UPPSALA
UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy 306

Development of analytical methods for the determination of the small molecule component of complex biological systems

KRISTIAN PIRTTILÄ

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022

ISSN 1651-6192
ISBN 978-91-513-1374-0
URN urn:nbn:se:uu:diva-461320

Dissertation presented at Uppsala University to be publicly examined in B22, BMC, Husargatan 3, Uppsala, Friday, 18 February 2022 at 08:00 for the degree of Doctor of Philosophy (Faculty of Pharmacy). The examination will be conducted in English. Faculty examiner: Professor Thomas Moritz (Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark).

Abstract
Pirttilä, K. 2022. Development of analytical methods for the determination of the small molecule component of complex biological systems. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy* 306. 59 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1374-0.

The research field of untargeted metabolomics aims to determine the relative abundance of all small metabolites in a biological system in order to find biomarkers or make biological inference with regards to the internal or external stimuli. This is no trivial aim, as the small metabolites are both vast in numbers and extremely diverse in their chemical properties. As such, no single analytical method exist that is able to capture the entire metabolome on its own. In addition, the data generated from such experiments is both immense in volume and very complex. This forces researchers to use algorithmic data processing methods to extract the informative part of this data. Such algorithms are, however, both difficult to parametrize and designed to be highly inclusive, the combination of which often leads to errors. One such algorithm is the peak picking procedures used to find chromatographic peaks in liquid chromatography-mass spectrometry (LC-MS) data.

In this thesis, four papers are included that focus both on the development of new methods for sample analysis and data processing as well as the application of such, and other, methods in two interdisciplinary research projects. The first paper describes the development and application of a protocol for LC-MS based untargeted analysis of guinea pig perilymph. The focus of the study was to investigate the biochemical processes underlying the protective effect of hydrogen gas on noise-induced hearing loss (NIHL) in guinea pigs exposed to impulse noise. This study sparked two research projects based on limitations observed during the analytical work. The first limitation was that of limited chemical coverage in the analysis when sample volumes are highly limited. The second paper describes the design and validation of a novel separation method for the sequential analysis of both hydrophilic and lipophilic compounds in biological samples. The second limitation observed was the abundance of false peaks reported by peak picking software. These have a negative effect on both downstream data processing as well as data analysis and metabolite identification. The third paper describes the development of a new algorithm for comprehensive peak characterization in untargeted analytical data with the purpose of filtering such false peaks. Both methods presented in the second and third paper were applied to the analysis of guinea pigs perilymph samples in a follow-up study on the attenuating effect of hydrogen gas on NIHL in guinea pigs exposed to broad band continuous noise.

Keywords: metabolomics, data processing, peak characterization, algorithm, multivariate data analysis, chromatography, sequential columns, HILIC, RPLC, liquid chromatography, mass spectrometry, LC, MS, noise-induced hearing loss, NIHL, guinea pig, perilymph, method development, validation, R, C++

Kristian Pirttilä, Department of Medicinal Chemistry, Analytical Pharmaceutical Chemistry, Box 574, Uppsala University, SE-75123 Uppsala, Sweden.

ISSN 1651-6192
ISBN 978-91-513-1374-0
URN urn:nbn:se:uu:diva-461320 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-461320)

If you don't know anything about computers, just remember that they are machines that do exactly what you tell them but often surprise you in the result.

— Richard Dawkins, *The Blind Watchmaker*

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I. **Pirttilä, K.***, Videhult Pierre, P., Haglöf, J., Engskog, M., Hedeland, M., Laurell, G., Arvidsson, T., Pettersson, C. (2019) An LCMS-based untargeted metabolomics protocol for cochlear perilymph: highlighting metabolic effects of hydrogen gas on the inner ear of noise exposed Guinea pigs. *Metabolomics*, 15, 138

II. **Pirttilä, K.***, Laurell, G., Pettersson, C., Hedeland, M. (2021) Automated Sequential Analysis of Hydrophilic and Lipophilic Fractions of Biological Samples: Increasing Single-Injection Chemical Coverage in Untargeted Metabolomics. *Metabolites*, 11, 295

III. **Pirttilä, K.***, Rainer, J., Pettersson, C., Hedeland, M., Brunius, C., Comprehensive Peak Characterization (CPC) in untargeted LC-MS analysis. *Manuscript submitted to Metabolites*

IV. **Pirttilä, K.***, Fransson, A.E., Videhult Pierre, P., Laurell, G., Pettersson, C., Hedeland, M. LCMS-based untargeted metabolomics of guinea pig perilymph using a novel separation method for sequential analysis of hydrophilic and lipophilic compounds: Studying the effect of hydrogen gas on noise-induced hearing loss. *Manuscript*

* Corresponding author

Additional papers not included in this thesis:

Fransson, A. E., Kisiel, M., **Pirttilä, K.**, Pettersson, C., Videhult Pierre, P. and Laurell, G. F. E. (2017) Hydrogen Inhalation Protects against Ototoxicity Induced by Intravenous Cisplatin in the Guinea Pig. *Front. Cell. Neurosci.*, 11:280. doi: 10.3389/fncel.2017.00280

Contents

Abbreviations

ACN	Acetonitrile
AUROC	Area under the receiver operating characteristic curve
CPC	Comprehensive peak characterization
DA	Discriminant analysis
ESI	Electrospray ionization
GC-MS	Gas chromatography-mass spectrometry
HRMS	High-resolution mass spectrometry
LC-MS	Liquid chromatography-mass spectrometry
NIHL	Noise-induced hearing loss
NMR	Nuclear magnetic resonance
OPLS-DA	Orthogonal projection to latent structures-discriminant analysis
PCA	Principal component analysis
RF	Random Forest

Introduction

All living organisms are composed of beautifully complex systems of interlinked chemical processes. Representing the end-product of these systems are the small metabolites, together referred to as the metabolome [1]. There have been various attempts to provide a definition of the metabolome, for example it has been described as all small molecules with a molecular weight less than 2000 Da [2]. These compounds are generally considered better representations of the phenotype than the genes and proteins of the biochemical response to various stimuli [3]. To that end, the research field of metabolomics aim at the comprehensive analysis of all such compounds in a biological system. While the idea of using the abundance of metabolites to study biochemical processes is not necessarily new [4], the field of metabolomics as it is known today has only been around since the start of the 21st century. The fast development during the last two decades can be primarily attributed to two factors; the performance advances in the analytical instrumentation available in addition to the development of algorithms enabling the timely processing and analysis of the high dimensional data generated. While originally defined separately and differing in key aspects [5–7], the terms metabolomics [1], metabonomics [3], and metabolic profiling [6,8] are today used interchangeably. In short, the general aim of the field is to relate the state or change of state, of the metabolome to a biological research question. To that end, metabolomics has been applied in numerous different fields such as cancer research [9–11], hearing research [12,13], pharmaceutical research [14–17], and nutritional research [18–20] to name a few.

The end result of a metabolomics study may differ depending on the aim of the research. In essence, the field is often discussed in terms of biomarker discovery or hypothesis generating prospective studies. In biomarker discovery, the end goal is to find metabolites, or combinations of metabolites, that can be used for physiological inference [21], e.g. diagnosis of disease [11,22–24] or dietary exposure [19,25]. On the other hand, in hypothesis generation, the goal is instead that of studying the system in order to better understand the biochemical processes underlying a phenomenon. In such a way, metabolomics has been applied successfully to study biological phenomena such as treatment mechanisms [26–28], disease development [29–31], effects of life style and nutrition [32,33], and toxicity [34–36]. Furthermore, the field can also be

divided in into targeted and untargeted metabolomics based on the methodologies used [7,37]. In targeted metabolomics, a specific subset of metabolites or compound classes, selected based on prior knowledge, are studied [5]. The analytical problem thus heavily resemble that of classical chemical analysis as selective quantitative methods can be developed and validated for the specific metabolites of interest. In untargeted metabolomics, no such prior selection of metabolites or compound classes is made. Instead the aim becomes that of comprehensive analysis of all metabolites in the studied system [5]. For this reason, care must be taken to infuse as little bias as possible by performing a minimum of sample preparation and using analytical methods with as wide coverage of chemical space as possible. Furthermore, developing quantitative methods for all compounds in a sample is impossible, and instead relative abundance is used to compare samples from different study groups or time points. Consequently, the importance of high reproducibility and robust methods in every stage of an untargeted metabolomics study is paramount. This thesis will mainly focus on hypothesis generating studies using untargeted metabolomics.

This lack of target compounds make method development in untargeted metabolomics challenging. What is more, the thousands of chemical compounds [7,38] in biological samples are incredibly diverse. The concentrations may range from picomolar to millimolar levels in addition to an immense diversity in terms of chemical properties [7]. This makes the selection of suitable methods in all steps from sample collection to data analysis very difficult, not to mention the evaluation of method performance. In the perspective of analytical methods, no single method available today allow complete coverage of an entire metabolome [5,39]. This means, in general, that compromises will have to be made with regards to chemical coverage. While in targeted analysis, all the steps of the analytical chain can be selected and optimized with direct respect to the compounds of interest. As mentioned, this is not possible in untargeted analysis where such targets do not exist and any optimization toward specific compounds, even in large numbers, is only truly representative of those compounds. While there have been efforts to specify guidelines for method validation in untargeted analysis [40,41], there is still no consensus. Often, proxy measures such as number of detected peaks or number of detected features are used to evaluate the methods. While they do provide some information related to the performance of the sample preparation and chromatographic separation, these measures are also highly dependent on fragmentation and adduct formation in the mass spectrometer. Thus, more detected peaks do not necessarily translate to better method performance and vice versa. Instead, a better approach is to use a combination of different ways to estimate method performance. In this thesis, the various aspects of untargeted metabolomics studies will be discussed from the basis of the analytical chain with examples from the included studies.

Aims

The overarching aim of this doctoral project was to develop new and existing analytical and data processing methods used in untargeted metabolomics. During the course of the project, a focus has been on increasing the quantity and improving the quality of the information obtained. To this end, new methods were developed for the analysis of guinea pig perilymph samples from two separate studies of the attenuating effect of hydrogen gas on noise-induced hearing loss (NIHL) in guinea pigs.

The specific aims of the included papers were to …

- … develop and apply an untargeted metabolomics protocol on *Scala tympani* perilymph from guinea pigs in a study on the protective effect of hydrogen gas on NIHL in guinea pigs exposed to impulse noise (**Paper I**). This work led to the identification of two urgent development areas, chemical coverage per sample injection, and the presence of falsely reported peaks in the data which sparked two development projects with the aim to …
- … increase the chemical coverage per sample injection through development of a separation method using liquid chromatography for singe-injection sequential analysis of both hydrophilic and lipophilic compounds in biological samples (**Paper II**) and …
- … reduce the number of false peaks reported during the data processing through development of an algorithm for comprehensive characterization of peaks in untargeted analysis to be used for filtering of false peaks reported by peak detection software (**Paper III**). And to …
- … apply the two developed methods in an untargeted metabolomics study of the protective effect of hydrogen gas on NIHL in guinea pigs exposed to broad band continuous noise using perilymph as sample material (**Paper IV**).

The analytical chain in untargeted metabolomics

In untargeted metabolomics the analytical chain can be divided into a number of steps (Figure 1): (i) sample generation, (ii) sample preparation, (iii) sample analysis, (iv) data processing, (v) data analysis, and (vi) identification of metabolites, and (vii) biological interpretation. Care must be taken in all these steps as they all have a strong influence on the overall quality of the data obtained after the analysis [42].

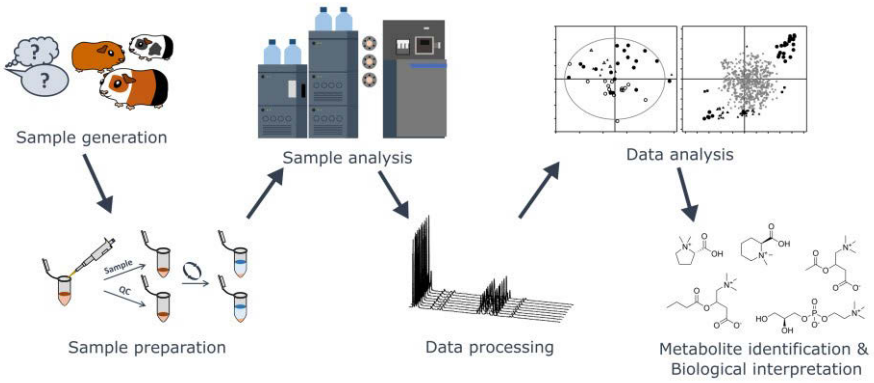

Figure 1 The different steps of the analytical chain in an untargeted metabolomics study.

This summary will discuss each section of the analytical chain in terms of important factors to consider and their implication on the end result. During the doctoral project presented in this thesis development work has primarily been performed in sample analysis (**Paper II**) and data processing (**Paper III**). Thus, **Paper II** and **Paper III** are discussed in detail under the corresponding sections. **Paper I** and **Paper IV** are applications of the untargeted metabolomics workflow detailed here. In **Paper I** a method for untargeted metabolomics of the hydrophilic component of guinea pig perilymph was presented and applied to samples taken from a study of the protective effect of hydrogen gas on NIHL in guinea pigs exposed to impulse noise. **Paper IV** details the application of the methods developed in **Paper II** and **Paper III** to

the analysis of perilymph samples taken from guinea pigs in a follow-up study, where the guinea pigs were instead exposed to broadband continuous noise.

Sample generation

Any chemical analysis can be corrupted by a poor sample generation [43]. For this reason, care must be taken to ensure that the sampling procedures used are reproducible. In this respect, all steps of the procedure from the selection of sample material to the time delays between sampling and processing and even laboratory material choices have an influence on the results of a metabolomics study [44]. In order to acquire answers to the research question posed, representative samples must also be acquired that relate to the system studied. Moreover, the number of replicate samples also needs careful thought to ensure that the analysis has sufficient statistical power [45]. Due to the ease and relative non-invasiveness of the sampling, among the most common sample materials in *in vivo* studies are plasma/serum and urine [43]. In addition to these, other sample material such as cells and tissue samples are also common. A benefit of using blood as sample material is that is comes into contact with all major organs of the organism [46]. This means that a lot of information about the organism studied can be obtained from a single sample [47]. This is especially beneficial when little is known about the phenomenon being studied, such as a poorly investigated disease or a new treatment. However, when inherently localized afflictions are studied, such as those studied in **Paper I** and **Paper IV**, where the condition is relatively localized to the tissues of the inner ear, using tissues or liquids sampled in high proximity to the afflicted region may be better [43]. As blood contacts all organs, this means that blood represent more of an average metabolic state of the entire organism [46] which might obscure such localised effects. In **Paper I** and **Paper IV**, *Scala tympani* perilymph was instead used as this is a relatively localized liquid that is in indirect contact with the sensory cells or hearing, and may therefore offer a better representation of the local metabolic effects studied. However, perilymph is not readily available in human studies as sampling would result in loss of hearing to the patient. Sampling is also difficult in animals, e.g. if contamination with cerebrospinal fluid is to be avoided, only 1 µl can be collected from guinea pigs [48].

Sample preparation

As previously described, in untargeted metabolomics, the aim is comprehensive analysis of all small metabolites. For this reason, the sample preparation performed should be as non-selective as possible in order to avoid infusing undue selection bias [49]. In addition, since relative feature intensity is used

to compare the samples, the method used needs to be highly reproducible. Moreover, the sample analysis method to be used must also be considered when choosing the sample preparation method. For instance, when working with NMR, very little sample preparation is necessary, whilst GC-MS often requires more involved procedures such as derivatisation of non-volatile and/or thermolabile compounds [45]. In LC-MS based analysis, derivatisation is generally not necessary but for sample materials containing proteins, these must be removed as they will interfere with the chromatographic separation. Within LC-MS based metabolomics, a commonly applied method for sample preparation is therefore protein precipitation by addition of an organic solvent such as acetonitrile or methanol [50]. In addition to protein precipitation, numerous other methods have been explored such as solid-phase extraction (SPE), liquid-liquid extraction (LLE), and more [44]. Protein precipitation with acetonitrile was the method of choice in all studies included in this thesis as it is relatively non-selective, simple to perform, and provides reproducible results. In contrast, in the case of urine, which does not normally contain proteins, a common approach, when working with LC-MS, is to just dilute the sample, commonly referred to as *dilute-and-shoot* [51].

As mentioned, in untargeted metabolomics, and untargeted analysis as a whole, the assessment of method performance is quite difficult due to the lack of specific target analytes. In targeted analysis, the method is often evaluated using a combination of internal standards and quality control samples that can be evaluated based on the analytes of interest. Since the analytes of interest are most often unknown in untargeted analysis and due to the vast amount of compounds present, it is practically impossible to validate the method for each and every metabolite in the samples [52]. A quality control (QC) sample in untargeted metabolomics is instead often made during sample preparation as a pooled mixture of aliquots from all study samples [53]. This sample is highly useful during the sample analysis and downstream data processing and data analysis to ensure instrumental suitability and assess the quality of the collected data.

Sample analysis

As mentioned previously, the aim of comprehensive profiling of all small metabolites is complicated by the immense diversity of these compounds. No single analytical method exist that can achieve complete coverage of an entire metabolome. This is due both to limitations in the chemical coverage of the separation and detection methods as well as requirements on the dynamic range of the detection. In order to maximize coverage, the approach is generally to analyse the samples using multiple different complementary analytical methods. To that end, the most commonly applied techniques today are nuclear magnetic resonance (NMR), gas chromatography-mass spectrometry

(GC-MS), and liquid chromatography-mass spectrometry (LC-MS). Each of these techniques offer both advantages and disadvantages when contrasted with each other [5,39].

In the early days of metabolomics, NMR was the most commonly used method. NMR represents a natural choice for untargeted metabolomics as it require a minimal amount of sample preparation, is non-destructive, and perhaps most importantly, is non-selectively sensitive to all hydrogen containing compounds [37]. Thus, it fulfils many of the requirements mentioned above. The drawback of NMR, however, in contrast with MS based methods, is that it is less sensitive overall and the slow data collection makes hyphenation to chromatographic separation methods difficult. This generally results in complicated spectra that can be difficult to interpret, which hinders the detection and identification of interesting metabolites [39]. GC, on the other hand, provide chromatographic separation with very high efficiency and, when hyphenated to mass spectrometry (GC-MS), offers very high sensitivity of detection in combination with an abundance of structural information. Furthermore, thanks to the high reproducibility of GC-MS fragmentation between different laboratories and instruments, large scale searchable mass spectral libraries are available, greatly aiding metabolite identification. However, GC, as the name suggests, operates at elevated temperatures with separation based mainly on the vapour pressure of the analytes. Thus, it is limited to volatile and thermally stable compounds or such compounds that can be derivatised to fulfil those requirements [45]. LC, on the other hand, provide a wider range of chemical coverage without requiring derivatisation. In addition, as the selectivity of LC separations is based mainly on chemical interactions with the stationary phase, the selectivity can be drastically altered by changing the column chemistry. The hyphenation of LC to MS is complicated, however, as separation is performed in the liquid phase and mass spectrometry is performed in the gas phase at high vacuum. This means that the effluent need to be evaporated using so called atmospheric pressure ionization methods (API) before MS detection can be performed. The most commonly used API sources today are electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) with ESI being the more common of the two, at least within metabolomics. Within all projects presented in this thesis, LC-ESI-MS was used to analyse the samples.

LC-MS based metabolomics

To achieve a large chemical coverage in LC-MS based metabolomics, the samples are often analysed multiple times, using different separation methods with complementary selectivity along with detection in both positive and negative ionization mode [50]. A suitable and often applied combination of chromatographic methods are reversed phase liquid chromatography (RPLC) and hydrophilic interaction liquid chromatography (HILIC) [54]. In this respect,

RPLC is especially suitable for the hydrophobic compounds and conversely HILIC offers excellent selectivity for more hydrophilic compounds, which often exhibit poor retention in RPLC. Care must be taken, however, to ensure that the instrument delivers stable and reliable results before and during the analysis of the samples. Due to contamination of both the LC system and MS detector when large numbers of injections of biological sample are made, the analysis using LC-MS is highly susceptible to drift in both retention times and MS response during the analytical run [42,53]. While there are recommendations [52,53,55], at the time of writing, no consensus with regards to specific validation criteria are available for untargeted metabolomics. As previously mentioned, the comprehensive monitoring of this drift for all metabolites in a sample is difficult. In addition, the first 5-10 injections performed are usually highly variable and a number of conditioning injections are necessary in order for the instrument to stabilise [42,53]. These are often performed using the QC sample and should not be included in the overall analytical performance evaluation. However, during the conditioning and throughout the analytical run, a selection of metabolites are often monitored in terms of retention time and peak area variation. This approach represent something of a compromise, as the stability of the selected metabolites will not correspond to the stability of every single compound in the sample, but will provide information about the instrument health. In addition to this, a multivariate approach can be used to get a birds-eye view of the overall performance and will be explained further later in this thesis.

Expanding per injection chemical coverage

Compounds with poor retention will generally experience a large degree of co-elution early in the chromatogram which will affect the ionisation and can lead to ion suppression or enhancement in addition to a large degree of variability [56]. This is the reason for combining different separation methods that complement each other as compounds that are poorly retained in one method will then hopefully be retained in another. Thus, an increased chemical coverage can be obtained. However, as mentioned, analysing the samples using multiple different analytical methods mean that the samples need to be analysed multiple times. This increases the need for sufficient sample volume in addition to increasing the instrument runtime which add costs. As such, when the available sample volume is limited, compromises must be made. During the analytical method development in **Paper I**, the highly limited sample volume resulted in the use of only HILIC separation, leaving the lipophilic component of the samples unexplored. It then became obvious to us that the chemical coverage, per injection, needed to be expanded for the, then under planning, follow-up study that is presented in **Paper IV**. To that end, column-switching techniques offer a promising approach and have been applied for a variety of different applications. Examples include online SPE [57–59], serial

[60–63] and parallel [64] coupling of columns, as well as comprehensive two-dimensional LC [65]. These systems can also be applied to sequential analysis, whereby a fraction of the effluent from a first column, containing the compounds of interest, is collected and stored pending conclusion of the first separation. The main issue with this approach, however, when working with HILIC and RPLC, is the incompatibility of the mobile phase solvents in terms of elution strength. The early effluent in a HILIC gradient contains a high proportion of acetonitrile which has a high elution strength in RPLC. Thus, as the idea with a method like this is to collect the unretained compounds in a HILIC separation, the collected effluent would need to be subjected to extensive dilution in order to achieve retention of the collected compounds on an RPLC column. A number of solutions to this problem have been proposed in the past. Notably, Wang et al. [66] and Cabooter et al. [67] both used a large-volume static mixer system, prefilled with diluent, to achieve the necessary dilution of the collected fraction. While mostly effective in solving the problem, the main issue with such a system is that a rather large volume is necessary to achieve sufficient dilution which, among other aspects, takes a long time to empty into the trap column. Additionally, their solution was shown to provide insufficient dilution for the more polar compounds tested, such as caffeine [67]. A different approach was taken by Kittlaus et al. who instead used a secondary binary pump to dilute the effluent and elute the secondary column [68]. More recently, Loos et al. presented another procedure utilising a mixer composed of capillaries with different inner diameter (Figure 2A) [69]. By splitting the flow from the diluent pump between a high resistance (HR, Figure 2A) and a low resistance (LR, Figure 2A) capillary, a flow differential across the two channels could be obtained. Thus, the content of the sample loop, connected to the HR capillary could be slowly displaced into the much higher flow coming via the LR capillary. They show in their paper that a preselected dilution ratio could be obtained by careful selection of the lengths and inner diameters of the two capillaries according to the law of Hagen-Poiseuille. This type of mixer was originally tested in the development work for **Paper II** (see Figure 2A for a schematic representation). However, in order to modify the dilution ratio, one or both capillaries needed to be replaced. For this reason, a more modular approach was designed, whereby one of the binary pumps (Figure 2B) instead was used to slowly displace the contents of the sample loop into a high water flow coming from the diluent pump (Figure 2B). This allowed more control over the dilution ratio without needing to replace components, a benefit that during development work outweighed the drawback of having to utilize two pumps.

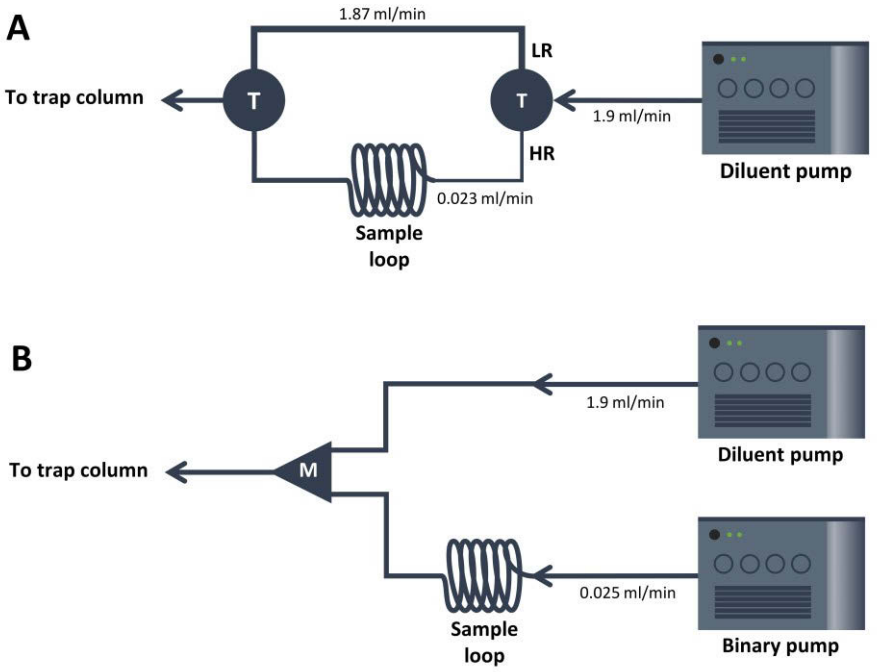

Figure 2 Schematic representations of the mixer designs investigated during the development work of **Paper II**. The first design (A) is based on the paper by Loos et al. [69] and utilises two capillaries with different inner diameter to obtain a flow differential between the low resistance (LR) and high resistance (HR) capillary. Two T-pieces (T) are used to split and subsequently combine the flows. This allows the content of the sample loop to be displaced slowly into the higher flow coming through the LR capillary. The second design (B) instead utilise two pumps to create a flow differential. One of the binary pumps of the system will slowly displace the content of the sample loop into an inline mixer (M) where it is combined with a high flow coming from the diluent pump, thereby diluting the collected fraction.

Using this mixer system for dilution of the HILIC effluent, an automated method was developed for sequential separation of both the hydrophilic and lipophilic component by combining HILIC and RPLC separation with a single injection. This method is presented in detail in **Paper III**. The instrument (Figure 3) is composed entirely of commercially available components, based on an ACQUITY I-class UPLC instrument. The benefit of using commercially available equipment is that the instrumental setup will be available for any laboratory to implement without requiring custom built parts. Furthermore, this solution allows all the components to be controlled from a single software, aiding both method development and sample analysis.

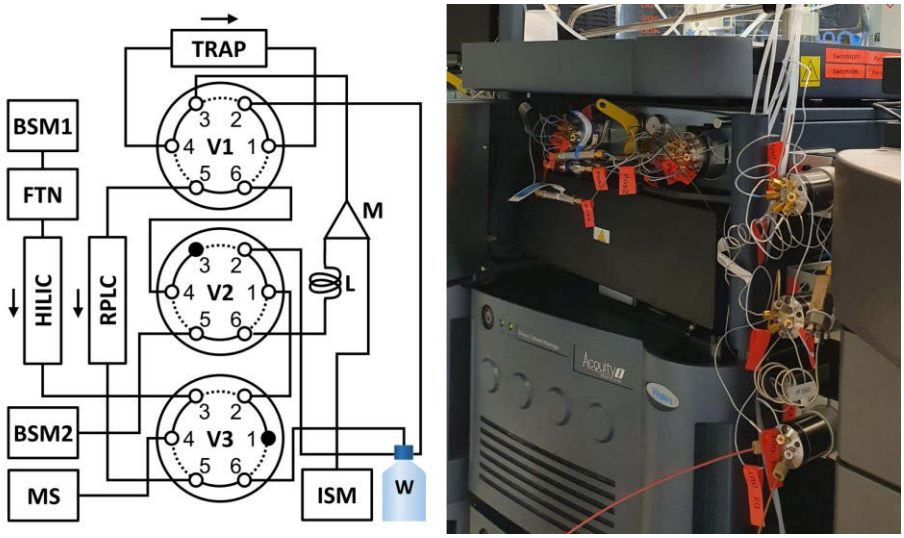

Figure 3 Schematic of the instrument configuration used in the sequential separation method presented in **Paper II** (left) and a picture of the instrument (right) showing the valves, the customized column manger, and couplings. The instrument is composed of two binary pumps (BSM1 and BSM2), an isocratic pump (ISM), a flow-through-needle sample manager (FTN), a custom configured column manager allowing selection of trap column (TRAP), three 6-port VICI valves (V1-3), a 200 µl sample loop (L), a 50 µl in-line mixer (M), and a Synapt G2S HRMS instrument (MS). Reused from **Paper II** with permission from all authors (CC BY).

In the final method (Figure 4), the unretained compounds eluting close to the solvent front in the HILIC separation are collected by diverting the effluent into a sample loop during the time window in which they elute (Figure 4B). While the HILIC separation is being run, the collected fraction can be diluted and loaded into the trap column (Figure 4D). During this step, the acetonitrile-rich fraction collected is simultaneously displaced toward the trap column using a very low flow of 100% water from secondary binary pump (BSM2, Figure 4) and combined in the mixer unit (M, Figure 4) with a high flow of 100% water using the isocratic pump (ISM, Figure 4). By modulating the BSM2 and ISM flow rates, the dilution ratio can be modified based on the situation. In the presented method, BSM2 and ISM flow rates of 0.025 ml/min and 1.9 ml/min, respectively, were used giving a dilution ratio of ca 77 times which was sufficient for this application. When the HILIC separation is finished, the compounds collected in the trap column can be injected onto the analytical RPLC column by back-flushing using a normal gradient with BSM2.

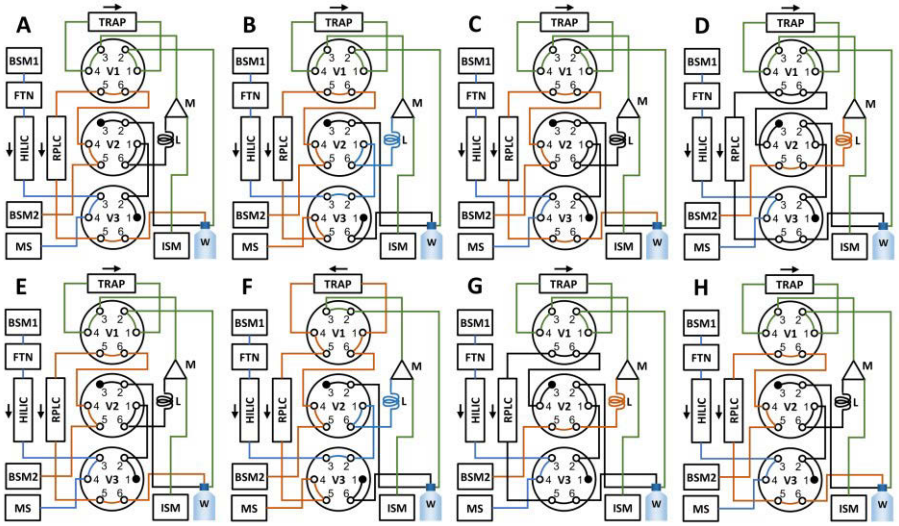

Figure 4 Schematic showing the valve configurations during each step of the sequential separation method. First the sample is injected on the HILIC column (A). At a predetermined time V3 is switched so that the HILIC effluent is collected in the sample loop (B). After collecting the wanted fraction, the HILIC effluent is again directed to the MS detector by switching V3 (C) and loading of the sample loop content to the trap column is commenced by switching V2 (D). When the sample loop has been completely emptied to the trap column, V2 is switched again and the flow of ISM is stopped (E). After the HILIC separation is finished and the RPLC column is equilibrated, V1 and V3 is switched to allow back-flushing of the trap column to the RPLC column and direct the RPLC effluent to the MS detector (F). After the RPLC separation is finished and the HILIC column is equilibrated, the sample loop and trap system is flushed with water (G) and then the instrument is reset for the next injection (H). Reused from **Paper II** with permission from all authors (CC BY).

Unfortunately, chemical compounds cannot be divided cleanly into polar and apolar subsets. The result is that some compounds will exhibit low retention in both HILIC and RPLC, herein referred to as semi-polar compounds. Those compounds within this category that exhibit greater affinity for RPLC than HILIC, such as caffeine and theophylline, offer the greatest challenge in terms of separation efficiency due to the difficulty in trapping and subsequently focusing them at the head of the RPLC column. During the development of the method, which is described in detail in **Paper II**, considerable effort was therefore spent at improving the efficiency of the RPLC separation, especially that of semi-polar compounds.

When selecting the trap column, the main issue was that the selected stationary phase need to afford strong retention not only of highly lipophilic compounds but also the semi-polar ones. A number of different trap column chemistries were investigated, including an Xbridge C8, an Xbridge C18, an Xbridge phenyl, an HSS T3, and an Oasis HLB. For more lipophilic compounds, such as diclofenac, all columns performed well but the only stationary

phase that performed sufficiently well with the semi-polar compounds was the Oasis HLB.

In addition to the trap column, the mobile phase gradients proved to be very important for peak efficiency. In a normal separation experiment, the gradient would be designed, usually with a specific set of compounds in mind, for optimal separation in as little time as possible. In untargeted analysis the method development outcome is harder to define. Moreover, the HILIC gradient also needs to be designed with respect to the unretained compounds as these need to elute in an as narrow band as possible, to minimise the volume of the collected fraction. The work therefore focused on maintaining peak distribution in the HILIC separation while eluting the poorly retained compounds in as short a time as possible. The RPLC gradient also proved to be more important and more difficult to design, than it would have been in a normal experiment. By logical deduction it was initially believed that the majority of focusing occur within the trap column during the loading of the collected fraction (Figure 4D). In an experiment where the RPLC column was removed from the system it was, however, observed that the compounds collected in the trap column eluted as broad bands, as opposed to focused peaks. This lead to the conclusion that the focusing instead occurred at the head of the analytical RPLC column. Here, not only the gradient, but also column temperature and flow rate proved to be important. The screening of these parameters is exemplified in Figure 5 with the separation of the regioisomers theobromine and theophylline. The separation of these two compounds is very challenging as they are highly similar and exhibit relatively low retention in the RPLC separation. Changing the flow rate did not affect this separation to a large extent (Figure 5A), however it did have an effect on the peak efficiencies of other compounds as is described in **Paper II**. For the separation of theobromine and theophylline, optimisation of the column temperature (Figure 5B) and gradient steepness (Figure 5C) had the overall greatest effect. This makes sense if the compounds are actually focused at the head of the RPLC column. In that situation, the compounds need to be eluted from the trap column in as short a time as possible in order to avoid band broadening, and this is also what we observed when developing this method. With a steeper gradient and lower column temperature of the RPLC column, the peak efficiencies, especially for the semi-polar compounds, increased dramatically.

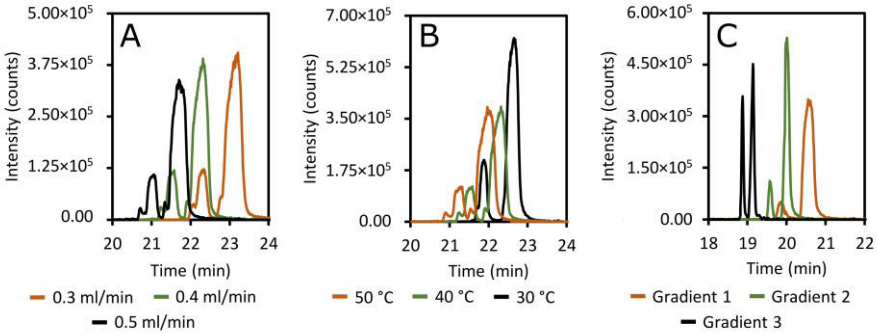

Figure 5 Chromatographic separation of theobromine (earlier eluting) and theophylline (later eluting) during the parameter screening performed for flow rate (A), column temperature (B), and gradient (C). The retention time reported on the x-axis is counted from the sample injection and includes the HILIC separation. The injection of the trapped compounds occurred at 15 minutes. In the screening of gradients, the steepness was increased from gradient 1 to gradient 4. Gradient 1: 0.00–1.00 min 100%A (isocratic), 1.00–5.00 min 100%A–85%A (linear gradient), 5.00–9.00 min 85%A–2%A (linear gradient). Gradient 2: 0.00–1.00 min 100%A (isocratic), 1.00–5.00 min 100%A–70%A (linear gradient), 5.00–9.00 min 70%A–2%A (linear gradient). Gradient 3: 0.00–1.00 min 100%A (isocratic), 1.00–9.00 min 100%A–2%A (linear gradient). Modified from **Paper II** with permission from all authors (CC BY).

Considering that the application of the method was not a targeted protocol, no full validation, according to standard practice within analytical chemistry, was performed. There are, however, important performance aspects that need to be considered if the method is to be used in untargeted metabolomics. As discussed briefly above, these factors mostly condense into ensuring consistent high quality results throughout a long analytical run of biological samples. This is of special importance here due the increased complexity of the system. The presence of additional components and many narrow channels and capillaries, offer new opportunities for failure and performance drift which may severely impact the resulting data. To ensure that the developed method is fit-for-purpose to be used in untargeted metabolomics, the repeatability in terms of both peak intensity and retention time, linearity, back pressure stability, and carry-over was investigated over an analytical run of more than 140 injections of protein precipitated blood plasma. Standard mixtures composed of nine compounds with varying concentration was spiked into the plasma samples, four of which (acetylcholine, acetylcarnitine, serotonin, and phenylalanine) was suitable for separation in the HILIC column and five (caffeine, theobromine, theophylline, diclofenac, and 2-aminobenzoic acid) that were collected in the early effluent and subsequently separated on the RPLC column. The method exhibited excelled repeatability in both peak areas and retention times of the tested compounds in both the HILIC separation (Figure 6A,C) and the RPLC separation (Figure 6B,D). A slight drift in both these metrics is to be

expected over a large number of injections as the sample composition will affect both the chromatographic retention as well as the sensitivity of the mass spectrometer. The first 20 injections shown in Figure 6 are the conditioning injections performed to allow the instrument to reach a steady state in performance. Here a large drift is initially seen which then stabilizes over the last five conditioning injections. Throughout the analytical run, the expected drift is then observed in peak areas as the instrument sensitivity changes over the large number of injections (Figure 6A,B). The retention times of all compounds remain stable through the entire injection sequence in both the HILIC and RPLC separation (Figure 6C,D).

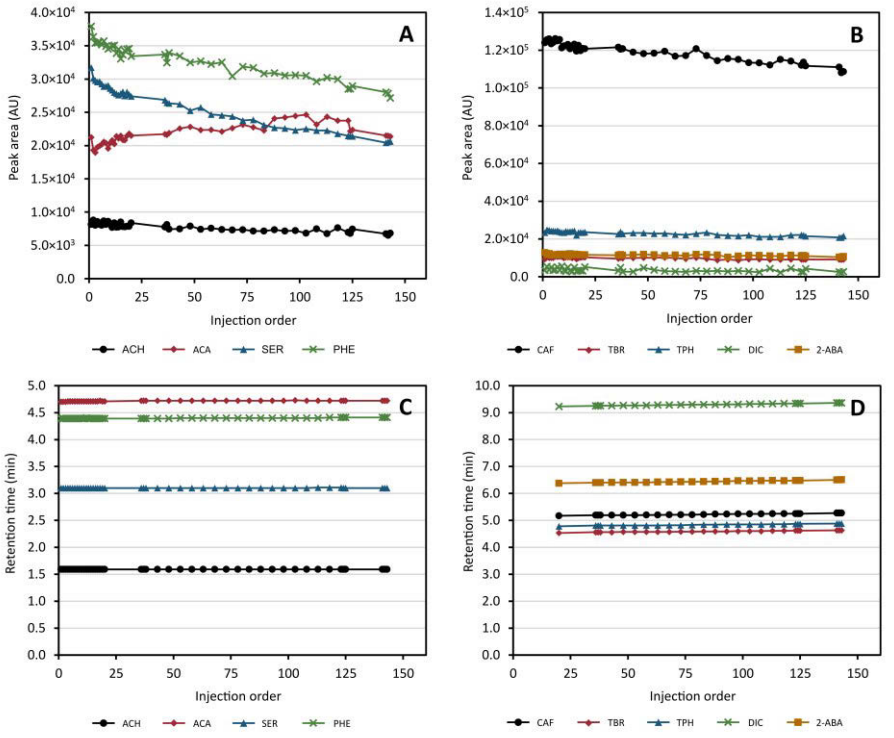

Figure 6 Results from the validation experiments for repeatability in peak intensities (A and B) and retention times (C and D). A moderate drift in peak areas is observed in most of the compounds in the HILIC separation (A) and the RPLC separation (B). The peak areas remain highly stable in the HILIC separation (C) and a very slight upwards trend is observed in the RPLC separation (D). Reused from **Paper II** with permission from all authors (CC BY).

The additional instrumental components such as the valves and numerous narrow capillaries offer ample opportunities for sample material, such as proteins, and salts to accumulate. To ensure that this is not an issue, the back pressure of the pumps was monitored throughout the validation run. In Figure 7 the

back pressure at the time-point of maximum pressure is plotted as a function of injection number for the three pumps used. The back pressures of all the pumps remain highly stable throughout 140+ injections of protein precipitated human blood plasma.

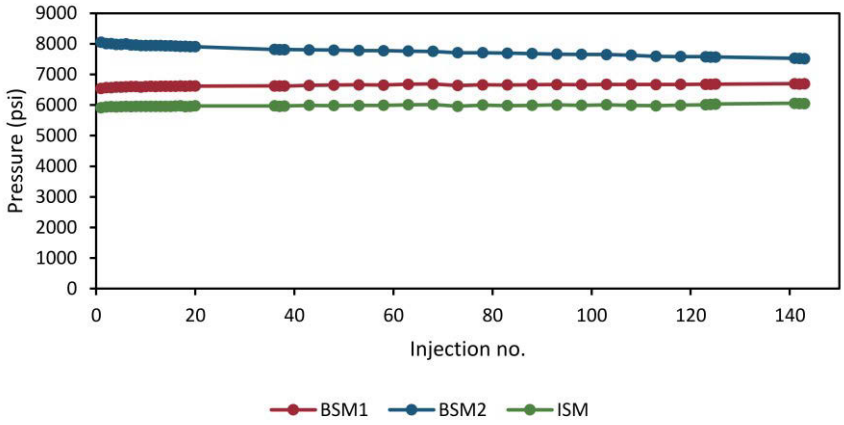

Figure 7 Results from the validation experiments for pump back pressures. The point of maximum back pressure is plotted for the BSM1 pump (red) used to elute the HILIC column, the BSM2 pump (blue) used to load the sample loop and elute the RPLC column, and the ISM pump (green), used to dilute the collected effluent as it is loaded to the trap column. Adapted from **Paper II** with permission from all authors (CC BY).

While the method is not meant to be used for validated quantitative analysis, the linearity of response is still important since the relative responses will be used to discriminate between study groups. Moreover, it serves as an indicator of the trapping performance as quantitative trapping should yield linear response functions. The linearity of the compounds in the standard mixture spiked to three different concentrations into the plasma samples was therefore included in the validation experiments. This is exemplified here with the linear models of theobromine (Figure 8A) and theophylline (Figure 8B) with more examples in **Paper II**.

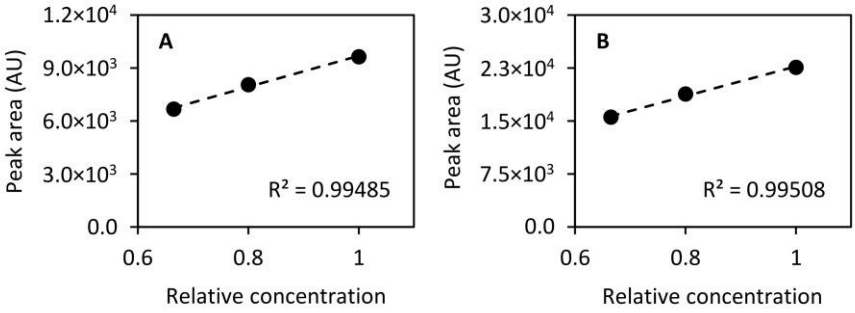

Figure 8 Linearity of response of theobromine (A) and theophylline (B) spiked to different concentrations into human plasma as they are collected from the initial effluent of a HILIC separation and subsequently separated using RPLC in the developed method presented in **Paper II**. Adapter from **Paper II** with permission from all authors (CC BY).

Finally, given the many added components, there was a big risk of carry-over, especially in the trapping components. To ensure this was not an issue, the carry-over was tested during the validation experiments by comparing the peak intensity of each compound against a blank injection. These experiments showed a very low level of carry-over for all tested compounds as described in **Paper II**.

In conclusion, the developed method allows collecting the poorly retained compounds eluting close to the solvent front in a HILIC separation and subsequently separating them on an RPLC column with a single sample injection. The validation performed showed that the method is fit-for-purpose to be used in untargeted metabolomics with biological samples. The method was applied practically to analyse the guinea pig perilymph samples in **Paper IV**.

Data processing

One of the biggest challenges in LC-MS based untargeted metabolomics is the complexity and sheer amount of experimental data generated. Each data point along the time axis, making up the total ion chromatogram, consist of one scan over the range of mass-to-charge (m/z) selected by the operator. When high-resolution mass spectrometry (HRMS) is used, this will amount to millions of mass peaks per experimental file. What is more, the majority of this data (often as much as 80%) holds no information, but originates from instrumental and chemical noise. Thus, before any data analysis can be performed, the complexity of the data needs to be reduced. In targeted analysis, the task is merely that of creating a peak integration method aimed at the target compounds within, for instance, the instrumental software. However, in untargeted analysis, no prior information exist to guide such a peak integration method. Since

the overall aim is the comprehensive profiling of all small compounds in the sample, the task here is to find all chromatographic peaks originating from such compounds. Due to the large amount of data, manual peak detection within any reasonable amount of time is impossible. Instead researchers use various algorithms to automate the process. For this, there is a plethora of proprietary software available such as MarkerLynx (Waters Corp.), MassHunter (Agilent), Progenesys (Waters Corp.), as well as free open-source alternatives such as XCMS [70], MZmine [71,72], and OpenMS [73] to name a few. XCMS has been applied for data processing in all studies included in this thesis. The general workflow of feature detection with XCMS, and similar software, can be summarized in a number of discrete steps: (i) detection of coherent ion traces where chromatographic peaks may be present, (ii) chromatographic peak detection within the identified ion traces, (iii) chromatographic retention drift correction (also referred to as retention alignment), (iv) grouping of peak across samples into so called features (feature detection), and (v) peak filling by correspondence integration in samples where the detected peaks of a feature are missing. All these steps require algorithmic choices and difficult parametrisation that will generally have a profound effect on the outcome.

Before data analysis, it is often necessary to perform drift and batch correction to the data, often referred to as normalisation, in order to remove variation in the feature intensities that is unrelated to biological variation [74]. As the samples are analysed, the response in the mass spectrometer can shift over time. Additionally, due to volumetric variation during sample preparation and sample injection, it is often obvious before correction which sample preparation and/or analytical batch the samples were part of. It has been shown that the choice of normalization method can have a significant impact on the results of a data analysis [74]. Numerous different ways to perform normalization have been proposed in the literature [42,75–77]. In all studies included in this thesis, the method of median fold change, representing one approach of probabilistic quotient normalization (PQN) [78] was used. In this method, the fold change of each feature to that of a representative sample is calculated. The median of the feature-wise fold changes in each sample is then calculated and used as a correction factor for the respective sample.

Finally, before proceeding to analyze the data, the features are generally subjected to a filtering step. This step generally serves to remove low-quality features exhibiting high variation and low intra-group detection rates [52]. The general inclusion criteria used in **Paper I** and **Paper IV** was 80% detection rate in at least one group and relative standard deviation (RSD) <30% calculated across the QC sample injections.

Development of a new algorithm for comprehensive peak characterization and filtering of false peaks

In an effort to not miss important information, the algorithms used for chromatographic peak detection are designed to be as inclusive as possible. This has the inadvertent side effect, however, of sometimes leading to an abundance of false peaks being reported by the software (Figure 9). This is a problem that has been noted by multiple authors in the past [79,80], but no definitive solution has been presented.

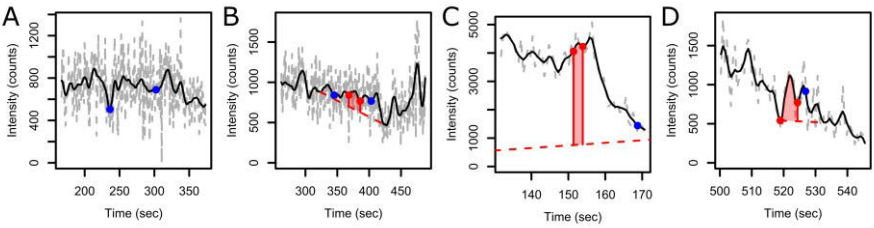

Figure 9 Examples of peaks removed by the CPC algorithm for different reasons. In this example, peaks were filtered if they were not detected (A), had a signal-to-noise ratio less than 10 (B), had fewer than 7 points across the peak (C), or a peak intensity less than 2000 (D). The peak boundaries reported by XCMS are marked as blue dots in the plots. Red dots and red shaded areas are determined by the algorithm described herein. The red dashed line is the baseline determined by the algorithm described herein.

While not in and of themselves a direct problem, these false peaks will affect both downstream data processing and data analysis as these are often based on density and/or parametric modelling. In addition, should these peaks be retained throughout the data processing they may also complicate the already time consuming step of metabolite identification. As an example, one of the feature detection algorithms in XCMS uses a density based approach to find peak groups across samples and may be severely affected by an increased number of randomly located peaks. Thus, the removal of these false peaks from the dataset holds great appeal from the perspective of data quality.

To improve the situation with regards to the false peak reporting by peak picking algorithms, a novel algorithm for comprehensive characterization of the reported peaks using the raw data was developed. The resulting algorithm (comprehensive peak characterization, CPC, Figure 10) is described in detail in **Paper III** and is written in the R and C++ programming languages and made available as the R-package 'cpc'. The premise for the development of the algorithm was that XCMS already has provided a list of peak candidates, which means that the heavy lifting in terms of computation is already done. Thus, a more careful, and thereby potentially slower, peak characterization algorithm, more akin to those used by instrumental software, can be used with

the peaks proposed by XCMS as targets. For this reason, the algorithm was written as an extension to XCMS and is fully integrated into a normal XCMS workflow, allowing easy implementation by the end users. Furthermore, the parameters set for XCMS are often very abstract and can be hard to comprehend as they relate more to the choice of algorithm than to analytical chemistry. In the implementation of CPC, quality metrics familiar to analytical chemists, such as tailing factor and full-width at half maxima (FWHM), are used. The main aim of the algorithm is to allow end-users to filter out false peaks reported by XCMS and thereby improve the data quality overall by minimizing the influence of such peaks on further data processing and data analysis.

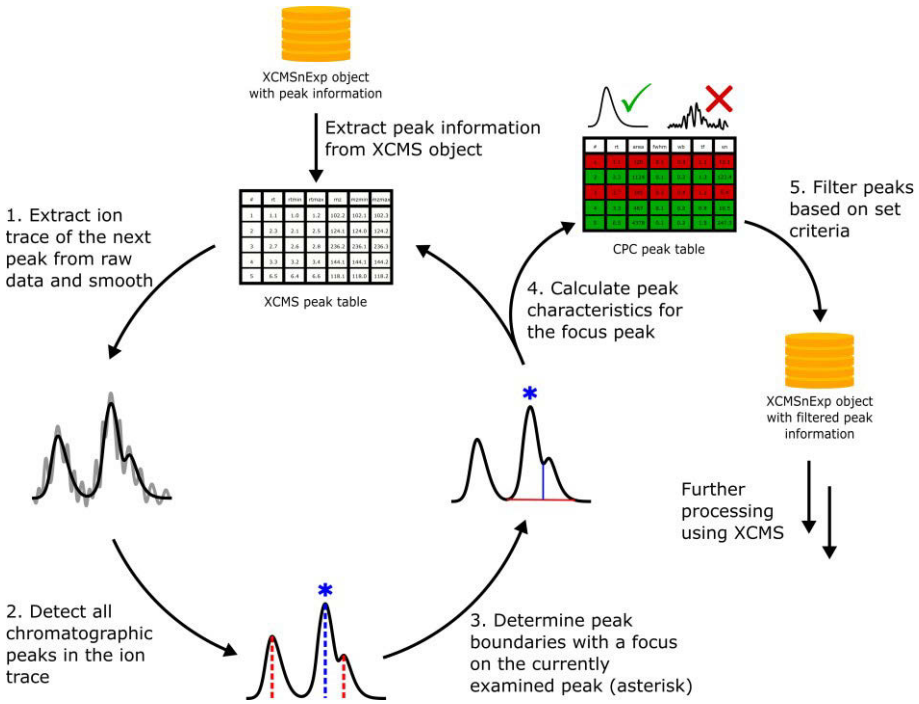

Figure 10 Overview of the CPC algorithm as integrated into an XCMS workflow. The algorithm works on an XCMSnExp object returned by XCMS. Initially, the peak information reported by XCMS is extracted from the XCMSnExp object. For each peak, the raw extracted ion trace (XIC) is retrieved from the raw data files. The XIC is then smoothed (step 1) and subjected to an adaptation of the ApexTrack [81] algorithm (steps 2 and 3) including (i) detection of all peaks using the second derivative, (ii) identification of the target peak as specified by XCMS, (iii) characterization of the peaks in the XIC, and (iv) calculation of the peak characteristics (step 4) using the baseline expansion algorithm presented in ref [81]. The calculated characteristics are then used to filter peaks that do not meet the user-specified criteria (step 5). The output from the algorithm is a new XCMSnExp object containing only those peaks that meet the criteria that can be used for further processing with XCMS. Figure adapted from **Paper III** with permission from all authors.

For each peak, the algorithm uses the m/z value and retention time reported by XCMS to generate an extracted ion chromatogram (XIC) from the raw data files. The XIC is then (optionally) smoothed using either a moving mean or Savitzky-Golay [82] smoother according to the user specified parameters (step 1, Figure 10). After smoothing, the second derivative trace of the XIC is determined. All chromatographic peaks along the XIC are then detected by searching for peak signature regions in the second derivative which manifest as negative minima cradled by inflection points (step 2, Figure 10). To determine the peak boundaries for the detected peaks (step 3, Figure 10), a baseline expansion algorithm adapted from the ApexTrack algorithm [81] is used (Figure 11). In short, the algorithm assumes an initial baseline (Figure 11A) set between the corresponding inflection points in the second derivative. The slope differences between the slope of the XIC at the front and tail scan points and the slope of the initial baseline is calculated. Slope difference thresholds for the front and tail baseline bound are then calculated using the liftoff and touchdown parameters supplied by the user as percentage of the initial slope differences. The algorithm then proceeds to expand the front and tail baseline bounds giving a new baseline in each iteration (Figure 11B), until the slope differences at the new baseline bounds subceeds the thresholds for both the front and tail bounds (Figure 11C) yielding the final baseline bounds for the peak.

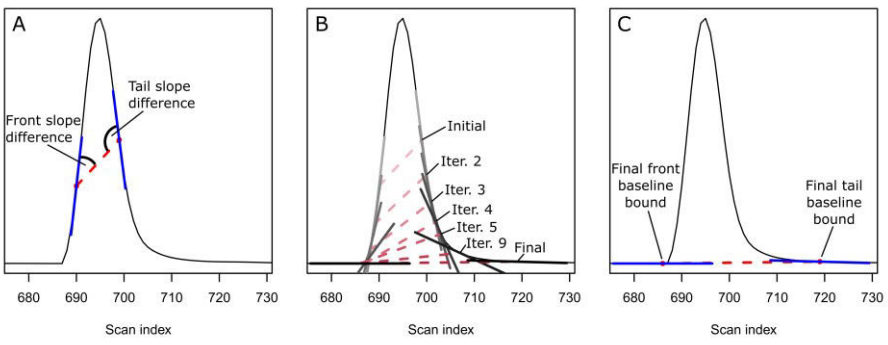

Figure 11 Illustration of the peak expansion algorithm adapted from the ApexTrack algorithm [81] in **Paper III**. An initial baseline is set between the second derivative inflection points of the front and tail (red dashed line in A) and a front and tail slope difference threshold is calculated based the slopes at the inflection points (blue lines in A) and the slope between the inflection points (initial baseline). These thresholds are used as break points during the expansion (B) which expands the baseline by one scan point per iteration until the absolute difference between the slope at the front and tail bounds and the current expanded baseline is met (C). Figure adapted from **Paper III** with permission from all authors.

Co-eluting peaks can be detected after baseline expansion as they will have overlapping baseline bounds. Three types of co-eluting peak bounds are recognized in the algorithm (Figure 12); valley bounds (Figure 12A), shoulder bounds (Figure 12B), and a special form of shoulder bounds in which the peaks are almost completely overlapping and of similar height (not shown), referred to herein as rounded peaks. The rounded peaks differ from the shoulder peaks in that no inflection point is present between the peak apices along the second derivative. When co-eluting peaks are detected, they are added to clusters, and a second round of baseline expansion is performed on the entire cluster using the outermost inflection points as starting points to obtain the baseline bounds for the entire cluster. The bounds between the peaks are then set depending on the type of boundary; for valley bounds the lowest point between the two peaks is used (Figure 12A), for shoulder bounds and rounded peaks, the highest point along the second derivative between the two apices is used (Figure 12B and Figure 12D).

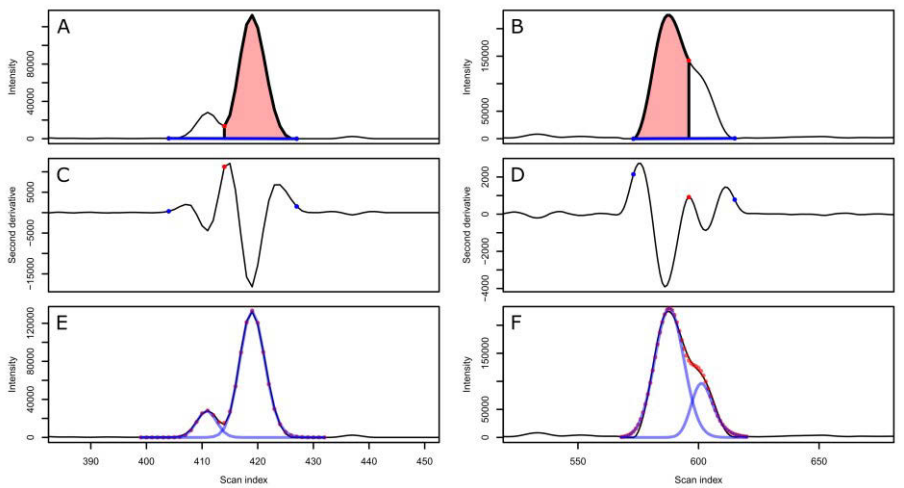

Figure 12 Examples of peak clusters with two different peak bounds, valley bounded (A, C, and E) and shoulder bounded (B, D, and F). In both cases, both peaks are detected as negative minima in the second derivative (C and D), cradled by inflection points (zero crossings). When valley bounds are detected, the peak bound of the valley is set to the lowest intensity point along the ion trace between the two peak apices (red point in A). In the case of shoulder peaks, the lower intensity peak does not have an apex. Here, the peak bound is set to the highest point along the second derivative instead (red point in D). The third type (not shown here) occur when two peaks with similar intensity co-elute almost completely resulting in a rounded peak shape and can be detected as they result in two peaks that share the same inflection points. If the true peak bounds of peaks in clusters are needed, peak deconvolution can be performed using least-squares fitting of exponentially modified Gaussian functions (E and F). Figure adapted from **Paper III** with permission from all authors.

As the method used for peak characterization in the developed package works directly on the second derivative curve, co-elution disturbs the detection of true peak boundaries. This is not an issue in peak filtering, however when peaks are to be characterized for other usage, it may be interesting to obtain the real peak boundaries. In these cases, it is possible to apply peak deconvolution on the selected peaks by least squares fitting of an exponentially modified Gaussian (EMG) function [83,84] (Figure 12E-F).

The algorithm is described in detail in **Paper III** where it is also applied to a real metabolomics dataset. In the presented example roughly 35% of the peaks reported by XCMS were removed. Interestingly, the majority of removed peaks were filtered due to having a too low signal-to-noise ratio. An example of such a removed peak can be seen in Figure 9B. An overall observation from viewing a large number of peaks removed by CPC is that many of the peaks, even when not removed due to low signal-to-noise ratio, still appear to be just noise. The algorithm was also applied to the data processing in the study presented in **Paper IV**. Here, similar numbers were observed in the HILIC data where 35% and 34% were removed in the positive and negative data, respectively. In the RPLC data the numbers were even higher and 54% and 57% were removed in the positive and negative mode data, respectively. Also here, the primary reason for filtering was low signal-to-noise ratio. Based on these results, it seems that the determination of signal-to-noise ratio used in XCMS is too optimistic since the signal-to-noise parameters (snthresh) in XCMS was set to the same value as that used in CPC (10).

In summary, an algorithm for comprehensive peak characterization was developed to address the issue of false peaks reported by software like XCMS. The algorithm uses methods and parameters familiar to analytical chemists, both for the peak characterization as well as the quality metrics used to filter the peaks. The algorithm is fully integrated into the XCMS workflow and made available as the R-package ‘cpc’.

Data analysis

The data analysis goal in metabolomics is to find metabolites with intensities that correlate with the biological question being studied. As previously mentioned, the aim could be either to find biomarkers that can be used for early diagnosis and/or hypothesis generation to learn more about the metabolic processes underlying a disease or treatment. In **Paper I** and **Paper IV**, the focus was on the protective effect of hydrogen gas on NIHL in guinea pigs. In **Paper I**, the animals were exposed to high intensity impulse noise to induce hearing loss, a subset of animals were then subjected to hydrogen gas by inhalation as an attenuating treatment. In contrast, in the study described in **Paper IV**, a broad band continuous noise was used instead. Thus, the data analysis performed was aimed at finding metabolites that can discriminate between those

animals that received hydrogen gas following noise exposure (*Noise+H2*) and those that did not (*Noise*).

Before multivariate data analysis, the data are generally mean-centered, scaled and transformed. The purpose of scaling is primarily to equalise the importance of the absolute intensities of the features and transformation is applied to correct for heteroscedasticity in the noise of the feature intensities. There are recommendations for which type of scaling to use in difference circumstances [85] and among the most commonly used are auto scaling (also called unit variance scaling), pareto scaling, and range scaling [85]. In auto scaling, the intensity of the features are divided by their standard deviation and in pareto scaling the square root of the standard deviation is used instead. The main difference in the outcome is that autoscaling will make the entire range of peak intensities equally important in the models which may increase the influence of noise, an effect which is reduced when applying pareto scaling. In range scaling the variables are divided instead by the intensity range. All these methods will, to an extent, equalise the importance of small and large fold changes in the dataset. Failure to perform such corrections may lead to missed discoveries as without scaling, features with high intensities will be given much higher importance in models than features with low intensities [85]. Moreover, if the noise profile of the data is heteroscedastic, which is often the case in LC-MS data, multivariate modelling methods that assume homoscedasticity may be adversely affected [86].

Multivariate analysis

Since thousands of features might be detected in the data, performing univariate testing on all features is not recommended as it may lead to spurious results. Moreover, they allow only single feature differences to be detected between the groups without consideration of feature interactions. Instead, various multivariate methods are generally applied instead including principal component analysis (PCA) [87], orthogonal projection to latent structures discriminant analysis (OPLS-DA) [88], and Random Forest discriminant analysis (RF-DA) [89]. Often a combination of modelling methodologies are applied to the data.

PCA is an unsupervised projection method that reduces the dimensionality of the high-dimensional dataset (**X**) under analysis [90]. The first component of a PCA model describes the largest variation in the data and each successive component is linearly orthogonal to the previous components calculated. In this way a PCA model can describe the variation in a dataset with many variables using the linear combination of only a few principal components. The sample scores can be used to visualise the relationships between the samples in this lower dimensional space while the loadings gives information on variable importance in each component. This makes PCA very useful in untar-

geted metabolomics for visualising systematic variation in the data. For instance, PCA is very useful for quality control purposes as it allows outlying samples and variables to be detected in addition to assessing residual unwanted variation originating from batch effects and performance drift. As an example of this, in **Paper I** PCA was used during the conditioning to evaluate the stability of the instrument performance. Moreover, PCA was used in all papers presented in this thesis both for quality control and to visualise structures with respect to the study groups. In **Paper I**, a clear difference between the study groups related to the administration of hydrogen gas in noise exposed guinea pigs could be seen (Figure 13). The individuals in that study were divided into the groups *Control* (n = 2) which received no intervention, *H2* (n = 4) which was not exposed to noise but was given hydrogen gas, *Noise* (n = 10) which was exposed to noise but received no hydrogen gas, and *Noise+H2* (n = 10) which was exposed to noise and was subjected to hydrogen gas. Additionally, perilymph samples were taken from both the left (Figure 13A) and right ear (Figure 13B). In both the left and right ear samples, the *Noise* samples deviate from the other groups in the second component, whereas the *Noise+H2*, *H2*, and *Control* samples are mixed together. This indicates a systematic difference in the perilymph metabolome of the *Noise* animals which is mitigated by the hydrogen gas in the *Noise+H2* animals.

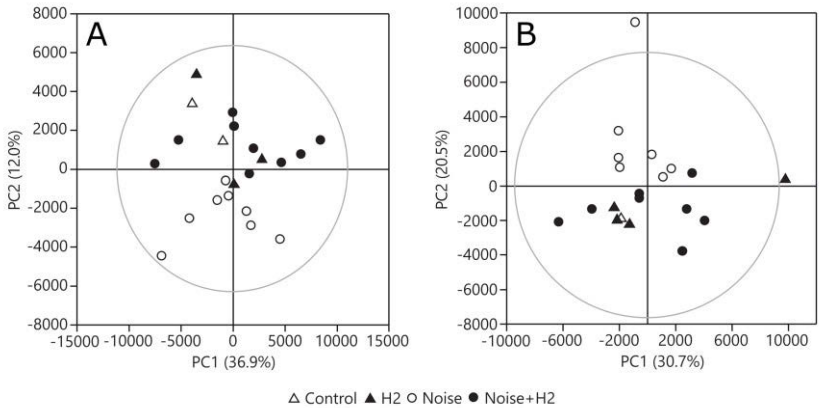

Figure 13 PCA scores plots from models fitted on data acquired from the left (noise exposed, A) and right (B) ear perilymph samples of noise exposed guinea pigs presented in **Paper I**. Four groups were included in the model; *Control* that received no intervention (open triangles), *H2* that were not exposed to noise but subjected to hydrogen gas (filled triangles), *Noise* which were exposed to noise but not subjected to hydrogen gas (open circles), and *Noise+H2* which were both exposed to noise and subjected to hydrogen gas (filled circles). The study focused on the attenuating effect of hydrogen gas on noise-induced hearing loss in guinea pigs exposed to impulse noise. Adapted from **Paper I** with permission from all authors (CC BY).

Like PCA, OPLS-DA is a projection method, however, in contrast to PCA, OPLS-DA is a supervised method. This means that, as opposed to modelling only the variation in **X**, as is done in PCA, instead the covariance of **X** and an outcome variable (**Y**), in our case group membership, is sought. As the name suggests, OPLS-DA is a development of projection to latent structures (PLS) where the variation in **X** that is uncorrelated, or orthogonal, to **Y** is removed. This allows easier interpretation of the models as there is only one, so called, predictive component for each **Y** variable. OPLS-DA was used in **Paper I** as a search tool to find metabolites with discriminant intensities between the study groups, and therefore would be correlated with the treatment with hydrogen gas. In combination with PCA, OPLS-DA is a very powerful tool for discovering such discriminant metabolites. However, the sensitivity of OPLS-DA makes it prone to overfitting and the models need to be carefully validated. For that reason, the models fitted in **Paper I** were validated using a combination of different methods. First off, permutation testing was performed where the order of the **Y** variable is randomized. If a true correlation is present, the reordering of **Y** should most of the time yield weaker models. Secondly, the models were subjected to a so called leave-one-out cross validation where one sample is kept out when the model is fitted and the fitted model is then used to predict the group membership of the held-out sample. A correct classification rate can then be determined based on the classification results from each round of cross validation. From the models presented in **Paper I**, variables were selected based on the S plot (Figure 14B,C) and the shared-and-unique-structures (SUS) plot (Figure 14A) [91]. The S plot shows the correlation to group membership as a function of the predictive loadings and takes on an S form when the data is pareto-scaled. This is very useful for finding highly important variables. In the SUS plot the correlation in two models is plotted on the x and y axis. This allows easy assessment of differences in variable importance between models. In Figure 14, the models fitted on the left (noise exposed) and right ear data were compared using the SUS-plot. The symbols are based on the model in which the feature showed importance with circles being common to both models, triangles showing importance only in the right ear models and squares showing importance in only the left ear models.

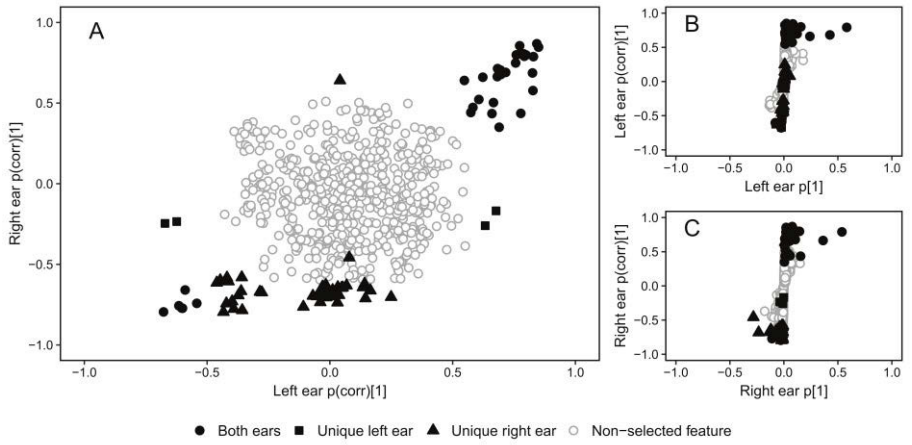

Figure 14 The shared-and-unique structures (SUS) plot (A) and S-plots (B and C) from two OPLS-DA models of the *Noise* group versus the *Noise+H2* group, fitted on data from the left (noise exposed) ear (x-axis in A, and B) and right ear (y-axis in A, and C) perilymph samples. The SUS-plot is a scatter plot of the variable correlation to group classification, and the point symbols represent the relationship of the variables between the models. In the SUS-plot, variables with similar correlation in both models end up in the lower left and upper right corners (filled circles), variables with opposite correlation between the models end up in the upper left and lower right corner (none present), and variables with correlation in only one of the models end up on the respective axis (filled squares with correlation in only left ear samples and filled triangles with correlation in only right ear samples). Variables with low correlation in both ears end up in the middle of the plot (open circles). Reused from **Paper I** with permission from all authors (CC BY).

RF-DA is different from the projection methods previously described [89]. PCA, OPLS, and similar methods are parametric methods akin to linear regression. RF-DA is instead based on building large numbers of decision trees with random selection of variables in each tree. This should, in theory, be less susceptible to overfitting than, for instance, projection methods. An issue with using RF-DA for classification is the difficulty in interpreting the classifier in terms of important variables making variable selection difficult. In **Paper IV**, RF-DA was used to discover discriminant metabolites. The study was a follow-up of the one presented in **Paper I** where instead of impulse noise, the animals were exposed to a broad-band continuous noise. To fit the models, the MUVR algorithm [92] was used which will continuously reduce the number of variables while monitoring the number of misclassifications until an optimised number of variables remain. At each iteration, the variables are ranked using the variable importance metric *mean decrease in Gini index* [89,92]. The model performance was assessed using a combination of the number of misclassifications, the area under the receiver operating characteristic curve (AUROC) of the classifier, and a permutation test p-value like that used in

OPLS-DA. Only variables from classifiers with sufficient strength were considered further.

In **Paper IV** volcano plots were used in combination with RF-DA, for feature selection. In volcano plots, the negative base-10 logarithm of the p-values from a t test is plotted as a function of the base-2 logarithm of the fold changes in the within-group mean intensity [93]. Plotted this way, it provides a combined analysis of both the absolute change in variable intensity (fold change) between the groups in addition to accounting for the variance within the groups (p-value). This method for variable selection is commonly used in mRNA microarray analysis [93] but is not as commonly used in metabolomics. A drawback to this is that it does not allow variable interactions to be analysed, as the projection methods do. From the volcano plots, a total of 258 unique features were selected from the positive and negative mode data. Those RF-DA models that exhibited high model performance metrics were fitted primarily between the time points in the *Noise* animals and features were only selected from such models. In total, the selected features from the combined use of volcano plots and RF-DA models resulted in six individual interesting metabolites.

Metabolite identification

Often, a large number of features are selected during the data analysis step. Each feature give information about the m/z value of the ion detected as well as a retention time. This can be used to match features originating from the same metabolite using software such as the CAMERA R package or similar [94]. However, the work is sometimes complicated by the false peaks discussed previously, which may sometimes also lead to spuriously correlating features which need to be filtered out. This, together with the immense number of compounds present in biological samples with both endogenous sources as well as exogenous compounds originating from the environment, make metabolite identification very time consuming work. In fact, metabolite identification still remain one of the main bottlenecks in untargeted metabolomics.

For well characterized biological systems, such as human blood and urine, the work is alleviated to some extent by prior knowledge. There are a number of online databases that can be used as reference material when trying to identify unknown metabolites. Good examples of such databases, most of which were used to successfully identify a number of compounds in both **Paper I** and **Paper IV** is the human metabolome database (HMDb) [2,38,95], KEGG [96], PubChem (https://pubchem.ncbi.nlm.nih.gov/), and METLIN [97]. In HMDb and METLIN, the m/z values detected can be matched to different adducts for known metabolites, and often reference MS spectra are available that can be used, to some extent, for confirmation. In addition, there are also MS spectral databases available, such as MassBank [98], often linked to from

HMDb and METLIN, which contain such reference MS spectra for large numbers of metabolites. However, when working with LC-MS/MS, as previously mentioned, the fragmentation patterns obtained between instruments can differ which makes identification using generalised databases difficult. None of these methods are sufficient for a confirmed identification, and even when a strong candidate is found, it can often be difficult to determine the exact identity of various regioisomers. It is therefore important to report the confidence of the identification. Some different models of this have been proposed which are all based on similar ways of thinking [37,40,99,100], for example that presented by the metabolomics standardization initiative (MSI) [40]. In both **Paper I** and **Paper IV** this method was used, which define a four level confidence scale. Increasing level correspond to decreasing confidence in the identification: (1) confirmed identification, which can only be made with authentic standards analysed under the same conditions and at least two orthogonal measurements, such as retention time and the reference mass spectra, (2) putatively annotated compounds identified using spectral library searches or similar methods, (3) putatively characterized compounds where only the compound class can be determined based on for instance characteristic fragments, or (4) unknown compounds for which only retention time, m/z value and/or putative molecular formula is presented. For such well-characterized metabolomes as described above, metabolites and even exogenous compounds may be readily available as authentic standards, however, for less well-known biological systems this can be very difficult to obtain. In **Paper I**, four out of fifteen compounds could be positively identified (MSI level 1) using authentic standards. In that work, a mixture of authentic standards was analysed in sequence with both a spiked QC sample where the same mixture had been added and the original QC sample. In **Paper IV**, two out the six metabolites selected during the data analysis could be putatively annotated (MSI level 2), however none were positively identified using authentic standards.

Biological interpretation

The final part of the analytical chain of untargeted metabolomics is to put the samples into the context of the biological question being studied. For instance, in **Paper I** and **Paper IV**, the mitigating effect of hydrogen gas on NIHL was being studied. Thus, it was important to discuss the role of the discovered metabolites in relation to the biological processes underlying NIHL. When exposed to intense or prolonged noise, depending on the intensity of the noise, NIHL can be caused by a combination of direct mechanical damage brought on by very intense noise as well as metabolic decomposition [101,102]. In the latter case, the process involves the formation reactive oxygen species (ROS) in the cells. The occurrence of ROS have been attributed to a few different pathways leading to the formation of superoxide inside the mitochondria

[101]. For example, when the mitochondria are overdriven by the need for excess amounts of energy due to the noise stimulation, it leads to excessive levels of superoxide formed which may leak into the cell where it can react with other compounds to form more potent ROS. In **Paper I**, multiple acylcarnitines were identified with higher levels in the *Noise* group in contrast with the *Noise+H2* and control groups not exposed to noise. In the cell, acylcarnitines are involved in the transportation of fatty acids across the mitochondrial membrane. This indicates an increased oxidative stress which was attenuated by the hydrogen gas treatment. In addition, two osmoprotectants, previously not described in this context, stachydrine and homostachydrine were shown to be present in much lower levels in the *Noise* group in both ears of the animals. Osmoprotectants serve to regulate the cell volume in response to osmotic pressure [103]. As an example, stachydrine has been described as an osmoprotectant for *E. coli* and other bacteria [104]. This indicates a possible osmotic response, not present when the animals are subjected to hydrogen gas. In the follow-up study presented in **Paper IV**, these compounds were not implicated as important, however, another osmoprotectant, glycerophosphorylcholine (GPC), was detected with higher levels in the animals not subjected to hydrogen gas. GPC is also a known antioxidant and has been shown to have a positive influence on oxygen consumption in the mitochondria [105]. In addition to the formation of ROS, NIHL is also connected to disruptions in the potassium transportation which is very important for hearing function [106]. Disruption of this transportation system and related genes have been linked to hearing loss and deafness in multiple species [106–108]. In **Paper IV** potassium, detected as a potassium formate cluster [109–112], was also linked with hydrogen gas treatment. A lower level of potassium ions in the *Noise* groups could indicate reduced potassium recycling as a response to the noise exposure which was attenuated by the hydrogen gas treatment.

To summarize, the metabolite identification and biological interpretation of untargeted metabolomics data is a difficult and time consuming work. It is also a very important aspect of any metabolomics study when the aim is to produce further knowledge of the metabolic processes involved. In both **Paper I** and **Paper IV**, metabolites with discriminant capability were detected in perilymph taken from guinea pigs exposed to noise with and without hydrogen gas treatment to study the effect of such treatment.

Concluding remarks

Untargeted metabolomics is the field that aims to analyse all small metabolites in a biological system in order to answer questions of biological origin. Twenty years in, the field is still full of challenges, owing to the ambitious nature of this goal. Achieving the comprehensive analysis of all small molecules in a biological system remains an elusive dream, but advances toward this goal have been made. These advances are mainly fuelled by improvements to analytical instrumentation and data processing algorithms. In short, each step in the analytical chain of an untargeted metabolomics study need to be carefully controlled if data of high quality is to be obtained that can be used for downstream biological inference. When the amount of available sample material is limited, the acquisition of wide coverage, high quality data is even more challenging. In addition, due to the vast and incredibly complex data generated in these studies, especially when using LC-MS, researchers employ a range of algorithms to enable the timely processing and analysis of the data. Each such algorithmic step adds choices and parametrisation that researcher has to make which will have a large impact on the end result. In this thesis, four papers are presented which relate to the field of untargeted metabolomics. The overarching goal during this doctoral project was to improve upon and develop new methods to be used in untargeted metabolomics studies. To this end, focus was on improving chemical coverage in LC-MS based untargeted metabolomics in addition to increasing data quality through improved data processing.

In **Paper I**, the development and application of an untargeted metabolomics protocol for guinea pigs perilymph was described. The protocol was applied in a study of the attenuating effect of hydrogen gas on noise-induced hearing loss (NIHL) in guinea pigs exposed to impulse noise. In the study, clear differences in the perilymph metabolome was observed between noise exposed animals subjected to hydrogen gas as a treatment for NIHL and those who were not. Differences in the abundance of a number of metabolites were observed between animals exposed to noise with and without subsequent hydrogen gas inhalation, for instance the levels of both acylcarnitines and two osmoprotectants. Additionally, the protective effect of hydrogen gas was underlined by the similarity of the metabolome in animals exposed to noise and subjected to hydrogen gas and the control animals which had not been exposed to noise. During the study, two primary development areas were recognised.

First, the highly limited volume of available sample material (1 µl) enforced compromises in the number of analysis that could be performed. This resulted in the use of only HILIC separation which means that the lipophilic component of the samples was never analysed. Thus, the chemical coverage in such studies needed to be increased. Secondly, numerous features were selected from the models, which upon inspection turned out to be only noise and baseline shifts, reported by the peak picking software as chromatographic peaks. Such peaks, not only affect the downstream data analysis and complicate metabolite identification, as mentioned, but also affect the data processing methods as they are often based on parametric models. As such, removal of such false peaks from the data after the peak picking step is necessary.

In **Paper II**, the development and validation of a separation method for the single-injection sequential analysis of both the lipophilic and hydrophilic compounds was described. HILIC separation offer excellent selectivity for hydrophilic compounds, however compounds with lower polarities exhibit very poor retention in such separation systems. These compounds often experience high levels of co-elution which may cause severe ionisation effects such as ion suppression or enhancement and this data is often discarded prior to data analysis. For this reason, a combination of HILIC and RPLC is often used when analysing samples in LC-MS based untargeted metabolomics. Lipophilic compounds with poor retention on HILIC often exhibit excellent retention in RPLC. In the developed method, the early eluting compounds in a HILIC separation was collected and focused on a trap column and subsequently separated on an RPLC column after the HILIC separation was concluded. The method showed excellent performance in terms of repeatability and robustness during the validation experiments and was applied to the analysis of guinea pig perilymph samples in a follow-up study presented in **Paper IV**.

The second development area was the reporting of false positive peaks by the peak picking software used. This sparked the development of a peak characterisation algorithm that could be used to filter such peaks, in addition to allowing researchers to use peak characteristics more familiar to analytical chemists to assess the quality of peaks. The algorithm was developed using a combination of the R and C++ programming languages and made available as the R-package ‘cpc’, and is fully integrated it into a normal XCMS workflow. During the testing and validation of the algorithm, it quickly became obvious that a large proportion of the peaks reported by XCMS were of very low quality with low signal-to-noise ratios and overall peak intensity. When applied to real metabolomics data, around 40-50% of the reported peaks are generally removed, depending on the parameter settings used.

In **Paper IV**, the application of the chromatographic separation and data processing methods described in **Paper II** and **Paper III**, to the analysis of guinea pig perilymph samples taken from a follow-up study to that described in **Paper I**, is presented. In this study of the attenuating effect of hydrogen gas

on NIHL, the guinea pigs were exposed to broad band continuous noise, analogous to industrial noise, in order to induce NIHL. Following noise exposure, a subset of animals were subjected to hydrogen gas through inhalation. Perilymph samples were taken at three different time points, two hours after exposure, one week after exposure and two weeks after exposure. Six chemical species were detected with differential trends between the animals subjected to hydrogen gas and those which were not. Among these potassium was identified with higher abundance in the animals treated with hydrogen gas than those that were not, indicating some disruption of the potassium recycling. This system is central to normal hearing function. In addition, glycerophosphorylcholine (GPC) was also identified with higher abundance in the hydrogen gas treated animals. GPC is known to be involved in the osmotic regulation of renal cells, in addition to its role as an antioxidant. This is of particular interest as one of the main causes of NIHL is oxidative stress induced by the overstimulation of the sensory cells of hearing.

To conclude, two of the included papers in this thesis (**Paper I** and **Paper IV**) describe applications of untargeted metabolomics to the study of hydrogen gas as a treatment for NIHL. Issues identified during the analytical work of **Paper I** sparked development presented in **Paper II** and **Paper III** with focus on the chromatographic chemical coverage and the performance of the data processing methods used.

Populärvetenskaplig sammanfattning

Levande organismer är uppbyggda av celler som sköter om alla de funktioner som behövs för att upprätthålla liv. Dessa funktioner styrs av kemiska processer som involverar allt från vårt DNA, RNA, våra proteiner och små molekyler, även kallade metaboliter. Då en organism, såsom en människa, till exempel blir sjuk, äter mat, tar läkemedel eller utsätts för någon annan yttre eller inre påverkan, ger det följdeffekter hos alla dessa processer i olika utsträckning. För att till exempel kunna utveckla bättre behandlingar mot sjukdomar eller ställa tidigare diagnoser, är forskare ofta intresserade av hur sjukdomen påverkar de olika delarna av vår biokemi. Inom forskningsfältet metabolomik fokuserar forskarna på de små kemiska ämnena, så kallade metaboliter, hos organismen. Målet är då ofta att bestämma hur mycket av alla dessa metaboliter som finns närvarande i ett prov för att på så sätt kunna jämföra till exempel friska individer med sjuka, eller för att kunna se skillnader som uppstår av en läkemedelsbehandling. Man kan dela in fältet i sådana studier som söker efter biomarkörer, till exempel för att påvisa sjukdom, och sådana som kallas hypotesgenererande, där man snarast söker efter nya bättre frågor att ställa i vidare studier. Dessutom kan studier vara endera riktade, där man tittar på specifika metaboliter, och oriktade där man försöker bestämma förändringar i mängden av så många metaboliter som möjligt. I den här avhandlingen diskuteras oriktad metabolomik för hypotesgenererande studier.

En svårighet i oriktad metabolomik är att forskarna på förhand inte vet vad de letar efter. Det kan finnas tusentals till hundratusentals metaboliter i ett biologiskt system som alla har olika kemiska egenskaper och en stor bredd i koncentration. Med andra ord behöver de analytiska metoderna som används vara väldigt generella för att man ska kunna se så mycket som möjligt. Det finns idag inga enskilda metoder som klarar av att analysera alla metaboliter utan normalt kombineras olika metoder för att på så sätt öka bredden. Det finns en rad olika analysmetoder som användas men den bredaste tekniken är vätskekromatografi kopplat till masspektrometri (LC-MS). Inom LC-MS separeras först substanser baserat på deras interaktioner med packningsmaterialet i en kolonn, ju starkare interaktioner, desto längre kommer en substans att stanna kvar i kolonnen, kallat retention. Beroende på vilken typ av kolonn som används kan olika klasser av substanser separeras från varandra och genom att kombinera användandet av olika typer av kolonner kan man öka bredden i sin analys. De vanligaste kolonntyperna inom metabolomik kallas för HILIC

(hydrophilic interaction liquid chromatography) och RPLC (reversed-phase liquid chromatography) kolonner. HILIC är en bra typ av kolonn för att analysera polära substanser medan RPLC kolonner lämpar sig bra för opolära substanser och tillsammans ger de därför en bra täckning. Efter separationen används masspektrometern som en teknik för att väga molekylerna vilket underlättar då man försöker identifiera dem.

Vidare ställs olika krav på metoderna beroende på vilken typ av prov som används. Vanligt förekommande provmaterial är exempelvis urin, blod (oftast i form av plasma eller serum), celler, eller olika typer av vävnader. Urin behöver normalt bara spädas för att kunna analyseras med LC medan till exempel plasma kräver att proteiner tas bort. I **artikel I** och **IV** var målsättningen att bättre förstå den skyddande effekten av vätgas på hörselnedsättning hos marsvin orsakad av olika typer av buller. Denna typ av skada, som är isolerad till innerörat, kan vara svår att studera om man tittar på exempelvis blod eller urin. Provmaterialet i dessa två studier var därför perilymfa som är en av de vätskorna som finns i innerörat. En svårighet här är att endast en väldigt liten volym (1 mikroliter hos marsvin) av denna vätska kan tas om man vill undvika kontamination av andra närliggande vätskor. Som kontrast kan här nämnas att en normal provvolym då man arbetar med plasma eller serum är runt 100-200 mikroliter. Den extremt låga provvolymen begränsar kraftigt hur många analyser som kan utföras då varje analys normalt förbrukar en del av provet. I **artikel I** analyserade därför bara de polära metaboliterna med hjälp av HILIC. För att öka bredden i metaboliter per analys då provmängden är begränsad behövs därför nya metoder.

I **artikel II** beskrivs utvecklingen av en ny separationsmetod som gör det möjligt att analysera prover med både HILIC och RPLC utan att kräva mer prov. Substanser som inte interagerar i så stor utsträckning med HILIC kolonner kommer komma ut ur kolonnen i stort sett omedelbart utan att separeras från varandra. I den utvecklade metoden tas dessa substanser om hand och sparas till dess att HILIC analysen är klar. De sparade substanserna kan sedan analyseras med en RPLC kolonn direkt efter, och därmed får man dessa ”på köpet” i en och samma analys utan att behöva mer prov.

En annan stor utmaning inom oriktad metabolomik är att experimenten genererar stora mängder av väldigt komplex data. För att forskarna ska kunna få ut användbar information krävs därför användningen av komplicerade algoritmer. En typ av algoritm som används letar efter signaler från substanser som separerats i kromatografin, så kallade kromatografiska toppar. Det är storleken på dessa toppar som sedan används för att dra slutsatser om hur mycket av en substans som finns i provet. Ett problem med dessa algoritmer är att målsättningen med deras design är att de ska hitta alla kromatografika toppar som finns i datan. Detta medför dock att de ibland blir för känsliga och rapporterar så kallade falska toppar, som vid inspektion inte alls är kromatografiska toppar. Dessa falska toppar har sedan en negativ inverkan på i stort sett all vidare processning och analys av datan. För att förbättra denna situation utvecklades

därför en algoritm som fungerar som ett komplement till de toppletande algoritmerna. Funktionen och utvecklingen av denna algoritm, kallad comprehensive peak characterization (CPC), är beskriven i **artikel III**. Den går systematiskt genom de toppar som rapporterats och kontrollerar om de verkligen är toppar. De toppar som bedöms vara falska kommer sedan tas bort innan man arbetar vidare med datan. Då algoritmen tillämpats på ett antal riktiga dataset från bland annat metabolomikstudier har vi sett att ungefär 30-50% av topparna som rapporteras endera inte finns alls, eller inte uppfyller grundläggande krav på exempelvis storlek. Både separationsmetoden (**artikel II**) och den nya algoritmen (**artikel III**) applicerades sedan i studien som presenteras i **artikel IV**.

Acknowledgements

All the research work presented in this thesis was conducted at the Department of Medicinal Chemistry at Uppsala University. I would like to express my sincere gratitude towards the following persons who have helped and motivated me towards finishing my PhD studies.

First of all, I want to thank my supervisors, **Professor Mikael Hedeland** for your invaluable knowledge and feedback, sometimes within impossible time frames, **Professor emeritus Curt Pettersson** for believing in me and taking me on as a PhD student, and **Torbjörn Arvidsson** for your enthusiasm and support.

A special thanks to **Associate Professor Carl Brunius** for your contagious enthusiasm in the field of data processing and for your guidance and in making the CPC package a reality.

Furthermore, I would like to thank all my co-authors in the papers presented in this thesis. Especially, **Professor Göran Laurell** and **Pernilla Videhult Pierre** for allowing me to take part in your interesting research and always providing your deep expertise.

I would also like to thank **Apotekarsocieteten** for making it possible for me to present my work at international conferences.

Furthermore, I would like to express my gratitude towards past and present colleagues at the Department of Medicinal Chemistry and in particular everyone in the research group of Analytical Pharmaceutical Chemistry.

Jakob, thank you for always being available to chat about teaching and research alike, and for giving me the opportunity to hold lectures. **Micke**, thank you for your never-ending enthusiasm and guidance. **Ylva**, thank you for all your help with both struggling GC instruments and teaching. **David**, it feels like we have always shared an office and you have helped me more than you admit, I look forward to more collaboration with you in the future. **Ida**, thank you for always being available when I need to ventilate, you are a big reason I am still sane. **Alfred**, probably the smartest people I know, thank you for all

the discussions and beers. **Albert**, thank you for great company during the first years. **Annelie**, thank you for your no-BS attitude, you are an inspiration. **Malin**, thank you for always sparing a smile and always being there to help. **Jenny**, I don't think I've ever seen you not smiling, keep it up! **Emelie**, always happy, even when you're dissecting bugs, I envy you! **Jens**, **Karin**, **Astrid**, **Erik**, **Camilla**, **Luke R.**, **Luke S.**, **Blazej** and **Quentin** thank you for all the great days and nights, you helped make my time as a PhD student memorable.

And finally, I want to express my gratitude to all my friends and family for you love and support. **Mamma** och **Pappa**, tack för att ni alltid finns där och tack för all flytthjälp. **Syrran**, tack för att du alltid finns där när jag behöver prata och för att du visar mig att det går, bara man verkligen vill.

Nicole, thank you for being my rock during the craziest, and thanks to you, best years of my life and for dragging me out of the door every once in a while to see the sun.

References

[1] O. Fiehn, Metabolomics — the link between genotypes and phenotypes, in: C. Town (Ed.), Functional Genomics, Springer Netherlands, Dordrecht, 2002: pp. 155–171. https://doi.org/10.1007/978-94-010-0448-0_11.

[2] D.S. Wishart, T. Jewison, A.C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat, E. Dong, S. Bouatra, I. Sinelnikov, D. Arndt, J. Xia, P. Liu, F. Yallou, T. Bjorndahl, R. Perez-Pineiro, R. Eisner, F. Allen, V. Neveu, R. Greiner, A. Scalbert, HMDB 3.0—The Human Metabolome Database in 2013, Nucleic Acids Research. 41 (2012) D801–D807. https://doi.org/10.1093/nar/gks1065.

[3] J.K. Nicholson, J.C. Lindon, E. Holmes, "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data, Xenobiotica. 29 (1999) 1181–1189. https://doi.org/10.1080/004982599238047.

[4] L. Pauling, A.B. Robinson, R. Teranishi, P. Cary, Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography, PNAS. 68 (1971) 2374–2376. https://doi.org/10.1073/pnas.68.10.2374.

[5] R. Goodacre, S. Vaidyanathan, W.B. Dunn, G.G. Harrigan, D.B. Kell, Metabolomics by numbers: acquiring and understanding global metabolite data, Trends in Biotechnology. 22 (2004) 245–252. https://doi.org/10.1016/j.tibtech.2004.03.007.

[6] O. Fiehn, Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks, Comparative and Functional Genomics. 2 (2001) 155–168. https://doi.org/10.1002/cfg.82.

[7] W.B. Dunn, N.J.C. Bailey, H.E. Johnson, Measuring the metabolome: current analytical technologies, Analyst. 130 (2005) 606. https://doi.org/10.1039/b418288j.

[8] M.S. Rashed, M.P. Bucknall, D. Little, A. Awad, M. Jacob, M. Alamoudi, M. Alwattar, P.T. Ozand, Screening blood spots for inborn errors of metabolism by electrospray tandem mass spectrometry with a microplate batch process and a computer algorithm for automated flagging of abnormal profiles, Clinical Chemistry. 43 (1997) 1129–1141. https://doi.org/10.1093/clinchem/43.7.1129.

[9] J.L. Griffin, R.A. Kauppinen, Tumour Metabolomics in Animal Models of Human Cancer, J. Proteome Res. 6 (2007) 498–505. https://doi.org/10.1021/pr060464h.

[10] J.L. Spratlin, N.J. Serkova, S.G. Eckhardt, Clinical Applications of Metabolomics in Oncology: A Review, Clinical Cancer Research. 15 (2009) 431–440. https://doi.org/10.1158/1078-0432.CCR-08-1059.

[11] E.G. Armitage, C. Barbas, Metabolomics in cancer biomarker discovery: Current trends and future perspectives, Journal of Pharmaceutical and Biomedical Analysis. 87 (2014) 1–11. https://doi.org/10.1016/j.jpba.2013.08.041.

[12] T. Fujita, D. Yamashita, Y. Irino, J. Kitamoto, Y. Fukuda, G. Inokuchi, S. Hasegawa, N. Otsuki, M. Yoshida, K. Nibu, Metabolomic profiling in inner ear fluid by gas chromatography/mass spectrometry in guinea pig cochlea, Neuroscience Letters. 606 (2015) 188–193. https://doi.org/10.1016/j.neulet.2015.09.001.

[13] L. Ji, H.-J. Lee, G. Wan, G.-P. Wang, L. Zhang, P. Sajjakulnukit, J. Schacht, C.A. Lyssiotis, G. Corfas, Auditory metabolomics, an approach to identify acute molecular effects of noise trauma, Sci Rep. 9 (2019) 9273. https://doi.org/10.1038/s41598-019-45385-8.

[14] J.C. Lindon, E. Holmes, J.K. Nicholson, Metabonomics in pharmaceutical R & D, The FEBS Journal. 274 (2007) 1140–1151. https://doi.org/10.1111/j.1742-4658.2007.05673.x.

[15] R. Kaddurah-Daouk, B.S. Kristal, R.M. Weinshilboum, Metabolomics: A Global Biochemical Approach to Drug Response and Disease, Annu. Rev. Pharmacol. Toxicol. 48 (2008) 653–683. https://doi.org/10.1146/annurev.pharmtox.48.113006.094715.

[16] D.S. Wishart, Emerging applications of metabolomics in drug discovery and precision medicine, Nat Rev Drug Discov. 15 (2016) 473–484. https://doi.org/10.1038/nrd.2016.32.

[17] T. Andrew Clayton, J.C. Lindon, O. Cloarec, H. Antti, C. Charuel, G. Hanton, J.-P. Provost, J.-L. Le Net, D. Baker, R.J. Walley, J.R. Everett, J.K. Nicholson, Pharmaco-metabonomic phenotyping and personalized drug treatment, Nature. 440 (2006) 1073–1077. https://doi.org/10.1038/nature04648.

[18] M.J. Gibney, M. Walsh, L. Brennan, H.M. Roche, B. German, B. van Ommen, Metabolomics in human nutrition: opportunities and challenges, The American Journal of Clinical Nutrition. 82 (2005) 497–503. https://doi.org/10.1093/ajcn/82.3.497.

[19] A. Scalbert, L. Brennan, C. Manach, C. Andres-Lacueva, L.O. Dragsted, J. Draper, S.M. Rappaport, J.J. van der Hooft, D.S. Wishart, The food metabolome: a window over dietary exposure, Am. J. Clin. Nutr. 99 (2014) 1286–1308. https://doi.org/10.3945/ajcn.113.076133.

[20] S. Rezzi, Z. Ramadan, L.B. Fay, S. Kochhar, Nutritional Metabonomics: Applications and Perspectives, J. Proteome Res. 6 (2007) 513–525. https://doi.org/10.1021/pr060522z.

[21] C.H. Johnson, J. Ivanisevic, G. Siuzdak, Metabolomics: beyond biomarkers and towards mechanisms, Nat. Rev. Mol. Cell Biol. 17 (2016) 451–459. https://doi.org/10.1038/nrm.2016.25.

[22] W.E. Gall, K. Beebe, K.A. Lawton, K.-P. Adam, M.W. Mitchell, P.J. Nakhle, J.A. Ryals, M.V. Milburn, M. Nannipieri, S. Camastra, A. Natali, E. Ferrannini, for the the R.S. Group, α-Hydroxybutyrate Is an Early Biomarker of Insulin Resistance and Glucose Intolerance in a Nondiabetic Population, PLOS ONE. 5 (2010) e10883. https://doi.org/10.1371/journal.pone.0010883.

[23] W. Arlt, M. Biehl, A.E. Taylor, S. Hahner, R. Libé, B.A. Hughes, P. Schneider, D.J. Smith, H. Stiekema, N. Krone, E. Porfiri, G. Opocher, J. Bertherat, F. Mantero, B. Allolio, M. Terzolo, P. Nightingale, C.H.L. Shackleton, X. Bertagna, M. Fassnacht, P.M. Stewart, Urine Steroid Metabolomics as a Biomarker Tool for Detecting Malignancy in Adrenal Tumors, The Journal of Clinical Endocrinology & Metabolism. 96 (2011) 3775–3784. https://doi.org/10.1210/jc.2011-1565.

[24] T. Soga, R. Baran, M. Suematsu, Y. Ueno, S. Ikeda, T. Sakurakawa, Y. Kakazu, T. Ishikawa, M. Robert, T. Nishioka, M. Tomita, Differential Metabolomics Reveals Ophthalmic Acid as an Oxidative Stress Biomarker Indicating Hepatic Glutathione Consumption*, Journal of Biological Chemistry. 281 (2006) 16768–16776. https://doi.org/10.1074/jbc.M601876200.

[25] M. Jenab, N. Slimani, M. Bictash, P. Ferrari, S.A. Bingham, Biomarkers in nutritional epidemiology: applications, needs and new horizons, Hum Genet. 125 (2009) 507–525. https://doi.org/10.1007/s00439-009-0662-5.

[26] R. Kaddurah-Daouk, J. McEvoy, R.A. Baillie, D. Lee, J.K. Yao, P.M. Doraiswamy, K.R.R. Krishnan, Metabolomic mapping of atypical antipsychotic effects in schizophrenia, Mol Psychiatry. 12 (2007) 934–945. https://doi.org/10.1038/sj.mp.4002000.

[27] J.V. Li, H. Ashrafian, M. Bueter, J. Kinross, C. Sands, C.W. le Roux, S.R. Bloom, A. Darzi, T. Athanasiou, J.R. Marchesi, J.K. Nicholson, E. Holmes, Metabolic surgery profoundly influences gut microbial–host metabolic cross-talk, Gut. 60 (2011) 1214–1223. https://doi.org/10.1136/gut.2010.234708.

[28] W. Li, H. Zhang, Y.G. Assaraf, K. Zhao, X. Xu, J. Xie, D.-H. Yang, Z.-S. Chen, Overcoming ABC transporter-mediated multidrug resistance: Molecular mechanisms and novel therapeutic drug strategies, Drug Resistance Updates. 27 (2016) 14–29. https://doi.org/10.1016/j.drup.2016.05.001.

[29] A. Malmendal, J. Overgaard, J.G. Bundy, J.G. Sørensen, N.Chr. Nielsen, V. Loeschcke, M. Holmstrup, Metabolomic profiling of heat stress: hardening and recovery of homeostasis in Drosophila, American Journal of Physiology-Regulatory, Integrative and Comparative Physiology. 291 (2006) R205–R212. https://doi.org/10.1152/ajpregu.00867.2005.

[30] C. Chen, Y.M. Shah, K. Morimura, K.W. Krausz, M. Miyazaki, T.A. Richardson, E.T. Morgan, J.M. Ntambi, J.R. Idle, F.J. Gonzalez, Metabolomics Reveals that Hepatic Stearoyl-CoA Desaturase 1 Downregulation Exacerbates Inflammation and Acute Colitis, Cell Metabolism. 7 (2008) 135–147. https://doi.org/10.1016/j.cmet.2007.12.003.

[31] S. Wu, G. Tan, X. Dong, Z. Zhu, W. Li, Z. Lou, Y. Chai, Metabolic Profiling Provides a System Understanding of Hypothyroidism in Rats and Its Application, PLOS ONE. 8 (2013) e55599. https://doi.org/10.1371/journal.pone.0055599.

[32] E. Holmes, R.L. Loo, J. Stamler, M. Bictash, I.K.S. Yap, Q. Chan, T. Ebbels, M. De Iorio, I.J. Brown, K.A. Veselkov, M.L. Daviglus, H. Kesteloot, H. Ueshima, L. Zhao, J.K. Nicholson, P. Elliott, Human metabolic phenotype diversity and its association with diet and blood pressure, Nature. 453 (2008) 396–400. https://doi.org/10.1038/nature06882.

[33] T.R. Koves, J.R. Ussher, R.C. Noland, D. Slentz, M. Mosedale, O. Ilkayeva, J. Bain, R. Stevens, J.R.B. Dyck, C.B. Newgard, G.D. Lopaschuk, D.M. Muoio, Mitochondrial Overload and Incomplete Fatty Acid Oxidation Contribute to Skeletal Muscle Insulin Resistance, Cell Metabolism. 7 (2008) 45–56. https://doi.org/10.1016/j.cmet.2007.10.013.

[34] R.J. Mortishire-Smith, G.L. Skiles, J.W. Lawrence, S. Spence, A.W. Nicholls, B.A. Johnson, J.K. Nicholson, Use of Metabonomics to Identify Impaired Fatty Acid Metabolism as the Mechanism of a Drug-Induced Toxicity, Chem. Res. Toxicol. 17 (2004) 165–173. https://doi.org/10.1021/tx034123j.

[35] J. Zhang, E.H. Herman, D.G. Robertson, M.D. Reily, A. Knapton, H.V. Ratajczak, N. Rifai, R. Honchel, K.T. Blanchard, R.E. Stoll, F.D. Sistare, Mechanisms and Biomarkers of Cardiovascular Injury Induced by Phosphodiesterase Inhibitor III SK&F 95654 in the Spontaneously Hypertensive Rat, Toxicol Pathol. 34 (2006) 152–163. https://doi.org/10.1080/01926230600588562.

[36] Y.-Y. Zhao, H.-L. Wang, X.-L. Cheng, F. Wei, X. Bai, R.-C. Lin, N.D. Vaziri, Metabolomics analysis reveals the association between lipid abnormalities and oxidative stress, inflammation, fibrosis and Nrf2 dysfunction in aristolochic acid-induced nephropathy, Sci Rep. 5 (2015) 12936. https://doi.org/10.1038/srep12936.

[37] A.C. Schrimpe-Rutledge, S.G. Codreanu, S.D. Sherrod, J.A. McLean, Untargeted Metabolomics Strategies—Challenges and Emerging Directions, J. Am. Soc. Mass Spectrom. 27 (2016) 1897–1905. https://doi.org/10.1007/s13361-016-1469-y.

[38] D.S. Wishart, D. Tzur, C. Knox, R. Eisner, A.C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.-A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, David.D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G.E. Duggan, G.D. MacInnis, A.M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B.D. Sykes, H.J. Vogel, L. Querengesser, HMDB: the Human Metabolome Database, Nucleic Acids Research. 35 (2007) D521–D526. https://doi.org/10.1093/nar/gkl923.

[39] E.M. Lenz, I.D. Wilson, Analytical Strategies in Metabonomics, J. Proteome Res. 6 (2007) 443–458. https://doi.org/10.1021/pr0605217.

[40] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.-M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI), Metabolomics. 3 (2007) 211–221. https://doi.org/10.1007/s11306-007-0082-2.

[41] S. Naz, M. Vallejo, A. García, C. Barbas, Method validation strategies involved in non-targeted metabolomics, J. Chromatogr. A. 1353 (2014) 99–105. https://doi.org/10.1016/j.chroma.2014.04.071.

[42] M.K.R. Engskog, J. Haglöf, T. Arvidsson, C. Pettersson, LC–MS based global metabolite profiling: the necessity of high data quality, Metabolomics. 12 (2016) 114. https://doi.org/10.1007/s11306-016-1058-x.

[43] W.B. Dunn, D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, M. Brown, J.D. Knowles, A. Halsall, J.N. Haselden, A.W. Nicholls, I.D. Wilson, D.B. Kell, R. Goodacre, Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry, Nat Protoc. 6 (2011) 1060–1083. https://doi.org/10.1038/nprot.2011.335.

[44] V.L. Stevens, E. Hoover, Y. Wang, K.A. Zanetti, Pre-Analytical Factors that Affect Metabolite Stability in Human Urine, Plasma, and Serum: A Review, Metabolites. 9 (2019) 156. https://doi.org/10.3390/metabo9080156.

[45] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass Spectrometry-Based Metabolomics, Mass Spectrom. Rev. 26 (2007) 51–78. https://doi.org/10.1002/mas.20108.

[46] N. Psychogios, D.D. Hau, J. Peng, A.C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T.L. Pedersen, S.R. Smith, F. Bamforth, R. Greiner, B. McManus, J.W. Newman, T. Goodfriend, D.S. Wishart, The Human Serum Metabolome, PLoS ONE. 6 (2011) e16957. https://doi.org/10.1371/journal.pone.0016957.

[47] D.B. Kell, M. Brown, H.M. Davey, W.B. Dunn, I. Spasic, S.G. Oliver, Metabolic footprinting and systems biology: the medium is the message, Nat Rev Microbiol. 3 (2005) 557–565. https://doi.org/10.1038/nrmicro1177.

[48] A. Hara, A.N. Salt, R. Thalmann, Perilymph composition in scala tympani of the cochlea: Influence of cerebrospinal fluid, Hearing Research. 42 (1989) 265–271. https://doi.org/10.1016/0378-5955(89)90150-0.

[49] D. Vuckovic, Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry, Anal Bioanal Chem. 403 (2012) 1523–1548. https://doi.org/10.1007/s00216-012-6039-y.

[50] G.A. Theodoridis, H.G. Gika, E.J. Want, I.D. Wilson, Liquid chromatography–mass spectrometry based global metabolite profiling: A review, Analytica Chimica Acta. 711 (2012) 7–16. https://doi.org/10.1016/j.aca.2011.09.042.

[51] E.J. Want, I.D. Wilson, H. Gika, G. Theodoridis, R.S. Plumb, J. Shockcor, E. Holmes, J.K. Nicholson, Global metabolic profiling procedures for urine using UPLC–MS, Nat. Protoc. 5 (2010) 1005–1018. https://doi.org/10.1038/nprot.2010.50.

[52] D. Broadhurst, R. Goodacre, S.N. Reinke, J. Kuligowski, I.D. Wilson, M.R. Lewis, W.B. Dunn, Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies, Metabolomics. 14 (2018). https://doi.org/10.1007/s11306-018-1367-3.

[53] T. Sangster, H. Major, R. Plumb, A.J. Wilson, I.D. Wilson, A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis, Analyst. 131 (2006) 1075–1078. https://doi.org/10.1039/b604498k.

[54] K. Spagou, H. Tsoukali, N. Raikos, H. Gika, I.D. Wilson, G. Theodoridis, Hydrophilic interaction chromatography coupled to MS for metabonomic/metabolomic studies, J. Sep. Sci. 33 (2010) 716–727. https://doi.org/10.1002/jssc.200900803.

[55] H.G. Gika, G.A. Theodoridis, J.E. Wingate, I.D. Wilson, Within-Day Reproducibility of an HPLC−MS-Based Method for Metabonomic Analysis: Application to Human Urine, J. Proteome Res. 6 (2007) 3291–3303. https://doi.org/10.1021/pr070183p.

[56] D.-P. Kloos, H. Lingeman, W.M.A. Niessen, A.M. Deelder, M. Giera, O.A. Mayboroda, Evaluation of different column chemistries for fast urinary metabolic profiling, J. Chromatogr. B. 927 (2013) 90–96. https://doi.org/10.1016/j.jchromb.2013.02.017.

[57] S.M.R. Stanley, H.C. Foo, Screening for basic drugs in equine urine using direct-injection differential-gradient LC–LC coupled to hybrid tandem MS/MS, J. Chromatogr. B. 836 (2006) 1–14. https://doi.org/10.1016/j.jchromb.2006.03.034.

[58] W.H. Kwok, T.L.S. Choi, Y.Y.K. Tsoi, G.N.W. Leung, T.S.M. Wan, Screening of over 100 drugs in horse urine using automated on-line solid-phase extraction coupled to liquid chromatography-high resolution mass spectrometry for doping control, J. Chromatogr. A. 1490 (2017) 89–101. https://doi.org/10.1016/j.chroma.2017.02.020.

[59] S. Wang, L. Zhou, Z. Wang, X. Shi, G. Xu, Simultaneous metabolomics and lipidomics analysis based on novel heart-cutting two-dimensional liquid chromatography-mass spectrometry, Anal. Chim. Acta. 966 (2017) 34–40. https://doi.org/10.1016/j.aca.2017.03.004.

[60] S. Louw, A.S. Pereira, F. Lynen, M. Hanna-Brown, P. Sandra, Serial coupling of reversed-phase and hydrophilic interaction liquid chromatography to broaden the elution window for the analysis of pharmaceutical compounds, J. Chromatogr. A. 1208 (2008) 90–94. https://doi.org/10.1016/j.chroma.2008.08.058.

[61] S. Falasca, F. Petruzziello, R. Kretz, G. Rainer, X. Zhang, Analysis of multiple quaternary ammonium compounds in the brain using tandem capillary column separation and high resolution mass spectrometric detection, J. Chromatogr. A. 1241 (2012) 46–51. https://doi.org/10.1016/j.chroma.2012.04.002.

[62] G. Greco, S. Grosse, T. Letzel, Serial coupling of reversed-phase and zwitterionic hydrophilic interaction LC/MS for the analysis of polar and nonpolar phenols in wine: Liquid Chromatography, J. Sep. Science. 36 (2013) 1379–1388. https://doi.org/10.1002/jssc.201200920.

[63] M. Rajab, G. Greco, C. Heim, B. Helmreich, T. Letzel, Serial coupling of RP and zwitterionic hydrophilic interaction LC-MS: Suspects screening of diclofenac transformation products by oxidation with a boron-doped diamond electrode: Liquid Chromatography, J. Sep. Science. 36 (2013) 3011–3018. https://doi.org/10.1002/jssc.201300562.

[64] K. Ortmayr, S. Hann, G. Koellensperger, Complementing reversed-phase selectivity with porous graphitized carbon to increase the metabolome coverage in an on-line two-dimensional LC-MS setup for metabolomics, Analyst. 140 (2015) 3465–3473. https://doi.org/10.1039/C5AN00206K.

[65] A.F.G. Gargano, M. Duffin, P. Navarro, P.J. Schoenmakers, Reducing Dilution and Analysis Time in Online Comprehensive Two-Dimensional Liquid Chromatography by Active Modulation, Anal. Chem. 88 (2016) 1785–1793. https://doi.org/10.1021/acs.analchem.5b04051.

[66] Y. Wang, R. Lehmann, X. Lu, X. Zhao, G. Xu, Novel, fully automatic hydrophilic interaction/reversed-phase column-switching high-performance liquid chromatographic system for the complementary analysis of polar and apolar compounds in complex samples, J. Chromatogr. A. 1204 (2008) 28–34. https://doi.org/10.1016/j.chroma.2008.07.010.

[67] D. Cabooter, K. Choikhet, F. Lestremau, M. Dittmann, G. Desmet, Towards a generic variable column length method development strategy for samples with a large variety in polarity, J. Chromatogr. A. 1372 (2014) 174–186. https://doi.org/10.1016/j.chroma.2014.11.006.

[68] S. Kittlaus, J. Schimanke, G. Kempe, K. Speer, Development and validation of an efficient automated method for the analysis of 300 pesticides in foods using two-dimensional liquid chromatography–tandem mass spectrometry, J. Chromatogr. A. 1283 (2013) 98–109. https://doi.org/10.1016/j.chroma.2013.01.106.

[69] G. Loos, K. Shoykhet, M. Dittmann, D. Cabooter, Restriction capillaries as an innovative mixing unit for intermediate mobile phase exchange in multidimensional analysis, J. Chromatogr. A. 1497 (2017) 70–80. https://doi.org/10.1016/j.chroma.2017.03.036.

[70] C.A. Smith, E.J. Want, G. O'Maille, R. Abagyan, G. Siuzdak, XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification, Anal. Chem. 78 (2006) 779–787. https://doi.org/10.1021/ac051437y.

[71] M. Katajamaa, J. Miettinen, M. Oresic, MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data, Bioinformatics. 22 (2006) 634–636. https://doi.org/10.1093/bioinformatics/btk039.

[72] T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data, BMC Bioinformatics. 11 (2010) 395. https://doi.org/10.1186/1471-2105-11-395.

[73] M. Sturm, A. Bertsch, C. Gröpl, A. Hildebrandt, R. Hussong, E. Lange, N. Pfeifer, O. Schulz-Trieglaff, A. Zerck, K. Reinert, O. Kohlbacher, OpenMS – An open-source software framework for mass spectrometry, BMC Bioinformatics. 9 (2008) 163. https://doi.org/10.1186/1471-2105-9-163.

[74] R. Di Guida, J. Engel, J.W. Allwood, R.J.M. Weber, M.R. Jones, U. Sommer, M.R. Viant, W.B. Dunn, Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling, Metabolomics. 12 (2016) 93. https://doi.org/10.1007/s11306-016-1030-9.

[75] K.A. Veselkov, L.K. Vingara, P. Masson, S.L. Robinette, E. Want, J.V. Li, R.H. Barton, C. Boursier-Neyret, B. Walther, T.M. Ebbels, I. Pelczer, E. Holmes, J.C. Lindon, J.K. Nicholson, Optimized Preprocessing of Ultra-Performance Liquid Chromatography/Mass Spectrometry Urinary Metabolic Profiles for Improved Information Recovery, Anal. Chem. 83 (2011) 5864–5872. https://doi.org/10.1021/ac201065j.

[76] E. Saccenti, Correlation Patterns in Experimental Data Are Affected by Normalization Procedures: Consequences for Data Analysis and Network Inference, J. Proteome Res. 16 (2017) 619–634. https://doi.org/10.1021/acs.jproteome.6b00704.

[77] S.M. Kohl, M.S. Klein, J. Hochrein, P.J. Oefner, R. Spang, W. Gronwald, State-of-the art data normalization methods improve NMR-based metabolomic analysis, Metabolomics. 8 (2012) 146–160. https://doi.org/10.1007/s11306-011-0350-z.

[78] F. Dieterle, A. Ross, G. Schlotterbeck, H. Senn, Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics, Analytical Chemistry. 78 (2006) 4281–4290. https://doi.org/10.1021/ac051632c.

[79] J.B. Coble, C.G. Fraga, Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery, J. Chromatogr. A. 1358 (2014) 155–164. https://doi.org/10.1016/j.chroma.2014.06.100.

[80] O.D. Myers, S.J. Sumner, S. Li, S. Barnes, X. Du, Detailed Investigation and Comparison of the XCMS and MZmine 2 Chromatogram Construction and Chromatographic Peak Detection Methods for Preprocessing Mass Spectrometry Metabolomics Data, Analytical Chemistry. 89 (2017) 8689–8695. https://doi.org/10.1021/acs.analchem.7b01069.

[81] ApexTrack Integration: Theory and Application, Waters Corp., 2016.

[82] Abraham. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures., Anal. Chem. 36 (1964) 1627–1639. https://doi.org/10.1021/ac60214a047.

[83] E. Grushka, Characterization of exponentially modified Gaussian peaks in chromatography, Anal. Chem. 44 (1972) 1733–1738. https://doi.org/10.1021/ac60319a011.

[84] Y. Kalambet, Y. Kozmin, K. Mikhailova, I. Nagaev, P. Tikhonov, Reconstruction of chromatographic peaks using the exponentially modified Gaussian function, J. Chemometrics. 25 (2011) 352–356. https://doi.org/10.1002/cem.1343.

[85] R.A. van den Berg, H.C. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data, BMC Genomics. 7 (2006) 142. https://doi.org/10.1186/1471-2164-7-142.

[86] E. Zelena, W.B. Dunn, D. Broadhurst, S. Francis-McIntyre, K.M. Carroll, P. Begley, S. O'Hagan, J.D. Knowles, A. Halsall, I.D. Wilson, D.B. Kell, Development of a Robust and Repeatable UPLC−MS Method for the Long-Term Metabolomic Study of Human Serum, Anal. Chem. 81 (2009) 1357–1364. https://doi.org/10.1021/ac8019366.

[87] L. Eriksson, T. Byrne, E. Johansson, J. Trygg, C. Vikström, Multi- and Mega-variate Data Analysis Basic Principles and Applications, Umetrics Academy, 2013.
[88] J. Trygg, S. Wold, Orthogonal projections to latent structures (O-PLS), J. Chemometrics. 16 (2002) 119–128. https://doi.org/10.1002/cem.695.
[89] L. Breiman, Random Forests, Machine Learning. 45 (2001) 5–32. https://doi.org/10.1023/A:1010933404324.
[90] J. Trygg, E. Holmes, T. Lundstedt, Chemometrics in Metabonomics, J. Proteome Res. 6 (2007) 469–479. https://doi.org/10.1021/pr060594q.
[91] S. Wiklund, E. Johansson, L. Sjöström, E.J. Mellerowicz, U. Edlund, J.P. Shockcor, J. Gottfries, T. Moritz, J. Trygg, Visualization of GC/TOF-MS-Based Metabolomics Data for Identification of Biochemically Interesting Compounds Using OPLS Class Models, Anal. Chem. 80 (2008) 115–122. https://doi.org/10.1021/ac0713510.
[92] L. Shi, J.A. Westerhuis, J. Rosén, R. Landberg, C. Brunius, Variable selection and validation in multivariate modelling, Bioinformatics. 35 (2019) 972–980. https://doi.org/10.1093/bioinformatics/bty710.
[93] W. Li, Volcano Plots In Analyzing Differential Expressions With mRNA Microarrays, J. Bioinform. Comput. Biol. 10 (2012) 1231003. https://doi.org/10.1142/S0219720012310038.
[94] C. Kuhl, R. Tautenhahn, C. Böttcher, T.R. Larson, S. Neumann, CAMERA: An Integrated Strategy for Compound Spectra Extraction and Annotation of Liquid Chromatography/Mass Spectrometry Data Sets, Anal. Chem. 84 (2012) 283–289. https://doi.org/10.1021/ac202450g.
[95] D.S. Wishart, C. Knox, A.C. Guo, R. Eisner, N. Young, B. Gautam, D.D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. Xia, L. Jia, J.A. Cruz, E. Lim, C.A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. De Souza, A. Zuniga, M. Dawe, Y. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhutdinov, L. Li, H.J. Vogel, I. Forsythe, HMDB: a knowledgebase for the human metabolome, Nucleic Acids Research. 37 (2009) D603–D610. https://doi.org/10.1093/nar/gkn810.
[96] M. Kanehisa, S. Goto, KEGG: Kyoto Encyclopedia of Genes and Genomes, Nucleic Acids Research. 28 (2000) 27–30. https://doi.org/10.1093/nar/28.1.27.
[97] C.A. Smith, G.O. Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custodio, R. Abagyan, G. Siuzdak, METLIN: A Metabolite Mass Spectral Database, Therapeutic Drug Monitoring. 27 (2005) 747–751. https://doi.org/10.1097/01.ftd.0000179845.53213.39.
[98] R. Taguchi, M. Nishijima, T. Shimizu, Basic Analytical Systems for Lipidomics by Mass Spectrometry in Japan, in: Methods in Enzymology, Academic Press, 2007: pp. 185–211. https://doi.org/10.1016/S0076-6879(07)32008-9.
[99] E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, J. Hollender, Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence, Environ. Sci. Technol. 48 (2014) 2097–2098. https://doi.org/10.1021/es5002105.

[100] S. Alseekh, A. Aharoni, Y. Brotman, K. Contrepois, J. D'Auria, J. Ewald, J. C. Ewald, P.D. Fraser, P. Giavalisco, R.D. Hall, M. Heinemann, H. Link, J. Luo, S. Neumann, J. Nielsen, L. Perez de Souza, K. Saito, U. Sauer, F.C. Schroeder, S. Schuster, G. Siuzdak, A. Skirycz, L.W. Sumner, M.P. Snyder, H. Tang, T. Tohge, Y. Wang, W. Wen, S. Wu, G. Xu, N. Zamboni, A.R. Fernie, Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices, Nat Methods. 18 (2021) 747–756. https://doi.org/10.1038/s41592-021-01197-1.

[101] D. Henderson, E.C. Bielefeld, K.C. Harris, B.H. Hu, The Role of Oxidative Stress in Noise-Induced Hearing Loss, Ear & Hearing. 27 (2006) 1–19. https://doi.org/10.1097/01.aud.0000191942.36672.f3.

[102] T.N. Le, L.V. Straatman, J. Lea, B. Westerberg, Current insights in noise-induced hearing loss: a literature review of the underlying mechanism, pathophysiology, asymmetry, and management options, J of Otolaryngol - Head & Neck Surg. 46 (2017) 41. https://doi.org/10.1186/s40463-017-0219-x.

[103] F. Lang, Mechanisms and Significance of Cell Volume Regulation, Journal of the American College of Nutrition. 26 (2007) 613S-623S. https://doi.org/10.1080/07315724.2007.10719667.

[104] S.T. Chambers, C.M. Kunin, Isolation of glycine betaine and proline betaine from human urine. Assessment of their role as osmoprotective agents for bacteria and the kidney., J Clin Invest. 79 (1987) 731–737. https://doi.org/10.1172/JCI112878.

[105] G. Strifler, E. Tuboly, A. Görbe, M. Boros, D. Pécz, P. Hartmann, Targeting Mitochondrial Dysfunction with L-Alpha Glycerylphosphorylcholine, PLOS ONE. 11 (2016) e0166682. https://doi.org/10.1371/journal.pone.0166682.

[106] A.A. Zdebik, P. Wangemann, T.J. Jentsch, Potassium Ion Movement in the Inner Ear: Insights from Genetic Disease and Mouse Models, Physiology. 24 (2009) 307–316. https://doi.org/10.1152/physiol.00018.2009.

[107] M.J. Dixon, J. Gazzard, S.S. Chaudhry, N. Sampson, B.A. Schulte, K.P. Steel, Mutation of the Na-K-Cl Co-Transporter Gene Slc12a2 Results in Deafness in Mice, Human Molecular Genetics. 8 (1999) 1579–1584. https://doi.org/10.1093/hmg/8.8.1579.

[108] Y. Takiguchi, G. Sun, K. Ogawa, T. Matsunaga, Long-lasting changes in the cochlear K+ recycling structures after acute energy failure, Neuroscience Research. 77 (2013) 33–41. https://doi.org/10.1016/j.neures.2013.06.003.

[109] Y. Guo, A. Huang, A HILIC method for the analysis of tromethamine as the counter ion in an investigational pharmaceutical salt, J. Pharm. Biomed. Anal. 31 (2003) 1191–1201. https://doi.org/10.1016/S0731-7085(03)00021-9.

[110] Z. Huang, M.A. Richards, Y. Zha, R. Francis, R. Lozano, J. Ruan, Determination of inorganic pharmaceutical counterions using hydrophilic interaction chromatography coupled with a Corona® CAD detector, Journal of Pharmaceutical and Biomedical Analysis. 50 (2009) 809–814. https://doi.org/10.1016/j.jpba.2009.06.039.

[111] A. Elmsjö, J. Haglöf, M.K.R. Engskog, I. Erngren, M. Nestor, T. Arvidsson, C. Pettersson, Method selectivity evaluation using the co-feature ratio in LC/MS metabolomics: Comparison of HILIC stationary phase performance for the analysis of plasma, urine and cell extracts, Journal of Chromatography A. 1568 (2018) 49–56. https://doi.org/10.1016/j.chroma.2018.05.007.

[112] I. Erngren, J. Haglöf, M.K.R. Engskog, M. Nestor, M. Hedeland, T. Arvidsson, C. Pettersson, Adduct formation in electrospray ionisation-mass spectrometry with hydrophilic interaction liquid chromatography is strongly affected by the inorganic ion concentration of the samples, Journal of Chromatography A. 1600 (2019) 174–182. https://doi.org/10.1016/j.chroma.2019.04.049.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy 306

Editor: The Dean of the Faculty of Pharmacy

A doctoral dissertation from the Faculty of Pharmacy, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Pharmacy".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-461320

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2022