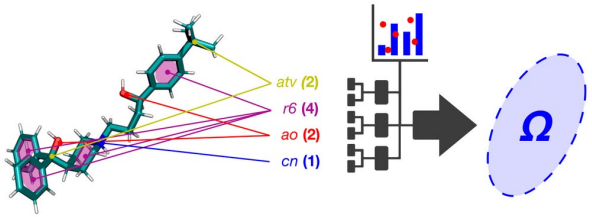
Luxembourg National Research Fund

Metabolomics Association of North America
3rd Summer Symposium 2021
September 10, 2021
"Novel analytical strategies for scoring confidence in compound annotation"

UNIVERSITÉ DU LUXEMBOURG

Identifying *Exposome* Chemicals: *Measured* Data, *Metadata*, *Metabolism* and More …

PubChemLite EXPOSOMICS



CCSbase

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

MANA Summer Symposium, Sept. 10th, 2021 (Virtual)
Novel analytical strategies for scoring confidence in compound annotation

Slides available at
DOI: 10.5281/zenodo.5500083

Intro & Credits!

R3

Luxembourg National Research Fund

Adelene Lai
@adelene.lai
Given access 1 year ago

Anjana Elapavalore
@anjana.elapavalore
Given access 1 year ago

Begoña Talavera Andújar
@begona.talavera
Given access 4 months ago

Corey Griffith
@corey.griffith
Given access 1 year ago

Dagny Aurich
@dagny.aurich
Given access 3 months ago

Emma Schymanski @emma.schymanski
Given access 1 year ago

German Andres Preciat Gonzales
@german.preciat
Given access 1 year ago

Hiba Mohammed-Taha @hiba.mohammetaha
Given access 4 months ago

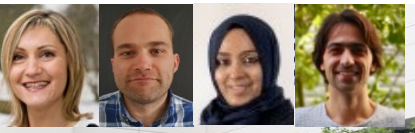
Jessy Krier @jessy.krier
Given access 1 year ago

Lorenzo Favilli
@lorenzo.favilli
Given access 1 year ago

Mira Narayanan
@mira.narayanan
Given access 2 months ago

Parviel Chirsir
@parviel.chirsir
Given access 1 month ago

Todor Kondic
@todor.kondic
Given access 1 year ago

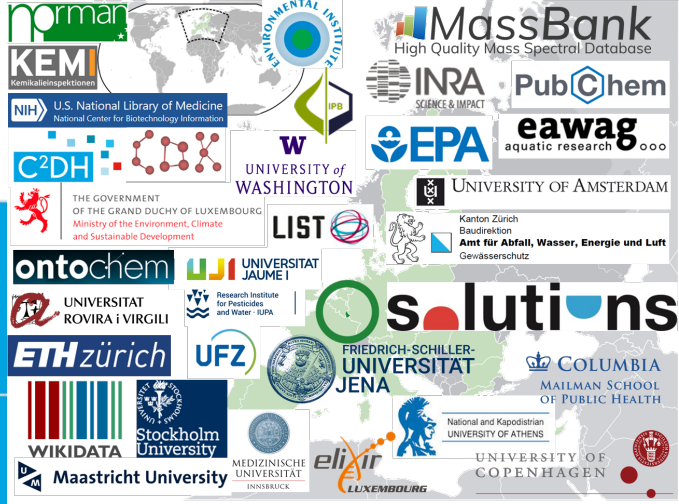


ECI @ LCSB

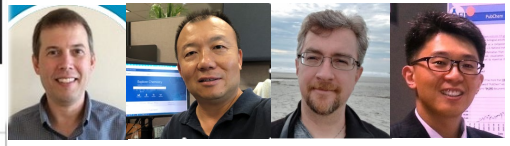
Genes
Medicine
Cohorts
Innovation
Microbes
Environment
Healthcare
Modeling

VANDERBILT
UNIVERSITAT JAUME I
Universiteit Antwerpen

SLE Lists S50, S61, S79



PubChem



… and team

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

IPB

MetFrag

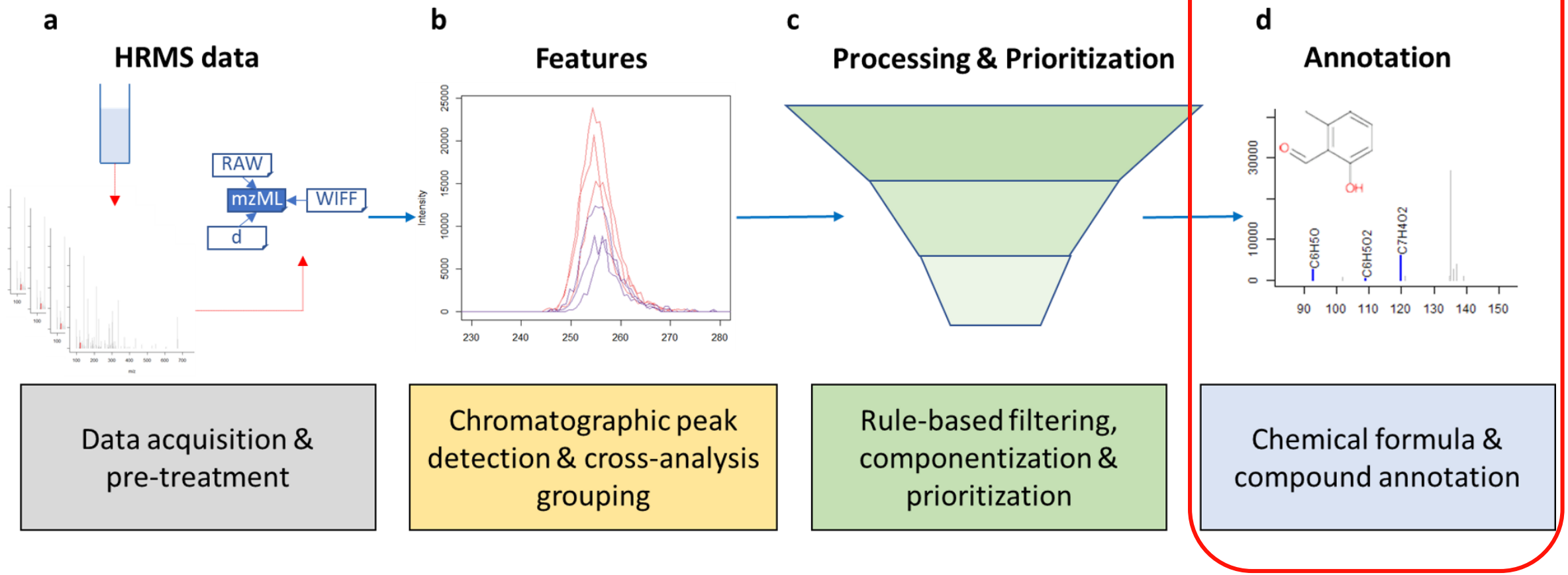
CCSbase
Xu Lab

UNIVERSITY of WASHINGTON

wi.lu

LCSB

Background: Identification with HR-MS



Helmus *et al* (2021). patRoon. *J. Cheminformatics*. DOI: 10.1186/s13321-020-00477-w

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

CAS
A DIVISION OF THE AMERICAN CHEMICAL SOCIETY

180 million

PubChem

111 million

ChemSpider
Search and share chemistry

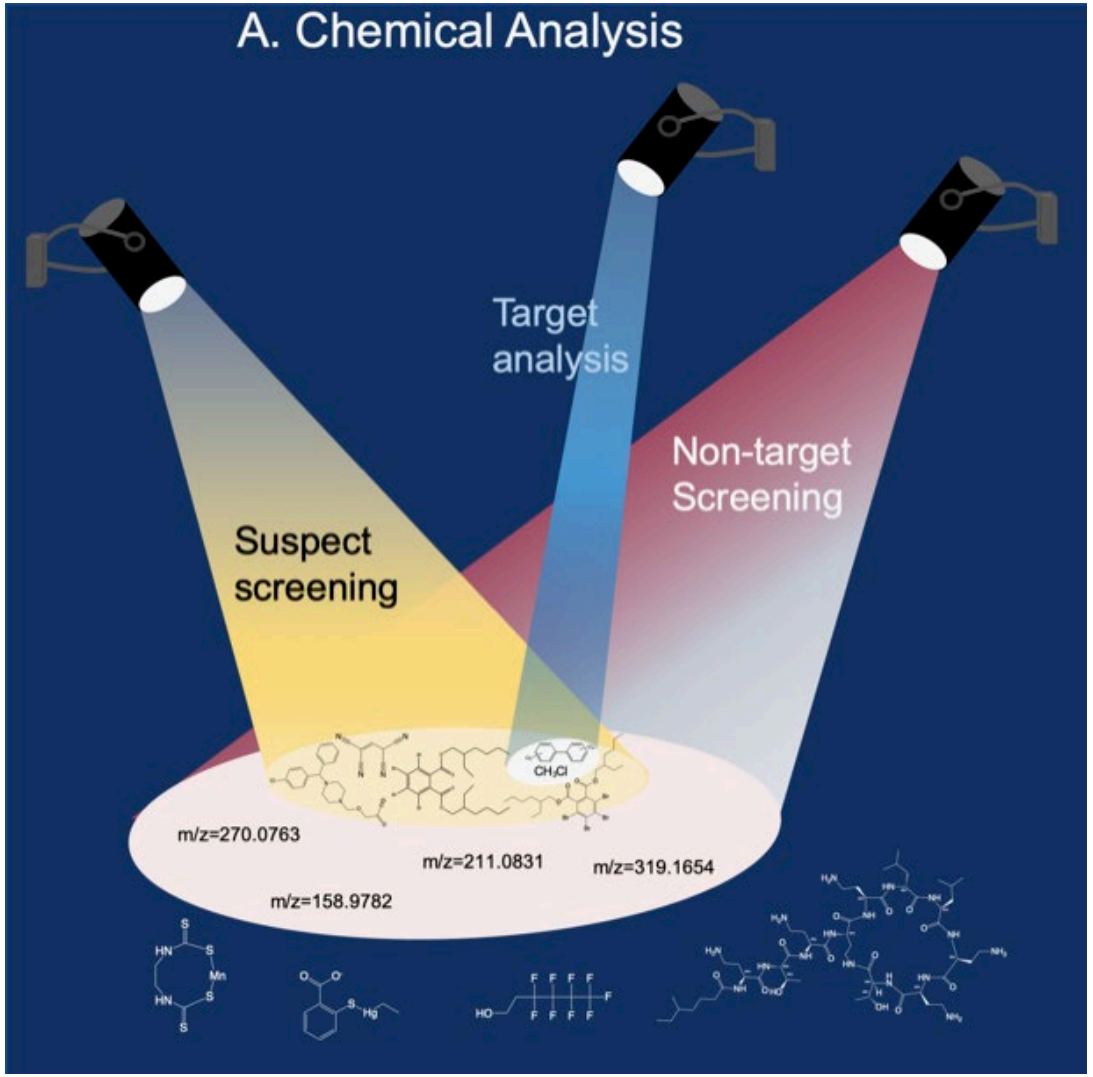
110 million

EPA Chemicals

883,000

norman suspects

113,000

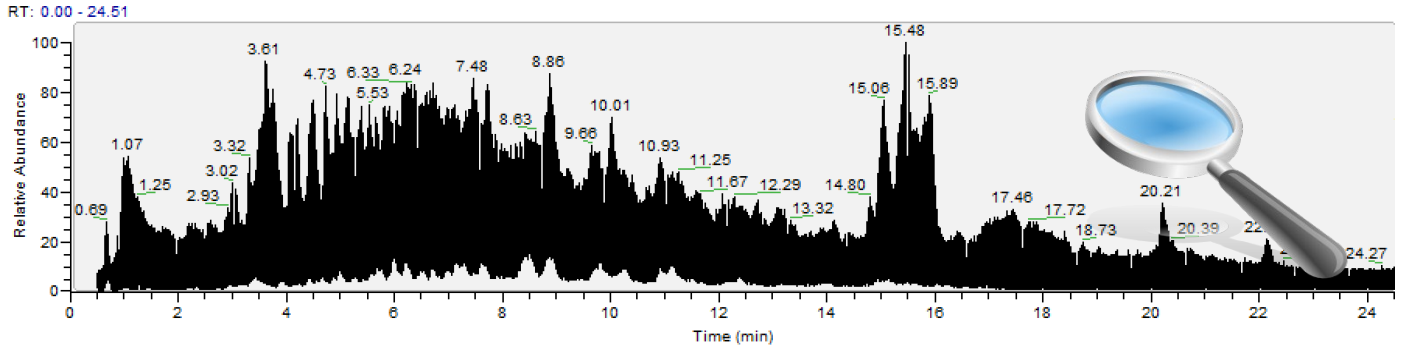


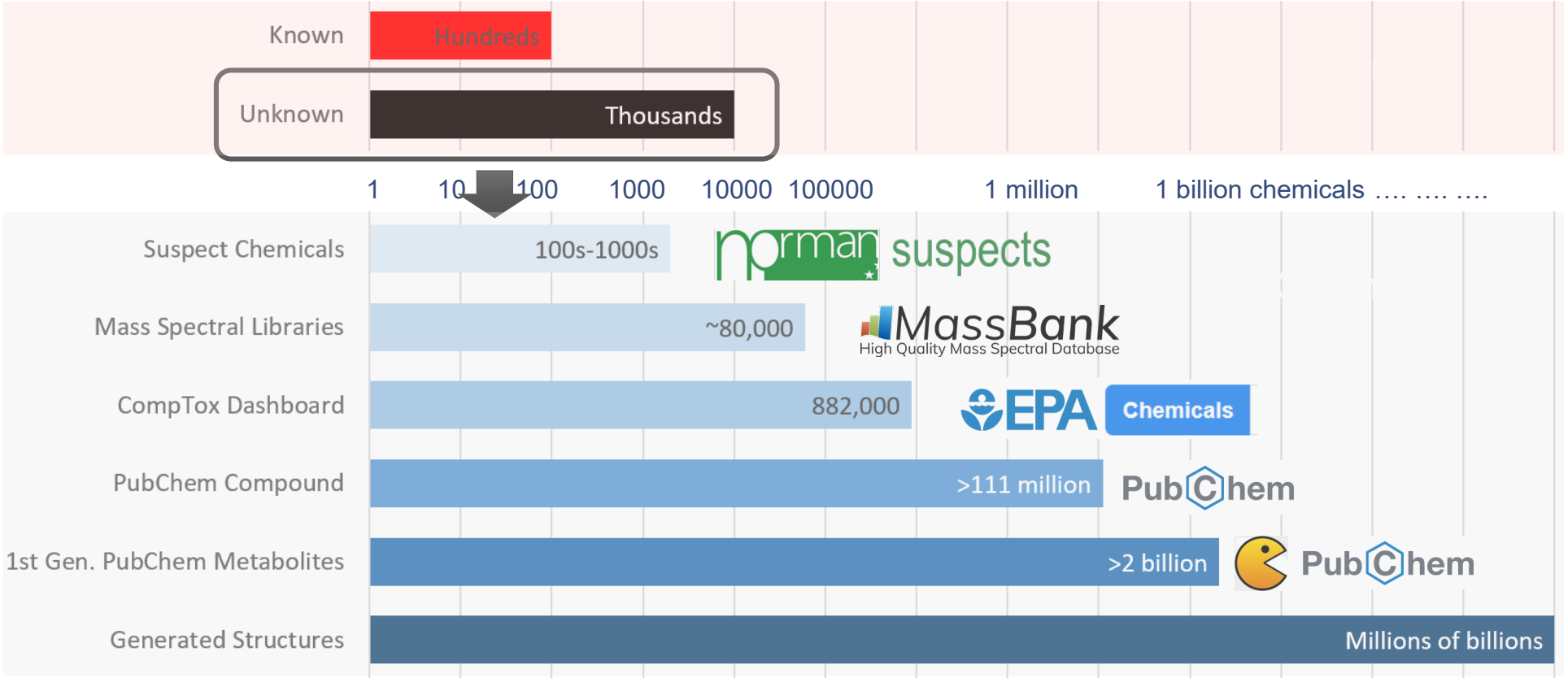
Source: ESO/IDA/Danish 1.5 m http://www.eso.org/public/images/potw1015a/ Mod. from Escher *et al* (2020). Science. DOI: 10.1126/science.aay6636

Background: High Resolution Mass Spec & Cheminformatics

High resolution mass spectrometry

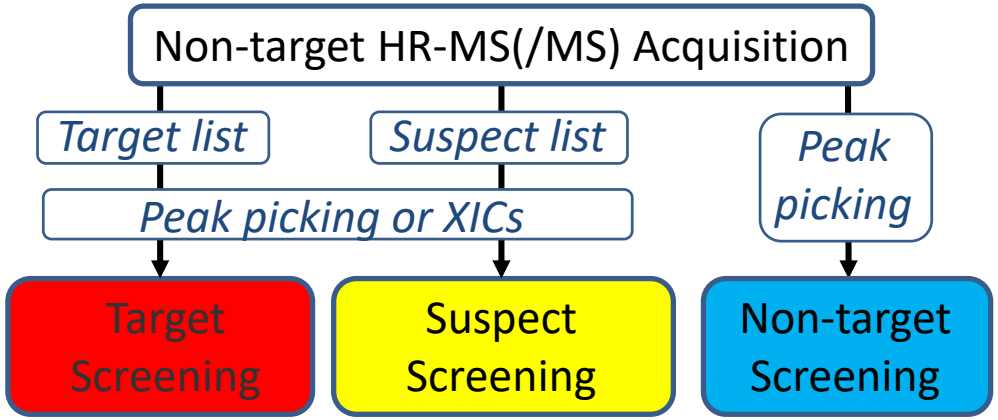
AND connecting chemical knowledge





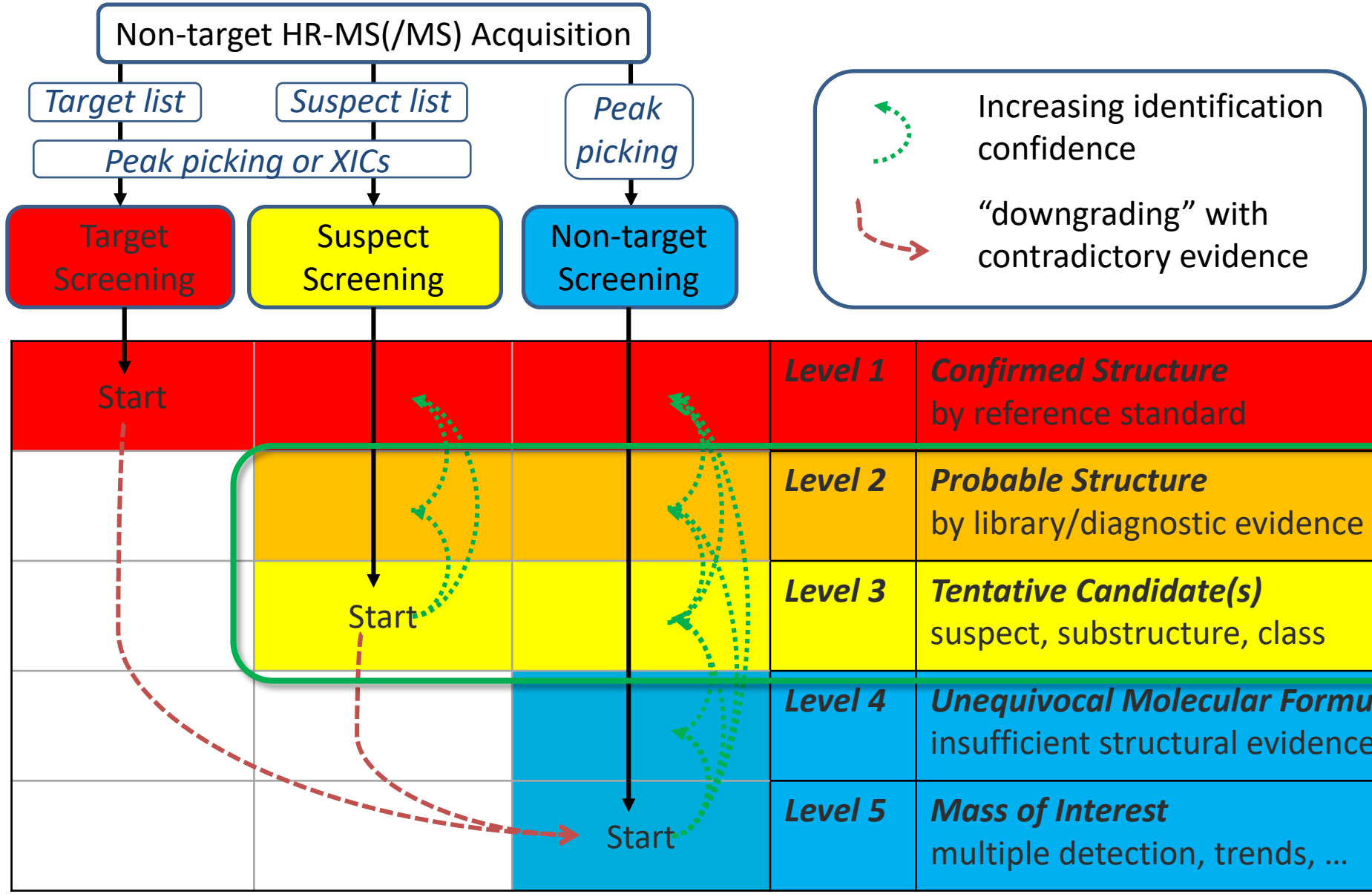
Schymanski *et al.* (2014) DOI: 10.1021/es4044374; Vermeulen *et al.* (2020) DOI: 10.1126/science.aay3164

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



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)



MassBank
High Quality Mass Spectral Database

MetFrag

eawag aquatic research

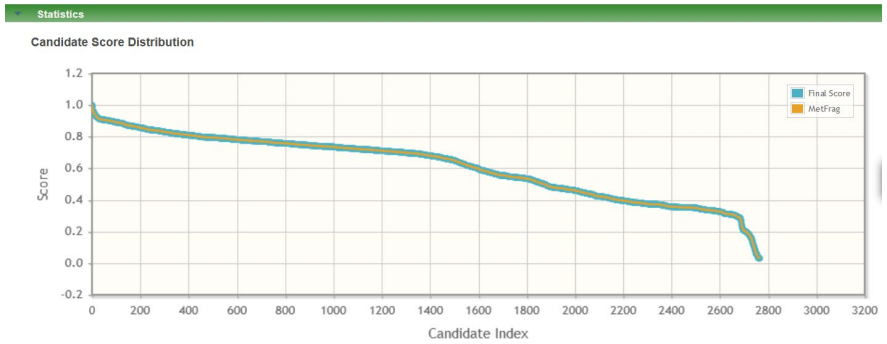
Schymanski *et al*, 2014, ES&T. DOI: 10.1021/es5002105 & Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

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

mz [M−H]⁻
213.9637
± 5 ppm

5 ppm
0.001 Da

Ranked Candidates



MS/MS

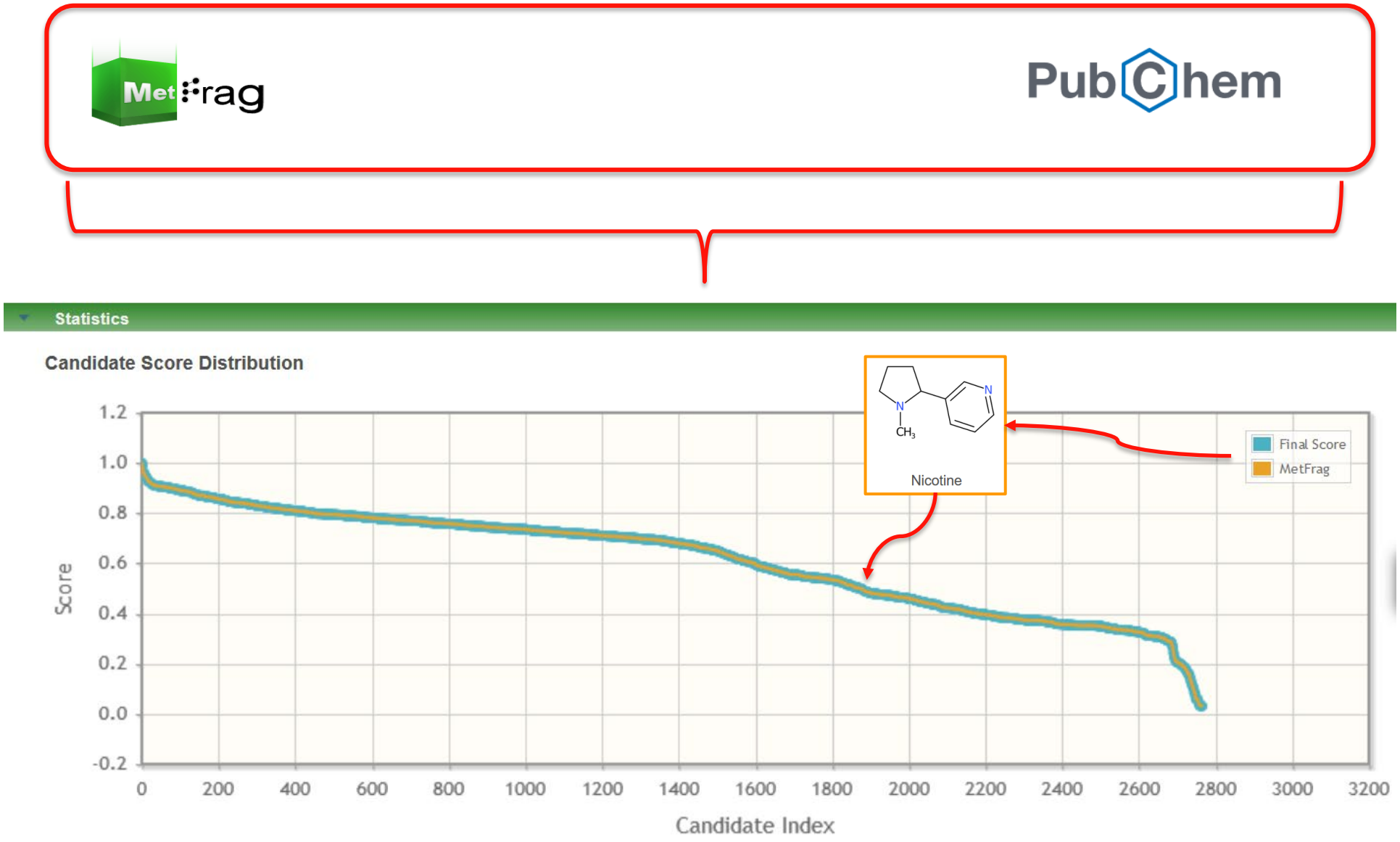
134.0054	339689
150.0001	77271
213.9607	632466

Ruttkies, Schymanski, Wolf, Hollender, Neumann (2016) *J. Cheminf.*, DOI: 10.1186/s13321-016-0115-9

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

Schymanski, Kondic, Neumann, Thiessen, Zhang & Bolton (2021) *J. Cheminf.*, DOI: 10.1186/s13321-021-00489-0
Ruttkies, Schymanski, Wolf, Hollender, Neumann (2016) *J. Cheminf.*, DOI: 10.1186/s13321-016-0115-9

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



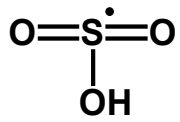
MetFrag + PubChem + Formula Search + https://massbank.eu/MassBank/RecordDisplay.jsp?id=EQ300804&dsn=Eawag

Status Quo in 2016: MetFrag Relaunched …

mz [M-H]⁻
213.9637
± 5 ppm

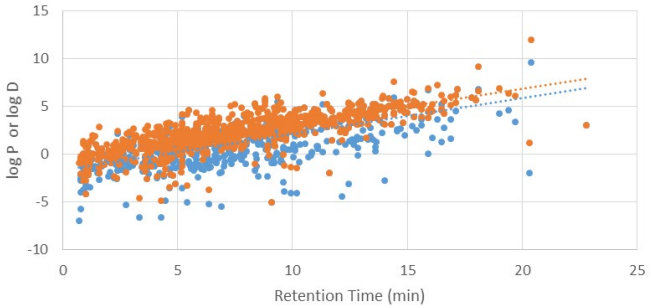
Elements: C,N,S

5 ppm
0.001 Da



RT: 4.54 min

355 InChI/RTs



EPA Chemicals
or
PubChem

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

Literature Patents

Suspect Lists

TOFF IDENT

norman suspects

EPA Chemicals

MoNA
MassBank of North America
MassBank
High Quality Mass Spectral Database

MS/MS

134.0054	339689
150.0001	77271
213.9607	632466



IPB

eawag aquatic research

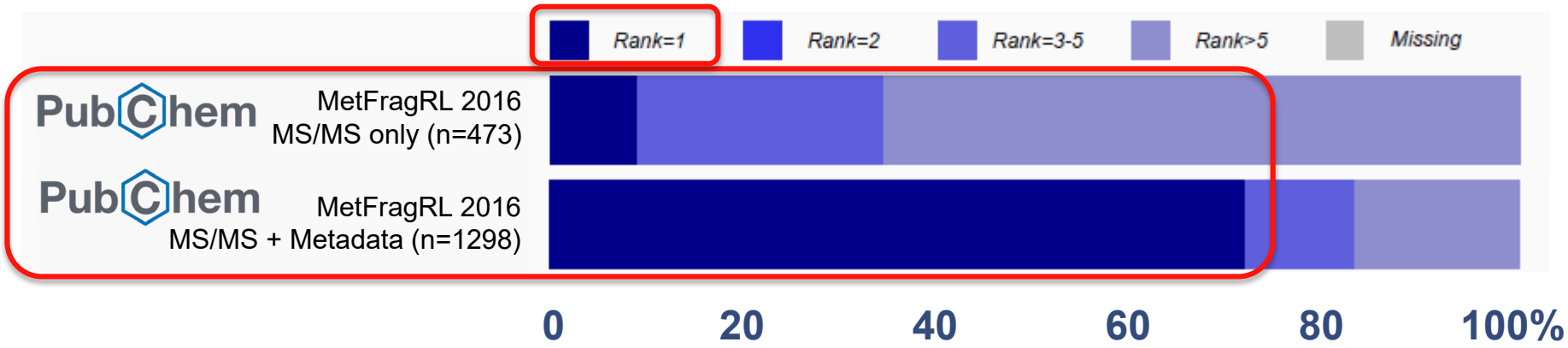
solutions

uni.lu | LCSB

Ruttkies, Schymanski, Wolf, Hollender, Neumann (2016) *J. Cheminf.*, DOI: 10.1186/s13321-016-0115-9

MetFragRL + PubChem + MS/MS + Metadata

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



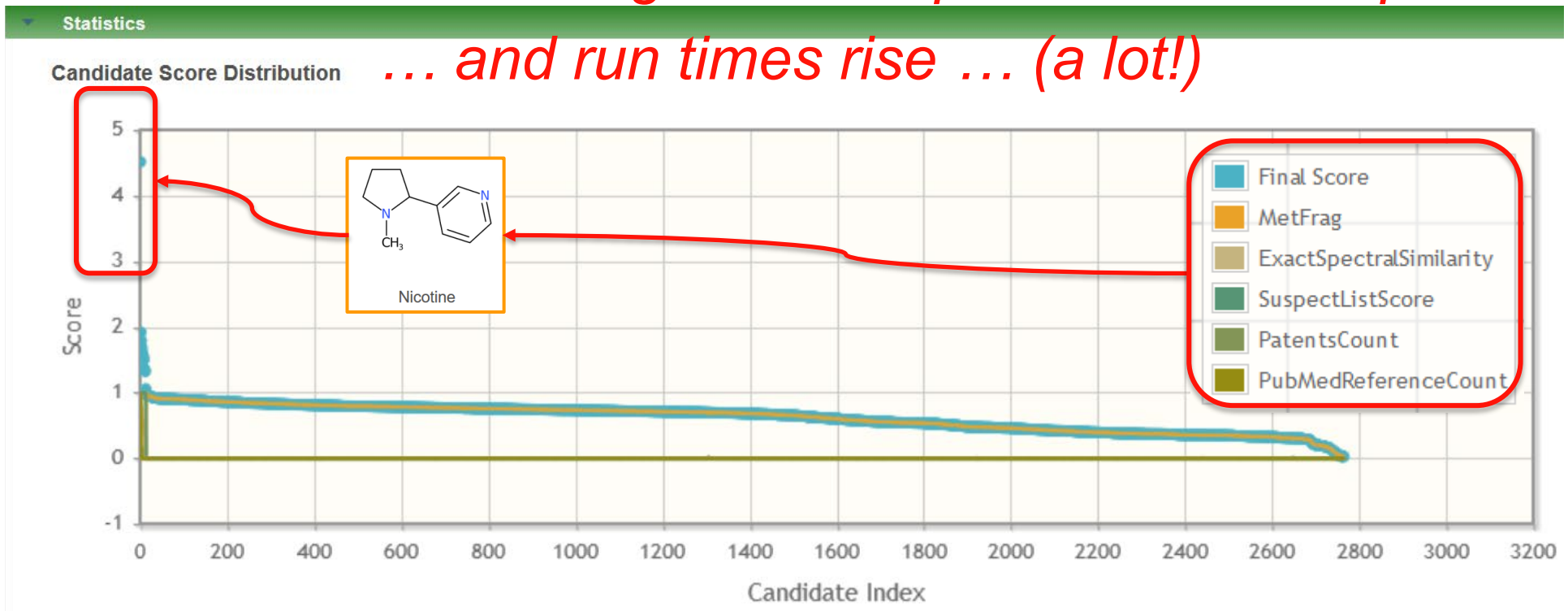
Schymanski, Kondic, Neumann, Thiessen, Zhang & Bolton (2021) *J. Cheminf.*, DOI: 10.1186/s13321-021-00489-0

Ruttkies, Schymanski, Wolf, Hollender, Neumann (2016) *J. Cheminf.*, DOI: 10.1186/s13321-016-0115-9

MetFragRL + PubChem + MS/MS + Metadata

BUT …databases grow … ID performance drops

… and run times rise … (a lot!)



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

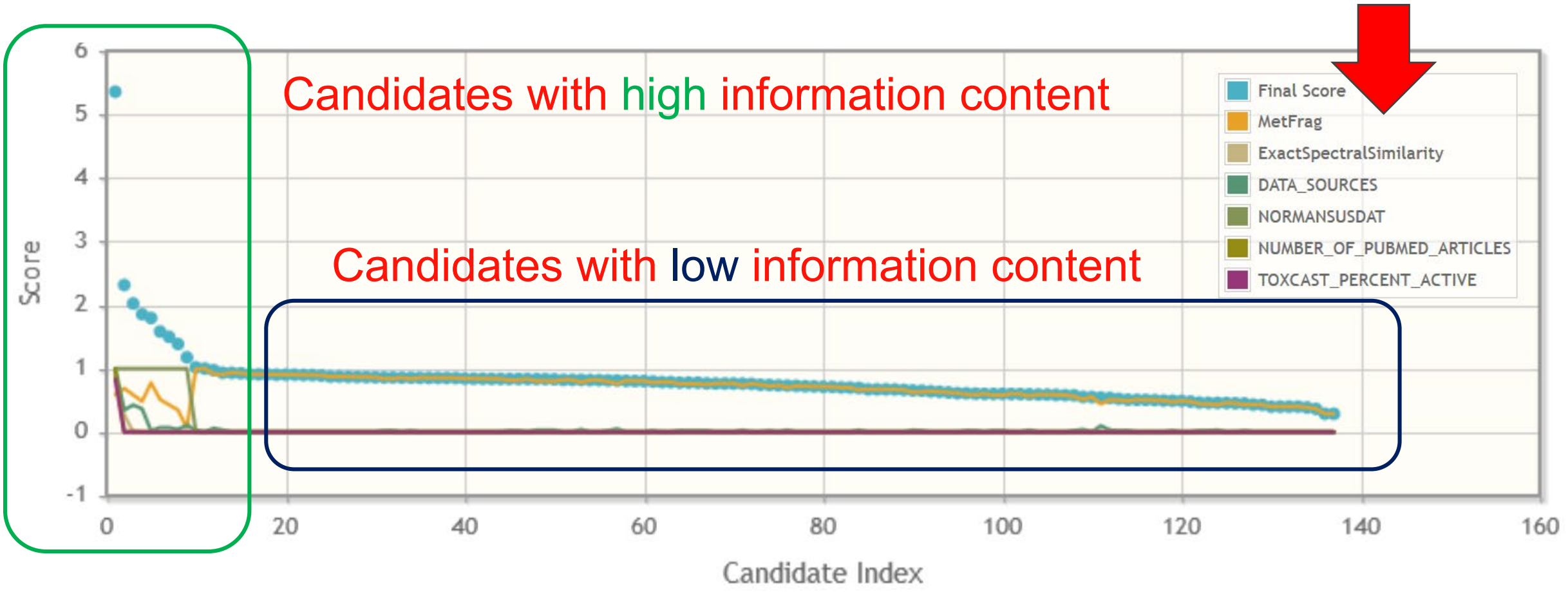
Problem: Exposomics “Chemical Space” is too big!

CAS A DIVISION OF THE AMERICAN CHEMICAL SOCIETY — 180 million

PubChem — 111 million

ChemSpider Search and share chemistry — 110 million

EPA Chemicals — 883,000



Can we break down PubChem into useful bits?

- PubChem Compound TOC 49,493,641
  - Agrochemical Information 3,045
  - Associated Disorders and Diseases 20,847
  - Biologic Description 2,056,521
  - Biological Test Results 3,622,920
  - Biomolecular Interactions and Pathways 125,253
  - Chemical and Physical Properties 263,015
  - Classification 1,454,824
  - Drug and Medication Information 17,922
  - Food Additives and Ingredients 8,414
  - Identification 4,968
  - Information Sources 20,271,277
  - Literature 1,833,941
  - Names and Identifiers 1,275,170
  - Patents 36,351,418
  - Pharmacology and Biochemistry 110,628
  - Related Records 9,224,590
  - Safety and Hazards 149,319
  - Spectral Information 480,730
  - Structures 9,117,635
  - Toxicity 114,012
  - Use and Manufacturing 115,321
  - Chemical Safety 147,023

- Agrochemical Information 3,045
  - Agrochemical Category 1,903
  - Agrochemical Transformations 1,383
  - EU Pesticides Data 1,213
  - USDA Pesticide Data Program 579
- Use and Manufacturing 115,321
  - Uses 27,314
  - Consumption Patterns 1,168
  - Formulations/Preparations 4,742
  - General Manufacturing Information 52,353
  - Impurities 870
  - Methods of Manufacturing 6,192
  - Overview 1,179
  - Sampling Procedures 579
  - U.S. Exports 1,038
  - U.S. Imports 1,547
  - U.S. Production 5,500
  - Use Classification 63,558

PubChem Furathiocarb (Compound) CONTENTS

7 Agrochemical Information

7.1 Agrochemical Category

Insecticides

- EU Pesticides Database

7.2 Agrochemical Transformations

Furathiocarb has known environmental transformation products that include carbofuran.

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

- NORMAN Suspect List Exchange

7.3 EU Pesticides Data

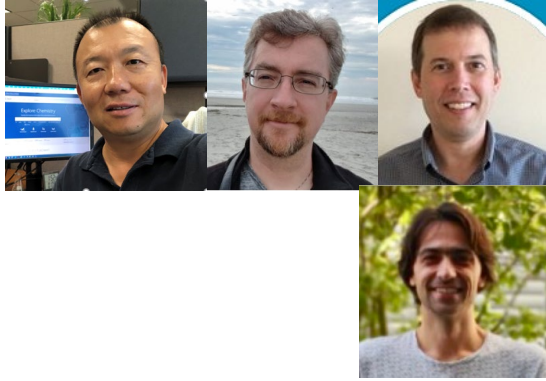
Active Substance	furathiocarb
Status	Not Approved [Reg. (EC) No 1107/2009]
Categories	Insecticides

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

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

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

Introducing …



- PubChem Compound TOC ? 49,493,641
  - Agrochemical Information ? 3,045
  - Associated Disorders and Diseases ? 20,847
  - Biologic Description ? 2,056,521
  - Biological Test Results ? 3,622,920
  - Biomolecular Interactions and Pathways ? 125,253
  - Chemical and Physical Properties ? 263,015
  - Classification ? 1,454,824
  - Drug and Medication Information ? 17,922
  - Food Additives and Ingredients ? 8,414
  - Identification ? 4,968
  - Information Sources ? 20,271,277
  - Literature ? 1,833,941
  - Names and Identifiers ? 1,275,170
  - Patents ? 36,351,418
  - Pharmacology and Biochemistry ? 110,628
  - Related Records ? 9,224,590
  - Safety and Hazards ? 149,319
  - Spectral Information ? 480,730
  - Structures ? 9,117,635
  - Toxicity ? 114,012
  - Use and Manufacturing ? 115,321

Chemical Safety ? 147,023

PubChemLite EXPOSOMICS

~370,000 entries “small”

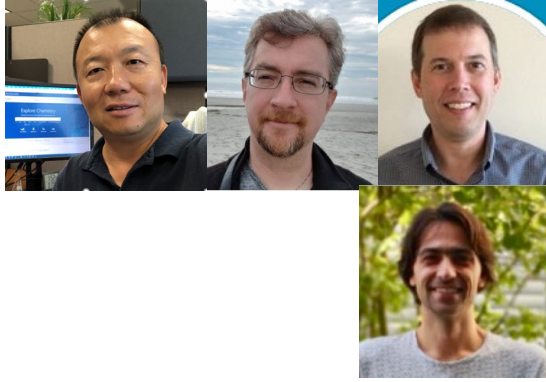
Schymanski *et al*. (2021) DOI: 10.1186/s13321-016-0115-9

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

uni.lu | LCSB

Introducing …

PubChemLite EXPOSOMICS



Collapsed by InChIKey First Block (skeleton) *and* by presence of annotation content

~370,000 entries “small”

PubChem INCSWYKICIYAHB

Compounds (6) | Substances (2)

Searching chemical names and synonyms including IUPAC names and InChIKeys across the compound collection. Note that annotations text searched. Read More...

6 results | Filters | SORT BY Compound C

2,3-dihydroxy-2,3-dihydrobenzoic Acid; 5,6-dihydroxycyclohe
Acid; 100459-00-5; 2,3-dihydro-2,3-dihydroxybenzoic Acid; ACM

Compound CID (sort by): 3
MF: $C_7H_8O_4$ MW: 156.14g/mol
InChIKey: INCSWYKICIYAHB-UHFFFAOYSA-N
IUPAC Name: 5,6-dihydroxycyclohexa-1,3-diene-1-carboxylic acid
Create Date: 2004-09-16

Summary | Similar Structures Search | Related Records

(2S,3S)-2,3-dihydroxy-2,3-dihydrobenzoic Acid; 176487-06-2; (2S
2,3-dihydroxybenzoate; (5S,6S)-5,6-dihydroxycyclohexa-1,3-dien
(2S,3S)-2,3-Dihydro-2,3-dihydroxybenzoate; ...

Compound CID (sort by): 9964159
MF: $C_7H_8O_4$ MW: 156.14g/mol
InChIKey: INCSWYKICIYAHB-WDSKDSINSA-N
IUPAC Name: (5S,6S)-5,6-dihydroxycyclohexa-1,3-diene-1-carboxylic acid
Create Date: 2006-10-25

MetFrag

Database Settings

Database: PubChemLite_31Oct2020_

Neutral Mass: 229.10948 Search ppm: 5

Formula: C9H16ClN5

Identifiers:

Retrieve Candidates | 4 Candidates

zenodo | Search | Upload | Communities | emma.schymanski@uni.lu

October 31, 2020 | Dataset | Open Access

PubChemLite for Exposomics

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

PubChemLite is a subset of PubChem (https://pubchem.ncbi.nlm.nih.gov/) selected from major
Contents page at the PubChem Classification Browser
.nih.gov/classification/#hid=72). With this release, there is now just one
h is the former tier1 plus two new categories (Associated Disorders &
):

" is 371,663 compounds (31 Oct 2020) com
es:
y, DrugMedicInfo, FoodRelated, PharmacoIn
Info,
se, Identification.

collapsed by InChIKey first block, reporting t
most
l CIDs. Entries that will be ignored by MetFra
's (e.g. transition metals) have been remove
Med ID
files on the PubChem FTP site. The "AnnoTypeCount" term counts how
represented, the subsequent column (named per category) counts the
gories available in the next sub-category of the TOC entry.

R3

Edit | New version

Communities
LCSB Environmental Cheminformatics Group | Remove

1,372 views | 1,477 downloads

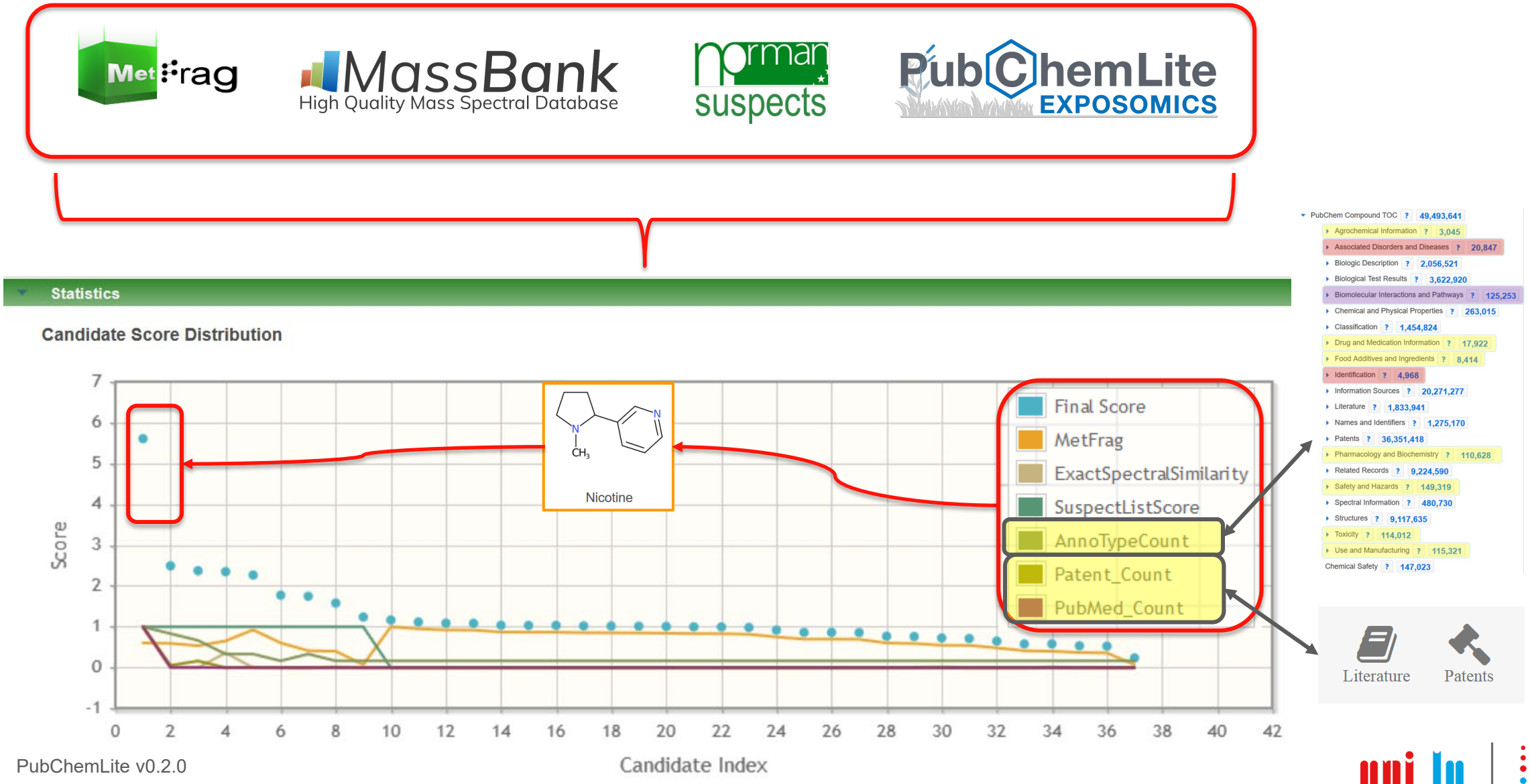
See more details...

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

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

uni.lu | LCSB

MetFragRL + PubChemLite: tailor-made database + metadata

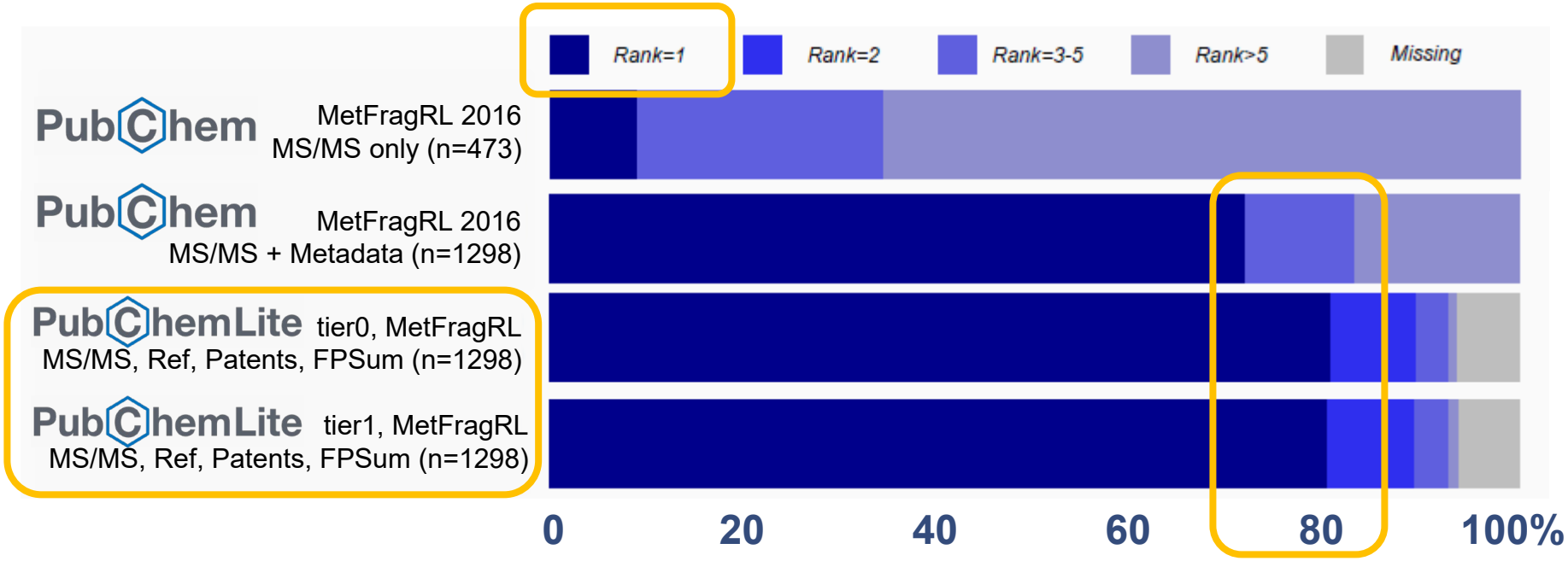


PubChemLite v0.2.0

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

How does PubChemLite perform?

- ~110 M => ~300 K … how does this influence performance?

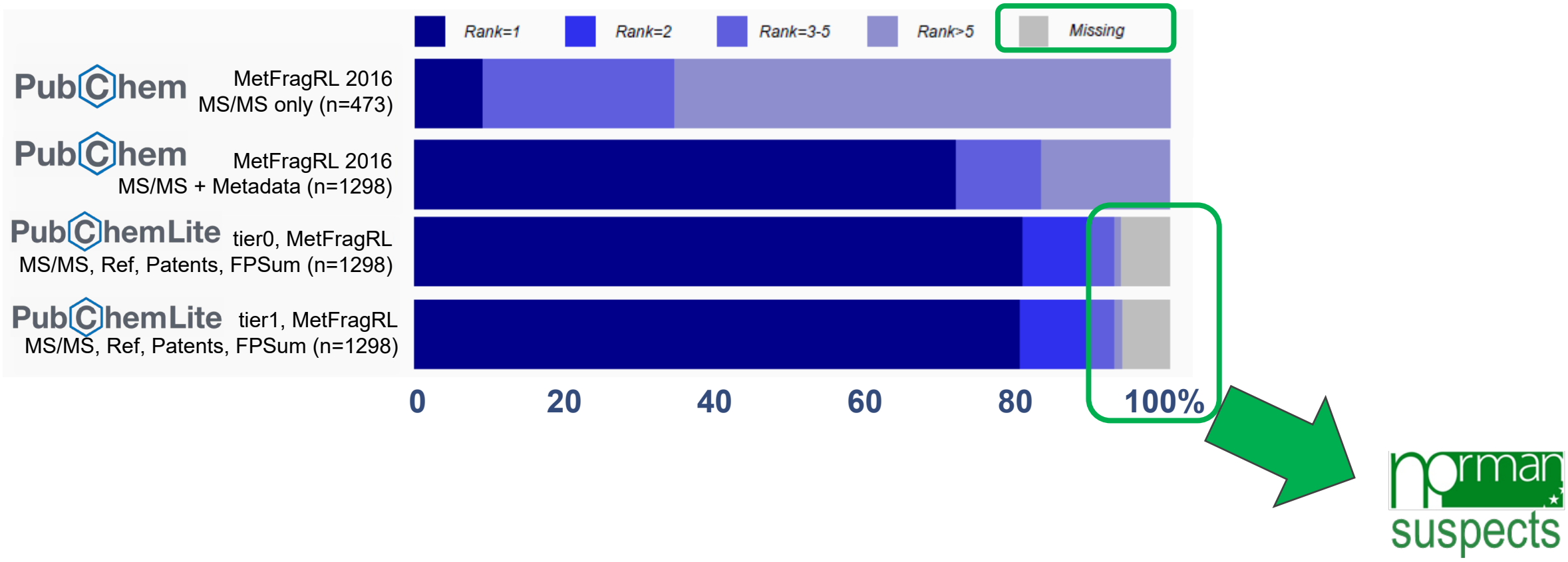


Schymanski, Kondic, Neumann, Thiessen, Zhang & Bolton (2021) *J. Cheminf.*, DOI: 10.1186/s13321-021-00489-0

Ruttkies, Schymanski, Wolf, Hollender, Neumann (2016) *J. Cheminf.*, DOI: 10.1186/s13321-016-0115-9

How does PubChemLite perform?

- ~110 M => ~300 K … how does this influence performance?



Schymanski, Kondic, Neumann, Thiessen, Zhang & Bolton (2021) *J. Cheminf.*, DOI: 10.1186/s13321-021-00489-0

Ruttkies, Schymanski, Wolf, Hollender, Neumann (2016) *J. Cheminf.*, DOI: 10.1186/s13321-016-0115-9

Expert Knowledge: NORMAN Database System

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

NORMAN Database System

NORMAN organises the development and maintenance of various web-based databases for the collection & evaluation of data / information on emerging substances in the environment

SEARCH All Databases

Searching for individual substance or group(s) of substances in all databases

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

SARS-CoV-2 in sewage

A database with the latest information on SARS-CoV-2 in sewage across Europe and internationally; including a common protocol for sample collection, storage, extraction, analysis and data sharing to support the development of an international comparable data set.

Substance Database

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

Chemical Occurrence Data

A database of geo-referenced monitoring data on emerging substances

Ecotoxicology

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

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

MassBank Europe

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

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

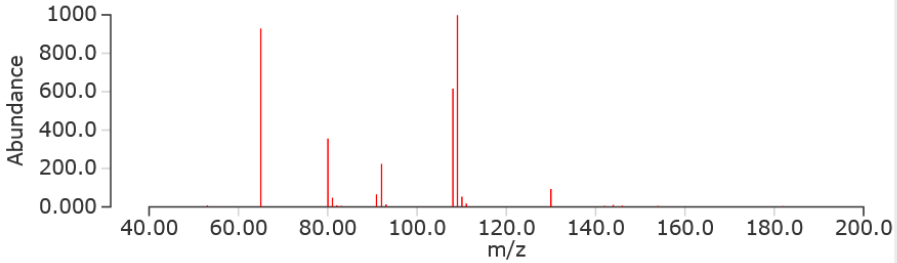
MassBank
High Quality Mass Spectral Database

>> Search Spectra

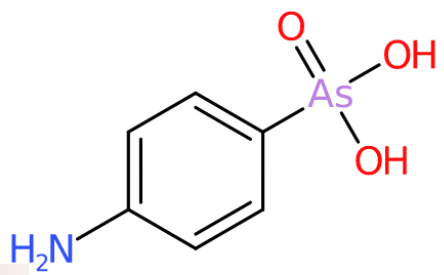
MassBank Record: LU040605

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

Mass Spectrum



Chemical Structure



Search for:

Basic Search | Peak List | Peaks | Peak Differences

Compound Information

Compound name

Exact Mass

AND

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

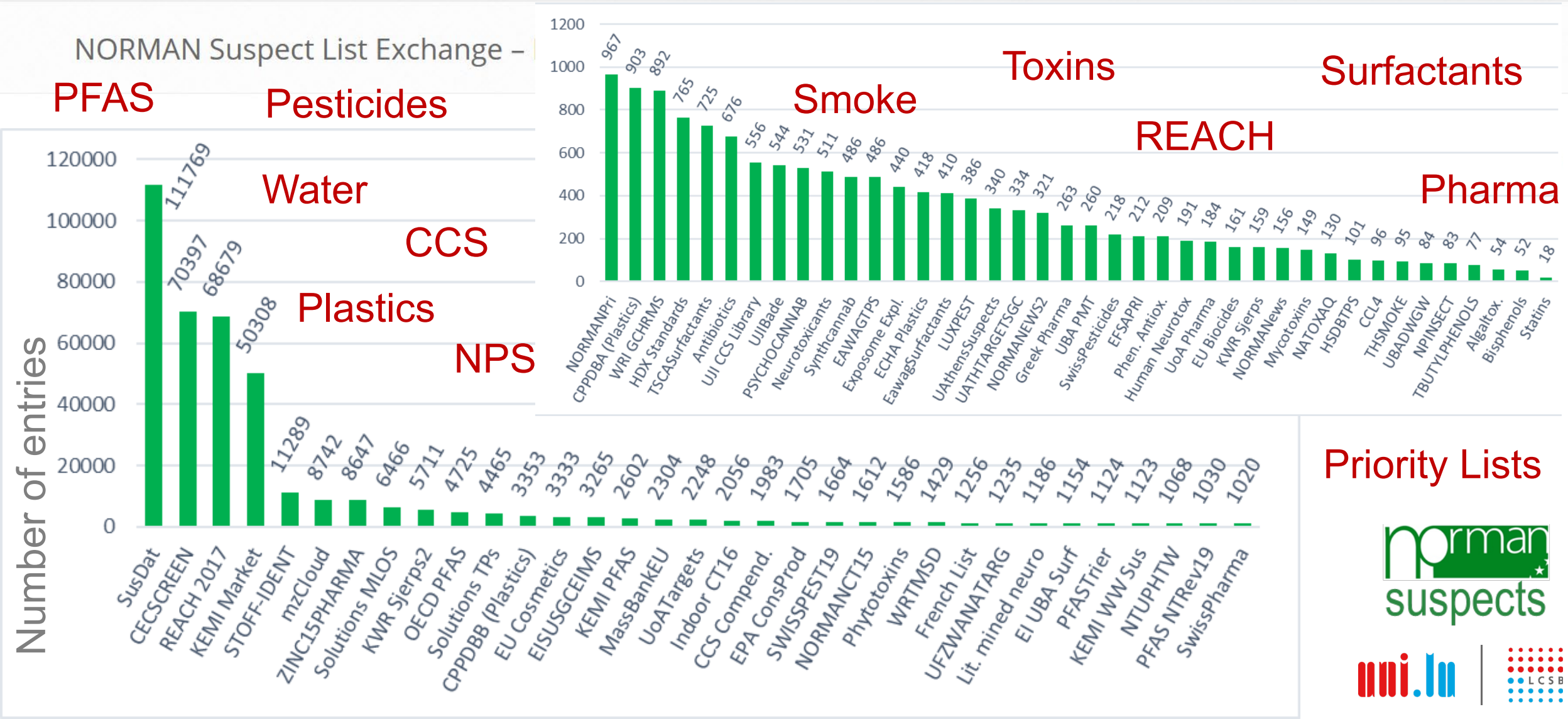
AND

Search

mi.lu LCSB

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

https://www.norman-network.com/nds/SLE/ https://zenodo.org/communities/norman-sle



Priority Lists

norman suspects

uni.lu LCSB

Filling Gaps: Integrating NORMAN-SLE

https://www.norman-network.com/nds/SLE/ => https://pubchem.ncbi.nlm.nih.gov/source/23819

PubChem NORMAN Suspect List Exchange

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	114,573 Live Substances 16,423 Annotations 1 Classification
Last Updated	2021/04/15

Filling Gaps: Integrating NORMAN-SLE

NORMAN-SLE Classification: https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101

PubChem Classification Browser

Help

Browse PubChem data using a classification of interest, or search for PubChem records annotated with the desired classification/term (e.g., MeSH: phenylpropionates, or Gene Ontology: DNA repair). More...

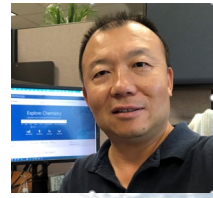
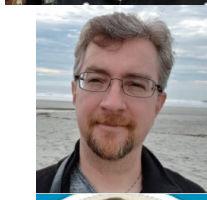
Select classification NORMAN Suspect List Exchange

Search selected classification by Keyword — Enter desired search term — Search

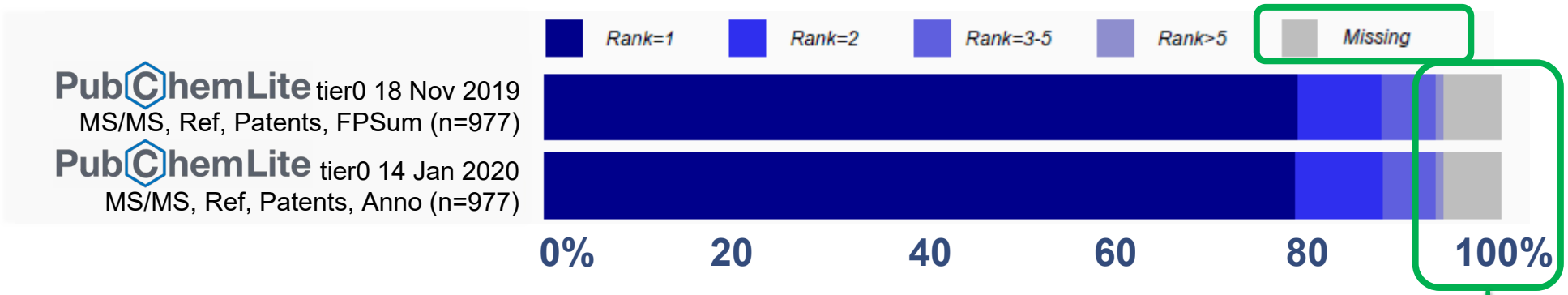
Classification description (from NORMAN Suspect List Exchange)

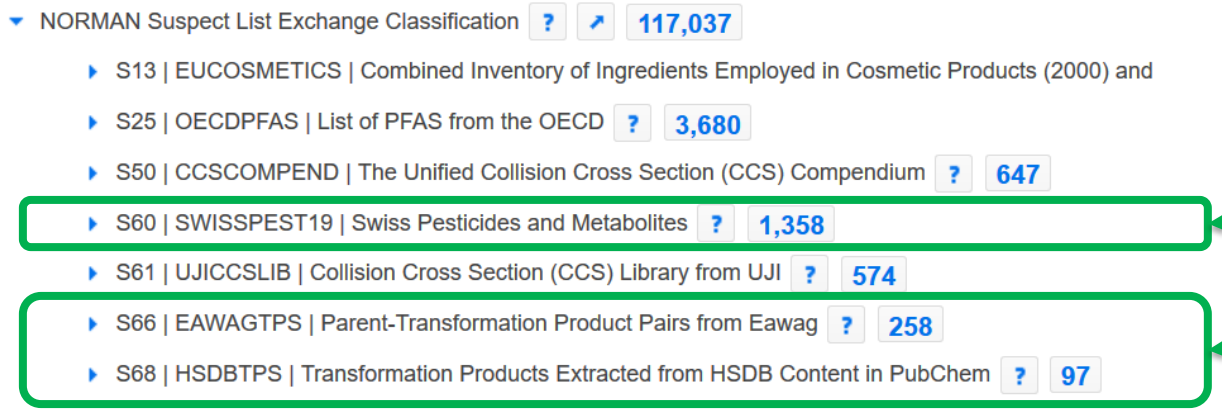
The NORMAN Suspect List Exchange (NORMAN-SLE) is a central access point for NORMAN members (and others) to find suspect lists relevant for their environmental monitoring questions. More...

- NORMAN Suspect List Exchange Classification ? 113,080
  - S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) ? 3,859
  - S25 | OECDPFAS | List of PFAS from the OECD ? 3,677
  - S36 | UBAPMT | Potential Persistent, Mobile and Toxic (PMT) substances ? 254
  - S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ? 885
  - S60 | SWISSPEST19 | Swiss Pesticides and Metabolites from Kiefer et al 2019 ? 1,343
  - S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI ? 574
  - S66 | EAWAGTPS | Parent-Transformation Product Pairs from Eawag ? 258
  - S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem ? 102
  - S69 | LUXPEST | Pesticide Screening List for Luxembourg ? 386
  - S72 | NTUPHTW | Pharmaceutically Active Substances from National Taiwan University ? 1,068
  - S75 | CyanoMetDB | Comprehensive database of secondary metabolites from cyanobacteria ? 2,088


Assessing the Missing Entries in PubChemLite





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

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

Transformation Products: Filling the Data Gaps!

PubChem NORMAN Suspect List Exchange

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

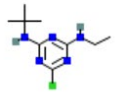
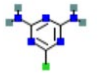
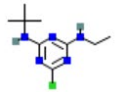
- Pharmacology and Biochemistry ? 112,039
  - Human Metabolite Information ? 64,199
  - Metabolism/Metabolites ? 8,204
  - Transformations ? 5,857

PubChem Terbuthylazine (Compound)

8.5 Transformations

Page 3 of 25 items View More Rows & Details Download

SORT BY Please Choose One

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

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

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

uni.lu LCSB

Transformation Products: Filling the Data Gaps!

PubChem Terbuthylazine (Compound)

7 Agrochemical Information

7.1 Agrochemical Category

Pesticides -> Herbicides -> Triazine herbicides -> Chlorotriazine herbicides

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

NORMAN Suspect List Exchange

7.2 Agrochemical Transformations

Terbutylazine has known environmental transformation products that include Terbutylazine-2-hydroxy, Terbutylazine-desethyl, and Terbutylazine-desethyl-2-hydroxy.

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

NORMAN Suspect List Exchange

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

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

NORMAN Suspect List Exchange

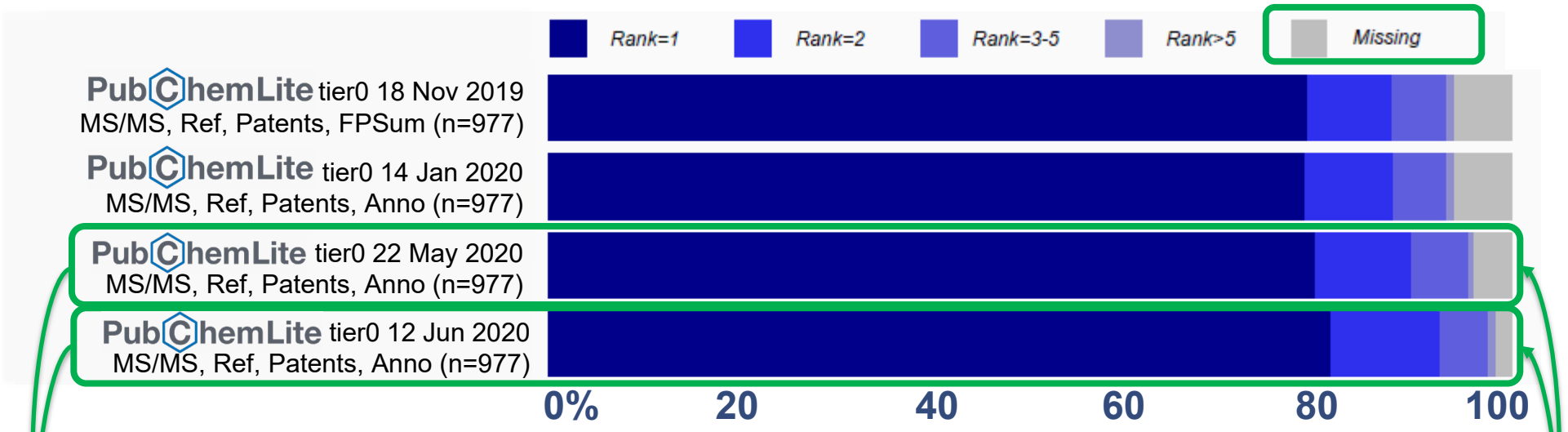
- Agrochemical Information ? 3,045
  - Agrochemical Category ? 1,903
  - Agrochemical Transformations ? 1,383
  - EU Pesticides Data ? 1,213
  - USDA Pesticide Data Program ? 579

PubChemLite EXPOSOMICS

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

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

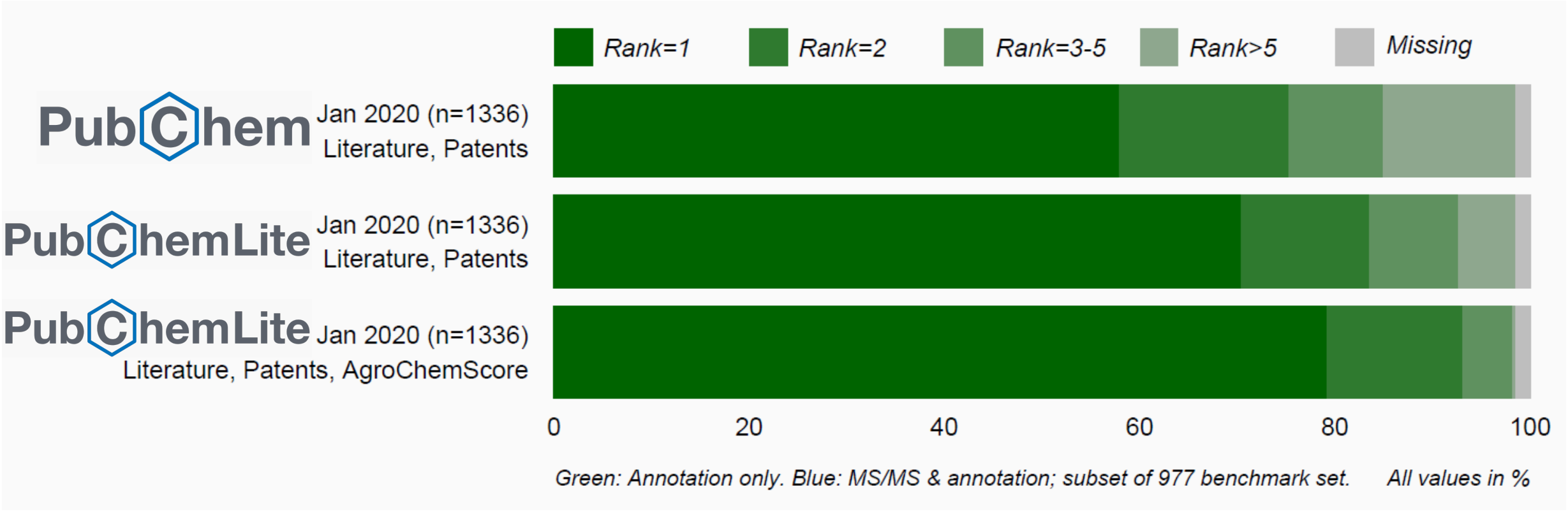
Assessing the Missing Entries in PubChemLite



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

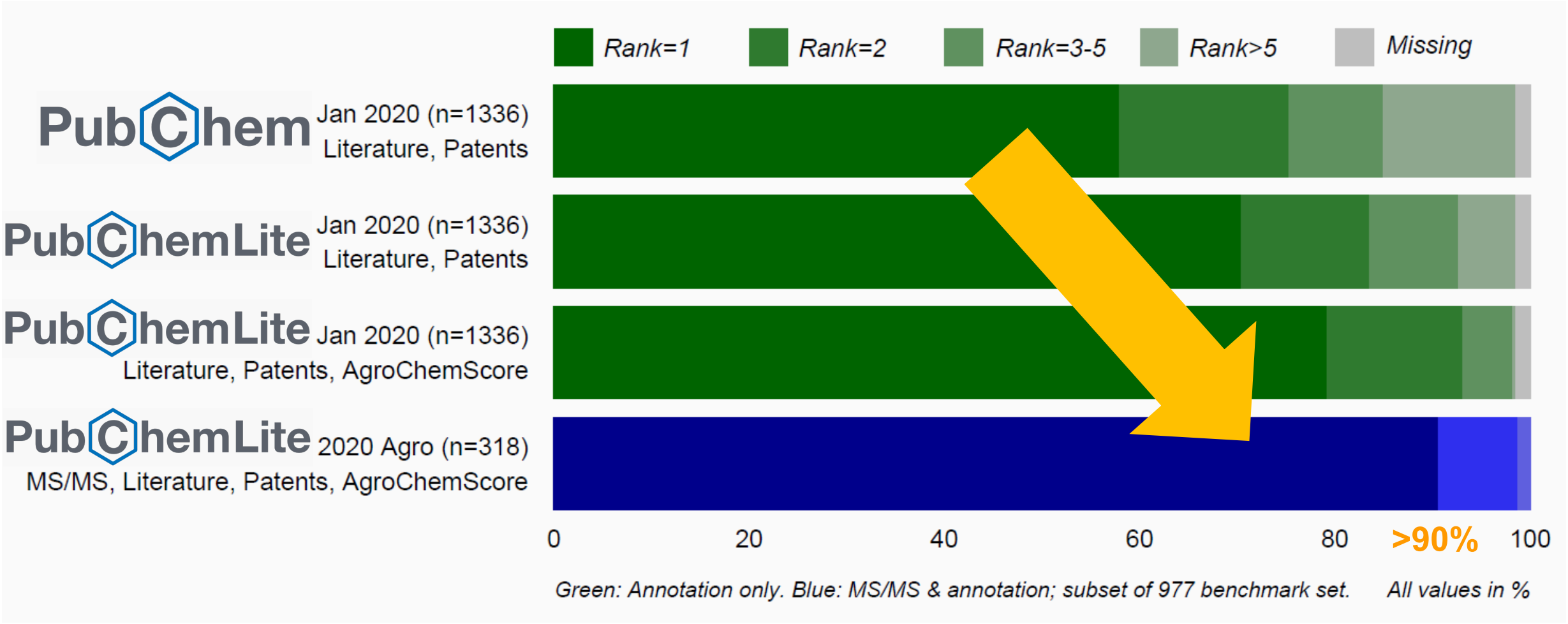
U.S. National Library of Medicine
National Center for Biotechnology Information

Influence of the Annotation Content in PubChemLite



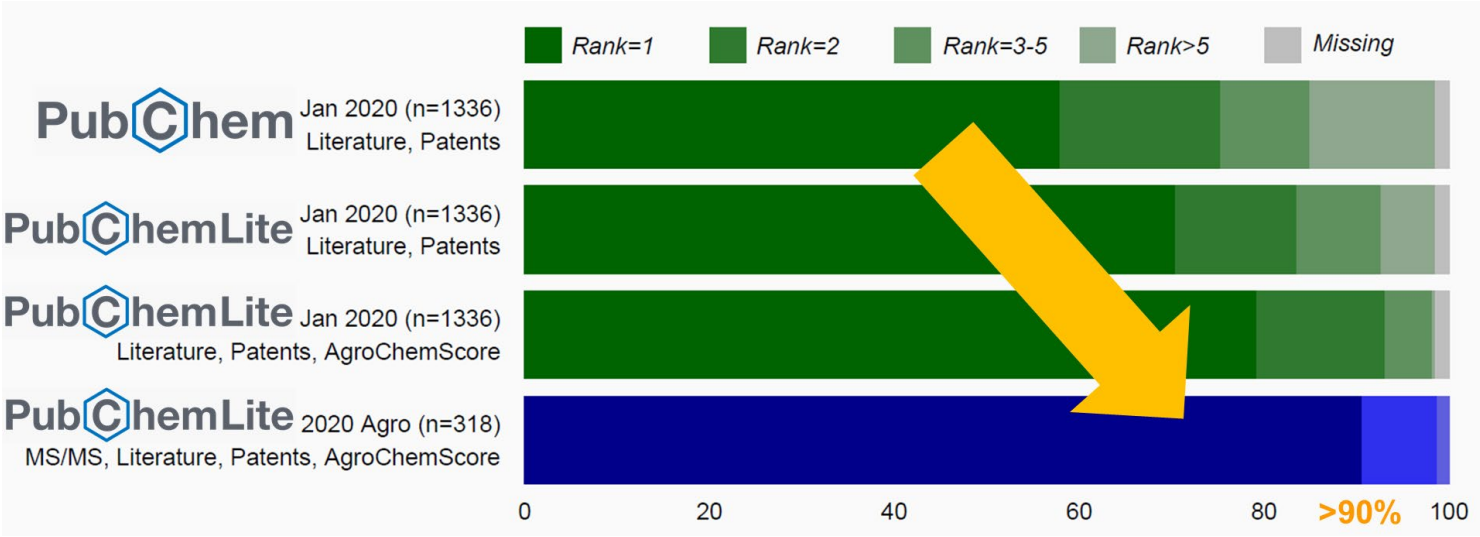
Schymanski *et al*. (2021) DOI: 10.1186/s13321-016-0115-9

Influence of the Annotation Content in PubChemLite



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

Influence of the Annotation Content in PubChemLite



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

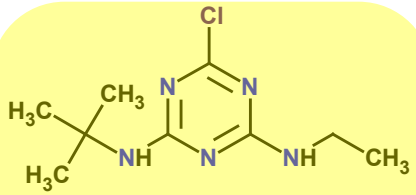
>97% are Top 2 … 100 % Top 3 …

Can we do even better?

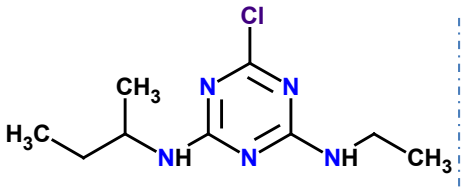
Schymanski *et al*. (2021) DOI: 10.1186/s13321-016-0115-9

MetFrag Example: Isobars, Metadata & Measured Data

$C_9H_{16}ClN_5$ *m/z 229.1094 Da*

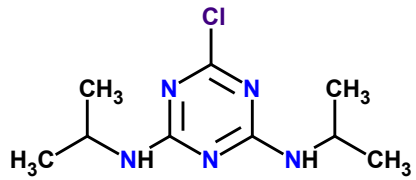


Terbutylazine
Detects: 12; # Refs: 220



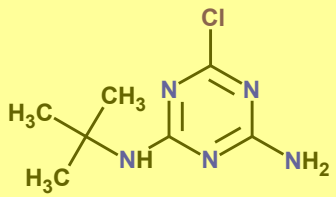
Sebutylazine
Detects: 3; # Refs: 51

(no related compound at this mass)

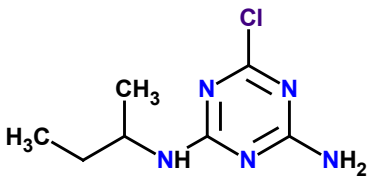


Propazine
Detects: 3; # Refs: 201

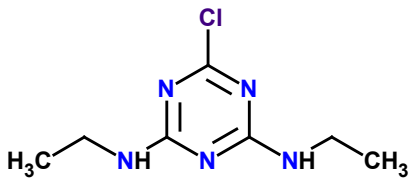
$C_7H_{12}ClN_5$ *m/z 201.0781 Da*



Terbutylazine-desethyl
Detects: 9; # Refs: 92



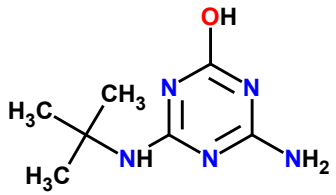
Sebutylazine-desethyl
Detects: 1; # Refs: 14



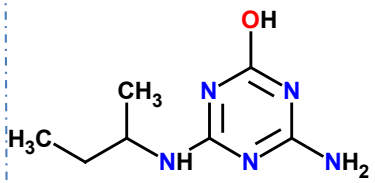
Simazine
Detects: 4; # Refs: 518

(no related compound at this mass)

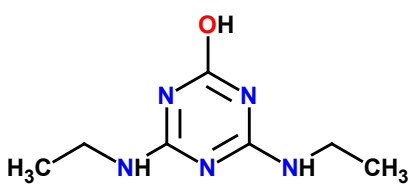
$C_7H_{13}N_5O$ *m/z 183.1120 Da*



Terbutylazine-desethyl-2-hydroxy
Detects: 2; # Refs: 57



Sebutylazine-desethyl-2-hydroxy
Detects: 0; # Refs: 3

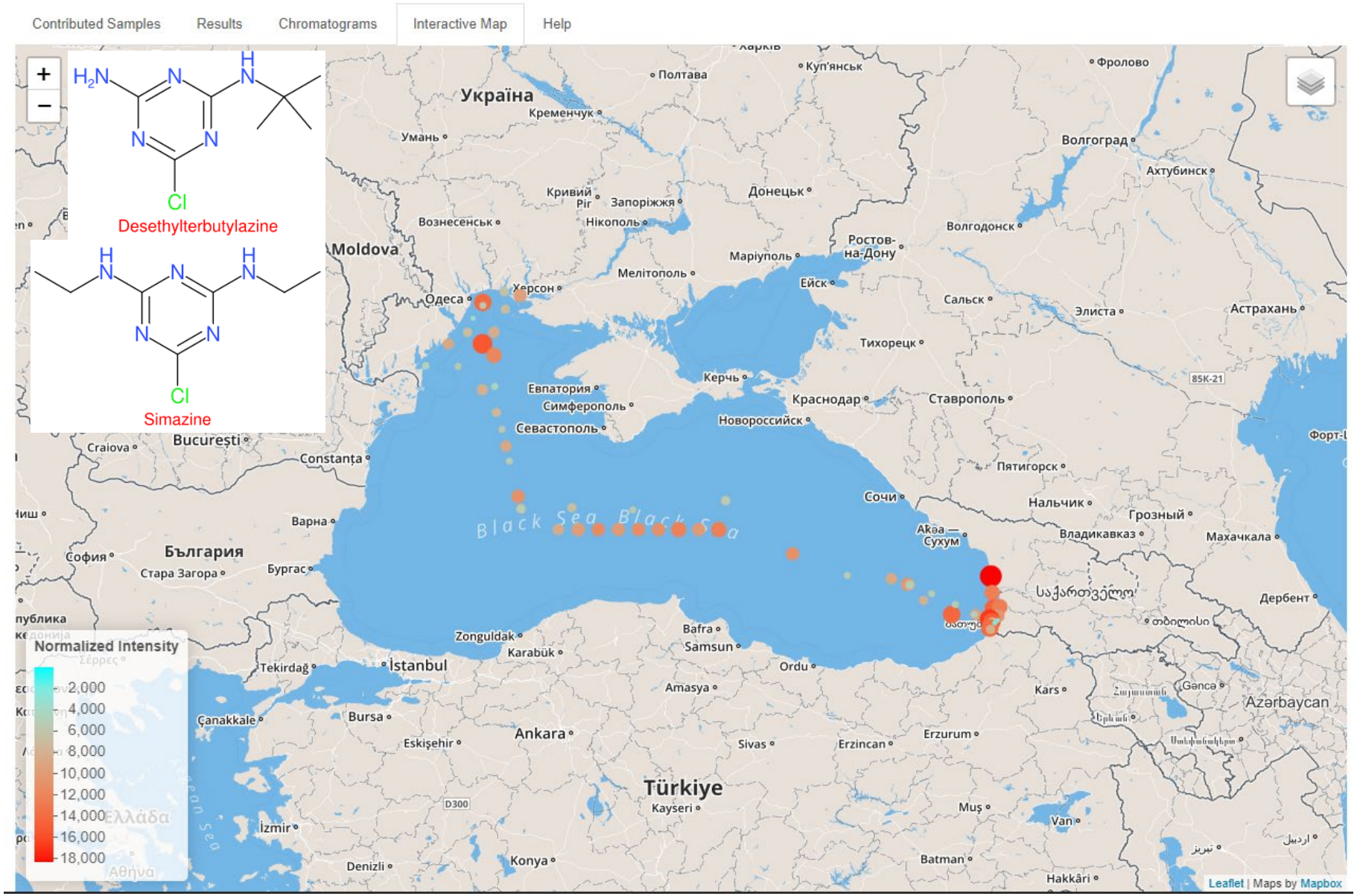


Simazine-2-hydroxy
Detects: 2; # Refs: 66

(no related compound at this mass)

Schymanski *et al*. 2015, ABC,
DOI: 10.1007/s00216-015-8681-7

MetFrag Example: Isobars, Metadata & Measured Data



Digital Sample Freezing Platform

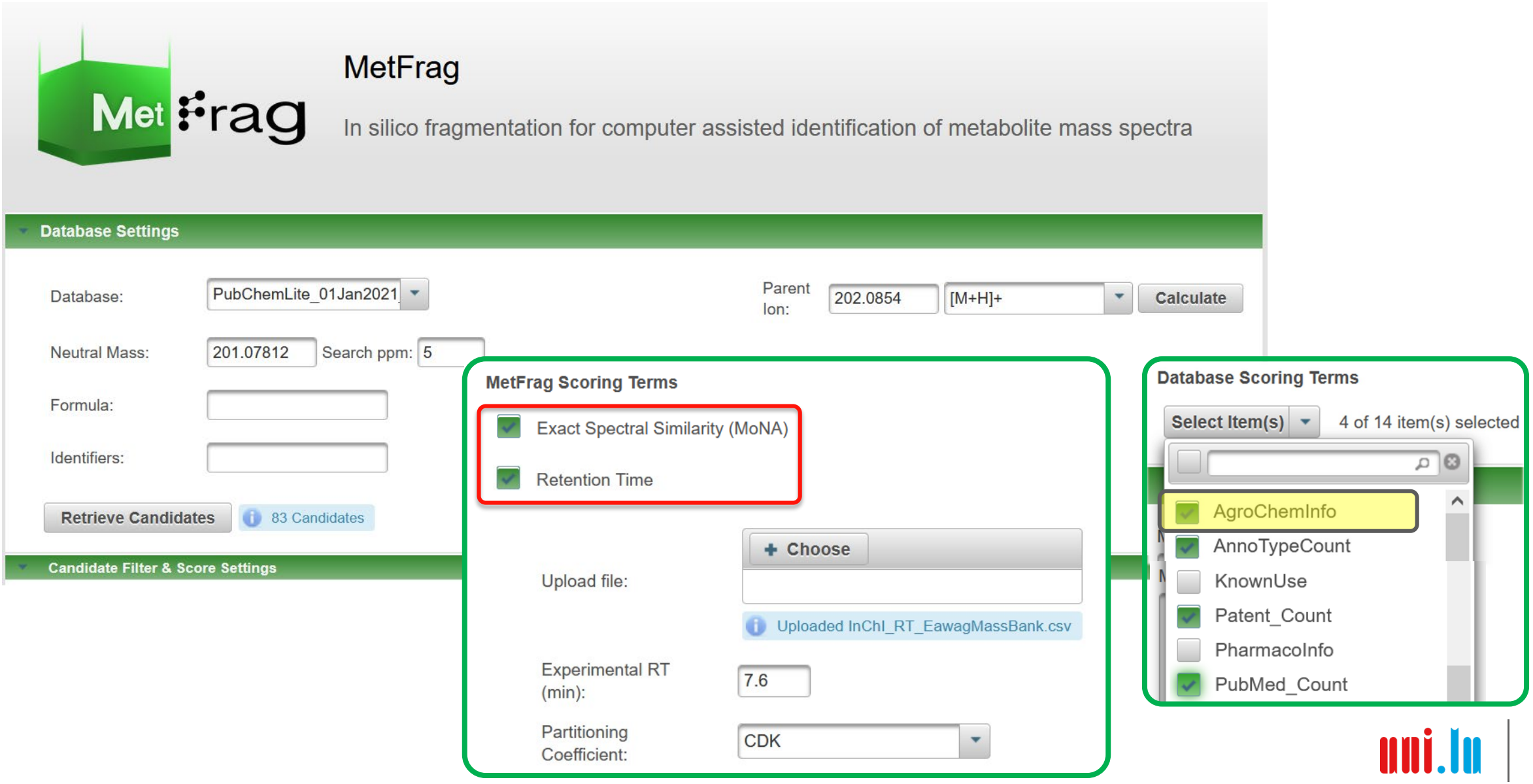
A database of mass chromatograms obtained by LC-HR-MS for retrospective screening of environmental samples

Two very common isomers …

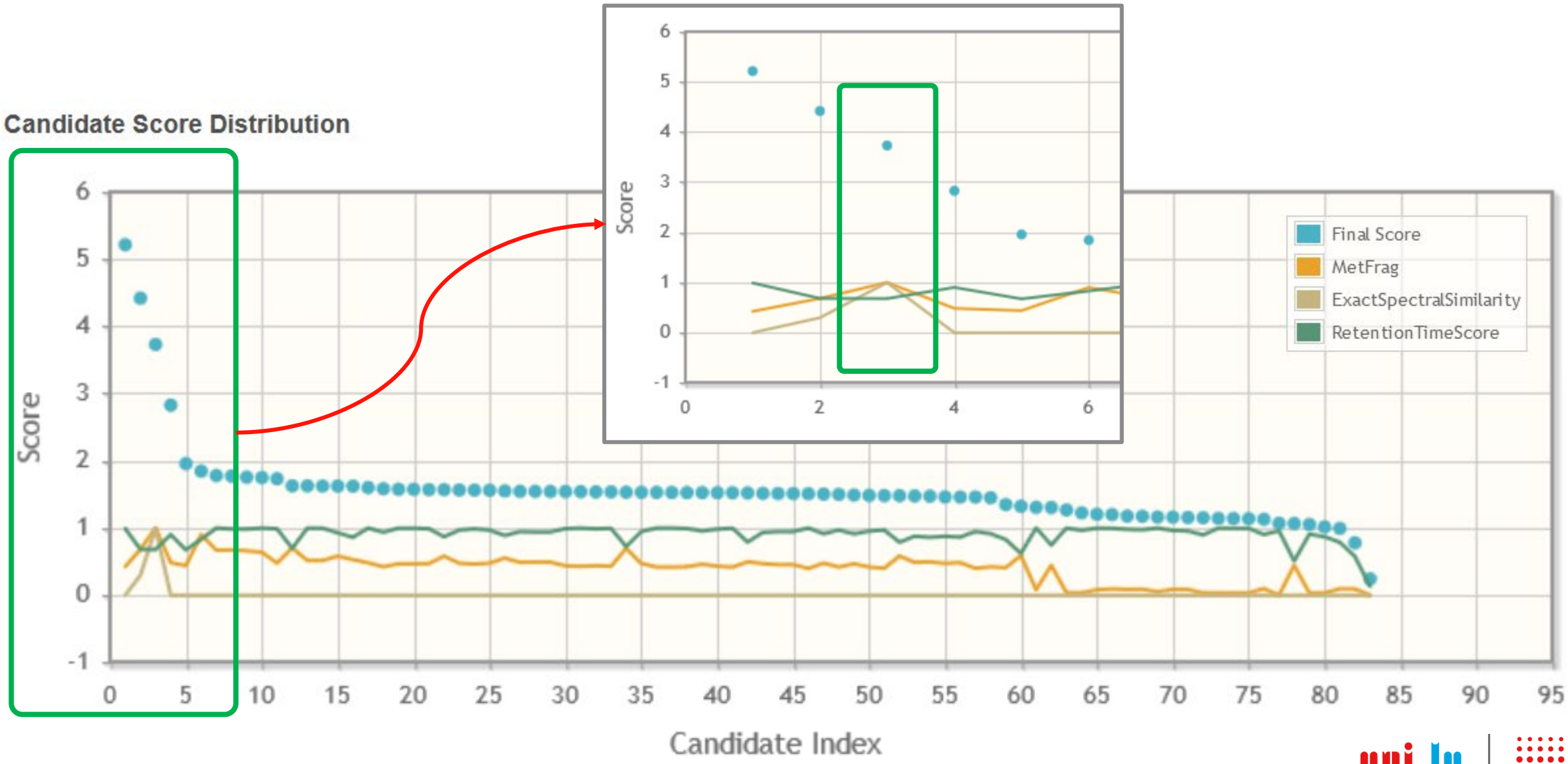
Joint Black Sea Survey 2016

Image provided by Nikiforos Alygizakis. DSFP: Alygizakis *et al*, 2019, TrAC, DOI: 10.1016/j.trac.2019.04.008

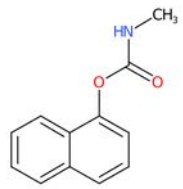
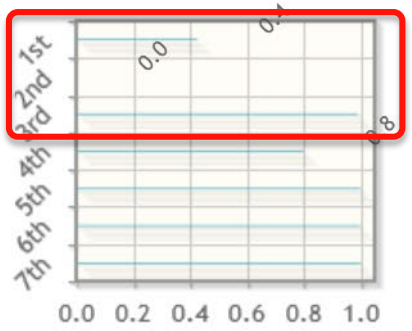
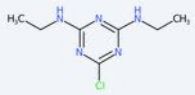
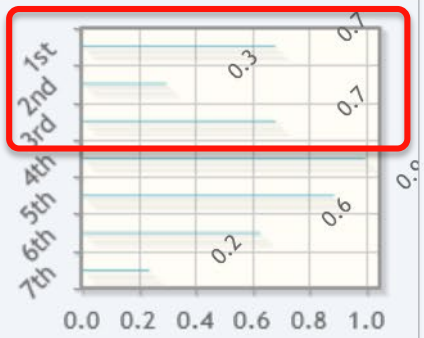
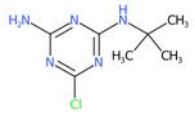
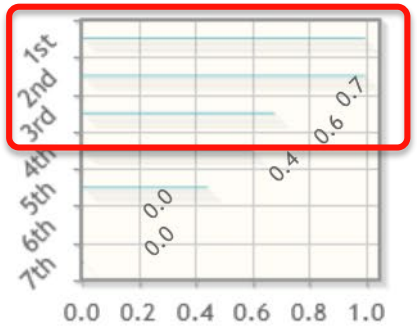
MetFrag & Measured Data – Retention Time and logP



MetFrag & Measured Data – Retention Time and logP



MetFrag & Measured Data – Retention Time and logP

#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1	 naphthalen-1-yl N-methylcarbamate	InChIKeyBlock1 = CVXBEEMKQHEXEN				5.2153	Peaks: 2 / 8 Fragments Scores Download
2	 6-chloro-2-N,4-N-diethyl-1,3,5-triazine-2,4-diamine	InChIKeyBlock1 = ODCWYMIRDDJXKW				4.4191	Peaks: 4 / 8 Fragments Scores Download
3	 2-N-tert-butyl-6-chloro-1,3,5-triazine-2,4-diamine	InChIKeyBlock1 = LMKQNTMFZLAJDV				3.7316	Download

Fragmenter Score = 0.4
Spectral Match = 0,
Retention Time = 0.99 (of 1)

Fragmenter Score = 0.68
Spectral Match = 0.2976
Retention Time = 0.7 (of 1)

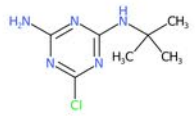
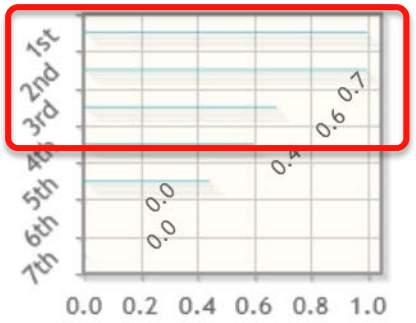
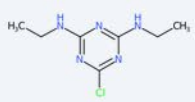
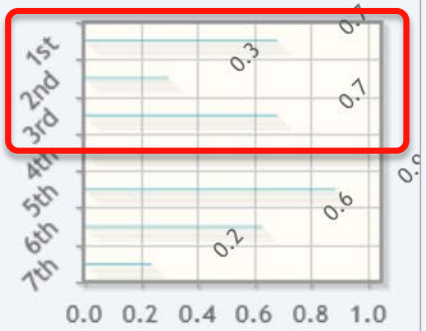
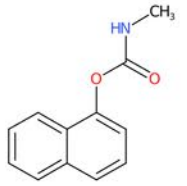
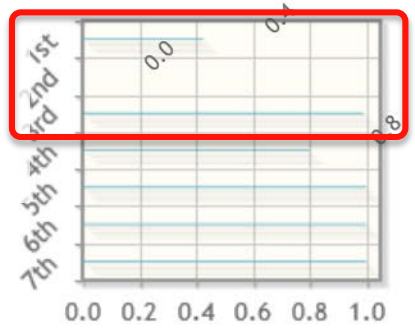
Fragmenter Score = 1.0
Spectral Match = 1.000
Retention Time = 0.7 (of 1)

Weights	
MetFrag (1st)	100 %
ExactSpectralSimilarity (2nd)	100 %
RetentionTimeScore (3rd)	100 %
AgroChemInfo (4th)	100 %
AnnoTypeCount (5th)	100 %
Patent_Count (6th)	100 %
PubMed_Count (7th)	100 %

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

Schymanski *et al*. (2021) DOI: 10.1186/s13321-016-0115-9 and https://massbank.eu/MassBank/RecordDisplay?id=EA067112

MetFrag & Measured Data – Retention Time and logP

#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1	 2-N-tert-butyl-6-chloro-1,3,5-triazine-2,4-diamine	LMKQNTMFZLAJDV				3.7242	Download
2	 6-chloro-2-N,4-N-diethyl-1,3,5-triazine-2,4-diamine	ODCWYMIRDDJXKW				3.5522	Peaks: 4 / 8 Fragments Scores Download
3	 naphthalen-1-yl N-methylcarbamate	CVXBEEMKQHEXEN				3.2153	Peaks: 2 / 8 Fragments Scores Download

Fragmenter Score = 1.0
Spectral Match = 1.000
Retention Time = 0.7 (of 1)

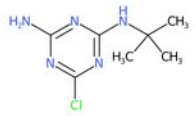
Fragmenter Score = 0.68
Spectral Match = 0.2976
Retention Time = 0.7 (of 1)

Fragmenter Score = 0.4
Spectral Match = 0
Retention Time = 0.99 (of 1)

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

Schymanski *et al*. (2021) DOI: 10.1186/s13321-016-0115-9 and https://massbank.eu/MassBank/RecordDisplay?id=EA067112

MetFrag & Measured Data – Retention Time and logP

#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1	 2-N-tert-butyl-6-chloro-1,3,5-triazine-2,4-	108201 InChIKeyBlock1 = LMKQNTMFZLAJDV	201.07812	$C_7H_{12}ClN_5$		3.7242	Download

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

PubChem Desethylterbuthylazine (Compound)

7 Agrochemical Information

7.1 Agrochemical Category

Pesticides -> Herbicides -> Triazine herbicides -> Chlorotriazine herbicides -> Transformation products

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

NORMAN Suspect List Exchange

7.2 Agrochemical Transformations

Terbutylazine-desethyl has known environmental transformation products that include Terbutylazine-desethyl-2-hydroxy.

Terbutylazine-desethyl is a known environmental transformation product of Terbutylazine.

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

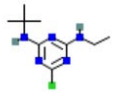
NORMAN Suspect List Exchange

PubChem Desethylterbuthylazine (Compound)

8.1 Transformations

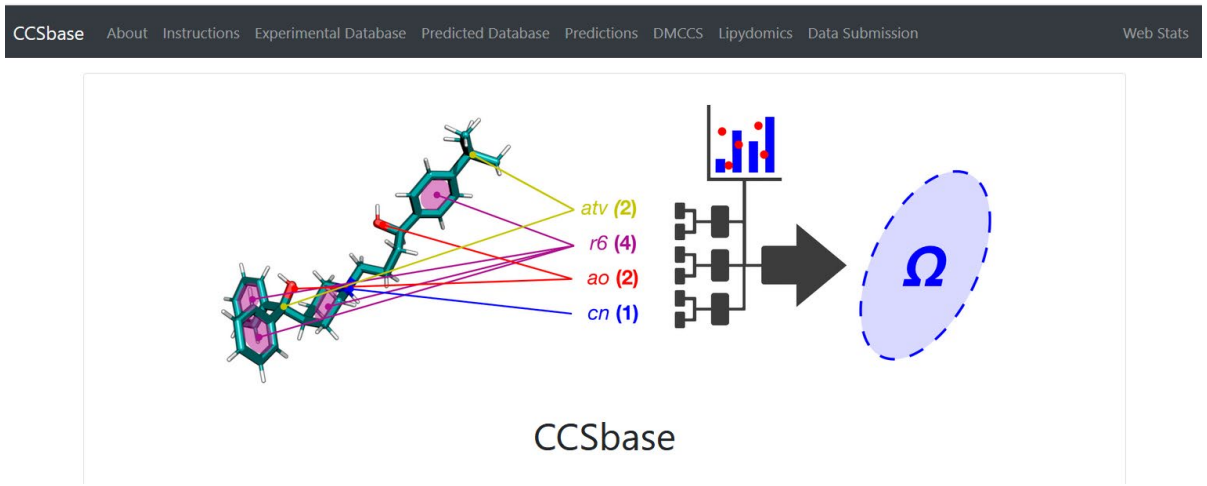
7 items View More Rows & Details Download

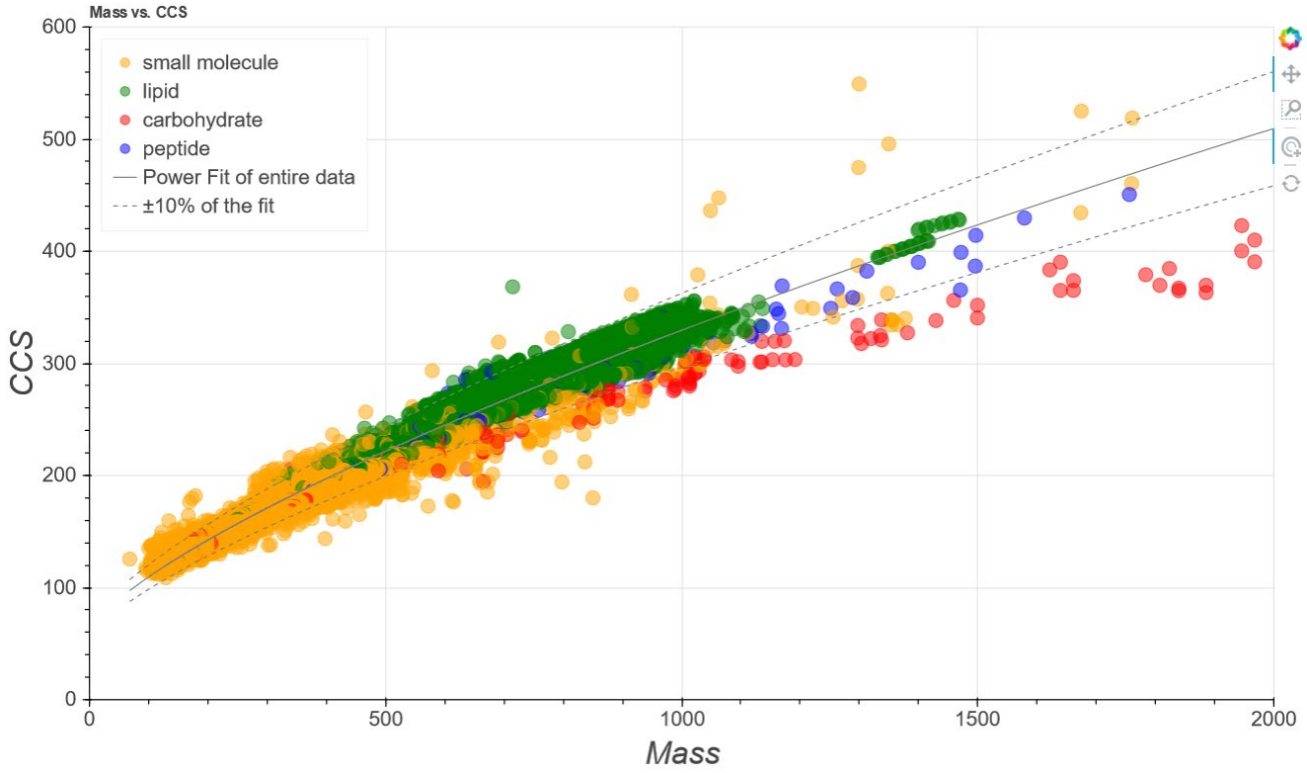
SORT BY Please Choose One

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme
	terbuthylazine		desethyl-terbuthylazine	Environmental	

Schymanski *et al*. (2021) DOI: 10.1186/s13321-016-0115-9 and https://massbank.eu/MassBank/RecordDisplay?id=EA067112

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

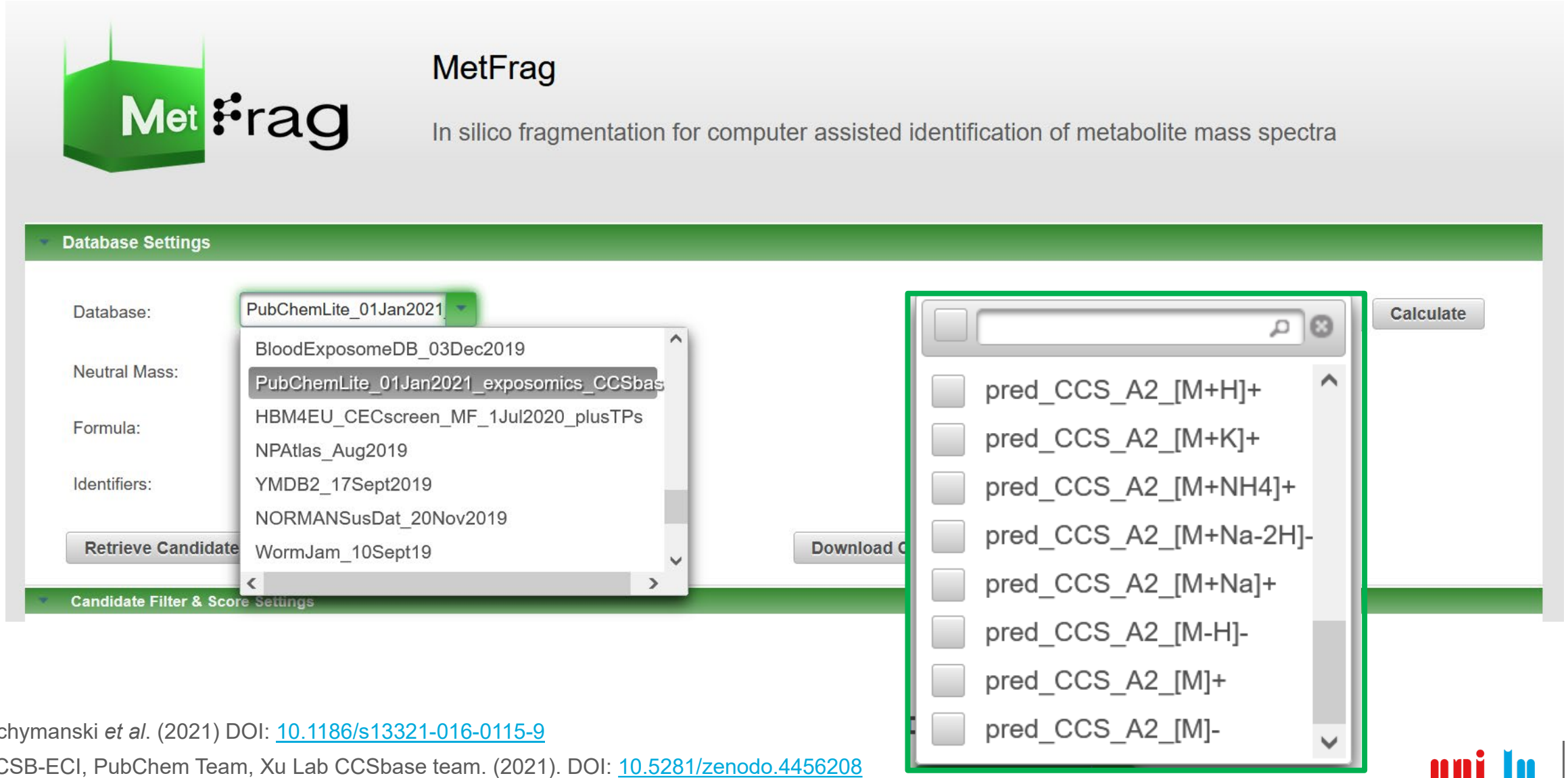
See more details...

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

Isobars: MetFrag + PubChemLite + CCSbase

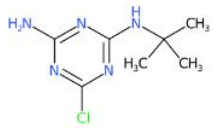
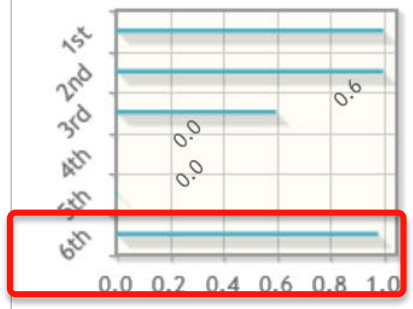
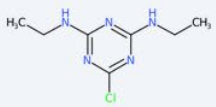
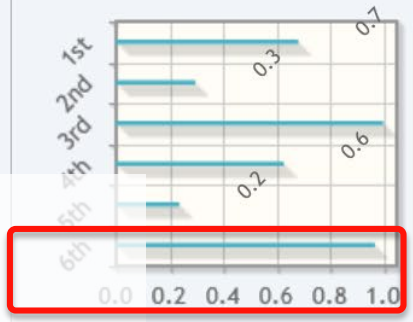
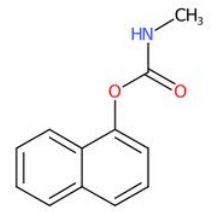
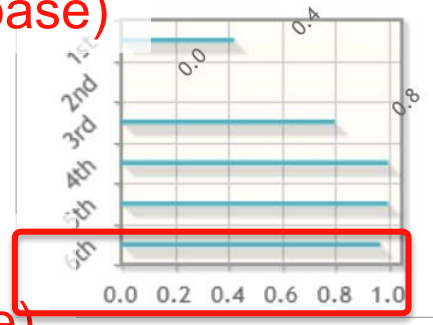


Schymanski *et al*. (2021) DOI: 10.1186/s13321-016-0115-9

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.

Isobars: MetFrag + PubChemLite + CCSbase

#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1	 2-N-tert-butyl-6-chloro-1,3,5-triazine-2,4-diamine	108201 InChIKeyBlock1 = [illegible]	201.07812	$C_7H_{12}ClN_5$		2.6	Download
2	 6-chloro-2-N,4-N-diethyl-1,3,5-triazine-2,4-diamine	5216 InChIKeyBlock1 = [illegible]	201.07812	$C_7H_{12}ClN_5$		1.98	Peaks: 4 / 8 Fragments Scores Download
3	 naphthalen-1-yl N-methylcarbamate	6129 InChIKeyBlock1 = CVXBEEMKQHEXEN	201.07898	$C_{12}H_{11}NO_2$		1.2241	

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

CCS (pred) = 143.7
CCS (exp) = 144.71 (UJI)

CCS (pred) = 141.9
CCS (exp) = 143 (UJI)
CCS (exp) = 142.09 (CCSbase)

CCS (pred) = 142.2
CCS (exp) = 150 (CCSbase)

CCS values all $[M+H]^+$

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

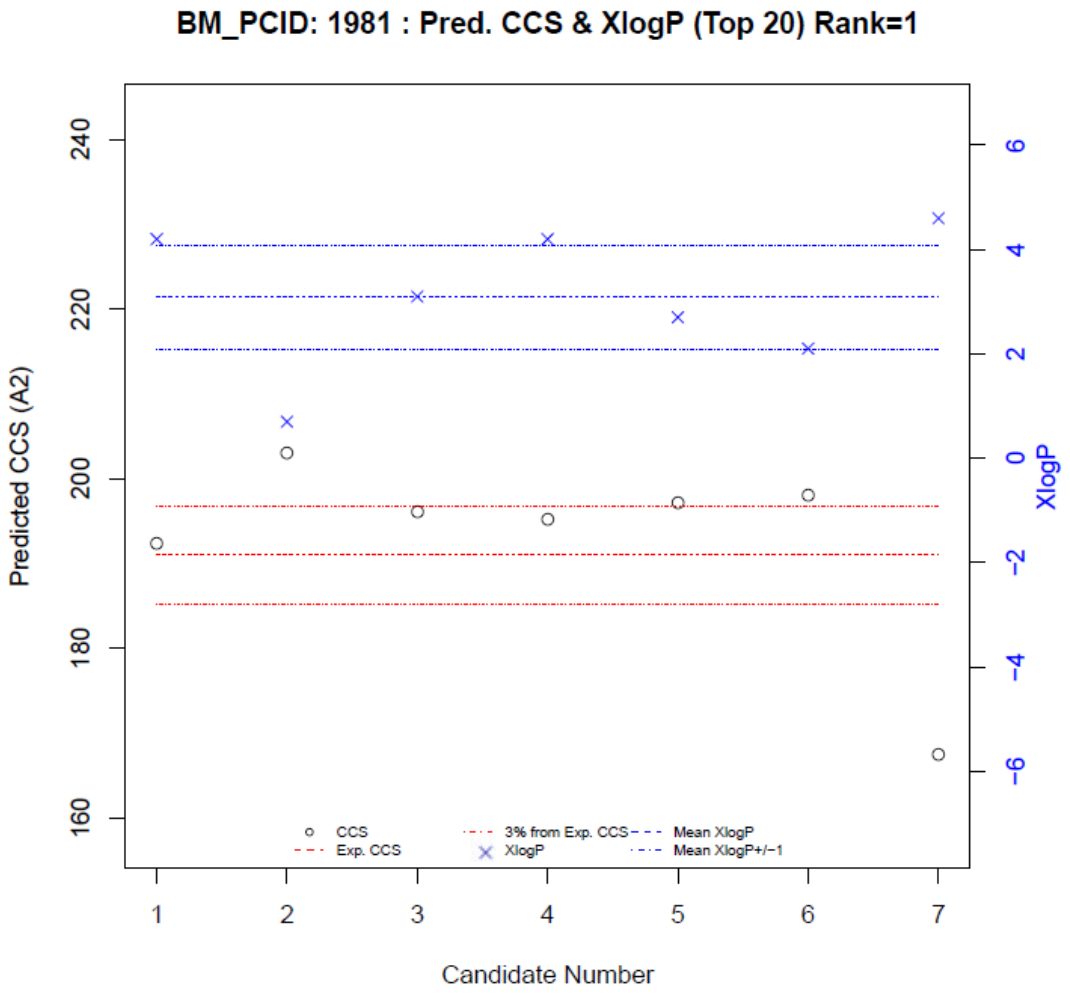
Incorporating CCS in Confidence Level Scheme

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

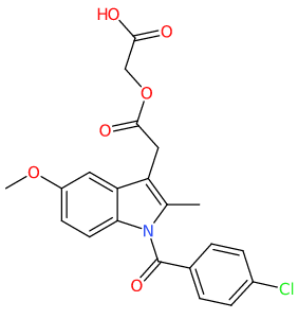
Confidence

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

PubChemLite EXPOSOMICS … plus CCSbase!



PubChem Acemetacin (Compound)

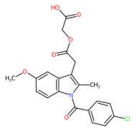
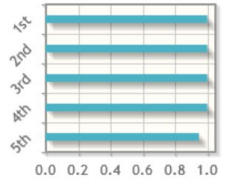
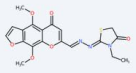
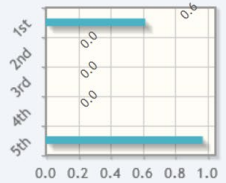


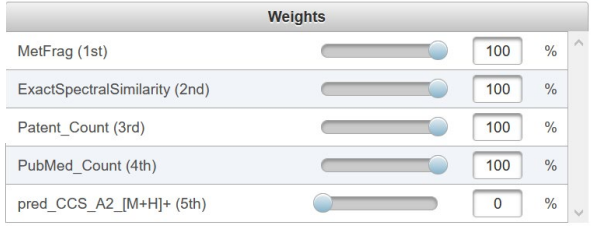
3.2.6 Collision Cross Section

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	$C_{21}H_{18}ClNO_6$		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	$C_{19}H_{17}N_3O_6S$		0.6166	Peaks: 2 / 6 Fragments Scores Download

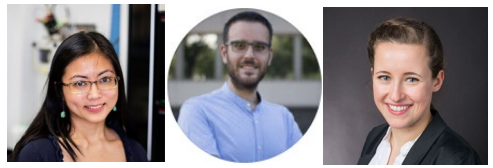


Schymanski *et al*. (2021) JCheminform. DOI: 10.1186/s13321-016-0115-9

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

CCS Values in PubChem


PubChem Annotations


SLE Lists S50, S61, S79


CCSbase

PubChem Classification Browser

Browse PubChem data using a classification of interest, or search for PubChem records annotated with the desired classification/term (e.g., MeSH: phenylpropionates, or Gene Ontology: DNA repair). More...

Select classification

PubChem: Aggregated CCS Classification

Search selec

Keyw

Classification description (from PubChem)

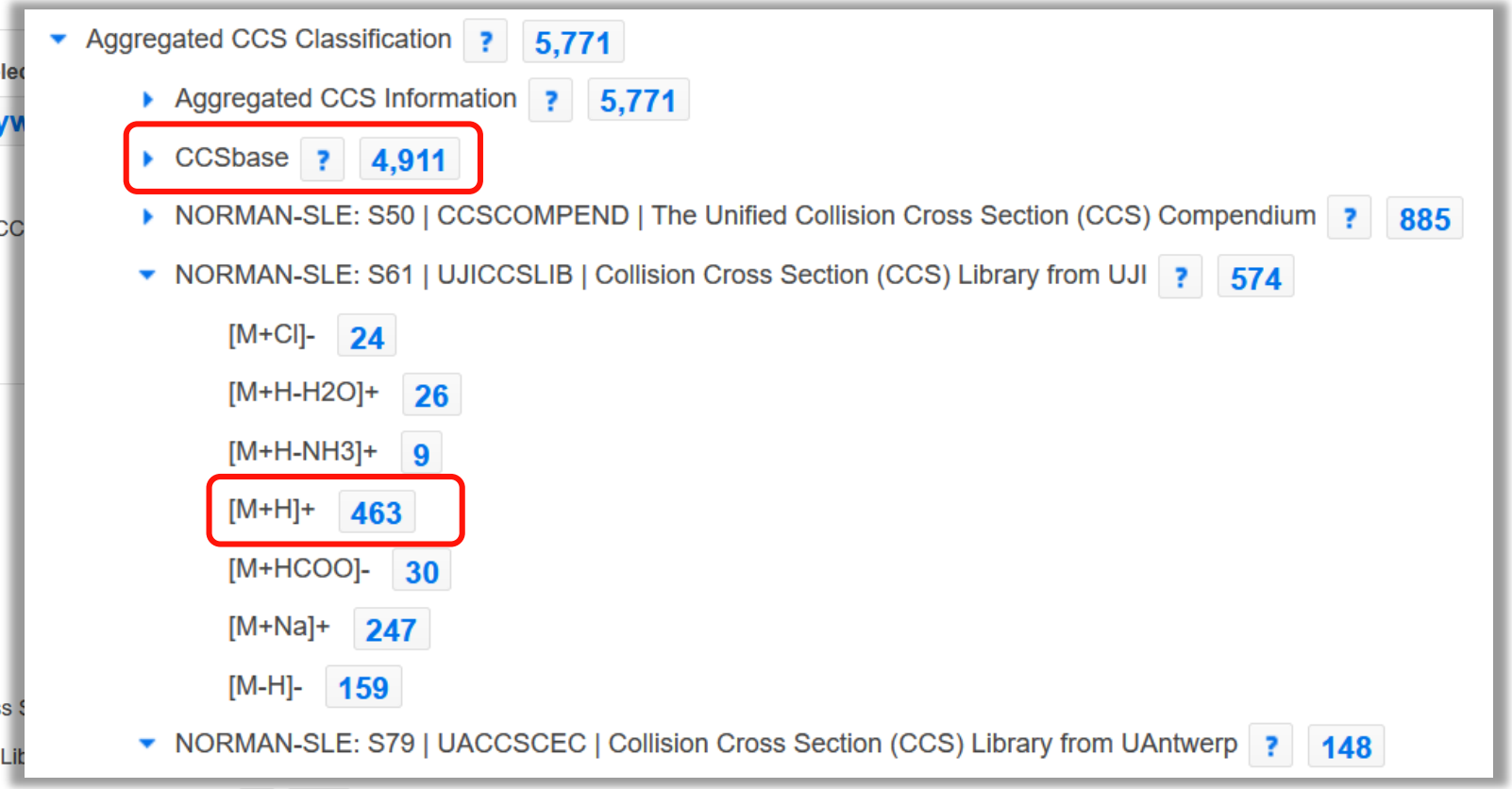
This classification aggregates experimental CCS data available in PubChem, from CC

Data type counts to display None | Compound

Display zero count nodes? Yes | No

Browse PubChem: Aggregated CCS Classification Tree

- Aggregated CCS Classification ? 5,771
  - Aggregated CCS Information ? 5,771
  - CCSbase ? 4,911
  - NORMAN-SLE: S50 | CCSCOMPEND | The Unified Collision Cross S
  - NORMAN-SLE: S61 | UJICCSLIB | Collision Cross Section (CCS) Lib
  - NORMAN-SLE: S79 | UACCSCEC | Collision Cross Section (CCS) Library from UAntwerp ? 148



https://pubchem.ncbi.nlm.nih.gov/classification/#hid=106

norman suspects

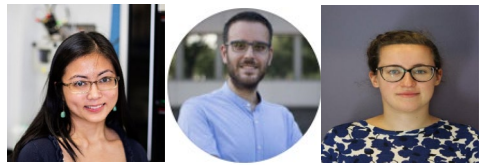
UNIVERSITY of WASHINGTON

uni.lu

LCSB

CCS Values in PubChem

PubChem Annotations



SLE Lists S50, S61, S79

CCSbase

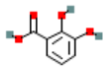
SEARCH FOR

PubChem: Aggregated CCS Classification: CCSbase

Treating this as a previously computed list of identifiers.

Compounds

4,911 results | Filters | SORT BY Relevance



2,3-Dihydroxybenzoic Acid; 303-38-8; Pyrocatechuic Acid; O-Pyrocatechuic Acid; 2-Pyrocatechuic Acid; ...

Compound CID: 19

MF: $C_7H_6O_4$ MW: 154.12g/mol

IUPAC Name: 2,3-dihydroxybenzoic acid

Isomeric SMILES: C1=CC(=C(C(=C1)O)O)C(=O)O

InChIKey: GLDQAMYCGOIJDV-UHFFFAOYSA-N

InChI: InChI=1S/C7H6O4/c8-5-3-1-2-4(6(5)9)7(10)11/h1-3,8-9H,(H,10,11)

Create Date: 2004-09-16

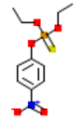
SEARCH FOR

PubChem: Aggregated CCS Classification: [M+H]+

Treating this as a previously computed list of identifiers.

Compounds

463 results | Filters | SORT BY Relevance | Download



Parathion; Ethyl Parathion; Thiophos; Parathion-ethyl; 56-38-2; ...

Compound CID: 991

MF: $C_{10}H_{14}NO_5PS$ MW: 291.26g/mol

IUPAC Name: diethoxy-(4-nitrophenoxy)-sulfanylidene-lambda5-phosphane

Isomeric SMILES: CCOP(=S)(OCC)OC1=CC=C(C=C1)[N+](=O)[O-]

InChIKey: LCCNCVORNKJIRZ-UHFFFAOYSA-N

InChI: InChI=1S/C10H14NO5PS/c1-3-14-17(18,15-4-2)16-10-7-5-9(6-8-10)11(12)13/h5-8H,3-4H2,1-2H3

Create Date: 2004-09-16

ACTIONS ON RESULTS WITH ID TYPE:

Compounds

Push to Entrez | Save for Later | Linked Data Sets

https://pubchem.ncbi.nlm.nih.gov/classification/#hid=106

norman suspects

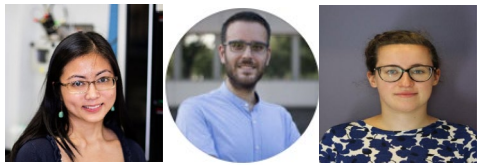
UNIVERSITY of WASHINGTON

uni.lu

LCSB

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

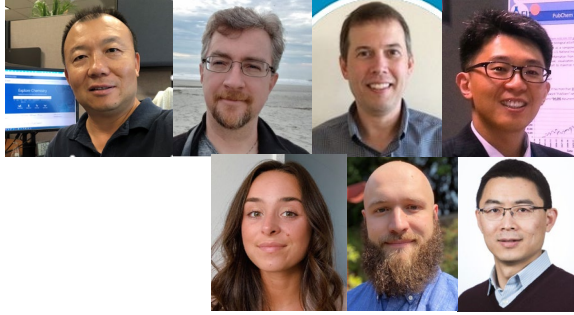
https://pubchem.ncbi.nlm.nih.gov/compound/2256#section=Collision-Cross-Section

norman suspects

UNIVERSITY of WASHINGTON

LCSB

Calculated CCS Values in PubChem ????



PubChem Atrazine (Compound)

3.1 Computed Properties

Property Name	Property Value	Reference
Molecular Weight	215.68	Computed by PubChem 2.1 (PubChem release 2021.05.07)
XLogP3	2.6	Computed by XLogP3 3.0 (PubChem release 2021.05.07)
Hydrogen Bond Donor Count	2	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Hydrogen Bond Acceptor Count	5	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Rotatable Bond Count	4	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)
Exact Mass	215.0937732	Computed by PubChem 2.1 (PubChem release 2021.05.07)
Monoisotopic Mass	215.0937732	Computed by PubChem 2.1 (PubChem release 2021.05.07)
Topological Polar Surface Area	62.7 Å²	Computed by Cactvs 3.4.8.18 (PubChem release 2021.05.07)

PubChem Atrazine (Compound)

3.2.16 Collision Cross Section

150.74 Å² [M+H]⁺ [CCS Type: DT, Method: stepped-field]

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

CCSbase

140.53 Å² [M+H]⁺

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

NORMAN Suspect List Exchange

149.26 Å² [M+H]⁺

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

NORMAN Suspect List Exchange

149.8 Å² [M+H]⁺

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

NORMAN Suspect List Exchange

Are we ready for calculated CCS in PubChem?
(we have the values already …)

https://pubchem.ncbi.nlm.nih.gov/compound/2554#section=Collision-Cross-Section
https://pubchem.ncbi.nlm.nih.gov/compound/2256#section=Collision-Cross-Section

norman suspects

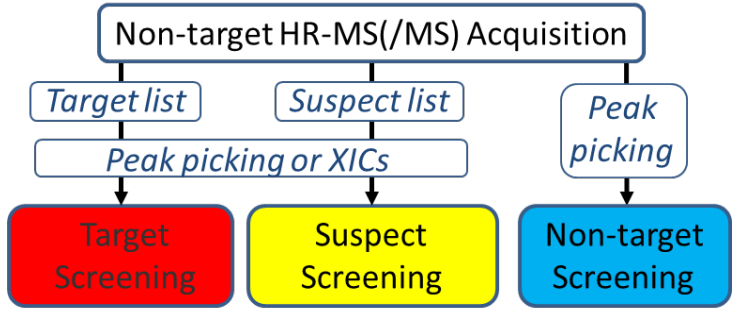
UNIVERSITY of WASHINGTON

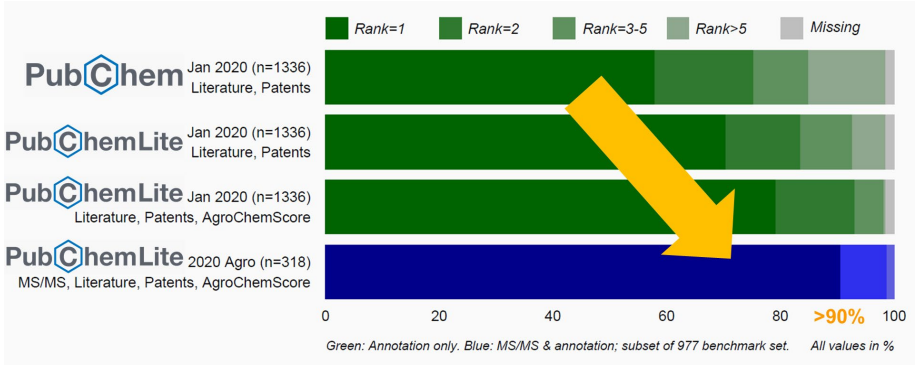
uni.lu LCSB

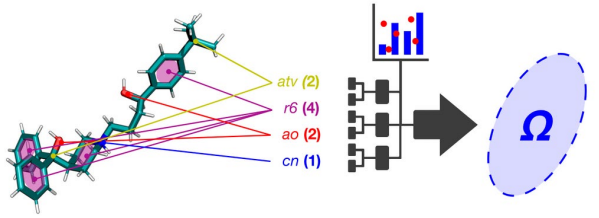
“Take home” Messages

PubChemLite EXPOSOMICS

- Introduction to HR-MS, cheminformatics and identification confidence level scheme
- Identification with MetFrag and PubChemLite
  - Mass spectral libraries help deliver Level 2a IDs
  - Annotation content is extremely powerful in assisting interpretation
- Measured data is critical!
  - log P / retention time can be integrated
  - Collision Cross Section (CCS) integration available
- ***Help contribute by adding your knowledge!***







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

Expert Knowledge is YOUR Knowledge!

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

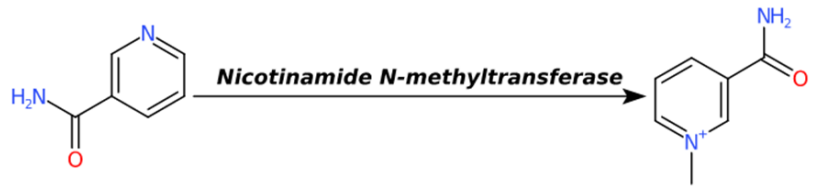
- Help us help you! Add data with FAIR templates

Chemical Structures

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

Transformations

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



Nicotinamide to MNAM DOI:10.1124/dmd.112.049734

FAIR chemical structures in the Journal of Cheminformatics

Emma L. Schymanski and Evan E. Bolton

Letter to the Editor | 7 July 2021

The Letter Response to this article has been published in *Journal of Cheminformatics* 2021 **13**:49

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

Rajarshi Guha, Nina Jeliazkova, Egon Willighagen and Barbara Zdrazil

Letter Response | 7 July 2021

The Letter to the Editor to this article has been published in *Journal of Cheminformatics* 2021 **13**:50

Schymanski & Bolton (submitted) FAIR-ifying the Exposome: Preprint DOI: 10.5281/zenodo.5495109

Schymanski & Bolton (2021) FAIR Chemical Structures. *J. Cheminform*. DOI: 10.1186/s13321-021-00520-4

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

Expert Knowledge is YOUR Knowledge!

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

- Help us help you! Add data with FAIR templates

  CCS values!

Name	SMILES	PubChem_CID	CCS_A2	Adduct	Comment	Reference_ID
1-Methylnicotinamide	C[N+]1=CC=CC(=C1)C(=O)N	457	126.40	[M]+	CCS Type: DT, Method: single field calibrated with Agilent tu...	DOI:10.1021/acs.analchem.6b03091
Taurine	C(CS(=O)(=O)O)N	1123	135.40	[M+H-H2O]+	CCS Type: DT, Method: single field calibrated with Agilent tu...	DOI:10.1021/acs.analchem.6b03091
PS(18:3(6Z,9Z,12Z)/22:2(13Z,16Z))	CCCCC/C=C\C/C=C\CCCCCCCCCCCC(=O)O[C	52925504	285.70	[M-H]-	CCS Type: DT, Method: single field calibrated with Agilent tu...	DOI:10.1021/acs.analchem.7b02625
N-Palmitoyldihydrosphingomyelin	CCCCCCCCCCCCCCC[C@H]([C@H](COP(=O)(	9939965	282.30	[M+HCOO]-	CCS Type: DT, Method: single field calibrated with Agilent tu...	DOI:10.1021/acs.analchem.7b02625
Succinic acid	C(CC(=O)O)C(=O)O	1110	122.66	[M+H]+	CCS Type: DT, Method: stepped-field	DOI:10.1039/C7SC03464D
Paxilline	C[C@]12CC[C@H]3C(=CC(=O)[C@H](O3)C(C	105008	201.00	[M+H]+	CCS Type: TW, Method: calibrated with polyalanine	DOI:10.1016/j.aca.2018.01.047
HT-2 Toxin	CC1=C[C@@H]2[C@](C[C@@H]1OC(=O)CC	10093830	203.00	[M+NH4]+	CCS Type: TW, Method: calibrated with polyalanine	DOI:10.1016/j.aca.2018.01.047
Cortodoxone	C[C@]12CCC(=O)C=C1CC[C@@H]3[C@@H]	440707	193.62	[M+Cl]-	CCS Type: TW, Drift Gas: nitrogen	DOI:10.1021/acs.est.0c05713

https://ftp.ncbi.nlm.nih.gov/pubchem/Other/Submissions/CCS_Template.csv

Schymanski & Bolton (submitted) FAIR-ifying the Exposome: Chemical Structures and Transformations. Submitted to *Exposome*. Preprint DOI: 10.5281/zenodo.5495109
Schymanski & Bolton (2021) FAIR Chemical Structures. *J. Cheminform*. DOI: 10.1186/s13321-021-00520-4

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

Thank you!

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

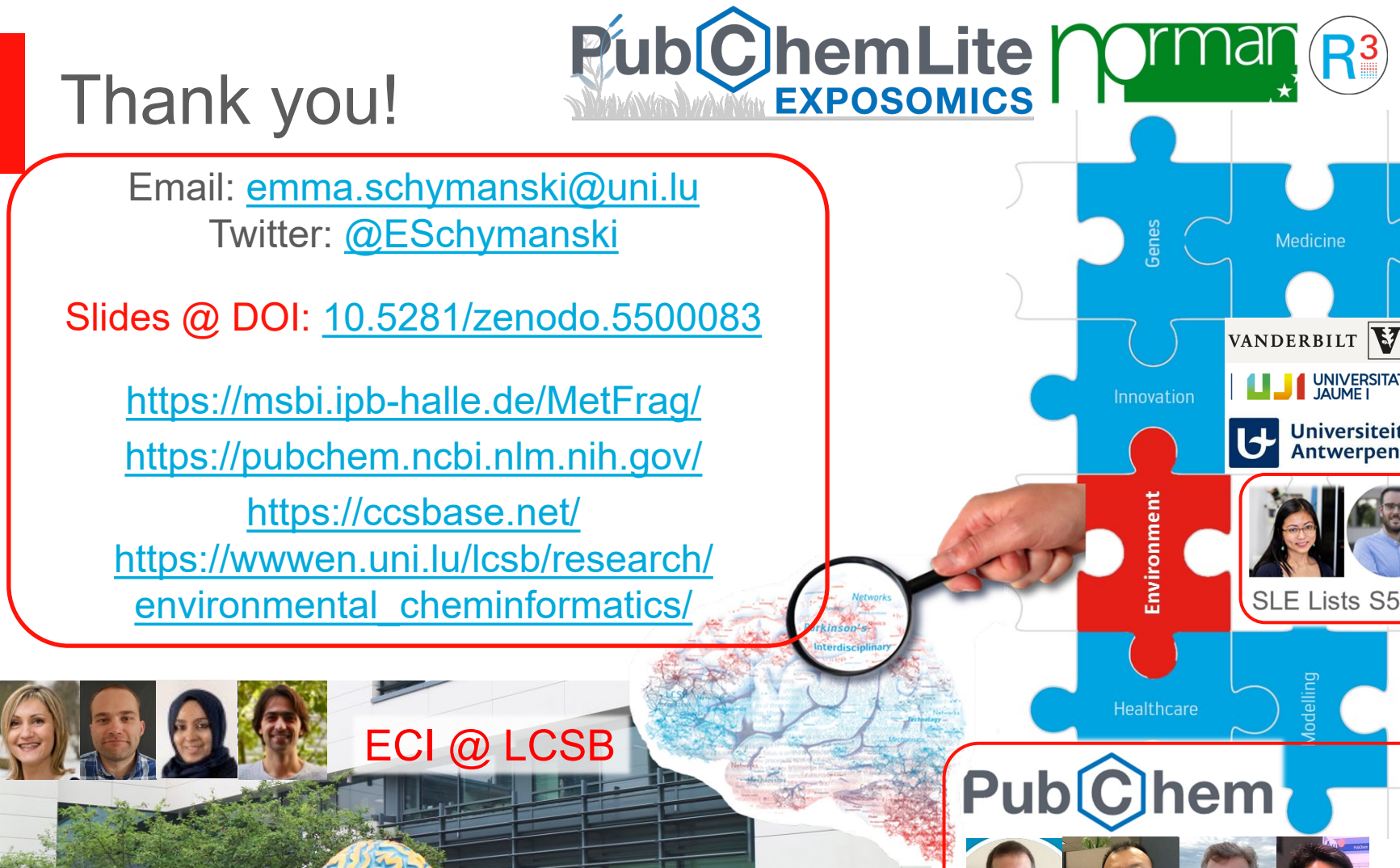
Slides @ DOI: 10.5281/zenodo.5500083

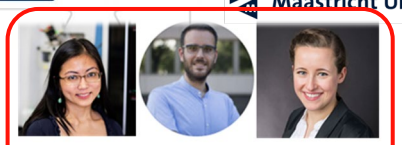
https://msbi.ipb-halle.de/MetFrag/

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

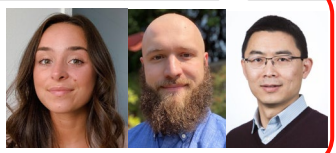
https://ccsbase.net/

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

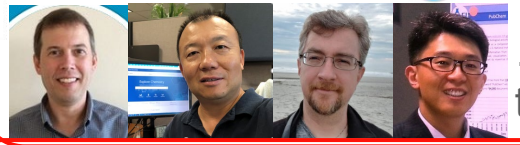



SLE Lists S50, S61, S79

CCSbase
Xu Lab



PubChem



… & team

MetFrag

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

Luxembourg National Research Fund