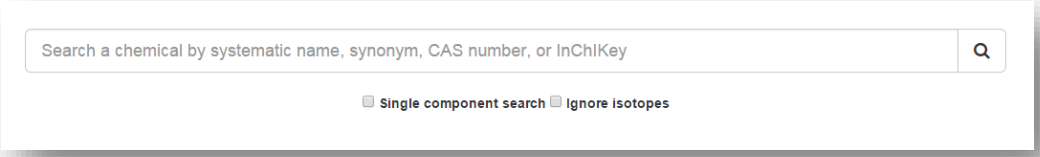
CompTox Chemistry Dashboard

Overview

The CompTox Chemistry Dashboard provides information regarding chemical compounds and substances of potential interest to environmental scientists. It contains information about substances, associated chemical structures, experimental and predicted physicochemical and toxicity data, bioassay data, exposure data and additional links to relevant websites and applications. The dashboard provides information for over 700,000 chemical substances and continues to expand in terms of the number of chemicals represented. Iterative review of the data is ongoing and additional data types are introduced with each release.

Basic Search

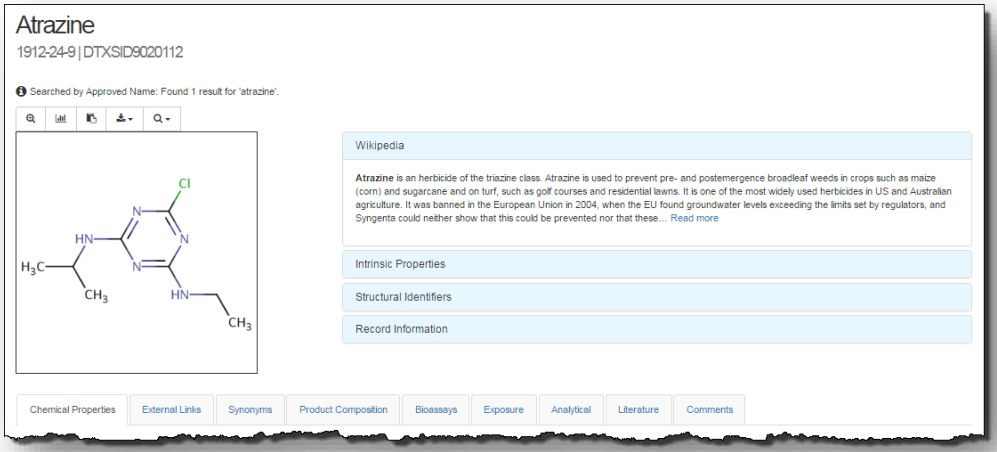
There are a variety of search capabilities presently available on the dashboard including by chemical name or identifier, by mass or molecular formula and a batch search capability for searching larger collections of data. Future work will introduce structure, substructure and similarity searching. The text search box allows a user to search using a number of chemical "identifiers" including chemical name, common name, CAS Number or InChIKey (SMILES and InChIString searching will be introduced in a later version). For the time being, the search is for an EXACT Search only (but can be filtered using the checkboxes below the entry box to return only chemical entities that have a single component or exclude isotopically-labeled compounds (e.g. no deuterium- or 13C-labeled compounds).



If a hit is identified in the database then the search will return a detailed results page with associated information for the chemical. For example, a search for "atrazine" returns the chemical at http://comptox.epa.gov/dashboard/DTXSID9020112, discussed in more detail below in the Chemical Results Page section.

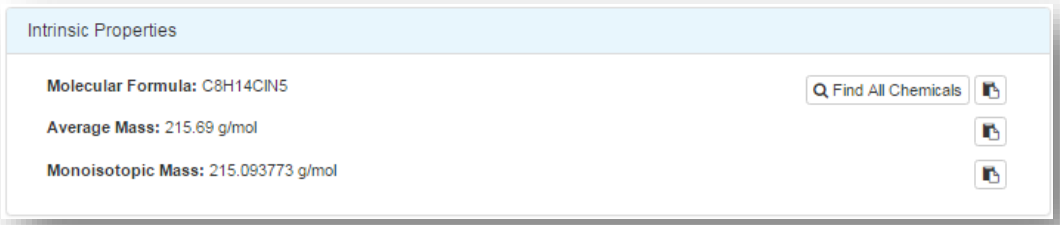
Chemical Results Page

The chemical results page for atrazine is given at http://comptox.epa.gov/dashboard/DTXSID9020112. An image of the results page is shown below.

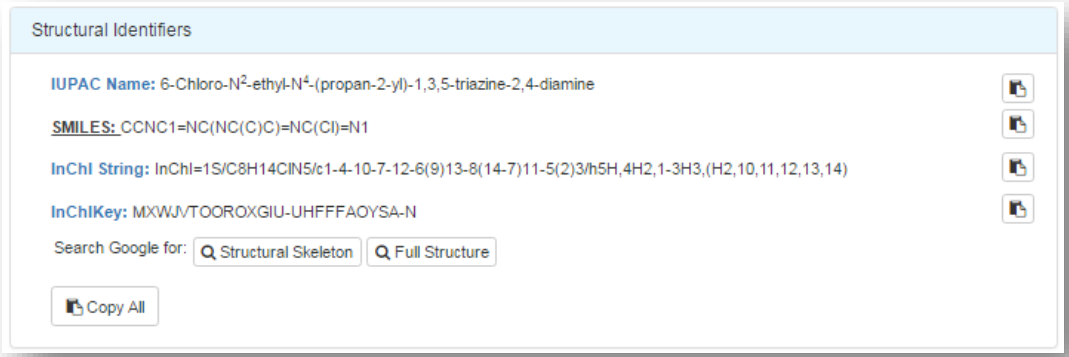


The page shows an image of the chemical structure and associated information. When a Wikipedia article is available then a snippet of that article is shown with a link out via the “Read More” link. There are over 700,000 substances on the dashboard and around 15,000 substances linked to Wikipedia only.

The intrinsic properties include molecular formula, average mass and monoisotopic mass, properties that are valuable to analytical scientists involved in structure identification and, in particular, to support targeted and non-targeted screening identification of environmental chemicals (for example, as reported by Sobus *et al.*). “Find All Chemicals” will perform a formula search of the dashboard and will display all results in a Tile or Table view depending on the user selection. This will be covered in more detail later in the help manual. All properties can be copied to the clipboard, individually, using the Copy button.

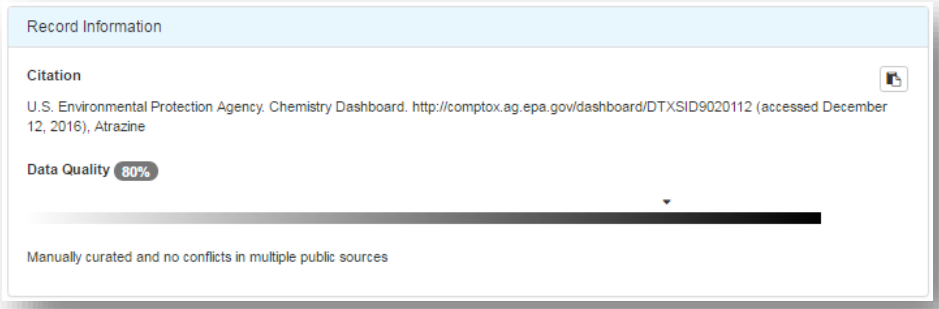


The structural identifiers include the IUPAC Name, SMILES, InChIString and InChIKey.



The InChIKey can be used to perform a search across the internet using the Google Search Engine. For example, for the chemical structure of Atrazine the InChIKey displayed above results in over 500 "exact structure" hits online: https://www.google.com/#q=MXWJVTOOROXGIU-UHFFFAOYSA-N. An InChIkey skeleton search ignores stereochemistry, charge and isotopes when searching for a particular chemical.

The Record Information tab includes a citation that can be utilized when reporting a record, as well as information regarding data quality.



The different levels that can be listed are as follows:

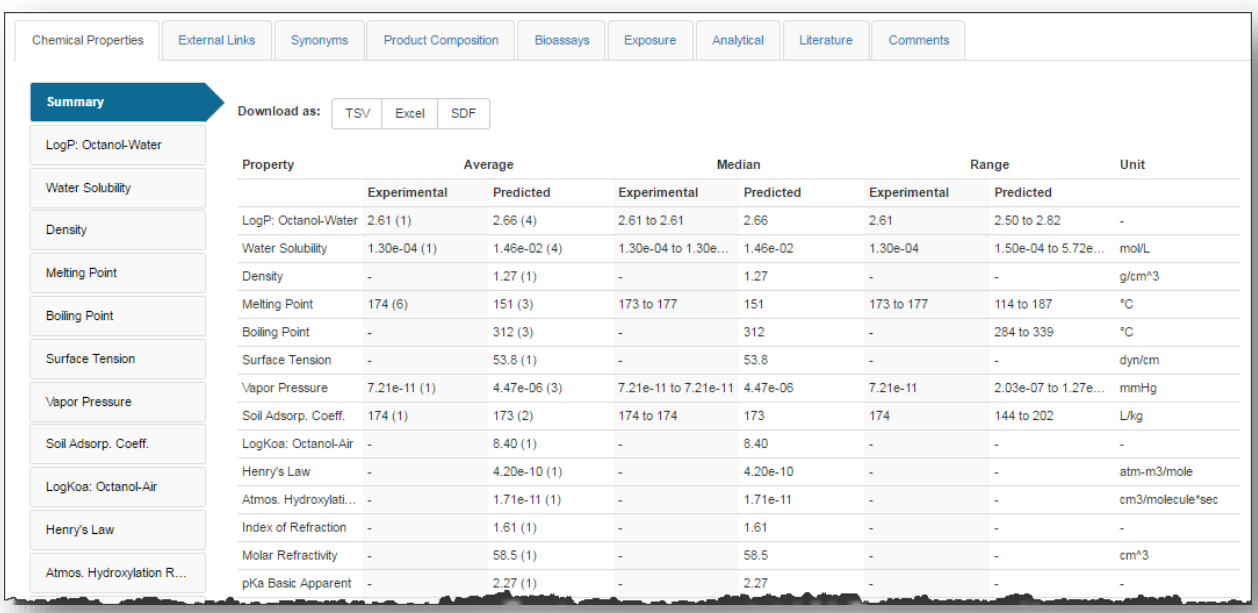
Data Quality	Definition
100%	Manually curated with highest confidence
80%	Manually curated and no conflicts in multiple public sources
60%	
40%	Programmatically curated and no conflicts in selected public sources
20%	Programmatically collected from a single public source

Chemical Search Results Page

For each chemical, a series of tabs are displayed providing access to additional information about the chemical. The list of tabs is shown below. Each of these will be discussed in order.

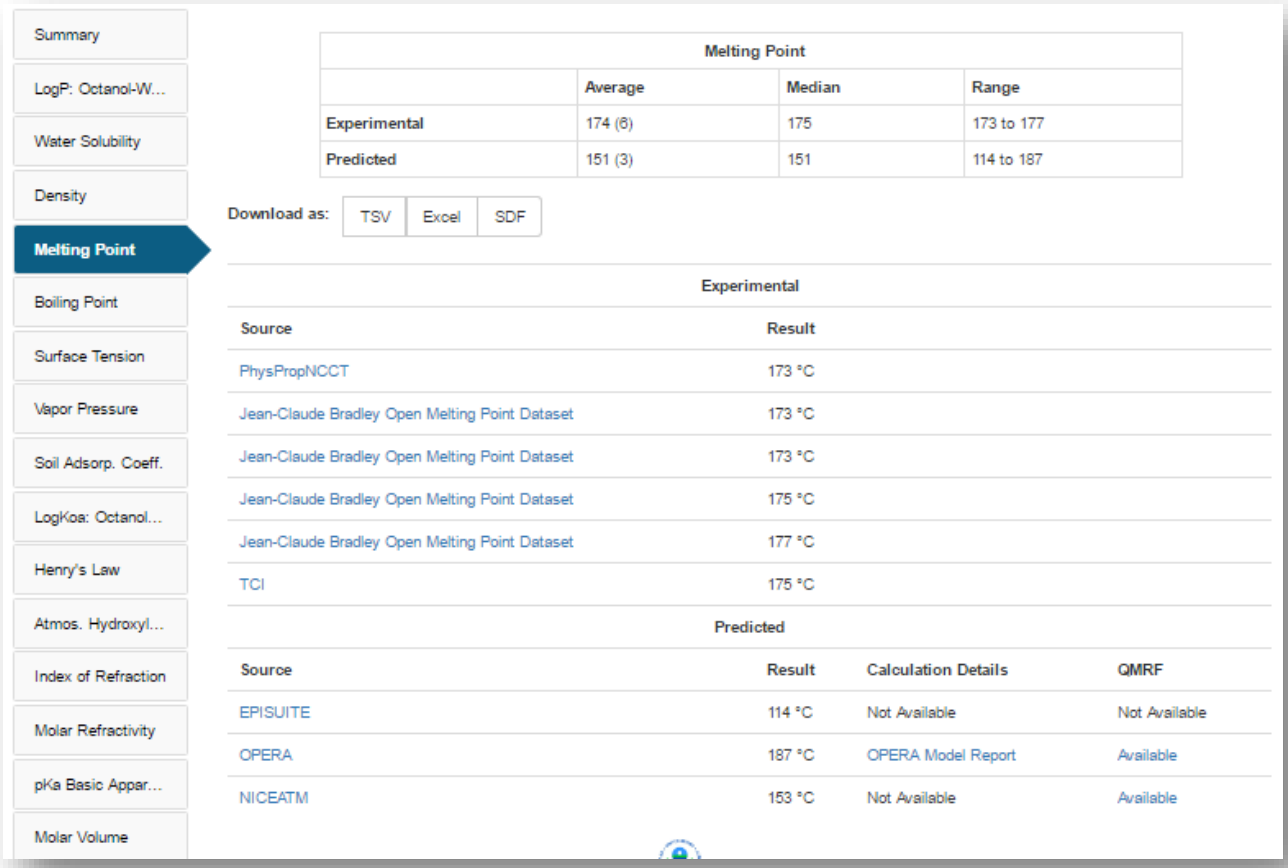
Chemical Properties

The Chemical Properties listed in the table include both experimental and predicted property data. The experimental data are sourced from various sources including data curated from PHYSPROP downloadable files (explained in detail in the paper by Mansouri *et al*) as well as data from other public websites. Five forms of predicted property data are available. These include data from EPISuite, from the OPERA models (Mansouri *et* al. publication in preparation), from the NICEATM models (Zang *et al*, manuscript accepted for publication), from the Toxicity Estimation Software Tool (TEST) Models and harvested from the Open PHACTS project as predicted by ACD/Labs. The summary tab displays a table with a list of all available properties and, if available, the average experimental property and range of experimental properties. The numbers in parentheses indicate the number of each of the properties available.



Property	Average		Median		Range		Unit
	Experimental	Predicted	Experimental	Predicted	Experimental	Predicted	
LogP: Octanol-Water	2.61 (1)	2.66 (4)	2.61 to 2.61	2.66	2.61	2.50 to 2.82	-
Water Solubility	1.30e-04 (1)	1.46e-02 (4)	1.30e-04 to 1.30e...	1.46e-02	1.30e-04	1.50e-04 to 5.72e...	mol/L
Density	-	1.27 (1)	-	1.27	-	-	g/cm^3
Melting Point	174 (6)	151 (3)	173 to 177	151	173 to 177	114 to 187	°C
Boiling Point	-	312 (3)	-	312	-	284 to 339	°C
Surface Tension	-	53.8 (1)	-	53.8	-	-	dyn/cm
Vapor Pressure	7.21e-11 (1)	4.47e-06 (3)	7.21e-11 to 7.21e-11	4.47e-06	7.21e-11	2.03e-07 to 1.27e...	mmHg
Soil Adsorp. Coeff.	174 (1)	173 (2)	174 to 174	173	174	144 to 202	L/kg
LogKoa: Octanol-Air	-	8.40 (1)	-	8.40	-	-	-
Henry's Law	-	4.20e-10 (1)	-	4.20e-10	-	-	atm-m3/mole
Atmos. Hydroxylati...	-	1.71e-11 (1)	-	1.71e-11	-	-	cm3/molecule*sec
Index of Refraction	-	1.61 (1)	-	1.61	-	-	-
Molar Refractivity	-	58.5 (1)	-	58.5	-	-	cm^3
pKa Basic Apparent	-	2.27 (1)	-	2.27	-	-	-

For the example above, the Average Melting Point is listed as 174 (6) indicating that there are 6 melting points available covering a range of 173 to 177 degrees. Clicking on the individual property button on the left hand side opens up a detailed property table for review. The table of Melting Points listed below shows the information for Atrazine. Hovering over the Source names will display details regarding the Source information.



Melting Point			
	Average	Median	Range
Experimental	174 (6)	175	173 to 177
Predicted	151 (3)	151	114 to 187

Experimental

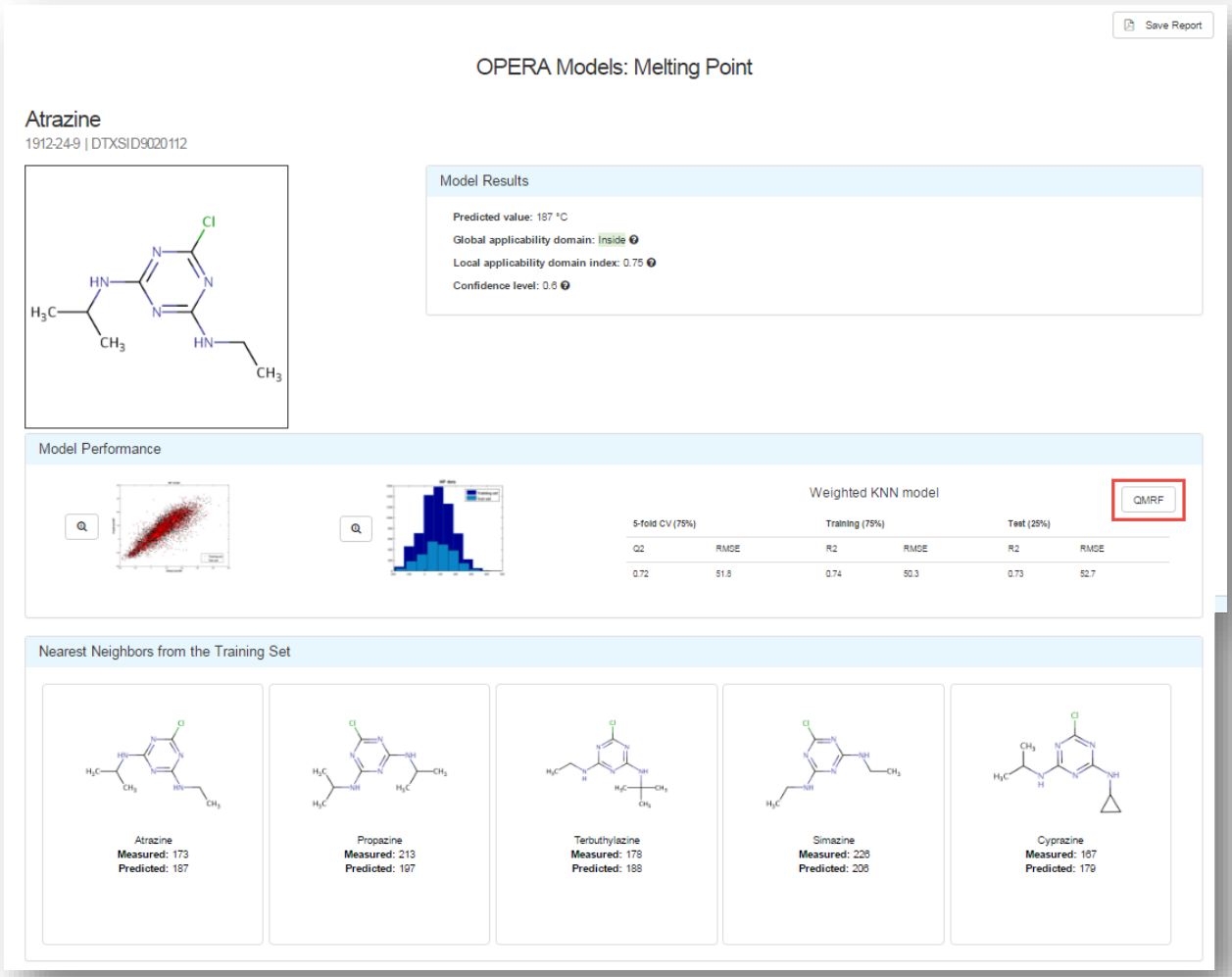
Source	Result
PhysPropNCCT	173 °C
Jean-Claude Bradley Open Melting Point Dataset	173 °C
Jean-Claude Bradley Open Melting Point Dataset	173 °C
Jean-Claude Bradley Open Melting Point Dataset	175 °C
Jean-Claude Bradley Open Melting Point Dataset	177 °C
TCI	175 °C

Predicted

Source	Result	Calculation Details	QMRF
EPISUITE	114 °C	Not Available	Not Available
OPERA	187 °C	OPERA Model Report	Available
NICEATM	153 °C	Not Available	Available

In both cases, the summary table of properties and the detailed table of values, it is possible to download the file in either tab-separated value (TSV), Excel table format or as a Structure Data File (SDF).

In order to deliver total transparency in terms of how the property predictions are performed a report regarding the *Calculation Details* is available for the OPERA calculations. An example is shown here for Atrazine and is displayed below.

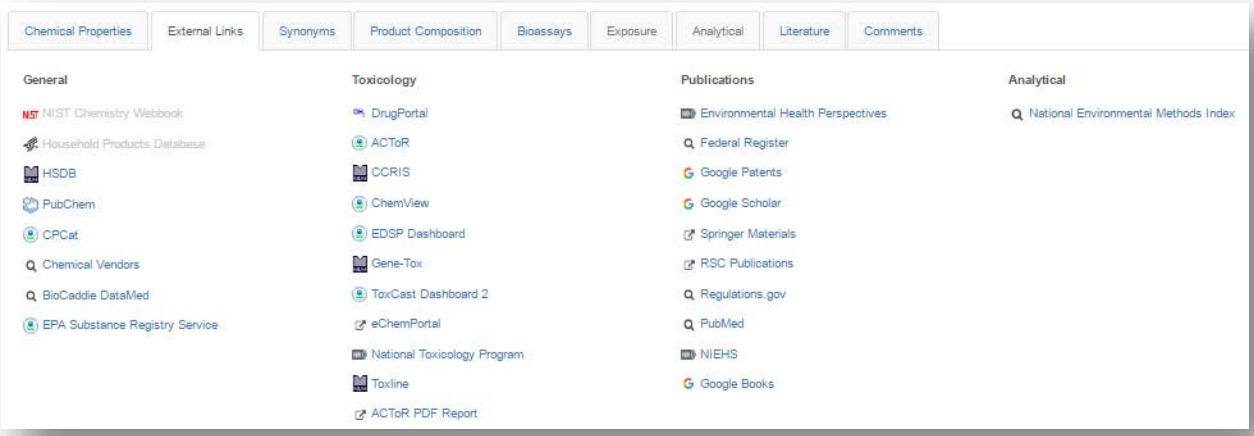


The details regarding how the prediction models were developed are given in the QMRF report that is directly available as a PDF file that can be downloaded. Each of the available OPERA models has its own QMRF detailing the software used to develop the models as well as the individual descriptors parameterizing a specific property. In the Calculation Report shown above the predicted value is the value produced by the prediction model and the Global Applicability Domain describes whether the chemical structure is contained Inside or Outside of the domain. The Local Applicability Domain Index is a value between 0 and 1 based on the similarity between the structure and its five nearest neighbors. The Confidence level, a value between 0 and 1, is an accuracy index of prediction based on the five nearest neighbors. More details are contained within the QMRF report.

Detailed prediction reports are available from the TEST predictors. An example is shown here.

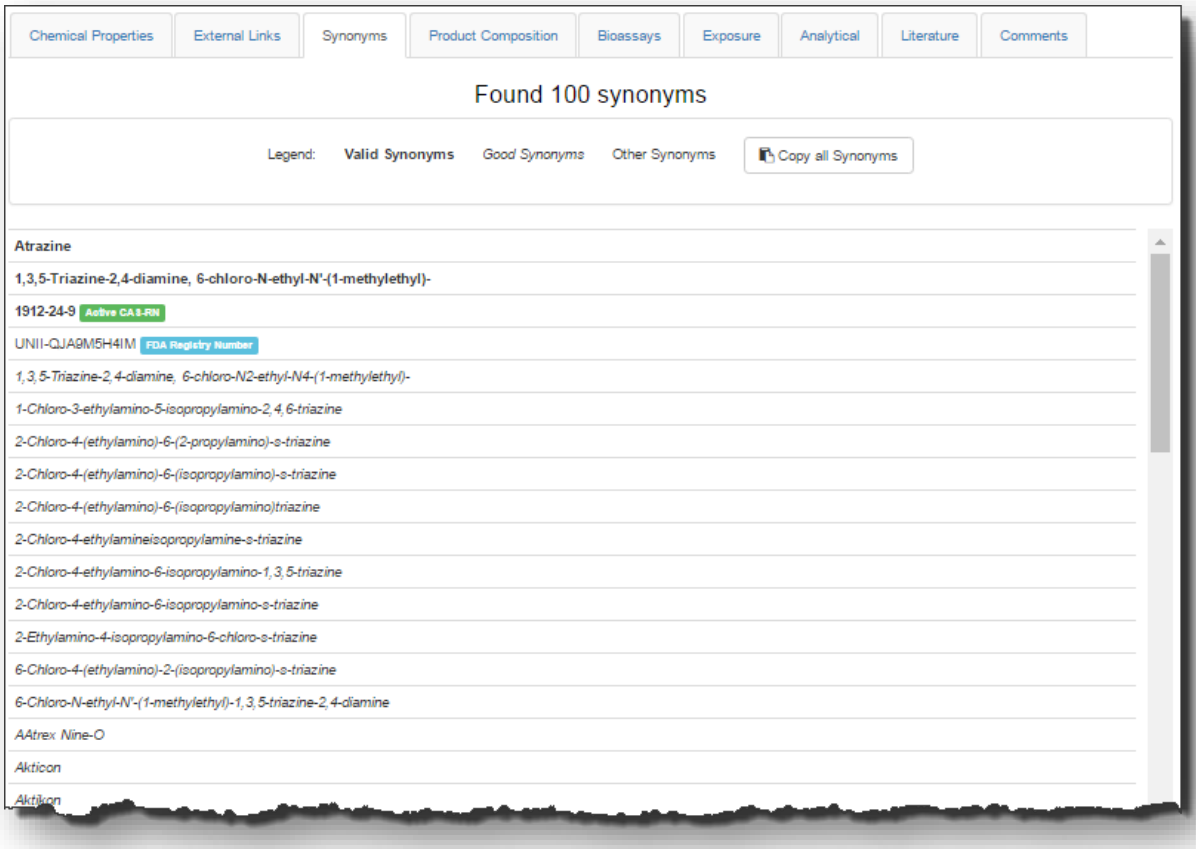
External Links

There is an abundance of information available for many of the chemicals contained within the CompTox Dashboard. In order to assist the user in navigating to these external links we have produced a click-through integration allowing the user to navigate directly to the site. Hovering over the names of the sources will display a tool tip with information about the specific resource. When links cannot be identified then the appropriate resource is grayed out and not linked. In the example below, there is no information for the particular chemical in the NIST Chemistry Webbook or Household Products Database. This approach is not exhaustive however and some of the links may not have data associated. We continue to invest in approaches to improve these external links.



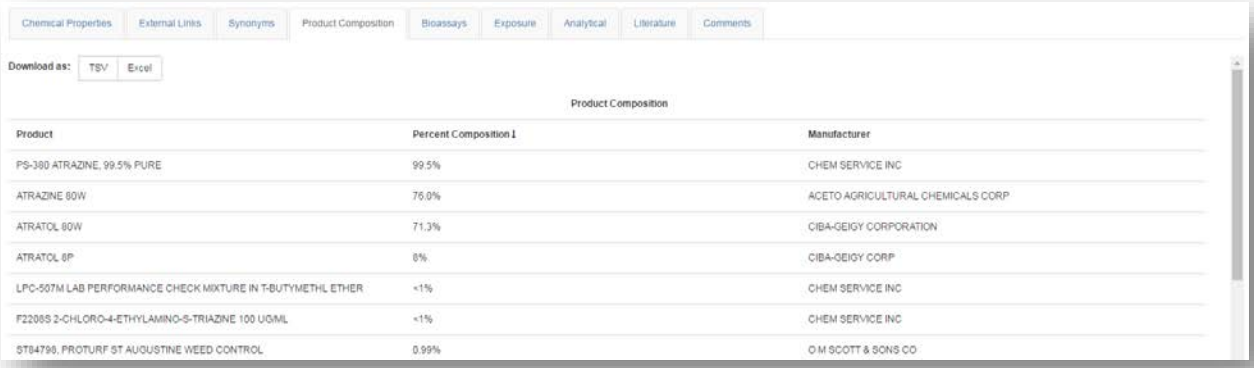
Synonyms

The database contains over a million synonyms obtained from various resources. These are displayed with the relevant quality flags as shown below. The list of synonyms can be copied to the clipboard using the “Copy All Synonyms” button. **Valid Synonyms** are those that have been manually curated by members of our curation team. *Good Synonyms* have been obtained from a variety of sources and using complex workflows for cross-referencing are ranked as being of higher quality than the remaining collection of "Other Synonyms". The synonyms are not warranted as perfect or complete but are the results of our best efforts.



Product Composition

The product composition tab provides a listing products, percent composition and associated manufacturers from the CPCat Database. The data can be downloaded in either TSV or Excel format.

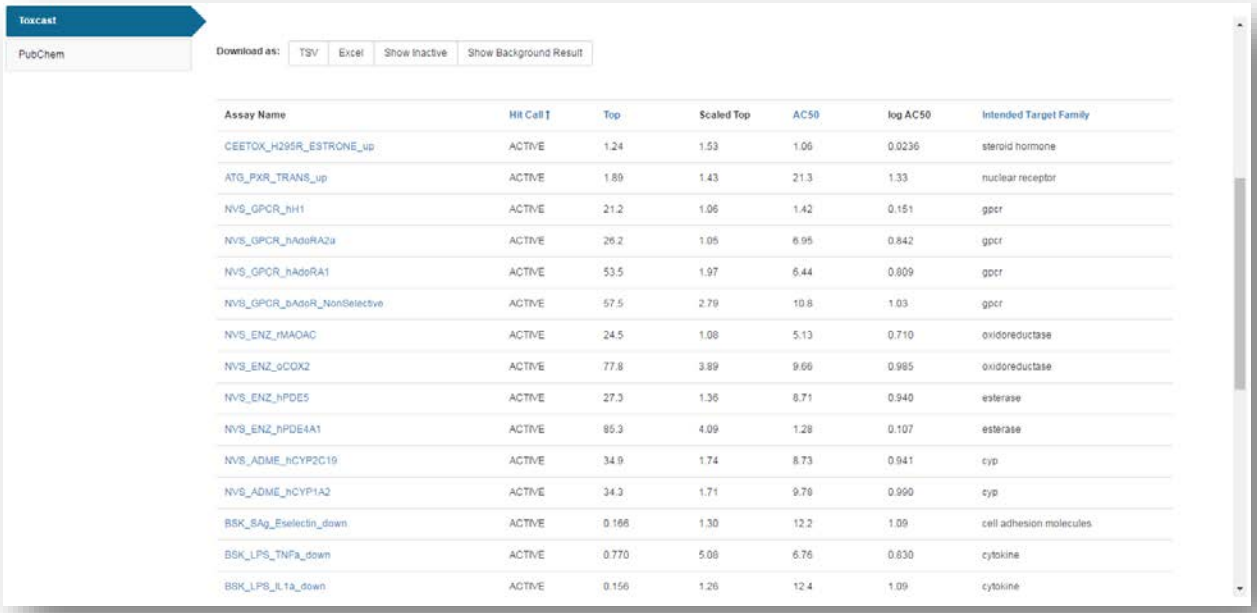


Product	Percent Composition	Manufacturer
PS-380 ATRAZINE, 99.5% PURE	99.5%	CHEM SERVICE INC
ATRAZINE 80W	76.0%	ACETO AGRICULTURAL CHEMICALS CORP
ATRATOL 80W	71.3%	CIBA-GEIGY CORPORATION
ATRATOL 6P	6%	CIBA-GEIGY CORP
LPC-507M LAB PERFORMANCE CHECK MIXTURE IN T-BUTYMETHL ETHER	<1%	CHEM SERVICE INC
F2206S 2-CHLORO-4-ETHYLAMINO-6-TRIAZINE 100 UG/ML	<1%	CHEM SERVICE INC
ST84796, PROTURF ST AUGUSTINE WEED CONTROL	0.99%	O M SCOTT & SONS CO

Bioassays

The Bioassays tab provides access to two different types of bioassay data. This includes the ToxCast data and an embedded view of PubChem Bioassay data. The Toxcast data shows the bioassay data represented as a graphic and the target families displayed on the right hand side. These can be displayed one family at a time and switched using Show/Hide All as necessary. The color coding represents the ACTIVE hit calls as a function of target family.

The table below the plot lists the tabular data associated with the Toxcast data. The Inactives and “Background Results” are *not* shown by default but can be displayed by selecting the buttons above the table. The columns can be ranked by selecting the appropriate column header (in blue) and hovering over the assay name provides the details of the assay. Selecting the download in either TSV or Excel will download the file *including* the background results and inactives.



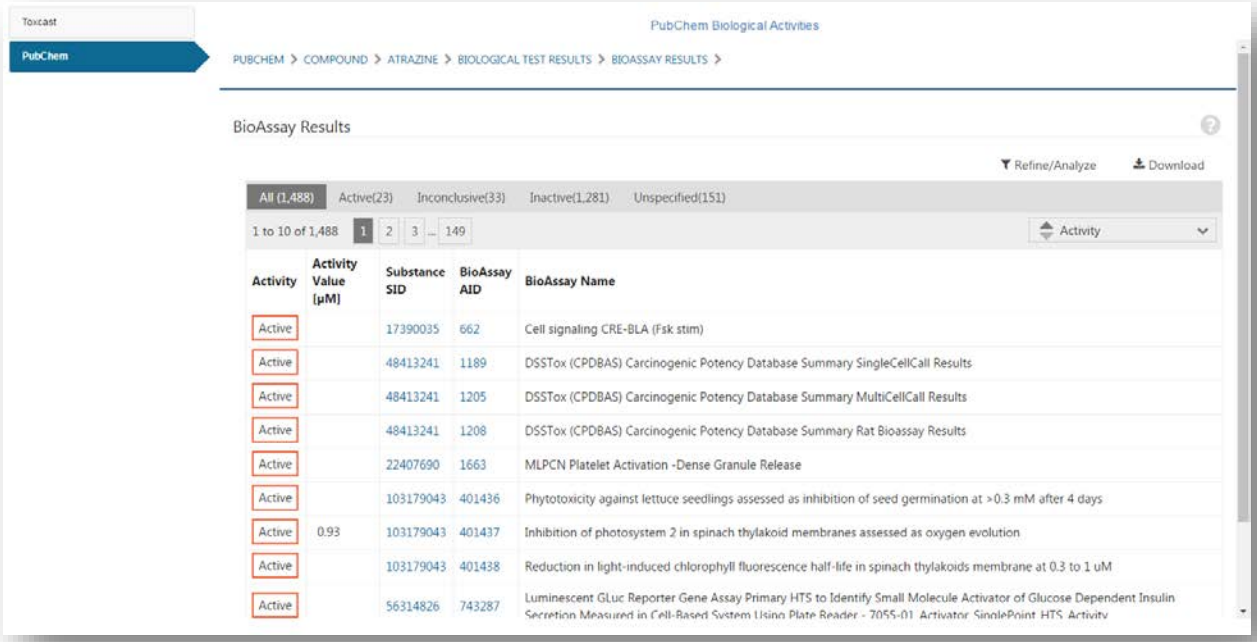
Toxcast

PubChem

Download as: TSV | Excel | Show Inactive | Show Background Result

Assay Name	Hit Call	Top	Scaled Top	AC50	log AC50	Intended Target Family
CEETOX_H295R_ESTRONE_up	ACTIVE	1.24	1.53	1.06	0.0236	steroid hormone
ATG_PXR_TRANS_up	ACTIVE	1.89	1.43	21.3	1.33	nuclear receptor
NVS_GPCR_hH1	ACTIVE	21.2	1.06	1.42	0.151	gpcr
NVS_GPCR_hAdoRA2a	ACTIVE	26.2	1.05	6.95	0.842	gpcr
NVS_GPCR_hAdoRA1	ACTIVE	53.5	1.97	6.44	0.809	gpcr
NVS_GPCR_bAdoR_NonSelective	ACTIVE	57.5	2.79	10.8	1.03	gpcr
NVS_ENZ_rMAOAC	ACTIVE	24.5	1.08	5.13	0.710	oxidoreductase
NVS_ENZ_oCOX2	ACTIVE	77.8	3.89	9.66	0.985	oxidoreductase
NVS_ENZ_hPDE5	ACTIVE	27.3	1.36	8.71	0.940	esterase
NVS_ENZ_hPDE4A1	ACTIVE	85.3	4.09	1.28	0.107	esterase
NVS_ADME_hCYP2C19	ACTIVE	34.9	1.74	8.73	0.941	cyp
NVS_ADME_hCYP1A2	ACTIVE	34.3	1.71	9.78	0.990	cyp
BSK_SAg_Eselectin_down	ACTIVE	0.166	1.30	12.2	1.09	cell adhesion molecules
BSK_LPS_TNFa_down	ACTIVE	0.770	5.08	6.76	0.830	cytokine
BSK_LPS_IL1a_down	ACTIVE	0.156	1.26	12.4	1.09	cytokine

The PubChem selection button displays a PubChem widget allowing the user to navigate through the PubChem data sourced using the embedded widget and real time service calls against PubChem to retrieve the latest data.



Toxcast

PubChem

PubChem Biological Activities

PUBCHEM > COMPOUND > ATRAZINE > BIOLOGICAL TEST RESULTS > BIOASSAY RESULTS >

BioAssay Results

Refine/Analyze | Download

All (1,488) | Active(23) | Inconclusive(33) | Inactive(1,281) | Unspecified(151)

1 to 10 of 1,488 | 1 | 2 | 3 | ... | 149 | Activity

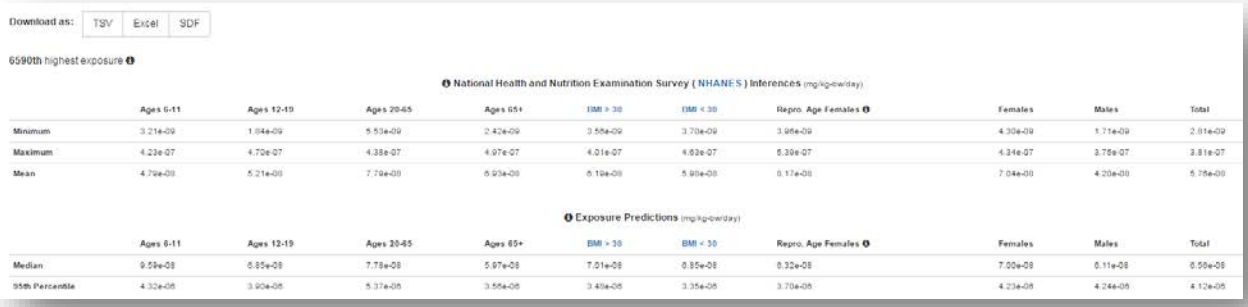
Activity	Activity Value [µM]	Substance SID	BioAssay AID	BioAssay Name
Active		17390035	662	Cell signaling CRE-BLA (Fsk stim)
Active		48413241	1189	DSSTox (CPDBAS) Carcinogenic Potency Database Summary SingleCellCall Results
Active		48413241	1205	DSSTox (CPDBAS) Carcinogenic Potency Database Summary MultiCellCall Results
Active		48413241	1208	DSSTox (CPDBAS) Carcinogenic Potency Database Summary Rat Bioassay Results
Active		22407690	1663	MLPCN Platelet Activation -Dense Granule Release
Active		103179043	401436	Phytotoxicity against lettuce seedlings assessed as inhibition of seed germination at >0.3 mM after 4 days
Active	0.93	103179043	401437	Inhibition of photosystem 2 in spinach thylakoid membranes assessed as oxygen evolution
Active		103179043	401438	Reduction in light-induced chlorophyll fluorescence half-life in spinach thylakoids membrane at 0.3 to 1 uM
Active		56314826	743287	Luminescent GLuc Reporter Gene Assay Primary HTS to Identify Small Molecule Activator of Glucose Dependent Insulin Secretion Measured in Cell-Based System Using Plate Reader - 7055-01 Activator SinglePoint HTS Activity

Navigation of the PubChem widget should be fairly self-explanatory

Exposure

The Exposure tab lists inferred exposure data associated with the National health and Nutrition Examination Survey (NHANES) (where available for a chemical) as well as predicted exposure levels, where available. The available data

is very limited across the dashboard with only 106 data tables available for NHANES and ~8000 predicted exposure levels. The data can be downloaded as TSV, Excel or SDF files.



Download as: TSV | Excel | SDF

6590th highest exposure

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	3.21e-09	1.84e-09	5.53e-09	2.42e-09	3.55e-09	3.70e-09	3.96e-09	4.30e-09	1.71e-09	2.81e-09
Maximum	4.23e-07	4.70e-07	4.38e-07	4.97e-07	4.01e-07	4.62e-07	5.39e-07	4.34e-07	3.75e-07	3.81e-07
Mean	4.79e-08	5.21e-08	7.79e-08	6.93e-08	6.19e-08	5.98e-08	6.17e-08	7.04e-08	4.20e-08	5.78e-08

Exposure Predictions (mg/kg-bw/day)

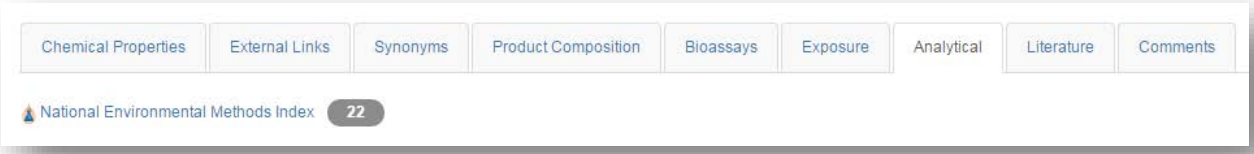
	Ages 6-11	Ages 12-19	Ages 20-65	Ages 65+	BMI > 30	BMI < 30	Repro. Age Females	Females	Males	Total
Median	9.59e-08	6.85e-08	7.78e-08	5.97e-08	7.01e-08	6.85e-08	6.32e-08	7.00e-08	6.11e-08	6.50e-08
95th Percentile	4.32e-06	3.90e-06	5.37e-06	3.55e-06	3.48e-06	3.35e-06	3.70e-06	4.23e-06	4.24e-06	4.12e-06

NHANES: The Centers for Disease Control (CDC) National Health and Nutrition Examination Survey (NHANES) is a program of studies designed to assess the health and nutritional status of adults and children in the United States. As part of the NHANES, the CDC monitors biomarkers of chemical exposure (chiefly metabolites) in the blood and urine to quantify the levels of chemical compounds present in U.S. residents. Each two-year study cycle comprises approximately 10,000 individuals, with exposure biomarker data for any given chemical available from only a subset of roughly 1/3 of the individuals in the study. In the 2014 publication "High Throughput Heuristics for Prioritizing Human Exposure to Environmental Chemicals" EPA scientists inferred steady state exposure rates (mg/kg bodyweight) for **106 chemicals** from the NHANES urine samples.

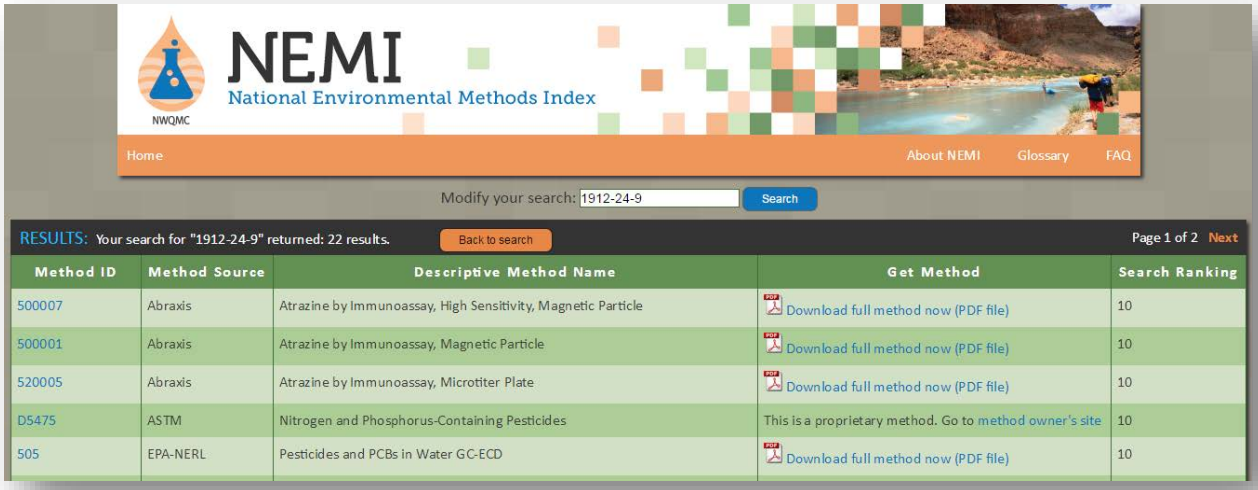
Predicted Exposure Data: In Figure 4 of the 2014 publication "High Throughput Heuristics for Prioritizing Human Exposure to Environmental Chemicals" EPA scientists ranked 7968 chemicals based on the upper (highest) 95th percentile limit on the predicted average (geometric mean) exposure rate for the total U.S. population. This is one of those 7968 chemicals. The 1st highest would have the highest predicted possible rate, while the 7968th highest would have the lowest predicted possible rate. Note that because of limited information available on exposure to most chemicals the uncertainties are large and we are only 95% confident that the exposure rate is equal to or lower than the upper 95th percentile limit. The ranking for a chemical is shown on the top left hand side as the XXXXth Highest Exposure out of 7968 chemicals.

Analytical

The analytical data is a simple count of the number of methods reported in the National Environmental Methods Index (https://www.nemi.gov/home/). The CAS number associated with the substance of interest is searched against the database and returns a count. Clicking on the count will open the NEMI page with the list of associated methods.



Links through to the NEMI website and lists the associated methods. An example is for Atrazine.



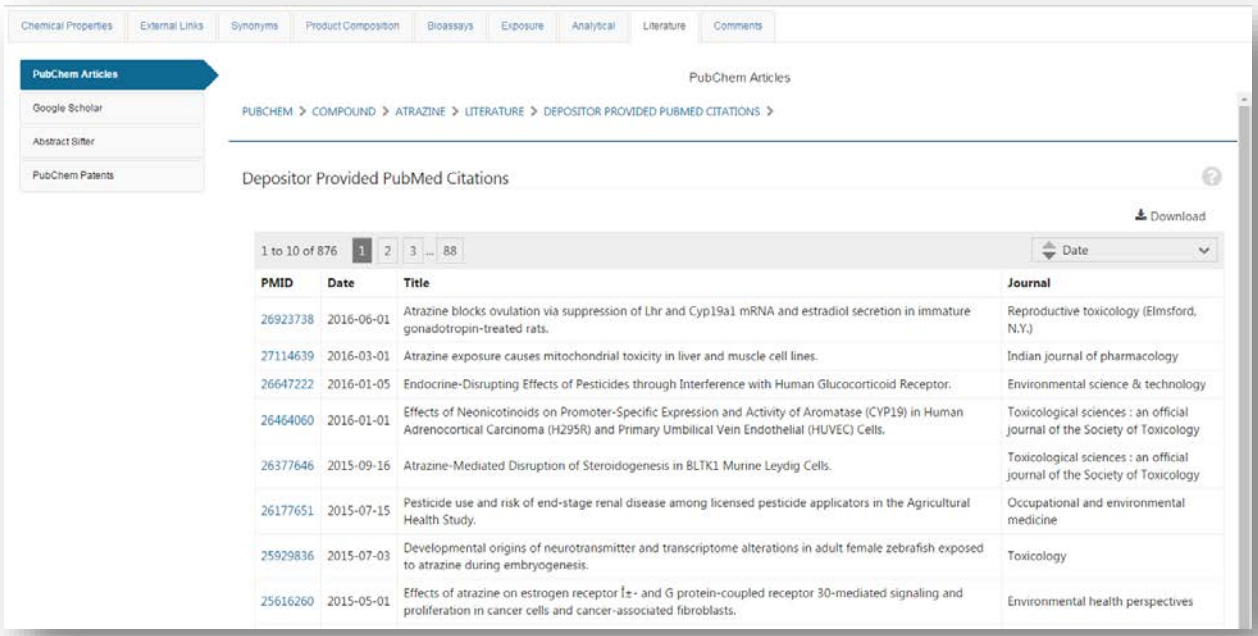
Method ID	Method Source	Descriptive Method Name	Get Method	Search Ranking
500007	Abraxis	Atrazine by Immunoassay, High Sensitivity, Magnetic Particle	Download full method now (PDF file)	10
500001	Abraxis	Atrazine by Immunoassay, Magnetic Particle	Download full method now (PDF file)	10
520005	Abraxis	Atrazine by Immunoassay, Microtiter Plate	Download full method now (PDF file)	10
D5475	ASTM	Nitrogen and Phosphorus-Containing Pesticides	This is a proprietary method. Go to method owner's site	10
505	EPA-NERL	Pesticides and PCBs in Water GC-ECD	Download full method now (PDF file)	10

Literature

The Literature Tab provides access to four different ways to access data we class as literature for the purpose of review within the dashboard. It includes access to PubMed literature, searches against Google Scholar, an Abstract Sifter that can be used for specific queries against Pubmed articles using MeSH based searches and access to PubChem Patents.

PubChem Articles

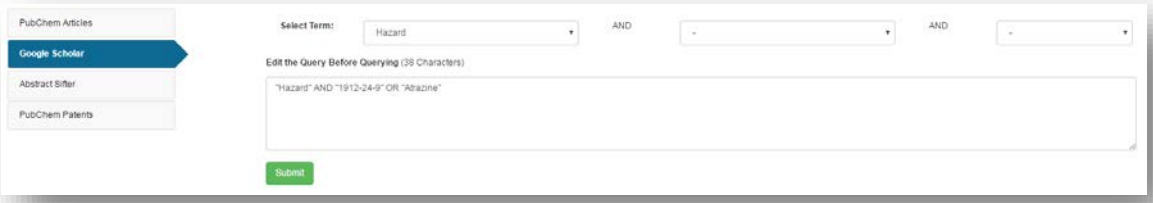
The PubMed widget provides access to the latest articles reported in Pubmed with the chemical in question flagged for inclusion using the MeSH Ontology mappings in PubChem. The articles are ranked with most recent listed first.



PMID	Date	Title	Journal
26923738	2016-06-01	Atrazine blocks ovulation via suppression of Lhr and Cyp19a1 mRNA and estradiol secretion in immature gonadotropin-treated rats.	Reproductive toxicology (Elmsford, N.Y.)
27114639	2016-03-01	Atrazine exposure causes mitochondrial toxicity in liver and muscle cell lines.	Indian journal of pharmacology
26647222	2016-01-05	Endocrine-Disrupting Effects of Pesticides through Interference with Human Glucocorticoid Receptor.	Environmental science & technology
26464060	2016-01-01	Effects of Neonicotinoids on Promoter-Specific Expression and Activity of Aromatase (CYP19) in Human Adrenocortical Carcinoma (H295R) and Primary Umbilical Vein Endothelial (HUVEC) Cells.	Toxicological sciences : an official journal of the Society of Toxicology
26377646	2015-09-16	Atrazine-Mediated Disruption of Steroidogenesis in BLTK1 Murine Leydig Cells.	Toxicological sciences : an official journal of the Society of Toxicology
26177651	2015-07-15	Pesticide use and risk of end-stage renal disease among licensed pesticide applicators in the Agricultural Health Study.	Occupational and environmental medicine
25929836	2015-07-03	Developmental origins of neurotransmitter and transcriptome alterations in adult female zebrafish exposed to atrazine during embryogenesis.	Toxicology
25616260	2015-05-01	Effects of atrazine on estrogen receptor [illegible]- and G protein-coupled receptor 30-mediated signaling and proliferation in cancer cells and cancer-associated fibroblasts.	Environmental health perspectives

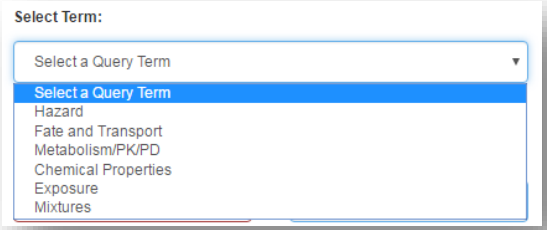
Google Scholar

The Google Scholar search function allows for terms to be selected from the drop d owns. These are appended into a search using the CAS Number and Preferred Name. For the query shown below a search of Google Scholar will return a list of articles with *Atrazine* and *Hazard* in the query results. The query can also be edited in the search box as necessary.

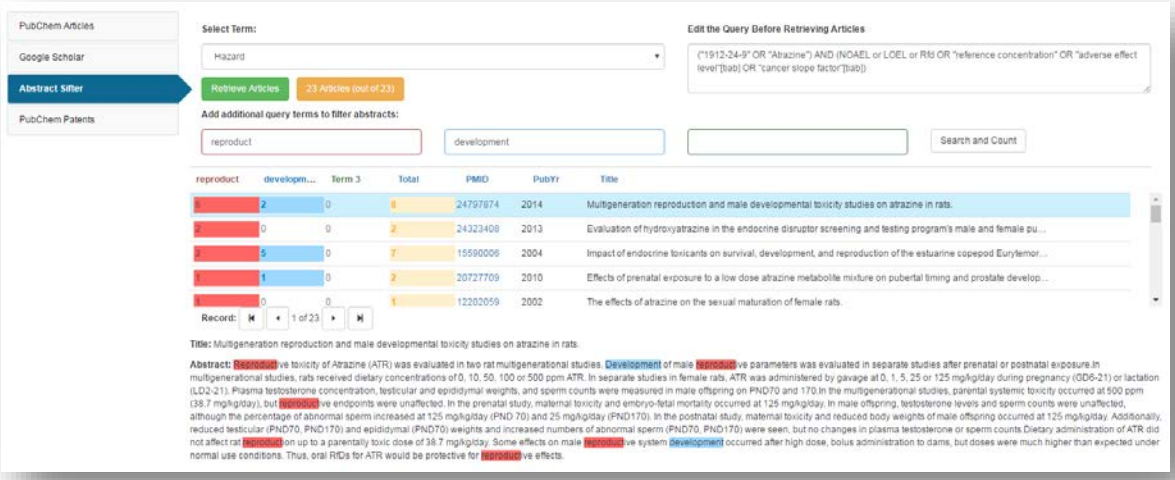


Abstract Sifter

The abstract sifter functionality uses pre-defined queries to initiate a query against PubMed. The list of queries is:

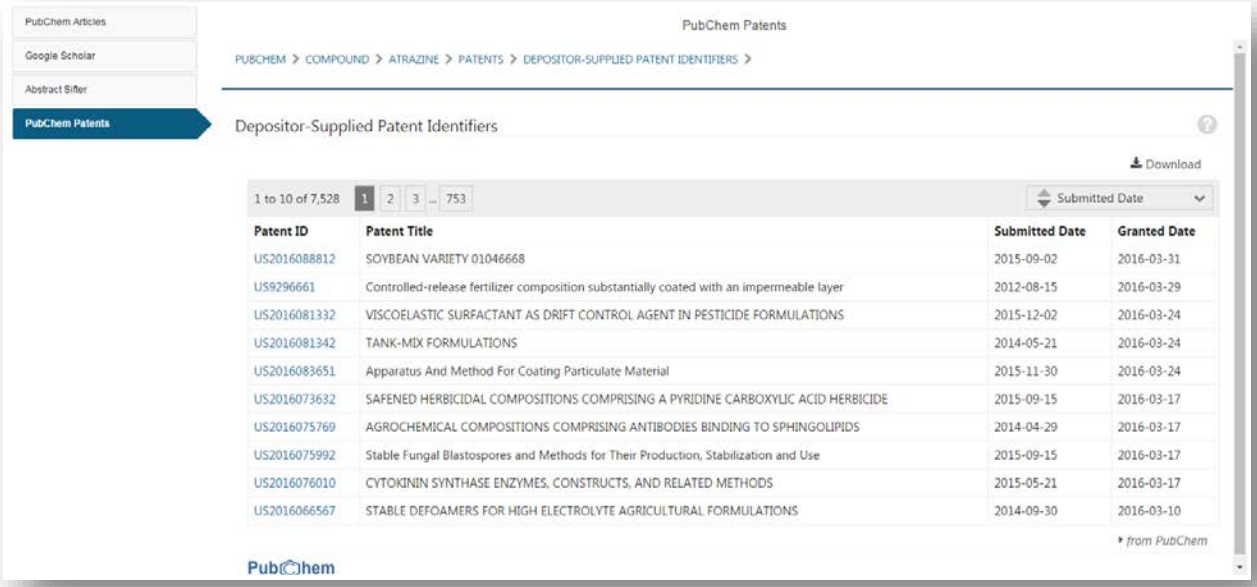


Selection of one of these queries creates the MeSH query required to perform the search. Retrieve articles (the green button) will return the number of articles that are identified with a maximum of 1000 available for download. For the example shown below the abstracts are retrieved and the filter boxes can be used to search and mark-up the abstracts based on substring searches. Up to three terms can be searched (using the Search and Count button) and clicking on the individual terms in the table will rank order the results based on the specific search terms. When one or more abstracts are identified for further reading then the PMID can be selected to open up the abstract in the PubMed interface.



Patents

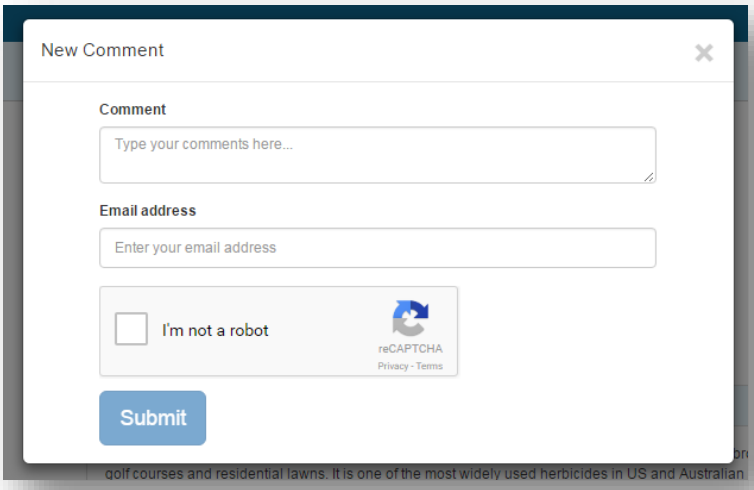
The PubChem Patent Widget lists patents associated with a particular chemical and indexed through PubChem. The list of patents is ranked with most recent patent listed first. The data can be downloaded directly from the widget using the Download Button at the top right hand corner.



Patent ID	Patent Title	Submitted Date	Granted Date
US2016088812	SOYBEAN VARIETY 01046668	2015-09-02	2016-03-31
US9296661	Controlled-release fertilizer composition substantially coated with an impermeable layer	2012-08-15	2016-03-29
US2016081332	VISCOELASTIC SURFACTANT AS DRIFT CONTROL AGENT IN PESTICIDE FORMULATIONS	2015-12-02	2016-03-24
US2016081342	TANK-MIX FORMULATIONS	2014-05-21	2016-03-24
US2016083651	Apparatus And Method For Coating Particulate Material	2015-11-30	2016-03-24
US2016073632	SAFENED HERBICIDAL COMPOSITIONS COMPRISING A PYRIDINE CARBOXYLIC ACID HERBICIDE	2015-09-15	2016-03-17
US2016075769	AGROCHEMICAL COMPOSITIONS COMPRISING ANTIBODIES BINDING TO SPHINGOLIPIDS	2014-04-29	2016-03-17
US2016075992	Stable Fungal Blastospores and Methods for Their Production, Stabilization and Use	2015-09-15	2016-03-17
US2016076010	CYTOKININ SYNTHASE ENZYMES, CONSTRUCTS, AND RELATED METHODS	2015-05-21	2016-03-17
US2016066567	STABLE DEFOAMERS FOR HIGH ELECTROLYTE AGRICULTURAL FORMULATIONS	2014-09-30	2016-03-10

Comments

The comments tab will list any comments made by users of the system in terms of potential data issues. Each chemical record can be annotated with comments regarding incorrect synonyms, quality of experimental property data etc. This allows crowd sourcing curation of the data on the database and engages the community of users in facilitating expansion and validation of the data. Comments can be submitted by selecting the Select Comment button on the top right hand side of the interface. This opens up a modal dialog box for the user to add their comments. The users email must be included in order to submit a comment as this will allow us to respond. The email will not be displayed on the record and will be kept private. A captcha is included to prevent spamming.

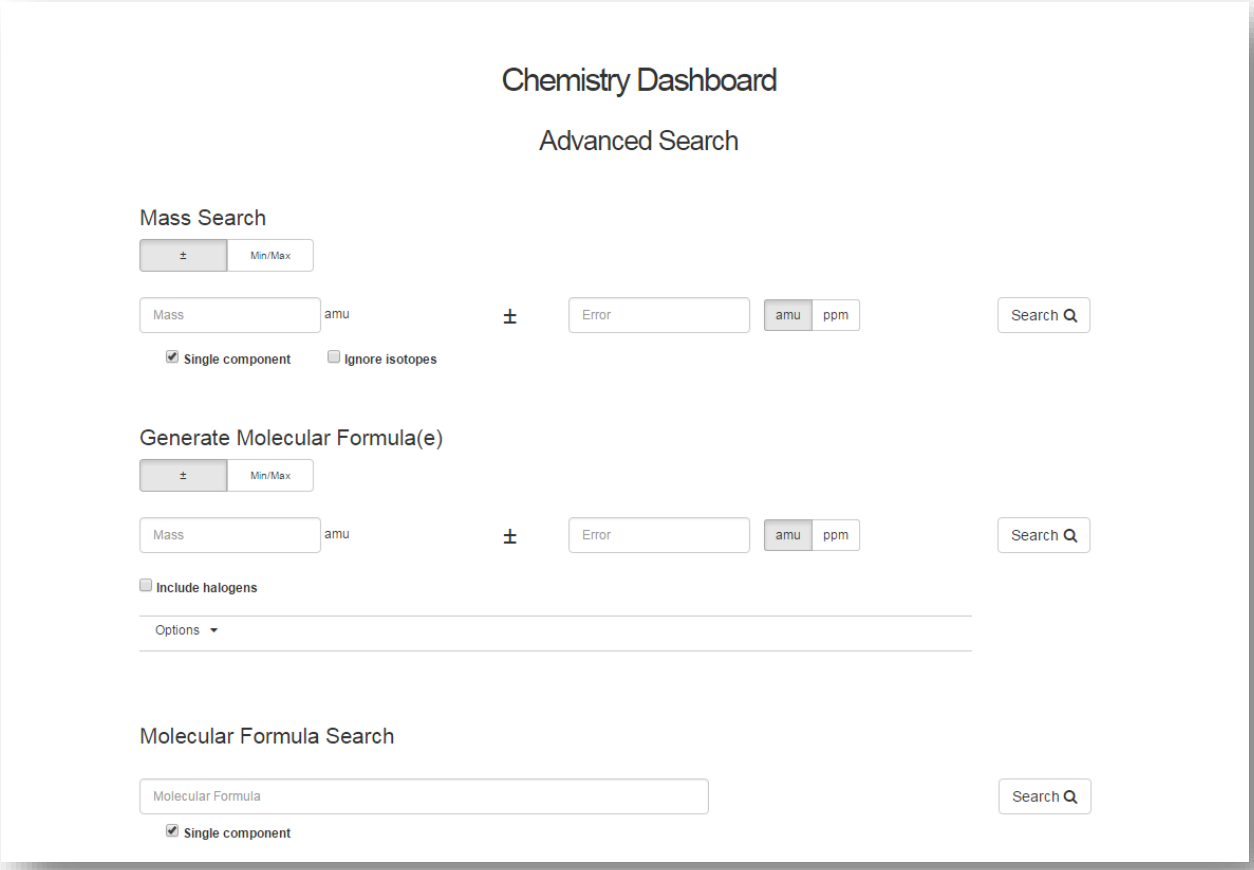


Advanced Searches

A number of additional searches are possible via the Advanced Search page at http://comptox.epa.gov/dashboard/dsstoxdb/advanced_search, or accessed on the top left hand side of the page. There are two blocks to the advanced search: Searches to support Mass spectrometry (formula and mass) and batch-based searches.

Mass Spectrometry Based Searches

The MS-based searches allow for searches based on inputting masses or formulae as shown in the interface entry forms below.

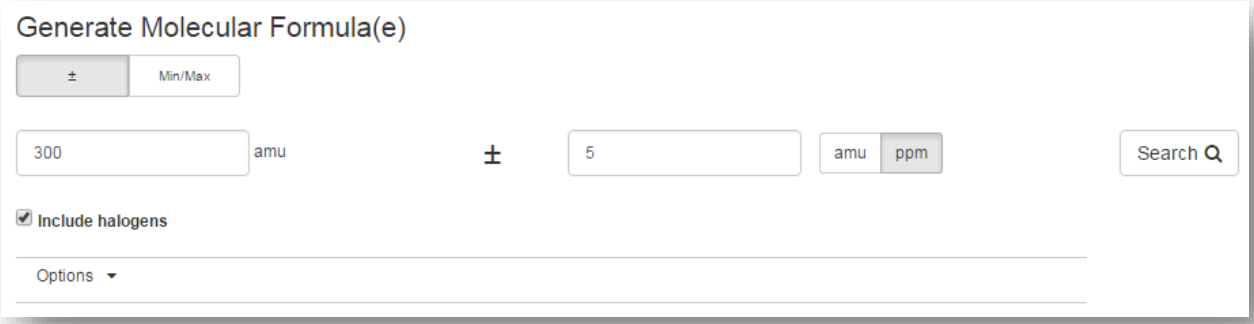


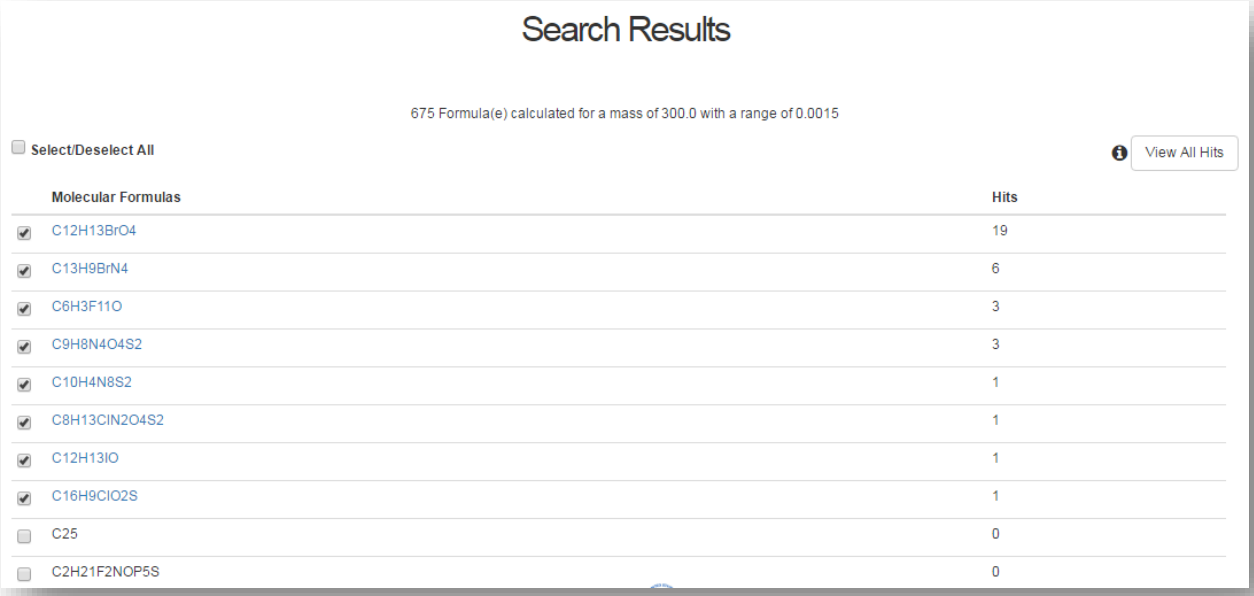
Mass Search

This search is primarily incorporated to allow a user to search the database for chemicals matching a particular monoisotopic mass. This search will be of particular value for screening and non-targeted analysis users. A search can be performed on a single mass +/- error, either in amu (Da) or ppm. The search can be performed ignoring multicomponent structures and ignoring isotopes. An example search such as 228.1 +/- 5 ppm gives 215 results in a Table or Tile format. Selection of the single component or ignore isotopes filters means that only chemicals containing one component will be returned (i.e. no multi component mixtures, no compounds with solvents of hydration, etc.) Pre-selection of the isotope filter removes all isotopically labeled compounds from the search results.

Generate Molecular Formulae

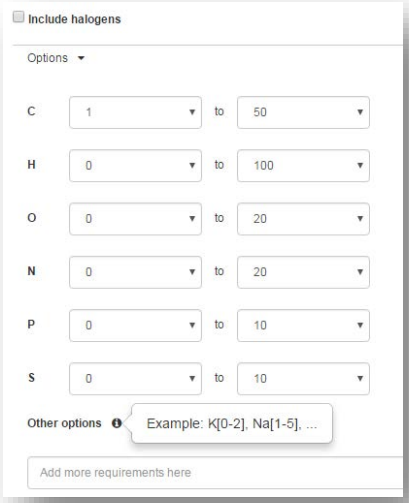
Generating one or more molecular formulae from a mass and error (a maximum error of 0.05 amu is allowed) generates all possible formulae depending on the elemental option settings and then ranks these formulae based on the number of matching hits in the database. For example, the search below for 300 amu +/- 5ppm produces a hit list of 657 formulae with only a small list of associated matches, eight formulae in total, as shown below.





Selecting on a particular formula will open a new tab with the associated chemical structures shown. It is possible to View All Hits also by selecting the button (top right hand side)

The pre-defined options for the molecular formulae are as shown below. The default options ignore halogens. Other elements can be included by adding to the options input box.

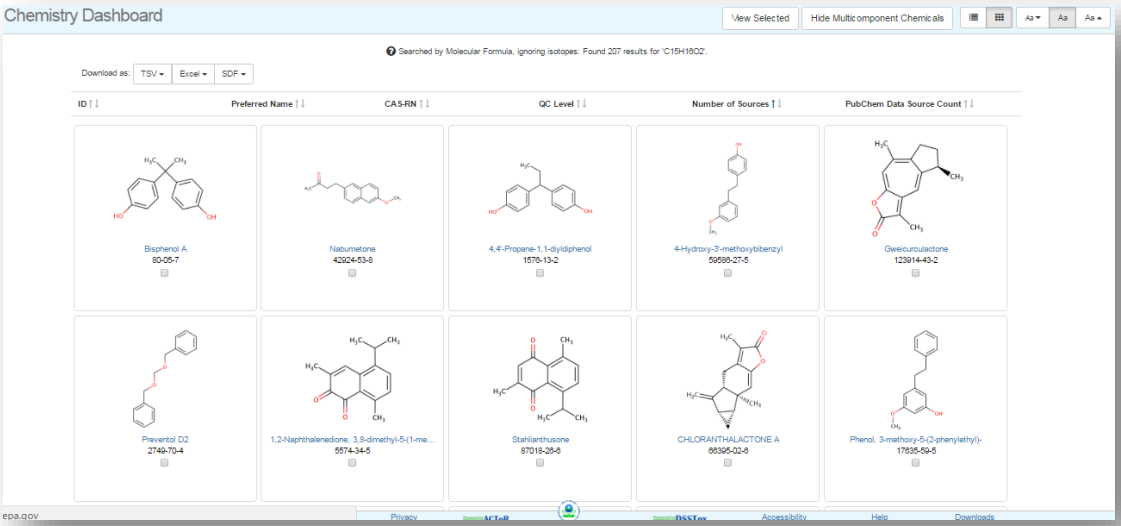


Molecular Formula Search

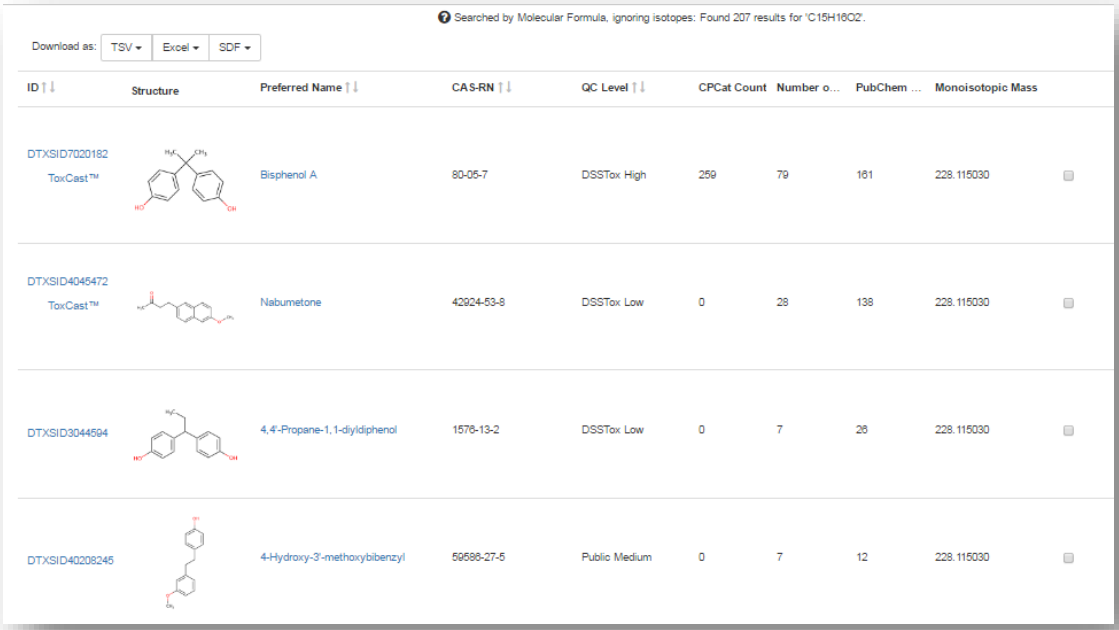
This search is primarily incorporated to allow a user to search the database for chemicals matching a particular molecular formula. This search will be of particular value for non-targeted analysis users who have already generated a potential molecular formula from an unknown observed chemical. As with a mass-based search it can be performed on a molecular formula for a single component and multicomponent structures can be ignored. An example search such as C15H16O2 yields 207 results.

Advanced Search Results Display

The results from a hit list can be shown as either a Table or in Tile form. The tile form of the hit list form is shown below. The tile list can be pruned using the check box below the structure and choosing the “View Selected” button. The hit list can be downloaded in the form of a TSV file, an Excel spreadsheet or an SDF file.



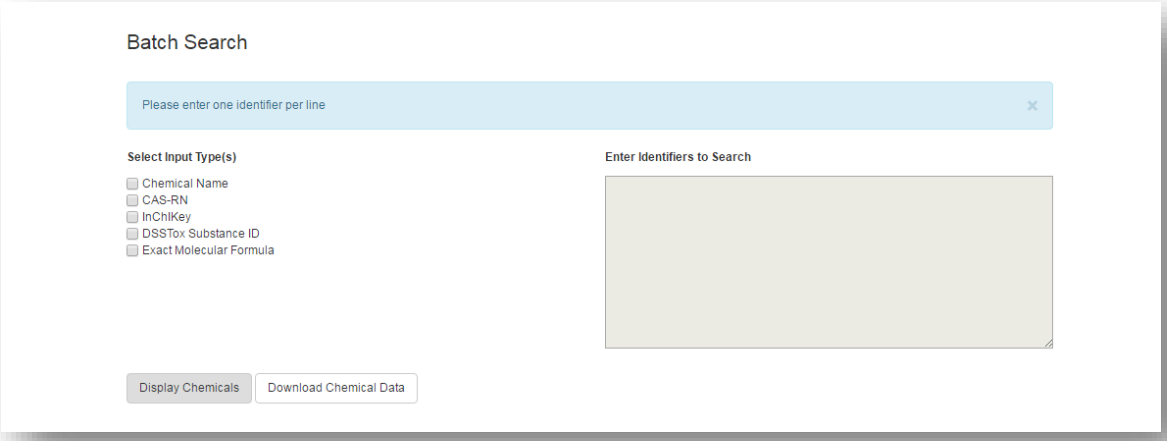
Using the buttons on the top right hand side allows the user to switch between table and tile mode and to show or hide multicomponent chemicals.



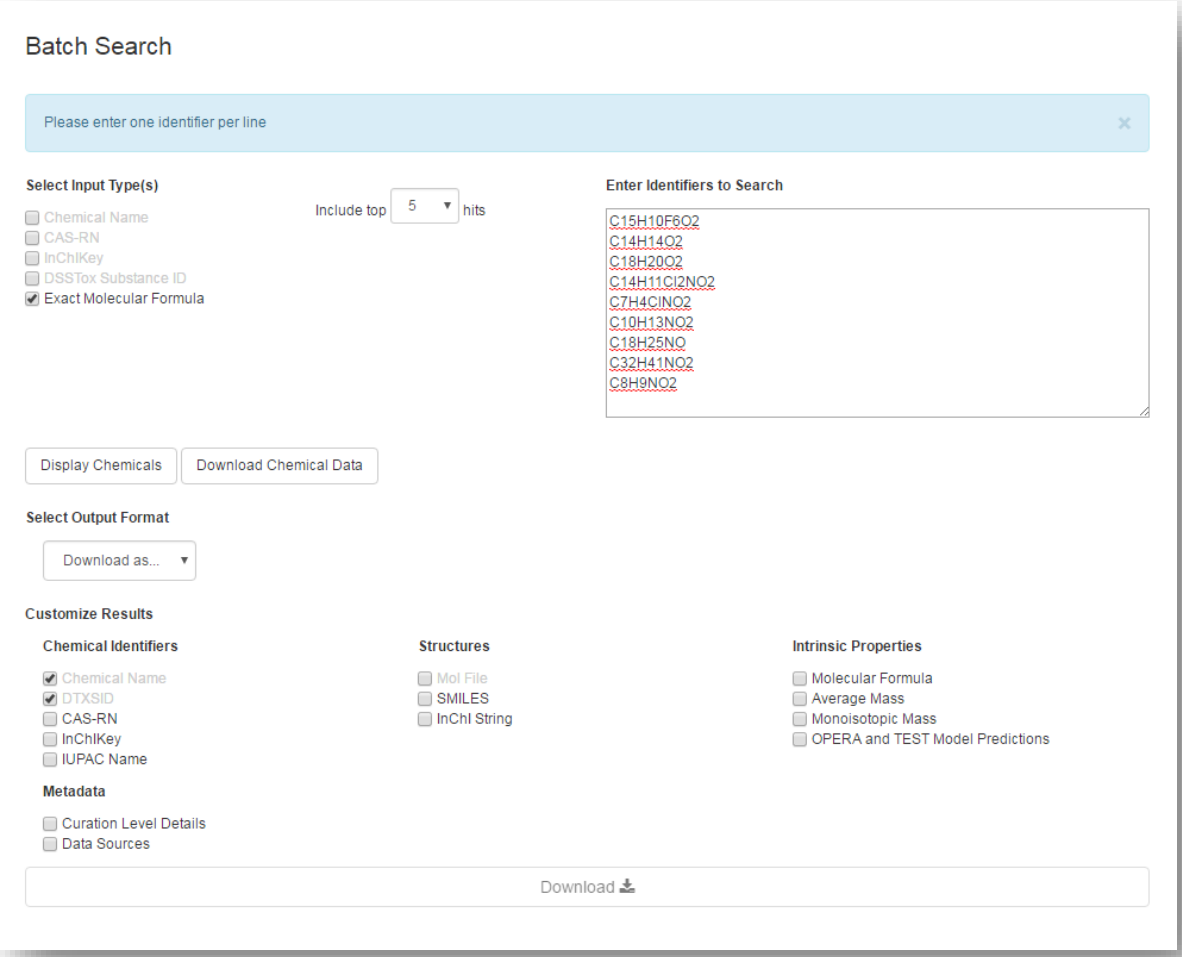
Data displayed for each record returned in the Search Results include DTXSID, structure, Preferred Name, and CAS. Additional valuable data for suspect screening and non-targeted analysis are the number of data sources, number of PubChem data sources, and the product occurrence count (CPCat Count) of each chemical returned in the results. It has been documented that ordering the number of data sources from high to low, the most likely chemicals observed in a sample rise to the top of the search results list (McEachran *et al*, 2016). Results can be downloaded, sorted, and investigated one a time. Additional data displayed as the ToxCast symbol on the left hand side indicates whether or not ToxCast data is available for the record. Hovering on the label will display the overall percentage of active assays out of the total for that chemical (Bisphenol A says “14% Active Calls: 174 Active Assays/1220 Total”).

Batch Searching

An additional manner in which to search the Dashboard is through Batch Searching, where a user can enter lists of many chemicals at a time and return results.



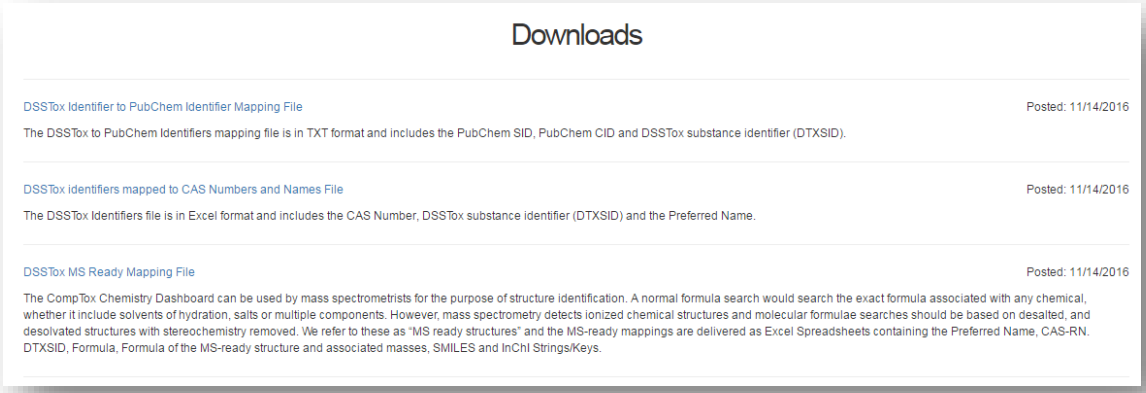
A user can search based on a list of CAS Numbers, Chemical Names, InChIKeys, DSSTox Substance Identifiers (DTXSIDs) or molecular formulae. Mixed searches, for example a mixture of names and CAS Numbers, is possible by selecting the input type and pasting a list of identifiers into the entry box. The user then selects whether they wish to display the chemicals hit list or whether to download the chemical data (see below for details). Of particular interest to suspect screening and non-targeted analysis is the ability to search many molecular formula(e) at once. Having identified a potential chemical formula of an unknown chemical compound, a user can enter the formula into the box on the right after checking the "Molecular Formula" box. This query searches the Dashboard for each formula entered and returns 1-10 results per formula (identified by the user) rank-ordered by the number of data sources. Results can be displayed or downloaded. When downloading results, the user selects data to be included in the download file (below). This enables the user to return the most likely candidate chemicals for a long list of unknown chemicals with associated identifiers, structures, and chemical properties.



Of further interest to the mass spectrometry community is the creation and inclusion of "MS-Ready" structures in the Dashboard. In this manner, an analyst attempting to identify unknowns searches the formula generated from spectra against de-salted, neutral, single component chemicals ("MS-Ready"). As one MS-Ready structure may map to several substances, the download of this file can rapidly expand to large files.

Data Downloads

Various types of data contained within the database can be downloaded via the Downloads Page listed in the website footer or via this URL.



Downloads

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

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

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.

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.

Future Work

Future versions of the CompTox Dashboard will include addition of the following functionality:

- Structure, substructure and similarity searching
- Additional visualization widgets for displaying search results
- Real-time prediction of properties
- Categorization of chemicals using various forms of tagging and ontology-based classification