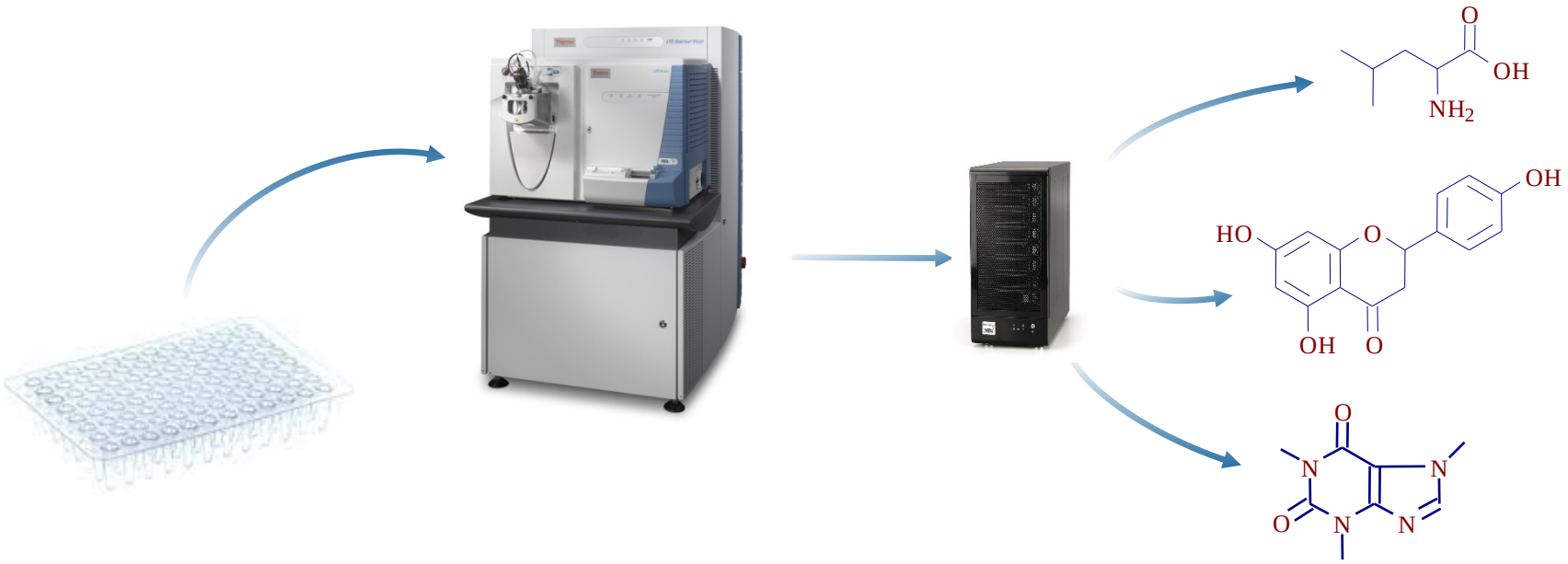
# Advanced Identification of Unknown Compounds



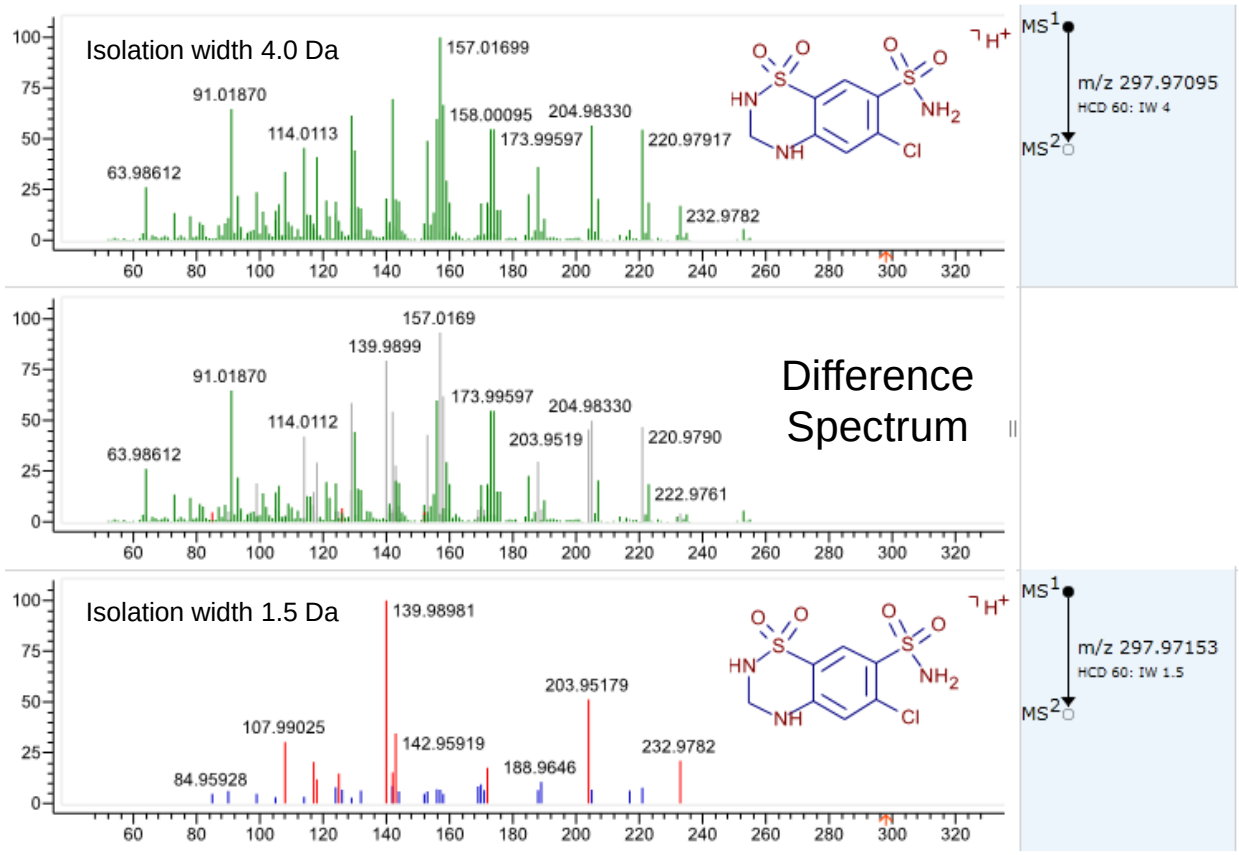

Robert Mistrik
HighChem
Bratislava, Slovakia

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

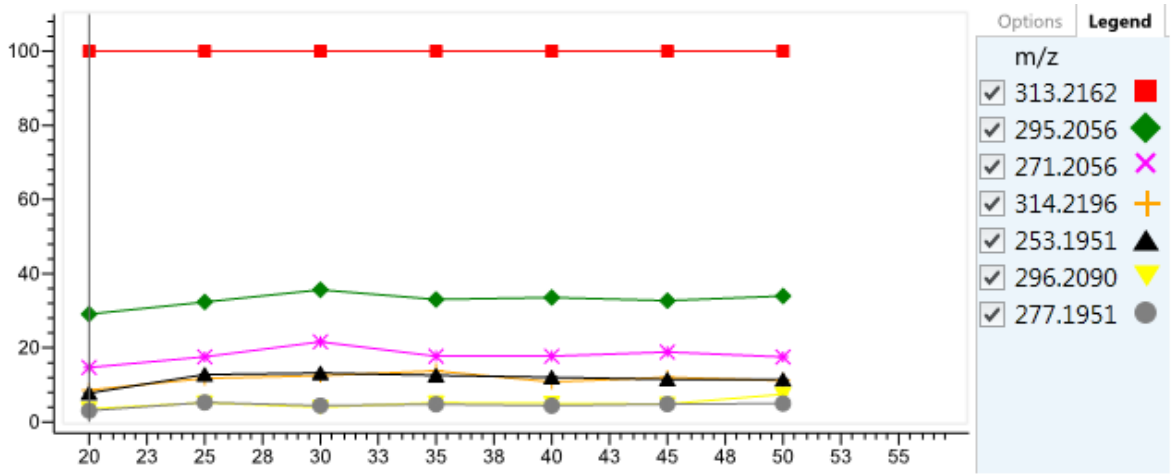



Spectrum courtesy of M. Stravs, E. Schymanski and H. Singer, EAWAG

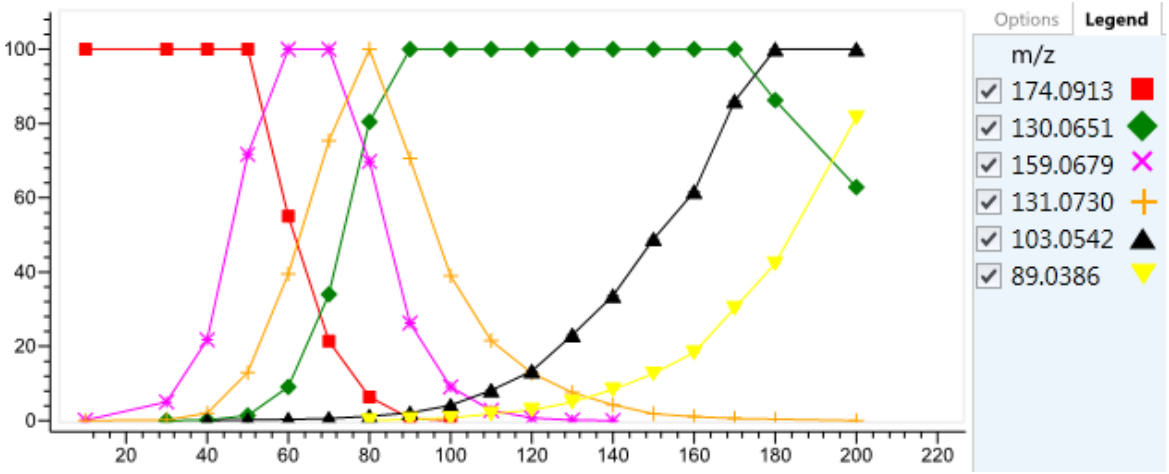
- **Spectra reproducibility**
  **Experimental variability**
- Spectra quality
- Spectra specificity
- Information ambiguity
- Isomers differentiation
- Structural diversity
- Accuracy issues
- Spectrum/Structure relationship

# Problems Associated with LCMS Library Search

17α Hydroxyprogesterone, CID $MS^2$ $[M+H]^+$



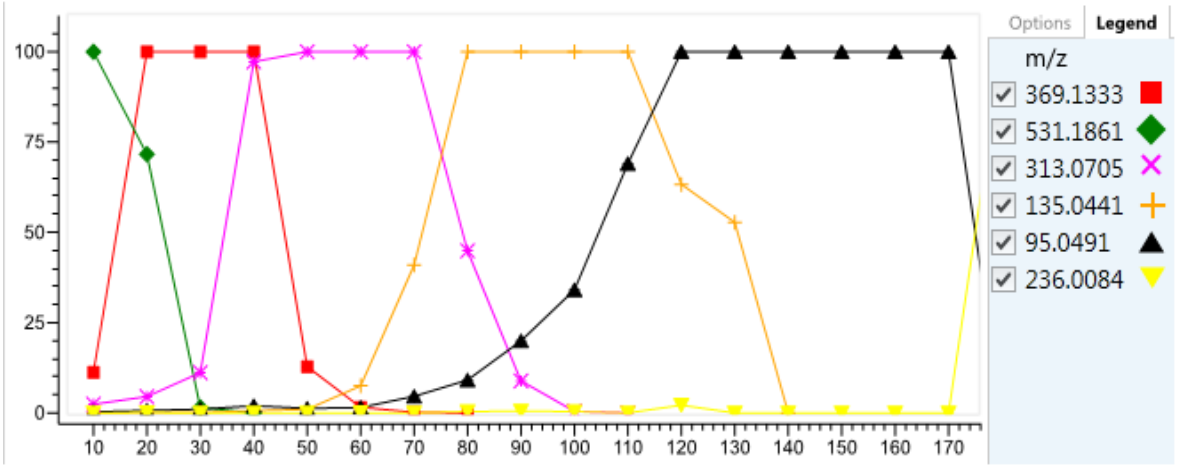

Serotonin, CID $MS^3$ $[M+H]^+ \rightarrow$ m/z 174.09134



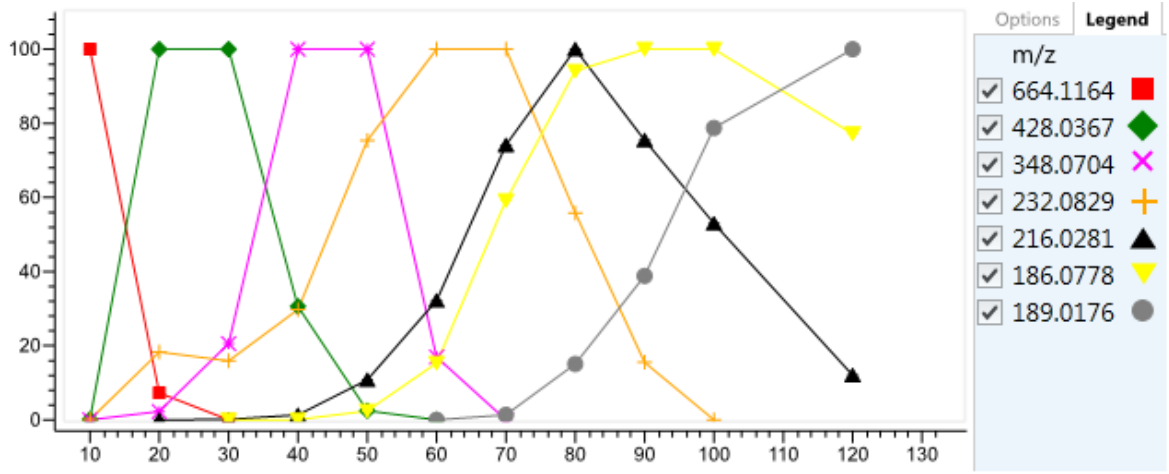

- **Spectra reproducibility**
  **Experimental variability**
- Spectra quality
- Spectra specificity
- Information ambiguity
- Isomers differentiation
- Structural diversity
- Accuracy issues
- Spectrum/Structure relationship

# Problems Associated with LCMS Library Search

Epimedin C, HCD $MS^2$ $[M+H]^+$

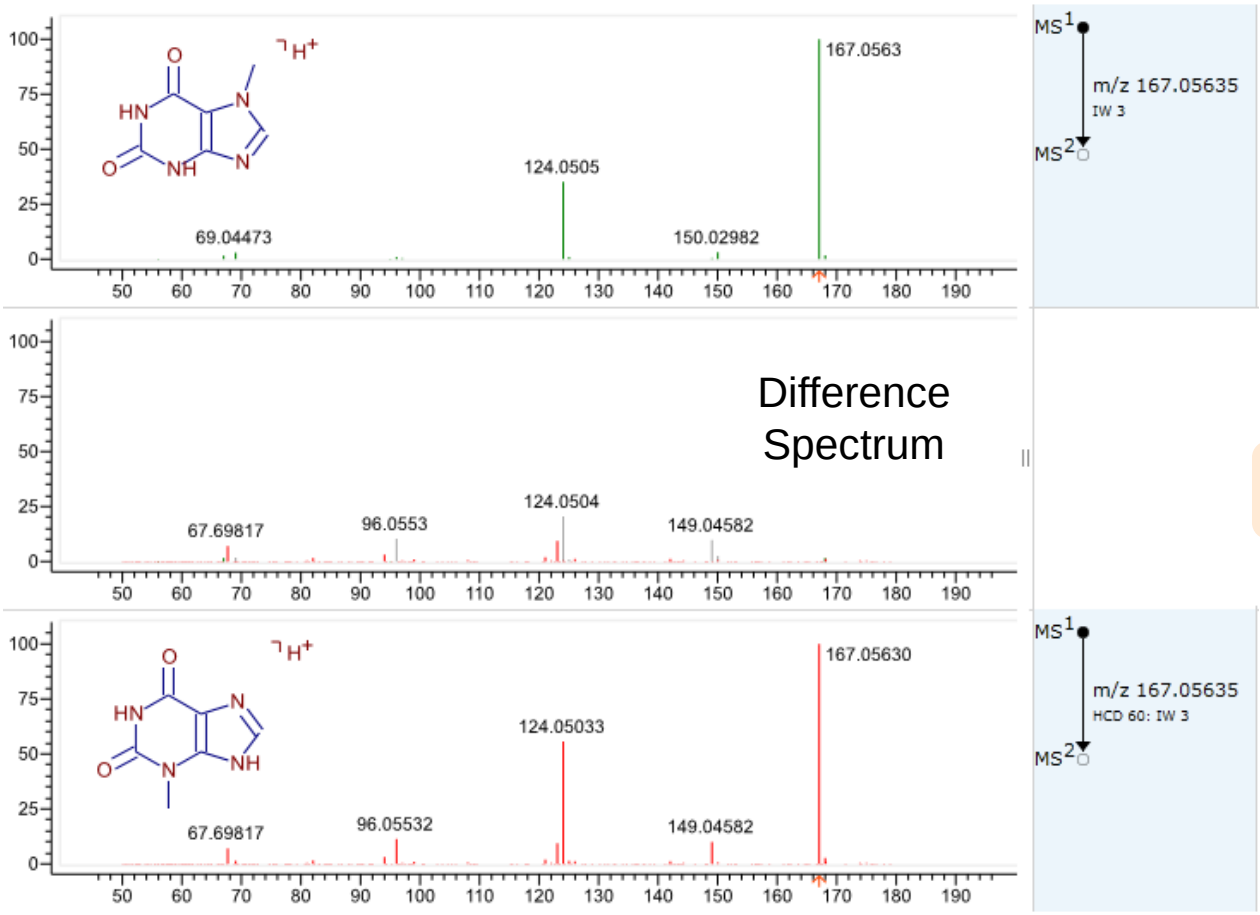



NAD+, HCD MS2 $[M+H]^+$



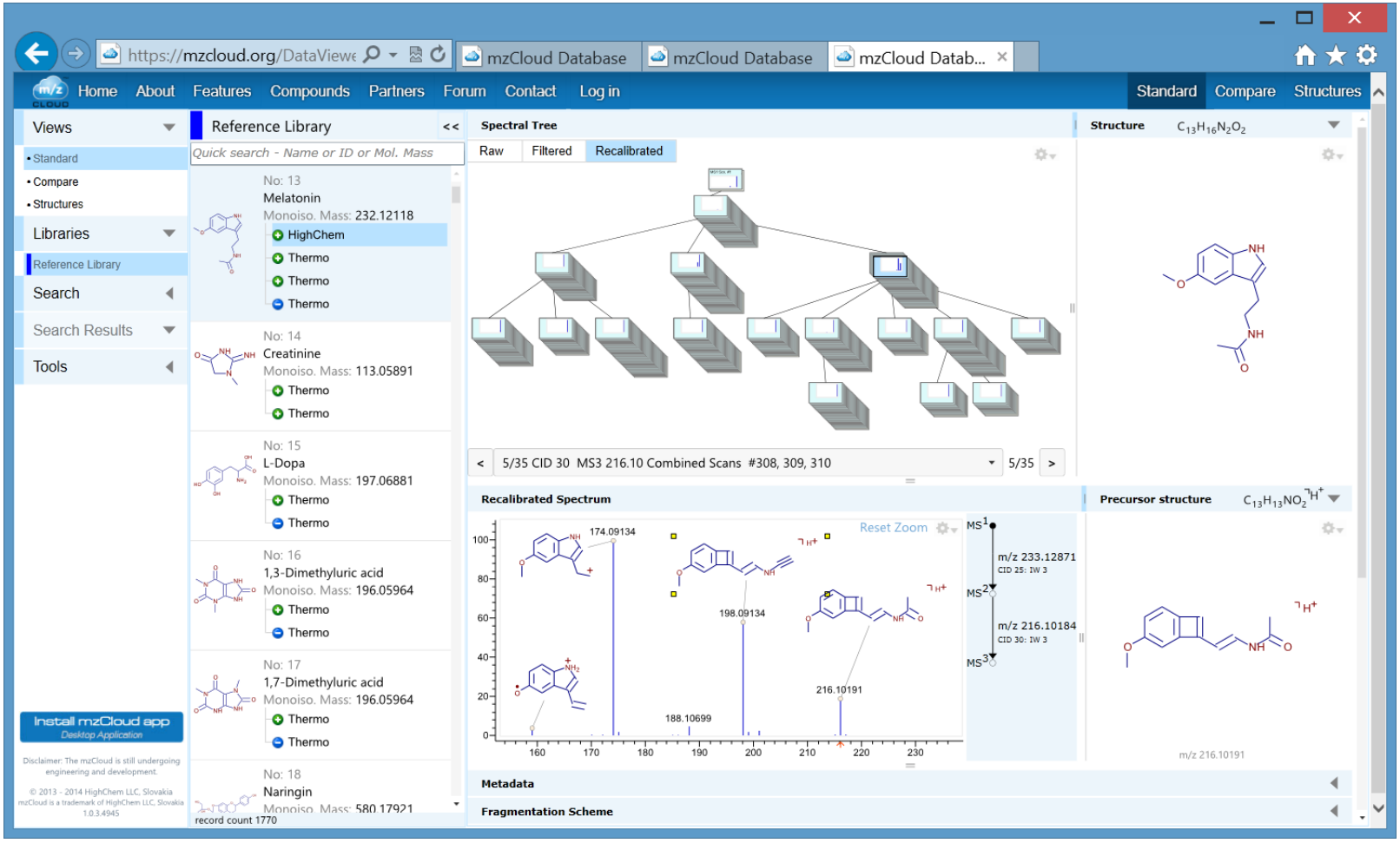

- **Spectra reproducibility**
  **Experimental variability**
- Spectra quality
- Spectra specificity
- Information ambiguity
- Isomers differentiation
- Structural diversity
- Accuracy issues
- Spectrum/Structure relationship

# 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

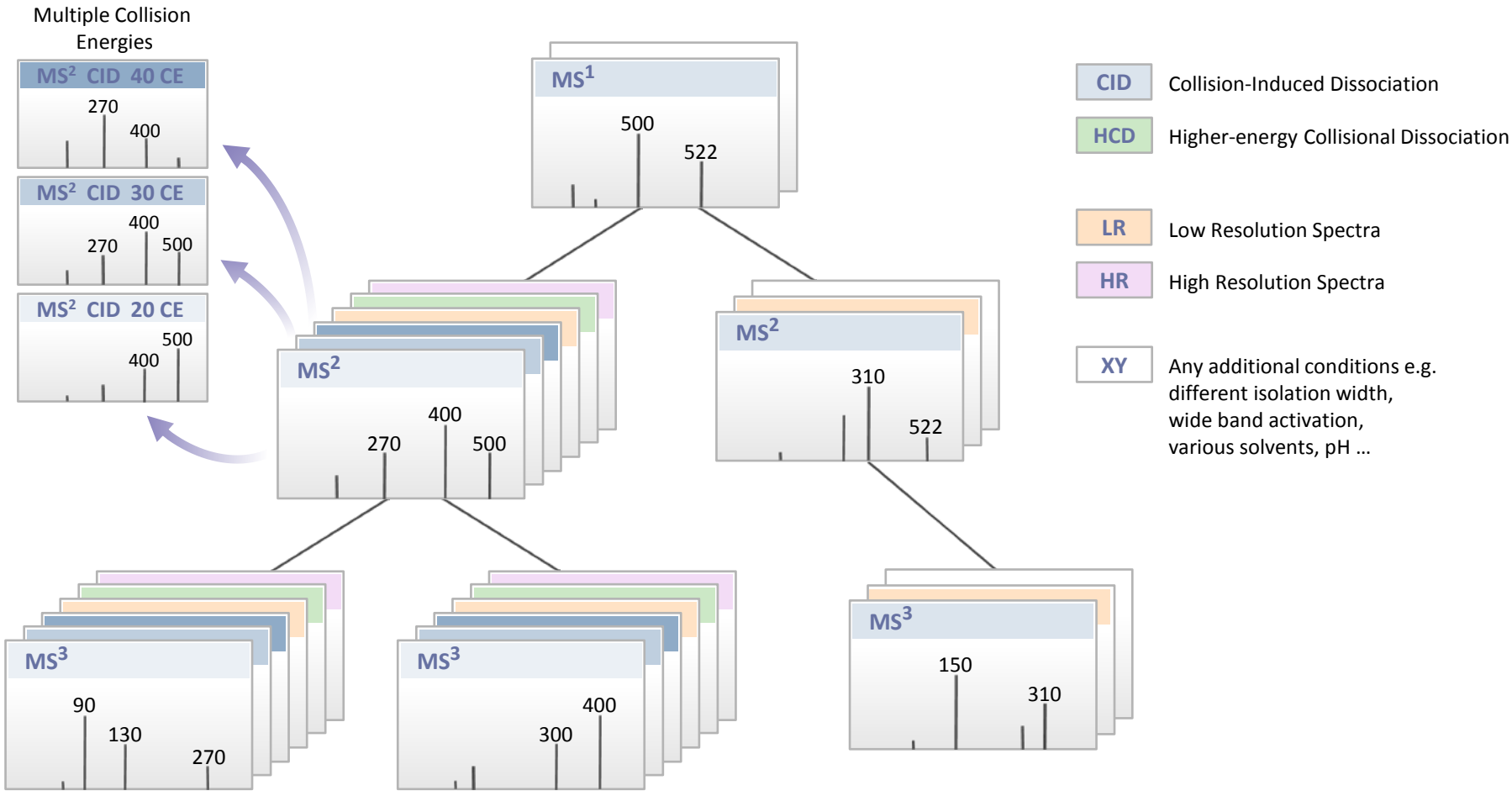
# mzCloud - a freely searchable collection of high resolution/accurate spectra



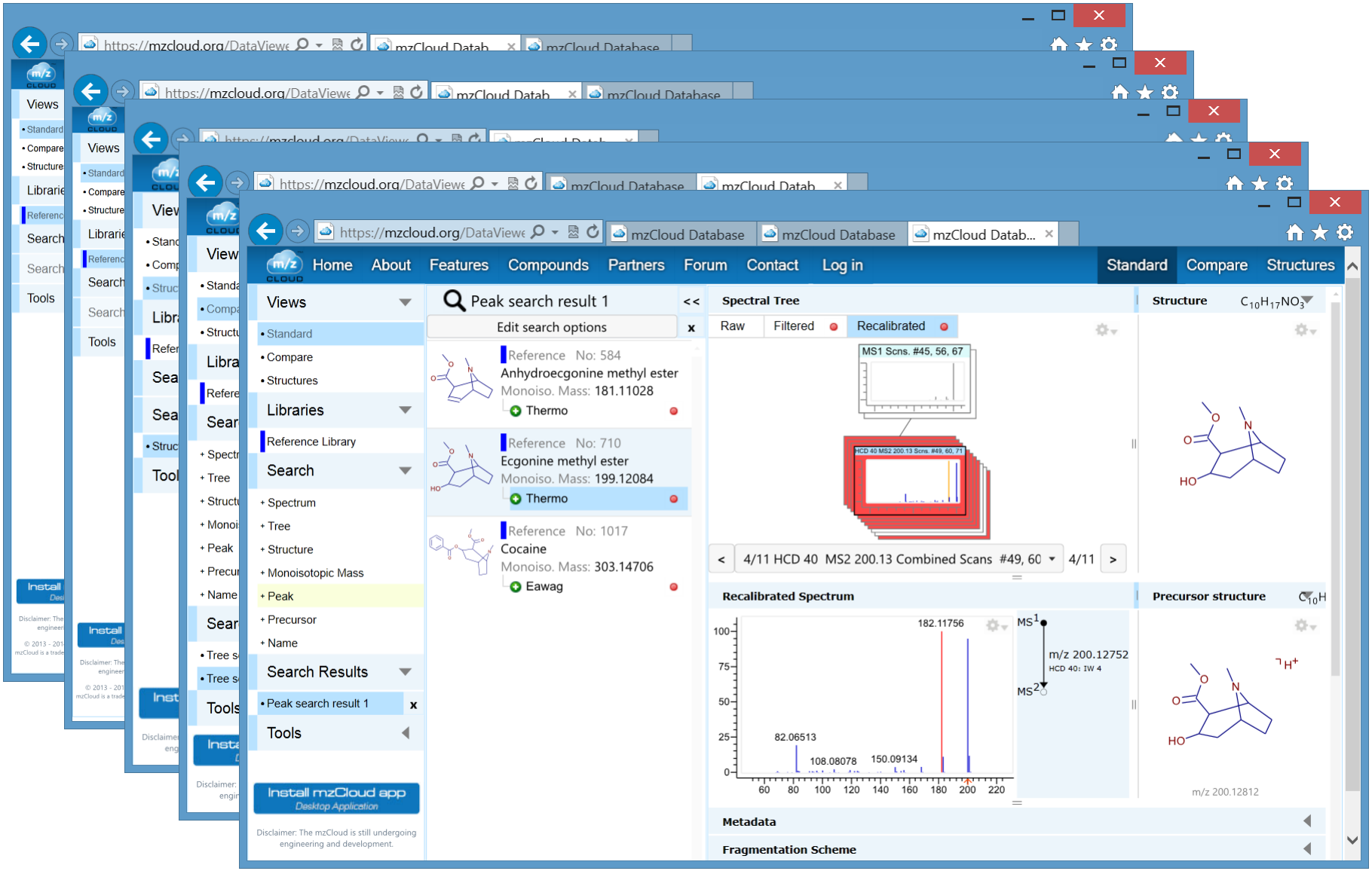
# mzCloud features

- 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

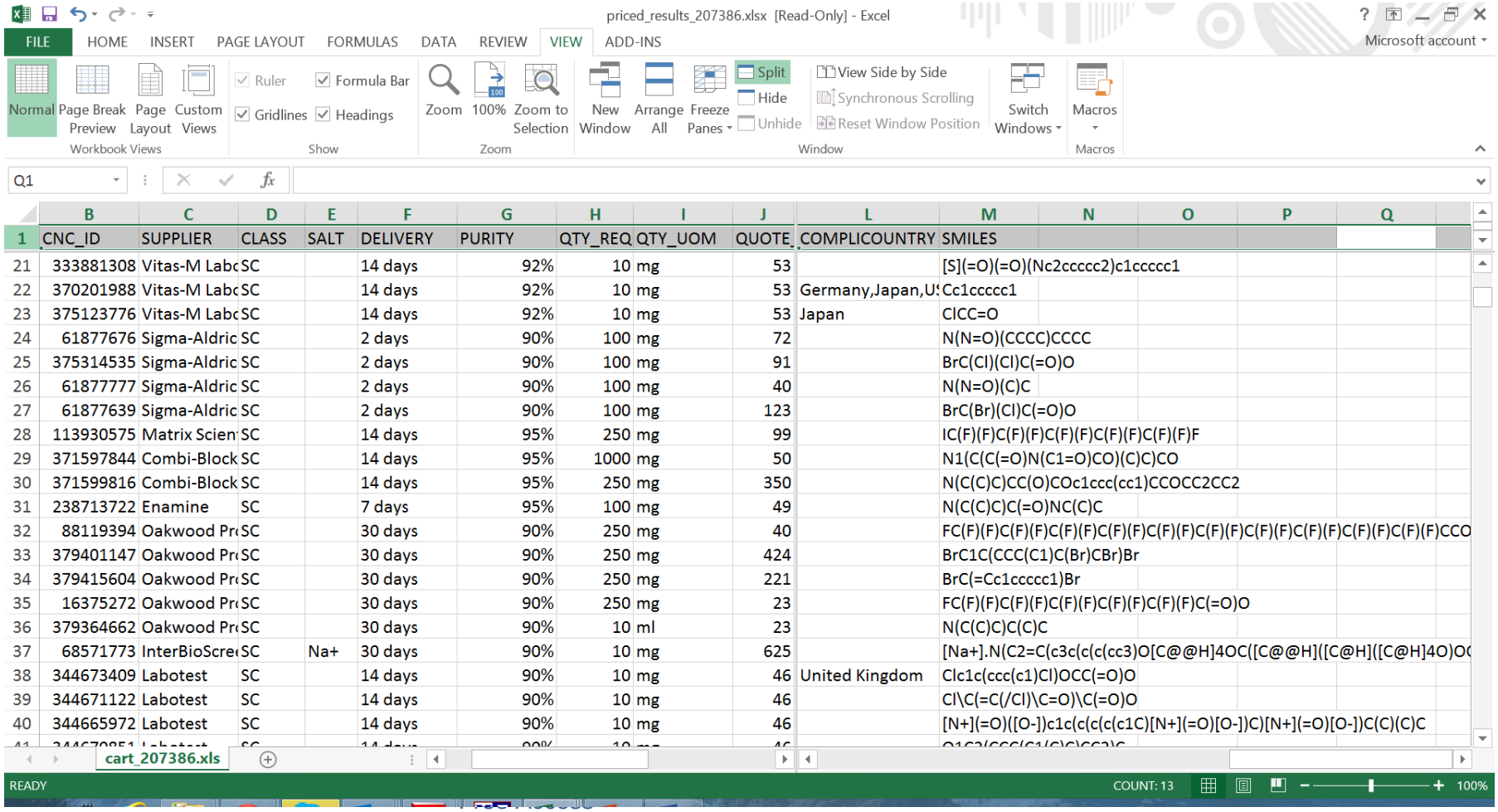
# Spectral Trees



# mzCloud.org


https://mzcloud.org/DataViewe
mzCloud Database
mzCloud Datab...
Home
About
Features
Compounds
Partners
Forum
Contact
Log in
Standard
Compare
Structures
Views
Standard
Compare
Structures
Libraries
Reference Library
Search
Spectrum
Tree
Structure
Monoisotopic Mass
Peak
Precursor
Name
Search Results
Peak search result 1
Tools
Install mzCloud app
Desktop Application
Disclaimer: The mzCloud is still undergoing engineering and development.
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
MS1 Scns. #45, 56, 67
4/11 HCD 40 MS2 200.13 Combined Scans #49, 60
4/11
Structure
$C_{10}H_{17}NO_3$
Recalibrated Spectrum
182.11756
82.06513
108.08078
150.09134
m/z 200.12752
HCD 40: IW 4
Precursor structure
m/z 200.12812
Metadata
Fragmentation Scheme

# www.mzCloud.org




| | B | C | D | E | F | G | H | I | J | L | M | N | O | P | Q |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | CNC_ID | SUPPLIER | CLASS | SALT | DELIVERY | PURITY | QTY_REQ | QTY_UOM | QUOTE_ | COMPLICOUNTRY | SMILES | | | | |
| 21 | 333881308 | Vitas-M Labc | SC | | 14 days | 92% | 10 | mg | 53 | | [S](=O)(=O)(Nc2ccccc2)c1ccccc1 | | | | |
| 22 | 370201988 | Vitas-M Labc | SC | | 14 days | 92% | 10 | mg | 53 | Germany,Japan,U | Cc1ccccc1 | | | | |
| 23 | 375123776 | Vitas-M Labc | SC | | 14 days | 92% | 10 | mg | 53 | Japan | ClCC=O | | | | |
| 24 | 61877676 | Sigma-Aldric | SC | | 2 days | 90% | 100 | mg | 72 | | N(N=O)(CCCC)CCCC | | | | |
| 25 | 375314535 | Sigma-Aldric | SC | | 2 days | 90% | 100 | mg | 91 | | BrC(Cl)(Cl)C(=O)O | | | | |
| 26 | 61877777 | Sigma-Aldric | SC | | 2 days | 90% | 100 | mg | 40 | | N(N=O)(C)C | | | | |
| 27 | 61877639 | Sigma-Aldric | SC | | 2 days | 90% | 100 | mg | 123 | | BrC(Br)(Cl)C(=O)O | | | | |
| 28 | 113930575 | Matrix Scien | SC | | 14 days | 95% | 250 | mg | 99 | | IC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F | | | | |
| 29 | 371597844 | Combi-Block | SC | | 14 days | 95% | 1000 | mg | 50 | | N1(C(C(=O)N(C1=O)CO)(C)C)CO | | | | |
| 30 | 371599816 | Combi-Block | SC | | 14 days | 95% | 250 | mg | 350 | | N(C(C)C)CC(O)COc1ccc(cc1)CCOCC2CC2 | | | | |
| 31 | 238713722 | Enamine | SC | | 7 days | 95% | 100 | mg | 49 | | N(C(C)C)C(=O)NC(C)C | | | | |
| 32 | 88119394 | Oakwood Pr | SC | | 30 days | 90% | 250 | mg | 40 | | FC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)CCO | | | | |
| 33 | 379401147 | Oakwood Pr | SC | | 30 days | 90% | 250 | mg | 424 | | BrC1C(CCC(C1)C(Br)CBr)Br | | | | |
| 34 | 379415604 | Oakwood Pr | SC | | 30 days | 90% | 250 | mg | 221 | | BrC(=Cc1ccccc1)Br | | | | |
| 35 | 16375272 | Oakwood Pr | SC | | 30 days | 90% | 250 | mg | 23 | | FC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(=O)O | | | | |
| 36 | 379364662 | Oakwood Pr | SC | | 30 days | 90% | 10 | ml | 23 | | N(C(C)C)C(C)C | | | | |
| 37 | 68571773 | InterBioScre | SC | Na+ | 30 days | 90% | 10 | mg | 625 | | [Na+].N(C2=C(c3c(c(c(cc3)O[C@@H]4OC([C@@H]([C@H]([C@H]4O)O( | | | | |
| 38 | 344673409 | Labotest | SC | | 14 days | 90% | 10 | mg | 46 | United Kingdom | Clc1c(ccc(c1)Cl)OCC(=O)O | | | | |
| 39 | 344671122 | Labotest | SC | | 14 days | 90% | 10 | mg | 46 | | Cl\C(=C(/Cl)\C=O)\C(=O)O | | | | |
| 40 | 344665972 | Labotest | SC | | 14 days | 90% | 10 | mg | 46 | | [N+](=O)([O-])c1c(c(c(c(c1C)[N+](=O)[O-])C)[N+](=O)[O-])C(C)(C)C | | | | |

# mzCloud

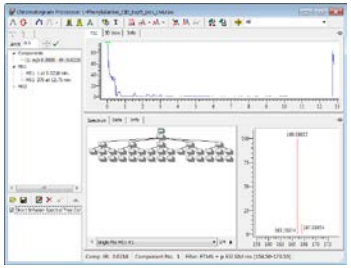
## Data Acquisition

Reference Compound

Orbitrap

TreeRobot

## Data Processing



Mass Frontier 7.0

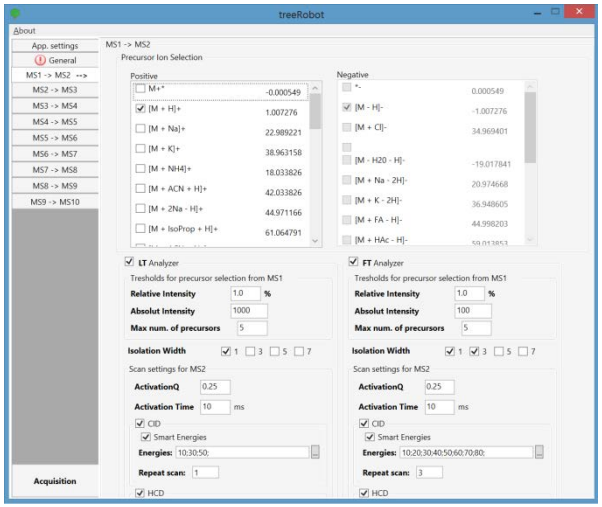
IMPORTER

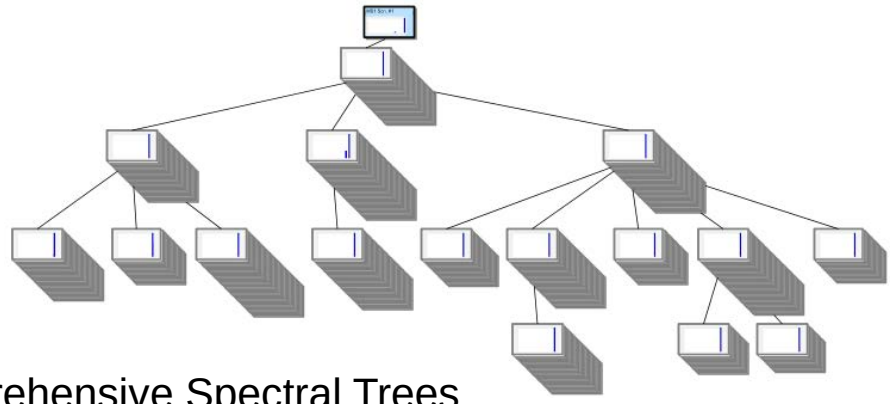
# Spectral Trees Acquisition

Orbitrap

+

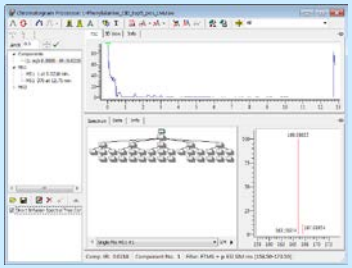
TreeRobot



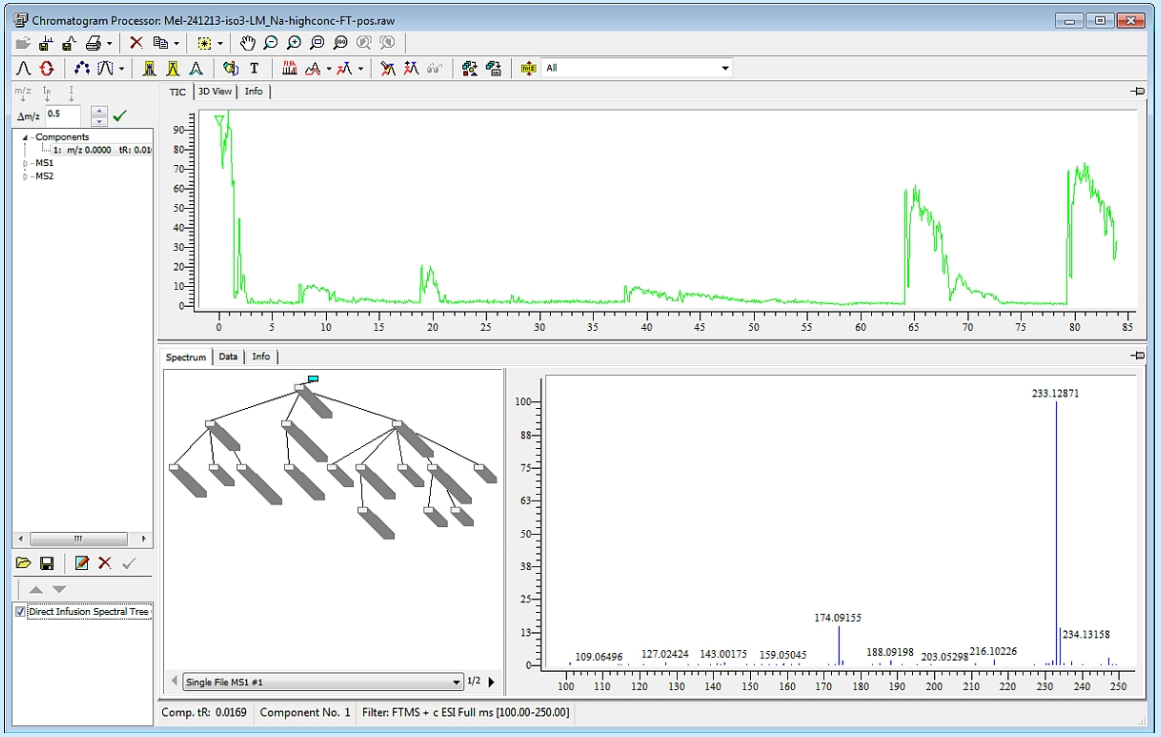
Comprehensive Spectral Trees

# mzCloud: Rigorously Curated Data

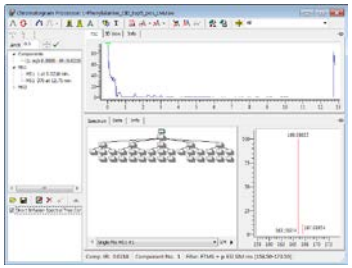


Mass Frontier 7.0

IMPORTER

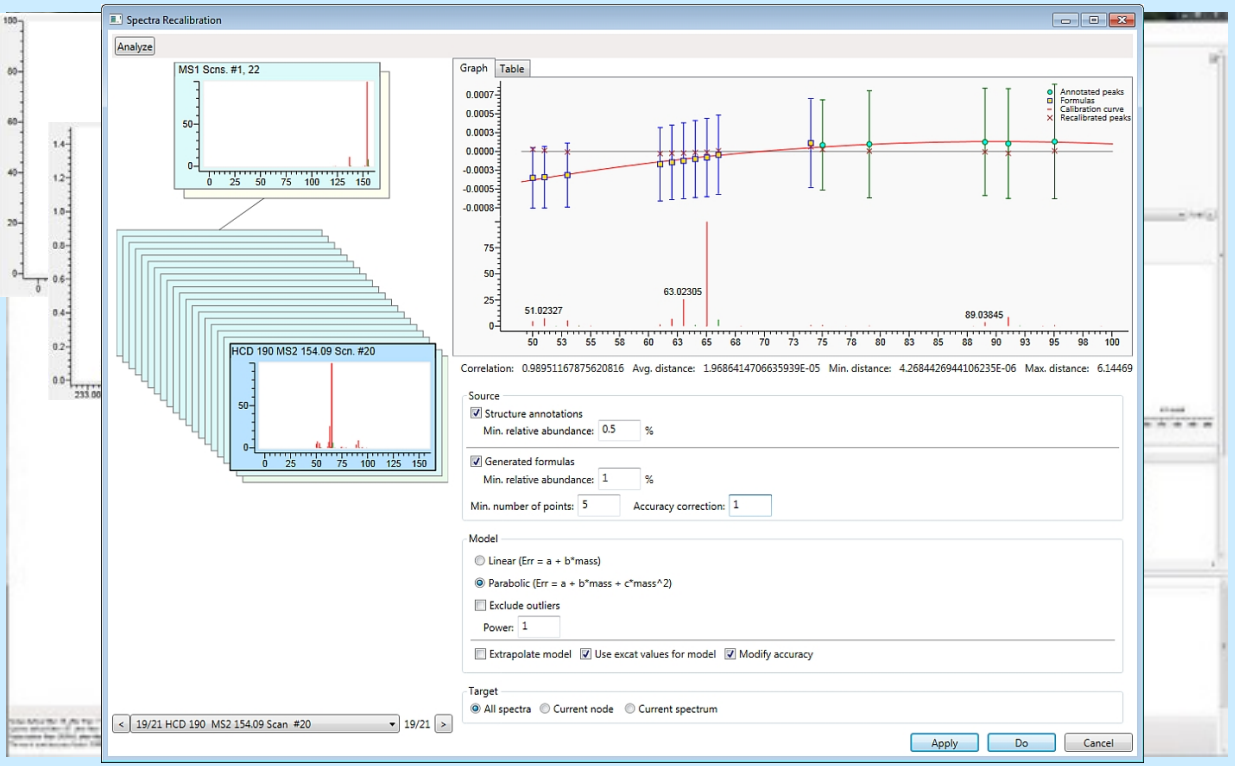


# mzCloud: Rigorously Curated Data

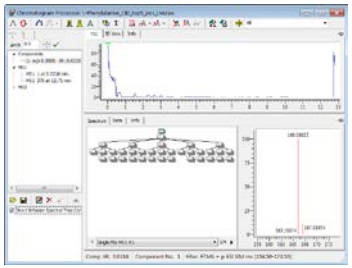


Mass Frontier 7.0

IMPORTER

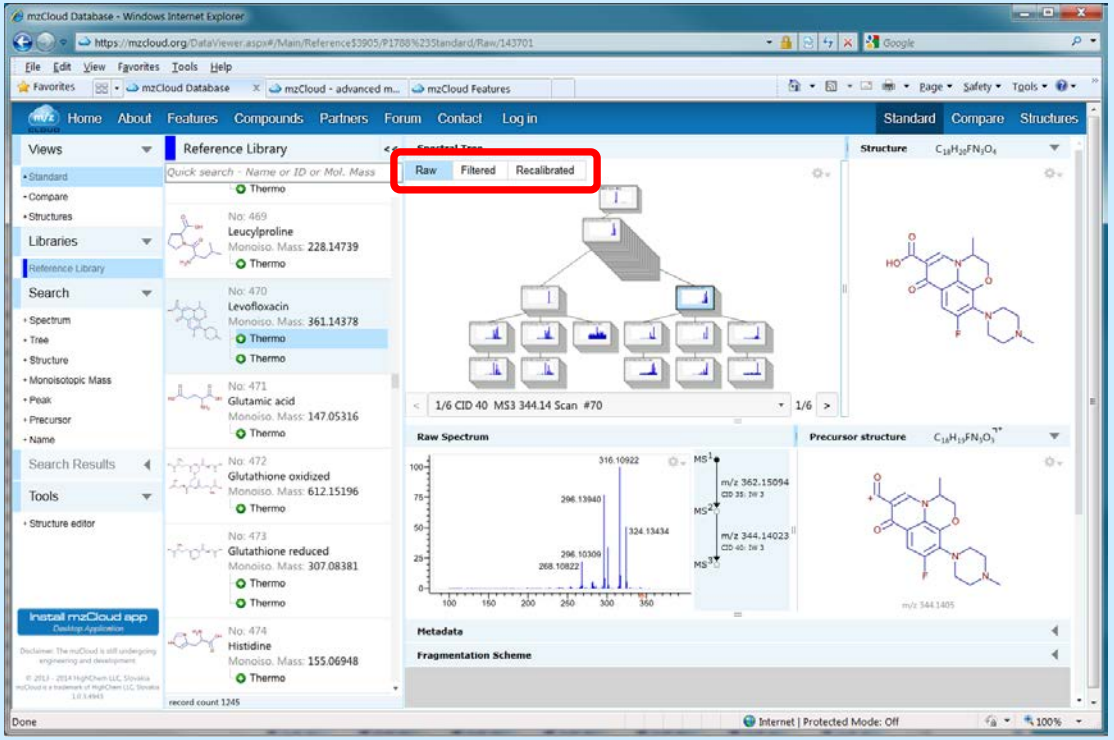


# mzCloud: Rigorously Curated Data

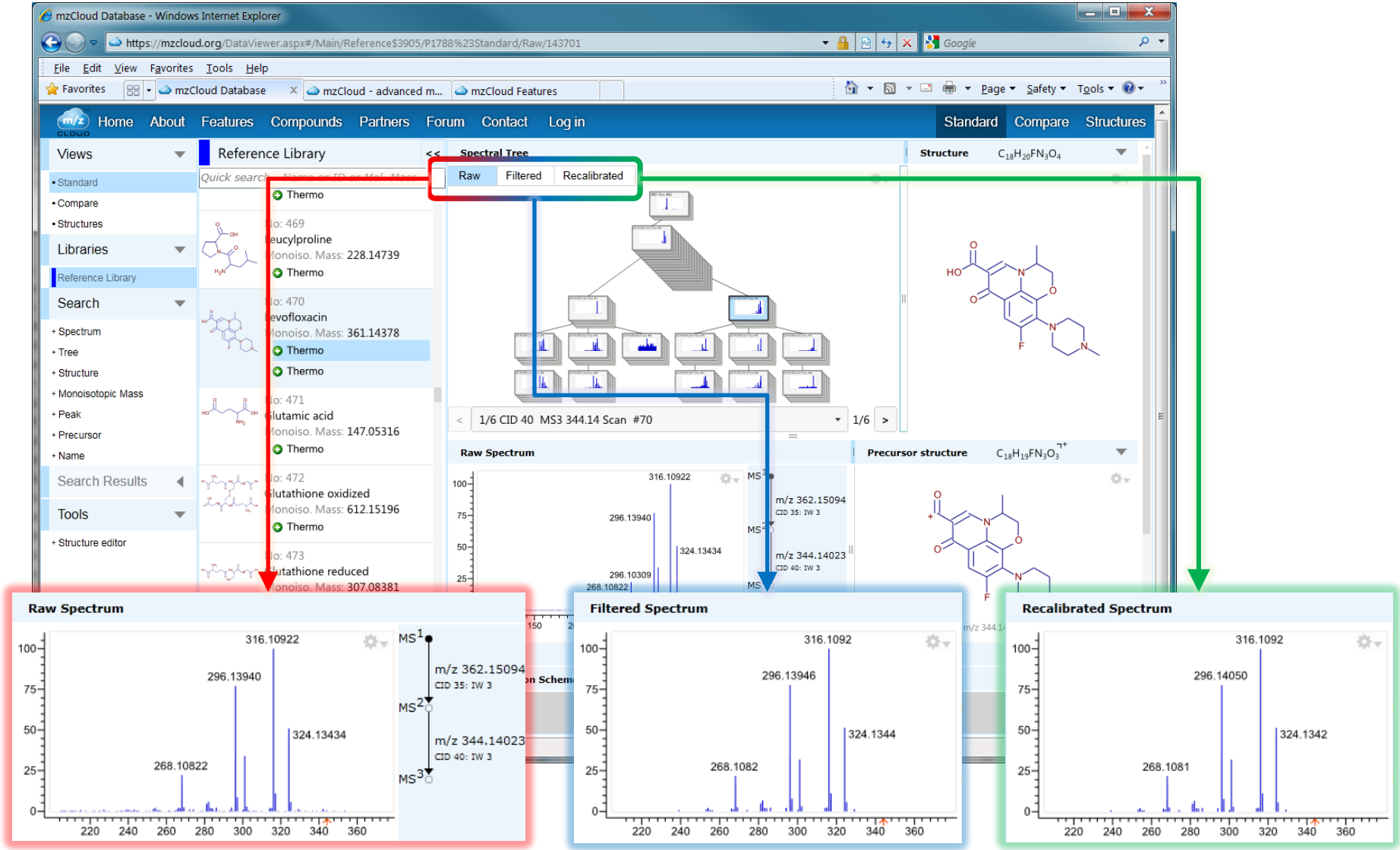


Mass Frontier 7.0

IMPORTER



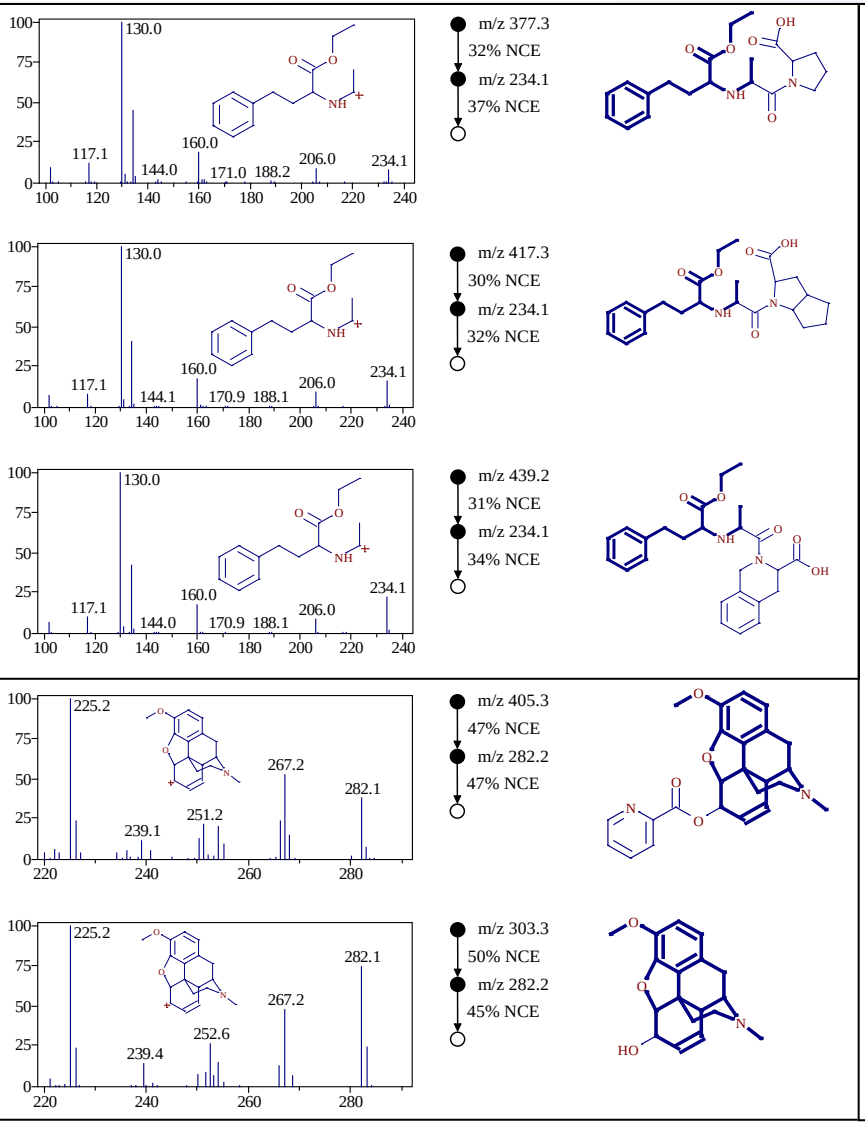
# mzCloud – Raw, Filtered and Recalibrated Spectra



# Work in progress

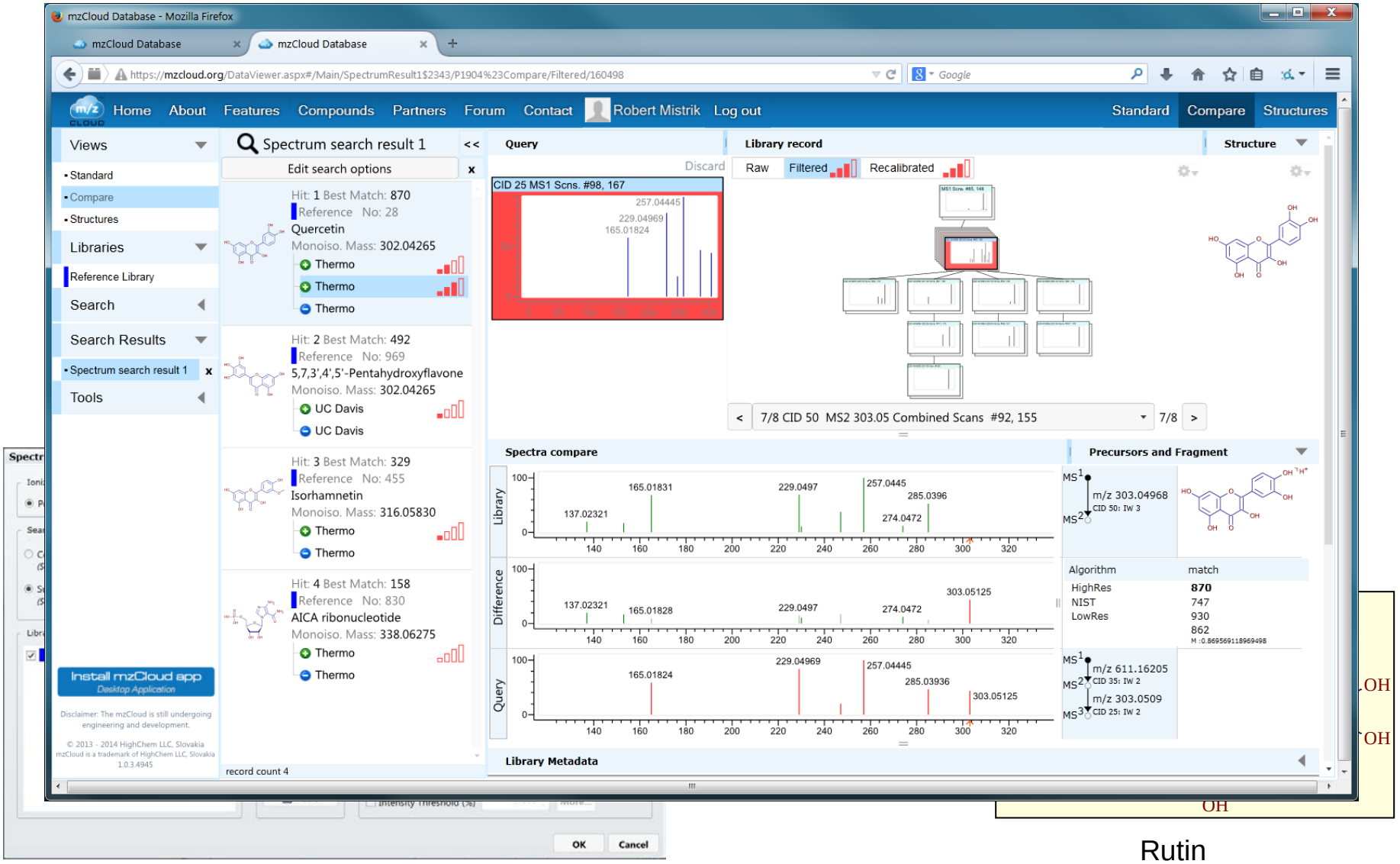
- Fully searchable sample database
- Predicted fragments for suspect compounds
- Precursor Ion Fingerprinting
- Real time validation of search algorithm
- Breakdown curves

# Precursor Ion Fingerprinting (PIF)



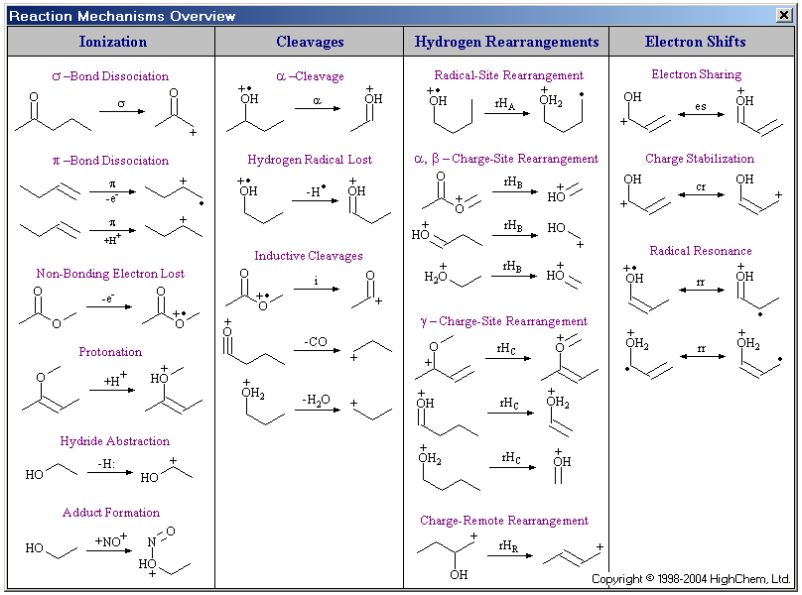

Sheldon, Michelle T., Robert Mistrik, and Timothy R. Croley. "Determination of ion structures in structurally related compounds using precursor ion fingerprinting." Journal of the American Society for Mass Spectrometry 20.3 (2009): 370-376.

# Metabolite Identification Using PIF/mzCloud



# Knowledge based prediction of fragmentation pathways

## General fragmentation rules

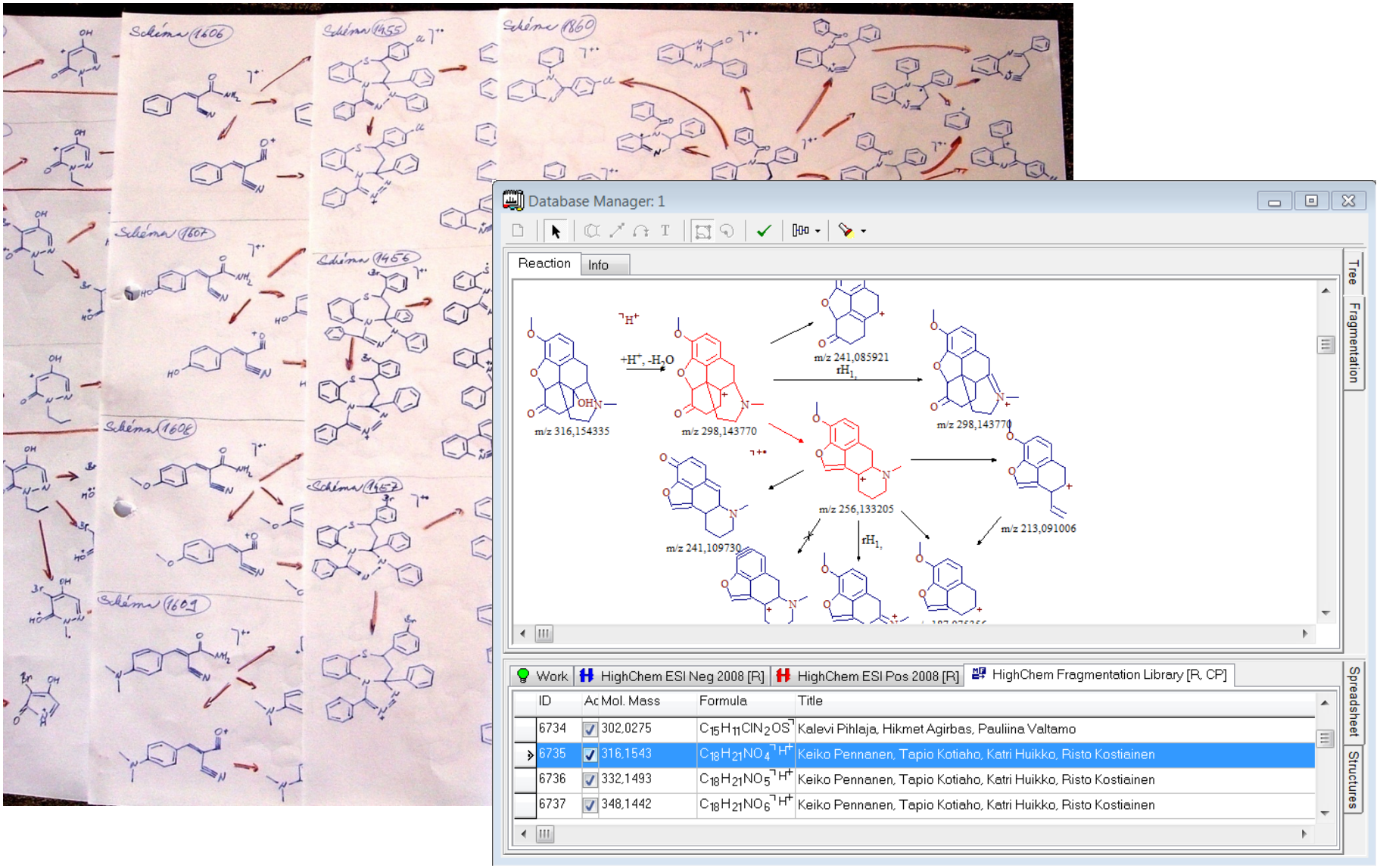



## Fragmentation Library

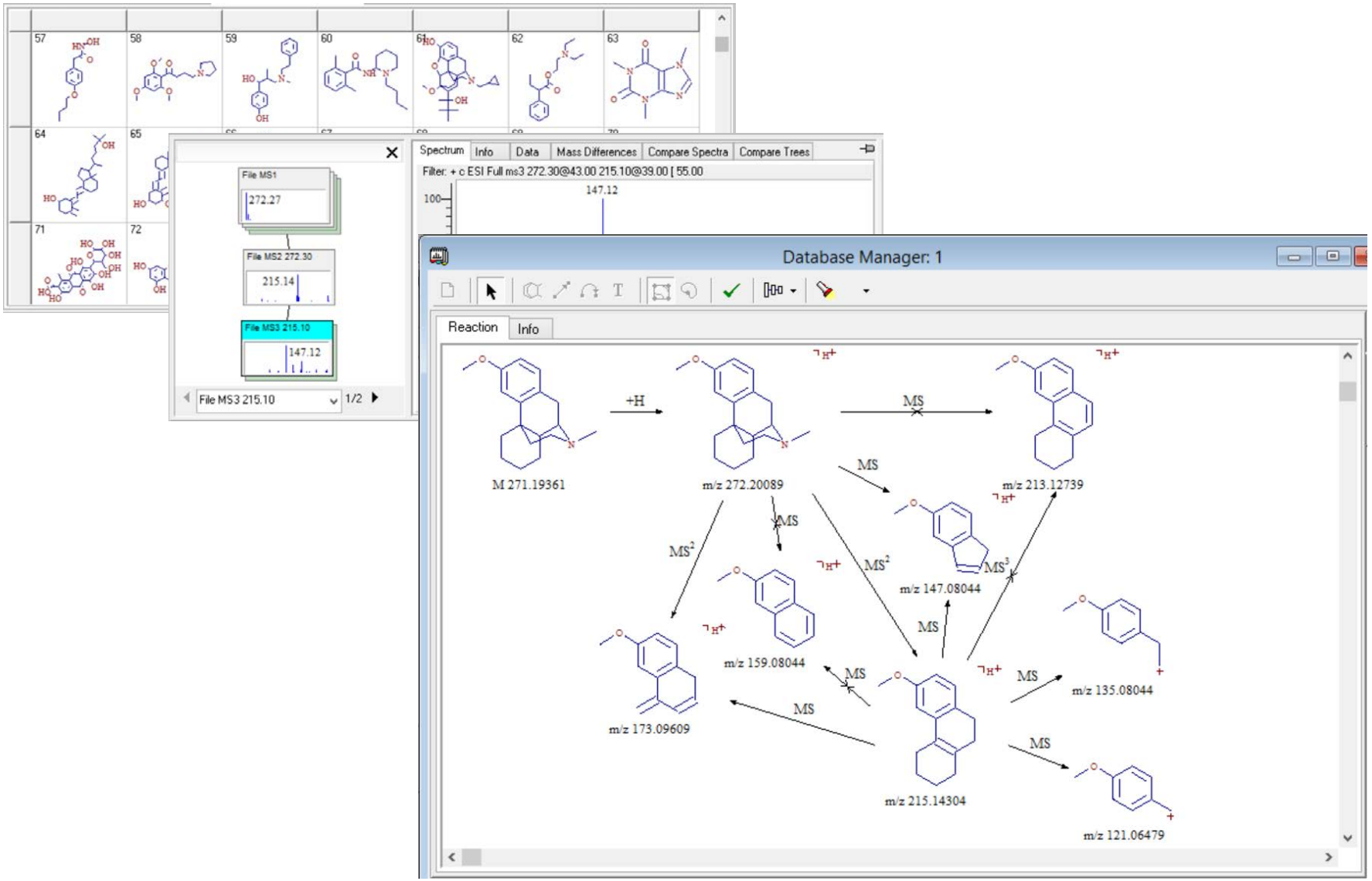
| Total number of | Mass Frontier 7.0 |
|---|---|
| Fragmentation Schemes | 30.936 |
| Individual Reactions | 129.229 |
| Chemical Structures | 151.762 |

| | Source | Volume | Year |
|---|---|---|---|
| 1. | **JASMS** (Journal of the American Society for Mass Spectrometry) | 1-17 | 1990-2006 |
| 2. | **IJMSIP** (International Journal of Mass Spectrometry and Ion Physics) | 1-53 | 1968-1983 |
| | **IJMSIP** (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<br>29-30<br>37-48<br>50-53 | 1964-1979<br>1981-1982<br>1989-2000<br>2002-2005 |
| 7. | **MSR** ( Mass Spectrometry Reviews) | 1-25 | 1981-2006 |
| 8. | **EJMSBMER** (European Journal of Mass Spectrometry in Biochemical, Medicine, and Environmental. Research) | 1-2 | 1980-1982 |
| 9. | **BMS** (Biomedical Mass Spectrometry)<br>**BEMS** (Biomedical and Environmental Mass Spectrometry)<br>**BMS** (Biological Mass Spectrometry) | 1-12<br>14-19<br><br>20-23 | 1974-1985<br>1987-1990<br><br>1991-1994 |
| 10. | **JC** (Journal of Chromatography) | 181-536 | 1980-1991 |
| 11. | **EJMS** (European Journal of Mass Spectrometry) | 4 | 1998 |

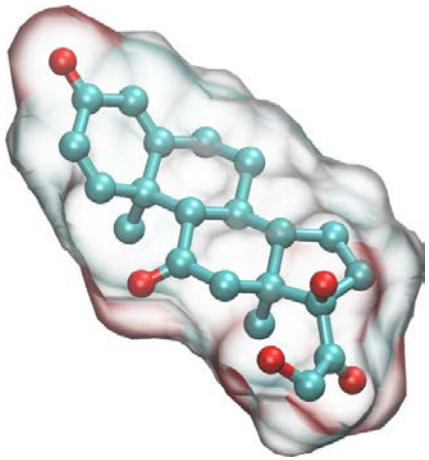
# Knowledge based prediction of fragmentation pathways



# Knowledge based prediction of fragmentation pathways



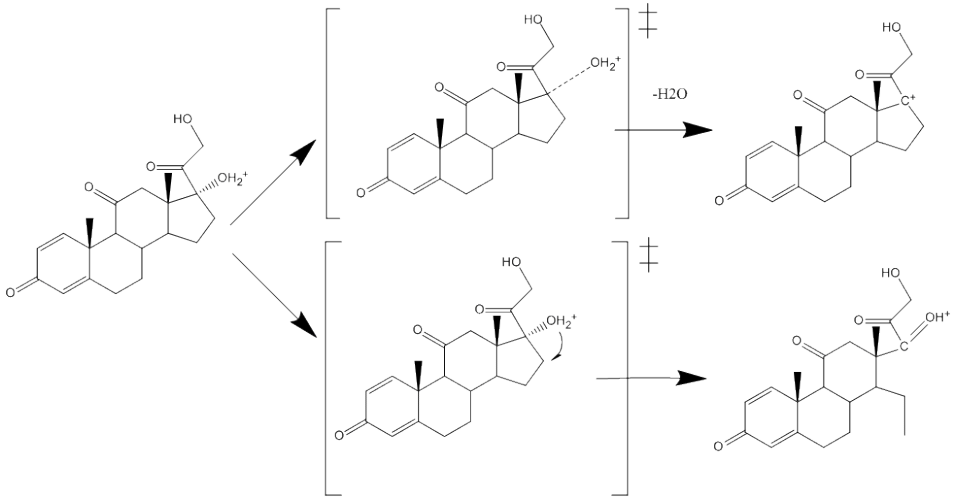
# Evaluation of fragmentation pathways by high-level quantum-mechanical calculations based on DFT method



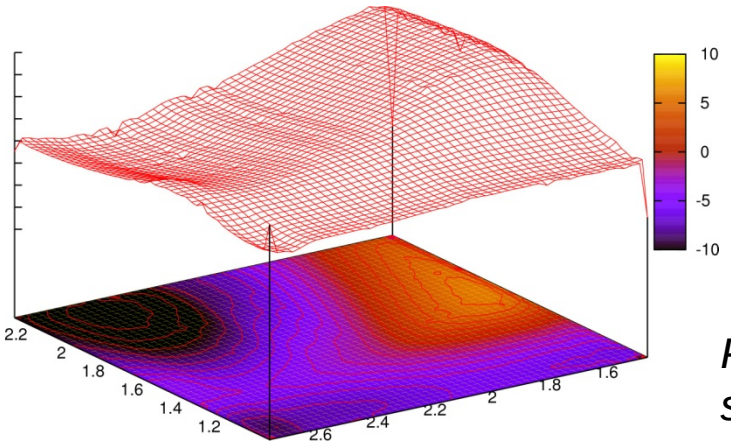
*Prednisone*

*(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

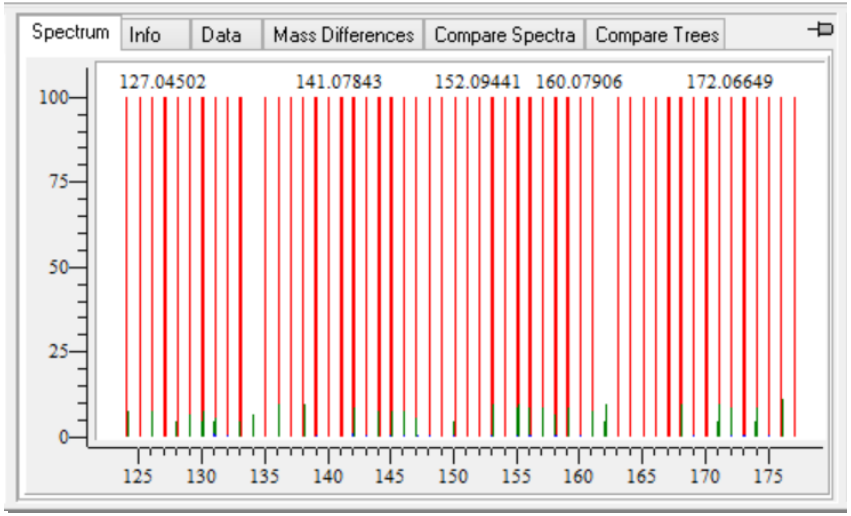



*Example for competitive reaction in the fragmentation pathway*

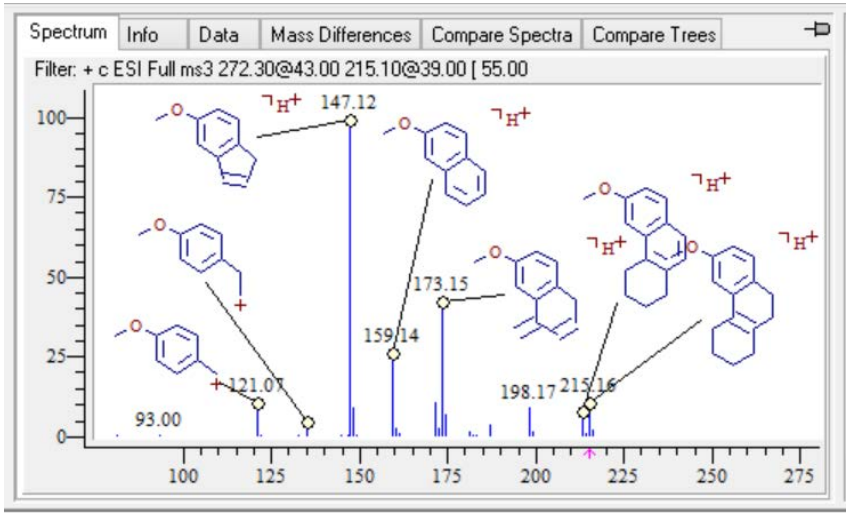



*PES for corresponding reaction, followed by transition state localization and reaction rate calculation*

# Use of Predicted Fragments

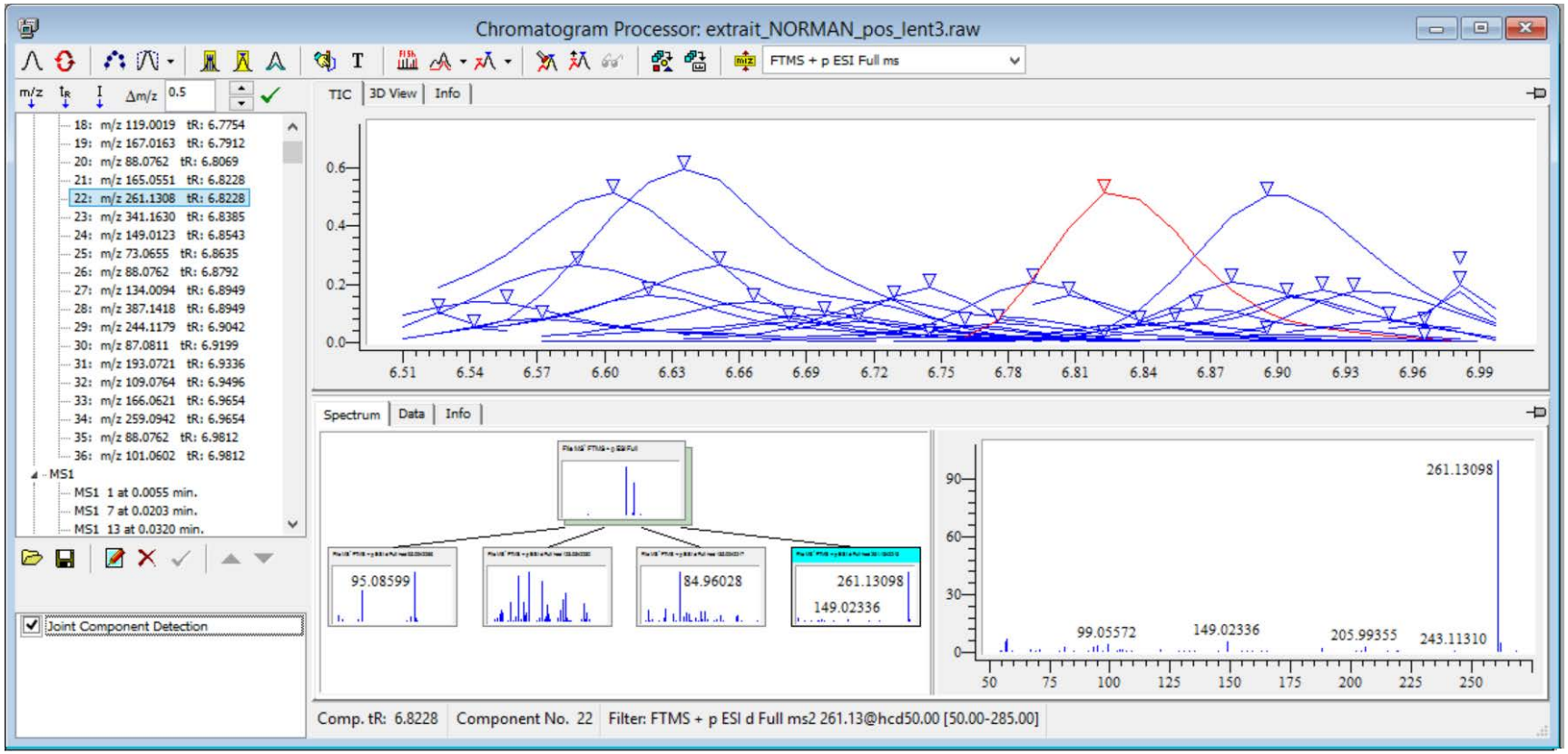



- **Matching *in silico* m/z values against experimental spectra**




- Fragment annotation

# What are the compounds?

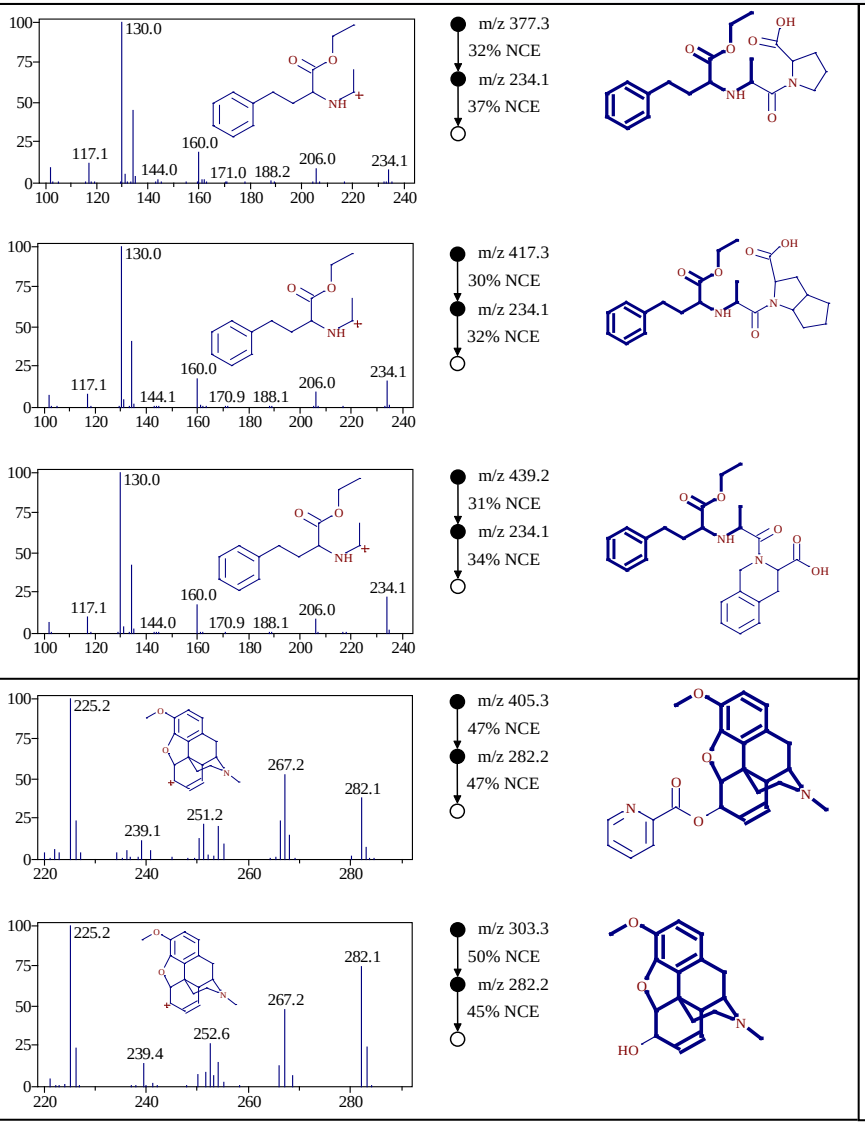


# BINGO Initiative

- 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)

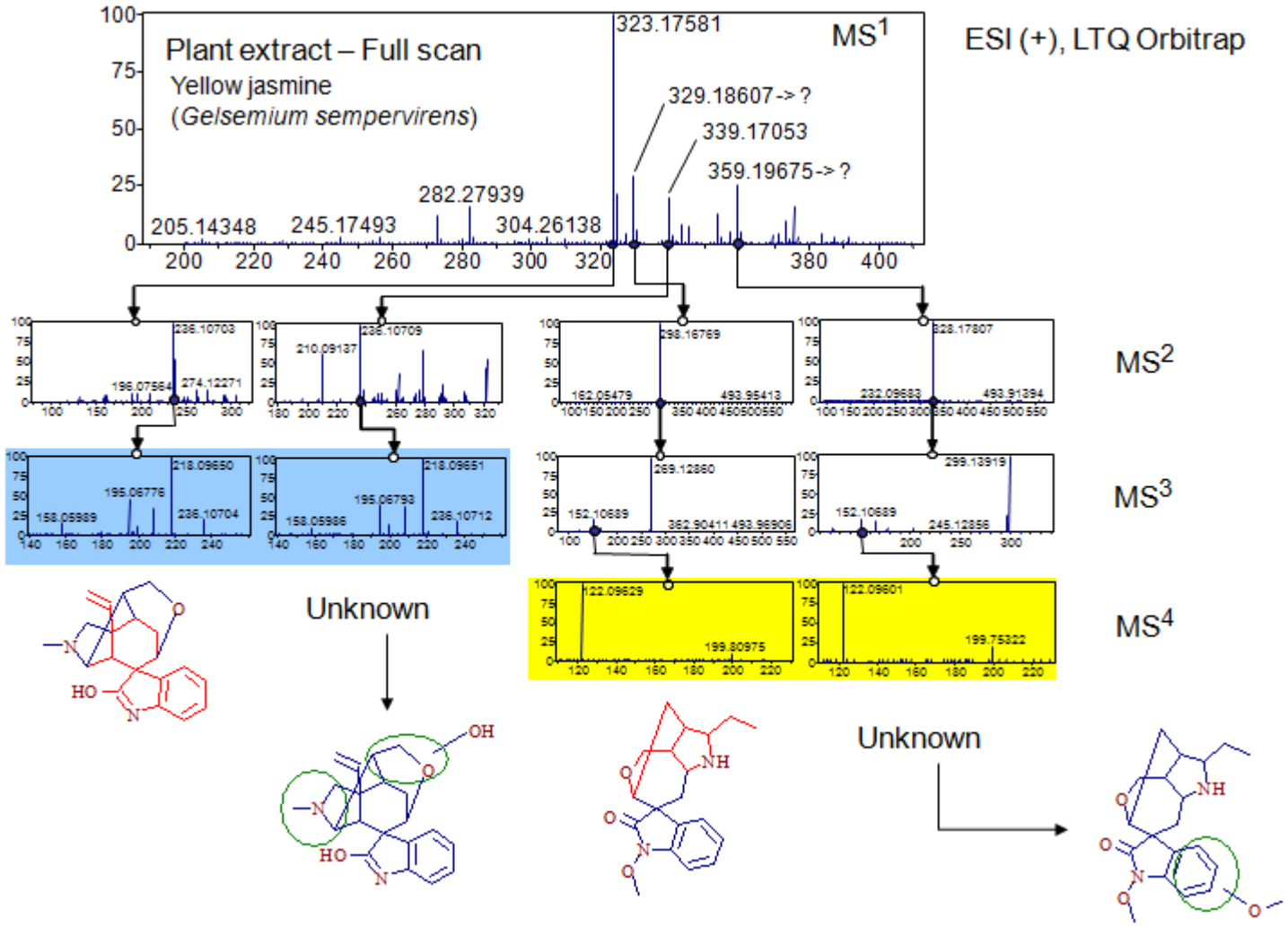
**Let’s make our dream come true**

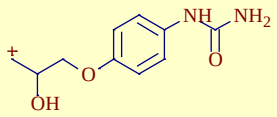
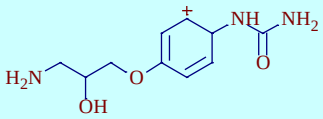
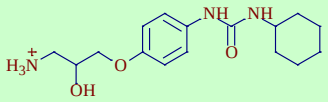
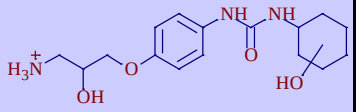
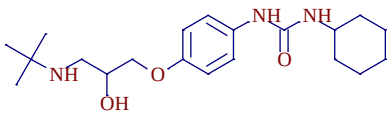
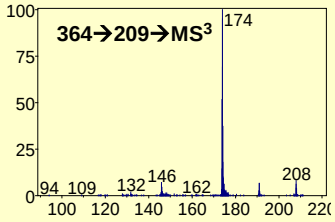
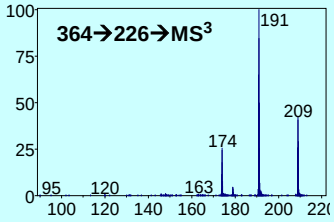
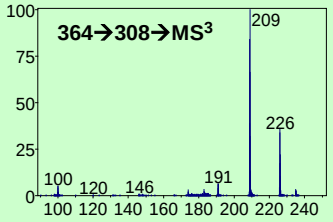
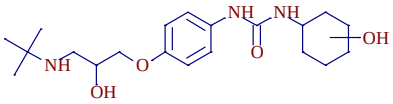
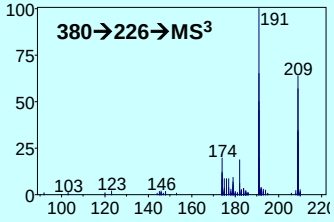
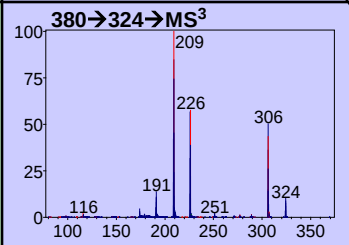
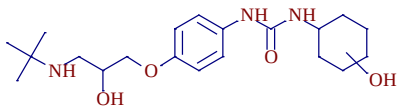
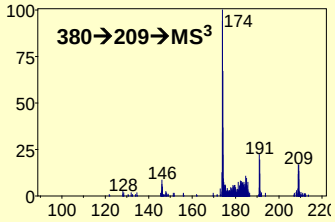
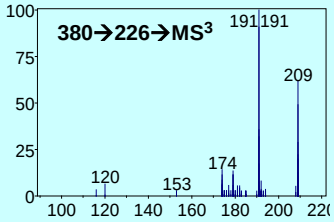
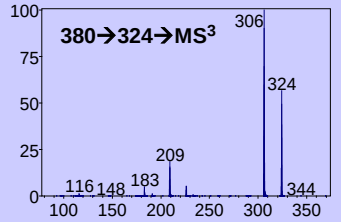
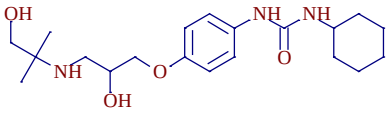
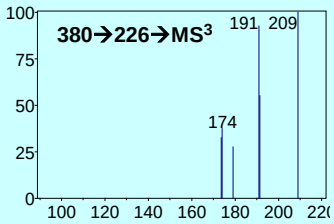
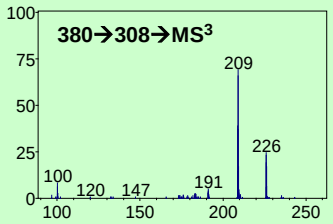
# Precursor Ion Fingerprinting (PIF)




Sheldon, Michelle T., Robert Mistrik, and Timothy R. Croley. "Determination of ion structures in structurally related compounds using precursor ion fingerprinting." Journal of the American Society for Mass Spectrometry 20.3 (2009): 370-376.

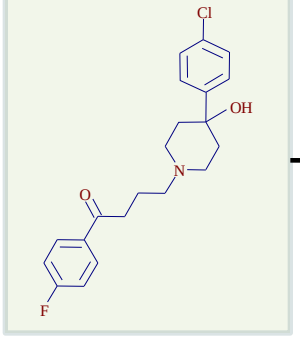
# Gelsemine Metabolite Identification by PIF



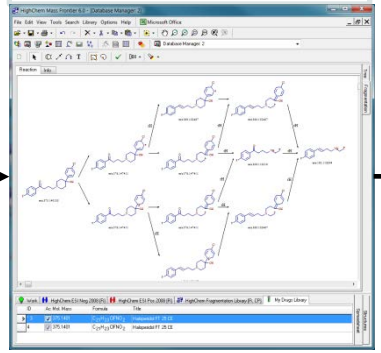
| | | m/z 209 | m/z 226 | m/z 308 | m/z 324 |
|---|---|---|---|---|---|
| | Metabolite Structure Assignment |  |  |  |  |
| Talinolol |  |   |   |   | |
| Metabolite C01 |  | |   | |   |
| Metabolite C02 |  |   |   | |   |
| Metabolite C09 |  | |   |   | Intensity differences due to different hydroxyl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

# Fragment Ion Search (FISh)

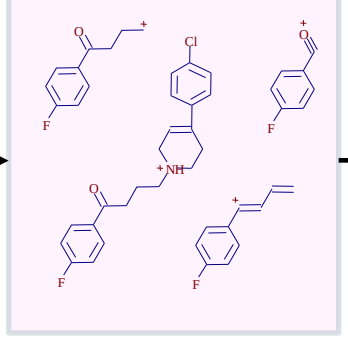
Parent Drug




Fragment Predict.



Predicted Ions

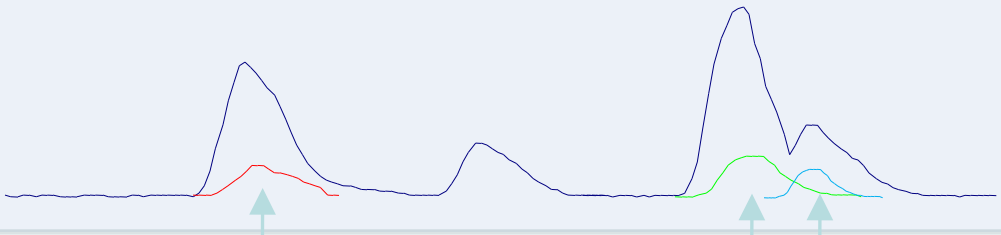



Accurate m/z

| | | | |
|---|---|---|---|
| 30.03383 | 109.04480 | 176.03873 | 233.09658 |
| 31.01784 | 109.99178 | 177.04655 | 234.10440 |
| 39.02293 | 110.99960 | 178.10265 | 235.13669 |
| 40.03075 | 112.07569 | 179.07409 | 236.14452 |
| 41.03858 | 113.01525 | 180.08192 | 237.09149 |
| 42.01002 | 114.09134 | 181.04147 | 238.09932 |
| 43.01784 | 118.04132 | 182.03672 | 239.10714 |
| 44.04948 | 119.02915 | 183.04454 | 244.08875 |
| 45.03349 | 120.03698 | 184.05237 | 245.09658 |
| 51.02293 | 121.04480 | 186.06802 | 246.12887 |
| 53.03858 | 122.01624 | 190.04180 | 247.13669 |
| 54.03383 | 123.02407 | 191.04963 | 248.14452 |
| 55.01784 | 124.03189 | 192.05745 | 249.09149 |
| 56.04948 | 125.03972 | 193.06528 | 250.09932 |
| 57.06988 | 133.04480 | 194.07310 | 251.10714 |
| 58.06513 | 134.05263 | 195.05712 | 252.11497 |
| 59.04914 | 135.06045 | 196.05237 | 254.09423 |

Search in LC/MS$^n$

Fragment Ion Search

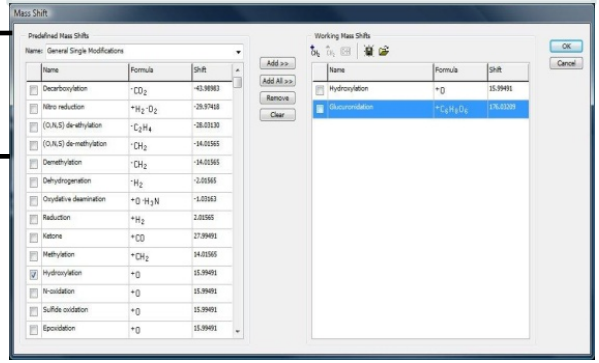


Metabolite I

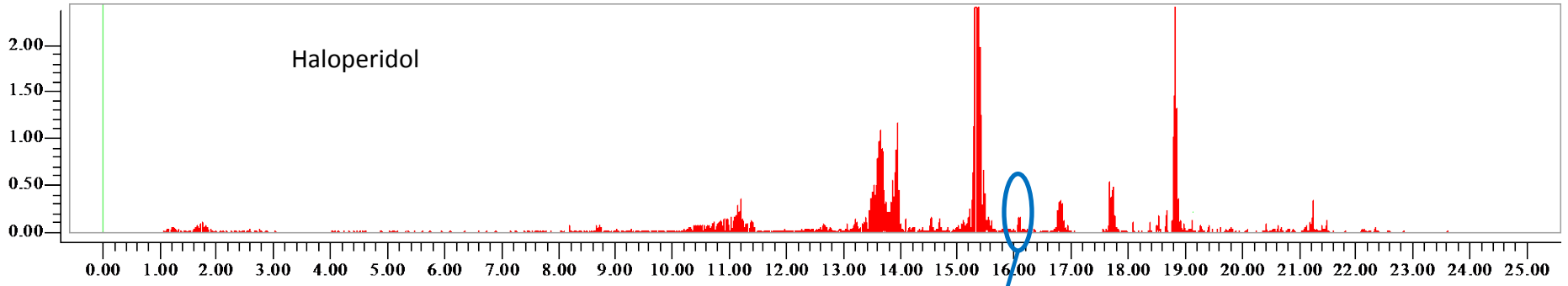
Metabolite II

Metabolite III

Biotransformation Mass Shifts

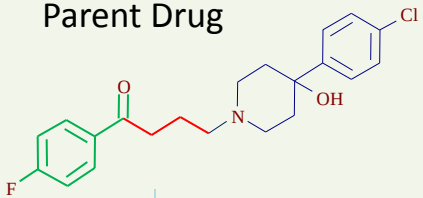


# Metabolite identification using FISh

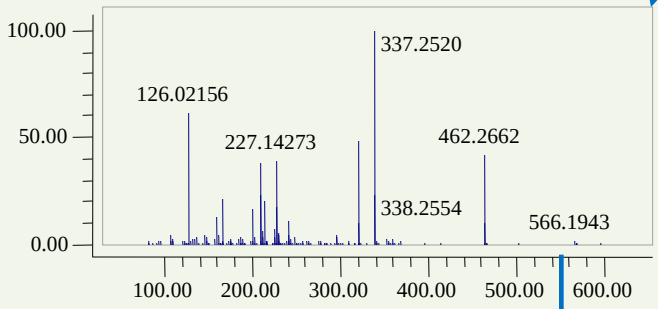


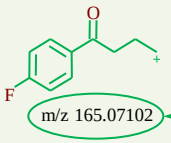

## Common Ions of Parent Drug and Unknown Metabolite

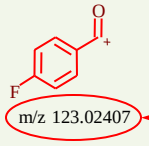
Parent Drug



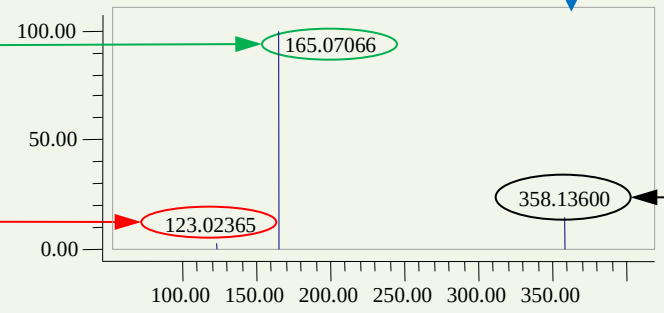

HCD $MS^2$ of unknown metabolite



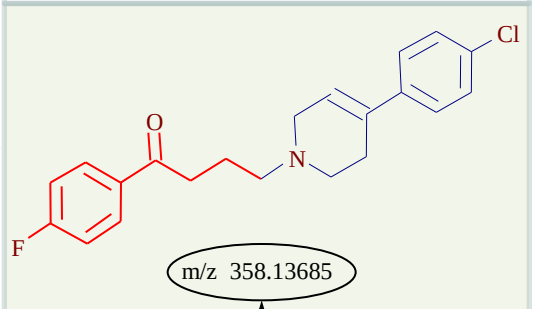



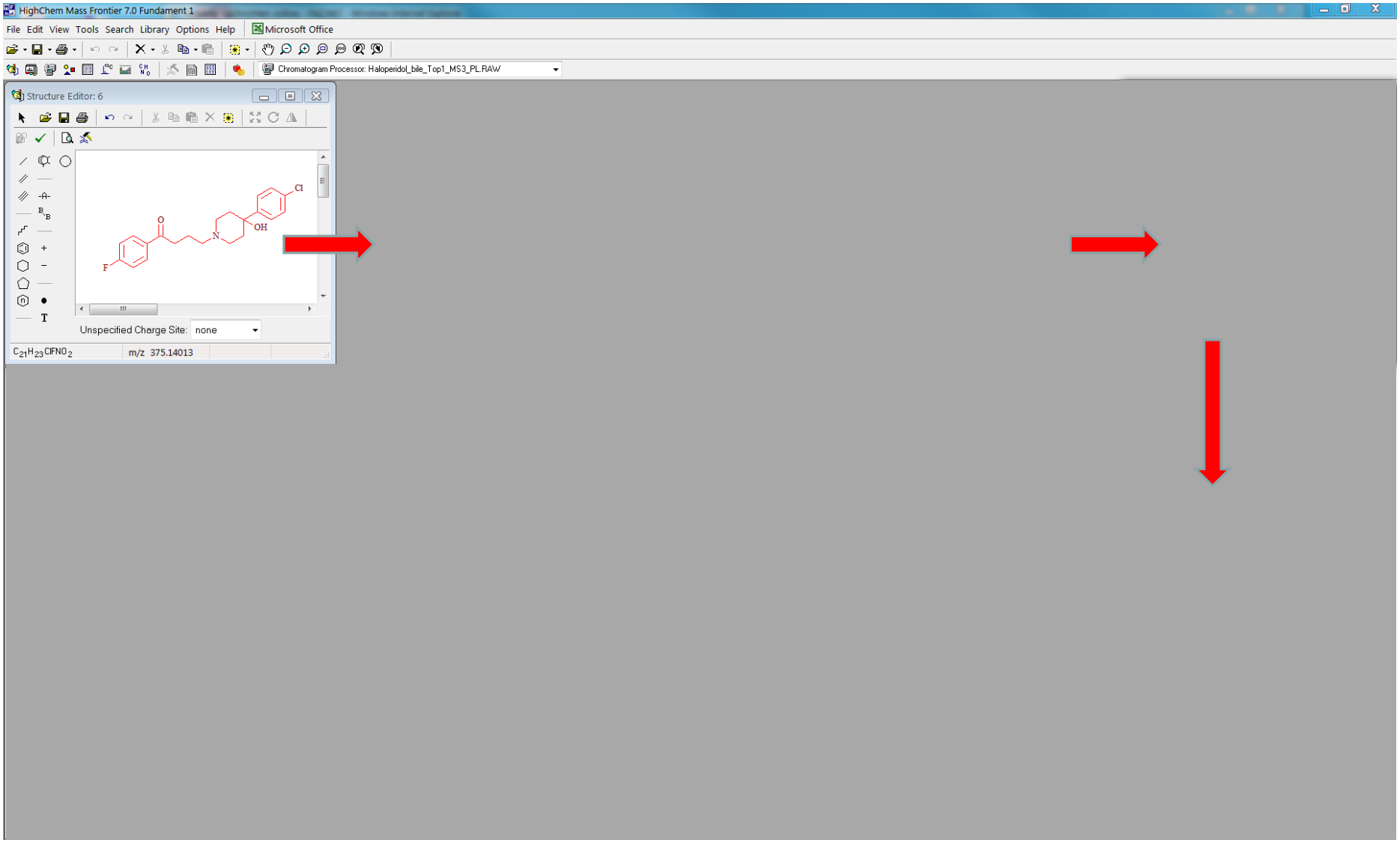

FISh filtered HCD $MS^2$




## Proposed Metabolite Structure



# Metabolite identification using FISh



# 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**