

Introduction to the US EPA CompTox Chemicals Dashboard

Antony John Williams

williams.antony@epa.gov

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

November 9th and 10th 2020

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

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

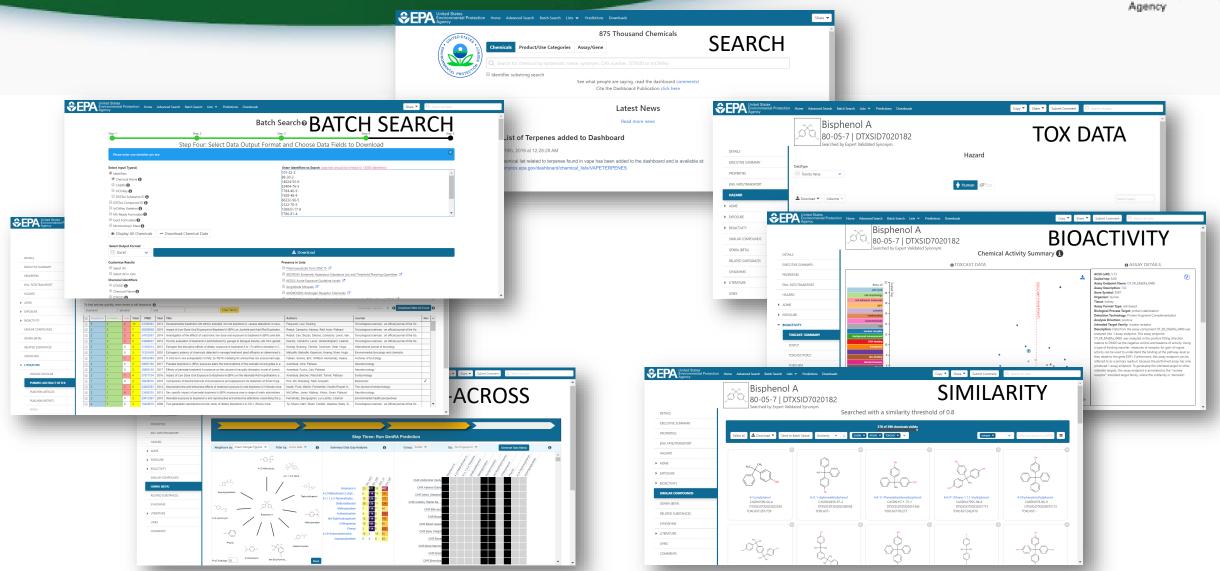


A publicly accessible website delivering access:

- ~882,000 chemicals with related property data
- Experimental and predicted physicochemical property data
- Experimental Human and Ecological hazard data
- Integration to "biological assay data" for 1000s of chemicals
- Information regarding consumer products containing chemicals
- Links to other agency websites and public data resources
- "Literature" searches for chemicals using public resources
- "Batch searching" for thousands of chemicals

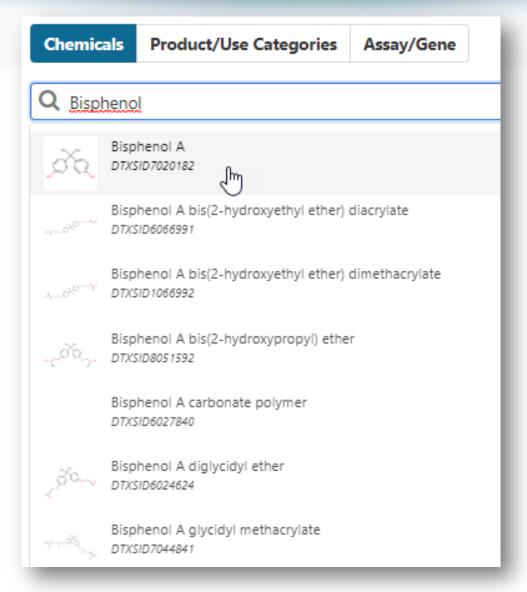
CompTox Chemicals Dashboard https://comptox.epa.gov/dashboard





BASIC Search

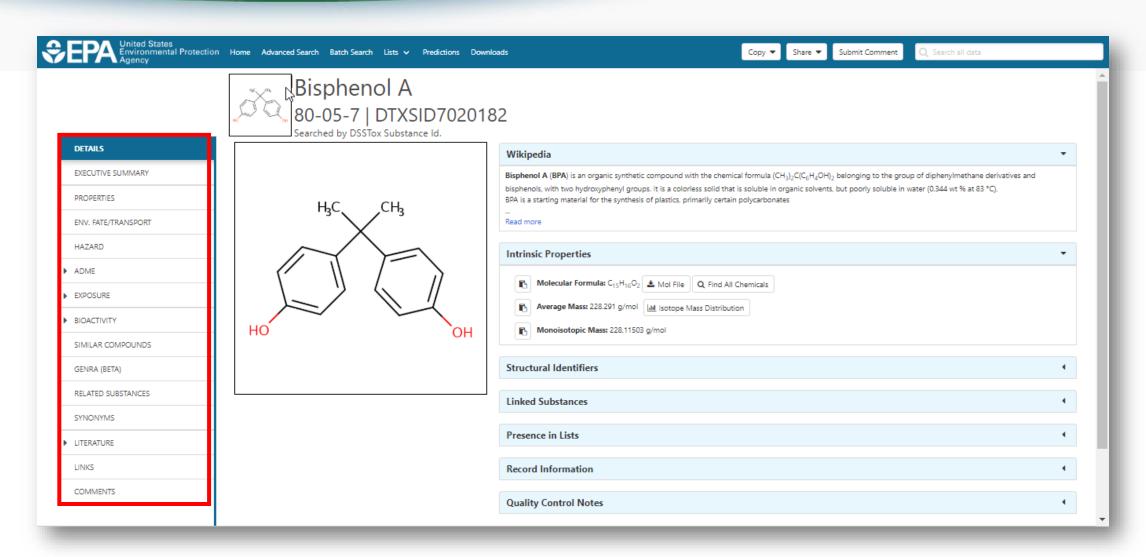




- Three searches from the home page
 - Chemicals (Names, CASRN, Substring)
 - Product/Use Categories
 - Assay Gene
- Searching millions of synonyms and includes Active, Deleted and Alternate CASRNs

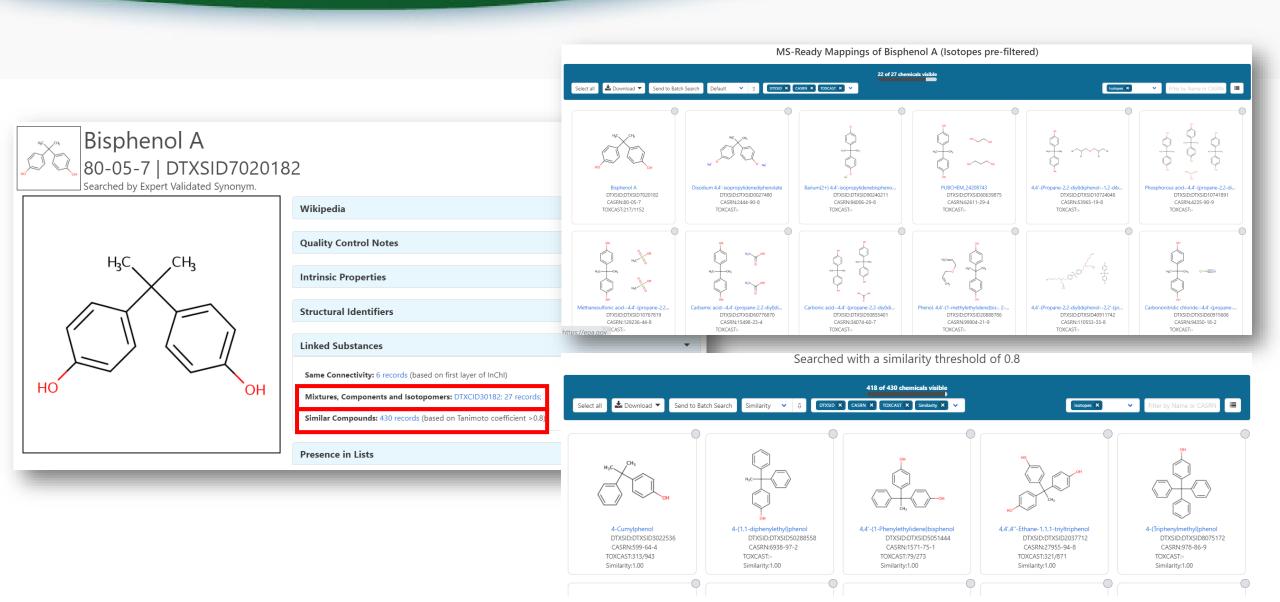
Detailed Chemical Pages





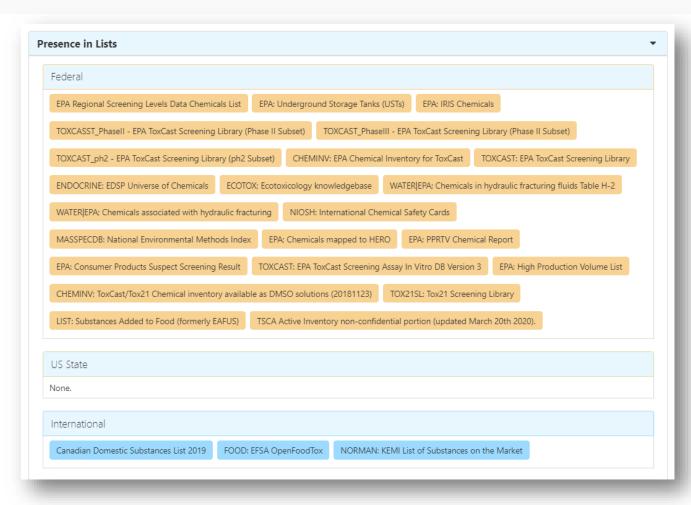
Linked Substances: Mixtures and Components, Similar Compounds





Presence in Lists



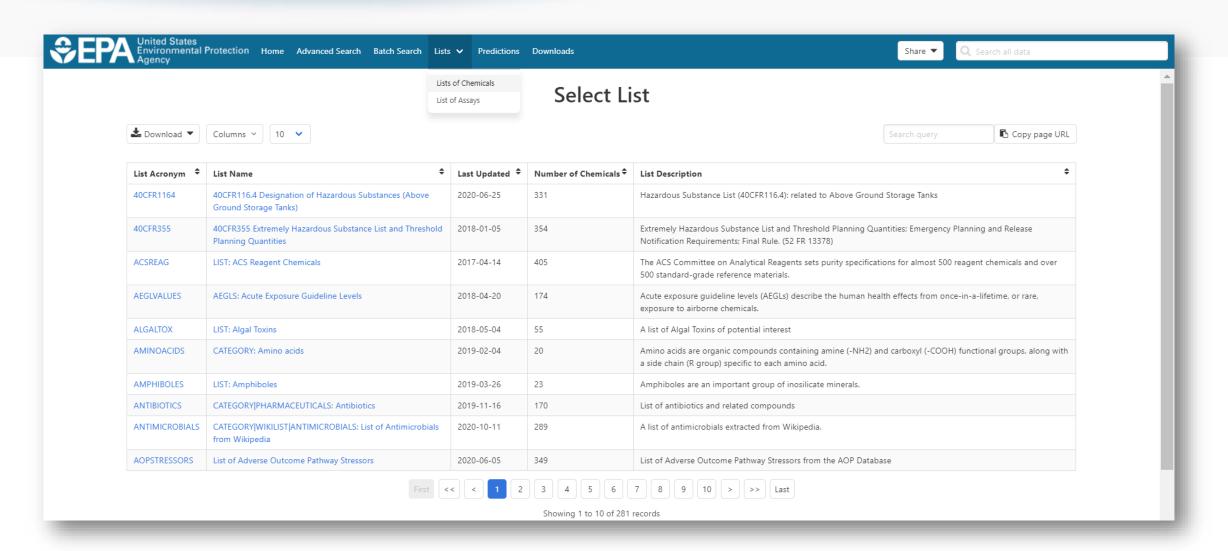


 ~882,000 chemicals are associated with > 280 lists

- Lists are segregated into
 - Federal
 - US State
 - International
 - Other
- Chemicals can be present in multiple lists – with clickthrough

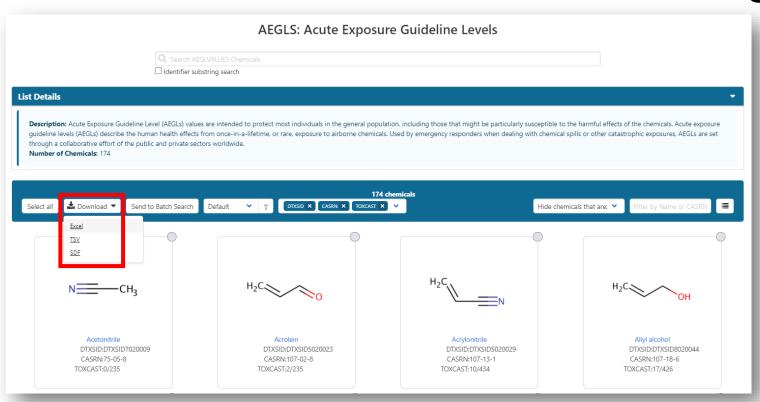
Lists of Lists of Chemicals >280 lists and growing





List of Chemicals Example: Acute Exposure Guideline Levels





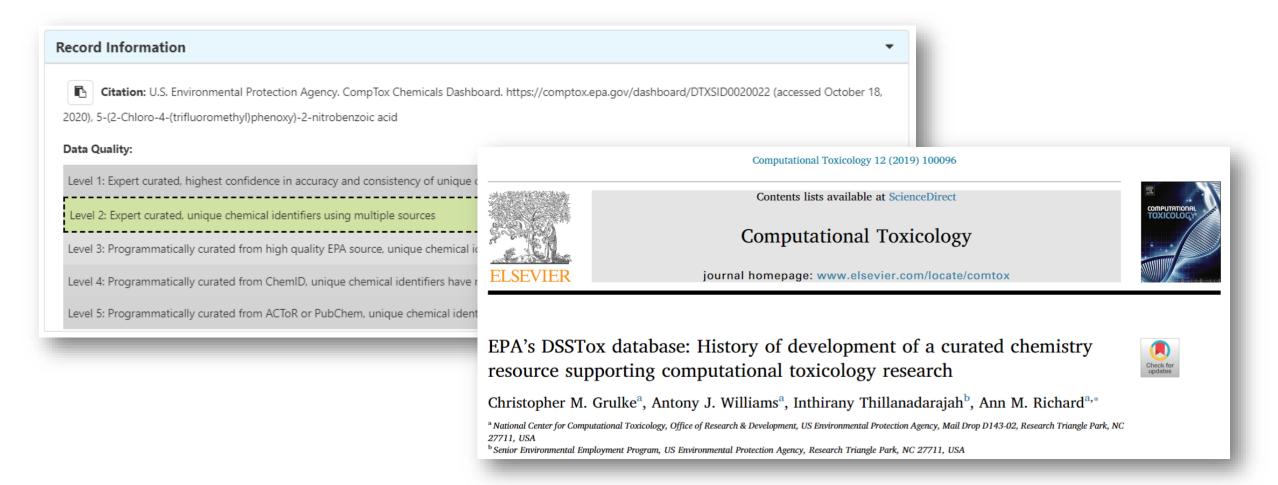
Chemical List Examples

- Disinfection by-products
- EPA Pesticide Search DB
- Consumer Products DB
- Chemicals in Biosolids
- ATSDR Toxicological Profiles
- Canadian Domestic Substances
- Chemical Inventory for ToxCast
- EFSA OpenFoodTox
- EDSP Universe of Chemicals
- Hydraulic Fracturing Chemicals

Quality

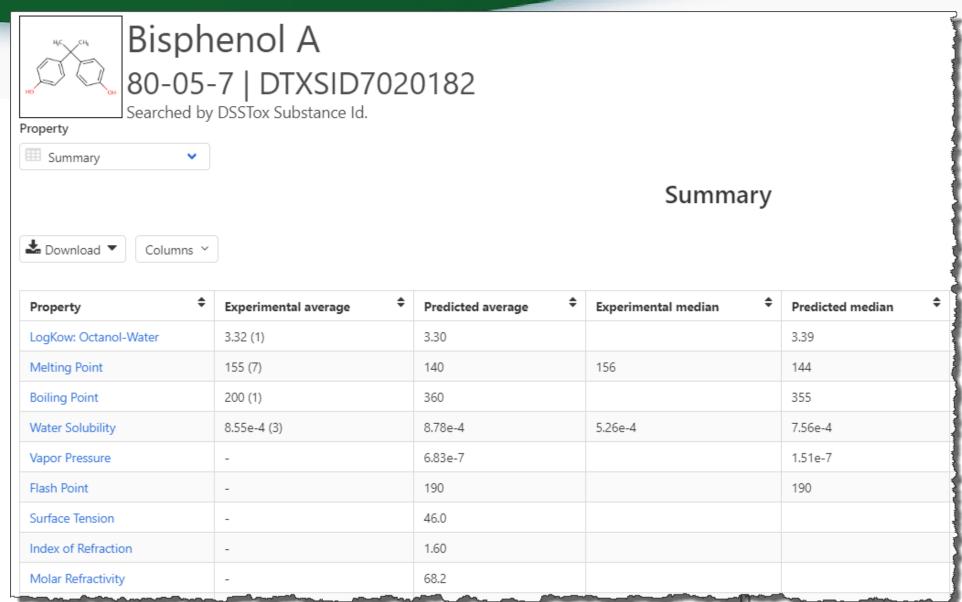


Five full time curators register and curate data to elevate quality



Properties, Fate and Transport





Properties, Fate and Transport e.g. Solubility



≛ Download Experimental Data ▼		
Source	\$	Result \$
PhysPropNCCT		5.26e-4
Tetko et al. J. Chem. Inf. and Comp. Sci. 41.6 (2001): 1488-1493		1.51e-3
Kovdienko, et. al. Molecular informatics 29.5 (2010): 394-406.		5.25e-4

Source	Result	Calculation Details
EPISUITE	7.56e-4	Not Available
NICEATM	1.31e-3	Not Available
TEST	1.24e-3	TEST Report
OPERA	5.44e-4	OPERA Model Report [Inside AD]
OPERA2	5.35e-4	Not Available

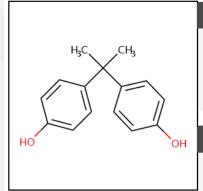
OPERA Models





Properties, Fate and Transport e.g. logP





Model Results

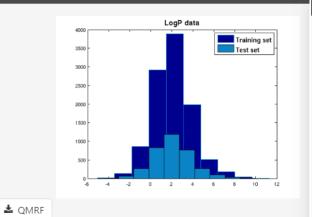
Predicted value: 3.35

Global applicability domain: Inside

Local applicability domain index: 0.877

Confidence level: 0.748

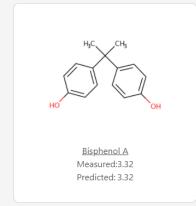
Model Performance

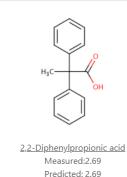


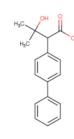
5-fold CV (75%)		
Q2	RMSE	
0.850	0.690	

od.epa.gov/dashboard/advanced search/index

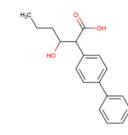
Nearest Neighbors from the Training Set



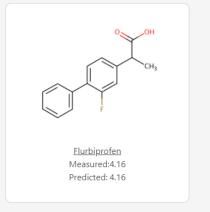




BUTANOIC ACID,2-(4-BIPHENYLYL)-3-HYDROXY-3-METHY Measured:3.25 Predicted: 3.25



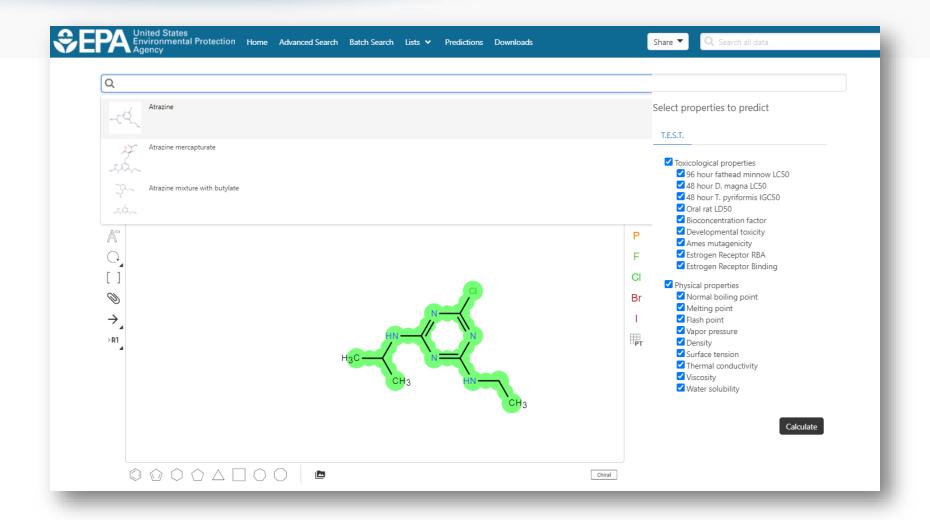
3-OH-2-(4-BIPHENYLYL)HEXANOIC ACID Measured: 3.75 Predicted: 3.75



Predictions for New Chemicals

https://comptox.epa.gov/dashboard/predictions/index





Predictions for New Chemicals

https://comptox.epa.gov/dashboard/predictions/index



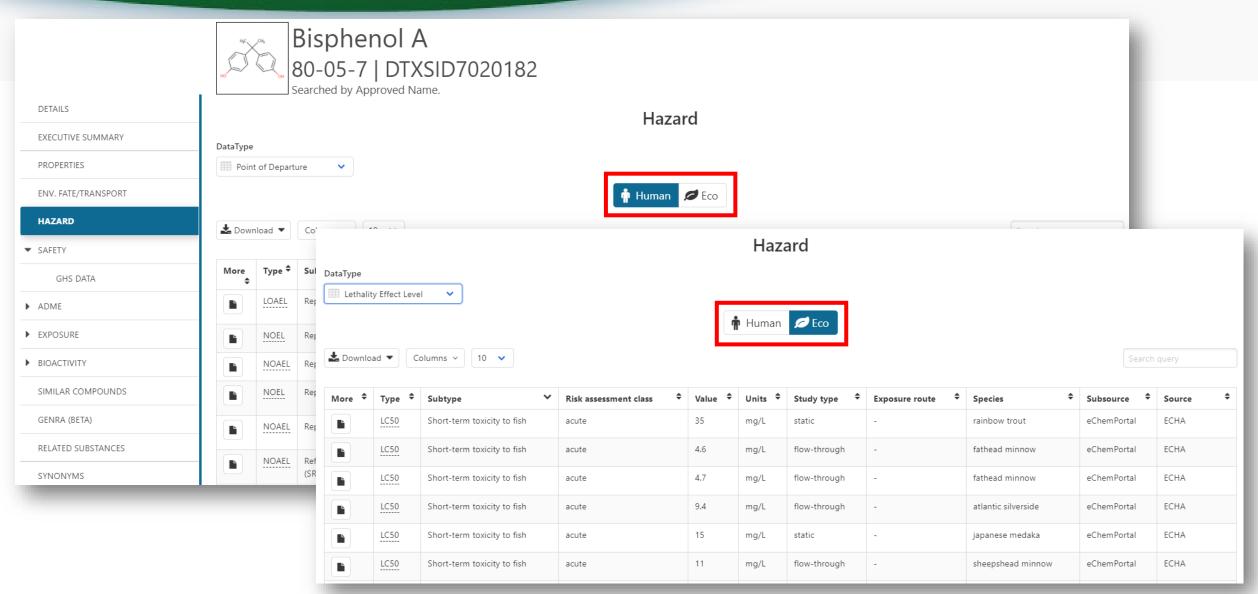
Provid	ler: `	T.E.:	S.T.

🚣 D	ownload	Summary	•
-----	---------	---------	---

Property	Experimental Value	Consensus	Hierarchical clustering	Single model	Group contribution	Nearest neighbor
96 hour fathead minnow LC50		3.969 -Log10(mol/L) 22.469 mg/L	3.936 -Log10(mol/L) 24.234 mg/L	3.742 -Log10(mol/L) 37.882 mg/L	4.135 -Log10(mol/L) 15.355 mg/L	4.064 -Log10(mol/L) 18.079 mg/L
48 hour D. magna LC50		3.791 -Log10(mol/L) 33.887 mg/L	3.568 -Log10(mol/L) 56.592 mg/L	4.243 -Log10(mol/L) 11.976 mg/L	3.832 -Log10(mol/L) 30.824 mg/L	3.521 -Log10(mol/L) 63.123 mg/L
48 hour T. pyriformis IGC50			2.986 -Log10(mol/L) 216.359 mg/L			
Oral rat LD50		2.439 -Log10(mol/kg) 761.182 mg/kg	2.355 -Log10(mol/kg) 924.254 mg/kg			2.524 -Log10(mol/kg) 626.882 mg/kg
Bioconcentration factor		0.847 Log10 7.038	0.906 Log10 8.051	0.690 Log10 4.903	0.820 Log10 6.607	0.974 Log10 9.409
Developmental toxicity		true	true	true		
Ames mutagenicity		false	true			false
Estrogen Receptor RBA						
Estrogen Receptor Binding		false	false	false	false	
Normal boiling point		315.9 °C	318.0 °C		345.4 °C	284.3 °C
Melting point		114.0 °C	105.0 °C		79.4 °C	157.7 °C
Flash point		176.1 °C	180.1 °C		172.5 °C	175.7 °C
Vapor pressure		-6.072 Log10(mmHg) 8.464*10^-7 mmHg	-6.164 Log10(mmHg) 6.851*10^-7 mmHg		-5.679 Log10(mmHg) 2.093*10^-6 mmHg	-6.374 Log10(mmHg) 4.23*10^-7 mmHg

Hazard Data (Human and Eco)

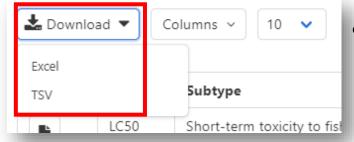




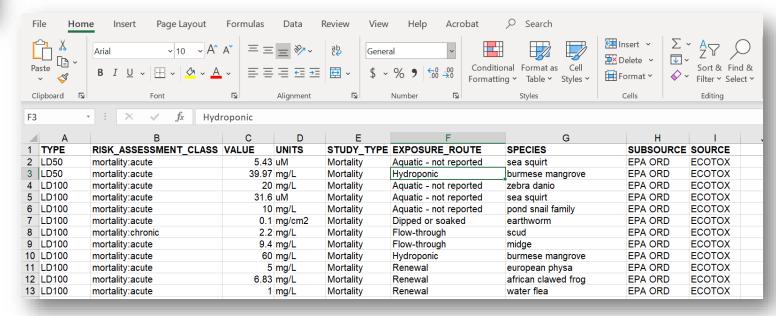
Hazard Data



 Data are harvested from > 30 data sources and provides both human and ecological data for >50,000 chemicals



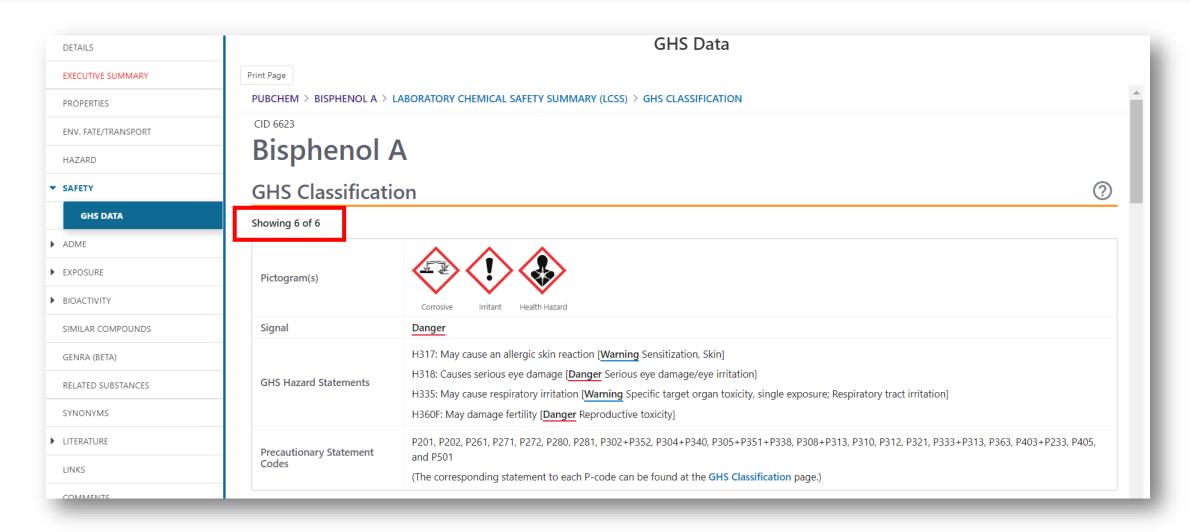
 General Feature – anywhere you see a table you can download the data…



Safety Data

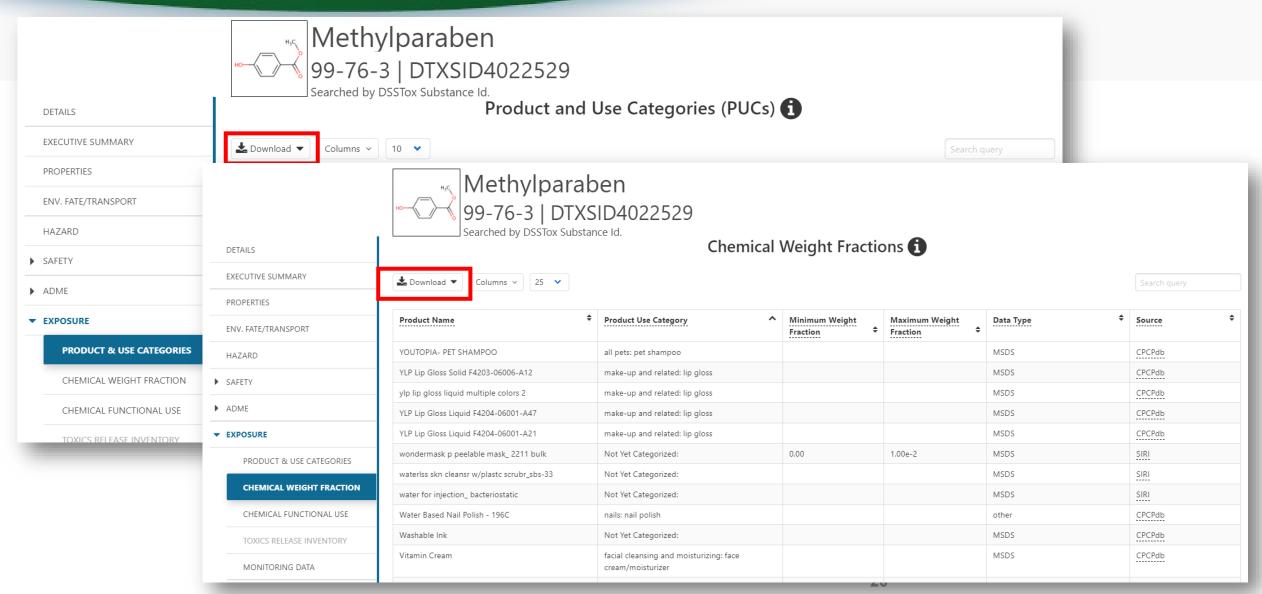


Safety Data (Global Harmonization System (GHS)) from PubChem



Sources of Exposure to Chemicals (More later – Kristin Isaacs)

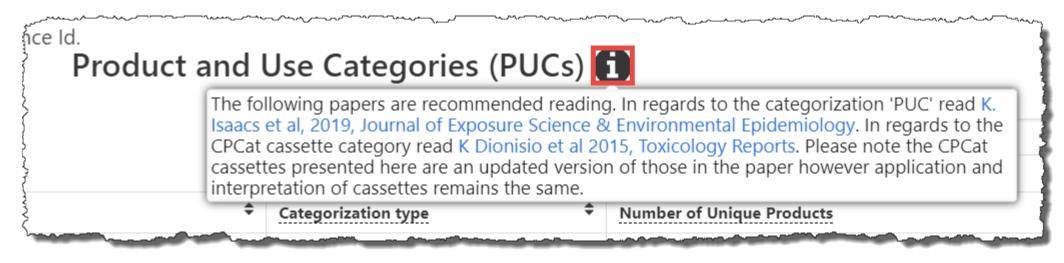




General Feature On-hover Informational Icons



Watch out for the on-hover informational icons to gather info



Descriptive text or video about the data, associated publications etc.

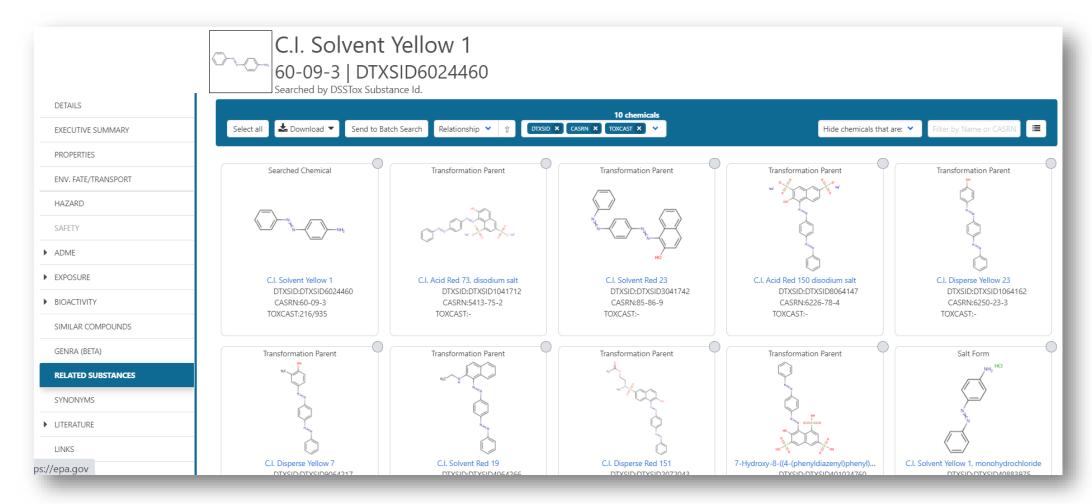
Advanced Search capabilities presently support mass spectrometry at this point with mass and formula searching. For an overview of capabilities we recommend you watch this short video.



Related Substances e.g. Aromatic amine to Azo Dyes

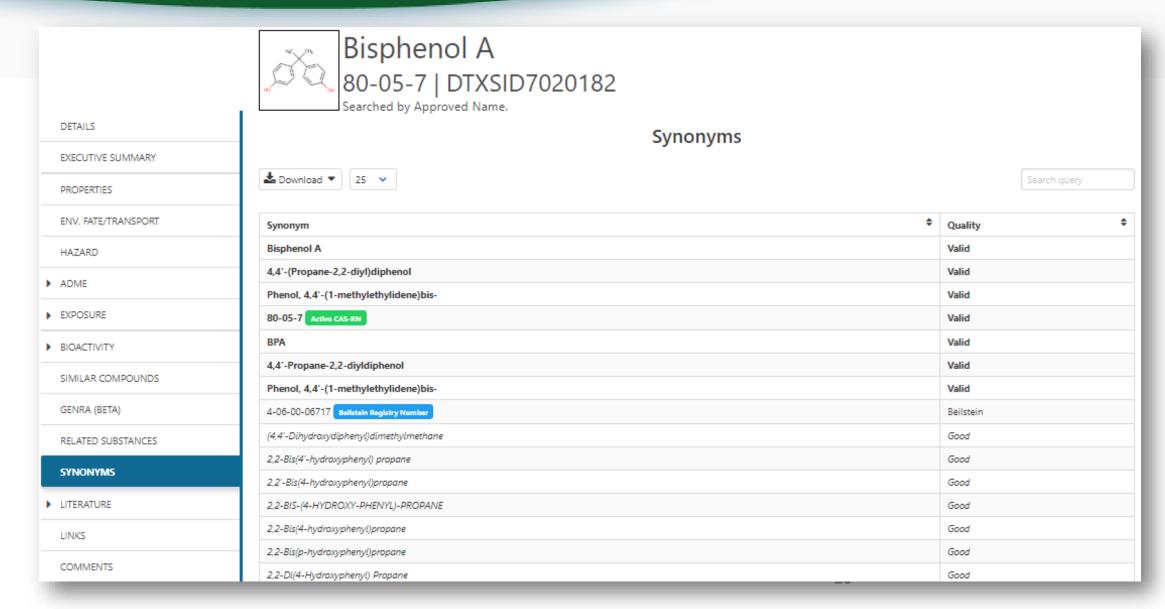


 Related substances – Parent to Transformation product: metabolites, degradants; Monomer to polymer;



Identifiers to Support Searches

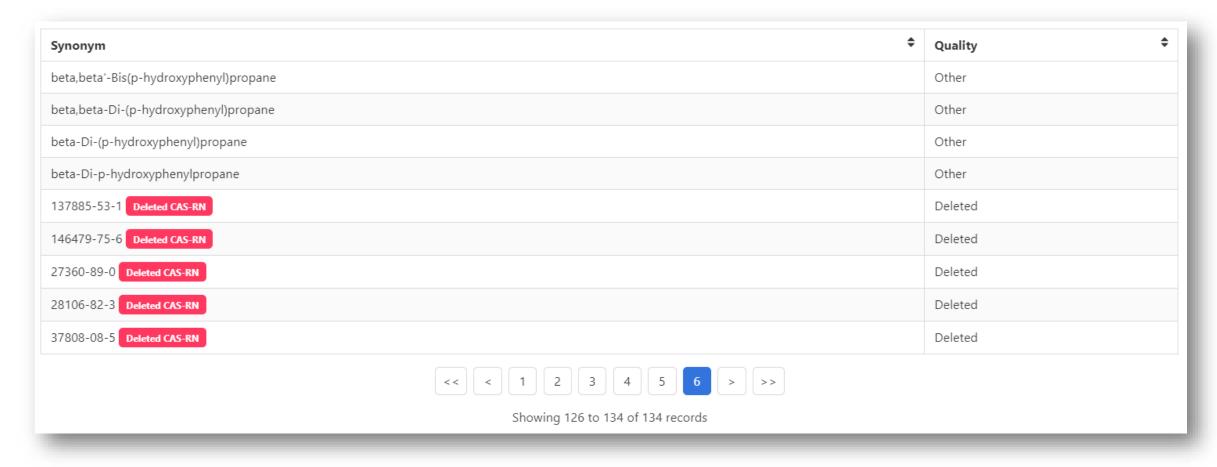




Identifiers to Support Searches



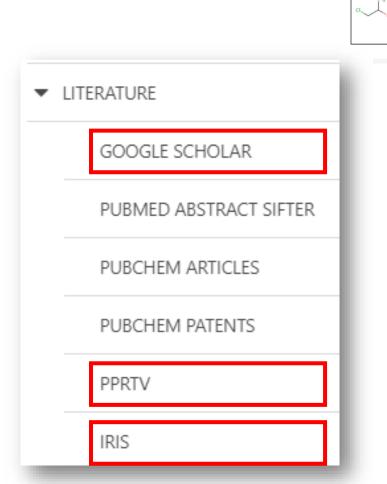
 Identifiers include systematic names, trade names, CASRNs (Active, Alternate and Deleted)



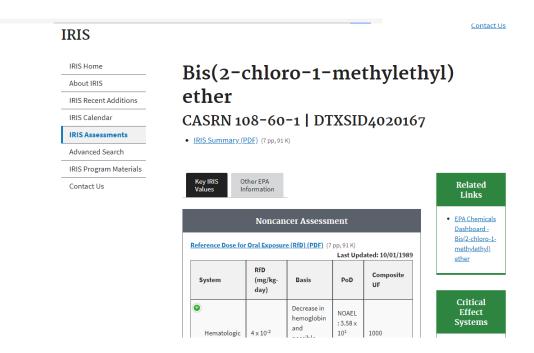
Identifiers are used in the app



Identifiers are used to feed and link into "Literature"



bis(2-Chloro-1-methylethyl) ether 108-60-1 | DTXSID4020167



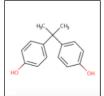
External Links – Also use Identifiers Names, CASRN, PubChem IDs...



	Bisphenol 80-05-7 DT Searched by Approved	XSID7020182			
DETAILS	General	Toxicology	Publications	Analytical	Prediction
EXECUTIVE SUMMARY	● EPA Substance Registry Service		Toxline	☐ FOR-IDENT	3 2D NMR HSQC/HMBC Predict
PROPERTIES	Household Products Database Chemical Entities of Biological Interest	OH _E DrugPortal CCRIS	Environmental Health Perspectives NIEHS	NEMI: National Environmental Methods Index RSC Analytical Abstracts	Carbon-13 NMR Prediction Proton NMR Prediction
ENV. FATE/TRANSPORT	(ChEBI)	ChemView	National Toxicology Program	△ Tox21 Analytical Data	ChemRTP Predictor
HAZARD	Chemspider	G CTD	G Google Books G Google Scholar	MONA: MassBank North America mzCloud	■ LSERD
ADME	● CPCat	Gene-Tox	G Google Patents	NST NIST IR Spectrum	
EXPOSURE	Amp HMDB	HSDB ToxCast Dashboard 2	PPRTVWEB PubMed	MST NIST MS Spectrum	
BIOACTIVITY	W Wikipedia Q MSDS Lookup	LactMed International Toxicity Estimates for Risk	■ IRIS Assessments ■ EPA HERO		
SIMILAR COMPOUNDS	© ChEMBL	ATSDR Toxic Substances Portal	NIOSH Skin Notation Profiles		
GENRA (BETA)	Chemical Vendors CalEPA Office of Environmental Health	Superfund Chemical Data matrix NIOSH IDLH Values	NIOSH Pocket Guide RSC Publications		
RELATED SUBSTANCES	Hazard Assessment NIOSH Chemical Safety Cards	ACTOR PDF Report	■ RioCaddie DataMed		
SYNONYMS	ToxPlanet	Toxics Release Inventory	Springer Materials		
LITERATURE	ACS Reagent Chemicals W Wikidata	CREST National Air Toxics Assessment	Federal Register Regulations.gov		
LINKS	ChemHat: Hazards and Alternatives Toolbox		Bielefeld Academic Search Engine		
COMMENTS	₩ Wolfram Alpha ScrubChem		CORE Literature Search		

External Links





Bisphenol A 80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

General

EPA Substance Registry Service



Chemspider





W Wikipedia

Q MSDS Lookup

ChEMBL

ToxPlanet

ACS Reagent Chemicals

ChemHat: Hazards and Alternatives Toolbox

Toxicology



он₂ DrugPortal



ChemView





eChemPortal





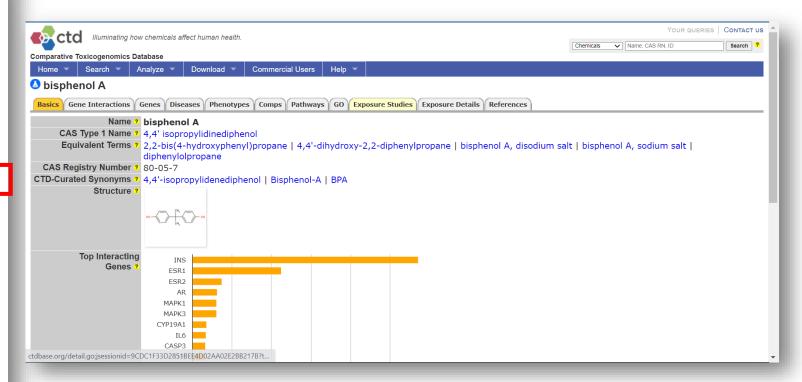


● ACToR PDF Report

CREST

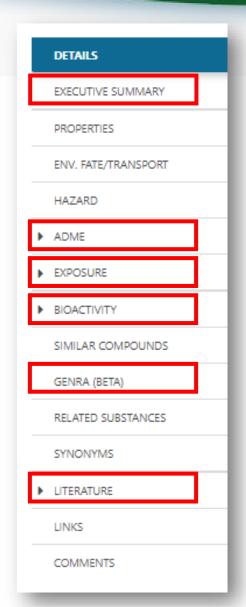
National Air Toxics
Assessment

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



More details will be covered on...



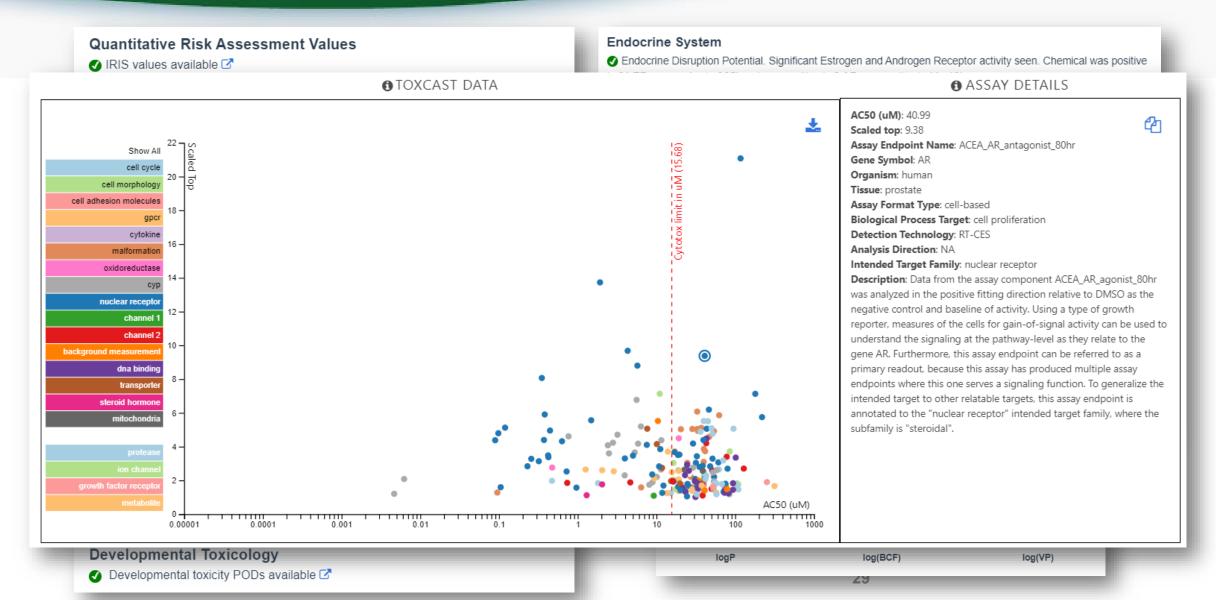


- ADME:IVIVE will be covered by John Wambaugh
- Exposure will be covered by Kristin Isaacs
- Bioactivity will be covered by Katie Paul-Friedman
- GenRA will be covered by Grace Patlewicz
- Literature: Abstract Sifter will be covered by Nancy Baker

 ...and much of this data will show up in the Executive Summary for a chemical substance

Executive Summary Summary view of relevant data



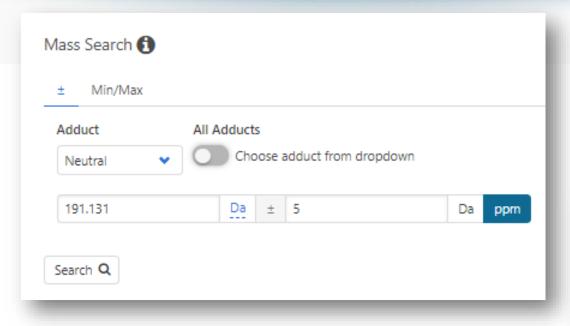


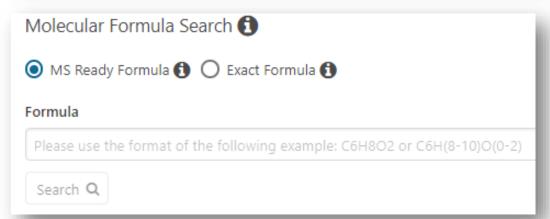


Advanced Searching: Mass and Formula

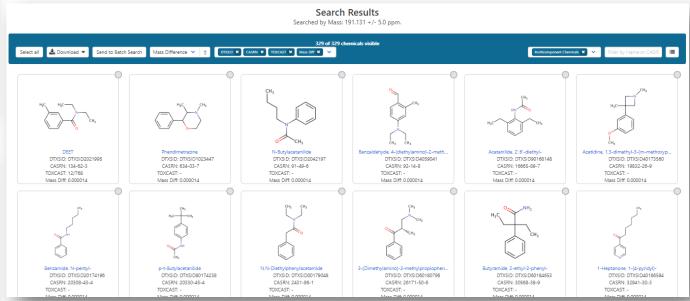
Advanced Searches Mass and Formula Searches







Supports our suspect screening and non-targeted analysis mass spectrometry research efforts

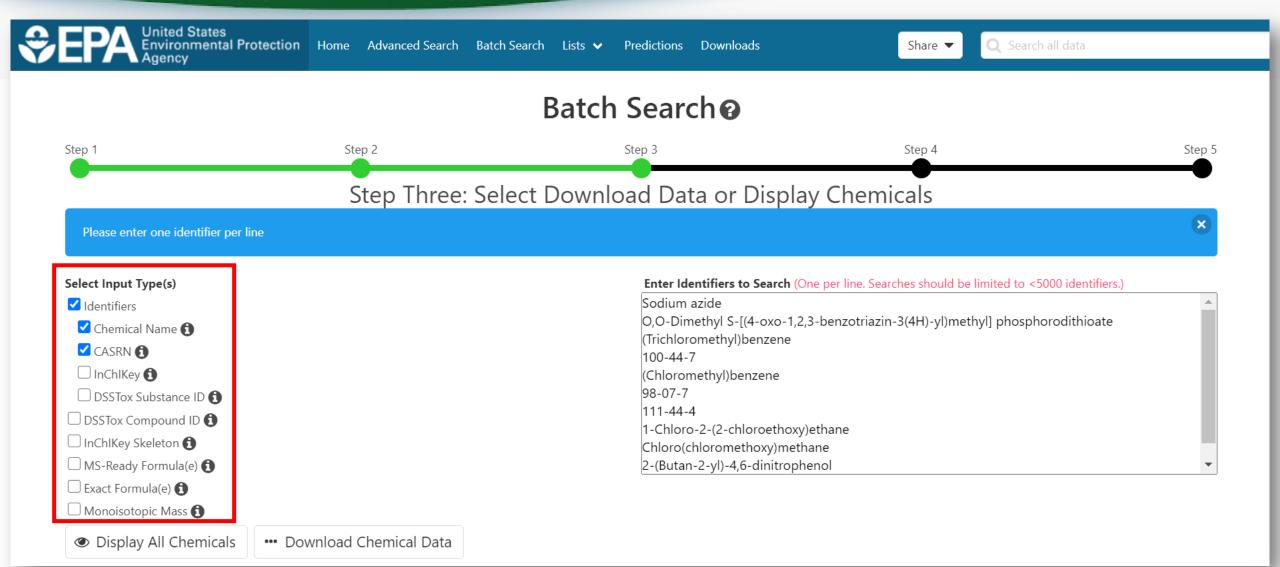




Batch Searching

Access data *en masse* for thousands of chemicals....





Select Output Format and Content



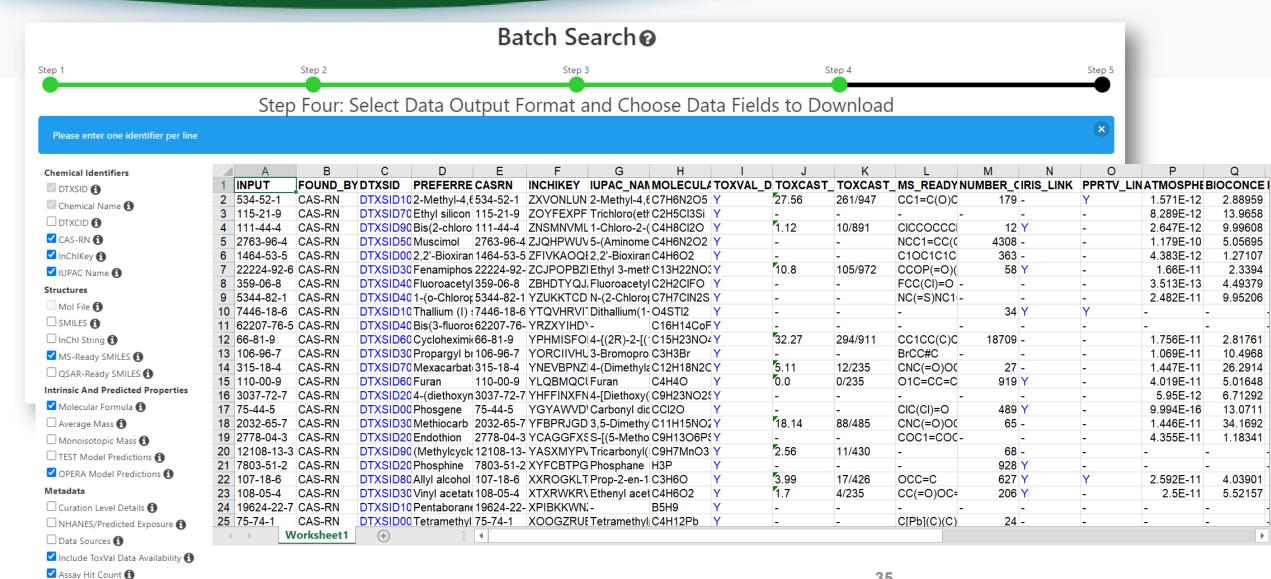
Step Four: Select Data Output Format and Choose Data Fields to Download

Excel 🗸		
Customize Results	Intrinsic And Predicted Properties	Presence in Lists:
Select All	☐ Molecular Formula 1	☐ 40CFR116.4 Designation of Hazardous Substances (Above Ground Storage Tanks) 🗹
Select All in Lists	Average Mass 1	☐ 40CFR355 Extremely Hazardous Substance List and Threshold Planning Quantities 🗹
hemical Identifiers	☐ Monoisotopic Mass 1	☐ AEGLS: Acute Exposure Guideline Levels 🗹
DTXSID 🚯	☐ TEST Model Predictions 1	ANDROGEN: Androgen Receptor Chemicals 🖸
Chemical Name 🐧	OPERA Model Predictions 1	ARTICLE; Bench-Mark Dose Human Health Assessment List (Wignall et al., 2014)
DTXCID 🚯	Metadata	☐ ARTICLE: Collaborative Estrogen Receptor Activity Prediction Project (CERAPP) ☑
CAS-RN 1	Curation Level Details (1)	☐ ARTICLE: Collaborative Estrogen Receptor Activity Prediction Project (COMPARA) ☐
InChlKey 1	NHANES/Predicted Exposure (1)	☐ ATSDR Toxicological Profiles ☑
IUPAC Name	☐ Data Sources 1	☐ ATSDR: Minimal Risk Levels (MRLs) for Hazardous Substances ☑
ructures	☐ Include ToxVal Data Availability 1	☐ ATSDR: Toxic Substances Portal Chemical List 🇹
Mol File 1	Assay Hit Count 1	☐ California Office of Environmental Health Hazard Assessment ☐ California Office of Environmental Health Hazard Assessment ☐ California Office of Environmental Health Hazard Assessment
SMILES (1)	Number of PubMed Articles 1	
InChl String 1	PubChem Data Sources 1	Canadian Domestic Substances List 2019
MS-Ready SMILES 1	CPDat Product Occurrence Count 1	CATEGORY: Amino acids
QSAR-Ready SMILES (1)	☐ IRIS 1	CATEGORY: Color Index dyes
J QJAIN-Neady SIVILLS (PPRTV (1)	☐ CATEGORY: Flame Retardants 🗹
	☐ Wikipedia Article	
	QC Notes 1	
	\square Include links to ACToR reports - SLOW! (BETA) $oldsymbol{1}$	

Batch Search CASRNs

✓ Number of PubMed Articles ♠







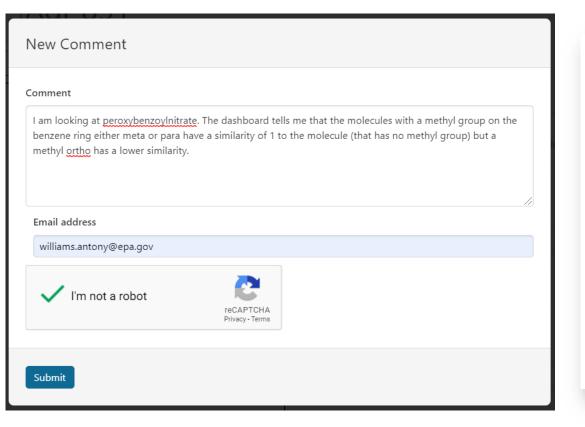
How Can You Help?

How can you help?



Please submit comments for curation and feedback. We will respond.





	Crowdsourced Comments							
25 🔻								
Chemical \$	Structure	Date 🗸	Comment	Resolution				
Didecyldimethylammonium chloride	-	2020- 10-11	I don't think the "Product Use Category" sorting function (on table) is working. The other columns seem to work.	Resolved				
Carotenoids No Chemical Structure Associated with this Substance No Chemical 2020- 36-88-4 is r			36-88-4 is not a CASRN	Resolved				
2-(Dimethylamino)-1- phenylethan-1-ol hydrogen c	N	2020- 10-07	This entry is a rare instance of disagreement for CASRN between EPA Chemistry Dashboard and ChemIDplus. ChemIDplus associates CASRN 99614-01-4 with Ondansetron hydrochloride: https://chem.nlm.nih.gov/chemidplus/rn/99614-01-4 Could someone check and let me know which is correct? Thanks, AAB	Resolved				
2-Hexenal, 5-methyl-2-(1- methylethyl)-	H,C CN ₀	2020- 09-25	Isodihydroavandulyl aldehyde BAD spelling lacks letter I - also present correctly! IUPAC name 2-isopropyl-5-methylhex-2-enal RTECS is MP6450000 REAXYS/BRN is 1752384 Several other corrections! 2-Hexen-1-al, 5-methyl-2-(1-methylethyl)- is a muddle of IUPAC & CAS names - remove?? Barrie	Resolved				
2-Hexenal, 5-methyl-2-(1- methylethyl)-	H,C CN ₀	2020- 09-25	Isodihydroavandulyl aldehyde BAD spelling lacks letter I - also present correctly! IUPAC name 2-isopropyl-5-methylhex-2-enal RTECS is MP6450000 REAXYS/BRN is 1752384 Several other corrections! 2-Hexen-1-al, 5-methyl-2-(1-methylethyl)- is a muddle of IUPAC & CAS names - remove?? Barrie	Resolved				



Prototypes in Progress

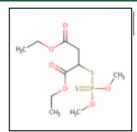
New AnalyticalQC Database



- The analytical QC data on the dashboard is based on a snapshot from ~5 years ago
- The Tox21 browser has continued to be updated with new data including NMR data
- New database is presently in construction hosting ALL QC data

Example of unsynchronized data

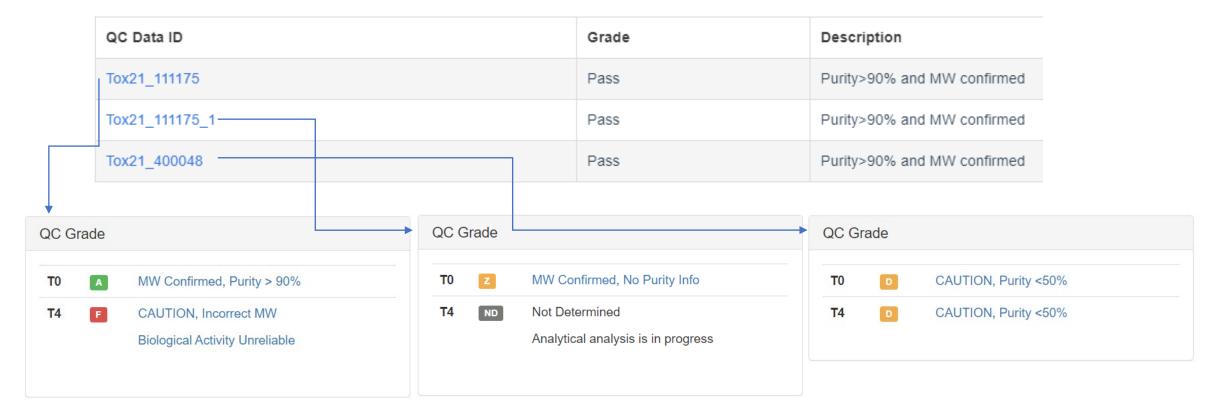




Malathion 121-75-5 | DTXSID4020791

Searched by Approved Name.

ToxCast/Tox21



AnalyticalQC Database

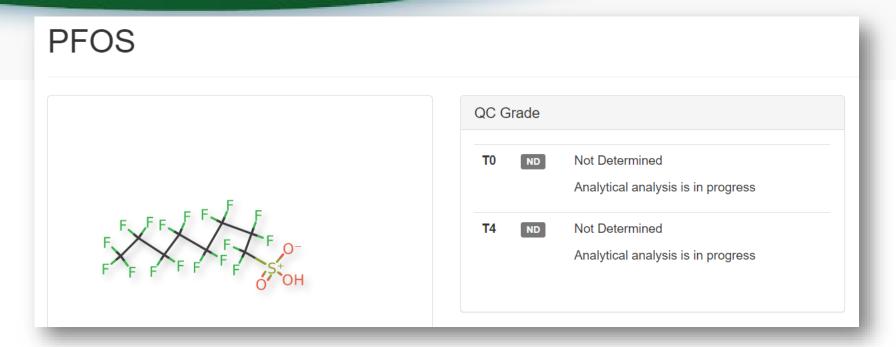


- Single database integrating all analytical data at a substance level
- Hyperlinked to all individual spectral report files

	$\underline{\mathrm{DTXSID4020791}}$													
	121-75-5													
	Malathion													
how 25 v entries Copy Excel CSV									Search:					
FileDate 🛊	ExpDate	SampleId \$	PubChemSid \$	PubChemCid \$	Tox21Id ♦	filename	study	Timepoint \$	batch \$	well	grade ▼	purity \$	PassFail	comment
2015-02-23	2014-07-29	NCGC00091902-08	144214009	4004	400048	Tox21_400048-12_04_LCMS.pdf	OPANS LCMS	T4	Batch5- SP123647	L19	Z	IW Confirmed, No urity Info	Fail	
2015-02-23	2014-07-30	NCGC00091902-08	144214009	4004	400048	Tox21_400048-13_04_LCMS.pdf	OPANS LCMS	T4	Batch5- SP123647	P20		MW Confirmed, No urity Info	Fail	
2019-02-13	2016-11-01	NCGC00091902-05	144204726	4004	<u>111175</u>	Tox21_111175_04_NMR.pdf	OPANS NMR	T4		A7	F	mpure	Fail	CAUTION, Incorrect MW
2014-11-07	2011-09-10	NCGC00091902-08	144214009	4004	400048	Tox21_400048_00_LCMS1.pdf	OPANS LCMS	T0	Batch2- SP109801	H24	D	:50%	Fail	
2015-02-23	2014-07-26	NCGC00091902-08	<u>144214009</u>	4004	400048	Tox21_400048-4_04_LCMS.pdf	OPANS LCMS	T4	Batch5- SP123646	C24	D	50%	Fail	
2015-02-23	2014-07-27	NCGC00091902-08	<u>144214009</u>	4004	400048	Tox21_400048-5_04_LCMS.pdf	OPANS LCMS	T4	Batch5- SP123646	H20	D	:50%	Fail	
2015-02-23	2014-08-02	NCGC00091902-08	144214009	4004	400048	Tox21_400048-3_04_LCMS.pdf	OPANS LCMS	T4	Batch5- SP123649	D24	D	50%	Fail	
2015-02-23	2014-09-04	NCGC00091902-08	144214009	4004	400048	Tox21_400048_04_LCMS.pdf	OPANS LCMS	T4	Batch5- SP123651	K 5	D	:50%	Fail	
2015-02-23	2014-09-29	NCGC00091902-08	<u>144214009</u>	4004	400048	Tox21_400048-1_04_LCMS.pdf	OPANS LCMS	T4	Batch5- SP123653	C12	D	:50%	Fail	
2015-02-23	2014-09-30	NCGC00091902-08	144214009	4004	400048	Tox21_400048-2_04_LCMS.pdf	OPANS LCMS	T4	Batch5- SP123653	P11	D	50%	Fail	
2014-10-24		NCGC00091902-05	144204726	4004	111175	Tox21_111175_00_GCMS1.pdf	NIST GCMS	то				/IW Confirmed, urity>90%	Pass	

"Analysis in Progress"

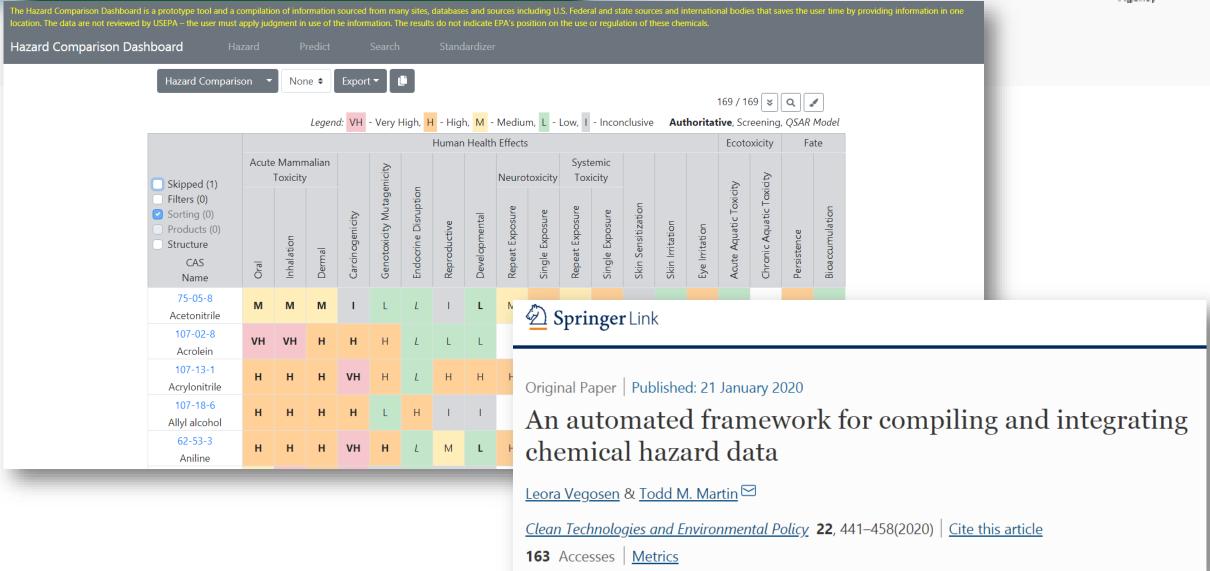




- Some chemicals remain "analysis in progress" after many years
- We have examined the data for many of these and the issue is in the analysis method. Examples are: Retention time is longer than run time, cannot get HNMR spectra for chemicals without H

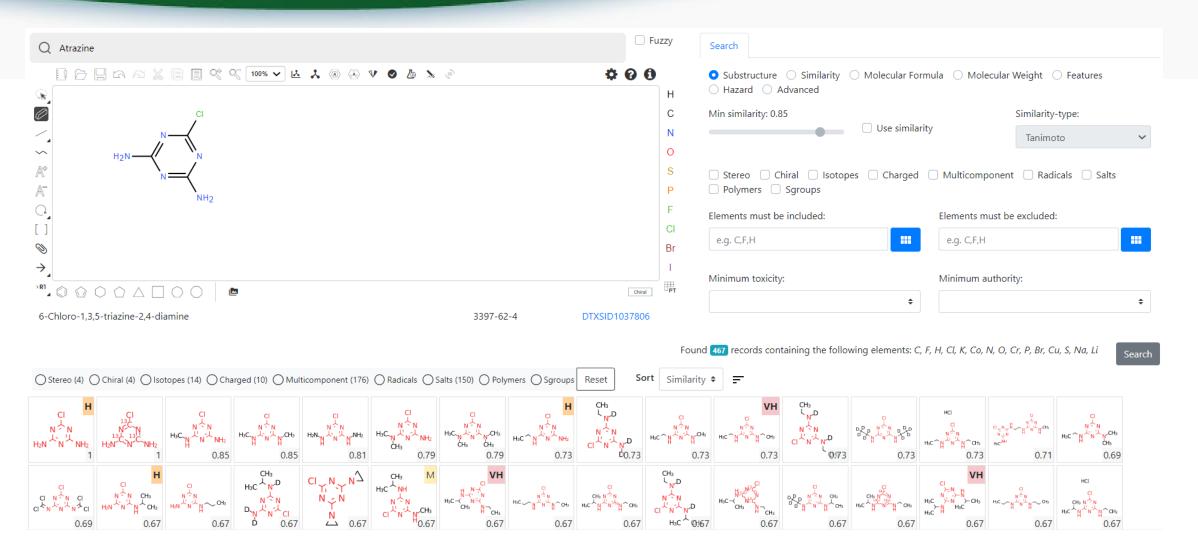
PROTOTYPE Hazard Comparison Dashboard





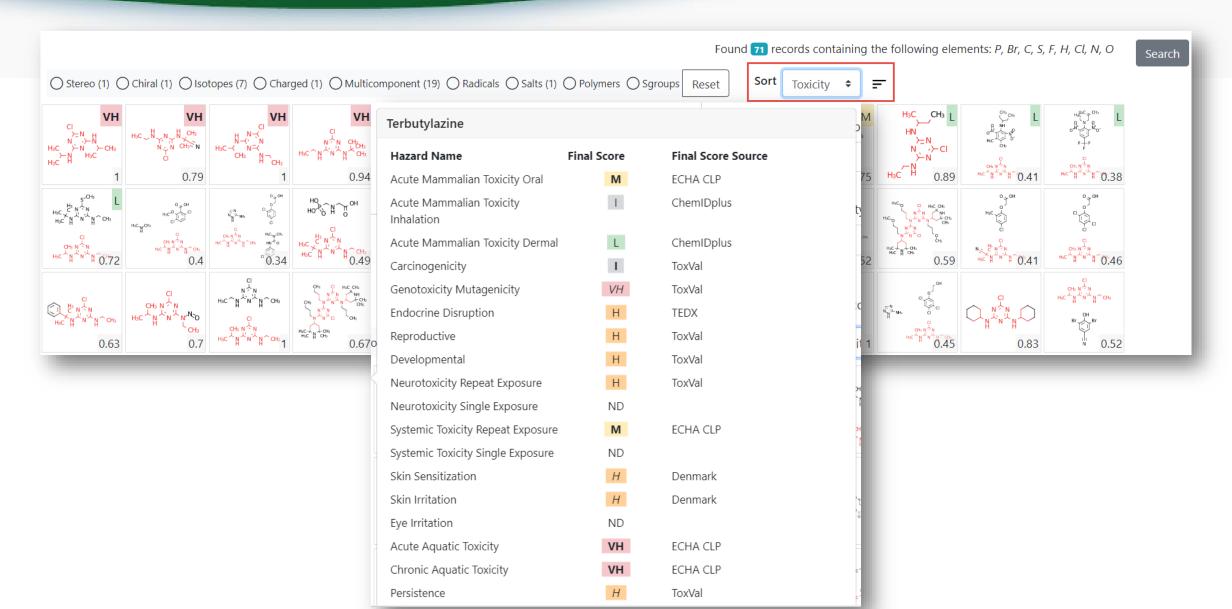
PROTOTYPE Structure/Substructure/Similarity Search available





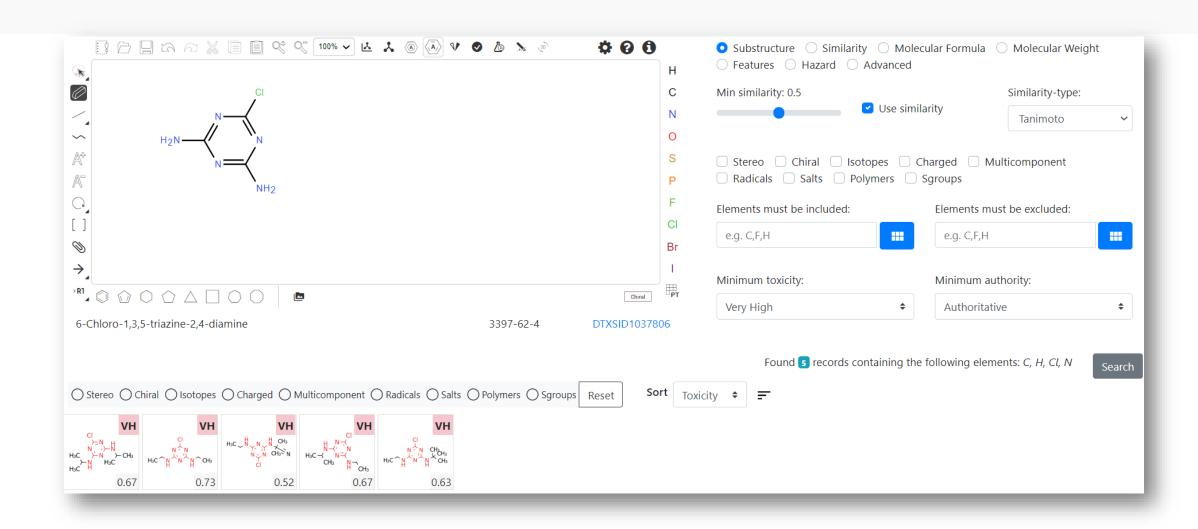
Hit Sets based on Show and Sort on Toxicity





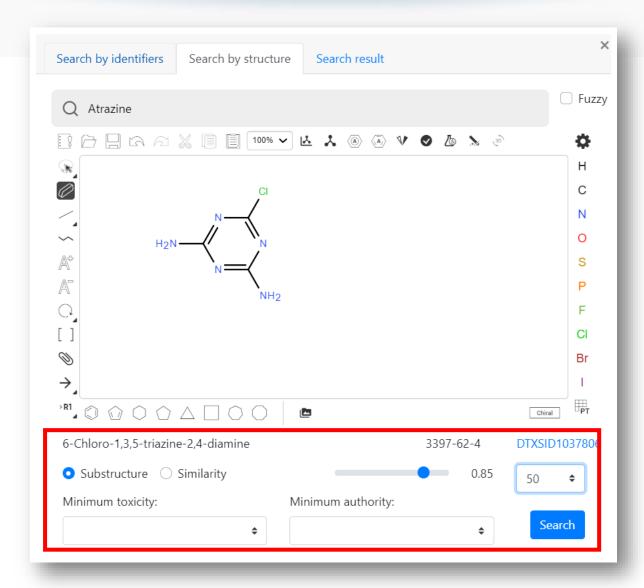
Substructure + "Very High Toxicity" from Authoritative Sources





Substructure/similarity search in HCD

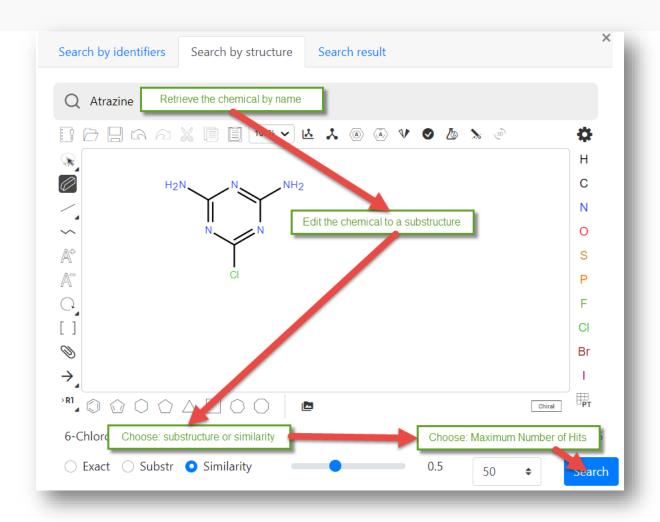




- In HCD search based on substructure and similarity
- Search results can be used as input to HCD
- Then the application can:
 - Run Hazard Comparison
 - Filter based on "available experimental data"

Choices to Create a Set of Hits



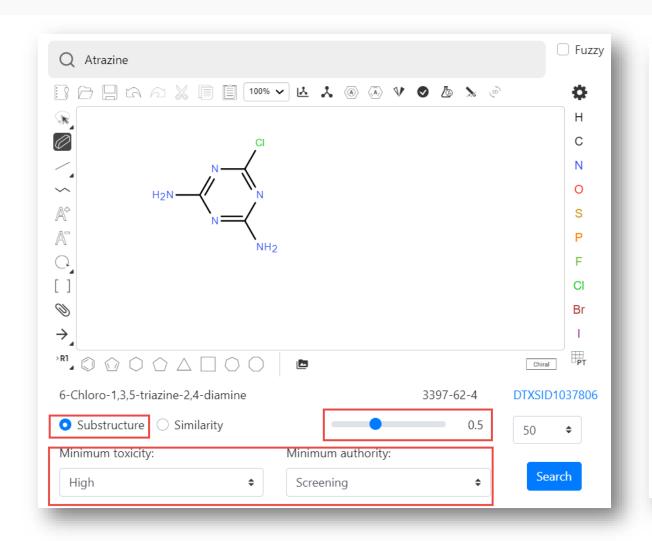


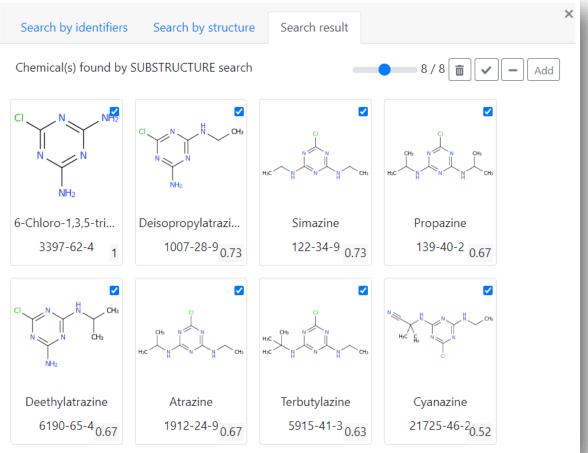
- Looking for a set of chemicals to add to HCD
 - Select structure by name or draw structure
 - Edit the chemical if necessary
 - Choose substructure or similarity
 - Choose number of hits
 - Choose Toxicity "thresholds"

Minimum toxicity:	Minimum authority:
\$	\$

Choose substructure, similarity, thresholds

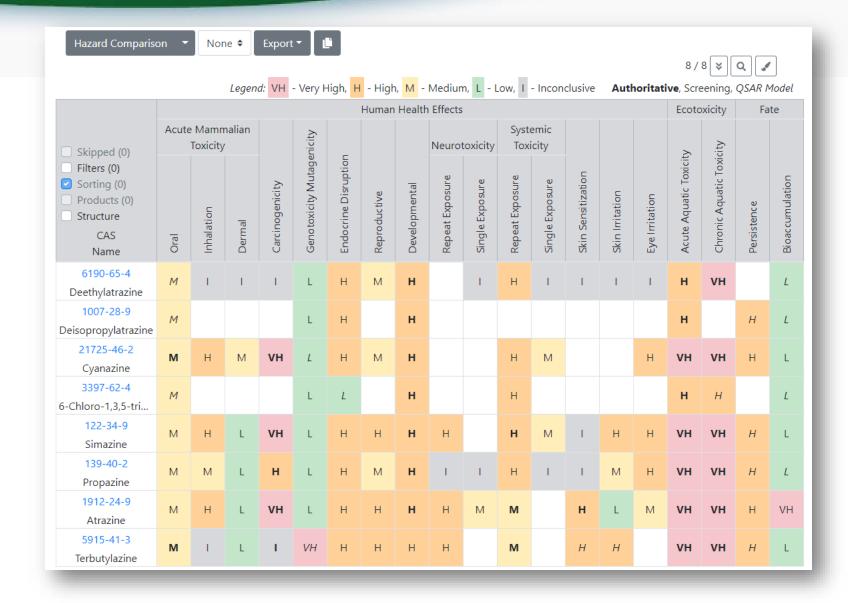






Choose substructure, similarity, thresholds







Online Materials

You want to know more...



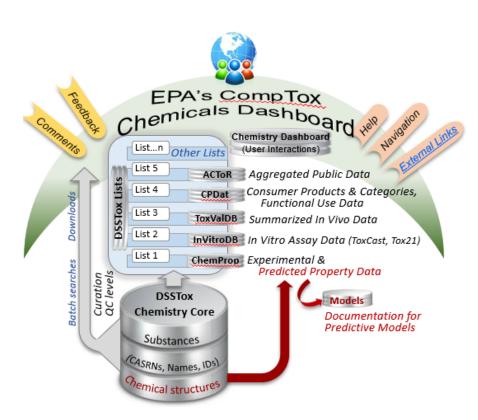
- Lots of resources available
 - Presentations: https://tinyurl.com/w5hqs55
 - Communities of Practice Videos: https://rb.gy/qsbno1
 - Manual: https://rb.gy/4fgydc
 - Latest News: https://comptox.epa.gov/dashboard/news info



Conclusion



- Dashboard access to data for ~882,000 chemicals (and growing)
- Data aggregation for ALL data sources continues unabated
- Flexible search capabilities continue to expand release-to-release
- Expansion of new modules continues with prototype development
- The application is being totally rearchitected at present to also develop a Public API
- ...and now for the deeper dives into the relevant sub-tabs and modules...



Acknowledgments



- Contact: Williams.Antony@epa.gov
- Feedback and follow-up is welcomed! Your questions help
- The dashboard is based on the efforts of many more team members than us. Many collaborators provide data also.
- Thank you NURA for hosting us



EPA's Center for Computational Toxicology and Exposure