

Cheminformatics and Mass Spectrometry meets Clinical Metabolomics



Assoc. Prof. Dr. Emma L. Schymanski
(plus many, many colleagues and collaborators!)

Environmental Cheminformatics Group,
Luxembourg Centre for Systems Biomedicine, University of Luxembourg

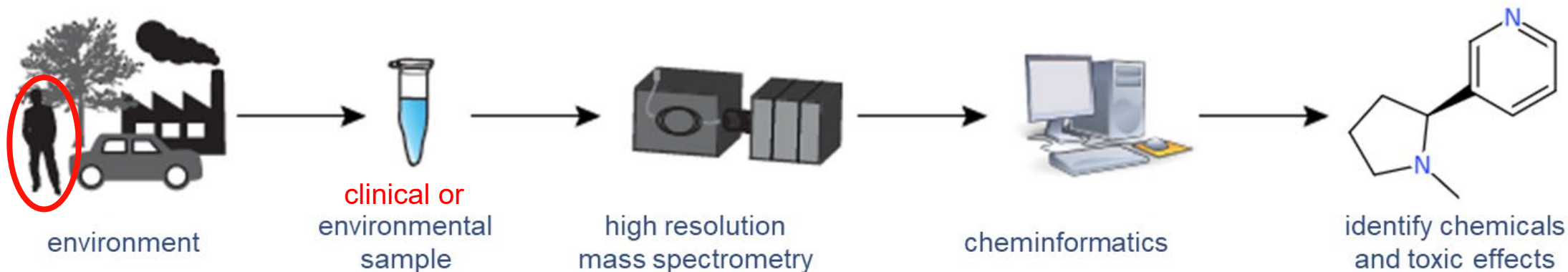
Email: emma.schymanski@uni.lu and [@ESchymanski](https://twitter.com/ESchymanski)

Web: https://www.uni.lu/lcsb/research/environmental_cheminformatics/

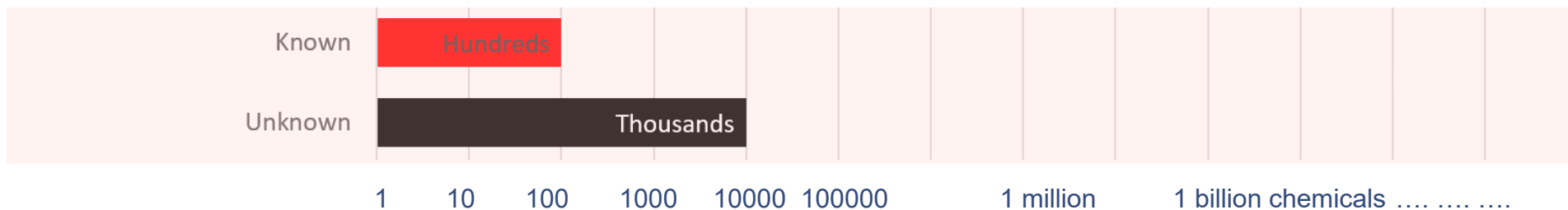
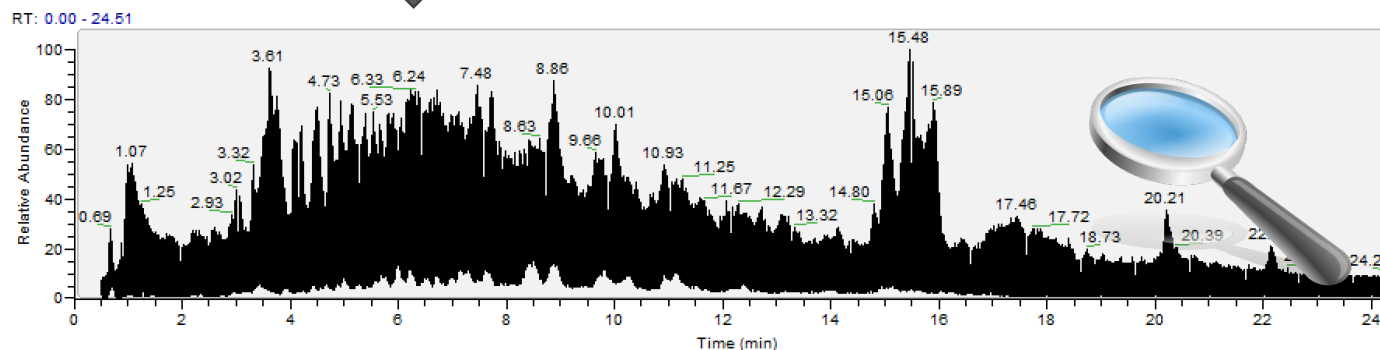
Clinical Metabolomics Copenhagen 2021
[Digital Event](#), 26-27 August 2021

Slides available at
DOI: [10.5281/zenodo.5354574](https://doi.org/10.5281/zenodo.5354574)

Background: Environmental Cheminformatics & HR-MS



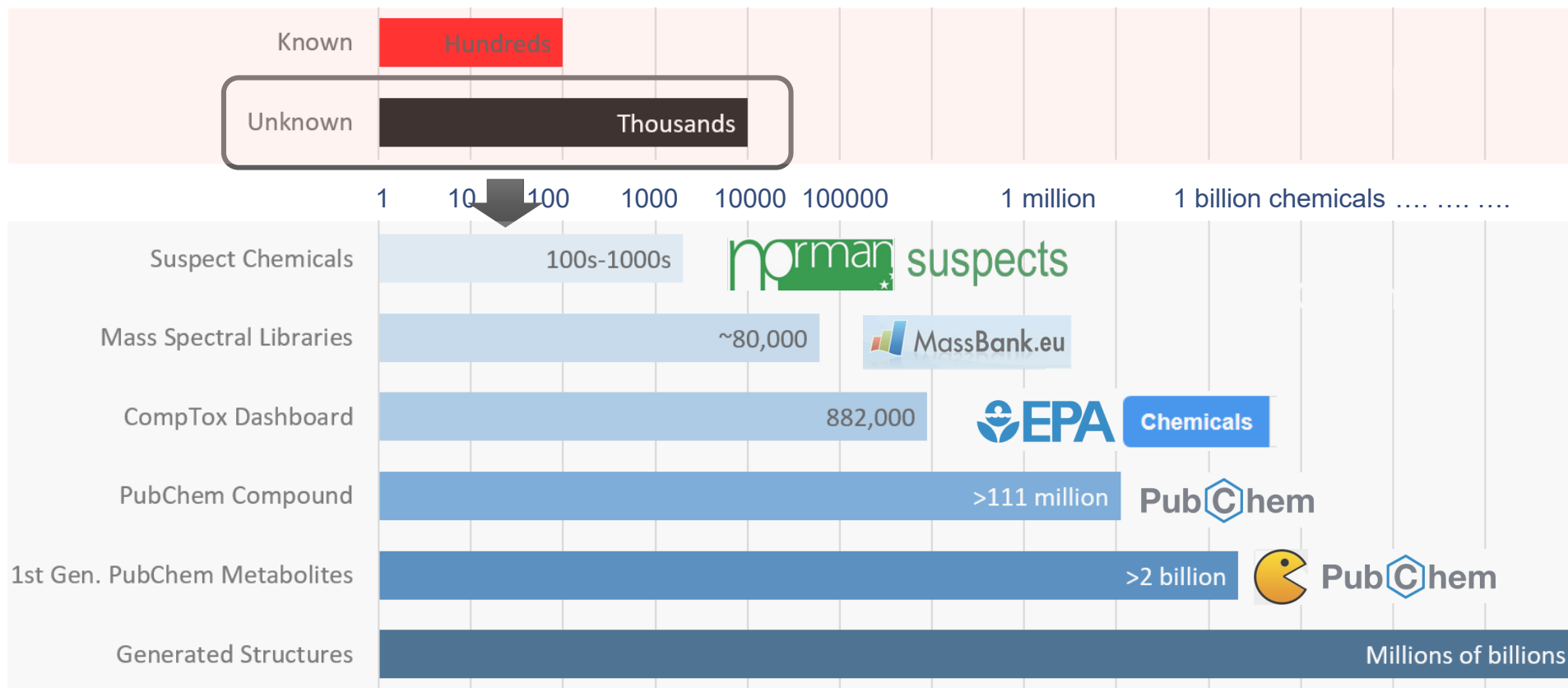
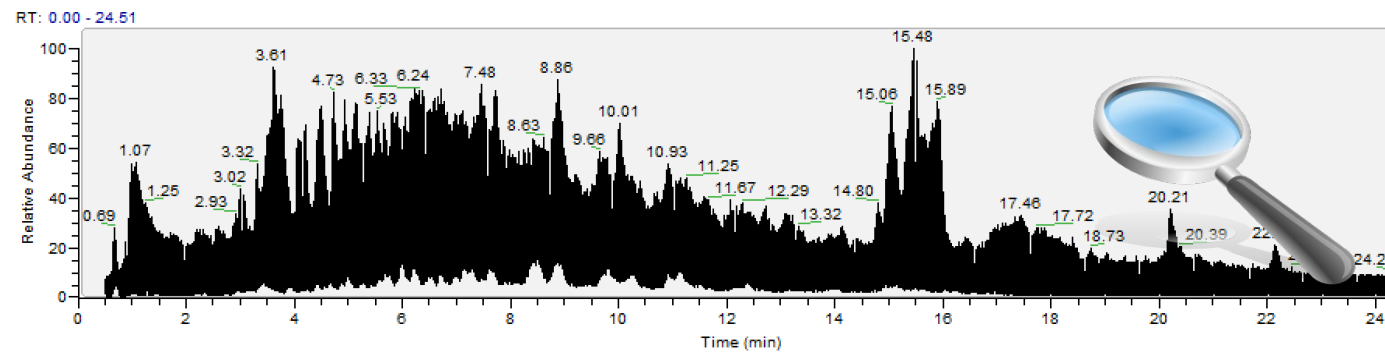
High resolution mass spectrometry



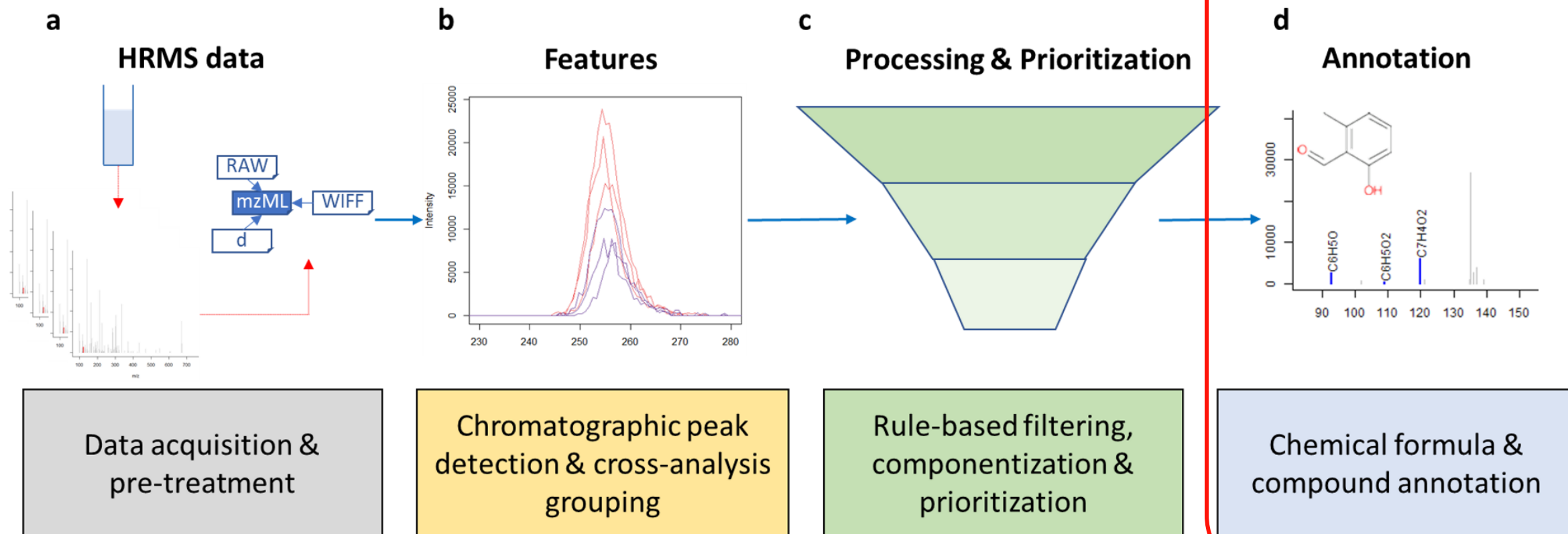
Background: Environmental Cheminformatics & HR-MS

High resolution
mass spectrometry

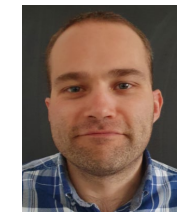
AND connecting
chemical knowledge



Background: Identification with HR-MS



Open Source Workflows for NT-HRMS: patRoon



Rick Helmus

Software | [Open Access](#) | [Published: 06 January 2021](#)

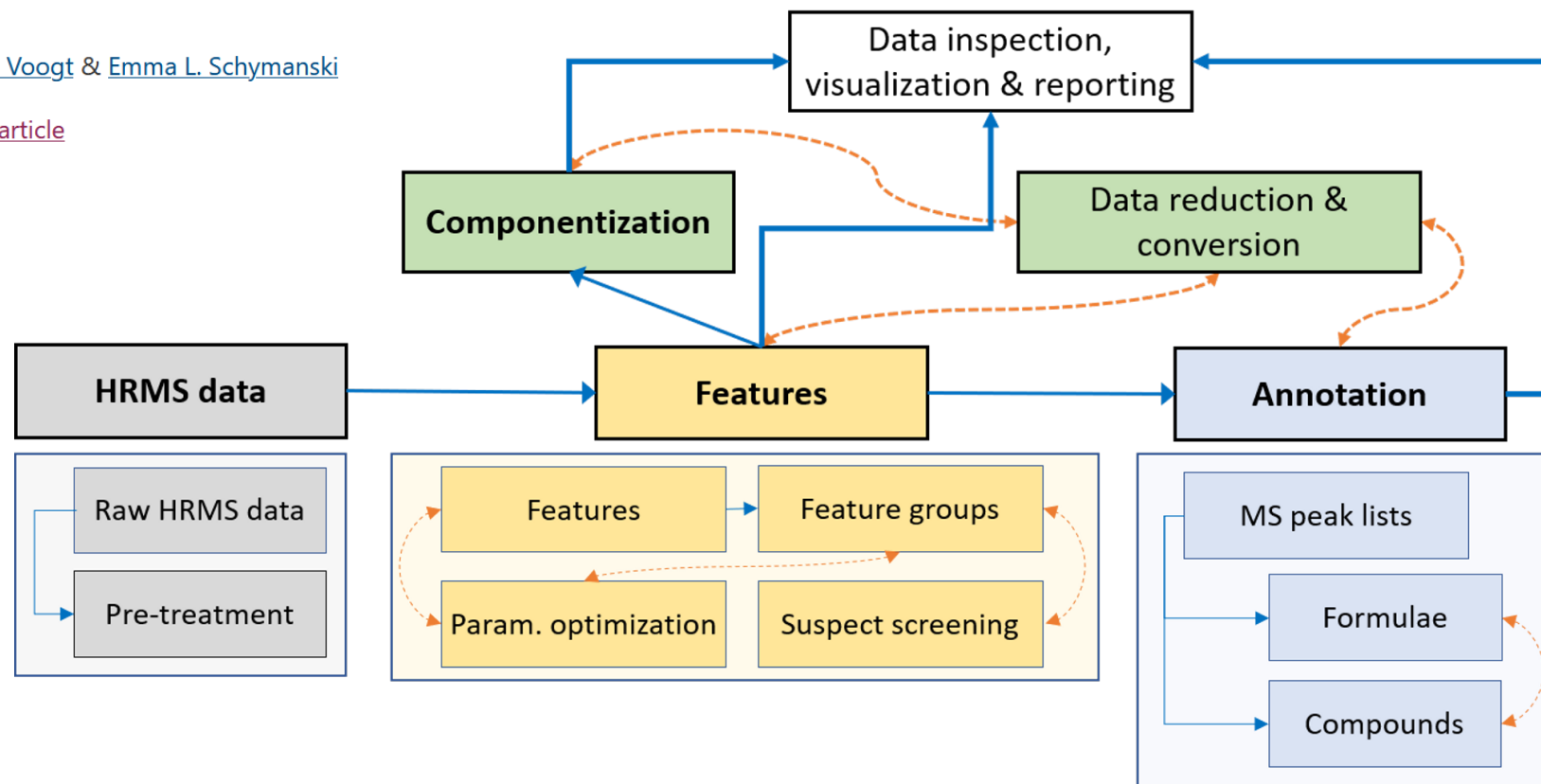
patRoon: open source software platform for environmental mass spectrometry based non-target screening

[Rick Helmus](#) , [Thomas L. ter Laak](#), [Annemarie P. van Wezel](#), [Pim de Voogt](#) & [Emma L. Schymanski](#)

Journal of Cheminformatics **13**, Article number: 1 (2021) | [Cite this article](#)

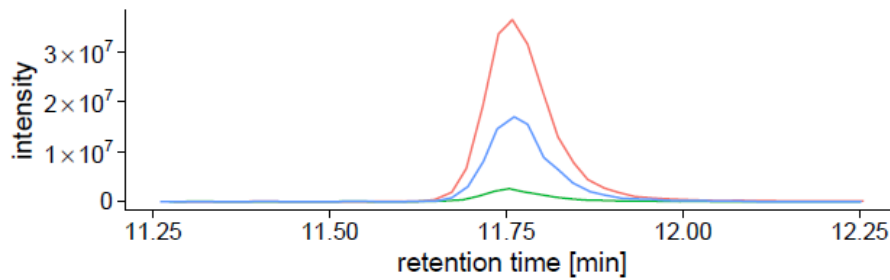
1394 Accesses | 2 Citations | 20 Altmetric | [Metrics](#)

Journal of Cheminformatics



Open Source Workflows for NT-HRMS: Shinyscreen

100 EIC (m/z = 182.0816)



peak retention time (MS1)

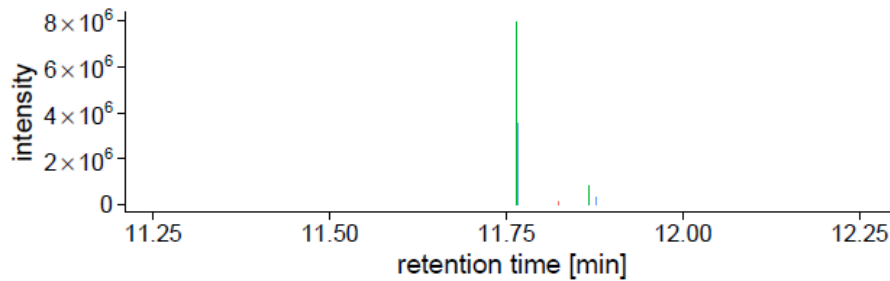
- Std ; rt= 11.76 min
- KO ; rt= 11.75 min
- WT ; rt= 11.76 min

<https://git-r3lab.uni.lu/eci/shinyscreen>



Anjana Elapavalore, Mira Narayanan,
Todor Kondic, Jessy Krier,
Hiba Mohammed Taha.

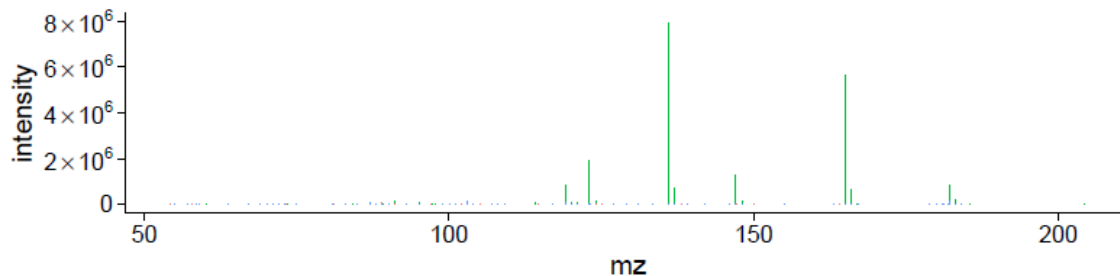
MS2



peak retention time (MS2)

- KO ; rt= 11.82 min
- Std ; rt= 11.76 min
- WT ; rt= 11.77 min

MS2



tag

- KO
- Std
- WT



Mass Spectral Libraries: MassBank (Open Source & Data!)

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

<https://github.com/MassBank/MassBank-data/>

MassBank

Search

Contents

Download

Accession

Go

More ▼

MassBank Europe

MassBank

High Quality Mass Spectral Database

>> Search Spectra

>> Learn More

Search for:

→ Mass calculator

Basic Search

Peak List

Peaks

Peak Differences

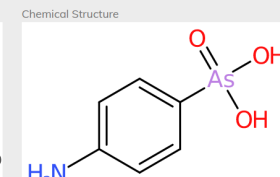
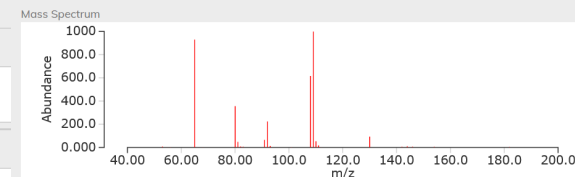
InChIKey

SPLASH

MassBank Record: LU040605

Compound Information

(4-Aminophenyl)arsonic acid; LC-ESI-QFT; MS2; CE: 75; R=17500; [M+H]⁺



Compound name

Exact Mass

Tolerance

AND

Formula (e.g. C6H7N5, C5H*N5, C5*)

AND

Search



Expert Knowledge: NORMAN Database System

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



NORMAN Database System

The screenshot shows the NORMAN Substance Database interface. At the top, there is a navigation bar with the NORMAN logo and links for 'NORMAN WEBSITE', 'NORMAN DATABASE SYSTEM', 'HOME', and 'LOGIN'. Below the navigation bar, the page title is 'NORMAN Suspect List Exchange - NORMAN SLE'. A search bar is visible on the left side. The main content area contains a paragraph: 'The NORMAN Suspect List Exchange (NORMAN-SLE) was established in 2015 as a central access point for NORMAN members (and others) to find suspect lists relevant for their environmental monitoring question. This Exchange documents all individual collections that form a part of NORMAN SusDat, the merged NORMAN Substance Database (DOI: 10.5281/zenodo.2664077).'



Substance Database

A merged list of NORMAN substances; Central Database to access various lists of substances for suspect screening and prioritisation



Suspect List Exchange

Central Database to access various lists of substances for suspect screening and prioritisation



Chemical Occurrence Data

A database of geo-referenced monitoring data on emerging substances



Antibiotic Resistance Bacteria/Genes

A database of ARBs/ARGs in environmental matrices



Ecotoxicology

A platform for systematic collection and evaluation of ecotoxicity studies for harmonised derivation of environmental quality standards



MassBank Europe

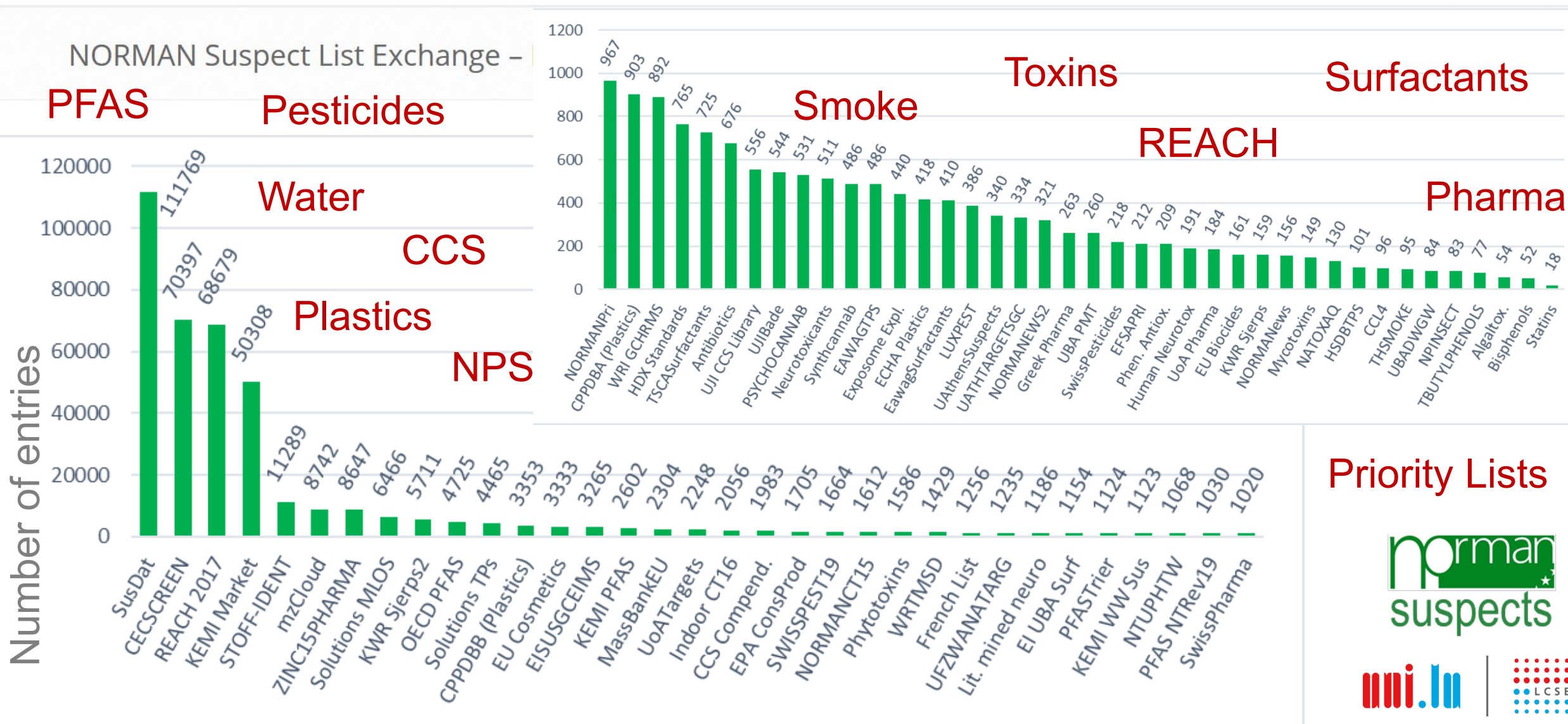
A database of mass spectra of emerging substances to support identification of unknown substances



Expert Knowledge: NORMAN Suspect List Exchange (>80 lists!)

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

<https://zenodo.org/communities/norman-sle>

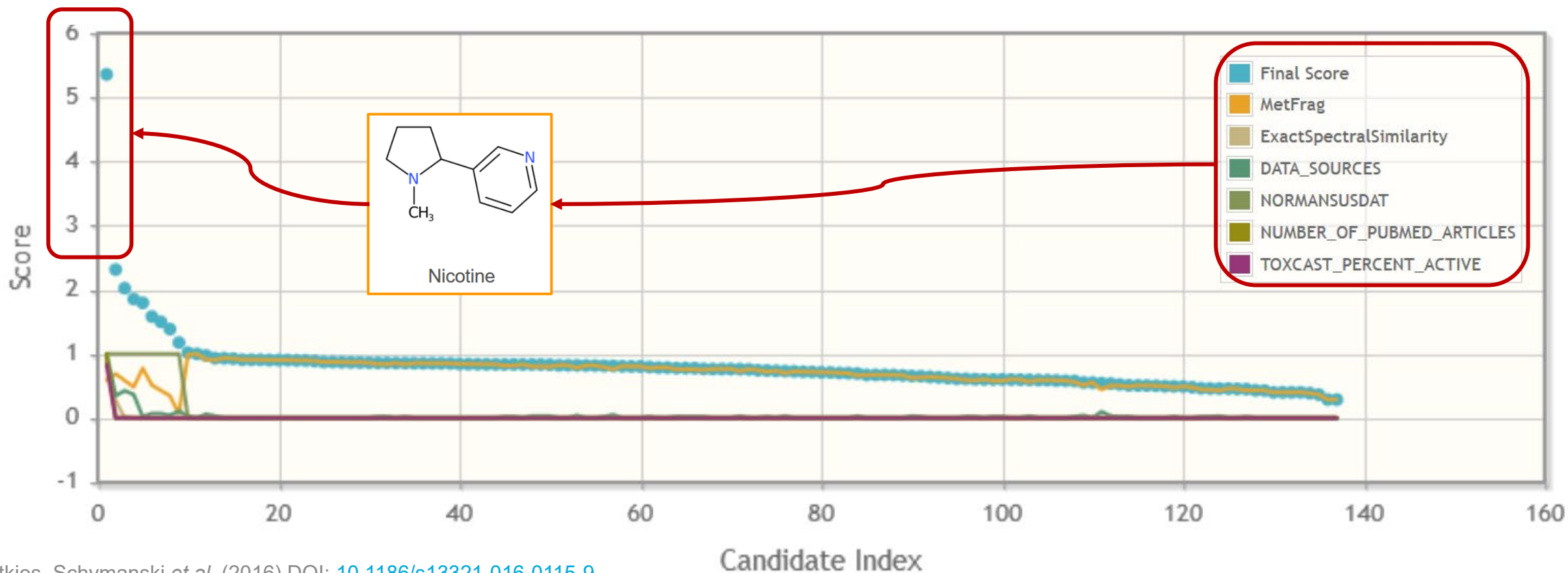


Connecting Knowledge for Chemical Identification: MetFrag



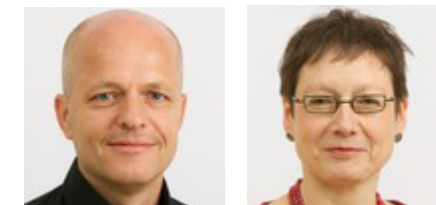
<https://msbi.ipb-halle.de/MetFrag/>

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

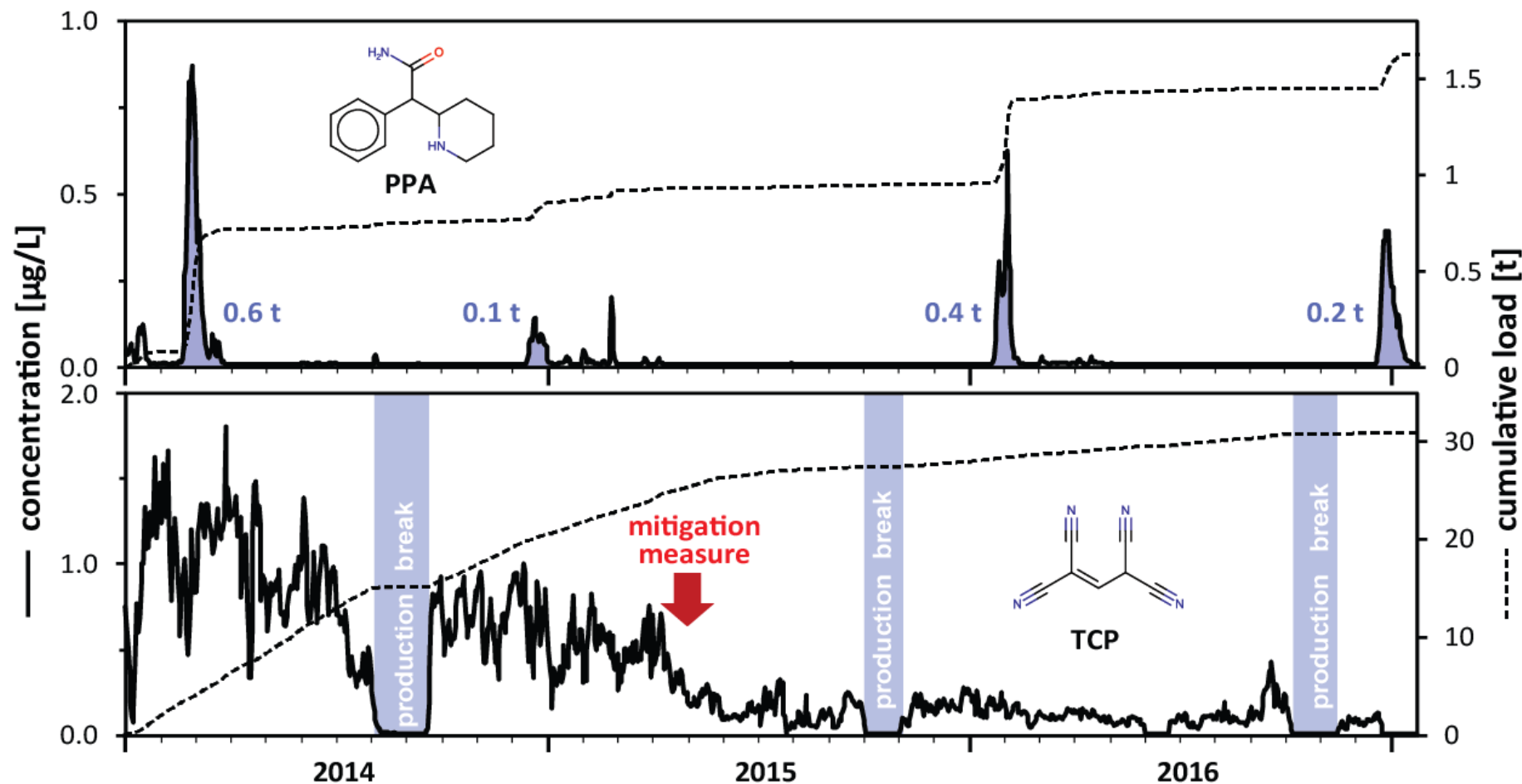


Examples: Real Time Monitoring of the Rhine River

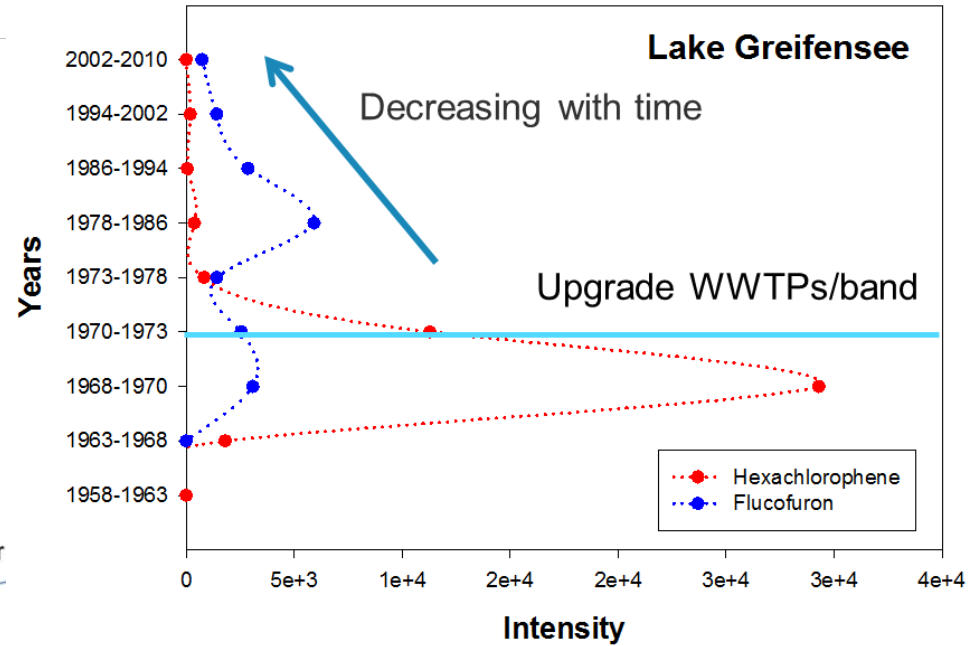
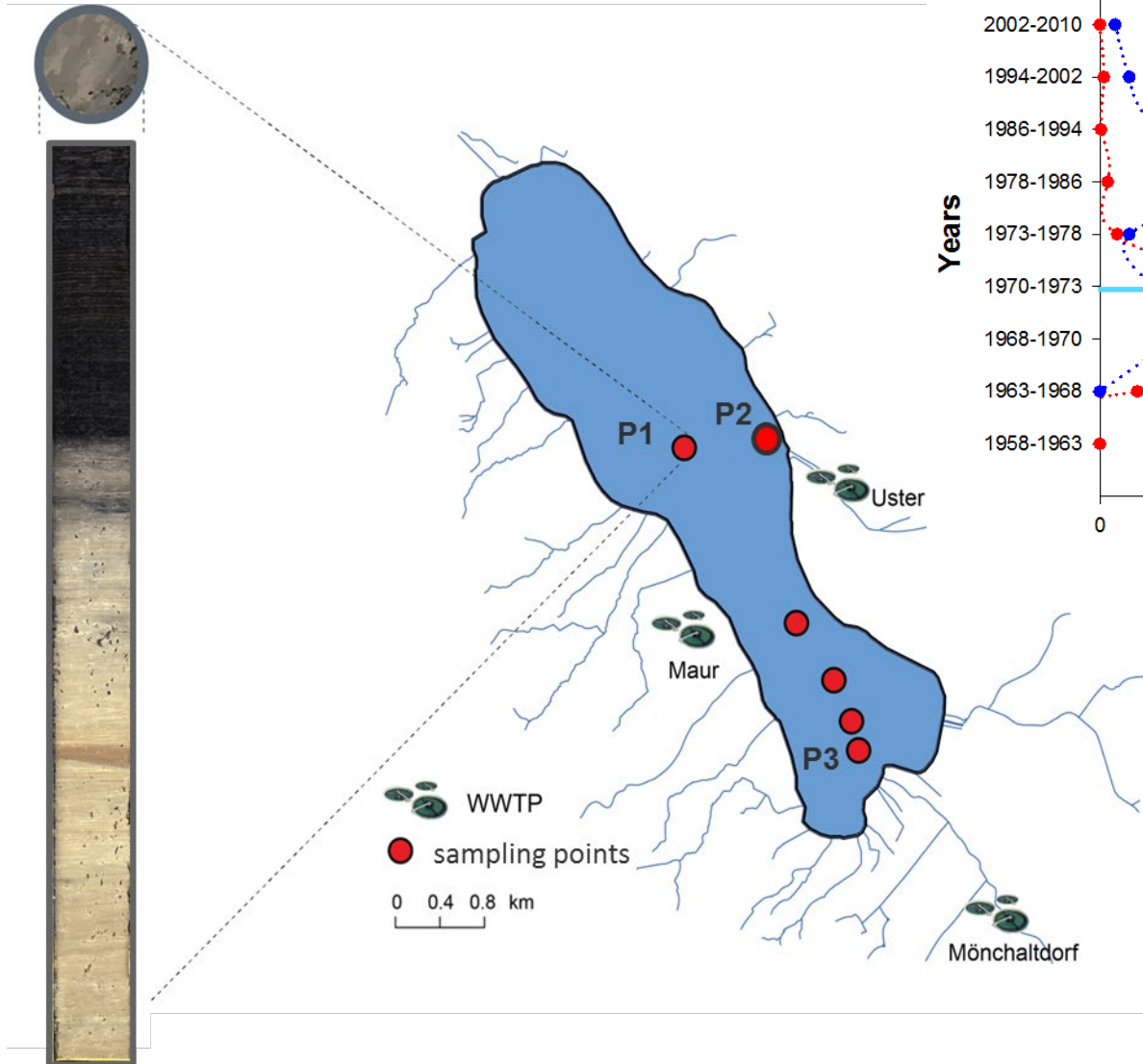
Previously unknown chemicals detected due to “stand-out” patterns



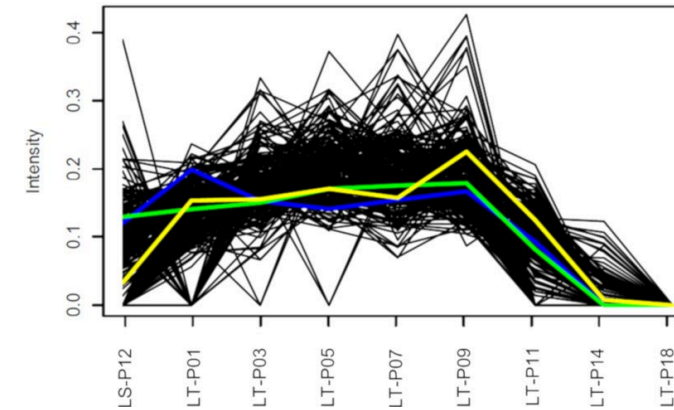
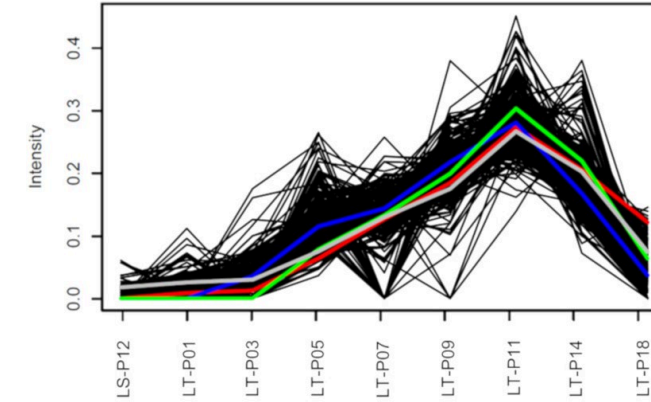
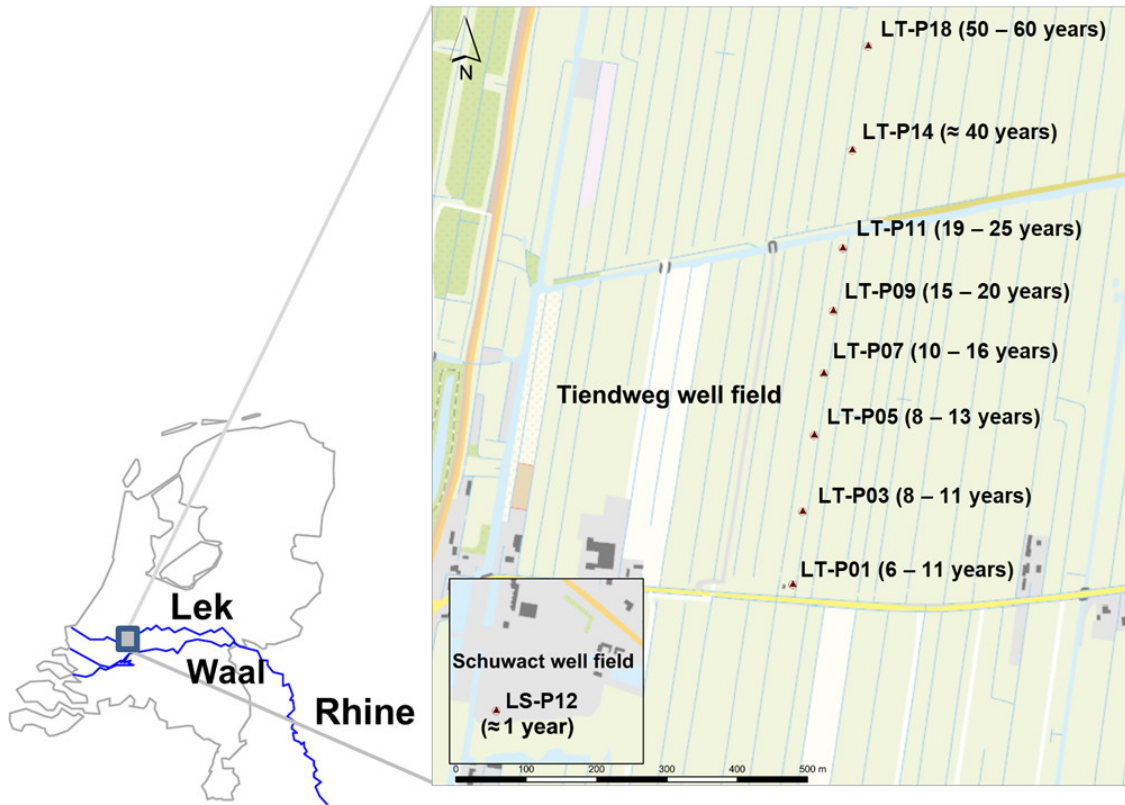
eawag
aquatic research



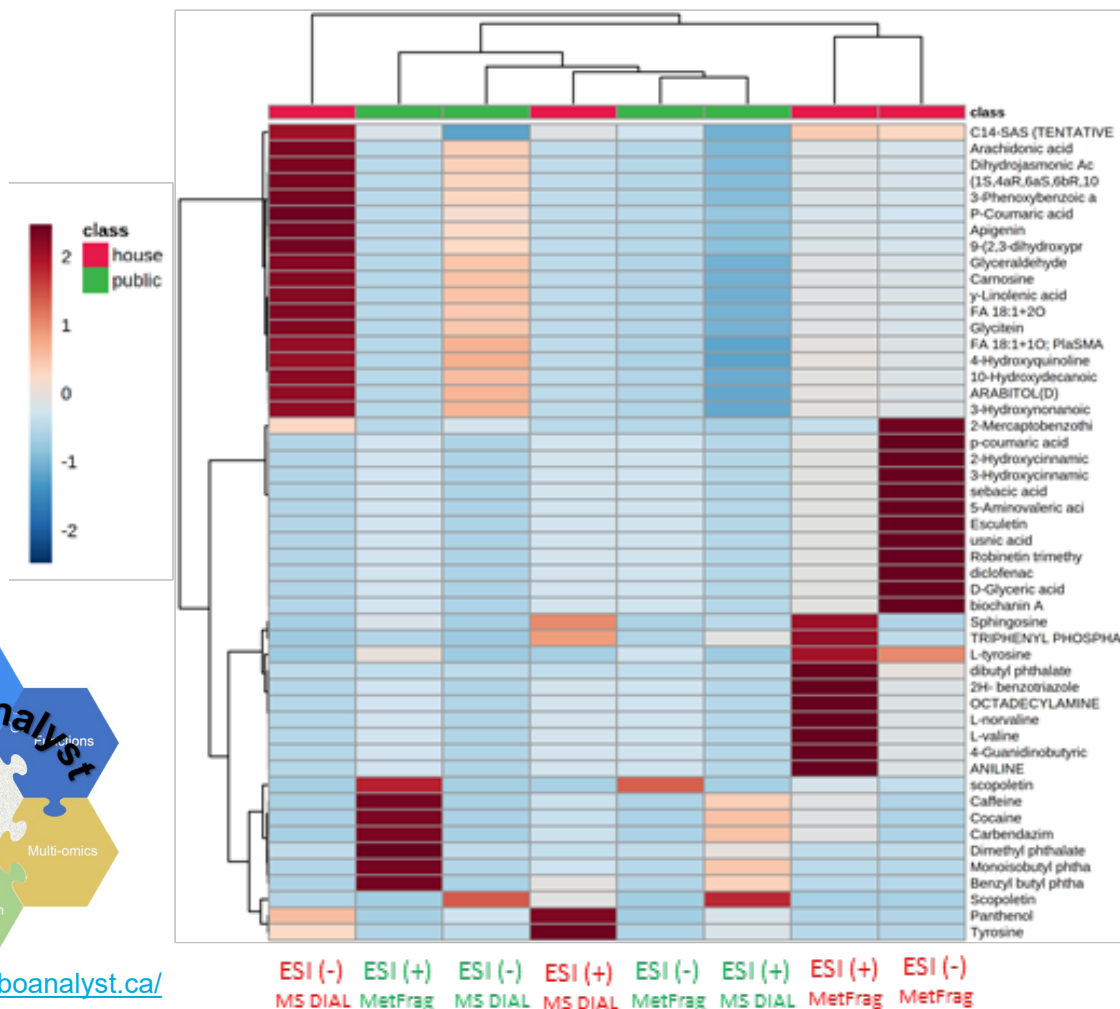
Examples: Historical Contamination in Lake Sediments



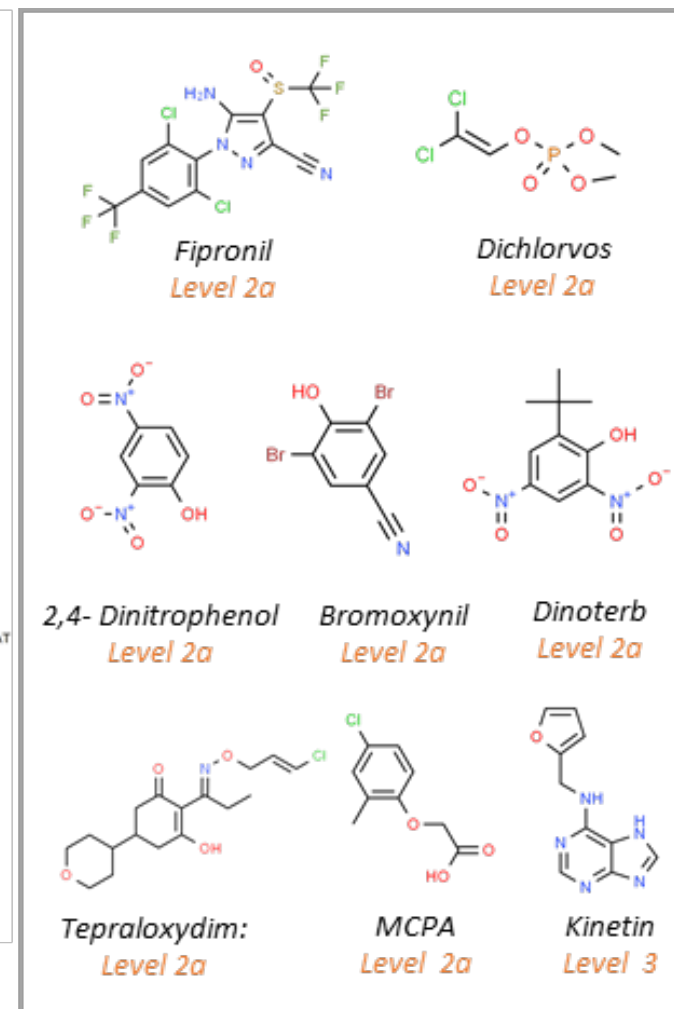
Examples: Time Trends in Riverbank Filtration Systems



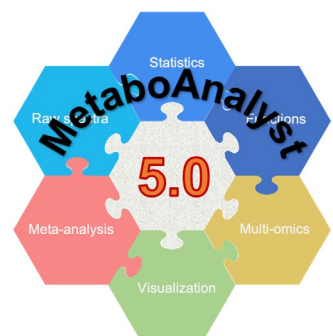
Examples: NORMAN Indoor Dust Collaborative Trial



Heatmap showing the coverage of the software used for the data analysis in ESI (+) and ESI (-). Only the top 50 chemicals, ranked by T-test, are represented in this heatmap.



Chemicals classified as environmental and health hazards and that were not reported in the first dust NORMAN trial. The identification levels are indicated in orange.



<https://www.metaboanalyst.ca/>



MS-DIAL



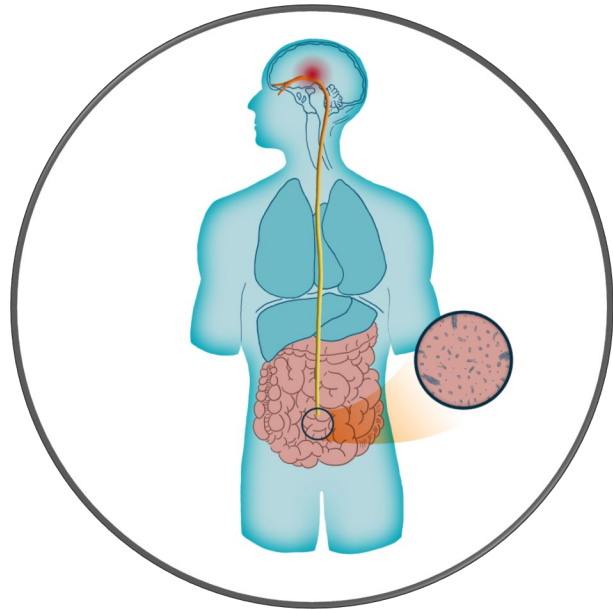
<http://prime.psc.riken.jp/comps/msdial/main.html> and
<https://msbi.ipb-halle.de/MetFrag/>



Examples (WiP): Application to Patient Microbiome Studies



MICROH

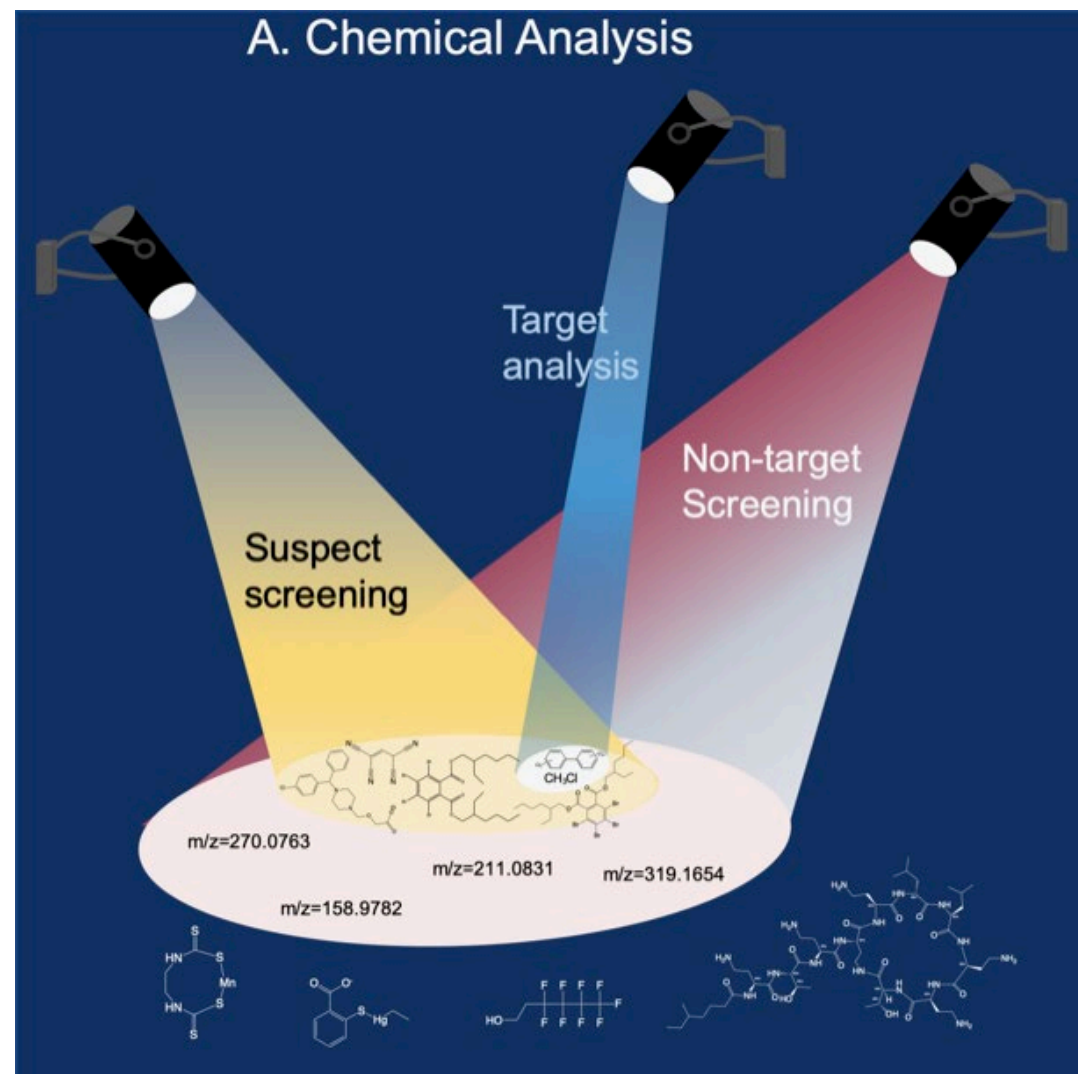
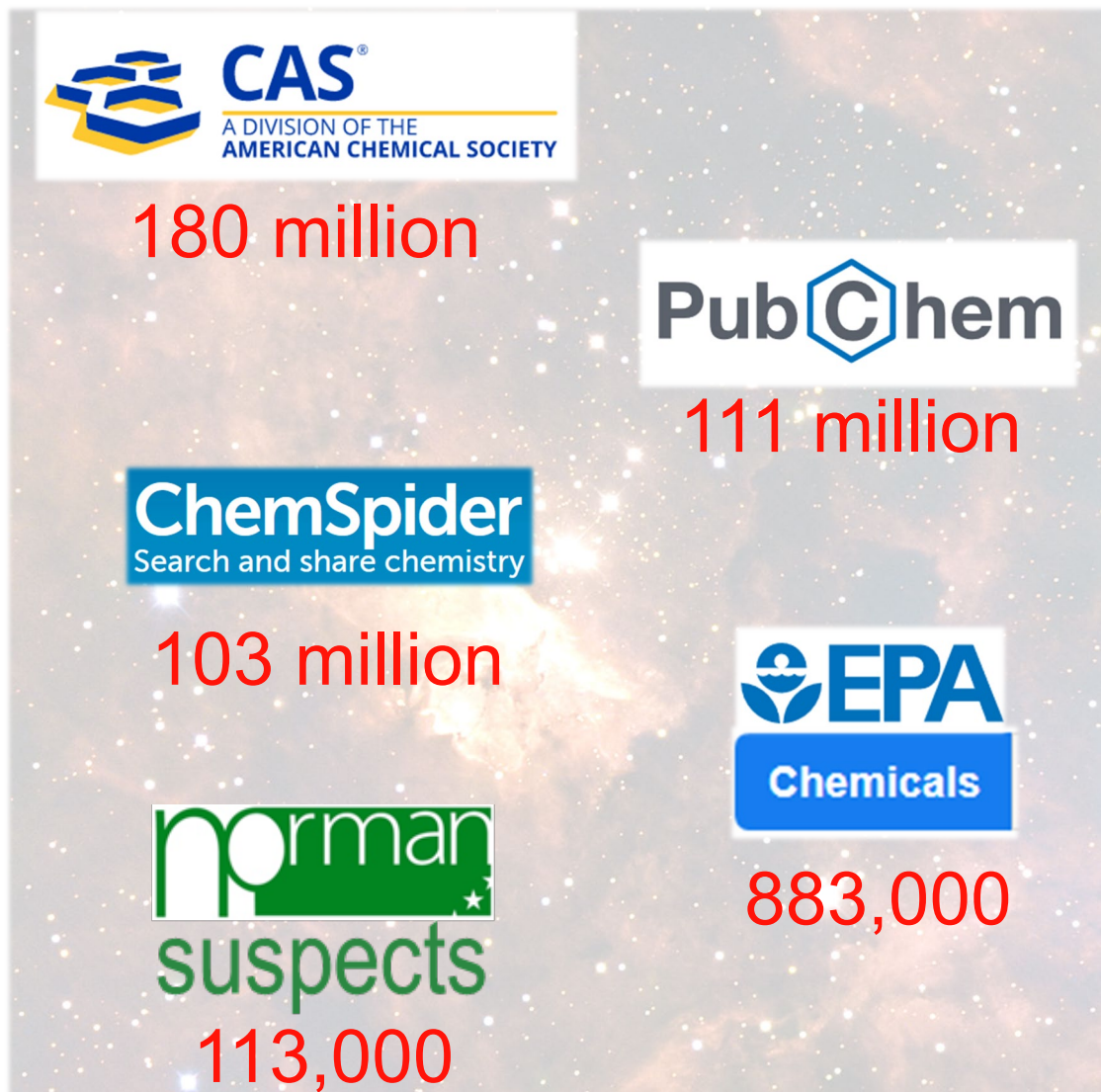


This study revealed and confirmed differential abundances of gut microbial taxa in PD patients vs Control

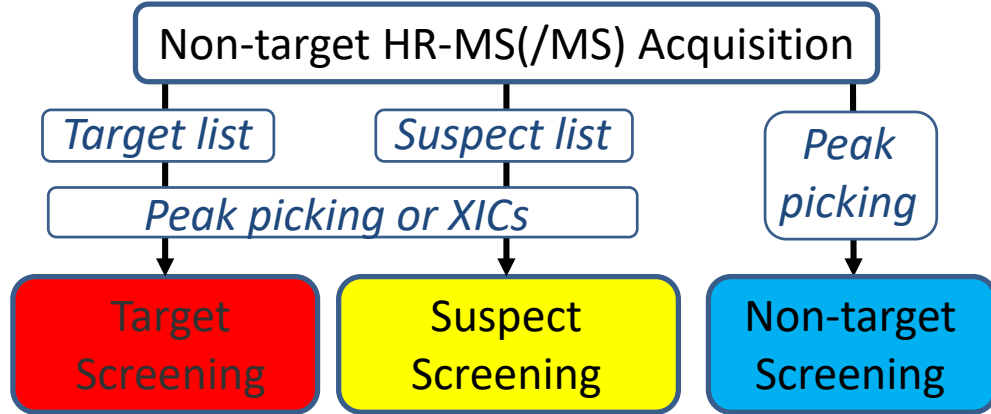
Hypothesis: Changes in the microbial community structure and function in the gastrointestinal tract accompany Parkinson's disease (PD) from its onset.



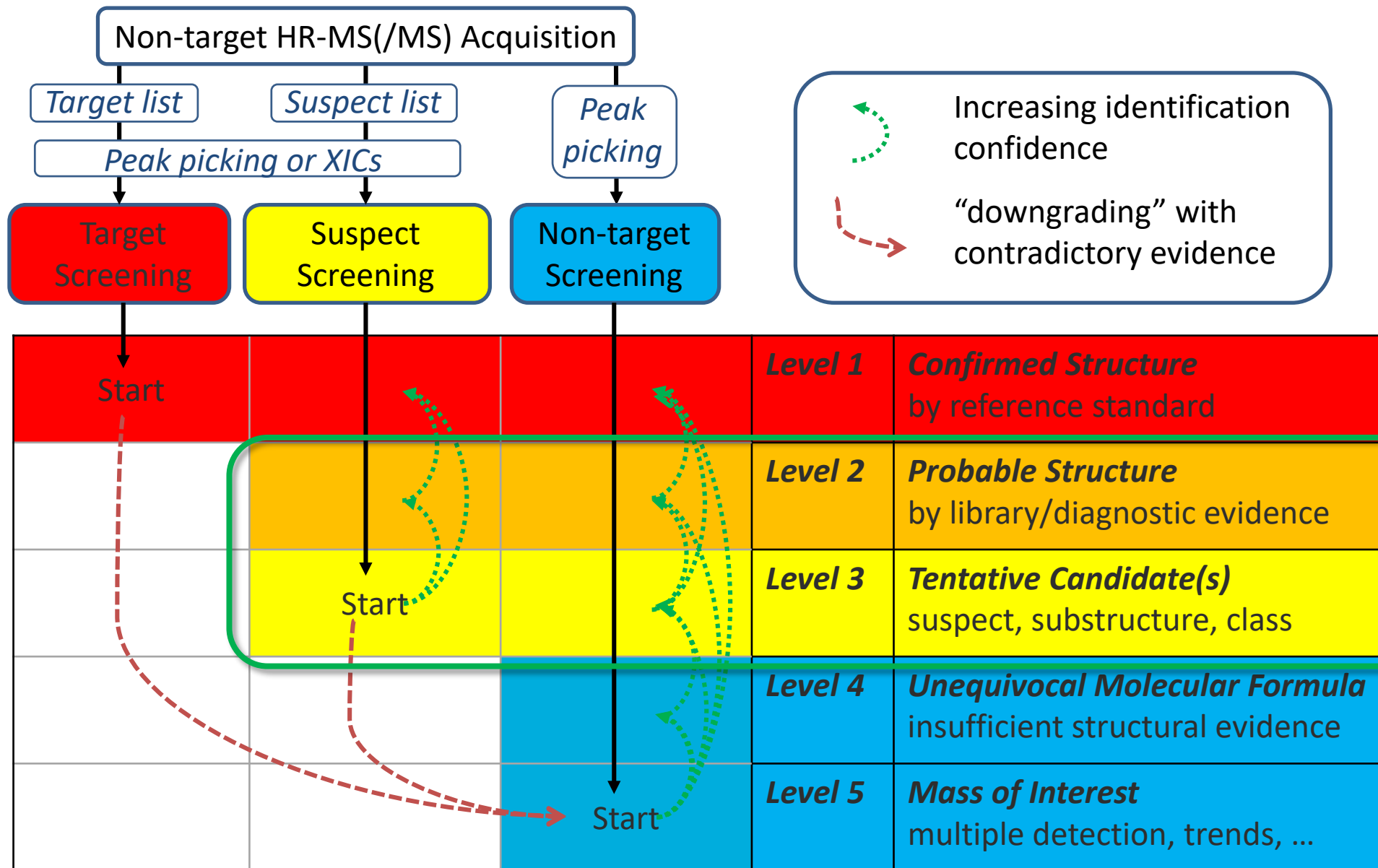
The Problem: Which chemicals are relevant? How to find them?



Identification Strategies and Confidence in NT-HRMS(/MS)



Identification Strategies and Confidence in NT-HRMS(/MS)



MassBank
High Quality Mass Spectral Database



eawag
aquatic research



Introduction to MetFrag

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

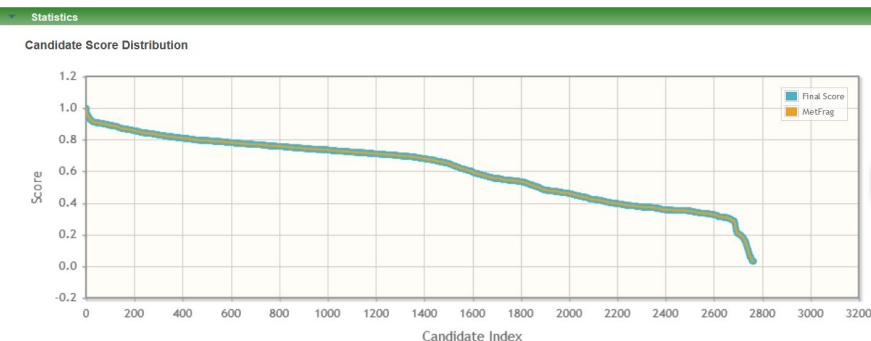
m/z $[M-H]^-$
213.9637
 ± 5 ppm

5 ppm
0.001 Da

PubChem



Ranked Candidates



MS/MS

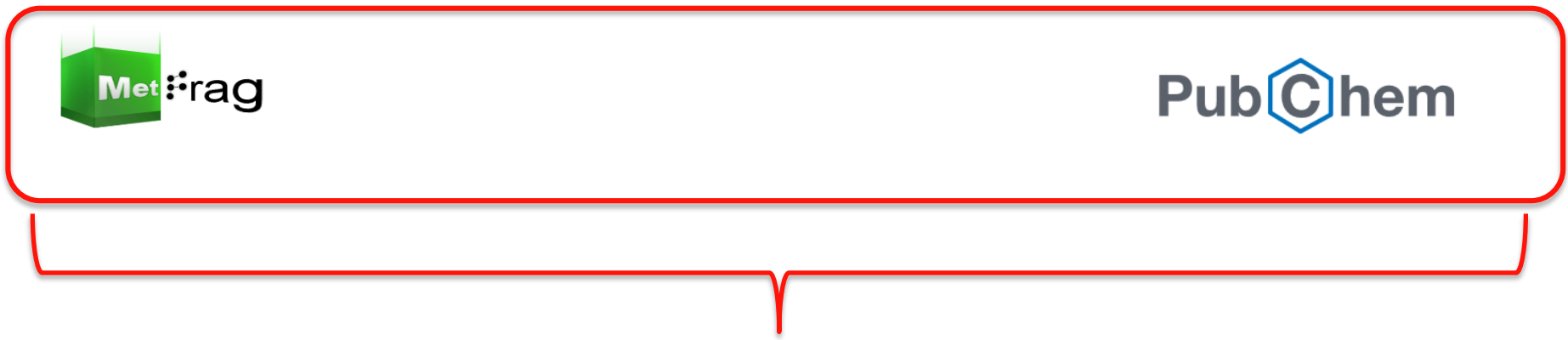
134.0054	339689
150.0001	77271
213.9607	632466



Key Challenge: MS and MS/MS alone is not enough!

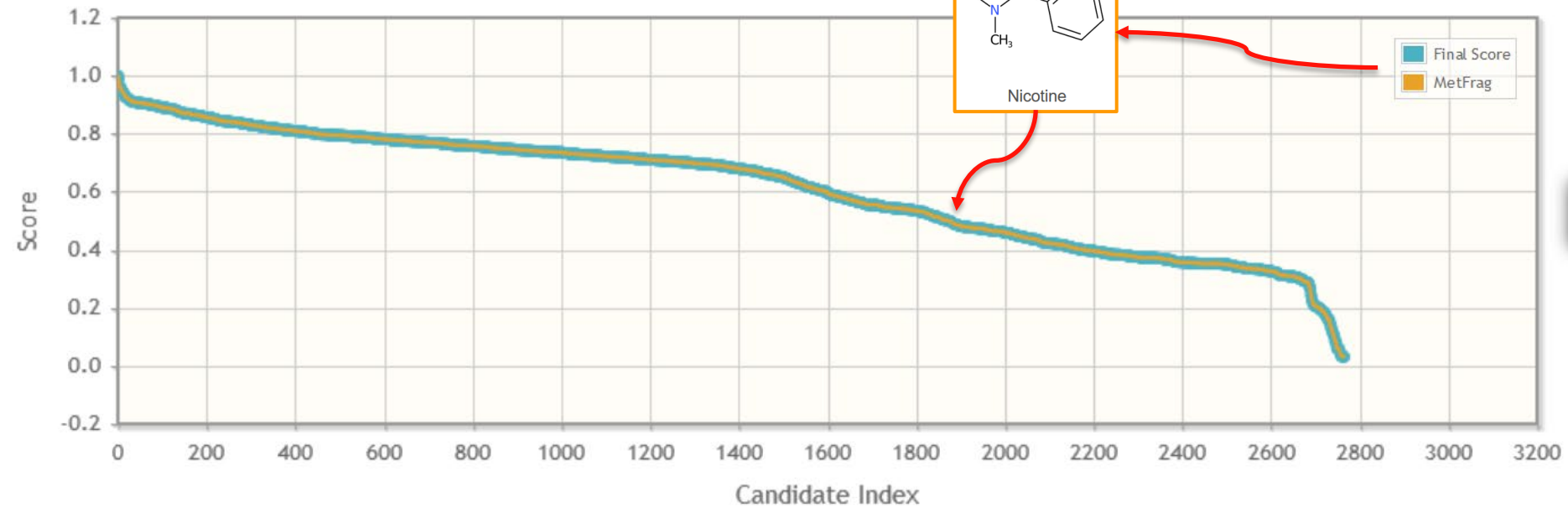


Key Challenge: MS and MS/MS alone is not enough!



Statistics

Candidate Score Distribution



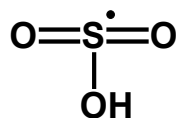
Status Quo in 2016: MetFrag Relunched ...



m/z $[M-H]^-$
213.9637
 ± 5 ppm

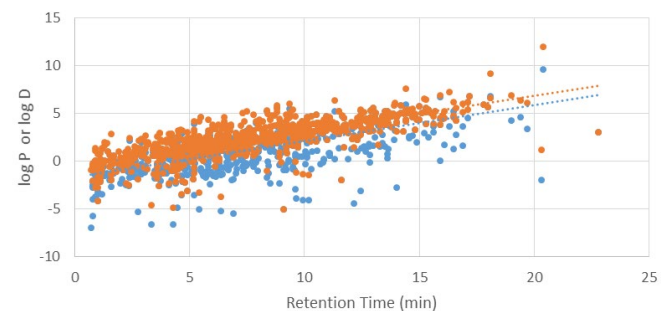
Elements: C, N, S

5 ppm
0.001 Da



RT: 4.54 min

355 InChI/RTs



or



References
Tox. Data
Data Sources
Exposure Info
MS-ready links



Suspect Lists



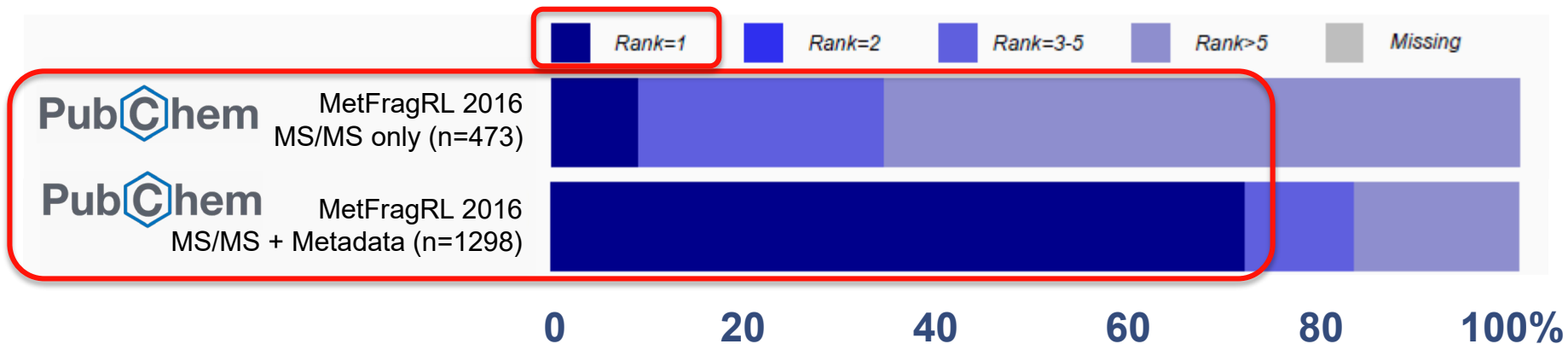
MS/MS

134.0054	339689
150.0001	77271
213.9607	632466



MetFragRL + PubChem + MS/MS + Metadata

- Adding literature, references & RT boosts to ~71 % rank 1!



MetFragRL + PubChem + MS/MS + Metadata



BUT ...databases grow ... ID performance drops

... and run times rise ... (a lot!)



Can we break down PubChem into useful bits?



PubChem Compound TOC ? 49,493,641

▶ Agrochemical Information ? 3,045

▶ Associated Disorders and Diseases ? 20,847

▶ Biologic Description ? 2,056,521

▶ Biological Test Results ? 3,622,920

▶ Biomolecular Interactions and Pathways ? 125,253

▶ Chemical and Physical Properties ? 263,015

▶ Classification ? 1,454,824

▶ Drug and Medication Information ? 17,922

▶ Food Additives and Ingredients ? 8,414

▶ Identification ? 4,968

▶ Information Sources ? 20,271,277

▶ Literature ? 1,833,941

▶ Names and Identifiers ? 1,275,170

▶ Patents ? 36,351,418

▶ Pharmacology and Biochemistry ? 110,628

▶ Related Records ? 9,224,590

▶ Safety and Hazards ? 149,319

▶ Spectral Information ? 480,730

▶ Structures ? 9,117,635

▶ Toxicity ? 114,012

▶ Use and Manufacturing ? 115,321

Chemical Safety ? 147,023

▼ Agrochemical Information ? 3,045

Agrochemical Category ? 1,903

Agrochemical Transformations ? 1,383

EU Pesticides Data ? 1,213

USDA Pesticide Data Program ? 579

▼ Use and Manufacturing ? 115,321

▶ Uses ? 27,314

Consumption Patterns ? 1,168

Formulations/Preparations ? 4,742

General Manufacturing Information ? 52,353

Impurities ? 870

Methods of Manufacturing ? 6,192

Overview ? 1,179

Sampling Procedures ? 579

U.S. Exports ? 1,038

U.S. Imports ? 1,547

U.S. Production ? 5,500

Use Classification ? 63,558

PubChem Furathiocarb (Compound)

CONTENTS ↑

7 Agrochemical Information

7.1 Agrochemical Category

Insecticides

▶ EU Pesticides Database

7.2 Agrochemical Transformations

Furathiocarb has known environmental transformation products that include [carbofuran](#).

S60 | *SWISSPEST19* | *Swiss Pesticides and Metabolites from Kiefer et al 2019* | DOI:10.5281/zenodo.3544759

▶ NORMAN Suspect List Exchange

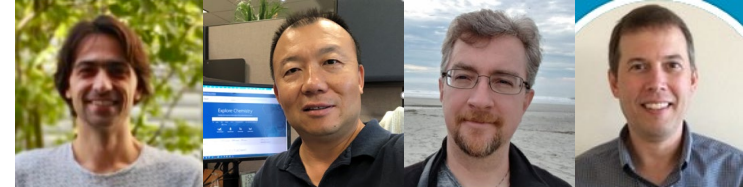
7.3 EU Pesticides Data

Active Substance	furathiocarb
Status	Not Approved [Reg. (EC) No 1107/2009]
Categories	Insecticides

<https://pubchem.ncbi.nlm.nih.gov/compound/Furathiocarb#section=Agrochemical-Information>

Schymanski et al. (2021) DOI: [10.1186/s13321-016-0115-9](https://doi.org/10.1186/s13321-016-0115-9)

Introducing ...



PubChem Compound TOC ? 49,493,641

Agrochemical Information ? 3,045

Associated Disorders and Diseases ? 20,847

Biologic Description ? 2,056,521

Biological Test Results ? 3,622,920

Biomolecular Interactions and Pathways ? 125,253

Chemical and Physical Properties ? 263,015

Classification ? 1,454,824

Drug and Medication Information ? 17,922

Food Additives and Ingredients ? 8,414

Identification ? 4,968

Information Sources ? 20,271,277

Literature ? 1,833,941

Names and Identifiers ? 1,275,170

Patents ? 36,351,418

Pharmacology and Biochemistry ? 110,628

Related Records ? 9,224,590

Safety and Hazards ? 149,319

Spectral Information ? 480,730

Structures ? 9,117,635

Toxicity ? 114,012

Use and Manufacturing ? 115,321

Chemical Safety ? 147,023

PubChemLite EXPOSOMICS

~370,000 entries "small"



zenodo

Search



Upload

Communities

emma.schymanski@uni.lu

October 31, 2020

Dataset Open Access

Edit

New version

PubChemLite for Exposomics

by [Todor Kondic](#), [Paul Thiessen](#), [Jeff Zhang](#)

(<https://pubchem.ncbi.nlm.nih.gov/classification/#hid=72>). With this release, there is now just one tier1 plus two new categories (Associated Disorders &

13 compounds (31 Oct 2020) compiled from 10 categories: cinfo, FoodRelated, PharmacolInfo, SafetyInfo, ToxicityInfo, etc.

17 InChIKey first block, reporting the structure from the most es that will be ignored by MetFrag (salts, disconnected sition metals) have been removed. The Patent and PubMed ID PubChem FTP site. The "AnnoTypeCount" term counts how d, the subsequent column (named per category) counts the able in the next sub-category of the TOC entry.

Database Settings

Database: PubChemLite_31Oct2020

Neutral Mass: 229.10948 Search ppm: 5

Formula: C9H16ClN5

Identifiers:

Retrieve Candidates 4 Candidates

Schymanski et al. (2021) DOI: [10.1186/s13321-016-0115-9](https://doi.org/10.1186/s13321-016-0115-9)



U.S. National Library of Medicine
National Center for Biotechnology Information

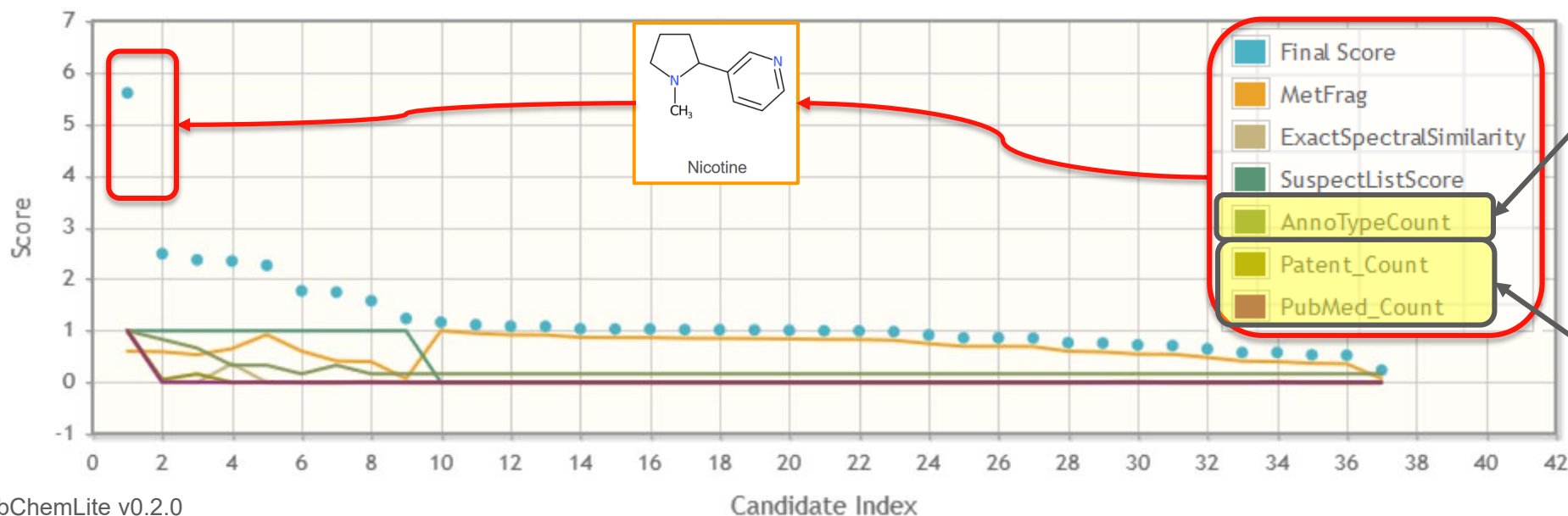


MetFragRL + PubChemLite: tailor-made database + metadata



Statistics

Candidate Score Distribution



PubChem Compound TOC ? 49,493,641

- Agrochemical Information ? 3,045
- Associated Disorders and Diseases ? 20,847
- Biologic Description ? 2,056,521
- Biological Test Results ? 3,622,920
- Biomolecular Interactions and Pathways ? 125,253
- Chemical and Physical Properties ? 263,015
- Classification ? 1,454,824
- Drug and Medication Information ? 17,922
- Food Additives and Ingredients ? 8,414
- Identification ? 4,968
- Information Sources ? 20,271,277
- Literature ? 1,833,941
- Names and Identifiers ? 1,275,170
- Patents ? 36,351,418
- Pharmacology and Biochemistry ? 110,628
- Related Records ? 9,224,590
- Safety and Hazards ? 149,319
- Spectral Information ? 480,730
- Structures ? 9,117,635
- Toxicity ? 114,012
- Use and Manufacturing ? 115,321
- Chemical Safety ? 147,023



PubChemLite v0.2.0

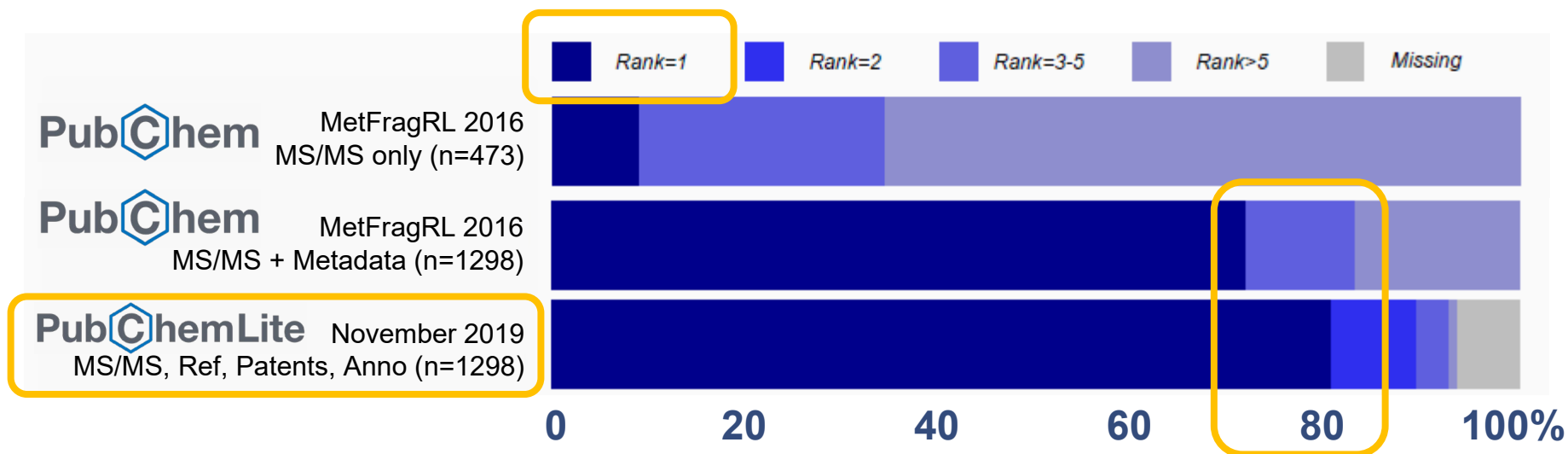
Candidate Index

MetFrag+PubChemLite+Formula+MoNA+SusDat+Pat+Refs+Anno + <https://massbank.eu/MassBank/RecordDisplay?id=EQ300804>



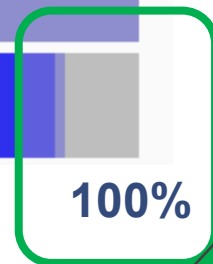
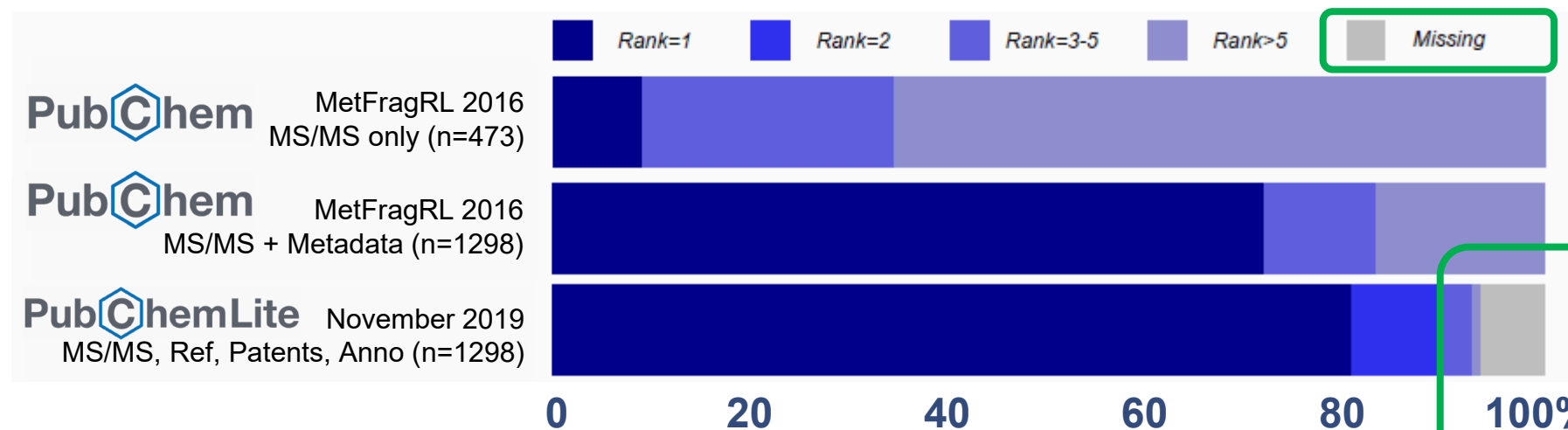
How does PubChemLite perform?

- ~110 M => ~370 K ... how does this influence performance?



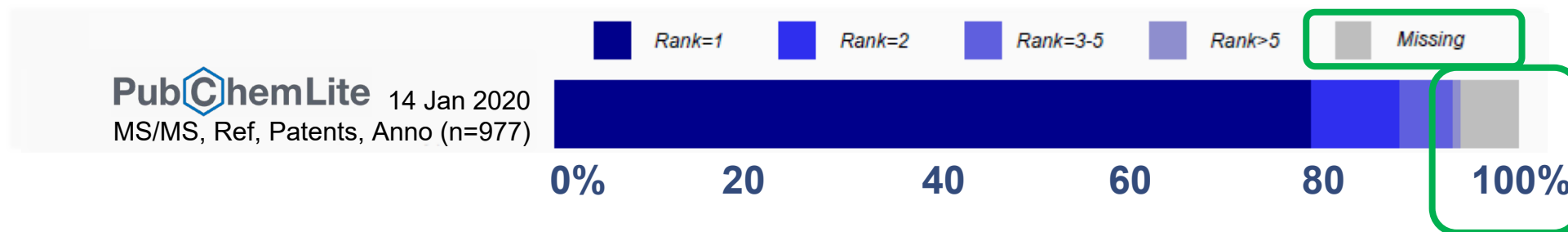
How does PubChemLite perform?

- ~110 M => ~370 K ... how does this influence performance?



norman
suspects

Assessing the Missing Entries in PubChemLite



- ▼ NORMAN Suspect List Exchange Classification ? ↗ 117,037
 - ▶ S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and
 - ▶ S25 | OECDPFAS | List of PFAS from the OECD ? 3,680
 - ▶ S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ? 647
 - ▶ S60 | SWISSPEST19 | Swiss Pesticides and Metabolites ? 1,358
 - ▶ S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI ? 574
 - ▶ S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag ? 258
 - ▶ S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem ? 97

Transformation Products: Filling the Data Gaps!



PubChem NORMAN Suspect List Exchange

- NORMAN Suspect List Exchange Classification **113,080**
 - S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) **3,85**
 - S25 | OECDPFAS | List of PFAS from the OECD **3,677**
 - S36 | UBAPMT | Potential Persistent, Mobile and Toxic (PMT) substances **254**
 - S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium **885**
 - S60 | SWISSPEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 **1,343**
 - S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI **574**
 - S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag **258**
 - S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem **102**
 - S69 | LUXPEST | Pesticide Screening List for Luxembourg **386**
 - S72 | NTUPHTW | Pharmaceutically Active Substances from National Taiwan University **1,068**
 - S75 | CyanoMetDB | Comprehensive database of secondary metabolites from cyanobacteria **2,088**
- S00 | SUSDAT | Merged NORMAN Suspect List: SusDat **99,130**
- S01 | MASSBANK | NORMAN Compounds in MassBank EU **7,164**
- S02 | STOFFIDENT | HSWT/LfU STOFF-IDENT Database of Water-Relevant Substances **11,261**
- S03 | NORMANCT15 | NORMAN Collaborative Trial Targets and Suspects **624**
- S04 | UJIBADE | Target List from UJI used in Bade et al 2015 **542**

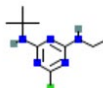

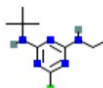

- Pharmacology and Biochemistry **112,039**
 - Human Metabolite Information **64,199**
 - Metabolism/Metabolites **8,204**
 - Transformations **5,857**

PubChem Terbutylazine (Compound)

8.5 Transformations

Page 3 of 25 items View More Rows & Details Download

SORT BY Please Choose One

Predecessor Image	Predecessor Name	Transformation	Successor Image	Successor Name	Evidence DOI
	Terbutylazine	Mammalian metabolism		6-Chloro-1,3,5-triazine-2,4-diamine	10.5281/zenodo.382
	Terbutylazine	Deethylation		Terbutylazine-desethyl	10.1007/s13361-017-

Transformation Products: Filling the Data Gaps!

PubChem Terbutylazine (Compound)

7 Agrochemical Information

7.1 Agrochemical Category

Pesticides -> Herbicides -> [Triazine herbicides](#) -> Chlorotriazine herbicides

S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag | DOI:10.5281/zenodo.3754448

▶ [NORMAN Suspect List Exchange](#)

7.2 Agrochemical Transformations

Terbutylazine has known environmental transformation products that include [Terbutylazine-2-hydroxy](#), [Terbutylazine-desethyl](#), and [Terbutylazine-desethyl-2-hydroxy](#).






S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag | DOI:10.5281/zenodo.3754448

▶ [NORMAN Suspect List Exchange](#)

Terbutylazine has known environmental transformation products that include CSAA036479, CSAA04949, CSCD648241, CSCD692760, GS31398, MT1, GS 26379, MT13, GS 23158, Terbutylazine metabolite MT14, Terbutylazine metabolite MT23, and Terbutylazine metabolite MT24.

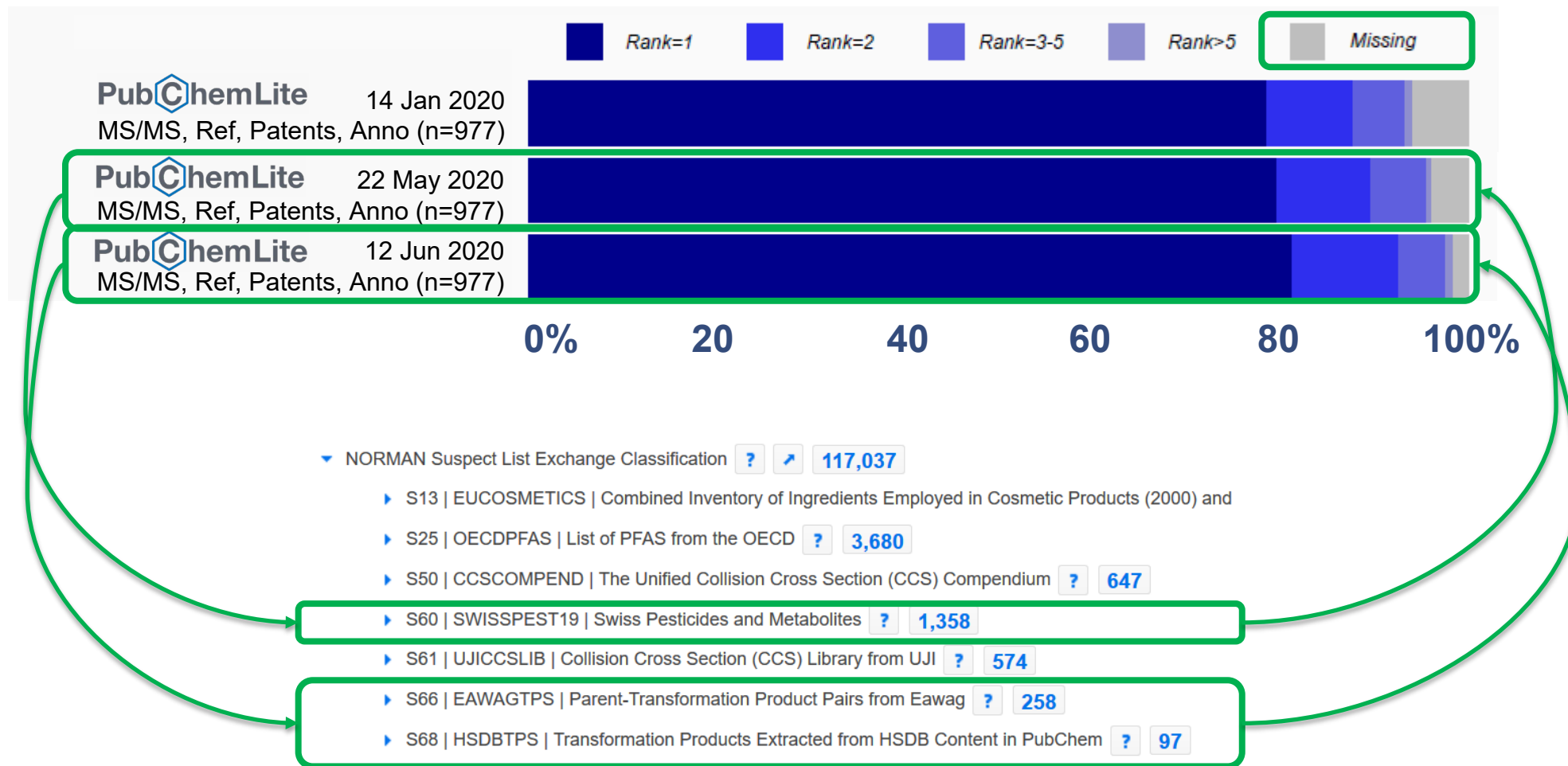
S60 | SWISSPEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 | DOI:10.5281/zenodo.3544759

▶ [NORMAN Suspect List Exchange](#)

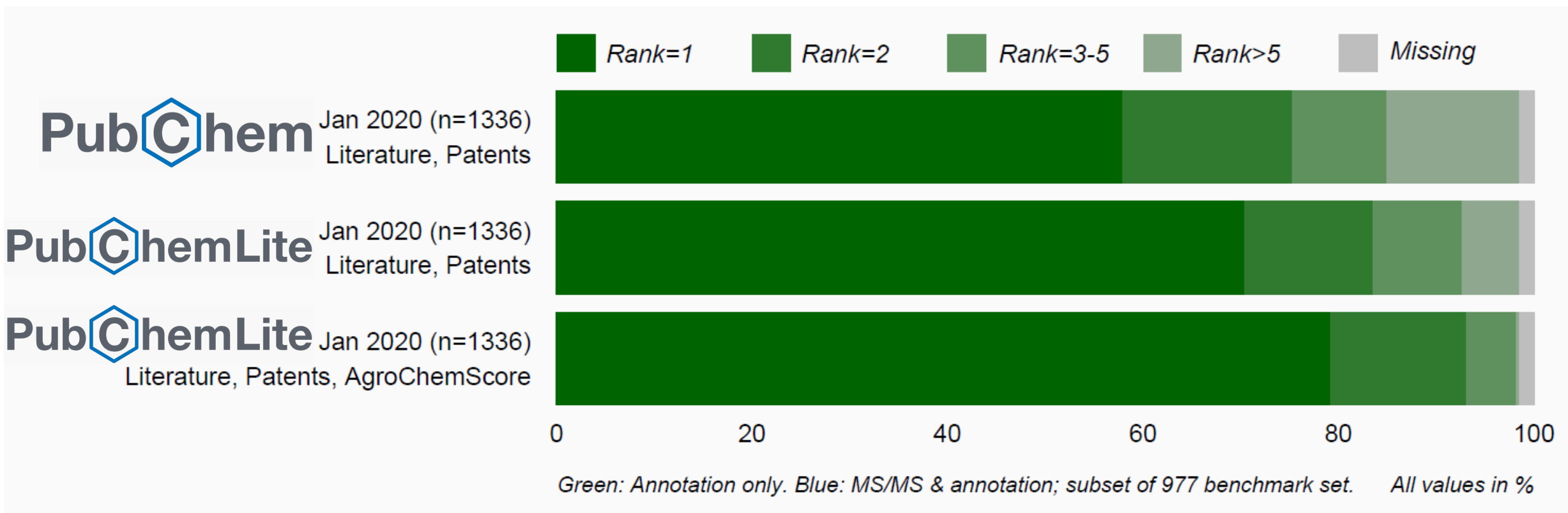
▼ Agrochemical Information 	3,045
Agrochemical Category 	1,903
Agrochemical Transformations 	1,383
EU Pesticides Data 	1,213
USDA Pesticide Data Program 	579



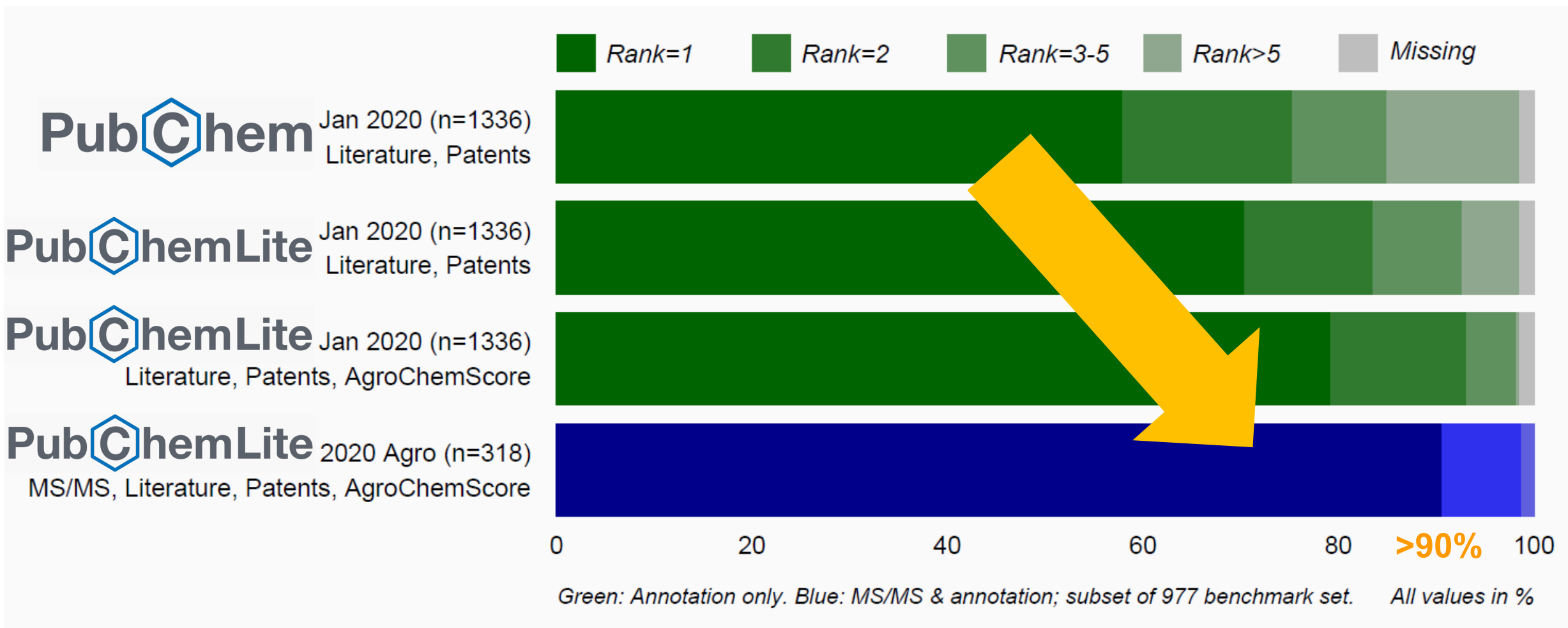
Assessing the Missing Entries in PubChemLite



Influence of the Annotation Content in PubChemLite



Influence of the Annotation Content in PubChemLite



MetFrag Example: Nicotine and Disease Associations

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



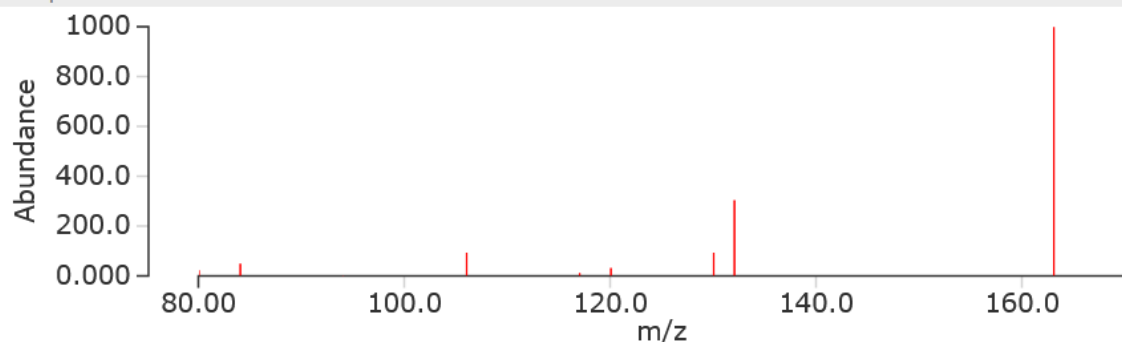
A screenshot of the MetFrag web interface. The top left features the MetFrag logo, a green 3D cube with the text "MetFrag". To its right, the text "MetFrag" is displayed in a large font, followed by the subtitle "In silico fragmentation for computer assisted identification of metabolite mass spectra". Below this is a green horizontal bar labeled "Database Settings". Underneath, there are input fields for "Database:", "Neutral Mass:", "Formula:", and "Identifiers:". The "Database:" field is currently set to "PubChemLite_01Jan2021" and has a dropdown menu open showing several options, with "PubChemLite_01Jan2021_exposomics" highlighted. To the right of these fields is a "Parent Ion:" field with a dropdown menu set to "[M+H]+" and a "Calculate" button. Below the input fields are two buttons: "Retrieve Candidate" and "Download Candidates". At the bottom left, there is a link for "Candidate Filter & Score Settings". A red arrow points from the "PubChemLite_01Jan2021_exposomics" option in the dropdown menu to the "PubChemLite EXPOSOMICS" logo in the top right corner of the slide.

MetFrag Example: Nicotine and Disease Associations

MassBank Record: EQ300801

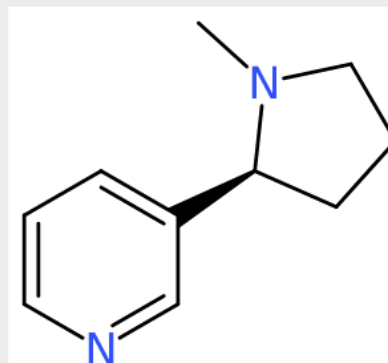
Nicotine; LC-ESI-QFT; MS2; CE: 15; R=35000; [M+H]⁺

Mass Spectrum



[metabolomics-usi visualisation](#)

Chemical Structure



PK\$NUM_PEAK: 9

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

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

//

ACCESSION: EQ300801

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

DATE: 2015.08.25

AUTHORS: Beck B, Stravs M, Schymanski E, Singer H, Department of Environmental



MetFrag Example: Nicotine and Disease Associations



Candidate Filter & Score Settings

Candidate Filters

- Element
- Element
- Substru
- Substru
- Substru
- Substru
- Minimum
- Maximum
- Suspec

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 14 item(s) selected

Select Item(s) 4 of 14 item(s) selected

Fragmentation Settings & Processing



MetFrag Example: Nicotine and Disease Associations



MassBank
High Quality Mass Spectral Database

PubChemLite
EXPOSOMICS

Results

Weights

MetFrag (1st)	<input type="range"/>	100	%
ExactSpectralSimilarity (2nd)	<input type="range"/>	100	%
AnnoTypeCount (3rd)	<input type="range"/>	100	%
DisorderDisease (4th)	<input type="range"/>	100	%
Patent_Count (5th)	<input type="range"/>	100	%
PubMed_Count (6th)	<input type="range"/>	100	%

Experimental evidence / values

Download Results

Filter Candidates by explained MS/MS Peaks

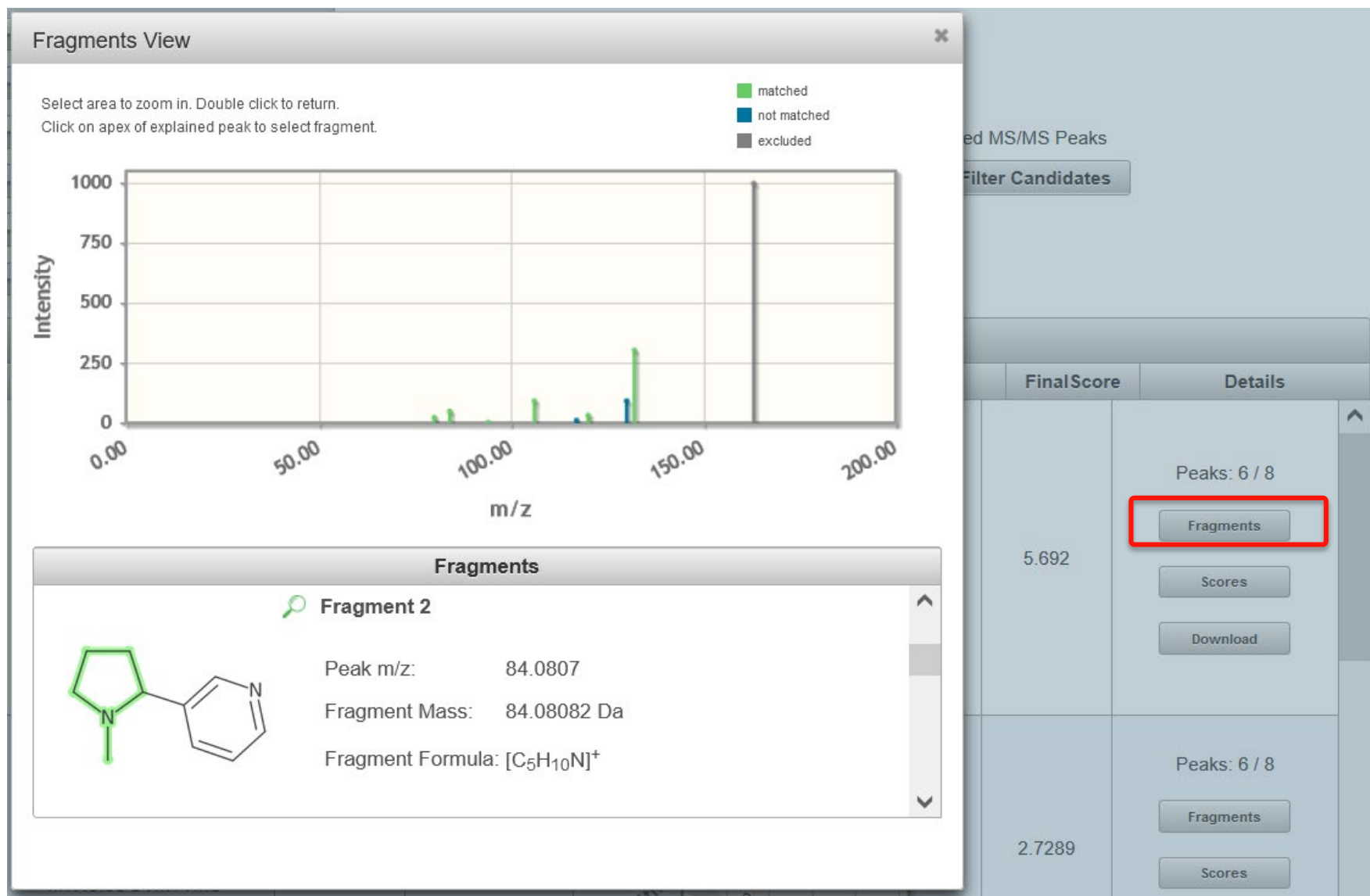
MS/MS Peaks

Disease/Disorder information available

#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1	 3-[(2S)-1-methylpyrrolidin-2-yl]pyridine	89594 InChIKeyBlock1 = SNICXCGAKADSCV	162.1157	C ₁₀ H ₁₄ N ₂		5.692	Peaks: 6 / 8 <input type="button" value="Fragments"/> <input type="button" value="Scores"/> <input type="button" value="Download"/>
2	 3-piperidin-2-ylpyridine	2181 InChIKeyBlock1 = MTXSIJUGVMTTMU	162.1157	C ₁₀ H ₁₄ N ₂		2.7289	Peaks: 6 / 8 <input type="button" value="Fragments"/> <input type="button" value="Scores"/> <input type="button" value="Download"/>



MetFrag Example: Nicotine and Disease Associations



MetFrag Example: Nicotine and Disease Associations



Level 2: Probable structure

- a) by library spectrum match
- b) by diagnostic evidence

MS, MS², Library MS² >0.9 sim
MS, MS², Exp. data

Scores View

Candidate Name: 3-[(2S)-1-methylpyrrolidin-2-yl]pyridine
Candidate Identifier: 89594|47593

	Name	Normalized Value	Raw Value
▶	MetFrag	0.692	246.1726
▶	ExactSpectralSimilarity	1.0	1.0
▶	AnnoTypeCount	1.0	10.0
▶	DisorderDisease	1.0	2.0
▶	Patent_Count	1.0	120010.0
▶	PubMed_Count	1.0	35984.0

Normalized Scores

FinalScore: 5.692

Details

Peaks: 6 / 8

Fragments

Scores

Download



MetFrag Example: Nicotine and Disease Associations



MassBank
High Quality Mass Spectral Database

PubChemLite
EXPOSOMICS

Results

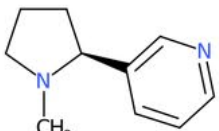
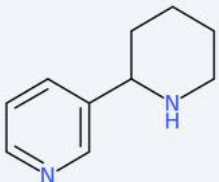
Weights	
MetFrag (1st)	100 %
ExactSpectralSimilarity (2nd)	100 %
AnnoTypeCount (3rd)	100 %
DisorderDisease (4th)	100 %
Patent_Count (5th)	100 %
PubMed_Count (6th)	100 %

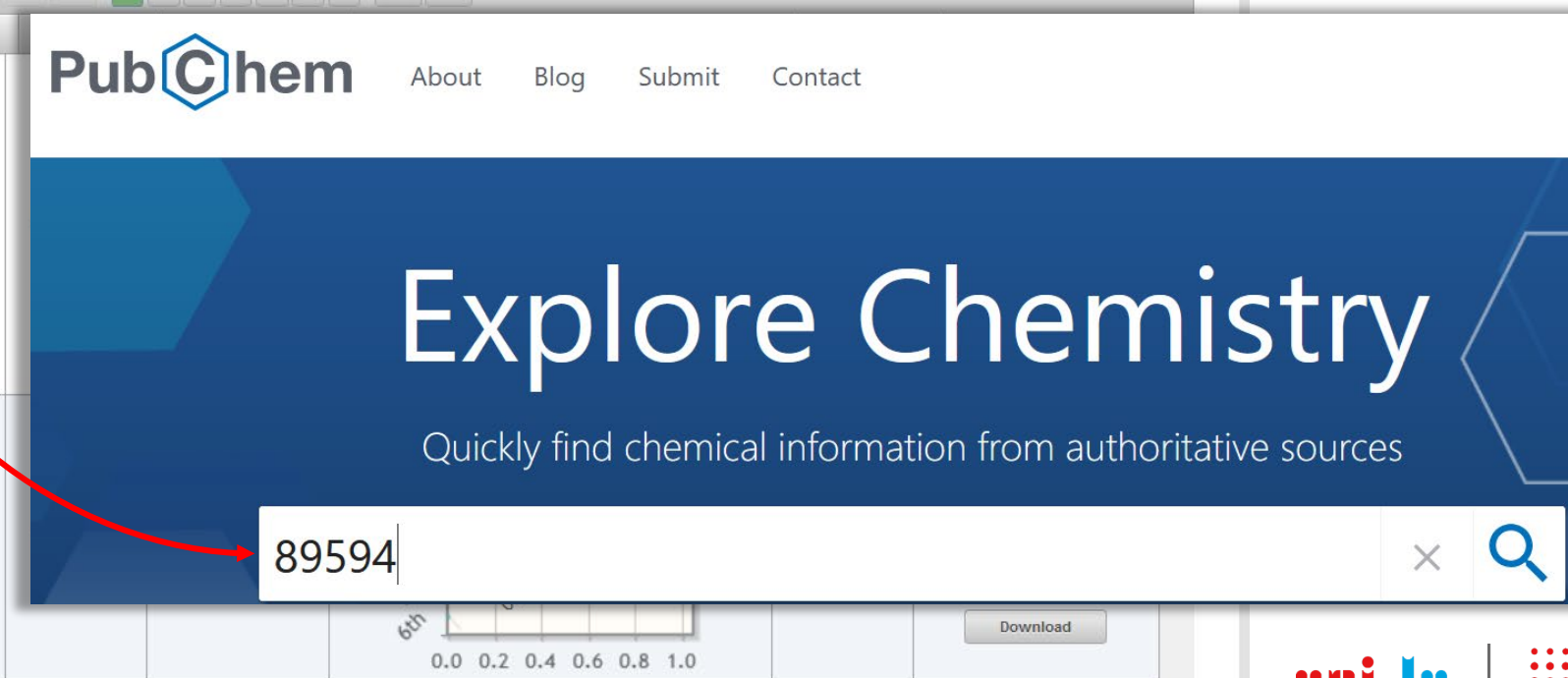
Download Results

Filter Candidates by explained MS/MS Peaks

MS/MS Peaks Filter Candidates

Disease/Disorder information available

#	Molecule	Identifier
1	 3-[(2S)-1-methylpyrrolidin-2-yl]pyridine	89594 InChIKeyBlock1 = SNICXCGAKADSCV
2	 3-piperidin-2-ylpyridine	2181 InChIKeyBlock1 = MTXSIJUGVMTTMU



PubChem About Blog Submit Contact

Explore Chemistry

Quickly find chemical information from authoritative sources

89594



PubChem Annotations: Disease Associations

PubChem Nicotine (Compound)

14 Associated Disorders and Diseases

246 items View More Rows & Details

Download

Cite

Download

Disease	Evidence Type	Evidence PMID
Abdominal Pain	marker/mechanism	9179092
Abnormalities, Drug-Induced	marker/mechanism	16193507
Acute Kidney Injury	marker/mechanism	21511693
Airway Obstruction	marker/mechanism	33082140
Akinetic Mutism	therapeutic	19455735

1 2 3 ... 50 Next >

▶ Comparative Toxicogenomics Database (CTD)

Disease	References
	PubMed: 9137998 , 8294547 , 10636262 , 6890513 ,

CONTENTS

7 Drug and Medication Information

8 Agrochemical Information

9 Pharmacology and Biochemistry

10 Use and Manufacturing

11 Identification

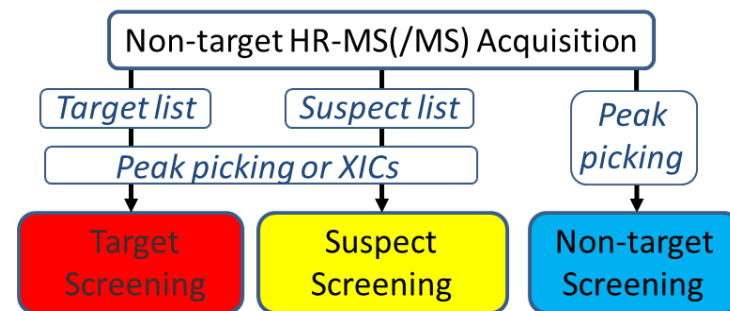
12 Safety and Hazards

13 Toxicity

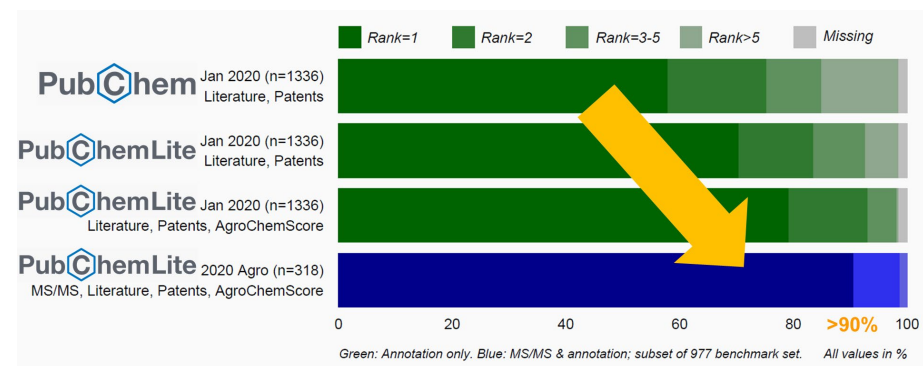
14 Associated Disorders and Diseases

“Take home” Messages

- Background of HR-MS and Cheminformatics
- Introduction to MetFrag and PubChemLite
 - Mass spectral libraries help deliver Level 2a IDs
 - Annotation content is extremely powerful
- The tools are there for clinical applications
 - Non-target HR-MS can deliver quick information
- Help contribute by adding your knowledge!



 **MassBank**
High Quality Mass Spectral Database



Expert Knowledge is YOUR Knowledge!

<https://ftp.ncbi.nlm.nih.gov/pubchem/Other/Submissions/>

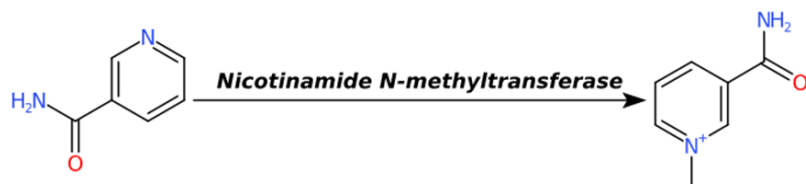
- Help us help you! Add data with FAIR templates

Chemical Structures

PubChem_CID	Name	SMILES	InChIKey
2256	Atrazine	CCNC1=NC(=NC(=N1)Cl)NC(C)C	MXWJVTOOROXGIU-UHFFFAOYSA-N
2328	Bentazone	CC(C)N1C(=O)C2=CC=CC=C2NS1(=O)=O	ZOMSMJKLGFBRBS-UHFFFAOYSA-N
3030	Dicamba	COC1=C(C=CC(=C1C(=O)O)Cl)Cl	IWEDIXLBFLEXBO-UHFFFAOYSA-N
3120	Diuron	CN(C)C(=O)NC1=CC(=C(C=C1)Cl)Cl	XMTQQYYKAHVGBJ-UHFFFAOYSA-N

Transformations

Predecessor_CID	Predecessor_Name	Transformation	Successor_CID	Successor_Name	Biosystem
13101	6PPD	Ozone	154926030	6PPD-quinone	Environment
2256	Atrazine	Environmental	13878	Deisopropyl-atrazine	Soil
2256	Atrazine	Mammalian metabolism	135408770	Ammeline	Mammal
2256	Atrazine	Fungal metabolism	22563	Desethyl-atrazine	Fungus



FAIR chemical structures in the Journal of Cheminformatics

Emma L. Schymanski and Evan E. Bolton

Letter to the Editor | 7 July 2021

i The [Letter Response to this article](#) has been published in *Journal of Cheminformatics* 2021 **13**:49

Reply to "FAIR chemical structure in the Journal of Cheminformatics"

Rajarshi Guha, Nina Jeliaskova, Egon Willighagen and Barbara Zdrzil

Letter Response | 7 July 2021

i The [Letter to the Editor to this article](#) has been published in *Journal of Cheminformatics* 2021 **13**:50

Thank you!

PubChemLite EXPOSOMICS

Email: emma.schymanski@uni.lu

Twitter: [@Eschymanski](https://twitter.com/Eschymanski)

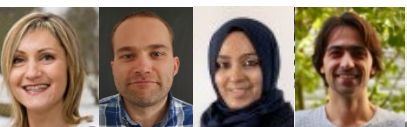
Slides @ DOI: [10.5281/zenodo.5354574](https://doi.org/10.5281/zenodo.5354574)

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

<https://pubchem.ncbi.nlm.nih.gov/>

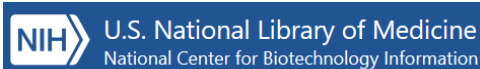
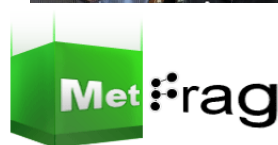
<https://massbank.eu/MassBank/>

https://wwwen.uni.lu/lcsb/research/environmental_cheminformatics/



PubChem

... and
team



Luxembourg National
Research Fund

