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

EPA United States Environmental Protection Agency

An Introduction to the EPA CompTox Chemicals Dashboard and GenRA Read-across Module

Antony John Williams and Grace Patlewicz

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

Nordic Training on New Approach Methodologies – November 9th 2021

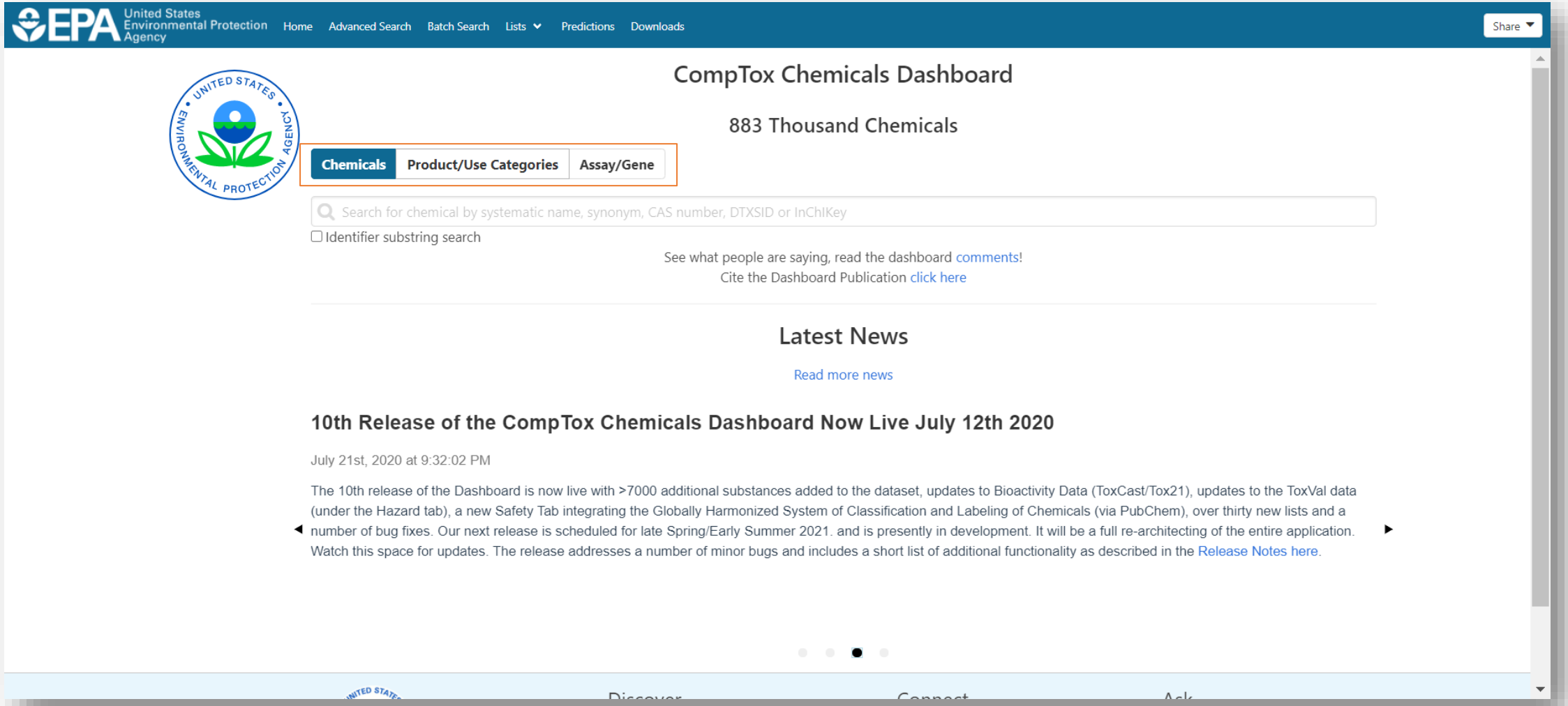
The Charge for the Dashboard

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

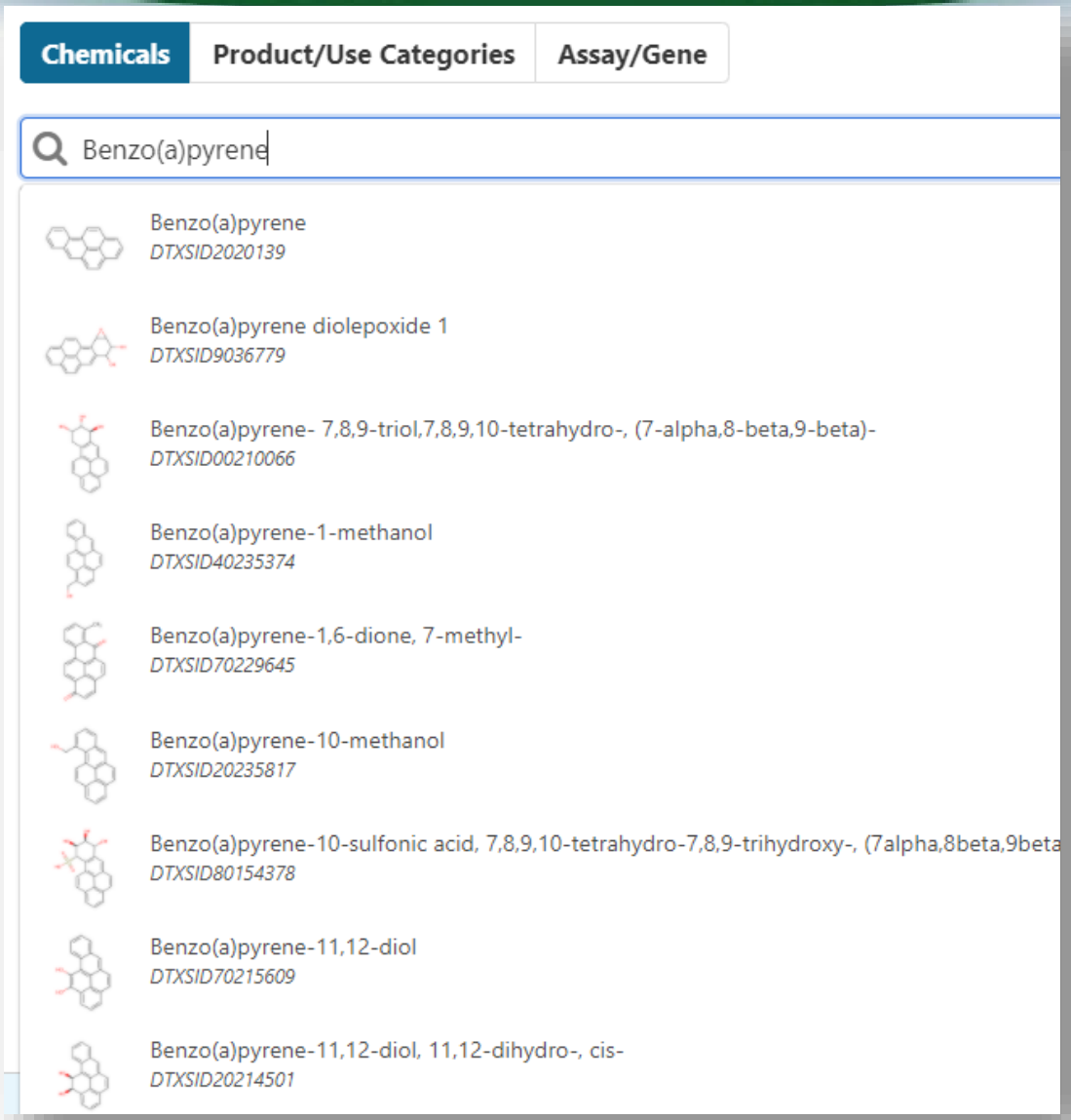
CompTox Chemicals Dashboard

EPA United States Environmental Protection Agency

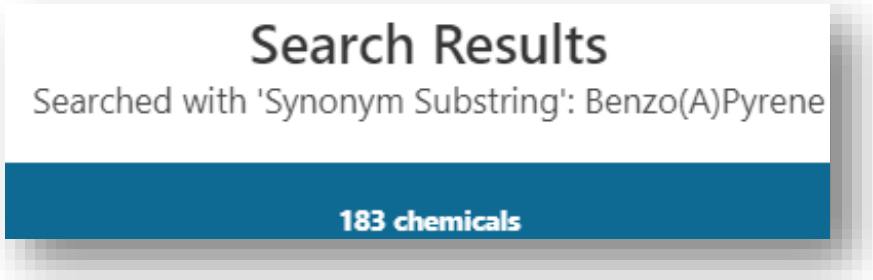
883k Chemical Substances



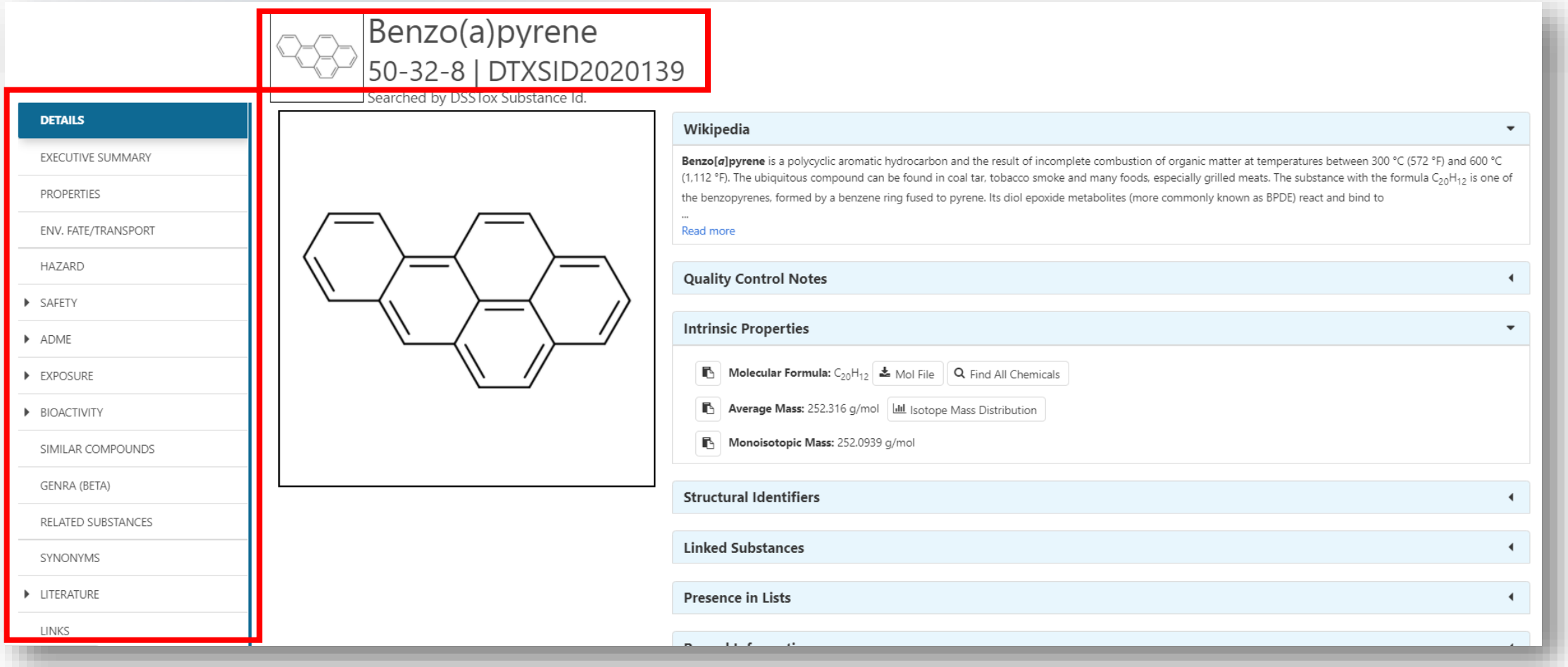
BASIC Search



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

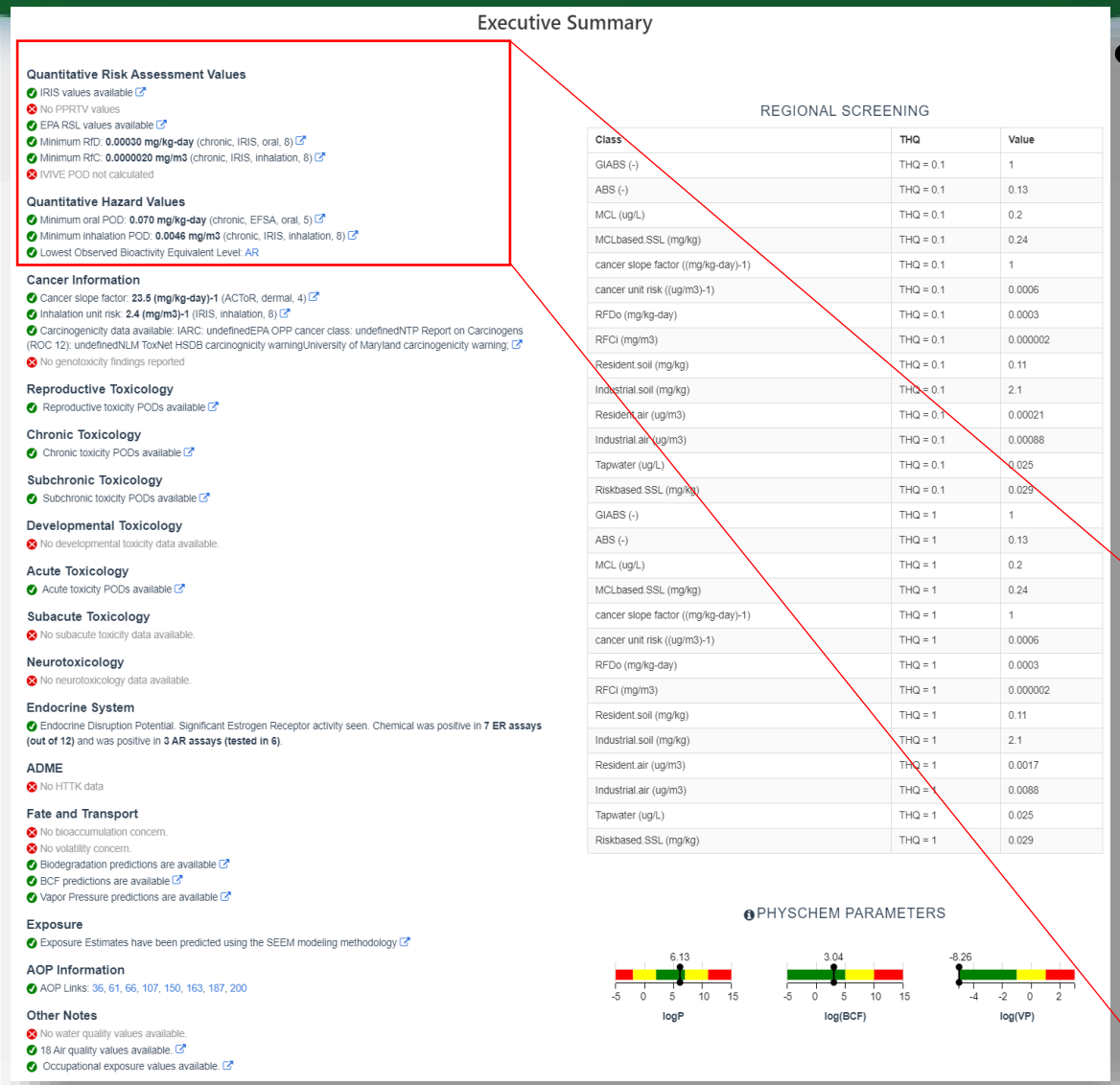


Detailed Chemical Pages
One more identifier – the DTXSID



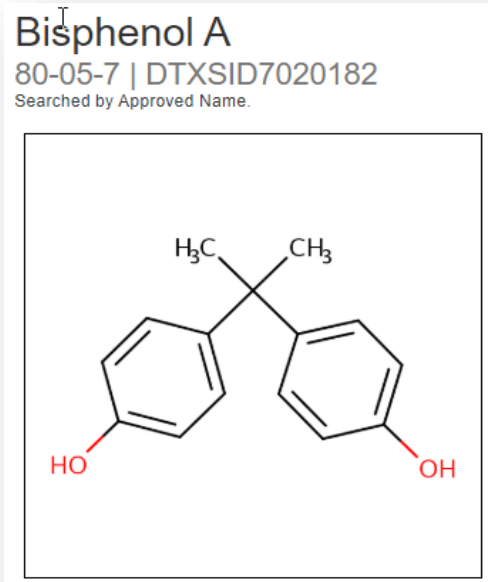
- Chemical page: Wikipedia snippet when available, intrinsic properties, structural identifiers, linked substances

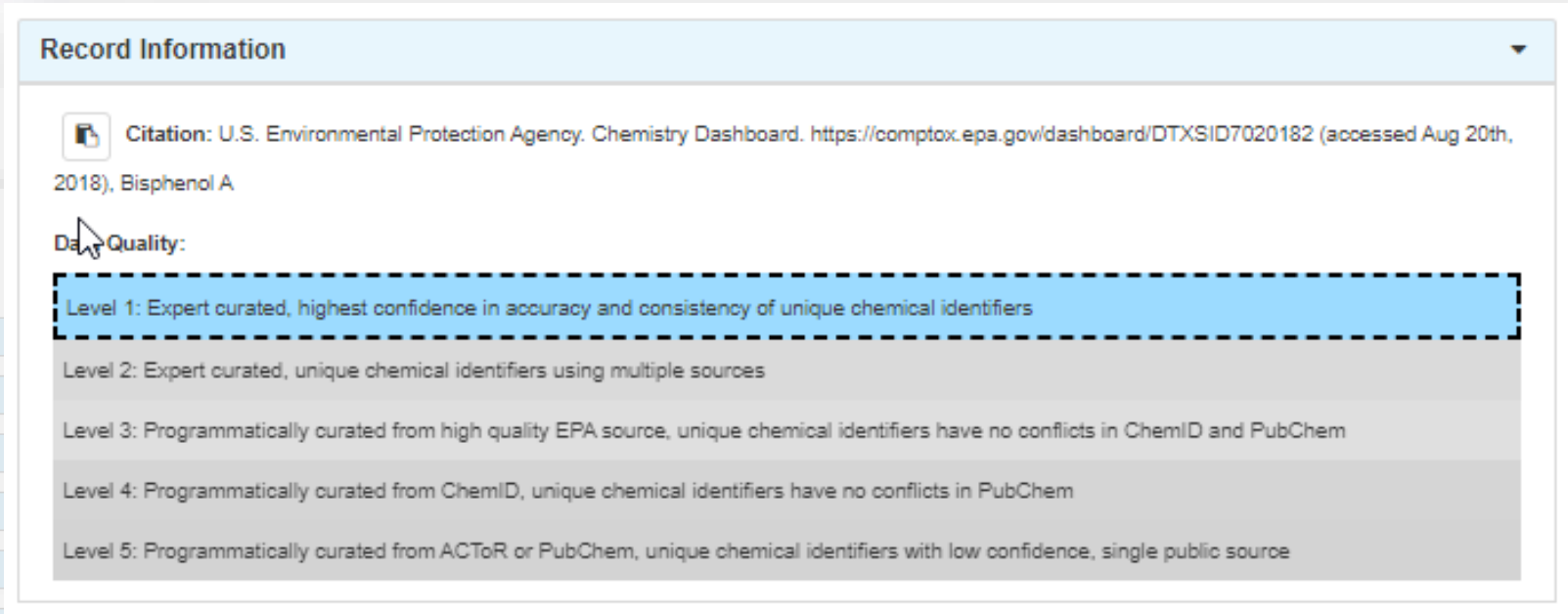
“Executive Summary”



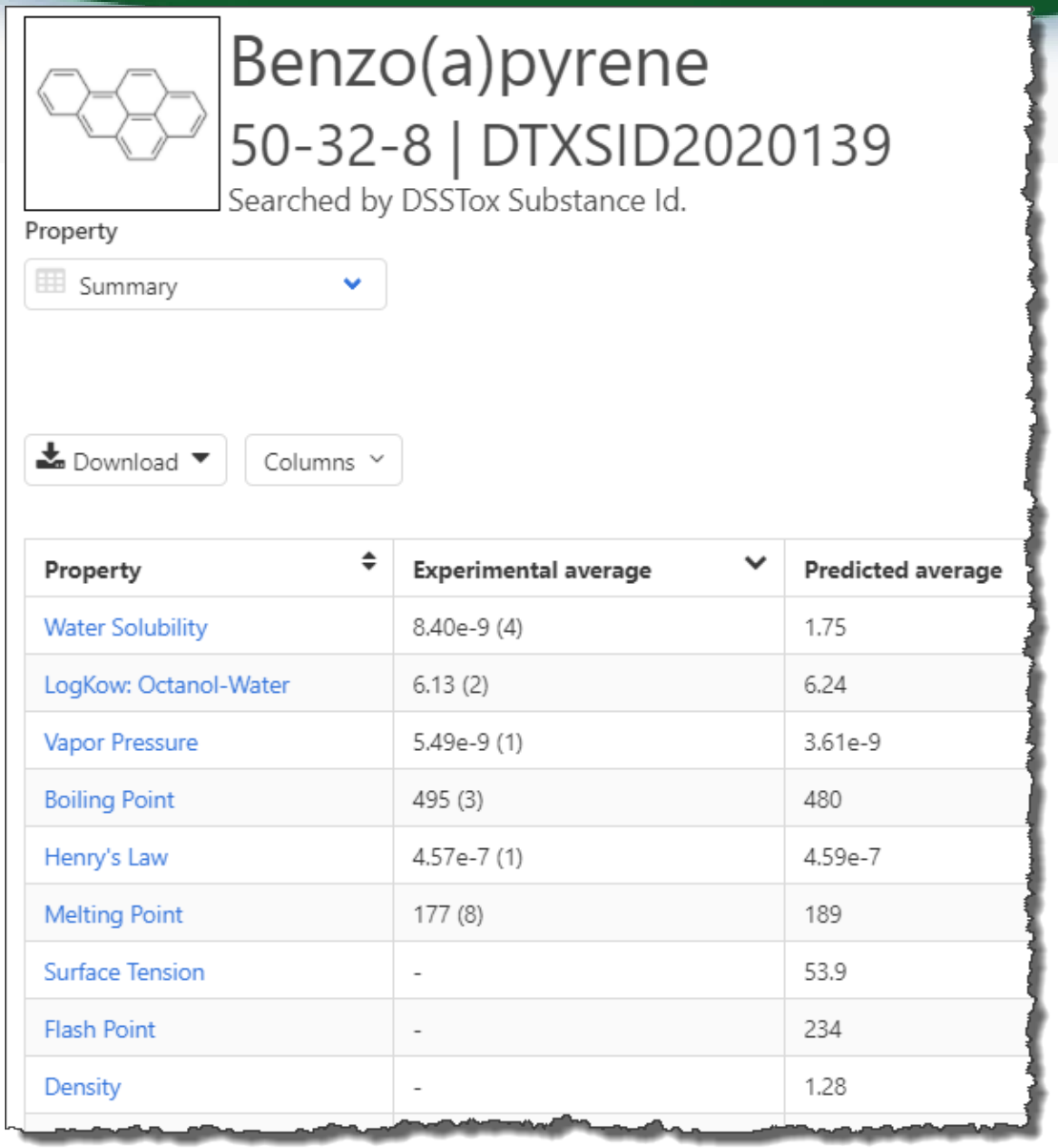
- Overview of toxicity-related info
  - Quantitative values
  - Info re. toxicology subsets
  - Physchem. and Fate & Transport
  - Adverse Outcome Pathway links
  - *In vitro* bioactivity summary plot

Record Information Quality Flags





Experimental and Predicted Data



Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

Property

Summary

Download Columns

Property	Experimental average	Predicted average
Water Solubility	8.40e-9 (4)	1.75
LogKow: Octanol-Water	6.13 (2)	6.24
Vapor Pressure	5.49e-9 (1)	3.61e-9
Boiling Point	495 (3)	480
Henry's Law	4.57e-7 (1)	4.59e-7
Melting Point	177 (8)	189
Surface Tension	-	53.9
Flash Point	-	234
Density	-	1.28

- 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

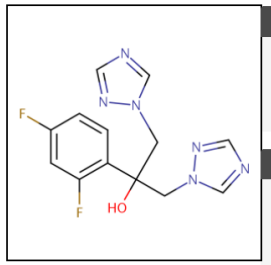
OPERA Reports

OPERA Models: LogKow: Octanol-Water

Fluconazole

86386-73-4 | DTXSID3020627

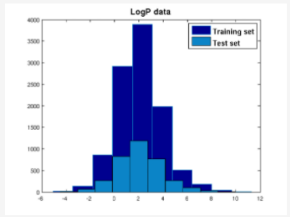
Print PDF

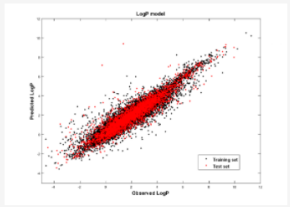


Model Results

Predicted value: 0.501
Global applicability domain: Inside
Local applicability domain index: 0.998
Confidence level: 0.732

Model Performance





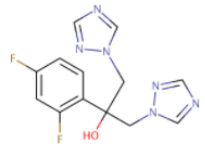
QMRF

Weighted KNN model

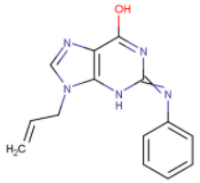
Weighted KNN model

5-fold CV (75%)		Training (75%)		Test (25%)	
Q2	RMSE	R2	RMSE	R2	RMSE
0.850	0.690	0.860	0.670	0.860	0.780

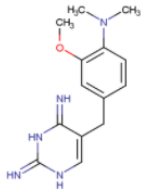
Nearest Neighbors from the Training Set



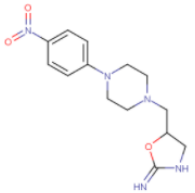
Fluconazole
Measured:0.5
Predicted: 0.50



GUANINEN2PHENYL9ALLYL
Measured:1.75
Predicted: 1.75

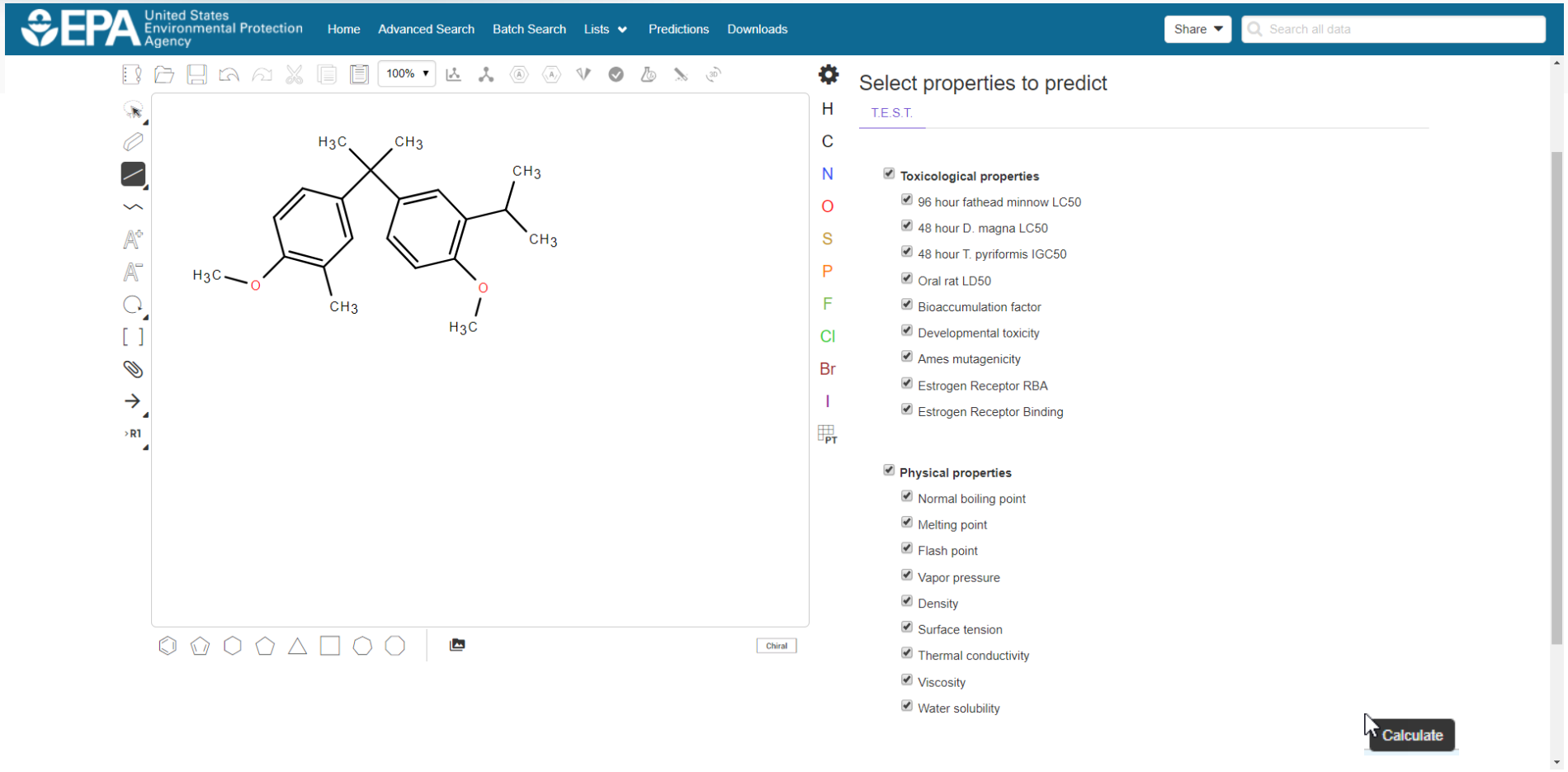


2,4-PYRIMIDINEDIAMINE, 5-[[4-(DIMETHYLAMINO)-3-M
Measured:1.87
Predicted: 1.87



5-(1-P-NITROPHENYL-4-PIPERAZINYL)METHYL-2-AMINO-
Measured:1.23
Predicted: 1.23

Real-Time Predictions

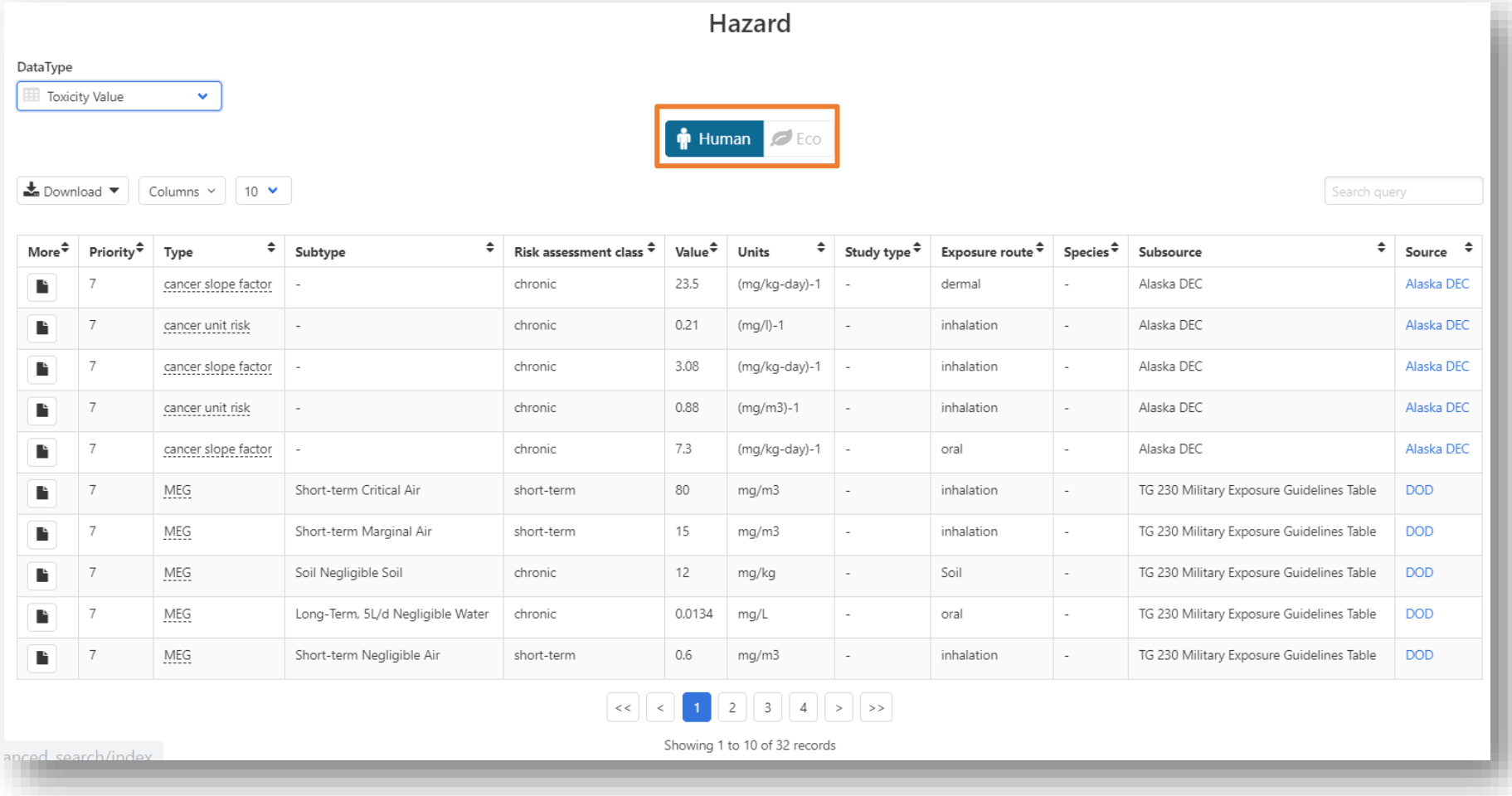


Chemical Hazard Data

EPA United States Environmental Protection Agency

ToxVal Database

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



More	Priority	Type	Subtype	Risk assessment class	Value	Units	Study type	Exposure route	Species	Subsource	Source
	7	cancer slope factor	-	chronic	23.5	(mg/kg-day)-1	-	dermal	-	Alaska DEC	Alaska DEC
	7	cancer unit risk	-	chronic	0.21	(mg/l)-1	-	inhalation	-	Alaska DEC	Alaska DEC
	7	cancer slope factor	-	chronic	3.08	(mg/kg-day)-1	-	inhalation	-	Alaska DEC	Alaska DEC
	7	cancer unit risk	-	chronic	0.88	(mg/m3)-1	-	inhalation	-	Alaska DEC	Alaska DEC
	7	cancer slope factor	-	chronic	7.3	(mg/kg-day)-1	-	oral	-	Alaska DEC	Alaska DEC
	7	MEG	Short-term Critical Air	short-term	80	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Short-term Marginal Air	short-term	15	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Soil Negligible Soil	chronic	12	mg/kg	-	Soil	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Long-Term, 5L/d Negligible Water	chronic	0.0134	mg/L	-	oral	-	TG 230 Military Exposure Guidelines Table	DOD
	7	MEG	Short-term Negligible Air	short-term	0.6	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	DOD

Safety Data

GHS Data

Print Page

PUBCHEM > BENZO[A]PYRENE > LABORATORY CHEMICAL SAFETY SUMMARY (LCSS) > GHS CLASSIFICATION

CID 2336

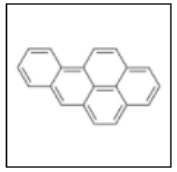
Benzo[a]pyrene

GHS Classification

Showing 6 of 6

Pictogram(s)	 Irritant, Health Hazard, Environmental Hazard
Signal	Danger
GHS Hazard Statements	H317: May cause an allergic skin reaction [Warning Sensitization, Skin] H340: May cause genetic defects [Danger Germ cell mutagenicity] H350: May cause cancer [Danger Carcinogenicity] H360FD: May damage fertility; May damage the unborn child [Danger Reproductive toxicity] H400: Very toxic to aquatic life [Warning Hazardous to the aquatic environment, acute hazard] H410: Very toxic to aquatic life with long lasting effects [Warning Hazardous to the aquatic environment, long-term hazard]
Precautionary Statement Codes	P201, P202, P261, P272, P273, P280, P281, P302+P352, P308+P313, P321, P333+P313, P363, P391, P405, and P501 (The corresponding statement to each P-code can be found at the GHS Classification page.)

Sources of Exposure to Chemicals



Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

Chemical Weight Fractions

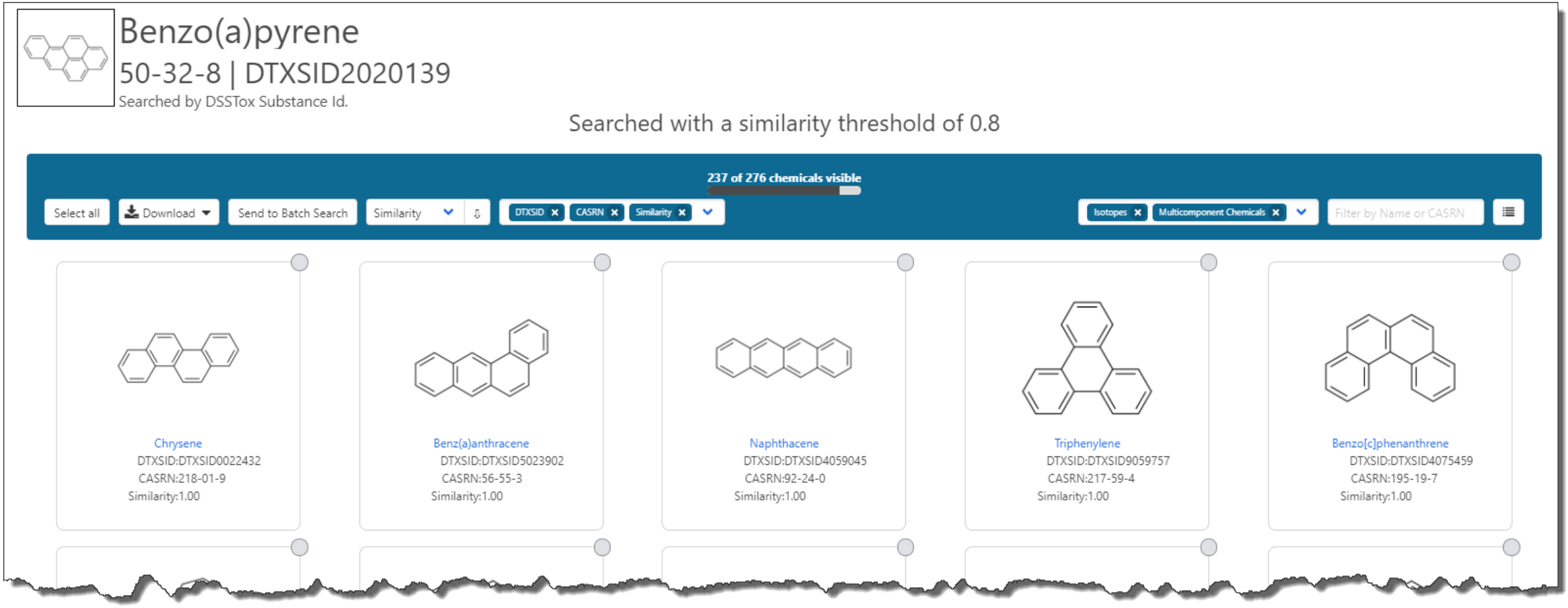
Download | Columns | 10 | Search query

Product Name	Product Use Category	Minimum Weight Fraction	Maximum Weight Fraction	Data Type	Source
m-525-1-5x pah mixtures 0.5 mg/ml for method 525	Not Yet Categorized:			MSDS	SIRI
mm6125 surface conditioner	Not Yet Categorized:			Health Product Declaration	Health Product Declaration Collaborative
monolithic membrane 6125 (mm6125) / monolithic membrane	Not Yet Categorized:			Health Product Declaration	Health Product Declaration Collaborative
organic potablewatr pw 32_component h:reg semi-volatile 690	Not Yet Categorized:	0.00	1.00e-3	MSDS	SIRI
polynuclear aromatic hydrocarbon mixture_ ep84627	Not Yet Categorized:			MSDS	SIRI
prestone(r) power steering fluid	engine maintenance: auto fluids and additives			MSDS	CPCPdb
r-12 shield tite wet surface coating	Not Yet Categorized:	0.00	0.500	MSDS	SIRI
sea tar 1010_ 0028	Not Yet Categorized:			MSDS	SIRI
supelpreme-hc kit pah mix_ 48909	Not Yet Categorized:			MSDS	SIRI
supelpreme-hc pah mix 1ml_ 48905	Not Yet Categorized:			MSDS	SIRI

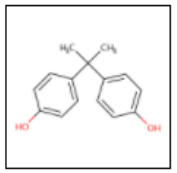
<< < 1 2 3 > >>

Showing 11 to 20 of 21 records

Similarity Searching



Related Substances



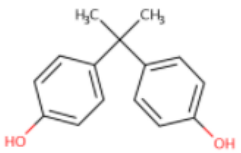
Bisphenol A

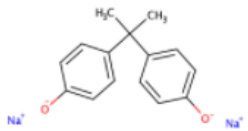
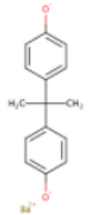
80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

9 chemicals

Select all | Download | Send to Batch Search | Relationship | DTXSID × | CASRN × | TOXCAST × | Hide chemicals that are: | Filter by Name or CASRN

Searched Chemical	Polymer	Polymer	Polymer	Predecessor: Component
	2 related chemical structures with this substance	2 related chemical structures with this substance	3 related chemical structures with this substance	4 related chemical structures with this substance
Bisphenol A DTXSID:DTXSID7020182 CASRN:80-05-7 TOXCAST:217/1152	Formaldehyde, polymer with bisphenol A DTXSID:DTXSID3049627 CASRN:25085-75-0 TOXCAST:-	Bisphenol A/ Epichlorohydrin resin DTXSID:DTXSID0050479 CASRN:25068-38-6 TOXCAST:-	Bisphenol A-epichlorohydrin-polyformal... DTXSID:DTXSID9050480 CASRN:28906-96-9 TOXCAST:-	Fatty acids, C18-unsatd., dimers, polyme... DTXSID:DTXSID40105886 CASRN:106906-26-7 TOXCAST:-

Predecessor: Component	Predecessor: Component	Salt Form	Salt Form
3 related chemical structures with this substance	2 related chemical structures with this substance		
...rohydrin-2,2',6,6'-tet...	Phenoxy Resin	Disodium 4,4'-isopropylidenediphenolate	Barium(2+) 4,4'-isopropylidenebispheno...

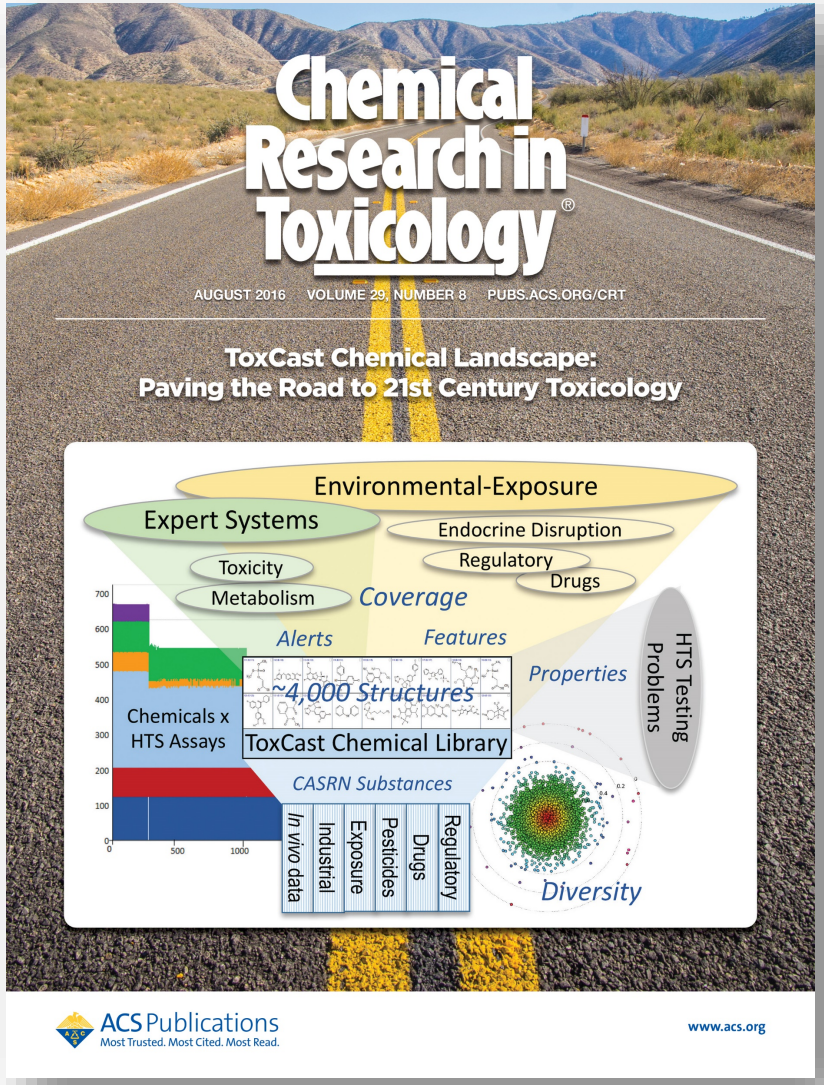
...anced_search/index

EPA
United States
Environmental Protection
Agency

Bioactivity Data

ToxCast

EPA United States Environmental Protection Agency



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

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

View Author Information

Cite this: *Chem. Res. Toxicol.* 2016, 29, 8, 1225–1251
Publication Date: July 1, 2016
https://doi.org/10.1021/acs.chemrestox.6b00135

ToxCast Chemicals and Assays

EPA United States Environmental Protection Agency

Select List

Download | Columns | toxcast_ | Copy Filtered Lists URL

List Acronym	List Name	Last Updated	Number of Chemicals	List Description
TOXCAST_PH1V2	TOXCAST_ph1v2 - EPA ToxCast Screening Library (ph1v2 Subset)	2016-01-25	293	TOXCAST_ph1v2 is the ph1v2 subset of TOXCAST, a reprocured subset of Phase I (ph1v1) chemicals moved into Phase II and later testing phases of the ToxCast program.
TOXCAST_PHASEI	TOXCAST_Phasei - EPA ToxCast Screening Library (Phase I subset)	2016-01-29	310	TOXCAST_PhaseI corresponds to the ph1v1 subset of TOXCAST (mostly pesticides) screened in Phase I of the ToxCast program.
TOXCAST_PH2	TOXCAST_ph2 - EPA ToxCast Screening Library (ph2 Subset)	2016-01-25	768	TOXCAST_ph2 is the ph2 subset of TOXCAST, added in Phase II of the ToxCast program to increase chemical diversity and coverage of chemicals of concern to EPA programs.
TOXCAST_E1K	TOXCAST_e1k - EPA ToxCast Screening Library (e1k Subset)	2016-01-25	799	TOXCAST_e1k is the e1k subset of TOXCAST, selected for screening in endocrine-related assays.
TOXCAST_PHASEII	TOXCASST_PhaseII - EPA ToxCast Screening Library (Phase II Subset)	2016-01-29	1864	TOXCAST_PhaseII is the full set of chemicals screened in Phase II of the ToxCast program, consisting of TOXCAST_ph1v2, ph2 and e1k sublists.
TOXCAST_PH3	TOXCAST_ph3 - EPA ToxCast Screening Library (ph3 subset)	2018-04-11	2678	TOXCAST_ph3 is the ph3 subset of TOXCAST, added to the most recent Phase III of the ToxCast program to further increase chemical diversity and coverage of chemicals of concern to EPA programs.
TOXCAST_PHASEIII	TOXCAST_PhaseIII - EPA ToxCast Screening Library (Phase II Subset)	2017-04-11	4584	TOXCAST_PhaseIII is the full set of chemicals available for screening in Phase III of the ToxCast program, consisting of the majority of chemicals screened in Phase II and newly added ph3

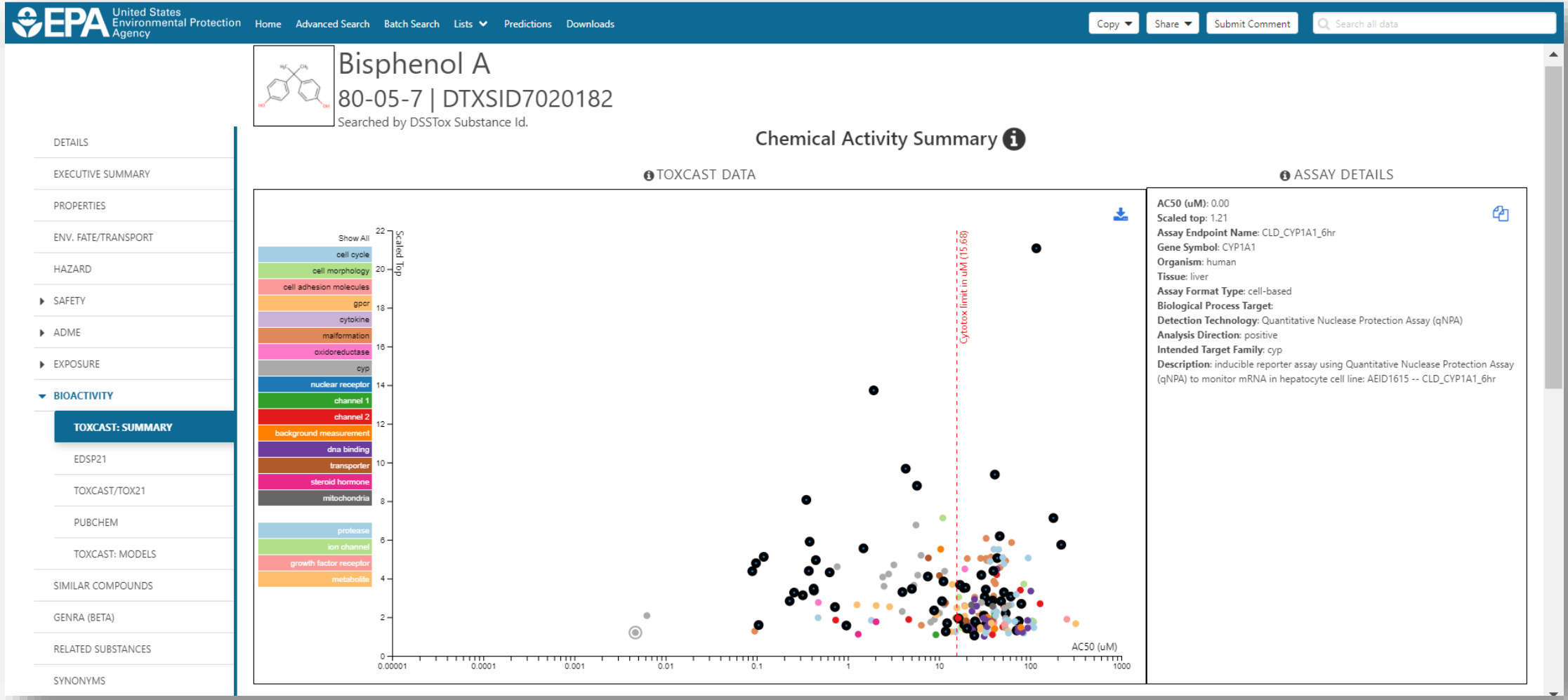
ToxCast covers a lot of biology but not all
ToxCast is growing over time.

EPA United States Environmental Protection Agency

*Invitrodb version 3.3 (released August 2020) contained **17 different assay sources**, covering (at least) **491 unique gene-related** targets with **1600 unique** assay endpoints.*

Assay source	Long name	Truncated assay source description	Some rough notes on the biology covered
ACEA	ACEA Biosciences	real-time, label-free, cell growth assay system based on a microelectronic impedance readout	Endocrine (ER-induced proliferation)
APR	Apredica	CellCiphr High Content Imaging system	Hepatic cells (HepG2)
ATG	Attagene	multiplexed pathway profiling platform	Nuclear receptor and stress response profile
BSK	Bioseek	BioMAP system providing uniquely informative biological activity profiles in complex human primary co-culture systems	Immune/inflammation responses
NVS	Novascreen	large diverse suite of cell-free binding and biochemical assays.	Receptor binding; transporter protein binding; ion channels; enzyme inhibition; many targets
OT	Odyssey Thera	novel protein:protein interaction assays using protein-fragment complementation technology	Endocrine (ER and AR)
TOX21	Tox21/NCGC	Tox21 is an interagency agreement between the NIH, NTP, FDA and EPA. NIH Chemical Genomics Center (NCGC) is the primary screening facility running ultra high-throughput screening assays across a large interagency-developed chemical library	Many – with many nuclear receptors
CEETOX	Ceetox/OpAns	HT-H295R assay	Endocrine (steroidogenesis)
CLD	CellzDirect	Formerly CellzDirect, this Contract Research Organization (CRO) is now part of the Invitrogen brand of Thermo Fisher providing cell-based in vitro assay screening services using primary hepatocytes.	Liver (Phase I/Phase II/ Phase III expression)
NHEERL_PADILLA	NHEERL Padilla Lab	The Padilla laboratory at the EPA National Health and Environmental Effects Research Laboratory focuses on the development and screening of zebrafish assays.	Zebrafish terata
NCCT	NCCT Simmons Lab	The Simmons Lab at the EPA National Center for Computational Toxicology focuses on developing and implementing in vitro methods to identify potential environmental toxicants.	Endocrine (thyroid - thyroperoxidase inhibition)
TANGUAY	Tanguay Lab	The Tanguay Lab, based at the Oregon State University Sinnhuber Aquatic Research Laboratory, uses zebrafish as a systems toxicology model.	Zebrafish terata/phenotypes
NHEERL_NIS	NHEERL Stoker & Laws	The Stoker and Laws laboratories at the EPA National Health and Environmental Effects Research Laboratory work on the development and implementation of high-throughput assays, particularly related to the sodium-iodide cotransporter (NIS).	Endocrine (thyroid - NIS inhibition)
UPITT	University of Pittsburgh	The Johnston Lab at the University of Pittsburgh ran androgen receptor nuclear translocation assays under a Material Transfer Agreement (MTA) for the ToxCast Phase 1, Phase 2, and E1K chemicals.	Endocrine (AR related)

ToxCast/Tox21 Data



Rich data tables – full transparency

EPA United States Environmental Protection Agency

217 active of 1152 assays

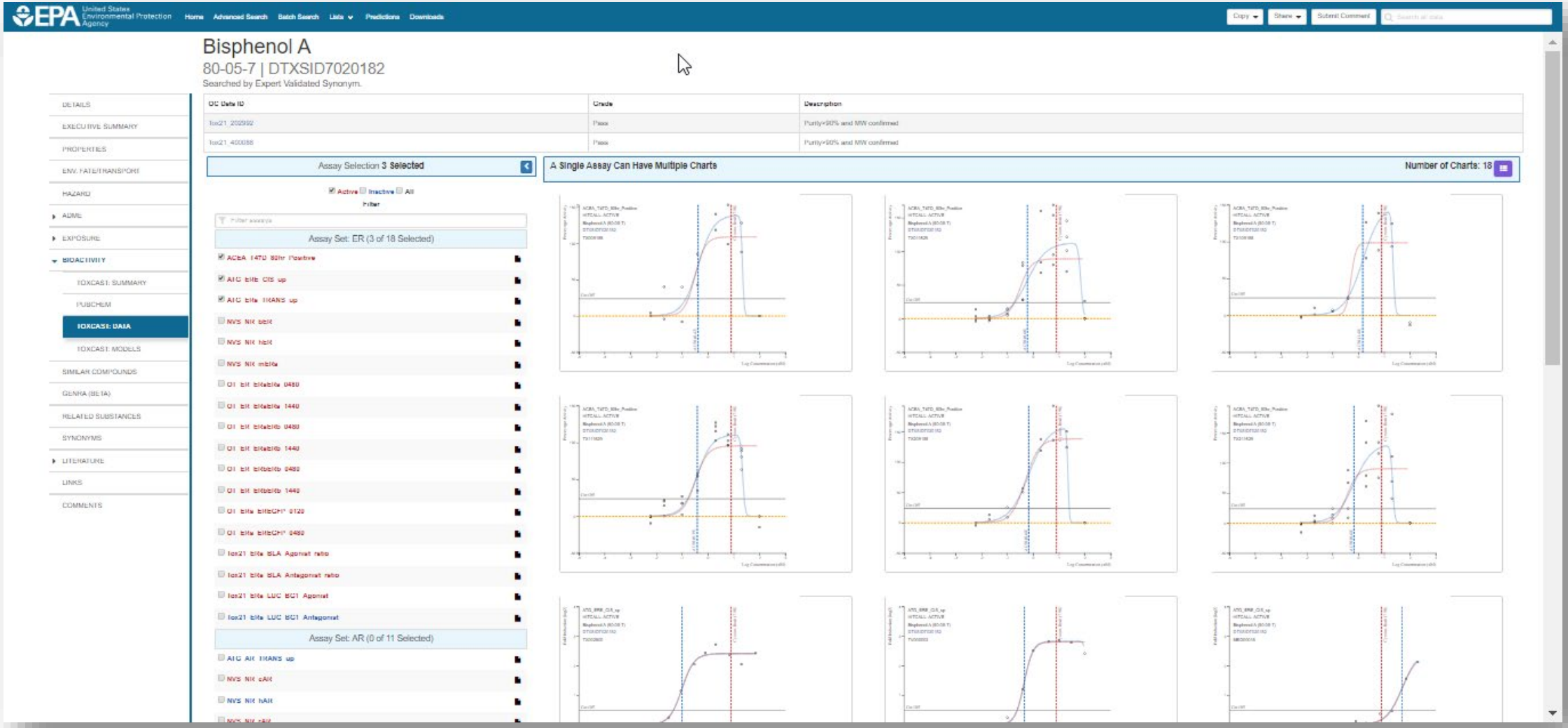
Download | Columns | 10 | Search query | Show Inactive | Show Background

Name	Modal	SeqAPASS	Gene Symbol	Gene Name	AOP	Event	Hit Call	Top	Scaled Top	AC50	logAC50	Bmad	Cutoff	Intended Target Family
ATG_Xbp1_CIS_up		NP_005071.2	XBP1	X-box binding protein 1	-	-	ACTIVE	1.08	1.83	93.8	1.97	0.118	0.590	dna binding
ATG_VDRE_CIS_up		NP_000367.1	VDR	vitamin D (1,25- dihydroxyvitamin D3) receptor	-	-	ACTIVE	1.71	2.35	8.84	0.946	0.146	0.728	nuclear receptor
BSK_hDFCGF_VCAM1_down		NP_001069.1	VCAM1	vascular cell adhesion molecule 1	-	-	ACTIVE	0.248	1.72	40.0	1.60	4.81e-2	0.144	cell adhesion molecules
BSK_LPS_VCAM1_down		NP_001069.1	VCAM1	vascular cell adhesion molecule 1	-	-	ACTIVE	0.107	1.35	40.0	1.60	2.31e-2	7.92e-2	cell adhesion molecules
TOX21_p53_BLA_p2_ratio		NP_000537.3	TP53	tumor protein p53	-	-	ACTIVE	35.5	1.78	73.7	1.87	0.959	20.0	dna binding
TOX21_p53_BLA_p4_ratio		NP_000537.3	TP53	tumor protein p53	-	-	ACTIVE	28.6	1.43	73.8	1.87	1.02	20.0	dna binding
ATG_p53_CIS_dn		NP_000537.3	TP53	tumor protein p53	-	-	ACTIVE	0.789	1.23	82.4	1.92	0.129	0.643	dna binding
NVS_MP_rPBR		NP_036647.1	Tspo	translocator protein	-	-	ACTIVE	61.5	1.96	15.3	1.19	5.24	31.4	transporter
ATG_AP_2_CIS_dn		NP_003211.1	TFAP2A	transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)	-	-	ACTIVE	0.458	1.44	93.8	1.97	6.35e-2	0.317	dna binding
ATG_TCF_b_cat_CIS_dn		NP_003193.2	TCF7	transcription factor 7 (T-cell specific, HMG-box)	-	-	ACTIVE	1.34	1.96	30.7	1.49	0.137	0.687	dna binding

First << < 1 2 3 4 5 6 7 8 9 10 > >> Last

Bioactivity Data (ToxCast/Tox21)

Data below for Bisphenol A



Use Models Derived from the Data

EPA United States Environmental Protection Agency

Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model

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

View Author Information

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

Vol. 124, No. 7 | Research

CERAPP: Collaborative Estrogen Receptor Activity Prediction Project

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

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

Development and Validation of a Computational Model for Androgen Receptor Activity

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

View Author Information

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

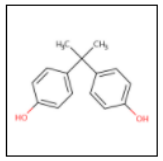
Vol. 128, No. 2 | Research

CoMPARA: Collaborative Modeling Project for Androgen Receptor Activity

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

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

For Endocrine (AR and ER) better to use summary models



Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

Positive ToxCast ER pathway agonist and ToxCast AR antagonist scores.

- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- SAFETY
- ADME
- EXPOSURE
- BIOACTIVITY
  - TOXCAST: SUMMARY
  - EDSP21
  - TOXCAST/TOX21
  - PUBCHEM
  - TOXCAST: MODELS

ToxCast: Models

ToxCast Model Predictions

Download ToxCast Model Predictions

Model	Receptor	Agonist	Antagonist	Binding
ToxCast Pathway Model (AUC)	Androgen	0.00	0.345	-
ToxCast Pathway Model (AUC)	Estrogen	0.450	0.00	-
COMPARA (Consensus)	Androgen	Inactive	Active	Active
CERAPP Potency Level (From Literature)	Estrogen	Active (Weak)	-	Active (Weak)
CERAPP Potency Level (Consensus)	Estrogen	Active (Weak)	Active (Strong)	Active (Weak)

CERAPP = consensus ER QSAR (from 17 groups)

COMPARA = consensus AR QSAR

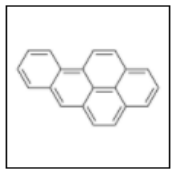
ToxCast Pathway Model AUC ER = full ER model (18 assays)

ToxCast Pathway Model AUC AR = full AR model (11 assays)

EPA
United States
Environmental Protection
Agency

Searching Literature and the Internet

Identifiers Support Searches in other systems



Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

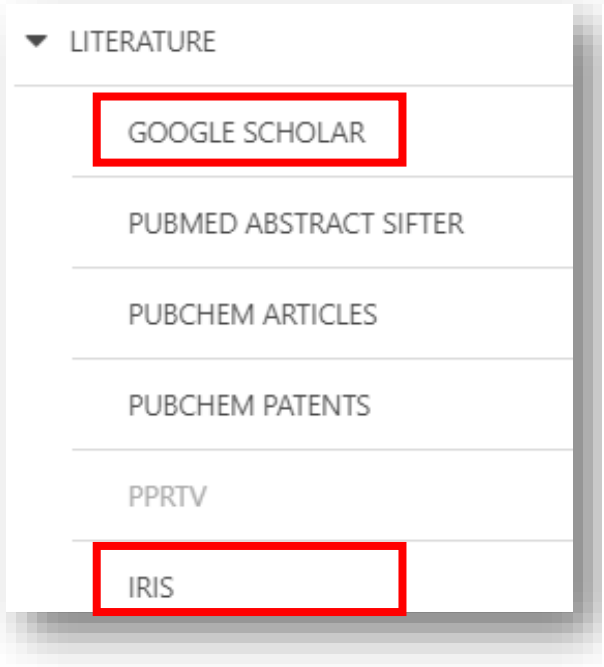
Synonyms

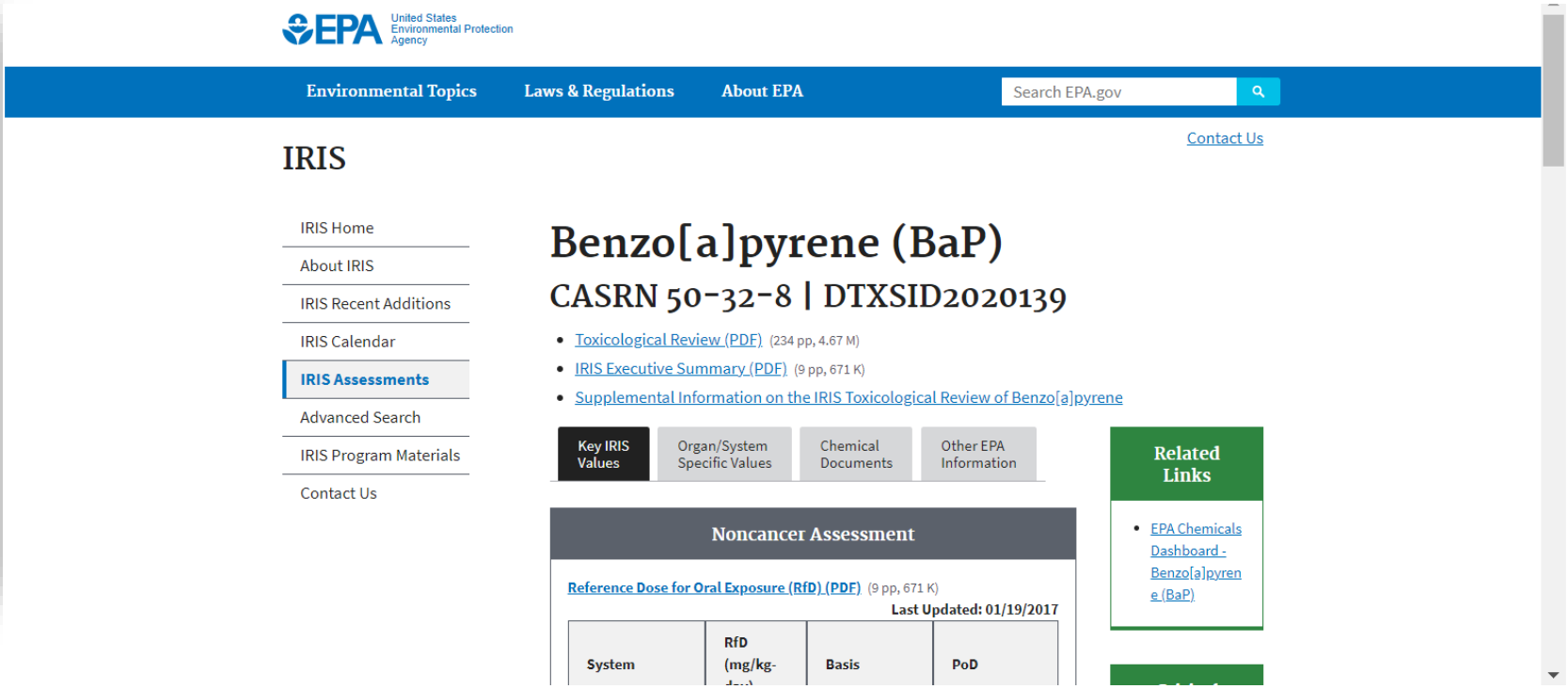
Download 25 Search query

Synonym	Quality
Benzo(a)pyrene	**Valid**
Benzo[pqr]tetraphene	**Valid**
Benzo[a]pyrene	**Valid**
50-32-8 Active CAS-RN	**Valid**
BaP	**Valid**
Benzo[a]pyrene	*Good*
3,4-Benz[a]pyrene	*Good*
3,4-Benzopyrene	*Good*
3,4-Benzpyrene	*Good*
6,7-Benzopyrene	*Good*
BENZ(A)PYREN	*Good*
Benz(a)pyrene	*Good*
Benz[a]pyrene	*Good*

Identifiers are used in the app

- Identifiers are used to feed and link into “Literature”





Literature Searching

Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

Abstract Sifter

1) Select PubMed starting point query then 2) click on Retrieve.

Hazard

Retrieve Articles

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

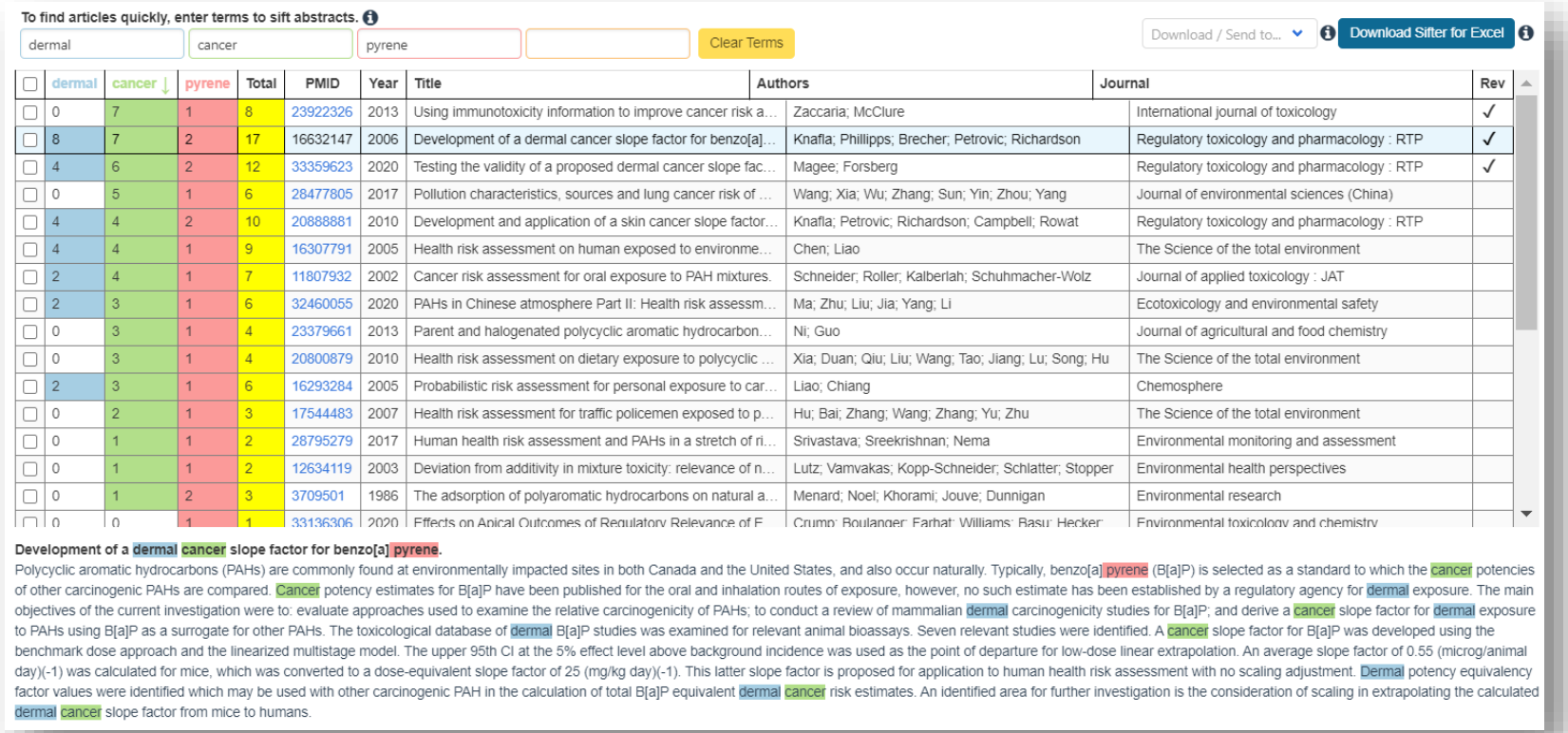
Optionally, edit the query before retrieving.

("50-32-8" OR "Benzo(a)pyrene") AND (NOAEL OR NOEL OR LOEL OR Rfd OR "reference dose" OR "reference concentration" OR "adverse effect level"[tiab] OR "cancer slope factor"[tiab])

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

Literature Searching

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



To find articles quickly, enter terms to sift abstracts.

dermal | cancer | pyrene | Clear Terms | Download / Send to... | Download Sifter for Excel

	dermal	cancer ↓	pyrene	Total	PMID	Year	Title	Authors	Journal	Rev
	0	7	1	8	23922326	2013	Using immunotoxicity information to improve cancer risk a…	Zaccaria; McClure	International journal of toxicology	✓
	8	7	2	17	16632147	2006	Development of a dermal cancer slope factor for benzo[a]…	Knafla; Phillipps; Brecher; Petrovic; Richardson	Regulatory toxicology and pharmacology : RTP	✓
	4	6	2	12	33359623	2020	Testing the validity of a proposed dermal cancer slope fac…	Magee; Forsberg	Regulatory toxicology and pharmacology : RTP	✓
	0	5	1	6	28477805	2017	Pollution characteristics, sources and lung cancer risk of …	Wang; Xia; Wu; Zhang; Sun; Yin; Zhou; Yang	Journal of environmental sciences (China)	
	4	4	2	10	20888881	2010	Development and application of a skin cancer slope factor…	Knafla; Petrovic; Richardson; Campbell; Rowat	Regulatory toxicology and pharmacology : RTP	
	4	4	1	9	16307791	2005	Health risk assessment on human exposed to environme…	Chen; Liao	The Science of the total environment	
	2	4	1	7	11807932	2002	Cancer risk assessment for oral exposure to PAH mixtures.	Schneider; Roller; Kalberlah; Schuhmacher-Wolz	Journal of applied toxicology : JAT	
	2	3	1	6	32460055	2020	PAHs in Chinese atmosphere Part II: Health risk assessm…	Ma; Zhu; Liu; Jia; Yang; Li	Ecotoxicology and environmental safety	
	0	3	1	4	23379661	2013	Parent and halogenated polycyclic aromatic hydrocarbon…	Ni; Guo	Journal of agricultural and food chemistry	
	0	3	1	4	20800879	2010	Health risk assessment on dietary exposure to polycyclic …	Xia; Duan; Qiu; Liu; Wang; Tao; Jiang; Lu; Song; Hu	The Science of the total environment	
	2	3	1	6	16293284	2005	Probabilistic risk assessment for personal exposure to car…	Liao; Chiang	Chemosphere	
	0	2	1	3	17544483	2007	Health risk assessment for traffic policemen exposed to p…	Hu; Bai; Zhang; Wang; Zhang; Yu; Zhu	The Science of the total environment	
	0	1	1	2	28795279	2017	Human health risk assessment and PAHs in a stretch of ri…	Srivastava; Sreekrishnan; Nema	Environmental monitoring and assessment	
	0	1	1	2	12634119	2003	Deviation from additivity in mixture toxicity: relevance of n…	Lutz; Vamvakas; Kopp-Schneider; Schlatter; Stopper	Environmental health perspectives	
	0	1	2	3	3709501	1986	The adsorption of polyaromatic hydrocarbons on natural a…	Menard; Noel; Khorami; Jouve; Dunnigan	Environmental research	
	0	0	1	1	33136306	2020	Effects on Apical Outcomes of Regulatory Relevance of E…	Crump; Boulanger; Farhat; Williams; Basu; Hecker	Environmental toxicology and chemistry	

Development of a dermal cancer slope factor for benzo[a] pyrene.

Polycyclic aromatic hydrocarbons (PAHs) are commonly found at environmentally impacted sites in both Canada and the United States, and also occur naturally. Typically, benzo[a] pyrene (B[a]P) is selected as a standard to which the cancer potencies of other carcinogenic PAHs are compared. Cancer potency estimates for B[a]P have been published for the oral and inhalation routes of exposure, however, no such estimate has been established by a regulatory agency for dermal exposure. The main objectives of the current investigation were to: evaluate approaches used to examine the relative carcinogenicity of PAHs; to conduct a review of mammalian dermal carcinogenicity studies for B[a]P; and derive a cancer slope factor for dermal exposure to PAHs using B[a]P as a surrogate for other PAHs. The toxicological database of dermal B[a]P studies was examined for relevant animal bioassays. Seven relevant studies were identified. A cancer slope factor for B[a]P was developed using the benchmark dose approach and the linearized multistage model. The upper 95th CI at the 5% effect level above background incidence was used as the point of departure for low-dose linear extrapolation. An average slope factor of 0.55 (microg/animal day)(-1) was calculated for mice, which was converted to a dose-equivalent slope factor of 25 (mg/kg day)(-1). This latter slope factor is proposed for application to human health risk assessment with no scaling adjustment. Dermal potency equivalency factor values were identified which may be used with other carcinogenic PAH in the calculation of total B[a]P equivalent dermal cancer risk estimates. An identified area for further investigation is the consideration of scaling in extrapolating the calculated dermal cancer slope factor from mice to humans.

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

- We know how complex chemicals identifiers are…
  - CASRN(s)
  - Hundreds of names (maybe)
  - SMILES
  - InChIs
  - EINECS, EC numbers

- What can WE do to help you navigate the internet?

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

Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

General

- EPA Substance Registry Service
- PubChem
- Chemspider
- CPCat
- DrugBank
- Wikipedia
- MSDS Lookup
- ChEMBL
- ToxPlanet
- ACS Reagent Chemicals
- Wolfram Alpha
- ECHA Infocard
- ChemAgora
- Consumer Product Information Database
- ChEBI
- NIST Chemistry Webbook
- WEBWISER
- PubChem Safety Sheet
- Consumer Product Information Database
- PubChem: Chemical Vendors

Toxicology

- ACToR
- DrugPortal
- CCRIS
- ChemView
- CTD
- eChemPortal
- Gene-Tox
- HSDB
- ACToR PDF Report
- CREST
- National Air Toxics Assessment
- ECOTOX
- ChemView
- Chemical Checker
- BindingDB
- CalEPA OEHHA
- NIOSH IDLH Values
- LactMed
- ECOTOX

Publications

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

Analytical

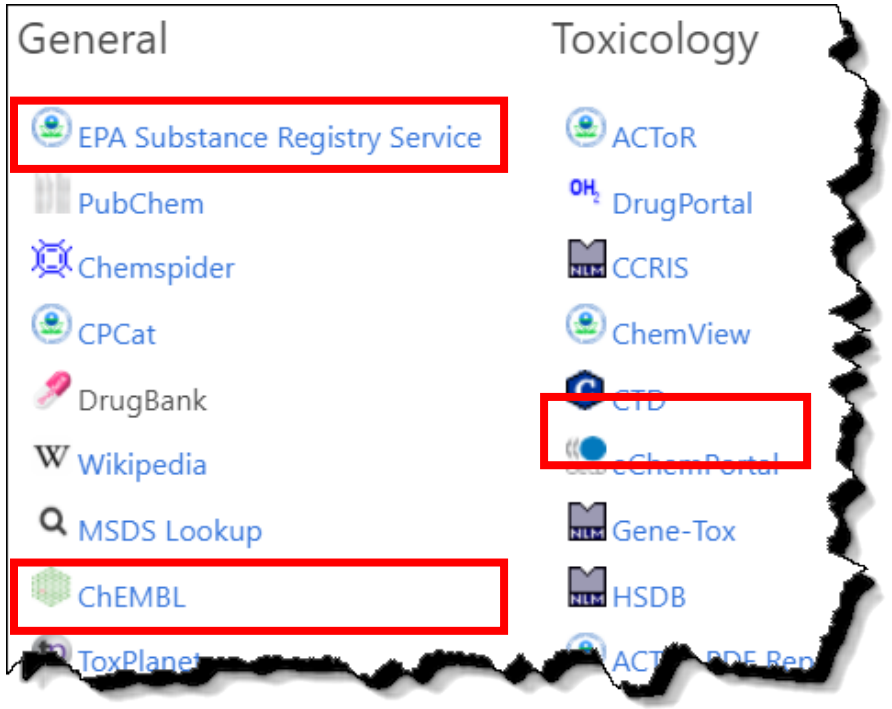
- RSC Analytical Abstracts
- Tox21 Analytical Data
- MONA: MassBank North America
- mzCloud
- NIST IR Spectrum
- NIST MS Spectrum
- MassBank
- NIST Antoine Constants
- IR Spectra on PubChem
- NIST Kovats Index values
- Protein DataBank
- National Environmental Methods Index

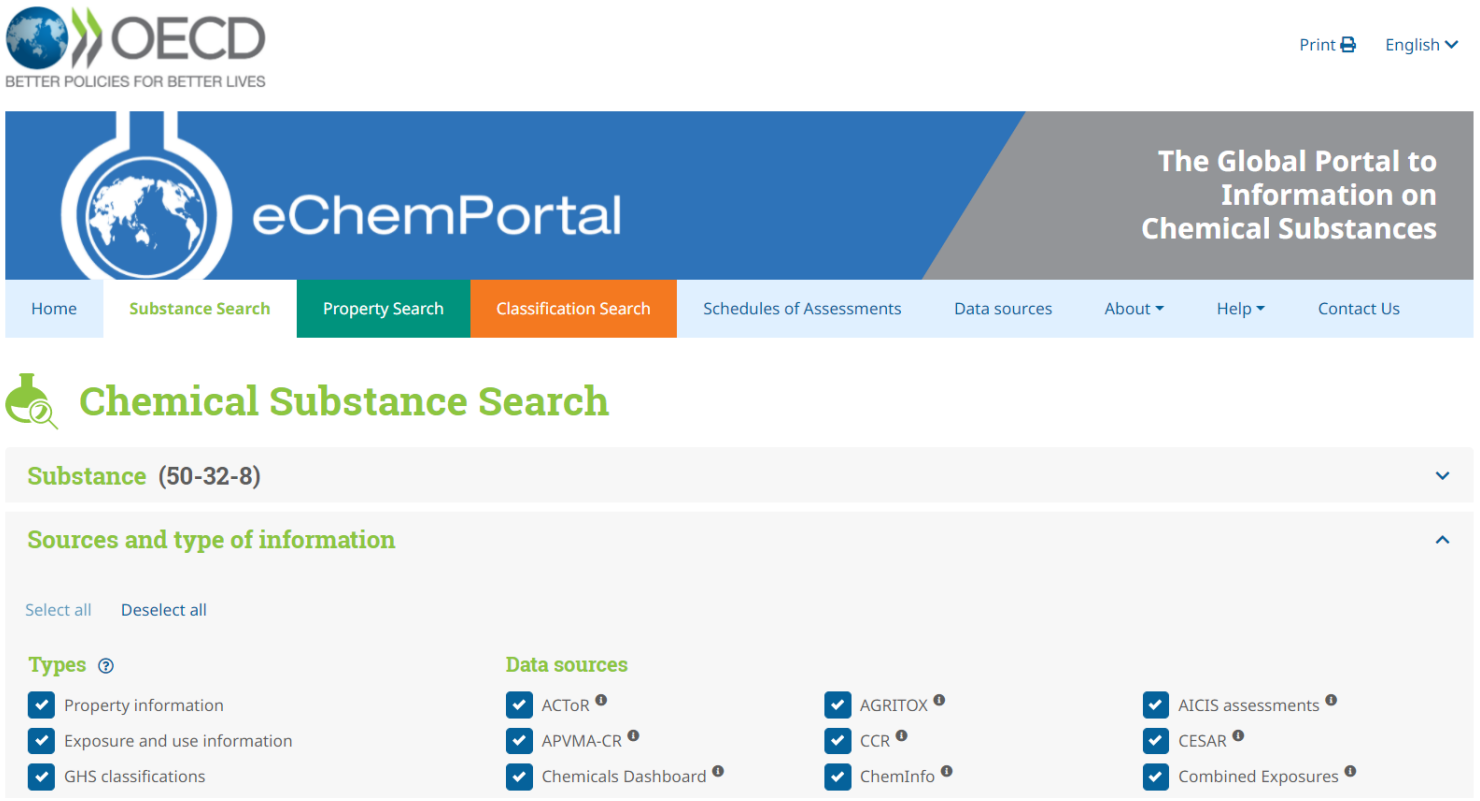
Prediction

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

External Links

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

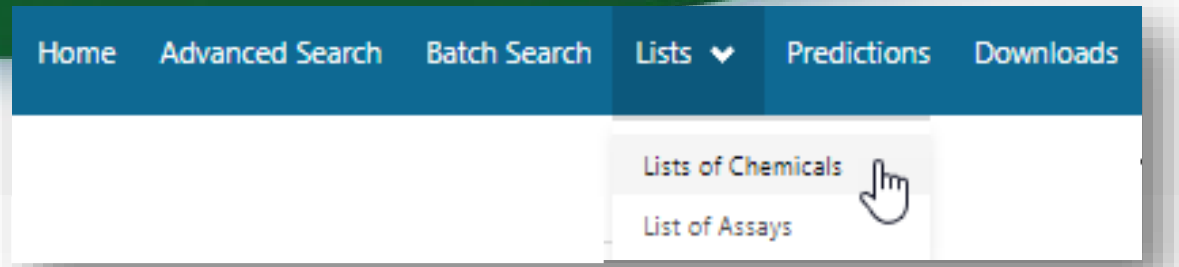




EPA
United States
Environmental Protection
Agency

Chemical Lists and Categories

Example: AEGLs list



AEGLS: Acute Exposure Guideline Levels

Search AEGLVALUES Chemicals

Identifier substring search

List Details

Description: Acute Exposure Guideline Level (AEGLs) values are intended to protect most individuals in the general population, including those that might be particularly susceptible to the harmful effects of the chemicals. Acute exposure guideline levels (AEGLs) describe the human health effects from once-in-a-lifetime, or rare, exposure to airborne chemicals. Used by emergency responders when dealing with chemical spills or other catastrophic exposures, AEGLs are set through a collaborative effort of the public and private sectors worldwide.
Number of Chemicals: 174

Select all | Download | Send to Batch Search | Name | CASRN | DTXSID | 174 chemicals | Hide chemicals that are: | Filter by Name or CASRN

(2-Chlorobenzylidene)propanedinitrile
CASRN:2698-41-1
DTXSID:DTXSID9020297

(E)-1,2-Dichloroethylene
CASRN:156-60-5
DTXSID:DTXSID7024031

(E)-Crotonaldehyde
CASRN:123-73-9
DTXSID:DTXSID6020351

(Z)-1,2-Dichloroethylene
CASRN:156-59-2
DTXSID:DTXSID2024030

PFAS lists of Chemicals

EPA United States Environmental Protection Agency

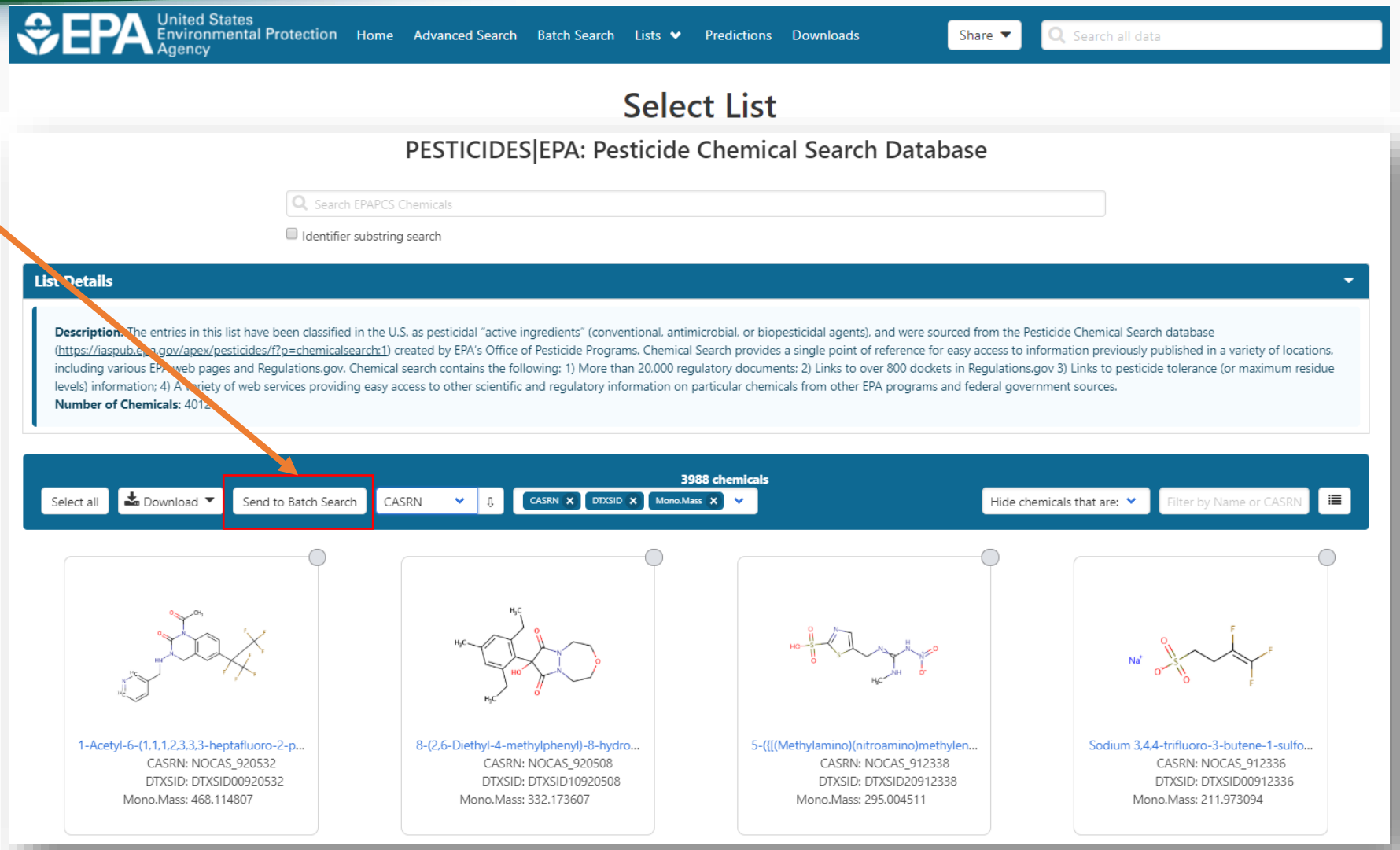
Select List

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)

Curated List of Pesticides

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



Batch Searching

Batch Searching

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

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

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

EPA United States Environmental Protection Agency

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

Batch Search

Step 1 — Step 2 — Step 3 — Step 4 — Step 5

Step Three: Select Download Data or Display Chemicals

Please enter one identifier per line

Select Input Type(s)

- [x] Identifiers
  - [x] Chemical Name
  - [x] CASRN
  - [ ] InChIKey
  - [ ] DSSTox Substance ID
- [ ] DSSTox Compound ID
- [ ] InChIKey Skeleton
- [ ] MS-Ready Formula(e)
- [ ] Exact Formula(e)
- [ ] Monoisotopic Mass

Enter Identifiers to Search (One per line. Searches should be limited to <5000 identifiers.)

```
Sodium azide
O,O-Dimethyl S-[(4-oxo-1,2,3-benzotriazin-3(4H)-yl)methyl] phosphorodithioate
(Trichloromethyl)benzene
100-44-7
(Chloromethyl)benzene
98-07-7
111-44-4
1-Chloro-2-(2-chloroethoxy)ethane
Chloro(chloromethoxy)methane
2-(Butan-2-yl)-4,6-dinitrophenol
```

Display All Chemicals | Download Chemical Data

Select Output Format and Content

United States Environmental Protection Agency

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

Select Output Format:

Excel

Download

Customize Results

- [ ] Select All
- [ ] Select All in Lists

Chemical Identifiers

- [x] DTXSID
- [x] Chemical Name
- [ ] DTXCID
- [ ] 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
- [ ] OPERA Model Predictions

Metadata

- [ ] Curation Level Details
- [ ] NHANES/Predicted Exposure
- [ ] Data Sources
- [ ] Include ToxVal Data Availability
- [ ] Assay Hit Count
- [ ] Number of PubMed Articles
- [ ] PubChem Data Sources
- [ ] CPDat Product Occurrence Count
- [ ] IRIS
- [ ] PPRTV
- [ ] Wikipedia Article
- [ ] QC Notes
- [ ] Include links to ACToR reports - SLOW! (BETA)

Presence in Lists:

- [ ] 40CFR116.4 Designation of Hazardous Substances (Above Ground Storage Tanks)
- [ ] 40CFR355 Extremely Hazardous Substance List and Threshold Planning Quantities
- [ ] AEGLS: Acute Exposure Guideline Levels
- [ ] ANDROGEN: Androgen Receptor Chemicals
- [ ] ARTICLE; Bench-Mark Dose Human Health Assessment List (Wignall et al., 2014)
- [ ] ARTICLE: Collaborative Estrogen Receptor Activity Prediction Project (CERAPP)
- [ ] ARTICLE: Collaborative Estrogen Receptor Activity Prediction Project (COMPARA)
- [ ] ATSDR Toxicological Profiles
- [ ] ATSDR: Minimal Risk Levels (MRLs) for Hazardous Substances
- [ ] ATSDR: Toxic Substances Portal Chemical List
- [ ] California Office of Environmental Health Hazard Assessment
- [ ] Canadian Domestic Substances List 2019
- [ ] CATEGORY: Amino acids
- [ ] CATEGORY: Color Index dyes
- [ ] CATEGORY: Flame Retardants

Batch Search CASRNs

United States Environmental Protection Agency

Batch Search

Step 1 — Step 2 — Step 3 — Step 4 — Step 5

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

Please enter one identifier per line

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

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

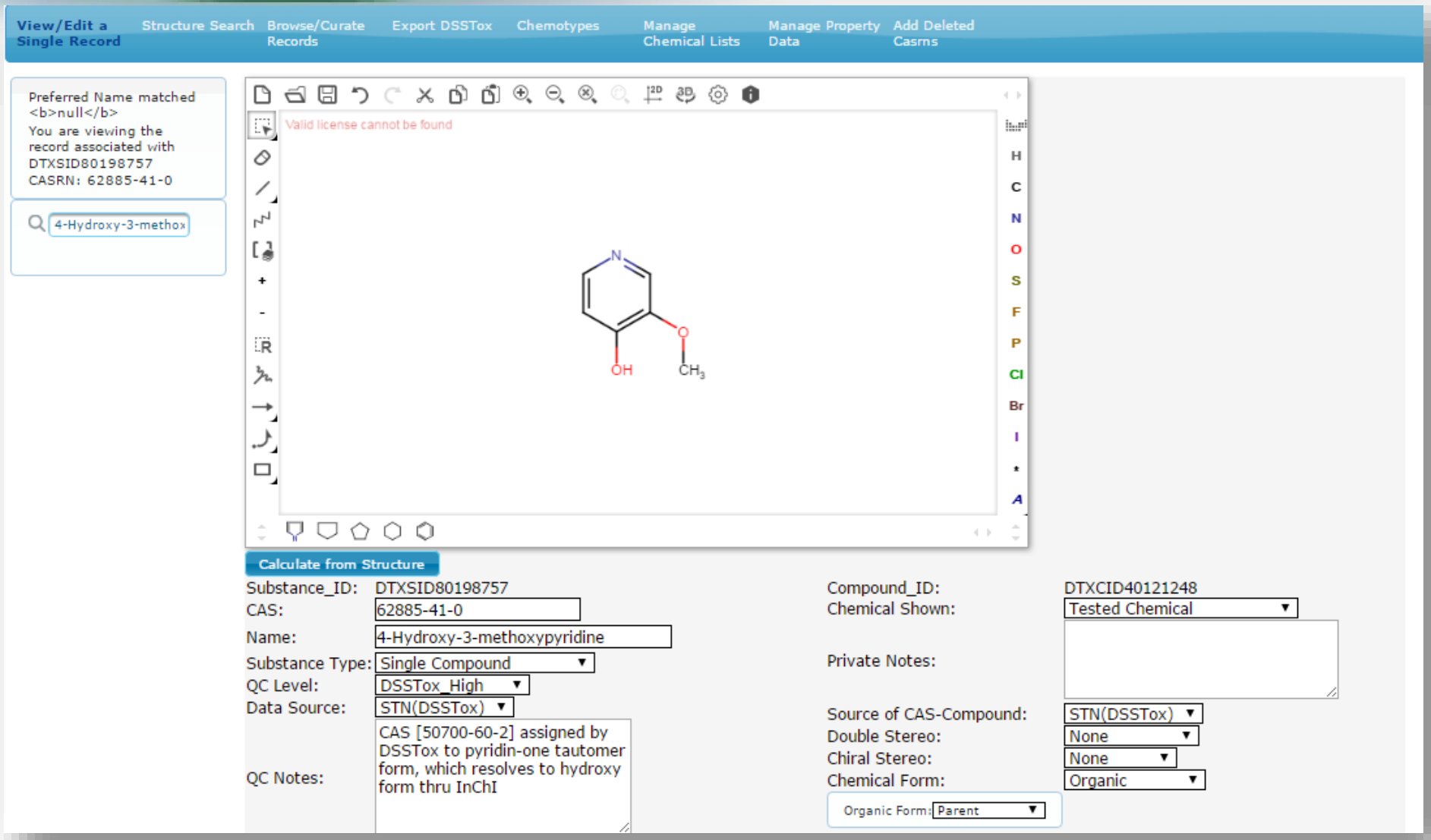
Intrinsic And Predicted Properties
- [x] Molecular Formula
- [ ] Average Mass
- [ ] Monoisotopic Mass
- [ ] TEST Model Predictions
- [x] OPERA Model Predictions

Metadata
- [ ] Curation Level Details
- [ ] NHANES/Predicted Exposure
- [ ] Data Sources
- [x] Include ToxVal Data Availability
- [x] Assay Hit Count
- [x] Number of PubMed Articles

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	INPUT	FOUND_BY	DTXSID	PREFERRE	CASRN	INCHIKEY	IUPAC_NAI	MOLECULA	TOXVAL_D	TOXCAST_	TOXCAST_	MS_READY	NUMBER_C	IRIS_LINK	PPRTV_LIN	ATMOSPHE	BIOCONCE
2	534-52-1	CAS-RN	DTXSID10	2-Methyl-4,6	534-52-1	ZXVONLUN	2-Methyl-4,6	C7H6N2O5	Y	27.56	261/947	CC1=C(O)C	179	-	Y	1.571E-12	2.88959
3	115-21-9	CAS-RN	DTXSID70	Ethyl silicon	115-21-9	ZOYFEXPF	Trichloro(eth	C2H5Cl3Si	Y	-	-	-	-	-	-	8.289E-12	13.9658
4	111-44-4	CAS-RN	DTXSID90	Bis(2-chloro	111-44-4	ZNSMNVML	1-Chloro-2-(	C4H8Cl2O	Y	1.12	10/891	ClCCOCCCl	12	Y	-	2.647E-12	9.99608
5	2763-96-4	CAS-RN	DTXSID50	Muscimol	2763-96-4	ZJQHPWUV	5-(Aminome	C4H6N2O2	Y	-	-	NCC1=CC((	4308	-	-	1.179E-10	5.05695
6	1464-53-5	CAS-RN	DTXSID00	2,2'-Bioxiran	1464-53-5	ZFIVKAOQE	2,2'-Bioxiran	C4H6O2	Y	-	-	C1OC1C1C	363	-	-	4.383E-12	1.27107
7	22224-92-6	CAS-RN	DTXSID30	Fenamiphos	22224-92-	ZCJPOPBZI	Ethyl 3-meth	C13H22NO3	Y	10.8	105/972	CCOP(=O)(	58	Y	-	1.66E-11	2.3394
8	359-06-8	CAS-RN	DTXSID40	Fluoroacetyl	359-06-8	ZBHDTYQJ	Fluoroacetyl	C2H2ClFO	Y	-	-	FCC(Cl)=O	-	-	-	3.513E-13	4.49379
9	5344-82-1	CAS-RN	DTXSID40	1-(o-Chlorop	5344-82-1	YZUKKTCD	N-(2-Chloro	C7H7ClN2S	Y	-	-	NC(=S)NC1	-	-	-	2.482E-11	9.95206
10	7446-18-6	CAS-RN	DTXSID10	Thallium (I) s	7446-18-6	YTQVHRVI	Dithallium(1+	O4STl2	Y	-	-	-	34	Y	Y	-	-
11	62207-76-5	CAS-RN	DTXSID40	Bis(3-fluoros	62207-76-	YRZXYIHD	-	C16H14CoF	Y	-	-	-	-	-	-	-	-
12	66-81-9	CAS-RN	DTXSID60	Cycloheximi	66-81-9	YPHMISFO	4-{(2R)-2-[(1	C15H23NO4	Y	32.27	294/911	CC1CC(C)C	18709	-	-	1.756E-11	2.81761
13	106-96-7	CAS-RN	DTXSID30	Propargyl br	106-96-7	YORCIIVHU	3-Bromopro	C3H3Br	Y	-	-	BrCC#C	-	-	-	1.069E-11	10.4968
14	315-18-4	CAS-RN	DTXSID70	Mexacarbat	315-18-4	YNEVBPNZ	4-(Dimethyla	C12H18N2O	Y	5.11	12/235	CNC(=O)OC	27	-	-	1.447E-11	26.2914
15	110-00-9	CAS-RN	DTXSID60	Furan	110-00-9	YLQBMQCU	Furan	C4H4O	Y	0.0	0/235	O1C=CC=C	919	Y	-	4.019E-11	5.01648
16	3037-72-7	CAS-RN	DTXSID20	4-(diethoxyn	3037-72-7	YHFFINXFN	4-[Diethoxy(	C9H23NO2S	Y	-	-	-	-	-	-	5.95E-12	6.71292
17	75-44-5	CAS-RN	DTXSID00	Phosgene	75-44-5	YGYAWVDI	Carbonyl dic	CCl2O	Y	-	-	ClC(Cl)=O	489	Y	-	9.994E-16	13.0711
18	2032-65-7	CAS-RN	DTXSID30	Methiocarb	2032-65-7	YFBPRJGD	3,5-Dimethy	C11H15NO2	Y	18.14	88/485	CNC(=O)OC	65	-	-	1.446E-11	34.1692
19	2778-04-3	CAS-RN	DTXSID20	Endothion	2778-04-3	YCAGGFXS	S-[(5-Metho	C9H13O6PS	Y	-	-	COC1=COC	-	-	-	4.355E-11	1.18341
20	12108-13-3	CAS-RN	DTXSID90	(Methylcyclc	12108-13-	YASXMYPV	Tricarbonyl(	C9H7MnO3	Y	2.56	11/430	-	68	-	-	-	-
21	7803-51-2	CAS-RN	DTXSID20	Phosphine	7803-51-2	XYFCBTPG	Phosphane	H3P	Y	-	-	-	928	Y	-	-	-
22	107-18-6	CAS-RN	DTXSID80	Allyl alcohol	107-18-6	XXROGKLT	Prop-2-en-1	C3H6O	Y	3.99	17/426	OCC=C	627	Y	Y	2.592E-11	4.03901
23	108-05-4	CAS-RN	DTXSID30	Vinyl acetate	108-05-4	XTXRWKRV	Ethenyl acet	C4H6O2	Y	1.7	4/235	CC(=O)OC=	206	Y	-	2.5E-11	5.52157
24	19624-22-7	CAS-RN	DTXSID10	Pentaborane	19624-22-	XPIBKKWN	-	B5H9	Y	-	-	-	-	-	-	-	-
25	75-74-1	CAS-RN	DTXSID00	Tetramethyl	75-74-1	XOOGZRUI	Tetramethyl	C4H12Pb	Y	-	-	C[Pb](C)(C)	24	-	-	-	-

Worksheet1

Underneath the Dashboard



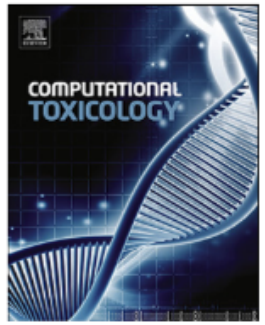
A little more about our data quality

Computational Toxicology 12 (2019) 100096

Contents lists available at ScienceDirect

Computational Toxicology

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



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

Christopher M. Grulke[a], Antony J. Williams[a], Inthirany Thillanadarajah[b], Ann M. Richard[a,*]

[a] National Center for Computational Toxicology, Office of Research & Development, US Environmental Protection Agency, Mail Drop D143-02, Research Triangle Park, NC 27711, USA

[b] Senior Environmental Employment Program, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA

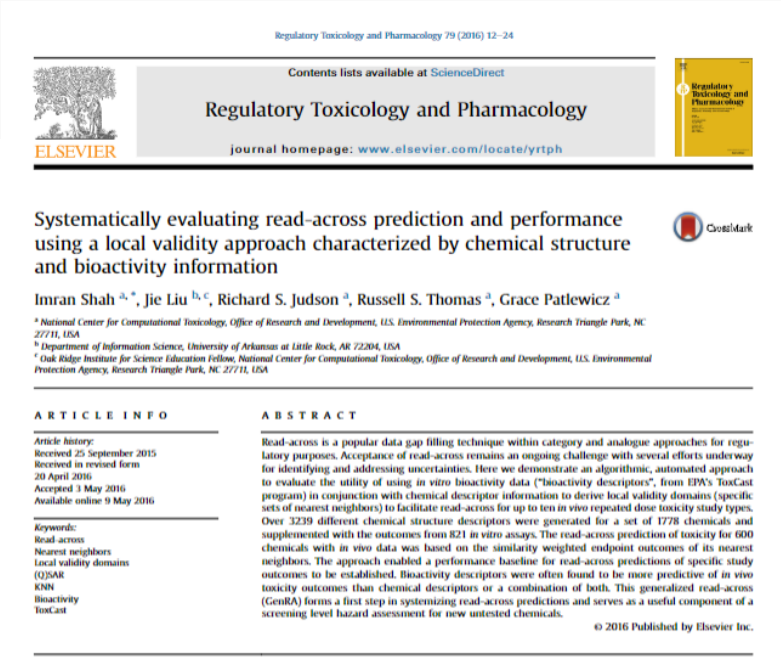
EPA
United States
Environmental Protection
Agency

Generalized Read-Across

GenRA (Generalised Read-Across)

EPA United States Environmental Protection Agency

- Predicting toxicity as a similarity-weighted activity of nearest neighbours based on chemistry and bioactivity descriptors (Shah et al, 2016)
- Goal: To establish an objective performance baseline for read-across and quantify the uncertainty in the predictions made



Regulatory Toxicology and Pharmacology 79 (2016) 12–24

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

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]

[a] National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA
[b] Department of Information Science, University of Arkansas at Little Rock, AR 72204, USA
[c] Oak Ridge Institute for Science Education Fellow, National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA

ARTICLE INFO

Article history:
Received 25 September 2015
Received in revised form
20 April 2016
Accepted 3 May 2016
Available online 9 May 2016

Keywords:
Read-across
Nearest neighbors
Local validity domains
(Q)SAR
KNN
Bioactivity
ToxCast

ABSTRACT

Read-across is a popular data gap filling technique within category and analogue approaches for regulatory purposes. Acceptance of read-across remains an ongoing challenge with several efforts underway for identifying and addressing uncertainties. Here we demonstrate an algorithmic, automated approach to evaluate the utility of using *in vitro* bioactivity data ("bioactivity descriptors", from EPA's ToxCast program) in conjunction with chemical descriptor information to derive local validity domains (specific sets of nearest neighbors) to facilitate read-across for up to ten *in vivo* repeated dose toxicity study types. Over 3239 different chemical structure descriptors were generated for a set of 1778 chemicals and supplemented with the outcomes from 821 *in vitro* assays. The read-across prediction of toxicity for 600 chemicals with *in vivo* data was based on the similarity weighted endpoint outcomes of its nearest neighbors. The approach enabled a performance baseline for read-across predictions of specific study outcomes to be established. Bioactivity descriptors were often found to be more predictive of *in vivo* toxicity outcomes than chemical descriptors or a combination of both. This generalized read-across (GenRA) forms a first step in systemizing read-across predictions and serves as a useful component of a screening level hazard assessment for new untested chemicals.

$$y_i^{\beta,\alpha} = \frac{\sum_j^k s_{ij}^{\alpha} x_j^{\beta}}{\sum_j^k s_{ij}^{\alpha}}$$

Jaccard similarity:

$$s_{ij} = \frac{\sum_l (x_{il} \wedge x_{jl})}{\sum_l (x_{il} \vee x_{jl})}$$

$\alpha \in \{chm, bio, bc\}$

$\beta \in \{bio, tox\}$

y_i = *predicted activity of chemical* (c_i)

x_j^{β} = *activity of* c_j in β

s_{ij}^{α} = *Jacccard similarity between* $\boldsymbol{x}_i^{\alpha}, \boldsymbol{x}_j^{\alpha}$

k = *up to* k *nearest neighbours*

Read-across workflow in GenRA v1.0

EPA United States Environmental Protection Agency

Decision Context

Screening level assessment of hazard based on toxicity effects from ToxRefDB v1

Analogue identification

Similarity context is based on structural characteristics

Data gap analysis for target and source analogues

Analogue evaluation

Evaluate consistency and concordance of experimental data of source analogues across and between endpoints

Read-across

Similarity weighted average – many to one read-across

Uncertainty assessment

Assess prediction and uncertainty using AUC and p value metrics

GenRA – Current research

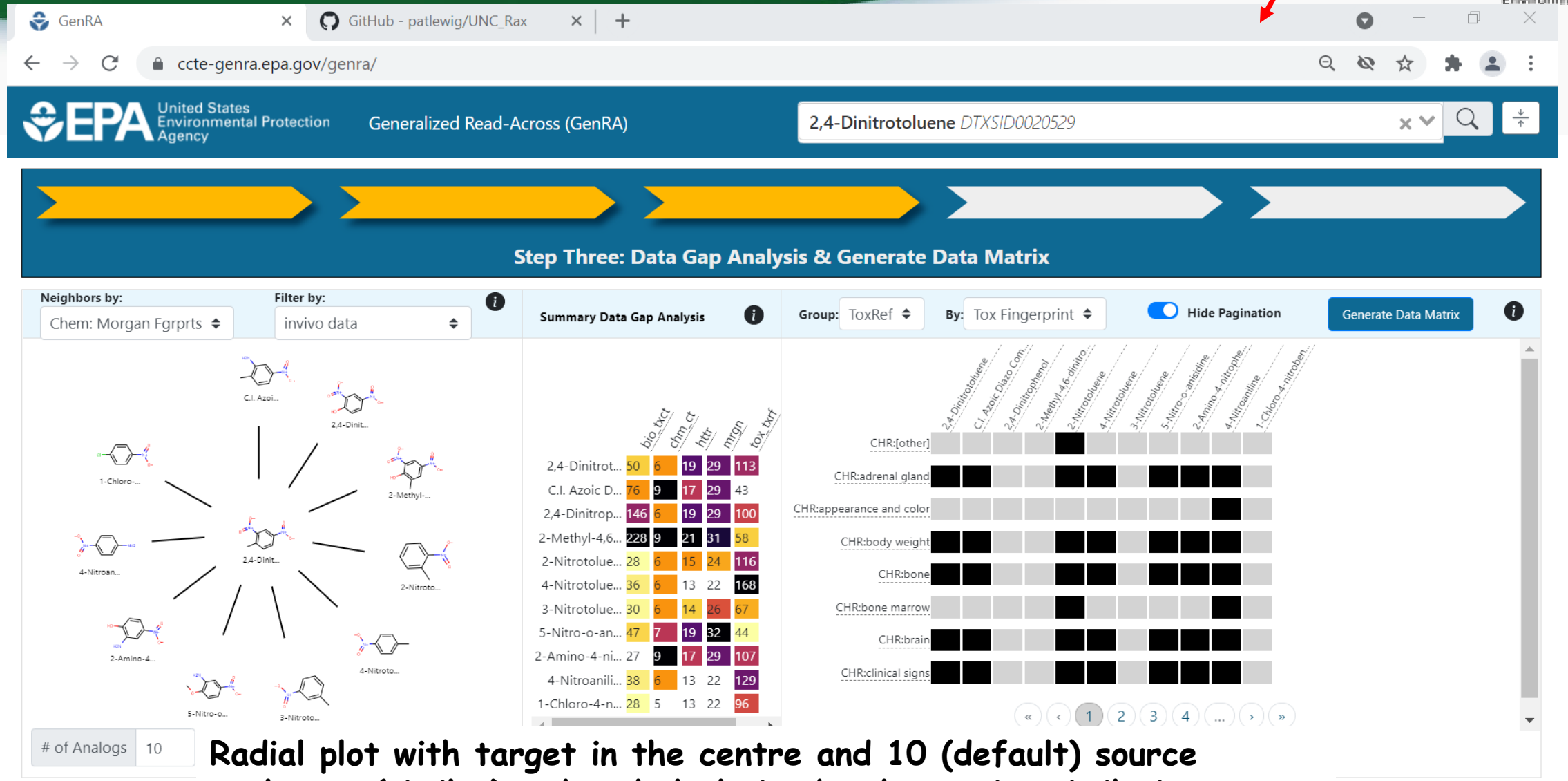
- Consideration of other information to define and refine the analogue selection & evaluation
  - physicochemical similarity (Helman et al 2018)
  - metabolic similarity (*in prep*),
  - reactivity similarity (Nelms et al 2018)
  - transcriptomics similarity (Tate et al, 2021 *in press*)
- Transitioning to quantitative predictions of toxicity
  - Using GenRA to predict Lowest Observed Adverse Effect Level (LOAEL), acute oral (median lethal dose) LD50 (Helman et al 2019a,b)
- Developing a compendium of expert driven read-across examples to investigate how data driven read-across with NAM data can mirror expert assessments (*in prep*)

GenRA Summary

- GenRA is an attempt to move towards an objective read-across approach where uncertainties and performance can be quantified. Provides opportunities for NAM data to be systemically incorporated.
- GenRA v1.0 established a baseline in performance. The approach relies on chemical descriptors to predict binary toxicity values but work continues to characterise other contexts of similarity (e.g. mechanistic, reactivity, metabolism) and quantify their contribution in predicting *in vivo* toxicity outcomes.
- GenRA v2.0 is anticipated to be released in the coming weeks as a standalone web application linked to the Dashboard. A python package (genra-py) was released (March 2021) to facilitate batch processing using user specific datasets – see https://github.com/i-shah/genra-py. The ability to try out the package using a sample dataset is available here https://github.com/patlewig/UNC_Rax

GenRA v2 in practice

Search for a chemical of interest (target) using the search box



Radial plot with target in the centre and 10 (default) source analogues (similar) ordered clockwise by decreasing similarity (Jaccard) on the basis of Morgan FP (default FP)

GenRA v2 in practice

ccte-genra.epa.gov/genra/

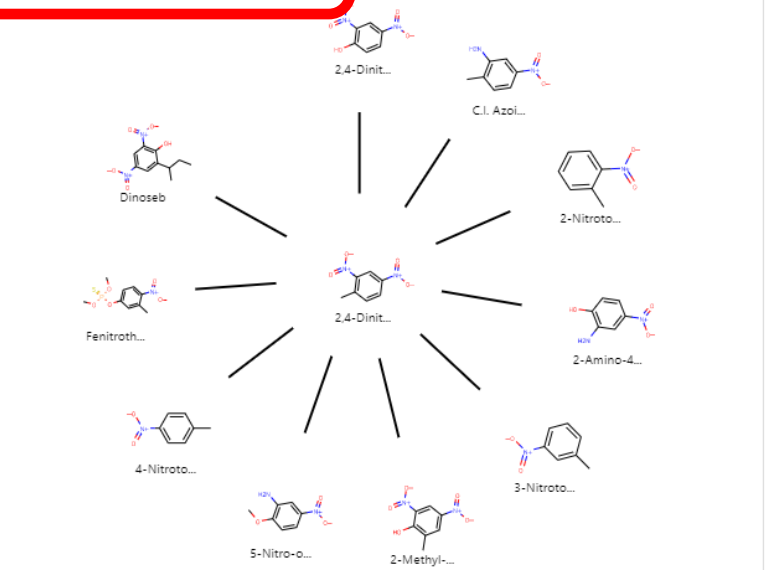
EPA United States Environmental Protection Agency | Generalized Read-Across (GenRA)

2,4-Dinitrotoluene *DTXSID0020529*

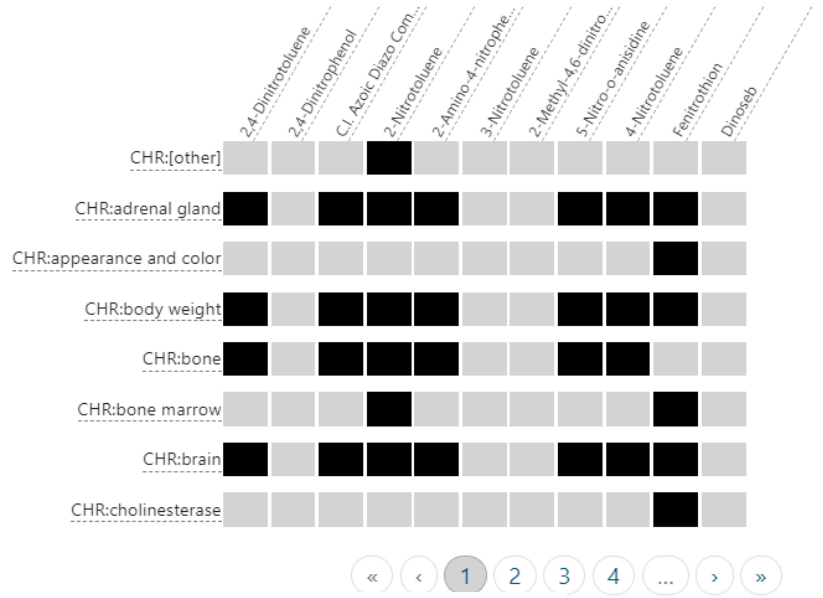
Two hybrid FP combinations added: Chem:Bioactivity

Step Three: Data Gap Analysis & Generate Data Matrix

Neighbors by: Morg2:Tors1:Bio1 | Filter by: invivo data | Summary Data Gap Analysis | Group: ToxRef | By: Tox Fingerprint | Hide Pagination | Generate Data Matrix



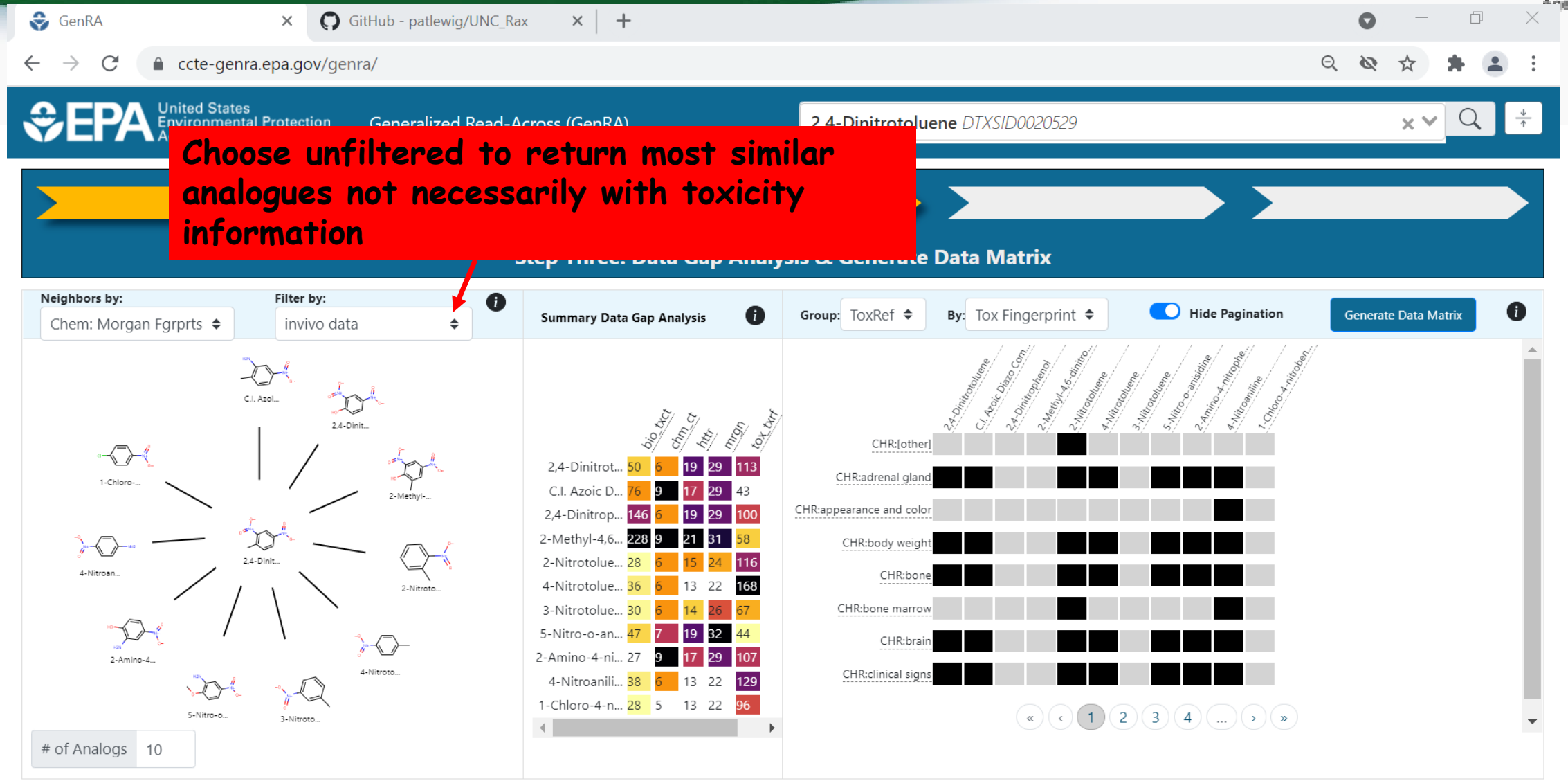
	bio_txct	chm_ct	httr	mrgn	tox_txrf
2,4-Dinitrot...	50	6	19	29	113
2,4-Dinitrop...	146	6	19	29	100
C.I. Azoic D...	76	9	17	29	43
2-Nitrotolue...	28	6	15	24	116
2-Amino-4-ni...	27	9	17	29	107
3-Nitrotolue...	30	6	14	26	67
2-Methyl-4,6...	228	9	21	31	58
5-Nitro-o-an...	47	7	19	32	44
4-Nitrotolue...	36	6	13	22	168
Fenitrothion	210	8	27	40	166
Dinoseb	252	12	27	39	32



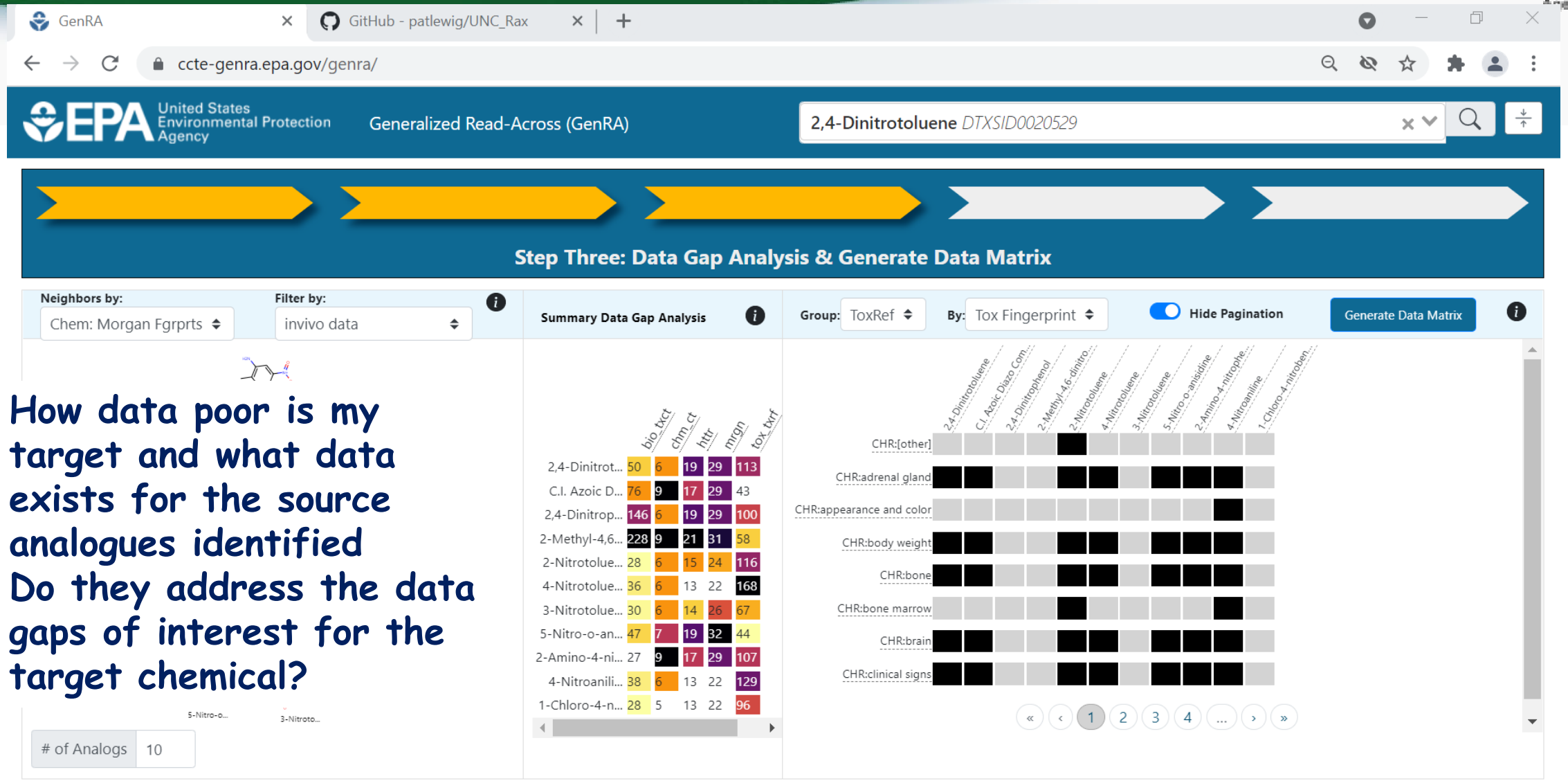
# of Analogs 10

Radial plot with target in the centre and source analogues (similar) ordered clockwise by decreasing similarity (Jaccard) on the basis of hybrid FP

GenRA v2 tool in practice



GenRA v2 tool in practice



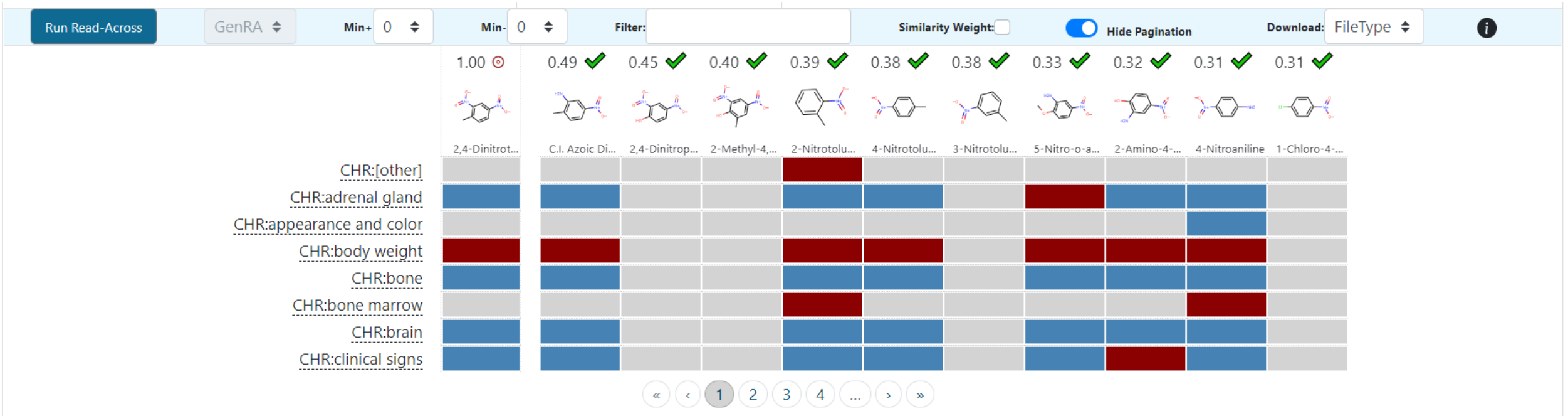
- How data poor is my target and what data exists for the source analogues identified
- Do they address the data gaps of interest for the target chemical?

GenRA v2 tool in practice

EPA United States Environmental Protection Agency

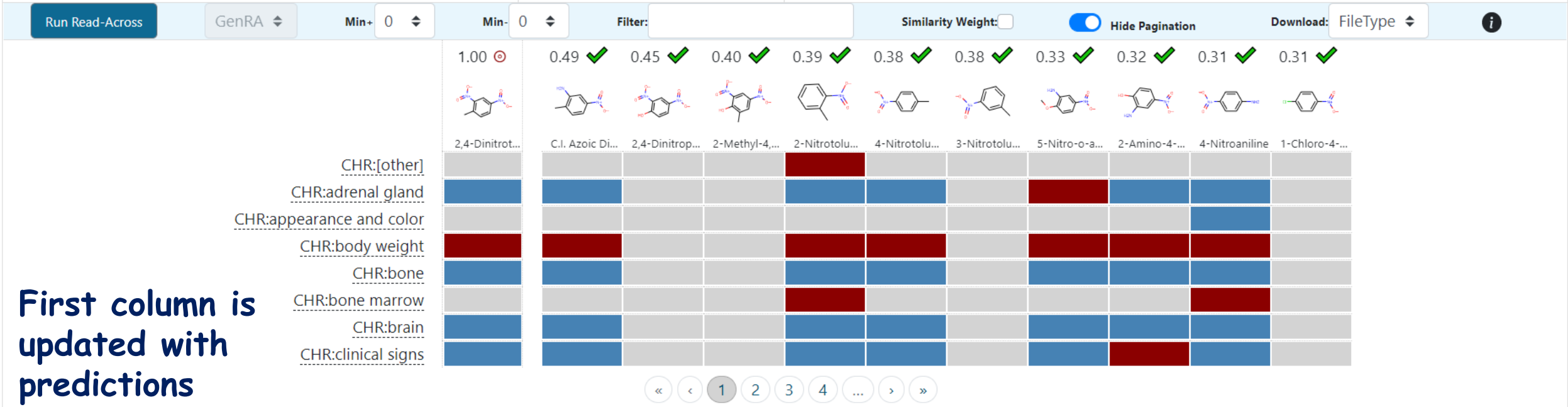
What is the consistency and concordance across my source analogues?
Should I deselect analogues from consideration from the entire set of predictions?
Should I consider subcategorising the analogues selected?

Toxicity data represented as binary outcomes – red (positive), blue (negative), grey (no data)

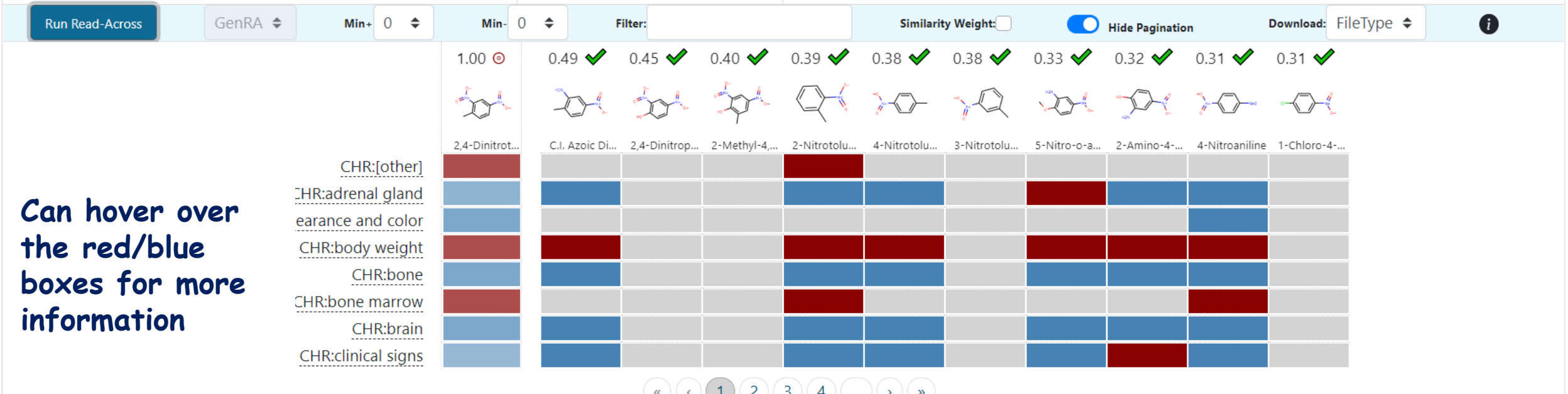


GenRA v2 tool in practice

EPA



First column is updated with predictions



Can hover over the red/blue boxes for more information

GenRA v2 Tool in practice

EPA United States Environmental Protection Agency

- Database underpinning GenRA v2.0: ToxRefDB v2
  - Different study types and effects within them are predicted e.g. chronic_liver is annotated as CHR_liver
  - Negative effects are inferred from guideline profiles which define the required tests for each study type. The assumption is that the study required an evaluation but no effects were reported.
  - Positive results – min dose at which toxicity effects are observed in a study
- Prediction: Similarity weighted activity
- Performance is categorised by the Area under the Curve (AUC) of the Receiving Operating Characteristic (ROC)
  - The significance was empirically estimated by constructing a null distribution by permuting the toxicity values 100 times and calculating the fraction of times the AUC was more extreme than what would be observed by chance (this is reported as the p-value).

GenRA v2 Tool in practice

- Ability to export the predictions as an excel file
- Output can be analysed in different ways

	A	B	C	D	
1	**role**	target	analog	analog	
2	**preferred name**	2,4-Dinitrotoluene	C.I. Azoic Diazo Component 12	2,4-Dinitrophenol	2-
3	**dsstox_sid**	DTXSID0020529	DTXSID4020959	DTXSID0020523	
4	**molecular weight**	182.135	152.153	184.107	
5	**similarity**	1.00001	0.487179487	0.45	
6	CHR:[other]	GenRA Pos Act=1 (1) AUC=0 p=1	no_data	no_data	
7	CHR:adrenal gland	GenRA TN Act=0 (0.147) AUC=0 p=0.845	no_effect	no_data	
8	CHR:appearance and color	GenRA Neg Act=0 (0) AUC=0 p=1	no_data	no_data	
9	CHR:body weight	GenRA TP Act=1 (1) AUC=0 p=1	2.5mg/kg/day	no_data	
10	CHR:bone	GenRA TN Act=0 (0) AUC=0 p=1	no_effect	no_data	
11	CHR:bone marrow	GenRA Pos Act=1 (1) AUC=0 p=1	no_data	no_data	
12	CHR:brain	GenRA TN Act=0 (0) AUC=0 p=1	no_effect	no_data	
13	CHR:clinical signs	GenRA TN Act=0 (0.144) AUC=0 p=0.785	no_effect	no_data	
14	CHR:clitoral gland	GenRA TN Act=0 (0.326) AUC=0 p=0.96	no_effect	no_data	
15	CHR:ear	GenRA Pos Act=1 (0.562) AUC=0 p=0.51	no_data	no_data	
16	CHR:epididymis	GenRA TN Act=0 (0.178) AUC=0 p=0.865	no_effect	no_data	
17	CHR:erythrocyte (rbc) count differential	GenRA Pos Act=1 (1) AUC=0 p=1	no_data	no_data	

GenRA v2 Tool in practice

- Rank order positive results based on AUC and p values
- Look at the distribution of positive vs negatives predictions
- Explore what effects are being identified for the source analogues – consider identifying the underlying data for source analogues (elsewhere on the Dashboard) – is there a critical effect that is driving the toxicity that should be compared with the target chemical predictions?
- ……
- Depends on the decision context and the level of uncertainty that can be tolerated.

GenRA v2 Tool in practice

- If for example, body weight effects were important based on the source analogues, we could look at the prediction in GenRA for our target chemical
- GenRA prediction is positive for body weight effects in a chronic CHR (reflected by the True Positive (TP)) as summarised in the Download report
- What if we wanted to predict a point of departure from the source analogues…?
- We can also compute a similarity weighted average *(recall the formula in an earlier slide..).*

GenRA v2 Tool in practice

EPA United States Environmental Protection Agency

role	target	analog	analog	analog
preferred name	2,4-Dinitrotoluene	C.I. Azoic Diazo Component 12	2,4-Dinitrophenol	2-Methyl-4,6-dinitrophenol
dsstox_sid	DTXSID0020529	DTXSID4020959	DTXSID0020523	DTXSID1022053
molecular weight	182.135	152.153	184.107	198.134
similarity	1	0.487179487	0.45	0.395348837
CHR:[other]	GenRA Pos Act=1 (1) AUC=0 p=1	no_data	no_data	no_data
CHR:adrenal gland	GenRA TN Act=0 (0.171) AUC=0 p=0.865	no_effect	no_data	no_data
CHR:body weight	GenRA TP Act=1 (1) AUC=0 p=1	2.5mg/kg/day	no_data	no_data
CHR:bone				
CHR:bone marr				
CHR:brain				
CHR:clinical sig				
CHR:clitoral gla				
CHR:ear				
CHR:epididymis				
CHR:esophagu				

	A	B	C	D	E	F	G	H
1	role	preferred name	dsstox_sid	molecular weight	similarity	CHR:body weight	CHR:body weight	-log10(toxicity/MW)
2	target	2,4-Dinitrotoluene	DTXSID0020529	182.135	1.00001	GenRA TP Act=1 (1) AUC=0 p=1	GenRA TP Act=1 (1) AUC=0 p=1	
3	analog	C.I. Azoic Diazo Component 12	DTXSID4020959	152.153	0.4871795	2.5mg/kg/day	2.5	1.784340511
4	analog	2,4-Dinitrophenol	DTXSID0020523	184.107	0.45	no_data	no_data	
5	analog	2-Methyl-4,6-dinitrophenol	DTXSID1022053	198.134	0.3953488	no_data	no_data	
6	analog	2-Nitrotoluene	DTXSID4025791	137.138	0.3947368	50mg/kg/day	50	0.438187807
7	analog	4-Nitrotoluene	DTXSID5023792	137.138	0.3783784	60mg/kg/day	60	0.359006561
8	analog	3-Nitrotoluene	DTXSID5021831	137.138	0.375	no_data	no_data	
9	analog	5-Nitro-o-anisidine	DTXSID0020943	168.152	0.326087	200mg/kg/day	200	-0.075327958

	similarity weighted activity	-log10 molar equivalent	toxicity in mg/kg/day
sum(pairwise similarity * toxicity)	1.547415827	0.69947419	36.38473532
sum(pairwise similarities)	2.21225579		

Target is predicted to have a NOAEL of 36 mg/kg/day in a CHR study on the basis of body weight effects

GenRA – Overall goal

- Quantify the contribution that different similarity contexts play in toxicity prediction and how that differs depending on the toxicity endpoint of interest, the chemical of interest and whether it mirrors expert driven read-across

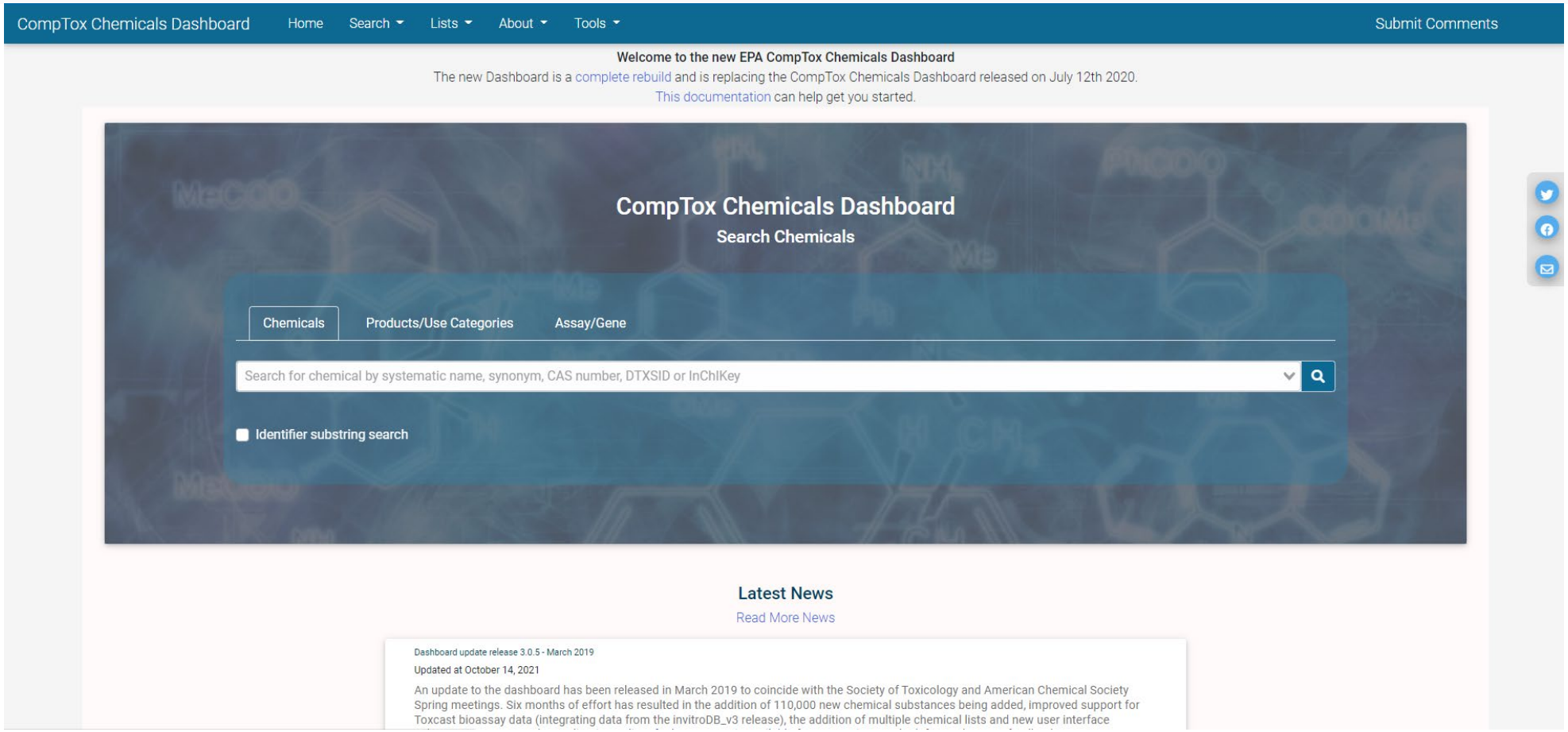
- Quantify level of confidence for prediction made

=> objective, reproducible read-across assessments

GenRA v3…new features..

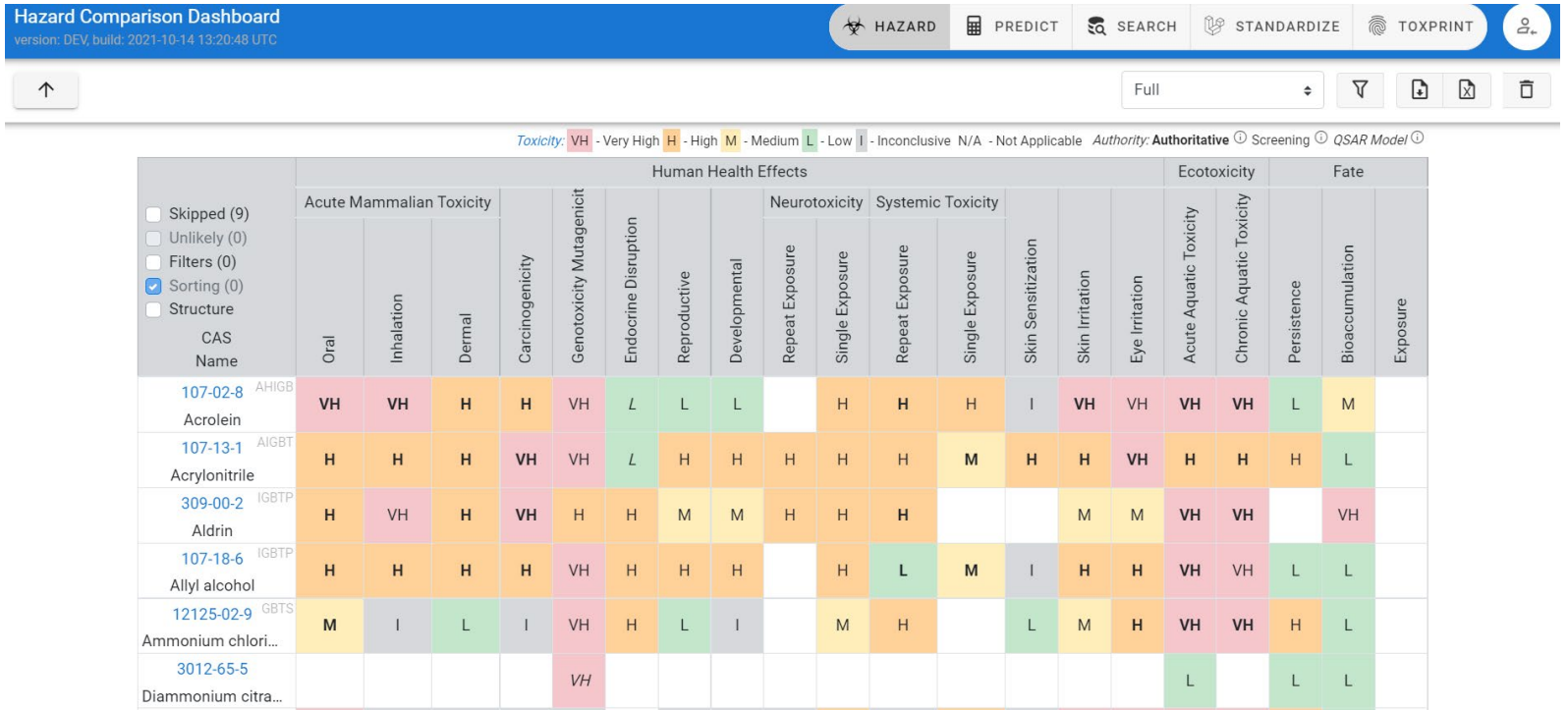
- Introduce structures not already registered in the Dashboard
- Custom fingerprints

Coming Soon….NEW DASHBOARD https://ccte-ccd.epa.gov/dashboard/

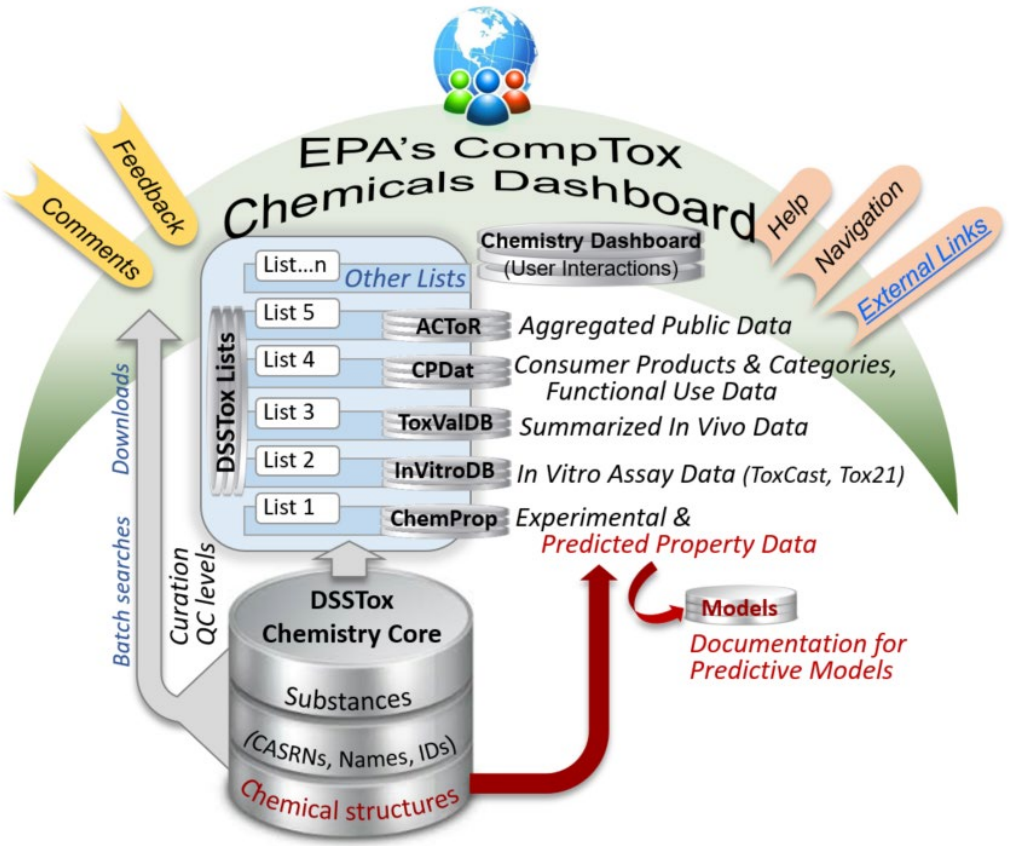


Proof-of-Concept Apps

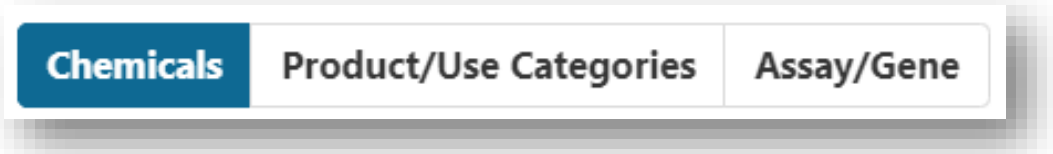
- Structure-substructure-similarity searching of Dashboard
- Batch prediction of physicochemical properties and toxicity
- “Hazard Comparison Dashboard” for comparing hazard profiles for chemicals



Summary and Conclusion



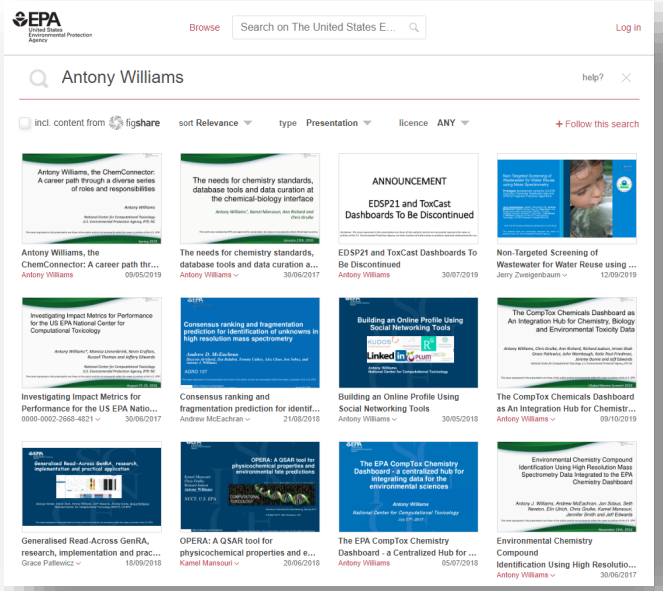
- CompTox Chemicals Dashboard - a central hub for environmental data
  - ~875k chemical substances
  - Integrating property data, hazard data, exposure data, *in vitro* bioactivity data
  - Interrogation of bioactivity data -
  - Multiple types of searches

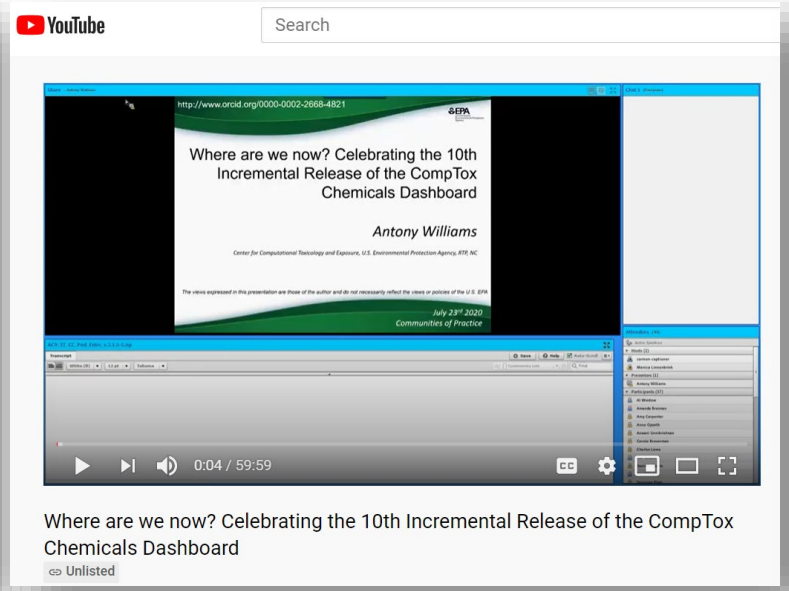


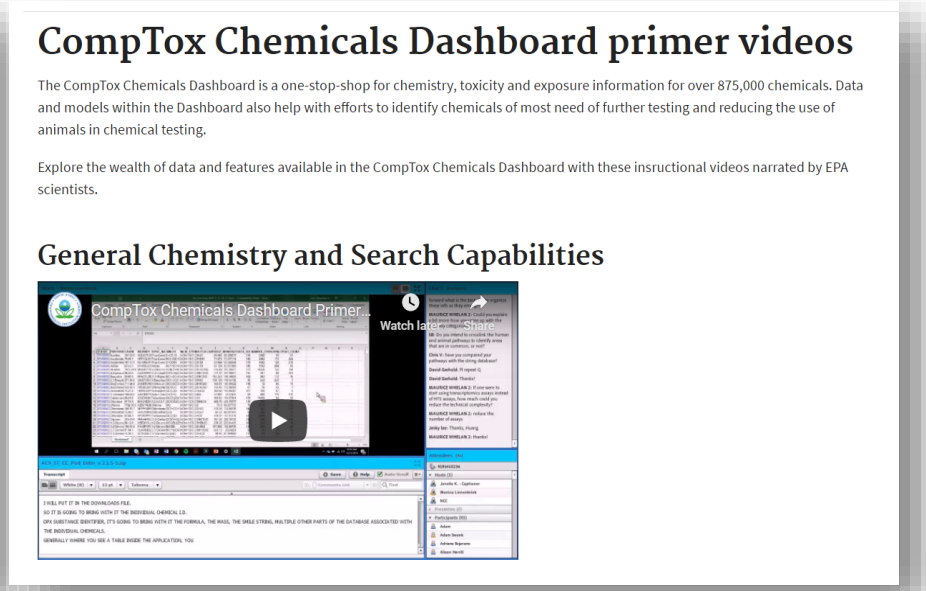
  - Batch search for thousands of chemicals
  - Real-time property and toxicity predictions
  - Downloadable files – CSV, TSV and Excel

You want to know more…

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







Acknowledgments

- Contacts:

Williams.Antony@epa.gov

Patlewicz.grace@epa.gov

- Feedback and follow-up is welcomed! Your questions help
- The dashboard is based on the efforts of many more team members than us. Many collaborators provide data also.

EPA's Center for Computational Toxicology and Exposure