

Accessing Environmental Chemistry Data via Data Dashboards

Antony J. Williams

Center for Computational Toxicology and Exposure, US-EPA, RTP, NC

BfR – German Federal Institute for Risk Assessment – April 7th 2022

The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA

Many other talks on SlideShare https://www.slideshare.net/AntonyWilliams/



Slides will be circulated after the presentation





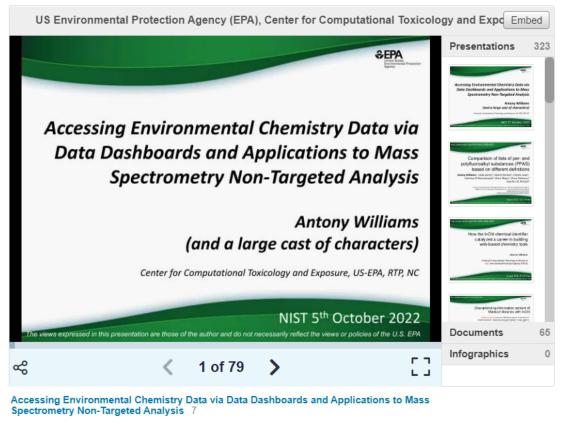
388 SlideShares 290 Followers 1 Clipboard

Follow

Raleigh-Durham, North Carolina Area, North Carolina, United States

www.chemconnector.com

Over the past decade I held many responsibilities including the direction of the development of scientific software applications for spectroscopy and general chemistry, directing marketing efforts, sales and business development collaborations for the company. Eight years experience of analytical laboratory leadership and management. Experienced in experimental techniques, implementation of new NMR technologies walk-up facility management



This talk is an overview



- This talk is a high-level overview only. We can provide trainings into the individual modules and data as required
- LOTS of training materials are available <u>https://www.epa.gov/chemical-research/new-approach-methods-nams-training</u>



20 Years of Curating Data in Our Team





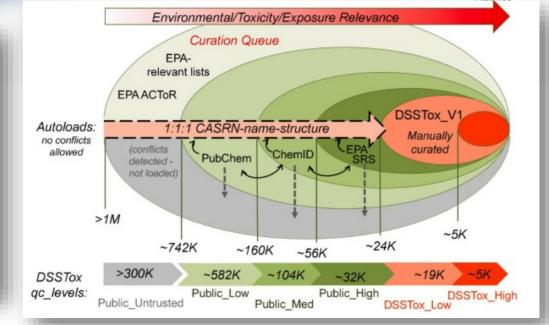
Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis Volume 499, Issue 1, 29 January 2002, Pages 27-52



Mutation Research Frontiers

Distributed structure-searchable toxicity (DSSTox) public database network: a proposal

Ann M. Richard ^a ^A [⊠], ClarLynda R. Williams ^{a, b}





Computational Toxicology Volume 12, November 2019, 100096

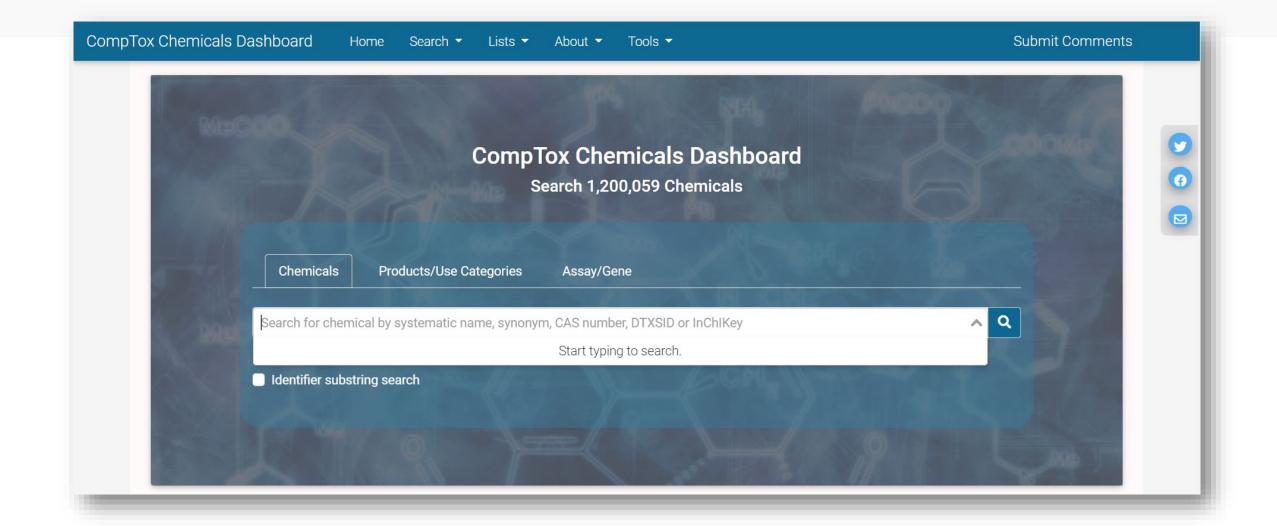


EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research

Christopher M. Grulke ª, Antony J. Williams ª, Inthirany Thillanadarajah ^b, Ann M. Richard ª Ӓ 🖾

CompTox Chemicals Dashboard >1.2 million chemicals



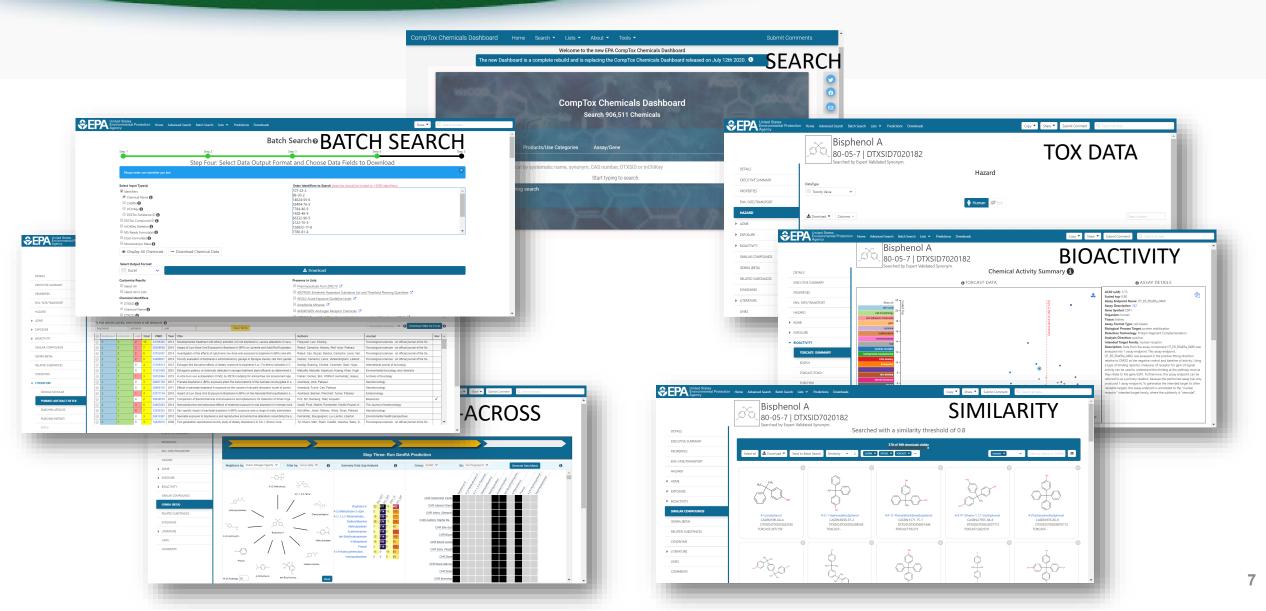




- Develop a "first-stop-shop" for environmental chemical data to support EPA and partner decision making:
 - Centralized location for relevant chemical data
 - Chemistry, exposure, hazard and dosimetry
 - Combination of existing data and predictive models
 - Publicly accessible, periodically updated, curated
- Easy access to data improves efficiency and ultimately accelerates chemical risk assessment

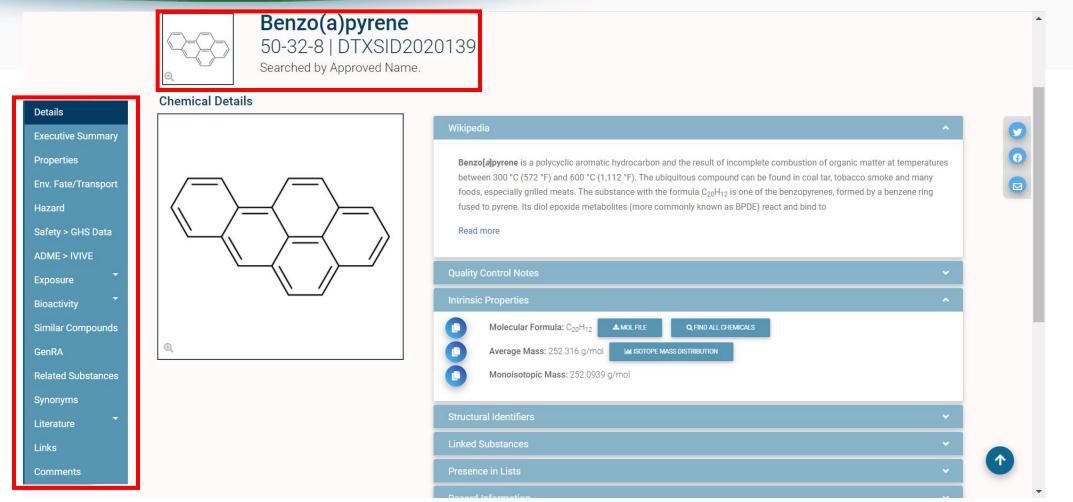
CompTox Chemicals Dashboard https://comptox.epa.gov/dashboard





Detailed Chemical Pages





• Chemical page: Wikipedia snippet when available, intrinsic properties, structural identifiers, linked substances

"Executive Summary"

Executive Summary

Quantitative Risk Assessment Values

🕑 IRIS values available 🗹

🖄 No PPRTV values

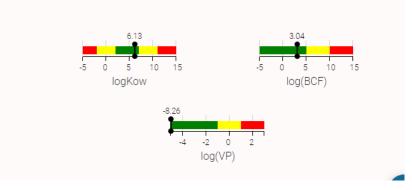
🔮 EPA RSL values available 🗹

🕑 Minimum RfD:0.0003 mg/kg-day (chronic,) 🗹

🕑 Minimum RfC:2e-06 mg/m3 (chronic,) 🗹

- Chronic toxicity PODs available 2*
- Subchronic Toxicology
 No subchronic toxicity data available
- Developmental Toxicology
 No developmental toxicity data available
- Acute Toxicology

No acute toxicity data available





- Overview of toxicityrelated info
 - Quantitative values
 - Physchem. and Fate & Transport
 - Adverse Outcome Pathway links
 - In vitro bioactivity summary plot

Experimental and Predicted Data



Summary	~	Q Search Chemical F
🛓 EXPORT 🝷		
Property	Experimental average	\equiv Predicted average \equiv
Polarizability	-	35.8 (1)
Henry's Law	4.57e-7 (1)	4.59e-7 (1)
Boiling Point	495 (3)	480 (4)
Flash Point	-	234 (2)
Melting Point	177 (8)	189 (3)
Molar Refractivity	-	90.3 (1)
Molar Volume	-	196 (1)
Surface Tension	-	53.9 (2)
Density	-	1.28 (2)
Vapor Pressure	5.49e-9 (1)	3.61e-9 (3)

- Physchem and Fate & Transport experimental and predicted data
- Data can be downloaded as Excel, TSV and CSV files

Chemical Hazard Data



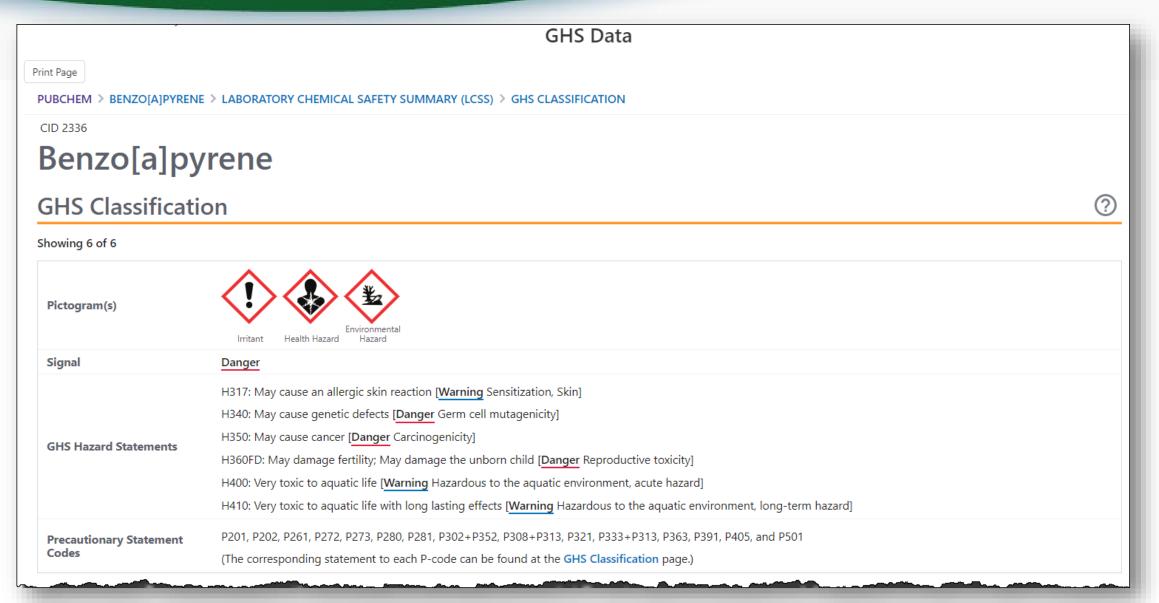
ToxVal Database

- >50k chemicals
- >770k tox. values
- >30 sources of data
- ~5k journals cited
- ~70k citations

	Export -												
	Priority 1	Source	≡ Туре	$\equiv \Big $ Subtype \equiv	Risk Assessment	$\equiv \left \begin{smallmatrix} Qualifi \\ er \end{smallmatrix} \right $	Value 🗄	E Units	Study Type	${}^{\rm Exposur}_{\rm e\ Route}\equiv$	Critical effect 🛛 🚍	Species \equiv	Year =
Ē	1	IRIS	LOAEL	2	chronic	=	9.10e-3	mg/m3	-	inhalation	reduced ovulation rate and ovary weight	2	-
L	1	IRIS	LOAEL	(-)	chronic	=	4.60e-3	mg/m3		inhalation	reduced embryo/fetal survival	=,	~
È	3	ECOTOX	NOEL	25)	chronic growth	57	100	mg/kg f	Growth	Food	Weight	norway	2000
È	3	ECOTOX	NOEL	921	chronic growth	120	25.0	ul/org	Growth	Topical,	Weight	house	1990
È	З	ECOTOX	NOEL	-	chronic growth		50.0	mg/kg f	Growth	Food	Weight	norway	2000
Ľ	3	ECOTOX	NOEL	1051	chronic growth	(73)	25.0	ul/org	Growth	Topical,	Weight	house	1990
ľ	З	ECOTOX	NOEL	17	chronic growth	673	100	mg/kg f	Growth	Food	Weight	norway	2000
È	3	ECOTOX	NOEL	10-1	chronic growth		50.0	mg/kg f	Growth	Food	Weight	norway	2000
È	3	ECOTOX	LOEL		chronic growth	8-8	100	mg/kg f	Growth	Food	Weight	norway	2000
ľ	3	ECOTOX	LOEL	873	chronic growth	1.53	100	mg/kg f	Growth	Food	Weight	norway	2000
È	3	ECOTOX	NOEL	22	chronic growth	22	25.0	ul/org	Growth	Topical,	Weight	house	1990

Safety Data





Sources of Exposure to Chemicals



Chemical Weight Fractions (CWF)

Search Chemical Weig	JIIIII													🛓 EXPOR	⁽¹
duct Name	=	Product Use Category	=	Categorization Subtype	=	Minimum Weight Fraction	=	Maximum Weight Fraction	≡	Data Type	=	Source	≡	Product Count	t =
	∇		∇		7		∇		∇		∇		∇		∇
3743 pah mixture		Not yet Categorized				-		-		reported		SIRI		1	
sphalt cement penetration 60-7	<u>o</u>	Not yet Categorized				-		-		reported		SIRI		1	
ase-neutral 4 1ml methylene ch		Not yet Categorized				-		-		reported		SIRI		1	
ase neutral calibration checkco.		Not yet Categorized				-		-		reported		SIRI		1	
enzo (a) pyrene_ 98%_ b1008-0)	Not yet Categorized				-		-		reported		SIRI		1	
enzo (a) pyrene_md-1956		Not yet Categorized				0.990		1.00		reported		SIRI		1	
asocut 2000 cf art no 875		Not yet Categorized				0.00		1.00e-3		reported		SIRI		1	
lasocut 2000 universal_ 870		Not yet Categorized				-		-		reported		SIRI		1	
lasocut 2000 universal art_ 870		Not yet Categorized				0.00		1.00e-3		reported		SIRI		1	
asocut 4000 strong_ 872		Not yet Categorized				-		-		reported		SIRI		1	
asocut 4000 universal art_ 872		Not yet Categorized				0.00		1.00e-3		reported		SIRI		1	
p-011a clp base/neutrals check		Not yet Categorized				-		-		reported		SIRI		1	



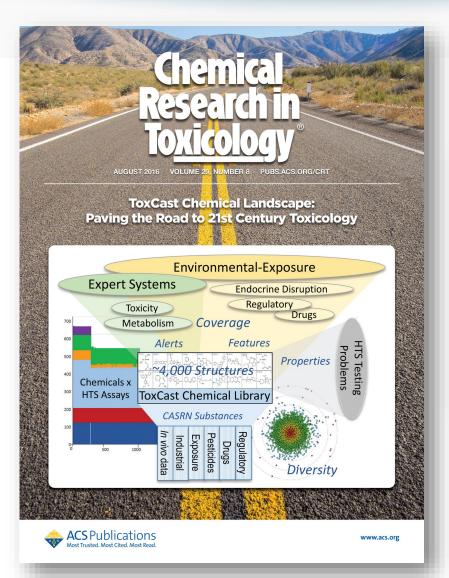
Bioactivity Data





Add to Export

RIS



ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology

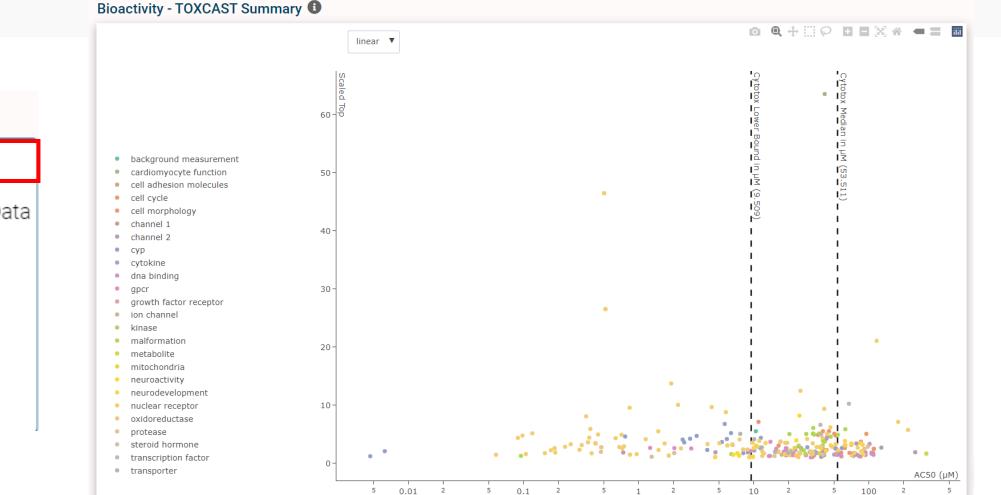
Ann M. Richard^{*†}, Richard S. Judson[†], Keith A. Houck[†], Christopher M. Grulke[†], Patra Volarath[‡], Inthirany Thillainadarajah[§], Chihae Yang^{∥⊥}, James Rathman^{⊥#}, Matthew T. Martin[†], John F. Wambaugh[†], Thomas B. Knudsen[†], Jayaram Kancherla[⊽], Kamel Mansouri[⊽], Grace Patlewicz[†], Antony J. Williams[†], Stephen B. Little[†], Kevin M. Crofton[†], and Russell S. Thomas[†]

View Author Information $^{\smallsetminus}$

Cite this: Chem. Res. Toxicol. 2016, 29, 8, 1225–	Article Views	Altmetric	Citations	Share
1251 Publication Date: July 1, 2016 ~	6687	36	244	
https://doi.org/10.1021/acs.chemrestox.6b00135	LEARN A	BOUT THESE ME	TRICS	

Bioactivity Data Summary views of >2000 Assay Endpoints





2

10

5

Bioactivity

ToxCast: Summary

Toxcast Conc. Response Data

HTTr: Summary

HTPP: Summary

PubChem

ToxCast: Models

Bioactivity Data Full transparency of data...



Bioactivity

ToxCast: Summary

Toxcast Conc. Response Data

HTTr: Summary

HTPP: Summary

PubChem

ToxCast: Models

Concentration Response Data ¹

Analytical Data on Tox21 Browser 🗹

🛓 EXPORT 👻

	Name 1	=	Description	=	Endpoint Name	≡	Active	=	Details	Rep. Plot	All Plots	Gene	=	Intended Target		⊟ Cell Form	iat 🤅
		∇		∇		∇		∇					7		♥	▽	
	ASSAY SOURCE: ACEA		ACEA Biosciences		ACEA_AR_agonist_80hr		Inactive		Ľ	₩.	⊞	AR		steroidal	prostate	cell line	2
	ASSAY SOURCE: ACEA		ACEA Biosciences		ACEA_AR_agonist_AUC_viability		Active		Ē	<u>⊷</u>	=	-		cytotoxicity	prostate	cell line	2
-	ASSAY SOURCE: ACEA		ACEA Biosciences		ACEA_AR_antagonist_80hr		Active		=	2	=	AR		steroidal	prostate	cell line	3
1	ASSAY SOURCE: ACEA		ACEA Biosciences		ACEA_AR_antagonist_AUC_viability		Active		=	<u>⊷</u>	=	-		cytotoxicity	prostate	cell line	3
	ASSAY SOURCE: ACEA		ACEA Biosciences		ACEA_ER_80hr		Active		-	<u>⊷</u>	=	ESR1		steroidal	breast	cell line	2
	ASSAY SOURCE: ACEA		ACEA Biosciences		ACEA_ER_AUC_viability		Inactive		=	2	=	-		cytotoxicity	breast	cell line	3
	ASSAY SOURCE: APR		Apredica		APR_HepG2_CellCycleArrest_1h_dn		Inactive		Ē	<u>12</u>	Ħ	-		proliferation	liver	cell line	2
	ASSAY SOURCE: APR		Apredica		APR_HepG2_CellCycleArrest_1h_up		Inactive		=	2	=	-		arrest	liver	cell line	3
	ASSAY SOURCE: APR		Apredica		APR_HepG2_CellCycleArrest_24h_dr	n	Inactive		=	₩	=	-		proliferation	liver	cell line	3
	ASSAY SOURCE: APR		Apredica		APR_HepG2_CellCycleArrest_24h_up	D	Inactive		Ē	<u>⊷</u>	=	-		arrest	liver	cell line	2
-	ASSAY SOURCE: APR		Apredica		APR_HepG2_CellCycleArrest_72h_dr	٦	Inactive		=	2	=	-		proliferation	liver	cell line	3
1	ASSAY SOURCE: APR		Apredica		APR_HepG2_CellCycleArrest_72h_up	þ	Inactive		-	<u>14</u>	Ħ	-		arrest	liver	cell line	;
	ASSAY SOURCE: APR		Apredica		APR_HepG2_CellLoss_1h_dn		Inactive		=	2	=	-		cytotoxicity	liver	cell line	3
	ASSAY SOURCE: APR		Apredica		APR_HepG2_CellLoss_1h_up		Inactive		=	<u>⊷</u>	=	-		proliferation	liver	cell line	3
	ASSAY SOURCE: APR		Apredica		APR_HepG2_CellLoss_24h_dn		Active		6	<u>~</u>	=	-		cytotoxicity	liver	cell line	2

Bioactivity Data ...including concentration-response



- -111

cutoff

cnst

hill

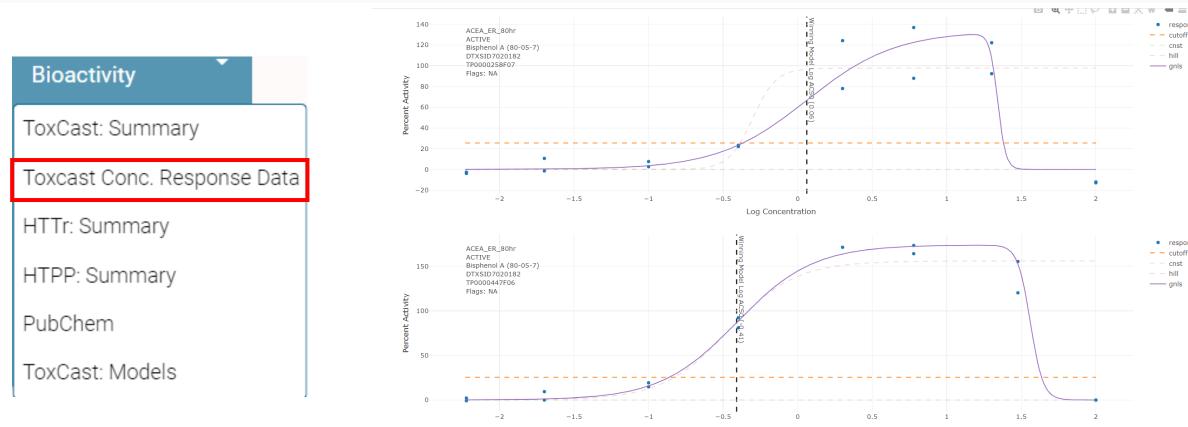
anl

response

cnst

hill

response



How can these data be used? Combine *in vitro* and httk



JOURNAL ARTICLE FEATURED

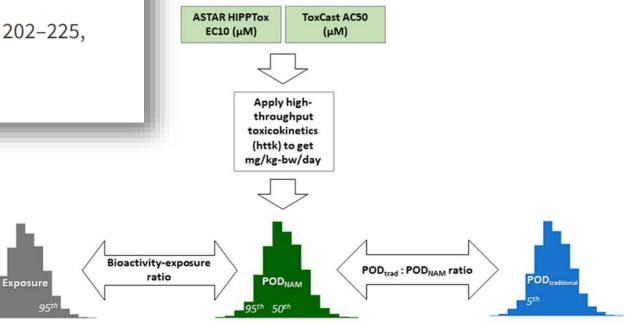
Utility of *In Vitro* Bioactivity as a Lower Bound Estimate of *In Vivo* Adverse Effect Levels and in Risk-Based Prioritization @

Katie Paul Friedman ➡, Matthew Gagne, Lit-Hsin Loo, Panagiotis Karamertzanis, Tatiana Netzeva, Tomasz Sobanski, Jill A Franzosa, Ann M Richard, Ryan R Lougee, Andrea Gissi ... Show more

Toxicological Sciences, Volume 173, Issue 1, January 2020, Pages 202–225, https://doi.org/10.1093/toxsci/kfz201

Published: 18 September 2019

Combine *in vitro* data (AC50s in uM) with httk data to get bioactivityexposure ratios and PODs



Bioactivity Data NEW High-Throughput Transcriptomics



BMD (µM)

100

P

ផ្ទើ

10

Bioactivity

ToxCast: Summary

Toxcast Conc. Response Data

HTTr: Summary

HTPP: Summary

PubChem

ToxCast: Models

Bioactivity - HTTr Summary 🕕

Biomolecular Process

- Blood
- CancerCancerlImmune
- Cardiovascular
- Cell Cycle
- Chemical Property
- CholinergicCongenital Disorder
- CYP
- EnzymeGPCR
- Growth Factor
- Hormone
 Immune
- Immune|Blood
- InfectionIon Channel
- Kinase
 Lipid
- Musculoskeletal
 NA
- NA
 Nervous System
- Nuclear Receptor
 Other
- Other
 Pesticide
- Pregnancy Complication
- Random
 Skin
- StressThyroid
- Transcription Factor

-0.6

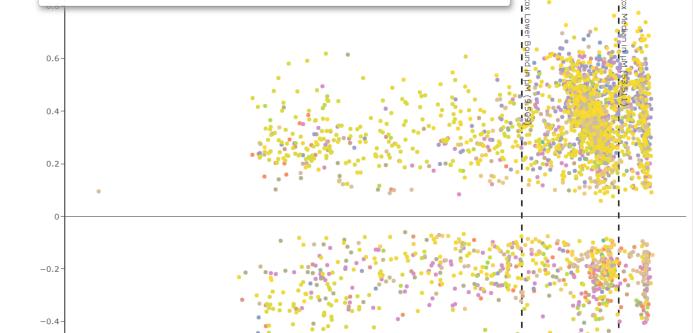
0.01

5

Translation/Transcription
 Transporter

values from active signatures only (nitcall>0.9). The x-axis is concentration in uni, and the y-axis is the top value from the curve fit. Data is shown at the signature level (one dot per signature or gene set). Signatures are organized into "super targets", which can be genes, gene families or higher order biological processes, up to human diseases. The process of curve fitting is described here and the overall signature scoring process is described here. https://doi.org/10.1093/toxsci/kfab00

0.1



Use Models Derived from the Data



Screening Chemicals for Estrogen Receptor Bioactivity Using a **Computational Model**

Patience Browne^{*†}, Richard S. Judson[‡], Warren M. Casey[§], Nicole C. Kleinstreuer^{II}, and Russell S. Thomas[‡]

View Author Information ~

Cite this: Environ. Sci. Technol. 2015, 49, 14, 8804-8814 Publication Date: June 12. 2015 ~ https://doi.org/10.1021/acs.est.5b02641

Article Views Altmetric Citations 27 157 3796 LEARN ABOUT THESE METRICS

Share <

Vol. 124, No. 7 Research

CERAPP: Collaborative Estrogen Receptor Activity Prediction Project

Kamel Mansouri, Ahmed Abdelaziz, Aleksandra Rybacka, Alessandra Roncaglioni, Alexander Tropsha, Alexandre Varnek, Alexey Zakharov, Andrew Worth, Ann M. Richard, Christopher M. Grulke, Daniela Trisciuzzi, Denis Fourches, Dragos Horvath, Emilio Benfenati, Eugene Muratov, Eva Bay Wedebye, Francesca Grisoni, Giuseppe F. Mangiatordi, ... See all authors 🛛 🗸

Published: 1 July 2016 https://doi.org/10.1289/ehp.1510267 Cited by: 76

CERAPP and CoMPARA available in OPERA

Vol. 128, No. 2 | Research

CoMPARA: Collaborative Modeling Project for Androgen Receptor Activity

Kamel Mansouri 🖂, Nicole Kleinstreuer, Ahmed M. Abdelaziz, Domenico Alberga, Vinicius M. Alves, Patrik L. Andersson, Carolina H. Andrade, Fang Bai, Ilya Balabin, Davide Ballabio, Emilio Benfenati, Barun Bhhatarai, Scott Boyer, Jingwen Chen, Viviana Consonni, Sherif Farag, Denis Fourches, Alfonso T. García-Sosa, Paola Gramatica, Francesca Grisoni, ... See all authors

Published: 7 February 2020 CID: 027002 https://doi.org/10.1289/EHP5580 Cited by: 2

Development and Validation of a Computational Model for Androgen Receptor Activity

Nicole C. Kleinstreuer*†, Patricia Ceger[‡], Eric D. Watt[§], Matthew Martin[§], Keith Houck[§], Patience Browne^{II}, Russell S. Thomas[§], Warren M. Casey[†], David J. Dix[⊥], David Allen[‡], Srilatha Sakamuru[#], Menghang Xia[#], Ruili Huang[#], and Richard Judson§

View Author Information ~

Cite this: Chem. Res. Toxicol. 2017, 30, 4, 946–964 Publication Date: November 18, 2016 ~ https://doi.org/10.1021/acs.chemrestox.6b00347 Copyright © 2016 American Chemical Society

Article Views Altmetric Citations 4338 27 LEARN ABOUT THESE METRICS

94





Searching Literature and the Internet

Literature Searching



Literature - PubMed Abstract Sifter

Abstract Sifter Instructions

 Select PubMed starting point query
Hazard 🗸 🗸
Choose Query Term
Hazard
Fate and Transport
Metabolism/PK/PD
Chemical Properties
Exposure
Mixtures
Male Reproduction
Androgen Disruption
Female Reproduction
GeneTox
Cancer
Clinical Trials
Embryo and embryonic development
Child (infant through adolescent)
Dust and Exposure
Food and Exposure
Water and Exposure
Algae
Disaster / Emergency

Optionally, enter any PubMed query or edit the query from step 1 ("50-32-8" OR "Benzo(a)pyrene") AND (NOAEL OR NOEL OR LOEL OR Rfd OR "reference dose" OR "reference concentration" OR "adverse effect level"[tiab] OR "cancer slope factor"[tiab])

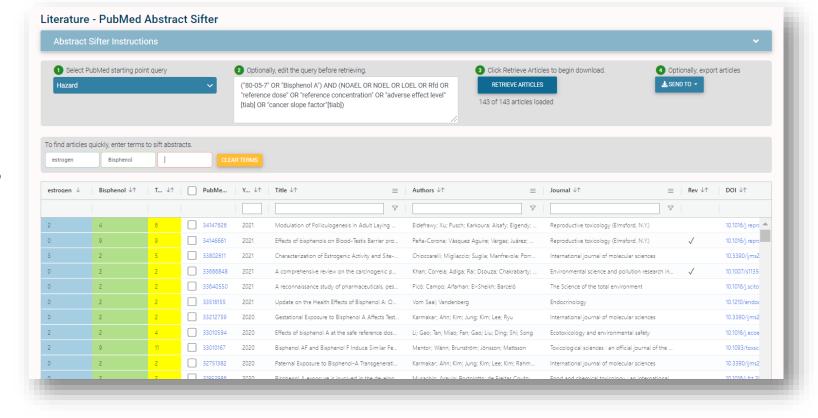


- Real-time retrieval of data from PubMed ~30 million abstracts and growing)
- Choose from set of pre-defined queries
- Adjust and fine tune queries based on interests

Literature Searching



- "Sifting" of results using multiple terms
- Frequency counting terms
- Color highlighting of terms
- Download list to Excel
- Send list to PubMed for downloading ref. file
- Direct link via PubMed ID



What's the best way to search the internet for chemical data?



- We know how complex chemicals identifiers are...
 - CASRN(s)
 - Hundreds of names (maybe)
 - SMILES
 - InChIs
 - EINECS, EC numbers
- What can WE do to help you navigate the internet?

External Links – Also use Identifiers Names, CASRN, PubChem IDs, InChIs.



Benzo(a)pyrene 50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

General

- (a) EPA Substance Registry Service
- PubChem
- Chemspider
- CPCat
- 🥖 DrugBank
- W Wikipedia
- Q MSDS Lookup
- ChEMBL
- toxPlanet
- ACS Reagent Chemicals
- 🌞 Wolfram Alpha
- 🔀 ECHA Infocard
- ChemAgora
- Consumer Product Information Database
- ChEBI
- NIST Chemistry Webbook
- **WEBWISER**
- PubChem Safety Sheet

PubChem: Chemical Vendors

Consumer Product Information Database

Toxicology

ACToR

- оң DrugPortal
- ChemView
- CTD
- Gene-Tox
- ACToR PDF Report
- CREST
- ECOTOX
- ChemView
- Chemical Checker
- BindingDB
- CalEPA OEHHA
- MIOSH IDLH Values
- LactMed
- ECOTOX

Publications

and Toxline

- PPRTVWEB
- PubMed
- IRIS Assessments
- 🖲 EPA HERO
- 🚾 NIOSH Skin Notation Profiles
- 💷 NIOSH Pocket Guide
- RSC Publications
- 🛋 BioCaddie DataMed
- 🖉 Springer Materials
- Bielefeld Academic Search Engine
- CORE Literature Search
- G Google Books (Text Search)
- Google Patents (Text search)
- G Google Scholar (Text search)
- G Google Patents (Structure search)
- Google Books (Structure Search)
- Google Scholar (Structure search)
- Federal Register

Analytical

RSC Analytical Abstracts

- 🗟 Tox21 Analytical Data
- 😬 MONA: MassBank North America
- imzCloud 🧆
- NIST IR Spectrum
- NIST MS Spectrum
- 🐗 MassBank
- NIST Antoine Constants
- IR Spectra on PubChem
- NIST Kovats Index values
- Protein DataBank
- 🍐 National Environmental Methods Index

Prediction

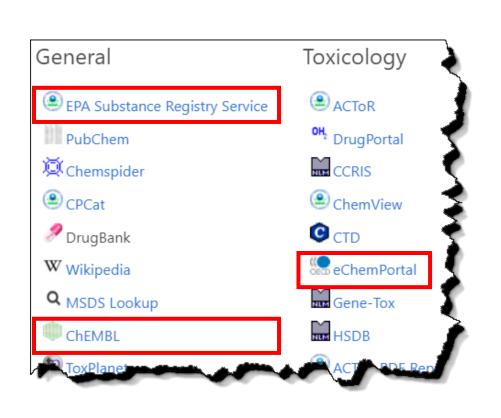
- 2D NMR HSQC/HMBC Prediction
- Carbon-13 NMR Prediction
- Proton NMR Prediction
- ChemRTP Predictor
- LSERD

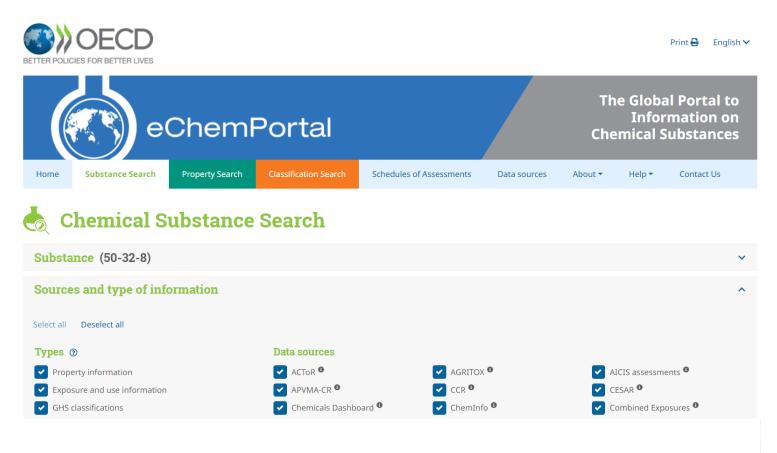
United States Environmental Protection Agency

External Links



Links to ~90 websites providing access to additional data on the chemical of interest







Chemical Lists and Categories

A List of Lists of Chemicals

https://comptox.epa.gov/dashboard/chemical_lists



Q Search Chem	ical List	S			EXPORT - COPY URL
				Showin	ig 36 of 319 Records
ist Acronym 🖓	$\equiv $	List Name	# Chemicals	Updated	List Description
PFAS	∇				
EPAPFASDW537		PFAS EPA WATER: Existing EPA DW Method	19	2019-11-16	EPA has recently revised method 537.1 for the PFAS on this list to detect them in drinking water.
EPAPFASDWTREAT		PFAS EPA WATER: Drinking Water Treatment	9	2019-11-16	EPA is gathering and evaluating treatment effectiveness and cost data for removing these PFAS from drinking water systems.
EPAPFASINSOL		PFAS EPA: Chemical Inventory Insoluble in	43	2021-11-21	Per- and Polyfluoroalkyl Substances (PFASs) in EPA's expanded ToxCast chemical inventory that were determined to be insoluble in DMSO above 5mM concentration. These PFAS chemicals were successfully procured from commercial suppliers (with a small number provided by National Toxicology Program partners) but deemed unsuitable for testing due to limited DMSO solubility. For a complete list of solubilized PFAS in EPA's inventory, see https://comptox.epa.gov/dashboard/chemical-lists/EPAPFASINV
EPAPFASINV		PFAS EPA: ToxCast Chemical Inventory	430	2021-11-21	Per- and Polyfluoroalkyl Substances (PFAS) included in EPA's expanded ToxCast chemical inventory and available for testing. These PFAS chemicals were successfully procured from commercial suppliers (with a small number provided by National Toxicology Program partners) and were deemed suitable for testing (i.e., solubilized in DMSO above 5mM, and not gaseous or highly reactive). All or portions of this inventory are being made available to EPA researchers and collaborators to be analyzed and tested in various high-throughput screening (HTS) and high-throughput toxicity (HTT) assays.
					The https://comptox.epa.gov/dashboard/chemical-lists/EPAPFAS75S1 list is a prioritized subset of this larger chemical inventory.

The OECD List of PFAS

http://www.oecd.org/chemicalsafety/portal-perfluorinated-chemicals/

PFAS: Listed in OECD Global Database

Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey

Identifier substring search

List Details

Description: OECD released a New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances (PFASs) listing more than 4700 new PFAS, including several new groups of PFASs that fulfill the common definition of PFASs (i.e. they contain at least one perfluoroalkyl moiety) but have not yet been commonly regarded as PFASs. The list can be used in conjunction with the methodology report summarising the major findings with respect to the total numbers and types of PFASs identified, the limitations, gaps and challenges identified, and opportunities for improving the future understanding of PFASs production, use on the global market, and presence in the environment, biota, and other matrices.

Source website: http://www.oecd.org/chemicalsafety/portal-perfluorinated-chemicals

A major effort was undertaken to register this list within DSSTox, adding chemical structures for as many PFAS entries as possible using both manual and auto-mapping (structures using CAS-matching) curation methods. The result is that approximately 1/3 of the list is curated at the highest two curation levels (DSSTox_High or DSSTox_Low) currently, whereas more than half of this list is registered at the Public_Low curation level (based on PubChem content). The PFASOECD list is undergoing continuous registration and curation.

Number of Chemicals: 4729

Search Results	~ <mark>~</mark>	SEND 4729 TO BATCH SEARCH	TILE INFO + FILTER +			Ł EXPORT +	
		s	showing 4729 of 4729 chemicals				
HHHHH		O related chemical structures with this substance	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~		HHHHH		
Q 2-(N-Methylperfluoroocta	Q 1-lodo-4-(tridecafluorohex	Copolymer of 2,3,3,3-tetra DTXSID : DTXSID60882687	© 1,1,1,2,2,3,3,4,4,5,5,6,6-Trid	Q Potassium perfluorooctan			



0

V Q

PFAS List Paper https://doi.org/10.3389/fenvs.2022.850019





Polyfluoroalkyl Substances (PFAS) to Support **Environmental Science Research**

🌉 Antony J. Williams1*, 🚊 Linda G. T. Gaines², 🚊 Christopher M. Grulke1t, 🚊 Charles N. Lowe1, 🚊 Gabriel F. B. Sinclair³, 🔄 Vicente Samano⁴, 🔄 Inthirany Thillainadarajah⁴, 🔄 Bryan Meyer⁴, 🌉 Grace Patlewicz¹ and 🎆 Ann M. Richard¹

- What makes our efforts
- MANUAL curation work
- Building lists, crossreferencing, mapping relationships, sourcing and curating data

What about PFAS?



9

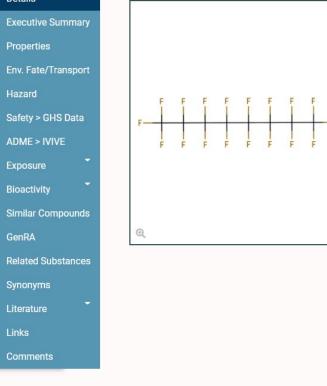
0



Chemical Details

Perfluorooctanesulfonic acid 1763-23-1 | DTXSID3031864 Searched by DTXSID3031864.

Details



Wikipedia

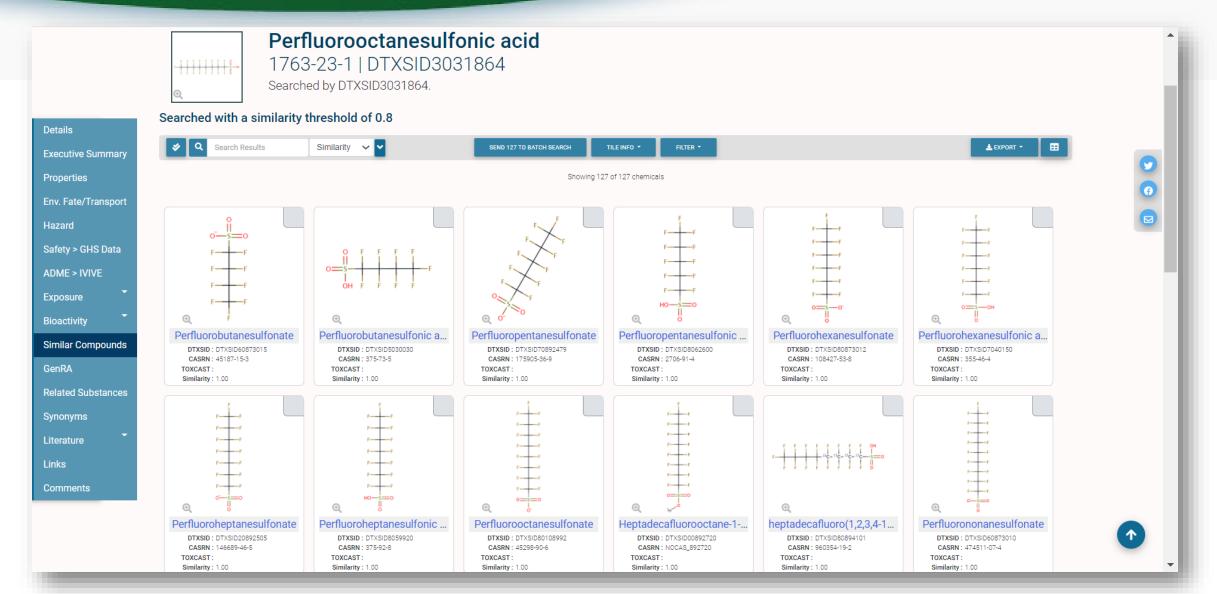
Perfluorooctanesulfonic acid (PFOS) (conjugate base perfluorooctanesulfonate) is an anthropogenic (human-made) fluorosurfactant, now regarded as a global pollutant. PFOS was the key ingredient in Scotchgard, a fabric protector made by 3M, and related stain repellents. In many contexts, PFOS refers to the parent sulfonic acid and its various salts of perfluorooctanesulfonate. These are all colorless or white, water soluble solids. Although of low acute toxicity, PFOS has

Read more



Are there Similar Compounds?

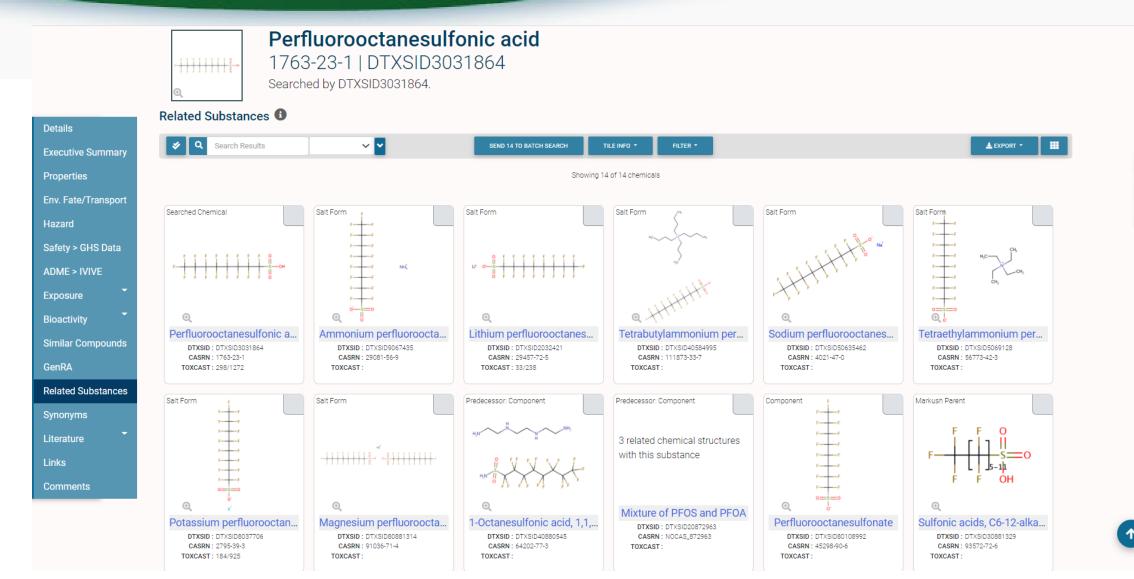




Relationships in the data



Ø

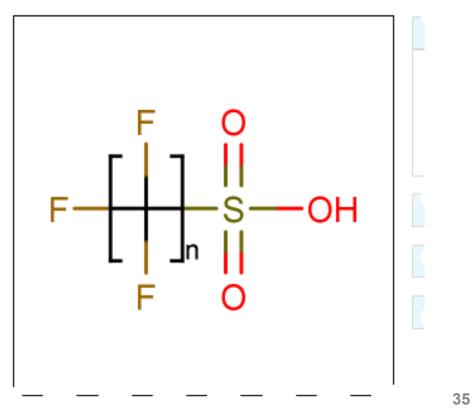


Markush Chemicals



• PFOS is a member of linear perfluoroalkyl sulfonates

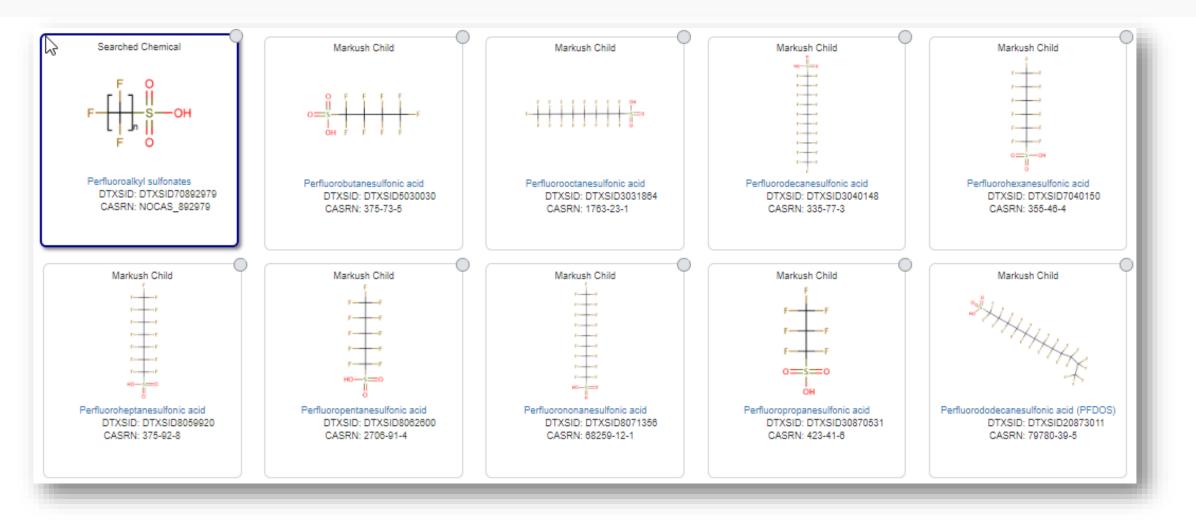
Perfluoroalkyl sulfonates NOCAS_892979 | DTXSID70892979 Searched by DSSTox Substance Id.



...and their Markush Children...



• Linear perfluoroalkyl sulfonates has children...







CompTox Chemicals Dashboard	Home Search - List	ts 🕶 About 👻 Tools 👻		Submit Comments	Search all data	<u>۷</u> ۹
FOOD: EFSA OpenF	oodTox					
Search for chemical by systematic	c name, synonym, CAS number, DT	XSID or InChIKey				~ Q
Identifier substring search						•
List Details						~ 🖂
Search Results	PubMed V		SEND 4237 TO BATCH SEARCH TILE INFO -	FILTER -	LEXPORT - PREFERRED VIE	w • 🖽
		Showir	ng 4237 of 4237 chemicals			
\mathbf{e} $\mathbf{h}_{c} \leftarrow \mathbf{h}_{c} \leftarrow $	0 related chemical structures with this substance Starch, phosphate DTXSID : DTXSID8093946 CASRW : 11120-02-8 TOXCAST :	H ₉ C \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow	OH 0 CH ₃ U 0 0 Cis-beta-Zearalenol 0 DTXSID: DTXSID201022837 CASRN: NOCAS_1022837 CASRN: NOCAS_1022837 TOXCAST:	$HO + H_3C + CH_3$ C C C $C(2,2,3-Trimethylcyclopen$ $DTXSID : DTXSID30852779$ $CASRN : 1901-38-8$ $TOXCAST :$	Furan, 2,2'-[thiobis(methyl DTXSID : DTXSID3065580 CASRN : 13678-67-6 TOXCAST :	

MTOX biomarkers



MTox700+ metabo	lic biomarkers				
Search for chemical by systemat	ic name, synonym, CAS number, DT)	(SID or InChIKey			✓ Q
Identifier substring search					
List Details					^
Description: List of metabolic bio Number of Chemicals: 627	markers associated with the publication	"Knowledge-Driven Approaches to Create	e the MTox700+ Metabolite Panel for P	redicting Toxicity" authored by Viant et	al.
Search Results	× ^	S	END 1000 TO BATCH SEARCH TILE INFO -	alter -	LEXPORT - PREFERRED VIEW -
		Showing 7	1254 of 1254 chemicals		
\mathbf{C}	Сну Сну Сну (8Z,11Z,14Z)-8,11,14-Еісо DTXSID: DTXSID00912351 ОКОВИ. 1700000		$(\mathbf{y}_{H}) = \mathbf{y}_{H}$	$CH_3 OH \\ H_3C V V V V V V V V V V V V V V V V V V V$	$H_3C \xrightarrow{N} H_2N \xrightarrow{N} H_2N \xrightarrow{N} H_2N \xrightarrow{N} H_3$



Batch Searching

Batch Searching



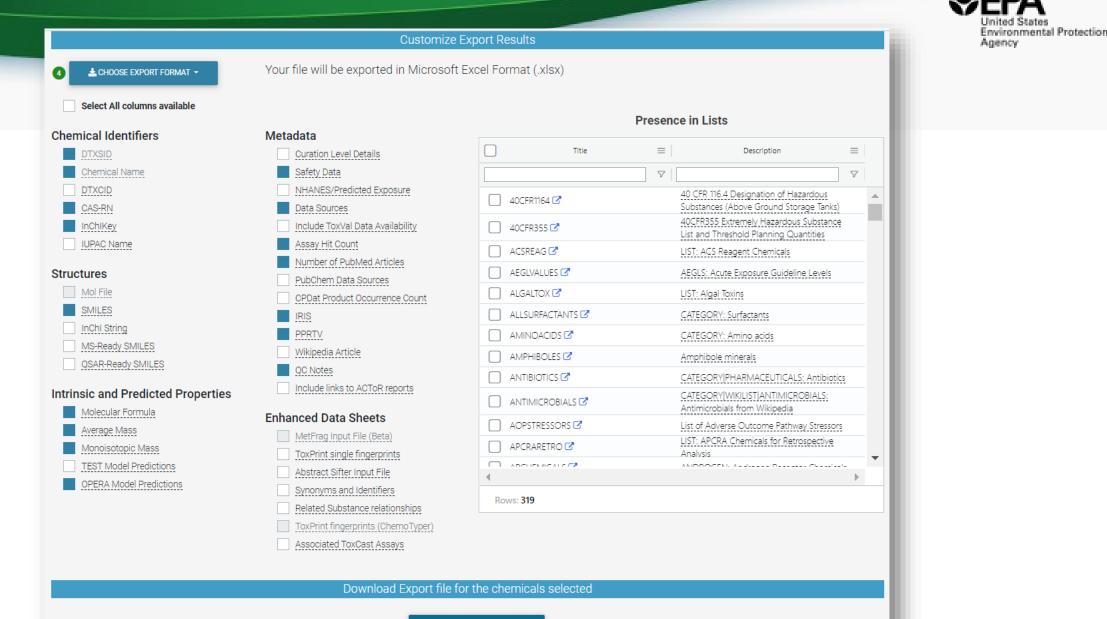
- Singleton searches are great but...
- ... we generally want data on LOTS of chemicals!
- Typical questions
 - What are the structures for a set of chemical names? Set of CASRNs?
 - Can I get chemical lists in Excel files? As a list of SMILES strings?
 Can I get an SDF file?
 - Can I include predicted properties? OPERA? TEST?
 - Are "these chemicals" screened in Toxcast?
 - I need masses and formulae for a list of chemicals

Batch Search



Batch Search	
Select Input Type(s) Substance Identifiers	2 Enter Identifiers to Search
Chemical Name CASRN InChIKey DSSTox Substance ID	(Please enter one identifier per line. Processing time increases with number of inputs.) DTXSID9020374 DTXSID9020827 DTXSID2022678 DTXSID4023381
 DSSTox Compound ID InChIKey Skeleton MS-Ready Formula(e) Exact Formula(e) Monoisotopic Mass 	DTXSID9044164 DTXSID7032004 DTXSID4022361 DTXSID8021771
Monoisotopic Mass 45% loaded	3 OISPLAY ALL CHEMICALS OF OCHOOSE EXPORT OPTIONS
REPLACE IDENTIFIERS WITH SELECTED CHEMICALS	
Structure DTXSID = Preferred Name	$\equiv CASRN \equiv Mono. Mass \equiv Mol. Formula \equiv$
DTXSID2022678 Bicalutamide	90357-06-5 430.061041 <u>C18H14F4N2O4S</u>
DTXSID3020621 (R,R)-Fenvalerate	67614-33-9 419.128821 <u>C25H22CINO3</u>

Batch Search – Excel, CSV, SDF file



A DOWNLOAD EXPORT FILE

5

Batch Search



AutoSave 💽 🕅 🏷 🖓 🗸 🔫	CCD-Batch-Search_2	022-03-27_05_36_52.xlsx 👻	✓ Search			Williams, A	ntony 🙀 🖻	- o x
File Home Insert Draw Page La	ayout Formulas	Data Review View	Developer He	elp			🖻 Share	Comments
$\begin{array}{c c} & & \\ & & \\ & & \\ & \\ & \\ & \\ & \\ & \\ $		ab c Wrap Text ⊡ Merge & Center ↓	General \$ ~ % 9	00 Conditional Fo	ormat as Cell able ~ Styles ~	Insert → Σ Delete → H Format →	∑ ~ AZV ZV Sort & Find & Filter * Select *	Sensitivity
Clipboard 🖸 Font	A IA	ignment 🛛	Number	Sty	les	Cells	Editing	Sensitivity 🔨
A2 \rightarrow : \times \checkmark f_x DTXS	ID9020299							~
A B	C D	E F	G	н і	J	K L	M N	0 P 🔺
1 DTXSID PREFERRED_NAME	INCHIKEY CASRN	SMILES MOLECULAR_FO	DRMULA AVERAGE	MONOISO SAFETY_E	DATA_SOUNU	MBER_IRIS_LINK	ATMOSPH BIOCONCE	BIODEGRA BOILIN
2 DTXSID9020299 Chlorobenzilate	RAPBNVD:510-15-6	CCOC(=O) C16H14Cl2O3	325.19	324.032 Y	154	16 Y	1.37E-11 477.542	4.6243 349.9
3 DTXSID6034712 Mesosulfuron-methyl	NIFKBBMC 208465-21	COC(=O)C C17H21N5O9S2	503.5	503.0781 Y	95	10	1.79E-11 3.2453	4.26547 254.0
4 DTXSID7034753 Foramsulfuron	PXDNXJSD 173159-57	COC1=CC(C17H20N6O7S	452.44	452.1114 Y	95		2.35E-11 3.84639	5.67465 265.1
5 DTXSID1033664 17-Methyltestosterone	GCKMFJBC58-18-4	C[C@]1(O C20H30O2	302.458	302.2246 Y	145	1377	3.99E-11 62.2298	97.9166 294.8
6 DTXSID8034401 Buprofezin	PRLVTUNV 69327-76-	CC(C)N1C(C16H23N3OS	305.44	305.1562 Y	134	42	1.38E-11 52.49	6.89035 353.7
7 DTXSID0020529 2,4-Dinitrotoluene	RMBFBMJ 121-14-2	CC1=C(C=(C7H6N2O4	182.135	182.0328 <mark>Y</mark>	198	379 <mark>Y</mark>	1.63E-12 9.12436	3.5609 299.8
8 DTXSID2034673 Iodosulfuron methyl ester	JUJFQMPK 144550-36	[Na+].COC C14H13IN5NaO	6S 529.24	528.9529 Y	88		1.77E-11 3.51252	4.73647 265.0
9 DTXSID7024247 Pentachlorobenzene	CEOCDNV 608-93-5	CIC1=CC(CC6HCI5	250.32	247.8521 <mark>Y</mark>	170	84 Y	2.25E-13 5620.75	6.16855 277.0
10 DTXSID0034227 Icaridin	QLHULAH(119515-38	CCC(C)OC C12H23NO3	229.32	229.1678 Y	111	67	1.82E-11 3.81105	5.17405 251.8
11 DTXSID0020440 Dichlorprop	MZHCENG 120-36-5	CC(OC1=C C9H8Cl2O3	235.06	233.985 Y	164	89	1.16E-11 3.54397	3.53597 298.4
12 DTXSID9034816 Monocrotophos	KRTSDMXI 6923-22-4	CNC(=O)\(C7H14NO5P	223.165	223.061 Y	152	274	2.77E-11 0.922318	4.13837 301.7
13 DTXSID8021301 Tamoxifen citrate	FQZYTYWI 54965-24-	OC(=O)CC C32H37NO8	563.647	563.2519 Y	90	17257	2.9E-11 1209.93	3.36316 419.3
14 DTXSID7032553 Flumetralin	PWNAWO 62924-70-	CCN(CC1= C16H12ClF4N3C	421.73	421.0452 Y	117		1.38E-11 35265.1	3.54617 347.7
15 DTXSID6024048 Difenzoquat metilsulfate	XQEMNBN 43222-48-	COS([O-])(C18H20N2O4S	360.43	360.1144 Y	79	20 Y	1.93E-11 565.107	13.7793 335
16 DTXSID3024104 Fluoranthene	GVEPBJHC 206-44-0	C1=CC2=C C16H10	202.256	202.0783 Y	211	398 Y	4.98E-11 3528.4	147.199 393.9
17 DTXSID8023890 Asulam	VGPYEHKC3337-71-1	COC(=O)N C8H10N2O4S	230.24	230.0361 Y	133	19 Y	1.21E-11 2.50573	4.63676 254.8
18 DTXSID4032532 Carfentrazone-ethyl	MLKCGVH 128639-02	CCOC(=O) C15H14Cl2F3N3	03 412.19	411.0364 Y	133	9	2.16E-11 192.141	4.88739 352.4
19 DTXSID5032498 Triclosan	XEFQLINVI 3380-34-5	OC1=C(OCC12H7Cl3O2	289.54	287.9512 Y	246	2221	1.74E-11 52.8927	4.50619 342.3
20 DTXSID1021160 Picloram	NQQVFXU 1918-02-1	NC1=C(CI) C6H3Cl3N2O2	241.45	239.926 Y	186	133 Y	8.15E-12 2.72427	4.09513 296.1
21 DTXSID9020160 Bifenthrin	OMFRMAI 82657-04-	CC1=C(C=(C23H22ClF3O2	422.87	422.126 Y	172	246 Y	3.32E-11 4990.65	3.54377 370.9
Cover Sheet Main Data (+)				100	4400		•
Ready 🐻						=	▣ 巴	+ 100%



Open Data Exchange

Since our data are Open...



• They flow into other systems for benefit ...

- ECHA eChemPortal
- ChemSpider
- EBI's UniChem
- PubChem

0 Use and Man	ufacturing	? 2
0.1 Uses		0 2
PA CPDat Chemical and Pro	duct Categories	
0 items View More 🛛		👤 Downloa
	SORT BY 🚖 Category	~
Category	Category Description	Categorization Type
Construction and building materials	Materials used for construction (e.g. flooring, tile, sinks, bathtubs, mirrors, wall materials/drywall, wall-to-wall carpets, insulation, playground su 🔄	Product Use Category (PUC)
binder		OECD Functional Use
binder		Reported Functional Use
catalyst		Reported Functional Use
epoxy hardener, curing agent		Reported Functional Use

PubChem bisphenol A (Compound) nttps://www.epa.gov/cnemicals-unaer-tsca 7. EPA DSSTox LICENSE https://www.epa.gov/privacy/privacy-act-laws-policies-and-resources Bisphenol A https://comptox.epa.gov/dashboard/DTXSID7020182 CompTox Chemicals Dashboard Chemical Lists https://comptox.epa.gov/dashboard/chemical-lists/ Classification ? 20.699.409 Ontologies ? 20.699.409 CAMEO Chemicals ? 4.856 CCSBase Classification ? 4.911 ChEBI Ontology ? 136,017 ChEMBL Target Tree ? 1,021,969 ChemIDplus ? 333,180 Consumer Product Information Database Classification ? 3,701 Drug Enforcement Administration (DEA) Classification ? 628 EPA CPDat Classification ? 9,625 EPA DSSTox Classification ? 40.613 EPA Safer Choice ? 994 FDA Drug Type and Pharmacologic Classification ? 3.929



Cheminformatics "PoC Modules" https://www.epa.gov/chemicalresearch/cheminformatics





- Hazard Comparison Profiling profile chemicals based on hazard
- Alerts structure, substructure, SMARTS based alerts and flags
- Predict batch prediction using WebTEST (100s of structures)
- Search structure/substructure/similarity searches
- Standardize convert structures into QSAR/MS-Ready forms
- ToxPrints generate ToxPrint substructural fragments and profile

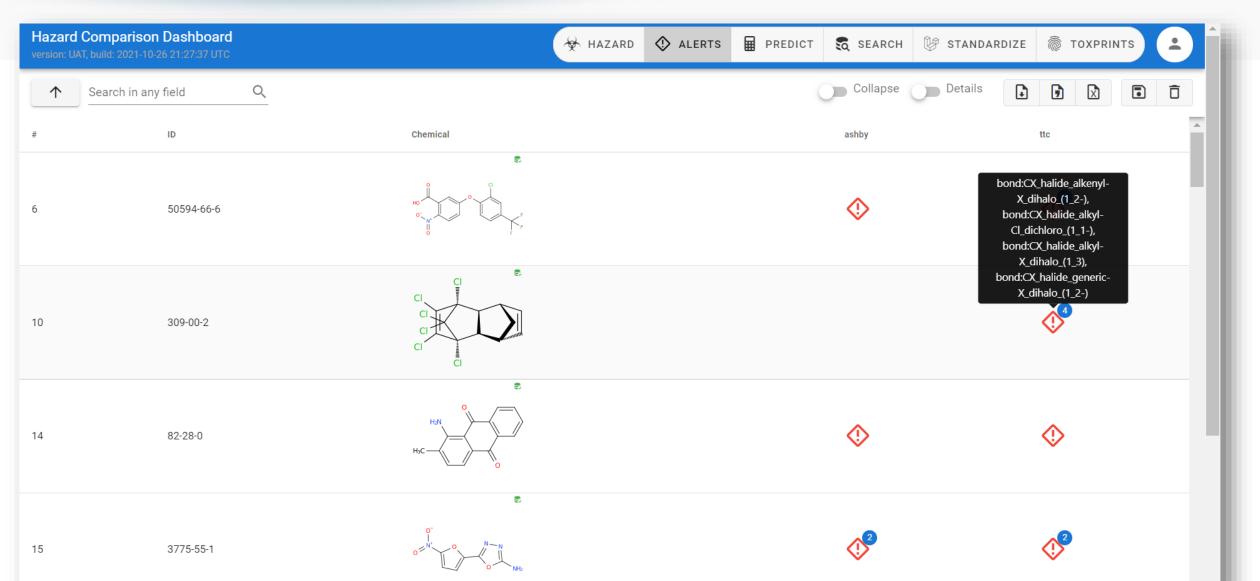
Module 1: Hazard Module



												🛠 HAZ	ARD	PR	EDICT	Concession of the local division of the loca	EARCH d assessn	~	TAND	ARC
																Full			¢	7
				Toxicity:	VH - Ver		- High <mark>M</mark> Human		11 C	w I - Inco	onclusive N	/A - Not App	licable	Authority.	Authorita	Emerg	m Jency Re pecific S		G	>
 Skipped (6) Unlikely (0) Filters (0) Sorting (0) Structure CAS Name 	Acute M	Iammaliar Inhalation	Dermal Dermal	Carcinogenicity	Genotoxicity Mutagenicit	Endocrine Disruption	Reproductive	Developmental	Repeat Exposure	Single Exposure	Systemic Kebeat Exposure	Single Exposure	Skin Sensitization	Skin Irritation	Eye Irritation	Acute Aquatic Toxicity	Chronic Aquatic Toxicity	Persistence	Bioaccumulation	
60-35-5 Acetamide	L	I	I	νн	VH	L	М	м	1	I	L	1	1	1	1	L	L	L	L	
107-13-1 Acrylonitrile	н	н	н	VH	VH	L	н	н	н	н	н	м	н	н	VH	н	н	н	L	
1912-24-9 Atrazine	м	н	L	νн	L	Н	Н	н	н	М	м		н	L	м	VH	νн	н	L	

Module 2: Alerts





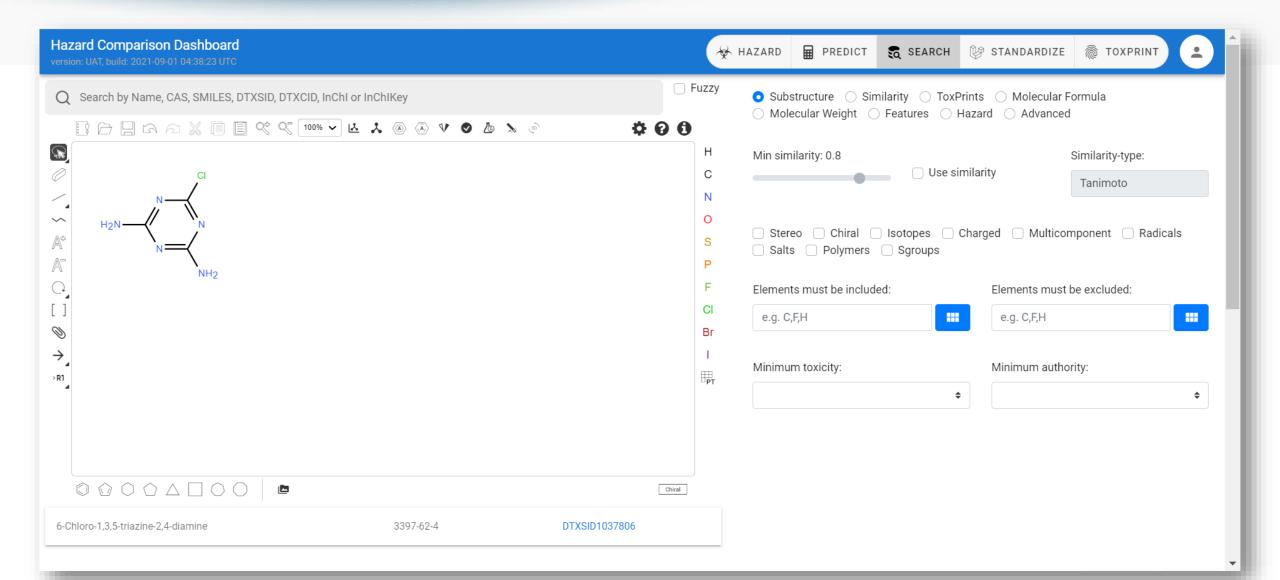
Module 3: WebTEST Batch Prediction



azard Comparison Dashboard rsion: DEV, build: 2021-09-22 14:04:34 UTC										₩ 1	IAZAR	D E	PRE	лст	🕄 si	EARCH	I U	STAND	ARDIZE	TOXPRINT
PREDICT	None 🗢	Expo	ort 🔻	RESI	ET										28 /	28	$\overline{\mathbf{v}}$	Q		
									Leg	end:A =	Active	, N = No	ot activ	e, I = In	conclus	sive, <u>Ex</u>	<u>(perime</u>	<u>ntal</u> , Pre	edicted	
 Filters Structure Products CAS Name 		Fathead minnow LC50 (96 hr)	Daphnia magna LC50 (48 hr)	T. pyriformis IGC50 (48 hr)	Oral rat LD50	Bioconcentration factor	Developmental Toxicity	Mutagenicity	Estrogen Receptor Binding	Estrogen Receptor RBA	Normal boiling point	Vapor pressure at 25°C	Melting point	Density	Flash point	Surface tension at 25°C	Thermal conductivity at 25°C	Viscosity at 25°C	Water solubility at 25°C	
18540-29-9 Chromium (VI) ior	HIGBT																			
375-85-9 Perfluoroheptanoi.	HGBTM	4.93	3.22	4.74	3.67	1		2.25				<u>177</u> 183	<u>-0.88</u> -0.22	-0	<u>1.74</u> 1.68	<u>51</u> 60	12.18	68.00	1	
375-95-1 Perfluorononanoic	GBTM	5.72	3.42	5.54		<u>N</u> /		2.51				200	-1.17	5	1.80	74	1	62.29	1	
1763-23-1 Perfluorooctanesu	GBTM	5.33	1	5.74		1		<u>3.73</u> 2.96				244	I	15	<u>1.84</u> 1.85	1	1	I		
96-18-4 1,2,3-Trichloropro	AIGBT	<u>3.41</u> 3.62	<u>3.65</u> 3.34	4.04	<u>2.99</u> 2.10	<u>1.02</u> 1.21	N	A A	N	I	<u>157</u> 145	<u>0.57</u> 0.82	<u>-15</u> -43	<u>1.39</u> 1.29	<u>74</u> 53	34.57	7 118.34	4 0.01	<u>1.93</u> 1.93	
16065-83-1	IIGBT																			
Chromium (III)	GBTM		1 96				٨	N	٨	2.00			170	1 17	210				1.88	

Module 4: Structure/Substructure/Similarity







(Work in Progress)





About

CTS Home

CTS Basic Information

Execute CTS Workflows

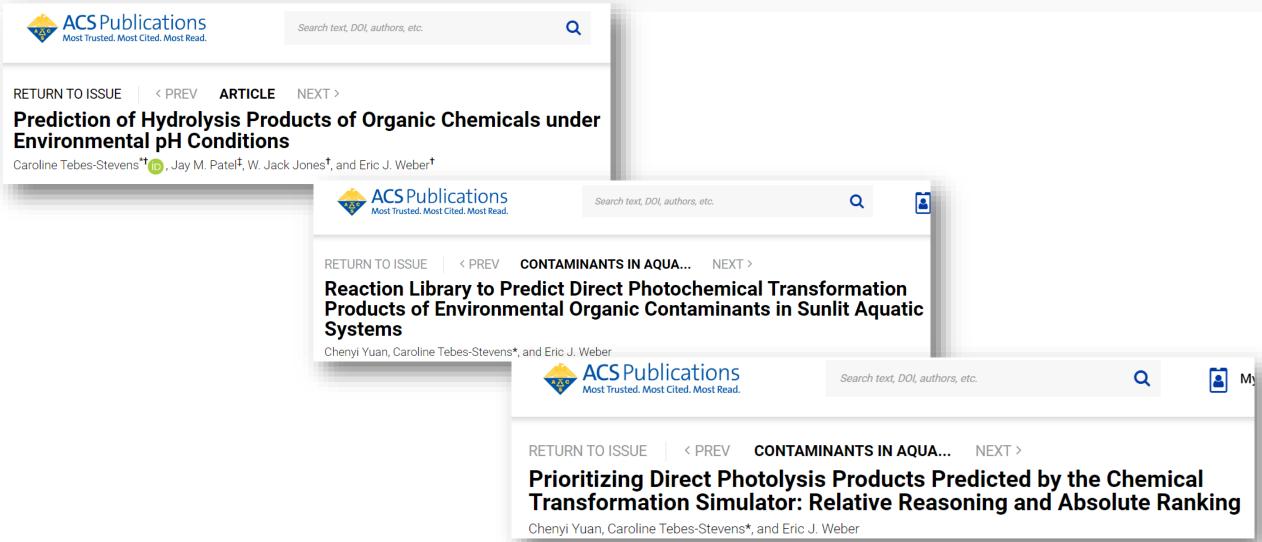
> Calculate Chemical Speciation

Calculate Physicochemical



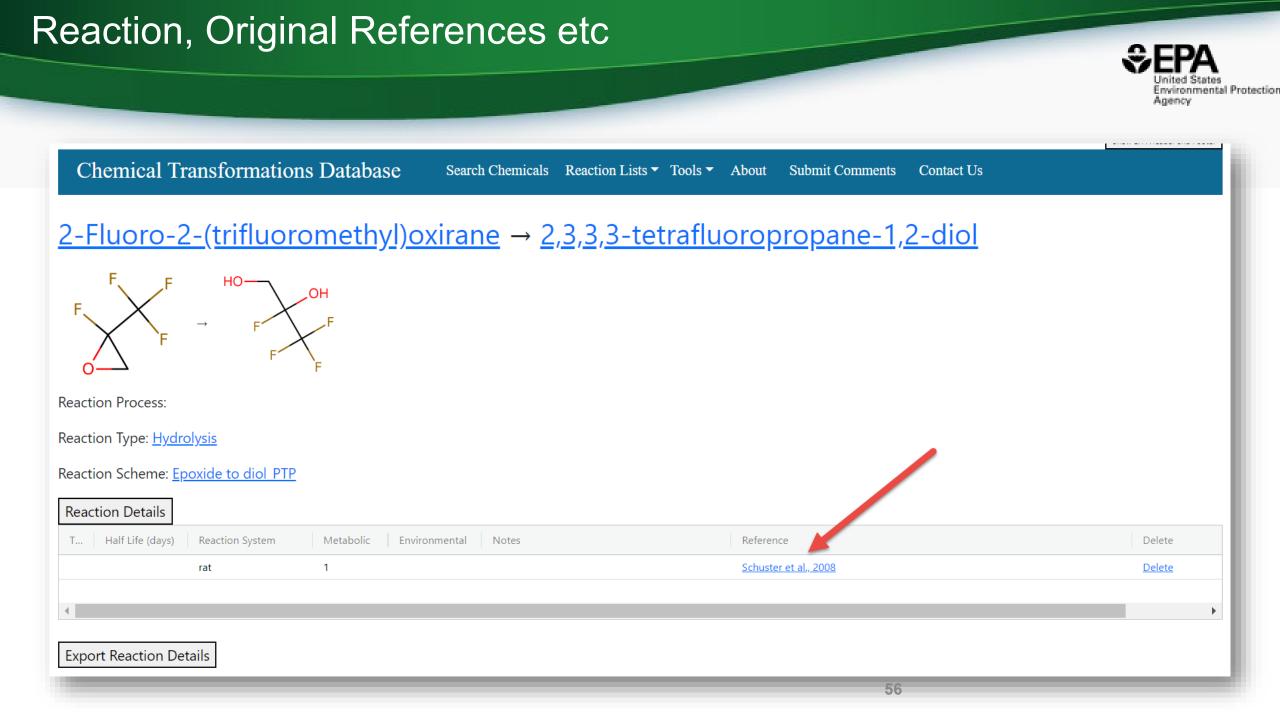


The database of reactions under the Chemical Transformation Simulator





Chemical Transform	mations Database	Search Chemicals Reaction Li	sts ▼ Tools ▼ About Subm	it Comments Contact Us	Show EPA Header and Footer
Search: Name, DTXSID, CASR	N, InChI key				
Table View Export Chemical	List				
Name: <u>(2Z)-4,4,5,5,6,6,7,7,8,</u>	Name: <u>2,3,3,3-tetrafluoropr</u>	Name: <u>γ-glutamyl-S-[(2Z)-4,</u>	Name: <u>γ-glutamyl-S-(3,3,3-t</u>	Name: <u>4,4,5,5,6,6,7,7,8,8,9,9,</u>	Name: <u>(2E)-4,4,5,5,6,6,7,7,8,</u>
	HO F F F F				r <u>+</u> ++++++
DTXSID: DTXSID301348999 >	DTXSID: DTXSID201349001 🗷	DTXSID: DTXSID601349003 🗡	DTXSID: DTXSID001349005 🗡	DTXSID: DTXSID301349004 🗡	DTXSID: DTXSID901349002 🗡
Name: <u>1-chloro-1,2,2,2-tetr</u>	Name: <u>γ-glutamyl-S-(3,3,3-t</u>	Name: <u>(2E)-4,4,5,5,6,6,7,7,7-</u>	Name: <u>3,4,4,5,5,6,6,7,7,8,8,9,</u>	Name: <u>2,2,3,3,5,6,6,7,7,8,8,8-</u>	Name: <u>30,30,31,31,32,32,33,</u>
					·+++++++



Migrating MetaPath data





Regulatory Toxicology and Pharmacology Volume 63, Issue 1, June 2012, Pages 84-96

MetaPath: An electronic knowledge base for collating, exchanging and analyzing case studies of xenobiotic metabolism

Richard C. Kolanczyk ^a $\stackrel{\sim}{\sim}$ $\stackrel{\boxtimes}{\sim}$, Patricia Schmieder ^a, William J. Jones ^b, Ovanes G. Mekenyan ^c, Atanas Chapkanov ^c, Stanislav Temelkov ^c, Stefan Kotov ^c,



METAPATH

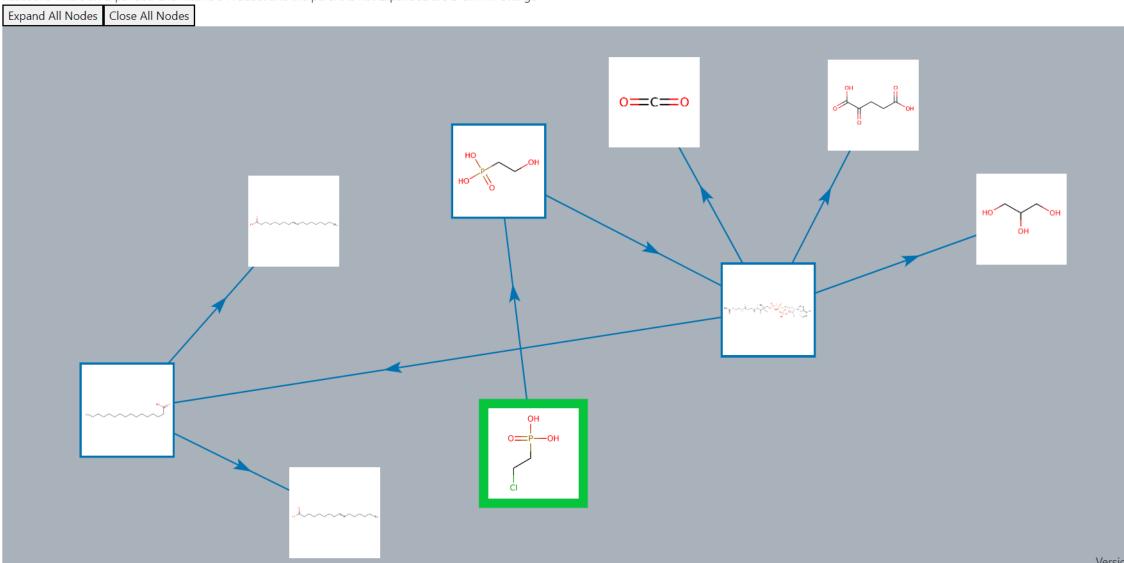
Consolidation and management of metabolism data

The MetaPath platform is developed to collect, organize and analyze experimental data on metabolism or catabolism, observed biotransformation pathways and crucial supporting metadata. Thus, the results from metabolism studies are compiled and organized into a systematic database. The system provides a set of tools for managing and accessing the information, including powerful search engine that can extract information based on chemical structure, biotransformation, experimental metadata and specific proteins responsible for a given biotransformation.

Work in Progress: MetaPath migration



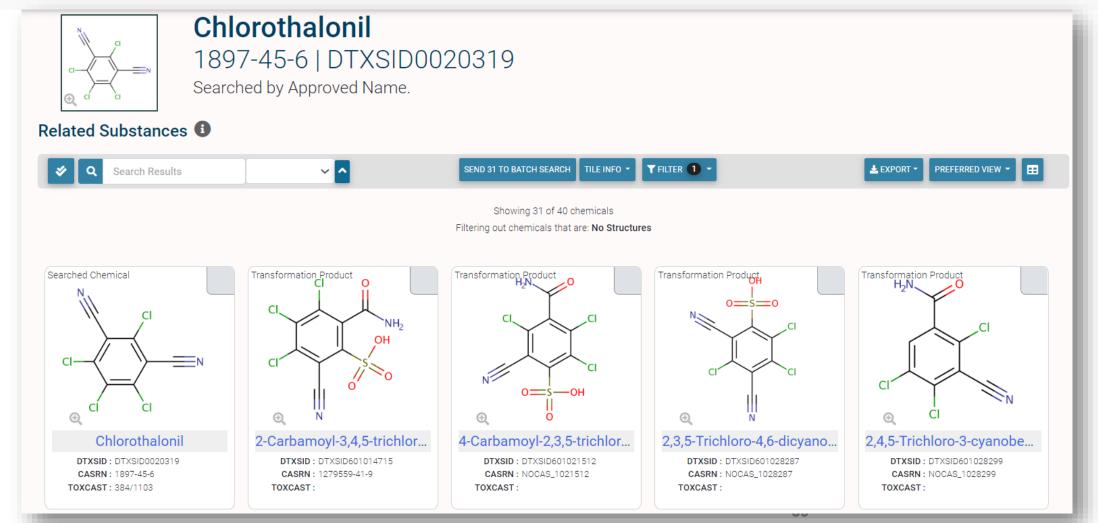
Reactions where an Expanded Chemical is a Product and the parent is not Expanded are Drawn in Orange







Adding Thousands of Parent-Product mappings from the Dashboard





Supporting Analytical Chemistry 1) Non-Targeted Analysis

- 2) Open Spectral Data
- 3) Analytical Methods

(Work in Progress)



Chemical Identification and Exposure





Suspect Screening Ana

Katherine A. Phillips,[†] Alice Yau,[‡] Christopher Grulke,^{||} Ann M. Rich, and John F. Wambaugh^{*,||}

[†]National Exposure Research Laboratory, Of Alexander Drive, Research Triangle Park, N [‡]Southwest Research Institute, San Antonio, [§]Oak Ridge Institute for Science and Educat ^{II}National Center for Computational Toxicol T. W. Alexander Drive, Research Triangle F

Supporting Information

ABSTRACT: A two-dimensional gas (GC×GC-TOF/MS) suspect screening a chemicals in 100 consumer products—wh articles (e.g., upholsteries, shower curtai broader efforts to prioritize chemicals base 4270 unique chemical signatures across the using the National Institute of Standards standards confirmed the presence of 11 chemicals, 1404 were not present in a pub Reported data and model predictions of c tentative chemical identifications. Estim manufacturer-reported values and other m data can now be used to improve estimate posed to human health and the environm



pubs.acs.org/est

Chemical Characterization Chemical Characterization

Charles N. Lowe, Katherine A. Phillips, Kri Antony J. Williams, Ashley J. Pfirrman, an

Cite This: Environ. Sci. Technol. 2021, 55, 11375–113

ACCESS | Metrics & More

ABSTRACT: Recycled materials are found in many part of a circular economy; however, the chemica products is generally uncharacterized. A suspect screen dimensional gas chromatography time-of-flight mass GC-TOFMS) was applied to 210 products (154 rec seven categories. Chemicals in products were tentat standard spectral library or confirmed using chemica 918 probable chemical structures identified (112 of in recycled materials versus 587 (110 confirmed Identified chemicals were characterized in terms of t structural class. Recycled paper products and o contained greater numbers of chemicals than virgin p chemicals had greater occurrence in recycled compa Products made from recycled materials contained gre results were clustered to identify groups of chemicals were prioritized for further study using high-throug indicative of risk, these results can be used to inforneglected in exposure assessments.

KEYWORDS: recycling, consumer products, human ex suspect screening



Full length article

Integrative exposomic, transcriptomic, epigenomic analyses of human placental samples links understudied chemicals to preeclampsia

Alex Chao^{a,*}, Jarod Grossman^b, Celeste Carberry^{c,d}, Yunjia Lai^c, Antony J. Williams^a, Jeffrey M. Minucci^e, S. Thomas Purucker^f, John Szilagyi^{c,d}, Kun Lu^{c,d,g}, Kim Boggess^h, Rebecca C. Fry^{c,d,g}, Jon R. Sobus^a, Julia E. Rager^{c,d,g,*}

^a U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Chemical Characterization and Exposure Division, Research Triangle Park, NC, USA

^b Agilent Technologies, Santa Clara, CA, USA

^c Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA ^d The Institute for Environmental Health Solutions, Gillings School of Global Public Health, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA ^e U.S. Environmental Protection Agency, Office of Research and Development, Center for Public Health and Environmental Assessment, Public Health and Environmental

Systems Division, Research Triangle Park, NC, USA ^f U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Great Lakes Toxicology and Ecology Division, Research Triangle Park, NC, USA

³ Curriculum in Toxicology and Environmental Medicine, School of Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

^h Department of Obstetrics and Gynecology, Division of Maternal Fetal Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Environment International 167 (2022) 107385

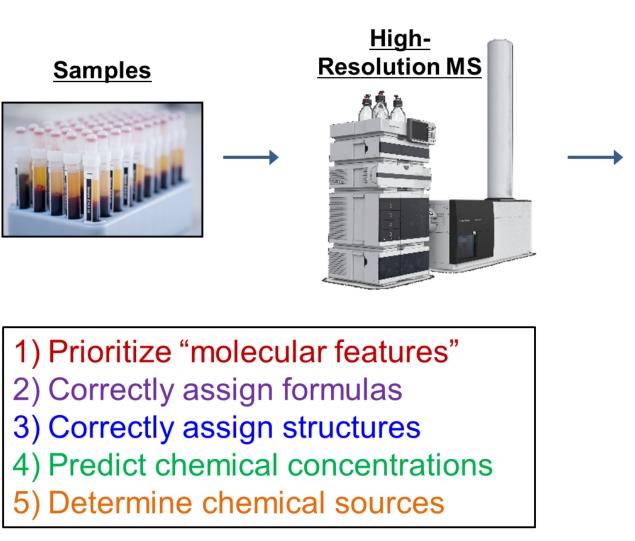
Contents lists available at ScienceDirect

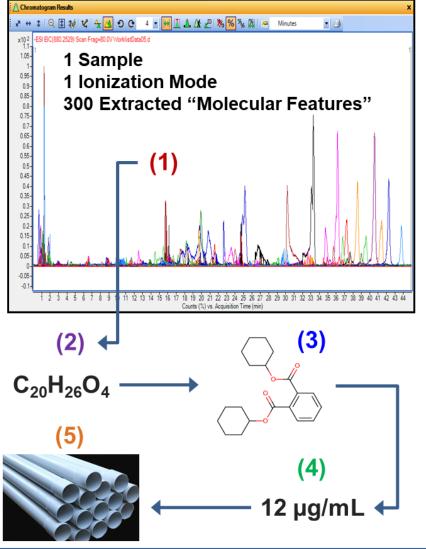
Environment International

journal homepage: www.elsevier.com/locate/envint

Rudimentary NTA Workflow







Non-Targeted Analysis Informatics



RAPID COMMUNICATION

Identifying known unknowns using the US EPA's CompTox Chemistry Dashboard

Andrew D. McEachran¹ · Jon R. Sobus² · Antony J. Williams³

Environmental Toxicology and Chemistry—Volume 41, Number 5—pp. 1117–1130, 2022 Received: 28 June 2021 | Revised: 26 July 2021 | Accepted: 17 August 2021

Critical Perspectives

A Framework for Utilizing High-Resolution Mass Spectrometry and Nontargeted Analysis in Rapid Response and Emergency Situations

Allison L. Phillips,^a Antony J. Williams,^b Jon R. Sobus,^b Elin M. Ulrich,^b Jennifer Gundersen,^c Christina Langlois-Miller,^d and Seth R. Newton^{b,*}

CrossMark



Article

Revisiting Five Years of CASMI Contests with EPA Identification Tools

Andrew D. McEachran ^{1,*}, Alex Chao ¹, Hussein Al-Ghoul ¹, Charles Lowe ², Christopher Grulke ², Jon R. Sobus ² and Antony J. Williams ^{2,*}

1117





The Next Phase Assembling Public Domain Spectra



- Assemble data from MoNA, MassBank, internal EPA data
- Already have CFM-ID data for 1.2M chemicals

scientific data

Explore content ~ About the journal ~ Publish with us ~

nature > scientific data > data descriptors > article

Data Descriptor Open Access Published: 02 August 2019

Linking *in silico* MS/MS spectra with chemistry data to improve identification of unknowns

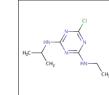
Andrew D. McEachran , Ilya Balabin, Tommy Cathey, Thomas R. Transue, Hussein Al-Ghoul, Chris Grulke, Jon R. Sobus & Antony J. Williams

 Scientific Data
 6, Article number: 141 (2019)
 Cite this article

 4853
 Accesses
 21
 Citations
 10
 Altmetric
 Metrics

• Experimental searching

92 Results for "atrazine"

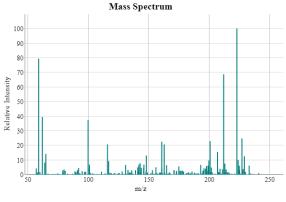


(Preferred) Name: Atrazine DTXSID: <u>DTXSID9020112</u> CASRN: 1912-24-9 InChIKey: MXWJVTOOROXGIU-UHFFFAOYSA-N Molecular Formula: C8H14CIN5 Mass: 215.69

Include Single Point Spectra

All Result	s (92) Spectra	ı (77)	Monographs (0)	Methods (15
Spectrum Type ↑	Source	Record Type	Information	
LC-MS	MoNA	<u>Spectrum</u>	ESI-QFT; MS2; CE:35HCD; Nega	ative; # PEAKS=538
LC-MS	MoNA	Spectrum	ESI-QFT; MS2; CE:45HCD; Nega	ative; # PEAKS=353
LC-MS	MoNA	Spectrum	ESI-QFT; MS2; CE:65HCD; Nega	ative; # PEAKS=259
LC-MS	MoNA	Spectrum	ESI-QFT; MS2; CE:35HCD; Posit	ive; # PEAKS=254
LC-MS	MoNA	Spectrum	ESI-QFT; MS2; CE:45HCD; Posit	tive; # PEAKS=172
LC-MS	MoNA	Spectrum	ESI-QFT; MS2; CE:65HCD; Posit	ive; # PEAKS=211
LC-MS	MoNA	Spectrum	Linear Ion Trap; MS2-MS5 Com	nposite; CE:35; Posit
LC-MS	MoNA	Spectrum	Linear Ion Trap; MS2; CE:35; Ne	egative; # PEAKS=4
LC-MS	MoNA	Spectrum	Linear Ion Trap; MS2-MS2 Com	nposite; CE:35; Neg;
LC-MS	MoNA	Spectrum	LC-ESI-QTOF; MS2; CE:10 V; Po	sitive; # PEAKS=6
LC-MS	MoNA	Spectrum	LC-ESI-QTOF; MS2; CE:20 V; Po	sitive; # PEAKS=7
LC-MS	MoNA	Spectrum	LC-ESI-QTOF; MS2; CE:40 V; Po	sitive; # PEAKS=6
LC/MS/MS	Environmental Chemistry Methods	Method	None	
LC/MS/MS	Environmental Chemistry Methods	Method	None	
LC/MS/MS	Environmental Chemistry Methods	Method	None	
LC/MS/MS	Environmental Chemistry Methods	Method	None	
LC/MS/MS	Environmental Chemistry Methods	Method	None	
LC/MS/MS	Environmental Chemistry Methods	Method	None	
LC/MS/MS	Environmental Chemistry Methods	Method	None	
Calcale	e e la companya de la		K 1	

Download Results



Information

Number of Points: 538 Spectral Entropy: 4.1850 Normalized Entropy: 0.7015 Rating: Noisy SPLASH: splash10-08mi-5890000000-4861d94f1816f880e7a8



- Simple Vision: Find the best method(s) for a particular chemical
- Approach:
 - Aggregate MS method documents
 - Extract chemistry (mostly CASRN and Names)
 - Map CASRN and Names to structures
 - Build a searchable database of chemicals mapped to methods
 - Chemical identifier search
 - Structure/substructure and similarity search

Where are there methods?



900 method documents

Related Topics: Pesticide Analytical Methods

CONTACT US

Environmental Chemistry Methods (ECM) Index - 0-9

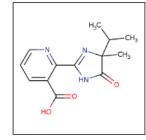
 $\textbf{0-9} \ | \ \underline{A} \ | \ \underline{B} \ | \ \underline{C} \ | \ \underline{D} \ | \ \underline{E} \ | \ \underline{E} \ | \ \underline{G} \ | \ \underline{H} \ | \ \underline{I} \ | \ \underline{K} \ | \ \underline{L} \ | \ \underline{M} \ | \ \underline{N} \ | \ \underline{O} \ | \ \underline{P} \ | \ \underline{Q} \ | \ \underline{R} \ | \ \underline{S} \ | \ \underline{T} \ | \ \underline{V} \ | \ \underline{Z}$

Analyte(s) by Pesticide	ECM MRID	Matrix	Method Date
<u>1,2,4-triazole</u>	49762553	Water	2/19/13
<u>1,3-dichloropropene & 1,2-dichloropropane</u>	44536511	Soil	3/27/98
<u>1,3-dichloropropene & 1,2-dichloropropane</u>	44536511	Water	3/27/98
<u>1,3-dichloropropene Degradate 3-chloroallyl Alcohol</u>	44536505	Water	12/12/97

Embedding old Method PDFs



Search Results for "Imazapyr"

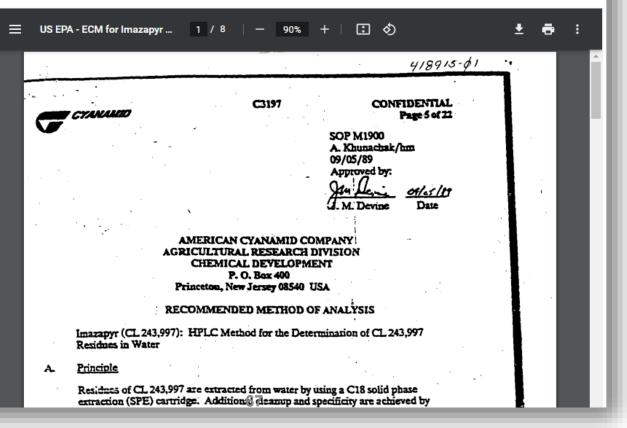


(Preferred) Name: Imazapyr DTXSID: <u>DTXSID8034665</u> CASRN: 81334-34-1 InChIKey: CLQMBPJKHLGMQK-UHFFFAOYNA-N Molecular Formula: C13H15N3O3 Mass: 261.281

Imazapyr

MRID: 41891501 Date: 10/1/89 Matrix: Water Registrant: American Cyanamide Co Analysis: HPLC/UV Limit of Quantitation: 5.0 µg/L

Spectrum Type 1	Source	Record Type	View
	Environmental Chemistry Methods	Method	PDF
	Environmental Chemistry Methods	Method	PDF
LC-MS+	MoNA	Spectrum	<u>Spectrum</u>
LC-MS+	MassBank EU	Spectrum	<u>Spectrum</u>
LC-MS+	MassBank EU	Spectrum	Spectrum
LC-MS+	MassBank EU	Spectrum	Spectrum
LC-MS+	MassBank EU	Spectrum	<u>Spectrum</u>
LC-MS+	MassBank EU	Spectrum	<u>Spectrum</u>



Vendor Methods

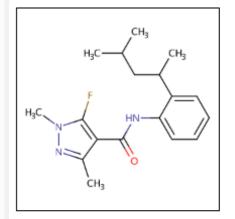


	States mental Protection y						Search EPA.gov
		1	Environmental Topics	Laws & Regulations	About EPA		
Search Term	Search					Home Monograph List	About this App Toggle Header/Footer
Search Res		ed) Name: Retinol): <u>DTXSID3023556</u> : 68-26-8 ey: FPIPGXGPPPQFEQ- ar Formula: C20H30O	-OVSJKPMPSA-N	Extraction wi Author: Hui Zi Focus/Analyte Synopsis: This (retinol), vitami matrices, includ extracted using were then simu	th LC/DAD : hao : Reliable sample pro- application note des in D3 (cholecalcifero ling infant formula, Agilent Chem Elut S ltaneously identified and Agilent 6470 tr	ble Vitamins in Foods Using and LC/MS/MS Triple-Qua eparation and identification/quantitation in v cribes a method for the determination of fat ol), vitamin D2 (ergocalciferol), and vitamir egg, canned tuna, and mushroom. Samples v S (Supported Liquid Extraction (SLE)) 12 m and quantified by an Agilent 1290 Infinity riple quadrupole LC/MS in series. Data were	drupole various food matrices -soluble vitamins, including vitamin A h E (α-tocopherol) in complex food were saponified as sample pretreatment, hL cartridges, and fat-soluble vitamins II LC coupled to an Agilent diode array
Spectrum Type ▽ ↑	Source	Record Type	View	PDF View	ver	Compounds (grid)	Compounds (table)
GC-MS	Spectrabase	Spectrum	External Link				
GC-MS	Spectrabase	Spectrum	External Link	📃 Determinatio	on of Fat-Solubl	1 / 13 - 100% + 🗄 🚸	• ± 🖶 :
GC-MS	Spectrabase	Spectrum	External Link				-
GC-MS	<u>Spectrabase</u>	Spectrum	External Link				
GC-MS	Spectrabase	Spectrum	External Link				
GC-MS	<u>Spectrabase</u>	Spectrum	External Link	Application	Note		Agilent
GC-MS	Spectrabase	Spectrum	External Link	Food Testin			Trusted Answers
GC-MS	Spectrabase	Spectrum	External Link	Agriculture			Husted Answers
LC-MS	Agilent	Method	PDF				
LC-MS+	MoNA	Spectrum	Spectrum				

Structure Tables/Tiles view



Search Results for "DTXSID9058107"



(Preferred) Name: Penflufen DTXSID: <u>DTXSID9058107</u> CASRN: 494793-67-8 InChIKey: GOFJDXZZHFNFLV-UHFFFAOYNA-N Molecular Formula: C18H24FN3O Mass: 317.408

Penflufen & Degradates

MRID: 48023715 Date: 1/8/10 Matrix: Water Registrant: Bayer CropScience Analysis: LC/MS/MS Limit of Quantitation: 0.1 µg/L

Spectrum Type 🗅	Source	Record Type	Info	PDF Viewe	r	Compo	ounds (grid)	Compounds (table)
LC/MS/MS	Environmental Chemistry Methods	Method (PDF)		Structure	DTXSID		CASRN	Compound Name
LC/MS/MS	Environmental Chemistry Methods	Method (PDF)		_Ph	UTASIC		CASIN	
LC/MS/MS	Environmental Chemistry Methods	Method (PDF)			DTXSID9	050107	494793-67-8	Penflufen
LC/MS/MS	Environmental Chemistry Methods	Method (<u>PDF</u>)		St. a	0173109	038107	494795-07-0	Pennulen
				475	DTXSID6	<u>01024859</u>	NOCAS_1024859	N-(2-Acetylphenyl)-5-fluoro-1,3-dime
						69		

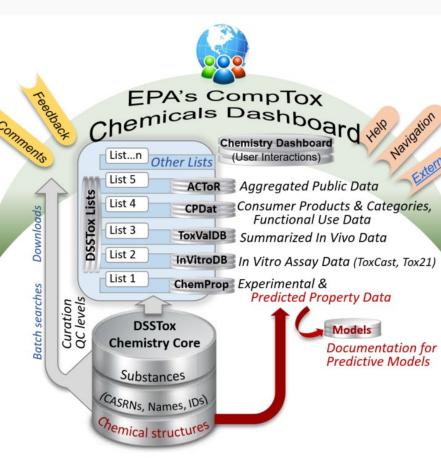
How many methods do we have?



- We have 2785 method documents to extract
- 40% have been extracted ca. 25 documents per day

B1	B1 · : × \checkmark f_x							
	А	В	С	D	E	F	G	
		Refresh	button to get all file					
1	Summary		names and counts from L:					
2					F		1	
3	Text file with chemi 🚽	Date textfile upd	Pdf file 🔹	Date pdf update	pdf size	Chen -	Folder	
4	GJ-014.txt	8/8/2022 12:45	CLG_AVR_1_03.pdf	8/4/2022 14:08	540323	7	_:\Lab\NCCT_Richard\GregoryJa	
5	GJ-005.txt	8/8/2022 12:20	~\$G-MRM3.03.pdf	8/2/2022 14:00	162	108	_:\Lab\NCCT_Richard\GregoryJa	
6	GJ-012.txt	8/8/2022 10:04	CLG-AGON1.pdf	8/4/2022 14:07	592359	5	_:\Lab\NCCT_Richard\GregoryJa	
7	GJ-011.txt	8/8/2022 9:53	CLG-ADD3.03.pdf	8/4/2022 14:05	212151	32	_:\Lab\NCCT_Richard\GregoryJa	
8	GJ-010.txt	8/8/2022 9:42	CLG-PST5.09.pdf	7/18/2022 15:25	2004966	117	.:\Lab\NCCT_Richard\GregoryJa	
9	GJ-009.txt	8/7/2022 16:30	CLG-AMG4.03.pdf	7/18/2022 15:24	514612	9	.:\Lab\NCCT_Richard\GregoryJa	
10	GJ-008.txt	8/7/2022 16:01	CLG-AMG2.08.pdf	7/18/2022 15:23	127547 <mark>8</mark>	9	_:\Lab\NCCT_Richard\GregoryJa	
11	GJ-007.txt	8/7/2022 15:24	CLG-MRM1.pdf	7/18/2022 15:22	4557209	99	_:\Lab\NCCT_Richard\GregoryJa	
12	GJ-024.txt	8/5/2022 15:26	CLG-PENG1.03.pdf	8/4/2022 13:13	433061	. 1	.:\Lab\NCCT_Richard\GregoryJa	
13	GJ-023.txt	8/5/2022 15:22	CLG-NFUR3.pdf	8/4/2022 15:12	224927	8	:\Lab\NCCT_Richard\GregoryJa	
14	GJ-022.txt	8/5/2022 13:44	CLG-MRM2.pdf	8/4/2022 14:56	73567	0	:\Lab\NCCT_Richard\GregoryJa	
4 -	extract Summary		CLC MCCV2 pdf	0/1/2022 11.21	- 	Л	·\Lab\NCCT_Dichard\Gragoryla	

Summary and Conclusion



CompTox Chemicals Dashboard - a central hub for environmental data

- 1.2M chemical substances integrating property data, hazard data, exposure data, *in vitro* bioactivity data
- Batch search for thousands of chemicals
- Real-time property and toxicity predictions

Proof-of-concept cheminformatics modules

- Informatics support for Non-Targeted Analysis
 - Open Spectral data
 - Analytical Methods database

Some Related Publications of Interest





Computational Toxicology Volume 12, November 2019, 100096



EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research

Christopher M. Grulke ^a, Antony J. Williams ^a, Inthirany Thillanadarajah ^b, Ann M. Richard ^a A 🖾

Show more 🧹

+ Add to Mendeley 😪 Share 🍠 Cite



RETURN TO ISSUE < PREV APPLICATION NOTE NEXT >

Enabling High-Throughput Searches for Multiple Chemical Data Using the U.S.-EPA CompTox Chemicals Dashboard

Charles N. Lowe* and Antony J. Williams*

 ♥ Cite this: J. Chem. Inf. Model. 2021, 61, 2, 565–570

 Publication Date: January 22, 2021 ~

 https://doi.org/10.1021/acs.jcim.0c01273

 This article not subject to U.S. Copyright. Published

 2021 by American Chemical Society

 RIGHTS & PERMISSIONS ✓ Subscribed

Article Views Altmetric Citations 802 2 7



Journal of Cheminformatics

Home About <u>Articles</u> Submission Guidelines About The Editors Calls For Papers

Database Open Access Published: 28 November 2017

The CompTox Chemistry Dashboard: a community data resource for environmental chemistry

Antony J. Williams ^{CI}, <u>Christopher M. Grulke</u>, Jeff Edwards, <u>Andrew D. McEachran</u>, <u>Kamel Mansouri</u>, <u>Nancy C. Baker</u>, <u>Grace Patlewicz</u>, <u>Imran Shah</u>, John F. Wambaugh, <u>Richard S. Judson</u> & <u>Ann M. Richard</u>

Journal of Cheminformatics 9, Article number: 61 (2017) Cite this article



Environment International Volume 154, September 2021, 106566



Review article

Sourcing data on chemical properties and hazard data from the US-EPA CompTox Chemicals Dashboard: A practical guide for human risk assessment

Antony J. Williams ^a A 🖾, Jason C. Lambert ^a, Kris Thayer ^b, Jean-Lou C.M. Dorne ^c

You want to know more...



- Lots of resources available
 - Presentations: <u>https://tinyurl.com/w5hqs55</u>

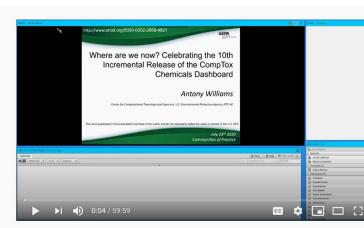
🔼 YouTube

Communities of Practice Videos: <u>https://rb.gy/qsbno1</u>

Search

- Manual: <u>https://rb.gy/4fgydc</u>
- Latest News: https://comptox.epa.gov/dashboard/news_info

Antony William	าร		help? $ imes$
📄 incl. content from 🎲 fig share	sort Relevance 🔻 type Prese	entation v licence ANY v	+ Follow this search
Antony Williams, the ChemConnector: A career path through a diverse series of roles and responsibilities Automatic and the series of the serie	The needs for chemistry standards, database tools and data curation at the chemical-biology interface Annu titing' technology and test and even over	ANNOUNCEMENT EDSP21 and ToxCast Dashboards To Be Discontinued	And the Ansatz of the Ansatz o
Antony Williams, the ChemConnector: A career path thr Antony Williams 09/05/2019	The needs for chemistry standards, database tools and data curation a Antony Williams ~ 30/06/2017	EDSP21 and ToxCast Dashboards To Be Discontinued Antony Williams 30/07/2019	Non-Targeted Screening of Wastewater for Water Reuse using Jerry Zweigenbaum ~ 12/09/2019
Investigating Impact Matrices for Performance for the US EPA National Center for Computational Tackbology Assay titlews* f. Aveau Earchest, extra citila- tic and hones and Agent Schemer Matrices and Agent Schemer Schemer Beneroorden Agent Schemer Bener	GDH Consensus ranking and fragmentation prediction for identification of unknowns in high resolution mass spectrometry and the factorian fragment and the factor of the many Applement for the factor of the factor of the Autor 102	Building an Online Profile Using Social Networking Tools Record Sector S	The CompTox Chemicals Dashboard as An insignation Hub for Chemistry, Bology and Environmental Toxicity Data Annu Killen, Chemistel, and Annue Annue Annue Annue Killen, Chemistel, and Annue Ann
Investigating Impact Metrics for Performance for the US EPA Natio 0000-0002-2668-4821 ~ 30/06/2017	Consensus ranking and fragmentation prediction for identif Andrew McEachran ~ 21/08/2018	Building an Online Profile Using Social Networking Tools Antony Williams ~ 30/05/2018	The CompTox Chemicals Dashboard as An Integration Hub for Chemistr Antony Williams v 09/10/2019
Array Constrained Actives Settled, research, sequences that any probability approximate any probabil	And the second s	The EPA Comp Tox Chemistry Destroyers - a contralized hub for integrating data to the other interneting data to the other interneting data to the other way Witten Access the other other data the other access to the other other access to the other access to the other other access to the other other access to the other access to the other other access to the other a	Environmental Chemistry Compound Identification Using High Researchers Mass Epochemistry Data Integrated to the EPA Chemistry Durabed Antroy, Lifetime, Antro Antrophysical Chemistry Manner, Christian and Status, and Status Manner, Christian and Status
Generalised Read-Across GenRA, research, implementation and prac Grace Patiewicz ~ 18/09/2018	OPERA: A QSAR tool for physicochemical properties and e Kamel Mansouri ~ 20/06/2018	The EPA CompTox Chemistry Dashboard - a Centralized Hub for Antony Williams 05/07/2018	Environmental Chemistry Compound Identification Using High Resolutio



Where are we now? Celebrating the 10th Incremental Release of the CompTox Chemicals Dashboard

CompTox Chemicals Dashboard primer videos

The CompTox Chemicals Dashboard is a one-stop-shop for chemistry, toxicity and exposure information for over 875,000 chemicals. Data and models within the Dashboard also help with efforts to identify chemicals of most need of further testing and reducing the use of animals in chemical testing.

Explore the wealth of data and features available in the CompTox Chemicals Dashboard with these insructional videos narrated by EPA scientists.

General Chemistry and Search Capabilities

bare - Manual Lacondanak		
		forward adupt is the best auguster these sells as they end
	mer Watch la	BAUBLE BURELAN 2: Could pay supplies a bit more how you come up with the () my compared Sign 2 ()
		SB: Do you intend to crosslink the human and animal pathways to identify areas that are in common, or ext?
Difference Difference <thdifference< th=""> Difference Differen</thdifference<>		Onia V: have you compared your pathways with the string database?
Constructions State Productions School State Sta		David Cerhold: 11 repeat Q
Emergina and another second se		David Serbald Thania!
		MAURCE WHELAN 2: If one were to start using transcriptomics assays instead of HTS assays, how much could you reduce the technical complexity?
		MAURICE WHELAN 2: reduce the number of assays
EPHERICIA-WERE ADDRESS OF THE TAXES OF TA		Jenky Ice: Thanks, Huang
transmission contractor terms of the second		MAURICE WHELAN 2: thanks!
4 > 0 # 6 6 8 8 8 0 8 8 8 0 8	4 (* 105.00 B)	Attendees (N)
ST_CC_Pod_Enter v.2.1.5-5.20		C. 10141412214
origit 0 far	•• O male M Auto-Scould #*	· Hods (3)
White (0) + 52 pt + Talenna +	extent + 10 Q fired	🙈 Jerelle K Captioner
· · · ·		A Marrice Linnerthriek
PUT IT IN THE DOWNLOADS FILE.	÷	A NCC
IS SOME TO BRING WITH IT THE INDIVIDUAL OFFICIAL LD.		
SUBSTANCE IDENTIFIER, IT'S GOING TO BRING WITH IT THE FORMULA, THE MASS, THE SHILE STRING, HIJLTIPLE OTHER PARTS OF	 Participants (90) 	
NONTINE OFFICES	and a second second second	Adam .
BRAILY KINERE VOU SEE A TABLE INSIDE THE APPLICATION. YOU		Adam Swank
Record Control of the second burger of the second control of the	5	Advance Bejarame

Acknowledgments



- Contact: Williams.Antony@epa.gov
- Feedback and follow-up is welcomed! Your questions help
- If anyone wants a deep dive into any of the software applications feel free to contact me



EPA's Center for Computational Toxicology and Exposure