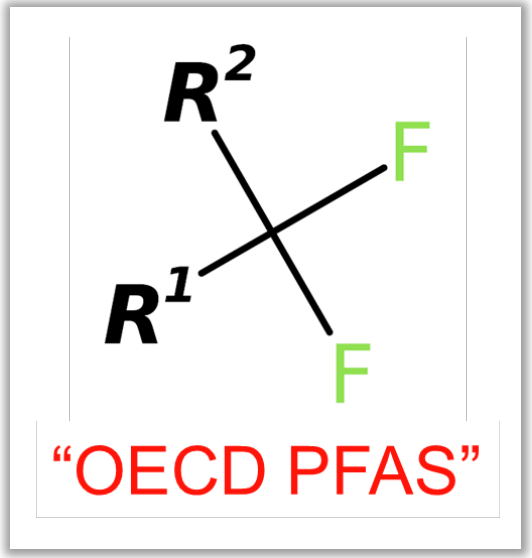
Luxembourg National Research Fund

BAM Bundesanstalt für Materialforschung und -prüfung

UNIVERSITÉ DU LUXEMBOURG

Funded by the European Union

Finding PFAS: Data Exchange to Support Suspect and Non-target Screening of PFAS



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

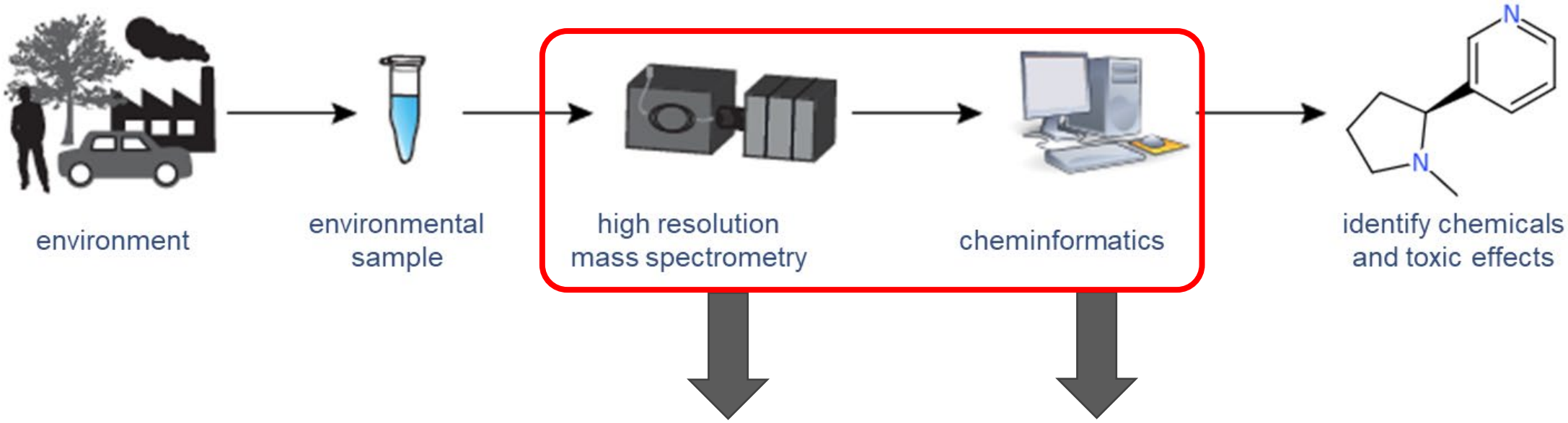
Environmental Cheminformatics Group,
Luxembourg Centre for Systems Biomedicine, University of Luxembourg
emma.schymanski@uni.lu / @ESchymanski / @schymane@mstdn.social
https://wwwen.uni.lu/lcsb/research/environmental_cheminformatics/

BAM PFAS Workshop (hybrid), 19 Sept. 2023, Berlin
https://www.bam.de/Content/EN/Events/2023/2023-09-19-pfas.html

Slides available at DOI:
10.5281/zenodo.8353248

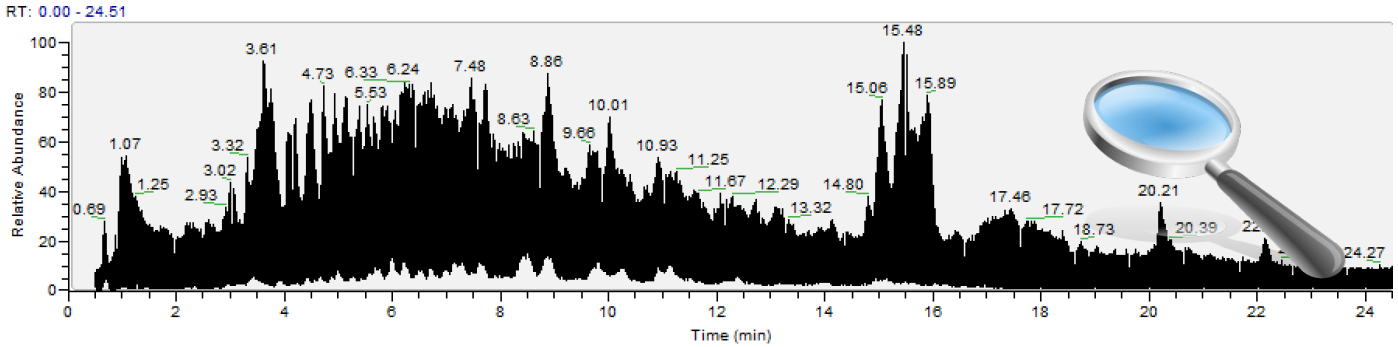
Today's slides
zenodo

Environmental Cheminformatics & Non-target HR-MS



High resolution mass spectrometry

AND connecting chemical knowledge

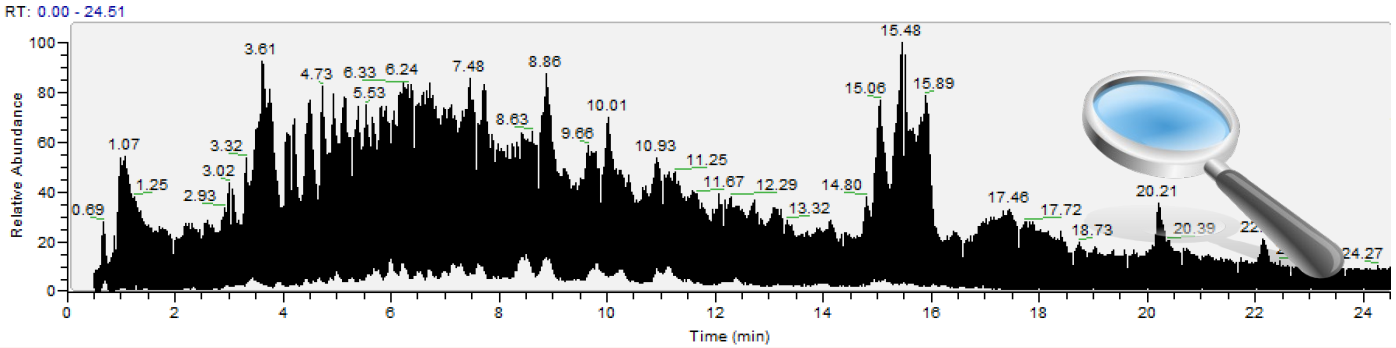


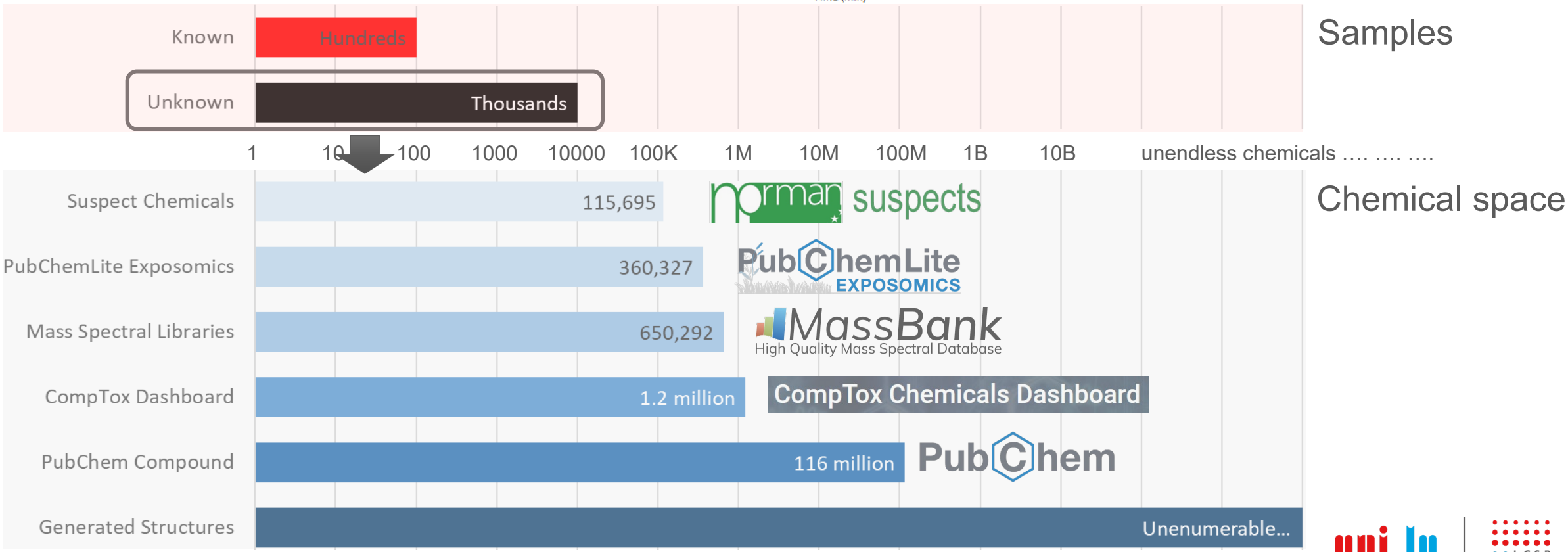
Samples

Environmental Cheminformatics & Non-target HR-MS

High resolution mass spectrometry

AND connecting chemical knowledge





Chromatogram: Schymanski *et al*. (2014) DOI: 10.1021/es4044374; Resource numbers updated 6 Sept. 2023.

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

https://massbank.eu/MassBank/ https://github.com/MassBank/MassBank-data/

MassBank Search Contents Download Accession Go More▼

MassBank Europe

MassBank
High Quality Mass Spectral Database

>> Search Spectra

Search for:

Basic Search | Peak List | Peaks

Compound Information

Compound name

Exact Mass

AND

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

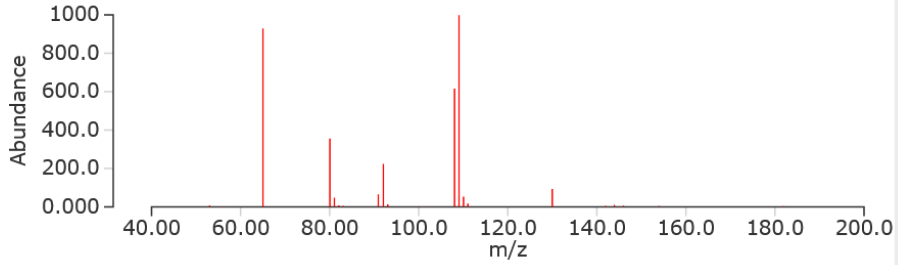
AND

Search

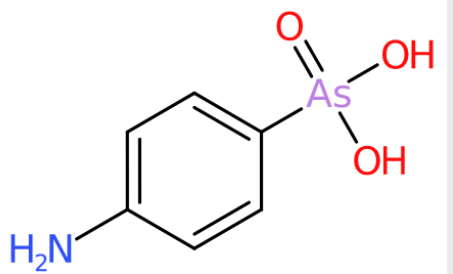
MassBank Record: LU040605

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

Mass Spectrum



Chemical Structure



Gathering Expert Knowledge: NORMAN-SLE

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

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


NORMAN Database System

NORMAN organises the development and maintenance of var

SEARCH All Databases

Searching for individual substance or group(s) of substances

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

NORMAN Suspect List Exchang

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

Comments and contributions are welcome - please email us a

Please refer to our documentation pages for: citation instructi

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

Substance Database

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

Suspect List Exchange

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

Antibiotic Resistance Bacteria/Genes

A database of ARBs/ARGs in environmen
matrices

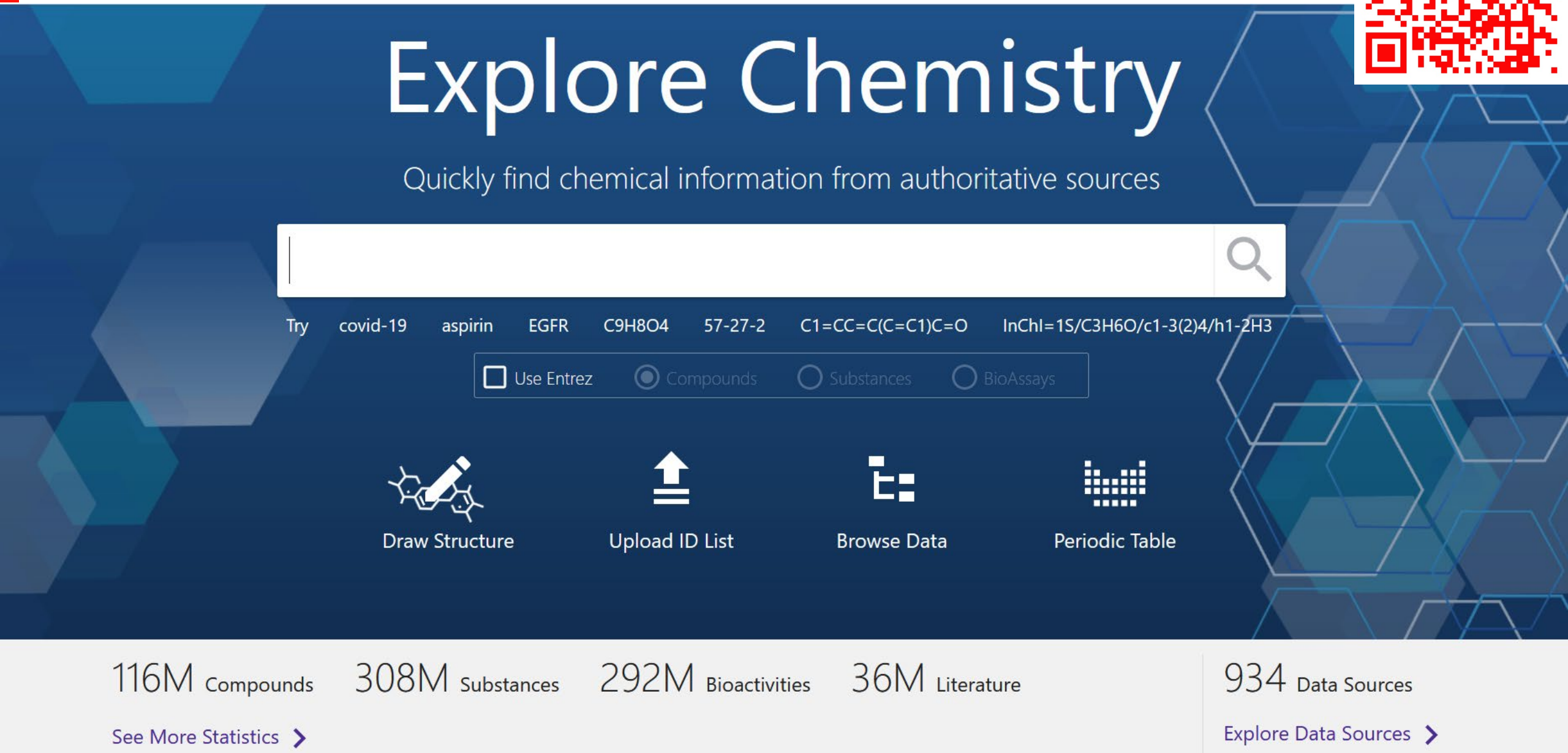
RESEARCH Open Access

Check for updates

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

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

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

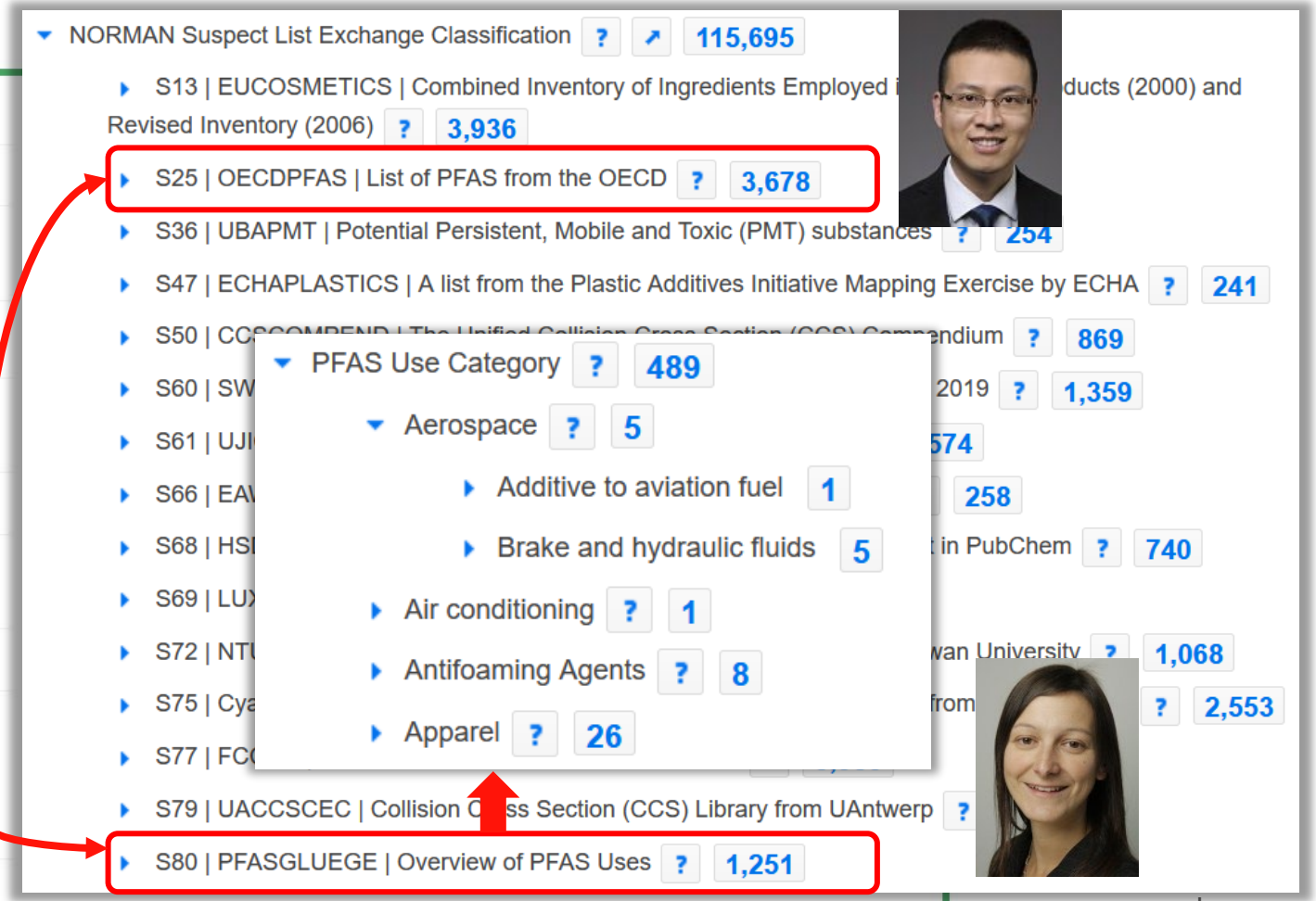


NORMAN-SLE in PubChem

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

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

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



https://pubchem.ncbi.nlm.nih.gov/classification/#hid=101 & https://pubchem.ncbi.nlm.nih.gov/source/23819

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

NORMAN guidance on suspect and non-target screening in environmental monitoring

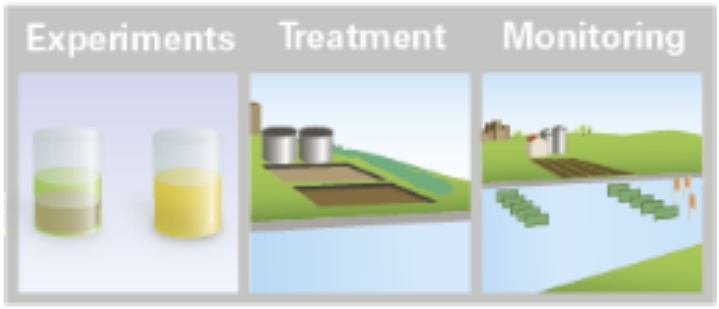
Review | Open Access | Published: 04 September 2023 | 35, Article number: 75 (2023)

Juliane Hollender, Emma L. Schymanski, Lutz Ahrens, Nikiforos Alygizakis, Frederic Béen, Lubertus Bijlsma, Andrea M. Brunner, Alberto Celma, Aurelie Fildier, Qiuguo Fu, Pablo Gago-Ferrero, Ruben Gil-Solsona, Peter Haglund, Martin Hansen, Sarit Kaserzon, Anneli Kruve, Marja Lamoree, Christelle Margoum, Jeroen Meijer, Sylvain Merel, Cassandra Rauert, Pawel Rostkowski, Saer Samanipour, Bastian Schulze, Tobias Schulze, Randolph R. Singh, Jaroslav Slobodnik, Teresa Steininger-Mairinger, Nikolaos S. Thomaidis, Anne Togola, Katrin Vorkamp, Emmanuelle Vulliet, Linyan Zhu & Martin Krauss

— Show fewer authors

Hollender *et al.* (2023) DOI: 10.1186/s12302-023-00779-4

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



Commercial / Open Software

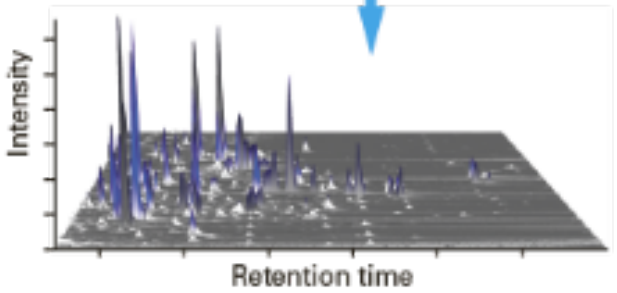
Databases

In Silico Prediction

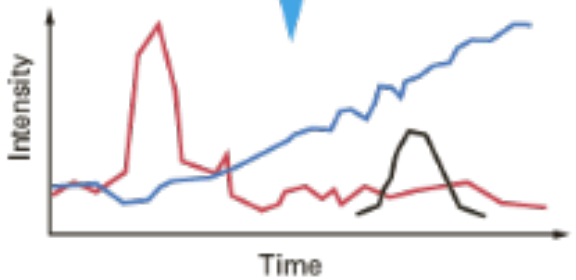
Sampling → Analysis → Data Pre-Processing → Prioritization → Identification

SAMPLE ARCHIVE

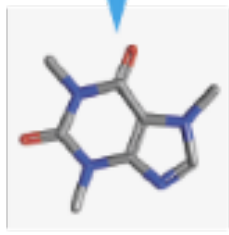
DIGITAL ARCHIVE



POLLUTION OVERVIEW



PATTERN



STRUCTURES

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

Open Source Workflows for NT-HRMS: patRoon

Software | Open Access | Published: 06 January 2021

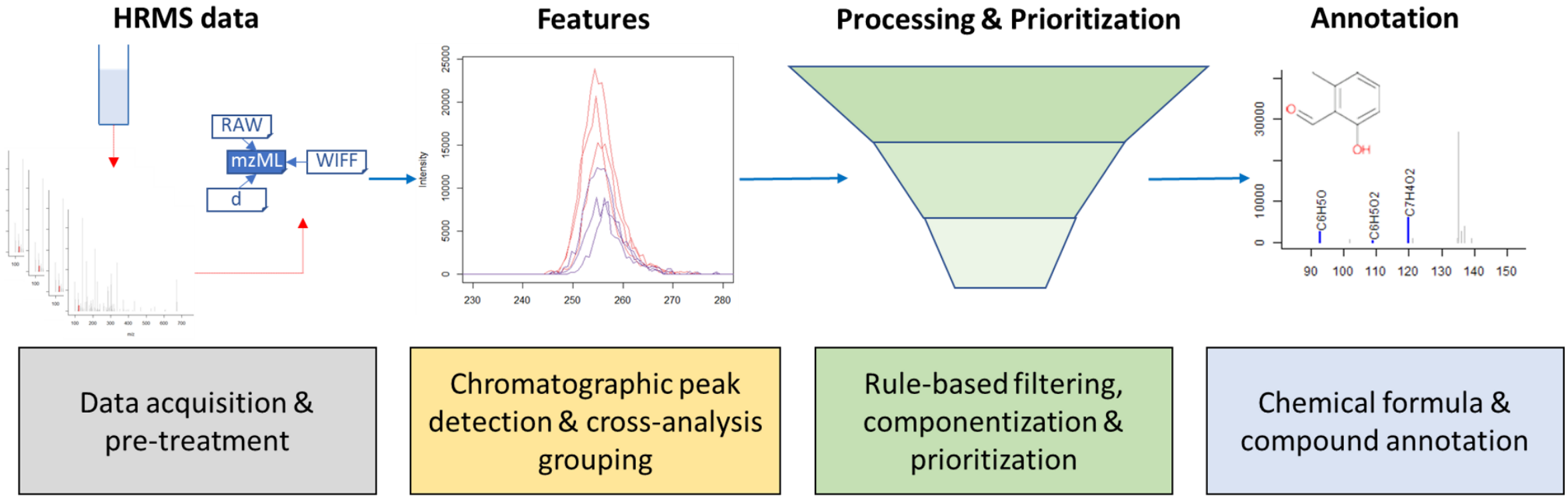
patRoon: open source software platform for environmental mass spectrometry based non-target screening

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

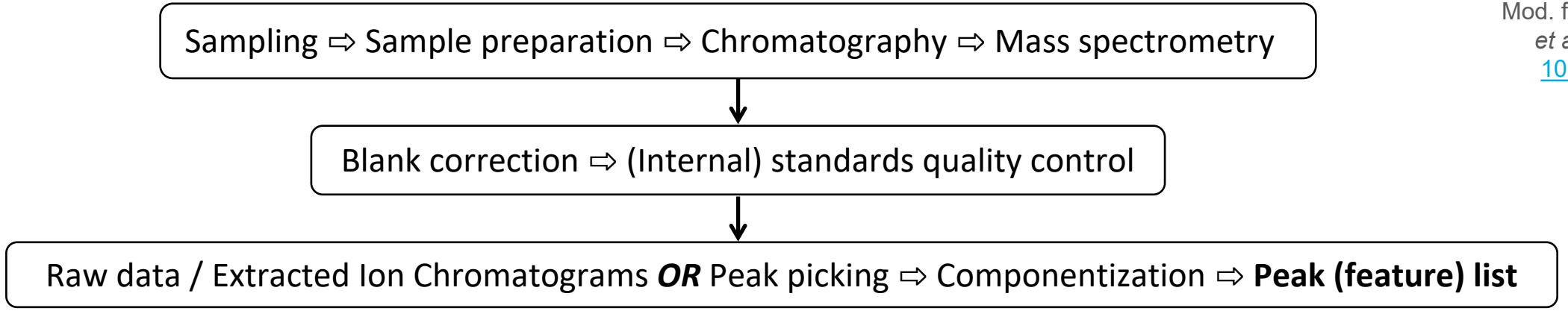
Rick Helmus, Thomas L. ter Laak, Annemarie P. van Wezel, Pim de Voogt & Emma L. Schymanski

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

Rick Helmus[1¶], Bas van de Velde[123], Andrea M. Brunner[24], Thomas L. ter Laak[12], Annemarie P. van Wezel[1], and Emma L. Schymanski[5]



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



Mod. from Hollender *et al*. (2023) DOI: 10.1186/s12302-023-00779-4

Sampling ⇨ Sample preparation ⇨ Chromatography ⇨ Mass spectrometry

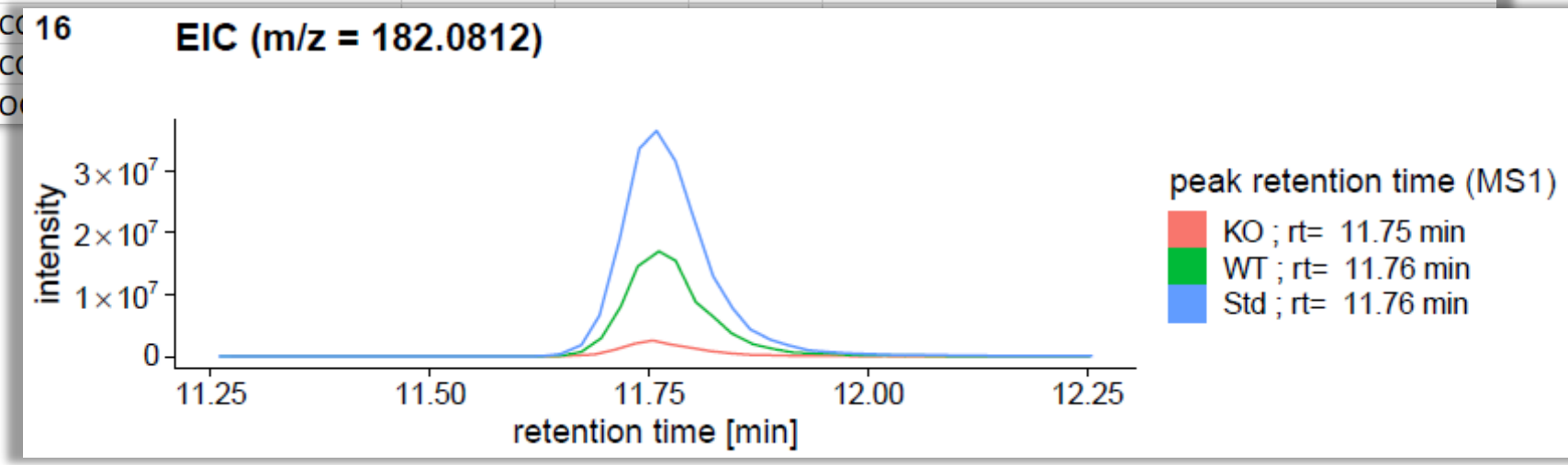
Blank correction ⇨ (Internal) standards quality control

Raw data / Extracted Ion Chromatograms ***OR*** Peak picking ⇨ Componentization ⇨ **Peak (feature) list**

Mod. from Hollender *et al.* (2023) DOI: 10.1186/s12302-023-00779-4

Target screening with target list

Name	SMILES	RT	Prec_mz	Adduct	MS/MS
N-Benzyladenine	C(NC1=C2N=CN=C2N=CN1)C1=C	13.340 min	226.1087	[M+H]+	91.0542;999:65.0386;115:63.0229;3:148.0617;2:119.(
Benodanil	IC1=C(C=CC=C1)C(=O)NC1=CC=C(	16.272 min	323.988	[M+H]+	94.0413;999:76.0307;885:81.0335;317:66.0464;147:2
Pimozide	FC1=CC=C(C=C1)C(CCCN1CCC(C(	15.883 min	462.2351	[M+H]+	109.0448;999:173.0708;358:147.0604;353:98.0964;28
Norflurazon	CNC1=C(Cl)C(=O)N(N=C1)C1=CC	16.641 min	304.0459	[M+H]+	140.0306;999:87.9948;888:102.0105;676:160.0368;47
Zenarestat	OC(=O)CN1C(=O)N(CC2=CC=C(B	18.674 min	438.9502	[M-H]-	78.9189;999:360.9386;317:166.0065;284:196.9924;27
Ametryn	CC				
Pyriproxyfen	CC				
Raloxifene	OC				



Confirmation with standard & quantification

Mod. from Hollender *et al.* (2023) DOI: 10.1186/s12302-023-00779-4

Sampling ⇨ Sample preparation ⇨ Chromatography ⇨ Mass spectrometry

Blank correction ⇨ (Internal) standards quality control

Raw data / Extracted Ion Chromatograms ***OR*** Peak picking ⇨ Componentization ⇨ **Peak (feature) list**

Target screening with target list

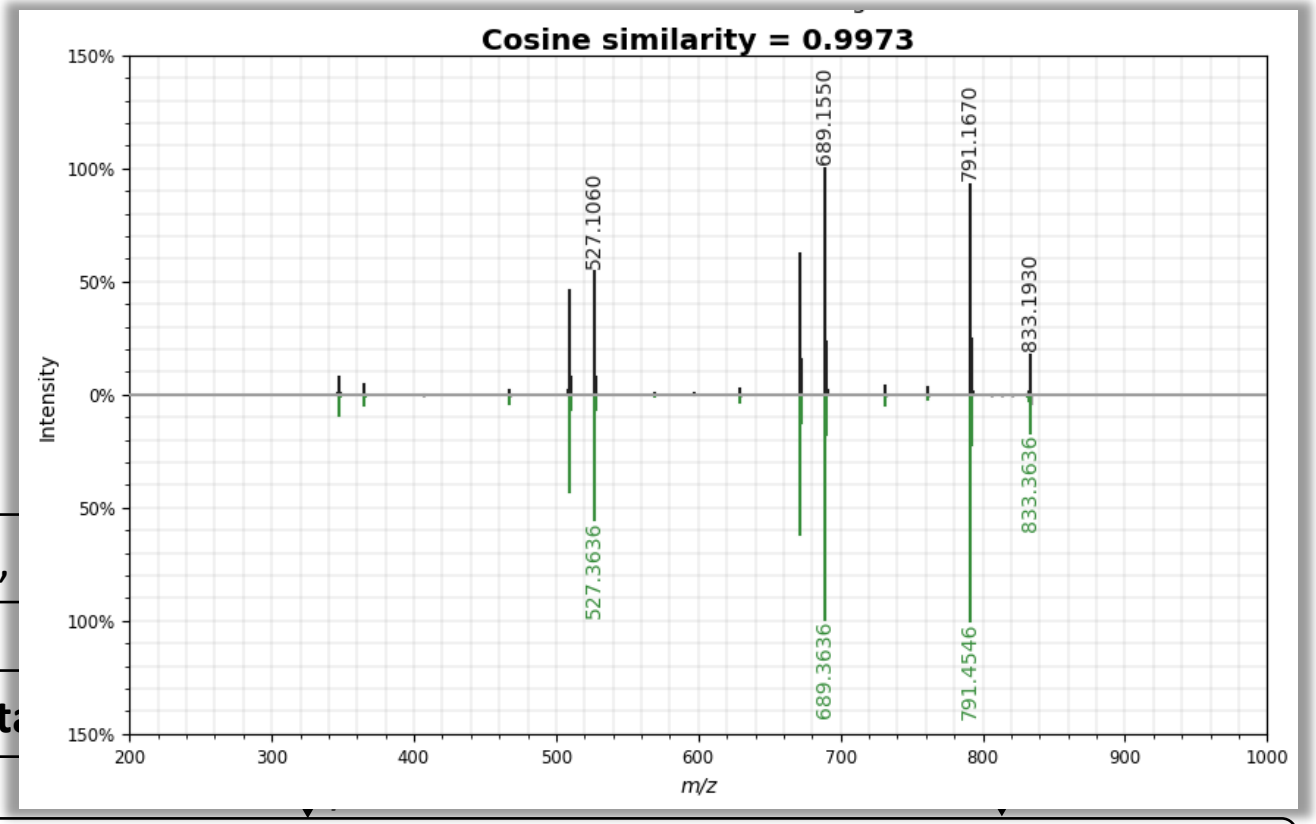
MS(MS) **library search**

Candidate filtering: MS/MS,

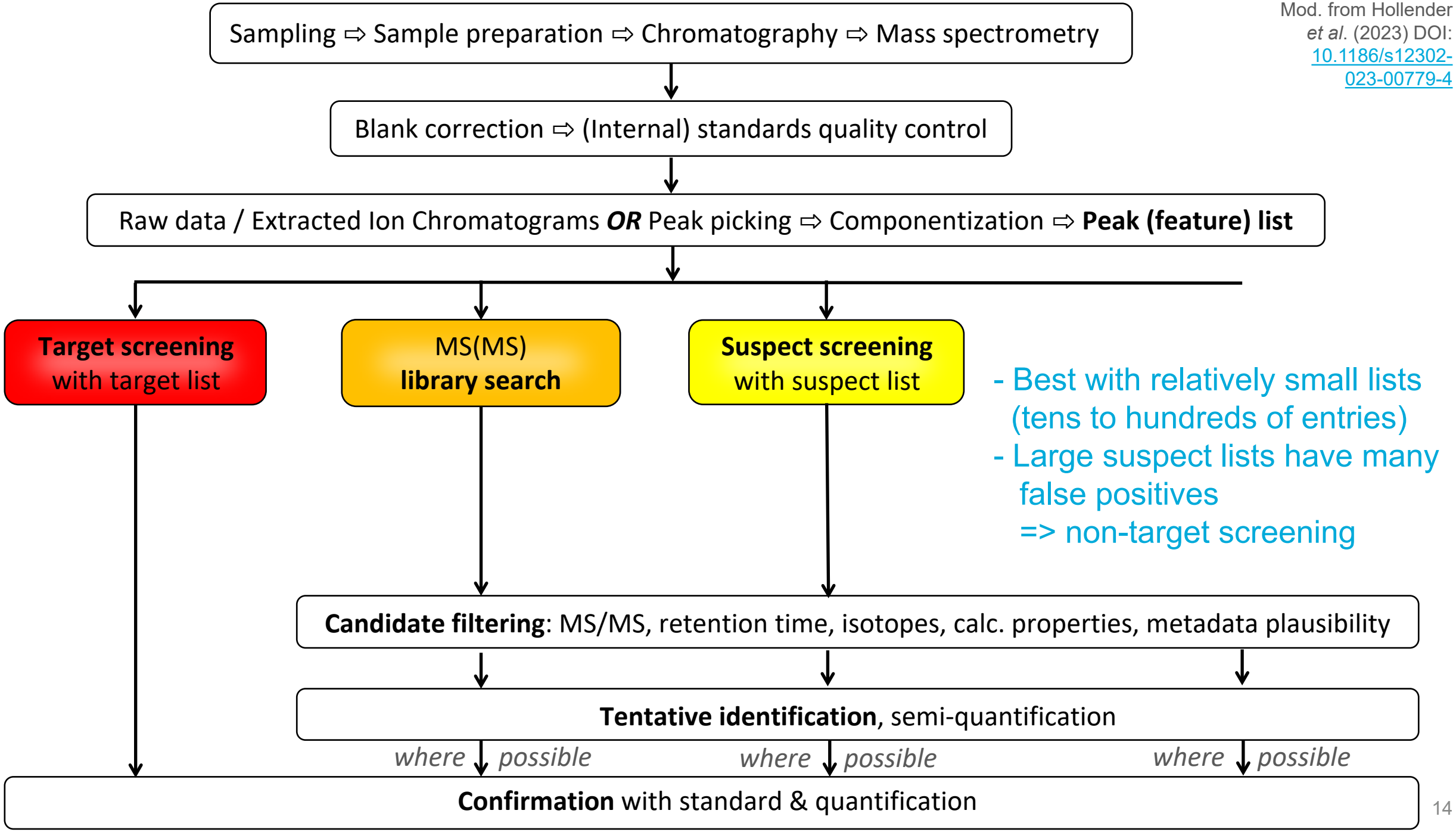
Tenta

where possible

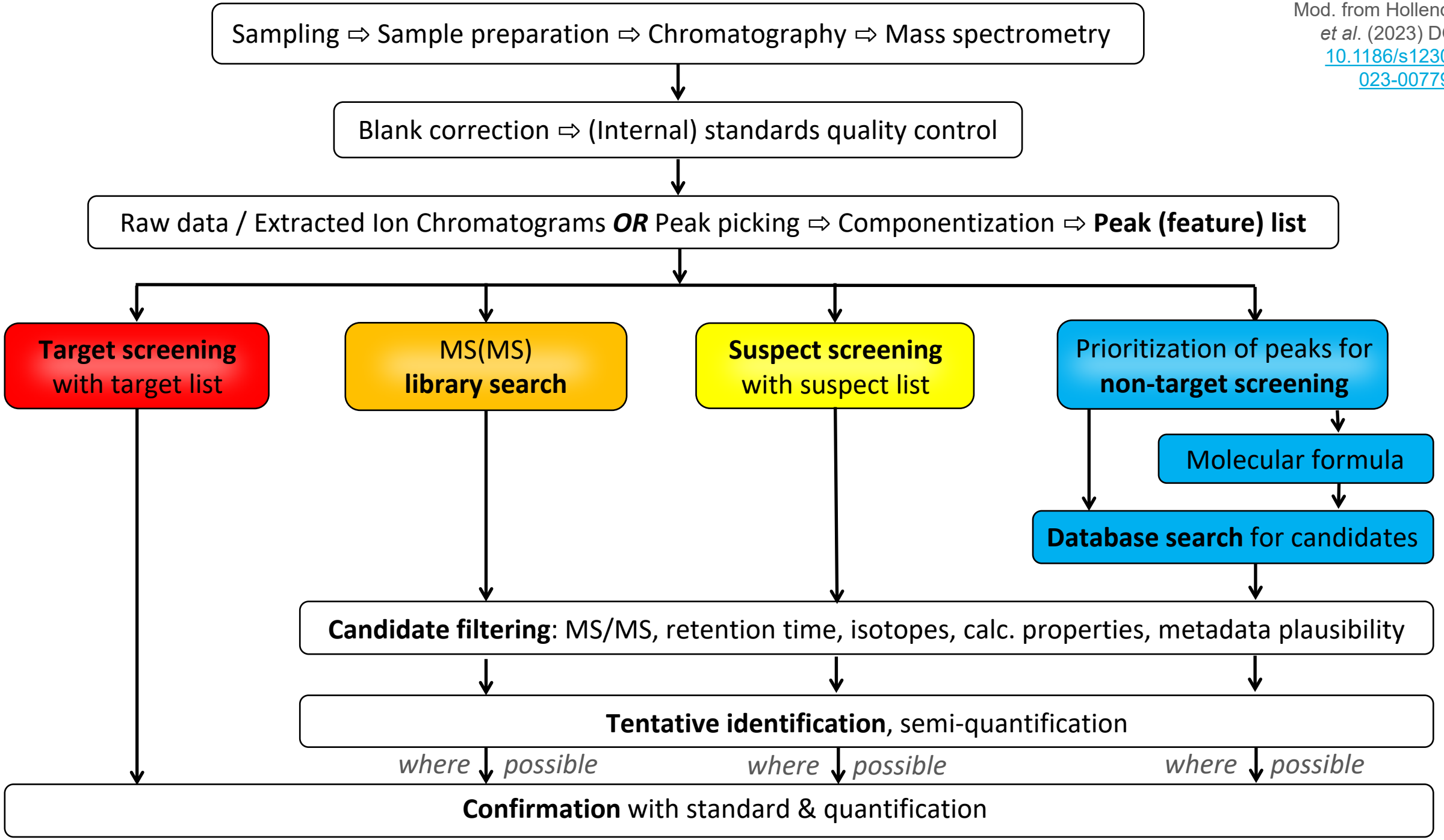
Confirmation with standard & quantification



GNPS Mirror Plot Example

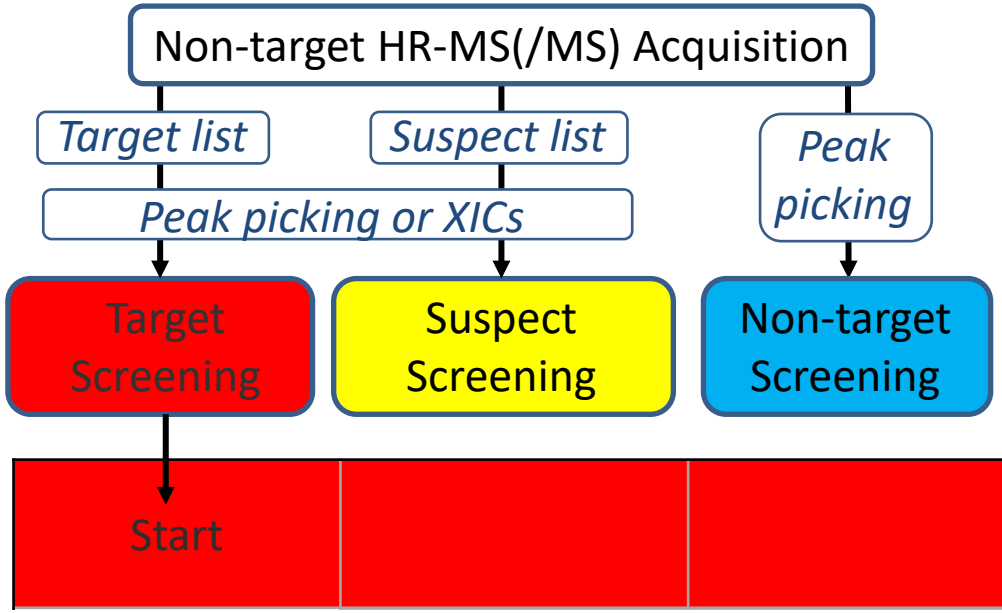


Mod. from Hollender *et al.* (2023) DOI: 10.1186/s12302-023-00779-4



Mod. from Hollender *et al.* (2023) DOI: 10.1186/s12302-023-00779-4

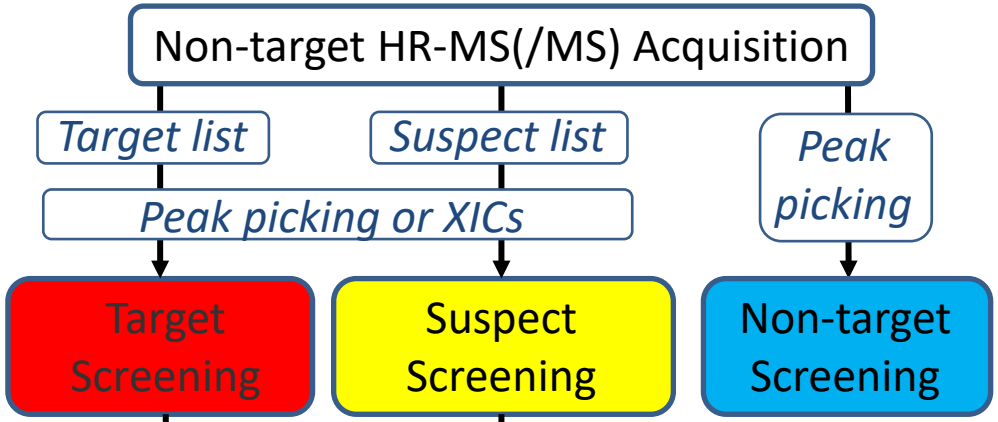
Identification Strategies and Confidence in NT-HRMS(/MS)



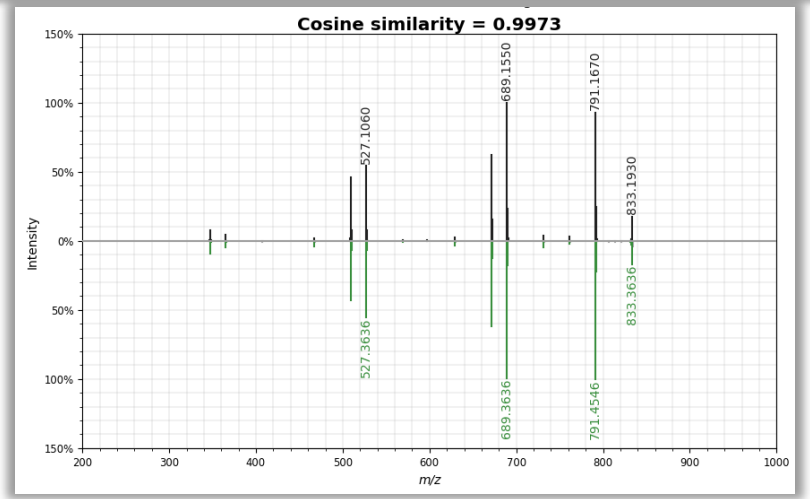
Limited by ref. std. availability …

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

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

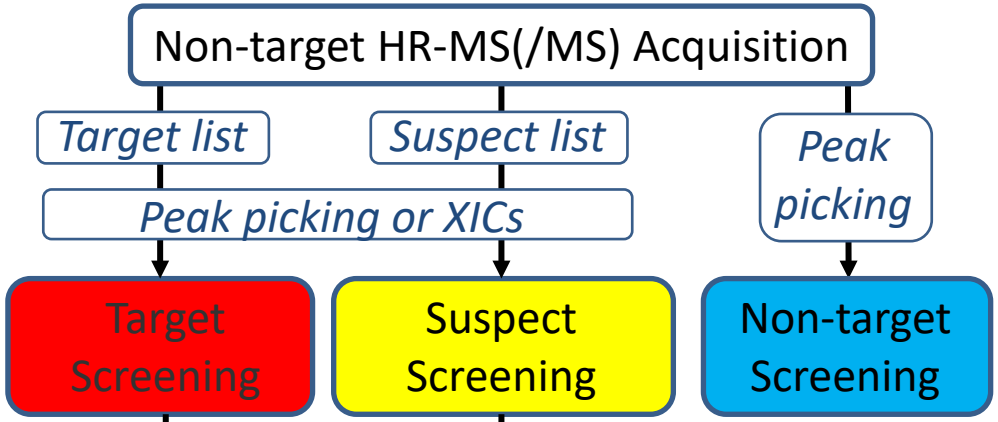


Level 1	**Confirmed Structure** by reference standard	*Limited by ref. std. availability …*
Level 2	**Probable Structure** by library/diagnostic evidence	MassBank High Quality Mass Spectral Database



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

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

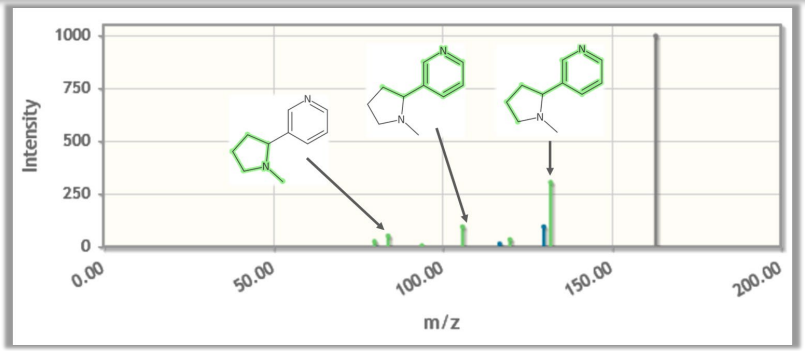


Start			**Level 1**	**Confirmed Structure** by reference standard
			Level 2	**Probable Structure** by library/diagnostic evidence
	Start		**Level 3**	**Tentative Candidate(s)** suspect, substructure, class

Limited by ref. std. availability …

MassBank
High Quality Mass Spectral Database

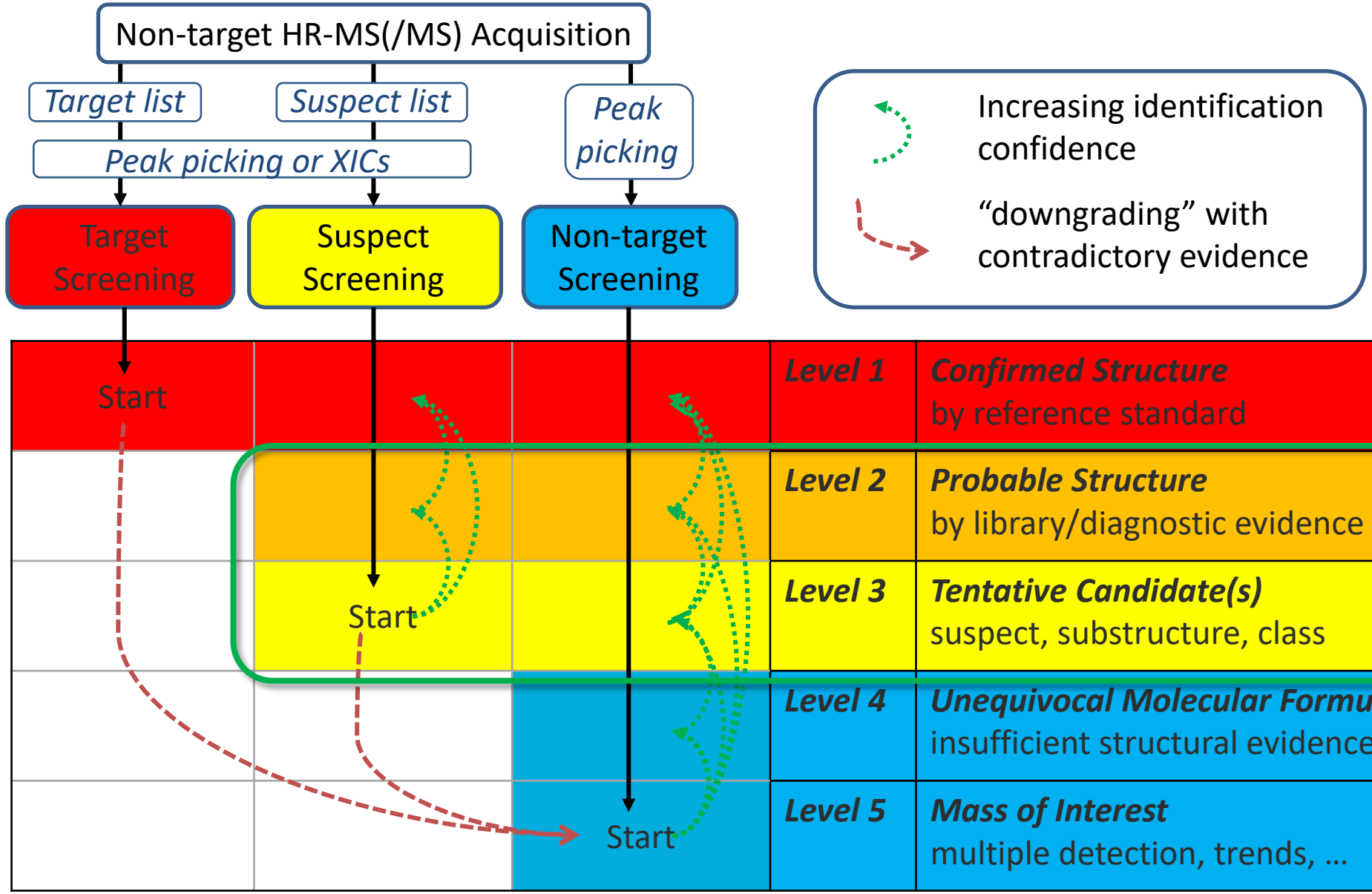
MetFrag



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

eawag aquatic research

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



Limited by ref. std. availability …

MassBank
High Quality Mass Spectral Database

MetFrag

The "dark matter"

eawag aquatic research

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

A PFAS-Specific Confidence Scale (adapting, not reinventing)

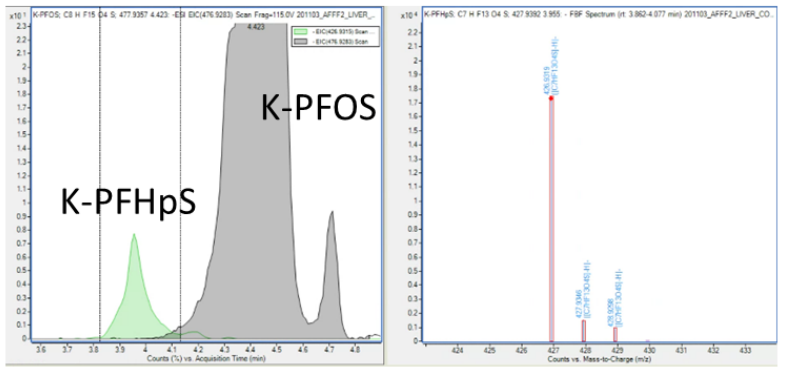
Level	Identification Confidence	Accurate Mass	Mass Defect	Isotopic Pattern Match	Consistent RT	Homologue (number; level)	MS^2 Fragments (number; type)	Library MS^2	Reference Standard
Level 1a	Confirmed by reference standard	✓	✓	✓	✓			✓	✓
Level 1b	Indistinguishable from reference standard	✓	✓	✓	✓			✓	✓
Level 2a	Probable by library spec. match	✓	✓	✓	✓			✓	
Level 2b	Probable by diagnostic fragmentation evidence	✓	✓	✓	✓	*≥ 1; ≥ level 3*	**≥ 3; diagnostic**		
Level 2c	Probable by diagnostic homologue evidence	✓	✓	✓	✓	**≥ 2; ≥ level 2a**	**≥ 2; diagnostic**		
Level 3a	Positional isomer candidates	✓	✓	✓	✓	*≥ 1; ≥ level 3*	**≥ 1; subclass-aligned**		
Level 3b	Fragmentation-based candidate	✓	✓	✓	✓	*≥ 1; ≥ level 3*	**≥ 1; subclass-aligned**		
Level 3c	Circumstantial candidate with fragmentation evidence	✓	✓	✓	✓	*≥ 1; ≥ level 3*	**≥ 1; subclass-aligned (in silico)**		
Level 3d	Circumstantial candidate with homologue evidence	✓	✓	✓	✓	**≥ 2; ≥ level 2a**			
Level 4	Unequivocal molecular formula	✓	✓	✓					
Level 5a	PFAS suspect screening exact mass match	**✓ (suspect list match)**							
Level 5b	Nontarget PFAS exact mass of interest	✓	✓			*≥ 3*	*≥ 2; subclass-aligned (in silico)*		

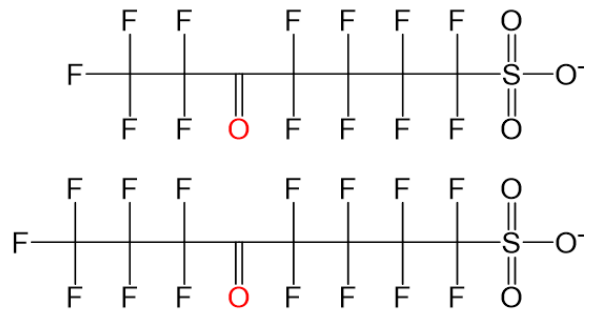
Charbonnet *et al*, 2022, ES&T Letters DOI: 10.1021/acs.estlett.2c0206

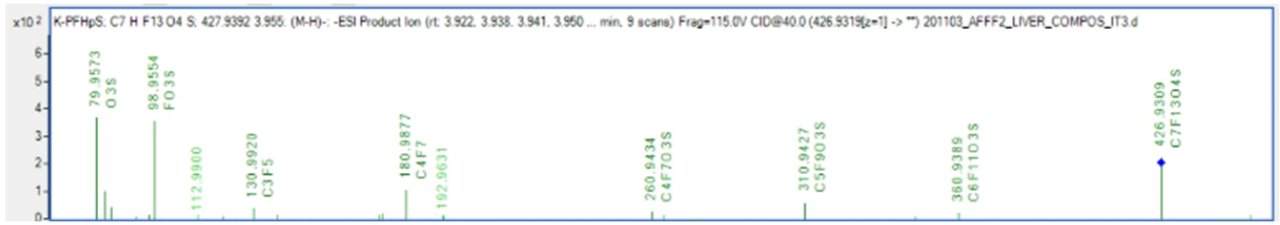
A PFAS-Specific Confidence Scale (adapting, not reinventing)

Level	Identification Confidence	Accurate Mass	Mass Defect	Isotopic Pattern Match	Consistent RT	Homologue (number; level)	MS^2 Fragments (number; type)	Library MS^2	Reference Standard
Level 1a	Confirmed by reference standard	✓	✓	✓	✓			✓	✓
Level 1b	Indistinguishable from reference standard	✓	✓	✓	✓			✓	✓
Level 2a	Probable by library spec. match	✓	✓	✓	✓			✓	
Level 2b	Probable by diagnostic fragmentation evidence	✓	✓	✓	✓	*≥ 1; ≥ level 3*	**≥ 3; diagnostic**		
Level 2c	Probable by diagnostic homologue evidence	✓	✓						
Level 3a	Positional isomer candidates	✓	✓						
Level 3b	Fragmentation-based candidate	✓	✓						
Level 3c	Circumstantial candidate with fragmentation evidence	✓	✓						
Level 3d	Circumstantial candidate with homologue evidence	✓	✓						
Level 4	Unequivocal molecular formula	✓	✓						
Level 5a	PFAS suspect screening exact mass match	**✓ (suspect list match)**							
Level 5b	Nontarget PFAS exact mass of interest	✓	✓						

Example: Ketone-Substituted PFHpS ($C_7F_{13}SO_4^-$)







Charbonnet *et al*, 2022, ES&T Letters DOI: 10.1021/acs.estlett.2c0206. Examples c/o J. Charbonnet

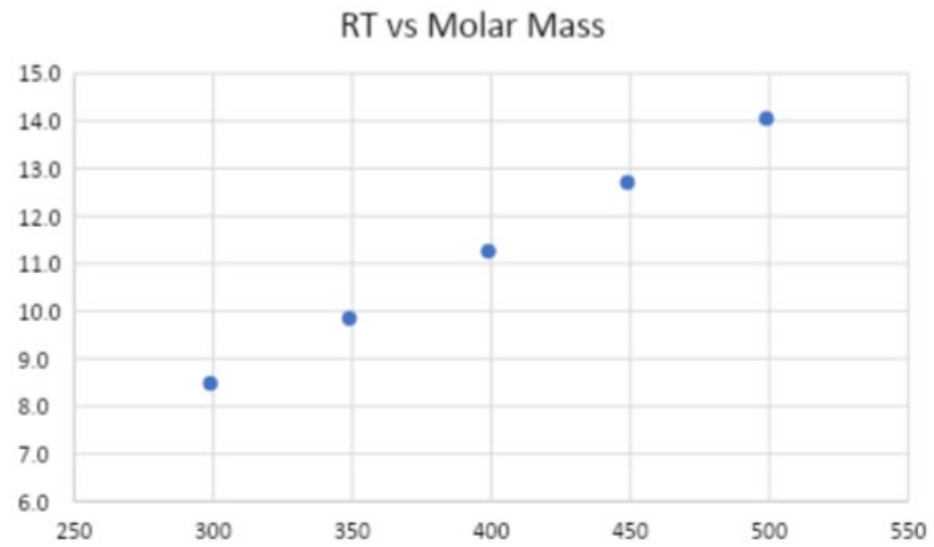
A PFAS-Specific Confidence Scale (adapting, not reinventing)

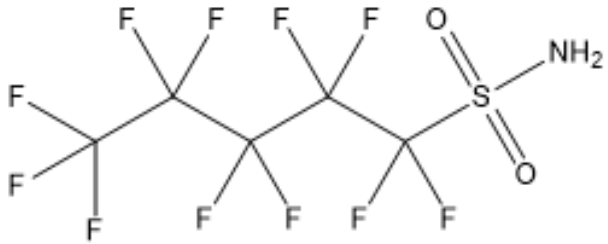
Level	Identification Confidence	Accurate Mass	Mass Defect	Isotopic Pattern Match	Consistent RT	Homologue (number; level)	MS^2 Fragments (number; type)	Library MS^2	Reference Standard
Level 1a	Confirmed by reference standard	✓	✓	✓	✓			✓	✓
Level 1b	Indistinguishable from reference standard								
Level 2a	Probable by library spec. match								
Level 2b	Probable by diagnostic fragmentation evidence								
Level 2c	Probable by diagnostic homologue evidence								
Level 3a	Positional isomer candidates								
Level 3b	Fragmentation-based candidate								
Level 3c	Circumstantial candidate with fragmentation evidence								
Level 3d	Circumstantial candidate with homologue evidence								
Level 4	Unequivocal molecular formula								
Level 5a	PFAS suspect screening exact mass match								
Level 5b	Nontarget PFAS exact mass of interest	✓	✓			≥ 3	≥ 2, subclass-aligned (in silico)		

Example: FPeSA ($C_5F_{11}SO_2NH_2$)

Fragment: 98.9790 in all series members [FSO_2NH_2]

FOSA confirmed with a standard





Charbonnet *et al*, 2022, ES&T Letters DOI: 10.1021/acs.estlett.2c0206. Examples c/o J. Charbonnet

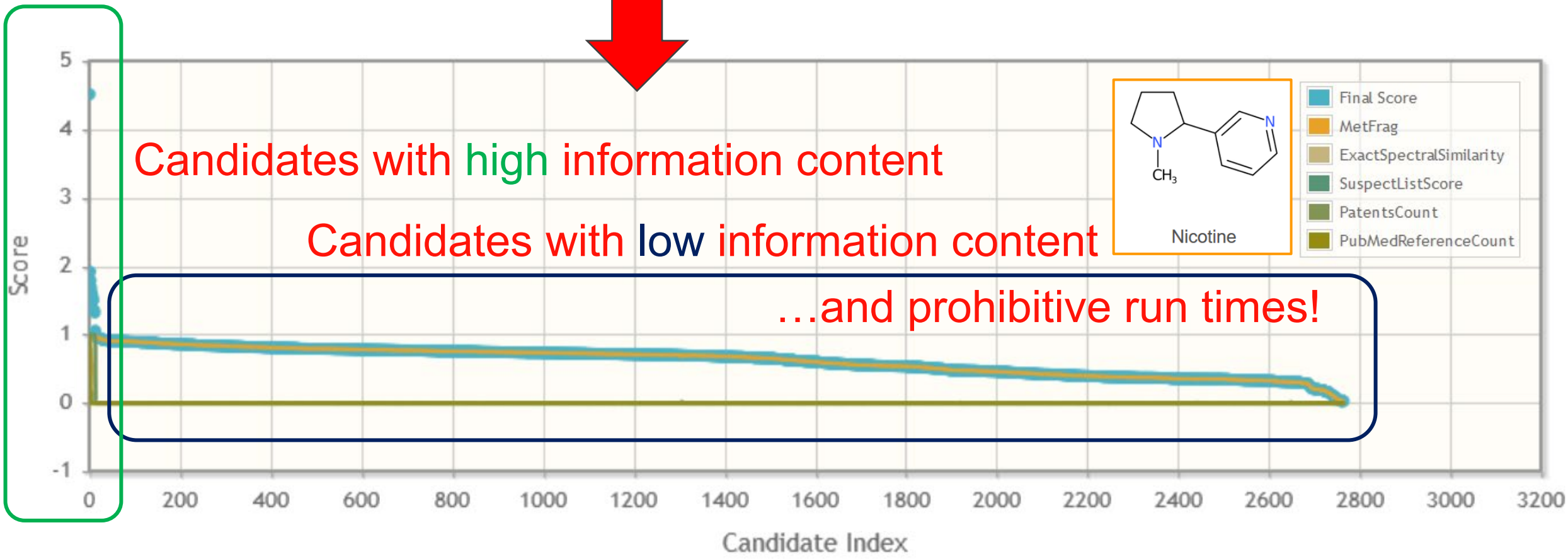
Grand Challenge: HR-MS “Chemical Space” is too big!

CAS® A DIVISION OF THE AMERICAN CHEMICAL SOCIETY — 204 million

PubChem — 116 million

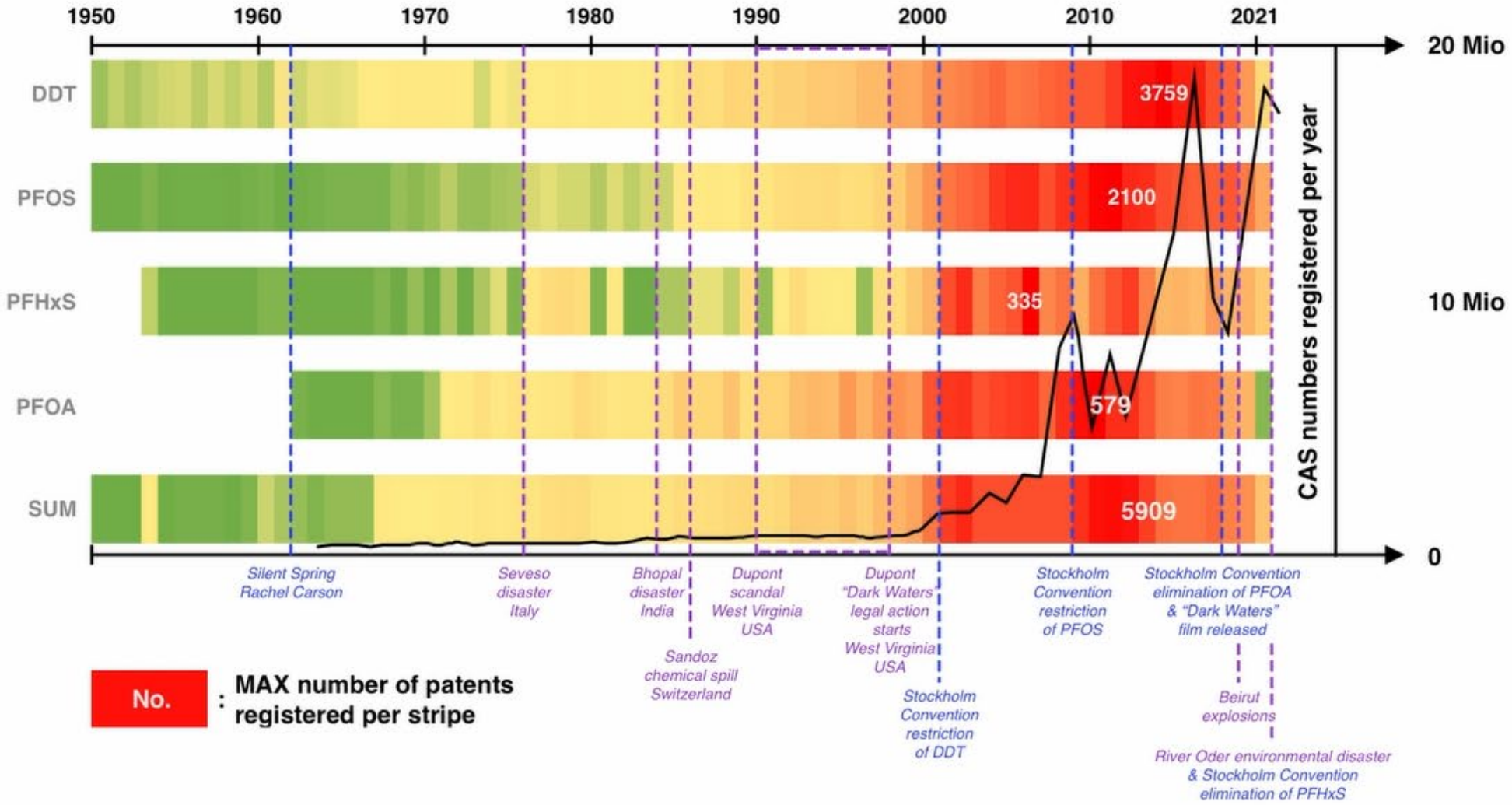
ChemSpider Search and share chemistry — 128 million

CompTox Chemicals Dashboard — 1.2 million



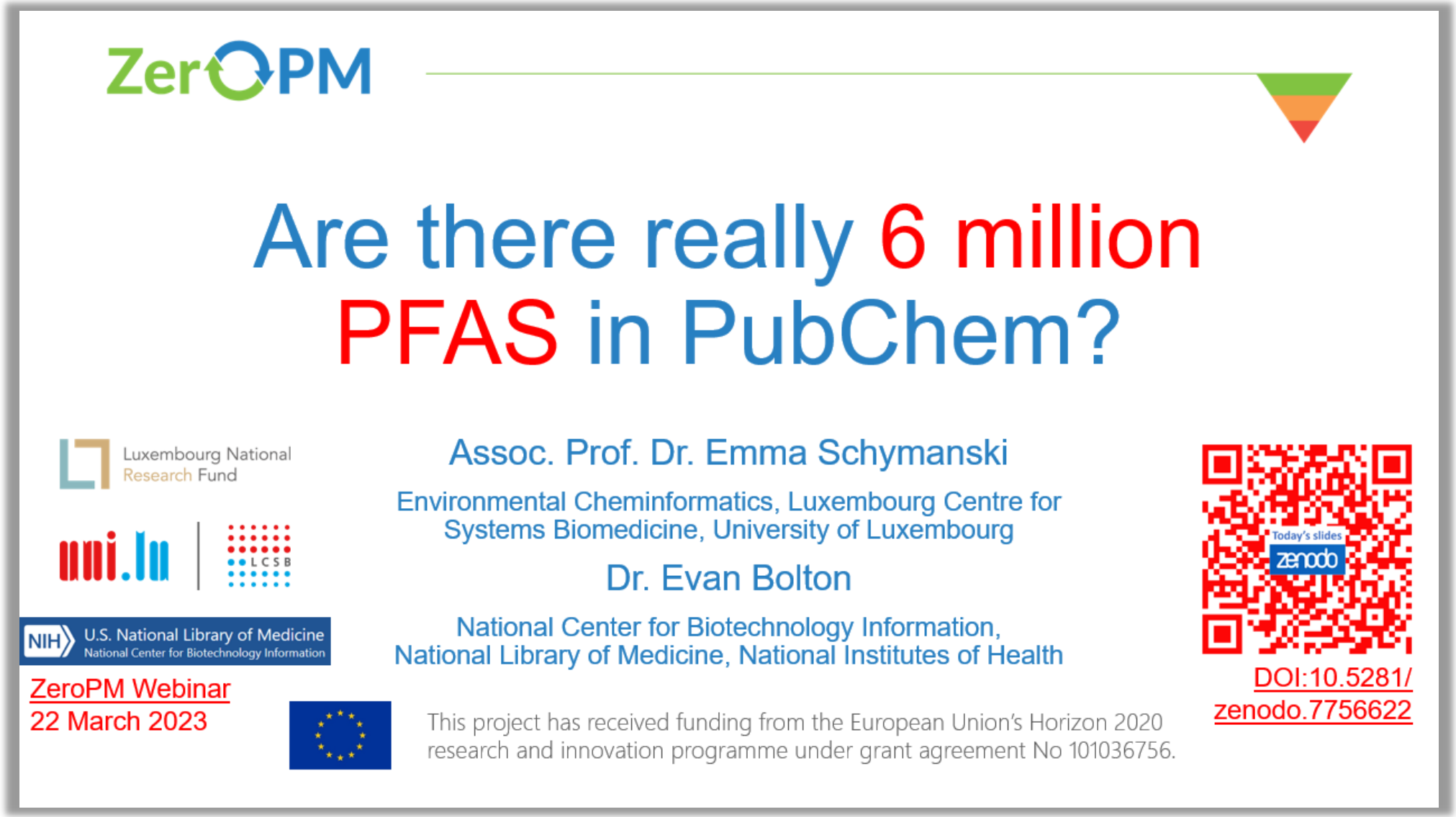
Our Chemical Past, Present, and Future

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



Composition & Animation: Jamie Perera; Figure: Dagny Aurich; mod. from Arp et al. (2023) DOI: 10.1021/acs.est.3c01735. Credits: Jamie Perera, Dagny Aurich, Paul Thiessen, Emma Schymanski

[Flashback 6 months ago …]

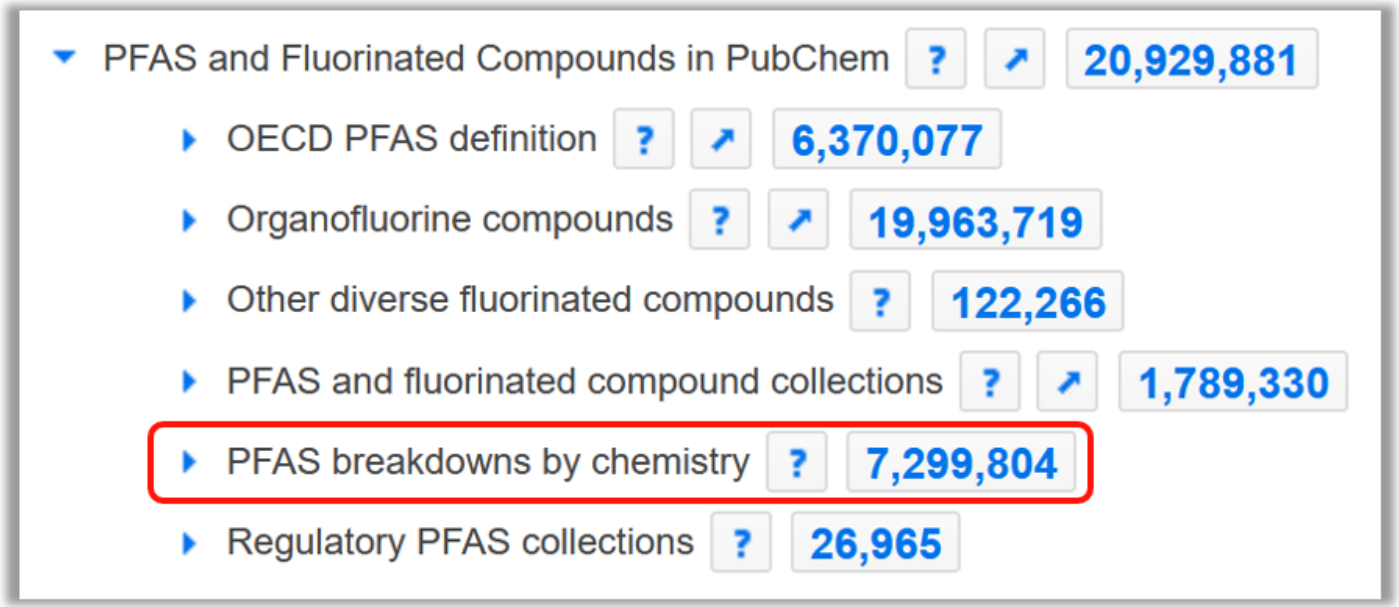


Schymanski & Bolton (2023) ZeroPM Webinar, Zenodo. DOI: 10.5281/zenodo.7756622

[Flashback 6 months ago …]

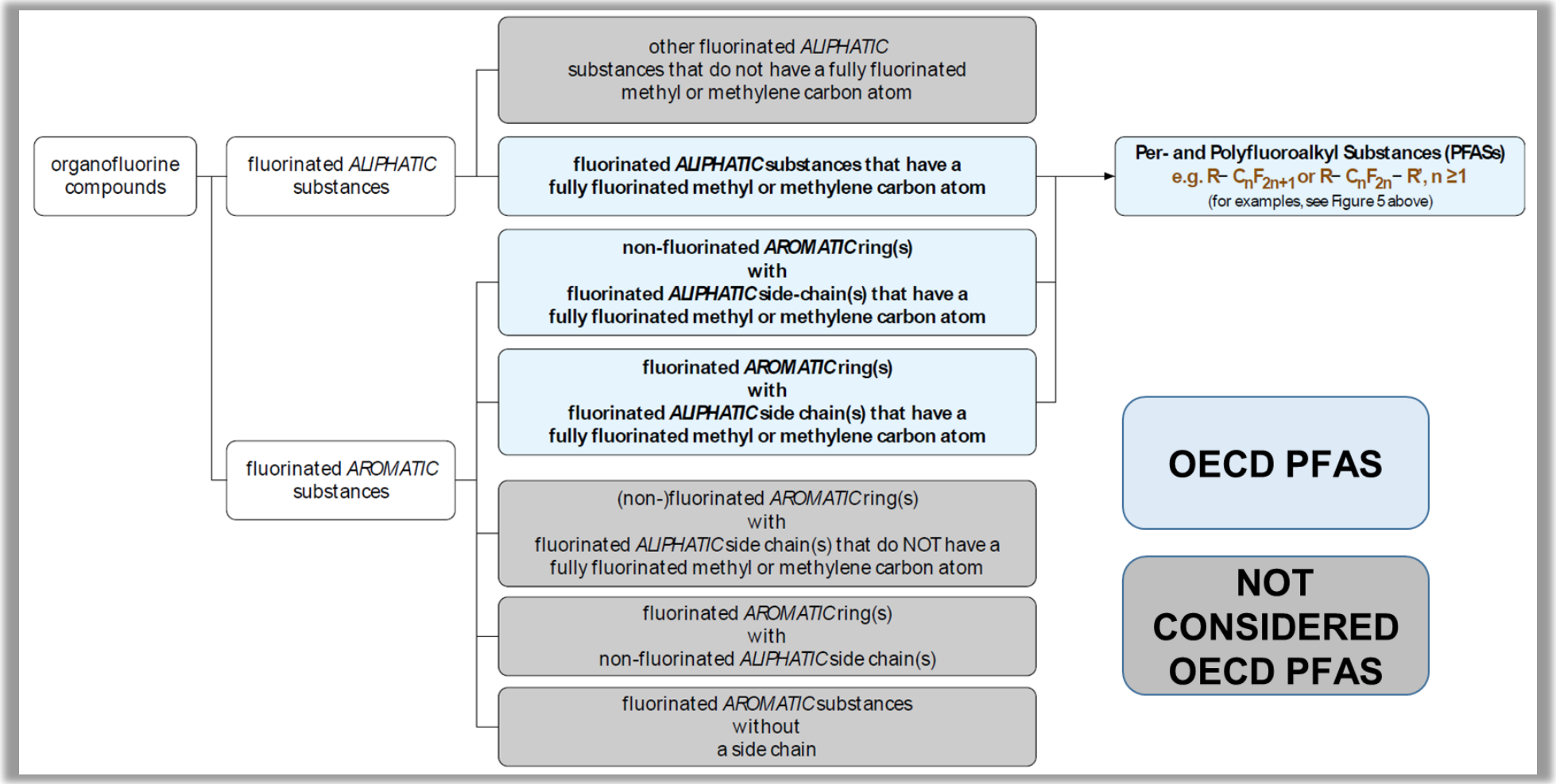
ZeroPM

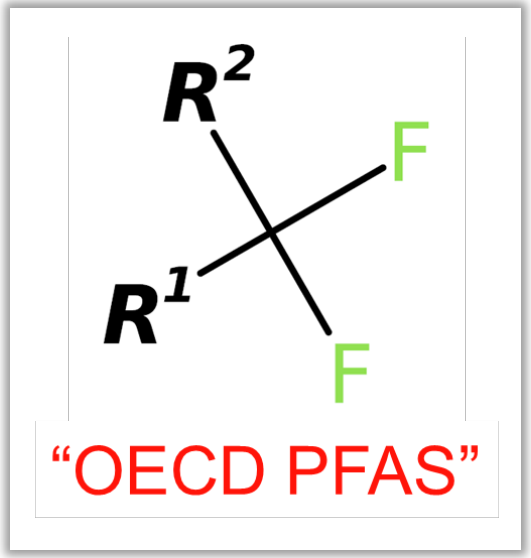
No…it's already 7 million!



Schymanski & Bolton (2023) ZeroPM Webinar, Zenodo. DOI: 10.5281/zenodo.7756622

Motivation: Updated OECD PFAS Definition in 2021



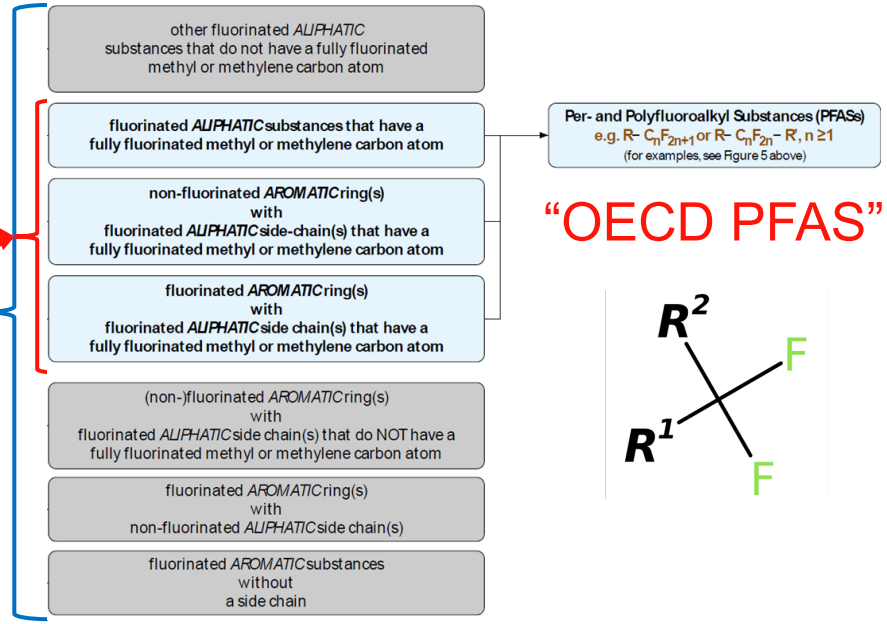


OECD Monograph ENV/CBC/MONO(2021)25 (9 July 2021) – Figure 8

The PubChem PFAS Tree

- PFAS and Fluorinated Compounds in PubChem — 21,411,181
  - OECD PFAS definition — 6,540,217
    - Molecule contains isolated CF2 — 675,776
    - Molecule contains isolated CF3 — 5,747,364
    - Molecule contains PFAS parts larger than CF2/CF3 — 229,607
  - Organofluorine compounds — 20,417,011
    - Fluorinated aliphatic substances — 904,417
    - Fluorinated aromatic substances — 19,439,533
    - Other fluorinated substances — 97,762
  - Other diverse fluorinated compounds — 125,621
  - PFAS and fluorinated compound collections — 1,789,296
  - PFAS breakdowns by chemistry — 7,497,376
  - Regulatory PFAS collections — 26,943

"OECD PFAS" (+salts/mixtures)



"OECD PFAS"

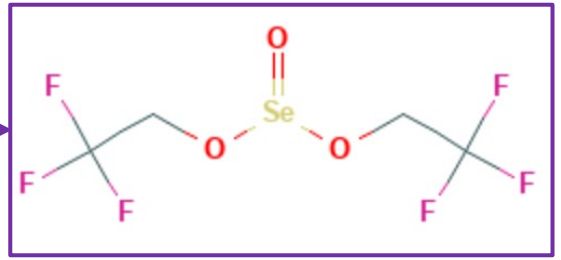
OECD Monograph ENV/CBC/MONO(2021)25 (9 July 2021)

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

The PubChem PFAS Tree

- PFAS and Fluorinated Compounds in PubChem — 21,411,181
  - OECD PFAS definition — 6,540,217
    - Molecule contains isolated CF2 — 675,776
    - Molecule contains isolated CF3 — 5,747,364
    - Molecule contains PFAS parts larger than CF2/CF3 — 229,607
  - Organofluorine compounds — 20,417,011
    - Fluorinated aliphatic substances — 904,417
    - Fluorinated aromatic substances — 19,439,533
    - Other fluorinated substances — 97,762
  - Other diverse fluorinated compounds — 125,621
  - PFAS and fluorinated compound collections — 1,789,296
  - PFAS breakdowns by chemistry — 7,497,376
  - Regulatory PFAS collections — 26,943

"OECD PFAS" (+salts/mixtures)



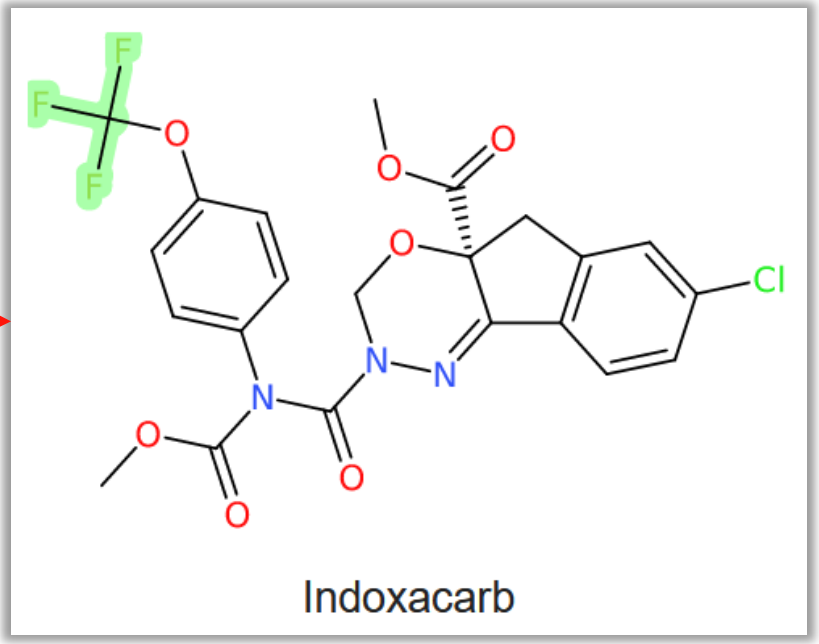
OECD Monograph ENV/CBC/MONO(2021)25 (9 July 2021)

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

Over 6 million OECD PFAS in PubChem ?!?!?!

- PFAS and Fluorinated Compounds in PubChem ? ↗ 21,411,181
  - OECD PFAS definition ? ↗ 6,540,217
    - Molecule contains isolated CF2 ? 675,776
      - Contains CF2 and larger PFAS parts ? 9,140
      - Contains only isolated CF2 ? 590,062
      - Contains only isolated CF2/CF3 ? 76,574
    - Molecule contains isolated CF3 ? 5,747,364
      - Contains CF3 and larger PFAS parts ? 26,816
      - Contains only isolated CF2/CF3 ? 76,574
      - Contains only isolated CF3 ? 5,643,974
    - Molecule contains PFAS parts larger than CF2/CF3 ? 229,607
      - Breakdown by isolated PFAS part count ? 229,607
      - Breakdown by isolated PFAS part type ? 229,607

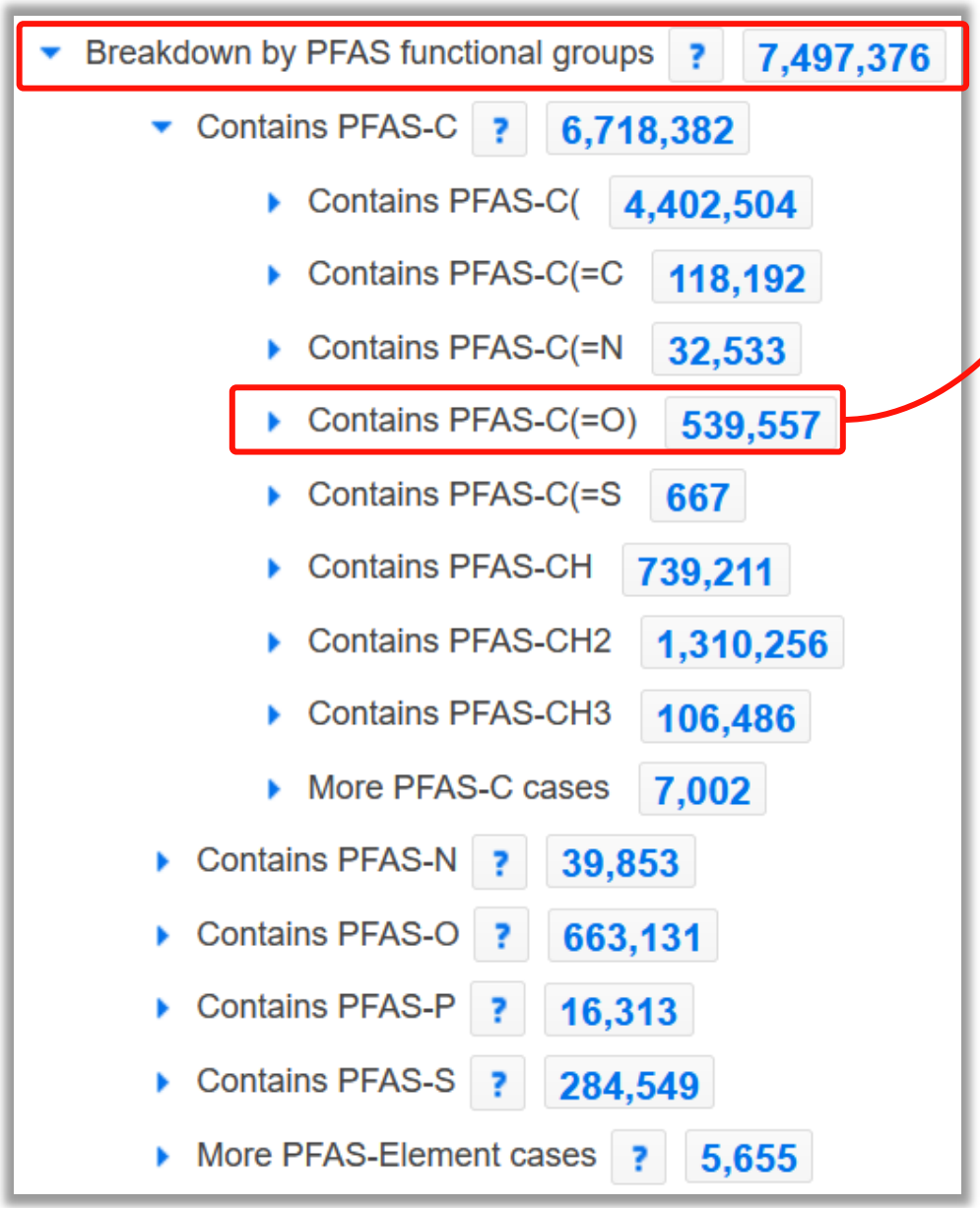
Note: this does not include mixtures and salts …

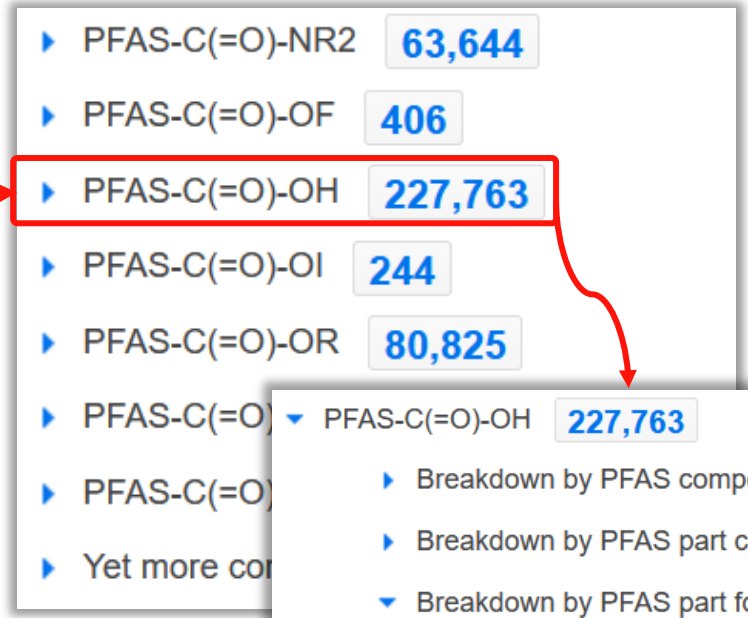


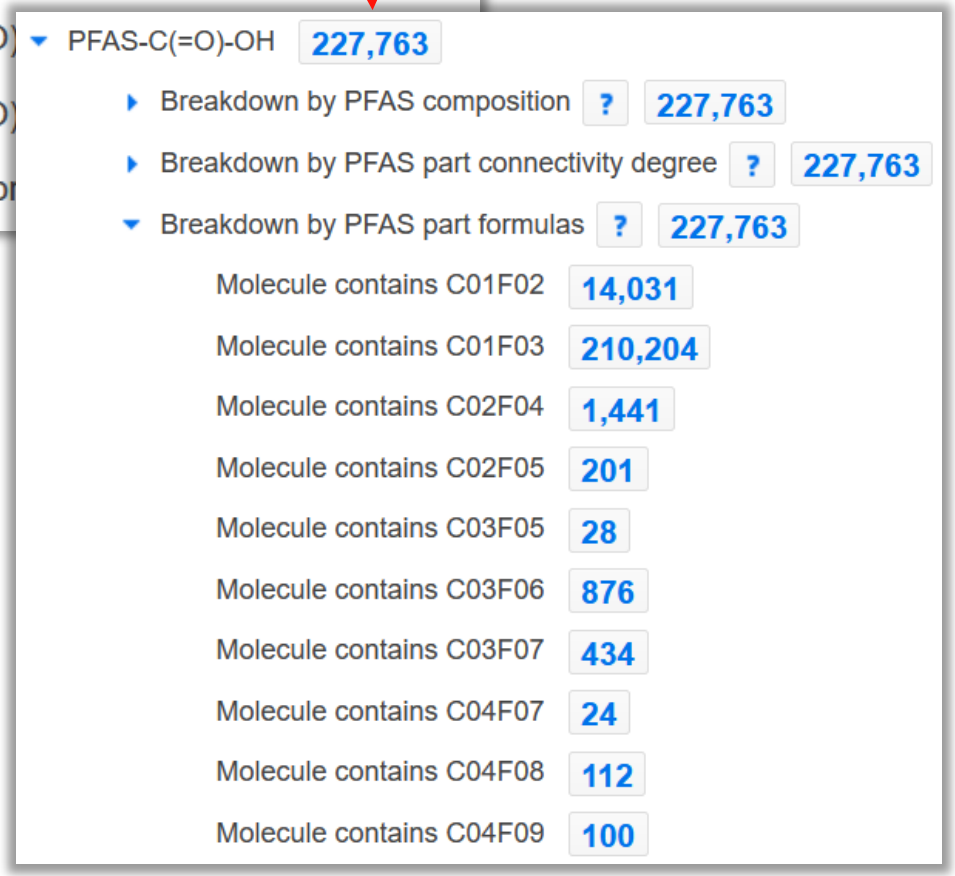
Indoxacarb

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

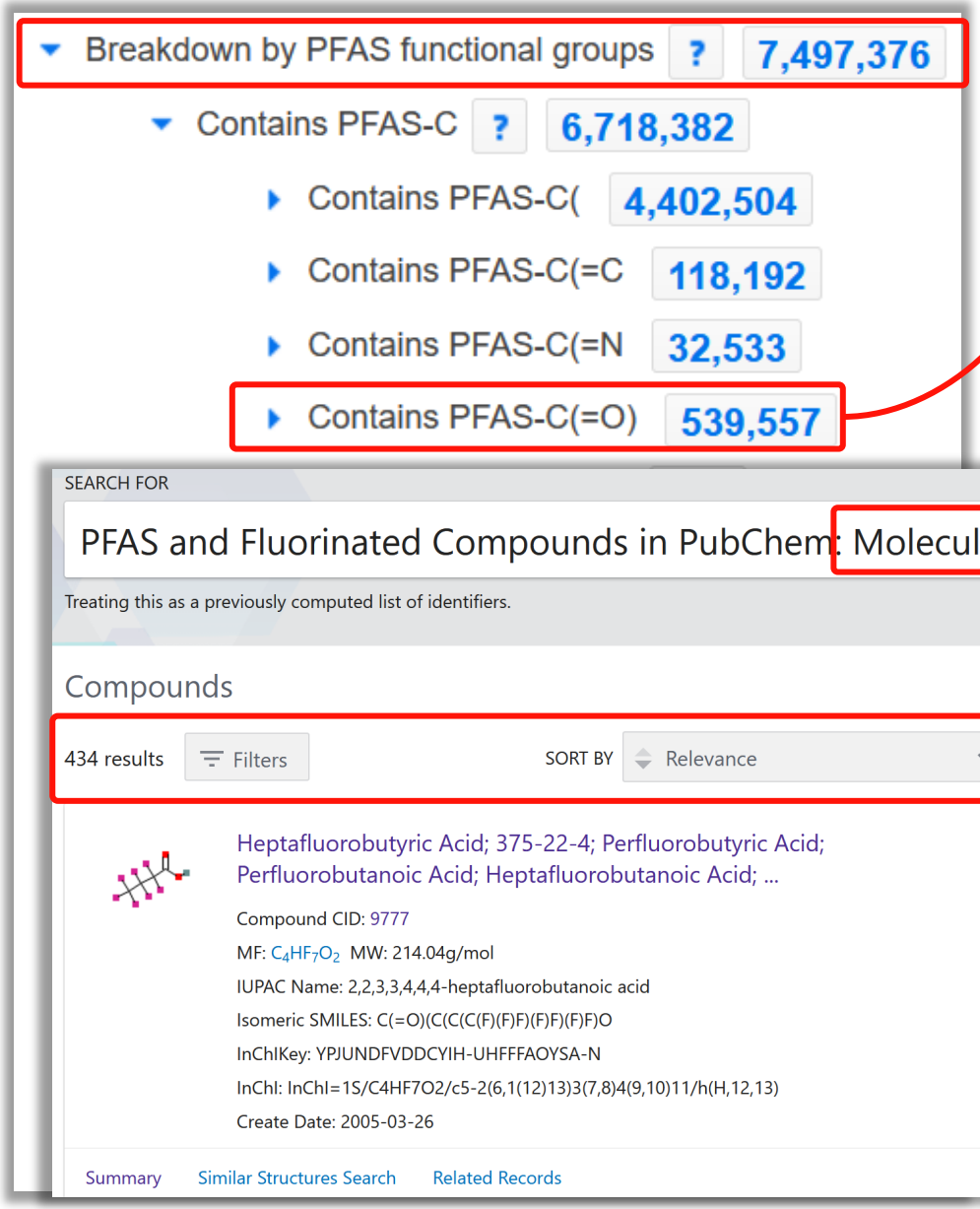
PFAS Breakdown by Chemistry

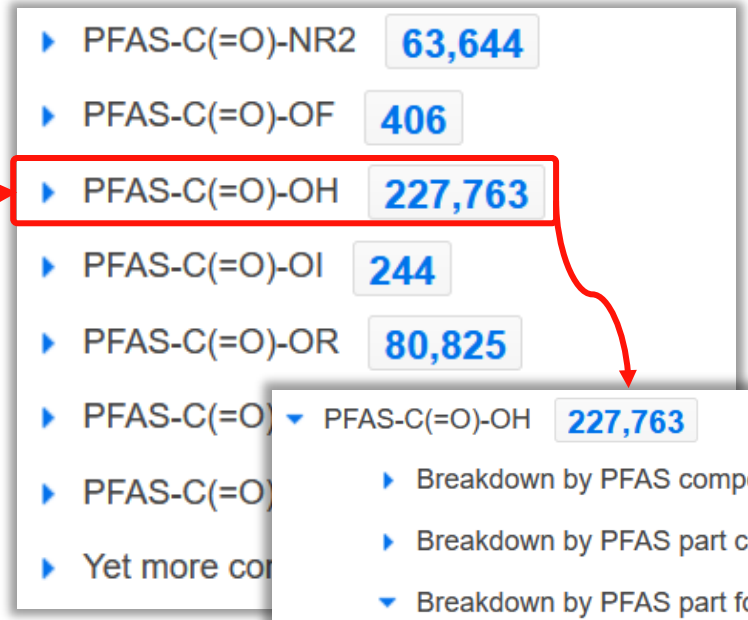


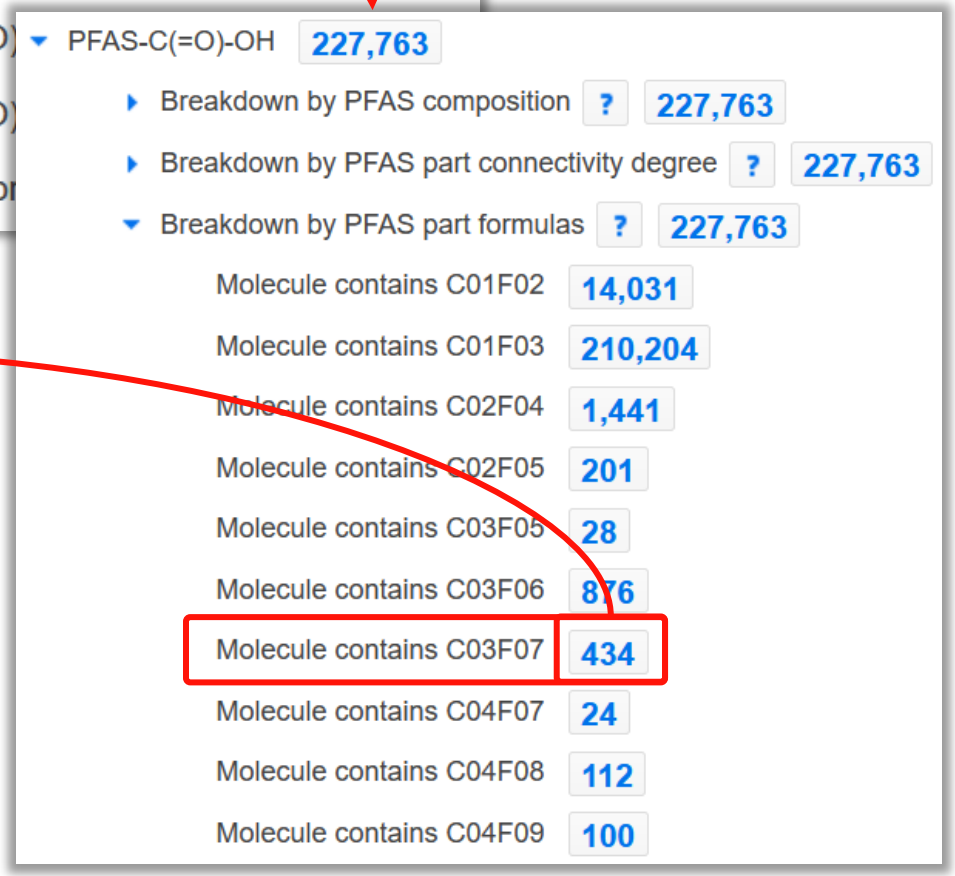




PFAS Breakdown by Chemistry

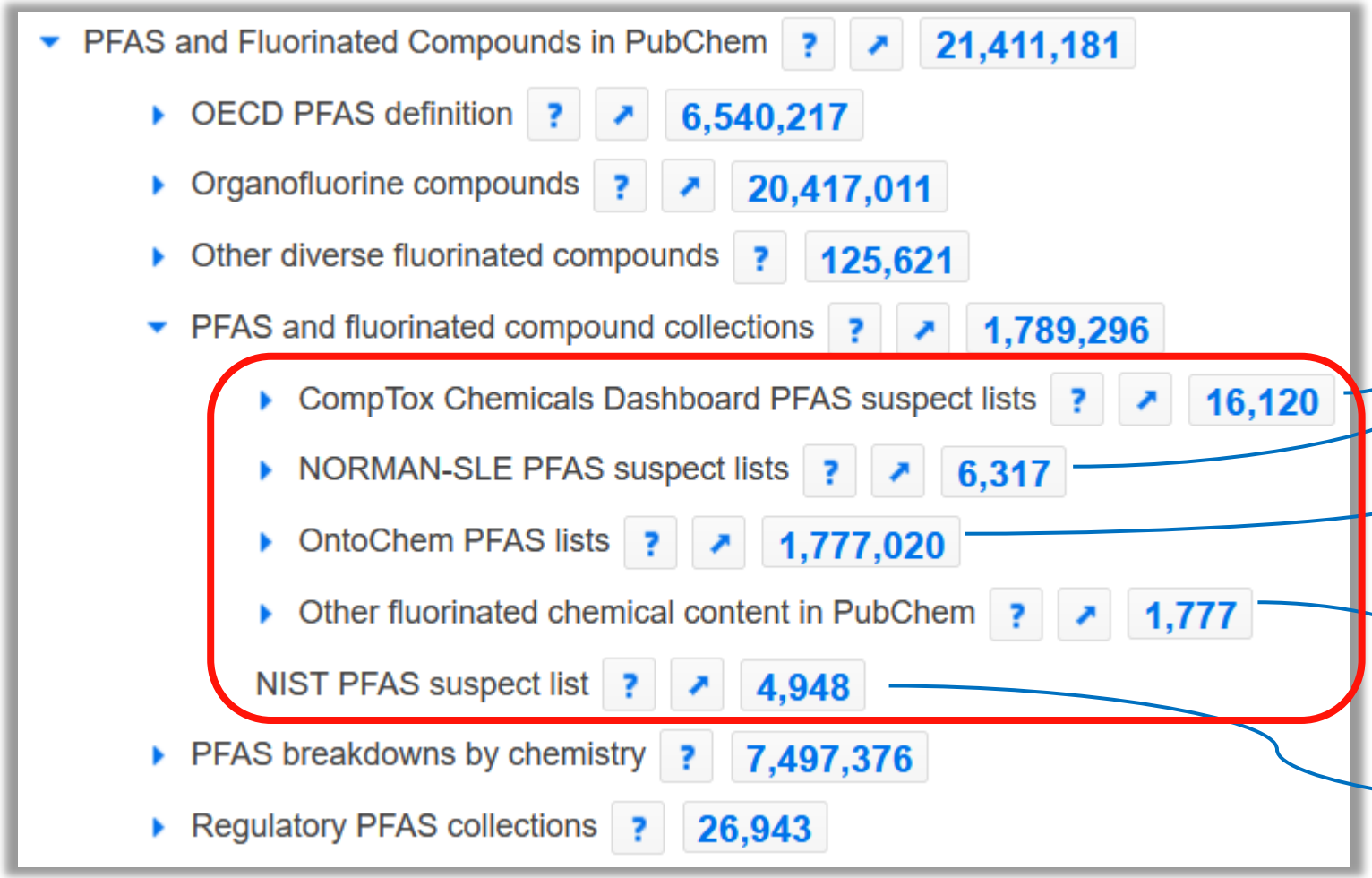






U.S. National Library of Medicine
National Center for Biotechnology Information

The PubChem PFAS Tree – Collection of Suspect Lists



EPA CompTox Chemicals Dashboard

norman suspects

Google ontochem

PubChem

NIST PUBLIC DATA REPOSITORY

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

PFAS Suspect Lists – NORMAN-SLE / NIST

- PFAS and fluorinated compound collections ? ↗ 1,789,296
  - CompTox Chemicals Dashboard PFAS suspect lists ? ↗ 16,120
  - NORMAN-SLE PFAS suspect lists ? ↗ 6,317
    - S09 | PFASTRIER | PFAS Suspect List of fluorinated substances from X. Trier and colleagues ? ↗ 468
    - S14 | KEMIPFAS | PFAS Highly Fluorinated Substances List from KEMI ? ↗ 1,344
    - S25 | OECDPFAS | List of PFAS from the OECD ? ↗ 3,692
    - S46 | PFASNTREV19 | List of PFAS reported in Non-Target HRMS Studies from Liu et al 2019 ? ↗ 680
    - S80 | PFASGLUEGE | Overview of PFAS Uses ? ↗ 1,250
    - S89 | PRORISKPFAS | List of PFAS Compiled from NORMAN SusDat ? ↗ 4,240
    - S92 | FLUOROPHARMA | List of 340 ATC classified fluoro-pharmaceuticals ? ↗ 290
    - S94 | FLUOROPEST | List of 423 FRAC/HRAC/IRAC classified fluoro-agrochemicals ? ↗ 318
    - S95 | PFASANEXCH | PFAS List from the NORMAN PFAS Analytical Exchange Activity ? ↗ 94
    - S96 | ECIPFAS | Updateable List to add PFAS Structures to Public Resources from ECI (UniLu) ? ↗ 257
    - S100 | PFASREACH | List of PFAS identified in REACH 2019 ? ↗ 429
    - S102 | PARCPFAS | List of PFAS from PARC WP4 ? ↗ 190
  - OntoChem PFAS lists ? ↗ 1,777,020
  - Other fluorinated chemical content in PubChem ? ↗ 1,777
  - NIST PFAS suspect list ? ↗ 4,948

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

RESEARCH Open Access

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

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

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

https://www.nist.gov/people/benjamin-place

1,232 new CIDs!

PFAS Suspect Lists – OntoChem+Google / PubChem

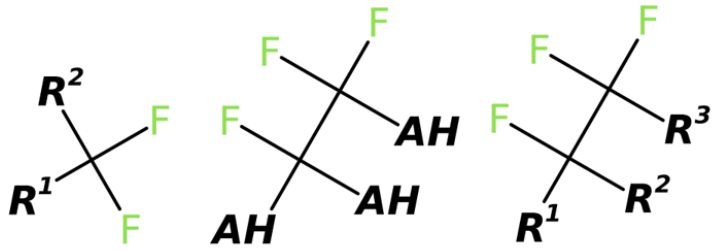
Digital Discovery Webinar

Today's slides

- OntoChem PFAS lists: 1,777,020
  - OntoChem PFAS from CORE - Definition A: 26,805
  - OntoChem PFAS from CORE - Definition B: 4,115
  - OntoChem PFAS from CORE - Definition C: 3,433
  - OntoChem PFAS from Google Patents - Definition A: 1,762,939
  - OntoChem PFAS from Google Patents - Definition B: 73,744
  - OntoChem PFAS from Google Patents - Definition C: 33,648
- Other fluorinated chemical content in PubChem: 1,777
  - MeSH: Fluorinated Hydrocarbons: 417
    - MeSH: Chlorofluorocarbons: 39
    - MeSH: Fluoroacetates: 30
    - MeSH: Fluorobenzenes: 104
    - MeSH: Fluorocarbons: 121
  - CAMEO Chemicals: Fluorinated Organic Compounds: 120
  - ChEBI: Organofluorine Compound: 1,372

Google ontochem

PubChem



# PFAS in:	Definition A	Definition B	Definition C
CORE Documents	27,958	4,139	3,457
Google Patents	1,783,651	75,108	34,197
Total	**1,797,831**	**77,441**	**36,788**

Barnabas *et al.* (2022) *Digital Discovery*.
DOI: 10.1039/D2DD00019A
More info: DOI: 10.5281/zenodo.7185579

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

PFAS Suspect Lists – CompTox – 42(!!!) PFAS Lists

CompTox Chemicals Dashboard

- CompTox Chemicals Dashboard PFAS suspect lists 16,132
  - [CCL5PFAS] WATER|EPA: Chemical Contaminants - CCL 5 PFAS subset 10,218
  - [EPAPFAS75S1] PFAS|EPA: List of 75 Test Samples (Set 1) 74
  - [EPAPFAS75S2] PFAS|EPA: List of 75 Test Samples (Set 2) 76
  - [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 427
  - [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

- [PFASINVITRO] PFAS|EPA: List of chemicals tested in in vitro methods 2019-2020 182
- [PFASKEMI] PFAS: List from the Swedish Chemicals Agency (KEMI) Report 1,499
- [PFASLCMSGCMS] PFAS: Collection of GC-MS and LC-MS standards: Food Contact Materials 38
- [PFASMASTER] PFAS Master List of PFAS Substances (Version 2) 10,740
- [PFASMASTERLISTV2] PFAS: V2 PFAS Master List of PFAS Substances 6,872
- [PFASNORDIC] PFAS: Nordic PFAS Report 2019 202
- [PFASNTREV19] PFAS: PFAS in Non-Target HRMS Studies (Liu et al 2019) 127
- [PFASOECD] PFAS: Listed in OECD Global Database 3,722
- [PFASOECDNA] NORMAN: List of PFAS from the OECD Curated by Nikiforos Alygizakis 3,203
- [PFASPACKAGING] PFAS|EPA PFAS Substances in Pesticide Packaging 8
- [PFASSTRUCT] Navigation Panel to PFAS Structure Lists 14,701
- [PFASSTRUCTV1] PFAS|EPA: PFAS structures in DSSTox (update March 2018) 4,333
- [PFASSTRUCTV2] PFAS|EPA: PFAS structures in DSSTox (update November 2019) 6,614
- [PFASSTRUCTV3] PFAS|EPA: PFAS structures in DSSTox (update August 2020) 8,121
- [PFASSTRUCTv4] PFAS|EPA: PFAS structures in DSSTox (update August 2021) 10,739
- [PFASSTRUCTV5] PFAS|EPA: PFAS structures in DSSTox (update August 2022) 14,701
- [PFASTDB] WATER|PFAS: PFAS Chemicals contained in the EPA Drinking Water Treatability Database 38
- [PFASTOXDB] PFAS: PFAS-Tox Database 43
- [PFASTRI] PFAS: PFAS to the Toxics Release Inventory (TRI) Program by the National Defense Authorization Act 98
- [PFASTRIER] PFAS Community-Compiled List (Trier et al. 2015) 592
- [PRORISKPFAS] NORMAN|List of PFAS Compiled from NORMAN-SusDat 3,360

Assembly and Curation of Lists of Per- and Polyfluoroalkyl Substances (PFAS) to Support Environmental Science Research


Antony J. Williams[1]*, Linda G. T. Gaines[2], Christopher M. Grulke[1†],
Charles N. Lowe[1], Gabriel F. B. Sinclair[3], Vicente Samano[4], Inthirany Thillainadarajah[4], Bryan Meyer[4], Grace Patlewicz[1] and 
Ann M. Richard[1]

Williams *et al.* (2022) DOI:10.3389/fenvs.2022.850019

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

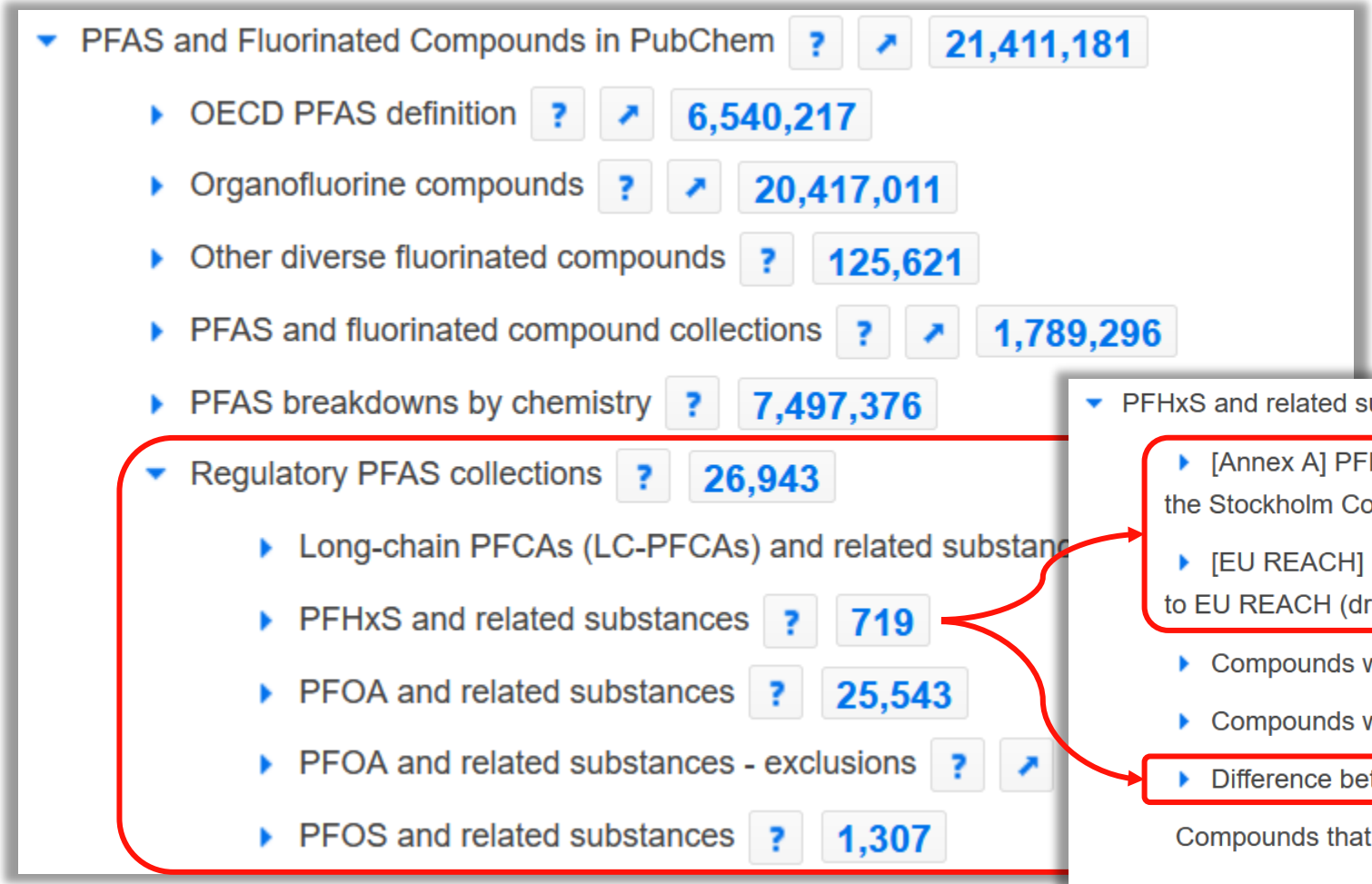
NIH U.S. National Library of Medicine
National Center for Biotechnology Information

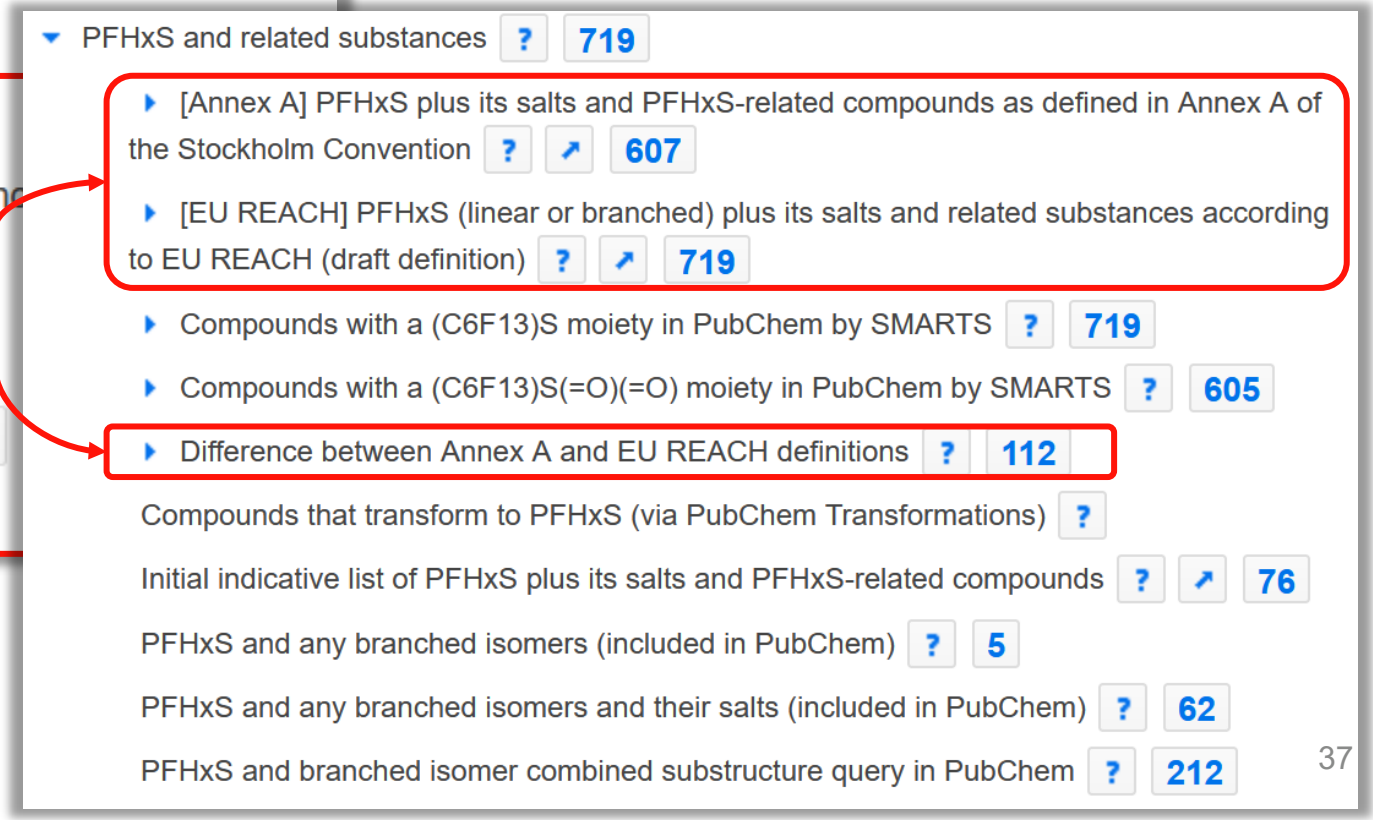
LCSB

Adding Regulatory Collections to the PFAS Tree

Today's slides zenodo

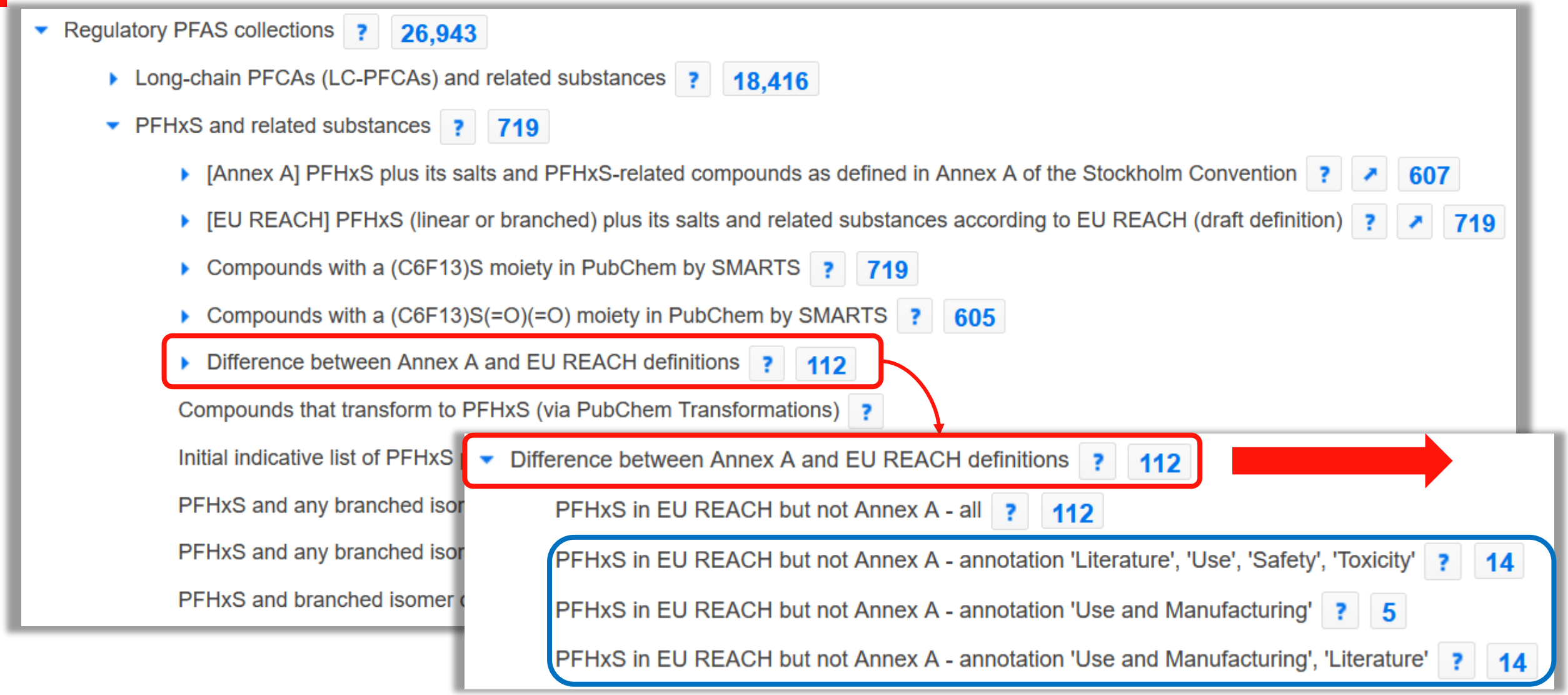
POPRC slides available at
DOI: 10.5281/zenodo.7118551





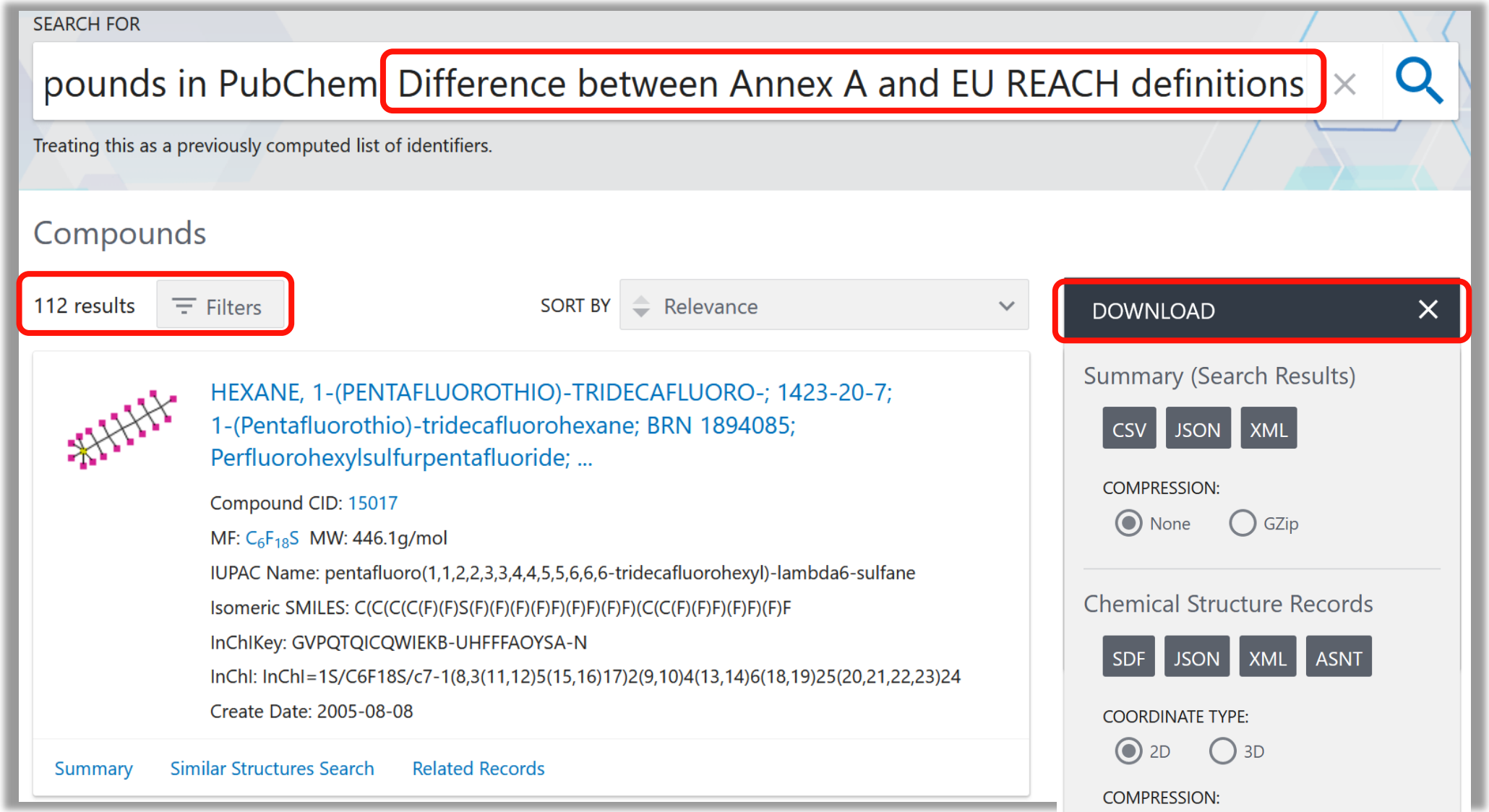
Thanks to Andreas Buser, FOEN, CH
https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120.

Example: PFHxS in Stockholm Convention vs EU REACH



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

PFHxS cont.: How can you find what is relevant for you?



Download files contain additional information…

Basic properties – name, identifiers, SMILES, XlogP, mass, formula etc.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	cid	cmpdname	cmpdsyno	mw	mf	polararea	complexity	xlogp	heavycnt	hbonddon	hbondacc	rotbonds	inchi	isosmiles	canonicals	inchikey	iupacname	exactmass	monoisotc
2	15017	Hexane, 1-	HEXANE, 1	446.1	C6F18S	1	530	8.2	25	0	18	4	InChI=1S/C	C(C(C(C(F)	C(C(C(C(F)	GVPQTQIC	pentafluor	445.943	445.943
3	2827766	4-methyl-I	NULL	597.4	C19H12F13	79.8	895	7.9	37	1	17	9	InChI=1S/C	CC1=CC=C(	CC1=CC=C(	XVAIFAXE	4-methyl-I	597.01	597.01
4	2827776	4,6-dichlor	NULL	591.2	C15H5Cl2F	76	732	8.8	35	1	18	8	InChI=1S/C	C1=CC=C(C	C1=CC=C(C	GAUUMFM	4,6-dichlor	589.94	589.94
5	2827778	6-chloro-2-	NULL	571.7	C15H7ClF1	102	741	7.5	35	2	19	8	InChI=1S/C	C1=CC=C(C	C1=CC=C(C	XICRCGPU	6-chloro-2-	570.99	570.99
6	2827780	2-N,2-N-di	NULL	580.4	C17H13F13	105	781	7	37	2	20	9	InChI=1S/C	CN(C)C1=N	CN(C)C1=N	KBFCJWZW	2-N,2-N-di	580.071	580.071
7	3021589	Tridecaflu	Tridecaflu	386.56	C6ClF13S	25.3	385	5.8	21	0	14	4	InChI=1S/C	C(C(C(C(F)	C(C(C(C(F)	BEDWVNV	1,1,2,2,3,3,	385.92	385.92
8	3023059	Bis(trideca	Bis(trideca	702.2	C12F26S2	50.6	832	10	40	0	28	11	InChI=1S/C	C(C(C(C(F)	C(C(C(C(F)	OCCOAYP	1,1,1,2,2,3,	701.903	701.903
9	10092845	Butyl(tride	SCHEMBL9	408.22	C10H9F13S	25.3	425	6.7	24	0	14	8	InChI=1S/C	CCCCSC(C(	CCCCSC(C(	UWLAEJRV	1-butylsulf	408.022	408.022
10	10971025	Tridecaflu	Tridecaflu	384.12	C6HF13O2S	56.5	452	3.9	22	1	16	5	InChI=1S/C	C(C(C(C(F)	C(C(C(C(F)	IYYLVCKN	1,1,2,2,3,3,	383.949	383.949
11	11048550	1-Fluoro-4	SCHEMBL9	446.2	C12H4F14S	25.3	517	7.1	27	0	15	6	InChI=1S/C	C1=CC(=CC	C1=CC(=CC	GUZURVM	1-fluoro-4-	445.981	445.981
12	11245889	1-Ethenyls	NULL	378.16	C8H3F13S	25.3	420	5.9	22	0	14	6	InChI=1S/C	C=CSC(C(C	C=CSC(C(C	ICKCZVFD	1-ethenyls	377.975	377.975
13	11327892	1-(2-Chlor	NULL	430.61	C8H4ClF13	36.3	485	4.4	24	0	15	7	InChI=1S/C	C(CCl)S(=C	C(CCl)S(=C	RMJKQZOT	1-(2-chlorc	429.946	429.946
14	11338428	2-(1,1,2,2,3	NULL	396.17	C8H5F13OS	45.5	415	4.8	23	1	15	7	InChI=1S/C	C(CSC(C(C(	C(CSC(C(C(	XXEIJTYHG	2-(1,1,2,2,3	395.985	395.985
15	11362001	1-(2-Chlor	NULL	414.61	C8H4ClF13	25.3	415	6	23	0	14	7	InChI=1S/C	C(CCl)SC(C	C(CCl)SC(C	NRIXEAPP	1-(2-chlorc	413.951	413.951
16	11990355	Copper;4,4	NULL	2822.7	C64H44CuF	238	3300	NULL	167	0	102	48	InChI=1S/C	C1CN(CCN	C1CN(CCN	BWMAOKV	copper;4,4	2821	2821
17	11990356	1,4,8,11-Te	1,4,8,11-Te	1776.9	C42H36F52	158	2660	15.1	106	0	64	32	InChI=1S/C	C1CN(CCN	C1CN(CCN	WXIBLQLL	1,4,8,11-te	1776.08	1776.08
18	12635301	1-Chloro-4	NULL	462.66	C12H4ClF1	25.3	522	7.6	27	0	14	6	InChI=1S/C	C1=CC(=CC	C1=CC(=CC	XOQJHQBC	1-chloro-4-	461.951	461.951
19	12996310	[(1beta,4b	[(1beta,4b	460.26	C13H9F13C	36.3	680	5.1	28	0	15	6	InChI=1S/C	C1[C@@H]	C1C2CC(C1	SFIGAKQS	(1R,4R,5R)-	460.017	460.017
20	12996311	(1R,4R,5S)-	NULL	460.26	C13H9F13C	36.3	680	5.1	28	0	15	6	InChI=1S/C	C1[C@@H]	C1C2CC(C1	SFIGAKQS	(1R,4R,5S)-	460.017	460.017
21	13213429	[(Tridecafl	[(Tridecafl	428.21	C12H5F13S	25.3	490	7	26	0	14	6	InChI=1S/C	C1=CC=C(C	C1=CC=C(C	OOARTWB	1,1,2,2,3,3,	427.99	427.99
22	13410186	(4-Methylp	(4-Methyl	442.24	C13H7F13S	25.3	517	7.4	27	0	14	6	InChI=1S/C	CC1=CC=C(	CC1=CC=C(	MBZVAVG	1-methyl-4	442.006	442.006
23	14544789	1,1,1,2,2,3,	SCHEMBL9	366.14	C7H3F13S	25.3	381	5.4	21	0	14	5	InChI=1S/C	CSC(C(C(C	CSC(C(C(C	PXEUPPDZ	1,1,1,2,2,3	365.975	365.975

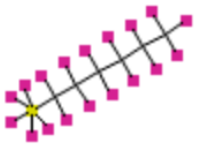
PubChem PFAS Tree: https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120
Docs: https://gitlab.lcsb.uni.lu/eci/pubchem-docs/-/raw/main/pfas-tree/PFAS_Tree.pdf?inline=false

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

Download files contain additional information…

Basic properties – name, identifiers, SMILES, XlogP, mass, formula etc.

cid	cmpdname	cmpdsyno	mw	mf	polararea	complexity	xlogp	heavycnt	hbonddon	hbondacc	rotbonds	inchi	isosmiles	canonicals	inchikey	iupacname	exactmass	monoisotc
15017	Hexane, 1-	HEXANE, 1	446.1	C6F18S	1	530	8.2	25	0	18	4	InChI=1S/C	C(C(C(C(F)	C(C(C(C(F)	GVPQTQIC	pentafluor	445.943	445.943
2827766	4-methyl-I	NULL	597.4	C19H12F13	79.8	895	7.9	37	1	17	9	InChI=1S/C	CC1=CC=C(	CC1=CC=C(	XVAIFAXE	4-methyl-I	597.01	597.01
2827776	4,6-dichlor	N													UMFM	4,6-dichlor	589.94	589.94
2827778	6-chloro-2-	N													RCGPU	6-chloro-2-	570.99	570.99
2827780	2-N,2-N-di	N													CJWZW	2-N,2-N-di	580.071	580.071
3021589	Tridecafluc	T													WVNV	1,1,2,2,3,3,	385.92	385.92
3023059	Bis(trideca	B													OAYP	1,1,1,2,2,3,	701.903	701.903
10092845	Butyl(tride	S													LAEJRV	1-butylsulf	408.022	408.022
10971025	Tridecafluc	T														,3,	383.949	383.949
11048550	1-Fluoro-4	S														-4-	445.981	445.981
11245889	1-Ethenyls	N														yls	377.975	377.975
11327892	1-(2-Chlor	N														orc	429.946	429.946
11338428	2-(1,1,2,2,3	N														2,3	395.985	395.985
11362001	1-(2-Chlor	N														orc	413.951	413.951
11990355	Copper;4,4	N														4,4	2821	2821
11990356	1,4,8,11-Te	1														-te	1776.08	1776.08
12635301	1-Chloro-4	N														-4-	461.951	461.951
12996310	[(1beta,4b	[(														R)-	460.017	460.017
12996311	(1R,4R,5S)-	N														S)-	460.017	460.017
13213429	[(Tridecafl	[(														,3,	427.99	427.99
13410186	(4-Methylp	(														yl-	442.006	442.006
14544789	1,1,1,2,2,3,	S														2,3	365.975	365.975



HEXANE, 1-(PENTAFLUOROTHIO)-TRIDECAFLUORO-; 1423-20-7; 1-(Pentafluorothio)-Tridecafluorohexane; BRN 1894085; Perfluorohexylsulfurpentafluoride; ...

Compound CID: 15017

MF: $C_6F_{18}S$ MW: 446.1g/mol

IUPAC Name: pentafluoro(1,1,2,2,3,3,4,4,5,5,6,6,6-tr

Isomeric SMILES: C(C(C(C(F)(F)S(F)(F)(F)(F)F)(F)F)(F)F)F

InChIKey: GVPQTQICQWIEKB-UHFFFAOYSA-N

InChI: InChI=1S/C6F18S/c7-1(8,3(11,12)5(15,16)17)

Create Date: 2005-08-08

Summary Similar Structures Search Related Records

Property Name	Property Value
Molecular Weight	446.10 g/mol
XLogP3-AA	8.2
Hydrogen Bond Donor Count	0
Hydrogen Bond Acceptor Count	18
Rotatable Bond Count	4
Exact Mass	445.9433281 g/mol

PubChem PFAS Tree: https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120
Docs: https://gitlab.lcsb.uni.lu/eci/pubchem-docs/-/raw/main/pfas-tree/PFAS_Tree.pdf?inline=false

Download files contain additional information…

Literature and patent counts, source & date, annotation & deposition categories

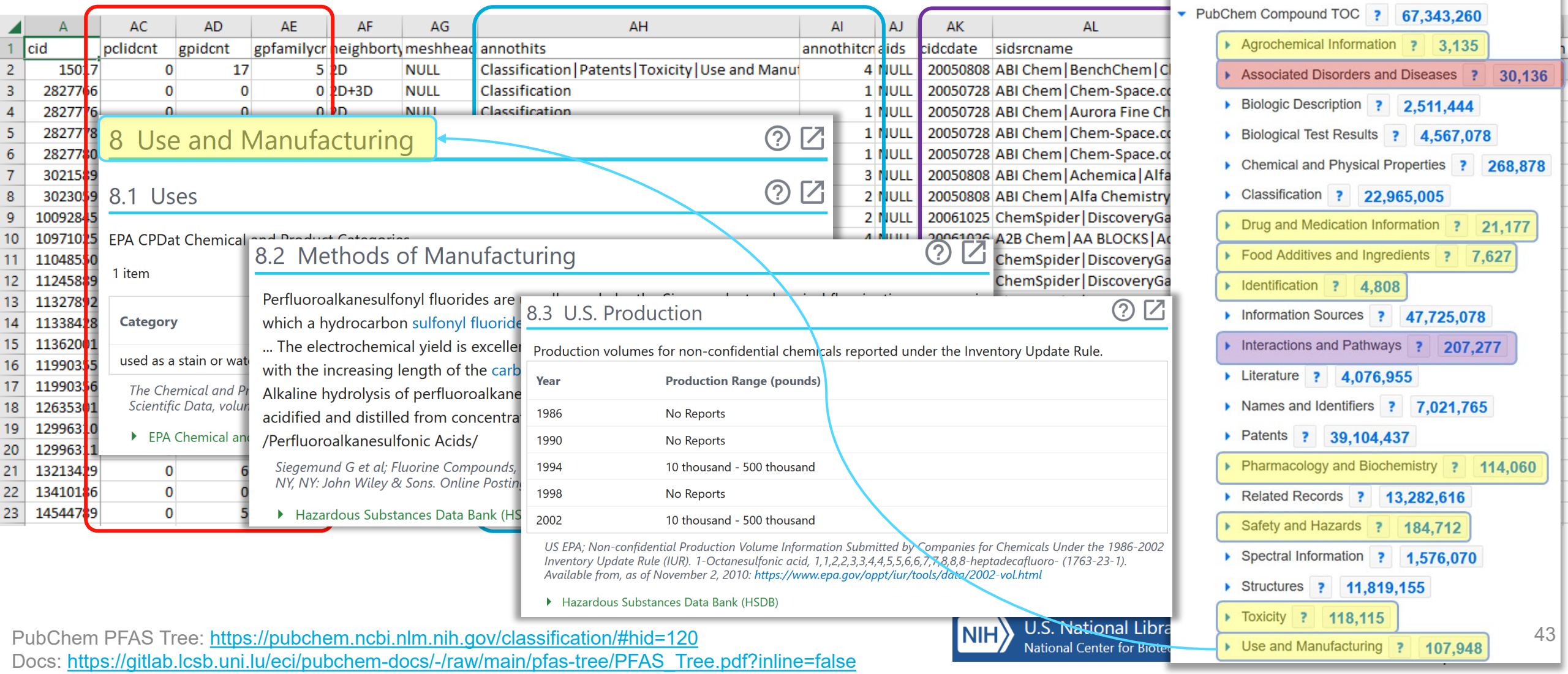
	A	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN
1	cid	pclidcnt	gpidcnt	gpfamilycr	neighborty	meshhead	annothits	annothitcn	aids	cidcdate	sidsrcname	depcatg	annotation
2	150[illegible]7	0	17	5	2D	NULL	Classification|Patents|Toxicity|Use and Manu	4	NULL	20050808	ABI Chem|BenchChem|Che	Chemical Vendors|Curation Efforts|Govern	NULL
3	2827766	0	0	0	2D+3D	NULL	Classification	1	NULL	20050728	ABI Chem|Chem-Space.com	Chemical Vendors|Legacy Depositors|Rese	NULL
4	2827776	0	0	0	2D	NULL	Classification	1	NULL	20050728	ABI Chem|Aurora Fine Chem	Chemical Vendors|Legacy Depositors|Rese	NULL
5	2827778	0	0	0	2D	NULL	Classification	1	NULL	20050728	ABI Chem|Chem-Space.com	Chemical Vendors|Legacy Depositors|Rese	NULL
6	2827780	0	0	0	2D	NULL	Classification	1	NULL	20050728	ABI Chem|Chem-Space.com	Chemical Vendors|Legacy Depositors|Rese	NULL
7	3021589	0	12	3	2D+3D	NULL	Classification|Patents|Use and Manufacturing	3	NULL	20050808	ABI Chem|Achemica|Alfa C	Chemical Vendors|Curation Efforts|Govern	NULL
8	3023059	0	0	0	2D	NULL	Classification|Use and Manufacturing	2	NULL	20050808	ABI Chem|Alfa Chemistry|C	Chemical Vendors|Curation Efforts|Govern	NULL
9	10092845	0	2	1	2D+3D	NULL	Classification|Patents	2	NULL	20061025	ChemSpider|DiscoveryGate	Curation Efforts|Governmental Organizatio	NULL
10	10971025	1	16	6	2D+3D	NULL	Classification|Literature|Patents|Use and Mar	4	NULL	20061026	A2B Chem|AA BLOCKS|Acc	Chemical Vendors|Curation Efforts|Govern	NULL
11	11048560	1	1	1	2D+3D	NULL	Classification|Literature|Patents	3	NULL	20061026	ChemSpider|DiscoveryGate	Curation Efforts|Journal Publishers|Legacy	NULL
12	11245839	0	0	0	3D	NULL	Classification	1	NULL	20061026	ChemSpider|DiscoveryGate	Journal Publishers|Legacy Depositors|Rese	NULL
13	11327892	0	0	0	2D+3D	NULL	Classification	1	NULL	20061026	ChemSpider|DiscoveryGate	Journal Publishers|Legacy Depositors|Rese	NULL
14	11338428	0	0	0	2D+3D	NULL	Classification	1	NULL	20061026	ChemSpider|DiscoveryGate	Journal Publishers|Legacy Depositors|Rese	NULL
15	11362001	0	0	0	2D+3D	NULL	Classification	1	NULL	20061026	ChemSpider|DiscoveryGate	Journal Publishers|Legacy Depositors|Rese	NULL
16	11990355	0	0	0	NULL	NULL	Classification	1	NULL	20070205	ChemSpider|DiscoveryGate	Journal Publishers|Legacy Depositors|Rese	NULL
17	11990356	0	0	0	2D	NULL	Classification	1	NULL	20070205	ChemSpider|DiscoveryGate	Governmental Organizations|Journal Publis	NULL
18	12635301	0	0	0	2D	NULL	Classification|Patents	2	NULL	20070208	PATENTSCOPE (WIPO)	Governmental Organizations	NULL
19	12996310	0	0	0	2D+3D	NULL	Classification	1	NULL	20070208	ChemSpider|Japan Chemica	Governmental Organizations|Legacy Depos	NULL
20	12996311	0	0	0	2D+3D	NULL	Classification	1	NULL	20070208	ChemSpider	Legacy Depositors|Research and Developm	NULL
21	13213429	0	6	1	2D+3D	NULL	Classification|Patents|Use and Manufacturing	3	NULL	20070208	ChemTik|DiscoveryGate|EC	Chemical Vendors|Curation Efforts|Govern	NULL
22	13410186	0	0	0	2D+3D	NULL	Classification	2	NULL	20070208	Japan Chemical Substance D	Governmental Organizations	NULL
23	14544789	0	5	2	2D+3D	NULL	Classification|Patents	3	NULL	20070209	ECI Group, LCSB, University	Curation Efforts|Research and Developmen	NULL

PubChem PFAS Tree: https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120
Docs: https://gitlab.lcsb.uni.lu/eci/pubchem-docs/-/raw/main/pfas-tree/PFAS_Tree.pdf?inline=false

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

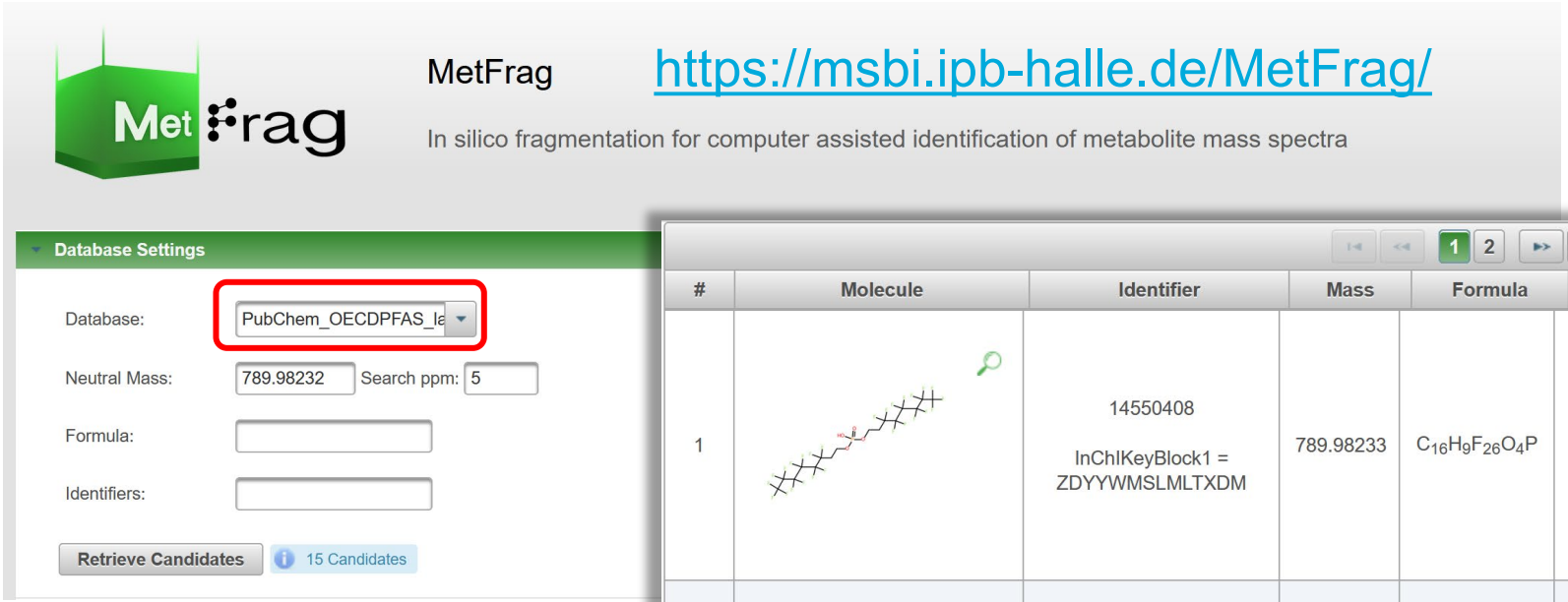
Download files contain additional information…

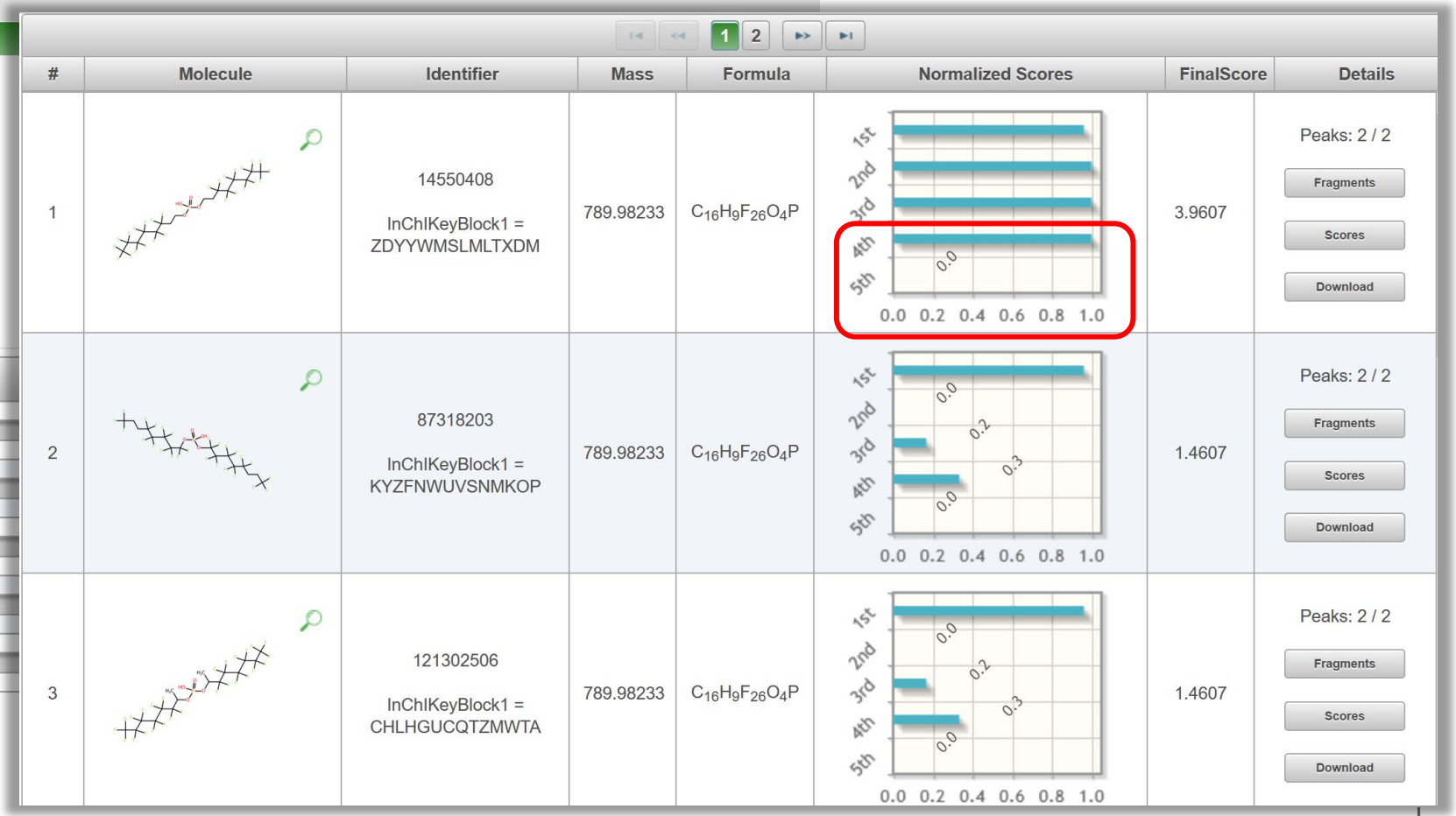
Literature and patent counts, source & date, annotation & deposition categories



PubChem PFAS Tree: https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120
Docs: https://gitlab.lcsb.uni.lu/eci/pubchem-docs/-/raw/main/pfas-tree/PFAS_Tree.pdf?inline=false

Integration of MetFrag and PubChem PFAS Tree





#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1		14550408 InChIKeyBlock1 = ZDYYWMSLMLTXDM	789.98233	$C_{16}H_9F_{26}O_4P$	1st 2nd 3rd 4th 5th 0.0 0.0 0.2 0.4 0.6 0.8 1.0	3.9607	Peaks: 2 / 2 Fragments Scores Download
2		87318203 InChIKeyBlock1 = KYZFNWUVSNMKOP	789.98233	$C_{16}H_9F_{26}O_4P$	1st 2nd 3rd 4th 5th 0.0 0.2 0.3 0.0 0.0 0.2 0.4 0.6 0.8 1.0	1.4607	Peaks: 2 / 2 Fragments Scores Download
3		121302506 InChIKeyBlock1 = CHLHGUCQTZMWTA	789.98233	$C_{16}H_9F_{26}O_4P$	1st 2nd 3rd 4th 5th 0.0 0.2 0.3 0.0 0.0 0.2 0.4 0.6 0.8 1.0	1.4607	Peaks: 2 / 2 Fragments Scores Download

Slides available at DOI:
10.5281/zenodo.6461325

SETAC EUROPE 32ND ANNUAL MEETING
15–19 MAY 2022 | COPENHAGEN, DENMARK + ONLINE

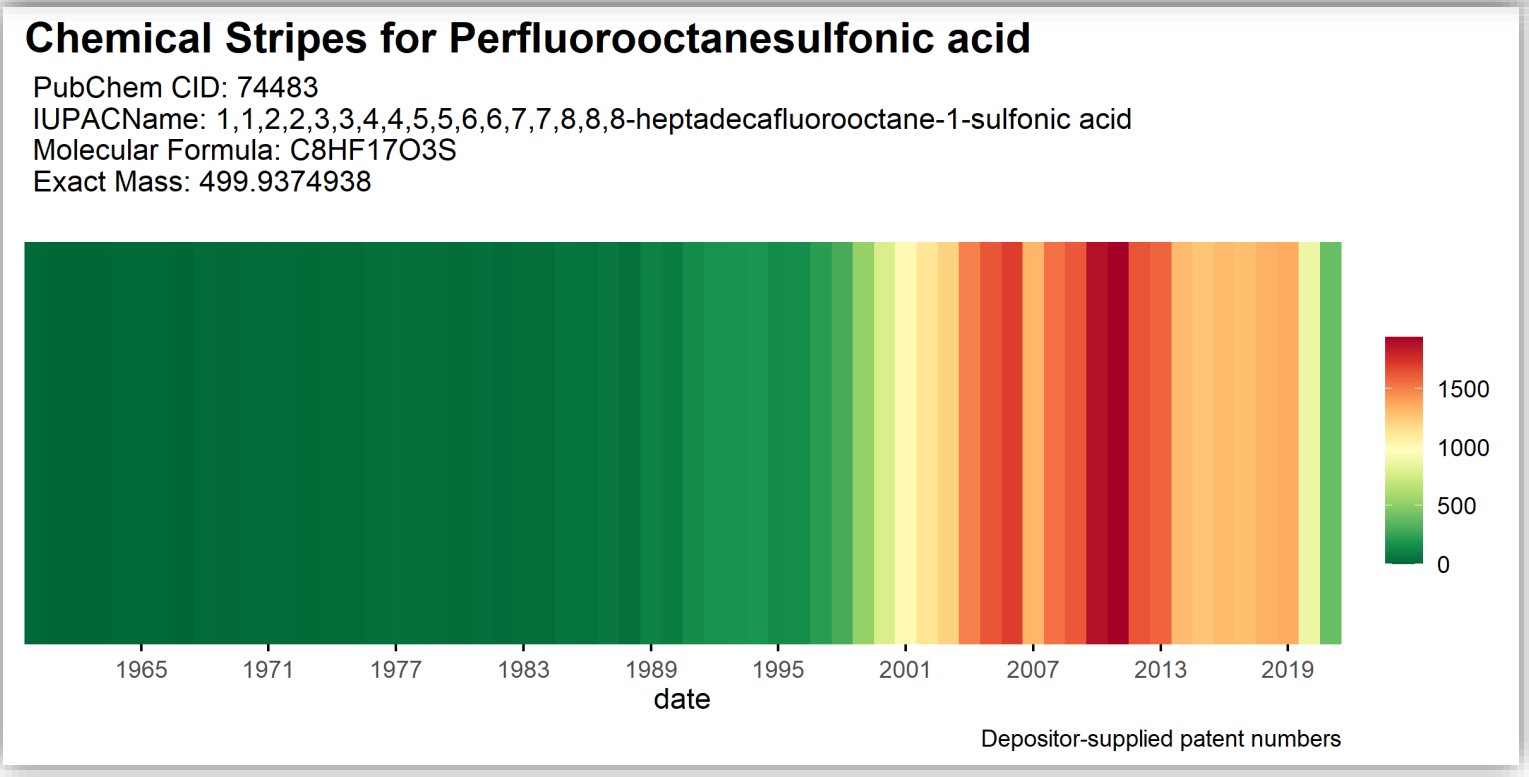
https://msbi.ipb-halle.de/MetFrag/ with
https://massbank.eu/MassBank/RecordDisplay?id=EA292203

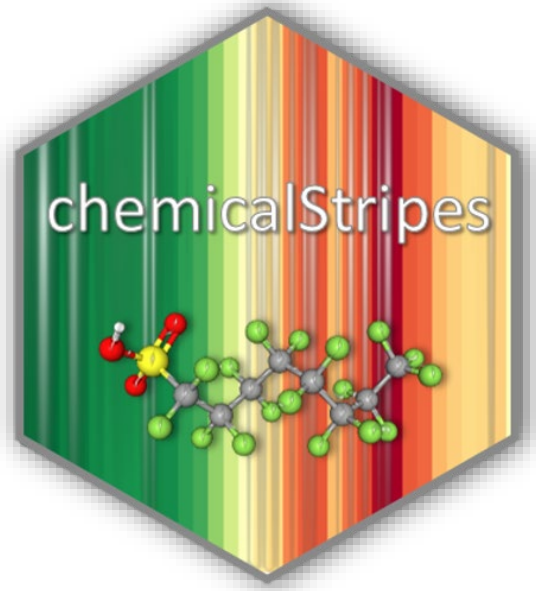
Ruttkies, Schymanski *et al.* (2016) DOI: 10.1186/s13321-016-0115-9

Chemical Stripes in R - for patents & literature

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

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





Aurich *et al* (2023). SETAC Presentation, DOI: 10.5281/zenodo.7885031 and https://gitlab.lcsb.uni.lu/eci/chemicalstripes

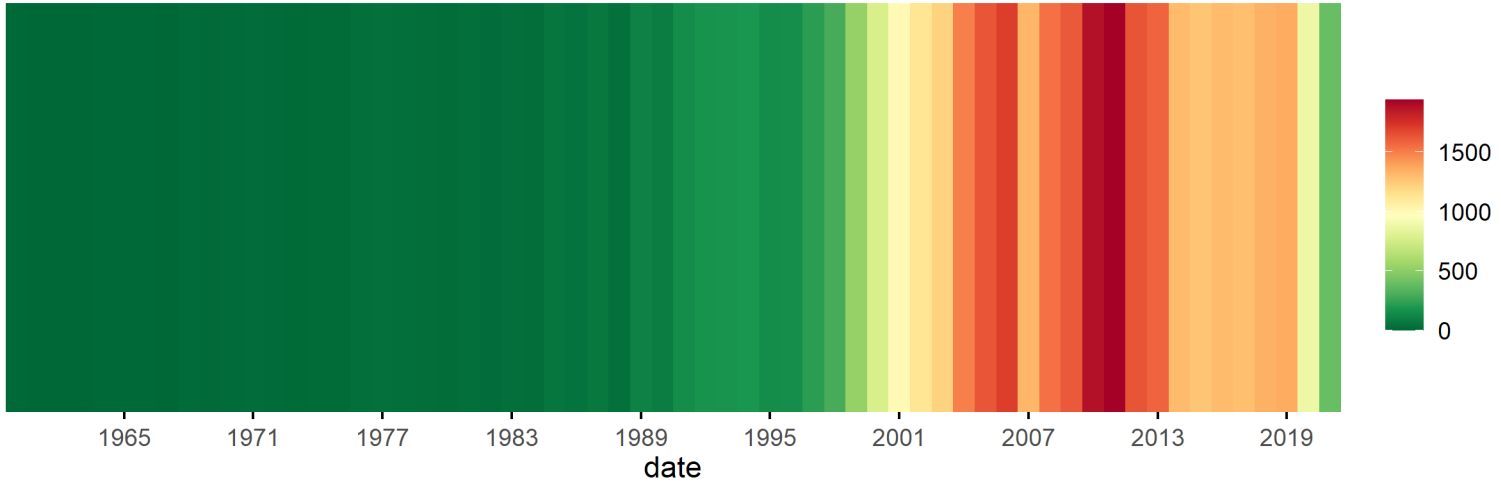
Chemical Stripes in R

Patents & literature

Chemical Stripes for Perfluorooctanesulfonic acid

PubChem CID: 74483
First patent: 1913

Patents

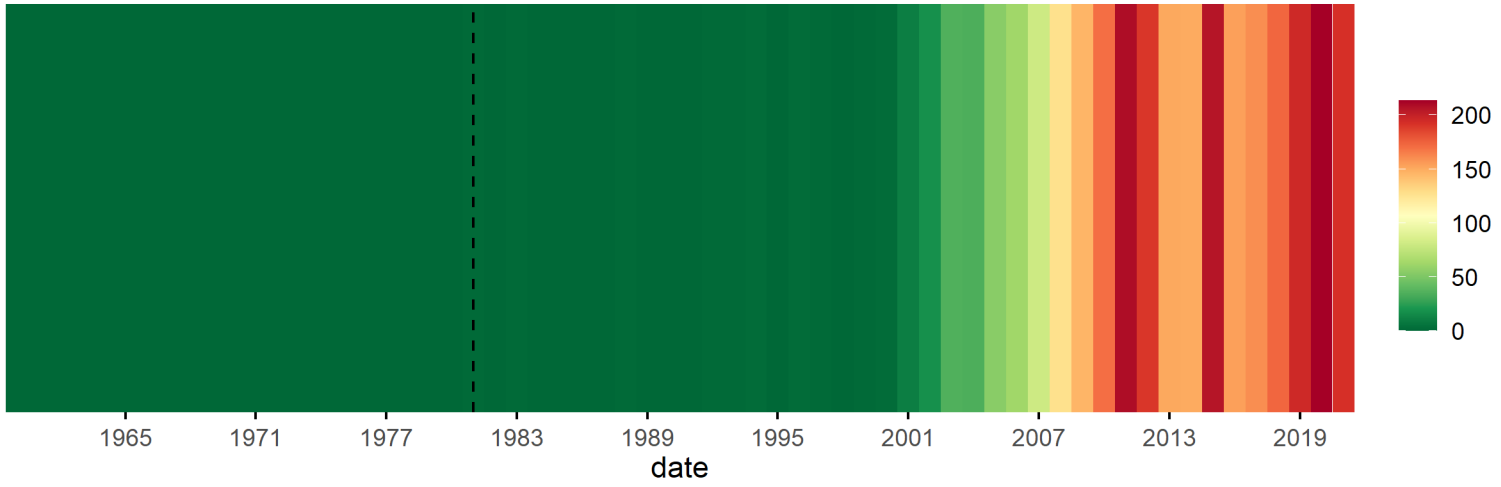


Depositor-supplied patent numbers

Chemical Stripes for Perfluorooctanesulfonic acid

PubChem CID: 74483
First reference: 1981

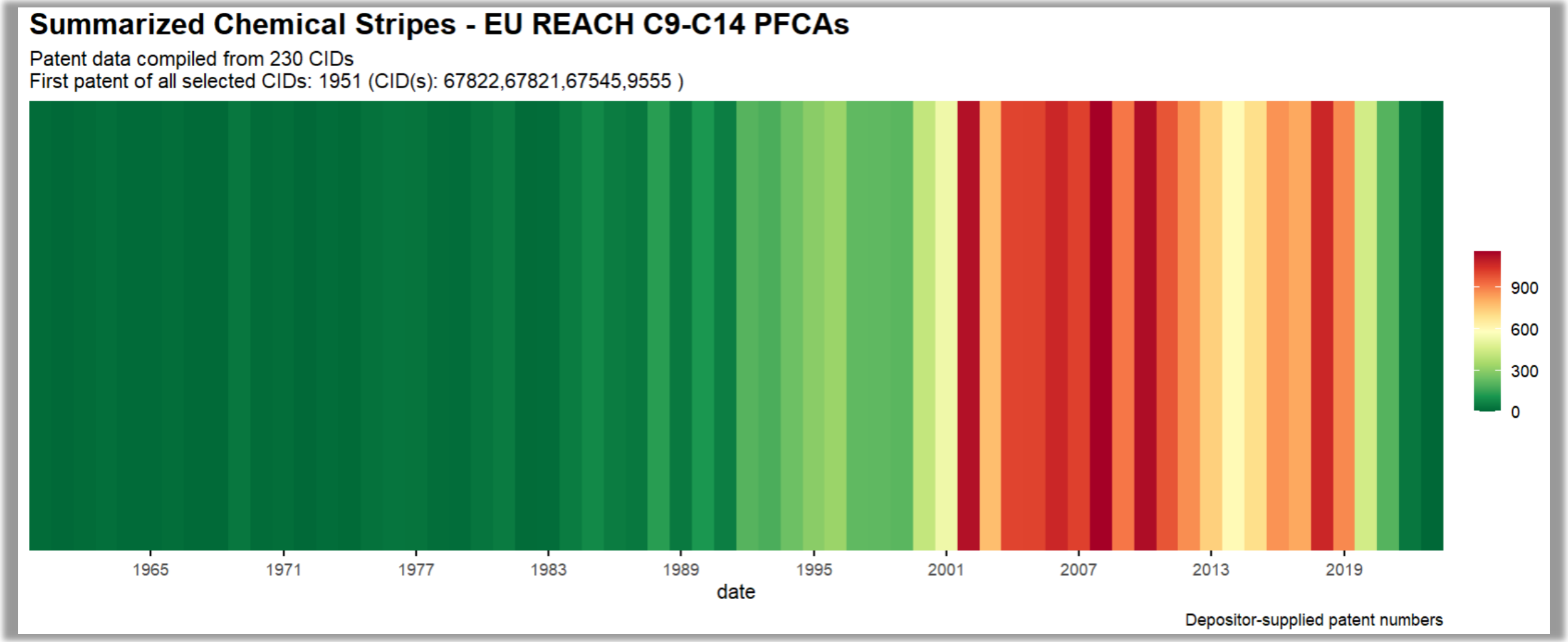
Literature



Consolidated reference numbers

Schymanski *et al.*(2023) DOI: 10.26434/chemrxiv-2023-j823z
https://gitlab.lcsb.uni.lu/eci/chemicalstripes

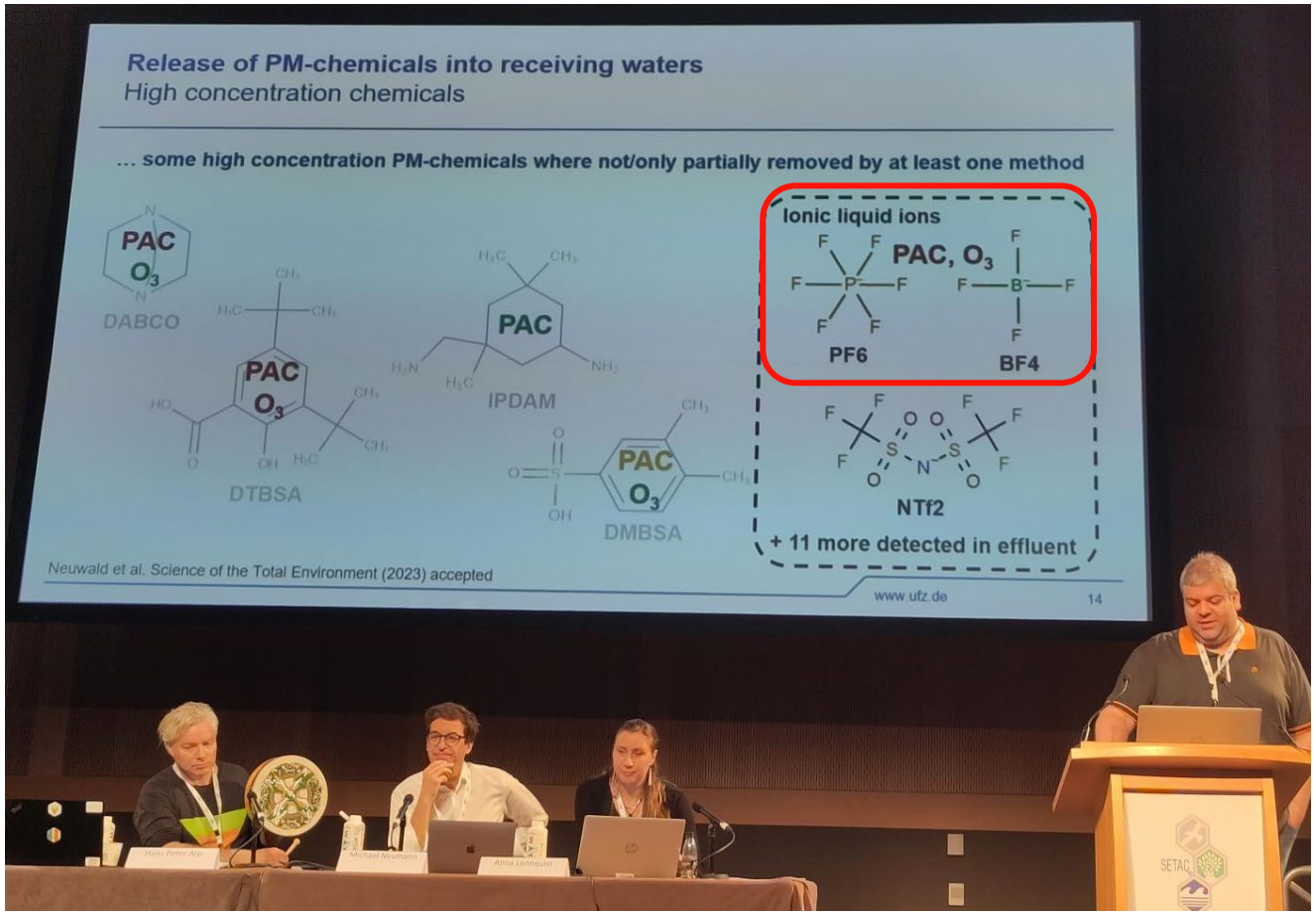
“Summarized” Chemical Stripes in R

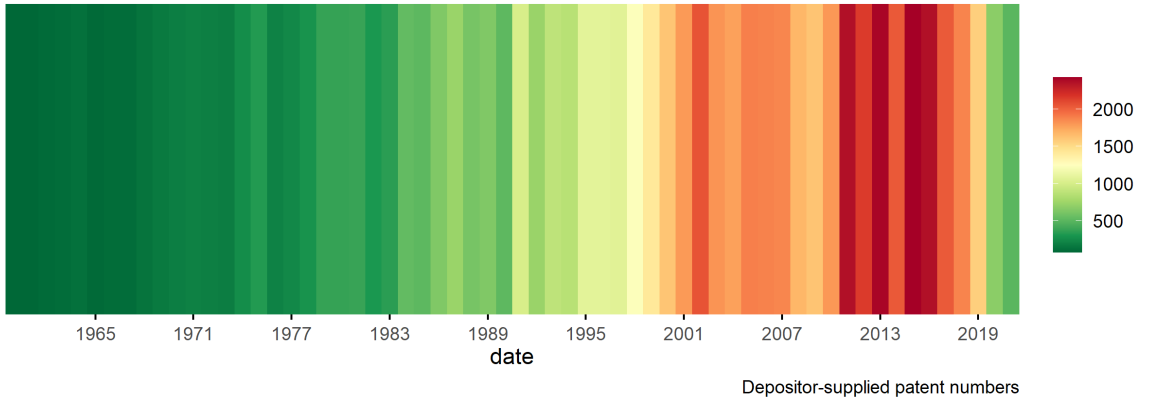


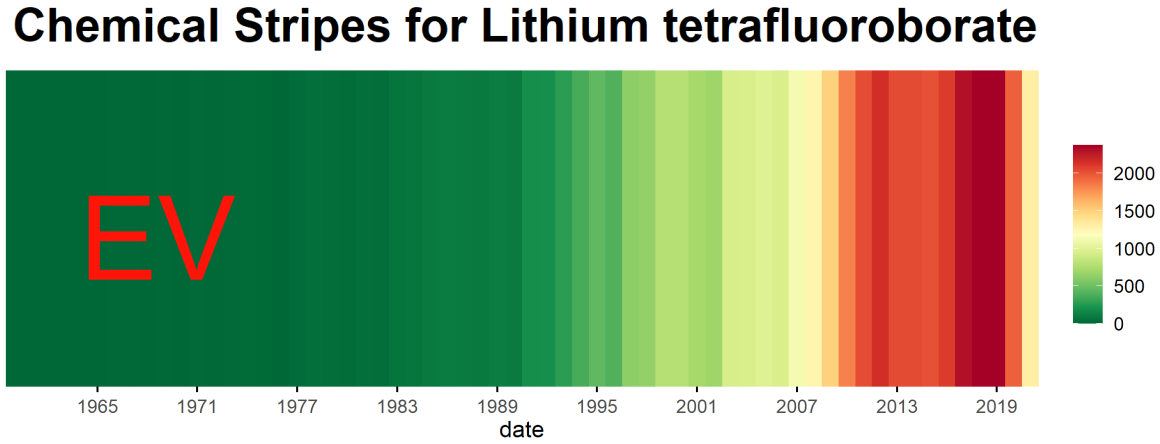
https://gitlab.lcsb.uni.lu/eci/chemicalstripes

The Chemical Stripes and Patent Data

(live stripe calculations during SETAC 2023)

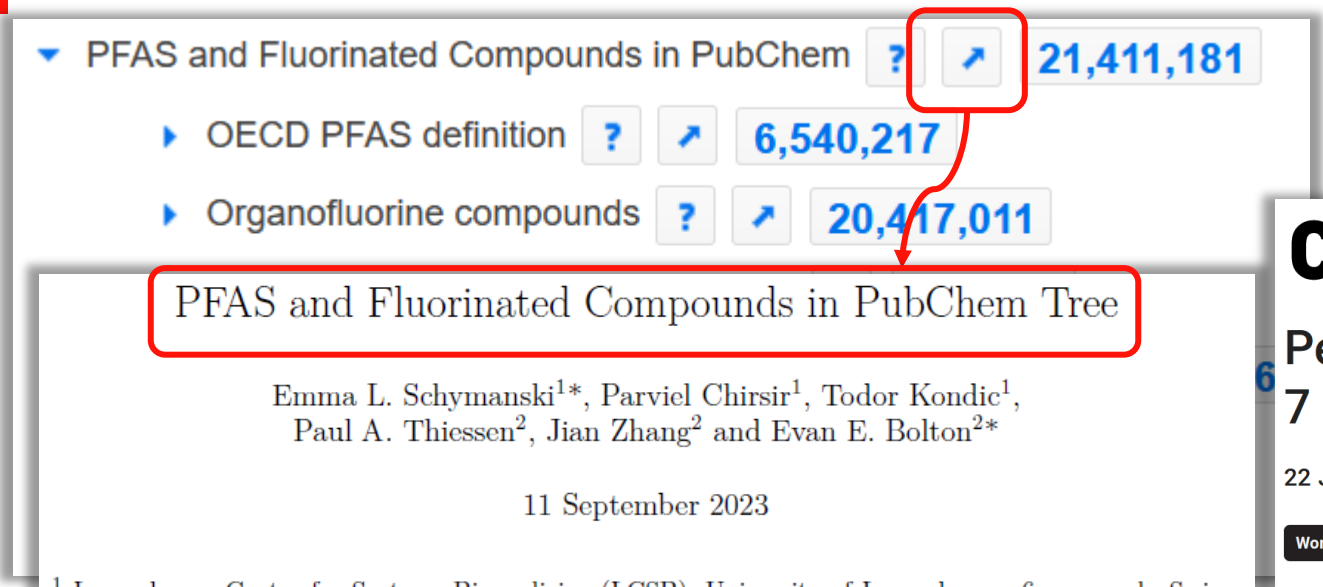






Left: Neuwald *et al, STOTEN*, DOI: 10.1016/j.scitotenv.2023.163921
Photo of Daniel Zahn, UFZ at SETAC Europe, 30 April – 4 May, 2023. Image reused with permission

Read/see more about the PubChem PFAS Tree



PFAS and Fluorinated Compounds in PubChem Tree

Emma L. Schymanski[1]*, Parviel Chirsir[1], Todor Kondic[1],
Paul A. Thiessen[2], Jian Zhang[2] and Evan E. Bolton[2]*

11 September 2023

[1] Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 avenue du Swing, 4367, Belvaux, Luxembourg. *ELS: emma.schymanski@uni.lu. ORCIDs: ELS: 0000-0001-6868-8145, PC: 0000-0002-9932-8609, TK: 0000-0001-6662-4375.

[2] National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), National Institutes of Health (NIH), Bethesda, MD, 20894, USA. *EEB: evan.bolton@nih.gov. ORCIDs: PAT: 0000-0002-1992-2086, JZ: 0000-0002-6192-4632, EEB: 0000-0002-5959-6190.

Preamble

This document describes the "PFAS and Fluorinated Compounds in PubChem Tree" (hereafter "PubChem PFAS Tree") in PubChem [1], developed jointly between PubChem (NCBI/NLM/NIH) and the Environmental Cheminformatics group (ECI) at the LCSB, University of Luxembourg, in consultation with several community representatives (see Contributions and Acknowledgements). The PubChem PFAS Tree (see Figure 1 and Contents listing) includes all compounds in PubChem satisfying various definitions, as explained later in this document. Note that each compound in PubChem has a PubChem Compound Identifier (CID), and the blue numbers next to each node header reflects the number of compounds (*i.e.* CIDs) in that node.

More details on the general PubChem Classification Brower features are given in the Section Exploring the Tree, via the PubChem documentation and help pages, or by reaching out to pubchem-help@ncbi.nlm.nih.gov for more information. Further information includes two videos on the ZeroPM YouTube channel, a ~23 min interactive walkthrough (Jun. 2022) and a ~1 hour webinar (Mar. 2023) [2], plus a preprint on ChemRxiv [3].

ChemRxiv®

How To Submit Browse About News

Per- and polyfluoroalkyl substances (PFAS) in PubChem: 7 million and growing

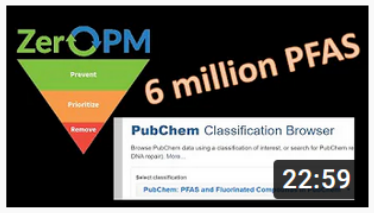
22 June 2023, Version 1

DOI: 10.26434/chemrxiv-2023-j823z

Working Paper

Emma Schymanski, Jian Zhang, Paul Thiessen, Parviel Chirsir, Todor Kondic, Evan Bolton

Show author details



Interactive walk-through (~23 min)
https://www.youtube.com/watch?v=g-sAazaagas
Detailed webinar (1 hr)
https://www.youtube.com/watch?v=jkdvCs4pGzU

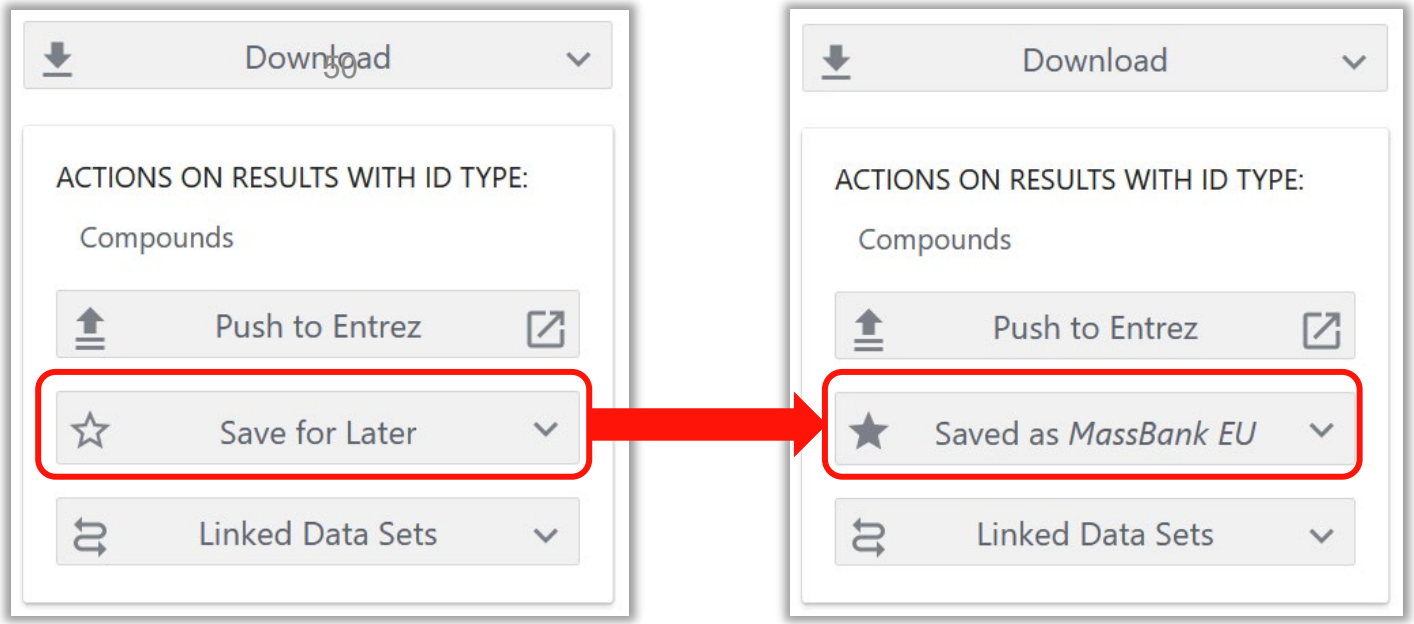
Advanced Queries

Schymanski *et al.*(2023) DOI: 10.26434/chemrxiv-2023-j823z

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

- PFAS and Fluorinated Compounds in PubChem ? 21,411,181
  - OECD PFAS definition ? 6,540,217
  - Organofluorine compounds ? 20,417,011
  - Other diverse fluorinated compounds ? 125,621
  - PFAS and fluorinated compound collections ? 1,789,296
  - PFAS breakdowns by chemistry ? 7,497,376
  - Regulatory PFAS collections ? 26,943

"Saved Searches" option



Advanced Queries

Schymanski *et al.*(2023) DOI: 10.26434/chemrxiv-2023-j823z

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

- PFAS and Fluorinated Compounds in PubChem 21,411,181
  - OECD PFAS definition 6,540,217
  - Organofluorine compounds 20,417,011
  - Other diverse fluorinated compounds 125,621
  - PFAS and fluorinated compound collections 1,789,296
  - PFAS breakdowns by chemistry 7,497,376
  - Regulatory PFAS collections 26,943

"Saved Searches" option

QUERY 1: Choose One | OPERATOR: AND | QUERY 2 (OF THE SAME ID TYPE AS QUERY 1): Choose One

Add to Saved | Reset

QUERY	ID TYPE	LIST SIZE	EXPIRES IN	ACTIONS ON YOUR RESULTS
OECD PFAS in PubChem AND Agrochemicals AND Collision Cross Section (CCS)	CID	27	7 hours	View Results / Delete
OECD PFAS in PubChem AND Agrochemicals AND MassBank EU	CID	71	7 hours	View Results / Delete
OECD PFAS in PubChem AND Agrochemicals	CID	306	7 hours	View Results / Delete
MassBank EU		16,255	7 hours	View Results / Delete
Collision Cross Section (CCS)		6,564	7 hours	View Results / Delete
Agrochemicals		3,135	7 hours	View Results / Delete
OECD PFAS in PubChem		7,497,376	7 hours	View Results / Delete

DOWNLOAD

Summary (Search Results)

CSV JSON XML

COMPRESSION: None GZip

The source of PFAS data in PubChem?

https://tarheels.live/bakerlab/

An overview of the uses of per- and polyfluoroalkyl substances (PFAS)†

Juliane Glüge, *[a] Martin Scheringer, [a] Ian T. Cousins, [b] Jamie C. DeWitt,[c] Gretta Goldenman,[d] Dorte Herzke, [ef] Rainer Lohmann, [g] Carla A. Ng, [h] Xenia Trier[i] and Zhanyun Wang[j]

Glüge *et al.* (2020) ESPI, DOI: 10.1039/d0em00291g

PubChem 13C3-PFHxS (Compound)

3.2.1 Collision Cross Section

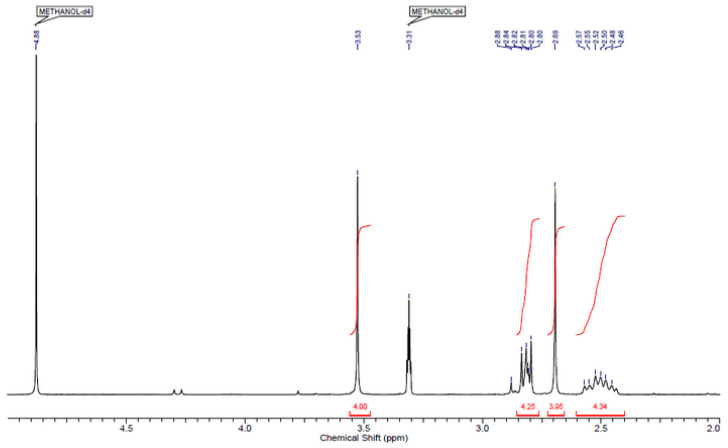
150.51 Å² [M-H]⁻ [CCS Type: DT; Buffer gas: N2; Dataset: PFAS]

DOI:10.1021/acs.est.2c00201

Baker Lab, Chemistry Department, The University of North Carolina at Chapel Hill

PubChem 6:2 FTMAP (Compound)

Reference for Source	B Bugsel, R Bauer, F Herrmann, ME Maier, C Zwiener (2022) Analytical and Bioanalytical Chemistry, 414, 1217-1225 doi:10.1007/s00216-021-03463-9



Reference for Dataset	S74 | REFTPS | Transformation Products and Reactions from Literature doi:10.5281/zenodo.4318838

NORMAN Suspect List Exchange

…and many more!

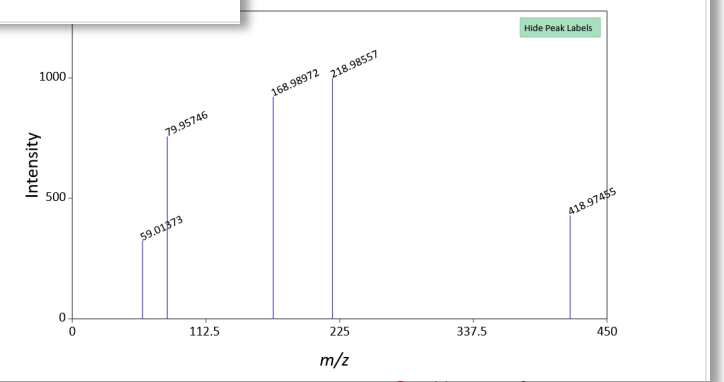
Bugsel *et al.* (2022) ABC, 414, 1217-1225. DOI:10.1007/s00216-021-03463-9

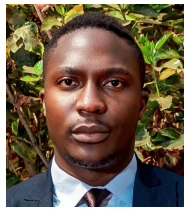
https://pubchem.ncbi.nlm.nih.gov/compound/156620404#section=1H-NMR-Spectra

PubChem Perfluorononanoic acid (Compound)

Accession ID	MSBNK-ACES_SU-AS000012
Authors	ACESx, Martin Group
Instrument	QExactive Orbitrap HF-X (Thermo Scientific)
Instrument Type	LC-ESI-QFT

391ed2b0ea221bd037

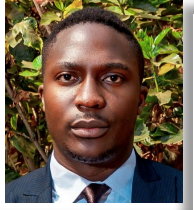




Parviel Chirsir @PChirsir

https://twitter.com/AcademicTox/status/1605997310726443009

FAIRifying PFAS TPs



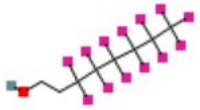
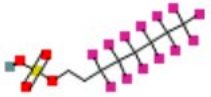
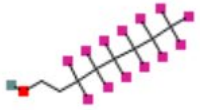
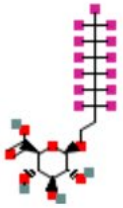
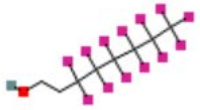
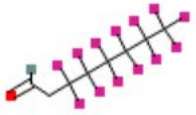
3,3,4,4,5,5,6,6,7,7,8,8,8-Tridecafluorooctan-1-ol (Compound)

7.1 Transformations

29 items

Search

SORT BY Predecessor Name -

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme	Evidence D
	6:2 FTOH		6:2 FTOH-Sulfate	Conjugation		10.1016/j.cbi.2005.06.007
	6:2 FTOH		6:2 FTOH-Gluc	Conjugation		10.1016/j.cbi.2005.06.007
	6:2 FTOH		6:2 FTAL	Oxidation	24296	10.1016/j.cbi.2005.06.007

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

ELSEVIER

Chemico-Biological Interactions

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

Metabolic products and pathways of fluorotelomer alcohols in isolated rat hepatocytes

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

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

Show more

+ Add to Me

https://doi.org/1

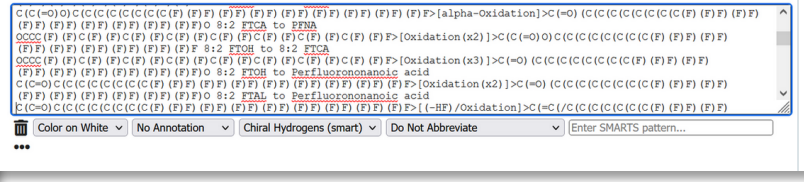
Jon Martin
@AcademicTox

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

@PChirsir

Emma Schymanski @ESchymanski · Jan 8

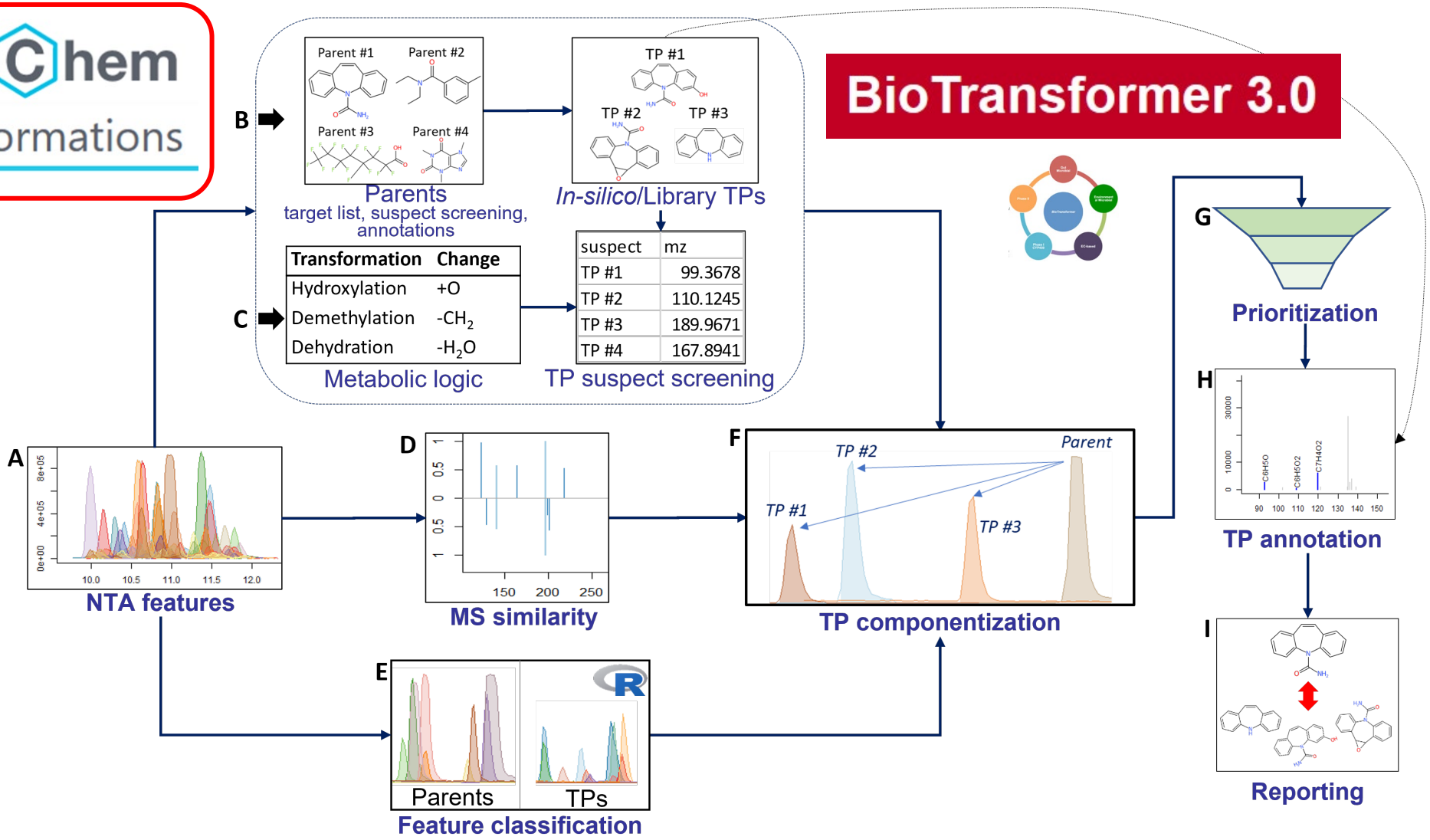
Spent a rainy afternoon FAIRifying #PFAS data from past @AcademicTox work, thanks to @PChirsir efforts we've committed the first set to @pubchem incl. new structures - hope to bring the 2005 PDF images to #OpenScience life soon! Again @the_cdk Depict makes it so much fun @jwmay !



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

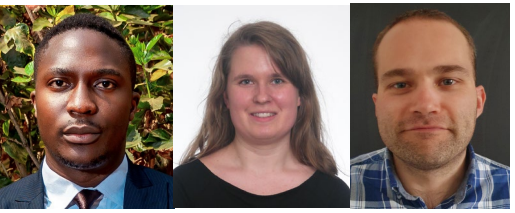
Open Transformation Products Workflows in patRoon 2.0

PubChem Transformations



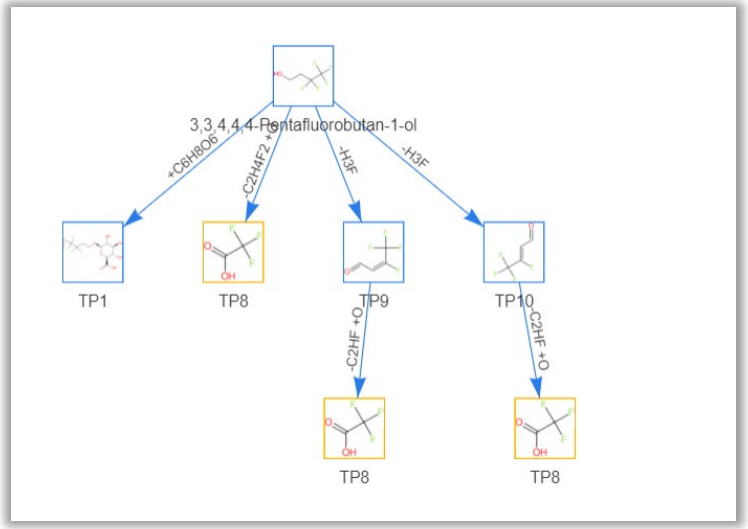
Schymanski *et al*. (2021) DOI: 10.1186/s13321-021-00489-0; Helmus *et al* (2022) DOI: 10.21105/joss.04029
Schymanski, Bolton, Cheng, Thiessen, Zhang, Helmus (2021) Transformations in PubChem, DOI: 10.5281/zenodo.5644560

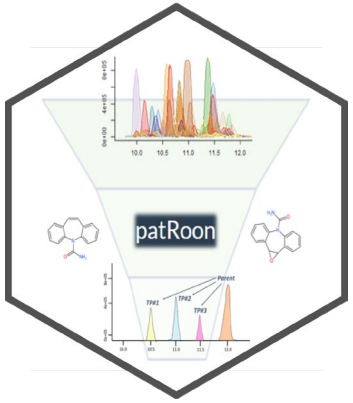
Applying TP NT-HRMS Workflows with patRoon

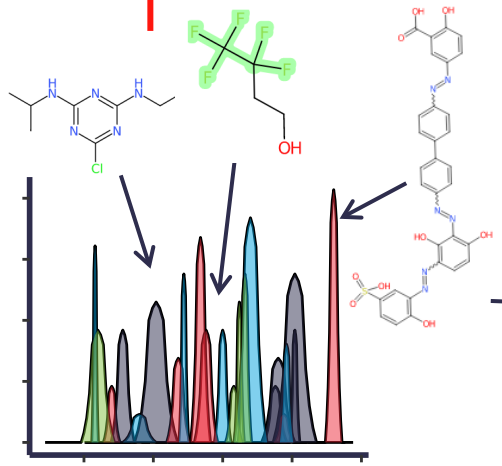


ShinyTPs

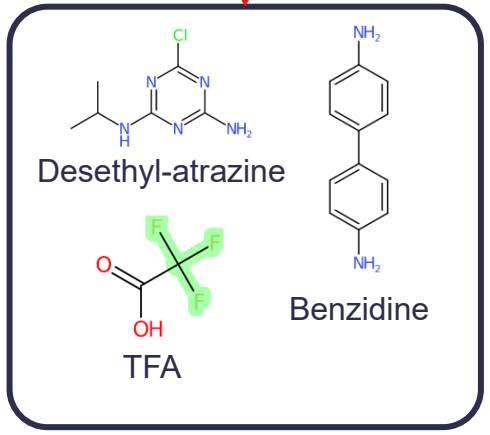
PubChem Transformations



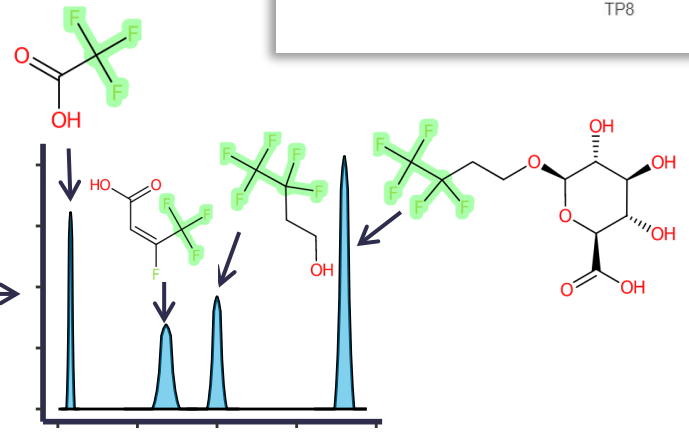




Suspect/Non-target screening



TP library

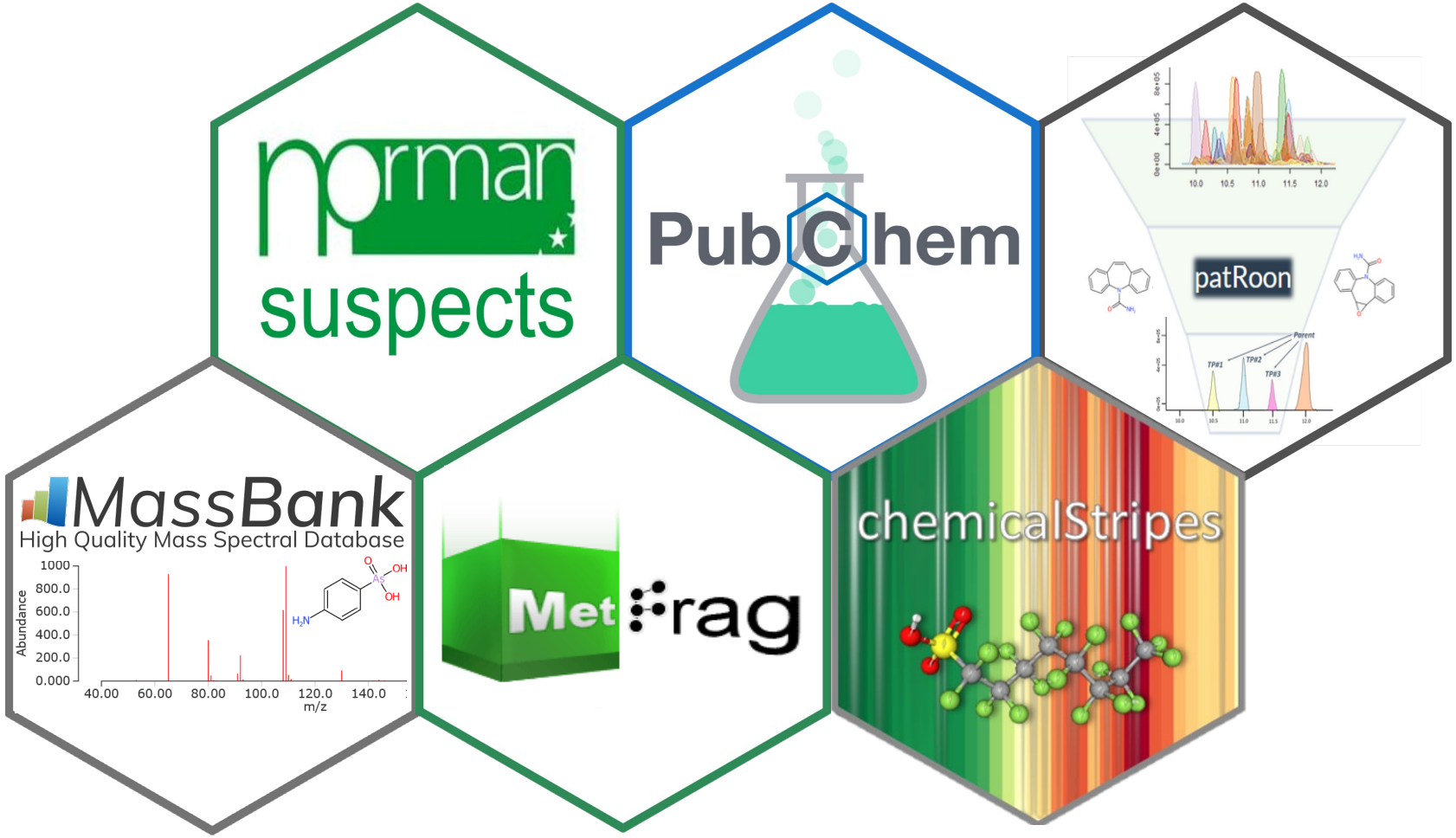


TP suspect screening

Mod. from Palm *et al* (2023) ShinyTPs, DOI: 10.26434/chemrxiv-2023-xm41h, Palm *et al* (2023). SETAC Presentation, DOI: 10.5281/zenodo.7885829

Take home messages

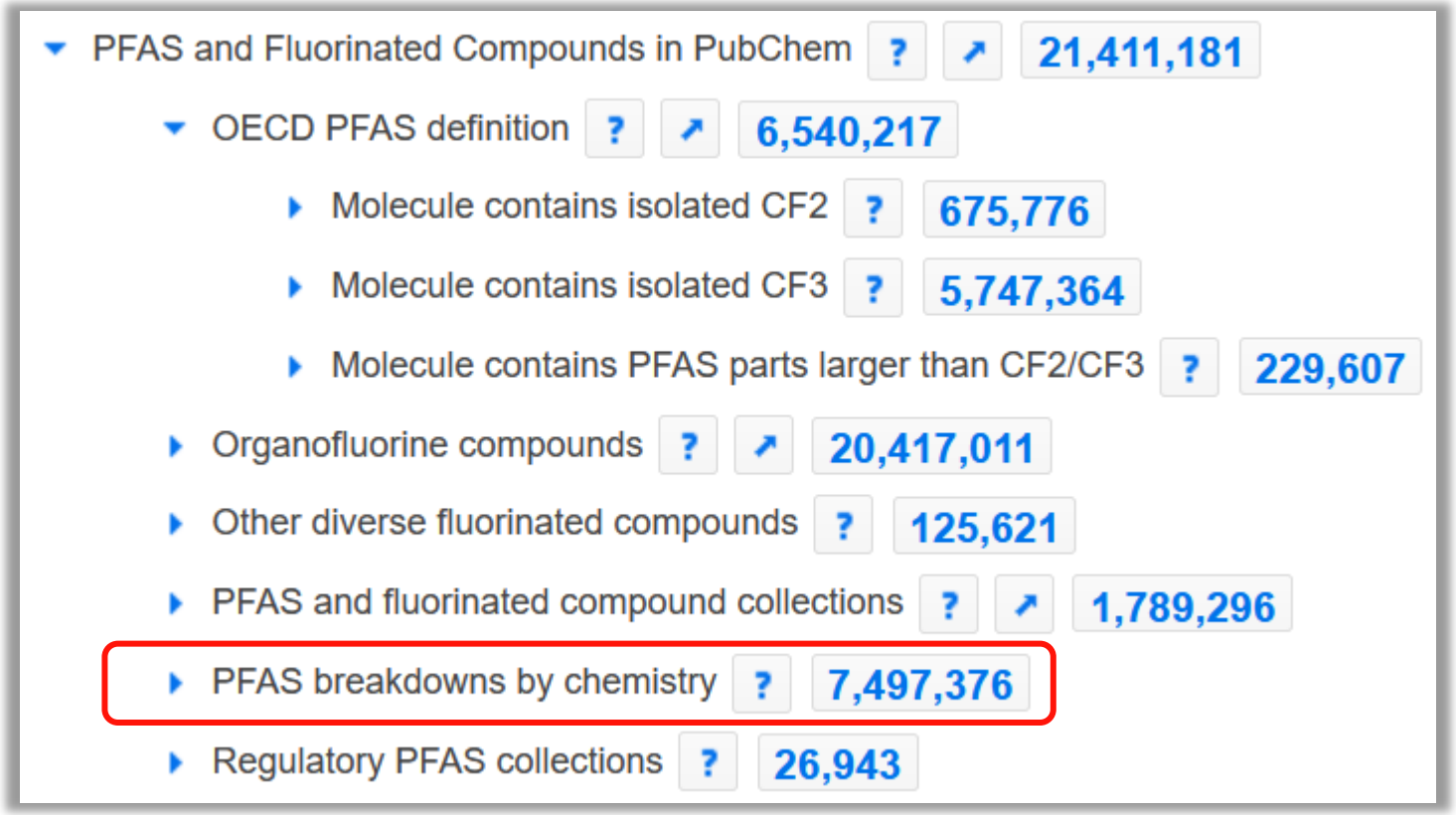
- Open data & suspect lists support NT-HRMS workflows



Take home messages

- Open data & suspect lists support NT-HRMS workflows

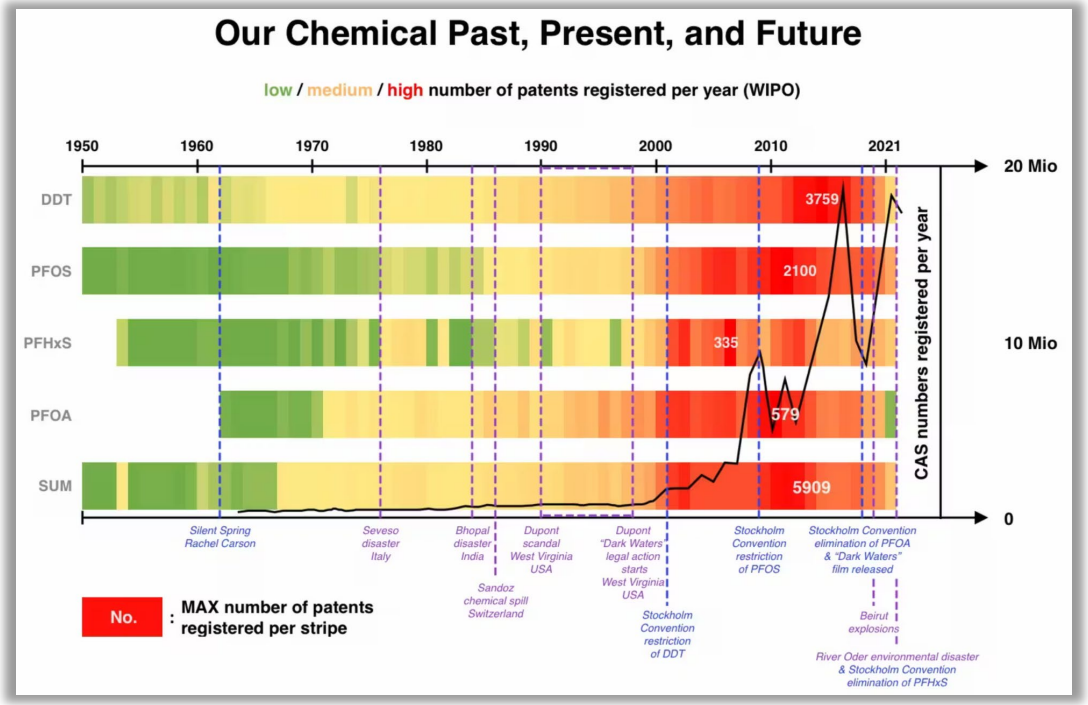
- There really are >7 million PFAS in PubChem!



Schymanski *et al* (2023) DOI: 10.26434/chemrxiv-2023-j823z & https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120

Take home messages

- Open data & suspect lists support NT-HRMS workflows
- There really are >7 million PFAS in PubChem!
- Annotation content helps find and interpret the relevant entries (and this comes from YOU!)



- Video: https://vimeo.com/jpmlmusic/ourchemicalpastpresentandfuture
- Soundtrack: https://soundcloud.com/jamieperera/our-chemical-past-present-and-future

Arp, Aurich *et al* (2023) ES&T, DOI: 10.1021/acs.est.3c01735

Take home messages

- Open data & suspect lists support NT-HRMS workflows
- There really are >7 million PFAS in PubChem!
- Annotation content helps find and interpret the relevant entries (and this comes from YOU!)
- Workflows are available in patRoon (and more)
- Help avoid the next Silent Spring & share your data!

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

Schymanski *et al* (2023) DOI: 10.26434/chemrxiv-2023-j823z / Arp, Aurich *et al* (2023) ES&T, DOI: 10.1021/acs.est.3c01735

Acknowledgements!

Today's slides:

DOI: 10.5281/zenodo.8353248

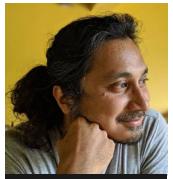
Email: emma.schymanski@uni.lu

Twitter/X: @ESchymanski

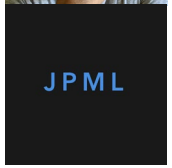
@schymane@mstdn.social

ORCID: 0000-0001-6868-8145

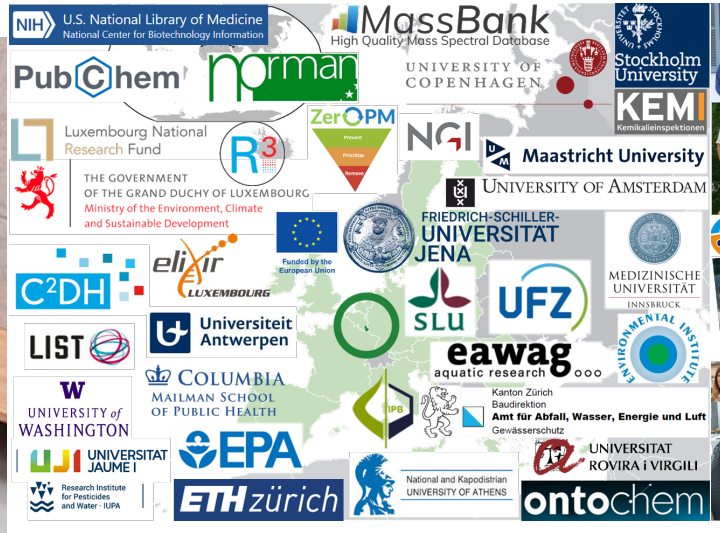
Today's slides zenodo



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



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



Funded by the European Union

H2020: 101036756