



# *Enhancing Spectral Exchange: Tentative and Unknown Spectra, NIST and Crowdsourcing Trends*

**NORMAN MassBank Workshop; 17-18 September 2014**

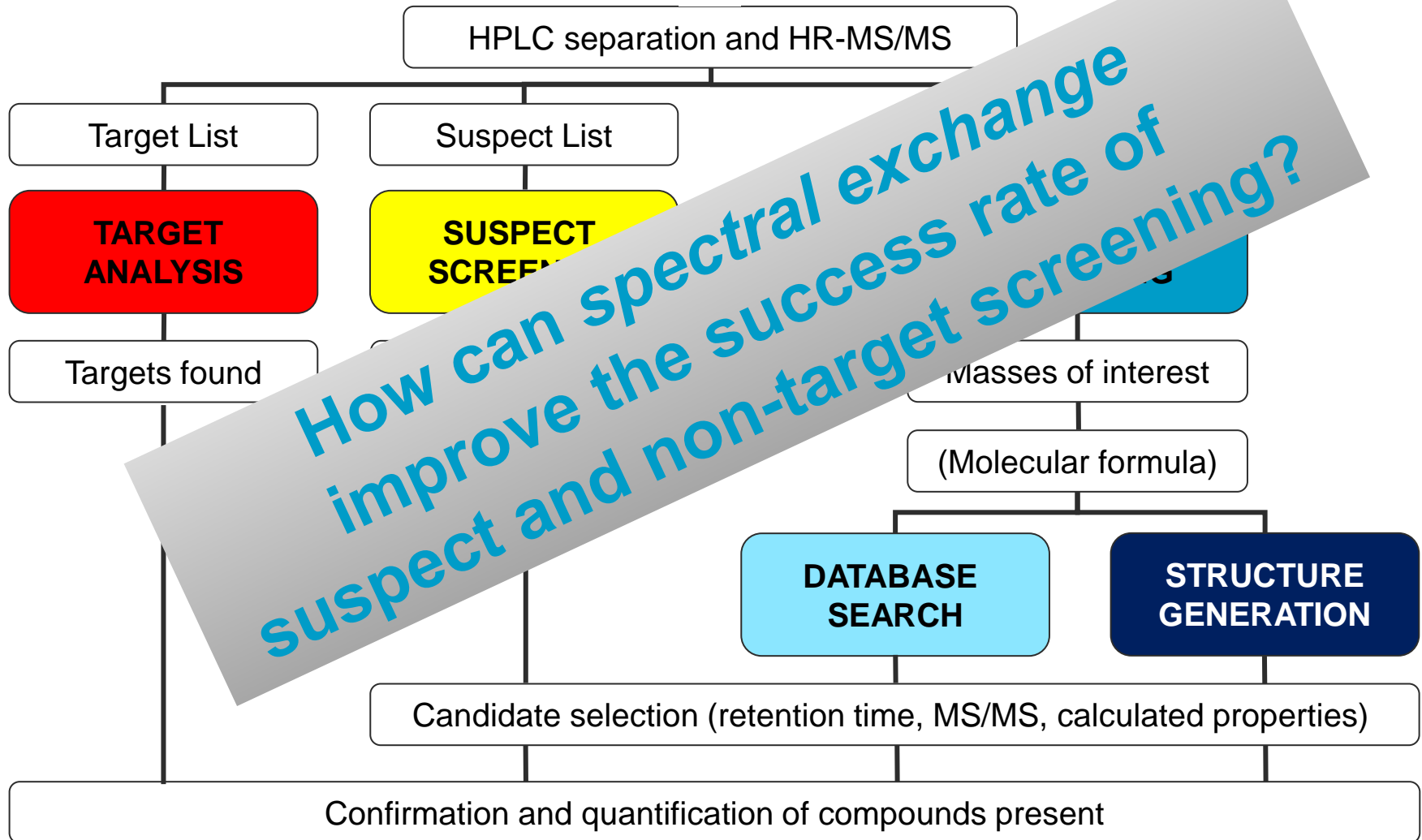
Emma Schymanski

Juliane Hollender, Heinz Singer und Environmental Chemistry Dept.

*Eawag: Swiss Federal Institute of Aquatic Science and Technology*

*Dübendorf, Switzerland*

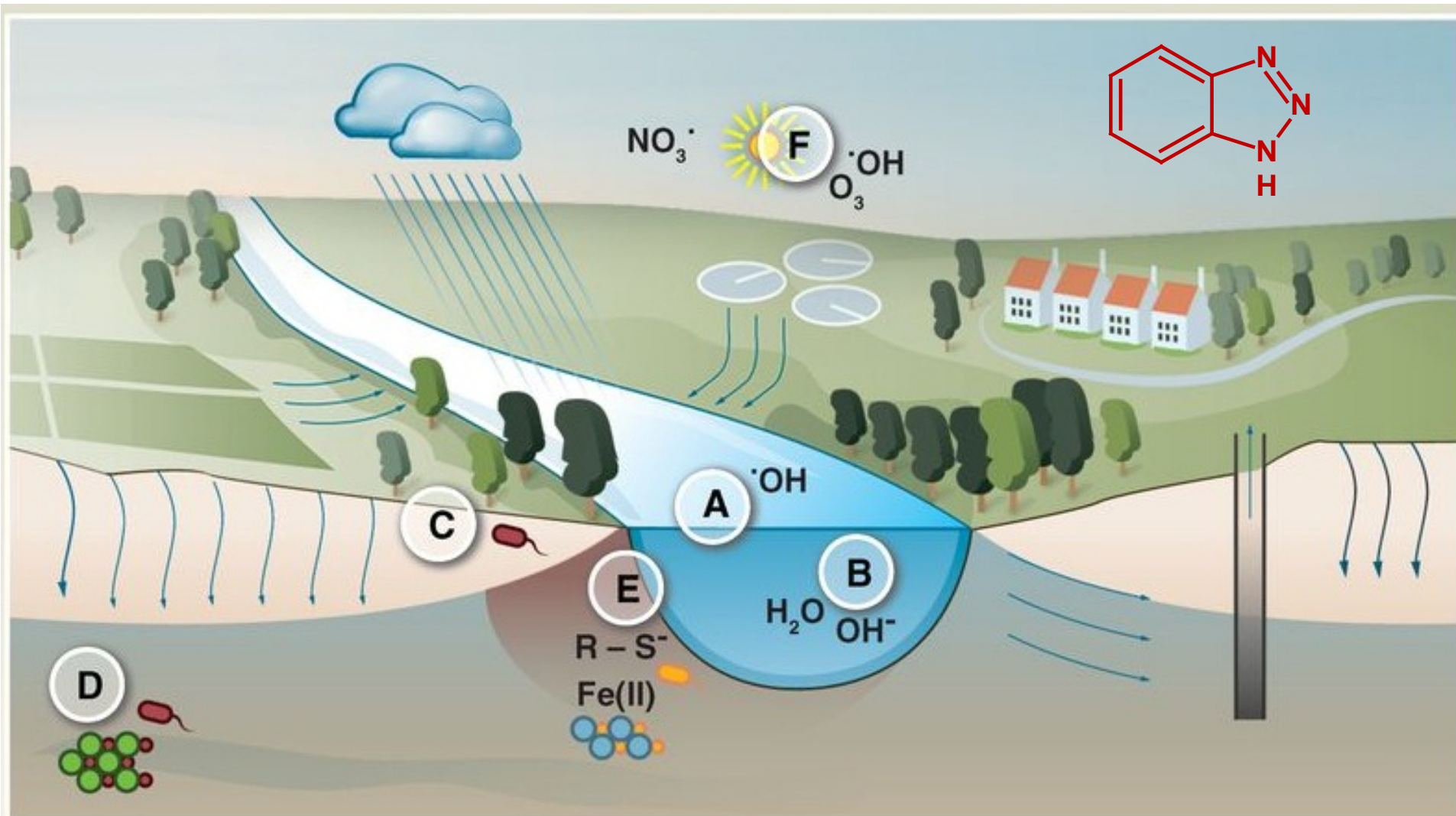
# Status quo of identification approaches



*How can spectral exchange improve the success rate of suspect and non-target screening?*

# The Reference Standard Dilemma<sup>1</sup>

...and why identification in the environment is such a challenge



# Accessing Mass Spectral Information

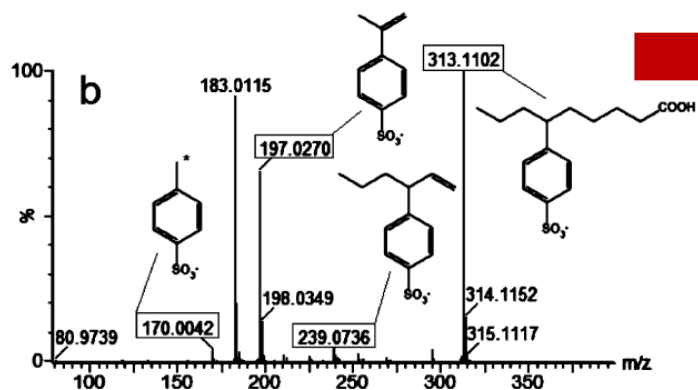
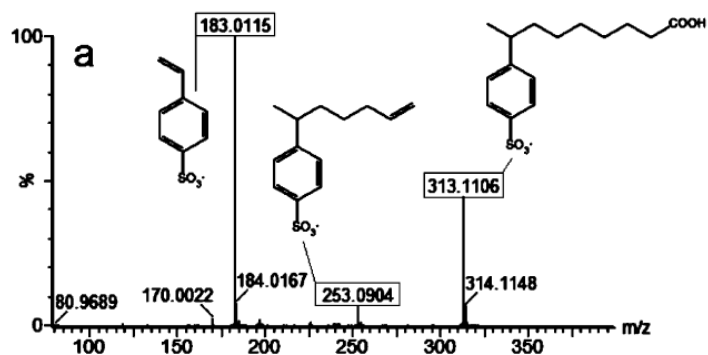
(Too) Many ways to access mass spectral information

- Mass Spectral Libraries
  - Curated, reference standard spectra: NIST, Wiley, mzCloud, METLIN
  - User-defined personal libraries
- Mass Spectral Repositories
  - User-contributed reference spectra: MassBank, HMDB, GOLM, GNPS
- Literature sources
  - Contain reference standard and elucidated spectra, but:
    - Tend to be in paper form, not digital!!!!!!
- ***Great need to improve the accessibility of spectral information!***

# Accessing Literature Spectra

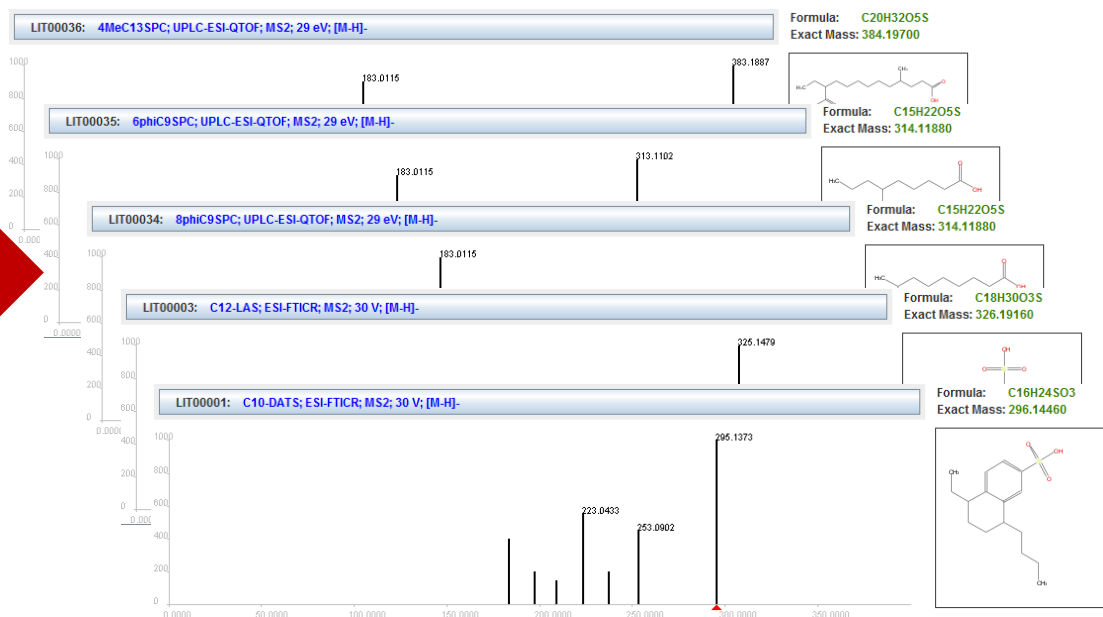
Enable access and comparison

Adding literature spectra to  
[www.massbank.eu/MassBank/](http://www.massbank.eu/MassBank/)



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

39 literature spectra (so far)



# User Contribution of Spectra as Supporting Information

Index Type :

Contributor: Eawag Additional Mass Spectra

[Edit / Resubmit Query](#)

Results : 50 Hit. ( 1 - 27 Displayed )

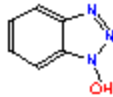
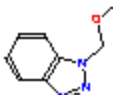
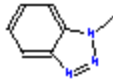

Open All Tree

Multiple Display

Spectrum Search

First Prev 1 2 Next Last ( Total 2 Page )

▼ Results End

<input type="checkbox"/>	Name		Formula / Structure	ExactMass	ID
<input type="checkbox"/>	<input checked="" type="checkbox"/> 1-Hydroxybenzotriazole	1 spectrum	C6H5N3O 	135.04330	
<input type="checkbox"/>	<a href="#">LC-ESI-ITFT; MS2; 80-110; R=7500; [M+H]<sup>+</sup></a>				ETS00117
<input type="checkbox"/>	<input checked="" type="checkbox"/> 1-Methoxymethylbenzotriazole	1 spectrum	C8H9N3O 	163.17660	
<input type="checkbox"/>	<a href="#">LC-ESI-ITFT; MS2; 80-110; R=7500; [M+H]<sup>+</sup></a>				ETS00116
<input type="checkbox"/>	<input checked="" type="checkbox"/> 1-Methylbenzotriazole	1 spectrum	C7H7N3 	133.06400	
<input type="checkbox"/>	<a href="#">LC-ESI-ITFT; MS2; 80-110; R=7500; [M+H]<sup>+</sup></a>				ETS00115
<input type="checkbox"/>	<input checked="" type="checkbox"/> 10phiC10SPC (STANDARD)	2 spectra	C16H24O5S 	328.13440	
<input type="checkbox"/>	<a href="#">LC-ESI-QFT; MS2; HCD50; [M-H]<sup>-</sup></a>				ETS00019
<input type="checkbox"/>	<a href="#">LC-ESI-QFT; MS2; HCD80; [M-H]<sup>-</sup></a>				ETS00020



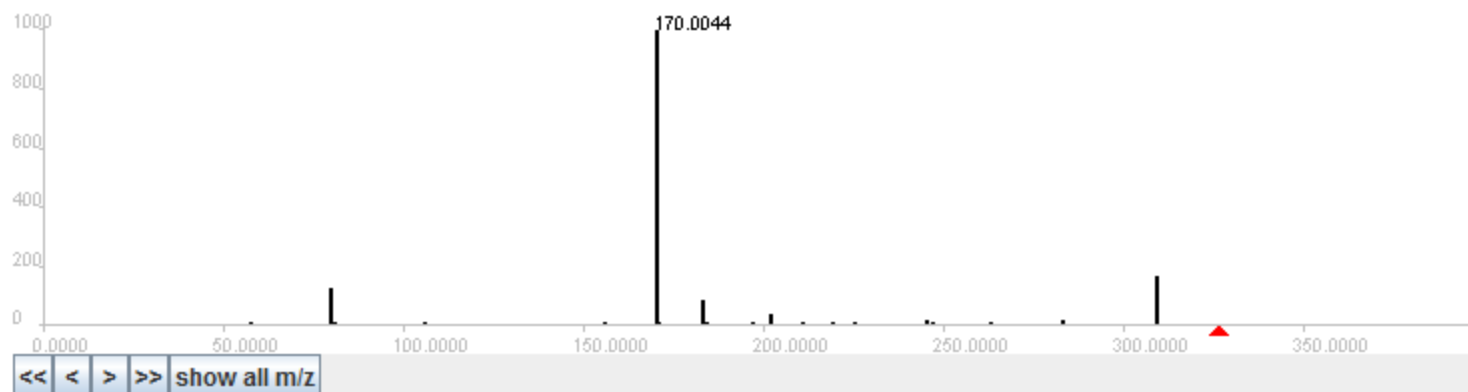
# User Contribution of Spectra as Supporting Information

## MassBank Record: ETS00019

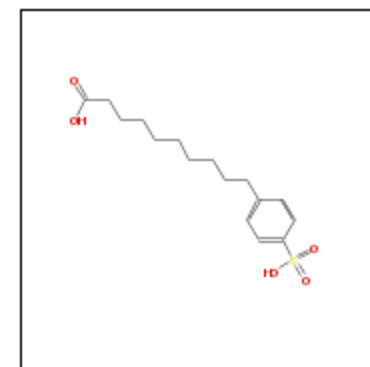
[Home](#) | [Spectrum](#) | [Quick](#) | [Peak](#) | [Substructure](#) | [Browser](#) | [Batch](#) | [Browse](#) | [Index](#) | MassBank ID:

10phiC10SPC (STANDARD); LC-ESI-QFT; MS2; HCD50; [M-H]-

Mass Spectrum



Chemical Structure



ACCESSION: ETS00019

RECORD\_TITLE: 10phiC10SPC (STANDARD); LC-ESI-QFT; MS2; HCD50; [M-H]-

DATE: 2014.06.26

AUTHORS: E. Schymanski, Dept. of Environmental Chemistry, Eawag, Switzerland

LICENSE: CC BY-SA

COPYRIGHT: Copyright (C) 2013 Eawag, Duebendorf, Switzerland

PUBLICATION: EL Schymanski, HP Singer, P Longree, M Loos, M Ruff, M Stravs, C Ripolles Vidal, J Hollender (2014) Environ

COMMENT: Auto-extracted from 131001\_neg\_02.mzML

COMMENT: CONFIDENCE: synthesised standard

COMMENT: Kindly provided by Jennifer Field, Oregon State University

COMMENT: Synthesis according to P.W. Taylor and G. Nickless, J. Chromotography, 178 (1979) 259-269.

CH\$NAME: 10phiC10SPC (STANDARD)

# User Contribution of Spectra as Supporting Information

ENVIRONMENTAL  
Science & Technology

Article

pubs.acs.org/est

## Strategies to Characterize Biotransformation of Benzotriazoles: Insights from Transformation Wastewater: Exploring the Product Identification and Compound-Specific Isotope Analysis Spectrometry

Sebastian Huntscha,<sup>†,‡</sup> Thomas B. Hofstetter,<sup>†,‡</sup> Emma L. Schymanski,<sup>†</sup> Stephanie Spahr,<sup>†</sup>

Emma L. Schymanski,<sup>†</sup> Heinz P. Singer and Juliane Hollender<sup>\*,†,‡</sup>

Michael A. Stravs,<sup>†,§</sup> Cristina Ripollés<sup>‡</sup>

<sup>†</sup>Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

<sup>‡</sup>Eawag: Swiss Federal Institute of Aquatic Science and Technology, Institute of Biogeochemistry and Pollutant Dynamics (IBP), ETH Zurich, 8092 Zurich, Switzerland

<sup>§</sup>University Jaume I, Department of Physical and

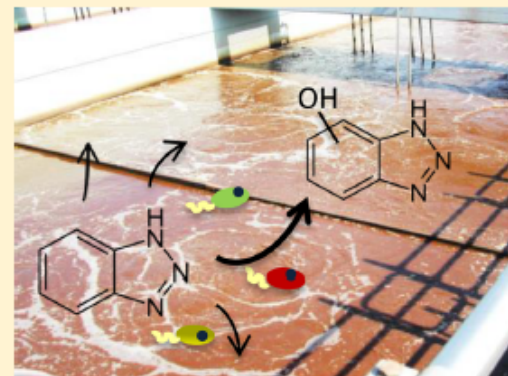
<sup>§</sup>Institute of Biogeochemistry and Pollutant Dynamics

### Supporting Information

### Supporting Information

**ABSTRACT:** Wastewater effluents contain a wide range of contaminants and transformation products, which are not covered by target analysis alone. High accuracy, high resolution data were explored with novel untargeted data (enviMass, nontarget, and RMassBank) to complement target analysis in initial “all in one” measurements. The detected peaks from 10 Swiss wastewater effluents were assigned to target compounds, with 30 available. Corrosion inhibitors, artificial sweeteners exhibited the highest concentrations. After blank subtraction, 70% of the peaks remained and were grouped into 10 clusters. These components had adduct and/or isotope information. Intensity-based prioritization revealed that only 40 of the top 30 most intense peaks (negative mode), tentatively identified via suspect screening for benzothiazole-2-sulfonate, an oxidation product of benzothiazole, were identified using tailor-made nontarget processing methods (all of heteroatom-containing compounds in the en-

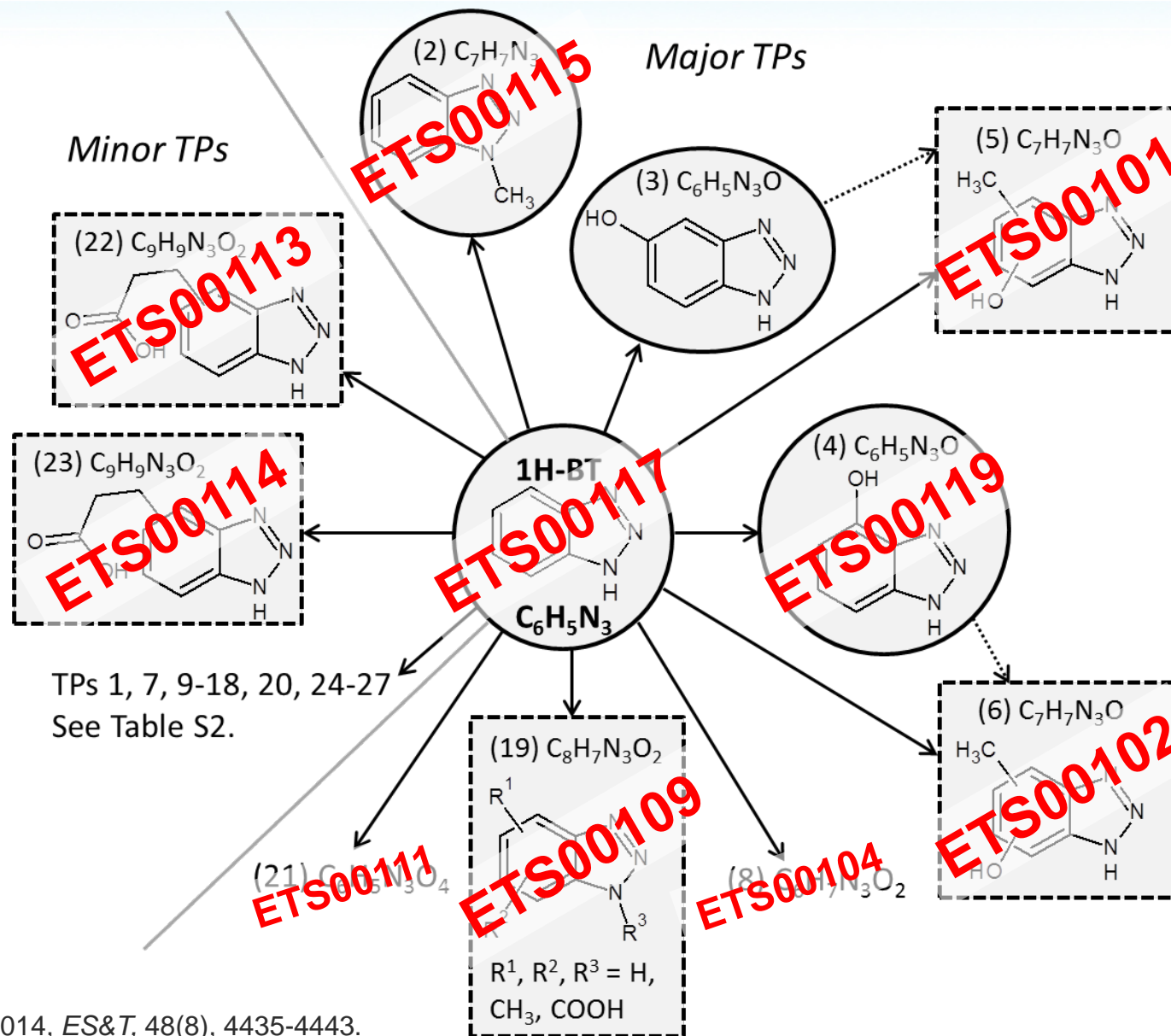
**ABSTRACT:** Benzotriazoles are widely used domestic and industrial corrosion inhibitors and have become omnipresent organic micropollutants in the aquatic environment. Here, the range of aerobic biological degradation mechanisms of benzotriazoles in activated sludge was investigated. Degradation pathways were elucidated by identifying transient and persistent transformation products in batch experiments using liquid chromatography–high-resolution tandem mass spectrometry (LC-HR-MS/MS). In addition, initial reactions were studied using compound-specific isotope analysis (CSIA). Biodegradation half-lives of 1.0 days for 1H-benzotriazole, 8.5 days for 4-methyl-1H-benzotriazole, and 0.9 days for 5-methyl-1H-benzotriazole with activated sludge confirmed their known partial persistence in conventional wastewater treatment. Major transformation products were identified as 4- and 5-hydroxy-1H-benzotriazole for the degradation of 1H-benzotriazole, and 1H-benzotriazole-5-carboxylic acid for the degradation of 5-methyl-1H-benzotriazole. These transformation products were found in wastewater effluents, showing their environmental relevance. Many other candidate transformation products, tentatively identified by interpretation of HR-MS/MS spectra, showed the broad range of possible reaction pathways including oxidation, alkylation, hydroxylation and indicate the significance of cometabolic processes for micropollutant degradation in biological wastewater treatment in general. The combination of evidence from product analysis with the significant carbon and nitrogen isotope fractionation suggests that aromatic monohydroxylation is the predominant step during the biotransformation of 1H-benzotriazole.





# Supporting Information in MassBank

1H-BT



# Identification of Plant Metabolites of Environmental Contaminants by UPLC-QToF-MS: The in Vitro Metabolism of Triclosan in Horseradish

André Macherius,<sup>†</sup> Bettina Seiwert,<sup>†</sup> Peter S. and Thorsten Reemtsma<sup>\*,†</sup>

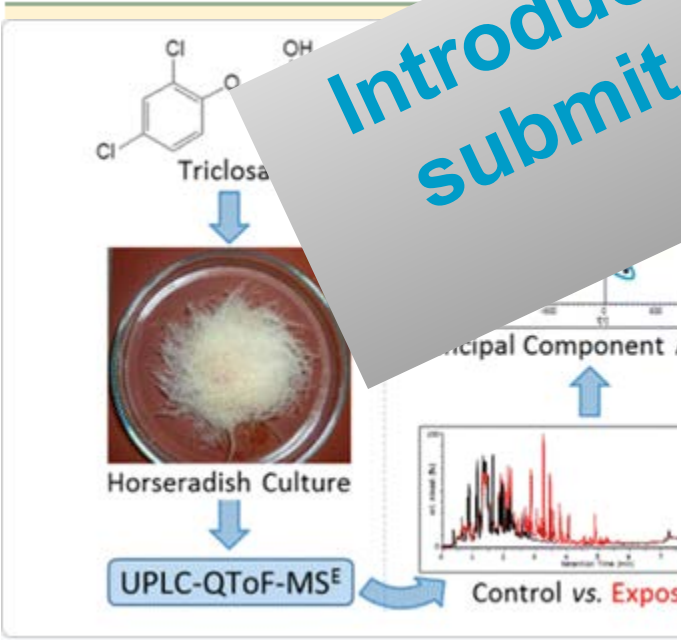
**ENVIRONMENTAL  
 Science & Technology**

ARTICLE  
 pubs.acs.org/est

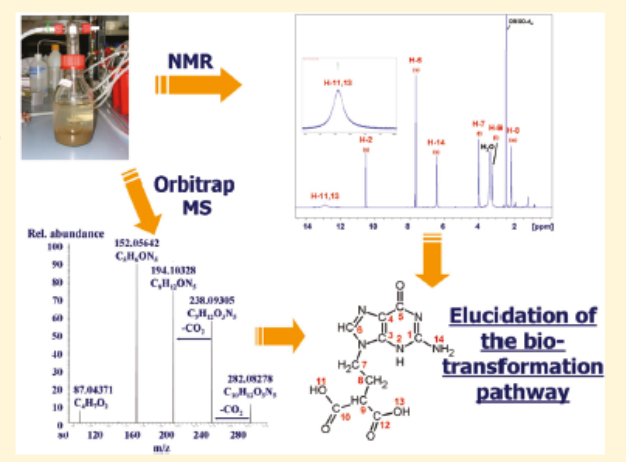
<sup>†</sup>Department of Analytical Chemistry, Helmholtz Centre for Environmental Research, 37071 Braunschweig, Germany  
<sup>§</sup>Department Microbe-Plant Interactions, Helmholtz Centre for Environmental Research, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany  
<sup>#</sup>Institute of Chemistry, Food Chemistry and Environmental Chemistry, University of Applied Sciences, Kurt-Mothes-Strasse 2, D-06120 Halle (Saale), Germany

**Biotransformation of Acyclovir and Penciclovir in Activated Sludge**  
 ...ernes<sup>\*,†</sup>  
 ... Landau, Germany  
 ... many

**Introduce journal requirements to submit spectra electronically to repositories?**



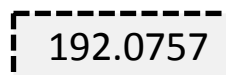
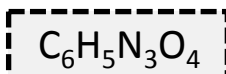
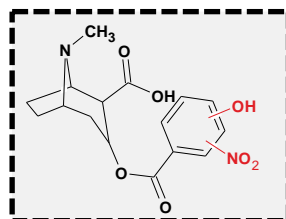
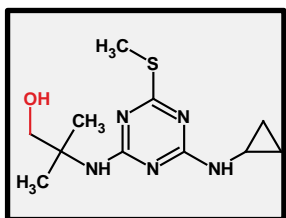
...ation of the two antiviral drugs, Acyclovir (ACV) and Penciclovir (PCV), was investigated in activated sludge. Biodegradation kinetics were determined, and transformation products (TPs) were identified using Hybrid Linear Ion Trap- FT Mass Spectrometry (LTQ Orbitrap Velos) and 1D (<sup>1</sup>H NMR, <sup>13</sup>C NMR) and 2D (<sup>1</sup>H,<sup>1</sup>H-COSY, <sup>1</sup>H-<sup>13</sup>C-HSQC) NMR Spectroscopy. ACV and PCV rapidly dissipated in the activated sludge batch systems with half-lives of 5.3 and 3.4 h and first-order rate constants in relation to the amount of suspended solids (SS) of  $4.9 \pm 0.1 \text{ L g}_{\text{SS}}^{-1} \text{ d}^{-1}$  and  $7.6 \pm 0.3 \text{ L g}_{\text{SS}}^{-1} \text{ d}^{-1}$ , respectively. For ACV only a single TP was found, whereas eight TPs were identified for PCV. Structural elucidation of TPs exhibited that transformation only took place at the side chain leaving the guanine moiety unaltered. The oxidation of the primary hydroxyl group in ACV resulted in the formation of carboxy-acyclovir (Carboxy-ACV). For PCV, transformation was more diverse with several enzymatic reactions taking place such as the oxidation of terminal hydroxyl groups and  $\beta$ -oxidation followed by acetate cleavage. Analysis of different environmental samples revealed the presence of Carboxy-ACV in surface and drinking water with concentrations up to  $3200 \text{ ng L}^{-1}$  and  $40 \text{ ng L}^{-1}$ , respectively.



# User Contribution of Spectra as Supporting Information

What about the tentatively identified and unknown spectra?

Example



*Identification confidence*

*Minimum data requirements*

**Level 1: Confirmed structure**  
by reference standard

MS, MS<sup>2</sup>, RT, Reference Std.

**Level 2: Probable structure**  
a) by library spectrum match  
b) by diagnostic evidence

MS, MS<sup>2</sup>, Library MS<sup>2</sup>  
MS, MS<sup>2</sup>, Exp. data

**Level 3: Tentative candidate(s)**  
structure, substituent, class

MS, MS<sup>2</sup>, Exp. data

**Level 4: Unequivocal molecular formula**

MS isotope/adduct


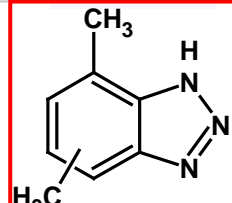
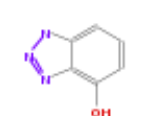
**Level 5: Exact mass** of interest

MS

# Tentative and Unknown Spectra

Need new ways to display the “uncertainty” correctly

## Molecular formula only

<p>1H-Benzotriazole TP9 (Tentative)</p> <p>LC-ESI-ITFT; MS2; 80-110; R=7500; [M+H]<sup>+</sup></p>	<p>1 spectrum</p> <p><b>C5H5N3O3</b></p> <p>Not Available</p>	<p>155.03300</p>	<p>ETS00105</p>
<p>1H-Benzotriazole-4-carboxylic acid (Tentative)</p> <p>LC-ESI-ITFT; MS2; 80-110; R=7500; [M+H]<sup>+</sup></p>	<p>1 spectrum</p> <p><b>C7H5N3O2</b></p> 	<p>163.03820</p>	<p>ETS00107</p>
<p>4,(6or7)-dimethyl benzotriazole (Tentative)</p> <p>LC-ESI-ITFT; MS2; 80-110; R=7500; [M+H]<sup>+</sup></p>	<p>1 spectrum</p> <p><b>C8H9N3</b></p> 	<p>147.07960</p>	<p>ETS00100</p>
<p>4-Hydroxybenzotriazole</p> <p>LC-ESI-ITFT; MS2; 80-110; R=7500; [M+H]<sup>+</sup></p>	<p>1 spectrum</p> <p><b>C6H5N3O</b></p> 	<p>135.04330</p>	<p>ETS00119</p>

Not (yet) possible to display Markush structures

# Exchanging Tentative and Unknown Spectra

- Need to be able to save and display “uncertain” structures
  - Possible in some applications; not universal
  
- Need an InChI form to express “uncertain” structures<sup>1</sup>
  - InChI is main exchange form of batch database retrieval services
  - Not yet possible – working group founded to address this
  
- Compromise: encouraging exchange vs. maintaining quality
  
- ***Enter a new era of “crowdsourcing” or “social mass spectrometry”?***<sup>2</sup>

<sup>1</sup>Creek, Dunn, Fiehn, Griffin, Hall, Lei, Mistrik, Neumann, Schymanski, Sumner, Trengove & Wolfender, 2014, *Metabolomics*, 10:350–353; DOI 10.1007/s11306-014-0656-8

<sup>2</sup>Towards Open Access Mass Spectral Libraries Workshop, IMSC, Geneva, Aug. 2014. Hosts: Steve Stein & Enrico Davoli



# Summary: Towards Open Access Mass Spectral Libraries

## MassBank ↔ NIST

- Make MassBank spectra more accessible to more users
  - Download public dataset – but not much software uses MassBank format
- NIST
  - MSP is a more “accepted” and used format
  - Already incorporated in many vendor software systems
  - A commercial, reference library – cannot accept distribute our spectra
  - BUT: can provide MassBank in NIST format for easy download
- Cooperation with Steve Stein (NIST) at the IMSC, Geneva
  - “massbank\_ei” – all EI spectra up to July 2012
  - “massbank\_msms” – all MS/MS spectra up to July 2012
  - “massbank\_eawag” – all Eawag spectra (basis for further collaboration)
- Very easy install for NIST users!

# Summary: Towards Open Access Mass Spectral Libraries

MassBank ↔ NIST

ATENOLOLMH45P2672
Clear
a-z
massbank\_eawag

Asulam [M-H]- 15% P=229

Asulam [M-H]- 30% P=229

Asulam [M-H]- 30% P=229

Asulam [M-H]- 35% P=229

Asulam [M-H]- 35% P=229

Asulam [M-H]- 45% P=229

Asulam [M-H]- 45% P=229

Asulam [M-H]- 60% P=229

Asulam [M-H]- 60% P=229

Atenolol [M+H]+ 15% P=267.2

Atenolol [M+H]+ 15% P=267.2

Atenolol [M+H]+ 30% P=267.2

Atenolol [M+H]+ 30% P=267.2

Atenolol [M+H]+ 35% P=267.2

Atenolol [M+H]+ 35% P=267.2

Atenolol [M+H]+ 45% P=267.2

Atenolol [M+H]+ 45% P=267.2

Atenolol [M+H]+ 60% P=267.2

Atenolol [M+H]+ 60% P=267.2

Atenolol [M+H]+ 75% P=267.2

Atenolol [M+H]+ 75% P=267.2

Atenolol [M+H]+ 90% P=267.2

Atenolol [M+H]+ 90% P=267.2

Atenolol acid [M+H]+ 15% P=268.2

Atenolol acid [M+H]+ 15% P=268.2

Atenolol acid [M+H]+ 30% P=268.2

Atenolol acid [M+H]+ 30% P=268.2

Atenolol acid [M+H]+ 35% P=268.2

Atenolol acid [M+H]+ 35% P=268.2

Atenolol acid [M+H]+ 45% P=268.2

Atenolol acid [M+H]+ 45% P=268.2

Atenolol acid [M+H]+ 60% P=268.2

Atenolol acid [M+H]+ 60% P=268.2

Atenolol acid [M+H]+ 75% P=268.2

Atenolol acid [M+H]+ 75% P=268.2

Atenolol acid [M+H]+ 90% P=268.2

Atenolol acid [M+H]+ 90% P=268.2

(massbank\_eawag) Atenolol

**Name:** Atenolol  
**Formula:** C<sub>14</sub>H<sub>22</sub>N<sub>2</sub>O<sub>3</sub>  
**MW:** 266 **Exact Mass:** 266.163 **CAS#:** 29122-68-7 **ID#:** 1042 **DB:** massbank\_eawag  
**Other DBs:** None  
**Contributor:** Stravs M, Schymanski E, Singer H, Department of Environmental Chemistry, Eawag  
**Comment:** ID=EA016904 License="CC BY-SA" Record\_title="Atenolol; LC-ESI-ITFT; MS2; CE: 45%; R=7500; [M+H]+" RT="2.0 min" D  
**AUX:** CC(C)NCC(O)COc1ccc(cc1)CC(N)=O  
**Collision energy:** 45%  
**Instrument:** LTQ Orbitrap XL Thermo Scientific  
**Instrument type:** LC-ESI-ITFT  
**Precursor m/z:** 267.1703  
**Precursor type:** [M+H]<sup>+</sup>  
**Ion mode:** P  
**Spectrum type:** ms2  
**10 largest peaks:**  
267.17 999 | 190.086 764 | 145.065 434 | 74.06 383 | 116.107 328 |

Names
Structures

Lib. Search
Other Search
Names
Compare
Librarian
MSMS

# Summary: Towards Open Access Mass Spectral Libraries

NIST14 and "SOS" (*images c/o Manor Askenazi*)

The screenshot displays a mass spectrometry software interface with the following components:

- Top Panel:** Shows the search name "1. MYSTERY\_MOL" and various icons for file operations and zooming.
- Table 1 (Source Scans):**

#	Src.	Name
1	A	MYSTERY_MOL
3	A	Scan 1467 (28.976 min) of JAN310301003.d 10149
4	A	Scan 1573 (30.855 min) of JAN310501005.d 10156
5	A	Scan 1488 (29.348 min) of JAN210101001.d 10164
6	A	Scan 1339 (26.715 min) of JAN210401004.d 10167
- Mass Spectrum Plot (Top):** Shows relative intensity vs. m/z. Major peaks are labeled at m/z 43, 71, 99, and 317. A question mark is placed above the spectrum.
- Table 2 (Search Results):**

#	Lib.	Match	R.Match	Prob. (%)	RI	Name
1	R	450	690	4.28	2...	Octadecane
2	M	448	682	3.95	-	Sulfurous
3	R	440	673	2.94	1...	Dodecane
4	M	439	662	2.83	-	Pentadecane
5	R	436	689	2.50	1...	Dodecane
6	R	434	683	2.30	2...	Heptacosane
7	M	434	637	2.30	-	3,3-Diethyl
8	M	433	719	2.22	1...	Decane, 2
- Mass Spectrum Plot (Middle):** Shows a reference mass spectrum for "MYSTERY\_MOL" with peaks at m/z 43, 57, 71, 85, 99, 113, 141, and 155.
- Mass Spectrum Plot (Bottom):** Shows a reference mass spectrum for "(replib) Octadecane, 1-iodo-" with peaks at m/z 43, 57, 71, 85, 99, 155, 197, and 253.
- Table 3 (10 Largest Peaks):**

57 999	43 849	71 631	41 566
55 399	29 269	69 158	99 144
- Table 4 (Synonyms):**

1. Octadecyl iodide
2. Stearal iodide
- Context Menu:** A right-click menu is open over the top spectrum, with the "SOS" option highlighted. Other options include "Library Search", "Structure Similarity Search", "Zoom Out", "Neutral Loss Display", "Set Anchors", "Copy Plot", "Copy", "Paste", "Send To", "Copy Structure to Clipboard", "Attach Clipboard Structure", "MW Estimation", "CI and/or Br Estimation", "Print", "Print Preview", "Change Splitter Orientation", and "Properties".
- Bottom Panel:** Shows navigation buttons for "Lib. Search", "Other Search", "Names", "Compare", "Librarian", and "MSMS".

# Summary: Towards Open Access Mass Spectral Libraries

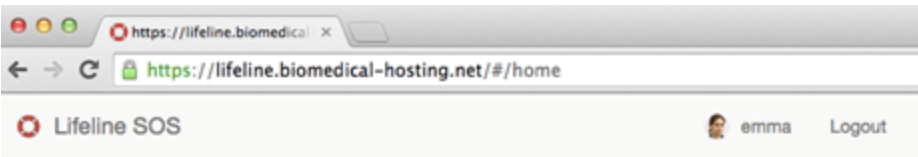
NIST14 and “SOS” (*images c/o Manor Askenazi*)

The screenshot shows the Lifeline SOS web application interface. At the top left, there is a logo for Lifeline SOS. At the top right, there is a user profile for 'manor' with options for 'Logout' and 'Help'. Below the logo, there is a circular profile picture of a man with a beard, surrounded by a red lifebuoy. To the right of the profile picture, the text reads 'Lifeline: Social Online Spectrometry'. In the center of the page, there is a network diagram consisting of several nodes and connections. The nodes are represented by blue and orange circles, with some larger blue circles representing spectral clusters. A legend in the bottom left corner explains the colors: a light blue circle for 'My spectra!', a dark blue circle for 'Spectral Cluster (Group)', and an orange circle for 'Someone else's submissions...'. At the bottom of the page, there is a URL: [https://lifeline.biomedical-hosting.net/#/group/GROUP\\_813985937](https://lifeline.biomedical-hosting.net/#/group/GROUP_813985937).

# Summary: Towards Open A

## NIST14 and "SOS"

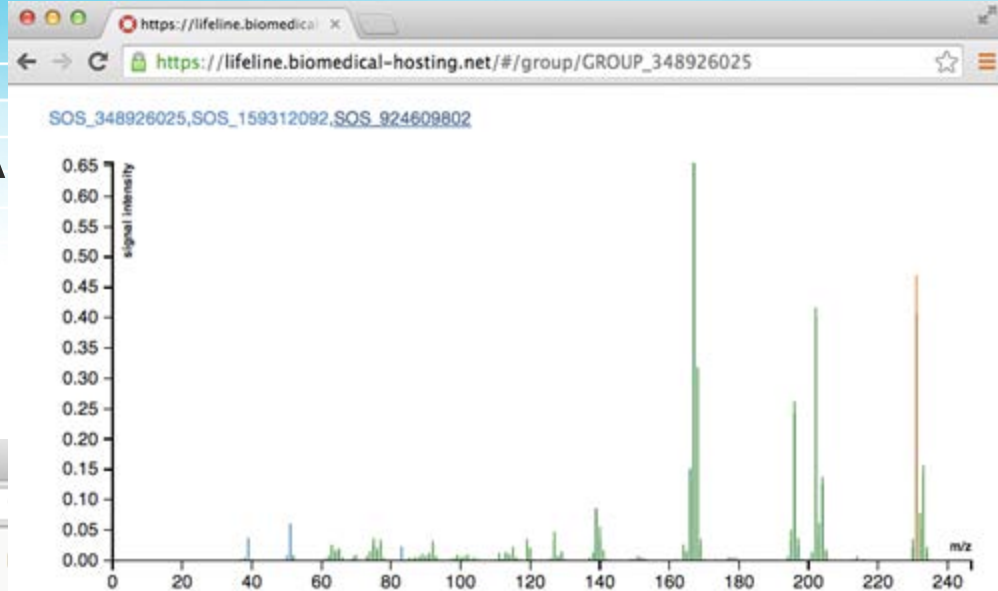
(images c/o Manor Askenazi)



Lifeline: **S**ocial **O**nline **S**pectrometry



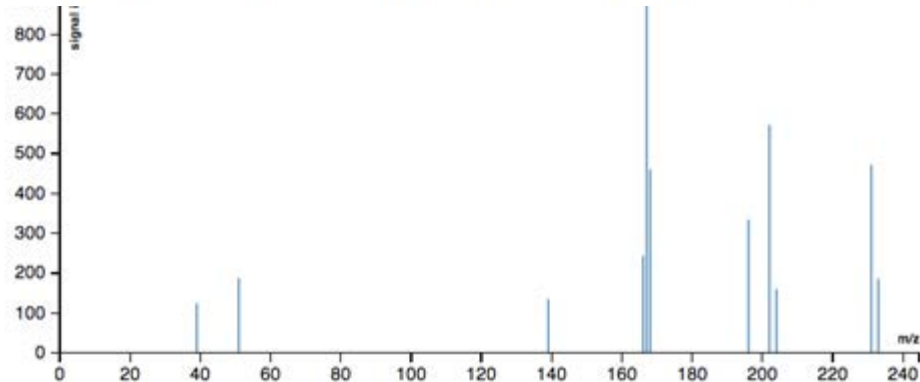
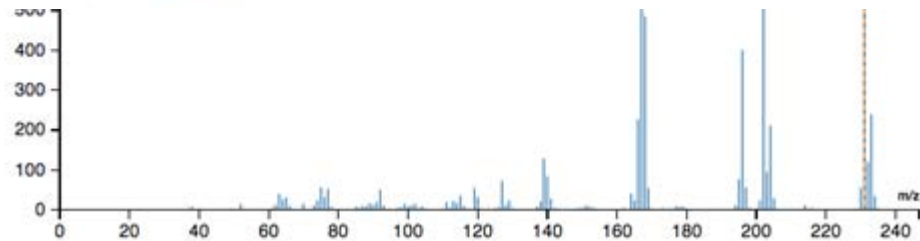
- My spectra!
- Spectral Cluster (Group)
- Someone else's submissions...



### Comments/Thoughts/Hypotheses:

We're synthesizing the standards now...

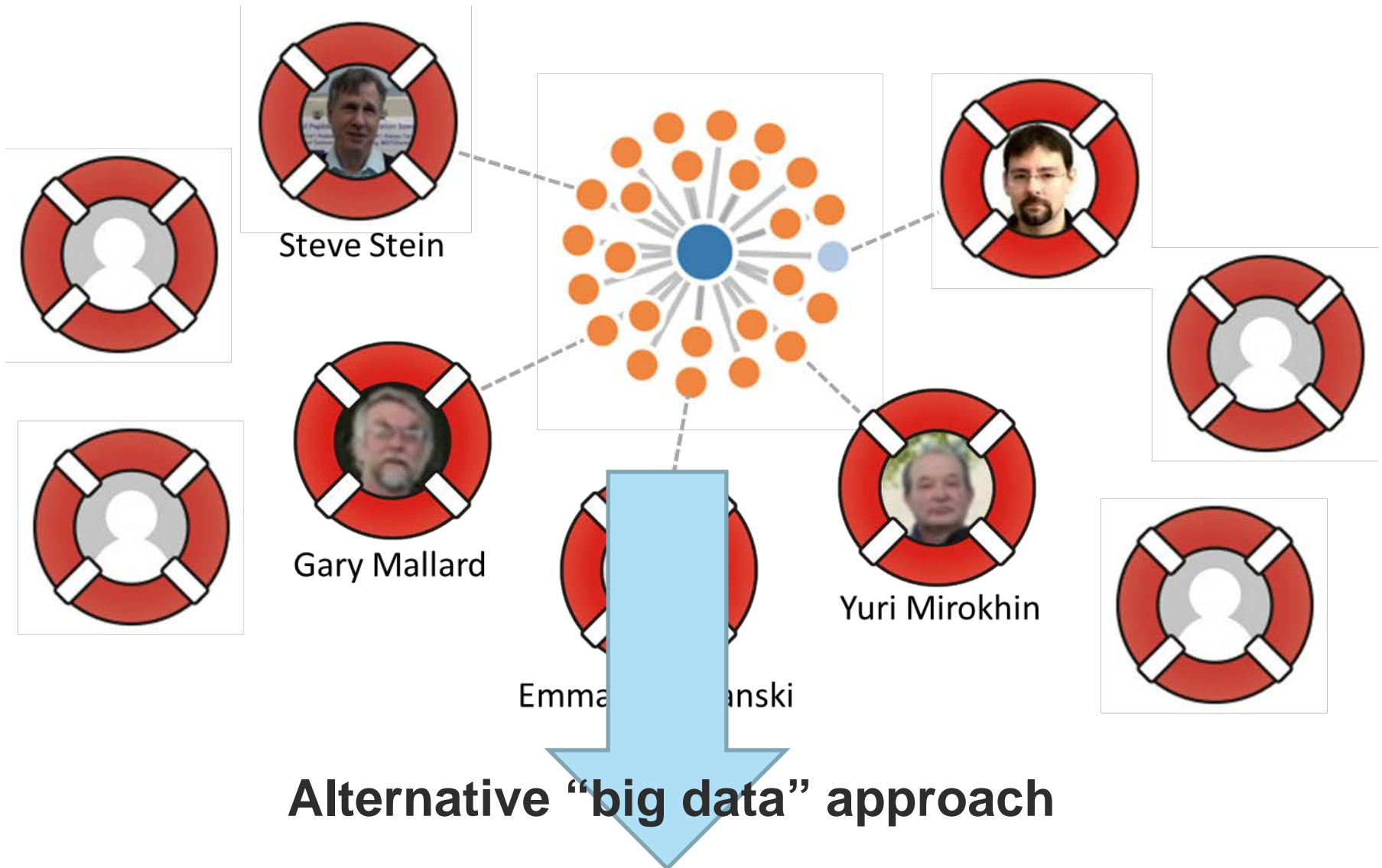
posted by emma (delete)





# Summary: Towards Open Access Mass Spectral Libraries

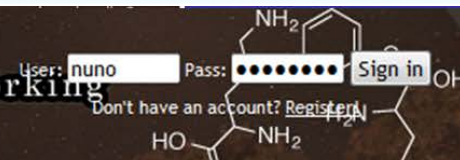
NIST14 and “SOS” – *Expert, Manual Co-Curation* (*images Manor Askenazi*)



# Summary: Towards Open Access Mass Spectral Libraries

## GnPS: Global Natural Products Social Molecular Networking

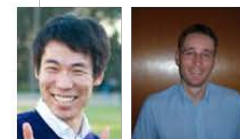
MassIVE Datasets | General Info | UCSD Proteomics | Future Tools | Demo | Contact



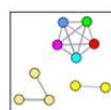
[gnps.ucsd.edu](http://gnps.ucsd.edu)

First MassIVE Knowledge Base, open March 2014

Scan	Precursor Ion	RT	Parent Mass
4238	52981.000	2409.300	445.130
4138	51753.000	2406.120	445.130
4239	53027.000	2401.010	445.130
4139	50048.000	2531.740	445.121
4140	51463.000	2372.360	445.130
4141	51112.000	2372.360	445.130
4142	51113.000	2370.870	445.130
4143	51114.000	2370.870	445.131
4144	51115.000	2370.870	445.130



Co-analyze private+public data



### Data Analysis

The [Data Analysis](#) portal will allow you to organize and visualize your mass spectrometry data. Leveraging the molecular networking techniques, there are additional tools to aid in understanding the unknowns in your sample. For a documentation click [here](#). Further, a separate [dereplication workflow](#) is provided as a standalone workflow.

### Create Public MassIVE Datasets

[Submit](#) your own data to be made public MassIVE datasets. These MassIVE datasets must be prefixed with GNPS to be visible to other GNPS users. Take advantage of continuous identification to learn more about your dataset after publication automatically. New hits to the community curated libraries and related datasets are reported. [Documentation](#)



Share data

### Contribute to Libraries



Be a part of the collaborative community effort to [create](#) the definitive collection of natural products MS/MS spectra. GNPS gives the power to add spectra, update annotations, and facilitate dialog around these spectra, to provide a truly collaborative and open natural product MS/MS database. For documentation and definition of quality requirements click [here](#). To make corrections to and comments on existing spectra in the libraries, users should refer to [this documentation](#).

### MassIVE Public GNPS Datasets



[Browse](#) publicly available datasets. Here you can download these datasets as well as comment on them so others in the community can see any updates or any new analysis. Additionally, users can subscribe to the datasets and get updates when new identifications are made via GNPS's continuous identification. To read further on how to take advantage of the subscriptions to MassIVE datasets and other social networking features [click here](#).

Crowdsourced curated libraries



### Browse Community Spectral Library



[Browse](#) the community contributed and community curated spectral libraries of natural products. These MS/MS libraries are community contributed and community curated. Users can peak at the inside of these libraries, as well as use them for data analysis. If corrections need to be made, users should refer to [this documentation](#).

### Molecule Explorer



Bridge the connection between molecules and datasets. [Explore](#) exactly where certain molecules are found in all the publicly available dataset at GNPS. Powered by GNPS's continuous identifications, users are able to see not only which datasets contain what compound, but how many known and unknown analogs exist in all datasets!

Explore unknown molecules

GNPS & Crowdsourcing Annotation of MS

Slides c/o Nuno Bandeira

# Summary: Towards Open Access Mass Spectral Libraries

GNPS & Crowdsourcing Annotation of MS (*c/o Nuno Bandeira*)

*Since March 2014*

- *Over 130 datasets,*
- *>1,200 users*
- *Data from **30+** countries*
- *>400M spectra, **11,000+** runs, approaching 1 TB of data*

## Public spectral libraries

- *Antibiotics and natural products*
- *FDA approved drugs*
- *Over 8,500 compound spectra*
  - *5,000 more running now*
- *Dozens of curators*
  - *Gold, Silver & Bronze levels*
- *Creating specialized communities*

## Previously in natural products

- *Hardly any spectra publicly available*
- *No available spectral libraries*



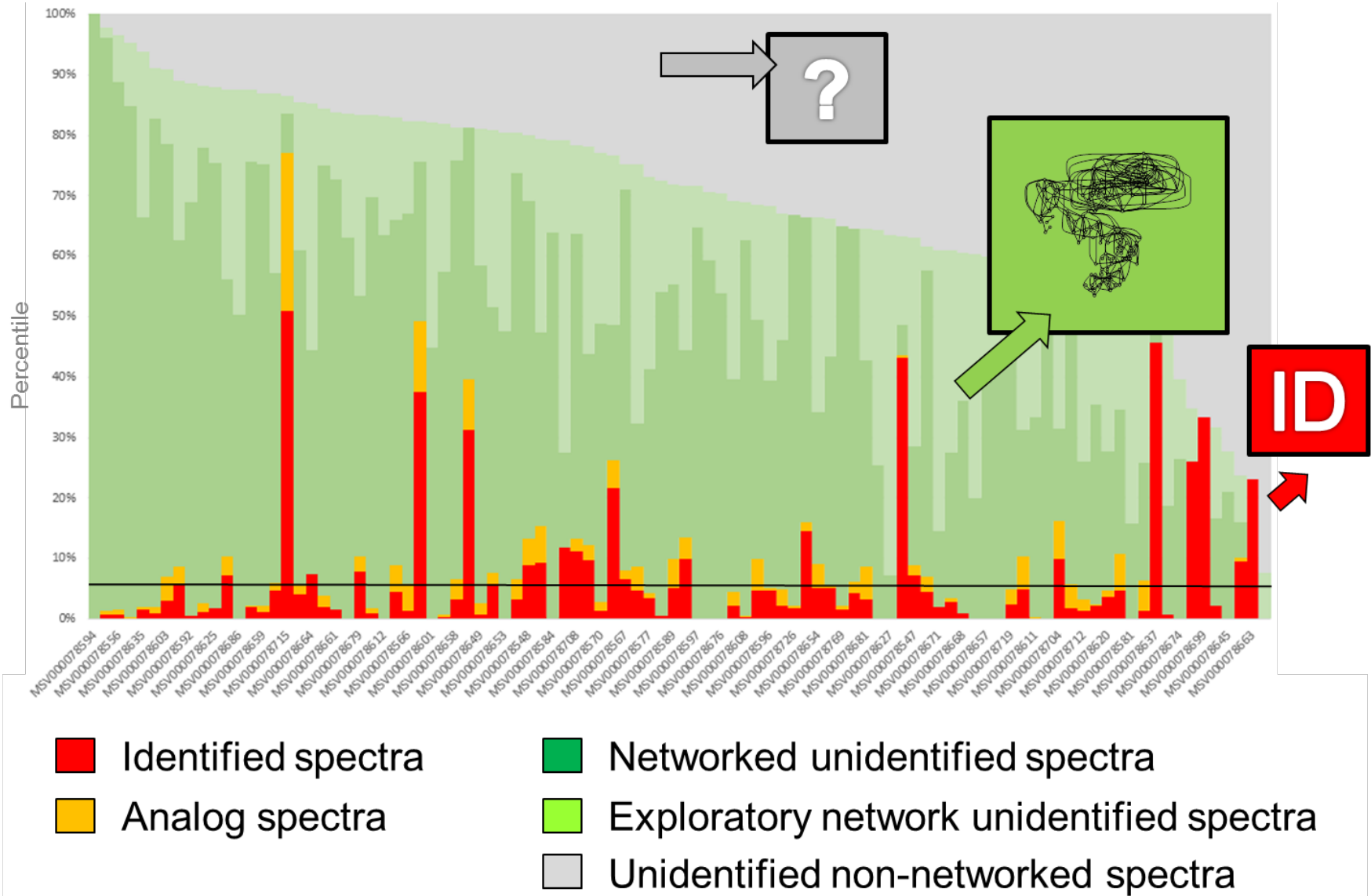
## Quality levels:

- *Gold: approved users, synthetic compounds*
- *Silver: published compounds*
- *Bronze: everything else*



# Summary: Towards Open Access Mass Spectral Libraries

GNPS & Crowdsourcing Annotation of MS (*c/o Nuno Bandeira*)



# Summary: Towards Open Access Mass Spectral Libraries

GNPS & Crowdsourcing - Email updates! (*c/o Nuno Bandeira*)

[GNPS] Updated Continuous Identification Results for MassIVE Datasets 06-02-2014 - Message (Plain Te...

File Message PDF Architect

Ignore X Delete Reply Reply All Forward Junk Delete

All Mail - bandei... To Manager Team E-mail

Move Mark Unread Follow Up

Translate Zoom Add to Evernote 5 Evernote

This message has extra line breaks.

From:  ccms@ucsd.edu Sent: Mon 6/2/2014 3:08 PM

To:  bandeira@ucsd.edu

Cc:

Subject: [GNPS] Updated Continuous Identification Results for MassIVE Datasets 06-02-2014

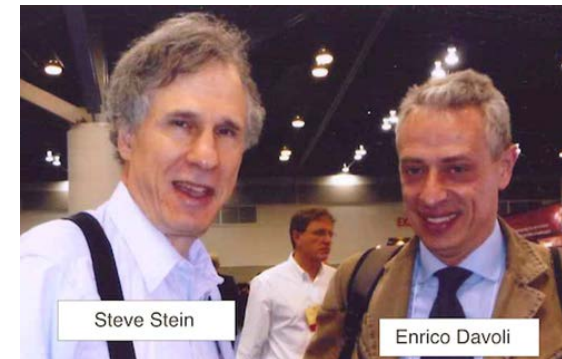
MassIVE Dataset MSV000078556 GNPS - Topobiographical molecular analysis of the skin from a man and a woman. MS/MS Data Reporting 22.051283% more IDs and 43 more IDs  
1.5384616% different IDs and 3 different IDs  
0.0% deleted IDs and 0 deleted IDs  
with 195 total

Results are available at:  
<http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=ed79a3b5652d4f55b839d4a774c6e8cd>

The dataset is available at:  
[http://gnps.ucsd.edu/ProteoSAFe/result.jsp?task=6b9dcff3899e4d5f89f0daf9489a3a5e&view=group\\_all\\_annotations](http://gnps.ucsd.edu/ProteoSAFe/result.jsp?task=6b9dcff3899e4d5f89f0daf9489a3a5e&view=group_all_annotations)



# Acknowledgements

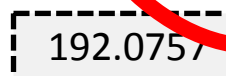
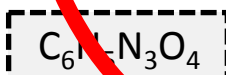
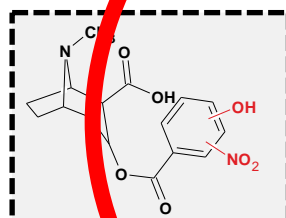
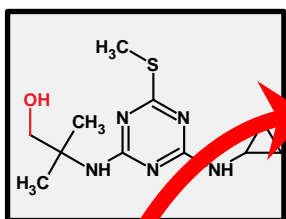




# Identification with (LC-HR)MS/MS – Confidence?

Proposed levels for MS and MS/MS data

Example



Identification confidence

Minimum data requirements

**Level 1: Confirmed structure**  
by reference standard

MS, MS<sup>2</sup>, RT, Reference Std.

**Level 2: Probable structure**  
a) by library spectrum match  
b) by diagnostic evidence

MS, MS<sup>2</sup>, Library MS<sup>2</sup>  
MS, MS<sup>2</sup>, Exp. data

**Level 3: Tentative candidate(s)**  
structure, substituent, class

MS, MS<sup>2</sup>, Exp. data

**Level 4: Unequivocal molecular formula**

MS isotope/adduct

**Level 5: Exact mass of interest**

MS

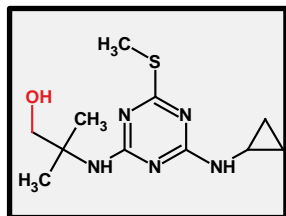
# Identification with (LC-HR)MS/MS – Confidence?

*Any questions?*

Example

Identification confidence

Minimum data requirements

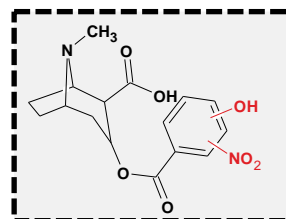


**Level 1: Confirmed structure**  
by reference standard

MS, MS<sup>2</sup>, RT, Reference Std.

**Level 2: Probable structure**  
a) by library spectrum match  
b) by diagnostic evidence

MS, MS<sup>2</sup>, Library MS<sup>2</sup>  
MS, MS<sup>2</sup>, Exp. data



**Level 3: Tentative candidate(s)**  
structure, substituent, class

MS, MS<sup>2</sup>, Exp. data

$C_6H_5N_3O_4$

**Level 4: Unequivocal molecular formula**

MS isotope/adduct

192.0757

**Level 5: Exact mass** of interest

MS