

CompTox Chemicals Dashboard v3 and invitroDB v3.1

<http://comptox.epa.gov/dashboard>

Katie Paul Friedman and Antony Williams
National Center for Computational Toxicology

Disclaimer: The views expressed in this presentation are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or products represent endorsement for use.

Overview of changes since August 2018 release

- Data
 - 875k chemicals total. An additional 110k chemical substances added
 - *InvitroDBv3.1* – including updated assay descriptions
 - ToxVal v7 data integrated – includes enormous curation effort
 - New OPERA predictions
- New User Interface elements
 - Reworked tables across the application
 - Reworked multiple chemical results page
 - Navigating concentration response plots for all AEIDs in *invitroDB_v3* data, not just the EDSP21 assays
 - Enhanced batch search capabilities

QC Notes for Chemicals

DETAILS

RELATED SUBSTANCES

SYNONYMS

LINKS

BIOACTIVITY

EXPOSURE

HAZARD

COMMENTS

PROPERTIES

LITERATURE

Toxaphene

8001-35-2 | DTXSID7021368

Searched by DSSTox Substance Id.

Wikipedia

Toxaphene was an insecticide used primarily for cotton in the southern United States during the late 1960s and 1970s. Toxaphene is a mixture of over 670 different chemicals and is produced by reacting chlorine gas with camphene. It can be most commonly found as a yellow to amber waxy solid. Toxaphene was banned in the United States in 1990 and was banned globally by the 2001 Stockholm Convention on Persistent Organic Pollutants. It is a very persistent chemical that can remain in ...
[Read more](#)

Presence in Lists

Record Information

Quality Control Notes

Complex, but reproducible mixture of at least 175 distinct C10-chloro compounds, having an approximate overall empirical formula of C10 H10 Cl8; the 2 most active components are a C10H10Cl8 compound and a C10H11Cl7 compound which had been elucidated as 2,2,5-endo,6-exo,8,9,10-heptachlorobomane. Produced by the chlorination of camphene to 67-69% chlorine by weight and made up of compds. of C10 H8 Cl10, C10 H18-n Cl n (mostly polychloroboranes) and C10 H16-n Cl n (polychloroboranes and/or polychlorotricyclenes) with n = 6 to 9 See Merck Index

Accessing QC Notes for Chemicals

Examples

- Toxaphene

Quality Control Notes

Complex, but reproducible mixture of at least 175 distinct C₁₀-chloro compounds, having an approximate overall empirical formula of C₁₀H₁₀Cl₈; the 2 most active components are a C₁₀H₁₀Cl₈ compound and a C₁₀H₁₁Cl₇ compound which had been elucidated as 2,2,5-endo,6-exo,8,9,10-heptachlorobornane. Produced by the chlorination of camphene to 67-69% chlorine by weight and made up of compds. of C₁₀H₈Cl₁₀, C₁₀H_{18-n}Cl_n (mostly polychlorobornanes) and C₁₀H_{16-n}Cl_n (polychlorobornanes and/or polychlorotricyclenes) with n = 6 to 9 See Merck Index

- Antimycin A

Quality Control Notes

mixture of antimycin A1; A2; A3 and A4

- Safflower Oil

Quality Control Notes

Extractives and their physically modified derivatives. It consists primarily of the glycerides of the fatty acid linoleic. (*Carthamus tinctorius*).

pKa experimental data added – no predictions

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

▶ EXPOSURE

▶ BIOACT

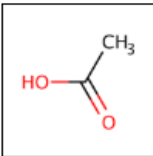
SIMILAR

GENRA

RELATE

SYNONYMS

LITERATURE



Acetic acid
64-19-7 | DTXSID5024394
Searched by Approved Name.

Property

pKa Acidic Apparent

Download Summary

Type	Average	Media
Experimental	4.70	-
Predicted	-	-

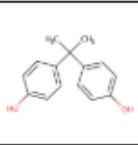
Experimental

Result	Experimental Data
4.70	

This pKa data was from the DataWarrior application (<http://www.openmolecules.org/>). A file named "pKaInWater.dwar" containing the pKa data is included in the DataWarrior download and contains experimentally-measured pKa values in water for 7912 chemicals along with SMILES strings. The providers of the original file collected and compiled pKa values representing different protonation states. Unfortunately, there are no literature references to support the pKa values. Most of these values are given as an average of multiple experimental values.

A new “Structure Zoom”

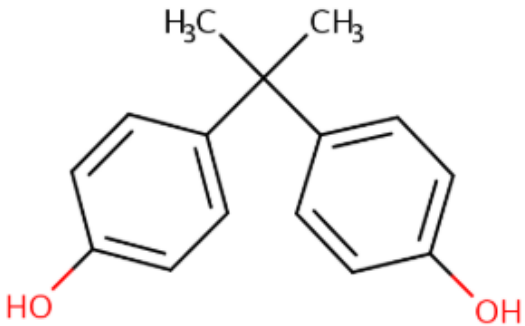

- On-click hover all over the dashboard as well as structure thumbnail



Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.





Wikipedia


Bisphenol A (BPA) is an organic synthetic compound with the chemical formula ($C_{15}H_{16}O_2$, 83 °C). BPA is a starting material for the synthesis of plastics, primarily certain polycarbonates.


[Read more](#)


Intrinsic Properties


 **Molecular Formula:** $C_{15}H_{16}O_2$

 **Mol File**

 **Find All Chemicals**

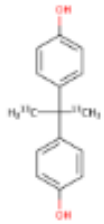


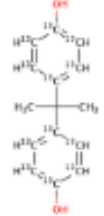


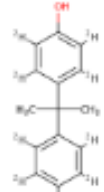


 **Average Mass:** 228.291 g/mol

 **Isotope Mass Distribution**

 **Monoisotopic Mass:** 228.11503 g/mol

Structural Identifiers

Linked Substances

Structure	DTXSID	Preferred Name
 	DTXSID30747173 	4,4'-[(1,3- $^{13}C_2$)Propane-2,2-diyl]diphenol
 	DTXSID10675703 	4,4'-(Propane-2,2-diyl)di($^{13}C_6$)phenol
 	DTXSID40662328 	4,4'-(Propane-2,2-diyl)di(2H_4)phenol

Reworked multiple results page

Searched with a similarity threshold of 0.8

390 chemicals

Select all

Download

Send to Batch Search

Similarity

DTXSID

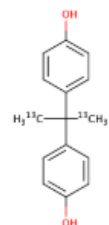
CASRN

TOXCAST

Similarity

Hide chemicals that are:

Filter by Name or CASRN

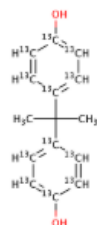


4,4'-[(1,3-¹³C₂)Propane-2,2-diyl]diphenol

DTXSID: DTXSID30747173

CASRN: 263261-64-9

TOXCAST: -

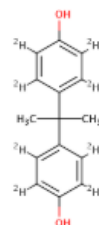


4,4'-(Propane-2,2-diyl)di(¹³C₆)phenol

DTXSID: DTXSID10675703

CASRN: 263261-65-0

TOXCAST: -

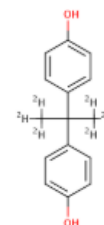


4,4'-(Propane-2,2-diyl)di(²H₄)phenol

DTXSID: DTXSID40662328

CASRN: 92739-58-7

TOXCAST: -

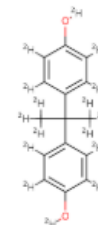


4,4'-[(²H₆)Propane-2,2-diyl]diphenol

DTXSID: DTXSID00584370

CASRN: 86588-58-1

TOXCAST: -



4,4'-[(²H₆)Propane-2,2-diyl]di(²H₅)phenol

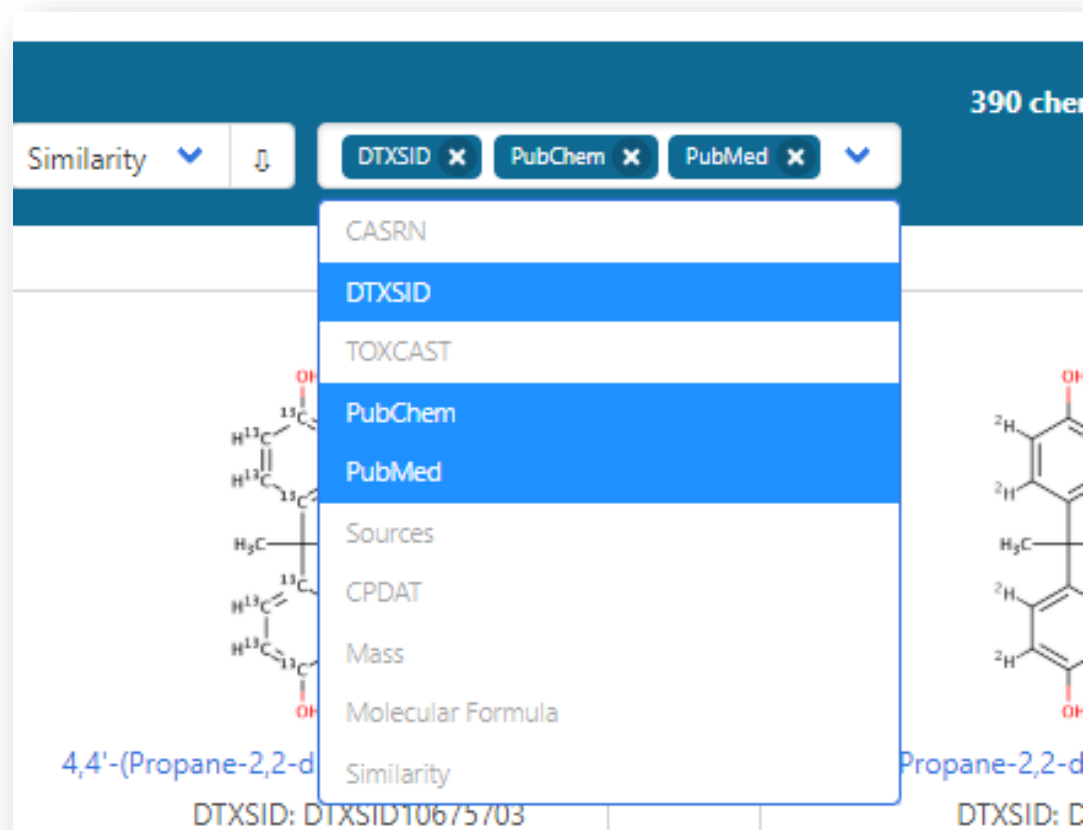
DTXSID: DTXSID40583721

CASRN: 96210-87-6

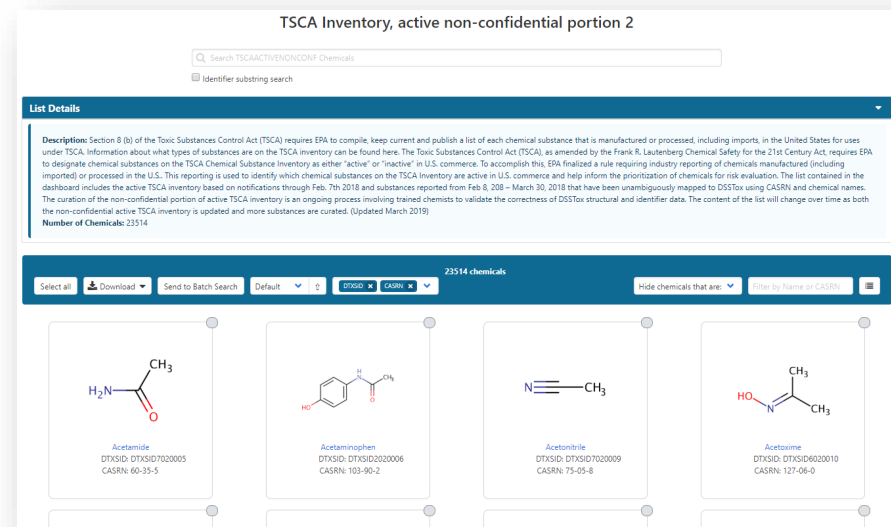
TOXCAST: -

Reworked multiple results page

- Use Ctrl to select multiple display
- Improved visual cues for loading large lists of chemicals












- Loading of Large lists RETAINS ordering



Batch Search













- New Search input - DTXCID

Select Input Type(s)








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

- New Search Outputs

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 
- ☒ QC Notes 
Clicking on QC Notes will include manual annotation notes added to a record during the chemical registration process.
- ☐ Include links to Action Reports 

Enhanced Data Sheets

- ☐ MetFrag Input File (Beta) 
- ☐ ToxPrint single fingerprints 
- ☐ Abstract Sifter Input File (Beta) 
- ☐ Synonyms and Identifiers 
- ☐ Related Substance relationships 
- ☐ ToxPrint fingerprints (ChemoTyper format - CSV/TSV only) 
- ☒ Associated ToxCast Assays 

Related Substance Relationships

Enhanced Data Sheets

☐ MetFrag Input File (Beta)

☐ ToxPrint single fingerprints

☐ Abstract Sifter Input File (Beta)

☐ Synonyms and Identifiers

☒ Related Substance relationships

☐ ToxPrint fingerprints (ChemoTyper f

☐ Associated ToxCast Assays

☐ EPA: National-Scale Air

☐ EPA: PPRTV Chemical R

☐ EPA: Provisional Adviso

☐ EPA: Safer Choice Chem

Selecting this checkbox provides a separate Excel worksheet containing the relationship between two chemicals. The output file includes the DTXSIDs and names/CASRN between the input list and the related chemical. Relationships include, for example, polymer, components, salt form, transformation product and other relationships.

	A	B	C	D	E	F	G	H
1	INPUT	DTXSID	PREFERRED_NAME	HAS_RELATIONSHIP_WITH	RELATED_DTXSID	RELATED_PREFERRED_NAME	RELATED_CASRN	
2	xlenes	DTXSID2021446	Xlenes	Transformation Product	DTXSID40176394	N-Benzoylalanine	2198-64-3	
3	xlenes	DTXSID2021446	Xlenes	Component	DTXSID6026298	m-Xylene	108-38-3	
4	xlenes	DTXSID2021446	Xlenes	Component	DTXSID3021807	o-Xylene	95-47-6	
5	xlenes	DTXSID2021446	Xlenes	Component	DTXSID2021868	p-Xylene	106-42-3	
6	xlenes	DTXSID2021446	Xlenes	Predecessor: Component	DTXSID9021421	Xlenes; defined mixture 1	NOCAS_21421	
7	xlenes	DTXSID2021446	Xlenes	Predecessor: Component	DTXSID7021447	Xlenes; defined mixture 2	NOCAS_21447	
8	xlenes	DTXSID2021446	Xlenes	Predecessor: Component	DTXSID30891529	Total Petroleum Hydrocarbons (TPH)	NOCAS_891529	
9	xlenes	DTXSID2021446	Xlenes	Markush Child	DTXSID3021807	o-Xylene	95-47-6	
10	xlenes	DTXSID2021446	Xlenes	Markush Child	DTXSID6026298	m-Xylene	108-38-3	
11	xlenes	DTXSID2021446	Xlenes	Markush Child	DTXSID2021868	p-Xylene	106-42-3	

Bioactivity Data

▼ BIOACTIVITY
TOXCAST: SUMMARY
EDSP21
TOXCAST/TOX21
PUBCHEM
TOXCAST: MODELS

- Summary data now has “enhanced tables”
- EDSP21 subset of assays has grown
- Toxcast/Tox21 “all data” has been integrated
- PubChem data widget – no change
- Subset of ToxCast “Models” – extended to include “COMPARA” data

Tables Reworked – Column Selection

- Ability to select columns to show added for tables –Bioactivity most important – Pick your own preferred display

185 active of 839 assays

Download Columns 10

Search query Show Inactive Show Background

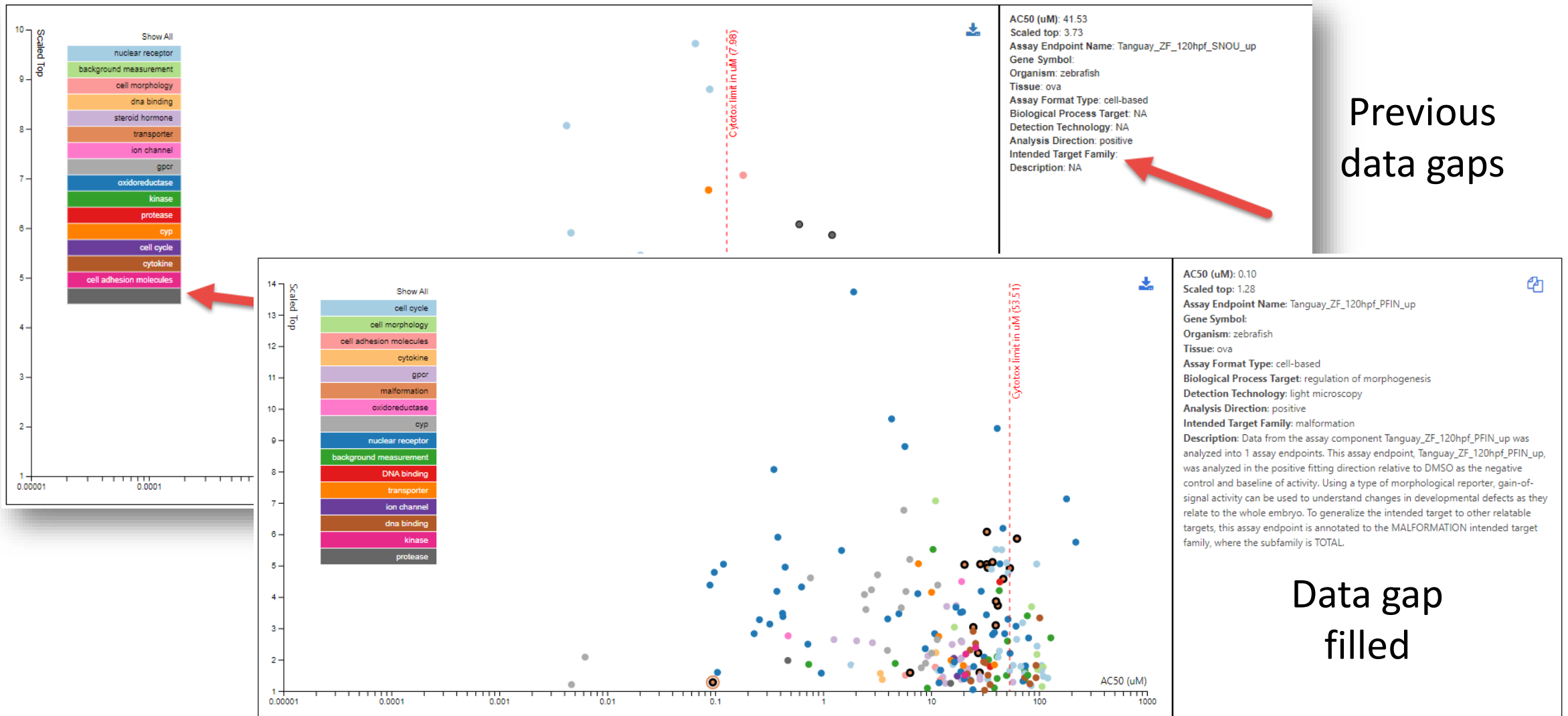
Name	Modal	Description	SeqaPASS	Gene Name	AOP	Event	Hit Call	Top	AC50	logAC50	MaxMed	Cutoff	ModIAcc	Intended Target Family
ACEA_ER_80hr		2	NP_000116.2	ESR1	200	1181	ACTIVE	112	0.373	-0.428	113	26.9	-0.686	nuclear receptor
APR_HepG2_Cell		-	-	-	-	-	ACTIVE	1.20	106	2.02	1.20	0.663	2.04	cell cycle
APR_HepG2_Mitc		-	-	-	-	-	ACTIVE	0.874	109	2.04	0.867	0.496	2.05	cell morphology
APR_HepG2_Mitc		-	-	-	-	-	ACTIVE	5.92	11.0	1.04	6.45	0.838	0.813	cell morphology
APR_HepG2_Oxic		-	-	-	-	-	ACTIVE	1.20	110	2.04	1.19	0.819	2.08	cell cycle
APR_HepG2_Cell		-	-	-	-	-	ACTIVE	4.49	95.2	1.98	4.43	0.889	1.75	cell cycle
APR_HepG2_Mitc		-	-	-	-	-	ACTIVE	2.71	85.3	1.93	2.26	0.733	1.70	cell morphology
APR_HepG2_Mitc		-	-	-	-	-	ACTIVE	1.66	84.7	1.93	1.44	1.42	2.29	cell cycle
APR_HepG2_Oxic		-	-	-	-	-	ACTIVE	1.80	106	2.02	1.60	1.10	2.08	cell cycle
ATG_Ahr_CIS_up		-	NP_001612.1	AHR	150	18	ACTIVE	1.31	23.4	1.37	1.28	0.994	1.56	dna binding

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

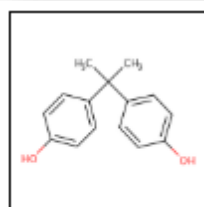
Showing 1 to 10 of 185 records

Discover. Connect. Ask.

Assay Annotations Filled a Lot of Gaps!



Toxcast: Models – COMPARA added




Bisphenol A




80-05-7 | DTXSID7020182

Searched by Expert Validated Synonym.

ToxCast: Models

ToxCast Model Predictions

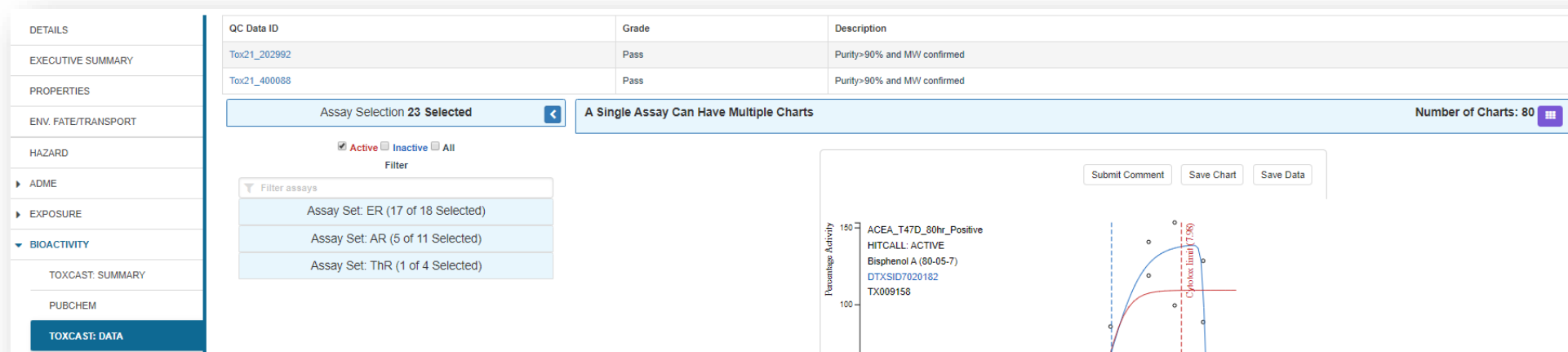
 Download ToxCast Model Predictions ▼

Model	Receptor	Agonist	Antagonist	Binding
 ToxCast Pathway Model (AUC)				-
 CoMPARA is a larger scale collaboration between 35 international groups, using QSAR models to predict androgen receptor activity using a common training set of 1746 compounds provided by U.S. EPA. A key result is a consensus model of AR agonist and antagonist activity that is run against the DSSTox chemical library. These results are intended to be used in prioritization for compounds for follow-up testing. More details about the project are available on ResearchGate .				-
 CERAPP Potency Level (Consensus)				Active
				Active (Weak)
				Active (Weak)

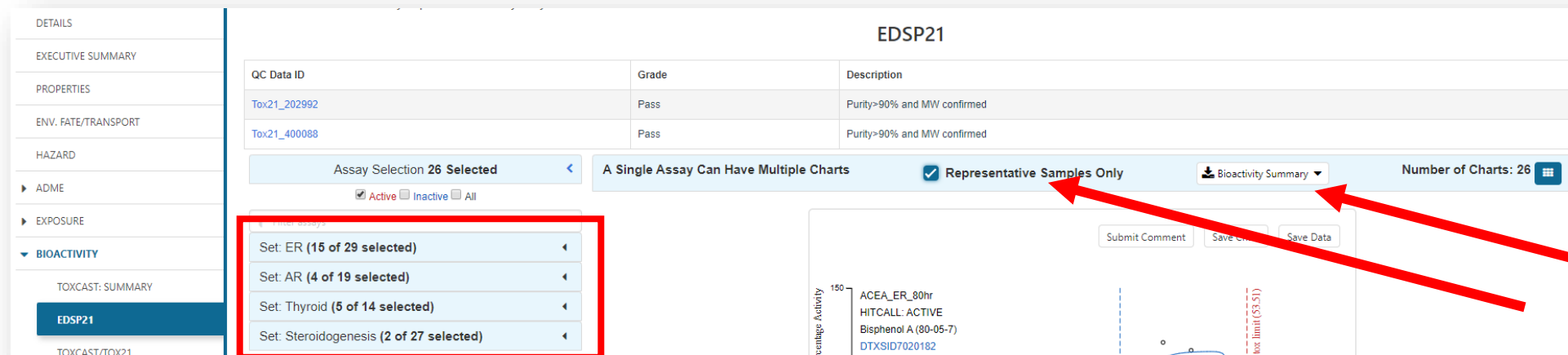
“EDSP Subset”

- New assays added – expanded all subsets. New set of steroidogenesis assays – including CEETOX data

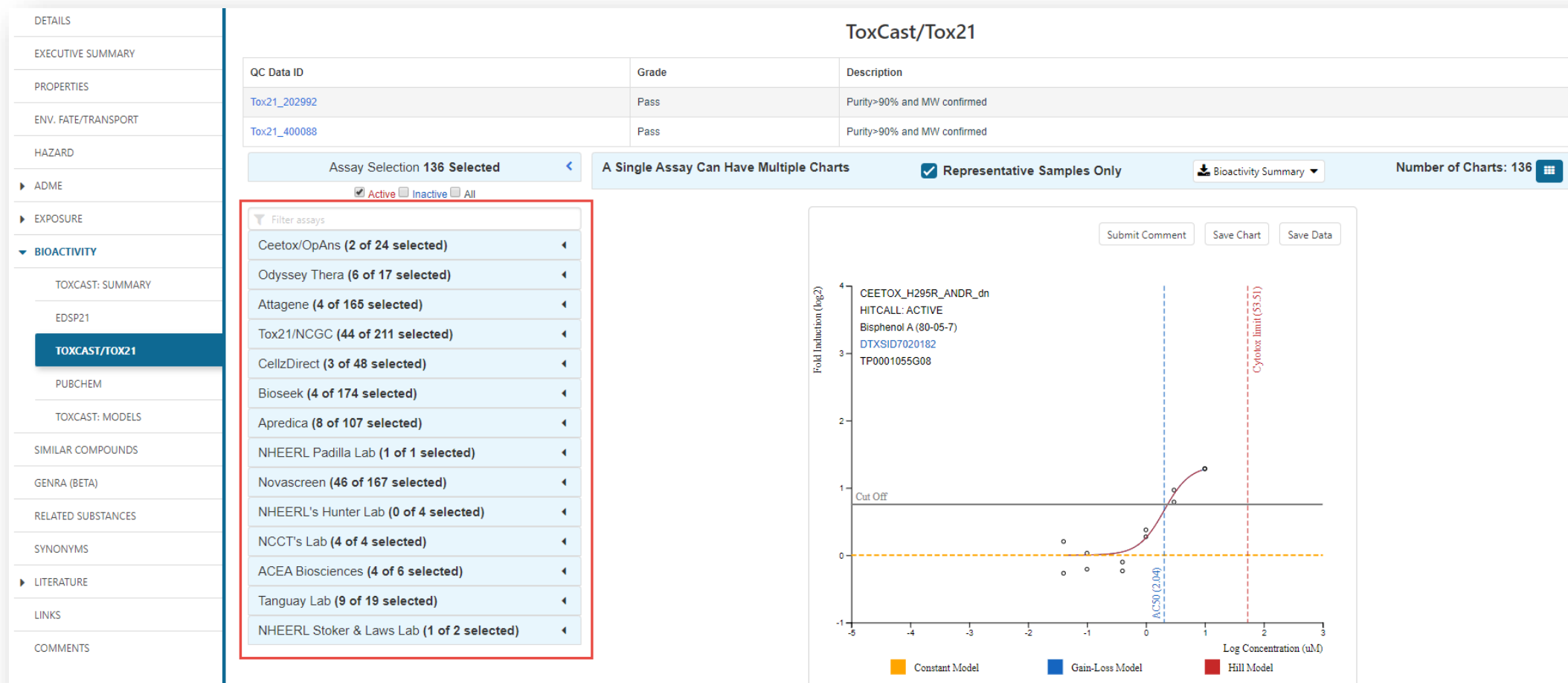
- Previous



- Update



ToxCast/Tox21 Data – All data from invitroDBv3



Filtering – Gene annotation added

Filter assays

Ceetox/OpAns (2 of 24 selected)

Odyssey Thera (6 of 17 selected)

Attagene (4 of 165 selected)

<input type="checkbox"/>	ATG_PBREM_CIS_up	NR1I3	
<input type="checkbox"/>	ATG_E2F_CIS_dn	E2F1	
<input type="checkbox"/>	ATG_HSE_CIS_dn	HSF1	
<input type="checkbox"/>	ATG_EGR_CIS_dn	EGR1	
<input type="checkbox"/>	ATG_ISRE_CIS_dn	IRF1	
<input type="checkbox"/>	ATG_GR_TRANS_dn	NR3C1	
<input type="checkbox"/>	ATG_p53_CIS_up	TP53	
<input type="checkbox"/>	ATG_Oct_MLP_CIS_dn	POU2F1	
<input type="checkbox"/>	ATG_Ets_CIS_up	ETS1	
<input type="checkbox"/>	ATG_EGR_CIS_up	EGR1	
<input type="checkbox"/>	ATG_RARb_TRANS_dn	RARB	
<input type="checkbox"/>	ATG_TGFB_CIS_up	TGFB1	
<input type="checkbox"/>	ATG_PPARG_TRANS_dn	PPARG	
<input type="checkbox"/>	ATG_M_10_CIS_up		
<input type="checkbox"/>	ATG_PXRE_CIS_up	NR1I2	

PPARG

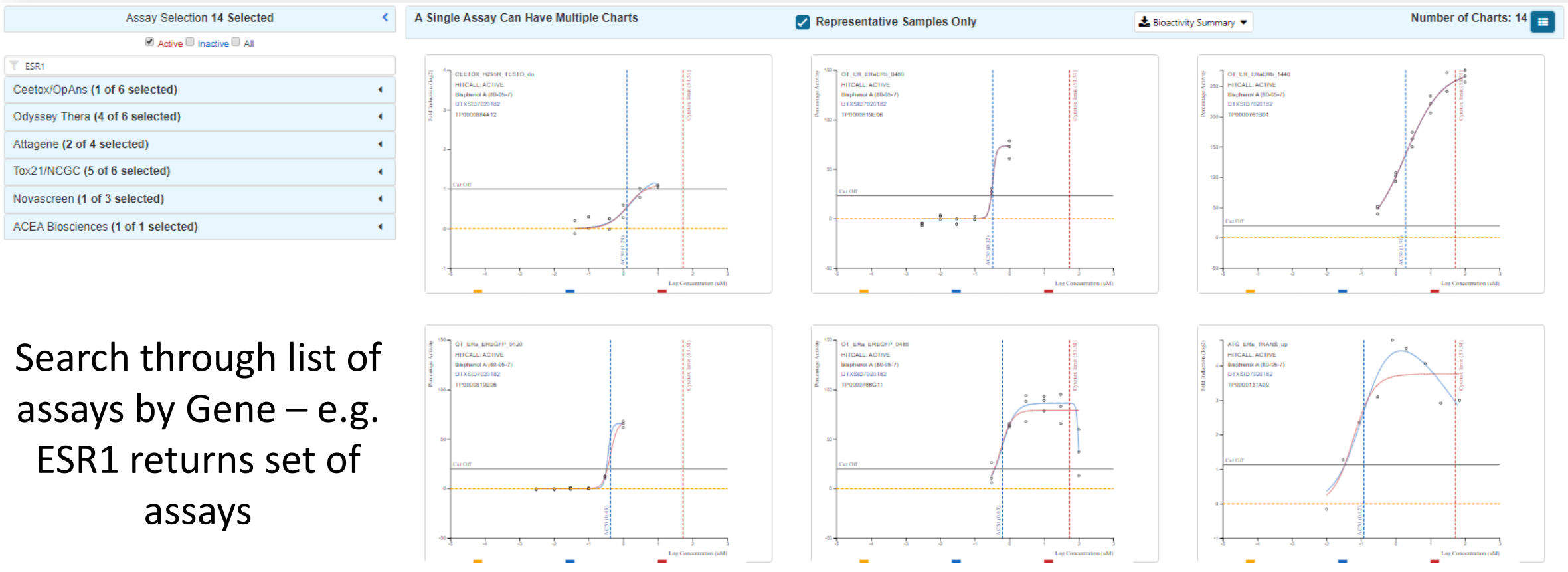
Attagene (0 of 4 selected)

<input type="checkbox"/>	ATG_PPARG_TRANS_dn	PPARG	
<input type="checkbox"/>	ATG_PPARG_CIS_up	3 Genes	
<input type="checkbox"/>	ATG_PPARG_TRANS_up	PPARG	
<input type="checkbox"/>	ATG_PPARG_CIS_dn	3 Genes	

Novascreen (1 of 1 selected)

<input checked="" type="checkbox"/>	NVS_NR_hPPARG	PPARG	
-------------------------------------	---------------	-------	--

Filtering



Lists of Chemicals/Lists of Assays

- Reworked Chemical List page – lots of lists added including segregation
 - LIST: Algal Toxins, Amino Acids, Bisphenol Compounds, PAHs, Synthetic Cannabinoids and Psychoactives, Vitamins, PCBs, PBDEs, Hair Dyes
 - WIKILIST: Additives in Cigarettes, Extremely Hazardous Substances
 - EPA: PALs, HPV list, Chemical Contaminants, PPRTV Reports etc, Pesticides Chemical Search
- Helps cluster in the batch search and as a query on the lists page (see figure)
- invitroDbv3 assays added to assay list



Lists of Chemicals

- Download the “list of lists” as Excel or TSV
- Subset of lists from query – “what are all PFAS lists?”

PFAS

Copy Filtered Lists URL

http://comptox.epa.gov/dashboard/chemical_lists/?search=PFAS

Select List

Download

Columns

10

Search query

Copy page URL

List Acronym	List Name	Last Updated	Number of Chemicals	List Description
40CFR355	40CFR355 Extremely Hazardous Substance List and Threshold Planning Quantities	2018-01-05	354	Extremely Hazardous Substance List and Threshold Planning Quantities; Emergency Planning and Release Notification Requirements; Final Rule. (52 FR 13378)
ACSREAG	LIST: ACS Reagent Chemicals	2017-04-14	405	The ACS Committee on Analytical Reagents sets purity specifications for almost 500 reagent chemicals and over 500 standard-grade reference materials.
AEGLVALUES	AEGLs: Acute Exposure Guideline Levels	2018-04-20	174	Acute exposure guideline levels (AEGLs) describe the human health effects from once-in-a-lifetime, or rare, exposure to airborne chemicals.
ALGALTOX	LIST: Algal Toxins	2018-05-04	54	A list of Algal Toxins of potential interest
AMINOACIDS	LIST: Amino Acids	2019-02-04	0	Amino acids are organic compounds containing amine (-NH ₂) and carboxyl (-COOH) functional groups, along with a side chain (R group) specific to each amino acid.
APCRA_PRO	LIST: APCRA Chemicals for Prospective Analysis	2018-02-14	204	The APCRA prospective case study list of approximately 200 chemicals as of January 2018, developed by ECHA in consultation with EPA and other partners

List of Assays

- Download list of all assays (Excel or TSV)
- Filter by Vendor or Multiple Vendors
- Subset of lists from query based on substring search e.g.

http://comptox.epa.gov/dashboard/assay_endpoints/?search=ESR1

Assay List

APR X NVS X CEETOX X NHEERL_PADILLA X NHEERL_HUNTER X

Search query

Copy page URL


		Gene Symbols
assay con ellCycleAr if activity. stand the a primary eneralize t et family, v	cellArrest_1hr was analyzed into 2 assay endpoints. This assay endpoint, the negative fitting direction relative to DMSO as the negative control reporter, measures of all nuclear dna for loss-of-signal activity can be referred as they relate to the gene . Furthermore, this assay endpoint can be has produced multiple assay endpoints where this one serves a signaling relatable targets, this assay endpoint is annotated to the "cell cycle" feration".	
assay con ellCycleAr ivity. Using the signa y readout, the intend where the	cellArrest_1hr was analyzed into 2 assay endpoints. This assay endpoint, the positive fitting direction relative to DMSO as the negative control and orter, measures of all nuclear dna for gain-of-signal activity can be used they relate to the gene . Furthermore, this assay endpoint can be referred luced multiple assay endpoints where this one serves a signaling function. targets, this assay endpoint is annotated to the "cell cycle" intended	
assay con ellLoss_Th	ss_1hr was analyzed into 2 assay endpoints. This assay endpoint, gative fitting direction relative to DMSO as the negative control and	

For a Specific Assay List

- All the advantages of the new “Multiple Results Page” plus...

Assay Endpoint Name: ACEA_ER_80hr

Assay Details

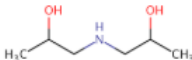
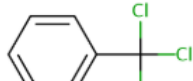

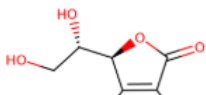
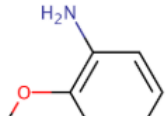
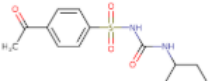
Assay Endpoint Name: ACEA_ER_80hr 

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

Histograms

425 of 3031 chemicals visible

Select all Download Send to Batch Search Default DTXSID CASRN Inactive Filter by Name or CASRN



Specific Assay List

- Reworked assay table – more details available including AOP Wiki link

All Chemicals in Assay Endpoint: [ACEA_ER_80hr](#)

[Annotations](#) [Citations](#) [tcpl Processing](#) [Reagents](#) [AOPs](#) [Excel](#)

	Assay Run Type	Level Applied	Method Name	Description
1	MULTI	2	none	apply no level 2 method
2	MULTI	3	pval.apid.medpcbyconc.max	plate-wise median response of positive control (max)
3	MULTI	3	resp.pc	response percent activity
4	MULTI	3	bval.apid.nwllslowconc.med	Take the median cval of the n wells and the first two concentrations, by apid
5	MULTI	3	resp.shiftneg.3bmad	Make values below baseline zero.
6	MULTI	4	bmad.aeid.lowconc.twells	bmad based on two lowest concentration of treatment wells
7	MULTI	5	bmad3	Add a cutoff value of 3*bmad.
8	MULTI	5	pc20	Add a cutoff value of 20.
9	MULTI	6	singlept.hit.high	Look for single point hits with activity only at the highest conc tested
10	MULTI	6	singlept.hit.mid	Look for single point hits with activity not at highest conc tested
11	MULTI	6	multipt.neg	Look for inactives with multiple medians above baseline
12	MULTI	6	noise	Look for noisy curves, relative to the assay
13	MULTI	6	border.hit	Look for actives with borderline activity
14	MULTI	6	border.miss	Look for inactives with borderline activity
15	MULTI	6	modlga.lowconc	AC50 less than lowest concentration tested
16	MULTI	6	gnls.lowconc	Look for low concentration gnls winners
17	MULTI	6	overfit.hit	Flag hit-calls that would get changed after doing the small N correction to the aic values.
18	MULTI	6	efficacy.50	Flag hit-calls with efficacy values less than 50% -- intended for biochemical assays.

All Chemicals in Assay Endpoint: [ACEA_ER_80hr](#)

[Annotations](#) [Citations](#) [tcpl Processing](#) [Reagents](#) [AOPs](#) [Excel](#)

AOP ID	AOP Title
200	Estrogen receptor activation leading to breast cancer

[AOP ID: 200](#)
AOP TITLE: Estrogen receptor activation leading to breast cancer
AUTHOR STATUS: Under development: Not open for comment. Do not cite
SAAOP STATUS: Under Development

Specific Assay List – Histogram summary view








Specific Assay List – Histogram summary view

- Display specific subset of data from histogram – switch to Table Mode



292 of 3031 chemicals visible

Structure	DTXSID	Preferred Name	# ToxCast Active	% ToxCast Active	Hit Call	Top	Scaled Top	AC50 (uM)	logAC50 (uM)
	DTXSID7020007 ToxCast™	Acetohexamide	4/376	1%	Active	29.3	1.09	4.49	0.652
	DTXSID00020105 ToxCast™	Sodium L-ascorbate	20/662	3%	Active	66.0	2.46	6.68	0.825
	DTXSID8020121 ToxCast™	Sodium azide	26/644	4%	Active	55.5	2.07	96.4	1.98
	DTXSID1020148 ToxCast™	Benzotrichloride	7/646	1%	Active	52.0	1.93	68.7	1.84
	DTXSID8020179 ToxCast™	Diisopropanolamine	9/399	2%	Active	82.9	3.08	8.25	0.916

The Underlying ToxCast data:
Release of invitrodb, version 3.1

Accessing data downloads via FTP

ToxCast Data and Information

• [About ToxCast](#)

- **ToxCast & Tox21 Summary Files for invitroDBv3.1.** Data for a single chemical endpoint pair for thousands of chemicals and assay endpoints for 20 variables such as the activity or hit call, activity concentrations, whether the chemical was tested in a specific assay, etc.
 - [Download ToxCast Summary Information](#)
 - [Download ReadMe](#)
- **ToxCast & Tox21 Data Spreadsheet for invitroDBv3.1.** A spreadsheet of EPA's analysis of the chemicals screened through ToxCast and the Tox21 collaboration which includes EPA's activity calls from the screening of over 1,800 chemicals.
 - [Download Data](#)
 - [Download ReadMe](#)
- **ToxCast Data Pipeline R Package (tcpl released with invitroDBv3.1).** The R computer programming package used to process and model all EPA ToxCast and Tox21 chemical screening data. The files include the R programming package as well as documents that provide overviews of the data analysis pipeline used and the R package. Users will need experience with R to use these files.
 - [Download Package](#)
 - [TCPL Overview](#)
- **ToxCast Database (invitroDBv3.1):** Database of EPA's analysis of chemicals screened through ToxCast assays, includes a MySQL database and the R package used to process the data.
 - [Download Database for MAC](#)
 - [Download Database for Windows](#)
 - [Download ReadMe](#)
- **ToxCast & Tox21 Concentration Response Plots for invitroDBv3.1.** Concentration response plots for all of the ToxCast and Tox21 assays.
 - [Download Concentration Response Plots](#)
- **ToxCast & Tox21 Chemicals Distributed Structure-Searchable Toxicity Database (DSSTox files) for invitroDBv3.** Chemical lists and information for 9,403 unique substances (DTXIDs) and DSSTox standard chemical fields (chemical name; CASRN; structure; etc.). These chemicals have been evaluated through the ToxCast and Tox21 high-throughput screening efforts.
 - [ToxCast Chemicals: Data Management and Quality Considerations Overview](#)
 - [Download ToxCast Chemical Information](#)
 - [Download ReadMe](#)
- **ToxCast & Tox21 High-Throughput Assay Documentation.** ToxCast high-throughput assay documentation including descriptions, target information, study design information and quality statistics.
 - [Assay Descriptions \(work in progress\)](#) - Descriptions and guidelines for ToxCast assays in format outlined by the OECD Guidance Document 211 for describing non-guideline in vitro test methods. The intent of GD 211 is to harmonize non-guideline, in vitro method descriptions to allow assessment of the relevance of the test method for biological responses of interest and the quality of the data produced. This document is organized by assay platform providers. You can also find descriptions for endocrine-related assays. It is a

<https://www.epa.gov/chemical-research/exploring-toxcast-data-downloadable-data>

Click here for SQL of invitrodb_v3.1

File(s) stored somewhere else

ftp://newftp.epa.gov/comptox/High_Throughput_Screening_Data/InvitroDB_V3.1/MySQL_Data/INVITRODB_V3_1_MYSQL.zip

Please Note: Unless noted as NOT stored on The United States Environmental Protection Agency's National Center for Computational Toxicology, and we can't guarantee its availability, quality, security or accept any liability.

Cite Share Embed + Collect (you need to log in first)

ToxCast Database (invitroDB)

Version 3 Dataset posted on 18.03.2019, 09:50 by EPA's National Center for Computational Toxicology

1357 views | 1 downloads | 0 citations

ToxCast high-throughput assay information including assay annotation user guide, assay target information, study design information and quality statistics on the assays.

EPA United States Environmental Protection Agency

Current snapshots, March 28, 2019

Accessing invitrodb_v3.1 download via FTP

← → ↻ ⓘ Not secure | ftp://newftp.epa.gov/comptox/High_Throughput_Screening_Data/InVitroDB_V3.1/

Apps Files - OneDrive toxrefdb - OneDrive Google Scholar Altmetric it!

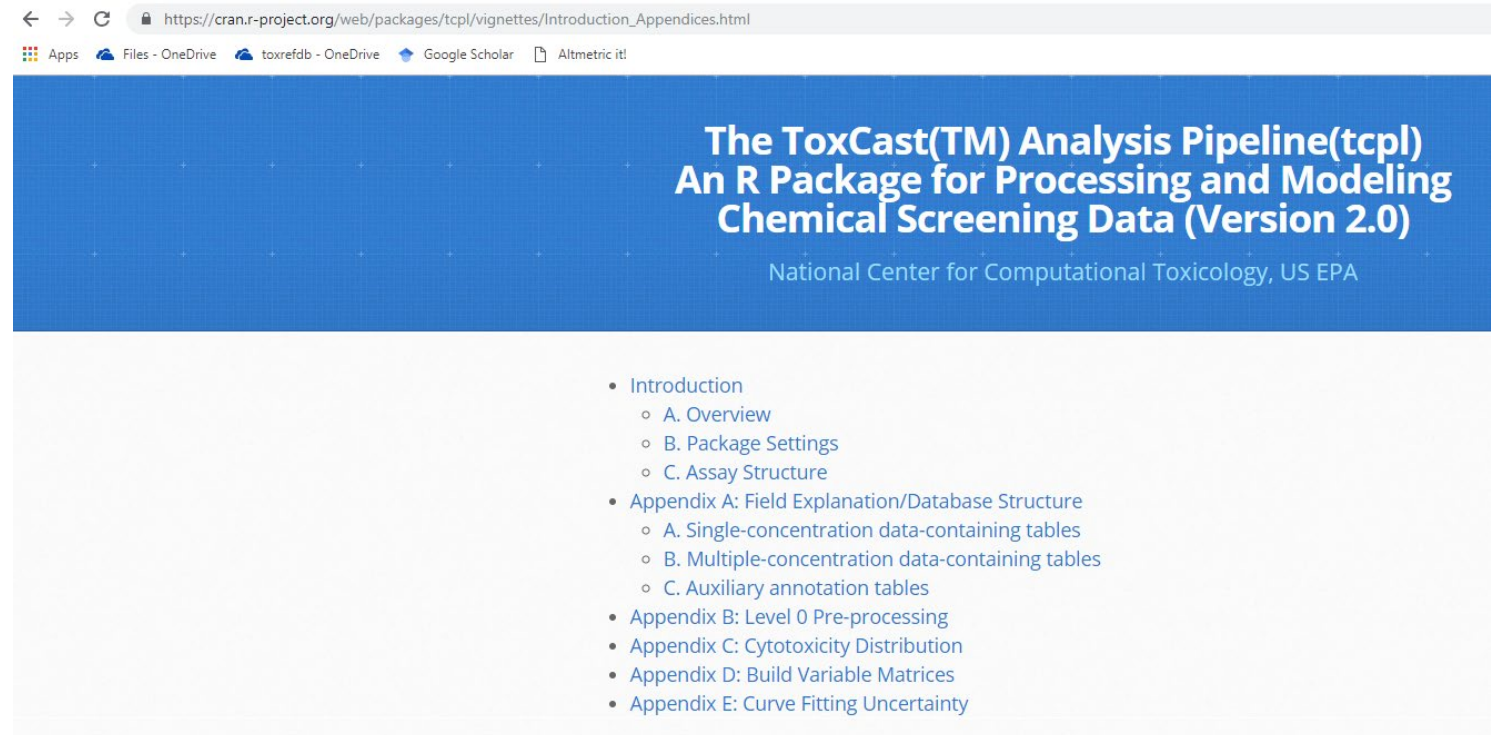
Index of /comptox/High_Throughput_Screening_Data/InVitroDB_V3.1/

📁 [parent directory]

Name	Size	Date Modified	
📁 Assay_Information/		3/8/19, 7:58:00 AM	← A searchable PDF of all assay descriptions available
📁 MySQL_Data/		3/27/19, 6:22:00 AM	← Invitrodb_v3.1 download as a .sql file
📁 Summary_Files/		3/8/19, 7:56:00 AM	← Familiar set of flat files, like version 2
📁 ToxCast_Concentration_Response/		3/8/19, 7:56:00 AM	← Plots
📁 ToxCast_Data_March_2019/		3/8/19, 7:58:00 AM	← Mc5 and mc6 export by vendor/source

Extracting information from invitrodb_v3.1

- Can use tcpl version 2.0.1 live now on CRAN <https://cran.r-project.org/web/packages/tcpl/>
- Refer to our rewritten vignettes that explain invitrodb and the ToxCast Pipeline, example:



The screenshot shows a web browser window displaying the vignette page for the ToxCast(TM) Analysis Pipeline (tcpl) R package. The browser's address bar shows the URL: https://cran.r-project.org/web/packages/tcpl/vignettes/Introduction_Appendices.html. The page has a blue header with the title "The ToxCast(TM) Analysis Pipeline(tcpl) An R Package for Processing and Modeling Chemical Screening Data (Version 2.0)" and the subtitle "National Center for Computational Toxicology, US EPA". Below the header, there is a table of contents listing the following sections:

- Introduction
 - A. Overview
 - B. Package Settings
 - C. Assay Structure
- Appendix A: Field Explanation/Database Structure
 - A. Single-concentration data-containing tables
 - B. Multiple-concentration data-containing tables
 - C. Auxiliary annotation tables
- Appendix B: Level 0 Pre-processing
- Appendix C: Cytotoxicity Distribution
- Appendix D: Build Variable Matrices
- Appendix E: Curve Fitting Uncertainty

Additional tables in invitrodb_v3.1

- Cytotox includes the “burst” information
- EDSP21-related models: model_generic_chemical_ar_scores, model_generic_chemical_cerapp_score, model_generic_chemical_compara_scores, model_generic_chemical_er_scores, model_generic_chemical_hth295r_scores
- A number of tables aimed at increased assay annotation:
 - Structured assay description tables (assay_component_descriptions, assay_component_endpoint_descriptions, assay_descriptions)
 - Assay lists (assay_list and assay_list_aeid) used in research to group assays (e.g., endocrine-relevant assays for the EDSP21 part of the dashboard)
 - Assay ontology tables for future dashboard searching (ontology and ontology_invitrodb) based on Bioassay Ontology codes
 - Gene/intended target tables updated
- Mc7: uncertainty information obtained using *toxboot* (<https://github.com/ericwatt/toxboot>)

Steroidogenesis model information is further described in Haggard et al. 2018 (PMID 29216406).

Mifepristone

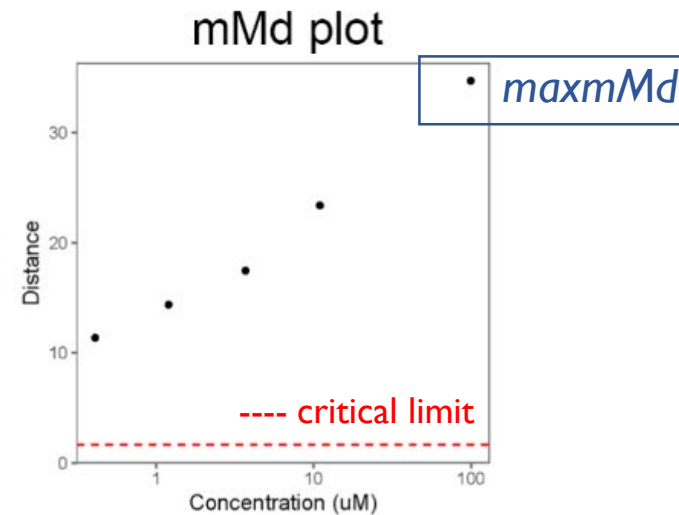
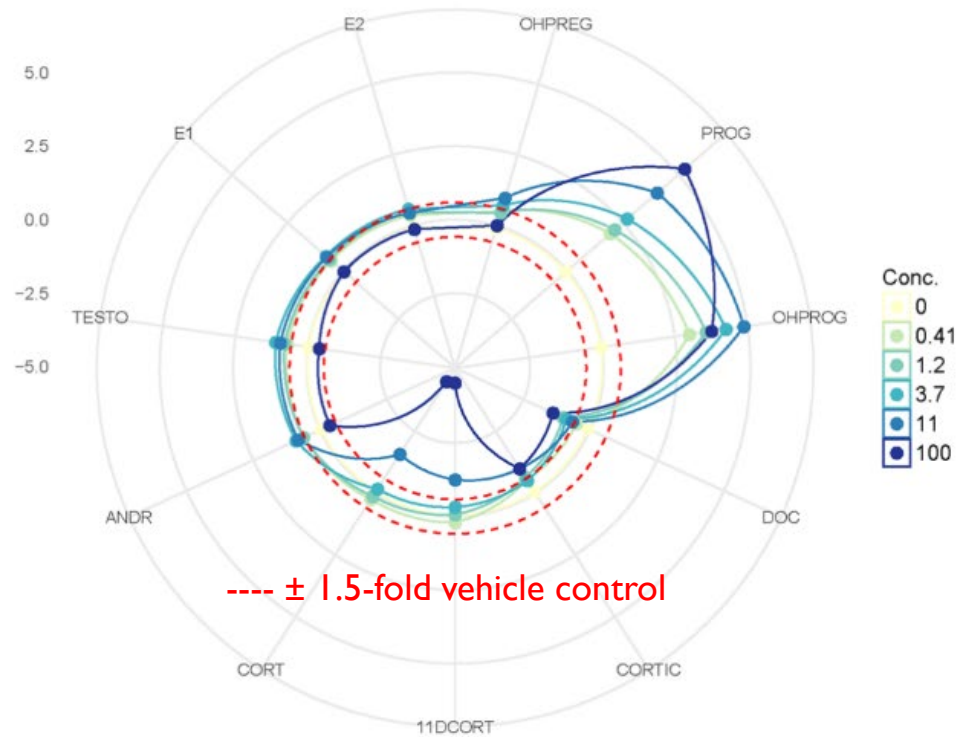


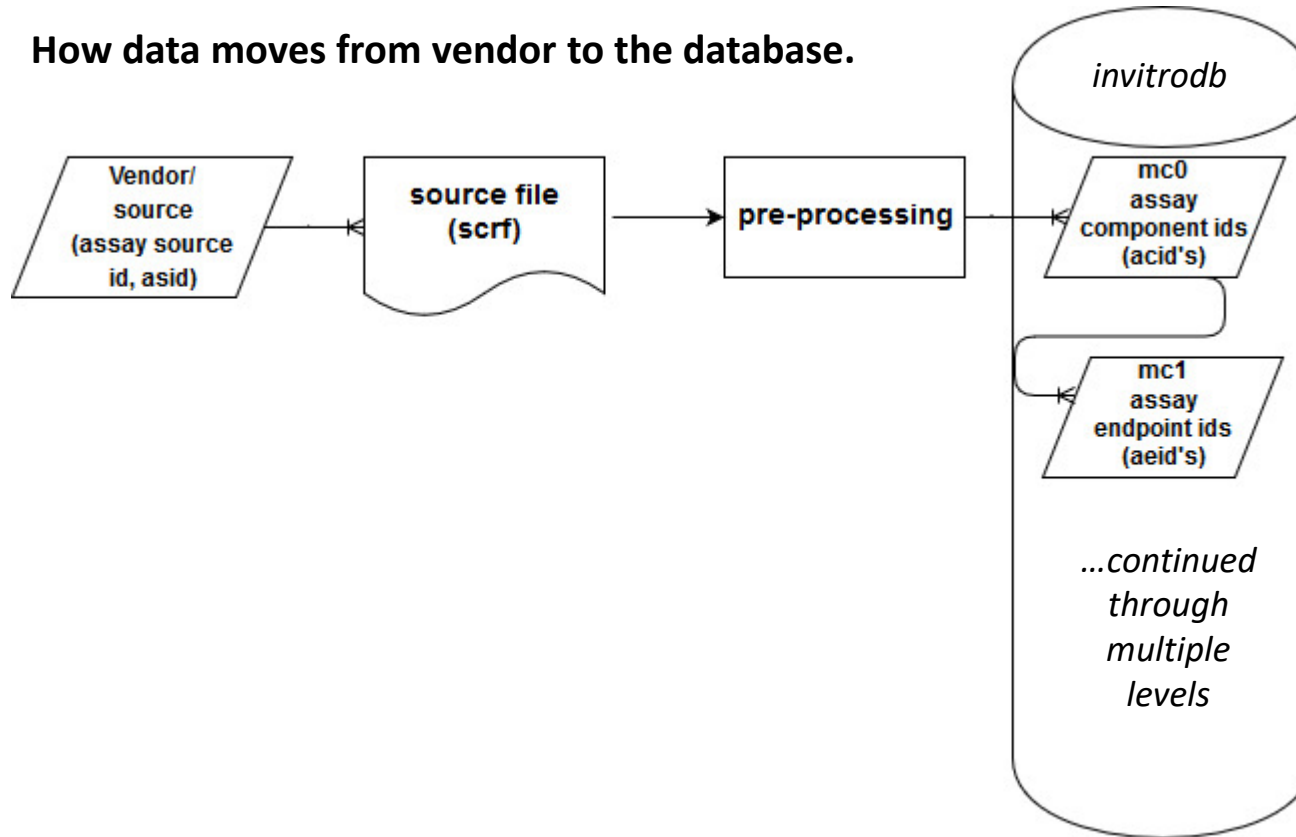
Figure 5, Haggard et al. (2018).

Mifepristone strongly modulated progestagens with significant effects on progesterone and OH-progesterone and moderate but non-significant trends on corticosteroids and androgens, resulting in a relatively high adjusted maxmMd of 33.

- Reduced an 11-dimensional question to a single dimension.
- Selection of the maxmMd appeared to provide a reproducible, quantitative approximation of the magnitude of effect on steroidogenesis.

Organization of data entering invitrodb

How data moves from vendor to the database.



Example: asid to acid to aeid.

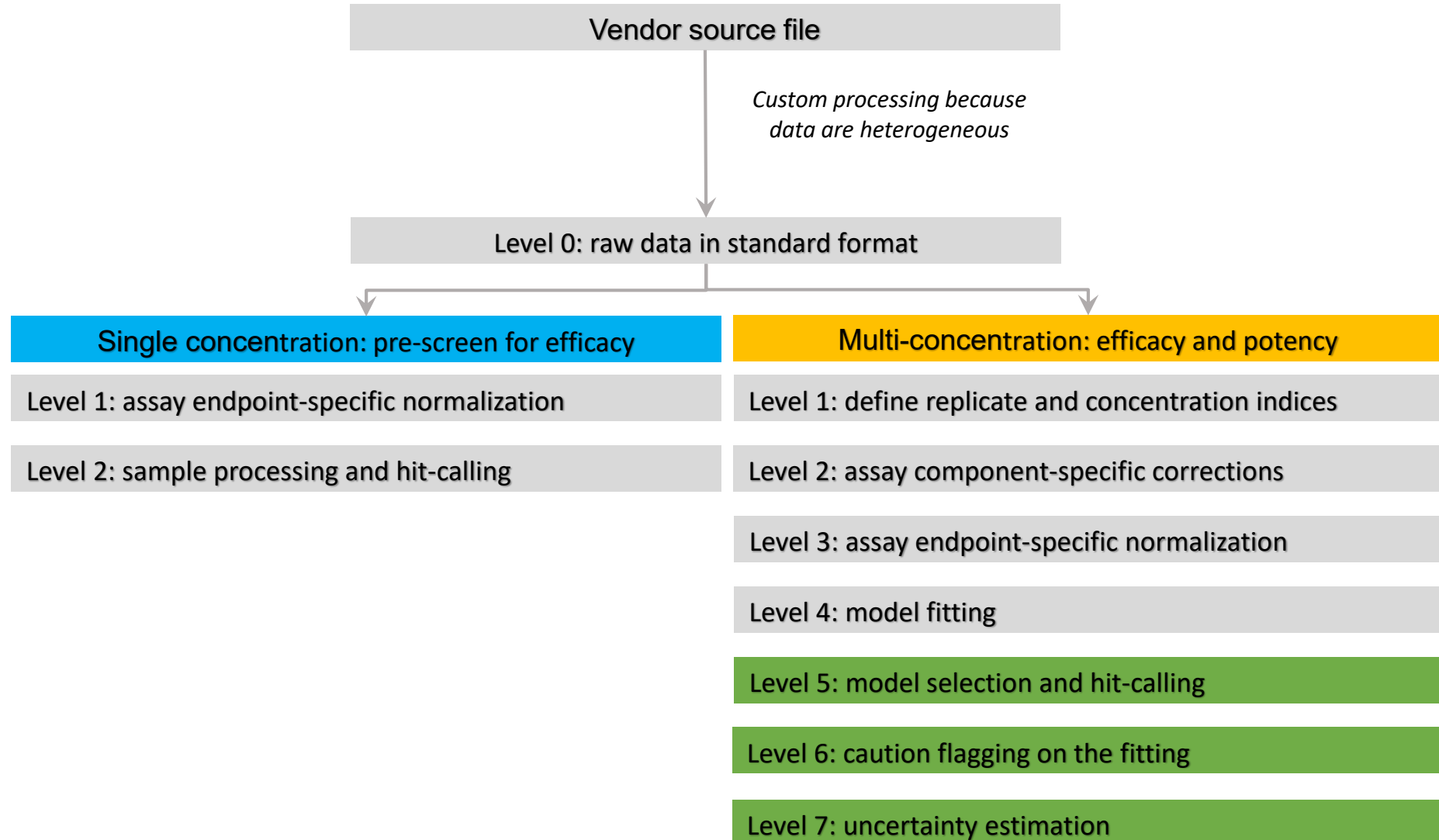
acid can be 1:1 or 1:many with aeid.

```

> tcplLoadAsid()
  asid      asnm
1:      1      ACEA
2:      2      APR
3:      3      ATG
4:      4      BSK
5:      5      NVS
6:      6      OT
7:      7      TOX21
8:      8      CEETOX
9:     11      CLD
10:     12 NHEERL_PADILLA
11:     17  NCCT_SIMMONS
12:     13  TANGUAY
> tcplLoadAcid(fld='asid', val=8)
  asid acid      acnm
1:      8  586 CEETOX_H295R_11DCORT
2:      8  587 CEETOX_H295R_OHPREG
3:      8  588 CEETOX_H295R_OHPROG
4:      8  589 CEETOX_H295R_ANDR
5:      8  591 CEETOX_H295R_CORTISOL
6:      8  593 CEETOX_H295R_DOC
7:      8  594 CEETOX_H295R ESTRADIOL
8:      8  595 CEETOX_H295R ESTRONE
9:      8  597 CEETOX_H295R_PROG
10:      8  598 CEETOX_H295R_TESTO
> tcplLoadAeid(fld='acid', val=586)
  acid aeid      aenm
1:   586  890 CEETOX_H295R_11DCORT_dn
2:   586  891 CEETOX_H295R_11DCORT_up
  
```

- Assay sources or vendors may send many files, which are pre-processed.
- The mc0 data in invitrodb is at the assay component level.
- At mc1, assay endpoints are defined, but it is not until normalization at mc3 that data are retrieved by assay endpoint.

Outline of the ToxCast pipeline

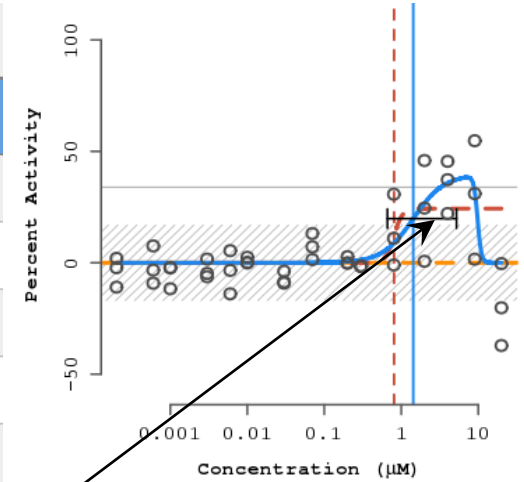


Mc7 summarizes uncertainty information for a curve-fit

- Reference vignette for description and table/fields:https://cran.r-project.org/web/packages/tcpl/vignettes/Introduction_Appendices.html#appendix-e-curve-fitting-uncertainty
- Watt, E. D. and R. S. Judson (2018). "Uncertainty quantification in ToxCast high throughput screening." PLoS One 13(7): e0196963.
- Manuscript that examines some ways to use this and other information to filter ToxCast data is in preparation (Brown, Judson, Paul Friedman, *in prep*)

Table 17: Fields in mc7 table.

Field	Description
m4id	Level 4 ID
Aeid	Assay endpoint ID
Aenm	Assay endpoint name
Asid	Assay source ID
Acid	Assay component ID
Hit_pct	Total percent of hit calls made after 1000 bootstraps
Total_hitc	Total number of hit calls made after 1000 bootstraps
Modl_ga_min	Low bound of the 95% confidence interval for the AC ₅₀
Modl_ga_max	Upper bound of the 95% confidence interval for the AC ₅₀
Modl_ga_med	Median AC ₅₀ after 1000 bootstraps
Modl_gw_med	Median gain Hill coefficient for 1000 bootstraps
Modl_ga_delta	AC ₅₀ confidence interval width in log units
Cnst_pct	Percent of 1000 bootstraps that the constant model was selected as the winning model
Hill_pct	Percent of 1000 bootstraps that the Hill model was selected as the winning model
Gnls_pct	Percent of 1000 bootstraps that the gain-loss was selected as the winning model



WORK IN PROGRESS

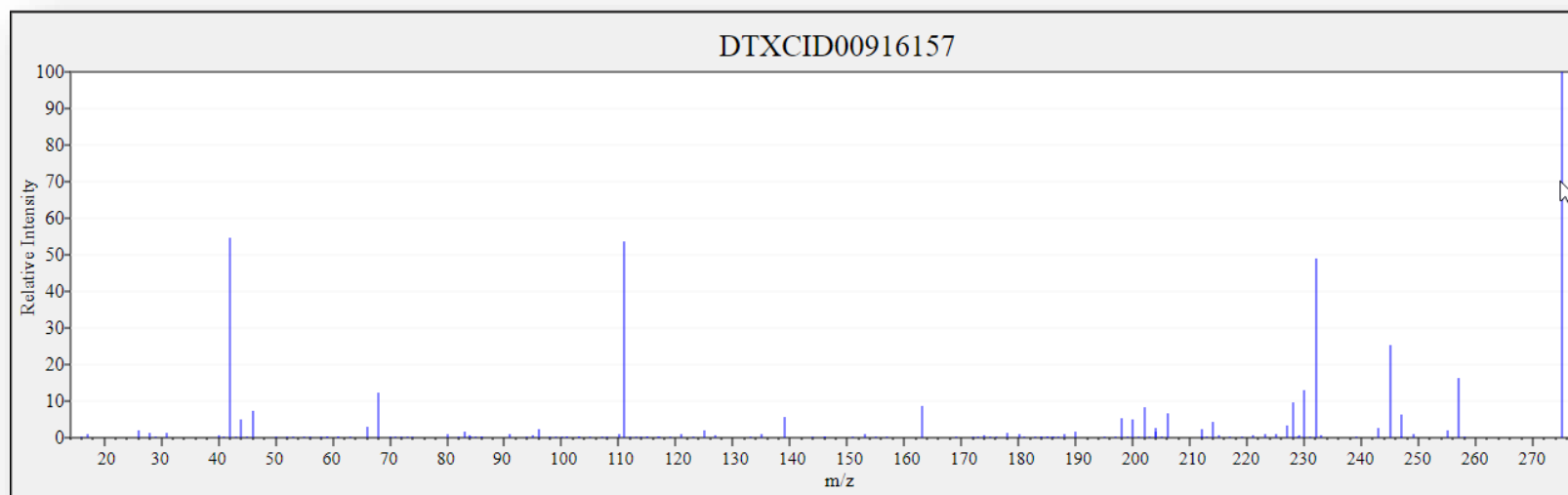
Work in Progress

- Predicted Spectra for candidate ranking
 - Viewing and Downloading pre-predicted spectra
 - Search spectra against the database


Predicted Mass Spectra



- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >800,000 structures, to be accessible via Dashboard



Search Expt. vs. Predicted Spectra

 United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share Search all data

Non Target Analysis Prototype

Mass Search

Da

Molecular Formula Search

Mass or Formula must be entered before searching spectrum

Ionization Type

ESI+ ▼

- ESI+
- ESI-
- EI

Spectra Input

304.1332052 11.8199475		
198.0913404 7.308439899		
123.0440559 6.538348292		
198.0758904 5.269483115		
216.1019051 4.700461978		
200.1080005 4.800144384		

Peak Match Window:

Search Expt. vs. Predicted Spectra

EPA United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share Search all data

Spectra Input

Single Energy

304.1332052 11.61
198.0913404 7.30
123.0440559 6.53
196.0756904 5.28
216.1019051 4.70
200.1080005 4.80

Peak Match

Search

TSV CSV Excel

Chemical Structure ID

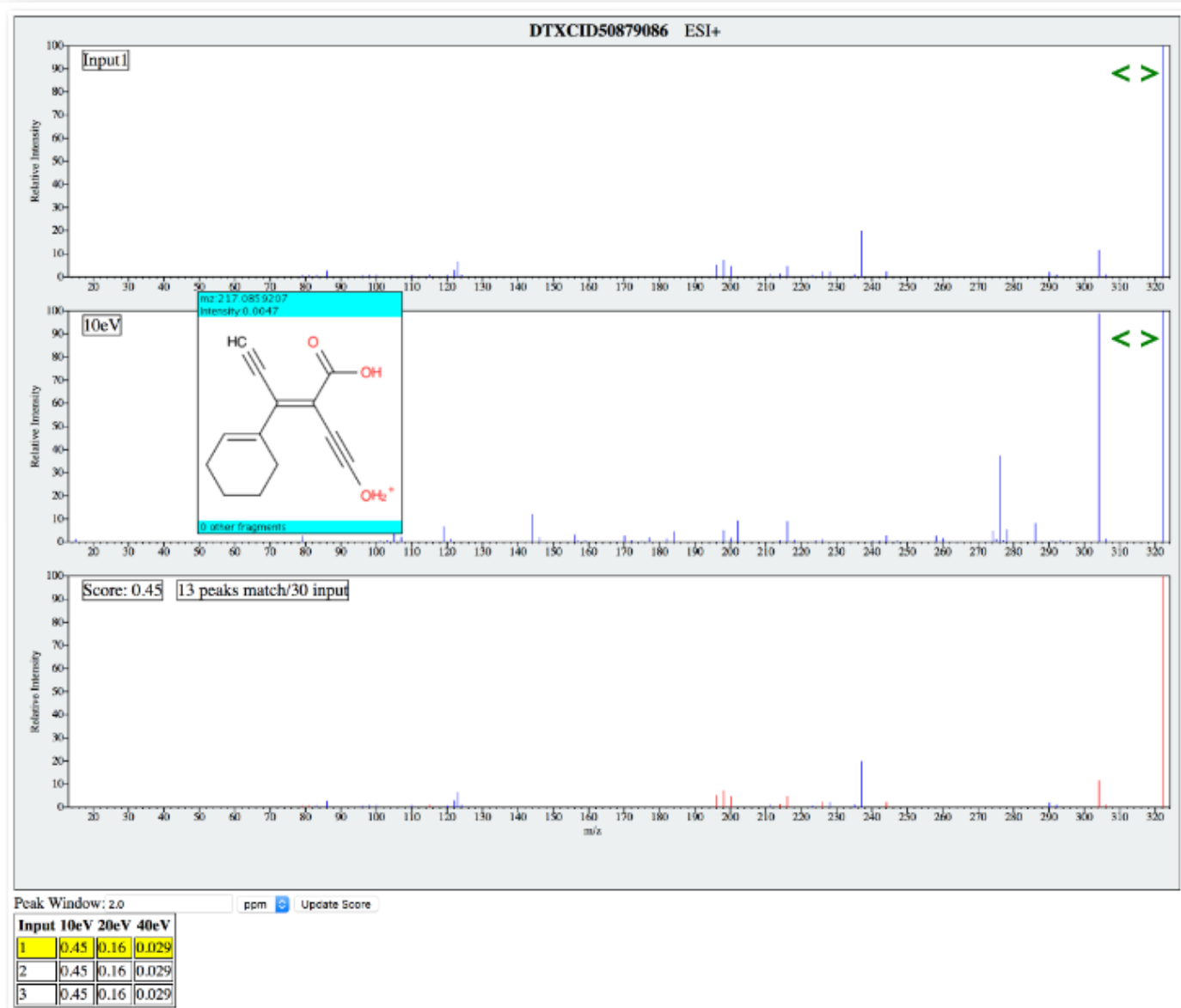
DTXCID101048191
DTXCID101181567
DTXCID50879086
DTXCID60686349
DTXCID00830900
DTXCID10971176
DTXCID60301242
DTXCID40703048
DTXCID60349982
DTXCID10316649

Showing 1 to 10 of 38 entries

Chemical Structure ID	Score (10eV)
DTXCID101048191	0.22
DTXCID101181567	0.19
DTXCID50879086	0.17
DTXCID60686349	0.14
DTXCID00830900	0.13
DTXCID10971176	0.12
DTXCID60301242	0.12
DTXCID40703048	0.11
DTXCID60349982	0.11
DTXCID10316649	0.09

1 2 3 4 Next

Spectral Viewer Comparison



Work in Progress

- Predicted Spectra for candidate ranking
 - Viewing and Downloading pre-predicted spectra
 - Search spectra against the database
- Retention Time Index Prediction

Including Relative Retention Times



Journal of Hazardous Materials

Volume 363, 5 February 2019, Pages 277-285



Development and application of retention time prediction models in the suspect and non-target screening of emerging contaminants

Reza Aalizadeh, Maria-Christina Nika, Nikolaos S. Thomaidis  

 [Show more](#)

<https://doi.org/10.1016/j.jhazmat.2018.09.047>

[Get rights and content](#)

Work in Progress

- Predicted Spectra for candidate ranking
 - Viewing and Downloading pre-predicted spectra
 - Search spectra against the database
- Retention Time Index Prediction
- Structure/substructure/similarity search

Prototype Development

AADashboard

atrazine

Search

Select properties to predict

H T.E.S.T. 18 OPERA Search

C

N Exact

O Substructure

Search result 2540

Show ☐ Isotopically Labeled ☐ Charged ☐ Salts or Mixtures

Sort Similarity

Chemical structure visualization and search results for atrazine.

Search result 2540

Show ☐ Isotopically Labeled

The screenshot displays the AADashboard web application. At the top, a search bar contains the word 'atrazine'. Below the search bar is a toolbar with various icons for file operations and editing. The main workspace shows the chemical structure of atrazine, which is a 1,3,5-triazine-2,4,6-trimethylamine derivative. To the right of the workspace, there are options to 'Select properties to predict', with 'T.E.S.T. 18' and 'OPERA' selected. Below this, there are radio buttons for 'Exact' and 'Substructure'. A 'Search' button is also present. The bottom section shows a 'Search result 2540' and a grid of 50 chemical structures, each with a similarity score. The structures are arranged in a 5x10 grid. The first structure is atrazine itself, with a similarity score of 1.0. The other structures are various derivatives of atrazine, such as 1,3,5-triazine-2,4,6-trimethylamine, 1,3,5-triazine-2,4,6-trimethylamine hydrochloride, and 1,3,5-triazine-2,4,6-trimethylamine sulfate. The similarity scores range from 0.38 to 0.57.

Prototype Development

atrazine Search

100%

Select properties to predict

T.E.S.T. 18 OPERA Search

Exact
Substructure
Similarity
Molecular Formula
Molecular Weight

Input formula (e.g. C6 H6):
C15H16O2

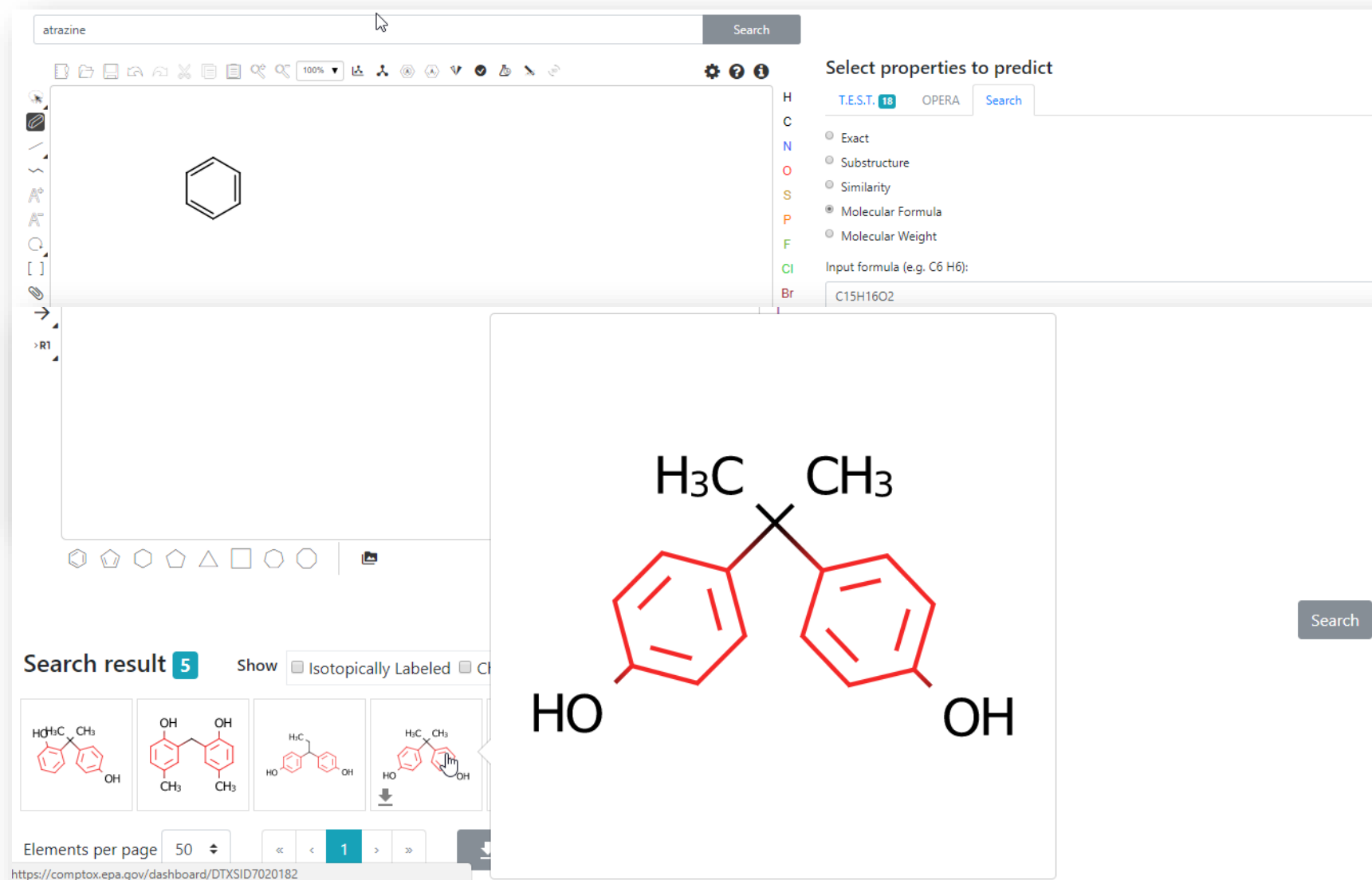
Search

Search result 5 Show ☐ Isotopically Labeled ☐ Ch

Elements per page 50

1

<https://comptox.epa.gov/dashboard/DTXSID7020182>



Work in Progress

- Predicted Spectra for candidate ranking
 - Viewing and Downloading pre-predicted spectra
 - Search spectra against the database
- Retention Time Index Prediction
- Structure/substructure/similarity search
- Access to API and web services for programmatic access

API services and Open Data

- Groups waiting on our API and web services
- Mass Spec companies instrument integration
- Release will be in iterations but for now our data are available

The screenshot shows the EPA website's DSSTox section. The header includes the EPA logo and navigation links: Home, Advanced Search, Batch Search, Lists, Predictions, Downloads, and a Share button. A search bar is also present. The main content area features three links to data files, each with a brief description and a post date.

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

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

	casrn	dsstox_substance_id	preferred_name
1			
2	26148-68-5	DTXSID7020001	A-alpha-C
3	107-29-9	DTXSID2020004	Acetaldehyde oxime
4	60-35-5	DTXSID7020005	Acetamide
5	103-90-2	DTXSID2020006	Acetaminophen
6	968-81-0	DTXSID7020007	Acetohexamide
7	18523-69-8	DTXSID2020008	Acetone[4-(5-nitro-2-furyl)-2-thiazolyl] hydrazone
8	75-05-8	DTXSID7020009	Acetonitrile
9	127-06-0	DTXSID6020010	Acetoxime
10	65734-38-5	DTXSID6020012	N'-Acetyl-4-(hydroxymethyl) phenylhydrazine

DSSTox MS Ready Mapping File Posted: 11/14/2016

The CompTox Chemistry Dashboard can be used by mass spectrometrists for the purpose of structure identification. A normal formula search would search the exact formula associated with any chemical, whether it include solvents of hydration, salts or multiple components. However, mass spectrometry detects ionized chemical structures and molecular formulae searches should be based on desalted, and desolvated structures with stereochemistry removed. We refer to these as "MS ready structures" and the MS-ready mappings are delivered as Excel Spreadsheets containing the Preferred Name, CAS-RN, DTXSID, Formula, Formula of the MS-ready structure and associated masses, SMILES and InChI Strings/Keys.

DSSTox SDF File Posted: 12/14/2016

This zip file contains the entire chemical structure collection of over 700,000 chemicals from the DSSTox database contained in one large SDF file. The file contains the structure, The DSSTox Structure Identifier (DTXCID), The DSSTOX Substance Identifier (DTXSID listed as PubChem External Data Source), the associated Dashboard URL, associated synonyms and Quality Control Level details. In order to view an SDF file you will need to have access to the appropriate piece of software to open an SDF file. Examples include ChemAxon JChem

Other Work in Progress and Future Plans

- **Work in Progress**

- New manual in development
- Training “videos” will be posted to YouTube in the future
- New lists are in preparation to add to the lists page

- **Future Work**

- Integrate ToxRefDB database views
- Realtime OPERA predictions for physchem/fate and transport
- Integrate H295R model views and other models as available

Acknowledgements for v3 release

- The NCCT IT development team – led by Jeff Edwards and Jeremy Dunne
- NCCT PIs –Chris Grulke, Keith Houck, Richard Judson, Grace Patlewicz, Ann Richard, Imran Shah, John Wambaugh
- ORISE/SSC – Jason Brown, Andrew McEachran, Nathaniel Rush, Anita Simha, Mahmoud Shobair

Feedback welcome

paul-friedman.katie@epa.gov

williams.antony@epa.gov