

# Exploring Millions of PFAS ...with FAIR and Open Science

**Emma Schymanski**, Dagny Aurich, Parviel Chirsir, Todor Kondic, Emma Palm  
(and the entire Environmental Cheminformatics team @ LCSB, UniLu)

Contact: [emma.schymanski@uni.lu](mailto:emma.schymanski@uni.lu) Twitter/X: [@ESchymanski](https://twitter.com/ESchymanski) Fedi: [@schymane@mstdn.social](https://mstdn.social/@schymane)

Rick Helmus (IBED, UvA), Steffen Neumann (IPB Halle), Jamie Perera (JPML),  
Evan Bolton, Paul Thiessen, Jian Zhang & PubChem Team (NCBI, NLM, NIH)  
(plus many, many other colleagues and collaborators through the years!)



# FAIR and Open Science

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

## Water science must be Open Science

Emma L. Schymanski & Stanislaus J. Schymanski

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

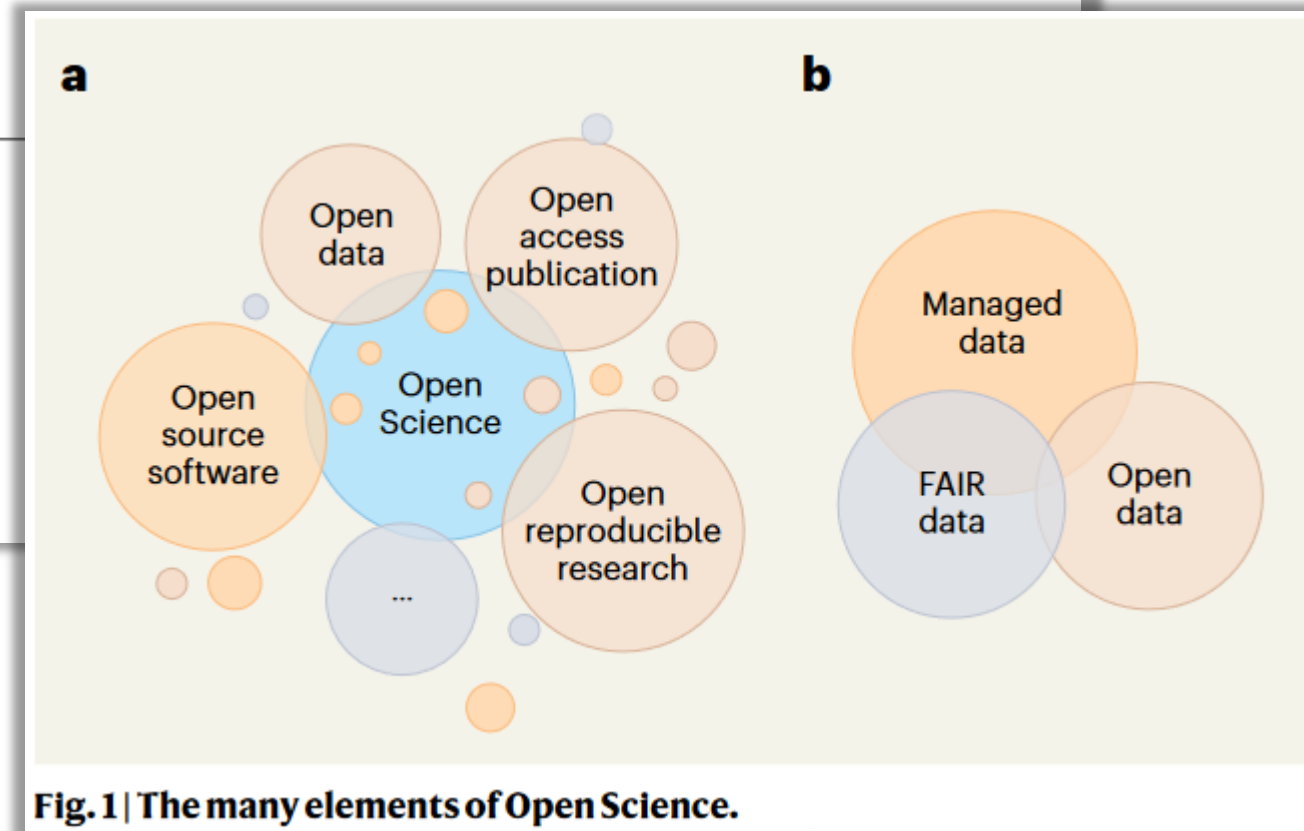
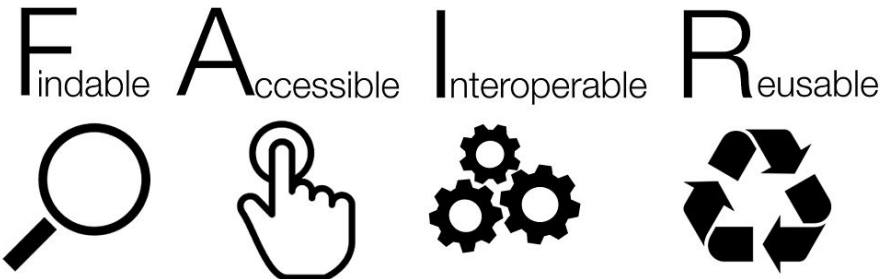
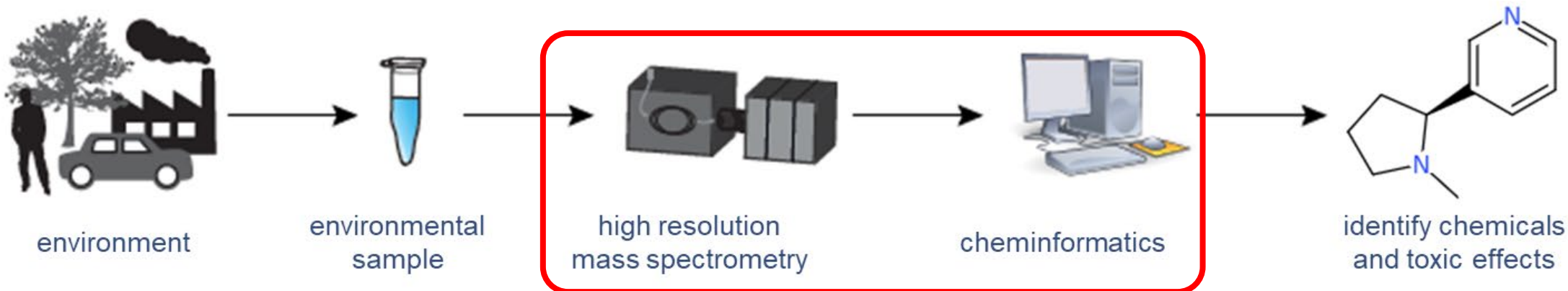
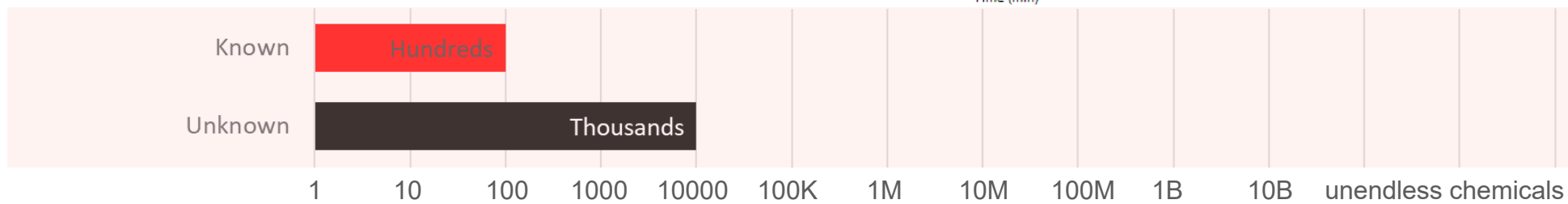
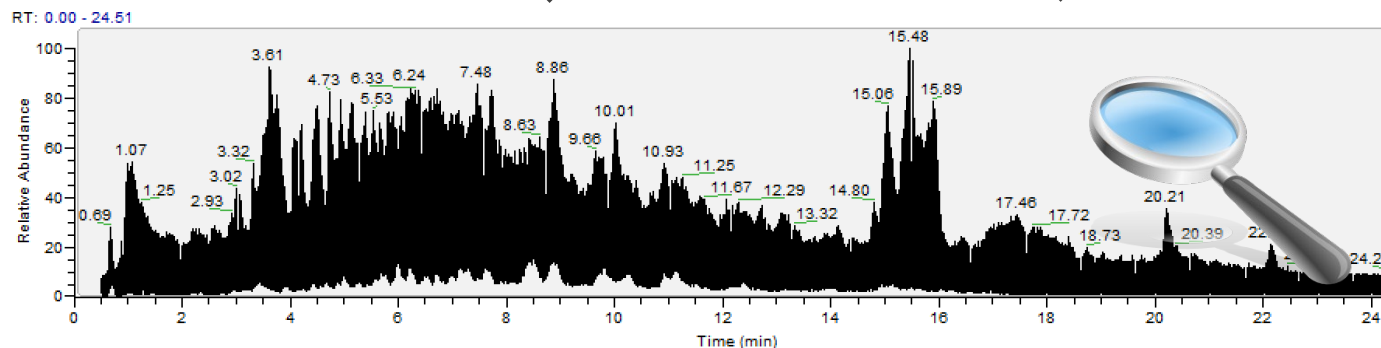


Fig. 1 | The many elements of Open Science.

# Environmental Cheminformatics & Non-target HR-MS



High resolution mass spectrometry  
AND connecting chemical knowledge



Samples

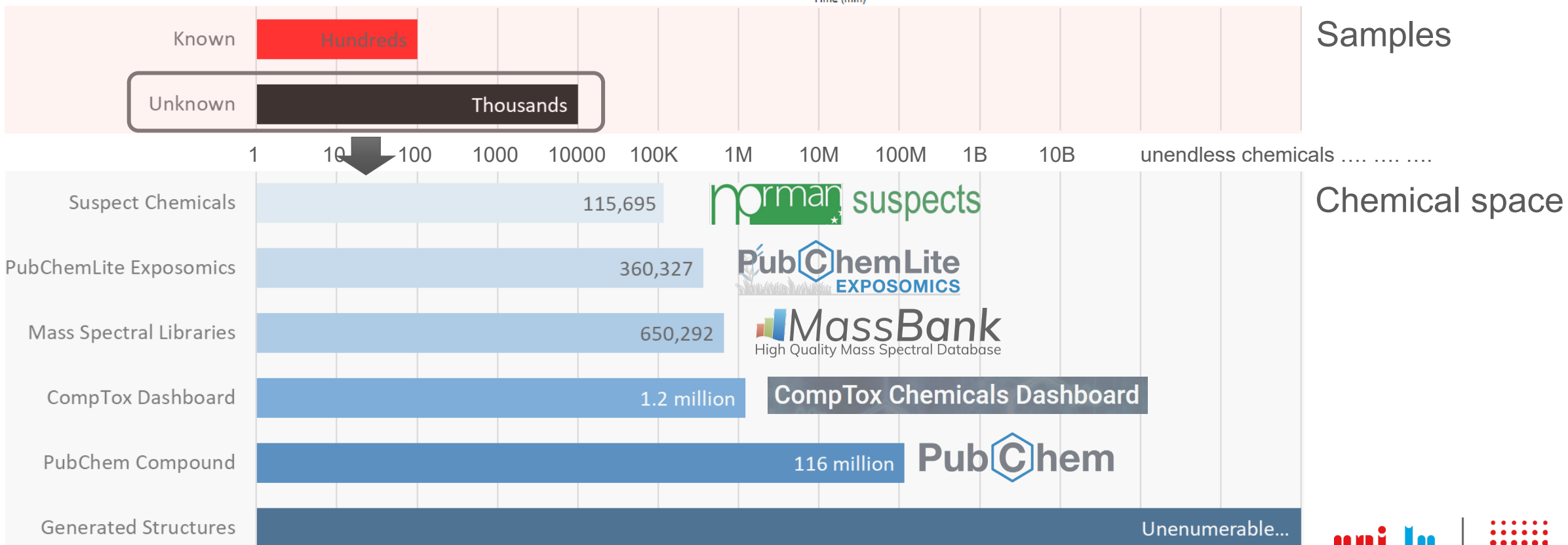
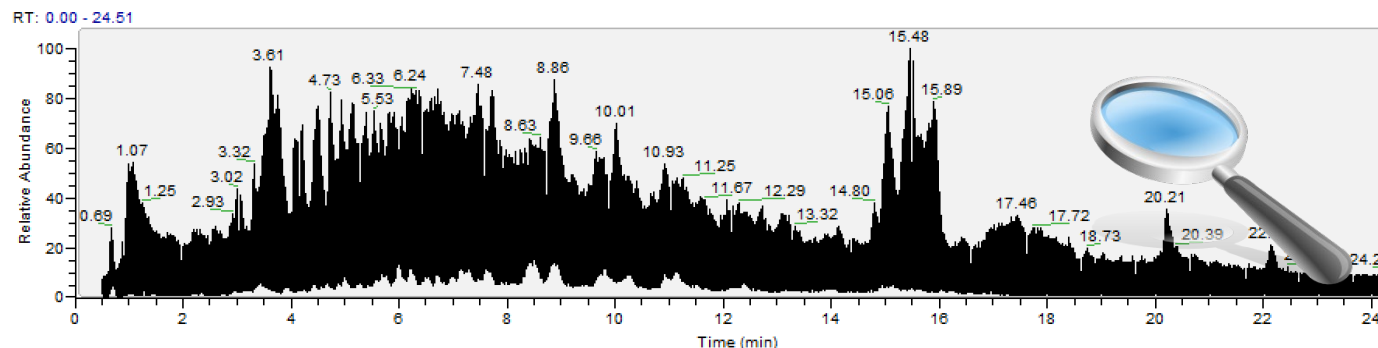
1 10 100 1000 10000 100K 1M 10M 100M 1B 10B unendless chemicals ....



# Environmental Cheminformatics & Non-target HR-MS

High resolution  
mass spectrometry

AND connecting  
chemical knowledge





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

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

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

MassBank Europe

MassBank High Quality Mass Spectral Database

>> Search Spectra

MassBank Record: LU040605

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

Search for:

Basic Search Peak List Peaks

Compound Information

Compound name

Exact Mass

AND

Formula ( e.g. C<sub>6</sub>H<sub>7</sub>N<sub>5</sub>, C<sub>5</sub>H<sup>\*</sup>N<sub>5</sub>, C<sub>5</sub><sup>\*</sup> )

AND

Search

Mass Spectrum

Abundance

m/z

Chemical Structure

Nc1ccc(cc1)S(=O)(=O)O

# MassBank: Cross resource integration & new article!

## Environmental Science Processes & Impacts



PAPER

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

Check for updates

Cite this: DOI: 10.1039/d3em00181d

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

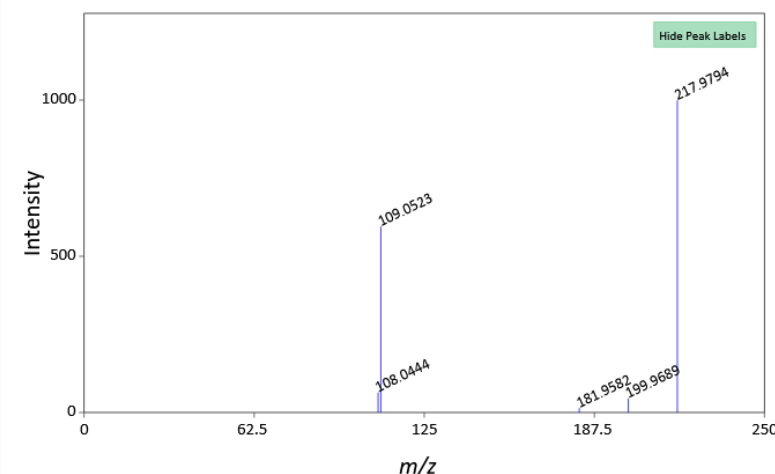
Anjana Elapavalore, \*<sup>a</sup> Todor Kondić, <sup>a</sup> Randolph R. Singh, <sup>ab</sup>  
Benjamin A. Shoemaker, <sup>c</sup> Paul A. Thiessen, <sup>c</sup> Jian Zhang, <sup>c</sup> Evan E. Bolton <sup>c</sup>  
and Emma L. Schymanski \*<sup>a</sup>

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

Arsanilic acid (Compound)

Top 5 Peaks	m/z
	217.9794 999
	109.0523 595
	108.0444 64
	199.9689 45
	181.9582 14

splash10-066r-0690000000-79ae659f03bdc3340757



CC BY

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

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



# Gathering Expert Knowledge: NORMAN-SLE

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



norman  
suspects

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



NORMAN Database System



NORMAN Suspect List Exchange

The NORMAN Suspect List Exchange (NORMAN-SLE) was established to facilitate the exchange of information on suspected substances. The NORMAN-SLE documents all individual collections of substances (see Source column in SusDat). NORMAN-SLE versions are updated regularly. Comments and contributions are welcome - please email us at [nds@norman-network.com](mailto:nds@norman-network.com). Please refer to our [documentation](#) pages for: [citation](#) instructions.

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

Antibiotic Resistance Bacteria/Genes

A database of ARBs/ARGs in environmental matrices

RESEARCH

Open Access



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

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



SEARCH All Databases

Searching for individual substance or group(s) of substances

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



Substance Database

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



Suspect List Exchange

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



Antibiotic Resistance Bacteria/Genes

A database of ARBs/ARGs in environmental matrices

# Suspect List(s)

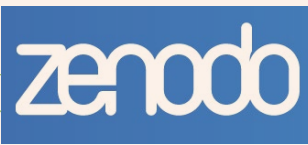
Eawag_ID	Name	CAS	Protected	logDpH7	SMILES	InChi	InChiKey	Molecular	ExactMass
249	N4-Acetyl	127-76-4	CAS_RN12	0.69	CC(=O)Nc1ccc(C)cc1	InChi=1S/C6H7NO	KXNWXWJN	C1H11N3	297.0242
236	4-Acetami	83-15-8	CAS_RN83	0.15	CN(C)C(=O)C	InChi=1S/C2H5NO	OLAGWXXK	C13H15N3	245.1164
245	N4-Acetyl	24341-30-8	CAS_RN24	1	COc1ccc(C)cc1	InChi=1S/C7H7NO	QDQWIKBK	C14H16N4	352.0841
247	N4-Acetyl	100-90-3	CAS_RN10	0.41	CC(=O)Nc1ccc(C)cc1	InChi=1S/C8H9NO	LUKAKWDL	C14H16N4	320.0943

Substrate	Substrate	Substrate	Substrate	Enzyme	Reaction
BIOTID0001	acetamino	1983	RZVAJNK	InChi=1S/CYP1A2	CYP Oxidation
BIOTID0001	pipoline	1958	QVMDECO	InChi=1S/CYP2D6	O-aryl de
BIOTID0001	3-Triflu	4296	KKIMDKM	InChi=1S/CYP2D6	CYP Aromatic
BIOTID0001	2-Amino-3	53462	ARZWATD	InChi=1S/CYP1A2	CYP N-Hydroxy

CAS Number	Name	SMILES	InChi	InChiKey	Molecular	ExactMass
283	Diuron-de	BIOTID0001	(R)-haloth			
294	Diuron-de	BIOTID0001	(R)-haloth			
118	Ethofume	BIOTID0001	(R)-haloth			
4	Metamitr	BIOTID0001	(R)-3,4-Me			
668	Simazine	BIOTID0001	(R)-3,4-Me			
671	Terbutylal	BIOTID0001	(4R)-5-(m			
672	Terbutylal	BIOTID0001	Homo-Tie			
268	Fipronil-si	BIOTID0001	(R)-1-(4-M			
2670	Atenolol	BIOTID0001	1-Methyl			
126	Methyl-2	BIOTID0001	(S)-2-(2,6			
296	N,N-Dime	BIOTID0001	Raberao			
3244	Betameth	BIOTID0001	Dehydro			
		BIOTID0001	Dehydro			
		BIOTID0001	2-Methoxy			
		BIOTID0001	3-(furan-2-			

## Suspect List Exchange

### (NORMAN-SLE)



DOI
10.5281/zenodo.6770176

12,835  
views

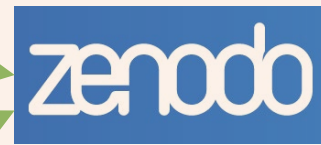
11,095  
downloads

# Suspect List(s)

Eawag_ID	Name	CAS	Protected	logDpH7	SMILES	InChi	InChiKey	Molecular	ExactMass
249	N4-Acetyl	127-76-4	CAS_RN12	0.69	CC(=O)NC1=CC=CC=C1	CC(=O)Nc1cccnc1	CC(=O)Nc1cccnc1	133.07	297.0242
236	4-Acetyl	83-15-8	CAS_RN83	0.15	CN1C=C(C)C=CC=C1	CN1C=C(C)C=CC=C1	CN1C=C(C)C=CC=C1	135.07	245.1164
245	N4-Acetyl	24341-30-8	CAS_RN24	1	COc1c(C)cc(C)cc1	COc1c(C)cc(C)cc1	COc1c(C)cc(C)cc1	150.08	352.0841
247	N4-Acetyl	100-90-3	CAS_RN10	0.41	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	CC(=O)Nc1ccc(C)cc1	151.08	320.0943

CAS Number	Name	Priority	Substance	Source	Category	Notes
1983	RZVAJINKF	InChi=1S/CYP1A2; CYP2D6	Oxidation	1983	1983	1983
1989	CYMDXQJ	InChi=1S/CYP2D6	O-aryl det	1989	1989	1989
4296	KKIMDKME	InChi=1S/CYP2D6; CYP2C19	Aromatic	4296	4296	4296
53462	ARZWATD	InChi=1S/CYP1A2; CYP2D6	N-Hydroxy	53462	53462	53462



DOI [10.5281/zenodo.6770176](https://doi.org/10.5281/zenodo.6770176)

12,835 views

11,095 downloads

Suspect List Exchange (NORMAN-SLE)

## NORMAN Database System (NDS)

### Ecotoxicology

Substance Database (NORMAN SusDat)

SARS-CoV-2 in sewage

Passive Sampling

Antibiotic Resistance Bacteria/Genes

Indoor Environment

MassBank Europe

Bioassays Monitoring Data

Substance Factsheets

Digital Sample Freezing Platform

Prioritisation

Chemical Occurrence Data





# Suspect List(s)

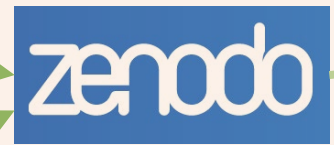
Eawag_ID	Name	CAS	Protected	logDpH7	SMILES	InChI	InChIKey	Molecular	ExactMass
249	N4-Acetyl	127-76-4	CAS_RN12	0.69	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C11H11N3	297.0242	
236	4-Acetyl	83-15-8	CAS_RN83	0.15	CN1C=C(C)O	CN1C=C(C)O	C13H15N3	245.1164	
245	N4-Acetyl	24341-30-8	CAS_RN24	1	COc1cnc1	COc1cnc1	C14H16N4	352.0841	
247	N4-Acetyl	100-90-3	CAS_RN10	0.41	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C14H16N4	352.0943	

BIOTID0001	acetamino	1983	RZVAJNKf	InChI=1S/C	CYP1A2;	CYP2D6	Oxidation
BIOTID0002 <td>paroline</td> <td>1988</td> <td>QYMDXQ</td> <td>InChI=1S/C <td>CYP2A6</td> <td></td> <td>(m Hydrolysis)</td> </td>	paroline	1988	QYMDXQ	InChI=1S/C <td>CYP2A6</td> <td></td> <td>(m Hydrolysis)</td>	CYP2A6		(m Hydrolysis)
BIOTID0003 <td>Trifluo</td> <td>4296</td> <td>KKIMDKMf</td> <td>InChI=1S/C <td>CYP2D6; <td>CYP2A6</td> <td>C Aromatic</td> </td></td>	Trifluo	4296	KKIMDKMf	InChI=1S/C <td>CYP2D6; <td>CYP2A6</td> <td>C Aromatic</td> </td>	CYP2D6; <td>CYP2A6</td> <td>C Aromatic</td>	CYP2A6	C Aromatic
BIOTID0004 <td>2-Amino-3</td> <td>53462</td> <td>ARZWATDf</td> <td>InChI=1S/C <td>CYP1A2; <td>CYP2D6</td> <td>C N-Hydroxy</td> </td></td>	2-Amino-3	53462	ARZWATDf	InChI=1S/C <td>CYP1A2; <td>CYP2D6</td> <td>C N-Hydroxy</td> </td>	CYP1A2; <td>CYP2D6</td> <td>C N-Hydroxy</td>	CYP2D6	C N-Hydroxy

CAS Number	Name	RT	SMILES	InChI	InChIKey	Molecular	ExactMass	Retention	Reference
283	Diuron-de	11.8	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C11H11N3	297.0242		11.8	Diuron-de
294	Diuron-de	11.8	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C11H11N3	297.0242		11.8	Diuron-de
118	Ethofome	11.8	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C11H11N3	297.0242		11.8	Ethofome
4	Metamitr	11.8	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C11H11N3	297.0242		11.8	Metamitr
668	Simazine	11.8	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C11H11N3	297.0242		11.8	Simazine
671	Terbutylat	11.8	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C11H11N3	297.0242		11.8	Terbutylat
672	Terbutylat	11.8	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C11H11N3	297.0242		11.8	Terbutylat
2668	Flpronil-st	11.8	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C11H11N3	297.0242		11.8	Flpronil-st
2670	Atenolol	11.8	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C11H11N3	297.0242		11.8	Atenolol
126	Methyl-2	11.8	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C11H11N3	297.0242		11.8	Methyl-2
296	N,N-Dime	11.8	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C11H11N3	297.0242		11.8	N,N-Dime
3244	Betameth	11.8	CC(=O)Nc1cnc1	CC(=O)Nc1cnc1	C11H11N3	297.0242		11.8	Betameth



DOI [10.5281/zenodo.6770176](https://doi.org/10.5281/zenodo.6770176)

12,835 views

11,095 downloads

www.norman-data.eu

NORMAN Digital Sample Freezing Platform

Main Page Batch mode

Contributed Samples Results Chromatograms Interactive Map Help

Choose Emerging Substance or input mass of interest and experimental RTI

Substance name or CAS or StdInChIKey

DEET [ 134-62-3 ]  
[MMOZXZBCLCQITDF-UHFFFAOYSA-N]

Choose Ionization

Positive

Adduct

[M+H]<sup>+</sup>

Predicted RTI Positive

523

Reset

## NORMAN Database System

Ecotoxicology

SARS-CoV-2 in sewage

Passive Sampling

Indoor Environment

Bioassays Monitoring Data

Digital Sample Freezing Platform

Chemical Occurrence Data



# Suspect List(s)

Eawag_ID	Name	CAS	Protected	logDpH7	SMILES	InChI	InChIKey	Molecular	ExactMass
249	N4-Acetyl	127-76-4	CAS_RN12	0.69	CC(=O)NC1=CC=CC=C1	CC(=O)Nc1ccccc1	CN1C=CC=CC=C1	123.07242	123.07242
236	4-Acetyl	83-15-8	CAS_RN83	0.15	CN1=CC=C(C=C1)C(=O)C	CN1=CC=C(C=C1)C(=O)C	CN1=CC=C(C=C1)C(=O)C	123.07242	123.07242
245	N4-Acetyl	24341-30-8	CAS_RN24	1	COc1ccc(cc1)C(=O)C	CC(=O)C1=CC=C(C=C1)C=C1	COC(=O)C1=CC=C(C=C1)C=C1	352.0841	352.0841
247	N4-Acetyl	100-90-3	CAS_RN10	0.41	CC(=O)NC1=CC=C(C=C1)C=C1	CC(=O)Nc1ccc(cc1)C=C1	CN1=CC=C(C=C1)C=C1	320.0943	320.0943

Substrate	substrate	substrate	substrate	enzyme	reaction
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP1A2	CYP1A2; CYP1A2
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2A6	(m-Hydroxylated)
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2D6	O-aryl
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C8	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C9	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C19	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C18	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C7	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C17	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C10	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C9	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C8	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C9	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C10	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C11	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C12	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C13	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C14	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C15	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C16	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C17	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C18	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C19	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C20	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C21	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C22	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C23	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C24	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C25	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C26	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C27	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C28	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C29	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C30	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C31	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C32	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C33	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C34	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C35	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C36	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C37	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C38	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C39	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C40	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C41	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C42	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C43	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C44	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C45	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C46	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C47	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C48	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C49	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C50	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C51	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C52	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C53	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C54	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C55	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C56	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C57	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C58	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C59	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C60	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C61	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C62	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C63	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C64	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C65	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C66	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C67	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C68	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C69	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C70	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C71	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C72	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C73	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C74	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C75	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C76	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C77	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C78	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C79	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C80	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C81	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C82	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C83	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C84	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C85	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C86	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C87	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C88	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C89	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C90	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C91	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C92	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C93	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C94	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C95	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C96	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C97	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C98	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C99	
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/CYP2C100	



DOI [10.5281/zenodo.6770176](https://doi.org/10.5281/zenodo.6770176)

12,835 views

11,095 downloads

**Suspect List Exchange (NORMAN-SLE)**

## NORMAN Database System (NDS)

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

**CompTox Chemicals Dashboard**

Chemical Lists

Search: NORMAN

Showing 74 of 300 Records

List Acronym	List Name	# Chemicals	Updated	List Description
ATHENSUS				
BSPHENOLS				
CCSCOMPEND				

**Metoprolol acid**  
56592-18-4 | DTXSID70881080

Chemical Details

- Quality Control Notes
- Intended Purposes
- Structural Identifiers
- Linked Substances
- Presence in Life
- Intentional
- Other

Metoprolol acid is a compilation of suspects, predicted transformation products and surfactants at University of Athens, as described in Gago-Ferreiro et al 2015, DOI: 10.1002/chem.201500454. The original data is available on the NORMAN Suspect List Exchange. Details are included here.

Metoprolol acid is a collection of biogenics available at NIB (Pawel Kocotowski) and from IAB (KTH) (Swedish Chemicals Agency, in Swedish with English summary), hosted on Met List Exchange (https://www.norman-network.com/nds/SLE). Dataset DOI: 10.377884

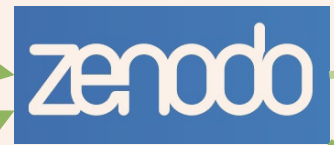
Metoprolol acid is included with > 3800 experimental collision cross section (CCS) values (drift tube IM-MS) in the NORMAN Suspect List Exchange by Jackie Pischke and John McLean, Vanderbilt University. Further details are available here.

# Suspect List(s)

Eawag_ID	Name	CAS	Protected	logDpH7	SMILES	InChI	InChIKey	Molecular	ExactMass
249	N4-Acetyl	127-76-4	CAS_RN12	0.69	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H13N3	297.0242		
236	4-Acetyl	83-15-8	CAS_RN83	0.15	CN1C(C=C)C(O)C=C1	InChI=1S/C11H15N3	245.1164		
245	N4-Acetyl	24341-30-8	CAS_RN24	1	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	352.0841		
247	N4-Acetyl	100-90-3	CAS_RN10	0.41	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	320.0943		

Substrate	substrate	substrate	substrate	enzyme	reaction
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/C11H15N3	CYP1A2; CYP2A6
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/C11H15N3	CYP2A6 (m-Hydroxylation)
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/C11H15N3	CYP2A6 (p-Hydroxylation)
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/C11H15N3	CYP2A6 (N-Hydroxylation)
BIOTID0001	acetaminol	1983	RZVAJNK	InChI=1S/C11H15N3	CYP2A6 (N-Acetylation)

CAS Number	Name	SMILES	InChI	InChIKey	Molecular	ExactMass
283	Diuron	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	297.0242		
294	Diuron	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	297.0242		
118	Ethofume	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	297.0242		
4	Metamit	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	297.0242		
668	Simazine	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	297.0242		
671	Terbutyl	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	297.0242		
672	Terbutyl	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	297.0242		
2668	Pipromid	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	297.0242		
2670	Atenolol	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	297.0242		
126	Methyl	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	297.0242		
296	N,N-Dime	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	297.0242		
3244	Betameth	CC(=O)Nc1ccc(C)cc1	InChI=1S/C11H15N3	297.0242		



DOI [10.5281/zenodo.6770176](https://doi.org/10.5281/zenodo.6770176)

12,835 views

11,095 downloads

## Suspect List Exchange (NORMAN-SLE)



NORMAN Suspect List Exchange Classification **115,260**

- S13 | EU-COSMETICS | Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006) **3,887**
- S25 | OECD PFAS | List of PFAS from the OECD **3,677**

3.2.10 Collision Cross Section

150.51 Å<sup>2</sup> [M-H]<sup>-</sup> [CCS Type: DT; Buffer gas: N<sub>2</sub>; Dataset: PFAS]  
DOI:10.1021/acs.est.2c02021

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

228.66 Å<sup>2</sup> [2M-H]<sup>-</sup>  
150.81 Å<sup>2</sup> [M-H]<sup>-</sup>

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

NORMAN Suspect List Exchange

S80 | PFASGLUE | Overview of PFAS Uses **1,250**

## NORMAN Database System (NDS)

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

## CompTox Chemicals Dashboard

Chemical Lists

Search: NORMAN

Showing 7/4 of 300 Records

List Acronym	List Name	Chemicals	Updated	List Description
ATHENSUS				
BSPHENOLS				
CCCOMPEND				

**Metoprolol acid**  
56592-14-4 [DTXSID7081080]  
Searched by Approved Name.

Quality Control Notes

- Internal Properties
- Structural Identifiers
- Linked Substances
- Presence in Life

Other

Metoprolol acid is a compilation of suspects, predicted transformation products and surfactants at University of Athens, as described in Gago-Ferreiro et al 2015, DOI: 10.1021/acs.est.5b01454. The original data is available on the NORMAN Suspect List Exchange. Please see included here.

Metoprolol acid is a collection of biogenics available at NIB (Pavel Kratochvíl) and from Lab K3M (Swedish Chemicals Agency, in Swedish with English summary), hosted on the NORMAN Suspect List Exchange (https://www.norman-network.eu/html/SLE). Dataset DOI: 10.5281/zenodo.4704648.

Metoprolol acid is included with > 3800 experimental collision cross section (CCS) values (drift tube IM-MS) from the NORMAN Suspect List Exchange by Jackie Pasche and John McLean, Vanderbilt University. The data is available on the NORMAN Suspect List Exchange (https://www.norman-network.eu/html/SLE). Dataset DOI: 10.5281/zenodo.4704648. Further curation ongoing. Further details available here.

7.5 Transformations

2 items View More Details

Sort By: Please Choose One

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Enzyme
	Nicotine		Norcotine	Thirdhand Smoke	
	Nicotine		Norcotine	Demethylation	

NORMAN Suspect List Exchange

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



# Explore Chemistry

Quickly find chemical information from authoritative sources

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

Use Entrez

Compounds

Substances

BioAssays



Draw Structure



Upload ID List



Browse Data



Periodic Table

116M Compounds

308M Substances

292M Bioactivities

36M Literature

934 Data Sources

[See More Statistics >](#)

[Explore Data Sources >](#)



# NORMAN-SLE in PubChem



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

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

The screenshot shows a hierarchical classification of NORMAN Suspect List Exchange (SLE) entries. At the top level, there are 115,695 entries. A dropdown menu is open for 'PFAS Use Category', showing 489 entries. The 'Aerospace' sub-category is expanded, showing 5 entries: 'Additive to aviation fuel' (1), 'Brake and hydraulic fluids' (5), 'Air conditioning' (1), 'Antifoaming Agents' (8), and 'Apparel' (26). Other categories visible include 'S13 | EUCOSMETICS | Combined Inventory of Ingredients Employed in Consumer Products (2000) and Revised Inventory (2006)' (3,936), 'S25 | OECDPFAS | List of PFAS from the OECD' (3,678), 'S36 | UBAPMT | Potential Persistent, Mobile and Toxic (PMT) substances' (254), 'S47 | ECHAPLASTICS | A list from the Plastic Additives Initiative Mapping Exercise by ECHA' (241), 'S50 | CCSCOMPEND | The Unified Collision Cross Section (CCS) Compendium' (869), 'S60 | SW' (1,359), 'S61 | UJI' (574), 'S66 | EA' (258), 'S68 | HSI' (740), 'S69 | LUX' (1,068), 'S72 | NT' (2,553), 'S75 | Cya', 'S77 | FCG', 'S79 | UACCSCEC | Collision Cross Section (CCS) Library from UAntwerp', and 'S80 | PFASGLUEGE | Overview of PFAS Uses' (1,251). Red boxes highlight the 'S25 | OECDPFAS' and 'S80 | PFASGLUEGE' categories, and a red arrow points from the 'Data in PubChem' section of the table to the 'S80' category.

# Creating Informative Subsets of PubChem



- 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

## PubChemLite EXPOSOMICS

~350,000 entries "small"



Updated PubChemLite to Aug 2023 version #127

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

**Database Settings**

Database: PubChemLite\_exposomics

Neutral Mass: 229.10948 Search ppm: 5

Formula: C9H16CIN5

Identifiers:

Retrieve Candidates 4 Candidates

zenodo

August 26, 2023

PubChemLite for Exposomics

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

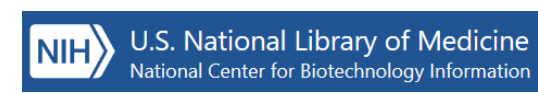
This is the repository for regular updates of the PubChemLite for Exposomics data collection. PubChemLite for Exposomics is a subset of PubChem selected from major categories of the Table of Contents page at the PubChem Classification Browser, described in DOI:10.1186/s13321-021-00489-0.

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

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


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

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



# MetFrag: *in silico* fragmentation for candidate selection





## MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

### Database Settings

Database:  Include references:

Parent Ion:

Neutral Mass:  Search ppm:

Formula:

Identifiers:

### Candidate Filter & Score Settings

### Fragmentation Settings & Processing

Mzppm:

Mzabs:

Mode:

Tree depth:

Group candidates

#### MS/MS Peak list

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



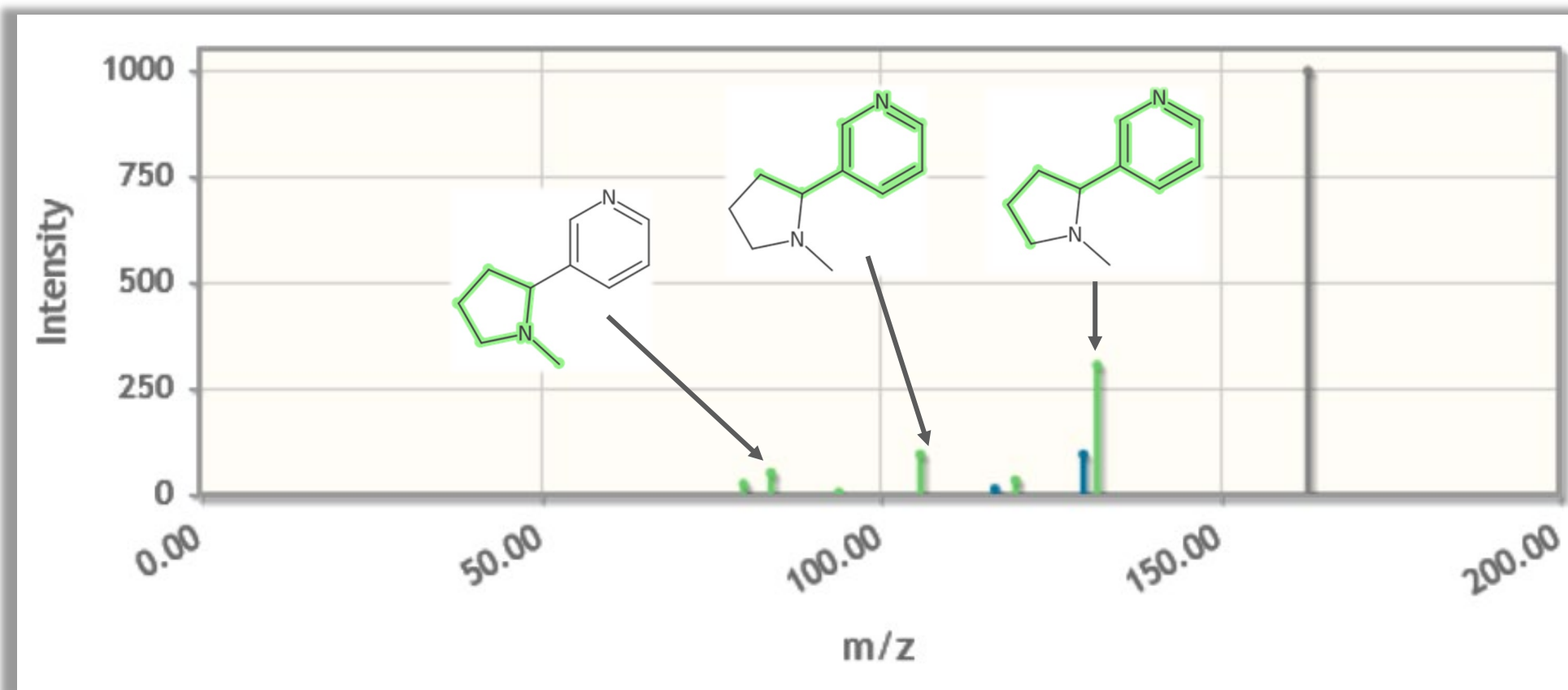


# MetFrag: *in silico* fragmentation for candidate selection



MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra



- matched
- not matched
- excluded



# MetFrag: Combining Open Resources for Annotation

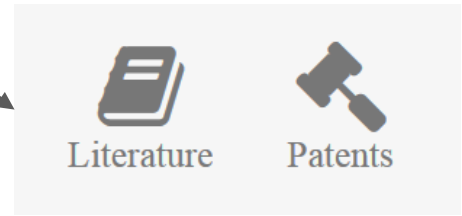
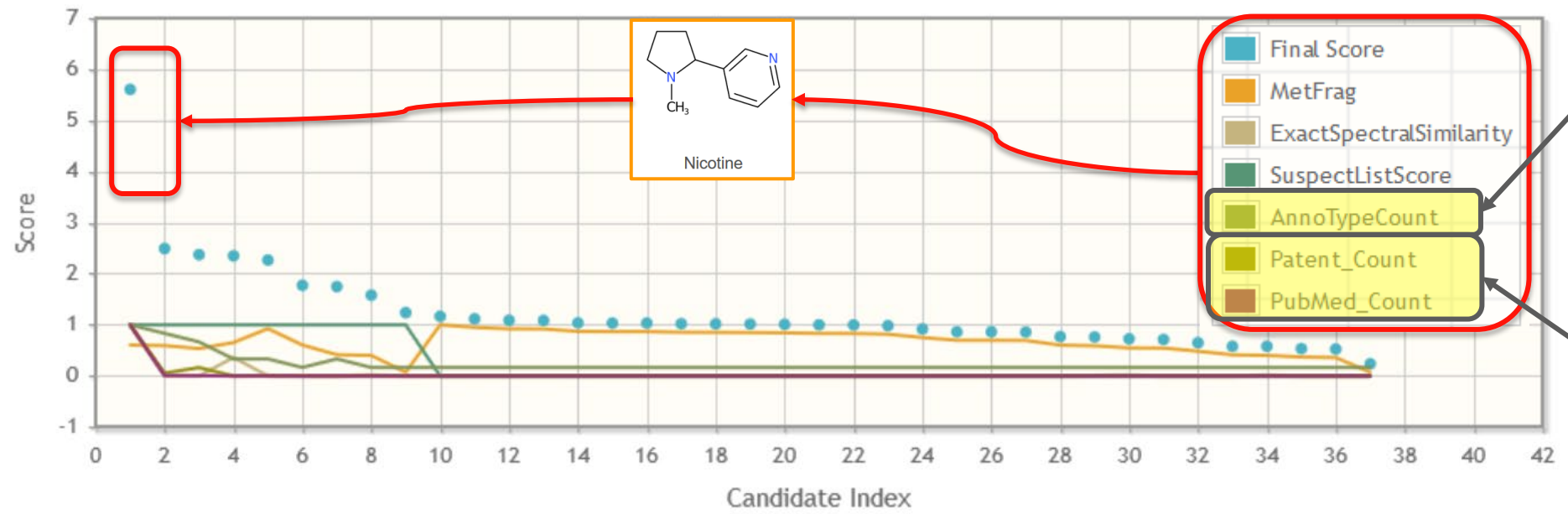


PubChem Compound TOC ? 67,343,260

- 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

## Statistics

Candidate Score Distribution



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

[Home](#) > [Environmental Sciences Europe](#) > [Article](#)

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

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

[Download PDF](#) ↓

✓ You have full access to this [open access](#) article



[Environmental Sciences Europe](#)

[Submit manuscript](#) →

[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](https://doi.org/10.1186/s12302-023-00779-4)

[Sections](#)

[Figures](#)

[References](#)

[Abstract](#)

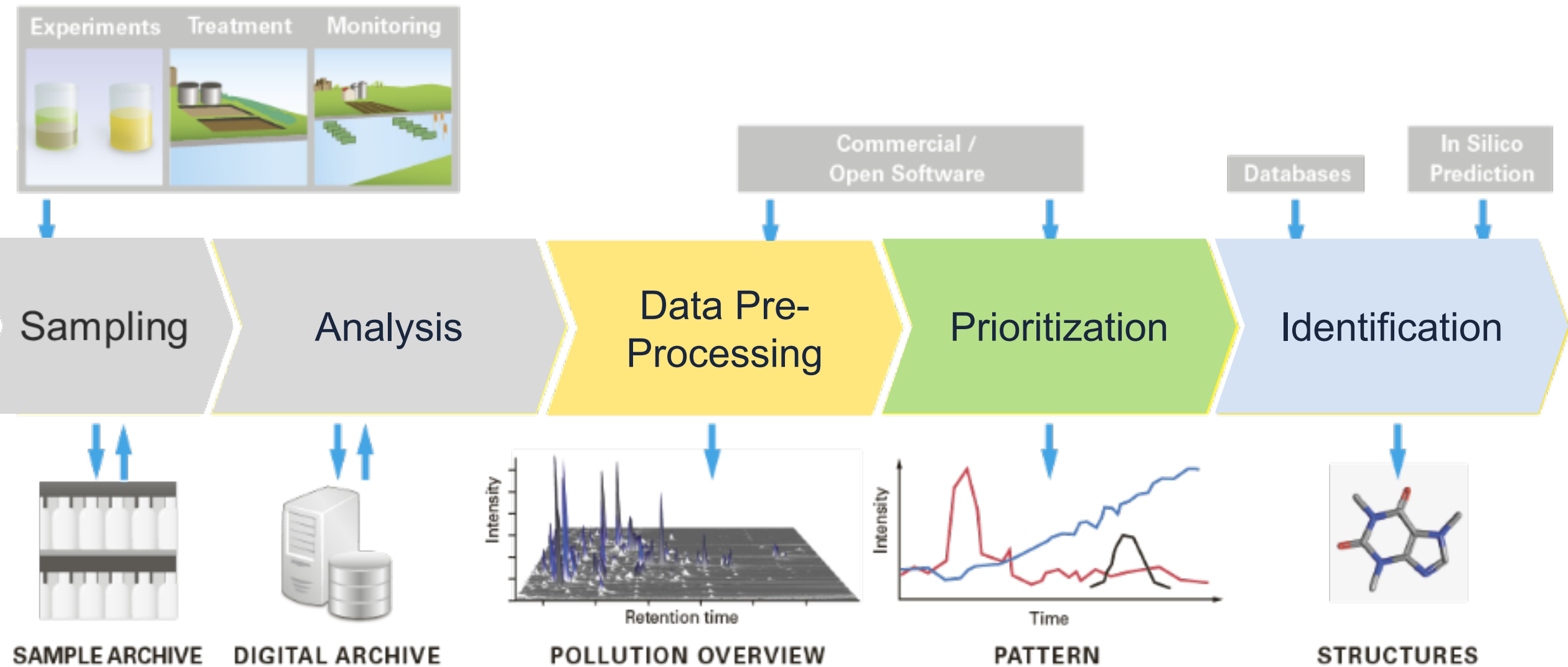
[Motivation for this guidance](#)

[Overview on analytical methods for NTS](#)

[Sampling and sample preparation for NTS](#)

[LC-HRMS/MS analysis](#)

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



# Open Source Workflows for NT-HRMS: patRoan



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

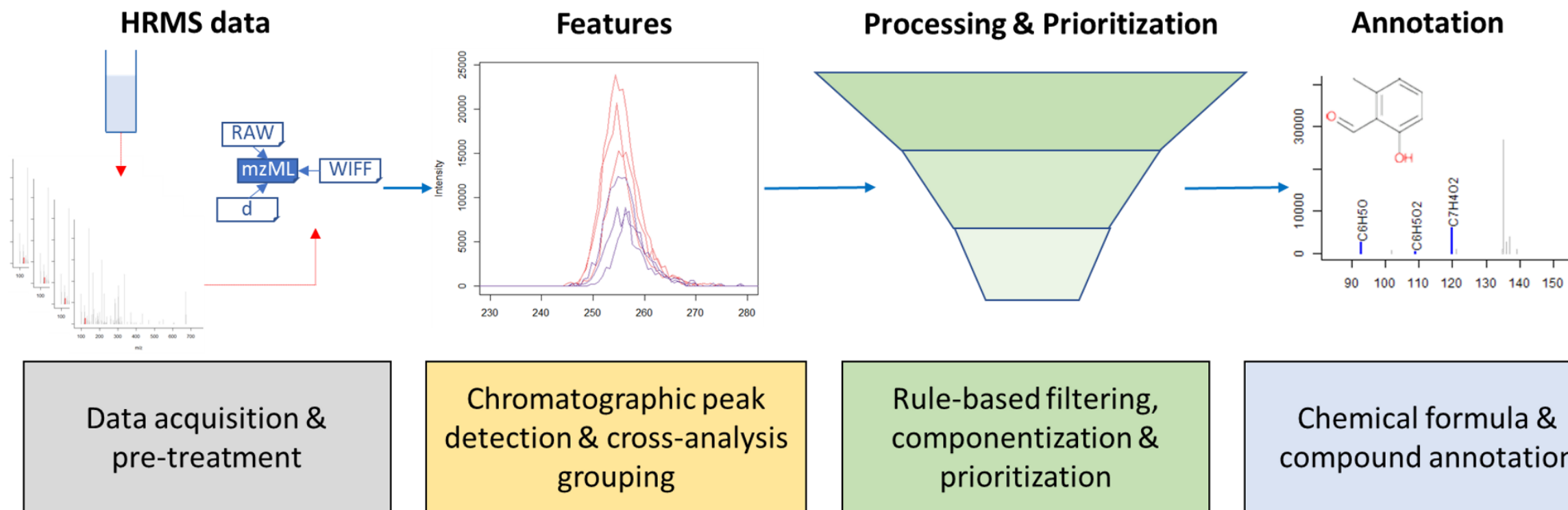
**patRoan: open source software platform for environmental mass spectrometry based non-target screening**

*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

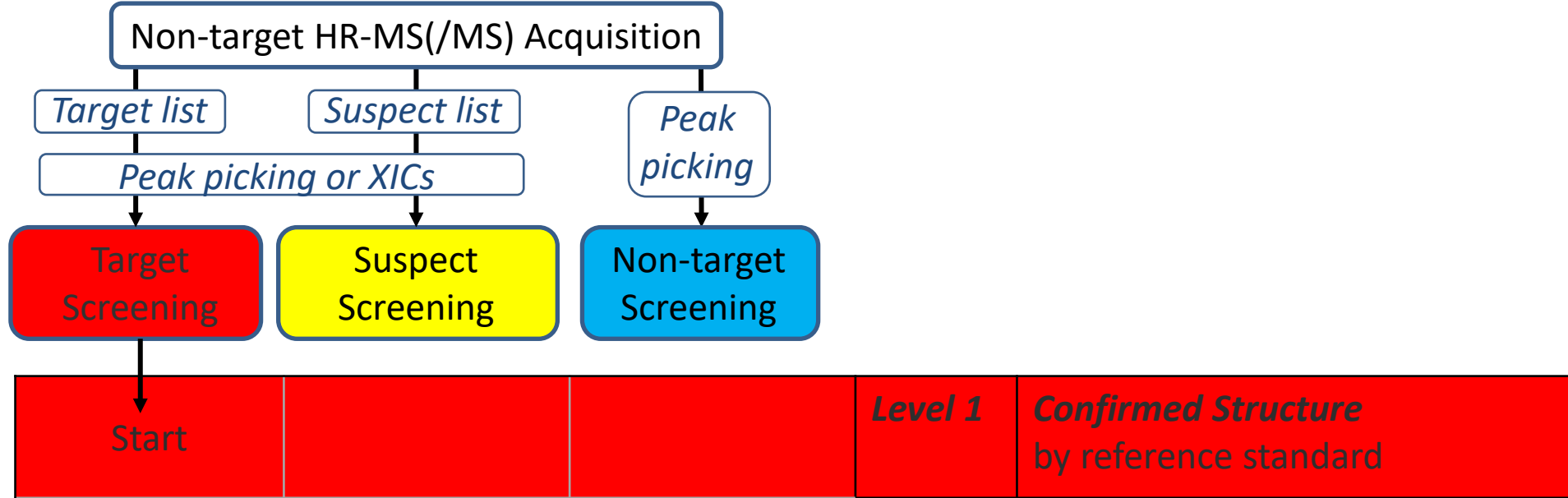
patRoan 2.0: Improved non-target analysis workflows including automated transformation product screening

Rick Helmus<sup>1†</sup>, Bas van de Velde<sup>123</sup>, Andrea M. Brunner<sup>24</sup>, Thomas L. ter Laak<sup>12</sup>, Annemarie P. van Wezel<sup>1</sup>, and Emma L. Schymanski<sup>5</sup>





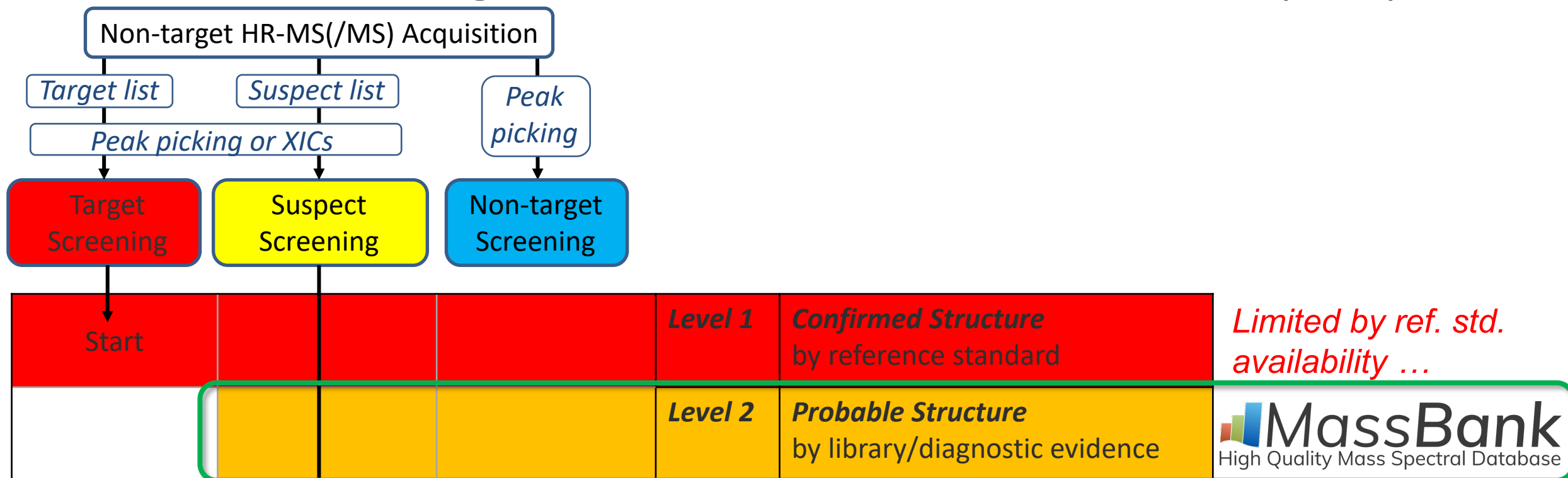
# Identification Strategies and Confidence in NT-HRMS(/MS)



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

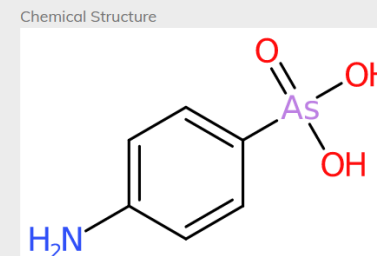
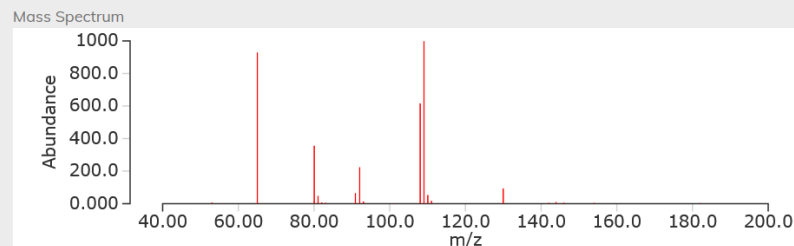


# Identification Strategies and Confidence in NT-HRMS(/MS)



## MassBank Record: LU040605

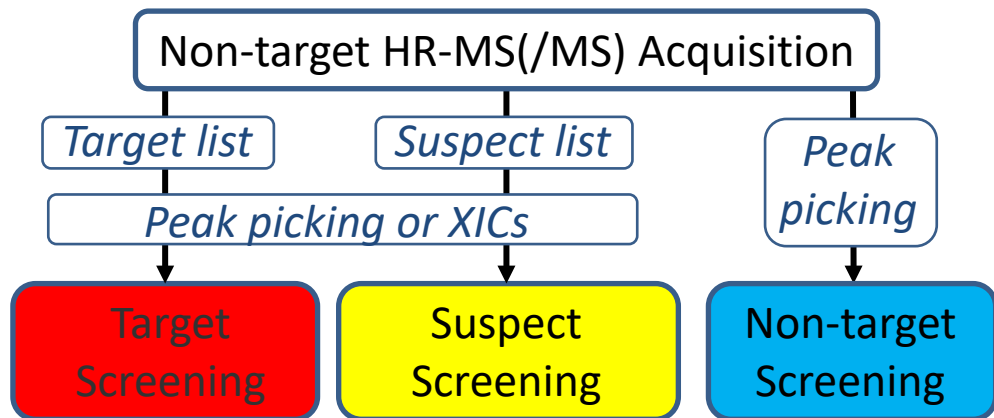
(4-Aminophenyl)arsonic acid; LC-ESI-QFT; MS2; CE: 75; R=17500; [M+H]<sup>+</sup>



**eawag**  
aquatic research



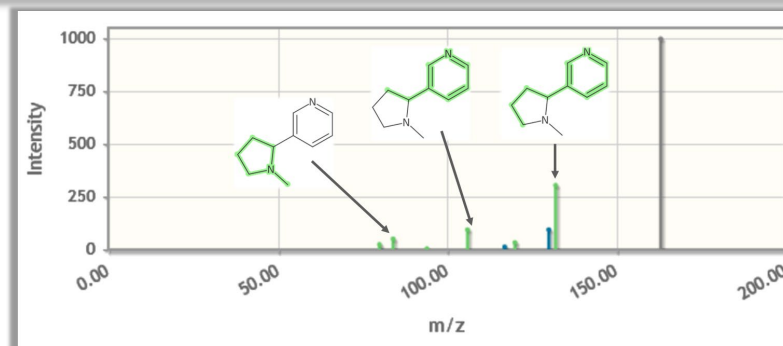
# Identification Strategies and Confidence in NT-HRMS(/MS)



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

Limited by ref. std. availability ...

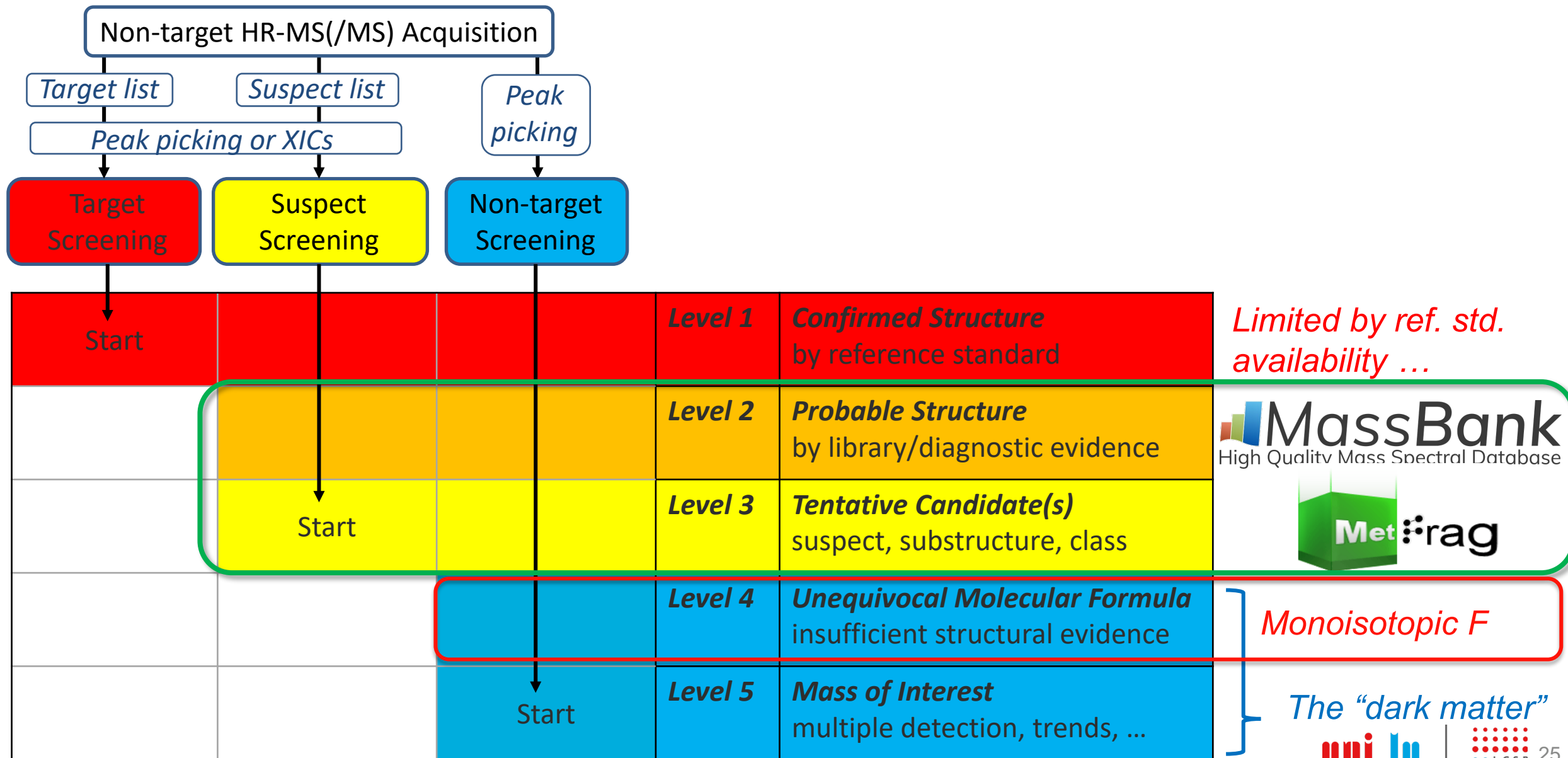
**MassBank**  
High Quality Mass Spectral Database



**eawag**  
aquatic research ooo



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



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



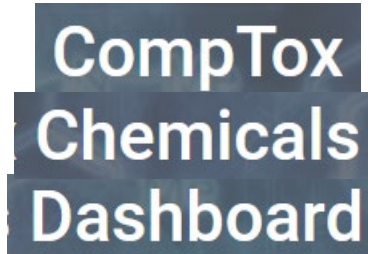
204 million



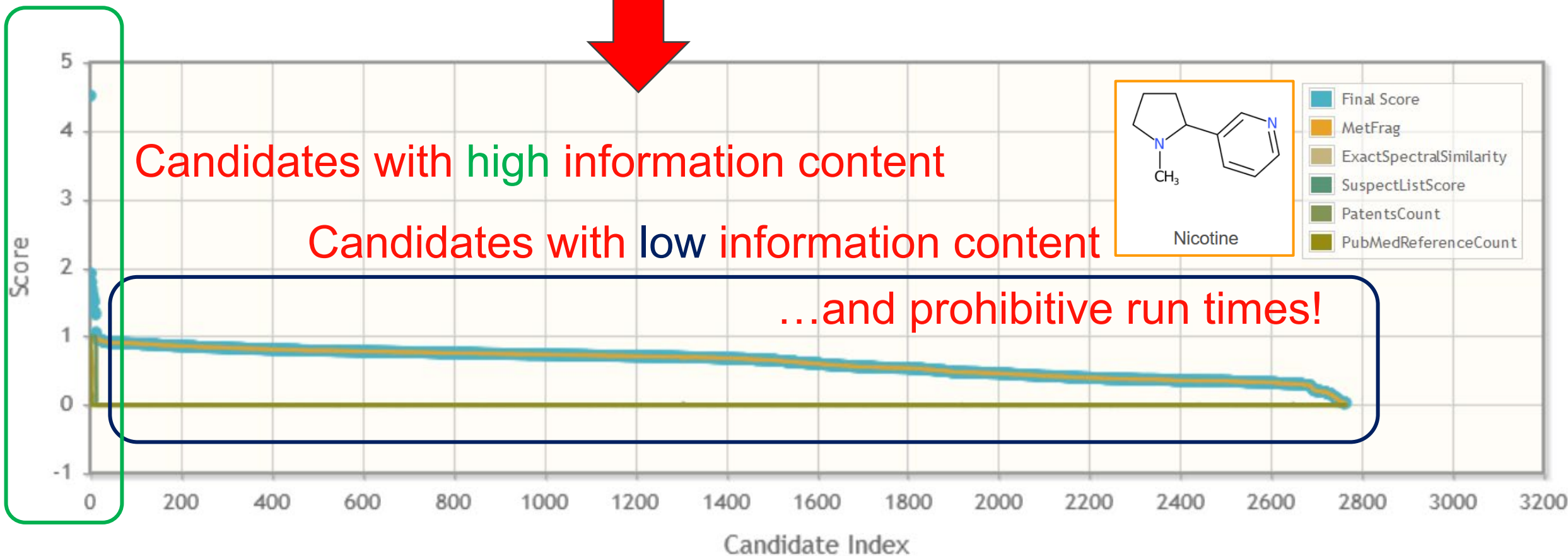
116 million



128 million



1.2 million



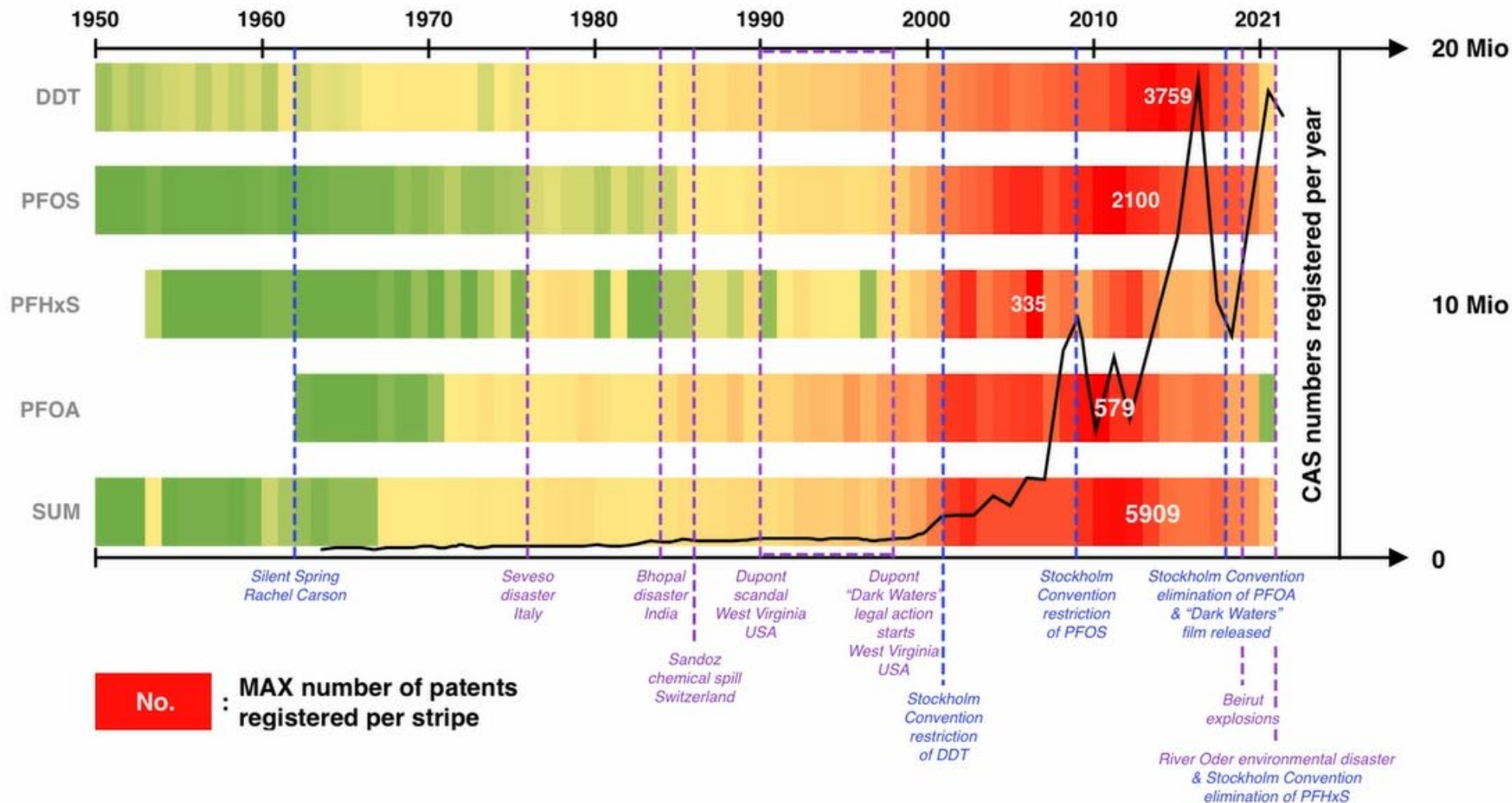


# Our Chemical Past, Present, and Future



JPML

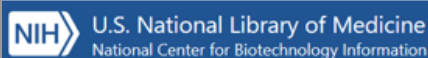
low / medium / high number of patents registered per year (WIPO)



[Flashback 6 months ago ...]

ZeroPM

# Are there really 6 million PFAS in PubChem?



ZeroPM Webinar  
22 March 2023



Assoc. Prof. Dr. Emma Schymanski  
Environmental Cheminformatics, Luxembourg Centre for  
Systems Biomedicine, University of Luxembourg  
Dr. Evan Bolton  
National Center for Biotechnology Information,  
National Library of Medicine, National Institutes of Health

This project has received funding from the European Union's Horizon 2020  
research and innovation programme under grant agreement No 101036756.



DOI:10.5281/  
zenodo.7756622





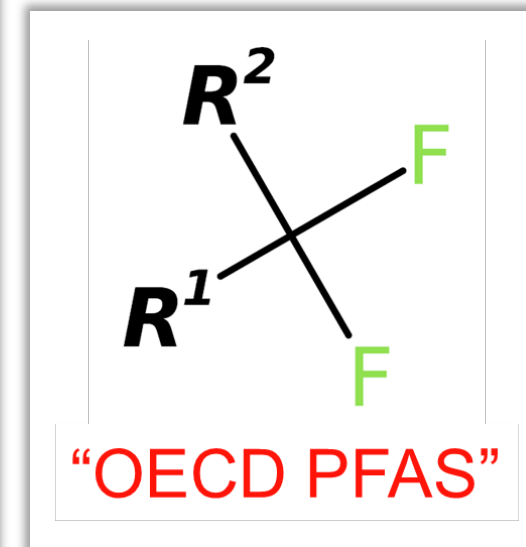
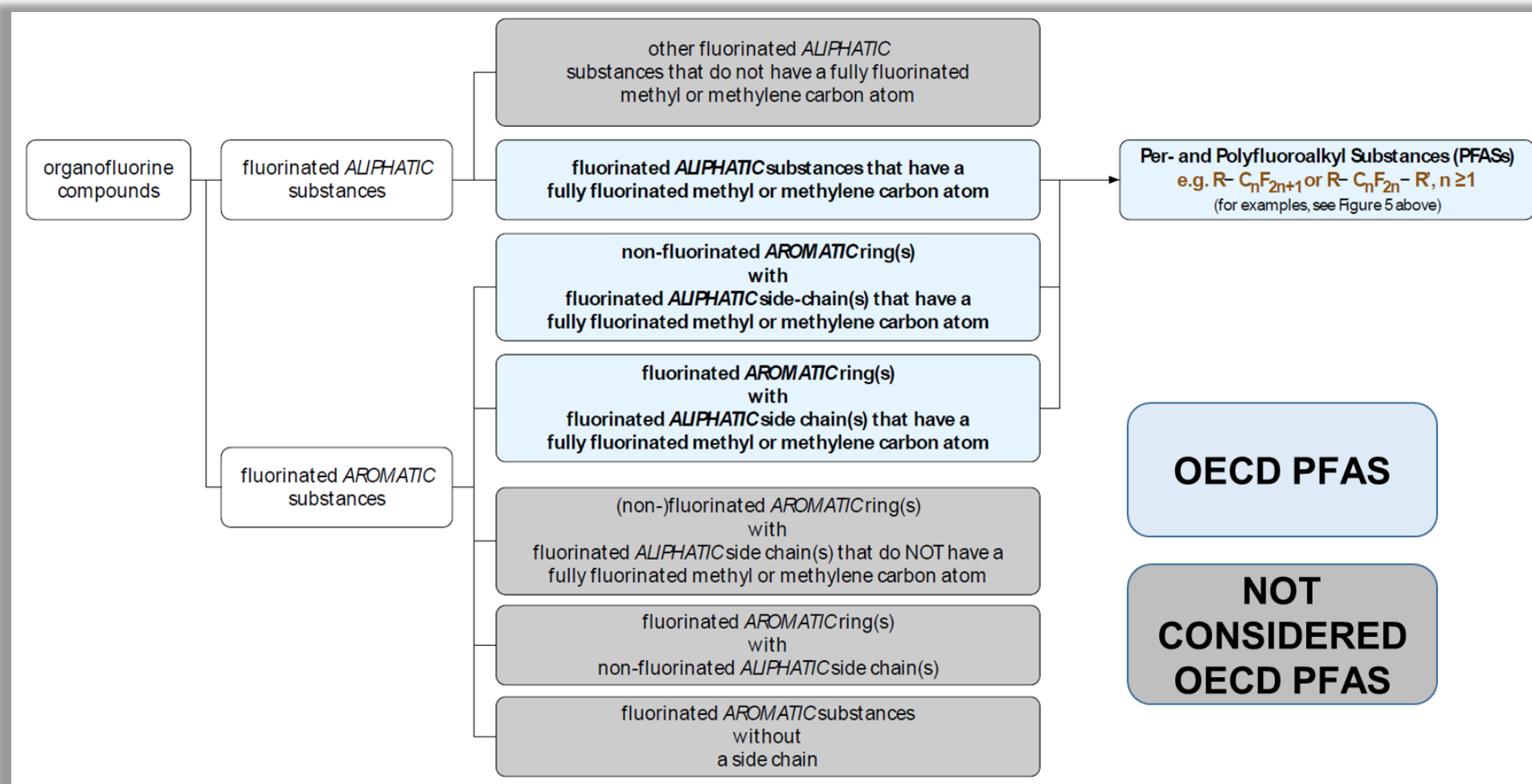
[Flashback 6 months ago ...]

ZeroPM

No...it's already **7 million!**

▼ PFAS and Fluorinated Compounds in PubChem	?	↗	20,929,881
▶ OECD PFAS definition	?	↗	6,370,077
▶ Organofluorine compounds	?	↗	19,963,719
▶ Other diverse fluorinated compounds	?		122,266
▶ PFAS and fluorinated compound collections	?	↗	1,789,330
▶ PFAS breakdowns by chemistry	?		7,299,804
▶ Regulatory PFAS collections	?		26,965

# Motivation: Updated OECD PFAS Definition in 2021



# Motivation: How to scale this to PubChem?

The screenshot shows the PubChem homepage with a search bar containing "PFAS". Below the search bar, there are suggestions for "Try" including "covid-19", "aspirin", "EGFR", "C9H8O4", "57-27-2", "C1=CC=C(C=C1)C=O", and "InChI=1S/C3H6O/c1-3(2)4/h1-2H3". There are also radio buttons for "Use Entrez", "Compounds", "Substances", and "BioAssays". Below these are icons for "Draw Structure", "Upload ID List", "Browse Data", and "Periodic Table". At the bottom, there are statistics: "116M Compounds", "308M Substances", "292M Bioactivities", "36M Literature", and "934 Data Sources".

Explore Chemistry

Quickly find chemical information from authoritative sources

PFAS

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

Use Entrez  Compounds  Substances  BioAssays

Draw Structure Upload ID List Browse Data Periodic Table

116M Compounds 308M Substances 292M Bioactivities 36M Literature 934 Data Sources

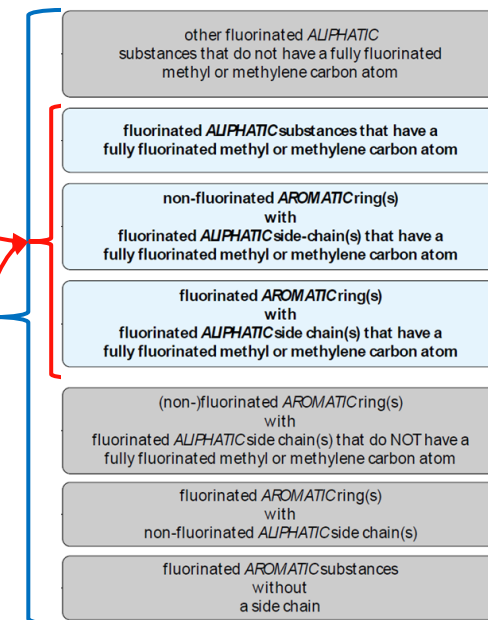
See More Statistics > Explore Data Sources >

# The PubChem PFAS Tree

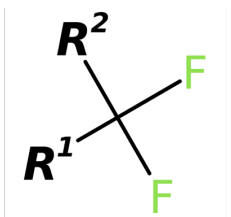


- ▼ PFAS and Fluorinated Compounds in PubChem ? ↗ **21,411,181**
  - ▼ OECD PFAS definition ? ↗ **6,540,217**
    - ▶ Molecule contains isolated CF<sub>2</sub> ? **675,776**
    - ▶ Molecule contains isolated CF<sub>3</sub> ? **5,747,364**
    - ▶ Molecule contains PFAS parts larger than CF<sub>2</sub>/CF<sub>3</sub> ? **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** **“OECD PFAS”**
  - ▶ Regulatory PFAS collections ? **26,943**

**(+salts/mixtures)**



**“OECD PFAS”**



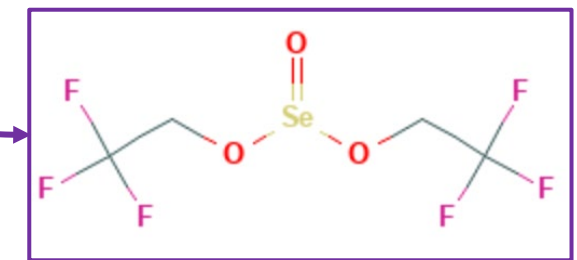
OECD Monograph [ENV/CBC/MONO\(2021\)25](#) (9 July 2021)



# 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 **“OECD PFAS”**
  - ▶ Regulatory PFAS collections ? 26,943 **(+salts/mixtures)**



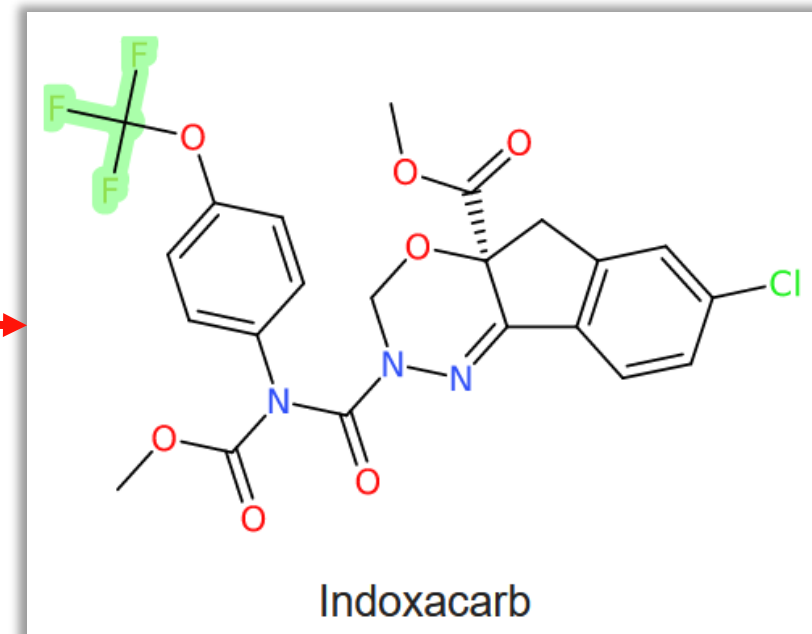
OECD Monograph [ENV/CBC/MONO\(2021\)25](#) (9 July 2021)

# 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 ...*



# PFAS Breakdown by Chemistry

- Breakdown by PFAS functional groups ? 7,497,376
  - Contains PFAS-C ? 6,718,382
    - Contains PFAS-C( 4,402,504
    - Contains PFAS-C(=C 118,192
    - Contains PFAS-C(=N 32,533
    - Contains PFAS-C(=O) 539,557
    - Contains PFAS-C(=S 667
    - Contains PFAS-CH 739,211
    - Contains PFAS-CH2 1,310,256
    - Contains PFAS-CH3 106,486
    - More PFAS-C cases 7,002
  - Contains PFAS-N ? 39,853
  - Contains PFAS-O ? 663,131
  - Contains PFAS-P ? 16,313
  - Contains PFAS-S ? 284,549
  - More PFAS-Element cases ? 5,655

- PFAS-C(=O)-NR2 63,644
- PFAS-C(=O)-OF 406
- PFAS-C(=O)-OH 227,763
- PFAS-C(=O)-OI 244
- PFAS-C(=O)-OR 80,825
- PFAS-C(=O)
- PFAS-C(=O)
- Yet more co

- PFAS-C(=O)-OH 227,763
  - Breakdown by PFAS composition ? 227,763
  - Breakdown by PFAS part connectivity degree ? 227,763
  - Breakdown by PFAS part formulas ? 227,763
    - Molecule contains C01F02 14,031
    - Molecule contains C01F03 210,204
    - Molecule contains C02F04 1,441
    - Molecule contains C02F05 201
    - Molecule contains C03F05 28
    - Molecule contains C03F06 876
    - Molecule contains C03F07 434
    - Molecule contains C04F07 24
    - Molecule contains C04F08 112
    - Molecule contains C04F09 100

# PFAS Breakdown by Chemistry

- Breakdown by PFAS functional groups ? 7,497,376
  - Contains PFAS-C ? 6,718,382
    - Contains PFAS-C( 4,402,504
    - Contains PFAS-C(=C 118,192
    - Contains PFAS-C(=N 32,533
    - Contains PFAS-C(=O) 539,557

- PFAS-C(=O)-NR2 63,644
- PFAS-C(=O)-OF 406
- PFAS-C(=O)-OH 227,763
- PFAS-C(=O)-OI 244
- PFAS-C(=O)-OR 80,825
- PFAS-C(=O) ?
- PFAS-C(=O) ?
- Yet more co

- PFAS-C(=O)-OH 227,763
  - Breakdown by PFAS composition ? 227,763
  - Breakdown by PFAS part connectivity degree ? 227,763
  - Breakdown by PFAS part formulas ? 227,763
    - Molecule contains C01F02 14,031
    - Molecule contains C01F03 210,204
    - Molecule contains C02F04 1,441
    - Molecule contains C02F05 201
    - Molecule contains C03F05 28
    - Molecule contains C03F06 876
    - Molecule contains C03F07 434
    - Molecule contains C04F07 24
    - Molecule contains C04F08 112
    - Molecule contains C04F09 100

SEARCH FOR

PFAS and Fluorinated Compounds in PubChem: Molecule contains C03F07

Treating this as a previously computed list of identifiers.

Compounds

434 results Filters SORT BY Relevance Download

ACTIONS ON RESULTS WITH ID TYPE: Compounds

- Push to Entrez
- Save for Later
- Linked Data Sets

Heptafluorobutyric Acid; 375-22-4; Perfluorobutyric Acid; Perfluorobutanoic Acid; Heptafluorobutanoic Acid; ...

Compound CID: 9777  
MF: C<sub>4</sub>HF<sub>7</sub>O<sub>2</sub> MW: 214.04g/mol  
IUPAC Name: 2,2,3,3,4,4,4-heptafluorobutanoic acid  
Isomeric SMILES: C(=O)C(C(C(F)(F)(F)F)F)O  
InChIKey: YPJUNDFVDDCYIH-UHFFFAOYSA-N  
InChI: InChI=1S/C4HF7O2/c5-2(6,1(12)13)3(7,8)4(9,10)11/h(H,12,13)  
Create Date: 2005-03-26

Summary Similar Structures Search Related Records



# The PubChem PFAS Tree – Collection of Suspect Lists



- ▼ 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
    - ▶ CompTox Chemicals Dashboard PFAS suspect lists ? ↗ 16,120
    - ▶ NORMAN-SLE PFAS suspect lists ? ↗ 6,317
    - ▶ OntoChem PFAS lists ? ↗ 1,777,020
    - ▶ Other fluorinated chemical content in PubChem ? ↗ 1,777
    - ▶ NIST PFAS suspect list ? ↗ 4,948
  - ▶ PFAS breakdowns by chemistry ? 7,497,376
  - ▶ Regulatory PFAS collections ? 26,943



CompTox Chemicals Dashboard



PubChem



# 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 | PNASNTREV19 | 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****



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

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



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

1,232 new CIDs!

# PFAS Suspect Lists – OntoChem+Google / PubChem



▼ 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**



# 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
<b>Total</b>	<b>1,797,831</b>	<b>77,441</b>	<b>36,788</b>



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

# 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<sup>1\*</sup>, Linda G. T. Gaines<sup>2</sup>, Christopher M. Grulke<sup>1†</sup>, Charles N. Lowe<sup>1</sup>, Gabriel F. B. Sinclair<sup>3</sup>, Vicente Samano<sup>4</sup>, Inthirany Thillainadarajah<sup>4</sup>, Bryan Meyer<sup>4</sup>, Grace Patlewicz<sup>1</sup> and Ann M. Richard<sup>1</sup>

Williams *et al.* (2022) DOI:[10.3389/fenvs.2022.850019](https://doi.org/10.3389/fenvs.2022.850019)

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



# Adding Regulatory Collections to the PFAS Tree



POPRC slides available at  
DOI: [10.5281/zenodo.7118551](https://doi.org/10.5281/zenodo.7118551)

## 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

- ▶ Long-chain PFCAs (LC-PFCAs) and related substances ? ↗
- ▶ PFHxS and related substances ? 719
- ▶ PFOA and related substances ? 25,543
- ▶ PFOA and related substances - exclusions ? ↗
- ▶ PFOS and related substances ? 1,307

## PFHxS and related substances ? 719

- ▶ [Annex A] PFHxS plus its salts and PFHxS-related compounds as defined in Annex A of the Stockholm Convention ? ↗ 607
- ▶ [EU REACH] PFHxS (linear or branched) plus its salts and related substances according to EU REACH (draft definition) ? ↗ 719

- ▶ Compounds with a (C6F13)S moiety in PubChem by SMARTS ? 719
- ▶ Compounds with a (C6F13)S(=O)(=O) moiety in PubChem by SMARTS ? 605
- ▶ Difference between Annex A and EU REACH definitions ? 112

Compounds that transform to PFHxS (via PubChem Transformations) ?

Initial indicative list of PFHxS plus its salts and PFHxS-related compounds ? ↗ 76

PFHxS and any branched isomers (included in PubChem) ? 5

PFHxS and any branched isomers and their salts (included in PubChem) ? 62

PFHxS and branched isomer combined substructure query in PubChem ? 212

# Example: PFHxS in Stockholm Convention vs EU REACH

- ▼ Regulatory PFAS collections ? **26,943**
  - ▶ Long-chain PFCAs (LC-PFCAs) and related substances ? **18,416**
  - ▼ PFHxS and related substances ? **719**
    - ▶ [Annex A] PFHxS plus its salts and PFHxS-related compounds as defined in Annex A of the Stockholm Convention ? ↗ **607**
    - ▶ [EU REACH] PFHxS (linear or branched) plus its salts and related substances according to EU REACH (draft definition) ? ↗ **719**
    - ▶ Compounds with a (C6F13)S moiety in PubChem by SMARTS ? **719**
    - ▶ Compounds with a (C6F13)S(=O)(=O) moiety in PubChem by SMARTS ? **605**
    - ▶ Difference between Annex A and EU REACH definitions ? **112**

Compounds that transform to PFHxS (via PubChem Transformations) ?

Initial indicative list of PFHxS

PFHxS and any branched isomer

PFHxS and any branched isomer

PFHxS and branched isomer

▼ Difference between Annex A and EU REACH definitions ? **112**

PFHxS in EU REACH but not Annex A - all ? **112**

PFHxS in EU REACH but not Annex A - annotation 'Literature', 'Use', 'Safety', 'Toxicity' ? **14**

PFHxS in EU REACH but not Annex A - annotation 'Use and Manufacturing' ? **5**

PFHxS in EU REACH but not Annex A - annotation 'Use and Manufacturing', 'Literature' ? **14**



# 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	cmpdsyn	mw	mf	polararea	complexit	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	CC(C(C(C(F)	C(C(C(C(F)	GVPQTQIC	pentafluor	445.943	445.943
3	2827766	4-methyl-	NULL	597.4	C19H12F13	79.8	895	7.9	37	1	17	9	InChI=1S/C	CC1=CC=C(C	CC1=CC=C(C	XVAIFAXE	4-methyl-	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	KBFCJWZM	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)	OCCOAYPC	1,1,1,2,2,3,	701.903	701.903
9	10092845	Butyl(tride	SCHEMBLS	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)	IYYLVCKNC	1,1,2,2,3,3,	383.949	383.949
11	11048550	1-Fluoro-4-	SCHEMBLS	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	ICKCZVFDU	1-ethenyls	377.975	377.975
13	11327892	1-(2-Chloro	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-chloro	429.946	429.946
14	11338428	2-(1,1,2,2,	NULL	396.17	C8H5F13O2S	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,	395.985	395.985
15	11362001	1-(2-Chloro	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-chloro	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	SFIGAKQSM	(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	SFIGAKQSM	(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-Methyl-	(4-Methyl-	442.24	C13H7F13S	25.3	517	7.4	27	0	14	6	InChI=1S/C	CC1=CC=C(C	CC1=CC=C(C	MBZVAVG	1-methyl-4	442.006	442.006
23	14544789	1,1,1,2,2,3,	SCHEMBLS	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

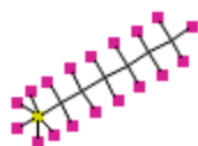


# Download files contain additional information...

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	cmpdsyn	mw	mf	polararea	complexit	xlogp	heavycnt	hbondon	hbondacc	rotbonds	inchi	isosmiles	canonicals	inchikey	iupacname	exactmass	monoisot	
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-1	NULL	597.4	C19H12F13	79.8	895	7.9	37	1	17	9	InChI=1S/C	CC1=CC=C	(CC1=CC=C	(XVAIFAXEE	4-methyl-1	597.01	597.01	
4	2827776	4,6-dichlor	NULL													UMFM	4,6-dichlor	589.94	589.94	
5	2827778	6-chloro-2- N														RCGPU	6-chloro-2-	570.99	570.99	
6	2827780	2-N,2-N-di N														CJWZM	2-N,2-N-di	580.071	580.071	
7	3021589	Tridecaflu	T													WVNV	1,1,2,2,3,3,	385.92	385.92	
8	3023059	Bis(trideca	B													OAYPC	1,1,1,2,2,3,	701.903	701.903	
9	10092845	Butyl(tride	S													LAJRV	1-butylsulf	408.022	408.022	
10	10971025	Tridecaflu	T															3,	383.949	383.949
11	11048550	1-Fluoro-4 S																-4	445.981	445.981
12	11245889	1-Ethenyls N																yls	377.975	377.975
13	11327892	1-(2-Chloro	N															orc	429.946	429.946
14	11338428	2-(1,1,2,2,3	N															2,3	395.985	395.985
15	11362001	1-(2-Chloro	N															orc	413.951	413.951
16	11990355	Copper;4,4 N																4,4	2821	2821
17	11990356	1,4,8,11-Te	1															te	1776.08	1776.08
18	12635301	1-Chloro-4 N																-4	461.951	461.951
19	12996310	[(1beta,4b	[(															R)	460.017	460.017
20	12996311	(1R,4R,5S)-	N															S)	460.017	460.017
21	13213429	[(Tridecafl	[(															,3,	427.99	427.99
22	13410186	(4-Methyl	(4															1-4	442.006	442.006
23	14544789	1,1,1,2,2,3, S																,3	365.975	365.975



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

Compound CID: 15017

MF: C<sub>6</sub>F<sub>18</sub>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

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

# 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	neighborborty	meshhead	annothis	annothiscr	ids	cidcdate	sidsrcname	depcatg	annotation
2	15027	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	282776	0	0	0	2D+3D	NULL	Classification	1	NULL	20050728	ABI Chem Chem-Space.com	Chemical Vendors Legacy Depositors Rese	NULL
4	282776	0	0	0	2D	NULL	Classification	1	NULL	20050728	ABI Chem Aurora Fine Chem	Chemical Vendors Legacy Depositors Rese	NULL
5	282778	0	0	0	2D	NULL	Classification	1	NULL	20050728	ABI Chem Chem-Space.com	Chemical Vendors Legacy Depositors Rese	NULL
6	282770	0	0	0	2D	NULL	Classification	1	NULL	20050728	ABI Chem Chem-Space.com	Chemical Vendors Legacy Depositors Rese	NULL
7	302159	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	302309	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 Acco	Chemical Vendors Curation Efforts Govern	NULL
11	11048500	1	1	1	2D+3D	NULL	Classification Literature Patents	3	NULL	20061026	ChemSpider DiscoveryGate	Curation Efforts Journal Publishers Legacy	NULL
12	1124589	0	0	0	2D	NULL	Classification	1	NULL	20061026	ChemSpider DiscoveryGate	Journal Publishers Legacy Depositors Rese	NULL
13	11327802	0	0	0	2D+3D	NULL	Classification	1	NULL	20061026	ChemSpider DiscoveryGate	Journal Publishers Legacy Depositors Rese	NULL
14	1133848	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	11990335	0	0	0	NULL	NULL	Classification	1	NULL	20070205	ChemSpider DiscoveryGate	Journal Publishers Legacy Depositors Rese	NULL
17	11990336	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	13410136	0	0	0	2D+3D	NULL	Classification	2	NULL	20070208	Japan Chemical Substance D	Governmental Organizations	NULL
23	14544739	0	5	2	2D+3D	NULL	Classification Patents	3	NULL	20070209	ECI Group, LCSB, University	Curation Efforts Research and Developmen	NULL



# 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
cid	pclidcnt	gpidcnt	gpfamilycr	neighborb	meshhead	annothits	annothitcr	cidcdate	sidsrcname	
15017	0	17	5	2D	NULL	Classification Patents Toxicity Use and Manu	4	NULL	20050808	ABI Chem BenchChem C
282776	0	0	0	2D+3D	NULL	Classification	1	NULL	20050728	ABI Chem Chem-Space.co
282776	0	0	0	D	NULL	Classification	1	NULL	20050728	ABI Chem Aurora Fine Ch
282778	0	0	0	D	NULL	Classification	1	NULL	20050728	ABI Chem Chem-Space.co
282778	0	0	0	D	NULL	Classification	1	NULL	20050728	ABI Chem Chem-Space.co
302159	0	0	0	D	NULL	Classification	3	NULL	20050808	ABI Chem Achemica Alfa
302309	0	0	0	D	NULL	Classification	2	NULL	20050808	ABI Chem Alfa Chemistry
100928	0	0	0	D	NULL	Classification	2	NULL	20061025	ChemSpider DiscoveryGa
109710	0	0	0	D	NULL	Classification	4	NULL	20061025	A2B Chem AA BLOCKS Ac
110485	0	0	0	D	NULL	Classification	4	NULL	20061025	ChemSpider DiscoveryGa
112458	0	0	0	D	NULL	Classification	4	NULL	20061025	ChemSpider DiscoveryGa
113278	0	0	0	D	NULL	Classification	4	NULL	20061025	ChemSpider DiscoveryGa
113384	0	0	0	D	NULL	Classification	4	NULL	20061025	ChemSpider DiscoveryGa
113620	0	0	0	D	NULL	Classification	4	NULL	20061025	ChemSpider DiscoveryGa
119903	0	0	0	D	NULL	Classification	4	NULL	20061025	ChemSpider DiscoveryGa
119903	0	0	0	D	NULL	Classification	4	NULL	20061025	ChemSpider DiscoveryGa
126353	0	0	0	D	NULL	Classification	4	NULL	20061025	ChemSpider DiscoveryGa
129963	0	0	0	D	NULL	Classification	4	NULL	20061025	ChemSpider DiscoveryGa
129963	0	0	0	D	NULL	Classification	4	NULL	20061025	ChemSpider DiscoveryGa
132134	0	0	0	D	NULL	Classification	4	NULL	20061025	ChemSpider DiscoveryGa
134101	0	0	0	D	NULL	Classification	4	NULL	20061025	ChemSpider DiscoveryGa
145447	0	0	0	D	NULL	Classification	4	NULL	20061025	ChemSpider DiscoveryGa

## 8 Use and Manufacturing

### 8.1 Uses

EPA CPDat Chemical and Product Categories

1 item

#### Category

used as a stain or water

*The Chemical and Physical Scientific Data, volume*

► EPA Chemical and Product Categories

### 8.2 Methods of Manufacturing

Perfluoroalkanesulfonyl fluorides are which a hydrocarbon sulfonyl fluoride ... The electrochemical yield is excellent with the increasing length of the carb Alkaline hydrolysis of perfluoroalkane acidified and distilled from concentrated Perfluoroalkanesulfonic Acids/

*Siegemund G et al; Fluorine Compounds, NY, NY: John Wiley & Sons. Online Posting*

► Hazardous Substances Data Bank (HSDB)

### 8.3 U.S. Production

Production volumes for non-confidential chemicals reported under the Inventory Update Rule.

Year	Production Range (pounds)
1986	No Reports
1990	No Reports
1994	10 thousand - 500 thousand
1998	No Reports
2002	10 thousand - 500 thousand

US EPA; Non-confidential Production Volume Information Submitted by Companies for Chemicals Under the 1986-2002 Inventory Update Rule (IUR). 1-Octanesulfonic acid, 1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,8-heptadecafluoro- (1763-23-1). Available from, as of November 2, 2010: <https://www.epa.gov/oppt/iur/tools/data/2002-vol.html>

► Hazardous Substances Data Bank (HSDB)

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

# Integration of MetFrag and PubChem PFAS Tree



MetFrag

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

In silico fragmentation for computer assisted identification of metabolite mass spectra

Slides available at DOI:  
[10.5281/zenodo.6461325](https://doi.org/10.5281/zenodo.6461325)

Database Settings

Database: PubChem\_OECDPFAS\_le

Neutral Mass: 789.98232 Search ppm: 5

Formula:

Identifiers:

Retrieve Candidates 15 Candidates

Weights

- MetFrag (1st)
- ExactSpectralSimilarity (2nd)
- AnnotHitCount (3rd)
- Patent\_Count (4th)
- PubMed\_Count (5th)

#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1		14550408 InChIKeyBlock1 = ZDYYWMSMLMTXDM	789.98233	C <sub>16</sub> H <sub>9</sub> F <sub>26</sub> O <sub>4</sub> P		3.9607	Peaks: 2 / 2 Fragments Scores Download
2		87318203 InChIKeyBlock1 = KYZFNWUVSNMKOP	789.98233	C <sub>16</sub> H <sub>9</sub> F <sub>26</sub> O <sub>4</sub> P		1.4607	Peaks: 2 / 2 Fragments Scores Download
3		121302506 InChIKeyBlock1 = CHLHGUCQTZMWTA	789.98233	C <sub>16</sub> H <sub>9</sub> F <sub>26</sub> O <sub>4</sub> P		1.4607	Peaks: 2 / 2 Fragments Scores Download

SETAC EUROPE 32<sup>ND</sup> 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](https://doi.org/10.1186/s13321-016-0115-9)





# Chemical Stripes in R - for patents & literature

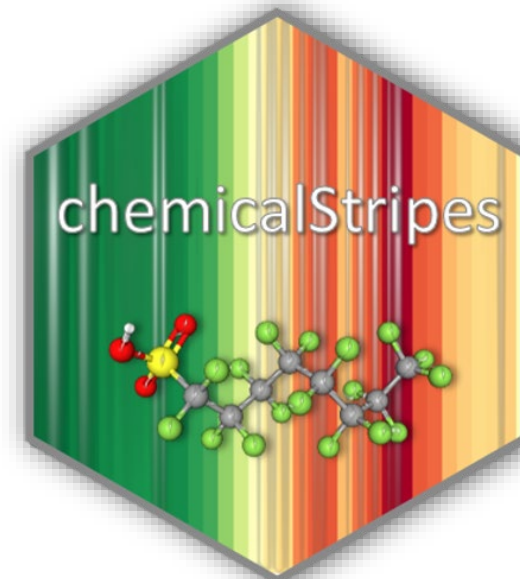
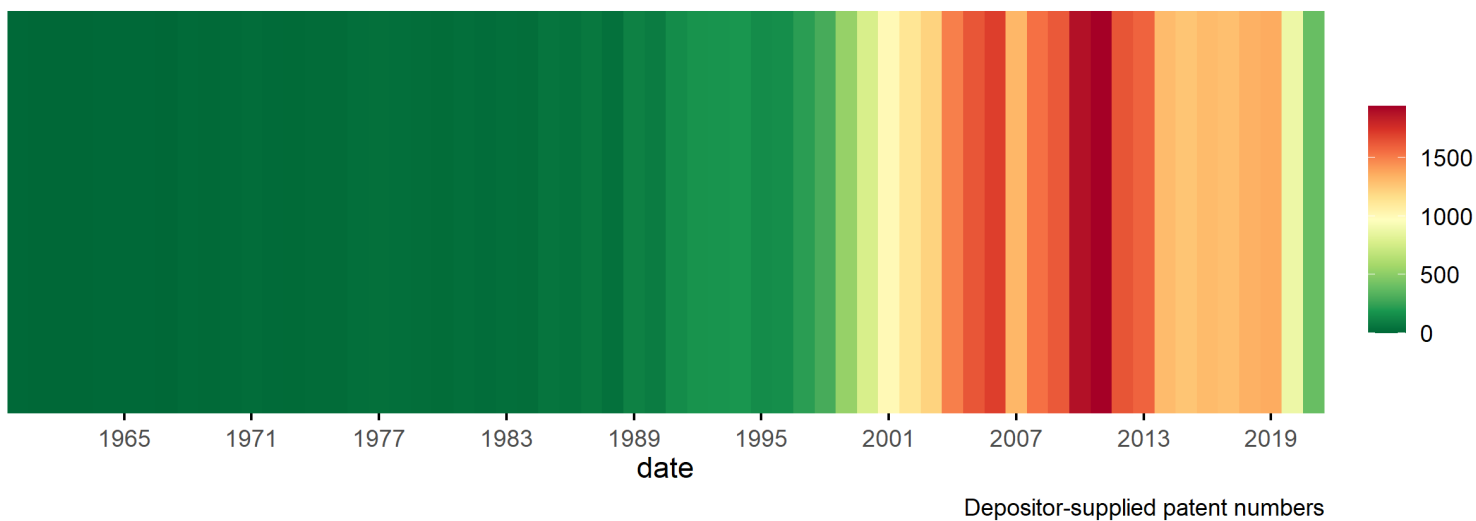


```
> chemical_stripes(74483)
Getting compound information
A total of 32461 patents were found for CID 74483
[=====]----->
Downloading patent data...
[=====]-----> 60% 7s
Processing patent data
32460 patents were processed for CID 74483
[=====]-----> 80% 3s
Plotting chemical stripes for the years between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

## Chemical Stripes for Perfluorooctanesulfonic acid

PubChem CID: 74483  
IUPACName: 1,1,2,2,3,3,4,4,5,5,6,6,7,7,8,8,8-heptadecafluorooctane-1-sulfonic acid  
Molecular Formula: C<sub>8</sub>HF<sub>17</sub>O<sub>3</sub>S  
Exact Mass: 499.9374938



# Chemical Stripes in R

## Patents & literature

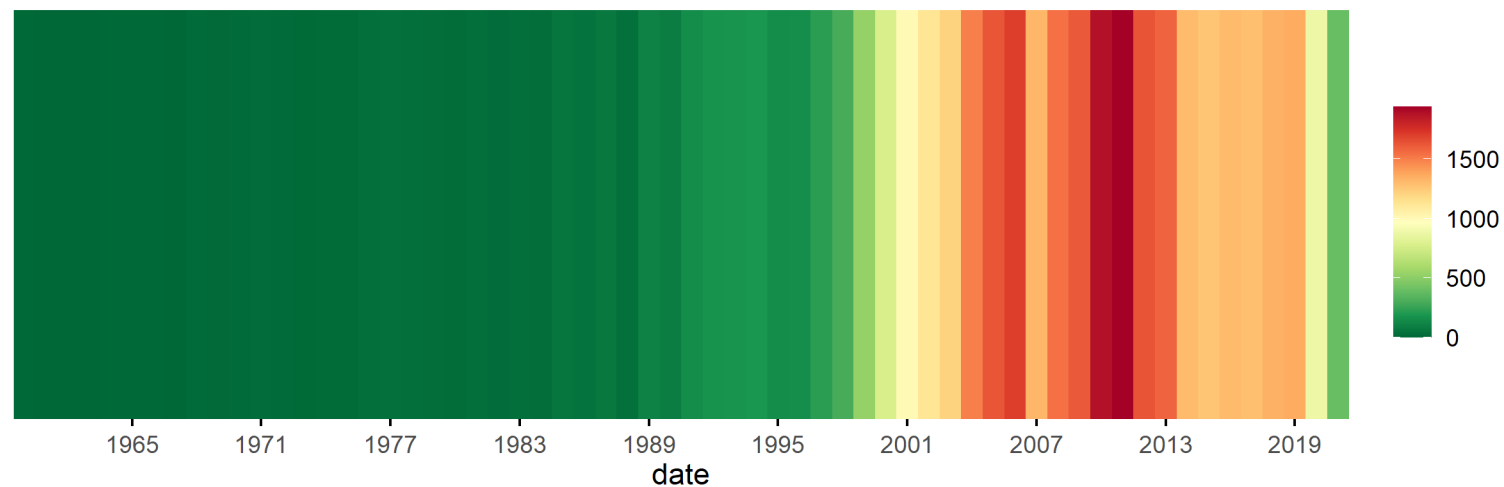


### Chemical Stripes for Perfluorooctanesulfonic acid

PubChem CID: 74483

First patent: 1913

Patents



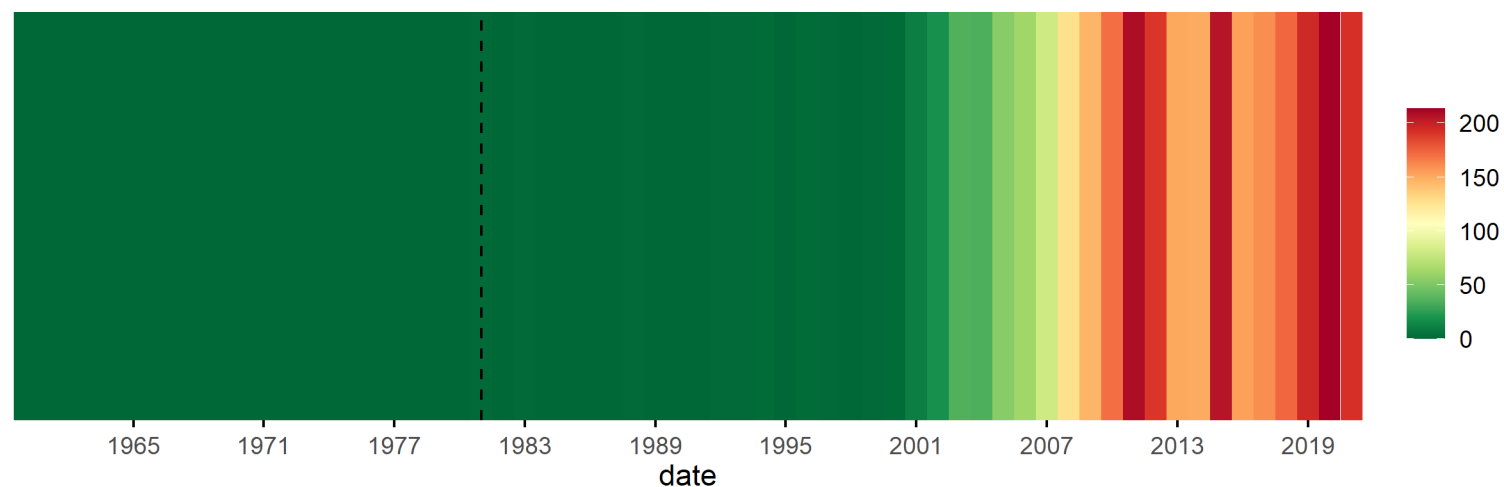
Depositor-supplied patent numbers

### Chemical Stripes for Perfluorooctanesulfonic acid

PubChem CID: 74483

First reference: 1981

Literature



Consolidated reference numbers

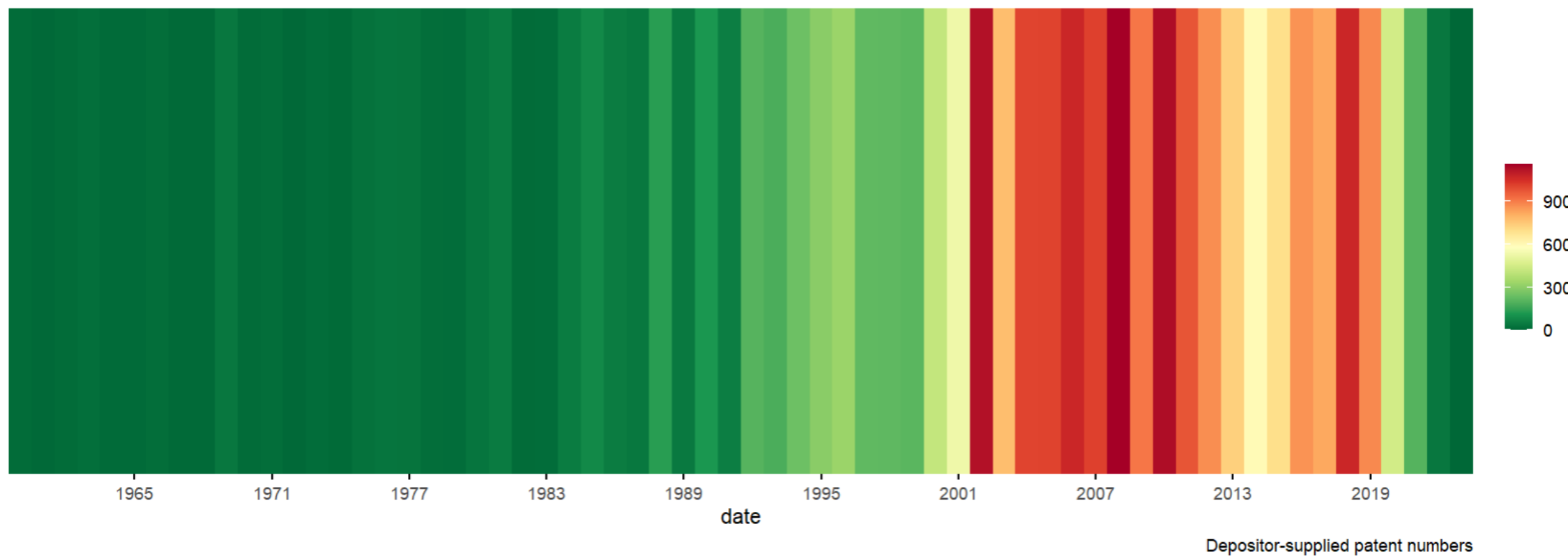
# “Summarized” Chemical Stripes in R



## Summarized Chemical Stripes - EU REACH C9-C14 PFCAs

Patent data compiled from 230 CIDs

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



# The Chemical Stripes and Patent Data

(live stripe calculations during SETAC 2023)

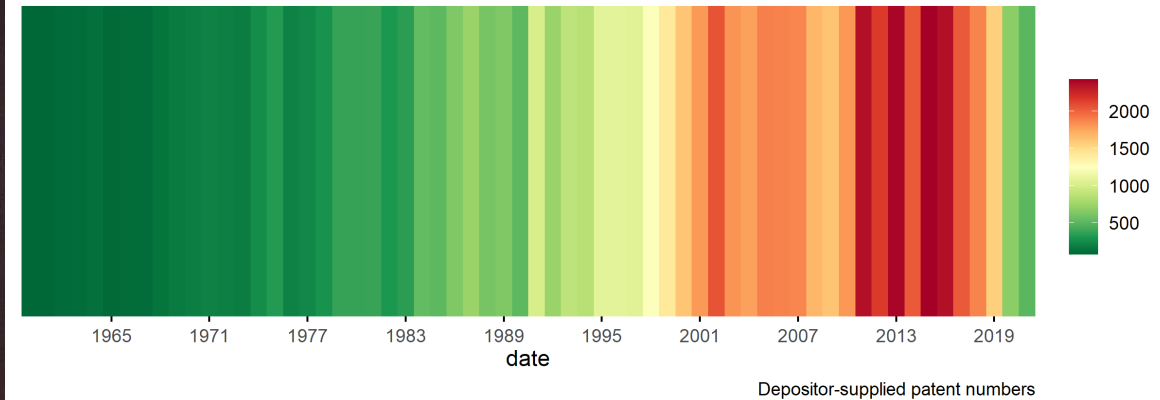


**Release of PM-chemicals into receiving waters**  
High concentration chemicals

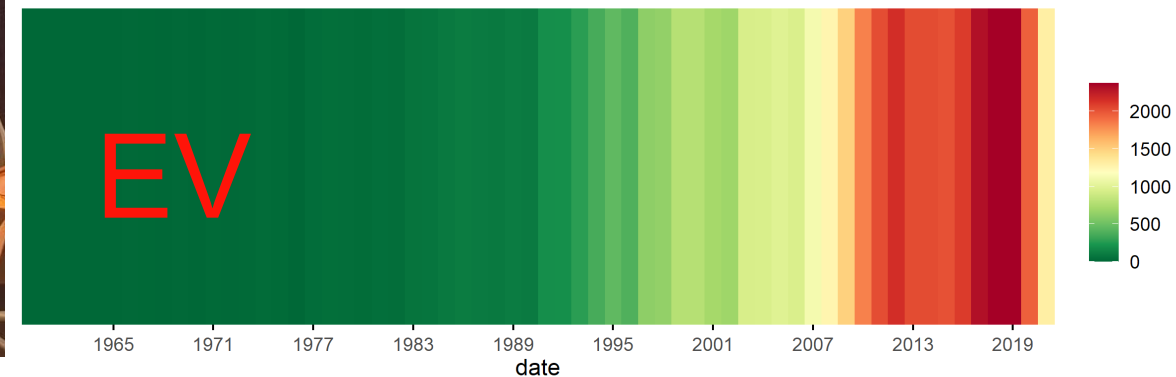
... some high concentration PM-chemicals where not/only partially removed by at least one method

Neuwald et al. Science of the Total Environment (2023) accepted  
www.ufz.de 14

## Chemical Stripes for Tetrafluoroboric acid



## Chemical Stripes for Lithium tetrafluoroborate



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

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



# Read/see more about the PubChem PFAS Tree



PFAS and Fluorinated Compounds in PubChem ? ↗ 21,411,181

OECD PFAS definition ? ↗ 6,540,217

Organofluorine compounds ? ↗ 20,417,011

## PFAS and Fluorinated Compounds in PubChem Tree

Emma L. Schymanski<sup>1\*</sup>, Parviel Chirsir<sup>1</sup>, Todor Kondic<sup>1</sup>,  
Paul A. Thiessen<sup>2</sup>, Jian Zhang<sup>2</sup> and Evan E. Bolton<sup>2\*</sup>

8 September 2023

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

<sup>2</sup> 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](mailto:evan.bolton@nih.gov). ORCID: 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](mailto: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<sup>®</sup>

[How To Submit](#) [Browse](#) [About](#) [News](#)

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

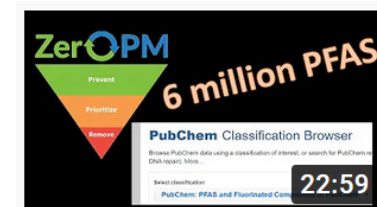
22 June 2023, Version 1

[DOI: 10.26434/chemrxiv-2023-j823z](https://doi.org/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

- 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

Download

ACTIONS ON RESULTS WITH ID TYPE:  
Compounds

- Push to Entrez
- Save for Later
- Linked Data Sets

Download

ACTIONS ON RESULTS WITH ID TYPE:  
Compounds

- Push to Entrez
- Saved as *MassBank EU*
- Linked Data Sets

# Advanced Queries

Schymanski *et al.* (2023) DOI: [10.26434/chemrxiv-2023-j823z](https://doi.org/10.26434/chemrxiv-2023-j823z)

“Saved Searches” option

- 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**

- 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**

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
<a href="#">OECD PFAS in PubChem AND Agrochemicals AND Collision Cross Section (CCS)</a>	CID	27	7 hours	<a href="#">View Results</a> <a href="#">Delete</a>
<a href="#">OECD PFAS in PubChem AND Agrochemicals AND MassBank EU</a>	CID	71	7 hours	<a href="#">View Results</a> <a href="#">Delete</a>
<a href="#">OECD PFAS in PubChem AND Agrochemicals</a>	CID	306	7 hours	<a href="#">View Results</a> <a href="#">Delete</a>
<a href="#">MassBank EU</a>		16,255	7 hours	<a href="#">View Results</a> <a href="#">Delete</a>
<a href="#">Collision Cross Section (CCS)</a>		6,564	7 hours	<a href="#">View Results</a> <a href="#">Delete</a>
<a href="#">Agrochemicals</a>		3,135	7 hours	<a href="#">View Results</a> <a href="#">Delete</a>
<a href="#">OECD PFAS in PubChem</a>		7,497,376	7 hours	<a href="#">View Results</a> <a href="#">Delete</a>

**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)<sup>†</sup>**

Juliane Glüge, <sup>id</sup>\*<sup>a</sup> Martin Scheringer, <sup>id</sup><sup>a</sup> Ian T. Cousins, <sup>id</sup><sup>b</sup> Jamie C. DeWitt,<sup>c</sup> Gretta Goldenman,<sup>d</sup> Dorte Herzke, <sup>id</sup><sup>e,f</sup> Rainer Lohmann, <sup>id</sup><sup>g</sup> Carla A. Ng, <sup>id</sup><sup>h</sup> Xenia Trier<sup>i</sup> and Zhanyun Wang<sup>j</sup>

Glüge *et al.* (2020) ESPI, DOI: [10.1039/d0em00291g](https://doi.org/10.1039/d0em00291g)

**PubChem** 13C3-PFHxS (Compound)

### 3.2.1 Collision Cross Section

150.51 Å<sup>2</sup> [M-H]<sup>-</sup> [CCS Type: DT; Buffer gas: N<sub>2</sub>; Dataset: PFAS]

DOI: [10.1021/acs.est.2c00201](https://doi.org/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!

**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

Parviel Chirsir  
[@PChirsir](https://twitter.com/PChirsir)

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

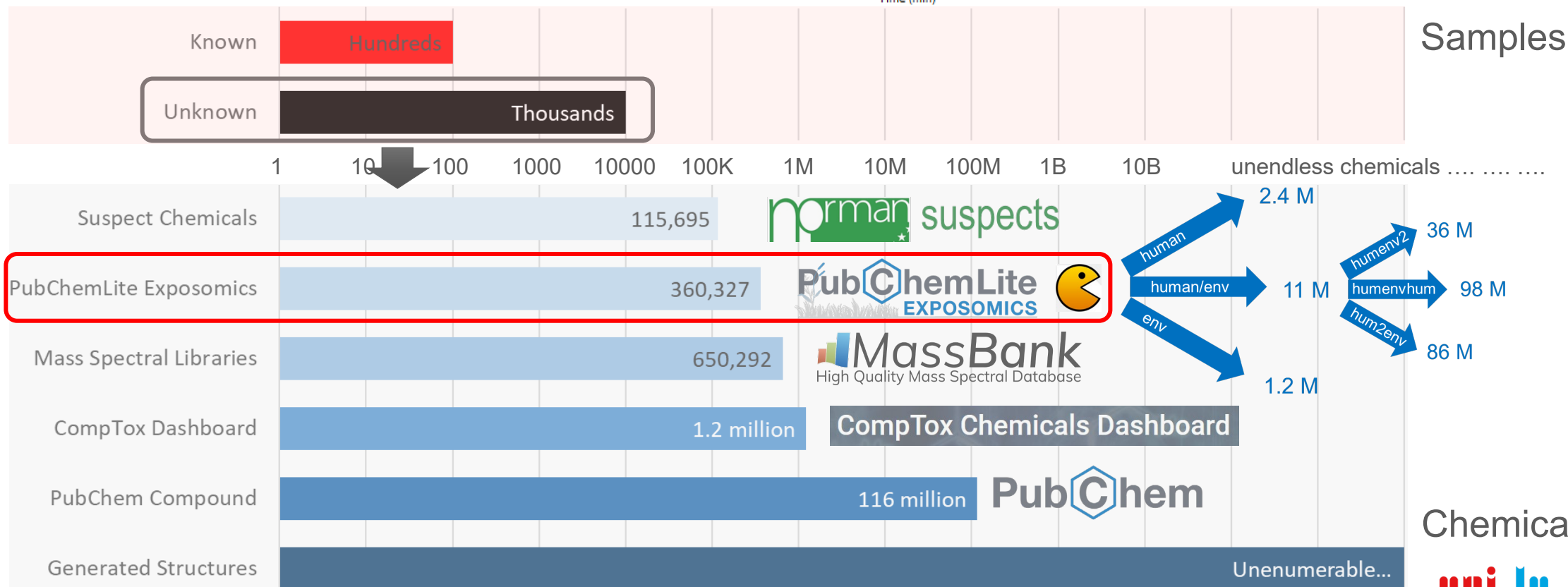
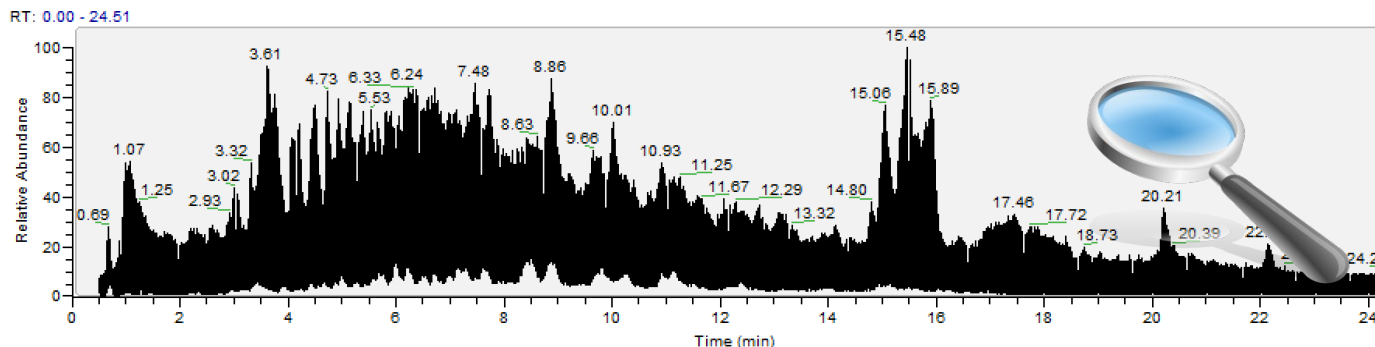
Bugsel *et al.* (2022) ABC, 414, 1217-1225. DOI: [10.1007/s00216-021-03463-9](https://doi.org/10.1007/s00216-021-03463-9)  
<https://pubchem.ncbi.nlm.nih.gov/compound/156620404#section=1H-NMR-Spectra>



# Environmental Cheminformatics, NT HR-MS & *Transformations*

High resolution  
mass spectrometry

AND connecting  
chemical knowledge



# Transforming PubChemLite with **BioTransformer 3.0**

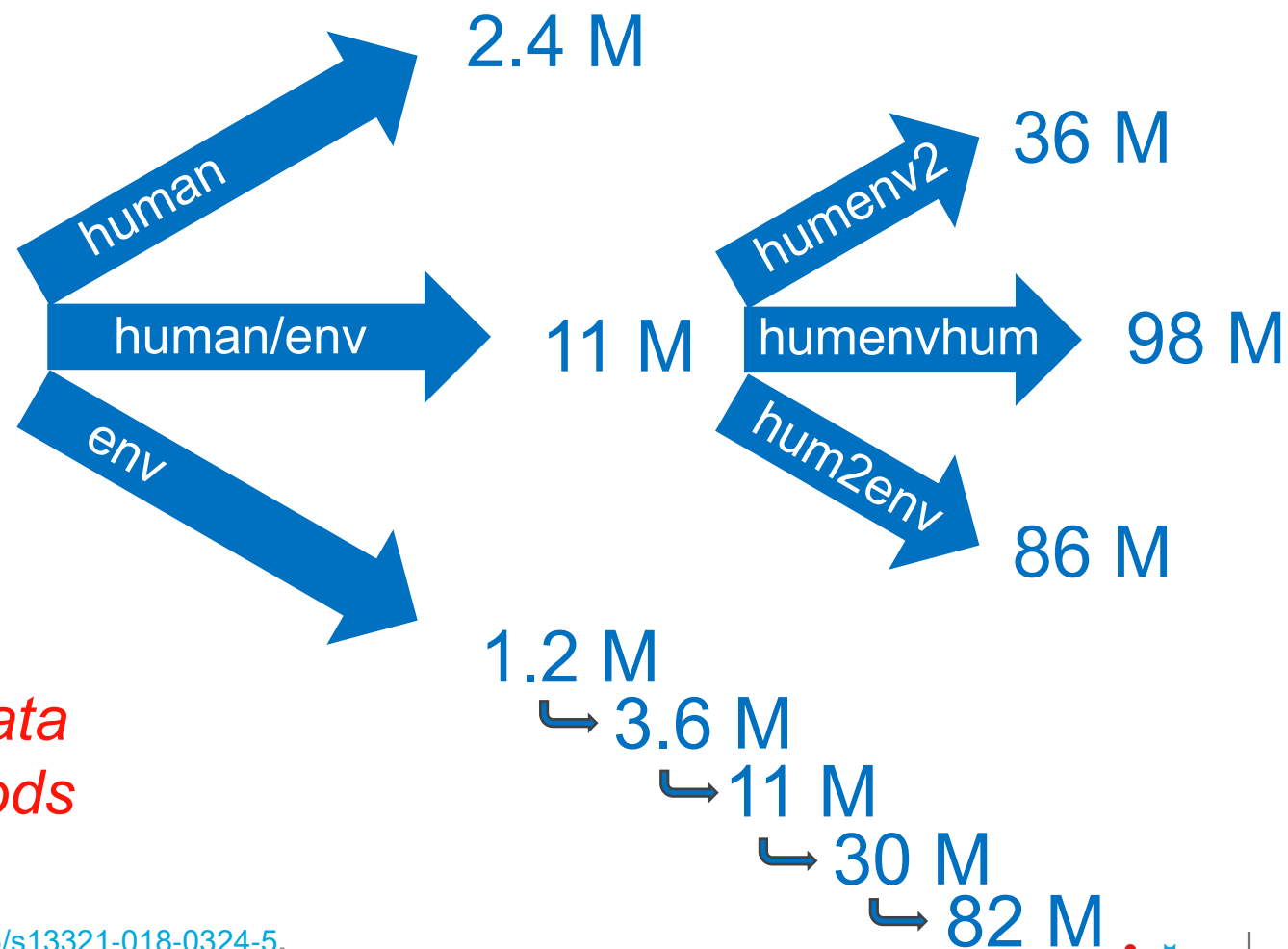
<http://biotransformer.ca/>

**PubChemLite**  
EXPOSOMICS

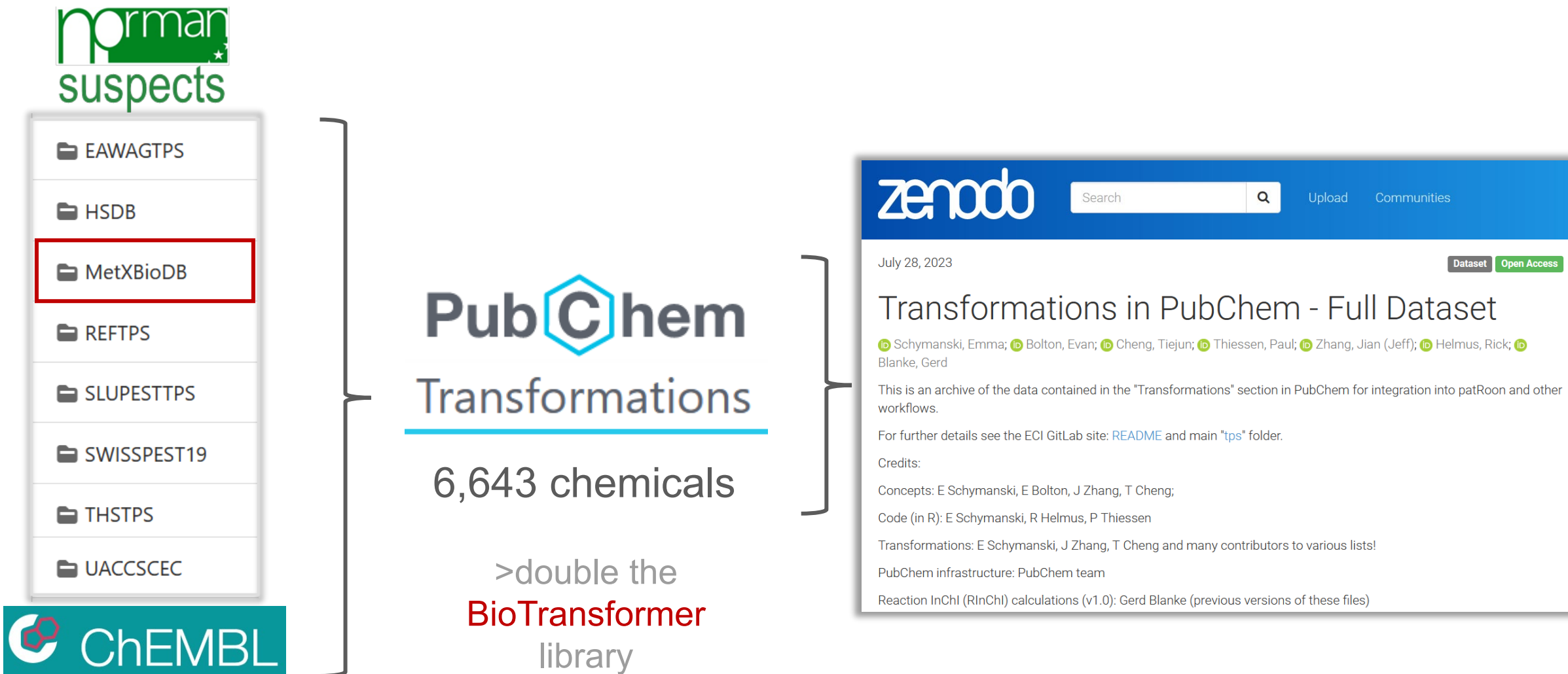
379,199 entries

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

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



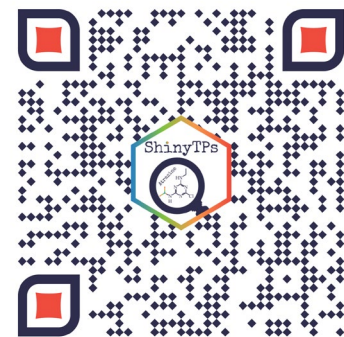
# FAIR Transformations in PubChem and NORMAN-SLE



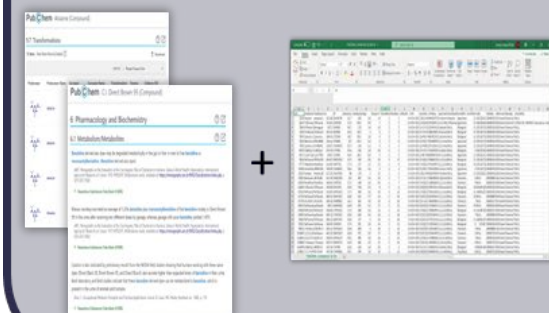




# ShinyTPs: Curating TPs from text mining results



## 1. Read in data



PubChem API interface showing search results and a data table with columns for compound names and IDs.

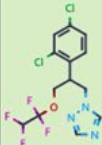
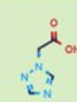
## 2. Launch ShinyTPs



## 3. Check available data

Select an input compound

Tetraconazole

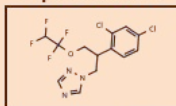
Parent	Parent CID	TP	TP CID
	80277		1810180

## 4. Curate reactions

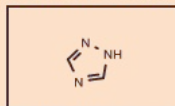
Select an input compound

Tetraconazole

Input structure



Potential TP structure



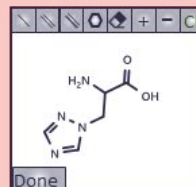
Tetraconazole results in the formation of /1,2,4-triazole (T), triazolyl alanine (TA), triazolyl acetic acid (TAA)/ as well as /triazolyl hydroxypropionic acid/ (THP).

## 5. Add missing entries

Select an input compound

Tetraconazole

Draw a structure

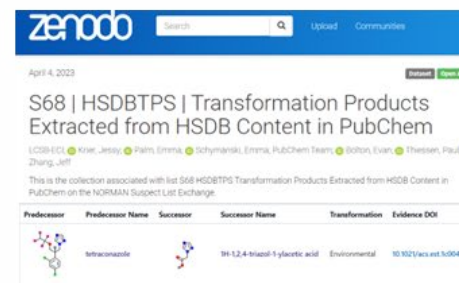


Tetraconazole results in the formation of /1,2,4-triazole (T), triazolyl alanine (TA), triazolyl acetic acid (TAA)/ as well as /triazolyl hydroxypropionic acid/ (THP).



SMILES:

NC(CN1C=NC=N1)C(=O)O

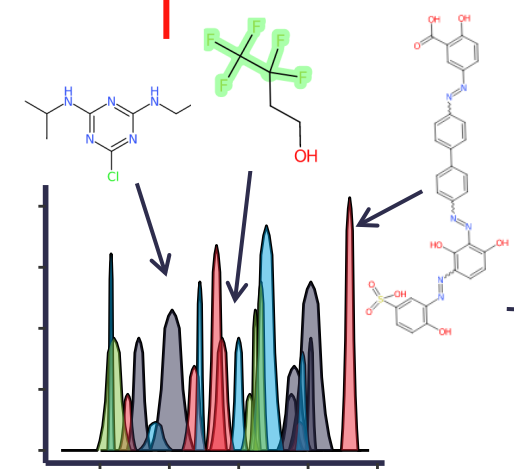
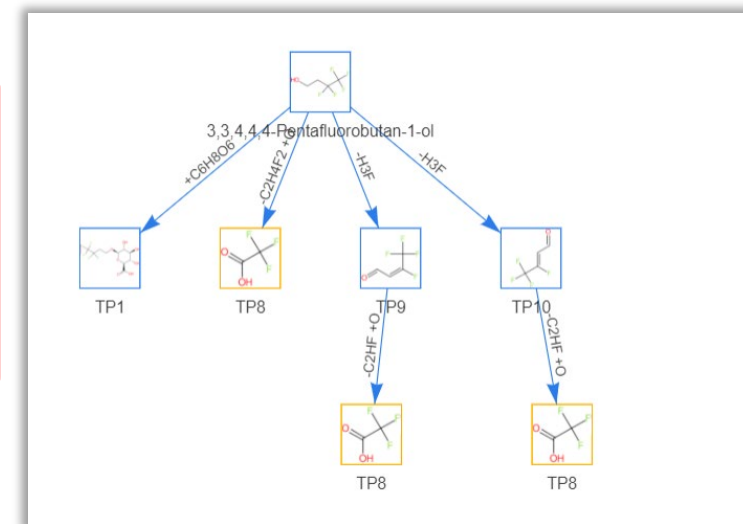
## 6. Export and upload



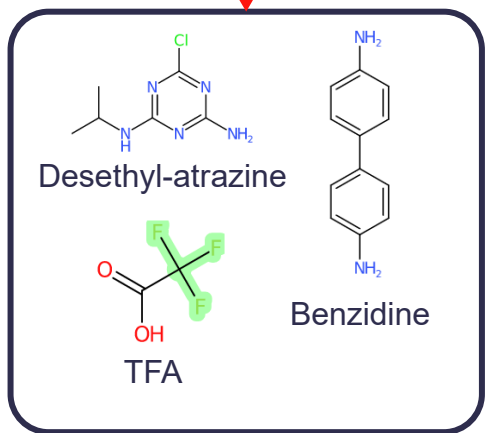
zenodo interface showing a collection titled "S68 | HSDBTPS | Transformation Products Extracted from HSDB Content in PubChem".

Predecessor	Predecessor Name	Successor	Successor Name	Transformation	Evidence DOI
	tetraconazole		1H-1,2,4-triazol-1-ylacetic acid	Environmental	10.1021/acs.est.5b0466

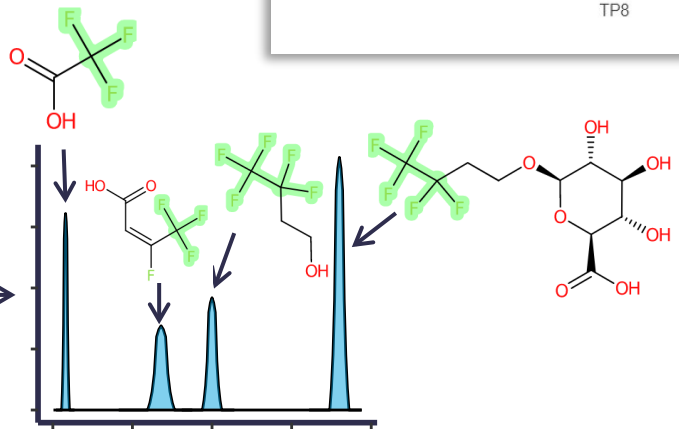
# Incorporating TPs into NT-HRMS Workflows with patRoan



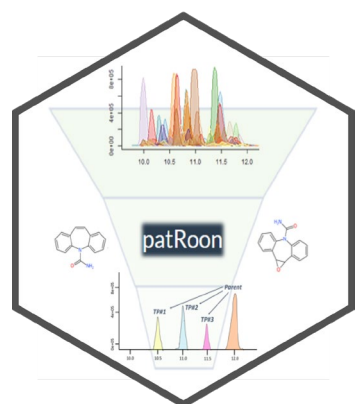
Suspect/Non-target screening



TP library



TP suspect screening



# Future topic: Expanding to Polymers/UVCBs

Functionality under development at PubChem

Data type counts to display  None  Compound Display zero count nodes?  Yes  No

*PubChem PFAS Tree is in the compound space*

Browse PubChem: PFAS and Fluorinated Compounds in PubChem Tree


UVCB = Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials

- 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
    - Breakdown by PFAS composition ? 7,497,376
      - Neutral ? 6,470,991
      - Salt/Mixture ? 1,026,397
    - Breakdown by PFAS functional groups ? 7,497,376
    - Breakdown by PFAS part connectivity degree ? 7,497,376
    - Breakdown by PFAS part formulas ? 7,497,376
  - Regulatory PFAS collections ? 26,943

PubChem About Docs Submit Contact Search PubChem

COMPOUND SUMMARY

## Polytetrafluoroethylene

See also:  Tetrafluoroethylene (has monomer).

PubChem CID Not available because this is not a discrete structure.

Cite Download

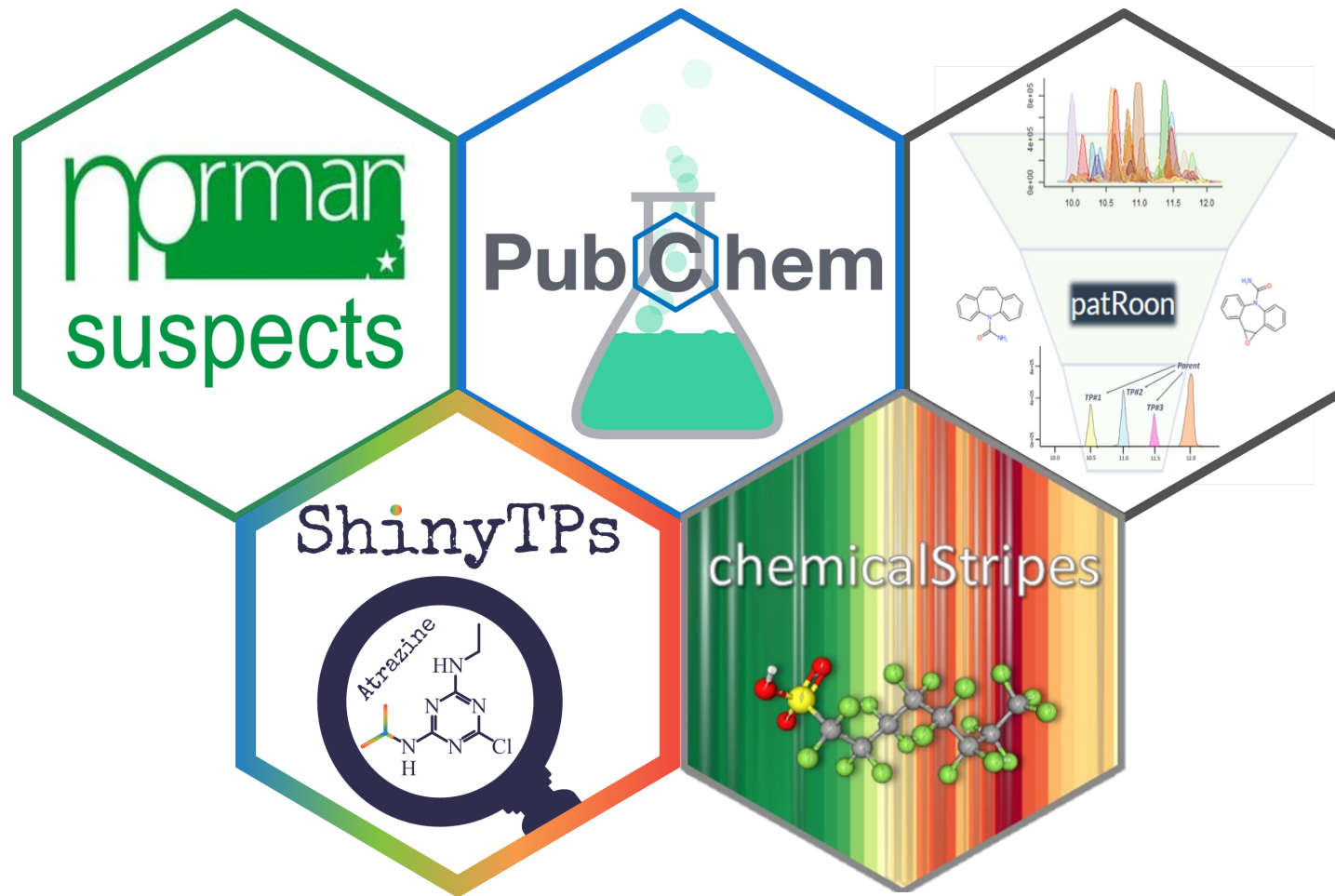
CONTENTS

- Title and Summary
- 1 Synonyms
- 2 Names and Identifiers
- 3 Chemical and Physical Properties
- 4 Related Records
- 5 Drug and Medication Information

<https://pubchem.ncbi.nlm.nih.gov/compound/Polytetrafluoroethylene>

# Take home messages

- Open, FAIR data supports NT-HRMS workflows





# Take home messages

- Open, FAIR data supports NT-HRMS workflows
- There really are **>7 million PFAS** in PubChem!



▼ 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
▶ 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

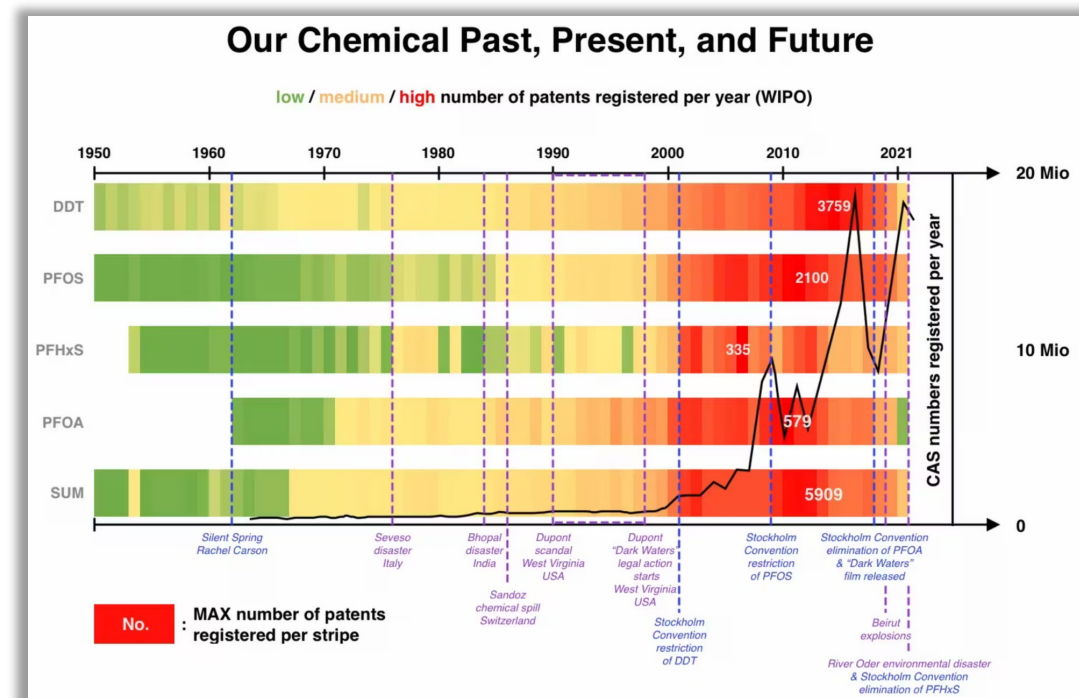
# Take home messages

- Open, FAIR data supports 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>



# Take home messages

- Open, FAIR data supports NT-HRMS workflows

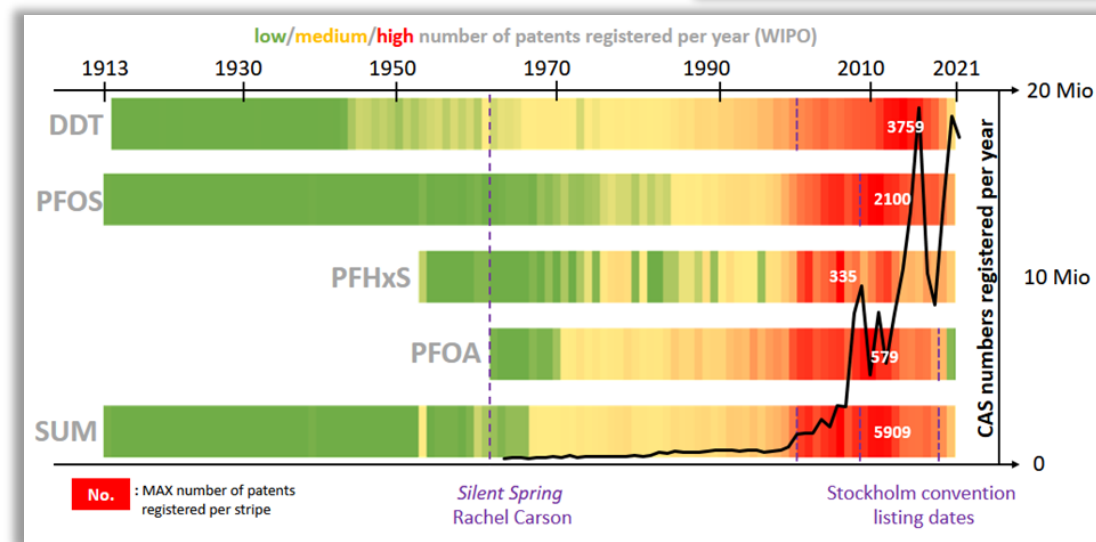
- There really are **>7 million PFAS** in PubChem!

- Annotation content helps find and interpret the **relevant** entries (and this comes from YOU!)

- Help avoid the next Silent Spring & share your data!

Emma: [emma.schymanski@uni.lu](mailto:emma.schymanski@uni.lu) or [@ESchymanski](https://twitter.com/ESchymanski)

PubChem: [pubchem-help@ncbi.nlm.nih.gov](mailto:pubchem-help@ncbi.nlm.nih.gov) or [@pubchem](https://twitter.com/pubchem)





# Acknowledgements!

Today's slides:

DOI: [10.5281/zenodo.8330405](https://doi.org/10.5281/zenodo.8330405)

Email: [emma.schymanski@uni.lu](mailto:emma.schymanski@uni.lu)

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

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

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



# PubChem

UNIVERSITY OF AMSTERDAM



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

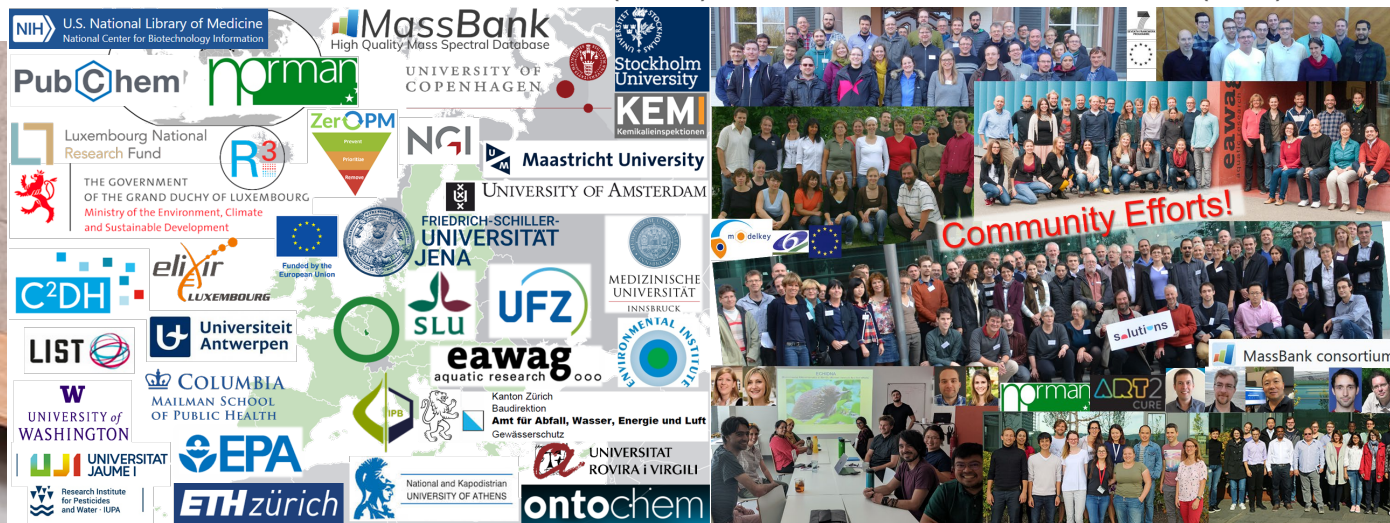


JPML

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



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



Funded by the European Union



H2020:  
101036756



Luxembourg  
National  
Research Fund



68