



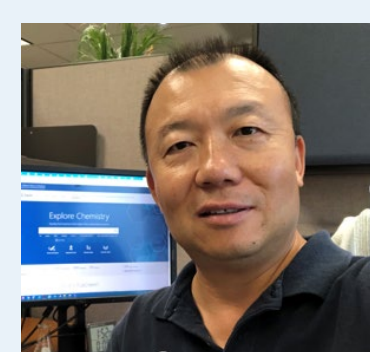
# The PubChem PFAS Tree



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



**Emma L. Schymanski**  
Environmental Cheminformatics,  
Luxembourg Centre for Systems Biomedicine,  
University of Luxembourg  
[emma.schymanski@uni.lu](mailto:emma.schymanski@uni.lu)



**Jian (Jeff) Zhang**  
National Center for Biotechnology  
Information, National Library of  
Medicine, National Institutes of Health  
[pubchem-help@ncbi.nlm.nih.gov](mailto:pubchem-help@ncbi.nlm.nih.gov)



**Paul A. Thiessen**  
National Center for Biotechnology  
Information, National Library of  
Medicine, National Institutes of Health  
[pubchem-help@ncbi.nlm.nih.gov](mailto:pubchem-help@ncbi.nlm.nih.gov)

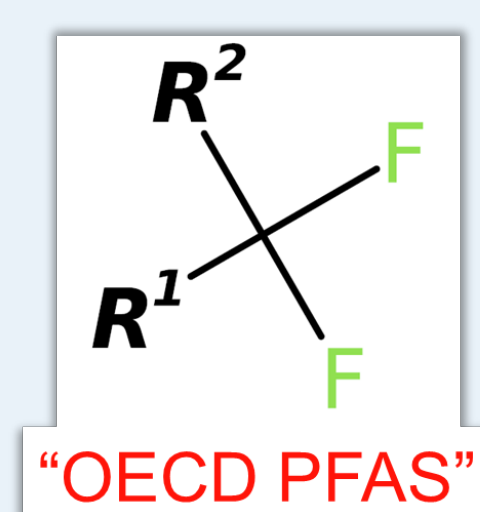


**Evan E. Bolton**  
National Center for Biotechnology  
Information, National Library of  
Medicine, National Institutes of Health  
[pubchem-help@ncbi.nlm.nih.gov](mailto:pubchem-help@ncbi.nlm.nih.gov)

## Motivation

In 2021, the OECD [1] revised the definition for per- and polyfluoroalkyl substances (PFAS) to include molecules containing **at least one saturated CF<sub>2</sub> or CF<sub>3</sub> moiety**.

Scaling this definition PubChem (114 million compounds) reveals a massive **6 million PFAS in PubChem!** Enter the PubChem PFAS Tree [2]



## Search and Download Contents of Nodes

Structures, properties, identifiers, metadata...

cid	mpdnam	cmpsynmw	mf	polararea	complexity	xlogp	heavynt	hbonddor	hbondacc	rotbonds
1	67812	Perfluoromethyl 375-62-2	266.04	CSF100	17.1	288	3.4	16	0	11
2	67814	Nonafluoromethyl 375-72-4	302.09	C4F10025	42.5	386	3.4	17	0	12
3	67815	Perfluoromethyl 375-73-5	300.1	C4HF9035	62.8	387	2.3	17	1	12
4	69494	Pentane, 1,6,9,10-tetrafluoro-	168.1	C5H8F4	11.2	112	1.2	12	0	11
5	69939	1-Butanes 812-94-2	100.1	C4H9	1.2	112	1.2	12	0	11
6	73893	2-Propene 1492-87-1	68.1	C3H6	1.2	112	1.2	12	0	11
7	74534	3,3,4,4,5,5-hexafluoropentane	338.1	C5F10	17.1	288	3.4	16	0	11
8	74883	1-Hexanol 2043-47-2	100.1	C6H14O	1.2	112	1.2	12	0	11
9	141140927	Patents								
10	105447	Classification Names and Identifiers Patents Safety								

PFAS and Fluorinated Compounds in PubChem **20,870,066**

- OECD PFAS definition **6,347,703**
  - Molecule contains isolated CF<sub>2</sub> **635,183**
  - Molecule contains isolated CF<sub>3</sub> **5,592,253**
  - Molecule contains PFAS parts larger than CF<sub>2</sub>/CF<sub>3</sub> **226,161**
- Organofluorine compounds **19,905,615**
  - Fluorinated aliphatic substances **878,562**
  - Fluorinated aromatic substances **18,956,093**
  - Other fluorinated substances **94,911**
- Other diverse fluorinated compounds **121,495**
  - Contains fluorine bond to non-carbon element **27,722**
  - Contains non-organic element **120,411**
- PFAS and fluorinated compound collections **1,789,330**
- PFAS breakdowns by chemistry **7,275,810**
  - Breakdown by PFAS composition **7,275,810**
  - Breakdown by PFAS functional groups **7,275,810**
  - Breakdown by PFAS part connectivity degree **7,275,810**
  - Breakdown by PFAS part formulas **7,275,810**
- Regulatory PFAS collections **25,826**
  - Long-chain PFCAs (LC-PFCAs) and related substances **18,393**
  - PFHxS and related substances **646**
  - PFOA and related substances **24,467**

## PFAS Suspect List Collections

PFAS and fluorinated compound collections **1,789,330**

- CompTox Chemicals Dashboard PFAS suspect lists **16,132**
- NORMAN-SLE PFAS suspect lists **6,300**
- OntoChem PFAS lists **1,777,053**
- Other fluorinated chemical content in PubChem **1,778**
- NIST PFAS suspect list **4,948**

## NEW: Explore Breakdowns by Chemistry

Breakdown by PFAS composition **7,275,810**

- Neutral **6,278,505**
- Salt/Mixture **997,317**

Breakdown by PFAS functional groups **7,275,810**

- Contains PFAS-C **6,517,325**
- Contains PFAS-N **38,992**
- Contains PFAS-O **645,033**
- Contains PFAS-P **15,603**
- Contains PFAS-S **277,352**
- More PFAS-Element cases **5,618**

Breakdown by PFAS part connectivity degree

- PFAS parts with 1 connection **6,570,638**
- PFAS parts with 2+ connections **819,553**

## More Information:

[1] OECD 2021 Report [ENV-CBC-MONO\(2021\)25](https://www.oecd.org/env/cbc/2021-report/)  
[2] PubChem PFAS Tree URL:  
<https://pubchem.ncbi.nlm.nih.gov/classification/#hid=120>

Find out more in our PubChem PFAS Video:



## Regulatory Collections

See our POPRC.18 slides for details:



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101036756. This work was supported in part by the National Center for Biotechnology Information of the National Library of Medicine (NLM), National Institutes of Health. ELS is funded by the Luxembourg National Research Fund (FNR), A18/BM/12341006

## Acknowledgements

We would like to acknowledge discussions with Zhanyun Wang (EMPA, CH), Hans Peter Arp (NGI, NO), Ian Cousins, Luc Miaz, Jon Martin (ACES, SE) and Andreas Buser (FOEN, CH), as well as discussions and contributions from various ECI (specifically Parviel Chirsir and Todor Kondic) and PubChem team members, along with other ZeroPM project members.

