

The EPA CompTox Chemicals Dashboard: A Web-Based Data Integration Hub for Environmental Chemistry and Toxicology Data

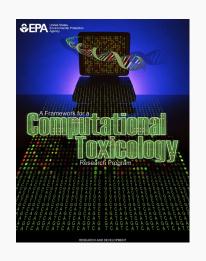
Antony Williams

U.S. Environmental Protection Agency, RTP, NC

This work was reviewed by the U.S. EPA and approved for presentation but does not necessarily reflect official Agency policy.

National Center for Computational Toxicology



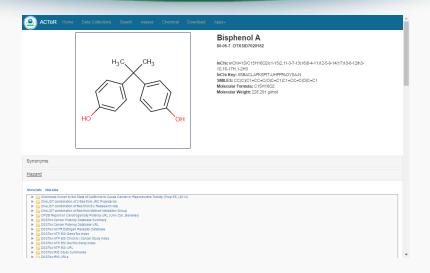


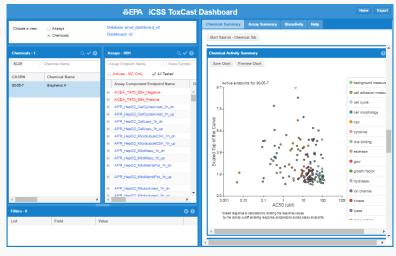


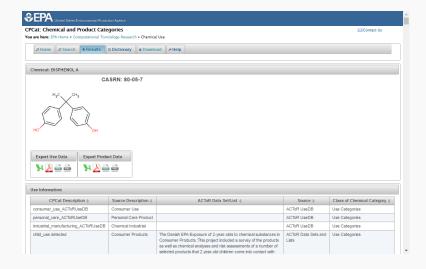
- National Center for Computational Toxicology established in 2005 to integrate:
 - High-throughput and high-content technologies
 - Modern molecular biology
 - Data mining and statistical modeling
 - Computational biology and chemistry
- Currently staffed by ~60 employees as part of EPA's Office of Research and Development
- Home of ToxCast & ExpoCast research efforts
- Key partner in U.S. Tox21 federal consortium
- Multiple cross-division collaborations (e.g. NERL, OPP, OPPT)

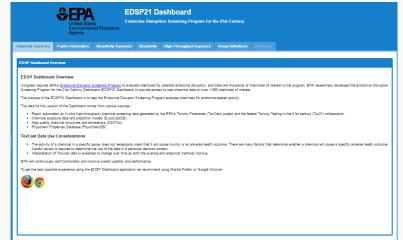
Earlier Dashboard Applications











The CompTox Portal https://comptox.epa.gov/





Environmental Topics

Laws & Regulations

About EPA

Search EPA.gov

.













The CompTox Portal https://comptox.epa.gov/





Environmental Topics

Laws & Regulations

About EPA

Search EPA.gov

,













The CompTox Chemicals Dashboard



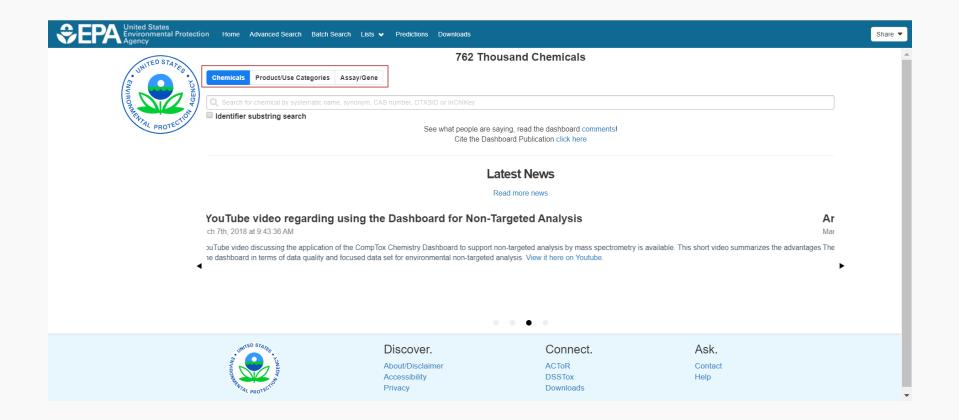
A publicly accessible website delivering access:

- ~765,000 chemicals with related property data
- Experimental and predicted physicochemical property 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
- DOWNLOADABLE Open Data for reuse and repurposing

CompTox Chemicals Dashboard

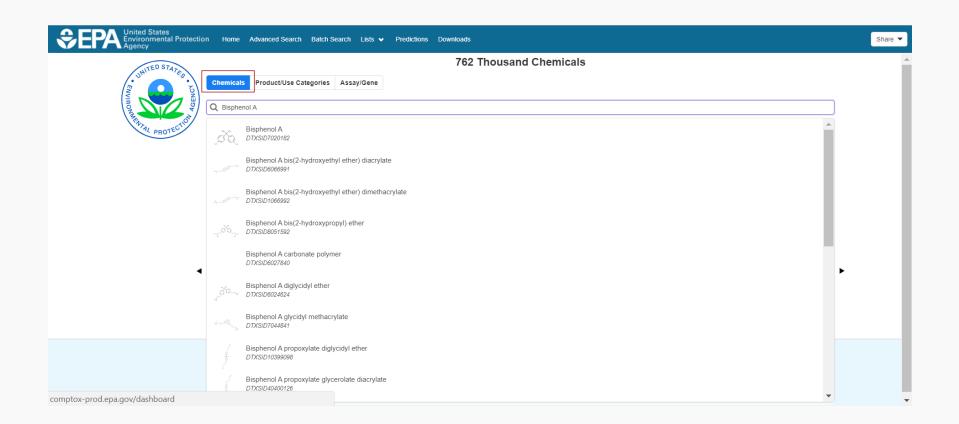
https://comptox.epa.gov/dashboard





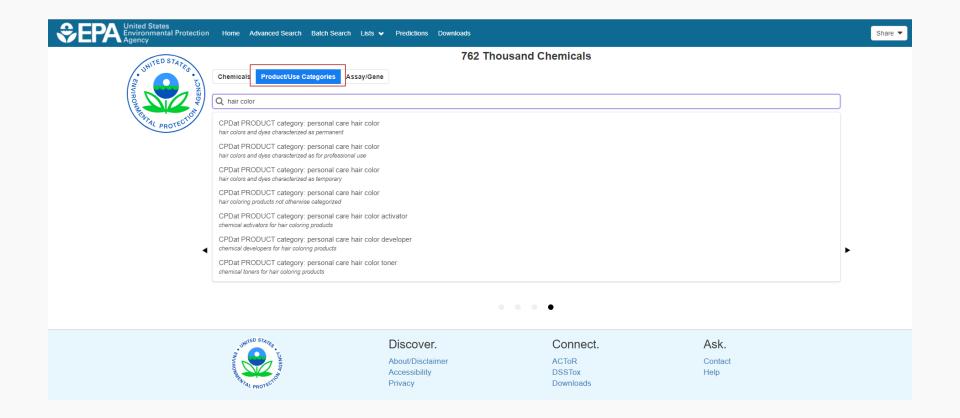
CompTox Chemicals Dashboard Chemicals





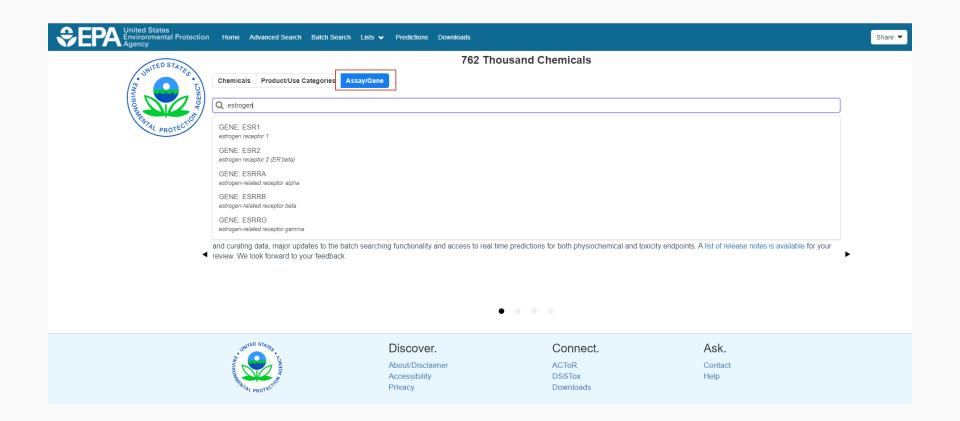
CompTox Chemicals Dashboard Products and Use Categories





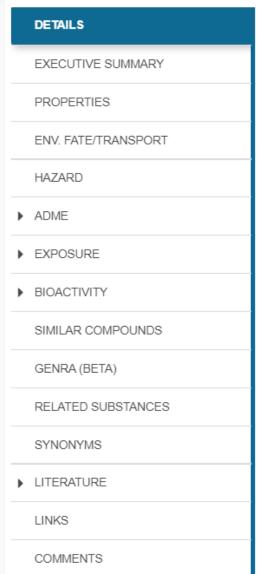
CompTox Chemicals Dashboard Assays and Genes

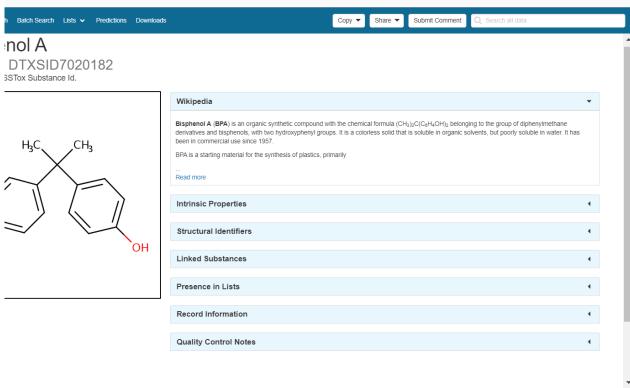




Detailed Chemical Pages

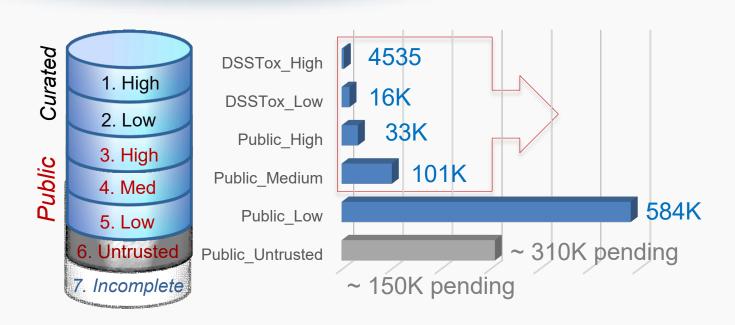






Approximately 15 Years of Data... Growing with daily curation





QC Levels

DSSTox_High: Hand curated and validated

DSSTox_Low: Hand curated and confirmed using multiple public sources

Public_High: Extracted from EPA SRS and confirmed to have no conflicts in ChemID and PubChem

Public_Medium: Extracted from ChemID and confirmed to have no conflicts in PubChem

Public_Low: Extracted from ACToR or PubChem

Public_Untrusted: Postulated, but found to have conflicts in public sources

Available Properties



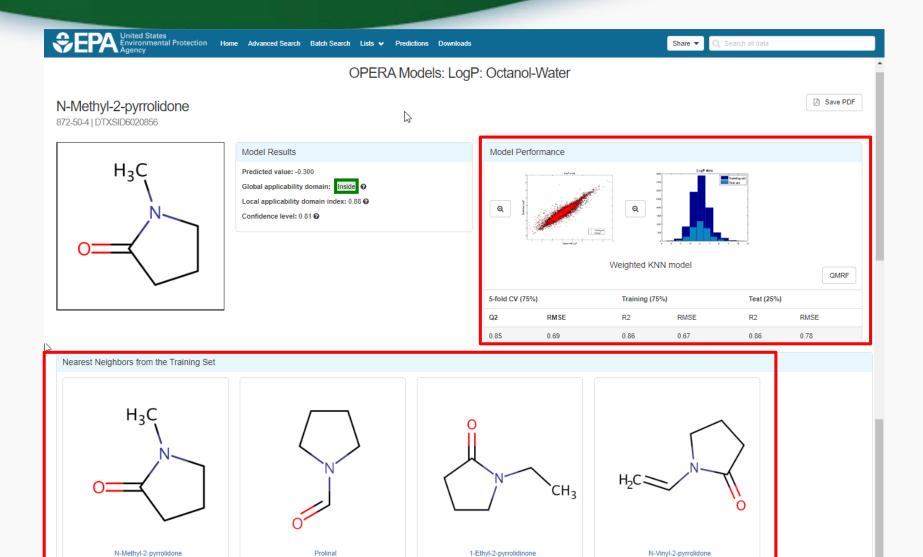
- Solubility
- Melting Point
- Boiling Point
- LogP (Octanol-water partition coefficient)
- Atmospheric Hydroxylation Rate
- LogBCF (Bioconcentration Factor)
- Biodegradation Half-life
- Henry's Law Constant
- Fish Biotransformation Half-life
- LogKOA (Octanol/Air Partition Coefficient)
- LogKOC (Soil Adsorption Coefficient)
- Vapor Pressure
- And more...

Developing "NCCT Models"



- When we don't have experimental data we predict the properties
- Our approach to modeling:
 - Obtain high quality training sets
 - Apply appropriate modeling approaches
 - Validate performance of models
 - Define the applicability domain and model limitations
 - Use models to predict properties across our full datasets
 - Release as Open Data and Open Models





Measured: -0.0400

Predicted: -0.211

Measured: 0.37

Predicted: 2.65e-01

Measured: -0.380

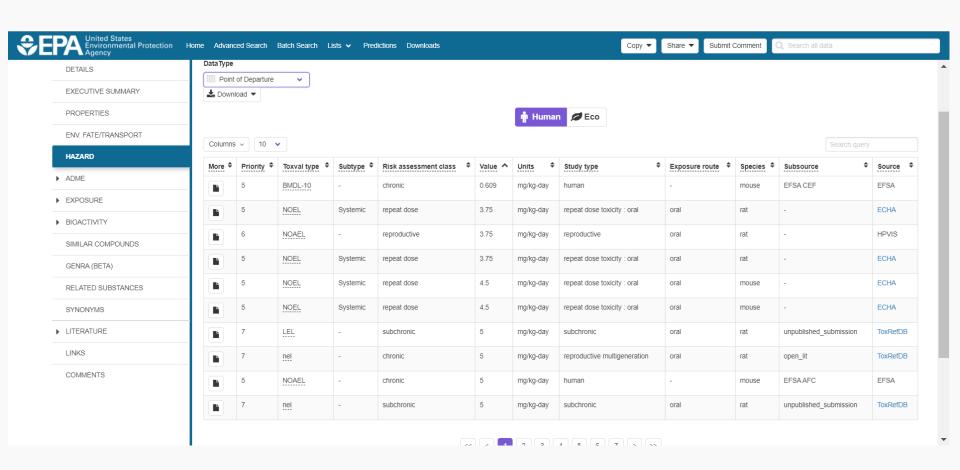
Predicted: -0.300

Measured: -0.320

Predicted: -0.226

Access to Chemical Hazard Data





Hazard Data from "ToxVal_DB"

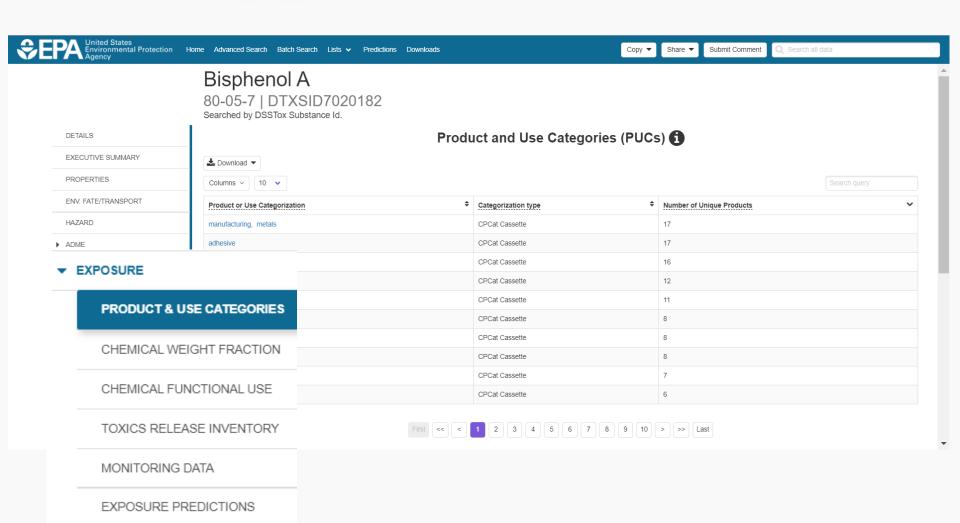


- ToxVal Database contains following data:
 - -30,050 chemicals
 - -772,721 toxicity values
 - -29 sources of data
 - -21,507 sub-sources
 - -4585 journals cited
 - -69,833 literature citations

Sources of Exposure to Chemicals

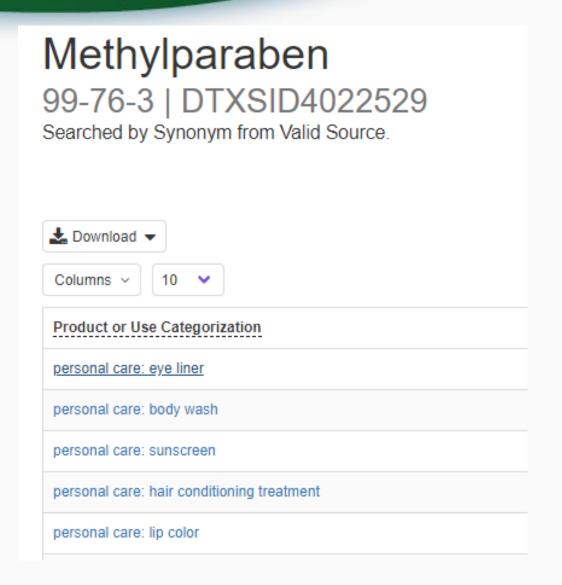
PRODUCTION VOLUME





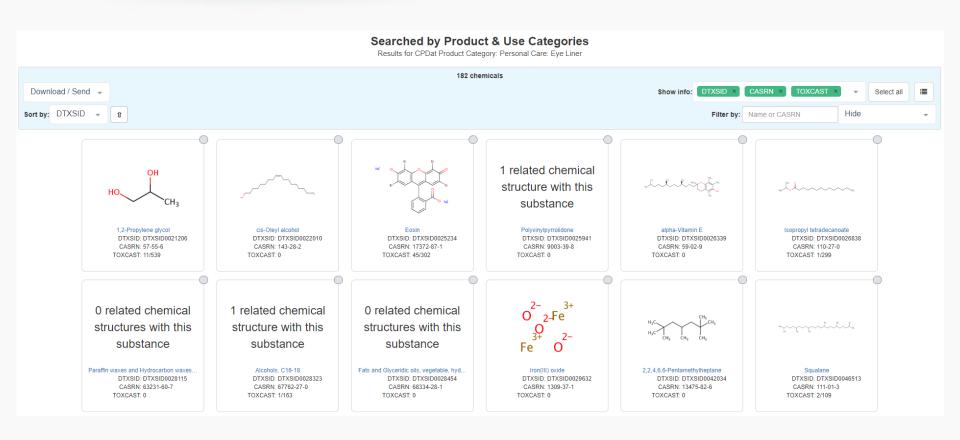
What chemicals in what product and use categories?





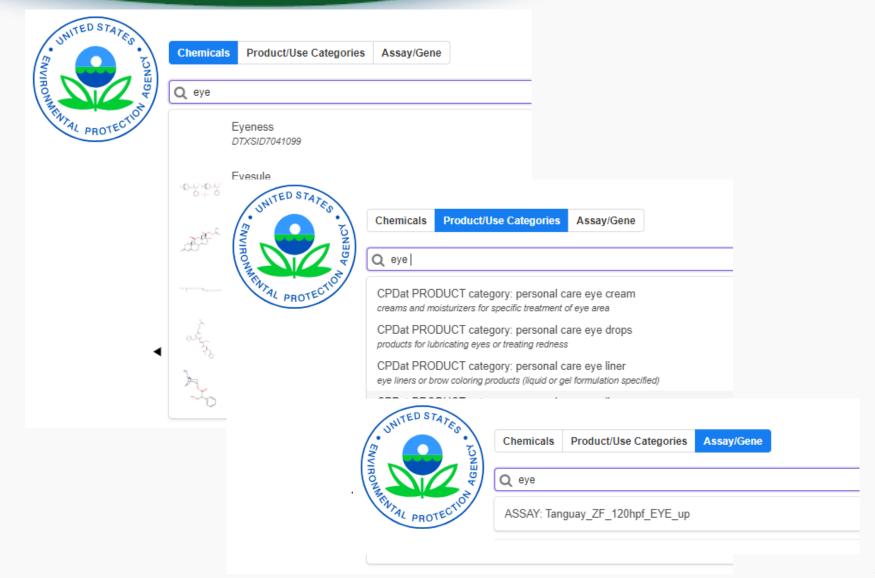
What chemicals in what product and use categories?





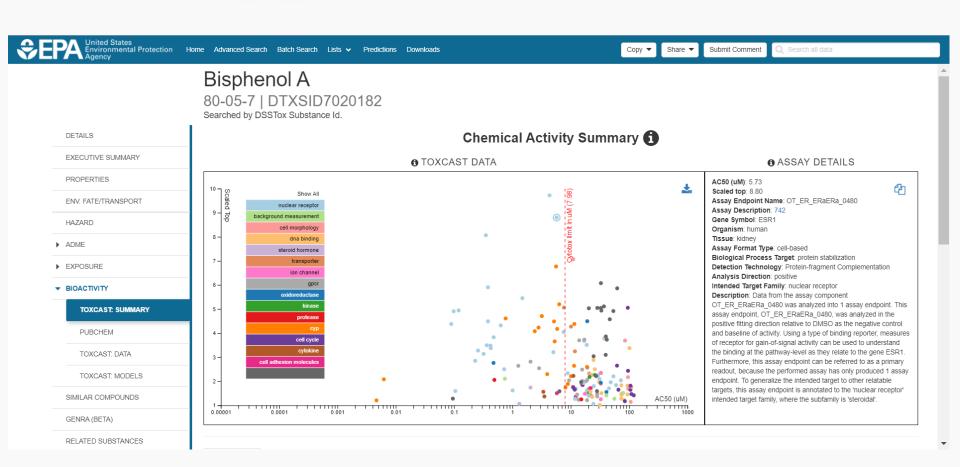
Remember home page searches Searching for "eye"...





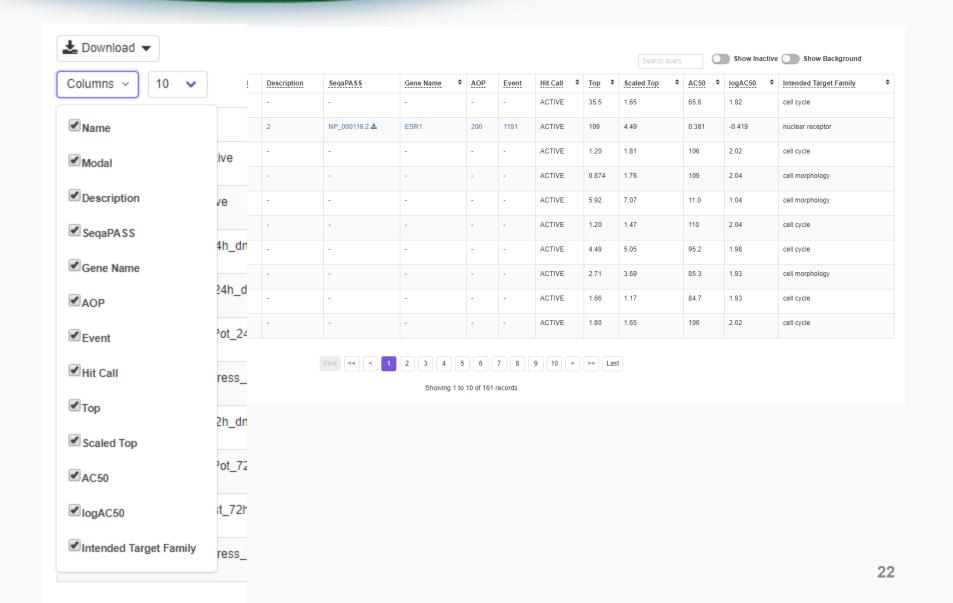






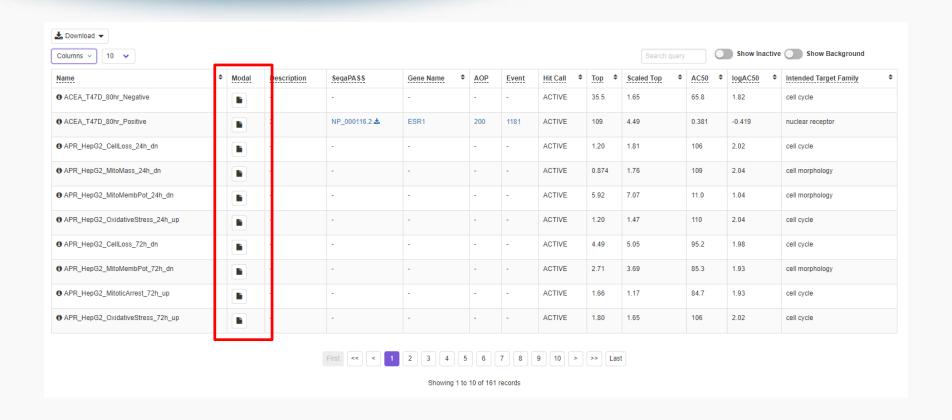
ToxCast and Tox21





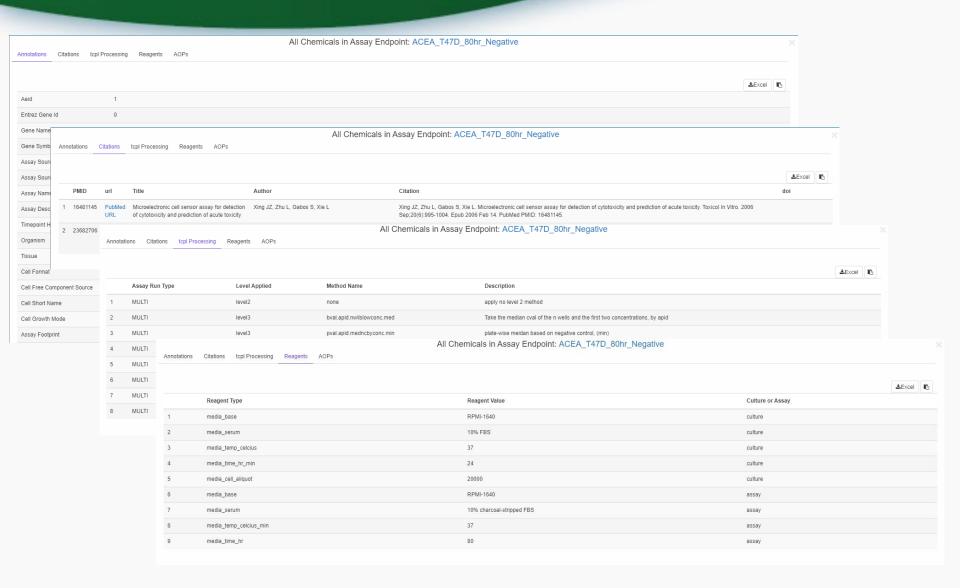
ToxCast and Tox21





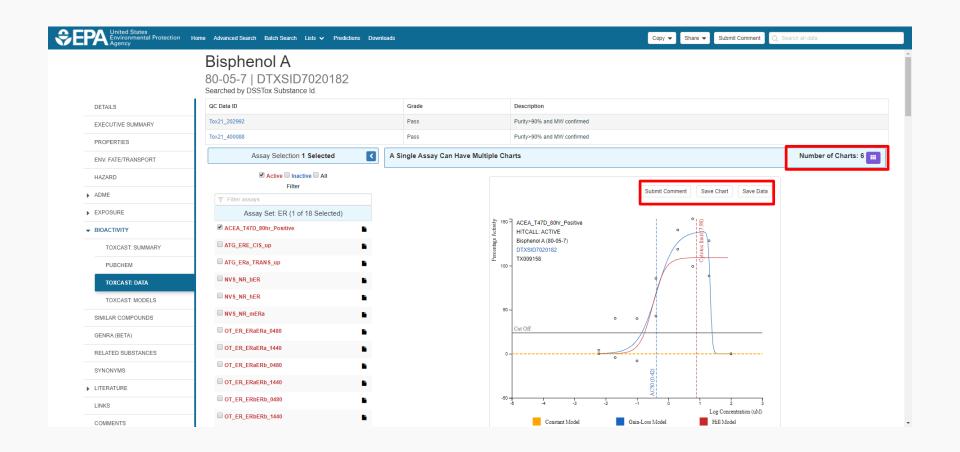
Assay Modal Details





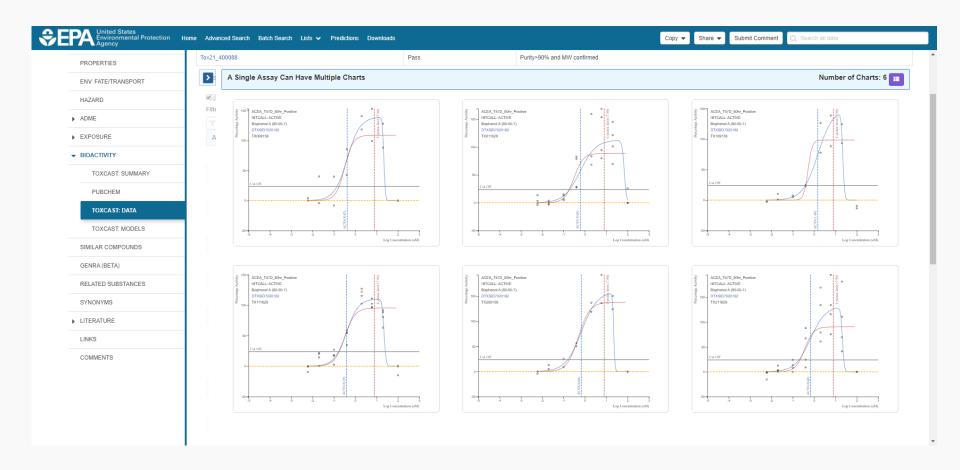
ToxCast and Tox21





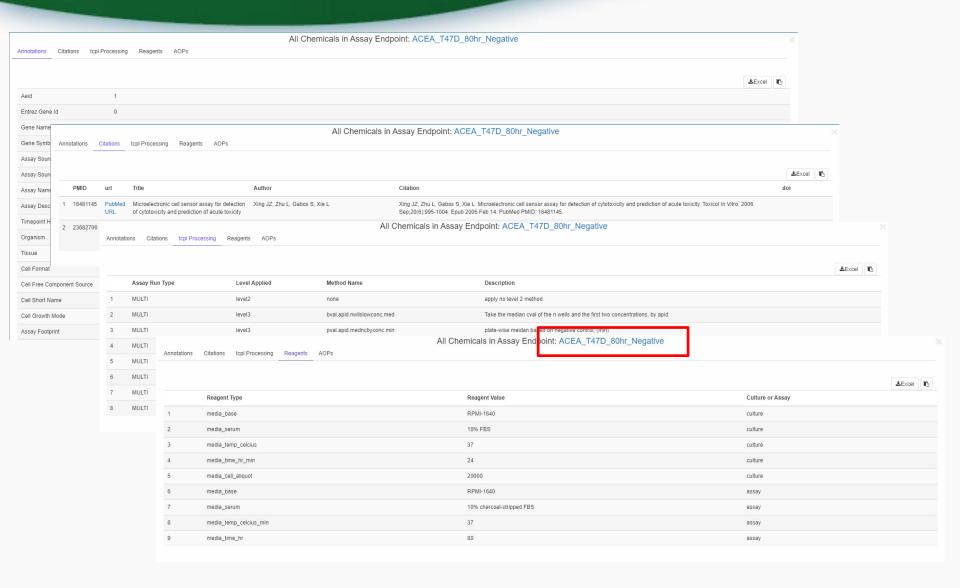
In Vitro Bioassay Screening Multi-chart Display





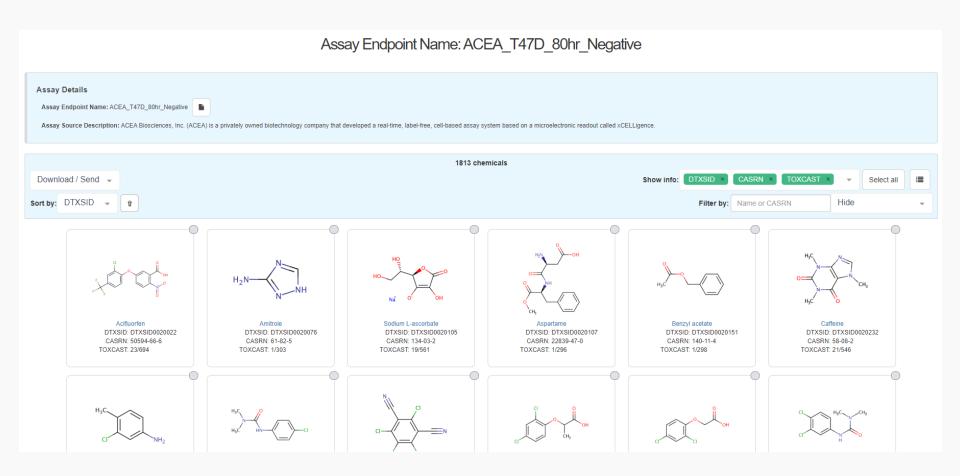
Assay Modal Details





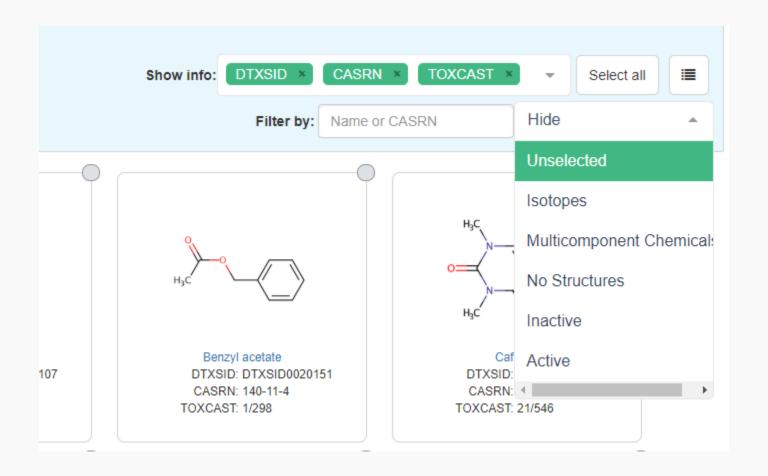
List of Chemicals for an Assay





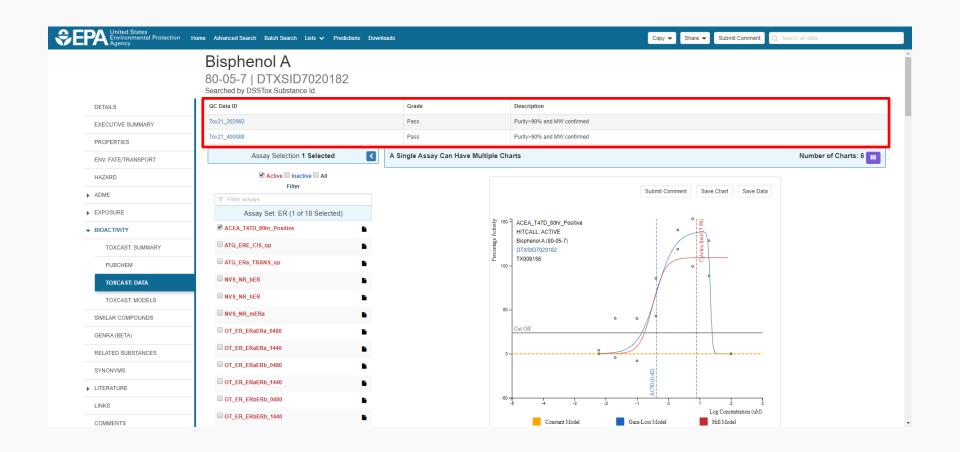
Choose Display Details





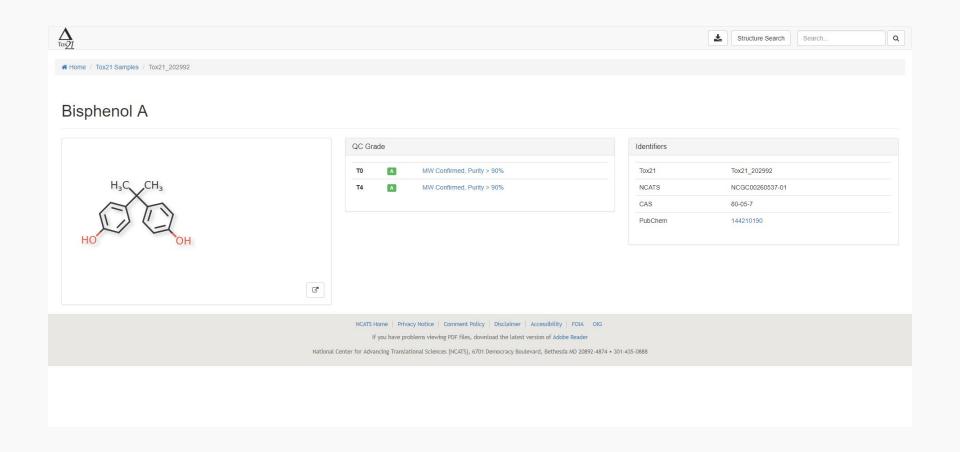
ToxCast and Tox21





Access to Analytical QC Data





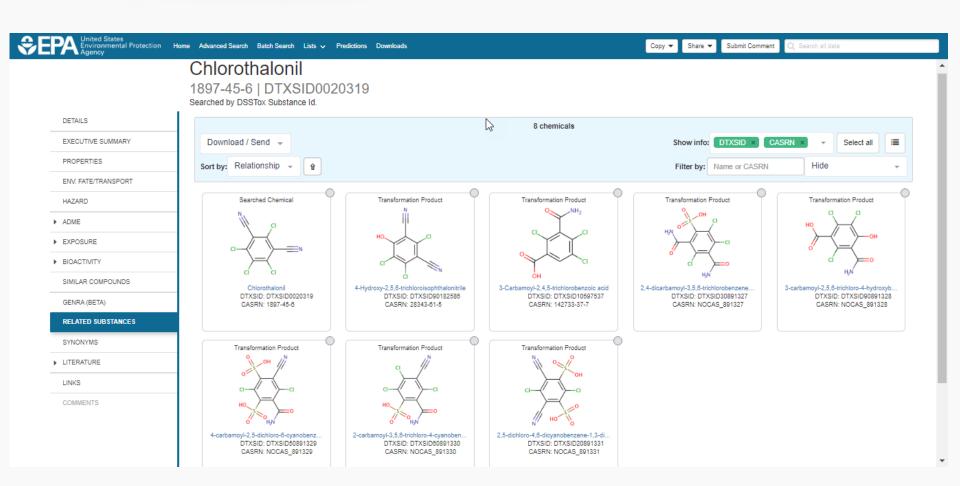
Access to Analytical QC Data





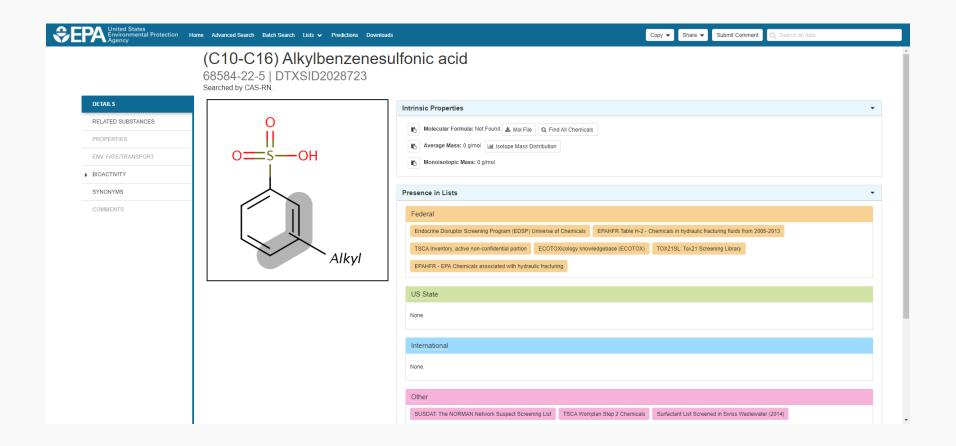
Related Substances e.g. Transformation Products





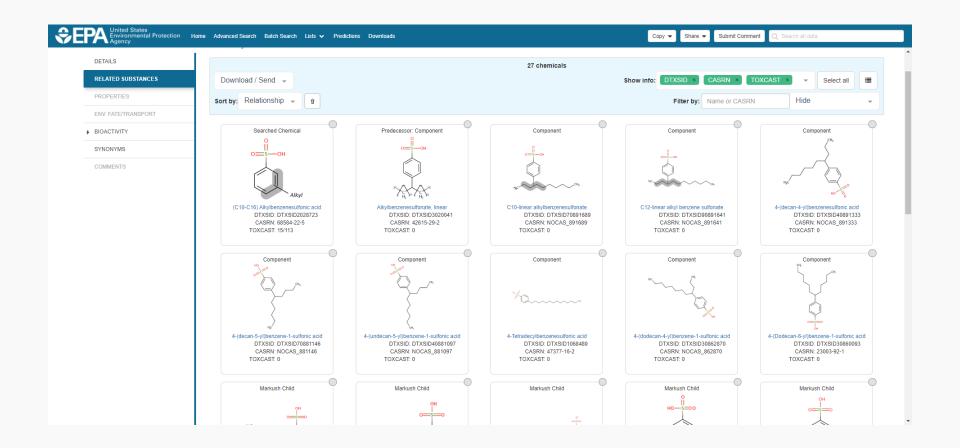
UVCB Chemicals





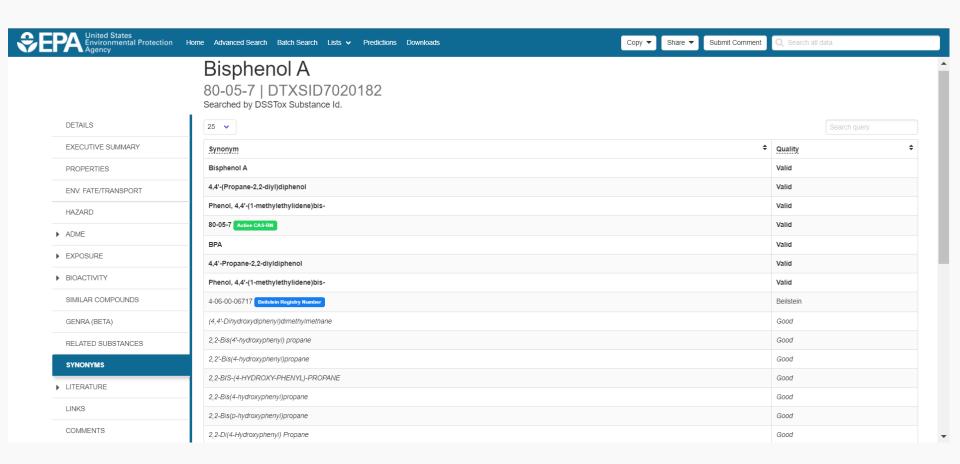
Related Substances for UVCB Chemicals





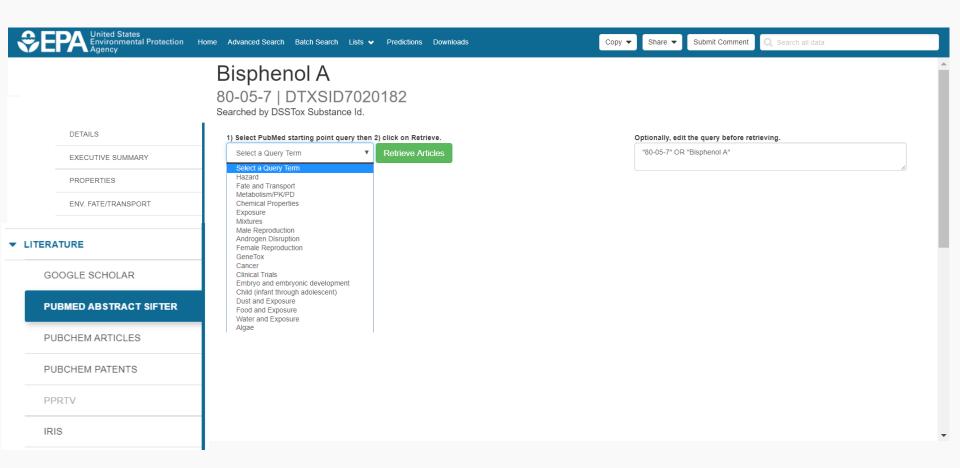
Identifiers to Support Searches





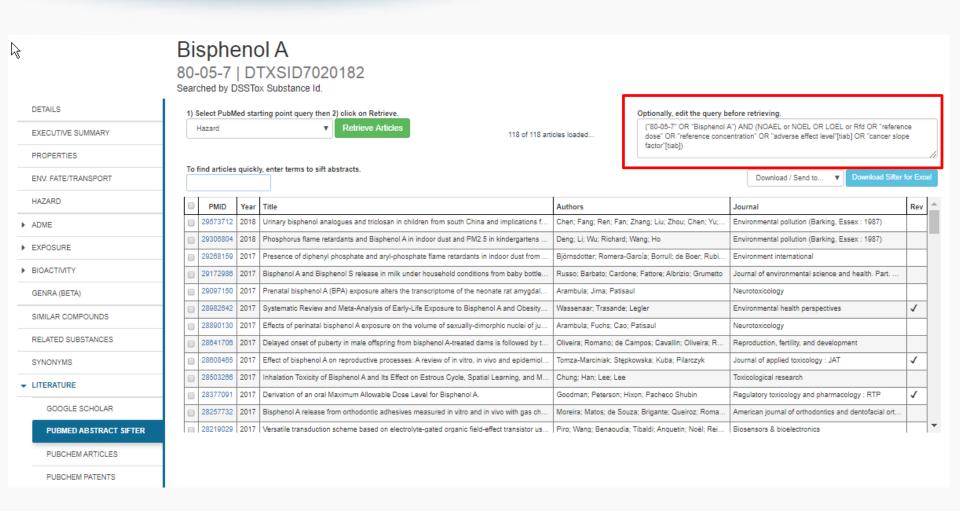
Literature Searches and Links





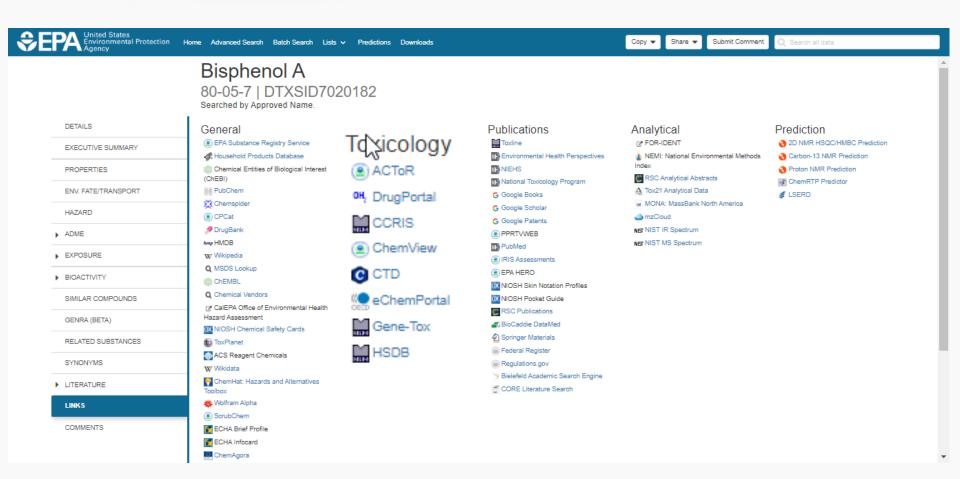
Abstract Sifter – PubMed Integration searching >28 million abstracts





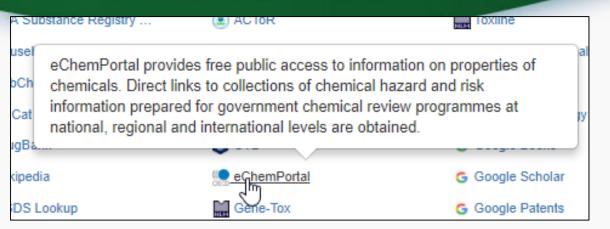
External Links to ~80 websites Growing list of out links -



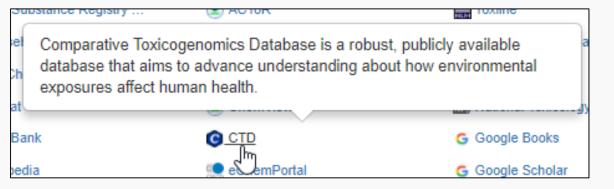


Integrated Linkouts





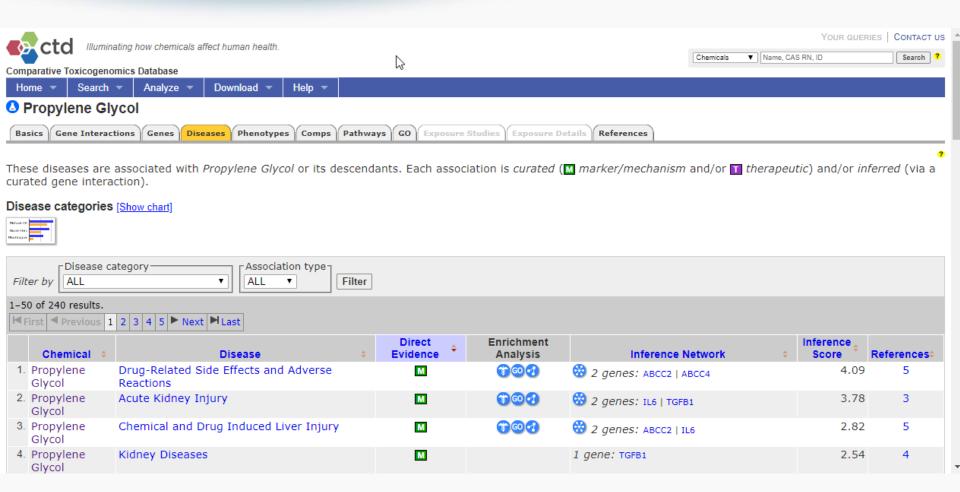




Integrated Linkouts

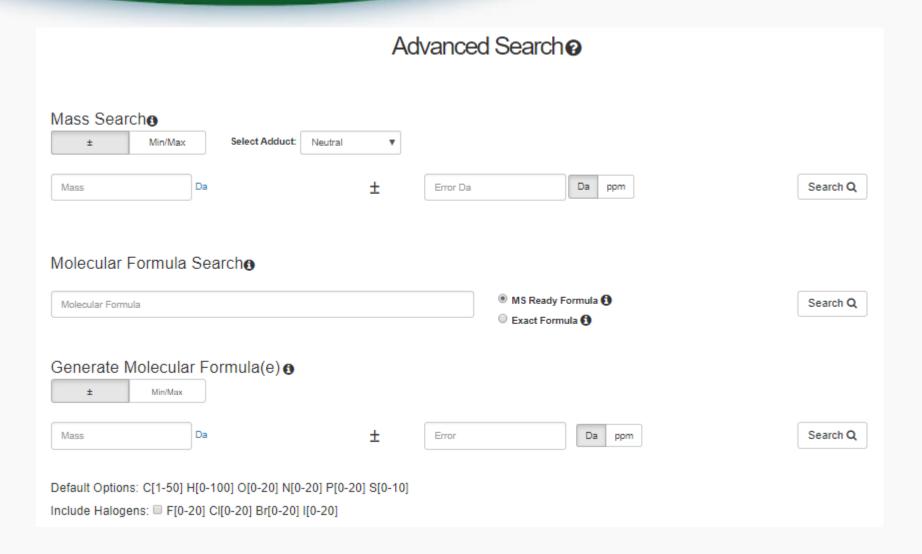
Comparative Toxicogenomics DB





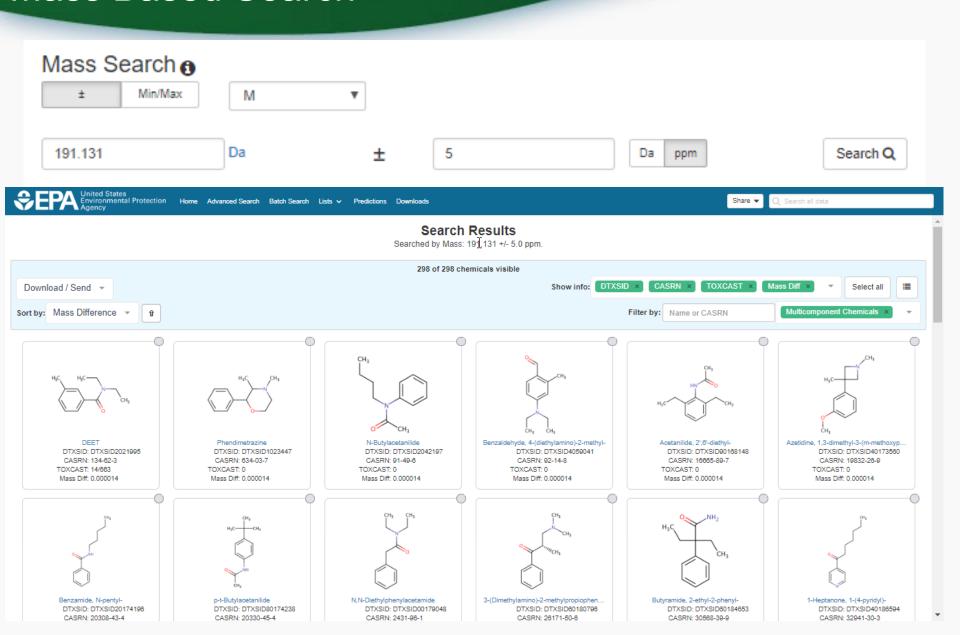
Mass and Formula Searches Supporting Mass Spectrometry





Advanced Searches Mass Based Search





Batch Searching



 Singleton searches are useful but we work with thousands of chemicals!

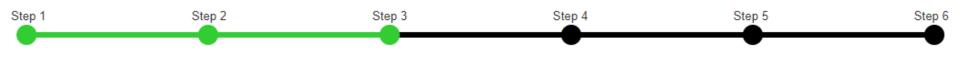
Typical questions

- What chemicals can I get for 5000 CAS Numbers?
- Can I get predicted properties for 1000 chemicals?
- What is the list of chemicals for the formula C_xH_yO_z?
- What is the list of chemicals for a mass +/- error ?
- Can I get chemical lists in Excel files? In SDF files?

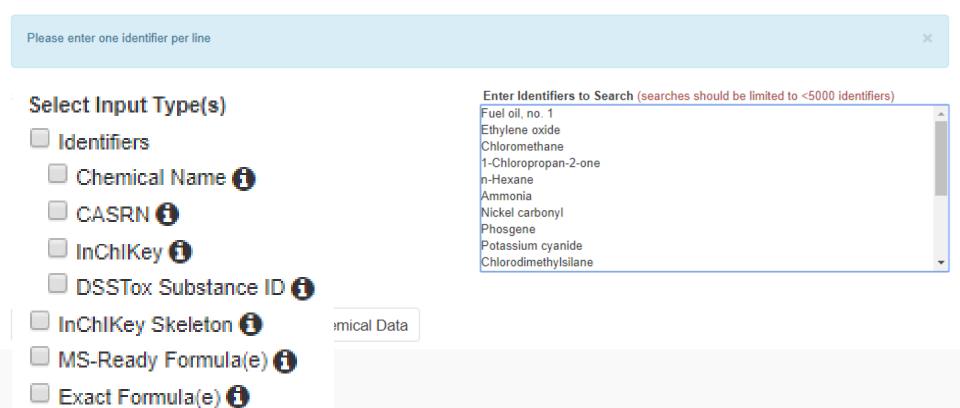
Batch Searching

Monoisotopic Mass





Step Three: Select Download Data or Display Chemicals



Batch Searching



lect Output Format:				
Excel •	♣ Download			
Customize Results Select All Select All in Lists Chemical Identifiers DTXSID Chemical Name CAS-RN	Presence in Lists: ICCVAM test method evaluation report: in vitro ocular toxicity test methods 40CFR355 A list of all PBDEs (Polybrominated diphenyl ethers) A list of all PCBs (Polychlorinated biphenyls) A list of polycyclic aromatic hydrocarbons Acute exposure guideline levels Algal Toxins			
□ InChlKey 1 □ IUPAC Name 1 Structures	■ Androgen Receptor Chemicals ■ APCRA Chemicals for Prospective Analysis			
Mol File SMILES InChI String MS-Ready SMILES MS-Ready SMILES	■ APCRA Chemicals for Retrospective Analysis ■ APCRA Chemicals for Retrospective Analysis_App_List_448_Chemicals ■ ATSDR Minimal Risk Levels (MRLs) for Hazardous Substances ■ ATSDR Toxic Substances Portal Chemical List ■ Bisphenol Compounds			
□ QSAR-Ready SMILES Intrinsic And Predicted Properties □ Molecular Formula □ Average Mass ①	 California Office of Environmental Health Hazard Assessment Chemicals with interesting names CMAP DNT Screening Library 			
Monoisotopic Mass TEST Model Predictions OPERA Model Predictions	□ Drinking Water Suspects, KWR Water, Netherlands □ EDSP Universe □ EPA Chemicals associated with hydraulic fracturing			

Excel Output

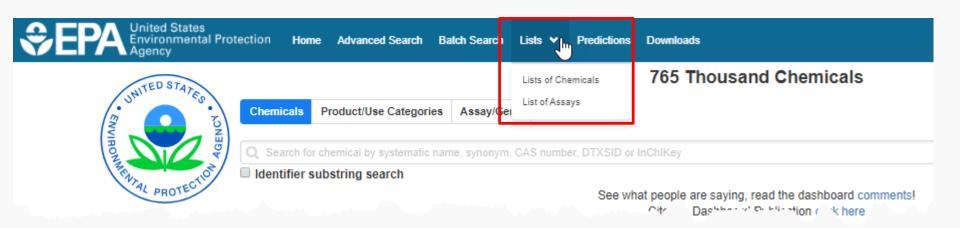


	Α	В	С	D	Е	F	G	Н
1	INPUT	FOUND_BY	DTXSID	PREFERRED_NAME	EXPOCAS	EXPOCAST	NHANES	TOXVAL_D
2	1445-75-6	CAS-RN	DTXSID5024051	Diisopropyl methylpho:	2.09e-08	Υ	-	Υ
3	50-00-0	CAS-RN	DTXSID7020637	Formaldehyde	1.32e-06	Υ	-	Υ
4	107-06-2	CAS-RN	DTXSID6020438	1,2-Dichloroethane	4.9e-06	Υ	-	Υ
5	57-12-5	CAS-RN	DTXSID6023991	Cyanide	-	-	-	Υ
6	7550-45-0	CAS-RN	DTXSID8042476	Titanium tetrachloride	-	-	-	Υ
7	79-01-6	CAS-RN	DTXSID0021383	Trichloroethylene	7.27e-06	Υ	-	Υ
8	121-82-4	CAS-RN	DTXSID9024142	Cyclonite	6.72e-08	Υ	-	Υ
9	108-05-4	CAS-RN	DTXSID3021431	Vinyl acetate	8.3e-05	Υ	_	Υ
10	7803-51-2	CAS-RN	DTXSID2021157	Phosphine	-	-	_	Υ
11	122-66-7	CAS-RN	DTXSID7020710	1,2-Diphenylhydrazine	1.49e-07	Υ	_	Υ
12	101-77-9	CAS-RN	DTXSID6022422	4,4'-Methylenedianiline	6.08e-06	Υ	-	Υ
13	14017-34-6	CAS-RN	DTXSID90161250	Selenium difluoride	-	-	-	-
14	75-44-5	CAS-RN	DTXSID0024260	Phosgene	-	-	-	Υ
15	621-64-7	CAS-RN	DTXSID6021032	N-Nitrosodipropylamine	4.55e-07	Υ	_	Υ
16	75-09-2	CAS-RN	DTXSID0020868		2.02e-06	Υ	-	Υ
17	100-41-4	CAS-RN	DTXSID3020596	Ethylbenzene	8.32e-05	Υ	-	Υ
18	7440-28-0	CAS-RN	DTXSID2036035	Thallium	-	-	-	Υ
19	108-88-3	CAS-RN	DTXSID7021360	Toluene	8.61e-05	Υ	-	Υ
20	111-44-4	CAS-RN	DTXSID9020168	Bis(2-chloroethyl) ethe	2.82e-07	Υ	-	Υ
21	7440-42-8	CAS-RN	DTXSID3023922	Boron	-	-	-	Υ
22	7440-29-1	CAS-RN	DTXSID6049800	Thorium	-	-	-	Υ

Lists of Lists



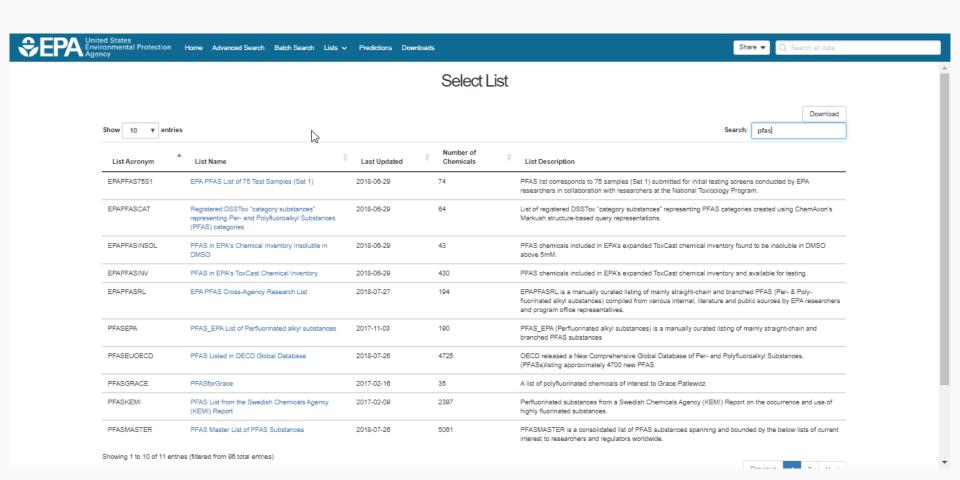
- Lists of chemicals ca. 100 lists
- List of ToxCast/Tox21 assays



11 PFAS Lists

http://comptox-prod.epa.gov/dashboard/chemical_lists





The OECD List of PFAS

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







HOME



The OEGD releases a new list of PFASs

The OECD releases a new list of Per- and Polyfluoroalkyl Substances (PFASs) based on a comprehensive analysis of information available in the public domain. In total, 4730 PFAS-related CAS numbers have been identified and categorised in this study, including several new groups of PFASs that fulfil the common definition of PFASs (i.e. they contain at least one perfluoroalkyl moiety) but have not yet been commonly regarded as PFASs.

This work has been conducted under the OECD/UN Environment Global PFC Group in support of the Strategic Approach to International Chemicals Management (SAICM) and shifting to safer alternatives for PFASs.

The New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances (PFASs) comes with a methodology report also detailing the major findings with respect to the total numbers and types of PFASs identified, the limitations, gaps and challenges identified in the development of the new list, and opportunities for improving the future understanding of PFASs production, use on the global market, and presence in the environment, biota, and other matrices.









The OECD List of PFAS

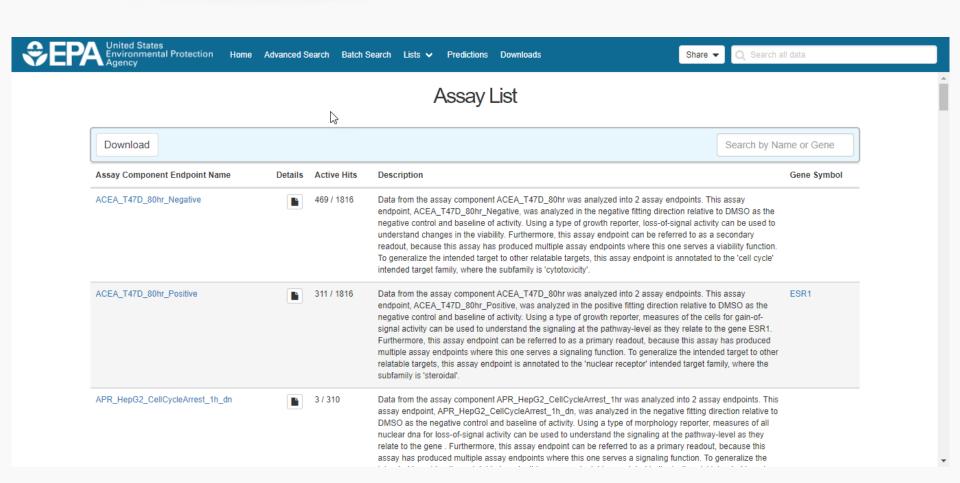




United States Environmental Protection Agency	Home Advanced Search Batch Search Lists ✔ Prediction	s Downloads	Share ▼ Q Search all data			
ß	PFAS Listed in OEC	D Global Database				
	Search PFASEUOECD Chemicals		Q			
	Substring search					
List Details						
that fulfill the common definition methodology report summarisin future understanding of PFASs Source website: http://www.oec A major effort was undertaken to curation methods. The result is	Description: OECD released a New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances (PFASs) listing approximately 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_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: 4725						
Download / Send	4725 chei		× ▼ Select all II			
Download / Send	Show info	CASRN * TOXCAST	▼ Select all			
Sort by: DTXSID	Û	Filter by: Name or CASRN	Hide ▼			

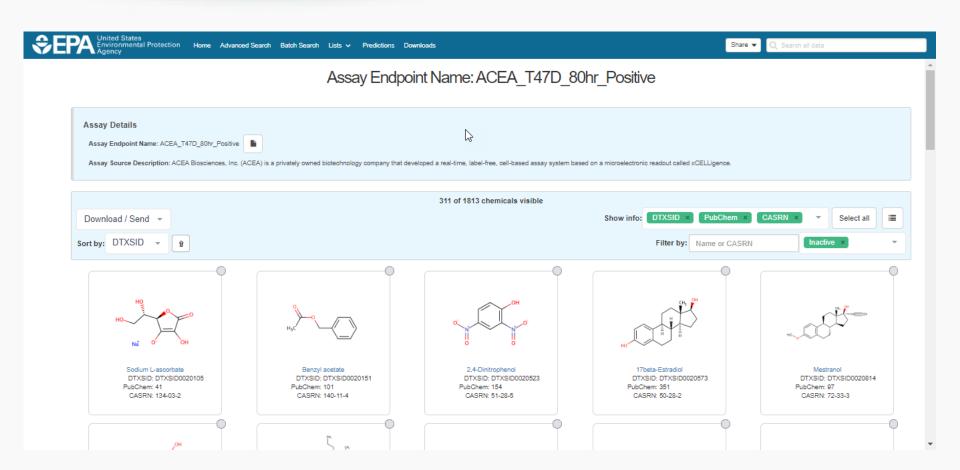
List of Assays





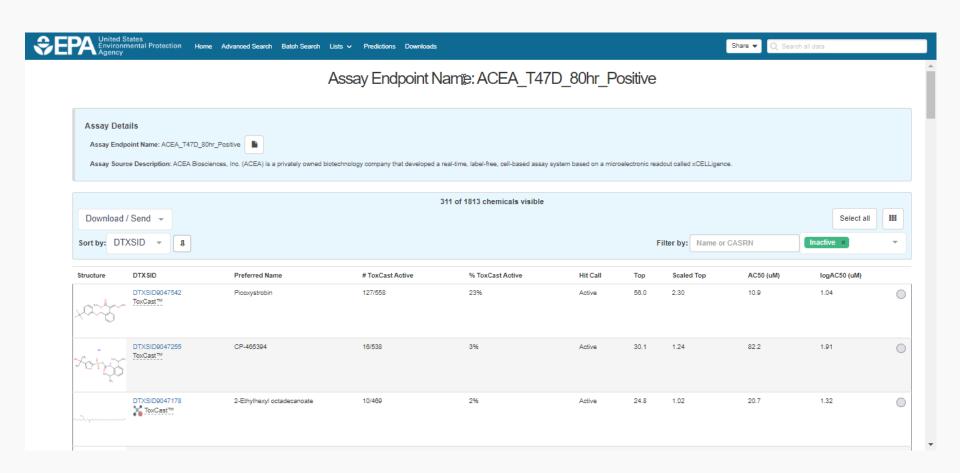
Select an Assay to Navigate Tile View





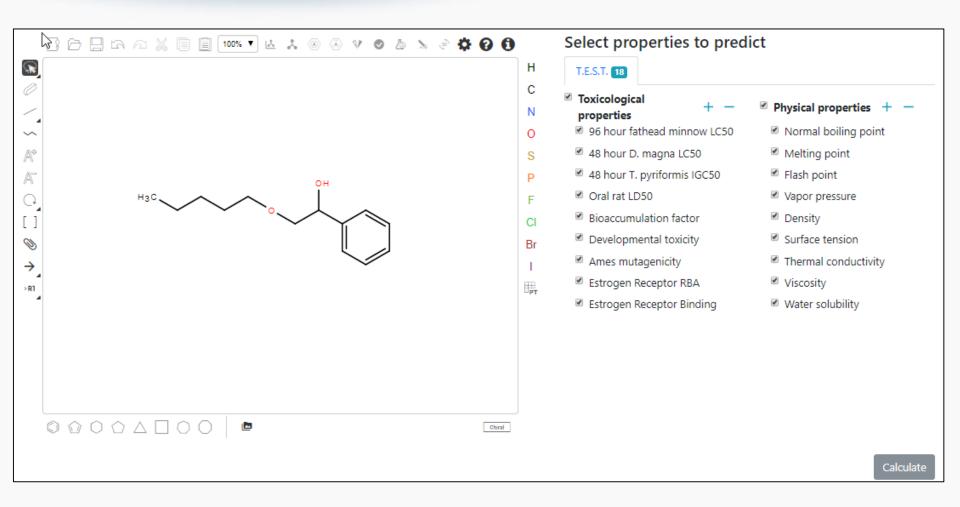
Select an Assay to Navigate Table View





Real-Time Predictions





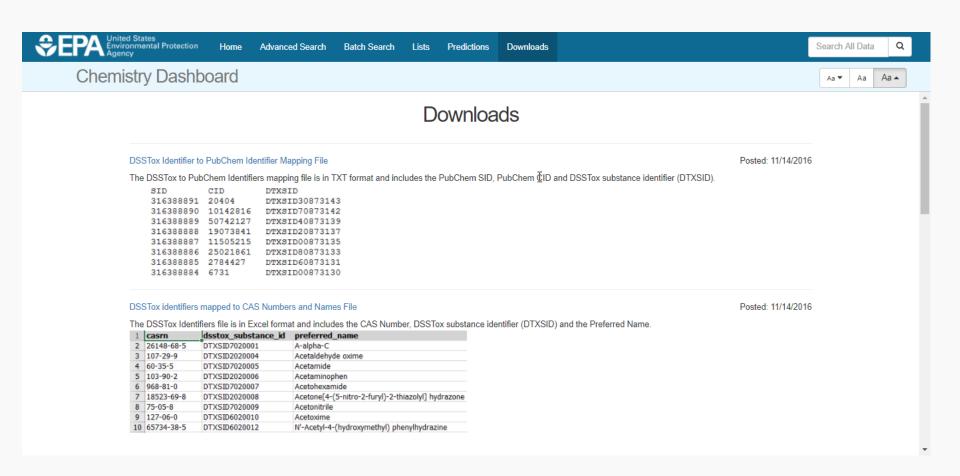
Real-Time Predictions



		Experimental	Prediction				
	Property	Value	Consensus	Hierarchical clustering	Single model	Group contribution	Nearest neighbor
V	96 hour fathead minnow LC50		4.477 -Log10(mol/L) 6.954 mg/L	4.195 -Log10(mol/L) 13.288 mg/L	3.994 -Log10(mol/L) 21.110 mg/L	3.478 -Log10(mol/L) 69.224 mg/L	6.238 -Log10(mol/L) 0.120 mg/L
	48 hour D. magna LC50		4.398 -Log10(mol/L) 8.328 mg/L	3.877 -Log10(mol/L) 27.677 mg/L	4.039 -Log10(mol/L) 19.026 mg/L	4.084 -Log10(mol/L) 17.173 mg/L	5.593 -Log10(mol/L) 0.532 mg/L
	48 hour T. pyriformis IGC50		4.063 -Log10(mol/L) 18.039 mg/L	3.731 -Log10(mol/L) 38.668 mg/L		3.386 -Log10(mol/L) 85.610 mg/L	5.070 -Log10(mol/L) 1.773 mg/L
	Oral rat LD50		1.758 -Log10(mol/kg) 3640.950 mg/kg	1.982 -Log10(mol/kg) 2172.756 mg/kg			1.533 -Log10(mol/kg) 6101.245 mg/kg
	Bioaccumulation factor		1.797 Log10 62.700	2.202 Log10 159.310	1.287 Log10 19.346	1.181 Log10 15.157	2.520 Log10 330.834
	Developmental toxicity		false	false	false		true
	Ames mutagenicity		false	false			false
	Estrogen Receptor RBA		-3.075 Log10 8.418*10 ⁻⁴	-3.078 Log10 8.356*10 ⁻⁴	-3.720 Log10 1.907*10 ⁻⁴		-2.427 Log10 0.004
	Estrogen Receptor Binding		true	true	true	false	true

Downloadable Data Being Updated





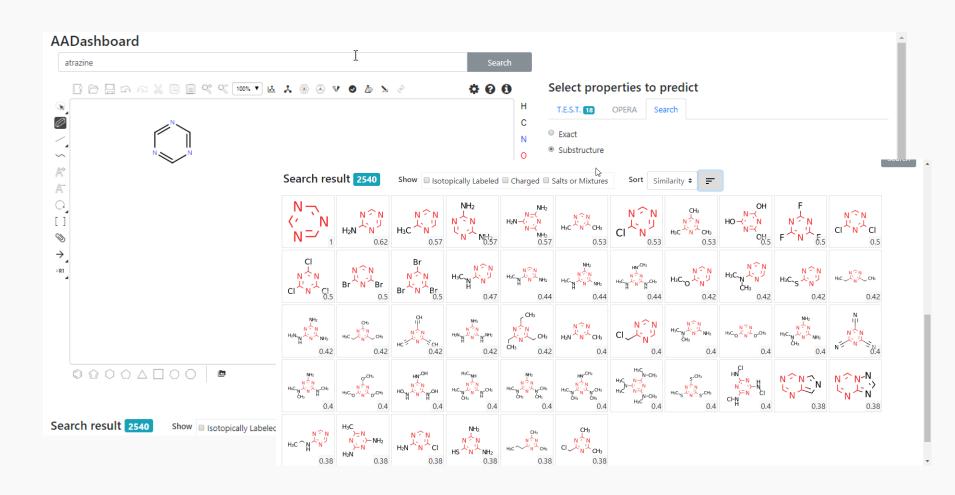
Work in Progress "InvitroDB_v3"



- The last public release of ToxCast data (invitroDB_v2) was in 3rd Quarter of 2015
- The next release invitroDB_v3 is Fall 2018
- Data includes new assays, new chemicals, new pipelining, results of data curation
- Data will also release via the Dashboard
- Data will be available at https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data

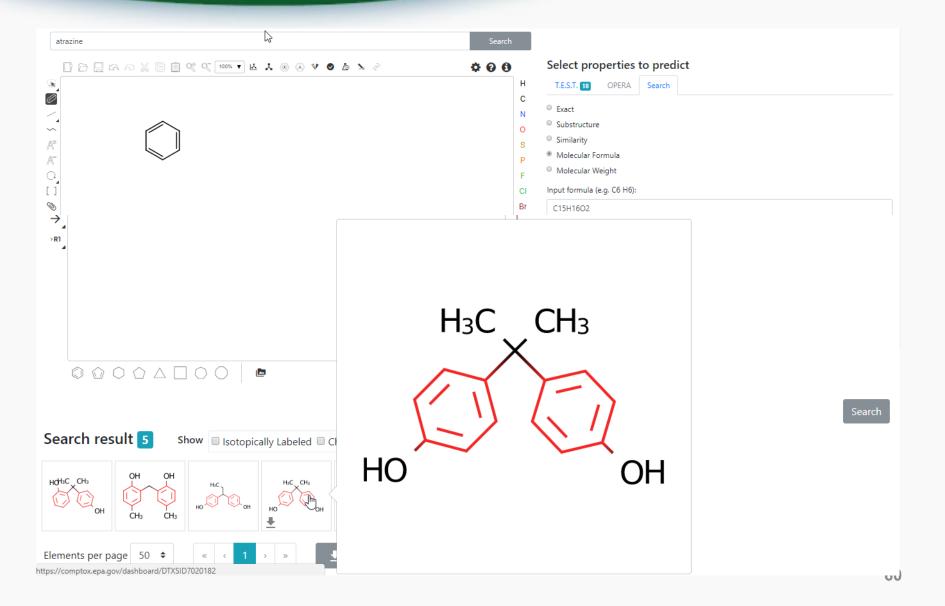
Prototype Development





Prototype Development

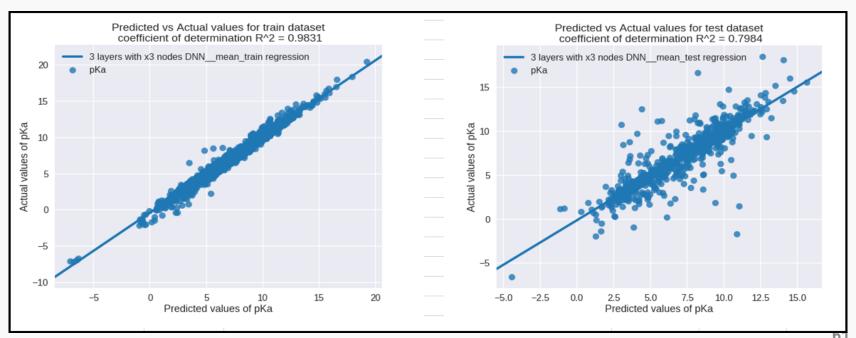




pKa Prediction Model



pKa prediction models based on Open
 Data Set of 8000 chemicals – acidic, basic
 and amphoteric chemicals



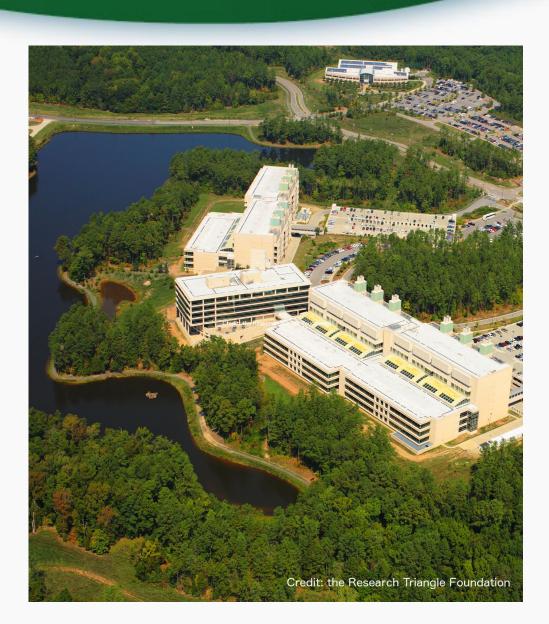
Conclusion



- The CompTox Chemistry Dashboard provides access to data for ~765,000 chemicals
- An expanding list of data types and sources has been integrated
- New searches based on Product Use and Categories and Assay and Gene
- The chemical lists of interest grows with each release
- Next release scheduled for Fall 2018 with InvitroDB v3 data – more chemicals, more assays

Acknowledgements





EPA-RTP

An enormous team of contributors from NCCT

and collaborators from NERL NHERL NRMRL



Antony Williams

NCCT, US EPA Office of Research and Development,

Williams.Antony@epa.gov

ORCID: https://orcid.org/0000-0002-2668-4821

