

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

Jeevan K. Prasain, PhD  
Department of Pharmacology and  
Toxicology, UAB  
[jprasain@uab.edu](mailto:jprasain@uab.edu)

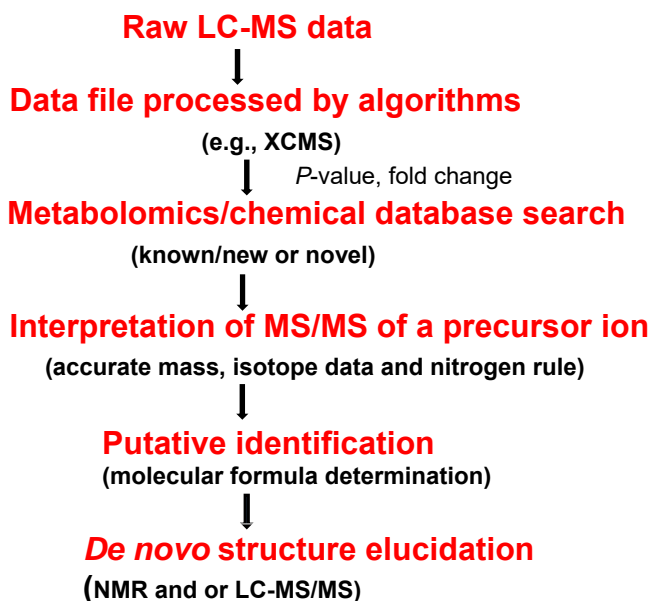
## Outline

- 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



## 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)

Moco et al. Trends in Analytical Chemistry, 2007

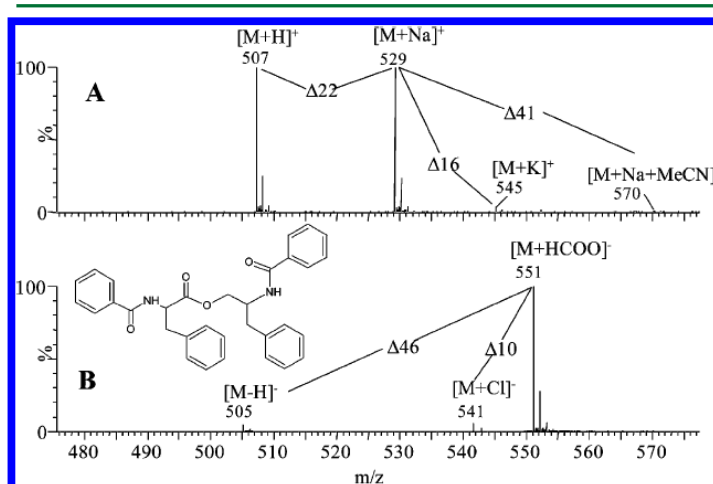
## LCMS-based metabolomics

- 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

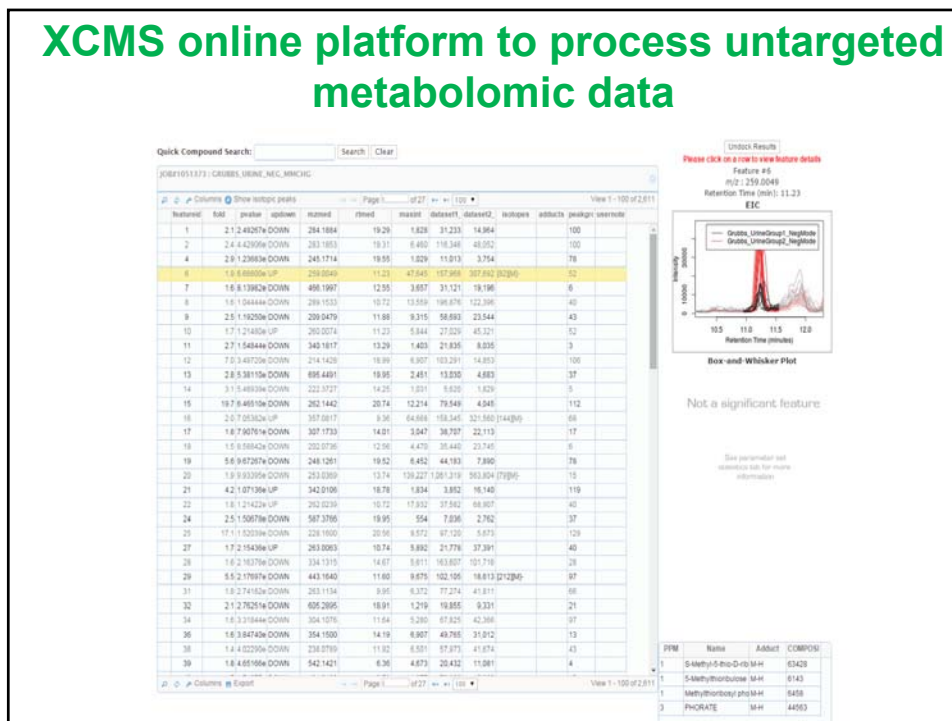
- XCMS (developed by the Siuzdak Lab at the Scripps Research Institute) Online, is a web-based version that allows users to easily upload and process LC-MS data. It is a bioinformatics platform to identify endogenous metabolites..
- METLIN (<http://metlin.scripps.edu>) is a metabolite database for metabolomics containing 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

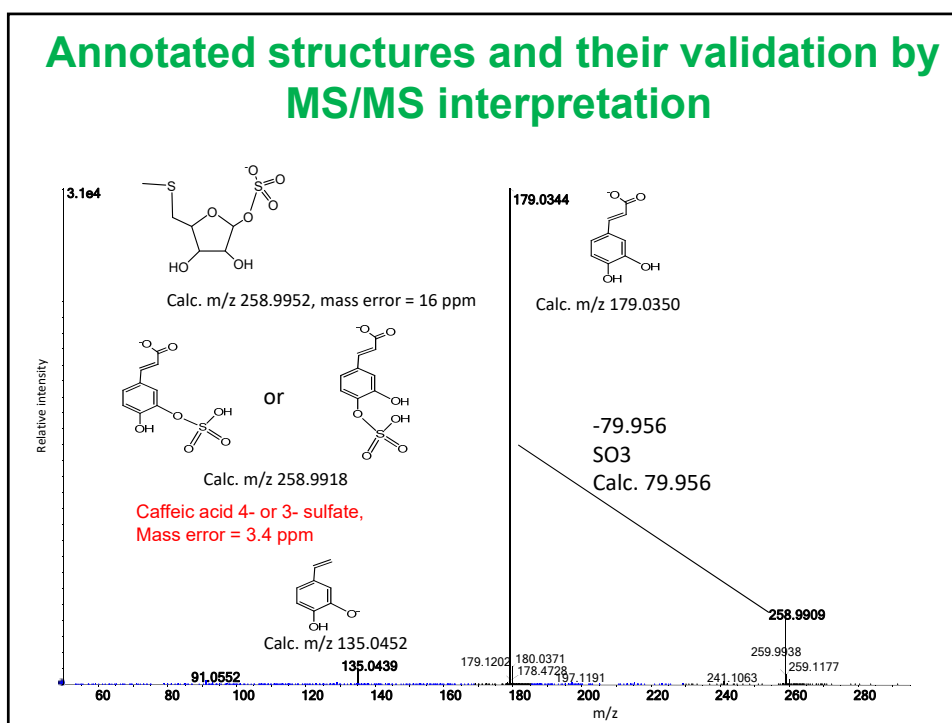


Nielsen et al., J Nat Prod. 2011

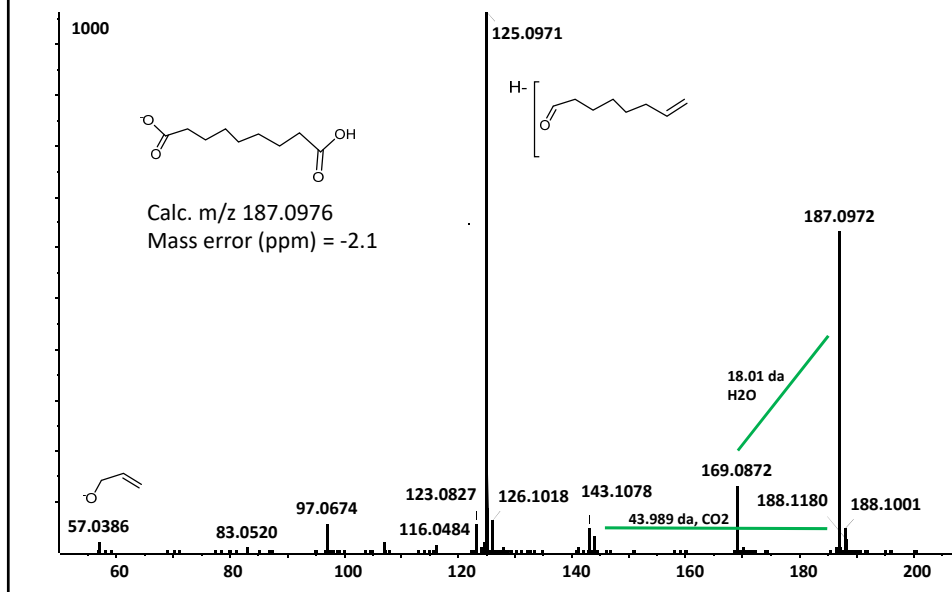
## XCMS online platform to process untargeted metabolomic data



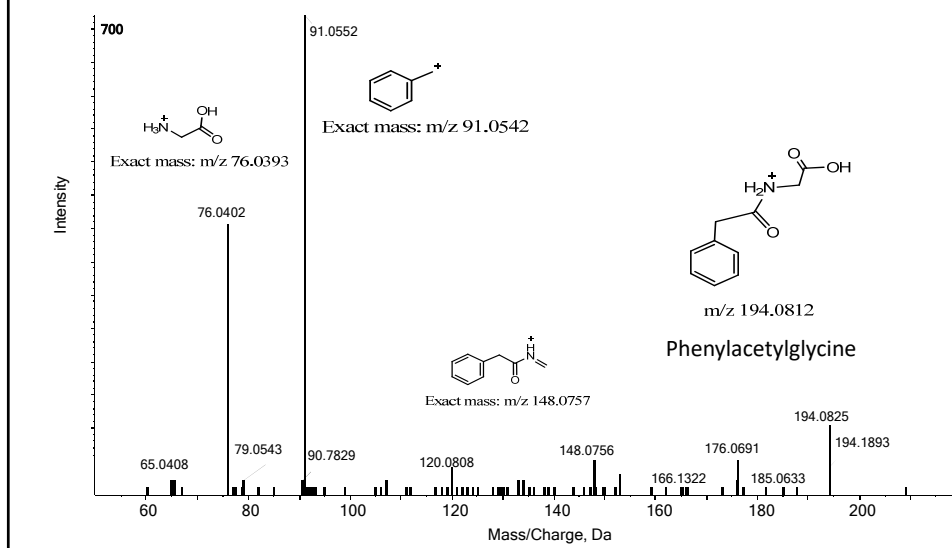
## Annotated structures and their validation by MS/MS interpretation



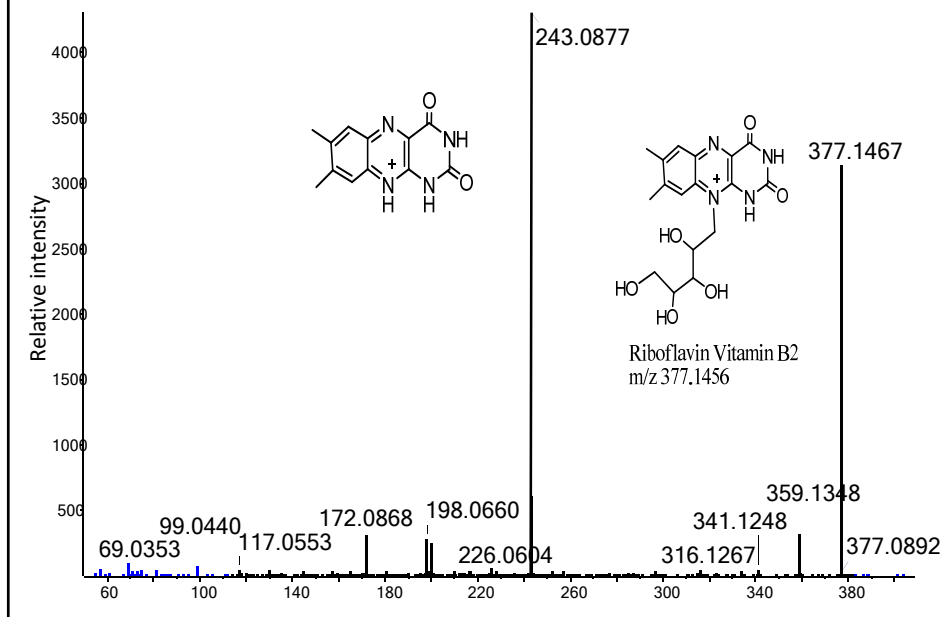
**M/z 187.0976 was identified as nonanedioic acid by comparing MS/MS profile between experimental and Metline data base**



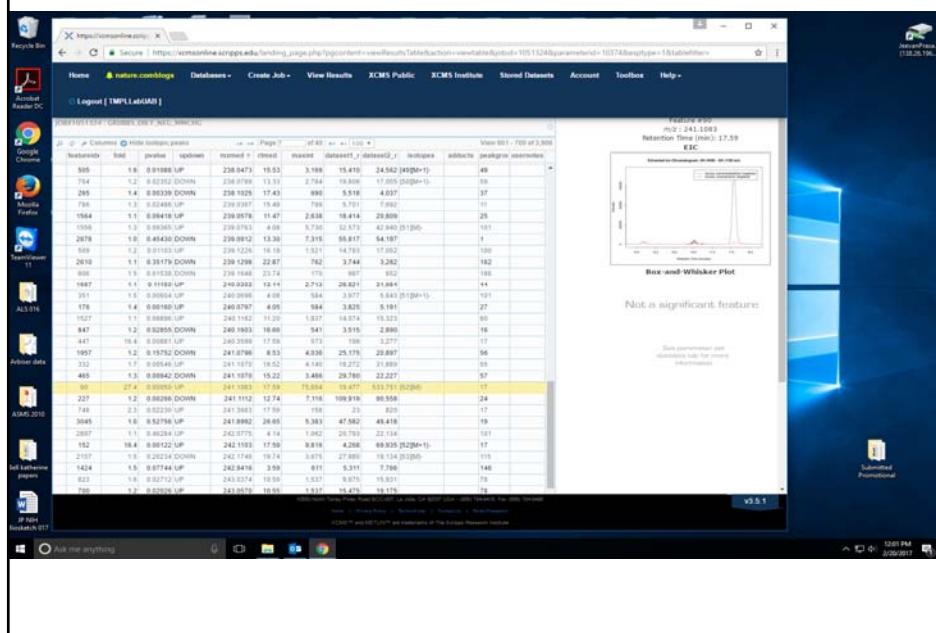
**Among the annotated list of compounds by Metlin- phenylacetyl-glycine's validation by MS/MS interpretation**

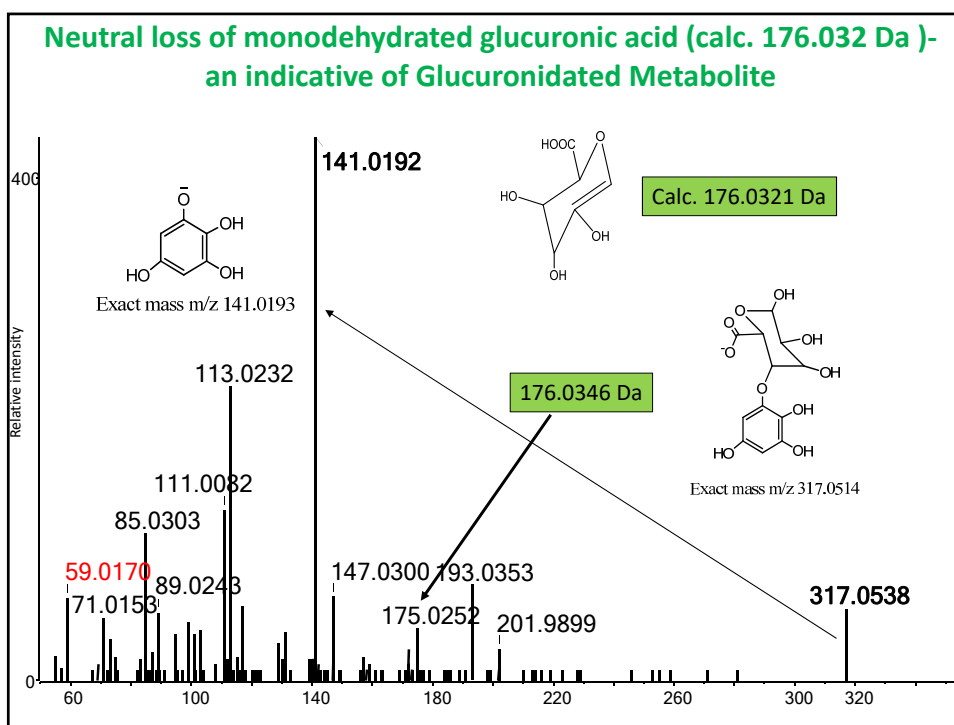
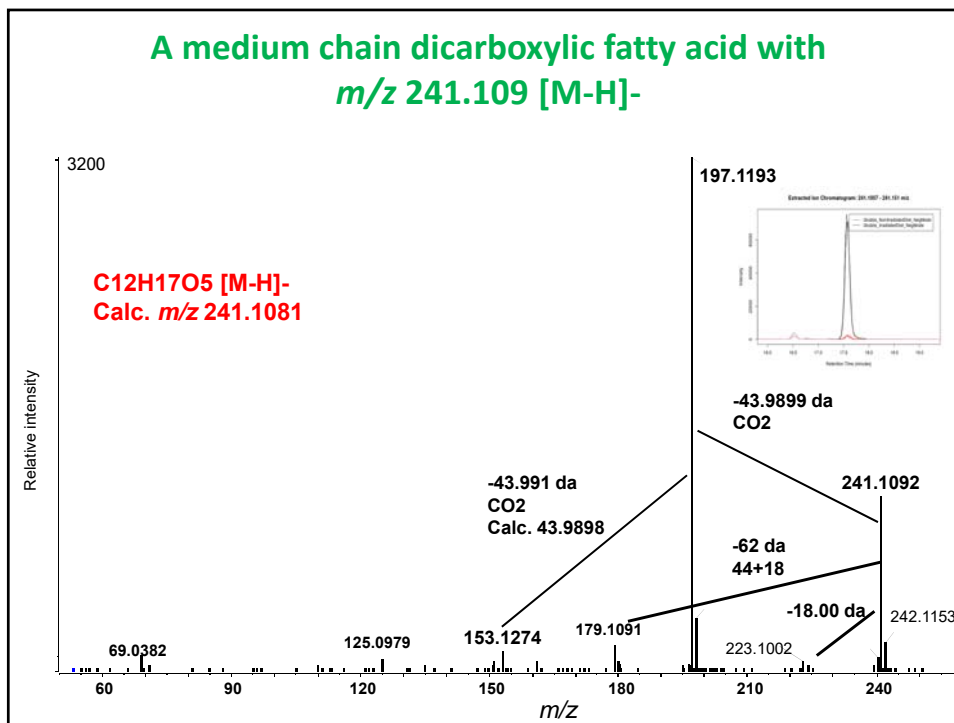


## Detection of Vitamin B2 (riboflavin) as urinary metabolite-fragmentation patterns matched with Metlin database



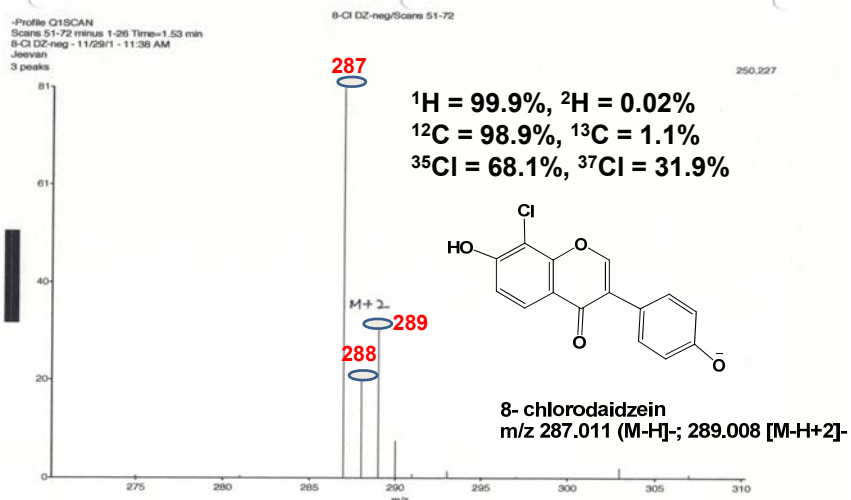
## Many metabolites, unidentified by the Metlin database







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



## Library search for eicosanoid <http://www.lipidmaps.org/>

LIPID MAPS -- LIPID Metabolites And Pathways Strategy

[Contact](#) | [Discussion](#) | [News](#) | [Publications](#) | [Site Map](#)



### LIPID Metabolites And Pathways Strategy

[About](#) | [Lipid Classification](#) | [Standards](#) | [Experimental Data](#)  
[Databases](#) | [Pathways](#) | [Tools](#) | [Protocols](#) | [Home](#)

#### LMSD: Lipid classification search results

Fatty Acyls [FA] (W) --&gt; Eicosanoids [FA03]

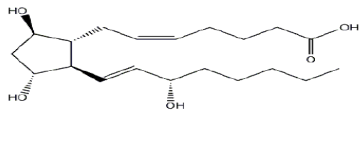
LM_ID	Common Name	Systematic Name	Formula	Mass
LMFA03000001	8(9)-EpETE	(+/-)-8(9)-epoxy-5Z,11Z,14Z,17Z-eicosatetraenoic acid	C <sub>20</sub> H <sub>30</sub> O <sub>2</sub>	316.22
LMFA03000002	11(12)-EpETE	(+/-)-11(12)-epoxy-5Z,8Z,14Z,17Z-eicosatetraenoic acid	C <sub>20</sub> H <sub>30</sub> O <sub>2</sub>	316.22
LMFA03000003	14(15)-EpETE	(+/-)-14(15)-epoxy-5Z,8Z,11Z,17Z-eicosatetraenoic acid	C <sub>20</sub> H <sub>30</sub> O <sub>2</sub>	316.22
LMFA03000004	17(18)-EpETE	(+/-)-17(18)-epoxy-5Z,8Z,11Z,14Z-eicosatetraenoic acid	C <sub>20</sub> H <sub>30</sub> O <sub>2</sub>	316.22
LMFA03000005	11(R)-HEDE	11R-hydroxy-12E,14Z-eicosadienoic acid	C <sub>20</sub> H <sub>30</sub> O <sub>2</sub>	324.27
LMFA03000006	17R,10S-EpETE	17R,10S-epoxy-5Z,6Z,11Z,14Z-eicosatetraenoic acid	C <sub>20</sub> H <sub>30</sub> O <sub>2</sub>	316.22
LMFA03000008	15(R)-HEDE	15R-hydroxy-11Z,13E-eicosadienoic acid	C <sub>20</sub> H <sub>30</sub> O <sub>2</sub>	324.27
LMFA03000009	11S-HEDE	11S-hydroxy-12E,14Z-eicosadienoic acid	C <sub>20</sub> H <sub>30</sub> O <sub>2</sub>	324.27
LMFA03010000	Prostanoic acid skeleton	-	-	-
LMFA03010001	6-keto-PGF1α	6-oxo-9S,11R,15S-trihydroxy-13E-prostanoic acid	C <sub>20</sub> H <sub>34</sub> O <sub>6</sub>	370.24
LMFA03010002	PGF2α	9S,11R,15S-trihydroxy-5Z,13E-prostadienoic acid	C <sub>20</sub> H <sub>34</sub> O <sub>6</sub>	354.24
LMFA03010003	PGE2 (W)	9-oxo-11R,15S-dihydroxy-5Z,13E-prostadienoic acid	C <sub>20</sub> H <sub>32</sub> O <sub>6</sub>	352.22
LMFA03010004	PGD2 (W)	9S,15S-dihydroxy-11-oxo-5Z,13E-prostadienoic acid	C <sub>20</sub> H <sub>32</sub> O <sub>6</sub>	352.22
LMFA03010005	PGA1	9-oxo-15S-hydroxy-10Z,13E-prostadienoic acid	C <sub>20</sub> H <sub>32</sub> O <sub>6</sub>	352.22
LMFA03010006	PGF2α-d4	9S,11R,15S-trihydroxy-5Z,13E-prostadienoic acid (3,3,4,4-d4)	C <sub>20</sub> H <sub>30</sub> D <sub>4</sub> O <sub>6</sub>	356.27
LMFA03010007	PGD2-d4	9S,15S-dihydroxy-11-oxo-5Z,13E-prostadienoic acid (3,3,4,4-d4)	C <sub>20</sub> H <sub>28</sub> D <sub>4</sub> O <sub>6</sub>	356.25
LMFA03010008	PGE2-d4	11R,15S-dihydroxy-9-oxo-5Z,13E-prostadienoic acid (3,3,4,4-d4)	C <sub>20</sub> H <sub>28</sub> D <sub>4</sub> O <sub>6</sub>	356.25
LMFA03010009	PGG2	9S,11R-epidoxo-15S-hydroperoxy-5Z,13E-prostadienoic acid	C <sub>20</sub> H <sub>32</sub> O <sub>6</sub>	366.22


**LIPID Metabolites And Pathways Strategy**

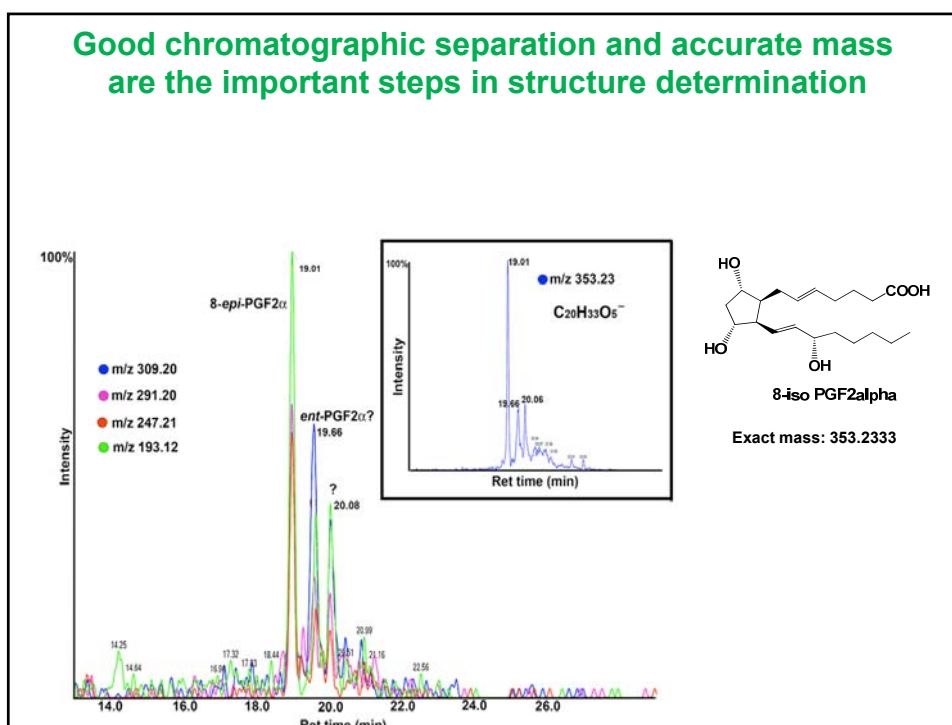
[About](#) | [Lipid Classification](#) | [Standards](#) | [Experimental Data](#) | [Databases](#) | [Pathways](#) | [Tools](#) | [Protocols](#) | [Home](#)

**Structure database (LMSD)**

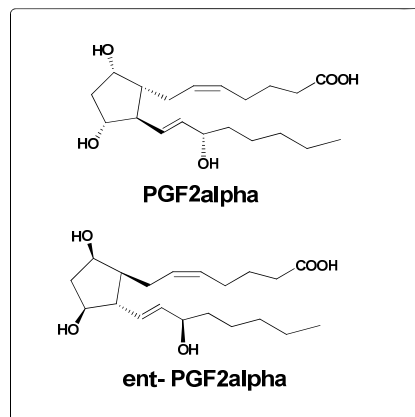
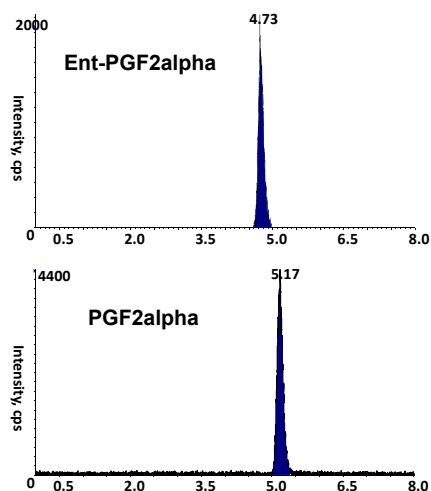
LMFA03010025



<b>LM ID</b>	LMFA03010025
<b>Common Name</b>	PGF2β
<b>Systematic Name</b>	9R,11R,15S-trihydroxy-5Z,13E-prostadienoic acid
<b>Synonyms</b>	-
<b>Exact Mass</b>	354.24
<b>Formula</b>	C <sub>20</sub> H <sub>34</sub> O <sub>5</sub>
<b>Category</b>	Fatty Acyls [FA]
<b>Main Class</b>	Eicosanoids [FA03]
<b>Sub Class</b>	Prostaglandins [FA0301]
<b>LIPIDBANK ID</b>	<a href="#">XPR1764</a>
<b>PubChem Substance ID (SID)</b>	<a href="#">4265968</a>
<b>KEGG ID</b>	-

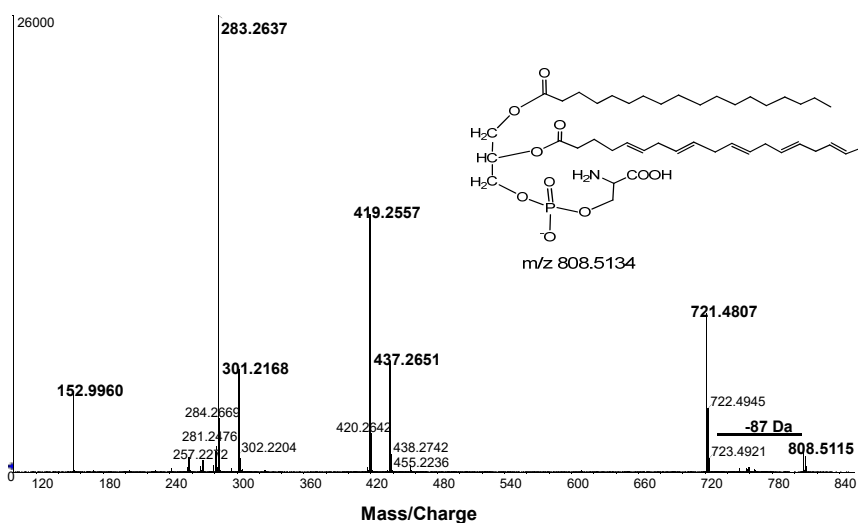


## Separation of stereoisomers by a chiral normal phase column

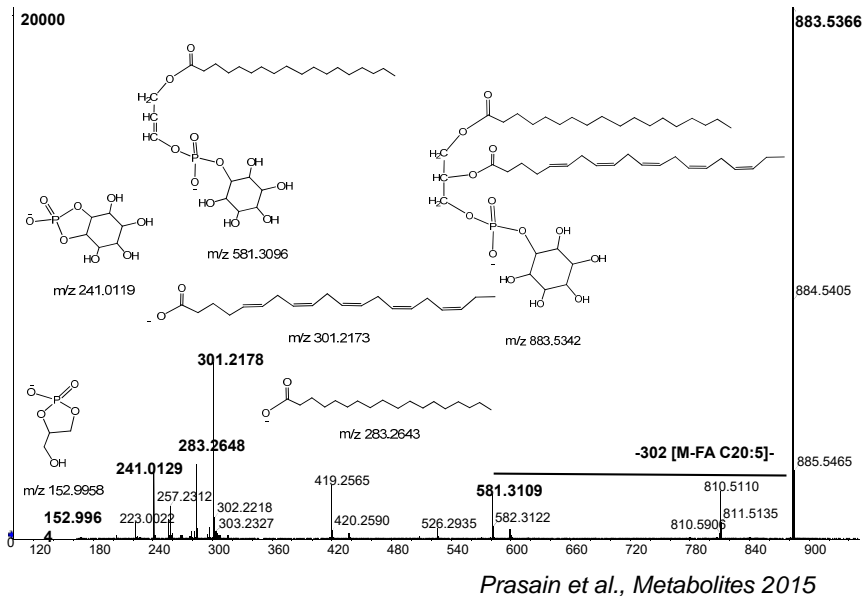


Hoang et al., PLOS Genetics. 2013

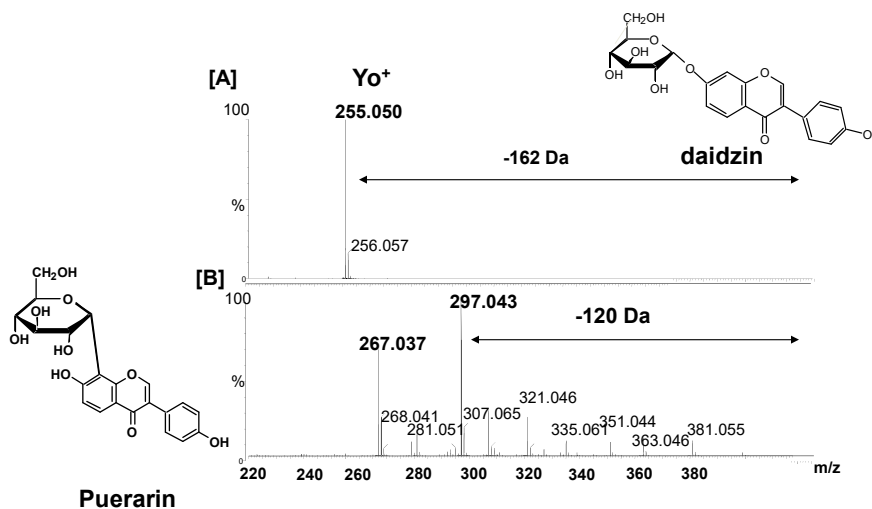
## Nitrogen rule- Odd number of nitrogens = odd MW No nitrogen or even nitrogens = even MW



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



## 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<sup>n</sup> 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.**