Identification of Metabolites tools and methods
Ways to identify compounds

• Accurate mass - Normally first step to finding the compound
  - database search using small ppm window

• Adducts :

• Retention time - running a standard or local method
  database lookup

• Drift :

• Fragmentation patterns -

• Data - incomplete databases
Accurate mass

• Features must be characterised to known neutral mass

• Mass Spectrometers have errors on mass - leads to larger search space.

• Too many isobaric compounds
Retention time

- Retention time against standards is a powerful tool
- Must have the standards
- Have to keep the same gradient for database lookup method
  - Columns and LC methods are changing
  - Don’t always use LC method - Rapid injection methods
Fragmentation

- Fragmentation is a (sort of) 2D representation of the compound
- Lots of issues
  - Specificity of isolation window
  - Scoring algorithms
  - Access to standards
Matching spectra to compounds
Scoring methods

- Either matching to a standard or a database
- Many different metrics some are more sensitive

- Cosine similarity suffers from low resolution issues and should be low sensitivity and specificity
- X-Rank - a very cool algorithm looks at peaks that matter the most
- Many more

\[
\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}
\]
Why it works

- Precursor selection with fragmentation data is highly specific
- The scoring is only needed for ranking purposes
Databases

Very hard to tell how many unique compounds Mass Bank has
Methods to matching MS-DIAL

**Targeted Quantitative Workflows**

**Top Two Techniques**

**SWATH™ Acquisition**

**MRM Workflow**

**MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis**

Hiroshi Tsugawa¹,², Tomas Cajka³, Tobias Kind³, Yan Ma³, Brendan Higgins⁴, Kazutaka Ikeda⁵,⁶, Mitsuhiro Kanazawa⁷, Jean VanderGheynst⁴, Oliver Fiehn³,⁸ & Masanori Arita¹,⁹

Data-independent acquisition (DIA) in liquid chromatography (LC) coupled to tandem mass spectrometry (MS/MS) provides comprehensive untargeted acquisition of molecular data. We provide an open-source software pipeline, which we call MS-DIAL, for DIA-based identification and quantification of small molecules by mass spectral deconvolution. For a reversed-phase LC-MS/MS analysis of nine algal strains, MS-DIAL using an enriched LipidBlast library identified 1,023 lipid compounds, highlighting the chemotaxonomic relationships between the algal strains.
Methods to matching
Similarity matching
De-novo methods

- General method is to take a structure of a molecule and break bonds (either randomly or rule based) to match to fragments masses.

MS Fragmentor
Mass-MetaSite
Agilent Technologies
Mass Frontier/
MetWorks
Thermo Scientific
MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

Database Settings
- Database: KEGG, PubChem, ChemSpider, Local SDF
- Neutral exact mass: 272.06847
- Molecular formula: C_{12}H_{12}O_{5}
- Only biological compounds: 1
- Limit # of structures: 100
- Database ID's: 15 hits

MetFrag Settings
- Mode: Charge, MZ of (e.g. 0.01)
- MZppm (e.g. 10): 10

Peaks:
- [M+H], [M-H], [M]
- pos. or neg.

View spectrum

Score: 1.0

Trivial Name: Naringenin chalcone
- Z',A',6'-Tetrahydroxychalcone
- Inositolpp
- Chalconaringenin

Exact Mass: 272.0685
Structure: C_{12}H_{12}O_{5}
Database ID: C06085
Actions: Fragments, Download

Score: 0.989

Trivial Name: p-Coumaroylactic acid lactone

Exact Mass: 272.0685
Structure: C_{12}H_{12}O_{5}
Database ID: C12087
Actions: Fragments, Download

Download complete table: Generate output files
NEW JOB SUBMISSION

Job Submitted
Job Id 8bd1f9437-50fd-4ea3-816d-94e62a8e651f

Email sent to hbenton@gmail.com with URL to track progress.

CURRENT STATUS

Job Submitted  Job In Queue  Job In Execution  Job Execution Complete

Copyright © ORNL 2014
MolGen - EI spectra

Low Resolution EI Mass Spectrum of an Unknown Chemical Compound

- MS-Classification (MSclass)
  - Structural Info
  - Intervals for Numbers of Atoms for each Chemical Element
- Elemental Composition Computation (ElCoCo)
  - Molecular Formulae
- Knowledgebase for Matchvalue & Molionmass
- Structure Generation (MOLGEN)
  - Structural Formulae
  - Generation of the MS Reaction Network
    - Comparison for Compatibility with the Experimental Spectrum
- Ranking of the Structural Formulae with Matchvalues that Describe the Compatibility with the Spectrum of the Unknown

160 MS Classifiers
92 MOLGEN Substructures
MS Reaction-mechanisms
Personalised Databases

- Crowd sourcing MS/MS data
- mzCloud - run by thermo collection of user based fragmentation data
- Metlin - Using XCMS Online each time someone searches MS/MS data the spectra is kept in their space. If approved by user as compound X, next time search on user’s personal MS/MS database.
Thank you! Questions?

Suggested reading material for more!!

[Article citation]
Computational mass spectrometry for small molecules
Kerstin Scheubert, Franziska Hufsky, and Sebastian Böcker