Luxembourg National Research Fund
UMEÅ UNIVERSITY
UNIVERSITÉ DU LUXEMBOURG
Funded by the European Union

Open, Dynamic Databases, Workflows and Transformations to Support Environmental Studies

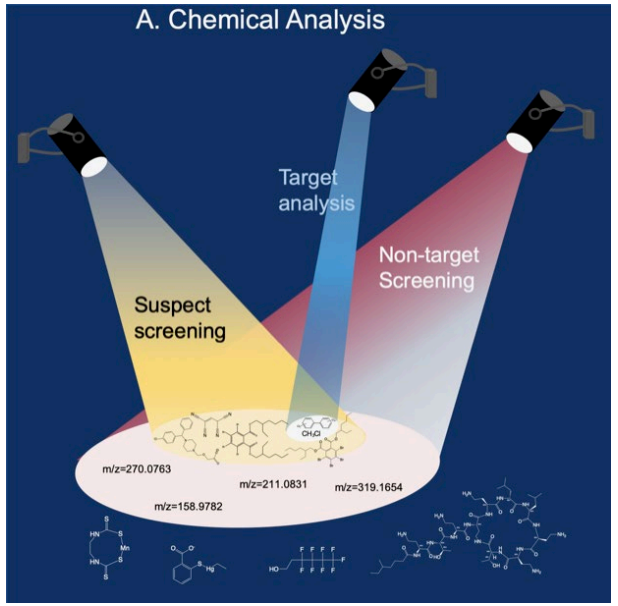


Image mod. from DOI: 10.1126/science.aay6636

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

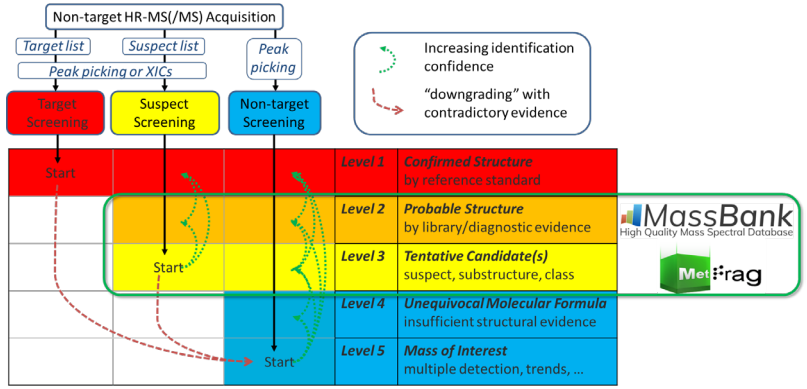
Today's slides zenodo

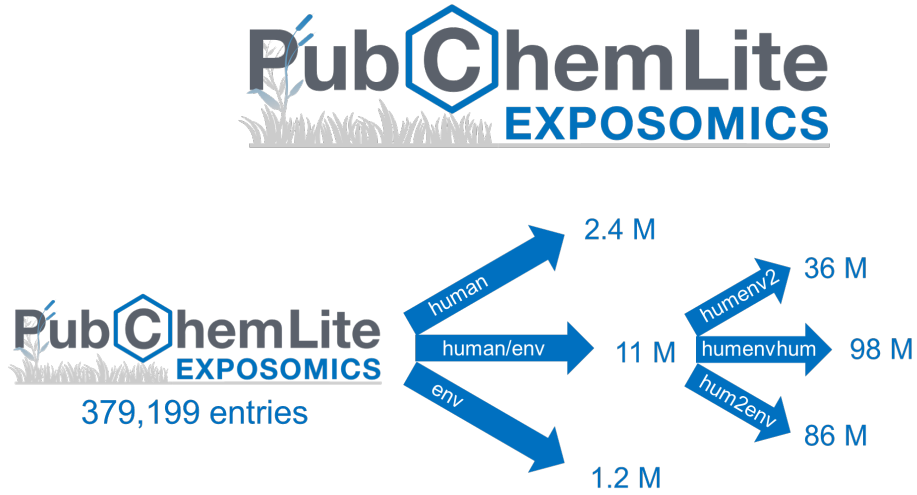
Environmental Chemistry and Biogeochemistry Seminar
Umeå University, 17 March 2023, Virtual Event

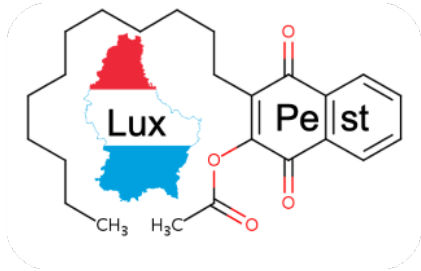
Slides available at DOI: 10.5281/zenodo.7744215

Outline

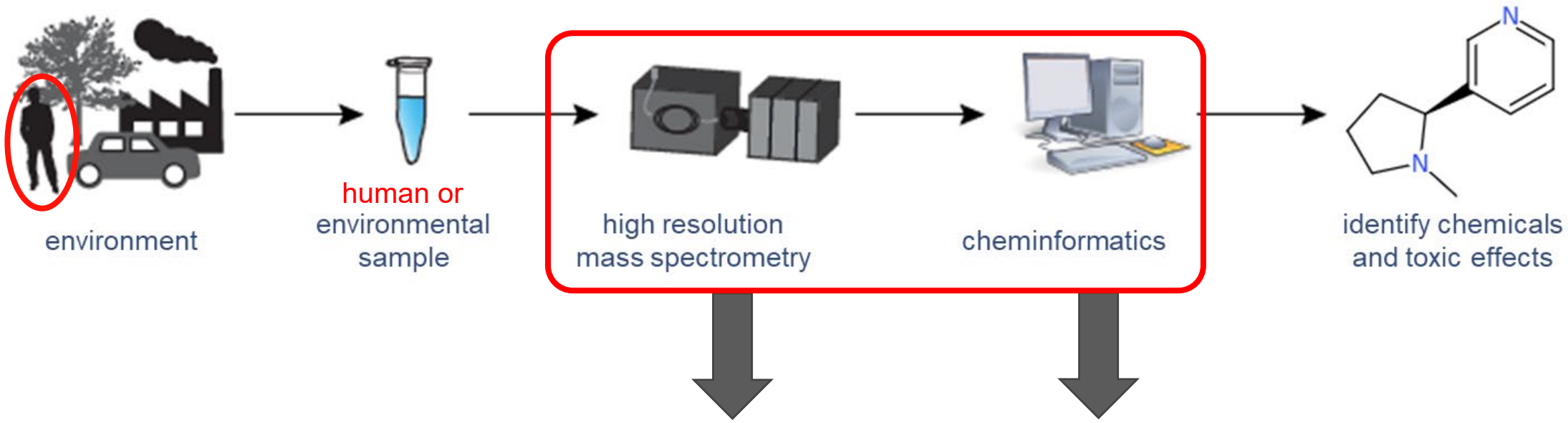
- Identifying Chemicals with NT-HRMS
  - Open Resources for Identification
  - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases
  - Defining Relevant Chemical Space
  - Transformations: Known & Predicted
- Applications in Environmental Studies
  - Including a little bit of CCS
- Perspectives





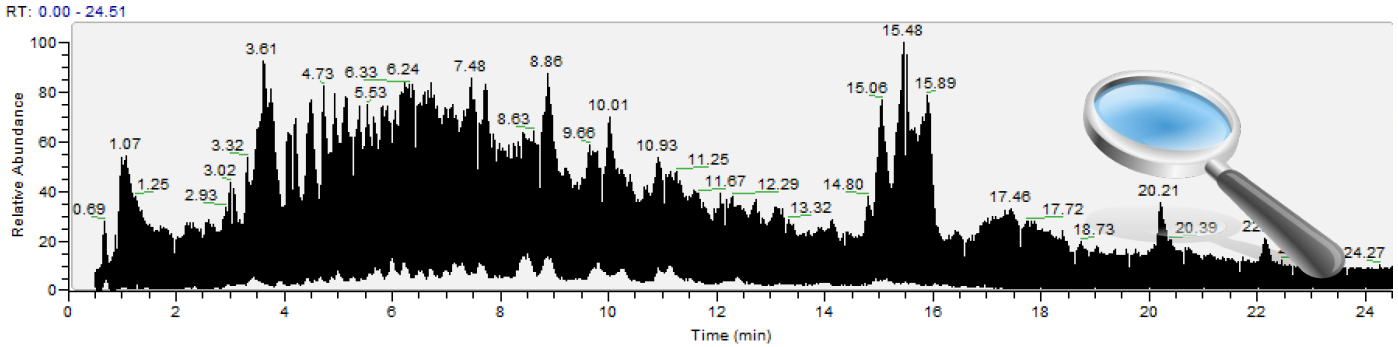


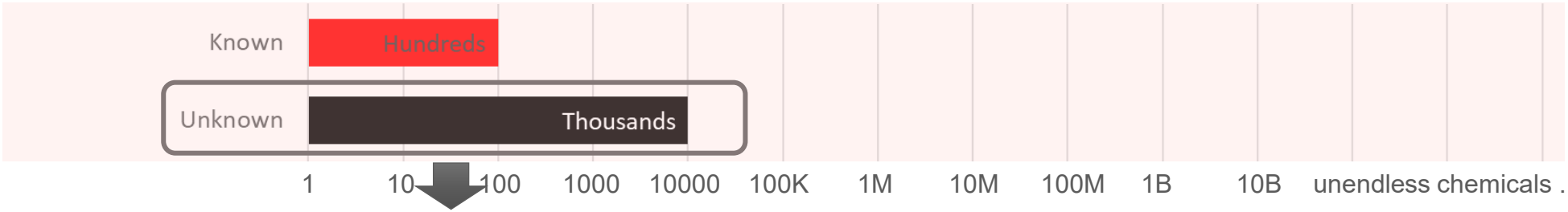
Environmental Cheminformatics & Non-target HR-MS



High resolution mass spectrometry

AND connecting chemical knowledge

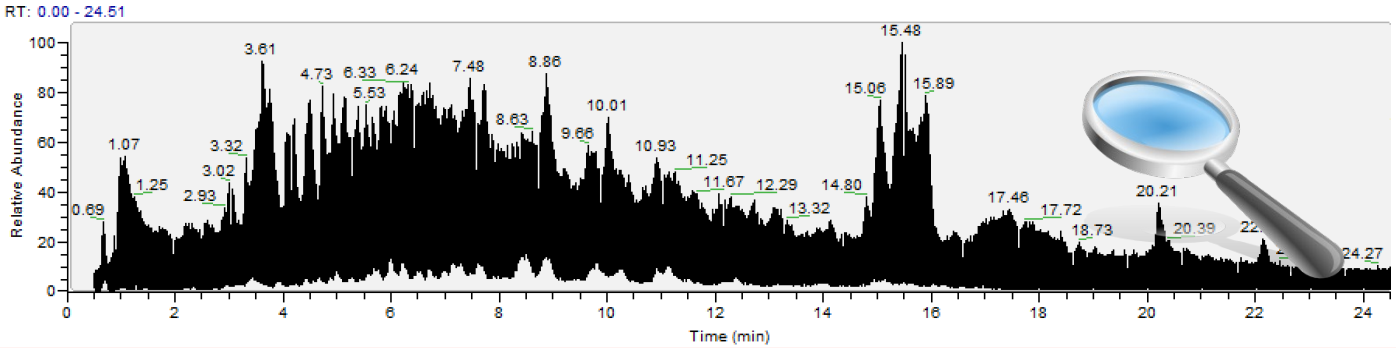


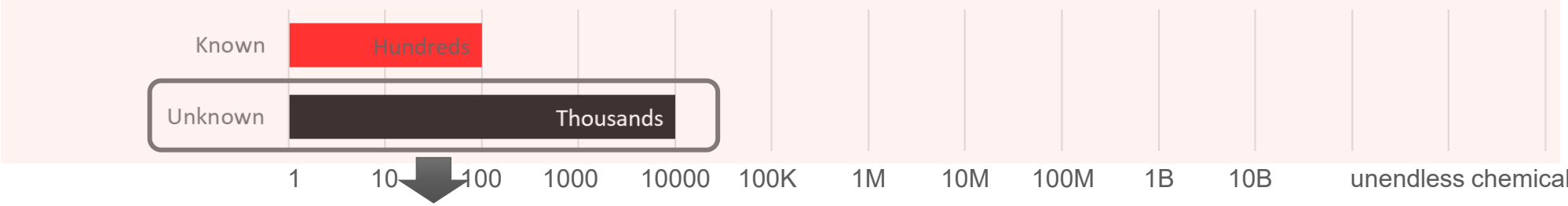


Environmental Cheminformatics & Non-target HR-MS

High resolution mass spectrometry

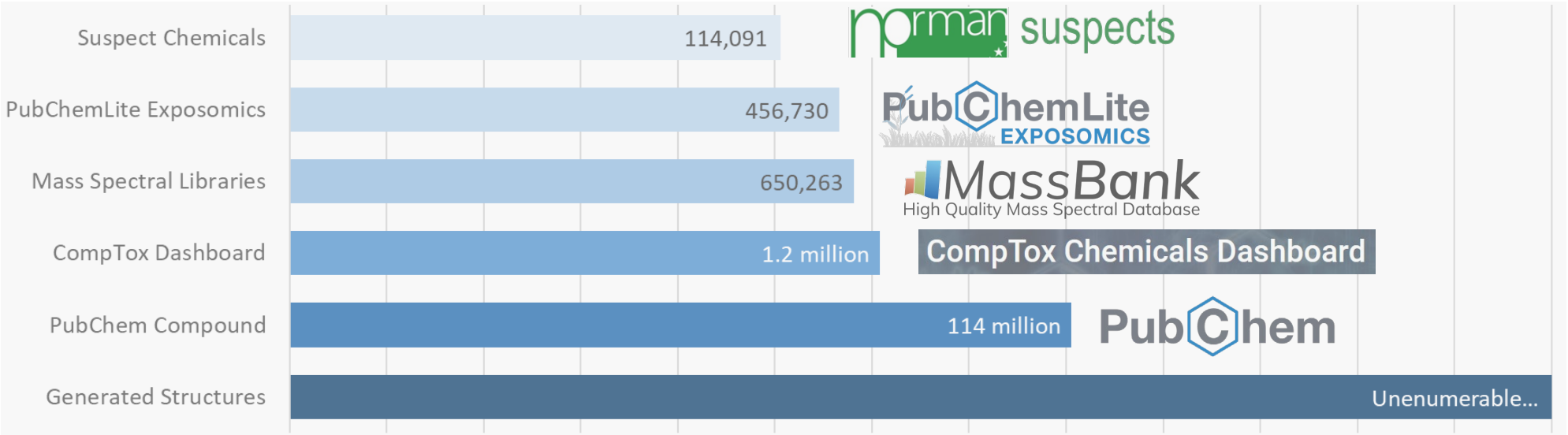
AND connecting chemical knowledge





Samples

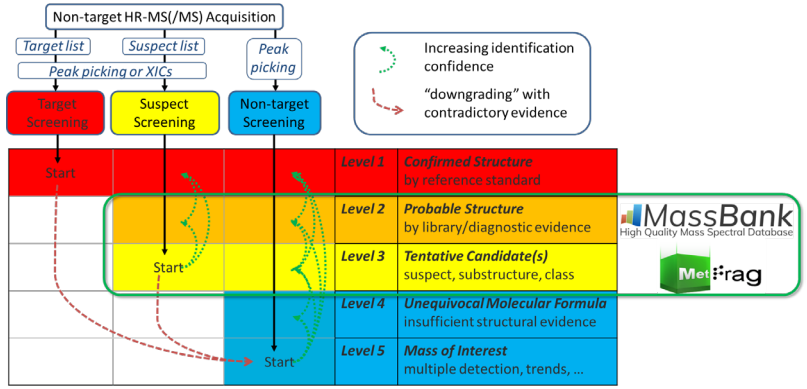
Chemical space

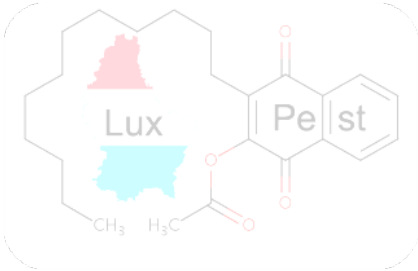


Chromatogram: Schymanski *et al*. (2014) DOI: 10.1021/es4044374; Resource numbers updated 15 March 2023.

Outline

- Identifying Chemicals with NT-HRMS
  - Open Resources for Identification
  - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases
  - Defining Relevant Chemical Space
  - Transformations: Known & Predicted
- Applications in Environmental Studies
  - Including a little bit of CCS
- Perspectives





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

Search for:

Basic Search | Peak List | Peaks | Peak Differences

Compound Information

Compound name

Exact Mass

AND

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

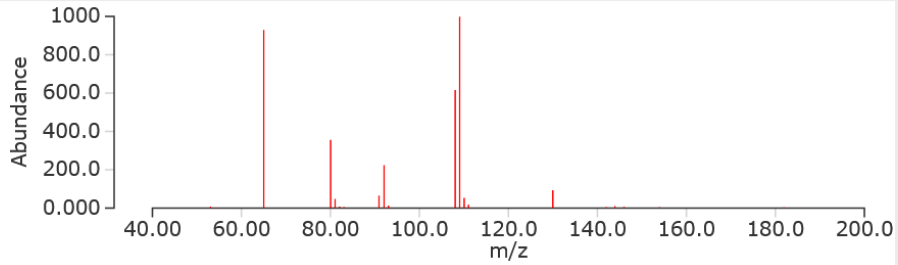
AND

Search

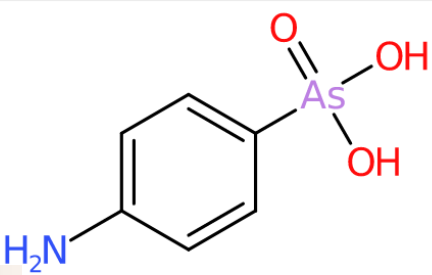
MassBank Record: LU040605

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

Mass Spectrum



Chemical Structure



Expert Knowledge: NORMAN Suspect List Exchange

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

NORMAN Database System

NORMAN organises the development and maintenance of var

SEARCH All Databases

Searching for individual substance or group(s) of substances

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

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

NORMAN Suspect List Exchang

The NORMAN Suspect List Exchange (NORMAN-SLE) was e
questions. The NORMAN-SLE documents all individual collec
(see Source column in SusDat). NORMAN-SLE versions are t

Comments and contributions are welcome - please email us a

Please refer to our documentation pages for: citation instructi

No.	Abbreviation	Description	Lin
S0	SUSDAT	**Merged NORMAN Suspect List: SusDat**	Inte Sus Met Cor

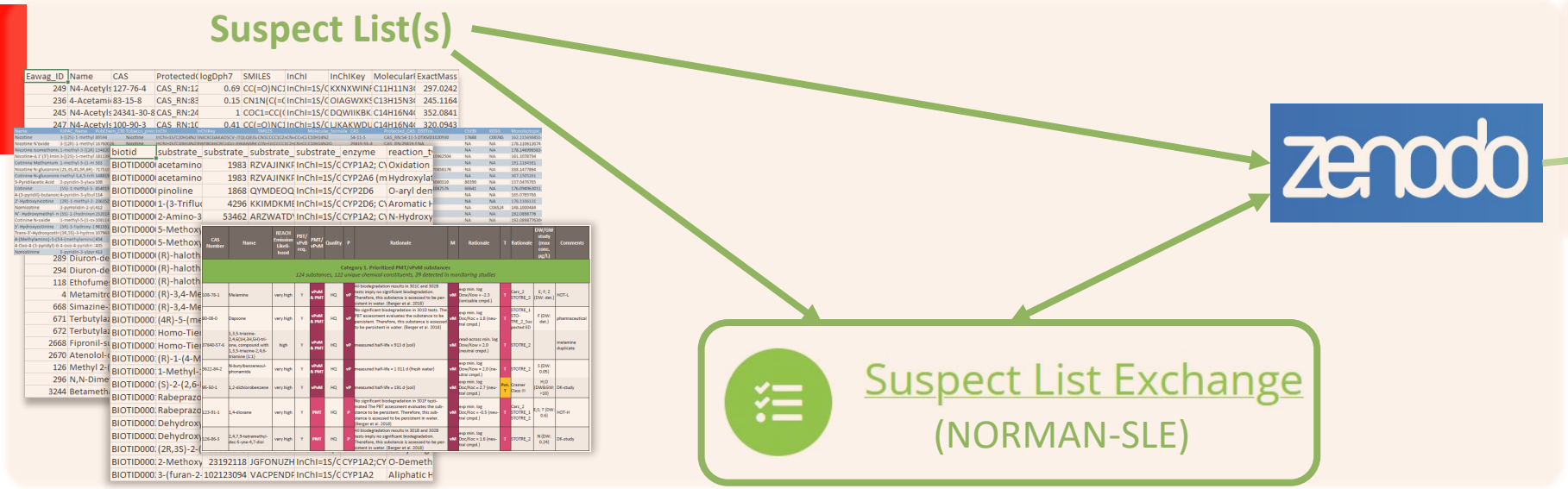
Antibiotic Resistance Bacteria/Genes

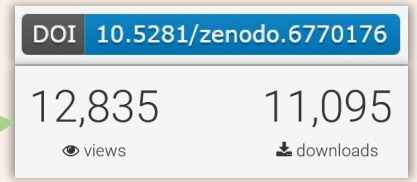
A database of ARBs/ARGs in environmen
matrices

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 Taha[1], Reza Aalizadeh[2], Nikiforos Alygizakis[3,2], Jean-Philippe Antignac[4], Hans Peter H. Arp[5,6], Richard Bade[7], Nancy Baker[8], Lidia Belova[9], Lubertus Bijlsma[10], Evan E. Bolton[11], Werner Brack[12,13], Alberto Celma[10,14], Wen-Ling Chen[15], Tiejun Cheng[11], Parviel Chirsir[1], Ľuboš Čirka[16,3], Lisa A. D'Agostino[17], Yannick Djoumbou Feunang[18], Valeria Dulio[19], Stellan Fischer[20], Pablo Gago-Ferrero[21], Aikaterini Galani[2], Birgit Geueke[22], Natalia Głowacka[3], Juliane Glüge[23], Ksenia Groh[24], Sylvia Grosse[25], Peter Haglund[26], Pertti J. Hakkinen[11], Sarah E. Hale[5], Felix Hernandez[10], Elisabeth M.-L. Janssen[24], Tim Jonkers[27], Karin Kiefer[24], Michal Kirchner[28], Jan Koschorreck[29], Martin Krauss[12], Jessy Krier[1], Marja H. Lamoree[27], Marion Letzel[30], Thomas Letzel[31], Qingliang Li[11], James Little[32], Yanna Liu[33], David M. Lunderberg[34,35], Jonathan W. Martin[17], Andrew D. McEachran[36], John A. McLean[37], Christiane Meier[29], Jeroen Meijer[38], Frank Menger[14], Carla Merino[39,40], Jane Muncke[22], Matthias Muschket[12], Michael Neumann[29], Vanessa Neveu[41], Kelsey Ng[3,42], Herbert Oberacher[43], Jake O'Brien[7], Peter Oswald[3], Martina Oswaldova[3], Jaqueline A. Picache[37], Cristina Postigo[44,14], Noelia Ramirez[45,39], Thorsten Reemtsma[12], Justin Renaud[46], Pawel Rostkowski[47], Heinz Rüdel[48], Reza M. Salek[41], Saer Samanipour[49], Martin Scheringer[23,42], Ivo Schliebner[29], Wolfgang Schulz[50], Tobias Schulze[12], Manfred Sengl[30], Benjamin A. Shoemaker[11], Kerry Sims[51], Heinz Singer[24], Randolph R. Singh[1,52], Mark Sumarah[46], Paul A. Thiessen[11], Kevin V. Thomas[7], Sonia Torres[39], Xenia Trier[53], Annemarie P. van Wezel[54], Roel C. H. Vermeulen[38], Jelle J. Vlaanderen[38], Peter C. von der Ohe[29], Zhanyun Wang[55], Antony J. Williams[56], Egon L. Willighagen[57], David S. Wishart[58], Jian Zhang[11], Nikolaos S. Thomaidis[2], Juliane Hollender[23,24], Jaroslav Slobodnik[3] and Emma L. Schymanski[1*]

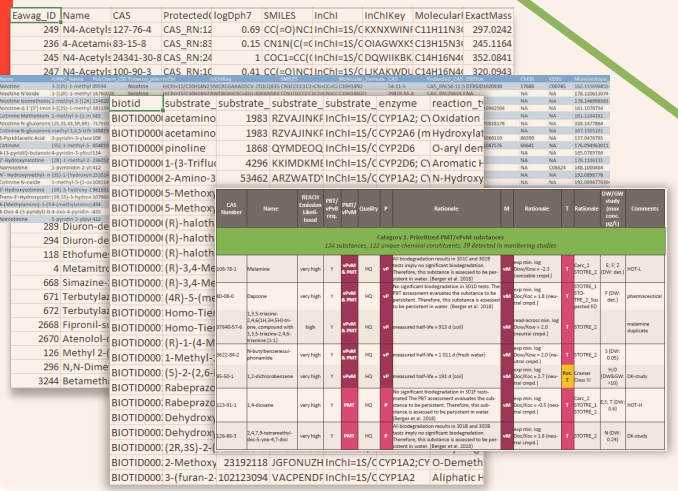


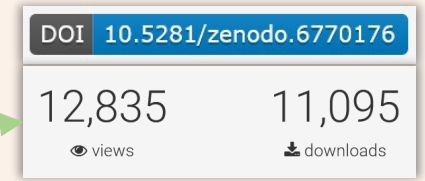


Mohammed Taha *et al.* (2022) DOI: 10.1186/s12302-022-00680-6

Mohammed Taha *et al.* (2022) DOI: 10.1186/s12302-022-00680-6

Suspect List(s)





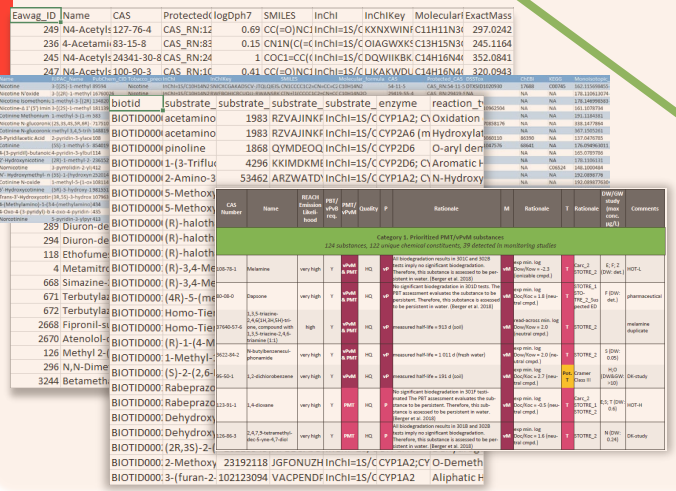
Suspect List Exchange (NORMAN-SLE)

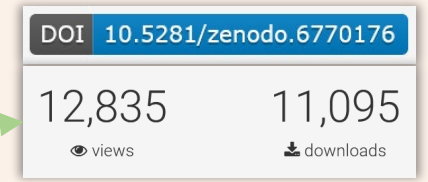
NORMAN Database System (NDS)

- Substance Database (NORMAN SusDat)
- Ecotoxicology
- SARS-CoV-2 in sewage
- Passive Sampling
- Antibiotic Resistance Bacteria/Genes
- Indoor Environment
- MassBank Europe
- Bioassays Monitoring Data
- Substance Factsheets
- Digital Sample Freezing Platform
- Prioritisation
- Chemical Occurrence Data

Mohammed Taha *et al.* (2022) DOI: 10.1186/s12302-022-00680-6

Suspect List(s)



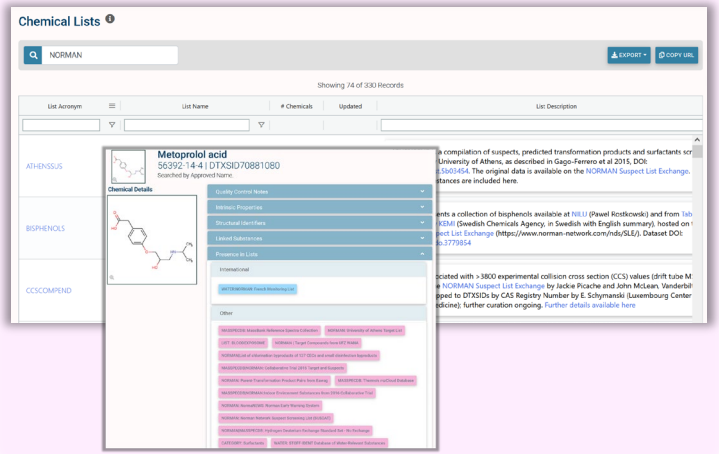


Suspect List Exchange (NORMAN-SLE)

NORMAN Database System (NDS)

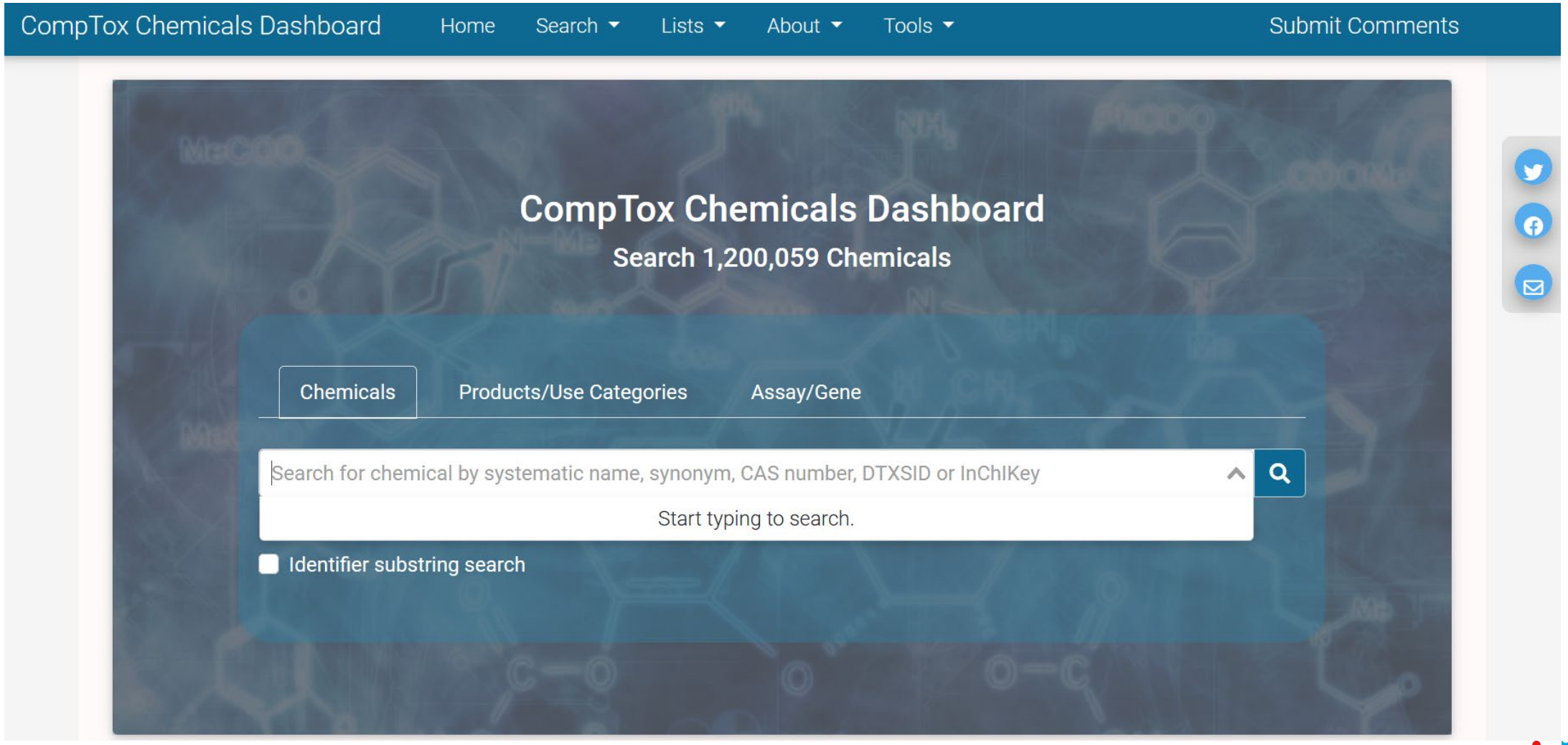
- Ecotoxicology
- SARS-CoV-2 in sewage
- Passive Sampling
- Indoor Environment
- Bioassays Monitoring Data
- Digital Sample Freezing Platform
- Chemical Occurrence Data
- Substance Database (NORMAN SusDat)
- Antibiotic Resistance Bacteria/Genes
- MassBank Europe
- Substance Factsheets
- Prioritisation

CompTox Chemicals Dashboard



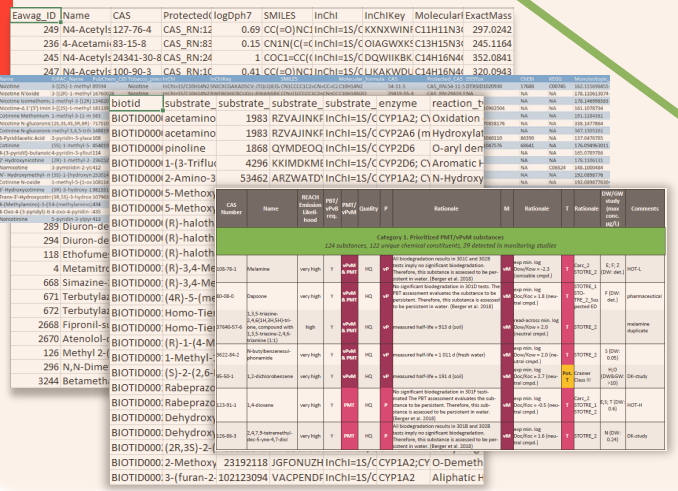
Toxicology: CompTox Chemicals Dashboard

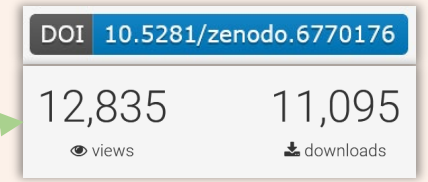
https://comptox.epa.gov/dashboard/



Mohammed Taha *et al.* (2022) DOI: 10.1186/s12302-022-00680-6

Suspect List(s)



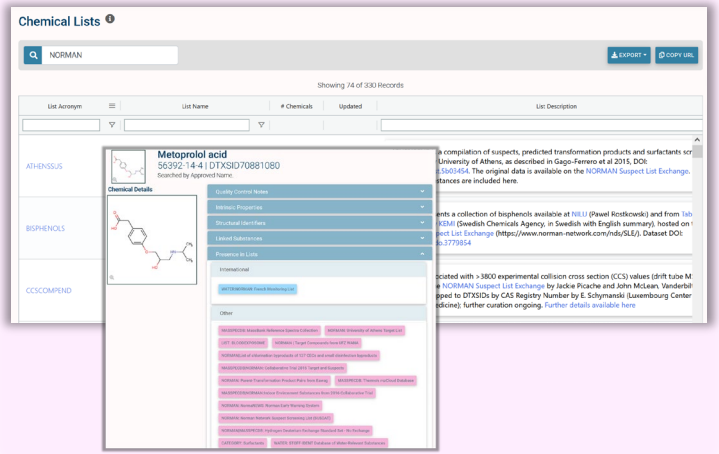


Suspect List Exchange (NORMAN-SLE)

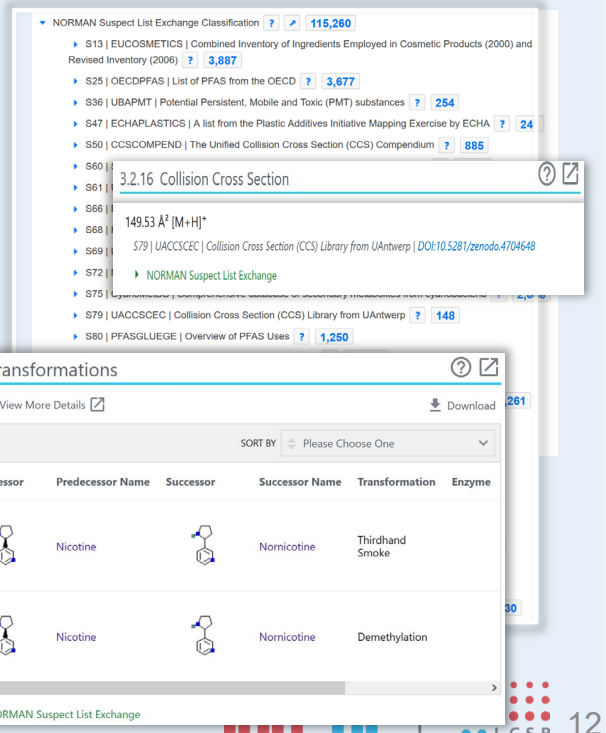
NORMAN Database System (NDS)

- Ecotoxicology
- SARS-CoV-2 in sewage
- Substance Database (NORMAN SusDat)
- Passive Sampling
- Antibiotic Resistance Bacteria/Genes
- Indoor Environment
- MassBank Europe
- Bioassays Monitoring Data
- Substance Factsheets
- Digital Sample Freezing Platform
- Prioritisation
- Chemical Occurrence Data

CompTox Chemicals Dashboard



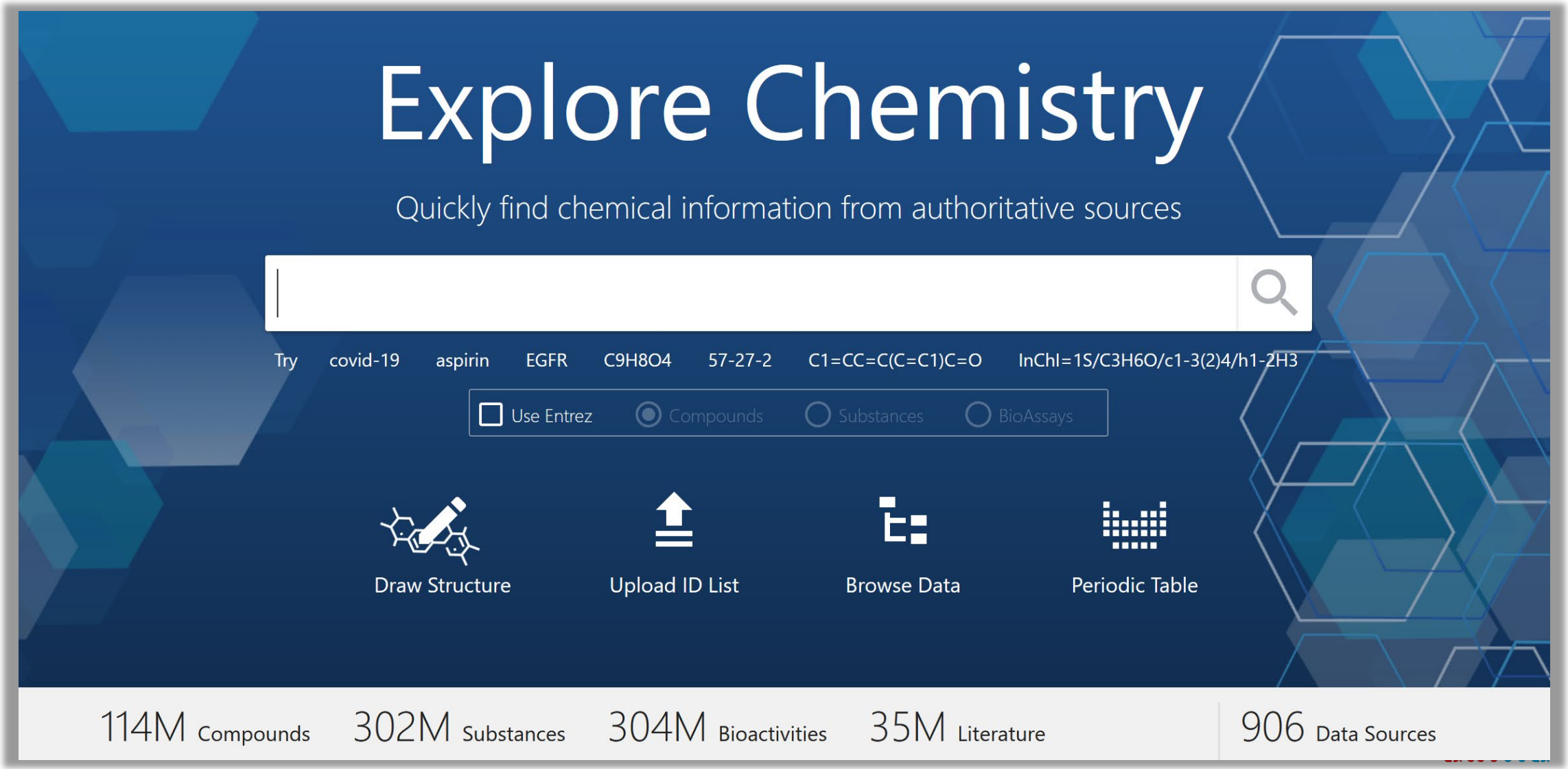
PubChem



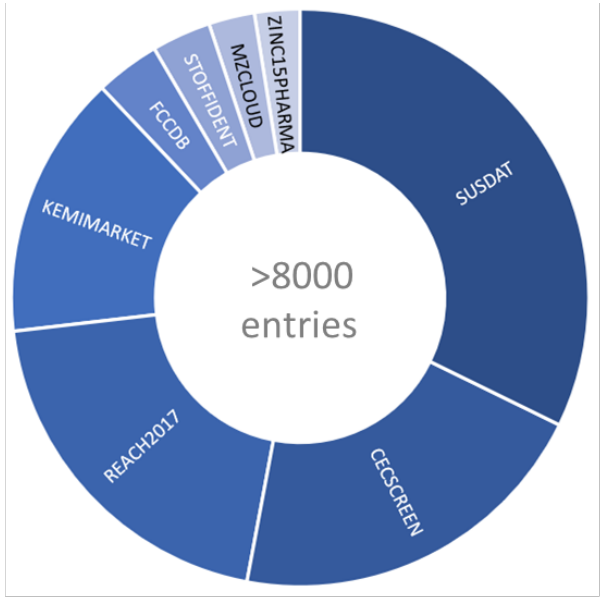
Open Chemistry Knowledge Bases: PubChem

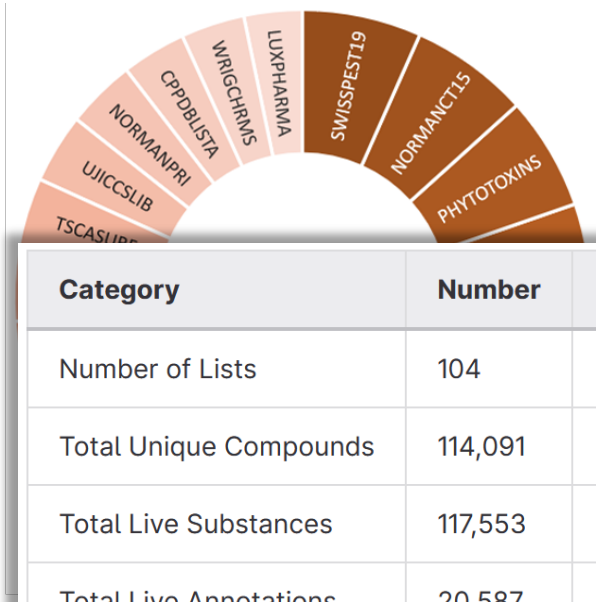
U.S. National Library of Medicine
National Center for Biotechnology Information

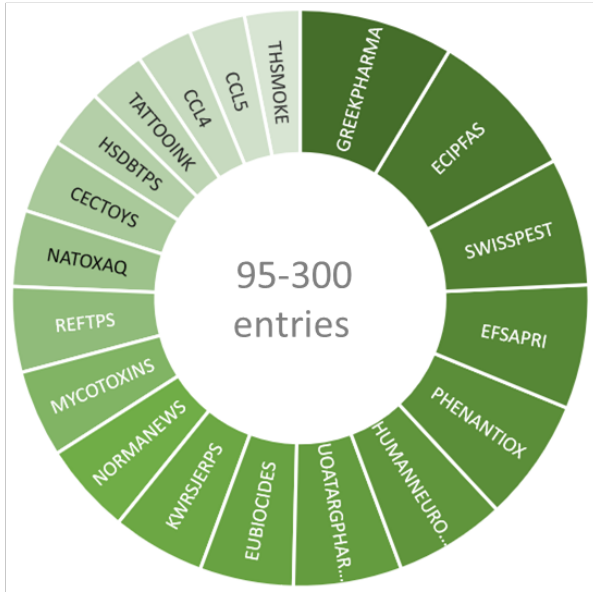
https://pubchem.ncbi.nlm.nih.gov/

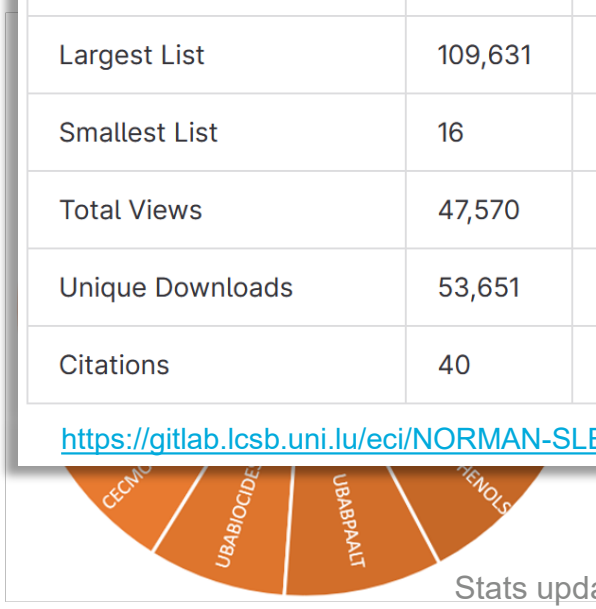


NORMAN-SLE: Now >100 lists!









Category	Number	Comment
Number of Lists	104	S00 to S103
Total Unique Compounds	114,091	From PubChem NORMAN-SLE Tree
Total Live Substances	117,553	From PubChem NORMAN-SLE Source Page
Total Live Annotations	20,587	From PubChem NORMAN-SLE Source Page
Largest List	109,631	S00 NORMAN-SusDat
Smallest List	16	S98 TIRECHEM (a work in progress)
Total Views	47,570	From Zenodo (see below)
Unique Downloads	53,651	From Zenodo (see below)
Citations	40	From Zenodo (see below)

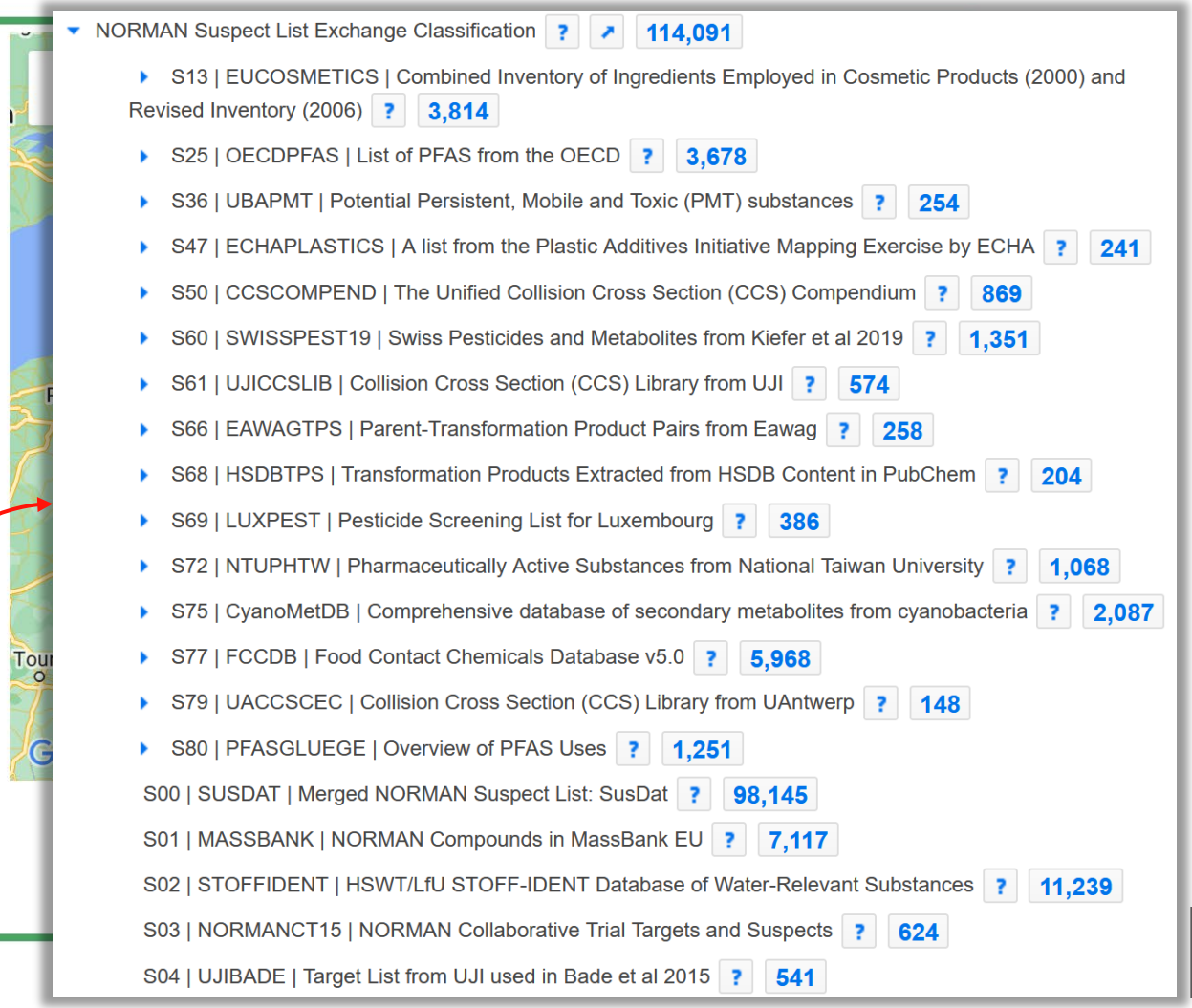
https://gitlab.lcsb.uni.lu/eci/NORMAN-SLE/-/tree/master/stats#norman-sle-summary

Stats updated Mar. 2023

NORMAN Suspect List Exchange in PubChem

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. It specifically seeks both to promote and to benefit from the synergies between research teams from different countries in the field of emerging substances.

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	117,553 Live Substances 20,587 Annotations 1 Classification
Last Updated	2023/03/09



https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101

- NORMAN Suspect List Exchange Classification ? 115,248
  - S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cos
    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 E
  - 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 Universit
  - S75 | CyanoMetDB | Comprehensive database of secondary metabolites from cyanoba
  - S79 | UACCSCEC | Collision Cross Section (CCS) Library from UAntwerp ? 148
  - S80 | PFASGLUEGE | Overview of PFAS Uses ? 1,250
  - S00
  - S01
  - S02
  - S03
  - S04

PubChem Phloroglucinol (Compound)

9.1.1 Use Classification

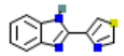
Cosmetics -> Hair dyeing

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

NORMAN Suspect List Exchange

9.6 Transformations

1 item View More Details Download

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme
	Thiabendazole		Benzimidazole	Environmental Transformation	

NORMAN Suspect List Exchange

PubChem Tri-O-cresyl phosphate (Compound)

3.2.18 Collision Cross Section

182.39 Å^2 [M+H]$^+$

192.43 Å^2 [M+Na]$^+$

263.75 Å^2 [2M+Na]$^+$

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

NORMAN Suspect List Exchange

PubChem 7-Demethylmicrocystin RR (Compound)

7 Taxonomy

1 item View More Details Download

Taxonomy	Evidence PMID	Evidence DOI
Microcystis aeruginosa	1440646	10.1016/0041-0101(92)90054-9

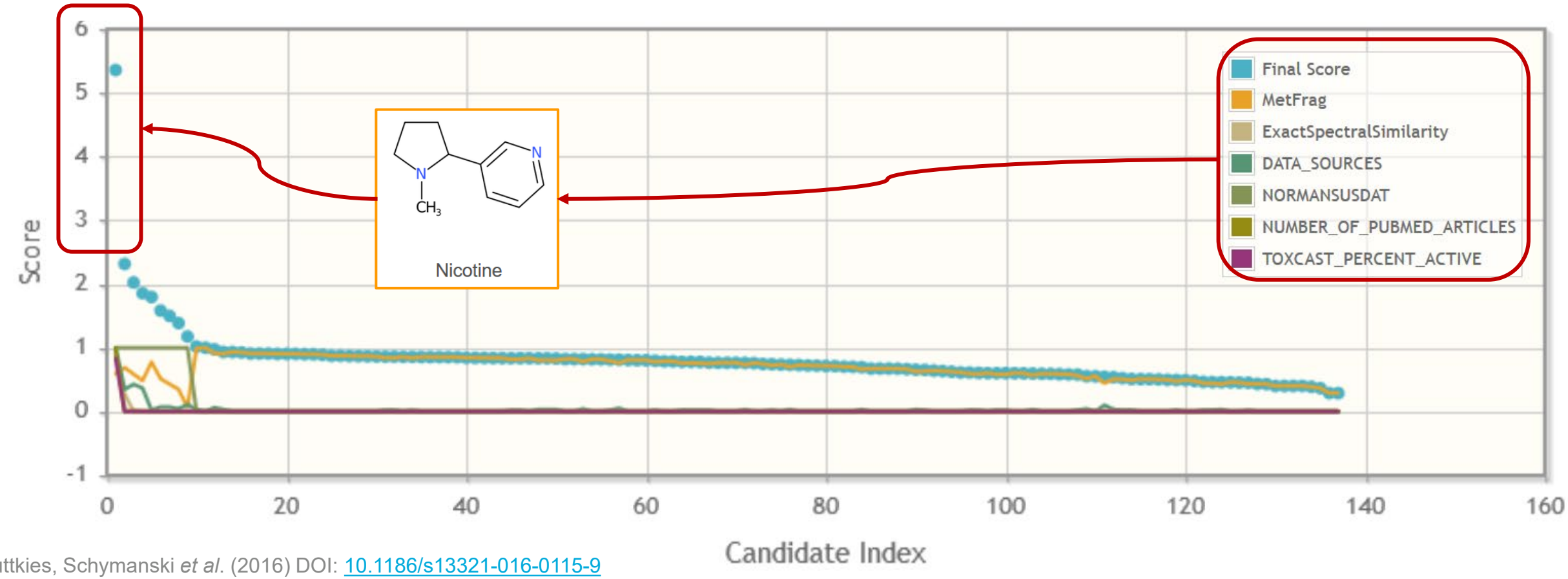
S75 | CyanoMetDB | Comprehensive database of secondary metabolites from cyanobacteria | DOI:10.5281/zenodo.4551528

NORMAN Suspect List Exchange

Mohammed Taha *et al*. (2022) DOI: 10.1186/s12302-022-00680-6

Connecting Knowledge for Chemical Identification: MetFrag

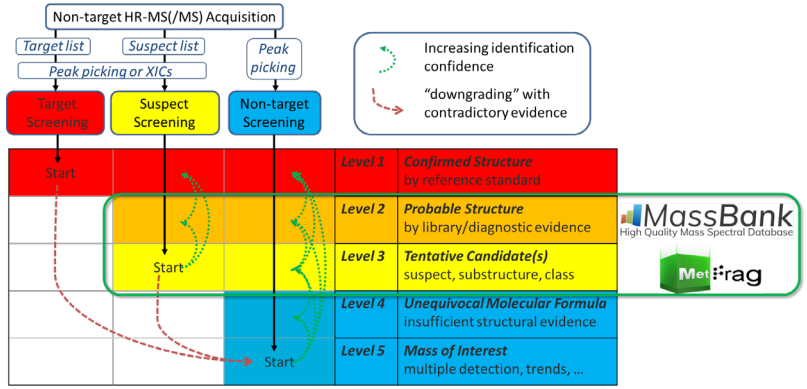
https://msbi.ipb-halle.de/MetFrag/ http://ipb-halle.github.io/MetFrag/

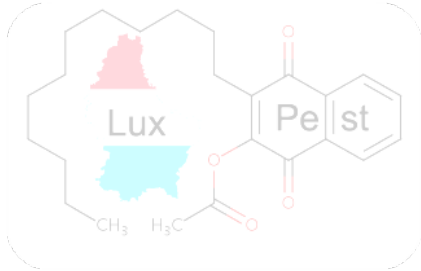


Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9

Outline

- Identifying Chemicals with NT-HRMS
  - Open Resources for Identification
  - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases
  - Defining Relevant Chemical Space
  - Transformations: Known & Predicted
- Applications in Environmental Studies
  - Including a little bit of CCS
- Perspectives





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

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

204 million

PubChem

114 million

CompTox Chemicals Dashboard

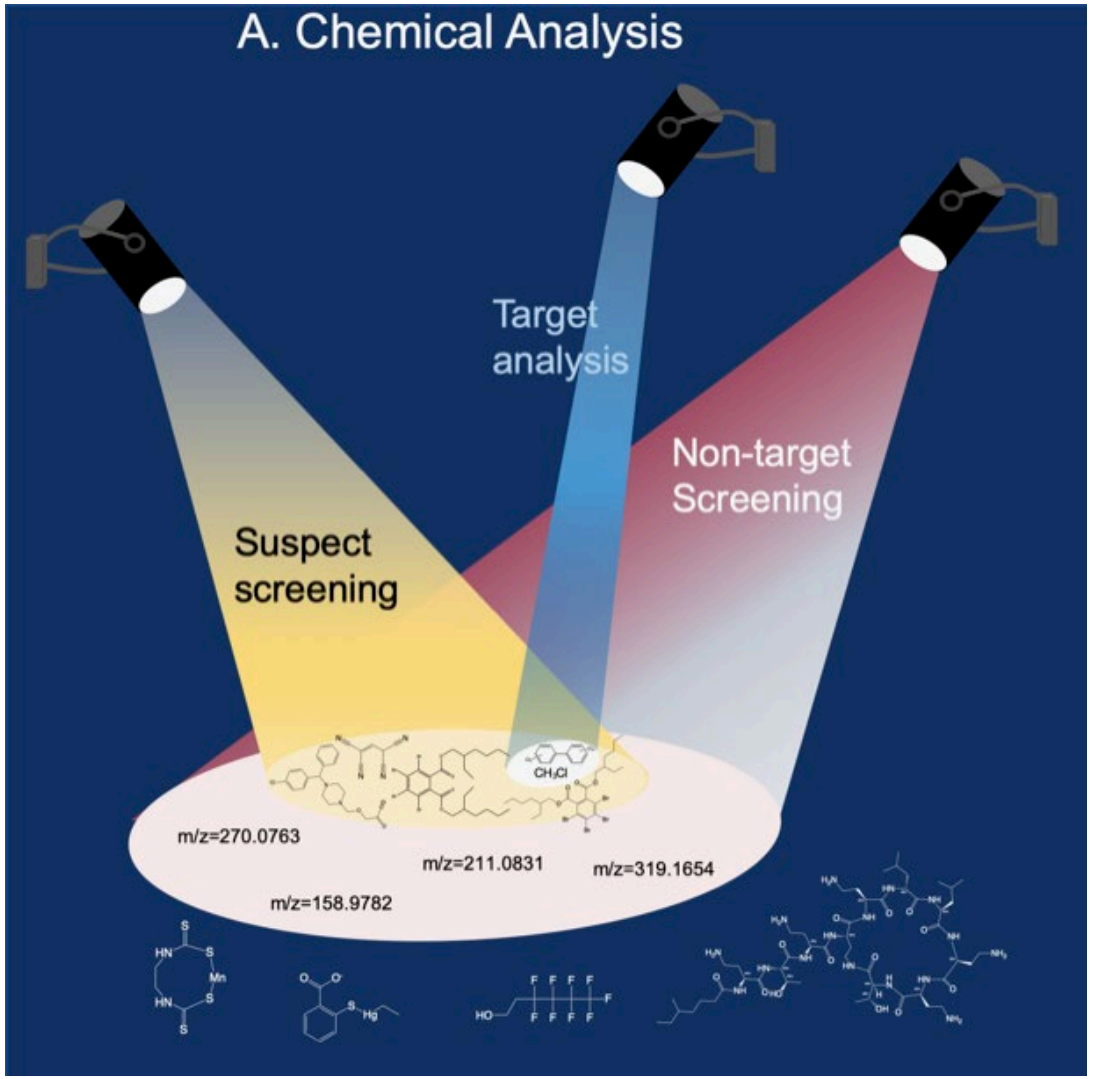
1.2 million

PubChemLite EXPOSOMICS

456,730

norman suspects

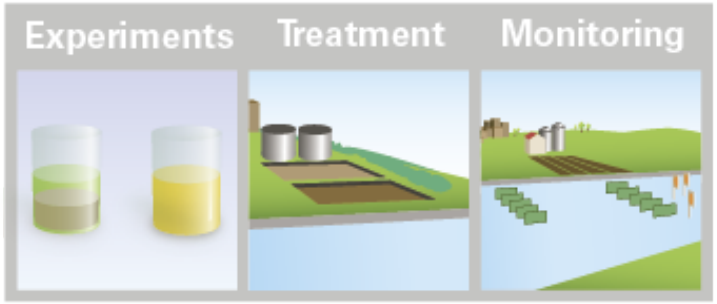
114,091



Numbers updated 15 Mar. 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

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

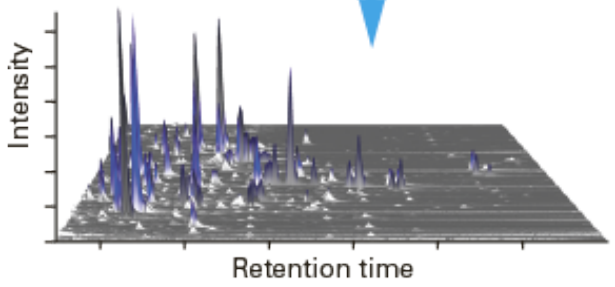


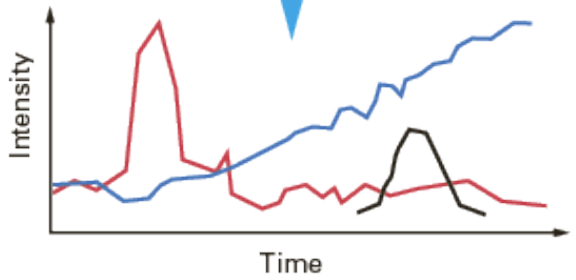
Commercial / Open Software

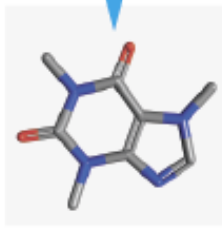
Databases

In Silico Prediction

Sampling → Analysis → Data Pre-Processing → Prioritization → Identification







SAMPLE ARCHIVE | DIGITAL ARCHIVE | POLLUTION OVERVIEW | PATTERN | STRUCTURES

Modified from: Hollender, Schymanski *et al*. 2018, ES&T Feature, 51:20, 11505-11512. DOI: 10.1021/acs.est.7b02184

Open Source Workflows for NT-HRMS: patRoon

Software | Open Access | Published: 06 January 2021

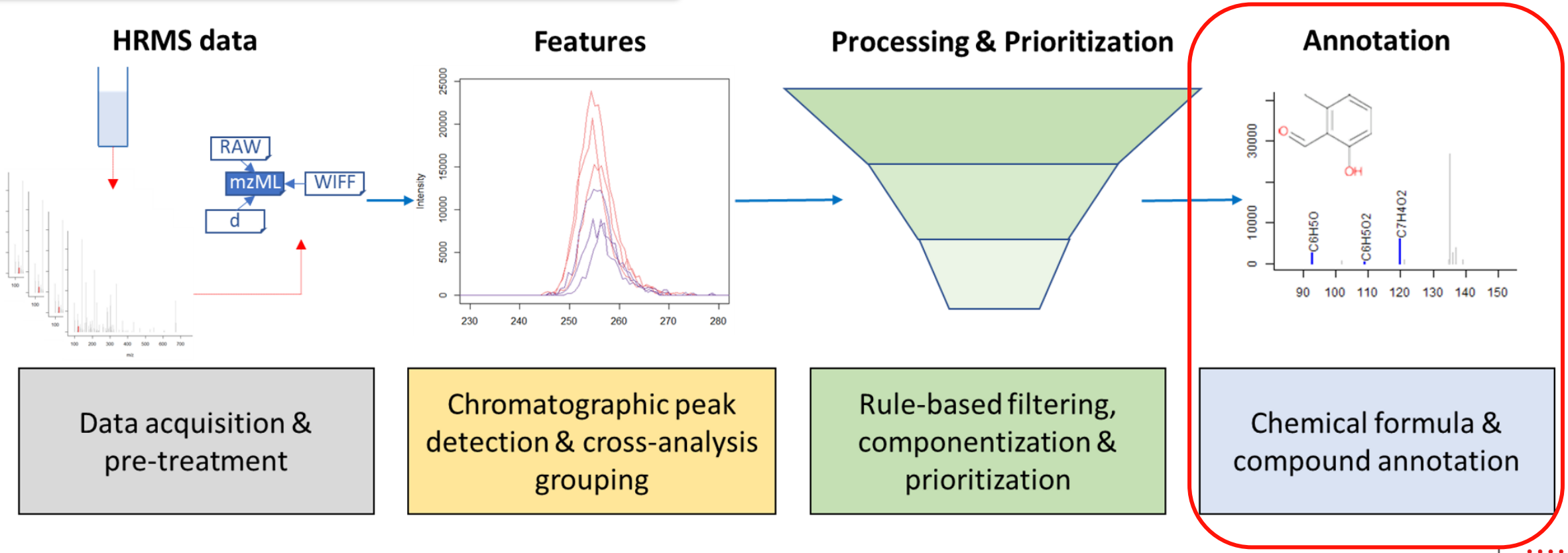
patRoon: open source software platform for environmental mass spectrometry based non-target screening

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

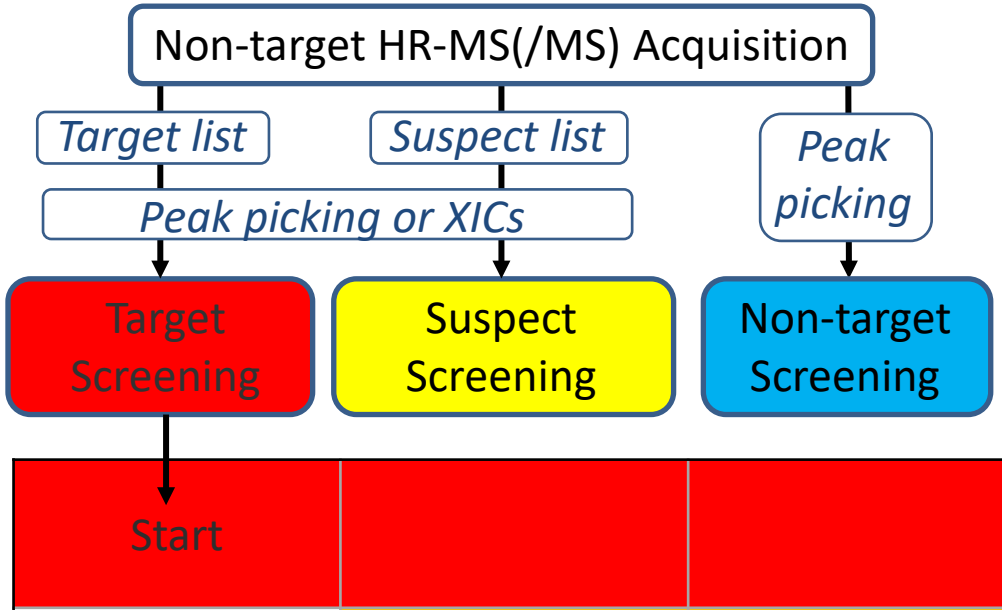
patRoon 2.0: Improved non-target analysis workflows including automated transformation product screening

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]



Helmus *et al.* (2022) patRoon 2.0 *J. Open Source Software*, DOI: 10.21105/joss.04029
Helmus *et al.* (2021) patRoon. *J. Cheminformatics*. DOI: 10.1186/s13321-020-00477-w

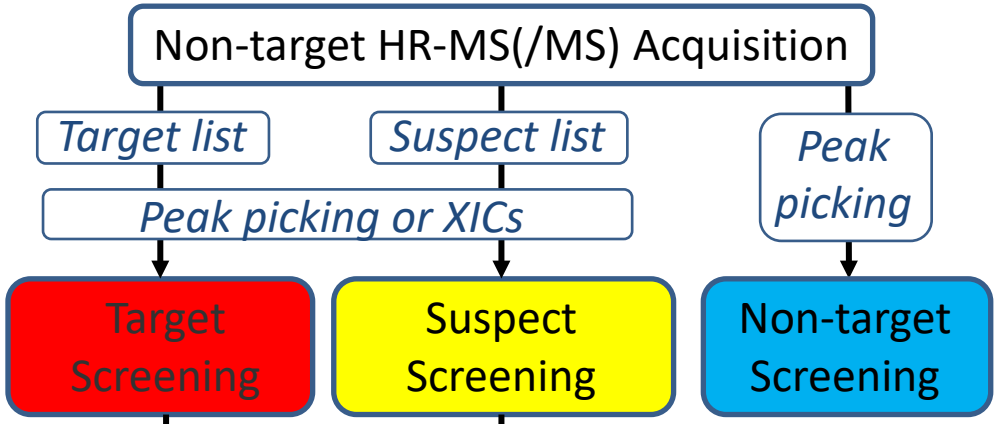
Identification Strategies and Confidence in NT-HRMS(/MS)



Limited by ref. std. availability …

Schymanski *et al*, 2014, ES&T. DOI: 10.1021/es5002105 & Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

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 …

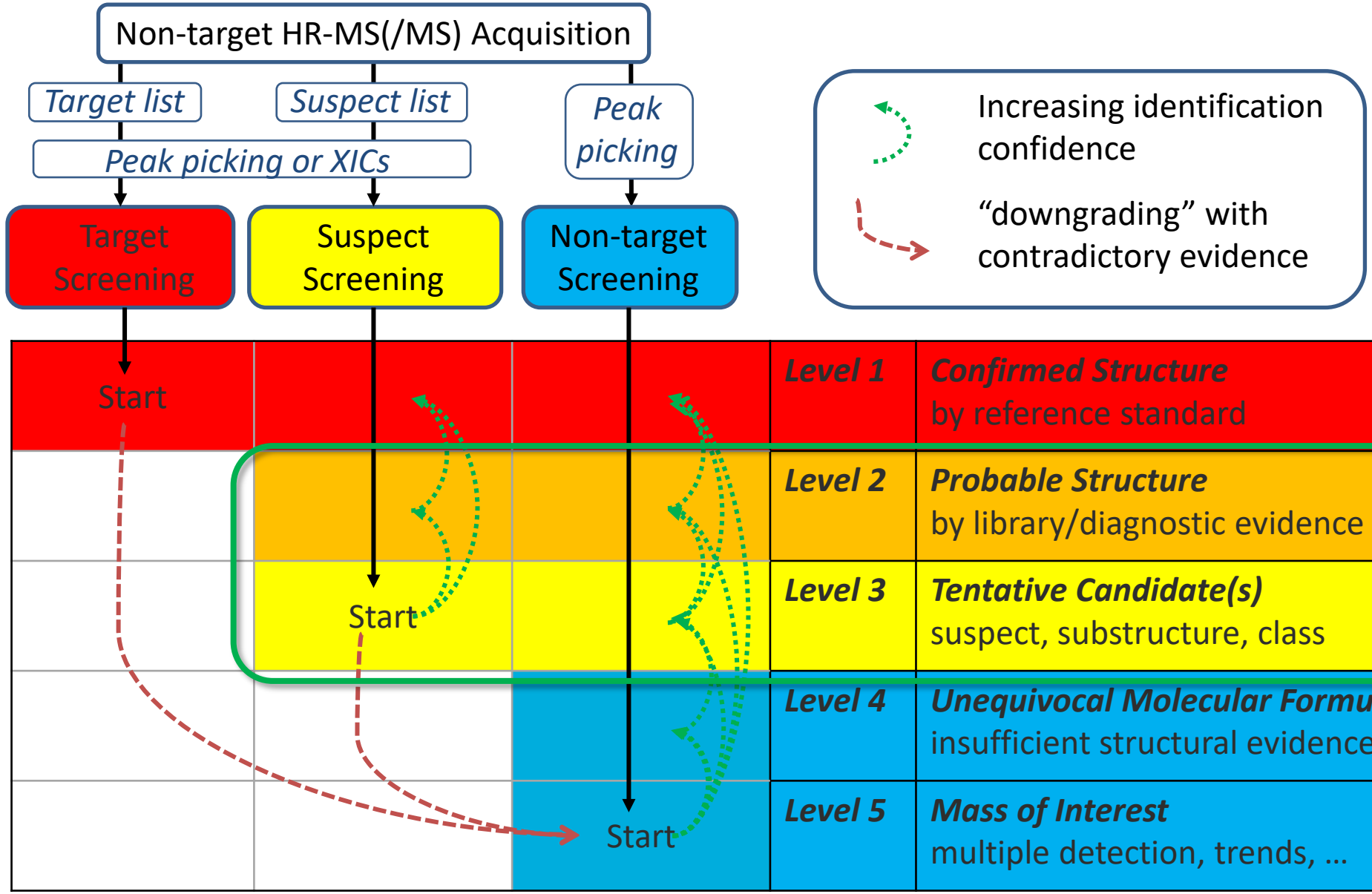
MassBank
High Quality Mass Spectral Database

MetFrag

eawag aquatic research

Schymanski *et al*, 2014, ES&T. DOI: 10.1021/es5002105 & Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

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



Limited by ref. std. availability …

MassBank
High Quality Mass Spectral Database

MetFrag

The "dark matter"

eawag aquatic research

Schymanski *et al*, 2014, ES&T. DOI: 10.1021/es5002105 & Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

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

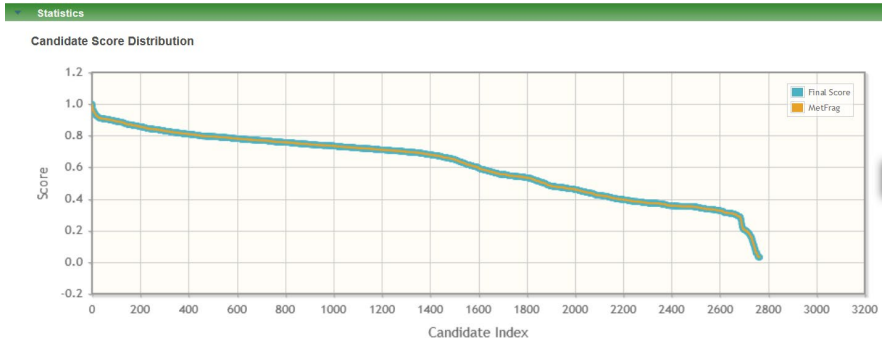
mz [M−H]⁻
213.9637
± 5 ppm

5 ppm
0.001 Da

PubChem

MetFrag

Ranked Candidates

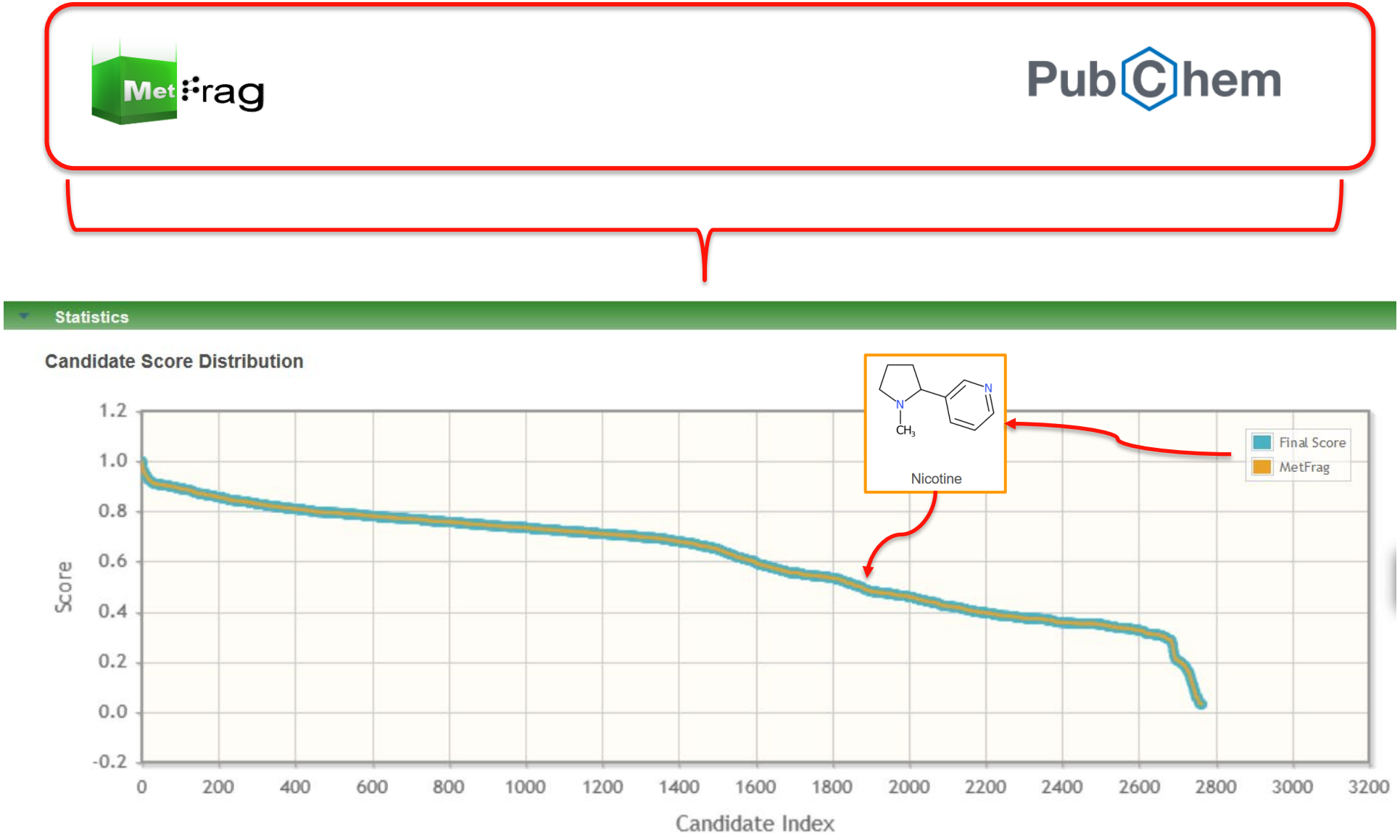


MS/MS

134.0054	339689
150.0001	77271
213.9607	632466

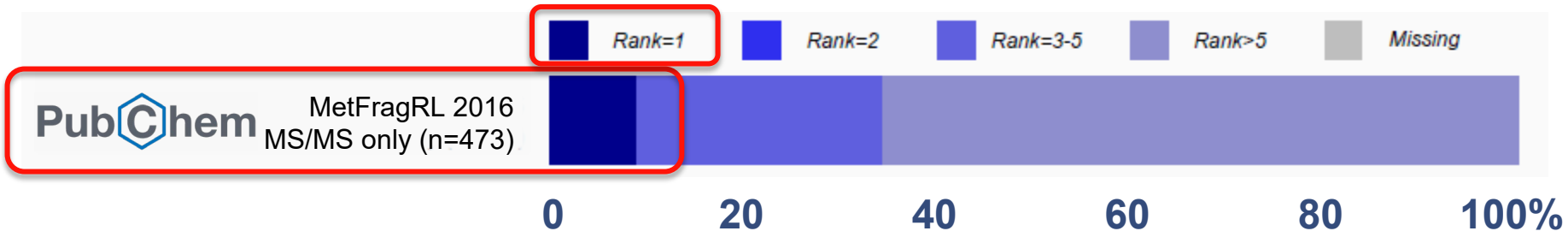
Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9

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



MetFrag + PubChem + Formula Search + https://massbank.eu/MassBank/RecordDisplay?id=EQ300804

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



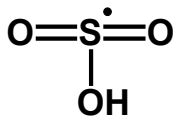
Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9 Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

Adding in more information: MetFrag Relaunched (2016)

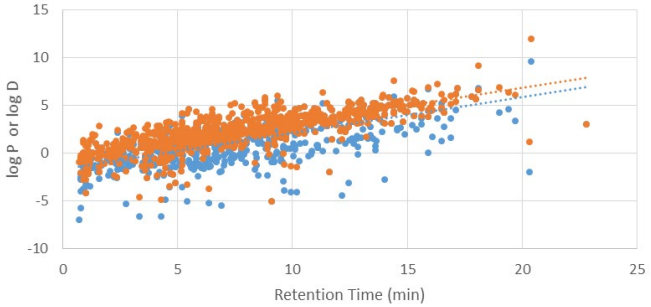
mz [M-H]⁻
213.9637
± 5 ppm

Elements: C,N,S

5 ppm
0.001 Da



RT: 4.54 min
355 InChI/RTs



CompTox Chemicals Dashboard
or
PubChem

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

Literature Patents

MetFrag

MoNA
MassBank of North America
MassBank
High Quality Mass Spectral Database

Suspect Lists

TOFF IDENT
norman suspects
CompTox Chemicals Dashboard

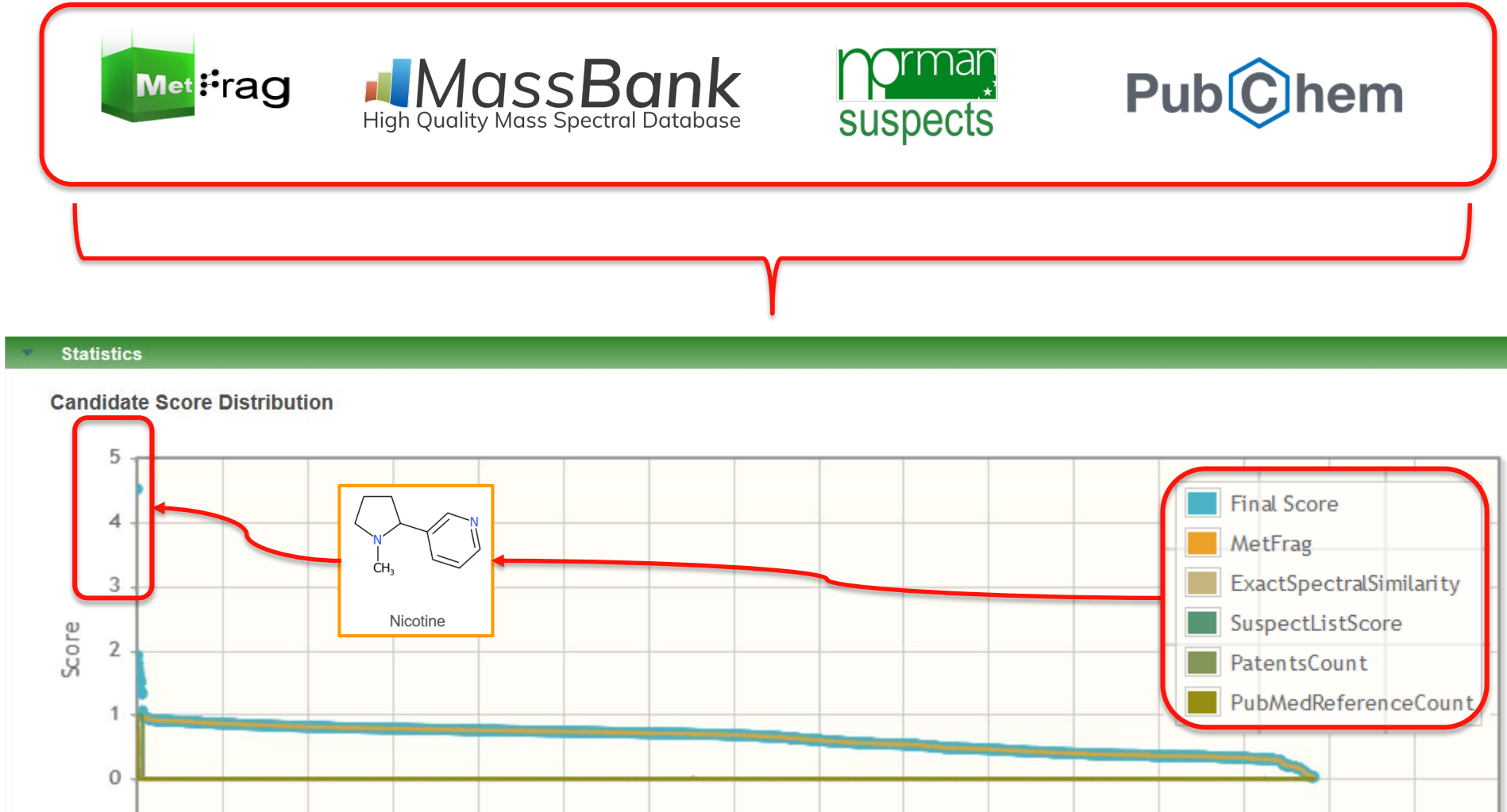
MS/MS

134.0054	339689
150.0001	77271
213.9607	632466

eawag aquatic research
solutions

Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9

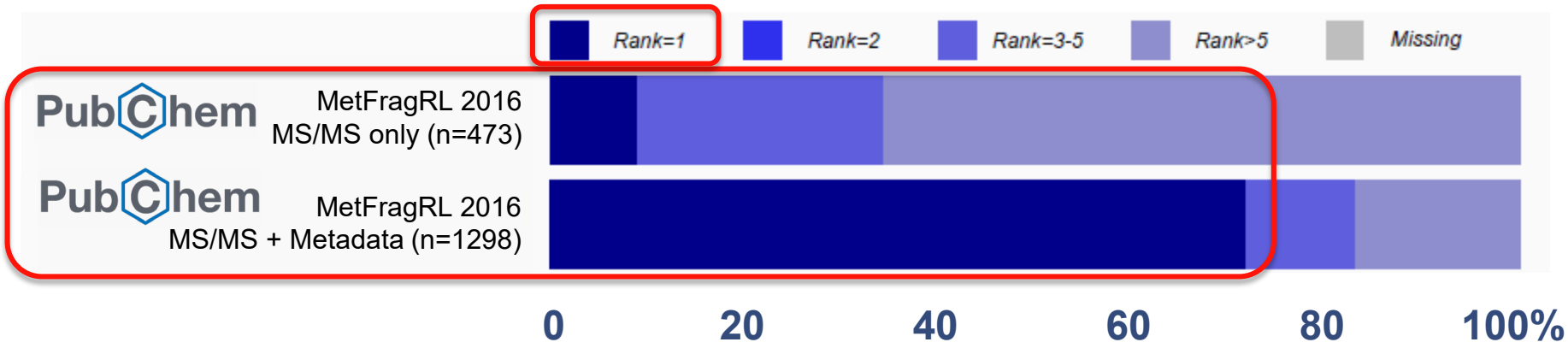
MetFragRL + PubChem + MS/MS + Metadata



MetFrag + PubChem + Formula + MoNA + SusDat + Pat + Refs + https://massbank.eu/MassBank/RecordDisplay?id=EQ300804

MetFragRL + PubChem + MS/MS + Metadata

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



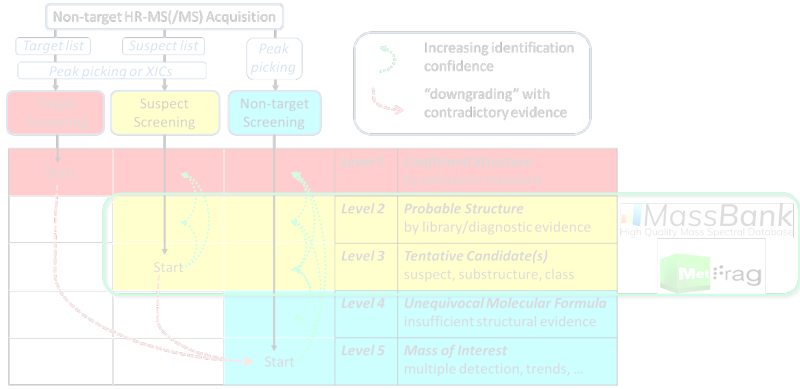
BUT …databases grow … identification performance drops

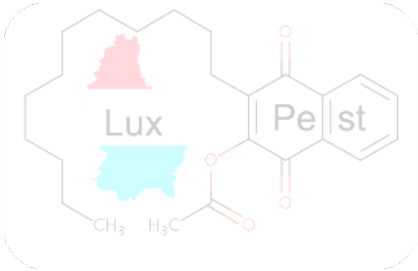
… and run times rise … (a lot!)

Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9 Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

Outline

- Identifying Chemicals with NT-HRMS
  - Open Resources for Identification
  - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases
  - Defining Relevant Chemical Space
  - Transformations: Known & Predicted
- Applications in Environmental Studies
  - Including a little bit of CCS
- Perspectives





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

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

204 million

PubChem

114 million

CompTox Chemicals Dashboard

1.2 million

PubChemLite EXPOSOMICS

456,730

norman suspects

114,091

Numbers updated 15 Mar. 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

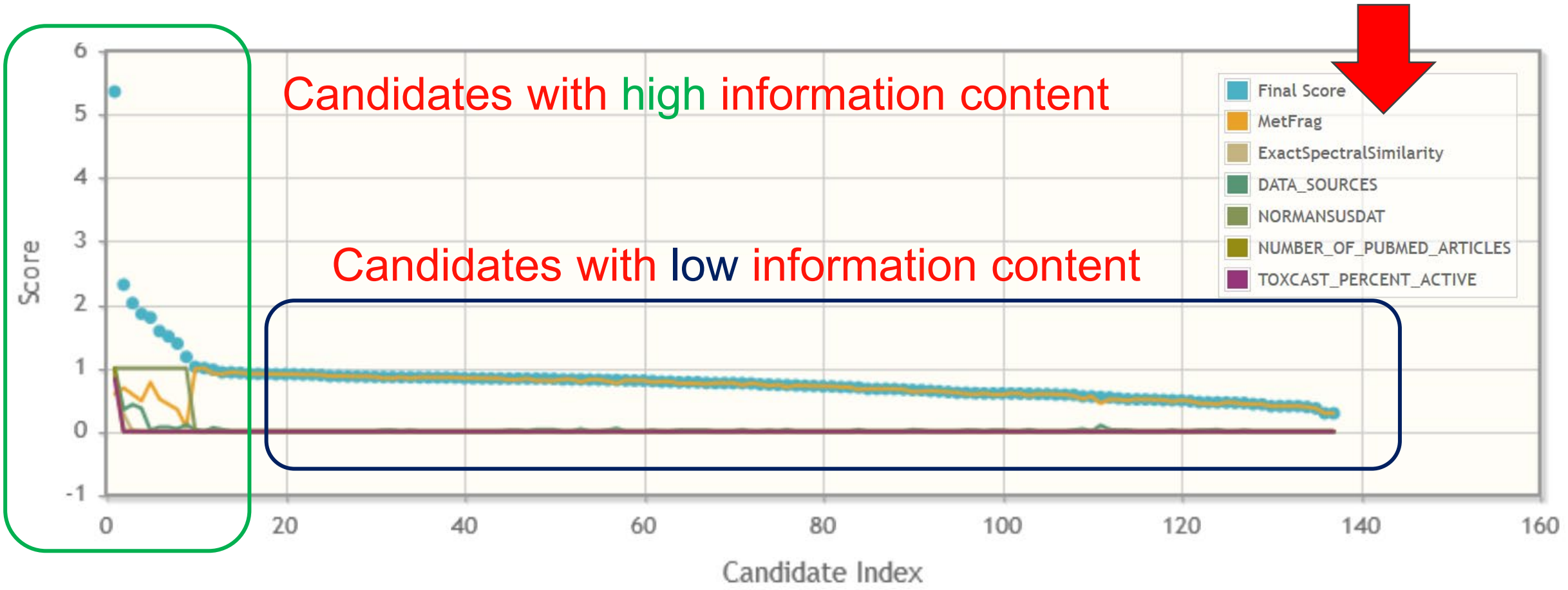
Problem: Exposomics “Chemical Space” is too big!

CAS A DIVISION OF THE AMERICAN CHEMICAL SOCIETY — 204 million

PubChem — 114 million

ChemSpider Search and share chemistry — 115 million

CompTox Chemicals Dashboard — 1.2 million



Can we break down PubChem into useful bits?

- PubChem Compound TOC 65,940,905
  - Agrochemical Information 3,090
  - Associated Disorders and Diseases 29,999
  - Biologic Description 2,464,498
  - Biological Test Results 4,163,774
  - Chemical and Physical Properties 267,905
  - Classification 22,519,280
  - Drug and Medication Information 19,171
  - Food Additives and Ingredients 7,337
  - Identification 4,719
  - Information Sources 45,543,957
  - Interactions and Pathways 204,206
  - Literature 2,093,131
  - Names and Identifiers 4,974,299
  - Patents 38,011,046
  - Pharmacology and Biochemistry 113,464
  - Related Records 13,106,636
  - Safety and Hazards 171,449
  - Spectral Information 1,535,947
  - Structures 11,758,914
  - Toxicity 116,718
  - Use and Manufacturing 105,769

- Agrochemical Information 3,090
  - Agrochemical Category 1,945
  - Agrochemical Transformations 1,475
  - EU Pesticides Data 1,230
  - USDA Pesticide Data Program 666

- Associated Disorders and Diseases 29,999
  - Disease and References 13,325

- Use and Manufacturing 105,769
  - Uses 71,631
    - Consumer Uses 3,787
    - Household Products 3,862
    - Industry Uses 5,938
    - Use Classification 59,082
  - Consumption Patterns 1,128
  - Formulations/Preparations 4,498
  - General Manufacturing Information 50,886

PubChem Furathiocarb (Compound)

CONTENTS

7 Agrochemical Information

7.3 EU Pesticides Data

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

PubChem Nicotine (Compound)

14 Associated Disorders and Diseases

250 items View More Rows & Details 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)

Schymanski *et al.* (2021) DOI: 10.1186/s13321-021-00489-0

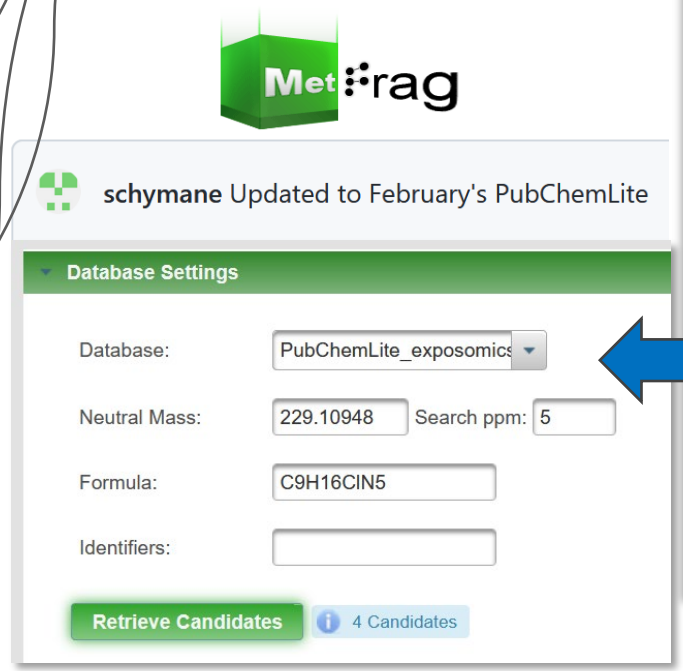
U.S. National Library of Medicine
National Center for Biotechnology Information

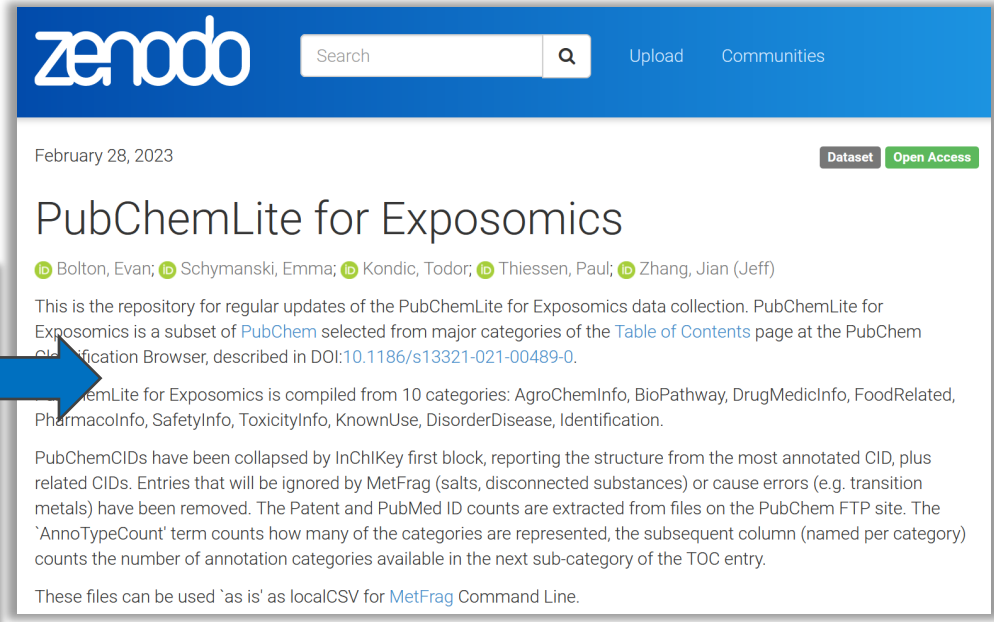
Introducing …

- PubChem Compound TOC ? 65,940,905
  - Agrochemical Information ? 3,090
  - Associated Disorders and Diseases ? 29,999
  - Biologic Description ? 2,464,498
  - Biological Test Results ? 4,163,774
  - Chemical and Physical Properties ? 267,905
  - Classification ? 22,519,280
  - Drug and Medication Information ? 19,171
  - Food Additives and Ingredients ? 7,337
  - Identification ? 4,719
  - Information Sources ? 45,543,957
  - Interactions and Pathways ? 204,206
  - Literature ? 2,093,131
  - Names and Identifiers ? 4,974,299
  - Patents ? 38,011,046
  - Pharmacology and Biochemistry ? 113,464
  - Related Records ? 13,106,636
  - Safety and Hazards ? 171,449
  - Spectral Information ? 1,535,947
  - Structures ? 11,758,914
  - Toxicity ? 116,718
  - Use and Manufacturing ? 105,769

PubChemLite EXPOSOMICS

~450,000 entries “small”



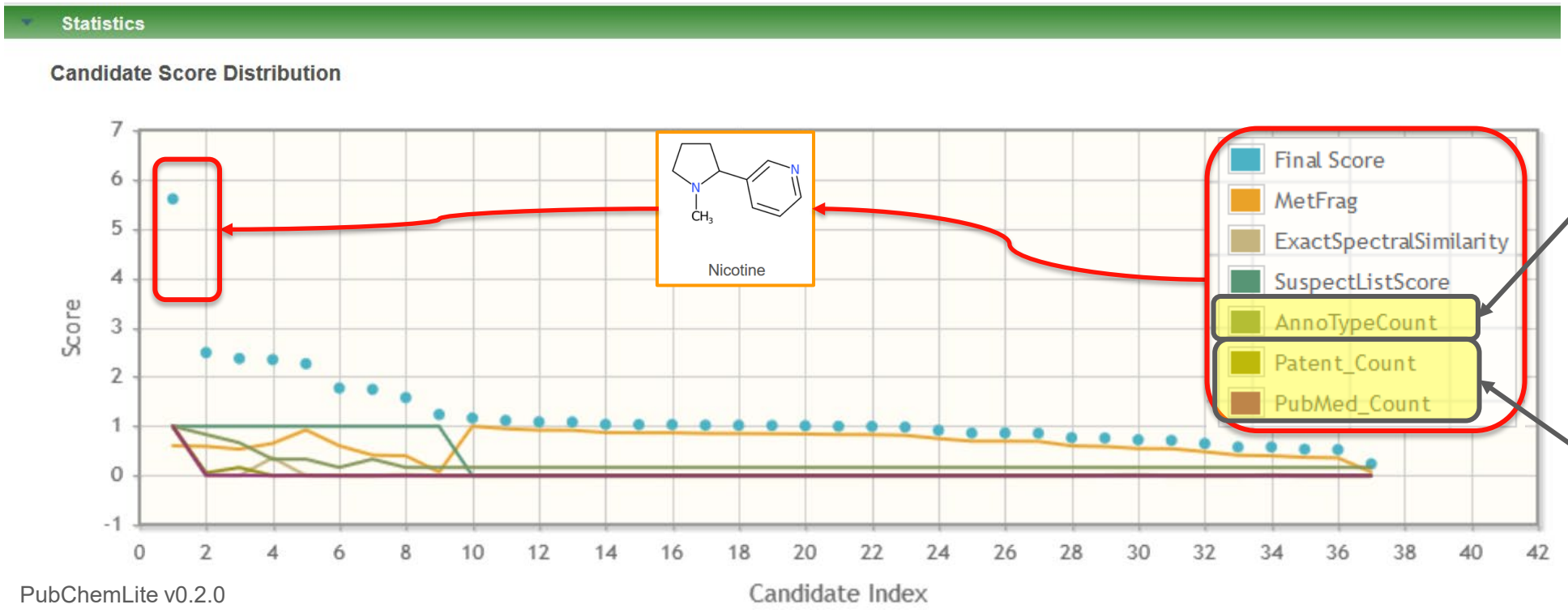


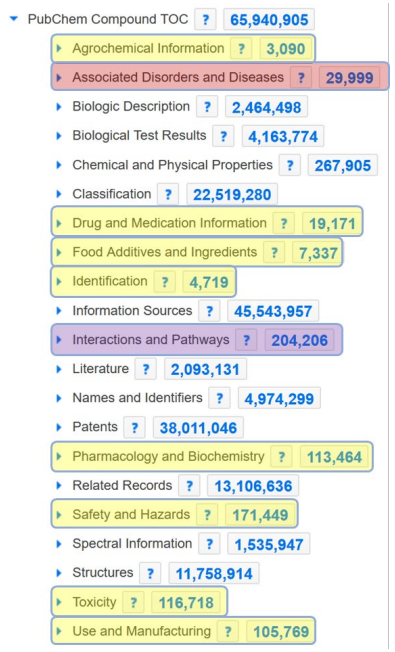
Schymanski *et al.* (2021) DOI: 10.1186/s13321-021-00489-0

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

PubChemLite: Fewer and more relevant candidates – with context

MetFrag · MassBank High Quality Mass Spectral Database · norman suspects · PubChemLite EXPOSOMICS



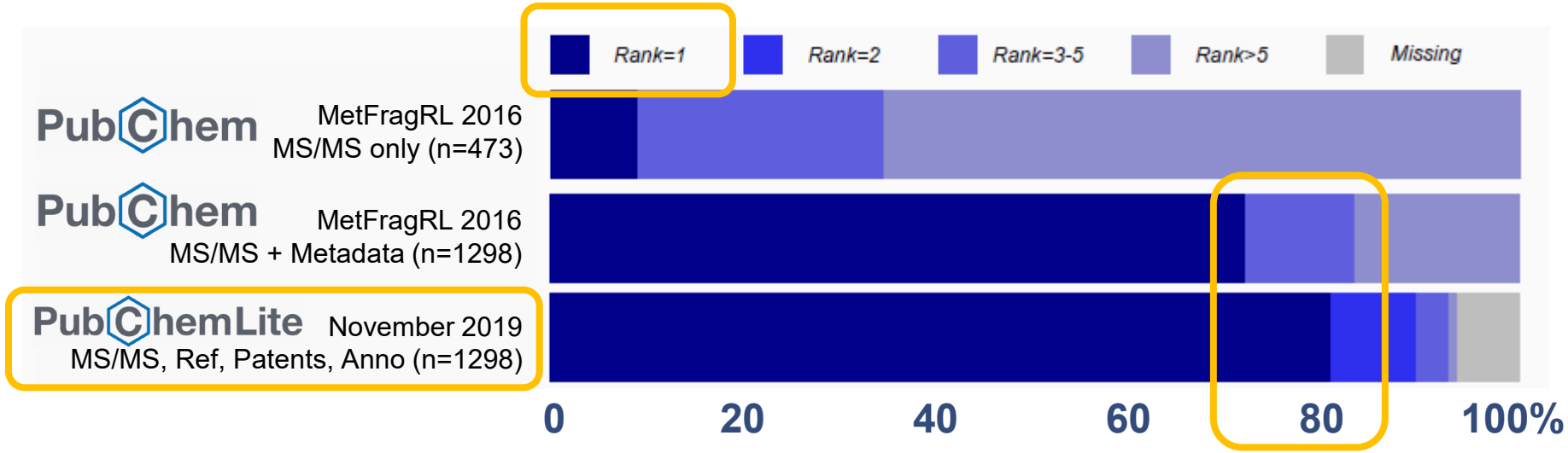


PubChemLite v0.2.0

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

How does PubChemLite perform?

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

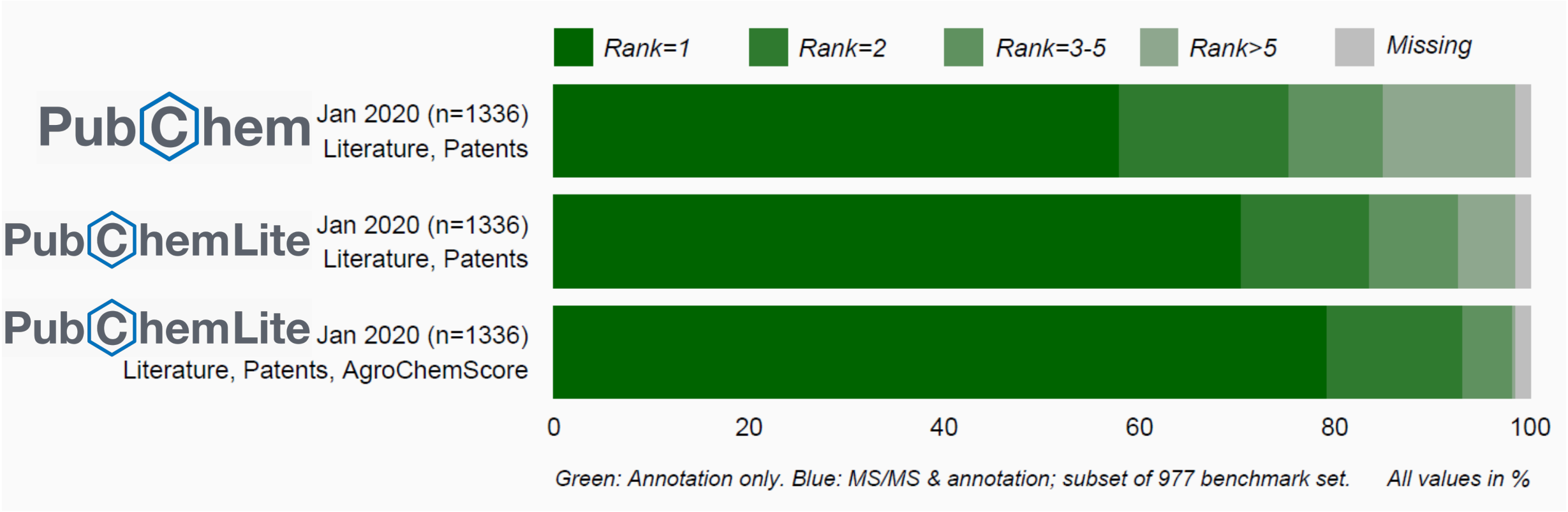


>80 % ranked in first place
~90 % ranked first or second!

} high throughput identification

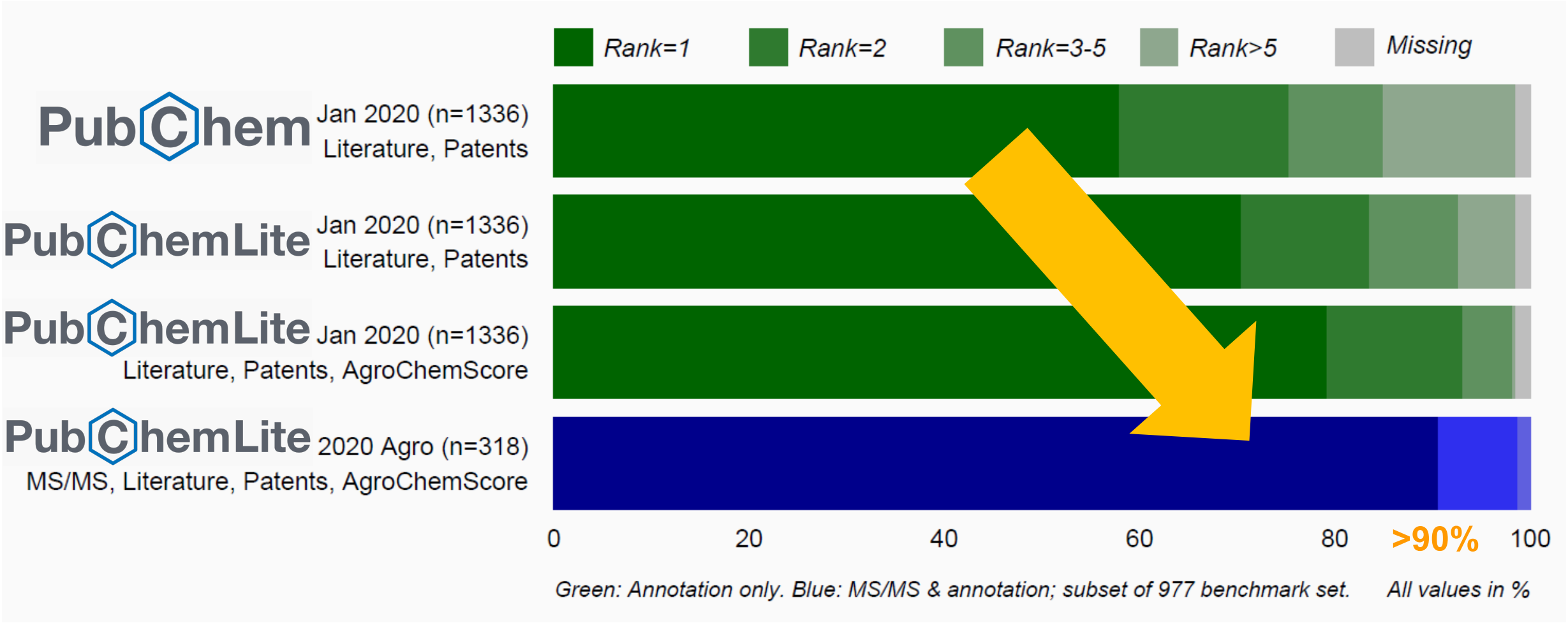
Ruttkies, Schymanski *et al*. (2016) DOI: 10.1186/s13321-016-0115-9 Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

Influence of the Annotation Content in PubChemLite



Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

Influence of the Annotation Content in PubChemLite

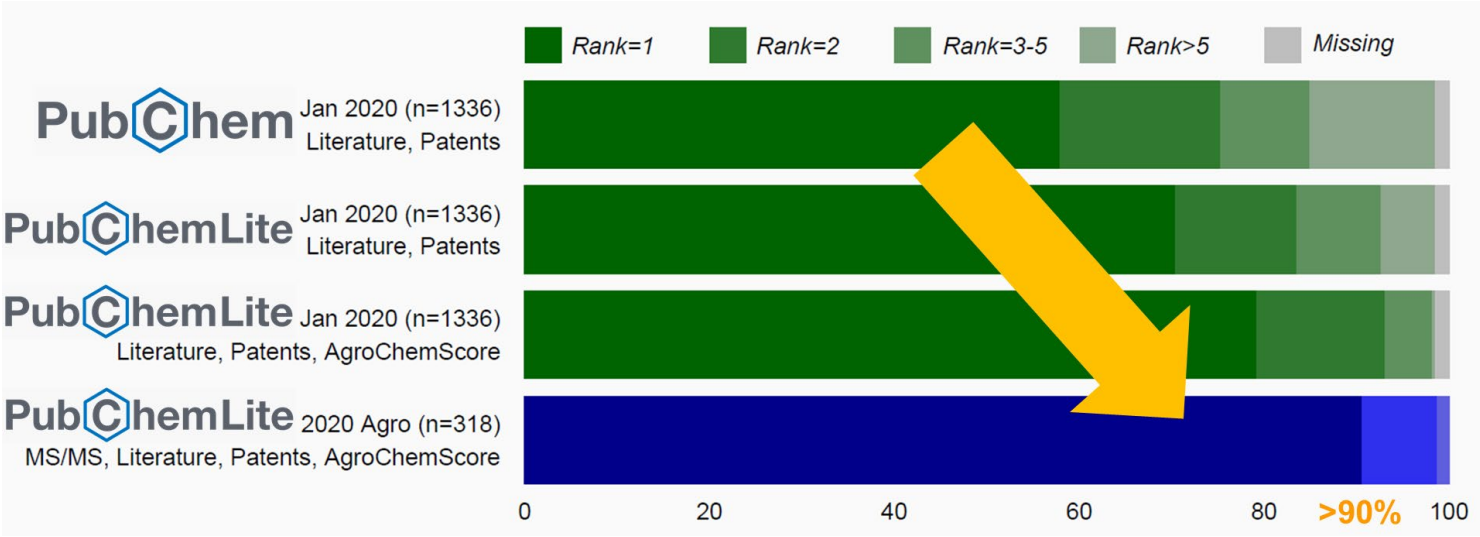


Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

PubChem

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

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!)

Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0

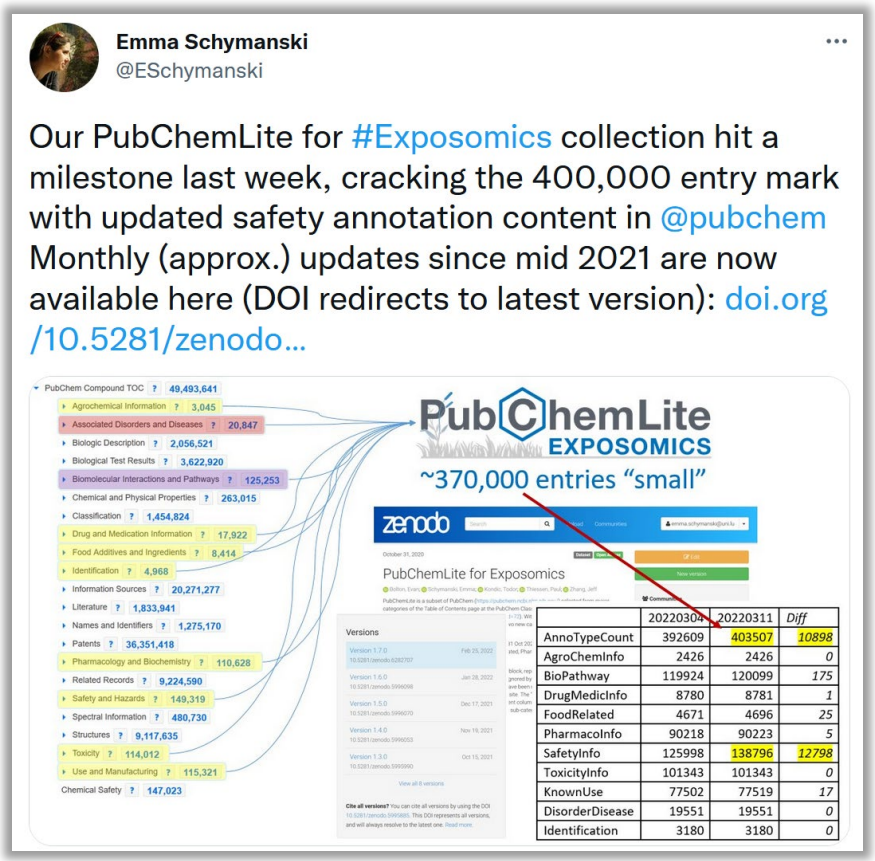
PubChem

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

uni.lu | LCSB

PubChemLite is a Dynamic Collection …

- … built every week, updated online every month
- … re-evaluated with every build for continuous quality control
- … growing as knowledge grows!



Emma Schymanski
@ESchymanski

Our PubChemLite for #Exposomics collection hit a milestone last week, cracking the 400,000 entry mark with updated safety annotation content in @pubchem Monthly (approx.) updates since mid 2021 are now available here (DOI redirects to latest version): doi.org /10.5281/zenodo...

	20220304	20220311	Diff
AnnoTypeCount	392609	403507	10898
AgroChemInfo	2426	2426	0
BioPathway	119924	120099	175
DrugMedicInfo	8780	8781	1
FoodRelated	4671	4696	25
PharmacoInfo	90218	90223	5
SafetyInfo	125998	138796	12798
ToxicityInfo	101343	101343	0
KnownUse	77502	77519	17
DisorderDisease	19551	19551	0
Identification	3180	3180	0

https://twitter.com/ESchymanski/status/1503260666613932033

Versions

Version 1.19.0 10.5281/zenodo.7684618	Feb 28, 2023
Version 1.18.0 10.5281/zenodo.7576412	Jan 27, 2023
Version 1.17.0 10.5281/zenodo.7498611	Jan 2, 2023
Version 1.16.0 10.5281/zenodo.7364484	Nov 26, 2022
Version 1.15.0 10.5281/zenodo.7261187	Oct 28, 2022

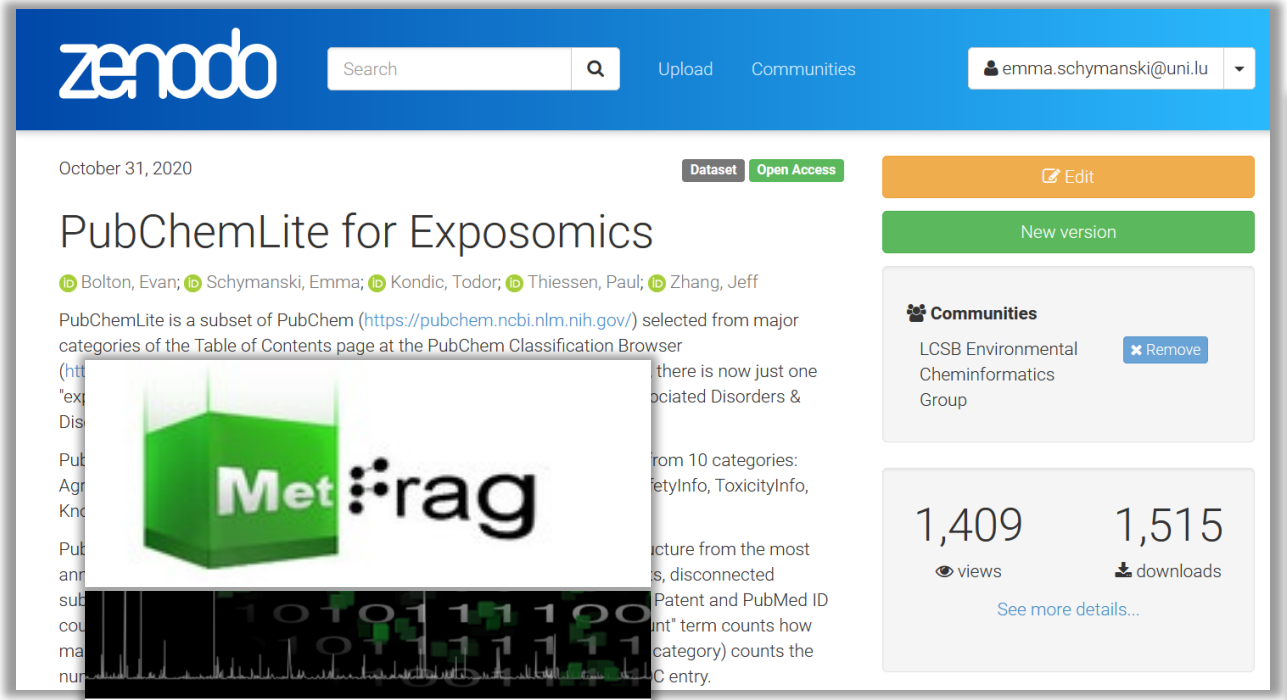
View all 20 versions

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.5995885. This DOI represents all versions, and will always resolve to the latest one. Read more.

DOI: 10.5281/zenodo.5995885

Putting PubChemLite into Practice?

- Downloadable & fully integrated in patRoon & MetFrag



DOI: 10.5281/zenodo.5995885

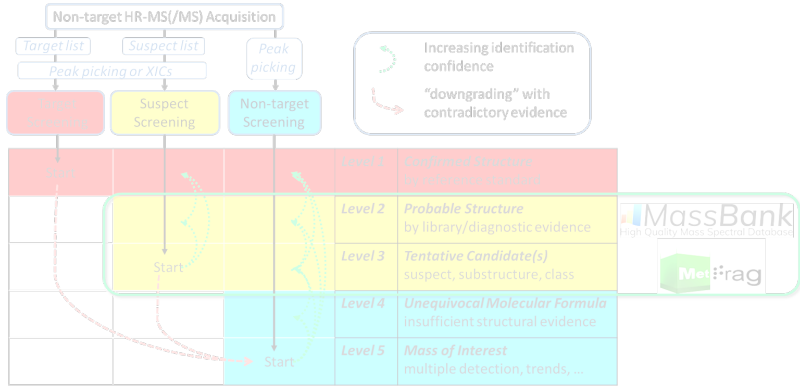
https://rickhelmus.github.io/patRoon/

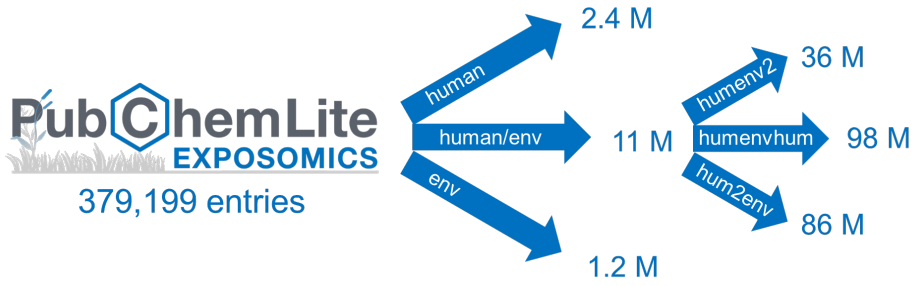
https://ipb-halle.github.io/MetFrag/

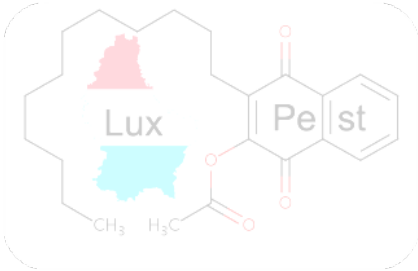
Schymanski *et al.* (2021) DOI: 10.1186/s13321-021-00489-0 & DOI: 10.5281/zenodo.5995885
Helmus *et al.* (2021) patRoon *J Cheminform* **13,** 1. DOI: 10.1186/s13321-020-00477-w

Outline

- Identifying Chemicals with NT-HRMS
  - Open Resources for Identification
  - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases
  - Defining Relevant Chemical Space
  - Transformations: Known & Predicted
- Applications in Environmental Studies
  - Including a little bit of CCS
- Perspectives



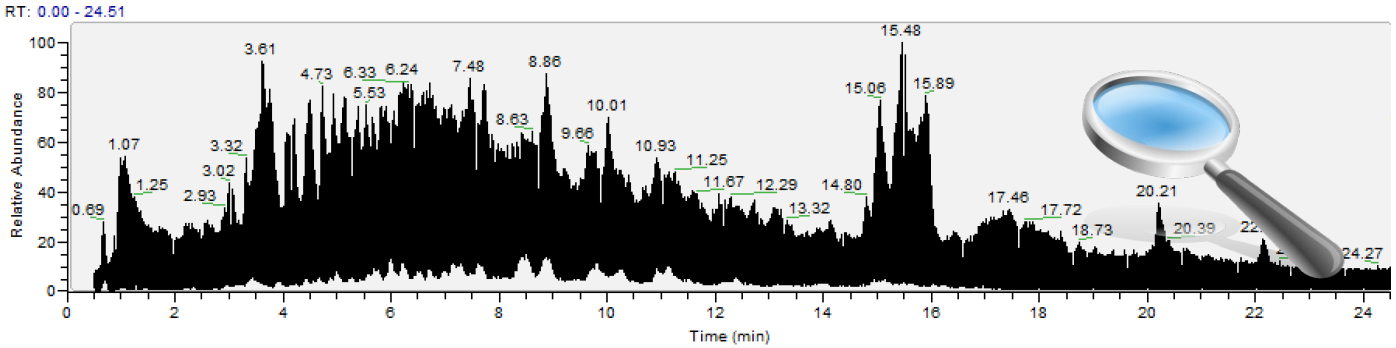


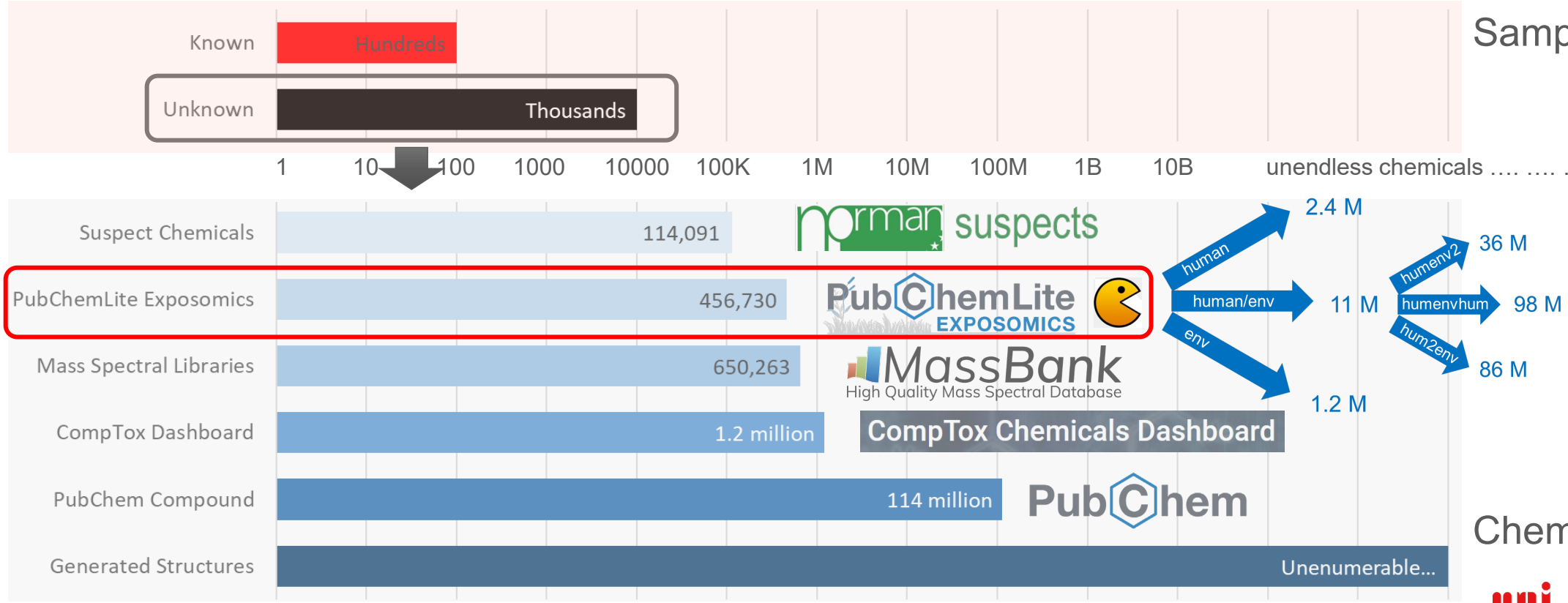


Environmental Cheminformatics, NT HR-MS & *Transformations*

High resolution mass spectrometry

AND connecting chemical knowledge





Chromatogram: Schymanski *et al*. (2014) DOI: 10.1021/es4044374; Resource numbers updated 15 March 2023.

Transforming PubChemLite with BioTransformer 3.0

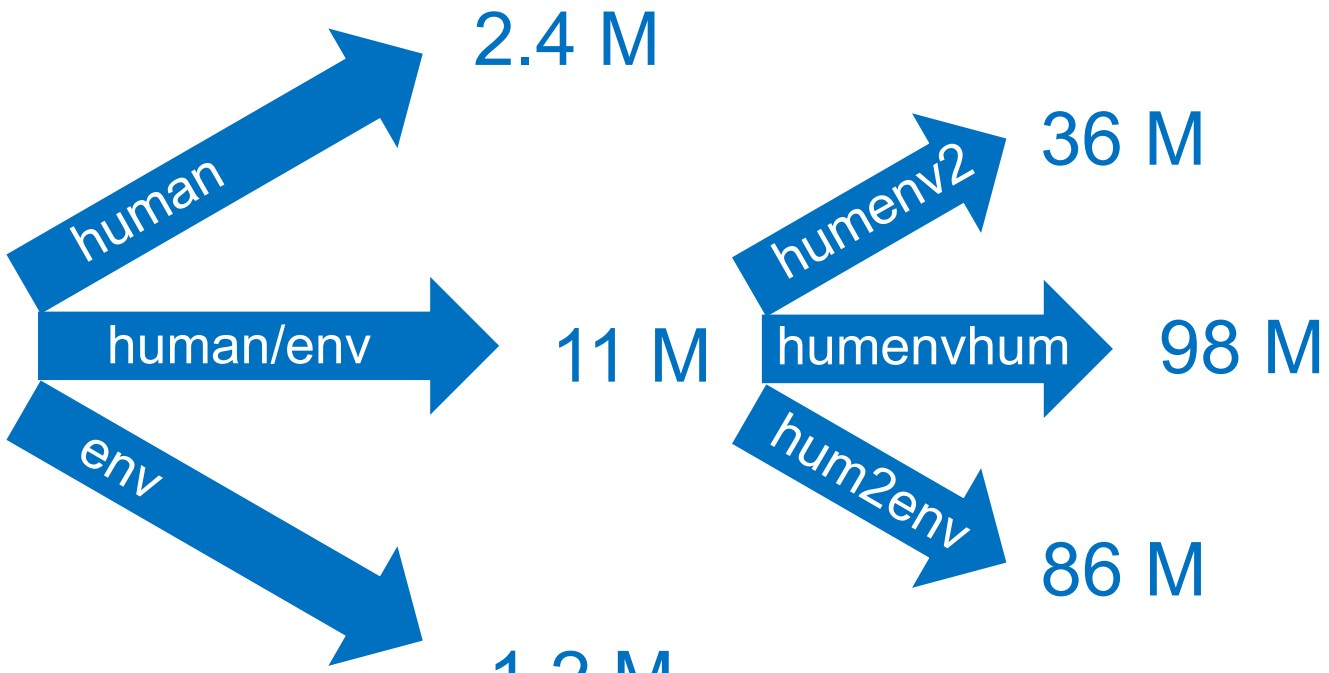
http://biotransformer.ca/

Wishart Lab
UNIVERSITY OF ALBERTA

PubChemLite EXPOSOMICS

379,199 entries

version 1.0.0; DOI: 10.5281/zenodo.5995886



Combinatorial explosion – more data needed to predict reaction likelihoods

Djoumbou Feunang *et al* (2019). BioTransformer, JCheminf. DOI: 10.1186/s13321-018-0324-5.
Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0; Bolton *et al* (2021) *preliminary calculations;*
Schymanski, Bolton, Cheng, Thiessen, Zhang, Helmus (2021) Transformations in PubChem, DOI: 10.5281/zenodo.5644560

FAIR Transformations in PubChem and NORMAN-SLE

norman suspects

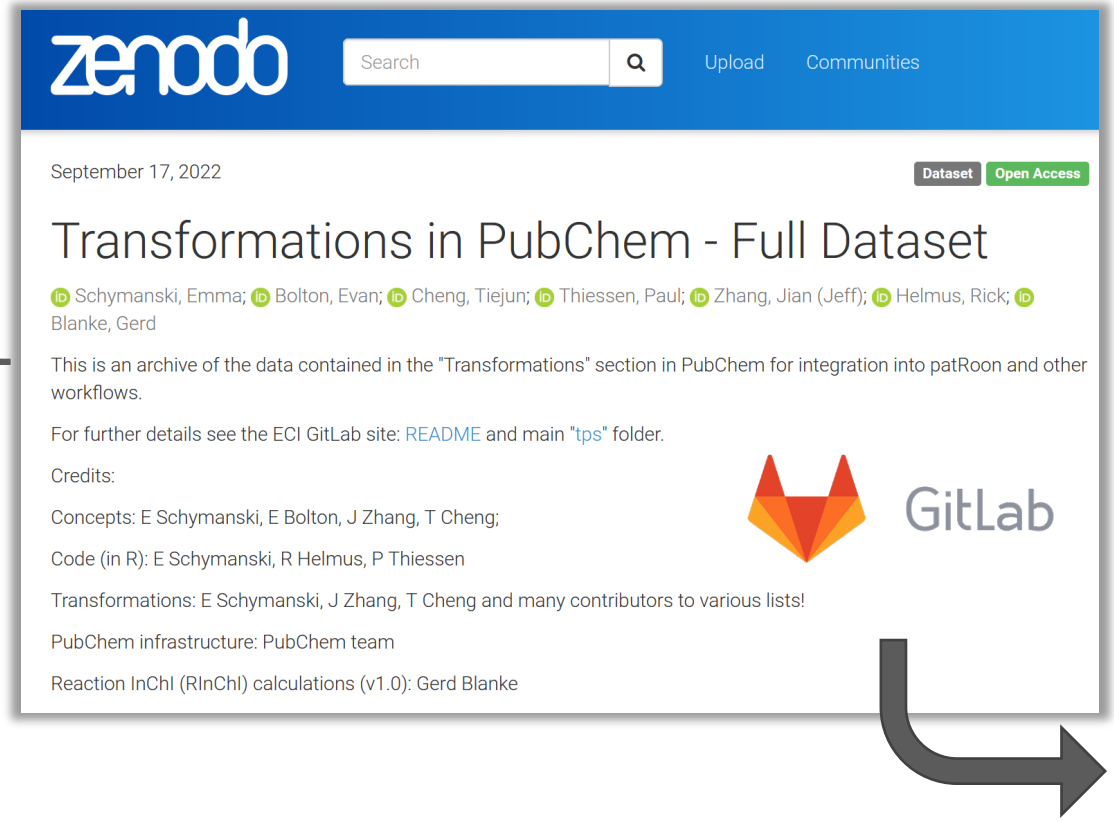
- EAWAGTPS
- HSDB
- MetXBioDB
- REFTPS
- SLUPESTTPS
- SWISSPEST19
- THSTPS
- UACCSCEC

ChEMBL

PubChem Transformations

6,130 chemicals

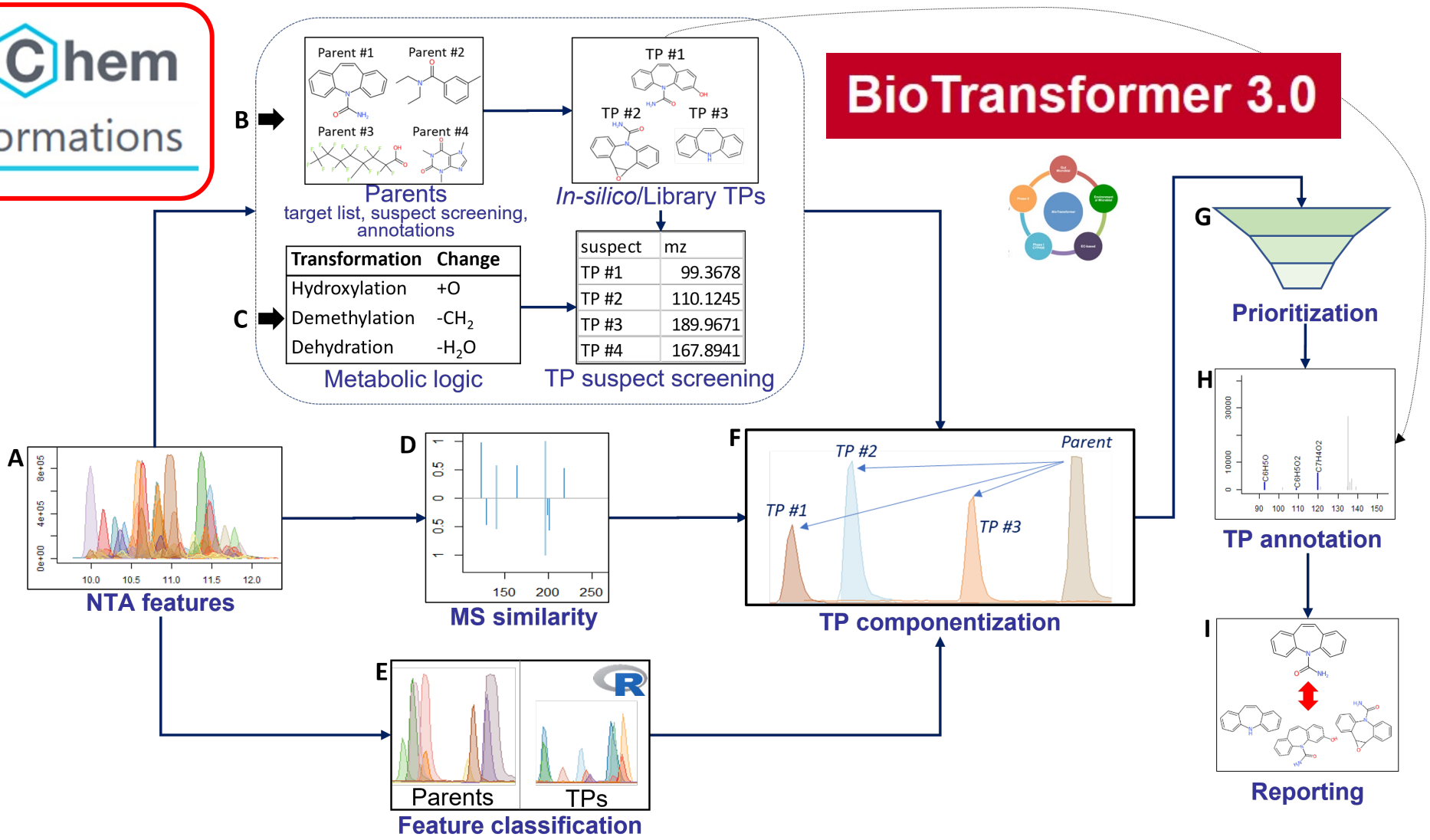
>double the BioTransformer library



Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0; Helmus *et al* (2022) DOI: 10.21105/joss.04029
Schymanski, Bolton, Cheng, Thiessen, Zhang, Helmus (2021) Transformations in PubChem, DOI: 10.5281/zenodo.5644560

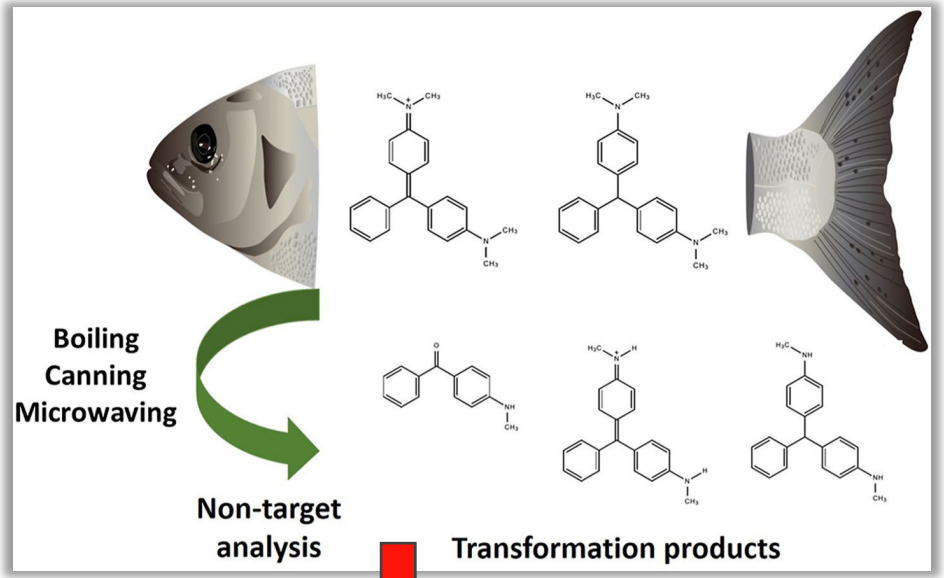
Open Transformation Products Workflows in patRoon 2.0

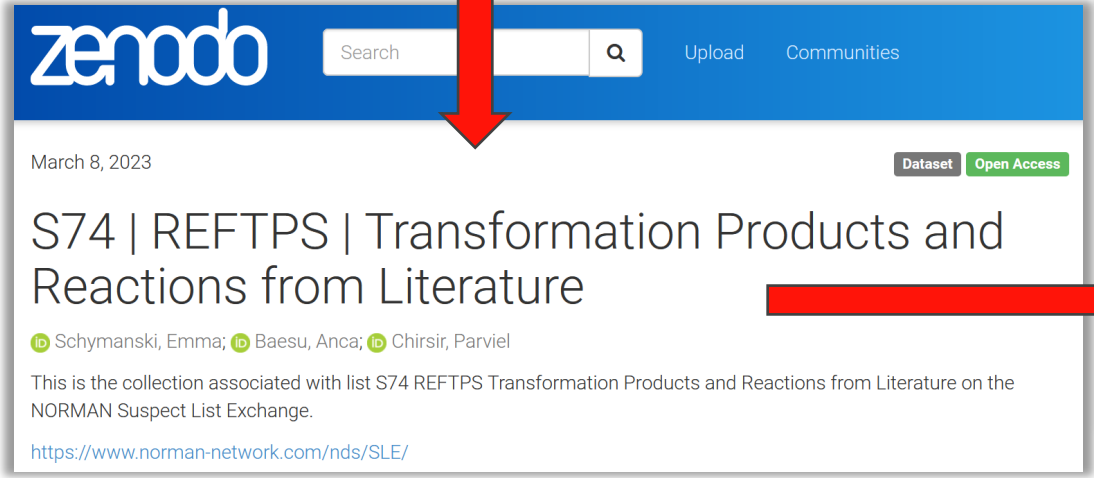
PubChem Transformations

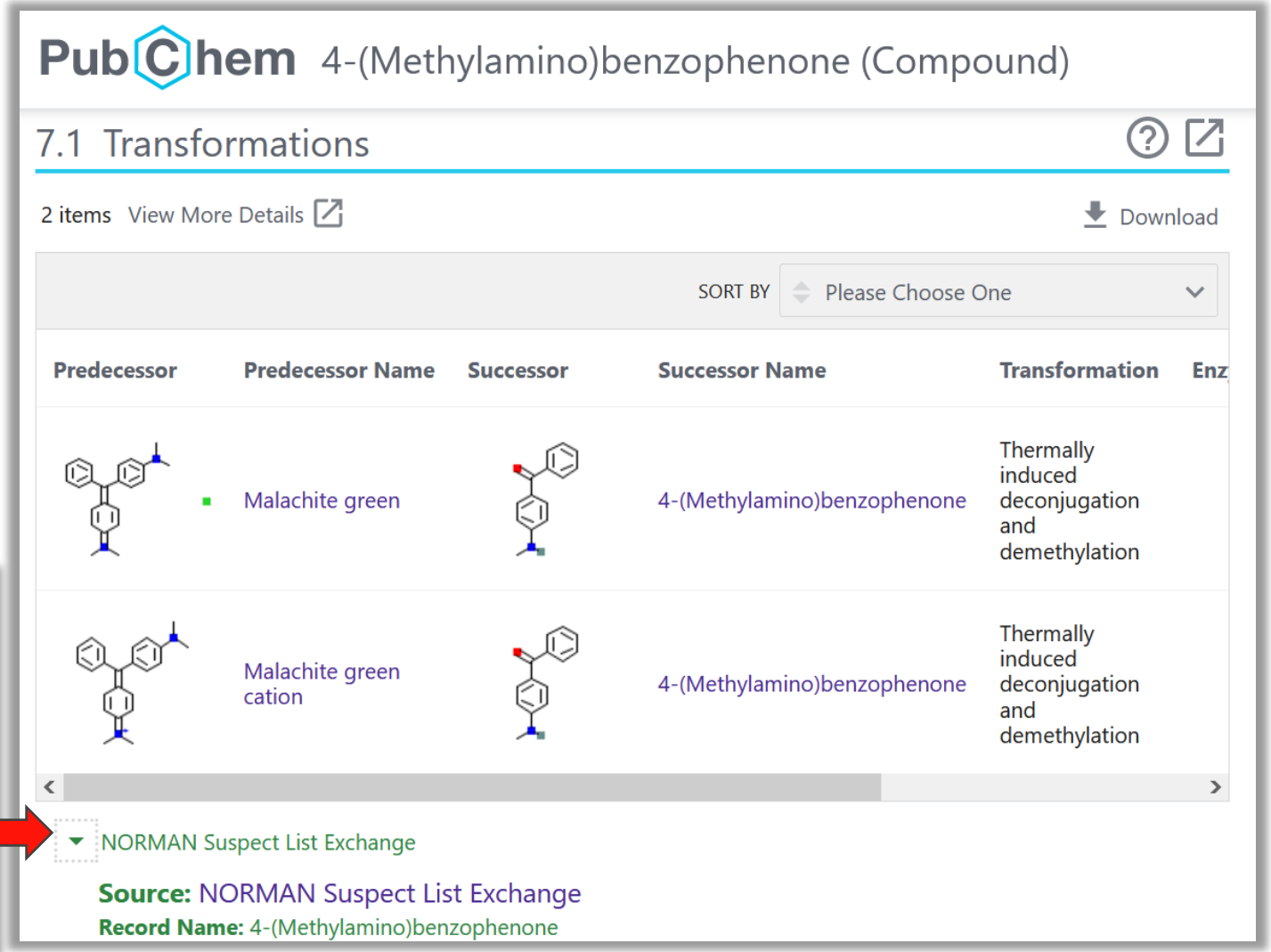


Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0; Helmus *et al* (2022) DOI: 10.21105/joss.04029
Schymanski, Bolton, Cheng, Thiessen, Zhang, Helmus (2021) Transformations in PubChem, DOI: 10.5281/zenodo.5644560

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







Baesu *et al*. (2021) DOI: 10.1016/j.crfs.2021.09.010. S74 | REFTPS (2021) DOI: 10.5281/zenodo.4318838

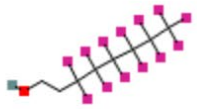
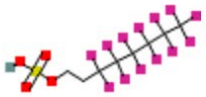
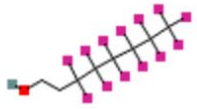
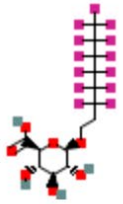
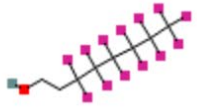
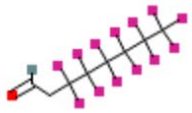
FAIRifying PFAS

PubChem 3,3,4,4,5,5,6,6,7,7,8,8,8-Tridecafluorooctan-1-ol (

7.1 Transformations

10 items View More Rows & Details

SORT BY Please Choose One

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzym
	6:2 FTOH		6:2 FTOH-Sulfate	Conjugation	
	6:2 FTOH		6:2 FTOH-Gluc	Conjugation	1(
	6:2 FTOH		6:2 FTAL	Oxidation	24296 1(

https://pubchem.ncbi.nlm.nih.gov/compound/69537#section=Pharmacology-and-Biochemistry

ELSEVIER

Chemico-Biological Interactions

Volume 155, Issue 3, 15 August 2005, Pages 165-180

Metabolic products and pathways of fluorotelomer alcohols in isolated rat hepatocytes

https://doi.org/10.1016/j.cbi.2005.06.007

Jonathan W. Martin [a], Scott A. Mabury [b], Peter J. O'Brien [a]

Show more

+ Add to Mendel

https://doi.org/10.10

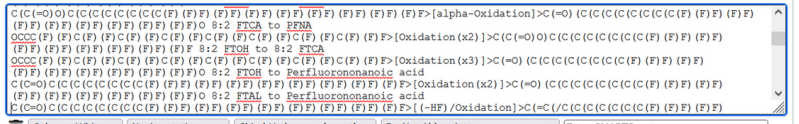
Jon Martin
@AcademicTox

Wow, this was a surprise. My postdoctoral work on metabolism of #PFAS has been FAIRified on @pubchem thanks to @ESchymanski and colleagues. Very exciting to think about how these huge efforts will improve future science for everyone.

@PChirsir

Emma Schymanski @ESchymanski · Jan 8

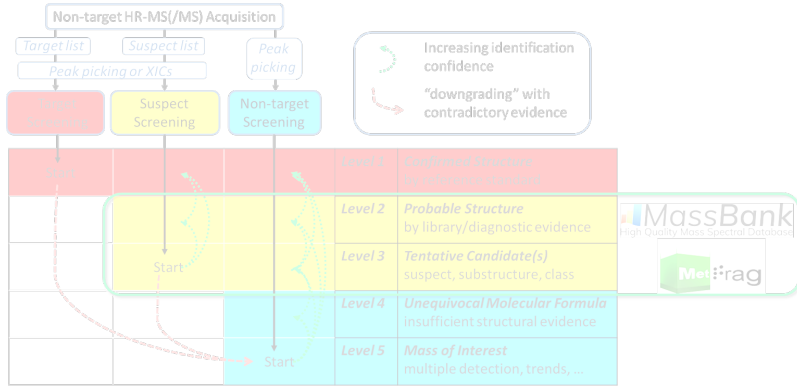
Spent a rainy afternoon FAIRifying #PFAS data from past @AcademicTox work, thanks to @PChirsir efforts we've committed the first set to @pubchem incl. new structures - hope to bring the 2005 PDF images to #OpenScience life soon! Again @the_cdk Depict makes it so much fun @jwmay !

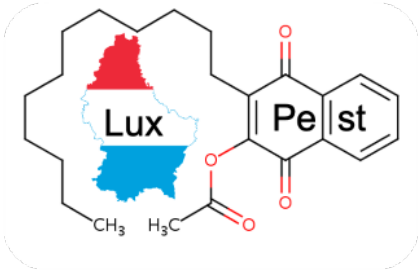


https://twitter.com/AcademicTox/status/1479841982679765001

Outline

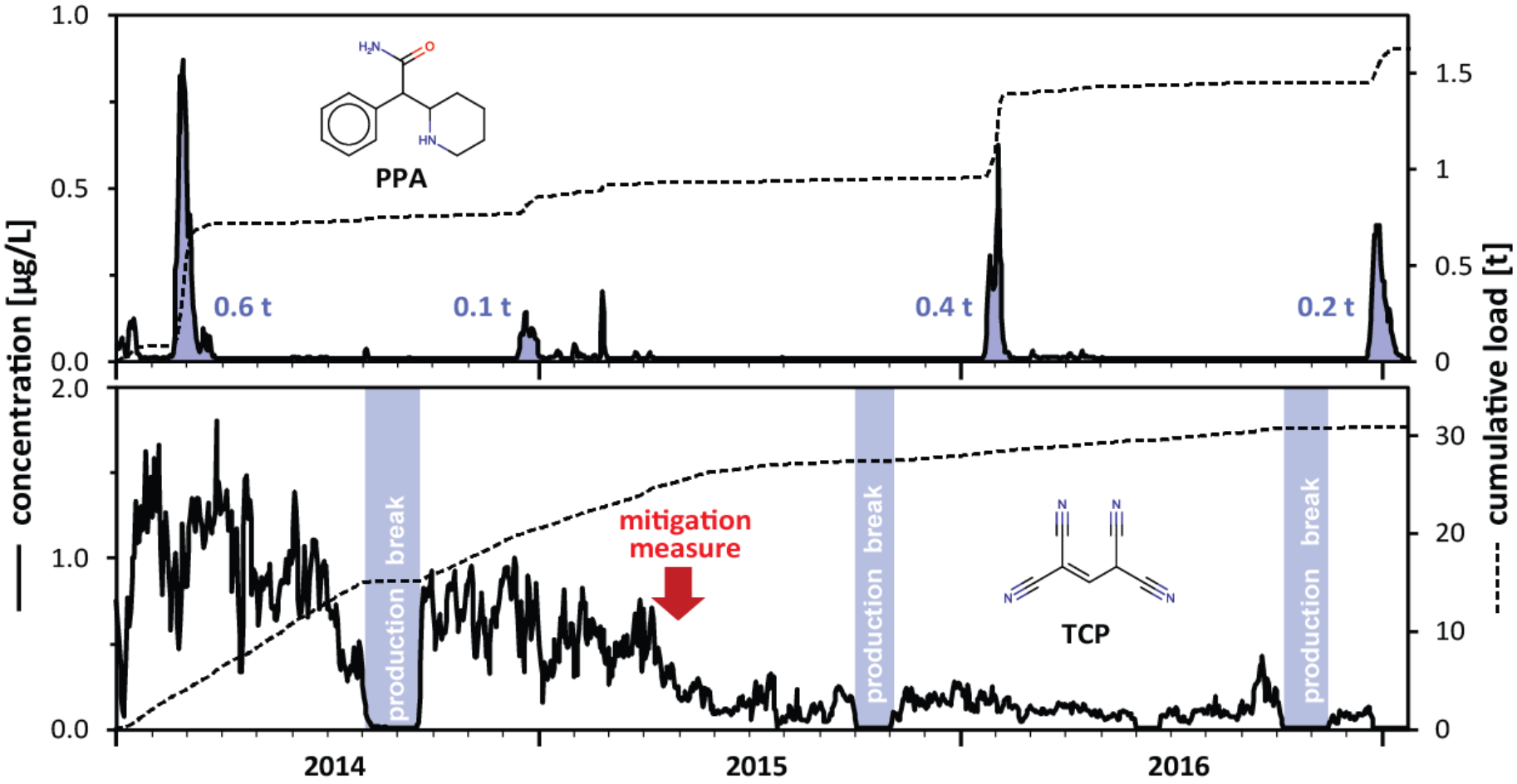
- Identifying Chemicals with NT-HRMS
  - Open Resources for Identification
  - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases
  - Defining Relevant Chemical Space
  - Transformations: Known & Predicted
- Applications in Environmental Studies
  - Including a little bit of CCS
- Perspectives





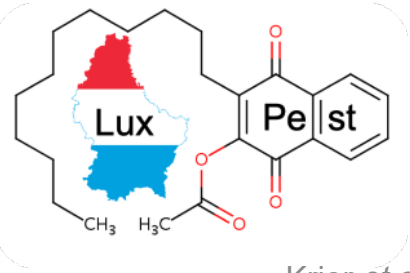
Examples: Real Time Monitoring of the Rhine River


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

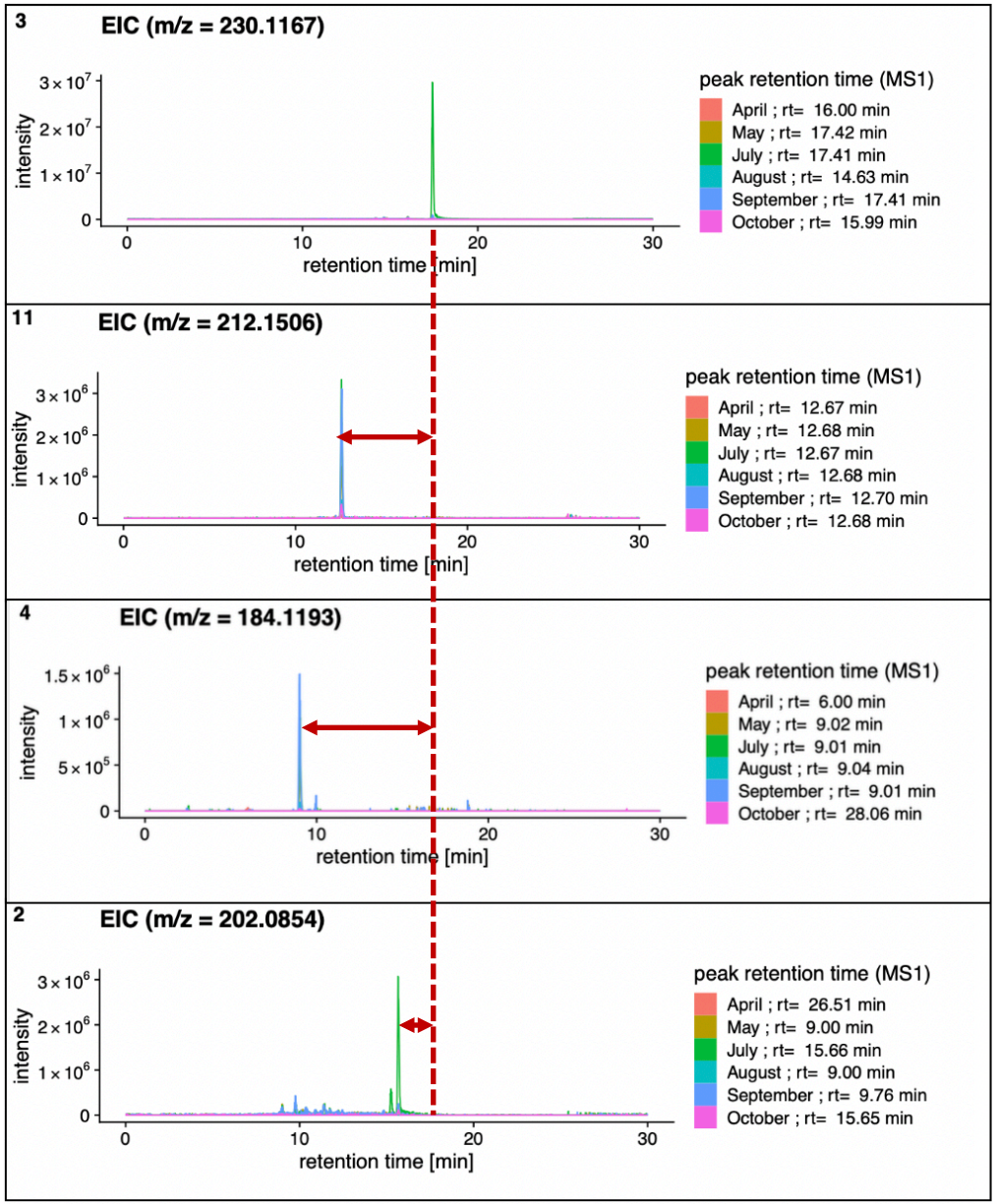


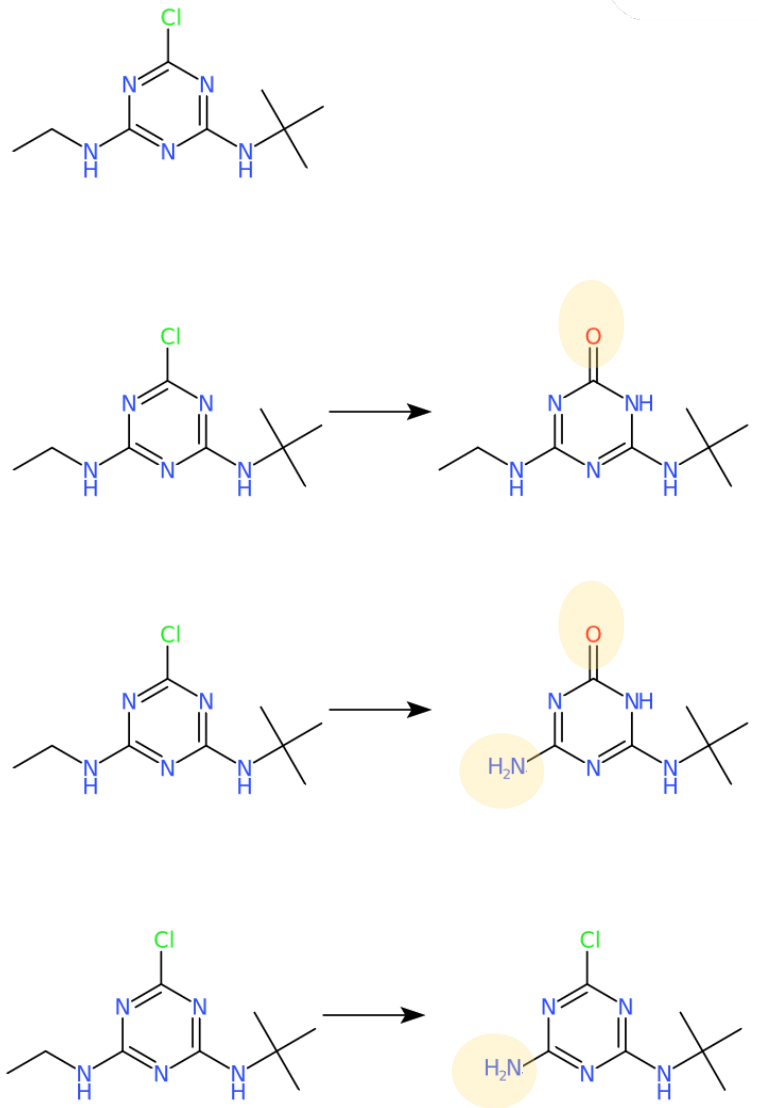
Hollender, Schymanski, Singer & Ferguson (2017) ES&T Feature, 51:20, 11505-11512. DOI: 10.1021/acs.est.7b02184

Examples: LuxPest and Transformation Products

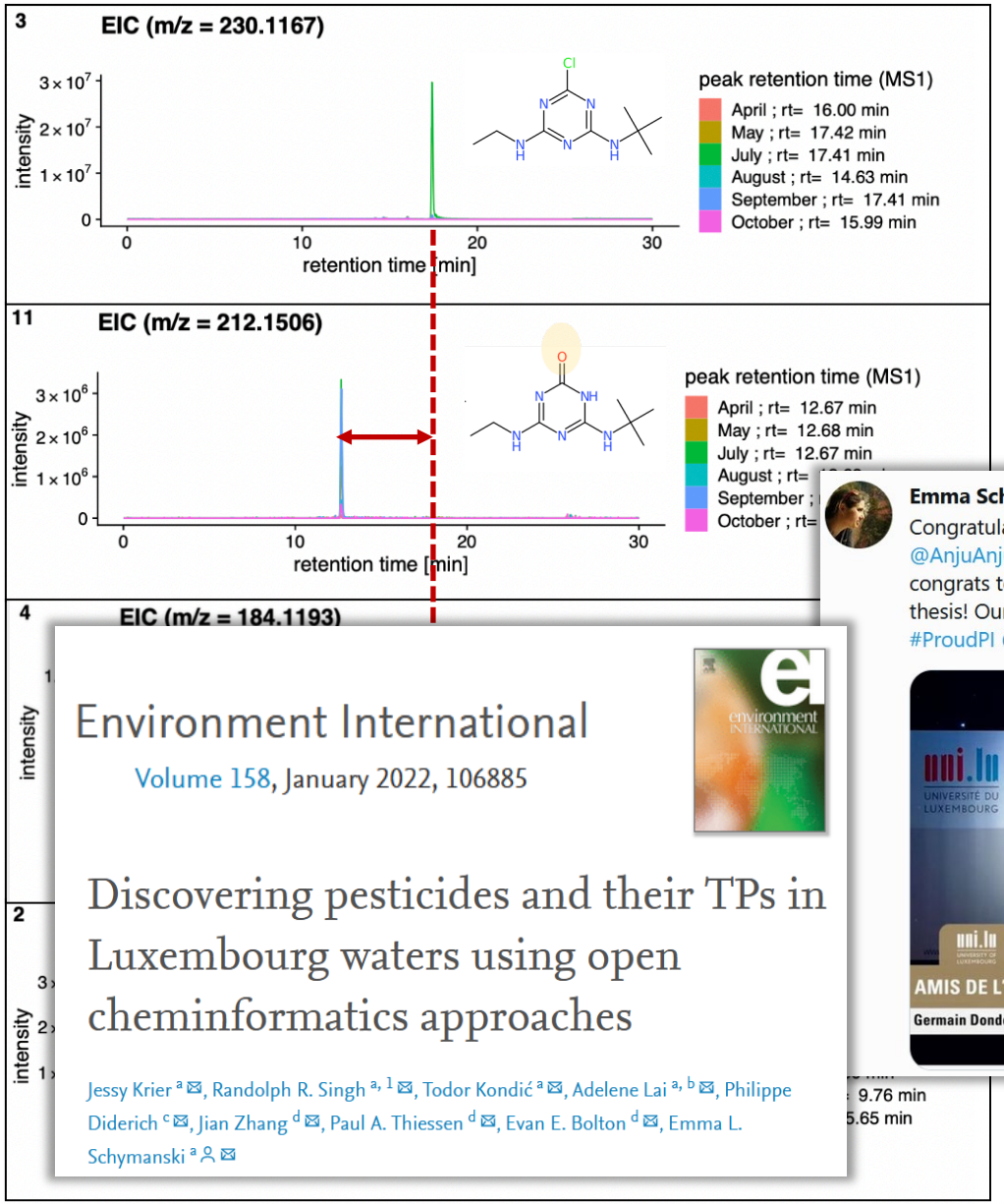


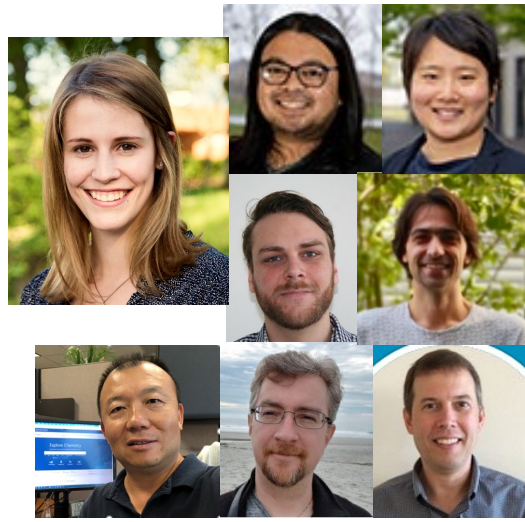
Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885





Pesticides & TPs in Luxembourg





Environment International

Volume 158, January 2022, 106885

Discovering pesticides and their TPs in Luxembourg waters using open cheminformatics approaches

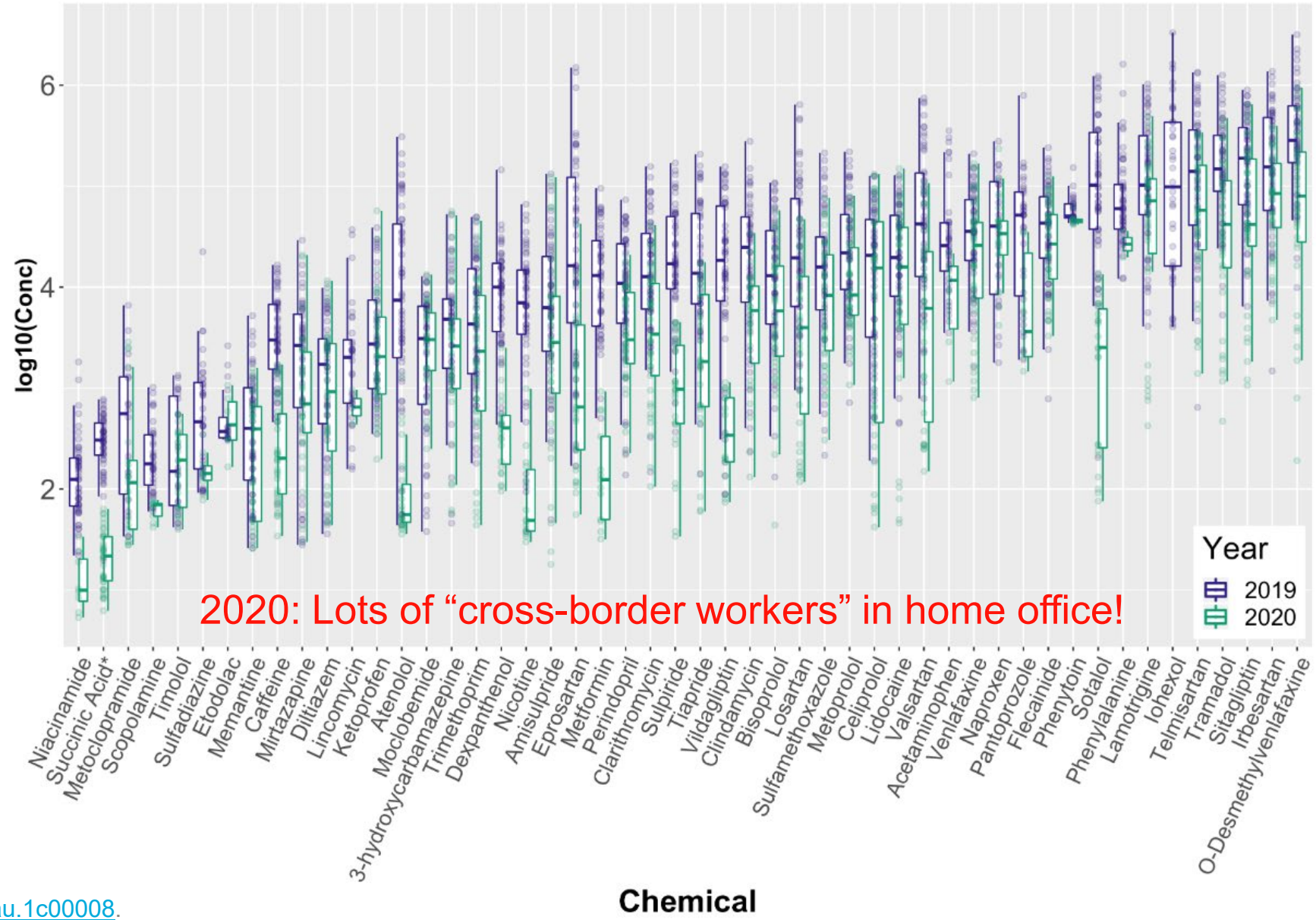
Jessy Krier [a] ✉, Randolph R. Singh [a, 1] ✉, Todor Kondić [a] ✉, Adelene Lai [a, b] ✉, Philippe Diderich [c] ✉, Jian Zhang [d] ✉, Paul A. Thiessen [d] ✉, Evan E. Bolton [d] ✉, Emma L. Schymanski [a] ✉

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

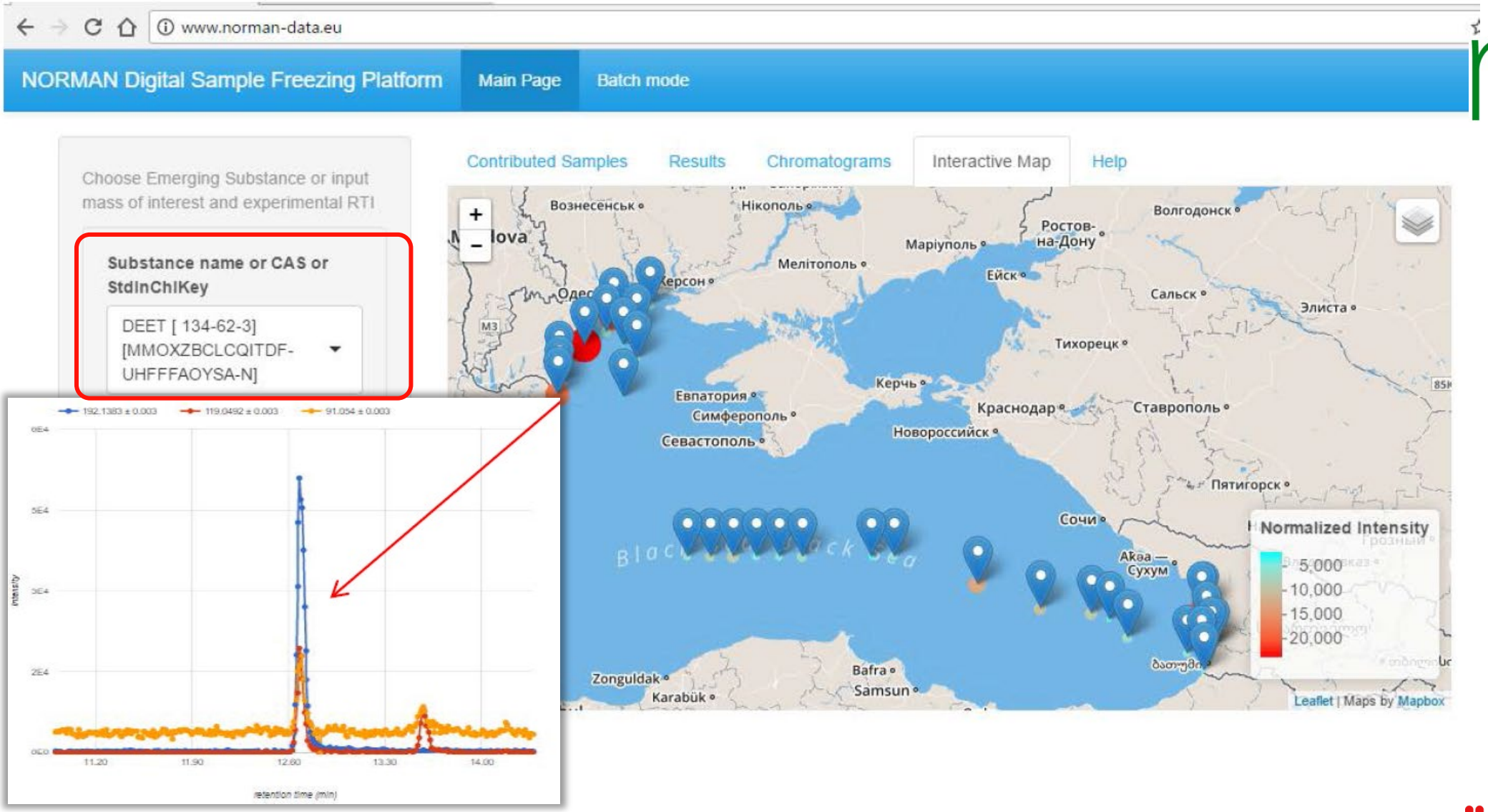
Krier *et al* (2022). DOI: 10.1016/j.envint.2021.106885.

Examples: LuxPharma – 2019 versus 2020 & COVID?



Singh *et al* (2021).
DOI: [10.1021/acsenvironau.1c00008](https://doi.org/10.1021/acsenvironau.1c00008).

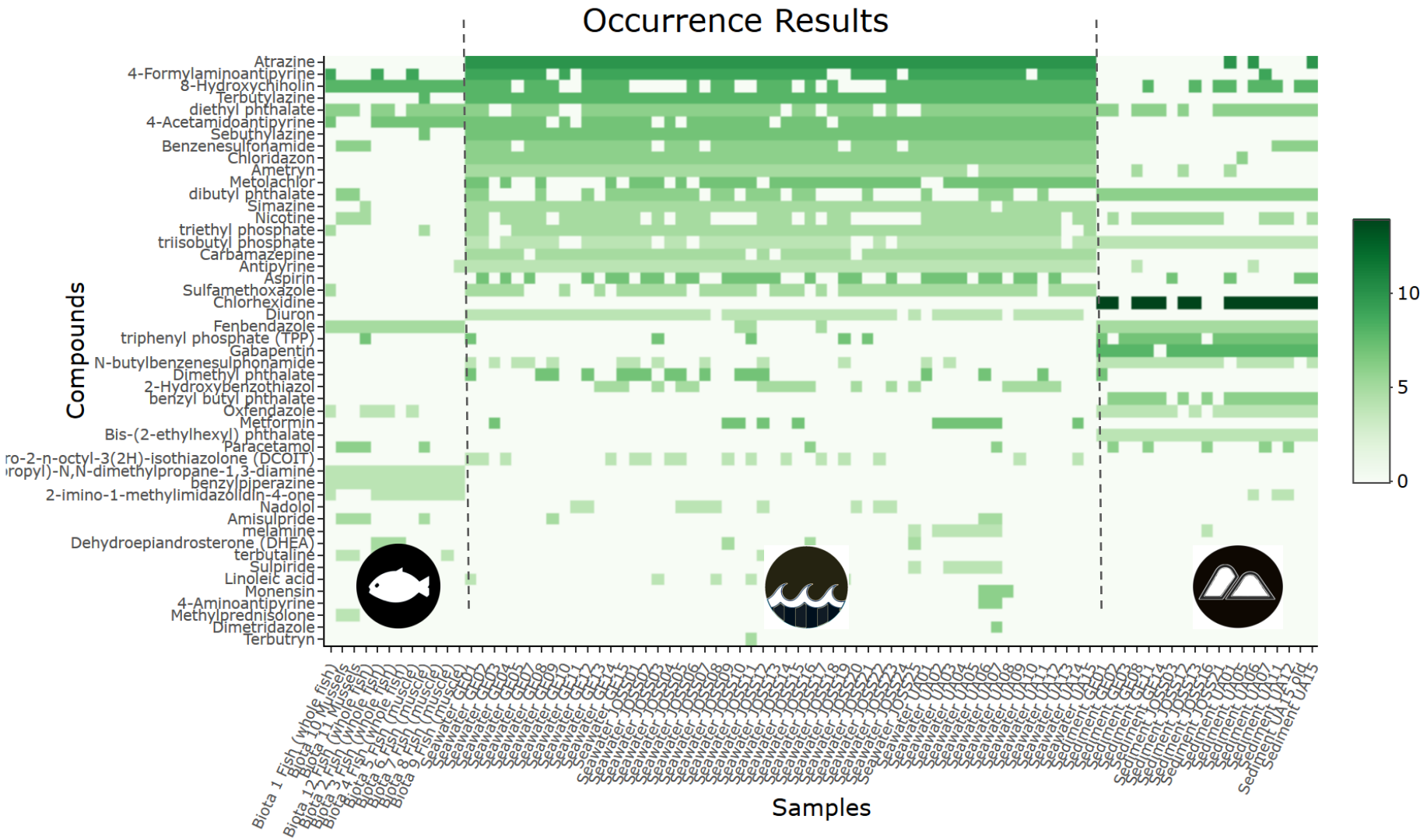
NORMAN DS – Digital Sample Freezing Platform



http://norman-data.eu/ and Alygizakis *et al*, 2019, TrAC, DOI: 10.1016/j.trac.2019.04.008

Examples: Distribution of Chemicals in Various Matrices

Retrospective screening of REACH chemicals in Black Sea samples

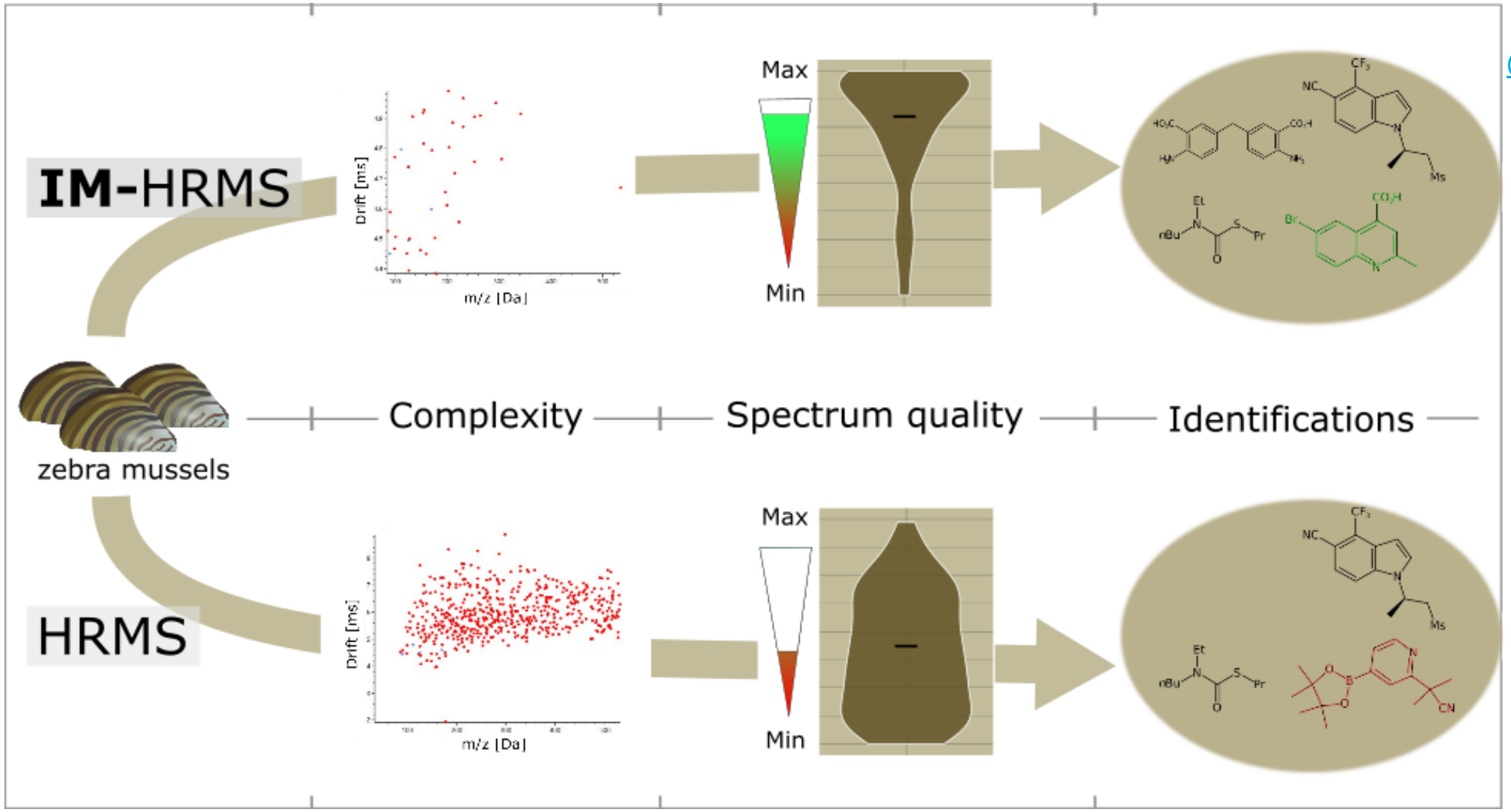


Alygizakis NA *et al* (2019) DOI: [10.1016/j.trac.2019.04.008](https://doi.org/10.1016/j.trac.2019.04.008)

Interactive heat map at http://norman-data.eu/NORMAN-REACH

CCS & MusselFiesta: Cleaner spectra; better identification

@MengerAcademic



Menger *et al* (2022) Environment International. DOI: 10.1016/j.envint.2022.107585

Incorporating CCS & RT in ID Confidence

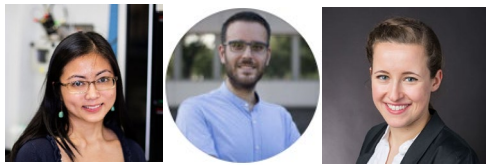
Target	Level 1. Confirmed structure with IP by reference standard	MS, MS^n (Precursor & diagnostic fragments)	RT (≤ 0.1 min)	CCS (≤ 2%)
Suspect	Level 2. Probable structure a) by library spectrum match b) by diagnostic evidence	MS, MS^n (from libraries) MS, MS^n (experimental data)	$RTI_{library}$ RTI, $RT_{Pred.}$	$CCS_{library}$ (≤ 2%) $CCS_{Pred.}$
Suspect	Level 3. Tentative candidate(s) structure, substituents, class	MS, MS^n (experimental data)	RTI, $RT_{Pred.}$	$CCS_{Pred.}$
Non-target	Level 4. Unequivocal molecular formula	MS isotope/adduct	-	CCS
Non-target	Level 5. Exact mass of interest	MS	-	CCS

Confidence

Celma, Sancho, Schymanski, Fabregat-Safont, Ibáñez, Goshawk, Barknowitz, Hernández & Bijlsma (2020) ES&T. DOI: 10.1021/acs.est.0c05713.

CCS Values in PubChem


PubChem Annotations


SLE Lists S50, S61, S79


CCSbase

PubChem Carbamazepine (Compound)

3.2.12 Collision Cross Section

150.3 Å^2 [M+H]$^+$ [CCS Type: TW, Method: Major Mix IMS/Tof Calibration Kit (Waters)]

https://www.sciencedirect.com/science/article/pii/S0021967318301894

▸ CCSbase

149 Å^2 [M+H]$^+$ [CCS Type: TW, Method: calibrated with polyalanine and drug standards]

https://pubs.acs.org/doi/abs/10.1021/acs.analchem.7b01709

▸ CCSbase

150 Å^2 [M+H]$^+$ [CCS Type: DT, Method: single field calibrated]

https://pubs.rsc.org/en/content/articlelanding/2018/ay/c7ay02808c

▸ CCSbase

149.11 Å^2 [M+H]$^+$

158.54 Å^2 [M+Na]$^+$

S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI | DOI:10.5281/zenodo.3549476

▸ NORMAN Suspect List Exchange

PubChem Atrazine (Compound)

3.2.16 Collision Cross Section

150.74 Å^2 [M+H]$^+$ [CCS Type: DT, Method: stepped-field]

https://pubs.rsc.org/en/content/articlelanding/2017/sc/c7sc03464d

▸ CCSbase

149.53 Å^2 [M+H]$^+$

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

▸ NORMAN Suspect List Exchange

149.26 Å^2 [M+H]$^+$

S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI | DOI:10.5281/zenodo.3549476

▸ NORMAN Suspect List Exchange

149.8 Å^2 [M+H]$^+$

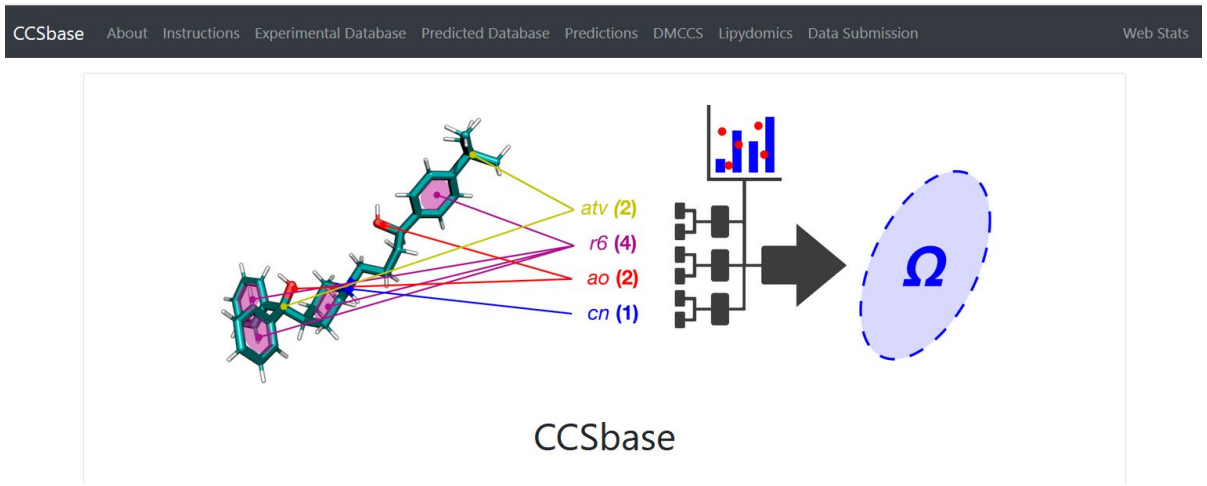
S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium | DOI:10.5281/zenodo.2658162

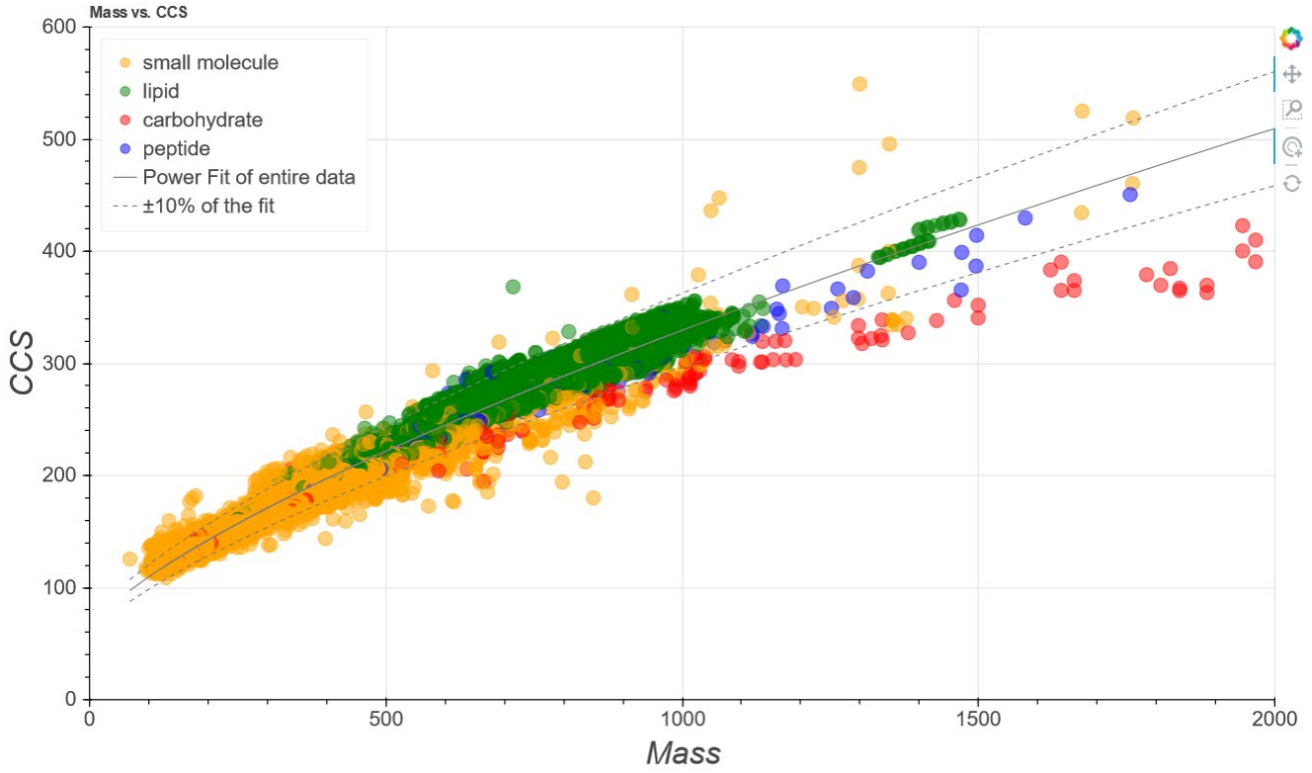
▸ NORMAN Suspect List Exchange

https://pubchem.ncbi.nlm.nih.gov/compound/2554#section=Collision-Cross-Section

https://pubchem.ncbi.nlm.nih.gov/compound/2256#section=Collision-Cross-Section

PubChemLite EXPOSOMICS … plus CCSbase!





Ross, D. H., Cho, J. H. & Xu, L. Anal. Chem. (2020). DOI: 10.1021/acs.analchem.9b05772.

UNIVERSITY of WASHINGTON

PubChemLite EXPOSOMICS … plus CCSbase!

zenodo Search Upload Communities emma.schymanski@uni.lu

January 22, 2021 Dataset Open Access

PubChemLite for Exposomics (1 Jan 2021) + predicted CCS from CCSbase

LCSB-ECI; Schymanski, Emma; Kondic, Todor; PubChem Team; Bolton, Evan; Thiessen, Paul; Zhang, Jeff; CCSbase Team; Krinsky, Ally; Ross, David H.; Xu, Libin

PubChemLite is a subset of PubChem (https://pubchem.ncbi.nlm.nih.gov/) selected from major categories of the Table of Contents page at the PubChem Classification Browser (https://pubchem.ncbi.nlm.nih.gov/classification/#hid=72). This version of PubChemLite for Exposomics (see original dataset here: DOI 10.5281/zenodo.4432124) has predicted collision cross section (CCS) values for 8 adducts provided by Libin Xu and team at CCSbase (https://ccsbase.net/).

PubChemLite ***exposomics*** is compiled from 10 categories: AgroChemInfo, BioPathway, DrugMedicInfo, FoodRelated, PharmacoInfo, SafetyInfo, ToxicityInfo, KnownUse, DisorderDisease, Identification

CCS adducts provided are: [M+H]+, [M+K]+, [M+NH4]+, [M+Na-2H]-, [M+Na]+, [M-H]-,[M]+,[M]-

Details on the CCS prediction are given here: Ross, D. H., Cho, J. H. & Xu, L. Anal. Chem. (2020). doi:10.1021/acs.analchem.9b05772

Edit

New version

Communities

LCSB Environmental Cheminformatics Group Remove

340 views 287 downloads

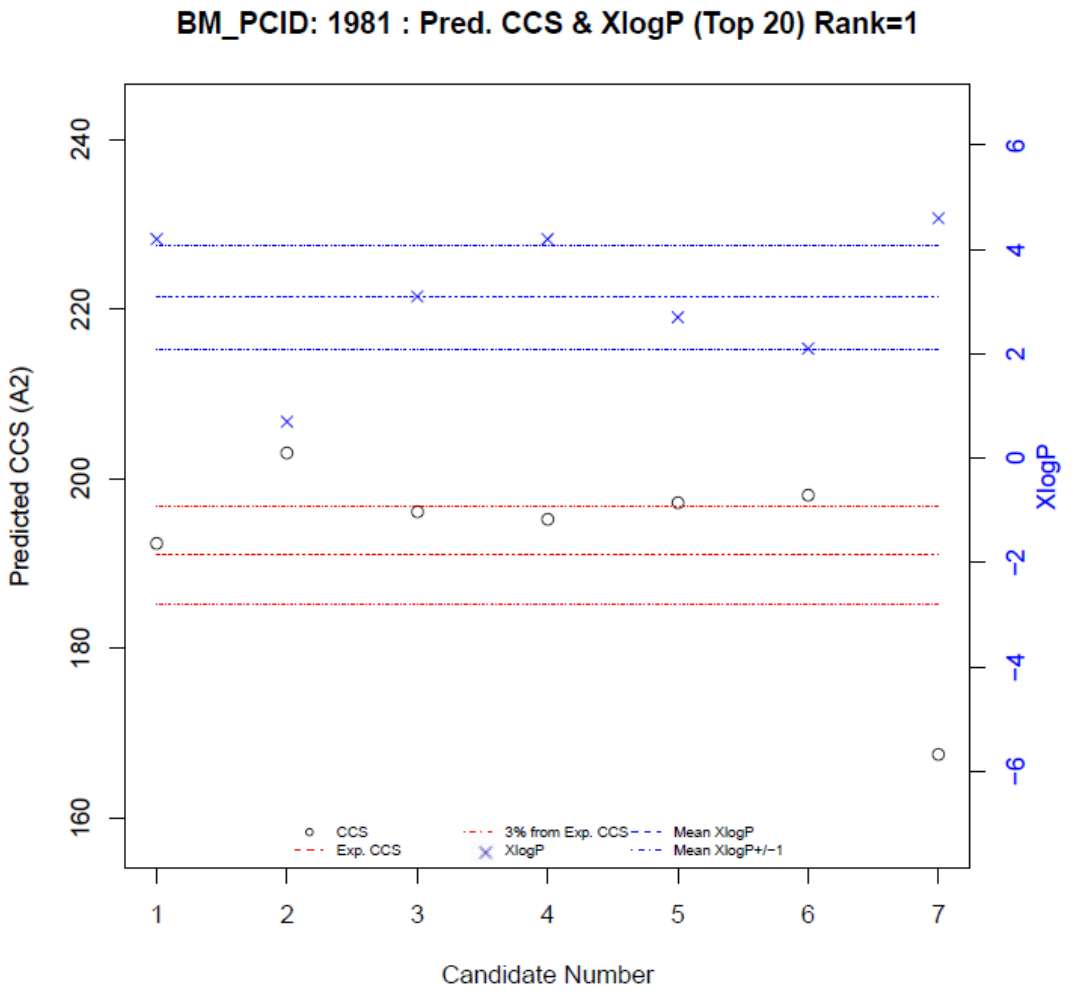
See more details...

LCSB-ECI, PubChem Team, Xu Lab CCSbase team. (2021). DOI: 10.5281/zenodo.4456208

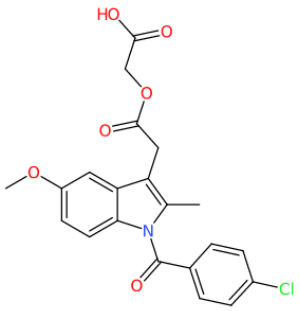
Ross, D. H., Cho, J. H. & Xu, L. Anal. Chem. (2020). DOI: 10.1021/acs.analchem.9b05772.

UNIVERSITY of WASHINGTON uni.lu LCSB

PubChemLite EXPOSOMICS … plus CCSbase!



PubChem Acemetacin (Compound)



3.2.6 Collision Cross Section

191 Å^2 [M+H]$^+$ [CCS Type: TW, Method: calibrated with polyalanine and drug standards

https://pubs.acs.org/doi/abs/10.1021/acs.analchem.7b01709

▸ CCSbase

#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1	2-[2-[1-(4-chlorobenzoyl)-5-methoxy-2-methylindol-3-yl]acetyl]oxyacetic acid	1981 InChIKeyBlock1 = FSQKKOOTNAMONP	415.08227	$C_{21}H_{18}ClNO_6$	1st 2nd 3rd 4th 5th 0.0 0.2 0.4 0.6 0.8 1.0	4.0	Peaks: 5 / 6 Fragments Scores Download
2	(2Z)-2-[(E)-(4,9-dimethoxy-5-oxofuro[3,2-g]chromen-7-yl)methylidenehydrazinylidene]-3-ethyl-1,3-thiazolidin-4-one	9588988 InChIKeyBlock1 = WPBSKCZFXIVOQF	415.08381	$C_{19}H_{17}N_3O_6S$	1st 2nd 3rd 4th 5th 0.0 0.2 0.4 0.6 0.8 1.0	0.6166	Peaks: 2 / 6 Fragments Scores Download

MetFrag

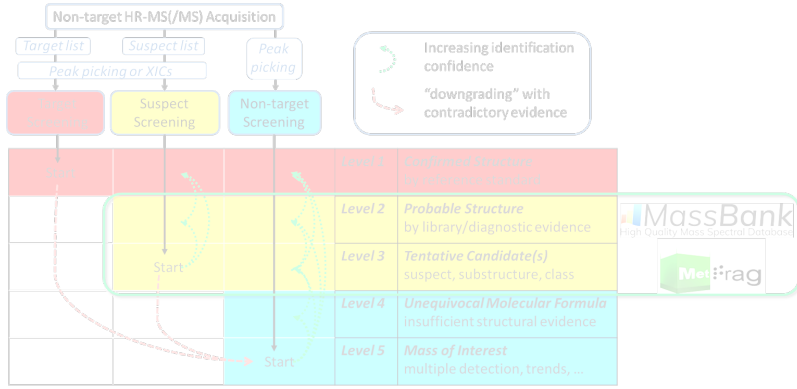
Weights		
MetFrag (1st)	100	%
ExactSpectralSimilarity (2nd)	100	%
Patent_Count (3rd)	100	%
PubMed_Count (4th)	100	%
pred_CCS_A2_[M+H]+ (5th)	0	%

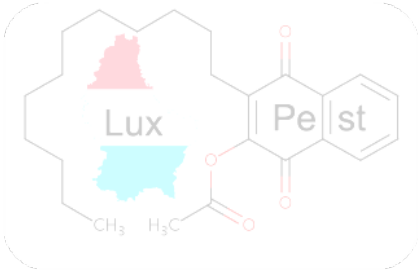
Schymanski *et al*. (2021) JCheminform. DOI: 10.1186/s13321-016-0115-9

Ross, D. H., Cho, J. H. & Xu, L. Anal. Chem. (2020). DOI: 10.1021/acs.analchem.9b05772.

Outline

- Identifying Chemicals with NT-HRMS
  - Open Resources for Identification
  - Workflows, Confidence, Limitations of MS/MS
- Dynamic Databases
  - Defining Relevant Chemical Space
  - Transformations: Known & Predicted
- Applications in Environmental Studies
  - Including a little bit of CCS
- Perspectives





Do we always need the entire chemical space?

PubChem

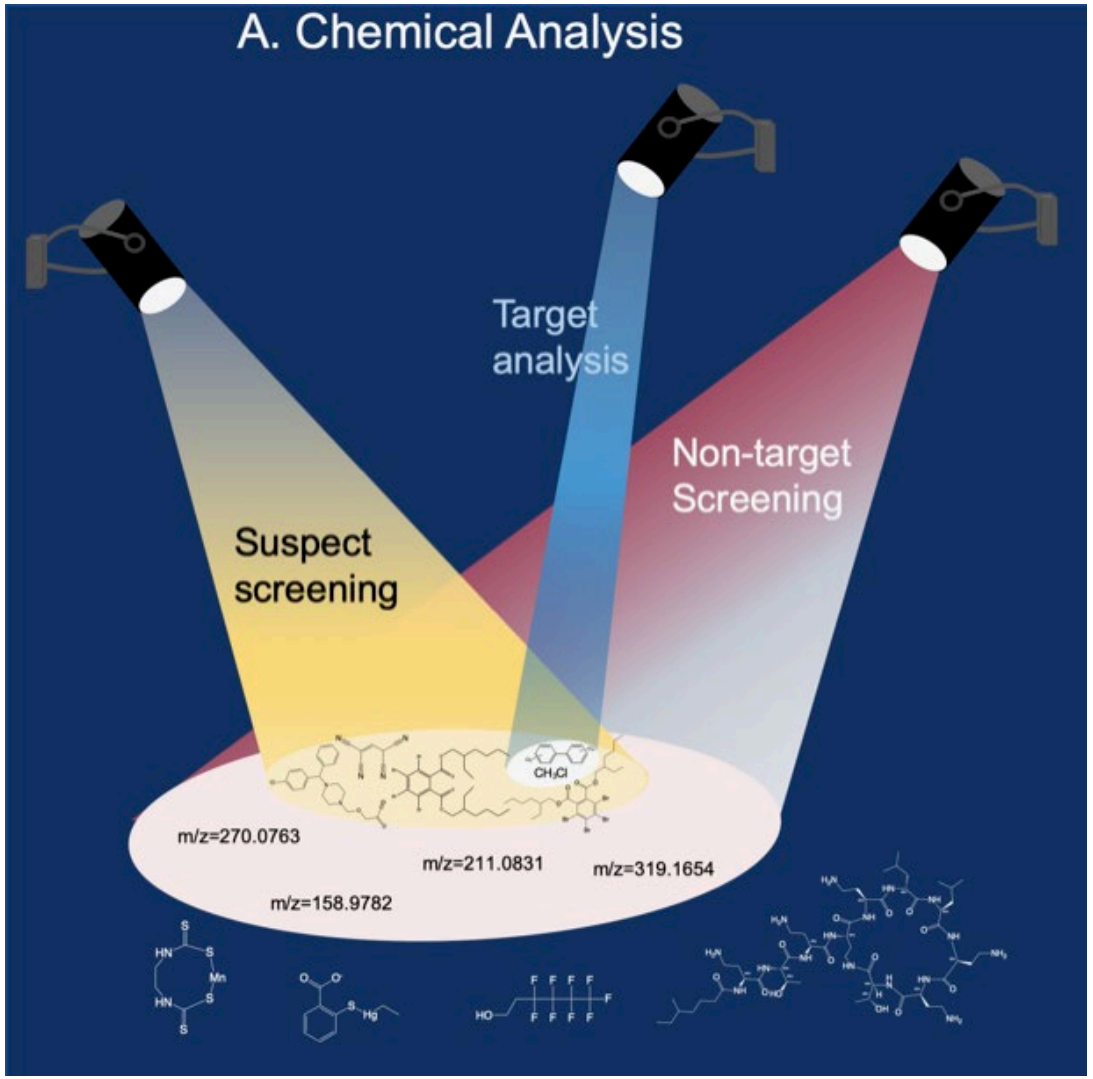
114 million

PubChemLite EXPOSOMICS

456,730

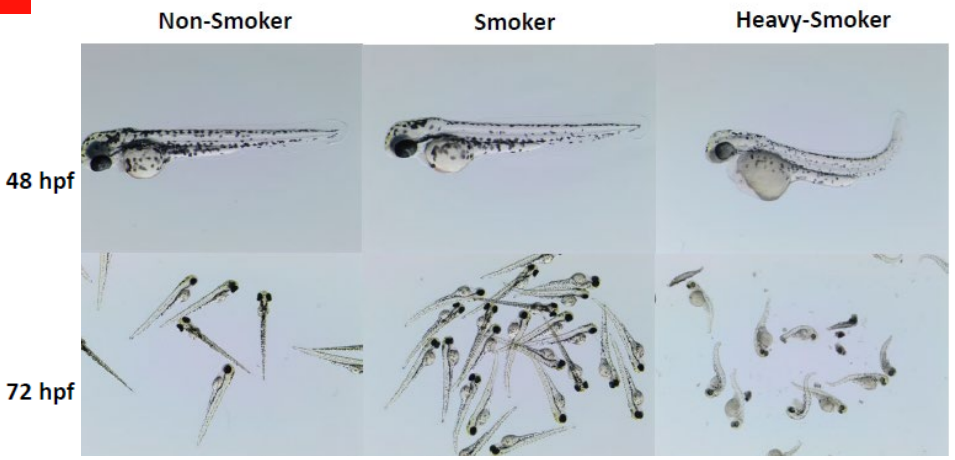
~1400 metabolites

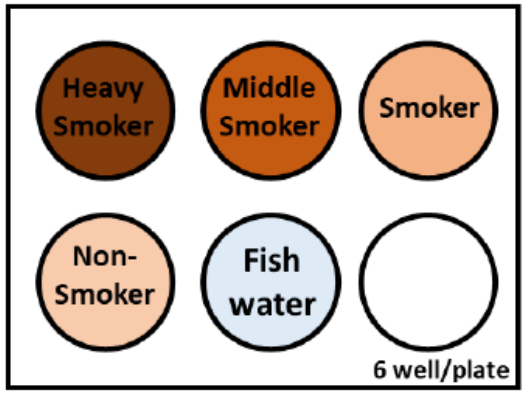
~3600 metabolites



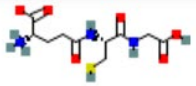
Source: ESO/IDA/Danish 1.5 m http://www.eso.org/public/images/potw1015a/ Mod. from Escher *et al* (2020). Science. DOI: 10.1126/science.aay6636

Organism-specific metabolomics?





PubChem Danio rerio (zebrafish) (Taxonomy)



Glutathione 10.1093/nar/gkaa1024

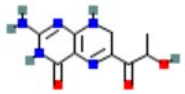
< Previous 1 ... 4 5 6 7 8 ... 85 Next >

Zebrafish Pathway Metabolite MetFrag Local CSV (Beta) | DOI:10.5281/zenodo.3457553

ECI Group, LCSB, University of Luxembourg

3.5 Natural Products

1 item View More Details Download

Structure	Compound	Evidence DOI
	2-amino-6-lactoyl-7,8-dihydropteridin-4(3H)-one	10.1034/J.1600-0749.2002.00045.X

The LOTUS Initiative for Open Natural Products Research: frozen dataset union wikidata (with metadata) | DOI:10.5281/zenodo.5794106

Taxonomy: Kim *et al*. (2022) DOI: 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 (als headers for MetFrag import.

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

Menu

C_elegans_metabolites.Rmd

9.79 KB

Edit

```
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.

Topic	Documentation
Taxonomy	*C. elegans* as PDF or RMarkdown

PubChem ECI GitLab Docs

PubChem Caenorhabditis elegans (Taxonomy)

1-methylnicotinamide 10.1080/21624054.2017.1373939

1 2 3 ... 233 Next >

WormJam Metabolites Local CSV for MetFrag | DOI:10.5281/zenodo.3403364

WormJam: A consensus C. elegans Metabolic Reconstruction and Metabolomics Community and Workshop Series, Worm, 6:2, e1373939, DOI:10.1080/21624054.2017.1373939

ECI Group, LCSB, University of Luxembourg

3.5 Natural Products

Page 3 of 29 items View More Rows & Details Download

Structure	Compound	Evidence DOI
	Pyrrolidin-1-ium-2-carboxylate	10.1186/1471-2164-13-36

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

Exposomics: Disease-specific databases and interpretation

MICROH

LC-HRMS analysis

Data processing and compound identification

Raw files

Data conversion

proteowizard

mzML. files

Non-target screening

MS- DIAL

Spectral libraries (2)

All public MS/MS spectral msp libraries for ESI(+) and ESI (-) (MSPs)

patRoon

Databases (6)

i. PD-MetFrag
ii. Constipation (D003248)
iii. Olfaction disorders (D00857)
iv. Anxiety Disorders (D001008)
v. Sleep Wake Disorders (D012893)
vi. PubChemLite for Exposomics (PCL)

Suspect Screening

patRoon

Suspect lists (4)

i. PD-CTD
ii. LITMINEDNEURO (LITMIN)
iii. Depression disorders (D003863)
iv. Dyssomnias (D020920)

Peak intensity tables

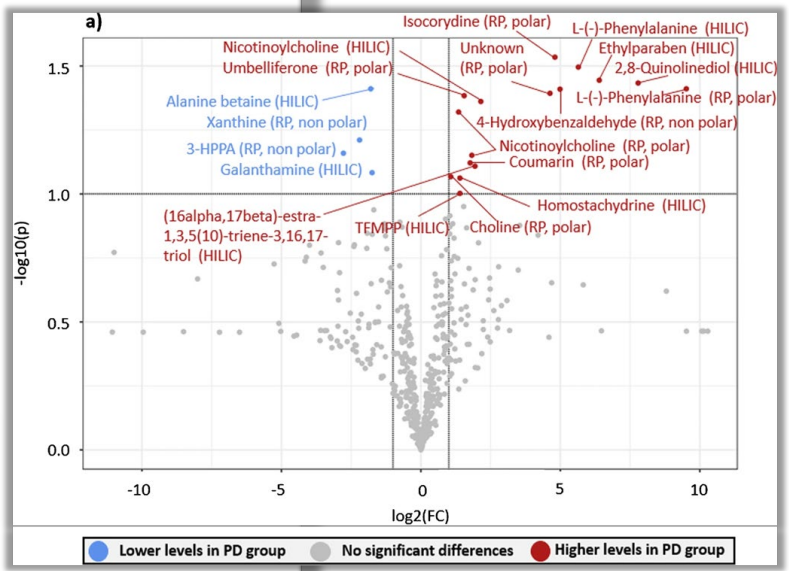
MetaboAnalyst 5.0

Statistical and Pathways analysis

Communication compound confidence levels

MetFrag

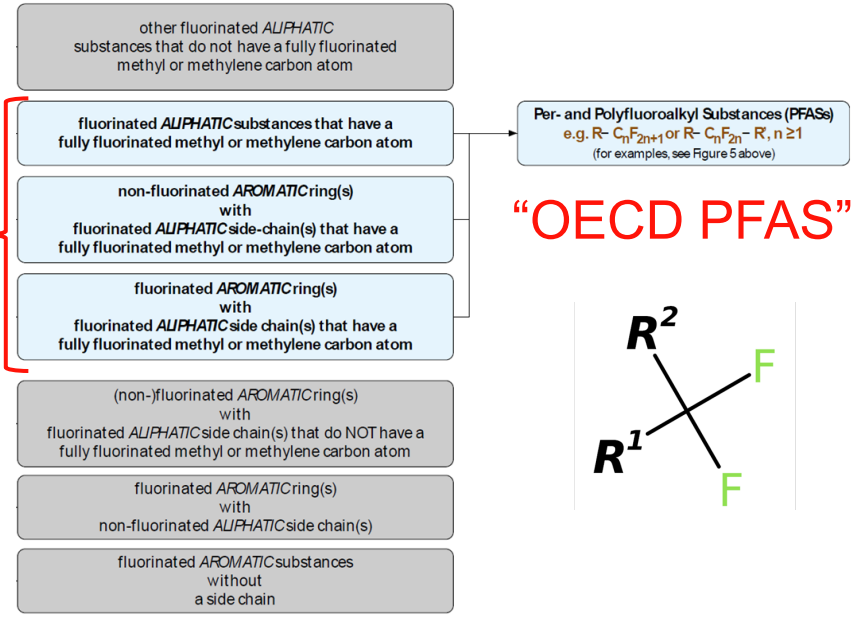
PubChem



Talavera Andújar B. et al. (2022). Anal Bioanal Chem. DOI: 10.1007/s00216-022-04207-z

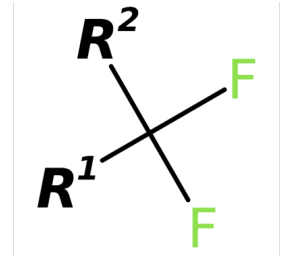
The PubChem PFAS Tree – OECD PFAS

- PFAS and Fluorinated Compounds in PubChem — 19,658,248
  - OECD PFAS definition — 6,191,469
    - Molecule contains isolated CF2 — 596,186
    - Molecule contains isolated CF3 — 5,476,352
    - Molecule contains PFAS parts larger than CF2/CF3 — 222,342
  - Organofluorine compounds — 19,508,893
    - Fluorinated aliphatic substances — 849,139
    - Fluorinated aromatic substances — 18,597,673
    - Other fluorinated substances — 83,882
  - Other diverse fluorinated compounds — 112,987
    - Contains fluorine bond to non-carbon element — 23,353
    - Contains non-organic element — 111,919
  - PFAS and fluorinated compound collections — 1,785,540



https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120

The PubChem PFAS Tree – PFAS Lists



"OECD PFAS"

PubChem PFAS Tree

- PFAS and Fluorinated Compounds in PubChem ? 20,929,770
  - OECD PFAS definition ? 6,370,060
    - Molecule contains isolated CF2 ? 639,815
    - Molecule contains isolated CF3 ? 5,610,808
    - Molecule contains PFAS parts larger than CF2/CF3 ? 226,435
  - Organofluorine compounds ? 19,963,609
  - Other diverse fluorinated compounds ? 122,245
  - PFAS and fluorinated compound collections ? 1,789,330
    - CompTox Chemicals Dashboard PFAS suspect lists ? 16,132
    - NORMAN-SLE PFAS suspect lists ? 6,300
    - OntoChem PFAS lists ? 1,777,053
    - Other fluorinated chemical content in PubChem ? 1,778
    - NIST PFAS suspect list ? 4,948
  - PFAS breakdowns by chemistry ? 7,299,786
  - Regulatory PFAS collections ? 26,965

EPA CompTox Chemicals Dashboard

norman

Google ontochem

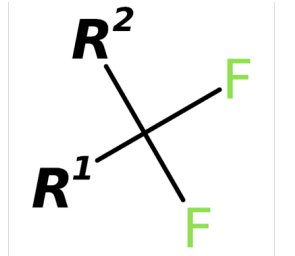
PubChem

NIST PUBLIC DATA REPOSITORY

https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120

uni.lu | LCSB

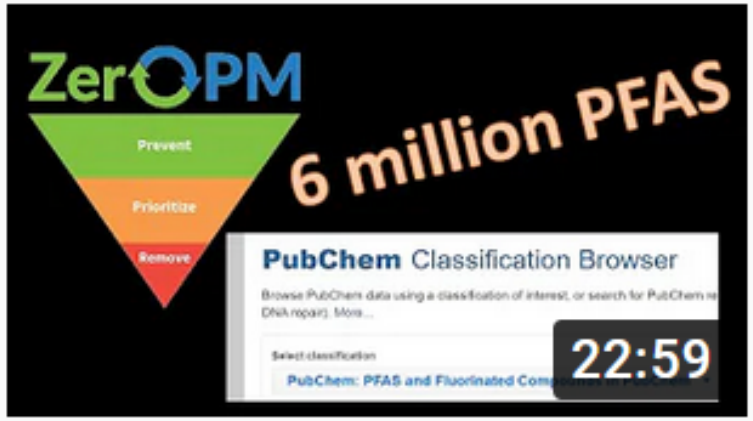
The PubChem PFAS Tree – More info…



"OECD PFAS"

PubChem PFAS Tree

- PFAS and Fluorinated Compounds in PubChem ? ↗ 20,929,770
  - OECD PFAS definition ? ↗ 6,370,060
    - Molecule contains isolated CF2 ? 639,815
    - Molecule contains isolated CF3 ? 5,610,808
    - Molecule contains PFAS parts larger than CF2/CF3 ? 226,435
  - Organofluorine compounds ?
  - Other diverse fluorinated compound
  - PFAS and fluorinated compound co
    - CompTox Chemicals Dashb
    - NORMAN-SLE PFAS suspe
    - OntoChem PFAS lists ?
    - Other fluorinated chemical c
  - NIST PFAS suspect list ? ↗ 4,948
  - PFAS breakdowns by chemistry ? 7,299,786
  - Regulatory PFAS collections ? 26,965



Navigating over 6 million PFAS! A wa…

ZeroPM - H2020

367 views •

2 months ago

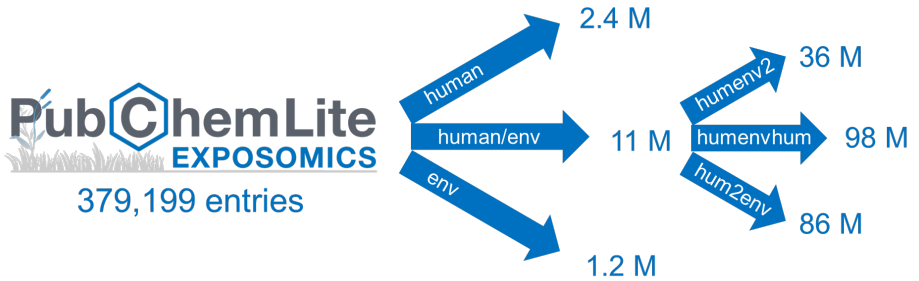
[ZeroPM Webinar: Are there really 6 million PFAS in PubChem?](https://zeropm.eu/news/)

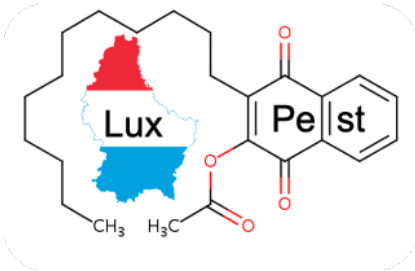
https://zeropm.eu/news/

https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120

“Take home” Messages

- Dynamic, Open Collection: PubChemLite EXPOSOMICS
  - Expert knowledge to fill the gaps
  - Annotation content is extremely powerful
- Transformations and Metabolism
  - More data needed to improve predictions
- Workflows available in patRoon
- NT-HRMS is supporting many environmental studies
  - More exciting exposomics work to come …
- Hopefully some of this helps support Umeå’s work too!





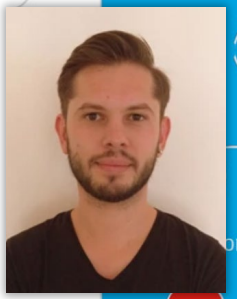
Acknowledgements!

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

Email: emma.schymanski@uni.lu
Twitter: @ESchymanski

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

Today's slides
zenodo



Community Efforts!

… & team

MassBank consortium

U.S. National Library of Medicine
National Center for Biotechnology Information
Luxembourg National Research Fund
uni.lu
LCSB