Challenges in ESI-MS-based Untargeted Metabolomics

Elena Tobolkina[§ab], Víctor González-Ruiz[§abc], Isabel Meister[ab], Miguel de Figueiredo[ab], Davy Guillarme[ab], Julien Boccard[abc], and Serge Rudaz*[abc]

Abstract: Untargeted metabolomics is now widely recognized as a useful tool for exploring metabolic changes taking place in biological systems under different conditions. In this article, we aim to provide a short overview of the liquid-phase separation methods hyphenated to MS to perform untargeted metabolomics of biological samples. Each approach is complemented by up-to-date literature to guide readers, as well as practical information for avoiding or fixing some of the most frequently encountered pitfalls. This article covers mainly data acquisition, but a short discussion is provided regarding signal processing and data treatment, as well as data analysis and its biological interpretation in the context of metabolomic studies.

Keywords: Annotation · Capillary electrophoresis · Liquid chromatography · Mass spectrometry · Metabolomics · Supercritical fluid chromatography · Toxicological and doping analysis

Serge Rudaz is Professor at the University of Geneva where he leads the biomedical and metabolomics analysis (BMA) group. He is President of the Swiss Metabolomics Society (SMS), vice-president of the Competence Center in Chemical and Toxicological Analysis (ccCTA) and member of the management Board of the Swiss Centre for Applied Human Toxicology (SCAHT) Foundation. He is interested in steroids analysis, UHPLC and CE coupled to MS, advances in sample preparation, analysis of pharmaceuticals and falsified medicines, clinical and preclinical studies, including metabolism and toxicological analysis. He is a (co)author of over 10 book chapters and more than 330 peer-reviewed papers.

Elena Tobolkina received her PhD on Gel Electrophoresis and Mass Spectrometry from École Polytechnique Fédérale de Lausanne (EPFL) in 2014, where she obtained a patent. Her PhD work led her to pursue a post-doc position in biochemistry at the University of Oxford in collaboration with L'Oreal. She worked later as a Team Leader at Michael J. Fox foundation on the discovery of biomarkers for Parkinson's disease. In 2018, she joined the School of Pharmaceutical Sciences at the University of Geneva as Senior Scientist. Her focus of research is the advanced untargeted metabolomics, human steroid metabolomics and analysis of falsified drugs.

Víctor González-Ruiz is a pharmacist (2008) and holds a M.S. in Pharmaceutical R&D (2009). He got a PhD in Analytical Chemistry (2014) from the Complutense University of Madrid, developing LC- and fluorescence-based analytical techniques to study antitumor drugs and their mechanisms of action. In 2015 he joined the Analytical Sciences lab at the School of Pharmaceutical Sciences, University of Geneva, where he currently works as a research associate. His main research interest is the development of toxicology-oriented analytical (mainly CE-MS and LC-MS) and data handling strategies for metabolomics, in collaboration with the Swiss Center for Applied Human Toxicology.

Isabel Meister is a researcher at the School of Pharmaceutical Sciences at the University of Geneva. She received her PhD in 2015 from the Swiss Tropical and Public Health Institute on LC-MS-based pharmacokinetic assays for anthelmintic drugs. In 2015 she was awarded the Novartis Excellence Scholarship. In 2017, she received a Japan Society for the Promotion of Science Postdoctoral fellowship and moved to the Karolinska International Open Laboratory at the University of Gunma, Japan, where she worked on untargeted metabolomics of asthma. In Geneva, she continues to advance untargeted metabolomics for clinical applications in the framework of the Swiss Center for Applied Human Toxicology.

Correspondence: Prof. S. Rudaz[abc], E-Mail: serge.rudaz@unige.ch

[a]School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland; [b]Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Geneva, Switzerland; [c]Swiss Centre for Applied Human Toxicology (SCAHT), Switzerland

[§]These authors contributed equally to the work

Davy Guillarme holds a PhD in analytical chemistry from the University of Lyon, France. He is now senior lecturer and research associate at the University of Geneva in Switzerland. He authored more than 300 journal articles related to pharmaceutical analysis. He is an associate editor of *Journal of Chromatography B* and editorial advisory board member of several journals. He was elected as one of the world's most influential analytical scientists by *Analytical Scientist* magazine in 2013, 2014, 2015, 2017, 2019, 2020 and 2021.

Julien Boccard is a specialist in chemometrics applied to Omics data. After studying biology and bioinformatics, he graduated with a PhD in 2009. He serves now as a lecturer at the School of Pharmaceutical Sciences of the University of Geneva. His research interests are focused on data mining applied to life sciences, including chemometrics, machine learning and bioinformatics. More specifically, he develops computational and chemometric methods to extract knowledge from megavariate data, such as metabolomic profiles, to combine multiple data sources through data fusion strategies, and incorporate biological or chemical knowledge into modeling workflows.
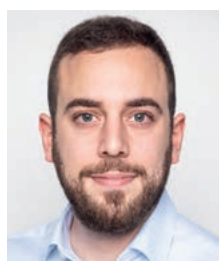
Miguel de Figueiredo is a forensic chemist specialized in chemometrics with a Master in forensic science from the School of Criminal Justice (UNIL). During his PhD, he investigated the contribution of chemometrics for arson investigation at the Police Prefecture Central Laboratory in Paris where he graduated in 2018. After a PostDoc at ETH Zürich in Zenobi group in real-time breath analysis, he serves now as a PostDoc within the Biomedical and Metabolomics Analysis group of the School of Pharmaceutical Sciences (UNIGE) developing innovative chemometrics and bioinformatics tools for the longitudinal profiling of endogenous steroids for the detection of doping.

1. Introduction

Metabolomics is one of the most suitable techniques for assessing biochemical processes at the phenotype level. While tracking a limited number of metabolites can be sufficient for diagnostic purposes, an extended set of compounds may be necessary to map and understand biochemical networks involved in complex or poorly known biological processes. Untargeted metabolomics has rapidly developed in the past ten years, but it still faces many challenges particularly those related to improving chemical coverage, metabolite identification, and data analysis and contextualization.

Firstly, untargeted monitoring of metabolites involves the detection of molecules of diverse chemical structures and physico-chemical properties, from very polar compounds (such as amino acids and sugars) to very apolar compounds (most lipids). Therefore, combining different analytical techniques with mass spectrometry (MS) detection is required to reach exhaustive metabolite coverage of biological samples. Hence, in recent years, research in metabolomics has seen the development of rapid, generic and high-resolution separation methods using ultra-high-performance liquid chromatography (UHPLC), supercritical fluid chromatography (SFC) and capillary electrophoresis (CE), which can be combined to broaden the chemical space that can be observed.

Secondly, metabolite annotation plays a central role since it bridges chemical analysis with biological interpretation. Nevertheless, it remains one of the main bottlenecks in the current biomarker discovery workflows due to the large degeneracy of signals obtained in electrospray ionization (ESI) and the difficulty to unequivocally assign the observed peaks. Annotation based on diverse and orthogonal parameters obtained thanks to separation techniques and MS(/MS) can be achieved not only through the in-house measurement of authentic standards but also through external repositories using both experimental and *in silico*-generated data.

The third major challenge for an improved coverage of the metabolome is the efficient treatment of data obtained from multiple analytical platforms. Processing and analyzing these datasets, integrating the results at the biological level, and attempting to establish relationships between the discriminant features are key aspects of future research.

In the present article, we will cover these three major challenges by first giving an overview of the main separation techniques used in MS-based metabolomics. We will then emphasize the relevance of combining different separation modes for MS-based metabolomics to obtain a more complete vision of the biological phenomena under study, finishing by an overview of the data mining approaches making it possible.

2. Liquid Chromatography

Due to its versatility, robustness and popularity within the analytical chemistry community, LC has become the leading separation technique in metabolomics. By providing a separation step that increases resolution and decreases matrix effects before MS detection, LC has remarkably contributed to improving the value of the information retrieved in metabolomics.[1,2]

The last two decades have witnessed the rise and spread of UHPLC conditions, allowing fast separations with a large peak capacity thanks to the use of shorter columns, smaller particles, and systems able to cope with the elevated back pressures thereby generated.[3,4] LC separates complex mixtures based on the different interactions of each component with the stationary and mobile phases. Thus, the nature of such LC elements will determine the driving mechanism of the separations. In LC-based metabolomic applications, metabolites are most commonly separated on the basis of their polarity. Hence, different separation modes have been adopted in this context, with reversed-phase liquid chromatography (RPLC) and hydrophilic interaction liquid chromatography (HILIC) being the most relevant modes but also including, although to a lesser extent, mixed-mode chromatography (MMC), ion-exchange chromatography (IEC), and normal-phase liquid chromatography (NPLC).

RPLC best fits applications in which moderately to highly apolar compounds must be separated. Due to its ruggedness, good retention time repeatability and straightforward method development, it remains the first option for many practitioners.[5] In most cases, compounds are eluted using $MeCN:H_2O$ gradients containing low concentrations of formic acid. In addition to classical stationary phases such as C18, other chemistries and endcappings presenting a certain polar character (and, thus, improving the retention of more hydrophilic compounds), such as Waters HSS T3, have become very popular. RPLC including isopropanol in the mobile phase is also a preferred technique for lipidomic profiling. RPLC coverage can be extended to more polar or charged compounds by means of two major approaches: analyte derivatization and the use of ion-pairing agents. In the former, polar analytes are linked to a derivatization agent, which renders them more hydrophobic, thus making their retention and separation possible under RPLC conditions. In the latter, an amphiphilic ion-pair reagent is added to the separation to facilitate the interactions between the ionic analytes and the hydrophobic stationary phase.

Nevertheless, these approaches are not exempt from drawbacks, such as poor repeatability in the derivatization reactions or the contamination of MS by ion-pairing agents.

To increase metabolite coverage, HILIC appeared to be a natural alternative to complement the selectivity of RPLC.[6] In this case, more polar compounds can be successfully separated, commonly using MeCN:H$_2$O gradients involving volatile additives to regulate pH and ionic strength, such as ammonium formate and acetate. Amide- and sulfobetaine-based zwitterionic stationary phases are the most widely used ones. While allowing the analysis of polar compounds, HILIC methods require a deeper knowledge of LC fundamentals and optimization, and longer equilibration times. Retention times are also less repeatable and reproducible, and harder to model. In regard to the analysis of polar compounds bearing carboxylate, phosphate and sulfate moieties, a commonly found issue is the loss of metabolites due to their adsorption on stainless steel surfaces. To minimize such problems, different options have been made commercially available, such as passivating the surfaces using a medronic acid solution as a mobile phase additive or applying a low-binding coating to instruments and columns.

MMC is another LC mode that is slowly gaining momentum in metabolomics as manufacturers release new column chemistries.[7] It is based on the coexistence of several ligands or combinations of ligands/endcappings on the same column, thus enabling different modes of interaction between the analytes and the particle surface.[8] They present the advantage of simultaneously retaining metabolites with different properties by interacting with them through diverse mechanisms. Nevertheless, the complex and assorted nature of the chromatographic processes concomitantly taking place in such columns requires a comprehensive method development process to ensure optimal performance.

Since each chromatographic mode will perform better when separating molecules within a certain polarity range, no single LC method can provide a complete overview of the whole metabolome. This fact has led to the search for approaches in which several LC modes are used to broaden the chemical space that can be covered by metabolomic experiments.[9] In this direction, the use of 2D chromatography should be a good candidate to combine the benefits of two orthogonal chemical selectivity types coming from different columns.[10] Nevertheless, a number of drawbacks, such as technical complexity, mobile phase compatibility, analysis time and analysis speed in the second dimension, have notably hampered the broad adoption of such an approach.

Alternatively, a much more widespread approach involves independent chromatographic analyses followed by the use of a number of resources to merge their individual outputs. In such cases, the same set of samples is analyzed under different LC–MS conditions (Fig. 1) (chromatographic modes, ionization polarity, *etc.*) and the resulting data are combined and jointly interpreted. The first and most straightforward option is the combination of orthogonal LC and ESI modes. For instance, the samples are run in RPLC with positive ESI, followed by a second sequence batch in which the samples are separated in HILIC conditions and ionized in the negative mode. This is a sensible combination, as some of the most relevant polar metabolites (such as those involved in energy metabolism) bear negative charges. Although this combination is based on an educated choice, it presents a bias towards *apolar, positively ionizable* compounds and *polar, negatively ionizable* compounds. Therefore, more comprehensive and systematic methodologies must be developed.

With this goal in mind, several studies have been conducted to quantitatively determine the value of the chemical information gathered by each LC–MS method in a multimode acquisition environment. Our group has recently proposed an algorithm allowing us to compare the performance of different LC–MS modes on a panel of benchmarking analytes.[11] This is done by means of a score assigned to the peak obtained for each compound on each of the analytical platforms. This score takes into account LC retention, peak shape, peak intensity and signal-to-noise ratio. Five different LC–MS methods were compared: RPLC with positive and negative ESI, amide-based HILIC (aHILIC) with positive and
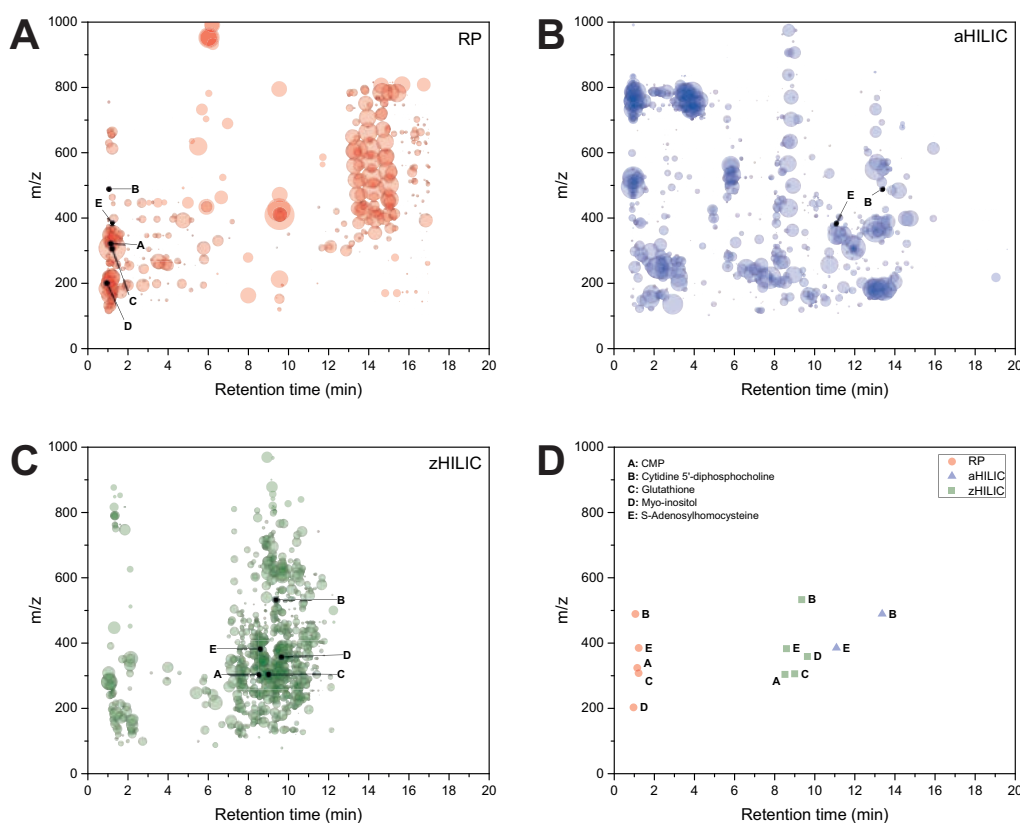

Fig. 1. An example of the complementarity achieved by combining several LC modes is shown. Some compounds poorly retained in RP (A) cannot be either qualitatively (because of lack of retention time) nor quantitatively (due to the extent of the ion suppression at the dead time) analyzed. Alternative LC methods showing complementary retention properties can thus be used. Amide HILIC (B) and zwitterionic HILIC (C) are two examples of this. Subplot D shows the position of each one of the compounds on the different m/z vs. retention time (RT) maps. In D, changes on the position of the same compound along the m/z axis are caused by different adducts detected with each technique. For A, B and C, the area of the bubbles is proportional to the intensity of each feature. Reproduced from ref. [2] with permission.

negative ESI, and zwitterion-based HILIC (zHILIC) with negative ESI. Since many metabolites can be detected on more than one LC–MS platform, the question arises as to which is the optimal combination of methods yielding the best coverage with the smallest number of analyses. By using the scoring approach, we found that the most efficient combination of methods was comprised of zHILIC ESI-, aHILIC ESI+ and RPLC ESI+, which together were capable of covering 95% of all detected metabolites with acceptable analytical behaviour. Interestingly, adding the results from the remaining two analytical platforms only marginally increased the coverage rate while involving a remarkable increase in the analysis time.

In addition to chemical diversity, LC-based metabolomics must face the challenge of improving its sensitivity to cope with low-concentration metabolites and volume-restricted samples. Sensitivity enhancement in LC–MS can be achieved by working in the nano-ESI mode, which dramatically increases the ionization yield. Notwithstanding, nano-ESI requires flow rates far below those typically used in LC; thus, it has led to the use of chromatographic columns with reduced internal diameters and eventually to the development of micro/nano-LC.[12,13] Such LC formats afford a remarkable gain in sensitivity by using reduced flow rates (10–1000 nL/min) on small internal diameter capillaries (from 10 μm for nano up to 1000 μm for micro) but at the expense of longer gradient times to retain adequate efficiency values under such operating conditions. Trapping columns are also commonly used in nano/micro-LC setups to load larger amounts of sample in the system without compromising the peak shape.

3. Capillary Electrophoresis

Arguably, capillary electrophoresis (CE) has been considered a less suitable technique for metabolomic analyses due to its technical challenges, and lower migration time reproducibility when compared to chromatographic techniques.[14–16] Among other reasons, CE presents technical challenges rendered to the fluctuations of electroosmotic flow (EOF) influenced by the coating of silica,[17] thus limiting its adoption. Remarkably, the issue of migration time variability has been successfully addressed in recent years using effective electrophoretic mobility (μ_{eff}), allowing the use of migration as a robust annotation and identification parameter.[18,19] Capillary zone electrophoresis (CZE), as the most popular mode of CE coupled to MS, is currently a mature analytical technique allowing the effective profiling of highly polar and ionizable compounds. A competitive advantage of CE-MS is the reduced influence of matrix effects. CE-MS is considerably less influenced by ion suppression thanks to the absence of leaking material and matrix build-up from a stationary phase. This results in avoiding the adverse effect on detector response due to reduced ionization efficiency for analytes of interest, resulting from the presence of species that compete for, or inhibit, their efficient ionization.[20,21]

Compared to the more widespread LC-MS, CE-MS remains less popular in untargeted metabolomics due to a series of drawbacks, such as sensitivity, CE-MS interfacing robustness and the lack of adapted solutions to deal with migration time variability. From an instrumentation and methodological standpoint, CE-MS in metabolomics has gone through a number of advancements aiming at helping to obtain a stable, reproducible and sensitive analysis. First, Soga et al. in Japan reported a method during which a few hundreds of charged metabolites were separated by cationic and anionic methods.[22–24] The separation and detection of anionic metabolites requires CE separation in the reverse-polarity mode and ionization in negative ESI. Under such conditions, stainless steel needles have shown corrosion issues due to oxidation, and thus, only platinum needles are adequate to perform such analyses. Further improvements were made by Boizard et al., who enhanced the stability of the CE-MS interface

by designing a bevelled capillary tip. When full coverage of highly polar and ionic metabolites must be attained, both ionic and cationic metabolites need to be analyzed in the CE-MS platform. Sheathless porous-tip interfaces and modified sheath-liquid (SL) interfaces are other promising candidates for the enhancement of coverage in anionic metabolic profiling, while the combination of low-pH separation conditions along with a fused-silica capillary is a robust method for cationic metabolic profiling.[25] A reliable interface is also equally important. Ramautar et al.[26] and Boizard et al.[27] developed strategies for improving the sensitivity, reliability and stability of the MS interface, which led to enhanced metabolite coverage. The former employed glioblastoma cell lines using a sheathless capillary interface to improve sensitivity and thus the overall coverage. In addition, Sweedler et al. employed a CE-ESI-MS technique, which matches well to single-cell assays thanks to its low sample-volume requirements at the picoliter range and its low detection limits.[28] More precisely, Sweedler and co-workers detected intracellular metabolites from a single neuron using a field-amplified sample injection (FASI) CE-ASI-MS. Throughout this approach they achieved between 100- and 300-fold enhancement in detection limit.

Remarkable contributions for the use of quality assurance and quality control (QC) samples, as well as for the identification of metabolites, have recently been made. McKibbin's group demonstrated the ability of CE-MS to analyze large cohorts with great robustness by including multiple samples, QCs and calibration standards in a single CE run by using a multiple-injection approach.[29] Drouin et al. proposed a generic CE-MS approach by performing normal and reverse-polarity CE separations to analyze 596 metabolites, and more than 450 compounds were detected. A large database relying on the μ_{eff} was built, allowing for the straightforward annotation of detected features in biological samples (Fig. 2).[30] Mamani-Huanca et al. presented an annotation tool based on their in-source fragmentation pattern and incorporated it into CEU Mass Mediator.[31] They also established a 5% tolerance of relative migration time (RMT) for all compounds they measured, whereas 10% tolerance was established for compounds separated with EOF. More recently, interlaboratory work[32] involving 20 different CE-MS platforms across 17 different laboratories assessed the reproducibility and identification capability of CE-MS by using μ_{eff} instead of RMT. It was demonstrated that the mismatch between experimentally measured μ_{eff} values and database entries was much smaller compared with the use of RMT.

Multiple CE-MS protocols have been developed to support quantitative metabolomic measurements in a wide range of samples, including urine,[33] bacteria,[34] neurons[35] and the brain.[36]
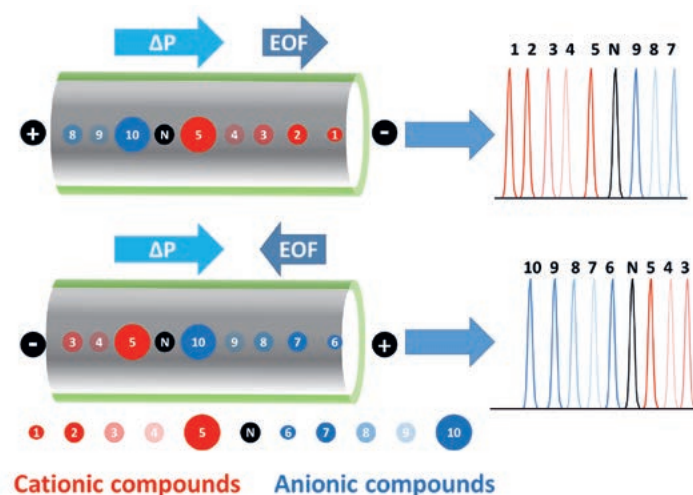

Fig. 2. Schematic representation of migration order under normal and reverse polarities. Reproduced with permission from ref. [30].

Kok *et al.* carried out a comparison study between CE-MS and HILIC-MS aimed at studying the main analytical parameters, including metabolite coverage, when measuring anions in urine.[37] The authors reported that when they analyzed rat urine with HILIC-MS, approximately 1360 molecular features were detected compared to 347 features with CE-MS. More interestingly, approximately 203 molecular features were exclusively found when CE-MS was employed, which proves the selectivity capabilities of this technique. The same team also studied the influence of various background electrolytes (BGEs) and SL additives to enhance metabolite coverage in the negative ionization mode.[38] Their findings showed that triethylamine (pH = 11.7) in the BGEs and SL resulted in lower limits of detection and greater metabolome coverage than common negative ionization CE-MS methods, where ammonia buffers are used.

The many advantages of using CE-MS in metabolomics have been shown in the recent years specially in the context of its ability to cope with volume-restricted biological samples. Another strong potential of CE-MS is the ability to be a complementary separation technique in multi-platform setups allowing the comprehensive analysis of highly polar and charged metabolites, especially amino acids, carbohydrates and nucleotides.

4. Supercritical Fluid Chromatography

In addition to the use of LC and CE previously discussed, supercritical fluid chromatography (SFC) can also be considered a valuable alternative analytical strategy in the field of metabolomics to improve metabolome coverage. Indeed, SFC offers excellent kinetic performance due to the low viscosity of the mobile phase and low environmental impact (the mobile phase is mostly composed of supercritical CO_2).[39] In SFC, any type of stationary phase can be used (*i.e.* polar, apolar, or aromatic), while the mobile phase is typically composed of supercritical CO_2 (the latter is used above its critical point, temperature >31°C and pressure >72 bar), mixed with any type of organic solvent (from hexane to methanol) and some additives (water and salts). Since it is possible to use any type of stationary phase and mobile phase in SFC, this technique is highly versatile.[40] However, in common practice, a polar stationary phase is frequently used today in combination with a mobile phase composed of a mixture of CO_2, methanol and additives. Under these conditions, the retention mechanism is based on polar interactions, including H-bond, dipole–dipole and ionic interactions.[41]

Since 2011, UHPLC-like instruments and columns packed with sub-2 μm particles have been introduced in SFC. This modern SFC approach has been described as UHPSFC, which stands for ultra-high performance supercritical fluid chromatography.[42] Moreover, thanks to the addition of a higher proportion of cosolvent in the mobile phase (up to 100%) and the addition of water (<5%), highly polar compounds can be successfully analyzed, and some applications also report the simultaneous analysis of lipophilic and hydrophilic substances in a single analysis, which is a significant advantage for metabolomic applications.[43] Another important aspect to make SFC readily applicable to metabolomics is the need for dedicated interfaces allowing the hyphenation of SFC and MS. There are two main reasons why SFC cannot be easily coupled to MS.[44] First, it is well known that supercritical fluid will depressurize/decompress when it is no longer under the control of the back pressure regulator (BPR), as in MS. This can lead to potential analyte precipitation in the tubing located before the MS inlet and the severe loss of chromatographic efficiency.[45] Second, when the mobile phase is mostly composed of supercritical CO_2 (at the beginning of the gradient), the MS sensitivity is expected to be poor with ESI sources due to the absence of a proton source. For these reasons, various interfaces were introduced over the years to circumvent these two limitations. Among those different interfaces, two are commercially available, namely, the pre-BPR splitter with sheath pump interface and the BPR and sheath pump with no splitter interface.[46] In our opinion, the pre-BPR splitter with a sheath pump interface is the most interesting splitter when using ESI as an ionization source. Indeed, with this interface, the amount of MeOH entering the ionization source is always between 150 and 400 μL/min regardless of the settings and analytical conditions, which is optimal for ESI operation.[47]

In the last few years, several applications of SFC-MS in the field of metabolomics have been reported, with UHPSFC-MS using a pre-BPR splitter with a sheath pump interface with a focus on lipid profiling and lipidomics.[48,49] These applications of UHPSFC-MS in metabolomics have been summarized in a recent review.[50] In the present contribution, we wanted to highlight the work made recently in our laboratory dealing with the characterization of the metabolome, from hydrophobic to hydrophilic metabolites.[51,52] The results reported in Fig. 3 were obtained with UHPSFC instrumentation using a bare silica core-shell column and a mobile phase varying from almost pure CO_2 to pure MeOH in the presence of water and additives, thus offering excellent flexibility in compound solubility and elution. It is, however, important to keep in mind that under such extreme gradient conditions, the mobile phase is transitioning from a supercritical to subcritical and finally liquid state. Since no phase transition is observed between these states, this is clearly not an issue. Fig. 3A shows the possibility of analyzing various types of hydrophilic and lipophilic molecules belonging to very different classes of metabolites under these conditions, with suitable peak shapes and adequate retention. This result is confirmed in Fig. 3B, showing the simultaneous analysis of a fatty acid (log P of 9.3) and a trisaccharide (log P of –6.3) under SFC conditions, with both metabolites having adequate retention.[51] This clearly highlights the huge potential of modern SFC in metabolomics as a replacement strategy for both RPLC and HILIC. Next, a larger library of standards, which encompasses 597 metabolites, was analyzed under UHPSFC-HRMS conditions. As illustrated in Fig. 3C, most of the metabolite categories were effectively analyzed. However, reduced performance was reported for phosphate-containing metabolites and nucleotides (also containing phosphate groups), explained by possible precipitation, adsorption of phosphorylated metabolites on the column walls and frits, or lack of elution of those metabolites still retained by the column. In the end, 66% of all library metabolites were successfully detected. In addition, it is important to mention that the retention of all these metabolites was always adequate, as reported in Fig. 3D (only two of the 394 detected metabolites were eluted in the first quarter of the SFC chromatogram, with limited retention).[52]

Despite these promising results, the potential of modern SFC-MS for untargeted metabolomics remains largely unexplored, as very few applications have been reported in the literature thus far. In addition, there is still no systematic comparison of SFC-MS and LC–MS performance in the field of metabolomics using the same MS device and samples.

5. Modes of Acquisition in MS Detection

With MS coupled to any of the aforementioned separation techniques, high-resolution full scan experiments allow the acquisition of a wealth of information about the composition of the investigated samples. Nevertheless, the acquisition of fragmentation data provides a number of additional advantages that can include structural information, increased sensitivity and selectivity. In such experiments, a trade-off has to be found between the amount and quality of information, acquisition speed and sensitivity.

In the case of untargeted metabolomics, the major aim of the fragmentation experiments is to retrieve structural information enabling a more reliable annotation of the features found in the
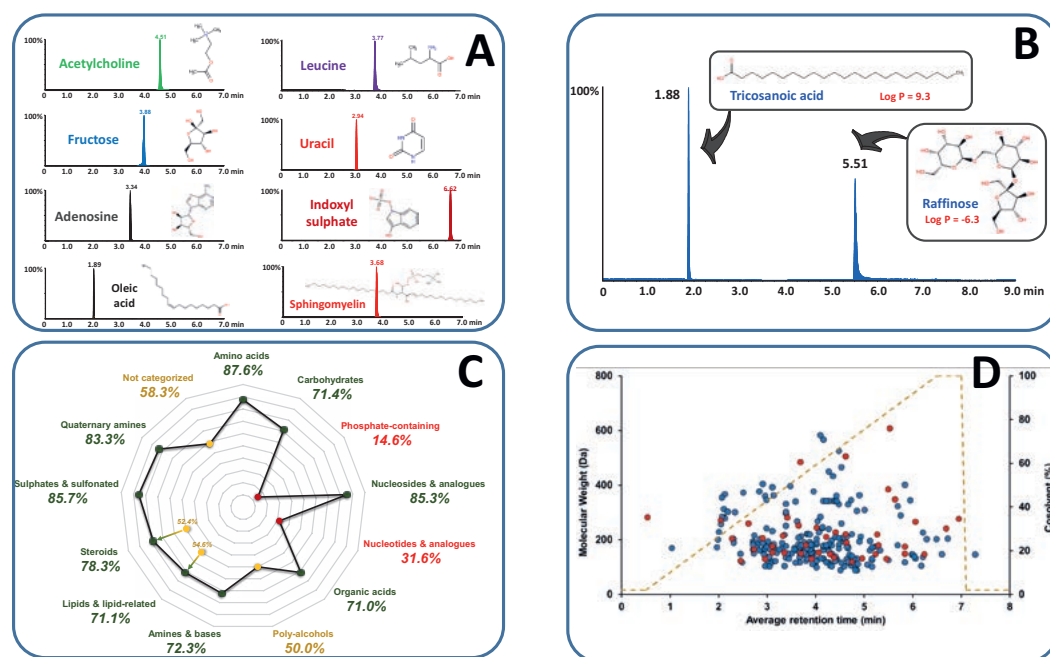
Fig. 3. Applications of SFC-MS highlighting the potential of the technique in metabolomics. (A) Analysis of various hydrophilic and lipophilic metabolites, (B) simultaneous analysis of hydrophilic and lipophilic metabolites, (C) metabolite coverage by metabolite families, for a library of 597 metabolites, (D) elution conditions for the library metabolites successfully eluted. Adapted from refs [51] and [52].

samples.[2] The different acquisition strategies developed to cope with this challenge can be grouped into data-dependent acquisition (DDA) and data-independent acquisition (DIA) approaches. In both cases, the aim is to enrich full scan MS1 data with fragmentation information acquired in MS2 experiments for as many compounds as possible.[53] Thus, a cycle of acquisition usually consists of a MS1 scan first, followed by one or more MS2 scans.

In DDA, the first MS1 scan is analyzed to look for a predefined number of the most intense ions, which are then isolated and fragmented in successive MS2 experiments to obtain their individual spectra.[54,55] The main advantage of DDA is that the link between the fragments and the precursor ion is established, as long as no isobaric species are simultaneously eluted and isolated for fragmentation. The main drawback of this approach is that only a limited number of precursors (usually the most abundant ones) can be selected in each cycle for MS2 analysis. Since the number of precursors that can be selected from each MS1 scan is limited by the acquisition speed, fast instruments (such as Q-Tof or orbitraps at low-resolution MS2 settings) are preferred. Considering that chromatographic peaks can only be built from MS1 data points, a typical consequence of DDA acquisition is the drop in quantitative performance due to the lower number of points per peak available in the LC dimension. This drawback is usually circumvented in metabolomics by acquiring only MS1 data in samples and keeping DDA acquisition with annotation purposes restricted to pooled QC samples. To increase the number of MS2 spectra acquired, it is common practice to use an exclusion list that will exclude masses present in blanks, such as contaminants, thus focusing fragmentation efforts on masses of biological interest. This concept can be taken even further by measuring the same sample several times and excluding at each new round masses previously fragmented, a process that is now fully automated in the AcquireX workflow (Thermo).

In DIA analyses, instead of individual precursors, sets of ions within broad m/z windows are simultaneously fragmented.[56] When the selected window covers the entire full mass range of the instrument, the approach is called either all ion fragmentation (AIF) or MS^e[57] depending on the manufacturer, the collision mechanism and the collision energy levels applied. Although fragmentation information is acquired for all the compounds in the sample, the obvious disadvantage of DIA approaches is the

complexity of linking a certain precursor observed in the MS1 scan to their corresponding fragments in the MS2 spectra, which usually requires complex algorithmic processing.[58] To reduce such complexity, narrower m/z isolation windows can be established to decrease the number of precursors fragmented in each MS2 acquisition. In the SWATH methodology[59] (Sequential Window Acquisition of All Theoretical Mass Spectra), a number of MS2 experiments are performed after a MS1 acquisition, sequentially moving the position of the isolation window in each MS2 acquisition until the whole mass range is covered. By doing so, spectral complexity drops since the precursors are limited to the m/z range selected by each isolation window, while all the possible precursors are fragmented in each cycle since the isolation windows move to cover the whole spectral range. SWATH is meant to afford a good balance between spectral complexity and information retrieval from the samples.[60]

Ion-mobility spectrometry (IMS) is gaining popularity among MS practitioners as it becomes available in an increasing number of instruments.[61] This additional separation dimension, based on the charge, shape and volume of the ions travelling in the gas phase, serves not only to determine the cross-collisional section (CCS) of the molecules but also to afford additional selectivity to the MS detection step.[62] When IMS is placed before the collision cell, it can help simplify MS2 spectra in AIF experiments, since fragments can be more easily linked to their precursors by means of their different IMS values (Fig. 4). In instruments where IMS separation takes place after the collision cell, it can help characterize the fragments themselves by allowing the determination of their CCS values.

6. Annotation of Untargeted Datasets

Annotation relies on the partial to complete structural elucidation of metabolic features, which are ions characterized by a m/z and by additional information such as RT, MS/MS spectra or CCS. The first step in the annotation process is to assign one or several plausible molecular formulas with a monoisotopic mass that most closely matches the experimental m/z of the feature. Depending on the resolving power of the mass spectrometer, the higher the value is, the lower the number of potential formulas. RT, MS/MS or other orthogonal parameters will provide further hints about the feature's identity. A typical untargeted metabolom-
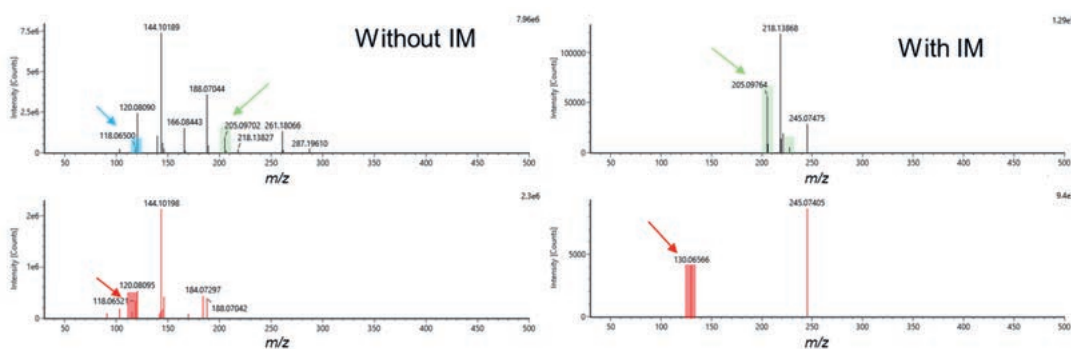
Fig. 4. Example of MS and MS/MS spectra for tryptophan (M+H)+ in hydrophilic interaction chromatography with amide (aHILIC) electrospray ionization (ESI)+ for all ion fragmentation (AIF) ramp 10–60 eV with and without IM. The precursor ion m/z 205.0976 (green), in-source fragment m/z 118.0650 (blue), and fragment ions (red) m/z 118.0652 (AIF) and m/z 130.06566 (IM-AIF) were retrieved using Waters data acquisition and processing software UNIFI. In-source fragment at m/z 118.0650 is filtered out when the IM cell is activated. Taken from ref. [62].

ic dataset usually results in a list of >20000 features, of which an estimated 3% are deemed unique and valid metabolites, while the rest are background ions, adducts, fragments, or isotopes.[63] The possibility of false-positives in the annotation process is therefore very important.

The Metabolomics Society has defined criteria for communicating the confidence level in annotation.[64] The scale spans from 4–fully unannotated feature to 1–identified structure. A recent proposal added the full structural elucidation of chiral compounds as level 0.[65] Therefore, a feature that has been annotated using only its m/z value will receive level 4. Confidence can increase to level 3 if there are hints that it belongs to a specific chemical class, for example, by elution at a similar RT as a standard of the same chemical class. Level 2 involves a putative identity that has not been confirmed by comparison to reference standards as in level 1. Level 1 confidence is granted after matching to two orthogonal parameters, *e.g.* accurate mass (AM) and RT or AM and MS/MS spectrum, measured in the same laboratory using a reference standard. Annotation based on AM and RT alone is, however, not sufficient in complex matrices where several metabolites elute very closely or coelute; it is therefore good practice to use a third orthogonal parameter to increase confidence.[66] Additionally, technological advances can provide higher structural resolution, calling for a more refined scale system between levels 1 and 0. The Metabolomics Standards Initiative metric is thus currently under revision to take these considerations into account.[67]

Robust annotation of features measured in complex samples cannot yet be fully automated and requires several rounds of curation to reduce the number of false-positives while expanding coverage. In the first step, features are usually matched to an in-house library containing experimentally acquired RTs and fragmentation spectra from standards as well as their theoretical AM. In general, deviations from library values should lie <10 ppm for AM, <0.2 min for RT and at least three fragments including the parent ion should be matched to the reference spectra.[68,69] Larger deviations might not necessarily exclude a particular annotation, but the reasons should be reported in detail. To ensure a high analytical quality, the curator should also examine the chromatographic shape, its consistent integration across samples, signal-to-noise ratio, which is usually >5, and concomitant presence of typical adducts[70] or fragments. Annotation of unknown features not present in the in-house library is possible by querying external repositories for retrieving AM and fragmentation patterns.[55,65] For example, MassBank,[71] Mass Bank of North America MoNA (https://mona.fiehnlab.ucdavis.edu), the Global Natural Products Social Molecular Networking GNPS,[72] the RIKEN MSn spectral database for phytochemicals ReSpect (http://spectra.psc.riken.jp) or HMDB[73] offer impressive collections of open-access experimental spectra, summing to >27'000 unique compounds.[74] The

commercial METLIN[75] and NIST (https://chemdata.nist.gov) libraries offer well-curated spectral coverage and can be accessed through several data processing vendor software. In complement to experimental data, *in silico* spectra can be generated through rule-based (MS-FINDER[76]), combinatorial (MetFrag[77]) or machine learning (CFM:ID[78] or CSI:FingerID[79,80]) approaches. The wide variety of structures encountered in untargeted metabolomics is nevertheless a challenge for *in silico* predictions, especially when targeting true unknowns, for which training sets for machine learning might be too small or fragmentation patterns still not well understood. To reduce the tremendous number of false-positives than can arise from using automated queries, prior examination of the correlation coefficient of chromatograms across samples is an efficient method to quickly discard artefacts and pinpoint uncovered areas in untargeted data,[55] as it is conveniently implemented in MSDIAL, for example.[81]

Annotations of untargeted lipidomics datasets are hampered by the high number of lipid structures found in living organisms. In the LIPIDMAPS database, greater than 45'000 individual lipids are reported, with half of them being curated and the others being computer-generated structures.[82] However, the total number of unique lipid structures is estimated to be 100'000 and even reaches 1'000'000 when taking into account oxidized species.[83] This complexity is a challenge in untargeted lipidomics, where a high number of very structurally similar lipids closely elute, requiring high-resolution mass spectrometers to assign a candidate formula.[84,85] Lipids display particular ionization behaviours characteristic of their families. For example, in the positive mode, neutral lipids such as diacylglycerols and triacylglycerols are more often encountered as adducts of ammonia or sodium, and steroids tend to display water losses, while in the negative mode, acetate and formate adducts are commonly depending on the mobile phase composition.[84] Similarly, acyl carnitines and endocannabinoids ionise in the positive mode, while the negative mode is favoured for fatty acids, phosphatidylinositols, sulfatides and phosphatidylglycosides. Reverse-phase chromatography enables the application of the equivalent carbon number (ECN) rule to support annotation.[84,86] The ECN is calculated as the total number of carbons in the fatty acyl chains minus 2x the number of double bonds. Hence, lipids with longer fatty acyl chains will elute later than lipids with shorter acyl chains, and the higher the number of double bonds, the shorter the retention time. Deviations from this pattern can be encountered for lipids with a combination of unsaturated and highly saturated acyl chains, which display later RTs, and oxidized lipids that elute earlier than their non-oxidized counterparts.[87] However, RT is far from sufficient for confident lipidomics annotation and requires the use of fragmentation patterns for structural elucidation. Lipids display characteristic fragments in the positive mode to confidently assign the correct lipid

class, such as ion 184 of choline-containing lipids (phosphatidyl-cholines and sphingomyelins) or the neutral loss of m/z 141 for phosphatidylethanolamines.[85] The level of resolution in the lipid structure thus greatly varies according to the amount of spectral information gathered. The LIPIDMAPS consortium has created a unified annotation nomenclature,[88] widely adopted by the lipidomics community and now integrated in most lipidomics workflows.[89] In contrast to generic metabolomics, lipidomics benefit from a high structural similarity within each lipid class, making it amenable for robust *in silico* fragmentation predictions. The LipidBlast[90] database generated in this way >120'000 *in silico* spectra covering all major lipid classes. When combining MS/MS similarity scoring using *in silico* spectra with decision trees based on the presence of specific lipid fragments and adducts, the annotation rates can dramatically increase. For example, MS-DIAL,[81] Lipostar[91] or the combination of MZmine[92] and Lipid Data Analyzer 2[93] offer such annotation workflows.[83]

7. Data Processing

Untargeted MS metabolomic data acquisition usually leads to very large raw datasets, including several thousand ion features. This massive volume of data potentially contains an important amount of information for characterizing the complexity of biological samples, but a workflow involving several steps is needed for knowledge discovery, as depicted in Fig. 5. Since unwanted sources of variability, such as instrumental variations and intrinsic biological diversity, may impact the measured profiles, preprocessing the raw data is mandatory to ensure comparability between samples and extract proper metabolic knowledge.[94] To this aim, an injection sequence including technical samples and designed according to a number of widely adopted best practices can serve as an efficient basis to correct or reduce these undesirable effects by removing poor quality or artefactual signals unrelated to the study samples and filtering data that do not meet specific analytical quality criteria.[2,53]
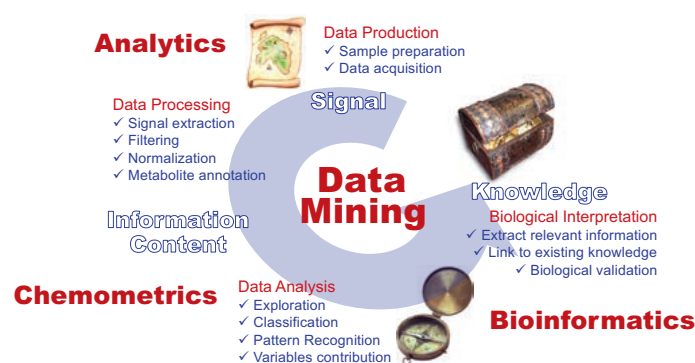

Fig. 5. Knowledge discovery workflow in untargeted metabolomics.

Monitoring data acquisition using QCs currently constitutes a widely adopted practice. QCs should be representative of the biological matrix subjected to analysis, and a pragmatic approach consists of taking a small volume from all samples of the study to create a synthetic average sample. QCs are then included in the analytical sequence at regular intervals and across batches to control analytical stability in real conditions. Feature-wise correction or filtering strategies are then applied to discard noisy or saturated signals.[95] This may include removing specific areas of the chromatogram (*e.g.* injection peak) or spectral ranges of poor chemical information, discarding features with a large proportion of missing values, and filtering features with high variability (*e.g.* relative standard deviation greater than 20%) in the repeated QC injec-

tions. QCs with a known dilution factor can also be used to assess the response to dilution as a criterion of signal reliability.[96] Since a gradual decline in overall signal intensity, potentially leading to artificial differences between samples, is often observed during long analytical sequences, QCs can be used as anchor points to correct such a drift using regression models.[97] Variations occurring between batches can be corrected using a similar strategy.[98] Intrinsic differences in concentration between the study samples, due to biological or technical variations, may also preclude the reliable comparison of metabolic signatures. To circumvent this issue, a normalization factor can be used to correct for such a size effect. It can be derived from a reference parameter (*e.g.* number of cells or amount of protein) or well-identified compound (*e.g.* creatinine) or based on a data-driven statistical approach.[99]

8. Data Analysis

The comparison of metabolic signatures can be performed using either univariate or multivariate statistical methods.[100] Univariate techniques encompass classical one-variable-at-a-time comparisons and are often carried out to investigate identified metabolites of specific interest. Fold changes can be used to investigate the ratio between average levels observed in different experimental groups, while hypothesis testing can be carried out to estimate the significance of the observed differences. Both parametric and nonparametric approaches can be used to compare groups of observations (*e.g.* case and control),[101] and the choice of the test depends on the characteristics of the data (dependence, distribution and scedasticity). Since a large number of features are measured, multiple hypothesis tests are often performed in parallel. As a consequence, correction strategies have to be applied to control the frequency of false-positive results (*e.g.* Bonferroni correction[102] or the Benjamini-Hochberg false discovery rate approach[103]).

Multivariate methods are particularly well suited to analyze untargeted metabolomic datasets, as they involve multivariate models accounting for all variables in the analysis, thus potentially revealing relationships between subsets of metabolites.[104] For that purpose, dimensionality reduction strategies are implemented to extract biologically meaningful trends that may be hidden in the mass of data and offer meaningful representations of the data.[105] In practice, scaling strategies are often used to ensure the comparability of the measured signals when the variables have very different intensity ranges. Different modelling strategies can then be chosen to explore metabolic variations in an unsupervised manner or to focus on differences between experimental groups. The first step of multivariate data analysis is often carried out using unsupervised methods to check data consistency. For that purpose, principal component analysis (PCA) is a widely used approach that aims at extracting the major structures of variability in the data by building new synthetic axes maximizing the variance explained. PCA results can be easily visualized as graphical outputs to assess the distribution of the samples, systematic trends or potential outliers. This represents an efficient exploratory tool, but specific modulations with relevant biological meaning may be masked by greater sources of variability in the data. In this context, supervised modelling approaches can be used to focus on a given part of the biochemical information (*e.g.* differences between experimental groups). With this aim, partial least squares (or projection to latent structures, PLS) regression is a potent method to efficiently handle datasets characterized by a large number of correlated variables but a limited number of observations.[106]

9. Biological Interpretation

The ultimate aim of metabolomics is to provide reliable mechanistic hypotheses to explain the observed changes.[107] It is there-

fore essential to place the statistical results obtained from metabolomic experiments in their biological context. To this aim, bioinformatic databases, *e.g.* KEGG[108] MetaCyc[109] and HMDB,[73] and metadata query tools designed to describe and explore metabolic pathways and networks constitute very potent solutions to translate lists of modulated metabolites into biological information.[110] These tools include pathway enrichment analysis, usually carried out based on overrepresentation to rank biological processes using a statistical criterion. Metabolic networks represented as graphs constitute today the gold standard to describe the hierarchy of biochemical reactions that define metabolism. An investigation of different topological features of biological meaning, such as path lengths, the degree of centrality and the clustering coefficient, can help to highlight hubs and modules of relevant metabolites.[111] Bioinformatic solutions for pathway mapping and network analysis include MetExplore,[112,113] MetaMapR,[114] BiNChE,[115] MBRole,[116] MPEA,[117] Mummichog,[118] and MetaboAnalyst.[119]

10. Conclusion

Analytical developments in metabolomics should improve the ability to detect a large number of metabolites at low concentrations as well as at high concentrations in a complex matrix. In that respect, separation techniques such as LC/UHPLC, SFC, and CE constitute mandatory tools to separate analytes and allow optimal conditions for their detection with MS. The study of fundamental aspects of separation science is therefore still essential to increase the overall coverage needed in metabolomics for various fields or research. Developments in MS instruments are also required to increase the annotation confidence, such as high resolution, capacity to increase the number of fragmentation events, and incorporation of different fragmentation strategies, such as ultraviolet photodissociation. As high-resolution MS is expected to help the identification of unknown analytes, untargeted approaches, instead of focusing on a restricted number of targeted compounds, still constitute promising tools for global monitoring.

Overall, state-of-the-art analytical technologies combined with computational techniques will make a difference in untargeted metabolomics, providing high-throughput analysis, large metabolite coverage, accurate quantification and low-cost analysis. Technical developments will bring progress in the field aiming at providing critical solutions in drug development and biotechnology. Finally, the integration of other omics data (genomics, transcriptomics, proteomics) in different pathway analysis tools will open up the way to overcome bioinformatic challenges in untargeted metabolomics.

[1] J. Ivanisevic, E. J. Want, *Metabolites* **2019**, *9*, 1, https://doi.org/10.3390/metabo9120308.
[2] J. Pezzatti, J. Boccard, S. Codesido, Y. Gagnebin, A. Joshi, D. Picard, V. González-Ruiz, S. Rudaz, *Anal. Chim. Acta* **2020**, *1105*, 28, https://doi.org/10.1016/j.aca.2019.12.062.
[3] S. Fekete, J. Schappler, J. L. Veuthey, D. Guillarme, *TrAC Trends Anal. Chem.* **2014**, *63*, 2, https://doi.org/10.1016/j.trac.2014.08.007.
[4] S. Fekete, J. L. Veuthey, D. Guillarme, *J. Chromatogr. A* **2015**, *1408*, 1, https://doi.org/10.1016/j.chroma.2015.07.014.
[5] 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, *Nat. Methods* **2021**, *18*, 747, https://doi.org/10.1038/s41592-021-01197-1.
[6] C. Virgiliou, I. Sampsonidis, H. G. Gika, N. Raikos, G. A. Theodoridis, *Electrophoresis* **2015**, *36*, 2215, https://doi.org/10.1002/elps.201500208.
[7] L. Perez de Souza, S. Alseekh, F. Scossa, A. R. Fernie, *Nat. Methods* **2021**, *18*, 733, https://doi.org/10.1038/s41592-021-01116-4.
[8] L. Wang, W. Wei, Z. Xia, X. Jie, Z. Z. Xia, *TrAC Trends Anal. Chem.* **2016**, *80*, 495, https://doi.org/10.1016/j.trac.2016.04.001.
[9] S. Wernisch, S. Pennathur, *Anal. Bioanal. Chem.* **2016**, *408*, 6079, https://doi.org/10.1007/s00216-016-9716-4.
[10] M. Grübner, A. Dunkel, F. Steiner, T. Hofmann, *Anal. Chem.* **2021**, *93*, 12565, https://doi.org/10.1021/acs.analchem.1c01857.
[11] J. Pezzatti, V. González-Ruiz, S. Codesido, Y. Gagnebin, A. Joshi, D. Guillarme, J. Schappler, D. Picard, J. Boccard, S. Rudaz, *J. Chromatogr. A* **2019**, *1592*, 47, https://doi.org/10.1016/j.chroma.2019.01.023.
[12] S. Geller, H. Lieberman, A. Kloss, A. R. Ivanov, *J. Chromatogr. A* **2021**, *1642*, 462047, https://doi.org/10.1016/j.chroma.2021.462047.
[13] B. He, W. Zhang, F. Guled, A. Harms, R. Ramautar, T. Hankemeier, *Microchem. J.* **2021**, *171*, 106794, https://doi.org/10.1016/j.microc.2021.106794.
[14] M. Shanmuganathan, M. O. Sarfaraz, Z. Kroezen, H. Philbrick, R. Poon, A. Don-Wauchope, M. Puglia, D. Wishart, P. Britz-McKibbin, *Front. Mol. Biosci.* **2021**, *8*, 1, https://doi.org/10.3389/fmolb.2021.676349.
[15] O. Begou, H. G. Gika, I. D. Wilson, G. Theodoridis, *Analyst* **2017**, *142*, 3079, https://doi.org/10.1039/C7AN00812K.
[16] L. Coulier, R. Bas, S. Jespersen, E. Verheij, M. J. van der Werf, T. Hankemeier, *Anal. Chem.* **2006**, *78*, 6573, https://doi.org/z10.1021/ac0607616.
[17] P. M. Nowak, M. Woźniakiewicz, M. Gładysz, M. Janus, P. Kościelniak, *Anal. Bioanal. Chem.* **2017**, *409*, 4383, https://doi.org/10.1007/s00216-017-0382-y.
[18] S. Codesido, N. Drouin, S. Ferré, J. Schappler, S. Rudaz, V. González-Ruiz, *Electrophoresis* **2021**, elps. 202000333, https://doi.org/10.1002/elps.202000333.
[19] V. González-Ruiz, Y. Gagnebin, N. Drouin, S. Codesido, S. Rudaz, J. Schappler, *Electrophoresis* **2018**, *39*, 1222, https://doi.org/10.1002/elps.201700427.
[20] W. Zhang, T. Hankemeier, R. Ramautar, *Curr. Opin. Biotechnol.* **2017**, *43*, 1, https://doi.org/10.1016/j.copbio.2016.07.002.
[21] C. R. Mallet, Z. Lu, J. R. Mazzeo, *Rapid Commun. Mass Spectrom.* **2004**, *18*, 49, https://doi.org/10.1002/rcm.1276.
[22] T. Soga, Y. Ohashi, Y. Ueno, H. Naraoka, M. Tomita, T. Nishioka, *J. Proteome Res.* **2003**, *2*, 488, https://doi.org/10.1021/pr034020m.
[23] T. Soga, K. Igarashi, C. Ito, K. Mizobuchi, H.-P. Zimmermann, M. Tomita, *Anal. Chem.* **2009**, *81*, 6165, https://doi.org/10.1021/ac900675k.
[24] Y. Ishibashi, S. Harada, A. Takeuchi, M. Iida, A. Kurihara, S. Kato, K. Kuwabara, A. Hirata, T. Shibuki, T. Okamura, D. Sugiyama, A. Sato, K. Amano, A. Hirayama, M. Sugimoto, T. Soga, M. Tomita, T. Takebayashi, *Sci. Rep.* **2021**, *11*, 7407, https://doi.org/10.1038/s41598-021-86600-9.
[25] M. Mever, T. Hankemeier, R. Ramautar, *Electrophoresis* **2019**, elps. 201900115, https://doi.org/10.1002/elps.201900115.
[26] M. C. Gulersonmez, S. Lock, T. Hankemeier, R. Ramautar, *Electrophoresis* **2016**, *37*, 1007, https://doi.org/10.1002/elps.201500435.
[27] F. Boizard, V. Brunchault, P. Moulos, B. Breuil, J. Klein, N. Lounis, C. Caubet, S. Tellier, J.-L. Bascands, S. Decramer, J. P. Schanstra, B. Buffin-Meyer, *Sci. Rep.* **2016**, *6*, 34453, https://doi.org/10.1038/srep34453.
[28] H.-W. Liao, S. S. Rubakhin, M. C. Philip, J. V. Sweedler, *Anal. Chim. Acta* **2020**, *1118*, 36, https://doi.org/10.1016/j.aca.2020.04.028.
[29] N. L. Kuehnbaum, A. Kormendi, P. Britz-McKibbin, *Anal. Chem.* **2013**, *85*, 10664, https://doi.org/10.1021/ac403171u.
[30] N. Drouin, J. Pezzatti, Y. Gagnebin, V. González-Ruiz, J. Schappler, S. Rudaz, *Anal. Chim. Acta* **2018**, *1032*, 178, https://doi.org/10.1016/j.aca.2018.05.063.
[31] M. Mamani-Huanca, A. G. de la Fuente, A. Otero, A. Gradillas, J. Godzien, C. Barbas, Á. López-González, *J. Chromatogr. A* **2021**, *1635*, 461758, https://doi.org/10.1016/j.chroma.2020.461758.
[32] N. Drouin, M. van Mever, W. Zhang, E. Tobolkina, S. Ferre, A.-C. Servais, M.-J. Gou, L. Nyssen, M. Fillet, G. S. M. Lageveen-Kammeijer, J. Nouta, A. J. Chetwynd, I. Lynch, J. A. Thorn, J. Meixner, C. Lößner, M. Taverna, S. Liu, N. T. Tran, Y. Francois, A. Lechner, R. Nehmé, G. Al Hamoui Dit Banni, R. Nasreddine, C. Colas, H. H. Lindner, K. Faserl, C. Neusüß, M. Nelke, S. Lämmerer, C. Perrin, C. Bich-Muracciole, C. Barbas, Á. L. Gonzálvez, A. Guttman, M. Szigeti, P. Britz-McKibbin, Z. Kroezen, M. Shanmuganathan, P. Nemes, E. P. Portero, T. Hankemeier, S. Codesido, V. González-Ruiz, S. Rudaz, R. Ramautar, *Anal. Chem.* **2020**, *92*, 14103, https://doi.org/10.1021/acs.analchem.0c03129.
[33] R. Ramautar, J.-M. Busnel, A. M. Deelder, O. A. Mayboroda, *Anal. Chem.* **2012**, *84*, 885, https://doi.org/10.1021/ac202407v.
[34] B. Timischl, K. Dettmer, H. Kaspar, M. Thieme, P. J. Oefner, *Electrophoresis* **2008**, *29*, 2203, https://doi.org/10.1002/elps.200700517.
[35] P. Nemes, S. S. Rubakhin, J. T. Aerts, J. V Sweedler, *Nat. Protoc.* **2013**, *8*, 783, https://doi.org/10.1038/nprot.2013.035.
[36] J. T. Aerts, K. R. Louis, S. R. Crandall, G. Govindaiah, C. L. Cox, J. V. Sweedler, *Anal. Chem.* **2014**, *86*, 3203, https://doi.org/10.1021/ac500168d.
[37] M. G. M. Kok, G. W. Somsen, G. J. de Jong, *Talanta* **2015**, *132*, 1, https://doi.org/10.1016/j.talanta.2014.08.047.
[38] M. G. M. Kok, G. J. de Jong, G. W. Somsen, *Electrophoresis* **2011**, *32*, 3016, https://doi.org/10.1002/elps.201100271.
[39] E. Lesellier, C. West, *J. Chromatogr. A* **2015**, *1382*, 2, https://doi.org/10.1016/j.chroma.2014.12.083.

[40] V. Desfontaine, D. Guillarme, E. Francotte, L. Nováková, *J. Pharm. Biomed. Anal.* **2015**, *113*, 56, https://doi.org/10.1016/j.jpba.2015.03.007.

[41] E. Lemasson, S. Bertin, C. West, *J. Sep. Sci.* **2016**, *39*, 212, https://doi.org/10.1002/jssc.201501062.

[42] L. Nováková, A. Grand-Guillaume Perrenoud, I. Francois, C. West, E. Lesellier, D. Guillarme, *Anal. Chim. Acta* **2014**, *824*, 18, https://doi.org/10.1016/j.aca.2014.03.034.

[43] K. Taguchi, E. Fukusaki, T. Bamba, *J. Chromatogr. A* **2014**, *1362*, 270, https://doi.org/10.1016/j.chroma.2014.08.003.

[44] A. Tarafder, *J. Chromatogr. B* **2018**, *1091*, 1, https://doi.org/10.1016/j.jchromb.2018.05.003.

[45] A. Grand-Guillaume Perrenoud, J.-L. Veuthey, D. Guillarme, *J. Chromatogr. A* **2014**, *1339*, 174, https://doi.org/10.1016/j.chroma.2014.03.006.

[46] G. L. Losacco, J.-L. Veuthey, D. Guillarme, *TrAC Trends Anal. Chem.* **2019**, *118*, 731, https://doi.org/10.1016/j.trac.2019.07.005.

[47] D. Guillarme, V. Desfontaine, S. Heinisch, J.-L. Veuthey, *J. Chromatogr. B* **2018**, *1083*, 160, https://doi.org/10.1016/j.jchromb.2018.03.010.

[48] M. Lísa, M. Holčapek, *Anal. Chem.* **2015**, *87*, 7187, https://doi.org/10.1021/acs.analchem.5b01054.

[49] H. Takeda, Y. Izumi, M. Takahashi, T. Paxton, S. Tamura, T. Koike, Y. Yu, N. Kato, K. Nagase, M. Shiomi, T. Bamba, *J. Lipid Res.* **2018**, *59*, 1283, https://doi.org/10.1194/jlr.D083014.

[50] B. van de Velde, D. Guillarme, I. Kohler, *J. Chromatogr. B* **2020**, *1161*, 122444, https://doi.org/10.1016/j.jchromb.2020.122444.

[51] V. Desfontaine, G. L. Losacco, Y. Gagnebin, J. Pezzatti, W. P. Farrell, V. González-Ruiz, S. Rudaz, J.-L. Veuthey, D. Guillarme, *J. Chromatogr. A* **2018**, *1562*, 96, https://doi.org/10.1016/j.chroma.2018.05.055.

[52] G. L. Losacco, O. Ismail, J. Pezzatti, V. González-Ruiz, J. Boccard, S. Rudaz, J.-L. Veuthey, D. Guillarme, *J. Chromatogr. A* **2020**, *1620*, 461021, https://doi.org/10.1016/j.chroma.2020.461021.

[53] J. Ivanisevic, E. J. Want, *Metabolites* **2019**, *9*, 308, https://doi.org/10.3390/metabo9120308.

[54] J. Guo, T. Huan, *Anal. Chem.* **2020**, *92*, 8072, https://doi.org/10.1021/acs.analchem.9b05135.

[55] W. J. Nash, W. B. Dunn, *TrAC Trends Anal. Chem.* **2019**, *120*, 115324, https://doi.org/10.1016/j.trac.2018.11.022.

[56] X. Zhu, Y. Chen, R. Subramanian, *Anal. Chem.* **2014**, *86*, 1202, https://doi.org/10.1021/ac403385y.

[57] R. S. Plumb, K. A. Johnson, P. Rainville, B. W. Smith, I. D. Wilson, J. M. Castro-Perez, J. K. Nicholson, *Rapid Commun. Mass Spectrom.* **2006**, *20*, 1989, https://doi.org/10.1002/rcm.2550.

[58] I. Tada, R. Chaleckis, H. Tsugawa, I. Meister, P. Zhang, N. Lazarinis, B. Dahlén, C. E. Wheelock, M. Arita, *Anal. Chem.* **2020**, *92*, 11310, https://doi.org/10.1021/acs.analchem.0c01980.

[59] R. Bonner, G. Hopfgartner, *TrAC Trends Anal. Chem.* **2019**, *120*, 115278, https://doi.org/10.1016/j.trac.2018.10.014.

[60] L. C. Gillet, P. Navarro, S. Tate, H. Röst, N. Selevsek, L. Reiter, R. Bonner, R. Aebersold, *Mol. Cell. Proteomics* **2012**, *11*, O111.016717, https://doi.org/10.1074/mcp.O111.016717.

[61] G. Paglia, A. J. Smith, G. Astarita, *Mass Spectrom. Rev.* **2021**, mas. 21686, https://doi.org/10.1002/mas.21686.

[62] J. Pezzatti, V. González-Ruiz, J. Boccard, D. Guillarme, S. Rudaz, *Metabolites* **2020**, *10*, 464, https://doi.org/10.3390/metabo10110464.

[63] N. G. Mahieu, G. J. Patti, *Anal. Chem.* **2017**, *89*, 10397, https://doi.org/10.1021/acs.analchem.7b02380.

[64] 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, *Metabolomics* **2007**, *3*, 211, https://doi.org/10.1007/s11306-007-0082-2.

[65] I. Blaženović, T. Kind, J. Ji, O. Fiehn, *Metabolites* **2018**, *8*, 31, https://doi.org/10.3390/metabo8020031.

[66] E. L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H. P. Singer, J. Hollender, *Environ. Sci. Technol.* **2014**, *48*, 2097, https://doi.org/10.1021/es5002105.

[67] F. Jourdan, *MetaboloNews* **2020**, *10*(2), 6.

[68] R. Chaleckis, I. Meister, P. Zhang, C. E. Wheelock, *Curr. Opin. Biotechnol.* **2019**, *55*, 44, https://doi.org/10.1016/j.copbio.2018.07.010.

[69] I. Meister, P. Zhang, A. Sinha, C. M. Sköld, Å. M. Wheelock, T. Izumi, R. Chaleckis, C. E. Wheelock, *Anal. Chem.* **2021**, *93*, 5248, https://doi.org/10.1021/acs.analchem.1c00203.

[70] A. Damont, M.-F. Olivier, A. Warnet, B. Lyan, E. Pujos-Guillot, E. L. Jamin, L. Debrauwer, S. Bernillon, C. Junot, J.-C. Tabet, F. Fenaille, *J. Mass Spectrom.* **2019**, *54*, 567, https://doi.org/10.1002/jms.4372.

[71] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, T. Nishioka, *J. Mass Spectrom.* **2010**, *45*, 703, https://doi.org/10.1002/jms.1777.

[72] M. Wang, J. J. Carver, V. V Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V Melnik, M. J. Meehan, W.-T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. Boya P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Ø. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, T. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein, N. Bandeira, *Nat. Biotechnol.* **2016**, *34*, 828, https://doi.org/10.1038/nbt.3597.

[73] D. S. Wishart, Y. D. Feunang, A. Marcu, A. C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach, A. Scalbert, *Nucleic Acids Res.* **2018**, *46*, D608, https://doi.org/10.1093/nar/gkx1089.

[74] M. Vinaixa, E. L. Schymanski, S. Neumann, M. Navarro, R. M. Salek, O. Yanes, *TrAC Trends Anal. Chem.* **2016**, *78*, 23, https://doi.org/10.1016/j.trac.2015.09.005.

[75] C. A. Smith, G. O. Maille, E. J. Want, C. Qin, S. A. Trauger, T. R. Brandon, D. E. Custodio, R. Abagyan, G. Siuzdak, *Ther. Drug Monit.* **2005**, *27*, 747, https://doi.org/10.1097/01.ftd.0000179845.53213.39.

[76] H. Tsugawa, T. Kind, R. Nakabayashi, D. Yukihira, W. Tanaka, T. Cajka, K. Saito, O. Fiehn, M. Arita, *Anal. Chem.* **2016**, *88*, 7946, https://doi.org/10.1021/acs.analchem.6b00770.

[77] C. Ruttkies, E. L. Schymanski, S. Wolf, J. Hollender, S. Neumann, *J. Cheminform.* **2016**, *8*, 3, https://doi.org/10.1186/s13321-016-0115-9.

[78] F. Allen, A. Pon, M. Wilson, R. Greiner, D. Wishart, *Nucleic Acids Res.* **2014**, *42*, W94, https://doi.org/10.1093/nar/gku436.

[79] K. Dührkop, H. Shen, M. Meusel, J. Rousu, S. Böcker, *Proc. Natl. Acad. Sci.* **2015**, *112*, 12580, https://doi.org/10.1073/pnas.1509788112.

[80] K. Dührkop, M. Fleischauer, M. Ludwig, A. A. Aksenov, A. V. Melnik, M. Meusel, P. C. Dorrestein, J. Rousu, S. Böcker, *Nat. Methods* **2019**, *16*, 299, https://doi.org/10.1038/s41592-019-0344-8.

[81] H. Tsugawa, K. Ikeda, M. Takahashi, A. Satoh, Y. Mori, H. Uchino, N. Okahashi, Y. Yamada, I. Tada, P. Bonini, Y. Higashi, Y. Okazaki, Z. Zhou, Z.-J. Zhu, J. Koelmel, T. Cajka, O. Fiehn, K. Saito, M. Arita, M. Arita, *Nat. Biotechnol.* **2020**, *38*, 1159, https://doi.org/10.1038/s41587-020-0531-2.

[82] M. Sud, E. Fahy, D. Cotter, A. Brown, E. A. Dennis, C. K. Glass, A. H. Merrill, R. C. Murphy, C. R. H. Raetz, D. W. Russell, S. Subramaniam, *Nucleic Acids Res.* **2007**, *35*, D527, https://doi.org/10.1093/nar/gkl838.

[83] Z. Ni, L. Goracci, G. Cruciani, M. Fedorova, *Free Radic. Biol. Med.* **2019**, *144*, 110, https://doi.org/10.1016/j.freeradbiomed.2019.04.027.

[84] H. C. Köfeler, T. O. Eichmann, R. Ahrends, J. A. Bowden, N. Danne-Rasche, E. A. Dennis, M. Fedorova, W. J. Griffiths, X. Han, J. Hartler, M. Holčapek, R. Jirásko, J. P. Koelmel, C. S. Ejsing, G. Liebisch, Z. Ni, V. B. O'Donnell, O. Quehenberger, D. Schwudke, A. Shevchenko, M. J. O. Wakelam, M. R. Wenk, D. Wolrab, K. Ekroos, *Nat. Commun.* **2021**, *12*, 4771, https://doi.org/10.1038/s41467-021-24984-y.

[85] J. P. Koelmel, C. Z. Ulmer, C. M. Jones, R. A. Yost, J. A. Bowden, *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **2017**, *1862*, 766, https://doi.org/10.1016/j.bbalip.2017.02.016.

[86] M. Holčapek, P. Jandera, P. Zderadička, L. Hrubá, *J. Chromatogr. A* **2003**, *1010*, 195, https://doi.org/10.1016/S0021-9673(03)01030-6.

[87] M. Ovčačíková, M. Lísa, E. Cífková, M. Holčapek, *J. Chromatogr. A* **2016**, *1450*, 76, https://doi.org/10.1016/j.chroma.2016.04.082.

[88] G. Liebisch, E. Fahy, J. Aoki, E. A. Dennis, T. Durand, C. S. Ejsing, M. Fedorova, I. Feussner, W. J. Griffiths, H. Köfeler, A. H. Merrill, R. C. Murphy, V. B. O'Donnell, O. Oskolkova, S. Subramaniam, M. J. O. Wakelam, F. Spener, *J. Lipid Res.* **2020**, *61*, 1539, https://doi.org/10.1194/jlr.S120001025.

[89] T. Züllig, H. C. Köfeler, *Mass Spectrom. Rev.* **2021**, *40*, 162, https://doi.org/10.1002/mas.21627.

[90] T. Kind, K.-H. Liu, D. Y. Lee, B. DeFelice, J. K. Meissen, O. Fiehn, *Nat. Methods* **2013**, *10*, 755, https://doi.org/10.1038/nmeth.2551.

[91] L. Goracci, S. Tortorella, P. Tiberi, R. M. Pellegrino, A. Di Veroli, A. Valeri, G. Cruciani, *Anal. Chem.* **2017**, *89*, 6257, https://doi.org/10.1021/acs.analchem.7b01259.

[92] T. Pluskal, S. Castillo, A. Villar-Briones, M. Orešič, *BMC Bioinformatics* **2010**, *11*, 395, https://doi.org/10.1186/1471-2105-11-395.

[93] J. Hartler, A. Triebl, A. Ziegl, M. Trötzmüller, G. N. Rechberger, O. A. Zeleznik, K. A. Zierler, F. Torta, A. Cazenave-Gassiot, M. R. Wenk, A. Fauland, C. E. Wheelock, A. M. Armando, O. Quehenberger, Q. Zhang, M. J. O. Wakelam, G. Haemmerle, F. Spener, H. C. Köfeler, G. G. Thallinger, *Nat. Methods* **2017**, *14*, 1171, https://doi.org/10.1038/nmeth.4470.

[94] C. Muschet, G. Möller, C. Prehn, M. H. de Angelis, J. Adamski, J. Tokarz, *Metabolomics* **2016**, *12*, 151, https://doi.org/10.1007/s11306-016-1104-8.

[95] M. M. W. B. Hendriks, F. A. va. Eeuwijk, R. H. Jellema, J. A. Westerhuis, T. H. Reijmers, H. C. J. Hoefsloot, A. K. Smilde, *TrAC Trends Anal. Chem.* **2011**, *30*, 1685, https://doi.org/10.1016/j.trac.2011.04.019.

[96] S. Naz, M. Vallejo, A. García, C. Barbas, *J. Chromatogr. A* **2014**, *1353*, 99, https://doi.org/10.1016/j.chroma.2014.04.071.

[97] M. A. Kamleh, T. M. D. Ebbels, K. Spagou, P. Masson, E. J. Want, *Anal. Chem.* **2012**, *84*, 2670, https://doi.org/10.1021/ac202733q.

[98] J. A. Kirwan, D. I. Broadhurst, R. L. Davidson, M. R. Viant, *Anal. Bioanal. Chem.* **2013**, *405*, 5147, https://doi.org/10.1007/s00216-013-6856-7.

[99] P. Filzmoser, B. Walczak, *J. Chromatogr. A* **2014**, *1362*, 194, https://doi.org/10.1016/j.chroma.2014.08.050.

[100] J. Boccard, J.-L. Veuthey, S. Rudaz, *J. Sep. Sci.* **2010**, *33*, 290, https://doi.org/10.1002/jssc.200900609.

[101] M. Vinaixa, S. Samino, I. Saez, J. Duran, J. J. Guinovart, O. Yanes, *Metabolites* **2012**, *2*, 775, https://doi.org/10.3390/metabo2040775.

[102] J. P. Shaffer, *Annu. Rev. Psychol.* **1995**, *46*, 561, https://doi.org/10.1146/annurev.ps.46.020195.003021.

[103] Y. Benjamini, Y. Hochberg, *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

[104] E. Saccenti, H. C. J. Hoefsloot, A. K. Smilde, J. A. Westerhuis, M. M. W. B. Hendriks, *Metabolomics* **2014**, *10*, 361, https://doi.org/10.1007/s11306-013-0598-6.

[105] J. Boccard, S. Rudaz, *J. Chemom.* **2014**, 28, 1 https://doi.org/10.1002/cem.2567.

[106] S. Wold, M. Sjöström, L. Eriksson, *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109, https://doi.org/10.1016/S0169-7439(01)00155-1.

[107] D. Kell, *Curr. Opin. Microbiol.* **2004**, *7*, 296, https://doi.org/10.1016/j.mib.2004.04.012.

[108] M. Kanehisa, *Nucleic Acids Res.* **2000**, *28*, 27, https://doi.org/10.1093/nar/28.1.27.

[109] R. Caspi, R. Billington, L. Ferrer, H. Foerster, C. A. Fulcher, I. M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, P. D. Karp, *Nucleic Acids Res.* **2016**, *44*, D471, https://doi.org/10.1093/nar/gkv1164.

[110] S. C. Booth, A. M. Weljie, R. J. Turner, *Comput. Struct. Biotechnol. J.* **2013**, *4*, e201301003, https://doi.org/10.5936/csbj.201301003.

[111] V. Lacroix, L. Cottret, P. Thebault, M.-F. Sagot, *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **2008**, *5*, 594, https://doi.org/10.1109/TCBB.2008.79.

[112] L. Cottret, D. Wildridge, F. Vinson, M. P. Barrett, H. Charles, M.-F. Sagot, F. Jourdan, *Nucleic Acids Res.* **2010**, *38*, W132, https://doi.org/10.1093/nar/gkq312.

[113] L. Cottret, C. Frainay, M. Chazalviel, F. Cabanettes, Y. Gloaguen, E. Camenen, B. Merlet, S. Heux, J.-C. Portais, N. Poupin, F. Vinson, F. Jourdan, *Nucleic Acids Res.* **2018**, *46*, W495, https://doi.org/10.1093/nar/gky301.

[114] D. Grapov, K. Wanichthanarak, O. Fiehn, *Bioinformatics* **2015**, *31*, 2757, https://doi.org/10.1093/bioinformatics/btv194.

[115] P. Moreno, S. Beisken, B. Harsha, V. Muthukrishnan, I. Tudose, A. Dekker, S. Dornfeldt, F. Taruttis, I. Grosse, J. Hastings, S. Neumann, C. Steinbeck, *BMC Bioinformatics* **2015**, *16*, 56, https://doi.org/10.1186/s12859-015-0486-3.

[116] M. Chagoyen, F. Pazos, *Bioinformatics* **2011**, *27*, 730, https://doi.org/10.1093/bioinformatics/btr001.

[117] M. Kankainen, P. Gopalacharyulu, L. Holm, M. Orešič, *Bioinformatics* **2011**, *27*, 1878, https://doi.org/10.1093/bioinformatics/btr278.

[118] S. Li, Y. Park, S. Duraisingham, F. H. Strobel, N. Khan, Q. A. Soltow, D. P. Jones, B. Pulendran, *PLoS Comput. Biol.* **2013**, *9*, e1003123, https://doi.org/10.1371/journal.pcbi.1003123.

[119] Z. Pang, J. Chong, G. Zhou, D. A. de Lima Morais, L. Chang, M. Barrette, C. Gauthier, P.-É. Jacques, S. Li, J. Xia, *Nucleic Acids Res.* **2021**, *49*, W388, https://doi.org/10.1093/nar/gkab382.