Supporting Information for Revealing Chemical Trends: Insights from Data-Driven Visualisation and Patent Analysis in Exposomics Research

Dagny Aurich*[1], Emma L. Schymanski*[1], Flavio de Jesus Matias[1,2], Paul A. Thiessen[3], Jun Pang[2]

[1] *Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 Avenue du Swing, Belvaux L-4367, Luxembourg.*

[2] *Faculty of Science, Technology and Medicine (FSTM), University of Luxembourg, 6 Avenue de la Fonte, L-4364 Esch-sur-Alzette, Luxembourg.*

[3]*National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.*

ORCIDs: DA: 0000-0001-8823-0596; ELS: 0000-0001-6868-8145; FdJM: 0009-0008-8585-7690; *PAT: 0000-0002-1992-2086; JP: 0000-0002-4521-4112.*

*Corresponding & co-first authors: dagny.aurich@uni.lu, emma.schymanski@uni.lu

Contents

Supplementary Text

Brief supplementary text is given on Page 2.

Supplementary Figures

Supplementary Text

The code to create the figures in the main text and this supplementary information is available on GitLab, with a README file describing the structure of the repository with cross reference to several of the figures:

https://gitlab.com/uniluxembourg/lcsb/eci/ULPatentTrends/-/tree/main#repository-contents

Supplementary Figures

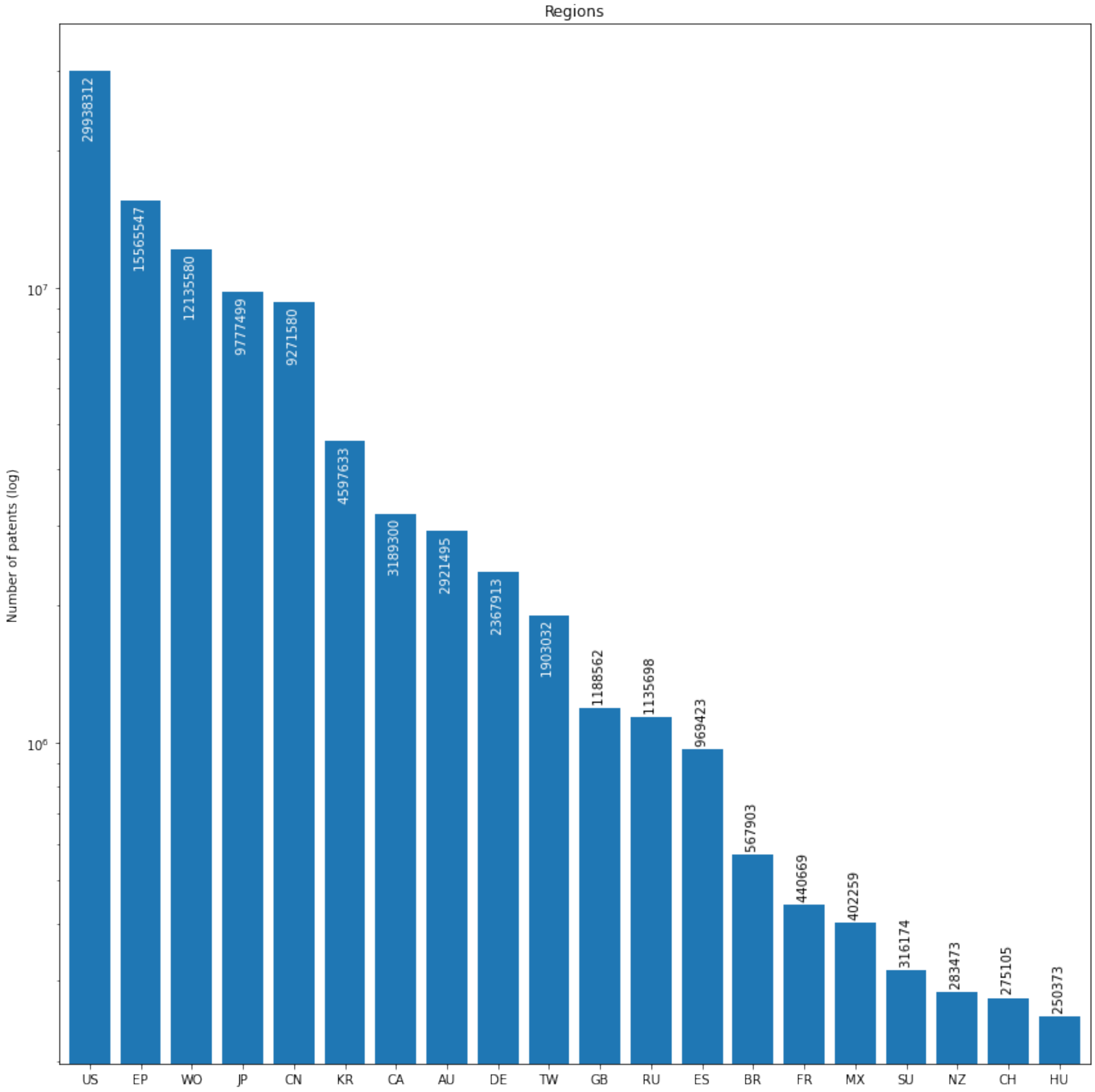


Figure S1: Patent counts by region of origin for 100M rows of the patent dataset. The top 20 of 103 regions are shown. Image extracted from visualize-CID-Patent.ipynb. Note that this distribution changes for patent families and for other row counts, this equivalent plot by family is included in visualize-ID-Date-Family.ipynb; other plots in result-100.000.000-rows.ipynb and result-250.000.000-rows.ipynb.

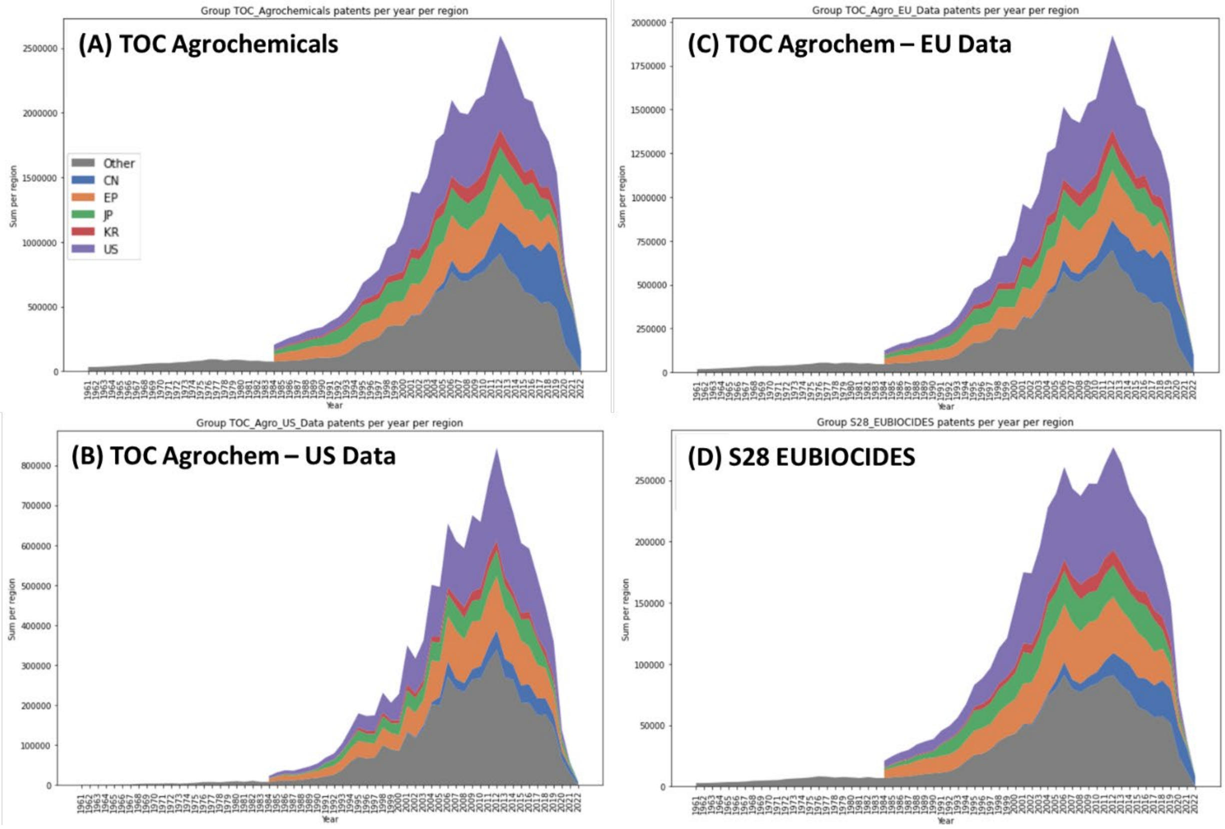


Figure S2: Patent counts for topic-based subsets of chemicals, with regional information - Agrochemical subset. (A) The PubChem Table of Contents (TOC) Agrochemicals category. (B) The TOC Agrochemicals category subset "USDA Pesticides Program". (C) The TOC Agrochemicals category subset "EU Pesticides Data". (D) The NORMAN-SLE S28 EUBIOCIDES list.

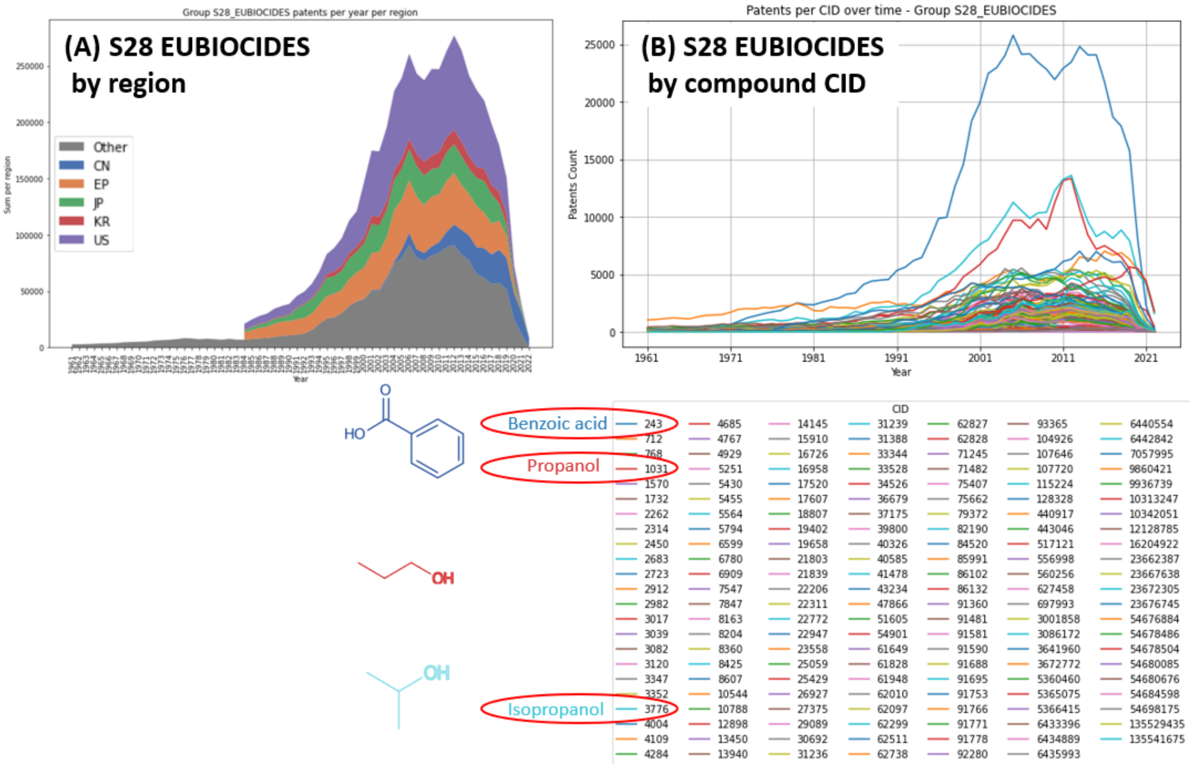


Figure S3: Patent counts for the S28 EUBIOCIDES list per compound (represented by the PubChem Compound Identifier CID). Top: Left: Image from Figure S1(D). Top Right: breakdown by CIDs, listed bottom right. Bottom left: structures for top 3 CIDs by patent count: benzoic acid (CID 243), propanol (CID 1031) and isopropanol (CID 3776).

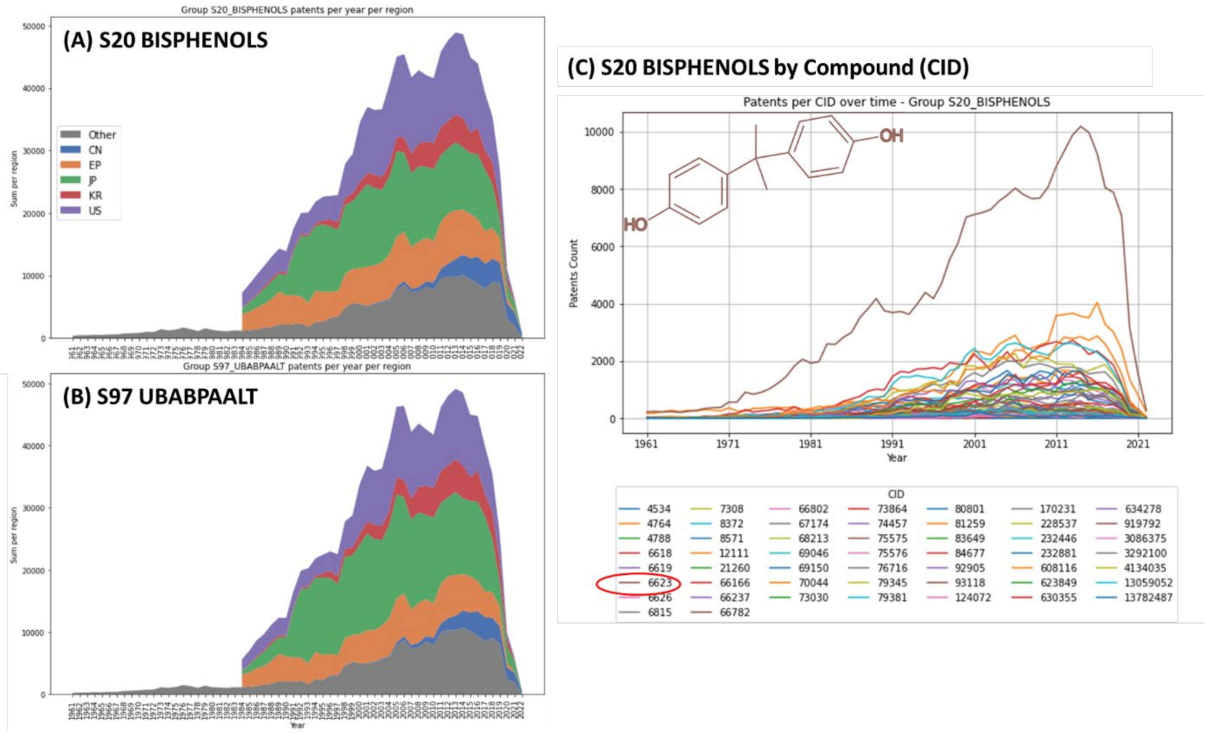


Figure S4: Patent counts for the bisphenol lists. (A) S20 BISPHENOLS by region. (B) S97 UBAPBAALT (BPA Alternatives list from UBA) by region. (C) The S20 BISPHENOLS list per compound (represented by the PubChem Compound Identifier CID), dominated by bisphenol A (BPA), CID 6623 (circled). Inset: structure of BPA.

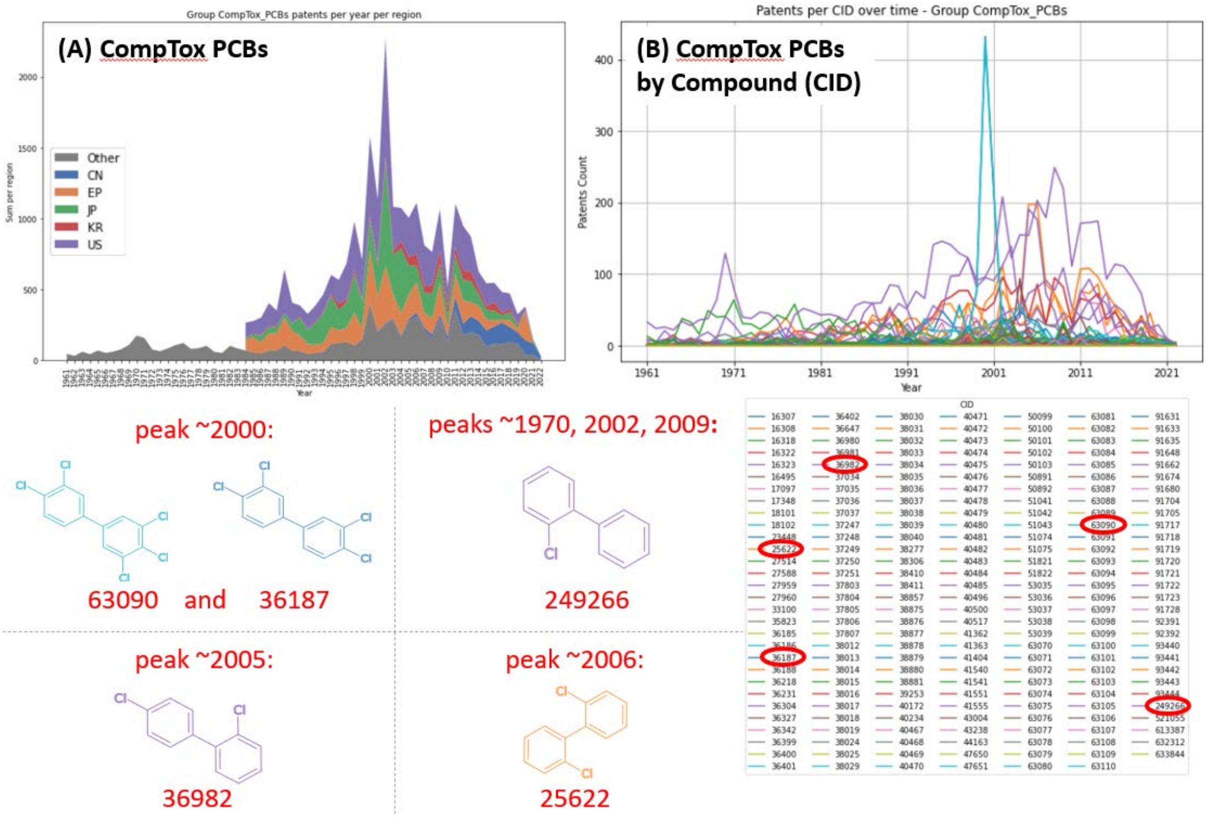


Figure S5: Patent counts for the PCBs. (A) CompTox PCBs by region. (B) CompTox PCBs by Compound (CID). Inset: structures of stand-out PCBs.

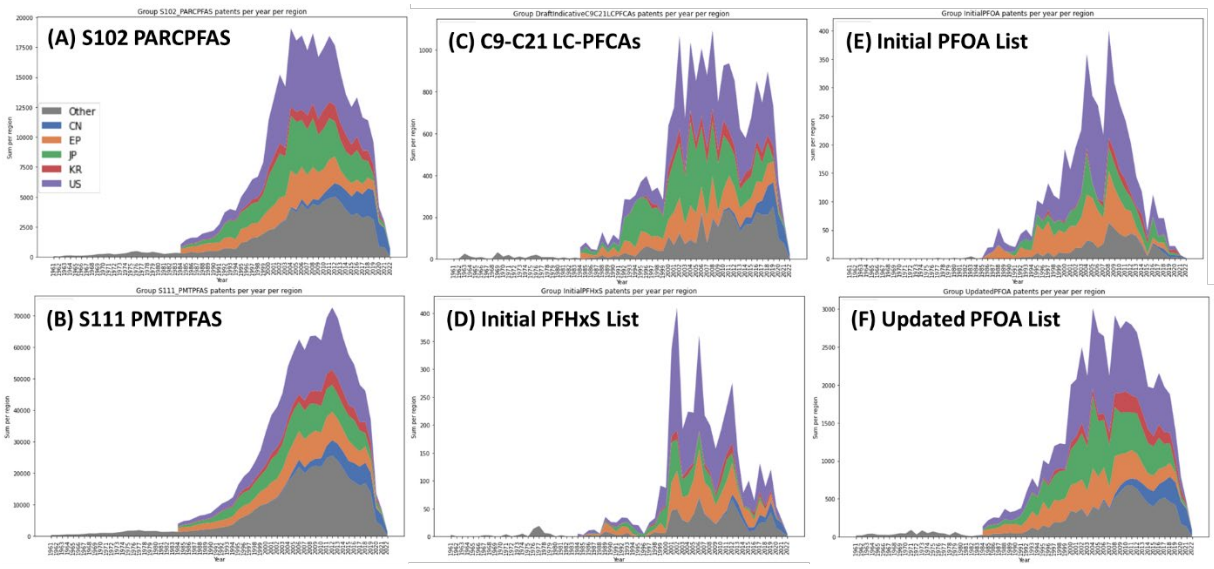


Figure S6: Patent counts for topic-based subsets of chemicals, with regional information - PFAS subset. (A) The NORMAN-SLE S102 PARCPFAS list. (B) The NORMAN-SLE S111 PMTPFAS list. (C) The C9-C21 LC-PFCAs (D) initial PFHxS (E) Initial PFOA and (F) Updated PFOA listing from the Stockholm Convention regulatory collection on the PubChem PFAS Tree. For further details see main text.

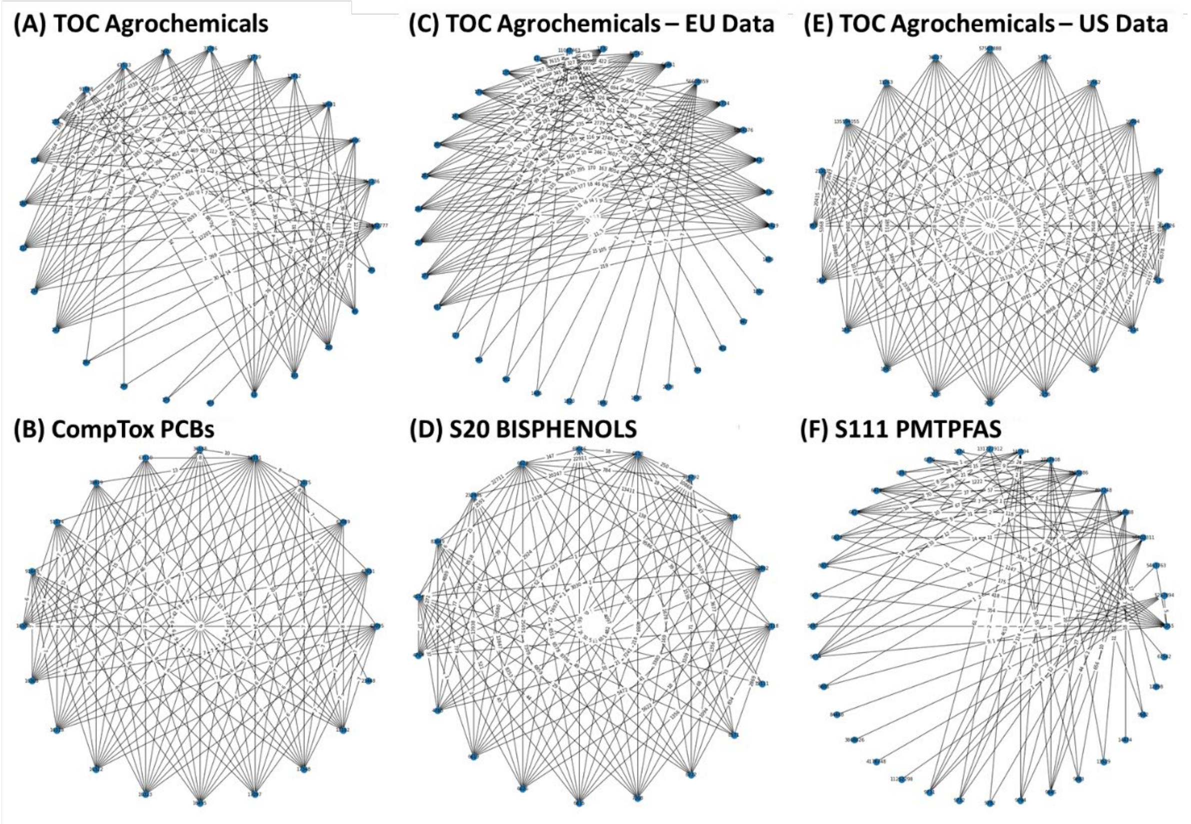


Figure S7: Connected network examples for several lists using 10 CIDs with 10 edges per CID, taken in top order. (A) The TOC Agrochemicals list; (B) CompTox PCBs, (C) TOC Agrochemicals – EU data; (D) S20 BISPHENOLS; (E) TOC Agrochemicals – US data and (F) S111 PMTPFAS. Further examples of connected networks are given in the GitHub repository.

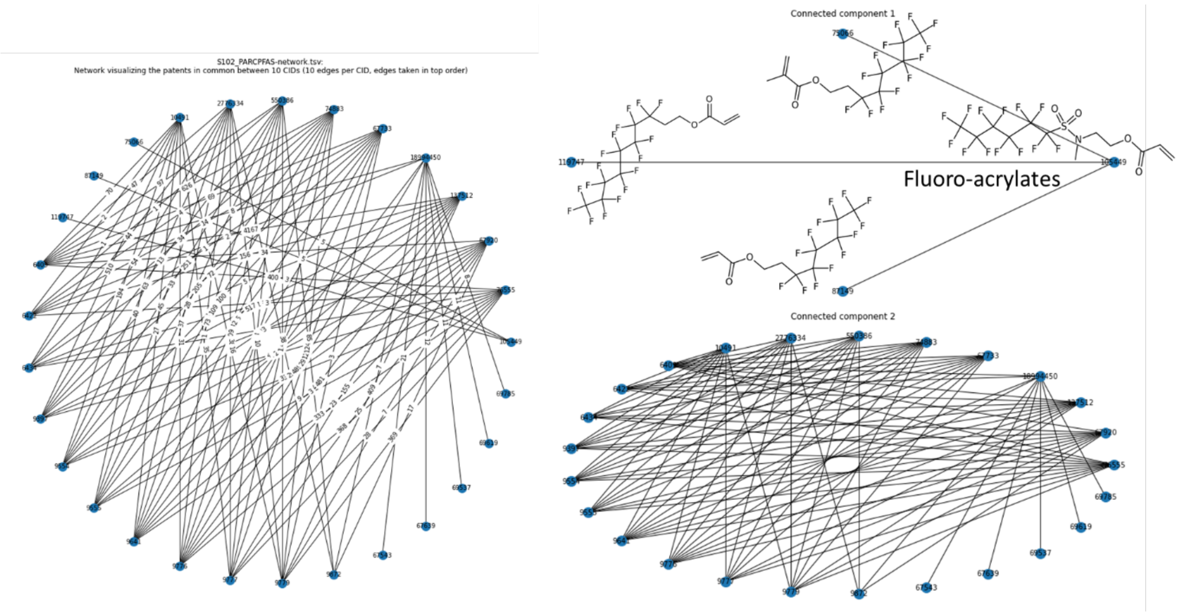
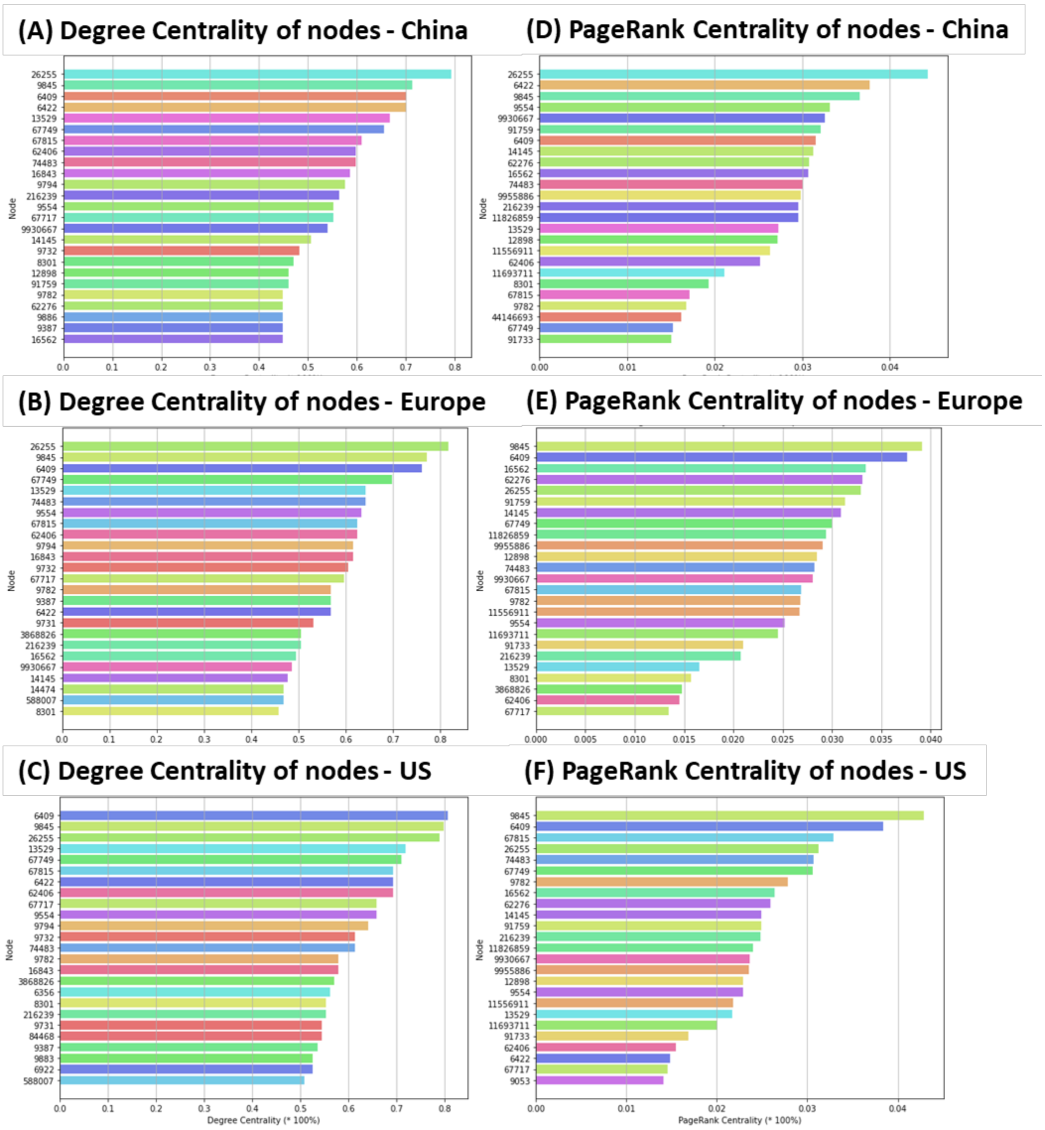

S102_PARCPFAS-network.tsv:
Network visualizing the patents in common between 10 CIDs (10 edges per CID, edges taken in top order)
Connected component 1
Fluoro-acrylates
Connected component 2

Figure S8: Example of a disconnected network for S102 PARCPFAS using 10 CIDs with 10 edges per CID, taken in top order.


(A) Degree Centrality of nodes - China

(D) PageRank Centrality of nodes - China

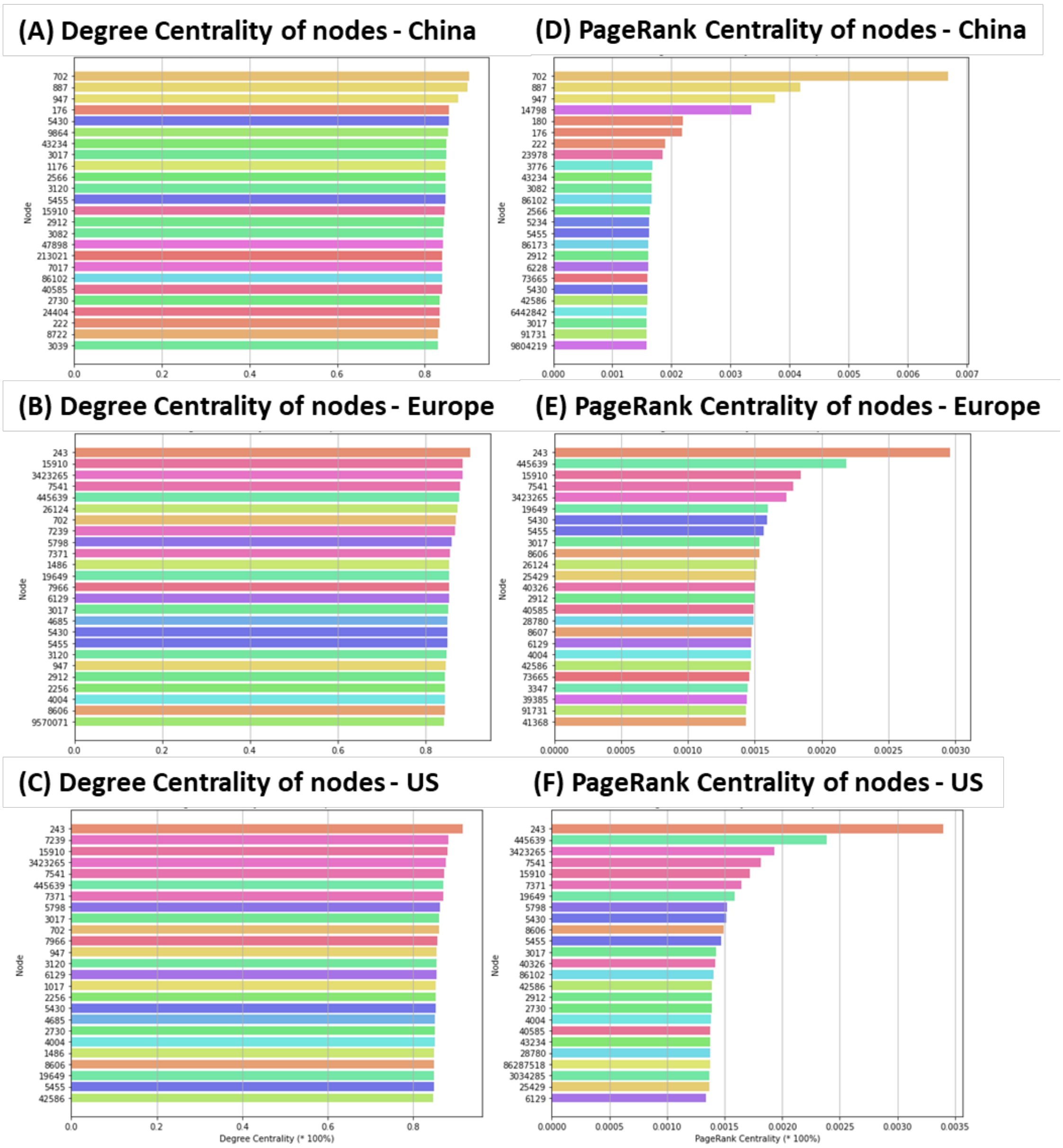
(B) Degree Centrality of nodes - Europe

(E) PageRank Centrality of nodes - Europe

(C) Degree Centrality of nodes - US

(F) PageRank Centrality of nodes - US

Figure S9: Network analysis on S111 PMTPFAS. (A-C) Degree centrality of nodes for China, Europe and the US, respectively. (D-F) PageRank centrality of nodes for China, Europe and the US, respectively. Further details in the main text.



(A) Degree Centrality of nodes - China

(D) PageRank Centrality of nodes - China

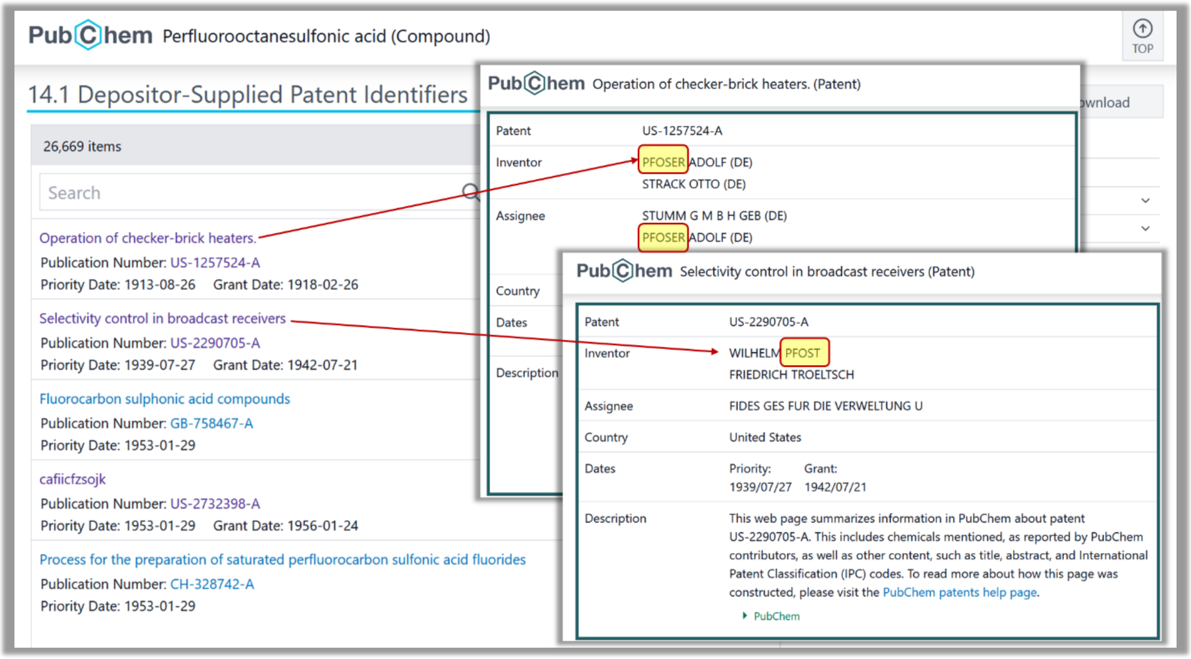
(B) Degree Centrality of nodes - Europe

(E) PageRank Centrality of nodes - Europe

(C) Degree Centrality of nodes - US

(F) PageRank Centrality of nodes - US

Figure S10: Network analysis on TOC Agrochemicals. (A-C) Degree centrality of nodes for China, Europe and the US, respectively. (D-F) PageRank centrality of nodes for China, Europe and the US, respectively. See text for details.



PubChem Perfluorooctanesulfonic acid (Compound)

14.1 Depositor-Supplied Patent Identifiers

26,669 items

Search

Operation of checker-brick heaters.
Publication Number: US-1257524-A
Priority Date: 1913-08-26 Grant Date: 1918-02-26

Selectivity control in broadcast receivers
Publication Number: US-2290705-A
Priority Date: 1939-07-27 Grant Date: 1942-07-21

Fluorocarbon sulphonic acid compounds
Publication Number: GB-758467-A
Priority Date: 1953-01-29

cafiicfzsojk
Publication Number: US-2732398-A
Priority Date: 1953-01-29 Grant Date: 1956-01-24

Process for the preparation of saturated perfluorocarbon sulfonic acid fluorides
Publication Number: CH-328742-A
Priority Date: 1953-01-29

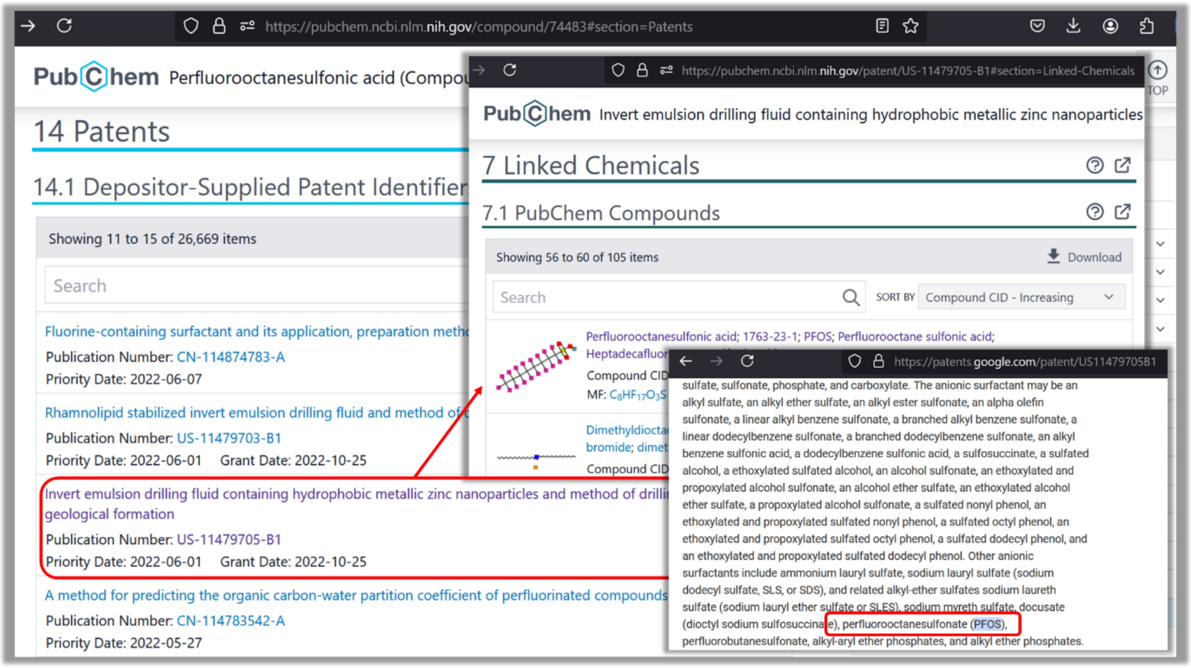
PubChem Operation of checker-brick heaters. (Patent)

Patent	US-1257524-A
Inventor	PFOSER ADOLF (DE) STRACK OTTO (DE)
Assignee	STUMM G M B H GEB (DE) PFOSER ADOLF (DE)
Country	
Dates	
Description	

PubChem Selectivity control in broadcast receivers (Patent)

Patent	US-2290705-A
Inventor	WILHELM PFOST FRIEDRICH TROELTSCH
Assignee	FIDES GES FUR DIE VERWELTUNG U
Country	United States
Dates	Priority: 1939/07/27 Grant: 1942/07/21
Description	This web page summarizes information in PubChem about patent US-2290705-A. This includes chemicals mentioned, as reported by PubChem contributors, as well as other content, such as title, abstract, and International Patent Classification (IPC) codes. To read more about how this page was constructed, please visit the PubChem patents help page. ▸ PubChem

Figure S11: Screenshots of the patent information for PFOS obtained through the PubChem interface on 29 June 2024 using the URL https://pubchem.ncbi.nlm.nih.gov/compound/74483#section=Depositor-Supplied-Patent-Identifiers (sort by "priority date oldest") and the corresponding patent information for patents US-1257524-A and US-2290705-A.



https://pubchem.ncbi.nlm.nih.gov/compound/74483#section=Patents

PubChem Perfluorooctanesulfonic acid (Compou

14 Patents

14.1 Depositor-Supplied Patent Identifier

Showing 11 to 15 of 26,669 items

Search

Fluorine-containing surfactant and its application, preparation meth
Publication Number: CN-114874783-A
Priority Date: 2022-06-07

Rhamnolipid stabilized invert emulsion drilling fluid and method of
Publication Number: US-11479703-B1
Priority Date: 2022-06-01 Grant Date: 2022-10-25

Invert emulsion drilling fluid containing hydrophobic metallic zinc nanoparticles and method of drilli geological formation
Publication Number: US-11479705-B1
Priority Date: 2022-06-01 Grant Date: 2022-10-25

A method for predicting the organic carbon-water partition coefficient of perfluorinated compounds
Publication Number: CN-114783542-A
Priority Date: 2022-05-27

https://pubchem.ncbi.nlm.nih.gov/patent/US-11479705-B1#section=Linked-Chemicals

PubChem Invert emulsion drilling fluid containing hydrophobic metallic zinc nanoparticles

7 Linked Chemicals

7.1 PubChem Compounds

Showing 56 to 60 of 105 items

Search

SORT BY Compound CID - Increasing

Perfluorooctanesulfonic acid; 1763-23-1; PFOS; Perfluorooctane sulfonic acid; Heptadecafluor
Compound CID
MF: $C_8HF_{17}O_3S$

Dimethyldiocta bromide; dimet
Compound CID

https://patents.google.com/patent/US11479705B1

sulfate, sulfonate, phosphate, and carboxylate. The anionic surfactant may be an alkyl sulfate, an alkyl ether sulfate, an alkyl ester sulfonate, an alpha olefin sulfonate, a linear alkyl benzene sulfonate, a branched alkyl benzene sulfonate, a linear dodecylbenzene sulfonate, a branched dodecylbenzene sulfonate, an alkyl benzene sulfonic acid, a dodecylbenzene sulfonic acid, a sulfosuccinate, a sulfated alcohol, a ethoxylated sulfated alcohol, an alcohol sulfonate, an ethoxylated and propoxylated alcohol sulfonate, an alcohol ether sulfate, an ethoxylated alcohol ether sulfate, a propoxylated alcohol sulfonate, a sulfated nonyl phenol, an ethoxylated and propoxylated sulfated nonyl phenol, a sulfated octyl phenol, an ethoxylated and propoxylated sulfated octyl phenol, a sulfated dodecyl phenol, and an ethoxylated and propoxylated sulfated dodecyl phenol. Other anionic surfactants include ammonium lauryl sulfate, sodium lauryl sulfate (sodium dodecyl sulfate, SLS, or SDS), and related alkyl-ether sulfates sodium laureth sulfate (sodium lauryl ether sulfate or SLES), sodium myreth sulfate, docusate (dioctyl sodium sulfosuccinate), perfluorooctanesulfonate (PFOS), perfluorobutanesulfonate, alkyl-aryl ether phosphates, and alkyl ether phosphates.

Figure S12: Verifying linked chemical information in patent documents. Background: patent table for PFOS compound entry in PubChem (https://pubchem.ncbi.nlm.nih.gov/compound/74483#section=Patents). First inset: corresponding patent page for patent US-11479705-B1 (https://pubchem.ncbi.nlm.nih.gov/patent/US-11479705-B1#section=Linked-Chemicals). Second inset: the corresponding Google Patent page where PFOS is mentioned in the patent text (https://patents.google.com/patent/US11479705B1).