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

Delivering The Benefits of Chemical-Biological Integration in Computational Toxicology at the EPA

Antony Williams

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

November 7th, 2016
German Cheminformatics Conference, Fulda, Germany

Overview

- An introduction to NCCT
- Our Data and Our Dashboards
- The CompTox Chemistry Dashboard
  - Architecture
  - Data and Models: e,g, Physicochemical Properties
  - Application to Non-Targeted Screening
- Coming Soon
- Future Work

Who is NCCT?

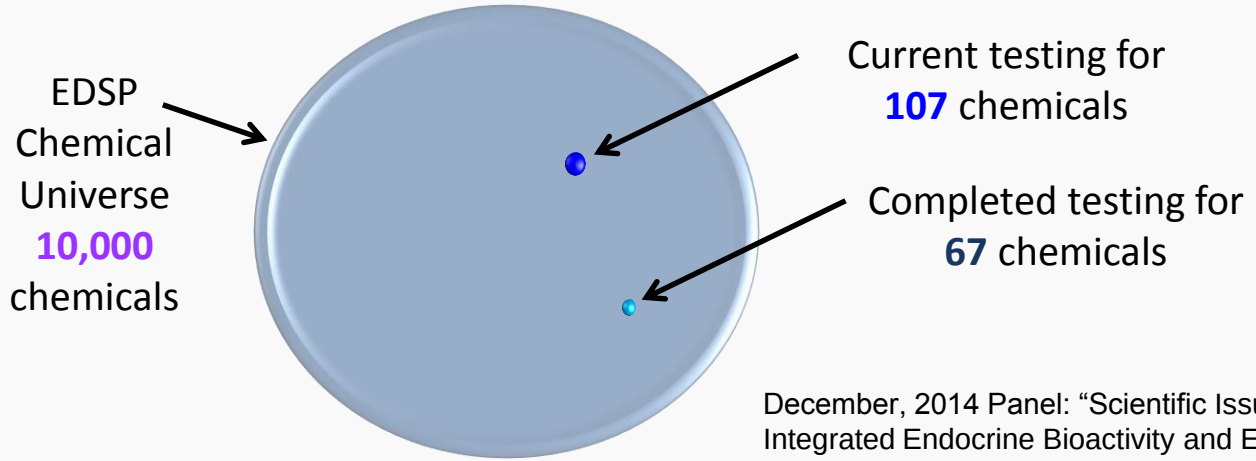
- National Center for Computational Toxicology – part of EPA's Office of Research and Development

- Research driven by EPA's *Chemical Safety for Sustainability Research Program*
  - Develop new approaches to **evaluate** the **safety** of chemicals
  - Integrate advances in biology, biotechnology, chemistry, exposure science and computer science

- Goal - To identify **chemical exposures** that may disrupt biological processes and cause adverse outcomes.

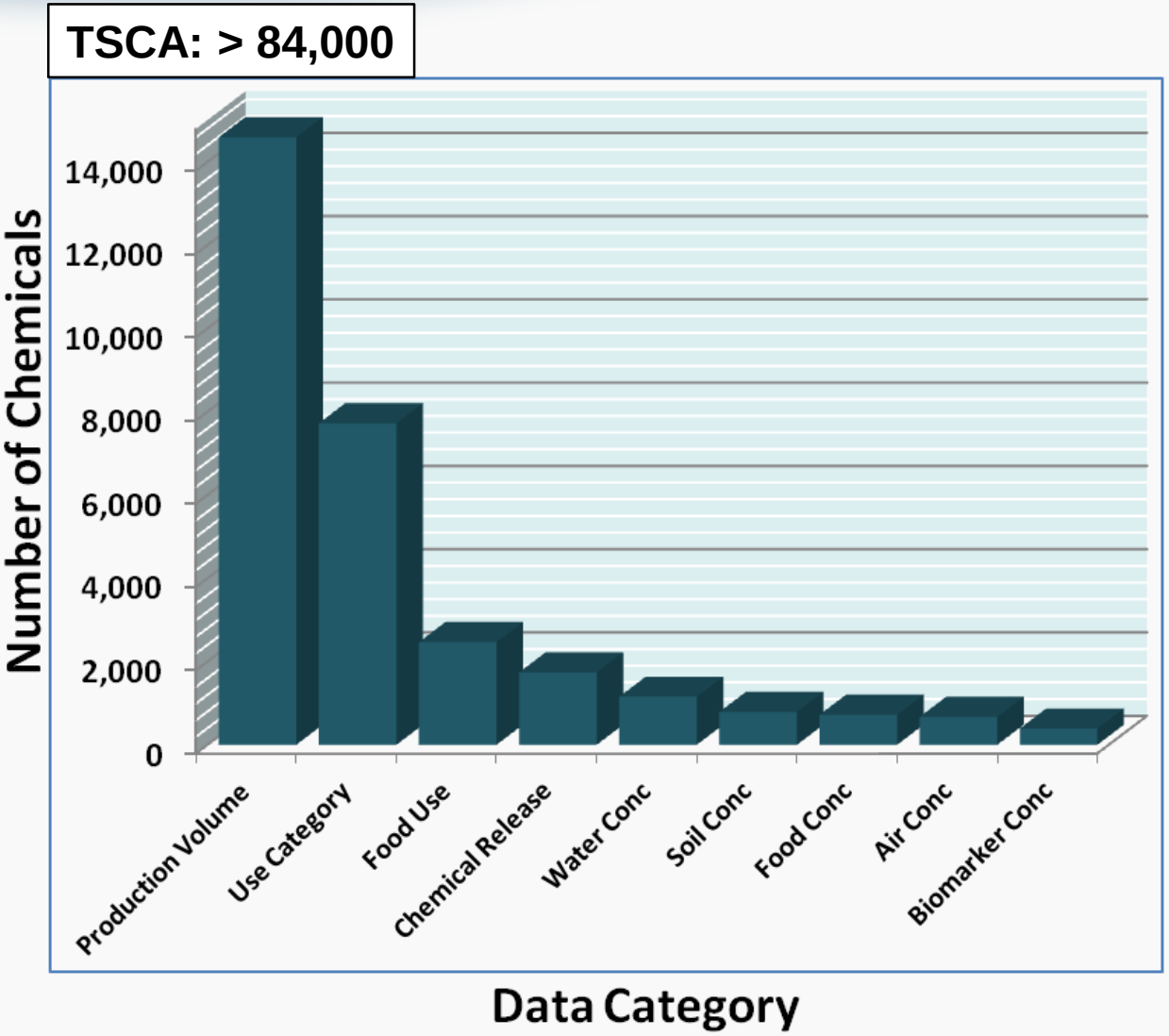
Number of Chemicals in Commerce Presents Regulatory Challenges

ONE LIST: EPA's Endocrine Disruptor Screening Program (EDSP) List	# of Compounds
Conventional Active Ingredients	838
Antimicrobial Active Ingredients	324
Biological Pesticide Active Ingredients	287
Non Food Use Inert Ingredients	2,211
Food Use Inert Ingredients	1,536
Fragrances used as Inert Ingredients	1,529
Safe Drinking Water Act Chemicals	3,616
TOTAL	**10,341**



December, 2014 Panel: "Scientific Issues Associated with Integrated Endocrine Bioactivity and Exposure-Based Prioritization and Screening" DOCKET NUMBER: EPA–HQ–OPP–2014–0614

Exposure Data Cannot Keep Pace with Regulatory Needs



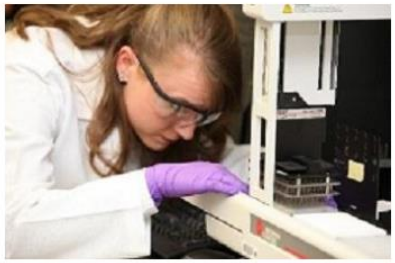
P.P. Egeghy et al. *Sci Total Environ.* 414 (2012) 159–166

We need more data and derivative models and algorithms

- Our outputs include a lot of data, models, algorithms and software applications
- We produce Open Data – we want people to interrogate it, learn from it, develop understanding

Toxicity Forecasting

Advancing the Next Generation of Chemical Evaluation



EPA needs rapid and efficient methods to prioritize, screen and evaluate thousands of chemicals. EPA's Toxicity Forecaster (ToxCast) generates data and predictive models on thousands of chemicals of interest to the EPA. ToxCast uses high-throughput screening methods and computational toxicology approaches to rank and prioritize chemicals. In fact, EPA's Endocrine Disruption Screening Program (EDSP) is working to use ToxCast to rank and prioritize chemicals.

- ToxCast has data on over 1,800 chemicals from a broad range of sources including industrial and consumer products, food additives, and potentially "green" chemicals that could be safer alternatives to existing chemicals.
- ToxCast screens chemicals in over 700 high-throughput assays that cover a range of high-

Downloadable Computational Toxicology Data

EPA's computational toxicology research efforts evaluate the potential health effects of thousands of chemicals. The process of evaluating potential health effects involves generating data that investigates the potential harm, or hazard of a chemical, the degree of exposure to chemicals as well as the unique chemical characteristics.

As part of EPA's commitment to share data, all of the computational toxicology data is publicly available for anyone to access and use.

High–throughput Screening Data

EPA researchers use rapid chemical screening (called high-throughput screening assays) to limit the number of laboratory animal tests while quickly and efficiently testing thousands of chemicals for potential health effects.

- ToxCast Data: High-throughput screening data on thousands of chemicals.

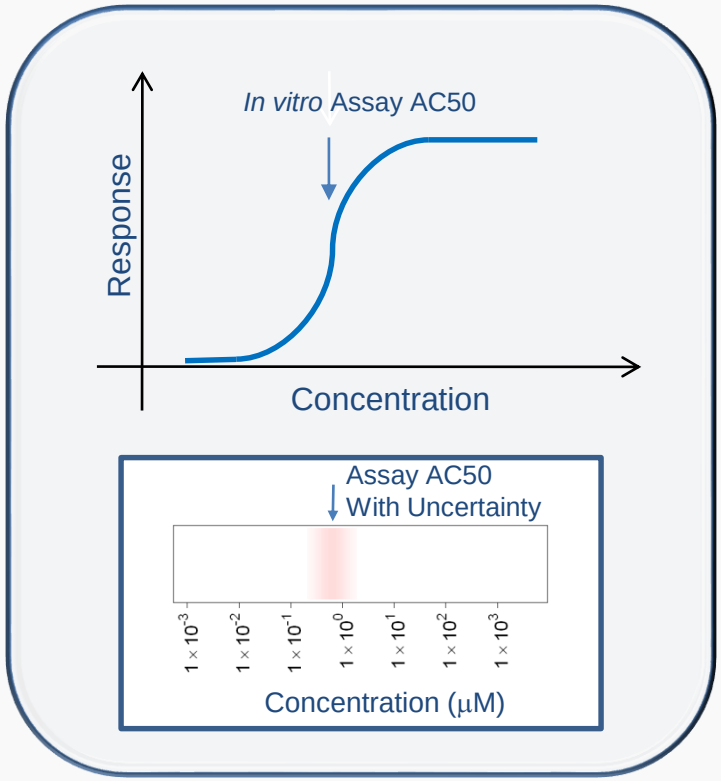
Rapid Exposure and Dose Data

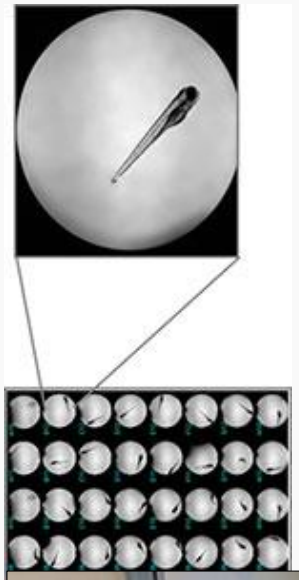
EPA researchers develop and use rapid exposure estimates to predict potential exposure for thousands of chemicals.

- High-throughput toxicokinetics data: It is important to link the external dose of a chemical to an internal blood or tissue concentration, this process is called toxicokinetics. EPA researchers measure the critical factors that determine the distribution

High-Throughput Bioactivity to Identify **Potential Hazards**

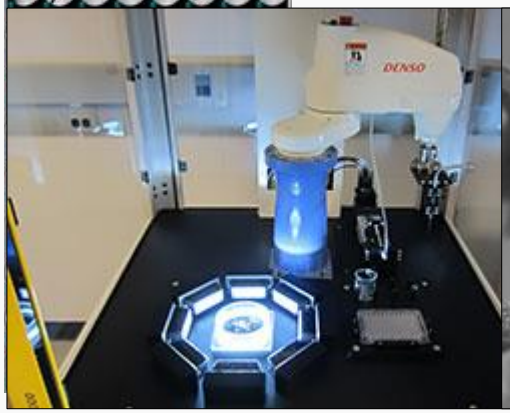
- Assays in dose-response format (50% activity concentration – AC50 – and efficacy if data described by a Hill function)

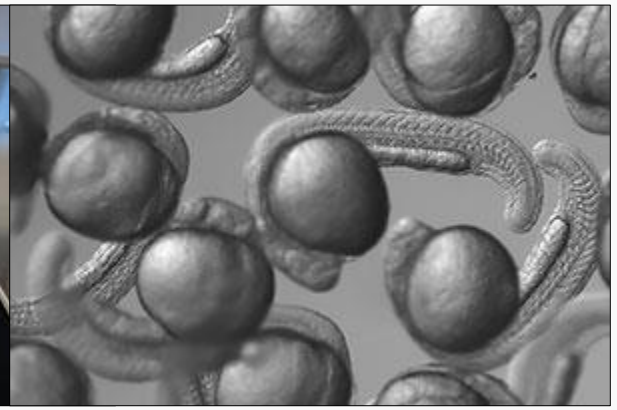




All data is made public:
http://actor.epa.gov/dashboard/

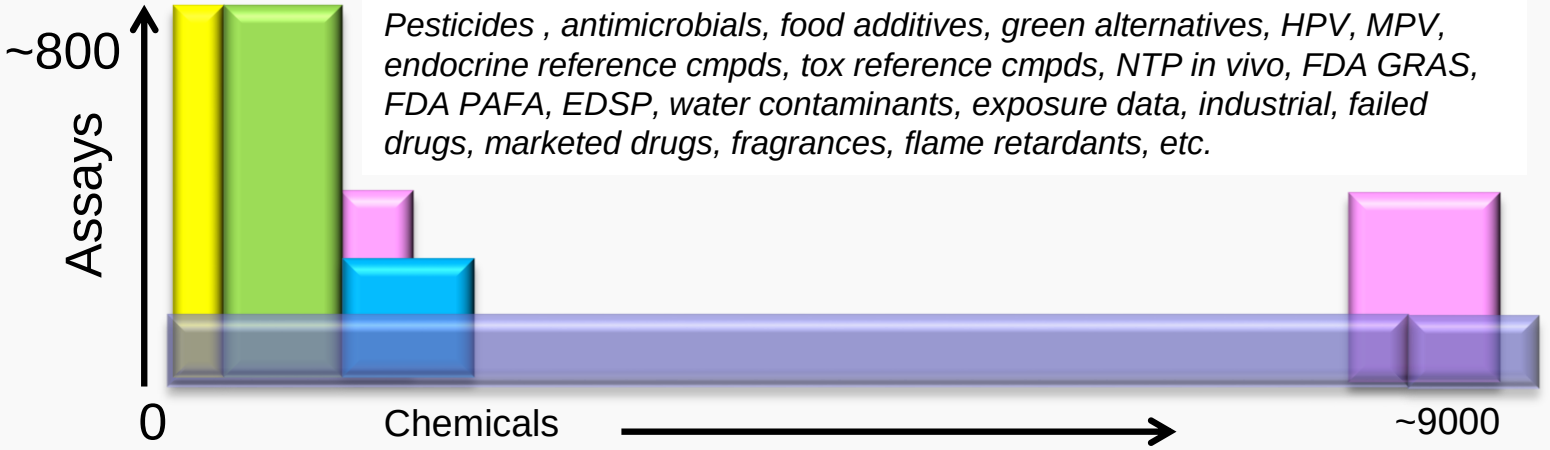
New datasets added continuously (for example, 1060 chemicals tested by Truong et al., 2014 zebrafish assay)



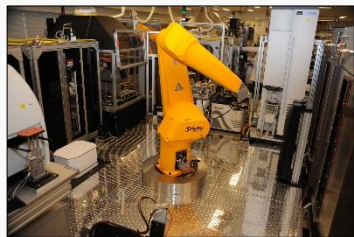


ToxCast & Tox21: Chemicals, Data and Release

Set	Chemicals	Assays	Endpoints	Completion	Available
ToxCast Phase I	293	~600	~700	2011	Now
ToxCast Phase II	767	~600	~700	03/2013	Now
ToxCast E1K	800	~50	~120	03/2013	Now
ToxCast Phase III	~900	~300	~300	In progress	2016
Tox21	~9000	~80	~150	In progress	ongoing



Pesticides , antimicrobials, food additives, green alternatives, HPV, MPV, endocrine reference cmpds, tox reference cmpds, NTP in vivo, FDA GRAS, FDA PAFA, EDSP, water contaminants, exposure data, industrial, failed drugs, marketed drugs, fragrances, flame retardants, etc.



High Throughput Measurement to Identify **Exposure**

EPA United States Environmental Protection Agency

EPA US Environmental Protection Agency

Español | 中文: 繁體版 | 中文: 简体版 | Tiếng Việt | 한국어

Learn the Issues | Science & Technology | Laws & Regulations | About EPA

Search EPA.gov

Related Topics: Safer Chemicals Research

Contact Us Share

Rapid Chemical Exposure and Dose Research

EPA is responsible for ensuring the safety of thousands of chemicals. Quantitative exposure data are available for only a small fraction of registered chemicals. This type of exposure data is needed to thoroughly evaluate chemicals for potential risks to humans, wildlife and ecosystems. EPA is developing innovative methods to develop exposure estimates for thousands of chemicals to better protect human health and the environment. These innovative methods are called rapid exposure and dose assessments.

Rapid Exposure Predictions

Rapid, also called high-throughput, exposure predictions or ExpoCast provide rapid exposure estimates for thousands of chemicals. ExpoCast quickly and efficiently looks at multiple routes of exposure to provide exposure estimates. ExpoCast uses and enhances two well-known exposure models to estimate chemical exposure.

Pictured Above: Farfield Exposure Examples

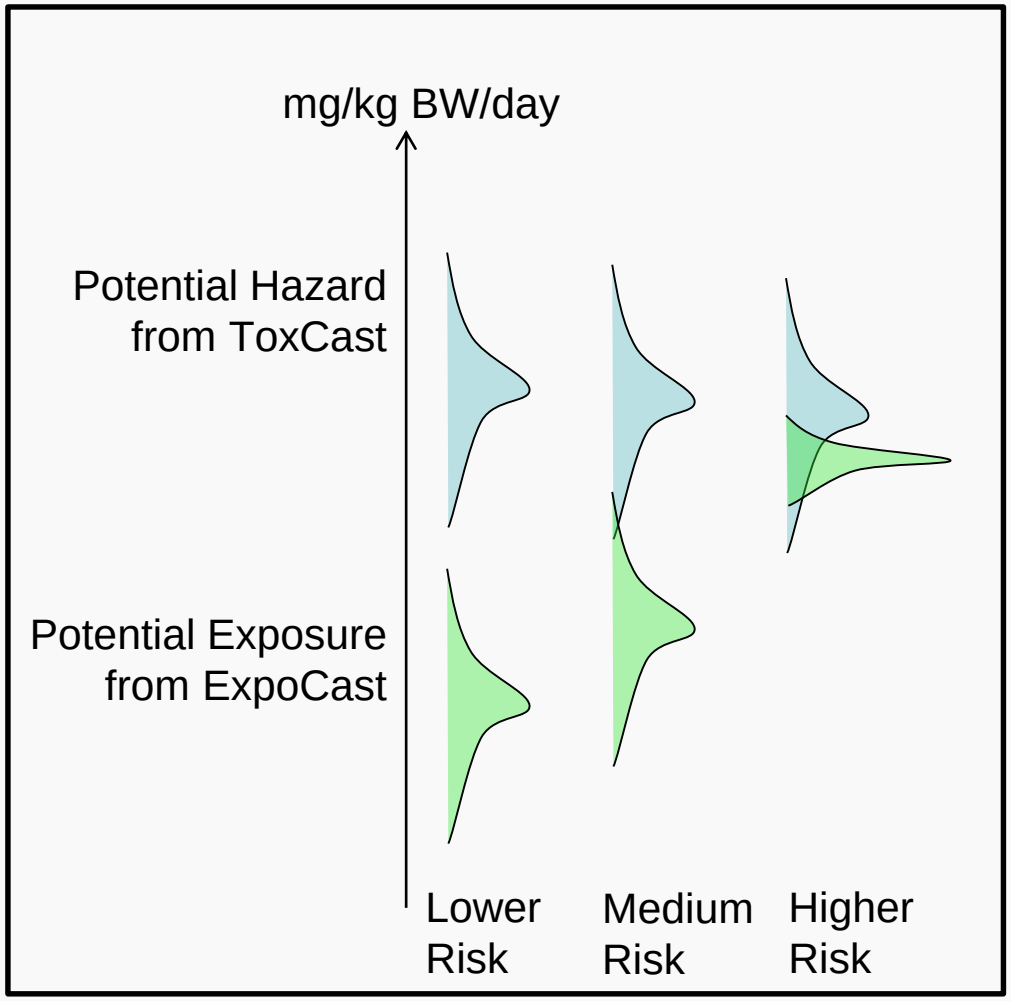
› Farfield Exposure Models

› Nearfield Exposure Models

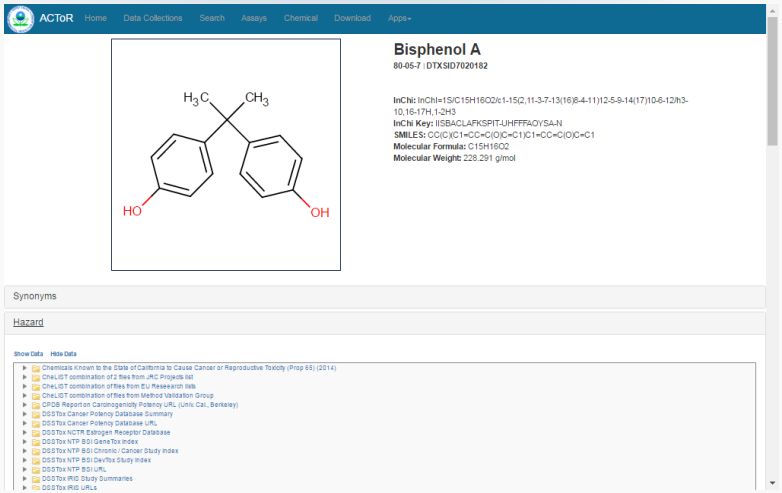
Evaluating High-throughput Exposure Predictions

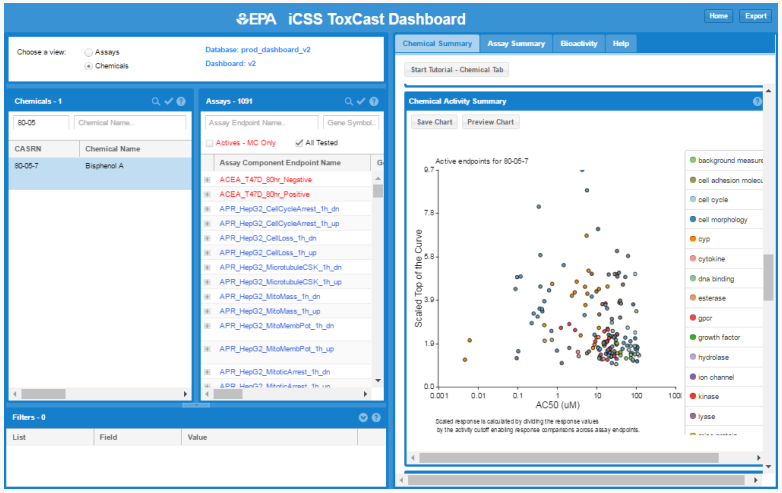
EPA is currently evaluating the effectiveness of high-throughput exposure models

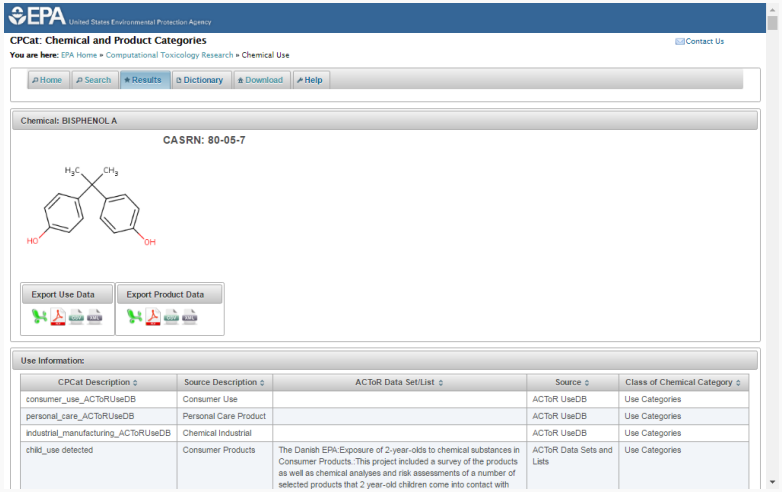
HT **Hazard and Exposure Combined** for Risk Assessment

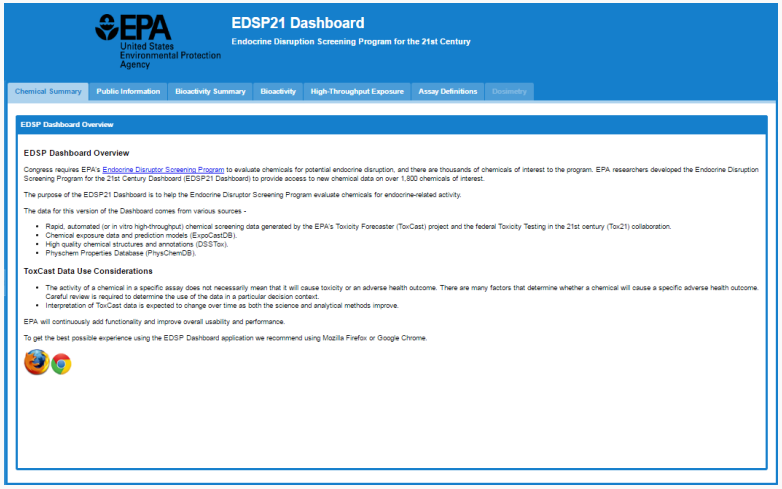


Delivering Data via Dashboards



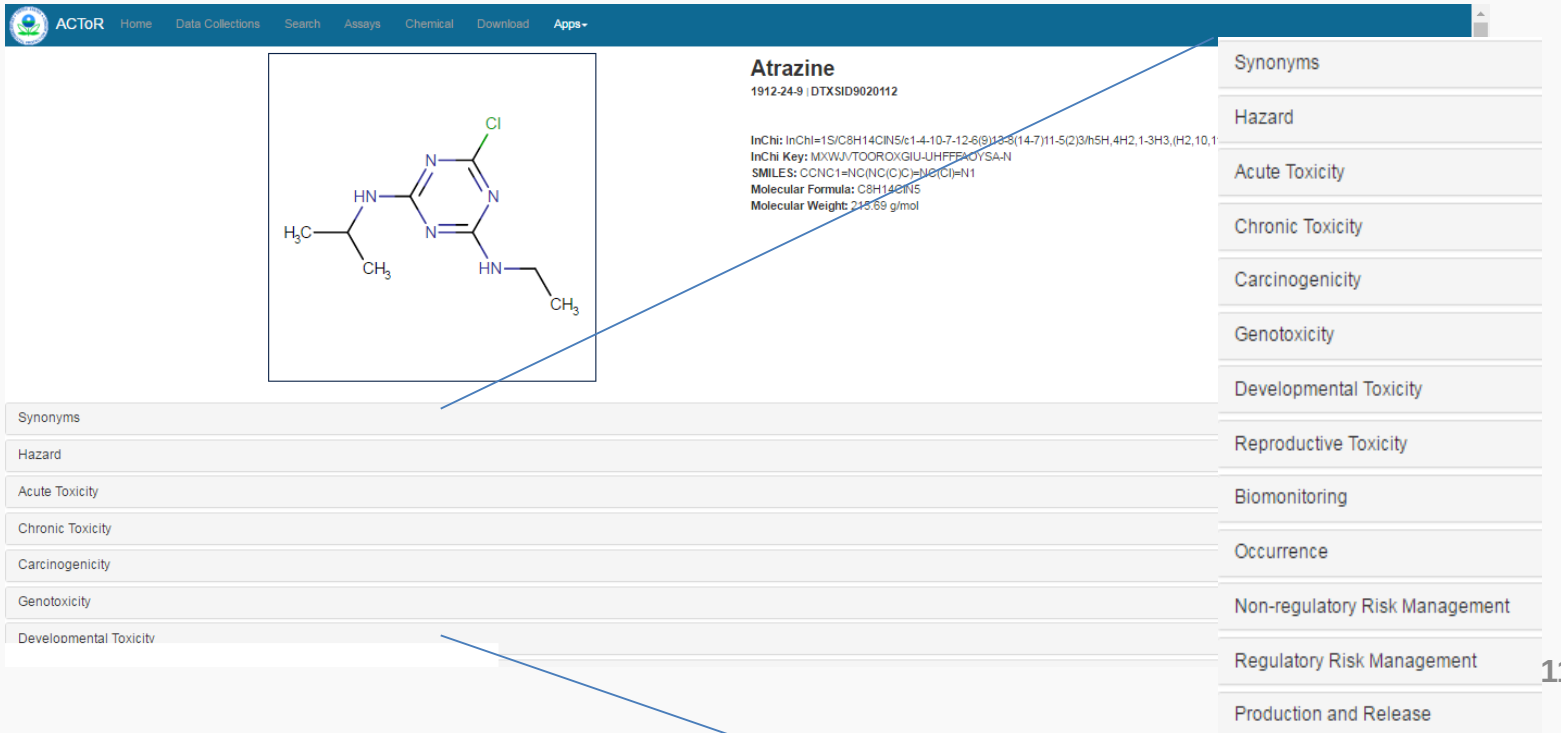






ACToR
https://actor.epa.gov/actor/

- Aggregated Computational Toxicology Resource
- Warehouse of available public toxicity data from >1000 public sources for >500,000 chemicals



ACToR
https://actor.epa.gov/actor/

EPA United States Environmental Protection Agency

Genotoxicity

Show Data Hide Data

- CheLIST combination of files from Method Validation Group
- DSSTox NTP BSI GeneTox Index
- DSSTox NTP BSI URL
- Genotoxicity Data In Vitro from EPA HPVIS
- NLM TOXNET CCRIS Data MSTU - MUTAGENICITY STUDIES

Genotoxicity Data In Vitro from EPA HPVIS

Result Group:

Component Name	Value
Conclusion	Negative
Conclusion (GeneTox)	This study was judged to have given a negative result since the observed increase in the first experiment was not reproducible.
Dose Remarks	30 - 50 g/mL with activation20 - 40 g/mL without activation
Genotoxic Effect	Negative
GLP	No Data
Metabolic Activation	With and Without
Method/Guideline Followed	No Data
Program Flag	HPVIS
Reliability	2
Results Remarks	With metabolic activation: 50 g/mLWithout metabolic activation: None reported In the first test with metabolic activation an increase in the percentage of metaphases with chromosome aberrations from bisphenol A treated cultures was observed only at the top dose in the presence of cytotoxicity; 14% at 50 g/mL compared to 3% in controls. In these high-dose cultures it was stated that cell confluence was reduced by approximately 70%. In the second test no significant increases were observed in with metabolic activation; only 3% of cells at the highest dose had aberrations. No significant increases in aberrations were observed without metabolic activation with bisphenol A evidently being tested up to "toxic levels." The positive controls produced clear increases in chromosome aberrations.
Species	Mammalian Cell Line
Sponsor Name	General Electric Company - Plastics
Sponsored Chemical Result Type	72162-28-8
Strain	Chinese Hamster Ovary (CHO)

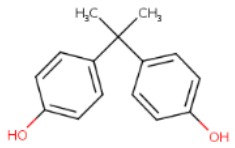
Ivett, J.L., Brown, B.M., Rodgers, C., Anderson, B.E., Ressnick, M.A. and Zeiger, E. 1989. Chromosomal aberrations and sister chromatid exchange tests in Chinese hamster ovary . Environ. Mol. Mutagen. 14:165-187. As cited in the EU Risk Assessment. Data entered into the HPVIS extracted from the robust

CPCat: Chemical and Product Categories https://actor.epa.gov/cpcat/

EPA United States Environmental Protection Agency

Chemical: BISPHENOL A

CASRN: 80-05-7



Export Use Data | Export Product Data

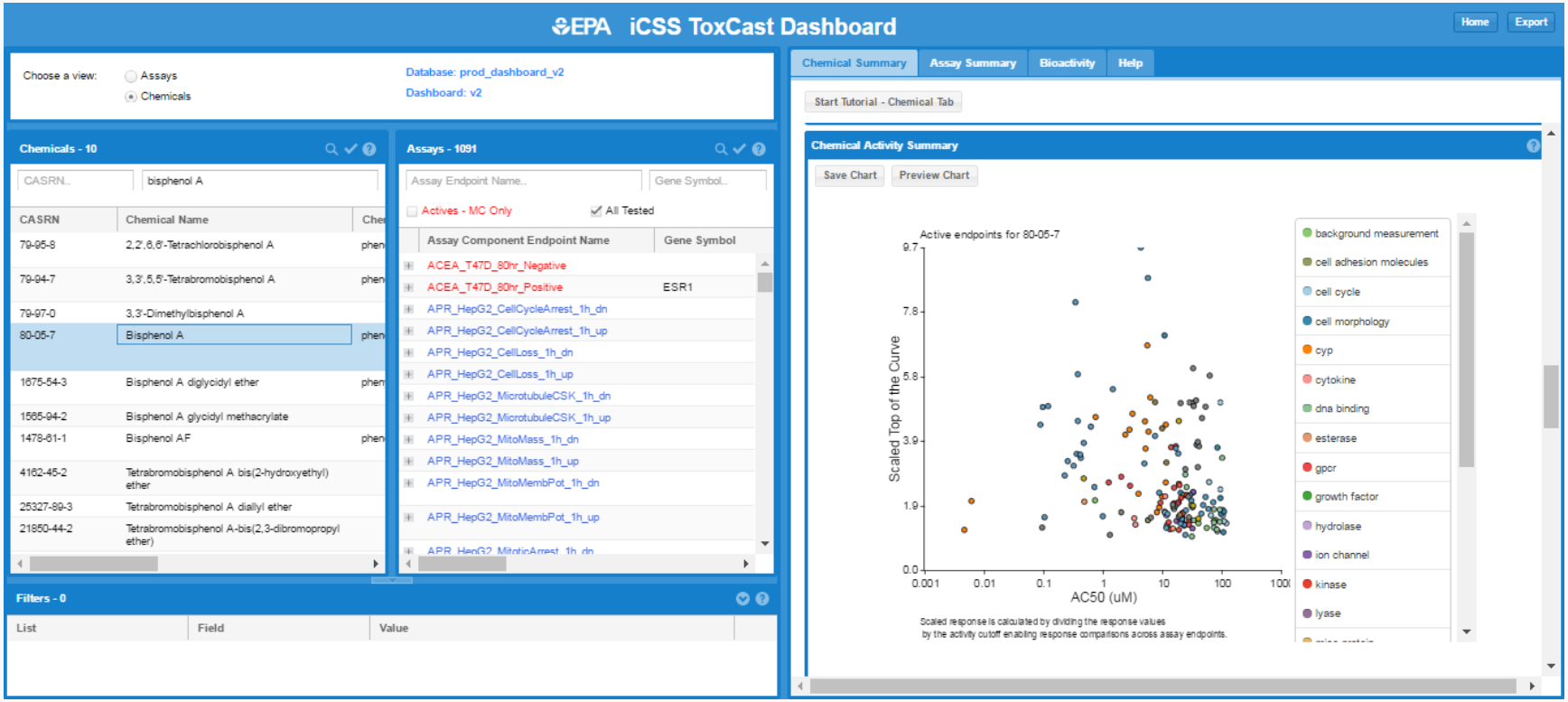
Chemical and Product Categories is a database containing information mapping **>43,000 chemicals** to a set of terms categorizing their usage or function.

Use Information:

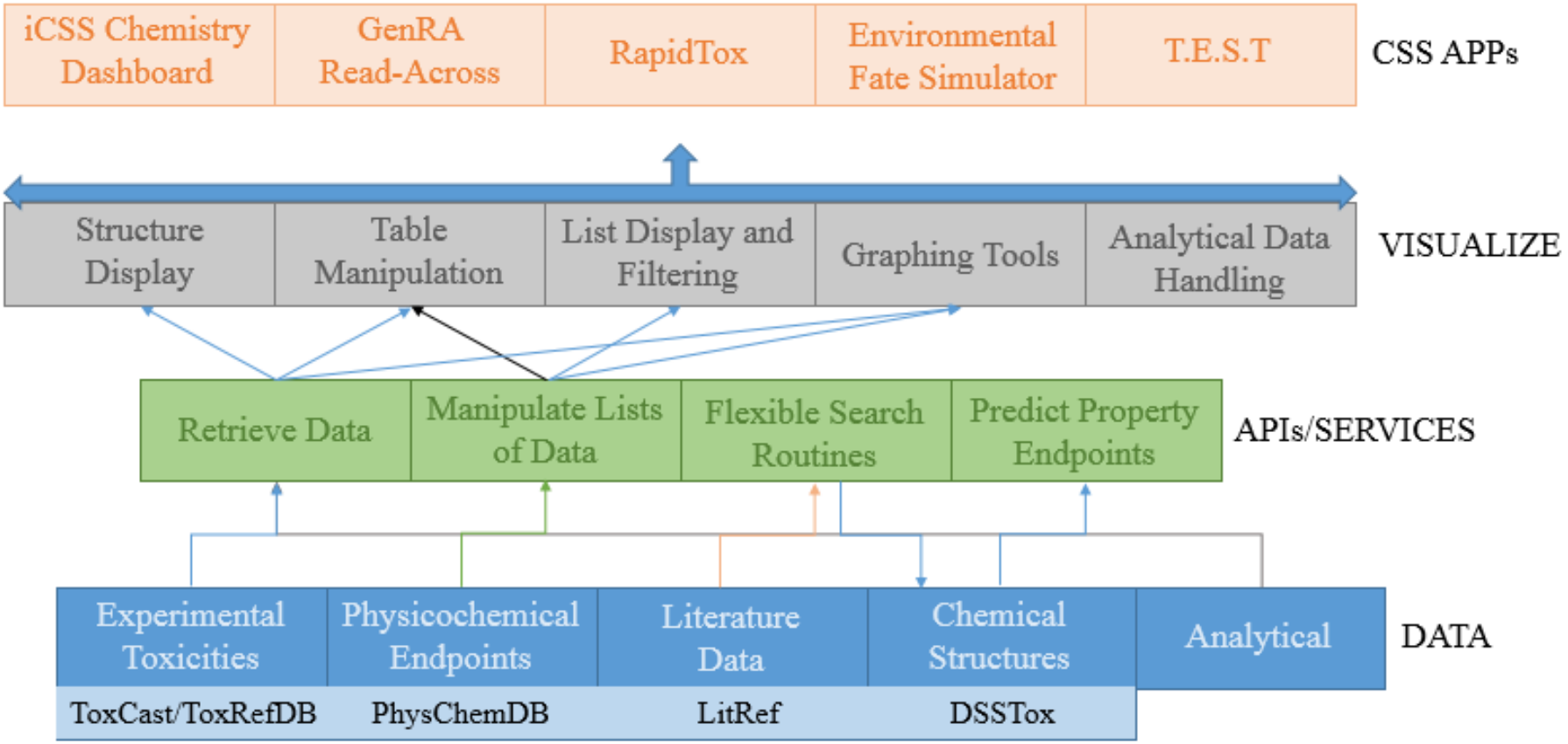
CPCat Description	Source Description	ACToR Data Set/List	Source	Class of Chemical Category
consumer_use_ACToRUseDB	Consumer Use		ACToR UseDB	Use Categories
personal_care_ACToRUseDB	Personal Care Product		ACToR UseDB	Use Categories
industrial_manufacturing_ACToRUseDB	Chemical Industrial		ACToR UseDB	Use Categories
child_use detected	Consumer Products	The Danish EPA:Exposure of 2-year-olds to chemical substances in Consumer Products.:This project included a survey of the products as well as chemical analyses and risk assessments of a number of selected products that 2 year-old children come into contact with throughout the course of a day. A total of 12 product groups were included in the survey phase. Selected products from 10 of these product groups were subsequently included in a screening phase and several problematic substances were subjected to quantitative analysis. A risk assessment was also performed for a number of problematic substances.	ACToR Data Sets and Lists	Use Categories
consumer_use detected	Consumer Products	TNO Nederlands Organisation for Applied Scientific Research:Hazardous Chemicals in Consumer Products:In this study 33 consumer products, including body care products, toys, textiles, deodorizers and cleaners, have been tested for the presence of bisphenol-A, alkylphenols and ethoxylates, phthalates, musks and organotin compounds.	ACToR Data Sets and Lists	Use Categories
child_use	Consumer Products:	(state of) Washington:Washington Childrens Safe Product Act (CSPA) Reporting Rule. List of chemical of high concern to children (CHCC)	ACToR Data Sets and	Use Categories

Toxcast Dashboard
https://actor.epa.gov/dashboard/

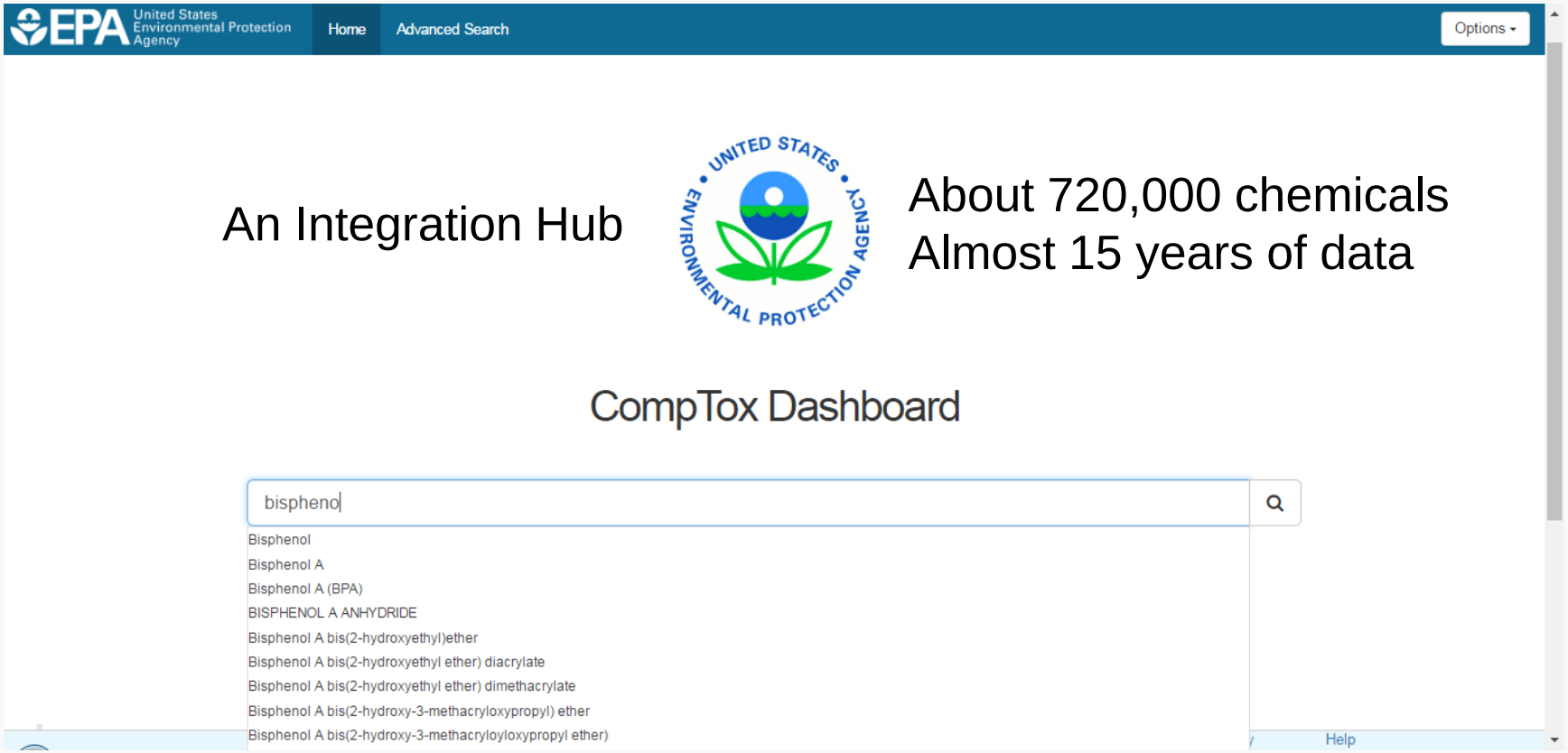
- Access and Interrogate chemical screening data from ToxCast and the Tox21 collaboration



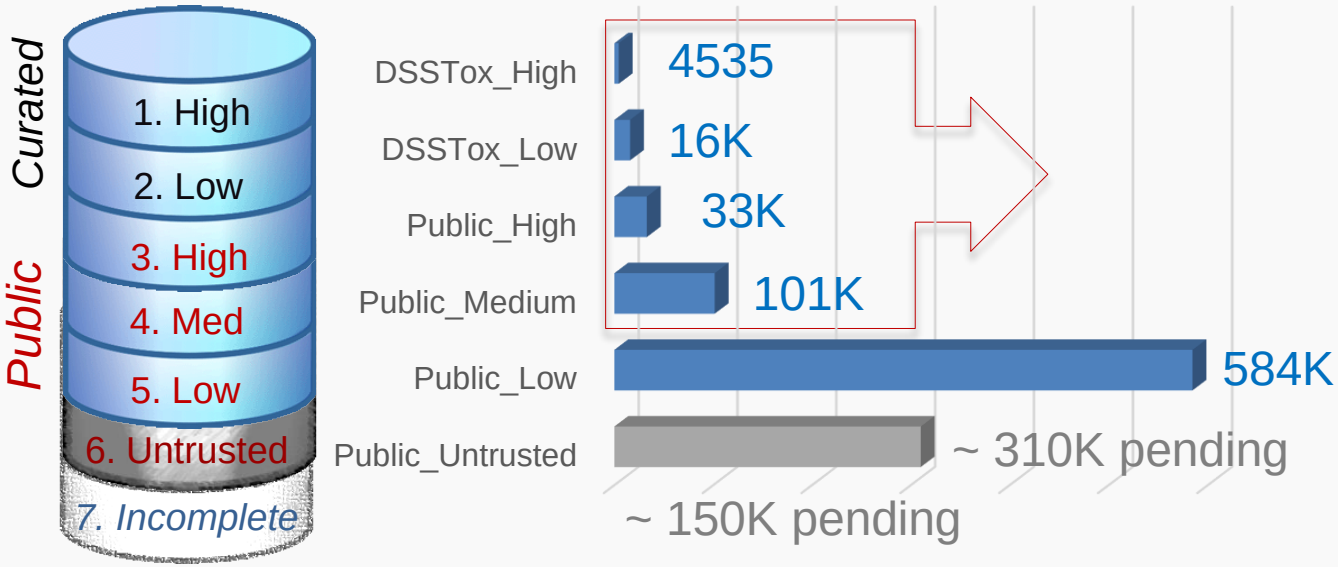
Our New Developing Architecture



Our **Latest** Dashboard
https://comptox.epa.gov



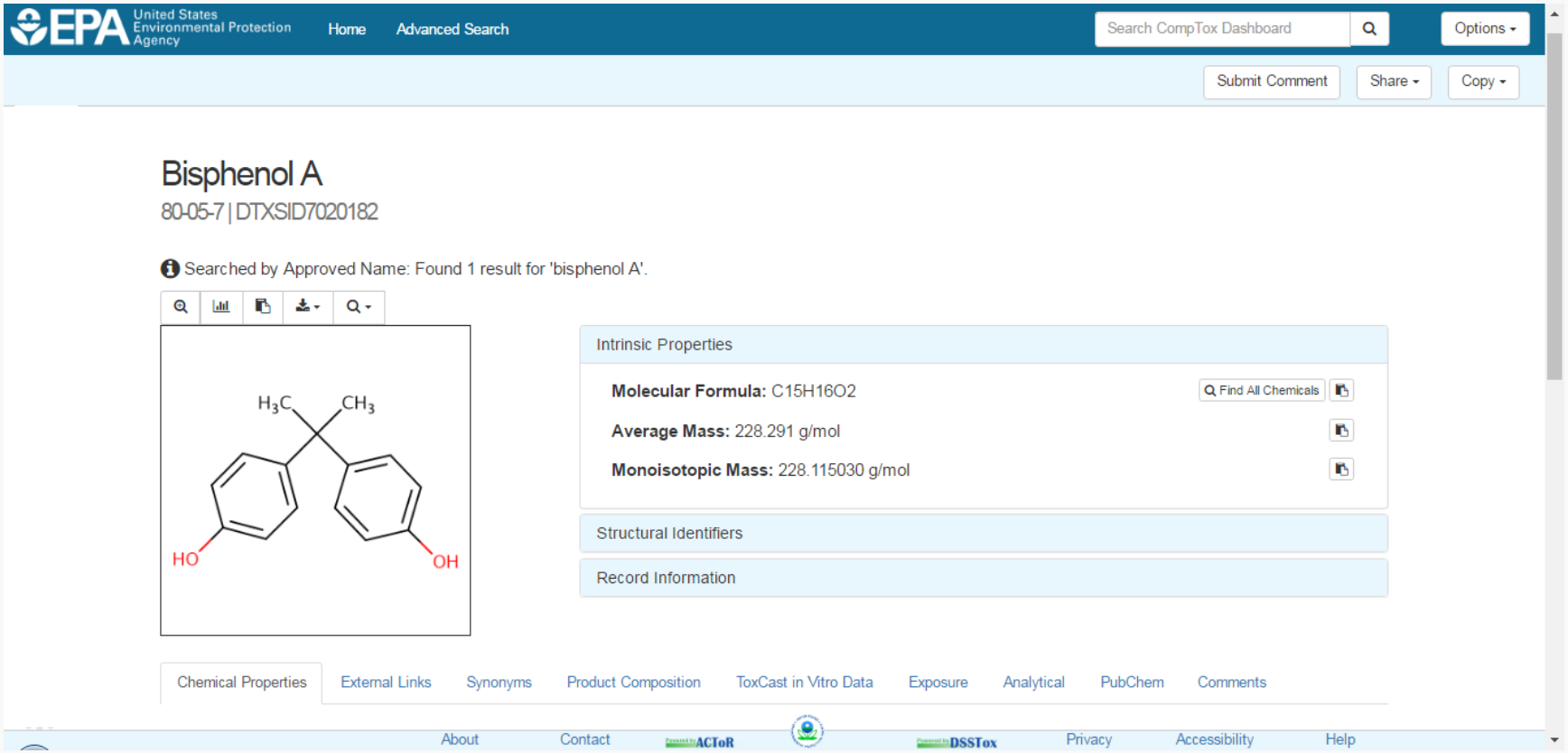
Approximately 15 Years of Data… Continually changing



QC Levels

DSSTox_High:	Hand curated and validated
DSSTox_Low:	Hand curated and confirmed using multiple public sources
Public_High:	Extracted from EPA SRS and confirmed to have no conflicts in ChemID and PubChem
Public_Medium:	Extracted from ChemID and confirmed to have no conflicts in PubChem
Public_Low:	Extracted from ACToR or PubChem
Public_Untrusted:	Postulated, but found to have conflicts in public sources

Bisphenol A (Accessing DSSTox Data)



Chemical Properties External Links Synonyms Product Composition ToxCast in Vitro Data Exposure PubChem Comments

Physicochemical Properties

EPA United States Environmental Protection Agency

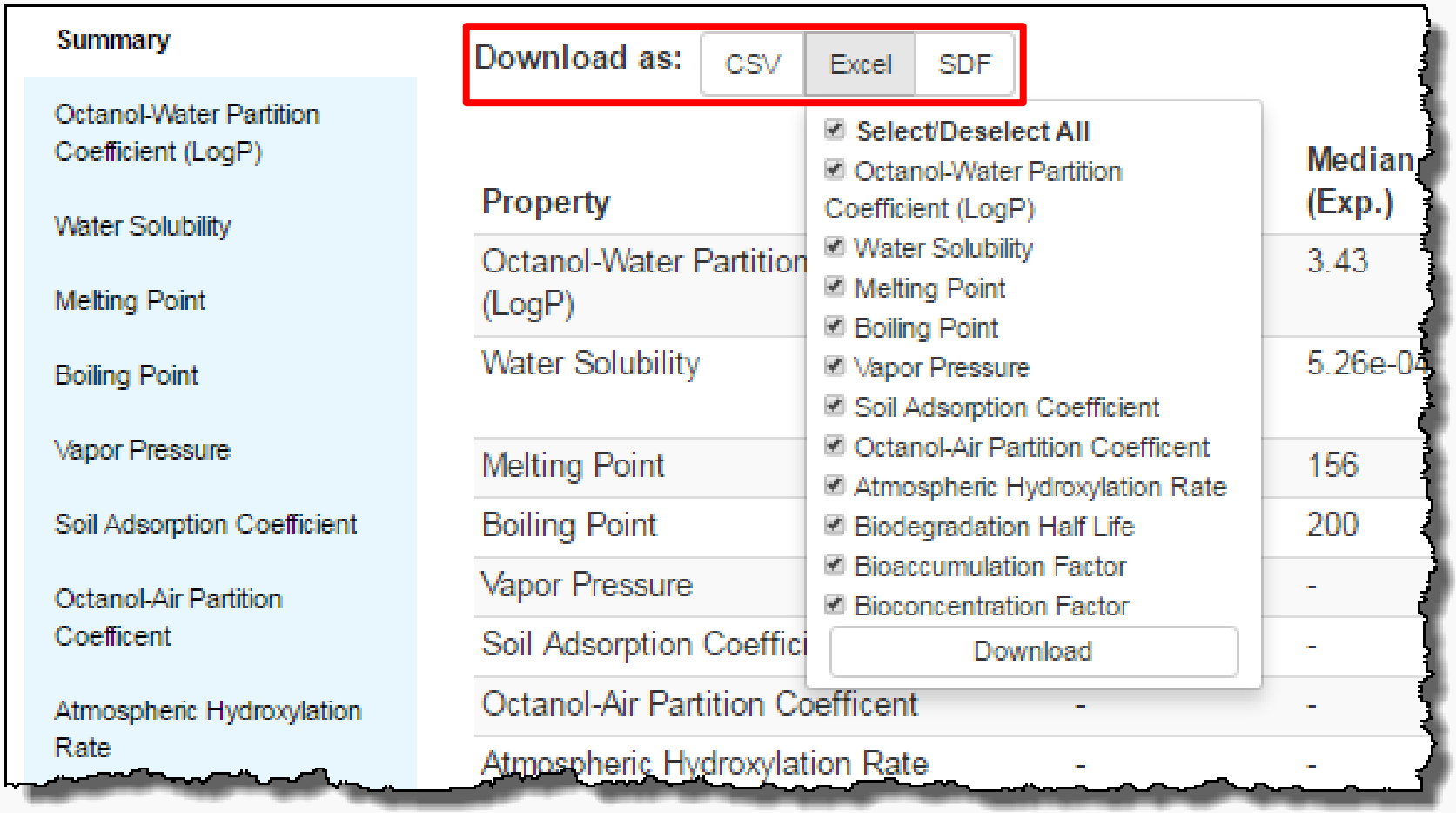
Summary

- Octanol-Water Partition Coefficient (LogP)
- Water Solubility
- Melting Point
- Boiling Point
- Vapor Pressure
- Soil Adsorption Coefficient
- Octanol-Air Partition Coefficent
- Atmospheric Hydroxylation Rate
- Biodegradation Half Life
- Bioaccumulation

Download as: CSV | Excel | SDF

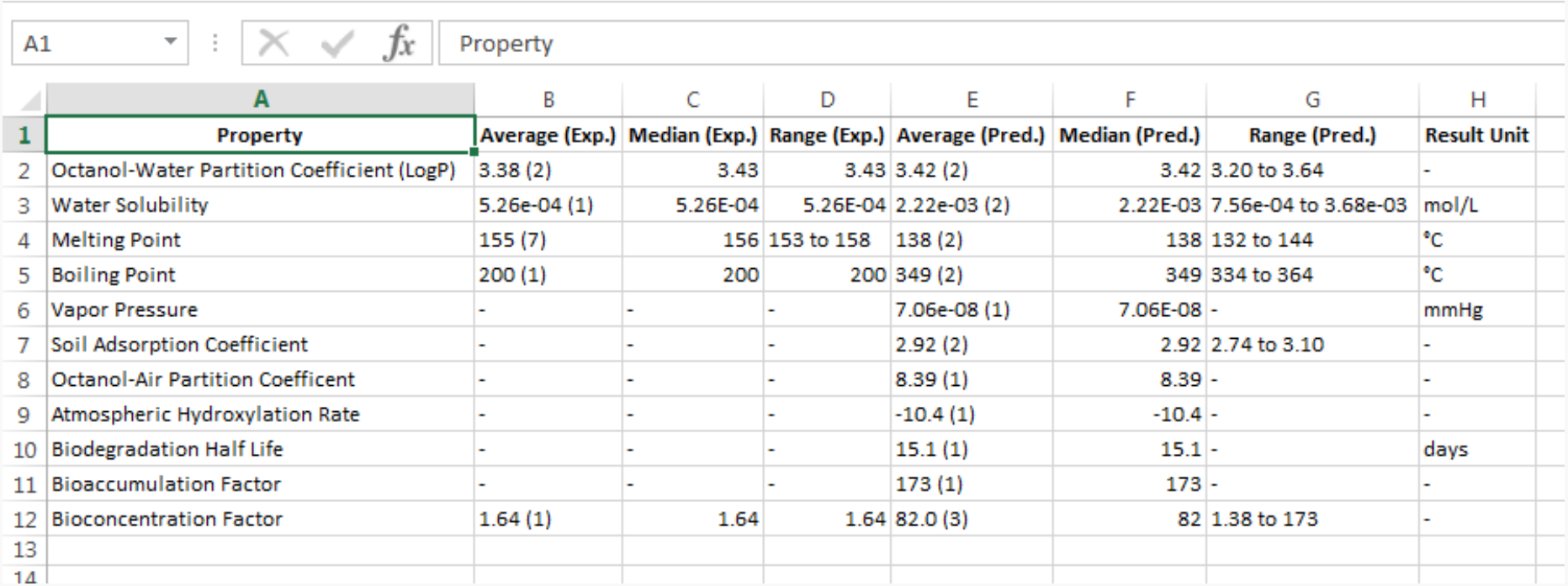
Property	Average (Exp.)	Median (Exp.)	Range (Exp.)	Average (Pred.)	Median (Pred.)	Range (Pred.)	Result Unit
Octanol-Water Partition Coefficient (LogP)	3.38 (2)	3.43	3.43	3.42 (2)	3.42	3.20 to 3.64	-
Water Solubility	5.26e-04 (1)	5.26e-04	5.26e-04	2.22e-03 (2)	2.22e-03	7.56e-04 to 3.68e-03	mol/L
Melting Point	155 (7)	156	153 to 158	138 (2)	138	132 to 144	°C
Boiling Point	200 (1)	200	200	349 (2)	349	334 to 364	°C
Vapor Pressure	-	-	-	7.06e-08 (1)	7.06e-08	-	mmHg
Soil Adsorption Coefficient	-	-	-	2.92 (2)	2.92	2.74 to 3.10	-
Octanol-Air Partition Coefficent	-	-	-	8.39 (1)	8.39	-	-
Atmospheric Hydroxylation Rate	-	-	-	-10.4 (1)	-10.4	-	-
Biodegradation Half Life	-	-	-	15.1 (1)	15.1	-	days
Bioaccumulation Factor	-	-	-	173 (1)	173	-	-
Bioconcentration Factor	1.64 (1)	1.64	1.64	82.0 (3)	82.0	1.38 to 173	-

Data Downloads



Chemical Properties | External Links | Synonyms | Product Composition | ToxCast in Vitro Data | Exposure | PubChem | Comments

Data Download: Excel



A1 | Property

	A	B	C	D	E	F	G	H
1	**Property**	**Average (Exp.)**	**Median (Exp.)**	**Range (Exp.)**	**Average (Pred.)**	**Median (Pred.)**	**Range (Pred.)**	**Result Unit**
2	Octanol-Water Partition Coefficient (LogP)	3.38 (2)	3.43	3.43	3.42 (2)	3.42	3.20 to 3.64	-
3	Water Solubility	5.26e-04 (1)	5.26E-04	5.26E-04	2.22e-03 (2)	2.22E-03	7.56e-04 to 3.68e-03	mol/L
4	Melting Point	155 (7)	156	153 to 158	138 (2)	138	132 to 144	°C
5	Boiling Point	200 (1)	200	200	349 (2)	349	334 to 364	°C
6	Vapor Pressure	-	-	-	7.06e-08 (1)	7.06E-08	-	mmHg
7	Soil Adsorption Coefficient	-	-	-	2.92 (2)	2.92	2.74 to 3.10	-
8	Octanol-Air Partition Coefficent	-	-	-	8.39 (1)	8.39	-	-
9	Atmospheric Hydroxylation Rate	-	-	-	-10.4 (1)	-10.4	-	-
10	Biodegradation Half Life	-	-	-	15.1 (1)	15.1	-	days
11	Bioaccumulation Factor	-	-	-	173 (1)	173	-	-
12	Bioconcentration Factor	1.64 (1)	1.64	1.64	82.0 (3)	82	1.38 to 173	-
13								
14								

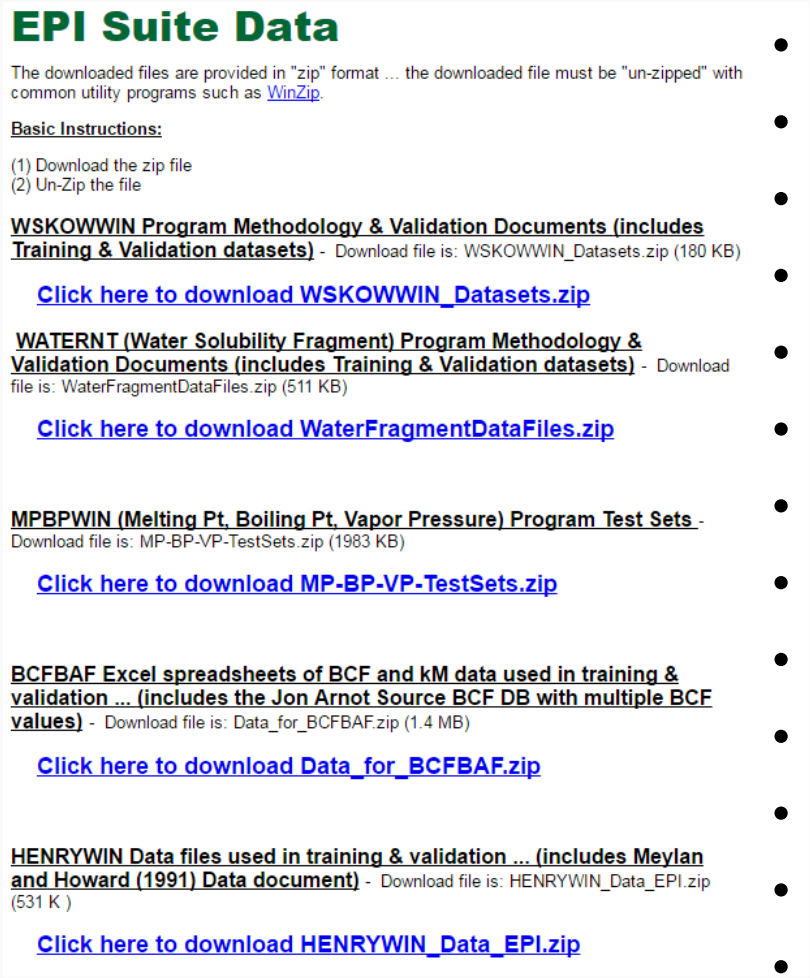
Developing “NCCT Models”

- Physchem properties for exposure modeling, augmented with ToxCast HTS *in vitro* data etc.

- Our approach to modeling:
  - Obtain **high quality** training sets
  - Apply appropriate modeling approaches
  - Validate performance of models
  - Define the applicability domain and limitations of the models
  - Use models to predict properties across our full datasets

PHYSPROP Data: Available from:

http://esc.syrres.com/interkow/EpiSuiteData.htm



EPI Suite Data

The downloaded files are provided in "zip" format ... the downloaded file must be "un-zipped" with common utility programs such as WinZip.

Basic Instructions:

(1) Download the zip file
(2) Un-Zip the file

WSKOWWIN Program Methodology & Validation Documents (includes Training & Validation datasets) - Download file is: WSKOWWIN_Datasets.zip (180 KB)

Click here to download WSKOWWIN_Datasets.zip

WATERNT (Water Solubility Fragment) Program Methodology & Validation Documents (includes Training & Validation datasets) - Download file is: WaterFragmentDataFiles.zip (511 KB)

Click here to download WaterFragmentDataFiles.zip

MPBPWIN (Melting Pt, Boiling Pt, Vapor Pressure) Program Test Sets - Download file is: MP-BP-VP-TestSets.zip (1983 KB)

Click here to download MP-BP-VP-TestSets.zip

BCFBAF Excel spreadsheets of BCF and kM data used in training & validation ... (includes the Jon Arnot Source BCF DB with multiple BCF values) - Download file is: Data_for_BCFBAF.zip (1.4 MB)

Click here to download Data_for_BCFBAF.zip

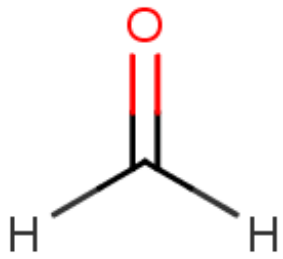
HENRYWIN Data files used in training & validation ... (includes Meylan and Howard (1991) Data document) - Download file is: HENRYWIN_Data_EPI.zip (531 K)

Click here to download HENRYWIN_Data_EPI.zip

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

Check and Curate Public Data

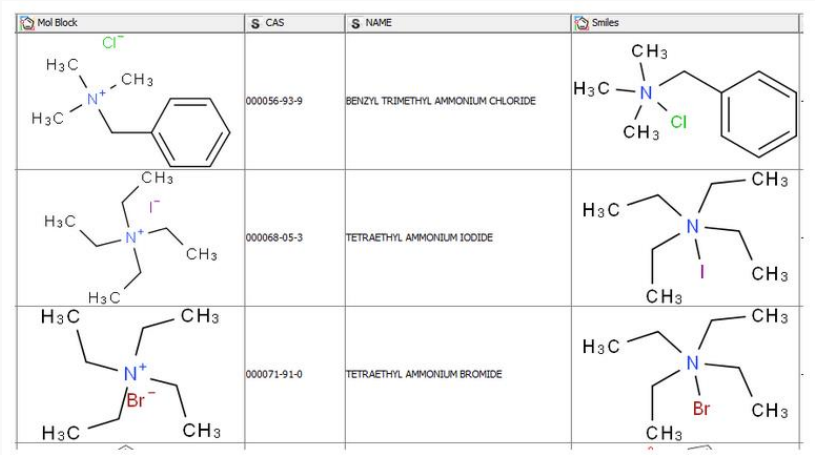
- Public data should be curated prior to modeling.
- The data files have **FOUR** representations of a chemical, plus the property value.

SDF Molecule	Mol Mol Block	S Smiles	S CAS	S NAME	D Kow
-ISIS- 09141018452D 4 3 0 0 0 0 0 0 0 0999 V2000 2.4667 -0.0833 0.0000 O 0 0 ... 2.4667 -0.9125 0.0000 C 0 0 ... 1.7500 -1.3292 0.0000 H 0 0 ... 3.1833 -1.3292 0.0000 H 0 0 ... 2 1 2 0 0 0 0 3 2 1 0 0 0 0 4 2 1 0 0 0 0 M END > <CAS> (000050-00-0) 000050-00-0 > <NAME> (000050-00-0) FORMALDEHYDE > <Kow> (000050-00-0) 3.500000000000000e-001		O=C	000050-00-0	FORMALDEHYDE	0.35

Check and Curate Public Data

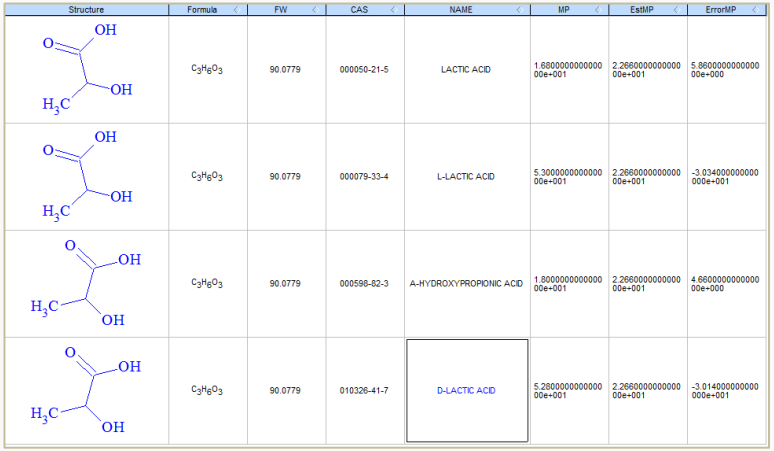
EPA United States Environmental Protection Agency

Public data should be curated prior to modeling



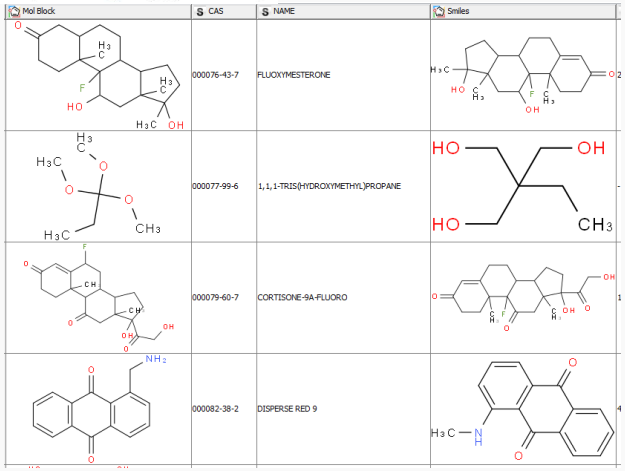
Mol Block	CAS	NAME	Smiles
	000056-93-9	BENZYL TRIMETHYL AMMONIUM CHLORIDE	
	000068-05-3	TETRAETHYL AMMONIUM IODIDE	
	000071-91-0	TETRAETHYL AMMONIUM BROMIDE	

Covalent Halogens



Structure	Formula	FW	CAS	NAME	MP	EstMP	ErrorMP
	$C_3H_6O_3$	90.0779	000050-21-5	LACTIC ACID	1.6800000000000e+001	2.2660000000000e+001	5.8600000000000e+000
	$C_3H_6O_3$	90.0779	000079-33-4	L-LACTIC ACID	5.3000000000000e+001	2.2660000000000e+001	-3.0340000000000e+001
	$C_3H_6O_3$	90.0779	000598-82-3	A-HYDROXYPROPIONIC ACID	1.8000000000000e+001	2.2660000000000e+001	4.6600000000000e+000
	$C_3H_6O_3$	90.0779	010326-41-7	D-LACTIC ACID	5.2800000000000e+001	2.2660000000000e+001	-3.0140000000000e+001

Identical Chemicals



Mol Block	CAS	NAME	Smiles
	000076-43-7	FLUOXYMESTERONE	
	000077-99-6	1,1,1-TRIS(HYDROXYMETHYL)PROPANE	
	000079-60-7	CORTISONE-9A-FLUORO	
	000082-38-2	DISPERSE RED 9	

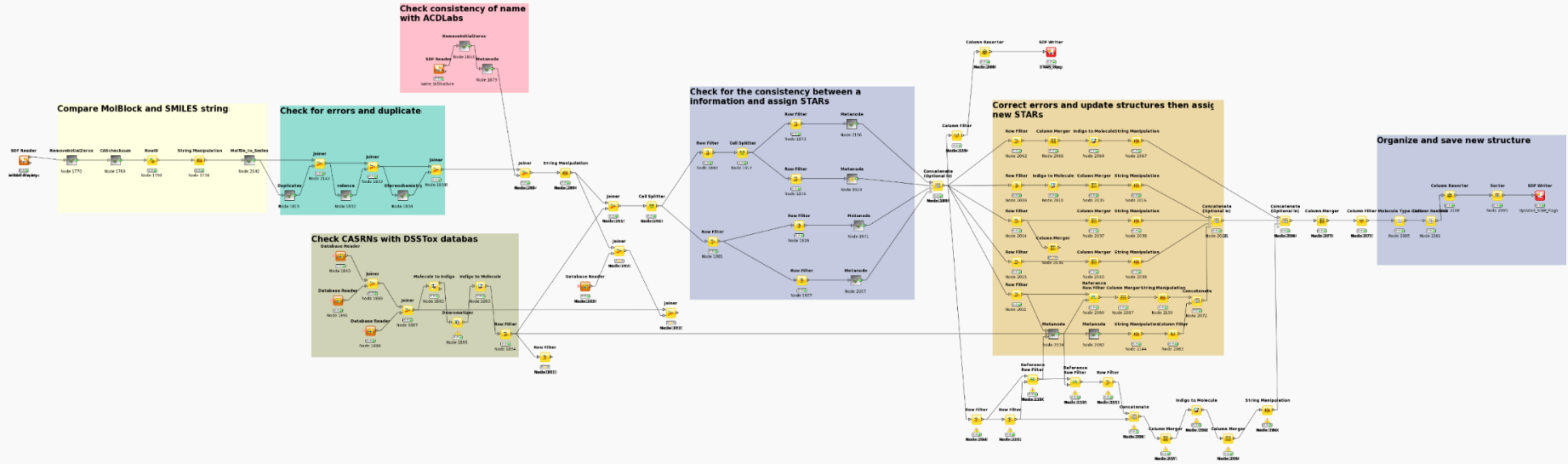
Mismatches

The Approach

- Our curation process
  - Choose the “chemical” by checking levels of consistency
  - Perform initial analysis manually to understand how to clean the data (chemical structure and ID)
  - Automate the process (and test iteratively)
  - Process all datasets using final method

  - We did **NOT** validate each measured **property value**

KNIME Workflow to Evaluate the Dataset



LogP dataset: 15,809 structures

- CAS Checksum: 12163 valid, 3646 **invalid** (>23%)
- **Invalid** names: 555
- **Invalid** SMILES 133
- Valence **errors**: 322 Molfile, 3782 SMILES (>24%)
- Duplicates check:
  - 31 DUPLICATE MOLFILES
  - 626 DUPLICATE SMILES
  - 531 DUPLICATE NAMES
- SMILES vs. Molfiles (structure check)
  - 1279 differ in stereochemistry (~8%)
  - 362 “Covalent Halogens”
  - 191 differ as tautomers
  - 436 are different compounds (~3%)

Curation to “QSAR Ready Files”

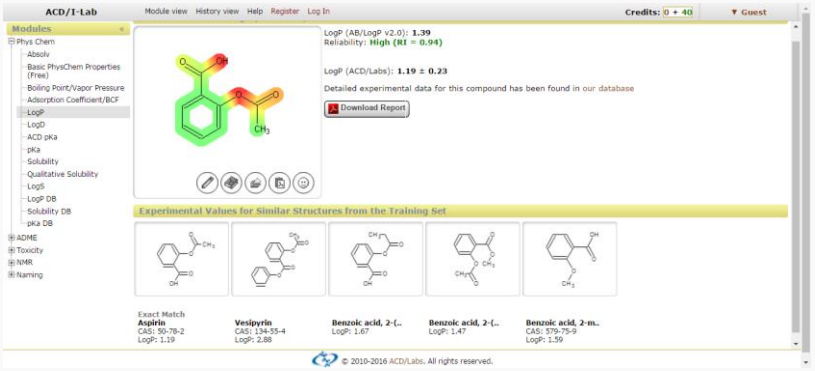
- “QSAR-Ready Structures”
  - Desalt/Neutralize, Desolvate, Remove stereochemistry

Property	Initial file	Curated Data	Curated QSAR ready
AOP	818	818	745
BCF	685	618	608
BioHC	175	151	150
Biowin	1265	1196	1171
BP	5890	5591	5436
HL	1829	1758	1711
KM	631	548	541
KOA	308	277	270
LogP	15809	~~14544~~	14041
MP	10051	9120	8656
PC	788	750	735
VP	3037	2840	2716
WF	5764	5076	4836
WS	2348	2046	2010

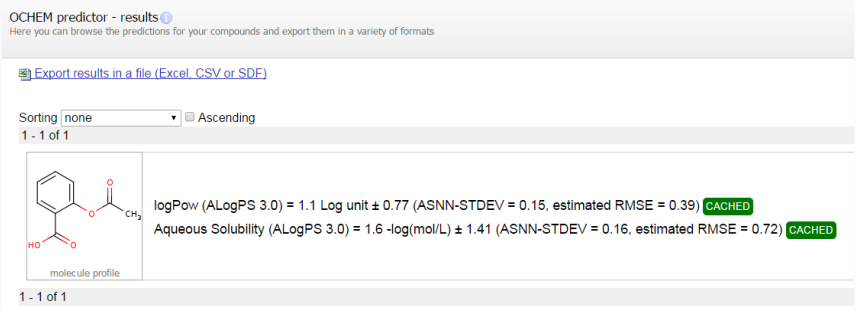
Communicating Transparency in Models to Users of an App

- Too often predicted values just give “numbers”
- Users have no real understanding of model performance
- There are good examples though! ACD/Ilab, T.E.S.T, OCHEM

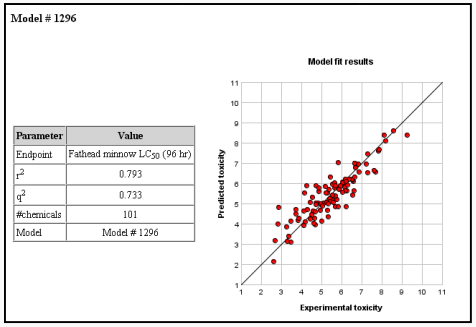
ACD/ILab



OCHEM



EPA T.E.S.T

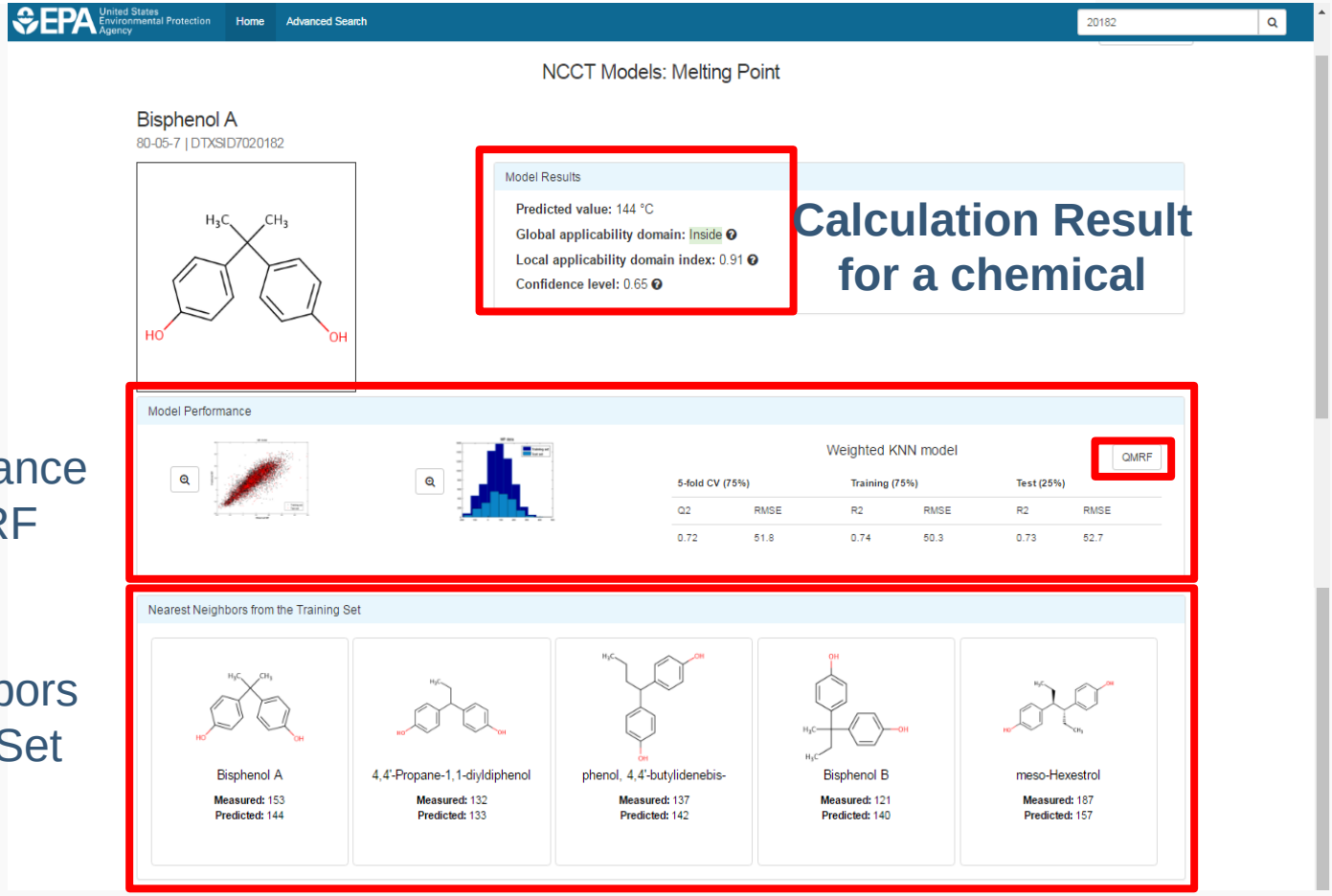


NCCT Models: What you would report in a paper…

Prop	Vars	5-fold CV (75%)		Training (75%)			Test (25%)		
		Q2	RMSE	N	R2	RMSE	N	R2	RMSE
BCF	10	0.84	0.55	465	0.85	0.53	161	0.83	0.64
BP	13	0.93	22.46	4077	0.93	22.06	1358	0.93	22.08
LogP	9	0.85	0.69	10531	0.86	0.67	3510	0.86	0.78
MP	15	0.72	51.8	6486	0.74	50.27	2167	0.73	52.72
VP	12	0.91	1.08	2034	0.91	1.08	679	0.92	1
WS	11	0.87	0.81	3158	0.87	0.82	1066	0.86	0.86
HL	9	0.84	1.96	441	0.84	1.91	150	0.85	1.82

Predicted Data

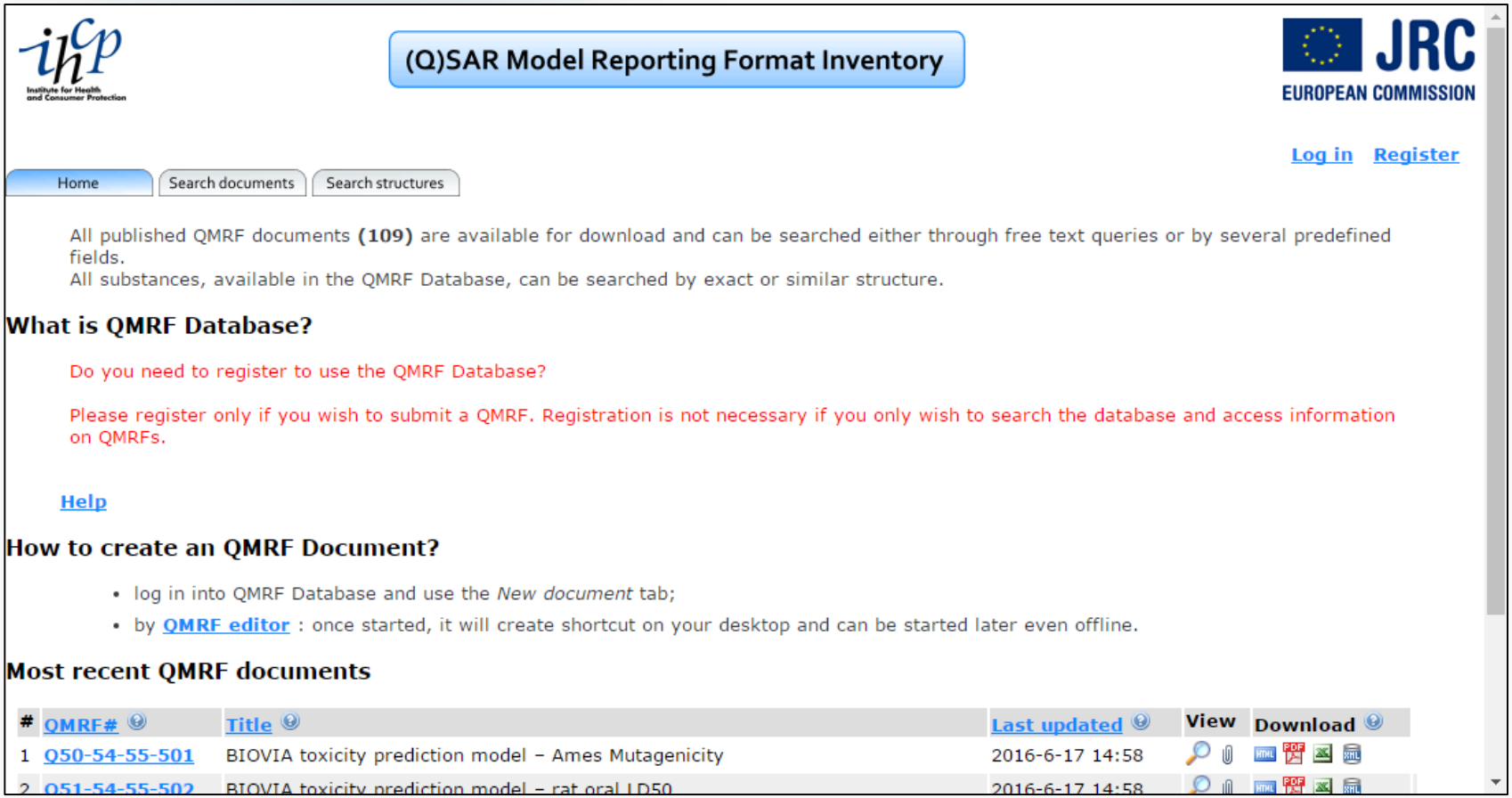
Predicted		
Source	**Result**	**Calculation Details**
EPISUITE	132 °C	Not Available
NCCT	144 °C	NCCT Model Report



Model Performance with full QMRF

Nearest Neighbors from Training Set

QMRF Reports



(Q)SAR Model Reporting Format Inventory

Log in Register

Home | Search documents | Search structures

All published QMRF documents **(109)** are available for download and can be searched either through free text queries or by several predefined fields.
All substances, available in the QMRF Database, can be searched by exact or similar structure.

What is QMRF Database?

Do you need to register to use the QMRF Database?

Please register only if you wish to submit a QMRF. Registration is not necessary if you only wish to search the database and access information on QMRFs.

Help

How to create an QMRF Document?

- log in into QMRF Database and use the *New document* tab;
- by QMRF editor : once started, it will create shortcut on your desktop and can be started later even offline.

Most recent QMRF documents

#	QMRF#	Title	Last updated	View	Download
1	Q50-54-55-501	BIOVIA toxicity prediction model – Ames Mutagenicity	2016-6-17 14:58		
2	Q51-54-55-502	BIOVIA toxicity prediction model – rat oral LD50	2016-6-17 14:58		

Prediction Details and QMRF Report

Model Results

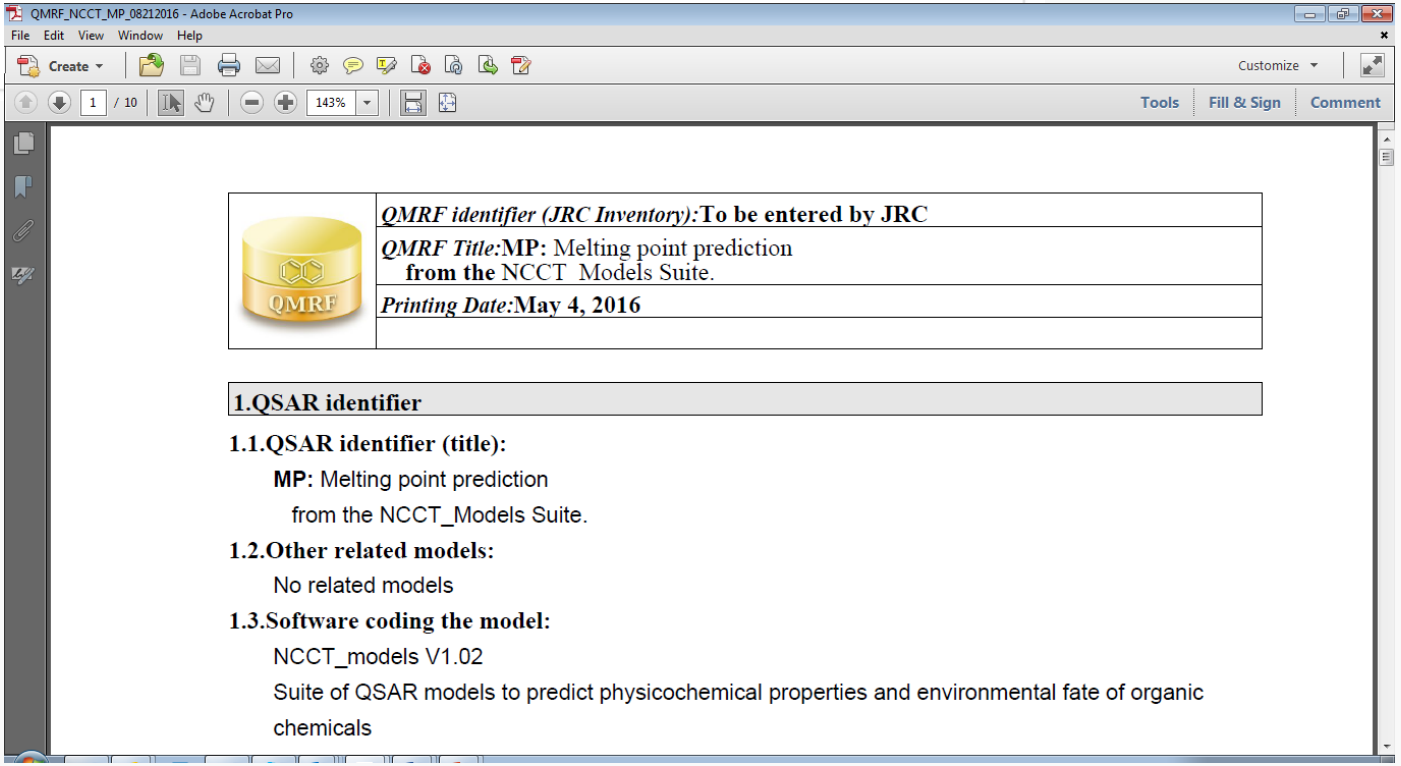
Predicted value: 144 °C

Global applicability domain: Inside

Applicability domain using the leverage approach. All training set space considered. More details in QMRF.

Local applicability domain index: 0.91

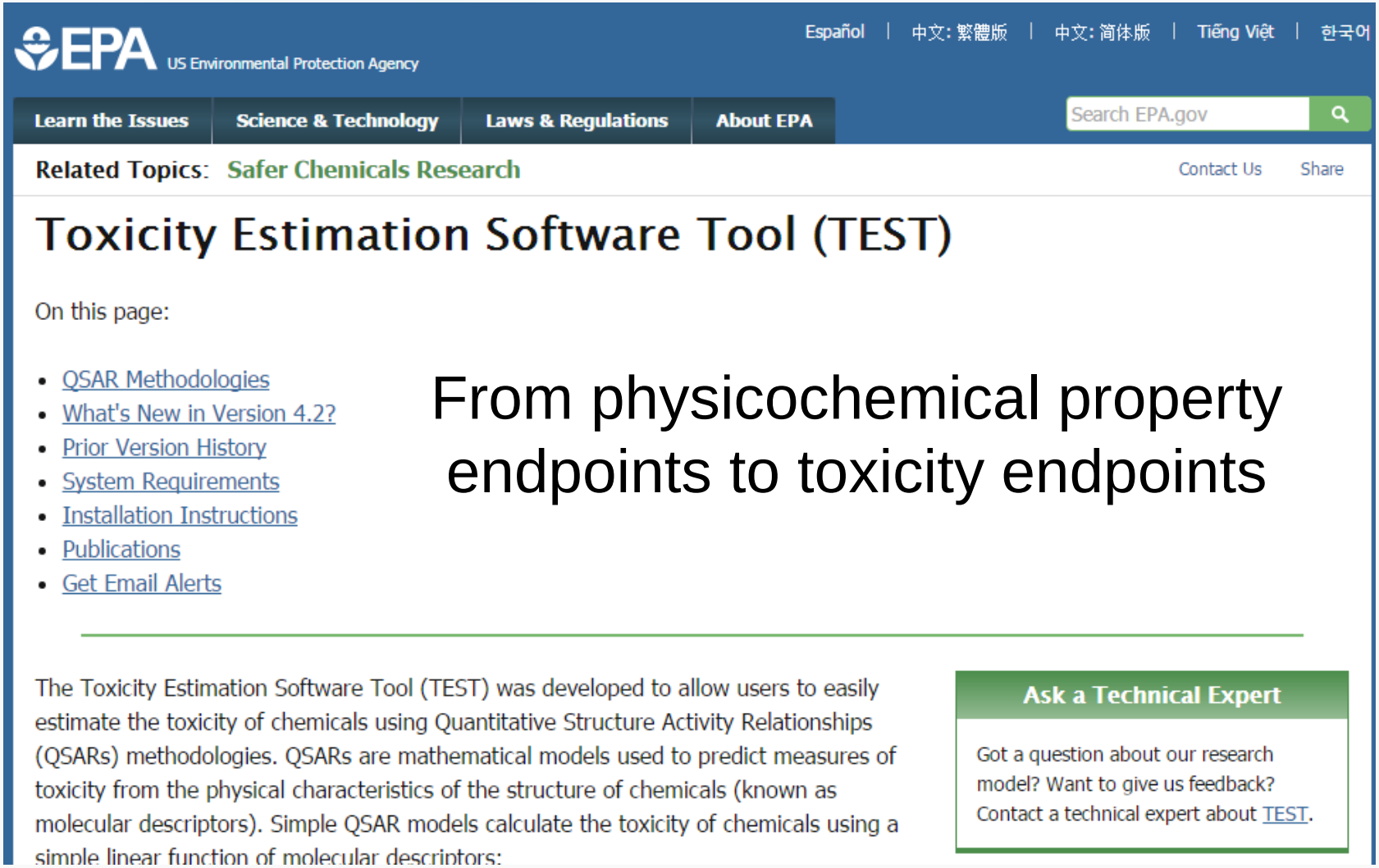
Confidence level: 0.65



- Accepted for publication to SAR and QSAR in Environmental Research

EPA T.E.S.T

https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test



Full transparency for each prediction

Toxicity prediction results for 333-41-5 for Hierarchical clustering method

Prediction results

Endpoint	Experimental value CAS: 333-41-5 Source: ECOTOX	Predicted value[a]	Prediction interval
Fathead minnow LC_{50} (96 hr) -Log(mol/L)	4.81	5.39	4.54 ≤ Tox ≤ 6.24
Fathead minnow LC_{50} (96 hr) mg/L	4.70	1.23	0.17 ≤ Tox ≤ 8.71

[a]Note: the test chemical was present in the external test set.

Cluster model predictions and statistics

Cluster model	Test chemical descriptor values	Prediction interval -Log(mol/L)	r^2	q^2	#chemicals
1296	Descriptors	6.010 ± 1.136	0.793	0.733	101
1300	Descriptors	5.458 ± 1.312	0.729	0.645	111
1301	Descriptors	5.136 ± 1.169	0.747	0.718	294
1302	Descriptors	4.922 ± 1.182	0.774	0.751	641

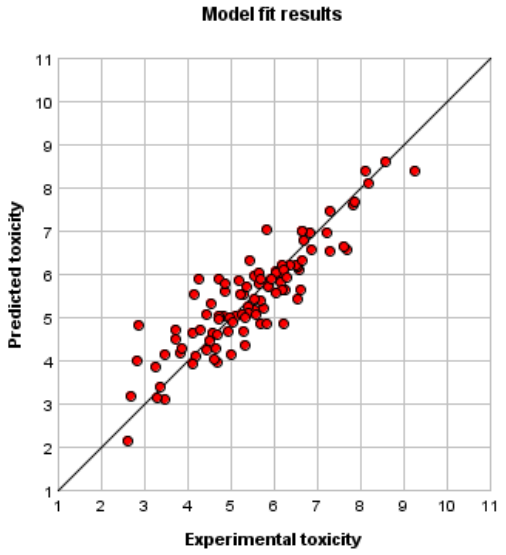
Model # 1296

Cluster models with violated constraints

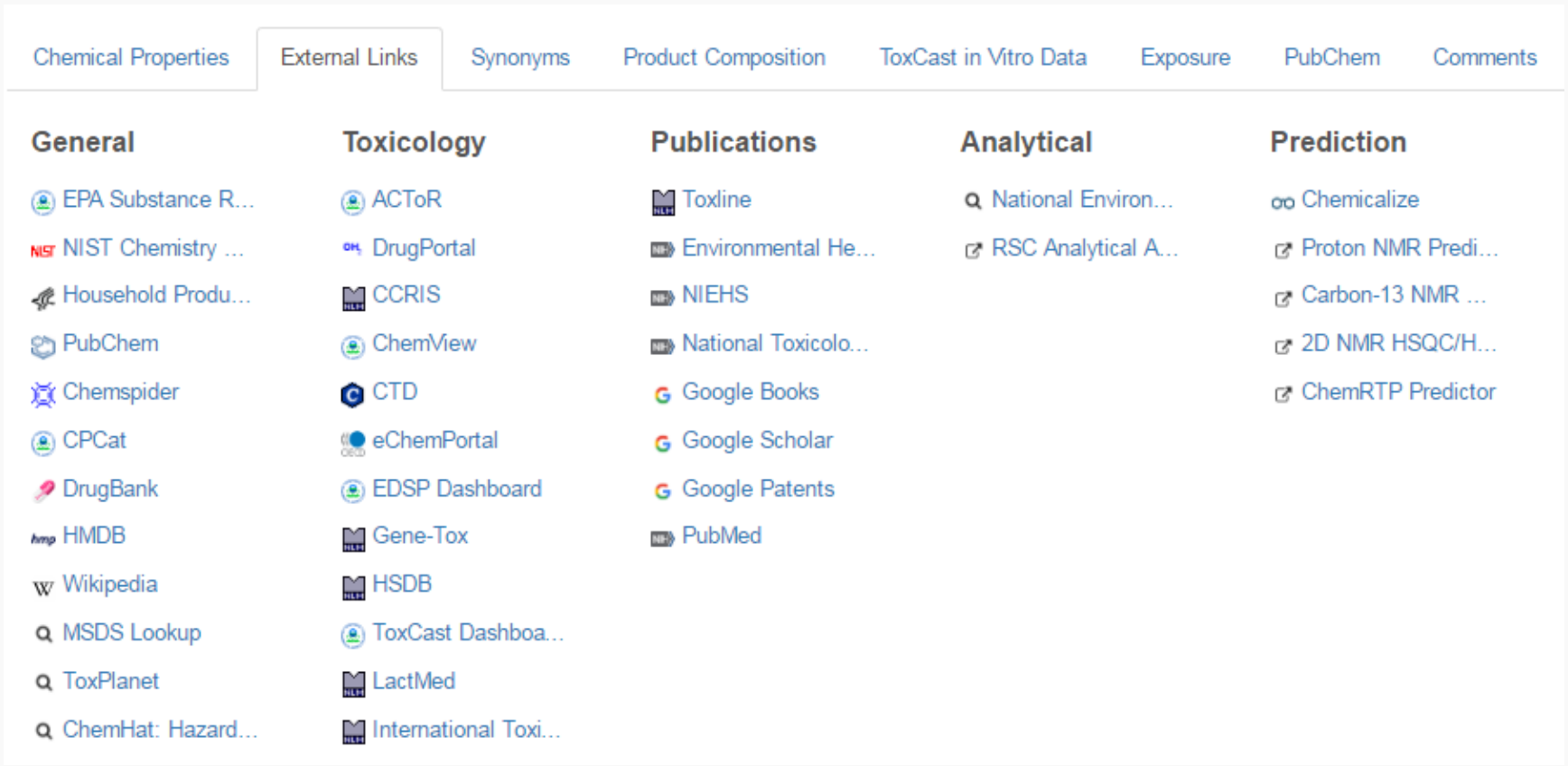
Cluster Model	r^2	q^2	# chemicals	Message
1121	0.810	0.576	10	Rmax constraint not met
1209	0.799	0.574	11	Fragment constraint not met
1247	0.919	0.647	20	Fragment constraint not met
1264	0.869	0.781	22	Fragment constraint not met
1268	0.675	0.553	24	Fragment constraint not met

Descriptor values for test chemical

Parameter	Value
Endpoint	Fathead minnow LC_{50} (96 hr)
r^2	0.793
q^2	0.733
#chemicals	101
Model	Model # 1296

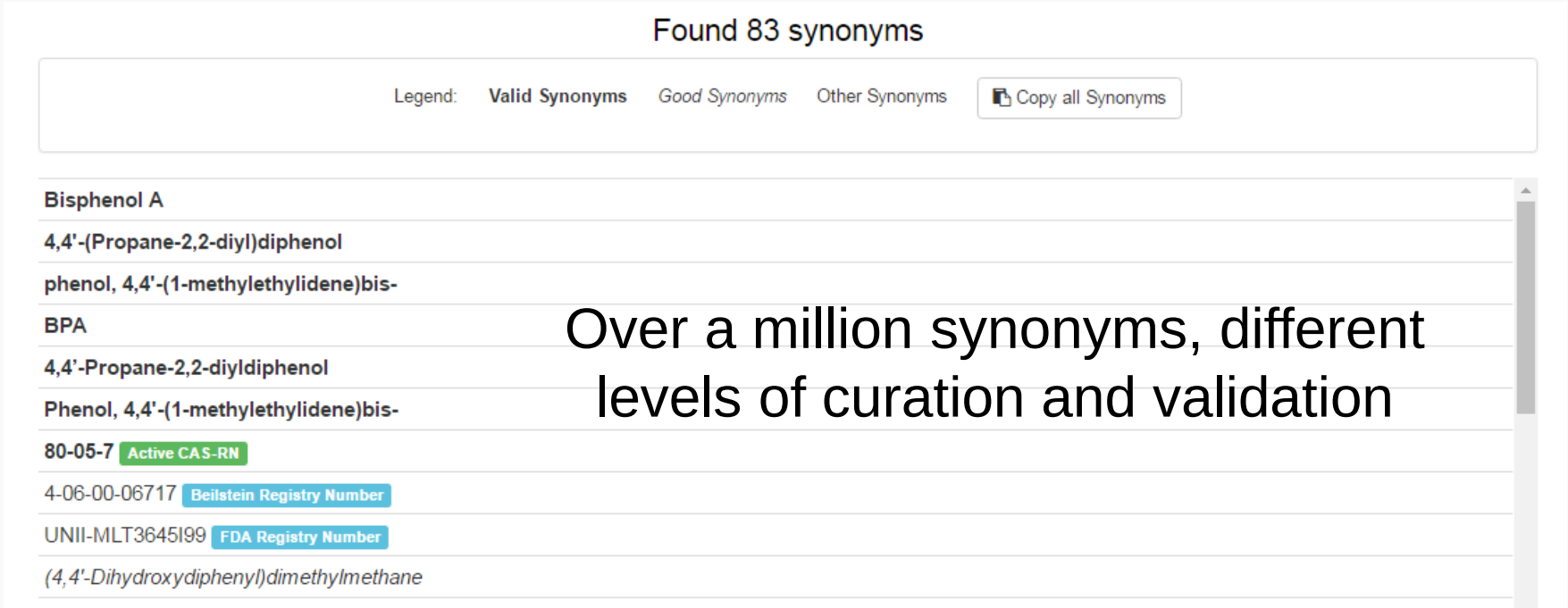


External Links



Chemical Properties External Links Synonyms Product Composition ToxCast in Vitro Data Exposure PubChem Comments

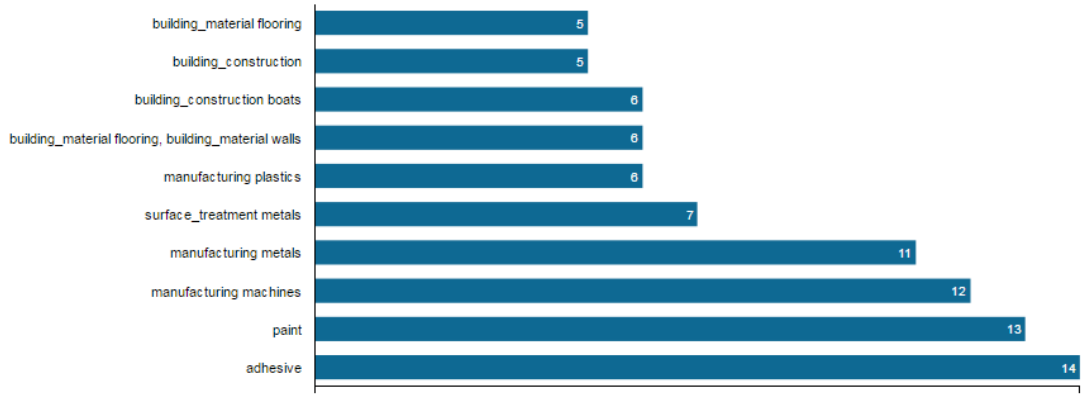
Synonyms



Over a million synonyms, different levels of curation and validation

Chemical Properties | External Links | Synonyms | Product Composition | ToxCast in Vitro Data | Exposure | PubChem | Comments

Functional Use and Composition (Integrating CPCat Data)



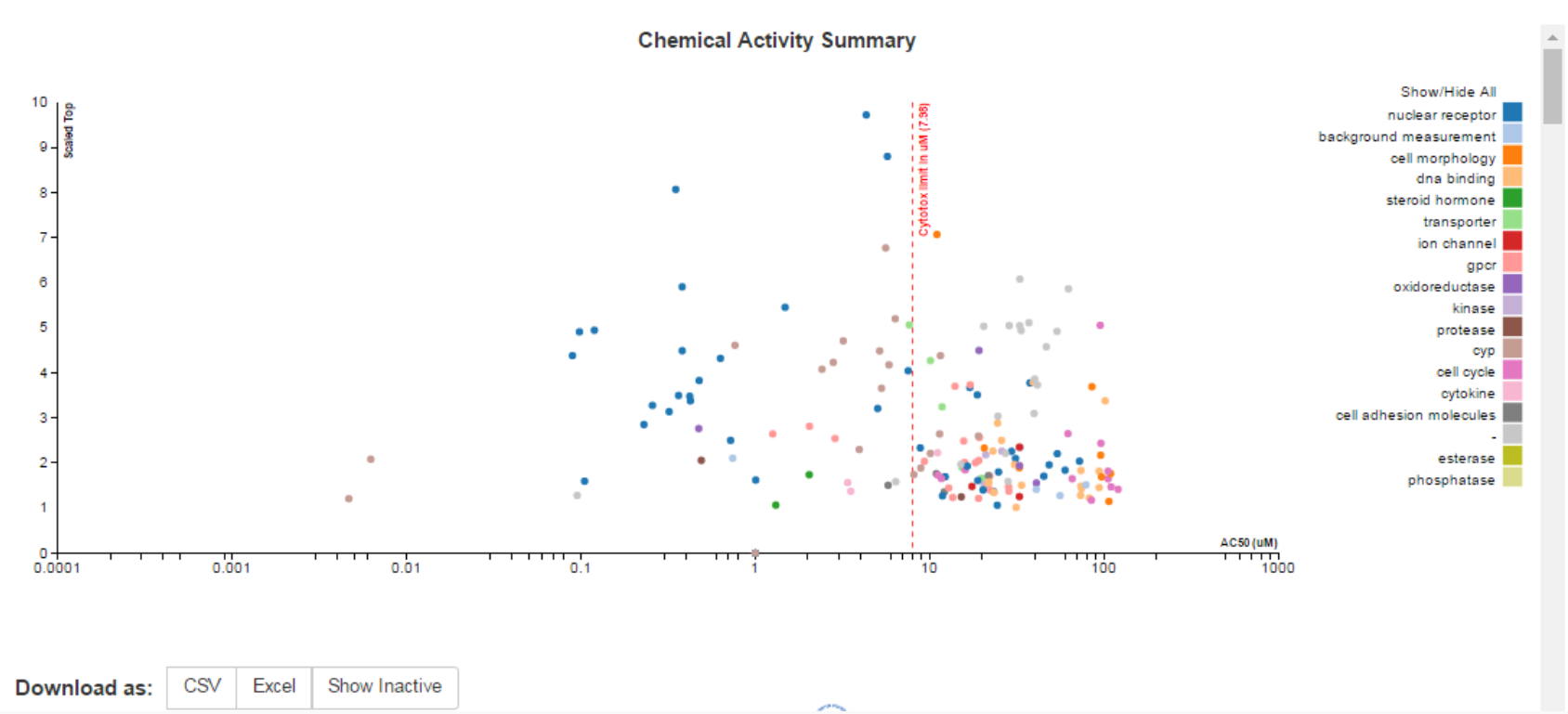
Download as: CSV Excel

Product Composition

Product	Percent Composition ↓	Manufacturer
BISPHENOL-A (BPA)	100%	GENERAL ELECTRIC COMPANY
EPOXY PASTE PIGMENTS, 3402-3408	<100%PPM	SYSTEM THREE RESINS
ISOPROPYLIDENEDIPHENOL, 99+%, 23965-8	99%+	ALDRICH CHEMICAL CO
BISPHENOL A (RESIN GRADE) (43106)	97.8%	SHELL OIL COMPANY
4,4-ISOPROPYLIDENEDIPHENOL, 97%, 13302-7	97%	ALDRICH CHEMICAL CO
ICO-PATCH EPOXY RESIN HARDENER, PART B	80%	INTERNATIONAL COATINGS CO
ADHESIVE-SCOTCH-WELD (R) 2216 B GRAY	72%	3M COMPANY
EPOCAST HARDENER 946, FPC 5000	45%	CIBA-GEIGY CORP
EPOLITE 1350 HANDENER	35-50	HEXCEL CORP, RESINS GROUP
EL CHEM NO 200 PRIMER PART B	32.5%	ELECTRO CHEMICAL ENGINEERING &

Chemical Properties | External Links | Synonyms | Product Composition | ToxCast in Vitro Data | Exposure | PubChem | Comments

Bioassay Screening Data (Integrating ToxCast Data)



Chemical Properties | External Links | Synonyms | Product Composition | ToxCast in Vitro Data | Exposure | PubChem | Comments

Bioassay Screening Data (Integrating ToxCast Data)

Assay Name	Hit Call	Top	Scaled Top	AC50 ↓	log AC50	Intended Target Family
APR_Hepat_CellLoss_48hr_dn						
APR_HepG2_OxidativeStress_24h_up						
APR_HepG2_MitoMass_24h_dn						
APR_Hepat_DNADamage_48hr_up						
APR_HepG2_CellLoss_24h_dn						
APR_HepG2_OxidativeStress_72h_up						
ATG_HSE_CIS_up						
APR_Hepat_DNADamage_24hr_up						
APR_Hepat_CellLoss_24hr_dn						
APR_Hepat_Steatosis_24hr_up						
APR_HepG2_CellLoss_72h_dn						

APR_HepG2_OxidativeStress_72h_up

Assay Name: APR_HepG2_OxidativeStress_72h_up

Gene Symbol: null

Organism: human

Tissue: liver

Assay Format Type: cell-based

Biological Process Target: oxidative phosphorylation

Detection Technology: Fluorescence

Analysis Direction: positive

Intended Target Family: cell cycle

Description: Data from the assay component APR_HepG2_OxidativeStress_72hr was analyzed into 2 assay endpoints. This assay endpoint, APR_HepG2_OxidativeStress_72h_up, was analyzed in the positive fitting direction relative to DMSO as the negative control and baseline of activity. Using a type of viability reporter, measures of protein for gain-of-signal activity can be used to understand the signaling at the pathway-level as they relate to the gene . Furthermore, this assay endpoint can be referred to as a primary readout, because this assay has produced multiple assay endpoints where this one serves a signaling function. To generalize the intended target to other relatable targets, this assay endpoint is annotated to the "cell cycle" intended target family, where the subfamily is "stress response".

Exposure Data – NHANES and ExpoCast Predictions

EPA United States Environmental Protection Agency

National Health and Nutrition Examination Survey (NHANES) Inferences (mg/kg-bw/day)

	Ages 6-11	Ages 12-19	Ages 20-65	Ages 65+	BMI > 30	BMI < 30	Repro. Age Females	Females	Males	Total
Minimum	3.80e-05	2.55e-05	2.79e-05	1.91e-05	2.38e-05	3.02e-05	2.83e-05	2.58e-05	2.94e-05	2.86e-05
Maximum	4.92e-05	3.38e-05	3.27e-05	2.31e-05	2.74e-05	3.30e-05	3.31e-05	3.03e-05	3.37e-05	3.08e-05
Mean	4.33e-05	2.93e-05	3.02e-05	2.10e-05	2.55e-05	3.16e-05	3.06e-05	2.80e-05	3.15e-05	2.97e-05

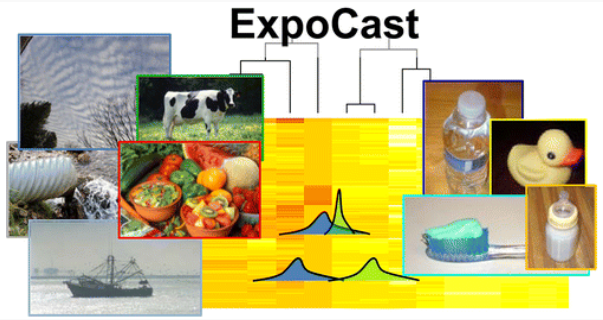
Exposure Predictions (mg/kg-bw/day)

	Ages 6-11	Ages 12-19	Ages 20-65	Ages 65+	BMI > 30	BMI < 30	Repro. Age Females	Females	Males	Total
Median	6.30e-05	2.68e-05	2.05e-05	1.61e-05	1.69e-05	2.67e-05	1.11e-05	1.11e-05	3.89e-05	2.11e-05
95th Percentile	5.82e-03	2.00e-03	1.61e-03	2.18e-03	1.45e-03	2.26e-03	1.57e-03	9.09e-04	3.34e-03	2.00e-03

National Health and Nutrition Examination Survey

High-Throughput Models for Exposure-Based Chemical Prioritization in the ExpoCast Project

Environ. Sci. Technol., **2013**, *47* (15), pp 8479–8488

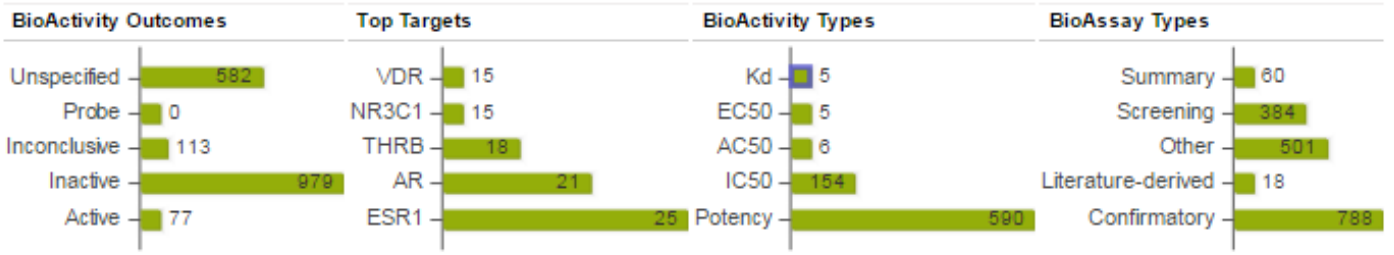


Chemical Properties | External Links | Synonyms | Product Composition | ToxCast in Vitro Data | Exposure | PubChem | Comments

PubChem integration

EPA United States Environmental Protection Agency

Biological Activities



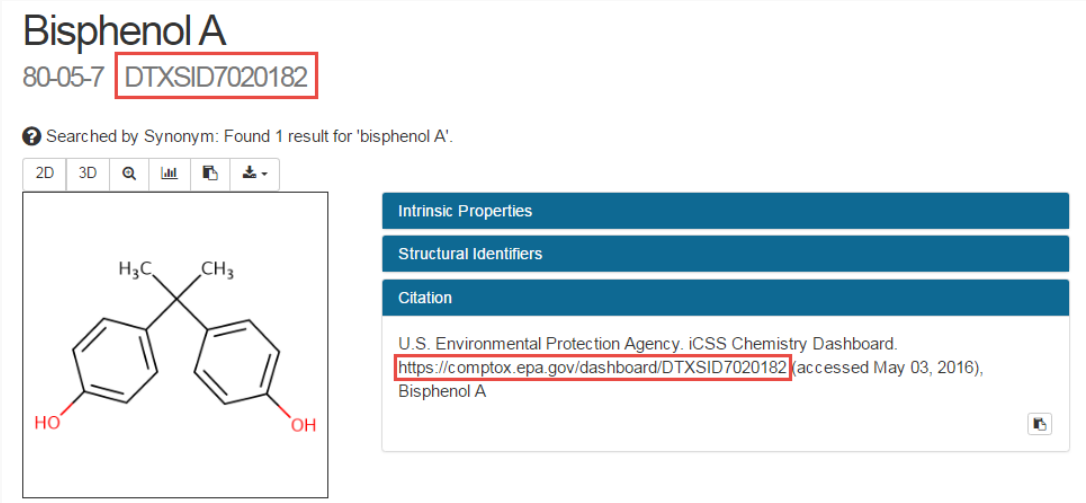
Structure	Substance SID	Activity			Compound Name	Bioassay Name
		Outcome	Type	Value [µM]		
	312534041	Inactive			Bisphenol A	Screen for inhibitors of RMI FANCM (MM2) intereaction [AID:1159607, Type: Other, PubMed ID:26962873]
	144214049	Inconclu...	Potency	62.1889	Bisphenol A	qHTS assay for small molecule agonists of the p53 signaling pathway [AID:651631, Type: Confirmatory]
	144214049	Inactive	Potency		Bisphenol A	qHTS assay for small molecules that induce genotoxicity in human embryonic kidney cells expressing luciferase-tagged ATAD5 [AID:651632, Type: Confirmatory]
	144214049	Inconclu...	Potency	55.4259	Bisphenol A	qHTS assay for small molecule agonists of the p53 signaling pathway - cell viability [AID:651633, Type: Confirmatory]
	144214049	Inactive	Potency		Bisphenol A	qHTS assay for small molecules that induce genotoxicity in human

Page 1 of 176 | Displaying 1 - 10 of 1751

Chemical Properties | External Links | Synonyms | Product Composition | ToxCast in Vitro Data | Exposure | PubChem | Comments

Connecting into the Dashboard

- Linkages into the Dashboard are simple: using the associated identifiers



- For integration use files of structures/identifiers mapped to DTXSIDs
- Integrated already - PubChem, EBI's UNICHEM, ChemSpider and whoever wants the files…

Our OPEN Data is available…

- Various types of data at FTP download site: ftp://newftp.epa.gov/COMPTOX/Sustainable_Chemistry_Data/Chemistry_Dashboard

Index of /COMPTOX/Sustainable_Chemistry_Data/Chemistry

Name	Size	Date Modified
[parent directory]		
PHYSPROP_Analysis/		9/6/16, 10:30:00 AM
DSSTOX_MS_Ready_Chemical_Structures.zip	280 MB	10/7/16, 1:32:00 PM
DSSTox_Mapping_20160701.zip	32.9 MB	8/25/16, 9:31:00 AM
DSSTox_Predicted_NCCT_Model.zip	343 MB	9/29/16, 10:13:00 AM
DSSTox_SDF_File_20160720.sdf.gz	162 MB	8/25/16, 9:34:00 AM
DSSTox_Synonyms_20161018.zip	145 MB	10/21/16, 3:31:00 PM
Dsstox_CAS_number_name.xlsx	32.1 MB	8/25/16, 9:36:00 AM
PubChem_DTXSID_mapping_file.txt	23.1 MB	9/27/16, 9:46:00 AM

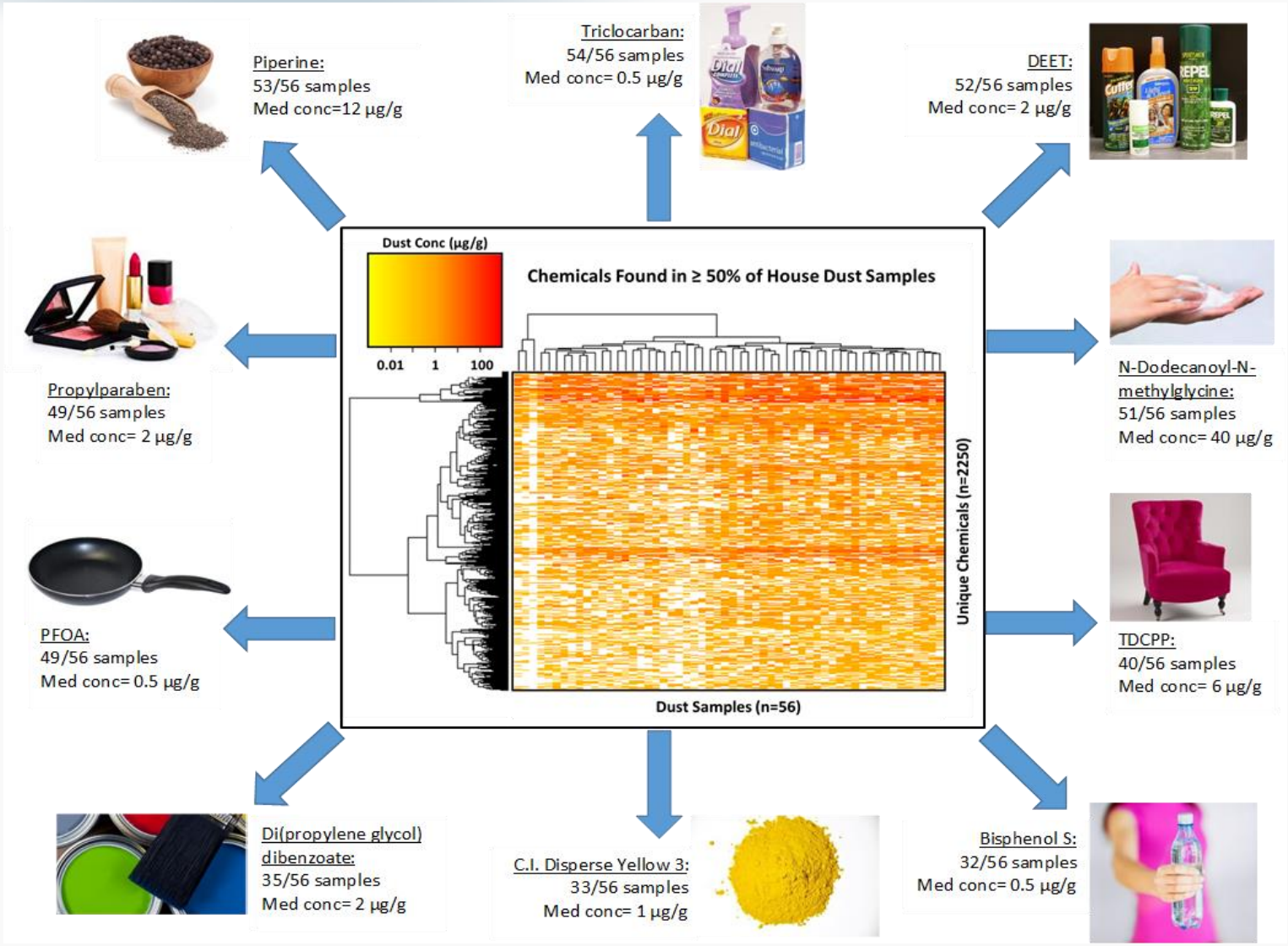
ONE Application of the Dashboard

- Targeted Analysis:
  - We know exactly what we're looking for
  - 10s – 100s of chemicals

- Suspect Screening Analysis (SSA):
  - We have chemicals of interest
  - 100s – 1,000s of chemicals

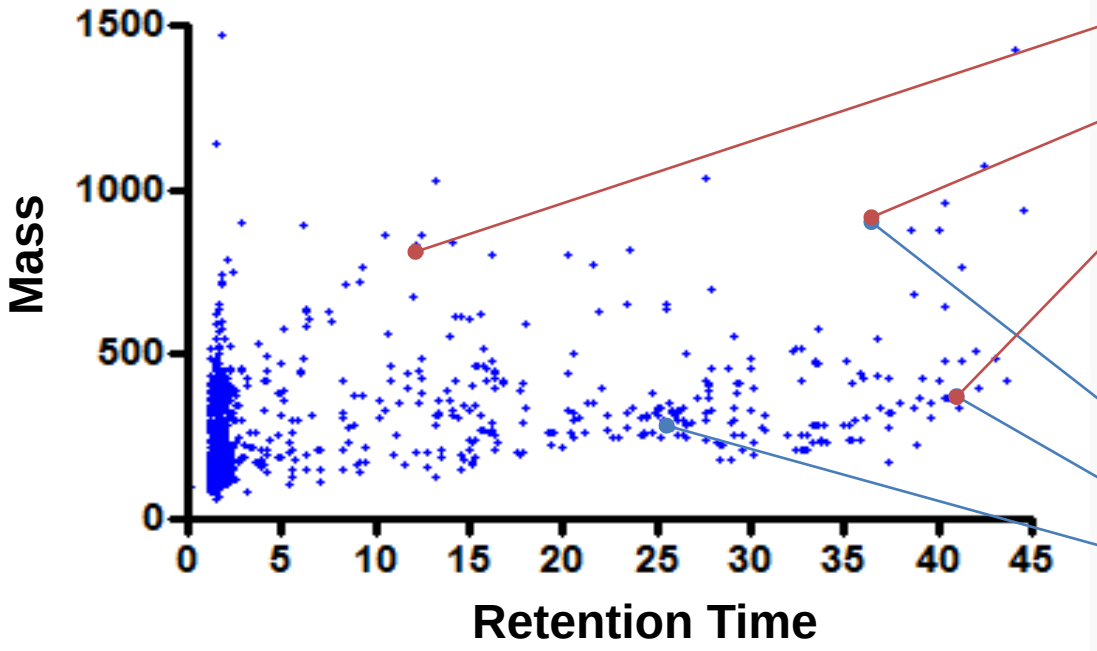
Suspect Screening Results

EPA United States Environmental Protection Agency



Suspect Screening in House Dust

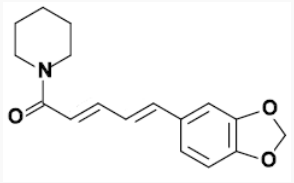
947 Peaks in an American Health Homes Dust Sample

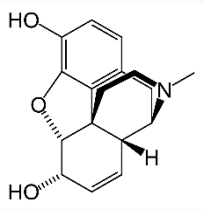


Liquid chromatography peaks corresponds to a chemical with an accurate mass and predicted formula:

$C_{17}H_{19}NO_3$

Multiple chemicals can have the same mass and formula:





Is chemical A present, chemical B, or both?

We are now expanding our identity libraries using reference samples of ToxCast chemicals

ONE Application of the Dashboard

- Targeted Analysis:
  - We know exactly what we're looking for
  - 10s – 100s of chemicals

- Suspect Screening Analysis (SSA)
  - We have chemicals of interest
  - 100s – 1,000s of chemicals

- Non-Targeted Analysis (NTA):
  - We have no preconceived lists
  - 1,000s – 10,000s of chemicals
  - In dust, soil, food, air, water, products, plants, animals, and…us!!

Previous Work with Suspect-Screening

Environment International 88 (2016) 269–280

Contents lists available at ScienceDirect

Environment International

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

Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring

Julia E. Rager [a], Mark J. Strynar [b], Shuang Liang [a], Rebecca L. McMahen [a], Ann M. Richard [c], Christopher M. Grulke [d], John F. Wambaugh [c], Kristin K. Isaacs [b], Richard Judson [c], Antony J. Williams [c], Jon R. Sobus [b,*]

[a] Oak Ridge Institute for Science and Education (ORISE) Participant, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States
[b] U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States
[c] U.S. Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States
[d] Lockheed Martin, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States

Rank-Ordering of "Known-Unknowns" using ChemSpider

EPA United States Environmental Protection Agency

RESEARCH ARTICLE

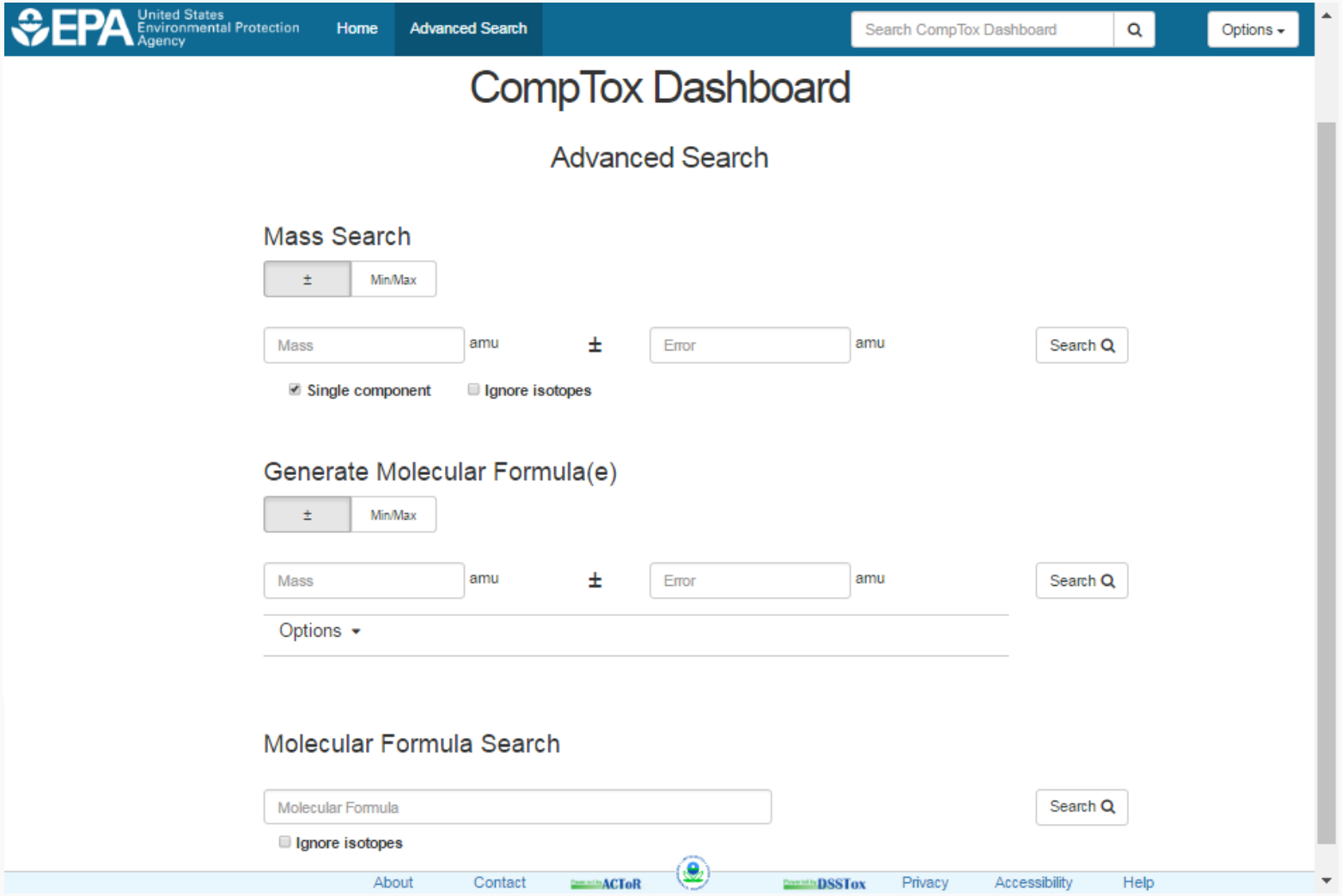
Identification of "Known Unknowns" Utilizing Accurate Mass Data and ChemSpider

Table 1. Searching ChemSpider by Elemental Composition then Sorting by Number of Associated References

Class of compounds	Number compounds in class	Position of compound sorted in de		
		#1	#2	#3
Drugs	45	43	1	1
Pesticides	8	7	1	
Toxins	2	2		
Polymer antioxidants	15	15		
Polymer UV stabilizers	10	8	1	1
Polymer clarifying agent (Irgaclear DM)	1			
Polyurethane additives	4	2	1	
Natural products	3	2		1
Herbicide (clofibric acid)	1	1		
Artificial sweetener (sucralose)	1	1		
Total compounds ChemSpider	90	81	4	3
Total compounds CAS Registry [1]	90	84	4	1

J. Am. Soc. Mass Spectrom. (2012) 23:179–185
DOI: 10.1007/s13361-011-0265-y

Advanced MS Searches



Does the Dashboard Add Value?

ChemSpider
Search and share chemistry

57 Million
chemical structures

CompTox Dashboard

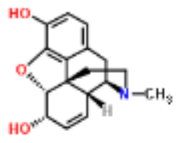
721k structures

Dilution Example… Morphine Skeleton

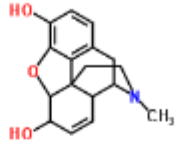
Found 32 results

Search term: **BQJCRHHNABKAKU** (Found by InChIKey (skeleton match))

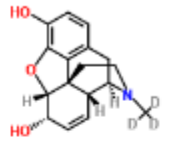
First | < | > | Last



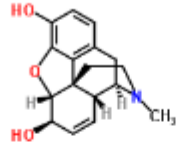
D-(-)-Morphine



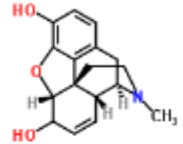
17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



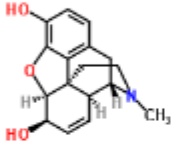
(5alpha,6alpha)-17-(~2~H_3_)Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



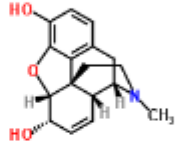
(5alpha,6beta)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



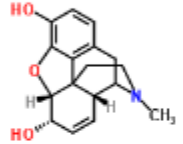
(5alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



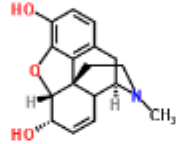
(5beta,6beta,9alpha,13alpha,14alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



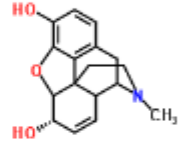
(5alpha,6alpha,9alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



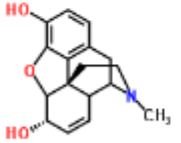
(5alpha,6alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



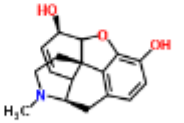
(5alpha,6alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



(6alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol

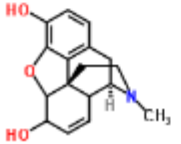


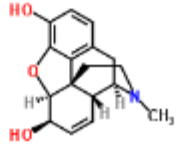
(6alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



(6beta)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol

(5beta, didehydr

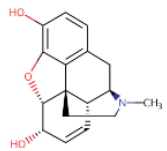




Searched by InchiKey Skeleton: Found 3 results for 'BQJCRHHNABKAKU'.

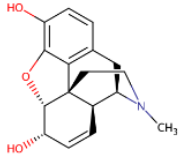
Download as: CSV | Excel | SDF

Preferred Name	CAS-RN	QC Level	Number of Sources	Monoisotopic Mass



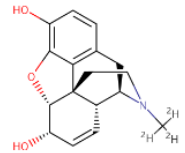
Morphine

57-27-2



Morphinan-3,6-d…

16206-77-2



Morphinan-3,6-a…

67293-88-3

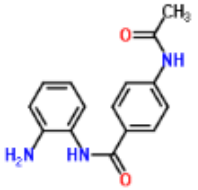
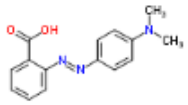
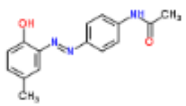
ChemSpider 6982 Results!!!
Search for C15H15N3O2

EPA United States Environmental Protection Agency

Found 6982 results

Search term: **MF = 'C_{15}H_{15}N_{3}O_{2}'**

1 2 3 4 5

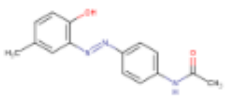
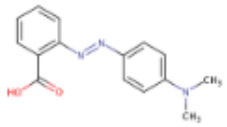
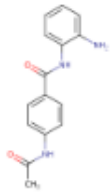
ID	Structure	Molecular Formula	Molecular Weight	# of Data Sources	
2644		$C_{15}H_{15}N_3O_2$	269.2985	86	**Tacedinaline**
9881		$C_{15}H_{15}N_3O_2$	269.2985	85	**Methyl Red**
10468668		$C_{15}H_{15}N_3O_2$	269.2985	33	**C.I Disperse Yellow 3**

Same top hits – different ranking
90 hits only versus **6926 hits**

EPA United States Environmental Protection Agency

Search Results

Searched by Molecular Formula, ignoring isotopes: Found 90 results for 'C15H15N3O2'.

Structure	Preferred Name	CAS-RN	QC Level	Number of Sources	Monoisotopic Mass
	**C.I Disperse Yellow 3**	2832-40-8	DSSTox Low	**18**	269.116427
	**Methyl Red**	493-52-7	DSSTox Low	**17**	269.116427
	**Tacedinaline**	112522-64-2	Public Medium	**4**	269.116427

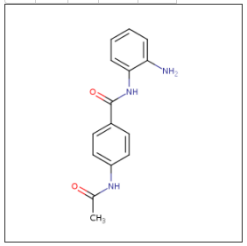
Using Meta-Data to Sort Candidates

EPA United States Environmental Protection Agency

Tacedinaline
112522-64-2 | DTXSID60150095

Anti-cancer Drug

Searched by Approved Name: Found 1 result for 'tacedinaline'.



Intrinsic Properties

Molecular Formula: C15H15N3O2 — Find All Chemicals

Average Mass: 269.304 g/mol

Monoisotopic Mass: 269.116427 g/mol

Structural Identifiers

Record Information

Chemical Properties | External Links | Synonyms | Product Composition | ToxCast in Vitro Data | Exposure | Analytical | PubChem | Comments

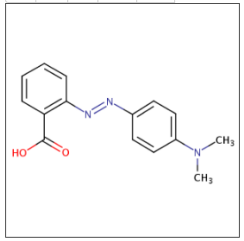
Frequent Uses and Functions

No product composition data found.

Methyl red
493-52-7 | DTXSID1042154

Microbiological Indicator Dye

Searched by Approved Name: Found 1 result for 'Methyl red'.



Intrinsic Properties

Molecular Formula: C15H15N3O2 — Find All Chemicals

Average Mass: 269.304 g/mol

Monoisotopic Mass: 269.116427 g/mol

Structural Identifiers

Record Information

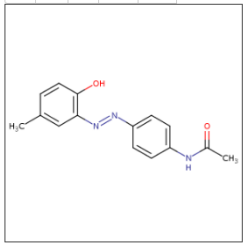
Chemical Properties | External Links | Synonyms | Product Composition | ToxCast in Vitro Data | Exposure | Analytical | PubChem | Comments

Frequent Uses and Functions

C.I. Disperse Yellow 3
2832-40-8 | DTXSID6021450

Textile/Product Dye

Searched by Approved Name: Found 1 result for 'C.I. Disperse Yellow 3'.



Intrinsic Properties

Molecular Formula: C15H15N3O2 — Find All Chemicals

Average Mass: 269.304 g/mol

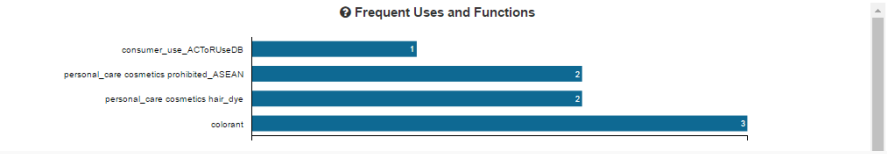
Monoisotopic Mass: 269.116427 g/mol

Structural Identifiers

Record Information

Chemical Properties | External Links | Synonyms | Product Composition | ToxCast in Vitro Data | Exposure | Analytical | PubChem | Comments

Frequent Uses and Functions



Dashboard vs ChemSpider Ranking Summary

- Selected peer-reviewed publications
- 162 total individual chemicals in search

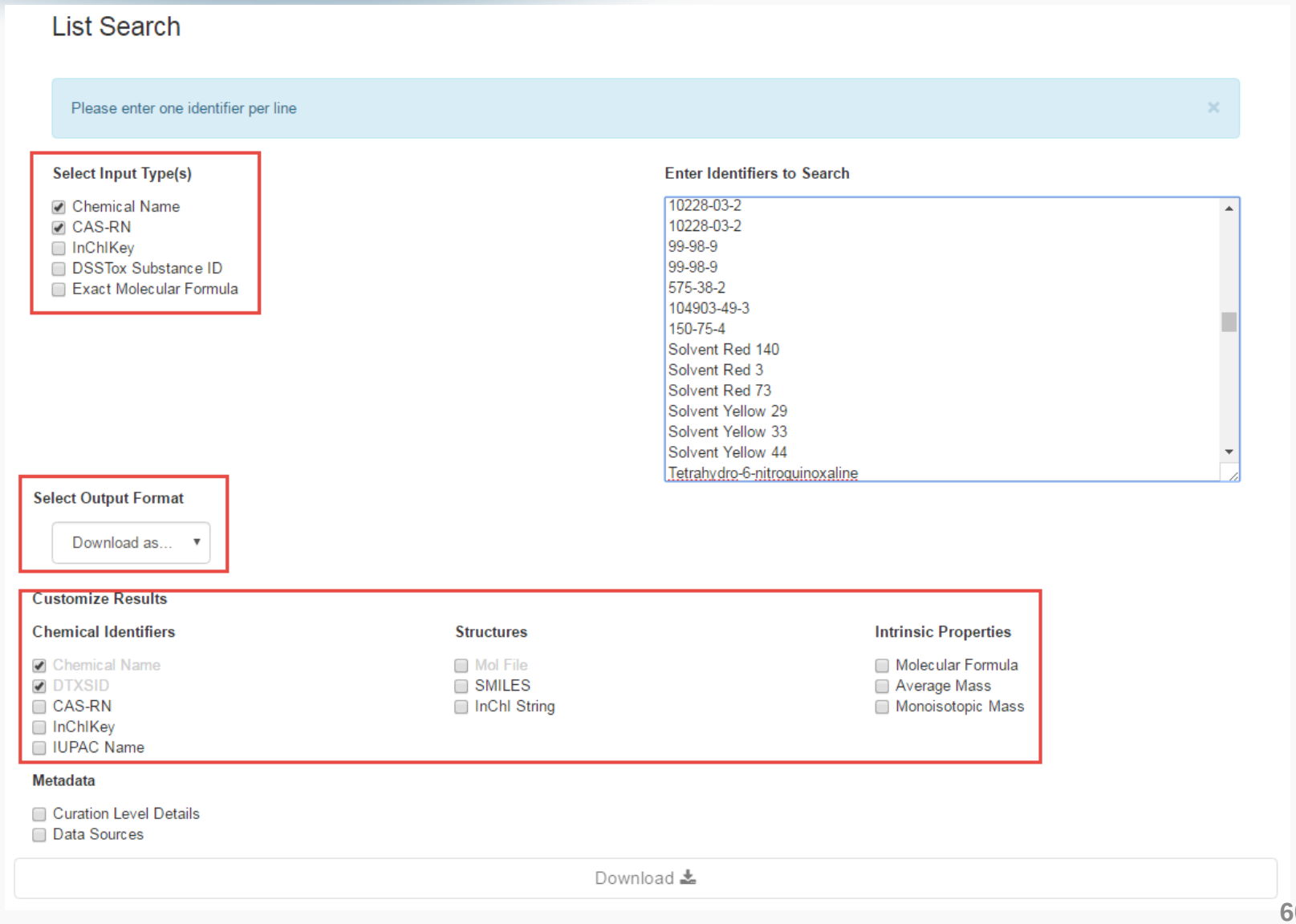
	Mass-based Searching		Formula Based Searching	
	Dashboard	*ChemSpider*	*Dashboard*	*ChemSpider*
Cumulative Average Position	1.3	2.2	1.2	1.4
% in #1 Position	85%	70%	88%	80%

Coming December 2016
Batch Searching Names/CASRNs

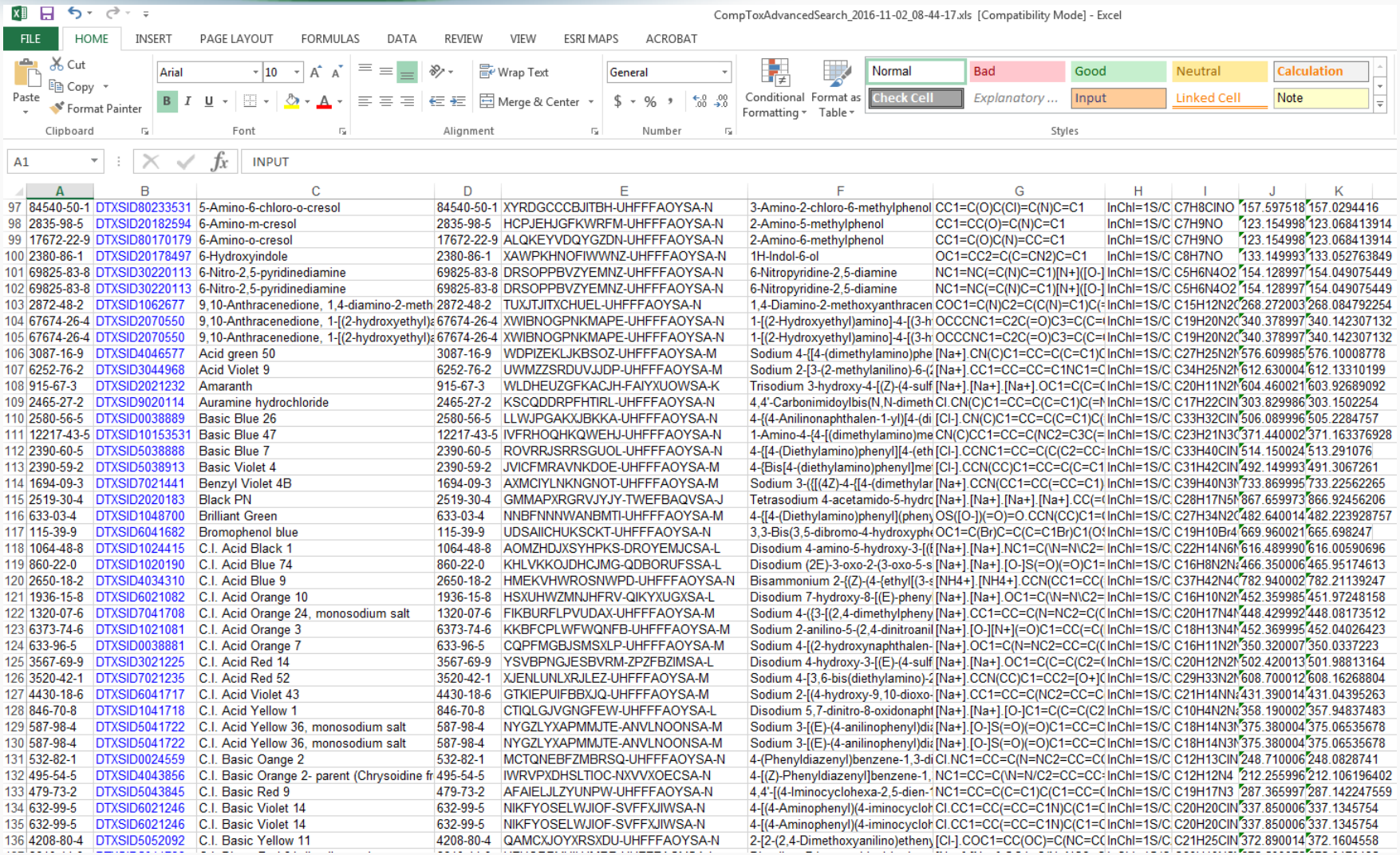
- What are these chemicals?

533-73-3	1,5-Naphthalenediol
95-54-5	2,7-Naphthalenediol
108-73-6	4-Amino-o-cresol
108-45-2	Disperse Blue 3
2051-85-6	Disperse Blue 377
106-50-3	3,4-Dihydroxyaniline
128-95-0	2,6-Dihydroxyethylaminotoluene
83-56-7	Catechol
27581-07-3	1-Acetoxy-2-methylnaphthalene
5697-02-9	Fast Green FCF
90-15-3	Acid Red 14
89-25-8	6-Nitro-2,5-pyridinediamine
92-44-4	Pyridine
15086-94-9	HC Orange 2
615-05-4	m-Aminophenol
5862-77-1	p-Aminophenol
5862-77-1	2-Methyl-4-nitroaniline
5862-77-1	o-Phenylenediamine
95-86-3	Acid Blue 1
95-80-7	2,3-Indolinedione
1004-75-7	1,2,4-Trihydroxybenzene
141-86-6	Solvent Red 23
141-86-6	2-Chloro-4-phenylenediamine
84540-47-6	2-Chloro-p-phenylenediamine
84540-47-6	p-Phenylenediamine
149330-25-6	Solvent Red 43
606-20-2	Acid Violet 9
823-40-5	Direct Blue 86
582-17-2	4-Hydroxyindole
16867-03-1	2,5,6-Triamino-4-pyrimidinol
16867-03-1	Solvent Red 1
603-85-0	HC Yellow 10
603-85-0	Oxazine 1
96-91-3	2,4-Diaminotoluene
95-85-2	m-Phenylenediamine

Coming December 2016
Batch Searching…

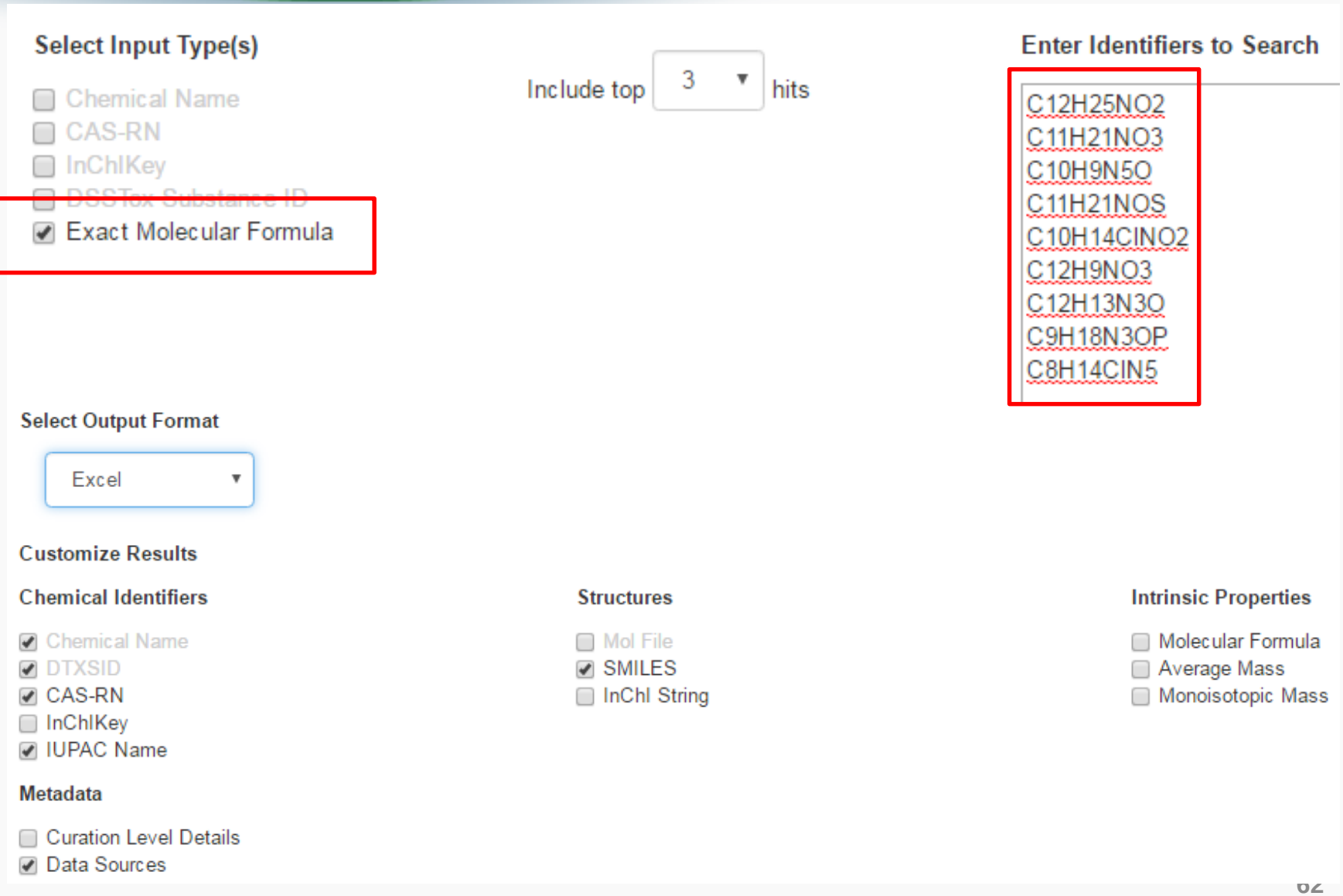


Coming December 2016
Download to Excel

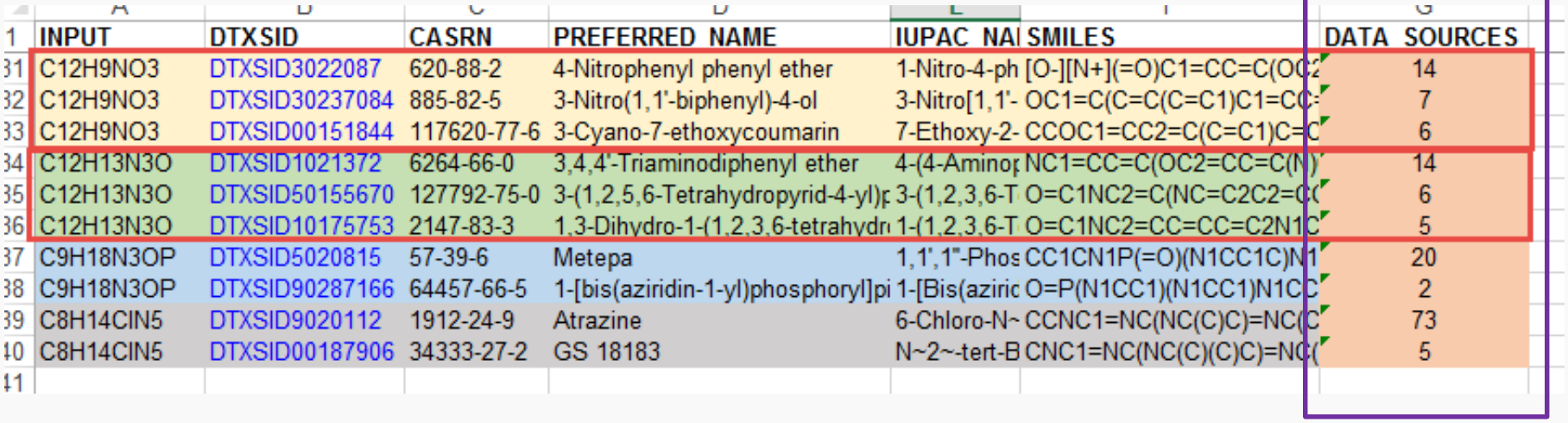


	A	B	C	D	E	F	G	H	I	J	K
97	84540-50-1	DTXSID80233531	5-Amino-6-chloro-o-cresol	84540-50-1	XYRDGCCCBJITBH-UHFFFAOYSA-N	3-Amino-2-chloro-6-methylphenol	CC1=C(O)C(Cl)=C(N)C=C1	InChI=1S/C	C7H8ClNO	157.597518	157.0294416
98	2835-98-5	DTXSID20182594	6-Amino-m-cresol	2835-98-5	HCPJEHJGFKWRFM-UHFFFAOYSA-N	2-Amino-5-methylphenol	CC1=CC(O)=C(N)C=C1	InChI=1S/C	C7H9NO	123.154998	123.068413914
99	17672-22-9	DTXSID80170179	6-Amino-o-cresol	17672-22-9	ALQKEYVDQYGZDN-UHFFFAOYSA-N	2-Amino-6-methylphenol	CC1=C(O)C(N)=CC=C1	InChI=1S/C	C7H9NO	123.154998	123.068413914
100	2380-86-1	DTXSID20178497	6-Hydroxyindole	2380-86-1	XAWPKHNOFIWWNZ-UHFFFAOYSA-N	1H-Indol-6-ol	OC1=CC2=C(C=CN2)C=C1	InChI=1S/C	C8H7NO	133.149993	133.052763849
101	69825-83-8	DTXSID30220113	6-Nitro-2,5-pyridinediamine	69825-83-8	DRSOPPBVZYEMNZ-UHFFFAOYSA-N	6-Nitropyridine-2,5-diamine	NC1=NC(=C(N)C=C1)[N+]([O-]	InChI=1S/C	C5H6N4O2	154.128997	154.049075449
102	69825-83-8	DTXSID30220113	6-Nitro-2,5-pyridinediamine	69825-83-8	DRSOPPBVZYEMNZ-UHFFFAOYSA-N	6-Nitropyridine-2,5-diamine	NC1=NC(=C(N)C=C1)[N+]([O-]	InChI=1S/C	C5H6N4O2	154.128997	154.049075449
103	2872-48-2	DTXSID1062677	9,10-Anthracenedione, 1,4-diamino-2-meth	2872-48-2	TUXJTJITXCHUEL-UHFFFAOYSA-N	1,4-Diamino-2-methoxyanthracen	COC1=C(N)C2=C(C(N)=C1)C(=	InChI=1S/C	C15H12N2C	268.272003	268.084792254
104	67674-26-4	DTXSID2070550	9,10-Anthracenedione, 1-[(2-hydroxyethyl)a	67674-26-4	XWIBNOGPNKMAPE-UHFFFAOYSA-N	1-[(2-Hydroxyethyl)amino]-4-[(3-h	OCCCNC1=C2C(=O)C3=C(C=C	InChI=1S/C	C19H20N2C	340.378997	340.142307132
105	67674-26-4	DTXSID2070550	9,10-Anthracenedione, 1-[(2-hydroxyethyl)a	67674-26-4	XWIBNOGPNKMAPE-UHFFFAOYSA-N	1-[(2-Hydroxyethyl)amino]-4-[(3-h	OCCCNC1=C2C(=O)C3=C(C=C	InChI=1S/C	C19H20N2C	340.378997	340.142307132
106	3087-16-9	DTXSID4046577	Acid green 50	3087-16-9	WDPIZEKLJKBSOZ-UHFFFAOYSA-M	Sodium 4-[[4-(dimethylamino)phe	[Na+].CN(C)C1=CC=C(C=C1)C	InChI=1S/C	C27H25N2N	576.609985	576.10008778
107	6252-76-2	DTXSID3044968	Acid Violet 9	6252-76-2	UWMZZSRDUVJJDP-UHFFFAOYSA-M	Sodium 2-[3-(2-methylanilino)-6-(2	[Na+].CC1=CC=CC=C1NC1=C(	InChI=1S/C	C34H25N2N	612.630004	612.13310199
108	915-67-3	DTXSID2021232	Amaranth	915-67-3	WLDHEUZGFKACJH-FAIYXUOWSA-K	Trisodium 3-hydroxy-4-[(Z)-(4-sulf	[Na+].[Na+].[Na+].OC1=C(C=(	InChI=1S/C	C20H11N2N	604.460021	603.92689092
109	2465-27-2	DTXSID9020114	Auramine hydrochloride	2465-27-2	KSCQDDRPFHTIRL-UHFFFAOYSA-N	4,4'-Carbonimidoylbis(N,N-dimeth	Cl.CN(C)C1=CC=C(C=C1)C(=N	InChI=1S/C	C17H22ClN	303.829986	303.1502254
110	2580-56-5	DTXSID0038889	Basic Blue 26	2580-56-5	LLWJPGAKXJBKKA-UHFFFAOYSA-N	4-[(4-Anilinonaphthalen-1-yl)[4-(di	[Cl-].CN(C)C1=CC=C(C=C1)C(	InChI=1S/C	C33H32ClN	506.089996	505.2284757
111	12217-43-5	DTXSID10153531	Basic Blue 47	12217-43-5	IVFRHOQHKQWEHJ-UHFFFAOYSA-N	1-Amino-4-[4-[(dimethylamino)me	CN(C)CC1=CC=C(NC2=C3C(=O	InChI=1S/C	C23H21N3C	371.440002	371.163376928
112	2390-60-5	DTXSID5038888	Basic Blue 7	2390-60-5	ROVRRJSRRSGUOL-UHFFFAOYSA-N	4-[[4-(Diethylamino)phenyl][4-(eth	[Cl-].CCNC1=CC=C(C(C2=CC=	InChI=1S/C	C33H40ClN	514.150024	513.291076
113	2390-59-2	DTXSID5038913	Basic Violet 4	2390-59-2	JVICFMRAVNKDOE-UHFFFAOYSA-M	4-[Bis[4-(diethylamino)phenyl]me	[Cl-].CCN(CC)C1=CC=C(C=C1	InChI=1S/C	C31H42ClN	492.149993	491.3067261
114	1694-09-3	DTXSID7021441	Benzyl Violet 4B	1694-09-3	AXMCIYLNKNGNOT-UHFFFAOYSA-M	Sodium 3-([[(4Z)-4-[[4-(dimethylar	[Na+].CCN(CC1=CC(=CC=C1)	InChI=1S/C	C39H40N3N	733.869995	733.22562265
115	2519-30-4	DTXSID2020183	Black PN	2519-30-4	GMMAPXRGRVJYJY-TWEFBAQVSA-J	Tetrasodium 4-acetamido-5-hydrc	[Na+].[Na+].[Na+].[Na+].CC(=(	InChI=1S/C	C28H17N5N	867.659973	866.92456206
116	633-03-4	DTXSID1048700	Brilliant Green	633-03-4	NNBFNNNWANBMTI-UHFFFAOYSA-M	4-[[4-(Diethylamino)phenyl](phen	OS([O-])(=O)=O.CCN(CC)C1=(	InChI=1S/C	C27H34N2O	482.640014	482.223928757
117	115-39-9	DTXSID6041682	Bromophenol blue	115-39-9	UDSAIICHUKSCKT-UHFFFAOYSA-N	3,3-Bis(3,5-dibromo-4-hydroxyphe	OC1=C(Br)C=C(C=C1Br)C1(O	InChI=1S/C	C19H10Br4	669.960021	665.698247
118	1064-48-8	DTXSID1024415	C.I. Acid Black 1	1064-48-8	AOMZHDJXSYHPKS-DROYEMJCSA-L	Disodium 4-amino-5-hydroxy-3-[(E	[Na+].[Na+].NC1=C(\N=N\C2=	InChI=1S/C	C22H14N6N	616.489990	616.00590696
119	860-22-0	DTXSID1020190	C.I. Acid Blue 74	860-22-0	KHLVKKOJDHCJMG-QDBORUFSSA-L	Disodium (2E)-3-oxo-2-(3-oxo-5-s	[Na+].[Na+].[O-]S(=O)(=O)C1=	InChI=1S/C	C16H8N2N	466.350006	465.95174613
120	2650-18-2	DTXSID4034310	C.I. Acid Blue 9	2650-18-2	HMEKVHWROSNWPD-UHFFFAOYSA-N	Bisammonium 2-[(Z)-(4-[ethyl[(3-s	[NH4+].[NH4+].CCN(CC1=CC(	InChI=1S/C	C37H42N4C	782.940002	782.21139247
121	1936-15-8	DTXSID6021082	C.I. Acid Orange 10	1936-15-8	HSXUHWZMNJHFRV-QIKYXUGXSA-L	Disodium 7-hydroxy-8-[(E)-pheny	[Na+].[Na+].OC1=C(\N=N\C2=	InChI=1S/C	C16H10N2N	452.359985	451.97248158
122	1320-07-6	DTXSID7041708	C.I. Acid Orange 24, monosodium salt	1320-07-6	FIKBURFLPVUDAX-UHFFFAOYSA-M	Sodium 4-([3-[(2,4-dimethylpheny	[Na+].CC1=CC=C(N=NC2=C(C	InChI=1S/C	C20H17N4N	448.429992	448.08173512
123	6373-74-6	DTXSID1021081	C.I. Acid Orange 3	6373-74-6	KKBFCPLWFWQNFB-UHFFFAOYSA-M	Sodium 2-anilino-5-(2,4-dinitroanil	[Na+].[O-][N+](=O)C1=CC(=C(	InChI=1S/C	C18H13N4N	452.369995	452.04026423
124	633-96-5	DTXSID0038881	C.I. Acid Orange 7	633-96-5	CQPFMGBJSMSXLP-UHFFFAOYSA-M	Sodium 4-[(2-hydroxynaphthalen-	[Na+].OC1=C(N=NC2=CC=C(C	InChI=1S/C	C16H11N2N	350.320007	350.0337223
125	3567-69-9	DTXSID3021225	C.I. Acid Red 14	3567-69-9	YSVBPNGJESBVRM-ZPZFBZIMSA-L	Disodium 4-hydroxy-3-[(E)-(4-sulf	[Na+].[Na+].OC1=C(C=C(C2=(	InChI=1S/C	C20H12N2N	502.420013	501.98813164
126	3520-42-1	DTXSID7021235	C.I. Acid Red 52	3520-42-1	XJENLUNLXRJLEZ-UHFFFAOYSA-M	Sodium 4-[3,6-bis(diethylamino)-2	[Na+].CCN(CC)C1=CC2=[O+]C	InChI=1S/C	C29H33N2N	608.700012	608.16268804
127	4430-18-6	DTXSID6041717	C.I. Acid Violet 43	4430-18-6	GTKIEPUIFBBXJQ-UHFFFAOYSA-M	Sodium 2-[(4-hydroxy-9,10-dioxo-	[Na+].CC1=CC=C(NC2=CC=C	InChI=1S/C	C21H14NN	431.390014	431.04395263
128	846-70-8	DTXSID1041718	C.I. Acid Yellow 1	846-70-8	CTIQLGJVGNGFEW-UHFFFAOYSA-L	Disodium 5,7-dinitro-8-oxidonaphl	[Na+].[Na+].[O-]C1=C(C=C(C2	InChI=1S/C	C10H4N2N	358.190002	357.94837483
129	587-98-4	DTXSID5041722	C.I. Acid Yellow 36, monosodium salt	587-98-4	NYGZLYXAPMMJTE-ANVLNOONSA-M	Sodium 3-[(E)-(4-anilinophenyl)dia	[Na+].[O-]S(=O)(=O)C1=CC=C	InChI=1S/C	C18H14N3N	375.380004	375.06535678
130	587-98-4	DTXSID5041722	C.I. Acid Yellow 36, monosodium salt	587-98-4	NYGZLYXAPMMJTE-ANVLNOONSA-M	Sodium 3-[(E)-(4-anilinophenyl)dia	[Na+].[O-]S(=O)(=O)C1=CC=C	InChI=1S/C	C18H14N3N	375.380004	375.06535678
131	532-82-1	DTXSID0024559	C.I. Basic Oange 2	532-82-1	MCTQNEBFZMBRSQ-UHFFFAOYSA-N	4-(Phenyldiazenyl)benzene-1,3-di	Cl.NC1=CC=C(N=NC2=CC=C(	InChI=1S/C	C12H13ClN	248.710006	248.0828741
132	495-54-5	DTXSID4043856	C.I. Basic Orange 2- parent (Chrysoidine fr	495-54-5	IWRVPXDHSLTIOC-NXVVXOECSA-N	4-[(Z)-Phenyldiazenyl]benzene-1,	NC1=CC=C(\N=N/C2=CC=CC=	InChI=1S/C	C12H12N4	212.255996	212.106196402
133	479-73-2	DTXSID5043845	C.I. Basic Red 9	479-73-2	AFAIELJLZYUNPW-UHFFFAOYSA-N	4,4'-[(4-Iminocyclohexa-2,5-dien-1	NC1=CC=C(C=C1)C(C1=CC=C(	InChI=1S/C	C19H17N3	287.365997	287.142247559
134	632-99-5	DTXSID6021246	C.I. Basic Violet 14	632-99-5	NIKFYOSELWJIOF-SVFFXJIWSA-N	4-[(4-Aminophenyl)(4-iminocyclol	Cl.CC1=CC(=CC=C1N)C(C1=C(	InChI=1S/C	C20H20ClN	337.850006	337.1345754
135	632-99-5	DTXSID6021246	C.I. Basic Violet 14	632-99-5	NIKFYOSELWJIOF-SVFFXJIWSA-N	4-[(4-Aminophenyl)(4-iminocyclol	Cl.CC1=CC(=CC=C1N)C(C1=C(	InChI=1S/C	C20H20ClN	337.850006	337.1345754
136	4208-80-4	DTXSID5052092	C.I. Basic Yellow 11	4208-80-4	QAMCXJOYXRSXDU-UHFFFAOYSA-N	2-[2-(2,4-Dimethoxyanilino)ethen	[Cl-].COC1=CC(OC)=C(NC=CC	InChI=1S/C	C21H25ClN	372.890014	372.1604558

Non-Targeted Analysis: In-testing

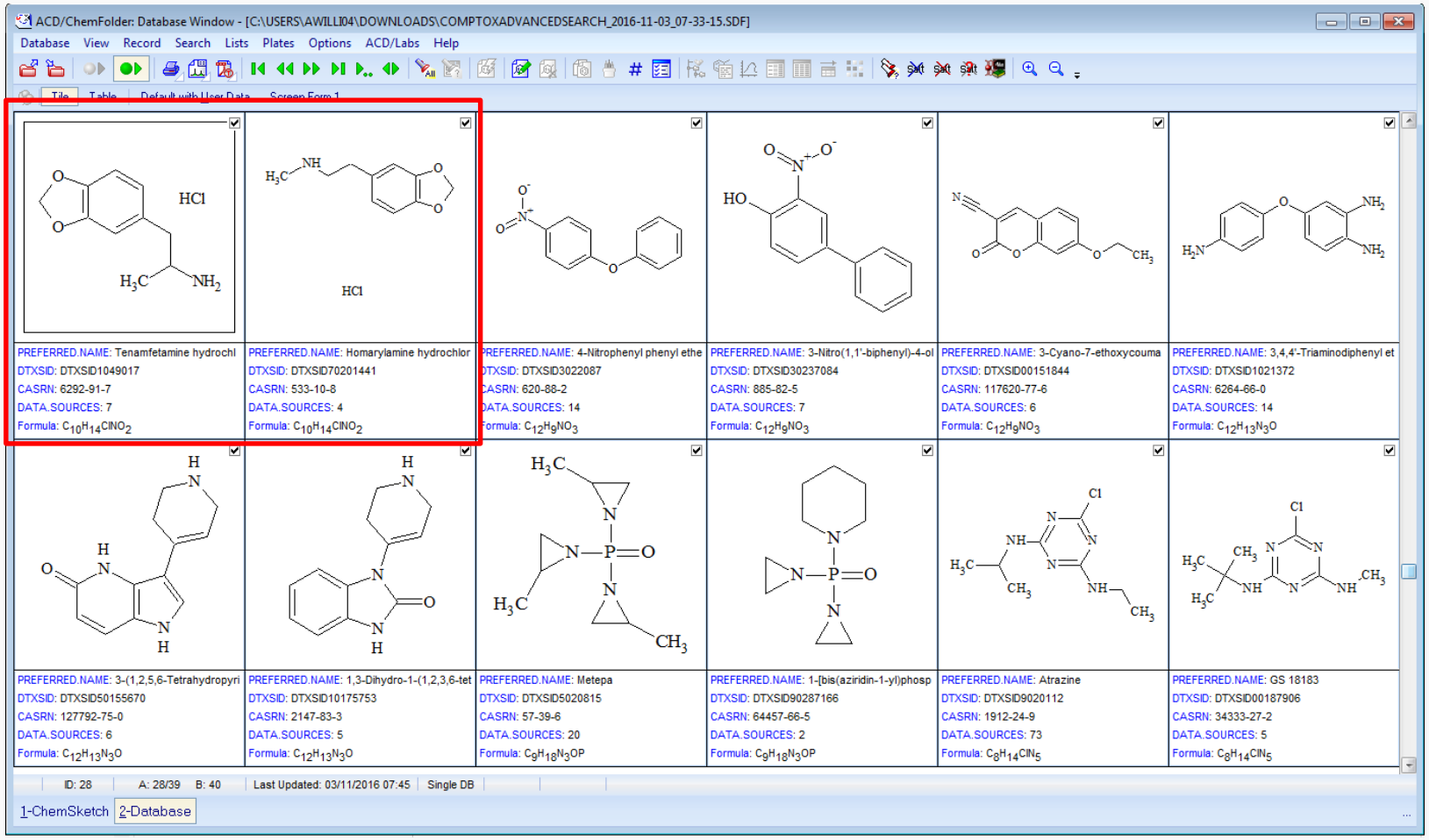


Metadata included for Ranking



	INPUT	DTXSID	CASRN	PREFERRED_NAME	IUPAC_NA	SMILES	DATA_SOURCES
31	C12H9NO3	DTXSID3022087	620-88-2	4-Nitrophenyl phenyl ether	1-Nitro-4-ph	[O-][N+](=O)C1=CC=C(OC2	14
32	C12H9NO3	DTXSID30237084	885-82-5	3-Nitro(1,1'-biphenyl)-4-ol	3-Nitro[1,1'-	OC1=C(C=C(C=C1)C1=CC	7
33	C12H9NO3	DTXSID00151844	117620-77-6	3-Cyano-7-ethoxycoumarin	7-Ethoxy-2-	CCOC1=CC2=C(C=C1)C=C	6
34	C12H13N3O	DTXSID1021372	6264-66-0	3,4,4'-Triaminodiphenyl ether	4-(4-Aminop	NC1=CC=C(OC2=CC=C(N)	14
35	C12H13N3O	DTXSID50155670	127792-75-0	3-(1,2,5,6-Tetrahydropyrid-4-yl)p	3-(1,2,3,6-T	O=C1NC2=C(NC=C2C2=C	6
36	C12H13N3O	DTXSID10175753	2147-83-3	1,3-Dihydro-1-(1,2,3,6-tetrahydr	1-(1,2,3,6-T	O=C1NC2=CC=CC=C2N1C	5
37	C9H18N3OP	DTXSID5020815	57-39-6	Metepa	1,1',1"-Phos	CC1CN1P(=O)(N1CC1C)N1	20
38	C9H18N3OP	DTXSID90287166	64457-66-5	1-[bis(aziridin-1-yl)phosphoryl]pi	1-[Bis(aziric	O=P(N1CC1)(N1CC1)N1CC	2
39	C8H14ClN5	DTXSID9020112	1912-24-9	Atrazine	6-Chloro-N~	CCNC1=NC(NC(C)C)=NC(C	73
40	C8H14ClN5	DTXSID00187906	34333-27-2	GS 18183	N~2~-tert-B	CNC1=NC(NC(C)(C)C)=NC(	5
41							

Need for “MS-Ready Structures”



“QSAR-Ready Structures”

- For the purpose of building QSAR Models we already “standardize” structures
  - Desalt/Neutralize
  - Desolvate
  - Remove stereochemistry

- Some minor tweaks gets us “MS-ready Structures”. ALREADY in our database.

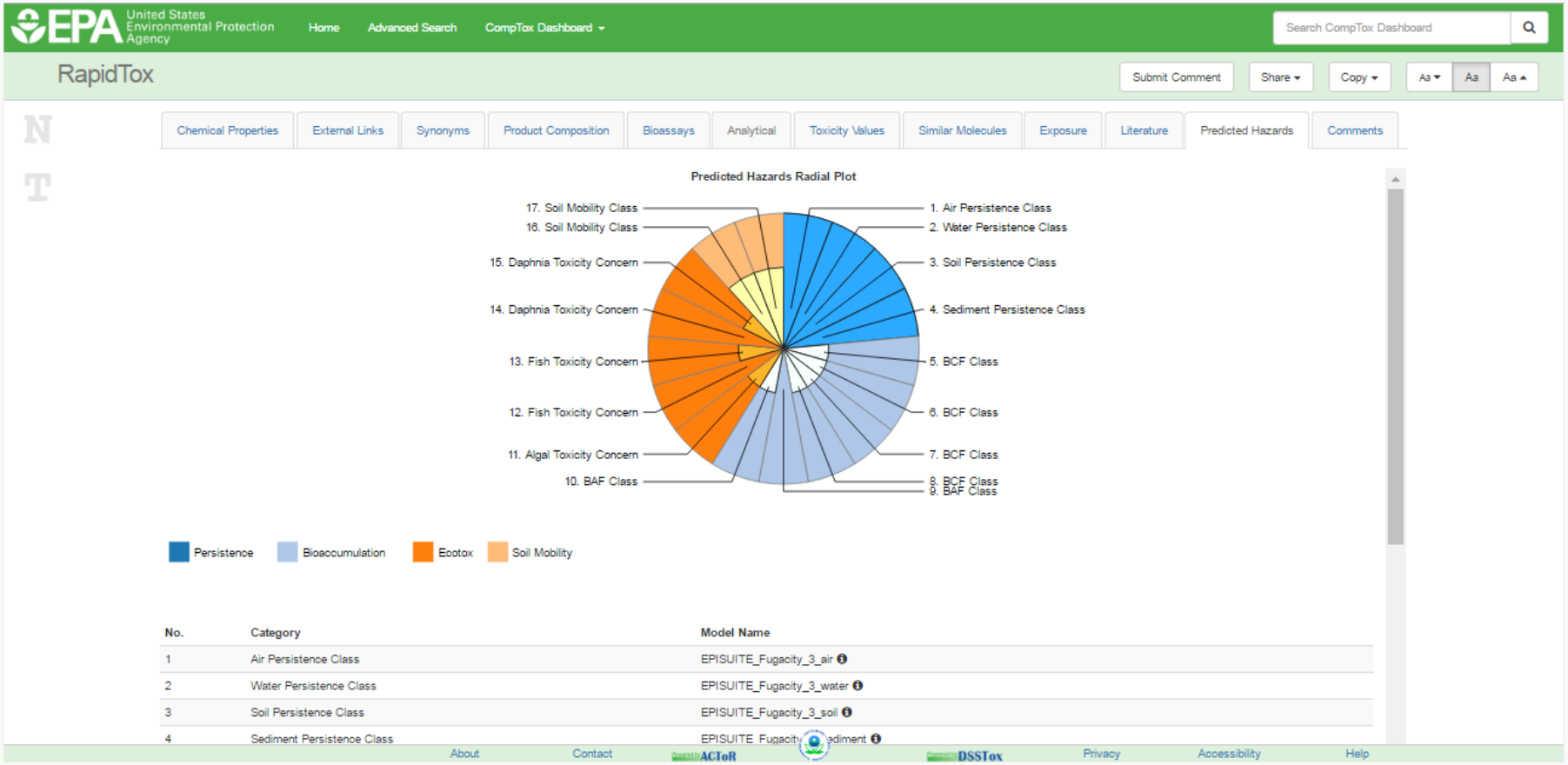
“QSAR-Ready Structures”

- Mass and Formula-based searches will be based on MS-ready structures but connected to the original chemical (with name, CAS, rank ordering)

	INPUT	DTXCID	DTXSID	CASRN	PREFERRED_NAME	MOL_FORMULA	MONOISOTOPI	DATA_SOURCES
2	C6H6O	DTXCID01234	DTXSID5021124	108-95-2	**Phenol**	C6H6O	**94.04186481**	**77**
3	C6H6O	DTXCID01234	DTXSID4027072	139-02-6	**Sodium phenolate**	C6H5NaO	**116.023809**	**19**
4	C6H6O	DTXCID01234	DTXSID10206632	5793-84-0	**Calcium phenoxide**	C12H10CaO2	**226.030671**	**5**
5	C6H6O	DTXCID01235	DTXSID60179347	24599-57-3	2,4-Cyclohexadienone	C6H6O	94.04186481	4

- MS-ready structures and substance mappings will be available as Open Data

Work in Progress: RAPIDTOX Predicted Hazards



Work in Progress: RAPIDTOX
Structural Analogs for Read-Across

EPA United States Environmental Protection Agency

Home | Advanced Search | CompTox Dashboard | Search CompTox Dashboard

RapidTox | Submit Comment | Share | Copy | Aa | Aa | Aa

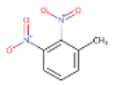
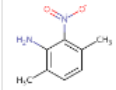
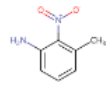
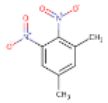
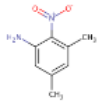
Chemical Properties | External Links | Synonyms | Product Composition | Bioassays | Analytical | Toxicity Values | Similar Molecules | Exposure | Literature | Predicted Hazards | Comments

Showing 20 of 44 Similar Molecules

Searched with a similarity threshold of 0.9

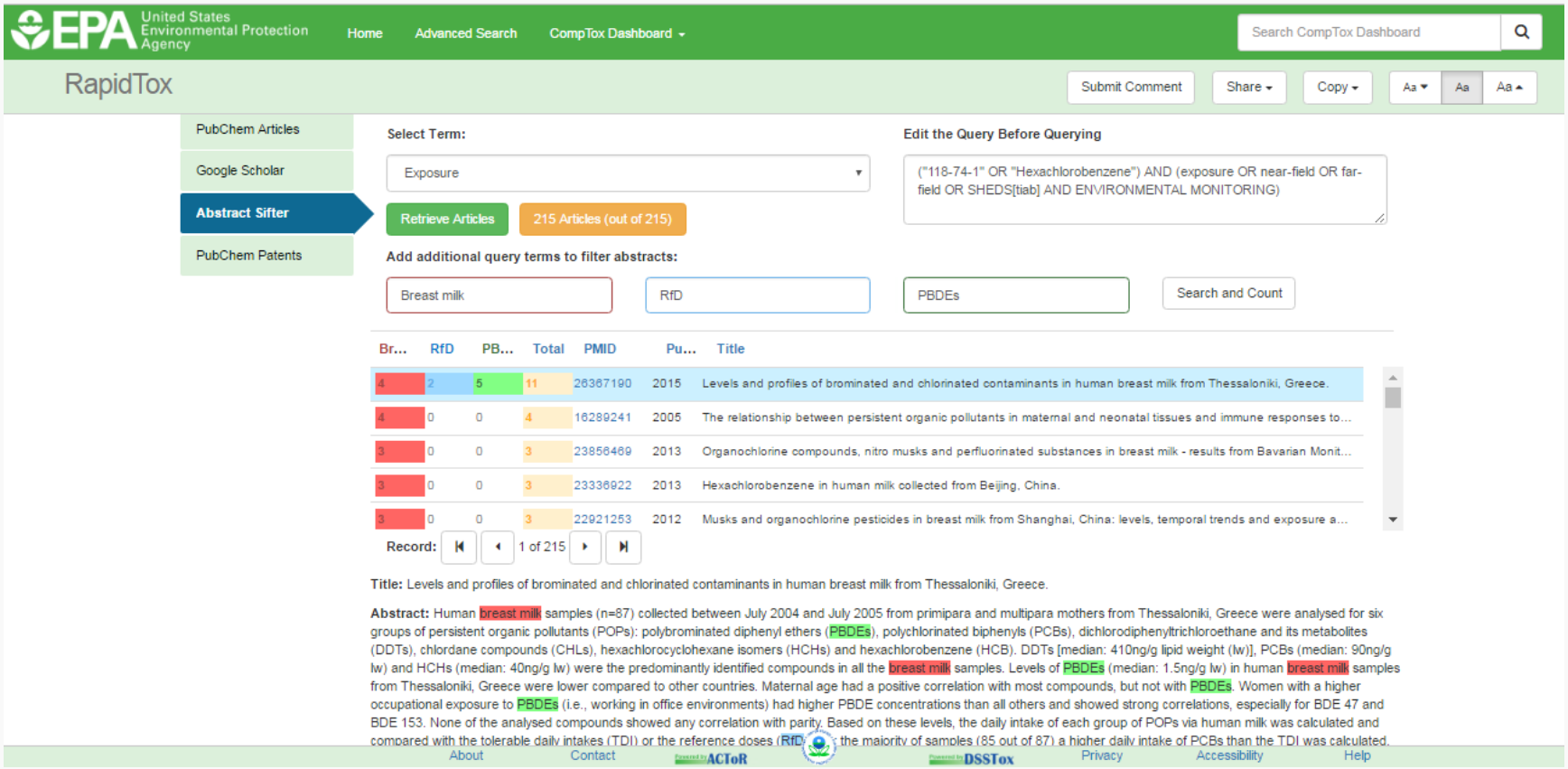
Download as: TSV | Excel

Properties | Chemicals

						
LogP: Octanol-Water	Experimental	-	-	-	-	-
	Predicted	2.00	2.50	1.95	2.16	1.72
LogWS: Water Solubility	Experimental	-	-	-	-	-
	Predicted	-2.00	-2.75	-2.29	-2.84	-2.70
Density	Experimental	-	-	-	-	-
	Predicted	1.41	-	-	-	-
Melting Point	Experimental	60.0	-	105	-	-
	Predicted	76.5	99.0	88.3	93.7	108
Boiling Point	Experimental	300	-	-	-	-
	Predicted	294	291	282	305	302

About | Contact | ACToR | DSSTox | Privacy | Accessibility | Help

Work in Progress: RAPIDTOX
PubMed "Abstract Sifting"



Future Work

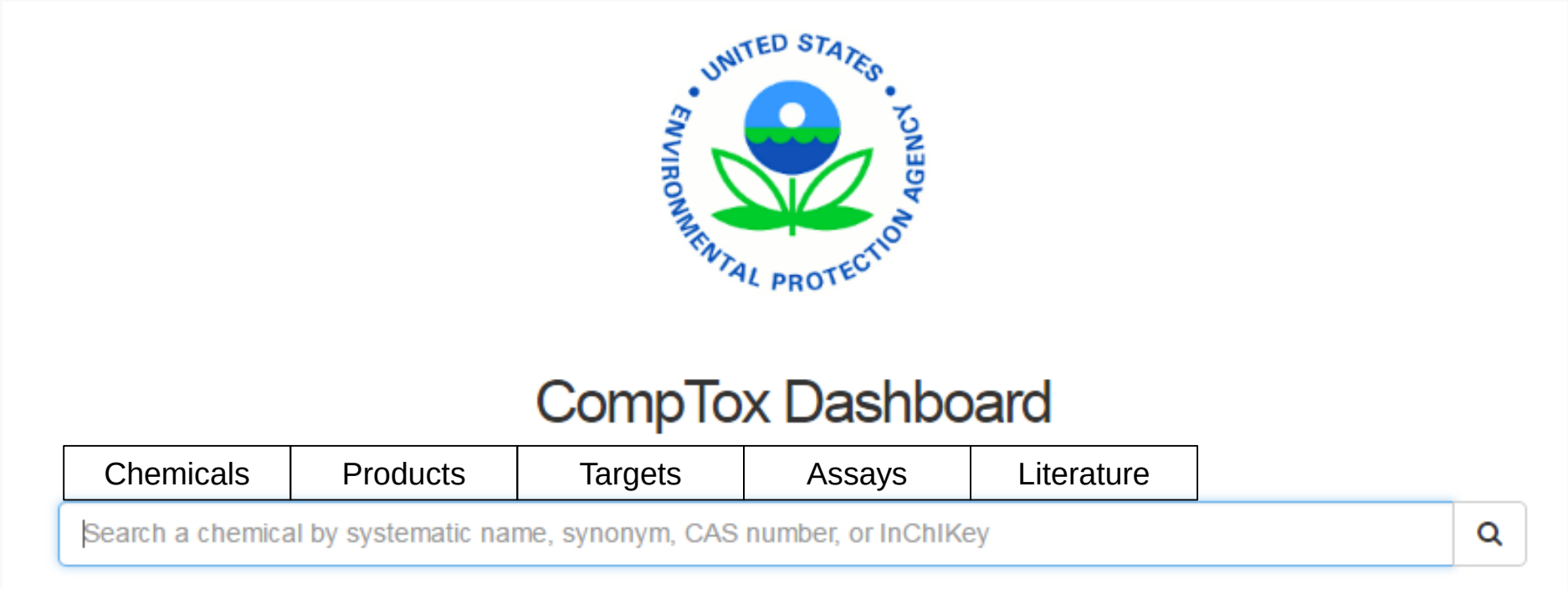
- Real-time Predictions
- API/Web Services in development
- Deeper integration to agency databases

- Focus on the challenge “Identify **chemical exposures** that may disrupt biological processes and cause adverse outcomes”

Conclusion

- NCCT: delivering data, algorithms, models and software tools for almost a decade.
- Developing a new flexible architecture to support multiple both internal *and* external apps
- A future concept for the CompTox dashboard…

Future Search Possibilities



Conclusion: We are focused on Integrating Research Efforts

HT research needs:

ToxCast:
Prioritizations for:
1) Parent chemicals
2) Mixtures
3) Metabolites

ExpoCast:
Measurement data for:
1) Model inputs
2) Model evaluation
3) Model refinement

NTA research needs:

1) Chemical databases
2) Chemical standards
3) Exposure forecasts
4) Bioactivity data
5) Functional use data
6) Prioritization methods
7) HT workflows

Acknowledgements

EPA-RTP

Chris Grulke
Kamel Mansouri
Ann Richard
Nancy Baker
Katharine Phillips
Kathie Dionisio
Jennifer Smith
Jeff Edwards
John Wambaugh
Grace Patlewicz
Imran Shah
Jon Sobus
Richard Judson