Current state-of-the-art of separation methods used in LC-MS based metabolomics and lipidomics

Eva-Maria Harrieder[1], Fleming Kretschmer[2], Sebastian Böcker[2], Michael Witting[1,3,4,*]

[1]Research Unit Analytical BioGeoChemistry, Helmholtz Zentrum München, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany

[2]Chair of Bioinformatics, Friedrich-Schiller-Universität Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany

[3]Metabolomics and Proteomics Core, Helmholtz Zentrum München, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany

[4]Chair of Analytical Food Chemistry, Technical University of Munich, Maximus-von-Imhof-Forum 2, 85354 Freising, Germany

*Corresponding author: Dr. Michael Witting, michael.witting@helmholtz-muenchen.de

This is a reprint of Harrieder *et al.*, Current state-of-the-art of separation methods used in LC-MS based metabolomics and lipidomics, *Journal of Chromatography B* 1188**:**123069, 2022**,** https://doi.org/10.1016/j.jchromb.2021.123069.

Abstract

Metabolomics deals with the large-scale analysis of metabolites, belonging to numerous compound classes and showing an extremely high chemical diversity and complexity. Lipidomics, being a subcategory of metabolomics, analyzes the cellular lipid species. Both require state-of-the-art analytical methods capable of accessing the underlying chemical complexity. One of the major techniques used for the analysis of metabolites and lipids is Liquid Chromatography-Mass Spectrometry (LC-MS), offering both different selectivities in LC separation and high sensitivity in MS detection. Chromatography can be divided into different modes, based on the properties of the employed separation system. The most popular ones are Reversed-Phase (RP) separation for non- to mid-polar molecules and Hydrophilic Interaction Liquid Chromatography (HILIC) for polar molecules. So far, no single analysis method exists that can cover the entire range of metabolites or lipids, due to the huge chemical diversity. Consequently, different separation methods have been used for different applications and research questions.

In this review, we explore the current use of LC-MS in metabolomics and lipidomics. As a proxy, we examined the use of chromatographic methods in the public repositories EBI MetaboLights and NIH Metabolomics Workbench. We extracted 1484 method descriptions, collected separation metadata and generated an overview on the current use of columns, eluents, etc. Based on this overview, we reviewed current practices and identified potential future trends as well as required improvements that may allow us to increase metabolite coverage, throughput or both simultaneously.

Introduction

Metabolomics is defined as the systematic study of all metabolites in a given biological system or supersystem at a given time. From all of the “omes” (genome, transcriptome, proteome and metabolome), the metabolome is closest to the observed phenotype [1]. In contrast to the genome, transcriptome or proteome, the full extent of the metabolome is still not known and remains elusive for

almost all organisms. Different from DNA, RNA and proteins, metabolites show a high chemical diversity and complexity and therefore, require the combination of different analytical approaches to achieve full coverage. Even lipids, sometimes believed to represent a homogenous group of molecules, show high structural diversity based on different combinations of acyl groups, head groups and further modifications. Unlike DNA and RNA, metabolites (and proteins) cannot be amplified, further aggravating their analysis.

Liquid Chromatography-Mass Spectrometry (LC-MS) is one of the major analytical techniques used in metabolomics. Other major techniques are Gas Chromatography-Mass Spectrometry, Capillary Electrophoresis-Mass Spectrometry and Nuclear Magnetic Resonance Spectroscopy (NMR). Chromatographic separation greatly enhances the analytical capabilities of MS: First, separation of molecules reduces ion suppression, leading to improved detection, especially of low abundant or only weakly ionizing molecules. Second, the obtained retention time (RT) represents an orthogonal information to MS and tandem MS (MS/MS) and represents potential information about the polarity of a compound (in Reversed-Phase (RP) early eluting substances are more polar, while in Hydrophilic Interaction Chromatography (HILIC) the earlier substances are more non-polar), which can be used in metabolite identification. Selectivity in LC is mostly based on the choice of the chromatographic column, followed by employed eluents and other parameters. In order to cover a wide range of metabolites, different separation modes have to be employed. One of the most-used techniques is RP separation for the analysis of mid- to non-polar metabolites. However, this method is not able to analyze polar metabolites, such as amino acids, organic acids, sugars, etc. HILIC represents an interesting alternative for the separation of these metabolites, but its use is not as widespread as RP.

For both types of chromatographic separation, several columns with different properties and selectivities are available. The chromatographic methods for metabolite and lipid analysis are far from being harmonized and standardized, with each metabolomics or lipidomics laboratory using a unique set of methods for their analyses. This review focuses on LC-MS based metabolomics and lipidomics and investigates current practices as well as new developments. The evaluation of current practices is based on a collection of chromatographic methods described in public metabolomics data repositories. We used two public metabolomics repositories, EBI MetaboLights and NIH Metabolomics Workbench, as a proxy for metabolomics experiments carried out in the community [2, 3]. We analyzed the available chromatographic metadata found in method descriptions to obtain an overview on employed columns, eluents and separation conditions. This data forms the basis for elected discussions on different aspects and needs of chromatography in metabolomics and lipidomics. This is followed by an overview of several recent improvements, application of novel chromatographic systems and integration of data from different separation methods to achieve comprehensive coverage of the metabolome and lipidome.

The task: dealing with chemical complexity of the metabolome

Chemical complexity increases throughout the hierarchy of the "omes". DNA and RNA represent linear polymers of four building blocks called nucleobases (Guanine (G), Cytosine (C), Adenine (A) and Thymine (T) in DNA and C, G, A and Uracil (U) in RNA). The different nucleobases can be further modified using different epigenetic marks (e.g. methylation). However, the principle underlying all analysis techniques in genomics and transcriptomics is the extreme structural confinement of these molecules. Furthermore, the Polymerase Chain Reaction allows us to duplicate these molecules at an exponential rate. Similar to DNA and RNA, proteins represent linear polymers but have substantially more building blocks: There exist 22 proteinogenic amino acids, which can also be chemically modified (post-translational modifications). Proteomics approaches digest proteins into smaller peptides for simplified chemical analysis. Although peptides show a huge diversity in sequence they are, chemically speaking, still a single compound class and can be analyzed under the same conditions. In contrast, the metabolome is composed of chemically very different structures belonging to multiple compound classes, including amino acids, carbohydrates, sterols and many more. This comparison is not meant to say that (epi)genomics, (epi)transcriptomics and proteomics are easy (they are definitely not), but the "starting conditions" in form of structural diversity for metabolomics and lipidomics are very different.

Metabolites have no common and unifying structural characteristics. Atoms in metabolites can be connected linearly, in rings, multiple rings and combinations thereof and by different types of bonds (single, double or triple bond). On top of the structural diversity, metabolites span several orders of magnitude in concentration from fM to mM [4]. Even lipids, which are generally believed to represent a rather homogenous group of small molecules with similar properties, show a high combinatorial and chemical diversity based on different building blocks such as different fatty acyls, headgroups, sphingoid bases and diverse chemical modifications [5]. From the metabolomics point of view, lipids may appear as a uniform group with the unifying property of being soluble in organic solvents such as chloroform. But a closer look shows that there is a high structural diversity within different lipid classes.

The requirement for separation techniques in metabolomics and lipidomics is two-fold. First, a large range of polarities and structures needs to be covered to enable detailed analysis of the full complement of metabolites and lipids. This includes the analysis of small organic acids such as pyruvate or lactate up to larger metabolites such as palmitoyl-CoA, phospholipids and beyond. While MS and MS/MS can only partially resolve isomeric structures, which show differential fragmentation, chromatography can resolve several cases of isomerism. Different molecular structures lead to differences in physicochemical properties and thus to different RTs, with sometimes very small differences for certain structure pairs. Second, small structural and molecular mass differences should ideally be resolved, to allow correct annotation, identification and interpretation of the obtained data. Both aspects are met by high and ultrahigh resolution MS and MS/MS, to a certain degree: For example, the overlap of isobars or isomers shows a different fragmentation (e.g. the overlap between monoisotopic and the M+2 isotope of lipids differs in a single double bond). Yet, certain pairs of structures require separation in the chromatographic dimension, such as isomers of dafachronic acid found in the model organism *Caenorhabditis elegans* (see below). Chromatography plays an important role in deconvolution and grouping of MS data. Using direct-infusion-MS we cannot determine if two signals differing by 18 Da are in fact two different metabolites or

the in-source loss of water. In contrast, using LC-MS two metabolites (usually) have different RTs, whereas the water loss shows perfect coelution with the intact molecule. Good chromatographic separation with well-defined peak shapes is important for the use and full exploitation of data-independent acquisition, where all ions within a certain m/z range are fragmented and detected [6].

Ultimately, RT represents valuable orthogonal information for metabolite identification, but this is often neglected in early stages of the metabolite annotation and identification workflow. RT is not as transferable across different labs as MS and MS/MS, which are physical properties or inherent to the structure. RT in contrast is a system property arising from the combination of the analyte and the employed chromatographic setup. This makes it complicated to compare RTs across different laboratories, but not impossible.

To get an overview on the range of polarity of metabolites that would be required to be separated in order to analyze the full metabolome or lipidome of a specific organism, metabolite structures from different databases have been downloaded. This includes the Human Metabolome Database (HMDB) [7], the Yeast Metabolome Database (YMDB) [8], the *Escherichia coli* Metabolome Database (ECMDB) [9], the genome-scale metabolic model WormJam [10] and the collection of biochemical pathways in *Arabidopsis* AraCyc [11]. Based on the obtained structures, the chemical formula, exact mass and logP have been calculated using ChemAxon JChem. The distribution of logP values within the different databases is shown in Figure 1. Compounds with a logP < 0 are generally regarded as polar, while compounds with a logP > 0 are non-polar. However, different other factors, such as charges, functional groups etc. influence the actual chromatographic behavior and the logP is only a rough estimator. The different databases are of different size, which in part reflects the complexity of the organism, but it mostly reflects the current state of curation. For example, HMDB contains a large number of predicted lipid structures.

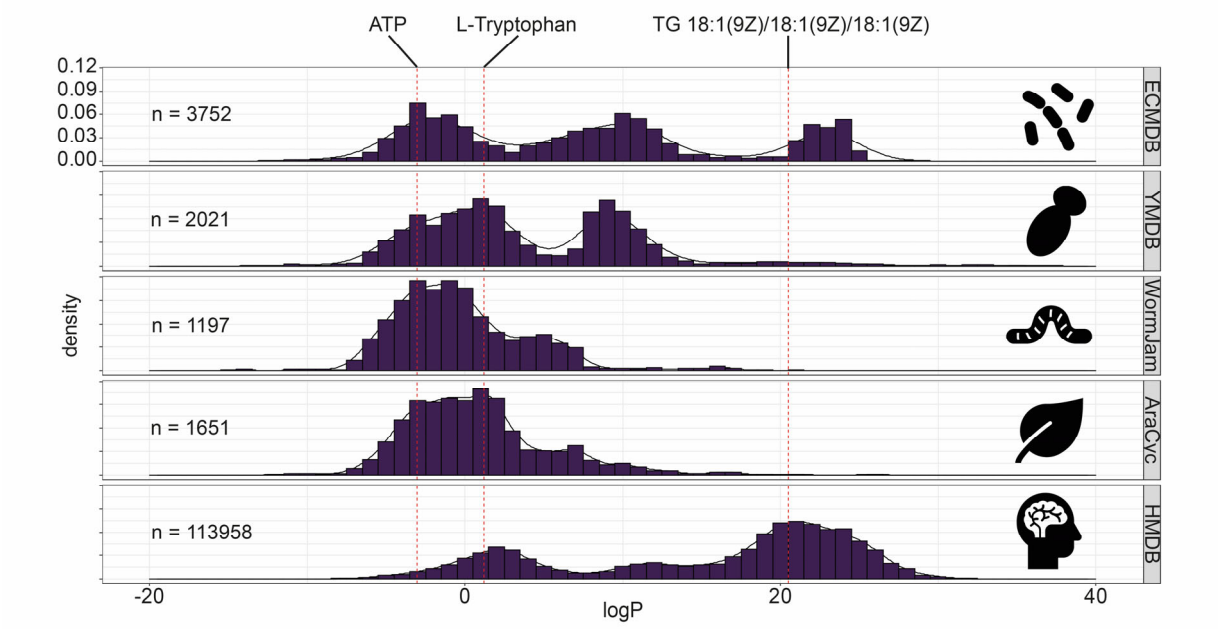

Figure 1 illustrates the polarity spread for different organisms (E. coli, yeast, C. elegans, Arabidopsis thaliana and human), measured as the logarithm of the water/octanol partition coefficient logP.

Metabolites from the respective databases span a large range of logP values. Since the polarity range is so wide, no single chromatographic method can currently cover this entire range. Instead, a combination of several separation methods is necessary. This plot should not be taken as a comparison of the different organisms since none of the utilized databases is complete and covers the entire metabolome. Rather, it illustrates that the distribution is independent of the organism. At present, only the E. coli and human databases contain also hydrophobic lipids. Vertical lines indicate the logP of ATP, tryptophan and TG(18:1(9Z)/18:1(9Z)/18:1(9Z)) as reference. Numbers indicate the size of the respective database. The complete data is found in the SI Tables 1-5.

Separation modes used in metabolomics

Chromatographic separation of molecules was originally described by Michael Tswett at the beginning of the 20th century [12]. Since then, different methods have been developed for different polarity ranges. Normal-Phase (NP) chromatography uses polar stationary phases such as polar modified or unmodified silica gel and apolar solvents such as hexane or ethyl acetate as mobile phases. Separation of compounds is based on adsorption of molecules on the polar surface of the stationary phase. Reversing the polarity of the stationary and mobile phase yields RP chromatography. Here, apolar modified stationary phases with ligands such as octadecylsilane are used in combination with polar mobile phases like water, methanol (MeOH), acetonitrile (ACN), 2-propanol (iPrOH) or mixtures of them. Metabolite separation is based on partition of analytes between the stationary and the mobile phase. Depending on the property of the analyte of interest and the separation that should be achieved, different modified RP phases are available offering additional selectivity, but generally, RP is used for mid- to non-polar molecules.

Polar metabolites can be analyzed on RP columns using ion pair chromatography [13]. In order to retain polar molecules, a hydrophobic salt with the opposite charge of the analytes of interest is added to the mobile phase. The salt and the analyte then form an ion pair that is retained on the RP column. Typical examples for the analysis of anions are tributylamine or higher alkyl amines. It has to be noted that if MS detection is used, the LC and often also the MS system are only used in a single ion mode, since the ion pair reagent can lead to large ion suppression in one of the ion modes (e.g. in positive mode for tributylamine). Therefore, often a dedicated LC-MS instrument is used for this type of chromatography.

Whereas RP is well suited for analysis of mid- to non-polar metabolites and lipids, HILIC can be used for analysis of polar metabolites without the need of ion pairing reagents. However, HILIC separation shows lower efficiency and resolution power compared to RP, with broad and often asymmetric chromatographic peaks. HILIC represents a special version of NP chromatography, using a polar stationary phase (typical for NP) and a polar mobile phase (typical for RP). Polar stationary phases are used to bind a layer of water on the stationary support, whereby the separation is based on partitioning between this polar water layer and the more hydrophobic mobile phase, such as ACN. However, since different ligands are used to immobilize the water, several secondary interactions such as dipole-dipole, electrostatic interactions or hydrogen bonding influence the separation. Therefore, the actual separation mechanism is not always fully understood, making it hard to predict HILIC elution order or to optimize methods.

Lastly, Ion (Exchange) Chromatography (IC) is a suitable alternative for analysis of ionic metabolites, such as intermediates from glycolysis and energy metabolism, but would miss zwitterionic or polar, but uncharged molecules. Recent progress in miniaturization and suppressor technology made it possible to

perform online coupling of IC to MS. A recent study analyzed anionic metabolites of the central carbon cycle with IC-MS and found that this method shows higher stability compared to HILIC [14].

Overview on currently employed columns and eluent systems

The ultimate goal of chromatography in LC-MS based metabolomics should be the baseline separation of all molecules with a sufficient chromatographic resolution *R*, in order to detect and quantify them independently from each other. If this will be ever achieved is questionable, but it can serve as guiding principle for method development and optimization. Several parameters like stationary phase, eluent composition, flow rate, temperature, etc. influence the actual separation of metabolites and lipids. A first step in the method development is the selection of a suitable column. Subsequently, the separation of pairs of molecules is optimized using the chromatographic resolution *R* as metric. According to equation 1, *R* depends on three parameters: the number of theoretical plates *N* (measure of column efficiency), the selectivity α (measure of separability of two analytes, equation 2) and the retention factor *k* (migration velocity of the analytes, equation 3).

$$R = \frac{\sqrt{N}}{4} \cdot (\alpha - 1) \cdot \frac{k}{1+k} \quad \text{(Equation 1)}$$

(with *R* = resolution, *N* = number of theoretical plates, α = selectivity and *k* = retention factor)

$$\alpha = \frac{\mathrm{k_2}}{\mathrm{k_1}} \quad \text{(Equation 2)}$$

(with α = selectivity, k_1 = retention factor of substance 1 (earlier eluted) and k_2 = retention factor of substance 2 (later eluted))

$$\mathrm{k} = \frac{\mathrm{tG \cdot F}}{\Delta\%\mathrm{B \cdot V_m \cdot S}} \quad \text{(Equation 3)}$$

(with *k* = retention factor, t_G = gradient time, *F* = flow rate, Δ%*B* = gradient range, V_m = column volume and *S* = constant for a given analyte)

$$\mathrm{n_c} = \frac{\mathrm{tRl - tRe}}{\mathrm{w}} \quad \text{(Equation 4)}$$

(with n_c = peak capacity, *t*Rl = retention time of the later eluted substance, *t*Re = retention time of the earlier eluted substance and ω = basic width)

From equation 1 it can be derived that the selectivity has the highest influence on R and thus presents the best starting point for method optimization. While in non-targeted experiments not all expected substances are known, the selectivity can only be used as a minor parameter for development and optimization, but it is more important to achieve high efficiency separation with a high peak capacity n_c (number of maximum peaks that are separated with a certain *R* within a given separation space, equation 4). In contrast to this, in targeted analysis, selectivity plays a major role, especially in the separation of isomeric structures.

In order to generate an overview on the separation methods used in LC-MS based metabolomics and lipidomics, data was curated from the two metabolomics data repositories EBI MetaboLights and NIH Metabolomics Workbench. Although the number of studies submitted to both repositories is comparably small to all publications in the field of metabolomics and lipidomics, it represents a suitable starting point. It might also happen that some working groups submit multiple studies using the same chromatographic methods. Based on the obtained data and comparison of methods we identified several duplications of methods. However, even in publications laboratory methods are re-used, so this might be not accounted as potential bias of our analysis. Still, the data and studies deposited in both repositories provide a good first overview and proxy for metabolomics and lipidomics studies. Data on chromatographic columns, eluents, etc. were extracted from the publicly available experimental descriptions and metadata submitted together with the studies. All data is summarized in SI Tables 6 and 7.

Currently, no standardized way of describing chromatographic metadata exists, so different column names or notation of gradients are used. Since the names of the columns can be written in different ways, we normalized column names using the exact naming supplied by the column vendors. Based on the United States Pharmacopeial (USP) convention code (grouping of columns based on the phase material, USP list can be downloaded on their webpage: https://www.usp.org/resources/chromatographic-columns) the columns were grouped into the categories RP, HILIC and other (e.g. pentafluorophenyl (PFP) columns are found in the latter category). An overview on the occurrence of the individual columns in the two repositories can be found in SI Figure 1 and 2.

In both repositories, the use of RP columns exceeds the two other categories (Figure 2) with 76.0% and 72.0% of all studies in MetaboLights and Metabolomics Workbench, respectively. This was followed by HILIC with 21.4% and 25.0%, respectively. Looking into the development of this proportion over time the use of HILIC increased in MetaboLights, while some fluctuation is observed in Metabolomics Workbench reaching even more than 50% descriptions using HILIC columns (Figure 2).

Within the category of RP columns, C18 columns were dominating with 86.0%, but also others such as C8 or C30 were employed. Although all C18 columns share the same chemical modification, each column offers (slightly) different selectivity due to differences in the used base material, carbon load, etc. Beside the typical C18 materials, phenyl, polyhydroxymethacrylate, graphic carbon or other alkyl chains are used for RP columns. One particular example is the *Restek Pinnacle DB Biphenyl* column, where the selectivity is based on the aromatic rings rather than classical alkyl chains. Longer alkyl chains such as C30 offer enhanced shape selectivity and increased retention for highly non-polar substances. Criscuolo *et al.* [15] performed a systematic comparison between different RP columns for the analysis of lipids. Human blood plasma was used to provide suggestions for lipid analysis based on different parameters. C30 columns showed high numbers of identified lipids with high chromatographic resolution.

Figure 2: Overview on columns reported in Metabolomics Workbench and MetaboLights in total or in the years 2012 to 2020. RP is the major separation mode found in the data sets, followed by HILIC and other separation modes (e.g. aqueous normal phase). An increased use of HILIC over time can be observed for MetaboLights.

However, since a large proportion of the metabolome from different organisms is polar with a logP < 0, there is an eminent need for the analysis of polar metabolites. In HILIC, different ligands for immobilization of a water layer on the surface of the stationary phase are used. Thicker layers are achieved by larger ligands that can bind water, e.g. by using zwitterionic ligands such as sulfobetaine groups. However, these ligands also show secondary interactions with metabolites, e.g. electrostatic interactions, often leading to broader peaks. Especially central energy and carbon metabolism possess many different highly polar metabolites, often of anionic and phosphorylated species and their analysis is difficult due to very strong electrostatic interactions. The generation of the water layer also requires longer equilibration times compared to RP. The number of different HILIC columns used is lower than the number of different RP columns, which might be because HILIC is not as widespread as RP in metabolomics. Different stationary phase chemistries such as silica, aminopropylsilane, alkyl amide and sulfobetaine groups are available for HILIC, each showing specifics for the analysis of certain classes of metabolites. A comparison of different HILIC columns has been carried out in an automated fashion using Derringer desirability functions and chromatographic parameters such as peak width, resolution and tailing factor as input [16]. One of the most commonly used methods applies an amide or amino column with a basic eluent (pH ~9). This method was first described by Bajad *et al.* [17] for the analysis of water-soluble metabolites from cellular extracts.

It was further described by Yuan *et al.* [18] in a protocol paper, which is often adopted for the analysis of polar metabolites. However, columns typically show a diminished lifetime at elevated pH values.

Selectivity is important for targeted analysis and different columns can be tested for optimization of separation. The base material of C18 columns of different manufacturers and brands differs slightly, leading to different selectivities. Furthermore, carbon load, end capping etc. have an influence on the selectivity. By modifying the "traditional" C18 RP column with additional functional groups, the separation can be further fine-tuned. A particular example found in MetaboLights and Metabolomics Workbench are the C18-PFP columns, which add the extra selectivity of PFP to the C18 based ligand. PFP shows weak anion exchange capabilities and aromatic selectivity.

Beside selectivity, column efficiency is another important factor. The use of sub-2-µm particles enables highly efficient separations, allowing either higher peak capacities or speed-up of methods without sacrificing results [19]. Nowadays, mostly ultra-high performance LC (UHPLC) with sub-2-µm particles is used in metabolomics and lipidomics. This trend is also reflected in the data curated from the two repositories. From a total of 399 method descriptions that contain information on the particle size, 252 used sub-2-µm particles in MetaboLights. Likewise, 581 out of 908 descriptions in Metabolomics Workbench used particles smaller than 2 µm.

Additionally, core-shell or superficially porous particles (SPP) emerged as an alternative to sub-2-µm particles. These particles consist of a solid core surrounded by a porous layer containing the stationary phase chemistry. Due to the manufacturing process, core-shell particles typically show a narrower size distribution compared to fully porous particles, leading to reduced Eddy diffusion. Furthermore, they show lower longitudinal diffusion and a shortened diffusion path, yielding better mass transfer. All parameters contribute to higher efficiency entailing higher theoretical plate numbers. Commercial core shell particles are available from 5 µm to even 1.3 µm [20, 21]. Such a very small SPP has been used for the separation of dafachronic acid isomers only differing in the position of a single double bond [22]. However, such highly efficient columns need to be used with low dead volumes in order not to sacrifice the separation, which is often not the case. Furthermore, the very small particle size causes very high backpressures. Nevertheless, it should be noted that *N* in equation 1 is found under the square root. This implies only a 1.4-fold improvement when doubling *N*.

Lastly, the retention influences the resolution. Increased retention can be achieved by increasing column length and/or changing the solvent conditions. However, increasing column length also leads to increased measurement times, conflicting with the often required high-throughput analysis.

Beside the actual stationary phase, the mobile phase has a huge influence on the polarity range that can be separated, as well as on the selectivity of the separation. A first rough classification of the polarity range was made by categorizing the studies into "metabolomics" and "lipidomics'' in case of RP separation. The category "metabolomics" contains all studies that use a gradient from water to typically ACN or MeOH, while the "lipidomics" category uses stronger solvents such as iPrOH for eluting triglycerides. However, the boundary between the two is blurry and certain metabolites and lipids might be analyzed under both conditions. In case of HILIC, such a differentiation is not possible. Based on column and solvents, 35 out of 331 RP descriptions are classified as lipid separations in MetaboLights, whereas in Metabolomics Workbench 141 out of 719 RP descriptions are classified as lipid separations.

Generally, rather simple solvent mixtures (mainly consisting of water, ACN, MeOH and iPrOH) with a selected number of additives are used in metabolomics and lipidomics analysis. Additives in the mobile phase are used as modulators for electrostatic interactions (e.g. hydrogen bonds) of the analyte and the stationary phase, ion pairing substances, complexation agents, chaotropes or to mask specific groups (e.g. silanols). In a stricter sense, buffers influence the electrostatic interactions as well, but they are not covered here. Formic or acetic acid in varying concentrations (mostly 0.1%) are mainly used for pH adjustment, although several studies add ammonium-based buffers. Addition of acids increases the retention of anions and decreases the retention of bases. Furthermore, acids can enhance the ionization of molecules with basic or other protonable compounds in positive ionization mode, while suppressing acidic compounds in negative ionization mode. As listed above, the use of additives has a great influence on the retention, which requires careful and correct selection of additives. Other additives in RP can be perfluoropentanoic acid or trifluoroacetic acid. In case of lipidomics, the addition of trace amounts of phosphoric acid has been shown to improve the peak shape of lipids with terminal phosphate groups such as phosphatidic acids [23]. In HILIC it is more common to use ammonium formate or acetate to buffer the eluents at specific pH values. Moreover, specific methods for anions and cations are often used in HILIC analysis with different eluent compositions optimized for each group of analytes. Based on our data, it seems to be generally accepted in metabolomics approaches to use the same chromatographic method for positive and negative ionization mode under RP conditions. This has the advantage that data from both ionization modes can be combined for identification of metabolites, but might lead to suboptimal performance of the MS. This is different in lipidomics where it appears to be more common to use different methods for positive and negative ionization modes, in particular for the mobile phase additives: ammonium formate and formic acid are used in positive mode, while ammonium acetate and acetic acid are used for negative mode. Our observations in lipidomics are consistent with those reported by Cajca and Fiehn [24]. Optimizing the separation conditions in metabolomics and lipidomics increase the coverage of metabolites as well as performance, leading to narrower and more defined peaks and improving MS sensitivity. However, several metabolites still show bad and irregular peak shapes, especially phosphorylated and poly-phosphorylated metabolites, as well as metabolites with more than one carboxylic acid group, representing highly charged metabolites. This comprises most of the glycolysis and tricarboxylic acid (TCA) cycle intermediates. The bad peak shape in HILIC is due to electrostatic interactions of the ionic analyte with the stationary phase. Using IC would be a better alternative for the analysis of these charged molecules, as mentioned above. One of the most used HILIC column types are zwitterionic phases, for which the addition of trace amounts of phosphate was shown to improve the peak shape of metabolites, potentially by masking strong positively charged groups on the stationary phase [25]. Good peak shapes are important, since most peak picking software for non-targeted metabolomics expects Gaussian or near-Gaussian peaks. Another cause of irregular peak shape are metal traces released from stainless steel of components of the LC system. These metal traces can be counteracted by the addition of chelating substances such as ethylenediaminetetraacetic acid (EDTA) to the mobile phases. However, EDTA can lead to ion suppression in LC-MS. As already mentioned, phosphorylated molecules often lack good peak shapes. Medronic acid has been suggested as a new additive for the analysis of those metabolites with multiple carboxylic acid functionalities and phosphorylated metabolites. The addition of 5 µM medronic acid improved the peak shape of e.g. succinate, malate, isocitrate, adenosine monophosphate, adenosine diphosphate and adenosine triphosphate and can be easily washed out of

the LC-MS instrumentation [26]. Beyond the use of additives, alternative approaches have been developed: Smith *et al.* [27] described the analysis of TCA cycle metabolites in human urine using mixed-mode separation between RP and anion-exchange chromatography. The column itself contained an inorganic-organic hybrid coating reducing the interaction of carboxylic acids with metals. Mixed mode columns are capable of separating polar, ionic and apolar analytes at the same time. Depending on the applied chromatographic conditions (e.g. gradient, solvent composition and pH), HILIC or RP separation may predominate. Wernisch and Pennathur [28] measured more than 750 authentic metabolite standards with seven different columns (one RP C18, two mixed modes and four HILIC columns). The diol-mixed mode column is used in RP- and in HILIC-mode, showing different selectivities. Amino acids and carbohydrates, having multiple polar hydroxyl groups, are better retained in HILIC-mode, whereas nucleosides prefer the RP-mode. Typically, gradient elution is employed in metabolomics and lipidomics due to the large complexity of the metabolome or lipidome and each laboratory has developed its own standard separation gradients including steps of linear and gradient elution. The optimal gradient time depends on e.g. flow rate and column volume (see equation 3). Generally, the length of the gradient increases with the length of the column. The column length and gradient time were found in 326 and 391 data sets of MetaboLights and Metabolomics Workbench, respectively. Gradient times of <5 minutes are very rare, with only 1.5% in MetaboLights and 3.8% in Metabolomics Workbench, whereas gradient times of more than 15 minutes are applied in 77.6% and 59.8% of the data sets, respectively. Looking only at the usual column lengths of 50, 100 and 150 mm, it is evident from particularly the MetaboLights data that longer gradients are used with longer columns (SI Figure 3). For 50 mm columns, 34.4% of the gradients are longer than 15 minutes, for 100 mm lengths 70.4% and for 150 mm lengths 97.6%. This trend is more difficult to see from the Metabolomics Workbench data with 50.0% for 50 mm, 50.7% for 100 mm and 80.2% for 150 mm column length (SI Figure 3). While we didn't check if targeted or non-targeted metabolomics was performed in the specific datasets, non-targeted generally uses longer columns and gradients to achieve a more comprehensive separation, while targeted metabolomics uses shorter and more specific gradients. Additionally, flow rate impacts the retention of metabolites. However, based on the data we didn't see any trend between the column length and flow rate settings.

The total time of a sample measurement consists of the gradient time and the time of the (re-)equilibrium of the column prior to the next injection. Column re-equilibration is of large importance for reproducible results. The equilibrium can be divided into two types: (1) full equilibrium, which means that further increasing time does not affect the retention of all peaks and (2) run-to-run or repeatable equilibrium, where the system is not fully equilibrated with the initial mobile phase but sufficient for a good repeatability of retention times. The equilibrium time is dependent on the stationary phase, solvents, flow rate, etc. In RP, as a rule of thumb, 10 – 20 column volumes are used to flush the column before the next sample for full equilibrium, while in HILIC twice that number is used due to the longer time required for generating the water layer on the stationary phase (especially if high proportions of water are used in the gradient). The column volumes can be calculated with equation 5. Schellinger *et al.* [29] and Heaton *et al.* [30] found that two and eight column volumes for RP and HILIC, respectively, are sufficient for run-to-run repeatability when the conditions are kept constant for all measurements.

$$V = 0.5 \cdot \frac{l \cdot d^2}{1000}$$ (Equation 5) [31]

(with V = column volume [mL], l = length of column [mm] and d = inner diameter [mm])

In 138 datasets (30.7%) of MetaboLights and 72 datasets (7.0%) of Metabolomics Workbench, the column volume of the (re-)equilibrium time could be calculated using equation 5. In RP, it appears to be common to use 10 – 20 column volumes for full equilibrium, instead of the two columns for run-to-run repeatability reported by Schellinger *et al.* [29]. Interestingly, the number of column volumes of the HILIC data is the lowest, even though it is known for longer equilibrium times. Our calculated numbers are even lower than the values of eight column volumes investigated for run-to-run repeatability by Heaton *et al.* [30]. For other modes, no values for equilibrium are found.

Combination of multiple chromatographic methods

Although different separation methods have different polarity ranges, an overlap between them exists and certain metabolites might be well retained under several conditions. An example for Tryptophane and Leucine separated on two columns with different selectivity's is shown in Figure 3. For the high coverage of different metabolites and lipids in a sample, often a combination of several methods is applied. Then, an important question is how data from the different methods can be integrated.

The combination of individual separation systems after data acquisition is often only possible for known metabolites. In case of unknowns one peak in one separation mode can potentially match multiple in the other, requiring additional work on identification and merging results.

Boudah *et al.* [32] and Pezzatti *et al.* [33] both demonstrated the comprehensive polarity ranges of different separation systems by analyzing metabolite standards in positive and negative ionization mode and merging the results afterwards. Boudah *et al.* [32] were able to identify unknown metabolites and to distinguish between isobaric molecules and some isomers in human serum by applying different chromatographic methods in positive and negative ionization mode. With the help of the three different methods, the number of metabolites detected and retained by at least one method was higher leading to an increased metabolites coverage. Pezzatti *et al.* [33] developed a scoring (ranging between 0 (worst) and 1 (best)) of the performance of an analytical system that indicates the quality of the LC separation and the MS sensitivity. Furthermore, it helps finding the best combination of methods in order to analytically detect as many metabolites as possible. For the development of the scoring, nearly 600 authentic metabolites were measured and analyzed under two RP and three HILIC conditions in positive and negative ESI mode. Afterwards, the metabolite retention factor, the signal-to-noise ratio, the peak shape and the intensity were each determined individually, logarithmically transformed and then combined into a total score that represents the quality of the separation system.

Combined analysis of a single injection on different columns can circumvent this since metabolites would only elute "on their specific column". Combination of different separation modes in a single analytical method is a great advantage. One of the first implementation of column coupling has been performed by Greco *et al.* [34] by serially coupling a RP and HILIC column and usage of two pumps for setting individual gradients. This allowed the analysis of polar and non-polar substances from a single injection. A similar setup was used by Haggarty *et al.* [35] for the analysis of polar and non-polar substances from beer as proof of principle. Still, both setups have the problem that the flow is always directly through both

columns, as they are placed one after the other. Hemmler *et al.* [36] developed a system using a short RP trapping column serially coupled with a HILIC column to pre-fractionate the metabolome into polar and non-polar metabolites. After trapping valves are switched in a way that the RP trapping column is in line with the RP separation column and the HILIC column with a second pump. Based on one additional switching valve the stream (either RP or HILIC) that is directed to the MS can be selected. This setup has the advantage that the two columns are isolated from each other during separation and one column can be equilibrated during the separation on the other column, which is especially useful for HILIC requiring longer equilibration times. Another advantage is that due to the prefractionation via the trap column metabolites either elute from the RP or the HILIC column.

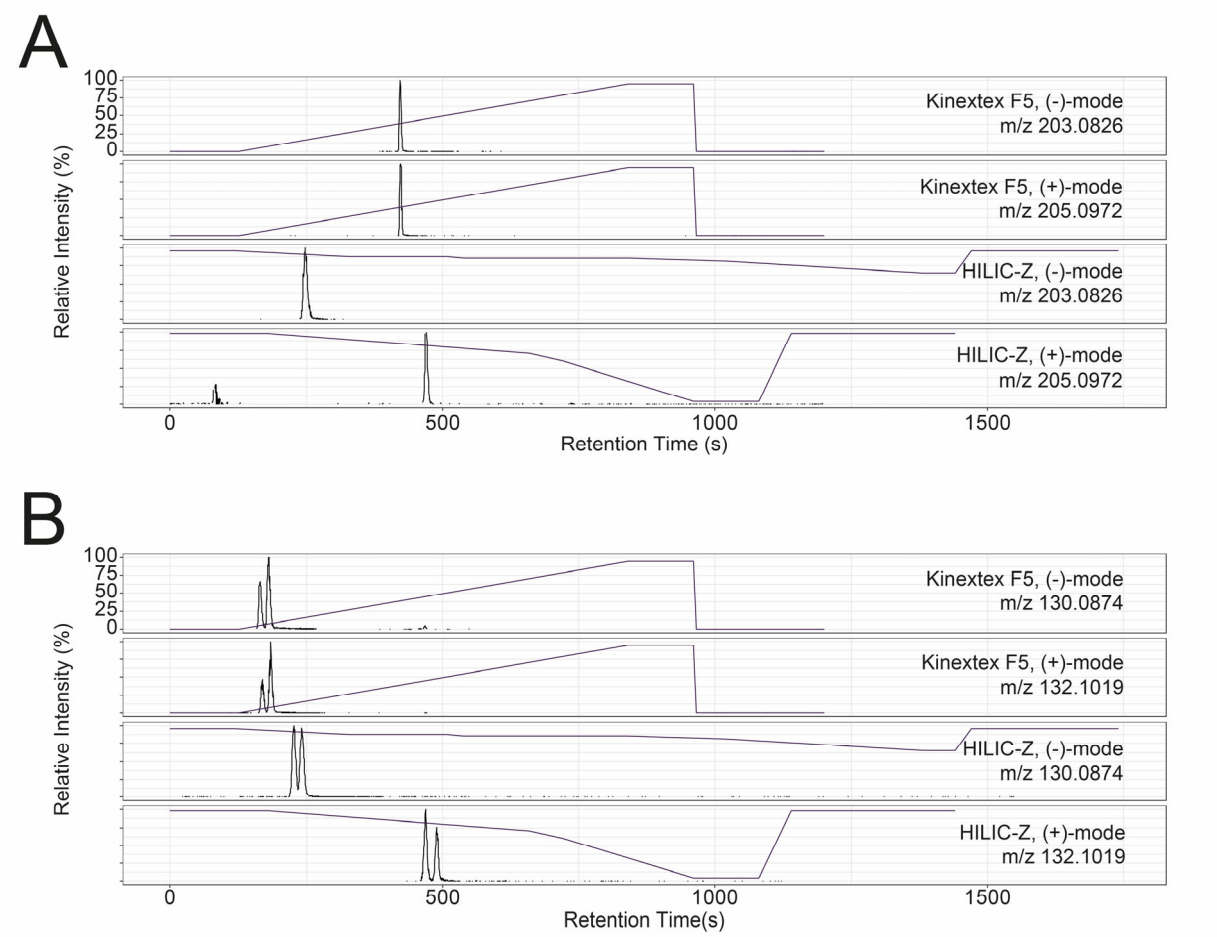

Figure 3: Data integration from multiple methods requires some caution. If in positive and negative ionization mode the same chromatographic method is used, which often is the case for RP separation, the retention times should be the same for a metabolite. (A) The first two chromatograms show an extracted ion chromatogram for Tryptophan from mouse serum samples measured on a Phenomenex Kinetex F5 column using an RP gradient from 100% H2O + 0.1% formic acid to 100% ACN + 0.1% formic acid. The retention time of Tryptophan is identical in both ion modes. In contrast to this, in HILIC often methods are optimized for the respective metabolite classes and therefore also the ion mode. The two lower chromatograms show Tryptophan measured in cell extracts using the method from Hsiao et al. Here different retention times due to different conditions are observed. However, since only a single

peak with the respective m/z is detected, mapping can be performed. (B) The chromatograms show the extracted ion chromatograms for Leucine and were obtained from the same analytical conditions. Two peaks are observed corresponding to Leucin and Isoleucine. While for the F5 column direct mapping between positive and negative mode is possible, this is not directly the case for HILIC, since reversal of elution order might happen.

Two-Dimensional Liquid Chromatography

Two-dimensional (2D) LC offers an increased separation power and helps in separation of very complex samples or when the analytes of interest are difficult to resolve in a single dimension. For that, two different, ideally orthogonal separation modes are used in the first (^{1}D) and the second dimension (^{2}D) (e.g. RP and IC or low pH RP and high pH RP). Effluent or unresolved analytes of the ^{1}D LC are transferred to a second column for further separation of the analytes. Different settings are possible in 2D LC: (1) Single heartcutting 2D LC (LC-LC), where a single fraction corresponding to a peak of interest of the ^{1}D is collected and then transferred to the ^{2}D for further separation. In this method the run-time of the separation in second dimension is independent of the run-time of the first column because no further following fraction from the same sample will be analyzed; (2) Comprehensive 2D LC (LC×LC), where the complete effluent of the ^{1}D is collected in fractions with intervals of 15 – 30 seconds, which are then successively transferred to the ^{2}D. This approach requires more complex instruments (parallel collection of a fraction while separating another) compared to LC-LC. Furthermore, separations in the two dimensions are dependent on each other and the sampling rate between the first and second dimension (modulation time). Hybrids of both approaches are developed (e.g. multiple heartcutting or selective comprehensive 2D LC). Compared to the normal LC, the LC×LC has a higher peak capacity, calculated as the product of the peak capacities of the two dimensions. However, this is only valid for truly orthogonal separation mechanisms and where the resolution of the ^{1}D separation is preserved with only little or no loss during collection and transferring of the fractions (no remixing of the fractions) and complete orthogonality of the chemistry of the two dimension [37]. Although 2D LC can improve the separation space and therefore the metabolome covered, data from a single injection, is spread across multiple chromatograms in the second dimension, which have to be combined afterwards for data analysis. This makes 2D LC data analysis more complicated and requiring special software.

Beside this, in the recent years, the number of comprehensive 2D LC×LC approaches in metabolomics and lipidomics has grown. In metabolomics, Navarro-Reig *et al.* [38] performed a HILIC×RP-MS separation on Japanese rice samples to study the changes in metabolism due to environmental factors. They were able to identify 139 metabolites with relative errors smaller than 10 ppm using in the first step data compression, followed by a chemometric method for the investigation and resolution of the pure elution profile and as a last step a database search and comparison of m/z values for identification. Using RP×RP-MS separations with two different stationary phase chemistries is quite common in metabolomics. For example, Wong *et al.* [39] applied this method coupled to a photodiode array detection using a cyano column in ^{1}D and a C18 column in ^{2}D for the analysis of a *Glycyrrhiza glabra* extract. For a better separation in ^{2}D, a multi-segmented shift gradient was used with different steepness steps and increasing intervals of organic solvent content. About 120 compounds were detected and 37 were tentatively identified. Cacciola *et al.* [40] analyzed the microbosphere of shikimate-producing *Escherichia coli* using a cyano

column in ^{1}D and three serially coupled C8 columns in ^{2}D. Compared to the one-dimensional chromatography, more compounds could be tentatively identified with 2D LC×LC. Yan *et al.* [41] used a combination of analytical and preparative LC×LC for the investigation of a fungal extract of *Chaetomium globosum*. Therefore, different type of RP phases (an octadecyl-, a cholesteryl- and a naphtylethyl-bonded silica) were tested for ^{1}D before applying the preparative RP×RP separation with an octadecyl-based silica for ^{1}D and a C18 column for ^{2}D. 12 compounds could be identified with the help of MS data and NMR measurements. In lipidomics, Li *et al.* [42] and Yang *et al.* [43] used 2D LC quadrupole time-of-flight MS for the investigation of human plasma of peritoneal dialysis (PD) patients and lacunar infarction (LI) patients, respectively. In both studies, ^{1}D was a NP LC for separating different lipid classes while ^{2}D was a RP LC for separation of lipid molecular species. Li at al. [42] were able to identify 190 endogenous lipid species of 10 lipid classes within a single run. Moreover 30 potential biomarkers for malnutrition, inflammation and atherosclerosis syndrome, which often occurs in PD patients, could be found. Yang *et al.* [43] identified 13 potential biomarkers for LI and established a plasma biomarker model for diagnosing LI. Holčapek *et al.* [44] developed another approach using RP as ^{1}D and HILIC as ^{2}D for the separation of lipid extracts of human plasma and porcine brain samples. They identified 143 lipid species of 10 lipid classes within one injection. By using 2D LC the sensitivity of LC-MS analysis can be enhanced. Helmer *et al.* [45] performed analysis of cardiolipin oxidation products in *C. elegans*. Under normal RP conditions cardiolipins would overlap with higher abundant lipids species and potentially be suppressed. In the developed approach they are first separated from other lipids classes by HILIC and then a single-heart cut is performed and individual species are analyzed on a RP column.

Recent advances and trends

In order to cover a large portion of the metabolome and lipidome the combination of different separation methods is required, which is time demanding. This conflicts with the ever-increasing number of samples in metabolomics studies. Since metabolomics is getting more and more employed in large-scale (epidemiological) or genome-wide studies (e.g. genome-wide RNAi screens), a high demand for high-throughput analysis exists. To handle the large number of samples, fast chromatographic methods using ballistic gradients have been developed in recent years. Gray *et al.* [46] developed a fast UHPLC based method requiring only 2.5 minutes. Columns were downscaled from a standard 2.1 mm x 100 mm format to a microbore 1 mm x 50 mm column. Further, the linear solvent velocity was increased. Using urine of control rats or rats administered with acute doses of 2-bromophenol or chronic acetaminophen as proof-of-principle, they showed that the fast method yields similar results compared to the standard method, for separation of samples using a principal component analysis. This method was based on RP separation using a C18 column with water – ACN gradient. As a complement, methods for the analysis of lipids as well as polar metabolites were developed and combined with ion mobility spectrometry (IMS). The HILIC separation required a total time of 3.3 minutes including column re-equilibration and was based on an amide column [47]. Lastly, a lipid separation method over 3.7 minutes with a C8 column was developed and combined with IMS [48].

Even if time is not a constraining factor, the amount of biological starting material might be a limiting factor for analysis on multiple methods. Therefore, another future trend is the miniaturization of separation systems to be able to analyze lower amounts of sample material, ideally going down to single

cells. While in proteomics the use of nanoLC is common practice, in metabolomics it is not or only partially used. However, some applications employing mostly RP have been published. Lipidomics represents one of the main applications since lipids are well-retained and can be easily enriched using trapping setups similar to proteomics or injected directly into nanocolumns. NanoLC has been applied to the analysis of the yeast lipidome by Danne-Rasche *et al.* [49], who found an intensity increase by factor 2 – 3 and an increased coverage of the lipidome. Another positive effect of using nanoLC is the reduction of solvents due to lower flow rates. Application of nanoLC even allowed the analysis of 18 polar metabolites from single HeLa cells [50]. A very recent development was the use of capillary columns with 1.1 µm particles and a length of 200 mm for the analysis of metabolites from human plasma extracts, showing high potential for high resolution separations in metabolomics [51].

Beside the actual improvement of separation systems to better cover different metabolomes, there is also an increased interest in the re-use of RT information. However, RT is a system property since it depends on the analyte itself as well as multiple experimental parameters like the chromatographic column, the dead volume, the eluents, etc. All these factors complicate the re-use of RT because there is no exact or correct RT and each laboratory has "its own RT". Accurate and complete reporting of chromatographic metadata is becoming an important issue to enable the re-use of the data and for building projections between similar systems. A prominent example for sharing of RT data is PredRet [52, 53]. Improved sharing of RT will allow the development of better RT prediction, which also enables better use of the RT dimension for metabolite identification as additional orthogonal information. Recent advances in RT prediction have been reviewed elsewhere [54]. Recently, multiple efforts were done to collect RT data in databases in order to allow (re-)use [55, 56]. However, efforts are still scattered and no central databases such as MassBank, MoNA or GNPS for RT data exist. We recently have started an effort to centralize RT data in a GitHub repository allowing version-controlled tracking of data (https://github.com/michaelwitting/RtPredTrainingData).

Conclusion

The metabolome and lipidome comprises a highly diverse set of chemical structures with a large dynamic range, spanning several orders of the concentration range and a yet unknown number of molecules. MS and NMR are key technologies in metabolomic studies. However, in order to fully exploit MS detection, additional and orthogonal chromatographic separation is required to be able to distinguish between isomeric and isobaric structures, which cannot be differentiated by MS or MS/MS alone. LC offers a broad range of stationary phase chemistries, column dimensions and solvent compositions. No single method is yet able to cover the entire range of the metabolome (and lipidome) and new phases are commercialized and released on a routine basis.

Different methods are employed in LC-MS based metabolomics and lipidomics and the individual methods are far from being standardized between labs. Based on data from the public metabolomics data repositories MetaboLights and Metabolomics Workbench, we investigated the current preferences and selections of columns and separation methods. RP separation represents the most dominant technique but is only able to cover mid- to non-polar metabolites and lipids. In contrast, HILIC better suits for the analysis of polar metabolites. We believe that with improved column technologies and shorter methods

(without the sacrifice of separation power), the combination of multiple chromatographic methods for high coverage metabolomics and lipidomics will be possible in the near future. Lastly, RT will become more and more important, even in early stages of the metabolite identification workflow, moving the interest in enhanced chromatographic methods, combination of methods and RT prediction to the center stage.

Acknowledgment

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project numbers BO 1910/23 and WI 4382/10 (Gepris number 425789784).

Supplementary Information

SI Tables 1-5: Metabolite databases used to generate Figure 1, including all structures and ClassyFire compound classification.

SI Table 6 and 7: Data descriptions from Metabolomics Workbench and MetaboLights used as basis for this review.

Literature

1. Fiehn, O., *Metabolomics - the link between genotypes and phenotypes.* Plant Molecular Biology, 2002. **48**(1): p. 155-171.
2. Haug, K., *et al.*, *MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data.* Nucleic Acids Research, 2013. **41**(D1): p. D781-D786.
3. Sud, M., *et al.*, *Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools.* Nucleic Acids Research, 2016. **44**(D1): p. D463-D470.
4. Bennett, B.D., *et al.*, *Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli.* Nature Chemical Biology, 2009. **5**(8): p. 593-599.
5. van Meer, G., *Cellular lipidomics.* The EMBO Journal, 2005(24): p. 3159-3165.
6. Yin, Y., *et al.*, *DecoMetDIA: Deconvolution of Multiplexed MS/MS Spectra for Metabolite Identification in SWATH-MS-Based Untargeted Metabolomics.* Analytical Chemistry, 2019(91): p. 11897-11904.
7. Wishart, D.S., *et al.*, *HMDB 4.0: the human metabolome database for 2018.* Nucleic Acids Research, 2018. **46**(D1): p. D608-D617.
8. Jewison, T., *et al.*, *YMDB: the Yeast Metabolome Database.* Nucleic Acids Research, 2012. **40**(D1): p. D815-D820.
9. Guo, A.C., *et al.*, *ECMDB: The E. coli Metabolome Database.* Nucleic Acids Research, 2012. **41**(D1): p. D625-D630.
10. Witting, M., *et al.*, *Modeling Meets Metabolomics - The WormJam Consensus Model as Basis for Metabolic Studies in the Model Organism Caenorhabditis elegans.* Frontiers in Molecular Biosciences, 2018. **5**(96).

11. Schläpfer, P., *et al.*, *Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants.* Plant Physiology, 2017. **173**(4): p. 2041-2059.

12. Sakodynskii, K., *The life and scientific works of Michael Tswett.* Journal of Chromatography A, 1972. **73**(2): p. 303-360.

13. Knee, J.M., *et al.*, *A novel ion pairing LC/MS metabolomics protocol for study of a variety of biologically relevant polar metabolites.* Journal of Chromatography B, 2013. **936**: p. 63-73.

14. Walsby-Tickle, J., *et al.*, *Anion-exchange chromatography mass spectrometry provides extensive coverage of primary metabolic pathways revealing altered metabolism in IDH1 mutant cells.* Communications Biology, 2020. **3**(1).

15. Criscuolo, A., *et al.*, *Rational selection of reverse phase columns for high throughput LC-MS lipidomics.* Chemistry and Physics of Lipids, 2019. **221**: p. 120-127.

16. Sampsonidis, I., *et al.*, *Computational analysis and ratiometric comparison approaches aimed to assist column selection in hydrophilic interaction liquid chromatography-tandem mass spectrometry targeted metabolomics.* Journal of Chromatography A, 2015. **1406**: p. 145-155.

17. Bajad, S.U., *et al.*, *Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry.* Journal of Chromatography A, 2006. **1125**(1): p. 76-88.

18. Yuan, M., *et al.*, *A positive/negative ion–switching, targeted mass spectrometry–based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue.* Nature Protocols, 2012. **7**(5): p. 872-881.

19. Wilson, I.D., *et al.*, *High Resolution "Ultra Performance" Liquid Chromatography Coupled to oa-TOF Mass Spectrometry as a Tool for Differential Metabolic Pathway Profiling in Functional Genomic Studies.* Journal of Proteome Research, 2005. **4**(2): p. 591-598.

20. Sanchez, A.C., *et al.*, *Pushing the performance limits of reversed-phase ultra high performance liquid chromatography with 1.3µm core–shell particles.* Journal of Chromatography A, 2013. **1311**: p. 90-97.

21. Fekete, S. and D. Guillarme, *Kinetic evaluation of new generation of column packed with 1.3µm core–shell particles.* Journal of Chromatography A, 2013. **1308**: p. 104-113.

22. Witting, M., *et al.*, *Fast separation and quantification of steroid hormones Delta4- and Delta7-dafachronic acid in Caenorhabditis elegans.* Journal of Chromatography B, 2015. **978-979**: p. 118-21.

23. Knittelfelder, O.L., *et al.*, *A versatile ultra-high performance LC-MS method for lipid profiling.* Journal of Chromatography B, 2014. **951-952**: p. 119-128.

24. Cajka, T. and O. Fiehn, *Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics.* Analytical Chemistry, 2016. **88**(1): p. 524-545.

25. Spalding, J.L., *et al.*, *Trace Phosphate Improves ZIC-pHILIC Peak Shape, Sensitivity, and Coverage for Untargeted Metabolomics.* Journal of Proteome Research, 2018. **17**(10): p. 3537-3546.

26. Hsiao, J.J., *et al.*, *Improved LC-MS Methods fro the Analysis of Anionic Analytes.* Agilent Technologies, 2018: p. 1-4.

27. Smith, K.M., I.D. Wilson, and P.D. Rainville, *Sensitive and Reproducible Mass Spectrometry-Compatible RP-UHPLC Analysis of Tricarboxylic Acid Cycle and Related Metabolites in Biological Fluids: Application to Human Urine.* Analytical Chemistry, 2021. **93**(2): p. 1009-1015.

28. Wernisch, S. and S. Pennathur, *Evaluation of coverage, retention patterns, and selectivity of seven liquid chromatographic methods for metabolomics.* Analytical and Bioanalytical Chemistry, 2016. **408**(22): p. 6079-6091.

29. Schellinger, A.P., D.R. Stoll, and P.W. Carr, *High speed gradient elution reversed-phase liquid chromatography.* Journal of Chromatography A, 2005. **1064**: p. 143-156.

30. Heaton, J.C., N.W. Smith, and D.V. McCalley, *Retention characteristics of some antibiotic and anti-retroviral compounds in hydrophilic interaction chromatography using isocratic elution, and gradient elution with repeatable partial equilibration.* Analytica Chimica Acta, 2019. **1045**: p. 141-151.

31. Dolan, J.W., *Be Careful of the Flow Rate.* LCGC North America, 2015. **33**(4): p. 248-253.

32. Boudah, S., *et al.*, *Annotation of the human serum metabolome by coupling three liquid chromatography methods to high-resolution mass spectrometry.* Journal of Chromatography B, 2014. **966**: p. 34-47.

33. Pezzatti, J., *et al.*, *A scoring approach for multi-platform acquisition in metabolomics.* Journal of Chromatography A, 2019. **1592**: p. 47-54.

34. Greco, G., S. Grosse, and T. Letzel, *Serial coupling of reversed-phase and zwitterionic hydrophilic interaction LC/MS for the analysis of polar and nonpolar phenols in wine.* Journal of Separation Science, 2013. **36**(8): p. 1379-1388.

35. Haggarty, J., *et al.*, *Serially coupling hydrophobic interaction and reversed-phase chromatography with simultaneous gradients provides greater coverage of the metabolome.* Metabolomics, 2015. **11**(5): p. 1465-1470.

36. Hemmler, D., *et al.*, *Tandem HILIC-RP liquid chromatography for increased polarity coverage in food analysis.* Electrophoresis, 2018. **39**(13): p. 1645-1653.

37. Stoll, D.R. and P.W. Carr, *Two-Dimensional Liquid Chromatography: A State of the Art Tutorial.* Analytical Chemistry, 2017. **89**(1): p. 519-531.

38. Navarro-Reig, M., *et al.*, *Untargeted Comprehensive Two-Dimensional Liquid Chromatography Coupled with High-Resolution Mass Spectrometry Analysis of Rice Metabolome Using Multivariate Curve Resolution.* Analytical Chemistry, 2017. **89**(14): p. 7675-7683.

39. Wong, Y.F., *et al.*, *Untargeted profiling of Glycyrrhiza glabra extract with comprehensive two-dimensional liquid chromatography-mass spectrometry using multi-segmented shift gradients in the second dimension: Expanding the metabolic coverage.* ELECTROPHORESIS, 2018. **39**(15): p. 1993-2000.

40. Cacciola, F., *et al.*, *Novel comprehensive multidimensional liquid chromatography approach for elucidation of the microbosphere of shikimate-producing Escherichia coli SP1.1/pKD15.071 strain.* Analytical and Bioanalytical Chemistry, 2018. **410**(15): p. 3473-3482.

41. Yan, X., *et al.*, *New on-line separation workflow of microbial metabolites via hyphenation of analytical and preparative comprehensive two-dimensional liquid chromatography.* Journal of Chromatography B, 2016. **1033-1034**: p. 1-8.

42. Li, M., *et al.*, *Lipid profiling of human plasma from peritoneal dialysis patients using an improved 2D (NP/RP) LC-QToF MS method.* Analytical and Bioanalytical Chemistry, 2013. **405**(21): p. 6629-6638.

43. Yang, L., *et al.*, *Lipidomic analysis of plasma in patients with lacunar infarction using normal-phase/reversed-phase two-dimensional liquid chromatography–quadrupole time-of-flight mass spectrometry.* Analytical and Bioanalytical Chemistry, 2017. **409**(12): p. 3211-3222.

44. Holčapek, M., *et al.*, *Continuous comprehensive two-dimensional liquid chromatography–electrospray ionization mass spectrometry of complex lipidomic samples.* Analytical and Bioanalytical Chemistry, 2015. **407**(17): p. 5033-5043.

45. Helmer, P.O., *et al.*, *Investigation of cardiolipin oxidation products as a new endpoint for oxidative stress in C. elegans by means of online two-dimensional liquid chromatography and high-resolution mass spectrometry.* Free Radical Biology and Medicine, 2021. **162**: p. 216-224.

46. Gray, N., *et al.*, *Development of a Rapid Microbore Metabolic Profiling Ultraperformance Liquid Chromatography–Mass Spectrometry Approach for High-Throughput Phenotyping Studies.* Analytical Chemistry, 2016. **88**(11): p. 5742-5751.

47. King, A.M., *et al.*, *Development of a rapid profiling method for the analysis of polar analytes in urine using HILIC–MS and ion mobility enabled HILIC–MS.* Metabolomics, 2019. **15**(2).

48. King, A.M., *et al.*, *Rapid profiling method for the analysis of lipids in human plasma using ion mobility enabled-reversed phase-ultra high performance liquid chromatography/mass spectrometry.* Journal of Chromatography A, 2020. **1611**: p. 1-10.

49. Danne-Rasche, N., C. Coman, and R. Ahrends, *Nano-LC/NSI MS Refines Lipidomics by Enhancing Lipid Coverage, Measurement Sensitivity, and Linear Dynamic Range.* Analytical Chemistry, 2018. **90**(13): p. 8093-8101.

50. Nakatani, K., *et al.*, *An Analytical System for Single-Cell Metabolomics of Typical Mammalian Cells Based on Highly Sensitive Nano-Liquid Chromatography Tandem Mass Spectrometry.* Mass Spectrometry, 2020. **9**(1).

51. Sorensen, M.J. and R.T. Kennedy, *Capillary ultrahigh-pressure liquid chromatography-mass spectrometry for fast and high resolution metabolomics separations.* Journal of Chromatography A, 2021. **1635**: p. 461706.

52. Stanstrup, J., S. Neumann, and U. Vrhovšek, *PredRet: Prediction of Retention Time by Direct Mapping between Multiple Chromatographic Systems.* Analytical Chemistry, 2015. **87**(18): p. 9421-9428.

53. Low, D.Y., *et al.*, *Data sharing in PredRet for accurate prediction of retention time: Application to plant food bioactive compounds.* Food Chemistry, 2021. **357**: p. 129757.

54. Witting, M. and S. Böcker, *Current status of retention time prediction in metabolite identification.* Journal of Separation Science, 2020. **43**(9-10): p. 1746-1754.

55. Naylor, B.C., *et al.*, *QSRR Automator: A Tool for Automating Retention Time Prediction in Lipidomics and Metabolomics.* Metabolites, 2020. **10**(6): p. 237.

56. Domingo-Almenara, X., *et al.*, *The METLIN small molecule dataset for machine learning-based retention time prediction.* Nature Communications, 2019. **10**(1).