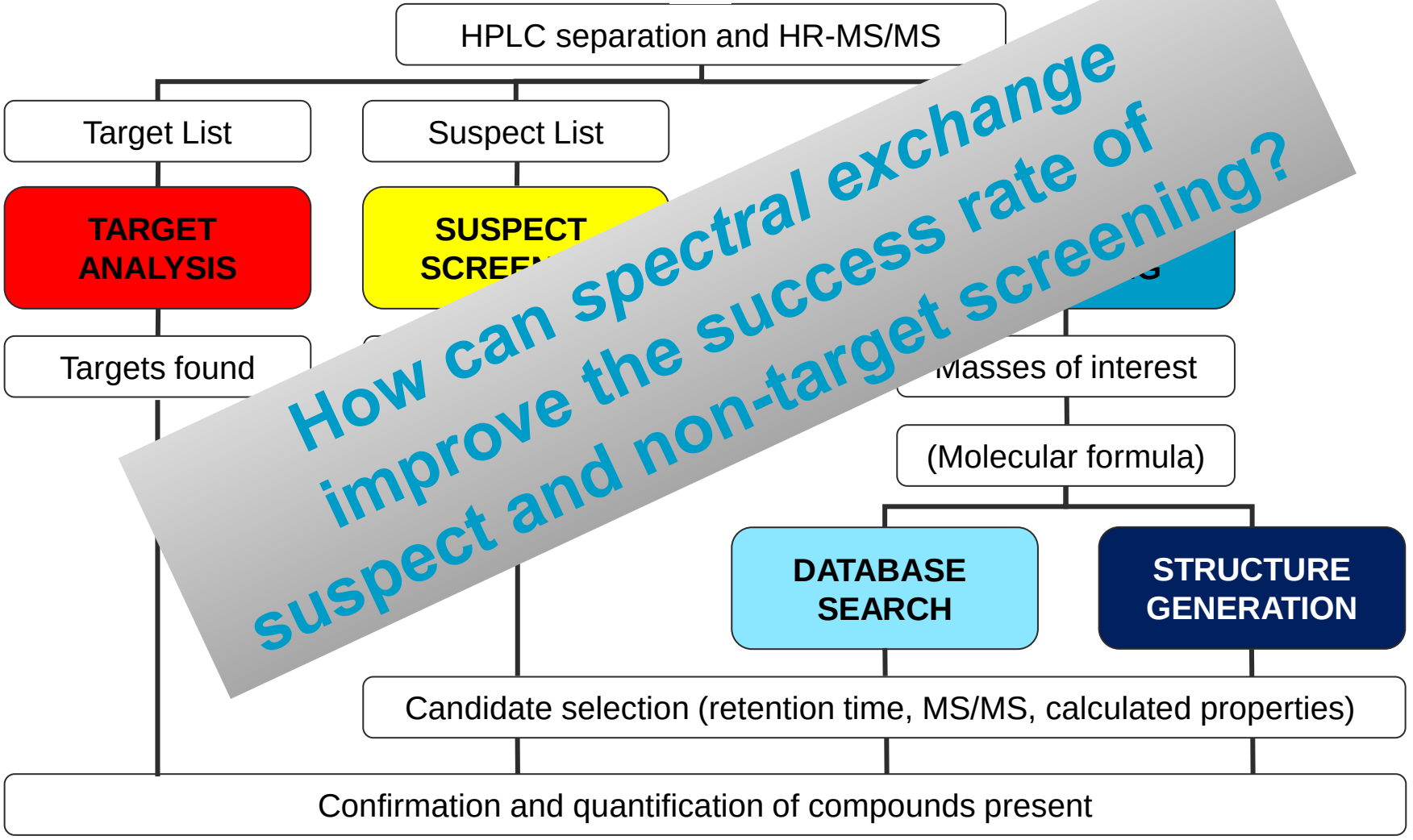
Enhancing Spectral Exchange: Tentative and Unknown Spectra, NIST and Crowdsourcing Trends

NORMAN MassBank Workshop; 17-18 September 2014

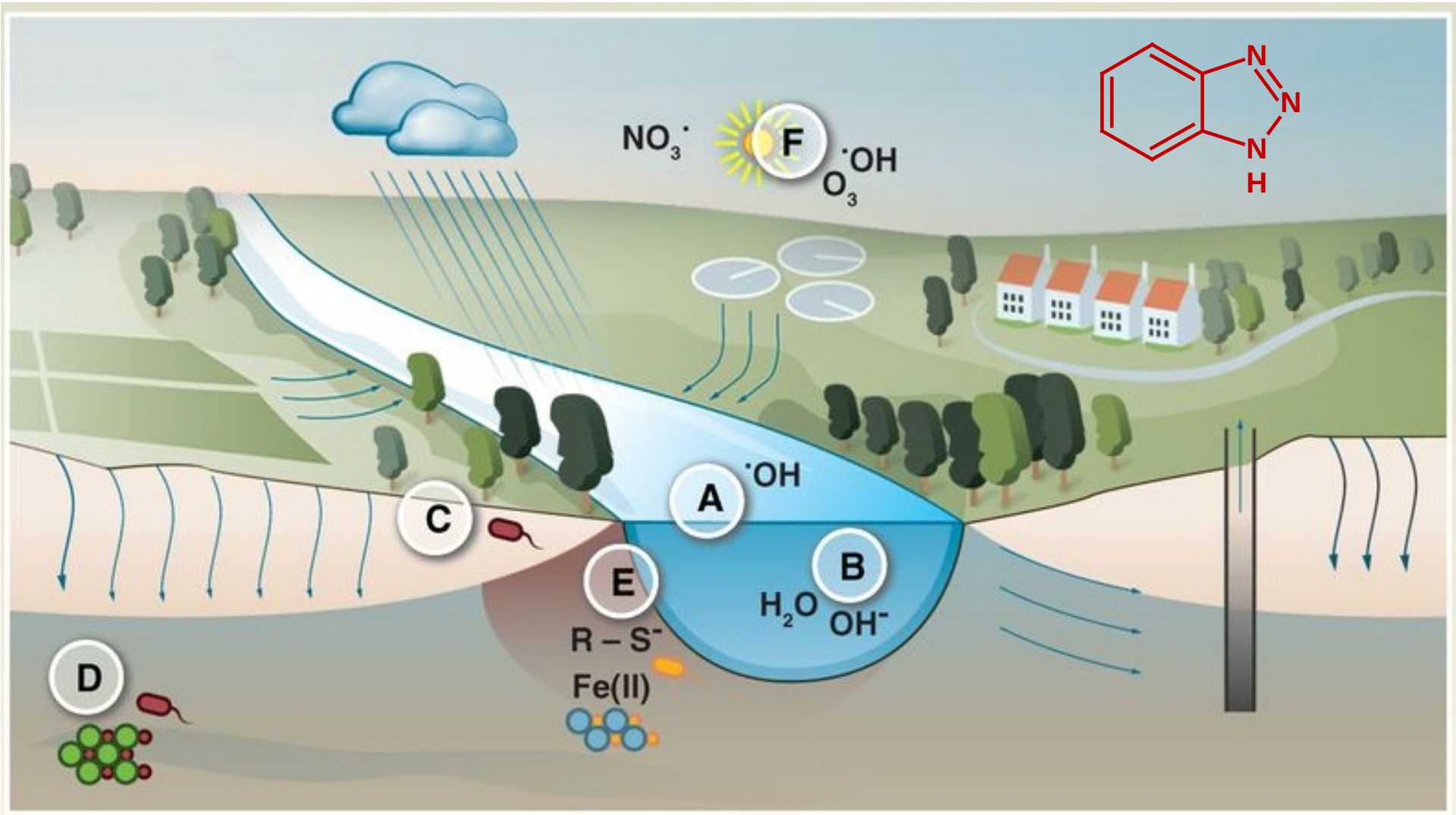
Emma Schymanski
Juliane Hollender, Heinz Singer und Environmental Chemistry Dept.
Eawag: Swiss Federal Institute of Aquatic Science and Technology
Dübendorf, Switzerland

Status quo of identification approaches



The Reference Standard Dilemma[1]

…and why identification in the environment is such a challenge



Source: Fenner *et al. (*2013) *Science*, 341(6147), 752-758;

[1]Moschet *et al.* (2013) *Analytical Chemistry*, 85(21) 10312-10320

Accessing Mass Spectral Information

(Too) Many ways to access mass spectral information

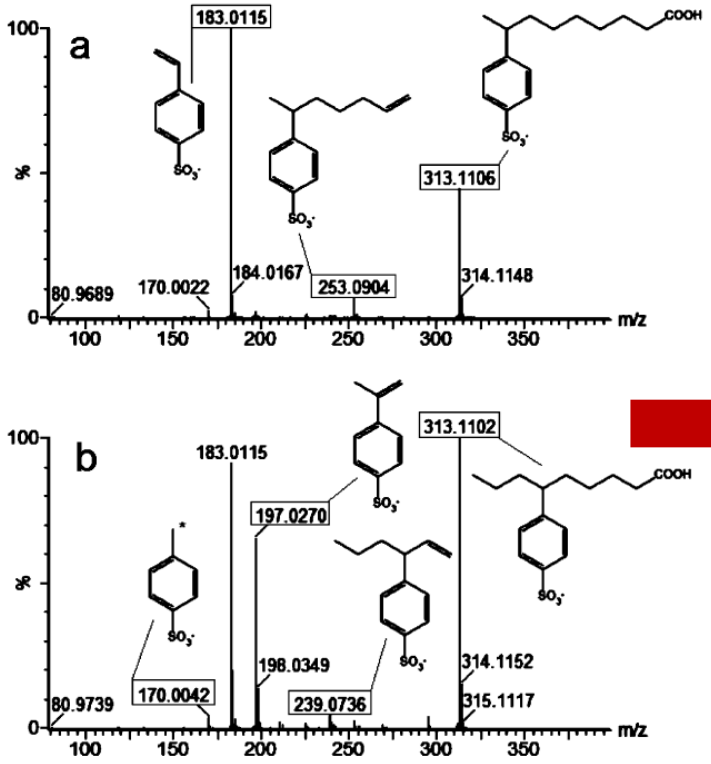
- Mass Spectral Libraries
  - Curated, reference standard spectra: NIST, Wiley, mzCloud, METLIN
  - User-defined personal libraries
- Mass Spectral Repositories
  - User-contributed reference spectra: MassBank, HMDB, GOLM, GNPS
- Literature sources
  - Contain reference standard and elucidated spectra,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

- ***Great need to improve the accessibility of spectral information!***

Accessing Literature Spectra

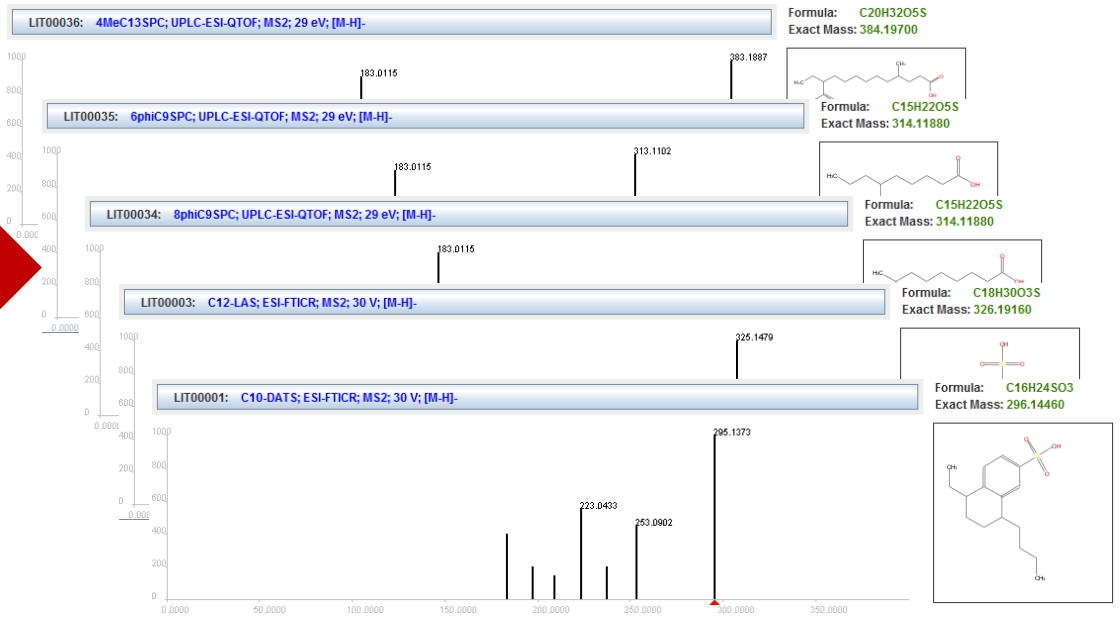
Enable access and comparison

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



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

39 literature spectra (so far)



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

User Contribution of Spectra as Supporting Information

Index Type :
Contributor: **Eawag Additional Mass Spectra**

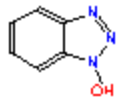
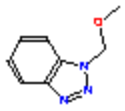
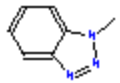
Edit / Resubmit Query

Results : 50 Hit. (1 - 27 Displayed)

Open All Tree | Multiple Display | Spectrum Search

First Prev **1** 2 Next Last (Total **2** Page)

▼ Results End

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

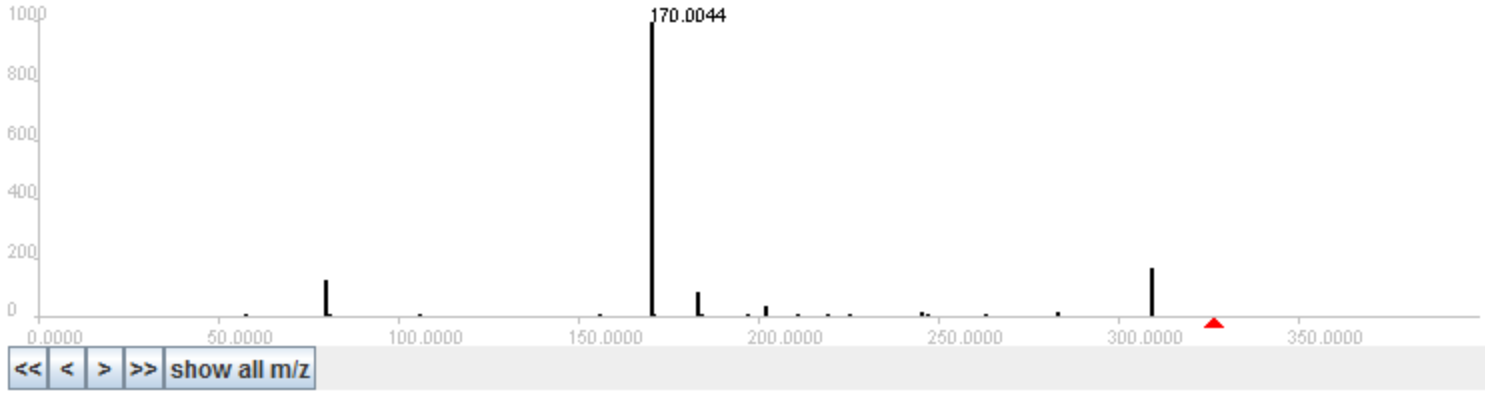
User Contribution of Spectra as Supporting Information

MassBank Record: ETS00019

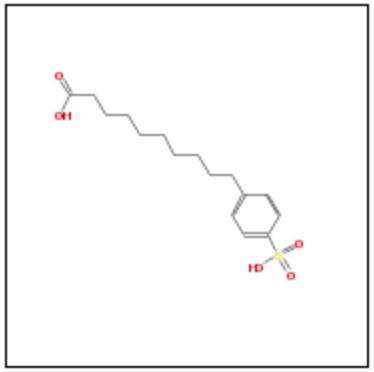
Home | Spectrum | Quick | Peak | Substructure | Browser | Batch | Browse | Index | MassBank ID: Go

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) Enviro
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)
```

eawag aquatic research

User Contribution of Spectra as Supporting Information

ENVIRONMENTAL Science & Technology

ENVIRONMENTAL Science & Technology

Article

pubs.acs.org/est

Strategies to Characterize Wastewater: Exploring the Spectrometry

Emma L. Schymanski,[†] Heinz P. Singer and Michael A. Stravs,[†,§] Cristina Ripollés V

[†]Eawag: Swiss Federal Institute of Aquatic Scienc

[‡]University Jaume I, Department of Physical and

[§]Institute of Biogeochemistry and Pollutant Dyna

Supporting Information

ABSTRACT: Wastewater effluents contain a contaminants and transformation products, whic target analysis alone. High accuracy, high resolu data were explored with novel untargeted data (enviMass, nontarget, and RMassBank) to co target analysis in initial "all in one" measureme the detected peaks from 10 Swiss wastewater were assigned to target compounds, with 3 available. Corrosion inhibitors, artificial sweeten exhibited the highest concentrations. After blan 70% of the peaks remained and were grouped i these components had adduct and/or isotope i intensity-based prioritization revealed that only 4 top 30 most intense peaks (negative mode), tentatively identified via suspect screening for s benzothiazole-2-sulfonate, an oxidation product tailor-made nontarget processing methods (all a of heteroatom-containing compounds in the en

Biotransformation of Benzotriazoles: Insights from Transformation Product Identification and Compound-Specific Isotope Analysis

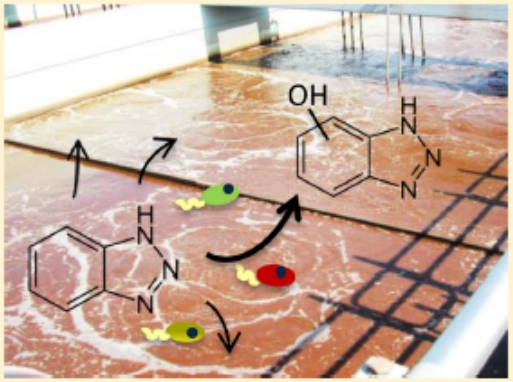
Sebastian Huntscha,[†,‡] Thomas B. Hofstetter,[†,‡] Emma L. Schymanski,[†] Stephanie Spahr,[†] and Juliane Hollender*[,†,‡]

[†]Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

[‡]Institute of Biogeochemistry and Pollutant Dynamics (IBP), ETH Zurich, 8092 Zurich, Switzerland

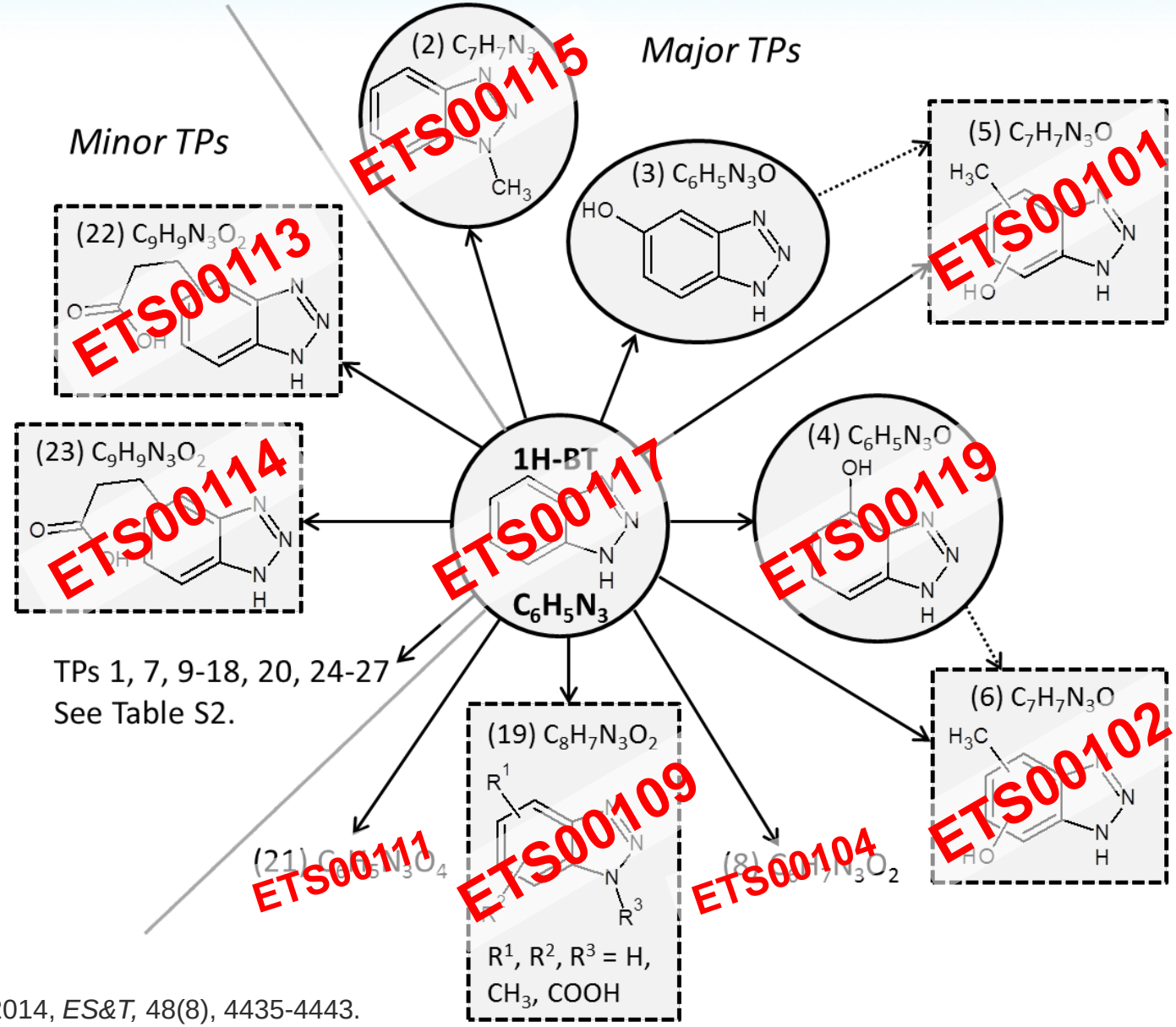
Supporting Information

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



Huntscha *et al.* 2014, *ES&T,* 48(8), 4435-4443.

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

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

André Macherius,[†] Bettina Seiwert,[†] Peter S
and Thorsten Reemtsma*,[†]

[†]Department of Analytical Chemistry, Helmholtz Centr
Germany

[§]Department Microbe-Plant Interactions, Helmholtz Ce
Ingolstädter Landstraße 1, D-85764 Neuherberg, Germ

[#]Institute of Chemistry, Food Chemistry and En
Kurt-Mothes-Strasse 2, D-06120 Halle (S

ENVIRONMENTAL Science & Technology

ARTICLE

pubs.acs.org/est

Biotransform
in Activ

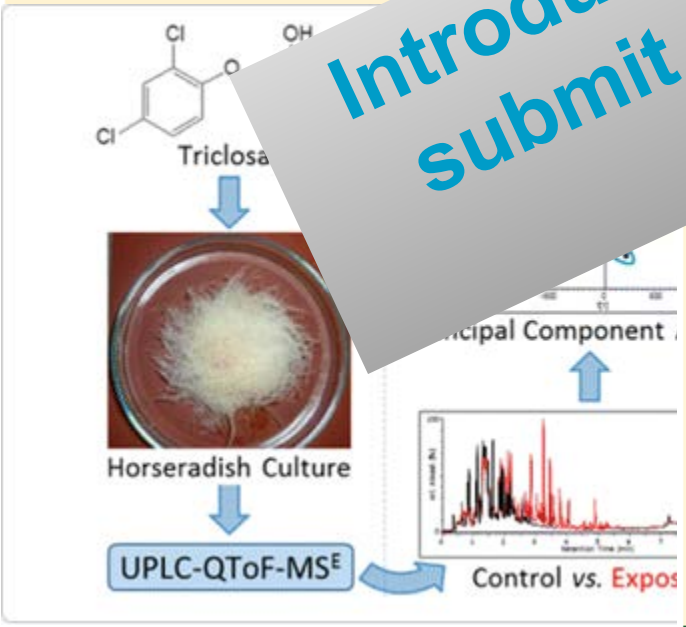
enciclovir

ernes*,[†]

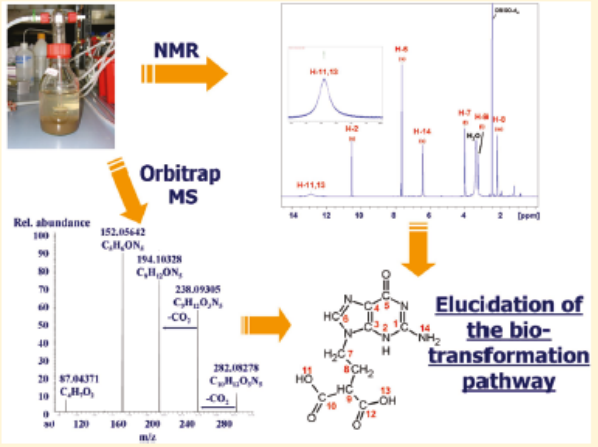
au, Landau, Germany

any

Introduce journal requirements to submit spectra electronically to respositories?

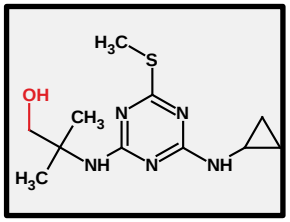
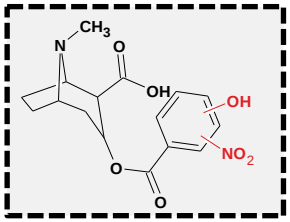


mation of the two antiviral drugs, enciclovir (PCV), was investigated in vated sludge. Biodegradation kinetics were , and transformation products (TPs) were identified Hybrid Linear Ion Trap- FT Mass Spectrometry (LTQ Orbitrap Velos) and 1D (^{1}H NMR, ^{13}C NMR) and 2D (^{1}H,^{1}H-COSY, ^{1}H-^{13}C-HSQC) 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 ± 0.1 L g_{ss}^{-1} d^{-1} and 7.6 ± 0.3 L g_{ss}^{-1} 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 β-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 ng L^{-1} and 40 ng L^{-1}, respectively.



User Contribution of Spectra as Supporting Information

What about the tentatively identified and unknown spectra?

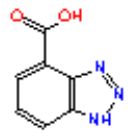
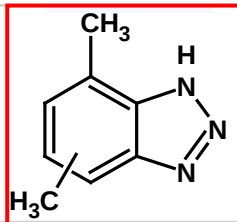
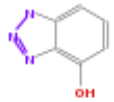
Example	*Identification confidence*	*Minimum data requirements*
	***Level 1: Confirmed structure*** by reference standard	MS, MS^2, RT, Reference Std.
	Level 2: Probable structure a) by library spectrum match b) by diagnostic evidence	MS, MS^2, Library MS^2 MS, MS^2, Exp. data
	***Level 3: Tentative 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

Tentative and Unknown Spectra

Need new ways to display the “uncertainty” correctly

Molecular formula only

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

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[1]
  - 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”?[2]***

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

[2]Towards Open Access Mass Spectral Libraries Workshop, IMSC, Geneva, Aug. 2014. Hosts: Steve Stein & Enrico Davoli

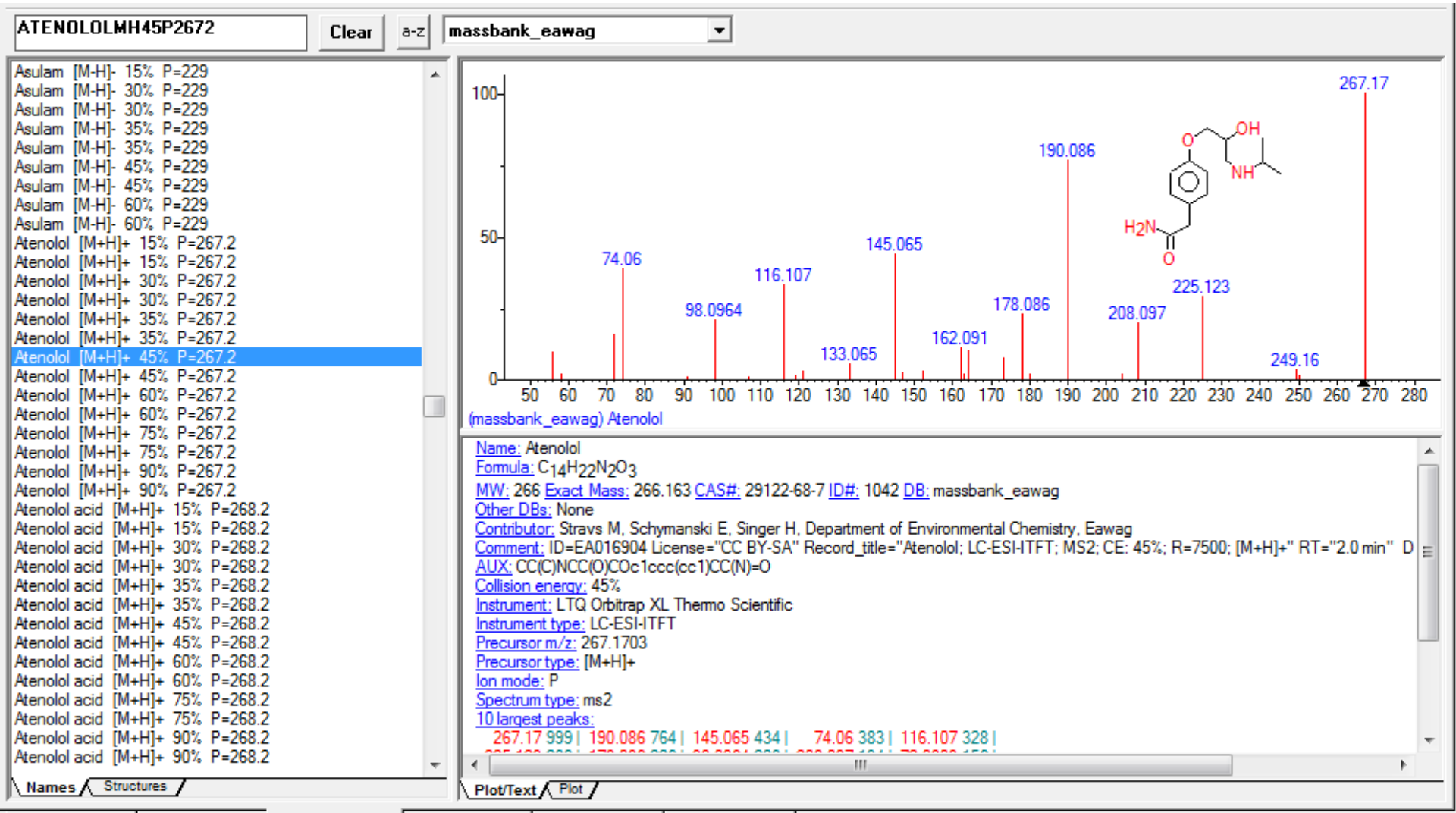
Summary: Towards Open Access Mass Spectral Libraries

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!

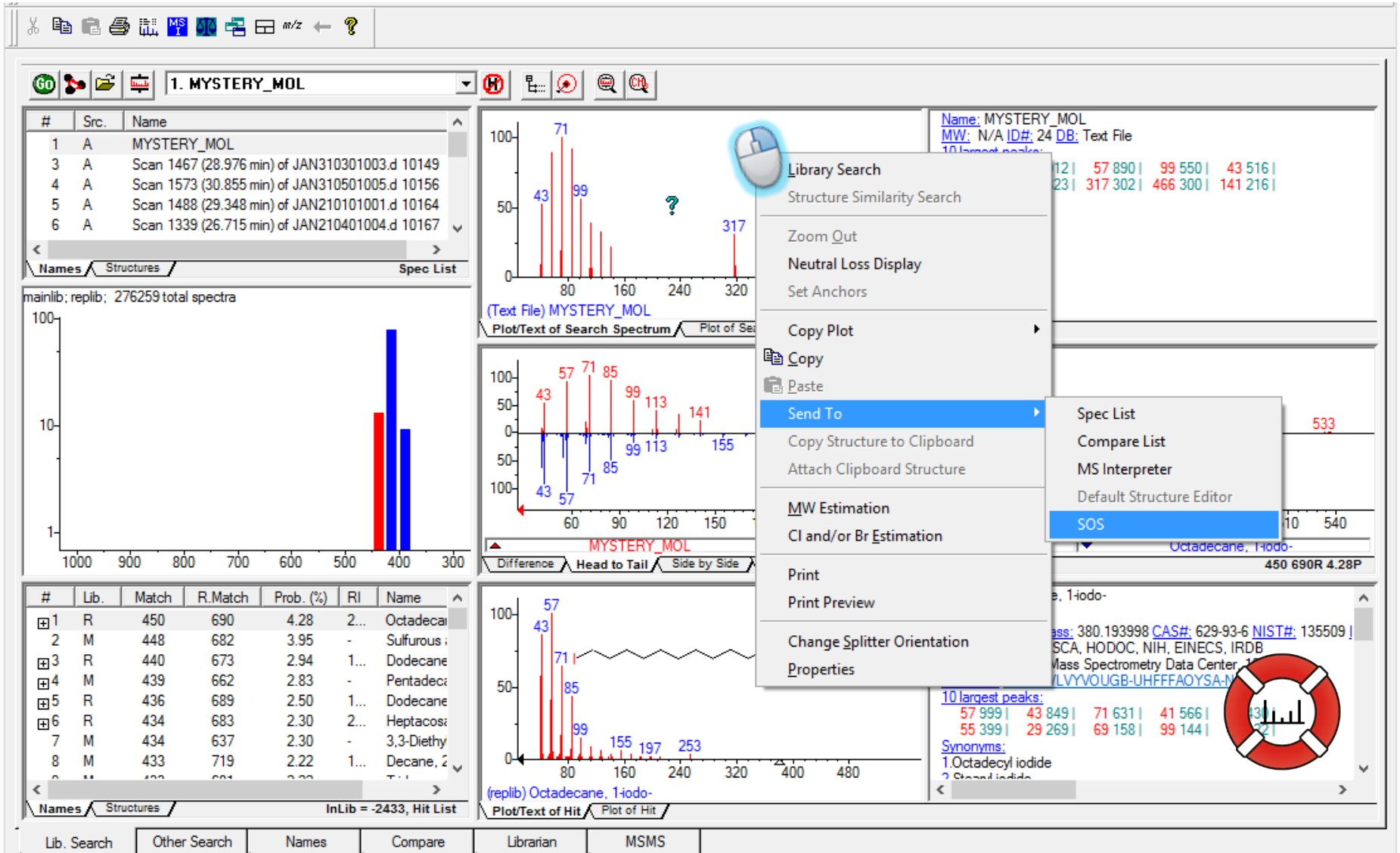
Summary: Towards Open Access Mass Spectral Libraries

MassBank ⇔ NIST



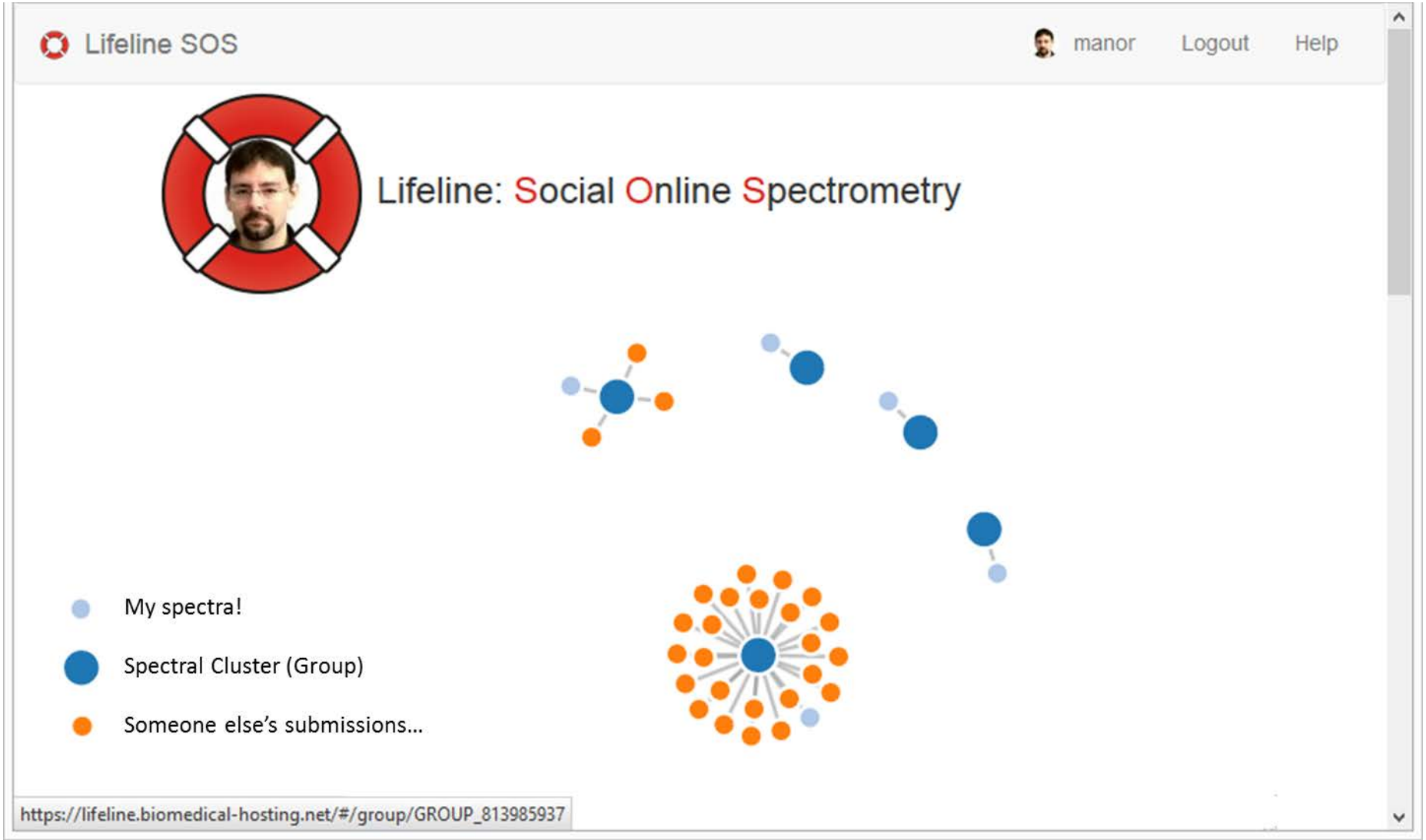
Summary: Towards Open Access Mass Spectral Libraries

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



Summary: Towards Open Access Mass Spectral Libraries

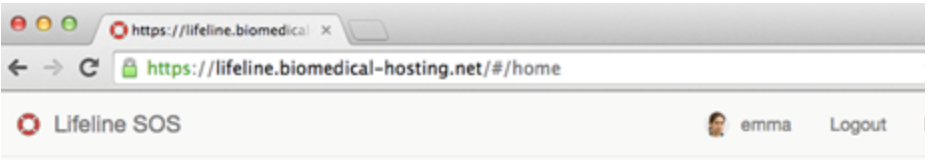
NIST14 and “SOS” (*images c/o Manor Askenazi*)



Summary: Towards Open A

NIST14 and “SOS”

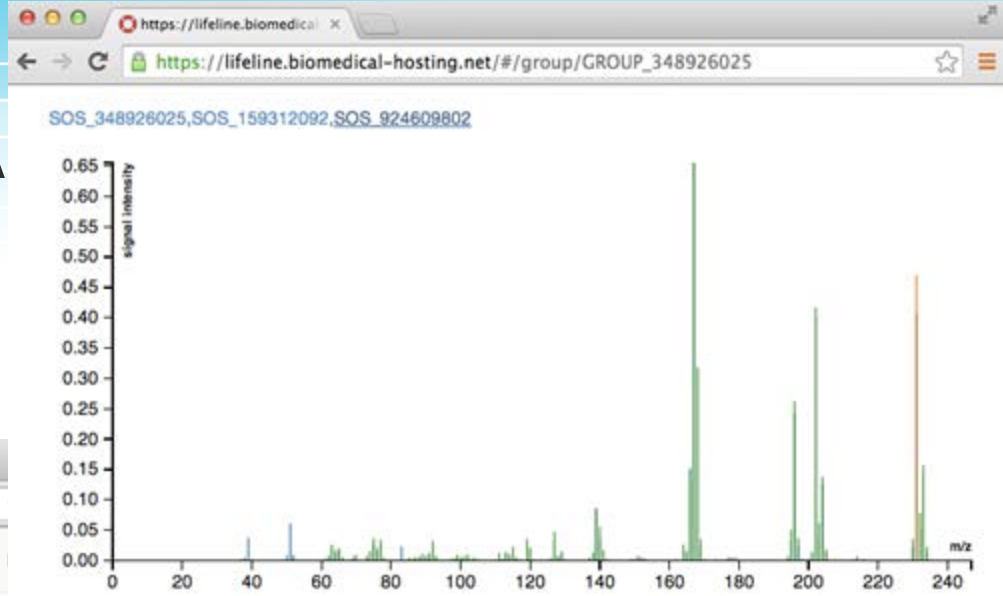
(*images c/o Manor Askenazi*)



My spectra!

Spectral Cluster (Group)

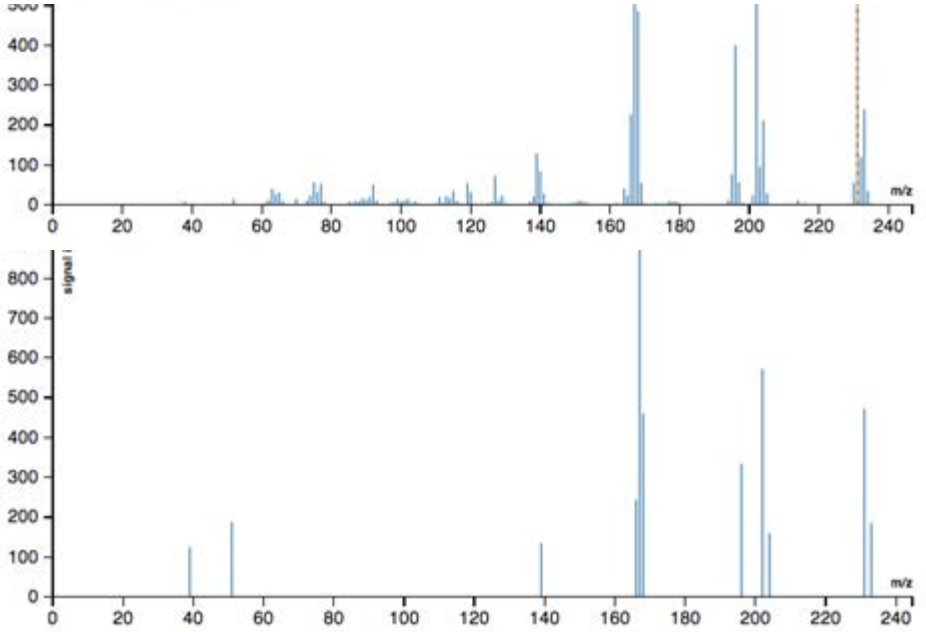
Someone else’s submissions...



Comments/Thoughts/Hypotheses:

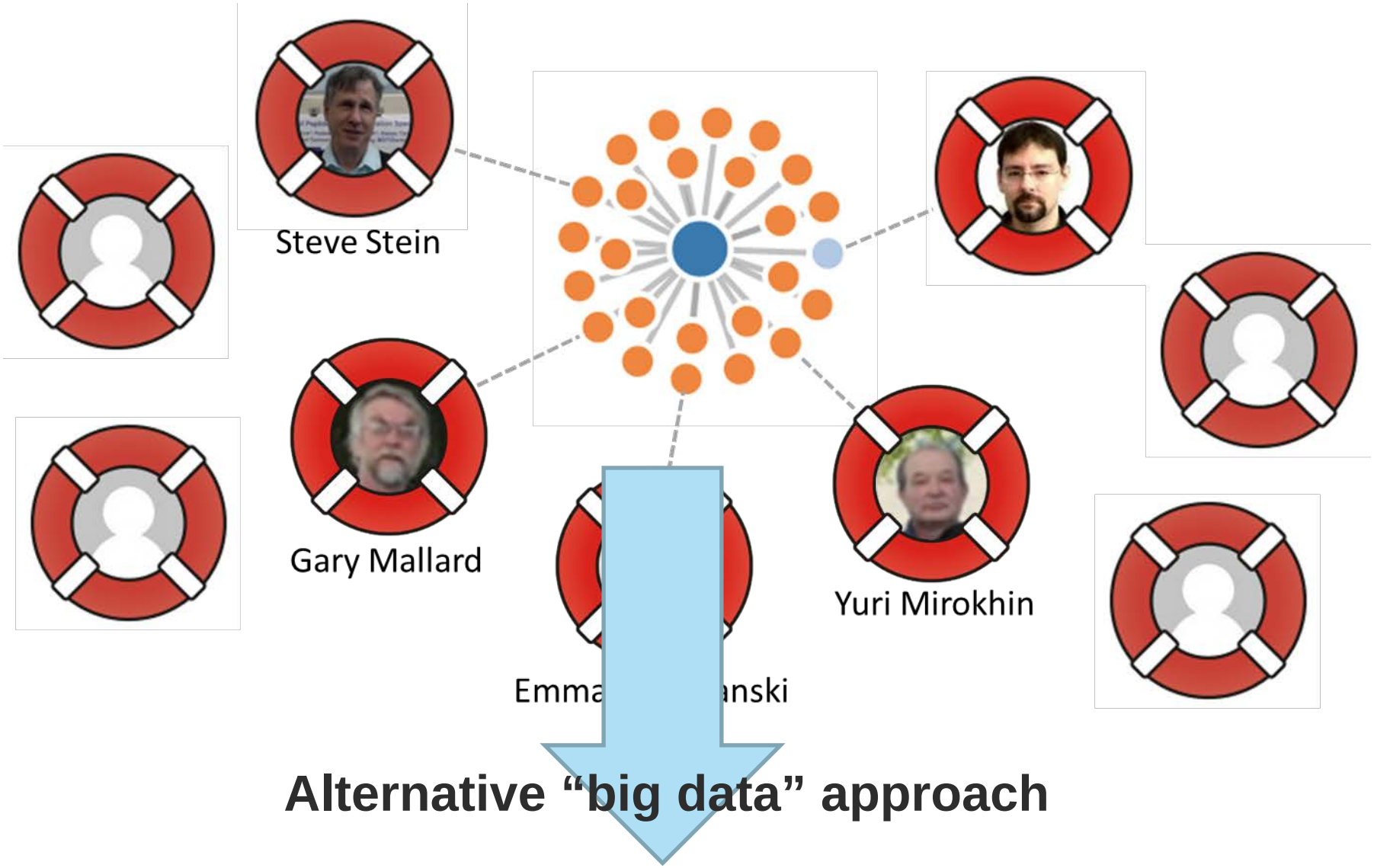
We’re synthesizing the standards now...

posted by emma (delete)



Summary: Towards Open Access Mass Spectral Libraries

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



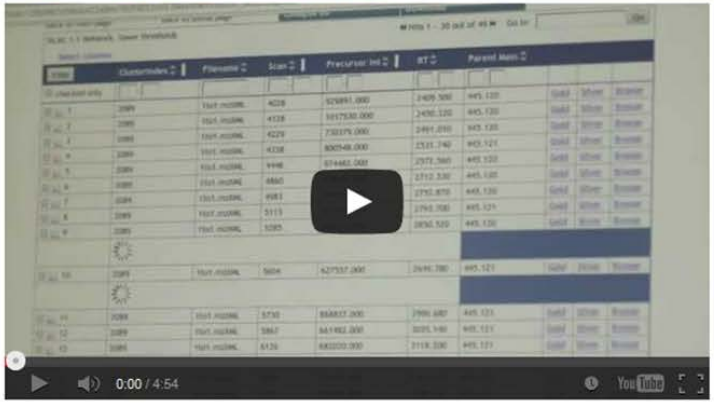
Alternative "big data" approach

Summary: Towards Open Access Mass Spectral Libraries

GNPS & Crowdsourcing Annotation of MS

gnps.ucsd.edu

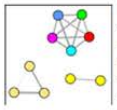
First MassIVE Knowledge Base, open March 2014



ProteoSAFe

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.

Contribute to Libraries

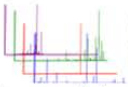
Be a part of the collaborative community effort to create the definitive collection of natural products MS/MS spectra. GNPS gives the power to add spectra, update annotations, and facilitate dialog around these spectra, to provide a truly collaborative and open natural product MS/MS database. For documentation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

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.

Crowdsourced curated libraries

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

MassIVE Public GNPS Datasets

Browse publically 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 .

Molecule Explorer

Bridge the connection between molecules and datasets. Explore exactly where certain molecules are found in all the publically 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!

MassIVE

Mass Spectrometry Interactive Virtual Environment

Share data

Explore unknown molecules

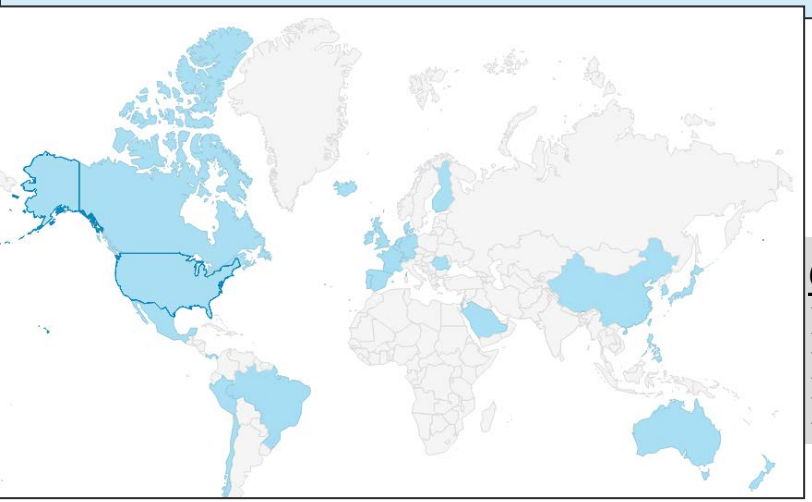
Slides c/o Nuno Bandeira

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

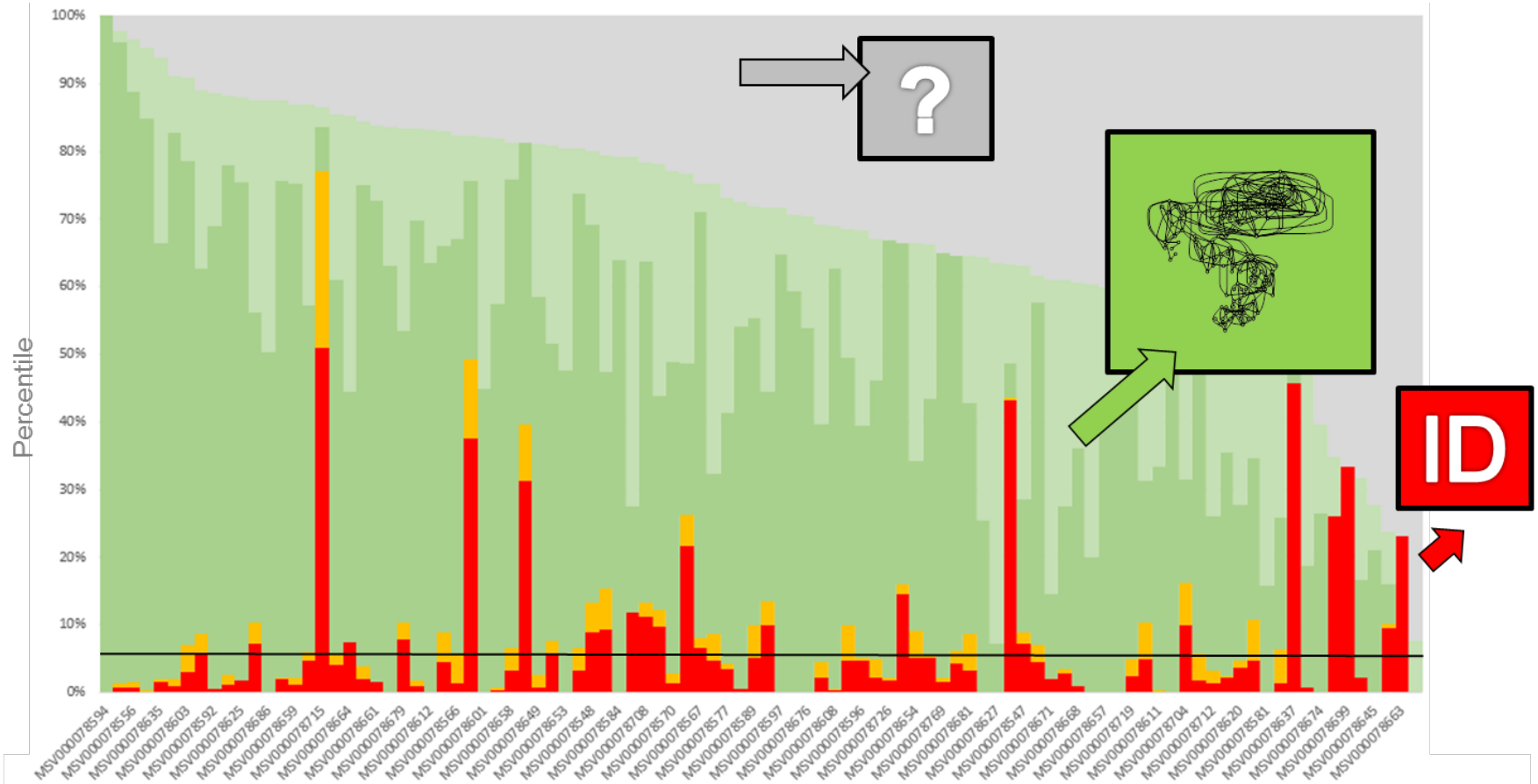
- *Hardly any spectra publicly available*
- *No available spectral libraries*

Quality levels:

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

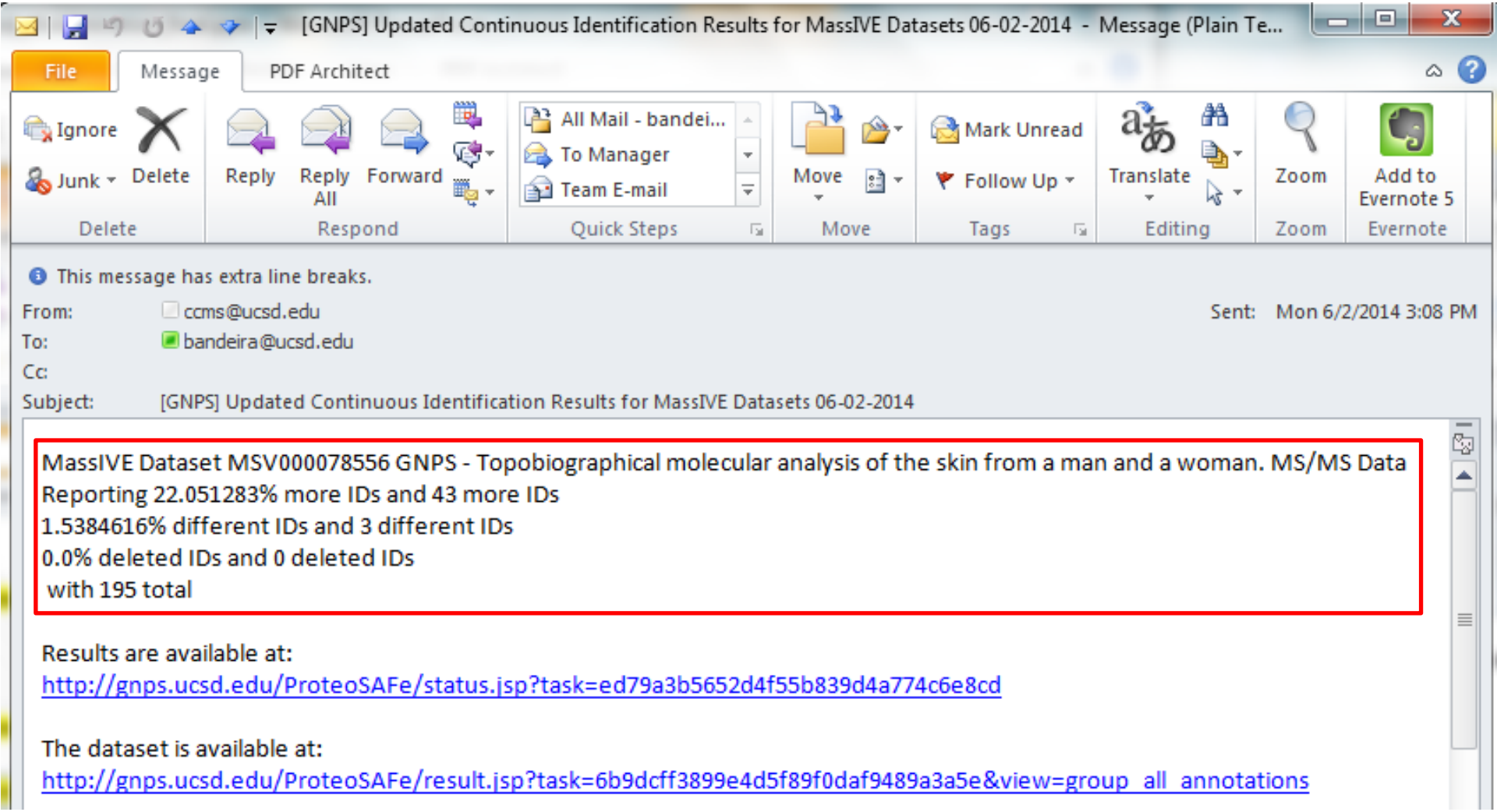
Summary: Towards Open Access Mass Spectral Libraries

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

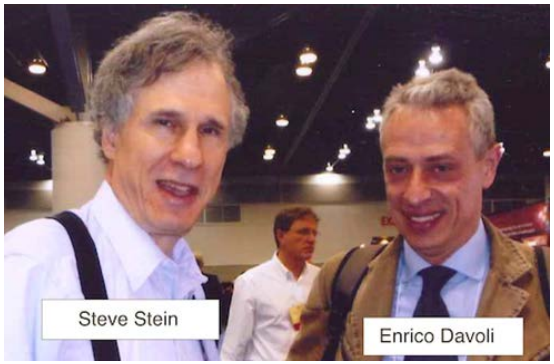


Summary: Towards Open Access Mass Spectral Libraries

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

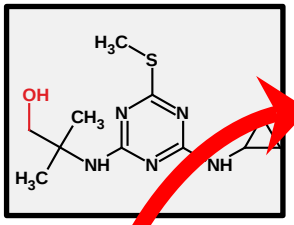
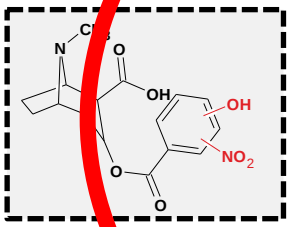


Acknowledgements



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

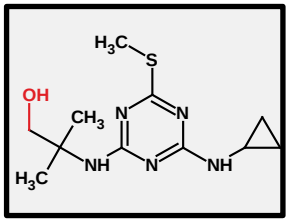
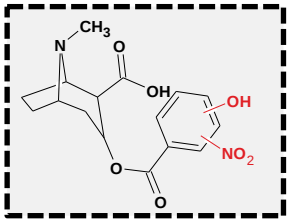
Proposed levels for MS and MS/MS data

Example	*Identification confidence*	*Minimum data requirements*
	***Level 1: Confirmed structure*** by reference standard	MS, MS^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

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