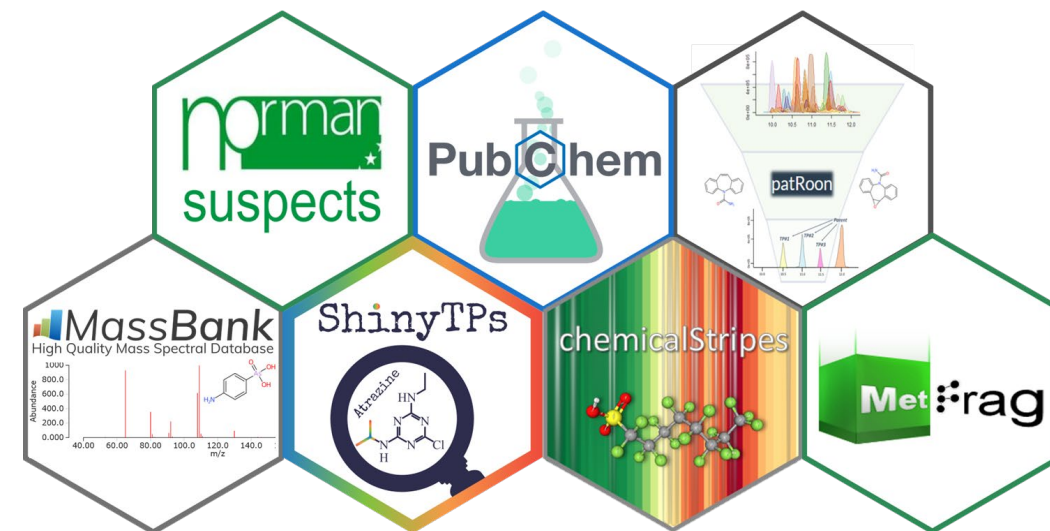


Exploring Environmental Chemical Space in R



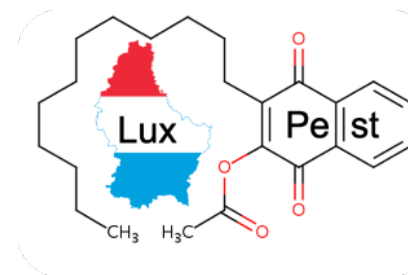
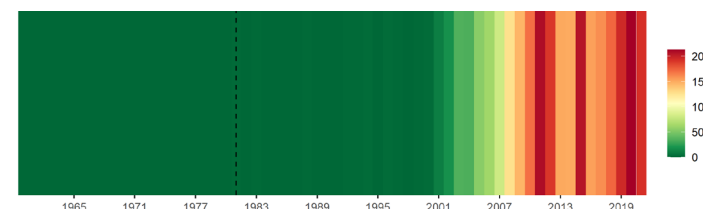
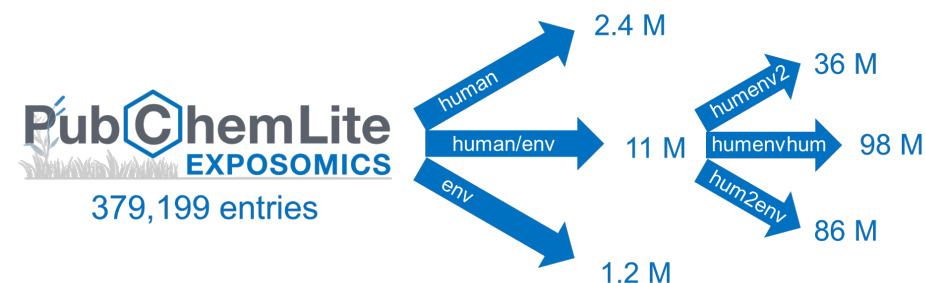
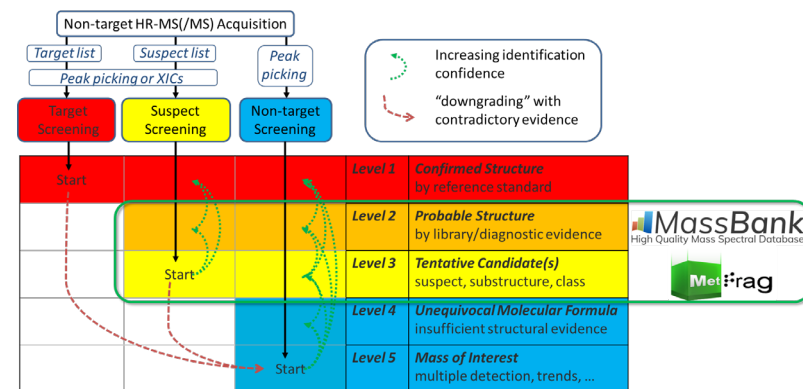
Prof. Dr. Emma L. Schymanski
(plus many, many colleagues and collaborators!)

Environmental Cheminformatics Group,
Luxembourg Centre for Systems Biomedicine, University of Luxembourg
emma.schymanski@uni.lu / [@ESchymanski](https://twitter.com/ESchymanski) / [@schymane@mstdn.social](https://www.instagram.com/schymane)
https://www.wen.uni.lu/lcsb/research/environmental_cheminformatics/

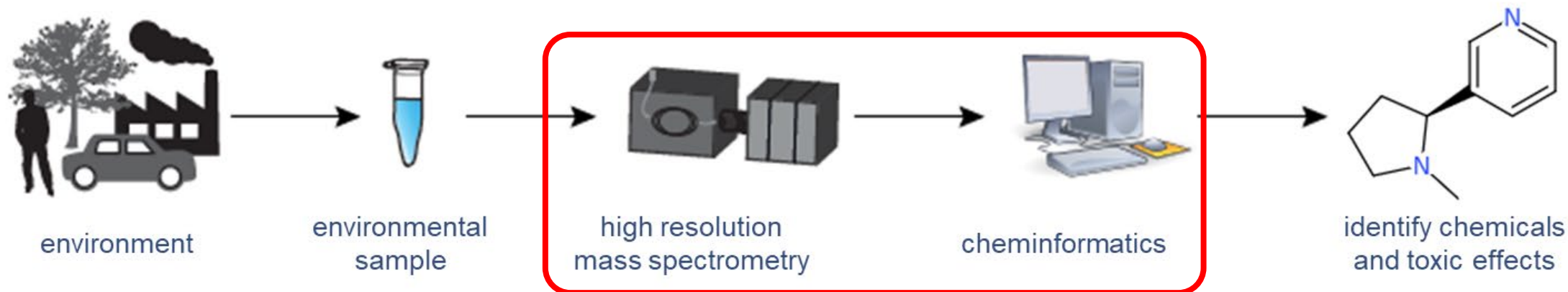


Outline

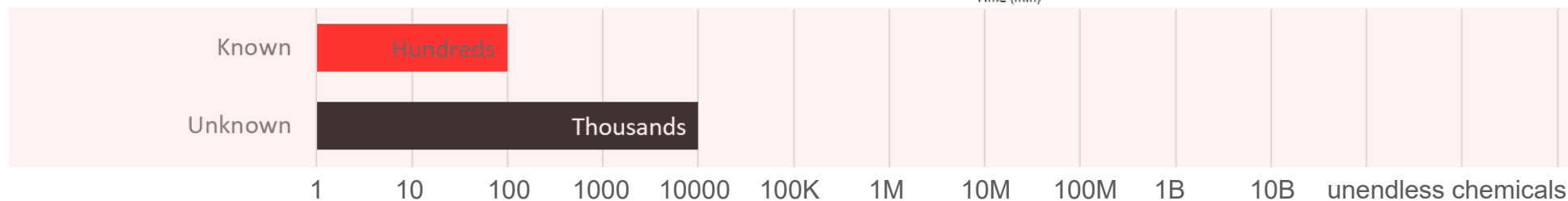
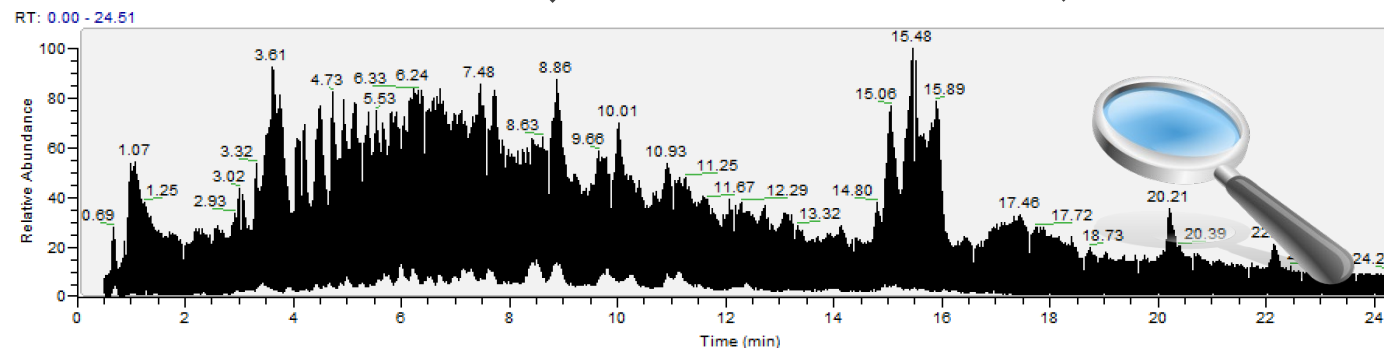
- Identifying Chemicals with NT-HRMS
 - Open Resources for Identification
 - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases & Workflows
 - Defining Relevant Chemical Space
 - Transformations: Known & Predicted
- No package too small?
- Applications in Environmental Studies



Environmental Cheminformatics & Non-target HR-MS



High resolution
mass spectrometry
AND connecting
chemical knowledge

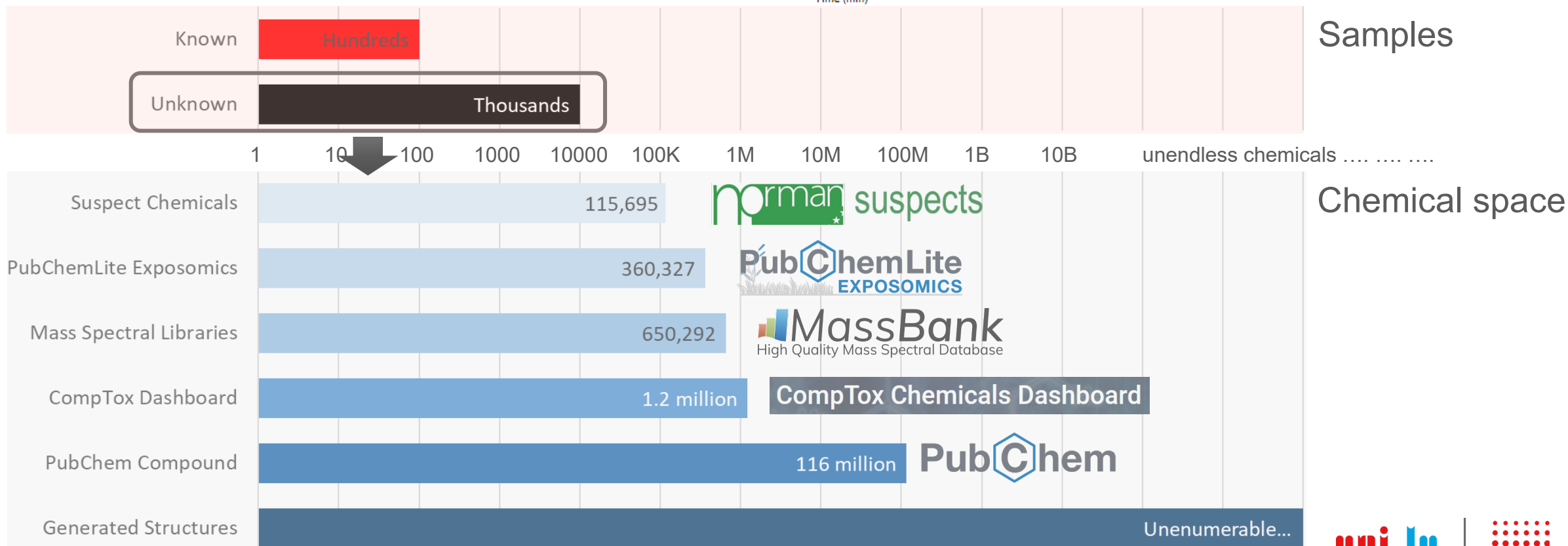
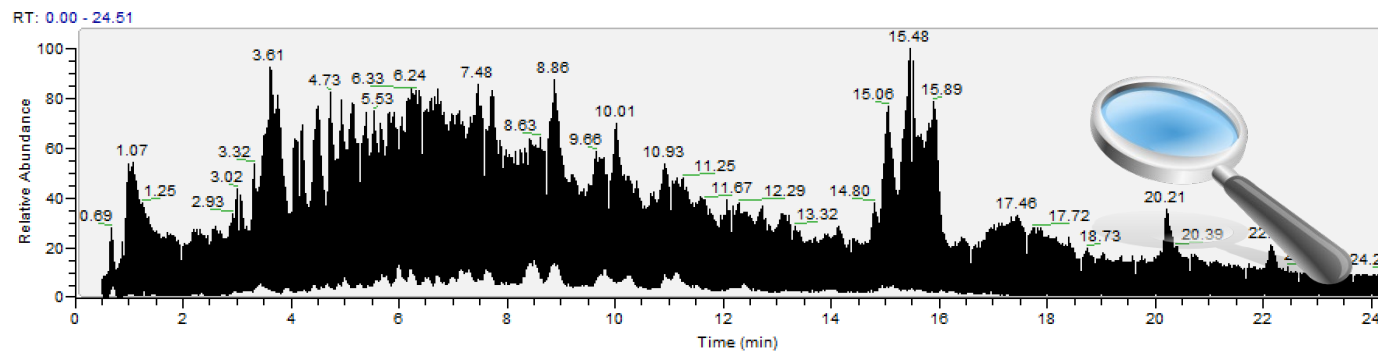


Samples

Environmental Cheminformatics & Non-target HR-MS

High resolution
mass spectrometry

AND connecting
chemical knowledge



FAIR and Open Science

<https://doi.org/10.1038/s44221-022-00014-z>

Water science must be Open Science

Emma L. Schymanski & Stanislaus J. Schymanski

Since water is a common good, the outcome of water-related research should be accessible to everyone. Since Open Science is more than just open access research articles, journals must work with the research community to enable fully open and FAIR science

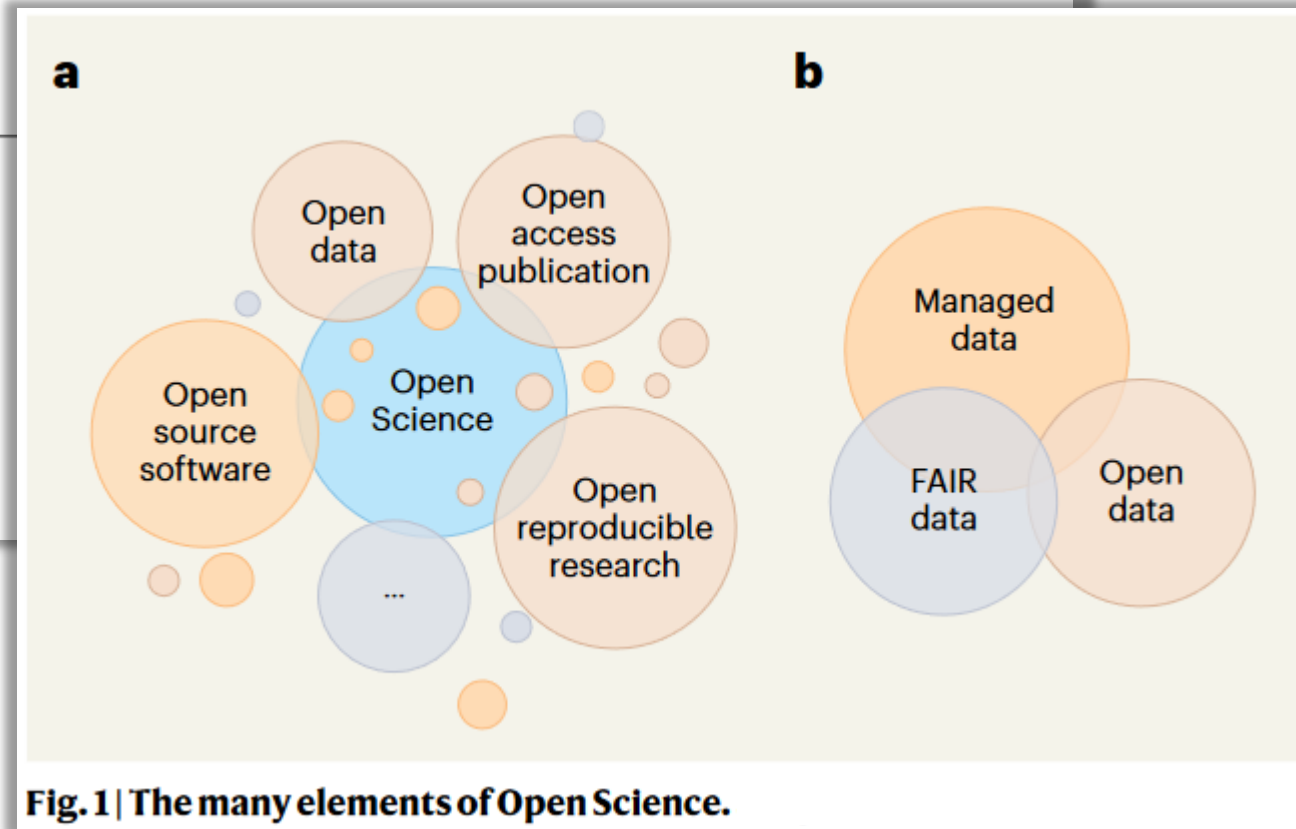
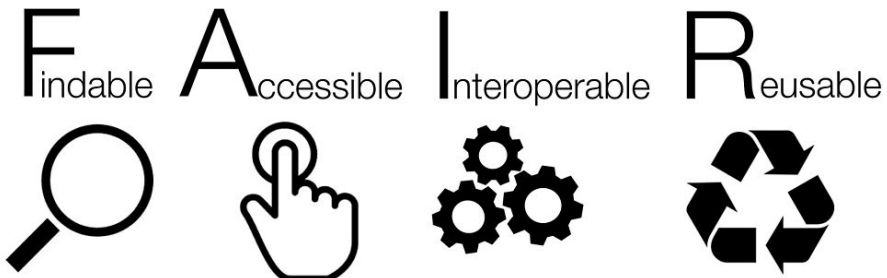
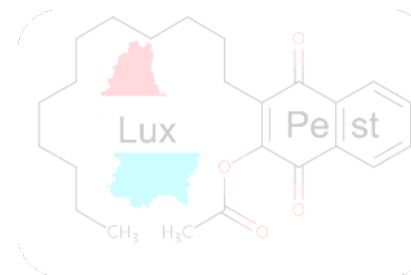
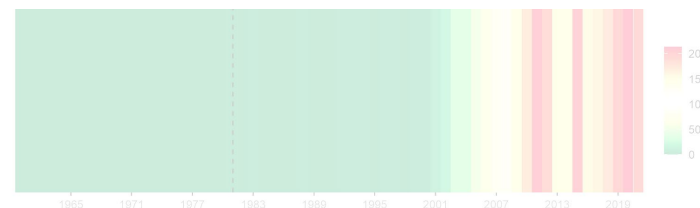
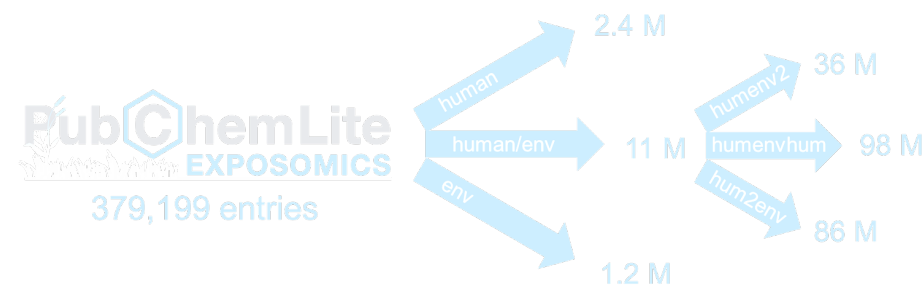
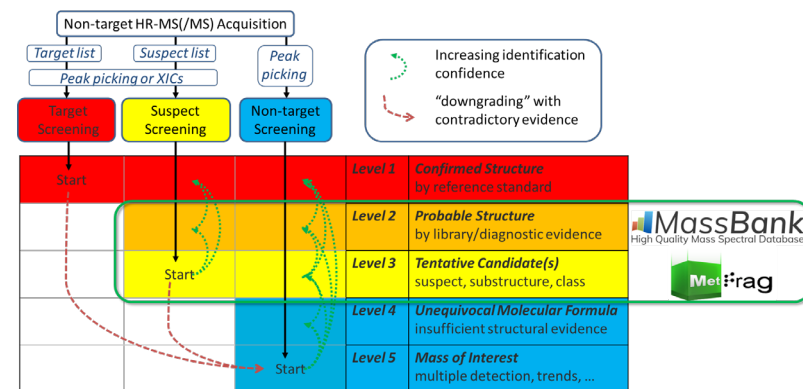


Fig. 1 | The many elements of Open Science.

Outline

- Identifying Chemicals with NT-HRMS
 - Open Resources for Identification
 - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases & Workflows
 - Defining Relevant Chemical Space
 - Transformations: Known & Predicted
- No package too small?
- Applications in Environmental Studies



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

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

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

 [Search](#) [Contents](#) [Download](#) [More ▼](#)

MassBank Europe


High Quality Mass Spectral Database

MassBank Record: LU040605

Search for:

Compound Information

Compound name

Exact Mass

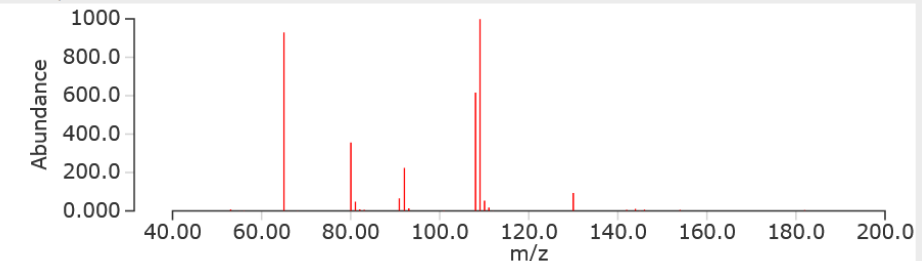
AND

Formula (e.g. C₆H₇N₅, C₅H⁺N₅, C₅⁺)

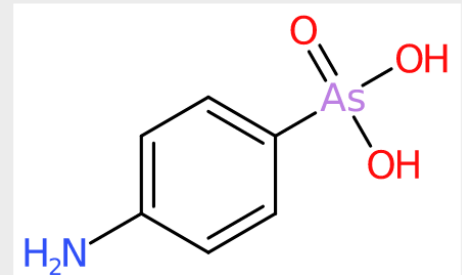
AND

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

Mass Spectrum



Chemical Structure









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

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

Overview

Repositories 18

Projects 1

Packages



MassBank consortium
The MassBank consortium develops an ecosystem of databases
10 followers <http://www.massbank.jp/> massbank-develop@massbank.jp

Pinned

 **MassBank-web** Public

The web server application and directly connected components for a MassBank web server

Java

13

22

 **MassBank-data** Public

Official repository of open data MassBank records

Shell

56

56

2 weeks ago

 meier-rene

2023.09

9b1255e

Compare

Release version 2023.09 Latest

This release features:

- 810 new spectra from University of Antwerp project METOX
- 2072 new spectra from UFZ Leipzig
- 242 new spectra from Eawag
- 56 new spectra from University of Birmingham
- minor fixes

Assets 6

 MassBank.json	328 MB	2 weeks ago
 MassBank.sql	420 MB	2 weeks ago
 MassBank_NIST.msp	95.2 MB	2 weeks ago
 MassBank_RIKEN.msp	114 MB	2 weeks ago
 Source code (zip)		2 weeks ago
 Source code (tar.gz)		2 weeks ago



8

RMassBank: Processing MS/MS in R for MassBank

RMassBank

platforms all rank 757 / 2230 support 0 / 0 in Bioc 11 years
build warnings updated before release dependencies 143

DOI: [10.18129/B9.bioc.RMassBank](https://doi.org/10.18129/B9.bioc.RMassBank)

Workflow to process tandem MS files and build MassBank records

Bioconductor version: Release (3.17)

Workflow to process tandem MS files and build MassBank records. Functions include automated extraction of tandem MS spectra, formula assignment to tandem MS fragments, recalibration of tandem MS spectra with assigned fragments, spectrum cleanup, automated retrieval of compound information from Internet databases, and export to MassBank records.

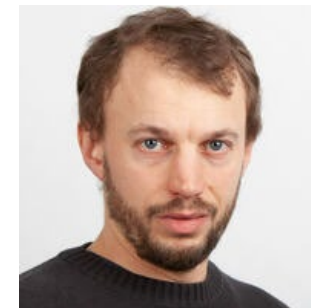
Author: Michael Stravs, Emma Schymanski, Steffen Neumann, Erik Mueller, with contributions from Tobias Schulze

Maintainer: RMassBank at Eawag <massbank at eawag.ch>

Citation (from within R, enter `citation("RMassBank")`):

Stravs MA, Schymanski EL, Singer H, Hollender J (2013). "Automatic Recalibration and Processing of Tandem Mass Spectra using Formula Annotation." *Journal of Mass Spectrometry*, **48**(1), 188.

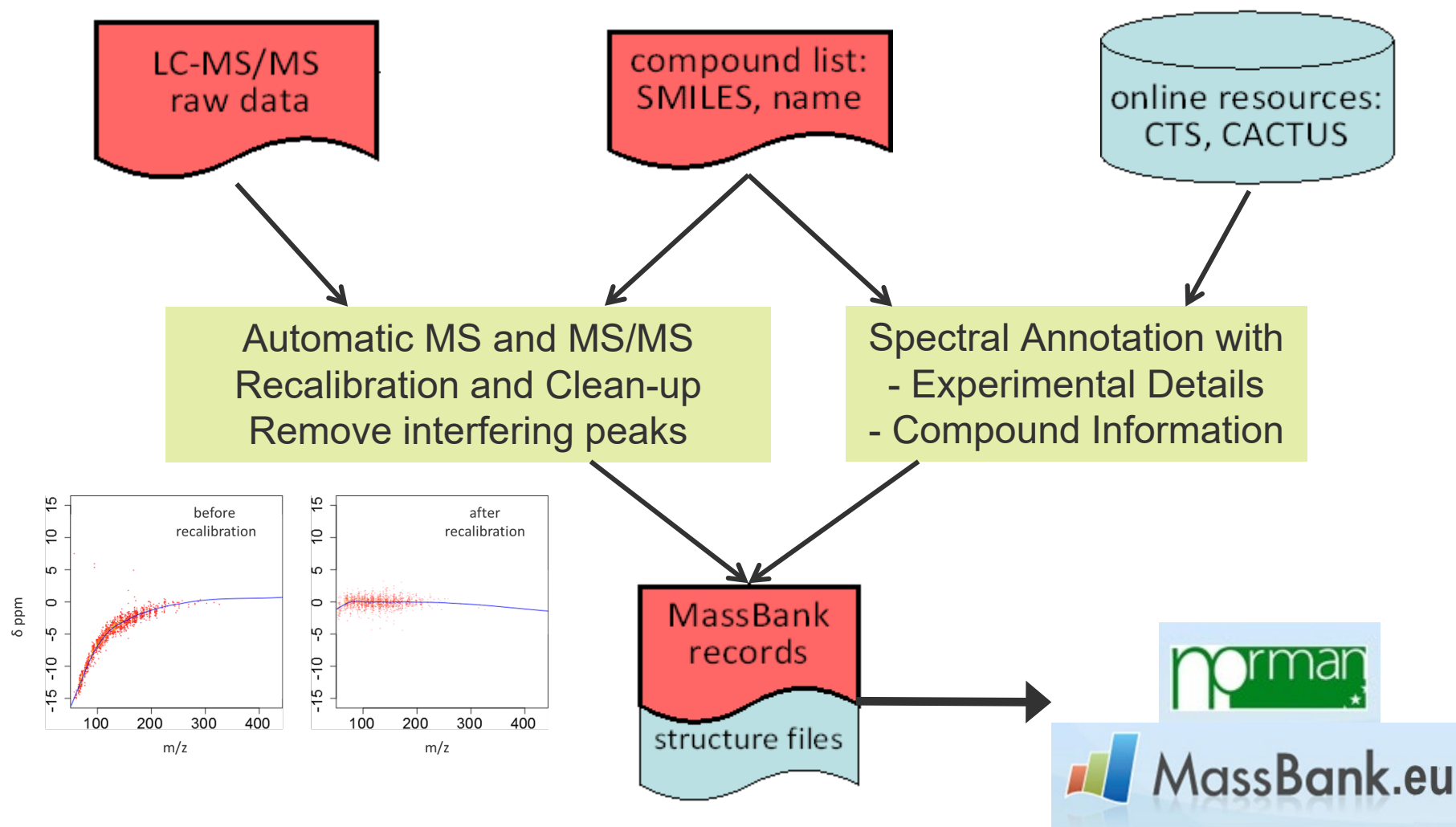
RMassBank



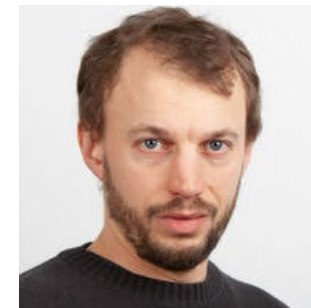
BioC reviewer:



RMassBank: Processing MS/MS in R for MassBank



RMassBank



(a slide from
the archives!)

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

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

MassBank: Cross Resource Integration (managed in R)

https://gitlab.lcsb.uni.lu/eci/pubchem/-/tree/master/massbank_eu

Data repo: <https://doi.org/10.5281/zenodo.5139996>

Environmental Science Processes & Impacts



PAPER

[View Article Online](#)
[View Journal](#)



Adding open spectral data to MassBank and PubChem using open source tools to support non-targeted exposomics of mixtures†

Anjana Elapavalore, ^{ID} *^a Todor Kondić, ^{ID} ^a Randolph R. Singh, ^{ID} ^{ab}
Benjamin A. Shoemaker, ^{ID} ^c Paul A. Thiessen, ^{ID} ^c Jian Zhang, ^{ID} ^c Evan E. Bolton ^{ID} ^c
and Emma L. Schymanski ^{ID} *^a

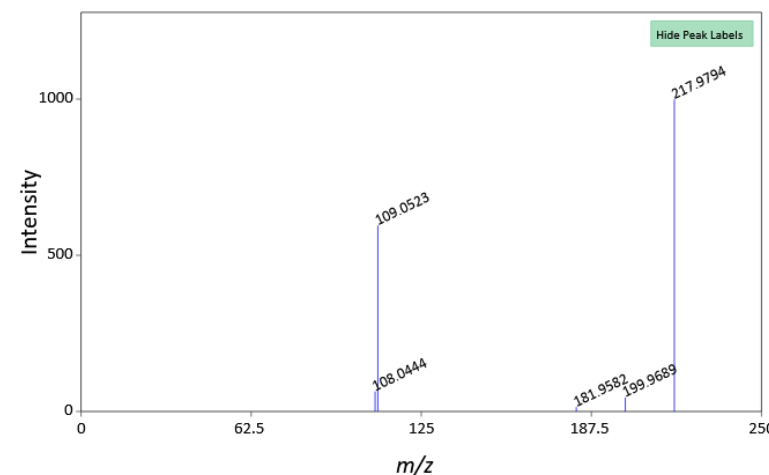
The term “exposome” is defined as a comprehensive study of life-course environmental exposures and the associated biological responses. Humans are exposed to many different chemicals, which can pose a major threat to the well-being of humanity. Targeted or non-targeted mass spectrometry techniques are widely used to identify and characterize various environmental stressors when linking exposures to human health. However, identification remains challenging due to the huge chemical space applicable to exposomics, combined with the lack of sufficient relevant entries in spectral libraries. Addressing these challenges requires cheminformatics tools and database resources to share curated open spectral data on chemicals to improve the identification of chemicals in exposomics studies. This article describes efforts

Arsanilic acid (Compound)

Top 5 Peaks

217.9794	999
109.0523	595
108.0444	64
199.9689	45
181.9582	14

[splash10-066r-0690000000-79ae659f03bdc3340757](#)



CC BY

Elapavalore, A.; Kondić, T.; et al., Adding Open Spectral Data to MassBank and

<https://pubchem.ncbi.nlm.nih.gov/compound/7389#section=LC-MS>



Gathering Expert Knowledge: NORMAN-SLE

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



NORMAN Database System

NORMAN organises the development and maintenance of various



SEARCH All Databases

Searching for individual substance or group(s) of substances

Note: Click on a link below to go to an individual database home



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



Antibiotic Resistance Bacteria/Genes

A database of ARBs/ARGs in environmental matrices



NORMAN Suspect List Exchange

The NORMAN Suspect List Exchange (NORMAN-SLE) was established to address the need for a centralised repository of suspect lists for screening and prioritisation. The NORMAN-SLE documents all individual collections of suspect lists (see Source column in SusDat). NORMAN-SLE versions are updated regularly. Comments and contributions are welcome - please email us at sle@norman-network.com. Please refer to our [documentation](#) pages for: [citation instructions](#)

No.	Abbreviation	Description	Link
S0	SUSDAT	Merged NORMAN Suspect List: SusDat	Interim Suspect List: Correlation

Mohammed Taha *et al.* (2022) DOI: [10.1186/s12302-022-00680-6](https://doi.org/10.1186/s12302-022-00680-6)

RESEARCH

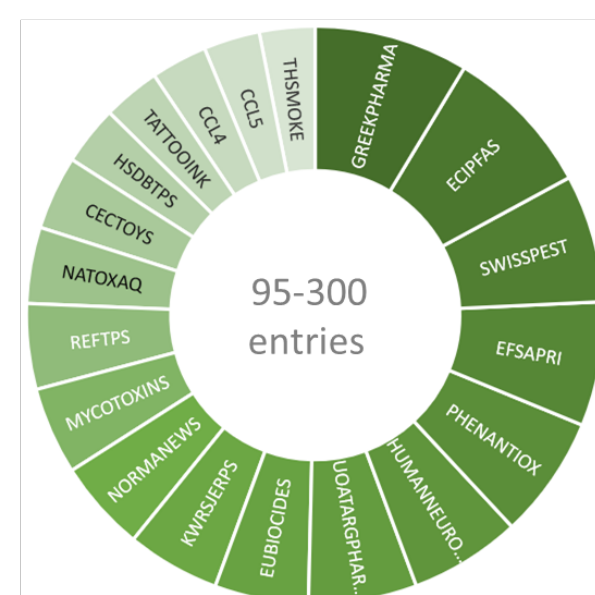
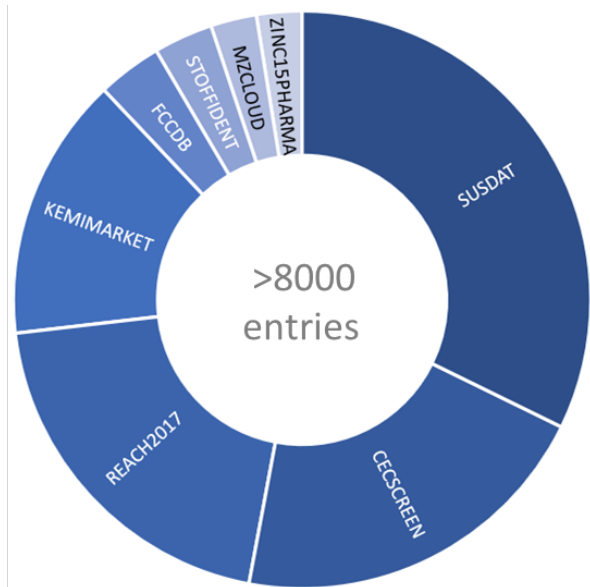
Open Access



The NORMAN Suspect List Exchange (NORMAN-SLE): facilitating European and worldwide collaboration on suspect screening in high resolution mass spectrometry

Hiba Mohammed Taha¹, Reza Aalizadeh², Nikiforos Alygizakis^{3,2}, Jean-Philippe Antignac⁴, Hans Peter H. Arp^{5,6}, Richard Bade⁷, Nancy Baker⁸, Lidia Belova⁹, Lubertus Bijlsma¹⁰, Evan E. Bolton¹¹, Werner Brack^{12,13}, Alberto Celma^{10,14}, Wen-Ling Chen¹⁵, Tiejun Cheng¹¹, Parviel Chirsir¹, Ľuboš Čirka^{16,3}, Lisa A. D'Agostino¹⁷, Yannick Djoumbou Feunang¹⁸, Valeria Dulio¹⁹, Stellan Fischer²⁰, Pablo Gago-Ferrero²¹, Aikaterini Galani², Birgit Geueke²², Natalia Głowacka³, Juliane Glüge²³, Ksenia Groh²⁴, Sylvia Grosse²⁵, Peter Haglund²⁶, Pertti J. Hakkinen¹¹, Sarah E. Hale⁵, Felix Hernandez¹⁰, Elisabeth M.-L. Janssen²⁴, Tim Jonkers²⁷, Karin Kiefer²⁴, Michal Kirchner²⁸, Jan Koschorreck²⁹, Martin Krauss¹², Jessy Krier¹, Marja H. Lamoree²⁷, Marion Letzel³⁰, Thomas Letzel³¹, Qingliang Li¹¹, James Little³², Yanna Liu³³, David M. Lunderberg^{34,35}, Jonathan W. Martin¹⁷, Andrew D. McEachran³⁶, John A. McLean³⁷, Christiane Meier²⁹, Jeroen Meijer³⁸, Frank Menger¹⁴, Carla Merino^{39,40}, Jane Muncke²², Matthias Muschket¹², Michael Neumann²⁹, Vanessa Neveu⁴¹, Kelsey Ng^{34,2}, Herbert Oberacher⁴³, Jake O'Brien⁷, Peter Oswald³, Martina Oswaldova³, Jaqueline A. Picache³⁷, Cristina Postigo^{44,14}, Noelia Ramirez^{45,39}, Thorsten Reemtsma¹², Justin Renaud⁴⁶, Pawel Rostkowski⁴⁷, Heinz Rüdel⁴⁸, Reza M. Salek⁴¹, Saer Samanipour⁴⁹, Martin Scherlinger^{23,42}, Ivo Schliebner²⁹, Wolfgang Schulz⁵⁰, Tobias Schulze¹², Manfred Sengl³⁰, Benjamin A. Shoemaker¹¹, Kerry Sims⁵¹, Heinz Singer²⁴, Randolph R. Singh^{1,52}, Mark Sumarah⁴⁶, Paul A. Thiessen¹¹, Kevin V. Thomas⁷, Sonia Torres³⁹, Xenia Trier⁵³, Annemarie P. van Wezel⁵⁴, Roel C. H. Vermeulen³⁸, Jelle J. Vlaanderen³⁸, Peter C. von der Ohe²⁹, Zhanyun Wang⁵⁵, Antony J. Williams⁵⁶, Egon L. Willighagen⁵⁷, David S. Wishart⁵⁸, Jian Zhang¹¹, Nikolaos S. Thomaidis², Juliane Hollender^{23,24}, Jaroslav Slobodnik³ and Emma L. Schymanski¹

NORMAN-SLE: Now >100 lists!



Category	Number	Comment
Number of Lists	109	S00 to S108
Total Unique Compounds	115,717	From PubChem NORMAN-SLE Tree
Total Live Substances	118,487	From PubChem NORMAN-SLE Source Page
Total Live Annotations	22,336	From PubChem NORMAN-SLE Source Page
Largest List	109,631	S00 NORMAN-SusDat
Smallest List	16	S98 TIRECHEM (a work in progress)
Total Views	86,005	From Zenodo (see below)
Unique Downloads	76,175	From Zenodo (see below)
Citations	61	From Zenodo (see below)

<https://gitlab.lcsb.uni.lu/eci/NORMAN-SLE/-/tree/master/stats#norman-sle-summary>

Stats updated Jul. 2023

Suspect List(s)

Eawag_ID	Name	CAS	Protected	logDph7	SMILES	InChi	InChiKey	Molecular	ExactMass
249	N4-Acetyl	127-76-4	CAS_RN12	0.69	CC(=O)NC1=CC=CC=C1	CC(=O)NC1=CC=CC=C1	CC(=O)NC1=CC=CC=C1	127.07242	297.0242
236	4-Acetamin	83-15-8	CAS_RN83	0.15	CN1C(C=C(C)C)C(=O)NC1=CC=CC=C1	CN1C(C=C(C)C)C(=O)NC1=CC=CC=C1	CN1C(C=C(C)C)C(=O)NC1=CC=CC=C1	151.15164	245.1164
245	N4-Acetyl	24341-30-8	CAS_RN24	1	COC1=CC(C=C(C)C)C(=O)NC1=CC=CC=C1	COC1=CC(C=C(C)C)C(=O)NC1=CC=CC=C1	COC1=CC(C=C(C)C)C(=O)NC1=CC=CC=C1	352.0841	352.0841
247	N4-Acetyl	100-90-3	CAS_RN10	0.41	CC(=O)NC1=CC=CC=C1	CC(=O)NC1=CC=CC=C1	CC(=O)NC1=CC=CC=C1	127.07242	297.0242

BIOTID0001	acetamino	1983	RZVAJINKF	InChi=15/CYP1A2;CYP1A2	Oxidation
BIOTID0001	acetamino	1983	RZVAJINKF	InChi=15/CYP1A2;CYP1A2	Oxidation
BIOTID0001	acetamino	1983	RZVAJINKF	InChi=15/CYP1A2;CYP1A2	Oxidation
BIOTID0001	acetamino	1983	RZVAJINKF	InChi=15/CYP1A2;CYP1A2	Oxidation
BIOTID0001	acetamino	1983	RZVAJINKF	InChi=15/CYP1A2;CYP1A2	Oxidation
BIOTID0001	acetamino	1983	RZVAJINKF	InChi=15/CYP1A2;CYP1A2	Oxidation
BIOTID0001	acetamino	1983	RZVAJINKF	InChi=15/CYP1A2;CYP1A2	Oxidation
BIOTID0001	acetamino	1983	RZVAJINKF	InChi=15/CYP1A2;CYP1A2	Oxidation
BIOTID0001	acetamino	1983	RZVAJINKF	InChi=15/CYP1A2;CYP1A2	Oxidation
BIOTID0001	acetamino	1983	RZVAJINKF	InChi=15/CYP1A2;CYP1A2	Oxidation

CAS Number	Name	Category	Priority	Quality	Ref	Rationale	M	Rationale	T	Rationale	Comments
289	Diuron-de	BIOTID0001 (R)-haloth	very high	1	1	1	1	1	1	1	1
294	Diuron-de	BIOTID0001 (R)-haloth	very high	1	1	1	1	1	1	1	1
118	Ethofume	BIOTID0001 (R)-haloth	very high	1	1	1	1	1	1	1	1
4	Metamitr	BIOTID0001 (R)-3,4-Me	very high	1	1	1	1	1	1	1	1
668	Simazine	BIOTID0001 (R)-3,4-Me	very high	1	1	1	1	1	1	1	1
671	Terbutylal	BIOTID0001 (R)-3,4-Me	very high	1	1	1	1	1	1	1	1
672	Terbutylal	BIOTID0001 (R)-3,4-Me	very high	1	1	1	1	1	1	1	1
2668	Fipronil-si	BIOTID0001 (R)-3,4-Me	very high	1	1	1	1	1	1	1	1
2670	Atenolol	BIOTID0001 (R)-3,4-Me	very high	1	1	1	1	1	1	1	1
126	Methyl-2	BIOTID0001 (R)-3,4-Me	very high	1	1	1	1	1	1	1	1
296	N,N-Dime	BIOTID0001 (R)-3,4-Me	very high	1	1	1	1	1	1	1	1
3244	Betameth	BIOTID0001 (R)-3,4-Me	very high	1	1	1	1	1	1	1	1



Suspect List Exchange
(NORMAN-SLE)



DOI	10.5281/zenodo.6770176
12,835	11,095
views	downloads

Suspect List Exchange (NORMAN-SLE)

[illegible]

Ecotoxicology

Substance Database (NORMAN SusDat)

Antibiotic Resistance

Bacteria/Genes

MassBank Europe

Substance Factsheets

Prioritisation



Mohammed
Taha *et al.*
(2022) DOI:
[10.1186/s12302-022-00680-6](https://doi.org/10.1186/s12302-022-00680-6)



12,835
views

11,095
downloads

Ecotoxicology

SARS-CoV-2 in sewage

Passive Sampling

Indoor Environment

Bioassays Monitoring Data

Digital Sample Freezing Platform

Chemical Occurrence Data

Enavg_ID	Name	CAS	Protected(logpH ⁷)	SMILES	InChI	InChIKey	Molecular	Enact	Mass
249	N-Acetyl-127-6-74	CAS RN312	0.69	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	127.0624	297.0242	
236	4-Acetamin-83-15-8	CAS RN383	0.15	CN(C)Cc1ccc(C(=O)N)cc1	CN(C)Cc1ccc(C(=O)N)cc1	CN(C)Cc1ccc(C(=O)N)cc1	154.1664	245.1164	
245	N-Acetyl-24341-30-8-CAS RN24		1	CCOC(=O)C1=CC=C(C=C1)C(=O)N	CCOC(=O)C1=CC=C(C=C1)C(=O)N	CCOC(=O)C1=CC=C(C=C1)C(=O)N	164.1604	352.0841	
247	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
248	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
249	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
250	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
251	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
252	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
253	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
254	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
255	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
256	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
257	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
258	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
259	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
260	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
261	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
262	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
263	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
264	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
265	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
266	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
267	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
268	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
269	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
270	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
271	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
272	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	100.0964	320.0943	
273	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1				



12,835 views 11,095 downloads

Suspect List Exchange (NORMAN-SLE)

Ecotoxicology

SARS-CoV-2 in sewage

Passive Sampling

Indoor Environment

 Bioassays Monitoring Data

Digital Sample Freezing Platform

Chemical Occurrence Data

Chemical Lists

Search: NORMAN

Showing 74 of 330 Records

List Acronym	List Name	# Chemicals	Updated	List Description
ATHENSUS	Metoprolol acid 56392-14-1 DTXSID70881060 Identified by approved name			...compilers of suspects, predicted transformation products and surfactants from University of Athens, as described in Gago-Fernandez et al. (2015, DOI: 10.1002/env.3454). The original data is available on the NORMAN Suspect List Exchange . Instances are included here.
BIPHENOLS				...ents a collection of biphenols available at HILLI (Pawel Kordicowski) and from ILMU (Swedish Chemical Agency) in Swedish with English summary) hosted on Suspect List Exchange (https://www.norman-network.com/nds/SLE/). Dataset DOI: 10.26434/chemrxiv-2019-07-2779854
CSCCOMPEND				...cited with >3800 experimental collision cross section (CCS) values (drift tube MS) ... NORMAN Suspect List Exchange by Jackie Pischke and John McLean. Vandalism applied to DTXSIDs by CAS Registry Number by E. Schymanski (Luxembourg Center for Toxicology) further curation ongoing. Further details available here

Envgw_ID	Name	CAS	Protected(logpH ⁷)	SMILES	INCln1	INClnHcy	Molecular	EnvgwMass
249	N-Acetyl-127-6-4	CAS RN312	0.69	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	297.0242			
236	4-Acetamyl-83-15-8	CAS RN383	0.15	CN(C)C(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	254.1164			
245	N-Acetyl-24341-30-8-CAS RN24		1	CCOC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	352.0841			
247	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
250	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
251	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
252	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
253	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
254	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
255	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
256	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
257	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
258	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
259	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
260	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
261	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
262	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
263	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
264	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
265	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
266	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
267	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
268	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
269	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
270	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
271	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
272	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
273	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
274	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
275	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
276	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
277	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
278	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
279	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
280	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
281	N-Acetyl-100-90-3	CAS RN130	0.41	CC(=O)Nc1nc1S/C(=O)NKKXWVn1c1H11N3	320.0943			
282	N-Acetyl-100-90-3	CAS RN130						



Ecotoxicology

SARS-CoV-2 in sewage

Substance Database (NORMAN SusDat)

Suspect List Exchange (NORMAN-SLE)

CompTox Chemicals Dashboard

formations

What is PubChem? <https://pubchem.ncbi.nlm.nih.gov/>



Explore Chemistry

Quickly find chemical information from authoritative sources



Try covid-19 aspirin EGFR C9H8O4 57-27-2 C1=CC=C(C=C1)C=O InChI=1S/C3H6O/c1-3(2)4/h1-2H3



Use Entrez



Compounds



Substances



BioAssays



Draw Structure



Upload ID List



Browse Data



Periodic Table

116M Compounds

308M Substances

292M Bioactivities

36M Literature

934 Data Sources

[See More Statistics >](#)

[Explore Data Sources >](#)

NORMAN-SLE in PubChem



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



The NORMAN network enhances the exchange of information on emerging environmental substances, and encourages the validation and harmonisation of common measurement methods and monitoring tools so that the requirements of risk assessors and risk managers can be better met. The NORMAN Suspect List Exchange (NORMAN-SLE) is a central access point to find suspect lists relevant for various environmental monitoring questions, described in [DOI:10.1186/s12302-022-00680-6](https://doi.org/10.1186/s12302-022-00680-6)

Organization	NORMAN Network (c/o UniLu)
Category	Research and Development
URL	https://www.norman-network.com/nds/SLE/
License Note	Data: CC-BY 4.0; Code (hosted by ECI, LCSB): Artistic-2.0
License URL	https://creativecommons.org/licenses/by/4.0/
Contact Name	Emma Schymanski
Address	6 avenue du Swing, Belvaux, Luxembourg, 4367
Data Source ID	23819
Data in PubChem	118,487 Live Substances 22,317 Annotations 1 Classification
Last Updated	2023/09/06

▼ NORMAN Suspect List Exchange Classification ? 115,695

- ▶ S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) ? 3,936
- ▶ S25 | OECDPFAS | List of PFAS from the OECD ? 3,678
- ▶ S36 | UBAPMT | Potential Persistent, Mobile and Toxic (PMT) substances ? 254
- ▶ S47 | ECHAPLASTICS | A list from the Plastic Additives Initiative Mapping Exercise by ECHA ? 241
- ▶ S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ? 869
- ▶ S60 | SWISSPEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 ? 1,359
- ▶ 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 ? 740
- ▶ 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 ? 2,553
- ▶ S77 | FCCDB | Food Contact Chemicals Database v5.0 ? 5,989
- ▶ S79 | UACCSCEC | Collision Cross Section (CCS) Library from UAntwerp ? 148
- ▶ S80 | PFASGLUEGE | Overview of PFAS Uses ? 1,251

NORMAN-SLE in PubChem



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



The NORMAN network enhances the exchange of information on emerging environmental substances, and encourages the validation and harmonisation of common measurement methods and monitoring tools so that the requirements of risk assessors and risk managers can be better met. The NORMAN Suspect List Exchange (NORMAN-SLE) is a central access point to find suspect lists relevant for various environmental monitoring questions, described in [DOI:10.1186/s12302-022-00680-6](https://doi.org/10.1186/s12302-022-00680-6)

Organization	NORMAN Network (c/o UniLu)
Category	Research and Development
URL	https://www.norman-network.org/ds/SLE/
License Note	Data: CC-BY 4.0; Code (hosted by ECI, LC) Artistic-2.0
License URL	https://creativecommons.org/licenses/by/4.0/
Contact Name	Emma Schymanski
Address	6 avenue du Swing, Belvaux, Luxembourg, 4367
Data Source ID	23819
Data in PubChem	118,487 Live Substances 22,317 Annotations 1 Classification
Last Updated	2023/09/06

pubchem

Project ID: 1565

1,071 Commits

7 Branches

1 Tag

22.8 MiB Project Storage

A project for interactions with PubChem

Entire integration managed in

GitLab

Update MassBank_File_Locations.txt

Emma Schymanski authored 1 week ago

59e5996e

master pubchem

History Find file

Clone

README Artistic License 2.0

Name	Last commit	Last update
annotations	updated Transformations	1 month ago
deposition	Updated S74 in mapping files	1 month ago
massbank_eu	Update MassBank_File_Locations.txt	1 week ago

- ▼ NORMAN Suspect List Exchange Classification ? ↗ 115,248
- ▶ S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) ? 3,887
 - ▶ S25 | OECDPFAS | List of PFAS from the OECD ? 3,677
 - ▶ S36 | UBAPMT | Potential Persistent, Mobile and Toxic (PMT) substances
 - ▶ S47 | ECHAPLASTICS | A list from the Plastic Additives Initiative Mapping Exercise
 - ▶ S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ?
 - ▶ S60 | SWISSPEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 ?
 - ▶ 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
 - ▶ S69 | LUXPEST | Pesticide Screening List for Luxembourg ? 386
 - ▶ S72 | NTUPHTW | Pharmaceutically Active Substances from National Taiwan University
 - ▶ S75 | CyanoMetDB | Comprehensive database of secondary metabolites from cyanobacteria
 - ▶ S79 | UACCSCEC | Collision Cross Section (CCS) Library from UAntwerp ? 148
 - ▶ S80 | PFASGLUECE | Overview of PFAS Uses ? 1,250

PubChem Phloroglucinol (Compound)

9.1.1 Use Classification

Cosmetics -> Hair dyeing

S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) | DOI:10.5281/zenodo.2624118

▶ NORMAN Suspect List Exchange

PubChem Thiabendazole (Compound)

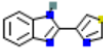
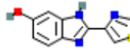
10.10 Transformations

10 items

Download

Search

SORT BY Predecessor Name - A to Z

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme	Evidence
	thiabendazole		5-hydroxythiabendazole	Aromatic hydroxylation of fused benzene ring / Human Phase I	CYP1A2	10.1186

PubChem 7-Desmethylnicrocystin RR (Compound)

7 Taxonomy

2 items

Download

Structure	Taxonomy	Evidence IDs	Data Source
	Microcystis aeruginosa	PMID:1440646 DOI:10.1016/0041-0101(92)90054-9	NORMAN Suspect List Exchange

PubChem Tri-O-cresyl phosphate (Compound)

3.2.18 Collision Cross Section

192.43 Å² [M+Na]⁺

182.39 Å² [M+H]⁺

263.75 Å² [2M+Na]⁺

S79 | UACCSCEC | Collision Cross Section (CCS) Library from UAntwerp | DOI:10.5281/zenodo.4704648

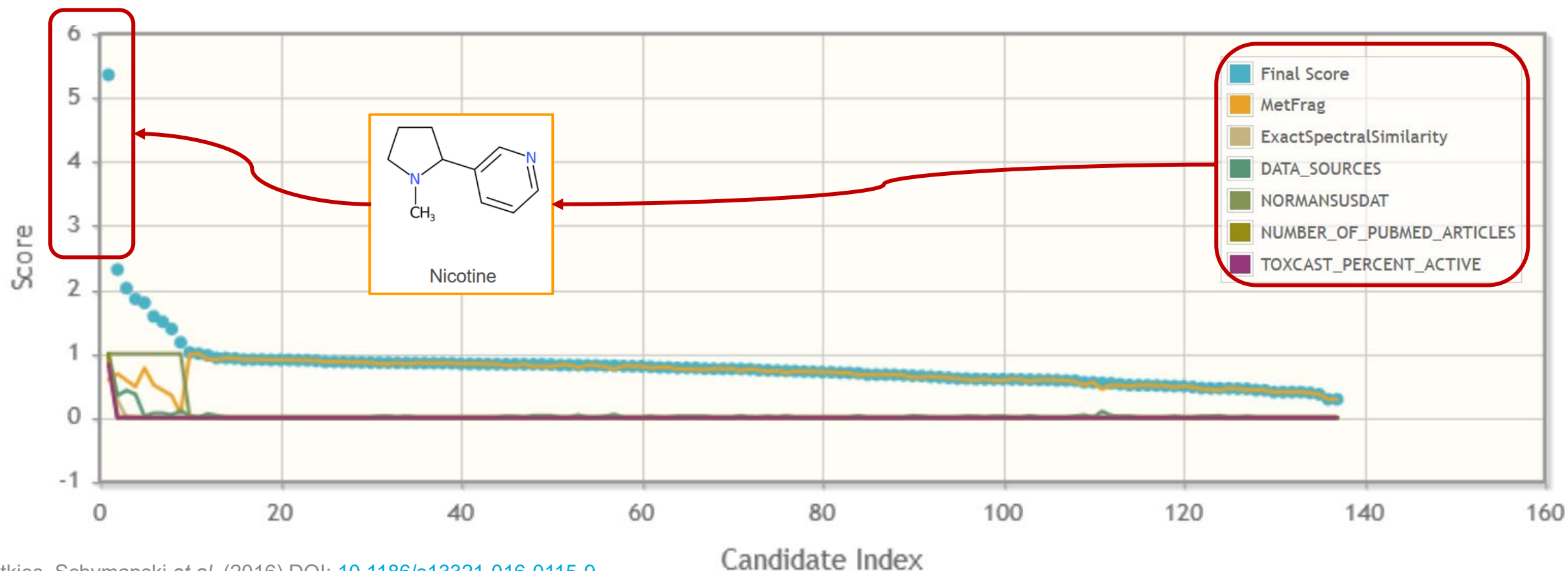
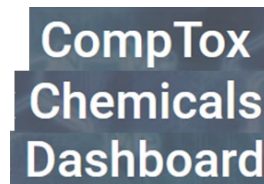
▶ NORMAN Suspect List Exchange

Connecting Knowledge for Chemical Identification: MetFrag



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

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



MetFrag: *In Silico* Fragmentation to Select Candidates





MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

Database Settings

Database: PubChem ☐ Include references: ☐

Neutral Mass: 162.11576 Search ppm: 5

Formula: C10H14N2

Identifiers:

Retrieve Candidates 3704 Candidates Download Candidates

Candidate Filter & Score Settings

Fragmentation Settings & Processing

Mzppm: 5

Mzabs: 0.001

Mode: [M+H]⁺

Tree depth: 2

Group candidates ☒

MS/MS Peak list

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

Show Spectrum

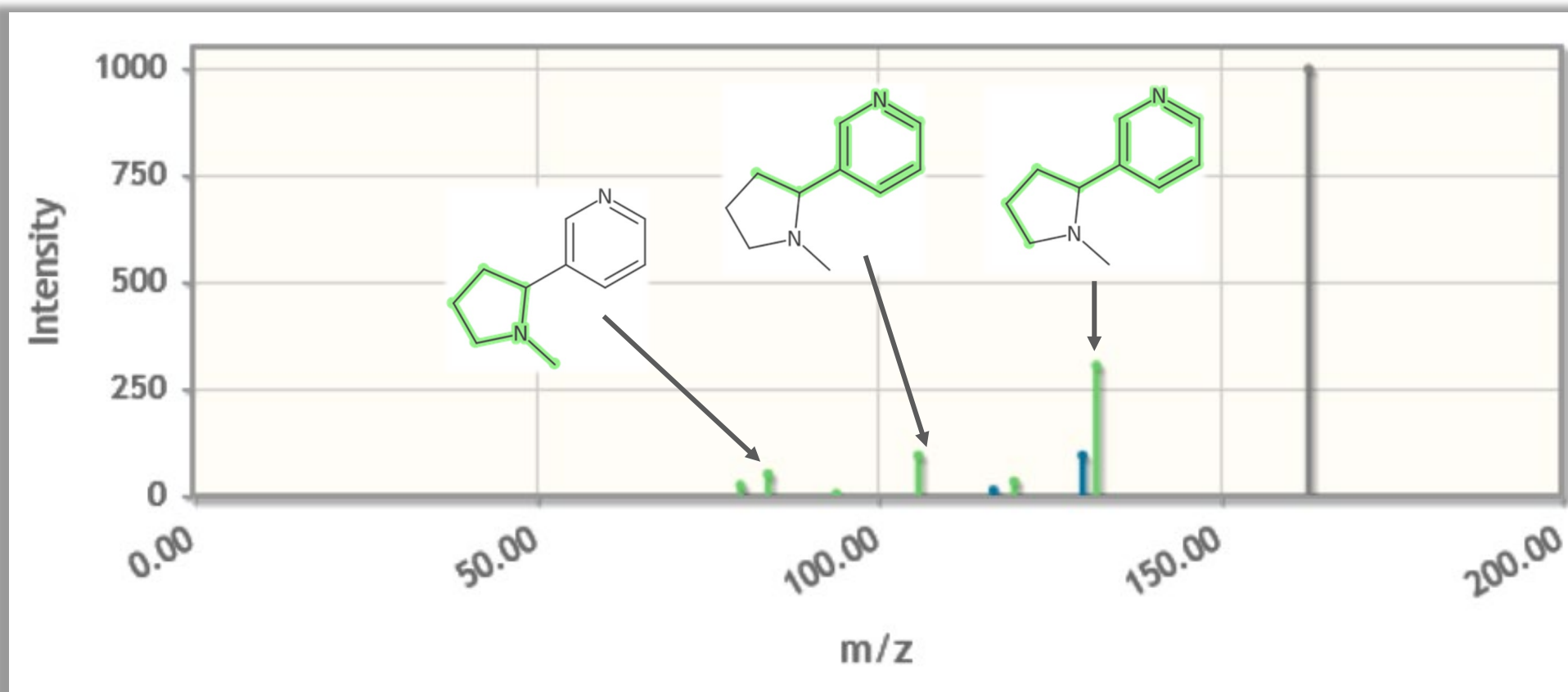


MetFrag: *In Silico* Fragmentation to Select Candidates



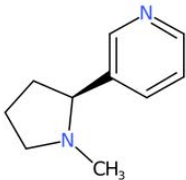
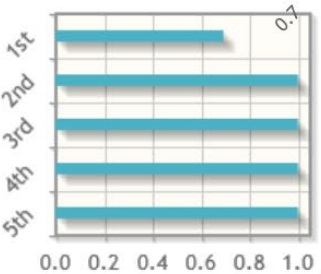
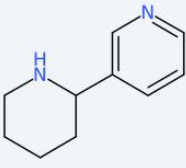
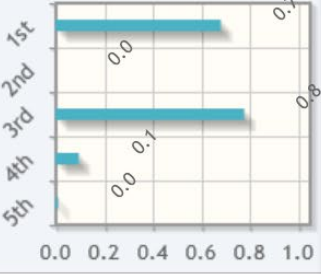
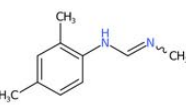
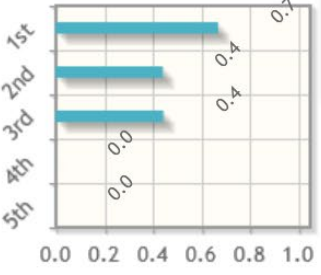
MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra



MetFrag: ...Plus Metadata for Context

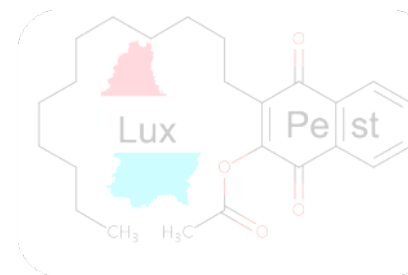
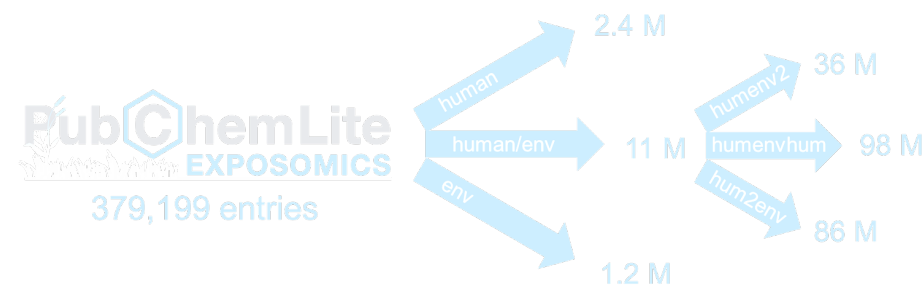
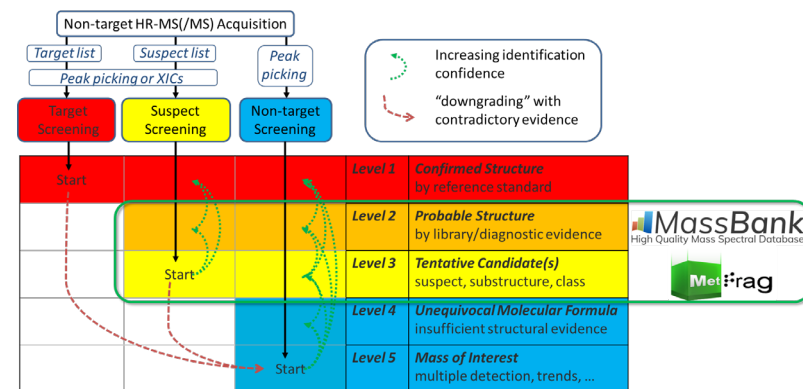


 MetFrag In silico fragmentation for computer assisted identification of metabolite mass spectra							
#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1	 3-[(2S)-1-methylpyrrolidin-2-yl]pyridine	89594 InChIKeyBlock1 = SNICXCGAKADSCV	162.1157	C ₁₀ H ₁₄ N ₂		4.692	Peaks: 6 / 8 Fragments Scores Download
2	 3-piperidin-2-ylpyridine	2181 InChIKeyBlock1 = MTXSIJUGVMTTMU	162.1157	C ₁₀ H ₁₄ N ₂			
3	 N-(2,4-dimethylphenyl)-N'-methylmethanimidamide	36326 InChIKeyBlock1 = JIIOLEGNERQDIP	162.1157	C ₁₀ H ₁₄ N ₂		1.5586	Scores Download

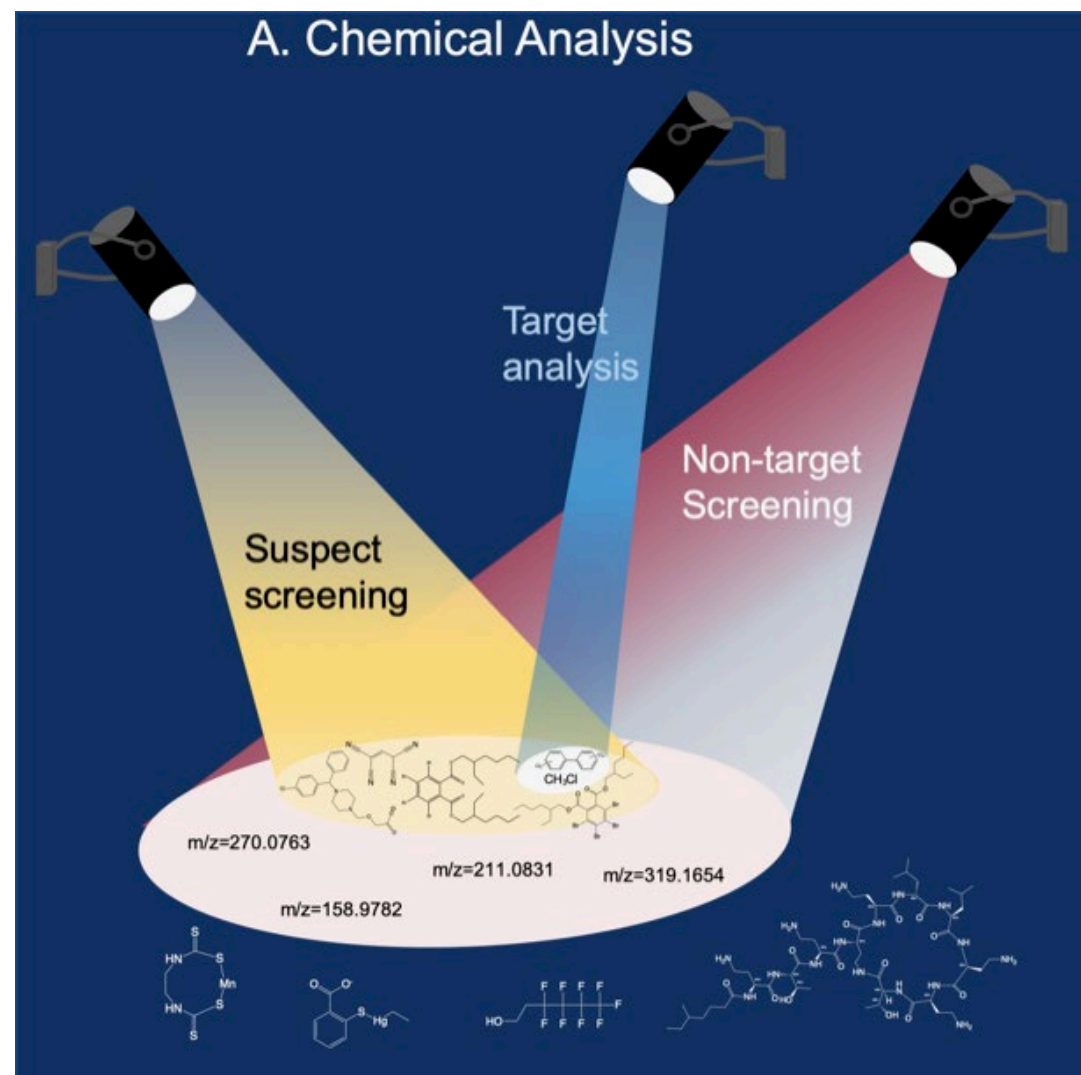
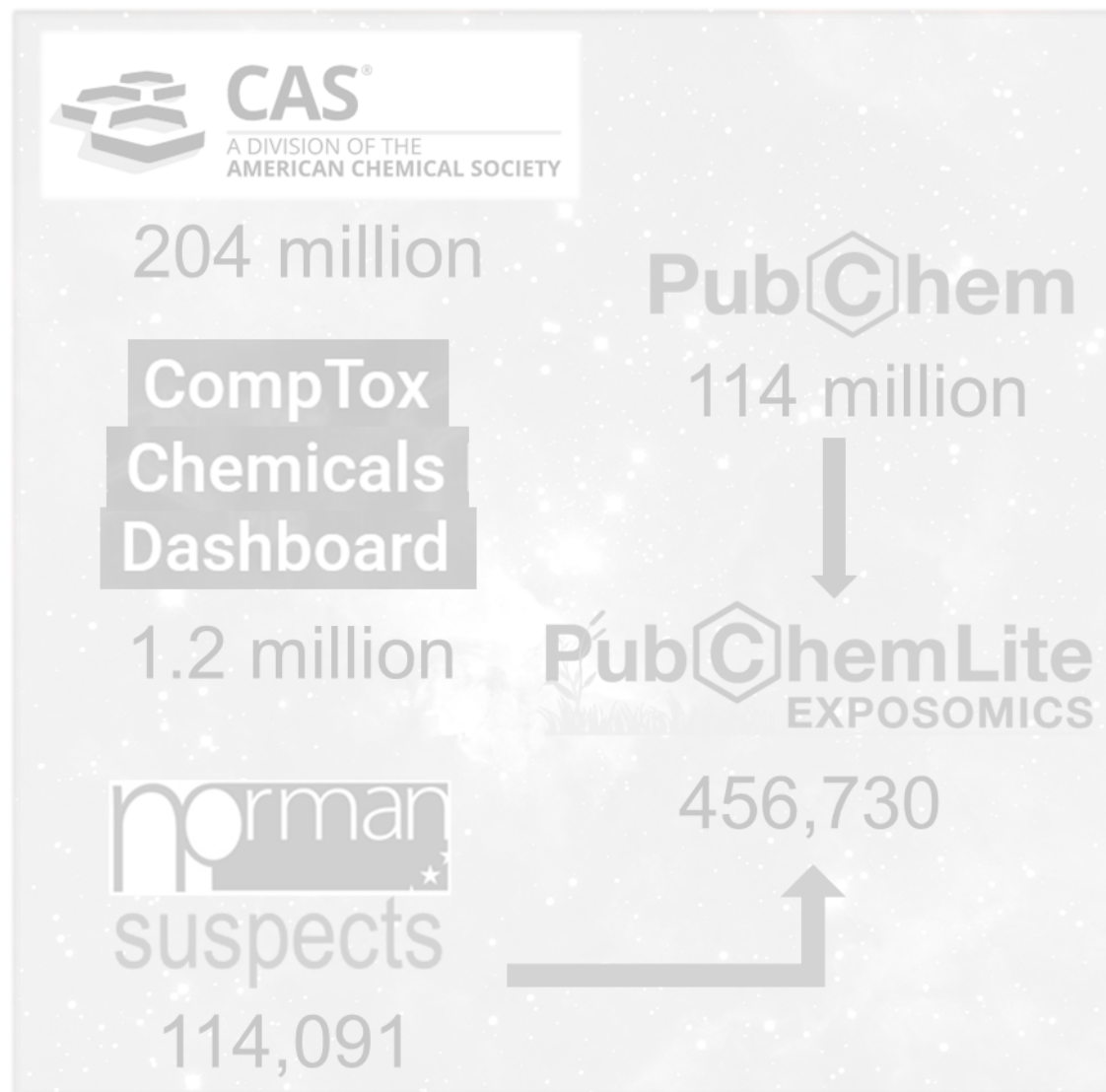
Scores View			
Candidate Name:		3-[(2S)-1-methylpyrrolidin-2-yl]pyridine	
Candidate Identifier:		89594 47445	
	Name	Normalized Value	Raw Value
1	MetFrag	0.692	246.1725
2	ExactSpectralSimilarity	1.0	1.0
3	AnnoTypeCount	1.0	9.0
4	Patent_Count	1.0	102416.0
5	PubMed_Count	1.0	38574.0

Outline

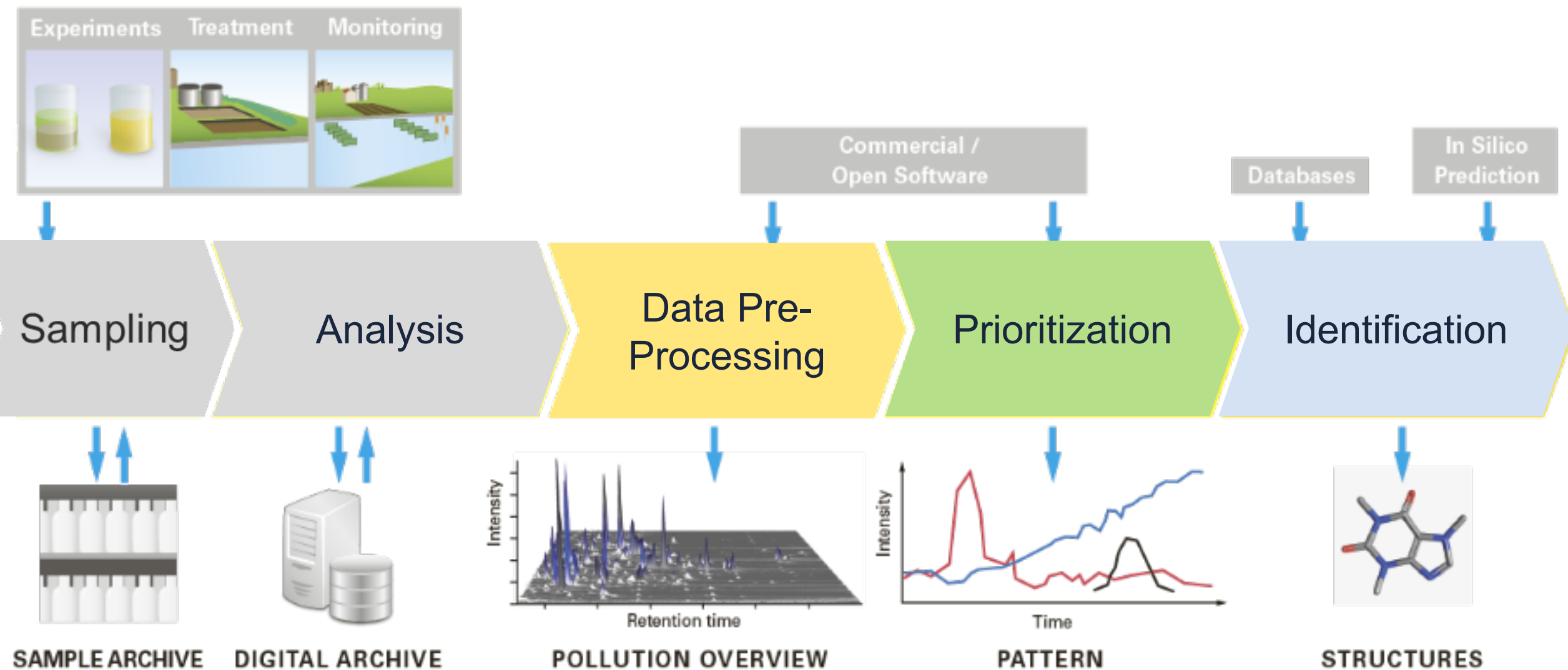
- Identifying Chemicals with NT-HRMS
 - Open Resources for Identification
 - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases & Workflows
 - Defining Relevant Chemical Space
 - Transformations: Known & Predicted
- No package too small?
- Applications in Environmental Studies



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



Non-target High Resolution Mass Spectrometry (NT-HRMS)



Open Source Workflows for NT-HRMS: patRoon



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

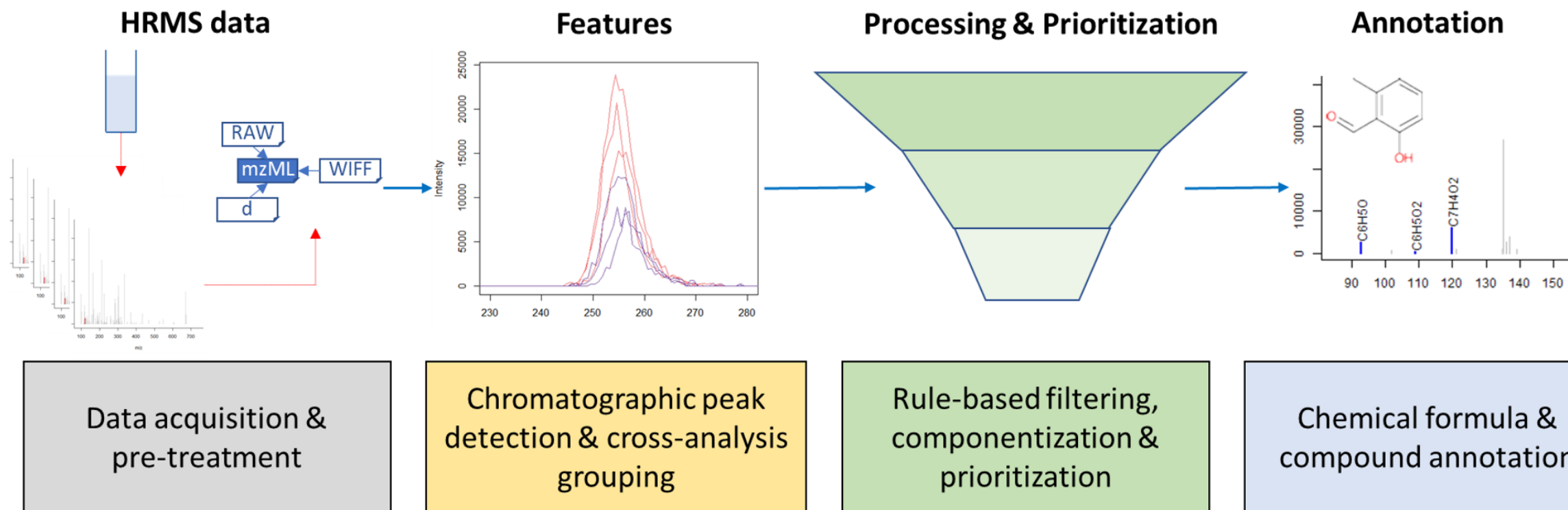
patRoon: open source software platform for environmental mass spectrometry based non-target screening

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

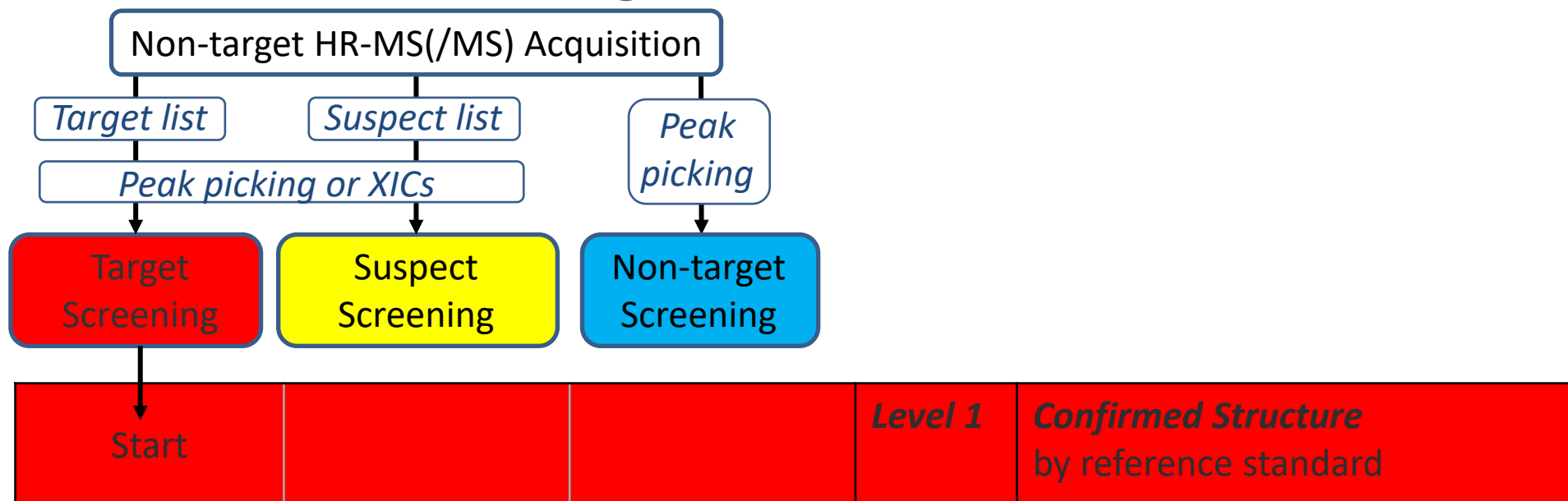
[Rick Helmus](#)✉, [Thomas L. ter Laak](#), [Annemarie P. van Wezel](#), [Pim de Voogt](#) & [Emma L. Schymanski](#)

patRoon 2.0: Improved non-target analysis workflows including automated transformation product screening

[Rick Helmus](#)^{1†}, [Bas van de Velde](#)¹²³, [Andrea M. Brunner](#)²⁴, [Thomas L. ter Laak](#)¹², [Annemarie P. van Wezel](#)¹, and [Emma L. Schymanski](#)⁵

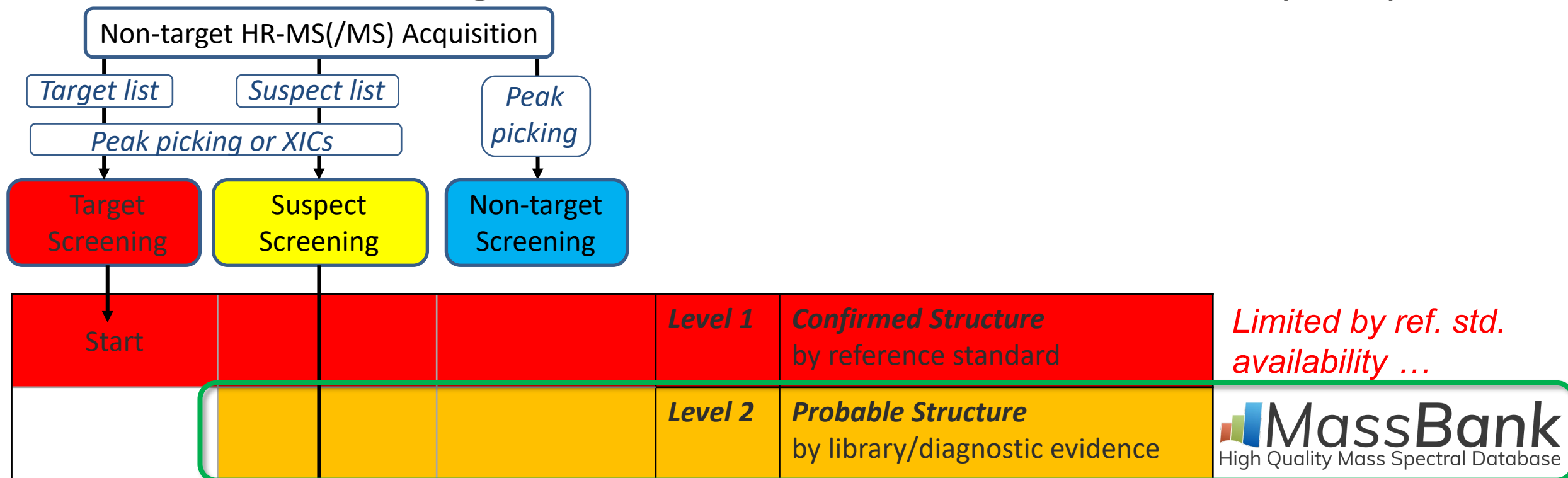


Identification Strategies and Confidence in NT-HRMS(/MS)



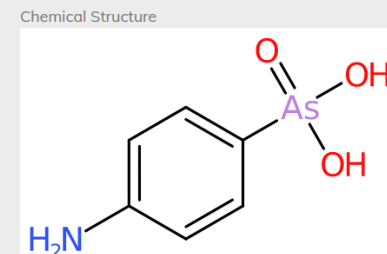
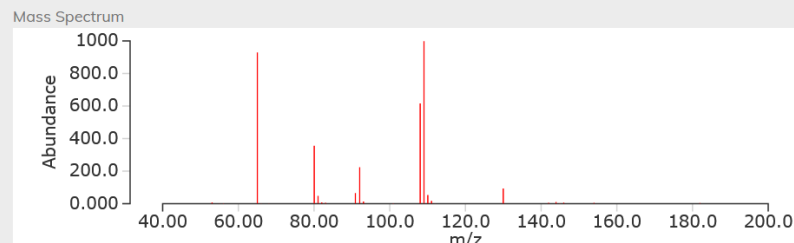
*Limited by ref. std.
availability ...*

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



MassBank Record: LU040605

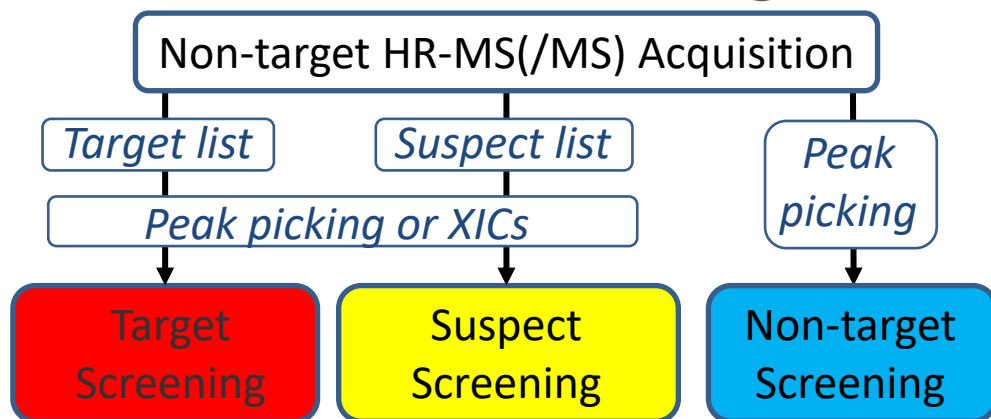
(4-Aminophenyl)arsonic acid; LC-ESI-QFT; MS2; CE: 75; R=17500; [M+H]⁺



eawag
aquatic research



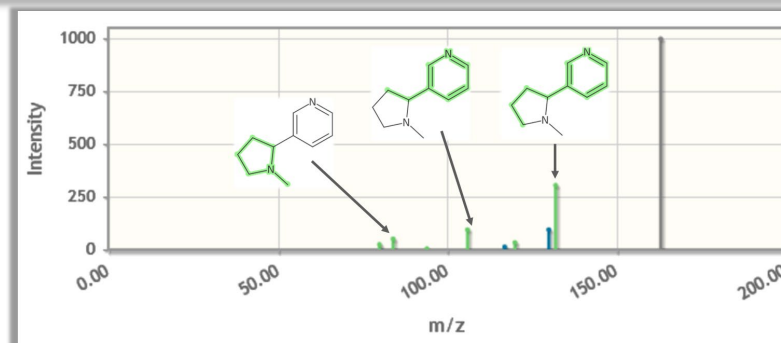
Identification Strategies and Confidence in NT-HRMS(/MS)



Start			Level 1	Confirmed Structure by reference standard
			Level 2	Probable Structure by library/diagnostic evidence
	Start		Level 3	Tentative Candidate(s) suspect, substructure, class

*Limited by ref. std.
availability ...*

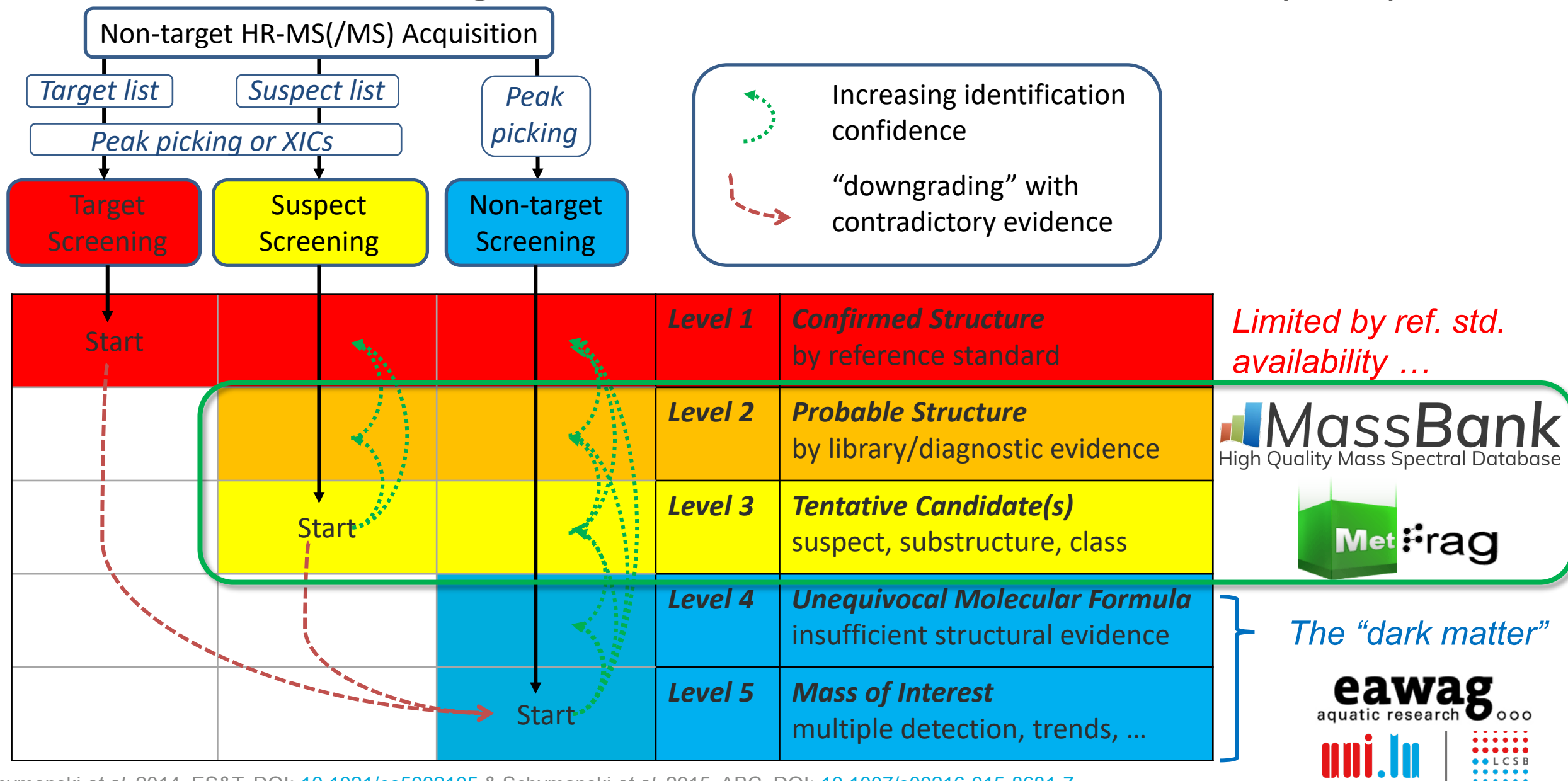
MassBank
High Quality Mass Spectral Database



eawag
aquatic research



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



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



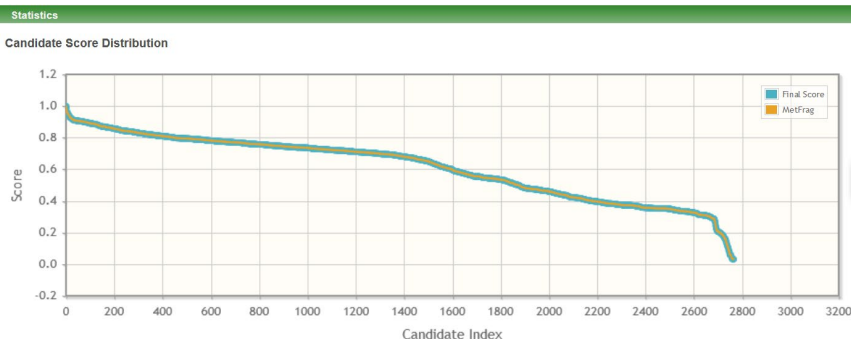
m/z $[M-H]^-$
213.9637
 ± 5 ppm

5 ppm
0.001 Da

PubChem



Ranked Candidates



MS/MS

134.0054	339689
150.0001	77271
213.9607	632466



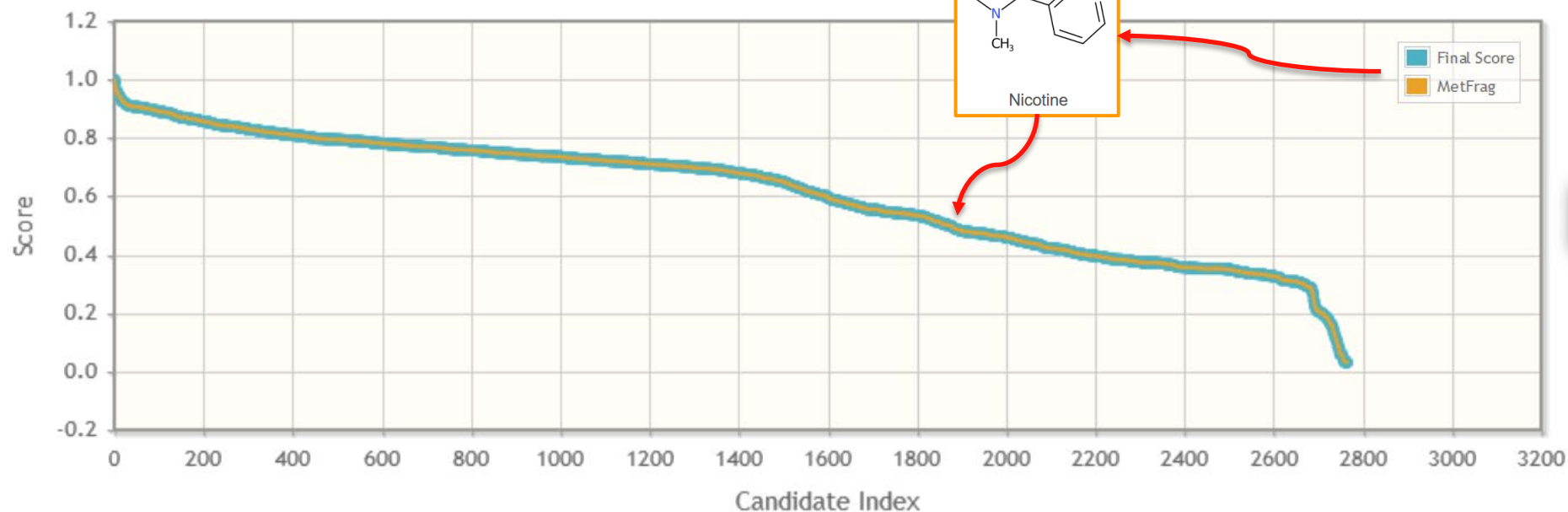
Key Challenge: MS and MS/MS alone is not enough!



PubChem

Statistics

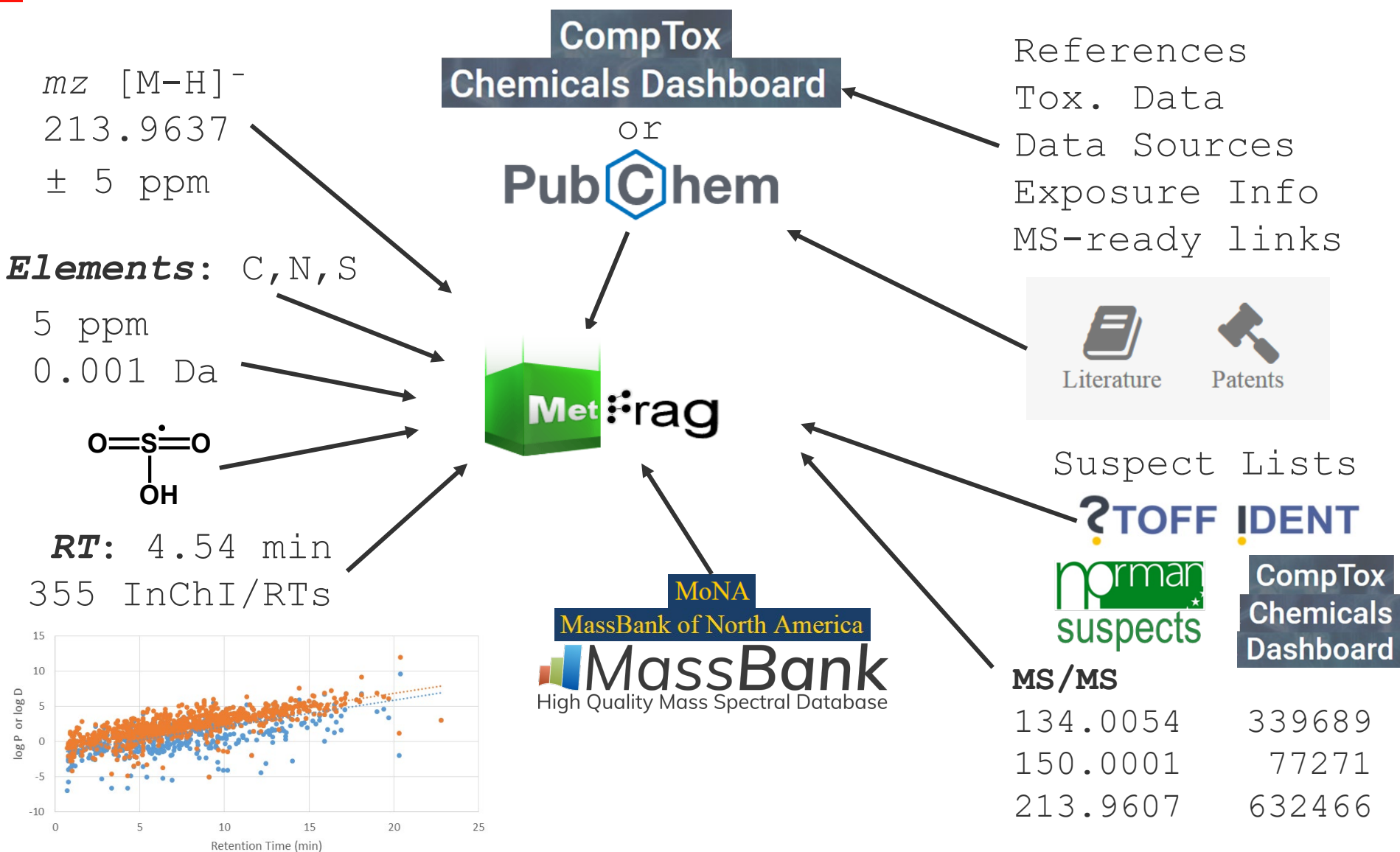
Candidate Score Distribution



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



Adding in more information: MetFrag Relunched (2016)



eawag
aquatic research

solutions

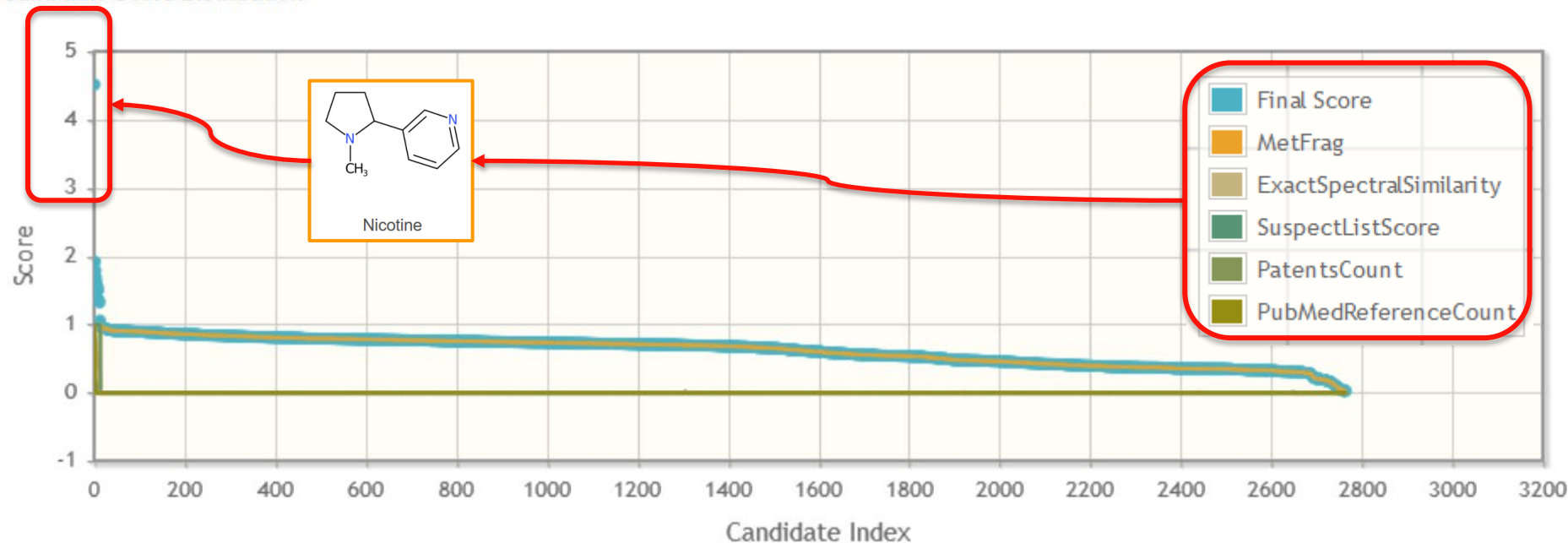


MetFragRL + PubChem + MS/MS + Metadata



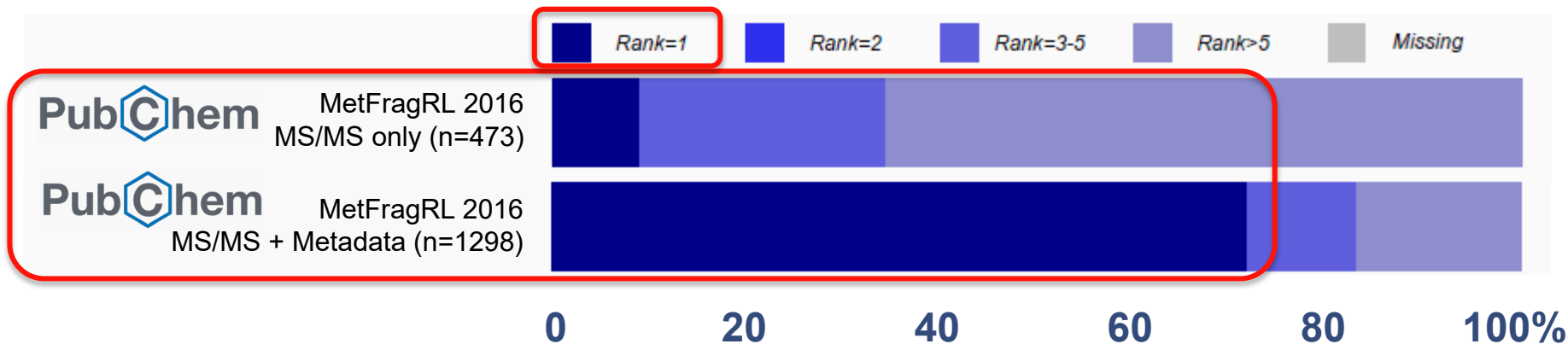
Statistics

Candidate Score Distribution



MetFragRL + PubChem + MS/MS + Metadata

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



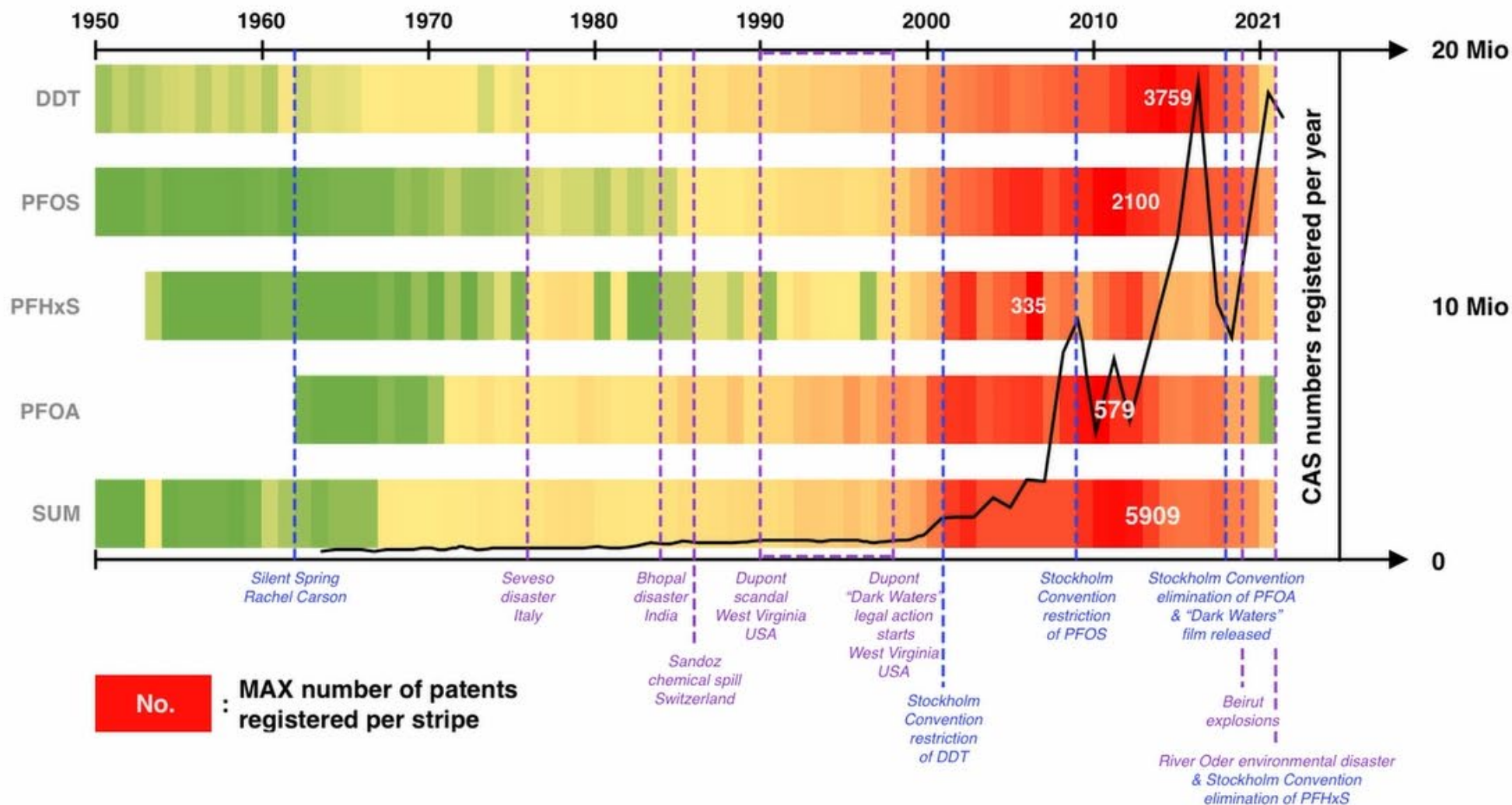
*BUT ...databases grow ... identification performance drops
... and run times rise ... (a lot!)*



JPML

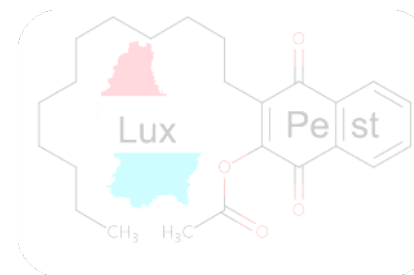
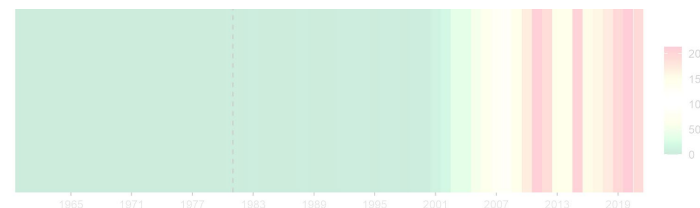
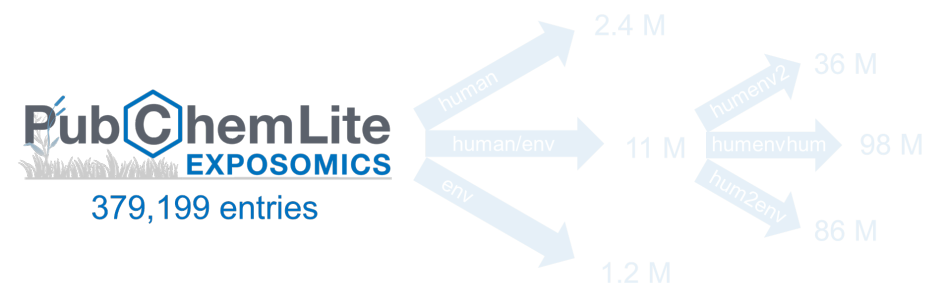
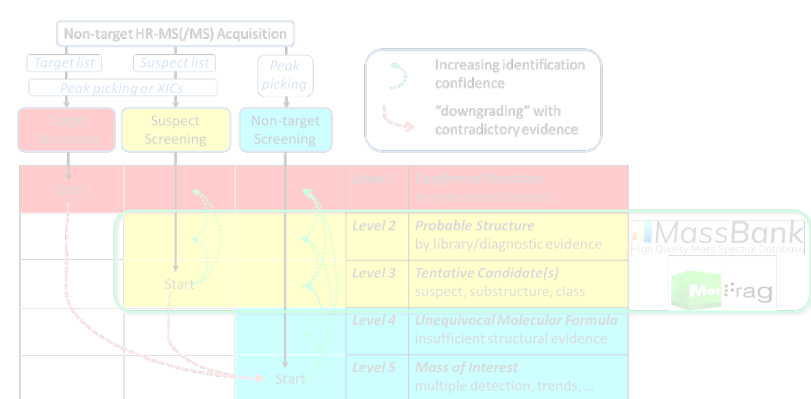
Our Chemical Past, Present, and Future

low / medium / high number of patents registered per year (WIPO)

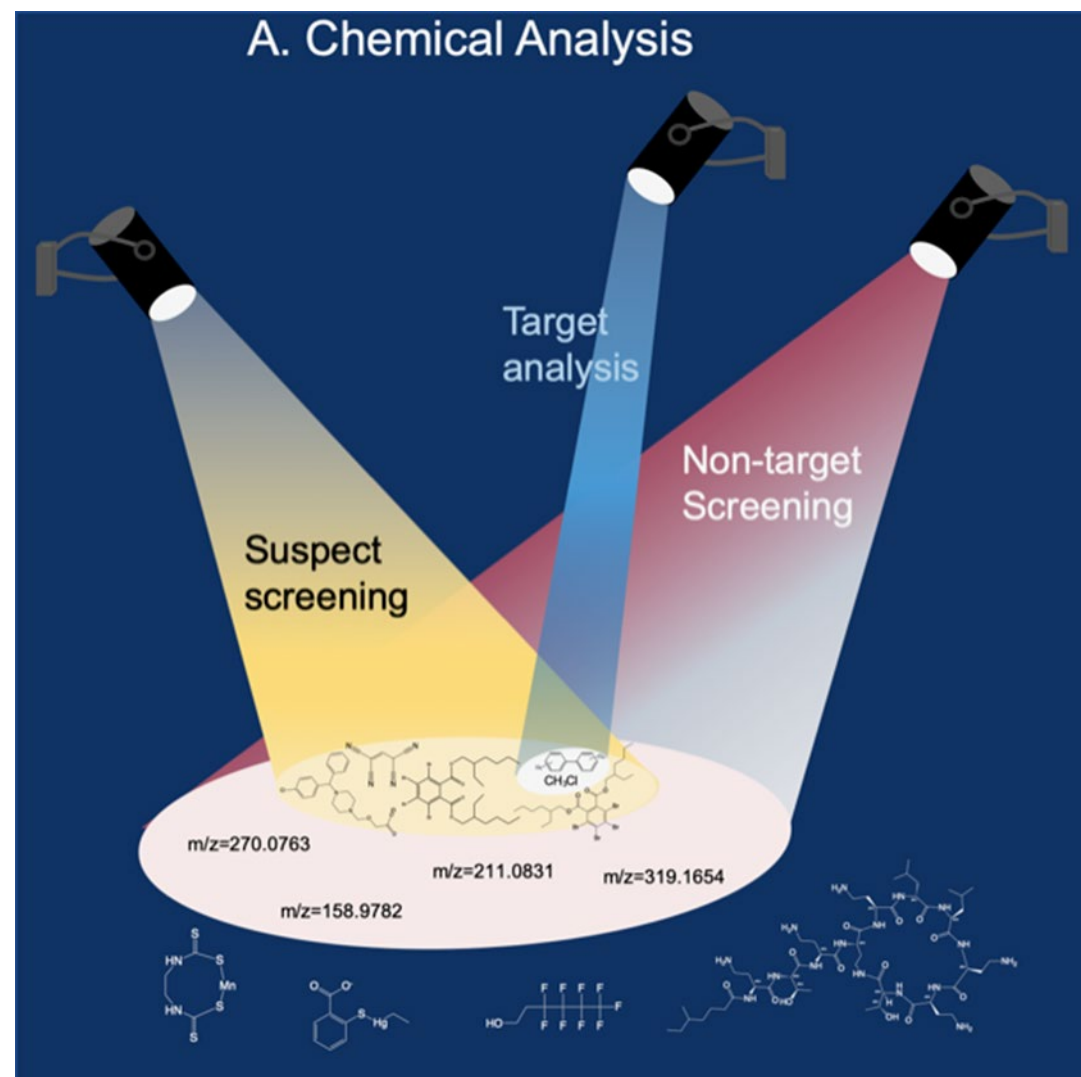
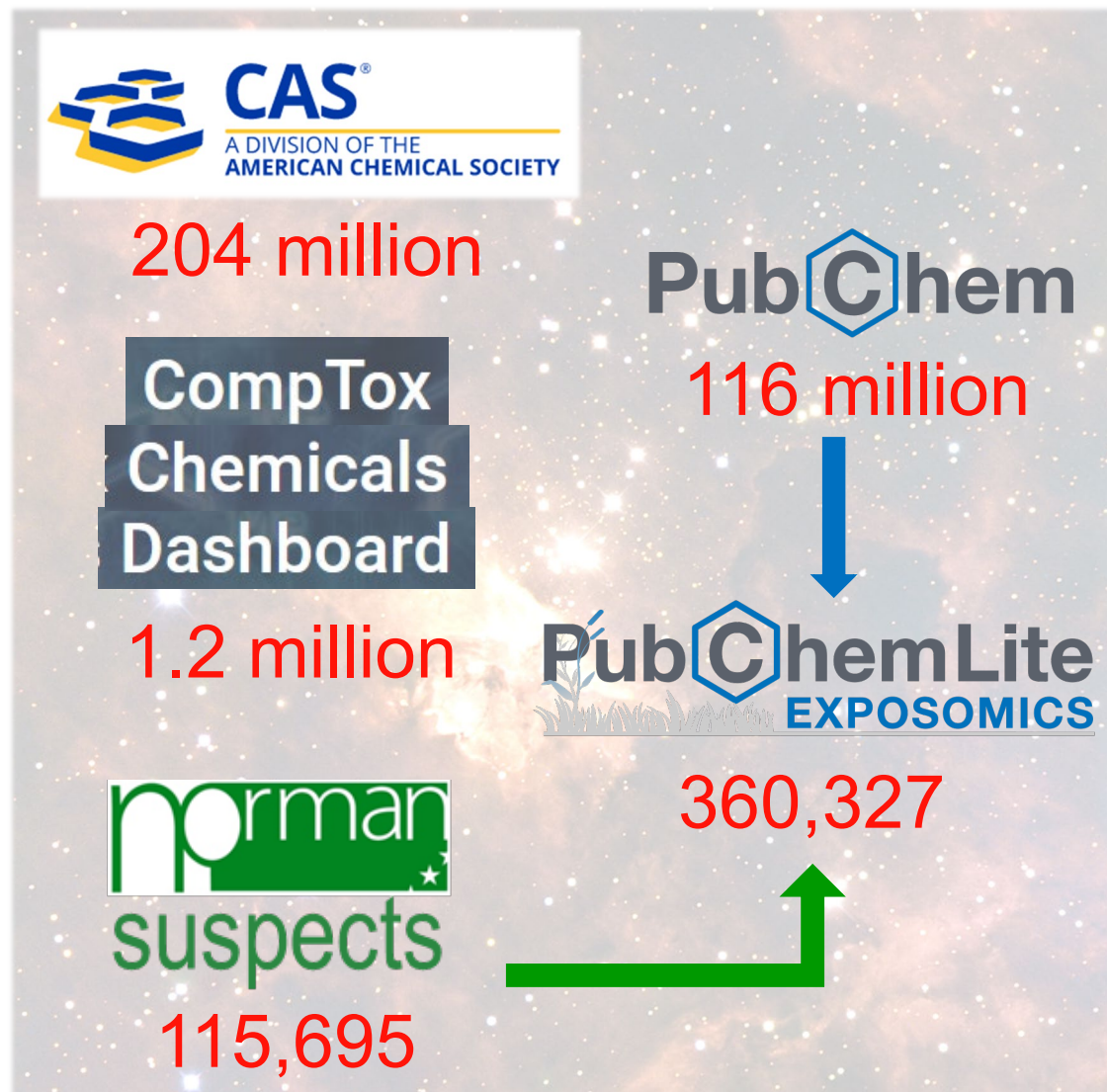


Outline

- Identifying Chemicals with NT-HRMS
 - Open Resources for Identification
 - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases & Workflows
 - Defining Relevant Chemical Space
 - Transformations: Known & Predicted
- No package too small?
- Applications in Environmental Studies



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



Numbers updated 9 Sept. 2023

Source: ESO/IDA/Danish 1.5 m <http://www.eso.org/public/images/potw1015a/> Mod. from Escher *et al* (2020). Science. DOI: [10.1126/science.aay6636](https://doi.org/10.1126/science.aay6636)

Grand Challenge: HR-MS “Chemical Space” is too Big!



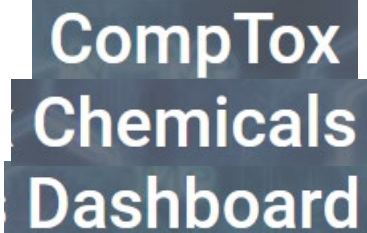
204 million



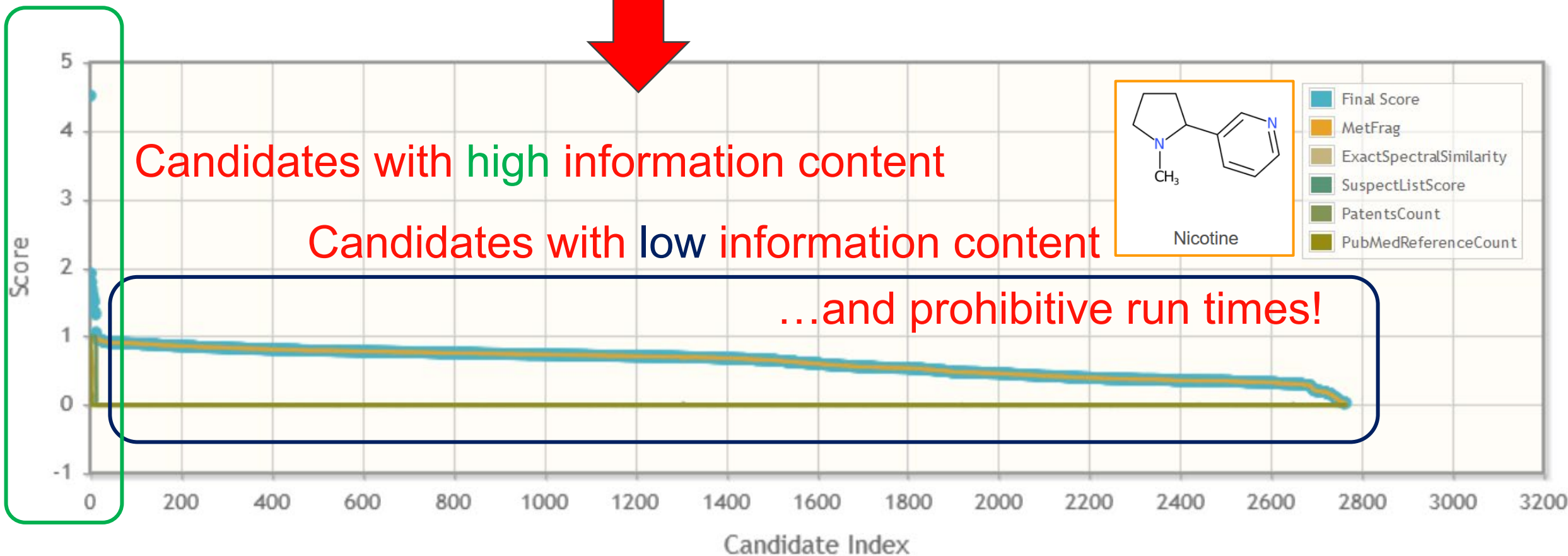
116 million



128 million



1.2 million



Creating Informative Subsets of PubChem



PubChemLite
EXPOSOMICS

PubChem Compound TOC ? 67,343,260

- ▶ Agrochemical Information ? 3,135
- ▶ Associated Disorders and Diseases ? 30,136
- ▶ Biologic Description ? 2,511,444
- ▶ Biological Test Results ? 4,567,078
- ▶ Chemical and Physical Properties ? 268,878
- ▶ Classification ? 22,965,005
- ▶ Drug and Medication Information ? 21,177
- ▶ Food Additives and Ingredients ? 7,627
- ▶ Identification ? 4,808
- ▶ Information Sources ? 47,725,078
- ▶ Interactions and Pathways ? 207,277
- ▶ Literature ? 4,076,955
- ▶ Names and Identifiers ? 7,021,765
- ▶ Patents ? 39,104,437
- ▶ Pharmacology and Biochemistry ? 114,060
- ▶ Related Records ? 13,282,616
- ▶ Safety and Hazards ? 184,712
- ▶ Spectral Information ? 1,576,070
- ▶ Structures ? 11,819,155
- ▶ Toxicity ? 118,115
- ▶ Use and Manufacturing ? 107,948

- ▼ Agrochemical Information ? 3,135
 - Agrochemical Category ? 1,977
 - Agrochemical Transformations ? 1,491
 - EU Pesticides Data ? 1,248
 - USDA Pesticide Data Program ? 663
- ▼ Associated Disorders and Diseases ? 30,136
 - Disease and References 13,386
- ▼ Interactions and Pathways ? 207,277
 - ▶ Molecular Imaging Information ? 1,910
 - ▶ Protein Bound 3D Structures ? 71,666
 - Chemical-Target Interactions ? 83,557
 - Drug-Drug Interactions ? 3,398
 - Drug-Food Interactions ? 1,246
 - Pathways ? 62,525

PubChem Furathiocarb (Compound)

7.3 EU Pesticides Data

Active Substance	furathiocarb
Status	Not approved [Reg. (EC) No 1107/2009]
Legislation	2002/2076

PubChem Nicotine (Compound)

14 Associated Disorders and Diseases

Showing 11 to 15 of 252 items [Download](#)

Disease	Evidence Type	Evidence PMID
Airway Obstruction	marker/mechanism	33082140
Alzheimer Disease	therapeutic	16627626 17135361

PubChem Acetylcarnitine (Compound)

12.1 Pathways

4 items [Download](#)

Oxidation of Branched-Chain Fatty Acids
Source: PathBank External ID: SMP0000030
Taxonomy: Homo sapiens (human)

Creating Informative Subsets of PubChem



PubChemLite EXPOSOMICS

~350,000 entries "small"



Updated PubChemLite to Aug 2023 version #127

Merged sneumann merged 2 commits into ipb-halle:master from schymane:master last week

Database Settings

Database: PubChemLite_exposomics

Neutral Mass: 229.10948 Search ppm: 5

Formula: C9H16ClN5

Identifiers:

Retrieve Candidates 4 Candidates

zenodo

Search Upload Communities

August 26, 2023 Dataset Open Access

PubChemLite for Exposomics

Bolton, Evan; Schymanski, Emma; Kondic, Todor; Thiessen, Paul; Zhang, Jian (Jeff)

This is the repository for regular updates of the PubChemLite for Exposomics data collection. PubChemLite for Exposomics is a subset of PubChem selected from major categories of the Table of Contents page at the PubChem Classification Browser, described in DOI:10.1186/s13321-021-00489-0.

PubChemLite for Exposomics is compiled from 10 categories: AgroChemInfo, BioPathway, DrugMedicInfo, FoodRelated, PharmacolInfo, SafetyInfo, ToxicityInfo, KnownUse, DisorderDisease, Identification.

PubChemCIDs have been collapsed by InChIKey first block, reporting the structure from the most annotated CID, plus related CIDs. Entries that will be ignored by MetFrag (salts, disconnected substances) or cause errors (e.g. transition metals) have been removed. The Patent and PubMed ID counts are extracted from files on the PubChem FTP site. The 'AnnoTypeCount' term counts how many of the categories are represented, the subsequent column (named per category) counts the number of annotation categories available in the next sub-category of the TOC entry.

These files can be used 'as is' as localCSV for MetFrag Command Line.

PubChem Compound TOC

Agrochemical Information	3,135
Associated Disorders and Diseases	30,136
Biologic Description	2,511,444
Biological Test Results	4,567,078
Chemical and Physical Properties	268,878
Classification	22,965,005
Drug and Medication Information	21,177
Food Additives and Ingredients	7,627
Identification	4,808
Information Sources	47,725,078
Interactions and Pathways	207,277
Literature	4,076,955
Names and Identifiers	7,021,765
Patents	39,104,437
Pharmacology and Biochemistry	114,060
Related Records	13,282,616
Safety and Hazards	184,712
Spectral Information	1,576,070
Structures	11,819,155
Toxicity	118,115
Use and Manufacturing	107,948

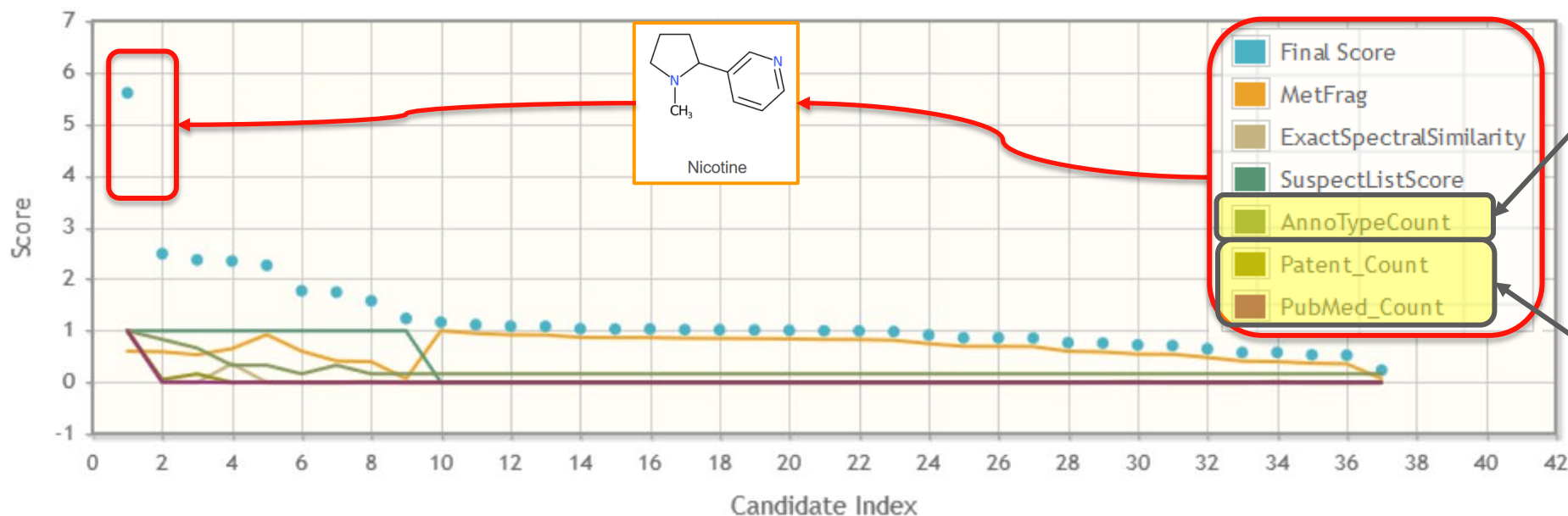
Schymanski et al. (2021) DOI: [10.1186/s13321-021-00489-0](https://doi.org/10.1186/s13321-021-00489-0)

PubChemLite: Fewer and more relevant candidates – with context



Statistics

Candidate Score Distribution



PubChem Compound TOC	67,343,260
Agrochemical Information	3,135
Associated Disorders and Diseases	30,136
Biologic Description	2,511,444
Biological Test Results	4,567,078
Chemical and Physical Properties	268,878
Classification	22,965,005
Drug and Medication Information	21,177
Food Additives and Ingredients	7,627
Identification	4,808
Information Sources	47,725,078
Interactions and Pathways	207,277
Literature	4,076,955
Names and Identifiers	7,021,765
Patents	39,104,437
Pharmacology and Biochemistry	114,060
Related Records	13,282,616
Safety and Hazards	184,712
Spectral Information	1,576,070
Structures	11,819,155
Toxicity	118,115
Use and Manufacturing	107,948



Literature



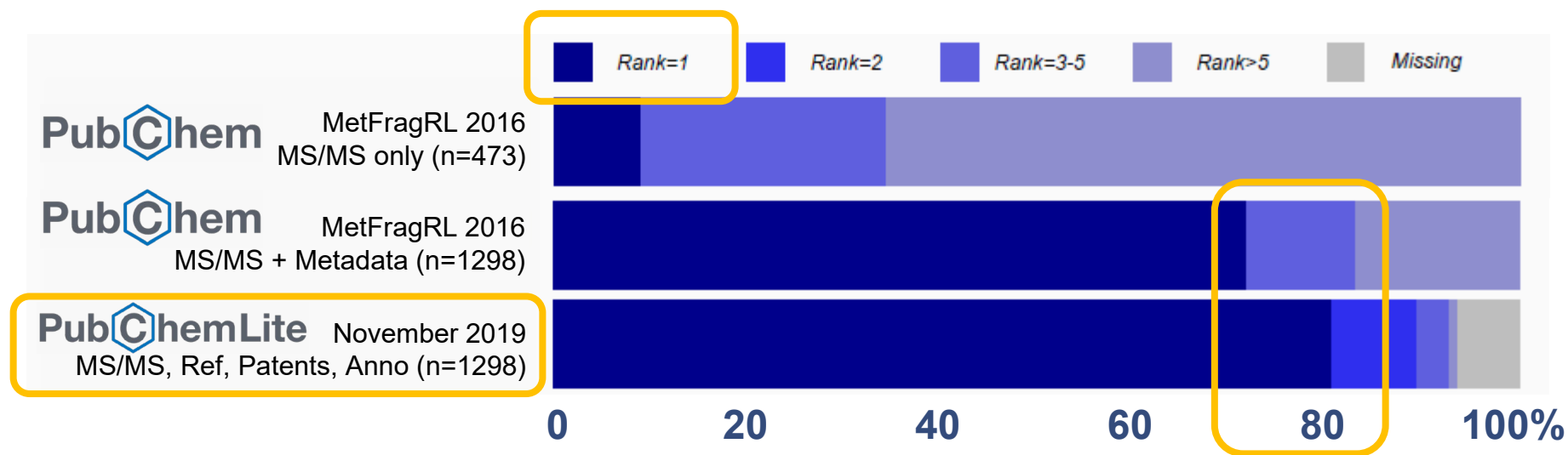
Patents



How does PubChemLite perform?



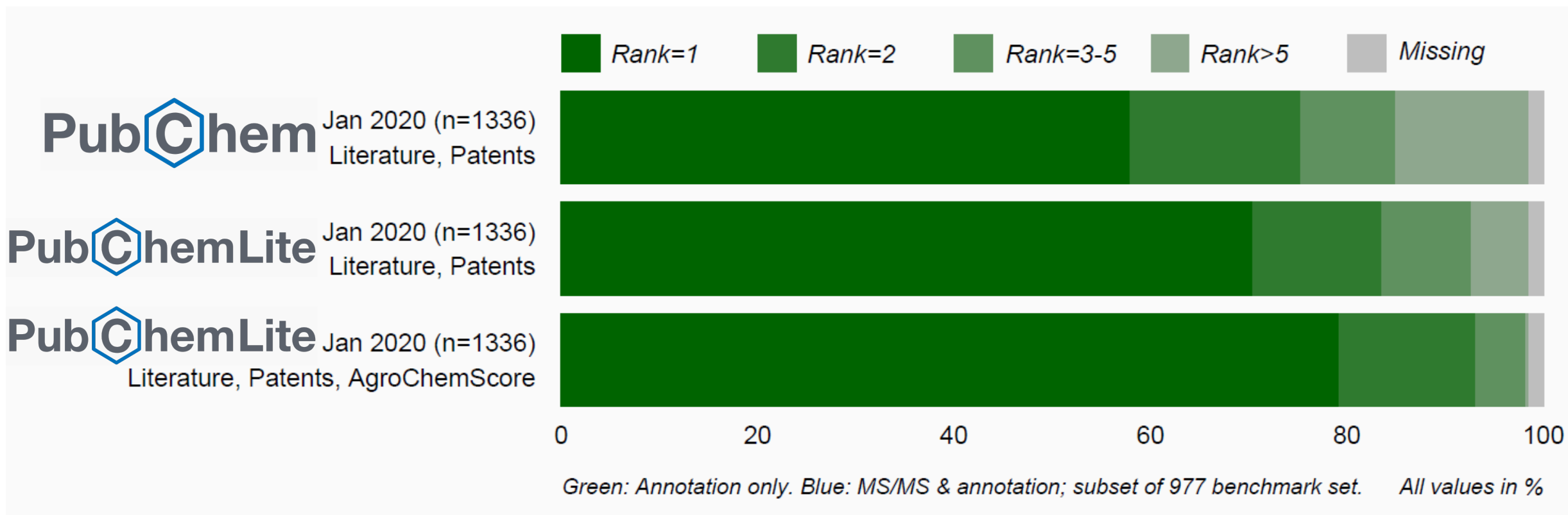
- ~111 M => ~400 K ... how does this influence performance?



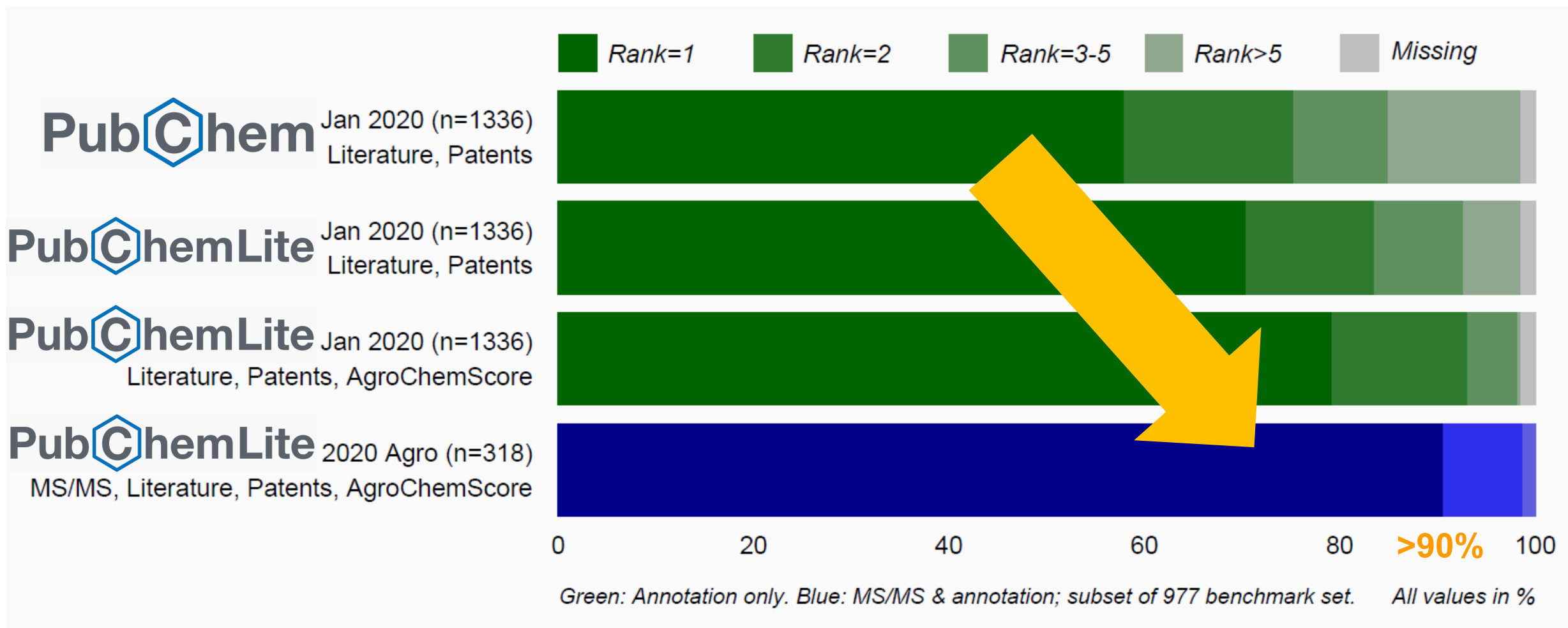
>80 % ranked in first place
~90 % ranked first or second!

} high throughput
identification

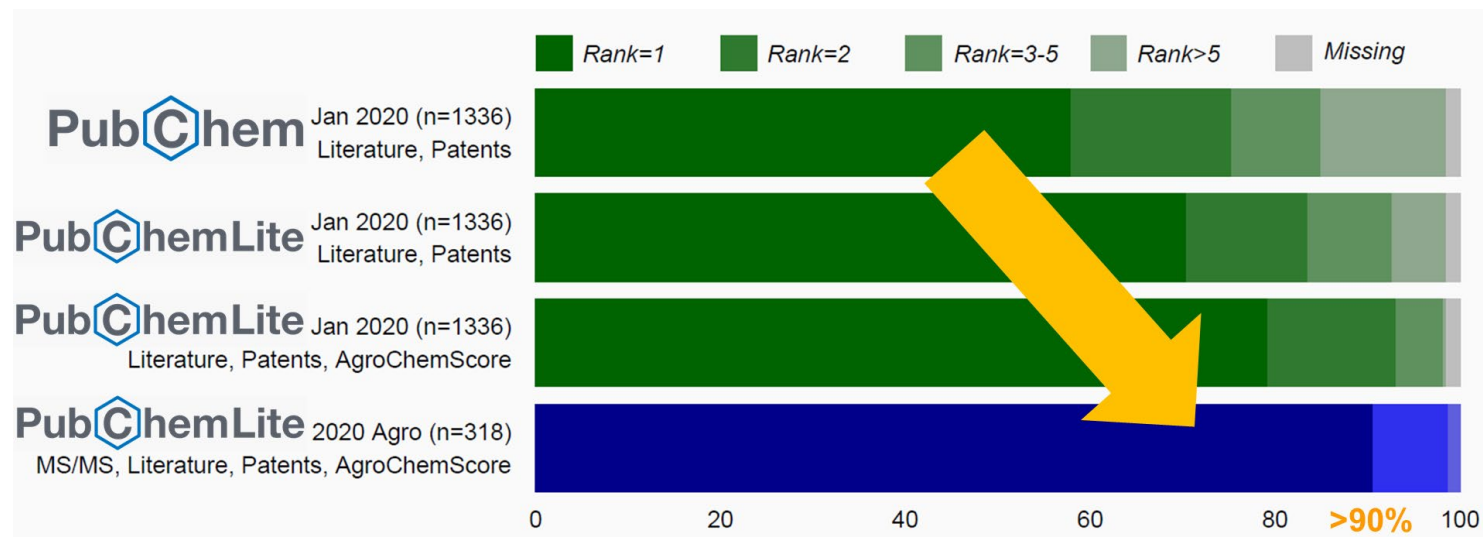
Influence of the Annotation Content in PubChemLite



Influence of the Annotation Content in PubChemLite



Influence of the Annotation Content in PubChemLite



>90% of (well known) entries are Top 1 rank

>97% are Top 2 ... 100 % Top 3 ...

***Exposomics: Top 1-3 matches will cover a lot!
(and quickly!)***

Putting PubChemLite into Practice?

- Downloadable & fully integrated in patRoön & MetFrag



zenodo

Search [] Upload Communities emma.schymanski@uni.lu

zenodo.org will be unavailable for 2 hours on September 29th from 06:00-08:00 UTC. See announcement.

August 26, 2023

PubChemLite for Exposomics

Dataset Open Access

Communities

LCSB Environmental Cheminformatics Group

2,087 views 1,778 downloads

See more details...

DOI: [10.5281/zenodo.5995885](https://doi.org/10.5281/zenodo.5995885)

MetFragWeb

Learn more

Use MetFrag directly from your browser which makes it pretty easy for beginners

MetFrag CL

Learn more

The new MetFrag commandline tool providing additional scoring terms for scoring candidates with MS/MS spectra

<https://rickhelmus.github.io/patRoön/>

patRoön 1.2.0

Reference Tutorial Handbook Changelog

Installation

patRoön itself can be installed as any other R package, however, depending on which algorithms you want to use in your workflow, some extra steps may be required to install all the necessary tools. Please see the [installation section in the handbook](#) for more information.

Getting start

For a very quick start:

```
library(patRoön)
newProject()
```

The `newProject()` function will pop-up a dialog screen (requires R Studio!) which will allow you to quickly select the analyses and common workflow options to subsequently generate a template R processing script.

However, for a better guide to get started it is recommended to read the [tutorial](#). Afterwards the [handbook](#) is a recommended read if you want to know more about advanced usage of patRoön. Finally, the [reference](#) outlines all the details of the patRoön package.

<https://github.com/rickhelmus/patRoön/issues>

License

GPL-3

Citation

Citing patRoön

Developers

Rick Helmus
Author, maintainer

All authors...

Dev status

PASSED

build passing

codecov 82%

image size 2.17 GB

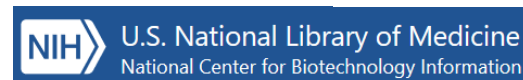
<https://ipb-halle.github.io/MetFrag/>

Schymanski *et al.* (2021) DOI: [10.1186/s13321-021-00489-0](https://doi.org/10.1186/s13321-021-00489-0) & DOI: [10.5281/zenodo.5995885](https://doi.org/10.5281/zenodo.5995885)

Helmus *et al.* (2021) patRoön *J Cheminform* **13**, 1. DOI: [10.1186/s13321-020-00477-w](https://doi.org/10.1186/s13321-020-00477-w)

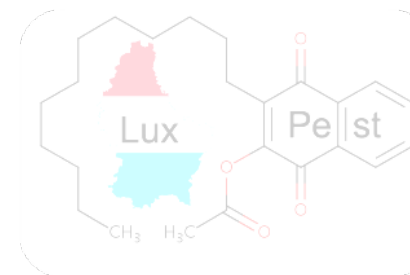
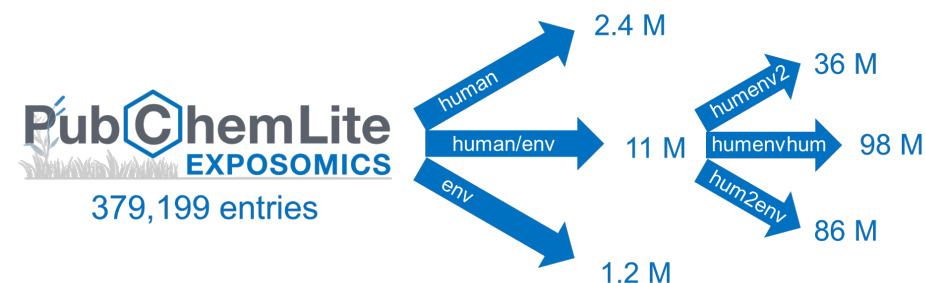
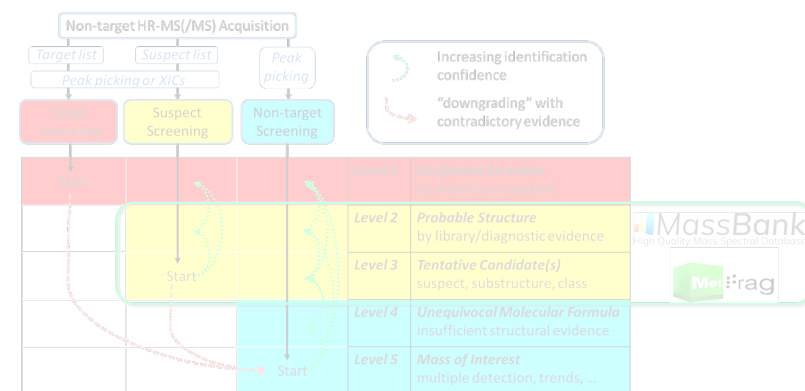


UNIVERSITY OF AMSTERDAM



Outline

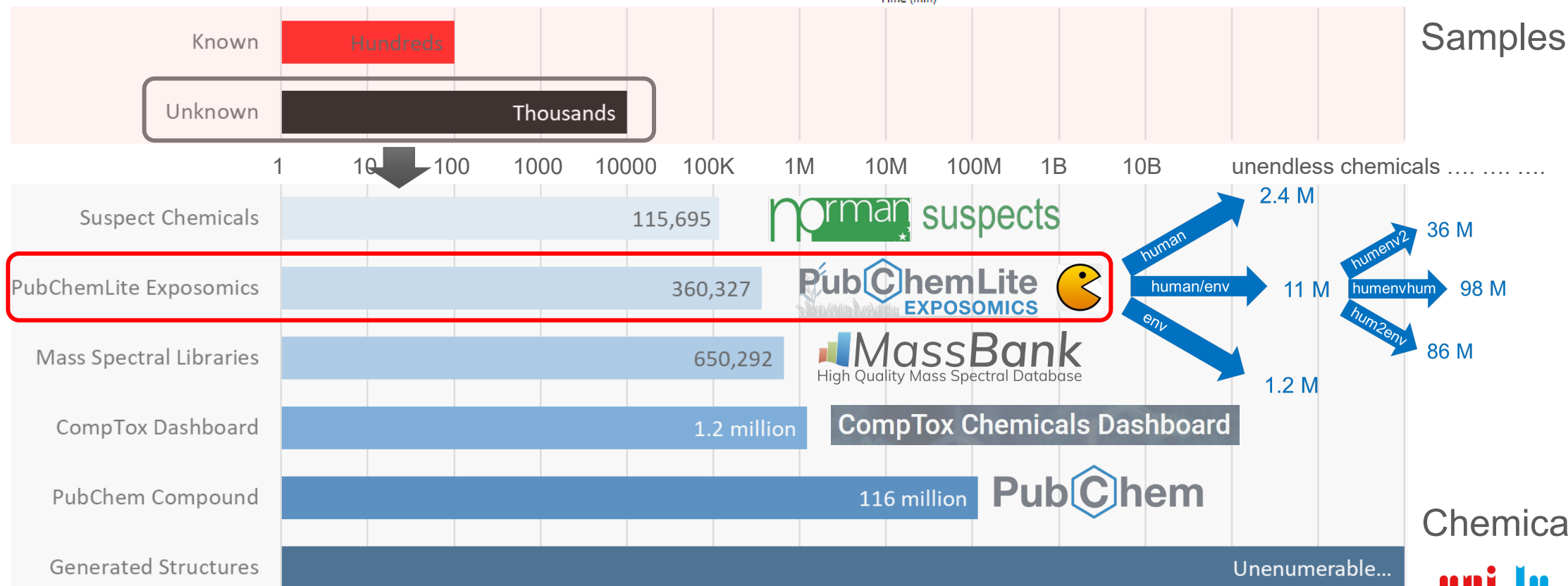
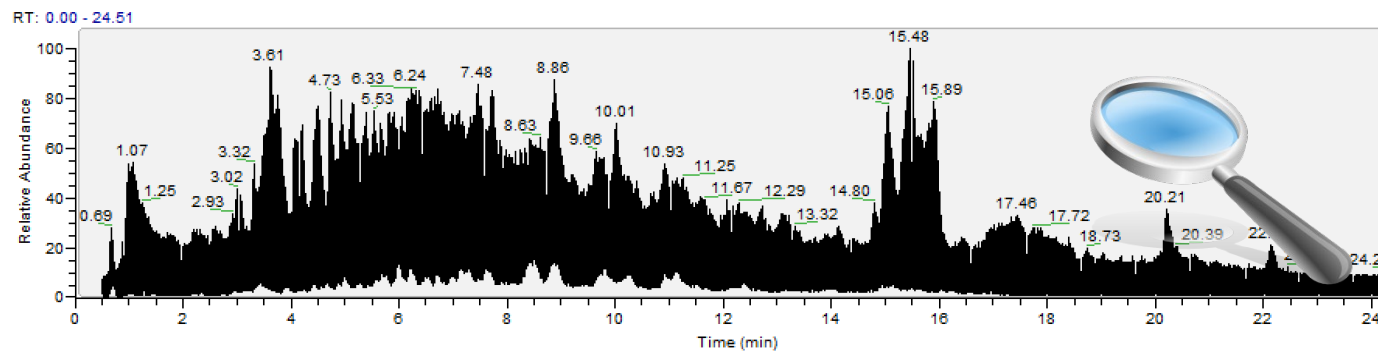
- Identifying Chemicals with NT-HRMS
 - Open Resources for Identification
 - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases & Workflows
 - Defining Relevant Chemical Space
 - Transformations: Known & Predicted
- No package too small?
- Applications in Environmental Studies



Environmental Cheminformatics, NT HR-MS & *Transformations*

High resolution
mass spectrometry

AND connecting
chemical knowledge



Transforming PubChemLite with **BioTransformer 3.0**

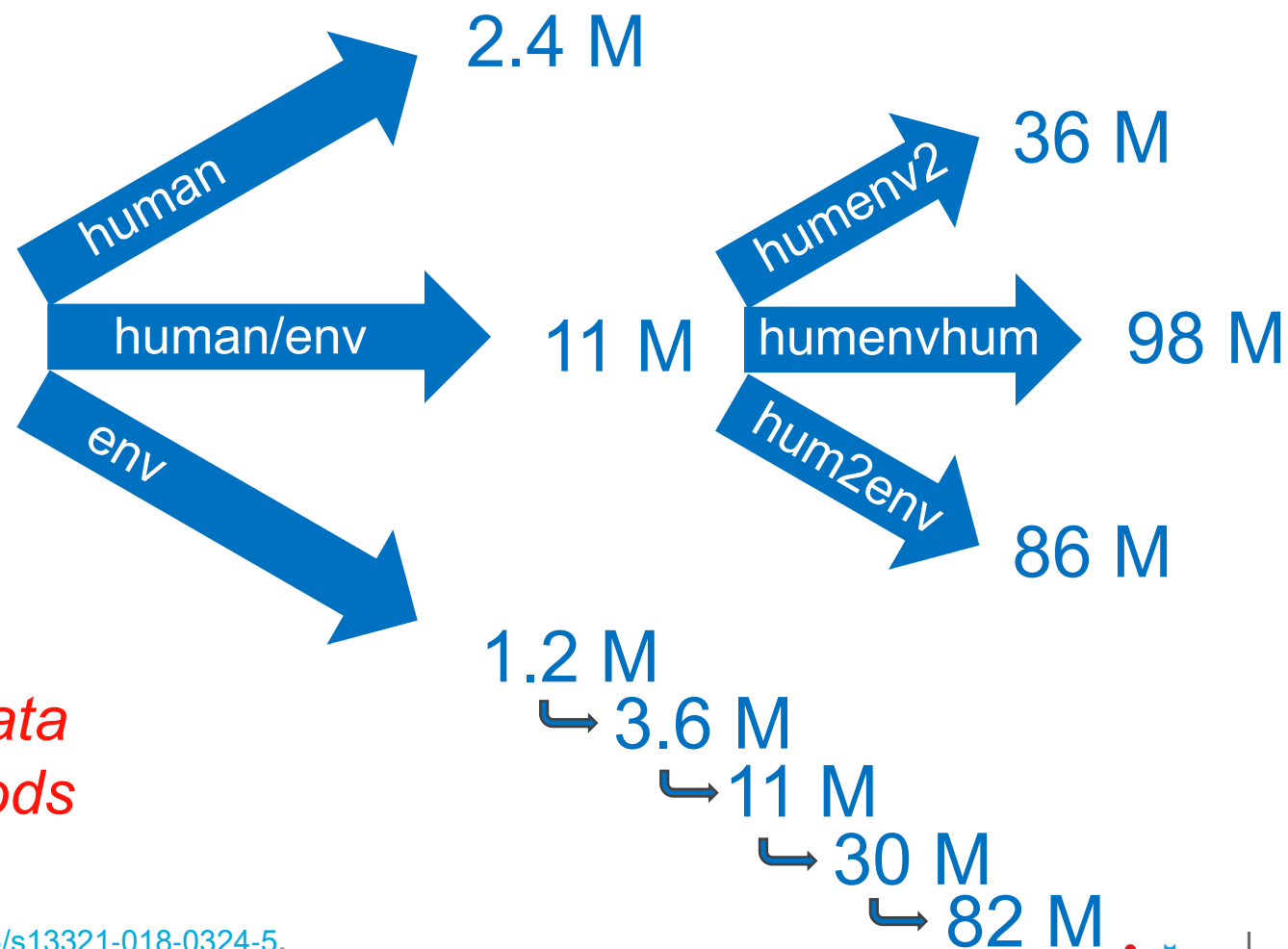
<http://biotransformer.ca/>

PubChemLite
EXPOSOMICS

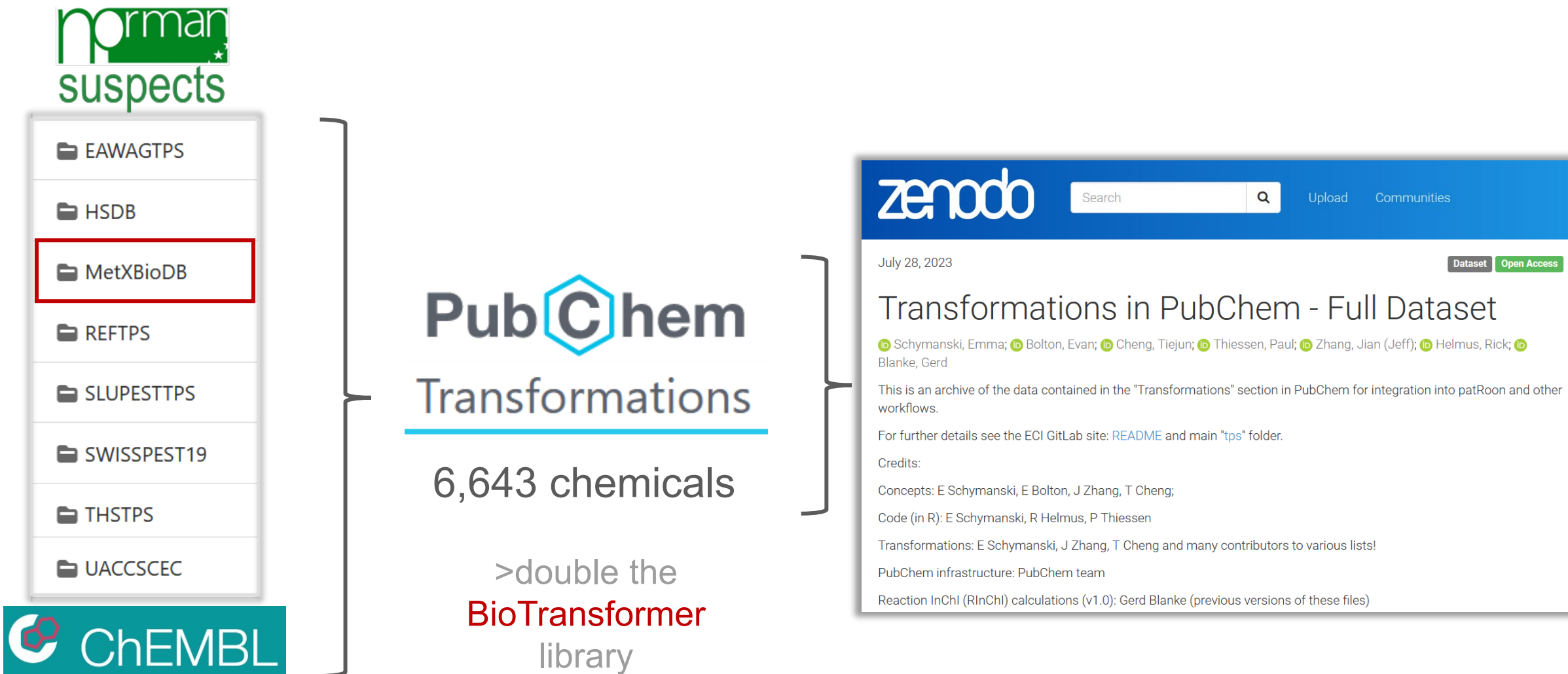
379,199 entries

version 1.0.0; DOI: [10.5281/zenodo.5995886](https://doi.org/10.5281/zenodo.5995886)

*Combinatorial explosion – more data
needed to predict reaction likelihoods*



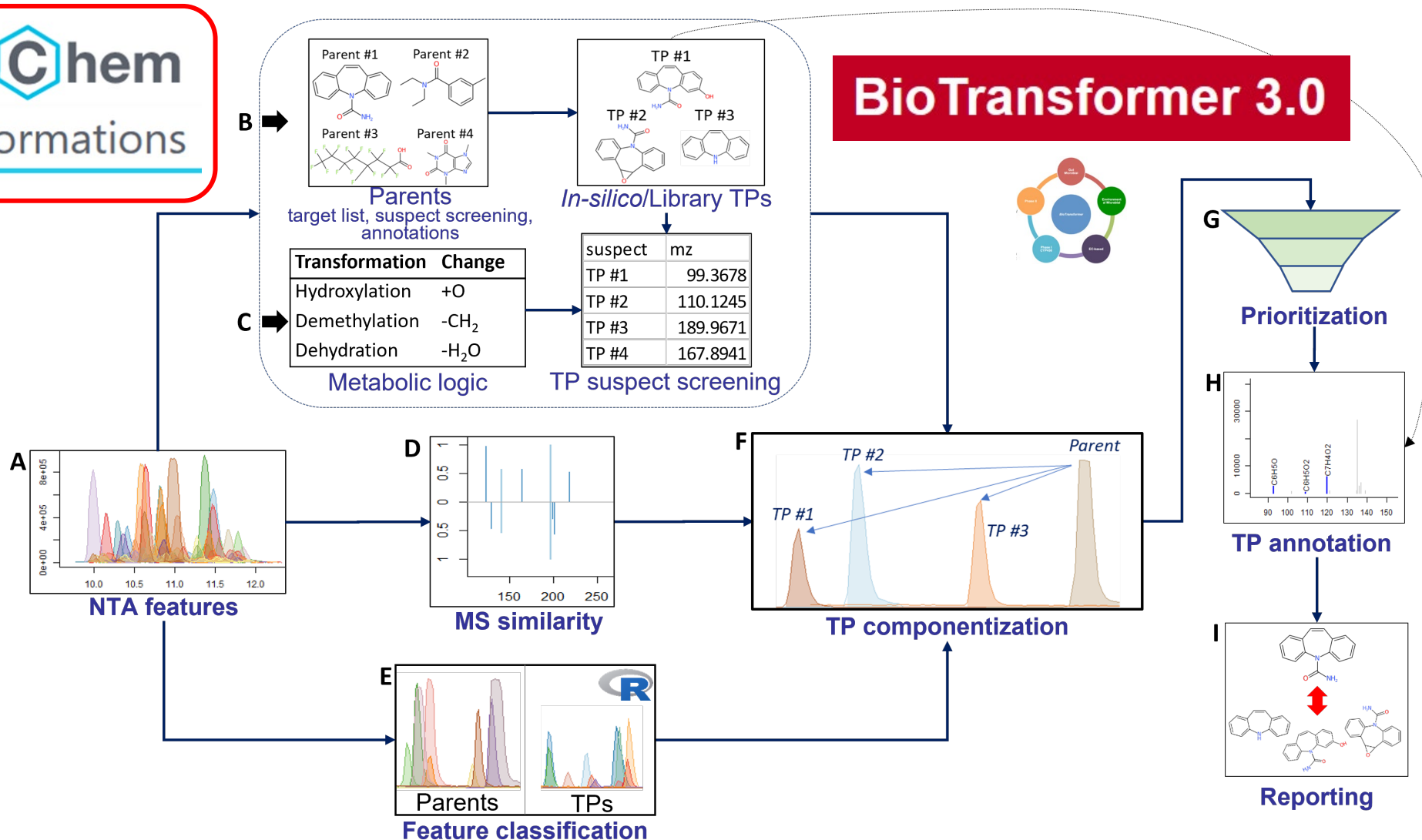
FAIR Transformations in PubChem and NORMAN-SLE



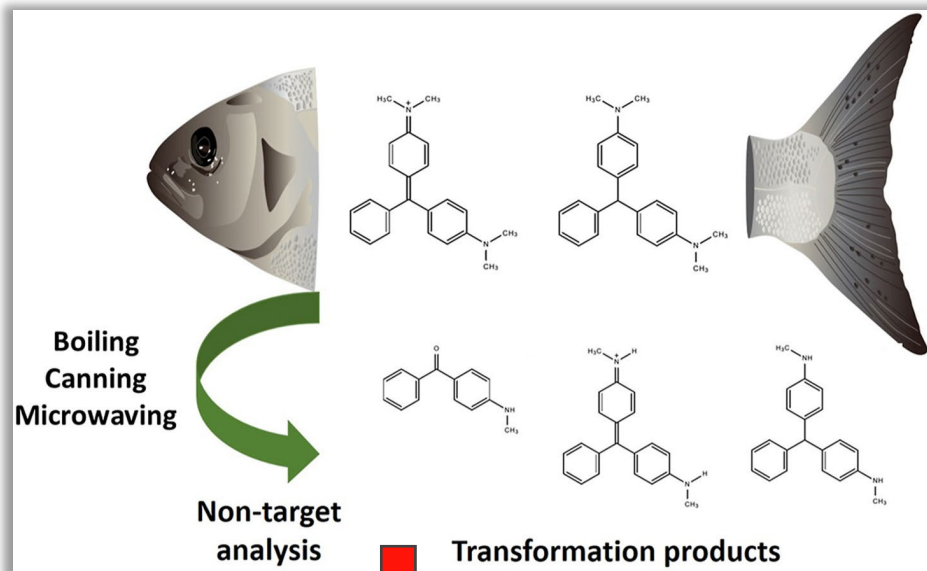
Open Transformation Products Workflows in patRoön 2.0



PubChem
Transformations



What about *new* information? Help add it in!



zenodo

Search Upload Communities

March 8, 2023 Dataset Open Access

S74 | REFTPS | Transformation Products and Reactions from Literature

Schymanski, Emma; Baesu, Anca; Chirsir, Parviel

This is the collection associated with list S74 REFTPS Transformation Products and Reactions from Literature on the NORMAN Suspect List Exchange.

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

PubChem (4-(Methylamino)phenyl)(phenyl)methanone (Compound)

7.1 Transformations

2 items Download

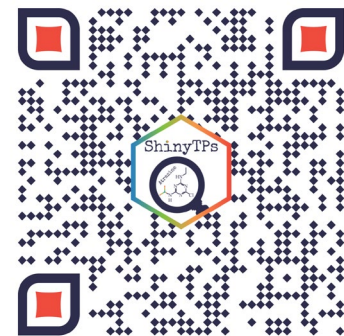
Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Evidence DO
	Malachite green		4-(Methylamino)benzophenone	Thermally induced deconjugation and demethylation	10.1016/j.crfs.
	Malachite green cation		4-(Methylamino)benzophenone	Thermally induced deconjugation and demethylation	10.1016/j.crfs.

< >

▼ NORMAN Suspect List Exchange

Source	NORMAN Suspect List Exchange
Record Name	4-(Methylamino)benzophenone

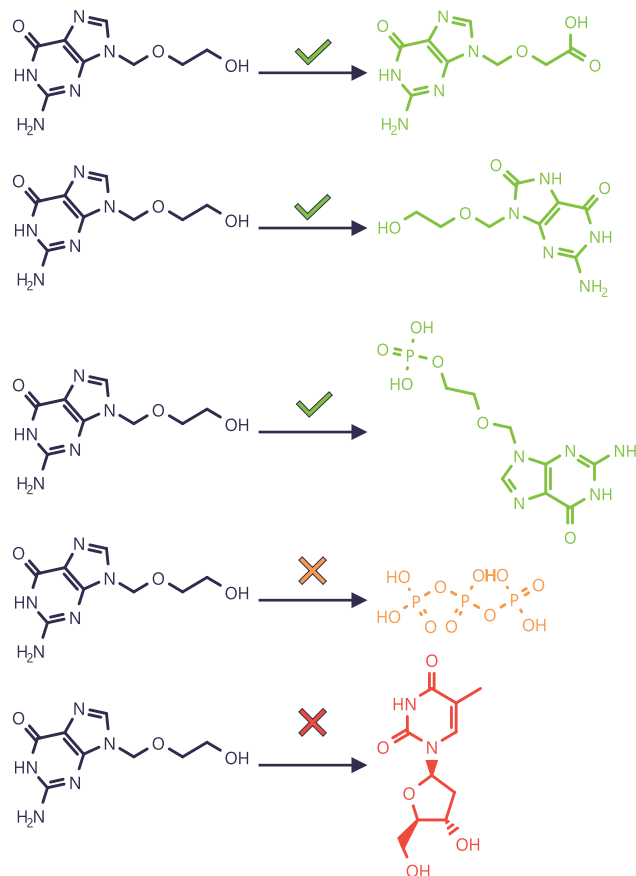
Curating Text-mined Content in PubChem



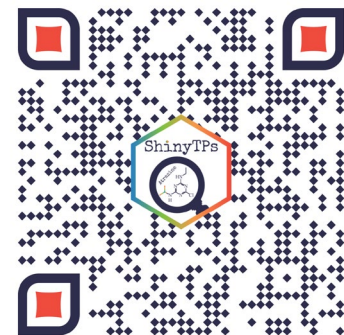
PubChem Acyclovir (Compound)

8.6 Metabolism/Metabolites

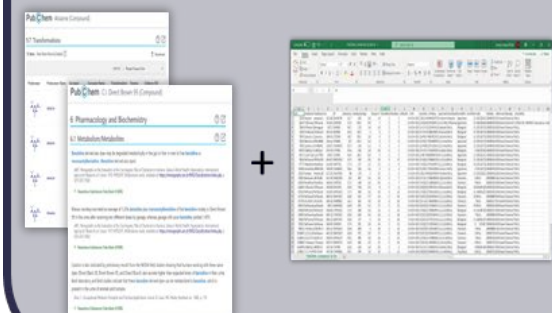
Acyclovir is metabolized partially to **9-carboxymethoxymethylguanine** and minimally to **8-hydroxy-9-(2-hydroxyethoxymethyl)guanine**. In vitro, **acyclovir** also is metabolized to **acyclovir monophosphate**, diphosphate, and **triphosphate** in cells infected with herpes viruses, principally by intracellular phosphorylation of the drug by virus coded **thymidine** kinase and several cellular enzymes.



ShinyTPs: Curating TPs from Text Mining Results



1. Read in data



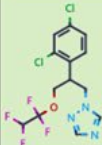
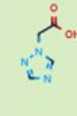
2. Launch ShinyTPs



3. Check available data

Select an input compound

Tetraconazole

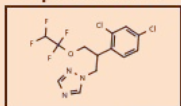
Parent	Parent CID	TP	TP CID
	80277		1810180

4. Curate reactions

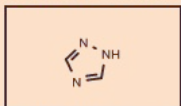
Select an input compound

Tetraconazole

Input structure



Potential TP structure



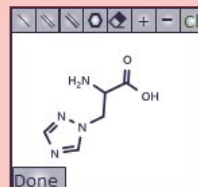
Tetraconazole results in the formation of /1,2,4-triazole (T), triazolyl alanine (TA), triazolyl acetic acid (TAA)/ as well as /triazolyl hydroxypropionic acid/ (THP).

5. Add missing entries

Select an input compound

Tetraconazole

Draw a structure

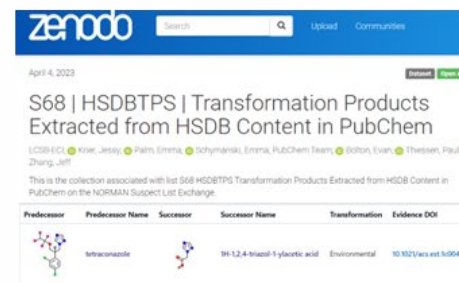


SMILES:

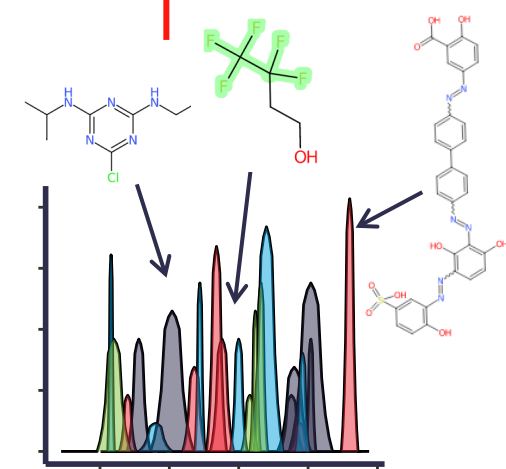
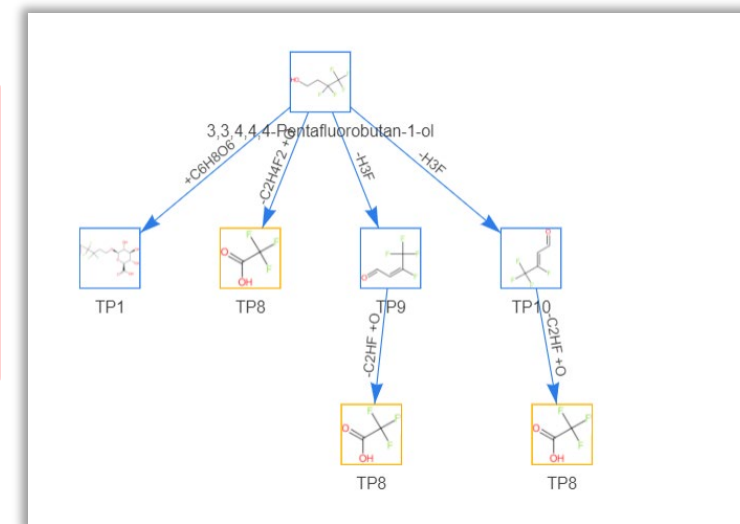
NC1C=NC=N1C(=O)O

Tetraconazole results in the formation of /1,2,4-triazole (T), triazolyl alanine (TA), triazolyl acetic acid (TAA)/ as well as /triazolyl hydroxypropionic acid/ (THP).

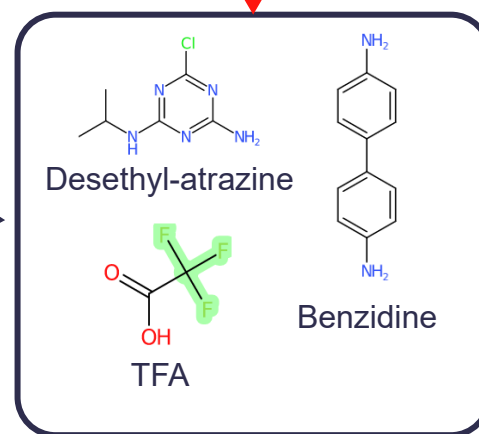
6. Export and upload



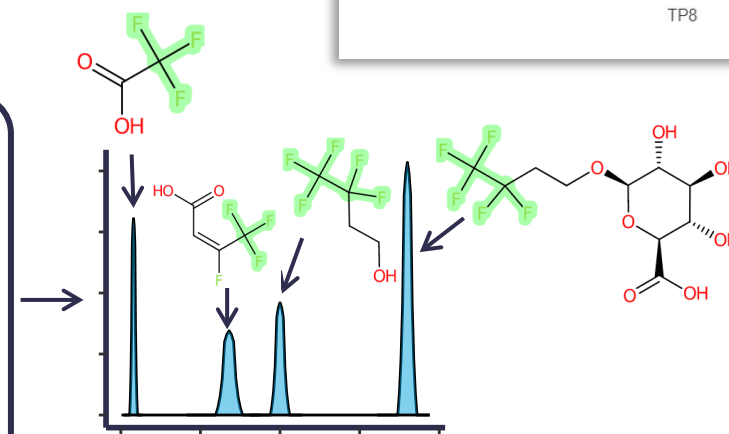
Applying TP NT-HRMS Workflows with patRoan



Suspect/Non-target screening



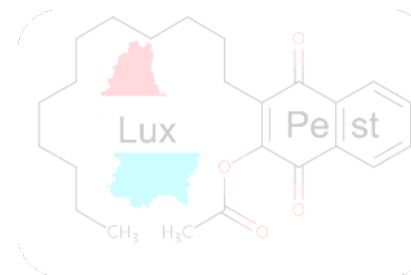
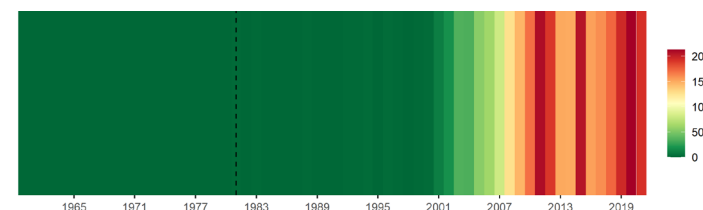
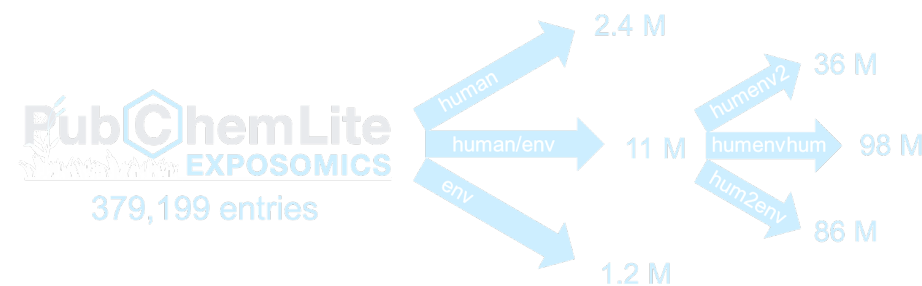
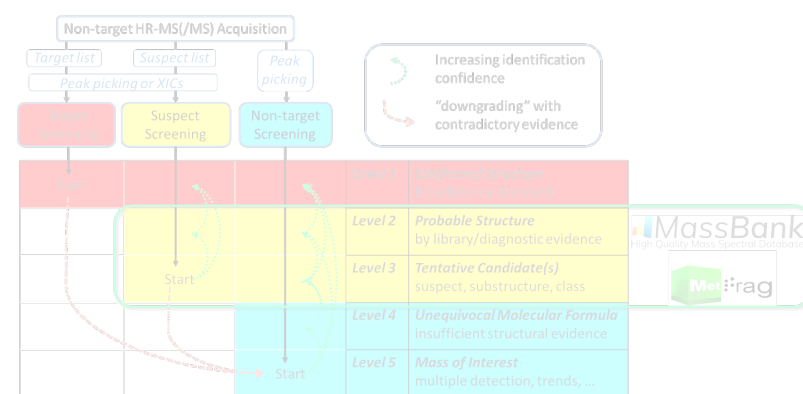
TP library



TP suspect screening

Outline

- Identifying Chemicals with NT-HRMS
 - Open Resources for Identification
 - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases & Workflows
 - Defining Relevant Chemical Space
 - Transformations: Known & Predicted
- No package too small?
- Applications in Environmental Studies



Chemical Stripes in R - for Patents & Literature



```
> chemical_stripes(74483)
Getting compound information
A total of 32461 patents were found for CID 74483
[=====>]
Downloading patent data...
[=====>] 60% 7s
Processing patent data
32460 patents were processed for CID 74483
[=====>] 80% 3s
Plotting chemical stripes for the years between 1960 and 2021

Your stripes have been saved as png_74483_1960_2021.png in your folder C:/Users/dagny.aurich/Documents/R_stripes/png_74483_1960_2021.png
```

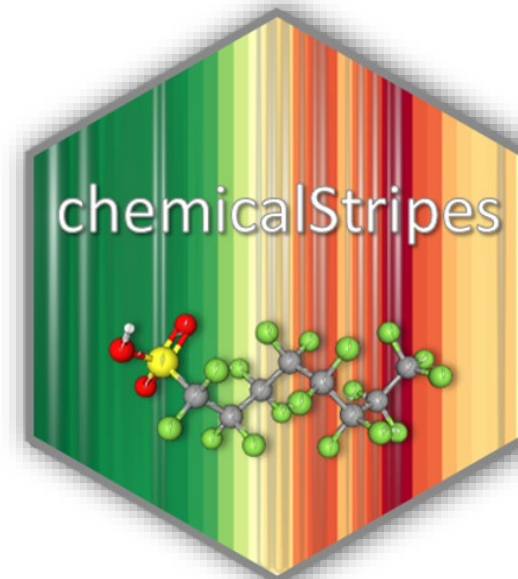
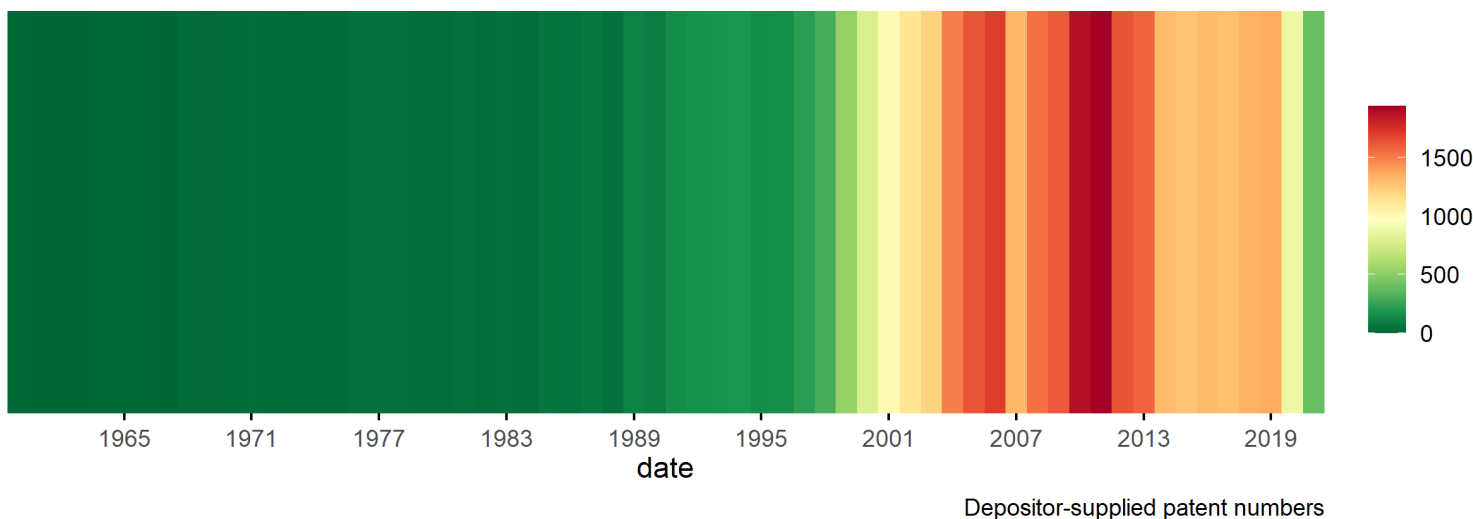
Chemical Stripes for Perfluorooctanesulfonic acid

PubChem CID: 74483

IUPACName: 1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,8-heptafluorooctane-1-sulfonic acid

Molecular Formula: C₈HF₁₇O₃S

Exact Mass: 499.9374938



Chemical Stripes in R

Patents & Literature

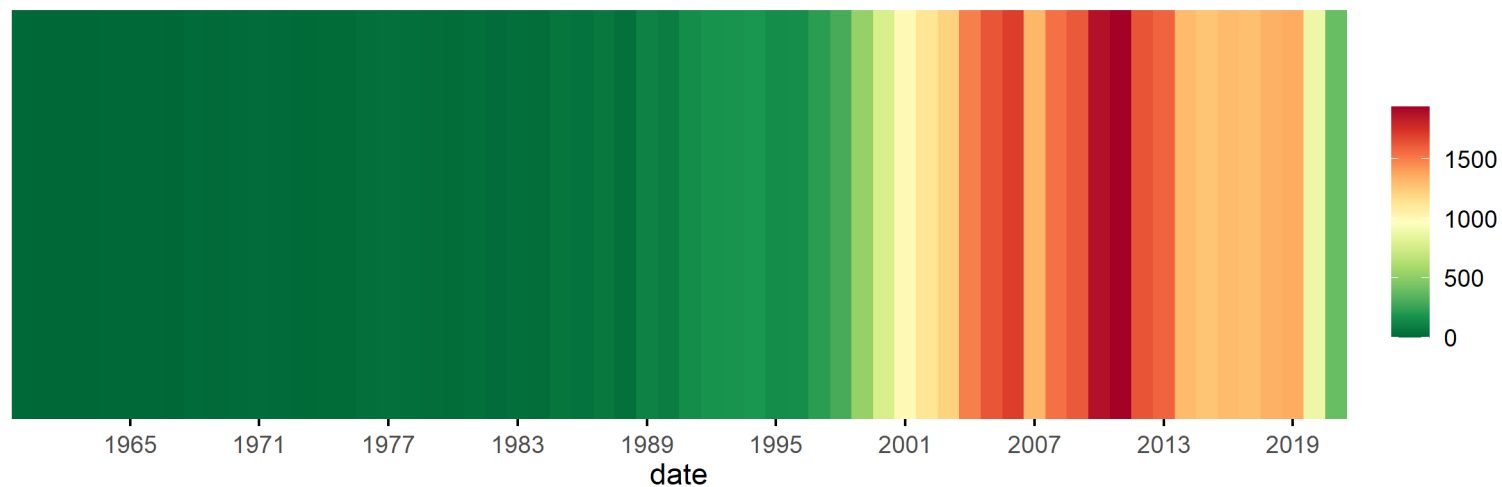


Chemical Stripes for Perfluorooctanesulfonic acid

PubChem CID: 74483

First patent: 1913

Patents



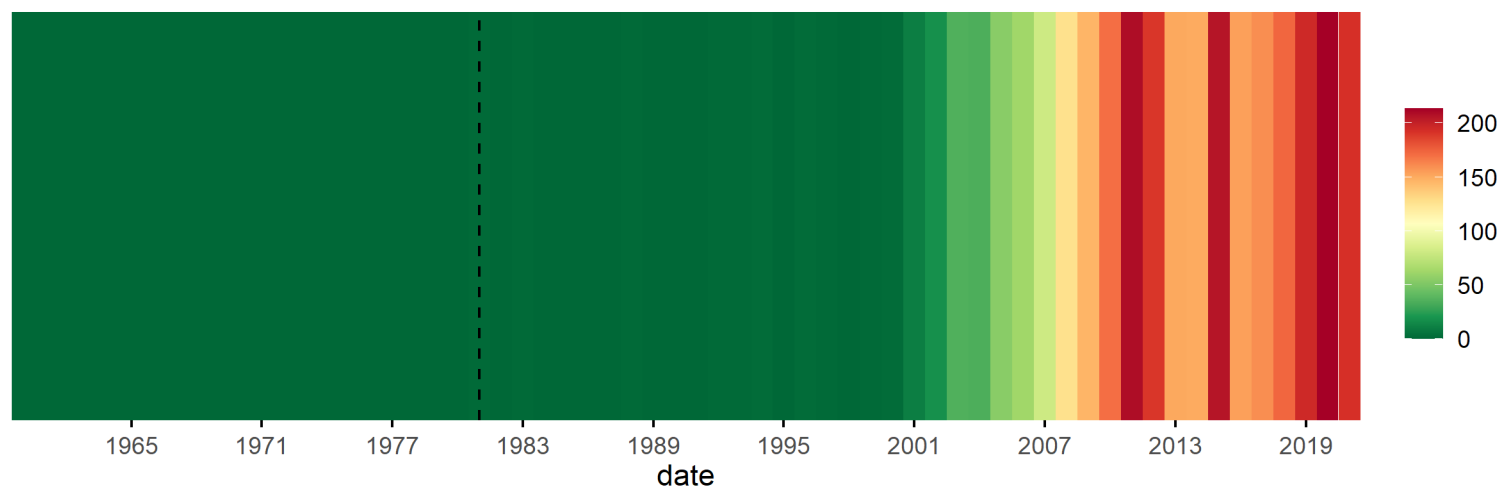
Depositor-supplied patent numbers

Chemical Stripes for Perfluorooctanesulfonic acid

PubChem CID: 74483

First reference: 1981

Literature



Consolidated reference numbers

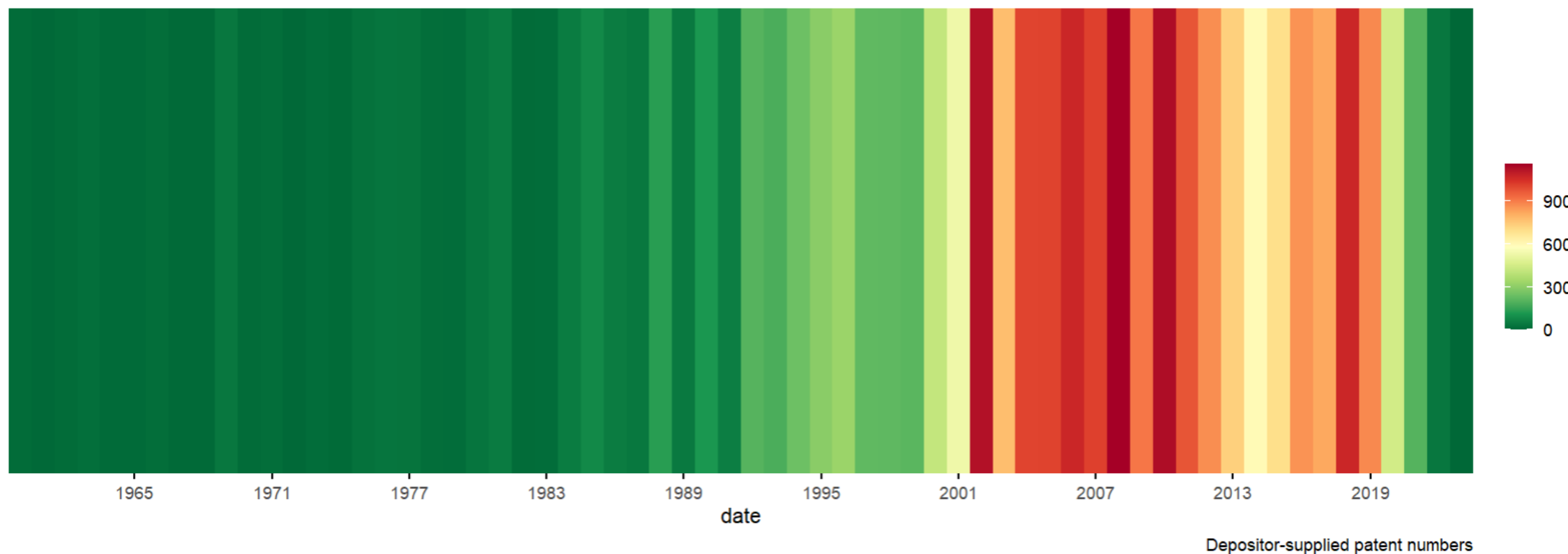
“Summarized” Chemical Stripes in R



Summarized Chemical Stripes - EU REACH C9-C14 PFCAs

Patent data compiled from 230 CIDs

First patent of all selected CIDs: 1951 (CID(s): 67822,67821,67545,9555)



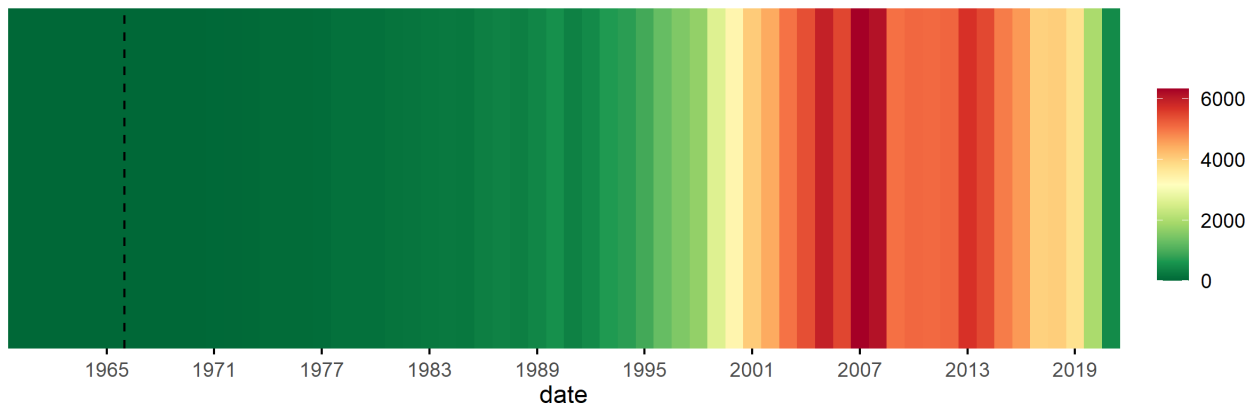
Variety of Stripe Patterns: Patents



Chemical Stripes for Caffeine

PubChem CID: 2519

First patent: 1966

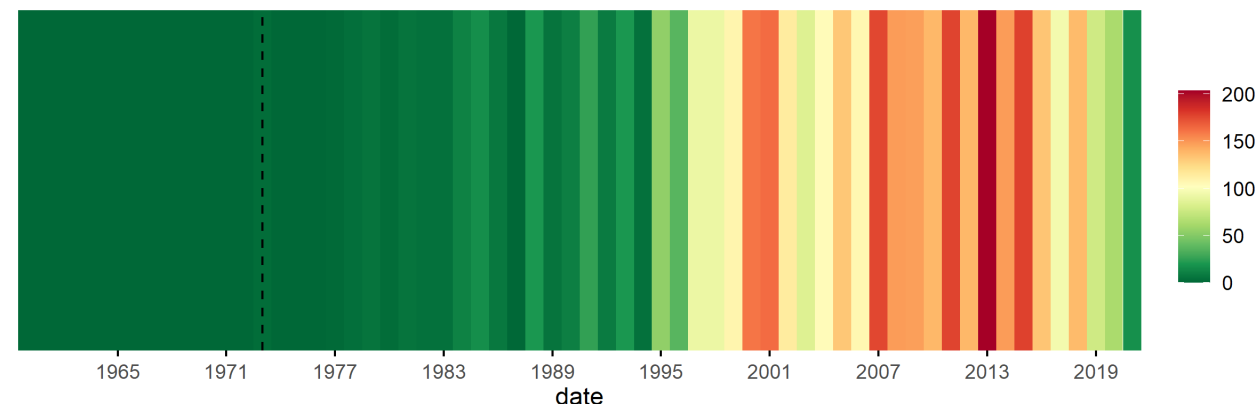


Depositor-supplied patent numbers

Chemical Stripes for Acetylcarnitine

PubChem CID: 1

First patent: 1973

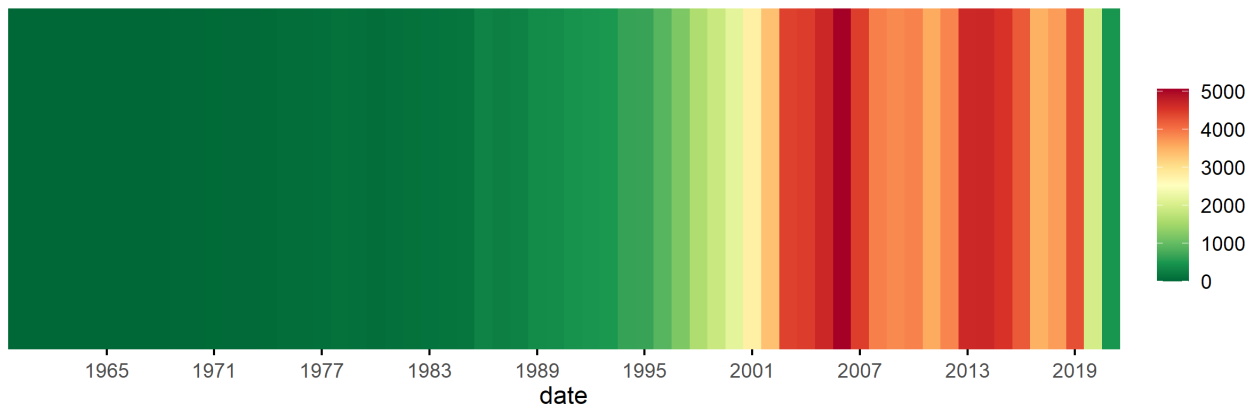


Depositor-supplied patent numbers

Chemical Stripes for Nicotine

PubChem CID: 89594

First patent: 1954

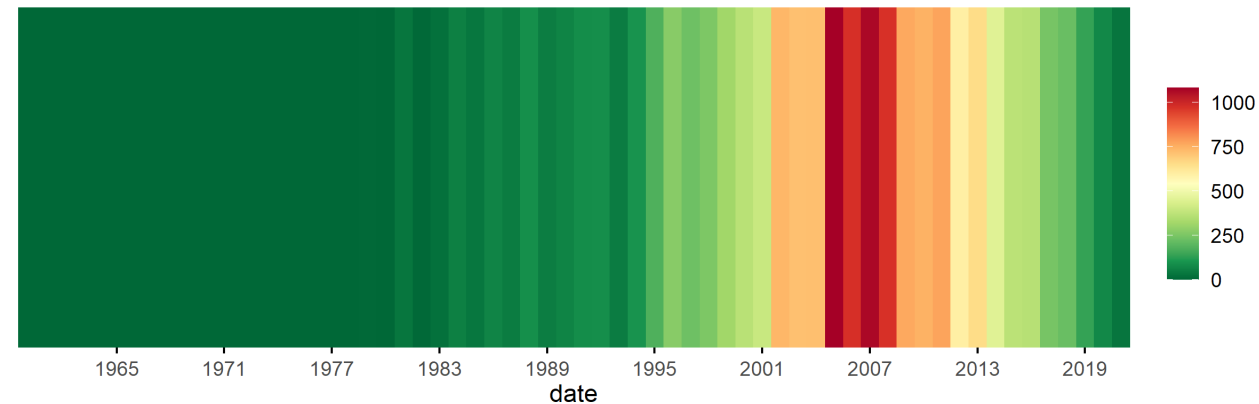


Depositor-supplied patent numbers

Chemical Stripes for Gallopamil

PubChem CID: 1234

First patent: 1951



Depositor-supplied patent numbers

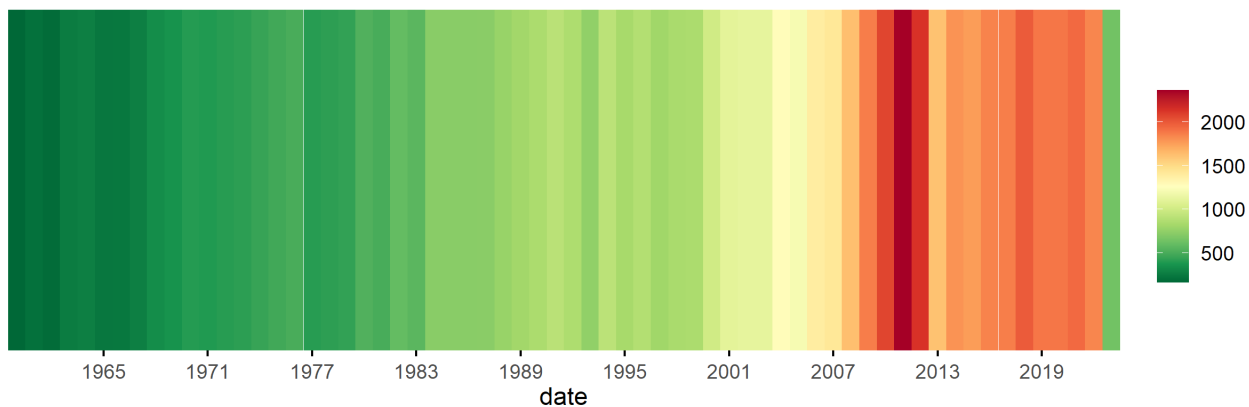
<https://gitlab.lcsb.uni.lu/eci/chemicalstripes>

Variety of Stripe Patterns: Literature



Chemical Stripes for Caffeine

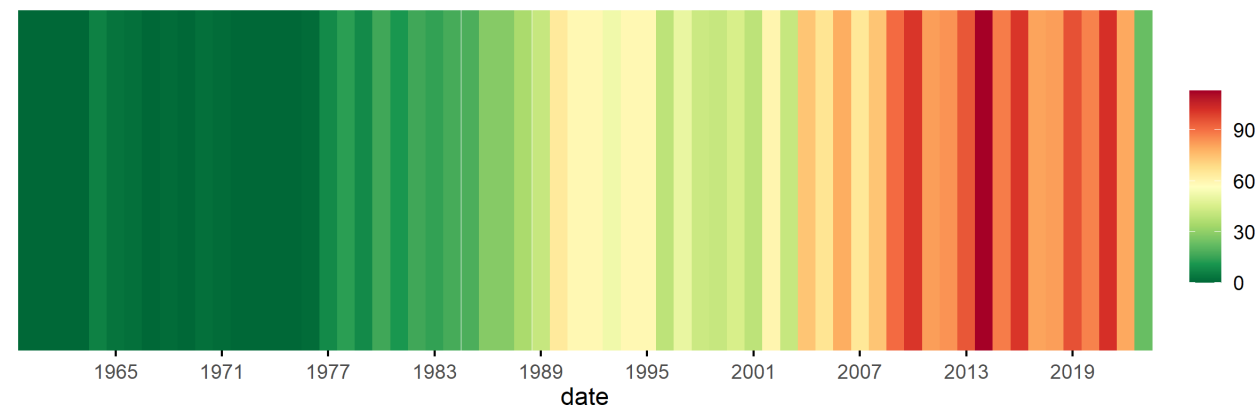
PubChem CID: 2519
First reference: 1851



Consolidated reference numbers

Chemical Stripes for Acetylcarnitine

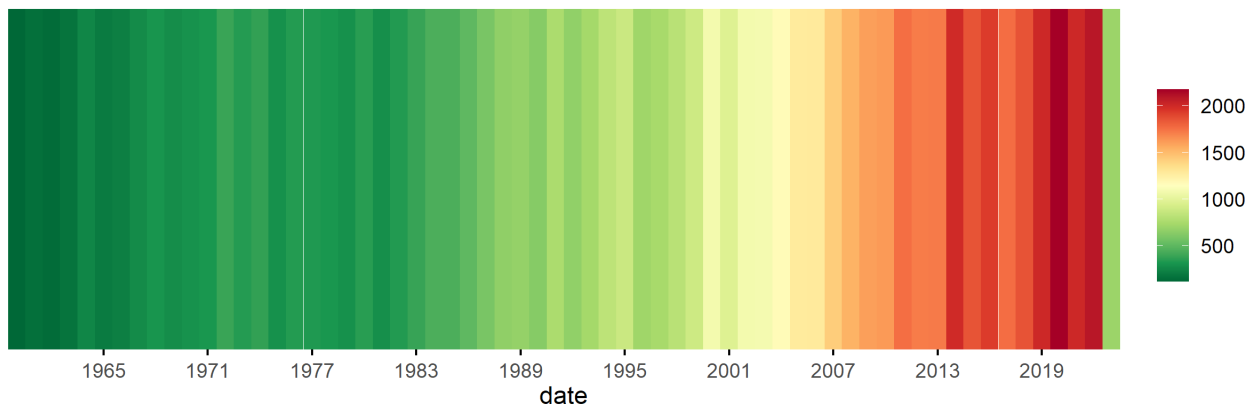
PubChem CID: 1
First reference: 1959



Consolidated reference numbers

Chemical Stripes for Nicotine

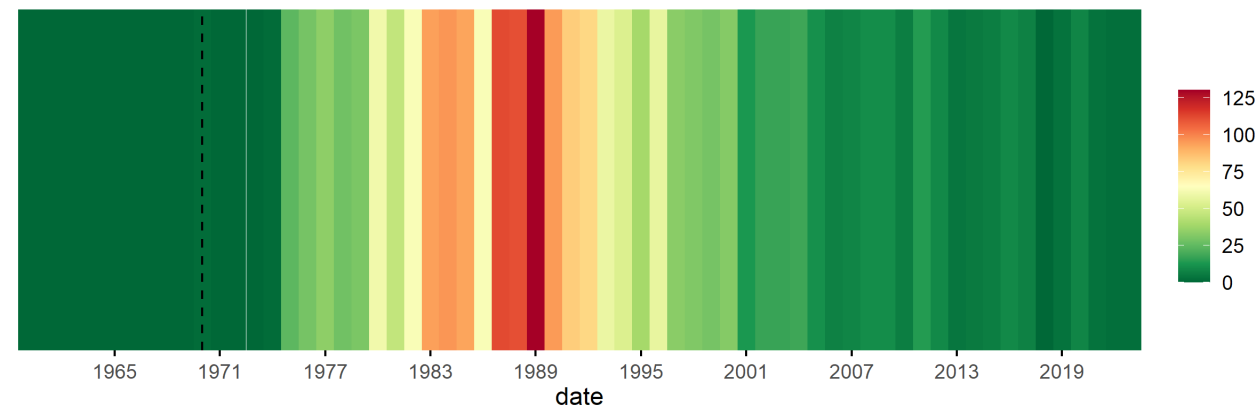
PubChem CID: 89594
First reference: 1854



Consolidated reference numbers

Chemical Stripes for Gallopamil

PubChem CID: 1234
First reference: 1970

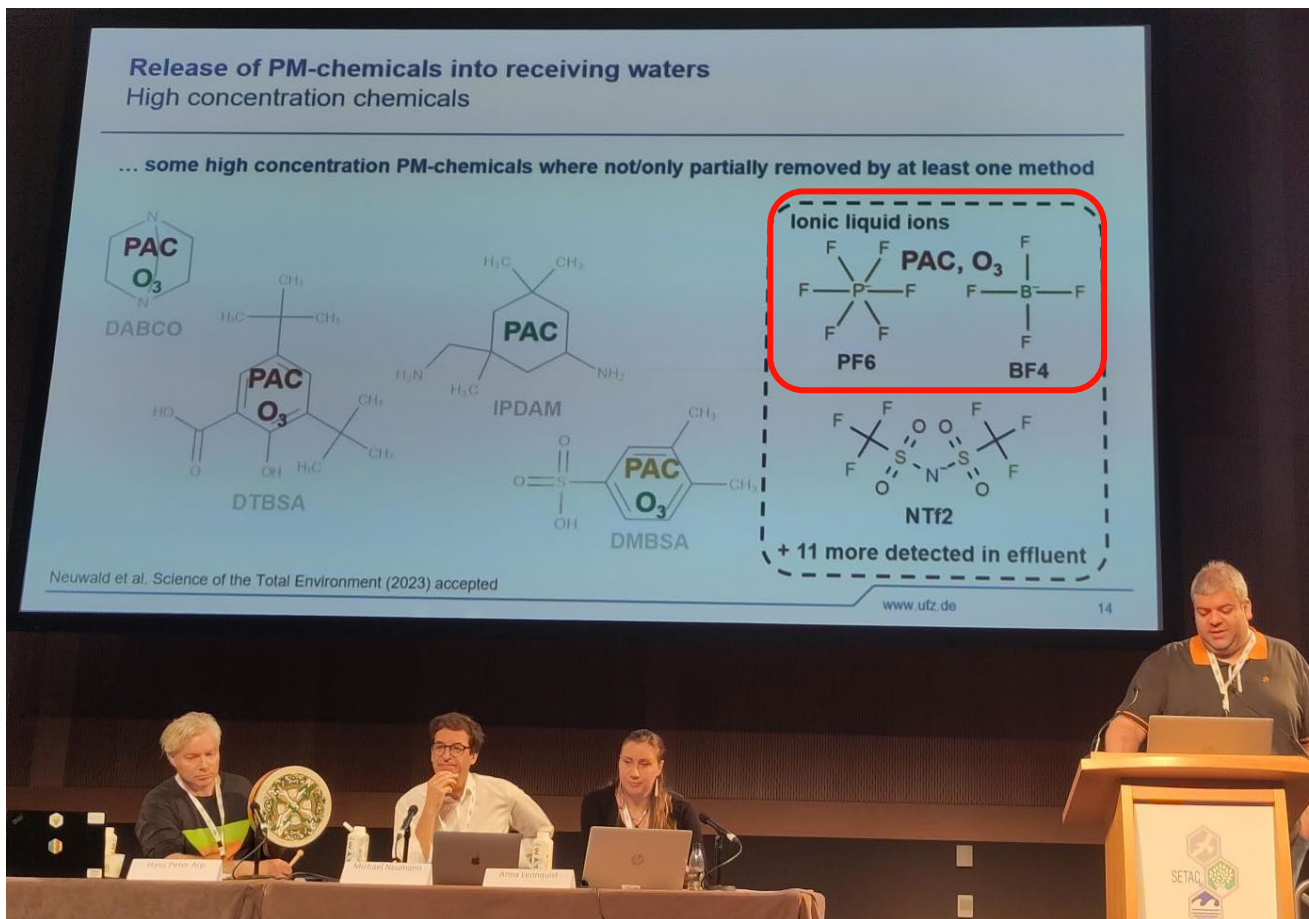


Consolidated reference numbers

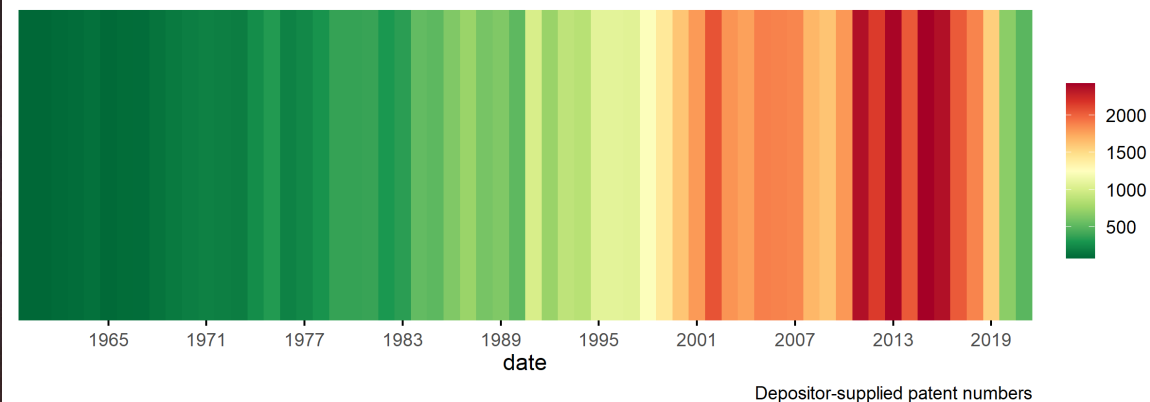
<https://gitlab.lcsb.uni.lu/eci/chemicalstripes>

The Chemical Stripes and Patent Data

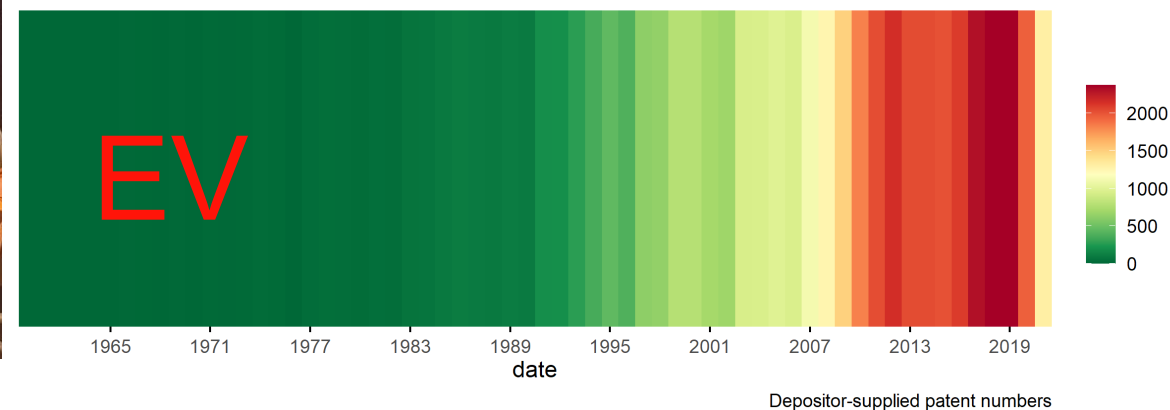
(live stripe calculations during SETAC 2023)



Chemical Stripes for Tetrafluoroboric acid



Chemical Stripes for Lithium tetrafluoroborate

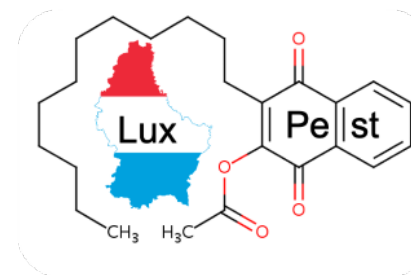
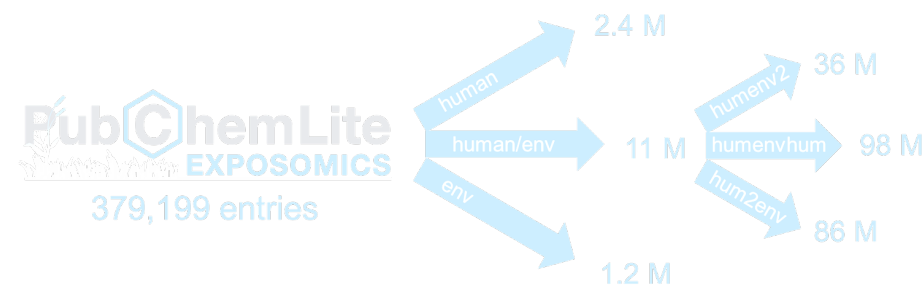
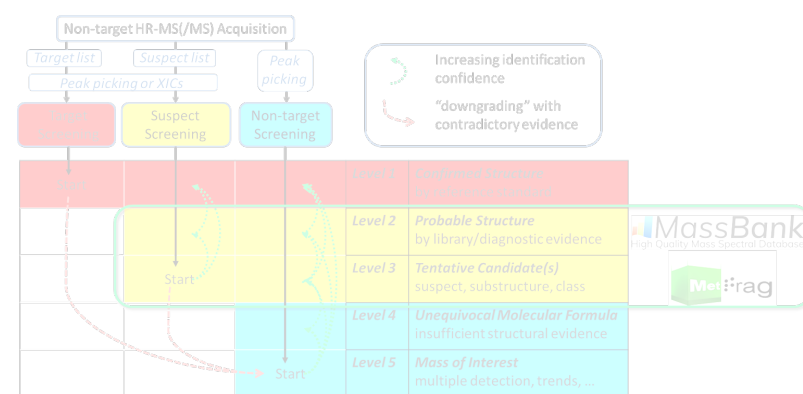


Left: Neuwald et al, STOTEN, DOI: [10.1016/j.scitotenv.2023.163921](https://doi.org/10.1016/j.scitotenv.2023.163921)

Photo of Daniel Zahn, UFZ at SETAC Europe, 30 April – 4 May, 2023. Image reused with permission

Outline

- Identifying Chemicals with NT-HRMS
 - Open Resources for Identification
 - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases & Workflows
 - Defining Relevant Chemical Space
 - Transformations: Known & Predicted
- No package too small?
- Applications in Environmental Studies

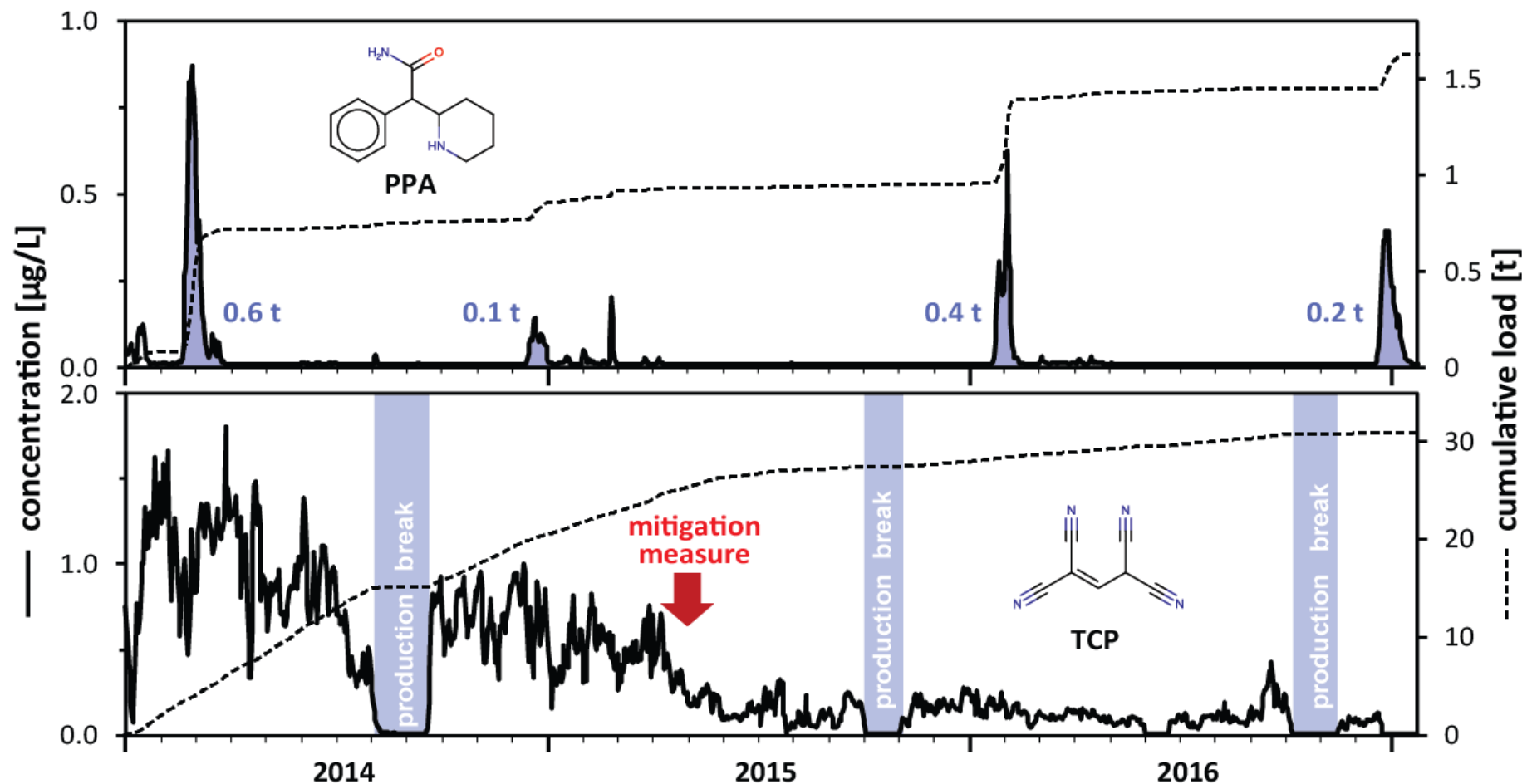


Examples: Real Time Monitoring of the Rhine River

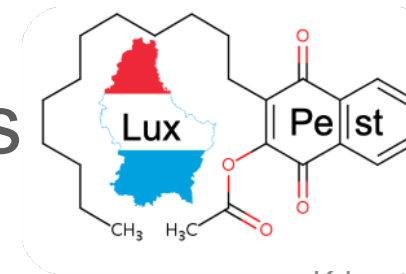
Previously unknown chemicals detected due to “stand-out” patterns



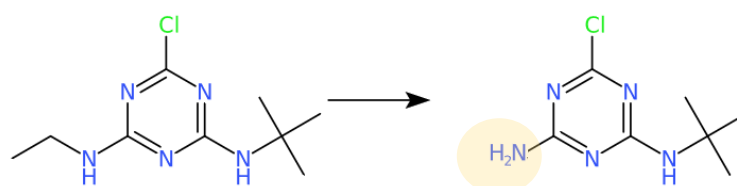
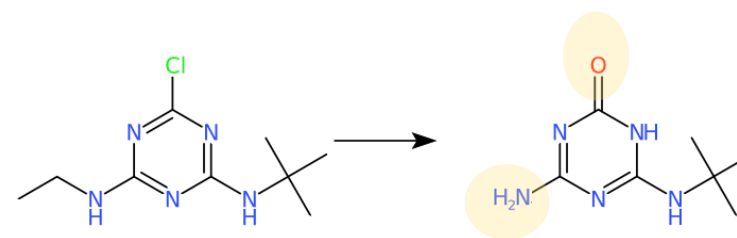
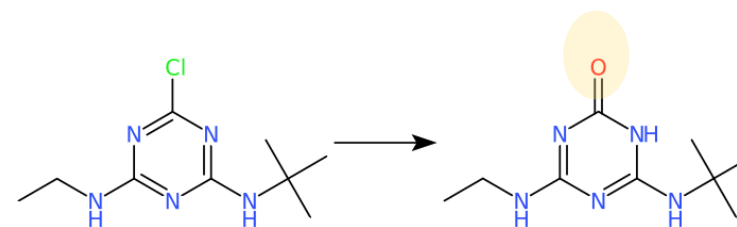
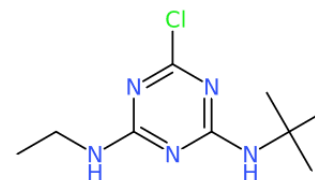
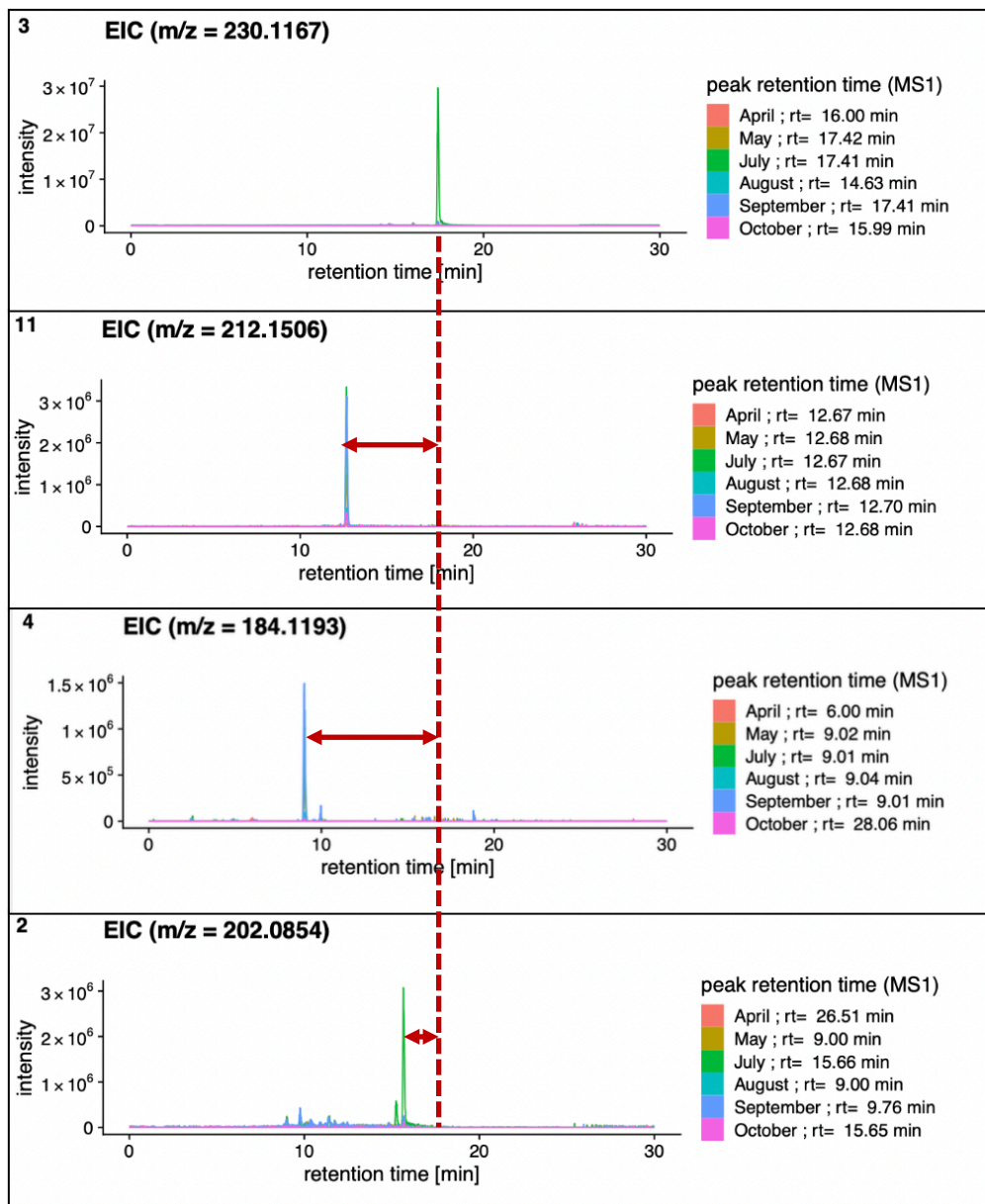
eawag
aquatic research



Examples: LuxPest and Transformation Products



Krier et al (2022). DOI:
[10.1016/j.envint.2021.106885](https://doi.org/10.1016/j.envint.2021.106885)

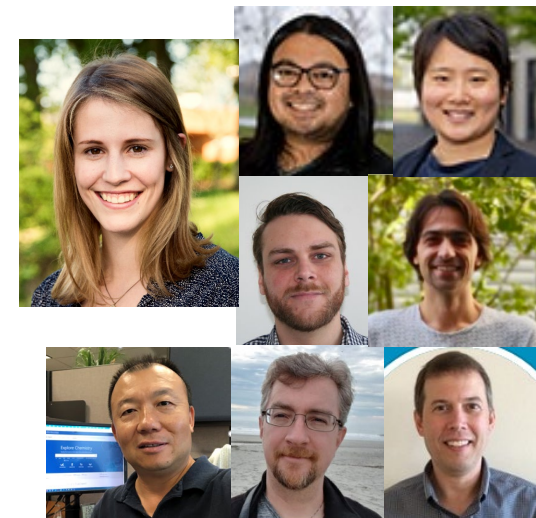
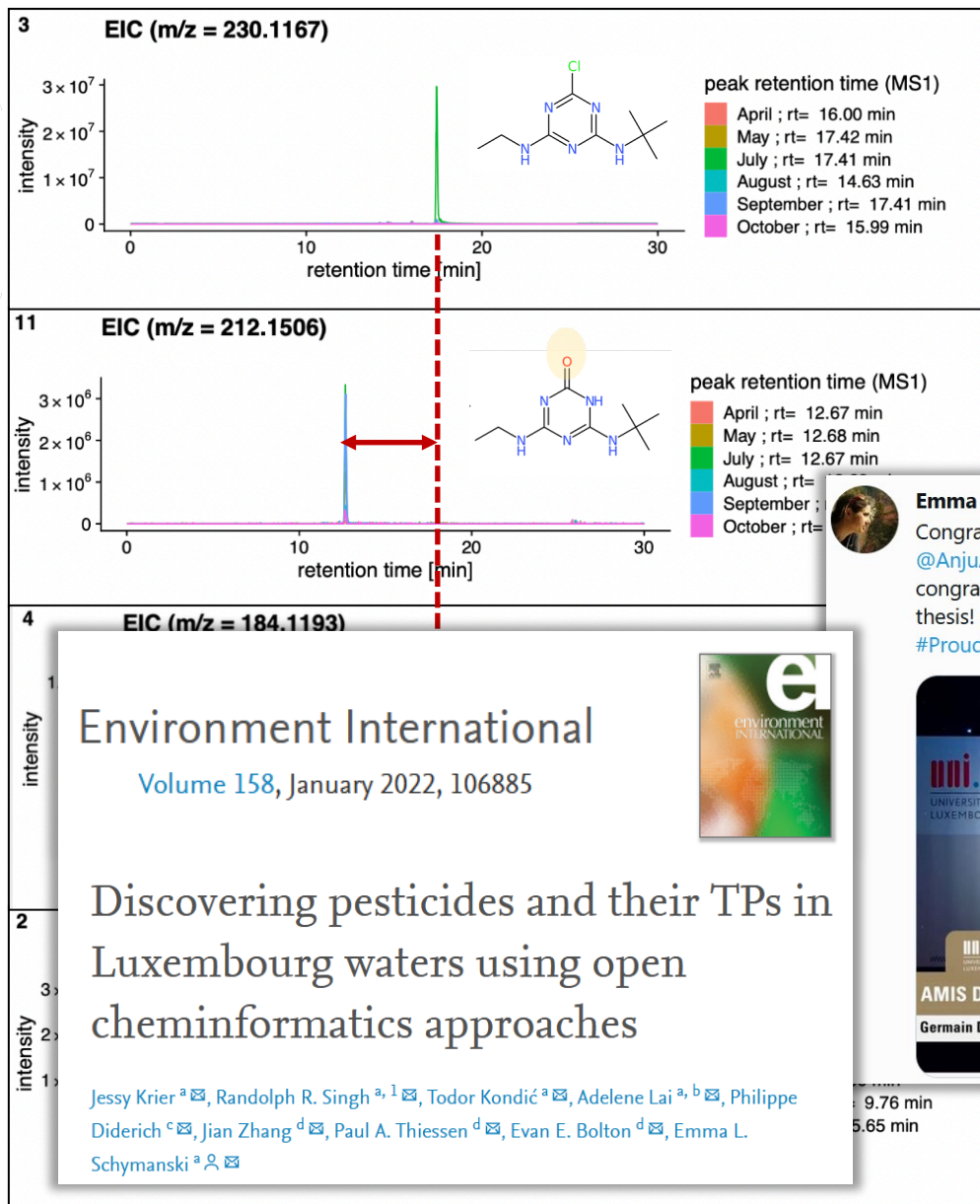
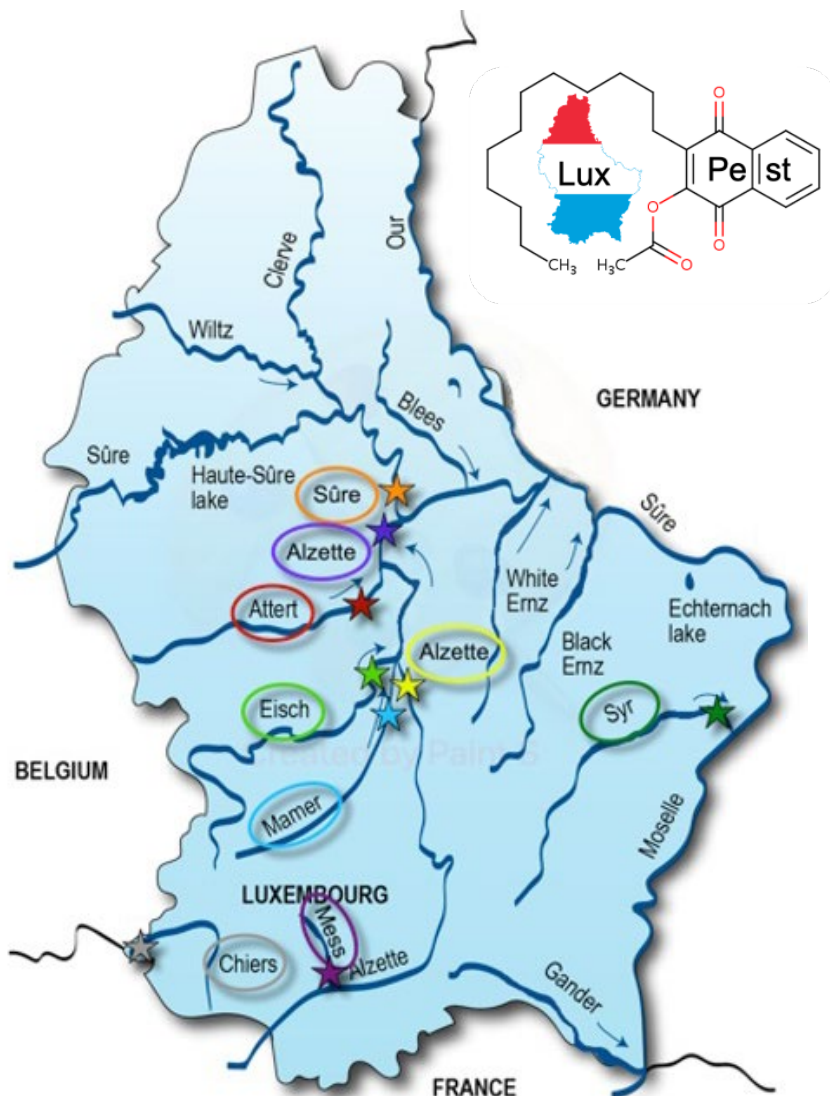


Pesticides & TPs in Luxembourg

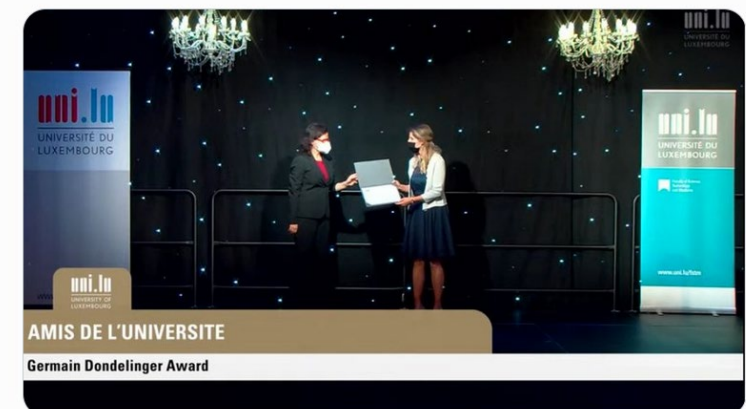


THE GOVERNMENT
OF THE GRAND DUCHY OF LUXEMBOURG
Ministry of the Environment, Climate
and Sustainable Development

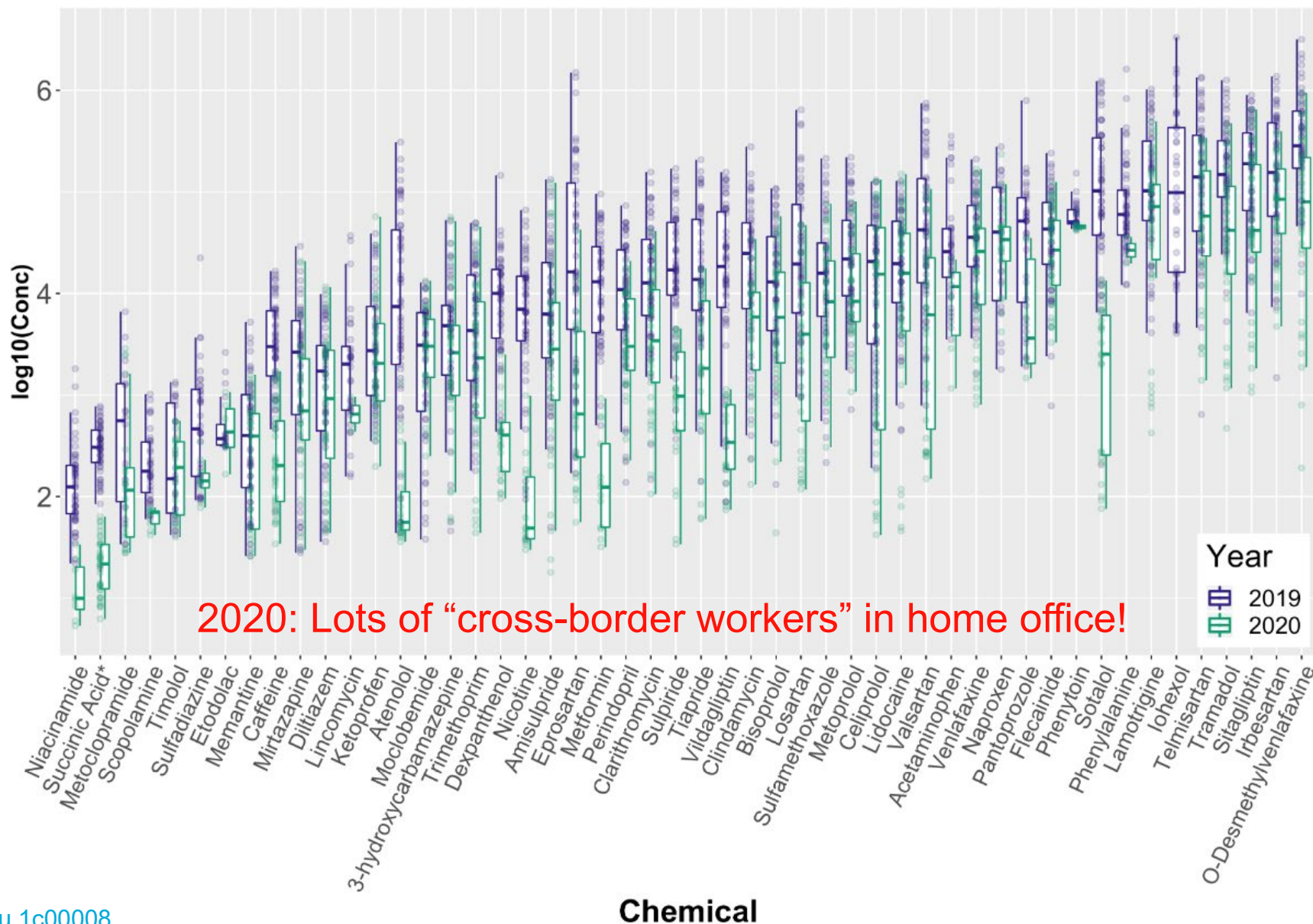
NIH National Library of Medicine
National Center for Biotechnology Information



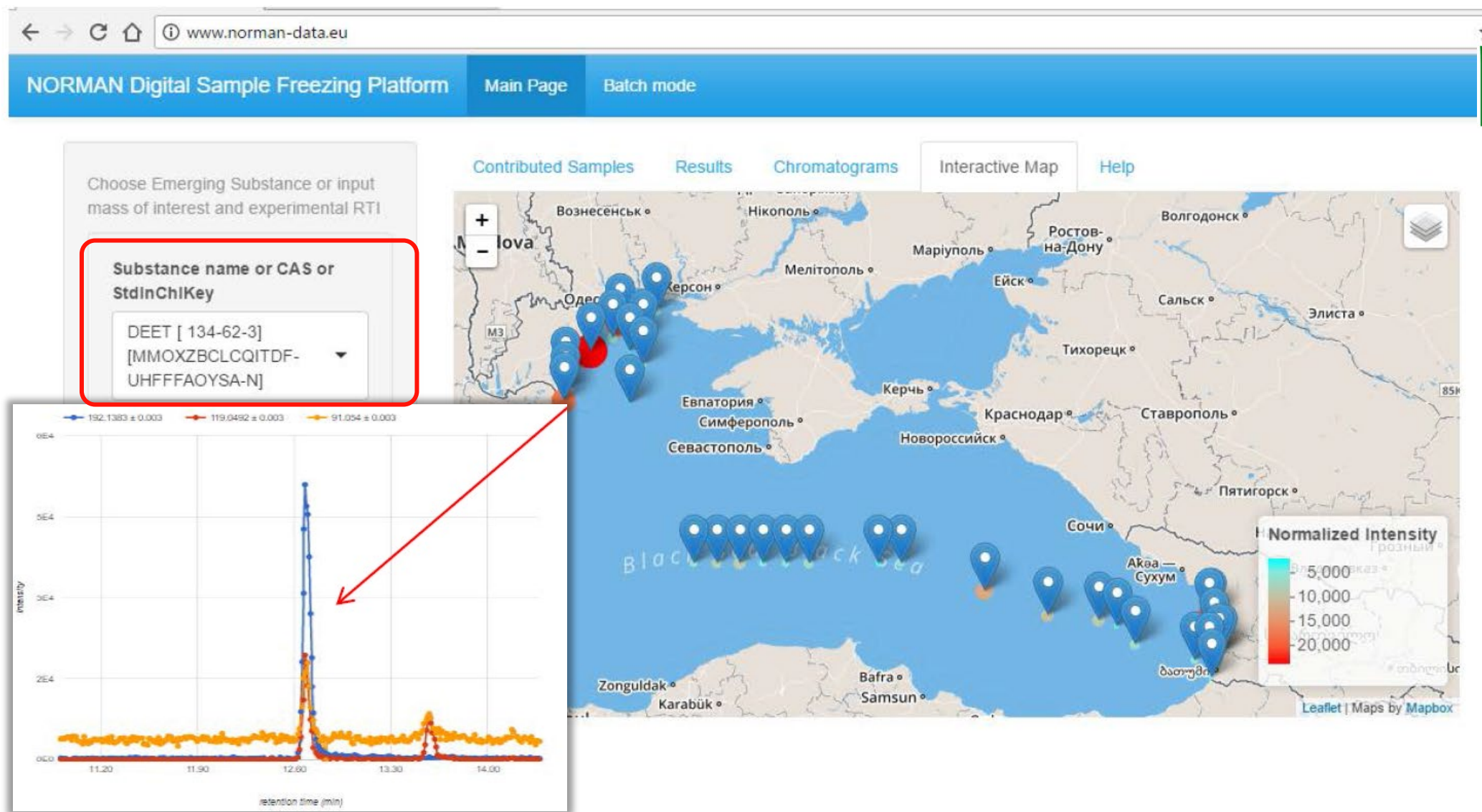
Emma Schymanski @ESchymanski · May 11
Congratulations to the #ECI LCSB @uni_lu #ClassOf2020 including @krije_ @AnjuAnjuraj15 @HibaMohamedTaha and @NarayananMira - special congrats to @krije_ for the Germain Dondelinger Award for her masters thesis! Our first graduates! #ProudPI @Fnrlux



Examples: LuxPharma – 2019 versus 2020 & COVID?



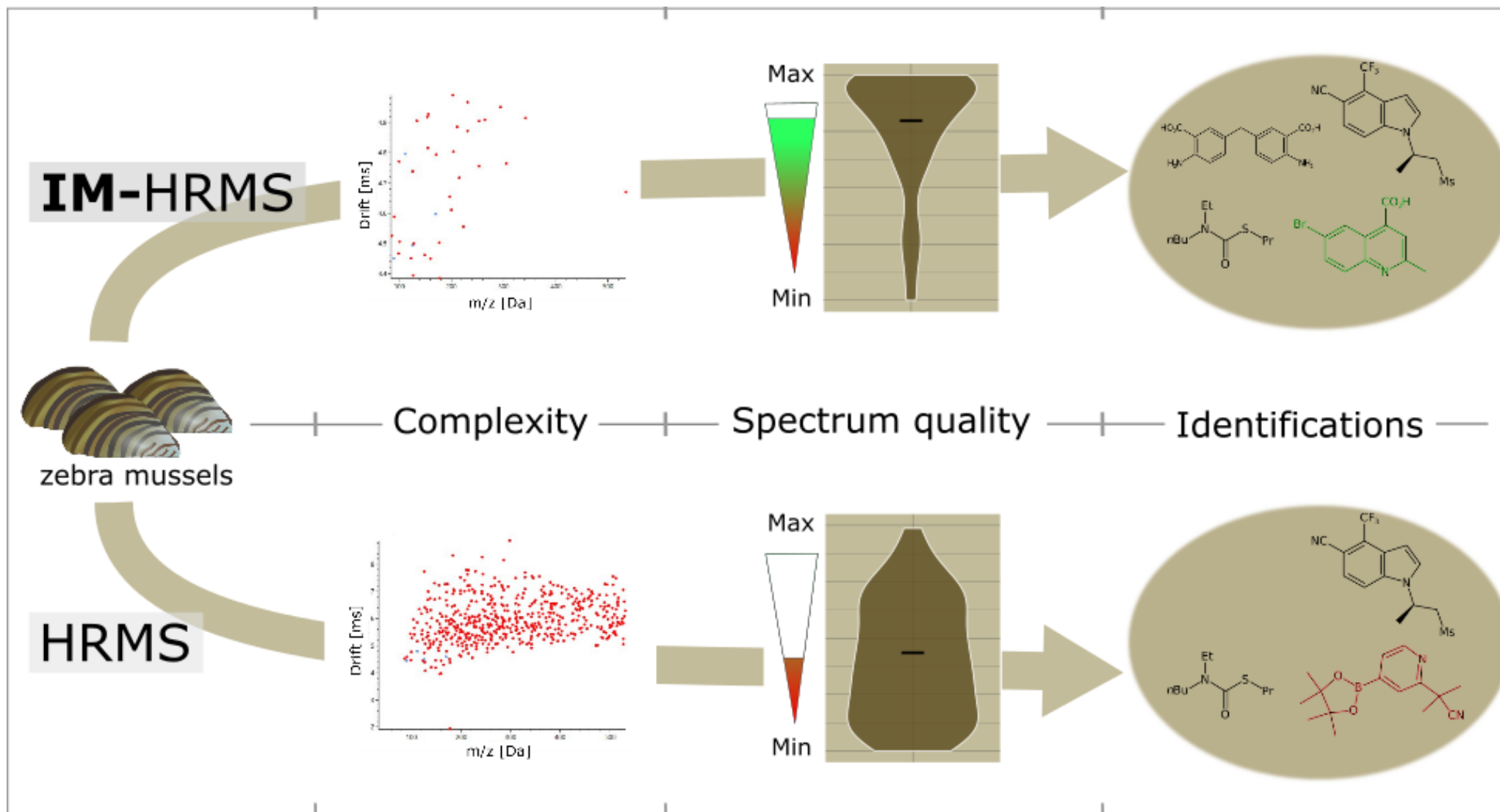
NORMAN DS – Digital Sample Freezing Platform



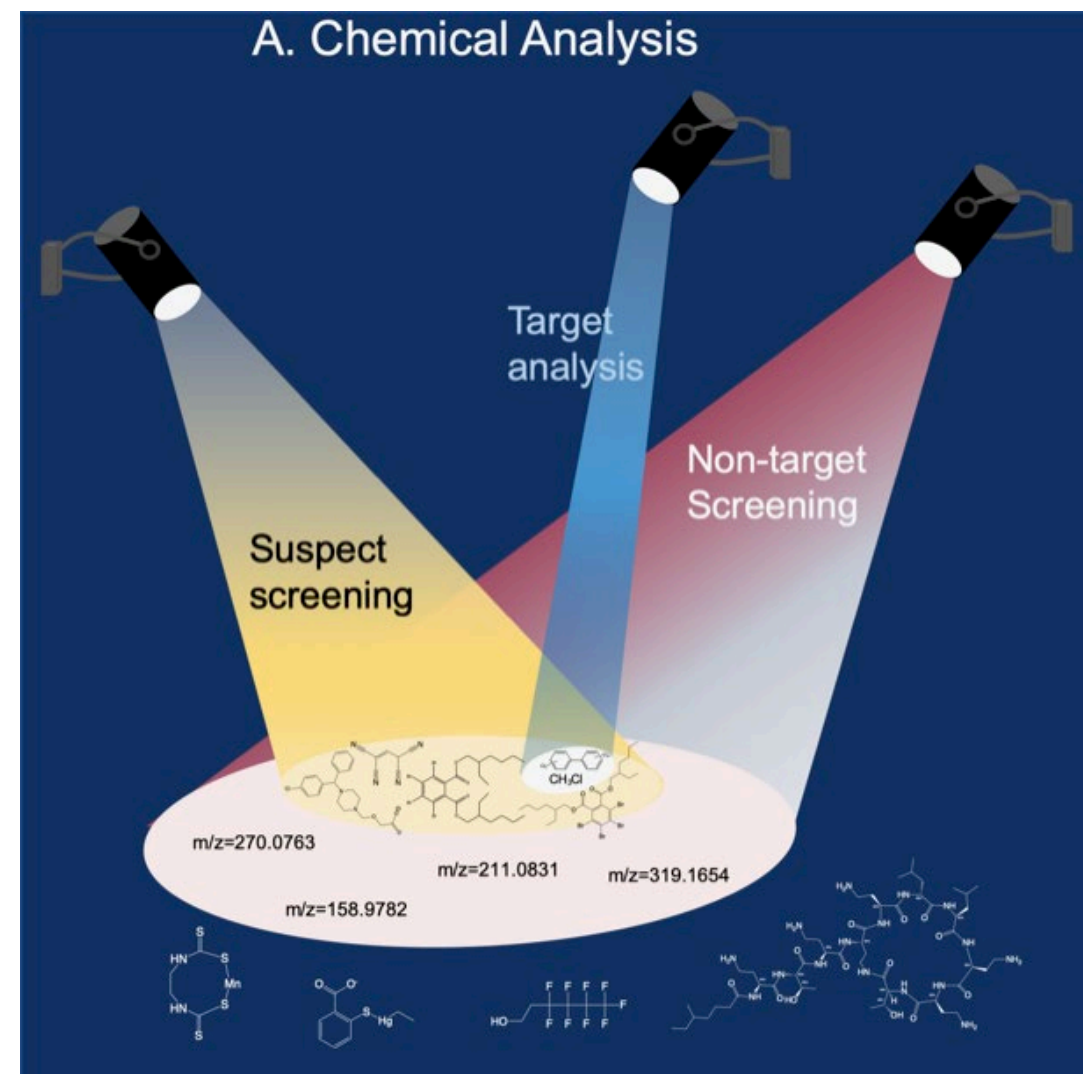
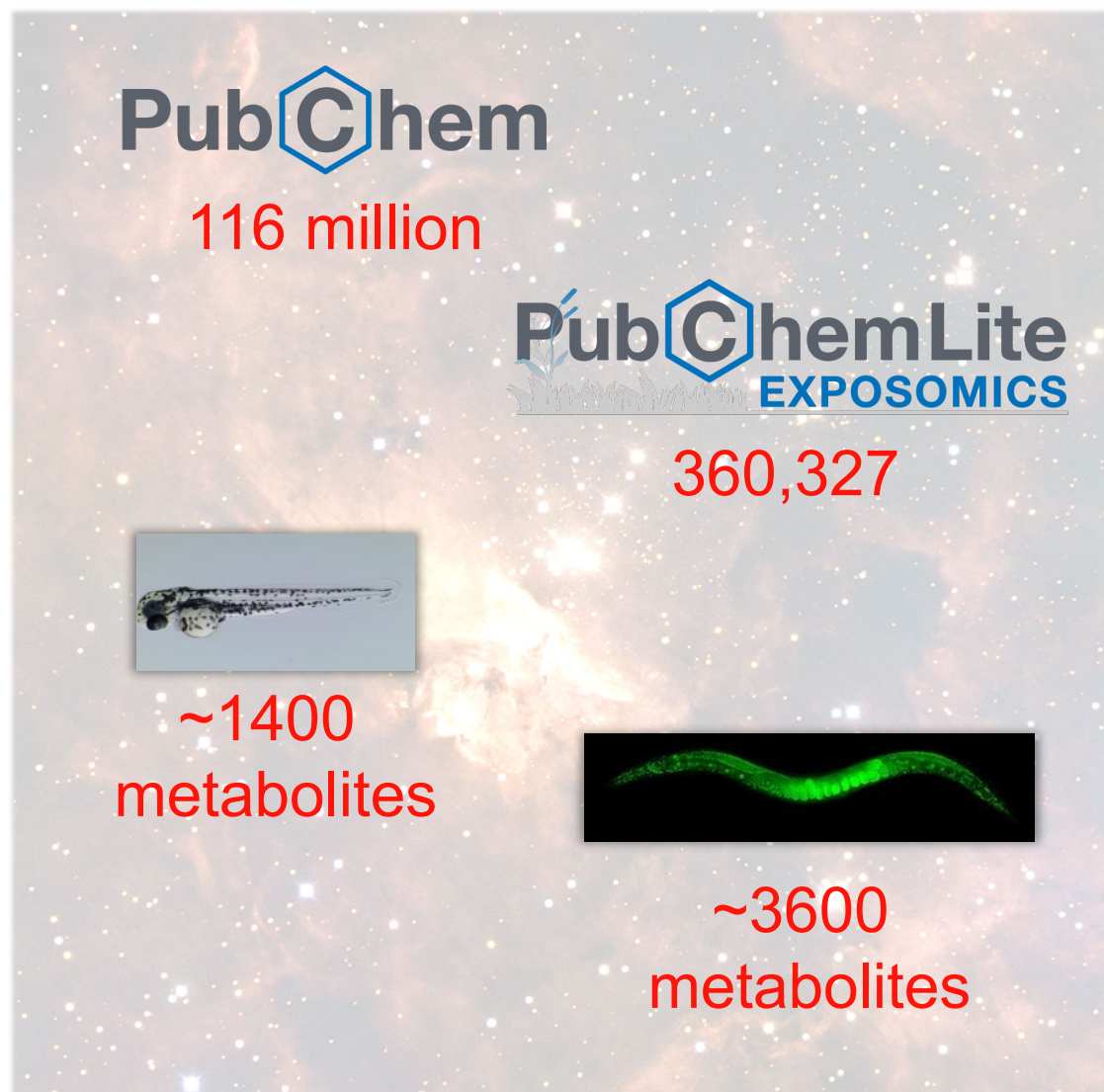
CCS & MusselFiesta: Cleaner spectra; better identification



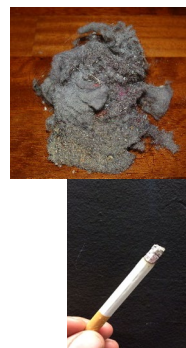
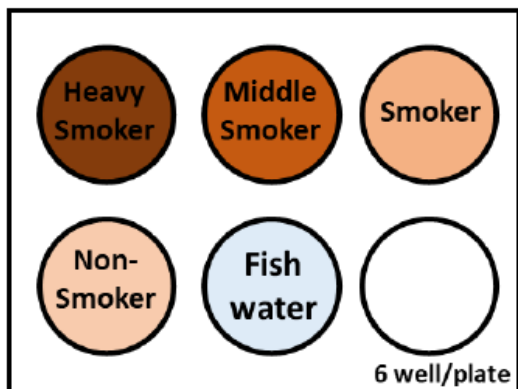
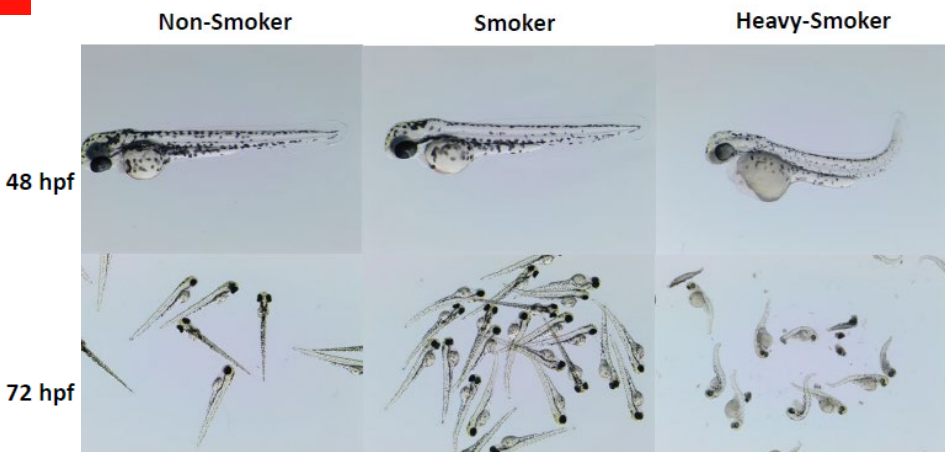
@MengerAcademic



Finally ... Do we always need the entire chemical space?



Organism-specific Metabolomics?



PubChem **Danio rerio (zebrafish) (Taxonomy)**

2-Phosphoglyceric acid [WikiPathways:WP1356](#)

CONTENTS

- Title and Summary
- 1 Names and Identifiers
- 2 Related Taxonomies
- 3 Chemicals and Bioactivities
 - 3.1 Whole-Organism Bioactivities
 - 3.2 Pathway Compounds
 - 3.3 Metabolites
 - 3.4 Natural Products
 - 3.5 Glycans

423 items

Search SORT BY Compound CID - Increasing

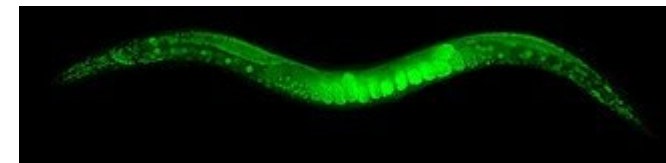
Structure	Compound CID	Compound	Evidence IDs	Data Source
	59	2-Phosphoglyceric acid	DOI:10.1093/nar/gkaa1024	ECI Group, LCSB, University of Luxembourg

Taxonomy: Kim *et al.* (2022) DOI: [10.1016/j.jmb.2022.167514](https://doi.org/10.1016/j.jmb.2022.167514)

<https://pubchem.ncbi.nlm.nih.gov/taxonomy/7955#section=Metabolites>

Image source: Carla Merino, Noelia Ramirez (URV). Merino *et al* (in prep), Torres *et al* (in prep).

Organism-specific Metabolomics?



zenodo Search Upload Communities

September 9, 2019 Dataset Open Access

WormJam Metabolites Local CSV for MetFrag

Witting, Michael; Schymanski, Emma

This is a local CSV file of WormJam (https://msbi.ipb-halle.de/MetFrag/).

The text file provided by Michael (also headers for MetFrag import).

This CSV file is for users wanting to use MetFrag online; please use the file in

Menu

C_elegans_metabolites.Rmd 9.79 KB

```
title: "Finding _C. elegans_ Metabolites on PubChem"
author: "Emma SCHYMANSKI"
date: "19/01/2022"
output: pdf_document
```

knitr::opts_chunk\$set(echo = TRUE)

Background

WormJam

In 2019, Michael Witting provided a copy of 1203 WormJam (Worm Jamboree) metabolites from the multi-author WormJam study (DOI:10.1080/21624054.2017.1373939). This was turned into a MetFrag database and put on Zenodo by the Environmental Cheminformatics group (ECI) at LCSB (DOI:10.5281/zenodo.3403364) for integration into MetFrag Web, and for download for MetFragCL users.

PubChem Caenorhabditis elegans (Taxonomy)

ECI Group, LCSB, University of Luxembourg

DOI:10.1080/21624054.2017.1373939

4-Methyl-2-oxopentanoic acid

70

Page 1 of 233

ECI Group, LCSB, University of Luxembourg

3.4 Natural Products

712 items Download

Search SORT BY Compound CID - Increasing

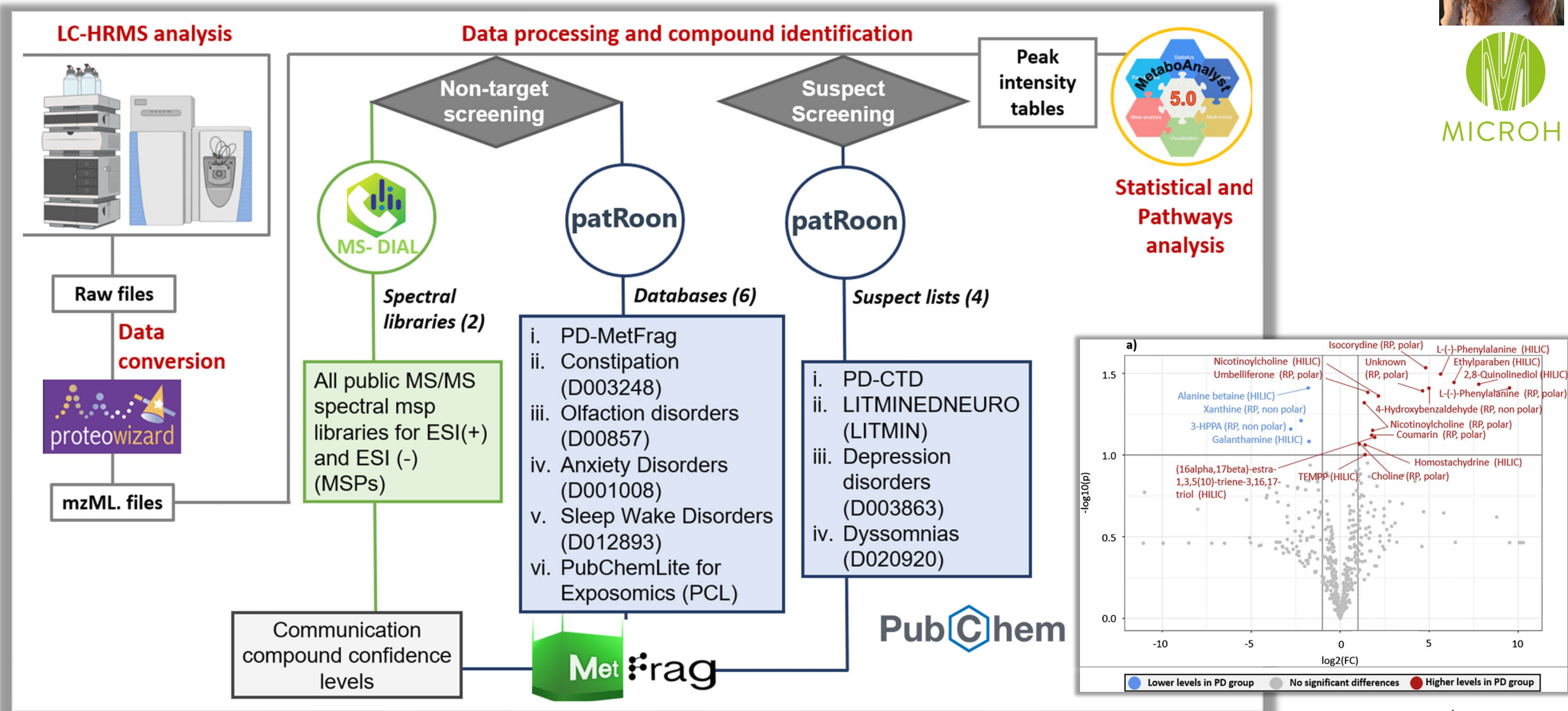
Structure	Compound CID	Compound	Evidence IDs	Data Source
	47	3-Methyl-2-oxovaleric acid	DOI:10.3389/FMOLB.2018.00096	LOTUS - the natural products occurrence database

Topic	Documentation
Taxonomy	C. elegans as PDF or RMarkdown



<https://pubchem.ncbi.nlm.nih.gov/taxonomy/6239#section=Metabolites>

Exposomics: Disease-specific Databases and Interpretation



Exposomics: Disease-specific Information in PubChem



PubChem Rotenone (Compound)

14 Associated Disorders and Diseases

83 items [Download](#)

Search

Disease	Evidence Type	Evidence PMID
Motor Disorders	marker/mechanism	26770656
		27016191
		29649621

PubChem Rotenone (Compound)

15.10 Chemical-Disease Co-Occurrences in Literature

Showing 3 of 25 [View More Co-Occurrence and Evidence Data](#) [Download](#)

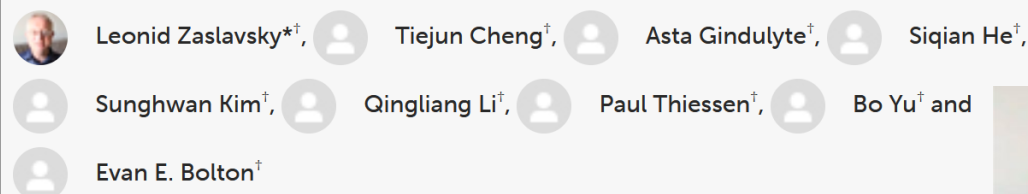
Disease	Selected evidence
Parkinson Disease	1,590 articles View All
	Rotenone-Induced Model of Parkinson's Disease: Beyond Mitochondrial Complex I Inhibition PMID 36593435; DOI 10.1007/s12035-022-03193-8; Molecular neurobiology 2023 Jan; ?(?): (Review Article) Name matches: parkinson's disease <i>rotenone</i>
	HR LC-MS/MS metabolomic profiling of Yucca aloifolia fruit and the potential neuroprotective effect on rotenone-induced Parkinson's disease in rats PMID 36854038; DOI 10.1371/journal.pone.0282246; PloS one 2023; 18(2):e0282246 Name matches: parkinson's disease <i>rotenone</i>

Database

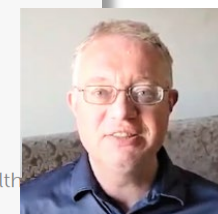
PubChem Protein, Gene, Pathway, and Taxonomy Data Collections: Bridging Biology and Chemistry through Target-Centric Views of PubChem Data

Sunghwan Kim[†], Tiejun Cheng[†], Siqian He, Paul A. Thiessen, Qingliang Li, Asta Gindulyte, Evan E. Bolton   

Discovering and Summarizing Relationships Between Chemicals, Genes, Proteins, and Diseases in PubChem

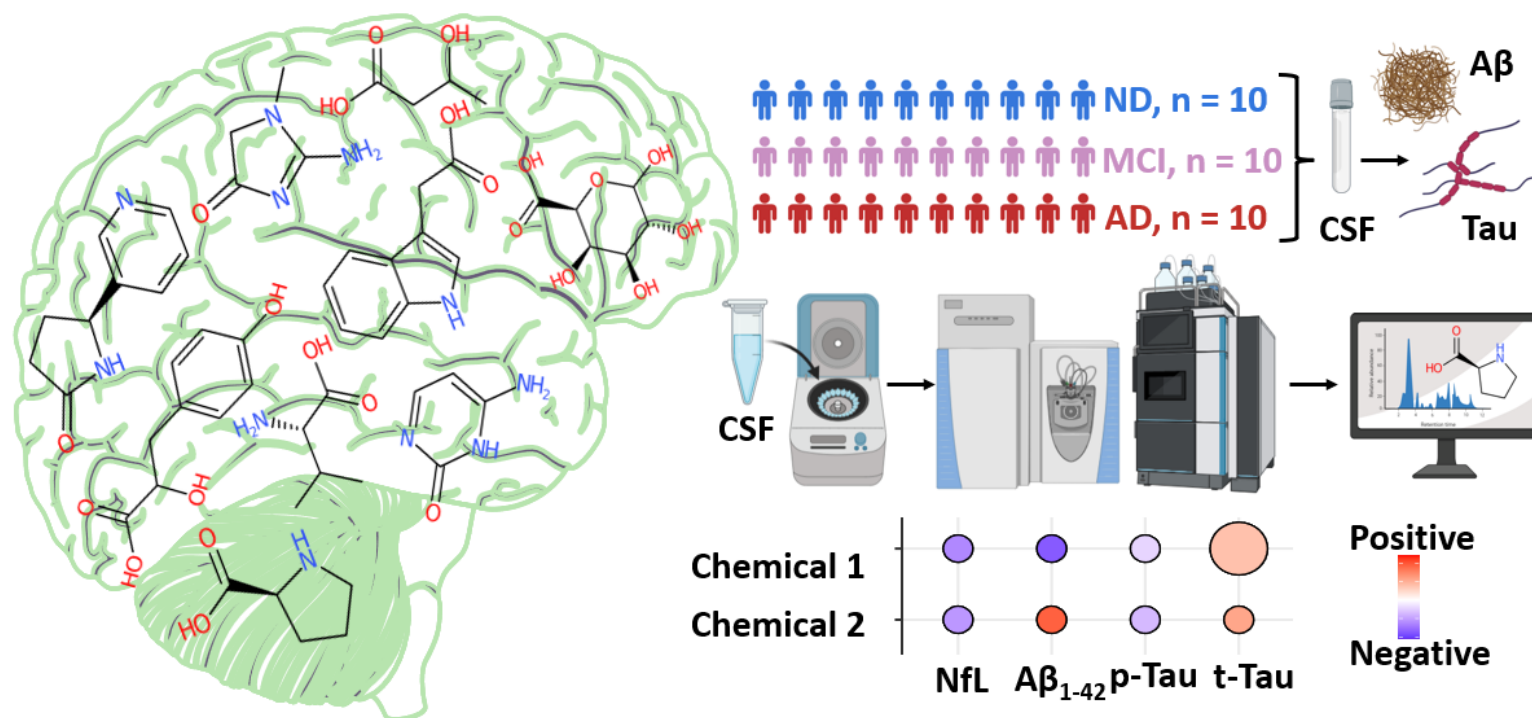


National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health
Bethesda, MD, United States



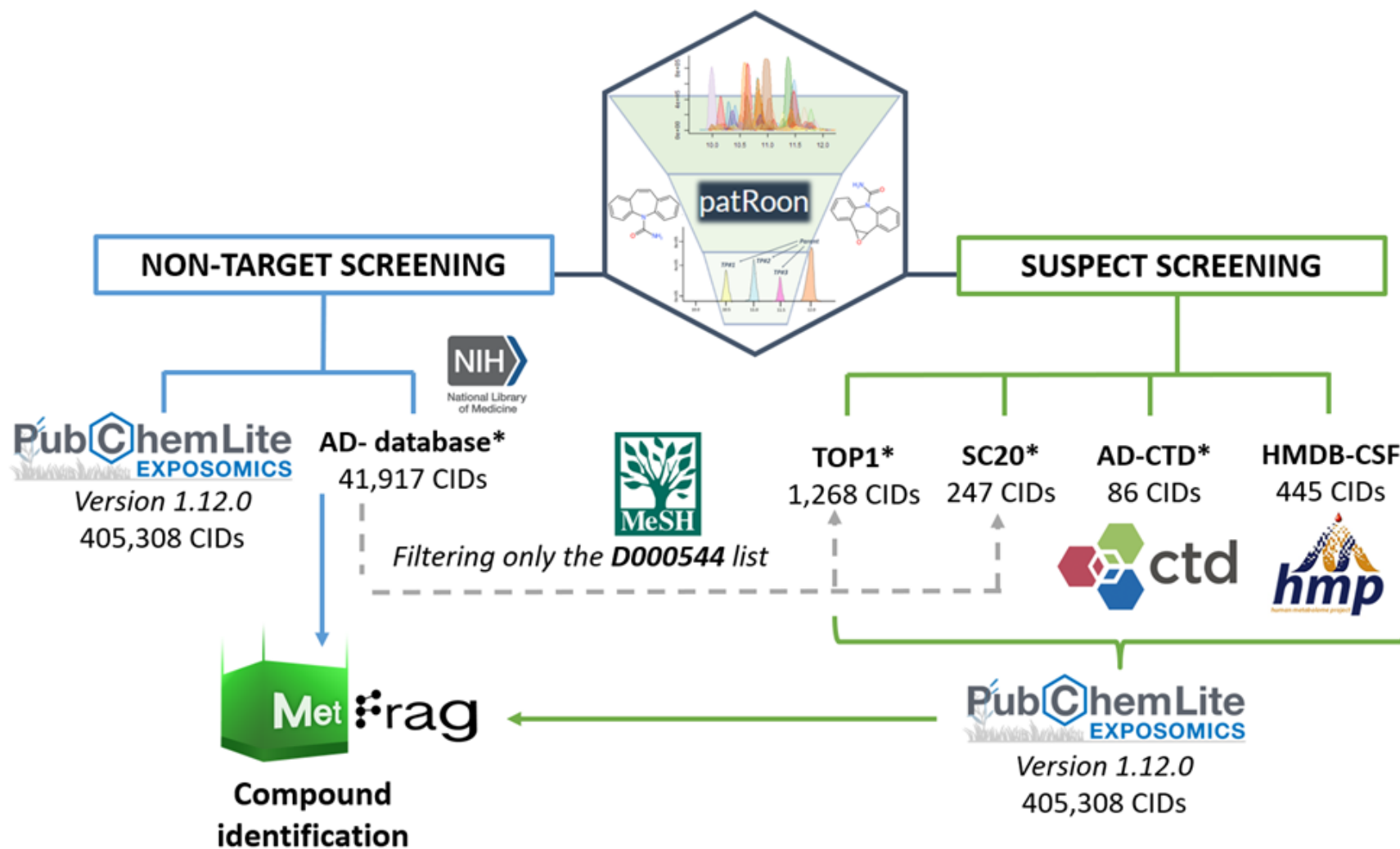
Exposomics: Disease-specific Databases and Interpretation

Information available varies widely depending on disease and sample



Exposomics: Disease-specific Databases and Interpretation

Information available varies widely depending on disease and sample

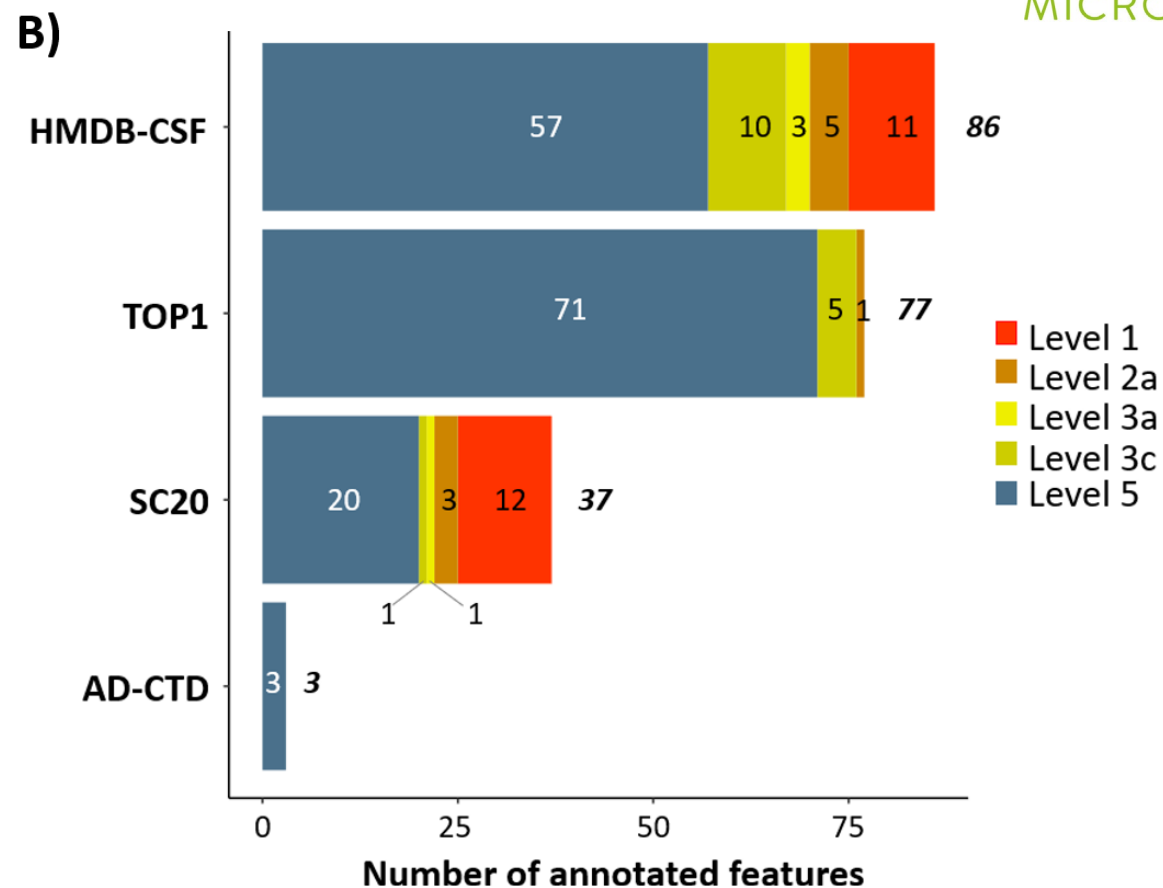
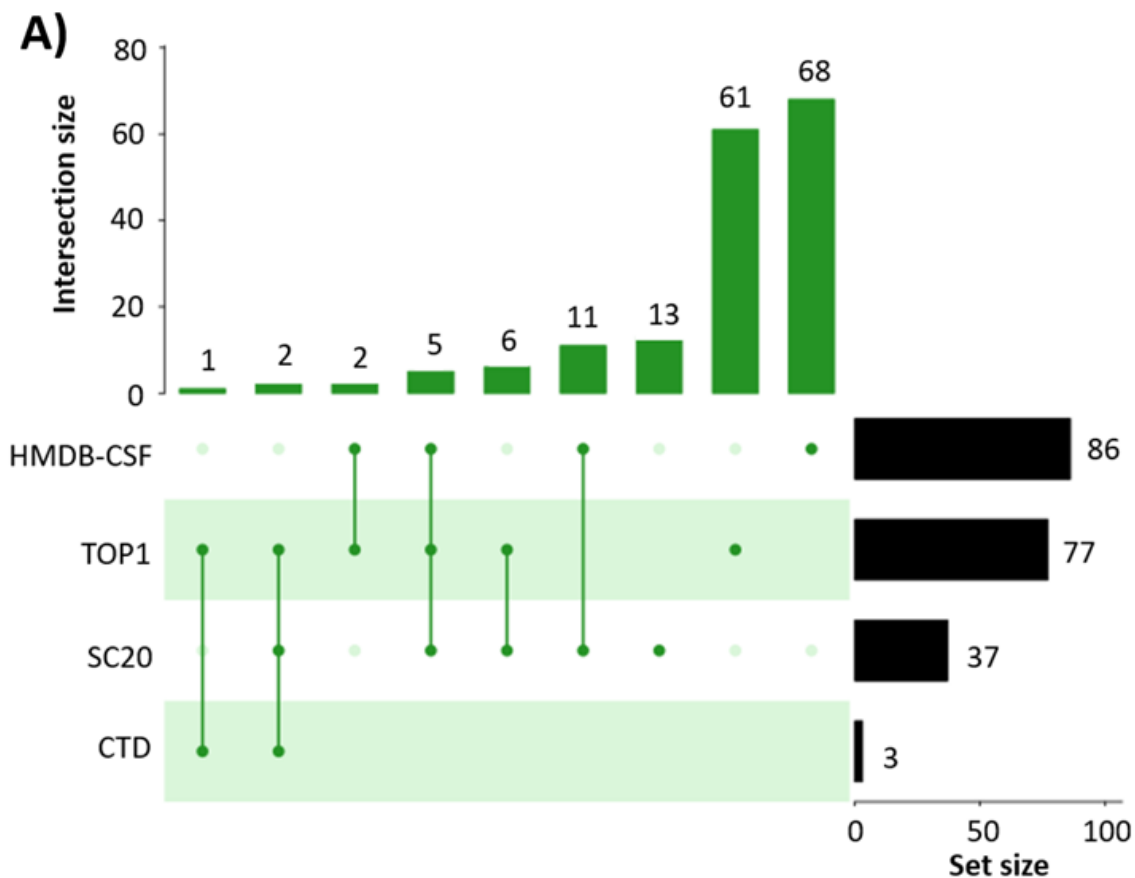


Exposomics: Disease-specific Databases and Interpretation

Information available varies widely depending on disease and sample

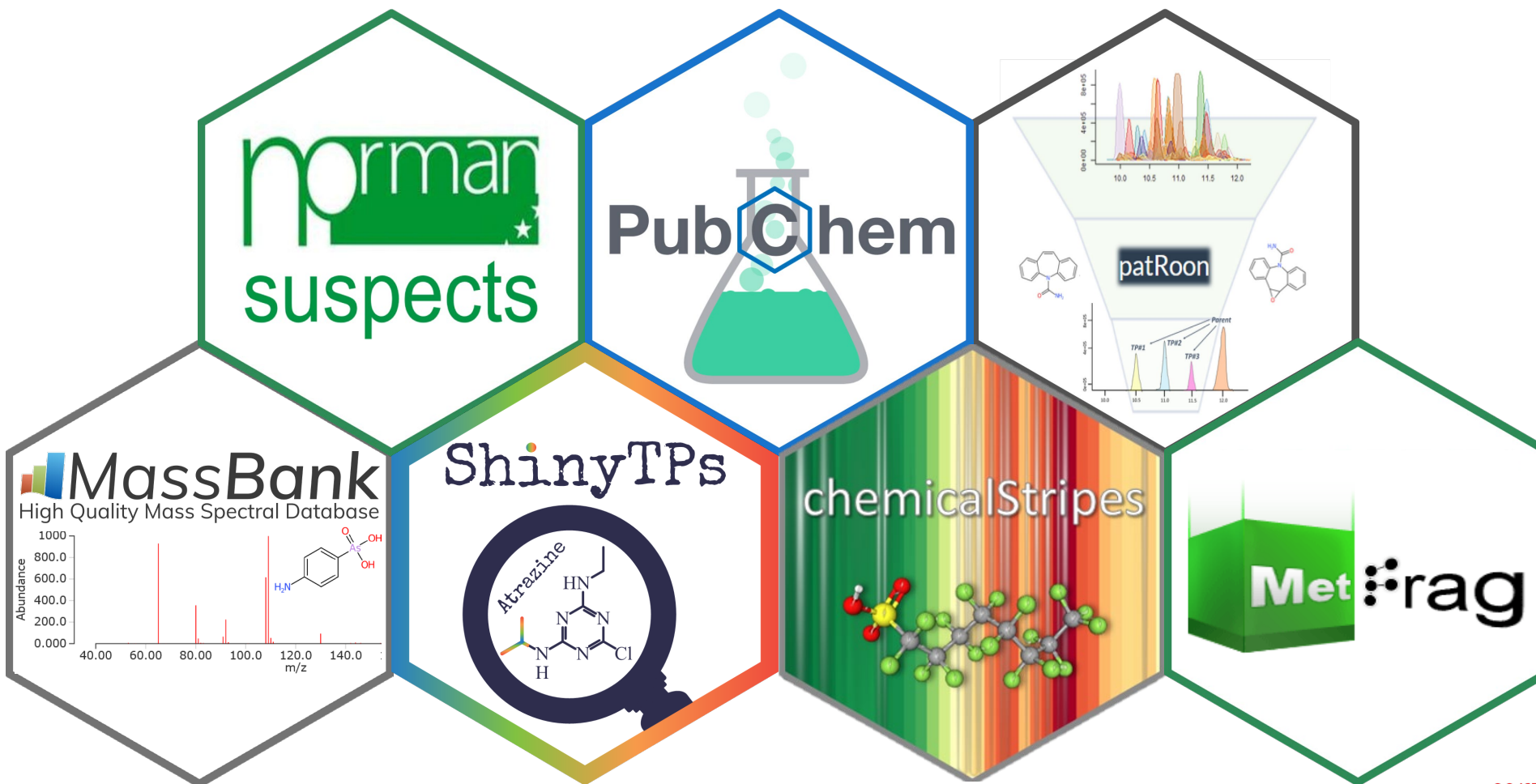


MICROH



“Take home” Messages

- Open, FAIR data supports NT-HRMS workflows



“Take home” Messages

- Dynamic, Open Collection:

- Expert knowledge to fill the gaps
- Annotation content is extremely powerful

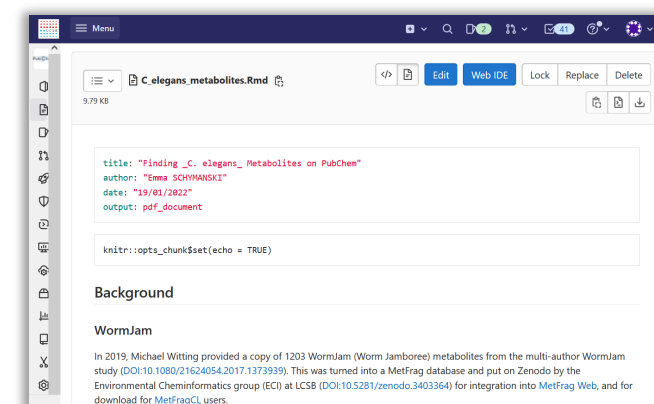
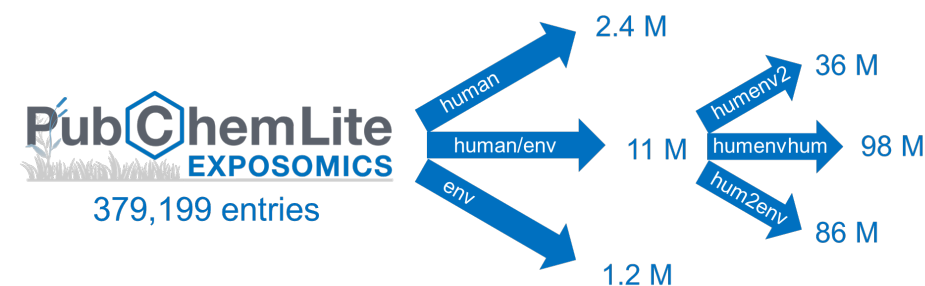
- Transformations and Metabolism

- More data needed to improve predictions

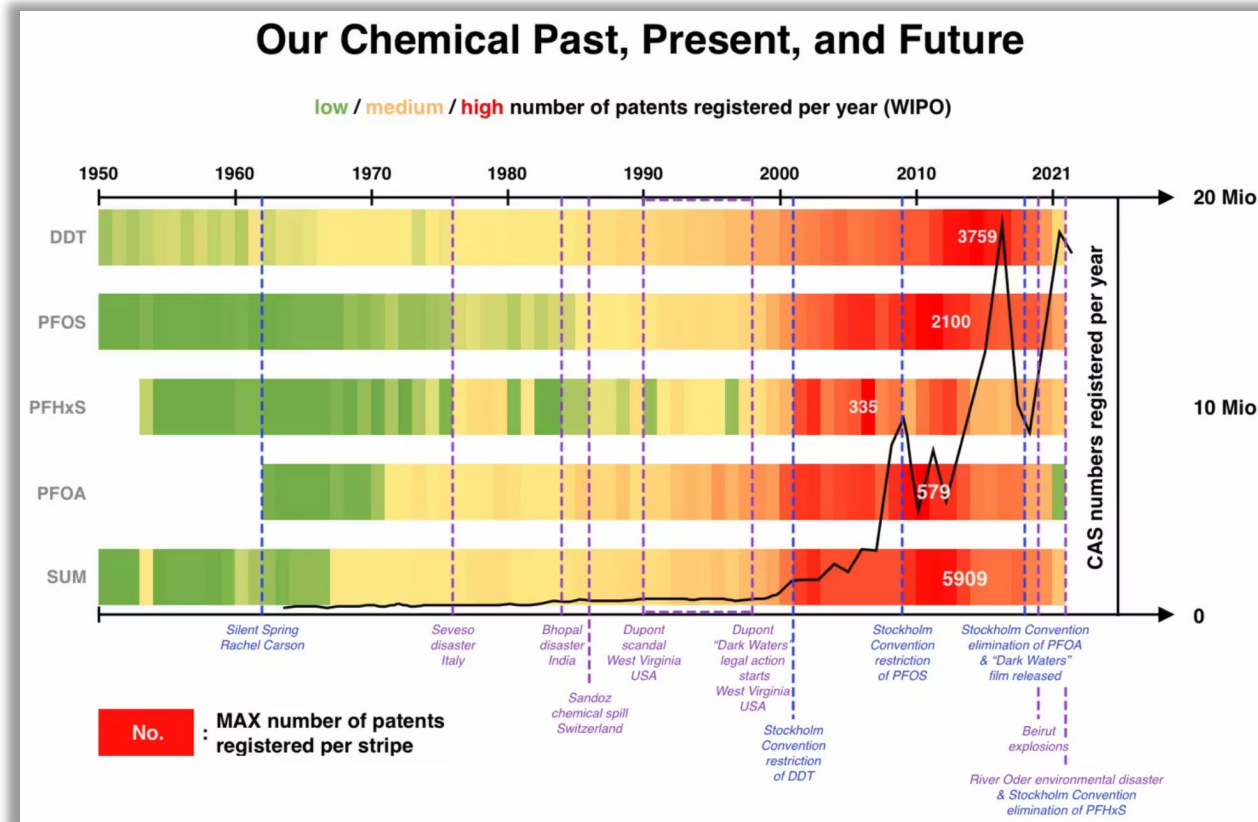
- Workflows available in patRoön

- Specific collections are easy to create for context

- Help contribute by adding your knowledge!



Today's Video / Soundtrack



Video: <https://vimeo.com/jpmlmusic/ourchemicalpastpresentandfuture>

Soundtrack: <https://soundcloud.com/jamieperera/our-chemical-past-present-and-future>

Acknowledgements!

Today's slides:

DOI: [10.5281/zenodo.8362117](https://doi.org/10.5281/zenodo.8362117)

Email: emma.schymanski@uni.lu

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

[@schymane@mstdn.social](https://mstdn.social/@schymane)

ORCID: [0000-0001-6868-8145](https://orcid.org/0000-0001-6868-8145)



PubChem

UNIVERSITY OF
AMSTERDAM

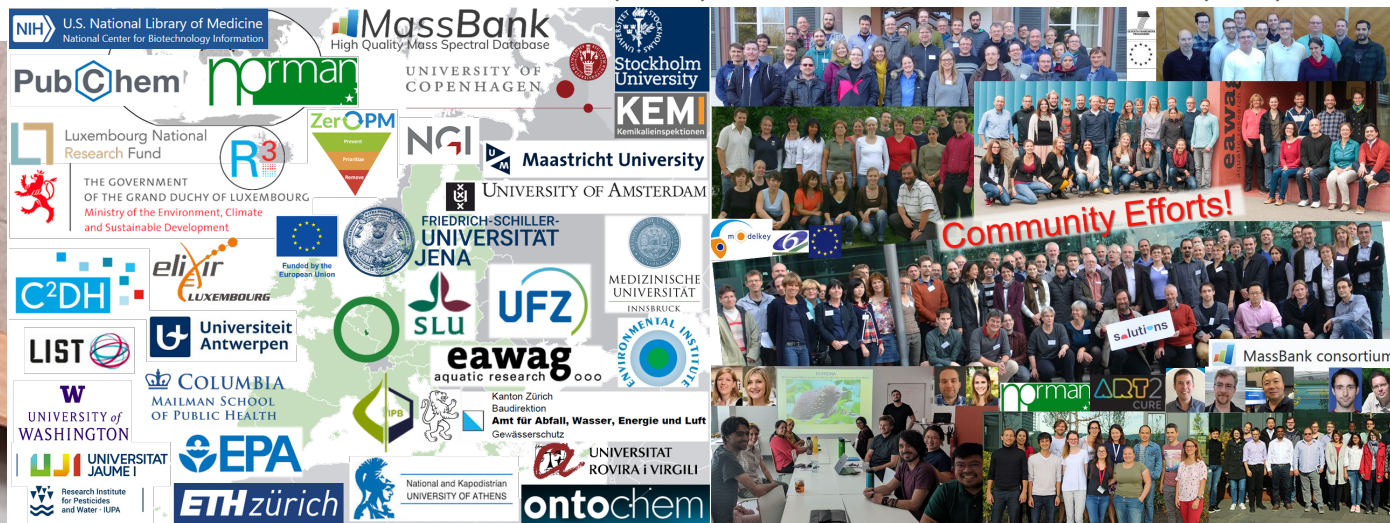


Evan Bolton, Jian (Jeff) Zhang,
Paul Thiessen, PubChem team

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



This work was supported in part by the National Center for
Biotechnology Information (NCBI) of the National Library of
Medicine (NLM), National Institutes of Health (NIH)



Funded by the
European Union



H2020:
101036756



Luxembourg
National
Research Fund

