Data Dictionary Fiehn laboratory NIH West Coast Metabolomics Center 07 2022

Analysis of biogenic amines and other small molecule metabolites by HILIC-ESI QTOF MS/MS

Glossary

HILIC hydrophilic interaction liquid chromatography, a variant of normal-phase chromatography UHPLC ultra high-pressure liquid chromatography

ESI electrospray ionization. The method uses both negative ESI and positive ESI for negatively charged and positively charged molecules.

QTOF quadrupole time-of-flight mass spectrometer

MS/MS tandem mass spectrometry. After soft ionization by electrospray, the precursor (intact) charged molecules are fragmented by collision with gas atoms, usually Helium. Fragments are then analyzed by time-of-flight mass spectrometry to obtain accurate mass information at high resolution.

Resolving power also called resolution. In MS, resolving power defines the ability to distinguish co-eluting masses that have the same nominal mass, but different accurate mass.

MTBE methyl-tertiary butyl ether

MeOH methanol

BEH amide bridged ethylene hybrid amide column

QC quality control

IS, istd internal standard

v/v volumetric ratio

InChI International Chemical Identifier key. Denotes the exact stereochemical and atomic description of chemicals and used as universal identifier in chemical databases.

rt retention time (minutes)

mz also m/z, or mass-to-charge ratio. In metabolomics, ions are almost exclusively detected as singly charged species.

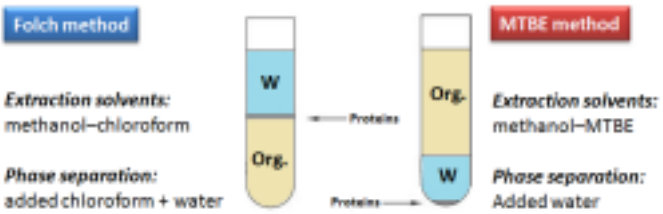
rt_mz identifier for individual metabolites in the MassHunter Quantification method consisting of the retention time and the m/z value of specific compounds.

IUPAC International Union of Pure and Applied Chemists

NIST National Institute of Standards and Technology

PCA Principal Component Analysis

Sample extraction



Blood plasma or serum is best extracted following the protocols first published in Matyash V. et al., *J. Lip. Res.* 49 (2008) 1137–1146. One of the major differences to the earlier protocols by Folch or Bligh Dyer is that in the Matyash protocol, lipid extracts (labeled 'org' in the figure on the right) are separated from proteins and from polar hydrophilic small molecules (in the methanol/water phase, labeled 'W' in the figure on the right) in a way that the lipids are found in the top layer of liquid-liquid separations, rather than in the bottom layer. Decanting the top layer therefore ensures that extracts are not contaminated by proteins or polar compounds. The top layer is used for lipidomics while the bottom layer (methanol/water phase) is very suitable for HILIC-MS investigations. In some cases, other extraction methods may be used. For example, a mixture of acetonitrile/water/isopropanol (2:2:3) can be used to precipitate proteins and extract metabolites. In such cases, complex lipids would be found in HILIC-QTOF MS chromatograms in addition to small molecule metabolites such as betaine, choline and TMAO. Details of extraction methods for individual studies may vary and are available by inquiry.

Data acquisition

HILIC samples are most frequently injected using an Agilent 1290 UHPLC/Sciex TripleTOF 6600 mass spectrometer. Occasionally, HILIC samples may be injected using an Agilent 1290 UHPLC/Agilent 6550/6530 mass spectrometer or a Thermo Vanquish UHPLC/Thermo QExactive mass spectrometer. LC parameters are displayed in the table below; MS parameters vary by instrument and are available by request. The analytical UHPLC column is protected by a short guard column. This chromatography method yields excellent retention and separation of various small molecule metabolites with narrow peak widths of 2-5 s and very good within-series retention time reproducibility typically less than 1 s absolute deviation of retention times.

Column	Waters Acquity Premier UPLC BEH Amide Column (1.7 µm, 2.1 mm x 50 mm)
Mobile phase A	Ultrapure water with 10 mM ammonium formate + 0.125% formic acid, pH 3
Mobile phase B	95:5 v/v acetonitrile:ultrapure water w/ 10 mM ammonium formate + 0.125% formic acid, pH 3
Column temperature	45°C
Gradient	0 min, 100% B; 0.5 min, 100% B; 1.95 min, 70% B; 2.55 min, 30% B; 3.15 min, 100% B; 3.8 min, 100% B
Flow rate	0.8 mL/min
Injection volume	1-5 µL

Data processing

Chromatograms first undergo a quality control check in which internal standards are examined for consistency of peak height and retention time. Raw data files are then processed using an updated version of MS-DIAL software which identifies and aligns peaks and then annotates peaks using both an in-house mzRT library and MS/MS spectral matching with NIST/MoNA libraries. All MS/MS annotations are then manually curated by a lab member to ensure that only high-quality compound identifications are included in the final report.

Data reporting

Data are reported including metadata. The 'identifier' column denotes the unique identifier for the technology platform, given as rt_mz. It is given for both identified and unidentified metabolites in the same manner. The 'metabolite name' column denotes the name of the metabolite, if the peak has been identified. A chemical name is not a unique identifier. We use names recognized by biologists instead of IUPAC nomenclature. The 'InChI key' column gives the unique chemical identifier defined by the IUPAC and NIST consortia. The 'average rt' and 'average mz' columns detail the retention time and mz values that were detected in a specific data processing sequence of chromatograms. These values may be slightly different from the values given in the 'identifier column'. The 'File ID' row denotes the name of the raw data file.

The actual data are given as peak heights for the quantification ion (mz value) at the specific retention time (rt value). We provide peak heights instead of peak areas because peak heights are more precise for low abundant metabolites than peak areas, due to the larger influence of baseline determinations on areas compared to peak heights. Also, overlapping (co-eluting) ions or peaks are harder to deconvolute in terms of precise determinations of peak areas than peak heights.

Raw data peak heights need to be normalized to reduce the impact of within-series drifts of instrument sensitivity, caused by machine maintenance, aging and tuning parameters. There are many different types of normalizations in the scientific literature. We usually provide a variant of a 'vector normalization' in which we calculate the sum of all peak heights for all identified internal standards for each sample, denoted iTIC. Peak heights for all samples are then normalized to the total average sum of the internal standards (iTIC).

Equation used for normalizations of metabolite *i* of sample *j*:

$$\text{metabolite}_{ij,\text{normalized}} = (\text{metabolite}_{ij,\text{raw}} / \text{iTIC}_{j}) * \text{iTIC}_{\text{average}}$$

In some cases, such as when variable amounts of samples were extracted, the sum of all annotated metabolites (mTIC) may be used in place of the sum of the internal standards (iTIC). For studies exceeding ~330 samples, SERRF normalization is used in place of iTIC normalization (for details, see Fan S, Kind T, Cajka T, et al. Systematic error removal using random forest for normalizing large-scale untargeted lipidomics data. Anal Chem. 2019;91(5):3590-3596). Reported peak heights are 'relative semi-quantifications', meaning they are normalized peak heights. Because the average iTIC/mTIC will be different between series of analyses that are weeks or months apart (due to differences in machine sensitivity, tuning, maintenance status and other parameters), additional normalizations need to be performed if comparisons between disparate data sets are needed. In such cases, please contact our staff for more information regarding appropriate normalizations.