

MetFrag in Practice



Assoc. Prof. Dr. Emma L. Schymanski

FNR ATTRACT Fellow and PI in Environmental Cheminformatics
Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg
Email: emma.schymanski@uni.lu and @ESchymanski

...and many colleagues who contributed to my science over the years!

(Molecule =>) Mass => Molecule => Metadata

- Background

- Molecule to Mass

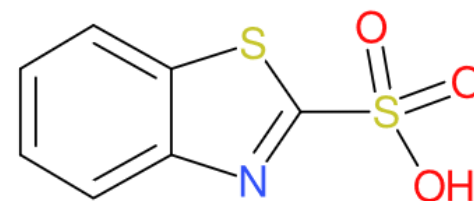
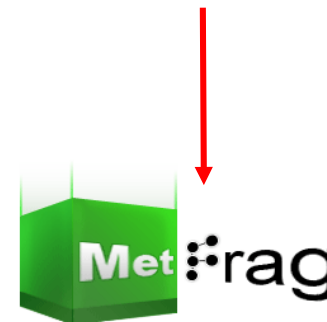
- Preparation for Mass to Molecule

- Gathering the “evidence”
 - Molecular Formula example

- Structure Elucidation with MetFrag

- Step-by-step with one example (MetFrag)
- Practice Session with Several Examples
- [Optional] Context and Perspectives
 - Potential for automated non-target workflows

213.9637



Before We Start ...

- ...and before you try your favourite example ...



- Servers, like people, need TLC

- And it's midnight in Halle where MetFrag sits...

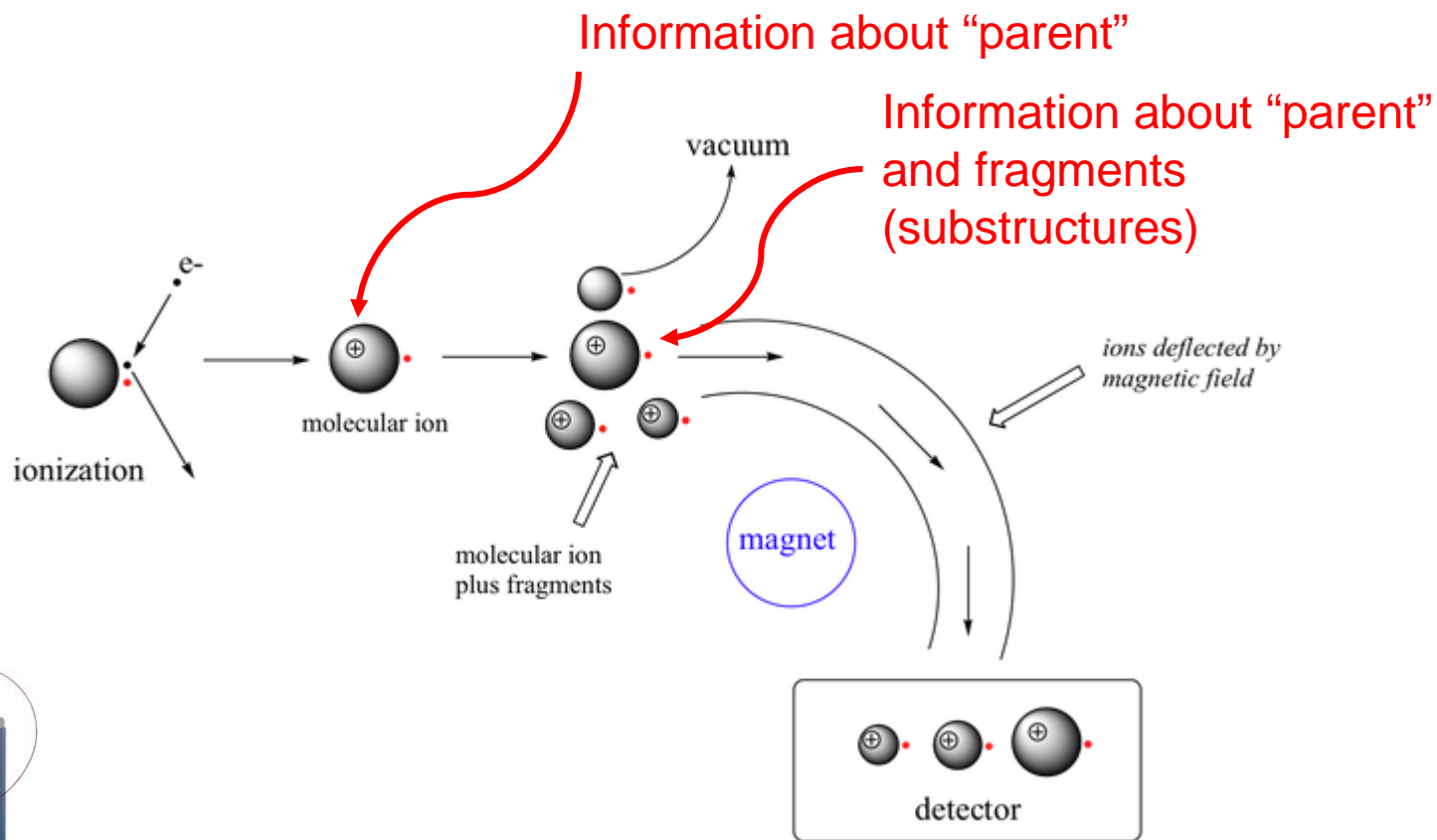


- I've carefully selected examples for today

- Try out these examples in groups to save bandwidth
- Please save new queries for the end of the workshop
- *Because ... it just takes a few of you to test the wrong formula in PubChem and we'll bring all the servers down...*

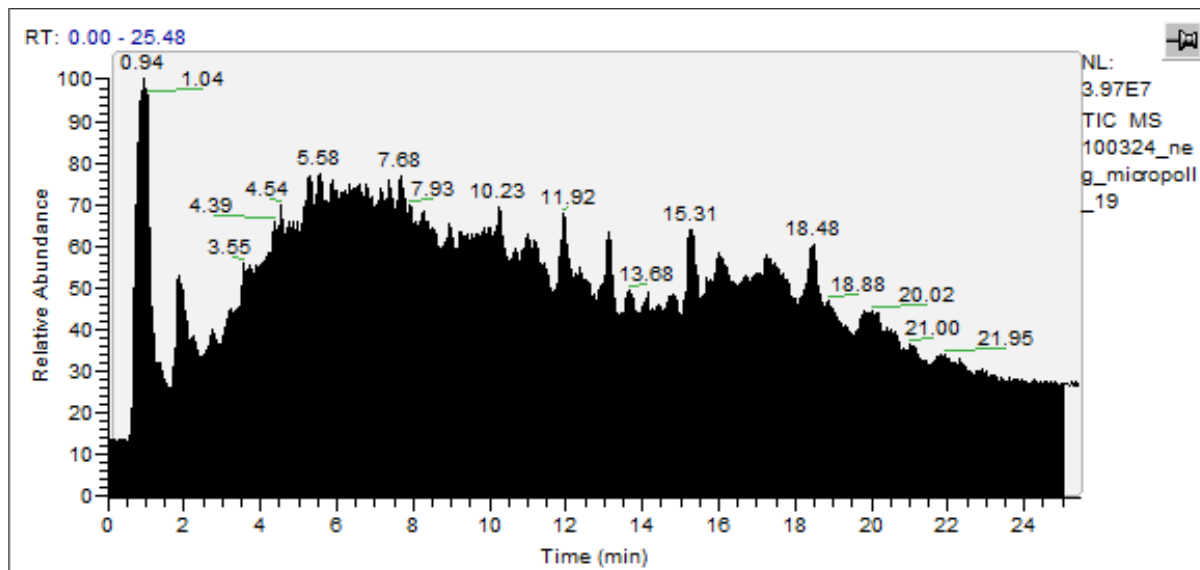
Background: From Molecule to Mass I

- General scheme of mass spectrometry

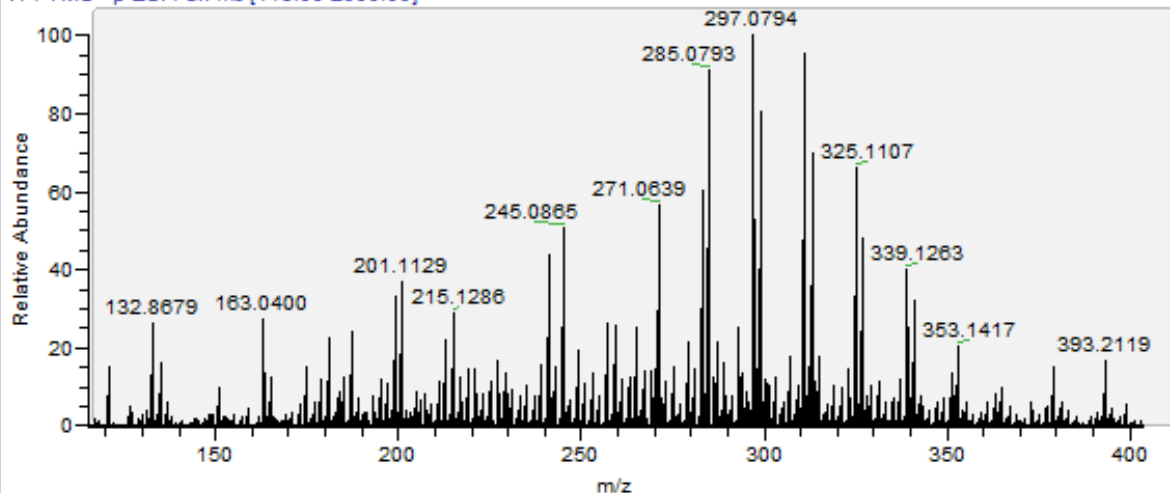


Background: From Molecule to Mass II

- This is what the output “really” looks like ...



100324_neg_micropoll_19 #656-1851 RT: 4.64-13.14 AV: 171 NL: 3.27E5
T: FTMS - p ESI Full ms [115.00-2000.00]



Background: From Molecule to Mass ...and back

- Identification = turning numbers into structures

| Centroid m/z | Intensity | Peak S/N | Scan Number | RT (min.) |
|--------------|-----------|----------|-------------|--------------|
| 161.98619 | 9951899 | 16703.46 | 267 | |
| 297 | | | | |
| Centroid m/z | Intensity | Peak S/N | Scan Number | RT (min.) |
| 301328 | 15995 | 28566120 | 10553.35 | |
| Centroid m/z | Intensity | Peak S/N | Scan Number | RT (min.) |
| 313.10996 | 11814826 | 19088.71 | 1331 | 9.4 6/9 4.77 |
| Centroid m/z | Intensity | Peak S/N | Scan Number | RT (min.) |
| 341 | | | | |
| 261418 | | | | |
| 327 | | | | |
| 313308 | | | | |
| 217 | | | | |
| 303344 | | | | |
| 261 | | | | |
| 285130 | | | | |
| 299 | | | | |
| 158322 | | | | |
| 161 | | | | |
| 245374 | | | | |
| 270 | | | | |
| 327462 | | | | |
| 309 | | | | |
| 201388 | | | | |
| 355 | | | | |
| 271256 | | | | |
| 285 | | | | |
| 311300 | | | | |
| 303 | | | | |
| 309149 | | | | |
| 297 | | | | |
| 257152 | | | | |
| 199 | | | | |
| 232119 | | | | |
| 295 | | | | |
| 341279 | | | | |
| 311 | | | | |
| 217240 | | | | |
| 315 | | | | |
| 121280 | | | | |
| 309 | | | | |
| 287432 | | | | |
| 209 | | | | |
| 163250 | | | | |
| 257 | | | | |
| 265399 | | | | |
| 181 | | | | |
| 209291 | | | | |
| 197 | | | | |
| 259195 | | | | |
| 271 | | | | |
| 506 | | | | |
| 232 | | | | |
| 476 | | | | |
| 283 | | | | |
| 158 | | | | |
| 192 | | | | |
| 418 | | | | |
| 22032950 | | | | |



MassBank

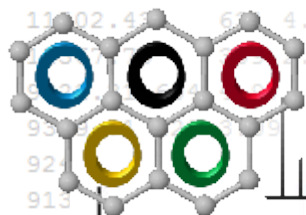
massbank.eu



enviPat: isotope pattern calculator

enviPat Web 1.9

ChemSpider
Search and share chemistry



MOLGEN
STRUCTURE ELUCIDATION



Chemistry Dashboard

RMassBank

MASS FRONTIER
Confident Path from Spectra to Structure

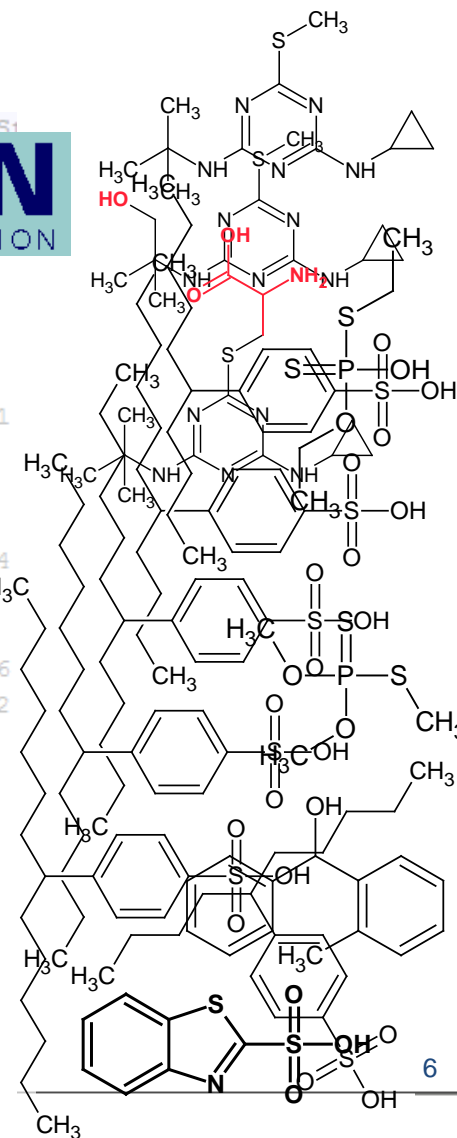


Metrag

Batch mode

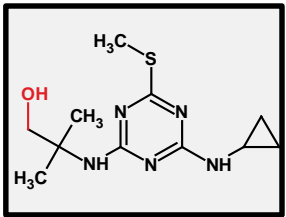
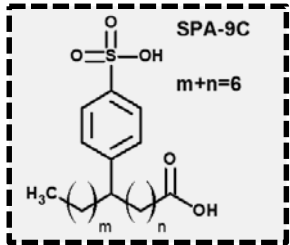
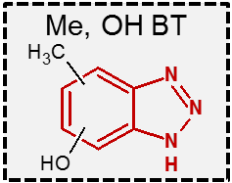
Run scanning in batch mode

| Box? | Status |
|---|--------|
| <input type="checkbox"/> Spark removal | |
| <input type="checkbox"/> Blank subtraction | |
| <input type="checkbox"/> Recalibration | |
| <input type="checkbox"/> Internal standard screening | |
| <input type="checkbox"/> Target screening | |
| <input type="checkbox"/> Target quantification | |
| <input type="checkbox"/> Adduct search for targets & int. stand. | |
| <input type="checkbox"/> Search for non-ionized isomeric peaks | |
| <input type="checkbox"/> Adduct search for non-ionized isomeric peaks | |
| <input type="checkbox"/> Filter sample peak list | |

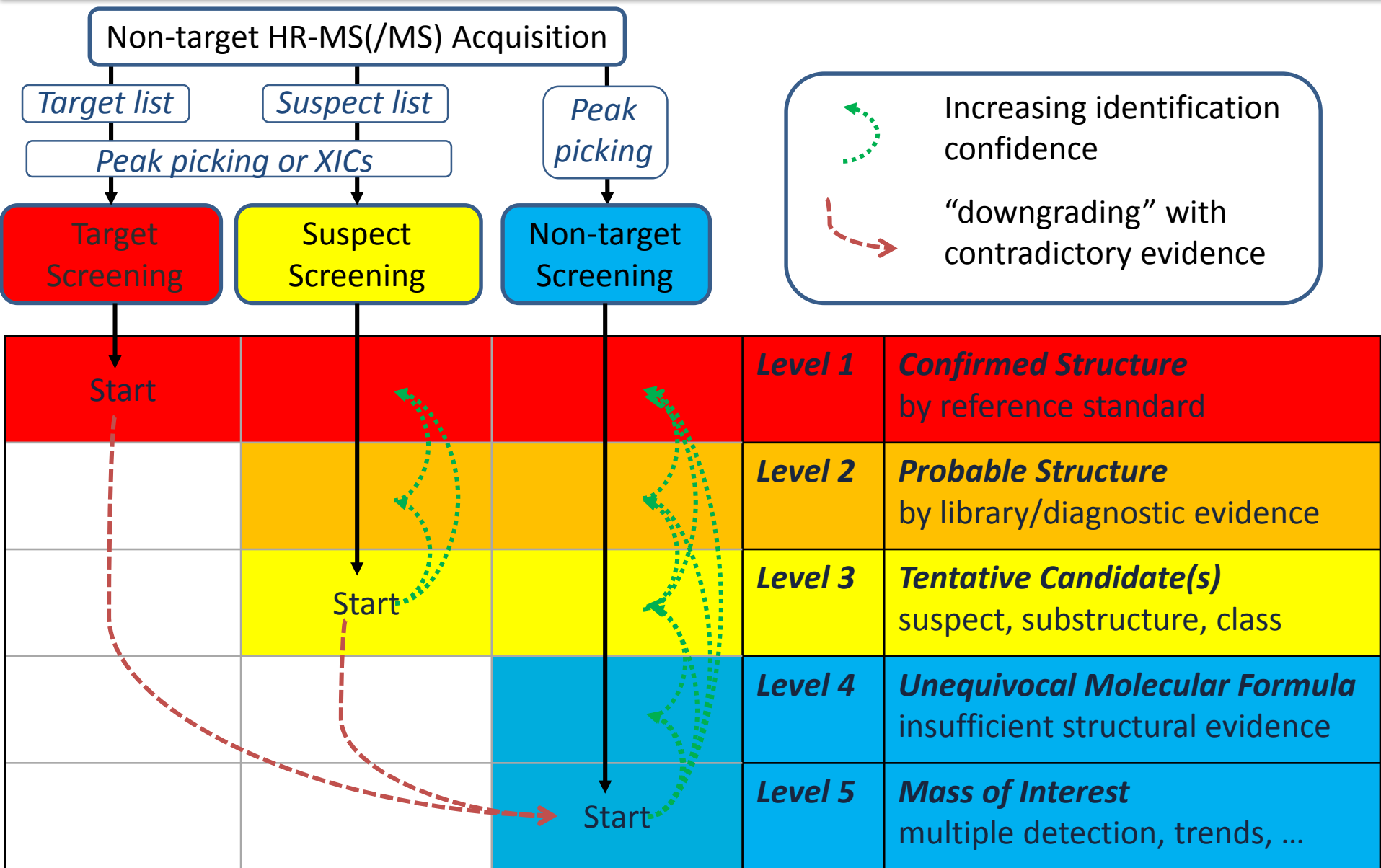


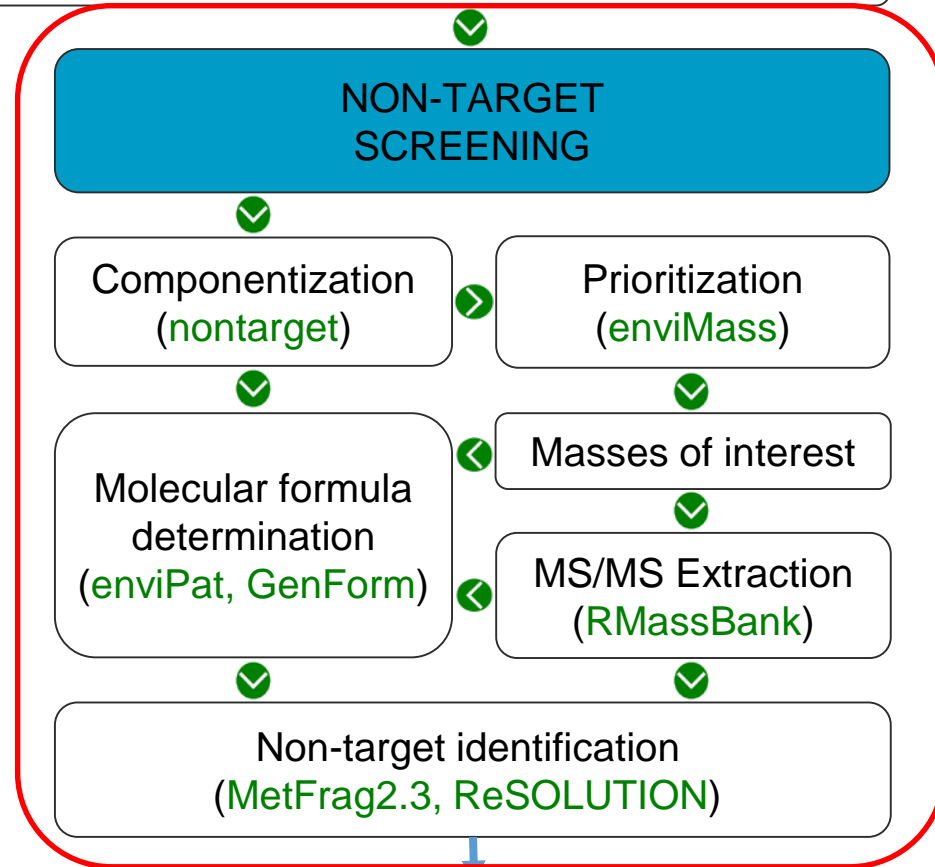
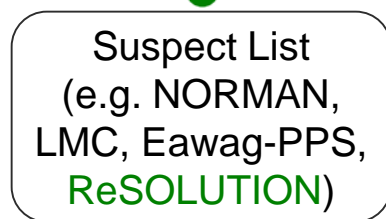
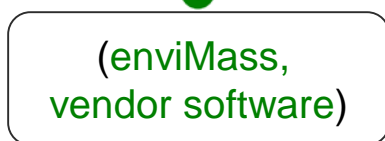
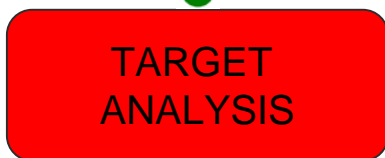
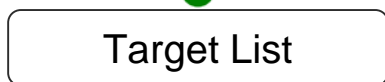
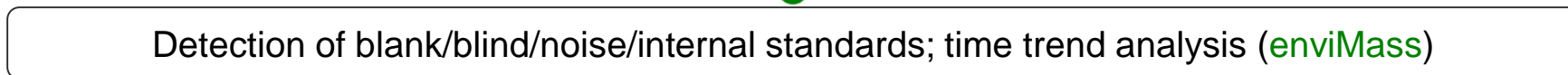
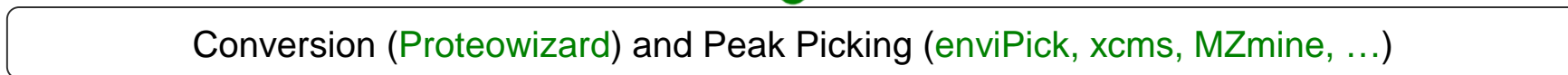
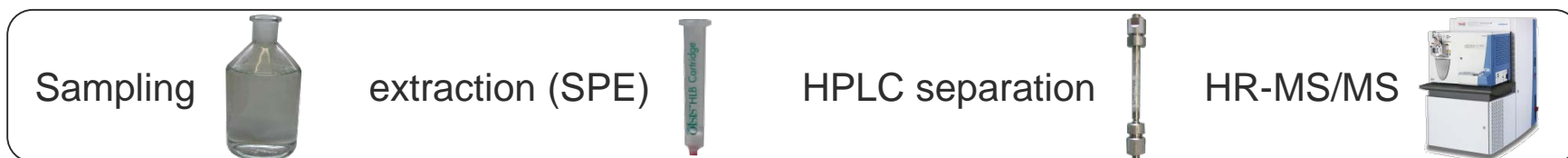
Confidence Levels for Tentative Structures

○ Annotation is the key to communicating information

| Example | Identification confidence | Minimum data requirements |
|---|---|--|
|  | Level 1: Confirmed structure by reference standard | MS, MS ² , RT, Reference Std. |
| | Level 2: Probable structure a) by library spectrum match b) by diagnostic evidence | MS, MS ² , Library MS ² MS, MS ² , Exp. data |
|  | Level 3: Tentative candidate(s) structure, substituent, class | MS, MS ² , Exp. data |
|  | Level 4: Unequivocal molecular formula | MS isotope/adduct |
| $C_6H_5N_3O_4$ | | |
| 192.0757 | Level 5: Exact mass of interest | MS |

Identification Approaches and Terminology





(Molecule =>) Mass => Molecule => Metadata

- Background

- Molecule to Mass

- Preparation for Mass to Molecule

- Gathering the “evidence”

- Molecular Formula example

- Structure Elucidation with MetFrag

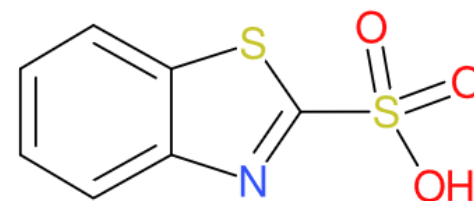
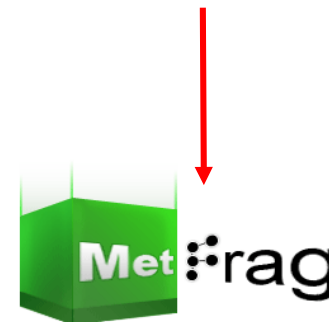
- Step-by-step with one example (MetFrag)

- Practice Session with Several Examples

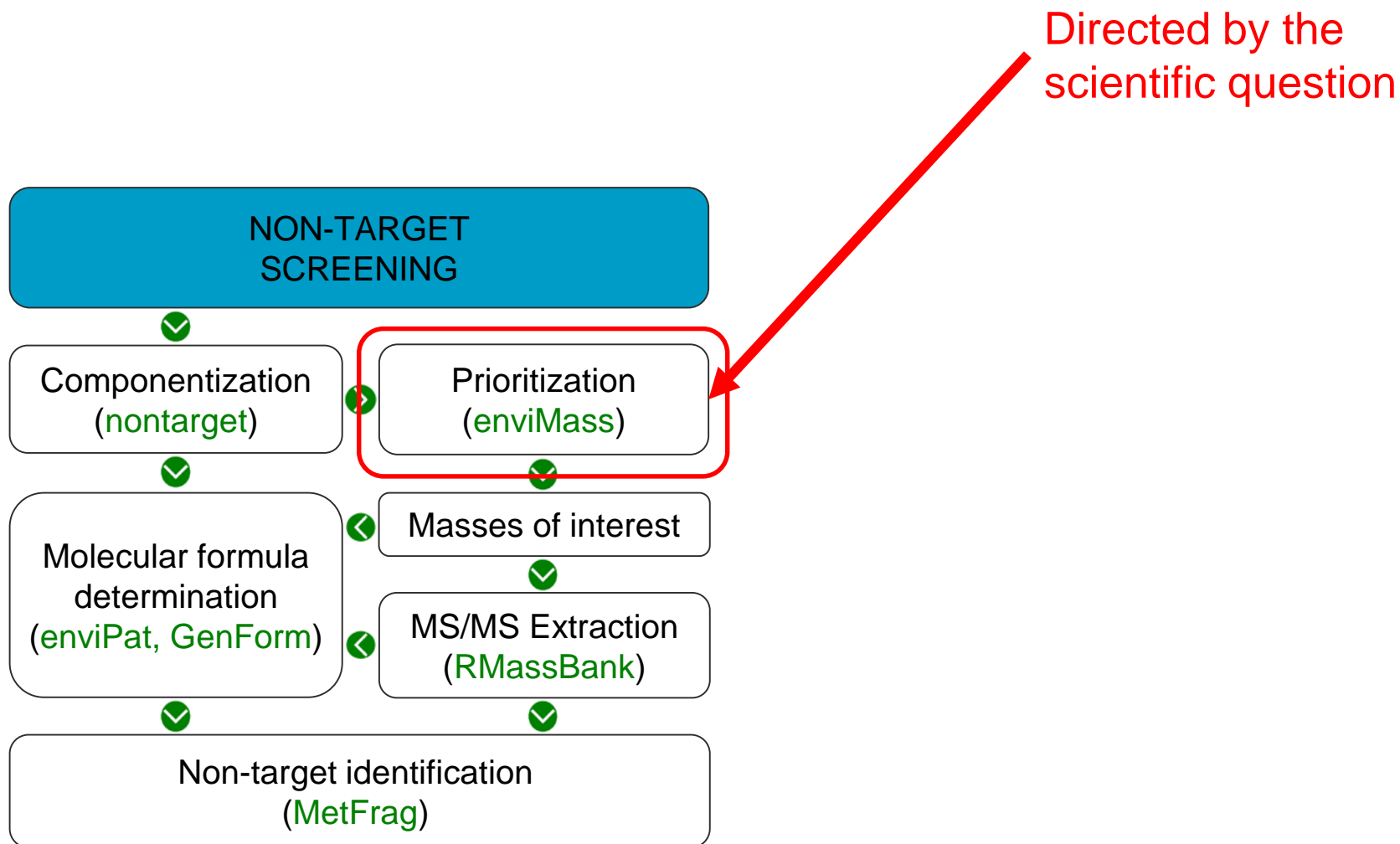
- [Optional] Context and Perspectives

- Potential for automated non-target workflows

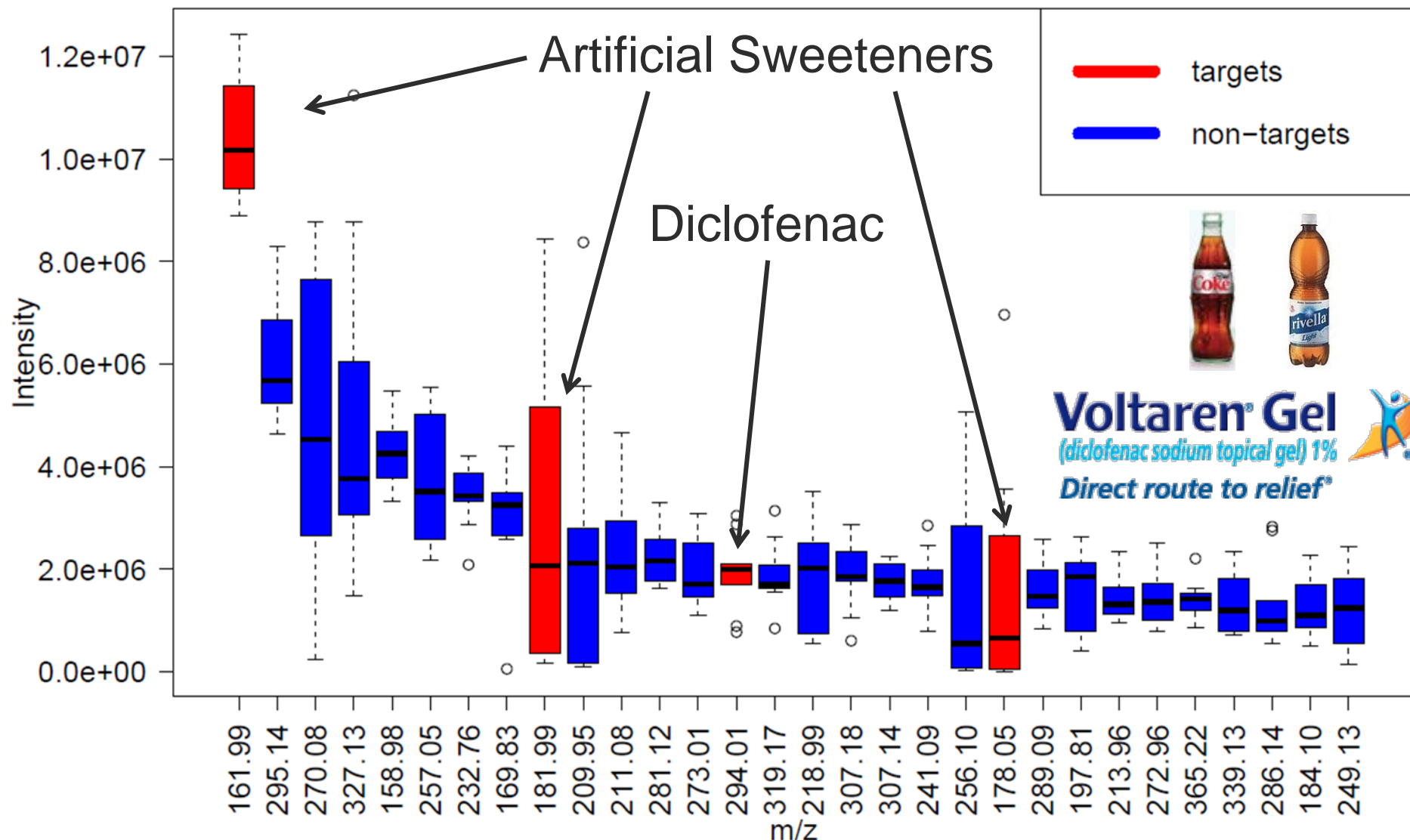
213.9637



Non-target Identification: Where to start?

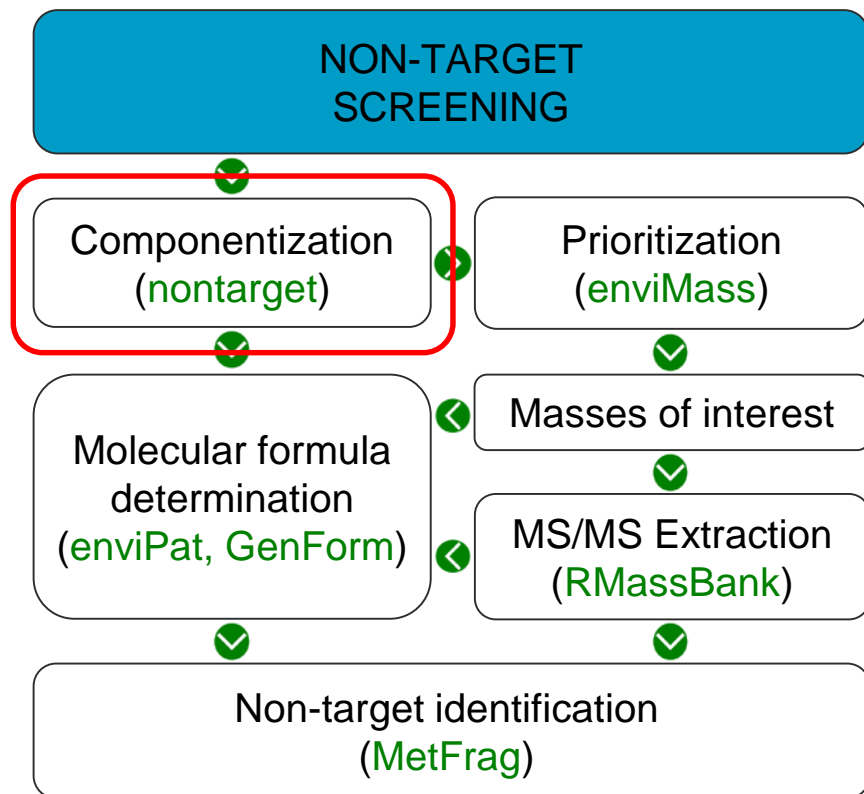


Targets, Non-targets and Isotopes (ESI-) by Intensity



Pictures: www.coca-cola-com; www.rivella.ch; www.voltargengel.com

Non-target Identification: Where to start?



| Centroid m/z | Intensity | Peak S/N | Scan Number | RT (min.) | Start Scan Number | Start RT (min.) | End Scan Number | End RT (min.) | Chrom. S/N |
|--------------|-----------|----------|-------------|-----------|-------------------|-----------------|-----------------|---------------|------------|
| 169.8336 | 3325231 | 5329.19 | 134 | 0.93 | 79 | 0.55 | 646 | 4.55 | 0.991 |
| 299.095 | 3170282 | 4921.75 | 1149 | 8.13 | 595 | 4.18 | 1597 | 11.31 | 0.973 |
| 365.1626 | 3203921 | 4917.57 | 2094 | 14.86 | 2017 | 14.31 | 2200 | 15.61 | 0.986 |
| 163.04 | 2891540 | 4754.39 | 792 | 5.58 | 79 | 0.55 | 3514 | 24.95 | 0.985 |
| 132.8678 | 3211435 | 4736.29 | 148 | 1.03 | 79 | 0.55 | 3514 | 24.95 | 0.938 |
| 309.1519 | 2837347 | 4531.15 | 2493 | 17.69 | 2233 | 15.84 | 2857 | 20.27 | 0.956 |
| 134.8651 | 3063484 | 4518.08 | 148 | 1.03 | 79 | 0.55 | 3514 | 24.95 | 0.951 |
| 234.7607 | 2748982 | 4266.01 | 155 | 1.08 | 79 | 0.55 | 415 | 2.92 | 0.994 |
| 311.095 | 2770672 | 4249.3 | 1114 | 7.88 | 763 | 5.37 | 1513 | 10.7 | 0.961 |
| 293.1747 | 2720875 | 4220.98 | 1688 | 11.95 | 1030 | 7.28 | 2185 | 15.5 | 0.988 |
| 257.0483 | 2577542 | 4130.58 | 750 | 5.28 | 316 | 2.21 | 1366 | 9.67 | 0.988 |
| 174.9557 | 2597792 | 4079.03 | 155 | 1.08 | 79 | 0.55 | 3514 | 24.95 | 0.864 |
| 213.9636 | 2352410 | 3890.96 | 645 | 4.54 | 631 | 4.44 | 778 | 5.48 | 0.995 |
| 167.8364 | 2376150 | 3808.14 | 134 | 0.93 | 79 | 0.55 | 1381 | 9.78 | 0.99 |
| 297.0798 | 2561181 | 3789.71 | 1170 | 8.27 | 679 | 4.78 | 1288 | 9.12 | 0.964 |
| 201.113 | 2299642 | 3706.57 | 1128 | 7.98 | 442 | 3.11 | 1696 | 12.01 | 0.984 |
| 310.1024 | 2335010 | 3645.29 | 1310 | 9.27 | 1288 | 9.12 | 1624 | 11.5 | 0.983 |
| 241.0716 | 2141622 | 3536.71 | 2514 | 17.84 | 1813 | 12.84 | 3514 | 24.95 | 0.855 |
| 374.1311 | 2580159 | 3382.28 | 771 | 5.43 | 709 | 5 | 898 | 6.34 | 0.991 |

Gather Evidence

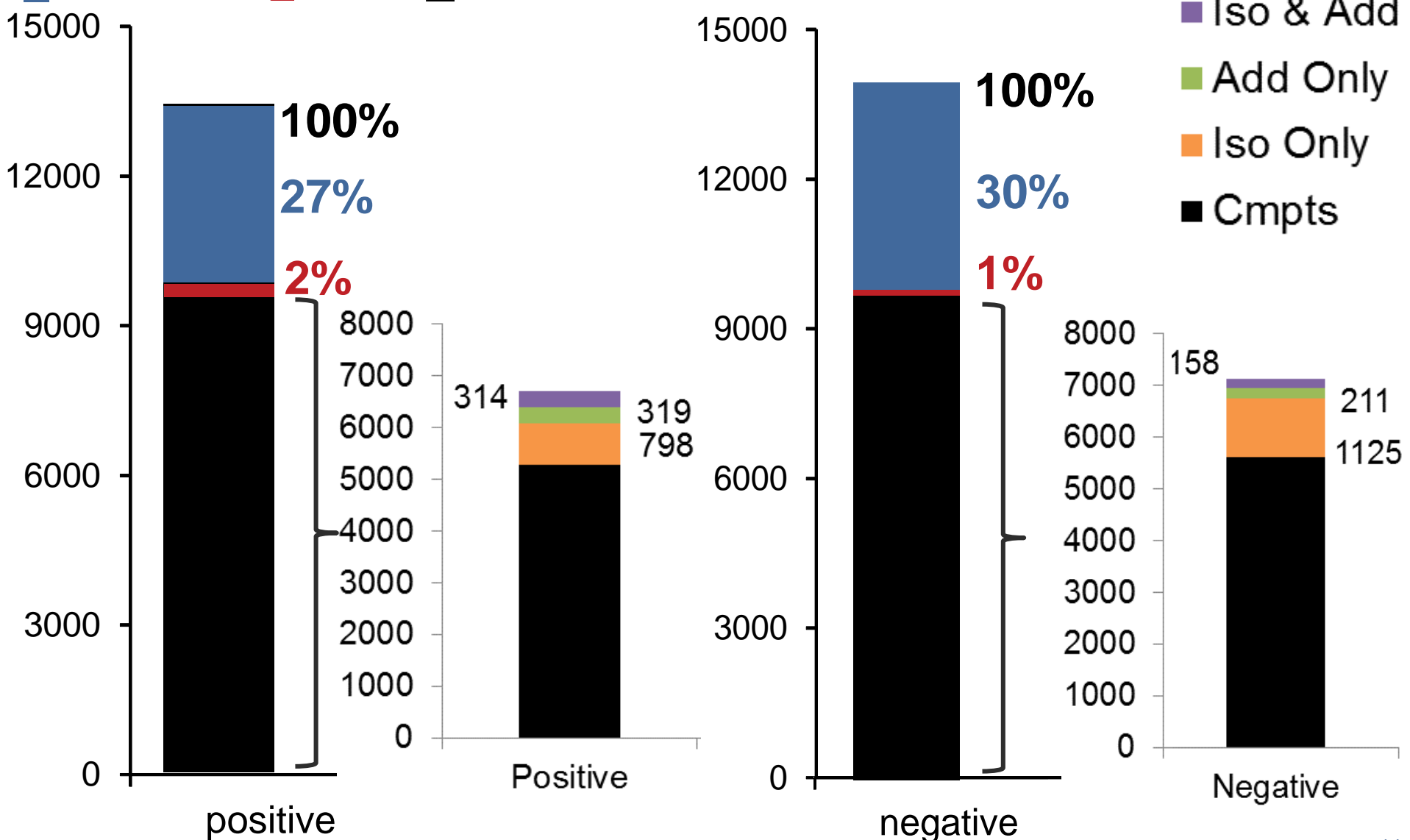
Grouping Adducts and Isotopes



nontarget

enviPat Web

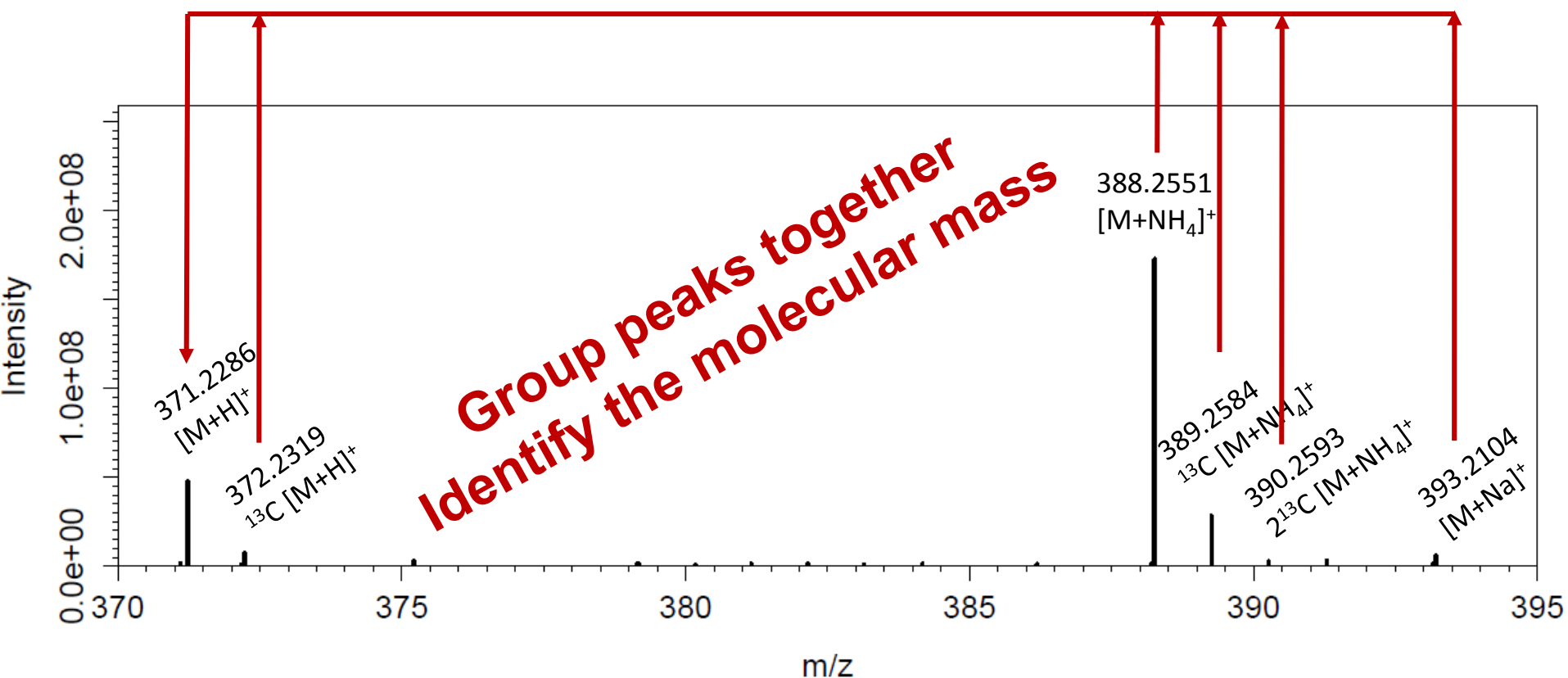
■ Noise/Blank ■ Targets ■ Non-targets



Gathering Evidence for Identification I



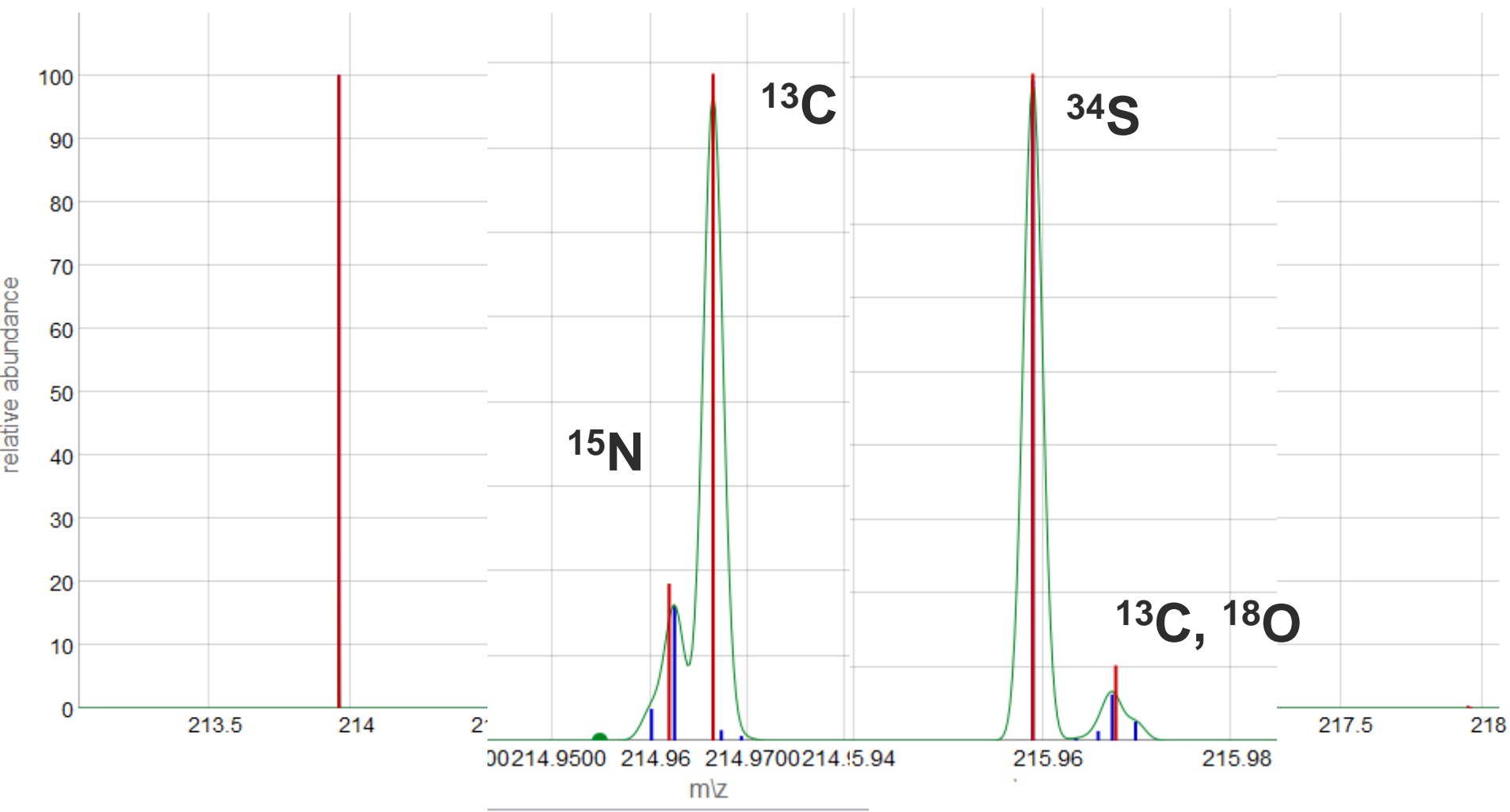
Determining the molecular mass and elements, adducts present **nontarget**



Schymanski *et al.* 2015, ABC,
DOI: 10.1007/s00216-015-8681-7
M. Loos, *et al.* 2015, DOI:
10.1021/acs.analchem.5b00941

<http://www.eawag.ch/forschung/uchem/software/>
<http://cran.r-project.org/web/packages/nontarget/>
<http://cran.r-project.org/web/packages/enviPat/>

Detecting the presence of elements with enviPat and nontarget



M. Loos, *et al.* 2015, DOI: 10.1021/acs.analchem.5b00941

Image © www.seanoakley.com/

<http://www.envipat.eawag.ch/>

<http://cran.r-project.org/web/packages/enviPat/>

Detecting Isotope Signals in Samples

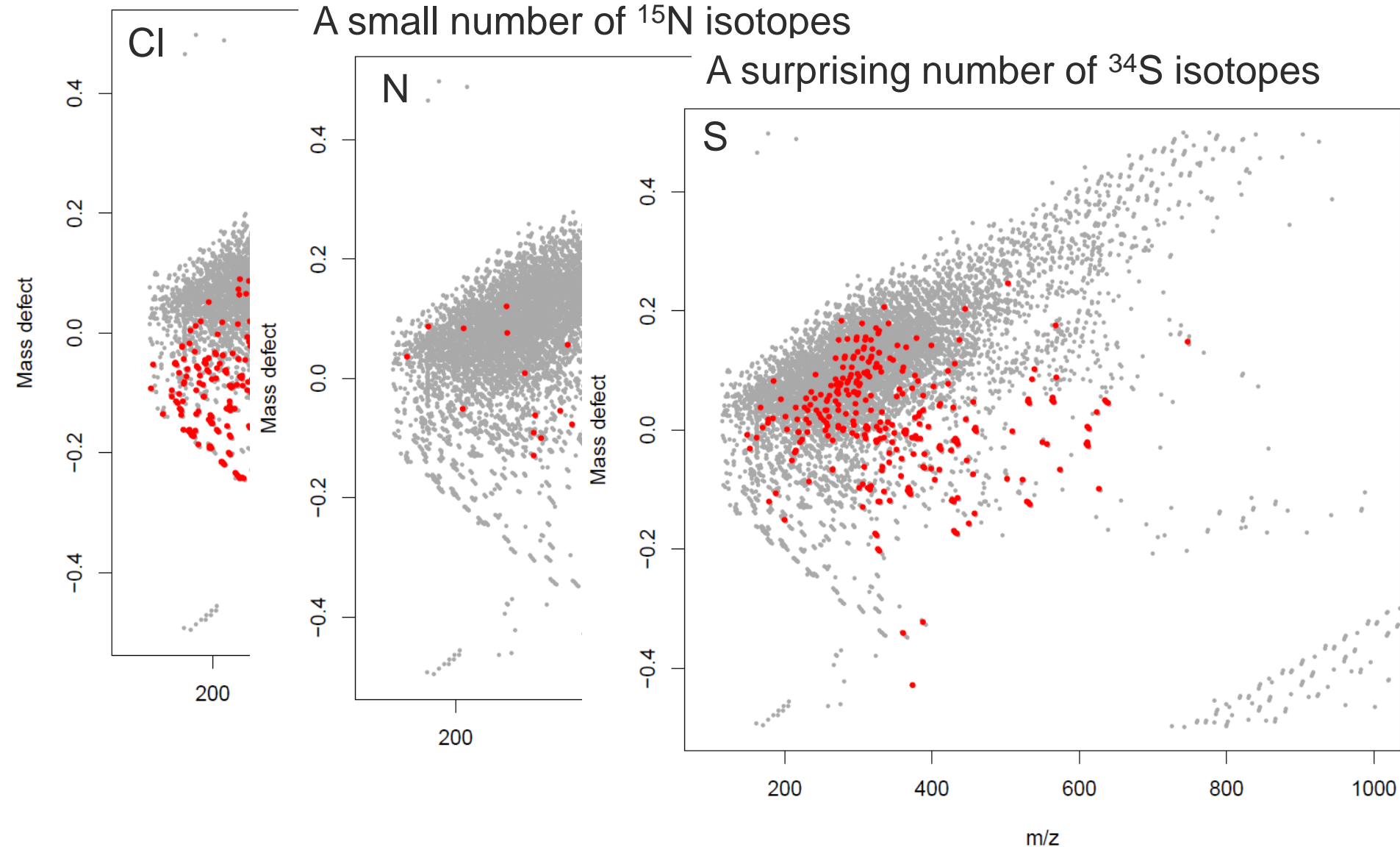


“Classic” environmental strategy: Cl isotopes

nontarget

A small number of ^{15}N isotopes

A surprising number of ^{34}S isotopes



Peak Grouping in Workflows

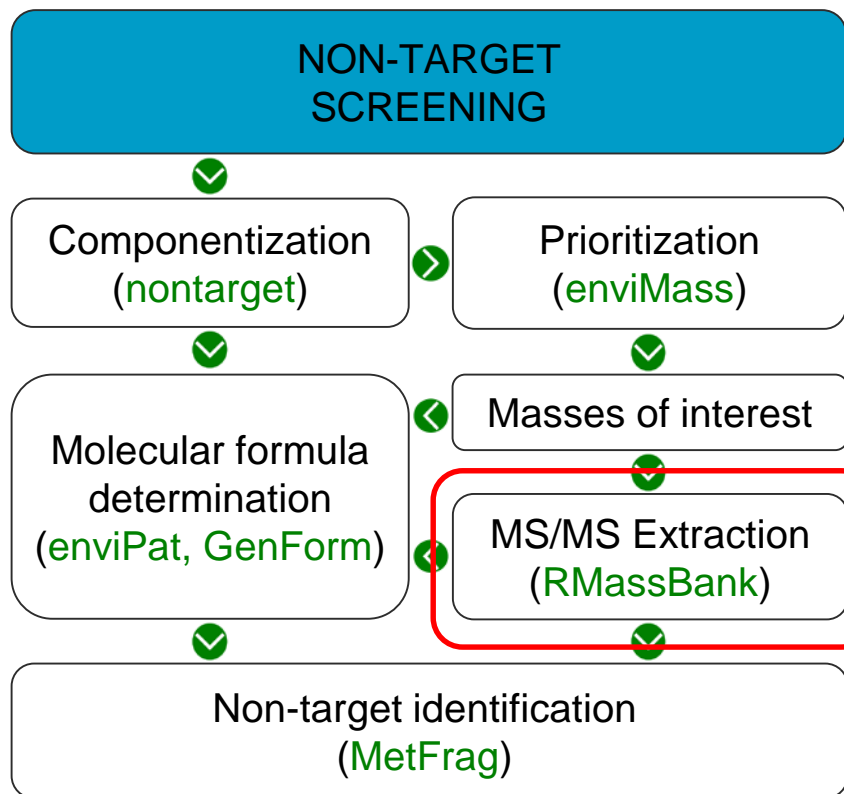
This can be done automatically with nontarget / enviMass (and others)

| m.z | max_int | RT | peak ID | group ID | isotope(s) |
|----------|----------|---------|---------|----------|------------|
| 589.3655 | 47715916 | 1281.84 | 40 | 304 | 13C |
| 329.2335 | 46938304 | 1232.32 | 41 | 0 | none |
| 209.0853 | 44881960 | 1204.91 | 42 | 29 | 13C/34S |

| m/z | int | ret | peak ID | adduct(s) |
|------|----------|----------|---------|--|
| 213 | 315.1726 | 1.73E+08 | 1035.33 | 74 none |
| 440. | 295.2268 | 1.71E+08 | 1399.63 | 75 M+H<->M+Na |
| 236. | 288.2532 | 1.69E+08 | 1359.84 | 76 M+H<->M+Na |
| 221 | 119.019 | 1.66E+08 | 570.906 | 77 none |
| 213 | 315.1057 | 1.63E+08 | 965.48 | 78 M+H<->M+Na |
| 213 | 212.1493 | 1.63E+08 | 608.964 | 79 M+NH4<->M+H//M+NH4<->M+Na//M+NH4<->M+K |
| 251. | 327.0081 | 1.6E+08 | 1244.89 | 80 M+H<->M+Na//M+H<->M+NH4 |
| 273. | 155.1591 | 1.6E+08 | 689.237 | 81 none |
| 246. | 375.2109 | 1.58E+08 | 684.805 | 82 none |
| 246. | 329.0052 | 1.58E+08 | 1211.29 | 83 M+H<->M+Na//M+H<->M+NH4 |
| 329. | 356.1768 | 1.57E+08 | 1000.85 | 84 none |
| | 240.1523 | 1.54E+08 | 682.152 | 85 M+H<->M+Na//M+H<->M+NH4//M+H<->M+K |
| | 300.2016 | 1.52E+08 | 686.578 | 86 none |
| | 404.2068 | 1.51E+08 | 1224.56 | 87 M+K<->M+H//M+NH4<->M+H//M+NH4<->M+Na//M+NH4<->M+K |
| | 219.1167 | 1.5E+08 | 1167.08 | 88 M+H<->M+Na |
| | 158.154 | 1.45E+08 | 1211.29 | 89 M+H<->M+Na |

This information can be used as input later...

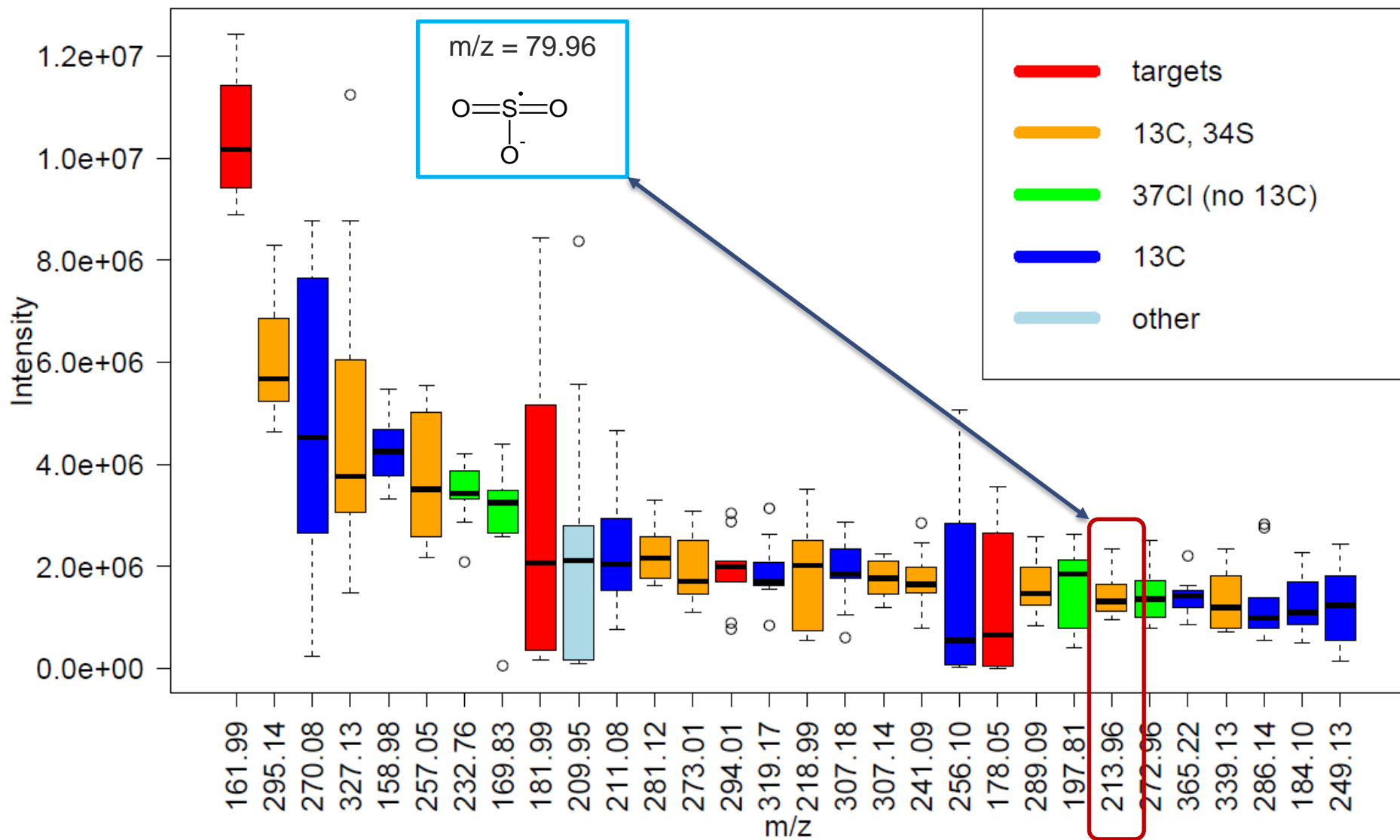
Non-target Identification: Where to start?

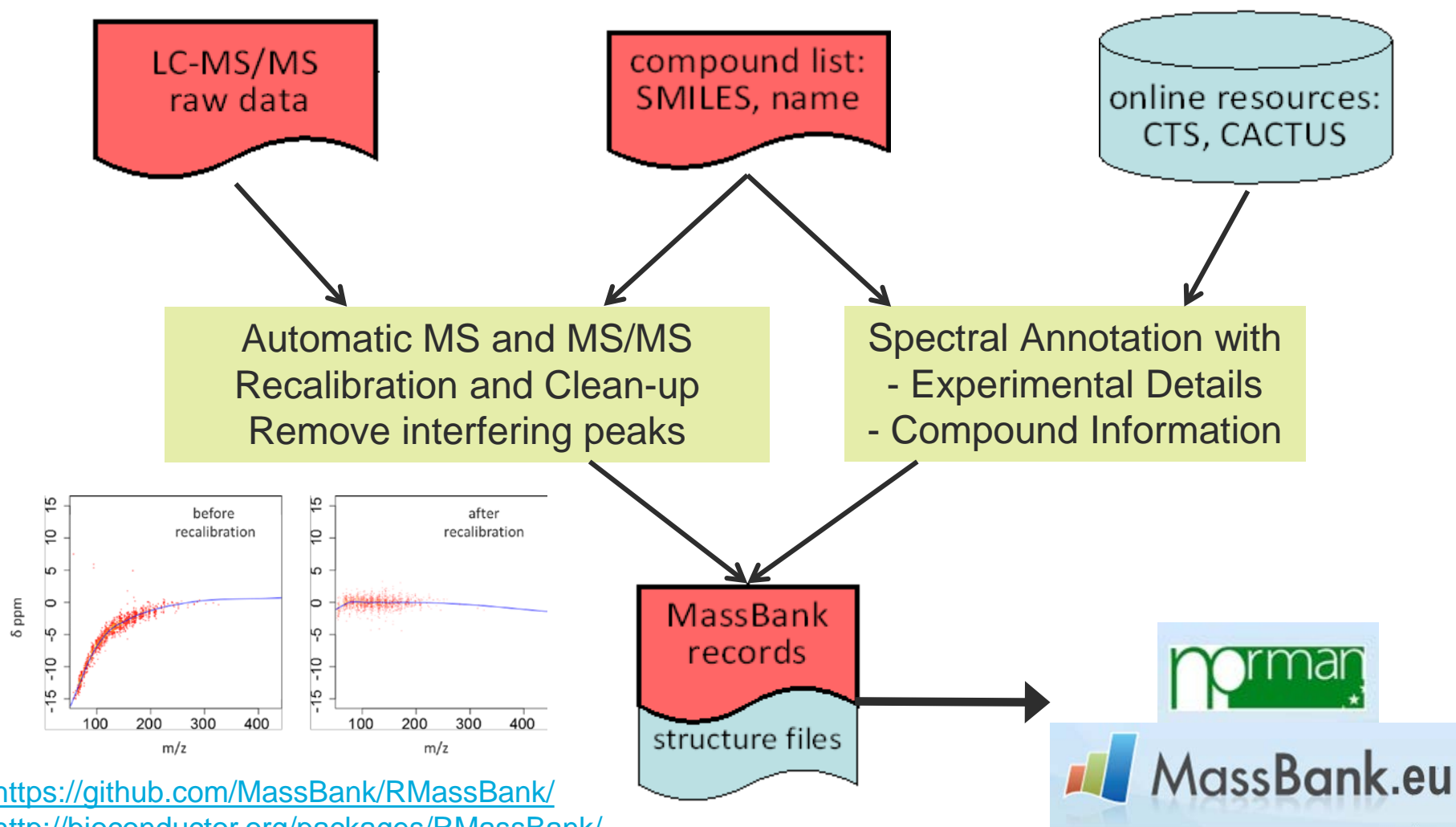


| Centroid m/z | Intensity | Peak S/N | Scan Number | RT (min.) | Start Scan Number | Start RT (min.) | End Scan Number | End RT (min.) | Chrom. S/N |
|-----------------|-----------|----------|----------------|-----------|----------------------|--------------------|--------------------|------------------|---------------|
| 169.8336 | 3325231 | 5329.19 | 134 | 0.93 | 79 | 0.55 | 646 | 4.55 | 0.991 |
| 299.095 | 3170282 | 4921.75 | 1149 | 8.13 | 595 | 4.18 | 1597 | 11.31 | 0.973 |
| 365.1626 | 3203921 | 4917.57 | 2094 | 14.86 | 2017 | 14.31 | 2200 | 15.61 | 0.986 |
| 163.04 | 2891540 | 4754.39 | 792 | 5.58 | 79 | 0.55 | 3514 | 24.95 | 0.985 |
| 132.8678 | 3211435 | 4736.29 | 148 | 1.03 | 79 | 0.55 | 3514 | 24.95 | 0.938 |
| 309.1519 | 2837347 | 4531.15 | 2493 | 17.69 | 2233 | 15.84 | 2857 | 20.27 | 0.956 |
| 134.8651 | 3063484 | 4518.08 | 148 | 1.03 | 79 | 0.55 | 3514 | 24.95 | 0.951 |
| 234.7607 | 2748982 | 4266.01 | 155 | 1.08 | 79 | 0.55 | 415 | 2.92 | 0.994 |
| 311.095 | 2770672 | 4249.3 | 1114 | 7.88 | 763 | 5.37 | 1513 | 10.7 | 0.961 |
| 293.1747 | 2720875 | 4220.98 | 1688 | 11.95 | 1030 | 7.28 | 2185 | 15.5 | 0.988 |
| 257.0483 | 2577542 | 4130.58 | 750 | 5.28 | 316 | 2.21 | 1366 | 9.67 | 0.988 |
| 174.9557 | 2597792 | 4079.03 | 155 | 1.08 | 79 | 0.55 | 3514 | 24.95 | 0.864 |
| 213.9636 | 2352410 | 3890.96 | 645 | 4.54 | 631 | 4.44 | 778 | 5.48 | 0.995 |
| 167.8364 | 2376150 | 3808.14 | 134 | 0.93 | 79 | 0.55 | 1381 | 9.78 | 0.99 |
| 297.0798 | 2561181 | 3789.71 | 1170 | 8.27 | 679 | 4.78 | 1288 | 9.12 | 0.964 |
| 201.113 | 2299642 | 3706.57 | 1128 | 7.98 | 442 | 3.11 | 1696 | 12.01 | 0.984 |
| 310.1024 | 2335010 | 3645.29 | 1310 | 9.27 | 1288 | 9.12 | 1624 | 11.5 | 0.983 |
| 241.0716 | 2141622 | 3536.71 | 2514 | 17.84 | 1813 | 12.84 | 3514 | 24.95 | 0.855 |
| 374.1311 | 2580159 | 3382.28 | 771 | 5.43 | 709 | 5 | 898 | 6.34 | 0.991 |

Gather
Evidence

Targets, Non-targets and Isotopes (ESI-)



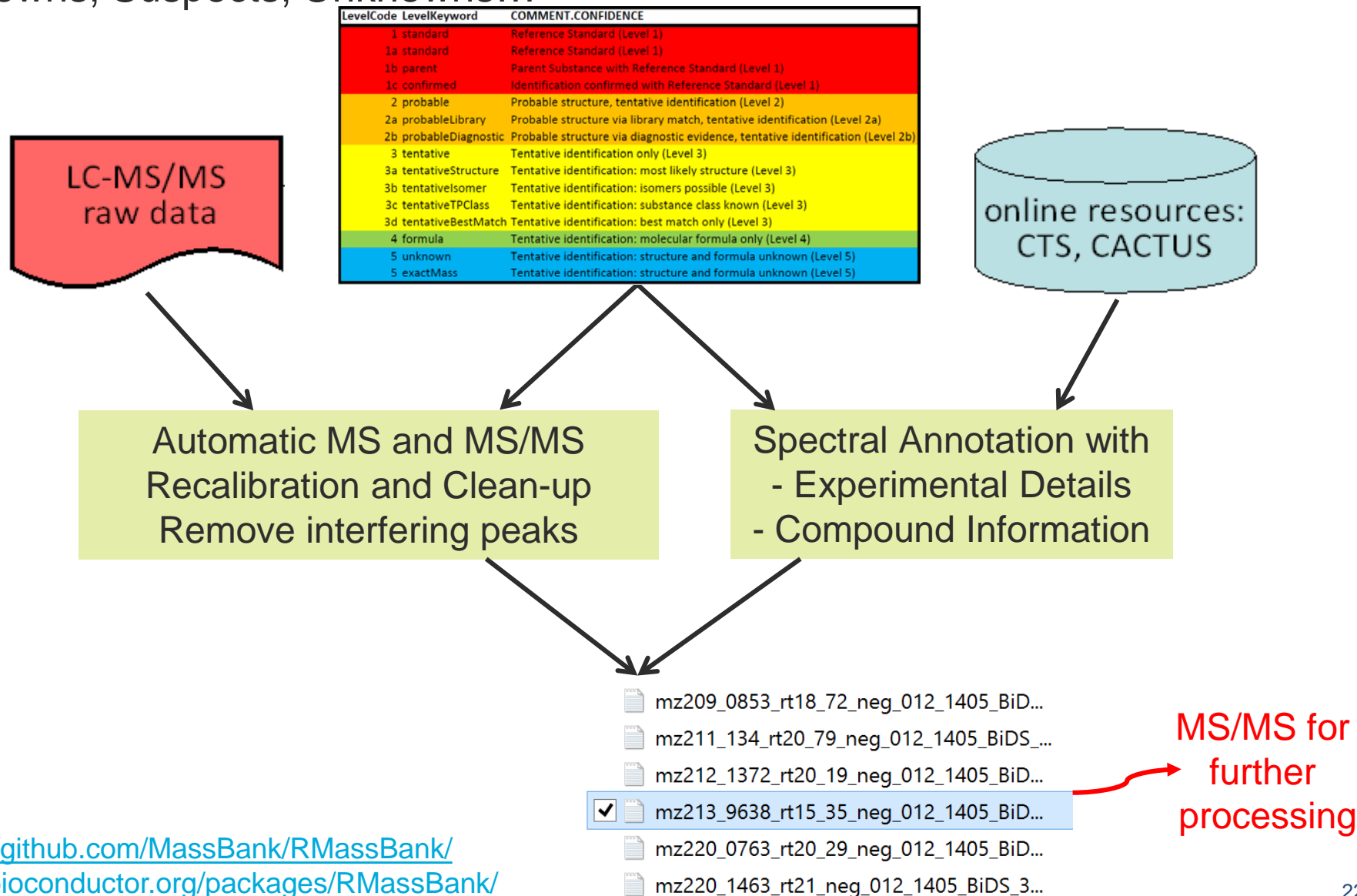


<https://github.com/MassBank/RMassBank/>
<http://bioconductor.org/packages/RMassBank/>

Stravs, Schymanski, Singer and Hollender, 2013,
Journal of Mass Spectrometry, 48, 89–99. DOI: 10.1002/jms.3131

MS2: Extracting Mass Spectra II

Knowns, Suspects, Unknowns...



(Molecule =>) Mass => Molecule => Metadata

- Background

- Molecule to Mass

- Preparation for Mass to Molecule

- Gathering the “evidence”

- Molecular Formula example

- Structure Elucidation with MetFrag

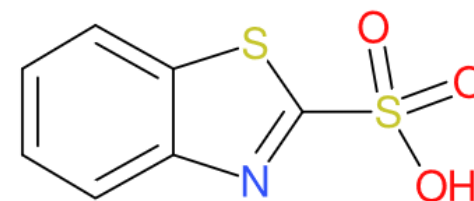
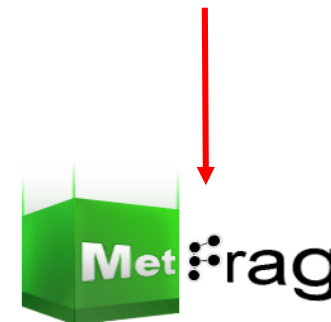
- Step-by-step with one example (MetFrag)

- Practice Session with Several Examples

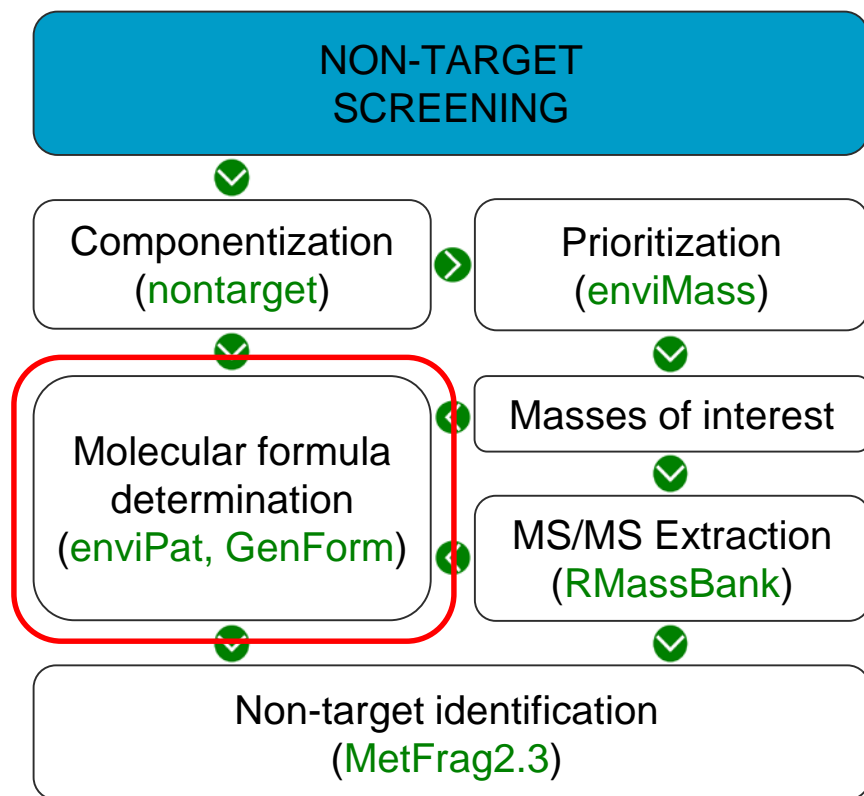
- [Optional] Context and Perspectives

- Potential for automated non-target workflows

213.9637



Non-target Identification: Molecular Formula



| Centroid m/z | Intensity | Peak S/N | Scan Number | RT (min.) | Start Scan Number | Start RT (min.) | End Scan Number | End RT (min.) | Chrom. S/N |
|-----------------|-----------|----------|----------------|-----------|----------------------|--------------------|--------------------|------------------|---------------|
| 169.8336 | 3325231 | 5329.19 | 134 | 0.93 | 79 | 0.55 | 646 | 4.55 | 0.991 |
| 299.095 | 3170282 | 4921.75 | 1149 | 8.13 | 595 | 4.18 | 1597 | 11.31 | 0.973 |
| 365.1626 | 3203921 | 4917.57 | 2094 | 14.86 | 2017 | 14.31 | 2200 | 15.61 | 0.986 |
| 163.04 | 2891540 | 4754.39 | 792 | 5.58 | 79 | 0.55 | 3514 | 24.95 | 0.985 |
| 132.8678 | 3211435 | 4736.29 | 148 | 1.03 | 79 | 0.55 | 3514 | 24.95 | 0.938 |
| 309.1519 | 2837347 | 4531.15 | 2493 | 17.69 | 2233 | 15.84 | 2857 | 20.27 | 0.956 |
| 134.8651 | 3063484 | 4518.08 | 148 | 1.03 | 79 | 0.55 | 3514 | 24.95 | 0.951 |
| 234.7607 | 2748982 | 4266.01 | 155 | 1.08 | 79 | 0.55 | 415 | 2.92 | 0.994 |
| 311.095 | 2770672 | 4249.3 | 1114 | 7.88 | 763 | 5.37 | 1513 | 10.7 | 0.961 |
| 293.1747 | 2720875 | 4220.98 | 1688 | 11.95 | 1030 | 7.28 | 2185 | 15.5 | 0.988 |
| 257.0483 | 2577542 | 4130.58 | 750 | 5.28 | 316 | 2.21 | 1366 | 9.67 | 0.988 |
| 174.9557 | 2597792 | 4079.03 | 155 | 1.08 | 79 | 0.55 | 3514 | 24.95 | 0.864 |
| 213.9636 | 2352410 | 3890.96 | 645 | 4.54 | 631 | 4.44 | 778 | 5.48 | 0.995 |
| 167.8364 | 2376150 | 3808.14 | 134 | 0.93 | 79 | 0.55 | 1381 | 9.78 | 0.99 |
| 297.0798 | 2561181 | 3789.71 | 1170 | 8.27 | 679 | 4.78 | 1288 | 9.12 | 0.964 |
| 201.113 | 2299642 | 3706.57 | 1128 | 7.98 | 442 | 3.11 | 1696 | 12.01 | 0.984 |
| 310.1024 | 2335010 | 3645.29 | 1310 | 9.77 | 1288 | 9.12 | 1624 | 11.5 | 0.983 |
| 241.0716 | 2141622 | 3536.71 | 2514 | 17.84 | 1813 | 12.84 | 3514 | 24.95 | 0.855 |
| 374.1311 | 2580159 | 3382.28 | 771 | 5.43 | 709 | 5 | 898 | 6.34 | 0.991 |

MS1 :

| | | |
|----------|-----------|--------------------|
| 213.9637 | 1168637.0 | [M-H] ⁻ |
| 214.9630 | 14790.0 | M+1/15N |
| 214.9670 | 94077.2 | M+1/13C |
| 215.9595 | 104643.6 | M+2/34S |
| 215.9679 | 7060.2 | M+2/13C |

MS / MS :

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |

Generating Molecular Formulas with GenForm

MS:

| | | |
|----------|-----------|--------------------|
| 213.9637 | 1168637.0 | [M-H] ⁻ |
| 214.9630 | 14790.0 | M+1/15N |
| 214.9670 | 94077.2 | M+1/13C |
| 215.9595 | 104643.6 | M+2/34S |
| 215.9679 | 7060.2 | M+2/13C |

MSMS:

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |

GenForm (formerly MOLGEN-MS/MS)¹

| Formulas | ppm | MS_MV | MSMS_MV | CombMV |
|------------|------|---------|---------|---------|
| C6HNO8 | -3.5 | 0.92525 | 0.33333 | 0.30842 |
| C9H2N3PS | -1.3 | 0.96281 | 0.33333 | 0.32094 |
| CH8N5P3S | 1.6 | 0.91823 | 0.33333 | 0.30608 |
| C7H5NO3S2 | 0.5 | 0.98487 | 1.00000 | 0.98487 |
| CH6N5O2PS2 | -4.3 | 0.96260 | 0.66667 | 0.64173 |

¹M. Meringer et al. (2011), *MATCH Commun. Math. Comput. Chem.* 65, 259-290.

Generating Molecular Formulas with GenForm

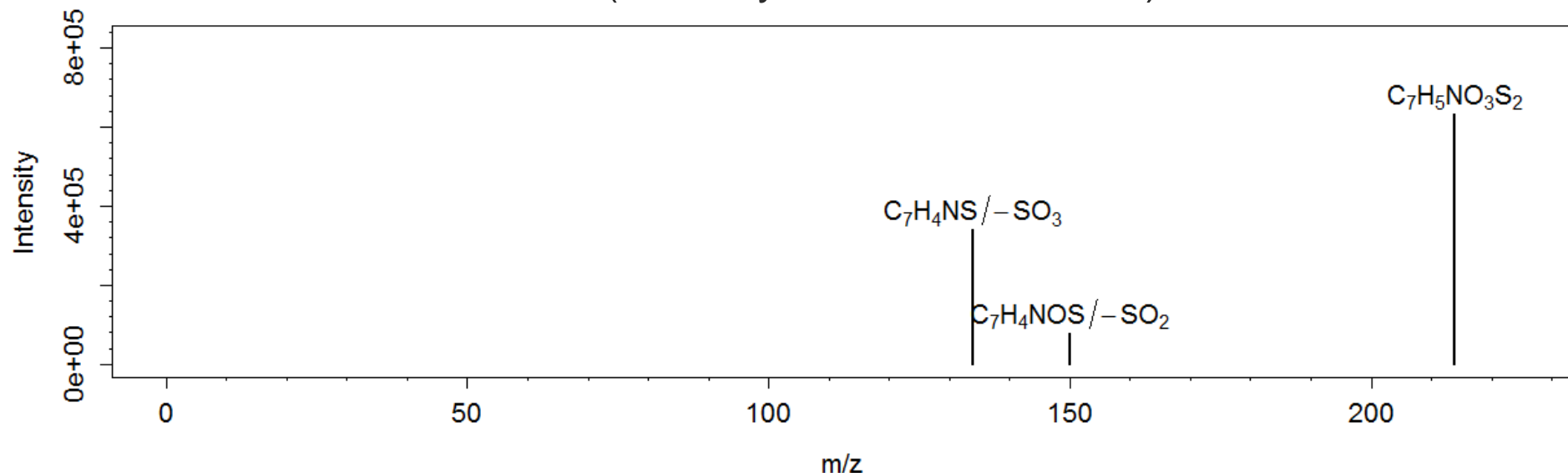
MS:

| | | |
|----------|-----------|----------------------|
| 213.9637 | 1168637.0 | [M-H] ⁻ |
| 214.9630 | 14790.0 | M+1/ ¹⁵ N |
| 214.9670 | 94077.2 | M+1/ ¹³ C |
| 215.9595 | 104643.6 | M+2/ ³⁴ S |
| 215.9679 | 7060.2 | M+2/ ¹³ C |

MSMS:

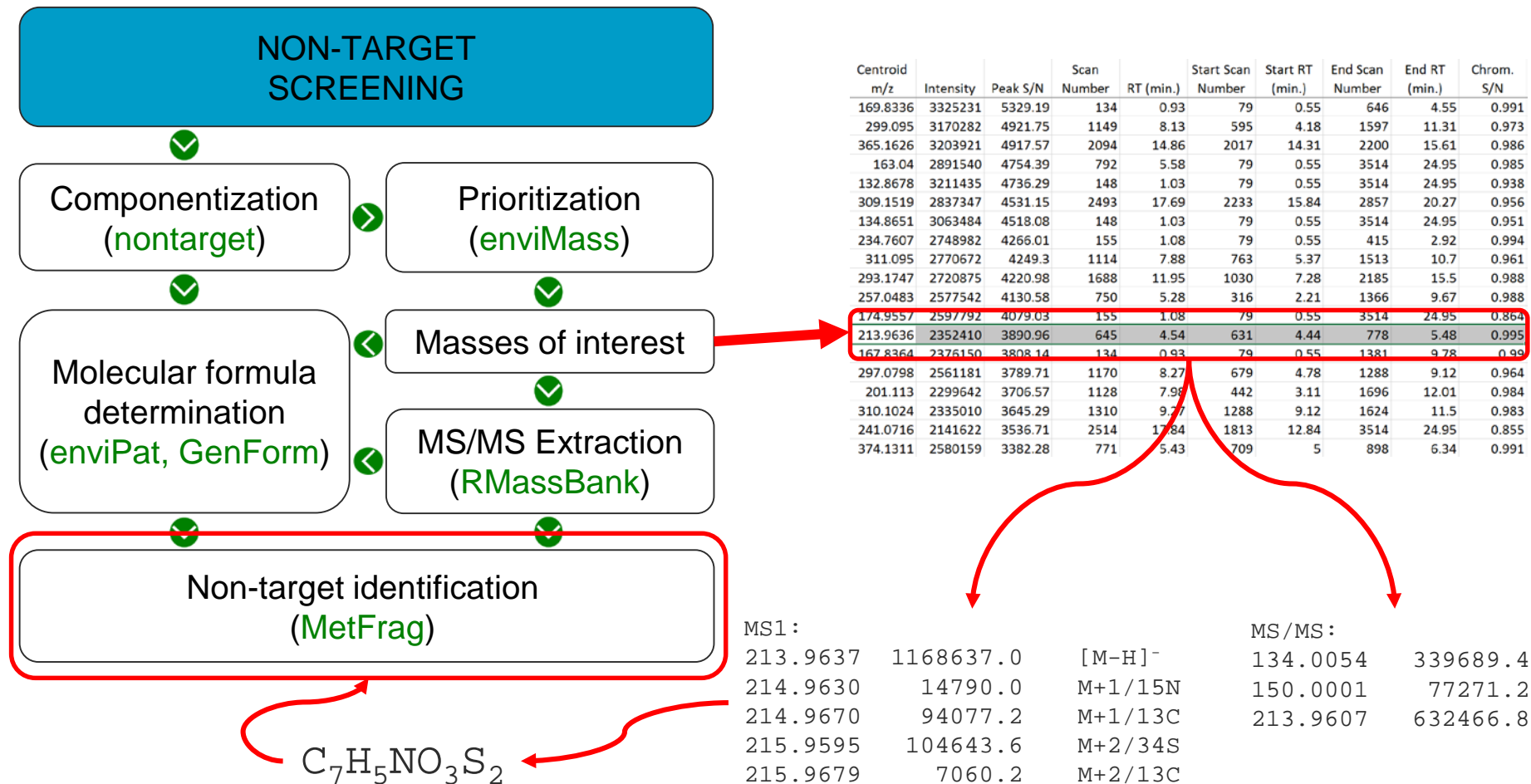
| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |

GenForm (formerly MOLGEN-MS/MS)¹



¹M. Meringer et al. (2011), *MATCH Commun. Math. Comput. Chem.* 65, 259-290.

Non-target Identification: Next Step: Candidates!



(Molecule =>) Mass => Molecule => Metadata

- Background

- Molecule to Mass

- Preparation for Mass to Molecule

- Gathering the “evidence”

- Molecular Formula example

- Structure Elucidation with MetFrag

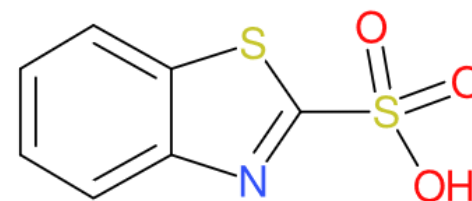
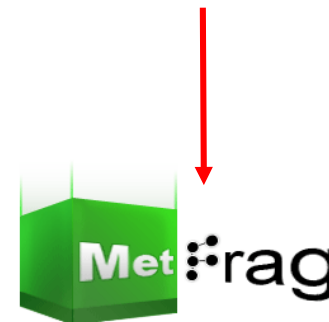
- Step-by-step with one example (MetFrag)

- Practice Session with Several Examples

- [Optional] Context and Perspectives

- Potential for automated non-target workflows

213.9637



MetFrag: *In silico* non-target identification

Status: 2010

m/z $[M-H]^-$
213.9637
 ± 5 ppm

5 ppm
0.001 Da

PubChem



135 Candidates

MS/MS

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |

Database Settings

Database: Include references: ☒ Parent Ion:

Neutral Mass: Search ppm:

Formula:

Identifiers:

 135 Candidates

Candidate Filter & Score Settings

Fragmentation Settings & Processing

Mzppm:

Mzabs:

Mode:

Tree depth:

Group candidates ☒

MS/MS Peak list

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |

Test set of 473 Eawag Target Substances

<http://ipb-halle.github.io/MetFrag/>

| | MetFrag 2010 | New MetFrag Fragments only | |
|-------------------------------|-----------------|----------------------------------|--|
| ChemSpider¹ | | | |
| Top 1 Ranks | 73 | 105 | |
| % Top 1 Ranks | 15 % | 22 % | |
| PubChem² | | | |
| Top 1 Ranks | - | 30 | |
| % Top 1 Ranks | | 6 % | |

Still some room for improvement!

¹www.chemspider.com; ~34 million entries

²<https://pubchem.ncbi.nlm.nih.gov/>; ~74 million entries

Database Settings

Database: Include references: ☒ Parent Ion:

Neutral Mass: Search ppm:

Formula:

Identifiers:

 135 Candidates

Candidate Filter & Score Settings

Fragmentation Settings & Processing

Mzppm:

Mzabs:

Mode:

Tree depth:

Group candidates ☒

MS/MS Peak list

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |

MetFrag: *In silico* non-target identification

Status: 2016

m/z $[M-H]^-$
213.9637
 ± 5 ppm

5 ppm
0.001 Da

PubChem

MetFrag



Literature



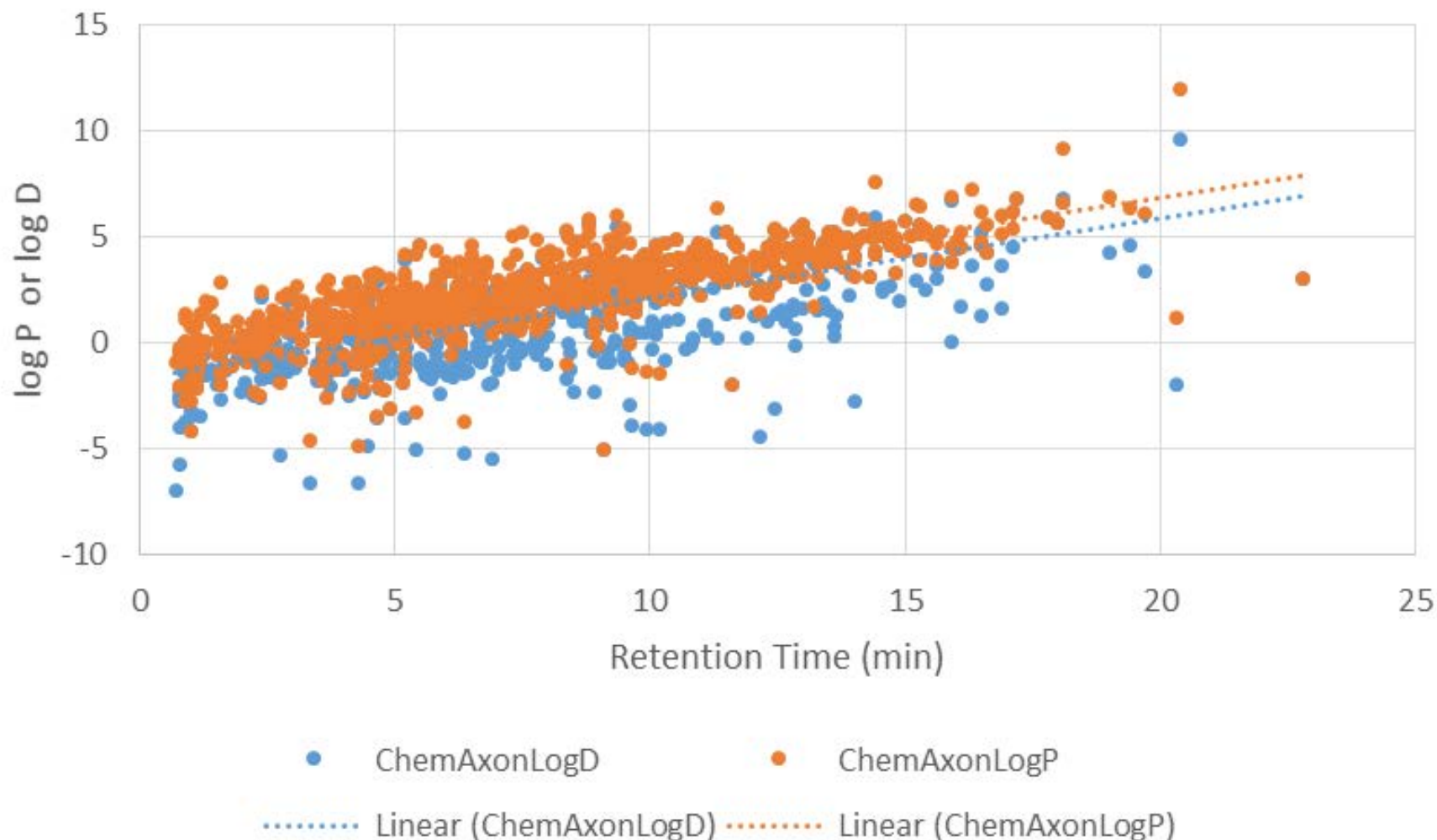
Patents

MS/MS

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |

Gathering Evidence for Identification

Retention time on Liquid Chromatography C₁₈ Column



MetFrag: *In silico* non-target identification

Status: 2016



Literature



Patents

m/z $[M-H]^-$
213.9637
 ± 5 ppm

PubChem

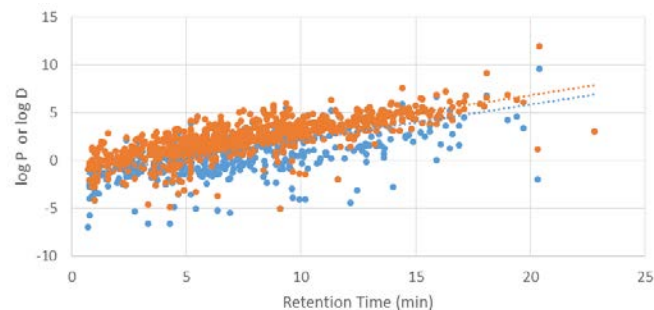
5 ppm
0.001 Da

MetFrag

RT: 4.58 min
355 InChI/RTs

MS/MS

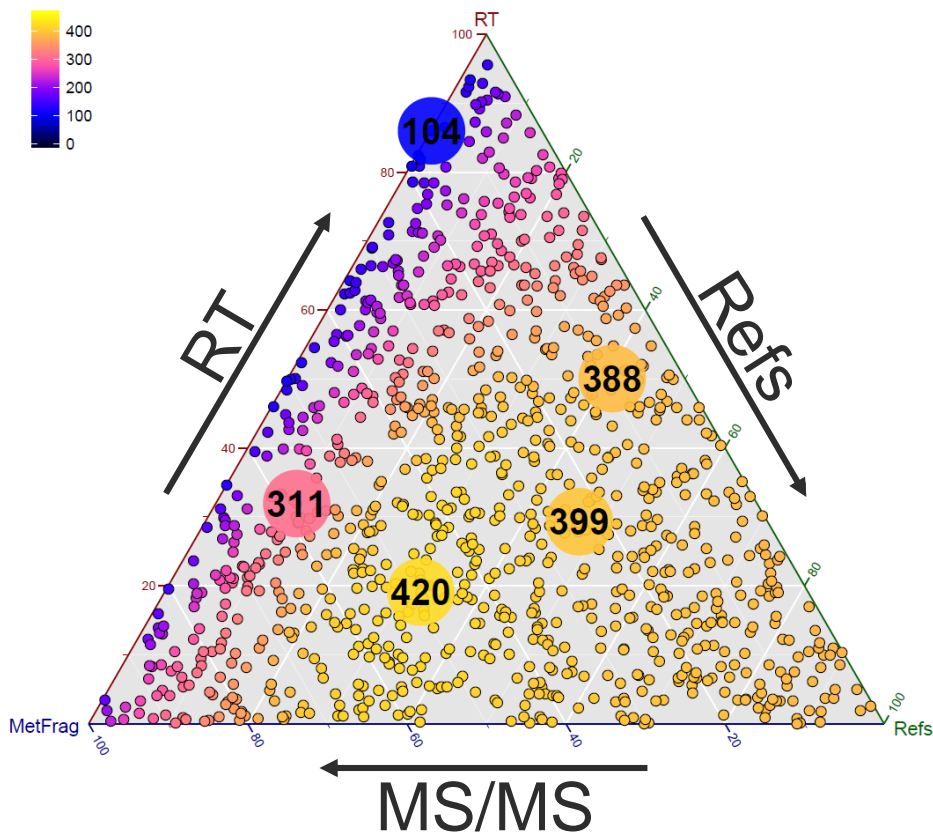
| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |



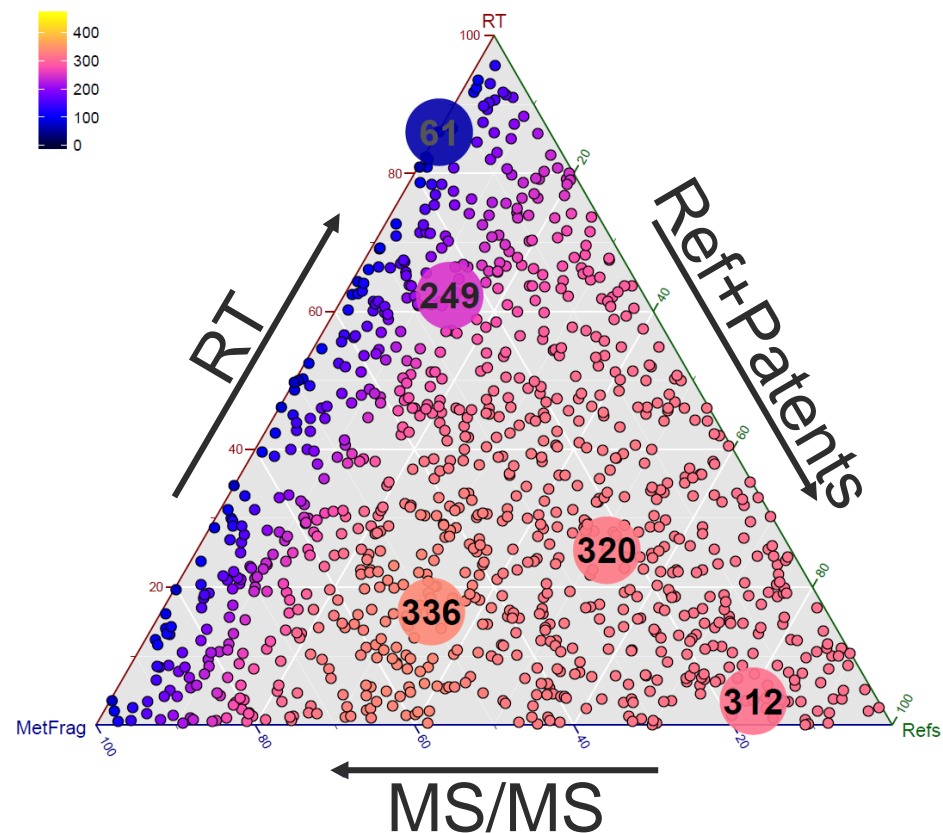
Contribution of References, RT and Fragments

ChemSpider
Search and share chemistry

PubChem | OPEN
CHEMISTRY
DATABASE



| MSMS | RT | Refs |
|------|------|------|
| 0.49 | 0.19 | 0.32 |



| MSMS | RT | Ref+Pats |
|------|------|----------|
| 0.50 | 0.16 | 0.34 |

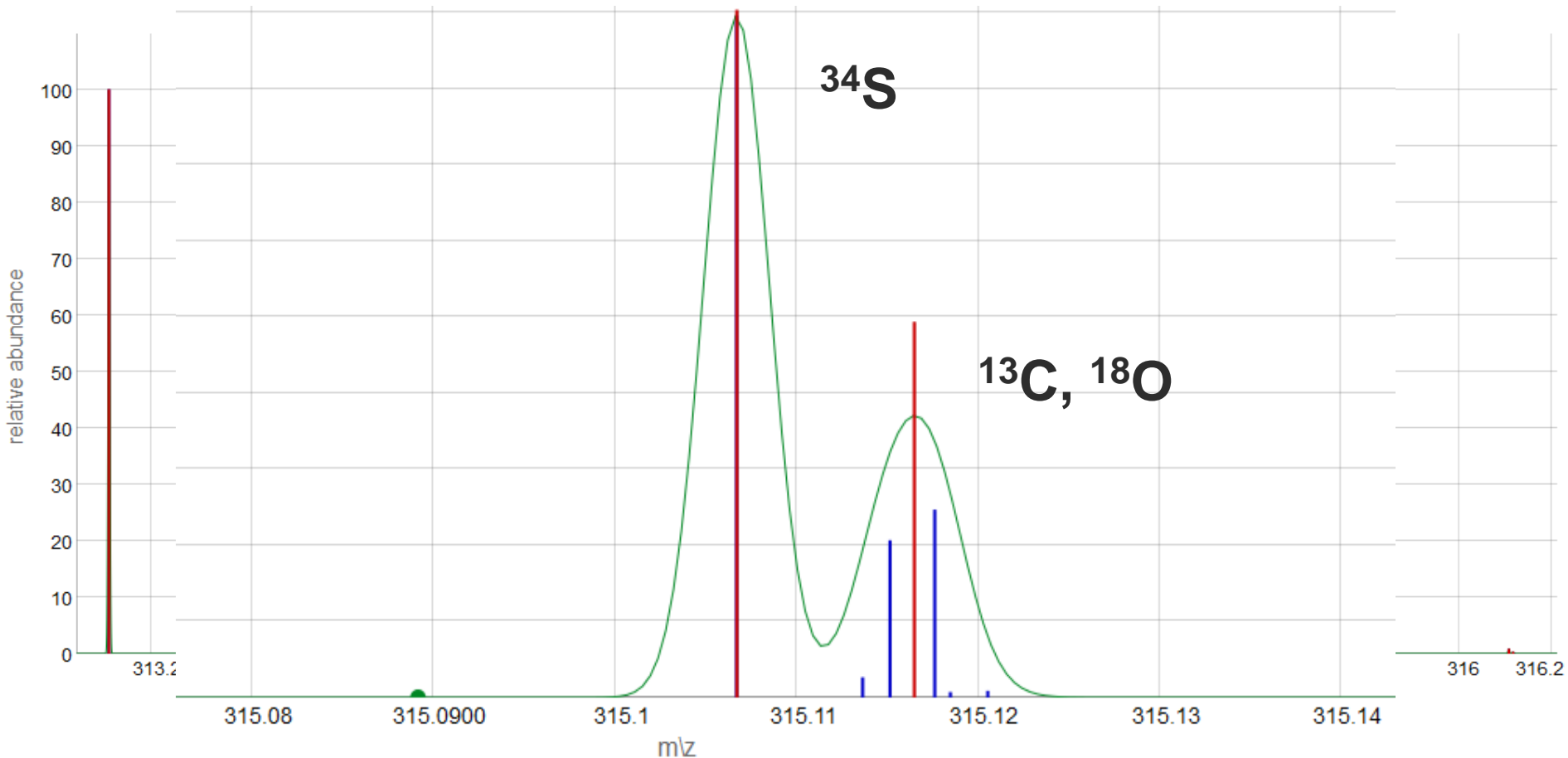
Test set of 473 Eawag Target Substances

| | MetFrag 2010 | New MetFrag Fragments only | New MetFrag +References +Retention time |
|-------------------------------|-----------------|----------------------------------|--|
| ChemSpider¹ | | | |
| Top 1 Ranks | 73 | 105 | 420 |
| % Top 1 Ranks | 15 % | 22 % | 89 % |
| PubChem² | | | |
| Top 1 Ranks | - | 30 | 336 |
| % Top 1 Ranks | - | 6 % | 71 % |

Similar results with 3 independent datasets of 310, 289 and 225 substances from Eawag and UFZ (www.massbank.eu)

¹www.chemspider.com; ~34 million entries

²<https://pubchem.ncbi.nlm.nih.gov/>; ~74 million entries



M. Loos, *et al.* 2015, DOI: 10.1021/acs.analchem.5b00941

<http://www.envipat.eawag.ch/>

<http://cran.r-project.org/web/packages/enviPat/>

Image © www.seanoakley.com/

MetFrag: *In silico* non-target identification

Status: 2016



Literature



Patents

m/z [M-H]⁻

213.9637

± 5 ppm

PubChem

Elements: C, N, S

5 ppm

0.001 Da

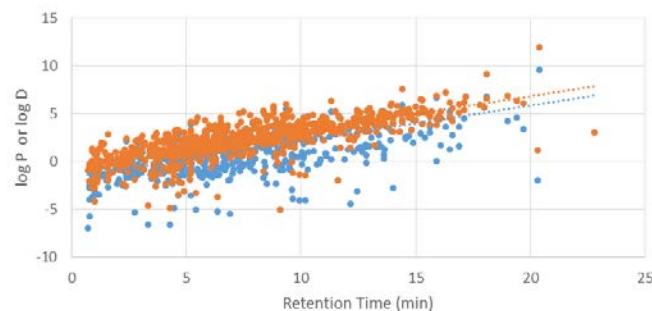
MetFrag

RT: 4.58 min

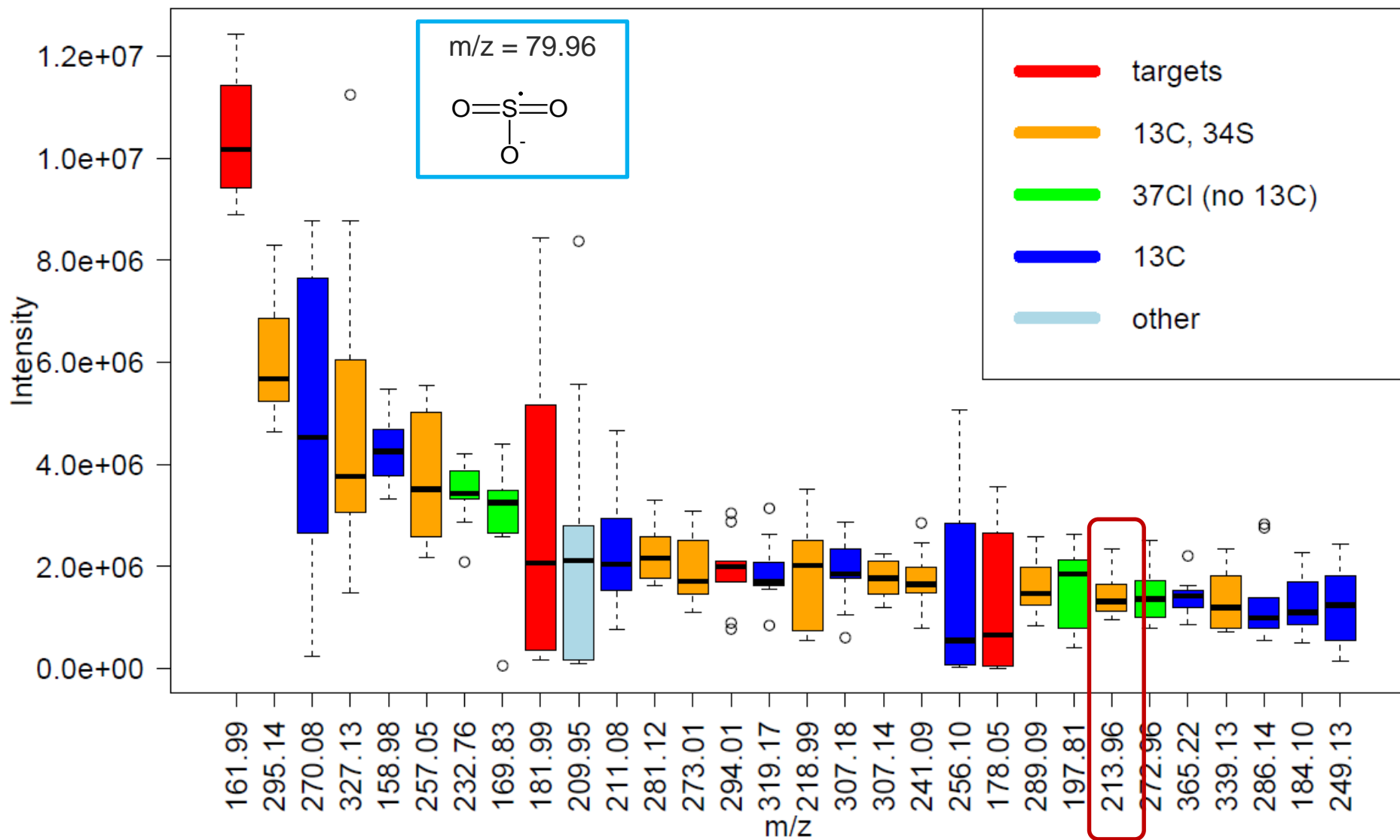
355 InChI/RTs

MS/MS

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |



Targets, Non-targets and Isotopes (ESI-)



Picture: www.momsteam.com

MetFrag: *In silico* non-target identification

Status: 2016



Literature



Patents

m/z $[M-H]^-$

213.9637

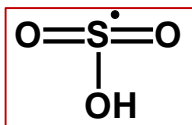
± 5 ppm

PubChem

Elements: C, N, S

5 ppm

0.001 Da



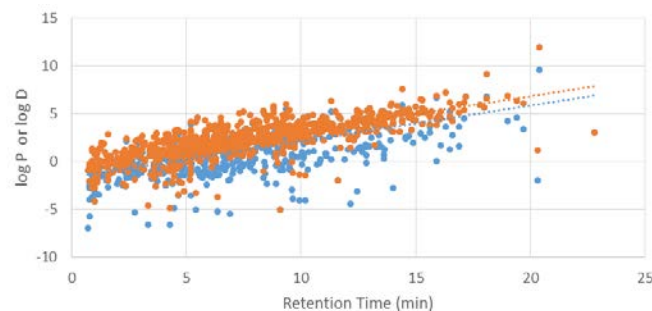
RT: 4.58 min

355 InChI/RTs

MetFrag

MS/MS

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |



Suspect Screening – [NORMAN SLE](#)



[NORMAN WEBSITE](#) |
 [NORMAN DATABASE SYSTEM](#) |
 [HOME](#) |
 [LOGIN](#)

NORMAN SUBSTANCE DATABASE

NORMAN Suspect List Exchange – NORMAN SLE

<https://www.norman-network.com/nds/SLE/>

| No. | Abbreviation | Description | Link to full list | Link to InChIKey list | References |
|-----|--------------|---|--|---|--|
| S0 | SUSDAT | Merged NORMAN Suspect List: SusDat | Interactive Data table (updating...) CompTox SUSDAT List | MS-ready InChIKeys (1/03/2018) | A merged list of >40,000 structures from suspect lists. See interactive version . Compiled by Reza Aalizadeh, University of Athens, including RTI and toxicity values, support by Nikiforos Alygizakis, EI. <i>Work in progress ... please report any issues!</i> DOI: 10.5281/zenodo.2664077 |
| S1 | MASSBANK | NORMAN Compounds in MassBank | CSV, XLSX with Fragments (3/10/2017) CompTox MassBank EU Reference List CompTox MassBank EU Special Cases CompTox Fragment Download | MassBankEUInChIKeys (17/06/2019) | www.massbank.eu Stravs <i>et al.</i> 2013. DOI: 10.1002/jms.3131 DOI: 10.5281/zenodo.2621390 |
| S2 | STOFFIDENT | HSWT/LfU STOFF-IDENT Database of Water-Relevant Substances | STOFF-IDENT Contents (6/09/2017) CompTox STOFF-IDENT List | STOFF-IDENT InChIKeys (6/09/2017) | The database enables the search for exact masses from target or unknown lists and the automatic use of a Retention Time Index. See: https://www.lfu.bayern.de/stoffident/#!home (single search for free; batch search after free registration). DOI: 10.5281/zenodo.2621451 |
| S3 | NORMANCT15 | NORMAN Collaborative Trial Targets and Suspects | LC-MS: CSV, XLSX (3/10/2017) GC-MS: CSV, XLSX (3/10/2017) CompTox NORMANCT15 List | LC-MS InChIKeys (31/10/2016) GC-MS InChIKeys | Schymanski <i>et al.</i> 2015. DOI: 10.1007/s00216-015-8681-7 DOI: 10.5281/zenodo.2621478 |

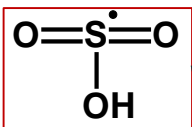
MetFrag: *In silico* non-target identification

Status: 2016

m/z [M-H]⁻
213.9637
± 5 ppm

Elements: C, N, S

5 ppm
0.001 Da



RT: 4.58 min
355 InChI/RTs

PubChem

MetFrag



Literature



Patents

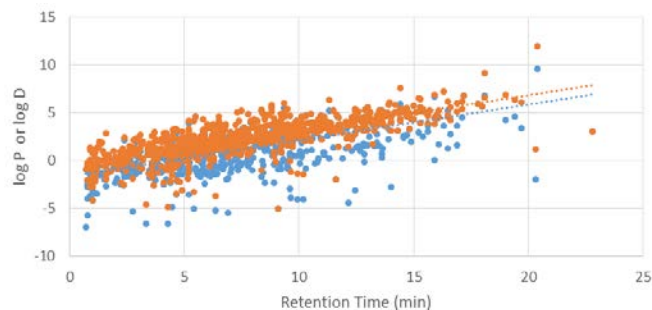
Suspect List(s)
InChIKeys

norman
suspects



MS/MS

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |



Adding in Spectral Libraries via MoNA

Welcome to MoNA Beta 2!

MassBank of America (MoNA) is a metadata-centric, auto-curating repository designed for efficient storage and querying of mass spectral records. It intends to serve as a the framework for a centralized, collaborative database of metabolite mass spectra, metadata and associated compounds. MoNA currently contains over 200,000 mass spectral records from experimental and in-silico libraries as well as from user contributions.

MoNA has recently been redesigned, with significant improvements to server-side architecture, query structure, and search speed. We are actively improving and adding features, so please be patient as functionality is added. If you notice any major issues, feel free to report them using the issue tracker linked below.

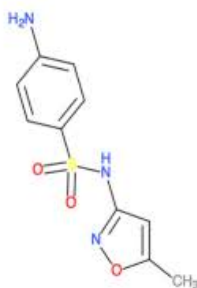
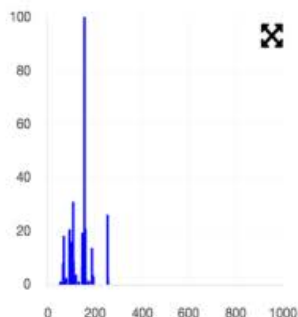
Q Search Spectra

📊 Browse Spectra

🚨 Issue Tracker

4-amino-N-(5-methyl-1,2-oxazol-3-yl)benzenesulfonamide

Score: ★★★★★★☆☆



| | |
|--------------------|---------------------------|
| Q accession | AU101801 |
| Q authors | Nikiforos Alygizakis, Ann |
| Q base peak | 254.0607 |
| Q collision energy | Ramp 19.9-29.9 eV |
| Q column name | Acclaim RSLC C18 2.2um, 2 |
| Q compound class | N/A; Environmental Standa |
| Q copyright | Copyright (C) 2015 Depart |
| Q data processing | REANALYZE Peaks with addi |
| Q data processing | WHOLE RMassBank 1.8.1 |
| Q data processing | RECALIBRATE identity on a |
| Q origin | AU101801.txt |
| Q retention time | 4.7 min |

LCMS ▾

massbank ▾

🔗 display full information

MetFrag: *In silico* non-target identification

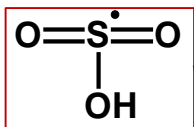
Status: 2016

m/z [M-H]⁻
213.9637
± 5 ppm

Elements: C, N, S

5 ppm

0.001 Da



RT: 4.58 min

355 InChI/RTs

PubChem



Literature



Patents

Suspect List(s)
InChIKeys

norman
suspects



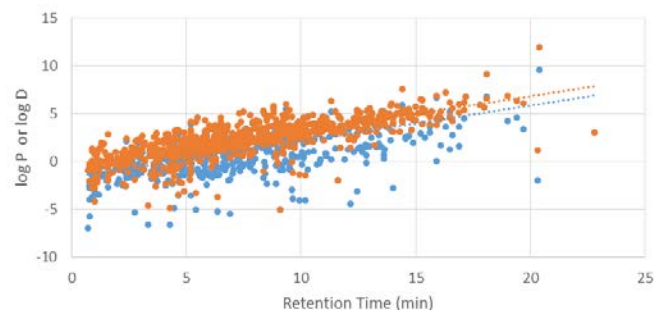
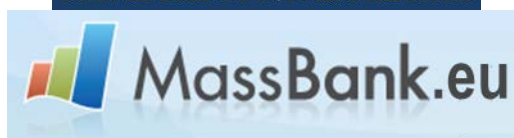
MetFrag

MS/MS

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |

MoNA

MassBank of North America



Example: <https://msbi.ipb-halle.de/MetFrag/>

Database Settings

Database: PubChem

Include references: ☒

Parent Ion:

213.9637

[M-H]-

Calculate

Neutral Mass: 214.97098 Search ppm: 5

Formula:

Identifiers:

Retrieve Candidates

 135 Candidates

Download Candidates

Candidate Filter & Score Settings

Fragmentation Settings & Processing

Mzppm: 5

Mzabs: 0.001

Mode: [M-H]-

Tree depth: 2

Group candidates ☒

Process Candidates

MS/MS Peak list

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |

Show Spectrum

Download Parameters

Example with MetFrag

m/z [M-H]⁻ 213.9637 [non-target]

m/z [M-H]⁻

213.9637

± 5 ppm

5 ppm

0.001 Da

PubChem



MS / MS

134.0054 339689.4

150.0001 77271.2

213.9607 632466.8

Database Settings

Database:

PubChem

Include references: ☒

Parent Ion:

213.9637

[M-H]⁻

Calculate

Neutral Mass:

214.97098

Search ppm: 5

Formula:

Identifiers:

Retrieve Candidates

 135 Candidates

Download Candidates

Example with MetFrag

m/z [M-H]⁻ 213.9637 [non-target]

Molecular

Formula:

C₇H₅NO₃S₂

5 ppm

0.001 Da

PubChem



MS / MS

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |

Database Settings

Database:

PubChem

Include references: ☒

Parent Ion:

213.9637

[M-H]⁻

Calculate

Neutral Mass:

214.97109

Search ppm: 5

Formula:

C7H5NO3S2

Identifiers:

Retrieve Candidates

 45 Candidates

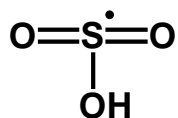
Download Candidates

Example with MetFrag

m/z [M-H]⁻ 213.9637 [non-target]

Elements: C, N, S

mz [M-H]⁻
213.9637
± 5 ppm



5 ppm
0.001 Da

RT: 4.58 min
355 InChI/RTs

PubChem

MoNA
MassBank of North America

MetFrag



Literature



Patents

MS/MS

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |

Suspect List(s)

InChIKeys

Remaining information is added in
“Candidate Filter & Score Settings”

Adding Elements

Database Settings

Database: Include references: ☒

Neutral Mass: Search ppm:

Formula:

Identifiers:

Candidate Filter & Score Settings

Candidate Filters

☒ Element Inclusion

Element List:

Filter Type

☒ Element Exclusion

Element List:

Adding Substructures

- Careful: there is one “pre-selected” if you activate this



Substructure Inclusion

Pre-defined Substructures ▼

1 Substructure selected



Carboxyl



Cyanamide



Fused benzene rings



Hydroxyl



Ketone



S



Thiol



Unfused benzene ring



Substructure Information

1cccc

Adding Substructures

- Deselect the active one, add in “S(=O)(=O)O” instead



Substructure Inclusion

Pre-defined Substructures



0 Substructures selected

Additional Substructures:

(Find out more about [SMARTS](#))

- NOTE: you can include substructures as “filters” or “scoring terms”

Adding Substructures

- NOTE: you can include substructures as “filters” or “scoring terms”

Candidate Filters

☐ Element Inclusion

☐ Element Exclusion

☒ Substructure Inclusion

Pre-defined Substructures ▼ 0 Substructures selected

Additional Substructures:

(Find out more about [SMARTS](#))

☐ Substructure Exclusion

MetFrag Scoring Terms

☒ Substructure Inclusion

Pre-defined Substructures ▼ 1 Substructure selected

Additional Substructures:

(Find out more about [SMARTS](#))

☐ Substructure Exclusion

☐ Retention Time

☐ Suspect Inclusion Lists

Adding Substructures

- You can always check your SMARTS on CDK Depict <https://www.simolecule.com/cdkdepict/depict.html>

```
n1c2ccccc2sc1(S(=O)(O)=O)
CN1C=NC2=C1C(=O)N(C(=O)N2C)C caffeine
[Cs+].[O-]C(=O)[O-].[Cs+] Cs2CO3
[Li+].[Al+3].[H-].[H-].[H-].[H-] LiAlH4
Cl[Pt@SP1](Cl)([NH3])[NH3] cis-platin
O=N[Co@]([NH3])([NH3])([NH3])([NH3])N(=O) trans-[Co(NH3)4(NO)2]
Cl*.Cl*.c1ccccc1-c1ccccc1 |m:1:4.5.6.7.8.9,3:10.11.12.13.14.15| dichlorobiphenyl
CCOCCOCCO |Sg:n:3.4.5::ht| PEGn
```

Color on White ▾

No Annotation ▾

Chiral Hydrogens (smart) ▾

Do Not Abbreviate ▾

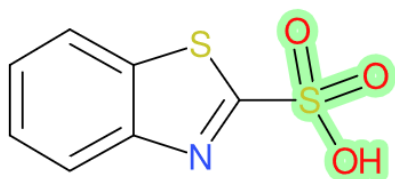
Zoom



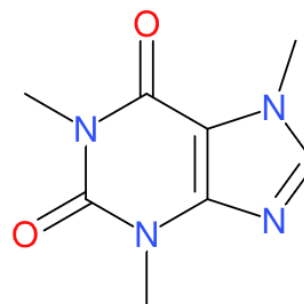
Show Title ☐

Highlight

S(=O)(=O)O



#1



caffeine

Influence of Elements, Substructure on Candidates

Neutral Mass:

Search ppm:

Formula:

Identifiers:

Retrieve Candidates

 135 Candidates



Processing finished

17 Candidates processed 118
Candidates filtered out

▼ Candidate Filter & Score Settings

Candidate Filters



Element Inclusion

Element List:

Filter Type

Exclusive

Optional



Element Exclusion

Element List:



Substructure Inclusion

Pre-defined Substructures

▼ 0 Substructures selected

Additional Substructures:

▼ Fragmentation Settings & Processing

Mzppm:

Mzabs:

Mode:

[M-H]-

Tree depth:

2

Group candidates



Process Candidates

 17 Candidates processed

MS/MS Peak list

| | |
|----------|----------|
| 134.0054 | 339689.4 |
| 150.0001 | 77271.2 |
| 213.9607 | 632466.8 |

Show Spectrum

Download Parameters

Adding Scoring Terms

○ Retention time

MetFrag Scoring Terms


☐ Substructure Inclusion

☐ Substructure Exclusion

☒ Retention Time

Upload file:

+ Choose

 Uploaded Eawag_rt_inchi.csv

Experimental RT
(min):

4.58

Partitioning
Coefficient:

CDK

Adding Scoring Terms

○ Suspect Inclusion Lists



Suspect Inclusion Lists

+ Choose

| Uploaded suspect lists | | |
|---------------------------------------|----------------|-------------------|
| Suspect List Name | Number Entries | |
| SusDat_MSReady_InChIKeys_01032018.txt | 40052 | <div>Remove</div> |

Number uploaded suspect lists: 1

Predefined Suspect Lists:

- ☐ FOR-IDENT (Find out more about [ForIdent](#))
- ☐ DSSTox (Find out more about [DSSTox](#))

Adding Scoring Terms

- Exact Spectral Similarity, Database Terms

☐ Spectral Similarity (MoNA)

☒ Exact Spectral Similarity (MoNA)

☐ Statistical Scoring

Database Scoring Terms

Select Item(s) ▼

2 of 2 item(s) selected



PubChemNumberPatents

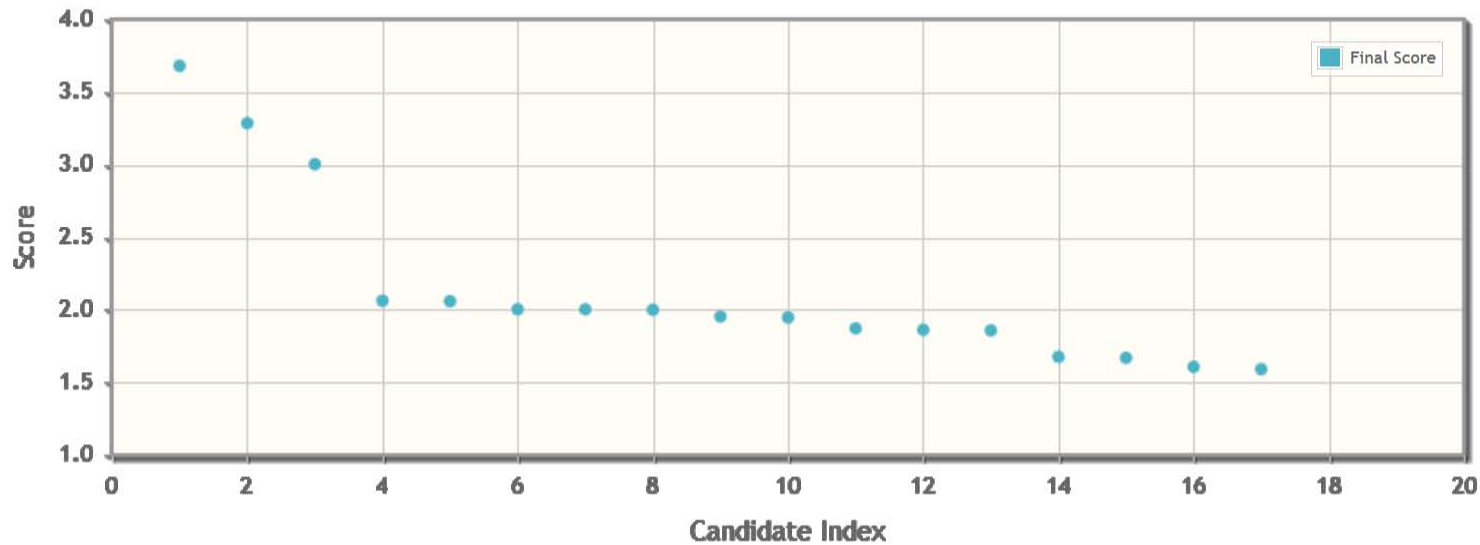


PubChemNumberPubMedReferences

Statistics Plot: Overview of results

Statistics

Candidate Score Distribution



Show Labels

Display Score Graphs

Scores

Export

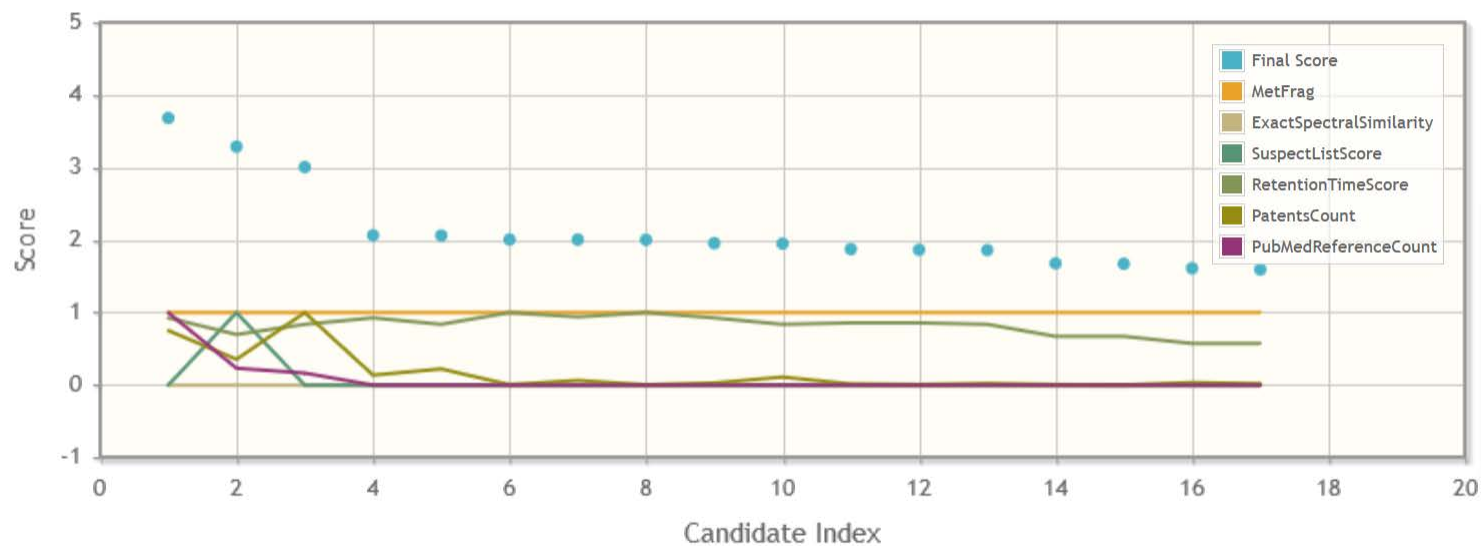
Select area to zoom in. Double click to return.

Click on dot to scroll to candidate in the Results tab.

Statistics Plot: Overview of results

Statistics

Candidate Score Distribution



Export

Select area to zoom in. Double click to return.
Click on dot to scroll to candidate in the Results tab.

Show Labels

Display Score Graphs

Scores

- ☒ MetFrag
- ☒ ExactSpectralSimilarity
- ☒ SuspectListScore
- ☒ RetentionTimeScore
- ☒ PatentsCount
- ☒ PubMedReferenceCount

Results Overview

Results

Weights

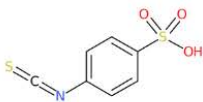
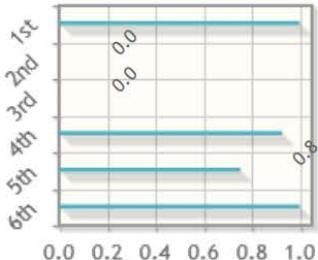
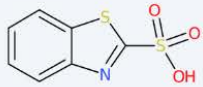
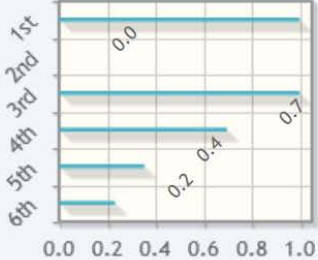
| | | | |
|-------------------------------|-----------------------|-----|---|
| MetFrag (1st) | <input type="range"/> | 100 | % |
| ExactSpectralSimilarity (2nd) | <input type="range"/> | 100 | % |
| SuspectListScore (3rd) | <input type="range"/> | 100 | % |
| RetentionTimeScore (4th) | <input type="range"/> | 100 | % |
| PatentsCount (5th) | <input type="range"/> | 100 | % |
| PubMedReferenceCount (6th) | <input type="range"/> | 100 | % |

Download Results

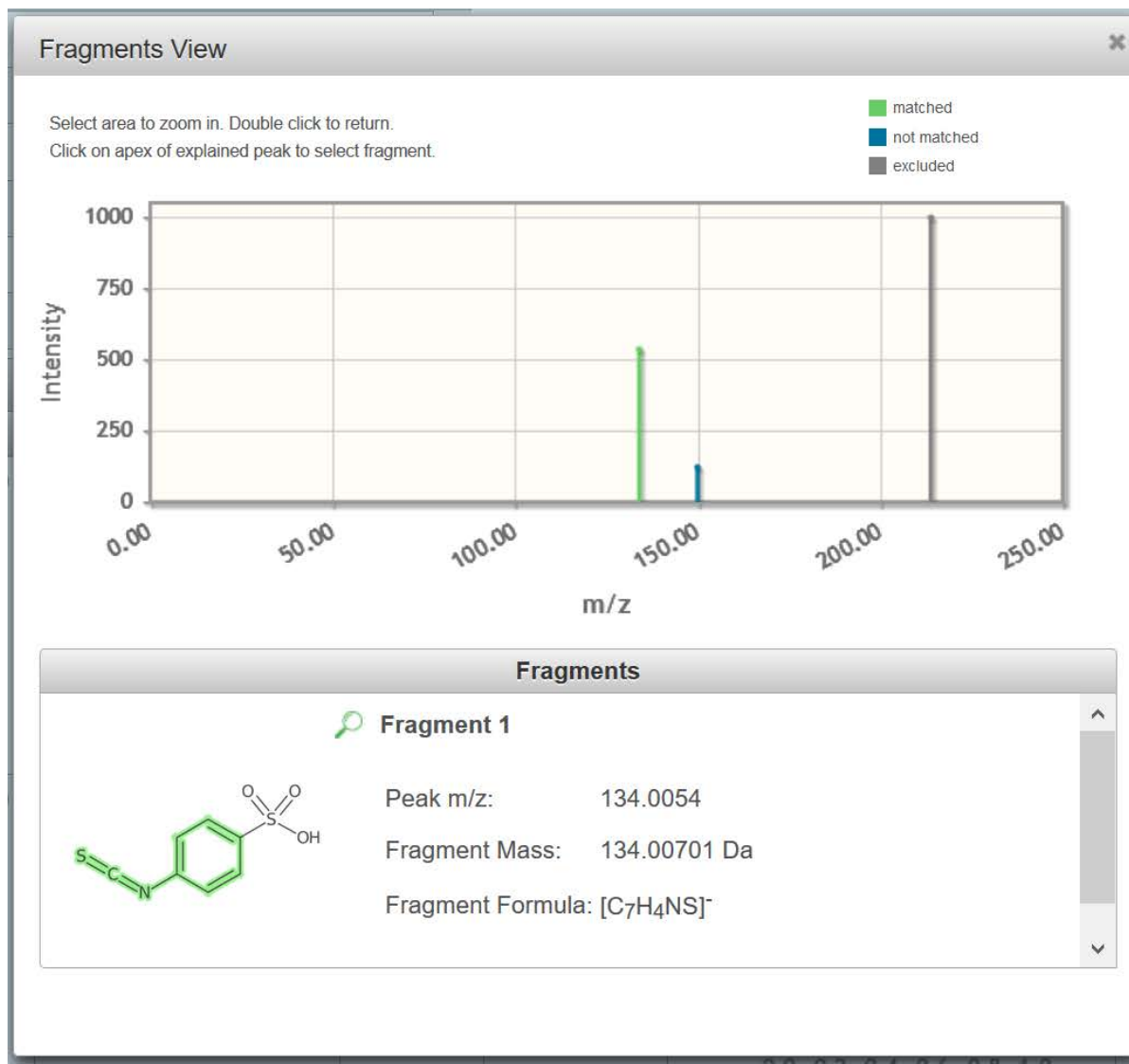
Filter Candidates by explained MS/MS Peaks

MS/MS Peaks

Filter Candidates

| # | Molecule | Identifier | Mass | Formula | Normalized Scores | FinalScore | Details |
|---|---|--|---------|--|---|------------|---|
| 1 |  4-isothiocyanatobenzenesulfonic acid | 87189 InChIKeyBlock1 = <u>LCYBJJOUISDNRJ</u> | 214.971 | C ₇ H ₅ NO ₃ S ₂ |  | 3.6804 | Peaks: 1 / 2 Fragments Scores Download |
| 2 |  1,3-benzothiazole-2-sulfonic acid | 30647 InChIKeyBlock1 = <u>ZCXGMSGCBDSEYO</u> | 214.971 | C ₇ H ₅ NO ₃ S ₂ |  | 3.2868 | Peaks: 1 / 2 Fragments Scores Download |

Fragments View



MS/MS Peaks

ter Candidates

FinalScore

Details

Peaks: 1 / 2

Fragments

Scores

Download

3.6804

Peaks: 1 / 2

Fragments

Scores

Download

3.2868

Scores View

Scores View

Candidate Name: 1,3-benzothiazole-2-sulfonic acid
Candidate Identifier: 30647

| | Name | Normalized Value | Raw Value |
|---|-------------------------|------------------|-----------|
| ▶ | MetFrag | 1.0 | 122.4115 |
| ▶ | ExactSpectralSimilarity | 0.0 | 0.0 |
| ▶ | SuspectListScore | 1.0 | 1.0 |
| ▶ | RetentionTimeScore | 0.6982 | 0.1844 |
| ▶ | PatentsCount | 0.3553 | 108.0 |
| ▶ | PubMedReferenceCount | 0.2333 | 7.0 |

Candidates

Score

Details

Scores

Download

Peaks: 1 / 2

Fragments

Scores

Download

30647

InChIKeyBlock1 =
ZCXGMSGCBDSEYOY

214.971

C7H5NO3S2

| Fragmentation Level | Relative Intensity |
|---------------------|--------------------|
| 2nd | 1.0 |
| 3rd | 0.7 |
| 4th | 0.4 |
| 5th | 0.2 |
| 6th | 0.1 |

3.2868

Download Results

Results

Weights

MetFrag (1st)

100

%

ExactSpectralSimilarity (2nd)

100

%

SuspectListScore (3rd)

100

%

RetentionTimeScore (4th)

100

%

PatentsCount (5th)

100

%

PubMedReferenceCount (6th)

100

%

Download Results

Filter Candidates by explained MS/MS Peaks

MS/MS Peaks

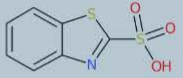
Filter Candidates

Select format

CSV

XLS

SDF

| # | Molecule | Identifier | Normalized Scores | FinalScore | Details |
|---|--|--|---|------------|--|
| 1 | 4-isothiocyanatobenzenesulfonic acid | InChIKeyBlock1 = <u>LBYBJJOUISDNRJ</u> | | 3.6804 | <div>Scores</div> <div>Download</div> |
| 2 |  1,3-benzothiazole-2-sulfonic acid | 30647 InChIKeyBlock1 = <u>ZCXGMSGCBDSEYO</u> | 214.971 C ₇ H ₅ NO ₃ S ₂ | 3.2868 | Peaks: 1 / 2 <div>Fragments</div> <div>Scores</div> <div>Download</div> |

Download Results

MetFragWeb_Candidates.xls [Protected View] - Excel

File Home Insert Page Layout Formulas Data Review View ACROBAT Tell me what you want to do...

PROTECTED VIEW Be careful—files from the Internet can contain viruses. Unless you need to edit, it's safer to stay in Protected View. Enable Editing

| A1 | Score | | | | | | | | | | | | | | | | | | | |
|----|--------|-----------|----------|-----------|-----------|----------|----------|-----------|-----------|------------|-----------|-----------------|-----------|-----------|---------|-----------|-----------|----------|-------------|-------|
| | Score | Monoisotc | SMILES | MetFrag | Retention | NumberPe | NoExplPe | InChI | ExactSpec | Identifier | ExplPeaks | Compound | PatentsCo | Molecular | PubMedR | Retention | MetFrag_V | Formulas | SuspectList | Score |
| 1 | 3.6804 | 214.971 | C1=CC(=C | 122.41154 | 1.44226 | 2.2 | 1 | InChI=1S/ | 0.0 | 87189 | 134.0054 | 4-isothiocy | 229.0 | C7H5NO3 | 30.0 | 0.2448048 | 272.0 | 134.0054 | [0.0 | |
| 2 | 3.2868 | 214.971 | C1=CC=C | 122.41154 | 1.44226 | 0.2 | 1 | InChI=1S/ | 0.0 | 30647 | 134.0054 | 1,3-benzot | 108.0 | C7H5NO3 | 7.0 | 0.1843501 | 272.0 | 134.0054 | [1.0 | |
| 3 | 3.0049 | 214.971 | C1=CC2=C | 122.41154 | 1.44226 | 0.2 | 1 | InChI=1S/ | 0.0 | 14397824 | 134.0054 | 1,3-benzot | 304.0 | C7H5NO3 | 5.0 | 0.2213244 | 272.0 | 134.0054 | [0.0 | |
| 4 | 2.0653 | 214.971 | C1=CC=C | 122.41154 | 1.44226 | 2.2 | 1 | InChI=1S/ | 0.0 | 22639215 | 134.0054 | 2-isothiocy | 42.0 | C7H5NO3 | 0.0 | 0.2448048 | 272.0 | 134.0054 | [0.0 | |
| 5 | 2.0619 | 214.971 | C1=CC2=C | 122.41154 | 1.44226 | 0.2 | 1 | InChI=1S/ | 0.0 | 21989207 | 134.0054 | 1,3-benzot | 68.0 | C7H5NO3 | 0.0 | 0.2213244 | 272.0 | 134.0054 | [0.0 | |
| 6 | 2.0066 | 214.971 | C1=CC2=C | 122.41154 | 1.44226 | 1.2 | 1 | InChI=1S/ | 0.0 | 89331297 | 134.0054 | 2,1-benzot | 2.0 | C7H5NO3 | 0.0 | 0.2640480 | 272.0 | 134.0054 | [0.0 | |
| 7 | 2.0063 | 214.971 | C1=CC=C | 122.41154 | 1.44226 | 0.2 | 1 | InChI=1S/ | 0.0 | 90883454 | 134.0054 | 1,2-benzot | 20.0 | C7H5NO3 | 0.0 | 0.2483497 | 272.0 | 134.0054 | [0.0 | |
| 8 | 2.0033 | 214.971 | C1=CC2=C | 122.41154 | 1.44226 | 1.2 | 1 | InChI=1S/ | 0.0 | 123227988 | 134.0054 | 2,1-benzot | 1.0 | C7H5NO3 | 0.0 | 0.2640480 | 272.0 | 134.0054 | [0.0 | |
| 9 | 1.9567 | 214.971 | C1=CC(=C | 122.41154 | 1.44226 | 2.2 | 1 | InChI=1S/ | 0.0 | 13954965 | 134.0054 | 3-isothiocy | 9.0 | C7H5NO3 | 0.0 | 0.2448048 | 272.0 | 134.0054 | [0.0 | |
| 10 | 1.95 | 214.971 | C1=CC2=C | 122.41154 | 1.44226 | 0.2 | 1 | InChI=1S/ | 0.0 | 20161840 | 134.0054 | 1,3-benzot | 34.0 | C7H5NO3 | 0.0 | 0.2213244 | 272.0 | 134.0054 | [0.0 | |
| 11 | 1.8749 | 214.971 | C1=CC=C | 122.41154 | 1.44226 | 0.2 | 1 | InChI=1S/ | 0.0 | 87765817 | 134.0054 | 2-thiocyan | 5.0 | C7H5NO3 | 0.0 | 0.2266606 | 272.0 | 134.0054 | [0.0 | |
| 12 | 1.865 | 214.971 | C1=CC(=C | 122.41154 | 1.44226 | 0.2 | 1 | InChI=1S/ | 0.0 | 86135707 | 134.0054 | 4-thiocyan | 2.0 | C7H5NO3 | 0.0 | 0.2266606 | 272.0 | 134.0054 | [0.0 | |
| 13 | 1.8612 | 214.971 | C1=CC2=C | 122.41154 | 1.44226 | 0.2 | 1 | InChI=1S/ | 0.0 | 21424572 | 134.0054 | 1,3-benzot | 7.0 | C7H5NO3 | 0.0 | 0.2213244 | 272.0 | 134.0054 | [0.0 | |
| 14 | 1.6791 | 214.971 | C1=CN=C | 122.41154 | 1.44226 | 0.2 | 1 | InChI=1S/ | 0.0 | 90901670 | 134.0054 | thieno[3,2-2 | 0 | C7H5NO3 | 0.0 | 0.1775756 | 272.0 | 134.0054 | [0.0 | |
| 15 | 1.6725 | 214.971 | C1=CN=C | 122.41154 | 1.44226 | 0.2 | 1 | InChI=1S/ | 0.0 | 21821183 | 134.0054 | thieno[2,3-0 | 0 | C7H5NO3 | 0.0 | 0.1775756 | 272.0 | 134.0054 | [0.0 | |
| 16 | 1.6108 | 214.971 | C1=CC2=C | 122.41154 | 1.44226 | -0.2 | 1 | InChI=1S/ | 0.0 | 18176360 | 134.0054 | thieno[3,2-11.0 | | C7H5NO3 | 0.0 | 0.1517260 | 272.0 | 134.0054 | [0.0 | |
| 17 | 1.5944 | 214.971 | C1=CC2=C | 122.41154 | 1.44226 | -0.2 | 1 | InChI=1S/ | 0.0 | 18176435 | 134.0054 | thieno[2,3-6.0 | | C7H5NO3 | 0.0 | 0.1517260 | 272.0 | 134.0054 | [0.0 | |
| 18 | | | | | | | | | | | | | | | | | | | | |
| 19 | | | | | | | | | | | | | | | | | | | | |
| 20 | | | | | | | | | | | | | | | | | | | | |
| 21 | | | | | | | | | | | | | | | | | | | | |
| 22 | | | | | | | | | | | | | | | | | | | | |
| 23 | | | | | | | | | | | | | | | | | | | | |
| 24 | | | | | | | | | | | | | | | | | | | | |
| 25 | | | | | | | | | | | | | | | | | | | | |
| 26 | | | | | | | | | | | | | | | | | | | | |
| 27 | | | | | | | | | | | | | | | | | | | | |

You can try this entire example soon – see Example 1 in documentation

(Molecule =>) Mass => Molecule => Metadata

- Background

- Molecule to Mass

- Preparation for Mass to Molecule

- Gathering the “evidence”
 - Molecular Formula example

- Structure Elucidation with MetFrag

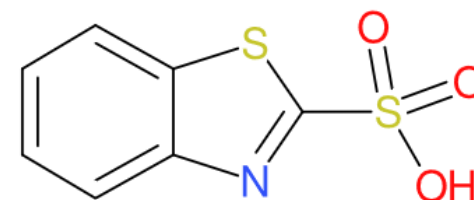
- Step-by-step with one example (MetFrag)

- Practice Session with Several Examples

- [Optional] Context and Perspectives

- Potential for automated non-target workflows

213.9637



Example 2

"MS-ready" Form for MetaData in MetFrag

| # | Molecule | Identifier | Mass | Formula | Normalized Scores | FinalScore | Details |
|---|--|--|-----------|--|-------------------|------------|---|
| 1 | Nicotine | DTXSID1020930 DTXSID8021725 DTXSID3048154 DTXSID0046351 DTXSID6020931 DTXSID00657553 DTXSID5075319 InChIKeyBlock1 = SNICXCGAKADSCV | 162.11576 | C ₁₀ H ₁₄ N ₂ | | 4.3349 | Peaks: 18 / 23 Fragments Scores Download |
| 2 | Phenylpiperazine | DTXSID40176612 DTXSID40193102 DTXSID90216632 DTXSID50291046 DTXSID00293111 DTXSID50296613 InChIKeyBlock1 = YZTJYBJCZXZGCT | 162.11576 | C ₁₂ H ₁₄ N ₂ | | | |
| 3 | N'-(2,4-Dimethylphenyl)-N-methylformamidin | DTXSID1037696 DTXSID10199510 InChIKeyBlock1 = JIIOLEGNERQDIP | 162.11576 | C ₁₂ H ₁₄ N ₂ | | | |

LEGEND: Name, SMILES
DTXSID | InChIKey 1st Block
CAS | MonoIs. Mass | logP | Sources
Data on: Toxicity | Exposure | Bioassays

Nicotine
CN1CCC[C@H]1C1=CN=CC=C1
 DTXSID0046351 | SNICXCGAKADSCV
 25162-00-9 | **162.1157** | 0.929 | **20**
 Tox: **no** | Expo: **yes** | Bioassay: **yes**

D-Nicotine
CN1CCC[C@@H]1C1=CN=CC=C1
 DTXSID0046351 | SNICXCGAKADSCV
 25162-00-9 | **162.1157** | 0.929 | **20**
 Tox: **no** | Expo: **yes** | Bioassay: **yes**


Nicotine hydrochloride
Cl.CN1CCC[C@H]1C1=CN=CC=C1
 DTXSID6020931 | HDJBTCAJIMNXEW
 2820-51-1 | **198.0924** | 0.929 | **9**
 Tox: **no** | Expo: **yes** | Bioassay: **yes**



MS-ready DL-Nicotine
CN1CCCC1C1=CN=CC=C1
 DTXSID3048154 | SNICXCGAKADSCV
 22083-74-5 | **162.1157** | 0.953 | **9**
 Tox: **yes** | Expo: **no** | Bioassay: **yes**




DL-Nicotine-d3
[2H]C([2H])([2H])N1CCCC1C1=CN=CC=C1
 DTXSID80442666 | SNICXCGAKADSCV
 69980-24-1 | **165.1345** | 0.929 | **1**
 Tox: **no** | Expo: **no** | Bioassay: **no**


Benzoic acid, 2-hydroxy-, compd. with 3-[(2S)-1-methyl-2-pyrrolidinyl]pyridine (1:1)
OC(=O)C1=C(O)C=CC=C1.CN1CCC[C@H]1C1=CN=CC=C1
 DTXSID5075319 | AIBWBPBUAKCMKNS
 29790-52-1 | **300.1474** | 0.929 | **6**
 Tox: **no** | Expo: **yes** | Bioassay: **no**





CompTox Dashboard + Nicotine





  <https://comptox.epa.gov/dashboard/>


  


 Search



 **EPA** United States Environmental Protection Agency






875 Thousand Chemicals

Chemicals

Product/Use Categories

Assay/Gene

 Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey

☐ Identifier substring search

See what people are saying, read the dashboard [comments!](#)
Cite the Dashboard Publication [click here](#)

Latest News


[Read more news](#)

Journal of Cheminformatics article regarding "MS-Ready structures"

March 9th, 2019 at 7:09:45 PM

A recent article describes "MS-Ready structures", what they are, how they are generated and details regarding the benefits of these structures in navigating structure relationships across the dashboard. The article is published in the Journal of Cheminformatics [here](#).

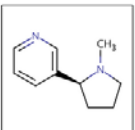
CompTox Dashboard + Nicotine

 United States Environmental Protection Agency

Home Advanced Search Batch Search Lists ▼ Predictions Downloads

Copy ▼ Share ▼ Submit Comment

Search all data



Nicotine

54-11-5 | DTXSID1020930

Searched by Approved Name.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

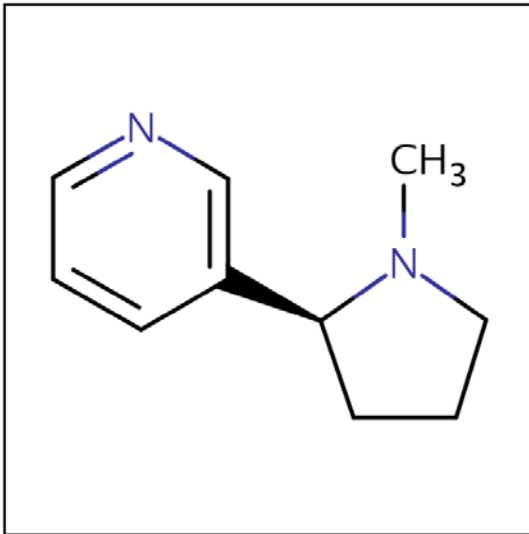
RELATED SUBSTANCES

SYNONYMS

▶ LITERATURE

LINKS

COMMENTS








Wikipedia


Nicotine is a stimulant and potent parasymphomimetic alkaloid that is naturally produced in the nightshade family of plants. It is used for the treatment of tobacco use disorders as a smoking cessation aid and nicotine dependence for the relief of withdrawal symptoms. Nicotine acts as a receptor agonist at most nicotinic acetylcholine receptors (nAChRs), except at two nicotinic receptor subunits (nAChRα9 and nAChRα10) where it acts as a receptor antagonist

...
[Read more](#)

Intrinsic Properties

 **Molecular Formula:** C₁₀H₁₄N₂  Mol File  Find All Chemicals

 **Average Mass:** 162.236 g/mol  Isotope Mass Distribution

 **Monoisotopic Mass:** 162.115698 g/mol

Structural Identifiers

Linked Substances

Presence in Lists

Federal

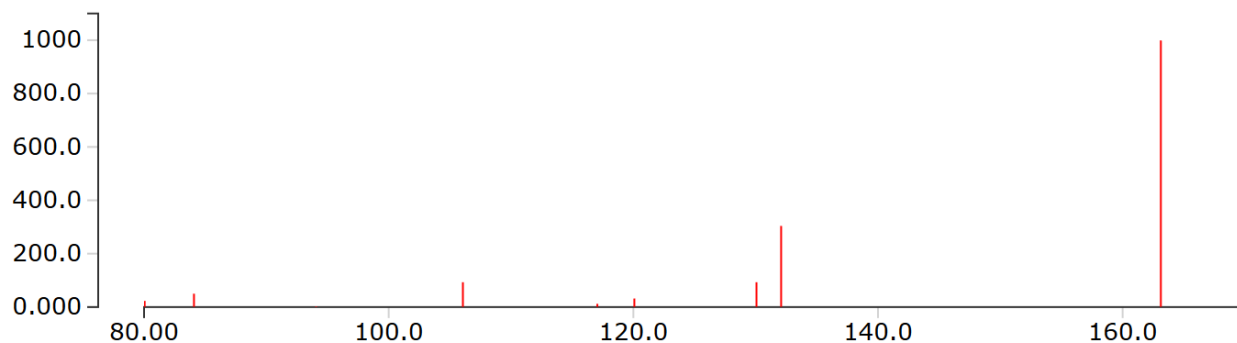
LIST:Hazardous Substances Data Bank

MassBank Record: EQ300801

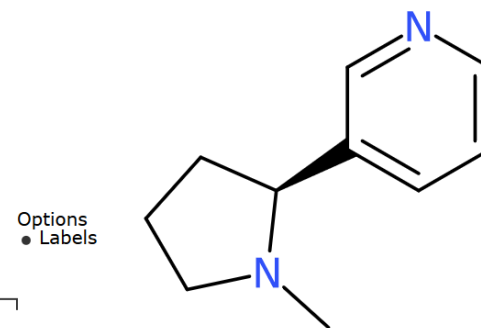
[Home](#) | [Search](#) | [Record Index](#) | [Data Privacy](#) | [Imprint](#) | MassBank ID:

Nicotine; LC-ESI-QFT; MS2; CE: 15; R=35000; [M+H]⁺

Mass Spectrum



Chemical Structure



ACCESSION: EQ300801

RECORD_TITLE:

Nicotine; LC-ESI-QFT; MS2; CE: 15; R=35000; [M+H]⁺

PK\$NUM PEAK: 9

PK\$PEAK: m/z int. rel.int.

| | | |
|----------|-------------|-----|
| 80.0494 | 6261028.7 | 23 |
| 84.0807 | 13197924.1 | 50 |
| 94.065 | 967625.9 | 3 |
| 106.065 | 24640249.3 | 93 |
| 117.0572 | 3192413.5 | 12 |
| 120.0807 | 8648923.7 | 32 |
| 130.0651 | 24669353.9 | 93 |
| 132.0807 | 80112590.7 | 304 |
| 163.1229 | 263120223.6 | 999 |

//

CompTox Dashboard + Nicotine in MetFrag

- CompTox is integrated as a “Local Database”



MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

Database Settings

Database:

CompTox_07March19_Se

Neutral Mass:

Formula:

Identifiers:

Retrieve Candidate

PSV

SDF

Local Databases

CompTox_01May18_AllMetaData

CompTox_01May18_SelectMetaData

CompTox_07March19_SelectMetaData

CompTox_01May18_SelectMetaDataPlu

Candidate Filter & Score Settings

Database Settings

Database:

CompTox_07March19_Se

Neutral Mass:

162.11576

Search ppm: 5

Formula:

C10H14N2

Identifiers:

Retrieve Candidates

225 Candidates

MetaData is included in CompTox

Candidate Filter & Score Settings

Candidate Filters

☐ Element Inclusion

☐ Element Exclusion

☐ Substructure Inclusion

☐ Substructure Exclusion

☐ Substructure Information

☐ Minimum

☐ Maximum

☐ Suspect

MetFrag Scoring Terms

☐ Substructure Inclusion

☐ Substructure Exclusion

☐ Retention Time

☐ Suspect Inclusion Lists

☐ Spectral Similarity (MoNA)

☒ Exact Spectral Similarity (MoNA)

☐ Statistical Scoring

Database Scoring Terms

Select Item(s) 0 of 10 item(s) selected

Select Item(s) 6 of 10 item(s) selected

☒ DATA_SOURCES

☐ ECOTOX

☒ EXPOCAST_MEDIAN_EXPOSURE_PREDICTION_MG/KG-BW/DAY

☐ KEMIMARKET

☐ MZCLOUD

☒ NORMANSUSDAT

☒ NUMBER_OF_PUBMED_ARTICLES

☒ PUBCHEM_DATA_SOURCES

☐ TOX21SL

☒ TOXCAST_PERCENT_ACTIVE

Include MS/MS and mode from MassBank Record

Fragmentation Settings & Processing

Mzppm:

Mzabs:

Mode:



Tree depth:



Group candidates



Process Candidates

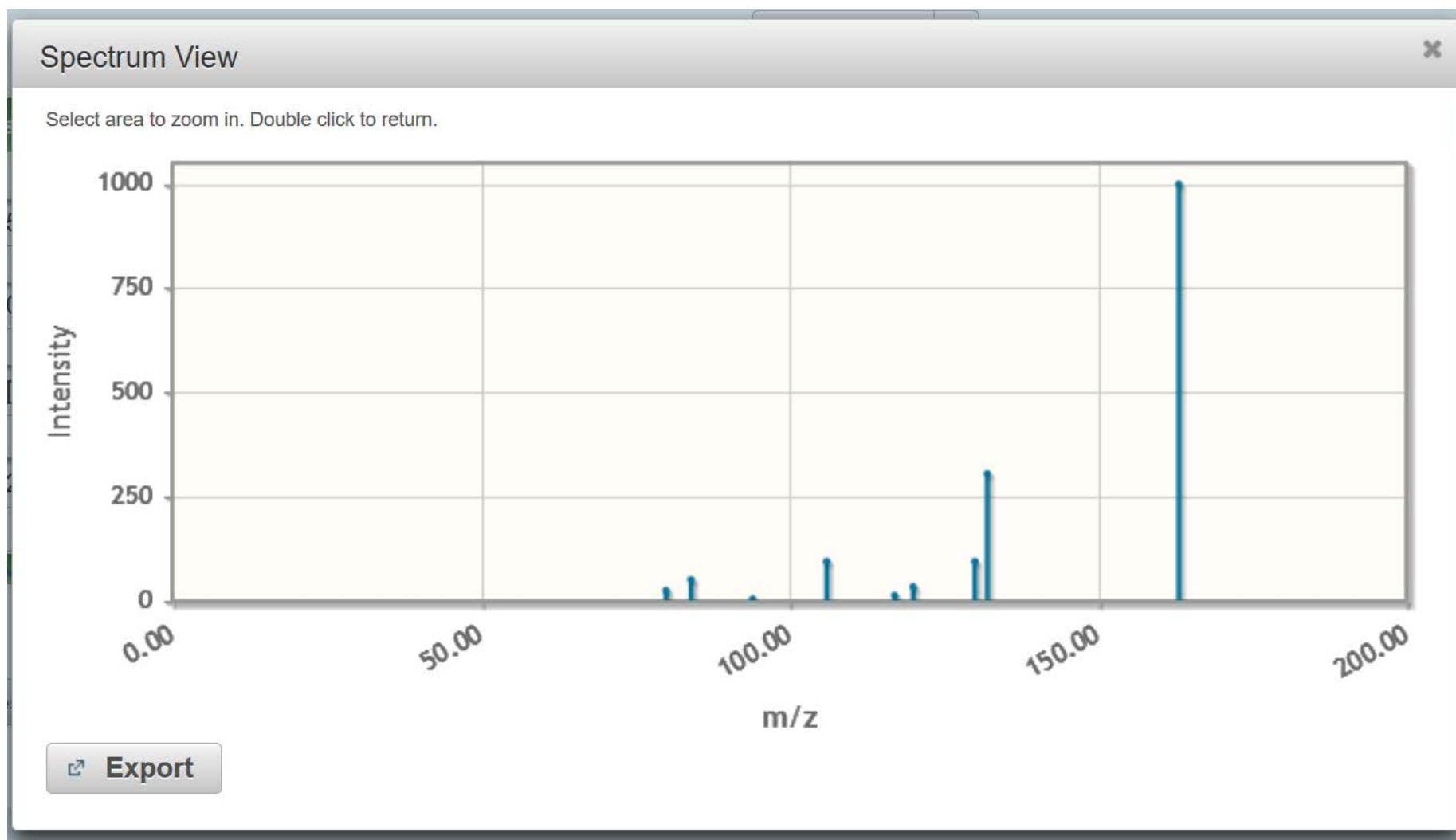
MS/MS Peak list

```
80.0494 6261028.7 23
84.0807 13197924.1 50
94.065 967625.9 3
106.065 24640249.3 93
117.0572 3192413.5 12
120.0807 8648923.7 32
130.0651 24669353.9 93
132.0807 80112590.7
304
163.1229 263120223.6
999
```

Show Spectrum

Download Parameters

Check Spectrum



Process Candidates (Grouped)

Fragmentation Settings & Processing

Mzppm:

Mzabs:

Mode:



Tree depth:



Group candidates



Process Candidates



225 Candidates processed

MS/MS Peak list

```
80.0494 6261028.7 23
84.0807 13197924.1 50
94.065 967625.9 3
106.065 24640249.3 93
117.0572 3192413.5 12
120.0807 8648923.7 32
130.0651 24669353.9 93
132.0807 80112590.7
304
163.1229 263120223.6
999
```

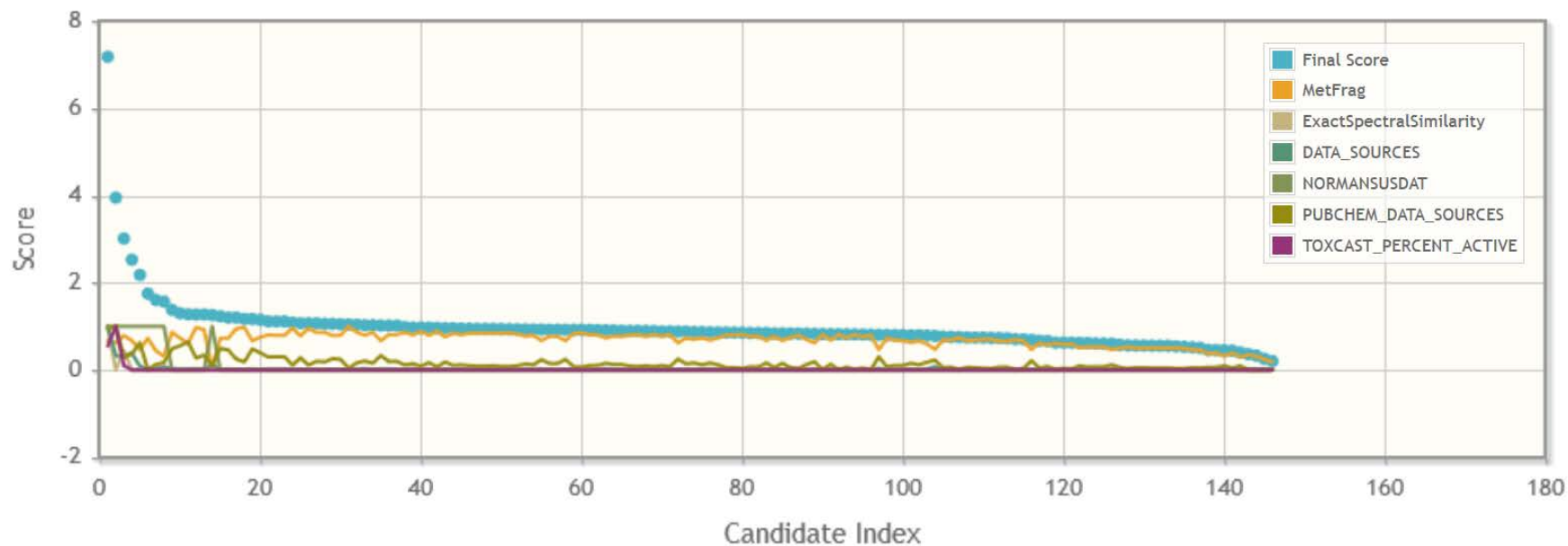
Show Spectrum

Download Parameters

Process Candidates (Grouped)

Statistics

Candidate Score Distribution

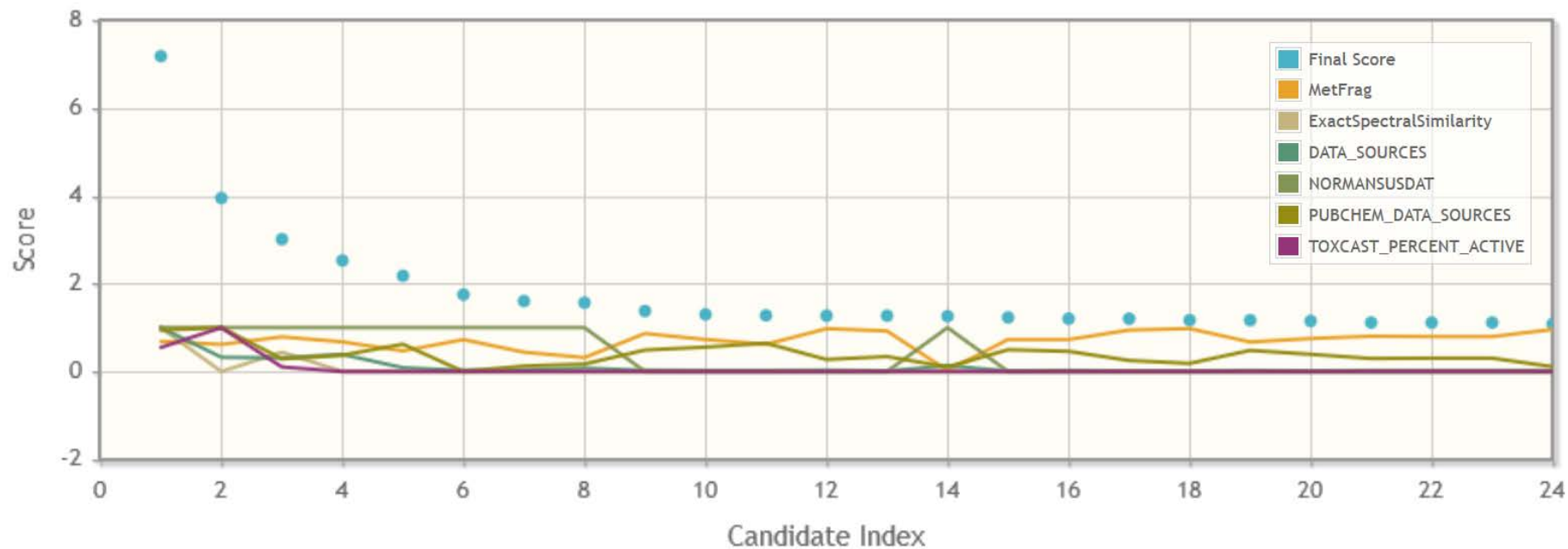


[Export](#)

Select area to zoom in. Double click to return.
Click on dot to scroll to candidate in the Results tab.

Process Candidates (Grouped)

Candidate Score Distribution



[Export](#)

Select area to zoom in. Double click to return.
Click on dot to scroll to candidate in the Results tab.

Process Candidates (Grouped)

Results

Weights

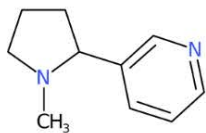
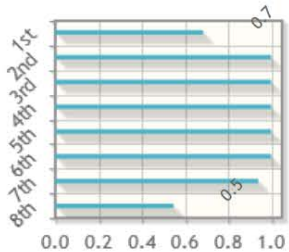
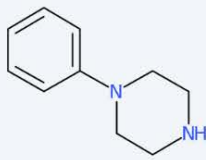
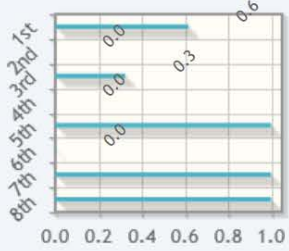
| | | | |
|--|-----------------------|-----|---|
| MetFrag (1st) | <input type="range"/> | 100 | % |
| ExactSpectralSimilarity (2nd) | <input type="range"/> | 100 | % |
| DATA_SOURCES (3rd) | <input type="range"/> | 100 | % |
| EXPOCAST_MEDIAN_EXPOSURE_PR EDICTION_MG/KG-BW/DAY (4th) | <input type="range"/> | 100 | % |
| NORMANSUSDAT (5th) | <input type="range"/> | 100 | % |
| NUMBER_OF_PUBMED_ARTICLES (6th) | <input type="range"/> | 100 | % |

Download Results

Filter Candidates by explained MS/MS Peaks

MS/MS Peaks

Filter Candidates

| # | Molecule | Identifier | Mass | Formula | Normalized Scores | FinalScore | Details |
|---|---|--|-----------|--|---|------------|---|
| 1 |  Nicotine | <u>DTXSID1020930</u> DTXSID8021725 DTXSID3048154 DTXSID20237235 DTXSID0046351 DTXSID6020931 DTXSID80442666 InChIKeyBlock1 = SNICXCGAKADSCV | 162.11576 | C ₁₀ H ₁₄ N ₂ |  | 7.1739 | Peaks: 6 / 8 Fragments Scores Download |
| 2 |  Phenylpiperazine | <u>DTXSID8057855</u> DTXSID40176612 DTXSID40193102 DTXSID90296632 DTXSID50293046 DTXSID00293011 DTXSID50296633 InChIKeyBlock1 = | 162.11576 | C ₁₀ H ₁₄ N ₂ |  | 3.9473 | Peaks: 6 / 8 Fragments Scores Download |

Process Candidates (Ungrouped)

Fragmentation Settings & Processing

Mzppm:

Mzabs:

Mode:

Tree depth:

Group candidates



Process Candidates

225 Candidates processed

MS/MS Peak list

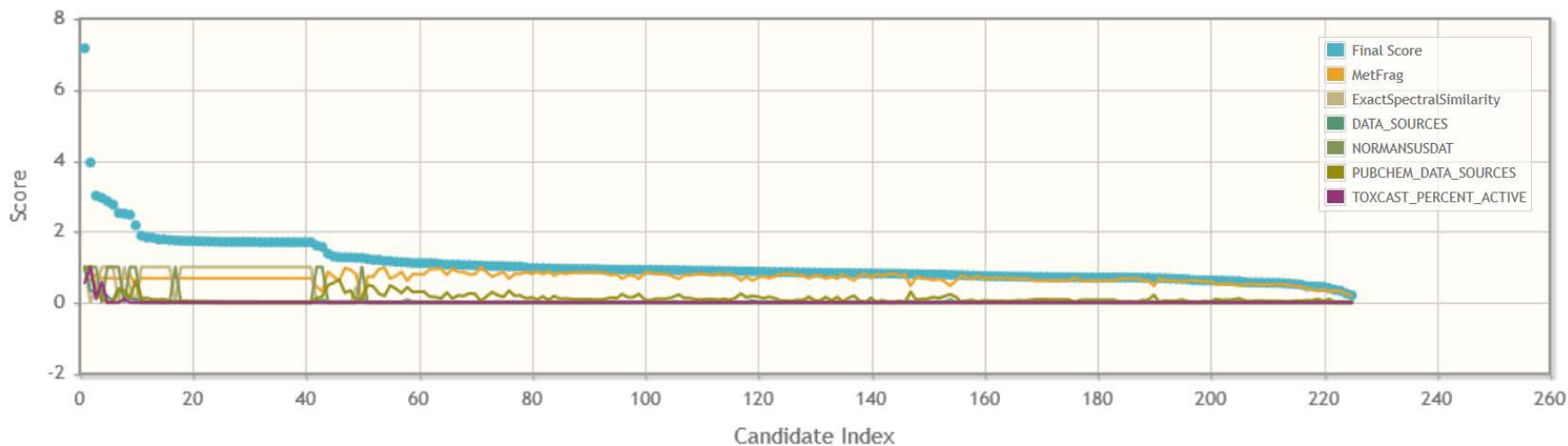
```
80.0494 6261028.7 23
84.0807 13197924.1 50
94.065 967625.9 3
106.065 24640249.3 93
117.0572 3192413.5 12
120.0807 8648923.7 32
130.0651 24669353.9 93
132.0807 80112590.7 304
163.1229 263120223.6
999
```

Show Spectrum

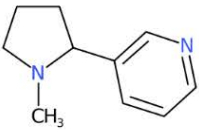
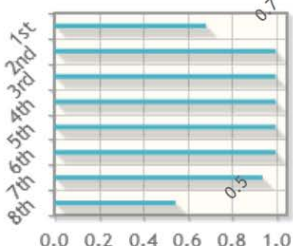
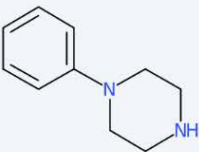
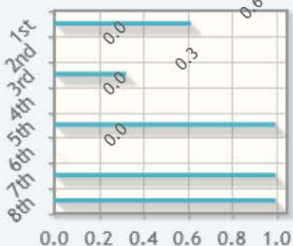
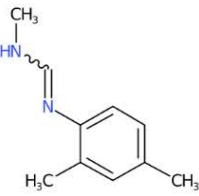
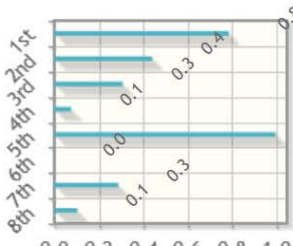
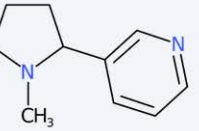
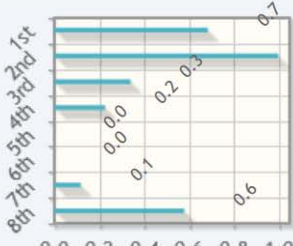
Download Parameters

Statistics

Candidate Score Distribution



Process Candidates (Ungrouped)

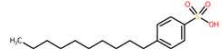
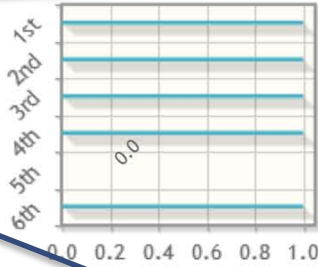
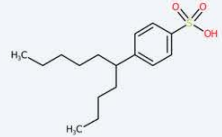
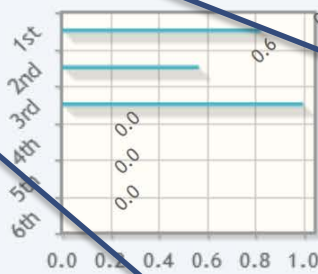
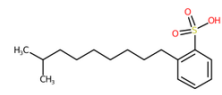
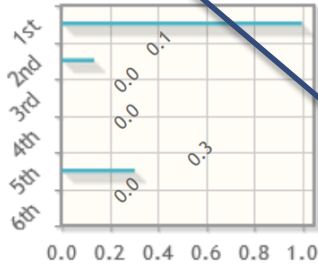
| <div> <div>1</div> <div>2</div> <div>3</div> <div>4</div> <div>5</div> <div>6</div> <div>7</div> <div>8</div> <div>9</div> <div>10</div> </div> | | | | | | | |
|---|--|---|-----------|--|---|------------|--|
| # | Molecule | Identifier | Mass | Formula | Normalized Scores | FinalScore | Details |
| 1 |  <p>Nicotine</p> | <p><u>DTXSID1020930</u></p> <p>InChIKeyBlock1 = <u>SNICXCGAKADSCV</u></p> | 162.11576 | C ₁₀ H ₁₄ N ₂ |  | 7.1739 | <p>Peaks: 6 / 8</p> <p>Fragments</p> <p>Scores</p> <p>Download</p> |
| 2 |  <p>Phenylpiperazine</p> | <p><u>DTXSID8057855</u></p> <p>InChIKeyBlock1 = <u>YZTJYBJCZXZGCT</u></p> | 162.11576 | C ₁₀ H ₁₄ N ₂ |  | 3.9473 | <p>Peaks: 6 / 8</p> <p>Fragments</p> <p>Scores</p> <p>Download</p> |
| 3 |  <p>N'-(2,4-Dimethylphenyl)-N-methylformamidine</p> | <p><u>DTXSID1037696</u></p> <p>InChIKeyBlock1 = <u>JIIOLEGNERQDIP</u></p> | 162.11576 | C ₁₀ H ₁₄ N ₂ |  | 3.0109 | <p>Peaks: 6 / 8</p> <p>Fragments</p> <p>Scores</p> <p>Download</p> |
| 4 |  <p>Nicotine sulfate</p> | <p><u>DTXSID8021725</u></p> <p>InChIKeyBlock1 = <u>SNICXCGAKADSCV</u></p> | 162.11576 | C ₁₀ H ₁₄ N ₂ |  | 2.9533 | <p>Peaks: 6 / 8</p> <p>Fragments</p> <p>Scores</p> <p>Download</p> |

Example 3

"MS-ready" Form for MetaData in MetFrag

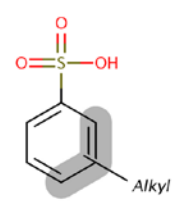
MetFrag <https://msbi.ipb-halle.de/MetFragMSready/> Search

Results

| # | Molecule | Identifier | Mass | Formula | Normalized Scores |
|---|---|---|-----------|--|---|
| 1 |  (C10-C16) Alkylbenzenesulfonic acid | DTXSID2028723 DTXSID7059696 DTXSID5041647 InChIKeyBlock1 = UASQKKHYUPBQJR | 298.16027 | C ₁₆ H ₂₆ O ₃ S |  |
| 2 |  Alkylbenzenesulfonate, linear | DTXSID3020041 DTXSID70881146 InChIKeyBlock1 = KIIODTIGNUGDLO | 298.16027 | C ₁₆ H ₂₆ O ₃ S |  |
| 7 |  Benzenesulfonic acid, isodecyl-, compd. with 2,2'-iminobis[ethanol] (1:1) | DTXSID2070762 InChIKeyBlock1 = CKNPXKLNOLAXDF | 298.16027 | C ₁₆ H ₂₆ O ₃ S |  |

(C10-C16) Alkylbenzenesulfonic acid
68584-22-5 | DTXSID2028723

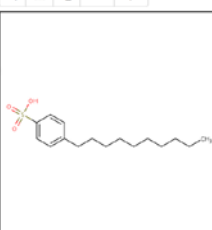
Search by DSSTox_Substance_id: Found 1 result for 'DTXSID2028723'.



Alkyl

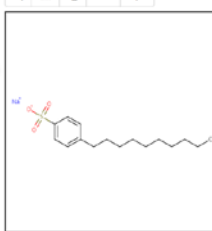
4-Decylbenzenesulfonic acid
140-60-3 | DTXSID7059696

Search by DSSTox_Substance_id: Found 1 result for 'DTXSID7059696'.



Sodium 4-decylbenzenesulfonate
2627-06-7 | DTXSID5041647

Search by DSSTox_Substance_id: Found 1 result for 'DTXSID5041647'.



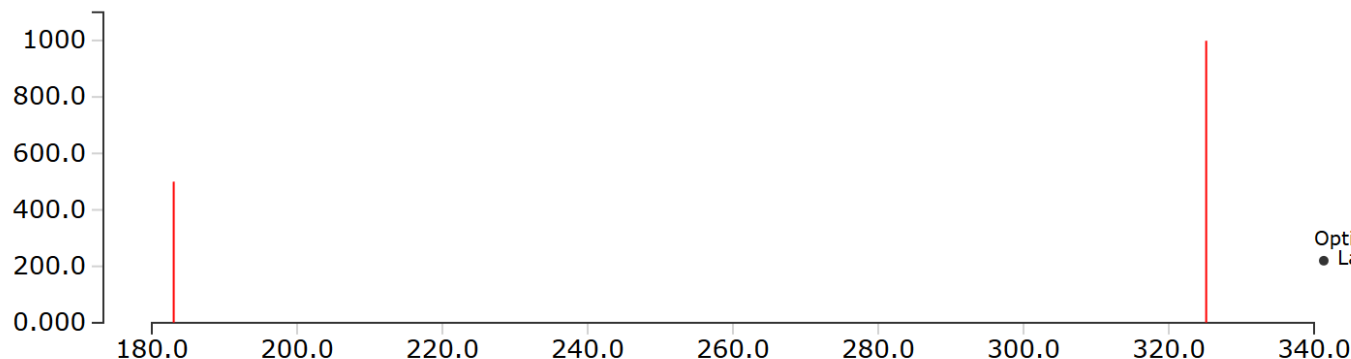
C12-LAS (Literature Spectrum in MassBank)

MassBank Record: LIT00003

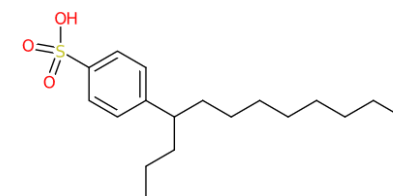
[Home](#) | [Search](#) | [Record Index](#) | [Data Privacy](#) | [Imprint](#) | MassBank ID:

C12-LAS; ESI-ITFT; MS2; 30 V; [M-H]-

Mass Spectrum



Chemical Structure



PK\$SPLASH: [splash10-0059-0409000000-8109f56885c4f79bf436](#)

PK\$ANNOTATION: m/z formula putative_smiles

183.01209 C8H7O3S- O=S([O-])(=O)c1ccc(C=C)cc1

325.14794 C17H25O4S- co-eluting_compound

PK\$NUM_PEAK: 2

PK\$PEAK: m/z int. rel.int.

183.01209 50 500

325.14794 100 999

//

MetFrag with C12-LAS and CompTox

▼ Database Settings

Database:

CompTox_07March19_Sel ▼

Parent Ion:

Neutral Mass:

326.1917

Search ppm: 5

Formula:

C₁₈H₃₀O₃S

Identifiers:

Retrieve Candidates

 112 Candidates

Download Candidates

► Candidate Filter & Score Settings

▼ Fragmentation Settings & Processing

Mzppm:

5

Mzabs:

0.001

Mode:

[M-H]- ▼

Tree depth:

2 ▼

Group candidates



MS/MS Peak list

183.01209 50 500
325.14794 100 999

Show Spectrum

Process Candidates

Download Parameters

MetFrag with C12-LAS and CompTox

☐ Spectral Similarity (MoNA)

☒ Exact Spectral Similarity (MoNA)

☐ Statistical Scoring

Database Scoring Terms

Select Item(s) ▼

4 of 7 item(s) selected

- ☒ DATA_SOURCES
- ☐ ECOTOX
- ☐ KEMIMARKET
- ☒ NORMANSUSDAT
- ☒ PUBCHEM_DATA_SOURCES
- ☐ TOX21SL
- ☒ TOXCAST_PERCENT_ACTIVE

MetFrag with C12-LAS and CompTox



MetFrag with C12-LAS and CompTox

Results

Weights

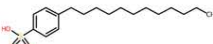
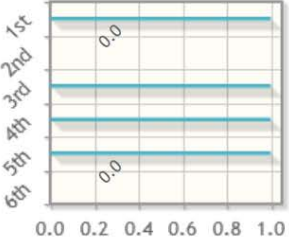
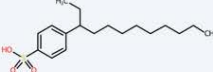
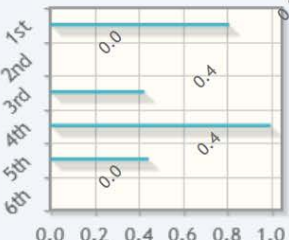
| | | | |
|-------------------------------|-----------------------|-----|---|
| MetFrag (1st) | <input type="range"/> | 100 | % |
| ExactSpectralSimilarity (2nd) | <input type="range"/> | 100 | % |
| DATA_SOURCES (3rd) | <input type="range"/> | 100 | % |
| NORMANSUSDAT (4th) | <input type="range"/> | 100 | % |
| PUBCHEM_DATA_SOURCES (5th) | <input type="range"/> | 100 | % |
| TOXCAST_PERCENT_ACTIVE (6th) | <input type="range"/> | 100 | % |

Download Results

Filter Candidates by explained MS/MS Peaks

MS/MS Peaks

Filter Candidates

| # | Molecule | Identifier | Mass | Formula | Normalized Scores | FinalScore | Details |
|---|--|--|----------|--|---|------------|--|
| 1 |  4-Dodecylbenzenesulfonic acid | <p>DTXSID8050443</p> <p>DTXSID9042413 DTXSID8069226 DTXSID9065786 DTXSID2065759 DTXSID5074264 DTXSID5074260</p> <p>InChIKeyBlock1 = KWXICGTUELOLSQ</p> | 326.1917 | C ₁₈ H ₃₀ O ₃ S |  | 4.0 | <p>Peaks: 1 / 1</p> <p>Fragments</p> <p>Scores</p> <p>Download</p> |
| 2 |  4-(3-Dodecanyl)benzenesulfonic acid | <p>DTXSID7058670</p> <p>DTXSID1058711 DTXSID8058669 DTXSID00891314 DTXSID50880399 DTXSID00873438</p> <p>InChIKeyBlock1 = QJRVOJKLQNSNDB</p> | 326.1917 | C ₁₈ H ₃₀ O ₃ S |  | 2.688 | <p>Peaks: 1 / 1</p> <p>Fragments</p> <p>Scores</p> <p>Download</p> |


Example 4 – own CSV file!

New MetaData: Disease-Specific Reference Counts

https://comptox.epa.gov/dashboard/chemical_lists/litminedneuro

| Chemical | CAS RN | DSSToxID | PMID Ct | Seizures | Nervous System Diseases | Peripheral Nervous System Diseases | Brain Diseases | Muscular Diseases | Basal Ganglia Diseases | Parkinson Disease, Secondary | Coma | Hallucinations | Tremor | Memory Disorders | Central Nervous |
|-----------------------------|------------|--------------------------------|---------|----------|-------------------------|------------------------------------|----------------|-------------------|------------------------|------------------------------|------|----------------|--------|------------------|-----------------|
| Cisplatin | 15663-27-1 | DTXSID4024983 | 1032 | 20 | 47 | 140 | 13 | 0 | 4 | 1 | 1 | 0 | 1 | 2 | 4 |
| Ethanol | 64-17-5 | DTXSID9020584 | 768 | 100 | 23 | 11 | 18 | 26 | 1 | 3 | 20 | 6 | 17 | 54 | 2 |
| Lead | 7439-92-1 | DTXSID2024161 | 740 | 28 | 107 | 68 | 102 | 4 | 2 | 2 | 1 | 3 | 4 | 19 | 30 |
| Lithium | 7439-93-2 | DTXSID5036761 | 689 | 30 | 50 | 9 | 22 | 5 | 36 | 13 | 25 | 6 | 93 | 12 | 15 |
| Valproic Acid | 76584-70-8 | DTXSID70227388 | 666 | 32 | 10 | 3 | 65 | 6 | 10 | 18 | 45 | 5 | 18 | 4 | 2 |
| 1-Methyl-4-phenylpiperazine | 28289-54-5 | DTXSID8040933 | 638 | 1 | 24 | 0 | 11 | 0 | 6 | 289 | 0 | 0 | 5 | 0 | 1 |
| Vincristine | 2068-78-2 | DTXSID8044331 | 567 | 17 | 59 | 125 | 15 | 5 | 1 | 1 | 5 | 3 | 2 | 1 | 8 |
| Phenytoin | 57-41-0 | DTXSID8020541 | 560 | 37 | 24 | 25 | 16 | 9 | 3 | 1 | 9 | 3 | 8 | 4 | 6 |
| Haloperidol | 52-86-8 | DTXSID4034150 | 555 | 6 | 6 | 1 | 10 | 6 | 153 | 51 | 4 | 4 | 11 | 1 | 0 |
| Cocaine | 50-36-2 | DTXSID2038443 | 530 | 151 | 16 | 0 | 8 | 0 | 2 | 3 | 3 | 8 | 6 | 12 | 11 |
| Aspirin | 50-78-2 | DTXSID5020108 | 489 | 8 | 3 | 0 | 3 | 2 | 2 | 0 | 9 | 4 | 1 | 0 | 5 |
| Paclitaxel | 33069-62-4 | DTXSID9023413 | 485 | 4 | 43 | 217 | 9 | 14 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| Aluminum | 7429-90-5 | DTXSID3040273 | 477 | 13 | 41 | 1 | 105 | 4 | 0 | 0 | 1 | 0 | 1 | 13 | 12 |
| Lidocaine | 6108-05-0 | DTXSID80209953 | 464 | 150 | 26 | 15 | 3 | 2 | 0 | 0 | 8 | 4 | 6 | 2 | 10 |
| Methotrexate | 59-05-2 | DTXSID4020822 | 451 | 17 | 25 | 1 | 79 | 4 | 0 | 1 | 5 | 0 | 1 | 9 | 18 |
| Mercury | 7439-97-6 | DTXSID1024172 | 450 | 6 | 79 | 22 | 23 | 2 | 3 | 5 | 2 | 2 | 38 | 7 | 25 |

MetFrag + Disease-based MetaData



MetFrag

In silico fragmentation for computer

▼ Database Settings

Database:

CSV ▼

Upload File:

+ Choose

i Uploaded PubMedNeurotox_14Jul2019_MSready_Disease...

Neutral Mass:

252.08992

Search ppm:

Formula:

C15H12N2O2

Identifiers:

Retrieve Candidates

i 3 Candidates

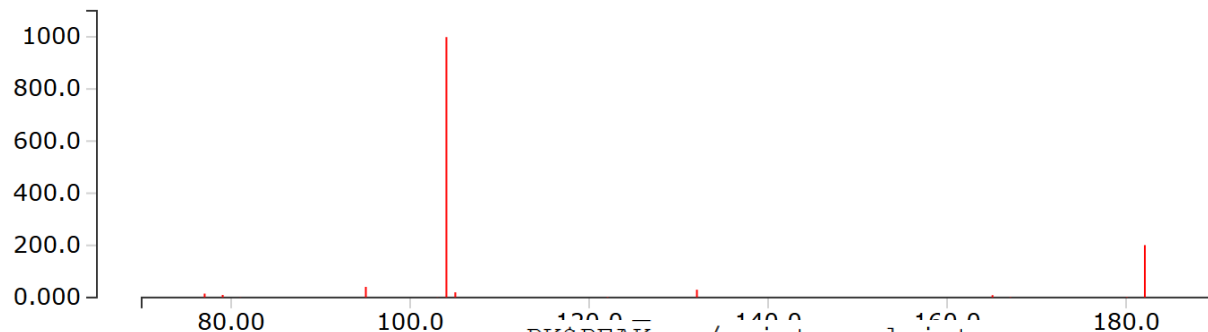
MetFrag + Disease-based MetaData

MassBank Record: EQ331905

[Home](#) | [Search](#) | [Record Index](#) | [Data Privacy](#) | [Imprint](#) | MassBank ID:

Phenytoin; LC-ESI-QFT; MS2; CE: 75; R=35000; [M+H]⁺

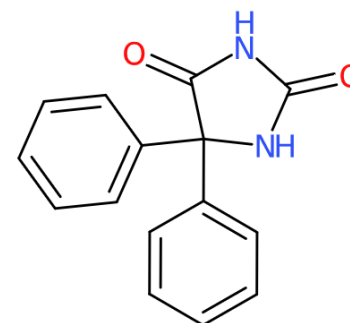
Mass Spectrum



| m/z | int. | rel.int. |
|----------|-----------|----------|
| 77.0385 | 5324013 | 15 |
| 79.0542 | 3557309.5 | 10 |
| 81.0334 | 388320.8 | 1 |
| 95.0491 | 14019273 | 41 |
| 103.0412 | 409314.6 | 1 |
| 104.0494 | 339629280 | 999 |
| 105.0333 | 6008052 | 17 |
| 105.0445 | 6911438 | 20 |
| 122.0599 | 695733.5 | 2 |
| 132.0443 | 10507750 | 30 |
| 165.0699 | 3174707.5 | 9 |
| 167.073 | 690844.3 | 2 |
| 180.0808 | 721253.2 | 2 |
| 182.0964 | 68547840 | 201 |

//

Chemical Structure



Options
● Labels

Select MetaData Terms ...

☒ Exact Spectral Similarity (MoNA)

☐ Statistical Scoring

Database Scoring Terms

Select Item(s) 5 of 14 item(s) selected

☐


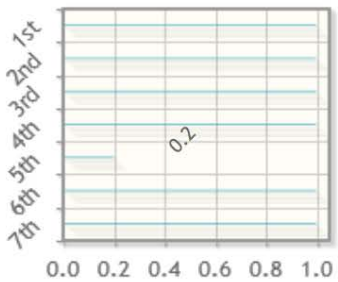
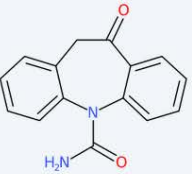
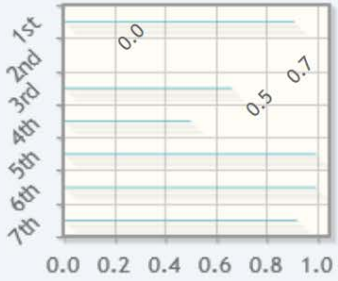
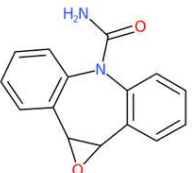
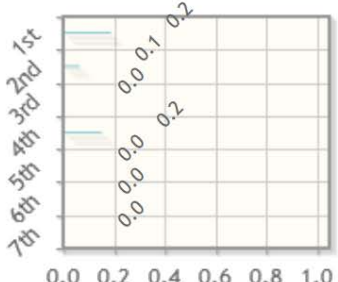
- ☐ Nervous System Diseases
- ☐ PUBCHEM_DATA_SOURCES
- ☒ Parkinson Disease, Secondary
- ☐ Peripheral Nervous System Diseases
- ☐ PubMedID_COUNT
- ☐ Seizures
- ☒ TOXCAST_PERCENT_ACTIVE

Show Spectrum

Weights

| | | | |
|--|-----------------------|-----|---|
| MetFrag (1st) | <input type="range"/> | 100 | % |
| ExactSpectralSimilarity (2nd) | <input type="range"/> | 100 | % |
| Basal Ganglia Diseases (3rd) | <input type="range"/> | 100 | % |
| DATA_SOURCES (4th) | <input type="range"/> | 100 | % |
| EXPOCAST_MEDIAN_EXPOSURE_PRE DICTION_MG/KG-BW/DAY (5th) | <input type="range"/> | 100 | % |
| Parkinson Disease, Secondary (6th) | <input type="range"/> | 100 | % |

MetFrag + Disease-based MetaData

| # | Molecule | Identifier | Mass | Formula | Normalized Scores | FinalScore | Details |
|---|---|---|-----------|---|--|------------|--|
| 1 |  <p>Phenytoin</p> | <p><u>DTXSID8020541</u></p> <p>InChIKeyBlock1 = <u>CXOFVDLJLONNDW</u></p> | 252.08992 | C ₁₅ H ₁₂ N ₂ O ₂ |  | 6.2006 | <p>Peaks: 9 / 14</p> <p>Fragments</p> <p>Scores</p> <p>Download</p> |
| 2 |  <p>oxcarbazepine</p> | <p><u>DTXSID0045703</u></p> <p>InChIKeyBlock1 = <u>CTRLABGOLIVAIY</u></p> | 252.08992 | C ₁₅ H ₁₂ N ₂ O ₂ |  | 5.0122 | <p>Peaks: 11 / 14</p> <p>Fragments</p> <p>Scores</p> <p>Download</p> |
| 3 |  <p>carbamazepine epoxide</p> | <p><u>DTXSID60891456</u></p> <p>InChIKeyBlock1 = <u>ZRWWEVEIOGMMT</u></p> | 252.08992 | C ₁₅ H ₁₂ N ₂ O ₂ |  | 0.4037 | <p>Peaks: 7 / 14</p> <p>Fragments</p> <p>Scores</p> <p>Download</p> |

MetFrag + Disease-based MetaData


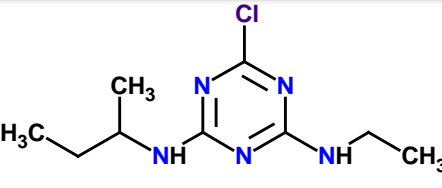
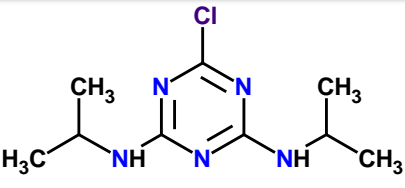
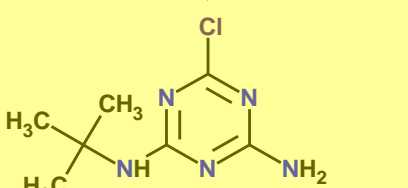
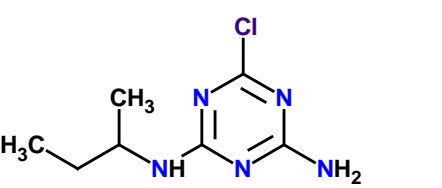
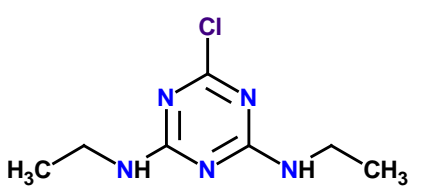
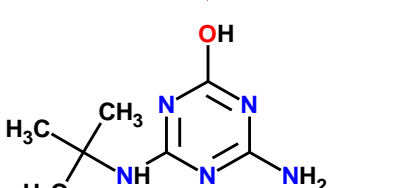
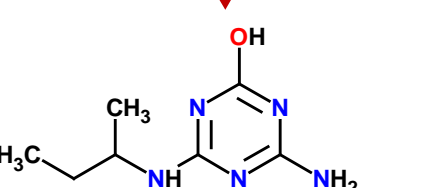
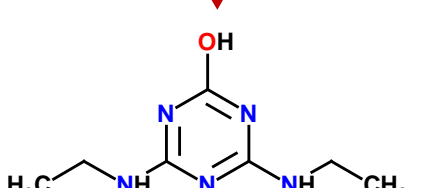
Scores View

Candidate Name: Phenytoin
Candidate Identifier: DTXSID8020541|5

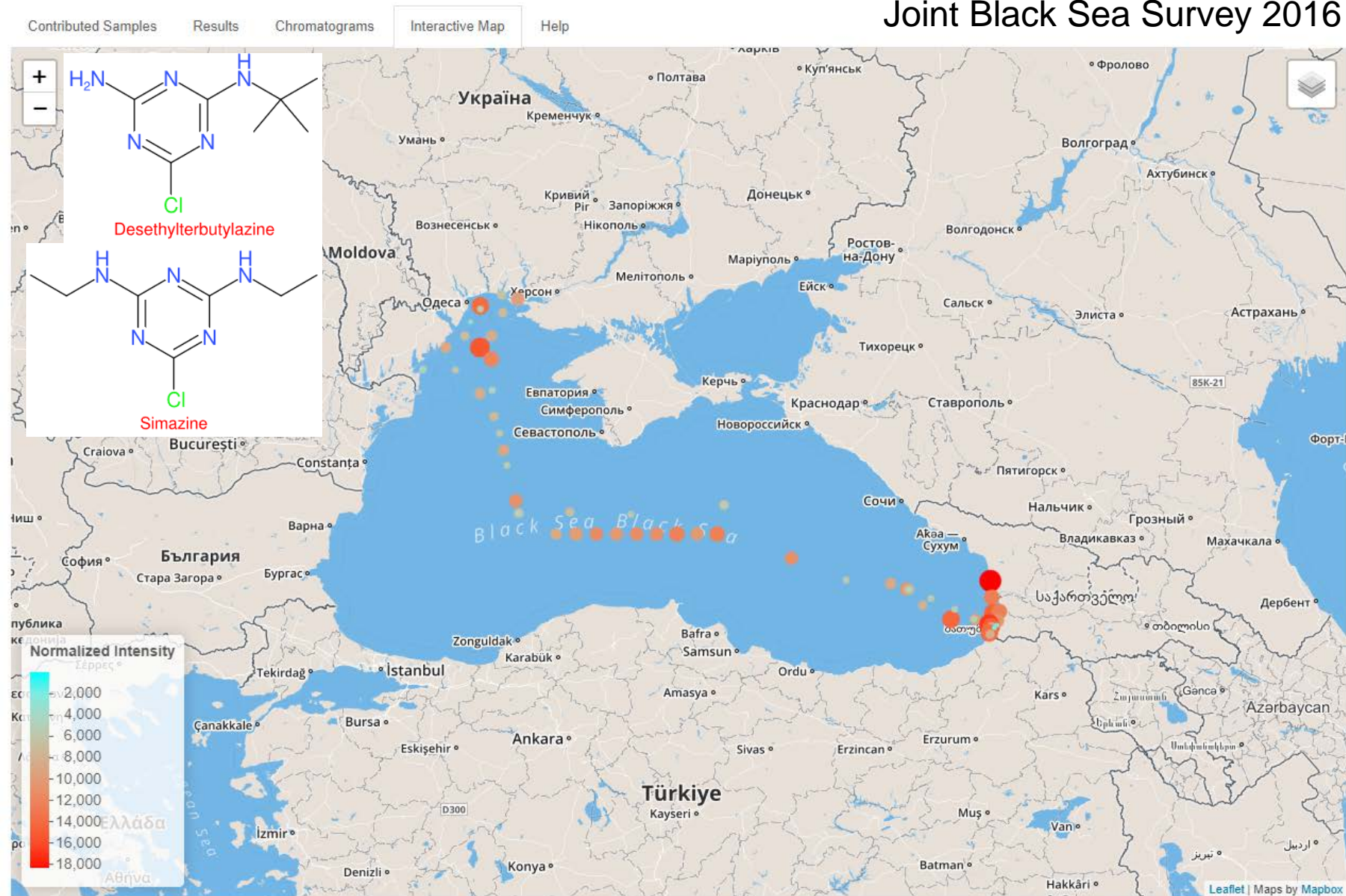
| | Name | Normalized Value | Raw Value |
|---|--|------------------|-----------|
| ▶ | MetFrag | 1.0 | 92.9258 |
| ▶ | ExactSpectralSimilarity | 1.0 | 1.0 |
| ▶ | Basal Ganglia Diseases | 1.0 | 3.0 |
| ▶ | DATA_SOURCES | 1.0 | 105.0 |
| ▶ | EXPOCAST_MEDIAN_EXPOSURE_PREDICTIO N_MG/KG-BW/DAY | 0.2006 | 0.0 |
| ▶ | Parkinson Disease, Secondary | 1.0 | 1.0 |
| ▶ | TOXCAST_PERCENT_ACTIVE | 1.0 | 2.5605 |

Example 5

NT Trial: The tricky cases!

| | | | | |
|---------------------------------------|--|---|--|---|
| $C_9H_{16}ClN_5$ m/z 229.1094 Da |  <p>Terbutylazine Detects: 12; # Refs: 220</p> |  <p>Sebutylazine Detects: 3; # Refs: 51</p> | <p>(no related compound at this mass)</p> |  <p>Propazine Detects: 3; # Refs: 201</p> |
| $C_7H_{12}ClN_5$ m/z 201.0781 Da |  <p>Terbutylazine-desethyl Detects: 9; # Refs: 92</p> |  <p>Sebutylazine-desethyl Detects: 1; # Refs: 14</p> |  <p>Simazine Detects: 4; # Refs: 518</p> | <p>(no related compound at this mass)</p> |
| $C_7H_{13}N_5O$ m/z 183.1120 Da |  <p>Terbutylazine-desethyl-2-hydroxy Detects: 2; # Refs: 57</p> |  <p>Sebutylazine-desethyl-2-hydroxy Detects: 0; # Refs: 3</p> |  <p>Simazine-2-hydroxy Detects: 2; # Refs: 66</p> | <p>(no related compound at this mass)</p> |

Joint Black Sea Survey 2016



MetFrag Web Interface – Retrieve Candidates



MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

Database Settings

Database:

PubChem

Include references:



Parent
Ion:

☐

Neutral Mass:

201.07816

Search ppm:

5

Formula:

C7H12ClN5

Identifiers:

Retrieve Candidates



73 Candidates

Download Candidates

MetFrag Web Interface – Add MS/MS Details



MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

Fragmentation Settings & Processing

Mzppm:

Mzabs:

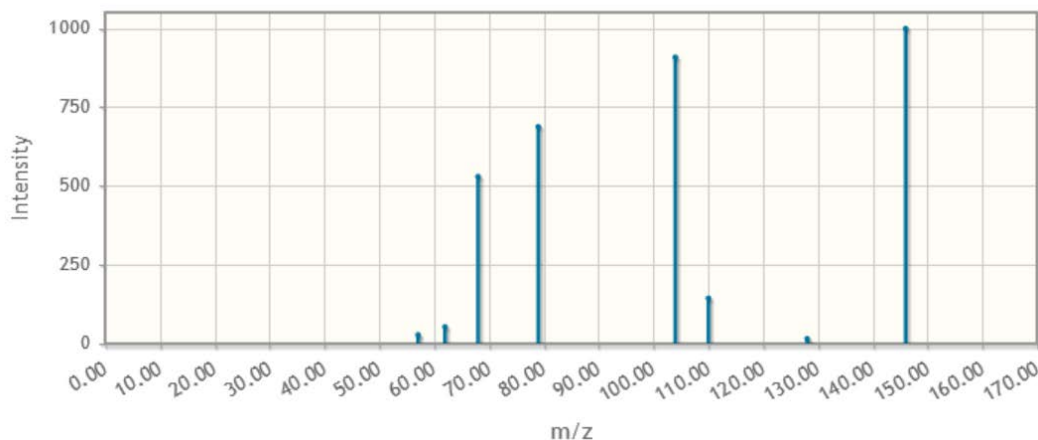
Mode:

Spectrum View

Select area to zoom in. Double click to return.

Tree depth:

Group candida



MS/MS Peak list

```
57.07 50398.6 27
61.9793 96225.8 53
68.0245 959131.3 529
79.0059 1246110.1 687
104.0012 1644724 907
110.0463 260579 143
128.0568 27853.4 15
146.023 1809905.6 999
```

Show Spectrum

Download Parameters

Process Ca

MetFrag Web Interface – Add Metadata Info



MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

Candidate Filter & Score Settings

Candidate Filters

- ☐ Element Inclusion
- ☐ Element Exclusion
- ☐ Substructure Inclusion
- ☐ Substructure Exclusion
- ☐ Substructure Information
- ☐ Minimum Number Elements

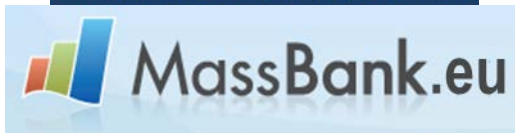
MetFrag Scoring Terms



☐ Suspect Inclusion Lists

MoNA
MassBank of North America

☐ Spectral Similarity (MoNA)



☐ Exact Spectral Similarity (MoNA)



☐ Statistical Scoring



Literature



Patents

Database Scoring Terms

Select Item(s)



0 of 2 item(s) selected



NORMAN Suspect List Exchange – NORMAN SLE

| No. | Abbreviation | Description | Link to full list | Link to InChIKey list | References |
|-----|--------------|---|--|--|---|
| S0 | SUSDAT | Merged NORMAN Suspect List: SusDat | Interactive Data table (updating...) CompTox SUSDAT List | MS-ready InChIKeys (1/03/2018) | A merged list of >40,000 structures from suspect lists. See interactive version . Compiled by Reza Aalizadeh, University of Athens, including RTI and toxicity values, support by Nikiforos Alygizakis, EI. <i>Work in progress ... please report any issues!</i> |
| S1 | MASSBANK | NORMAN Compounds in MassBank | CSV, XLSX with Fragments (3/10/2017) CompTox MassBank EU Reference List CompTox MassBank EU Special Cases CompTox Fragment Download | MassBankEUInChIKeys (11/04/2017) | www.massbank.eu Stravs <i>et al.</i> 2013. DOI: 10.1002/jms.3131 |
| S2 | STOFFIDENT | HSWT/Lfu STOFF-IDENT Database of Water-Relevant Substances | STOFF-IDENT Contents (6/09/2017) CompTox STOFF-IDENT List Further curation in progress... | STOFF-IDENT InChIKeys (6/09/2017) | The database enables the search for exact masses from target or unknown lists and the automatic use of a Retention Time Index. See: https://www.lfu.bayern.de/stoffident/#!home (single search for free; batch search after free registration). |
| S3 | NORMANCT15 | NORMAN Collaborative Trial Targets and Suspects | LC-MS: CSV, XLSX (3/10/2017) GC-MS: CSV, XLSX (3/10/2017) CompTox NORMANCT15 List | LC-MS InChIKeys (31/10/2016) GC-MS InChIKeys (31/10/2016) | Schymanski <i>et al.</i> 2015. DOI: 10.1007/s00216-015-8681-7 |

MetFrag Web Interface



MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra



Suspect Inclusion Lists

+ Choose

Uploaded suspect lists

Suspect List Name

Number Entries

SusDat_MSReady_
InChIKeys_010320
18.txt

40052

Remove

Number uploaded suspect lists: 1

Predefined Suspect Lists:

☐

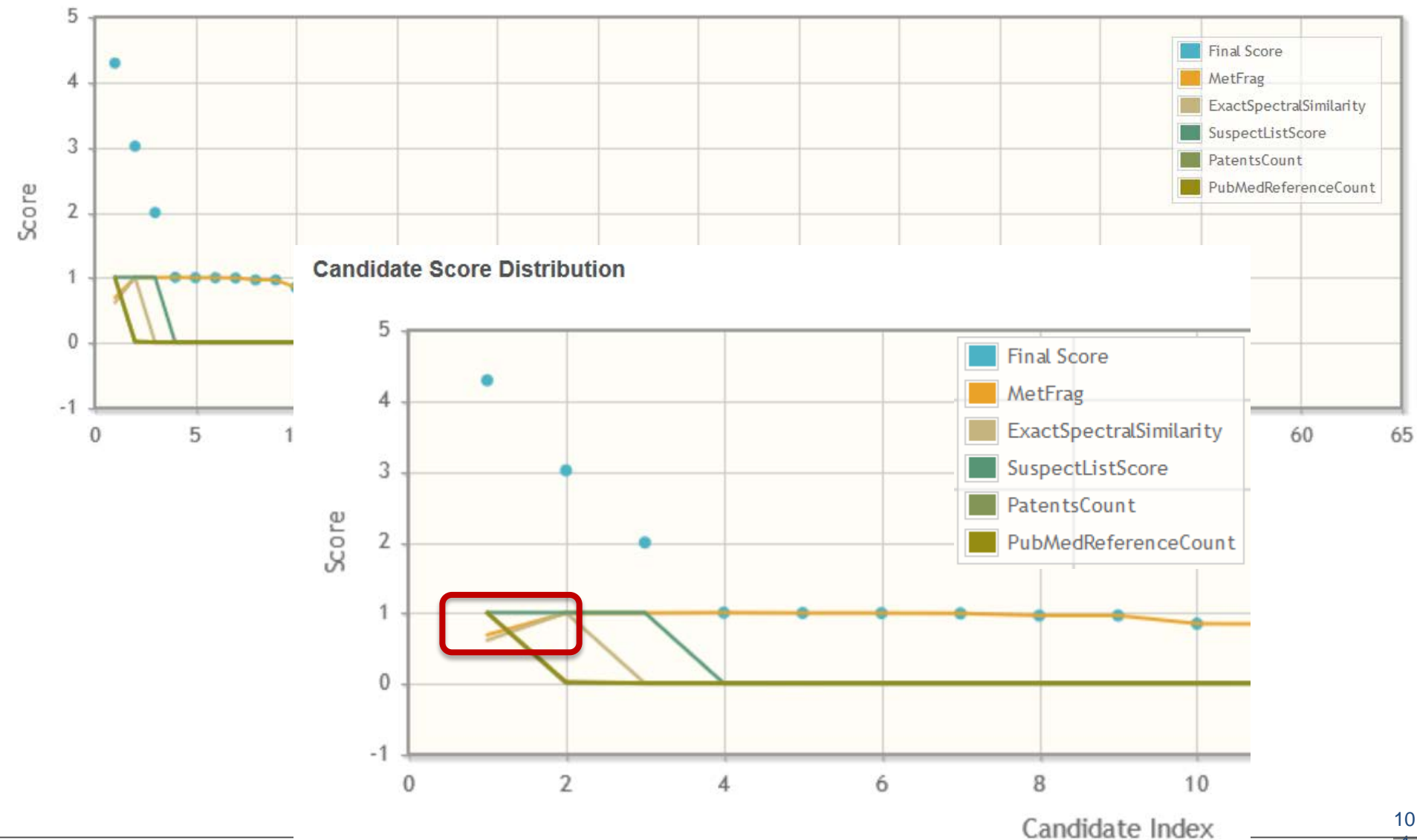
FOR-IDENT (Find out more about [ForIdent](#))

☐

DSSTox (Find out more about [DSSTox](#))

MetFrag Results

Candidate Score Distribution



MetFrag Results – Metadata is good ... but ...

Weights

MetFrag (1st) 100 %

ExactSpectralSimilarity (2nd) 100 %

SuspectListScore (3rd) 100 %

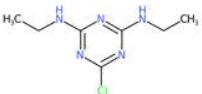
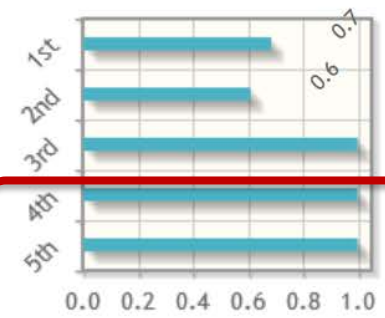
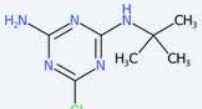

PatentsCount (4th) 100 %

PubMedReferenceCount (5th) 100 %

Download Results

Filter Candidates by explained MS/MS Peaks

MS/MS Peaks

| # | Molecule | Identifier | Mass | Formula | Normalized Scores | FinalScore | Details |
|---|--|---|---------|---|---|------------|---|
| 1 |  <p>6-chloro-N2,N4-diethyl-1,3,5-triazine-2,4-diamine</p> | <p>5216</p> <p>InChIKeyBlock1 = <u>ODCWYMIRDDJXKW</u></p> | 201.078 | C ₇ H ₁₂ ClN ₅ |  | 4.2927 | <p>Peaks: 4 / 8</p> <p><input type="button" value="Fragments"/></p> <p><input type="button" value="Scores"/></p> <p><input type="button" value="Download"/></p> |
| 2 |  <p>N2-tert-butyl-6-chloro-1,3,5-triazine-2,4-</p> | <p>108201</p> <p>InChIKeyBlock1 = <u>LMKQNTMFZLAJDV</u></p> | 201.078 | C ₇ H ₁₂ ClN ₅ |  | 3.0155 | <p>Peaks: 6 / 8</p> <p><input type="button" value="Fragments"/></p> <p><input type="button" value="Scores"/></p> <p><input type="button" value="Download"/></p> |

MetFrag Results - Consider Experimental Evidence!

Weights

MetFrag (1st) 100 %

ExactSpectralSimilarity (2nd) 100 %

SuspectListScore (3rd) 100 %

PatentsCount (4th) 0 %

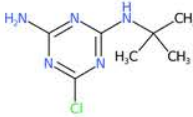
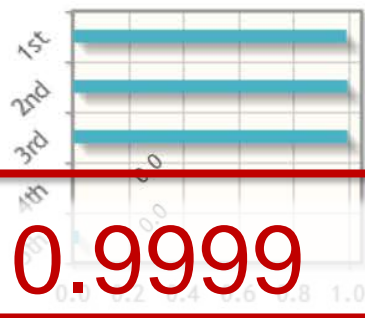
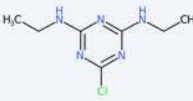
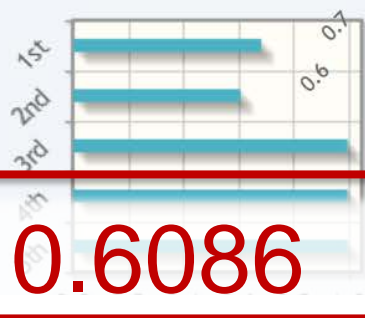
PubMedReferenceCount (5th) 0 %

Download Results

Filter Candidates by explained MS/MS Peaks

MS/MS Peaks

1 2 3 4 5 6

| # | Molecule | Identifier | Mass | Formula | Normalized Scores | FinalScore | Details |
|---|--|---|---------|---|---|------------|--|
| 1 |  N2-tert-butyl-6-chloro-1,3,5-triazine-2,4-diamine | <u>108201</u> InChIKeyBlock1 = LMKQNTMEZLAJDV | 201.078 | C ₇ H ₁₂ ClN ₅ |  | 2.9948 | Peaks: 6 / 8 <input type="button" value="Fragments"/> <input type="button" value="Scores"/> <input type="button" value="Download"/> |
| 2 |  6-chloro-N2,N4-diethyl-1,3,5-triazine-2,4- | <u>5216</u> InChIKeyBlock1 = ODCWYMIRDDJXKW | 201.078 | C ₇ H ₁₂ ClN ₅ |  | 2.2927 | Peaks: 4 / 8 <input type="button" value="Fragments"/> <input type="button" value="Scores"/> <input type="button" value="Download"/> |

(Molecule =>) Mass => Molecule => Metadata

- Background

- Molecule to Mass

- Preparation for Mass to Molecule

- Gathering the “evidence”

- Molecular Formula example

- Structure Elucidation with MetFrag

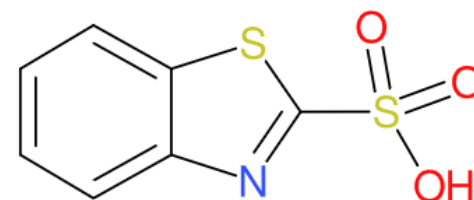
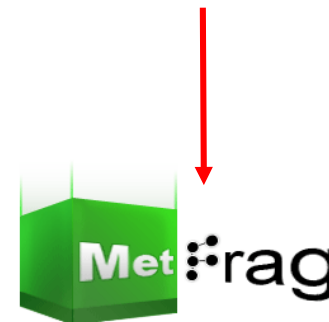
- Step-by-step with one example (MetFrag)

- Practice Session with Several Examples

- [Optional] Context and Perspectives

- Potential for automated non-target workflows

213.9637





- MS Processing:

RMassBank

- Screening List: ***Compound List of Mix Content***

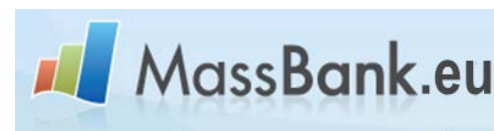
- Identification:



+

***Compound List
of Mix Content***











- Fragmenter Score
- Retention Time Score (RTs from Eawag Mix Standards)
- Offline MetFusion Score
- Offline Individual MONA Score



- **Maximum Score: 4**
- **NO REFERENCE INFORMATION (irrelevant)**

\\Disk (C:) > DATA > External > US_EPA > ENTACT > ENTACT_R > Unblinded_files

☐ Name

 Mix499_CmpdList.csv
 Mix500_CmpdList.csv
 Mix501_CmpdList.csv
 Mix502_CmpdList.csv
 Mix503_CmpdList.csv
 Mix504_CmpdList.csv
 Mix505_CmpdList.csv
 Mix506_CmpdList.csv
 Mix507_CmpdList.csv
 Mix508_CmpdList.csv

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|----|--------------|-----------|------------|------------|-----------|---------------------|------------|----------|----------|----|-------|------------|---|
| 1 | ID | Name | DTXSID | CAS | SMILES_DT | DTXCID | SMILES | InChIKey | Formula | mz | RT | Level | Source | |
| 2 | 1 | 4-Chloro-2- | DTXSID002 | 3165-93-3 | Cl.CC1=C(N | DTXCID902 | CC1=C(N)C: CXNVOWPI | C7H8ClN | | 141.0345 | | 1 | BF00173499 | |
| 3 | 2 | Ethionamic | DTXSID002 | 536-33-4 | CCCC1=NC=C | DTXCID005 | CCCC1=NC=C | AEOCXXJPG | C8H10N2S | 166.0565 | | 1 | BF00173499 | |
| 4 | 3 | Phenobarb | DTXSID002 | 57-30-7 | [Na+].CCC1 | DTXCID601 | CCC1(C(=O) | DDBREP KU | C12H12N2 | 232.0848 | | 1 | BF00173499 | |
| 5 | 4 | Sulfasalazir | DTXSID002 | 599-79-1 | OC(=O)C1=C | DTXCID401 | OC(=O)C1=C | NCEXYHBE | C18H14N4 | 398.0685 | | 1 | BF00173499 | |
| 6 | 5 | 2-Chloro-4- | DTXSID002 | 92-04-6 | OC1=CC(Cl) | DTXCID502 | OC1=CC(Cl) | MXORDJXB | C12H9ClO | 204.0342 | | 1 | BF00173499 | |
| 7 | 6 | Olivetol | DTXSID002 | 500-66-3 | CCCCCCC1=C | DTXCID205 | CCCCCCC1=C | IRMPFYJSH | C11H16O2 | 180.115 | | 1 | BF00173499 | |
| 8 | 7 | (Z)-Hexade | DTXSID004 | 373-49-9 | CCCCC\C=C | DTXCID501 | CCCCC\C=C | SECPZKHBE | C16H30O2 | 254.2246 | | 1 | BF00173499 | |
| 9 | 8 | cis-1,2,3,6- | DTXSID004 | 1469-48-3 | O=C1NC(=C | DTXCID106 | O=C1NC(=C | CIFFBTOJCK | C8H9NO2 | 151.0633 | | 1 | BF00173499 | |
| 10 | 9 | Monopota | DTXSID004 | 877-24-7 | [K+].OC(=O | DTXCID901 | OC(=O)C1=C | XNGIFLGAS | C8H6O4 | 166.0266 | | 1 | BF00173499 | |
| 11 | 10 | Octylparab | DTXSID004 | 1219-38-1 | CCCCCCCC | DTXCID002 | CCCCCCCC | RIKCMEDSE | C15H22O3 | 250.1569 | | 1 | BF00173499 | |
| 12 | 11 | 2-Amino-5- | DTXSID102 | 121-88-0 | NC1=C(O)C | DTXCID206 | NC1=C(O)C | DOPJTDJKZ | C6H6N2O3 | 154.0378 | | 1 | BF00173499 | |
| 13 | 12 | 4-Chloroph | DTXSID102 | 106-48-9 | OC1=CC=C | DTXCID201 | OC1=CC=C | WXNZTHHC | C6H5ClO | 128.0029 | | 1 | BF00173499 | |
| 14 | 13 | Stavudine | DTXSID102 | 3056-17-5 | CC1=CN([C] | DTXCID102 | CC1=CN(C2 | XNKLVCFA | C10H12N2 | 224.0797 | | 1 | BF00173499 | |
| 15 | 14 | Pirimicarb | DTXSID103 | 23103-98-2 | CN(C)C(=O) | DTXCID901 | CN(C)C(=O) | YFGYUFNIC | C11H18N4 | 238.143 | | 1 | BF00173499 | |
| 16 | 15 | 17-Methylt | DTXSID103 | 58-18-4 | C[C@]1(O) | DTXCID202 | CC1(O)CCC | GCKMFJBG | C20H30O2 | 302.2246 | | 1 | BF00173499 | |
| 17 | 16 | Tamoxifen | DTXSID103 | 10540-29-1 | CC\C=C(/C | DTXCID901 | CC\C=C(/C | NKANXQFJ | C26H29NO | 371.2249 | | 1 | BF00173499 | |
| 18 | 17 | Phenyl 1-h | DTXSID103 | 132-54-7 | OC1=C(C=C | DTXCID901 | OC1=C(C=C | QHDYIMW | C17H12O3 | 264.0786 | | 1 | BF00173499 | |

- Chemicals in each list had a unique ID (0XXX, 1XXX, 9XXX)
- Processing done on DTXCID SMILES
- Metals now included in RMassBank [PPGs done separately]
- No retention times given (=> isobars = Level 3!!!!!!)

DATA > External > US_EPA > ENTACT > ENTACT_R > MetFrag_unblinded > RMB_EIC_PDFs

☐ Name

☒ mH_20170106_dd_neg_03_EICscan.pdf

mH_20170106_dd_neg_03_RT.csv

mH_20170106_dd_neg_04_EICscan.pdf

mH_20170106_dd_neg_04_RT.csv

mH_20170106_dd_neg_05_EICscan.pdf

mH_20170106_dd_neg_05_RT.csv

mH_20170106_dd_neg_06_EICscan.pdf

mH_20170106_dd_neg_06_RT.csv

mH_20170106_dd_neg_07_EICscan.pdf

mH_20170106_dd_neg_07_RT.csv

mH_20170106_dd_neg_08_EICscan.pdf

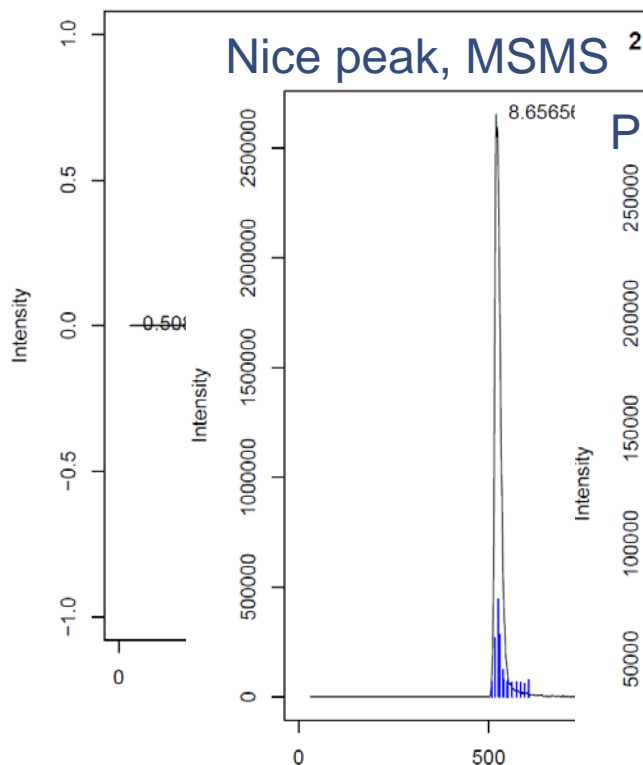
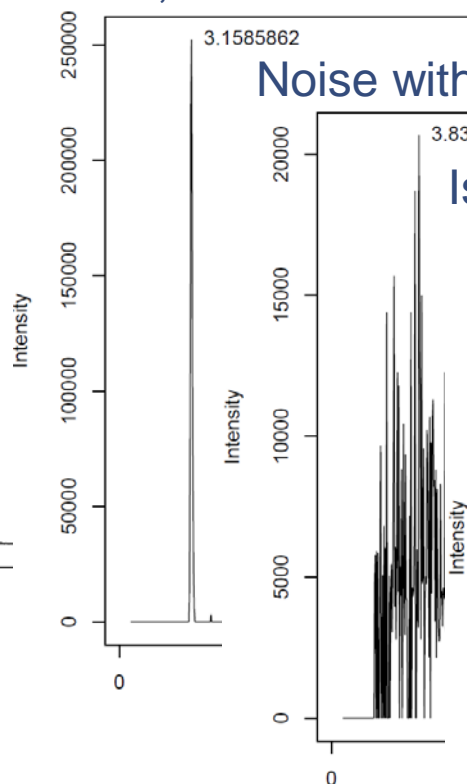
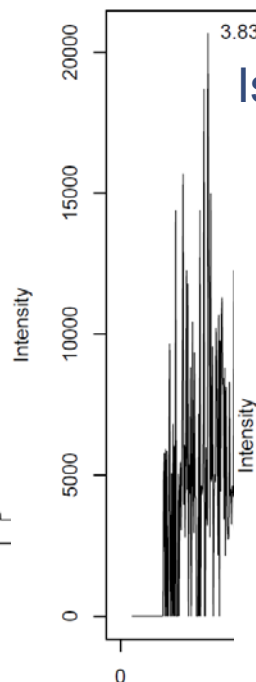
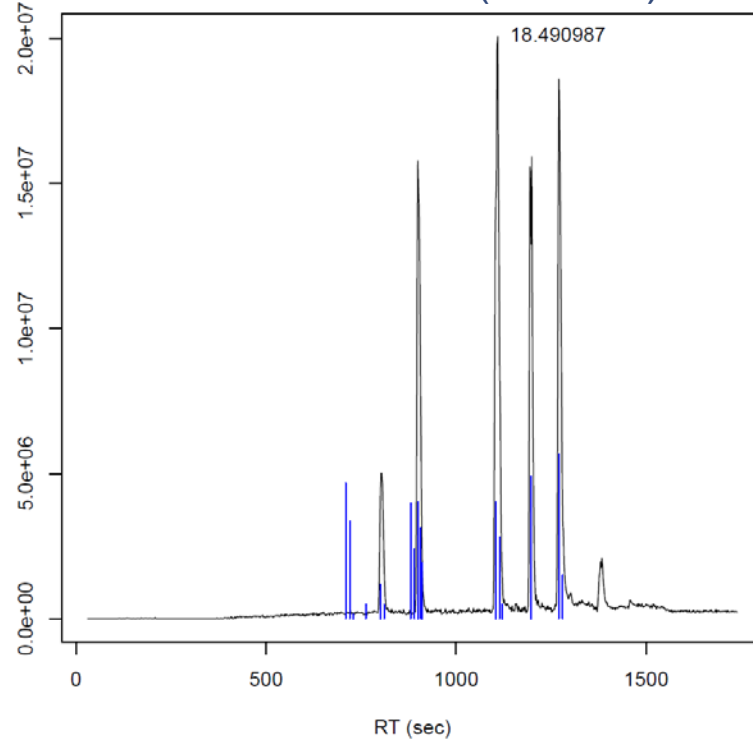
mH_20170106_dd_neg_08_RT.csv

Pre-Screening

- Auto-selects only those with MS/MS for RMassBank
- CSV and PDF output for quality control
- CSV basis for (part) of results summary
- **CAVEAT:** only takes most intense peak without RT
(=> Level 3 for isobars)
- [this can be addressed but I ran out of time ...]

No peak at all

1

Nice peak, MSMS²Peak, no MSMS²⁰Noise with MSMS⁶⁰ (careful!)Isobars with MSMS⁶³⁵³ (careful!)*

The screenshot displays the RMassBank results interface. On the left, a list of files is shown, including 20170106_dd_neg_01 through 20170106_dd_neg_06, and 20170106_dd_pos_01_FileList.csv. A file named 20170106_dd_pos_01_FileList_trim.csv is selected and highlighted in blue. A blue arrow points from this file to a Notepad window titled 'EP021502.txt - Notepad'. The Notepad window displays the following metadata for EP021502:

ACCESSION: EP021502
RECORD_TITLE: Monomethyl phthalate; LC-ESI-QFT; MS2; CE: Ramp 20,50,90; R=17500; [M+H]⁺
DATE: 2017.12.12
AUTHORS: E Schymanski, B Beck, J Hollender, Eawag
LICENSE: CC BY
COPYRIGHT: Copyright (C) 2017, Eawag
PUBLICATION:
COMMENT: CONFIDENCE Tentative identification only (Level 3)
COMMENT: INTERNAL_ID 215
CH\$NAME: Monomethyl phthalate
CH\$NAME: 2-(Methoxycarbonyl)benzoic acid
CH\$NAME: 2-methoxycarbonylbenzoic acid
CH\$COMPOUND_CLASS: N/A; Environmental Standard
CH\$FORMULA: C₉H₈O₄
CH\$EXACT_MASS: 180.0423
CH\$SMILES: COC(=O)C1=C(C(=CC=C1)C(=O)=O
CH\$IUPAC: InChI=1S/C₉H₈O₄/c1-13-9(12)7-5-3-2-4-6(7)8(10)11/h2-5H,1H3,(H,10,11)
CH\$LINK: CAS 4376-18-5
CH\$LINK: CHEBI 89749
CH\$LINK: PUBCHEM CID:20392
CH\$LINK: INCHIKEY FNJSWIPFHMKRAT-UHFFFAOYSA-N
CH\$LINK: CHEMSPIDER 19207
AC\$INSTRUMENT: Q Exactive Plus Orbitrap Thermo Scientific

On the right side of the interface, a list of text files is shown, including EP005801.txt, EP012402.txt, EP021502.txt, EP021605.txt, and EP022002.txt. A blue arrow points from the selected file to EP021502.txt.

- Information auto-retrieved via the unique ID from
 - Individual mixture lists (used also to create summaries)
 - MassBank records created during processing
 - Dummy record (mz=999, I=999) used if MS/MS was absent
- Screening List: ***LocalDB of Mix Content (localCSV)***
- Identification:
 - Fragmenter Score
 - Retention Time Score (RTs from Eawag Mix Standards)
 - Offline MetFusion Score
 - Offline Individual MONA Score
 - **Maximum Score: 4**
 - **NO REFERENCE INFORMATION (irrelevant)**

- Create Config Files (MetFragConfig)
- Run MetFrag (runMetFrag)






```
176 # now, run MetFrag and extract results for reporting into cmpd_info.
177
178 results_filename <- paste0(run_name,"_",cmpdID_char,"_",as.character(i))
179 if (isPos) {
180   config_file <- MetFragConfig(mass = ExactMass, adduct_type = 0, neutralPrecursorMass=TRUE,
181     results_filename = results_filename,
182     peaklist_path = MetFrag_msms, base_dir = results_run_dir,
183     DB = "LocalCSV", localDB_path=localDB,useMonaIndiv = T,useMoNAMetFusion = T,
184     IsPosMode = TRUE,filter_by_InChIKey = F,rt_file_path=MetFrag_rt_file,rt_exp=RT)
185 } else {
186   config_file <- MetFragConfig(mass = ExactMass, adduct_type = 0, neutralPrecursorMass=TRUE,
187     results_filename = results_filename,
188     peaklist_path = MetFrag_msms, base_dir = results_run_dir,
189     DB = "LocalCSV", localDB_path=localDB,useMonaIndiv = T,useMoNAMetFusion = T,
190     IsPosMode = FALSE,filter_by_InChIKey = F,rt_file_path=MetFrag_rt_file,rt_exp=RT)
191 }
192
193 runMetFrag(config_file, MetFrag_dir, CL_name = "MetFrag2.4.4-msready-CL.jar")
194
195 results_file <- paste0(results_run_dir,"/results/",results_filename,".xls")
```

○ Extract Results and Summarize


```
198 #extract results we need:
199 MetFrag_res <- read_excel(results_file)
200
201 compd_info$num_poss_IDs[i] <- length(MetFrag_res$Score)
202 compd_info$poss_IDs[i] <- paste(MetFrag_res$Name,collapse=";")
203 compd_info$poss_ID_scores[i] <- paste(MetFrag_res$Score,collapse=";")
204 compd_info$max_Score[i] <- max(MetFrag_res$Score)
205 compd_info$n_Score_GE3p5[i] <- length(which(MetFrag_res$Score>=3.5))
206 compd_info$n_Score_GE3[i] <- length(which(MetFrag_res$Score>=3))
207 compd_info$n_Score_GE2p5[i] <- length(which(MetFrag_res$Score>=2.5))
208 compd_info$poss_DTXSIDs[i] <- paste(MetFrag_res$DTXSID,collapse=";")
209 compd_info$poss_CAS[i] <- paste(MetFrag_res$CAS,collapse=";")
210 compd_info$MoNAScore[i] <- paste(MetFrag_res$OfflineIndividualMoNAScore,collapse=";")
211 compd_info$MaxMoNAScore[i] <- max(MetFrag_res$OfflineIndividualMoNAScore)
212
213 }
214
215
216 write.csv(compd_info,paste0(results_summary_dir,"/MetFragResultSummary_",run_name,".csv"),row.names = F)
```


Voila!

DATA > External > US_EPA > ENTACT > ENTACT_R > MetFrag

☐ Name MetFragResultSummary_20170106_dd_neg_04.csv MetFragResultSummary_20170106_dd_neg_05.csv MetFragResultSummary_20170106_dd_neg_06.csv MetFragResultSummary_20170106_dd_neg_07.csv MetFragResultSummary_20170106_dd_neg_08.csv

| | A | B | C | D | E | F | G | H | I | J | | | |
|-------|--------------|--------------|-----------|------------|------------|-----------|------------|------------|-------------|------------|-----------|--------|--------------|
| 1 | ID | Name | DTXSID | CAS | SMILES_DT | DTXCID | SMILES | InChIKey | Formula | exactMass | | | |
| 8 | 7 | (Z)-Hexade | DTXSID004 | 373-49-9 | CCCCC\C=C | DTXCID501 | CCCCC\C=C | SECPZKHBE | C16H30O2 | 254.2246 | | | |
| 9 | 8 | cis-1,2,3,6- | DTXSID004 | 1469-48-3 | O=C1NC(=C | DTXCID106 | O=C1NC(=C | CIFFBTOJCK | C8H9NO2 | 151.0633 | | | |
| 10 | 9 | Monopotas | DTXSID004 | 877-24-7 | [K+].OC(=O | DTXCID901 | OC(=O)C1= | XNGIFLGAS | C8H6O4 | 166.0266 | | | |
| 11 | 10 | Octylparab | DTXSID004 | 1219-38-1 | CCCCCCCCC | DTXCID002 | CCCCCCCCC | RIKCMEDSE | C15H22O3 | 250.1569 | | | |
| 12 | 11 | 2-Amino-5- | DTXSID102 | 121-88-0 | NC1=C(O)C | DTXCID206 | NC1=C(O)C | DOPJTDJKZ | C6H6N2O3 | 154.0378 | | | |
| 13 | 12 | 4-Chloroph | DTXSID102 | 106-48-9 | OC1=CC=C | DTXCID201 | OC1=CC=C | WXNZTHHC | C6H5ClO | 128.0029 | | | |
| 14 | 13 | Stavudine | DTXSID102 | 3056-17-5 | CC1=CN([C | DTXCID102 | CC1=CN([C | XNKLIVCAF | C10H12N2 | 224.0797 | | | |
| 15 | 14 | Pirimicarb | DTXSID103 | 23103-98-2 | CN(C)C(=O | DTXCID901 | CN(C)C(=O | YFGYUFNIC | C11H18N4 | 238.143 | | | |
| 16 | 15 | 17-Methylt | DTXSID103 | 58-18-4 | C[C@]1(O) | DTXCID202 | CC1(O)CCC | GCKMFJBG | C20H30O2 | 302.2246 | | | |
| 17 | 16 | Tamoxifen | DTXSID103 | 10540-29-1 | CC\C=C(C | DTXCID901 | CC\C=C(C | NKANXQFJ | C26H29NO | 371.2249 | | | |
| 18 | 17 | Phenyl 1-h | DTXSID103 | 132-54-7 | OC1=C(C=C | DTXCID901 | OC1=C(C=C | QHDYIMW | C17H12O3 | 264.0786 | | | |
| 19 | 18 | Propamoca | DTXSID104 | 24579-73-5 | CCCCC(=O | DTXCID902 | CCCCC(=O | WZZLDXDL | C9H20N2O | 188.1525 | | | |
| 20 | 19 | Clofibric ac | DTXSID104 | 882-09-7 | CC(C)(OC1= | DTXCID902 | CC(C)(OC1= | TXCGAZHT | C10H11ClO | 214.0397 | | | |
| | U | V | W | X | Y | Z | AA | AB | AC | AD | AE | AF | |
| poss_ | poss_IDs | poss_ID | sci_max | Score | n_Score | Gfn_Score | Gfn_Score | Gf | poss_DTXSID | poss_CAS | MoNAScore | RT_RMB | MaxMoNAScore |
| 1 | (Z)-Hexade | 3 | | 3 | | 0 | 1 | 1 | DTXSID004 | 373-49-9 | 0 | 19.125 | 0 |
| 1 | cis-1,2,3,6- | 3 | | 3 | | 0 | 1 | 1 | DTXSID004 | 1469-48-3 | 0 | 14.298 | 0 |
| 1 | Monopotas | 3 | | 3 | | 0 | 1 | 1 | DTXSID004 | 877-24-7 | 0 | 13.335 | 0 |
| 1 | Octylparab | 3 | | 3 | | 0 | 1 | 1 | DTXSID004 | 1219-38-1 | 0 | 12.575 | 0 |
| 1 | 2-Amino-5- | 3 | | 3 | | 0 | 1 | 1 | DTXSID102 | 121-88-0 | 0 | 13.364 | 0 |
| 1 | 4-Chloroph | 1 | | 1 | | 0 | 0 | 0 | DTXSID102 | 106-48-9 | 0 | | 0 |
| 1 | Stavudine | 3 | | 3 | | 0 | 1 | 1 | DTXSID102 | 3056-17-5 | 0 | 20.223 | 0 |
| 1 | Pirimicarb | 3.99747 | | 3.99747 | | 1 | 1 | 1 | DTXSID103 | 23103-98-2 | 0.99747 | 14.617 | 0.99747 |
| 1 | 17-Methylt | 3 | | 3 | | 0 | 1 | 1 | DTXSID103 | 58-18-4 | 0 | 20.368 | 0 |
| 1 | Tamoxifen | 3 | | 3 | | 0 | 1 | 1 | DTXSID103 | 10540-29-1 | 0 | 17.766 | 0 |
| 1 | Phenyl 1-h | 3 | | 3 | | 0 | 1 | 1 | DTXSID103 | 132-54-7 | 0 | 20.541 | 0 |
| 1 | Propamoca | 3.97525 | | 3.97525 | | 1 | 1 | 1 | DTXSID104 | 24579-73-5 | 0.97525 | 8.44 | 0.97525 |
| 1 | Clofibric ac | 1 | | 1 | | 0 | 0 | 0 | DTXSID104 | 882-09-7 | 0 | | 0 |
| 1 | 5,7-Dimeth | 3 | | 3 | | 0 | 1 | 1 | DTXSID104 | 487-06-9 | 0 | 17.853 | 0 |
| 1 | 1,3-Cyclohe | 3 | | 3 | | 0 | 1 | 1 | DTXSID104 | 504-02-9 | 0 | 1.929 | 0 |
| 1 | Potassium | 3 | | 3 | | 0 | 1 | 1 | DTXSID104 | 70321-85-6 | 0 | 8.004 | 0 |
| 1 | 4-Methoxy | 3 | | 3 | | 0 | 1 | 1 | DTXSID202 | 102-50-1 | 0 | 8.468 | 0 |
| 1 | Amiloride h | 3.00182 | | 3.00182 | | 0 | 1 | 1 | DTXSID202 | 2016-88-8 | 0.00182 | 10.038 | 0.00182 |
| 1 | Fenarimol | 3 | | 3 | | 0 | 1 | 1 | DTXSID203 | 60168-88-5 | 0 | 19.937 | 0 |
| 1 | Triflumizole | 3 | | 3 | | 0 | 1 | 1 | DTXSID203 | 68694-11-1 | 0 | 20.802 | 0 |
| 1 | CP-457677 | 3 | | 3 | | 0 | 1 | 1 | DTXSID204 | 214535-77 | 0 | 18.548 | 0 |
| 1 | Piperine | 3.76577 | | 3.76577 | | 1 | 1 | 1 | DTXSID302 | 94-62-2 | 0.76577 | 19.88 | 0.76577 |

 MetFragResultSummary_20170106_dd_pos_T2.csv

SummaryOfMetFragSummaries.xlsx

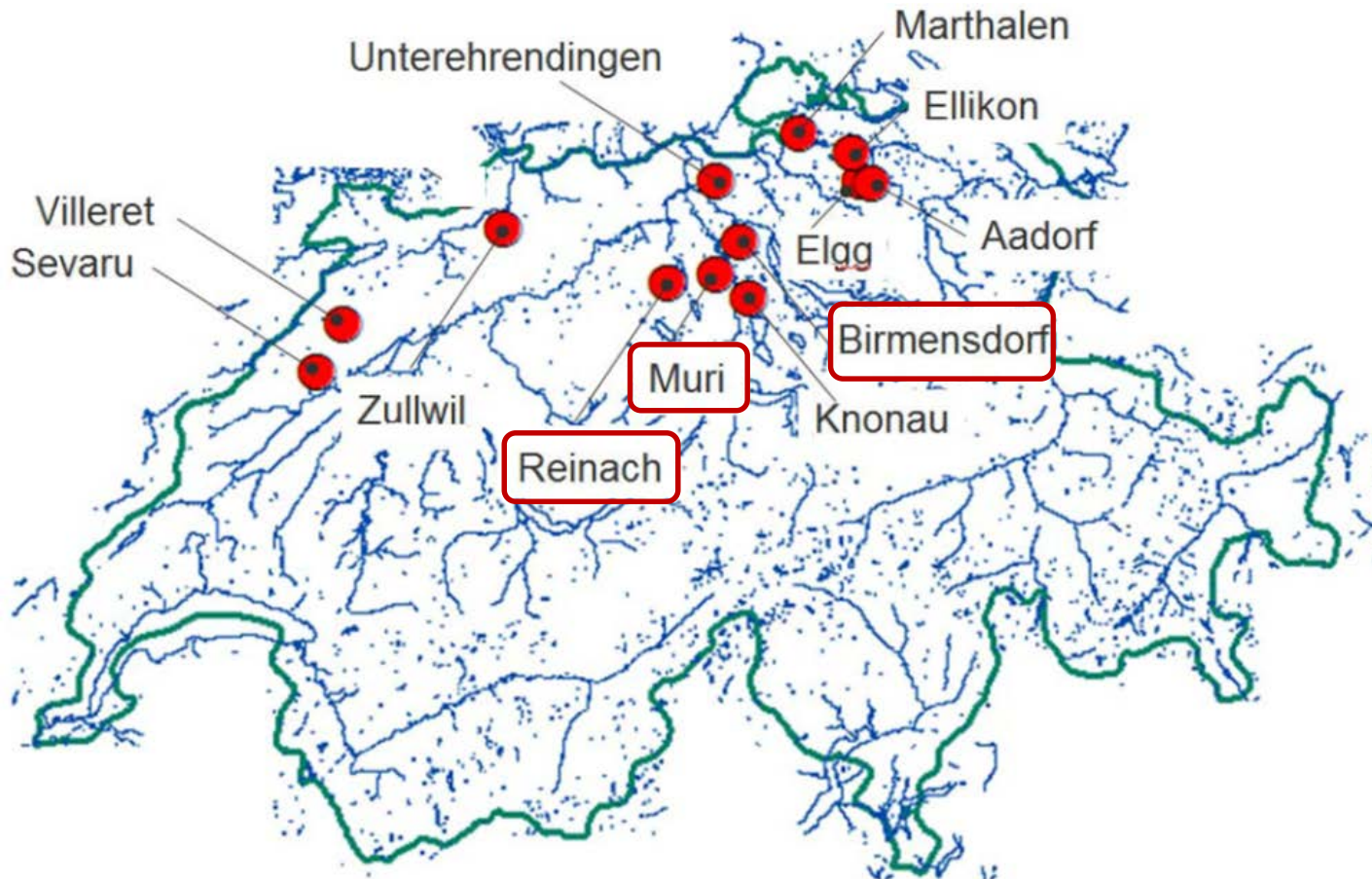


Application to Rhine Case Study

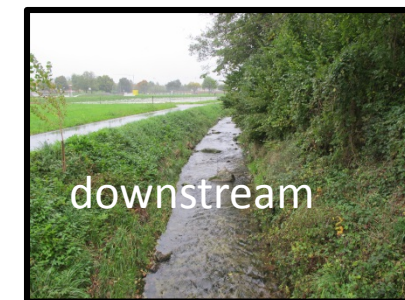
Upstream, effluent and downstream samples from 3 locations

solutions

EcoImpact



MURI





Conversion (**Proteowizard**) and Peak Picking (**enviPick**)

Detection of blank/blind/noise/internal standards; presence upstream/downstream (**enviMass**)

Target List

Suspect Surfactants

NON-TARGET SCREENING

TARGET ANALYSIS

SUSPECT SCREENING

By: Nicole Munz
(**enviMass**,
Trace Finder)

Gather evidence
(**nontarget**,
RMassBank)

Componentization
(**nontarget**)

Top 100 per sample
with isotope/adduct

DDA-MS/MS

159 masses positive
148 masses negative

EIC, MS/MS retrieval
(**RMassBank**)

Most intense MS/MS

Non-target identification
(**MetFrag**, **ReSOLUTION**)

Interpretation, confirmation, peak inventory, confidence and reporting

Reality check
for non-target ID

Presence of LAS
TPs downstream

"Industrial Chemical List" (Stellan Fischer, KEMI)

Exposure Score & Hazard Scores for Suspected chemicals

| Suspected chemical | Charge | Adducts | Observed mass (m/z) | Exposure Score (4-24) | Quantity Index (1-9) | Wide Use Index (1-9) | Release Potential Index (1-9) | Hazard Score EcoAcute (0-1) | Hazard Score EcoChronic (0-1) | Hazard Score HumAcute (0-1) | Hazard Score HumChronic (0-1) |
|--|--------|---------|---------------------|-----------------------|----------------------|----------------------|-------------------------------|-----------------------------|-------------------------------|-----------------------------|-------------------------------|
| 1,2-Benzenedicarboxylic acid, 1,2-diethyl | pos | -H | 223.0965 | 18 | 5.1 | 6.4 | 6.4 | 0.9 | 0.9 | 0.5 | 0.9 |
| Propanoic acid, 2-methyl-, 4-formyl-2-methyl | pos | -H | 223.0965 | 13 | 1 | 3.3 | 9 | 0.1 | 0.1 | 0.2 | 0.1 |
| 2-Propenoic acid, 2-hydroxy-3-phenoxy | pos | -H | 223.0965 | 11 | 6.4 | 3.3 | 1.3 | 0.1 | 0.2 | 0.3 | 0.1 |
| Oxirane, 2,2'-[1,3-phenylenebis(oxy)methoxy] | pos | -H | 223.0965 | 5 | 1.3 | 3.3 | 1.3 | 0.8 | 1 | 0.8 | 0.8 |
| 1,2-Benzenedicarboxylic acid, di-C4-13- | pos | -H | 223.0965 | 4 | 2 | 1 | 1 | 0.2 | 0.3 | 0.2 | 0.3 |
| 1,2-Benzenedicarboxylic acid, mono(2-r | pos | -H | 223.0965 | 4 | 2 | 1 | 1 | 0.3 | 0.3 | 0.3 | 0.3 |
| 1,3-Benzenedicarboxylic acid, 1,3-diethyl | pos | -H | 223.0965 | 4 | 2 | 1 | 1 | | | | |
| 1,4-Benzenedicarboxylic acid, 1,4-diethyl | pos | -H | 223.0965 | 4 | 1 | 1.8 | 1.8 | 0.1 | 0.1 | 0.2 | 0.1 |
| Benzoic acid, 2-hydroxy-3-methoxy-5-(2 | pos | -H | 223.0965 | 4 | 2 | 1 | 1 | | | | |

UVCB: "Unknown or Variable composition, Complex reaction products or Biological materials"

m/z values also from UVCB components

Substances with hazard scores

Upstream, effluent and downstream samples from 3 locations

- Top 100 masses with isotope (+/-) and adducts (+) from every sample taken
 - Positive: 159 masses ($<m/z$ 460) \Rightarrow 92 MS/MS
 - Negative: 148 masses ($<m/z$ 500) \Rightarrow 90 MS/MS
- Exposure and Hazard Screening (Stellan Fischer, KEMI)
 - 199 “hits” in the KEMI list \Rightarrow 149 chemicals \Rightarrow 54 masses (35 +, 19 -)
 - 28 positive and 16 negative masses with MS/MS

| | Total | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 |
|----------|-------|-----------|----------|-----------|---------|----------|
| | | Confirmed | Probable | Tentative | Formula | Ex. Mass |
| Positive | 28 | 2 | 2* | 19 | 3 | - |
| Negative | 16 | 3 | 3 | 8 | 1 | 1 |

*includes case where another isomer is possible

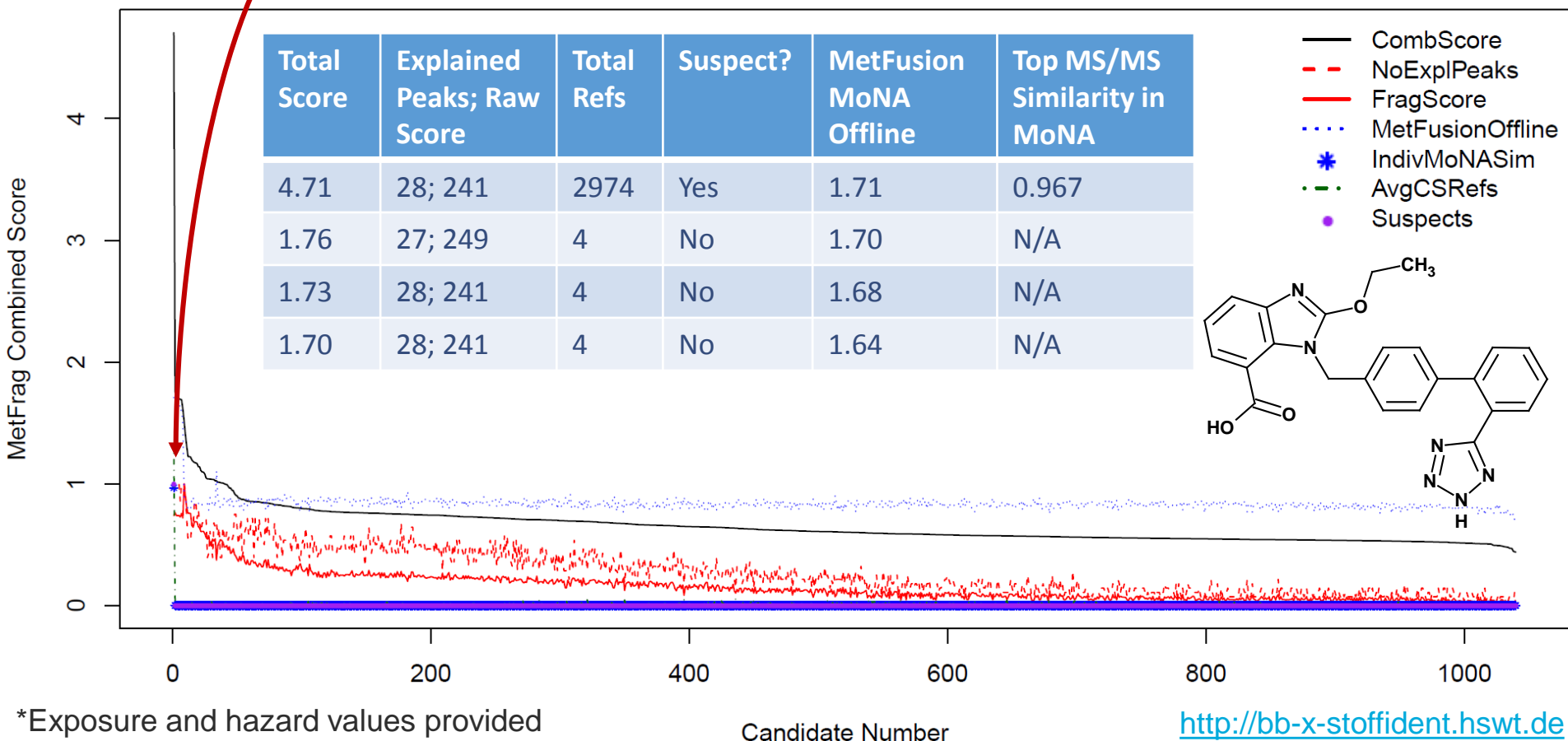
Proof of Concept: Candesartan

Combined and individual scores in MetFrag

“Exposure Score” 11 (of max. 27)*

“Hazard Score Sum” 0.75 (of max. 4)*

$\Delta RT(\text{online}) = 0.15 \text{ min}$



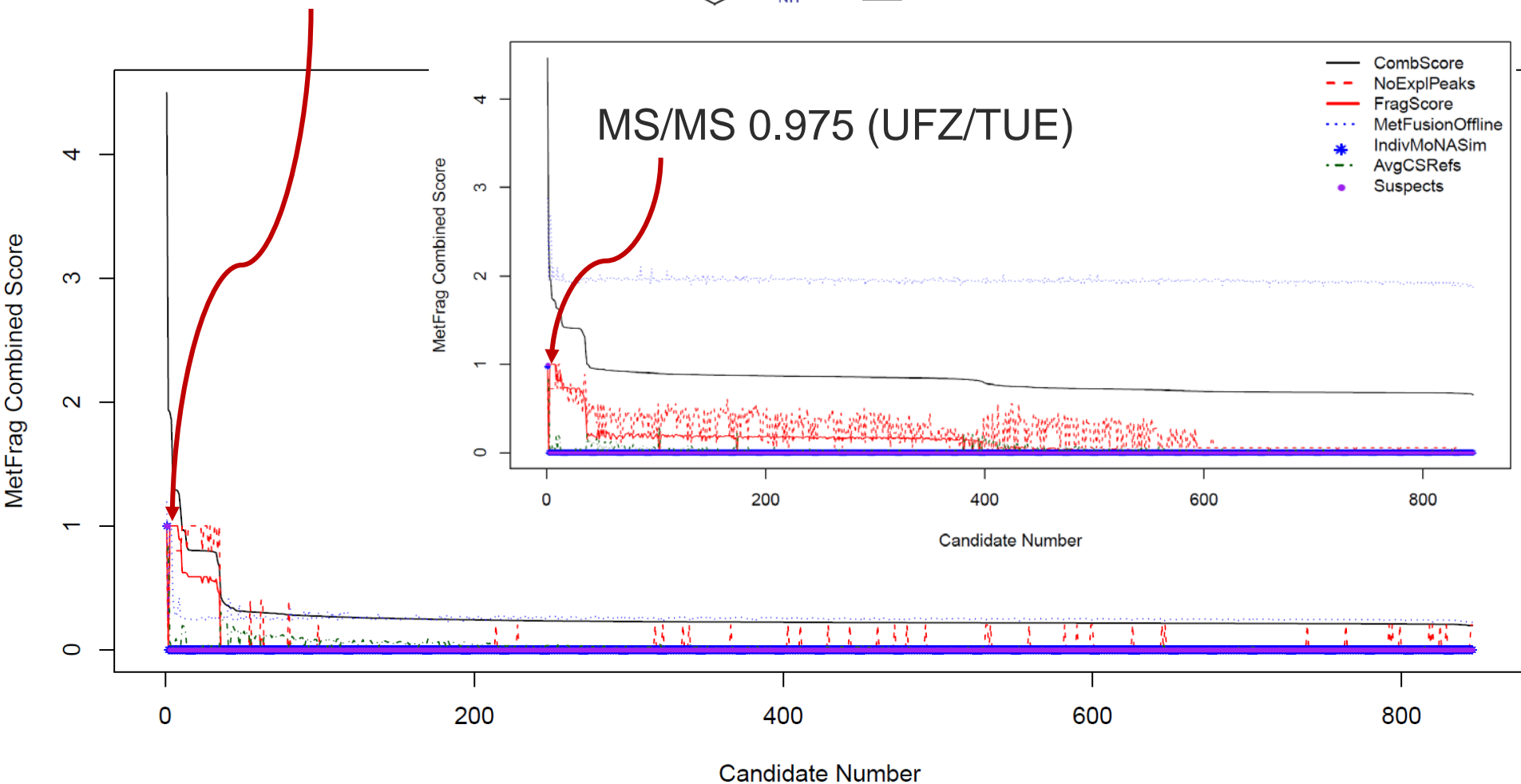
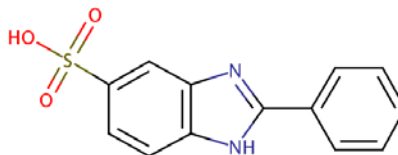
*Exposure and hazard values provided by Stellan Fischer, KEMI, Sweden

Candidate Number

<http://bb-x-stoffident.hswt.de>
<http://mona.fiehnlab.ucdavis.edu>

Non-target with Ref. Std. at Partner

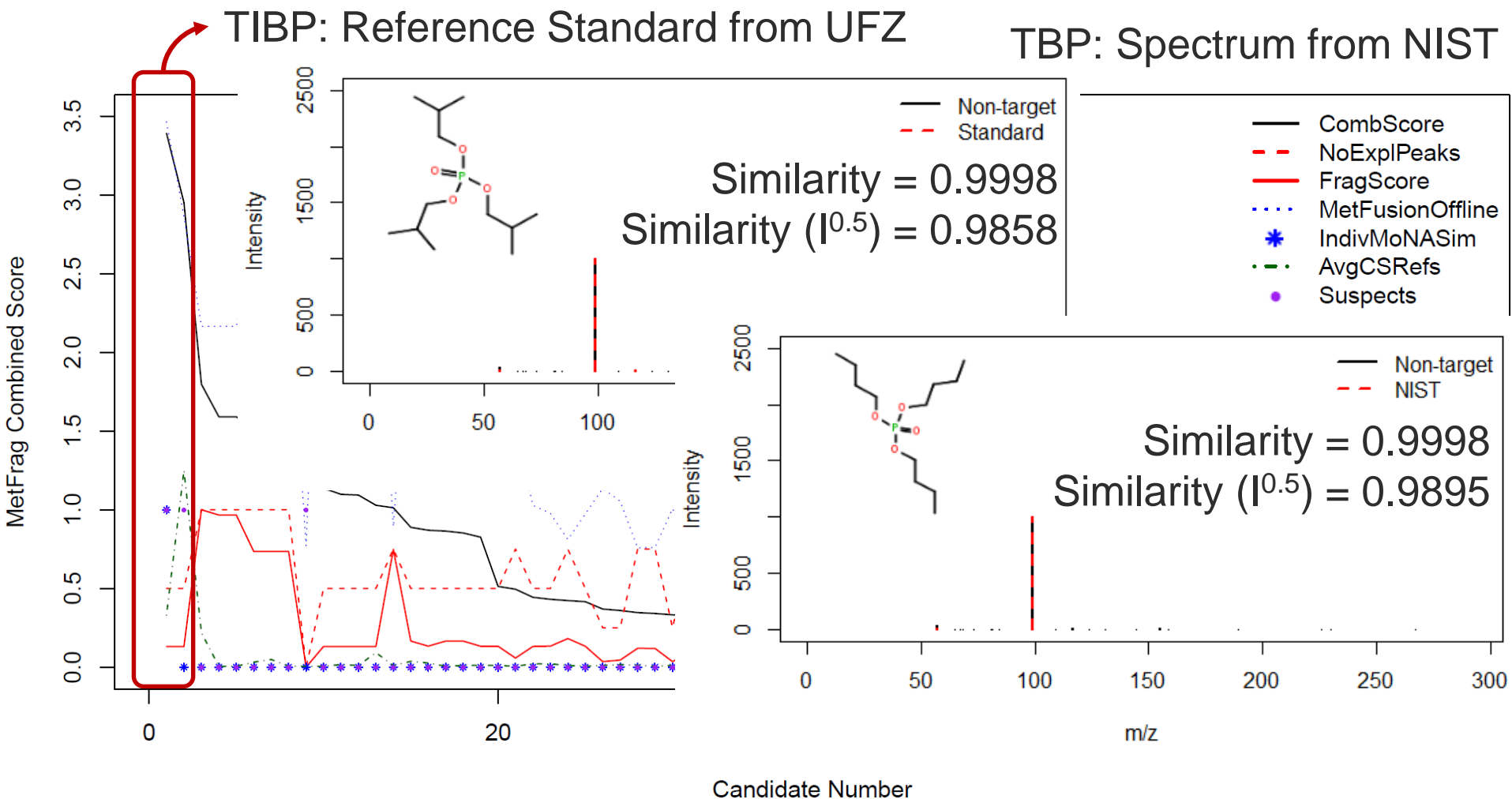
MS/MS 0.999 (UFZ/TUE)
Expo 14, Hazard N/A



Non-target m/z $[M+H]^+ = 267.1719$

TIBP(CAS 126-71-6): Expo Score 19; Hazard Score Sum 0.72

TBP (CAS 126-73-8): Expo Score 19; Hazard Score Sum: 2.2



Non-target screening made easier...

...but not always easy...

Exposure and toxicity estimations

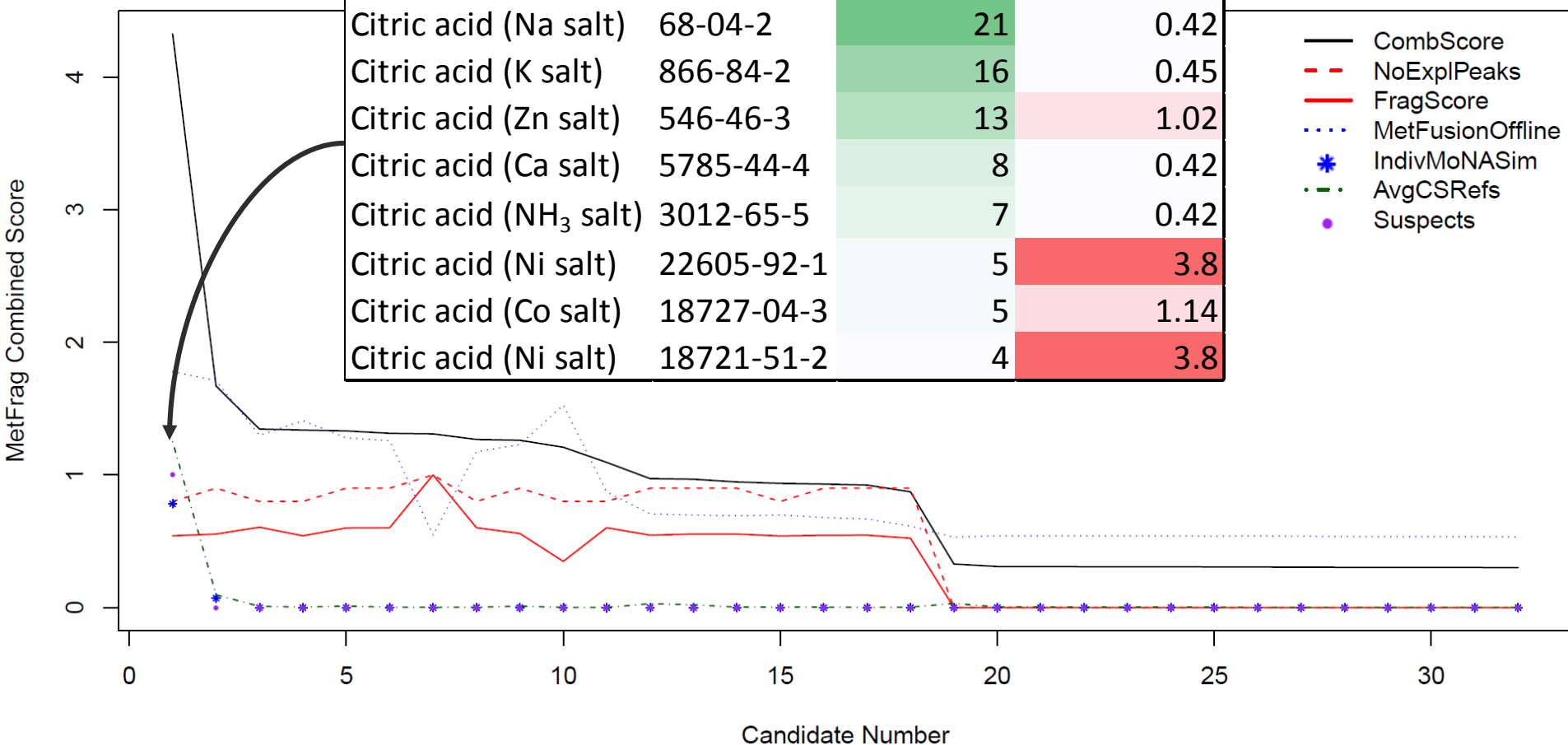
What is driving the toxicity?

m/z [M-H]⁻

191.0195

Citric acid

| Name | CAS | Expo Score | Hazard Score |
|------------------------------------|------------|------------|--------------|
| Citric acid | 77-92-9 | 23 | 0.48 |
| Citric acid (Na salt) | 68-04-2 | 21 | 0.42 |
| Citric acid (K salt) | 866-84-2 | 16 | 0.45 |
| Citric acid (Zn salt) | 546-46-3 | 13 | 1.02 |
| Citric acid (Ca salt) | 5785-44-4 | 8 | 0.42 |
| Citric acid (NH ₃ salt) | 3012-65-5 | 7 | 0.42 |
| Citric acid (Ni salt) | 22605-92-1 | 5 | 3.8 |
| Citric acid (Co salt) | 18727-04-3 | 5 | 1.14 |
| Citric acid (Ni salt) | 18721-51-2 | 4 | 3.8 |

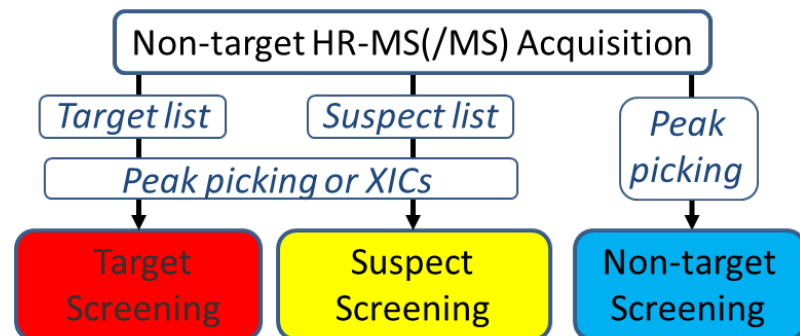


Identifying Small Molecules in Untargeted Data:

A case study of Thirdhand Smoke (THS)



+



=

| ID | mz | Name | RT | Int | MS/MS | RT_RMB | #Cand | MaxScore | SMILES | Name_maxScore | max | max |
|------|----------|--------|-------|-------|-------|--------|-------|----------|---------|----------------------------|-----------|---------|
| 3027 | 163.1221 | FT0532 | 1.073 | 9E+07 | TRUE | 1.052 | 137 | 7.270015 | CN1CCCC | Nicotine | 58.0658_9 | 0.99701 |
| 3008 | | | | +07 | TRUE | 1.01 | 41 | 7.22864 | CN1CC(= | Creatinine | 57.0454_1 | 0.54699 |
| 3131 | | | | +06 | TRUE | 19.102 | 71 | 7.028138 | CC(C=CC | all-trans-Retinoic acid | 57.0706_5 | 0.10337 |
| 3321 | | | | +07 | TRUE | 26.205 | 25 | 6.093 | CCCCCCC | Didecyl phthalate | 69.0704_2 | 0 |
| 3484 | | | | +06 | TRUE | 24.088 | 1 | 6 | CC(C)(C | 2,2'-Oxamidodiethyl bis[3- | 57.0706_3 | 0 |
| 3206 | 346.1096 | FT3193 | 21.08 | 3E+06 | TRUE | 21.144 | 68 | 5.973483 | CCCN(CC | Nitralin | NA | 0 |
| 3044 | | | | +07 | TRUE | 1.073 | 253 | 5.960103 | CCN(CC) | N,N-Diethylnicotinamide | 78.0342_4 | 0.01075 |
| 3006 | | | | +06 | TRUE | 1.287 | 20 | 5.94264 | OCCOCC | Diethylene glycol | NA | 0 |
| 3043 | | | | +07 | TRUE | 3.163 | 253 | 5.721703 | CCN(CC) | N,N-Diethylnicotinamide | 84.0811_1 | 0.00209 |
| 3055 | | | | +07 | TRUE | 11.202 | 114 | 5.712626 | COC1=C | Scopoletin | 53.0393_1 | 0.94271 |
| 3046 | 183.0796 | FT0741 | 24.77 | 1E+07 | TRUE | 24.744 | 72 | 5.689798 | O=C(C1= | Benzophenone | 50.0159_2 | 0 |
| 3039 | | | | +07 | TRUE | 7.635 | 219 | 5.467659 | CN1C(CC | Cotinine | 53.0393_1 | 0.9987 |
| 3038 | | | | +07 | TRUE | 8.017 | 219 | 5.354236 | CC1CN(N | 4-Methyl-1-phenylpyrazol | 53.0393_2 | 0.66408 |
| 3183 | | | | +07 | TRUE | 19.655 | 74 | 4.928842 | CCCCCCC | Dihexyl phthalate | 54.0452_2 | 0 |
| 3095 | | | | +08 | TRUE | 21.548 | 74 | 4.860782 | CC(C)CO | Triisobutyl phosphate | 57.0707_8 | 0.99613 |
| 3020 | 151.1111 | FT0427 | 13.16 | 2E+06 | FALSE | | 307 | 3.943886 | CC(C)(C | 4-tert-Butylphenol | NA | 0 |
| 3029 | | | | +05 | FALSE | | 242 | 3.021459 | COC1=CC | 1,2,4-Trimethoxybenzene | NA | 0 |
| 3123 | | | | +06 | FALSE | | 18 | 3 | CCCCCCC | Octadecyl isocyanate | NA | 0 |
| 3128 | | | | +06 | FALSE | | 13 | 3 | CCCCCCC | Octadecanoic acid, hydraz | NA | 0 |
| 3172 | | | | +06 | FALSE | | 15 | 3 | CCCCCCC | Stearylbenzene | NA | 0 |

Sampling



Extraction (AcN+
QuEChERS+ Z-Sep)



UHPLC separation



HR-MS/MS



Conversion (Proteowizard)



TSNA+ Target List



THS Suspect List



NON-TARGET SCREENING



SUSPECT SCREENING



Peak Picking (xcms)



Prioritization (xcms)

Masses of interest



Masses of interest



Statistics (in house)



RMassBank

MS/MS Extraction
(RMassBank)

RMassBank



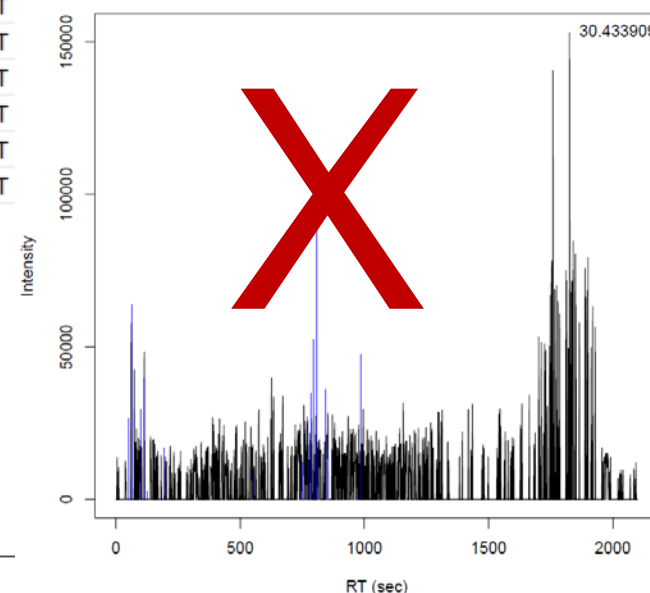
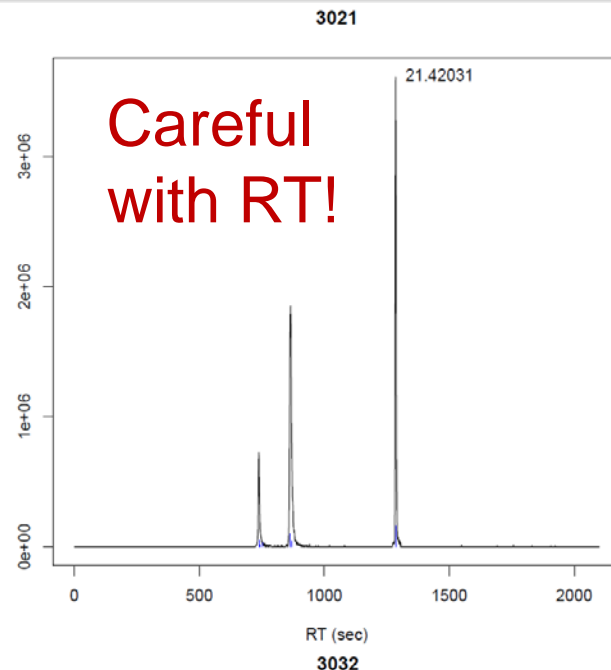
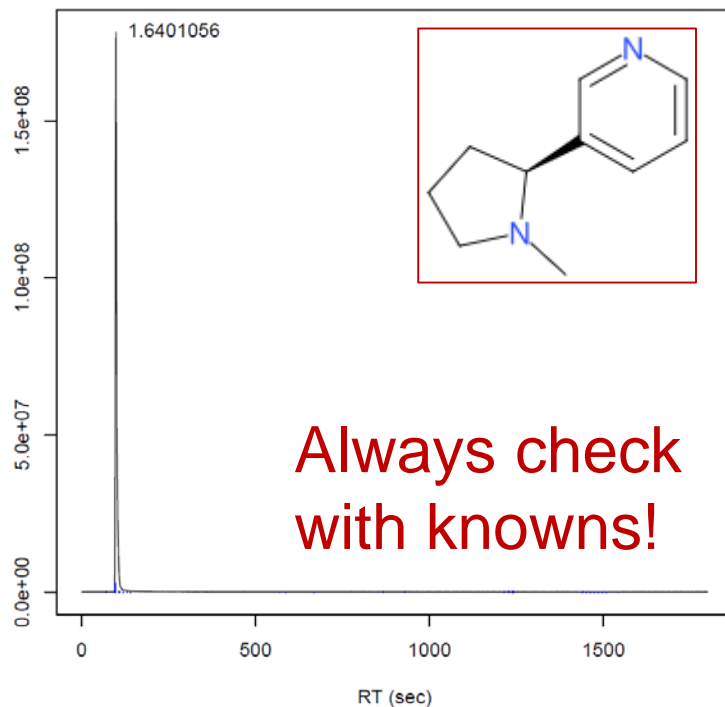
Non-target identification
(MetFrag, ReSOLUTION)



Interpretation, confirmation, peak inventory, confidence and reporting

MS/MS Extraction with RMassBank

| A | B | C | D | E | F | G | H | I | |
|------|---|-----------|----------------------------|-------------|-------|------------|---------|----|-----|
| ID | Name | Name2 | DTXSID | SMILES | CAS | RT | Formula | mz | Std |
| 112 | Nicotine (3 Nicotine | DTXSID102 | CN1CCC[C@H]1N | 54-11-5 | 1.62 | C10H14N2 | Std | | |
| 114 | Nicotine-d ₃ DL-Nicotine | DTXSID804 | [2H]C([2H])[2H]1CCC[C@H]1N | 69980-24-1 | 1.62 | C10H11D3N2 | Std | | |
| 117 | Cotinine (1 Cotinine | DTXSID104 | CN1[C@@H]1CCC(=O)N | 486-56-6 | 1.77 | C10H12N2O | Std | | |
| 119 | NNN (N'-NN'-Nitrosor | DTXSID402 | O=NN1CCC[C@H]1N | 16543-55-8 | 1.79 | C9H11N3O | Std | | |
| 122 | Cotinine-d ₃ Cotinine-d ₃ | DTXSID404 | [2H]C([2H])[2H]1CCC(=O)N | 110952-70-0 | 1.77 | C10H9D3N2O | Std | | |
| 134 | NNK (4-(Methyl-4-(N-Methy | DTXSID302 | CN(CCCC(=O)N)C1=CC=CC=C1 | 64091-91-4 | 11.33 | C10H13N3O2 | Std | | |
| 135 | NNA (4-(methyl-4-(methyln | DTXSID008 | CN(N=O)C1=CC=CC=C1 | 64091-90-3 | 1.77 | C10H13N3O2 | Std | | |
| 136 | NNAL (4-(methyl-4-(methyln | DTXSID802 | CN(CCCC(=O)N)C1=CC=CC=C1 | 76014-81-8 | 1.77 | C10H15N3O2 | Std | | |
| 157 | isoNNAL | | CN(C(CCCO)C1=CC=CC=C1 | | 1.77 | C10H15N3O2 | Std | | |
| 158 | NNN | | | 112 | | 7D4N3O | Std | | |
| 159 | NNA | | | | | H10D3N3O2 | Std | | |
| 160 | NNA | | | | | H12D3N3O2 | Std | | |
| 1001 | FTO | | | | | 60.04523 | NT | | |
| 1002 | FTO | | | | | 61.04228 | NT | | |
| 1003 | FTO | | | | | 90.97725 | NT | | |
| 1004 | FTO | | | | | 104.1072 | NT | | |
| 1005 | FTO | | | | | 109.1013 | NT | | |
| 1006 | FTO | | | | | 111.0216 | NT | | |
| 1007 | FTO | | | | | 114.0666 | NT | | |
| 1008 | FTO | | | | | 121.0398 | NT | | |
| 1009 | FTO | | | | | 125.0392 | NT | | |
| 1010 | FTO | | | | | 125.0959 | NT | | |



- Create config files (MetFragConfig)
- Run MetFrag (runMetFrag)

```
176 # now, run MetFrag and extract results for reporting into compd_info.
177
178 results_filename <- paste0(run_name,"_",compdID_char,"_",as.character(i))
179 if (isPos) {
180     config_file <- MetFragConfig(mass = ExactMass, adduct_type = 0, neutralPrecursorMass=TRUE,
181                                 results_filename = results_filename,
182                                 peaklist_path = MetFrag_msms, base_dir = results_run_dir,
183                                 DB = "LocalCSV", localDB_path=localDB,useMonaIndiv = T,useMoNAMetFusion = T,
184                                 IsPosMode = TRUE,filter_by_InChIKey = F,rt_file_path=MetFrag_rt_file,rt_exp=RT)
185 } else {
186     config_file <- MetFragConfig(mass = ExactMass, adduct_type = 0, neutralPrecursorMass=TRUE,
187                                 results_filename = results_filename,
188                                 peaklist_path = MetFrag_msms, base_dir = results_run_dir,
189                                 DB = "LocalCSV", localDB_path=localDB,useMonaIndiv = T,useMoNAMetFusion = T,
190                                 IsPosMode = FALSE,filter_by_InChIKey = F,rt_file_path=MetFrag_rt_file,rt_exp=RT)
191 }
192
193 runMetFrag(config_file, MetFrag_dir, CL_name = "MetFrag2.4.4-msready-CL.jar")
194
195 results_file <- paste0(results_run_dir,"/results/",results_filename,".xls")
```

○ Extract results and summarize

```
198 #extract results we need:
199 MetFrag_res <- read_excel(results_file)
200
201 compd_info$num_poss_IDs[i] <- length(MetFrag_res$Score)
202 compd_info$poss_IDs[i] <- paste(MetFrag_res$Name,collapse=";")
203 compd_info$poss_ID_scores[i] <- paste(MetFrag_res$Score,collapse=";")
204 compd_info$max_Score[i] <- max(MetFrag_res$Score)
205 compd_info$n_Score_GE3p5[i] <- length(which(MetFrag_res$Score>=3.5))
206 compd_info$n_Score_GE3[i] <- length(which(MetFrag_res$Score>=3))
207 compd_info$n_Score_GE2p5[i] <- length(which(MetFrag_res$Score>=2.5))
208 compd_info$poss_DTXSIDs[i] <- paste(MetFrag_res$DTXSID,collapse=";")
209 compd_info$poss_CAS[i] <- paste(MetFrag_res$CAS,collapse=";")
210 compd_info$MoNAScore[i] <- paste(MetFrag_res$OfflineIndividualMoNAScore,collapse=";")
211 compd_info$MaxMoNAScore[i] <- max(MetFrag_res$OfflineIndividualMoNAScore)
212
213 }
214
215
216 write.csv(compd_info,paste0(results_summary_dir,"/MetFragResultSummary_",run_name,".csv"),row.names = F)
```


CSV Summary Output for Results Interrogation

| ID | mz | Name | RT | Int | MS/MS | RT_RMB | #Cand | MaxScore | SMILES | Name_maxScore | ExplPeaks | max #Peaks | max #Peaks | max MoNA |
|------|-------------------------|--------|-------|-------|-------|--------|-------|----------|------------|----------------------------|-----------|---------------|---------------|-------------|
| 3027 | 163.1221 | FT0532 | 1.073 | 9E+07 | TRUE | 1.052 | 137 | 7.270015 | CN1CCCC | Nicotine | 58.0658_9 | 9 | 28 | 0.99701 |
| 3008 | Level 1 Target | | | +07 | TRUE | 1.01 | 41 | 7.22864 | CN1CC(= | Creatinine | 57.0454_1 | 3 | 14 | 0.54699 |
| 3131 | | | | +06 | TRUE | 19.102 | 71 | 7.028138 | CC(C=CC | all-trans-Retinoic acid | 57.0706_5 | 8 | 39 | 0.10337 |
| 3321 | | | | +07 | TRUE | 26.205 | 25 | 6.093 | CCCCCCC | Didecyl phthalate | 69.0704_2 | 22 | 71 | 0 |
| 3484 | | | | +06 | TRUE | 24.088 | 1 | | 6 CC(C)(C) | 2,2'-Oxamidodiethyl bis[3- | 57.0706_3 | 4 | 42 | 0 |
| 3206 | 346.1096 | FT3193 | 21.08 | 3E+06 | TRUE | 21.144 | 68 | 5.973483 | CCCN(CC | Nitralin | NA | 9 | 51 | 0 |
| 3044 | Level 2 MSMS Match | | | +07 | TRUE | 1.073 | 253 | 5.960103 | CCN(CC) | N,N-Diethylnicotinamide | 78.0342_4 | 9 | 44 | 0.01075 |
| 3006 | | | | +06 | TRUE | 1.287 | 20 | 5.94264 | OCCOCC | Diethylene glycol | NA | 0 | 8 | 0 |
| 3043 | | | | +07 | TRUE | 3.163 | 253 | 5.721703 | CCN(CC) | N,N-Diethylnicotinamide | 84.0811_1 | 7 | 24 | 0.00209 |
| 3055 | | | | +07 | TRUE | 11.202 | 114 | 5.712626 | COC1=C | Scopoletin | 53.0393_1 | 9 | 28 | 0.94271 |
| 3046 | 183.0796 | FT0741 | 24.77 | 1E+07 | TRUE | 24.744 | 72 | 5.689798 | O=C(C1= | Benzophenone | 50.0159_2 | 5 | 45 | 0 |
| 3039 | Level 3 Tentative ID | | | +07 | TRUE | 7.635 | 219 | 5.467659 | CN1C(CC | Cotinine | 53.0393_1 | 9 | 23 | 0.9987 |
| 3038 | | | | +07 | TRUE | 8.017 | 219 | 5.354236 | CC1CN(N | 4-Methyl-1-phenylpyrazoli | 53.0393_2 | 9 | 40 | 0.66408 |
| 3183 | | | | +07 | TRUE | 19.655 | 74 | 4.928842 | CCCCCCC | Dihexyl phthalate | 54.0452_2 | 9 | 51 | 0 |
| 3095 | | | | +08 | TRUE | 21.548 | 74 | 4.860782 | CC(C)CO | Triisobutyl phosphate | 57.0707_8 | 5 | 15 | 0.99613 |
| 3020 | 151.1111 | FT0427 | 13.16 | 2E+06 | FALSE | | 307 | 3.943886 | CC(C)(C) | 4-tert-Butylphenol | NA | 0 | 0 | 0 |
| 3029 | Level 5 No MS/MS | | | +05 | FALSE | | 242 | 3.021459 | COC1=CC | 1,2,4-Trimethoxybenzene | NA | 0 | 0 | 0 |
| 3123 | | | | +06 | FALSE | | 18 | | 3 CCCCCC | Octadecyl isocyanate | NA | 0 | 0 | 0 |
| 3128 | | | | +06 | FALSE | | 13 | | 3 CCCCCC | Octadecanoic acid, hydraz | NA | 0 | 0 | 0 |
| 3172 | | | | +06 | FALSE | | 15 | | 3 CCCCCC | Stearylbenzene | NA | 0 | 0 | 0 |

- ⇒ Approaching automatic assignment of confidence levels
- ⇒ Quick, high throughput prioritization for data reacquisition

Sampling



Extraction (AcN+
QuEChERS+ Z-Sep)



UHPLC separation



HR-MS/MS



Conversion (Proteowizard)



TSNA+ Target List



THS Suspect List



NON-TARGET SCREENING



SUSPECT SCREENING



Peak Picking (xcms)



Prioritization (xcms)

Masses of interest



Masses of interest



Statistics (in house)

RMassBank

MS/MS Extraction
(RMassBank)

RMassBank



Non-target identification
(MetFrag, ReSOLUTION)

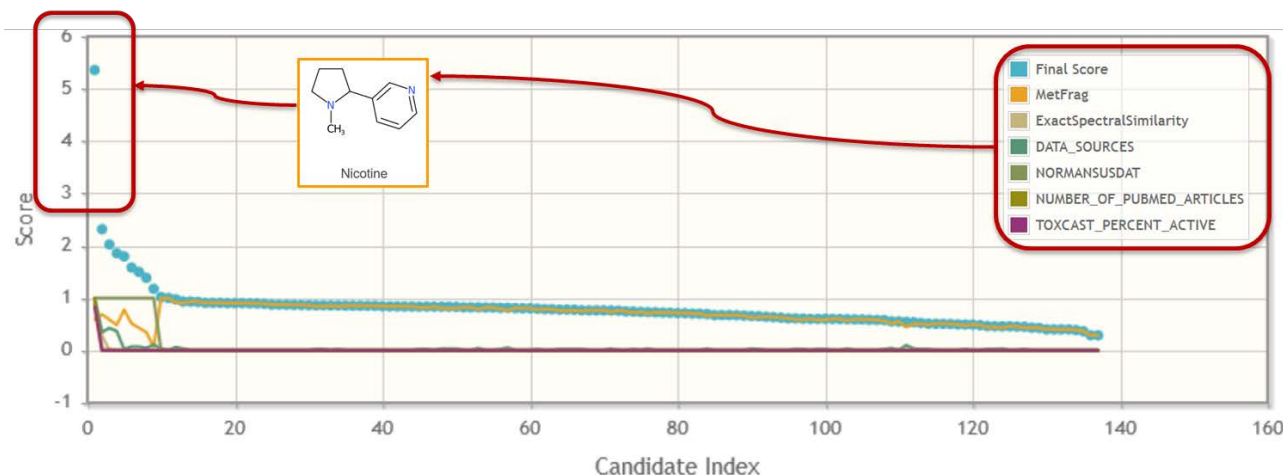
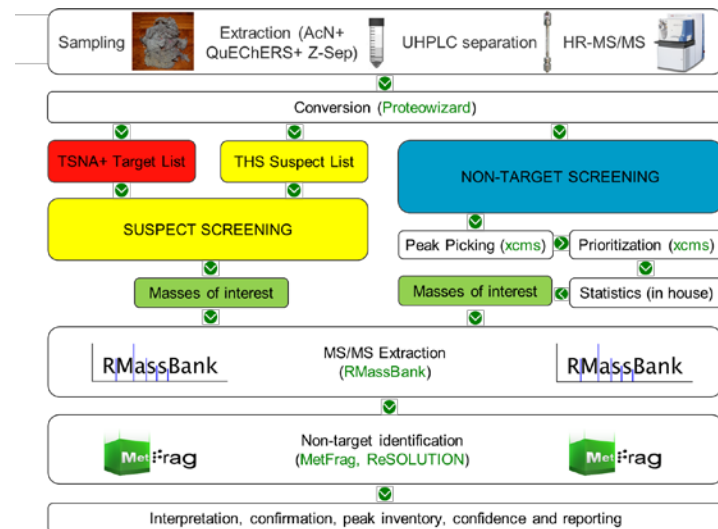


Interpretation, confirmation, peak inventory, confidence and reporting



Perspectives: Identifying Small Molecules in NTS

- Many comprehensive workflows
 - *I have presented just one of them!*
- Annotation of “known unknowns” is now relatively “quick”:
 - *Especially with well-chosen suspect screening and metadata*



- The bottleneck is still in expt. design and interpretation
 - *But in the meantime we can do some pretty neat things!*

Acknowledgements



THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate
and Sustainable Development



UNIVERSITAT
ROVIRA I VIRGILI



UNIVERSITY OF AMSTERDAM



National and Kapodistrian
UNIVERSITY OF ATHENS

solutions



EU Grant
603437



MEDIZINISCHE
UNIVERSITÄT
INNSBRUCK



emma.schymanski@uni.lu

@ESchymanski

Further Information:

<https://massbank.eu/MassBank/>

<https://ipb-halle.github.io/MetFrag/>

<https://www.norman-network.com/nds/SLE/>

[https://www.en.uni.lu/lcsb/research/
environmental_cheminformatics](https://www.en.uni.lu/lcsb/research/environmental_cheminformatics)

