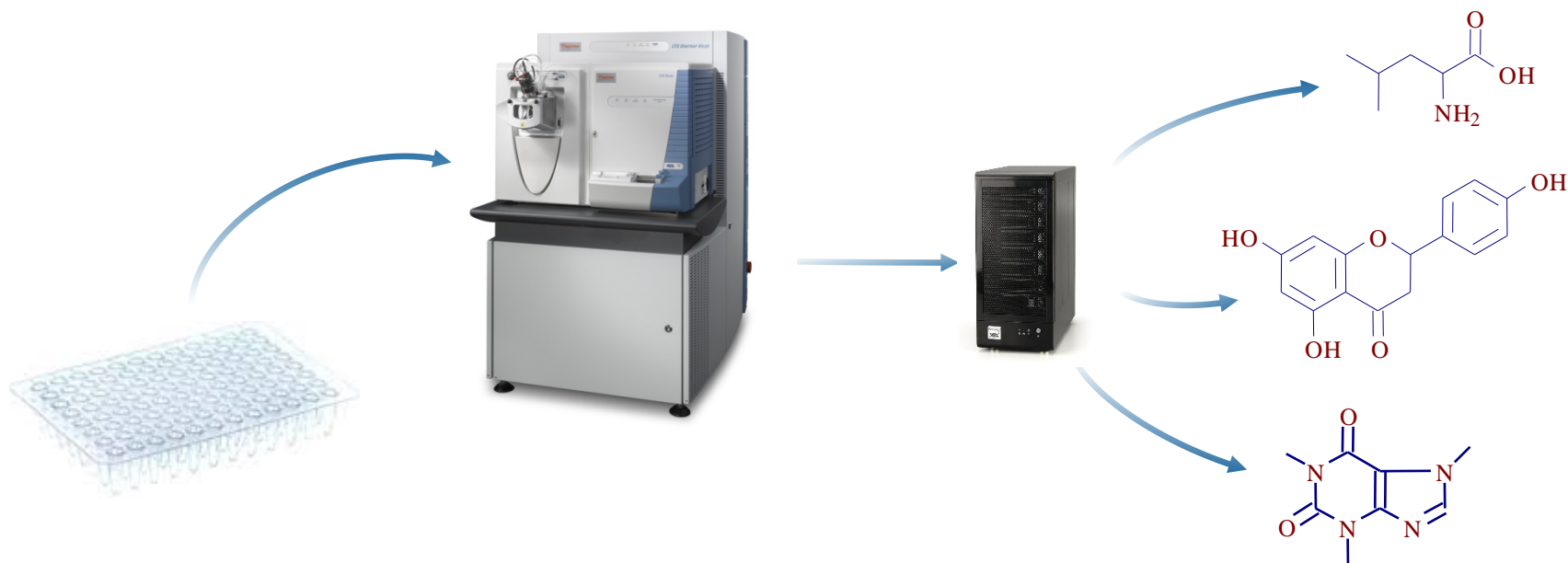


# Advanced Identification of Unknown Compounds



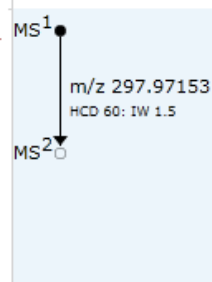
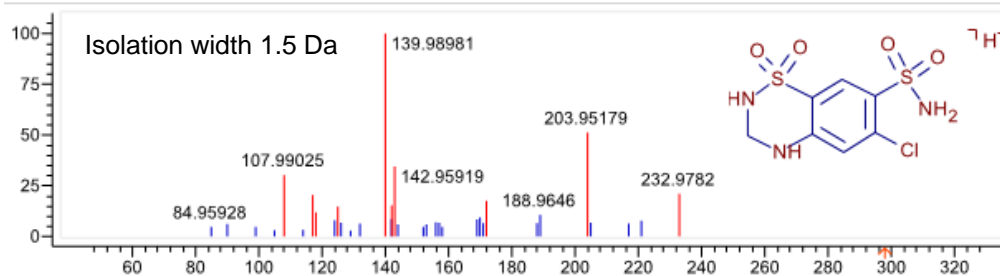
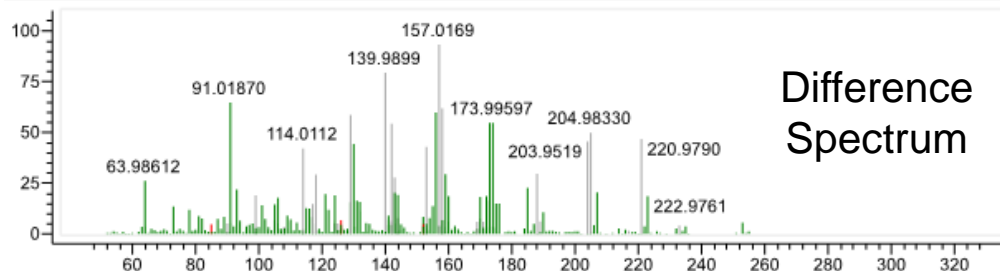
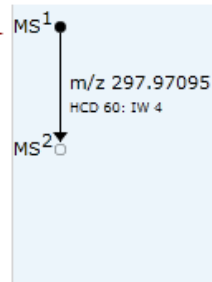
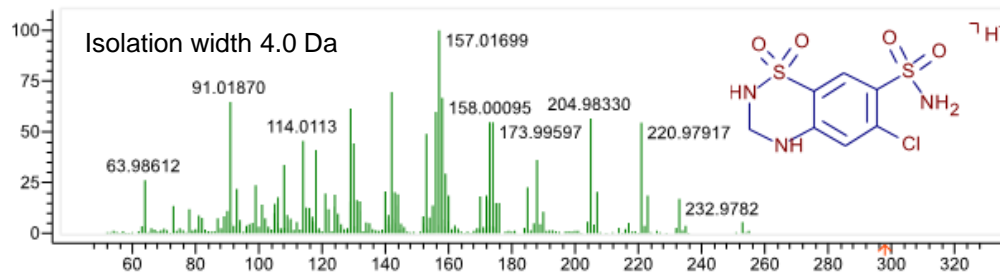
Robert Mistrik  
HighChem  
Bratislava, Slovakia

**HighChem**

## Why we cannot reliably identify majority of detected compounds using LC/MS?

- Nature of mass spectrometry (Inherent problems with structure elucidation)
- Numerous problems associated with library search
- Existing LCMS spectral databases are small and contains virtually the same compounds
- *Ad hoc* identification
- Promises made by computer assisted methods have not been fulfilled yet

# Problems Associated with LCMS Library Search



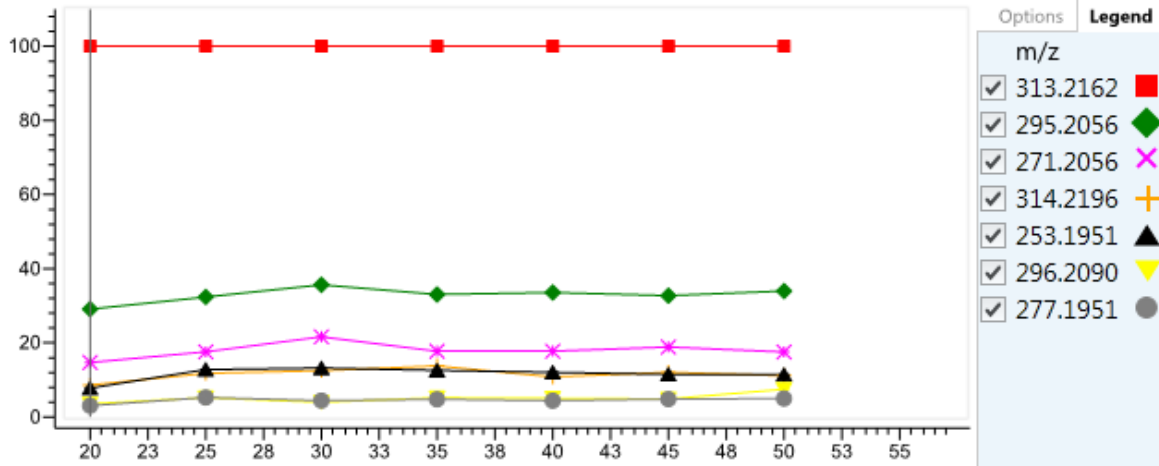
Spectrum courtesy of M. Stravs, E. Schymanski and H. Singer, EAWAG

## Spectra reproducibility Experimental variability

- Spectra quality
- Spectra specificity
- Information ambiguity
- Isomers differentiation
- Structural diversity
- Accuracy issues
- Spectrum/Structure relationship

# Problems Associated with LCMS Library Search

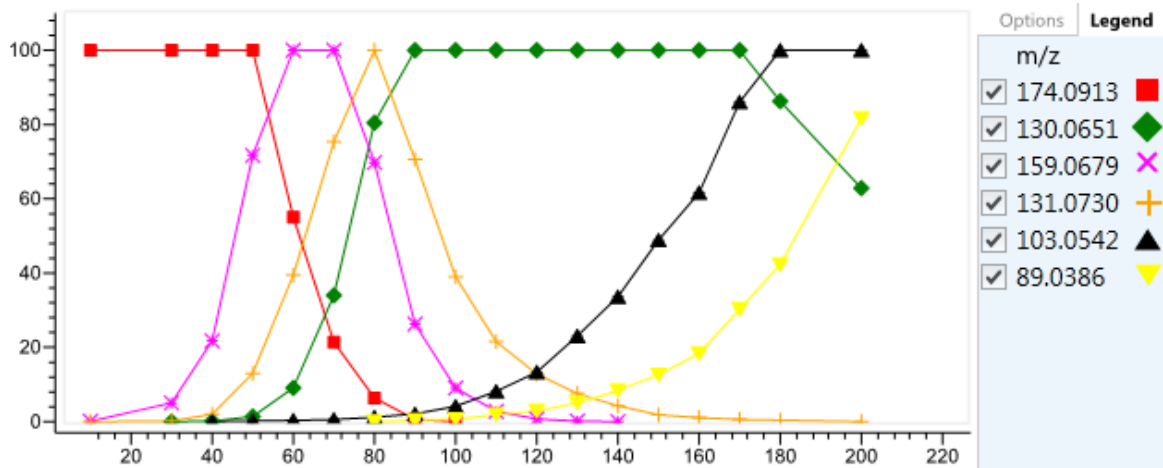
17 $\alpha$  Hydroxyprogesterone, CID MS<sup>2</sup> [M+H]<sup>+</sup>



• Spectra reproducibility  
Experimental variability

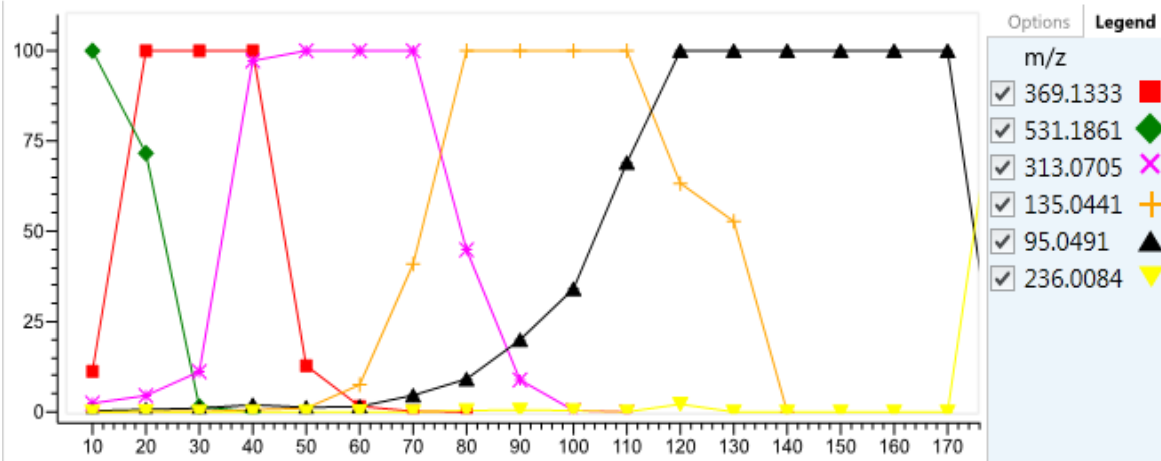
- Spectra quality
- Spectra specificity
- Information ambiguity
- Isomers differentiation
- Structural diversity
- Accuracy issues
- Spectrum/Structure relationship

Serotonin, CID MS<sup>3</sup> [M+H]<sup>+</sup> → m/z 174.09134

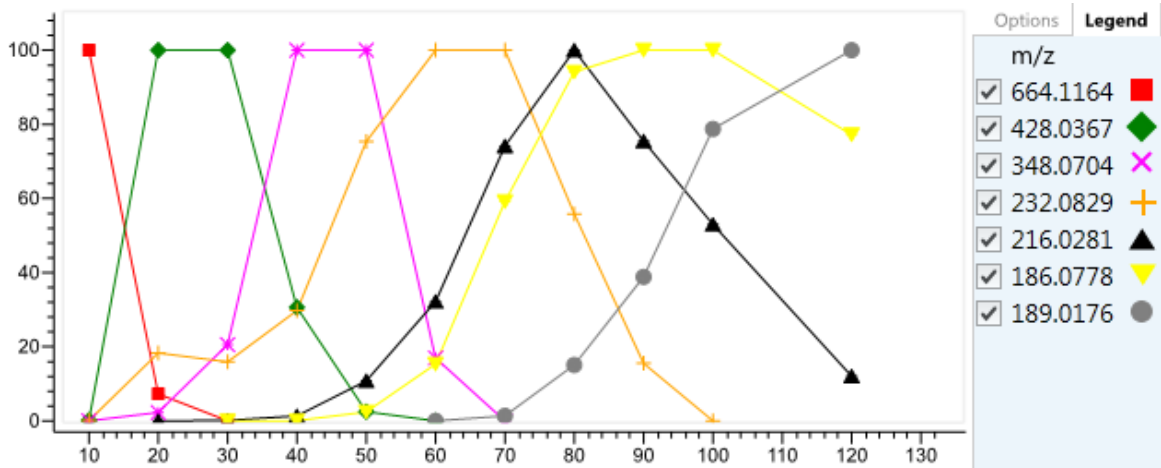


# Problems Associated with LCMS Library Search

Epimedin C, HCD MS<sup>2</sup> [M+H]<sup>+</sup>



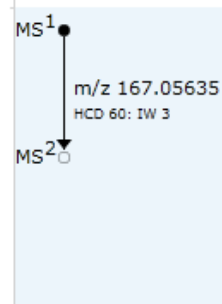
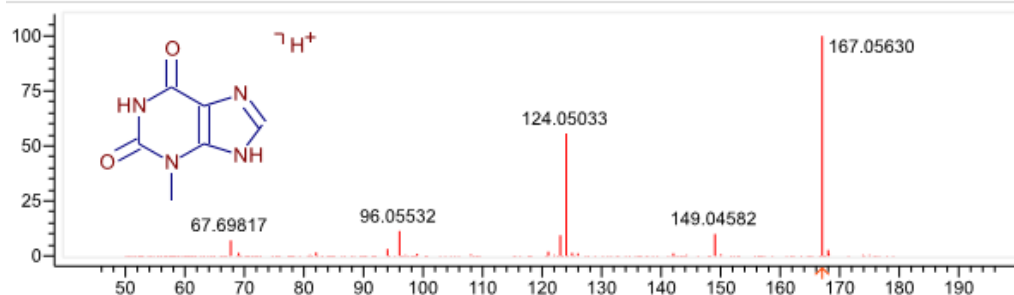
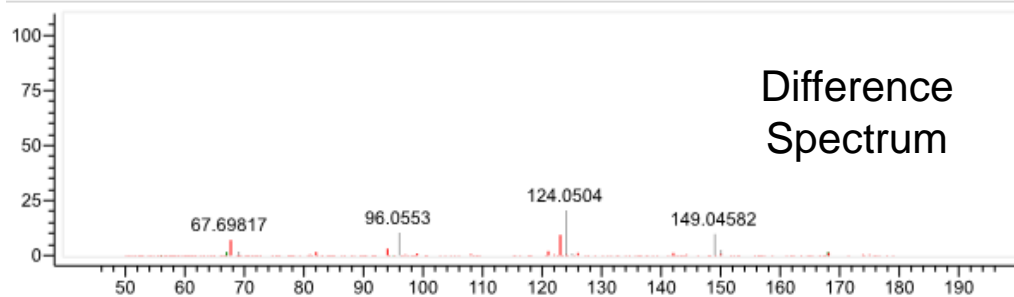
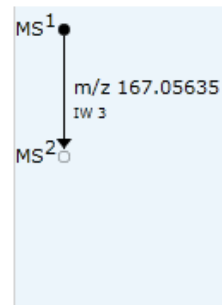
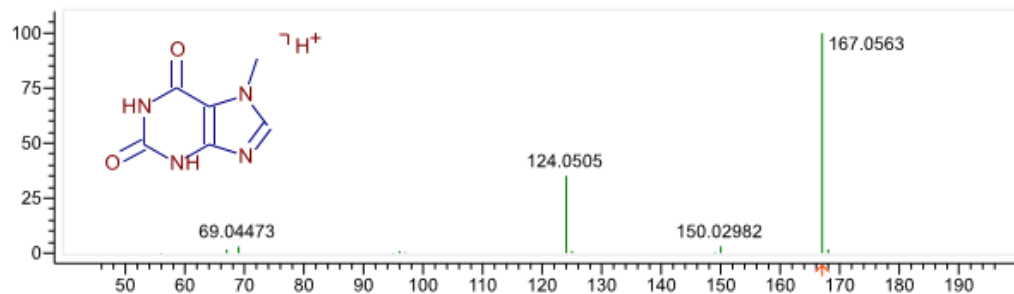
NAD<sup>+</sup>, HCD MS<sup>2</sup> [M+H]<sup>+</sup>



• Spectra reproducibility  
• Experimental variability

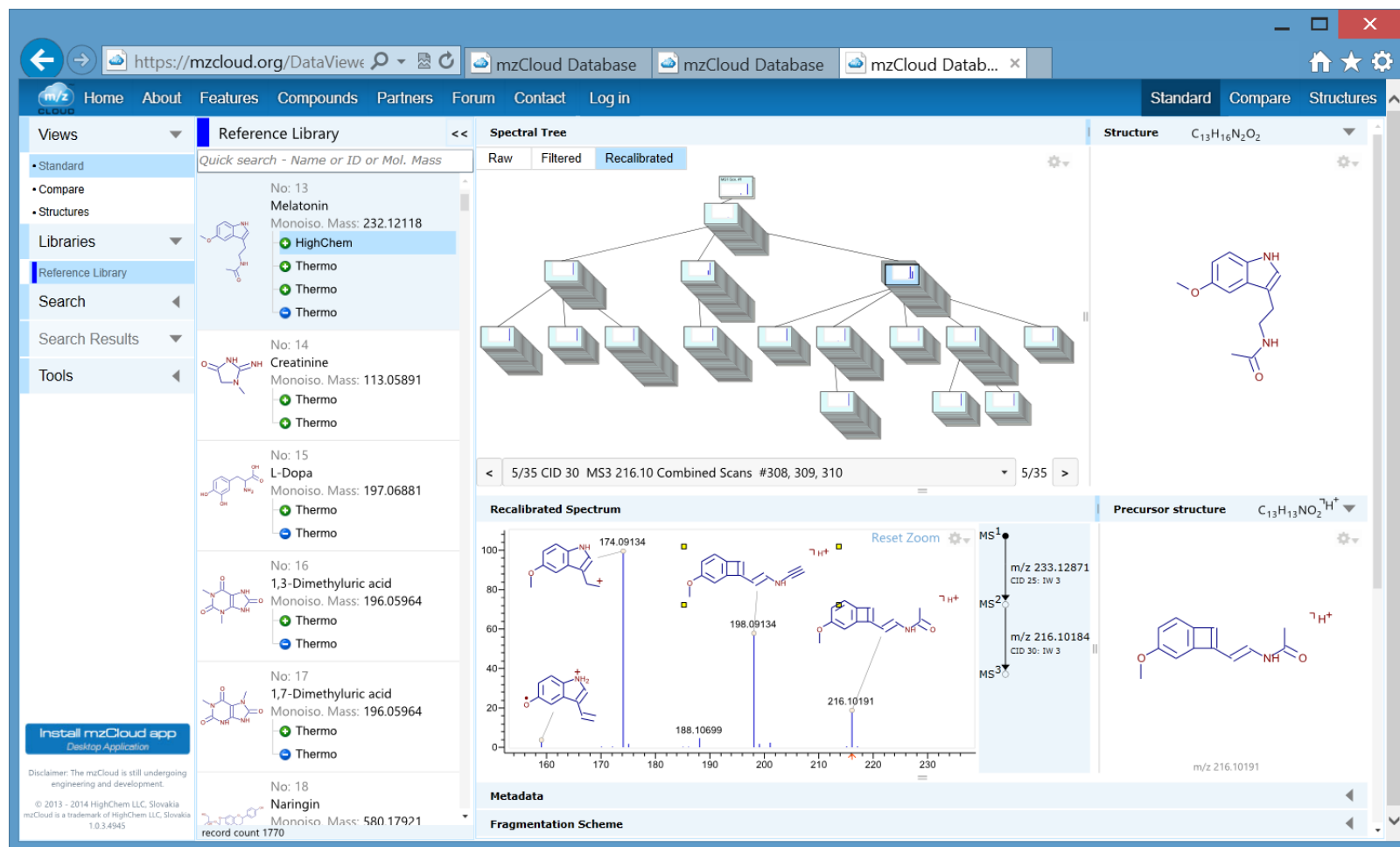
- Spectra quality
- Spectra specificity
- Information ambiguity
- Isomers differentiation
- Structural diversity
- Accuracy issues
- Spectrum/Structure relationship

# Problems Associated with LCMS Library Search



- Spectra reproducibility
- Experimental variability
- Spectra quality
- Spectra specificity
- Information ambiguity
- **Isomers differentiation**
- Structural diversity
- Accuracy issues
- Spectrum/Structure relationship

# mzCloud - a freely searchable collection of high resolution/accurate spectra



The screenshot displays the mzCloud web interface with the following components:

- Navigation:** Home, About, Features, Compounds, Partners, Forum, Contact, Log in.
- Views:** Reference Library (selected), Standard, Compare, Structures.
- Reference Library:**
  - No: 13 Melatonin (Monoiso. Mass: 232.12118) with HighChem, Thermo, and Thermo tags.
  - No: 14 Creatinine (Monoiso. Mass: 113.05891) with Thermo and Thermo tags.
  - No: 15 L-Dopa (Monoiso. Mass: 197.06881) with Thermo and Thermo tags.
  - No: 16 1,3-Dimethyluric acid (Monoiso. Mass: 196.05964) with Thermo and Thermo tags.
  - No: 17 1,7-Dimethyluric acid (Monoiso. Mass: 196.05964) with Thermo and Thermo tags.
  - No: 18 Naringin (Monoiso. Mass: 580.17921, record count 1770) with Thermo and Thermo tags.
- Spectral Tree:** A tree diagram showing the fragmentation path from the precursor to the base peak.
- Recalibrated Spectrum:** A mass spectrum plot with peaks at m/z 174.09134 (base peak), 198.09134, 216.10191, and 233.12871. The x-axis ranges from 160 to 230 m/z.
- Structure:** Chemical structure of Melatonin (C13H16N2O2).
- Precursor structure:** Chemical structure of the precursor (C13H13NO2H+).
- Metadata:** 5/35 CID 30 MS3 216.10 Combined Scans #308, 309, 310.
- Fragmentation Scheme:** A diagram showing the fragmentation of the precursor into the base peak.

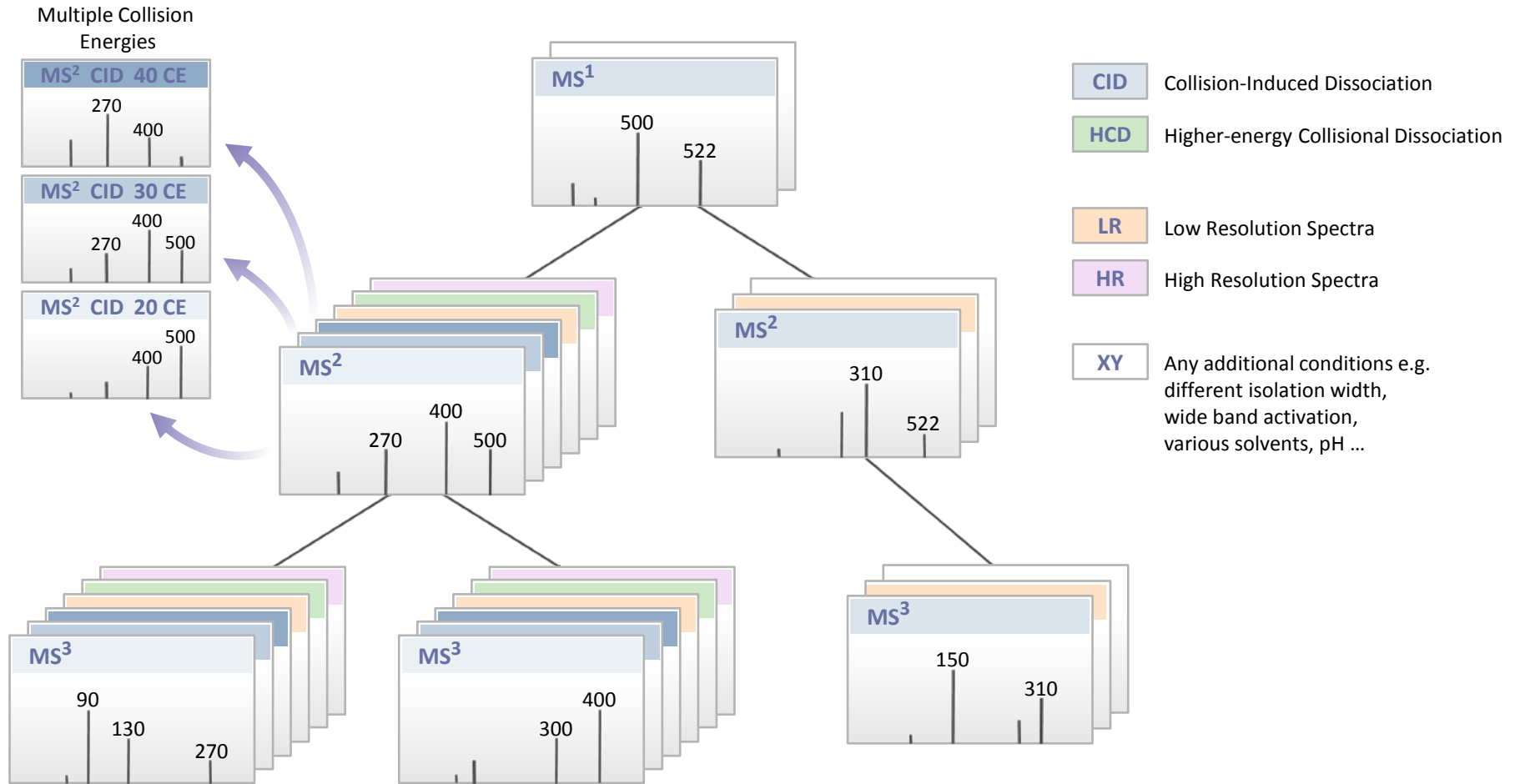


## mzCloud features

- Spectral trees – intuitively organized multi-stage tandem mass spectra
- Identification of compounds even if they are not represented in the library
- Comprehensive spectral data – a guarantee for correct compound identification
- Filtered and recalibrated spectra
- Third generation spectra correlation algorithm
- Structurally annotated spectral peaks using non-trivial prediction methods
- Substructure search capable of extracting spectra with common structural scaffolds
- Fragment structure support throughout the database
- Extensive experimental information and identifier metadata data
- Rigorously curated data
- Advanced cloud technology
- Multiple databases



# Spectral Trees





The screenshot displays the mzCloud.org web application interface, showing search results and spectral analysis for Cocaine. The interface is presented as a stack of overlapping browser windows.

**Search Results:**

- Peak search result 1:** Edit search options
- Reference No: 584:** Anhydroecgonine methyl ester, Monoiso. Mass: 181.11028, Thermo
- Reference No: 710:** Ecgonine methyl ester, Monoiso. Mass: 199.12084, Thermo
- Reference No: 1017:** Cocaine, Monoiso. Mass: 303.14706, Eawag

**Spectral Tree:**

- Raw
- Filtered
- Recalibrated

**Structure:** C10H17NO3

**Recalibrated Spectrum:**

- MS1 Scns. #45, 56, 67
- HCD 40 MS2 200.13 Scns. #49, 60, 71
- 4/11 HCD 40 MS2 200.13 Combined Scans #49, 60 4/11

**Recalibrated Spectrum Plot:**

- Y-axis: Relative intensity (0 to 100)
- X-axis: m/z (60 to 220)
- Major peaks: 82.06513, 108.08078, 150.09134, 182.11756 (base peak), 200.12752
- MS<sup>1</sup> and MS<sup>2</sup> labels are present.

**Precursor structure:** C10H17NO3 (7H<sup>+</sup>)

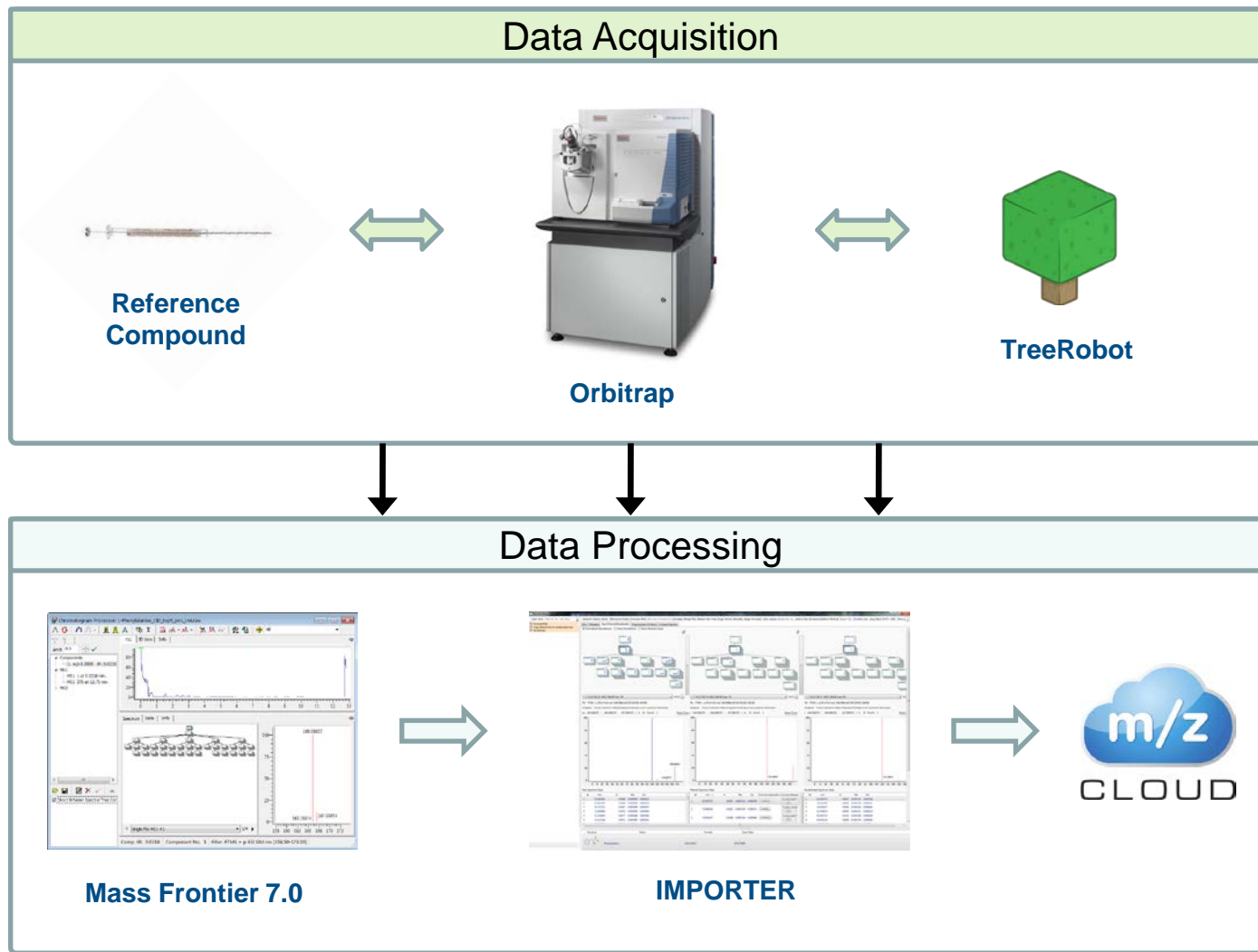
**Metadata:**

- Fragmentation Scheme

**Footer:**

- Install mzCloud app Desktop Application
- Disclaimer: The mzCloud is still undergoing engineering and development.

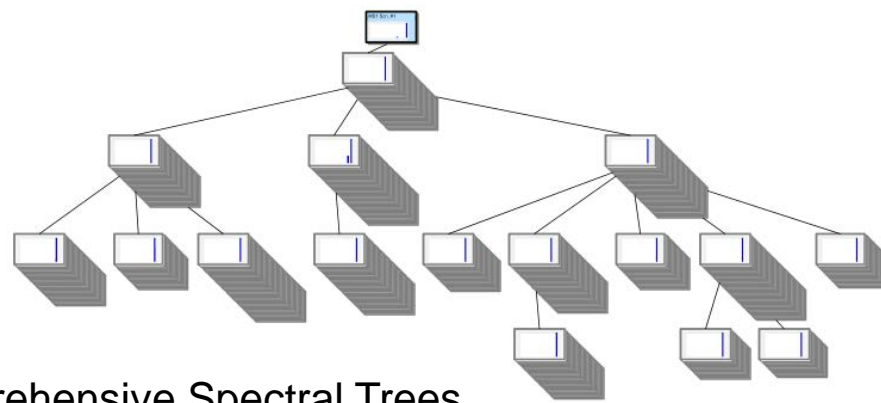
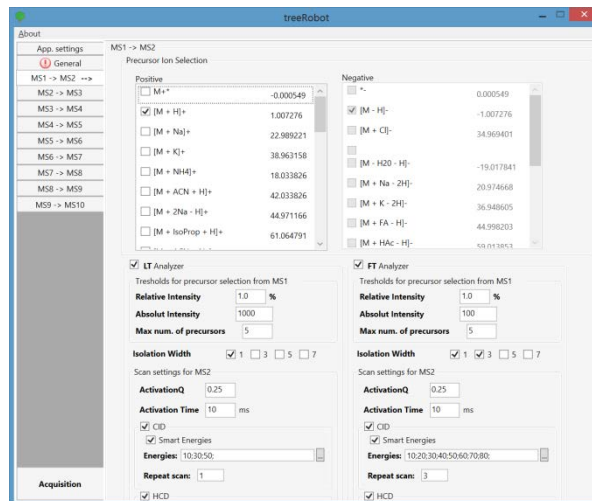
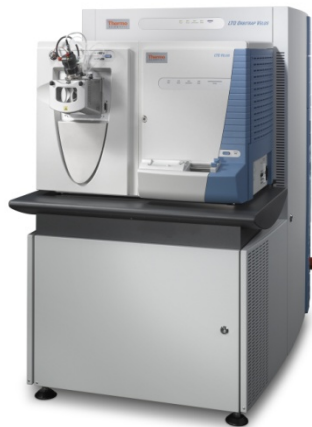




# Spectral Trees Acquisition

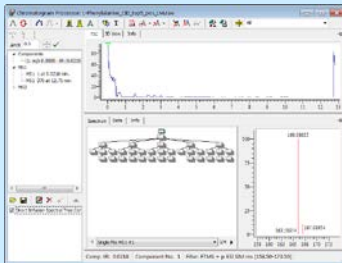
Orbitrap

TreeRobot



Comprehensive Spectral Trees

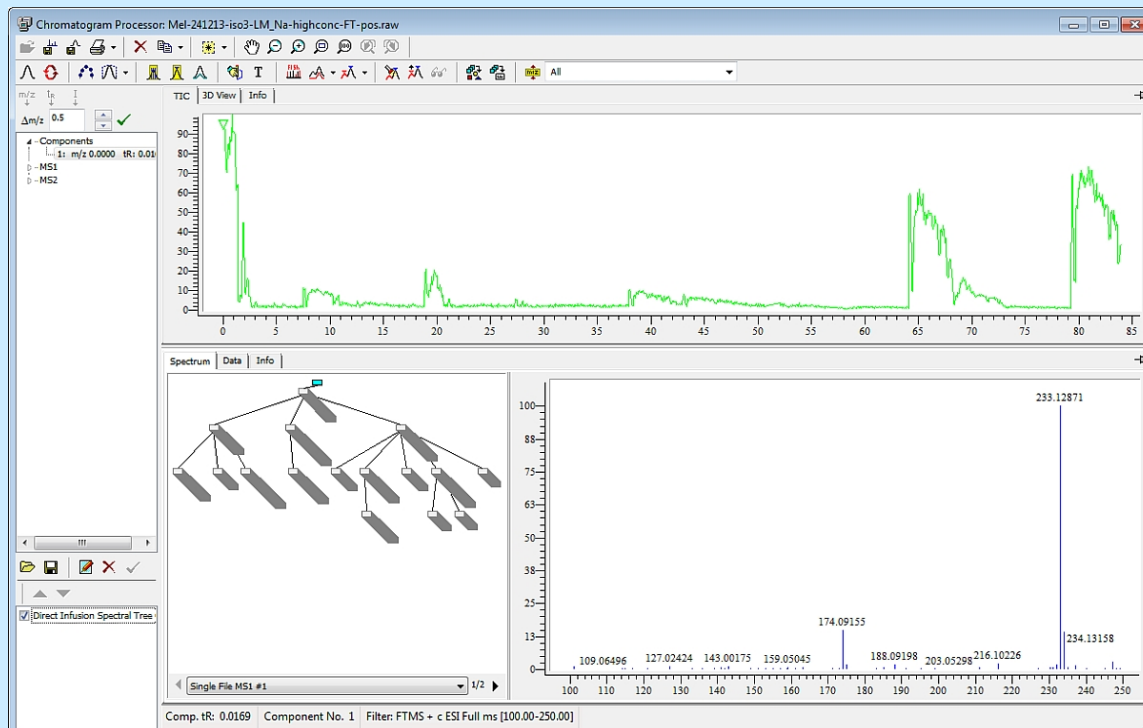
# mzCloud: Rigorously Curated Data



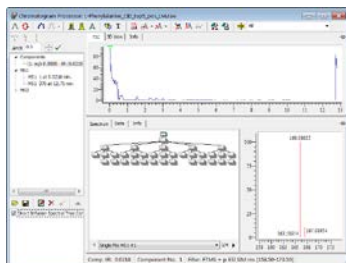
Mass Frontier 7.0



IMPORTER



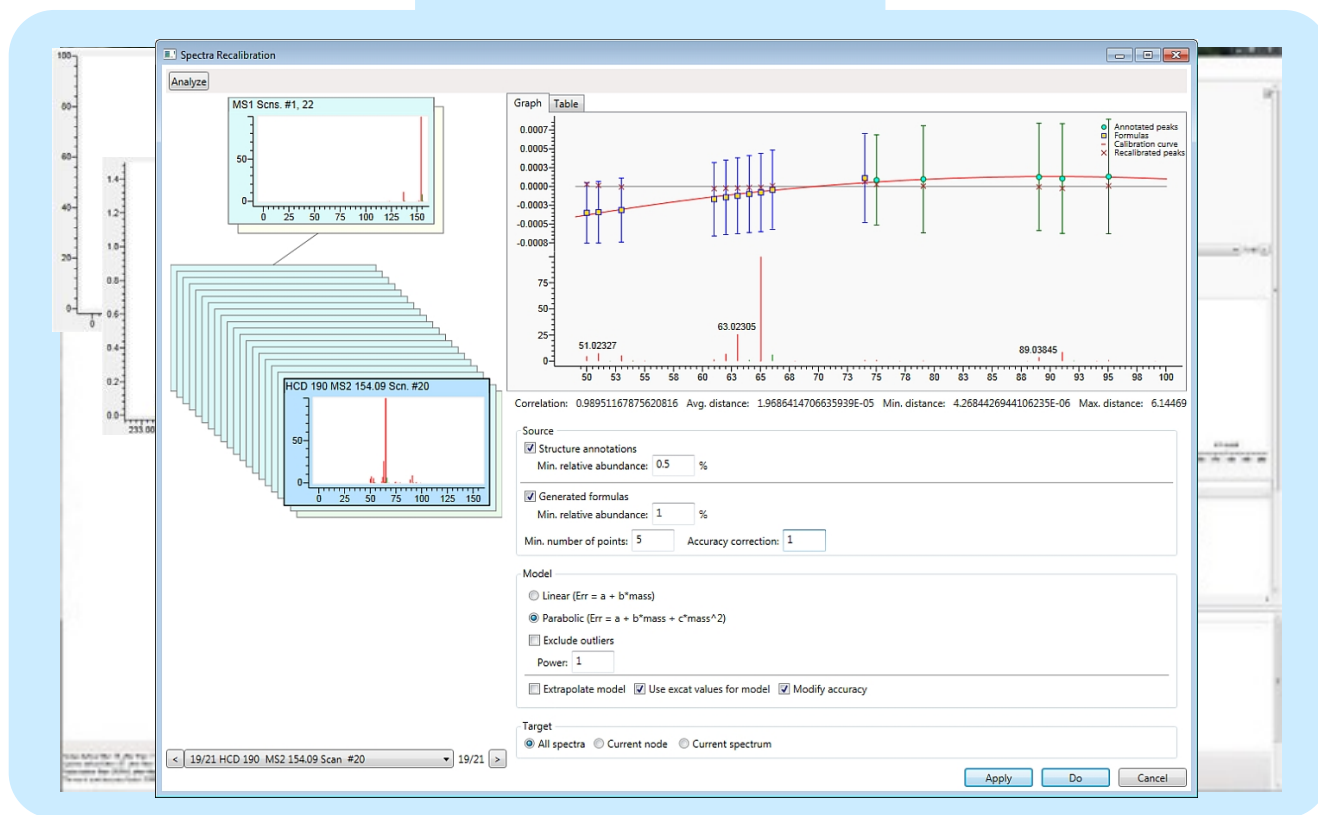
# mzCloud: Rigorously Curated Data



Mass Frontier 7.0

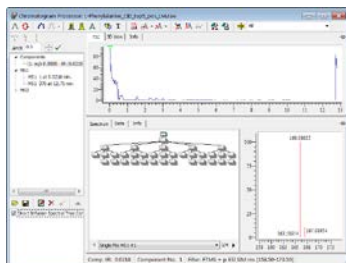


IMPORTER





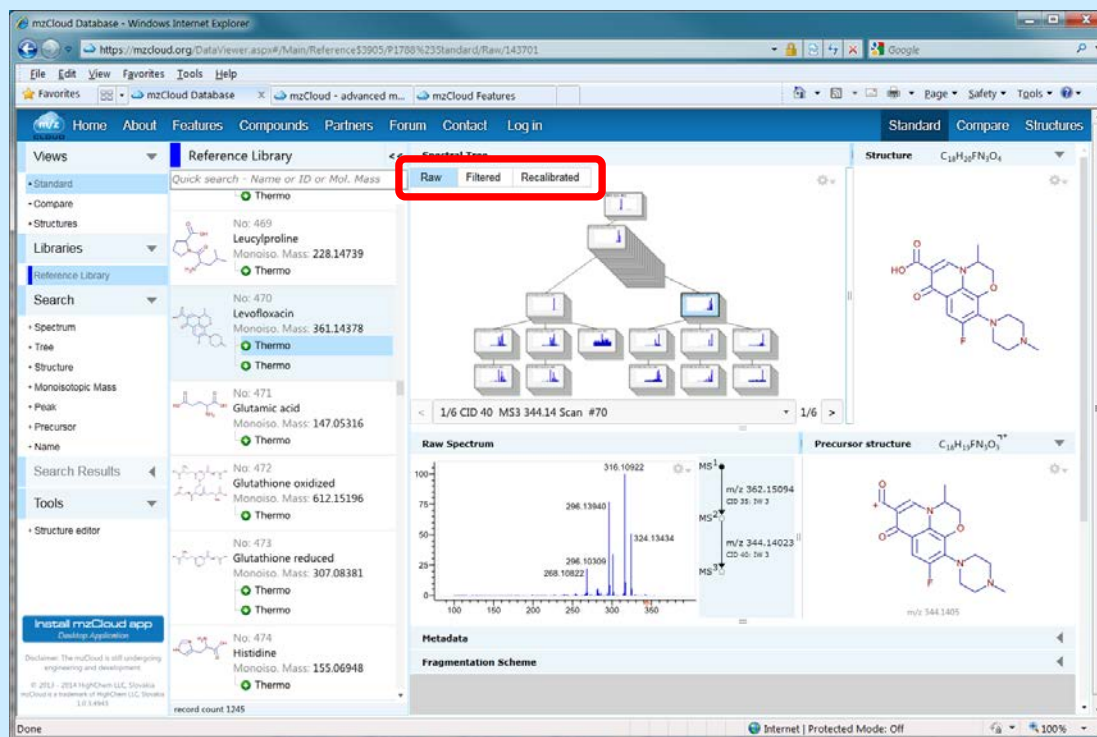
# mzCloud: Rigorously Curated Data



Mass Frontier 7.0



IMPORTER

mzCloud Database - Windows Internet Explorer

https://mzcloud.org/Data/Viewer.aspx#/Main/Reference5905/P1766%23Standard/Raw/143701

File Edit View Favorites Tools Help

Home About Features Compounds Partners Forum Contact Log in

Views Reference Library

Quick search - Name or ID or Mol. Mass

Raw Filtered Recalibrated

No: 469  
Leucylproline  
Monois. Mass: 228.14739

No: 470  
Levofloxacin  
Monois. Mass: 361.14378

No: 471  
Glutamic acid  
Monois. Mass: 147.05316

No: 472  
Glutathione oxidized  
Monois. Mass: 612.15196

No: 473  
Glutathione reduced  
Monois. Mass: 307.08381

No: 474  
Histidine  
Monois. Mass: 155.06948

Structure C<sub>14</sub>H<sub>19</sub>FN<sub>2</sub>O<sub>4</sub>

1/6 CID-40 MS3 344.14 Scan #70

Raw Spectrum

316.10922

296.13840

268.10822

296.10309

324.13434

344.14023

362.15094

MS<sup>1</sup>

MS<sup>2</sup>

MS<sup>3</sup>

Precursor structure C<sub>14</sub>H<sub>19</sub>FN<sub>2</sub>O<sub>4</sub><sup>+</sup>

m/z 344.1405

Metadata

Fragmentation Scheme

record count 1245

Internet | Protected Mode: Off





# mzCloud – Raw, Filtered and Recalibrated Spectra

mzCloud Database - Windows Internet Explorer

https://mzcloud.org/DataViewer.aspx#/Main/Reference\$3905/P1788%23Standard/Raw/143701

File Edit View Favorites Tools Help

Home About Features Compounds Partners Forum Contact Log in

Standard Compare Structures

Views Reference Library

Quick search

Thermo

Id: 469  
Leucylproline  
Monoisot. Mass: 228.14739

Thermo

Id: 470  
levofloxacin  
Monoisot. Mass: 361.14378

Thermo

Thermo

Id: 471  
glutamic acid  
Monoisot. Mass: 147.05316

Thermo

Id: 472  
glutathione oxidized  
Monoisot. Mass: 612.15196

Thermo

Id: 473  
glutathione reduced  
Monoisot. Mass: 307.08381

Spectral Tree

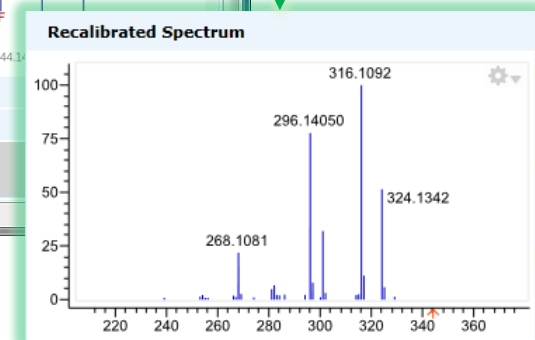
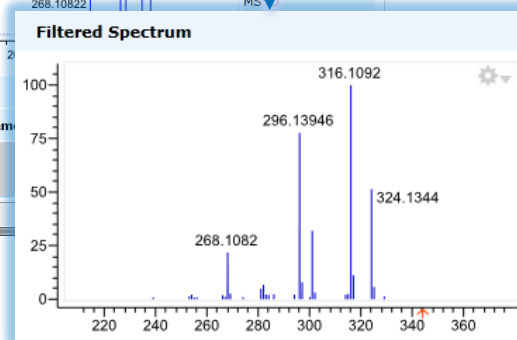
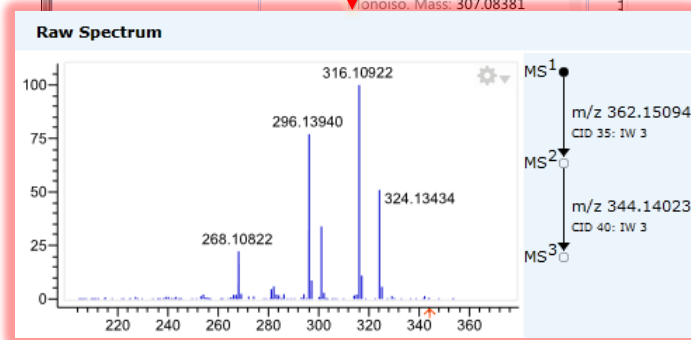
Raw Filtered Recalibrated

Structure C18H20FN3O4

1/6 CID 40 MS3 344.14 Scan #70

Raw Spectrum

Precursor structure C18H19FN3O3



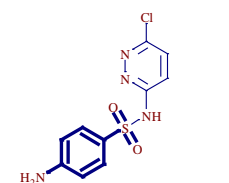
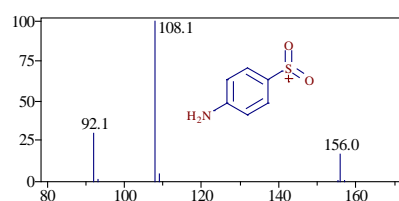
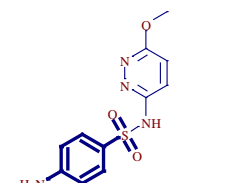
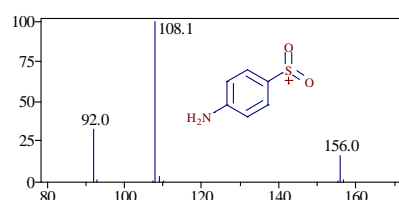
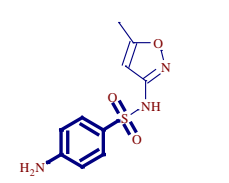
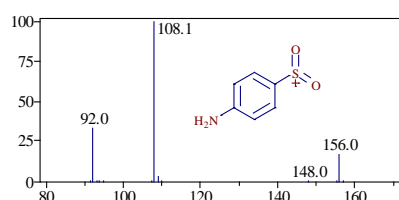
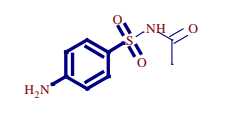
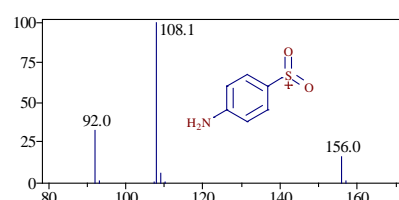
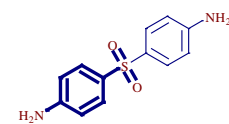
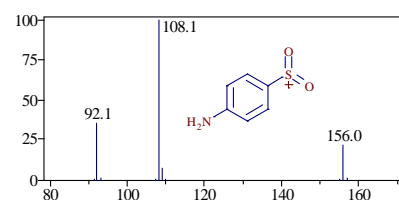
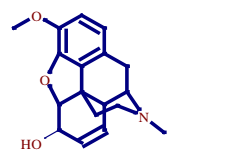
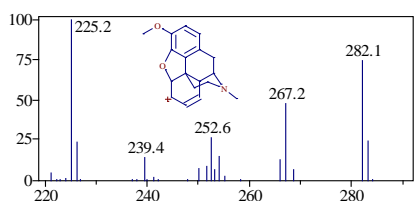
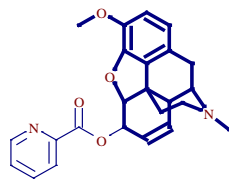
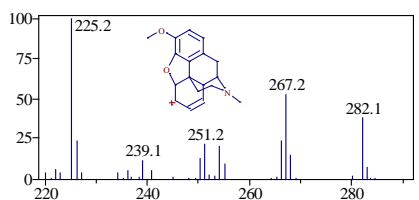
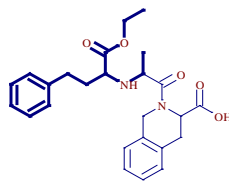
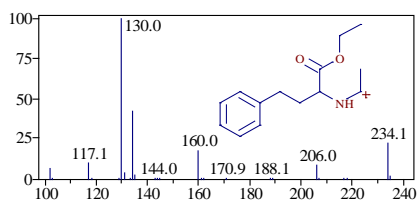
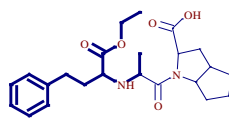
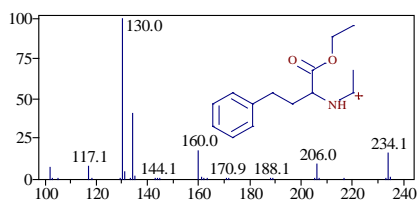
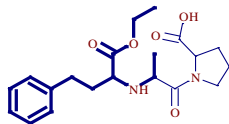
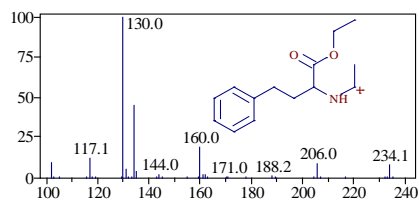


## Work in progress

- Fully searchable sample database
- Predicted fragments for suspect compounds
- Precursor Ion Fingerprinting
- Real time validation of search algorithm
- Breakdown curves



# Precursor Ion Fingerprinting (PIF)



mzCloud Database - Mozilla Firefox

mzCloud Database

https://mzcloud.org/DataViewer.aspx#/Main/SpectrumResult1\$2343/P1904%23Compare/Filtered/160498

Home About Features Compounds Partners Forum Contact Robert Mistrik Log out Standard Compare Structures

Views

- Standard
- Compare
- Structures

Libraries

- Reference Library
- Search

Search Results

- Spectrum search result 1

Tools

Spectrum search result 1

Query

Library record

Structure

Hit: 1 Best Match: 870  
Reference No: 28  
**Quercetin**  
Monoiso. Mass: 302.04265  
Thermo

Hit: 2 Best Match: 492  
Reference No: 969  
**5,7,3',4',5'-Pentahydroxyflavone**  
Monoiso. Mass: 302.04265  
UC Davis

Hit: 3 Best Match: 329  
Reference No: 455  
**Isorhamnetin**  
Monoiso. Mass: 316.05830  
Thermo

Hit: 4 Best Match: 158  
Reference No: 830  
**AICA ribonucleotide**  
Monoiso. Mass: 338.06275  
Thermo

Install mzCloud app Desktop Application

Disclaimer: The mzCloud is still undergoing engineering and development.

© 2013 - 2014 HighChem LLC, Slovakia  
mzCloud is a trademark of HighChem LLC, Slovakia  
1.0.3.4945

record count 4

Library Metadata

Spectra compare

Precursors and Fragment

MS<sup>1</sup> m/z 303.04968  
CID 50: IW 3

Algorithm match  
HighRes 870  
NIST 747  
LowRes 930  
862  
M: -0.869569118969498

MS<sup>1</sup> m/z 611.16205  
CID 35: IW 2  
MS<sup>2</sup> m/z 303.0509  
CID 25: IW 2

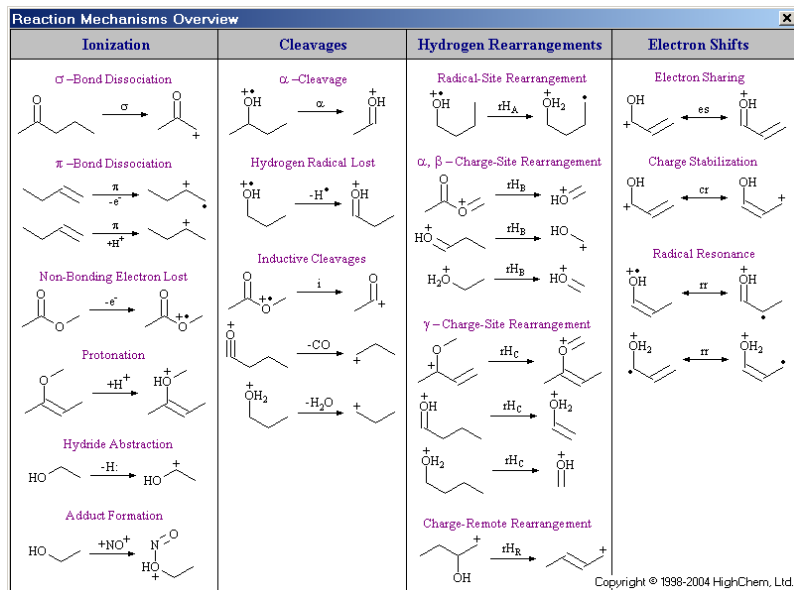
OH  
OH

Rutin



# Knowledge based prediction of fragmentation pathways

## General fragmentation rules



## Fragmentation Library

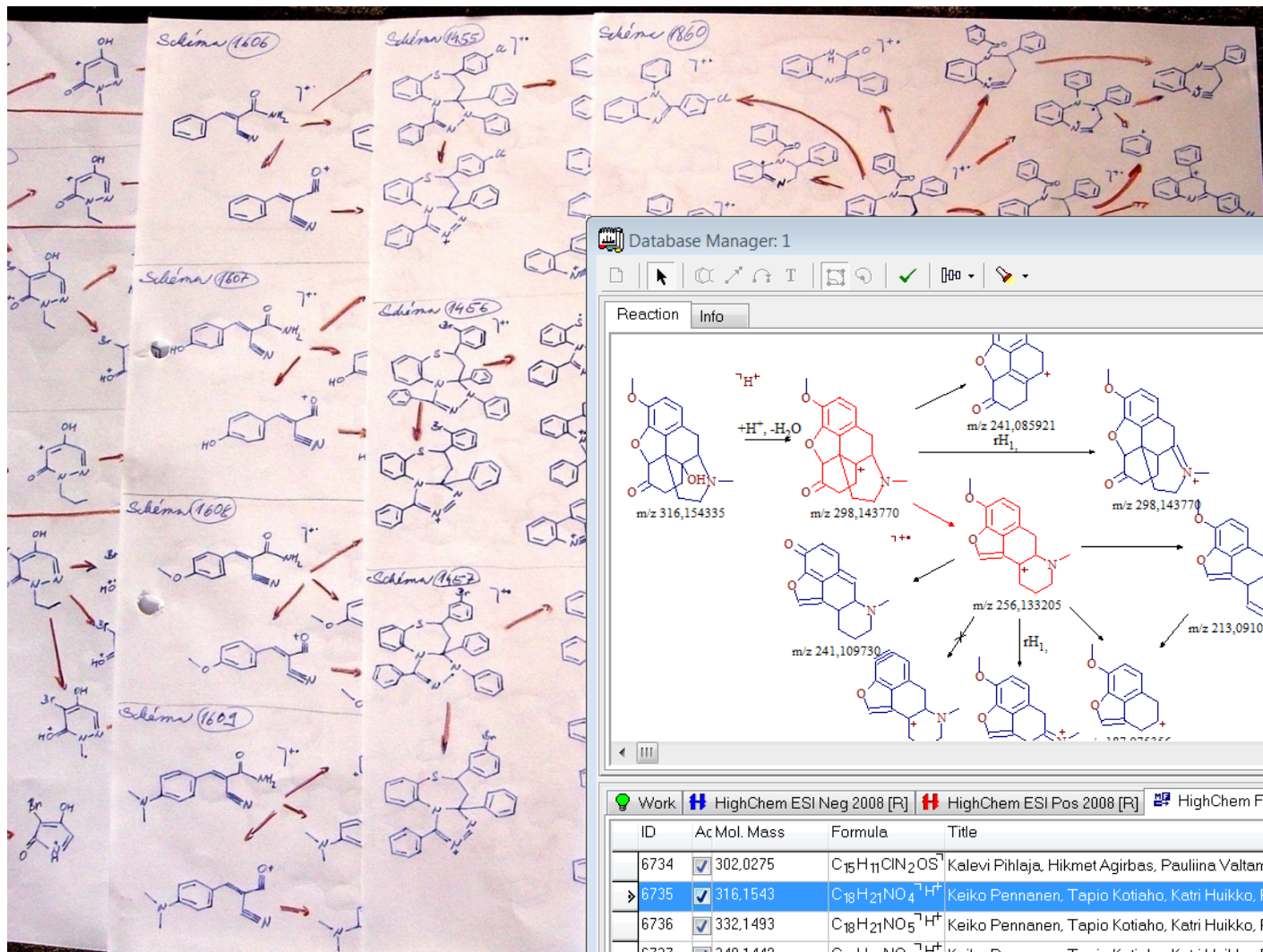
Total number of	Mass Frontier 7.0
Fragmentation Schemes	30.936
Individual Reactions	129.229
Chemical Structures	151.762

	Source	Volume	Year
1.	JASMS (Journal of the American Society for Mass Spectrometry)	1-17	1990-2006
2.	IJMSP (International Journal of Mass Spectrometry and Ion Physics)	1-53	1968-1983
	IJMSP (International Journal of Mass Spectrometry and Ion Processes)	54-175	1983-1998
	IJMS (International Journal of Mass Spectrometry)	176-255	1998-2006
3.	RCM (Rapid Communications in Mass Spectrometry)	1-20	1987-2006
4.	JMS (Journal of Mass Spectrometry)	30-41	1995-2006
5.	OMS (Organic Mass Spectrometry)	1-29	1968-1994
6.	JMSSJ (Journal of the Mass Spectrometry Society of Japan)	11-27 29-30 37-48 50-53	1964-1979 1981-1982 1989-2000 2002-2005
7.	MSR (Mass Spectrometry Reviews)	1-25	1981-2006
8.	EJMBSMER (European Journal of Mass Spectrometry in Biochemical, Medicine, and Environmental Research)	1-2	1980-1982
9.	BMS (Biomedical Mass Spectrometry) BEMS (Biomedical and Environmental Mass Spectrometry) BMS (Biological Mass Spectrometry)	1-12 14-19 20-23	1974-1985 1987-1990 1991-1994
10.	JC (Journal of Chromatography)	181-536	1980-1991
11.	EJMS (European Journal of Mass Spectrometry)	4	1998





# Knowledge based prediction of fragmentation pathways



Database Manager: 1

Reaction Info

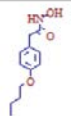
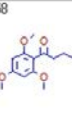
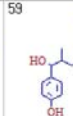
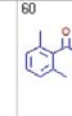
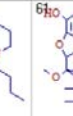

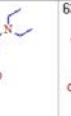
Work HighChem ESI Neg 2008 [R] HighChem ESI Pos 2008 [R] HighChem Fragmentation Library [R, CP]

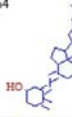

ID	Ac Mol. Mass	Formula	Title
6734	302,0275	C <sub>15</sub> H <sub>11</sub> ClN <sub>2</sub> O <sub>5</sub> <sup>1-</sup>	Kalevi Pihlaja, Hikmet Agirbas, Pauliina Valtamo
6735	316,1543	C <sub>18</sub> H <sub>21</sub> NO <sub>4</sub> <sup>1+</sup> H <sup>+</sup>	Keiko Pennanen, Tapio Kotiaho, Katri Huikko, Risto Kostainen
6736	332,1493	C <sub>18</sub> H <sub>21</sub> NO <sub>5</sub> <sup>1+</sup> H <sup>+</sup>	Keiko Pennanen, Tapio Kotiaho, Katri Huikko, Risto Kostainen
6737	348,1442	C <sub>18</sub> H <sub>21</sub> NO <sub>6</sub> <sup>1+</sup> H <sup>+</sup>	Keiko Pennanen, Tapio Kotiaho, Katri Huikko, Risto Kostainen

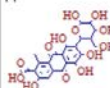
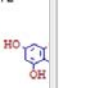




# Knowledge based prediction of fragmentation pathways

57  58  59  60  61  62  63 

64  65 

71  72 

File MS1  
272.27

File MS2 272.30  
215.14

File MS3 215.10  
147.12

File MS3 215.10 1/2

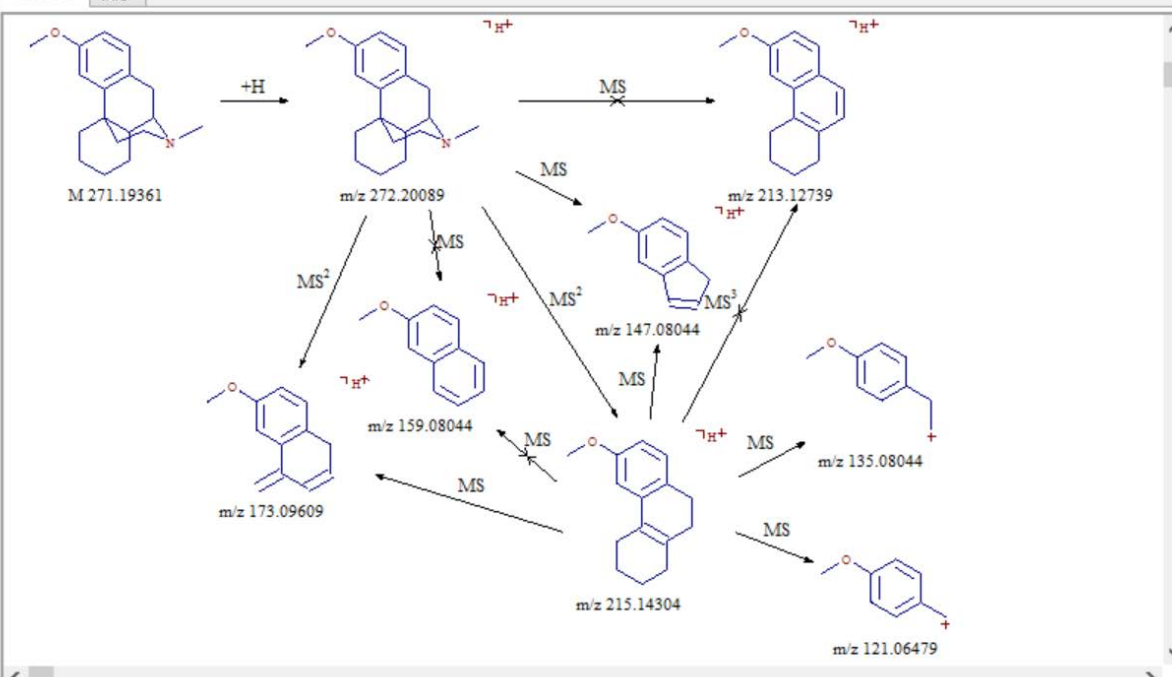
Spectrum Info Data Mass Differences Compare Spectra Compare Trees

Filter: + c ESI Full ms3 272.30@43.00 215.10@39.00 | 55.00

100 147.12

Database Manager: 1

Reaction Info



M 271.19361

$+H$

$m/z$  272.20089

MS

$m/z$  213.12739

MS

$m/z$  147.08044

MS<sup>2</sup>

$m/z$  159.08044

MS<sup>2</sup>

$m/z$  173.09609

MS

$m/z$  215.14304

MS

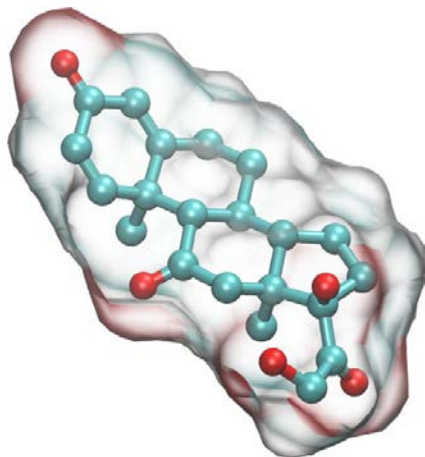
$m/z$  135.08044

MS

$m/z$  121.06479

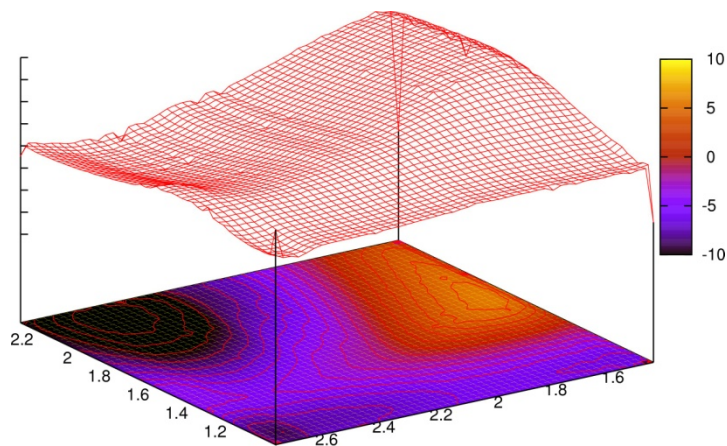


# Evaluation of fragmentation pathways by high-level quantum-mechanical calculations based on DFT method

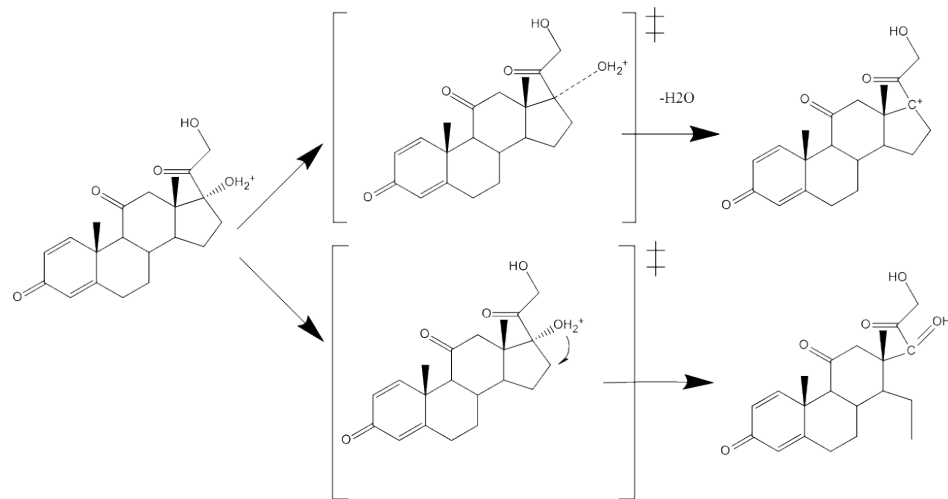


*Prednisone*

(17,21-dihydroxypregna-1,4-diene-3,11,20-trione)



- 3D conformer generator for fragments of concern
- Construction of competitive reaction mechanism
- Calculation of potential energy surface (PES)
- Transition state localization and vibrational analysis
- Reaction rate prediction based on RRKM method



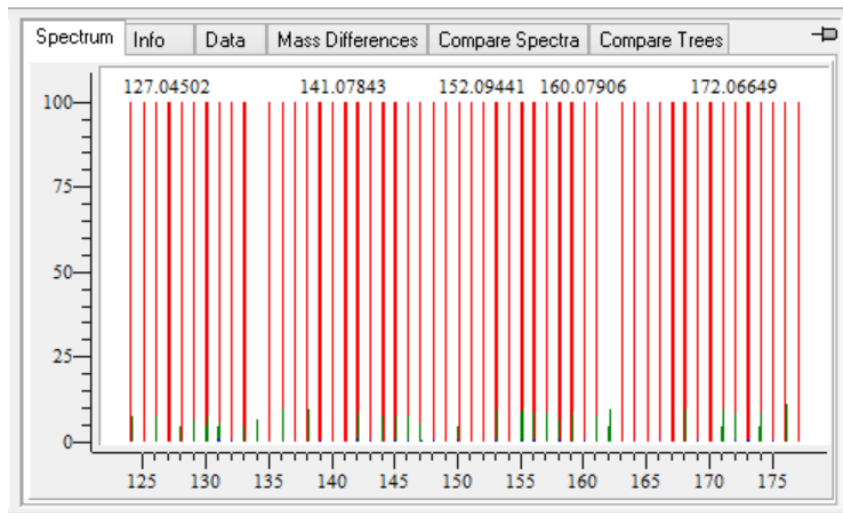
*Example for competitive reaction in the fragmentation pathway*

*PES for corresponding reaction, followed by transition state localization and reaction rate calculation*

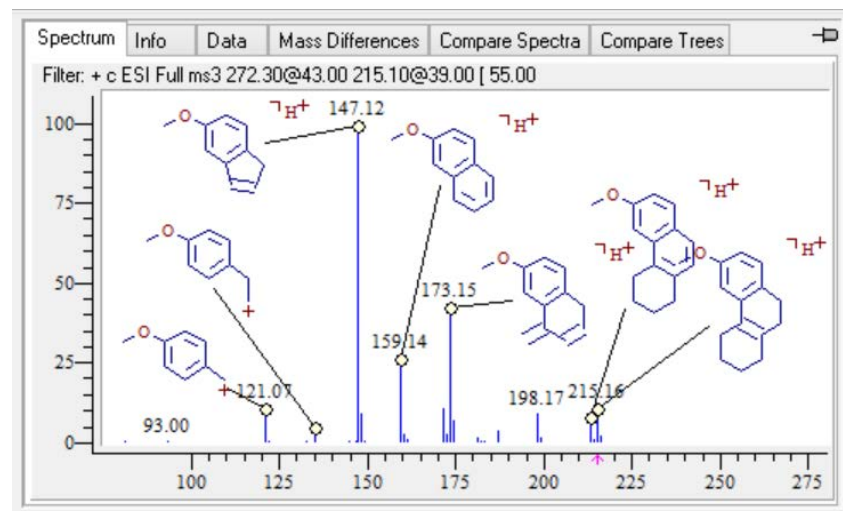




## Use of Predicted Fragments

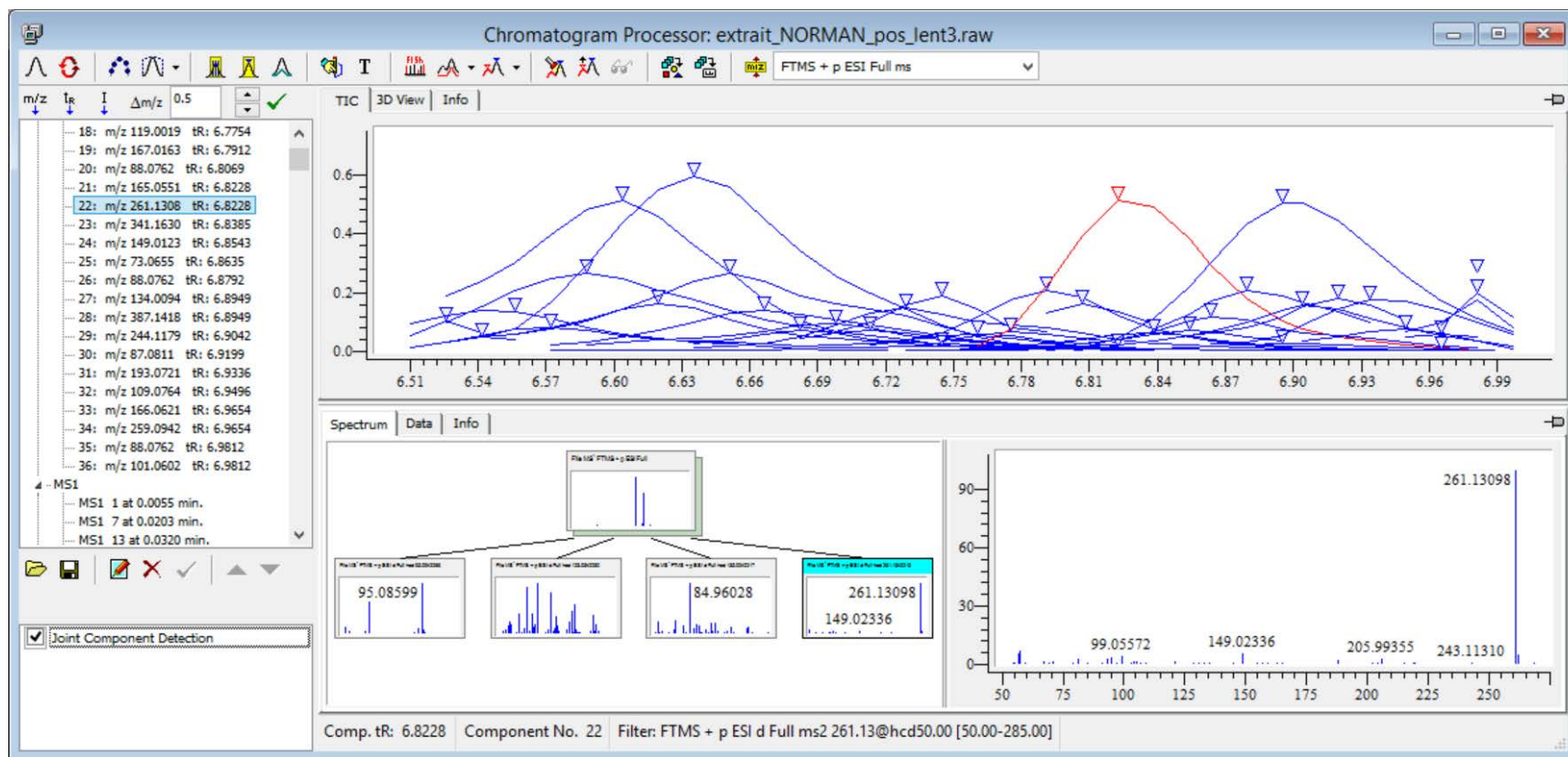


- **Matching *in silico* m/z values against experimental spectra**



- **Fragment annotation**

# What are the compounds?



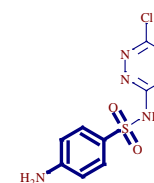
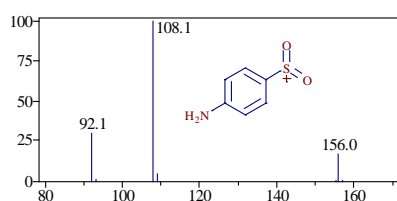
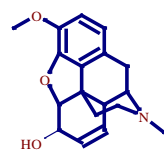
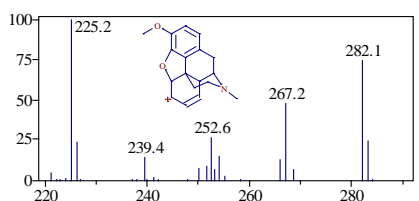
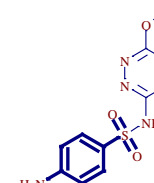
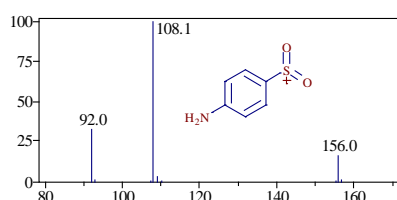
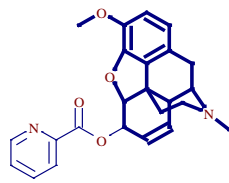
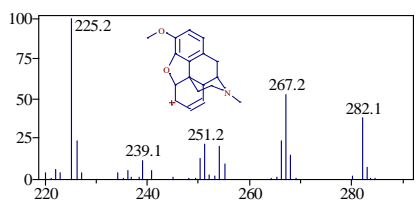
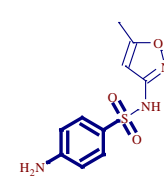
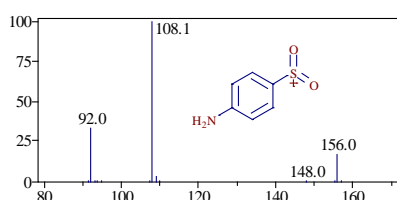
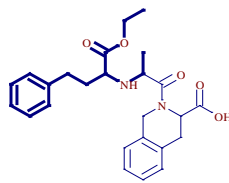
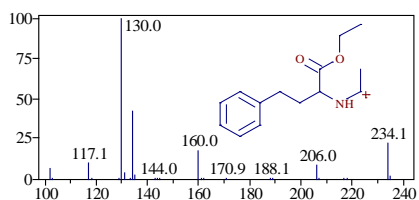
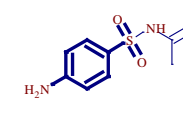
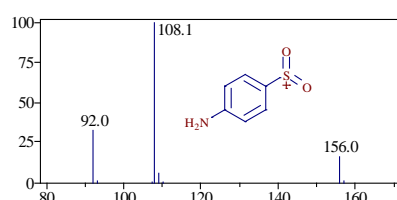
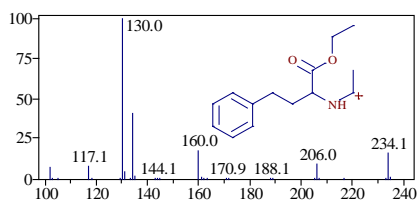
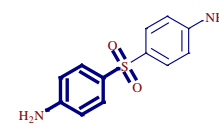
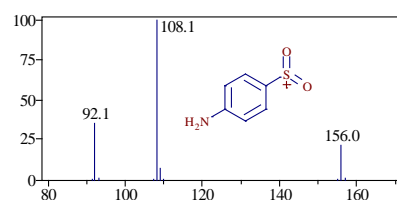
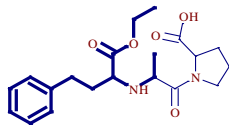
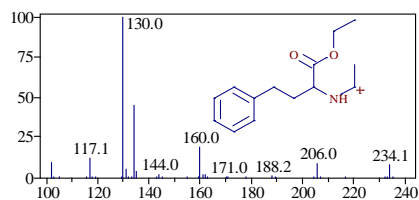
## BINGO Initiative

- Systematic identification of unknowns
- Inspired by Model Organism Metabolome initiative
- Advanced strategy
- Focused on limited number of samples
- Smart standards selection and acquisition
- Collaborative effort
- H2020 project proposal
- Everyone is welcome to join ([robert.mistrik@highchem.com](mailto:robert.mistrik@highchem.com))

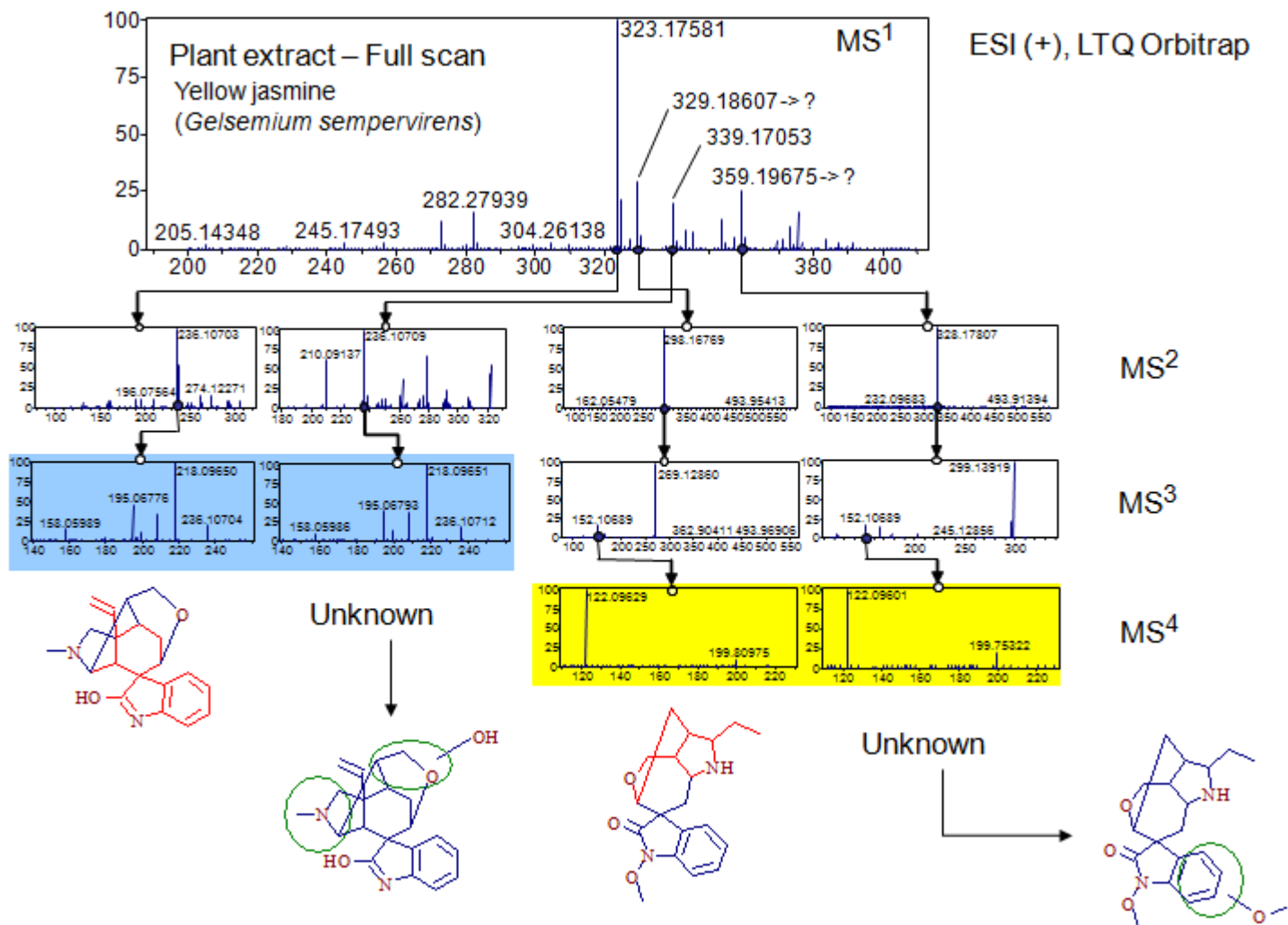
**Let's make our dream come true**

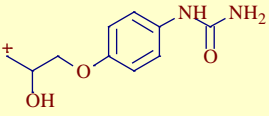
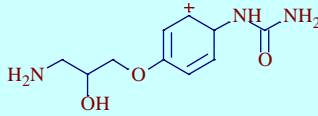
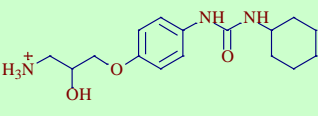
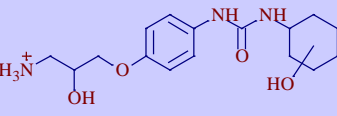
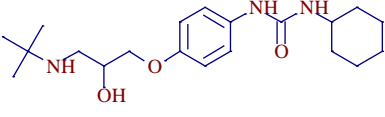
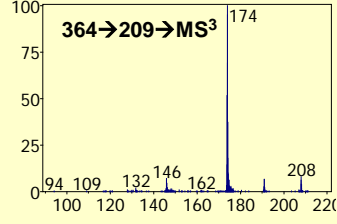
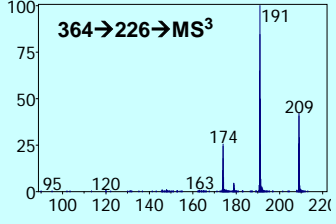
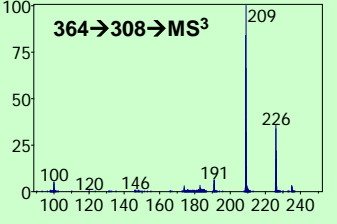
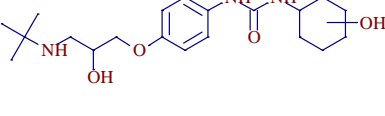
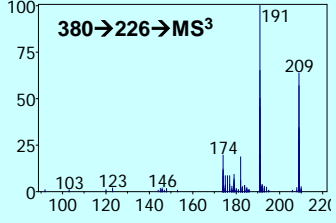
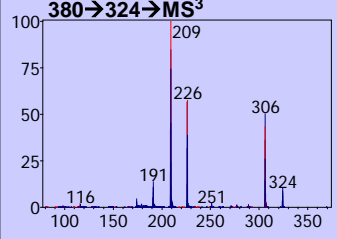
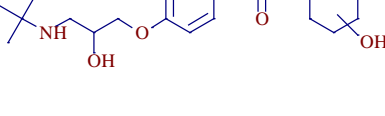
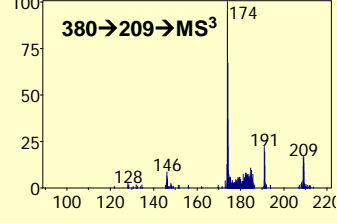
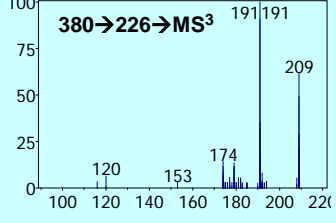
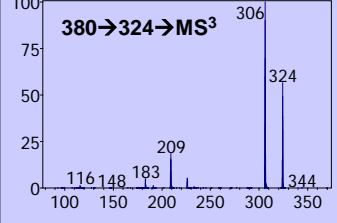
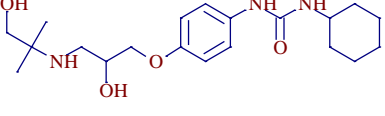
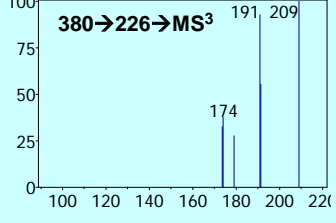
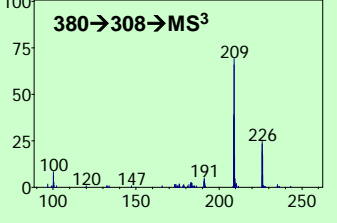


# Precursor Ion Fingerprinting (PIF)

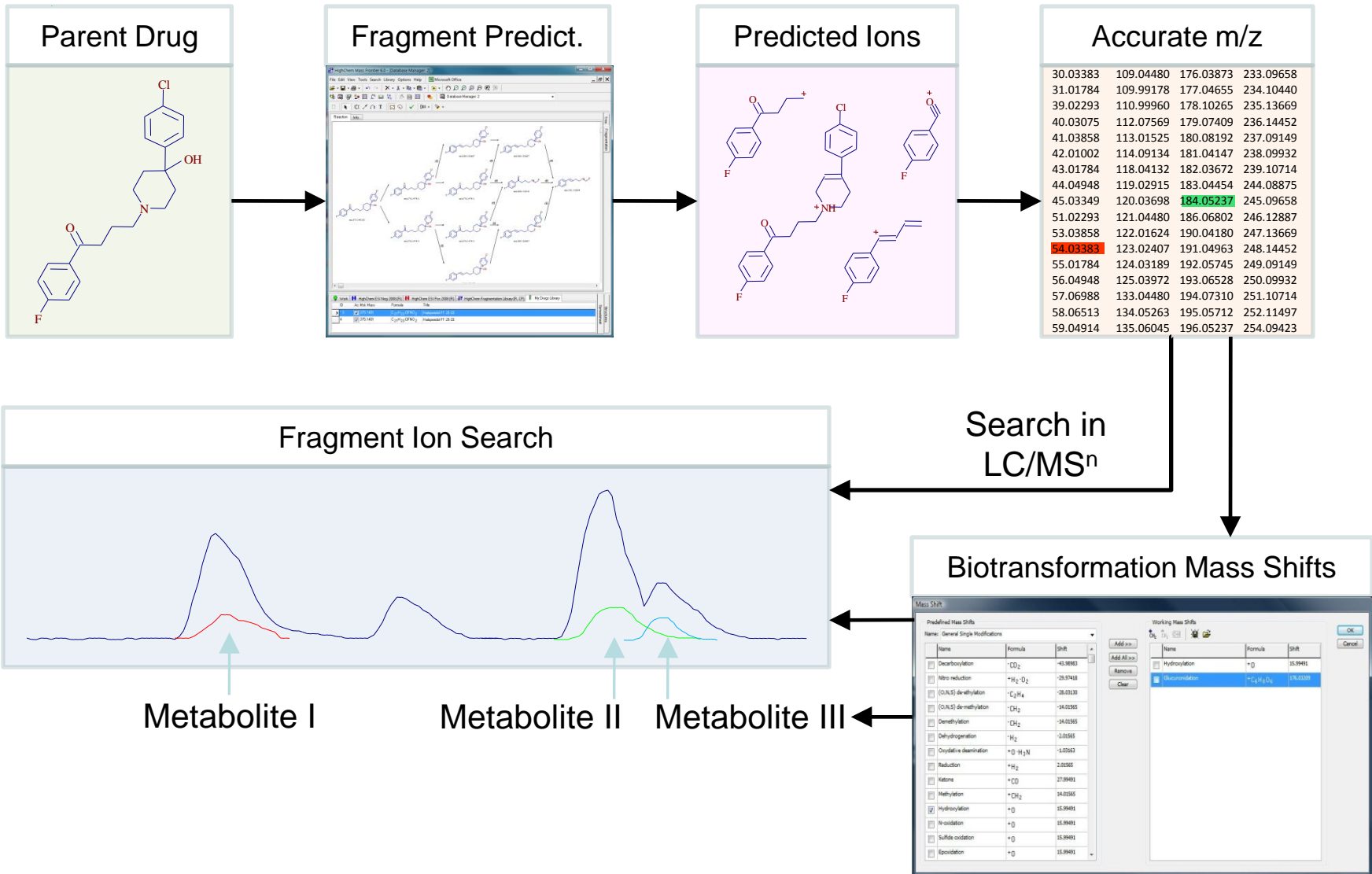


# Gelsemine Metabolite Identification by PIF

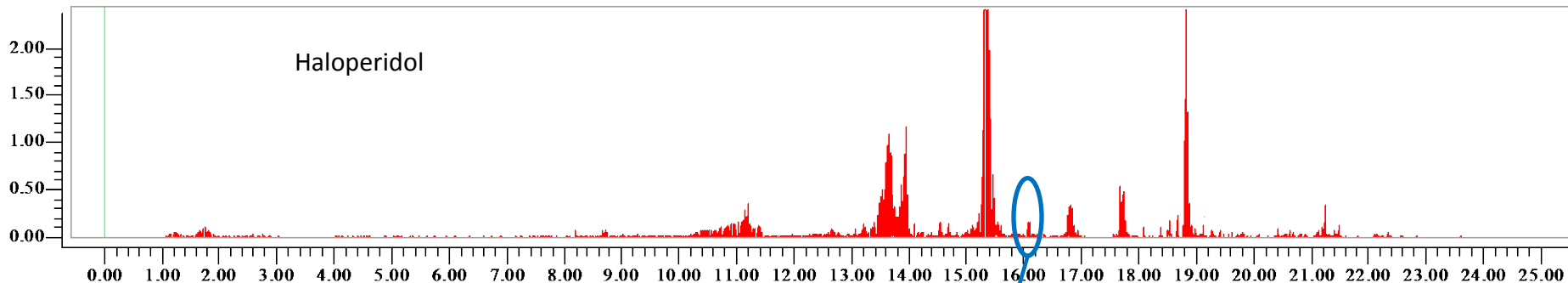


		m/z 209	m/z 226	m/z 308	m/z 324
	Metabolite Structure Assignment				
Talinolol					
Metabolite C01					
Metabolite C02					
Metabolite C09					<p>Intensity differences due to different hydroxyl group position</p>

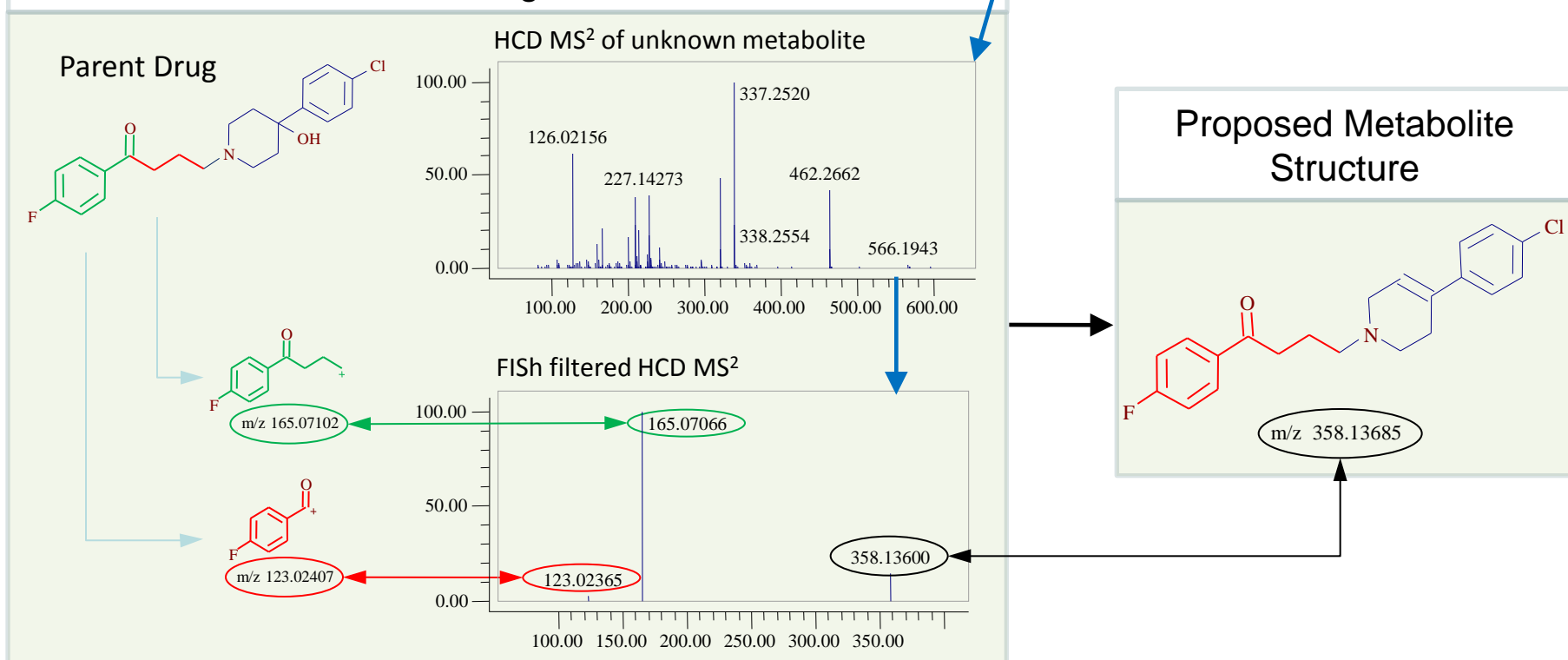
# Fragment Ion Search (FISh)



# Metabolite identification using FISh



## Common Ions of Parent Drug and Unknown Metabolite





# Metabolite identification using FISh

The image shows a screenshot of the HighChem Mass Frontier 7.0 software interface. The main window displays a chemical structure editor with a red chemical structure of a metabolite. The structure consists of a piperidine ring substituted with a 4-chlorophenyl group and a 4-fluorophenyl group, connected via a propyl chain to a carbonyl group. The molecular formula is  $C_{21}H_{23}ClFNO_2$  and the mass-to-charge ratio is  $m/z$  375.14013. A red arrow points from the structure editor to a large grey area, which then has a red arrow pointing down.

HighChem Mass Frontier 7.0 Fundament 1

File Edit View Tools Search Library Options Help Microsoft Office

Chromatogram Processor: HaloperidoL\_bile\_Top1\_MS3\_PL.RAW

Structure Editor: 6

Unspecified Charge Site: none

$C_{21}H_{23}ClFNO_2$   $m/z$  375.14013

# Acknowledgment

- **Juraj Lutisan, Milos Suchy, Alena Bednarikova, Andrea Belicova, Jakub Mezey, Martin Mandak, Michal Raab**  
HighChem, Bratislava, Slovakia
- **Ernst Pittenauer, Gunter Allmaier**  
Vienna University of Technology, Vienna, Austria
- **Yingying Huang, Mark Sanders, Tim Stratton, Junhua Wang, David Peake, Ralf Tautenhahn, Lukas Najdekr, Kate Comstock**  
Thermo Fisher Scientific, San Jose, CA

## Initial mzCloud data set

- **Eric Genin, Emma Schymanski, Heinz Singer, Arpana Vaniya, Vladimir Patoprsty, Silvia. Vlckova, Wei Wei, Christophe Junot**