THE UNIVERSITY OF ALABAMA AT BIRMINGHAM

Knowledge that will change your world

Metabolite identification in metabolomics: Metlin Database and interpretation of MSMS spectra

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

Outline

- **Introduction**
- **Metabolite annotation**
- **Putative structure identification**
  - database
  - *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 >200,000, there is a lack of commercial analytical standards (only a few thousands available) or comprehensive databases.**
- **In untargeted metabolomics, >30% of the compounds are rarely identified (*Blaženović et al., 2017*).**
- **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

↓

Data file processed by algorithms

(e.g., XCMS)

↓ *P*-value, fold change

Metabolomics/chemical database search

(known/new or novel)

↓

Metabolite annotation

↓

Putative structure identification

↓

***Validation, 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
- Product/precursor ion intensity ratio
- Comparison with authentic standards (definitive)

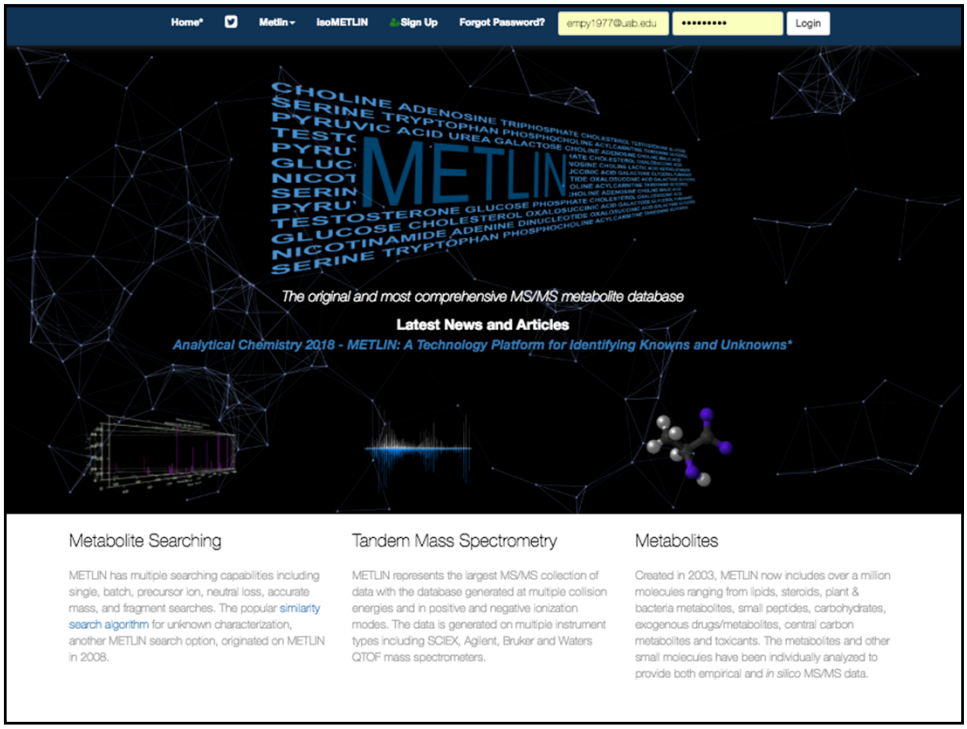
Moco et al. Trends in Analytical Chemistry, 2007

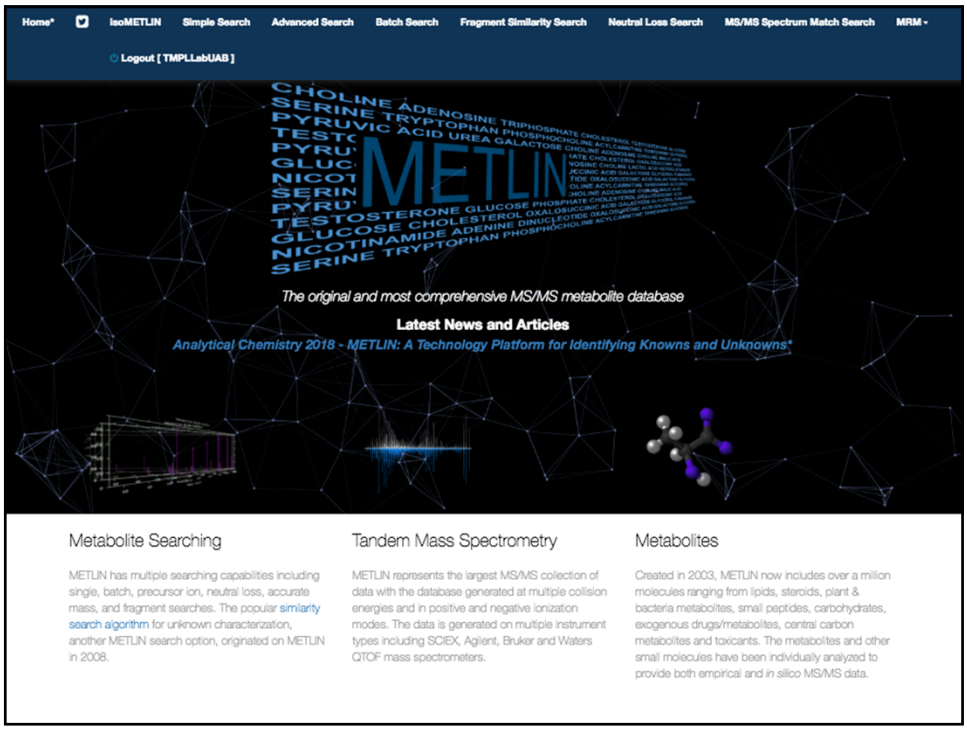
LCMS-based metabolomics

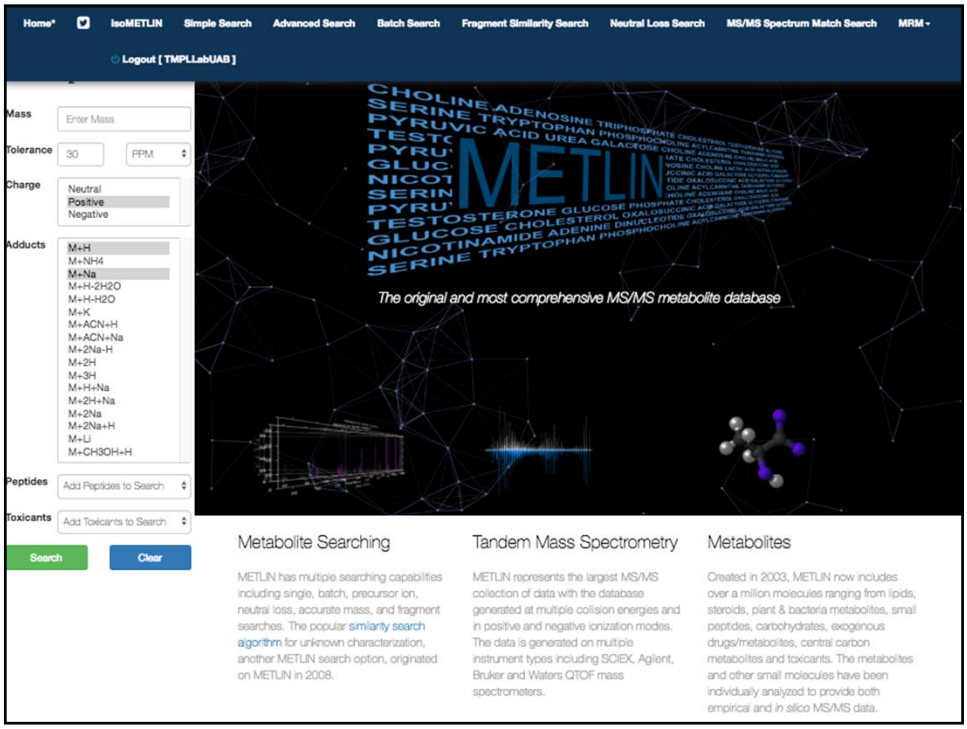
- Detection of intact molecular ions $[M+H]^+/[M-H]^-$ is possible with soft ionization such as ESI
- High mass accuracy of many instruments (<5 ppm, 0.0005%) helps identify isobaric compounds
- Enables the separation of complex mixtures and identification of molecular weight of pure compounds
- Substructures 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

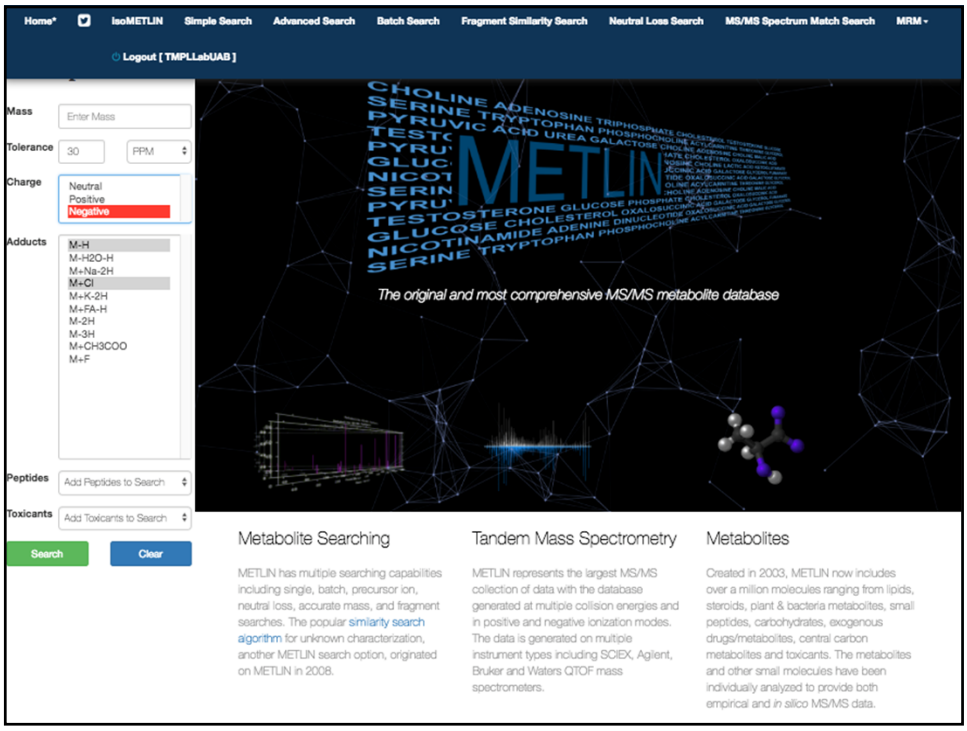
Platform to process untargeted metabolomic data

- **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) is a metabolite database for metabolomics containing 1 million compounds 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.

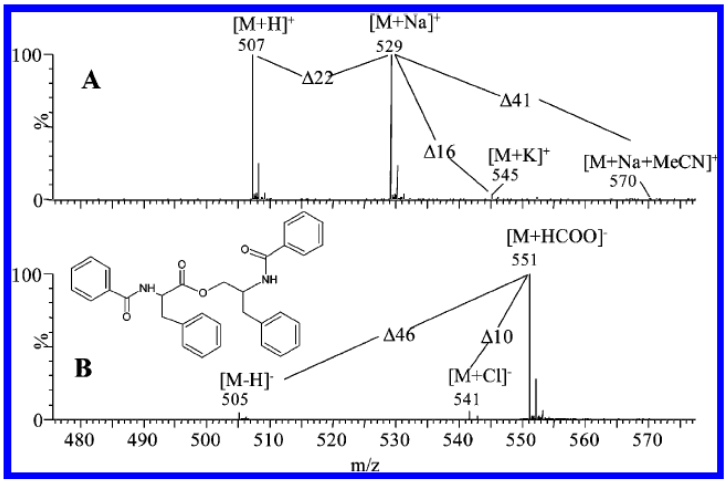



Home*
isoMETLIN
Simple Search
Advanced Search
Batch Search
Fragment Similarity Search
Neutral Loss Search
MS/MS Spectrum Match Search
MRM
Logout [TMPLLabUAB]
METLIN
The original and most comprehensive MS/MS metabolite database
Latest News and Articles
Analytical Chemistry 2018 - METLIN: A Technology Platform for Identifying Knowns and Unknowns*
Metabolite Searching
METLIN has multiple searching capabilities including single, batch, precursor ion, neutral loss, accurate mass, and fragment searches. The popular similarity search algorithm for unknown characterization, another METLIN search option, originated on METLIN in 2008.
Tandem Mass Spectrometry
METLIN represents the largest MS/MS collection of data with the database generated at multiple collision energies and in positive and negative ionization modes. The data is generated on multiple instrument types including SCIEX, Agilent, Bruker and Waters QTOF mass spectrometers.
Metabolites
Created in 2003, METLIN now includes over a million molecules ranging from lipids, steroids, plant & bacteria metabolites, small peptides, carbohydrates, exogenous drugs/metabolites, central carbon metabolites and toxicants. The metabolites and other small molecules have been individually analyzed to provide both empirical and in silico MS/MS data.


Home*
isoMETLIN
Simple Search
Advanced Search
Batch Search
Fragment Similarity Search
Neutral Loss Search
MS/MS Spectrum Match Search
MRM
Logout [TMPLLabUAB]
Mass
Enter Mass
Tolerance
30
PPM
Charge
Neutral
Positive
Negative
Adducts
M+H
M+NH4
M+Na
M+H-2H2O
M+H-H2O
M+K
M+ACN+H
M+ACN+Na
M+2Na-H
M+2H
M+3H
M+H+Na
M+2H+Na
M+2Na
M+2Na+H
M+Li
M+CH3OH+H
Peptides
Add Peptides to Search
Toxicants
Add Toxicants to Search
Search
Clear
METLIN
The original and most comprehensive MS/MS metabolite database
Metabolite Searching
METLIN has multiple searching capabilities including single, batch, precursor ion, neutral loss, accurate mass, and fragment searches. The popular similarity search algorithm for unknown characterization, another METLIN search option, originated on METLIN in 2008.
Tandem Mass Spectrometry
METLIN represents the largest MS/MS collection of data with the database generated at multiple collision energies and in positive and negative ionization modes. The data is generated on multiple instrument types including SCIEX, Agilent, Bruker and Waters QTOF mass spectrometers.
Metabolites
Created in 2003, METLIN now includes over a million molecules ranging from lipids, steroids, plant & bacteria metabolites, small peptides, carbohydrates, exogenous drugs/metabolites, central carbon metabolites and toxicants. The metabolites and other small molecules have been individually analyzed to provide both empirical and in silico MS/MS data.

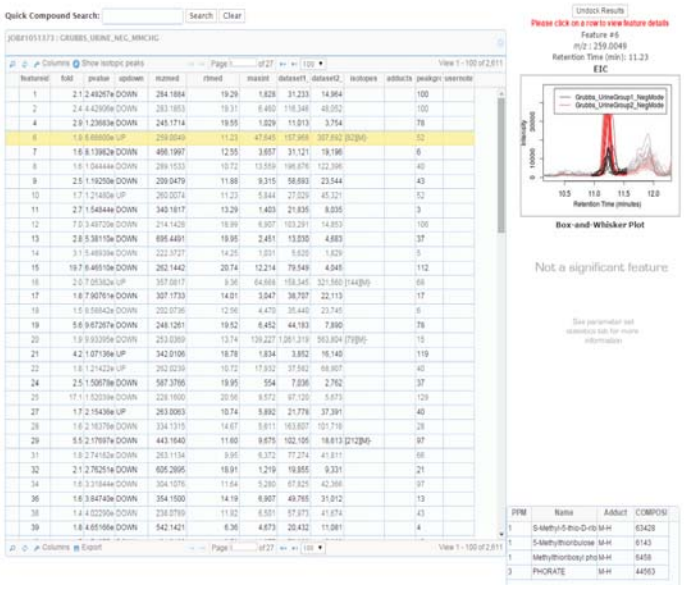


Not every peak represents individual metabolite: Adduct formation

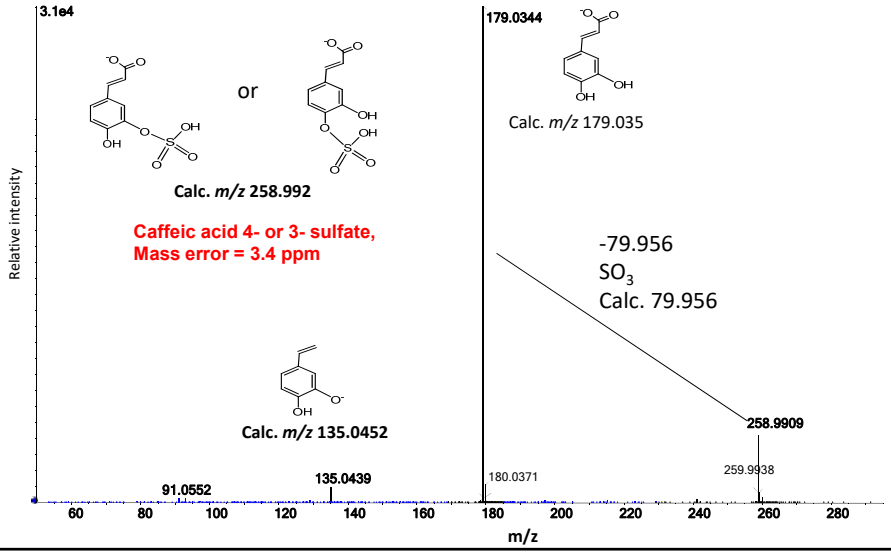


Nielsen et al., J Nat Prod. 2011

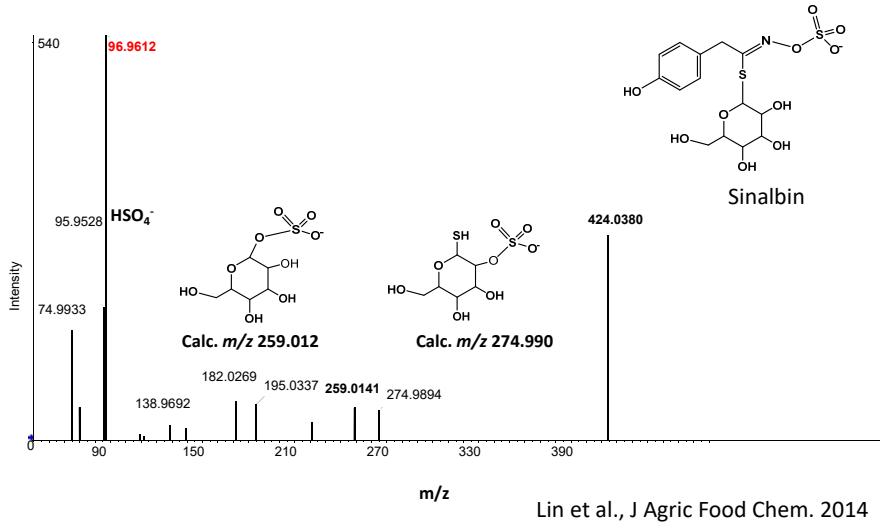
XCMS online platform to process untargeted metabolomic data



Validation of annotated metabolite by MS/MS interpretation

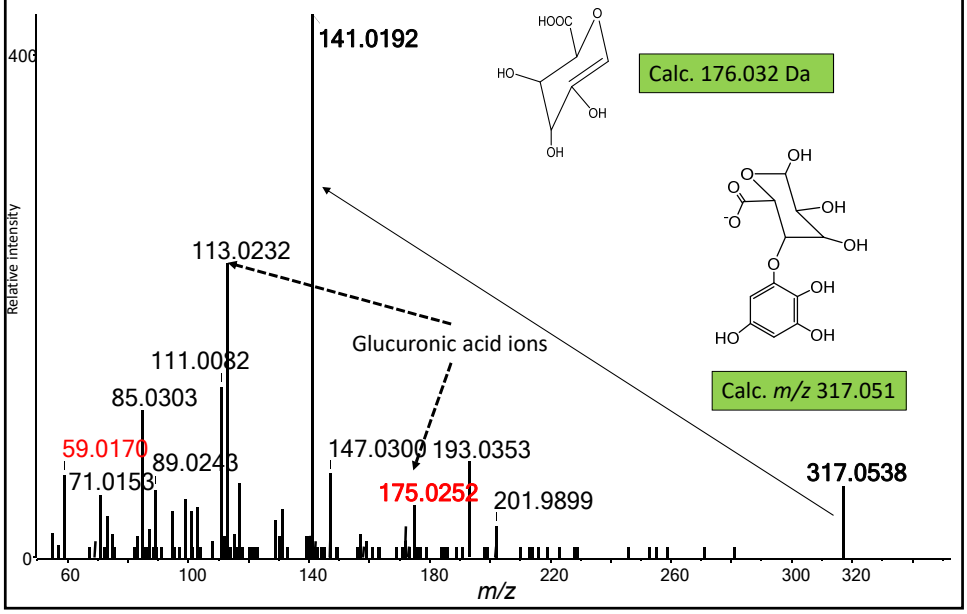


No experimental MS/MS of *m/z* 424.038, only *in silico* MS/MS data in positive ion mode available in Metlin, direct comparison not possible

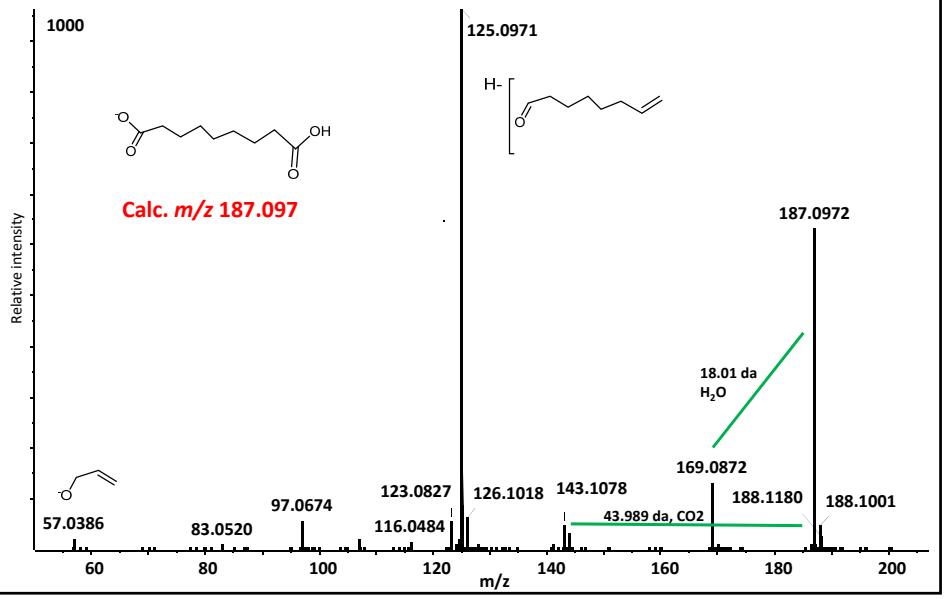


Lin et al., J Agric Food Chem. 2014

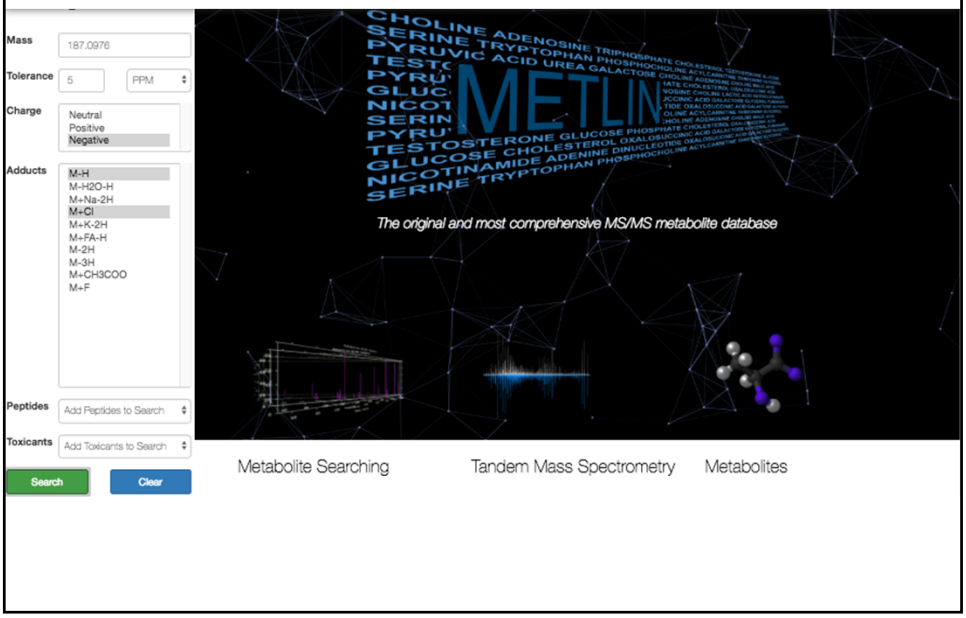
Neutral loss of monodehydrated glucuronic acid (calc. 176.032 Da)- an indicative of Glucuronidated Metabolite



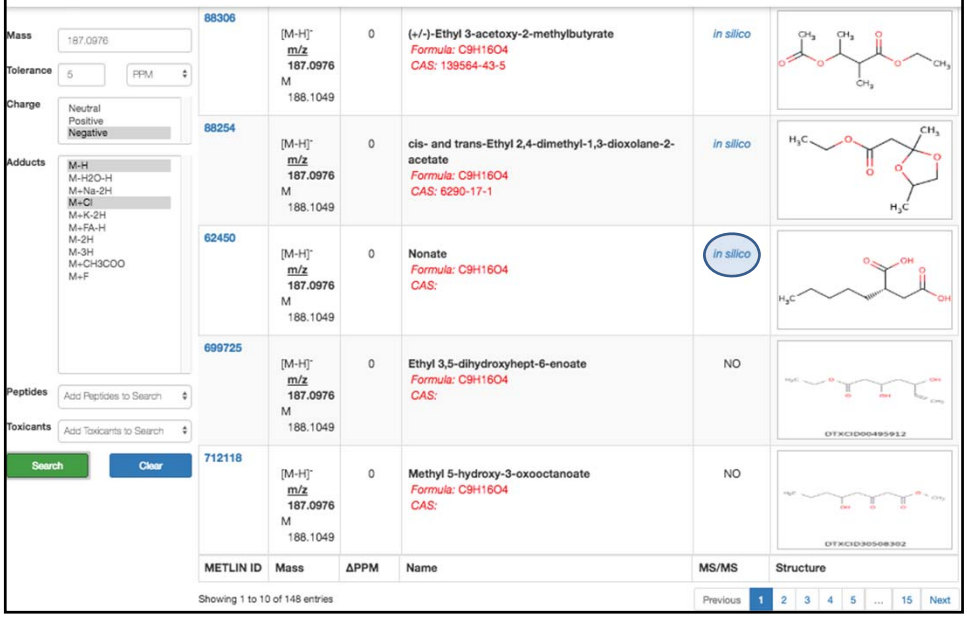
m/z 187.0976 identified as nonanedioic acid by comparing MS/MS profile between experimental and Metlin data base



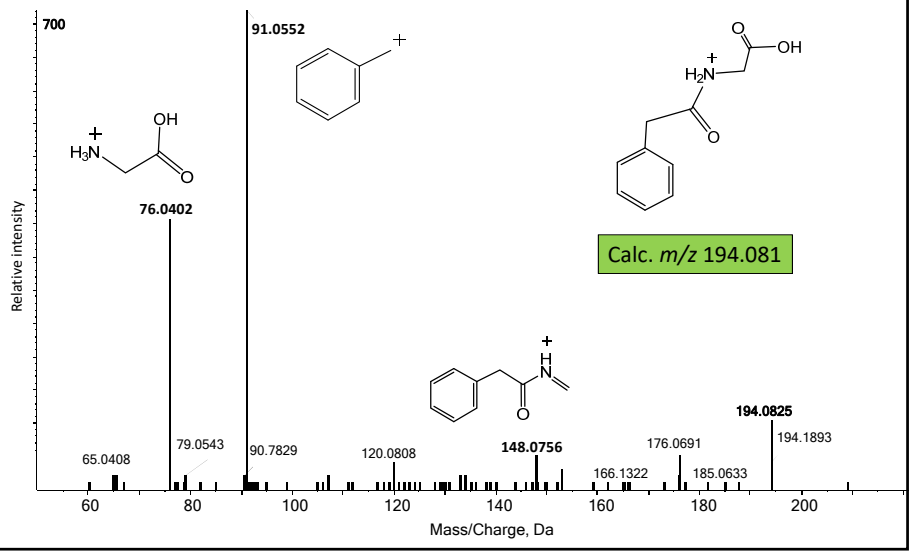
Searching METLIN for *m/z* 187.0976



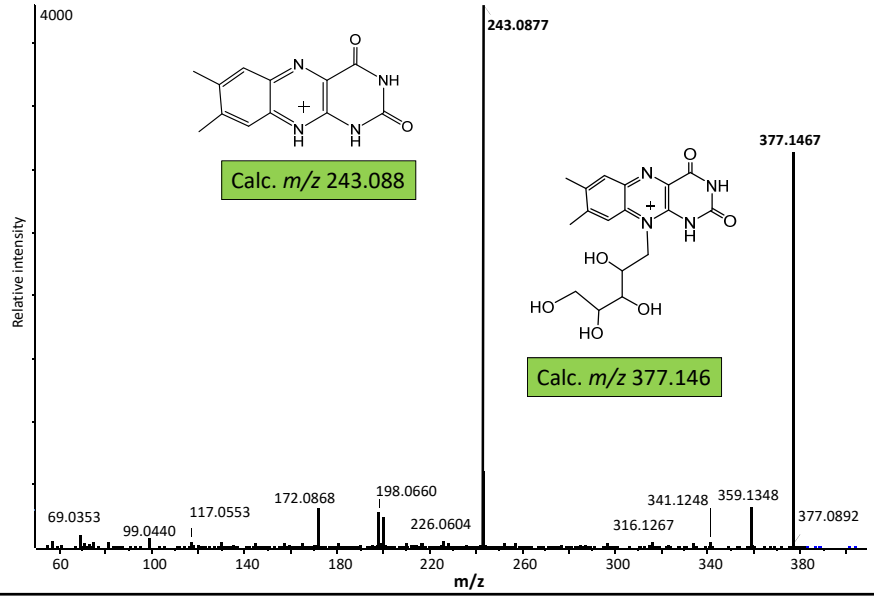
Output from Metlin



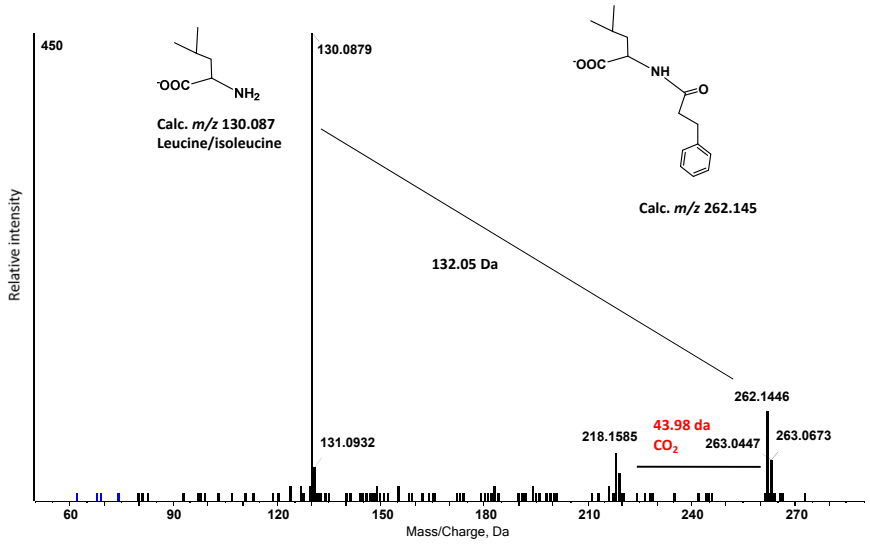
Among the annotated list of compounds by Metlin - phenylacetylglycine's validation by MS/MS interpretation



Detection of Vitamin B2 (riboflavin) as urinary metabolite-fragmentation patterns matched with Metlin database

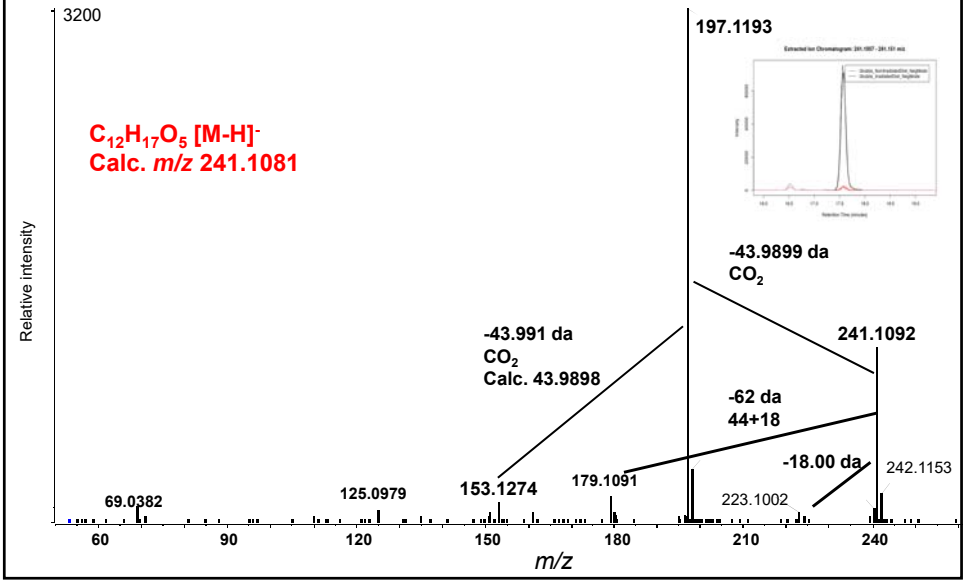


Fragment similarity search help propose structure of unknown metabolites

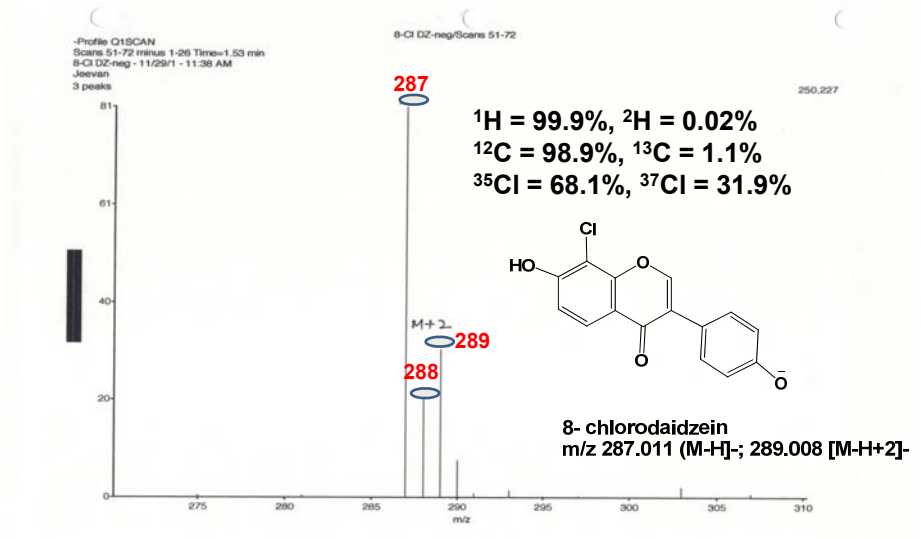


Many metabolites, unidentified by the Metlin database

A medium chain dicarboxylic fatty acid with *m/z* 241.109 [M-H]$^-$



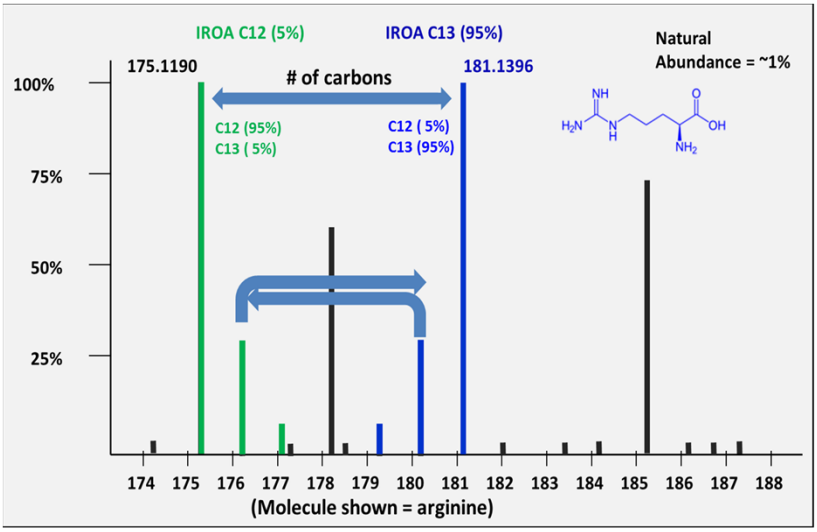
Isotopic pattern and intensity of ions indicates the number of carbons and hetero atoms in the molecular ion



Use of isotope pattern in identification of metabolites

- Very close in mass, but different in isotope patterns.
- Isotope ratio outlier analysis (IROA)
  - Used for LC-MS (and possibly GC-MS)
  - Designed to distinguish between metabolites of interest and background signals
  - Requires uniform labeling at the 95% and 5% ^{13}C-enrichment levels

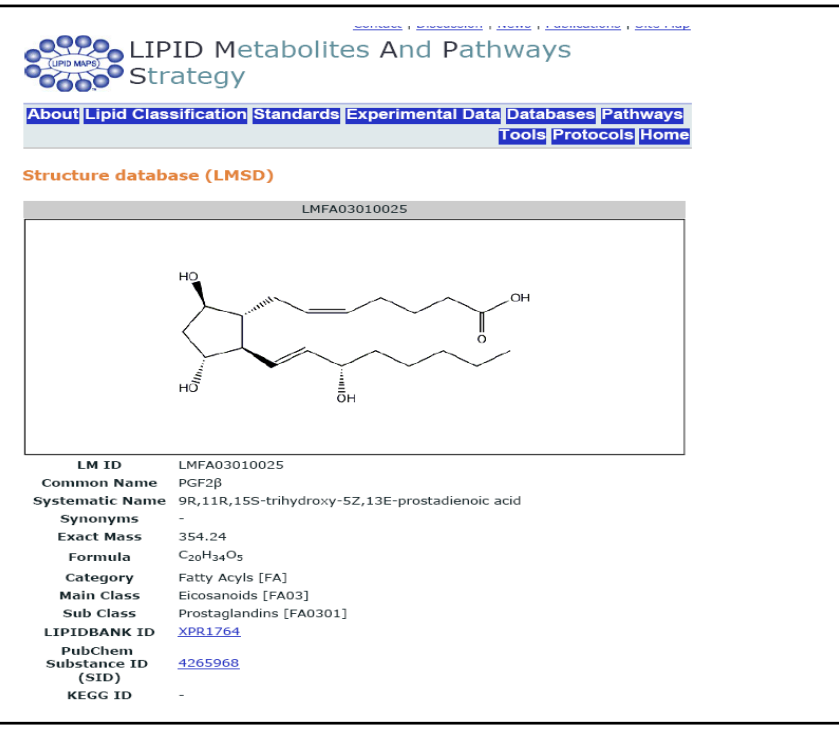
Pairing the 5% and 95% ^{13}C-labeling distinguishes artifactual molecules



Courtesy of Dr. Chris Beecher

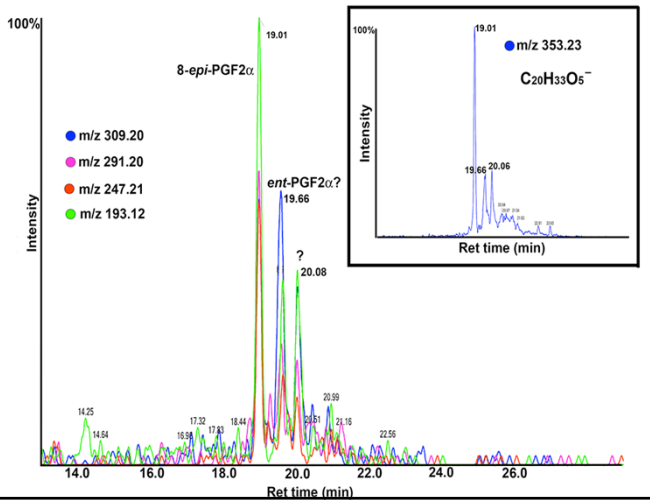
Library search for eicosanoid http://www.lipidmaps.org/

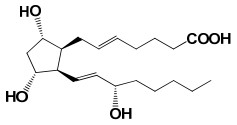
LM_ID	Common Name	Systematic Name	Formula	Mass
LMFA03000001	8(9)-EpETE	(+/-)-8(9)-epoxy-5Z,11Z,14Z,17Z-eicosatetraenoic acid	$C_{20}H_{30}O_3$	318.22
LMFA03000002	11(12)-EpETE	(+/-)-11(12)-epoxy-5Z,8Z,14Z,17Z-eicosatetraenoic acid	$C_{20}H_{30}O_3$	318.22
LMFA03000003	14(15)-EpETE	(+/-)-14(15)-epoxy-5Z,8Z,11Z,17Z-eicosatetraenoic acid	$C_{20}H_{30}O_3$	318.22
LMFA03000004	17(18)-EpETE	(+/-)-17(18)-epoxy-5Z,8Z,11Z,14Z-eicosatetraenoic acid	$C_{20}H_{30}O_3$	318.22
LMFA03000005	11(R)-HEDE	11R-hydroxy-12E,14Z-eicosadienoic acid	$C_{20}H_{36}O_3$	324.27
LMFA03000006	17R,18S-EpETE	17R,18S-epoxy-5Z,8Z,11Z,14Z-eicosatetraenoic acid	$C_{20}H_{30}O_3$	318.22
LMFA03000008	15(R)-HEDE	15R-hydroxy-11Z-13E-eicosadienoic acid	$C_{20}H_{36}O_3$	324.27
LMFA03000009	11S-HEDE	11S-hydroxy-12E,14Z-eicosadienoic acid	$C_{20}H_{36}O_3$	324.27
LMFA03010000	Prostanoic acid skeleton	-	-	-
LMFA03010001	6-keto-PGF1a	6-oxo-9S,11R,15S-trihydroxy-13E-prostenoic acid	$C_{20}H_{34}O_6$	370.24
LMFA03010002	PGF2a	9S,11R,15S-trihydroxy-5Z,13E-prostadienoic acid	$C_{20}H_{34}O_5$	354.24
LMFA03010003	PGE2 (W)	9-oxo-11R,15S-dihydroxy-5Z,13E-prostadienoic acid	$C_{20}H_{32}O_5$	352.22
LMFA03010004	PGD2 (W)	9S,15S-dihydroxy-11-oxo-5Z,13E-prostadienoic acid	$C_{20}H_{32}O_5$	352.22
LMFA03010005	PGA1	9-oxo-15S-hydroxy-10Z,13E-prostadienoic acid	$C_{20}H_{32}O_4$	336.23
LMFA03010006	PGF2a-d4	9S,11R,15S-trihydroxy-5Z,13E-prostadienoic acid (3,3,4,4-d4)	$C_{20}H_{30}D_4O_5$	358.27
LMFA03010007	PGD2-d4	9S,15S-dihydroxy-11-oxo-5Z,13E-prostadienoic acid (3,3,4,4-d4)	$C_{20}H_{28}D_4O_5$	356.25
LMFA03010008	PGE2-d4	11R,15S-dihydroxy-9-oxo-5Z,13E-prostadienoic acid (3,3,4,4-d4)	$C_{20}H_{28}D_4O_5$	356.25
LMFA03010009	PGG2	9S,11R-epidioxy-15S-hydroperoxy-5Z,13E-	$C_{20}H_{32}O_6$	368.22



LM ID	LMFA03010025
Common Name	PGF2β
Systematic Name	9R,11R,15S-trihydroxy-5Z,13E-prostadienoic acid
Synonyms	-
Exact Mass	354.24
Formula	$C_{20}H_{34}O_5$
Category	Fatty Acyls [FA]
Main Class	Eicosanoids [FA03]
Sub Class	Prostaglandins [FA0301]
LIPIDBANK ID	XPR1764
PubChem Substance ID (SID)	4265968
KEGG ID	-

Good chromatographic separation and accurate mass are the important steps in structure determination

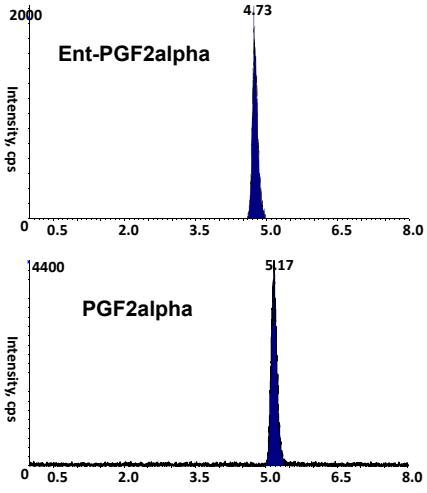


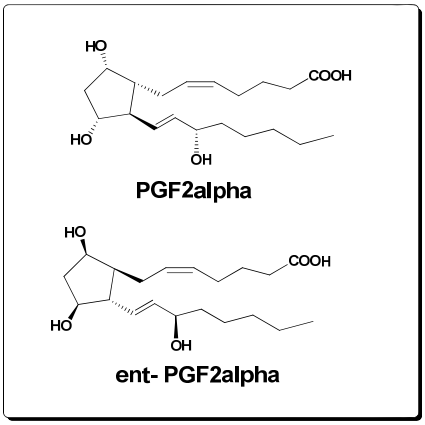


8-iso PGF2alpha

Exact mass: 353.2333

Separation of stereoisomers by a chiral normal phase column

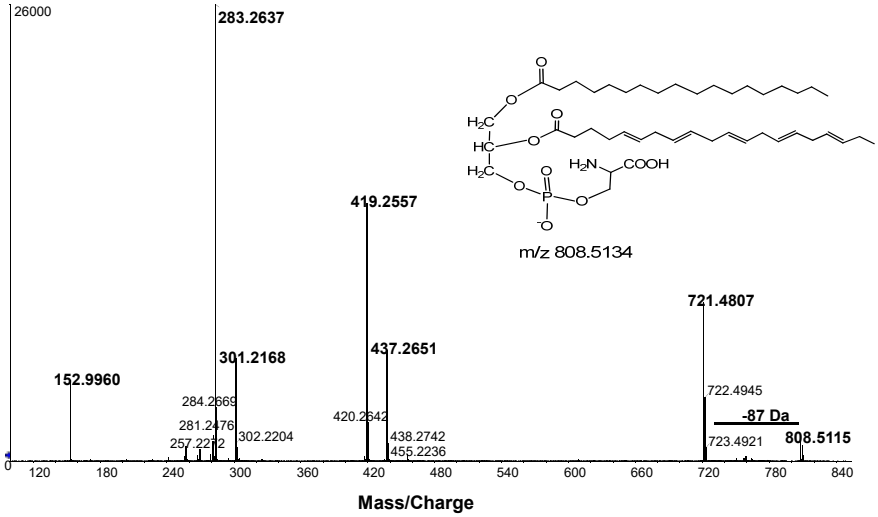




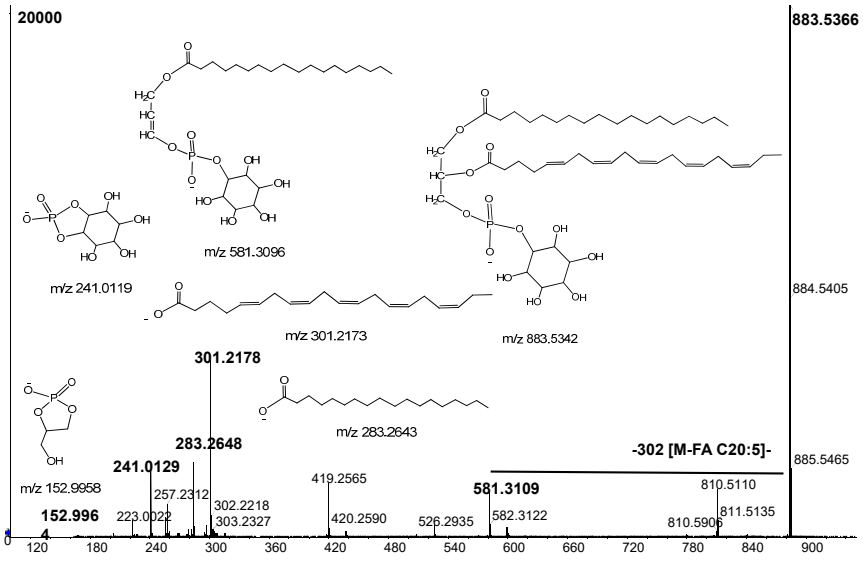
Hoang et al., PLOS Genetics. 2013

Nitrogen rule

Odd number of nitrogens = odd MW, even *m/z*
No nitrogen or even nitrogens = even MW, odd *m/z*

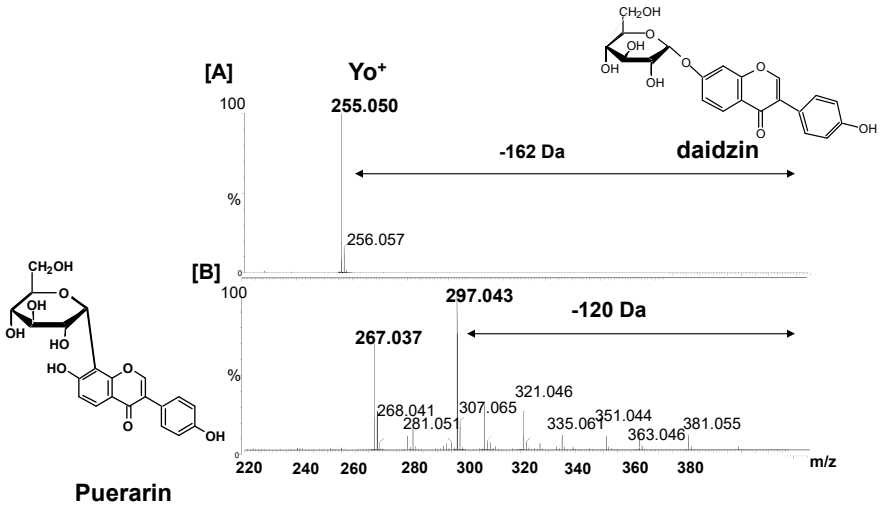


Accurate mass (<5 ppm), fragmentation patters help propose putative structures



Prasain et al., Metabolites 2015

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



Prasain et al., J. Agric. Food Chem., 2003

Conclusions

- **Identifying unknown metabolites is a significant analytical challenge in metabolomics and it requires integrated analytical and bio-informative approaches.**
- **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.**