

We are learning on
Noongar land



THE UNIVERSITY OF
WESTERN
AUSTRALIA

Collecting, Analysing and Interpreting Big Data in Biology

SCIE4001

Metabolomics

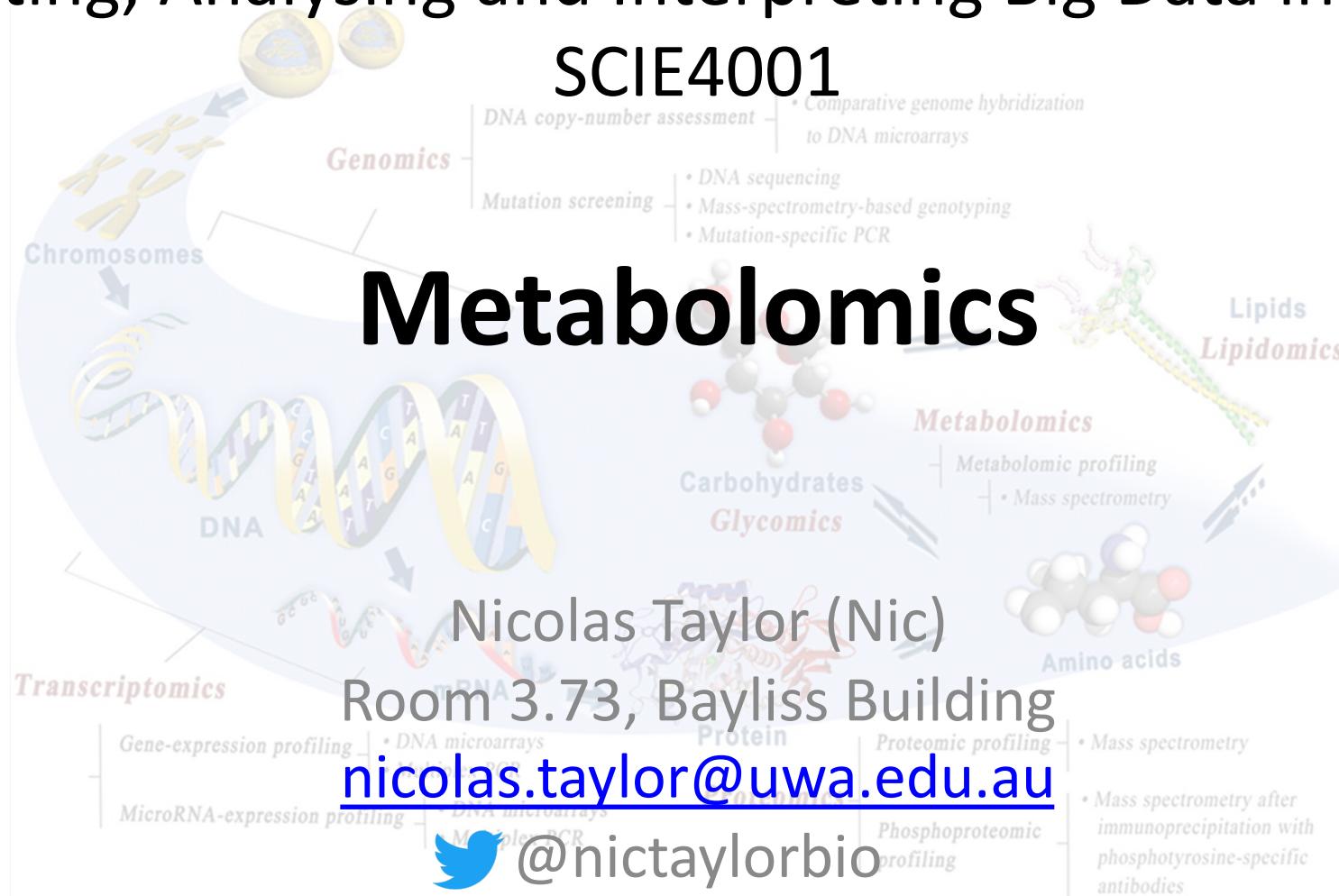
Nicolas Taylor (Nic)

Room 3.73, Bayliss Building

nicolas.taylor@uwa.edu.au



@nictaylorbio



THE UNIVERSITY OF
WESTERN
AUSTRALIA

Metabolomics

1. What is Metabolomics ?
2. What kind of biological questions does it answer ?
3. How does metabolomics fit with other omics approaches in to help us understand biology.



Genome - the complete set of genes in a system

Transcriptome - complete set of mRNAs in a system

Proteome - the complete set of proteins expressed in a system

Metabolome - the complete set of metabolites in a system

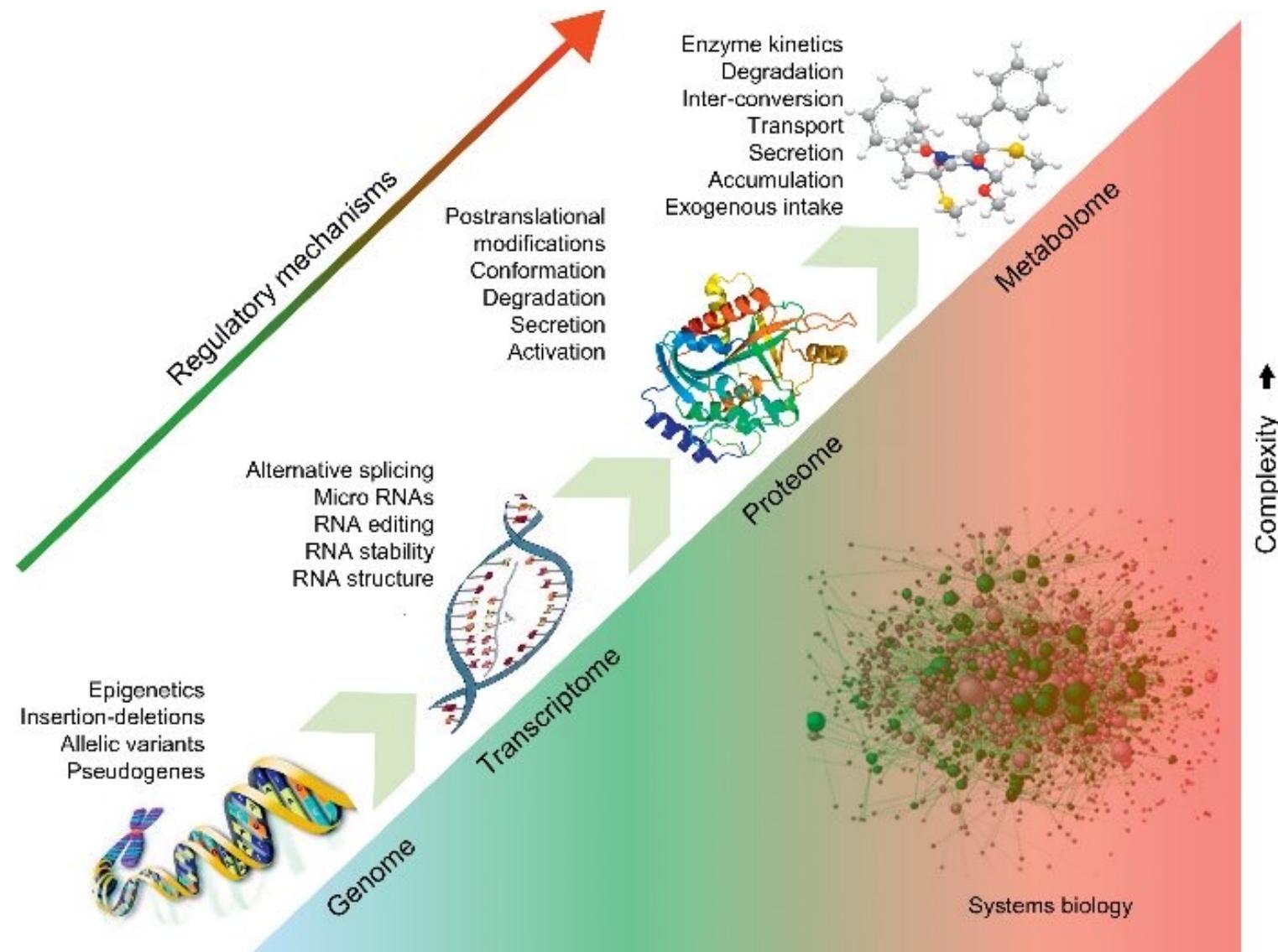
Metabolomics

The study of the ‘metabolites’

While an organism has only one genome,

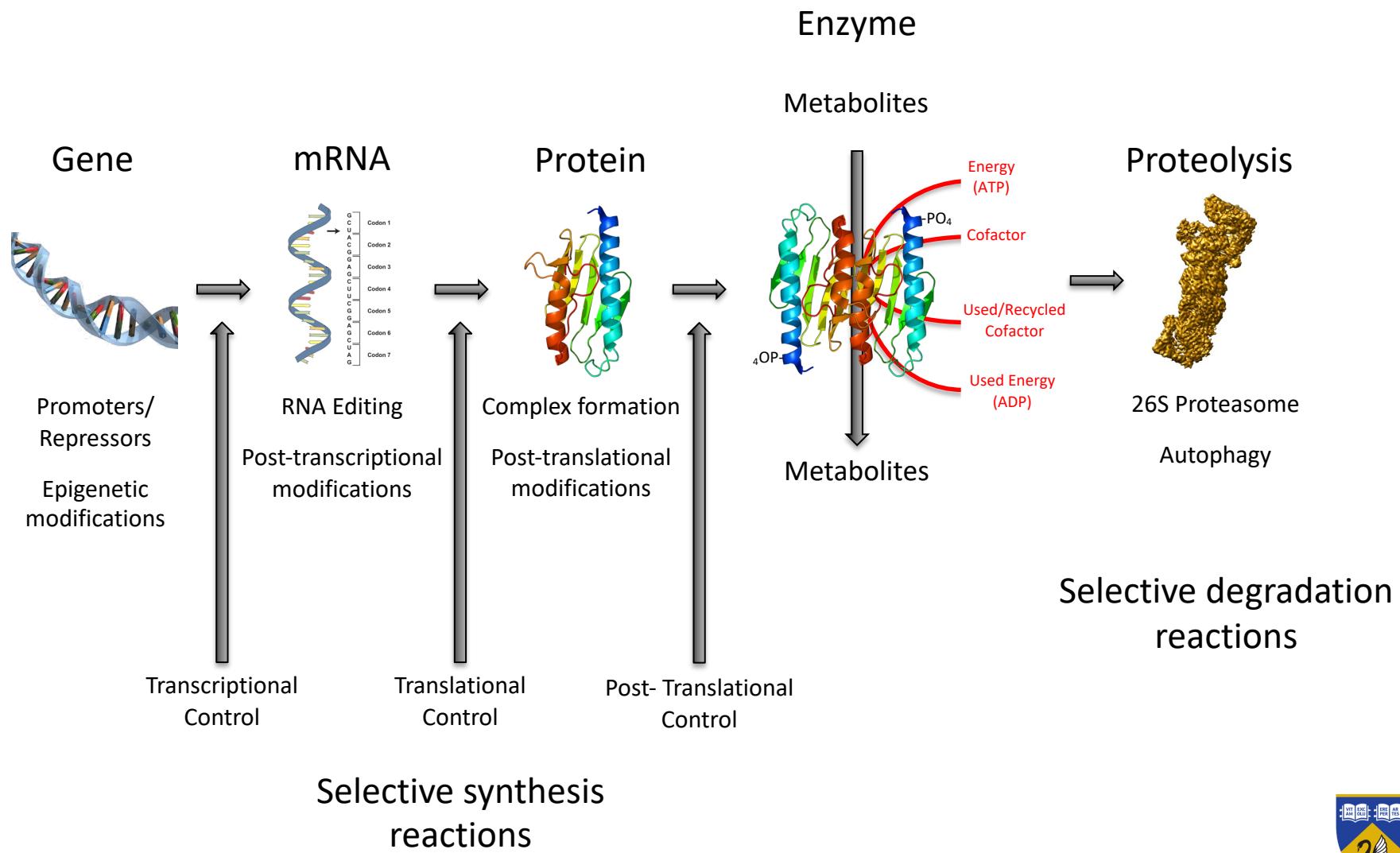
it has many transcriptomes, proteomes and metabolomes



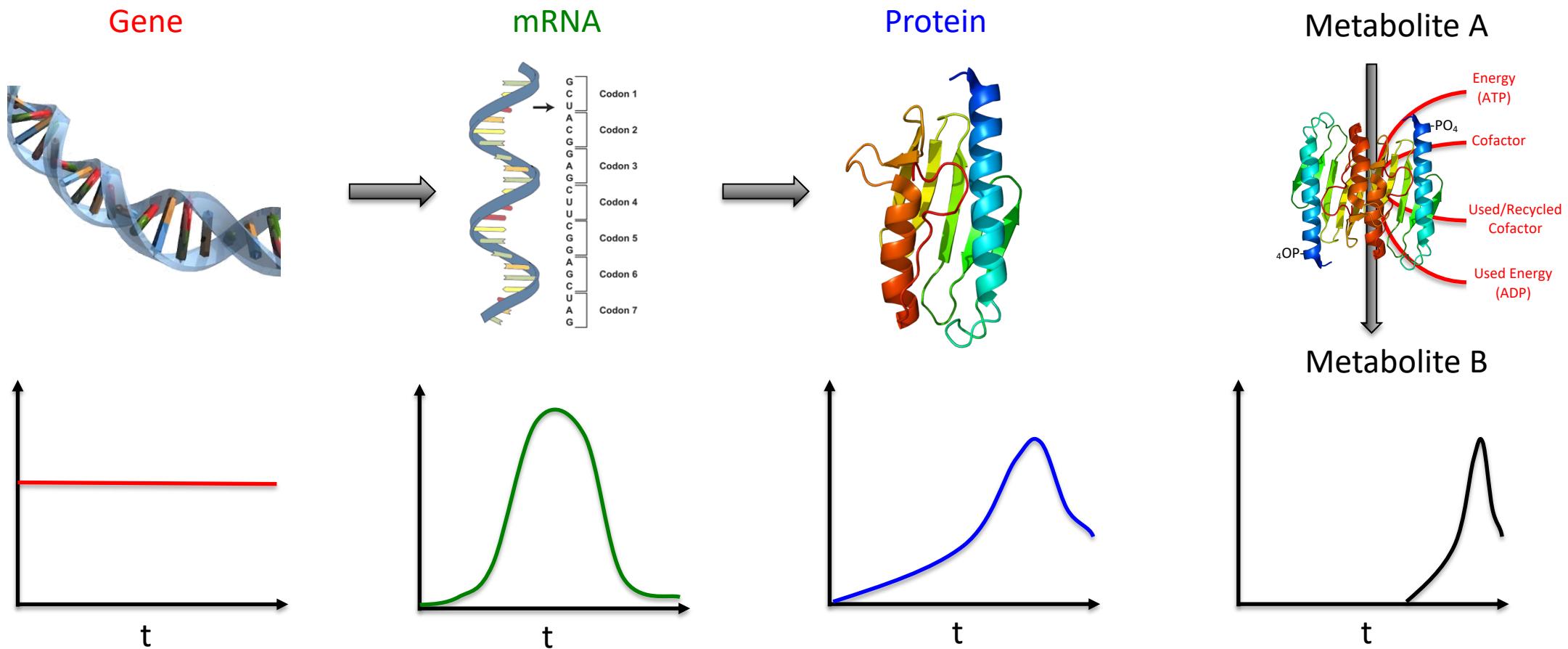


Approach	Component Examined	Approach
Genomics	Genes	NGS
Epigenomics	DNA methylation	Bisulphite Sequencing NGS
Transcriptomics	mRNA	DNA arrays GeneChip RNASeq
Proteomics	Proteins	2D-PAGE MALDI-MS LC-MS SRM
Metabolomics	Metabolites	GC-MS LC-MS SRM NMR

Protein Activity : Energy & Cofactors



Temporal changes in metabolites



When you measure expression is important

What are Metabolites

“Small molecules”

Primary and secondary metabolites have molecular weight

< 1000 Da

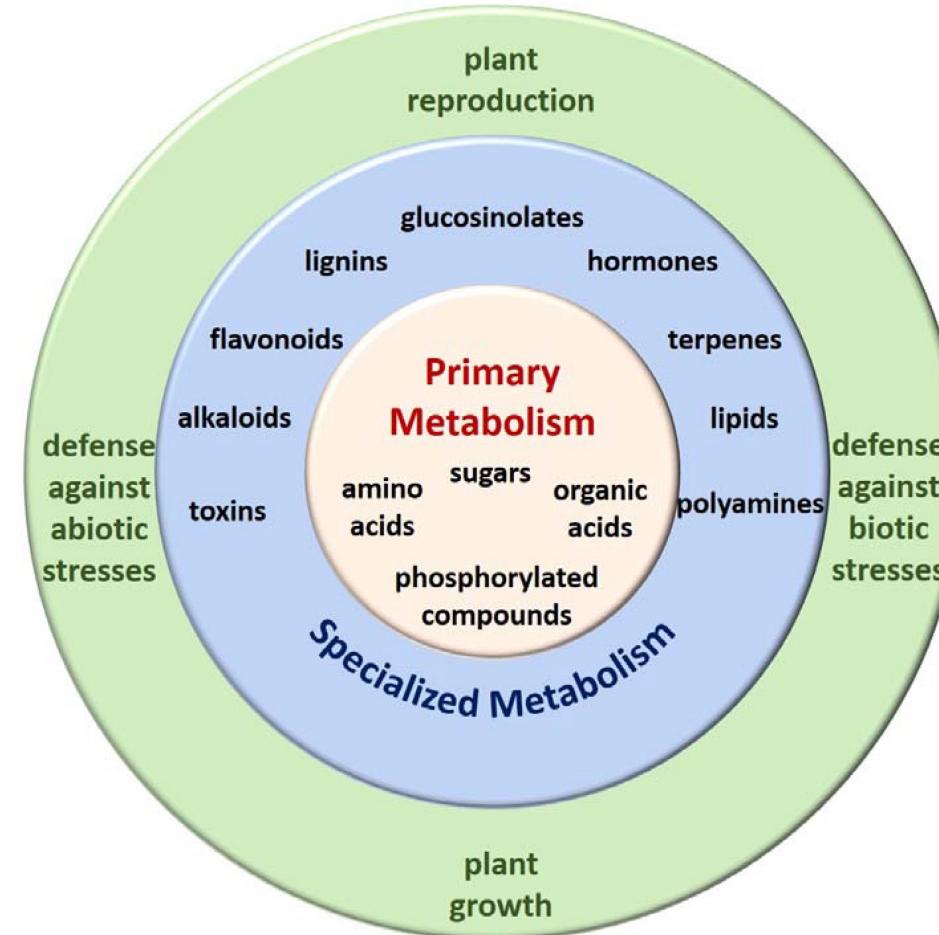
Examples small molecules

Lipids, give cells shape, form, structure

Amino acids

Sugars, source of all cellular energy

Cofactors and signaling molecules



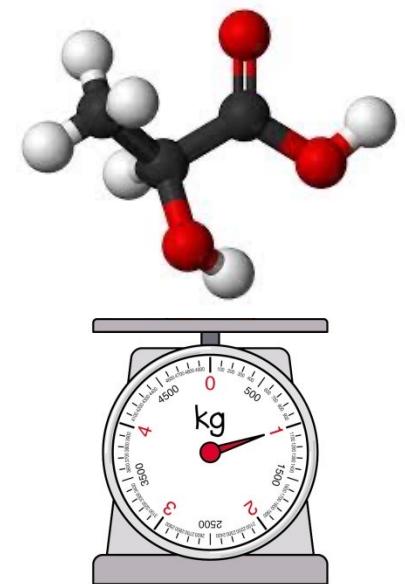
Untargeted Metabolomics

Metabolomics

Measure small molecule metabolites

Quantitative analysis of metabolites in complex specimens

An instantaneous snapshot of biological sample



Untargeted Metabolomics

Simultaneous measurements of thousands of metabolites

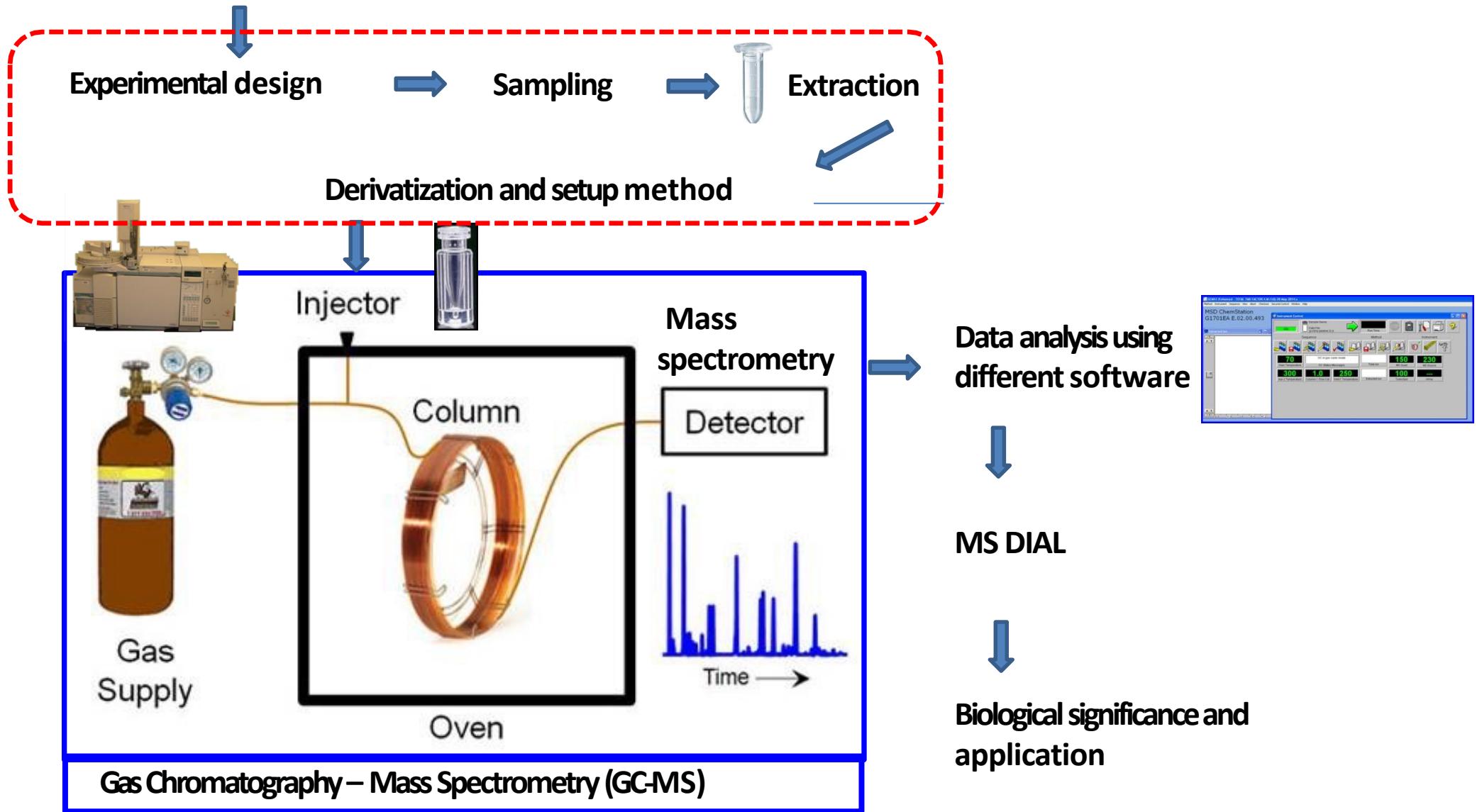
Relies on mass spec technology and databases

Targeted Metabolomics

Simultaneous measurements of thousands of metabolites

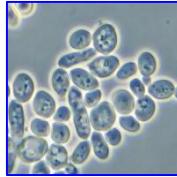
Relies on mass spec technology and a priori knowledge

GC-MS



Experimental Design

Collect samples under identical conditions



Randomize the treatment groups

(Make sure the effects you measure are due to the variable being testing)

Number of replicates... depends on what you want to find

- Large differences = small replication needed
- Small differences = large replication needed

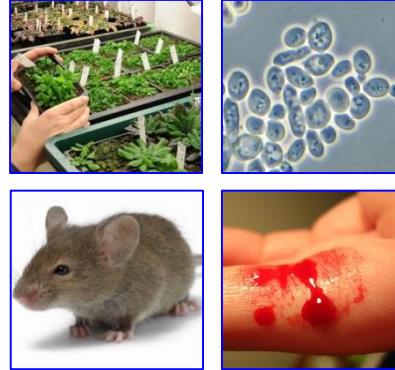
In general, six replicates for each treatment are needed.
(due to high biological variability)

Experimental Design

Uniform sample sizes (e.g. hole punches in leaves)

Be consistent

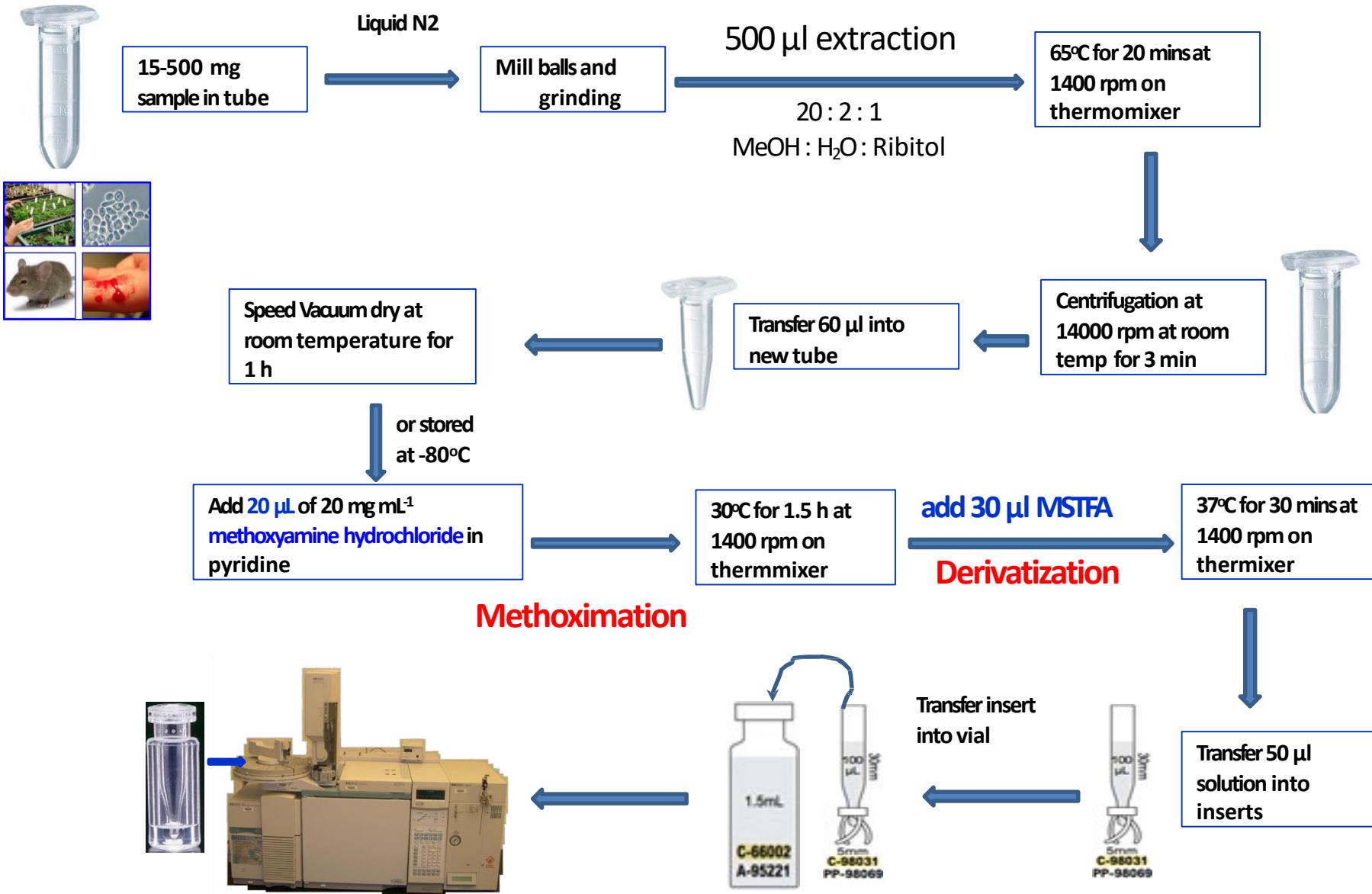
- similar tissue
- time of day



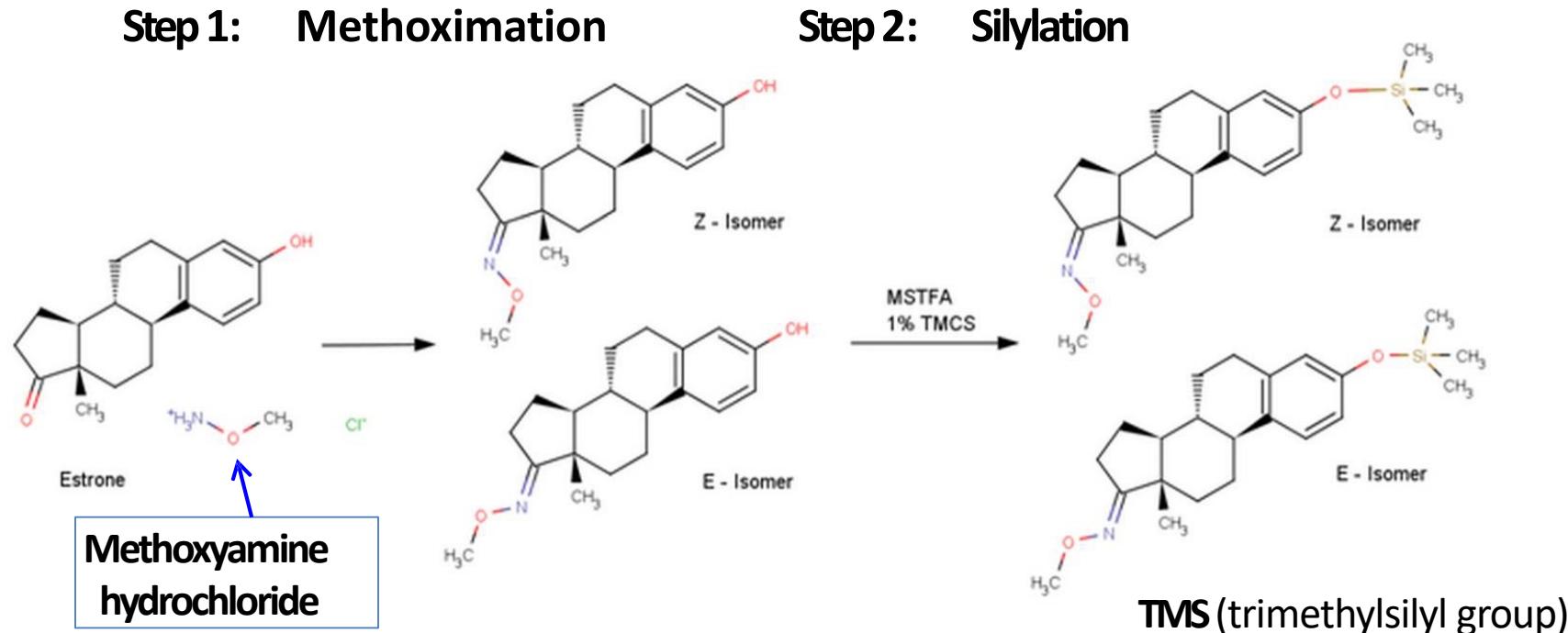
Quickly freeze sample in liquid nitrogen, store samples at -80°C (to avoid degradation or turnover).



Protocol used for extraction of metabolites and derivatization



Sample derivatization



Gas chromatography requires volatile compounds (two steps of derivatization in vial)

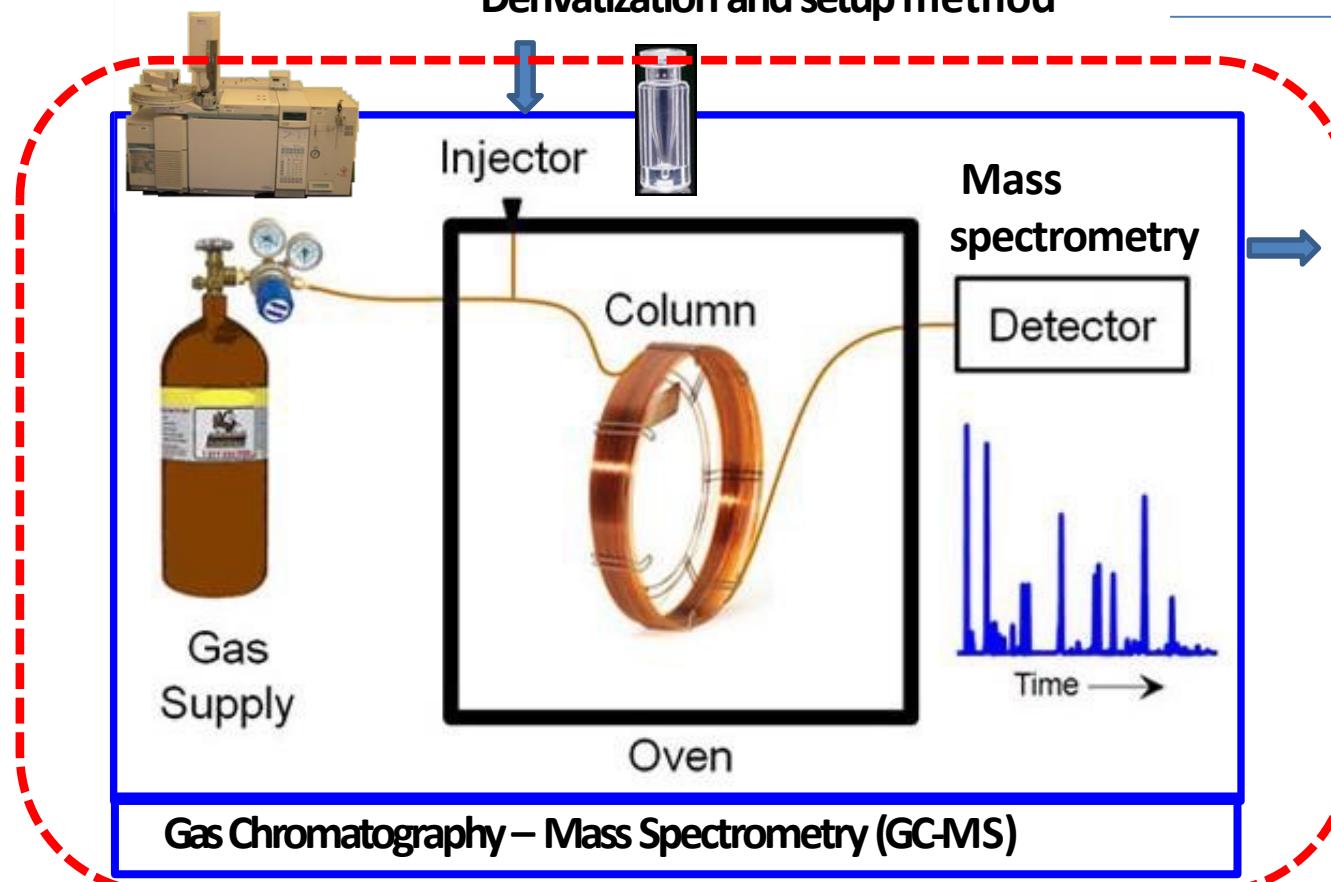
- 1) Methoximation of keto and aldehyde groups (primarily for opening reducing ring sugars).
- 2) Silylation of polar hydroxy, thiol, carboxy and amino groups with silylation agent MSTFA.
 - A single compound with multiple active groups will result in multiple peaks (1TMS, 2TMS, 3TMS)
 - GC-MS can distinguish between stereoisomers.

GC-MS

Definitions and background



Derivatization and setup method



Data analysis using different software



MS-DIAL

Biological significance and application

Chromotography

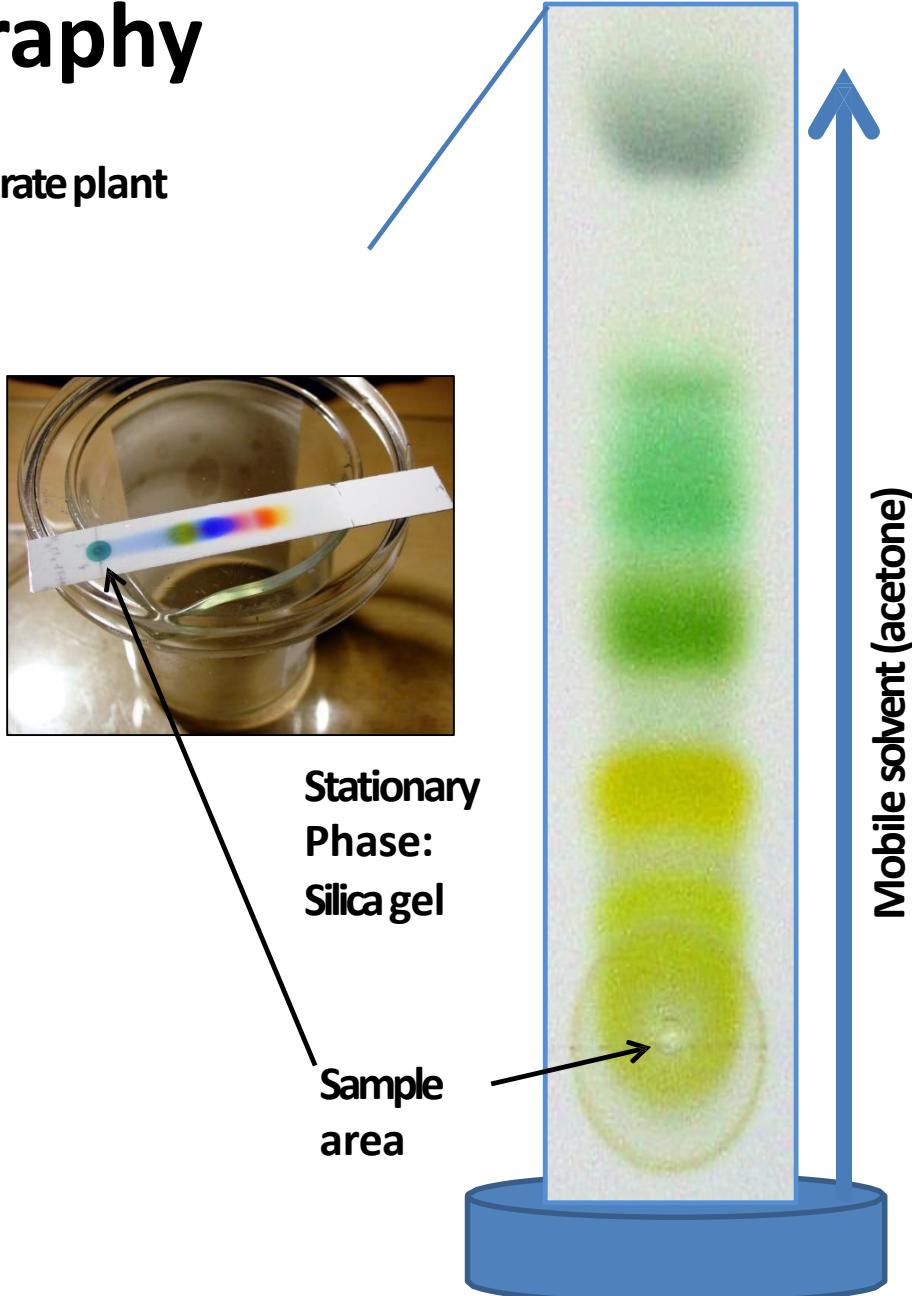
- Invented in 1900 by Mikhail Tsvet (used to separate plant pigments)

- There are several types of chromatography, but all consist of a stationary phase and a mobile phase.

Compounds are separated based on differential partitioning between the two phases.

- Types include:
 - LC(liquid chromatography)
 - GC(gas chromatography)

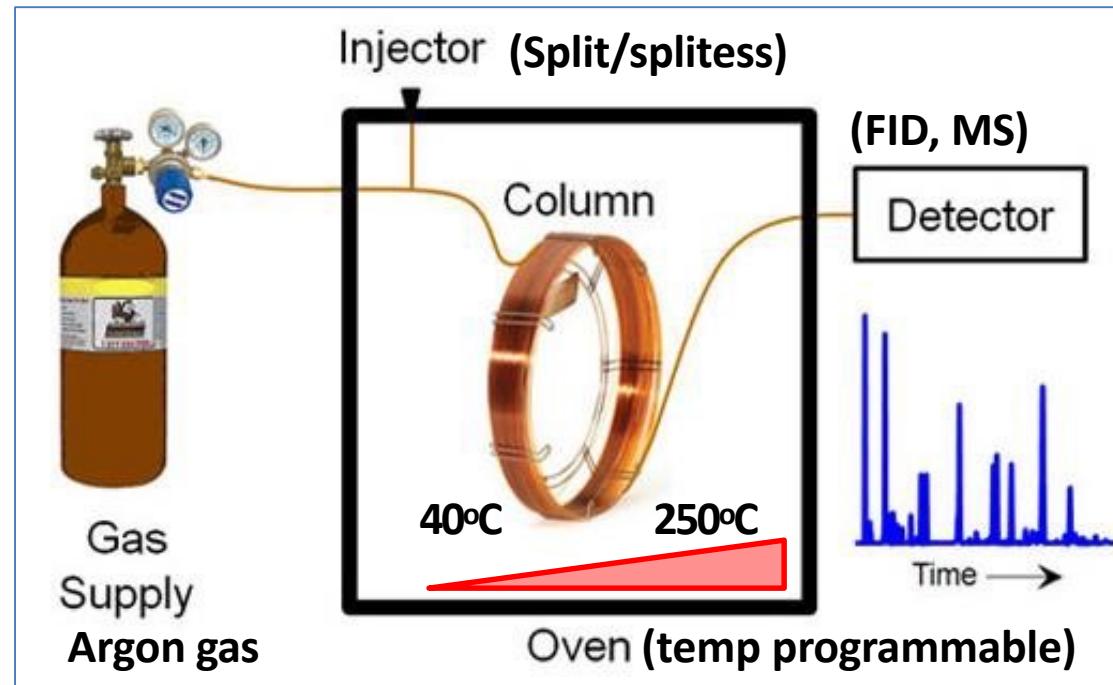
GC is routinely used in metabolomics.



Gas Chromotography

Routinely used for metabolomics

- GC= ‘good chromatography’
- optimized over several decades
- Capillary columns routinely used
(5% diphenyl/95% methyl siloxane)
- high reproducibility
Identification based on RT



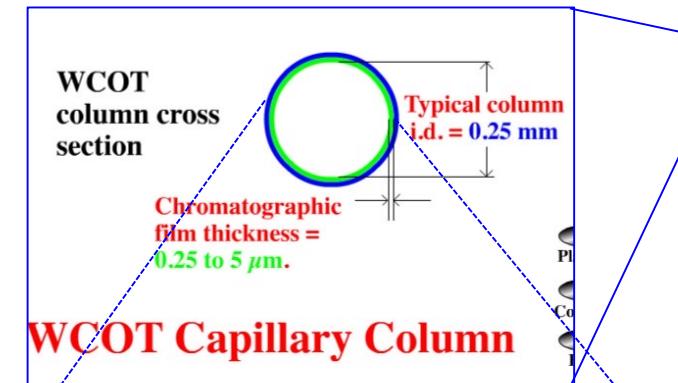
Limitations:

- high temperatures can destroy unstable compounds.
- polar compounds cannot ‘fly’ on GC columns and must first be derivatized.

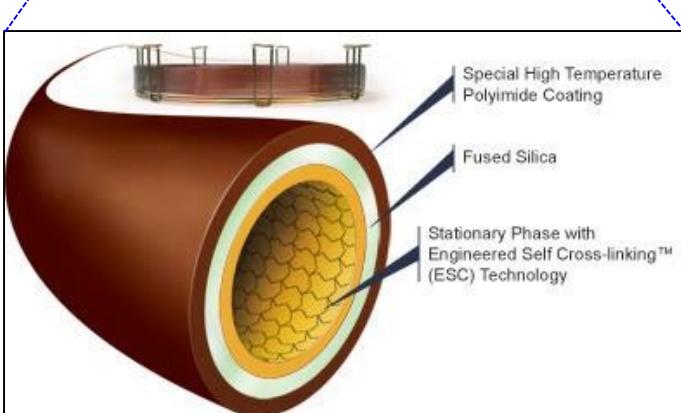
Retention time (RT)

GC Column

WCOT: Wall coated open tubular

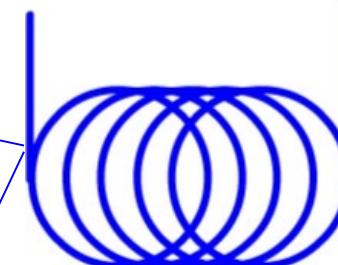


WCOT Capillary Column



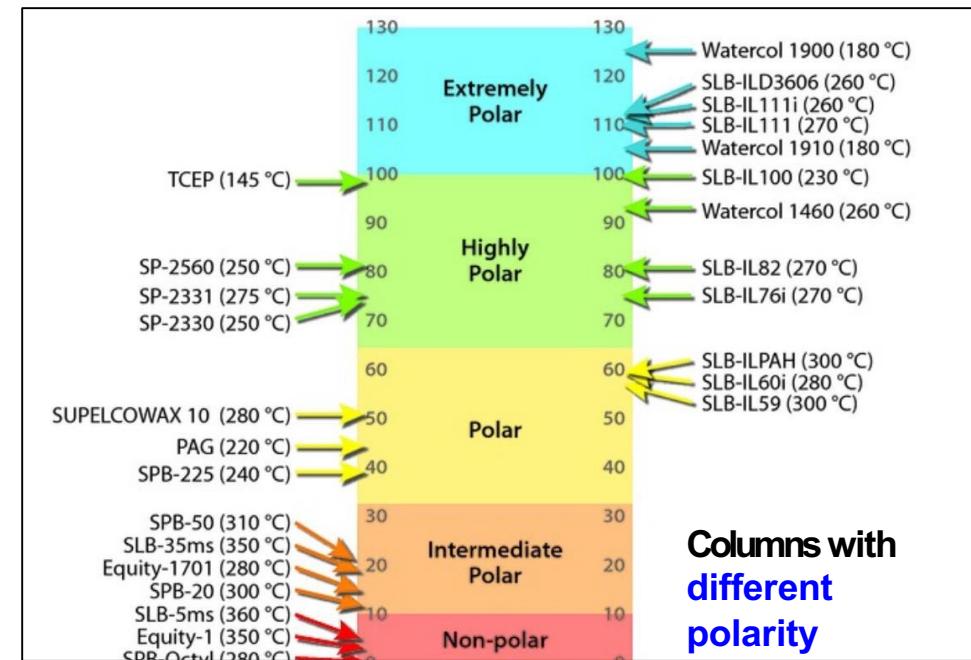
How GC columns work (video 2-3 mins)

<https://www.youtube.com/watch?v=q0pM-k0SvOQ>



(5% diphenyl/95% methyl siloxane) non-polar column

Typical capillary columns are 30 meters long, coiled up to fit in the GC oven.



Columns with different polarity

<https://www.sigmaaldrich.com/technical-documents/articles/analytical-applications/gc/select-proper-stationary-phase-gc.html>

GC Column

Analytes separate by two key factors when pass through column:

Sample Volatility: Volatile analytes move faster through the column.

In chemistry, **volatility** is a material quality which describes how readily a substance vaporizes

Analyte Polarity: Polar analytes interact strongly with polar column. Non-polar analytes interact strongly non-polar column.

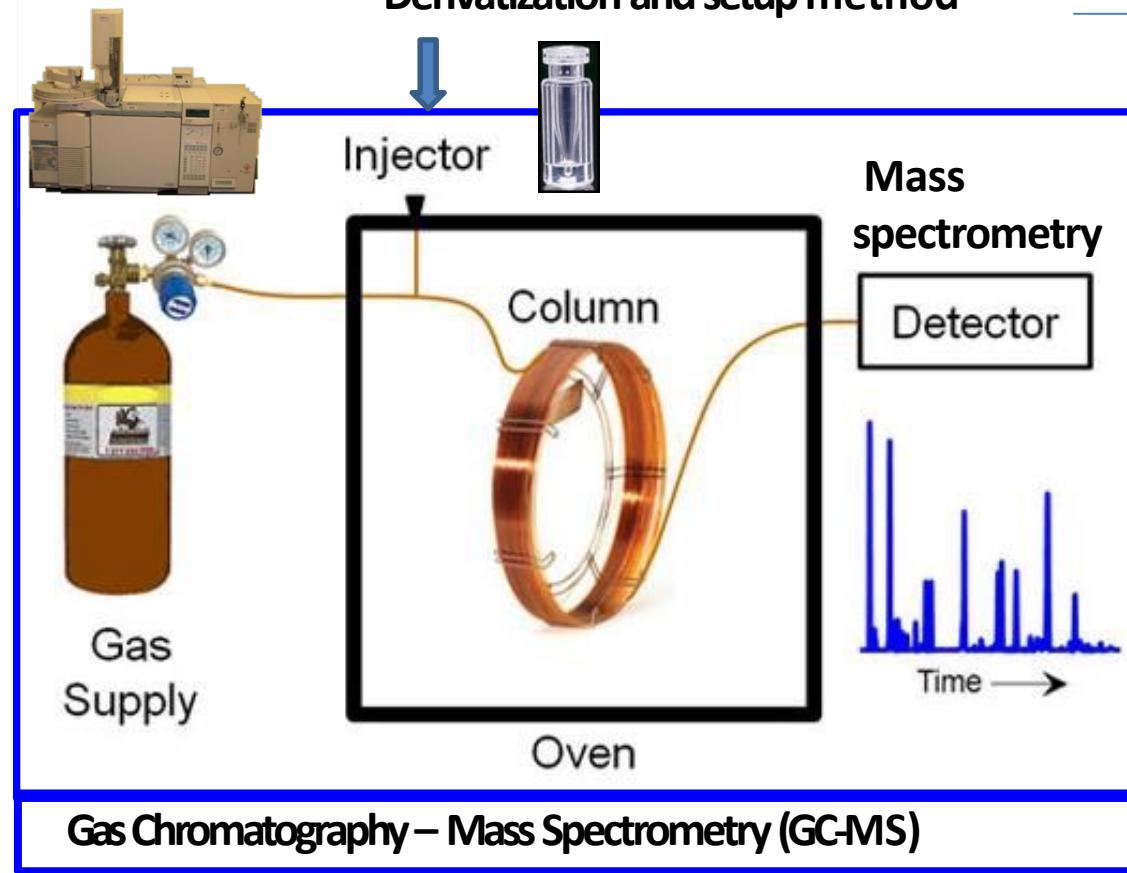
In chemistry, **polarity** is a separation of electric charge leading to a molecule or its chemical groups having an electric dipole moment, with a negatively charged end and a positively charged end.

GC-MS

Definitions and background



Derivatization and setup method



Data analysis using different software



MS-DIAL



MS-DIAL

Biological significance and application

MS-DIAL



► Software

► Database

Contact

Licence



MS-DIAL

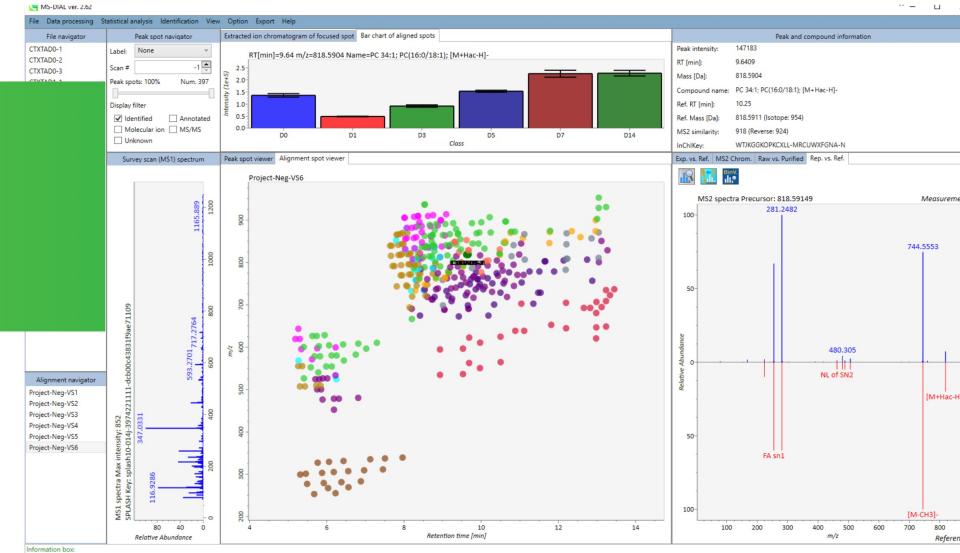


Objective

MS-DIAL was launched as a universal program for untargeted metabolomics that supports multiple instruments (GC/MS, GC/MS/MS, LC/MS, and LC/MS/MS) and MS vendors (Agilent, Bruker, LECO, Scieix, Shimadzu, Thermo, and Waters). Common data formats such as netCDF (AIA) and mzML, can also be managed in our project. In addition, we released several MSP files including both EI- and MS/MS spectra as a 'start-up kit'. Moreover, MS-DIAL internally has a version of Fiehn lab's GC/MS database (oriented by FAME RI index), and in silico retention time- and MS/MS database for LC/MS/MS based lipidomics. The isotope labeled tracking can also be executed in LC/MS project.

It features

- (1) spectral deconvolution for both GC/MS and data-independent MS/MS,
- (2) streamlined criteria for peak identification,
- (3) support of all data processing steps from raw data import to statistical analysis, and
- (4) user-friendly graphic user interface.



Citation

- ✓ General, and lipidomics: A lipidome atlas in MS-DIAL 4. *Nature Biotechnology*, 38, 1159–1163, 2020 [\[NPG link\]](#)
- Springer Nature 'SharedIt': [full-text access to a view-only version of this paper](#)
- ✓ Plant metabolomics and stable isotope labeling technology: A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms. *Nature Methods*, 16, 295–298, 2019 [\[NPG link\]](#)
- Springer Nature 'SharedIt': [full-text access to a view-only version of this paper](#)
- ✓ GC-MS project: Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nature Methods*, 15, 53–56, 2018 [\[NPG link\]](#)
- ✓ SWATH-MS/MS and DIA-MS: MS-DIAL: data independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature Methods*, 12, 523–526, 2015 [\[PubMed\]](#)

MS-DIAL

mtbinfo.github.io

msinfo manuals

[View the Project on GitHub](#)
mtbinfo-team/mtbinfo.github.io

This project is maintained by
[mtbinfo-team](#)

Hosted on GitHub Pages — Theme by [orderedlist](#)
MS-DIAL tutorial

Table of contents

Chapter 1: General introduction of MS-DIAL

Section 1-1: Downloading file converter to convert a vendor's format into ABF

Section 1-2: Check the conditions for file conversion

Section 1-3: File conversion

Section 1-4: Start up a project of MS-DIAL

Section 1-5: Centroid or Profile? The content of ABF data

Section 1-6: Database (MSP or Text) for compound identification

Section 1-6-1: MSP format for precursor- and MS/MS library

Section 1-6-2: Adduct ion format

Section 1-6-3: Text format library for retention time and accurate mass search (post identification)

Section 1-6-4: MSP format as GC/MS library

Section 1-6-5: Alkane- or FAME retention time dictionary for the calculation of retention index

Chapter 2: LC/MS/MS (data independent MS/MS) project with MS-DIAL's LipidBlast database (in silico retention time and MS/MS library for lipids)

Section 2-1: Starting up your project

Section 2-2: Importing ABF files

Section 2-3: Setting parameters

Section 2-3-1: Data collection tab

Section 2-3-2: Peak detection tab

Section 2-3-3: MS2Dec tab

Section 2-3-4: Identification tab

Section 2-3-5: Adduct tab

Section 2-3-6: Alignment tab

Section 2-4: Data curation for the reduction of false positive identifications

Chapter 3: LC/MS or LC/MS/MS (data dependent MS/MS) project with user-defined MS/MS database (MSP format) in MS-DIAL

Section 3-1: Starting up your project

Section 3-2: Importing ABF files

Section 3-3: Setting parameters

Section 3-3-1: Data collection tab

Section 3-3-2: Peak detection tab

Section 3-3-3: MS2Dec tab

MS-DIAL tutorial

Last edited in March. 12, 2020

ABSTRACT

MS-DIAL was launched as a universal program for untargeted metabolomics that supports multiple instruments (GC/MS, GC/MS/MS, LC/MS, and LC/MS/MS) and MS vendors (Agilent, Bruker, LECO, Sciex, Shimadzu, Thermo, and Waters). Common data formats such as netCDF (AIA) and mzML, can also be managed in our project. In addition, we released several MSP files including both EI- and MS/MS spectra as a 'start-up kit'. Moreover, MS-DIAL internally has a version of Fiehn lab's GC/MS database (oriented by FAME RI index), and in silico retention time- and MS/MS database for LC/MS/MS based lipidomics. The isotope labeled tracking can also be executed in LC/MS project.

It features (1) spectral deconvolution for both GC/MS and data-independent MS/MS, (2) streamlined criteria for peak identification, (3) support of all data processing steps from raw data import to statistical analysis, and (4) user-friendly graphic user interface.

MS-DIAL has been developed as the collaborative work between Prof. Masanori Arita team (RIKEN) and Prof. Oliver Fiehn team (UC Davis) supported by the JST/NSF SICORP "Metabolomics for the low carbon society" project.

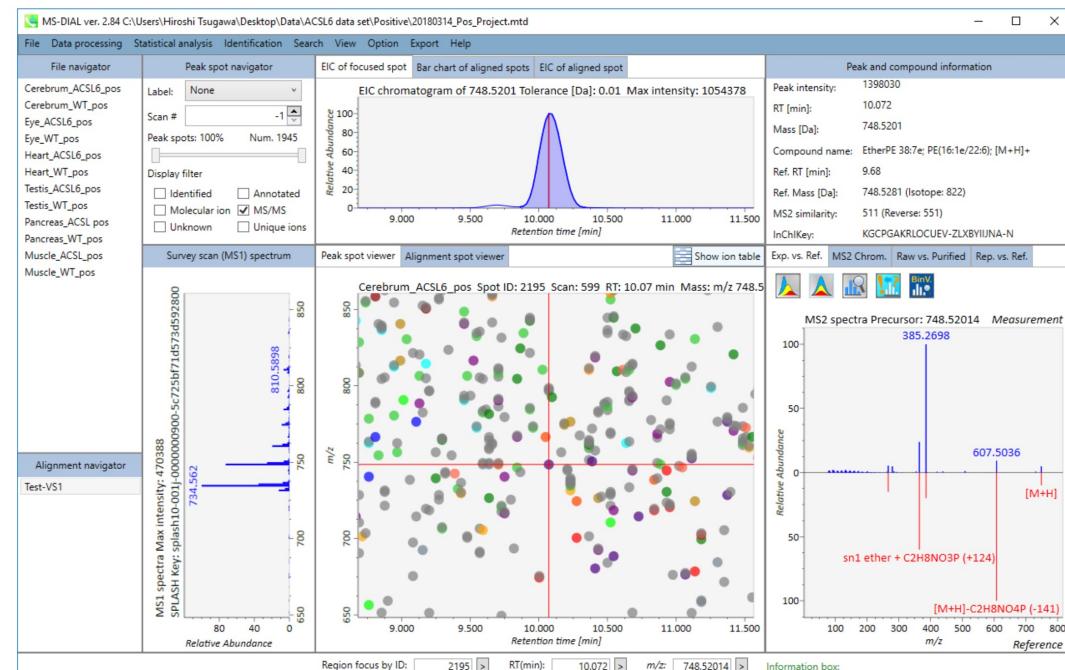
Hiroshi Tsugawa

RIKEN Center for Sustainable Resource Science

hiroshi.tsugawa@riken.jp

Lead developer: Hiroshi Tsugawa (RIKEN)

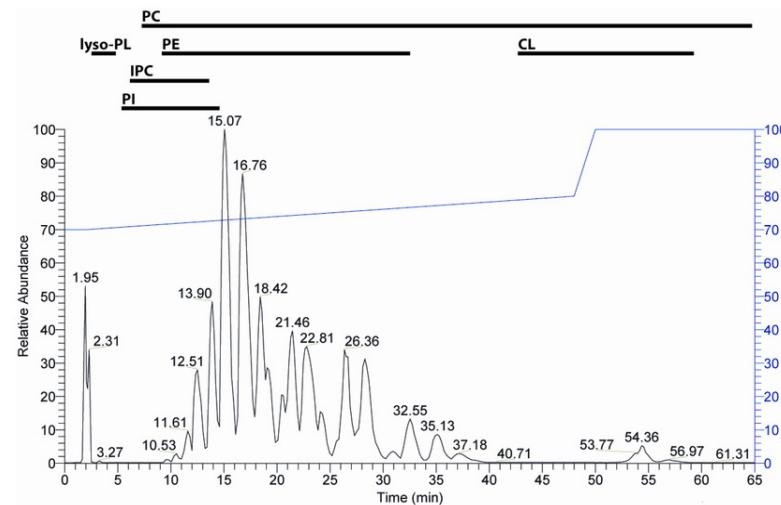
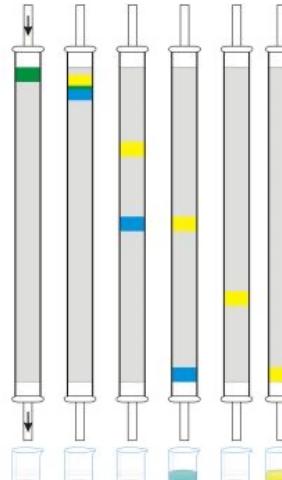
Main contributors: Diego Pedrosa (UC Davis), Tomas Cajka (Institute of Physiology CAS), Iputta Tada (NIG), Haruki Uchino (Keio)



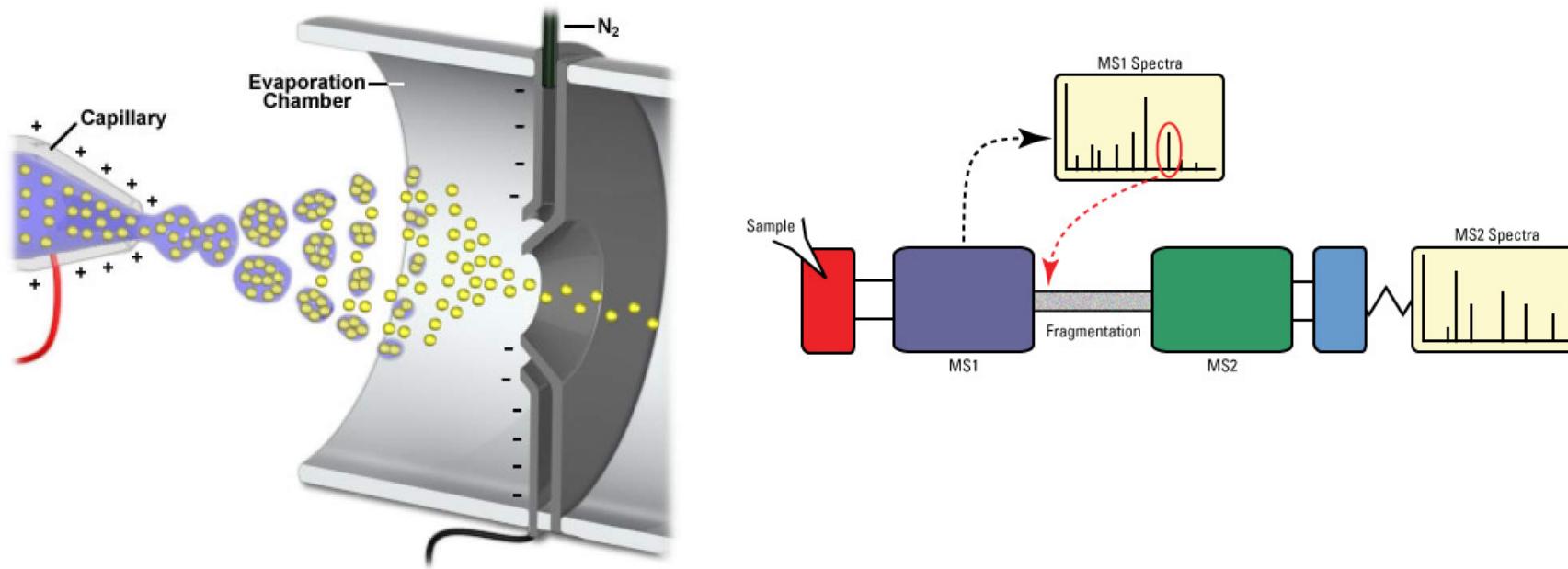
MS-DIAL screenshot

Untargeted and Targeted LC – MS Metabolomics

- Reverse Phase (C-18) HPLC
- Non-polar stationary phase (hydrophobic), polar mobile phase (hydrophilic) – Gradient $\text{H}_2\text{O} \rightarrow \text{Acetonitrile}$



ESI and Tandem MS (MS/MS)



Targeted and Non-Targeted Analysis

- Targeted approaches mainly focus on a single metabolite or group of metabolites.
 - i. Highly sensitive
 - ii. Less complex data handling and analysis
 - iii. Requires *a priori* knowledge of targeted metabolite
- Non-targeted approaches aim to cover the metabolome as broadly as possible.
 - i. Excellent for the discovery of novel or unexpected metabolites
 - ii. Very complex data handling and analysis
 - iii. May require further analysis for complete characterisation.

Targeted and Non-Targeted Analysis

- 4 modes of MSMS analysis

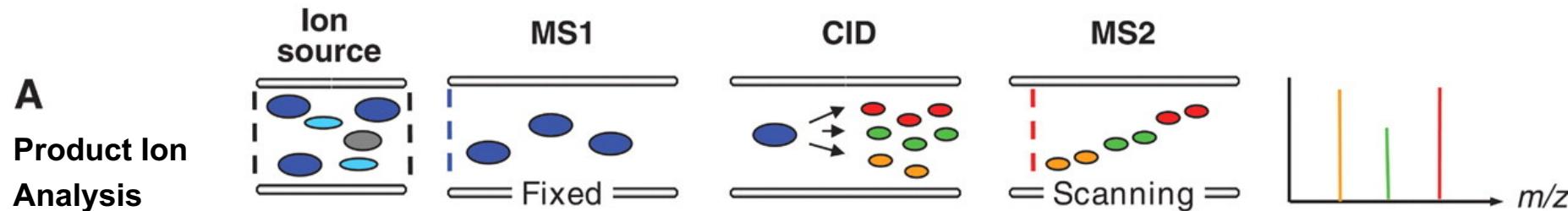
Non-Targeted

- Product-Ion analysis mode (PIA) (Q-ToF MS)

Targeted

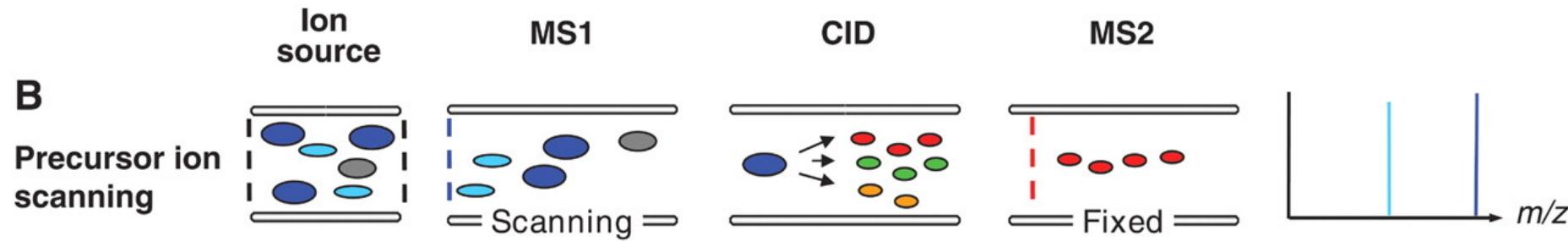
- Precursor-Ion scanning mode (PIS) (QqQ MS)
- Neutral-Loss scanning mode (NLS) (QqQ MS)
- Selected reaction monitoring mode (SRM) (QqQ MS)

Product Ion Analysis (PIA) - Untargeted



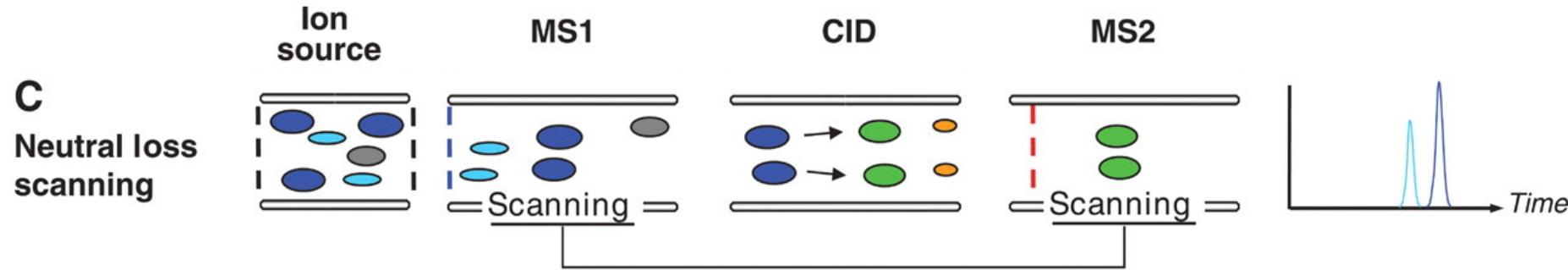
- In the product-ion analysis mode, the first mass analyzer selects a particular precursor ion of interest by setting the mass analyser to transmit only the particular ion of interest. The selected ion is fragmented in the collision cell with CID. The resultant product ions are analyzed with the second mass analyser in scanning mode

Precursor Ion Scanning (PIS) - Targeted



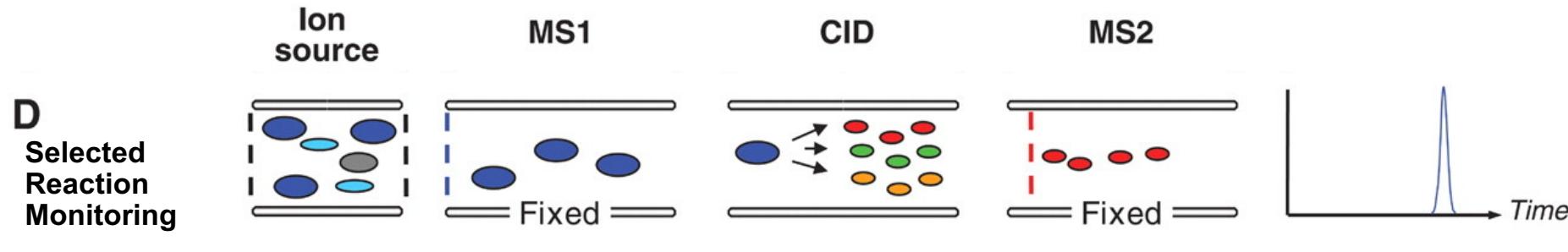
- In the precursor-ion scanning (PIS) mode, the second analyzer now focuses on a particular product ion of interest after CID while scanning the m/z ratios in the first mass analyzer. All of the precursor ions that produce the selected product ion after fragmentation are thus detected

Neutral Loss Scanning (NLS) - Targeted



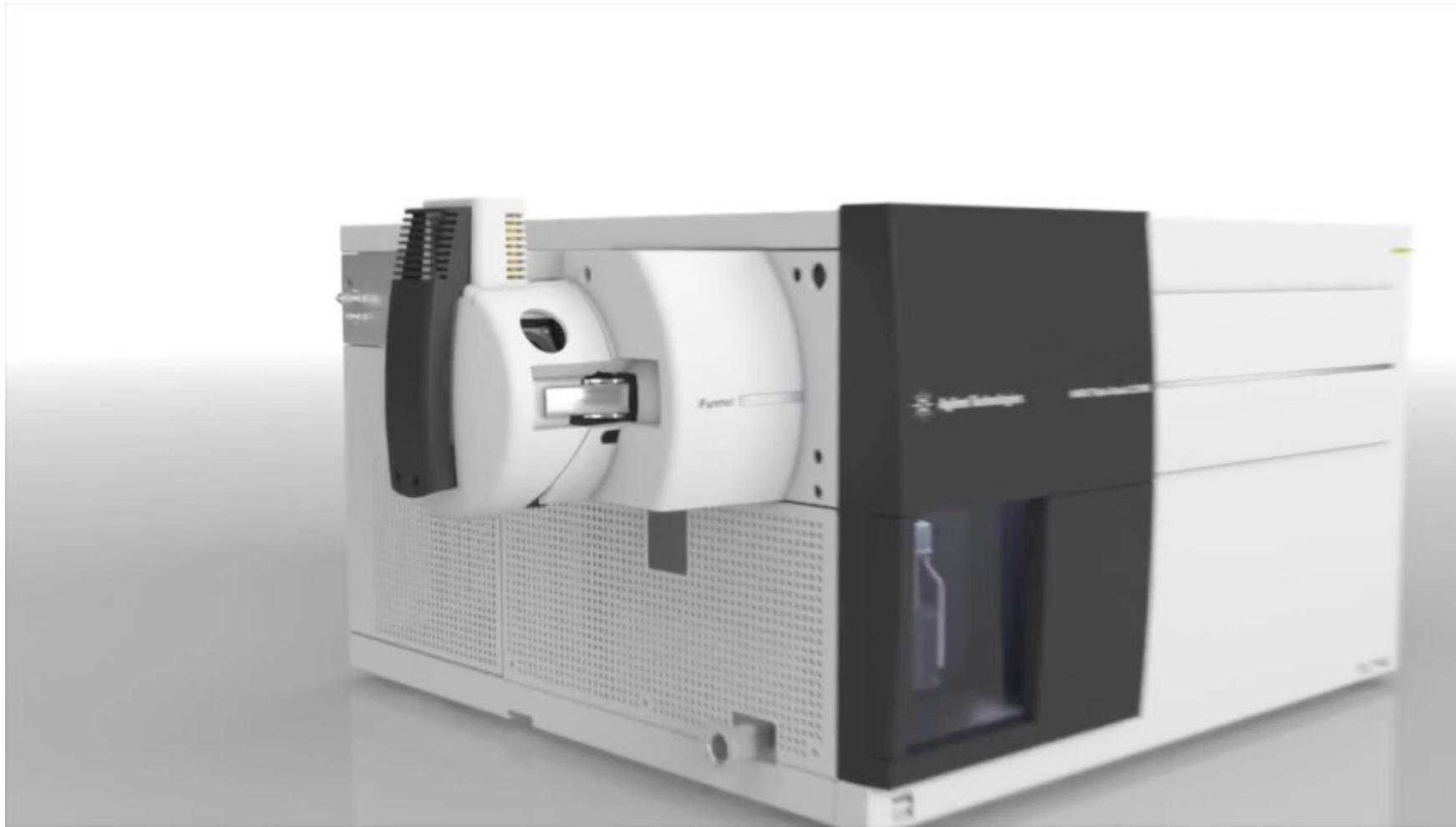
- In the neutral-loss scanning (NLS) mode, both the first and second mass analyzers are scanned simultaneously, but with a constant mass offset between the two. When a precursor ion is transmitted through the first mass analyzer, this ion is recorded if it yields a particular product ion that corresponds to the loss of a neutral fragment that has a mass from the precursor ion.

Selected Reaction Monitoring (SRM) – Targeted



- In the selected reaction monitoring (SRM) mode, transitions between the molecule ion and product ion must be previously known, and the first and second mass analyzers are both focused on the selected ions.

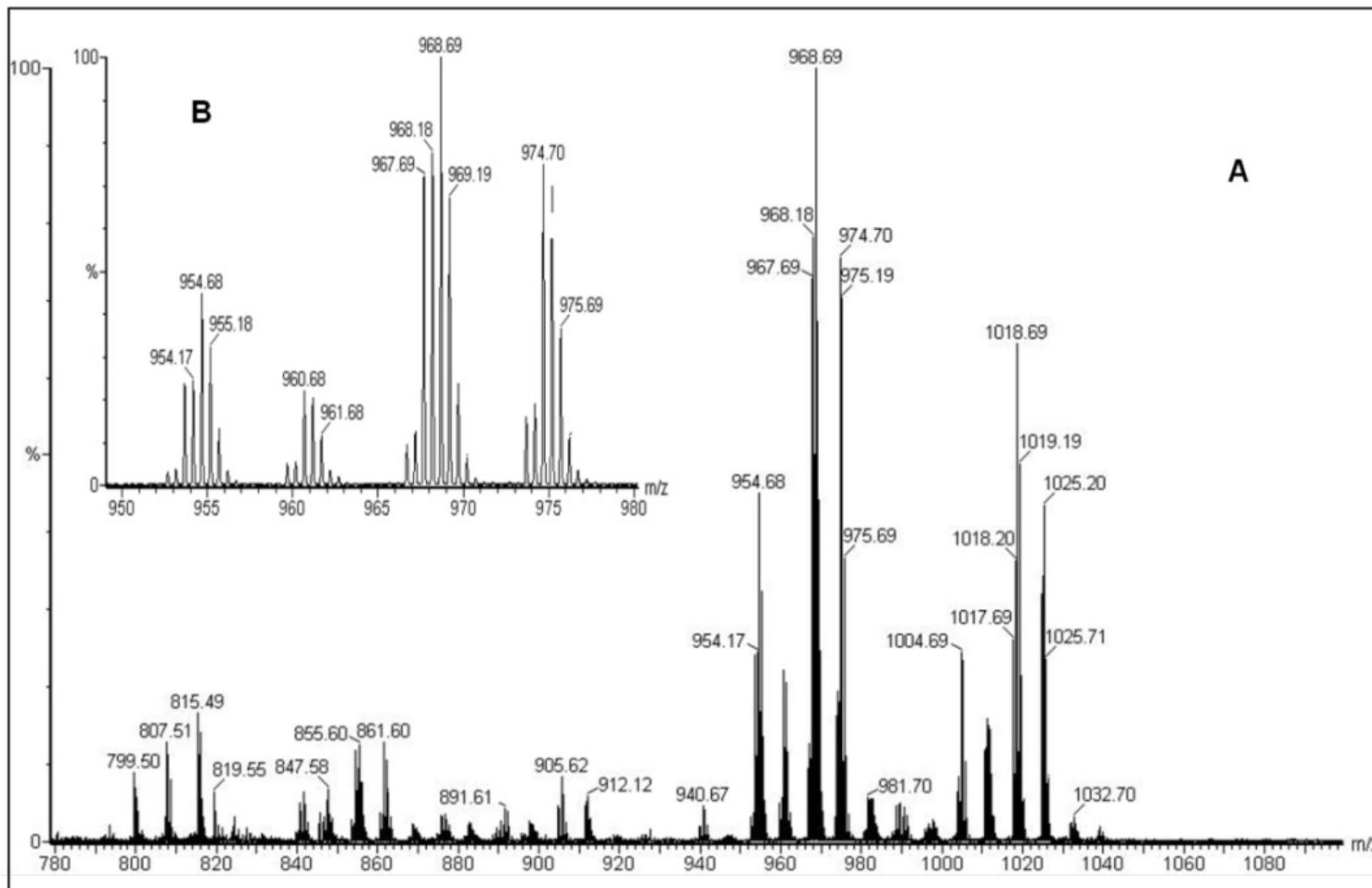
SRM in a Triple Quad (QqQ) mass spectrometer



Class Specific Scanning

Lipid class (ref.)	Ion format	Scans for class specific prescreen	Scans for identification of acyl chain and/or regioisomers
PC (Yang et al., 2009b)	$[M+Li]^+$	NLS189, -35 eV NLS 183, -35 eV	NLS(59+FA), -40 eV
lysoPC (Yang et al., 2009b)	$[M+Na]^+$	NLS59, -22 eV NLS205, -34 eV	PIS104, -34 eV PIS147, -34 eV
PE, lysoPE (Han et al., 2005)	$[M-H]^-$ $[M-H+Fmoc]^-$ ($[M+C_{15}H_9O_2]^-$)	PIS196, 50 eV for $[M-H]^-$ NLS222, 30 eV	PIS(FA-H), 30 eV
PI, lysoPI (Han et al., 2004b)	$[M-H]^-$	PIS241, 45 eV	PIS(FA-H), 47 eV
PS, lysoPS (Han et al., 2004b)	$[M-H]^-$	NLS87, 24 eV	PIS(FA-H), 30 eV
PG, PA, lysoPG, lysoPA (Han et al., 2004b)	$[M-H]^-$	PIS153, 35 eV	PIS(FA-H), 30 eV
Cardiolipin (CL), mono-lysoCL (Han et al., 2006a)	$[M-2H]^{2-}$	Full MS at high resolution	PIS(FA-H) at high resolution, 25 eV; NLS(FA-H ₂ O) at high resolution, 22 eV

Metabolite Identification



Other Mass Spectral Databases

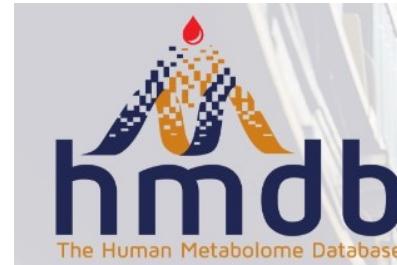
HMDB (Human metabolome database)

~42,000 metabolite entries

Combines with four databases, [DrugBank](#), [T3DB](#) (*Toxic Exposome Database*), [SMPDB](#) (The Small Molecule Pathway Database) and [FooDB](#) (food) are also part of the HMDB suite of databases

5,912 MS/MS spectra

chemical data, clinical data, and molecular biology/biochemistry data



MassBank

mass spectral database of experimentally acquired high resolution MSspectra of metabolites.

>41,000 MS/MS spectra



In Silico fragmentation – The Key Idea

Given a molecular weight and its spectrum, we would like to ID the metabolite

Find candidate matches based on molecular weight, and then computationally fragment each candidate

Using fragmentation data, evaluate how well we can explain peaks in the measured spectrum

Different techniques for identifying fragments

Fragmentation Trees

Assume multi-step fragmentation process

Systematically disconnect bonds, tracking parent fragment

Create “Fragmentation Trees”

Limit tree depth

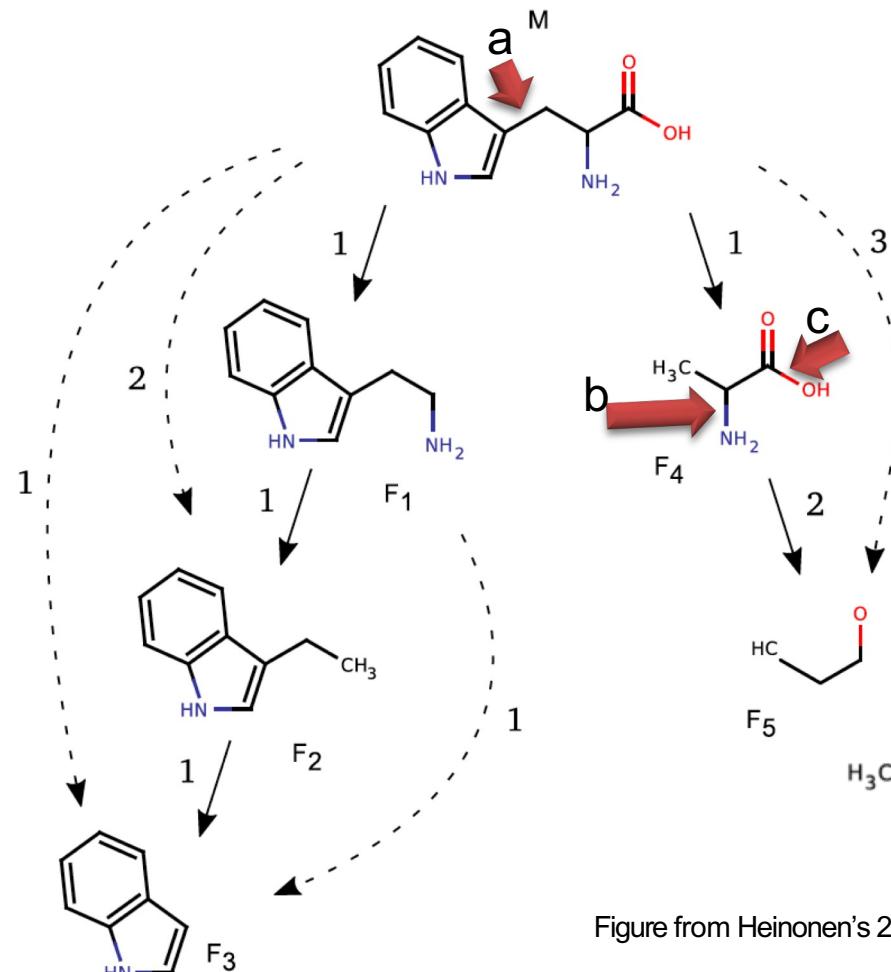
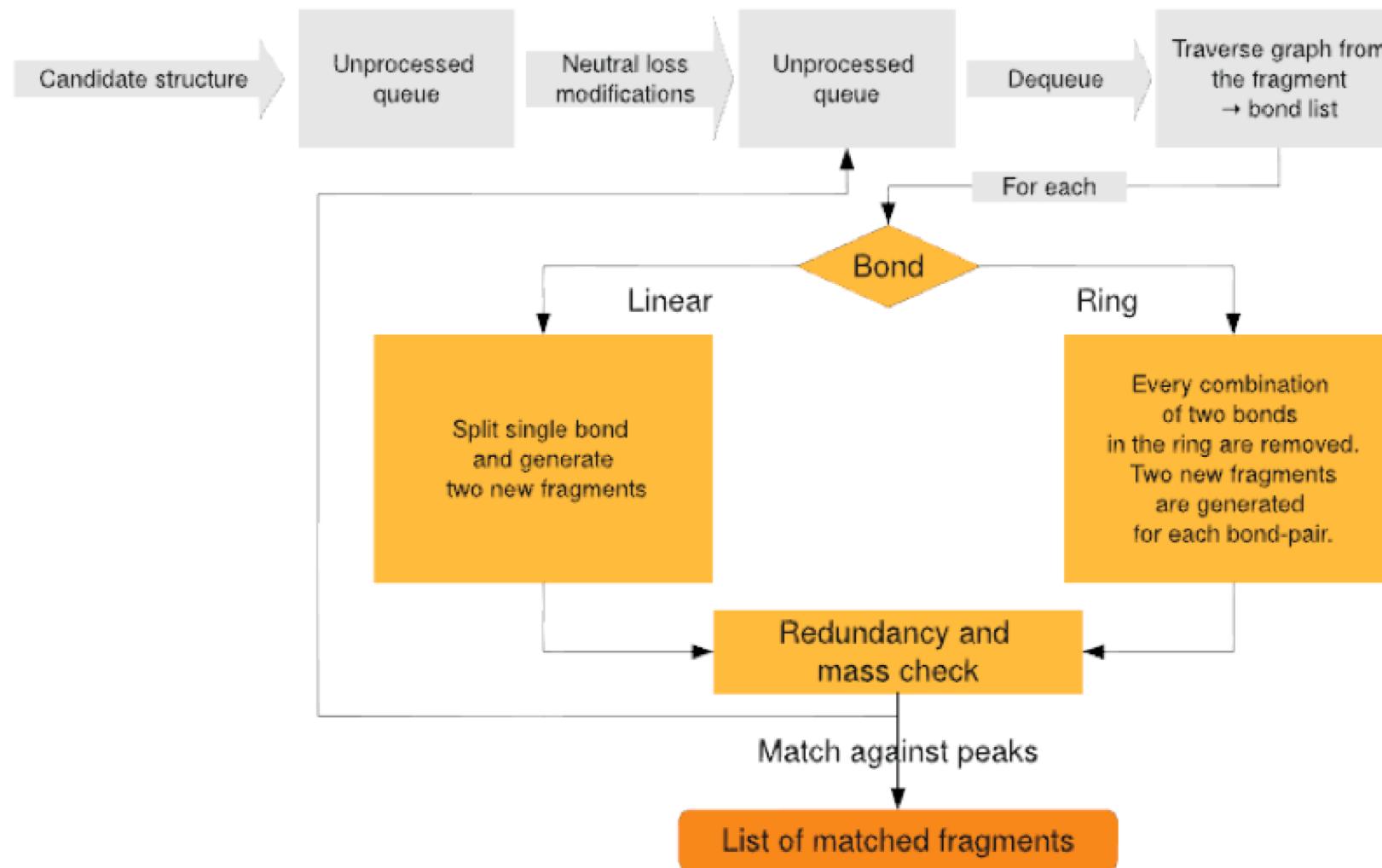


Figure from Heinonen's 2012 PhD Thesis

Generating Fragmentation Trees (MetFrag)





MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

MetFrag MzAnnotate Viewer About / News

Database Settings

Database: KEGG PubChem ChemSpider Local SDF

Neutral exact mass: Search PPM:

Molecular formula:

Only biological compounds:

Limit # of structures:

Database ID's:

MetFrag Settings

Mode: [M+H] [M-H] [M]

Charge: pos. neg.

Mzabs (e.g. 0.01):

Mzppm (e.g. 10):

Parent ion: Neutral

Peaks:

138.066 10877
123.043 350
110.071 2221
109.036 234
83.060 318
69.045 351
42.034 644

*Ruttkies, Christoph, et al. "MetFrag relaunched: incorporating strategies beyond *in silico* fragmentation." *Journal of cheminformatics* 8.1 (2016): 1.