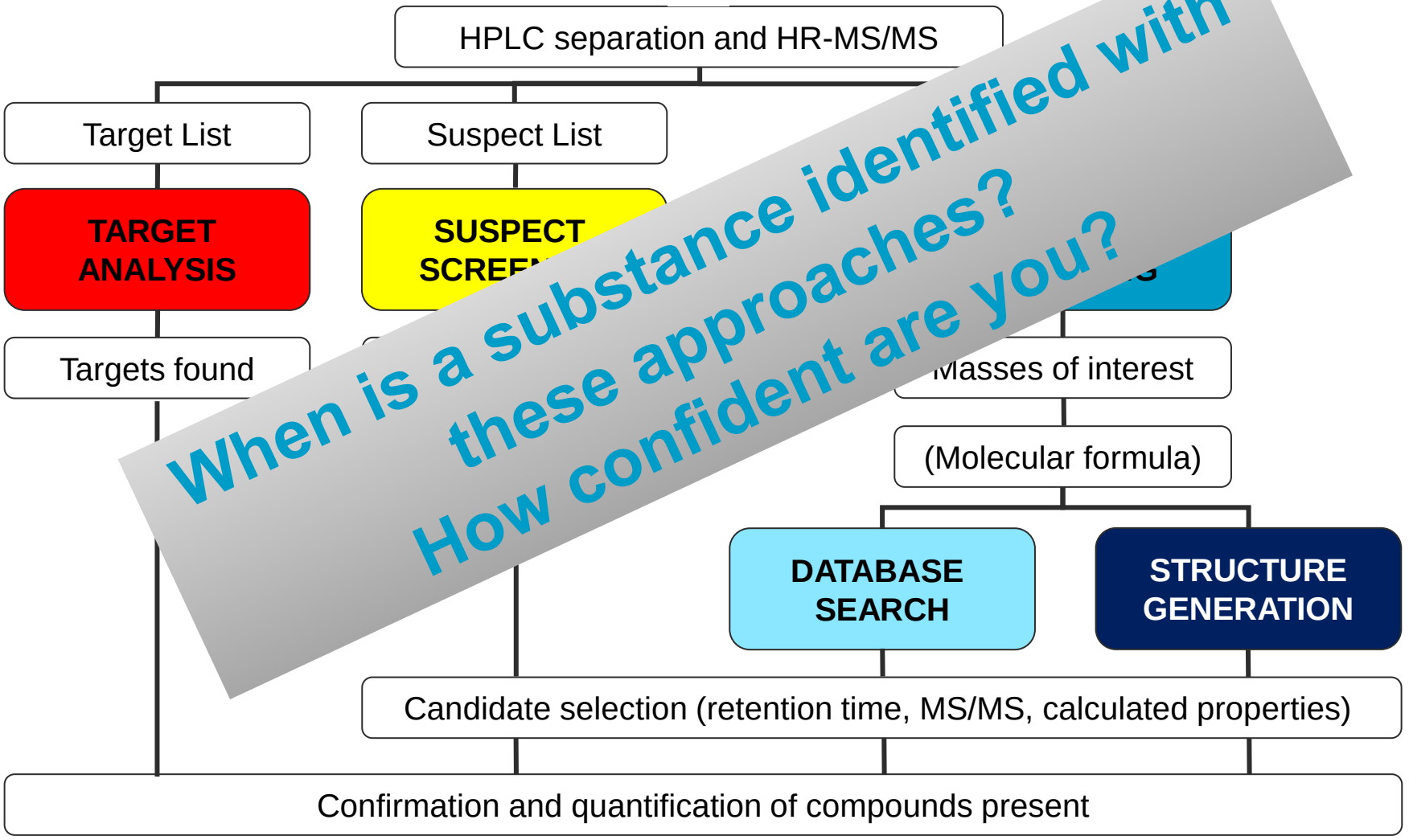
Non-Target Screening: Spectral Information and Identification Confidence

NORMAN SOLUTIONS Workshop on Non-Target Screening; 16-17 Sept. 2014

Emma Schymanski
Juliane Hollender, Heinz Singer und Environmental Chemistry Dept.
Eawag: Swiss Federal Institute of Aquatic Science and Technology
Dübendorf, Switzerland

Status quo of identification *approaches*



Identification levels in metabolomics

What is the confidence in the identification?

Metabolomics Standards Initiative (2007)

Level	Name	Minimum Requirements
1	Identified compounds	At least two independent and orthogonal data relative to an authentic compound under identical conditions
2	Putatively annotated compounds	Similar to level 1, but based on literature values reported for authentic samples by other laboratories
3	Putatively characterised compound class	Based upon characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class
4	Unknown compounds	These metabolites can still be differentiated and quantified based upon spectral data

The problem: large gaps between the levels; many cases not covered

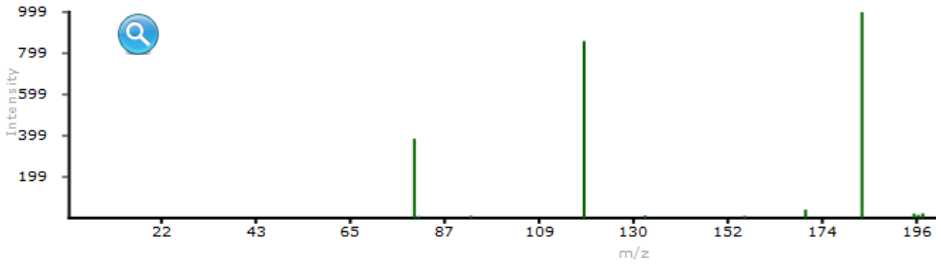
Sumner *et al.* 2007, Metabolomics, 3(3), 211-221; Neumann & Böcker 2010, Anal. Bioanal. Chem. 398, 2779-2788

Updating identification levels?

It is now time to reassess the current reporting standards for…identification[1]

- Simplicity is an advantage for users
- BUT, insufficient ability to distinguish many special cases
- Range of options:
  - Level system to show how well structure is identified
  - Scoring system to show the evidence behind the identification
  - Combination of both

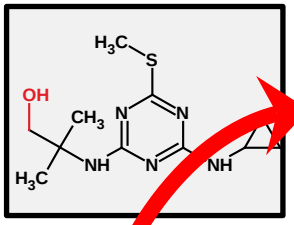
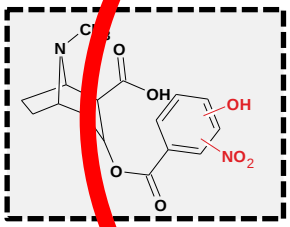
- Starting point for (most) environmental investigations is MS and MS/MS



[1] Creek, Dunn, Fiehn, Griffin, Hall, Lei, Mistrik, Neumann, Schymanski, Sumner, Trengove & Wolfender (2014) "Metabolite identification: are you sure? And how do your peers gauge your confidence?" *Metabolomics*, 10, 350-353.

Identification with (LC-HR)MS/MS – Confidence?

Proposed levels for MS and MS/MS data

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
$C_6H_5N_3O_4$	***Level 4: Unequivocal molecular formula***	MS isotope/adduct
192.0757	***Level 5: Exact mass*** of interest	MS

Schymanski, Jeon, Gulde, Fenner, Ruff, Singer & Hollender (2014) *ES&T*, 48 (4), 2097-2098. DOI: 10.1021/es5002105

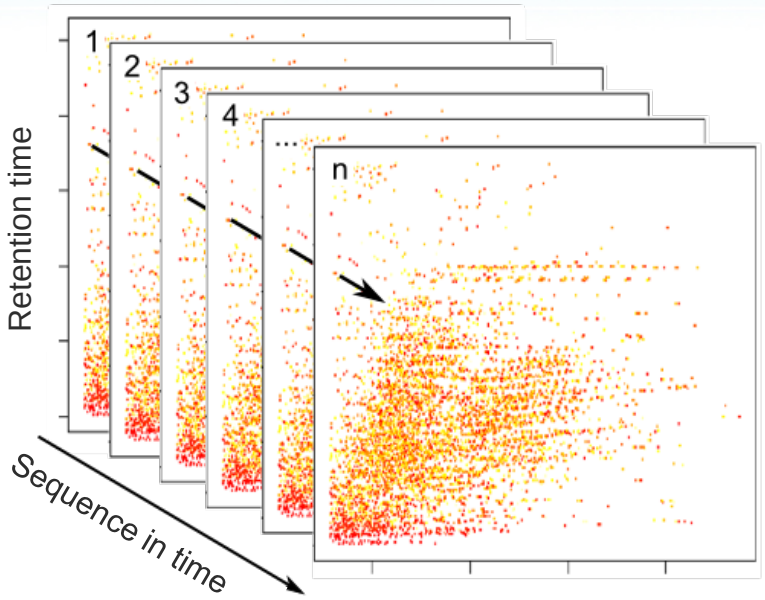
Level 5: Non-target mass of interest

"Unknowns"

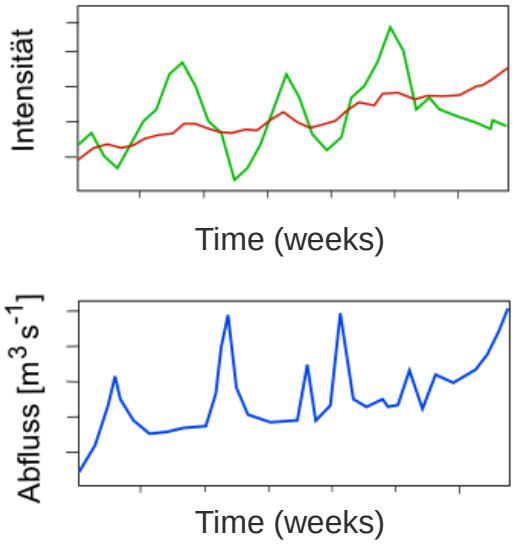
Start point for a non-target identification; various methods for prioritization

- See talks tomorrow morning by D. Stipanicev and M. Ruff

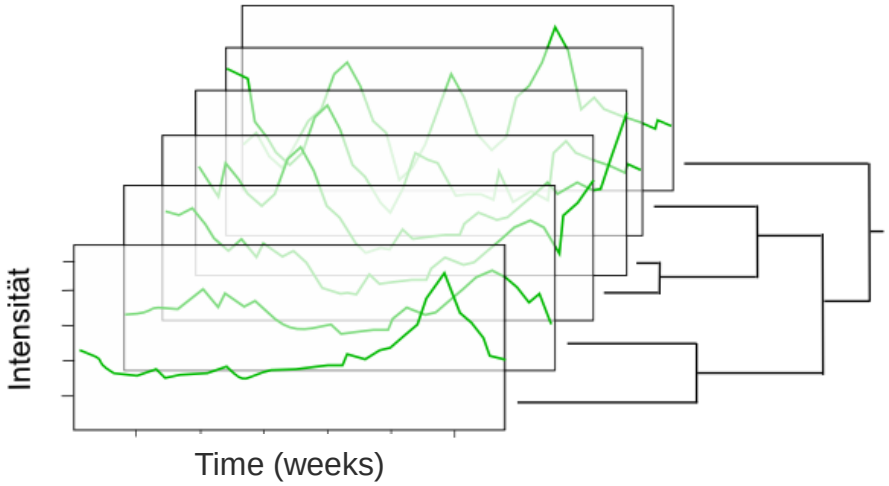
Pictures:
http://www.eawag.ch/forschung/uchem/software/



Trend Analysis



Peak Grouping



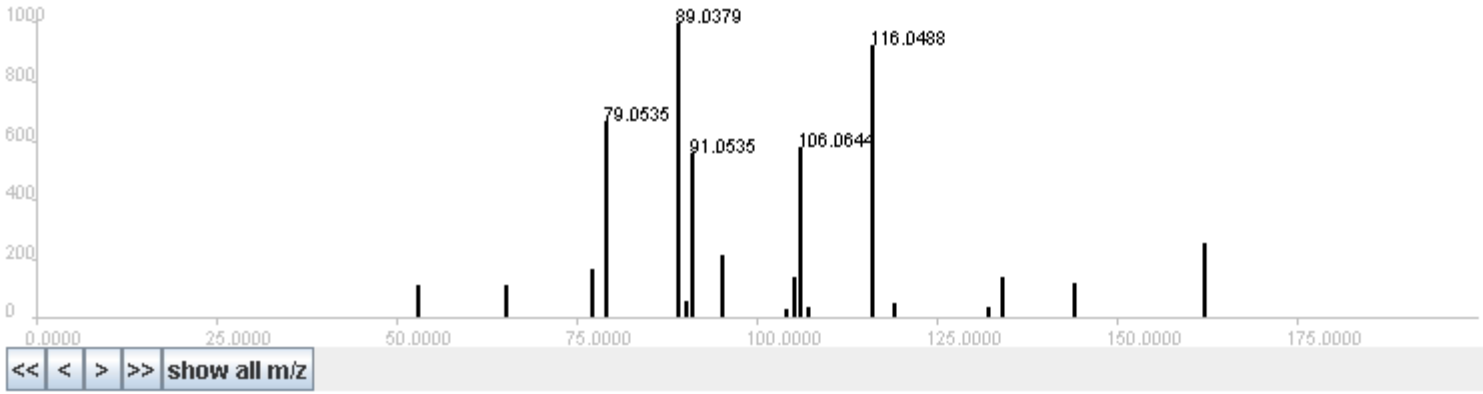
Level 5: Non-target mass of interest

MassBank and “unknown” spectra; exchanging masses of interest

MassBank Record: SMI00152

CASMI2012 LC Challenge 15; LC-APCI-ITFT; MS2; CE:120 HCD;

Mass Spectrum



Chemical Structure

Not Available

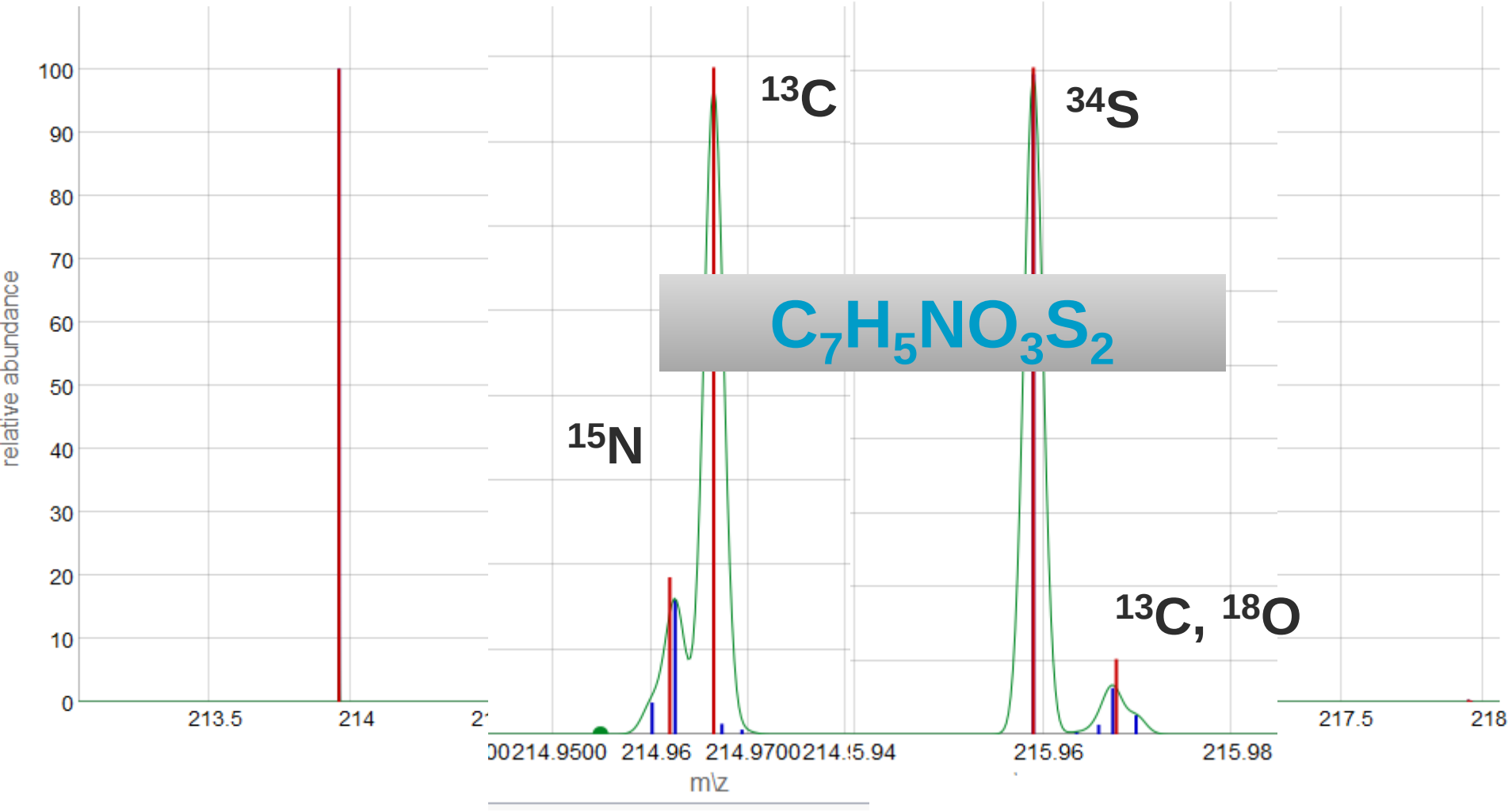
```
ACCESSION: SMI00152
RECORD_TITLE: CASMI2012 LC Challenge 15; LC-APCI-ITFT; MS2; CE:120 HCD;
DATE: 2012.08.31 (Created 2012.08.31)
AUTHORS: S. Neumann: IPB-Halle, Germany & E. Schymanski: Eawag, Switzerland
COPYRIGHT: CASMI2012
COMMENT: Unknown
COMMENT: http://casmi-contest.org/challenges-cat1-2.shtml

CH$NAME: CASMI2012 LC Challenge 15
CH$COMPOUND_CLASS: N/A; Unknown for CASMI
CH$FORMULA: N/A
CH$EXACT_MASS: N/A
CH$SMILES: N/A
CH$IUPAC: N/A
```

E. Schymanski & S. Neumann, 2013, CASMI: And the winner is…, *Metabolites,* DOI: 10.3390/metabo3020412

Level 4: Molecular formula assigned

Adduct state is known, one formula possible



M. Loos, H. Singer, C. Gerber, enviPat: http://www.envipat.eawag.ch/

M. Meringer, MOLGEN-MS/MS, http://www.molgen.de/

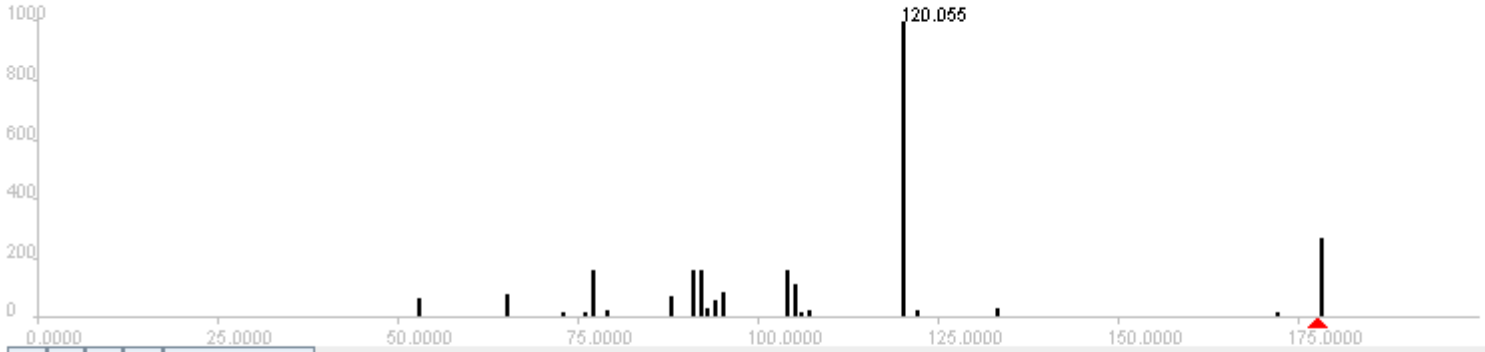
Image © www.seanoakley.com/

Level 4: Molecular formula assigned

MassBank Record: ETS00108

1H-Benzotriazole TP18 (Tentative); LC-ESI-ITFT; MS2; 80-110; R=7500; [M+H]+

Mass Spectrum



<< < > >> show all m/z

Chemical Structure

Not Available

```
ACCESSION: ETS00108
RECORD_TITLE: 1H-Benzotriazole TP18 (Tentative); LC-ESI-ITFT; MS2; 80-110; R=7500; [M+H]+
DATE: 2014.06.27
AUTHORS: Huntscha S, Schymanski E, Hofstetter TB, Spahr S, Hollender J, Department of Environmental Chemistry, Eawag
LICENSE: CC BY-SA
COPYRIGHT: Copyright (C) 2014 Eawag, Duebendorf, Switzerland
PUBLICATION: Huntscha S, Hofstetter TB, Schymanski E, Spahr S, Hollender J (2014) Environ. Sci. Technol, 48:4435-4443,
COMMENT: CONFIDENCE MOLECULAR FORMULA IDENTIFIED ONLY!
COMMENT: Source; 178m0605a_MSMS.txt

CH$NAME: 1H-Benzotriazole TP18 (Tentative)
CH$NAME: 1H-Benzotriazole Transformation Product 18 (TP18)
CH$COMPOUND_CLASS: N/A; Biotransformation Product
CH$FORMULA: C8H7N3O2
CH$EXACT_MASS: 177.0538
CH$SMILES: N/A
CH$IUPAC: N/A
```

Level 3: Tentative candidate(s)

Candidate selection: Substituents or class defined

A very subjective level; dependent on experiment

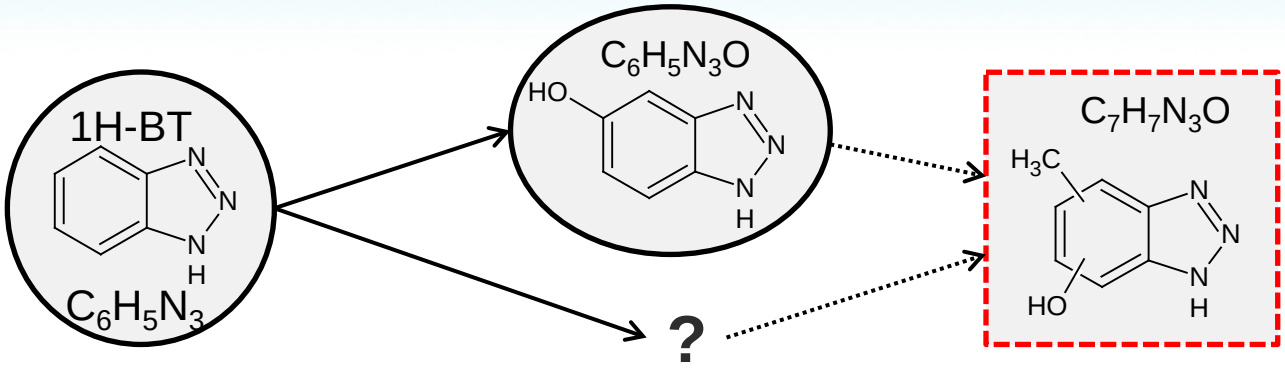
- Some knowledge of substructures, substituents, compound class
- Background knowledge or “hunch”
- *In silico* methods; candidate selection => “top rank”
- “There’s only one database entry with this mass anyway”
- …

If there is insufficient evidence for a complete identification => all Level 3

Level 3: Tentative candidate(s)

Examples

Transformation Products



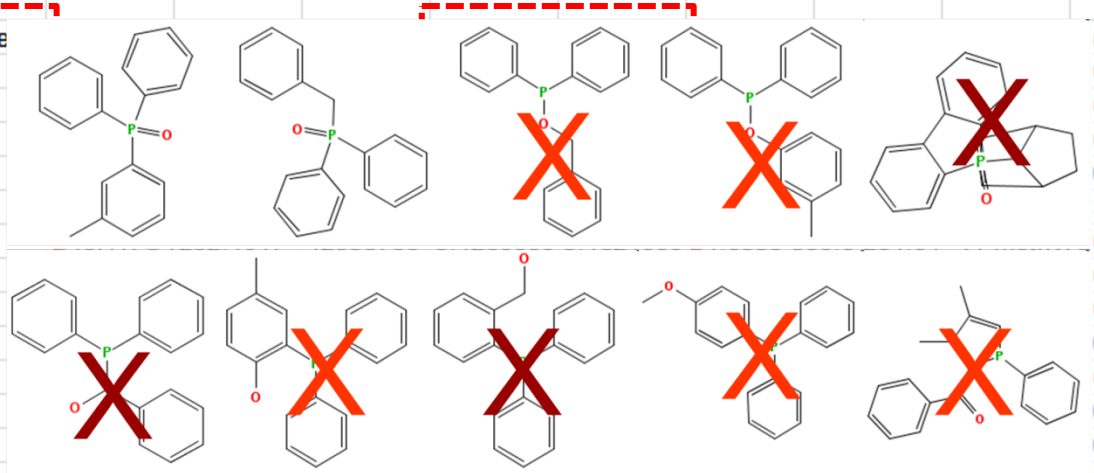
In silico methods

MetFragResults_1349071807740.xls [Compatibility Mode]

ChemSpider
Search and share chemistry

MetFrag

	A	B	C	D	E	F	G
1	Database:	Biologica	Mode : [M+H]+				
2	Exact ma:	Limit # of	Mzabs: 0.005				
3	Formula:	Database	Mzppm: 5				
4							
5	Rank	Score	# of Pe				
6	1	1.0	3				
7	2	1.0	3				
8	3	1.0	3				
9	4	1.0	3				
10	5	0.954	3				
11	6	0.954	3				
12	7	0.954	3				
13	8	0.954	3				
14	9	0.878	1				
15	10	0.05	2				
16	11	0.05	2				
17	12	0.05	2				
18	13	0.017	1				



miles

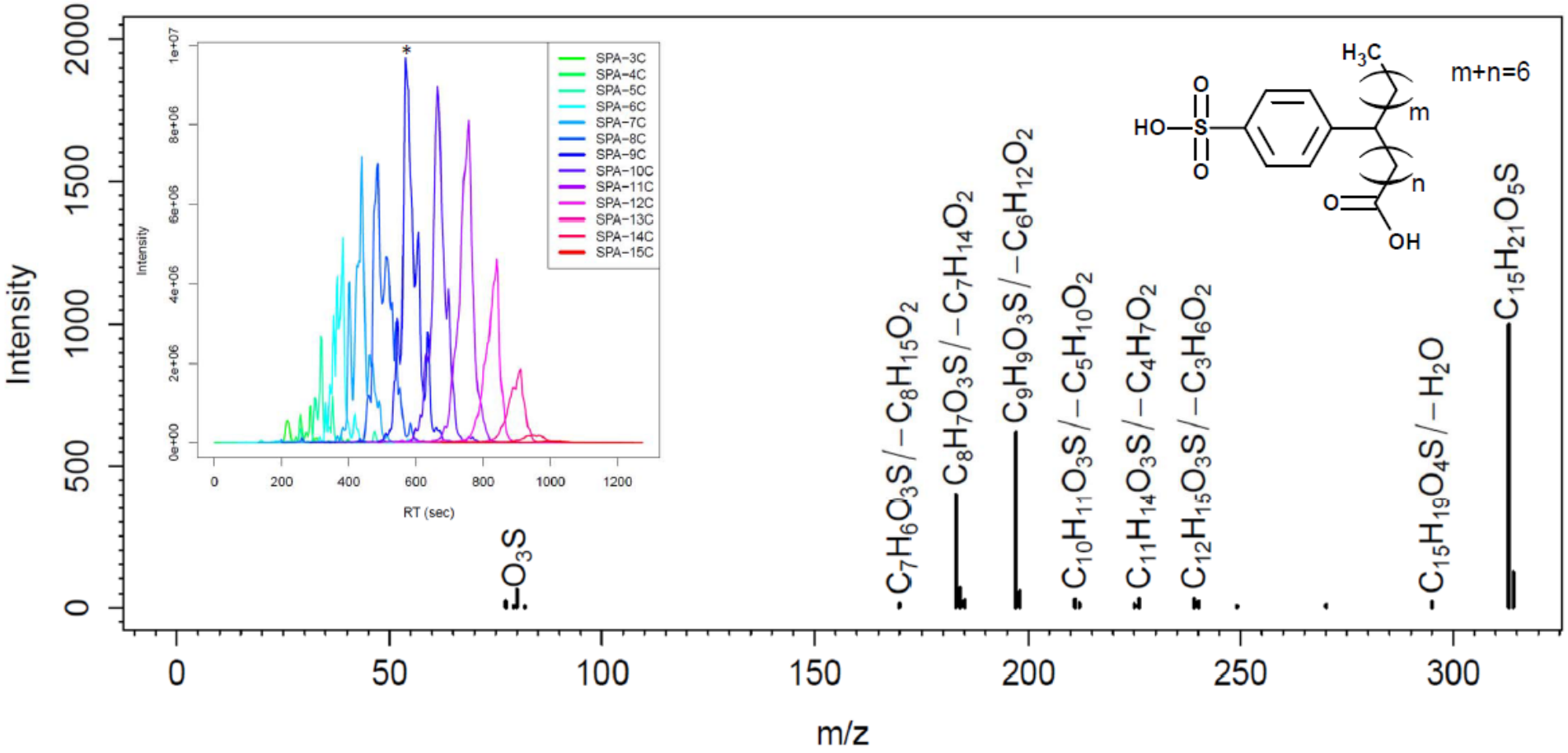
C1=CC=CC=C1P(=O)(C2=CC=CC=C
C=1C=CC=C(C=1)P(=O)(C2=CC=C
C1=CC=C(C=C1)P(=O)(C2=CC=CC
=P(CC1=CC=CC=C1)(C2=CC=CC=
1=CC=C(C=C1)COP(C2=CC=CC=C
C1=CC=C(C=C1)OP(C2=CC=CC=C
C=1C=CC=C(C=1)OP(C2=CC=CC=
C3=CC=CC=C3(OP(C1=CC=CC=C1
=P2(C3=CC=CC=C3(C1=CC=CC=C
C=1C=CC(O)=C(C=1)P(C2=CC=CC=
CC1=CC=CC=C1P(C2=CC=CC=C2)
C(C1=CC=CC=C1)P(C2=CC=CC=C
OC=1C=CC=C(C=1)P(C2=CC=CC=

Huntscha *et al.*, (2014) *ES&T*, 48(8) 4435-4443; Gallampois *et al. in prep.* Resources: www.chemspider.com; http://msbi.ipb-halle.de/MetFrag/

Level 3: Tentative candidate(s)

Examples: weight of evidence in sample

RMassBank



M. Stravs, E. Schymanski, H. Singer, J. Hollender (2013) *J. Mass Spectrom.,* 48(1), 89-99. DOI: 10.1002/jms.3131

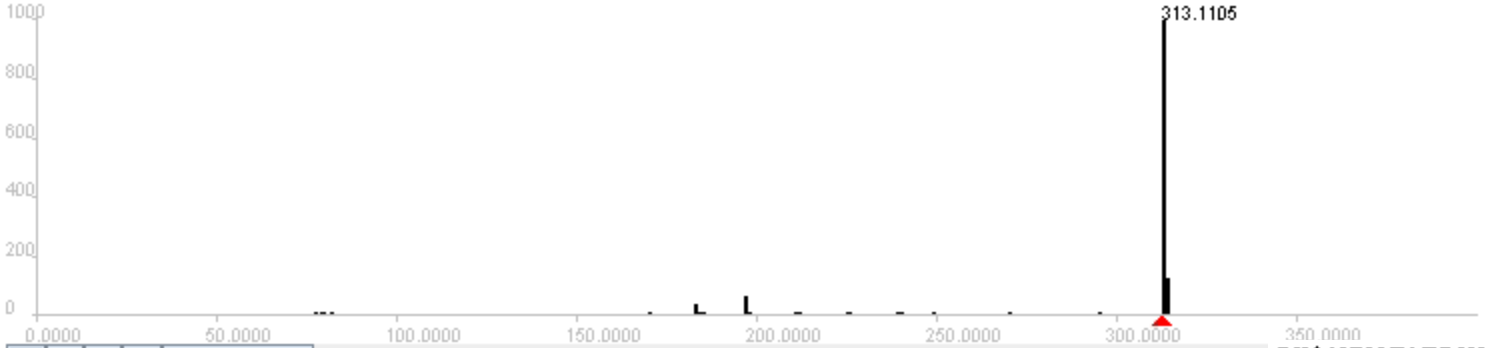
E. Schymanski, H. Singer, P. Longree, M. Loos, M. Ruff, M. Stravs, C. Ripolles Vidal & J. Hollender (2014) *ES&T*, 48(3), 1811-1819.

Level 3: Tentative candidate(s)

MassBank Record: ETS00002

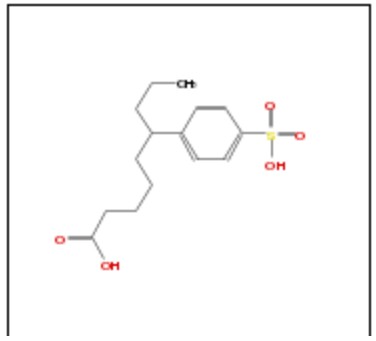
SPA-9C (TENTATIVE); LC-ESI-ITFT; MS2; HCD60; [M-H]-

Mass Spectrum



<< < > >> show all m/z

Chemical Structure



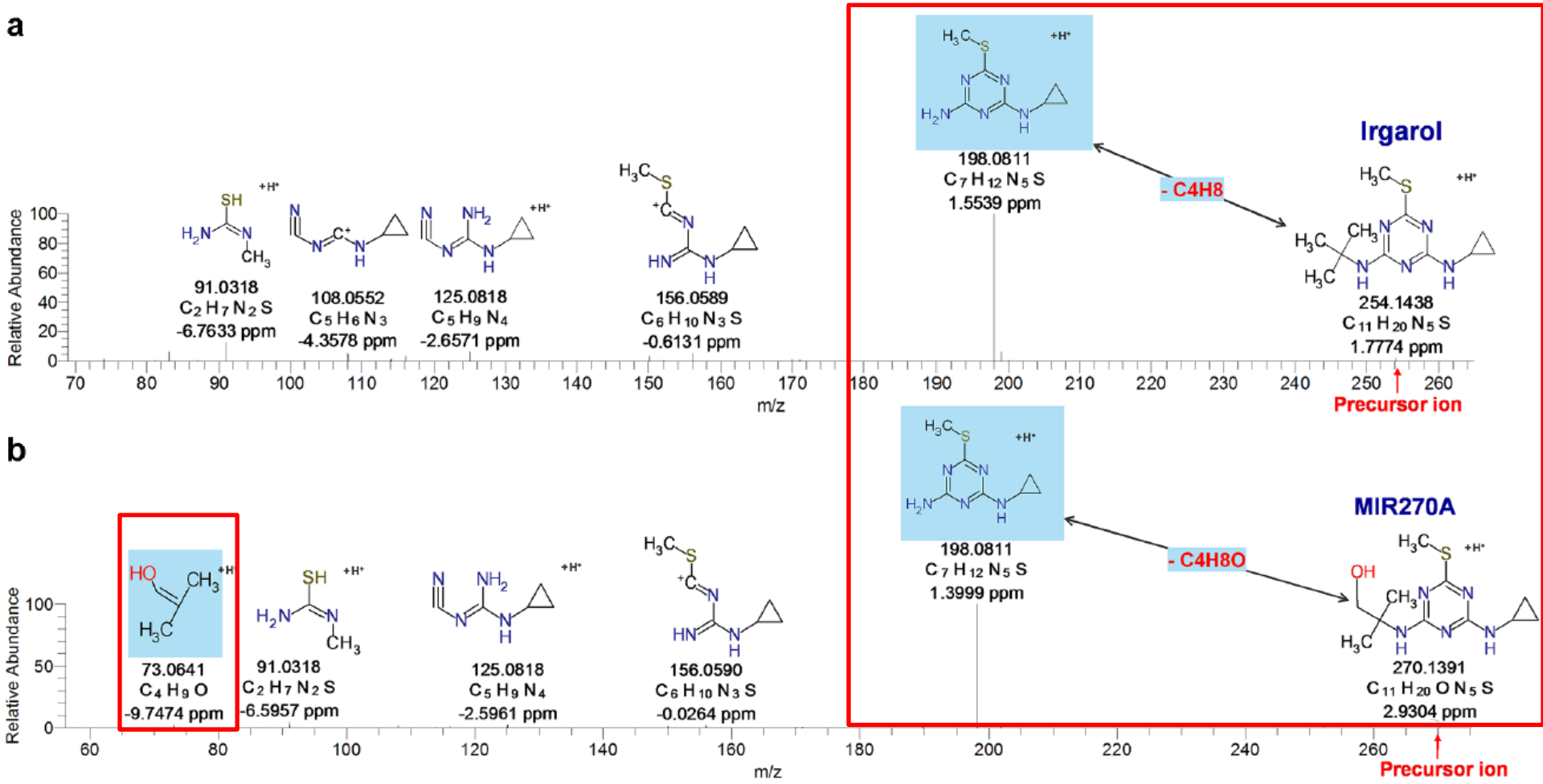
```
ACCESSION: ETS00002
RECORD_TITLE: SPA-9C (TENTATIVE); LC-ESI-ITFT; MS2; HCD60; [M-H]-
DATE: 2013.08.19
AUTHORS: E. Schymanski, Dept. of Environmental Chemistry, Eawag, Switzerland
LICENSE: CC BY-SA
COPYRIGHT: Copyright (C) 2013 Eawag, Duebendorf, Switzerland
COMMENT: TENTATIVELY IDENTIFIED SPECTRUM ONLY!!!!!!!!!
COMMENT: Extracted (without noise removal) from ZUE_N: mz313_11_rt10_56_HCD60_S
COMMENT: CONFIDENCE: one of several possible homologues;
COMMENT: Good match to 6phiC9SPC, LIT00035

CH$NAME: SPA-9C (TENTATIVE)
```

```
PK$ANNOTATION: m/z int. subformula loss
  79.9573 46351 O3S -C15H21O2
  170.0046 9793 C7H6O3S -C8H15O2
  183.0119 292054 C8H7O3S -C7H14O2
  184.0205 51077 C8H8O3S -C7H13O2
  185.0086 18731 C8H9O3S -C7H12O2
  197.0273 452737 C9H9O3S -C6H12O2
  198.0365 41943 C9H12O3S -C6H11O2
  211.0427 20440 C10H11O3S -C5H10O2
  212.0514 11308 C10H12O3S -C5H9O2
  225.0578 8340 C11H13O3S -C4H8O2
  226.067 21933 C11H14O3S -C4H7O2
  239.0739 22565 C12H15O3S -C3H6O2
  240.0823 15750 C12H16O3S -C3H5O2
  249.1476 5480 C15H21O3 -O2S
  295.0997 14684 C15H19O4S -H2O
  313.1105 7299560 C15H21O5S none
```

Level 2b: Probable structure - Diagnostic spectrum

Only one structure fits the experimental/spectral information available



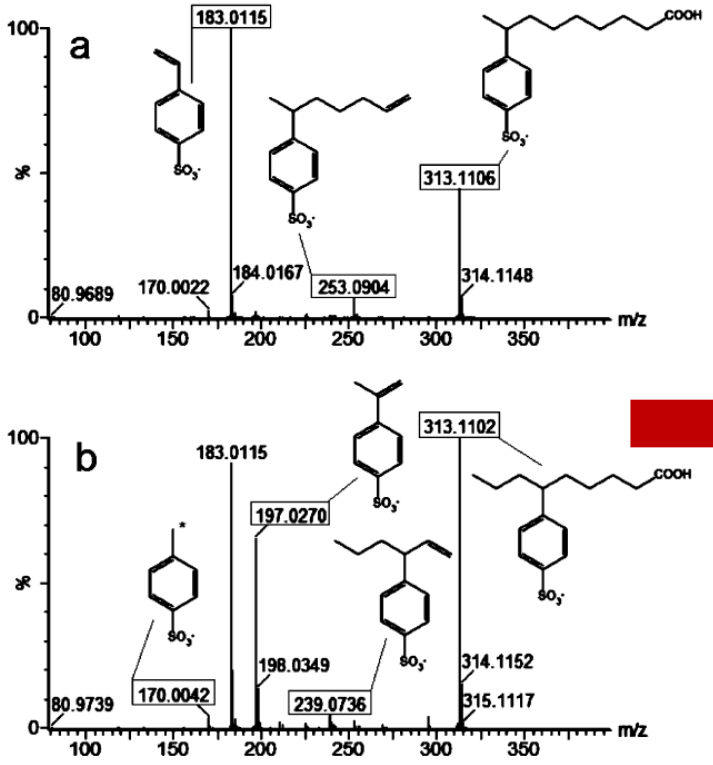
Source: Figure 3, J. Jeon, D. Kurth, J. Hollender; 2013. *Chem. Res. Toxicol.* 26, p313-324, DOI: 10.1021/tx300457f.

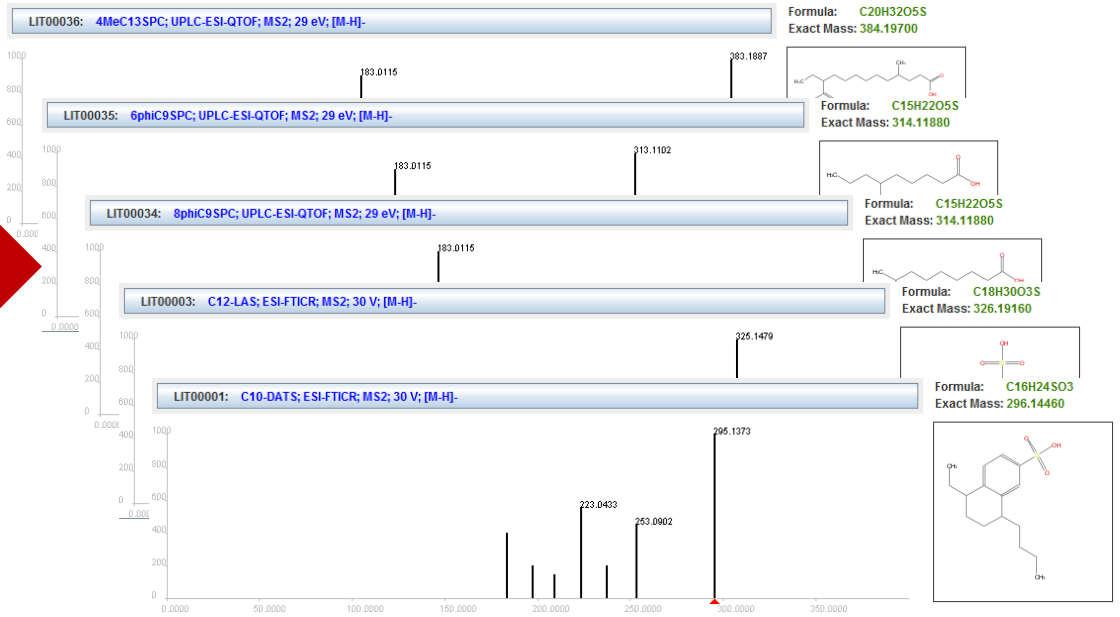
Level 2a: Probable structure - Library spectrum

Enable access and comparison

Adding literature spectra to
www.massbank.eu/MassBank/

39 literature spectra (so far)





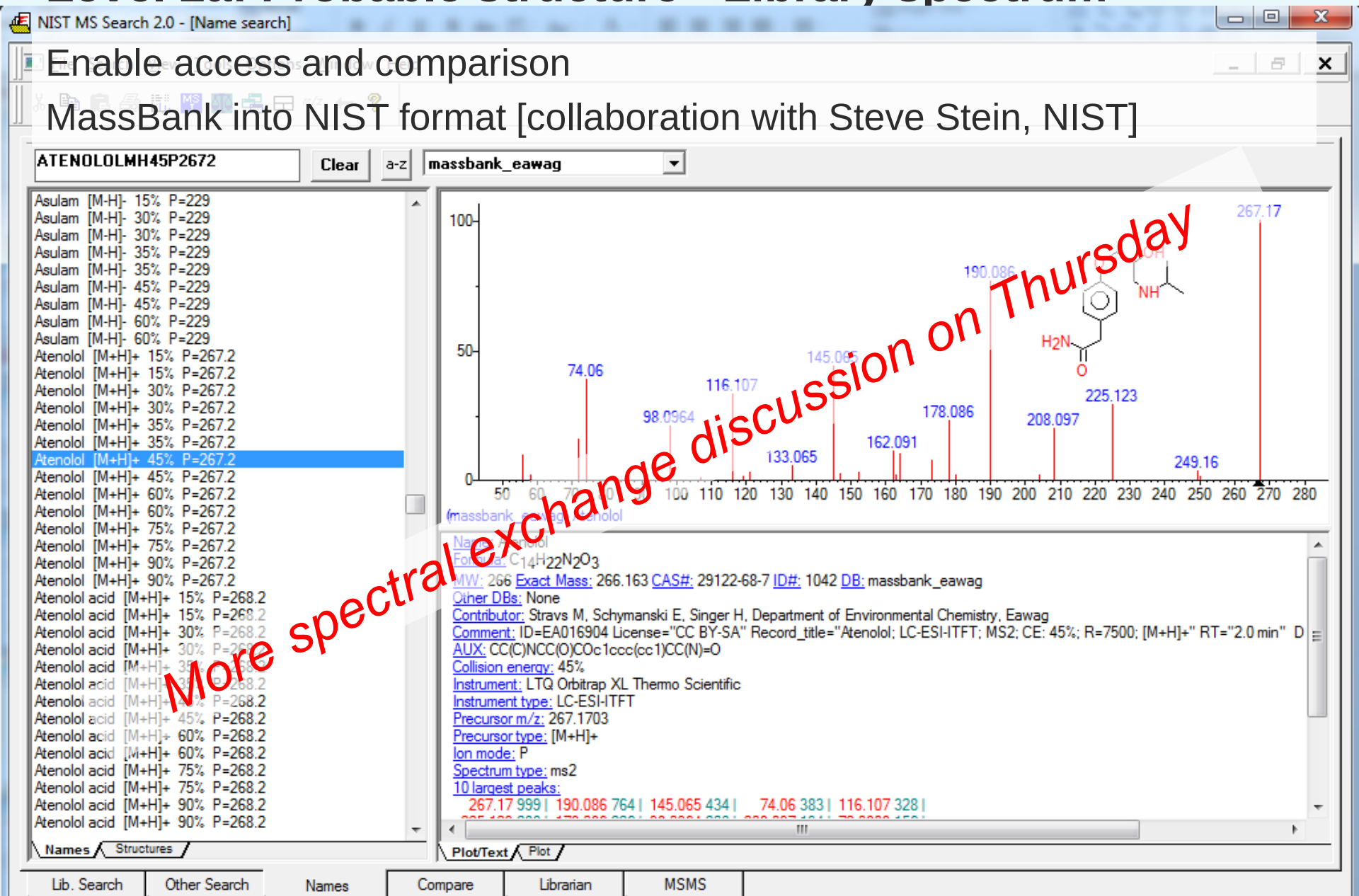
Lara-Martin et al. EST. 2010, 44: 1670-1676

E. Schymanski, H. Singer, P. Longree, M. Loos, M. Ruff, M. Stravs, C. Ripolles Vidal & J. Hollender (2014) *ES&T*, 48(3), 1811-1819.

Level 2a: Probable structure - Library spectrum

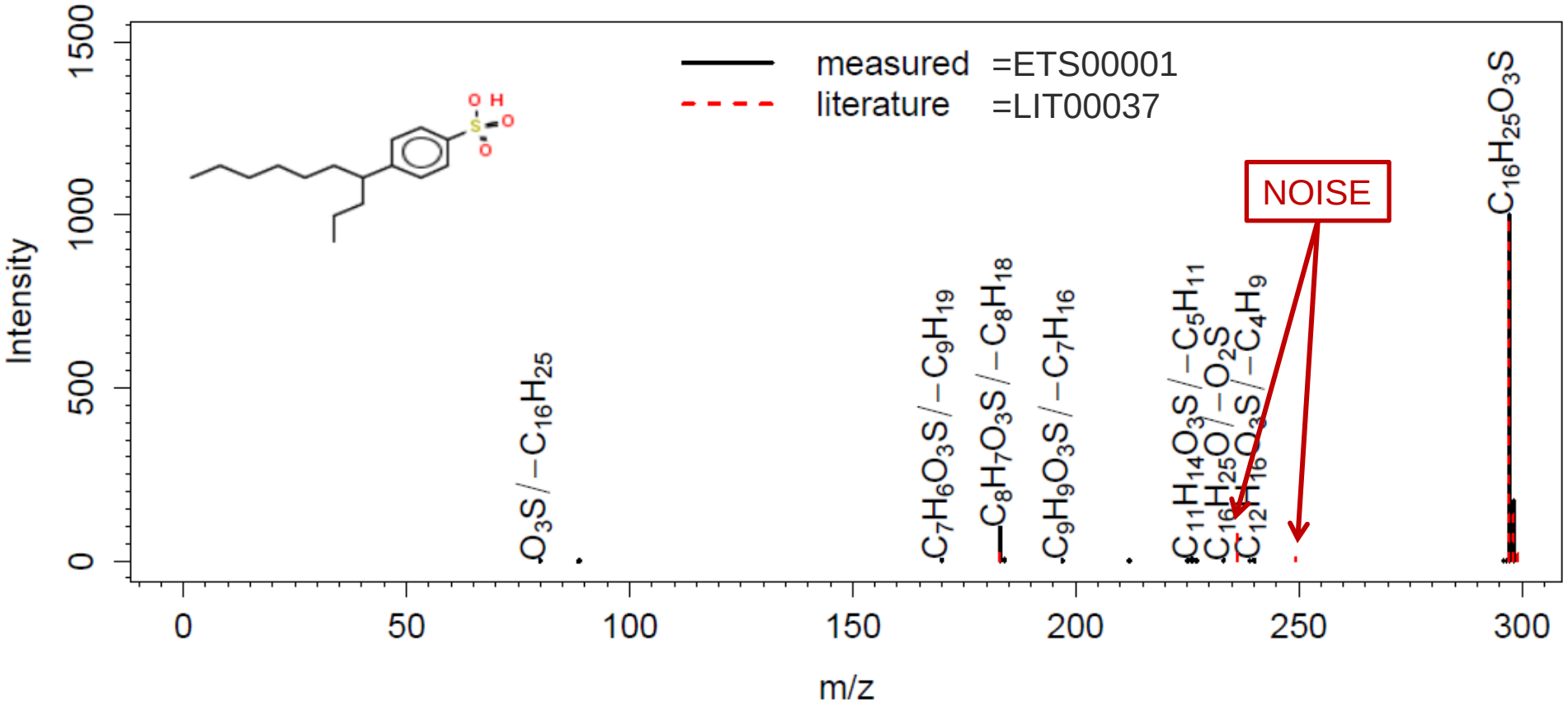
Enable access and comparison

MassBank into NIST format [collaboration with Steve Stein, NIST]



Level 2a: Probable structure - Library spectrum

Open question: how “good” is good?



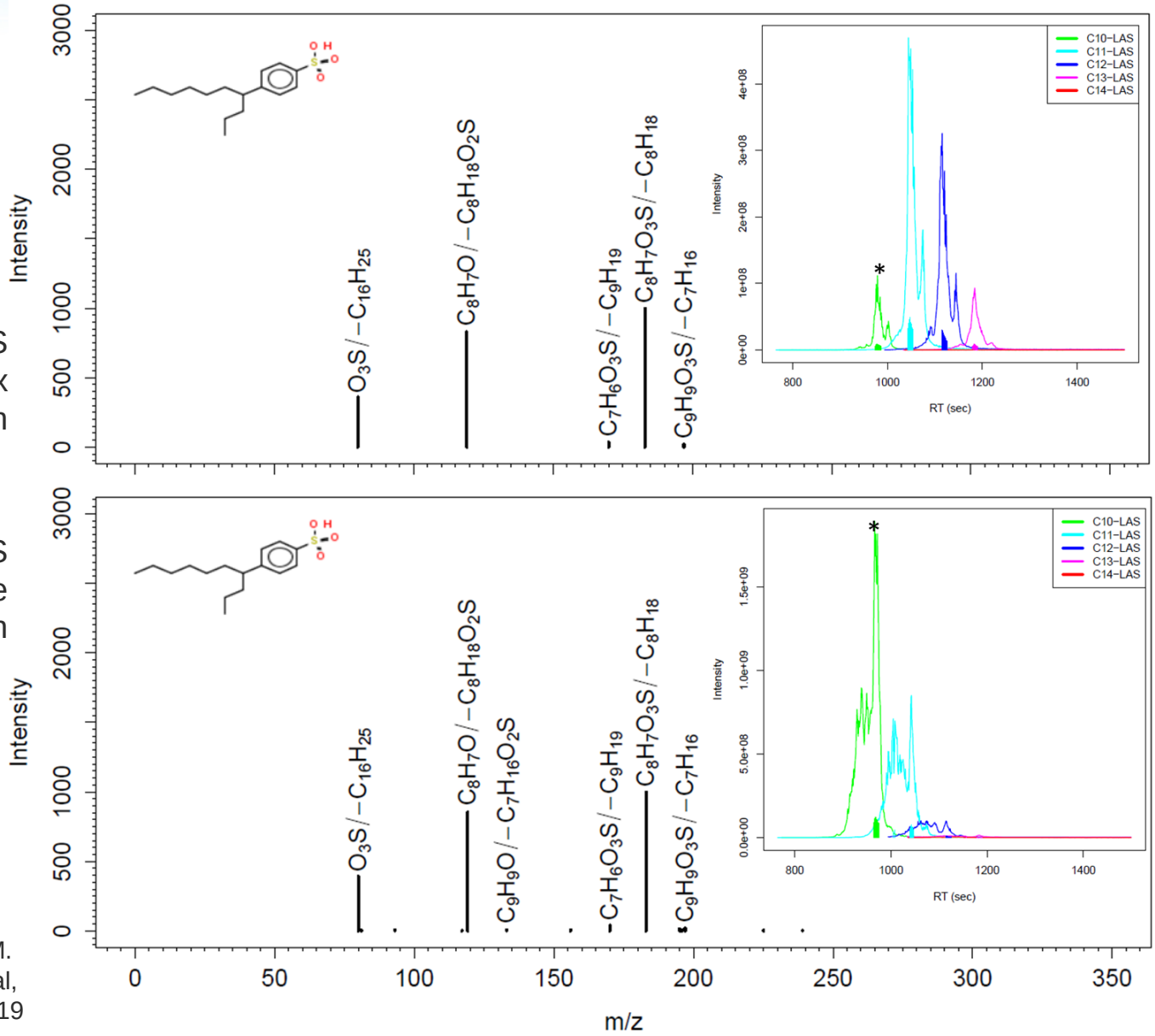
Literature spectrum from: Lara-Martin et al. *J. Chrom. A*. 2011, 1218 (30): 4799-4807
M. Stravs, E. Schymanski, H. Singer, J. Hollender, 2013, *J. Mass Spectrom.,* DOI: 10.1002/jms.3131
E. Schymanski, H. Singer, P. Longree, M. Loos, M. Ruff, M. Stravs, C. Ripolles Vidal & J. Hollender, (2014) *ES&T*, 48(3), 1811-1819.

Level 1: Confirmation with reference standard

Open question:
How many orthogonal methods?

HCD90, C10 LAS
Standard mix
m/z 297.15, 16.32 min

HCD90, C10 LAS
Sample
m/z 297.15, 16.11 min



M. Stravs, E. Schymanski, H. Singer, J. Hollender, 2013, *J. Mass Spectrom.*, DOI: 10.1002/jms.3131
E. Schymanski, H. Singer, P. Longree, M. Loos, M. Ruff, M. Stravs, C. Ripolles Vidal, J. Hollender (2014) *ES&T*, 48(3) 1811-1819

Level 1: Confirmation with reference standard

Open question: When is there sufficient evidence for a target?

m/z	RT	Name	Level	
120.0556	4.69	Benzotriazole	Level 1, 4.5 IP (HRMS+dds-MS/MS)	✓
=> fulfils all criteria				
134.0712	5.97	4&5-Methyl-Benzotriazole	Level 1, 4.5 IP (HRMS+dds-MS/MS)	?
=> co-eluting isomers: fulfils all identification criteria, but not a unique structure				
170.1288	3.49	Pyrimidinol	Level1, 4 IP (HRMS+DIA MS/MS)	?
=> MS/MS from data independent acquisition only, not isolated precursor				
123.0916	1.07	4-Dimethylaminopyridine	Level 1, 2 IP (HRMS only)	?
=> no MS/MS available in this analysis, but a target with reference standard				

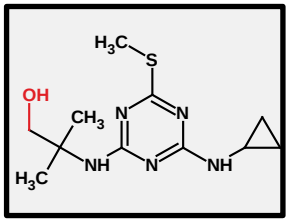
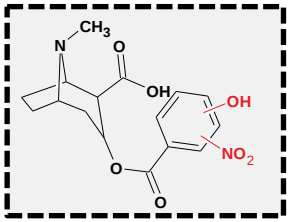
The combination of “level” and “score” could help represent structure and evidence

Acknowledgements

MassBank norman solutions

Identification with (LC-HR)MS/MS – Confidence?

Any questions?

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
$C_6H_5N_3O_4$	***Level 4: Unequivocal molecular formula***	MS isotope/adduct
192.0757	***Level 5: Exact mass*** of interest	MS

Schymanski, Jeon, Gulde, Fenner, Ruff, Singer & Hollender (2014) *ES&T*, 48 (4), 2097-2098. DOI: 10.1021/es5002105