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 vitamines. 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

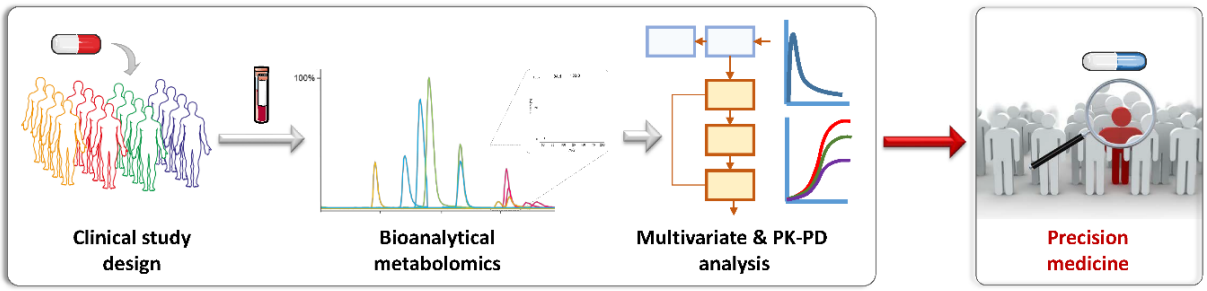

Figure 1. Precision medicine workflow using clinical metabolomics-based biomarker discovery and clinical pharmacology. Adapted from Kohler et al.[8]

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

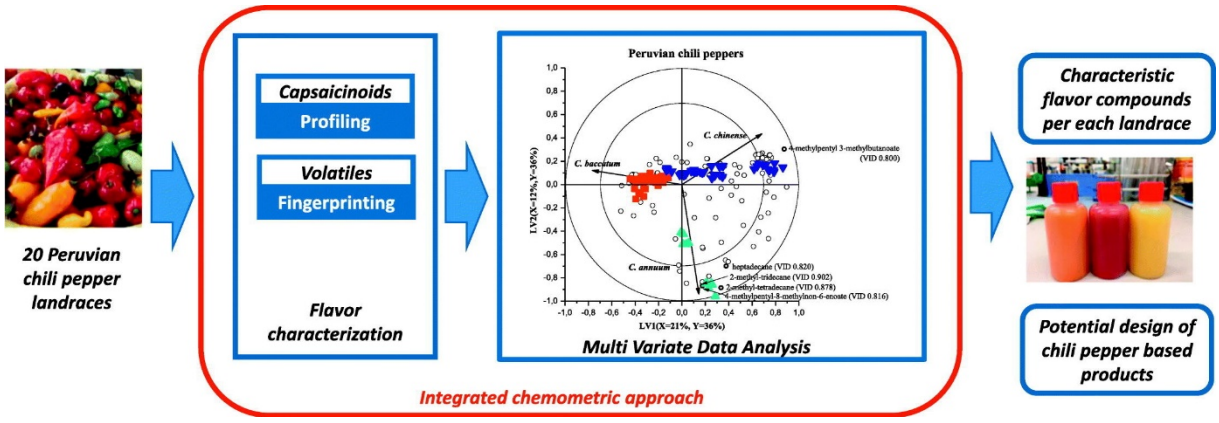

Figure 2. Pathway for the identification of taste-related features in food products. The profiling of several compound classes and flavor characterization can result in the identification of characteristic flavor compounds. Adapted from Morales-Soriano et al.[18]

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 powerfull 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 ^{1}H NMR. ^{1}H 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

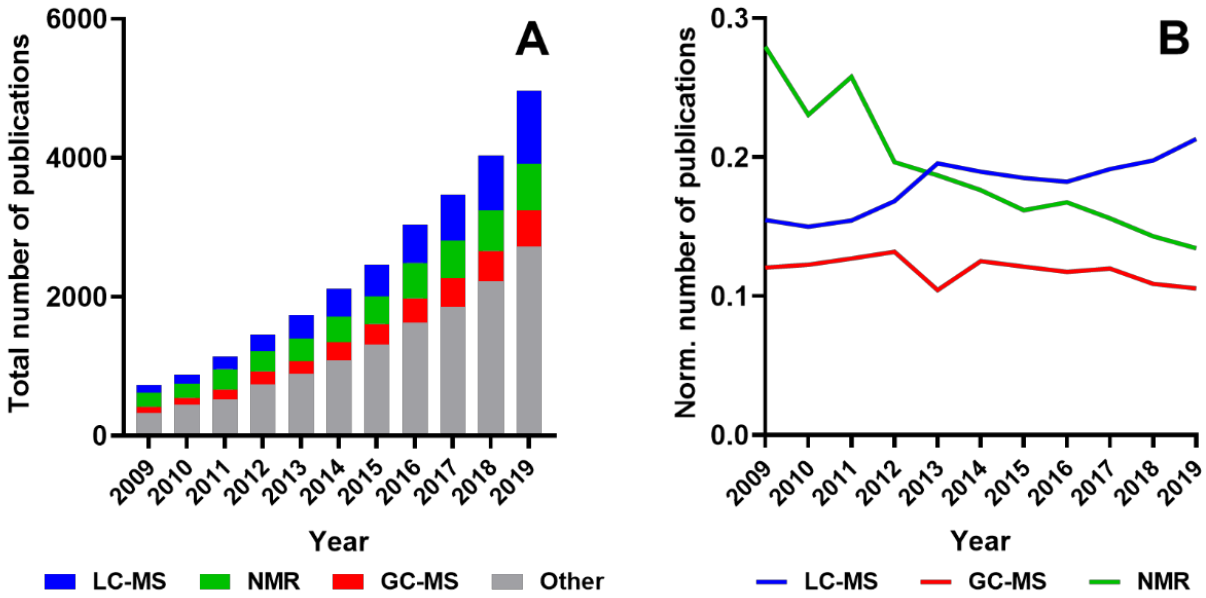

Figure 3. Overview of a literature survey on PubMed including metabolom or metabonom* in the title or abstract in a period between 2009 and 2019. (A) The total number of publications using the most popular metabolomics techniques, i.e. NMR, LC-MS and GC-MS, are indicated with different colors. (B) Normalized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

^{1}H 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. ^{1}H-^{1}H, and heteronuclear, e.g. ^{1}H-^{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 MMCD[29], BMRB[30], HMDB[31] and COLMAR[32], 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), HMDB[31], 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

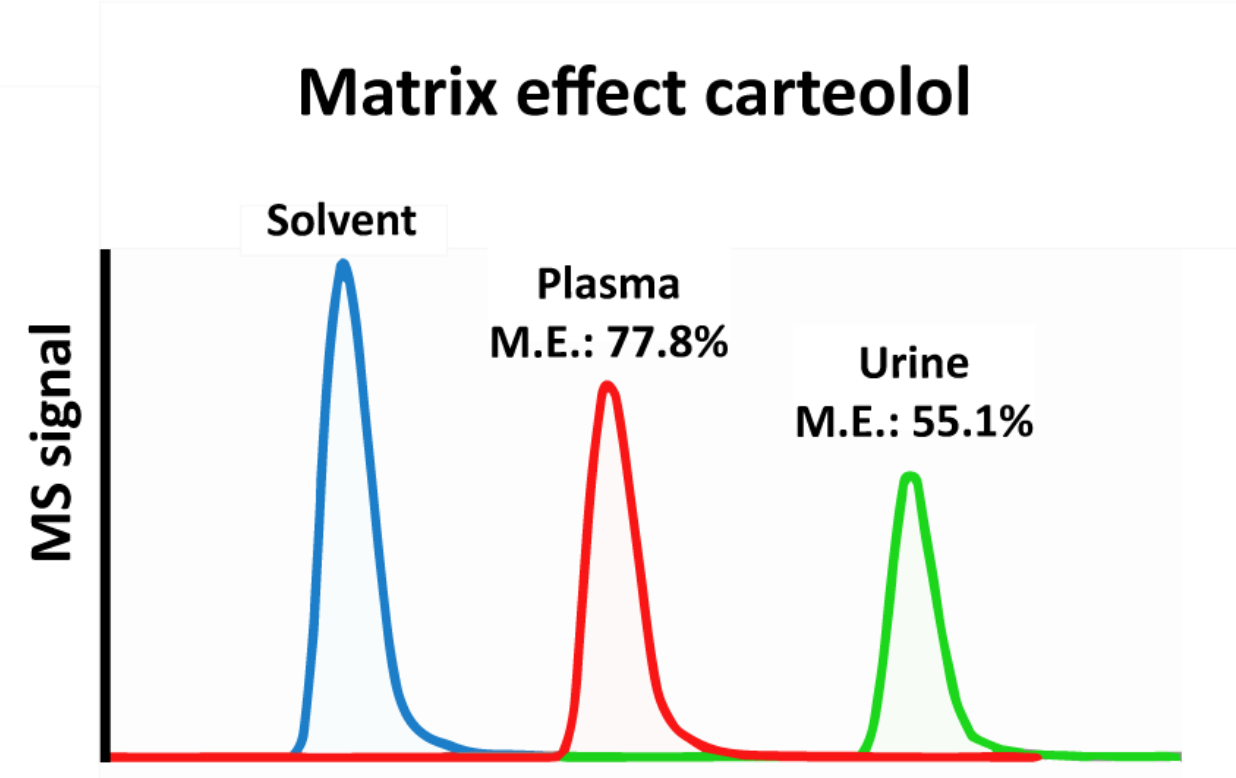

Figure 4. The MS signal of carteolol in different biological matrices. Matrix effect (M.E.) is observed in the plasma and urine analysis. The plasma and urine matrix reduced the MS signal of carteolol to 77.8% and 55.1% of its original value, respectively. Adapted from Gonzalez et al.[70]

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 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

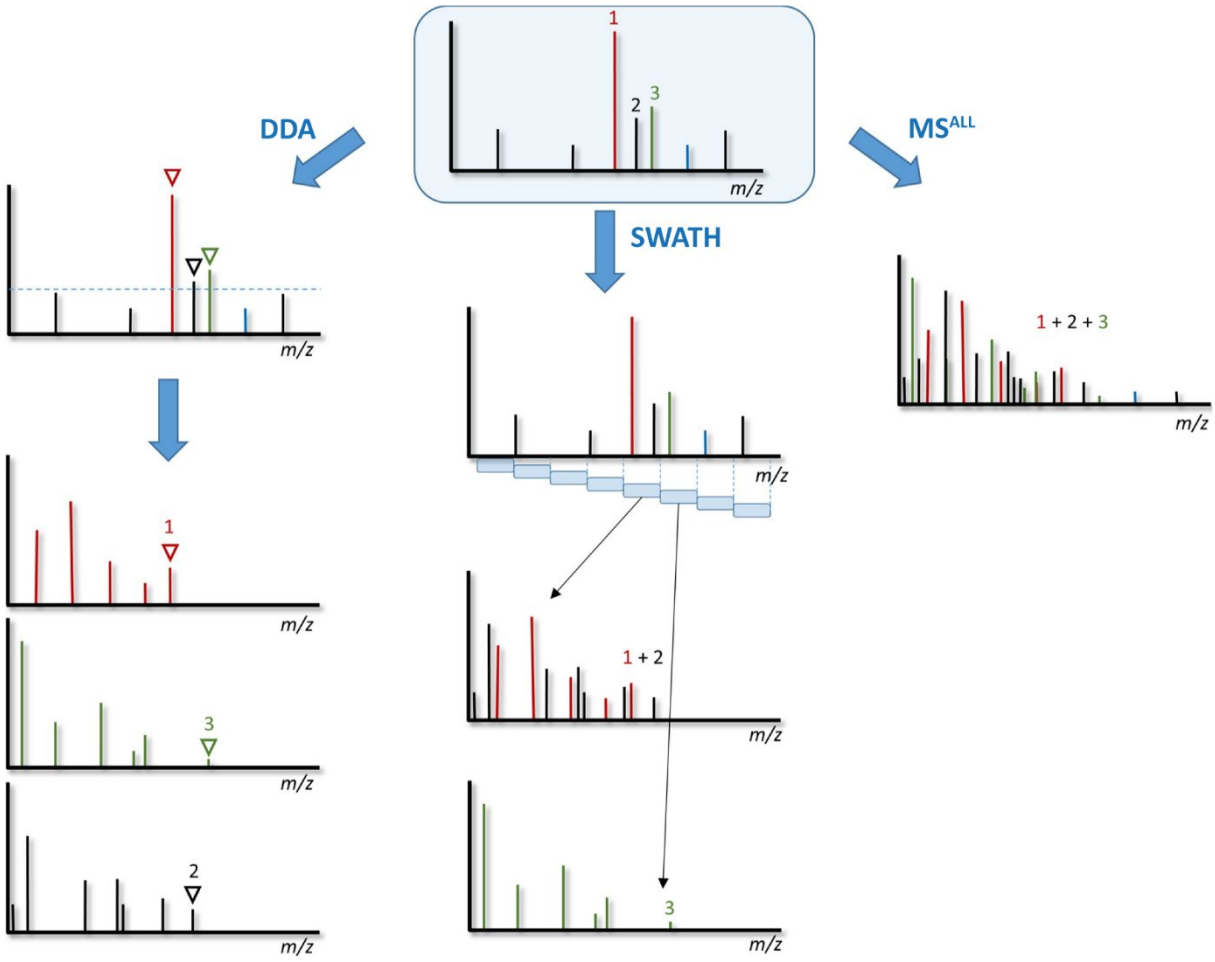

Figure 5. MS/MS procedures for data-dependent acquisition and data-independent acquisition, i.e. SWATH and MS^{ALL}. Adapted from Fenaille et al.[25]

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

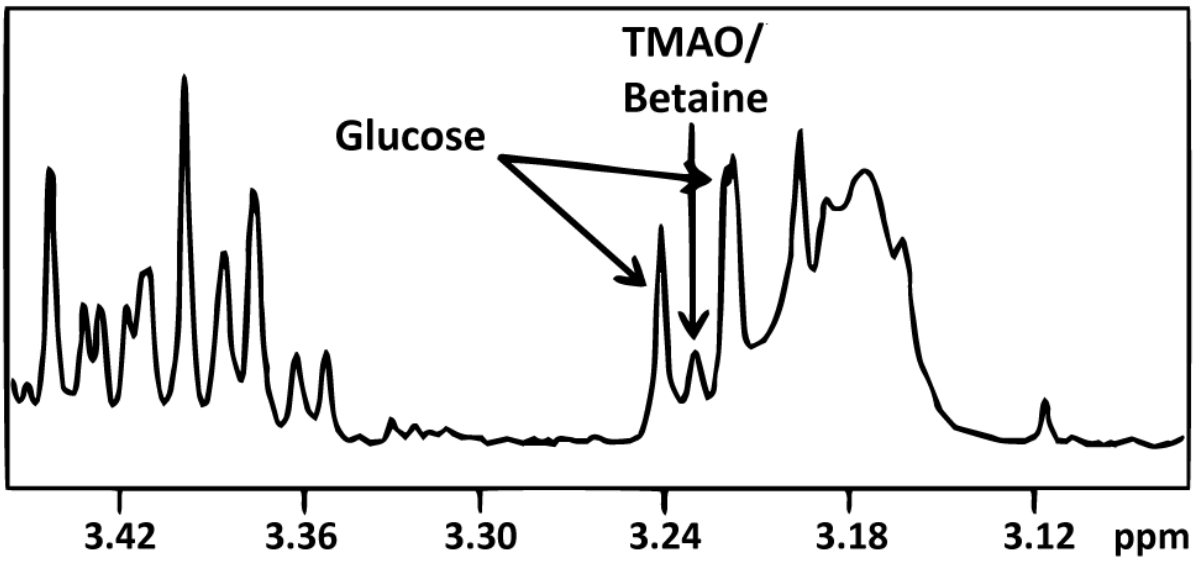

Figure 6. The analysis of a blood sample at physiological pH using a clinical ^{1}H NMR analyzer (400 MHz). The NMR spectra shows overlap of TMAO, betaine and glucose. Adapted from Garcia et al.[64]

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 ^{1}H 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. & Brüschweiler, 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., Brüschweiler-Li, L. & Brüschweiler, 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. & Ressom, 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. 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).
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]))