# Metabolomics Case Study

Jonathan E. Hunter - January 2025

## 1. Case Study Objectives

* End to end experimental approach for secondary metabolites.
* 20 plant extracts by MS based untargeted metabolomics.
* Elucidate structure of as many small molecules as possible.
* **Focus on QC and QA within Experimental Design.**
  + **Sample prep workflow, instrumentation selection and QC measures.**
* Provide a detailed description of the data preprocessing workflow.
  + Detailed description, **describe the objective of each step, outline how specific features will be selected for downstream analysis.**
* signals to be extracted, tools and methods to ID compounds.
  + Describe the tools, databases and methods you would use to ID structures.
  + **Outline additional cheminformatics/computational approaches you might use to improve structural elucidation further.**

I have provided a single example of a sensible approach based upon my experience, but every step is of course subject to alternatives and further method development and validation. I’ve tried to illustrate the most scientifically robust, but not necessarily the most expedient or efficient approach.

Solvents are LC-MS grade or better. Storage of sample biomass at –80°C, extracts for short term storage at -20°C.

## 2. Sample Preparation

10g of plant material is roughly cut into smaller pieces using stainless steel scissors, (washed with 70% MeOH(aq) between samples) and collected into labelled cryomill grinding jars. Fresh, washed as above, grinder beads are added to the jar and homogenization conducted on the cryomill.

A negative procedural surrogate blank matrix, for example spinach leaves, is prepared alongside the test samples in sufficient quantity.

Sample material is Lyophilized (freeze-dried) to remove excess moisture whilst preserving analyte content.

100 mg of each test sample, plus a sufficient count of blank matrix aliquots are weighed into 8mL glass screw top sample tubes. A representative mixture of stable isotope labelled metabolite internal markers, for example from Cambridge Isotopes Laboratories, is spiked onto the samples, at a concentration backcalculated from an equivalent final concentration of 100 pmol on column. A qualitative range, for example 10, 100, 1000 pmol o.c. could be prepared for semi-quantitative estimates of analyte abundance.

Butylated-hydroxytoluene is added for antioxidant activity. 2 matrix blanks (spinach) are extracted excluding internal standards (Double Blank), plus 2 matrix blanks including internal standards (Single Blank/Zero).

70% MeOH(aq) 2mL is added to all samples and blanks. Ultrasonic Assisted Extraction is conducted at 40°C for 15 minutes. Sample extracts are centrifuged at RCF 1000xG for 10 minutes at 5°C and the supernatant aspirated off with glass pasteur pipettes into 2mL LC sample vials. Sample extract is evaporated under flowing N2 or by centrifugal evaporator. Samples are topped with Argon inert atmosphere prior to storage at -20°C.

## 3. Analytical Instrumentation

A leading option for analytical instrumentation to analyse metabolomics samples are the Thermo Scientific Orbitrap series (eg. Q-Exactive, Exploris, Fusion etc.). These provide excellent mass resolution and accuracy at good but not great scan rates. An Orbitrap, for example a Q-Exactive, could be used here as follows:

Samples are reconstituted in 50% MeOH(aq) 500µL prior to analysis. 2x reconstitution solvent blanks (recon blanks) are prepared.

A Thermo Vanquish UHPLC system is used with a 1.7µm C18 core-shell stationary phase, 2.1mm x 15cm at 40°C. 10 µL of sample is injected onto column and separated during a gradient of 0.1% FA in H2O to 0.1% FA in MeOH at 400 µL/min over 20 minutes.

The eluate is analysed by the Thermo Q-Exactive MS in ESI mode. The system is configured to switch rapidly between positive and negative ion modes. MS1 full scan mass spectra (50-1000 m/z) and data-dependent MS2 spectra (20-precursor m/z) are acquired continuously. The base peak in a given MS1 scan is selectively filtered, fragmented by CID and the MS2 spectrum acquired. A dynamic exclusion window of 5 seconds is then applied to prevent reacquisition of the same analyte MS2 and the next most abundant precursor is selected. A stepped Normalised Collision Energy range of 25, 35, 45% is applied to generate consistent and high quality fragmentation.

A typical analytical batch structure would include recon, single and double blank samples at the beginning and end of the batch. Sample data would be reviewed for instrument/recon solvent contamination (recon blank); internal marker contamination (single blank) and procedural (extraction) contamination (double blank). The internal markers would act as an injection marker, indicator of consistent chromatography and MS instrument sensitivity, and finally as a approximate point of comparison for semi-quantification of analytes.

## 4. Data Preprocessing

Raw data files are preprocessed into an open format (eg. mzXML/mzML) using the ProteoWizard MSConvert docker image. For further data processing and interpretation see the associated Jupyter notebook (GitHub).

## 6. Alternative Analytical Options

Lipid Analysis - Bligh/Dyer lipid extraction, more hydrophobic LC conditions, similar MS conditions.

GC-MS/MS Orbitrap – Extraction of apolar compounds, derivatization (eg. silylation) and analysis by GC Orbitrap (eg. Thermo GC Exploris / GC Q-Exactive)

Ion Mobility/Q-ToF – Sample preparation and LC as previous but coupled to a Q-ToF with ion mobility, eg. Waters Vion or G2-Si. Ion mobility provides an additional dimension of chemical separation according to analyte cross sectional area between the Quadrupole mass filter and the ToF mass analyser.

HILIC or Ion Exchange Chromatography – to determine highly polar or charged analytes.

Alternative MS2 Databases – METLIN (mostly paywalled), mzCloud. (GNPS includes many libraries eg. HMDB, MassBank).

## 7. Further Work

De-novo simulation/determination of MS2 spectra/analyte structures.

* Deep learning model based upon the ingestion of a lot of annotated MS2 spectra.
* Or a mechanistic model based upon fragmentation chemistry.

Deep learning based peak picking algorithm

* Rather than the conventional peak picking algorithms that XCMS uses, some instrument vendor software now includes Deep Learning based peak picking algorithms that reduce the number of false positives, split peaks etc. Simplifying and improving the accuracy of the workflow.