Luxembourg National Research Fund

NORDIC METABOLOMICS SOCIETY

UNIVERSITÉ DU LUXEMBOURG

Defining a *Manageable*, *Dynamic* Chemical Space for *Exposomics*

PubChemLite EXPOSOMICS

Assoc. Prof. Dr. Emma L. Schymanski

(plus many, many colleagues and collaborators!)

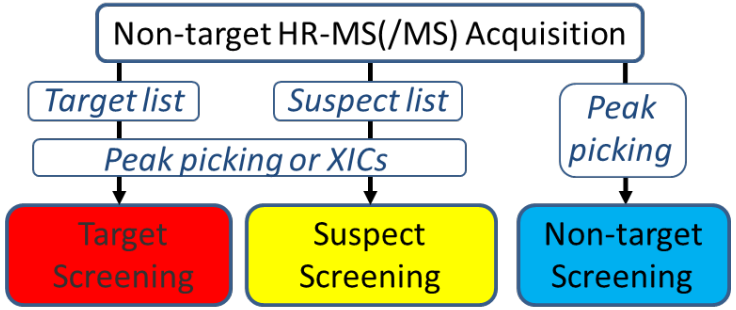
Environmental Cheminformatics Group,
Luxembourg Centre for Systems Biomedicine, University of Luxembourg
Email: emma.schymanski@uni.lu and @ESchymanski
Web: https://wwwen.uni.lu/lcsb/research/environmental_cheminformatics/

The Metabolome Meets the Exposome
Online Workshop, April 28, 2021

Slides available at
DOI: 10.5281/zenodo.4722508

Outline for today

- Problem formulation
  - Why is non-target HR-MS still so hard?



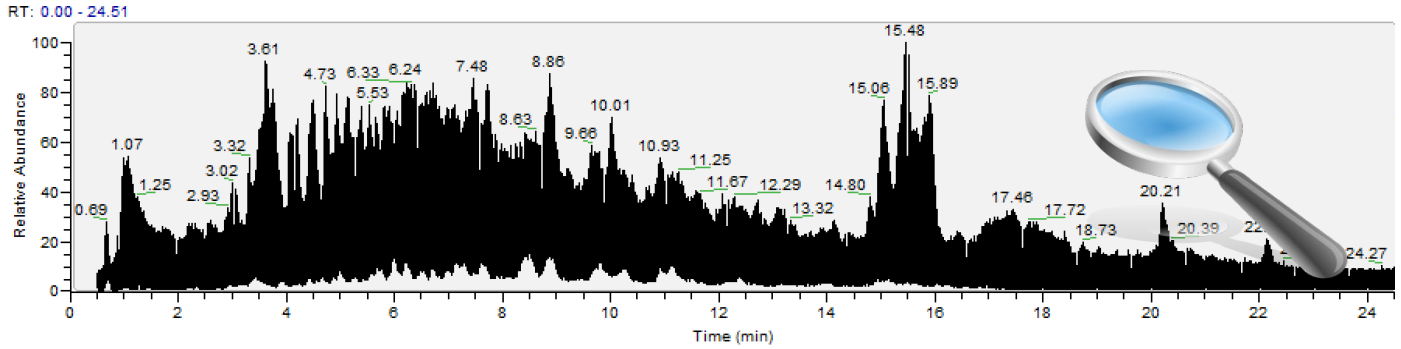
- Redefining our strategy – *i.e.* DIY Databases
  - Take the "good bits" of databases, and ignore the rubble
  - … in a way that keeps many players happy…
  - … and updates as knowledge updates …

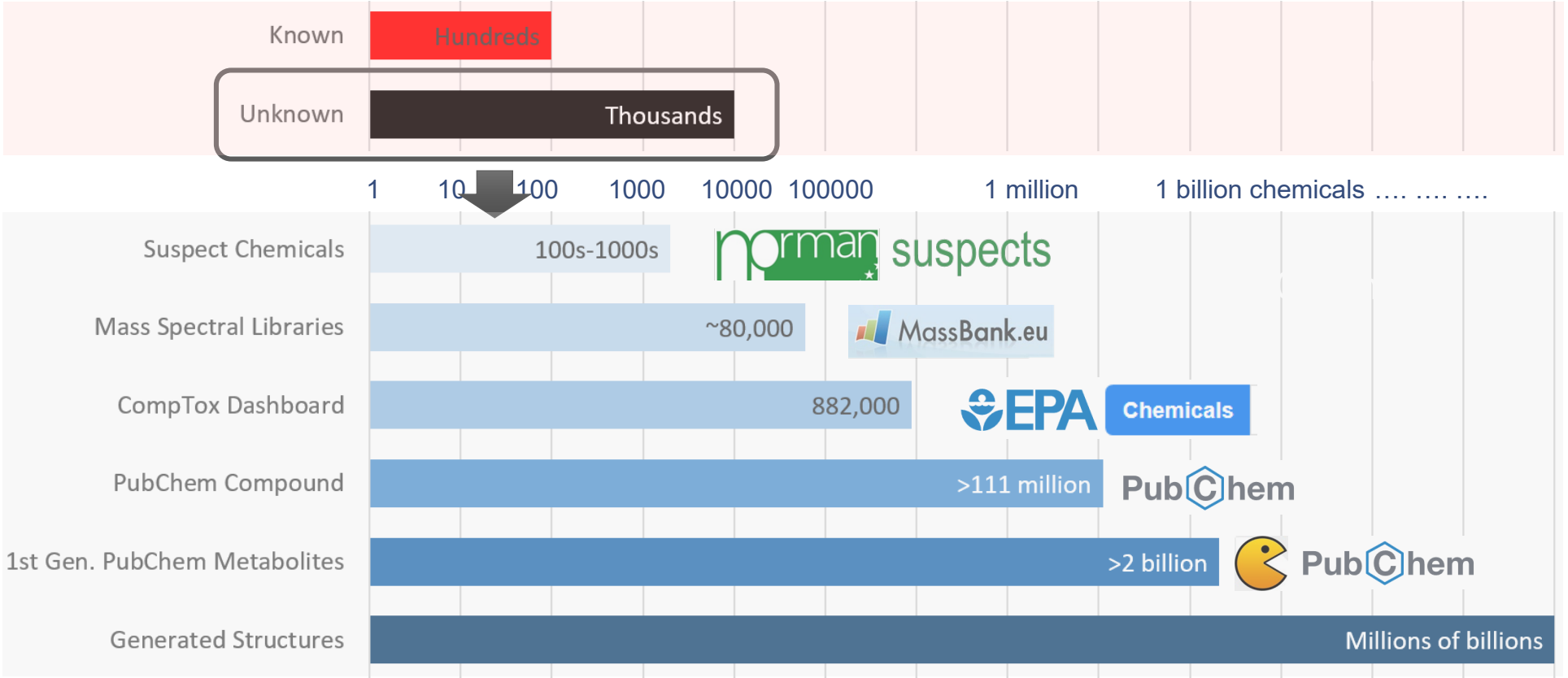
- What's next?

Metabolome/Exposome meets Environmental Cheminformatics

High resolution mass spectrometry

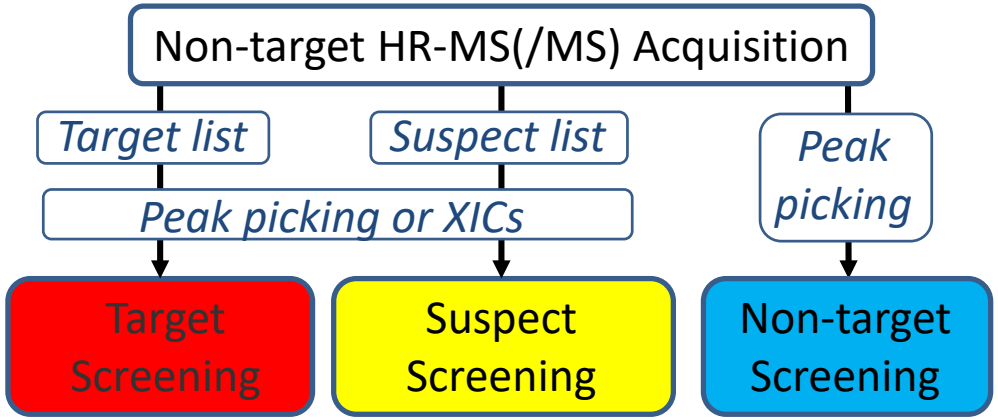
AND connecting chemical knowledge





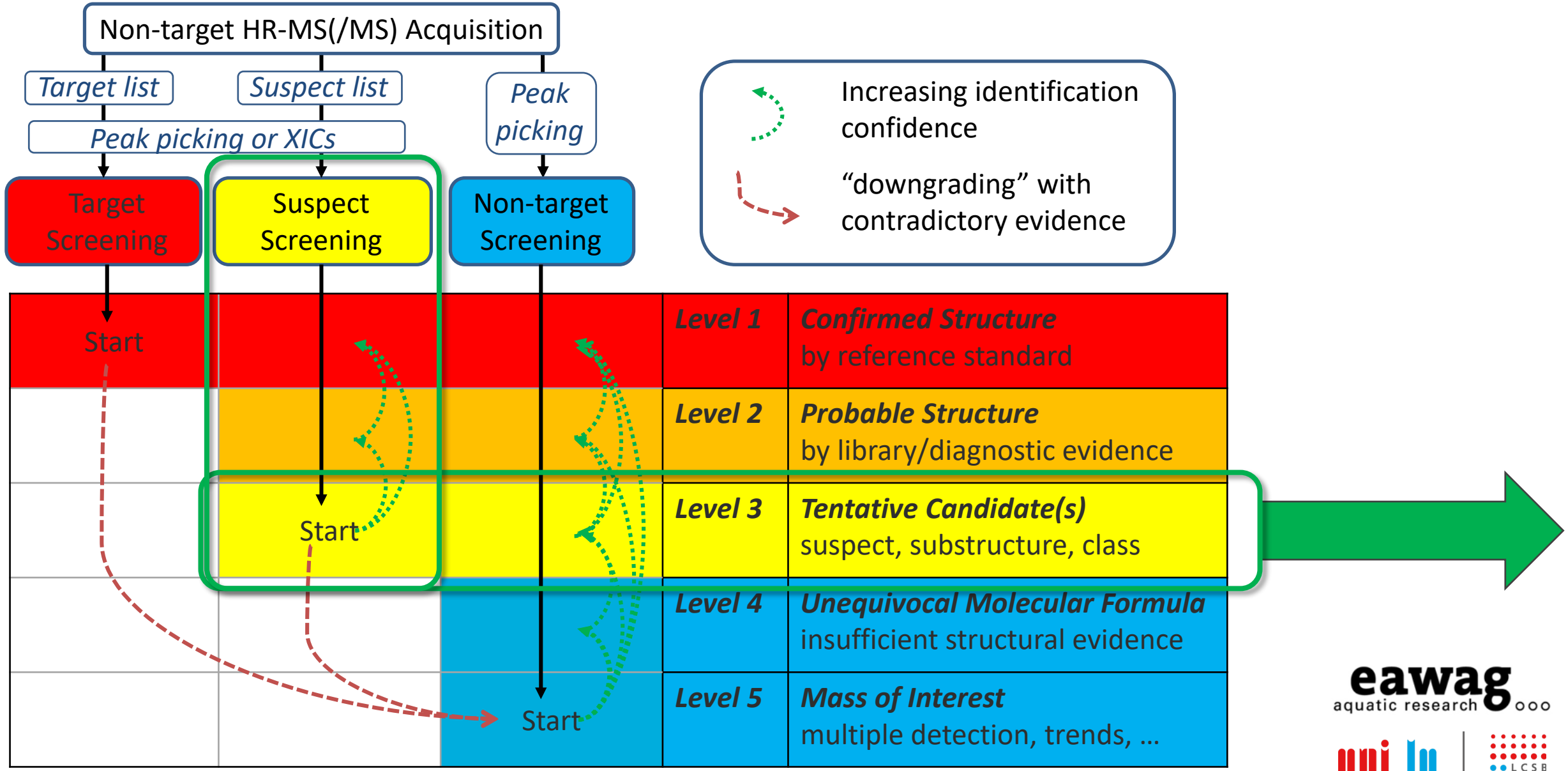
Schymanski *et al.* (2014) DOI: 10.1021/es4044374; Vermeulen *et al.* (2020) DOI: 10.1126/science.aay3164

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

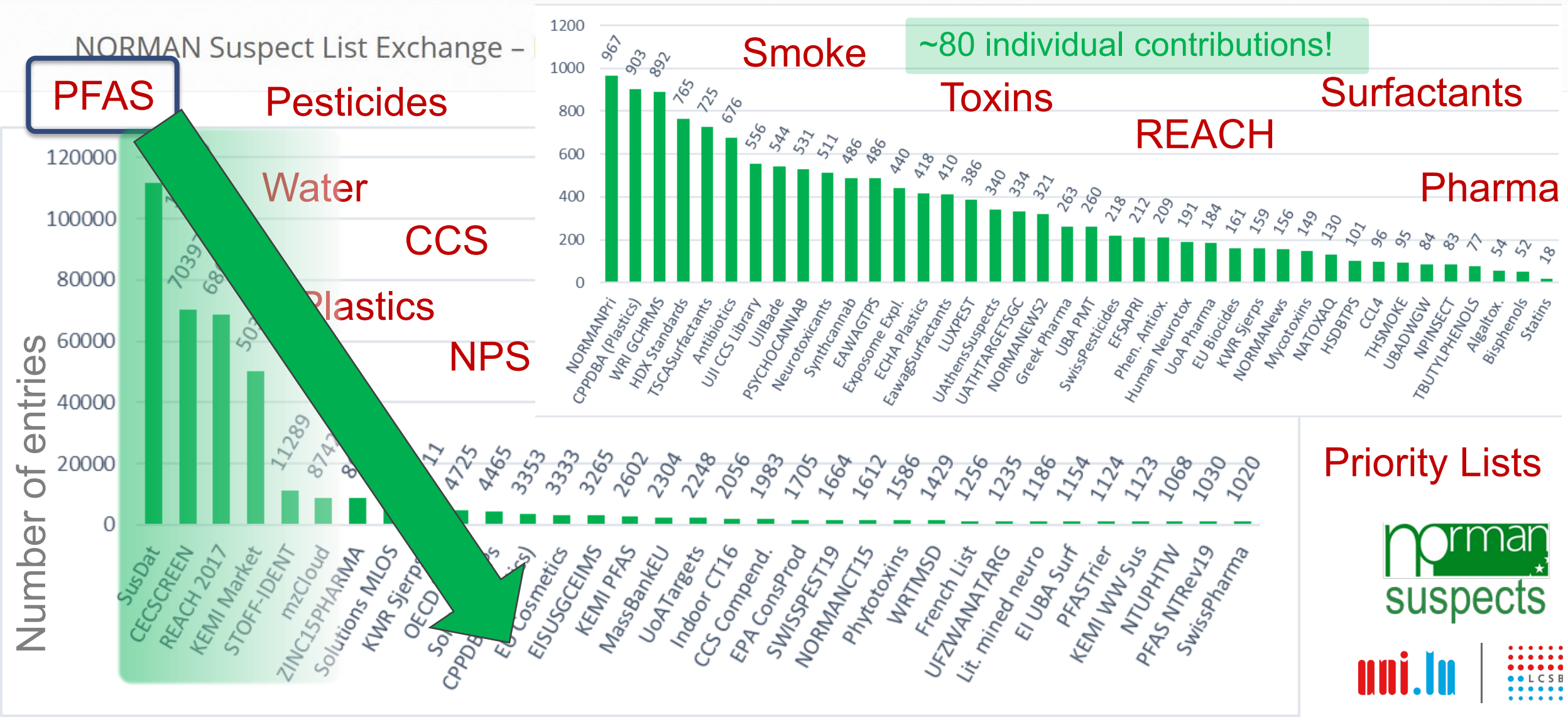
Identification Strategies and Confidence in NT-HRMS(/MS)



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

Growth of Suspect Lists: *e.g.* NORMAN Suspect List Exchange

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



“Just” how many PFAS lists?

EPA United States Environmental Protection Agency | Home | Advanced Search | Batch Search | Lists | Predictions | Downlo

Download | Columns | 10

List Acronym	List Name	Last Updated	Number of Chemicals	List Descripti
EPAPFAS75S1	PFAS|EPA: List of 75 Test Samples (Set 1)	2018-06-29	74	PFAS list corre researchers in
EPAPFAS75S2	PFAS|EPA: List of 75 Test Samples (Set 2)	2019-02-21	75	PFAS list corre by EPA resear
EPAPFASCAT	PFAS|EPA Structure-based Categories	2020-06-02	112	List of registe
EPAPFASDW	PFAS|EPA: New EPA Method Drinking Water	2019-11-16	26	
EPAPFASDW537	PFAS|EPA|WATER: Existing EPA DW Method 537.1	2019-11-16	19	EPA has
EPAPFASDWTREAT	PFAS|EPA|WATER: Drinking Water Treatment Technology	2019-11-16	9	EPA is gatheri drinking wate
EPAPFASINSOL	PFAS|EPA: Chemical Inventory Insoluble in DMSO	2018-06-29	43	PFAS chemica DMSO above
EPAPFASINV	PFAS|EPA: ToxCast Chemical Inventory	2018-06-29	430	PFAS chemica
EPAPFASINVIVO	PFAS|EPA: In Vivo Studies Available	2019-11-16	23	These PFAS ha
EPAPFASLITSEARCH	PFAS|EPA: Literature Search Completed:	2019-11-16	23	A literature re

https://pubchem.ncbi.nlm.nih.gov/classification/#hid=105

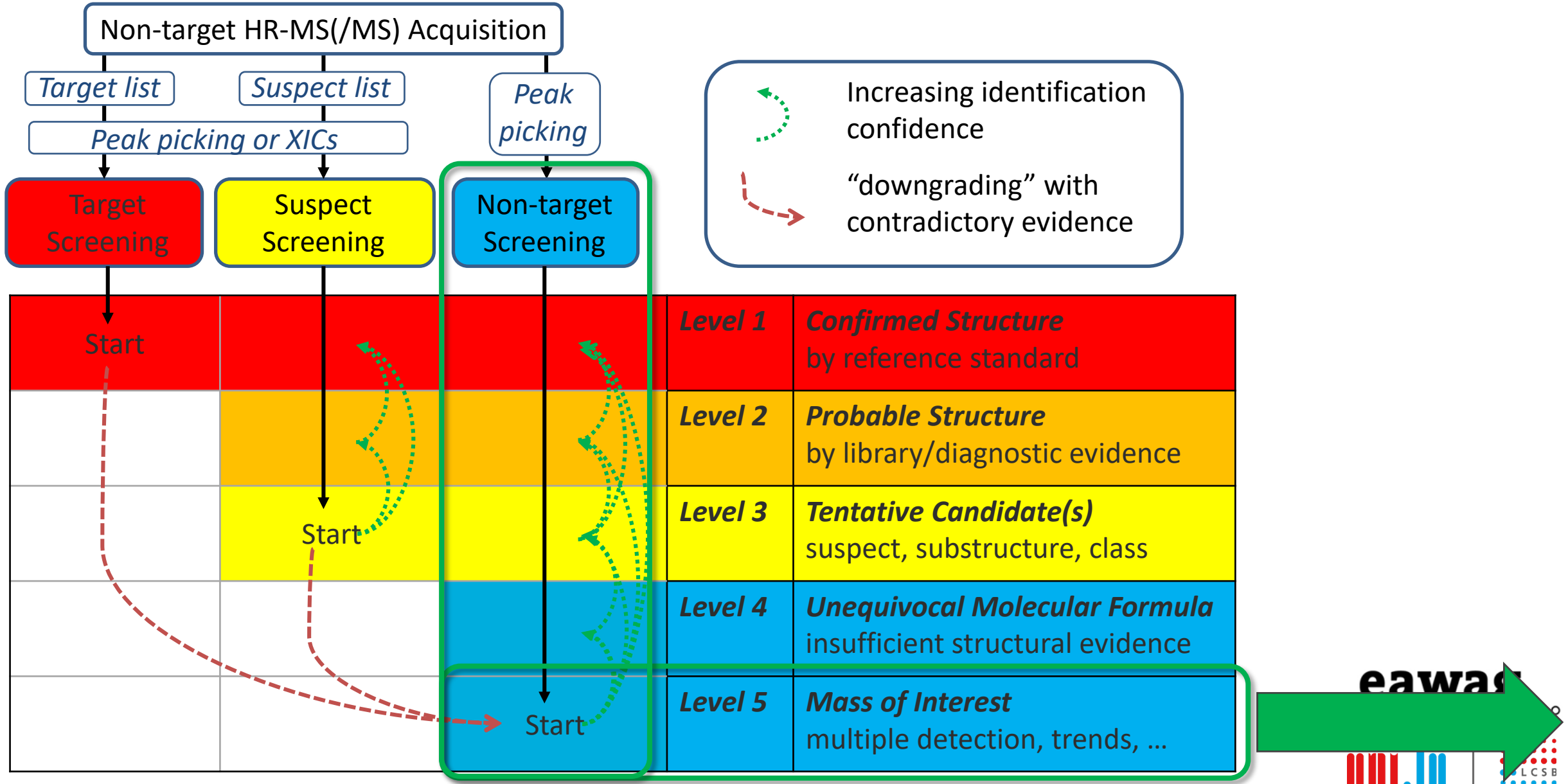
PFAS 8,334

- [EPAPFAS75S1] PFAS|EPA: List of 75 Test Samples (Set 1) 74
- [EPAPFAS75S2] PFAS|EPA: List of 75 Test Samples (Set 2) 75
- [EPAPFASCAT] PFAS|EPA Structure-based Categories 82
- [EPAPFASDW537] PFAS|EPA|WATER: Existing EPA DW Method 537.1 19
- [EPAPFASDW] PFAS|EPA: New EPA Method Drinking Water 26
- [EPAPFASDWTREAT] PFAS|EPA|WATER: Drinking Water Treatment Technology 9
- [EPAPFASINSOL] PFAS|EPA: Chemical Inventory Insoluble in DMSO 43
- [EPAPFASINV] PFAS|EPA: ToxCast Chemical Inventory 428
- [EPAPFASINVIVO] PFAS|EPA: In Vivo Studies Available 23
- [EPAPFASLITSEARCH] PFAS|EPA: Literature Search Completed: 23
- [EPAPFASNONDW] PFAS|EPA: New EPA Method Non-Drinking Water 24
- [EPAPFASRESEARCH] PFAS|EPA: EPA PFAS Research List 165
- [EPAPFASSP] PFAS|EPA: Cross-Agency Research List 193
- [EPAPFASTOX] PFAS|EPA: Toxicity Assessments 9
- [EPAPFASVALWATER] PFAS|EPA|WATER: PFAS with Validated EPA Drinking Water Methods 31
- [EPASDEVTOOLS] PFAS|EPA PFAS chemicals without explicit structures 31
- [EPASINVITRO] PFAS|EPA: List of chemicals tested in in vitro methods 2019-2020 182
- [PFASKEMI] PFAS: List from the Swedish Chemicals Agency (KEMI) Report 1,472
- [PFASLCMSGCMS] PFAS: Collection of GC-MS and LC-MS standards: Food Contact Materials 38
- [PFASMASTER] PFAS Master List of PFAS Substances (Version 2) 8,159
- [PFASNORDIC] PFAS: Nordic PFAS Report 2019 203
- [PFASNTREV19] PFAS: PFAS in Non-Target HRMS Studies (Liu et al 2019) 127
- [PFASOECD] PFAS: Listed in OECD Global Database 3,702
- [PFASOECDNA] NORMAN: List of PFAS from the OECD Curated by Nikiforos Alygizakis 3,206
- [PFASSTRUCT] PFAS|EPA: PFAS structures in DSSTox (update August 2020) 8,137
- [PFASTRI] PFAS: PFAS to the Toxics Release Inventory (TRI) Program by the National Defense Authorization Act 95
- [PFASTRIER] PFAS Community-Compiled List (Trier et al., 2015) 589
- [TONYPFASDASH] TONY'S Mockup of a PFAS Dashboard List 11

28 PFAS “Suspect Lists”… Where to start?

https://comptox.epa.gov/dashboard/chemical_lists/?search=PFAS
https://pubchem.ncbi.nlm.nih.gov/classification/#hid=105

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



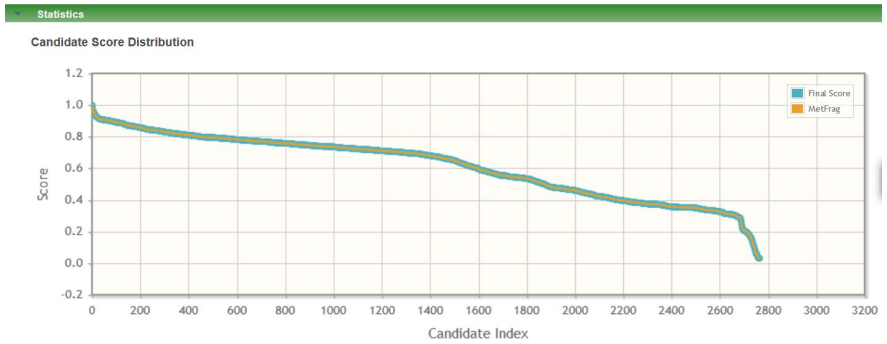
Schymanski *et al*, 2014, ES&T. DOI: 10.1021/es5002105 & Schymanski *et al*. 2015, ABC, DOI: 10.1007/s00216-015-8681-7

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

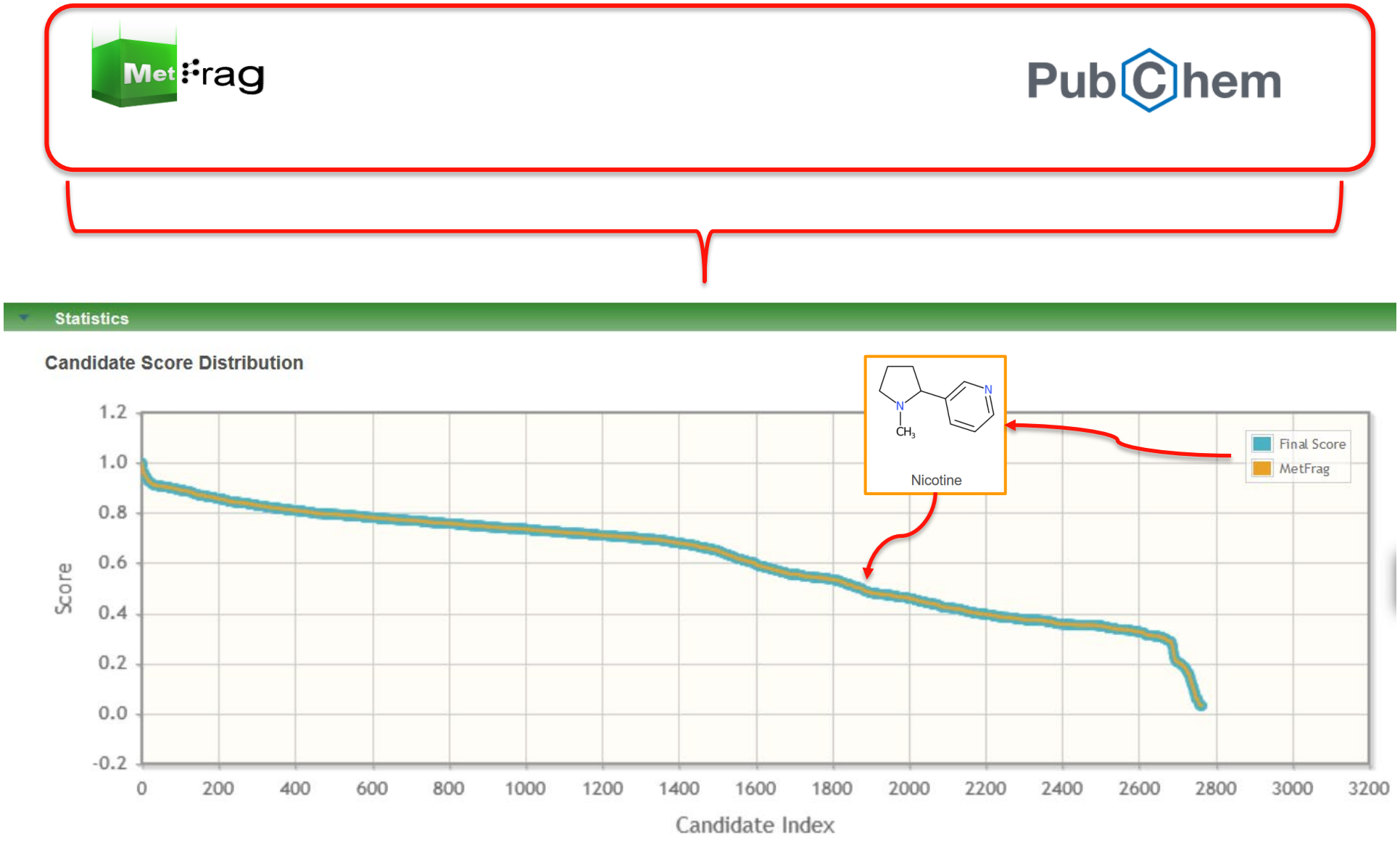
134.0054	339689
150.0001	77271
213.9607	632466

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!



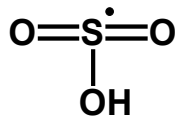
MetFrag + PubChem + Formula Search + https://massbank.eu/MassBank/RecordDisplay.jsp?id=EQ300804&dsn=Eawag

Status Quo in 2016: MetFrag Relaunched …

mz [M-H]⁻
213.9637
± 5 ppm

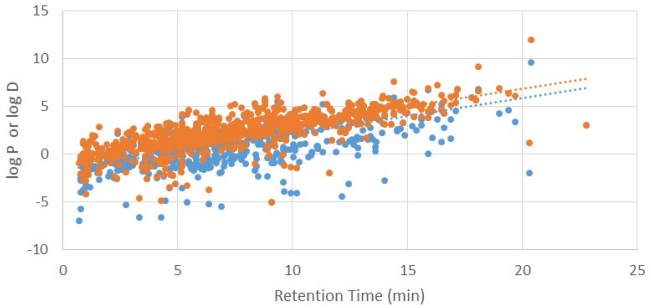
Elements: C,N,S

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

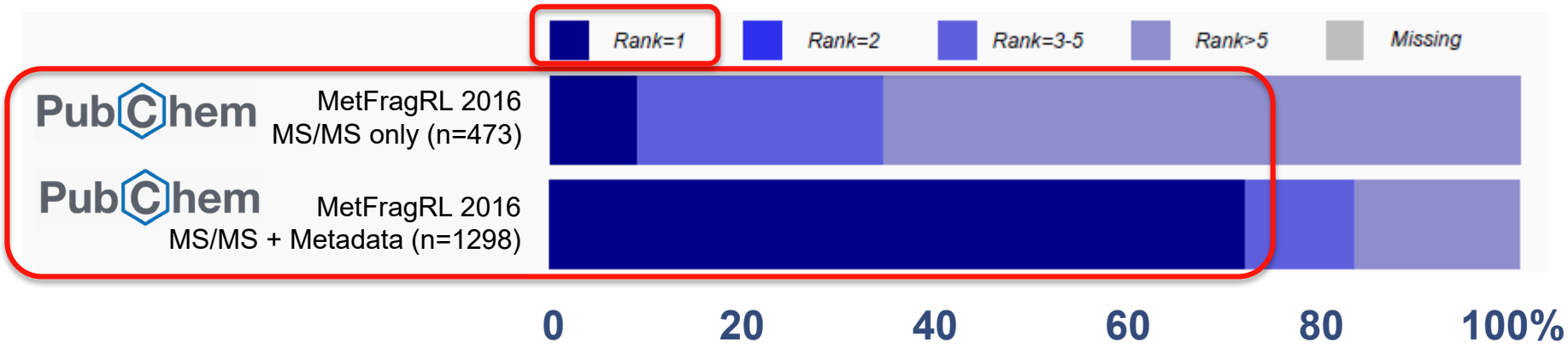
solutions

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!



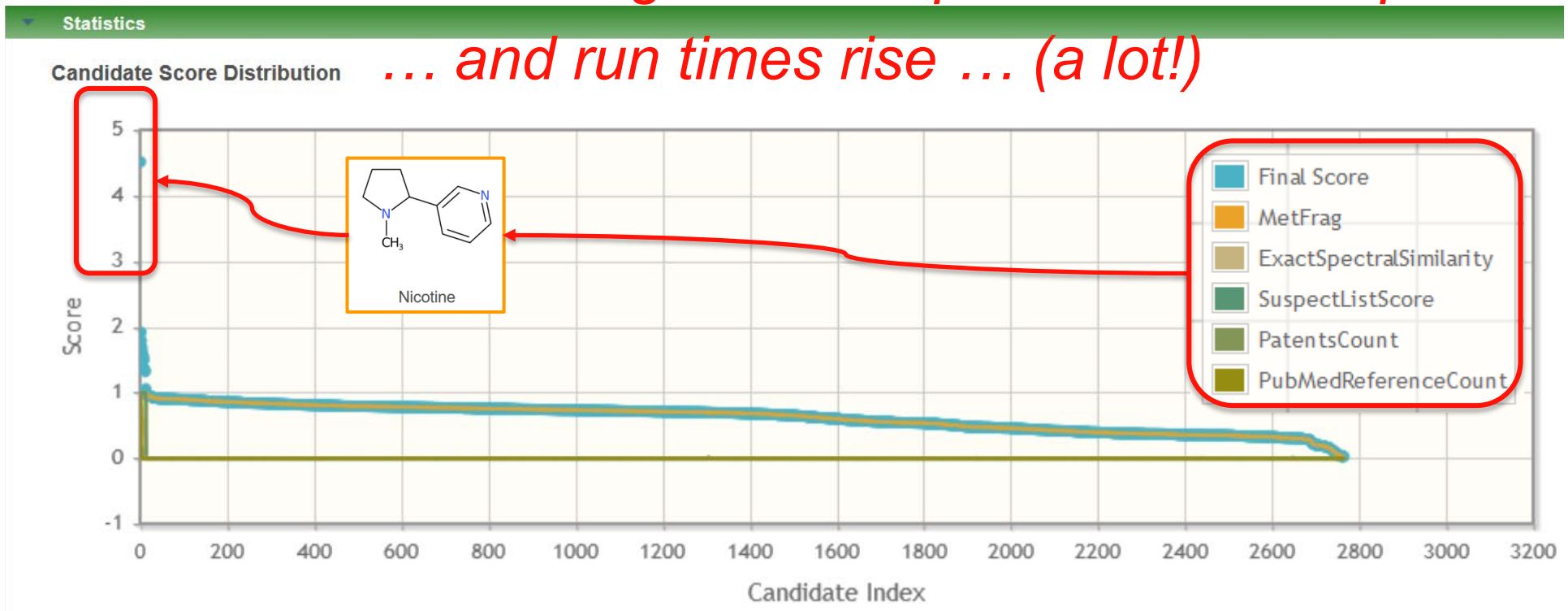
Schymanski, Kondic, Neumann, Thiessen, Zhang & Bolton (2021) *J. Cheminf.*, DOI: 10.1186/s13321-021-00489-0

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

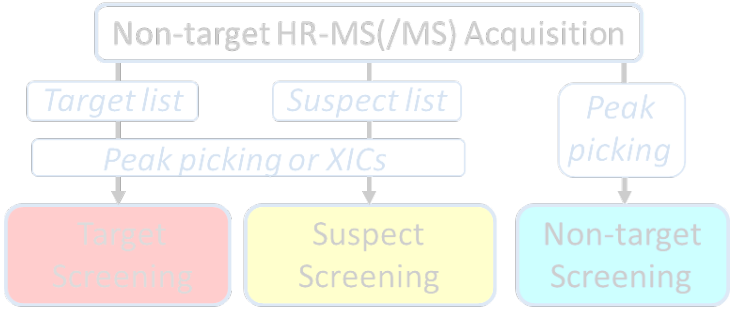
… and run times rise … (a lot!)



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

Outline for today

- Problem formulation
  - Why is non-target HR-MS still so hard?



- Redefining our strategy – *i.e.* DIY Databases
  - Take the "good bits" of databases, and ignore the rubble
  - … in a way that keeps many players happy…
  - … and updates as knowledge updates …

- What's next?

PubChemLite
EXPOSOMICS

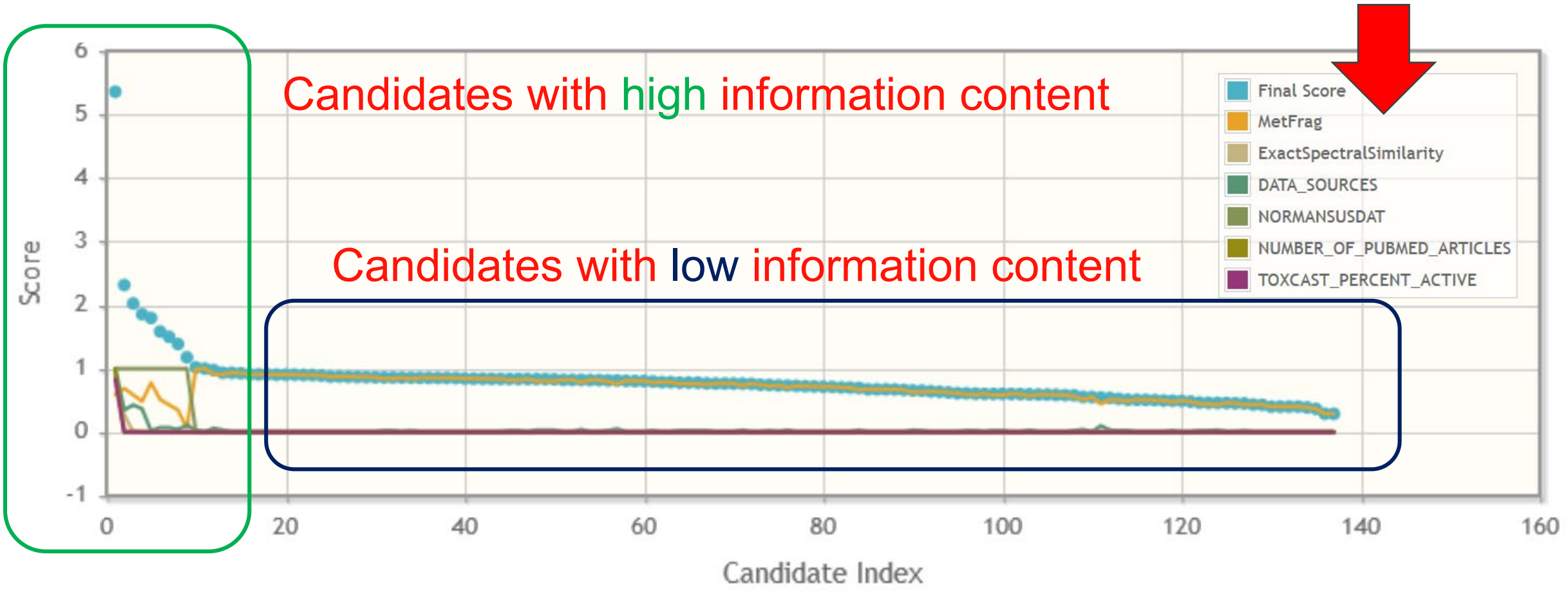
Problem: Exposomics “Chemical Space” is too big!

CAS A DIVISION OF THE AMERICAN CHEMICAL SOCIETY — 180 million

PubChem — 110 million

ChemSpider Search and share chemistry — 103 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

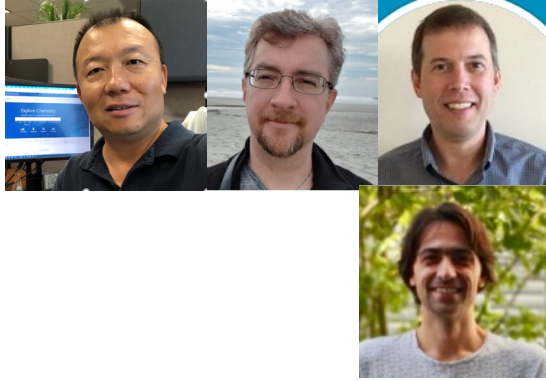
Active Substance	furathiocarb
Status	Not Approved [Reg. (EC) No 1107/2009]
Categories	Insecticides

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”

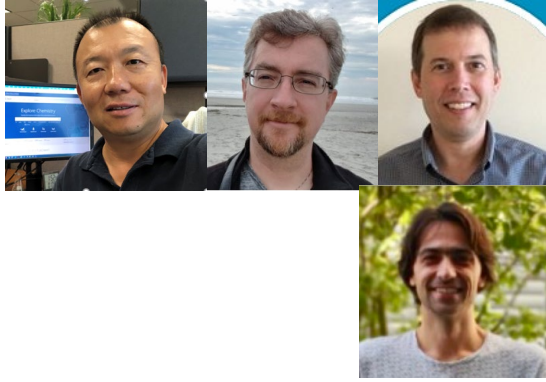
Schymanski *et al*. (2021) DOI: 10.1186/s13321-016-0115-9

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
ontents page at the PubChem Classification Browser
ih.gov/classification/#hid=72). With this release, there is now just one
is the former tier1 plus two new categories (Associated Disorders &

is 371,663 compounds (31 Oct 2020) compiled from 10 categories:
DrugMedicInfo, FoodRelated, PharmacoInf Info,
, Identification.

llapsed by InchIKey first block, reporting th most
CIDs. Entries that will be ignored by MetFra
(e.g. transition metals) have been removed Med ID
les on the PubChem FTP site. The "AnnoTy how
epresented, the subsequent column (named per category) counts the
ories available in the next sub-category of the TOC entry.

Edit | New version

Communities
LCSB Environmental Cheminformatics Group | Remove

1,084 views | 1,170 downloads

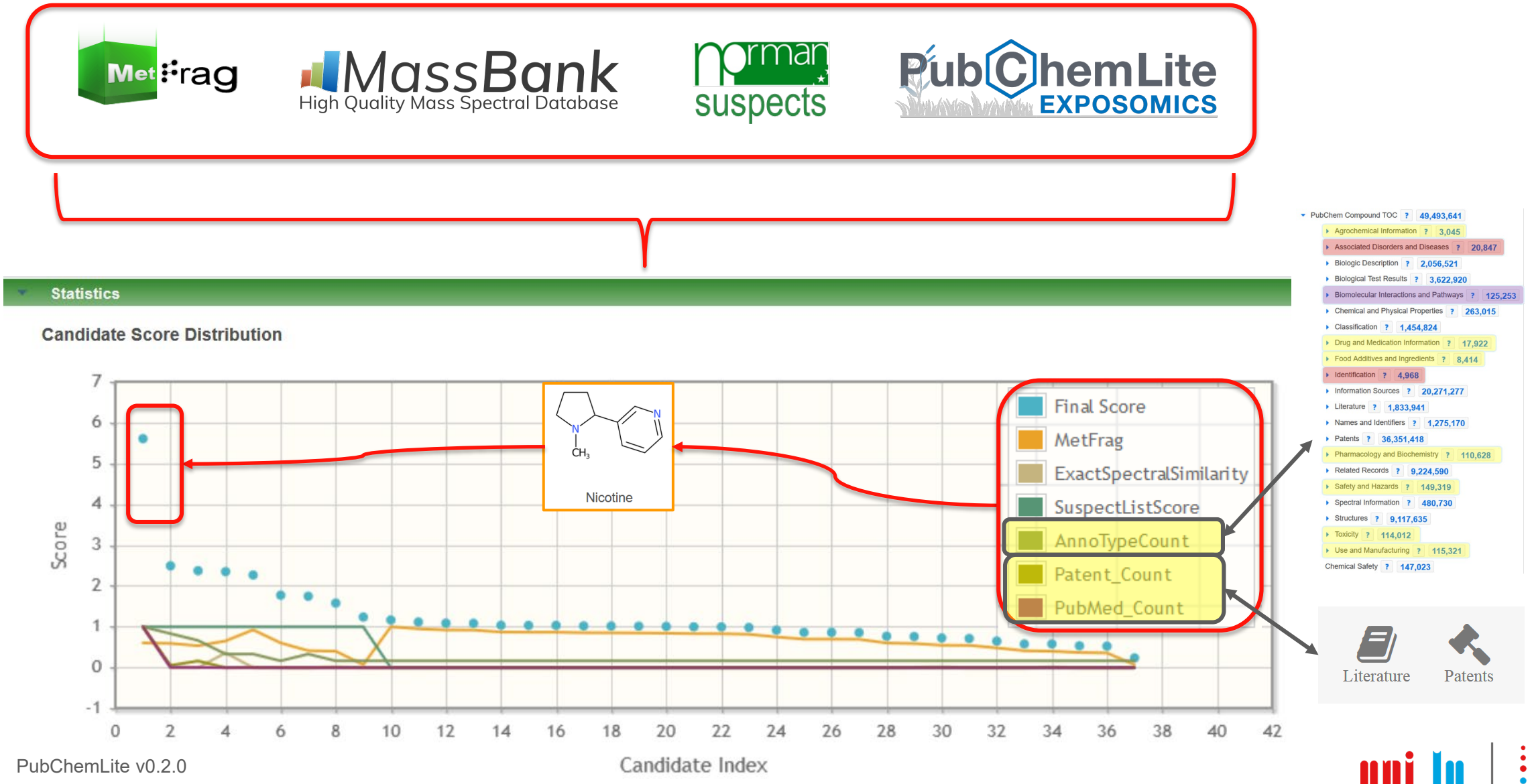
See more details...

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

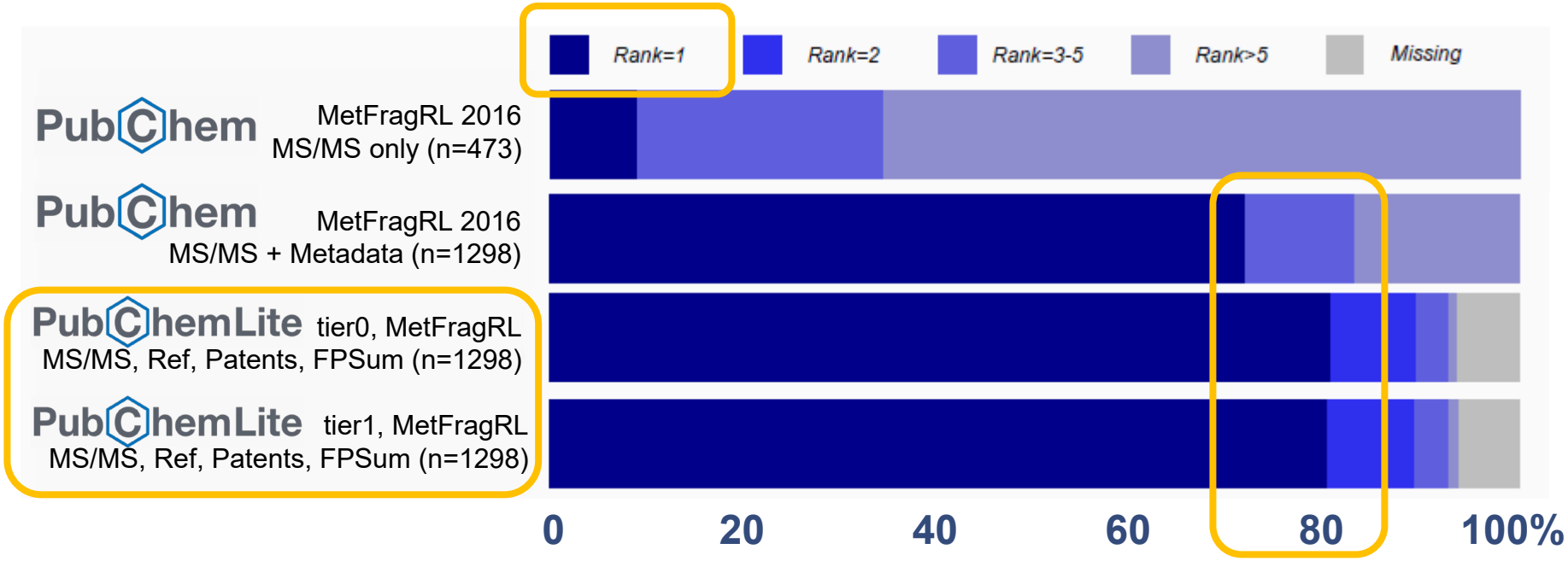


PubChemLite v0.2.0

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?

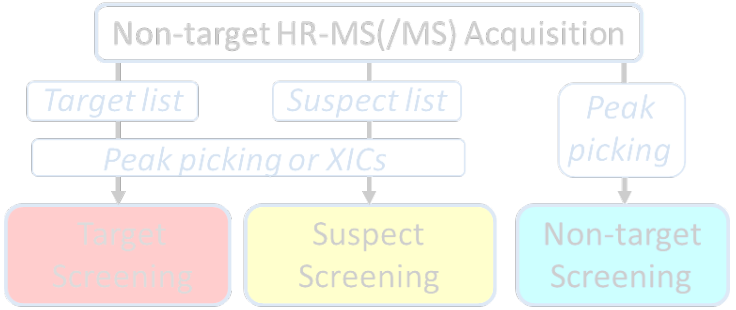


Schymanski, Kondic, Neumann, Thiessen, Zhang & Bolton (2021) *J. Cheminf.*, DOI: 10.1186/s13321-021-00489-0

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

Outline for today

- Problem formulation
  - Why is non-target HR-MS still so hard?

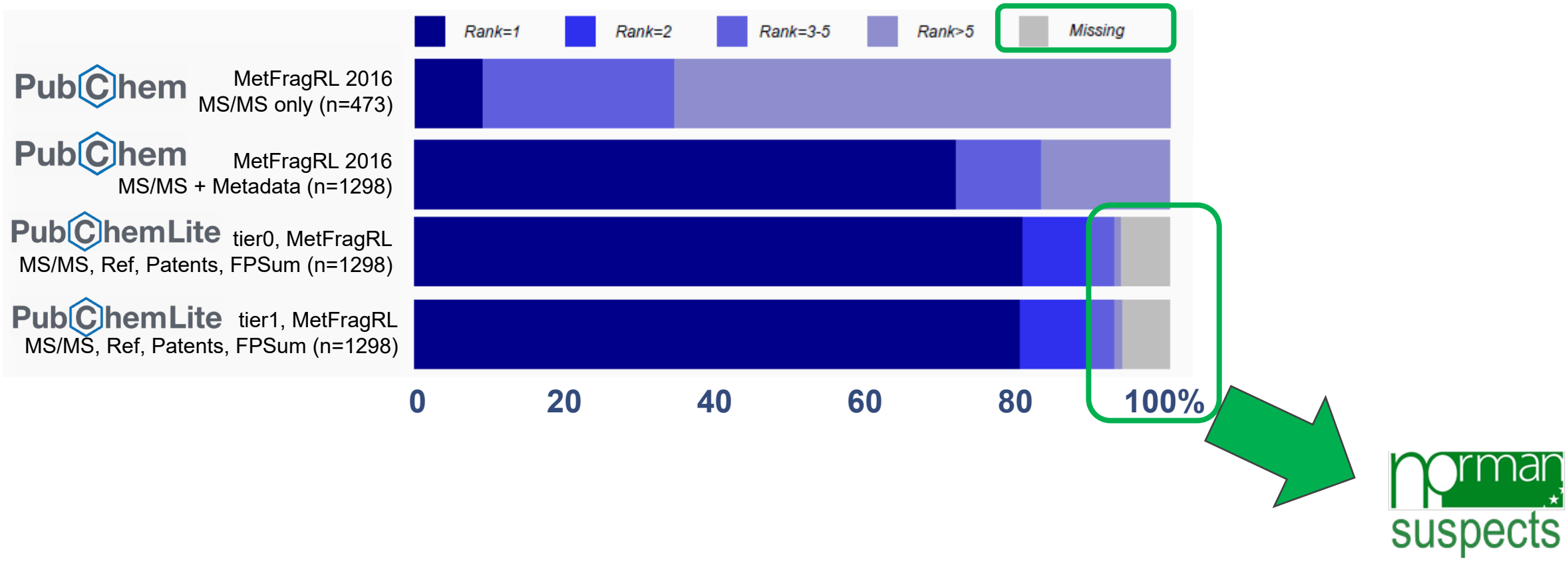


- Redefining our strategy – *i.e.* DIY Databases
  - Take the "good bits" of databases, and ignore the rubble
  - … in a way that keeps many players happy…
  - … and updates as knowledge updates …

- What's next?

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

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

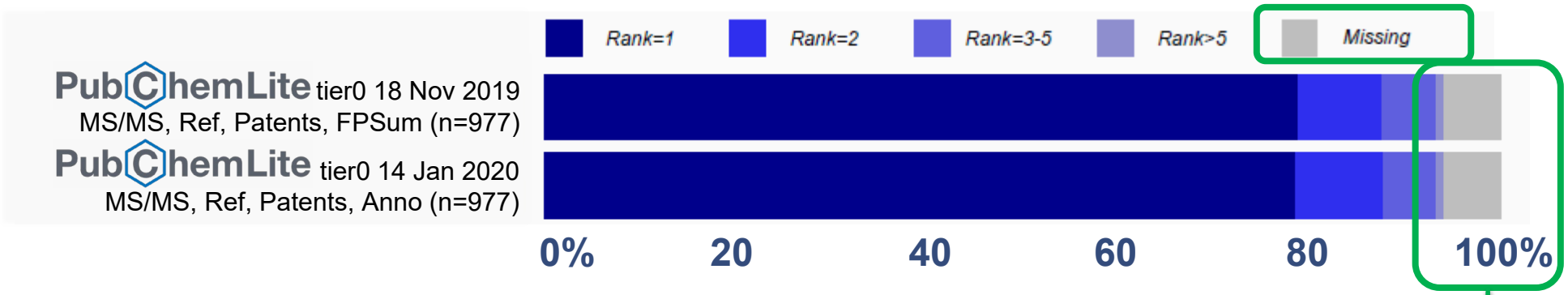
Search selected classification by Keyword | Enter desired search term | Search

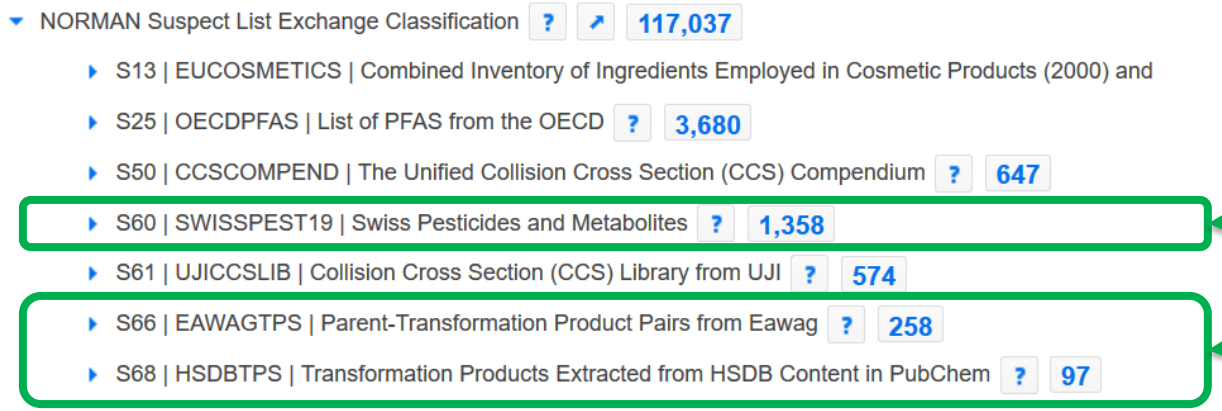
Classification description (from NORMAN Suspect List Exchange)

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

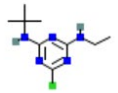
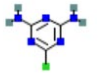
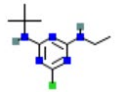
- Pharmacology and Biochemistry ? 112,039
  - Human Metabolite Information ? 64,199
  - Metabolism/Metabolites ? 8,204
  - Transformations ? 5,857

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

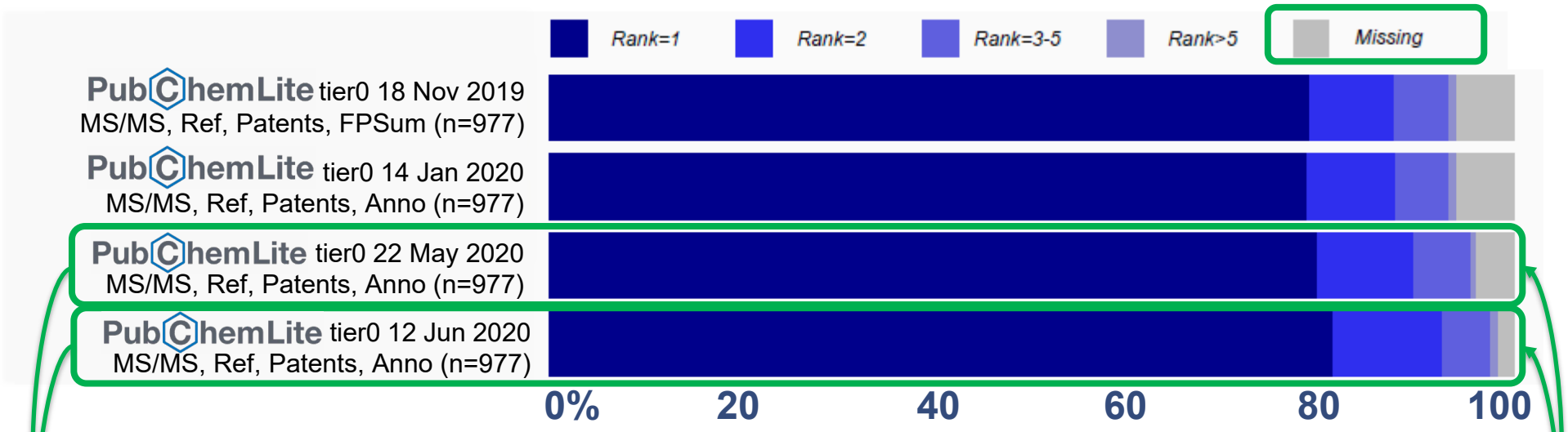
- Agrochemical Information ? 3,045
  - Agrochemical Category ? 1,903
  - Agrochemical Transformations ? 1,383
  - EU Pesticides Data ? 1,213
  - USDA Pesticide Data Program ? 579

PubChemLite EXPOSOMICS

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

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

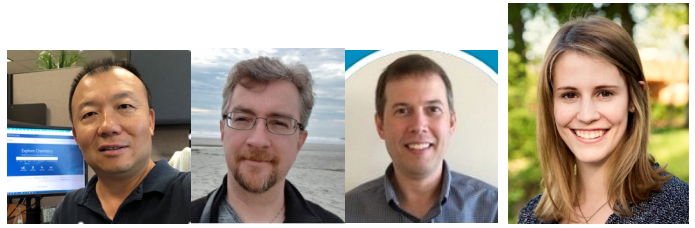
Assessing the Missing Entries in PubChemLite



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

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

“Circle of Data”: Literature Mining for Metabolites / TPs



PubChem Terbuthylazine (Compound)

8.3 Metabolism/Metabolites

Metabolism of terbuthylazine in rats is similar to other chloro-s-triazine herbicides. The major routes of metabolism are hydrolysis of the chlorine moiety and mono- or didealkylation. Hydroxylation of one or both of the dealkylated amine groups may also occur.

USEPA; Reregistration Eligibility Decision (RED) Database for Terbuthylazine (5915-41-3). EPA 738-R-95-005 p.12 (March 1995). Available from, as of October 11, 2012: http://www.epa.gov/pesticides/reregistration/status.htm

▸ Hazardous Substances Data Bank (HSDB)

Urine and feces contained up to 25 and 15 identified metabolites, respectively, most of which were polar. Degradation of the triazine ring did not occur. Ammeline and ammelide, 2 dechlorinated and dealkylated/hydroxylated metabolites common to all triazines, were identified in low amounts in the feces.

USEPA; Reregistration Eligibility Decision (RED) Database for Terbuthylazine (5915-41-3). EPA 738-R-95-005 p.13 (March 1995). Available from, as of October 11, 2012: http://www.epa.gov/pesticides/reregistration/status.htm

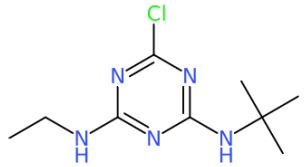
▸ Hazardous Substances Data Bank (HSDB)

In mammals, following oral administration, ...a de-ethyl metabolite forms rapidly, followed by conjugates of products formed by oxidation of one methyl group of the tert-butyl moiety. All are rapidly excreted.

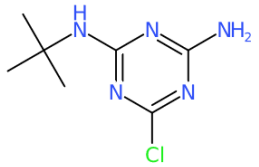
Tomlin CDS, ed. Terbuthylazine (5915-41-3). In: The e-Pesticide Manual, Version 2.2 (2002). Surrey UK, British Crop Protection Council.

▸ Hazardous Substances Data Bank (HSDB)

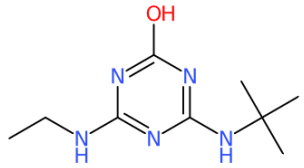
CDK DEPICT



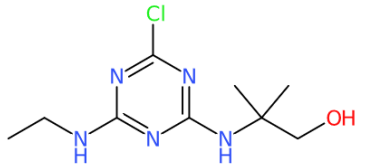
Terbutylazine CID:22206



desethyl-terbutylazine CID:108201



2-hydroxy-terbutylazine CID:135495928



(hydroxy-t-butyl)-Terbutylazine CID:779516

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

“Living data connections”

zenodo Search Upload Communities

June 11, 2020 Dataset Open A

S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem

LCSB-ECI; Krier, Jessy; Schymanski, Emma; PubChem Team; Bolton, Evan; Thiessen, Paul; Zhang, Jeff

This is the collection associated with list S68 HSDBTPS Transformation Products Extracted from HSDB Content in PubChem on the NORMAN Suspect List Exchange.

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

HSDBTPS is a list of metabolites / transformation products extracted from the "Metabolites/Metabolism" section from HSDB (Hazardous Substance Data Bank) in PubChem (https://pubchem.ncbi.nlm.nih.gov/source/11933). Dataset DOI: 10.5281/zenodo.3827487.

Preview

Predecessor_CID	Predecessor_Name	Successor_CID	Successor_Name	Transf
13450	Terbutryn	110189337	2-[[4-(Ethylamino)-6-methylsulfanyl-1,3,5-triazin-2-yl]amino]-2-methylpropanoic acid	mamma metabc
13450	Terbutryn	110189337	2-[[4-(Ethylamino)-6-methylsulfanyl-1,3,5-triazin-2-yl]amino]-2-methylpropanoic acid	mamma metabc

https://git-r3lab.uni.lu/eci/pubchem/

LCSB-ECI & PubChem Team. DOI 10.5281/zenodo.3890392

File Edit View Repository Branch Help

Current repository **pubchem** | Current branch **master** | **Fetch origin** Last fetched 2 minutes ago

Changes 2 | History

No branches to compare

Update extractAnnotations.R — Emma Schymanski • Jun 9, 2020

HSDB Ref Info — Emma Schymanski • Jun 8, 2020

added new CIDs to HSDBTPS — Emma Schymanski • Jun 8, 2020

Update PCLite_eval_support.R — Emma Schymanski • Jun 8, 2020

Added S69 LUXPEST — Emma Schymanski • May 28, 2020

Update P — Emma

Update u — Emma

Update P

added new CIDs to HSDBTPS

Emma Schymanski 13fdb18 5 changed files Hide Whitespace

Added newly registered CIDs to base HSDB files, HSDBTPS struct info and transformation tables.

annotations\tps\H...13450_selected.csv
annotations\tps\H..._3120_selected.csv
annotations\tps\H...31645_selected.csv
...\S68_HSDBTPS_StructureInfoOnly.csv
annotations\tps\H...formationTable.csv

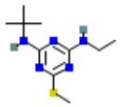
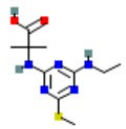
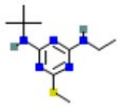
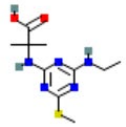
S-demethylation; N-deethylation; and disulfide formation.",13450|13450,13450|135495928|135612794,TRUE,mammalian metabolism,TRUE,"TPs added, rest are not yet in PubChem or too inspecific"

3 +HSDB,1525,TERBUTRYNE,13450,1,13450,"Menzie, C.M. Metabolism of Pesticides-Update III. Special Scientific Report- Wildlife No. 232. Washington, DC: U.S.Department of the Interior, Fish and Wildlife Service, 1980., p. 540","After administration of terbutryne to rats, urinary metabolites observed ... included: 2-hydroxy terbutryne; 2-amino-4-hydroxy-6-t-butylamino-s-triazine; 2-amino-4-t-butylamino-6-mercapto-s-triazine; two S-glucuronides and two t-butyl-O-glucuronides. Other metabolites were formed by one or a combination of the follow

8.5 Transformations

19 items View More Rows & Details Download

SORT BY Please Choose One

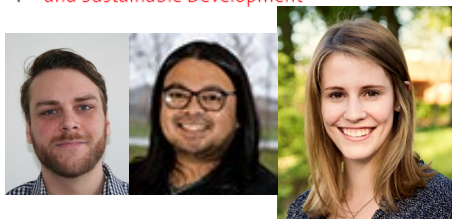
Predecessor Image	Predecessor Name	Transformation	Successor Image	Successor Name	Evidence DOI
	Terbutryn	Mammalian metabolism		2-[[4-(Ethylamino)-6-methylsulfanyl-1,3,5-triazin-2-yl]amino]-2-methylpropanoic acid	10.1002/bms.12000506
	Terbutryn	Mammalian metabolism		2-[[4-(Ethylamino)-6-methylsulfanyl-1,3,5-triazin-2-yl]amino]-2-methylpropanoic acid	10.5281/zenodo.38274

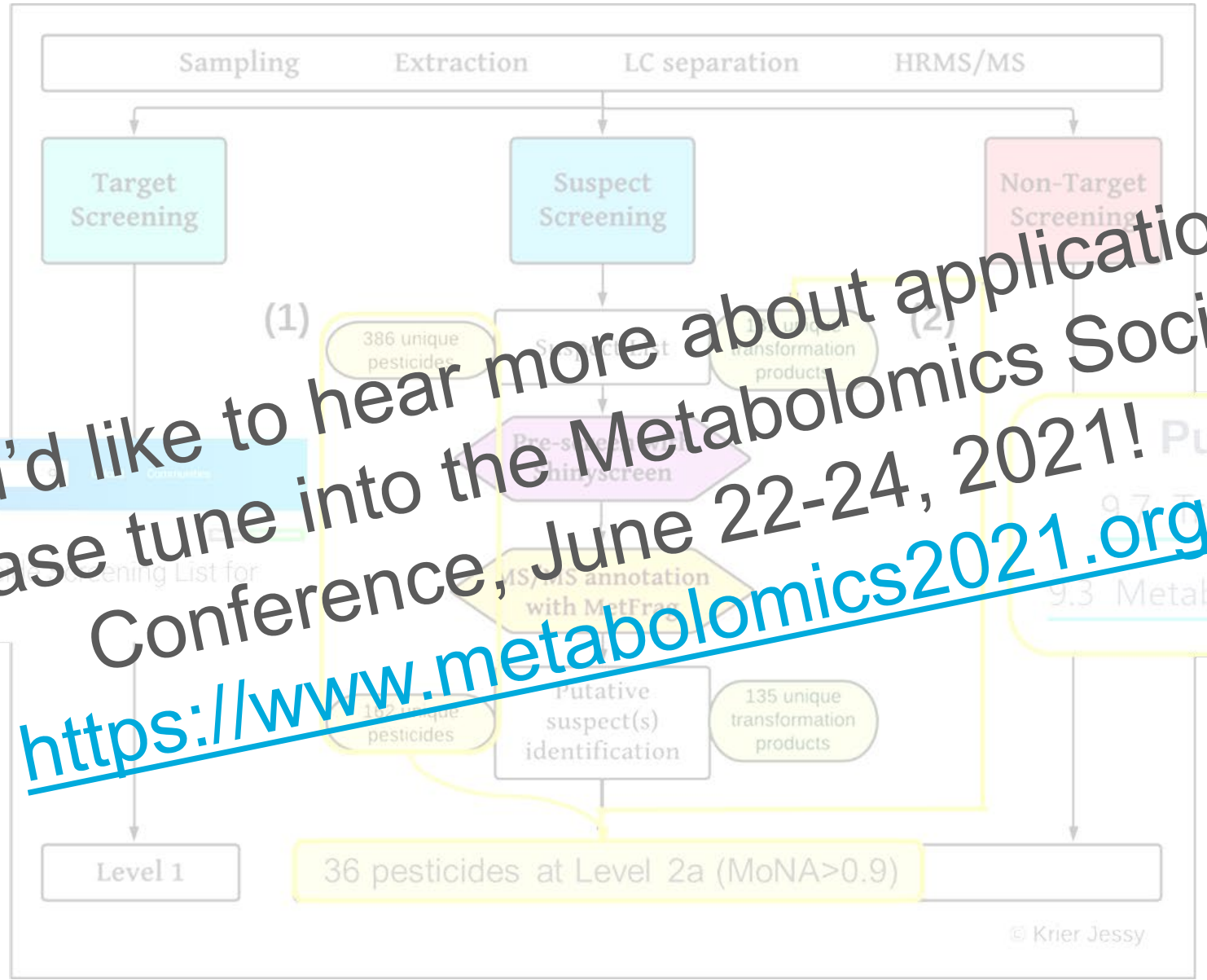
NIH U.S. National Library of Medicine National Center for Biotechnology Information

uni.lu | LCSB

Examples: Pesticides & Metabolites in Luxembourg

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



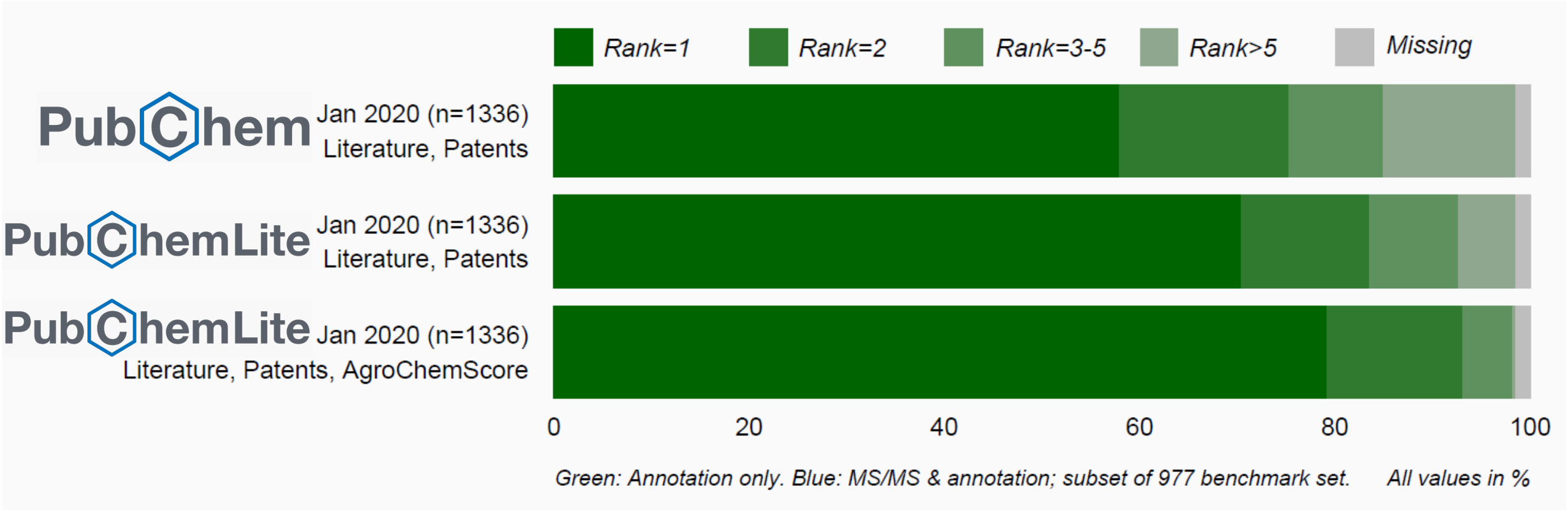


If you'd like to hear more about applications, please tune into the Metabolomics Society Conference, June 22-24, 2021!

https://www.metabolomics2021.org/

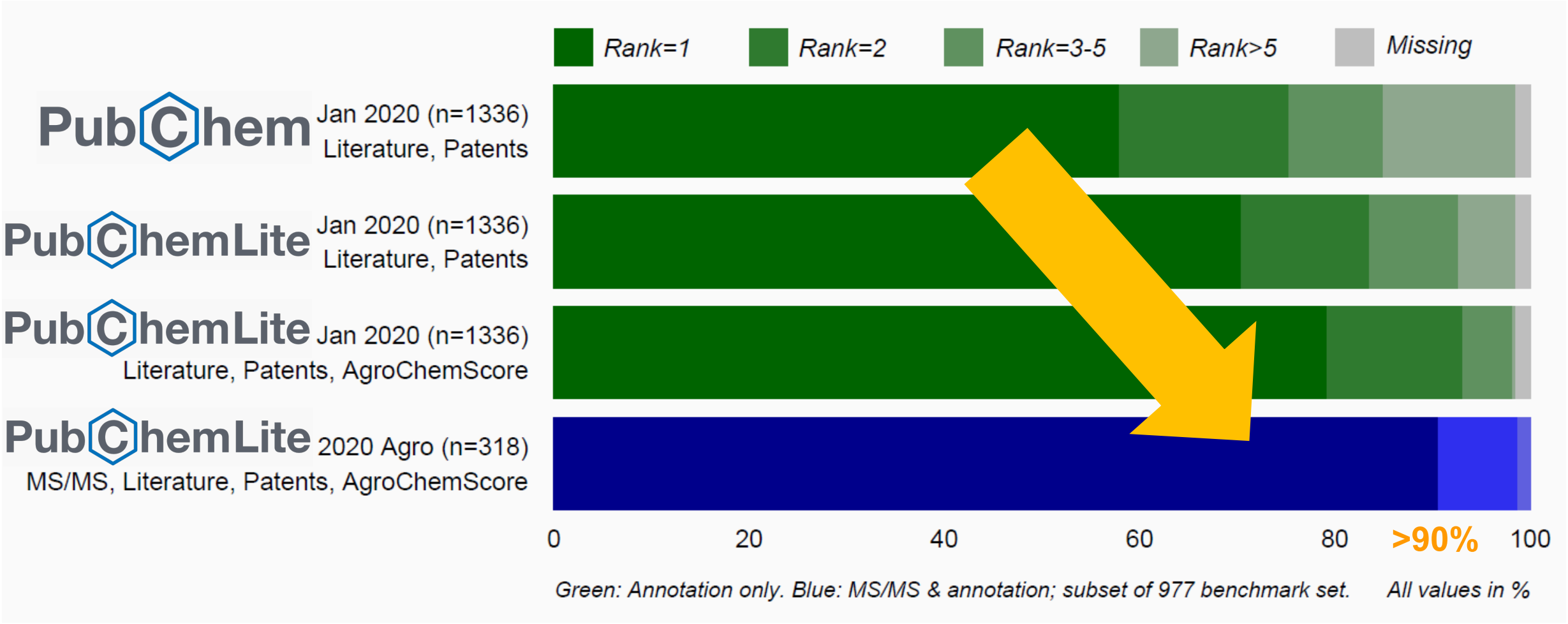
Jessy Krier (2020) Master's Thesis, defended July 2020. Figure 8 & Jessy Krier (2020) S69 | LUXPEST. DOI: 10.5281/zenodo.3862689

Influence of the Annotation Content in PubChemLite



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

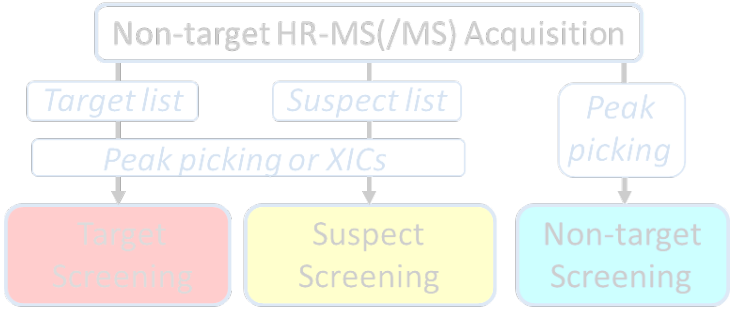
Influence of the Annotation Content in PubChemLite



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

Outline for today

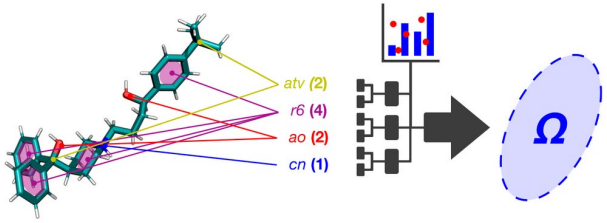
- Problem formulation
  - Why is non-target HR-MS still so hard?



- Redefining our strategy – *i.e.* DIY Databases
  - Take the "good bits" of databases, and ignore the rubble
  - … in a way that keeps many players happy…

  - … and updates as knowledge updates …

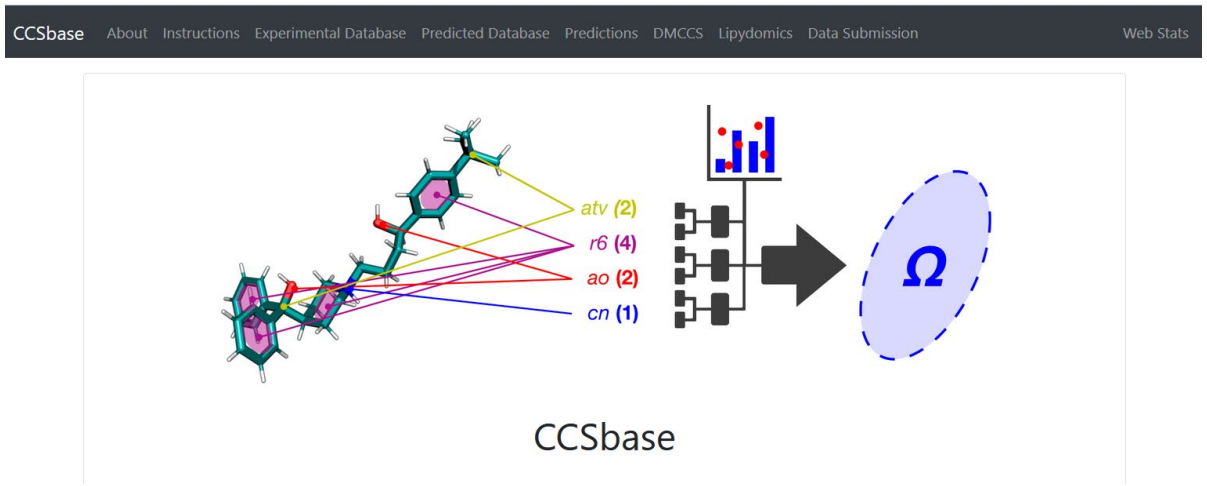
- What's next?

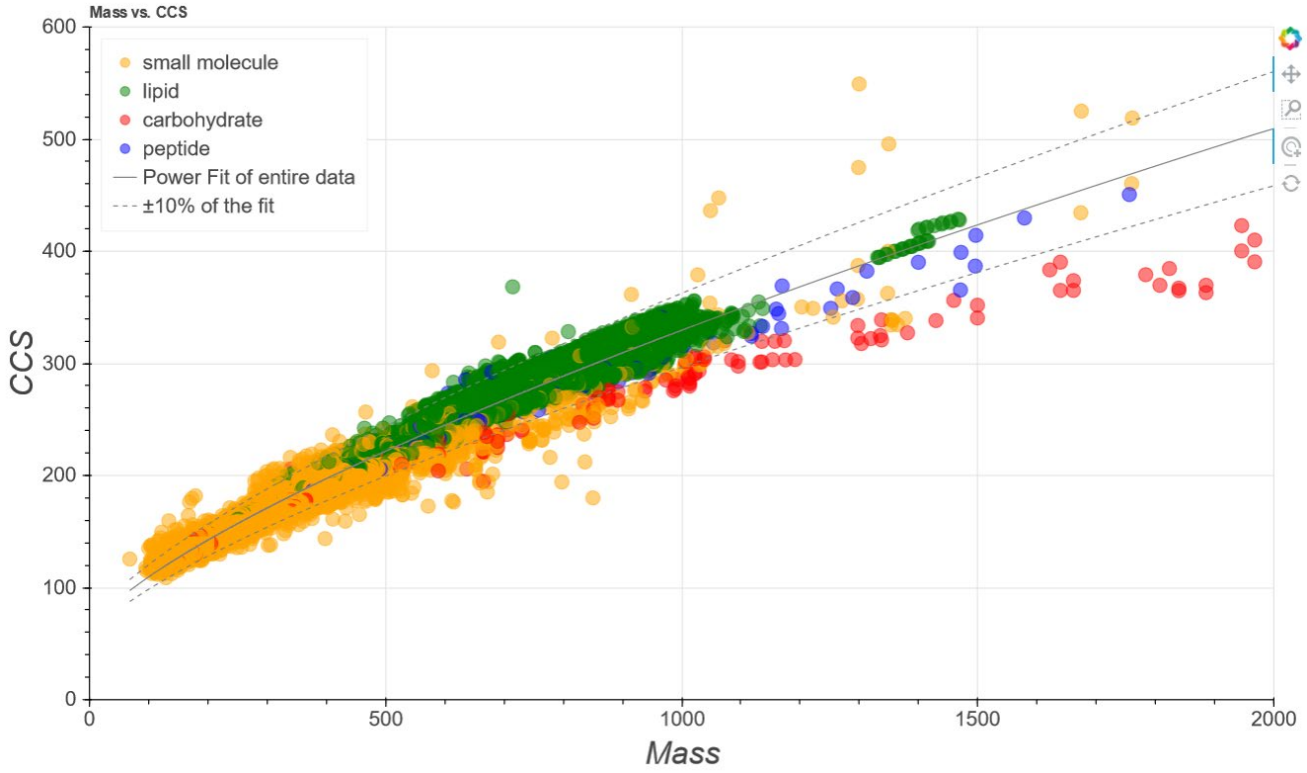


CCSbase

PubChemLite EXPOSOMICS

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

295 views 239 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 Classification Browser

Browse PubChem data using a classification of interest, or search for PubChem records annotated
phenylpropionates, or Gene Ontology: DNA repair). More...

Select classification

PubChem: Aggregated CCS Classification

Search selected classification by

Keyword | Enter

Classification description (from PubChem)

This classification aggregates experimental CCS data available in PubChem, from CCSbase (https://ccsbase.ne
More...

Data type counts to display None | Compound

Display zero count nodes? Yes | No

Browse PubChem: Aggregated CCS Classification Tree

- Aggregated CCS Classification ? 5,434
  - Aggregated CCS Information ? 5,434
  - CCSbase ? 4,911
  - NORMAN-SLE: S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium ? 630
  - NORMAN-SLE: S61 | UJICCSLIB | Collision Cross Section (CCS) Library from UJI ? 574

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

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

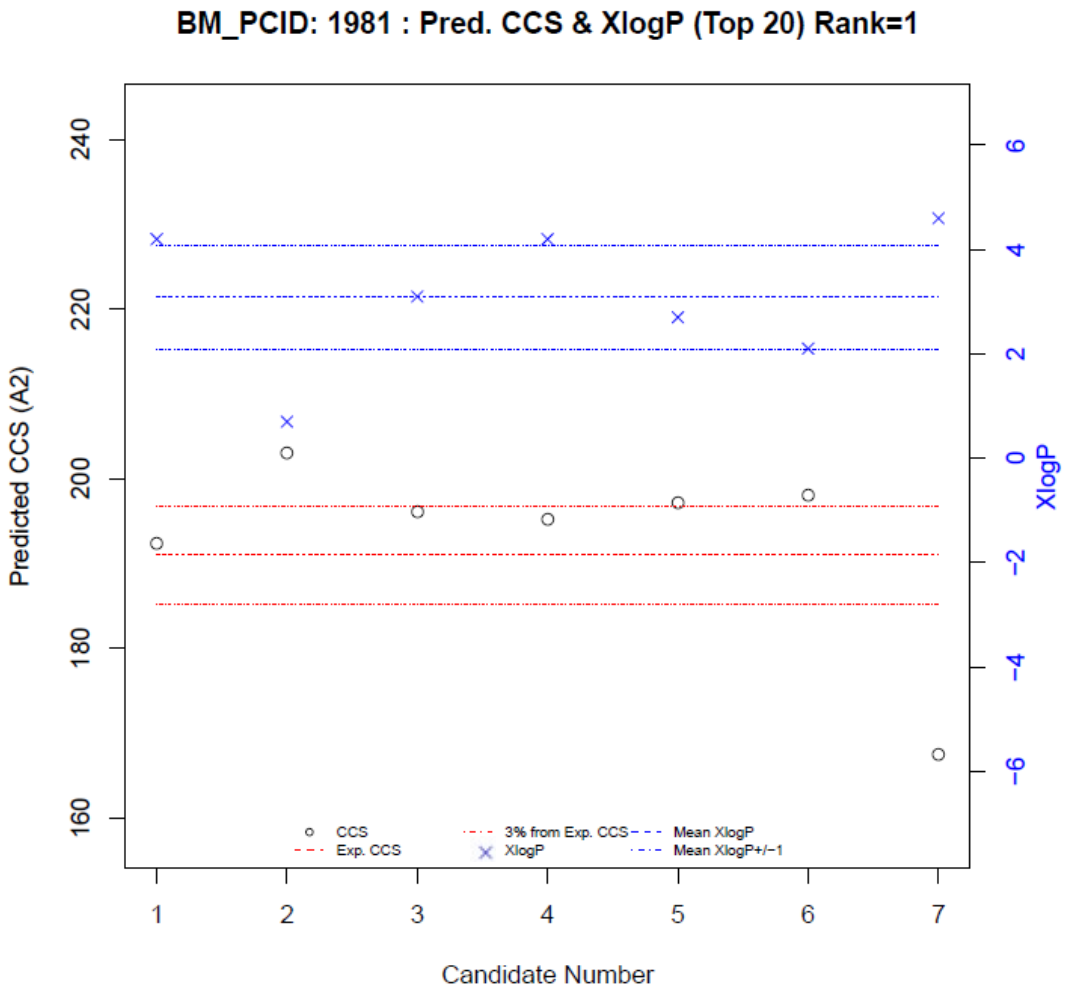
Ross, D. H., Cho, J. H. & Xu, L. Anal. Chem. (2020). DOI: 10.1021/acs.analchem.9b05772.

norman suspects

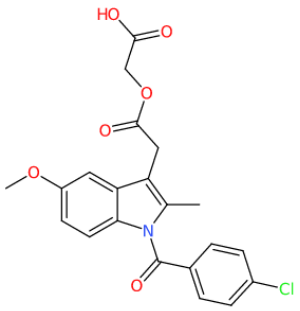
UNIVERSITY of WASHINGTON

LCSB

PubChemLite EXPOSOMICS … plus CCSbase!



PubChem Acemetacin (Compound)

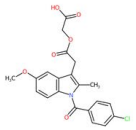
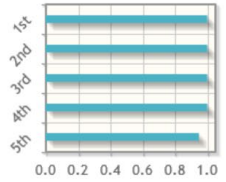
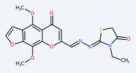
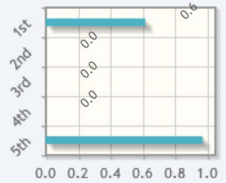


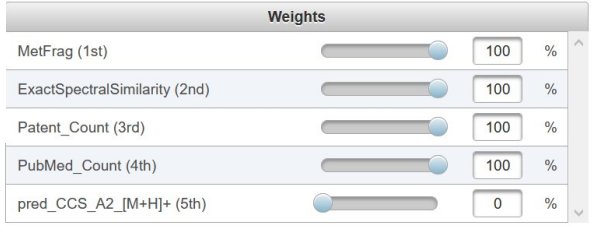
3.2.6 Collision Cross Section

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	$C_{21}H_{18}ClNO_6$		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

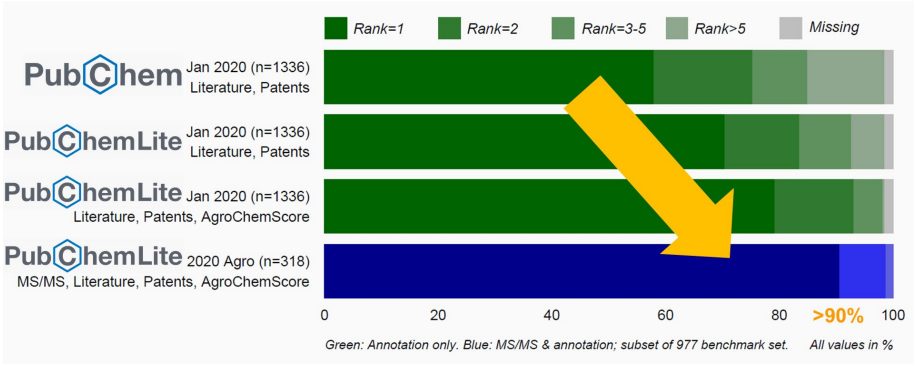


Schymanski *et al*. (2021) JCheminform. DOI: 10.1186/s13321-016-0115-9

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

“Take home” Messages

- Time will tell if PubChemLite EXPOSOMICS is useful
  - Users and feedback welcome!



- Annotation content is extremely powerful
  - Any category in the PubChem TOC can be added
  - (and new versions can be created)

filter_bits.lst 209 Bytes

```
281   Agrochemical Information
120   Drug and Medication Information
299   Food Additives and Ingredients
462   Pharmacology and Biochemistry
479   Safety and Hazards
534   Toxicity
470   Use and Manufacturing
```

- Coupling of experimental evidence is underway…

- … hopefully making NTS & HR-MS/MS & exposomics a little easier!

Thank you!

PubChemLite EXPOSOMICS

emma.schymanski@uni.lu and @Eschymanski

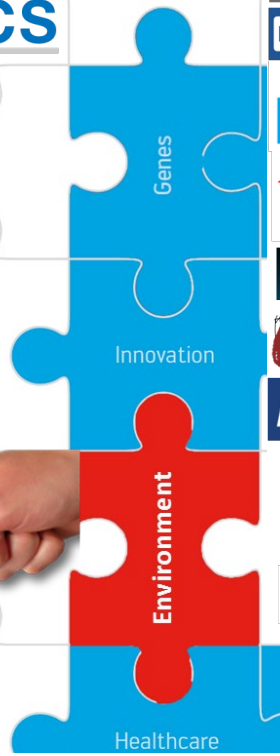
PCL+CCS: 10.5281/zenodo.4456208

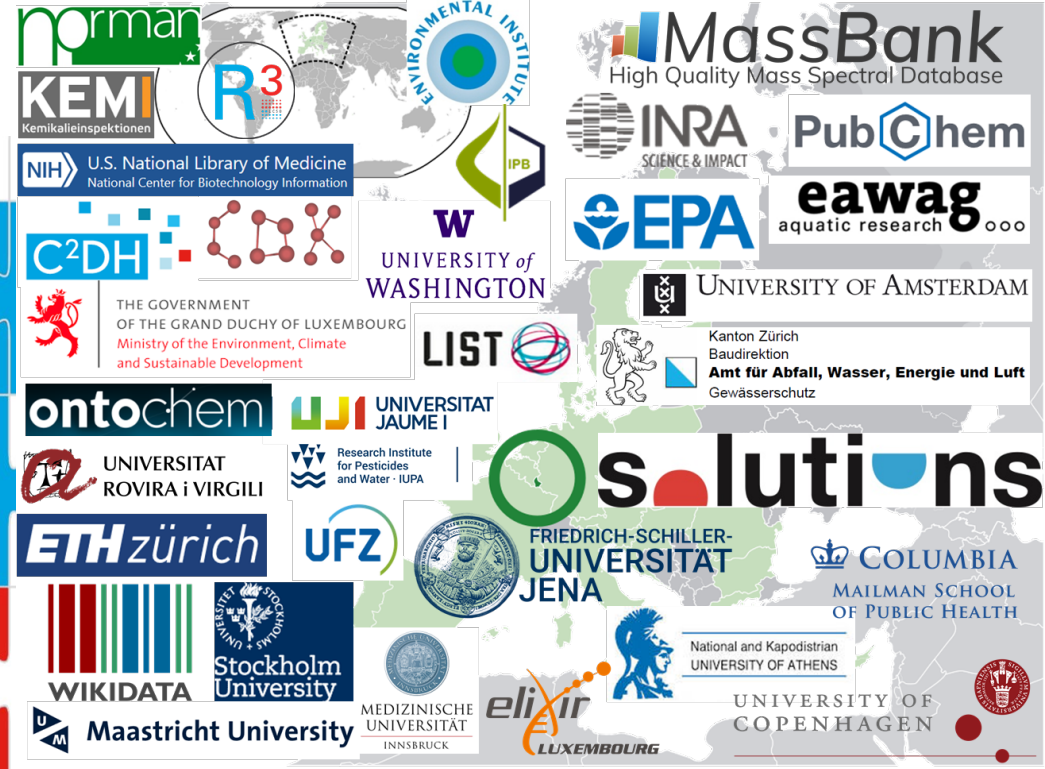
https://git-r3lab.uni.lu/eci/pubchem

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

https://zenodo.org/communities/lcsb-eci/

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





PubChem MetFrag

CCSbase Xu Lab

Slides available at DOI: 10.5281/zenodo.4722508

Luxembourg National Research Fund

uni.lu | LCSB