Metabolite identification in metabolomics: Database and interpretation of MSMS spectra

Jeevan K. Prasain, PhD
Department of Pharmacology and Toxicology, UAB
jprasain@uab.edu

Outline

• Introduction
• Putative structures identification - database or De novo structure determination by MS/MS
• Conclusions
Introduction

• Identification of metabolites at a molecular level is the biggest bottleneck in metabolomics due to their structural diversity (isobars and isomers) and dynamic metabolism.

• Considering the number of metabolites is >2000,000, there is a lack of commercial analytical standards (only a few thousands available) or comprehensive databases.
  – Note that there is the opportunity to make specific metabolite standards through the NIH Common Fund
  – Go to http://metabolomicsworkbench.org

• MS/MS interpretation is needed for validation of annotated structure and unknown determination.

• Inclusion of many artifacts in database.

• Structural complexity of metabolites.

Metabolite identification workflow

Raw LC-MS data
\[\downarrow\]
Data file processed by algorithms (e.g., XCMS)
\[\downarrow\]
\[P\text{-value, fold change}\]
Metabolomics/chemical database search (known/new or novel)
\[\downarrow\]
Interpretation of MS/MS of a precursor ion (accurate mass, isotope data and nitrogen rule)
\[\downarrow\]
Putative identification (molecular formula determination)
\[\downarrow\]
De novo structure elucidation (NMR and or LC-MS/MS)
Keys to identifying chemical structures (putative/definitive) by mass spectrometry

- Retention time in LC
- Accurate mass
- Isotope distribution
- Nitrogen rule
- Fragmentation pattern of a precursor ion
- Comparison with authentic standards (definitive)


LCMS-based metabolomics

- Detection of intact molecular ions \([M+H]^+/[M-H]^-\) is possible with soft ionization such as ESI
- High mass accuracy of many instruments (<5 ppm, 0.0005%) helps identify isobaric compounds
- Enables the separation of complex mixtures and identification of molecular weight of pure compounds
- Substructures of unknown metabolite may be proposed on the basis of LC retention time, exact mass measurement and interpretation of signature ions upon MS/MS of a precursor ion
Platform to process untargeted metabolomic data

- XCMS (developed by the Siuzdak Lab at the Scripps Research Institute) Online, is a web-based version that allows users to easily upload and process LC-MS data. It is a bioinformatics platform to identify endogenous metabolites.

- METLIN ([http://metlin.scripps.edu](http://metlin.scripps.edu)) is a metabolite database for metabolomics containing over 64,000 structures and it also has comprehensive tandem mass spectrometry data on over 10,000 molecules at different collision energies.

- Provides an annotated list of known metabolites, their masses, chemical forms and structures.

---

Not every peak represents individual metabolite: Adduct formation

Nielsen et al., J Nat Prod. 2011
XCMS online platform to process untargeted metabolomic data

Annotated structures and their validation by MS/MS interpretation

Calculated m/z 258.9952, mass error = 16 ppm

Calculated m/z 258.9918, mass error = 3.4 ppm

Caffeic acid 4- or 3- sulfate, Mass error = 3.4 ppm

Calculated m/z 179.0350

-79.956
SO3
Calc. 79.956
M/z 187.0976 was identified as nonanedioic acid by comparing MS/MS profile between experimental and Metline data base

M/z 187.0976 was identified as nonanedioic acid by comparing MS/MS profile between experimental and Metline data base.

Among the annotated list of compounds by Metlin- phenylacetylglycine’s validation by MS/MS interpretation.
Detection of Vitamin B2 (riboflavin) as urinary metabolite-fragmentation patterns matched with Metlin database

Many metabolites, unidentified by the Metlin database
A medium chain dicarboxylic fatty acid with

\[ m/z \ 241.109 \ [M-H]^{-} \]

\[
\text{C12H17O5} \ [M-H]^{-} \\
\text{Calc. } m/z \ 241.1081
\]

Neutral loss of monodehydrated glucuronic acid (calc. 176.032 Da) -
an indicative of Glucuronidated Metabolite

\[
\text{Calc. 176.0321 Da}
\]
Isotopic pattern and intensity of ions indicates the number of carbons and hetero atoms in the molecular ion

$^1$H = 99.9%, $^2$H = 0.02%
$^{12}$C = 98.9%, $^{13}$C = 1.1%
$^{35}$Cl = 68.1%, $^{37}$Cl = 31.9%

8-chlorodaidzein
m/z 287.011 (M-H)+; 289.008 [M-H+2]-

Library search for eicosanoid http://www.lipidmaps.org/
Good chromatographic separation and accurate mass are the important steps in structure determination.
Separation of stereoisomers by a chiral normal phase column

Hoang et al., PLOS Genetics. 2013

Nitrogen rule-
Odd number of nitrogens = odd MW
No nitrogen or even nitrogens = even MW
Accurate mass (<5 ppm), fragmentation patterns help propose putative structures

Prasain et al., Metabolites 2015

Structure determination: Accurate mass of a precursor ion is not enough, but MS/MS is

Conclusions

• Identifying unknown metabolites is a significant analytical challenge in metabolomics and it requires integrated analytical and bio-informative approaches.
• Data processing and data analysis are important for putative identifications.
• The use of high-resolution MS and MS\(^n\) provides important information to propose structures of fragment and precursor ions.
• Only an integrated approach can describes multitude of metabolites present in a biological sample.