US EPA CompTox Chemicals Dashboard

Antony Williams

US EPA, Center for Computational Toxicology and Exposure

Durham, NC

919-541-1033

williams.antony@epa.gov

Conflict of Interest Statement

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

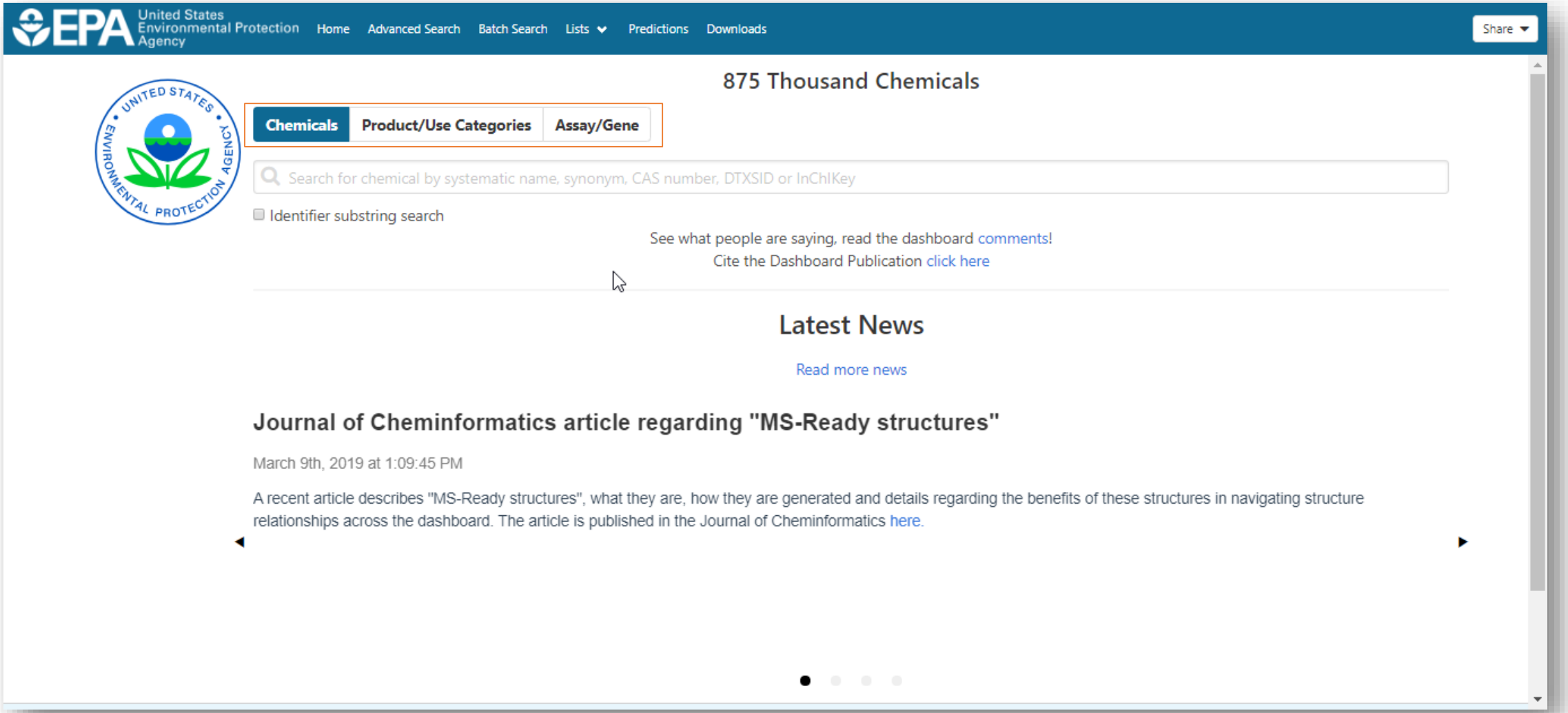
Abbreviations

- CompTox - Computational Toxicology
- DSSTox - Distributed Structure Searchable Toxicity DB
- CASRN - Chemical Abstracts Registry Number
- InChI - International Chemical Identifier
- QMRF – QSAR Model Report Format
- ToxVal - Toxicity Value Database
- OPERA - OPEn structure–activity Relationship App
- TEST - Toxicity Estimation Software Tool
- ToxCAST - Toxicity Forecaster
- CERAPP - Collaborative Estrogen Receptor Activity Prediction Project
- CoMPARA - Collaborative Modeling Project for Androgen Receptor Activity
- SDF - Structure data file

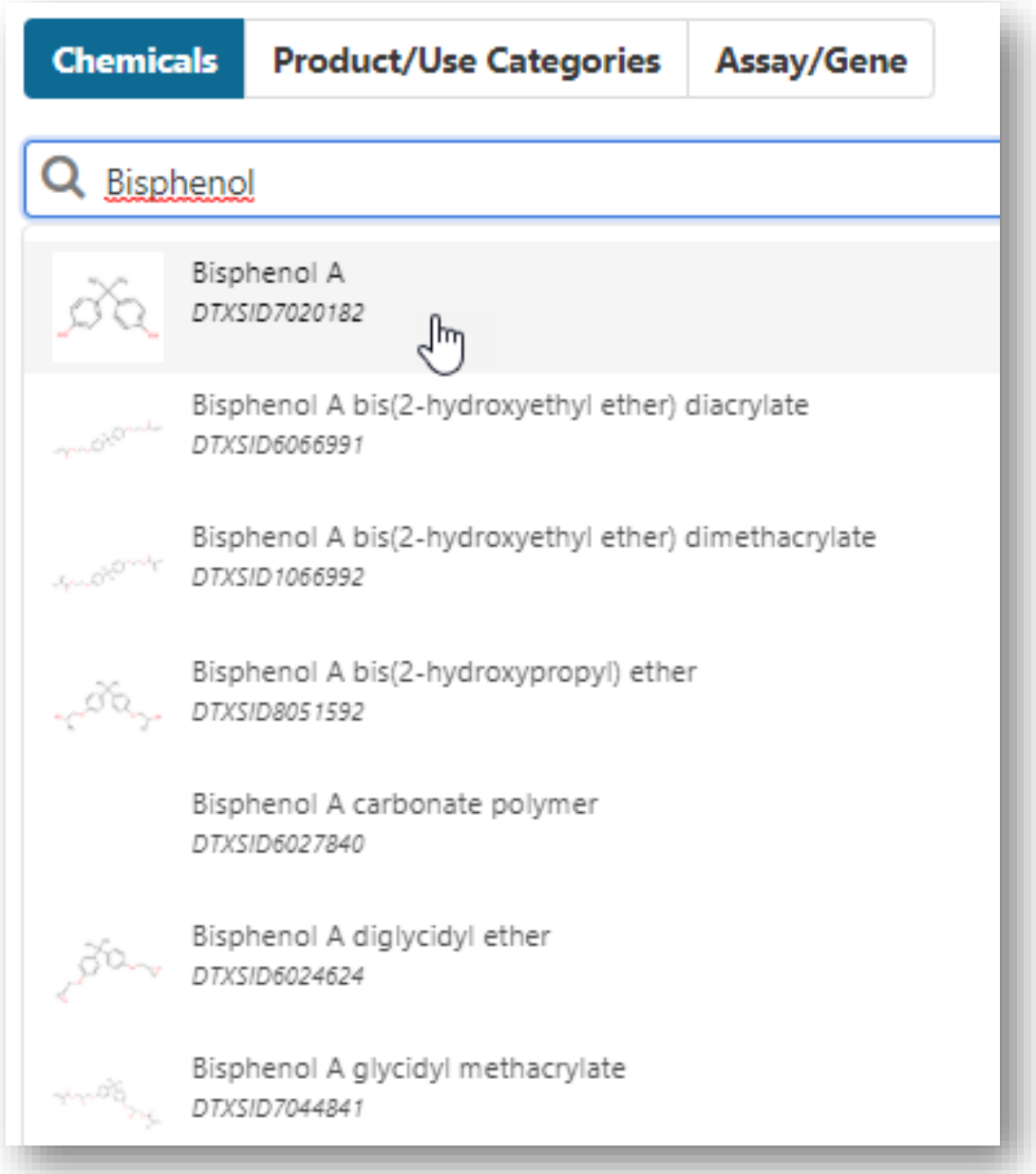
CompTox Chemicals Dashboard

https://comptox.epa.gov/dashboard

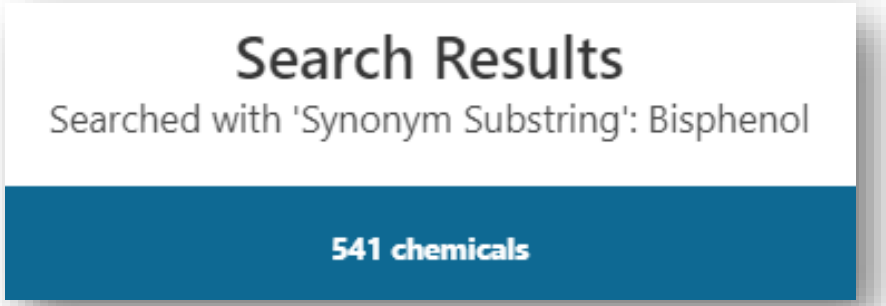
875k Chemical Substances



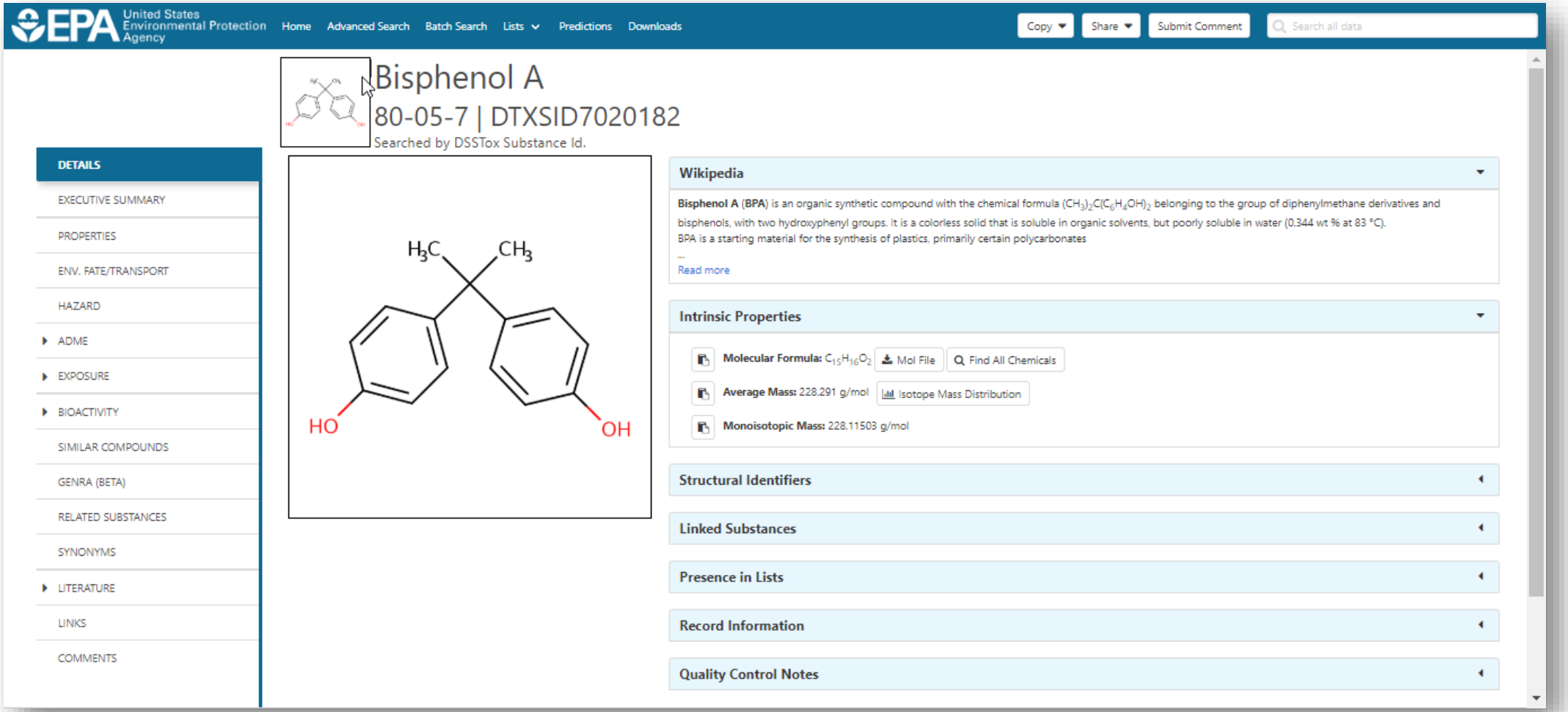
BASIC Search



- Type ahead search using Names, synonyms and CASRNs
- Millions of identifiers
- Substring search

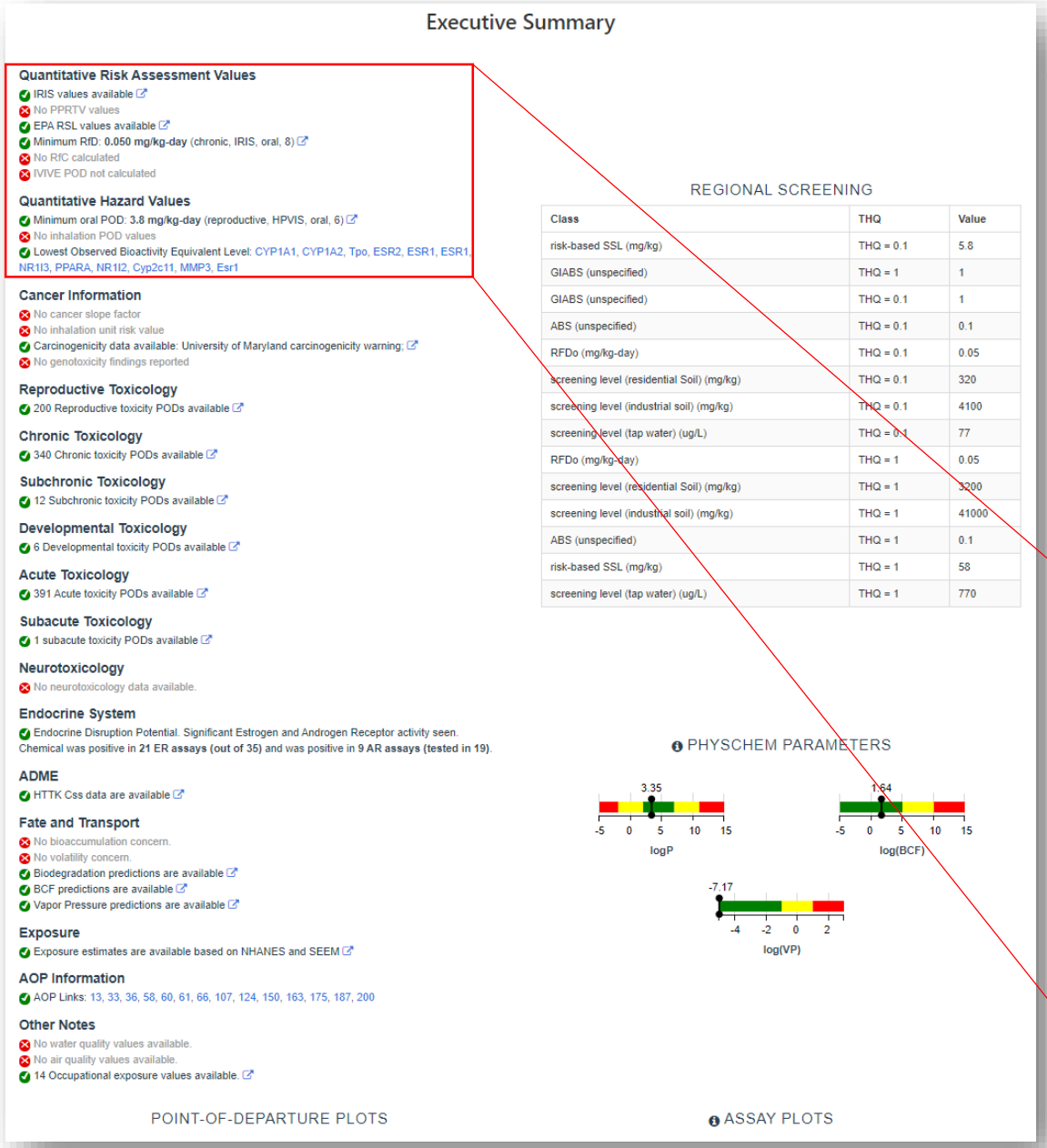


Detailed Chemical Pages



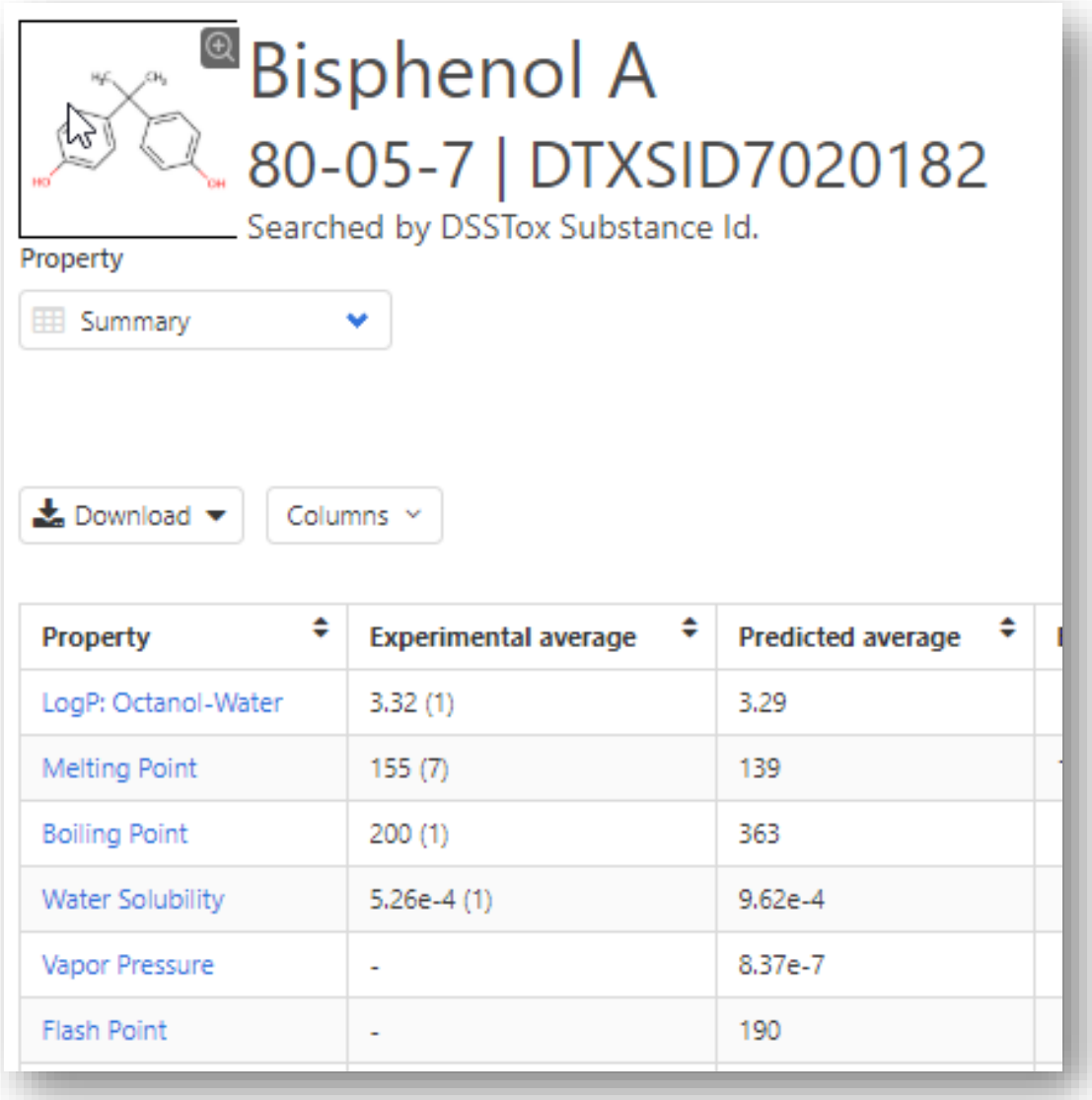
- Chemical page: Wikipedia snippet when available, intrinsic properties, structural identifiers, linked substances

“Executive Summary”



- Overview of toxicity-related info
  - Quantitative values
  - Info re. toxicology subsets
  - Physchem. and Fate & Transport
  - Adverse Outcome Pathway links
  - *In vitro* bioactivity summary plot

Experimental and Predicted Data



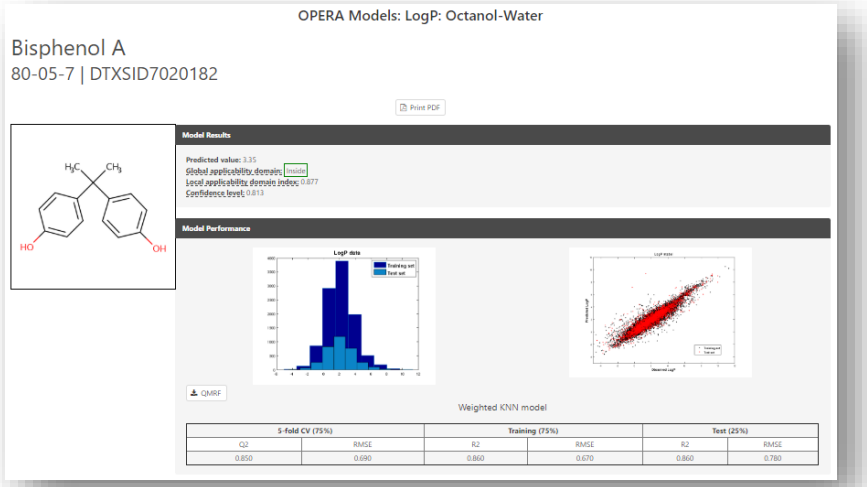
- Physchem and Fate & Transport experimental and predicted data
- Data can be downloaded as Excel, TSV and CSV files
- Predictions: multiple algorithms
  - EPI Suite: Estimation Program Interface
  - ACD/Labs (commercial)
  - TEST: **T**oxicity **E**stimation **S**oftware **T**ool
  - OPERA: **OPE**n structure–activity/ property **R**elationship **A**pp

Transparency for prediction models

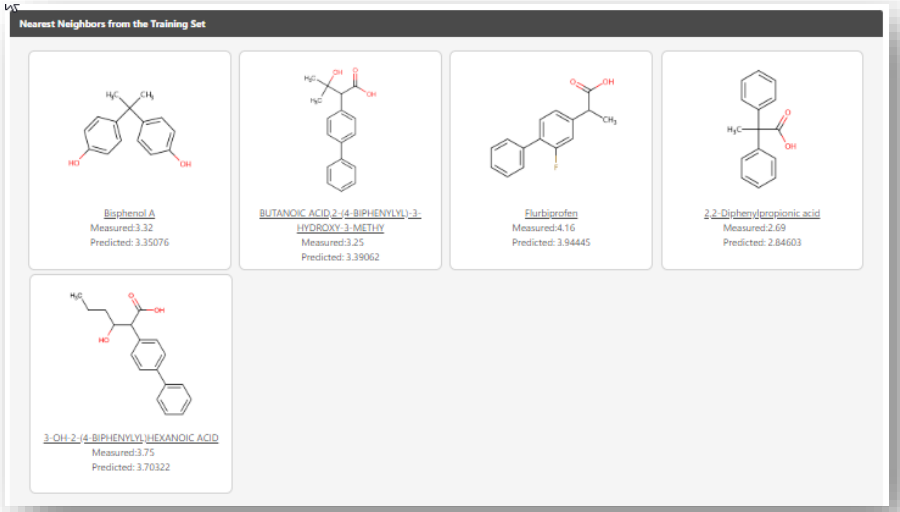
Predicted

Download Predicted Data

Source	Result	Calculation Details	QMRF
EPISUITE	3.64	Not Available	Not Available
NICEATM	2.40	Not Available	Available
ACD/Labs Consensus	3.63	Not Available	Not Available
ACD/Labs	3.43	Not Available	Not Available
OPERA	3.35	OPERA Model Report [Inside AD]	Available

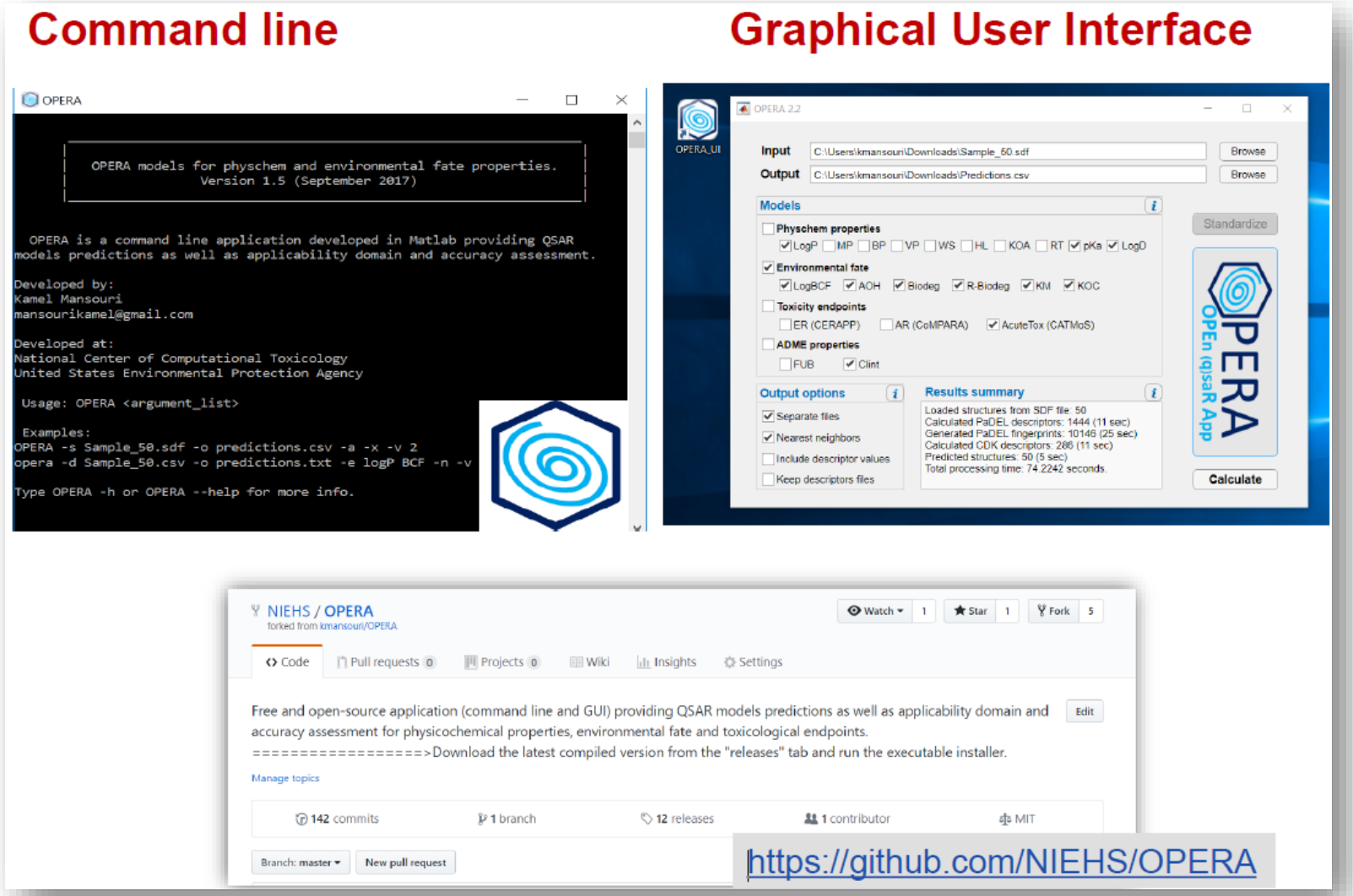


- QMRF – QSAR Model Report Format details
- Applicability Domain
- Plots of expt. vs pred.
- Nearest-neighbors



OPERA Standalone Application

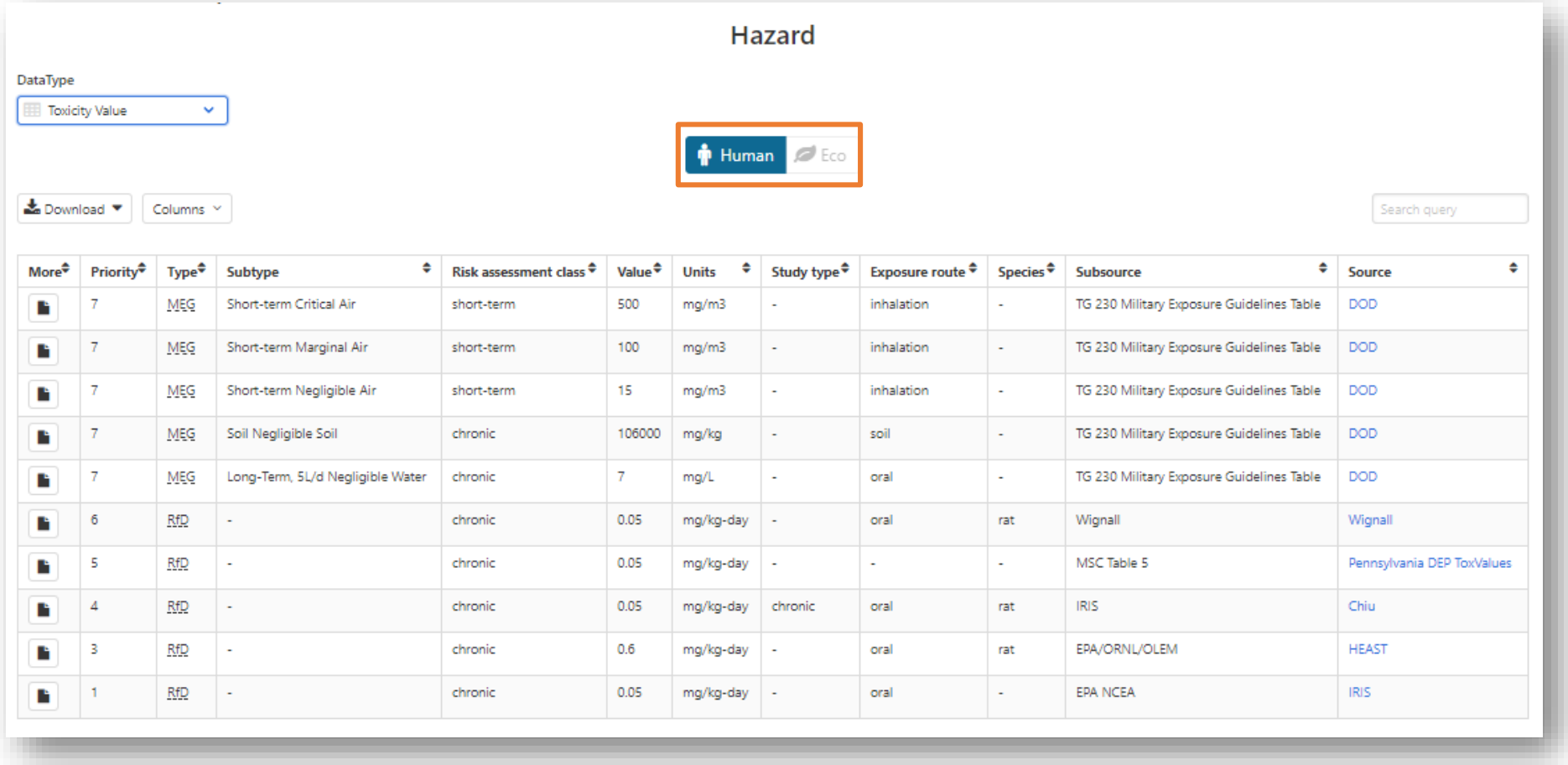
- OPERA predictions available on dashboard
- OPERA application available (from Github)
- Both GUI and command line versions available



Chemical Hazard Data

ToxVal Database

- ~30k chemicals
- >770k tox. values
- ~30 sources of data
- ~22k sub-sources
- ~5k journals cited
- ~70k citations



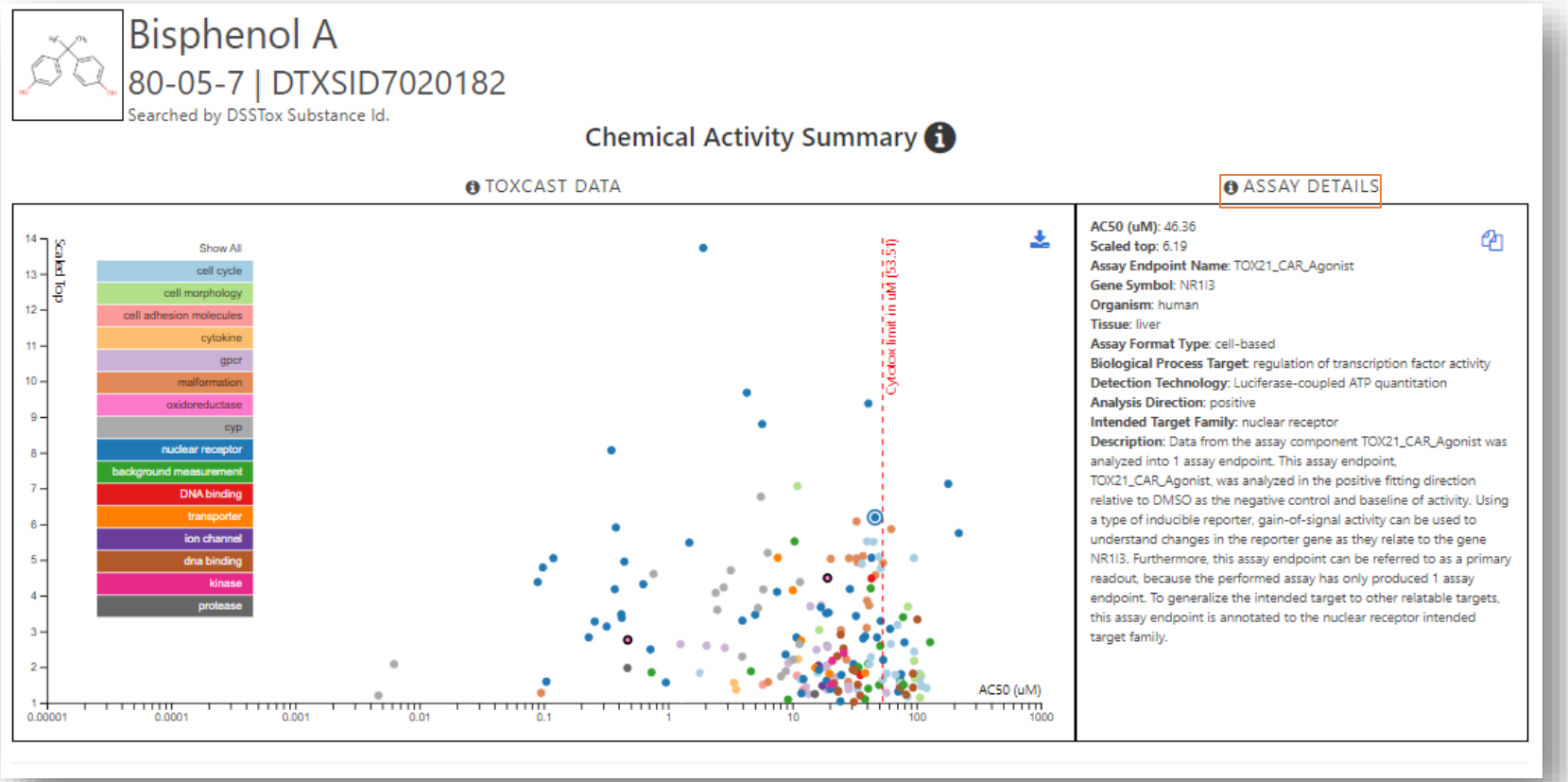
More	Priority	Type	Subtype	Risk assessment class	Value	Units	Study type	Exposure route	Species	Subsource	Source
	7	MEG	Short-term Critical Air	short-term	500	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Short-term Marginal Air	short-term	100	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Short-term Negligible Air	short-term	15	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Soil Negligible Soil	chronic	106000	mg/kg	-	soil	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Long-Term, 5L/d Negligible Water	chronic	7	mg/L	-	oral	-	TG 230 Military Exposure Guidelines Table	DOD
	6	RfD	-	chronic	0.05	mg/kg-day	-	oral	rat	Wignall	Wignall
	5	RfD	-	chronic	0.05	mg/kg-day	-	-	-	MSC Table 5	Pennsylvania DEP ToxValues
	4	RfD	-	chronic	0.05	mg/kg-day	chronic	oral	rat	IRIS	Chiu
	3	RfD	-	chronic	0.6	mg/kg-day	-	oral	rat	EPA/ORNL/OLEM	HEAST
	1	RfD	-	chronic	0.05	mg/kg-day	-	oral	-	EPA NCEA	IRIS

In Vitro Bioassay Screening

ToxCast and Tox21

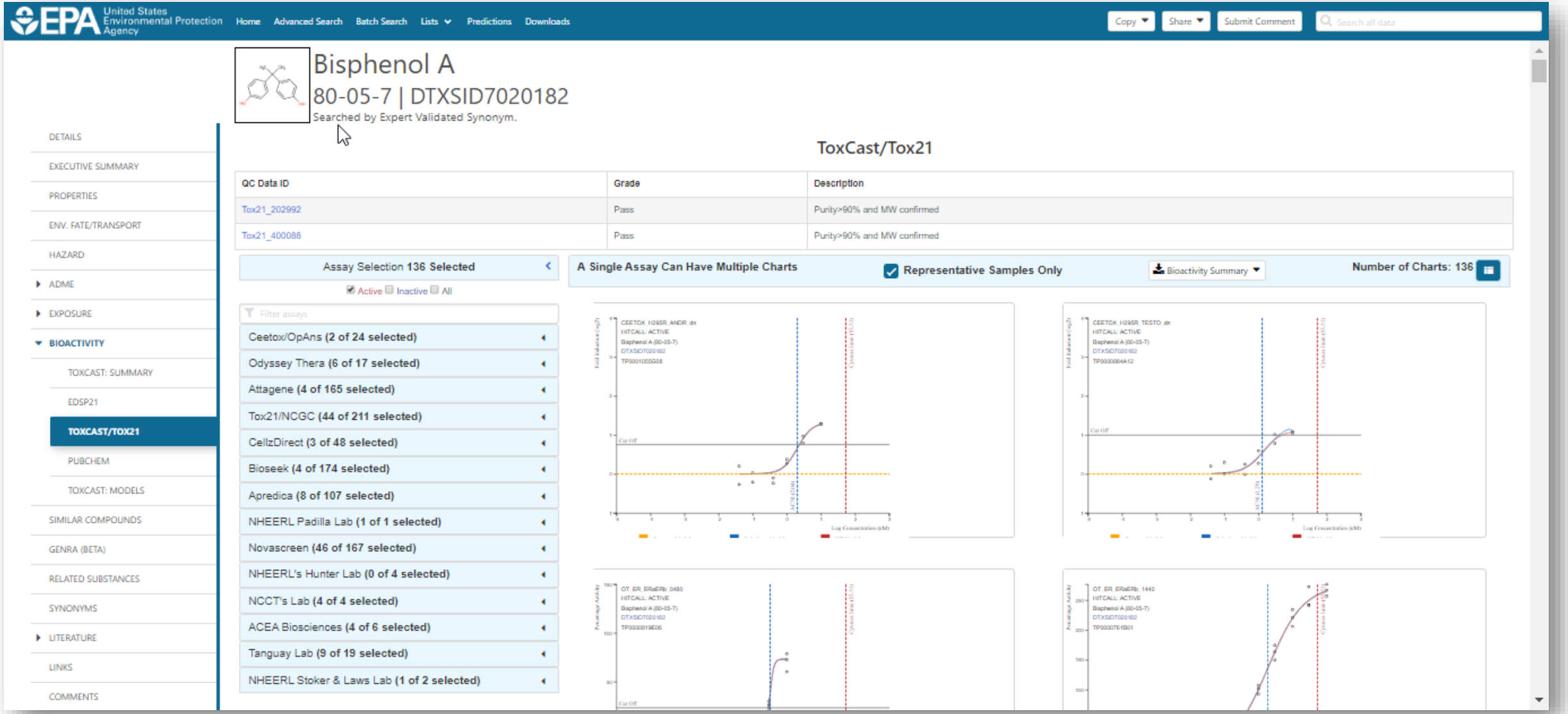
In vitro bioactivity

- 4K chemicals (ToxCast)
- 8k chemicals (Tox21)
- ~2k assay endpoints

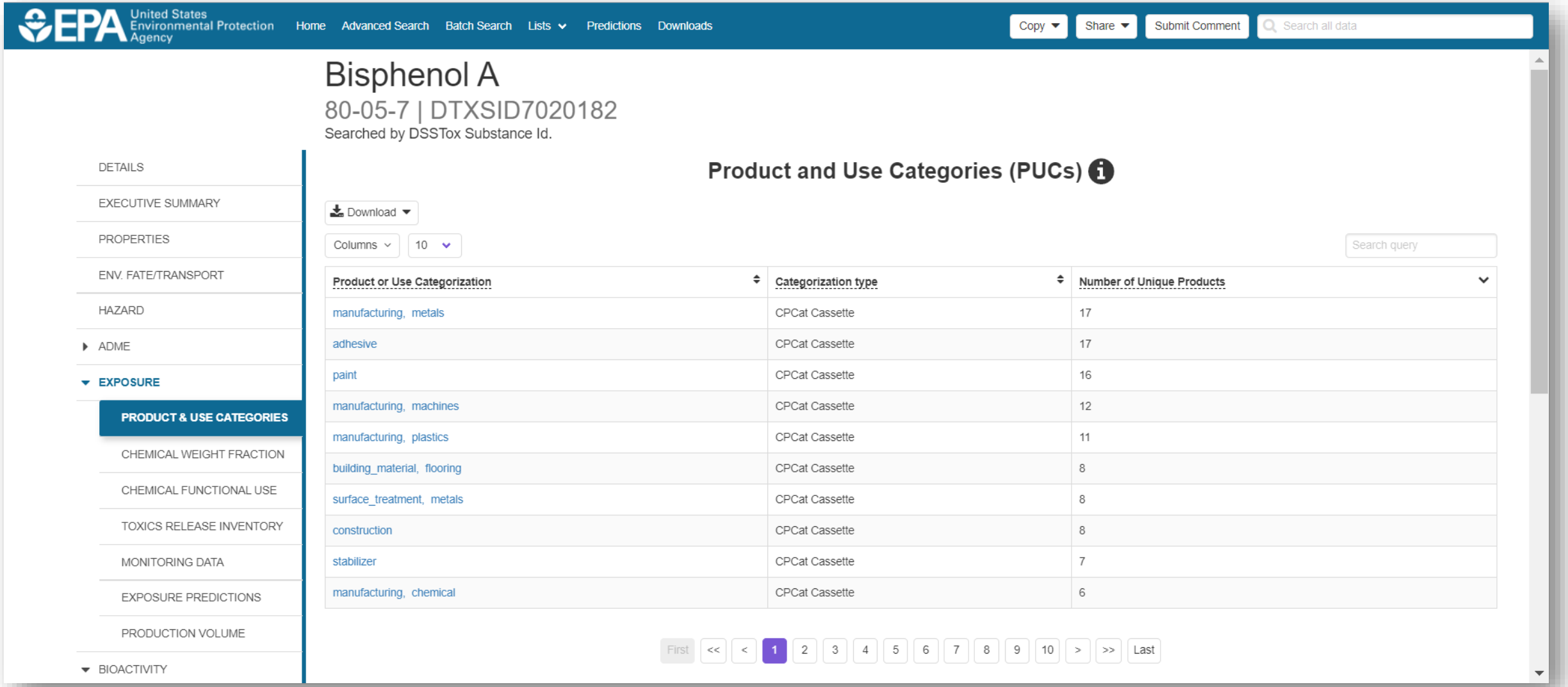


In Vitro Bioassay Screening

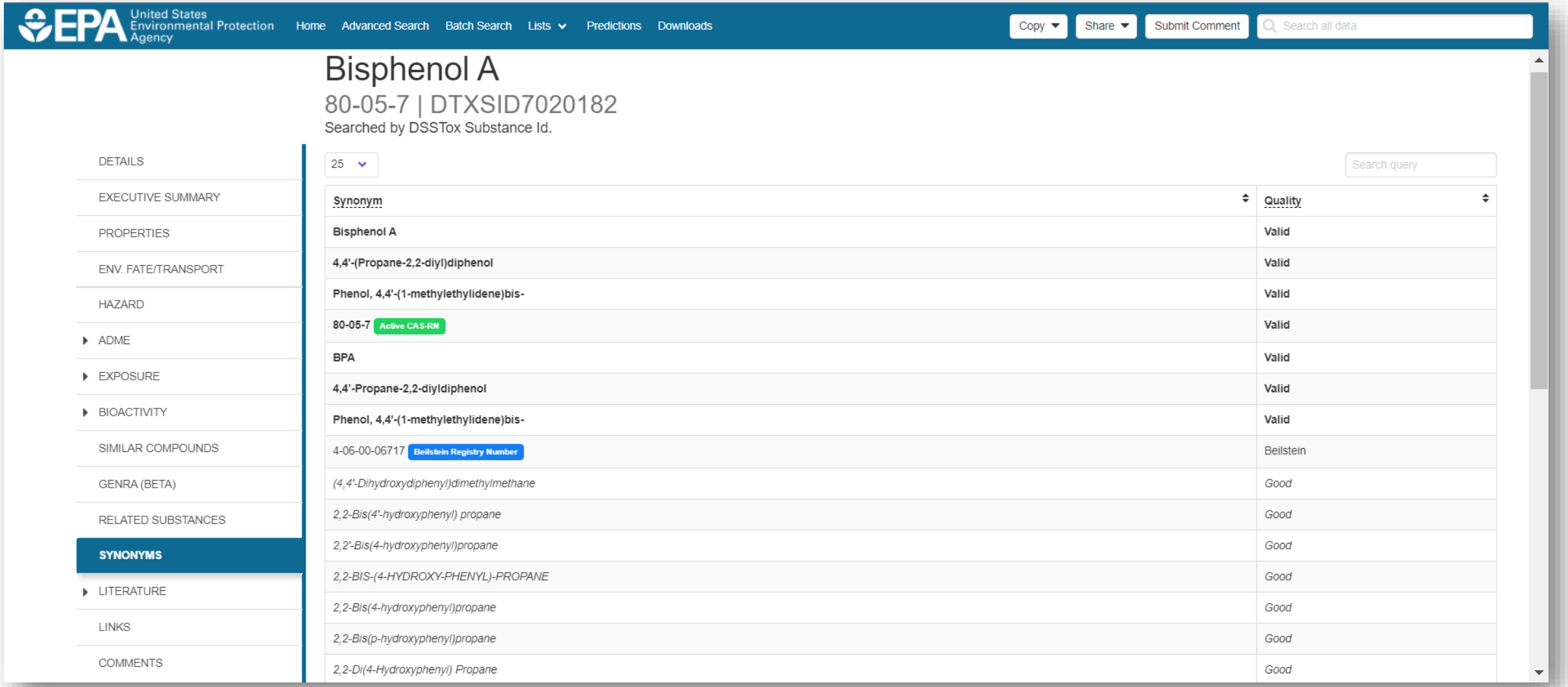
ToxCast and Tox21



Sources of Exposure to Chemicals



Identifiers to Support Searches

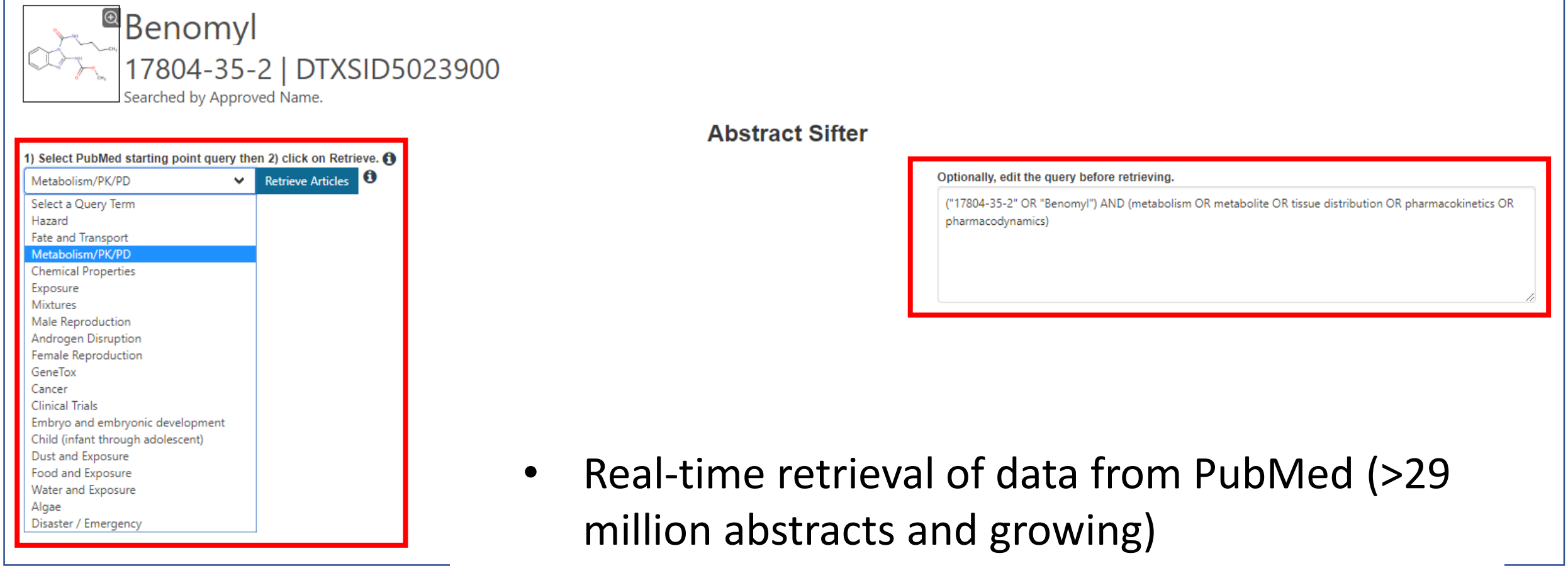


Chemicals Discussed Today

- During talks today we heard about these chemicals. All are on the dashboard.
  - 2,4-dihydroxybenzophenone, 131-56-6 (Mark)
  - 50-00-0, 6898-97-1, 17804-35-2, 1582-09-8, 789-02-6, 50-06-6, 57-30-7, 131-55-5 (Shannon)
  - Argatroban, Lepirudin (Thomas)
  - Arsenic trioxide, chlorpyrifos, cadmium, phorate, butylate, methyl bromide, Diazinon, Fonofos, Atrazine, Dichlorvos, Phorate, Parathion, 2-butenal, pyruvaldehyde, nicotine, formaldehyde, acetaldehyde, acetone, propionaldehyde (Carolyn)

BUILT-IN “MODULES”

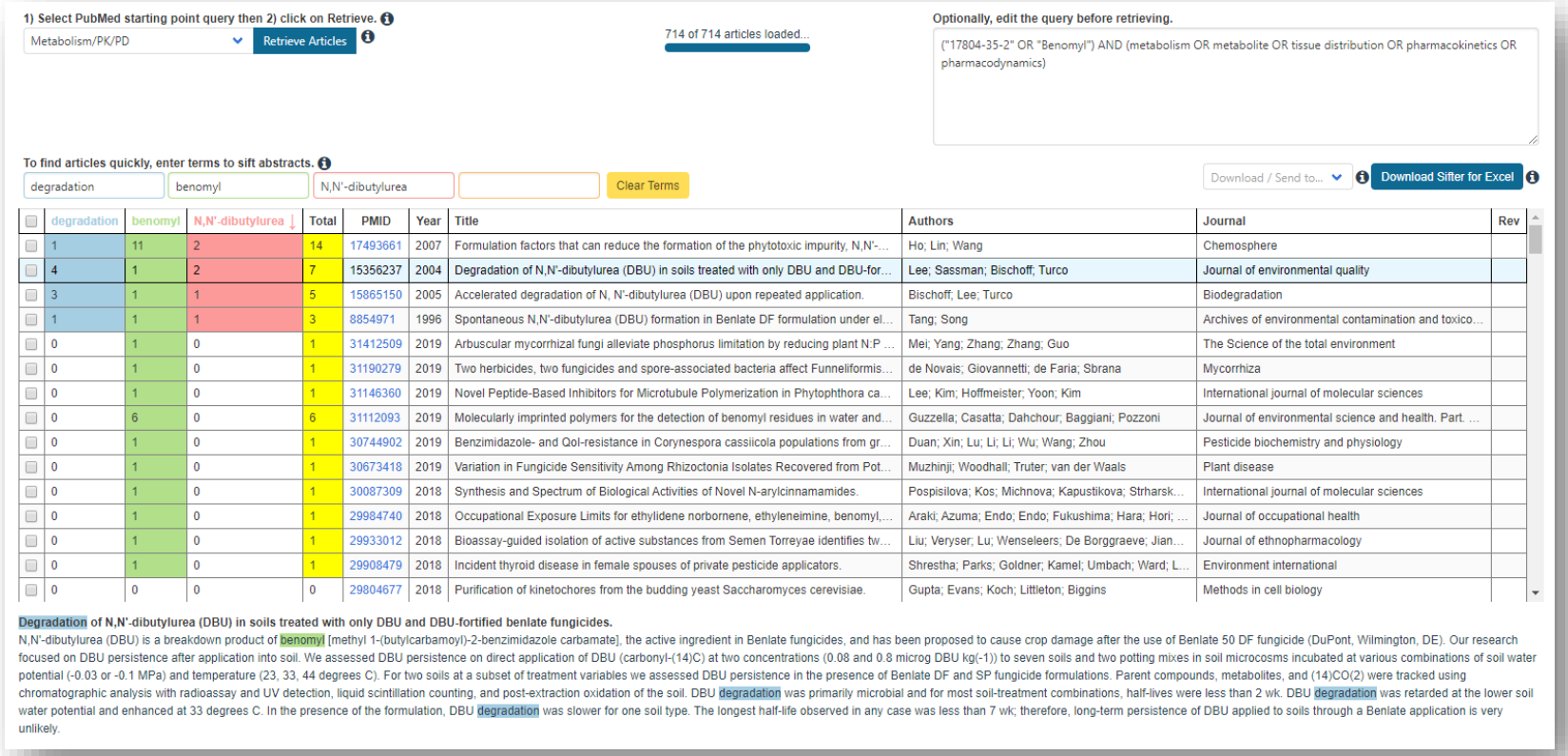
Literature Searching



- Real-time retrieval of data from PubMed (>29 million abstracts and growing)
- Choose from set of pre-defined queries
- Adjust and fine tune queries based on interests

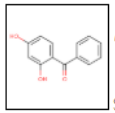
Literature Searching

- “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



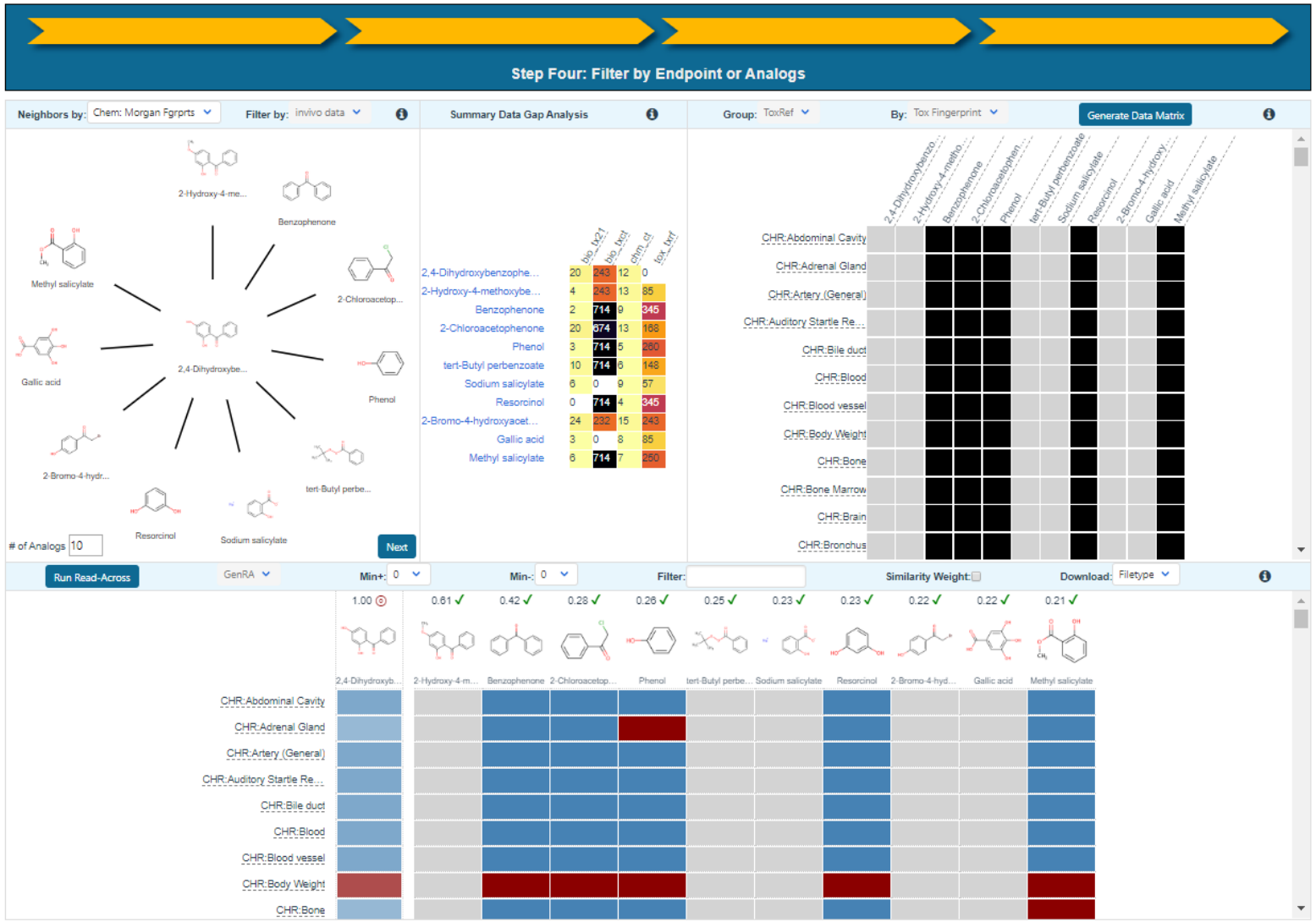
degradation	benomyl	N,N'-dibutylurea	Total	PMID	Year	Title	Authors	Journal	Rev
1	11	2	14	17493661	2007	Formulation factors that can reduce the formation of the phytotoxic impurity, N,N'-...	Ho; Lin; Wang	Chemosphere	
4	1	2	7	15356237	2004	Degradation of N,N'-dibutylurea (DBU) in soils treated with only DBU and DBU-for...	Lee; Sassman; Bischoff; Turco	Journal of environmental quality	
3	1	1	5	15865150	2005	Accelerated degradation of N, N'-dibutylurea (DBU) upon repeated application.	Bischoff; Lee; Turco	Biodegradation	
1	1	1	3	8854971	1996	Spontaneous N,N'-dibutylurea (DBU) formation in Benlate DF formulation under el...	Tang; Song	Archives of environmental contamination and toxico...	
0	1	0	1	31412509	2019	Arbuscular mycorrhizal fungi alleviate phosphorus limitation by reducing plant N:P ...	Mei; Yang; Zhang; Zhang; Guo	The Science of the total environment	
0	1	0	1	31190279	2019	Two herbicides, two fungicides and spore-associated bacteria affect Funneliformis...	de Novais; Giovannetti; de Faria; Sbrana	Mycorrhiza	
0	1	0	1	31146360	2019	Novel Peptide-Based Inhibitors for Microtubule Polymerization in Phytophthora ca...	Lee; Kim; Hoffmeister; Yoon; Kim	International journal of molecular sciences	
0	6	0	6	31112093	2019	Molecularly imprinted polymers for the detection of benomyl residues in water and...	Guzzella; Casatta; Dahchour; Baggiani; Pozzoni	Journal of environmental science and health. Part. ...	
0	1	0	1	30744902	2019	Benzimidazole- and QoI-resistance in Corynespora cassiicola populations from gr...	Duan; Xin; Lu; Li; Li; Wu; Wang; Zhou	Pesticide biochemistry and physiology	
0	1	0	1	30673418	2019	Variation in Fungicide Sensitivity Among Rhizoctonia Isolates Recovered from Pot...	Muzhinji; Woodhall; Truter; van der Waals	Plant disease	
0	1	0	1	30087309	2018	Synthesis and Spectrum of Biological Activities of Novel N-arylcinnamamides.	Pospisilova; Kos; Michnova; Kapustikova; Strharsk...	International journal of molecular sciences	
0	1	0	1	29984740	2018	Occupational Exposure Limits for ethylidene norbornene, ethyleneimine, benomyl,...	Araki; Azuma; Endo; Endo; Fukushima; Hara; Hori; ...	Journal of occupational health	
0	1	0	1	29933012	2018	Bioassay-guided isolation of active substances from Semen Torreyae identifies tw...	Liu; Veryser; Lu; Wenseleers; De Borggraeve; Jian...	Journal of ethnopharmacology	
0	1	0	1	29908479	2018	Incident thyroid disease in female spouses of private pesticide applicators.	Shrestha; Parks; Goldner; Kamel; Umbach; Ward; L...	Environment international	
0	0	0	0	29804677	2018	Purification of kinetochores from the budding yeast Saccharomyces cerevisiae.	Gupta; Evans; Koch; Littleton; Biggins	Methods in cell biology	

Generalized Read-Across (GenRA)

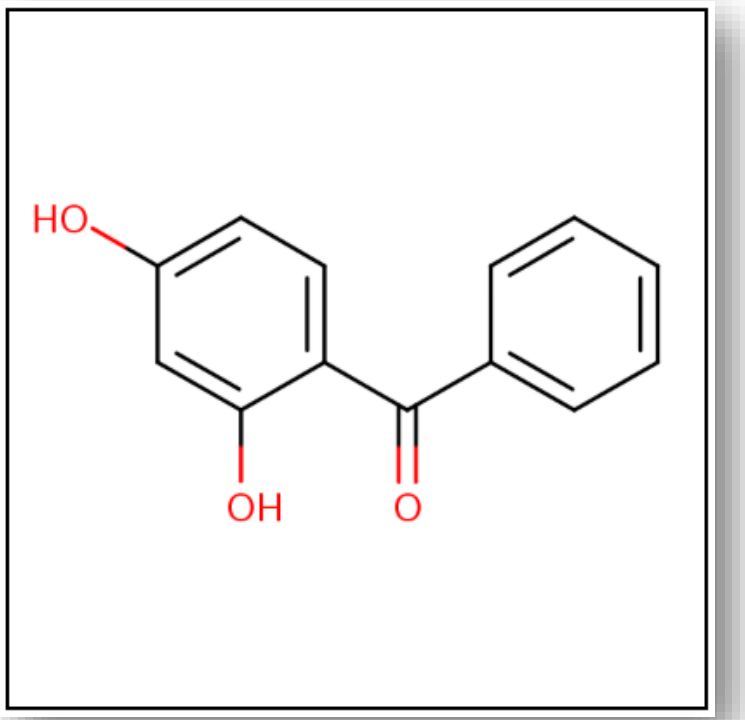


2,4-Dihydroxybenzophenone
131-56-6 | DTXSID8022406
Searched by DSSTox Substance Id.

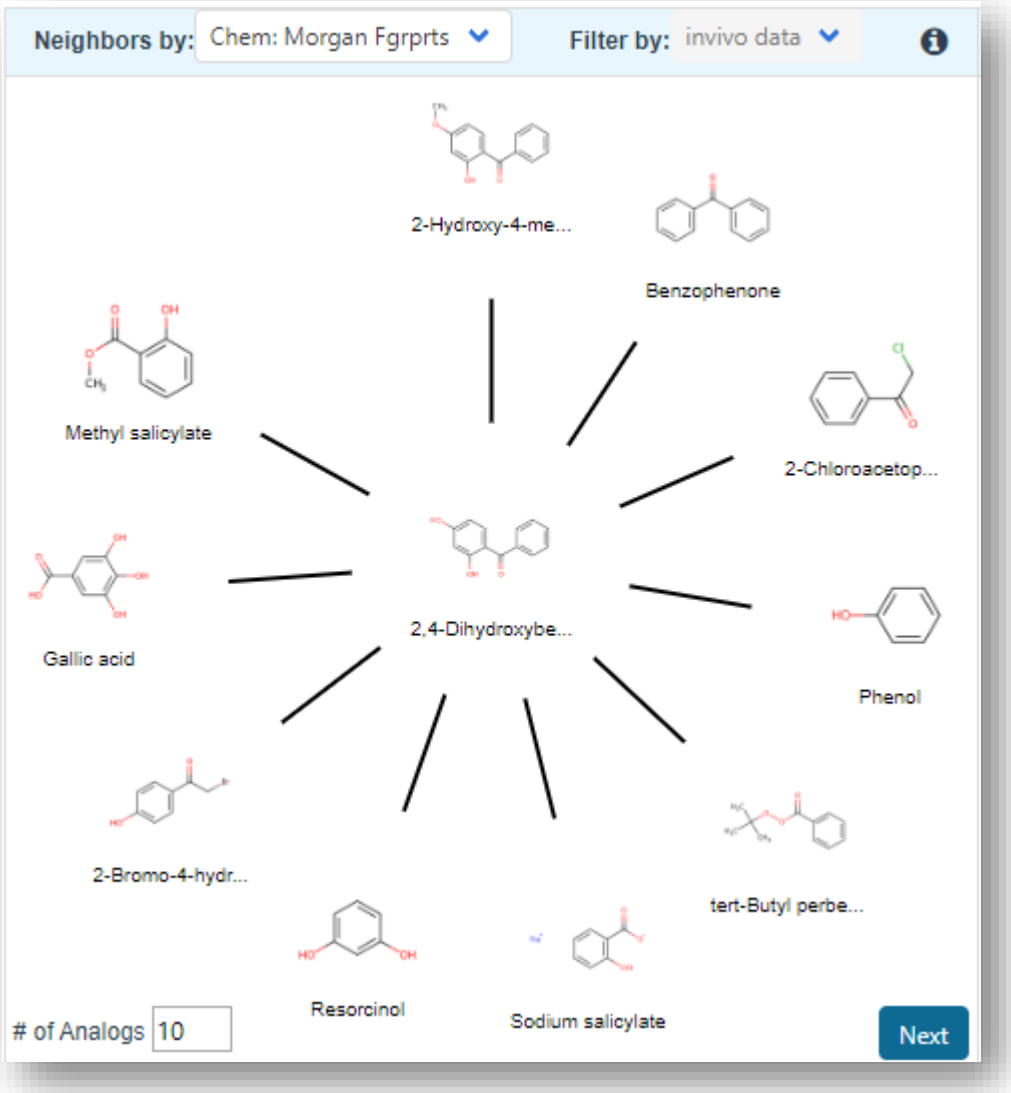
- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- ADME
- EXPOSURE
- BIOACTIVITY
- SIMILAR COMPOUNDS
- **GENRA (BETA)**
- RELATED SUBSTANCES
- SYNONYMS
- LITERATURE
- LINKS
- COMMENTS



GenRA in practice – step by step

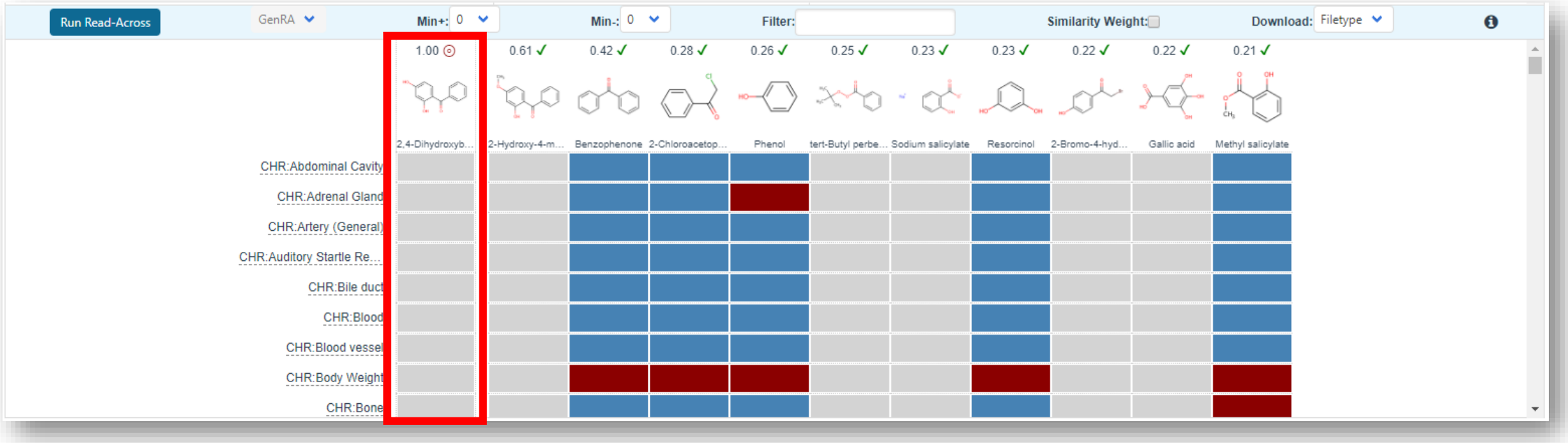


- Analogue identification: based on Morgan fingerprints and selecting 10 default analogues



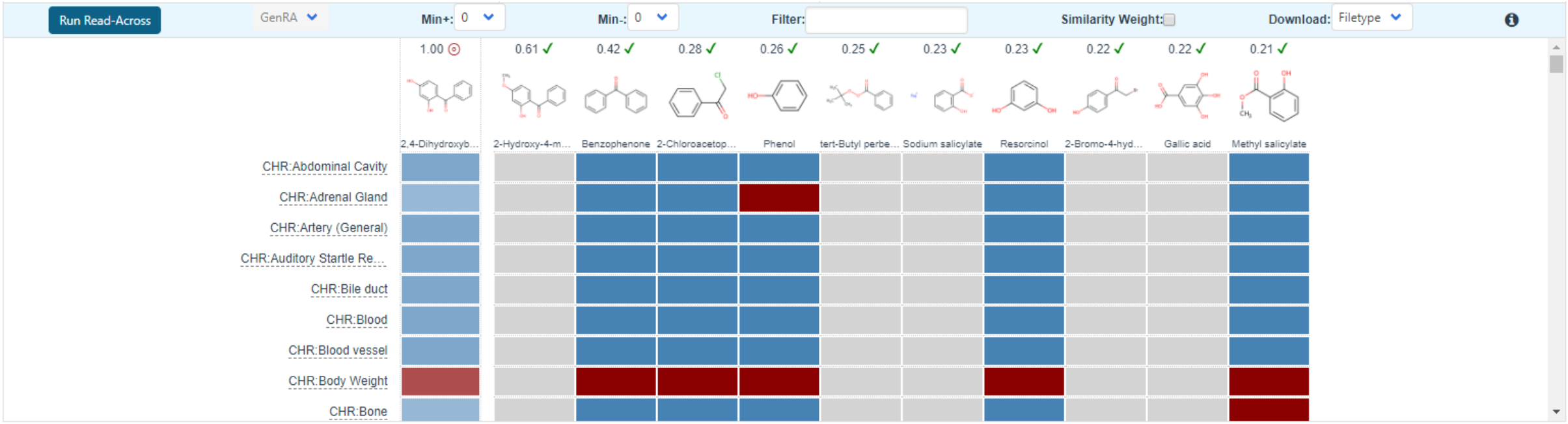
GenRA in practice – step by step

- Data matrix view of source analogues relative to target chemical



GenRA in practice – step by step

- Updated Data matrix view with GenRA predictions for target chemical



- Predictions are binary (yes/no) for toxicity effects
- Predictions summarized on study level basis. Red: “positive” and Blue: “negative”.

Related Publications

Chemical Research in Toxicology

Article

Cite This: Chem. Res. Toxicol. 2017, 30, 2046-2059 pubs.acs.org/crt

Predicting Organ Toxicity Using *in Vitro* Bioactivity Data and Chemical Structure

Jie Liu,[‡,§] Grace Patlewicz,[†] Antony J. Williams,[†] Russell S. Thomas,[†] and Imran Shah[*,†]

[†]National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, Durham, North Carolina 27711, United States

[‡]Department of Information Science, University of Arkansas at Little Rock, Arkansas 72204, United States

[§]Oak Ridge Institute for Science Education, National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, Durham, North Carolina 27711, United States

ELSEVIER

Regulatory Toxicology and Pharmacology

Volume 79, August 2016, Pages 12-24

Systematically evaluating read-across prediction and performance using a local validity approach characterized by chemical structure and bioactivity information

Imran Shah [a], Jie Liu [b, c], Richard S. Judson [a], Russell S. Thomas [a], Grace Patlewicz [a]

ELSEVIER

Computational Toxicology

Available online 23 July 2018

In Press, Corrected Proof

Extending the Generalised Read-Across approach (GenRA): A systematic analysis of the impact of physicochemical property information on read-across performance

George Helman [a, b], Imran Shah [b], Grace Patlewicz [b]

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.elsevier.com

ELSEVIER

Journal Cover Image

Navigating through the minefield of read-across frameworks: A commentary perspective

Grace Patlewicz[a, *], Mark T.D. Cronin[b], George Helman[a, c], Jason C. Lambert[d], Lucina E. Lizarraga[d], Imran Shah[a]

[a] National Center for Computational Toxicology (NCCT), Office of Research and Development, US Environmental Protection Agency (US EPA), 109 TW Alexander Dr, Research Triangle Park (RTP), NC 27711, USA

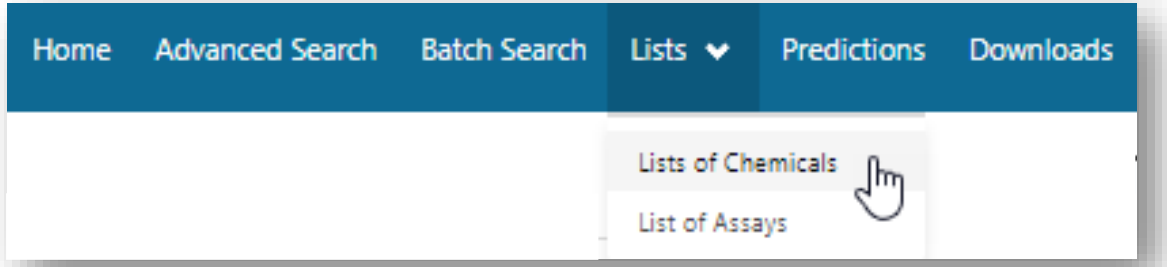
[b] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK

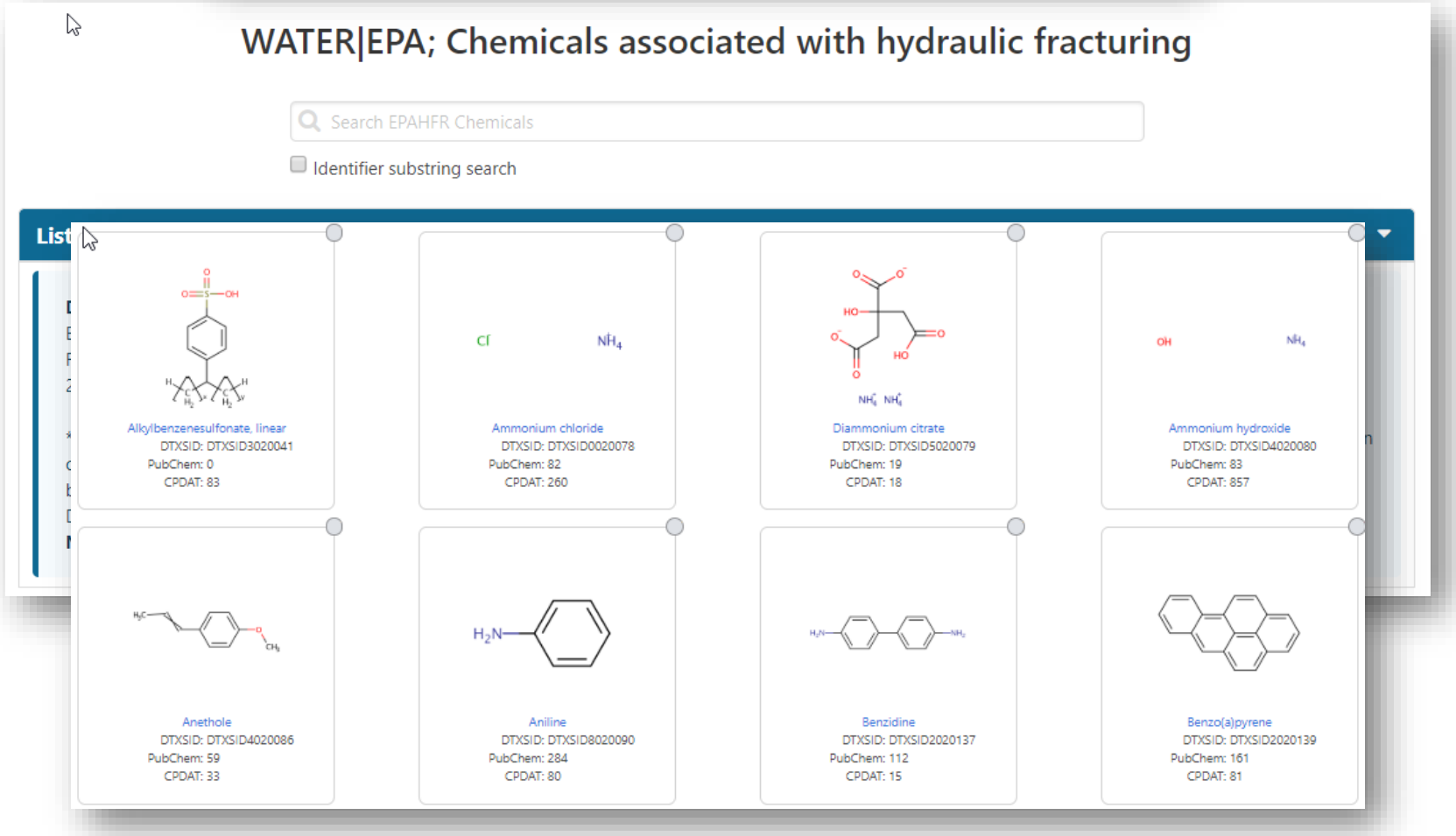
[c] Oak Ridge Institute for Science and Education (ORISE), 1299 Bethel Valley Road, Oak Ridge, TN 37830, USA

[d] National Center for Evaluation Assessment (NCEA), US Environmental Protection Agency (US EPA), 26 West Martin Luther King Dr, Cincinnati, OH 45268, USA

CHEMICAL LISTS AND CATEGORIES

EPAHFR: Hydraulic Fracturing





PFAS lists of Chemicals

Select List

Download | Columns | PFAS | Copy Filtered Lists URL

List Acronym	List Name	Last Updated	Number of Chemicals	List Description
EPAPFAS75S1	PFAS|EPA: List of 75 Test Samples (Set 1)	2018-06-29	74	PFAS list corresponds to 75 samples (Set 1) submitted for initial testing screens conducted by EPA researchers in collaboration with researchers at the National Toxicology Program.
EPAPFAS75S2	PFAS|EPA: List of 75 Test Samples (Set 2)	2019-02-21	75	PFAS list corresponds to a second set of 75 samples (Set 2) submitted for testing screens conducted by EPA researchers in collaboration with researchers at the National Toxicology Program.
EPAPFASCAT	PFAS|EPA Structure-based Categories	2018-06-29	64	List of registered DSSTox "category substances" representing PFAS categories created using ChemAxon's Markush structure-based query representations.
EPAPFASINSOL	PFAS|EPA: Chemical Inventory Insoluble in DMSO	2018-06-29	43	PFAS chemicals included in EPA's expanded ToxCast chemical inventory found to be insoluble in DMSO above 5mM.
EPAPFASINV	PFAS|EPA: ToxCast Chemical Inventory	2018-06-29	430	PFAS chemicals included in EPA's expanded ToxCast chemical inventory and available for testing.
EPAPFASRL	PFAS|EPA: Cross-Agency Research List	2017-11-16	199	EPAPFASRL is a manually curated listing of mainly straight-chain and branched PFAS (Per- & Poly-fluorinated alkyl substances) compiled from various internal, literature and public sources by EPA researchers and program office representatives.
PFASKEMI	PFAS: List from the Swedish Chemicals Agency (KEMI) Report	2017-02-09	2416	Perfluorinated substances from a Swedish Chemicals Agency (KEMI) Report on the occurrence and use of highly fluorinated substances.
PFASMASTER	PFAS Master List of PFAS Substances	2018-07-26	5061	PFASMASTER is a consolidated list of PFAS substances spanning and bounded by the below lists of current interest to researchers and regulators worldwide.
PFASOECD	PFAS: Listed in OECD Global Database	2018-05-16	4729	OECD released a New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances, (PFASs) listing more than 4700 new PFAS
PFASTRIER	PFAS Community-Compiled List (Trier et al., 2015)	2017-07-16	597	PFASTRIER community-compiled public listing of PFAS (Trier et al, 2015)

List of Assays

Assay List

Download | 10 | ATG × | OT × | NHEERL_PADILLA × | Search query | Copy page URL

- ACEA Biosciences
- Apredica
- Attagene
- Bioseek
- Novascreen
- Odyssey Thera
- Tox21/NCGC
- Ceetox/OpAns
- CellzDirect
- NHEERL Padilla Lab
- NCCT's Lab
- Tanguay Lab
- NHEERL's Hunter Lab
- NHEERL Stoker & Laws Lab
- NHEERL Med Lab

Assay Component Endpoint Name	Details	Active Hits	Description	Gene Symbols
ATG_Ahr_CIS_up		573 / 3807	Data from the assay component ATG_A…assay endpoint. This assay endpoint, ATG_Ahr_CIS_up, was analyz…ction relative to DMSO as the negative control and baseline of activit…reporter, measures of mRNA for gain-of-signal activity can be used to u…e at the transcription factor-level as they relate to the gene AHR. Furtherm…be referred to as a primary readout, because this assay has produ…s where this one serves a reporter gene function. To generalize the intend…targets, this assay endpoint is annotated to the "dna binding" intend…subfamily is "basic helix-loop-helix protein".	AHR
ATG_AP_1_CIS_up		658 / 3807	Data from the assay component ATG_A…assay endpoint. This assay endpoint, ATG_AP_1_CIS_up, was analy…rection relative to DMSO as the negative control and baseline of activit…reporter, measures of mRNA for gain-of-signal activity can be used to u…e at the transcription factor-level as they relate to the gene FOS and JUN. F…oint can be referred to as a primary readout, because this assay has produ…s where this one serves a reporter gene function. To generalize the intend…targets, this assay endpoint is annotated to the "dna binding" intended target family, where the subfamily is "basic leucine zipper".	FOS JUN
ATG_AP_2_CIS_up		47 / 3807	Data from the assay component ATG_AP_2_CIS was analyzed into 1 assay endpoint. This assay endpoint, ATG_AP_2_CIS_up, was analyzed in the positive fitting direction relative to DMSO as the negative control and baseline of activity. Using a type of inducible reporter, measures of mRNA for	TFAP2A TFAP2B TFAP2D

From Assay to Chemicals…

Assay Endpoint Name: ACEA_ER_80hr

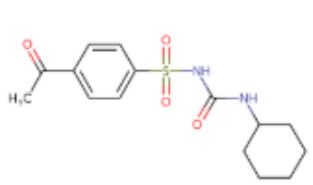
Assay Details

Assay Endpoint Name: ACEA_ER_80hr

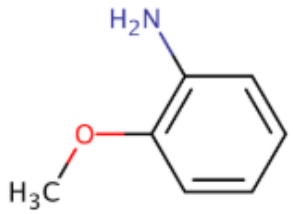
Assay Source Description: ACEA Biosciences, Inc. (ACEA) is a privately owned biotechnology company that developed a real-time, label-free, cell-based assay system based on a microelectronic readout called xCELLigence.

425 of 3031 chemicals visible

Select all | Download | Send to Batch Search | Default | CASRN | TOXCAST | PubMed | Inactive | Filter by Name or CASRN

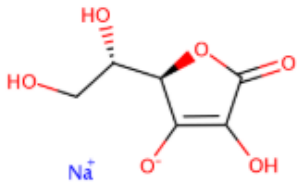


Acetohexamide
CASRN: 968-81-0
TOXCAST: 4/376
PubMed: 235



HCl

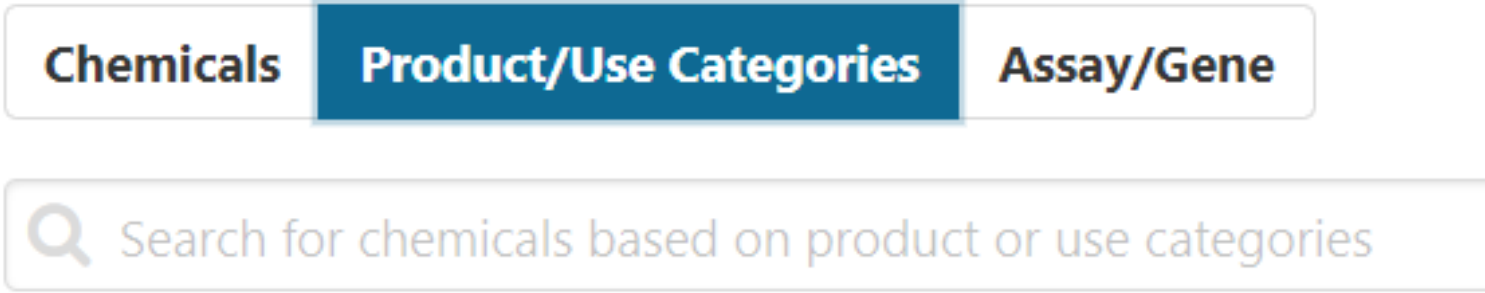
2-Methoxyaniline hydrochloride
CASRN: 134-29-2
TOXCAST: 17/385
PubMed: 46



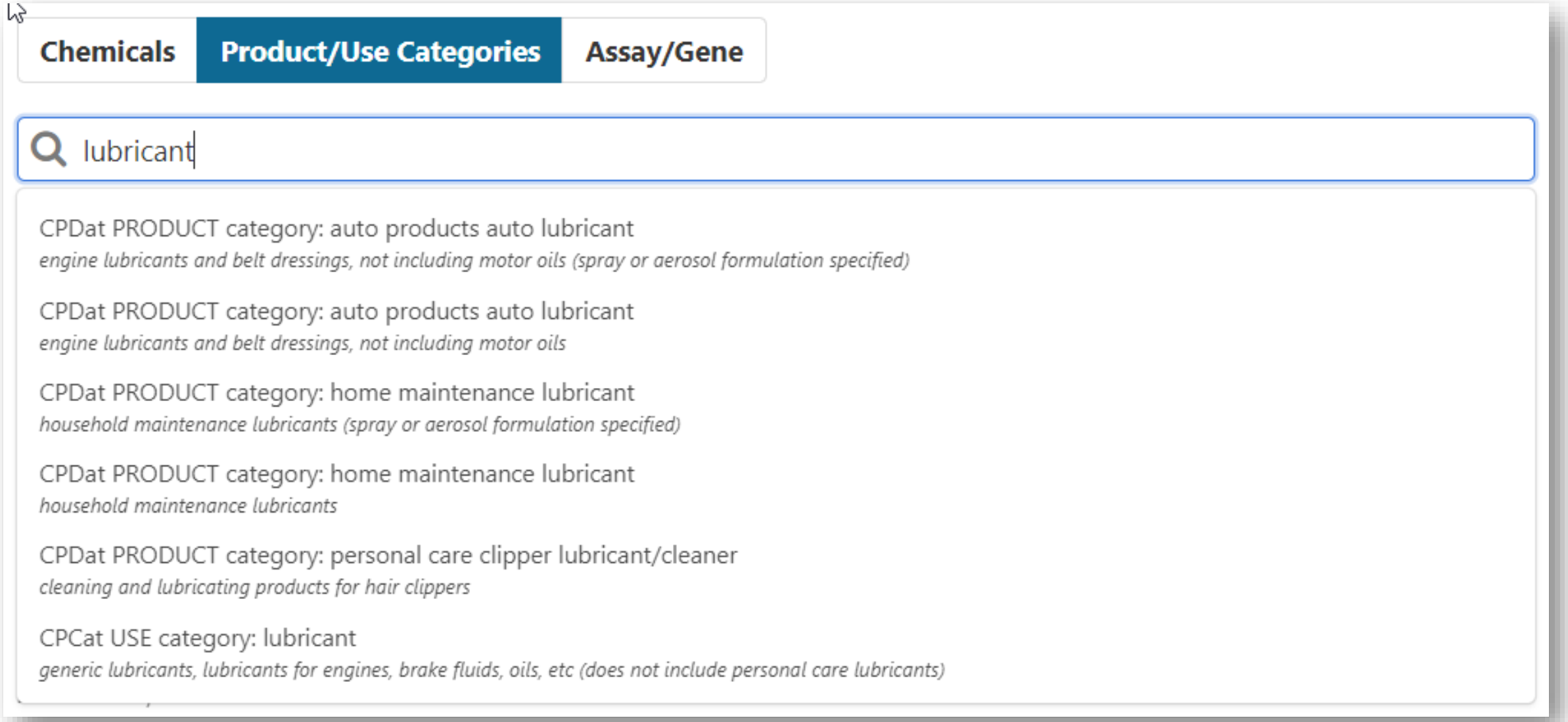
Sodium L-ascorbate
CASRN: 134-03-2
TOXCAST: 20/662
PubMed: 38756

Sodium azide
CASRN: 26628-22-8
TOXCAST: 26/644
PubMed: 1866

Other Searches



Product/Use Categories



Lubricant

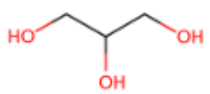
Searched by Product & Use Categories

Results for CPCat Use Category: CPCat term: Lubricant

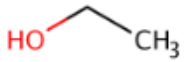
1738 chemicals

Select all | Download | Send to Batch Search | CPDAT | DTXSID | PubChem | CPDAT | Hide chemicals that are: | Filter by Name or CASRI

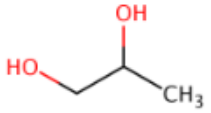
Water
DTXSID: DTXSID6026296
PubChem: 637
CPDAT: 8270



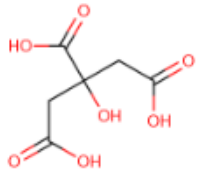
Glycerol
DTXSID: DTXSID9020663
PubChem: 13028
CPDAT: 4249



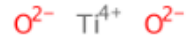
Ethanol
DTXSID: DTXSID9020584
PubChem: 1591
CPDAT: 3635



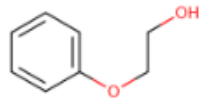
1,2-Propylene glycol
DTXSID: DTXSID0021206
PubChem: 472
CPDAT: 3193



Citric acid
DTXSID: DTXSID3020332
PubChem: 1516
CPDAT: 2560

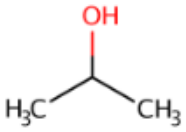


Titanium dioxide
DTXSID: DTXSID3021352
PubChem: 103
CPDAT: 2519



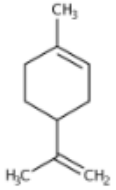
2-Phenoxyethanol
DTXSID: DTXSID9021976
PubChem: 140
CPDAT: 2314

Sodium chloride
DTXSID: DTXSID3021271
PubChem: 203
CPDAT: 2197

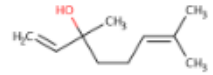


Isopropanol
DTXSID: DTXSID7020762
PubChem: 1267
CPDAT: 2176

Sodium hydroxide
DTXSID: DTXSID0029634
PubChem: 203
CPDAT: 2085

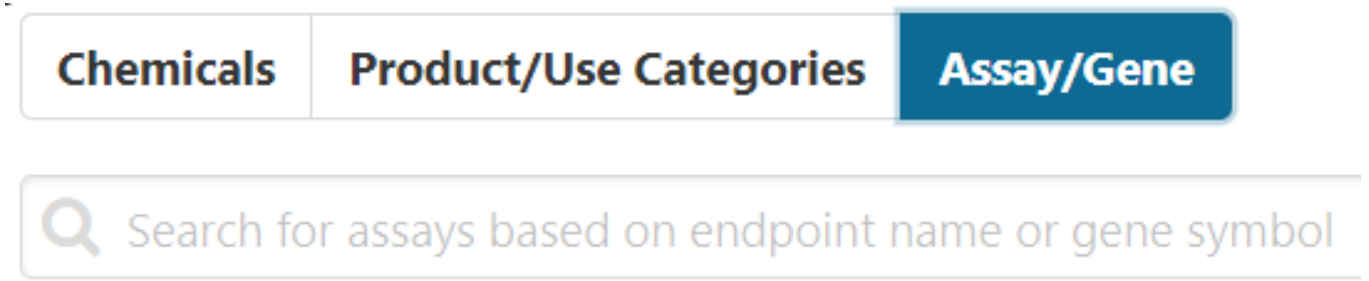


Limonene
DTXSID: DTXSID2029612
PubChem: 123
CPDAT: 2068



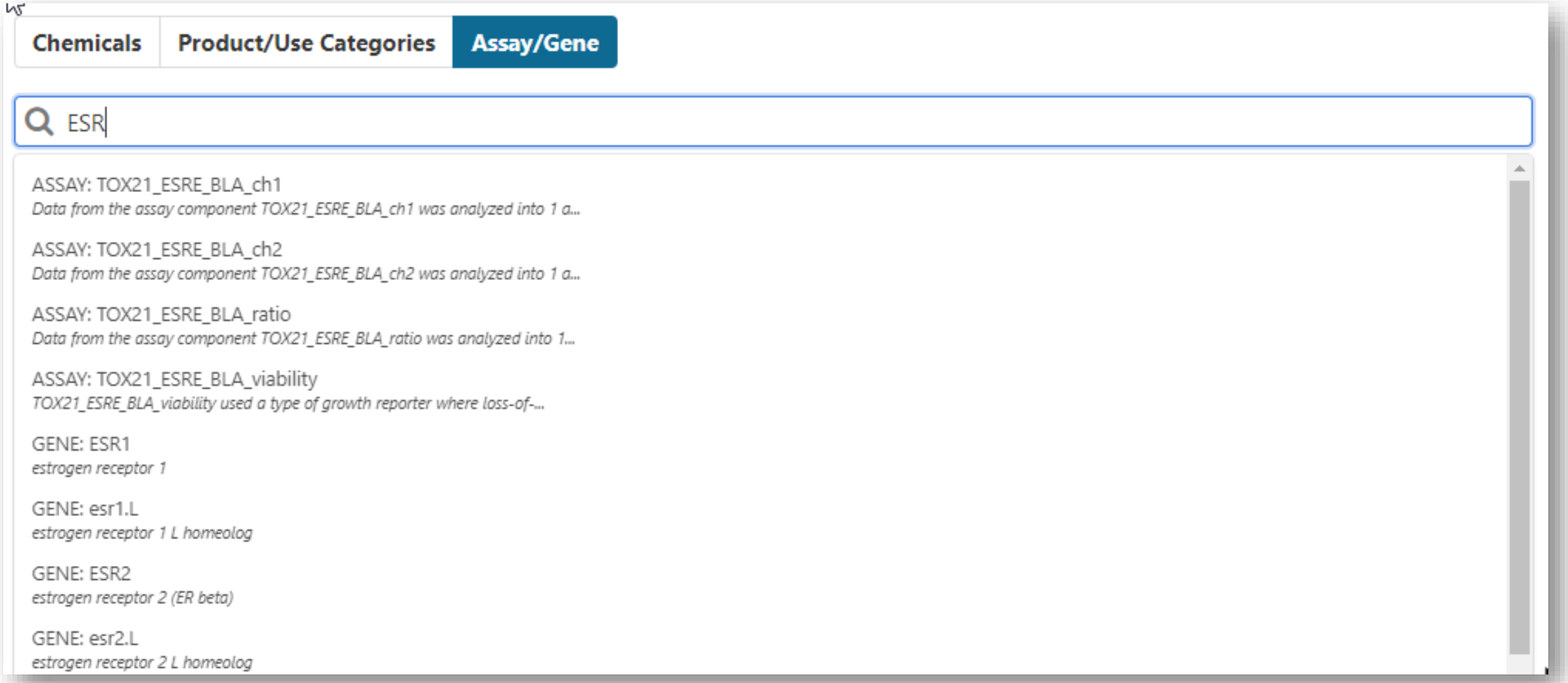
Linalool
DTXSID: DTXSID7025502
PubChem: 118
CPDAT: 1984

Other Searches



Chemical-Biology

Assay/Gene Search



Assay/Gene Search

Assay List

Download | 10 | Filter by vendor | ESR1 | Copy Filtered Lists URL

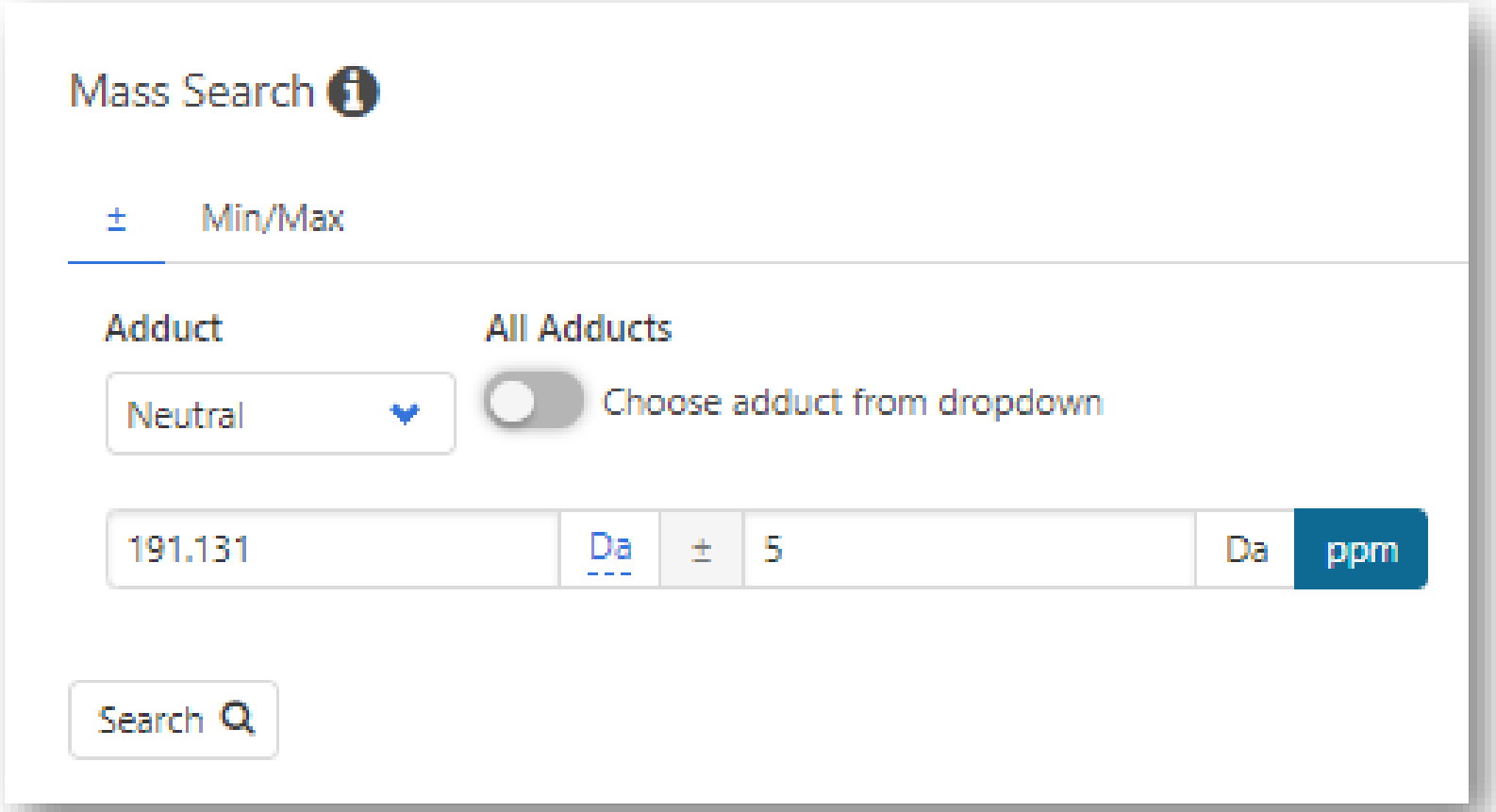
Assay Component Endpoint Name	Details	Active Hits	Description	Gene Symbols
ACEA_ER_80hr		425 / 3031	Data from the assay component ACEA_ER_80hr was analyzed into 2 assay endpoints. This assay endpoint, ACEA_ER_80hr_Positive, was analyzed in the positive fitting direction relative to DMSO as the negative control and baseline of activity. Using a type of growth reporter, measures of the cells for gain-of-signal activity can be used to understand the signaling at the pathway-level as they relate to the gene ESR1. Furthermore, this assay endpoint can be referred to as a primary readout, because this assay has produced multiple assay endpoints where this one serves a signaling function. To generalize the intended target to other relatable targets, this assay endpoint is annotated to the "nuclear receptor" intended target family, where the subfamily is "steroidal".	ESR1
ATG_ERE_CIS_up		992 / 3807	Data from the assay component ATG_ERE_CIS was analyzed into 1 assay endpoint. This assay endpoint, ATG_ERE_CIS_up, was analyzed in the positive fitting direction relative to DMSO as the negative control and baseline of activity. Using a type of inducible reporter, measures of mRNA for gain-of-signal activity can be used to understand the reporter gene at the transcription factor-level as they relate to the gene ESR1. Furthermore, this assay endpoint can be referred to as a primary readout, because this assay has produced multiple assay endpoints where this one serves a reporter gene function. To generalize the intended target to other relatable targets, this assay endpoint is annotated to the "nuclear receptor" intended target family, where the subfamily is "steroidal".	ESR1

<< < 1 2 3 4 > >>

Showing 1 to 10 of 32 records

Mass/Formula Searching

Advanced Searches
Mass Search



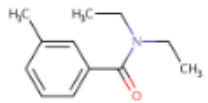
Advanced Searches
Mass Search

Search Results

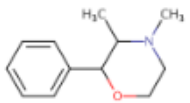
Searched by Mass: 191.131 +/- 5.0 ppm.

329 of 329 chemicals visible

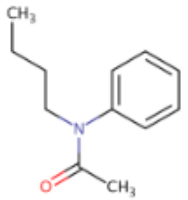
Select all | Download | Send to Batch Search | Mass Difference | DTXSID | CASRN | TOXCAST | Mass Diff | Multicomponent Chemicals | Filter by Name or CASRN



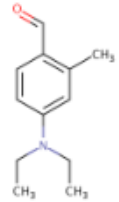
DEET
DTXSID: DTXSID2021995
CASRN: 134-62-3
TOXCAST: 12/768
Mass Diff: 0.000014



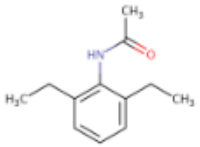
Phendimetrazine
DTXSID: DTXSID1023447
CASRN: 634-03-7
TOXCAST: -
Mass Diff: 0.000014



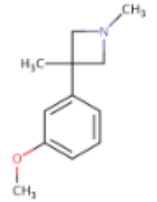
N-Butylacetanilide
DTXSID: DTXSID2042197
CASRN: 91-49-6
TOXCAST: -
Mass Diff: 0.000014



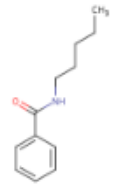
Benzaldehyde, 4-(diethylamino)-2-meth...
DTXSID: DTXSID4059041
CASRN: 92-14-8
TOXCAST: -
Mass Diff: 0.000014



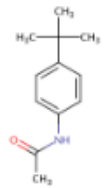
Acetanilide, 2',6'-diethyl-
DTXSID: DTXSID90168148
CASRN: 16665-89-7
TOXCAST: -
Mass Diff: 0.000014



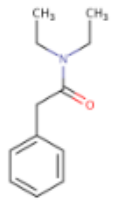
Azetidine, 1,3-dimethyl-3-(m-methoxyp...
DTXSID: DTXSID40173560
CASRN: 19832-26-9
TOXCAST: -
Mass Diff: 0.000014



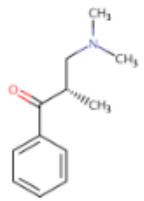
Benzamide, N-pentyl-
DTXSID: DTXSID20174196
CASRN: 20308-43-4
TOXCAST: -



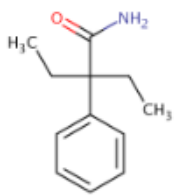
p-t-Butylacetanilide
DTXSID: DTXSID80174238
CASRN: 20330-45-4
TOXCAST: -



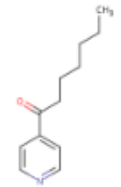
N,N-Diethylphenylacetamide
DTXSID: DTXSID00179048
CASRN: 2431-96-1
TOXCAST: -



3-(Dimethylamino)-2-methylpropiophen...
DTXSID: DTXSID60180796
CASRN: 26171-50-6
TOXCAST: -



Butyramide, 2-ethyl-2-phenyl-
DTXSID: DTXSID60184653
CASRN: 30568-39-9
TOXCAST: -



1-Heptanone, 1-(4-pyridyl)-
DTXSID: DTXSID40186594
CASRN: 32941-30-3
TOXCAST: -

Mass Spec Focused Applications

Analytical and Bioanalytical Chemistry (2019) 411:853–866
https://doi.org/10.1007/s00216-018-1435-6

RESEARCH PAPER

EPA's non-targeted analysis collaborative trial (ENTACT): genesis, design, and initial findings

Elin M. Ulrich[1] · Jon R. Sobus[1] · Christopher M. Grulke[2] · Ann M. Richard[2] · Seth R. Newton[1] · Mark J. Strynar[1] · Kamel Mansouri[3,4] · Antony J. Williams[2]

Analytical and Bioanalytical Chemistry (2019) 411:835–851
https://doi.org/10.1007/s00216-018-1526-4

RESEARCH PAPER

Using prepared mixtures of ToxCast chemicals to evaluate non-targeted analysis (NTA) method performance

Jon R. Sobus[1] · Jarod N. Grossman[2,3] · Alex Chao[2] · Randolph Singh[4] · Antony J. Williams[5] · Christopher M. Grulke[5] · Ann M. Richard[5] · Seth R. Newton[1] · Andrew D. McEachran[4] · Elin M. Ulrich[1]

Mass Spec Focused Applications

Journal of Exposure Science & Environmental Epidemiology (2018) 28:411–426
https://doi.org/10.1038/s41370-017-0012-y

REVIEW ARTICLE

Integrating tools for non-targeted analysis research and chemical safety evaluations at the US EPA

Jon R. Sobus[1] · John F. Wambaugh[2] · Kristin K. Isaacs[1] · Antony J. Williams[2] · Andrew D. McEachran[3] · Ann M. Richard[2] · Christopher M. Grulke[2] · Elin M. Ulrich[1] · Julia E. Rager[3,4] · Mark J. Strynar[1] · Seth R. Newton[1]

Environmental Science & Technology

Article

Cite This: *Environ. Sci. Technol.* 2018, 52, 3125–3135

pubs.acs.org/est

Suspect Screening Analysis of Chemicals in Consumer Products

Katherine A. Phillips,[†] Alice Yau,[‡] Kristin A. Favela,[‡] Kristin K. Isaacs,[†] Andrew McEachran,[§,∥] Christopher Grulke,[∥] Ann M. Richard,[∥] Antony J. Williams,[∥] Jon R. Sobus,[†] Russell S. Thomas,[∥] and John F. Wambaugh*[,∥]

Batch Searching

Batch Searching

- Singleton searches are great but…
- …we generally want data on LOTS of chemicals!

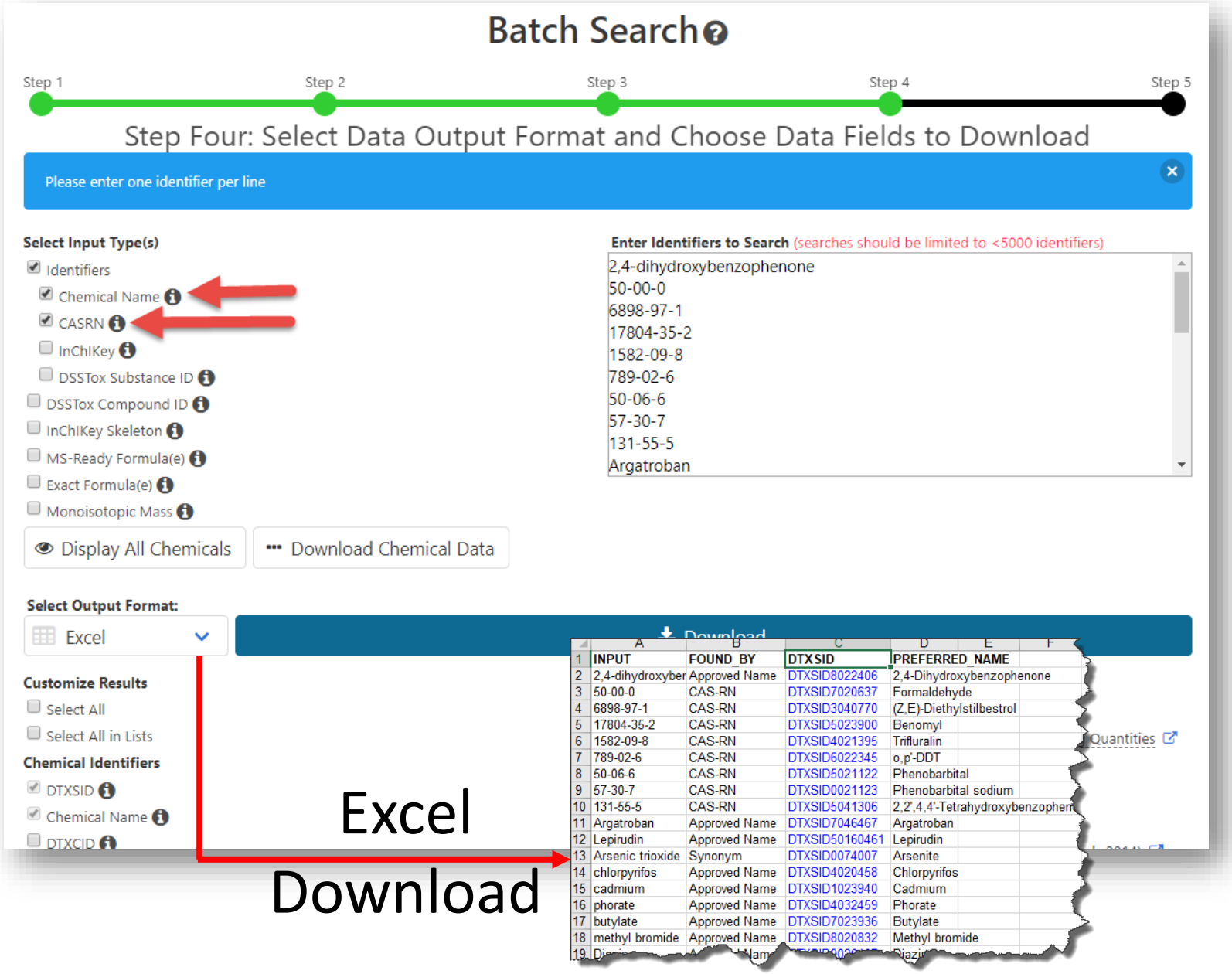
- Typical questions
  - What are the structures for a set of chemical names? Set of CASRNs?
  - Can I get chemical lists in Excel files? As a list of SMILES strings? Can I get an SDF file?
  - Can I include predicted properties in the download file? OPERA? TEST?
  - Are “these chemicals” screened in Toxcast?
  - I’m a mass spectrometrist and need masses and formulae for a list of chemicals

Chemicals Discussed Today

- During talks today we heard about….
  - 2,4-dihydroxybenzophenone, 131-56-6 (Mark)
  - 50-00-0, 6898-97-1, 17804-35-2, 1582-09-8, 789-02-6, 50-06-6, 57-30-7, 131-55-5 (Shannon)
  - Argatroban, Lepirudin (Thomas)
  - Arsenic trioxide, chlorpyrifos, cadmium, phorate, butylate, methyl bromide, Diazinon, Fonofos, Atrazine, Dichlorvos, Phorate, Parathion, 2-butenal, pyruvaldehyde, nicotine, formaldehyde, acetaldehyde, acetone, propionaldehyde (Carolyn)
- What information can we find and how fast???

Batch Search Identifiers

```
2,4-dihydroxybenzophenone
50-00-0
6898-97-1
17804-35-2
1582-09-8
789-02-6
50-06-6
57-30-7
131-55-5
Argatroban
Lepirudin
Arsenic trioxide
chlorpyrifos
cadmium
phorate
butylate
methyl bromide
Diazinon
Fonofos
Atrazine
Dichlorvos
Phorate
Parathion
2-butenal
pyruvaldehyde
nicotine
formaldehyde
acetaldehyde
acetone
propionaldehyde
```



	A	B	C	D	E	F
1	INPUT	FOUND_BY	DTXSID	PREFERRED_NAME		
2	2,4-dihydroxyber	Approved Name	DTXSID8022406	2,4-Dihydroxybenzophenone		
3	50-00-0	CAS-RN	DTXSID7020637	Formaldehyde		
4	6898-97-1	CAS-RN	DTXSID3040770	(Z,E)-Diethylstilbestrol		
5	17804-35-2	CAS-RN	DTXSID5023900	Benomyl		
6	1582-09-8	CAS-RN	DTXSID4021395	Trifluralin		
7	789-02-6	CAS-RN	DTXSID6022345	o,p'-DDT		
8	50-06-6	CAS-RN	DTXSID5021122	Phenobarbital		
9	57-30-7	CAS-RN	DTXSID0021123	Phenobarbital sodium		
10	131-55-5	CAS-RN	DTXSID5041306	2,2',4,4'-Tetrahydroxybenzophen		
11	Argatroban	Approved Name	DTXSID7046467	Argatroban		
12	Lepirudin	Approved Name	DTXSID50160461	Lepirudin		
13	Arsenic trioxide	Synonym	DTXSID0074007	Arsenite		
14	chlorpyrifos	Approved Name	DTXSID4020458	Chlorpyrifos		
15	cadmium	Approved Name	DTXSID1023940	Cadmium		
16	phorate	Approved Name	DTXSID4032459	Phorate		
17	butylate	Approved Name	DTXSID7023936	Butylate		
18	methyl bromide	Approved Name	DTXSID8020832	Methyl bromide		
19	Diazi	Nam		Diazi		

Include Data of Interest – then dive deeper…

Chemical Identifiers

- [x] DTXSID
- [x] Chemical Name
- [ ] DTXCID
- [x] CAS-RN
- [ ] InChIKey
- [ ] IUPAC Name

Structures

- [ ] Mol File
- [ ] SMILES
- [ ] InChI String
- [ ] MS-Ready SMILES
- [ ] QSAR-Ready SMILES

Intrinsic And Predicted Properties

- [ ] Molecular Formula
- [ ] Average Mass
- [ ] Monoisotopic Mass
- [ ] TEST Model Predictions
- [x] OPERA Model Predictions

Metadata

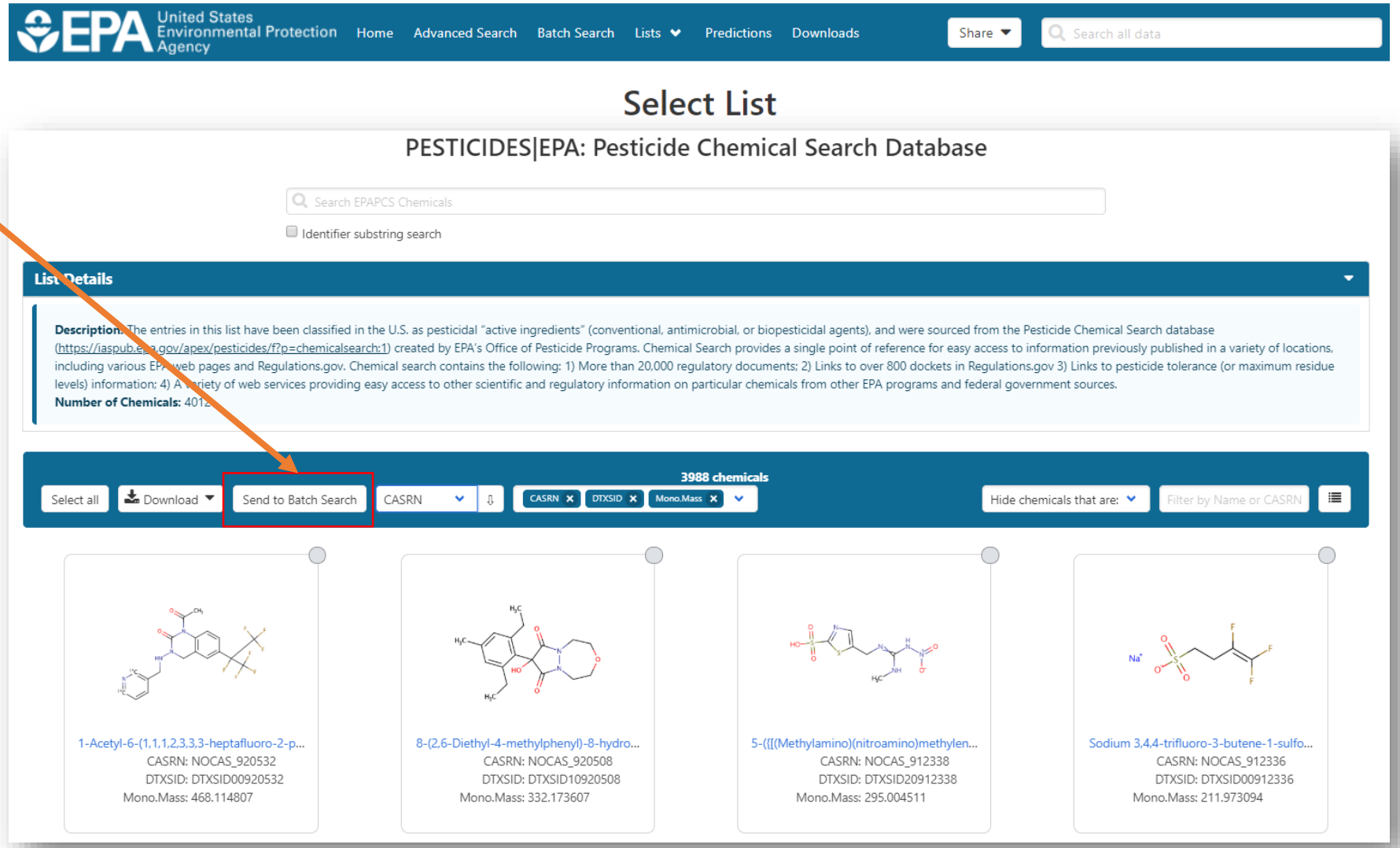
- [ ] Curation Level Details
- [x] NHANES/Predicted Exposure
- [ ] Data Sources
- [x] Include ToxVal Data Availability
- [x] Assay Hit Count
- [x] Number of PubMed Articles
- [ ] PubChem Data Sources
- [ ] CPDat Product Occurrence Count
- [x] IRIS

DTXSID	PREFERRED_NA	CASRN	EXPOCAST	EXPOCAST	NHANES	TOXVAL_DATA	TOXCAST_PERCENT_	TOXCAST_NUMBER	NUMBER_OF_PUBMED_ARTICL	IRIS_LINK
DTXSID80	2,4-Dihydroxyber	131-56-6	2.78e-07	Y	-	Y	14.02	92/656	25	-
DTXSID70	Formaldehyde	50-00-0	1.32e-06	Y	-	Y	-	-	19234	Y
DTXSID30	(Z,E)-Diethylstilb	6898-97-1	1.65e-07	Y	-	-	25.1	62/247	8326	-
DTXSID50	Benomyl	17804-35-2	1.11e-07	Y	-	Y	11.23	96/855	476	Y
DTXSID40	Trifluralin	1582-09-8	1.57e-06	Y	-	Y	10.44	87/833	259	Y
DTXSID60	o,p'-DDT	789-02-6	1.16e-07	Y	-	Y	32.87	239/727	125	-
DTXSID50	Phenobarbital	50-06-6	2.57e-08	Y	-	Y	1.7	4/235	18244	-
DTXSID00	Phenobarbital so	57-30-7	1.66e-07	Y	-	Y	2.15	15/698	17600	-
DTXSID50	2,2',4,4'-Tetrahyd	131-55-5	2.31e-07	Y	-	Y	18.31	139/759	25	-
DTXSID70	Argatroban	74863-84-6	3.3e-07	Y	-	-	0.43	1/235	868	-
DTXSID50	Lepirudin	138068-37	-	-	-	-	-	-	354	-
DTXSID00	Arsenite	15502-74-6	-	-	-	-	-	-	4003	-
DTXSID40	Chlorpyrifos	2921-88-2	2.3e-08	Y	Y	Y	18.73	124/662	2387	Y
DTXSID10	Cadmium	7440-43-9	-	-	-	Y	-	-	29395	Y
DTXSID40	Phorate	298-02-2	1.23e-08	Y	Y	Y	5.26	36/685	113	-
DTXSID70	Butylate	2008-41-5	6.43e-08	Y	-	Y	1.14	9/790	17	Y
DTXSID80	Methyl bromide	74-83-9	-	-	-	Y	-	-	500	Y
DTXSID90	Diazinon	333-41-5	1.02e-07	Y	Y	Y	6.9	59/855	890	-
DTXSID20	Fonofos	944-22-9	7.86e-08	Y	-	Y	2.13	5/235	41	Y
DTXSID90	Atrazine	1912-24-9	6.56e-08	Y	Y	Y	4.62	40/866	14356	Y
DTXSID50	Dichlorvos	62-73-7	1.37e-08	Y	Y	Y	7.13	58/814	1106	Y
DTXSID40	Phorate	298-02-2	1.23e-08	Y	Y	Y	5.26	36/685	113	-
DTXSID70	Parathion	56-38-2	8.86e-08	Y	Y	Y	13.05	116/889	2213	Y
DTXSID80	Crotonaldehyde	4170-30-3	8.64e-07	Y	-	Y	0.0	0/235	172	-
DTXSID00	Methyl glyoxal	78-98-8	-	-	-	Y	0.72	2/277	1553	-
DTXSID10	Nicotine	54-11-5	7.08e-07	Y	-	Y	2.39	17/711	22837	-
DTXSID70	Formaldehyde	50-00-0	1.32e-06	Y	-	Y	-	-	19234	Y
DTXSID50	Acetaldehyde	75-07-0	2.57e-06	Y	-	Y	35.5	142/400	5568	Y
DTXSID80	Acetone	67-64-1	4.72e-05	Y	-	Y	0.43	1/235	7176	Y
DTXSID20	Propanal	123-38-6	2.73e-05	Y	-	Y	0.0	0/235	1219	Y

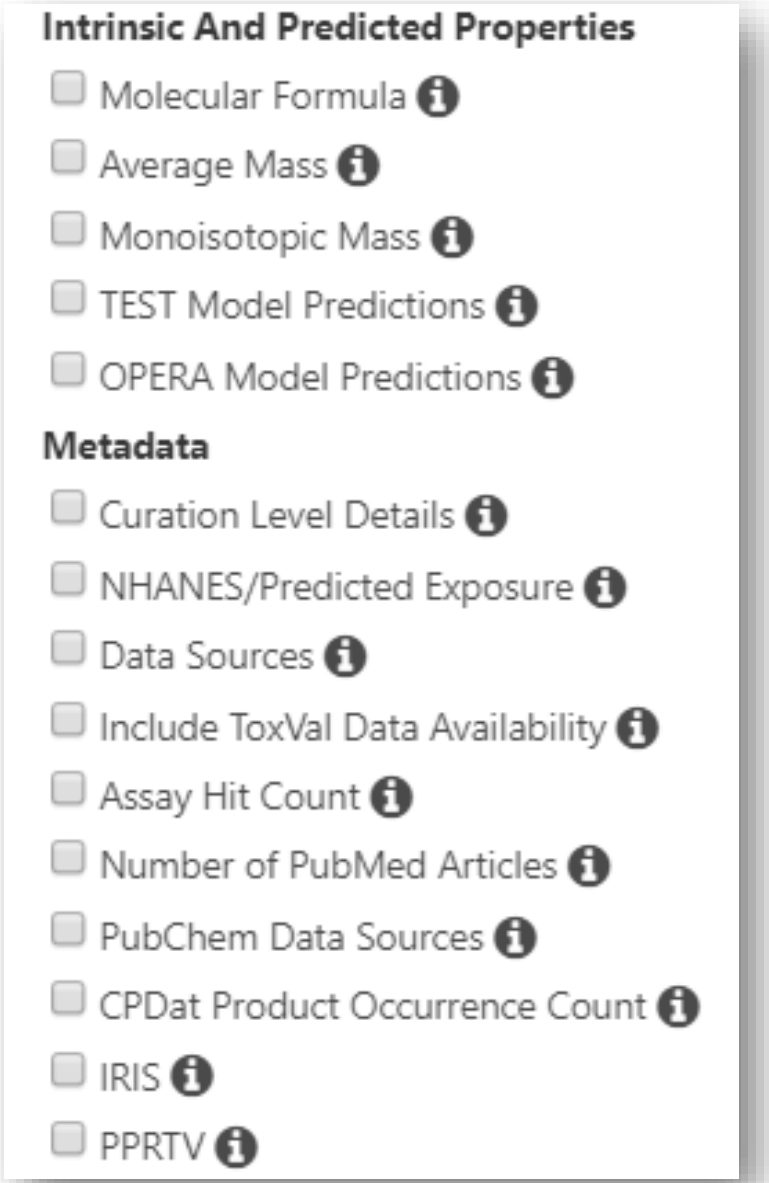
Batch collection of data for a set of pesticides

Curated List of Pesticides

- Find list of interest
- Select list and send to batch



Send to batch and select….



- A few seconds to assemble
  - ToxCast data - #actives/#assays and % active
  - # articles in PubMed
  - Links to IRIS or PPRTV reports
  - TEST or OPERA predictions
  - Exposure data: predictions and CPDat

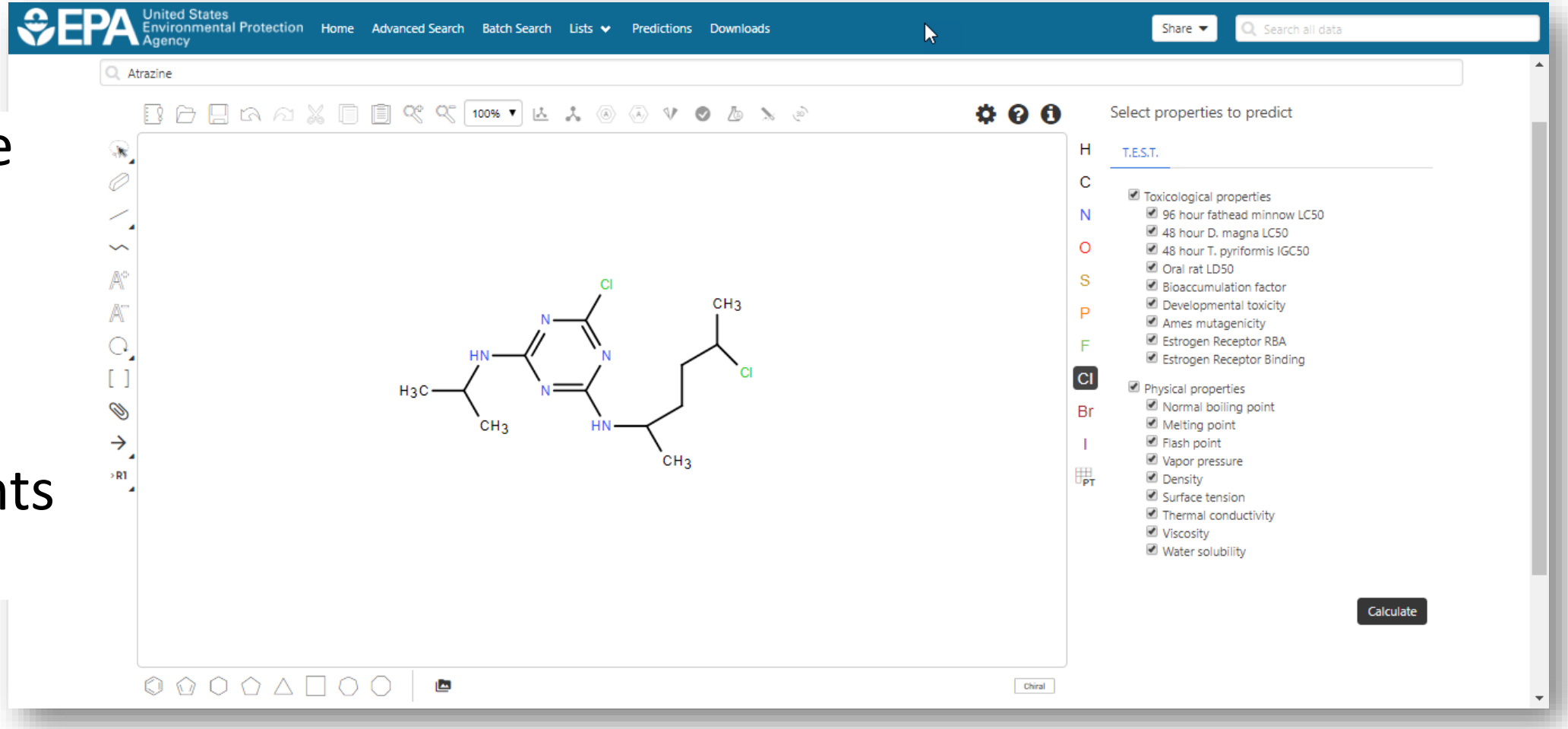
	A	B	C	D	E	F	G	H	I	J	K	L	M
	DTXSID	PREFERRED_NAME	EXPOCAST_MEI	EXPOCAST	NHANES	TOXVAL_DATA	TOXCAST_%_ACT	TOXCAST_	#PUBMED_	PUBCHEM_	CPDAT_COUI	IRIS_LINK	PPRTV_LINK
	DTXSID2021105	Pentachloronitrobenzene	1.14e-07	Y	Y	Y	11.8	99/839	69	96	164	Y	-
	DTXSID4022527	Propylparaben	1.4e-05	Y	Y	Y	13.77	99/719	201	121	1476	-	-
	DTXSID4024064	Dinex	8.29e-08	Y	-	Y	42.13	99/235	-	35	5	Y	-
	DTXSID0032493	Triadimenol	1.73e-08	Y	-	Y	10.54	98/930	163	74	83	-	-
	DTXSID4032667	Esfenvalerate	1.7e-06	Y	-	Y	11.45	98/856	483	45	198	-	-
	DTXSID6020561	Endrin	1.29e-07	Y	-	Y	14.02	98/699	284	16	98	Y	Y
	DTXSID6025355	Glutaraldehyde	2.03e-05	Y	-	Y	14.35	98/683	6515	139	1144	-	-
	DTXSID8032417	Isofenphos	1.87e-08	Y	-	Y	16.28	98/602	30	42	60	-	-
	DTXSID6032352	Chlorpyrifos-methyl	1.07e-07	Y	Y	Y	11.27	97/861	72	50	116	-	-
	DTXSID8020620	Fenthion	8.99e-08	Y	Y	Y	11.56	97/839	354	100	99	-	-
	DTXSID2020189	FD&C Blue No. 1	0.000178	Y	-	Y	13.72	97/707	174	49	672	-	-
	DTXSID7044843	Erythrosin B	6.3e-07	Y	-	-	24.25	97/400	14843	51	7	-	-
	DTXSID5041778	Chloropropylate	1.05e-07	Y	-	Y	40.93	97/237	-	36	12	-	-
	DTXSID5023900	Benomyl	1.11e-07	Y	-	Y	11.23	96/855	476	91	105	Y	-
	DTXSID9020247	Carbaryl	5.61e-08	Y	Y	Y	11.51	96/834	1135	117	245	Y	-
	DTXSID8024109	Flutolanil	1.63e-08	Y	-	Y	11.4	95/833	6	59	80	-	-
	DTXSID1023998	Cypermethrin	1.62e-06	Y	Y	Y	10.78	94/872	1148	148	246	-	-
	DTXSID2024242	Paclobutrazol	9.19e-08	Y	-	Y	11.11	94/846	139	-	40	Y	-
	DTXSID1020807	2-Mercaptobenzothiazole	4.7e-05	Y	-	Y	12.82	94/733	111	181	86	-	Y

Real-Time Predictions

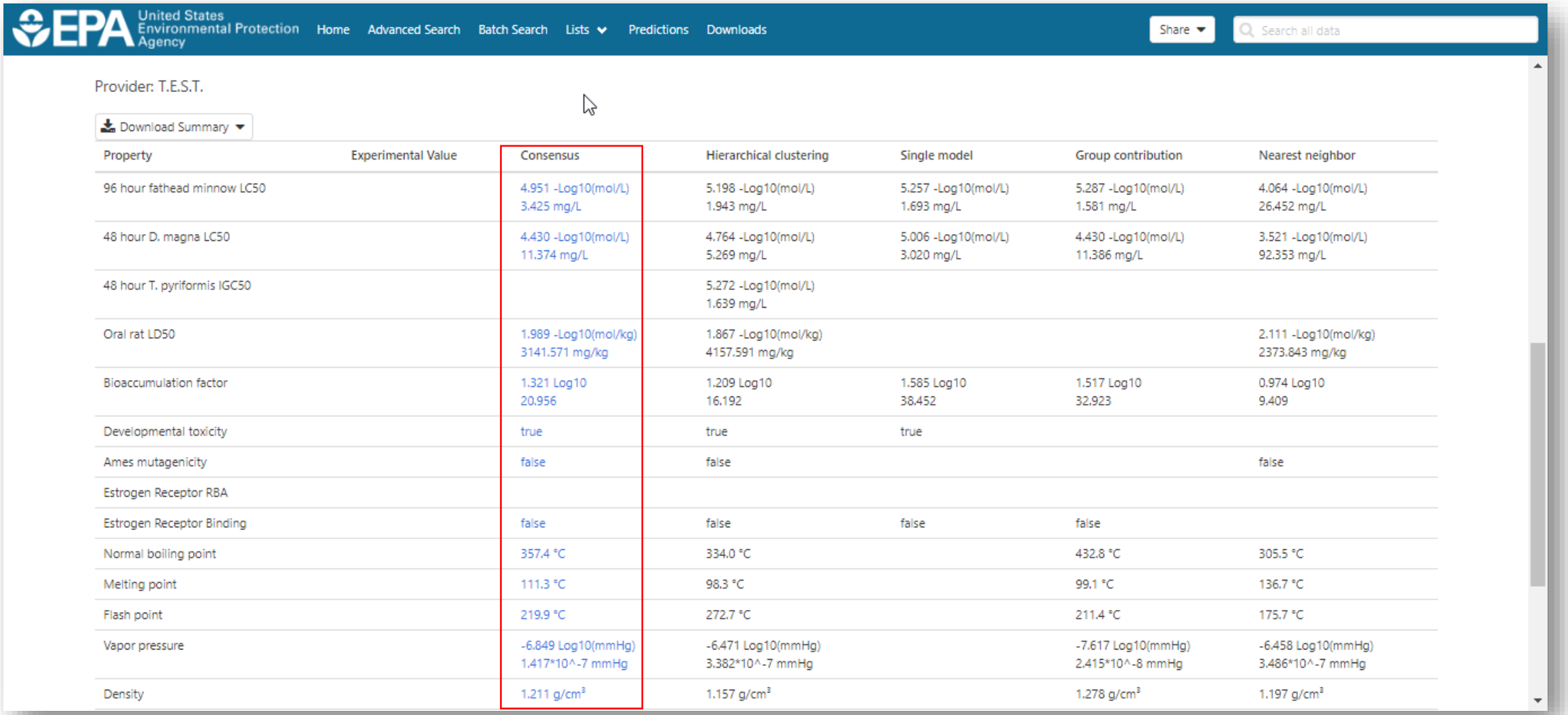
Input structure using sketcher

Edit chemical

Select endpoints for prediction



Real-Time Predictions with detailed calculation reports



Property	Experimental Value	Consensus	Hierarchical clustering	Single model	Group contribution	Nearest neighbor
96 hour fathead minnow LC50		4.951 -Log10(mol/L) 3.425 mg/L	5.198 -Log10(mol/L) 1.943 mg/L	5.257 -Log10(mol/L) 1.693 mg/L	5.287 -Log10(mol/L) 1.581 mg/L	4.064 -Log10(mol/L) 26.452 mg/L
48 hour D. magna LC50		4.430 -Log10(mol/L) 11.374 mg/L	4.764 -Log10(mol/L) 5.269 mg/L	5.006 -Log10(mol/L) 3.020 mg/L	4.430 -Log10(mol/L) 11.386 mg/L	3.521 -Log10(mol/L) 92.353 mg/L
48 hour T. pyriformis IGC50			5.272 -Log10(mol/L) 1.639 mg/L			
Oral rat LD50		1.989 -Log10(mol/kg) 3141.571 mg/kg	1.867 -Log10(mol/kg) 4157.591 mg/kg			2.111 -Log10(mol/kg) 2373.843 mg/kg
Bioaccumulation factor		1.321 Log10 20.956	1.209 Log10 16.192	1.585 Log10 38.452	1.517 Log10 32.923	0.974 Log10 9.409
Developmental toxicity		true	true	true		
Ames mutagenicity		false	false			false
Estrogen Receptor RBA						
Estrogen Receptor Binding		false	false	false	false	
Normal boiling point		357.4 °C	334.0 °C		432.8 °C	305.5 °C
Melting point		111.3 °C	98.3 °C		99.1 °C	136.7 °C
Flash point		219.9 °C	272.7 °C		211.4 °C	175.7 °C
Vapor pressure		-6.849 Log10(mmHg) 1.417*10^-7 mmHg	-6.471 Log10(mmHg) 3.382*10^-7 mmHg		-7.617 Log10(mmHg) 2.415*10^-8 mmHg	-6.458 Log10(mmHg) 3.486*10^-7 mmHg
Density		1.211 g/cm^3	1.157 g/cm^3		1.278 g/cm^3	1.197 g/cm^3

- Four individual models plus consensus model with calculation report

Real-Time Predictions with detailed calculation reports

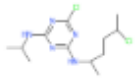
Predicted Vapor pressure at 25°C for ClC=1N=C(N=C(N1)NC(C)CCC(Cl)C)NC(C)C from Consensus method

Prediction results

Endpoint	Experimental value	Predicted value
Vapor pressure at 25°C Log10(mmHg)	N/A	-6.85
Vapor pressure at 25°C mmHg	N/A	1.42E-07

Individual Predictions

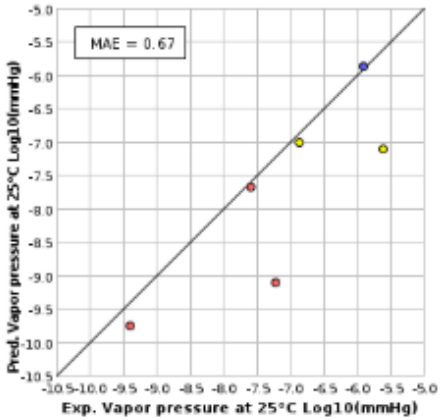
Method	Predicted value Log10(mmHg)
Hierarchical clustering	-6.47
Group contribution	-7.62
Nearest neighbor	-6.46



Predictions for the test chemical and for the most similar chemicals in the external test set

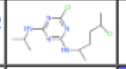
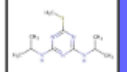
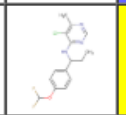
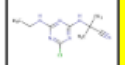
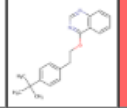
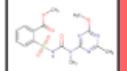
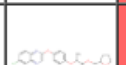
If the predicted value matches the experimental values for similar chemicals in the test set (and the similar chemicals were predicted well), one

Prediction results (colors defined in table below)



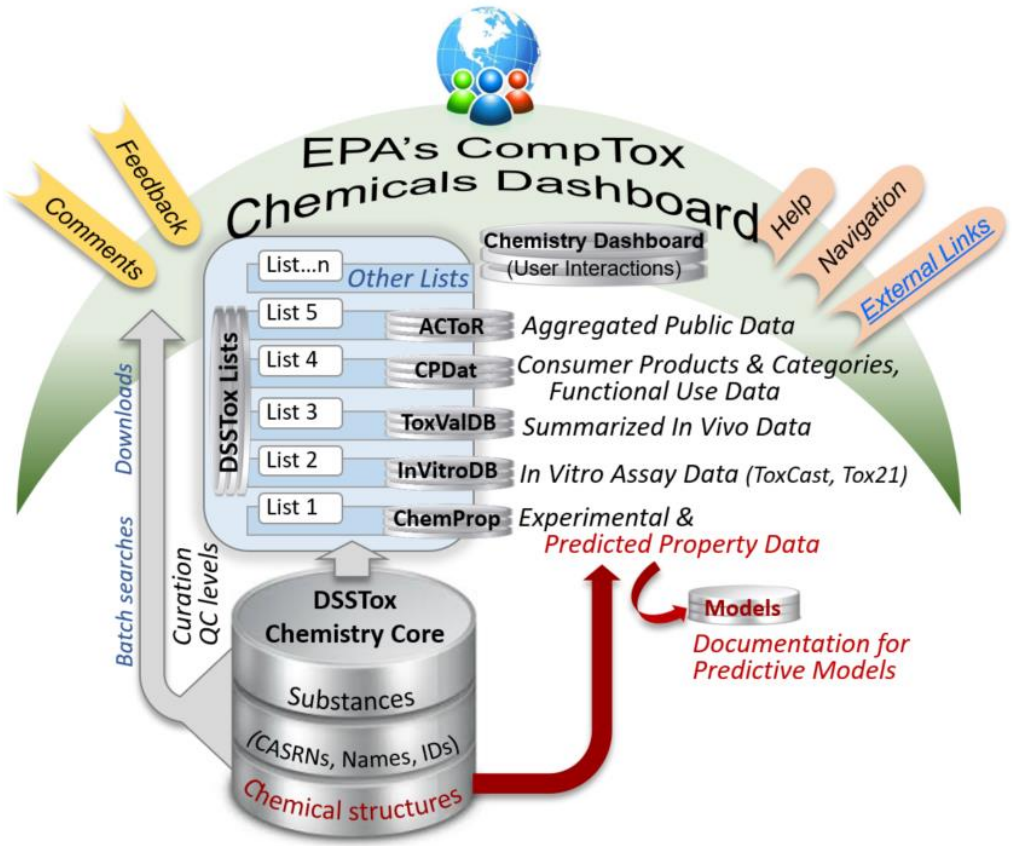
Chemicals	MAE*
Entire set	0.47
Similarity coefficient ≥ 0.5	0.67

*Mean absolute error in Log10(mmHg)

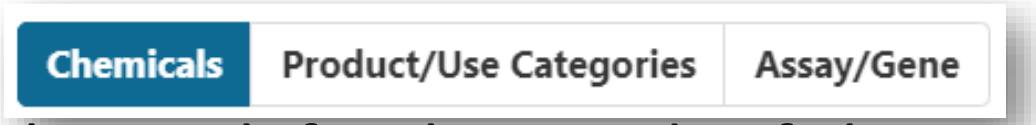
CAS	Structure	Similarity Coefficient	Experimental value Log10(mmHg)	Predicted value Log10(mmHg)
ClC=1N=C(N=C(N1)NC(C)CCC(Cl)C)NC(C)C (test chemical)			N/A	-6.85
7287-19-6		0.83	-5.91	-5.86
130339-07-0		0.77	-5.62	-7.11
21725-46-2		0.76	-6.86	-7.01
120928-09-8		0.58	-7.59	-7.67
101200-48-0		0.56	-9.41	-9.76
119738-06-6		0.55	-7.23	-9.11

- Full prediction report
- Shows chemicals used in training set

Summary and Conclusion



- CompTox Chemicals Dashboard - a central hub for environmental data
  - ~875k chemical substances
  - Integrating property data, hazard data, exposure data, *in vitro* bioactivity data
  - Interrogation of bioactivity data -
  - Multiple types of searches



  - Batch search for thousands of chemicals
  - Real-time property and toxicity predictions
  - Downloadable files – CSV, TSV and Excel

Acknowledgements

EPA-RTP

- *An enormous team of contributors from Center for Computational Toxicology and Exposure, especially the IT development team*
- *Multiple centers and laboratories across EPA*
- *Special Mentions:*
  - *Dashboard: Jeff Edwards and Chris Grulke*
  - *GenRA: Imran Shah & George Helman*
  - *Abstract Sifter: Nancy Baker*
  - *InvitroDB: Katie Paul-Friedman*

References

- The CompTox Chemistry Dashboard: a community data resource for environmental chemistry, *J. Cheminformatics*, **9**, 61 (2017)
- EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research, *Comp. Tox.* **12**, 100096 (2019)
- OPERA models for predicting physicochemical properties and environmental fate endpoints, *J. Cheminformatics*, **10**, 10 (2018)
- Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model, *Environ. Sci. Technol.* **49**, 8804-8814 (2015)
- ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology, *Chem. Res. Toxicol.* **29**, 1225-51 (2016)
- Development and Validation of a Computational Model for Androgen Receptor Activity, *Chem. Res. Toxicol.* **30**, 946-964 (2017)
- CERAPP: Collaborative Estrogen Receptor Activity Prediction Project, *Environ. Health Perspect*. **124**, 1023 (2016)
- Abstract Sifter: a comprehensive front-end system to PubMed, *F1000,* **6**, 2164 (2017)

LIVE DEMO IF TIME