Identifying Unknowns: A challenge for Metabolomics

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

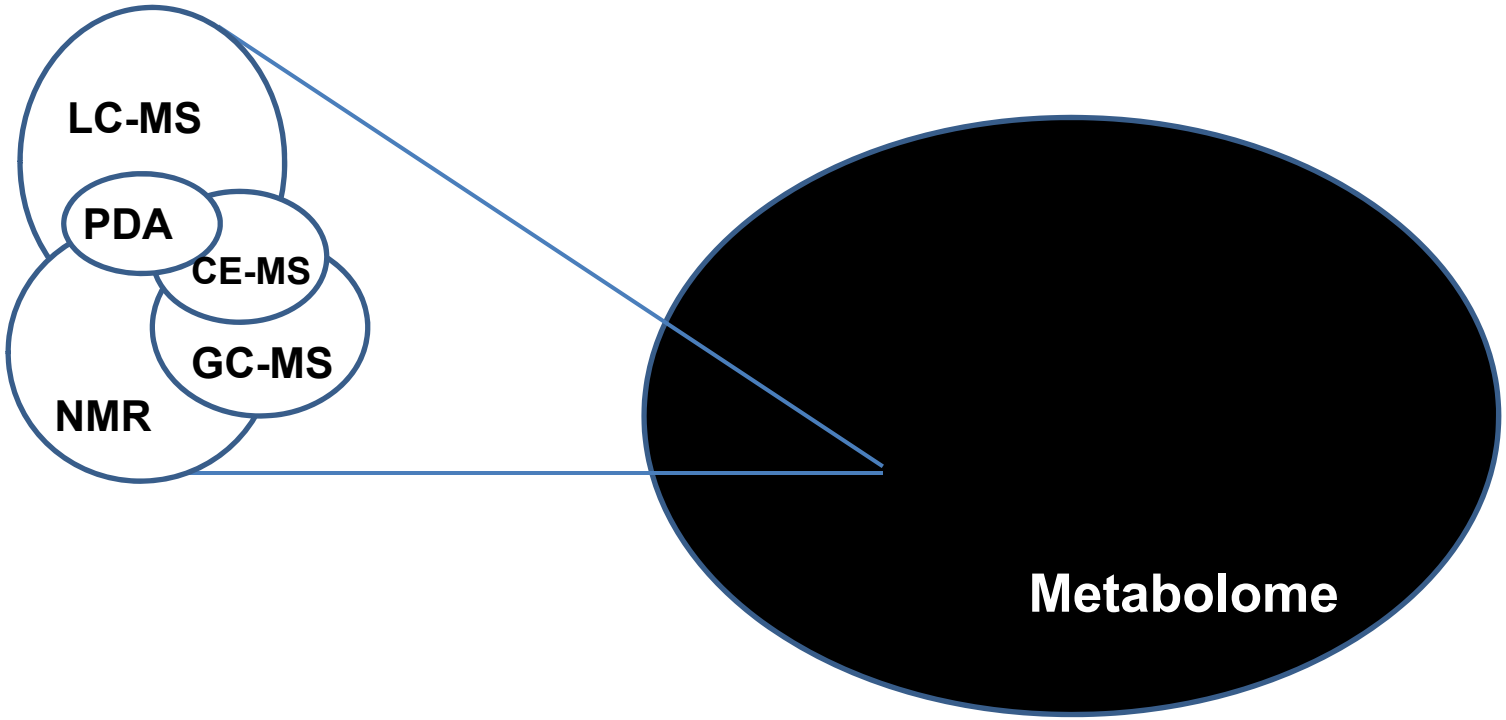
Outline

- Introduction
- How to interpret LC-MS and MS/MS data.
- Identification of some conjugated metabolites.
- Conclusions

Introduction

- **Identification of metabolites (lipids or any other metabolites) at a molecular level presents a great challenge 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**
- **Inclusion of many artifacts in database.**
- **Structural complexity of metabolites.**
- **Low concentrations and difficult to isolate.**

Majority of metabolites are yet to be identified; LC-MS and NMR are the most commonly used analytical techniques for metabolite identification

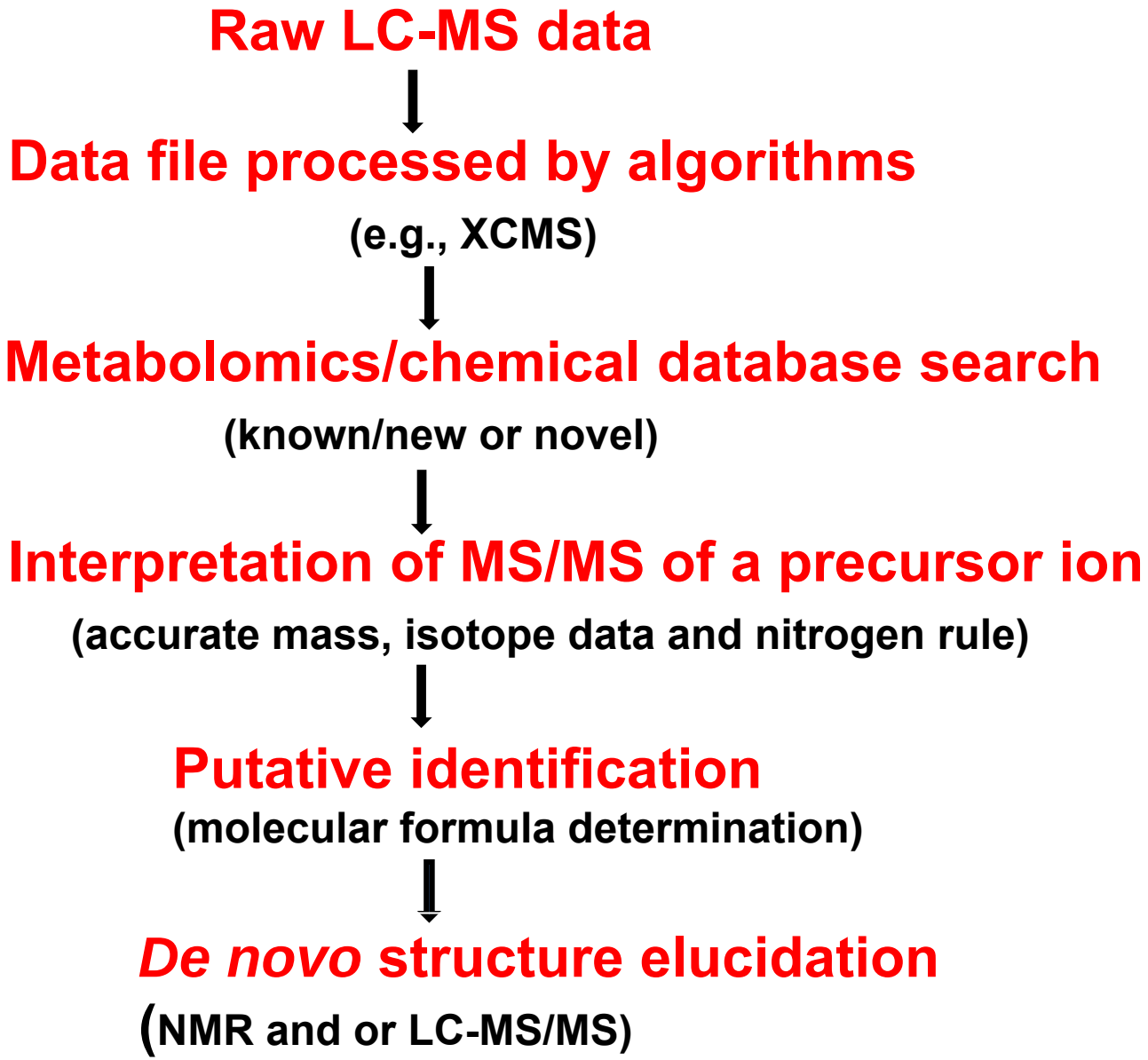


Adapted from Moco et al., Trend in Analytical Chemistry, 2007

Keys to identifying chemical structures by mass spectrometry

- **Combination of the following:**
  - **Retention time in LC**
  - **Accurate mass**
  - **Isotope distribution**
  - **MS/MS product ions of a precursor ion**

Metabolite identification workflow



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 provides links to METLIN database and produces a list of potential metabolites.**

- **METLIN (developed by the Siuzdak Lab.) 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.**

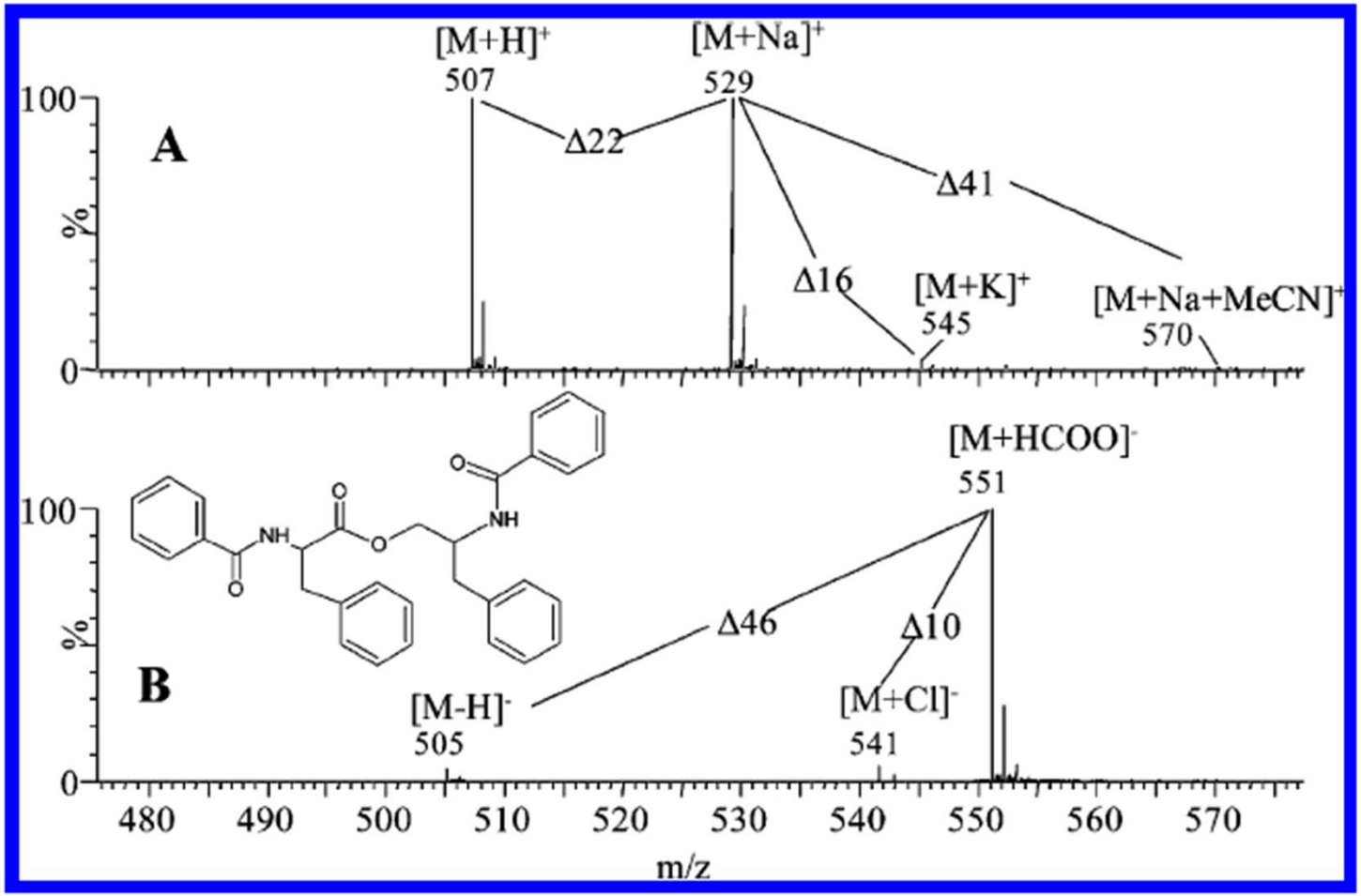
LCMS-based metabolomics

- 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

Points to be considered in LC-MS analysis

- Choice of ionization mode- ESI Vs APCI +ve/-ve modes
- Choice of eluting solvent- methanol Vs acetonitrile
- Additives/pH in mobile phase
- Molecular ion recognition (adduct formation)
- Chromatographic separation- stationary phase C8, C18 ..
- Evaluation of spectral quality- what to look for in a good quality spectra

Adduct formation might complicate metabolite identification

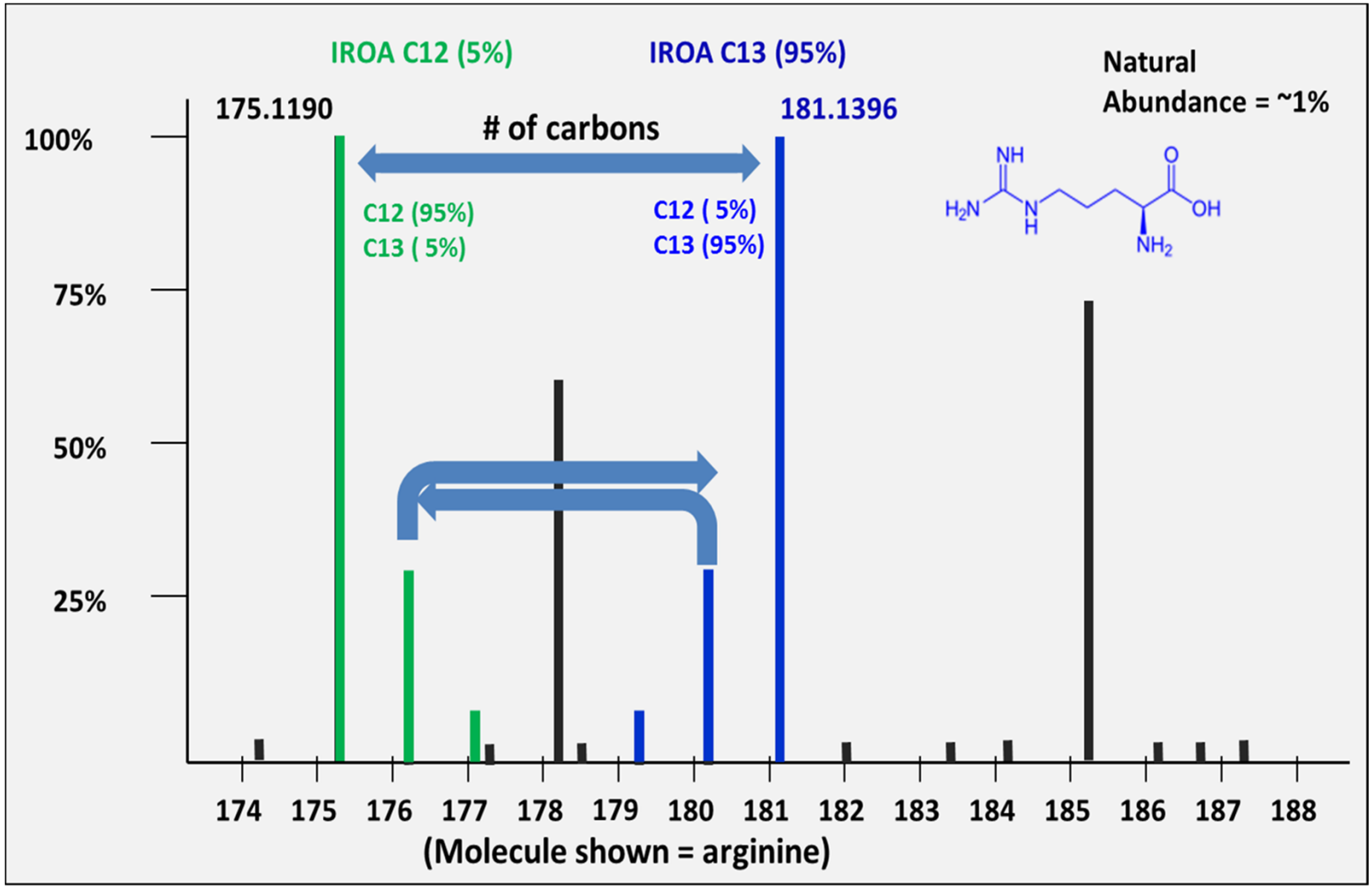


Nielsen et al., J Nat Prod. 2011

Use of isotope pattern in identification of metabolites

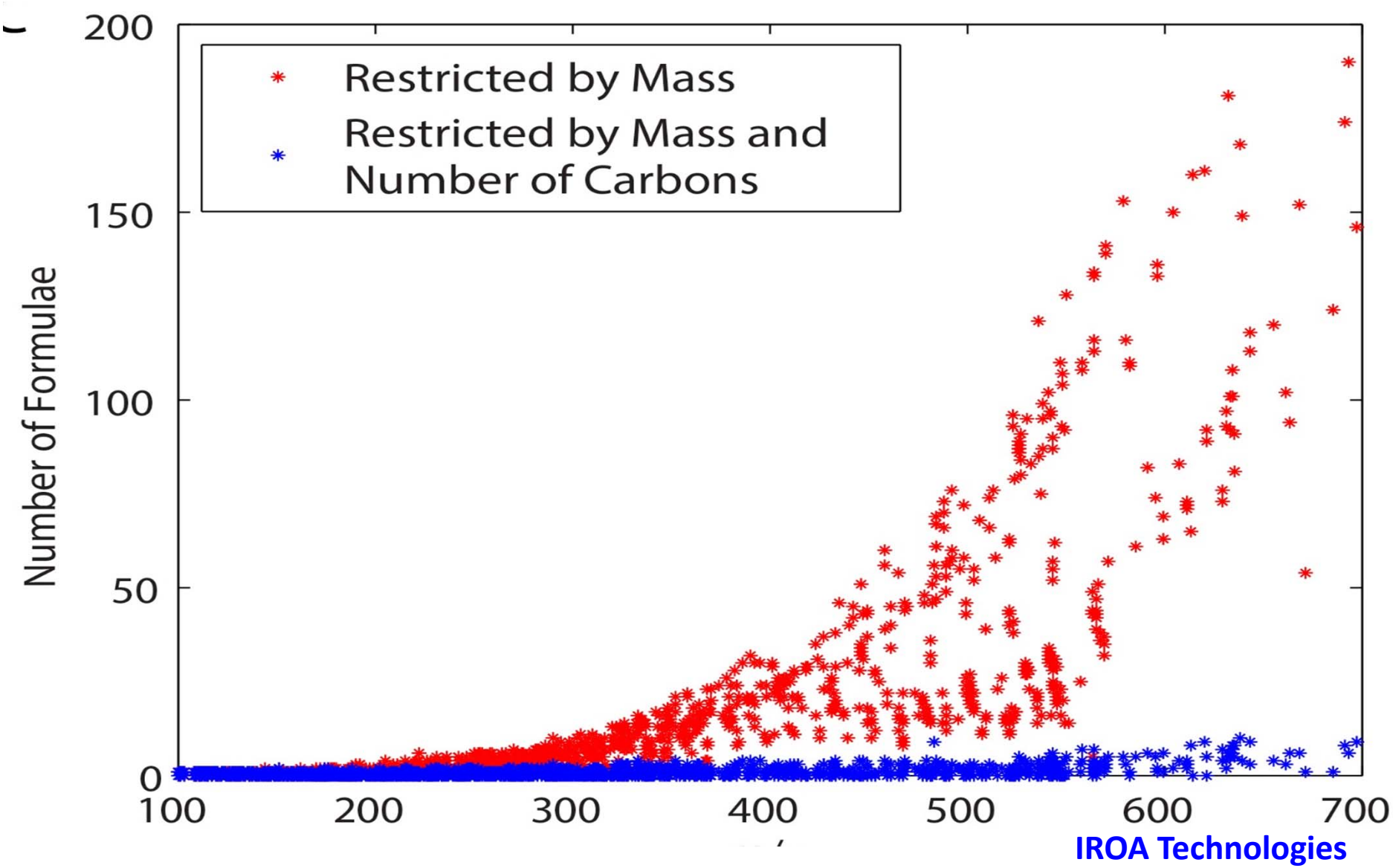
- **Very close in mass, but different in isotope patterns.**
- **Isotope ratio outlier analysis (IROA)**
  - **Used for LC-MS (and possibly GC-MS)**
  - **Designed to distinguish between metabolites of interest and background signals**
  - **Requires uniform labeling at the 95% and 5% ^{13}C-enrichment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

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



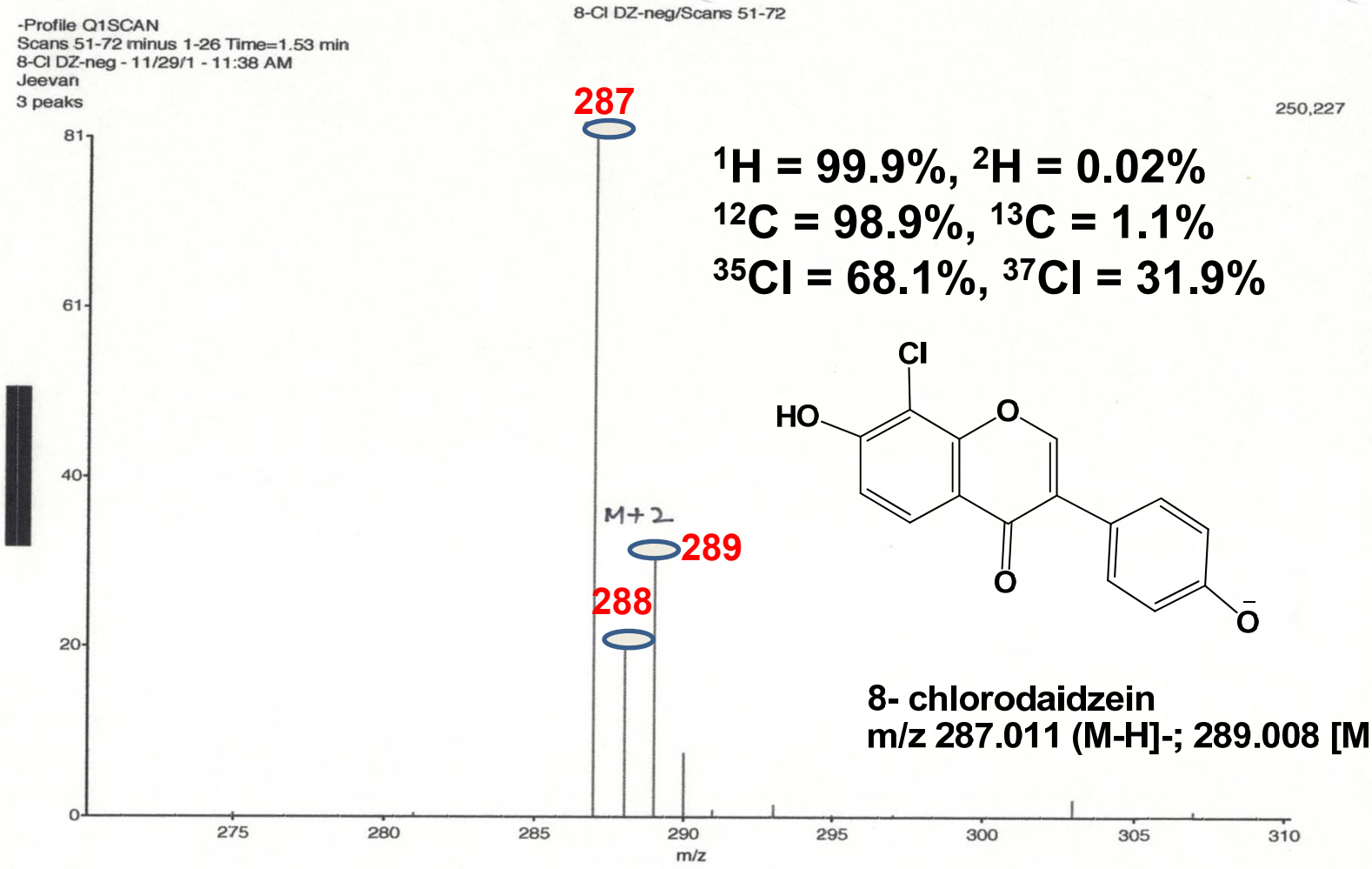
Courtesy of IROA Technologies

Value of knowing the carbon



IROA Technologies

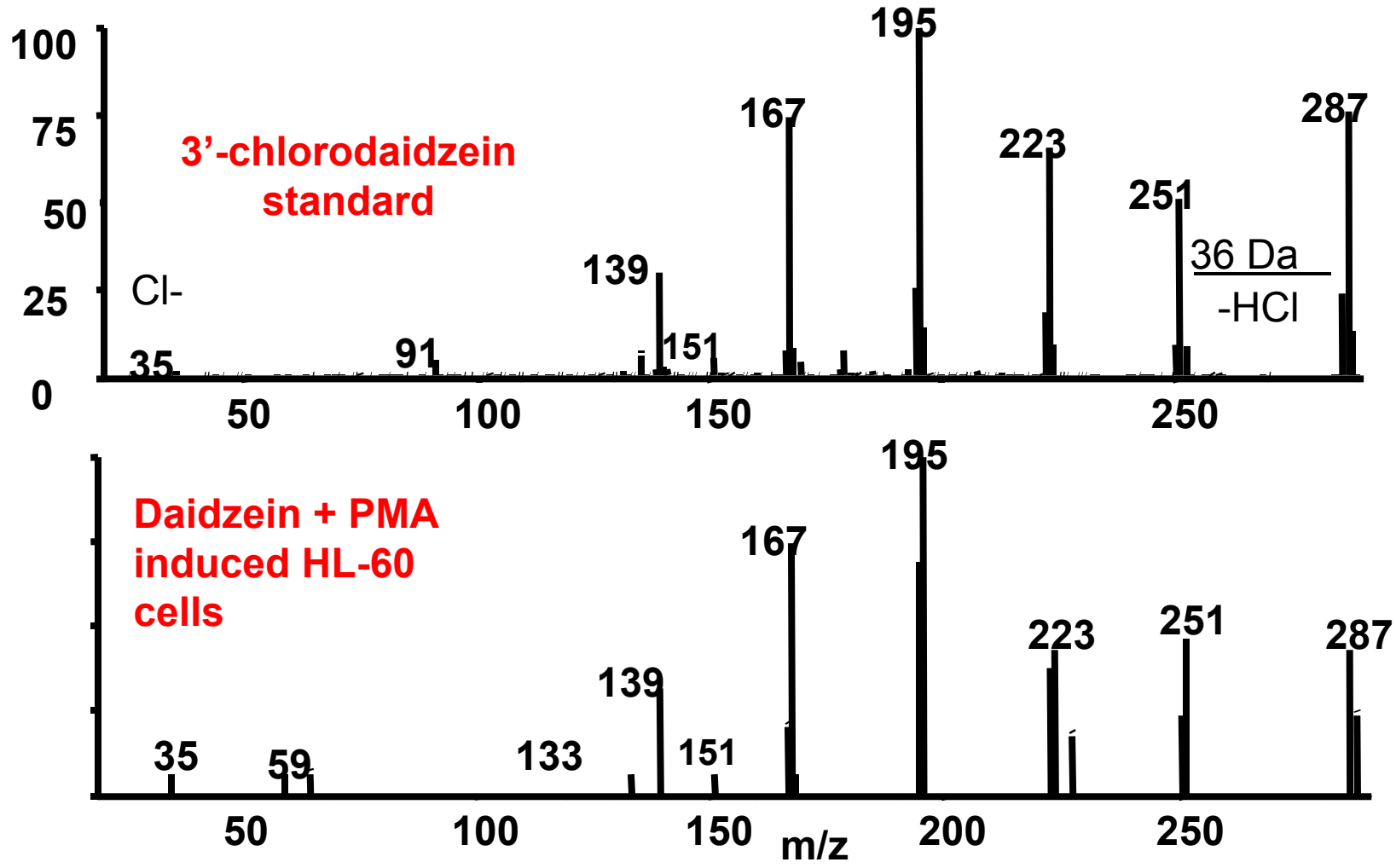
Isotopic pattern intensity of $[M-H]^-$ and ^{13}C and $[M-H+2]^-$ signals indicates the number of carbons and hetero atoms in the molecular ion



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

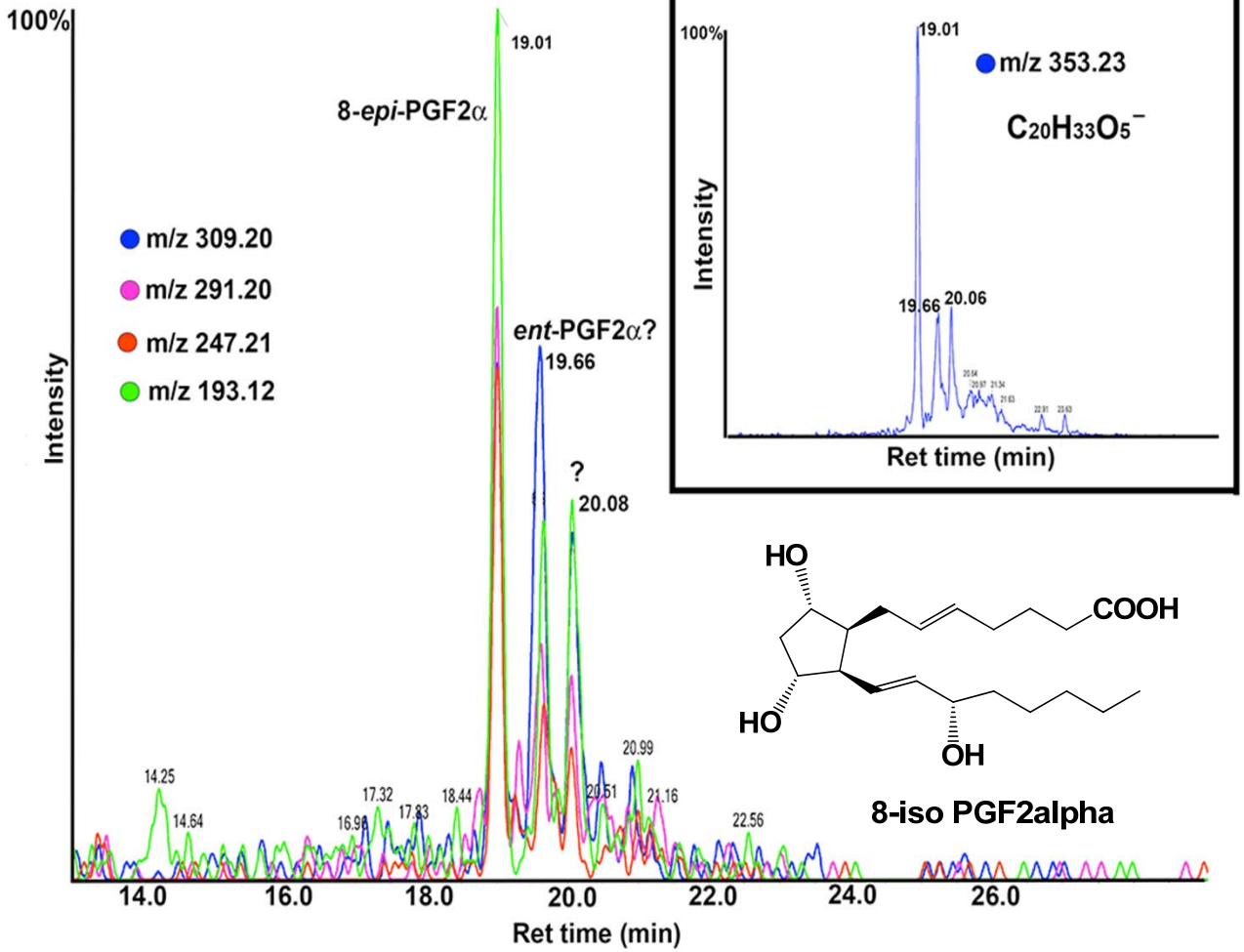
8- chlorodaidzein
m/z 287.011 (M-H]-; 289.008 [M-H+2]-

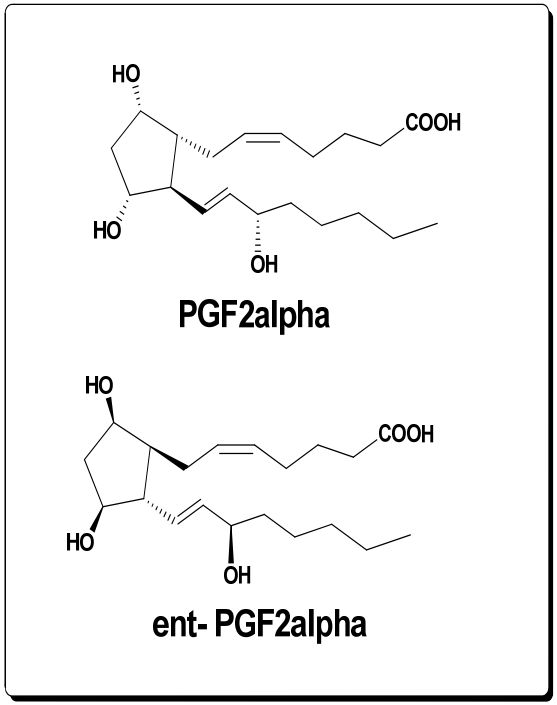
Comparison of product ions between known and unknown can help elucidate the unknown



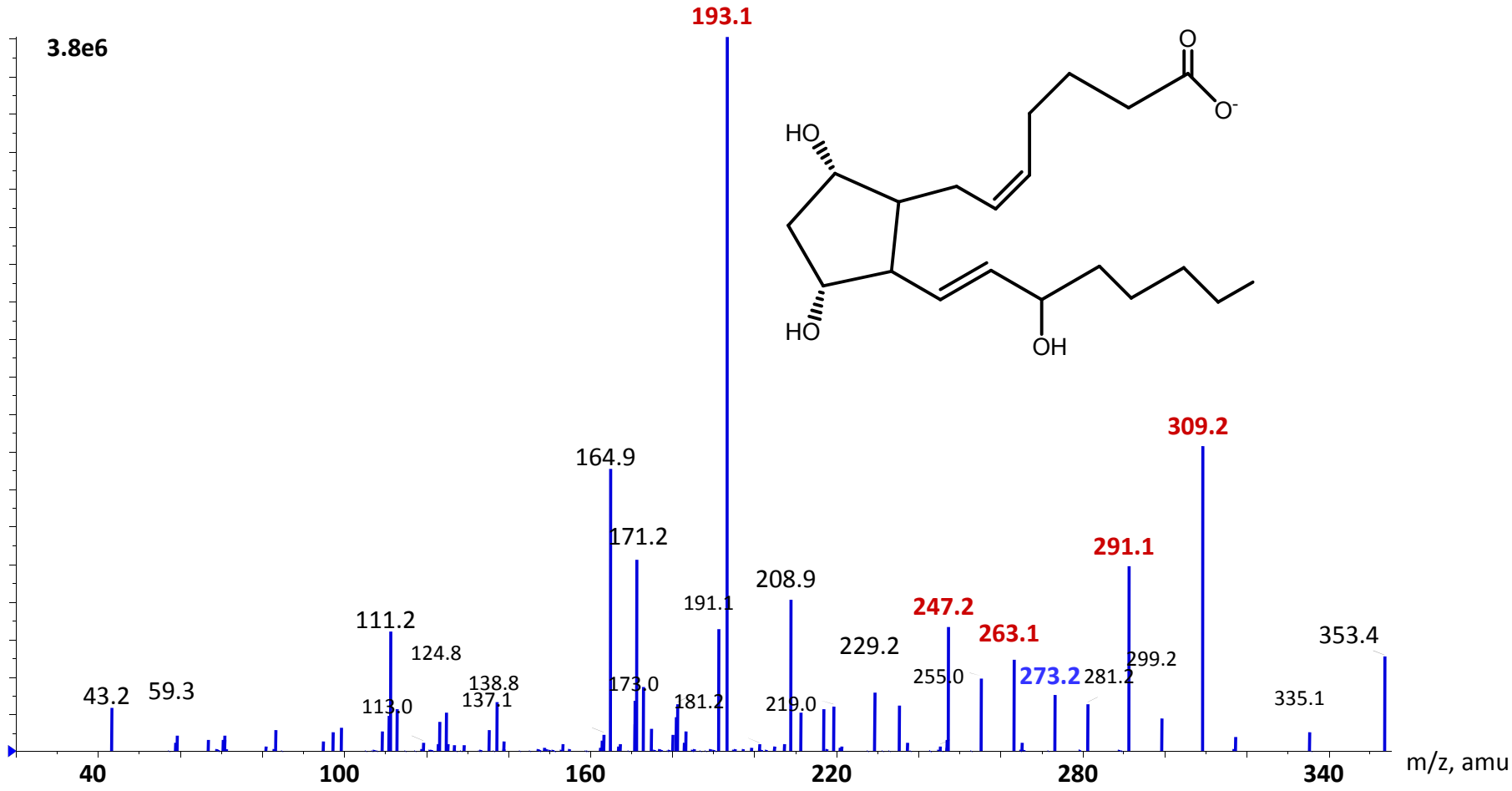
Prasain et al., 2003; Boersma et al., 2003

Good chromatographic separation and accurate mass indicate a number of diastereoisomer/enantiomer of PGF2alpha in worm extracts

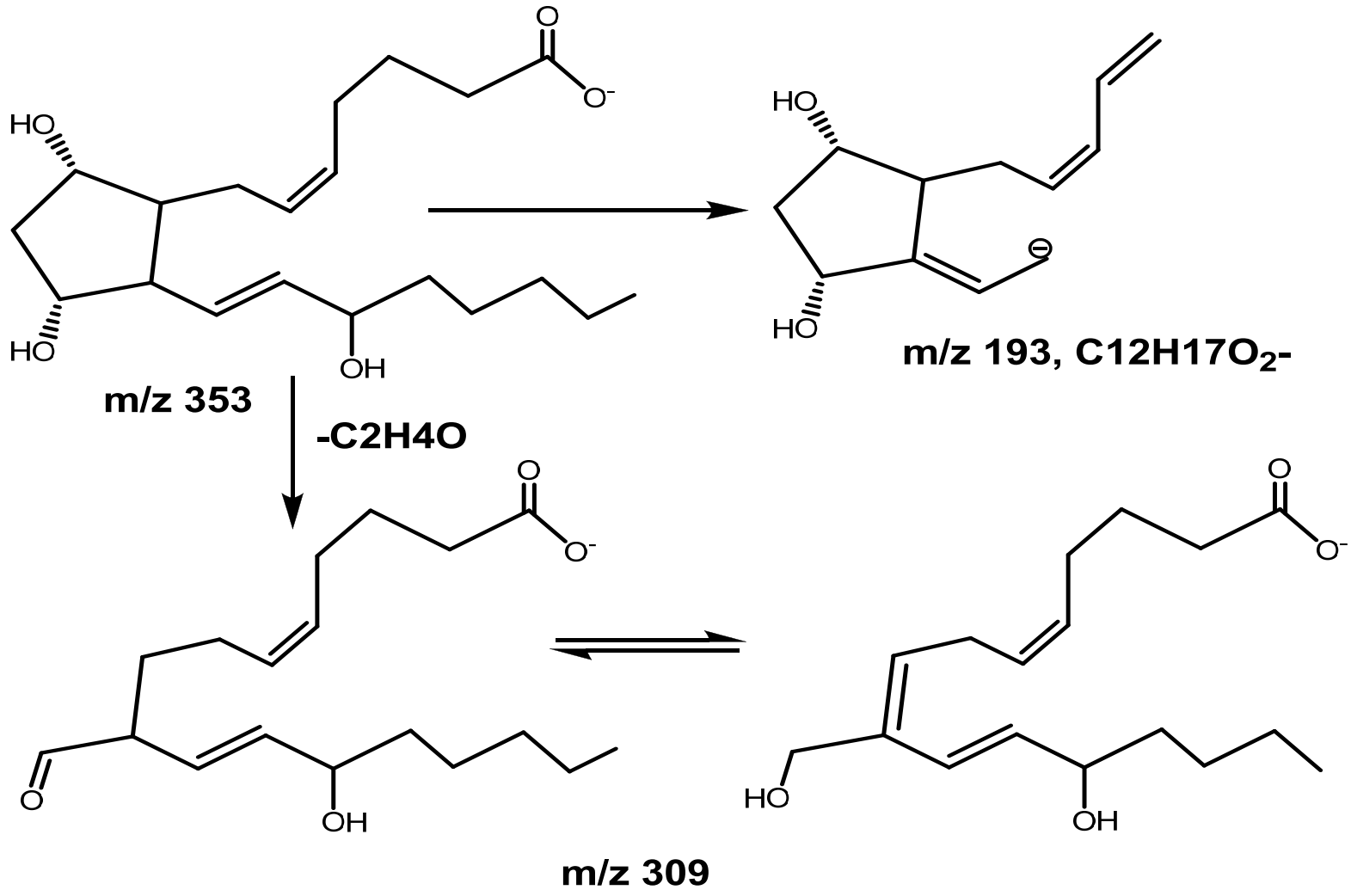




ESI-MS/MS of the [M-H]- from PGF2a m/z 353 using a quadrupole mass spectrometer

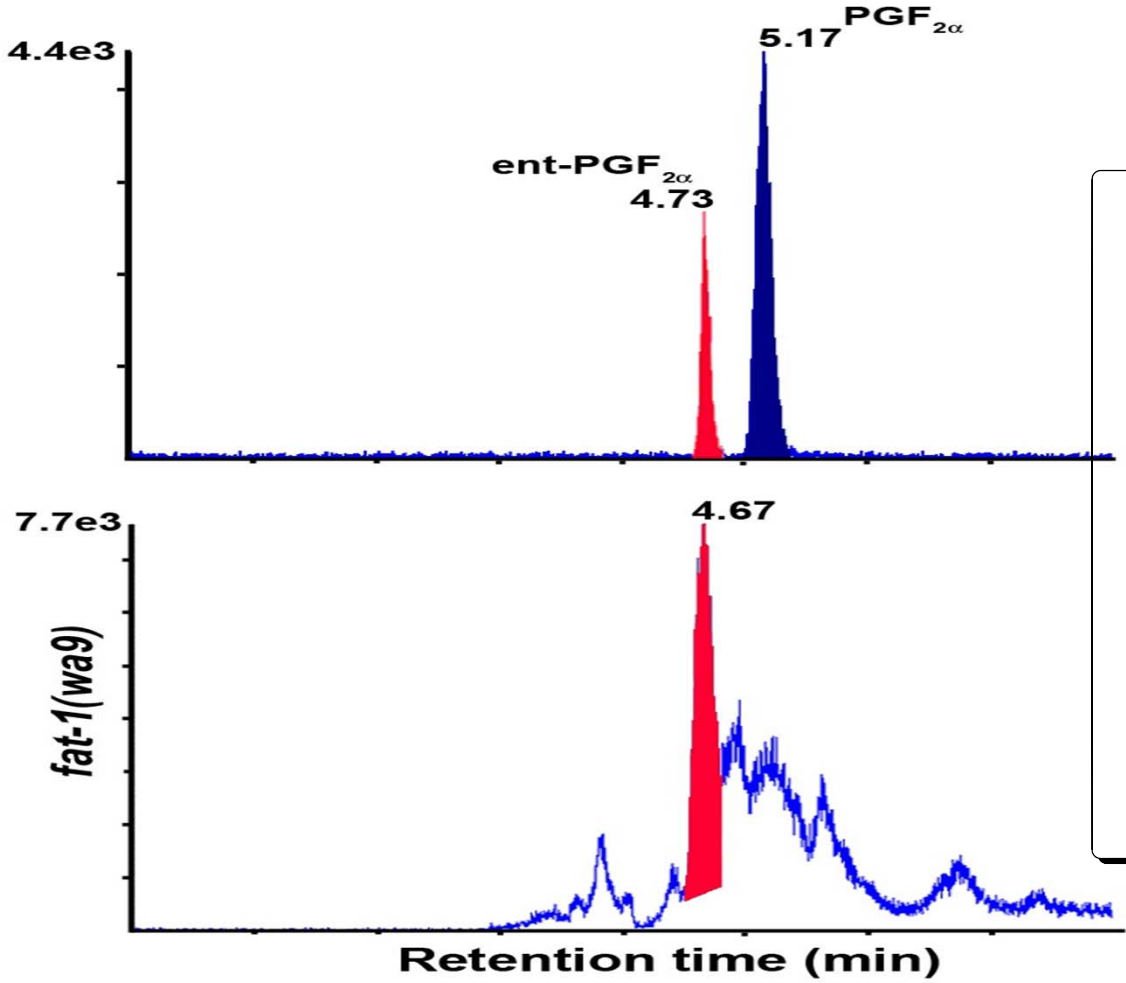


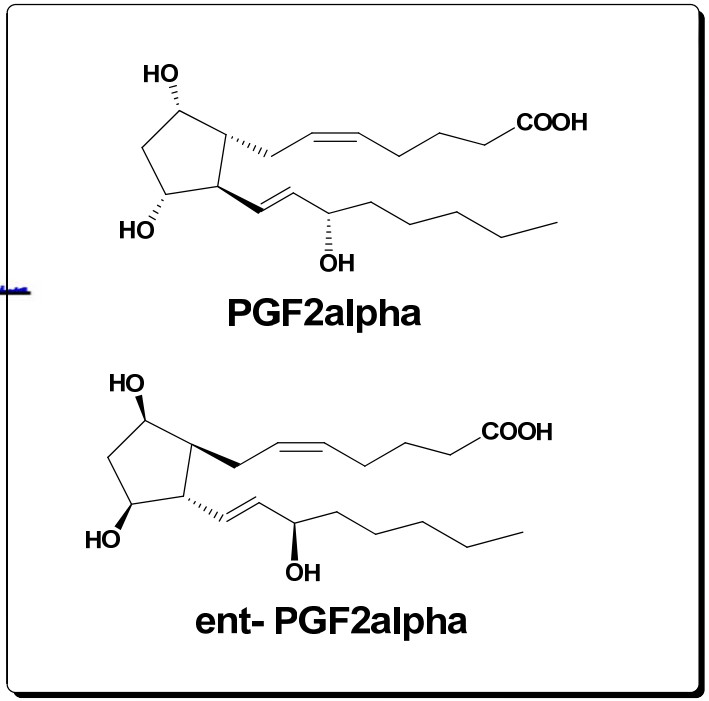
Fragmentation scheme of PGF2a $[M-H]^-$ m/z 353



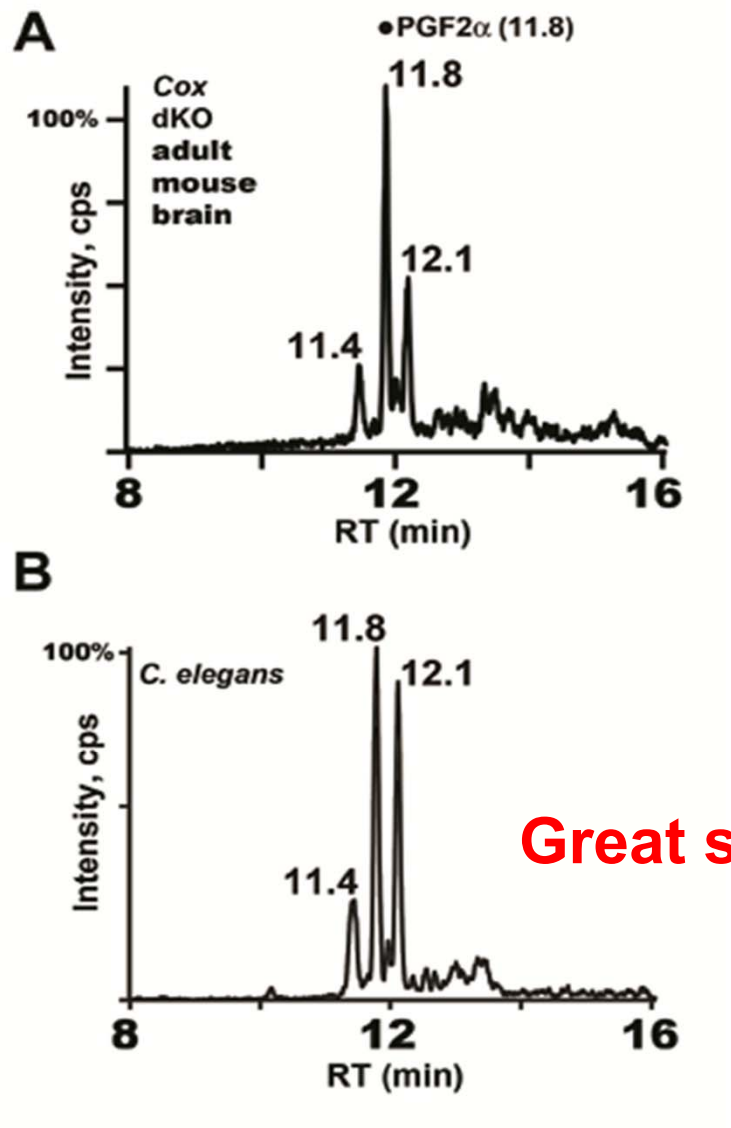
Ions m/z 309, 291, 273 and 193 are indicative of F2-ring

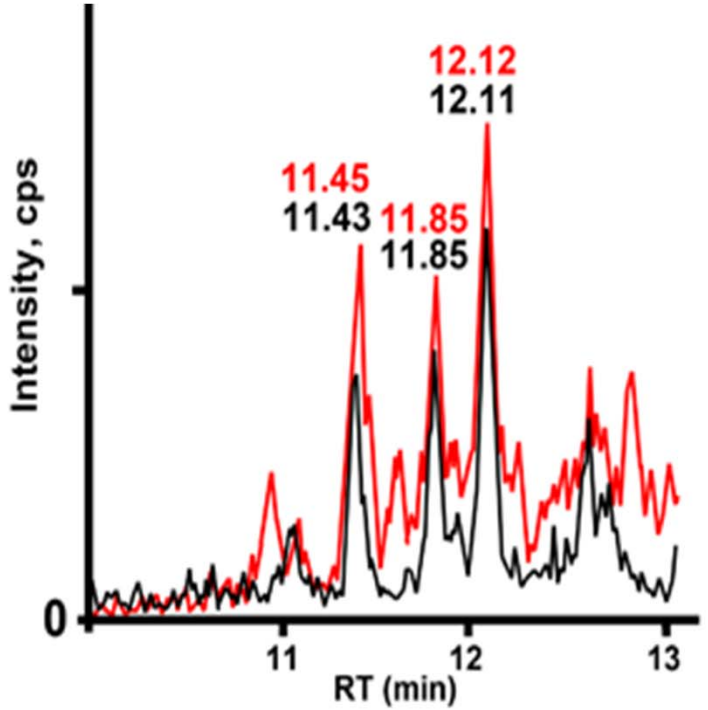
Only in chiral normal phase column,PGF2alpha and its enantiomer can be distinguished





Great similarities between *C. elegans* and Cox dKO pups PGF2 profile

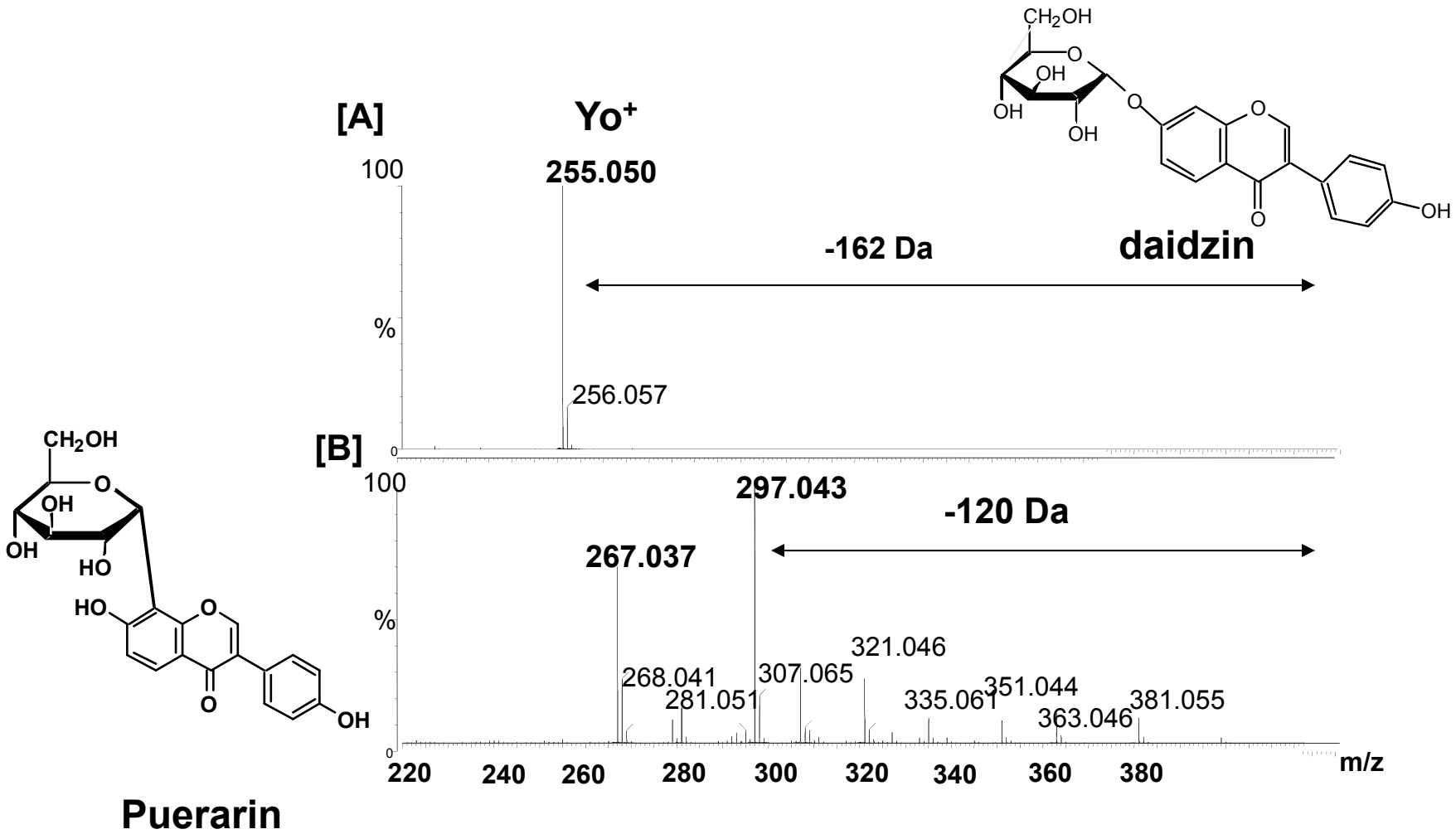




Great similarities between *C. elegans* and Cox dKO pups PGF2 profile

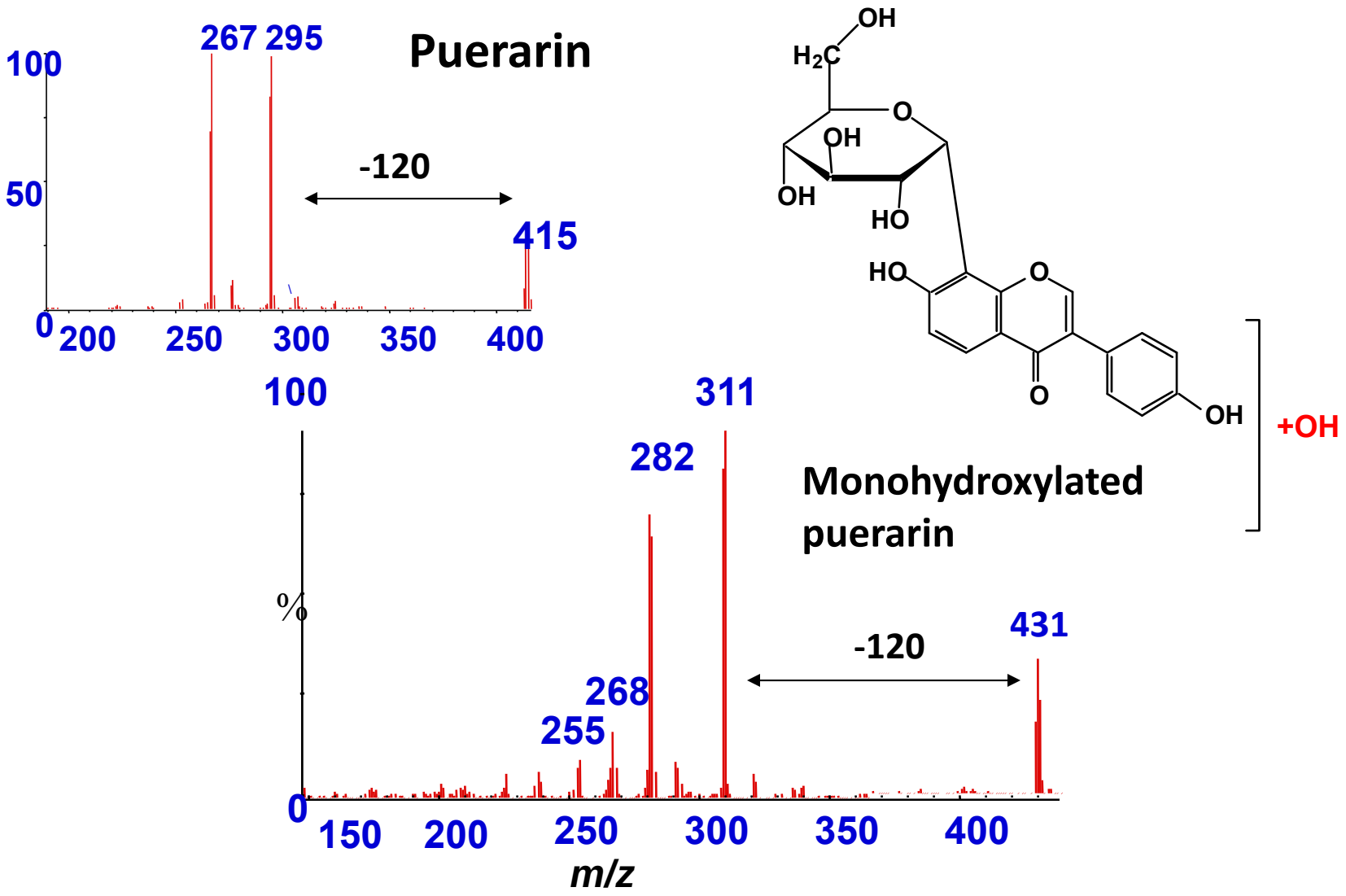
McKnight et al., (Science, 2014)

Ions break at weakest points; difference in O- and C-glucoside fragmentation



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

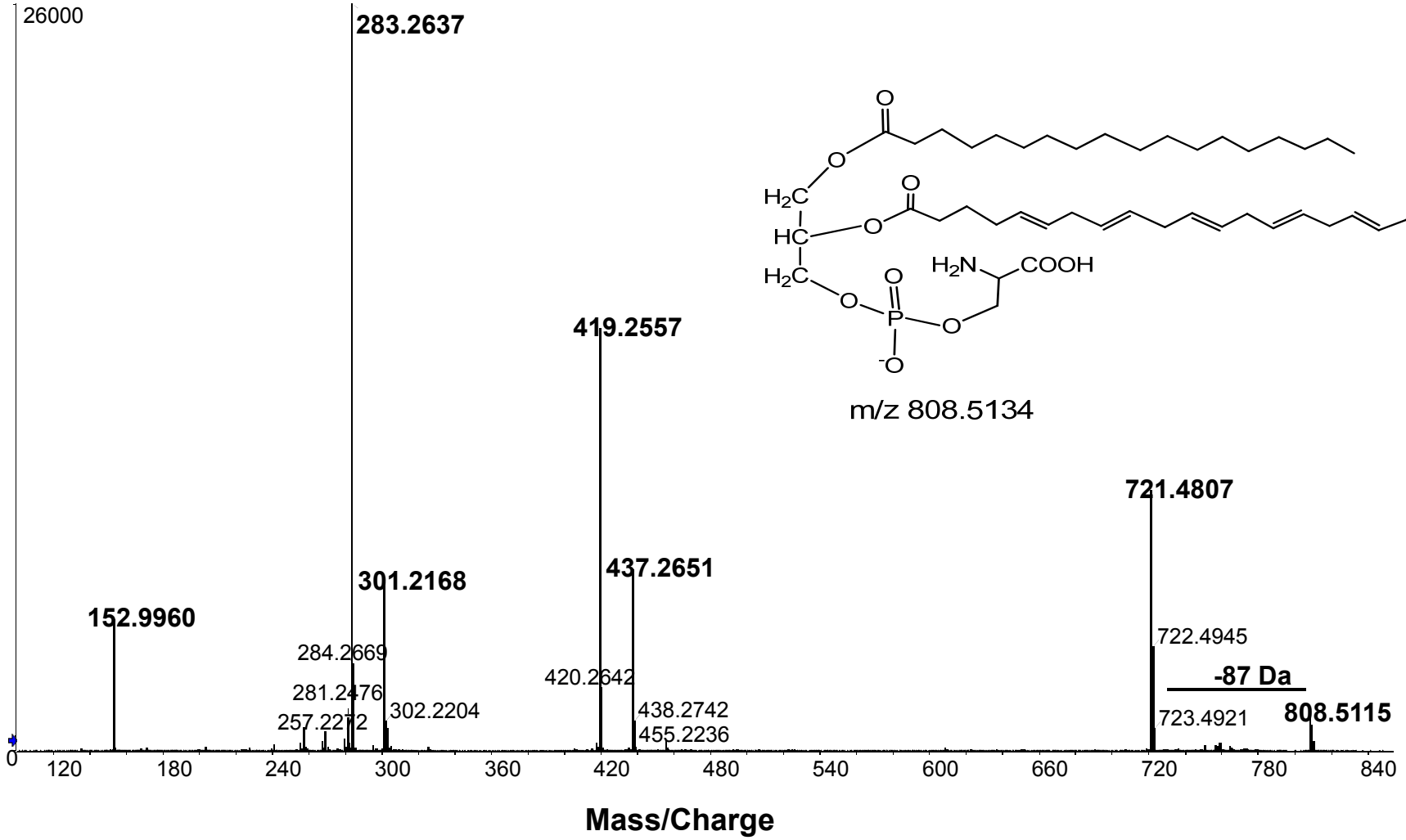
Identification of unknowns: comparing the MS/MS spectra of the known compound (puerarin)



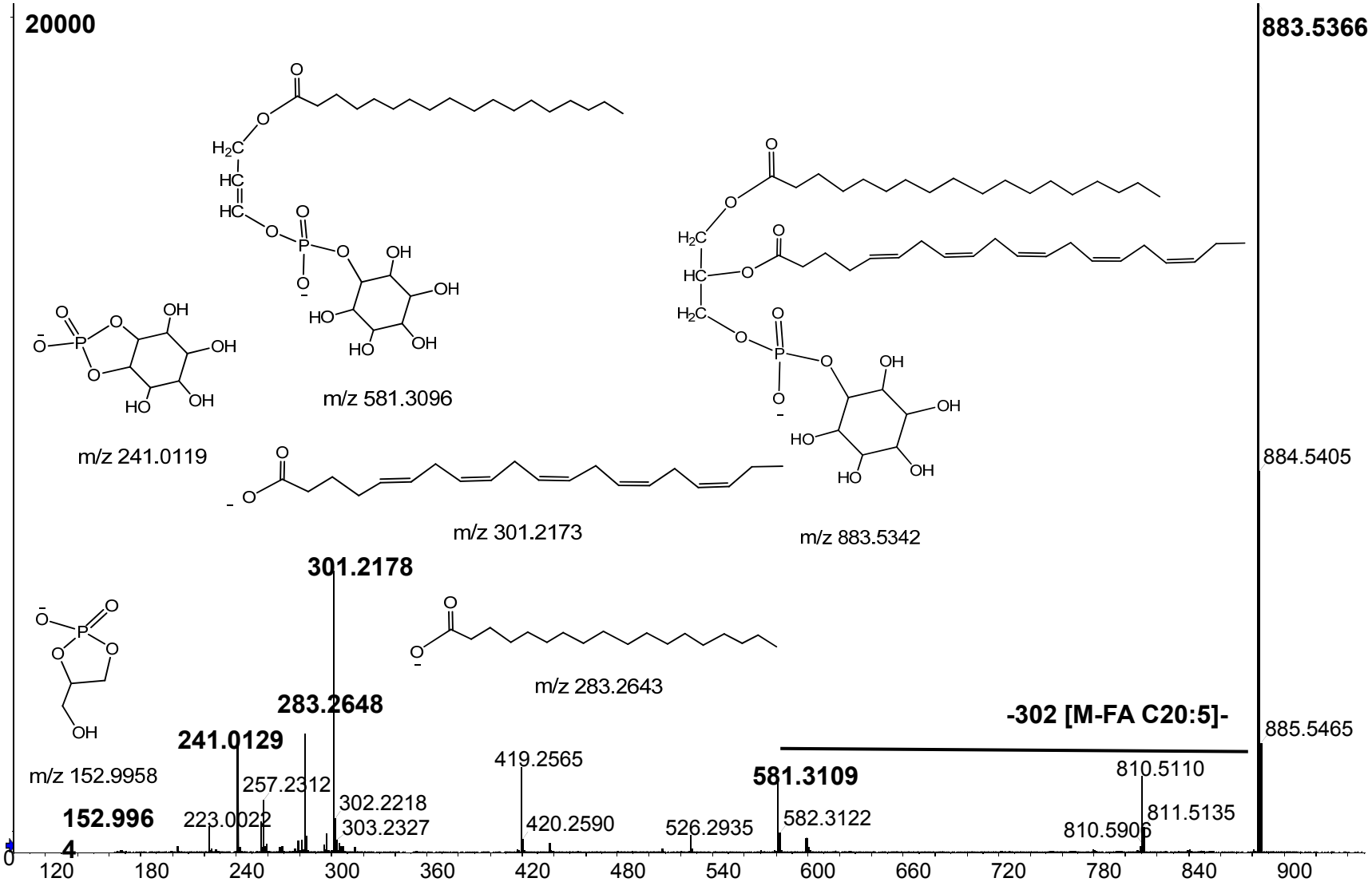
Nitrogen rule-

Odd number of nitrogens = odd MW

Even nitrogens = even MW



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



Prasain et al., unpublished results

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

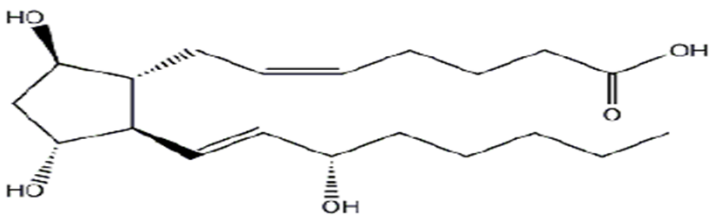
LMSD: Lipid classification search results

Fatty Acyls [FA] (W) --> Eicosanoids [FA03]

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-PGF1α	6-oxo-9S,11R,15S-trihydroxy-13E-prostenoic acid	$C_{20}H_{34}O_6$	370.24
LMFA03010002	PGF2α	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	PGF2α-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-prostadienoic acid	$C_{20}H_{32}O_6$	368.22

Structure database (LMSD)

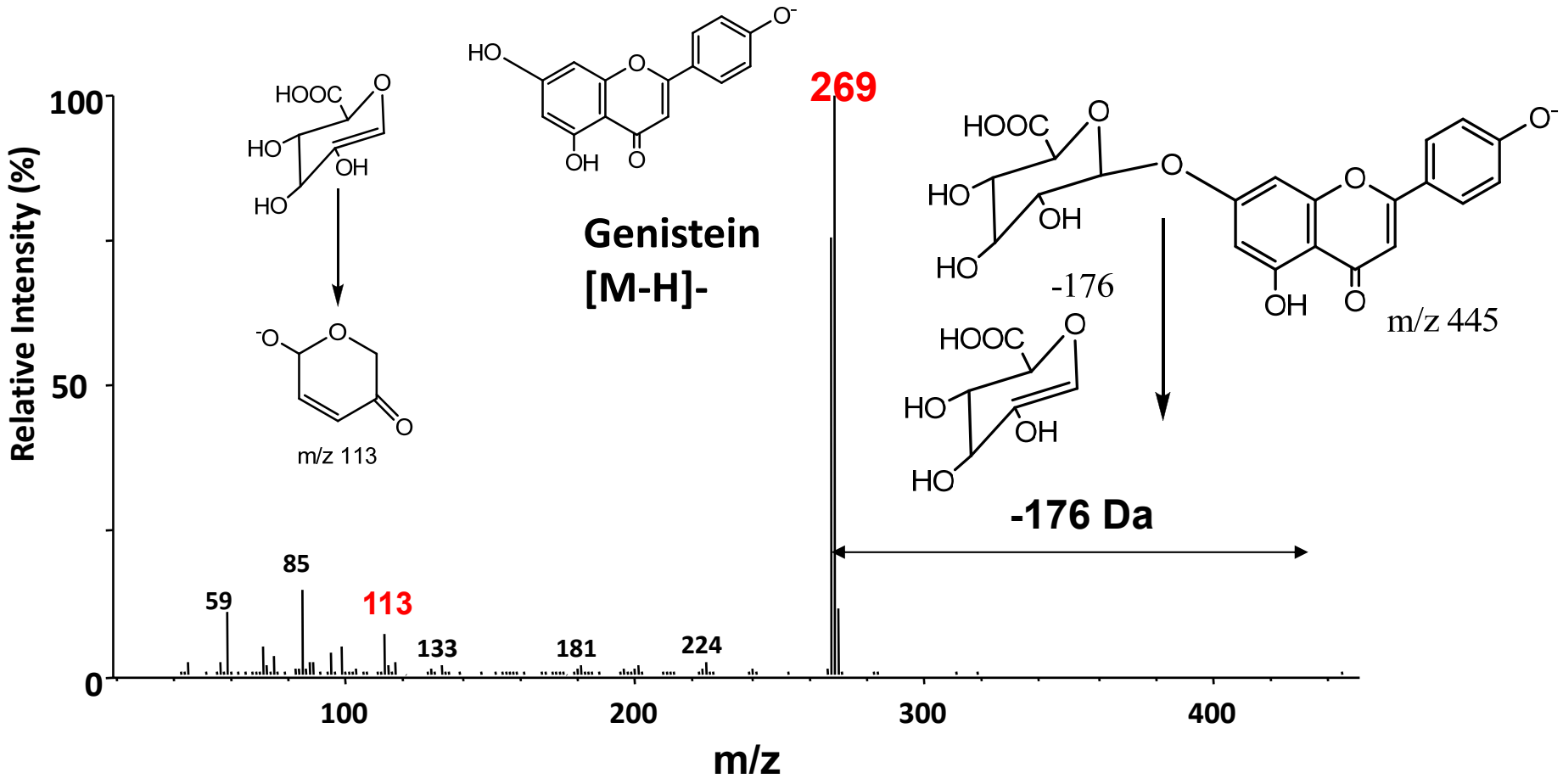
LMFA03010025



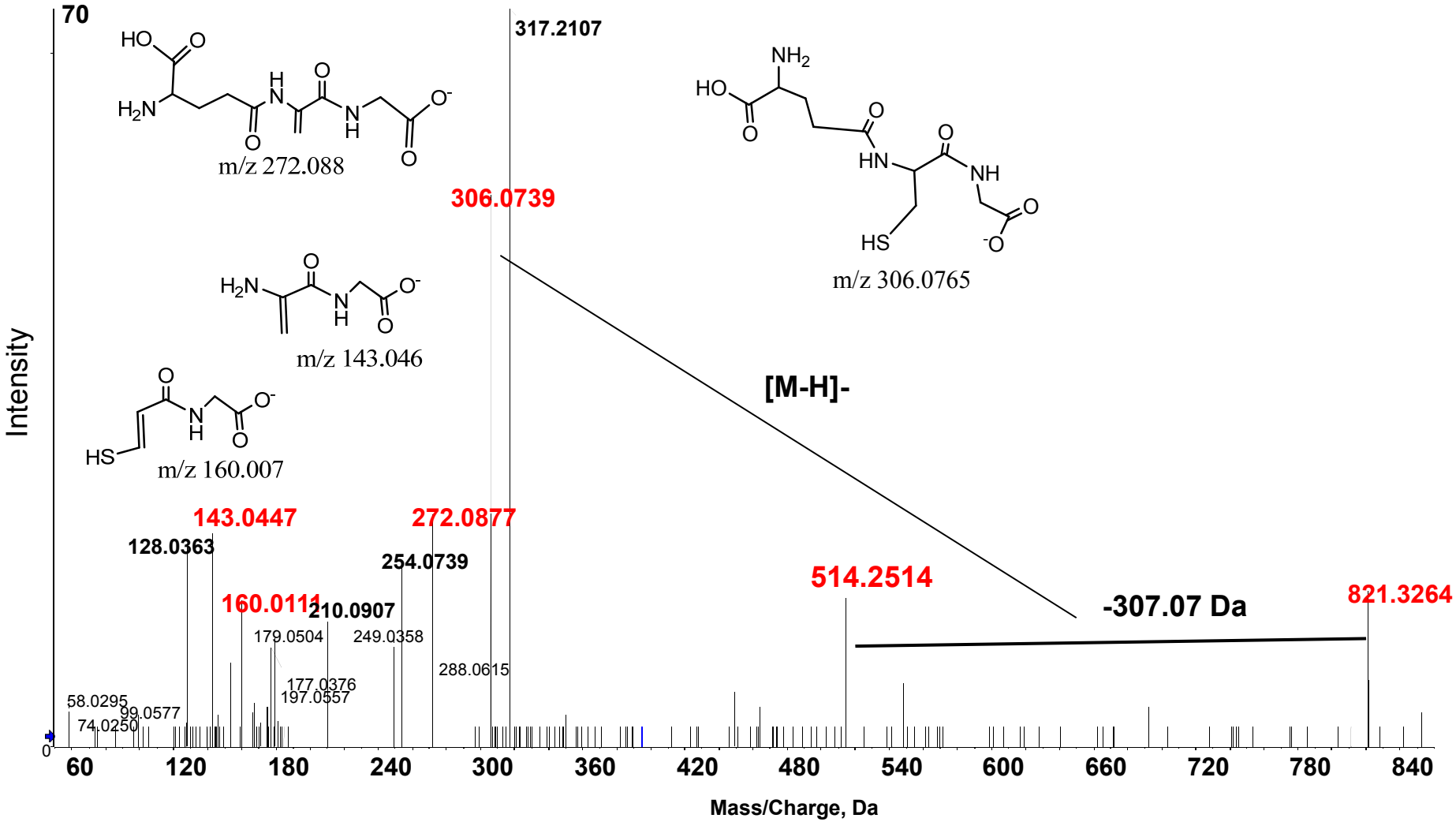
LM ID	LMFA03010025
Common Name	PGF2β
Systematic Name	9R,11R,15S-trihydroxy-5Z,13E-prostadienoic acid
Synonyms	-
Exact Mass	354.24
Formula	$C_{20}H_{34}O_5$
Category	Fatty Acyls [FA]
Main Class	Eicosanoids [FA03]
Sub Class	Prostaglandins [FA0301]
LIPIDBANK ID	XPR1764
PubChem Substance ID (SID)	4265968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

Targeted metabolomics: identification of conjugated metabolites (glucuronide/sulfate conjugates/glutathione-GSH)

The loss of anhydrous glucuronic acid (176 Da)- the characteristic of all glucuronides

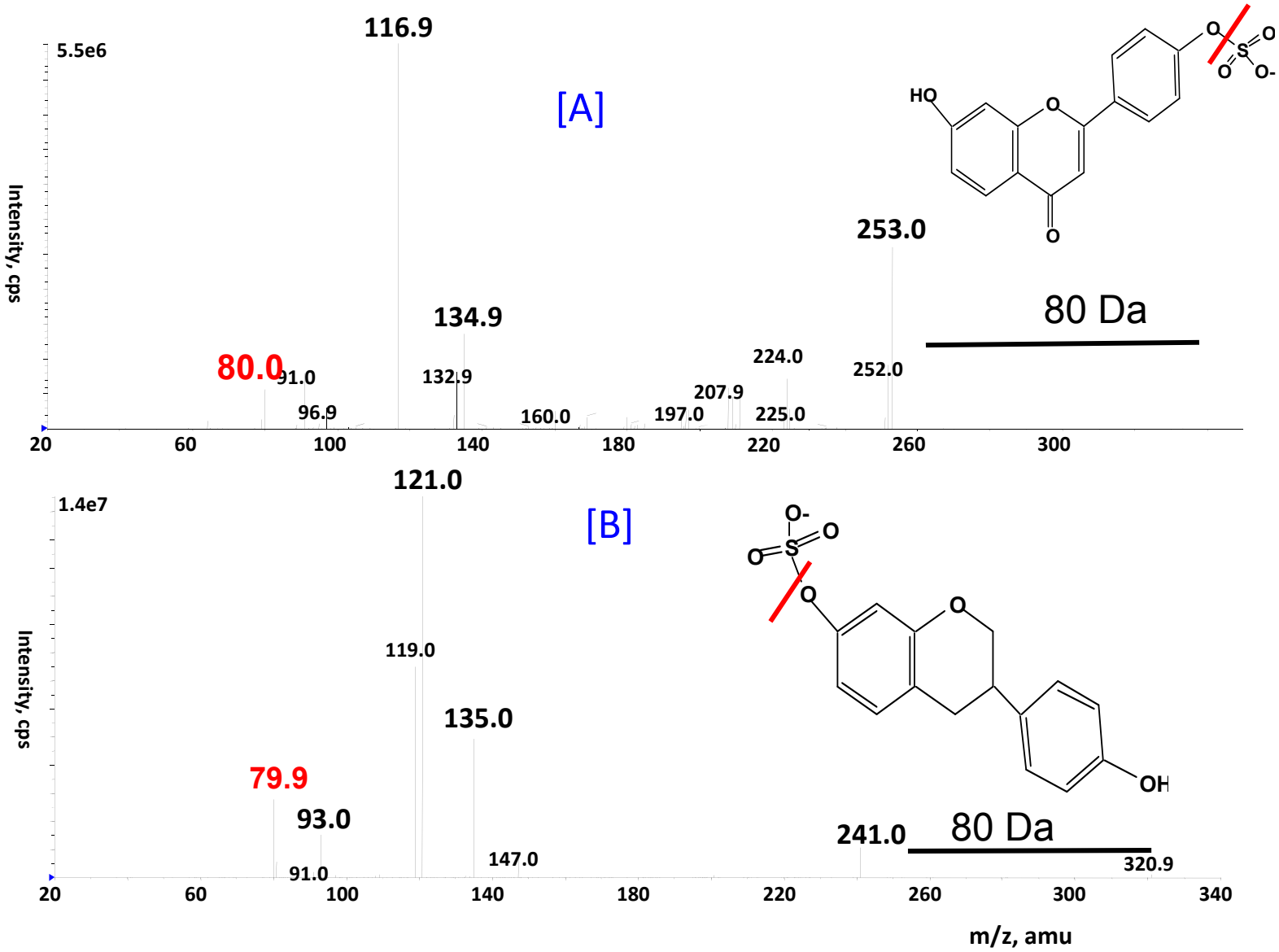


GSH conjugates can be identified with characteristic fragment ions *m/z* 306, 272, 254, 210, 179, 160 and 143 in -ve ion mode

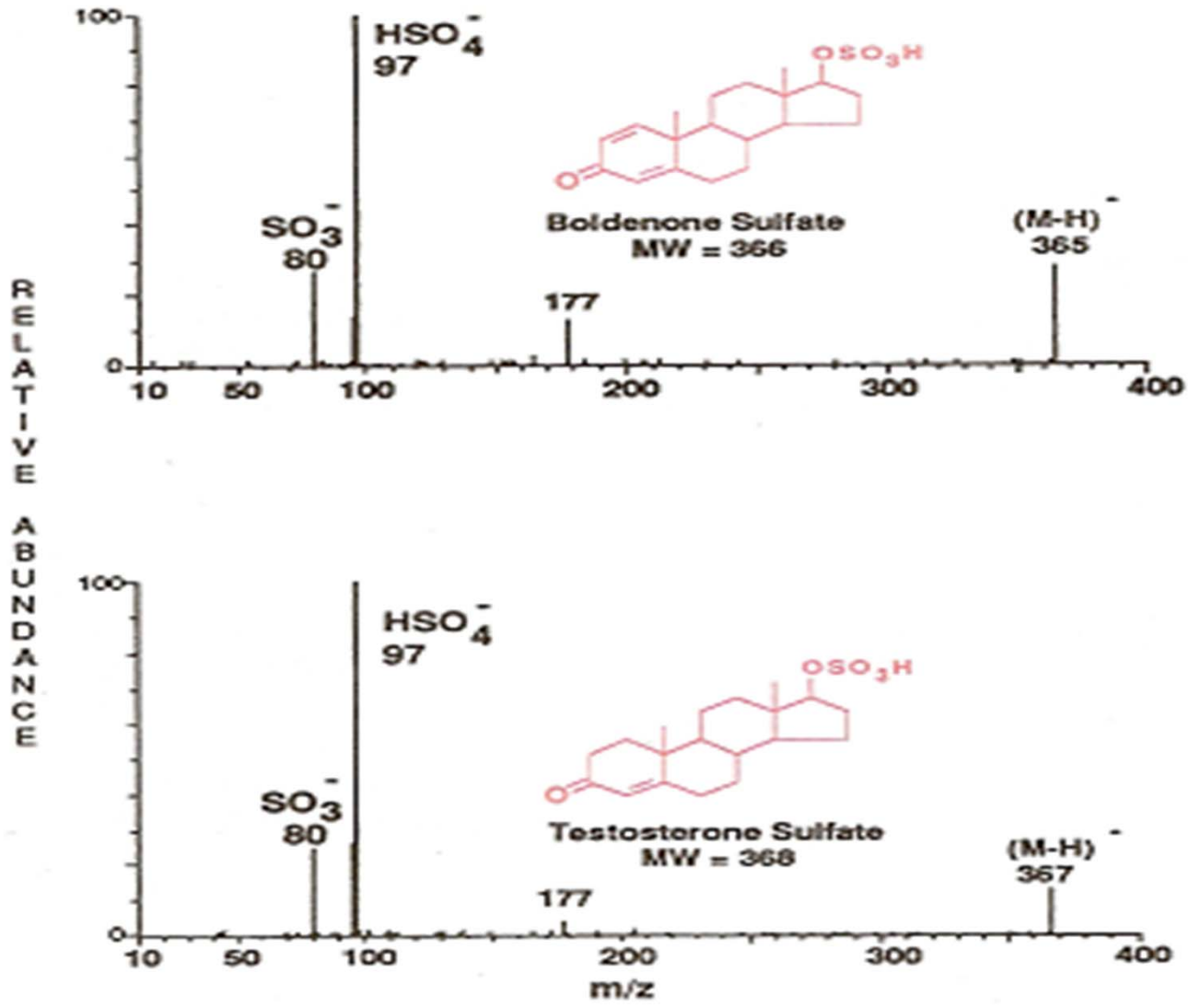


Jackson et al. unpublished results

The loss of 80 Da from the parent ion and the presence of *m/z* 80 in the product ion spectra indicates aromatic sulfate conjugates



Aliphatic sulfates typically show *m/z* 97 (HSO_4^-), whereas aromatics show *m/z* 80 (SO_3^-)



Weidolf et al. Biomed. and Environ. Mass Spec. 1988

Characteristic fragmentation of metabolite conjugates by MS/MS

Conjugate	Ionization mode	Scan
Glucuronides	pos/neg	NL 176 amu
Hexose sugar	pos/neg	NL 162 amu
Pentose sugar	pos/neg	NL 132 amu
Phenolic sulphate	pos	NL 80 amu
Phosphate	neg	Prec m/z 79, NL 80/98
GSH	Pos/neg	NL 129 amu
GSH	Neg	Prec m/z 272
taurines	Pos	Precursor of m/z 126
N-acetylcysteins	neg	NL 129 amu

NL = neutral loss.

Kostiainen et al., 2003

Conclusions

- **Identifying uncharacterized metabolites in biological samples is a significant analytical challenge and it requires integrated analytical approaches.**
- **A well separated metabolite signals in LC-MS/MS is crucial for unambiguous identification of new chemical entity.**
- **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 molecular ions.**
- **Combination of MS with NMR (eg. LC-NMR) is one of the most powerful analytical strategies for structure elucidation of novel metabolites.**

Acknowledgements

- Stephen Barnes, PhD (UAB)
- Michael Miller, PhD (UAB)
- Lalita A. Shevde, PhD (UAB)
- Ray Moore (ex-TMPL)
- Landon Wilson (TMPL)
- William P. Jackson (UAB)