2									131.5	7.38
Φ 3									130.8	7.34	
Proline	(C=O) C1	61.5	4.33	63.7	4.43	61.6	4.3	61.9	4.33	62	4.06
	C2	30.08	1.96 2.34	30	2 2.31	31	1.93 2.35	30.5	1.96 2.34	30.06	0.79 2.01
	C3	24.7	1.96 2.08	25.6	2 2.08	24.6	1.96 2.06	24.6	1.96 2.08	23.7	1.76
	NH-C4	48.2	3.56	48.2	3.56 3.46	48	3.65 3.5	48.1	3.56	47.8	3.34 3.5

*Figure S5. ^1H NMR spectrum of non-fractionated soy sauce*

Chapter 6

Conclusions and perspectives

CONCLUSIONS

The field of metabolomics is increasingly being implemented in various disciplines. This popularity results into an increased demand for metabolic profiles, which pushes the qualitative and quantitative performance of current analytical techniques to its limits. Mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy have shown to be very useful in acquiring metabolic profiles. These techniques have been used to answer the two most important questions in metabolomics: what is the identity and quantity of a metabolite in a mixture. Although technological advances have led to an unsurpassed performance of MS and NMR in the field of metabolomics, the potential of these techniques has not been fully exploited due to several analytical challenges. The presence of severe matrix effects hampers the fast quantification of metabolites in complex mixtures by mass spectrometry. In addition, the analysis complex mixtures will also result in complex NMR and MS/MS spectra which complicates the unambiguous identification of unknown metabolites. The aim of this thesis was to tackle these challenges by the development and application of innovative fractionation approaches and state-of-the-art MS and NMR analyses. We have used fractionation to facilitate the quantification and identification of metabolites in complex mixtures. Fractionation has shown to decrease the complexity of biological samples in a high-throughput fashion, which allowed for the fast quantification of metabolites using MS. Fractionation has also shown to be useful for in-depth purification and concentration of unknown features prior to NMR analysis. In this chapter, we will provide an overall conclusion with regards to the conducted research on fractionation. We will discuss the technical modifications needed to further increase the throughput of our fractionation platforms. We will also discuss a new data independent acquisition (DIA) technique and an advanced NMR coil, which could increase the selectivity and sensitivity of our fractionation platforms, respectively. We will finalize this chapter with a general conclusion on the power of metabolomics platforms in life sciences.

Throughput in metabolomics

Metabolomics is becoming increasingly important in several disciplines, ranging from the food industry to drug research and healthcare. Many metabolites have been identified that might have a diagnostic or prognostic value. Biomarker discovery in the field of metabolomics is mostly conducted by correlating metabolite levels with a certain physiological state, e.g. therapy response or disease progression. However, the underlying causal relationship is still often unknown and only a limited number of metabolites have been validated as clinical biomarkers. In order to achieve more mechanistic insights into metabolite biomarkers and validate more potential biomarkers, large-scale metabolomics studies are needed.^{1,2} When dynamic biomarkers are applied in the clinic, multiple time points have to be included in order to properly monitor disease

progression or drug response. When these dynamic biomarkers are used for common diseases, it drastically increases the demand for metabolic profiles. Ideally, these studies are analyzed by high-throughput platforms. This accelerates the biomarker discovery and application process and will most likely also reduce its costs because less instrument time and lab personnel is needed.

MS offers great possibilities for fast metabolic profiling. MS allows for a wide metabolic coverage and quantitative MS data can be obtained in principle in less than a second.³ However, the presence of matrix effect is currently limiting the speed of MS analyses. Severe matrix effects are especially occurring during the analysis of biological samples because of the complexity of this sample type. To minimize matrix effect, MS is often coupled to a separation technique, e.g. liquid chromatography, in order to decrease the complexity of a sample before the ionization source.⁴ However, performing this separation step is at the cost of analysis time.

Separation-free platforms, e.g. flow injection analysis (FIA), have been developed in order to improve the throughput of MS analyses. In these platforms, matrix effect is often tackled by a dilute-and-shoot approach or by using fast sample preparation methodologies. However, the cleanup efficiency of these methods is limited. The dilution of samples also dilutes the analytes while highly abundant matrix components can still cause suppression. In addition, commonly used sample preparation techniques, i.e. solid-phase extraction (SPE) and liquid-liquid extraction (LLE), usually result in two fractions. SPE allows for a retained and not retained fraction, whereas LLE allows for a polar and apolar fraction. These fractions are obtained either off-line in parallel or on-line and, when combined with FIA-MS, introduced into the MS without further separation. This allows for little cleanup efficiency and a high chance of severe ion suppression.

In this thesis, we have improved cleanup efficiency in a high-throughput fashion by the development of more extended fractionation approaches. We have employed high performance columns, which allow for within-fraction separation, and coupled multiple columns in series increasing the number of fractions. This makes the cleanup efficiency of our fractionation approaches higher than conventional sample preparation techniques, which only result in two fractions and do not allow for within-fraction separation. Our fractionation approaches are faster than conventional chromatography because time-consuming gradients are replaced by fast solvent switches. We have used shorter columns than conventionally applied to decrease the time needed for column flushing and equilibration. Therefore, fractionation provides an ideal balance between cleanup efficiency and throughput as is demonstrated in **chapter 3** and **4**.

Quantification in metabolomics

The quantitative analysis of metabolites in biological samples is essential in biomarker discovery and clinical decision-making.⁵ Quantitative data is needed in order to find potential biomarkers

that categorize patients and/or healthy volunteers into different clinical groups, e.g. good drug responder versus bad drug responder or healthy versus diseased. Eventually, a validated biomarker has to be quantified in patients and aid clinical decision-making. The accuracy of quantitation is of high importance to set clear boundaries between different clinical groups and to assign new patients to these groups. Inaccurate quantification can lead to the mislabeling of patients and all the associated consequences. Therefore, good quantitation is essential for a proper biomarker discovery and application pipeline. Quantitation can be impaired by the presence of matrix effect when samples consist of varying matrix composition, e.g. urine and blood samples taken at different times of the day. Variability in matrix composition will vary ion suppression/enhancement and, therefore, decrease the precision and repeatability of the analysis. Therefore, matrix effect should be properly controlled and ideally not present at all.

Fractionation can be used to obtain quantitative data in a high-throughput fashion. In **chapter 2**, we have accomplished this by employing fast reversed phase chromatography. We shifted the aim of general liquid chromatography from separating analytes to trapping known ion suppressors. Literature demonstrated that salts and phospholipids are notorious ion suppressors in blood samples.^{6,7} We evaluated the effect of these ion suppressors in a flow injection analysis (FIA) coupled to MS. In this study, phospholipids suppressed the MS signal to a greater extent than salts and therefore, we developed a fast LC-MS analysis that enabled the removal of these matrix components from the elution region of the analytes.

The metabolomics application of **chapter 2** realized the quantification of gut metabolites that are involved in the trimethylamine-N-oxide (TMAO) metabolism. Since these gut metabolites have been linked to an increased risk of cardiovascular disease (CVD), there is an increased interest to get quantitative data on these potential biomarkers.⁸ Since these gut metabolites are highly polar and phospholipids are apolar, the phospholipids are easily trapped by a conventional reversed phase column, leaving the gut metabolites in a cleaned flow-through. We could drastically decrease the LC gradient time because the analytes did not experience any retention on the column and the (phospho)lipids were eluted from the column by switching the mobile phase almost directly to 100% organic. The eluting lipids were directed to waste, resulting in less contamination of the MS. The fast LC-MS method successfully removed all ion suppression caused by the phospholipids in an analysis time of only three minutes.

The platform in **chapter 2** allowed for the quantification of five metabolites. The aim of **chapter 3** was to extend the fractionation approach in order to quantify 50 chemically diverse metabolite biomarkers. Although **chapter 2** demonstrated that phospholipids were the major source of ion suppression, salts also caused a considerable amount of ion suppression during the analysis of a blood plasma sample. The matrix effect profile of endogenous levels of phospholipids and salts

together already quite closely resembled the matrix effect experienced in analysis of a non-fractionated plasma sample. Therefore, it is hypothesized that the majority of signal suppression is caused by these two matrix components. To overcome the ion suppression of both phospholipids and salts, we have extended the fractionation approach of **chapter 2** by the addition of two extra columns in **chapter 3**. The additional columns included a cation and anion exchange column in order to trap positive and negative salts, respectively. The combination of reversed phase and ion exchange resulted in a fractionation that was based on hydrophobic interactions and charge. This allowed for a very comprehensive fractionation of the metabolome and, therefore, we were able to quantify fifty known prognostic and/or diagnostic biomarker metabolites from very diverse chemical classes, i.e. amino acids, amines, acylcarnitine, sugars, purines, organic acids and fatty acids. To target a very broad chemical range of metabolites, the approach in **chapter 3** used three serially coupled columns to allocate known ion suppressors, such as phospholipids and positive and negative ions from salts, over different fractions, minimizing their adverse effects during electrospray ionization.

The throughput of the comprehensive fractionation methodology in **chapter 3** was maintained by the use of fast solvent switches and short columns (≤ 1 cm). This allowed for an analysis time of only three minutes per MS polarity. Moreover, the use of high performance (particle size ≤ 5 μm) SPE columns allowed for a within-fraction separation. This resulted in the separation of phospholipids from acylcarnitines and fatty acids within the same fraction. The beneficial effect of the fractionation platform was demonstrated by a comparison with FIA. The fractionation decreased the ion suppression from 89% to 25% and allowed for a sensitivity that was sufficient to analyze endogenous concentrations of a wide range of metabolites in plasma. Although the fractionation method only took 3 minutes per MS polarity, it demonstrated a similar sensitivity in comparison with conventional LC-MS. LC-MS has a considerably lower throughput because of the increased gradient time (3-22 minutes per sample).^{9,10,11,12,13,14} Therefore, we have shown that by efficiently fractionating metabolites and known ion suppressors, it is possible to achieve a similar performance in comparison with LC-MS in a high-throughput fashion.

A challenge in a fractionation-based separation is the lack of separation of isomers. Isomers have the same m/z value, which means they cannot be distinguished on an MS^1 level. However, structural isomers are structurally different which means that the fragmentation of these isomers could lead to the production of unique product ions. When unique product ions are found, it is possible to distinguish structural isomers. In **chapter 5**, we have evaluated different types of DIA techniques in order to selectively quantify unique product ions of structural isomers. We have shown that variable sequential window acquisition of all theoretical mass spectra (SWATH) in

combination with hydrophilic interaction chromatography (HILIC) allowed for the selective quantification of five structural isomer pairs.

The use of fixed SWATH windows resulted in several inaccurate quantifications, which demonstrated that the customization of SWATH windows is of utmost importance in the prevention of product ion overlap. In MS^{ALL}, in which the Q1 filter covered the whole mass range of interest, peak integration was hampered by fluctuating baselines and the linearity and repeatability were drastically impaired in comparison with SWATH. The combination of DIA and fractionation resulted, in general, in low quality MS/MS scans. Even the use of variable SWATH windows could not prevent product ion overlap. This was most likely due to the relatively small peak widths that were a result of the high-throughput character of the separation. Smaller peaks allow for less fragmentation time, which means that the size of the SWATH windows has to increase. This results in the fragmentation of more precursor ions per SWATH window, which leads to more product ion overlap. The MS/MS performance is even further pushed by the presence of severe coelution which is often a problem in high-throughput platforms. Nevertheless, our results demonstrate that a conventional separation in combination with variable SWATH results in the accurate quantification of structural isomers.

Identification in metabolomics

In many metabolomics applications, the identity of the analytes of interest is not known in advance. An untargeted metabolic screening is used in these applications to find new interesting metabolic features that could potentially be used as valuable biomarkers. In order to translate an unknown metabolic feature to a structurally annotated metabolite, an identification step is necessary. However, the unambiguous identification of metabolites is considered as one of the most challenging aspects of present-day metabolomics. Methodologies used for metabolite identification are often sensitive to errors or very labour-intensive. MS and NMR are often combined to elucidate the structure of unknown metabolic features, because of their unique advantages and their complementary nature.

MS-based identification strategies are performed on the basis of high-quality MS/MS and MSⁿ scans. MS/MS scans are needed in order to obtain structural information about certain m/z and retention time features. The MS/MS scan can be deciphered by the user or searched in MS/MS libraries and eventually lead to a potential match. However, the acquisition of these MS/MS scans for unknown features remains challenging. In data dependent acquisition, the coverage of MS/MS scans is limited because precursor ions are selected based on, for example, intensity and an inclusion list. DIA provides a more comprehensive coverage since all precursor ions in the selected mass range are fragmented. However, the mass range of the first quadrupole (Q1) often has to be

increased in order to achieve this coverage. This leads to the formation of complex MS/MS spectra, which decreases the power of these scans substantially.

In **chapter 4**, we have used SWATH in order to obtain selective DIA fragmentation data. We have shown that by using a HILIC separation in combination with SWATH, the top hit from a NIST 2017 library search matched the known identity of 20 test features. The use of MS^{ALL} scans resulted in MS/MS scans that were too complex for metabolite identification. The combination of the fractionation platform of **chapter 3** and DIA also resulted in MS/MS spectra that were unsuitable for identification purposes. The data analysis software was unable to recognize specific fragmentation patterns from the spectral library in the crowded DIA MS/MS data. This shows that an extensive separation and selective fragmentation are essential in the acquisition of untargeted MS/MS spectra.

The certainty of metabolite identifications drastically improves when structural annotations are confirmed by two complementary techniques.¹⁵ In addition, when spectral matching does not result in a library hit (e.g. compound has not been identified before or does not fragment), NMR is crucial for de-novo structure elucidation. Therefore, we have developed an identification platform for complex mixtures including MS and NMR in **chapter 6**. We have used this platform to identify five taste-related unknown retention time and *m/z* features in soy sauce. Soy sauce is a complex mixture and the abundance of compounds of interest can be low, which makes structural elucidation by NMR a challenging task. For this, we developed a fractionation strategy for the purification of unknown features prior to NMR analysis. The fractionation employed a directed two-dimensional chromatography method, which was adaptable to differences in orthogonality of features and impurities. There are several chromatography phases known that offer orthogonal separations. Our results demonstrated that the combination of low and high pH reversed phase chromatography yielded the cleanest fractions for most unknown features. However, one feature highly benefited from the combination of reversed phase chromatography and HILIC. Fractions were collected off-line in order to concentrate the unknown features prior to the NMR analysis. The purified features were clean and concentrated enough for the acquisition of one and two-dimensional NMR spectra, which lead to the structural confirmation of the unknown features.

Although the use of one-dimensional chromatography and NMR is common practice, we have demonstrated that there are still highly abundant impurities present in first dimension fractions. These impurities might cause peak overlap with the unknown features preventing the recognition of typical spin patterns. Therefore, the use of a second dimension chromatography offers great potential in reducing the complexity further. However, the usefulness of the second dimension chromatography is highly dependent on its orthogonality to the first dimension. This

orthogonality is in turn highly dependent on the chemical properties of the unknown feature and the remaining impurities. Therefore, it is important to critically assess the composition of the first dimension fractions. This determines which second dimension chromatography results in the most efficient separation of the unknown feature from the remaining impurities.

FUTURE PERSPECTIVES

High-throughput quantification

In **chapter 2**, we demonstrated an unconventional use of high performance chromatography columns by trapping ion suppressors instead of employing a time-consuming separation of the analytes. In **chapter 3**, we extended this approach by adding multiple high performance columns in series, which enabled the allocation of ion suppressors and allowed for the analysis of a wide variety of chemical classes. We have shown that this approach decreased ion suppression drastically and that the sensitivity was sufficient for the analysis of a wide range of endogenous metabolites in biological samples.

The analysis time of the fractionation method in **chapter 3** was substantially lower in comparison with general LC-MS. However, in comparison with other high-throughput platforms, the analysis time is still relatively high. This can be explained by the hardware that is used for the fractionation method. A conventional LC-MS system was extended with an additional pump and two six-port valves. LC-MS systems include gradient pumps, which are used for the loading and extraction of LC columns. These pumps make use of a mixer with a relatively large dead-volume and are optimized for reliable gradient formation instead of fast solvent switches. Since run times are generally long, the autosampler of an LC-MS system is also not designed to deliver a high sample throughput. Moreover, the connections of three guard columns, six-port valves, pumps and an MS resulted in a relatively large dead-volume, which was disadvantageous for the throughput of the fractionation.

An example of a dedicated automated SPE device is the RapidFire system, which is manufactured by Agilent Technologies.¹⁶ In this system, the sample and solvent handling, tubing and positions of modules are all designed in a way that is most beneficial in terms of sample throughput. The analysis time for this method is dramatically lower than our proposed method. The analysis has been shown to be feasible within 8.5 s, whereas three minutes are needed in our fractionation platform. Where we perform three SPE procedures instead of one per analysis and achieve a significant reduction in matrix effect, the RapidFire is considerably faster in the execution of an SPE.

The coverage and cleanup efficiency of dedicated automated SPE devices are limited because only one SPE column is used at a time. Therefore, these systems would highly benefit from the inclusion of serially coupled high performance columns. When different sorbent chemistries are connected through six-port valves, a more comprehensive fractionation can be realized as is demonstrated in our fractionation platform. By extending already existing automated SPE devices with our proposed methodology, an optimized solvent handling and sample injection can be combined with a fractionation that is based on both polarity and charge. This would result in a fractionation platform that excels in both coverage and reduction of ion suppression as well as throughput.

Next generation data independent acquisition

In **chapter 5**, we demonstrated a promising global profiling method with a good qualitative and quantitative performance. The combination of variable SWATH and HILIC allowed for correct identifications and accurate quantifications. However, with increased Q1 mass windows, the occurrence of product ion overlap or complex MS/MS spectra remains a risk factor. When multiple compounds fall within the same SWATH window and have identical retention times, the association between precursor ions and product ions is lost. This makes it more difficult to deconvolute MS/MS spectra that are obtained with larger Q1 window sizes. Therefore, the selectivity of analytical methodologies using SWATH analysis should always be carefully examined. It might be possible that the performance of our variable SWATH and HILIC method only holds for the particular metabolomics application used in **chapter 5**. A new performance assessment has to be made when, for example, a different biological matrix is analyzed. Differences in sample types can substantially change the complexity and composition of the metabolites and it is possible that additional coeluting compounds result in more product ion overlap or complex MS/MS spectra. This could eventually lead to a lack of quantitative and qualitative performance. Moreover, SWATH was not selective enough for our high-throughput application due to the severe coelution of metabolites and the relatively large SWATH windows, which were a result of small peak widths.

Recently, Sciex introduced a new technique, which increases the selectivity of SWATH analysis: scanningSWATH. In scanningSWATH, the Q1 scans over the mass range of interest in steps of 0.1 Dalton per 0.1 ms within each MS cycle.¹⁷ The time-of-flight (TOF) pulses are synchronized with the Q1 scanning steps, which gives the MS data an extra dimension (Q1 m/z) next to the exact mass and retention time. This allows the user to plot the product ion intensity against Q1 m/z , which makes it possible to identify product ion overlap from other precursor ions. The extra dimension allows for a better association between product and precursor ions and, therefore, results in a more selective DIA type of fragmentation. The Q1 window size is in a similar range as

conventional SWATH, which means that Q1 windows encompassing the precursor ion of interest can be summed in order to increase the sensitivity of the acquired TOF scans.

In **chapter 5**, some DIA platforms demonstrated complex MS/MS spectra and product ion overlap. The lack of selectivity in current DIA techniques could be overcome by the use of scanningSWATH. Concerning the chromatography of the fractionation platform, the mass range of **chapter 5** is easily fragmented using the scanning speed of scanningSWATH. By introducing the extra Q1 m/z dimension, the product ions could more easily be traced back to their corresponding precursor ion, improving the selectivity of the MS/MS spectra. This results in cleaner MS/MS spectra for identification and a more accurate quantification of analytes. Eventually, this might lead to a high-throughput global profiling method, in which unknown features can be both identified as well as quantified.

The breadth of processing options for scanningSWATH data makes it challenging to develop data analysis software that exploits the full potential of scanningSWATH.¹⁸ When one wants to aim for the highest selectivity, it is key to sum TOF pulses that do not demonstrate product ion overlap with other coeluting precursor ions. However, when sensitivity becomes critical, it might be more important to sum all the data containing the concerned analyte. Eventually a compromise between sensitivity and selectivity should lead to data analysis software that allows for the implementation of scanningSWATH into the field of metabolomics.

Sensitive and affordable NMR analysis

In **chapter 6**, we demonstrated a comprehensive fractionation platform for the purification of unknown features in complex samples, resulting in purified features, which were clean and concentrated enough for the acquisition of one and two-dimensional NMR data. However, in order to have an adequate sensitivity during the NMR analysis, a sufficient amount of starting material is needed. Due to column capacity, ten consecutive injections were needed to realize this, which made the fractionation procedure relatively time-consuming. The amount of starting material could be reduced by the use of more sensitive NMR systems, e.g. higher field systems or cryogenically cooled probes. However, this increase in sensitivity comes at the cost of substantially higher instrument costs.

A cost-effective way of improving the sensitivity of NMR is the use of microcoils.¹⁹ The active volume of these particular coils is in the microliter range, which allows for a more sensitive and cheaper coil design. Moreover, the coil design makes it possible to position the sample closer to the detection coil, which in turn also increases the sensitivity. Olson *et al.* have shown that in comparison with a conventional NMR probe, the mass sensitivity of a microcoil probe is

approximately tenfold higher.²⁰ Because the sensitivity improvement is based on mass, the use of a microcoil is highly advantageous for mass-limited or pre-concentrated samples.

Our directed two-dimensional chromatography method could highly benefit from the use of microcoil NMR. The number of injections on the HPLC system could be reduced when the mass sensitivity in a microcoil is substantially higher. This will dramatically improve the throughput of the identification platform. Although there are clear advantages of the use of microcoil NMR, the sample introduction into the microcoil is less straightforward. Fractions have to be concentrated into a few microliters and subsequently positioned into the microcoil without dilution and diffusion, which is a challenging task. However, there are several analytical techniques which could make this possible. First of all, machine vision-controlled droplet evaporation has been shown to allow for solvent exchange and pre-concentration of samples into a submicroliter droplet.²¹ In this pre-concentration technique, samples are infused into a hanging droplet at a constant flow rate. The heating rates are adjusted by machine vision in order to maintain a stable droplet. The system allows for up to 90% solvent exchange, which can be used to replace the sample solvent by deuterated solvents for NMR analysis. Moreover, concentrating a sample into a droplet makes the sample volume and microcoil volume more compatible.

The transfer of concentrated microliter samples can be established by using flow NMR in combination with segmented flow analysis. When concentrated samples are injected into an immiscible carrier fluid, diffusion is prevented, which allows for the transfer of samples over long distances without causing dilutions. This is beneficial for flow NMR analysis, because the samples often have to be transferred over a few meters of tubing because the sampler is positioned outside the magnet's fringe field. Fluorocarbon fluid appeared to be a promising immiscible carrier fluid and it allowed for the transfer of microliter samples to a microcoil probe.²² The sample can be positioned and analyzed in the microcoil by a stop-flow analysis.

FINAL REMARKS

The impact of metabolomics in life sciences is becoming increasingly clear. At the heart of these developments are good analytical platforms, which allow for metabolite identification and quantification. MS and NMR have shown to be very powerful at providing qualitative and quantitative data. However, the potential of these techniques has not been fully exploited due to several analytical challenges. We have shown that these challenges can be overcome by using a fractionation approach. Fractionation can be applied in all shapes and sizes, which makes it highly flexible and adaptable. We have shown its applicability for purification prior to NMR analysis and high-throughput MS analyses. These methodologies demonstrated that, for the development of an analytical platform, knowledge about your matrix components is as important as knowledge

about your analytes. This knowledge allowed us to achieve an in-depth purification of metabolites from complex mixtures, which is essential for NMR identification. On-line fractionation realized a fast and comprehensive analysis of the metabolome while maintaining a performance that was comparable with conventional LC-MS. Therefore, we have made a big step in bringing large-scale metabolomics closer to reality. This progress can be continued by the implementation of some recent technological advances, which could highly benefit a fractionation-based analysis. We expect that fractionation coupled to a state-of-the-art analyzer will realize a global metabolic profiling method that allows for both selective identification and accurate quantification in a high-throughput fashion. A fast and global metabolic profiling method will facilitate the analysis of large-scale metabolomics studies and strengthen the impact of metabolomics in life sciences.

REFERENCES

1. Miggiels, P., Wouters, B., Westen, G. J. Van, Dubbelman, A. & Hankemeier, T. Novel technologies for metabolomics: more for less. *TrAC - Trends Anal. Chem.* 2018, 1–9 (2018).
2. Zampieri, M., Sekar, K., Zamboni, N. & Sauer, U. Frontiers of high-throughput metabolomics. *Curr. Opin. Chem. Biol.* 36, 15–23 (2017).
3. Sinclair, I. et al. Novel Acoustic Loading of a Mass Spectrometer: Toward Next-Generation High-Throughput MS Screening. *J. Lab. Autom.* 21, 19–26 (2016).
4. Gowda, G. A. N. & Djukovic, D. Overview of Mass Spectrometry-Based Metabolomics: Opportunities and Challenges. *Methods Mol Biol* 1198, 3–12 (2014).
5. Zhou, J. & Yin, Y. Strategies for large-scale targeted metabolomics quantification by liquid chromatography-mass spectrometry. *Analyst* 141, 6362–6373 (2016).
6. Trufelli, H., Palma, P., Famigliini, G. & Cappiello, G. An overview of matrix effects in liquid chromatography-mass spectrometry. *Indian J. Exp. Biol.* 30, 491–509 (2011).
7. Ismaiel, O. A., Zhang, T., Jenkins, R. G. & Karnes, H. T. Investigation of endogenous blood plasma phospholipids, cholesterol and glycerides that contribute to matrix effects in bioanalysis by liquid chromatography/mass spectrometry. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 878, 3303–3316 (2010).
8. Wang, Z. et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472, 57–63 (2011).
9. Trabado, S. et al. The human plasma-metabolome: Reference values in 800 French healthy volunteers; Impact of cholesterol, gender and age. *PLoS One* 12, 1–17 (2017).
10. Chuang, C. K. et al. A method for lactate and pyruvate determination in filter-paper dried blood spots. *J. Chromatogr. A* 1216, 8947–8952 (2009).
11. Giesbertz, P., Ecker, J., Haag, A., Spanier, B. & Daniel, H. An LC-MS/MS method to quantify acylcarnitine species including isomeric and odd-numbered forms in plasma and tissues. *J. Lipid Res.* 56, 2029–2039 (2015).
12. Matsunami, R. K., Angelides, K. & Engler, D. A. Development and validation of a rapid ¹³C6-Glucose Isotope Dilution UPLC-MRM Mass Spectrometry Method for Use in determining system accuracy and performance of blood glucose monitoring devices. *J. Diabetes Sci. Technol.* 9, 1051–1060 (2015).
13. TAKAHASHI, H. et al. Long-Chain Free Fatty Acid Profiling Analysis by Liquid Chromatography–Mass Spectrometry in Mouse Treated with Peroxisome Proliferator-Activated Receptor α Agonist. *Biosci. Biotechnol. Biochem.* 77, 2288–2293 (2013).
14. Liu, J. et al. Simultaneous targeted analysis of trimethylamine-N-oxide, choline, betaine, and carnitine by high performance liquid chromatography tandem mass spectrometry. *J.*

- Chromatogr. B Anal. Technol. Biomed. Life Sci. 1035, 42–48 (2016).
15. Pan, Z. & Raftery, D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Anal. Bioanal. Chem.* 387, 525–527 (2007).
16. Zhang, X. et al. SPE-IMS-MS: An automated platform for sub-sixty second surveillance of endogenous metabolites and xenobiotics in biofluids. *Clin. Mass Spectrom.* 2, 1–10 (2016).
17. Ivosev, G., Cox, D. M., Bloomfield, N., Wasim, F. & Leblanc, Y. Scanning SWATH® Acquisition Method for Improved Compound Screening. ASMS (2018).
18. Messner, C. et al. ScanningSWATH enables ultra-fast proteomics using high-flow chromatography and minute-scale gradients. *bioRxiv* 656793 (2019). doi:10.1101/656793
19. Grimes, J. H. & O'Connell, T. M. The application of micro-coil NMR probe technology to metabolomics of urine and serum. *J. Biomol. NMR* 49, 297–305 (2011).
20. Olson, D. L. et al. Microflow NMR: Concepts and capabilities. *Anal. Chem.* 76, 2966–2974 (2004).
21. Schoonen, J. W. et al. Solvent exchange module for LC-NMR hyphenation using machine vision-controlled droplet evaporation. *Anal. Chem.* 85, 5734–5739 (2013).
22. Kautz, R. A., Goetzinger, W. K. & Karger, B. L. High-throughput microcoil NMR of compound libraries using zero-dispersion segmented flow analysis. *J. Comb. Chem.* 7, 14–20 (2005).

Addendum

Nederlandse Samenvatting

Curriculum Vitae

Acknowledgements

NEDERLANDSE SAMENVATTING

Metabolomics is het onderzoeksveld dat zich richt op de analyse van kleine moleculen in een biologisch systeem. Deze kleine moleculen worden metabolieten genoemd en omvatten bijvoorbeeld aminozuren, vetzuren, suikers, hormonen en vitamines. Het metaboloom is de verzameling van alle metabolieten in een biologisch systeem en het zo goed mogelijk meten van het metaboloom is het hoofddoel van metabolomics.

Metabolietenconcentraties worden beïnvloed door een groot aantal chemische processen die zijn onder te verdelen in genetische factoren en omgevingsfactoren. Deze combinatie van factoren zorgt ervoor dat het metaboloom een goede weerspiegeling geeft van het fenotype van een biologisch systeem. Dit fenotype is specifiek voor elk persoon en kan daarom van grote toegevoegde waarde zijn voor gepersonaliseerde gezondheidszorg. Buiten de medische hoek worden metabolieten metingen ook toegepast in de voedselindustrie. Metabolietenprofielen kunnen bijvoorbeeld worden gebruikt voor het begrijpen van eigenschappen van voedselproducten, zoals voedingswaarde en smaak.

Metabolietenprofielen worden voornamelijk verkregen door massa spectrometrie (MS) en nucleaire magnetische resonantiespectroscopie (NMR). Deze technieken worden gebruikt voor het beantwoorden van de twee belangrijkste vragen in metabolomics: wat is de hoeveelheid van een metaboliet en wat is de identiteit van een metaboliet in een biologisch monster. Ondanks de ontwikkeling van hoog-geavanceerde analytische apparatuur, is de analyse van metabolieten onderhevig aan een aantal analytische uitdagingen. De grootste uitdaging bij MS-analyses is het optreden van matrix effect in complexe monsters. Matrix effect treedt op wanneer meerdere moleculen tegelijkertijd worden geïntroduceerd in de MS, waardoor ze met elkaar in competitie gaan voor het verkrijgen van een lading. Het verkrijgen van een lading wordt ook wel ionisatie genoemd en is nodig om gemeten te kunnen worden in de MS. Sommige metabolieten verliezen deze competitie, waardoor het signaal van deze metabolieten onderdrukt kan worden en onder de detectiegrens kan vallen. Daarnaast is de samenstelling van monsters niet altijd gelijk, waardoor de ladingcompetitie verandert en dus ook de herhaalbaarheid van de analyses verminderd wordt. Bij NMR-analyses vormen voornamelijk het gebrek aan gevoeligheid van de metingen en signaaloverlap de grootste problematiek.

In dit proefschrift tackelen wij de grootste uitdagingen van MS en NMR door middel van een fractioneringsaanpak. Het verdelen van een monster over fracties zorgt ervoor dat de complexiteit wordt verminderd voordat de analyse plaatsvindt. Dit zorgt aan de ene kant voor minder matrix effect tijdens de MS-analyse. Hierdoor kan er sneller gemeten worden omdat tijdrovende chromatografische scheidingen niet meer nodig zijn. Daarnaast kan de kwaliteit van NMR spectra drastisch worden verhoogd door het fractioneren van een complex monster omdat er minder

signaaloverlap optreed en de fracties kunnen worden geconcentreerd. Het onderzoek van dit proefschrift is uitgevoerd in het MI3 consortium en gefinancierd door COAST/NWO.

In **hoofdstuk 2** hebben wij een methode ontwikkeld waarmee matrix effect snel kan worden verminderd door middel van een onconventioneel gebruik van vloeistofchromatografie (LC). Het doel was in dit geval niet om analieten te scheiden maar om beruchte matrix componenten te binden. Tijdens de analyse van vijf potentiële biomarkers voor hart- en vaatziekten, zagen wij dat fosfolipiden de voornaamste bron waren voor signaalonderdrukking. Daarom maakten wij gebruik van een reversed phase kolom voor het binden van de (fosfo)lipiden, waardoor de polaire potentiële biomarkers in een schone flow-through konden worden gemeten. De analysetijd kon drastisch worden verlaagd omdat de analieten geen retentie ondervonden op de LC-kolom en de (fosfo)lipiden vervolgens snel van de LC-kolom konden worden geëluëerd door snelle gradiëntwisselingen. De vloeistofstroom van de eluerende (fosfo)lipiden werd naar een afvalvat gestuurd zodat de ionisatiebron van de MS minder vervuild werd. De snelle LC-MS methode verwijderde de signaalsuppressie die was veroorzaakt door de fosfolipiden in een analyse tijd van slechts drie minuten.

Hoofdstuk 2 toonde aan dat, naast fosfolipiden, ook zouten een belangrijke bron van signaalsuppressie zijn. Daarom hebben wij de fractioneringsaanpak van **hoofdstuk 2** uitgebreid met twee extra kolommen in **hoofdstuk 3**. De extra kolommen waren samengesteld uit een kationen- en anionenwisselaar. De combinatie van een reversed phase kolom en de ionwisselaars zorgde voor een fractionering die was gebaseerd op polariteit en lading. Hierdoor was het mogelijk om het aantal analieten uit te breiden naar vijftig bekende prognostische en diagnostische biomarker metabolieten. In deze methode konden wij geen fracties verwijderen omdat de analieten een veel breder chemisch bereik omvatten. Hierdoor was het doel van de fractionering om de bekende signaalonderdrukkers te verspreiden over verschillende fracties zodat hun negatieve effect tijdens de ionisatie kon worden geminimaliseerd. Het gebruik van drie in serie geschakelde kolommen zorgde hiervoor door fosfolipiden, positieve en negatieve zouten te verspreiden over verschillende fracties. De korte analysetijd van het fractioneringsplatform van **hoofdstuk 3** werd bereikt door het snel schakelen van mobiele fasen en het gebruik van korte kolommen (≤ 1 cm). Dit resulteerde in een analysetijd van drie minuten per MS polariteit. Bovenop de scheiding tussen fracties zorgde het gebruik van hoogwaardige vastefase-extractie kolommen (deeltjesgrootte ≤ 5 μm) voor een scheiding binnen de fracties. Hierdoor werden fosfolipiden van acylcarnitines en vetzuren gescheiden in dezelfde fractie. Het voordeel van fractioneren werd benadrukt door het vergelijk met een directe injectie in de MS en met LC-MS. In vergelijking met een directe injectie in de MS werd de signaalonderdrukking verminderd van 89% naar 25%. Daarnaast was de gevoeligheid van het fractioneringsplatform vergelijkbaar aan conventionele

LC-MS analyses. LC-MS analyses zijn doorgaans specifiek ontwikkeld voor een bepaalde stofklasse en de scheiding duurt aanzienlijk langer (3-22 minuten per monster). Wij laten zien dat fractionering het mogelijk maakt om meerdere stofklassen snel en in één platform te meten met een vergelijkbare gevoeligheid als LC-MS.

In **hoofdstuk 4** hebben wij een platform ontwikkeld dat zowel de kwantificatie als de identificatie van metabolieten mogelijk maakt. Om dit te bewerkstelligen hebben wij een MS techniek gebruikt die fragmentatiedata verkrijgt van alle stoffen in een meting. Deze techniek heet sequential window acquisition of all theoretical mass spectra (SWATH). Informatie over de fragmentatie van stoffen is van groot belang bij het bepalen van de structuurformule van onbekende stoffen. Daarnaast kan fragmentatie data gebruikt worden om stoffen met dezelfde massa, isomeren, van elkaar te onderscheiden. SWATH maakt het mogelijk om fragmentatie data van alle stoffen in een monster te verkrijgen op een selectieve manier. De massa's van de analieten worden in de massa spectrometer verdeeld in verschillende pakketten en vervolgens gefragmenteerd. De indeling van deze pakketten kan worden bepaald door een vast massabereik per pakket (vaste SWATH) of een variabel massabereik dat zorgt voor een evenredige verdeling van de analieten per pakket (variabele SWATH). Wij hebben laten zien dat variabele SWATH in combinatie met een conventionele chromatografische scheiding leidde tot correcte identificaties en accurate kwantificaties. Unieke fragmentatiepatronen konden worden bepaald voor de onderzochte isomeren, wat resulteerde in de accurate kwantificatie van vijf isomerenparen. Daarnaast hebben wij 20 vooraf bekende metabolieten kunnen identificeren met behulp van de vaste en variabele SWATH data. De fragmentatie data konden worden gebruikt om de 20 metabolieten terug te vinden in een MS spectra bibliotheek.

De betrouwbaarheid van de identificatie van onbekende metabolieten neemt aanzienlijk toe wanneer deze wordt uitgevoerd door twee complementaire technieken. Daarom hebben wij in **hoofdstuk 5** een identificatiemethode ontwikkeld met behulp van MS en NMR. Hiervoor hebben wij het probleem van signaaloverlap in NMR spectra en de lage gevoeligheid van NMR metingen aangepakt door middel van een fractioneringsmethode. Eendimensionale chromatografie wordt vaak gebruikt om onbekende stoffen te zuiveren voordat de NMR analyse plaatsvindt. Onze resultaten laten echter zien dat er nog veel vervuilingen aanwezig zijn in de fracties van een eendimensionale scheiding. Deze vervuilingen kunnen signaaloverlap veroorzaken met de onbekende stoffen en zo een juiste identificatie voorkomen. In onze identificatiemethode hebben wij gebruik gemaakt van een tweedimensionale scheiding om de fracties schoner te krijgen. Ten eerste werd achterhaald welke vervuilingen er nog aanwezig waren in de fracties. Daarna werd bekeken welke tweede dimensie chromatografie het meest geschikt was om deze vervuilingen te scheiden van de onbekende stof. Vervolgens werd deze scheiding gebruikt voor het verder

zuiveren van de onbekende stoffen. Dit resulteerde in fracties die geconcentreerd en puur genoeg waren voor het verkrijgen van zuivere NMR spectra.

Metabolomics wordt steeds vaker toegepast in de gezondheidszorg, voedselindustrie en vele andere sectoren. Het succes van metabolomics is voor een groot deel te danken aan goede analytische methodes die zorgen voor de identificatie en kwantificering van metabolieten in biologische monsters. MS en NMR hebben in het verleden bewezen erg krachtig te zijn in het verkrijgen van kwalitatieve en kwantitatieve data van metabolieten. Er zijn, echter, een aantal uitdagingen waardoor het potentieel van deze technieken niet volledig benut kan worden. Wij hebben deze uitdagingen aangepakt door middel van fractionering. Het fractioneren van biologische monsters stelde ons in staat om snelle MS metingen uit te voeren en zuivere NMR spectra te verkrijgen. Deze methodes lieten zien dat voor de ontwikkeling van een goede analytische methode, kennis over de matrix componenten net zo belangrijk is als kennis over de analieten. Wij verwachten dat fractionering in combinatie met MS en/of NMR zal zorgen voor een snel analysesysteem, dat kan worden gebruikt voor zowel de selectieve identificatie als de accurate kwantificatie van het metabool. Zo'n algemeen meetsysteem zal de analyse van grootschalige metabolomics onderzoeken faciliteren en de impact van metabolomics op levenswetenschappen versterken.

CURRICULUM VITAE

Tom van der Laan was born on February 8 1992 in Amersfoort. In 2010, he obtained his high school (Gymnasium) diploma at Johan van Oldenbarnevelt in Amersfoort. After his graduation, he moved to Leiden to continue his studies at Leiden University. He enrolled for the bachelor Biopharmaceutical Sciences in which he was introduced with analytical chemistry for the first time by Peter Lindenburg and Rawi Ramautar. During his bachelor internship, he developed his first bioanalytical platforms using capillary electrophoresis (CE) and mass spectrometry (MS).

In 2014, Tom started the master Biopharmaceutical Sciences, in which he choose the specialization track 'Analytical Biosciences'. His first internship was supervised by Rawi Ramautar and Rosilene Burgos and focused on correlating ultra-weak photon emission with metabolic changes using CE-MS. This research successfully identified seven metabolites which were inversely correlated with ultra-weak photon emission and resulted in his first publication.

For his second internship, Tom moved to Karlsruhe to work at Agilent Technologies. Under the supervision of Christian Wenz and Peter Lindenburg, he developed a method for a new CE-MS interface in which the dilution factor of the sheath liquid was minimized. The evaluation of the new interface demonstrated promising results with regards to the sensitivity of CE-MS analyses.

In 2016, Tom started as a PhD student in the group of Professor Thomas Hankemeier. Under the supervision of his co-promotors Anne-Charlotte Dubbelman and Amy Harms, he focused on the high-throughput quantification and unambiguous identification of metabolites in complex samples. He developed a fast and on-line fractionation technique, which lead to a patent application and a publication in Analytical Chemistry. In collaboration with Sciex, he evaluated different data independent acquisition techniques. He presented these results in a webinar hosted by the Analytical Scientist. In a joined academic and industrial consortium, he developed an identification platform for metabolites in complex samples. For this project, he collaborated with DSM, Unilever and Wageningen University, which lead to the successful identification of five taste-related metabolites in soy sauce.

After his PhD, Tom is going to work in the Analytical Assay group at Janssen Vaccines in Leiden. As a scientist, he will be responsible for the development of analytical methods for the characterization of new vaccines.

ACKNOWLEDGEMENTS

First of all, I would like to highlight the major role of Peter and Rawi during my university studies. They have introduced me with the beauty of science and inspired me to pursue a scientific career.

Thomas, I admire your vision and your critical mind. I have always been amazed by your drive to explore and implement the power of metabolomics. Our many fruitful discussions have formed me into the researcher that I am today.

Amy, you have supervised me since day one. Your door was always open and you were always willing to discuss the challenges that I faced. Anne-Charlotte, you have done a tremendous job as a first time co-promotor. Our weekly meetings really inspired and motivated me to bring our research to the next level and allowed me to see the bigger picture of my PhD.

I would like to thank my students Xiaochun, Joshua, Merve, Isabelle, Firda, Koen and Laurens for contributing to my PhD research and giving me the opportunity to develop my supervision skills. Special thanks to Joshua and Isabelle who performed most of the research of chapter 4.

I am happy to have worked with so many nice colleagues from Analytical Biosciences. The good laughs during lunch, conferences and borrels were a welcome distraction from my scientific activities. Special thanks to fight fight fight, salmon and my other office mates for creating an office that sometimes felt more like a social gathering place than an actual work area. I would also like to emphasize Tim's role in this thesis, since he laid the foundation of chapter 2.

I have also enjoyed the collaborations within the MI3 consortium with Wageningen University, Unilever and DSM as partners. I have learned a lot about the development and implementation of analytical methods for industrial applications.

Very special thanks to Milou, who always created a nice home after an intense day at the lab. She has always been supportive throughout my PhD and very patient in hearing my developments on 'stofjes meten'.

Finally, I would like to thank my parents, Paul and Francie, who have raised me with all the love and freedom I could wish for. Their support has been crucial in my development as a person and scientist.