Analytical and Bioanalytical Chemistry (2018) 410:1873–1884
https://doi.org/10.1007/s00216-018-0860-x

PAPER IN FOREFRONT

Metabolomic spectral libraries for data-independent SWATH liquid chromatography mass spectrometry acquisition

Tobias Bruderer[1] · Emmanuel Varesio[2] · Anita O. Hidasi[1] · Eva Duchoslav[3] · Lyle Burton[3] · Ron Bonner[4] · Gérard Hopfgartner[1]

Received: 3 November 2017 / Revised: 12 December 2017 / Accepted: 8 January 2018 / Published online: 6 February 2018

Abstract

High-quality mass spectral libraries have become crucial in mass spectrometry-based metabolomics. Here, we investigate a workflow to generate accurate mass discrete and composite spectral libraries for metabolite identification and for SWATH mass spectrometry data processing. Discrete collision energy (5–100 eV) accurate mass spectra were collected for 532 metabolites from the human metabolome database (HMDB) by flow injection analysis and compiled into composite spectra over a large collision energy range (e.g., 10–70 eV). Full scan response factors were also calculated. Software tools based on accurate mass and predictive fragmentation were specially developed and found to be essential for construction and quality control of the spectral library. First, elemental compositions constrained by the elemental composition of the precursor ion were calculated for all fragments. Secondly, all possible fragments were generated from the compound structure and were filtered based on their elemental compositions. From the discrete spectra, it was possible to analyze the specific fragment form at each collision energy and it was found that a relatively large collision energy range (10–70 eV) gives informative MS/MS spectra for library searches. From the composite spectra, it was possible to characterize specific neutral losses as radical losses using in silico fragmentation. Radical losses (generating radical cations) were found to be more prominent than expected. From 532 metabolites, 489 provided a signal in positive mode $[M+H]^+$ and 483 in negative mode $[M-H]^-$. MS/MS spectra were obtained for 399 compounds in positive mode and for 462 in negative mode; 329 metabolites generated suitable spectra in both modes. Using the spectral library, LC retention time, response factors to analyze data-independent LC-SWATH-MS data allowed the identification of 39 (positive mode) and 72 (negative mode) metabolites in a plasma pool sample (total 92 metabolites) where 81 previously were reported in HMDB to be found in plasma.

Keywords MS/MS libraries · High-resolution LC-MS · Data-independent acquisition · SWATH · Metabolomics

Introduction

Metabolomics ideally requires the analysis of all metabolites in a variety of body fluids for a large number of samples so that the statistical power is high and valid conclusions can be made about differences among sample groups. Although the most recent release of the human metabolome database (HMDB, version 4.0) [1] contains 113,872 metabolite entries including both water- and lipid-soluble metabolites, covering various classes of compounds, all analytical techniques have limitations which require compromises such as compound coverage, analyte concentration dynamic range, ease of sample preparation, and analysis time. Along with other techniques, tandem mass spectrometry combined with ultra high-performance liquid chromatography [2](UHPLC-MS/MS) is widely used for both qualitative and quantitative

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00216-018-0860-x) contains supplementary material, which is available to authorized users.

✉ Gérard Hopfgartner
gerard.hopfgartner@unige.ch

[1] Life Sciences Mass Spectrometry, Department of Inorganic and Analytical Chemistry, University of Geneva, 24, Quai Ernest Ansermet, 1211 Geneva 4, Switzerland

[2] School of Pharmaceutical Sciences, University of Geneva, University of Lausanne, Rue Michel-Servet 1, 1211 Geneva 4, Switzerland

[3] Sciex, 71 Four Valley Drive, Concord, ON L4K 4V8, Canada

[4] Ron Bonner Consulting, Newmarket, ON L3Y 3C7, Canada

analyses of metabolites. However, a number of considerations in both stages, for example, hydrophobicity, pKa, MS response, concentration dynamic range, sample volume, and analytical throughput as well as assay precision and accuracy, limit the number of analytes which can be covered simultaneously. Precise identification and accurate quantification of metabolites in various systems (e.g., human body fluids, cells, food, plants) has become essential to transform analytical data into biological knowledge. This can only be achieved with the use of high-quality annotated metabolite databases including MS1, MS/MS, retention times, response factors, and additional metadata based on high-resolution mass spectrometry [3, 4]. In silico fragmentation tools are becoming more popular to be used for MS/MS spectrum annotation or for compound identification based on compound database search [5]. LC-MS/MS high-resolution libraries including retention information are used in the field of forensic and environmental sciences [6, 7]. High-resolution mass spectrometry is also widely applied in metabolomics studies often using data-dependent acquisition (DDA) where molecular ions detected in an MS survey scan are automatically targeted for subsequent acquisition of product ions scan. The survey scan provides accurate mass and isotopic distributions, allowing elemental formula to be determined and it is used for quantification while the MS/MS scan is only used for identification. DDA is a powerful technique but suffers from several critical limitations, in particular the selection process is stochastic and therefore not reproducible across samples; also quantitation from precursor ions is less specific than using fragment ions and results in lower signal-to-noise ratios.

Consequently, data-independent acquisition (DIA) schemes, such as MS^{ALL} or SWATH, have attracted attention. Sequential windowed acquisition of all theoretical MS (SWATH) has emerged as a key technique for peptide quantification since it is non-stochastic and allows post-acquisition SRM-like quantification for almost any analyte previously identified [8]. While SWATH/MS has been mostly used for proteomics, these features are valuable for metabolomics, metabolism, and forensic studies [9–12] and provide qualitative and quantitative analysis in the same run (QUAL/QUAN workflow).

As authentic standards are only available for relatively few metabolites, metabolite identification is a major challenge in metabolomics and various approaches have been described including database searches [13] and MS/MS spectral interpretation [4, 14]. In contrast to the ionization techniques used in LC-MS/MS, electron impact (EI) spectra generation is well standardized and comprehensive and spectral libraries have been generated for use in GC-MS and applied to metabolite identification [15]. For LC-MS, various web-based libraries are available (such as Metlin, MassBank, HMDB, and LipidMaps [13]) but collision-induced dissociation (CID) spectra can be generated with various techniques and conditions, including trap CID, quadrupole CID, and high collision dissociation (HCD), which can challenge database searches [16]. Library building, especially by DDA, is further confounded by degradation of standard compounds and/or impurities, unexpected and unpredictable molecular ions, different adducts (e.g., $[M+H]^+$ vs. $[M+Na]^+$), and co-selection of background contaminant ions, and therefore, it can be time consuming and error-prone. Since spectral libraries are also required for all targeted quantitation methods, e.g., SWATH and SRM, the timely generation of high-quality spectral libraries, complete with collision energy information and retention times for different separation techniques in particular for liquid chromatography [17], has become critical.

In the present work, we describe the generation of a discrete and composite collision energy accurate mass spectral library of 532 metabolites, reported in the human metabolome database, using flow injection analysis complemented with retention time information. In addition, electrospray response factors were also determined and frequency of neutral loss and radical fragmentation in the set of analytes is reported. The applicability of the library to identify plasma metabolites by LC-MS/MS using SWATH data-independent acquisition has been investigated.

Material and methods

Chemicals and solvents

Water (Millipore), methanol, acetonitrile, and isopropanol (all HPLC grade) were provided by VWR (Darmstadt, Germany); formic acid, ammonium hydroxide, ammonium formate, and ammonium acetate were provided by Sigma-Aldrich (Buchs, Switzerland).

The 532 library compounds were obtained from the Human Metabolome Database (HMDB) and/or purchased from Sigma-Aldrich as powder or liquids (see Electronic Supplementary Material (ESM) Tables S1 and S2). Stock concentrations were made by dissolving the analytes in mixtures of water, methanol, acetonitrile, ethanol, isopropanol, or acetone to obtain a final concentration in the range 0.5–1 mg/ml. Working solutions of individual standards were prepared at 10 μg/ml in the same solvent as the stock solutions with further dilutions in steps of 10. A standard sample set composed of 52 metabolites was used as a quality control mix (ESM Table S3). The solutions were prepared taking into account the response factors of the different analytes.

Plasma samples

Plasma samples were obtained from the University Hospital of Geneva with 15 U/ml heparin and stored at about − 20 °C. A plasma pool was thawed on ice and aliquots of 200 μl plasma were taken and 600 μl methanol added followed by

centrifugation, evaporation to dryness, reconstitution at gradient initial conditions, and sonication. Plasma pool samples were diluted 10× with mobile phase at the gradient starting conditions.

Internal standards were added before analysis with a final concentration of 1 μg/ml for tryptophane-^{15}N2, 100 ng/ml for benzamide-^{15}N (Cambridge Isotope Laboratories, Andover MA, United States) and 1 μg/ml for phenylalanine-^{13}C, estrone-$^{13}C_3$, and myristic acid-^{13}C for ESI negative only and 10 ng/ml for testosterone-$^{13}C_3$ for ESI positive only (all provided by Sigma-Aldrich, Buchs, Switzerland, except as noted).

Flow injection analysis for generation of spectral library entries

For flow injection analysis (FIA), a μLC10AD pump (Shimadzu, Japan) was connected to a PAL autosampler (CTC, Switzerland). The mobile phases for analysis were as follows. Positive mode: solvent A: 5 mM ammonium formate in water, adjusted to pH = 3.0 with formic acid; negative mode: solvent B: 5 mM ammonium acetate in water, adjusted to pH = 8.0 with ammonium hydroxide for ESI negative mode.

The flow rate was 200 μl/min with an injection volume of 200 μl for TOF MS and 500 μl for TOF MS/MS acquisitions operating Q1 at unit mass. A TripleTOF 5600 mass spectrometer (AB Sciex, Concord ON, Canada) was used for MS detection. TOF MS and TOF MS/MS spectra for the library compounds were acquired with 16 collision energy steps (5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, and 100 eV) and an accumulation time of 100 msec for each step. In MS/MS mode, the transmission was set to 100%. Possible saturation of the MS/MS signal was monitored for all analytes by measuring the mass error over all fragments in the composite spectra and was not found to be problematic. The sprayer capillary and DP voltages were 5300 and 70 V in positive mode and − 4300 and − 80 V in negative mode. The source temperature was set at 300 °C, the curtain gas was 25, and gas 1 and gas 2 were 30. Batches for MS and MS/MS data collection were created either manually using the standard acquisition software (Analyst TF 1.6 or 1.7.1) or with an in-house software tool named MS/MS Batch Builder. Batch Builder generates files necessary to automate acquisition; the user selects the precursor type (e.g., $[M+H]^+$, $[M+Na]^+$, $[M+NH4]^+$) and provides a list of elemental compositions from which acquisition methods for multiple compounds and multiple collision energies are built. Additional analyte information (e.g., name, formula, CAS, CID) can also be added. The process and output format are illustrated in the ESM Fig. S1.

LC-MS for quality control and plasma analyses

Chromatographic separation was achieved on an UltiMate 3000 RSLC (Dionex, Germering, Germany) using a 2.1 × 150-mm, 2.5-μm T3 Xselect column HSS XP (Waters) at a flow rate of 300 μl/min with the column heated at 40 °C. Mobile phases for ESI positive were solvent A: 5 mM ammonium formate in water, adjusted to pH = 3.0 with formic acid; solvent B: methanol. The solvents for ESI negative mode were solvent C: 5 mM ammonium acetate in water, adjusted to pH = 8.0 with ammonium hydroxide and solvent D: methanol. The gradient used in for QC mix and plasma samples was 1% B (1 min), 5–95% B (20 min), 100% B (3 min), 1% B (4 min). The injection volume for plasma samples was 1–10 μl. Mass spectrometric detection was performed for plasma samples on a TripleTOF 5600 and on a TripleTOF 6600 for the quality control mix (QC mix). For SWATH acquisition, a single TOF MS scan was followed by 12 MS/MS experiments with variable Q1 windows with a mass range from *m*/*z* 80 to 800 for positive mode QC mix and from *m*/*z* 50 to 990 for positive mode and *m*/*z* 60 to 990 for negative mode for the pooled plasma (ESM Table S4). The variable Q1 windows were built with an in-house developed software tool (swathTUNER) [18]. The cycle time was adjusted to 831 ms based on an average LC peak width of 12 s to obtain at least 12 points/peak. A collision energy spread of 40 ± 30 eV was applied. The sprayer capillary voltage was of 5300 V, − 4300 V with a DP of ± 70 V and a source temperature of 450 °C for positive and negative modes. The curtain gas was set at 25 and the gas 1 and gas 2 at 40.

PeakView 2.2 with the MasterView 1.1 package (Sciex) was used for data evaluation and library search of QC mix and plasma samples. The MasterView search criteria were set as follows: XIC intensities above 100 counts or S:N > 10, XIC width 10 mmu, with the following library search parameters: confirmation search, precursor mass tolerance of 0.4 Da (smaller tolerance could also be used), a fragment mass accuracy of 5 ppm, polarity filter was applied, an intensity threshold of 1%, a minimal purity of 0.1% and an intensity factor of 100.

Database creation and curation

Acquired TOF product ion spectra were background subtracted, recalibrated using the residual precursor (assumed to be the closest significant peak within the precursor selection window) as a reference and stored in a spectral library (MS Access format) with data acquisition parameters and compound information including structures.

A prototype tool, "Populate Compound Library," was used to assess the completeness and quality of the MS/MS spectra. This tool allows review and management of spectral library entries and annotates accurate mass fragments and neutral

losses using the compound structure as a reference and enables an iterative process for library construction where low-quality spectra are identified and flagged for further consideration. Furthermore, the tool provides statistics about the overall library quality such as the total number of fragment ions and the fraction and the percentage of the summed ion intensity successfully annotated.

The first step in the process is assigning elemental compositions to the observed fragment ions using the formula of the precursor and adduct species such as H, Na, or K to constrain the atoms considered and a narrow *m*/*z* window (e.g., 3 mDa). If the summed intensity of the annotated peaks is less than a given portion (e.g., 90%) of the overall sum, the spectrum is flagged and examined manually to determine the cause:

i) An incorrect peak was used as a reference for recalibration; this can be identified by a consistent, large error for several peaks and corrected by recalibrating using the correct precursor.
ii) The signal and hence the mass accuracy were low; in this case the sample is re-analyzed at a higher concentration.

If the intensity of the assigned peaks remains below 90% after these steps, the spectrum is rejected from the library since either the standard has decomposed or the response is poor.

The second step is to determine if plausible structures can be assigned to the annotated ions. The tool uses a brute-force approach to generate all possible fragments of the target structure; the parameters are user-selectable, but generally allow up to three single or aromatic bonds, including C–C bonds, to be broken, and the resulting fragment to be even- or odd-electron, i.e., radicals are allowed. Rearrangement ions are not permitted since these are hard to predict. This approach was (or has been) implemented in the Fragment Pane (ESM Fig. S2). Predicted fragments matching elemental compositions determined in the first step are scored according to mass error, the number and type of bond cleavages, H-atom transfers required and electron parity; those with the highest scores are retained for neutral loss and radical loss analyses.

Results and discussion

In metabolomics, with a growing number of compounds of interest, the use of libraries is essential for compound identification and to build quantitative methods for multi-analyte analyses. Most of the available public libraries are built from discrete collision energy data or from collision energy range data and from different instruments with limited information on data quality. A workflow (Fig. 1) was developed to collect MS and MS/MS spectra over a large collision energy range (5–100 eV) for 532 metabolites reported by the HMDB database; the curated collection is reported as the Accurate Mass Metabolite Library (AMML).

Flow injection analysis (FIA) was used to obtain MS response factors (Fig. 1B) and to record 16 discrete MS/MS experiments in the range 10 to 100 eV (Fig. 1C); average spectra were extracted from the plateau of the FIA trace. FIA acquisition was selected to allow consistent ionization conditions and sufficient time to generate MS/MS spectra at multiple collision energies, and because a generic LC system is impossible given the large chemical space of the analytes. Saturation of the MS/MS fragment signals was not observed for the compounds investigated.

From 532 metabolites (ESM Table S2), 489 metabolites provided a signal in positive mode $[M+H]^+$ and 483 in negative mode $[M-H]^-$. MS/MS spectra were obtained for 399 compounds in positive mode and for 462 in negative mode; 329 metabolites generated usable spectra in both modes.

In electrospray-MS, the response is analyte dependent for given solvent conditions. In the present work, normalized FIA conditions were considered to obtain information about the response factor of the analyte useful for relative quantification. Figure 2 summarizes the MS response in positive and negative mode for (i) amino acids, peptides, and analogues; (ii) steroids and derivatives (lipids); (iii) benzoic acids and derivatives; (iv) organoheterocyclic compounds; (v) carbohydrates and conjugates; and (vi) nucleosides, nucleotides, and analogues (additional data in ESM Fig. S3). Beside the expected large response dynamic range, that has exceeded five orders of magnitude, adducts, such as ammonium and sodium in positive mode and formic acid and chlorine in negative mode, were observed. For some compound classes, e.g., carbohydrates and conjugates), the median adduct response exceeded that from the simple precursor ($[M+H]^+$ or $[M-H]^-$). Adducts formation is a critical point as the protonated or deprotonated molecule may not be observable and as the adduct ratio may change from sample to sample. While ammonium adducts generate good-quality CID spectra, sodium adducts do not fragment well. Despite the soft declustering conditions used, many metabolites also showed losses of water which may generate false-positive library hits. In CID, water loss is frequently observed from many metabolites and chemicals [19].

Database generation and quality control of the library

The fragment ions observed in collision induced dissociation are strongly dependent on the specific instrument and collision energy settings [20]. Many libraries are built either with a few discrete collision energy such as 10, 20, and 40 eV [21] or multi-collision energies (10 to 20 steps) [22] serving as a base for the generation of consensus spectra. Since AMML is also intended to assist in developing low- or high-resolution SRM acquisition methods, storing spectra acquired at discrete

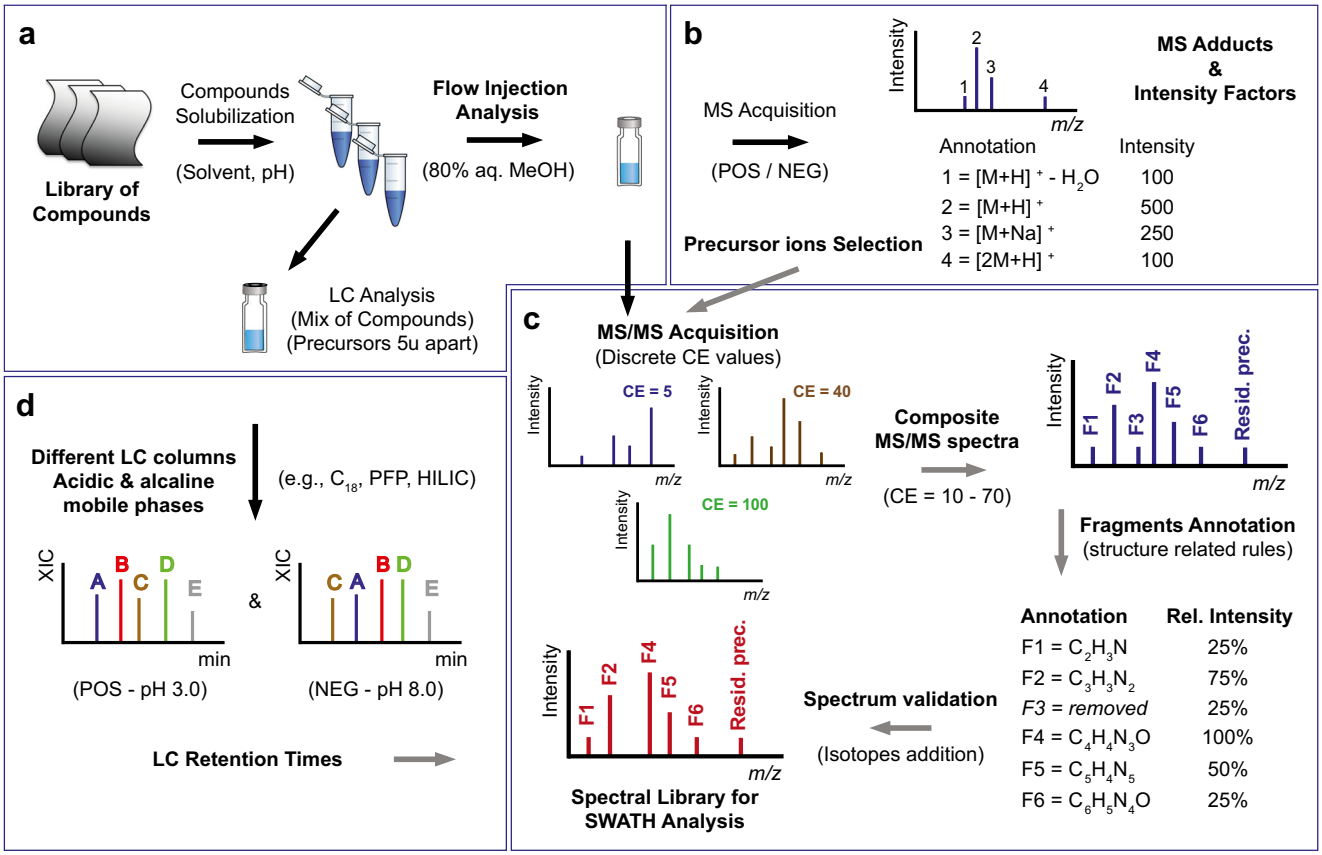

Fig. 1 Overall library generation workflow. (A) Sample preparation: compounds are analyzed individually by flow injection analysis (FIA) MS and pooled by LC-MS; (B) MS acquisition: MS spectra provide response and adduct information; (C) MS/MS acquisition and validation: MS/MS spectra are acquired at several CE values and checked for consistency with the known structure (see text); (D) LC-MS acquisition: pools of 5–10 non-isobaric compounds are acquired in positive and negative modes with different pH values to provide retention times

collision energies provides optimum precursor-product ion pairs and the corresponding collision energy.

For all metabolites present in the AMML library, a plot of the number of novel MS/MS fragments for a given collision energy relative to the previous collision energy is shown in Fig. 3 for the range 10 to 100 eV in positive and negative modes. In positive mode, novel fragment production maximizes around 25 eV and again around 70 eV. The lower maximum corresponds to neutral losses (e.g., water, ammonia) and single bond cleavages while the maximum at higher energy involves multiple bonds and ring cleavages. Higher energy fragments are often less intense and generally of lower *m/z* but can still be informative. In negative mode a maximum was found at −20 eV. Therefore, composite spectra using the ranges 10–70 or 20–80 eV will provide the most useful information. With unit mass precursor isolation and possible interfering ions, even in pure standards, quality control of MS/MS spectra is essential and several different approaches have been described including spectra annotation based on elemental formula [7, 23]. Analyte purity and stability is always a concern when building reference libraries independent of the source of material. Authenticity was confirmed on both MS and MS/MS data using elemental composition and fragment annotation. For MS/MS acquisition, the analyte concentration was adapted in such a way that more than 90% of the fragment signal could be associated with the analyte measured.

Mass spectra annotation is performed in two steps: first, similar to Stein et al. [23] elemental compositions are calculated using the compound's formula to define maxima for each element and a narrow error range; precursor ions where no composition can be found are tracked and spectra with unassigned fragments exceeding a given limit are reviewed, curated, or rejected. Recalibration of the MS/MS spectra is essential for providing the most accurate fragment *m/z* values and is easily performed as the residual precursor ion is always present in the composite spectra. To automate processing, the closest ion within the selected precursor window is assumed to be the correct precursor and used as the recalibration standard. This however, can result in fragment mass errors if the assumption made is incorrect. To detect this possibility, we compare the elemental formula assigned to each fragment based on the expected compound elemental composition [24] to in silico fragment structures generated from the structure of the precursor metabolite (ESM Fig. S2).

Overall, considering MS/MS fragments having relative intensity above 2% and using a mass tolerance of 3 mDa, 4821

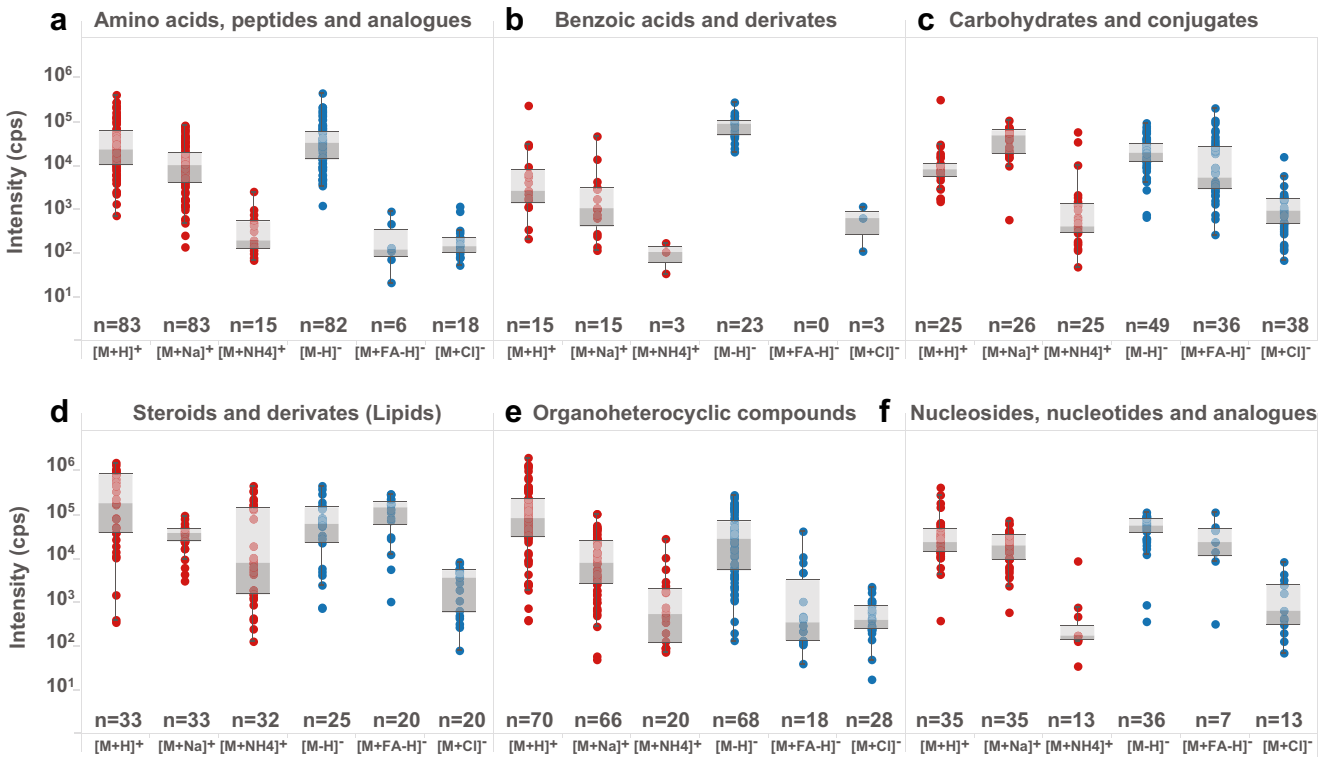

Fig. 2 Electrospray FIA MS responses at 10 ng/μL of selected library metabolite classes (HMDB definition) in positive (red) and negative (blue) mode for (A) amino acids, peptides, and analogues; (B) benzoic acids and derivatives; (C) carbohydrates and conjugates; (D) steroids and derivatives (lipids); (E) organoheterocyclic compounds; and F) nucleosides, nucleotides, and analogues

fragments were assigned in positive mode (explaining 99.4% TIC) and 2576 in negative mode (99.7% TIC). Allowing cleavage of up to three single or aromatic bonds, 83% of assigned positive peaks and 94% of assigned negative peaks were assigned fragment structures. In addition, 7074 accurate neutral losses were annotated in positive mode and 2758 in negative mode. The overall *m/z* precision and accuracy of annotated fragments (-0.1 ± 0.6 and 0.0 ± 0.4 mDa, respectively, see ESM Fig. S4) is consistent with the anticipated instrument performance and does not show any bias in any mass range segments. In positive mode, the number of MS/MS spectra with 100% annotation is 345 and in negative mode 432. After removing spectra where less than 90% of the TIC could be annotated, the final library represented 532 compounds. While these filtering criteria may appear very stringent, this was deemed an important feature of the library in order to reduce the likelihood of false-positive matches and incorrect targeting (ESM Table S5).

A standard sample set composed of 52 metabolites (QC mix) selected from those identified in urine and plasma was generated and used for library quality control using a different instrument than that for library generation (TripleTOF 5600 versus TripleTOF 6600, ESM Table S3). A representative chromatogram is presented in ESM Fig. S5 showing 18 analytes eluting in a 0.3-min window. All 52 metabolites could be positively identified based on mass accuracy of the precursor ion and library hit (median score 0.98 with MasterView candidate search), (ESM Table S3). Representative full scan and SWATH/MS spectra are presented in ESM Figs. S6 to S9. For histidine (ESM Fig. S6) two additional fragments were recorded at 130.0862 and 147.1123. Various spectra filtering procedure could be considered, but just using elemental formula filtering these two ions can be clearly eliminated. In the present experiment, calculated theoretical isotopes were added to the AMML library spectra [7]. This is particularly useful for confirmation of analytes containing Br and Cl as they showed a specific isotopic profile as presented for clotrimazole (ESM Fig. S8).

FIA was used to provide sufficient time to generate high-quality MS/MS data and to normalize the response factor, but retention time information can help metabolite identification, especially as many analyses are performed by LC-MS and isomers and isobaric compounds are likely. This is especially true if library data are to be used to target compounds in data acquired using SWATH since the wide precursor ion selection windows almost guarantee co-selection of multiple precursor ions. Due to the large chemical space of the metabolome, and as an alternative to using different column types, we performed LC characterization using a standard reversed phase column operated at two different pH values (3 and 8), with detection in positive and negative modes, respectively. It should be noted that including this information in MS/MS

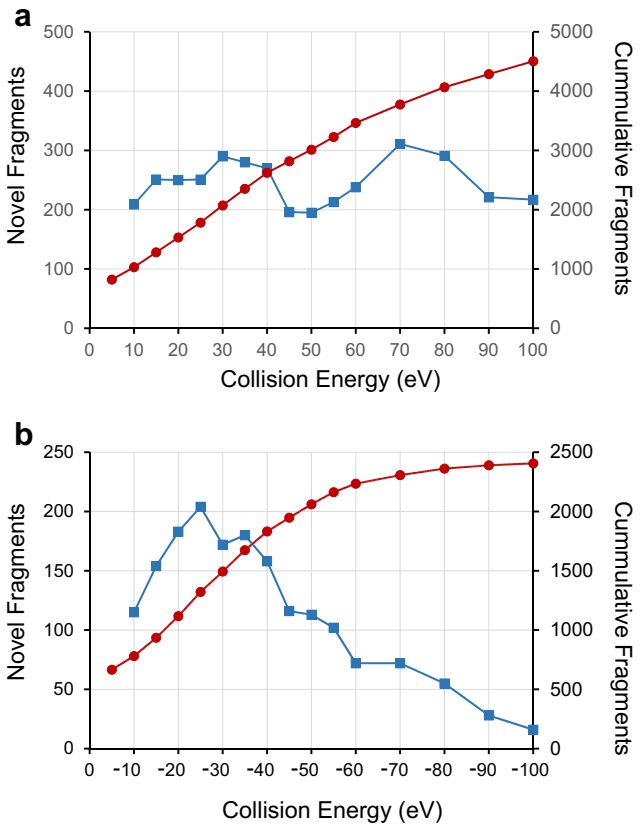

Fig. 3 Summary of novel fragments observed at each collision energy relative to previous collision energy in positive and negative modes (blue trace). (A) Positive mode. (B) Negative mode. The red trace shows the cumulative number of fragments

libraries, and performing appropriate data acquisition and spectrum matching, would increase the confidence of metabolite identification if the compound is detected in both modes. We were able to determine LC retention times for 339 compounds in positive mode, 410 in negative and 264 in both modes. Furthermore, some isomers co-eluted under one set of conditions but were separated under the other. For example, adenosine and deoxyguanoine eluted at 3.6 min at pH = 3 but appeared at 4.8 and 3.7 min, respectively, at pH = 8; the opposite behavior was observed for their monophosphates which eluted together at 1.4 min at pH = 8 but at 1.4 and 2.4 min at pH = 3 (ESM Table S2). In addition, the application of predicted LC retention time to support metabolite identification was evaluated for the AMML MS/MS database and was reported elsewhere [25].

The generation of composite MS/MS spectral library enabled us to perform fragment data analysis. Collision induced fragmentation of even electron ions commonly involves the loss of neutral fragments [19, 26]. Table 1 illustrates the 15 most common neutral losses (frequency) in positive and negative modes from precursors and fragments; water is the most frequent loss for positive and negative modes. Interestingly, 9 out 15 losses are common in both modes. 176 and 93 analytes showed direct H_2O loss from the precursor in positive and negative mode respectively, while 63 and 25 showed direct NH_3 loss from the precursor in positive and negative mode, respectively. Therefore, in-source fragmentation is a critical factor to consider for identification of the true precursor ion mass. The annotation of MS1 adducts is therefore essential to unambiguously annotate the right $[M+H]^+$ precursor [27]. The generation of radical cation fragments has been reported [26] but is not always taken into account when using in silico fragmentation. Table 2 illustrates the most common radical losses from the precursor and the fragments observed with the AMML library. Also, 9 radical losses were in common in positive and negative modes. The frequency is certainly lower compared to that of neutral losses but it is still significant and should be further investigated as they could be class-dependent and generate informative fragments.

The importance of MS/MS libraries for qualitative analysis is well-known, but recent developments suggest to replace spectral libraries by molecular structure databases [28]. Less well appreciated is their value for constructing targeted quantitation methods. Broadly, there are two types of targeted method: those, such as SRM, where the acquisition method targets selected compounds, and others where a more generic acquisition method is used and the data processing is targeted, for example, SWATH. In both cases precursor and product ion *m/z* values and expected retention times are required for the compounds of interest. MRM has great sensitivity but the number of compounds that can be monitored is limited and monitoring additional compounds requires developing a new method and re-acquiring the data, if the sample is still available. Because SWATH systematically collects all fragment ion data for all ions created from a sample, the number of target compounds is large (essentially unlimited) and more can be added simply by reprocessing the acquired data later.

Analysis of pooled human plasma samples by LC-SWATH/MS and metabolite identification

Several studies have been performed to characterize serum or plasma metabolites. Simon-Manso et al. [29] performed a qualitative metabolic profiling of a standard reference material for human plasma (SRM 1950) using GC-MS, LC-MS, and NMR. A total of 322 compounds were identified using four different LC-MS/MS platforms in positive ion mode. It includes 12 out of 20 amino acids; 84 carboxylic acids including 14 heavy fatty acids, 10 phospholipids, and 25 hormones and steroids; several dipeptides and tripeptides; and a variety of other small molecules. Most MS/MS were recorded in data-dependent mode, using survey scans from 50 to 1750 Da followed by MS/MS scans of the four most intense peaks, with a tolerance of 1.3 *m/z*. Boudah et al. [30] evaluated the relevance and versatility of liquid chromatography coupled to high-resolution mass spectrometry (LC-HRMS) for performing a qualitative and

Table 1 Frequency of neutral losses in positive and negative modes from precursor and fragments; fragments in green are common to both modes

POS Total fragments	7017	NEG Total fragments	2748
Neutral Loss	**Frequency**	**Neutral Loss**	**Frequency**
H2O	608	H2O	400
CH4	362	CH2O	225
CO	341	CO2	225
C2H2	323	CO	113
CH2O2	291	C2H4O2	99
C2H4	276	C2H2O	97
NH3	208	NH3	79
CH2O	198	C6H10O5	50
C3H4	142	C2H2	49
C2H2O	141	CH2O2	45
C3H6	135	CH4	38
C2H6	125	CH4O	34
CH4O	110	C6H12O6	34
C3H2	110	HPO3	32
C2H4O2	81	C3H4O2	27
LOSS H2O	176 analytes		93 analytes
LOSS NH3	63 analytes		25 analytes

comprehensive study of the human serum metabolome considering 657 representative metabolite standards. The LC-MS analyses were based on a reversed phase (RP), hydrophilic interaction chromatography (HILIC) and a pentafluorophenylpropyl (PFPP) stationary phase, with detection in both positive and negative electrospray modes. As a result, they reported the formal or putative identification of 266 metabolites.

In the present work, a pooled plasma sample was analyzed with the use of LC-MS/SWATH acquisition with large Q1 windows and processed with the MasterView software using the AMML library as the reference.

The MasterView software uses a list of reference masses to generate chromatograms and finds peaks in the chromatograms and, if any match the expected retention time, the peak area is recorded. If data acquisition included MS/MS, a spectrum obtained at the matching retention time is compared to the library spectrum and the score is used to confirm or reject the compound match. For SWATH data, the chromatograms are extracted from the window containing the expected precursor m/z and the MS/MS spectrum is obtained by background subtraction around the measured retention time in the specific SWATH window. Because of the stringent library curation, fragment masses are known accurately so narrow tolerances are used for spectral library matching. Furthermore, knowledge of the expected retention time allows the time range for chromatogram extraction to be constrained which

Table 2 Common radical losses in positive and negative modes for a frequency > 5. In green common radical losses in both modes

POS Elemental Formula	Accurate Mass	Frequency	NEG Elemental Formula	Accurate Mass	Frequency
CHO	29.0027	46	CHO	29.0027	16
C2H3	27.0235	39	C2H3	27.0235	12
CH3O	31.0184	36	CH3O	31.0184	30
C2H	25.0078	33	C2H	25.0078	7
CHO2	44.9977	20	CHO2	44.9977	15
CH2N	28.0187	18	CH2N	28.0187	6
CH3O2	47.0133	13	CH3O2	47.0133	5
C2H3O	43.0184	11	C2H3O	43.0184	24
C2H3O2	59.0133	10	C2H3O2	59.0133	6
C2H5	29.0391	21	CH4N	30.0344	13
C2H7	31.0548	11	C2HO2	56.9977	5
CH5O	33.0340	15	C2H5O2	61.0290	22
C3H5	41.0391	10	C2H4NO2	74.0242	6
C3H7	43.0548	13	C3H7O3	91.0395	5
C3H3O2	71.0133	14			
I	126.9045	12			

reduces the time needed for processing and improves the peak picking selectivity. Using this approach, a total of 92 metabolites from the library were detected by LC-MS (Fig. 4 and ESM Table S6) in either positive mode with pH = 3 (39 metabolites) or negative mode at pH = 8 (72 metabolites). The use of MS response factors to confirm identification was found also to be important. Out of the 92 metabolites detected, 81 of these metabolites were reported in the HMDB database to have been detected in human plasma. As the focus of the present work was on the investigation, construction, and use of MS/MS libraries for SWATH acquisition and was not a systematic investigation of human plasma metabolites, comparison with published work is beyond the scope. In some cases, the product ion spectra were not very informative calling for alternative fragmentation processes to enhance identification. One of the key features of data-independent acquisition, such as SWATH, is that data can be reprocessed at any time, in particular when expanding the MS/MS library. The library (discrete or composite) is also useful to perform quantitative analysis at the MS/MS level (HR-SRM).

Conclusions

Regardless of the intended use of an MS library, the quality of the MS/MS spectra is important, in particular, that the spectra are pure, i.e., that there are no fragments from other compounds, since these may lead to false positive identifications or incorrect targeting. Even when narrow precursor windows are used for acquisition, impurities or background masses may be co-selected and fragmented generating spurious peaks. Information content is another consideration—a larger number of peaks will generally result in a more specific search. In

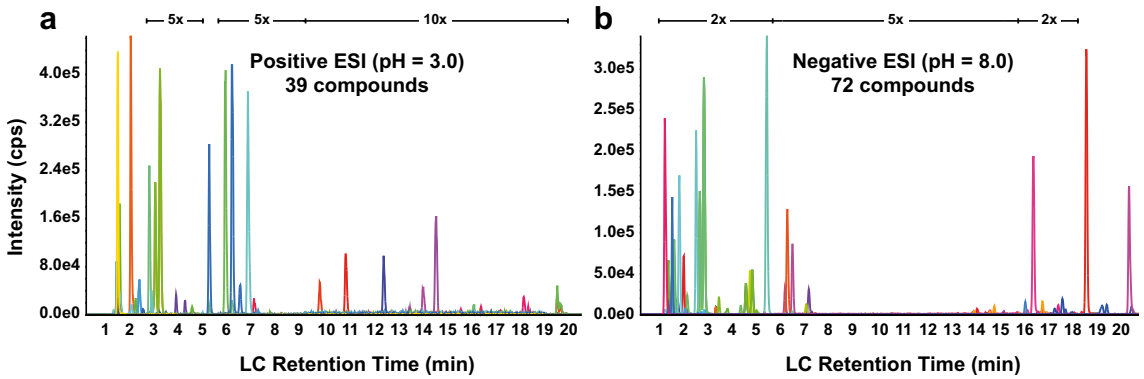

Fig. 4 Extracted ion current of metabolites identified in a plasma sample. (A) pH = 3., (B) pH = 8 (ESM Table S6)

the workflow described here, MS/MS spectra were acquired in flow injection mode at 16 different collision energy values which allows composite spectra appropriate for the acquisition to be constructed; this also provides information for optimizing SRM experiments. In addition, the confidence in weak fragments is increased if they are observed in several spectra obtained at adjacent CE values. A set of tools were developed to automatically check fragment ions based on the structure of the standard, reject those that cannot be explained, and reject spectra where less than 90% of the summed ion intensity can be annotated. Confidently assigning elemental compositions also allows the isotope pattern of the fragments to be assigned. This is valuable in SWATH acquisition since the wide windows allow entire isotope clusters to be selected for fragmentation, generating product ions with isotope patterns that can be compared to the theoretical values. Including isotopic information enhances the score and may help differentiate isobaric interferences and improve halogen recognition. As mentioned before, retention time information is useful for supporting identification and determining isomers, but essential for targeted assays such as SRM and SWATH. In the approach described here, a single column is used but the pH is changed for positive (pH = 3) and negative (pH = 8) mode acquisition. While simple, this has valuable benefits, for example, some isomers are only separated under one set of conditions and the different chromatographic behavior can provide additional confirmation of compounds that are detected in both modes.

The AMML library includes 532 metabolites, 489 metabolites provided a signal in positive mode $[M+H]^+$ and 483 in negative mode $[M-H]^-$. MS/MS spectra were obtained for 399 compounds in positive mode and for 462 in negative mode; 329 metabolites generated usable spectra in both modes. LC retention times were measured for 339 metabolites in positive mode at pH = 3 and 410 metabolites in negative mode at pH = 8. Using our library, the analysis by LC-SWATH/MS allowed the identification of 39 (positive mode) and 72 (negative mode) metabolites in a plasma sample (total 92). As the library is expanded with additional compounds, the SWATH experimental data can be reprocessed without the need of reacquisition.

Acknowledgements GH would like to thank SystemsX and the Swiss National Sciences Foundation for the financial support: Projects 51RTP0_151032 and 206021_170779.

Compliance with ethical standards Plasma samples were obtained from healthy voluntary who have consented that their donation or some of its components being used for medical research after final anonymization or in a coded form. The plasma samples were provided by the Centre de Transfusion Sanguine, University Hospital Geneva, Geneva, Switzerland. The Human Research Act (HRA) does not apply for the anonymized pooled plasma samples analyzed in the present work (Art. 2 para. 2 let. b and c).

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu YF, et al. HMDB 3.0—the Human Metabolome Database in 2013. Nucleic Acids Res. 2013;41(D1):D801–D7. https://doi.org/10.1093/nar/gks1065.
2. Rainville PD, Theodoridis G, Plumb RS, Wilson ID. Advances in liquid chromatography coupled to mass spectrometry for metabolic phenotyping. TRAC-Trend Anal Chem. 2014;61:181–91. https://doi.org/10.1016/j.trac.2014.06.005.
3. Brenton AG, Godfrey AR. Accurate mass measurement: terminology and treatment of data. J Am Soc Mass Spectrom. 2010;21(11):1821–35. https://doi.org/10.1016/j.jasms.2010.06.006.
4. Kind T, Tsugawa H, Cajka T, Ma Y, Lai Z, Mehta SS, et al. Identification of small molecules using accurate mass MS/MS search. Mass Spectrom Rev. 2017; https://doi.org/10.1002/mas.21535.
5. Ruttkies C, Schymanski EL, Wolf S, Hollender J, Neumann S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J Cheminformatics. 2016;8. https://doi.org/10.1186/s13321-016-0115-9.
6. Ojanpera I, Kolmonen M, Pelander A. Current use of high-resolution mass spectrometry in drug screening relevant to clinical and forensic toxicology and doping control. Anal Bioanal Chem. 2012;403(5):1203–20. https://doi.org/10.1007/s00216-012-5726-z.
7. Stravs MA, Schymanski EL, Singer HP, Hollender J. Automatic recalibration and processing of tandem mass spectra using formula annotation. J Mass Spectrom. 2013;48(1):89–99. https://doi.org/10.1002/jms.3131.
8. Gillet LC, Navarro P, Tate S, Rost H, Selevsek N, Reiter L, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics. 2012;11(6):O111 016717. https://doi.org/10.1074/mcp.O111.016717.
9. Hopfgartner G, Tonoli D, Varesio E. High-resolution mass spectrometry for integrated qualitative and quantitative analysis of pharmaceuticals in biological matrices. Anal Bioanal Chem. 2012;402(8):2587–96. https://doi.org/10.1007/s00216-011-5641-8.
10. Siegel D, Meinema AC, Permentier H, Hopfgartner G, Bischoff R. Integrated quantification and identification of aldehydes and ketones in biological samples. Anal Chem. 2014;86(10):5089–100. https://doi.org/10.1021/ac500810r.
11. Scheidweiler KB, Jarvis MJ, Huestis MA. Nontargeted SWATH acquisition for identifying 47 synthetic cannabinoid metabolites in human urine by liquid chromatography-high-resolution tandem mass spectrometry. Anal Bioanal Chem. 2015;407(3):883–97. https://doi.org/10.1007/s00216-014-8118-8.
12. Roemmelt AT, Steuer AE, Kraemer T. Liquid chromatography, in combination with a quadrupole time-of-flight instrument, with sequential window acquisition of all theoretical fragment-ion spectra acquisition: validated quantification of 39 antidepressants in whole blood as part of a simultaneous screening and quantification procedure. Anal Chem. 2015;87(18):9294–301. https://doi.org/10.1021/acs.analchem.5b02031.

13. Go EP. Database resources in metabolomics: an overview. J Neuroimmune Pharm. 2010;5(1):18–30. https://doi.org/10.1007/s11481-009-9157-3.
14. Scheubert K, Hufsky F, Bocker S. Computational mass spectrometry for small molecules. J Cheminformatics. 2013;5(1):12. https://doi.org/10.1186/1758-2946-5-12.
15. Ausloos P, Clifton CL, Lias SG, Mikaya AI, Stein SE, Tchekhovskoi DV, et al. The critical evaluation of a comprehensive mass spectral library. J Am Soc Mass Spectrom. 1999;10(4):287–99.
16. Oberacher H, Pitterl F, Siapi E, Steele BR, Letzel T, Grosse S, et al. On the inter-instrument and the inter-laboratory transferability of a tandem mass spectral reference library. 3. Focus on ion trap and upfront CID. J Mass Spectrom. 2012;47(2):263–70. https://doi.org/10.1002/jms.2961.
17. Lei Z, Jing L, Qiu F, Zhang H, Huhman D, Zhou Z, et al. Construction of an ultrahigh pressure liquid chromatography-tandem mass spectral library of plant natural products and comparative spectral analyses. Anal Chem. 2015;87(14):7373–81. https://doi.org/10.1021/acs.analchem.5b01559.
18. Zhang Y, Bilbao A, Bruderer T, Luban J, Strambio-De-Castillia C, Lisacek F, et al. The use of variable Q1 isolation windows improves selectivity in LC-SWATH-MS acquisition. J Proteome Res. 2015;14(10):4359–71. https://doi.org/10.1021/acs.jproteome.5b00543.
19. Duchoslav E, Burton L, Varesio E, Bonner R, Hopfgartner G, editors. Towards comprehensive, reliable and accurate mass data repositories 2014; Proceedings of the 69th ASMS Conference on Mass Spectrometry and Allied Topics, Baltimore, MD, 15–19 June, 2014.
20. Bazso FL, Ozohanics O, Schlosser G, Ludanyi K, Vekey K, Drahos L. Quantitative comparison of tandem mass spectra obtained on various instruments. J Am Soc Mass Spectrom. 2016;27(8):1357–65. https://doi.org/10.1007/s13361-016-1408-y.
21. Broecker S, Herre S, Wust B, Zweigenbaum J, Pragst F. Development and practical application of a library of CID accurate mass spectra of more than 2,500 toxic compounds for systematic toxicological analysis by LC-QTOF-MS with data-dependent acquisition. Anal Bioanal Chem. 2011;400(1):101–17. https://doi.org/10.1007/s00216-010-4450-9.
22. Yang X, Neta P, Stein SE. Quality control for building libraries from electrospray ionization tandem mass spectra. Anal Chem. 2014;86(13):6393–400. https://doi.org/10.1021/ac500711m.
23. Stein S. Mass spectral reference libraries: an ever-expanding resource for chemical identification. Anal Chem. 2012;84(17):7274–82. https://doi.org/10.1021/ac301205z.
24. Vinaixa M, Schymanski EL, Neumann S, Navarro M, Salek RM, Yanes O. Mass spectral databases for LC/MS- and GC/MS-based metabolomics: state of the field and future prospects. TRAC-Trend Anal Chem. 2016;78(Supplement C):23–35. https://doi.org/10.1016/j.trac.2015.09.005.
25. Bruderer T, Varesio E, Hopfgartner G. The use of LC predicted retention times to extend metabolites identification with SWATH data acquisition. J Chromatogr B. 2017; https://doi.org/10.1016/j.jchromb.2017.07.016.
26. Niessen WMA, RA CC. Interpretation of MS-MS mass spectra of drugs and pesticides, Wiley Series on Mass Spectrometry. Hoboken, New Jersey: Wiley; 2017.
27. Bonner R, Hopfgartner G. SWATH acquisition mode for drug metabolism and metabolomics investigations. Bioanalysis. 2016;8(16):1735–50. https://doi.org/10.4155/bio-2016-0141.
28. Bocker S. Searching molecular structure databases using tandem MS data: are we there yet? Curr Opin Chem Biol. 2017;36:1–6. https://doi.org/10.1016/j.cbpa.2016.12.010.
29. Simon-Manso Y, Lowenthal MS, Kilpatrick LE, Sampson ML, Telu KH, Rudnick PA, et al. Metabolite profiling of a NIST standard reference material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. Anal Chem. 2013;85(24):11725–31. https://doi.org/10.1021/ac402503m.
30. Boudah S, Olivier MF, Aros-Calt S, Oliveira L, Fenaille F, Tabet JC, et al. Annotation of the human serum metabolome by coupling three liquid chromatography methods to high-resolution mass spectrometry. J Chromatogr B. 2014; https://doi.org/10.1016/j.jchromb.2014.04.025.

Tobias Bruderer obtained his PhD under the supervision of Prof. Gérard Hopfgartner at the University of Geneva. He is an analytical chemist with focus on high-resolution mass spectrometry. His current research aims to develop molecular breath analysis for asthma preschool diagnosis and early infection screening in cystic fibrosis patients in the Moeller group at the Children's University Hospital Zurich and the Zenobi group at ETH Zurich.

Emmanuel Varesio is a senior lecturer at the School of Pharmaceutical Sciences and Executive Director of the Mass Spectrometry core facility of the Faculty of Sciences at University of Geneva. Besides his interest in new MS-based techniques and strategies like data-independent acquisition, he likes to teach and implement such new tools at the core facility to support scientists in their research.

Anita O. Hidasi is a postdoctoral associate in the Life Sciences Mass Spectrometry group in the Department of Inorganic and Analytical Chemistry at the University of Geneva. She is a biochemist with experience in toxicology, molecular biology, and bioanalytics including LC-MS-based omics approaches, furthermore working with the animal model zebrafish (*Danio rerio*) embryo. She is interested in exploring biomarkers of chemical exposures and diseases. Her current research focuses on data-independent acquisition and electron-based dissociation in metabolomics.

Eva Duchoslav works as a senior staff research scientist at SCIEX, focusing on new mass spectrometry data collection and processing concepts, enabling automation of small molecule qualitative applications, such as identification of endogenous and exogenous metabolites and characterization and profiling of lipids.

Lyle Burton is a senior research scientist at Sciex and is based near Toronto, Canada. He is interested in many aspects of LC-MS data processing to extract useful information from raw data with a particular focus on the areas of metabolomics, biologics, and quantitative data analysis.

Ron Bonner is the Principal of Ron Bonner Consulting which he started following a 40-year career in mass spectrometry research. He continues to explore mass spectrometry applications such as metabolomics and proteomics and is particularly interested in the analysis and interpretation of complex data and in compound identification.

Gérard Hopfgartner is Professor for Analytical Chemistry and Mass Spectrometry at the Department of Analytical and Inorganic Chemistry of the University of Geneva. His research interests focus on the development and application of novel hyphenated mass spectrometry in the field of life sciences.