http://www.orcid.org/0000-0002-2668-4821

US-EPA CompTox chemicals dashboard: A web-based data integration hub for environmental chemistry data

Antony Williams, *Chris Grulke, Richard Judson, John Wambaugh, Jeremy Dunne and Jeff Edwards*

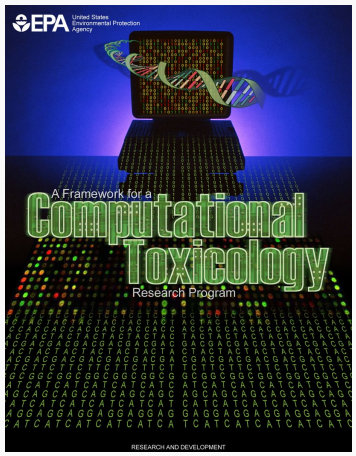
National Center for Computational Toxicology, U.S. Environmental Protection Agency, RTP, NC

Spring 2019
ACS Spring Meeting, Orlando

Abstract

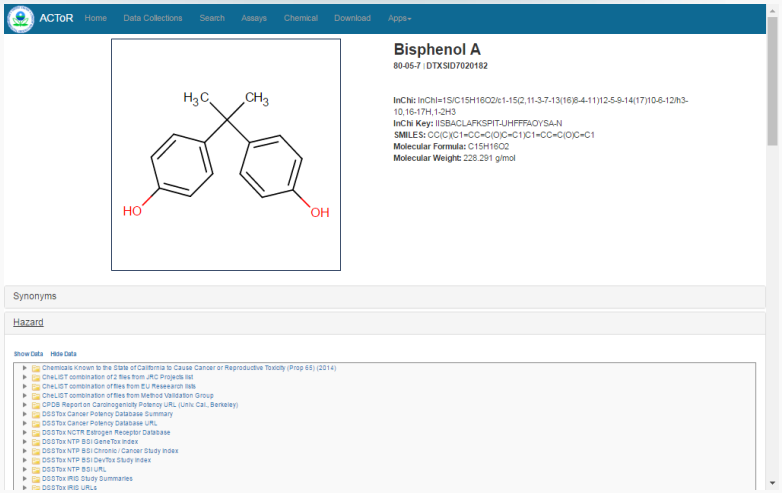
- The US EPA's CompTox Chemistry Dashboard provides access to various types of data associated with ~760,000 chemical substances. These data include experimental and predicted property data, high-throughput screening assay data and hazard and environmental exposure data. With millions of individual data points and annotations associated with hundreds of thousands of chemicals, data quality is a priority. With tens of thousands of individual users per month browsing the data on the dashboard, the ability of users to provide feedback has allowed us to identify, confirm and address issues in the data. This has required the implementation of novel approaches for data feedback via the user interface that include general feedback on the dashboard and down to individual data points contained in a table. We are presently investigating ways to garner feedback on our ToxCast bioassay data to facilitate the curation of tens of thousands of data points. This presentation will provide an overview of our existing capabilities in the CompTox Chemistry Dashboard for gathering crowdsourced data from the user base and its impact on assisting in the curation of data.

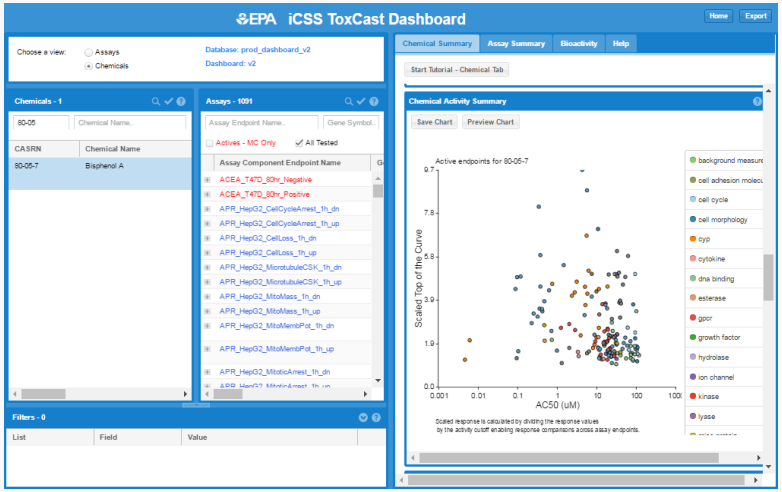
National Center for Computational Toxicology

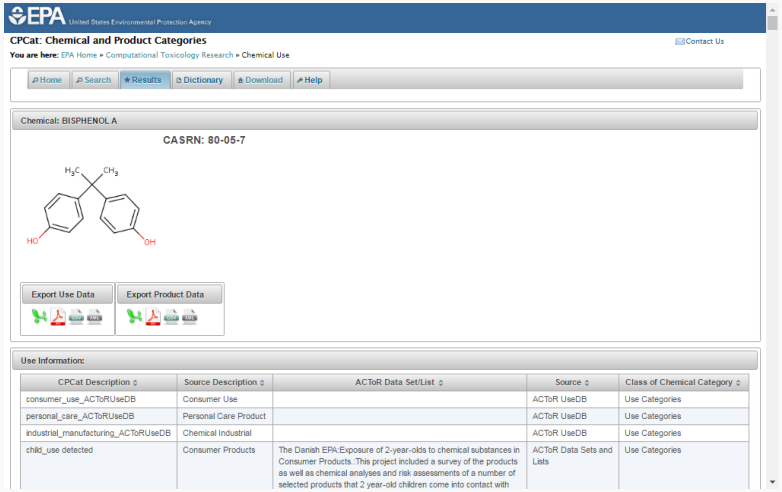


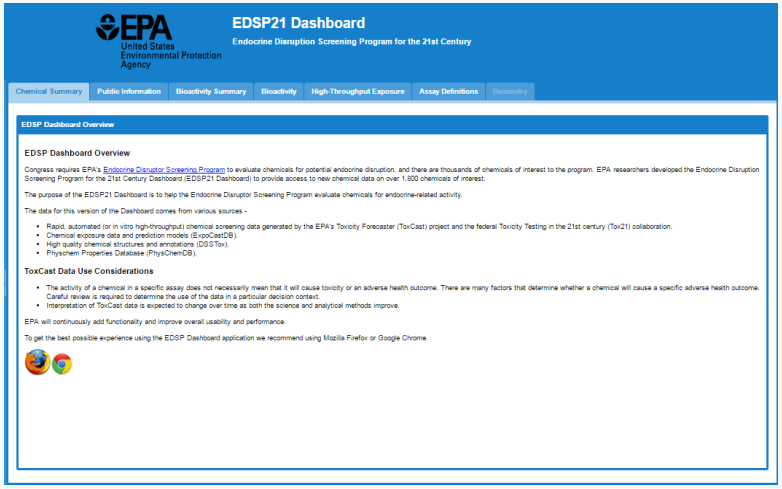
- National Center for Computational Toxicology established in 2005 to integrate:
  - High-throughput and high-content technologies
  - Modern molecular biology
  - Data mining and statistical modeling
  - Computational biology and chemistry
- Researching computational approaches to quickly evaluate the safety of chemicals for potential risk.
- Outputs: a lot of data, models, algorithms and software applications

Early Dashboard Applications





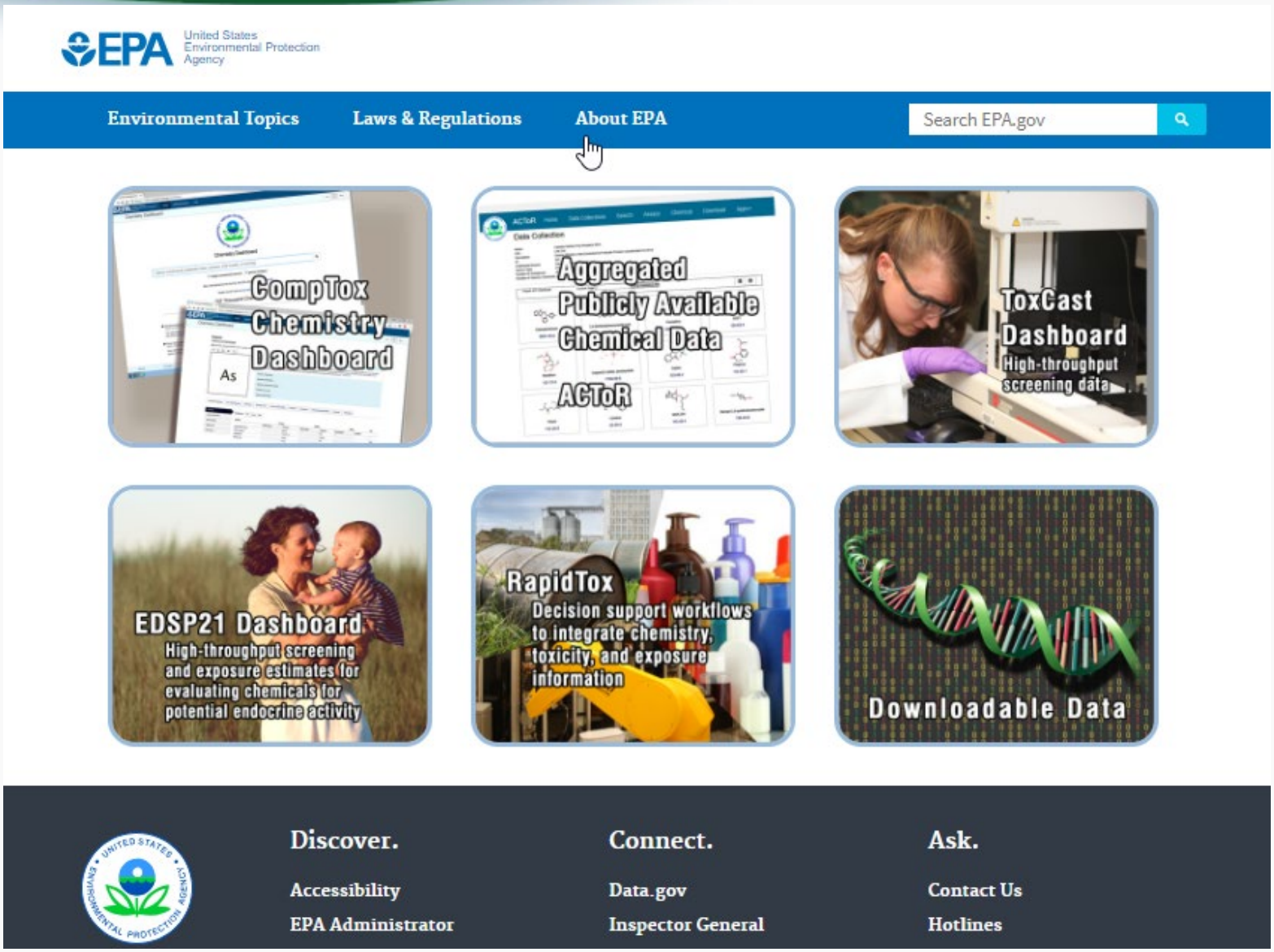




The CompTox ~~Chemistry~~ Dashboard

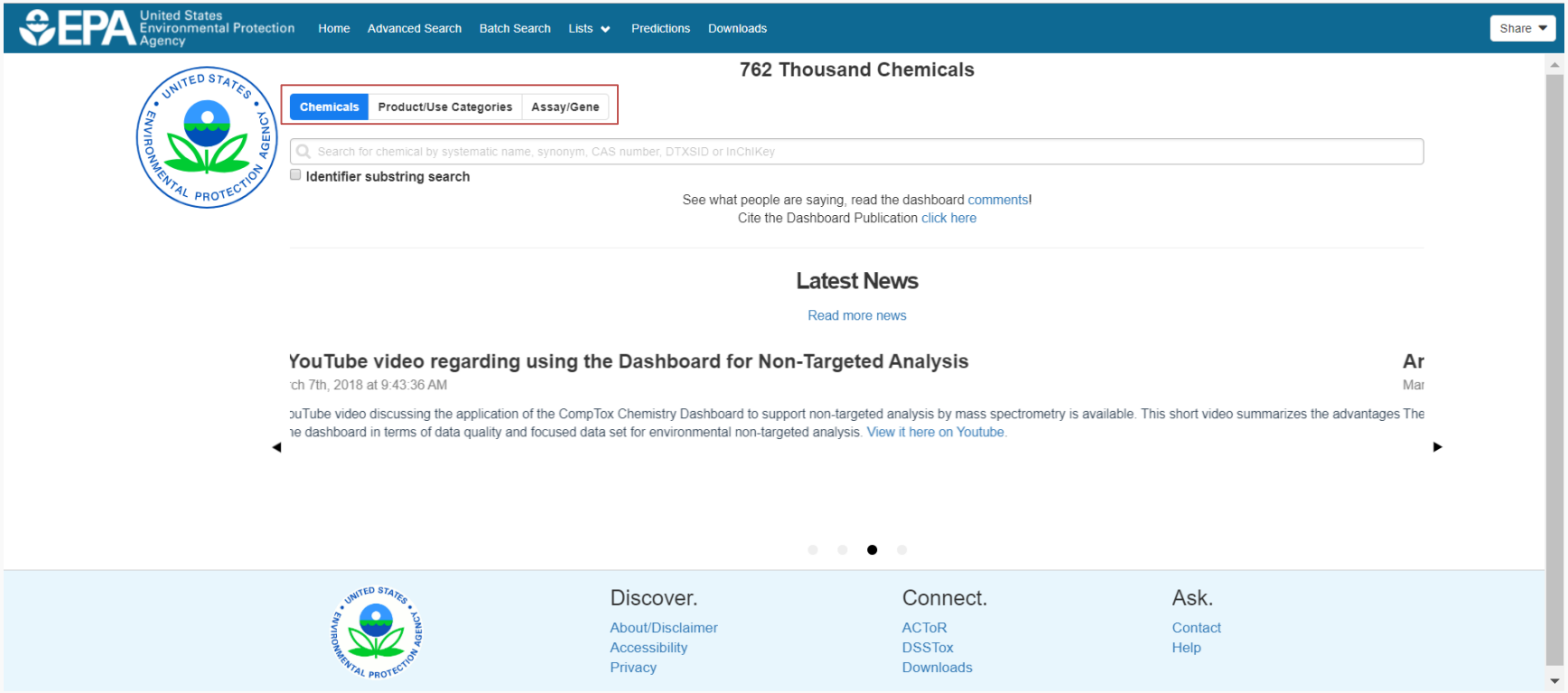
- A publicly accessible website delivering access:
  - New entry portal for all NCCT dashboards
  - **~762,000** chemicals with related property data
  - **Searchable by chemical, product use, gene and assay (ToxCast)**
  - Experimental and predicted physicochemical property data
  - **“Bioactivity data” for the ToxCast/Tox21 project**
  - **Generalized Read-Across (GenRA) module**
  - 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 Portal

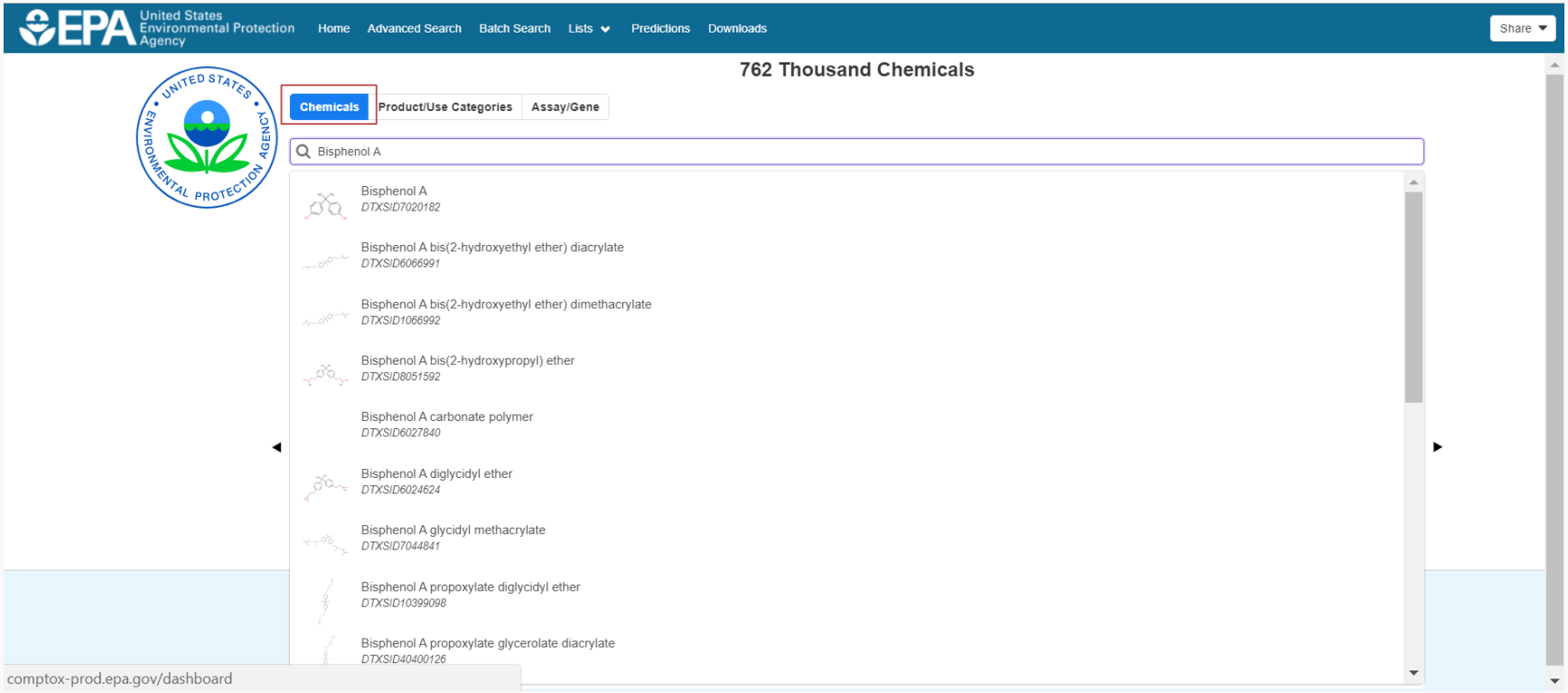


CompTox Dashboard

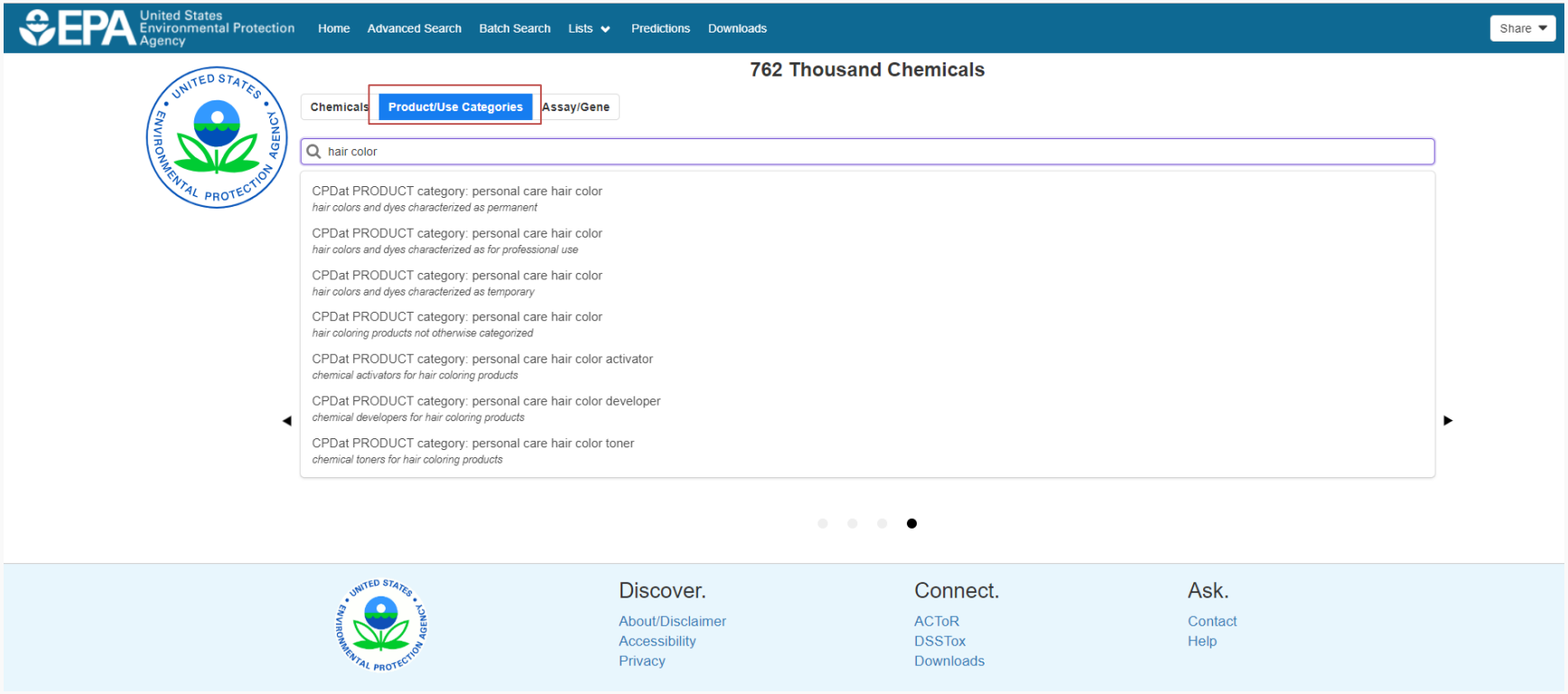
https://comptox.epa.gov/dashboard



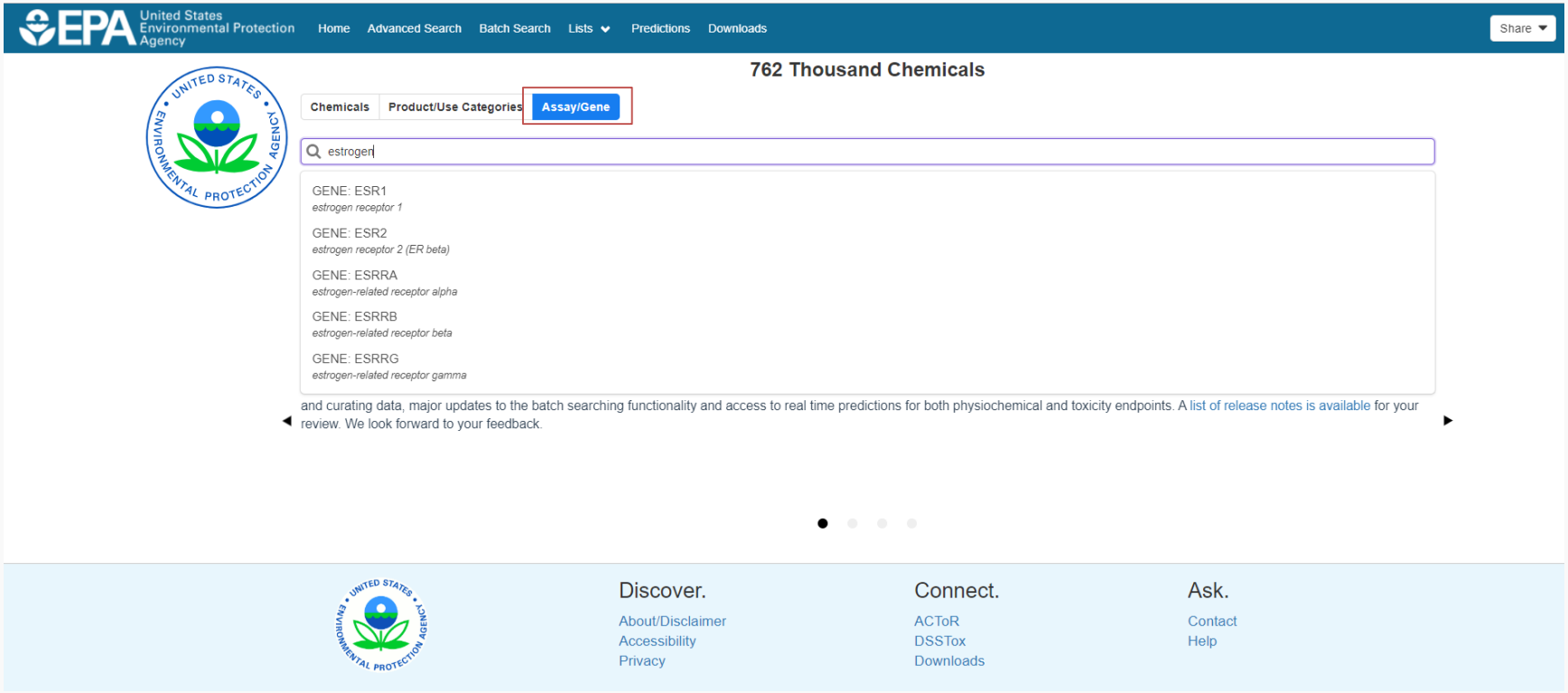
CompTox Dashboard Chemicals



CompTox Dashboard
Products and Use Categories



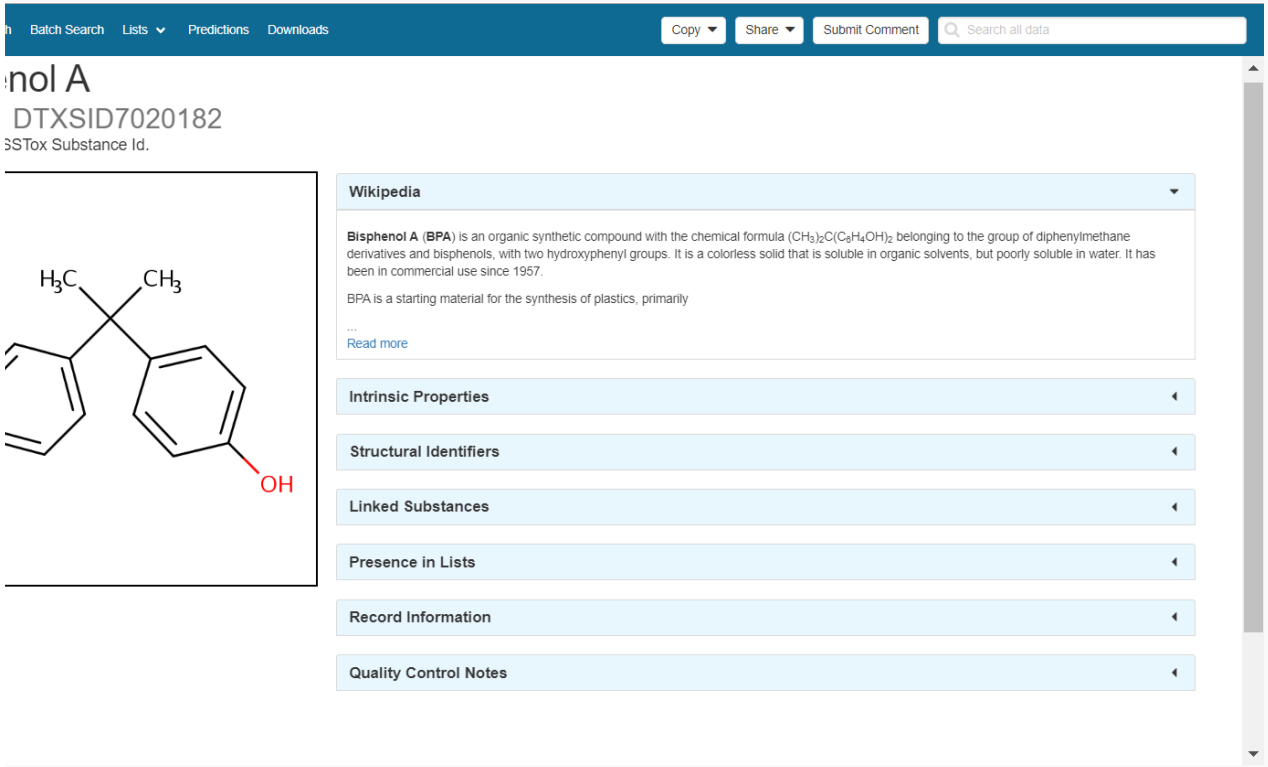
CompTox Dashboard
Assays and Genes



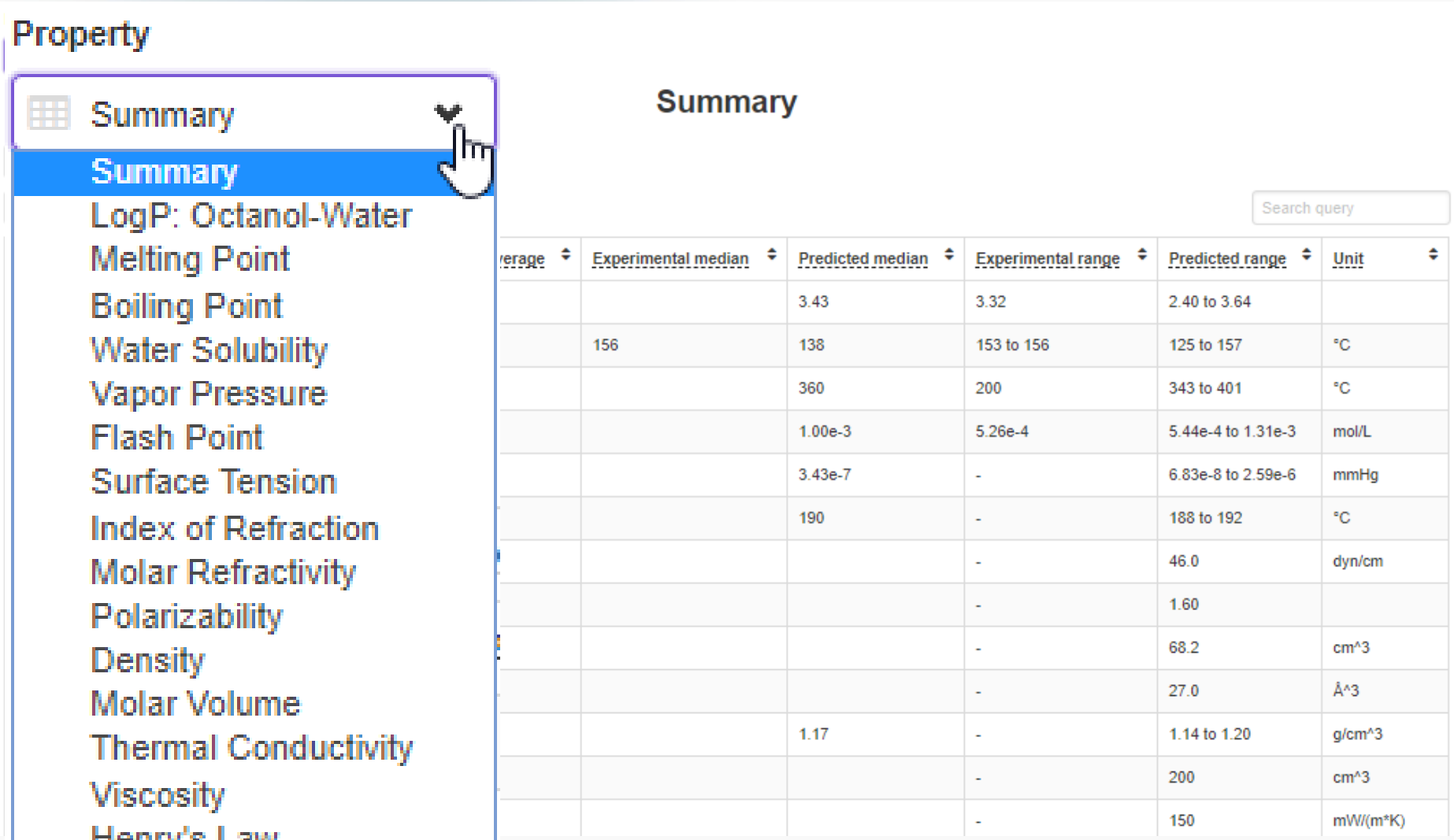
Detailed Chemical Pages

EPA United States Environmental Protection Agency

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



Physicochemical properties



OPERA Predicted Properties

EPA United States Environmental Protection Agency

An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling

K. Mansouri, C. M. Grulke, A. M. Richard, R. S. Judson & A. J. Williams

To cite this article: K. Mansouri, C. M. Grulke, A. M. Richard, R. S. Judson & A. J. Williams (2016)
An automate
datasets use
DOI: 10.108(

To link to th

Mansouri *et al. J Cheminform (2018) 10:10*
https://doi.org/10.1186/s13321-018-0263-1

Journal of Cheminformatics

RESEARCH ARTICLE Open Access

CrossMark

OPERA models for predicting physicochemical properties and environmental fate endpoints

Kamel Mansouri[1,2,3*], Chris M. Grulke[1], Richard S. Judson[1] and Antony J. Williams[1]

Detailed OPERA Prediction Reports

EPA United States Environmental Protection Agency

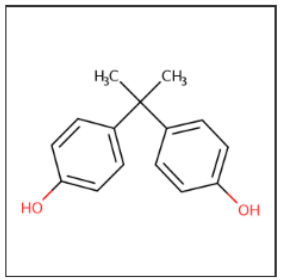
Source	Result	Calculation Details
Experimental Values		
PhysPropNCCT		
Predicted Values		
EPISUITE		
NICEATM		
ACD/Labs Conse		
ACD/Labs		
OPERA		

OPERA Models: LogP: Octanol-Water

Save PDF

Bisphenol A

80-05-7 | DTXSID7020182



Model Results

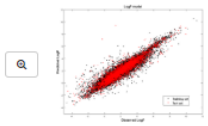
Predicted value: 3.35

Global applicability domain: Inside

Local applicability domain index: 0.88

Confidence level: 0.75

Model Performance

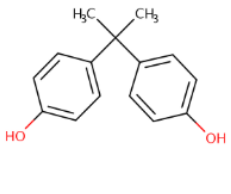


Weighted KNN model

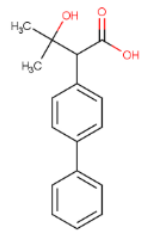
QMRF

5-fold CV (75%)		Training (75%)		Test (25%)	
Q2	RMSE	R2	RMSE	R2	RMSE
0.85	0.69	0.86	0.67	0.86	0.78

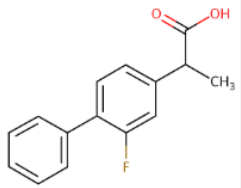
Nearest Neighbors from the Training Set



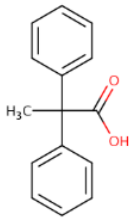
Bisphenol A
Measured: 3.32
Predicted: 3.35



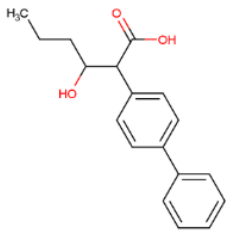
BUTANOIC ACID,2-(4-BIPHENYLYL)-3-HYDROX...
Measured: 3.25
Predicted: 3.45



Flurbiprofen
Measured: 4.16
Predicted: 3.83

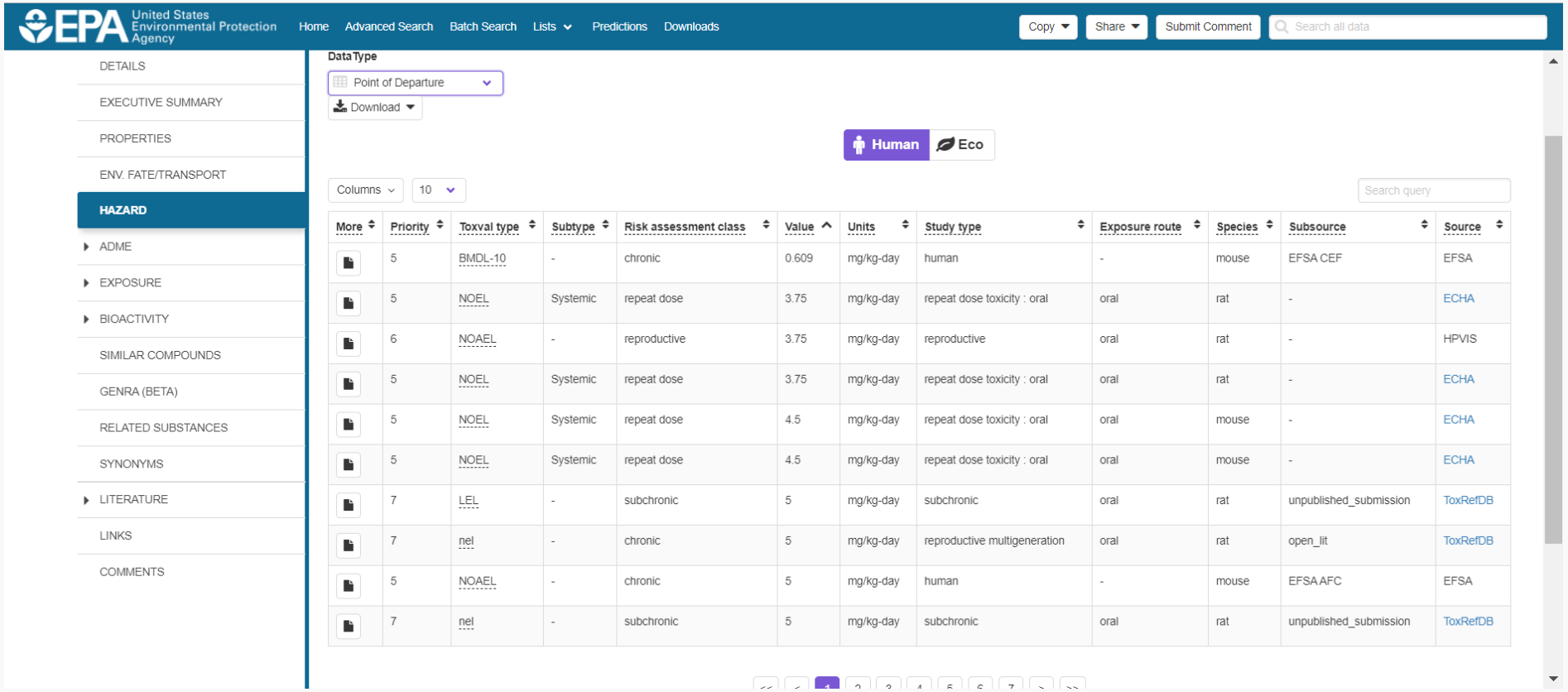


2,2-Diphenylpropionic acid
Measured: 2.69
Predicted: 2.93



3-OH-2-(4-BIPHENYLYL)HEXANOIC ACID
Measured: 3.75
Predicted: 3.68

Access to Chemical Hazard Data



More	Priority	Toxval type	Subtype	Risk assessment class	Value	Units	Study type	Exposure route	Species	Subsource	Source
	5	BMDL-10	-	chronic	0.609	mg/kg-day	human	-	mouse	EFSA CEF	EFSA
	5	NOEL	Systemic	repeat dose	3.75	mg/kg-day	repeat dose toxicity : oral	oral	rat	-	ECHA
	6	NOAEL	-	reproductive	3.75	mg/kg-day	reproductive	oral	rat	-	HPVIS
	5	NOEL	Systemic	repeat dose	3.75	mg/kg-day	repeat dose toxicity : oral	oral	rat	-	ECHA
	5	NOEL	Systemic	repeat dose	4.5	mg/kg-day	repeat dose toxicity : oral	oral	mouse	-	ECHA
	5	NOEL	Systemic	repeat dose	4.5	mg/kg-day	repeat dose toxicity : oral	oral	mouse	-	ECHA
	7	LEL	-	subchronic	5	mg/kg-day	subchronic	oral	rat	unpublished_submission	ToxRefDB
	7	nel	-	chronic	5	mg/kg-day	reproductive multigeneration	oral	rat	open_lit	ToxRefDB
	5	NOAEL	-	chronic	5	mg/kg-day	human	-	mouse	EFSA AFC	EFSA
	7	nel	-	subchronic	5	mg/kg-day	subchronic	oral	rat	unpublished_submission	ToxRefDB

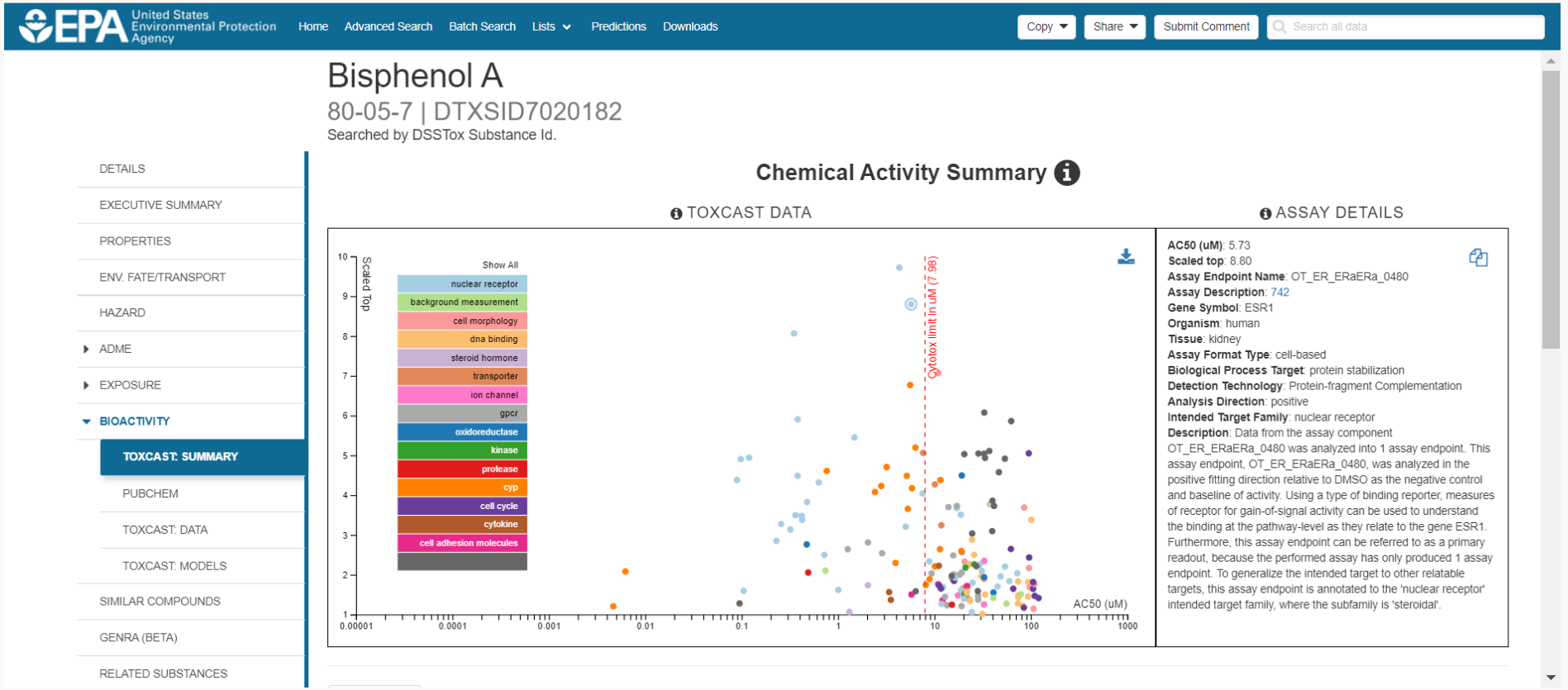
Hazard Data from “ToxVal_DB”

- 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

In Vitro Bioassay Screening

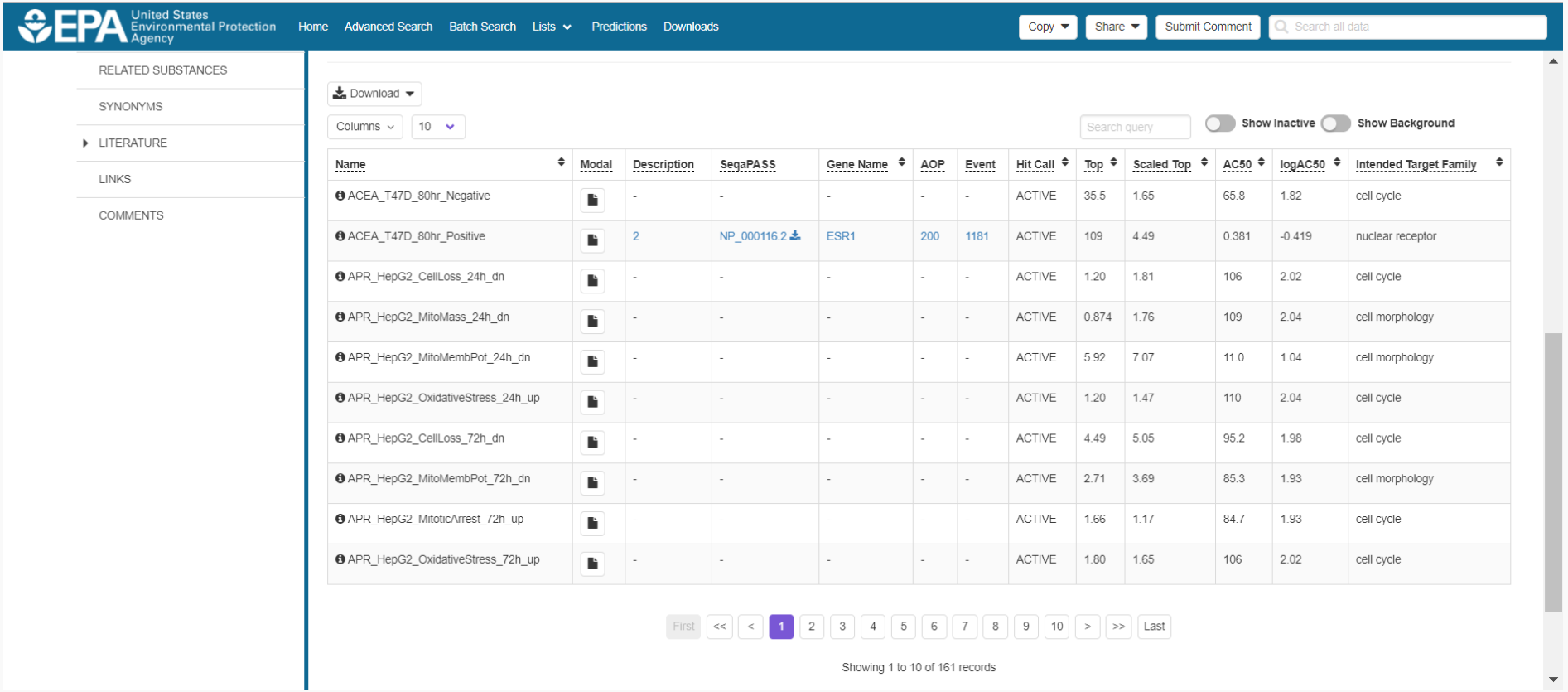
ToxCast and Tox21

EPA United States Environmental Protection Agency



In Vitro Bioassay Screening

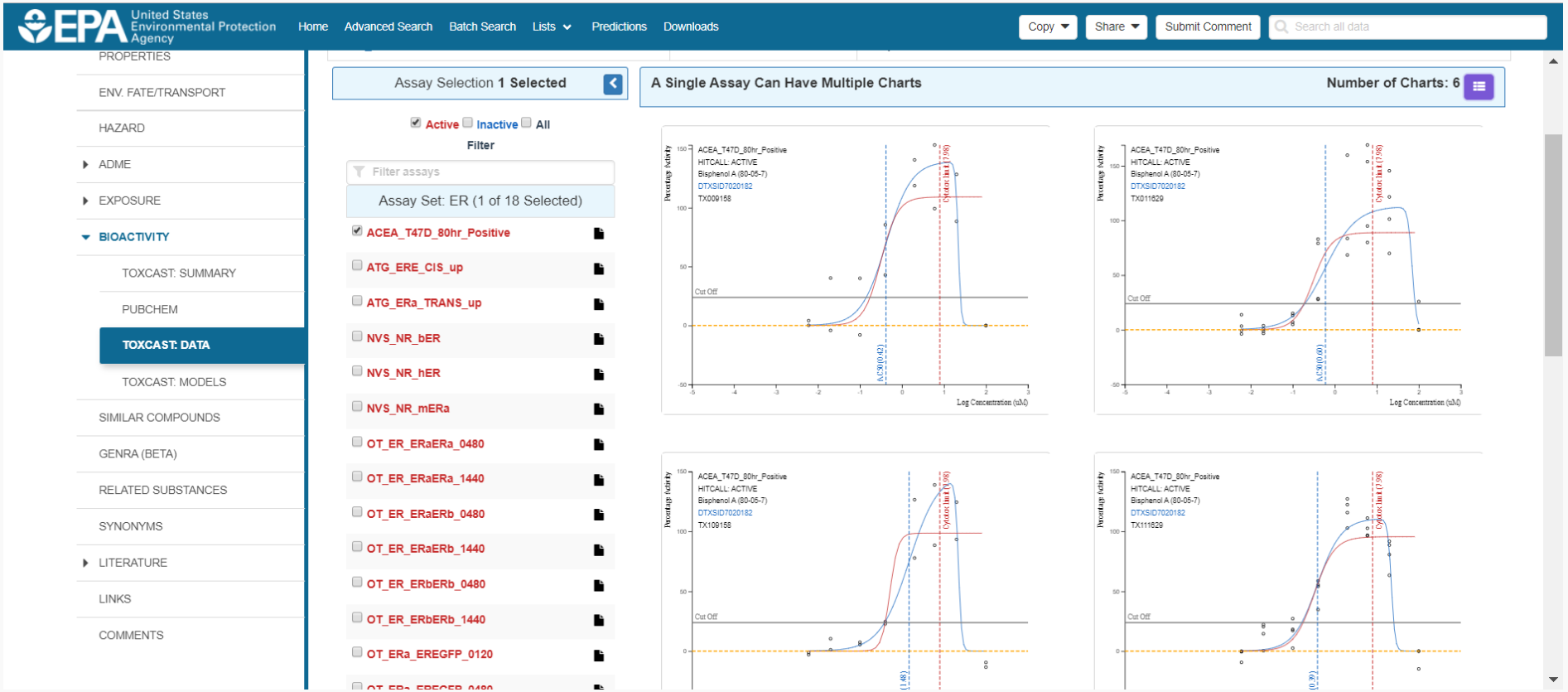
ToxCast and Tox21



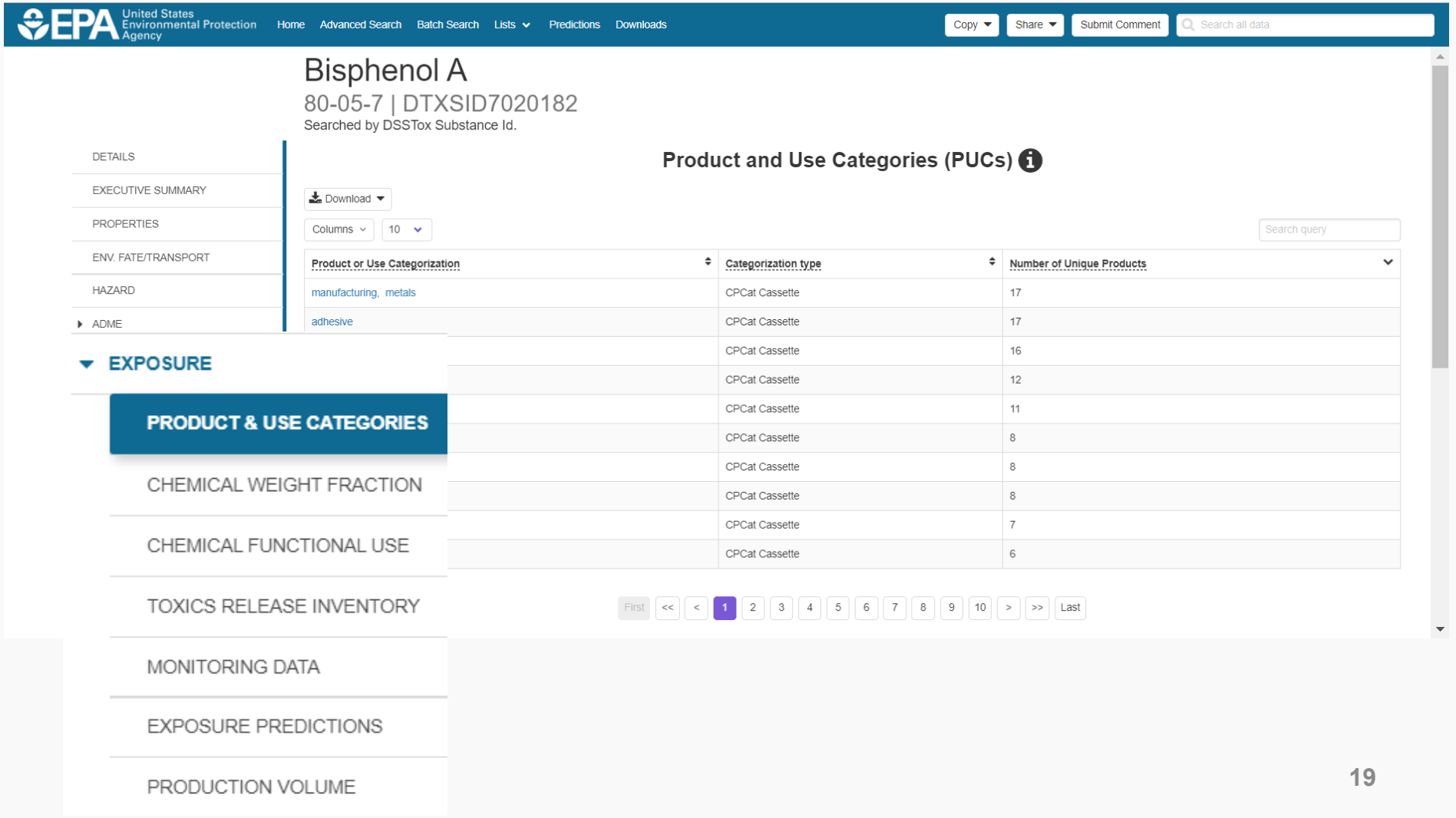
Name	Modal	Description	SegaPASS	Gene Name	AOP	Event	Hit Call	Top	Scaled Top	AC50	logAC50	Intended Target Family
ACEA_T47D_80hr_Negative		-	-	-	-	-	ACTIVE	35.5	1.65	65.8	1.82	cell cycle
ACEA_T47D_80hr_Positive		2	NP_000116.2	ESR1	200	1181	ACTIVE	109	4.49	0.381	-0.419	nuclear receptor
APR_HepG2_CellLoss_24h_dn		-	-	-	-	-	ACTIVE	1.20	1.81	106	2.02	cell cycle
APR_HepG2_MitoMass_24h_dn		-	-	-	-	-	ACTIVE	0.874	1.76	109	2.04	cell morphology
APR_HepG2_MitoMembPot_24h_dn		-	-	-	-	-	ACTIVE	5.92	7.07	11.0	1.04	cell morphology
APR_HepG2_OxidativeStress_24h_up		-	-	-	-	-	ACTIVE	1.20	1.47	110	2.04	cell cycle
APR_HepG2_CellLoss_72h_dn		-	-	-	-	-	ACTIVE	4.49	5.05	95.2	1.98	cell cycle
APR_HepG2_MitoMembPot_72h_dn		-	-	-	-	-	ACTIVE	2.71	3.69	85.3	1.93	cell morphology
APR_HepG2_MitoticArrest_72h_up		-	-	-	-	-	ACTIVE	1.66	1.17	84.7	1.93	cell cycle
APR_HepG2_OxidativeStress_72h_up		-	-	-	-	-	ACTIVE	1.80	1.65	106	2.02	cell cycle

In Vitro Bioassay Screening

ToxCast and Tox21



Sources of Exposure to Chemicals



Sources of Exposure to Chemicals

EPA United States Environmental Protection Agency

EPA United States Environmental Protection Agency | Home | Advanced Search | Batch Search | Lists | Predictions | Downloads | Copy | Share | Submit Comment | Search all data

Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- ADME
- **EXPOSURE**
  - PRODUCT & USE CATEGORIES
  - CHEMICAL WEIGHT FRACTION
  - CHEMICAL FUNCTIONAL USE
  - **TOXICS RELEASE INVENTORY**
  - MONITORING DATA
  - EXPOSURE PREDICTIONS
  - PRODUCTION VOLUME
- BIOACTIVITY

Toxics Release Inventory

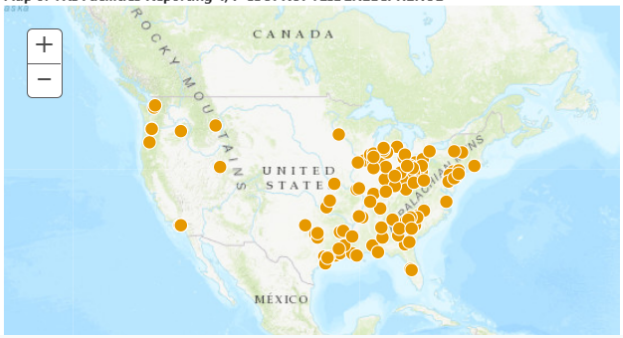
Print Page

2015 TRI Factsheet: Chemical - 4,4'-ISOPROPYLIDENEDIPHENOL, 000080057

Data Source: 2016 Dataset (released March 2018)

The Toxics Release Inventory (TRI) tracks the management of certain toxic chemicals that may pose a threat to human health and the environment. Certain industrial facilities in the U.S. must report annually how much of each chemical is recycled, combusted for energy recovery, treated for destruction, and disposed of or otherwise released on- and off-site. This information is collectively referred to as production-related waste managed.

Map of TRI Facilities Reporting 4,4'-ISOPROPYLIDENEDIPHENOL

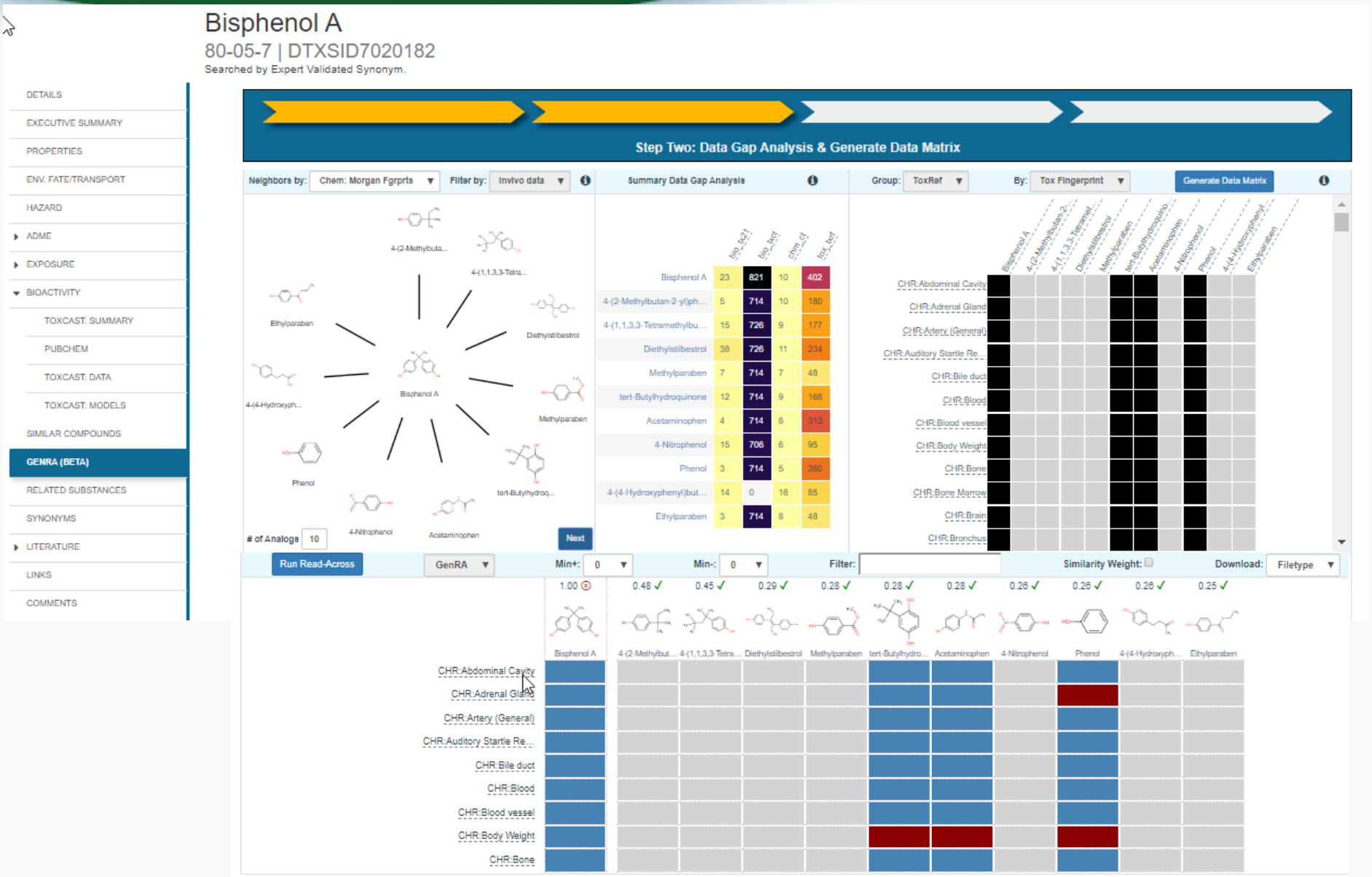


Quick Facts for 2015

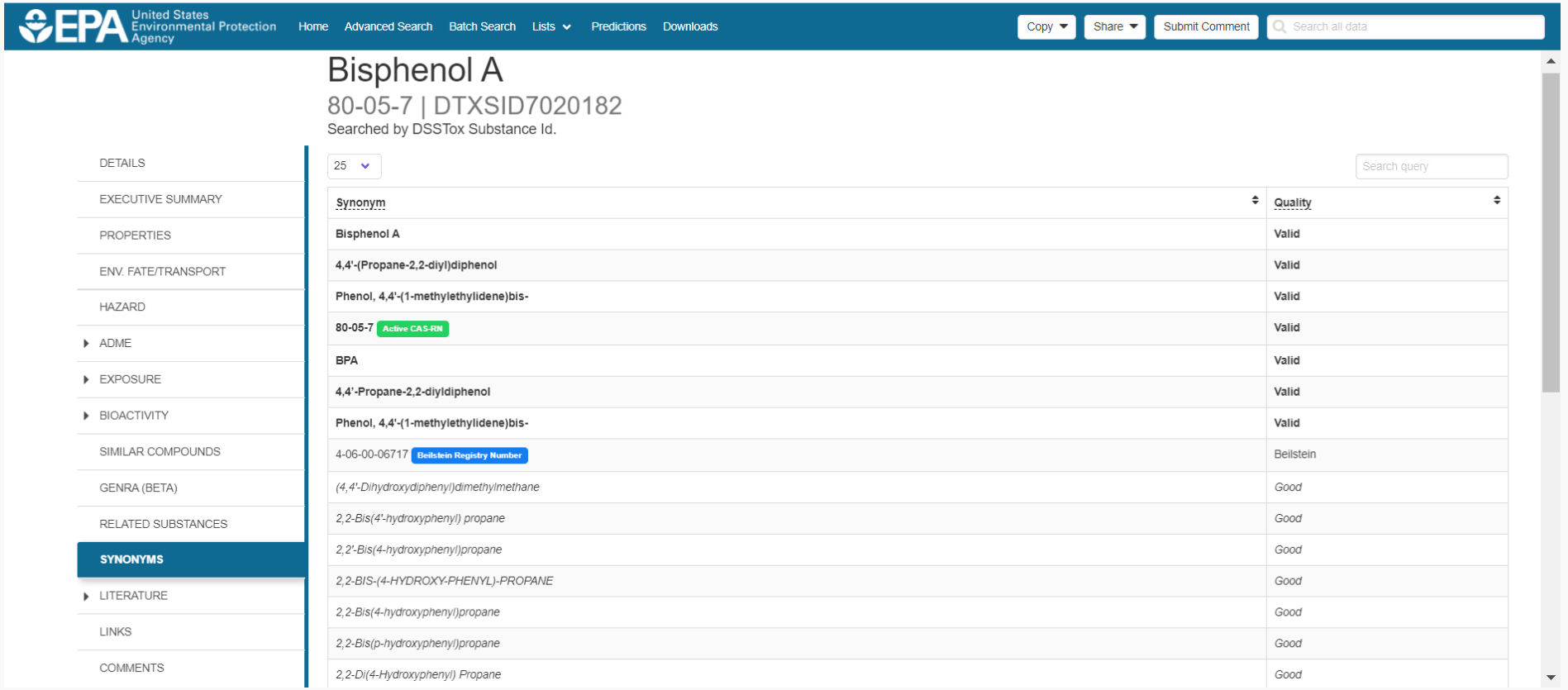
	Chemical	United States
Number of TRI Facilities:	120	22,130
Total Production-Related Waste Managed:	15.8 million lbs	27.1 billion lbs
Total On-site and Off-site Disposal or Other Releases:	2.5 million lbs	3.4 billion lbs
Total On-site:	39.4 thousand lbs	2.9 billion lbs
• Air:	28.7 thousand lbs	686.4 million lbs
• Water:	4.4 thousand lbs	198.2 million lbs
• Land:	6.2 thousand lbs	2.0 billion lbs

Generalized Read-Across (GenRA)

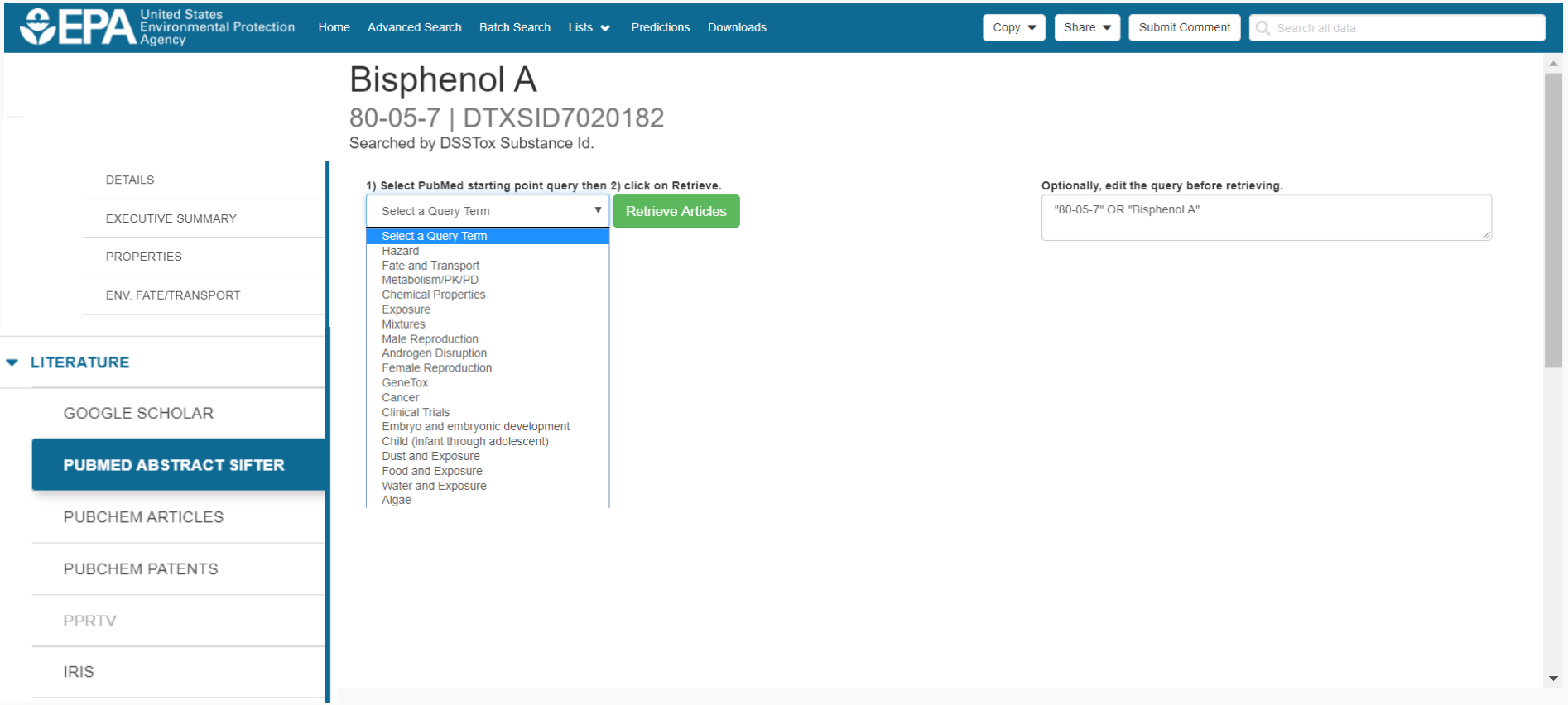
EPA United States Environmental Protection Agency



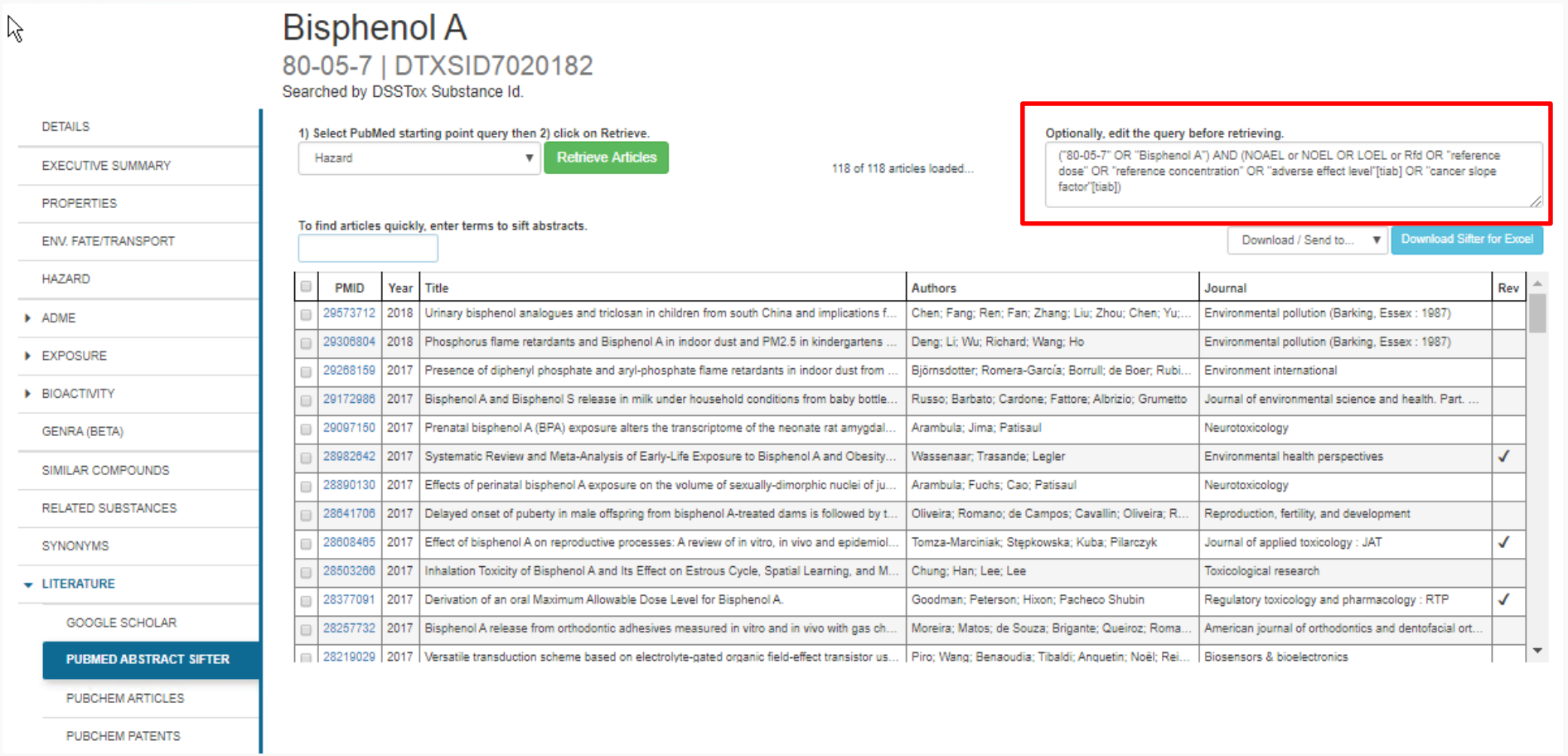
Identifiers to Support Searches



Literature Searches and Links



Abstract Sifter – PubMed Integration



Abstract Sifter for Excel

EPA United States Environmental Protection Agency

F1000Research Open for Science

Search

BROWSE GATEWAYS HOW TO PUBLISH ABOUT BLOG

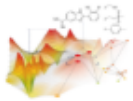
Check for updates

SOFTWARE TOOL ARTICLE

Abstract Sifter: a comprehensive front-end system to PubMed [version 1; referees: 2 approved]

Nancy Baker [1], Thomas Knudsen[2], Antony Williams [2]

Author details



This article is included in the Chemical Information Science gateway.

METRICS

629 VIEWS

118 DOWNLOADS

Get PDF

Get XML

Cite

Export

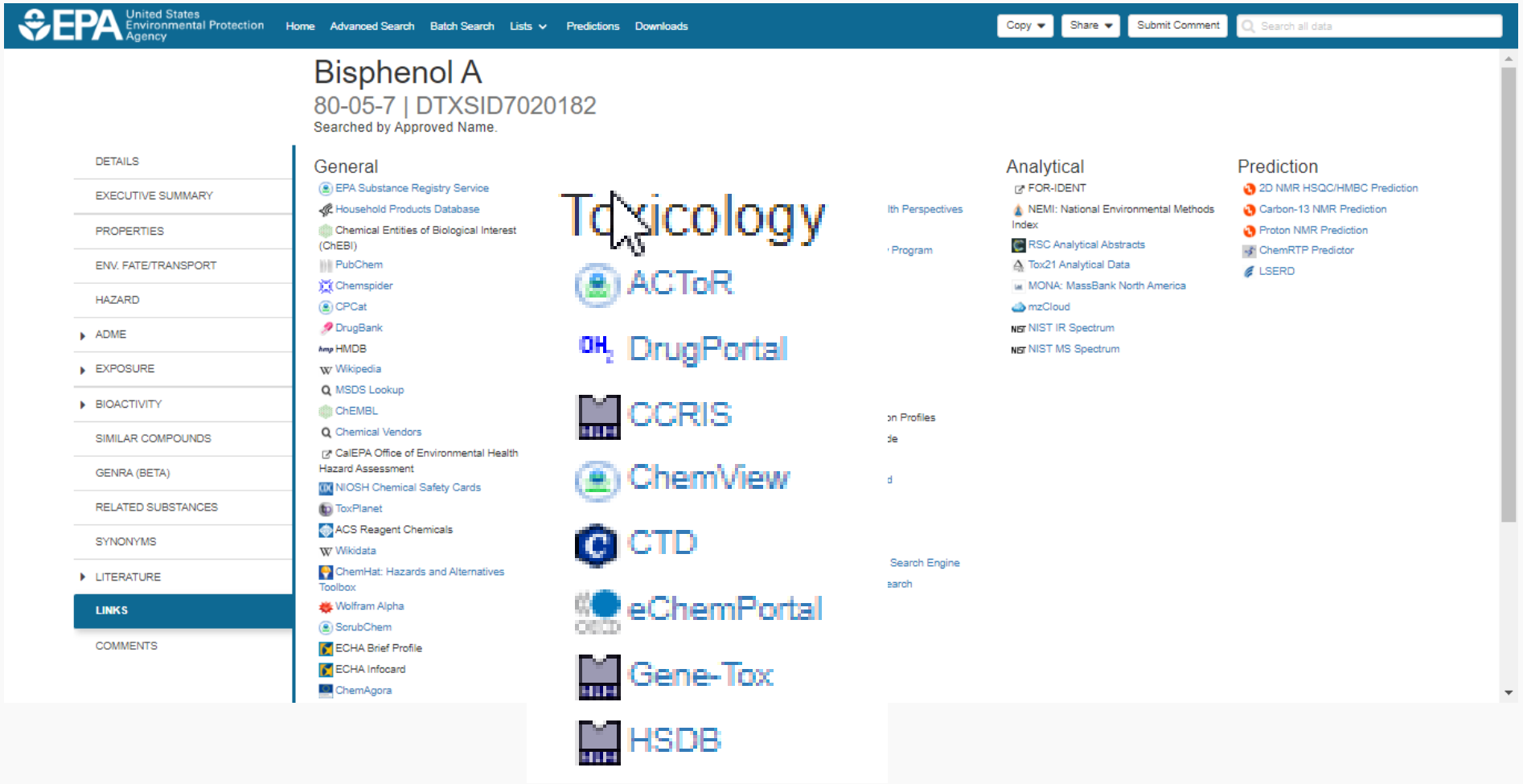
Track

Email

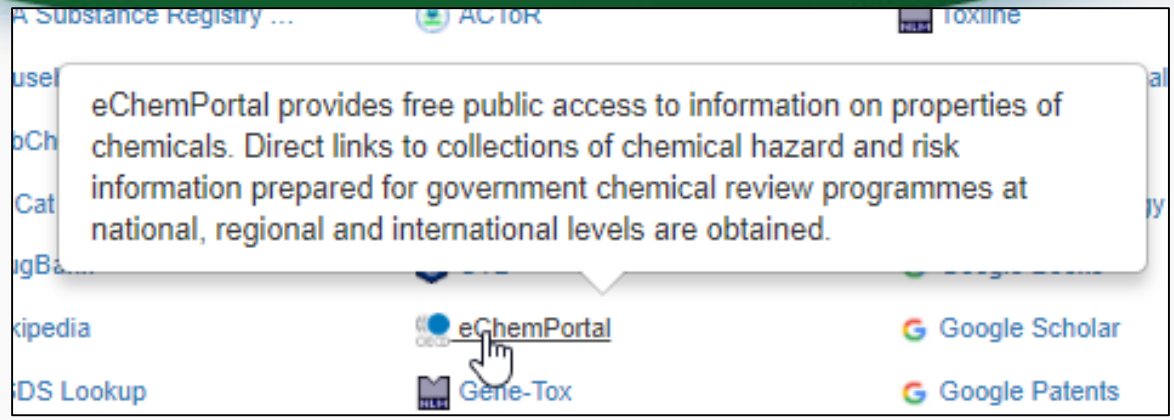
Abstract

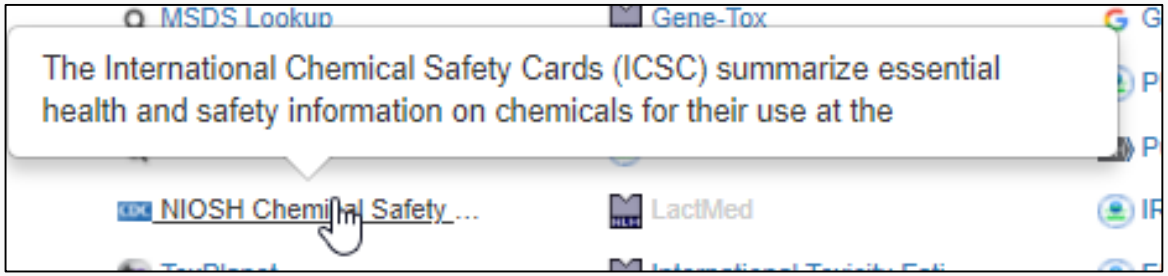
The Abstract Sifter is a Microsoft Excel based application that enhances existing search capabilities of PubMed. The Abstract Sifter assists researchers to search effectively, triage results, and keep track of articles of interest. The tool implements an innovative "sifter" functionality for relevance ranking, giving the researcher a way to find articles of interest quickly. The tool also gives

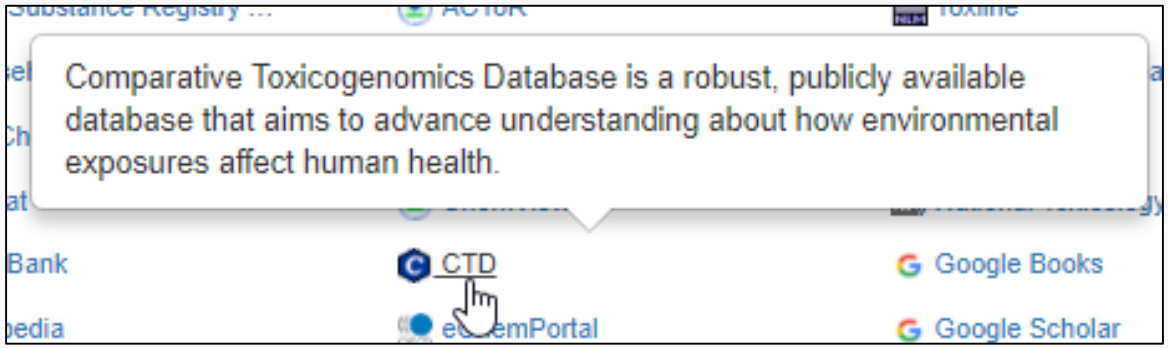
External Links to ~80 websites



Integrated Linkouts



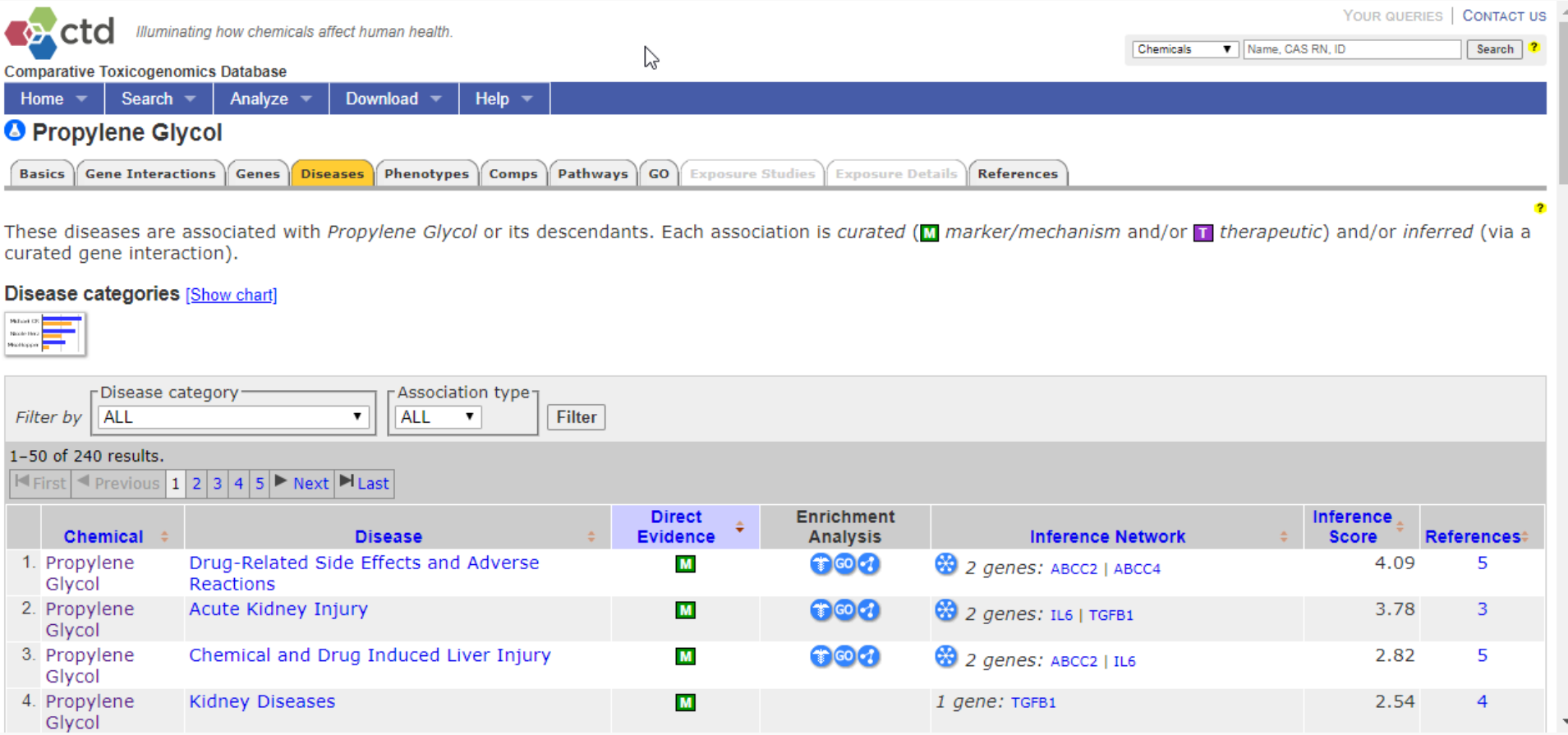




Integrated Linkouts

Comparative Toxicogenomics DB

EPA United States Environmental Protection Agency



	Chemical	Disease	Direct Evidence	Enrichment Analysis	Inference Network	Inference Score	References
1.	Propylene Glycol	Drug-Related Side Effects and Adverse Reactions	M	GO	2 genes: ABCC2 | ABCC4	4.09	5
2.	Propylene Glycol	Acute Kidney Injury	M	GO	2 genes: IL6 | TGFB1	3.78	3
3.	Propylene Glycol	Chemical and Drug Induced Liver Injury	M	GO	2 genes: ABCC2 | IL6	2.82	5
4.	Propylene Glycol	Kidney Diseases	M		1 gene: TGFB1	2.54	4

Not just chemical “structures”

- Chemicals in commerce, of interest to the EPA, are not all easily represented by structures
- Different chemical substances supported
  - Chemical structures
  - “UVCB chemicals” - Unknown or Variable Composition, Complex Reaction Products and Biological Materials
  - Metabolites and transformation products
  - Homologous series as Markush Structures
  - Curated classes of chemicals

UVCB Chemicals

(C10-C16) Alkylbenzenesulfonic acid

68584-22-5 | DTXSID2028723

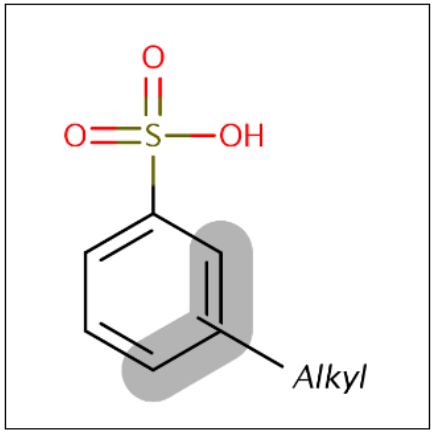
Searched by DSSTox Substance Id.

DETAILS

RELATED SUBSTANCES

PROPERTIES

COMMENTS



Intrinsic Properties

Presence in Lists

Federal

Safer Choice Chemical List | TOX21SL: Tox21 Screening Library | TSCA Workplan Step 2 Chemicals | EPAHFR - EPA Chemicals associated with hydraulic fracturing

US State

None.

International

OLEM RapidTox Chemicals | REACH Dossier Chemicals

Other

Surfactant List Screened in Swiss Wastewater (2014) | EDSP Universe | TSCAACTIVE | EPA Chemicals associated with hydraulic fracturing

Markush Structures

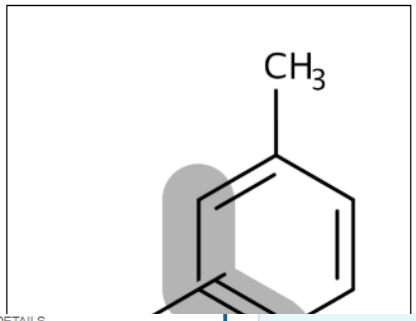
EPA United States Environmental Protection Agency

Xylenes

1330-20-7 | DTXSID2021446

Searched by DSSTox Substance Id.

DETAILS
RELATED SUBSTANCES
PROPERTIES
COMMENTS



Wikipedia

Xylene (from Greek ξύλο, *xylo*, "wood"), **xylol** or **dimethylbenzene** is any one of three isomers of dimethylbenzene, or a combination thereof. With the formula $(CH_3)_2C_6H_4$, each of the three compounds has a central benzene ring with two methyl groups attached at substituents. They are all colorless, flammable liquids, some of which are of great industrial value. The mixture is referred to as both xylene and, more precisely, xylenes

...
Read more

Intrinsic Properties

Presence in Lists

Record Information

DETAILS
RELATED SUBSTANCES
PROPERTIES
COMMENTS

8 chemicals

Download / Send

Show info: DTXSID × CASRN ×

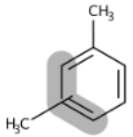
Select all

Sort by: Relationship

Filter by: Name or CASRN

Hide

Searched Chemical



Xylenes
DTXSID: DTXSID2021446
CASRN: 1330-20-7

Predecessor: Component

4 related chemical structures with this substance

Xylenes; defined mixture 1
DTXSID: DTXSID9021421
CASRN: NOCAS_21421

Predecessor: Component

4 related chemical structures with this substance

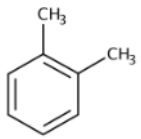
Xylenes; defined mixture 2
DTXSID: DTXSID7021447
CASRN: NOCAS_21447

Predecessor: Component

5 related chemical structures with this substance

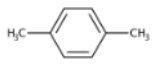
Total Petroleum Hydrocarbons (TPH)
DTXSID: DTXSID30891529
CASRN: NOCAS_891529

Markush Child



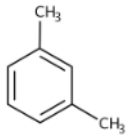
o-Xylene
DTXSID: DTXSID3021807
CASRN: 95-47-6

Markush Child



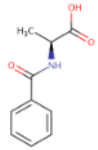
p-Xylene
DTXSID: DTXSID2021868
CASRN: 106-42-3

Markush Child



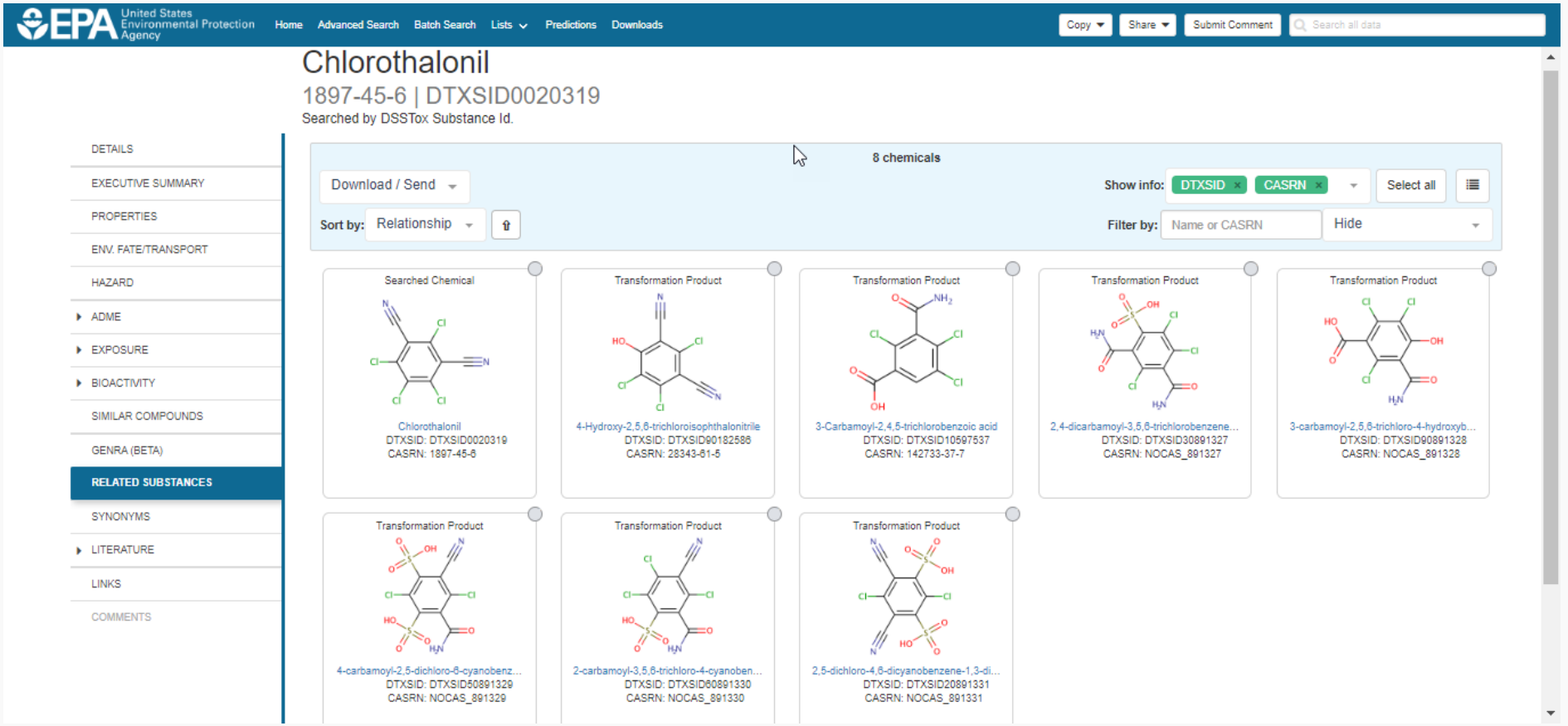
m-Xylene
DTXSID: DTXSID6026298
CASRN: 108-38-3

Transformation Product



N-Benzoylalanine
DTXSID: DTXSID40176394
CASRN: 2198-64-3

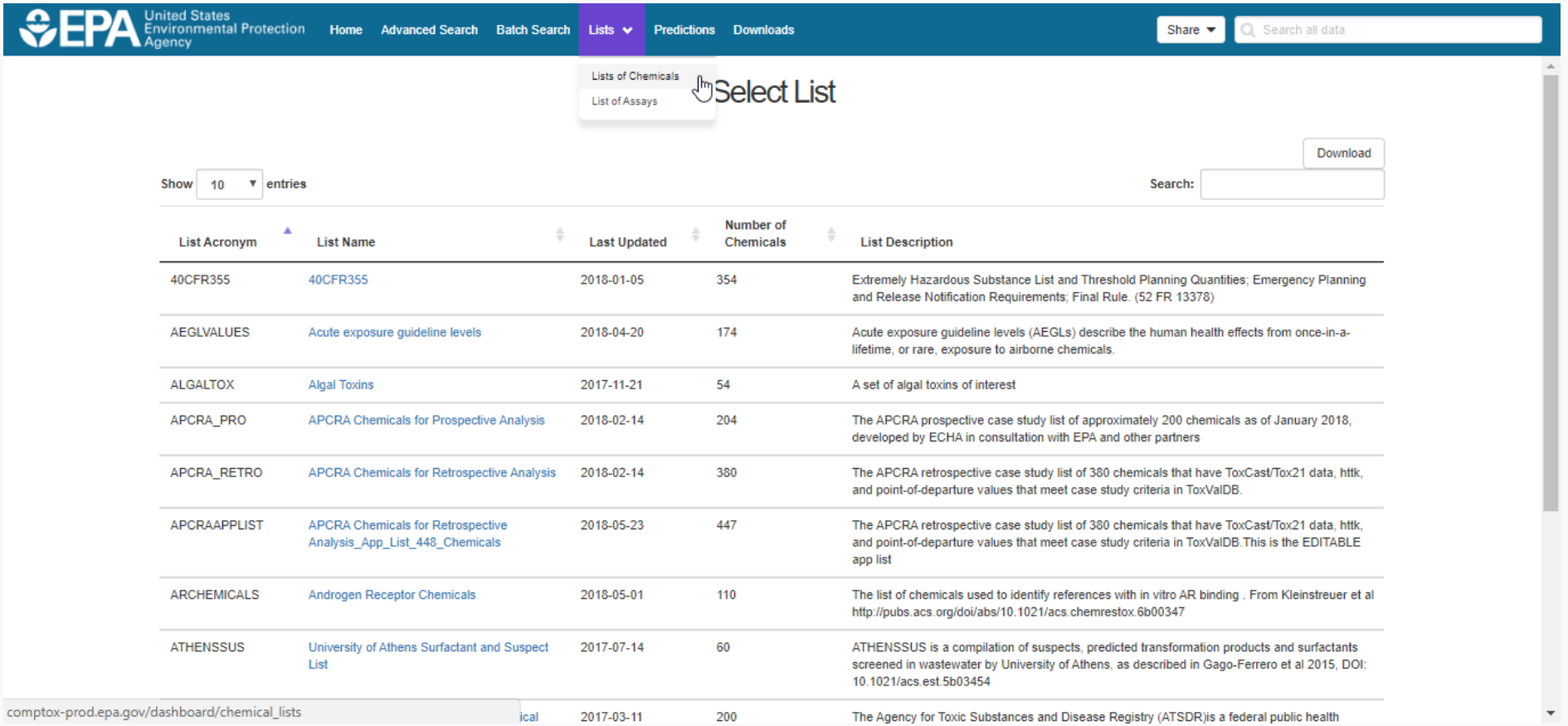
Transformation Products



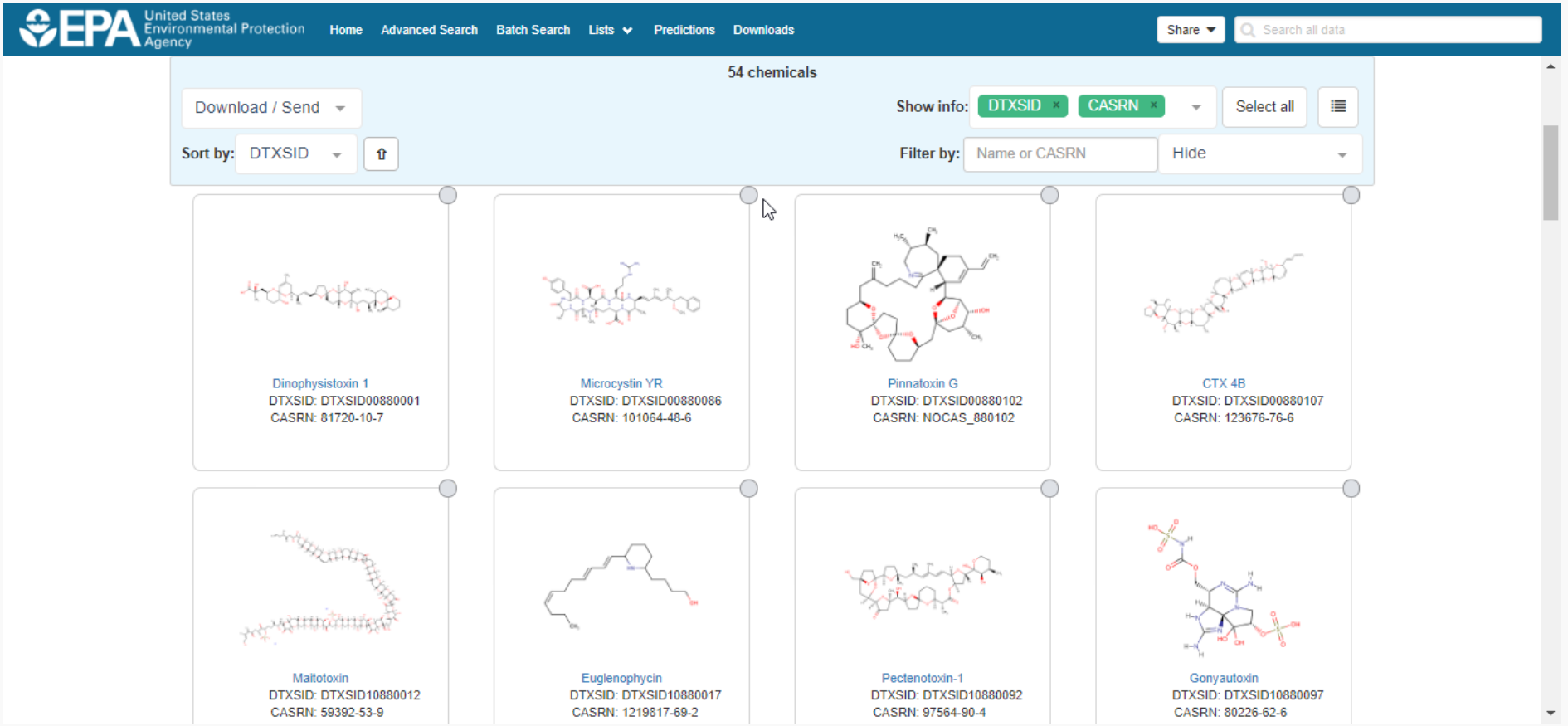
Not just chemical “structures”

- Different chemical substances supported
  - Chemical structures
  - “UVCB chemicals” - Unknown or Variable Composition, Complex Reaction Products and Biological Materials
  - Metabolites and transformation products
  - Homologous series as Markush Structures
  - Curated classes of chemicals
- Lists of chemicals
  - Submitted lists of chemicals – Federal, State, International and other general lists
  - Growing lists to support specific projects – e.g. ToxCast phases, algal toxins, our publication datasets

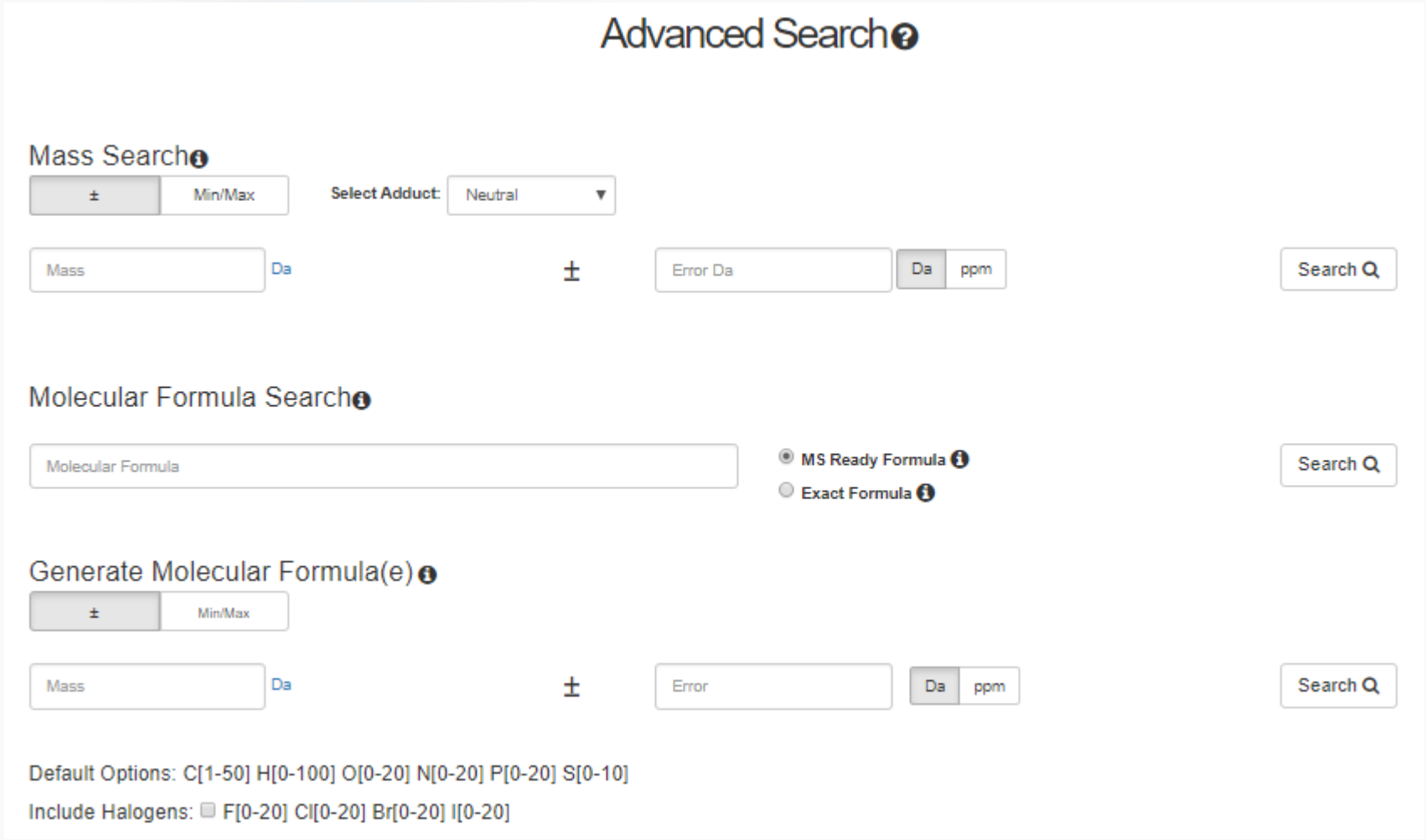
List of Chemicals



Algal Toxins



Mass and Formula Searches Supporting Mass Spectrometry



Advanced Searches
Mass Based Search

EPA United States Environmental Protection Agency

Mass Search

± | Min/Max

M

191.131 Da ± 5 Da | ppm

Search

EPA United States Environmental Protection Agency | Home | Advanced Search | Batch Search | Lists | Predictions | Downloads | Share | Search all data

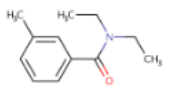
Search Results

Searched by Mass: 191.131 +/- 5.0 ppm.

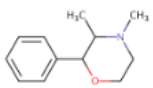
298 of 298 chemicals visible

Download / Send | Show info: DTXSID × CASRN × TOXCAST × Mass Diff × | Select all

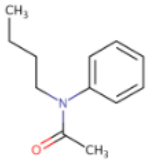
Sort by: Mass Difference | Filter by: Name or CASRN | Multicomponent Chemicals ×



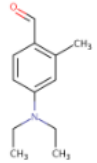
DEET
DTXSID: DTXSID2021995
CASRN: 134-62-3
TOXCAST: 14/663
Mass Diff: 0.000014



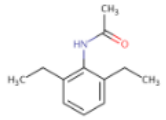
Phendimetrazine
DTXSID: DTXSID1023447
CASRN: 634-03-7
TOXCAST: 0
Mass Diff: 0.000014



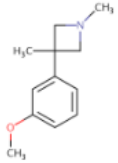
N-Butylacetanilide
DTXSID: DTXSID2042197
CASRN: 91-49-6
TOXCAST: 0
Mass Diff: 0.000014



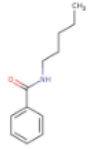
Benzaldehyde, 4-(diethylamino)-2-methyl-
DTXSID: DTXSID4059041
CASRN: 92-14-8
TOXCAST: 0
Mass Diff: 0.000014



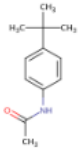
Acetanilide, 2',6'-diethyl-
DTXSID: DTXSID90168148
CASRN: 16665-89-7
TOXCAST: 0
Mass Diff: 0.000014



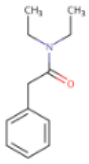
Azetidine, 1,3-dimethyl-3-(m-methoxyp...
DTXSID: DTXSID40173560
CASRN: 19832-26-9
TOXCAST: 0
Mass Diff: 0.000014



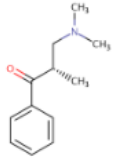
Benzamide, N-pentyl-
DTXSID: DTXSID20174196
CASRN: 20308-43-4



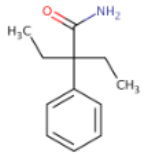
p-t-Butylacetanilide
DTXSID: DTXSID80174238
CASRN: 20330-45-4



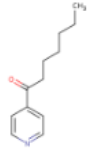
N,N-Diethylphenylacetamide
DTXSID: DTXSID00179048
CASRN: 2431-96-1



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

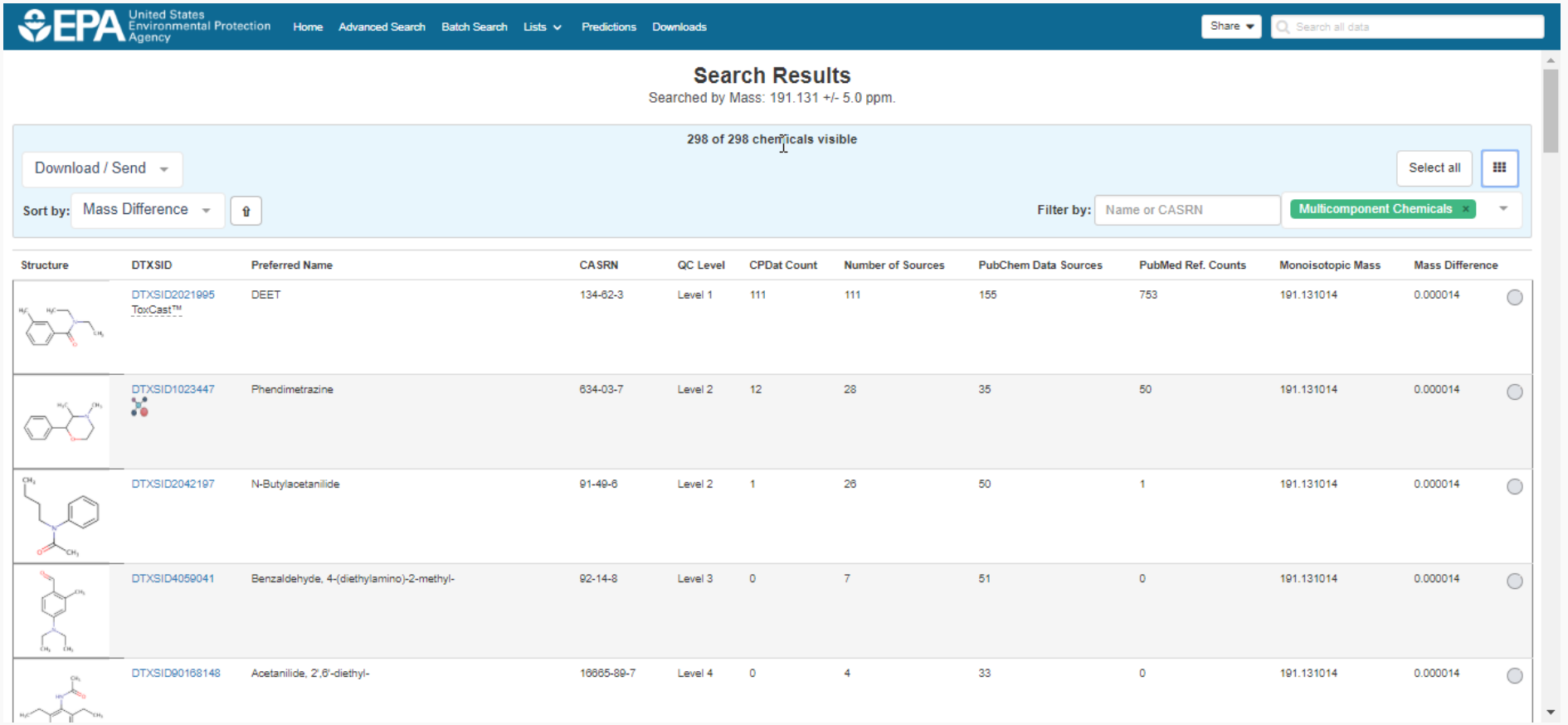


Butyramide, 2-ethyl-2-phenyl-
DTXSID: DTXSID60184653
CASRN: 30568-39-9



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

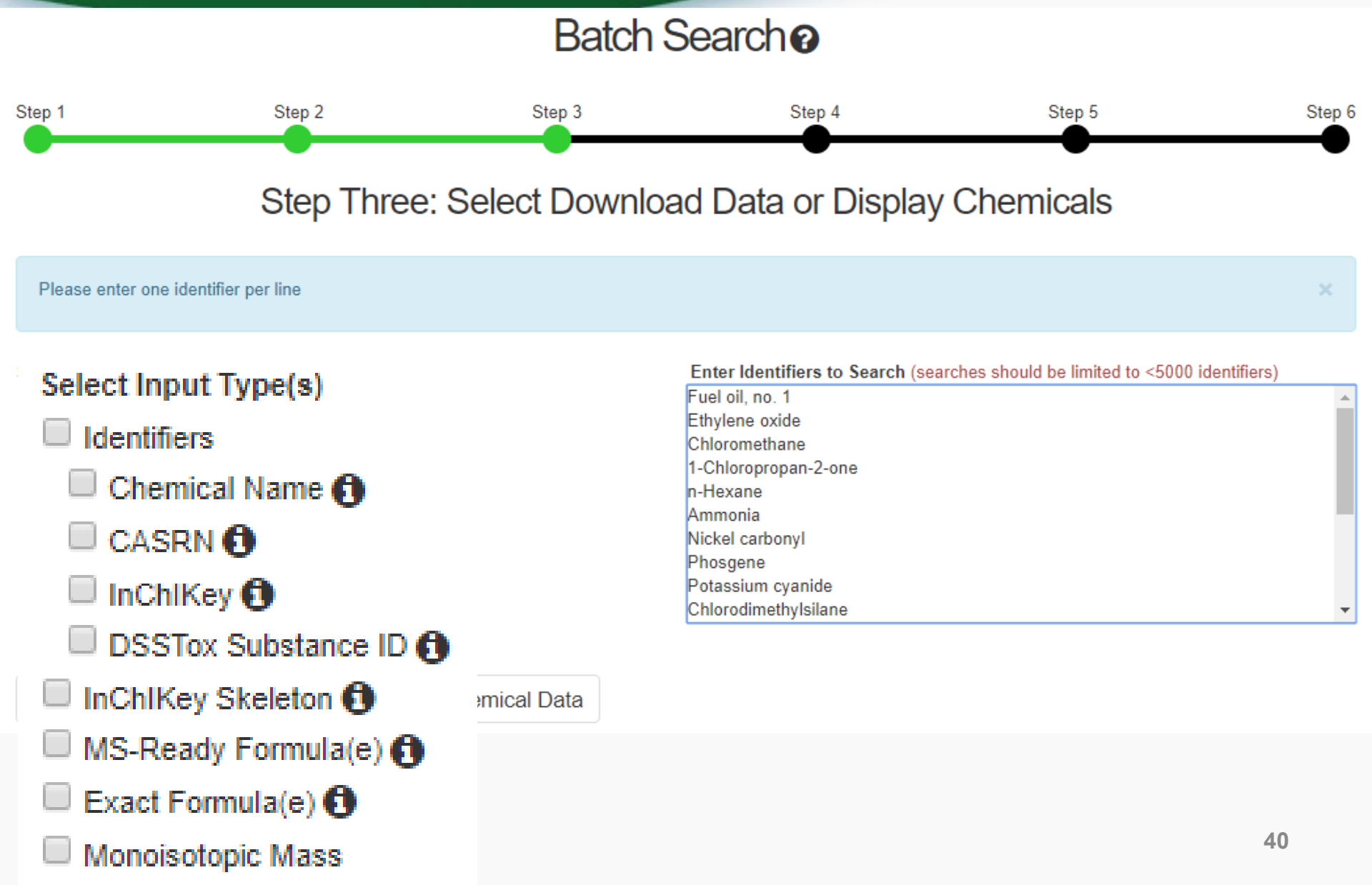
Advanced Searches
Mass Based Search



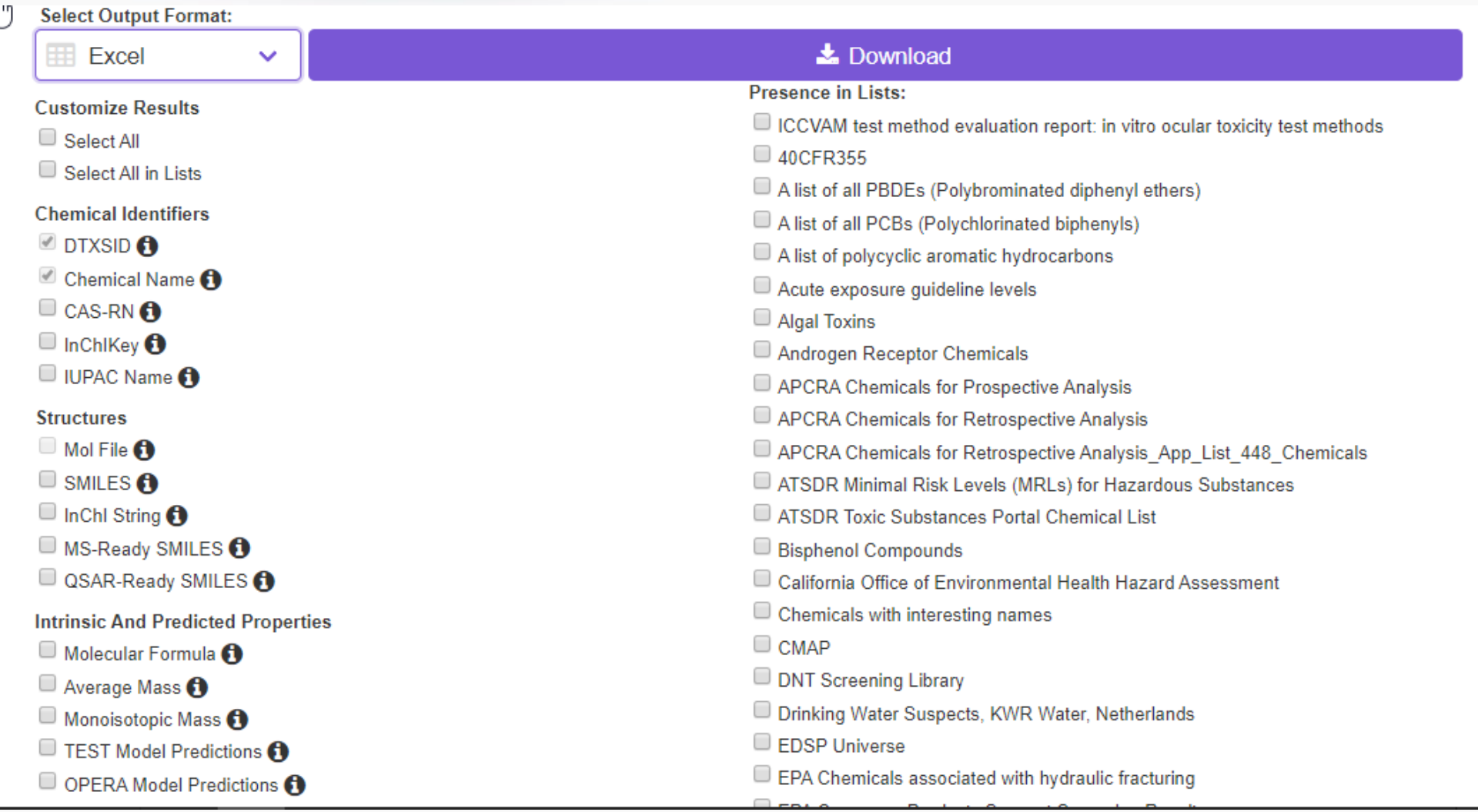
Batch Searching

- Singleton searches are useful but we work with thousands of chemicals!
- Typical questions
  - What are the SMILES strings for a list of 1000 chemicals?
  - Do any of this list of chemicals have XXX type of data?
  - What are the predicted logP values for a list of chemicals?
  - Can I get chemical lists in Excel files? In SDF files?

Batch Searching



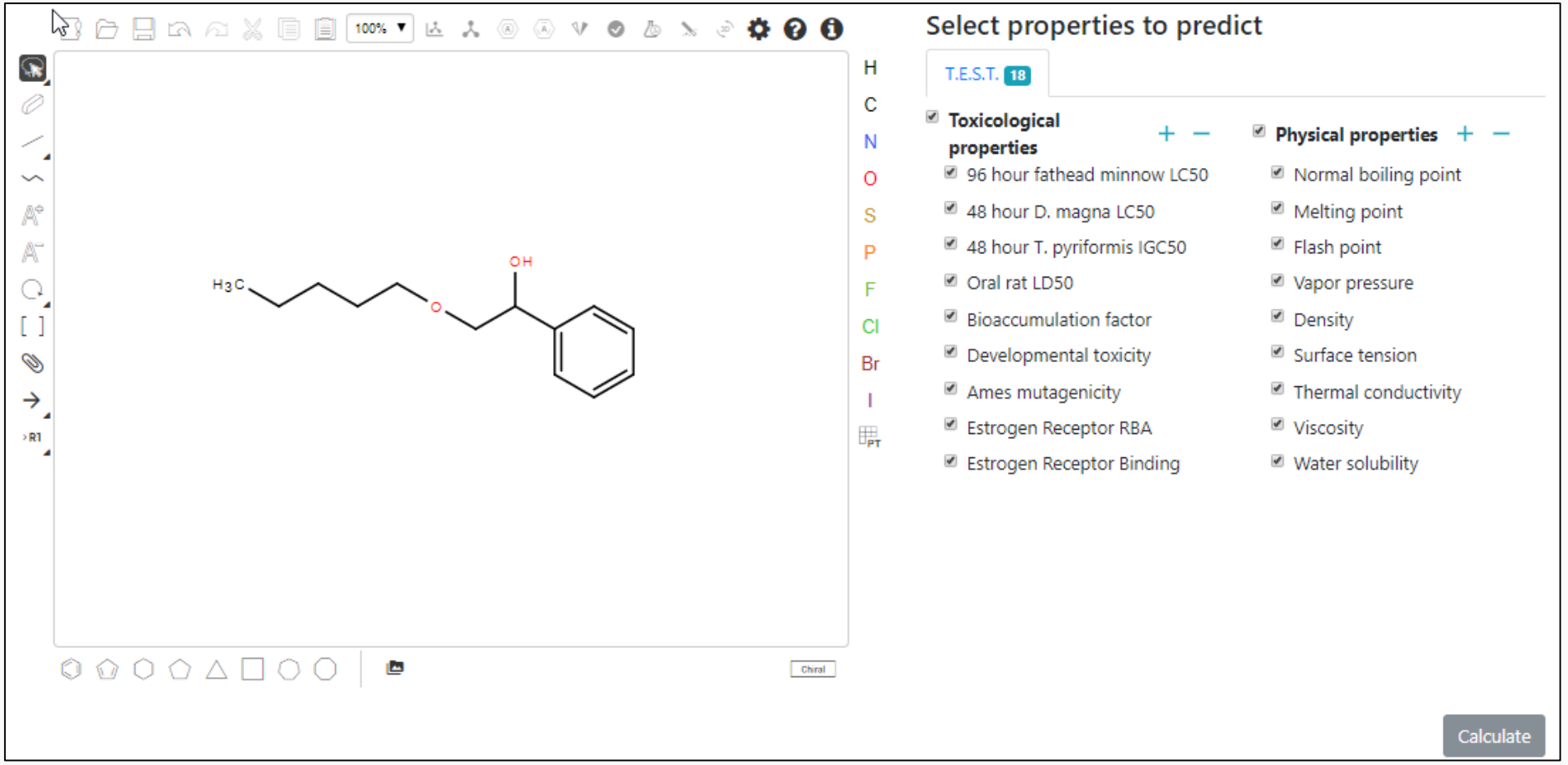
Batch Searching



Excel Output

INPUT	FOUND_BY	DTXCID_IN	DATA_SOU	TOXVAL_D	TOXCAST_	TOXCAST_	NUMBER_C	PUBCHEM_	STO
C6H12O3	MS Ready I	DTXCID701	91	Y	0.36	2/562	24	83	Y
C6H12O3	MS Ready I	DTXCID003	67	Y	0.36	1/276	376	80	Y
C6H12O3	MS Ready I	DTXCID106	65	Y	4.42	5/113	6	77	Y
C6H12O3	MS Ready I	DTXCID105	45	Y	0.0	0/163	3	94	-
C6H12O3	MS Ready I	DTXCID901	38	Y	-	-	14	110	Y
C6H12O3	MS Ready I	DTXCID402	34	Y	0.0	0/113	-	53	Y
C6H12O3	MS Ready I	DTXCID202	31	Y	-	-	-	36	Y
C6H12O3	MS Ready I	DTXCID202	30	-	2.54	7/276	-	54	-
C6H12O3	MS Ready I	DTXCID109	26	Y	-	-	-	46	-
C6H12O3	MS Ready I	DTXCID202	24	Y	0.0	0/113	-	47	-
C6H12O3	MS Ready I	DTXCID303	22	Y	-	-	-	89	-
C6H12O3	MS Ready I	DTXCID302	20	Y	-	-	2	25	Y
C6H12O3	MS Ready I	DTXCID407	19	Y	-	-	12	62	-
C6H12O3	MS Ready I	DTXCID704	17	Y	-	-	-	64	-
C6H12O3	MS Ready I	DTXCID704	16	Y	-	-	3	49	-

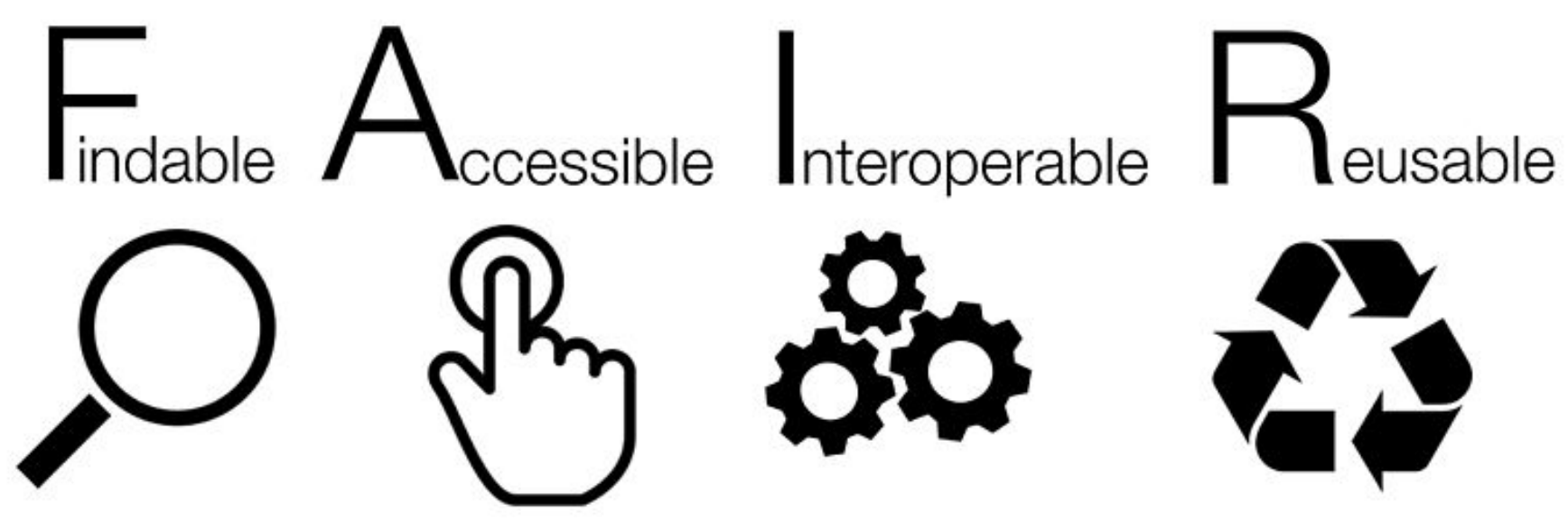
Real-Time Predictions



Real-Time Predictions

Property	Experimental Value	Prediction				
		Consensus	Hierarchical clustering	Single model	Group contribution	Nearest neighbor
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

Our support for FAIR Data



Downloadable Data

EPA United States Environmental Protection Agency

Home | Advanced Search | Batch Search | Lists | Predictions | Downloads | Share | Search all data

DSSTox Identifier to PubChem Identifier Mapping File — Posted: 11/14/2016

The DSSTox to PubChem Identifiers mapping file is in TXT format and includes the PubChem SID, PubChem CID and DSSTox substance identifier (DTXSID).

SID	CID	DTXSID
316388891	20404	DTXSID30873143
316388890	10142816	DTXSID70873142
316388889	50742127	DTXSID40873139
316388888	19073841	DTXSID20873137
316388887	11505215	DTXSID00873135
316388886	25021861	DTXSID80873133
316388885	2784427	DTXSID60873131
316388884	6731	DTXSID00873130

DSSTox identifiers mapped to CAS Numbers and Names File — Posted: 11/14/2016

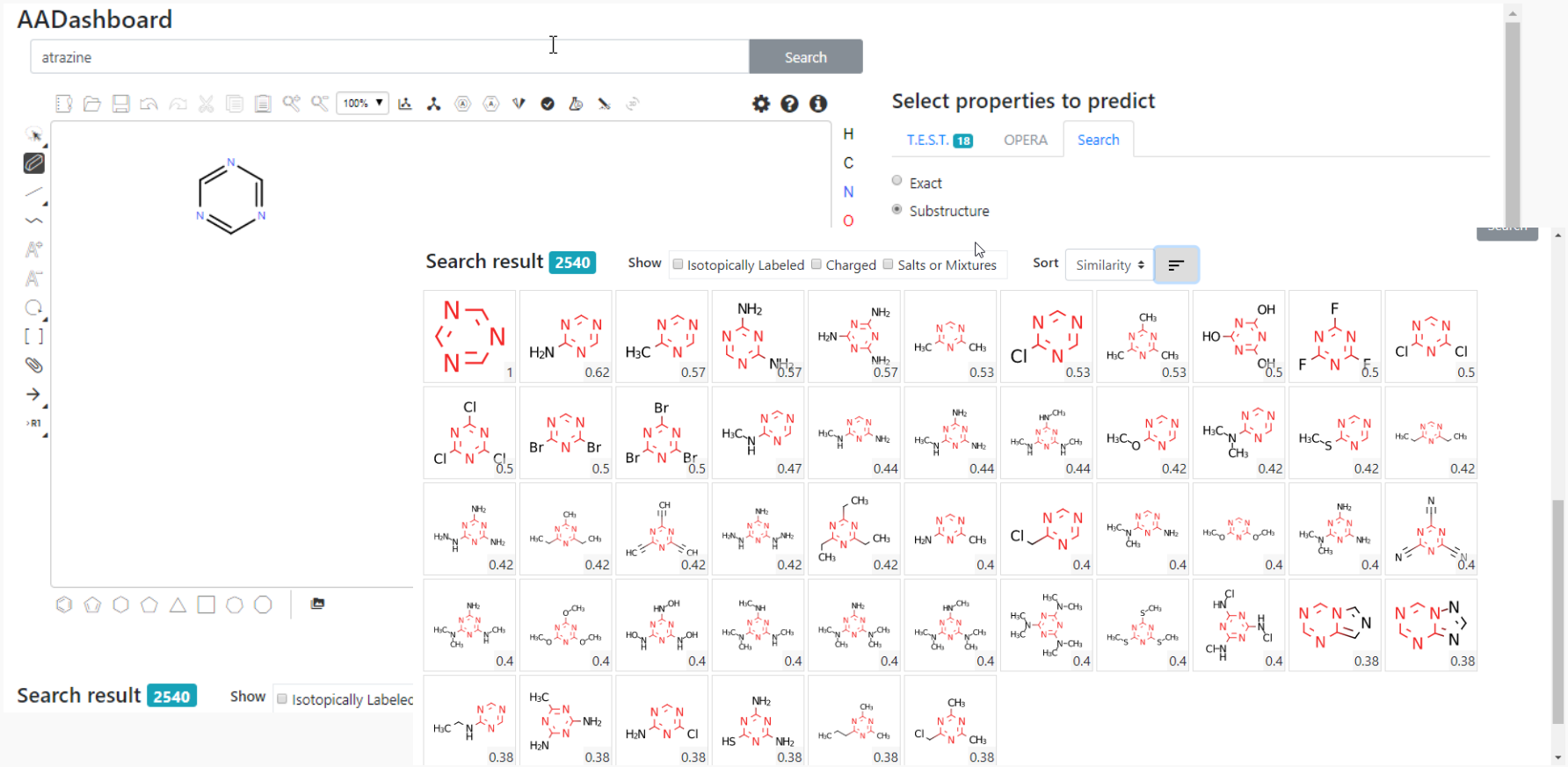
The DSSTox Identifiers file is in Excel format and includes the CAS Number, DSSTox substance identifier (DTXSID) and the Preferred Name.

	casrn	dsstox_substance_id	preferred_name
2	26148-68-5	DTXSID7020001	A-alpha-C
3	107-29-9	DTXSID2020004	Acetaldehyde oxime
4	60-35-5	DTXSID7020005	Acetamide
5	103-90-2	DTXSID2020006	Acetaminophen
6	968-81-0	DTXSID7020007	Acetohexamide
7	18523-69-8	DTXSID2020008	Acetone[4-(5-nitro-2-furyl)-2-thiazolyl] hydrazone
8	75-05-8	DTXSID7020009	Acetonitrile
9	127-06-0	DTXSID6020010	Acetoxime
10	65734-38-5	DTXSID6020012	N'-Acetyl-4-(hydroxymethyl) phenylhydrazine

Work in Progress

- Present work in development
  - Real time prediction using OPERA models
  - Structure/substructure/similarity search integration
  - Ongoing expansion of chemicals
  - Release of new ToxCast database (v3_2018)
  - Addition of products data from 10s of thousands of MSDS sheets
  - Analytical Data support
    - Integration of analytical data for ToxCast/Tox21 data
    - Spectral searching against predicted Mass Spectra

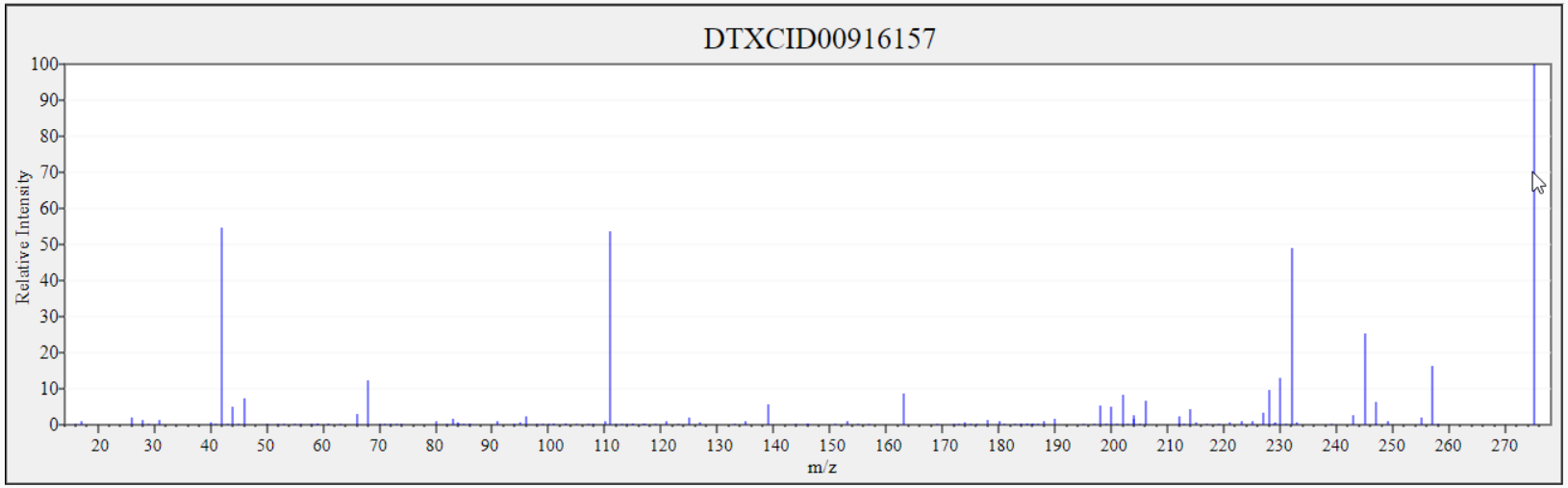
Prototype Development



Predicted Mass Spectra

http://cfmid.wishartlab.com/

- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >700,000 structures, to be accessible via Dashboard



Conclusion

- The EPA CompTox Dashboard provides access to data for ~875,000 chemicals, ToxCast assays and associated product use categories
- High quality data from ongoing curation efforts
- An integration hub for multiple “modules”
  - Experimental and predicted properties
  - Human and Ecological Hazard data
  - Exposure data – products, data in the environment
  - *In vitro* bioassay data – ToxCast/Tox21
  - Literature searching – Google Scholar and PubMed
  - Specialized searches – mass/formula for analytical support
  - Batch searching and Real Time Predictions
- Data and functionality increases with every release

Acknowledgments

- The NCCT CompTox Chemistry Dashboard Development Team
- NERL scientists (Jon Sobus, Elin Ulrich) – Mass Spectrometry
- Kamel Mansouri – OPERA models
- Todd Martin and Valery Tkachenko – TEST predictions
- Nancy Baker – Abstract Sifter

Contact

Antony Williams
US EPA Office of Research and Development
National Center for Computational Toxicology (NCCT)
Williams.Antony@epa.gov
ORCID: https://orcid.org/0000-0002-2668-4821