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

Regional Seminar

Antony Williams

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

August 2018

Regional Seminar, Boston

Contact

Antony Williams
US EPA Office of Research and Development
National Center for Computational Toxicology (NCCT)
Williams.Antony@epa.gov
919-541-1033

ORCID: https://orcid.org/0000-0002-2668-4821
LinkedIn: https://www.linkedin.com/in/antonywilliams/
Scholar: https://scholar.google.com/citations?user=O2L8nh4AAAAJ
Wikipedia: https://en.wikipedia.org/wiki/Antony_John_Williams

About Me

- NMR Spectroscopist (PhD London 1988)
- Postdoc (National Research Council Canada)
- NMR Facility Manager – Kodak 5.5 yrs
- Chief Science Officer – ACD/Labs 10 yrs
- Consultant & ChemSpider development – 2 yrs
- VP Strategic Development at RSC – 5.5 yrs
- EPA – 3 yrs

As you can see…

- Not trained as a toxicologist
- First opportunity for a multi-hour training
- Some modules are still in beta mode – we released the new version 2 weeks ago
  - GenRA is in beta and presently offline
  - Our Submit Comments layer is awaiting an update

- There are no silly questions
- I might not have all of the answers

2:6 Fluorotelomer alcohol

- “How can the dashboard be used to characterize risks from exposure to 2:6 Flourotelomer alcohol (CASRN 647427)?”

- The issues of Chemical Names and CAS RNs

Names and CASRNs

- In our domain most chemicals are text – chemical names and CAS Numbers

Attachment D (Method 3)
SIM quantitation ions and qualifiers for internal standards, references method analysis, and surrogates

Name of Compound	CAS No.	Quantitation Ion	Qualifier Ions
Phenol-d6 (SS)	13187-88-3	99	71, 42
Phenol	108-95-2	94	66
1,4-Dichlorobenzene	106-46-0	146	111. 75, 50
Acetophenone	98-86-2	105	77, 51, 120
Acenaphthene-d10 (IS)	15067-26-2	162	160, 80
p-Cresol	106-44-5	107	108, 77
Isophorone	78-59-1	82	138, 54
Camphor	76-22-2	95	81, 108, 152
Isoborneol	124-76-5	95	110, 121, 136
Menthol	89, 78, 1	71	81, 123, 138
Naphthalene	91-20-3	128	102, 51
Methyl salicilate	119-36-8	120	92, 152, 65

Generally problematic…

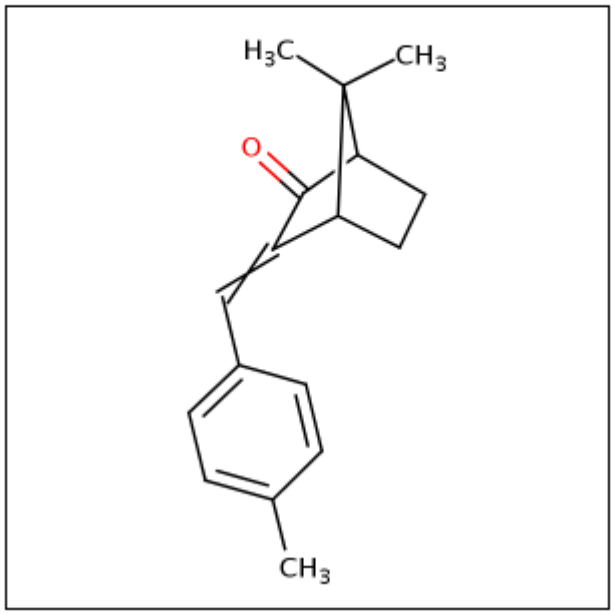
Name of Compound	CAS No.
Phenol-d6 (SS)	13187-88-3
Phenol	108-95-2
1,4-Dichlorobenzene	106-46-0
Acetophenone	98-86-2
Acenaphthene-d10 (IS)	15067-26-2
p-Cresol	106-44-5
Isophorone	78-59-1
Camphor	76-22-2
Isoborneol	124-76-5
Menthol	89, 78, 1
Naphthalene	91-20-3
Methyl salicilate	119-36-8

Active vs Deleted CASRN

Enzacamene

36861-47-9 | DTXSID8047896

Searched by Approved Name.



Synonym	Quality
Enzacamene	Valid
7,7-Dimethyl-3-[(4-methylphenyl)methylidene]bicyclo[2.2.1]heptan-2-one	Valid
Bicyclo[2.2.1]heptan-2-one, 7,7-dimethyl-3-[(4-methylphenyl)methylene]-	Valid
36861-47-9 Active CASRN	Valid
Bicyclo[2.2.1]heptan-2-one, 1,7,7-trimethyl-3-[(4-methylphenyl)methylene]-	Valid
EINECS 253-242-6	Other
Eusolex 6300	Other
Uvinul MBC 95	Other
Parsol 5000	Other
UNII-8I3XWY40L9	Other
4-Methylbenzylidenecamphor	Other
p-Methylbenzylidenecamphor	Other
38102-62-4 Deleted CASRN	Deleted
84055-65-2 Deleted CASRN	Deleted

This one has 316 Deleted CASRN

CAS Registry Number: 25068-38-6
$(C_{15}H_{16}O_2 . C_3H_5ClO)_x$
Phenol, 4,4'-(1-methylethylidene)bis-, polymer with 2-(chloromethyl)oxirane

Polymer
Polymer Class Terms: Epoxy resin
Alternate CAS Registry Numbers: 26402-79-9
Deleted CAS Registry Numbers: 1336-88-5, 1337-15-1, 8000-31-5, 9015-99-0, 9049-54-1, 9050-21-9, 9081-91-8, 9083-76-5, 9084-94-0, 9086-62-8, 9087-26-7, 9087-76-7, 11097-80-6, 11098-13-8, 11098-40-1, 11100-23-5, 11108-41-1, 11120-31-3, 11121-19-0, 11126-36-6, 20232-24-0, 35038-60-9, 36344-96-4, 36347-13-4, 36347-14-5, 37184-50-2, 37184-52-4, 37208-29-0, 37217-92-8, 37230-74-3, 37243-66-6, 37243-67-7, 37251-33-5, 37265-21-7, 37270-82-9, 37291-75-1, 37293-07-5, 37294-18-1, 37305-82-1, 37307-45-2, 37317-45-6, 37325-21-6, 37338-63-9, 37342-17-9, 37345-34-9, 37348-56-4, 37348-57-5, 37357-73-6, 37360-93-3, 39277-59-3, 39288-99-8, 39296-08-7, 39296-09-8, 39296-11-2, 39296-15-6, 39315-77-0, 39349-91-2, 39354-86-4, 39362-25-9, 39362-45-3, 39373-81-4, 39378-29-5, 39378-55-7, 39389-49-6, 39405-18-0, 39412-57-2, 39419-66-4, 39453-22-0, 39454-54-1, 39454-69-8, 39470-62-7, 42612-34-0, 42618-03-1, 50642-36-9, 50642-55-2, 50642-78-9, 51158-20-4, 51273-81-5, 51329-73-8, 51393-99-8, 51394-03-7, 51553-00-5, 52011-87-7, 52038-45-6, 52051-70-4, 52051-82-8, 52052-16-1, 52232-05-0, 52232-75-4, 52276-55-8, 52365-33-0, 52519-66-1, 52519-67-2, 52627-94-8, 52907-38-7, 53027-88-6, 53127-14-3, 53200-30-9, 53238-86-1, 53238-87-2, 53239-67-1, 53239-68-2, 53570-97-1, 53570-98-2, 53681-78-0, 53858-93-8, 54018-73-4, 54352-05-5, 55464-96-5, 55584-55-9, 55585-07-4, 55818-73-0, 56258-35-6, 56449-43-5, 56509-48-9, 57107-66-1, 57284-90-9, 57534-21-1, 57693-04-6, 58052-05-4, 58128-38-4, 58392-89-5, 58392-92-0, 58516-14-6, 58572-71-7, 59029-19-5, 59459-14-2, 59473-30-2, 59948-36-6, 60202-19-9, 60267-31-4, 60382-89-0, 60606-56-6, 60800-54-6, 60831-77-8, 60894-16-8, 61036-82-6, 61287-42-1, 61356-27-2, 61711-38-4, 61763-30-2, 61991-18-2, 62169-28-2, 62169-29-3, 62601-75-6, 62601-76-7, 62887-23-4, 63055-40-3, 63172-55-4, 63799-24-6, 63993-57-7, 63993-58-8, 64086-14-2, 64086-16-4, 64176-52-9, 64176-61-0, 64176-66-5, 64177-03-3, 65233-49-0, 65931-38-6, 65931-39-7, 66995-96-8, 67185-62-0, 68821-97-6, 69899-40-7, 70179-83-8, 70213-44-4, 70726-45-3, 71965-91-8, 72514-40-0, 73413-19-1, 74504-20-4, 74564-76-4, 75831-44-6, 78564-77-9, 79585-43-6, 80702-61-0, 81458-12-0, 81843-57-4, 81843-58-5, 81855-87-0, 82197-12-4, 82197-46-4, 83202-85-1, 84286-97-5, 84683-04-5, 84931-29-3, 85537-69-5, 86090-60-0, 88385-37-9, 88528-19-2, 88651-18-7, 89750-00-5, 91727-28-5, 91727-29-6, 92481-37-3, 95327-25-6, 96420-31-4, 96510-68-8, 97568-16-6, 97709-01-8, 99400-50-7, 101027-12-7, 102256-87-1, 103599-13-9, 103599-14-0, 104364-97-8, 104491-99-8, 105521-57-1, 106207-08-3, 106856-89-7, 107991-47-9, 108556-05-4, 108728-21-8, 110158-22-0, 111367-08-9, 111517-59-0, 114013-37-5, 115902-32-4, 117216-90-7, 117313-45-8, 117786-92-2, 118340-04-8, 120146-74-9, 120797-43-5, 121181-85-9, 121273-37-8, 121547-73-7, 123939-44-6, 125147-87-7, 127176-80-1, 127176-81-2, 128281-71-0, 132822-20-9, 132893-73-3, 135976-90-8, 137545-29-0, 138157-20-7, 138361-18-9, 139554-29-3, 142540-11-2, 144046-24-2, 144046-25-3, 144855-66-3, 149013-58-1, 150825-32-4, 157321-42-1, 157481-46-4, 158725-45-2, 160674-45-3, 161937-12-8, 162031-55-2, 167972-06-7, 168042-08-8, 179607-24-0, 183581-68-2, 183890-12-2, 187619-11-0, 188448-56-8, 189282-49-3, 191606-83-4, 220090-06-2, 222835-65-6, 222835-66-7, 222835-68-9, 222835-69-0, 222835-70-3, 222835-72-5, 222835-74-7, 222835-77-0, 309945-96-8, 339530-81-3, 353239-57-3, 367523-08-8, 383889-26-7, 383889-27-8, 395069-05-3, 470462-49-8, 681001-41-2, 848887-61-6, 913745-83-2, 917483-69-3, 922728-11-8, 934588-09-7, 945610-97-9, 950907-45-6, 1033821-54-3, 1034342-45-4, 1068160-75-7, 1082736-74-0, 1096473-97-0, 1114797-08-8, 1189565-70-5, 1190235-62-1, 1190729-68-0, 1192045-32-1, 1195324-26-5, 1196030-95-1, 1198291-96-1, 1199811-18-1, 1203835-26-0, 1206700-05-1, 1228639-00-6, 1245563-83-0, 1271727-39-9, 1300093-58-6, 1300102-07-1, 1305321-17-8, 1338071-08-1, 1446691-72-0, 1450839-98-1, 1620807-39-7, 1641551-32-7, 1807886-28-7, 1815624-46-4, 1815624-47-5

https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID0050479

Hazard Data from “ToxVal_DB”

- ToxVal Database contains following data:
  - 30,050 chemicals
  - 772,721 toxicity values
  - 29 sources of data
  - 21,507 sub-sources
  - 4585 journals cited
  - 69,833 literature citations

Chlorobenzene

- How can the dashboard assist in developing an aquatic life criterion for chlorobenzene?
- Can the dashboard supplement Ecotox database and other sources of data available in the literature?
- How can QSAR and Read-Across approaches for similar chemicals apply to fish and aquatic invertebrates?
- Can the dashboard provide insight into environmental fate and transport to reduce uncertainty concerning attenuating in the subsurface?

Chlorobenzene

- The Hazard Table contains ECOTOX data and other data via Los Alamos National Lab
- We have experimental and predicted environmental fate and transport data

- Show OPERA Models and TEST real time predictions

2:6 Fluorotelomer alcohol

- “How can the dashboard be used to characterize risks from exposure to 2:6 Flourotelomer alcohol (CASRN 647427)?”

- We have
  - Hazard Data
  - Exposure Predictions
  - Production Volume Data

Predicted Exposure Data

High Throughput Heuristics for Prioritizing Human Exposure to Environmental Chemicals

John F. Wambaugh*†, Anran Wang†§‖, Kathie L. Dionisio‡, Alicia Frame†‖, Peter Egeghy‡, Richard Judson†, and R. Woodrow Setzer†

†National Center for Computational Toxicology, and ‡National Exposure Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711, United States

§ North Carolina State University, Department of Statistics, Raleigh, North Carolina 27695-8203, United States

‖ Oak Ridge Institute for Science and Education Grantee, P.O. Box 117, Oak Ridge, Tennessee 37831-0117, United States

Environ. Sci. Technol., **2014**, *48* (21), pp 12760–12767

DOI: 10.1021/es503583j

Cite this: *Environ. Sci. Technol.* 48, 21, 12760-12767

- Five descriptors are capable of explaining roughly 50% of the variability in geometric means across 106 NHANES chemicals for all the demographic groups, including children aged 6–11

3-trifluoromethyl-4-nitrophenol

- TFM (3-trifluoromethyl-4-nitrophenol)
- Potential exposures through drinking water pathway. *In vivo* toxicology data are limited, 90d sub-chronic rat study underway.
- An RfD is available on the dashboard-how was it derived and can it be used to set a drinking water standard?

Data is under constant curation

- Our data is under constant curation and with this release there is no RfD listed under the Hazard Tab
- There are LOAEL data and details under the Modal

Details Under the Modal

ToxRefDB Details | Long Details

Download as: TSV | Excel

grouping ID	Study ID	study type	effect category	toxicity value type	qualifier	effect level	units	species	strain group	strain	exposure route	exposure method	reference	year	study source	data usability	guideline	purity
5660	428	SUB	systemic	lel	=	45.0	mg/kg/day	rat	sprague dawley	Sprague Dawley	Oral	Feed	US Dept of the Interior, Study Conducted by WARF Institure, Subchronic oral toxicity study in rats of TFM. (1971)	1971	opp_der	acceptable	Subchronic oral toxicity in rodents	90
5660	428	SUB	systemic	loael	=	145.8	mg/kg/day	rat	sprague dawley	Sprague Dawley	Oral	Feed	US Dept of the Interior, Study Conducted by WARF Institure, Subchronic oral toxicity study in rats of TFM. (1971)	1971	opp_der	acceptable	Subchronic oral toxicity in rodents	90
5660	428	SUB	systemic	nel	=	25.0	mg/kg/day	rat	sprague dawley	Sprague Dawley	Oral	Feed	US Dept of the Interior, Study Conducted by WARF	1971	opp_der	acceptable	Subchronic oral toxicity in rodents	90

OPERA MODELS AND DATA

Online Today

EPI Suite Data - ISIS/Base & SDF

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

... *Updated September 15, 2010*

Basic Instructions:

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

NOTE ... zipped files extract to **Folders** containing the individual data files ... Folders named **EPI_ISIS_Data** and **EPI_SDF_Data**

Substructure Searching Files:

ISIS™/Base & SD Files of the EPI Suite Program Experimental Data Files are now available ... The ISIS™/Base files require the commercial program for use ... The SD Files can be imported into other commercial chemical structure programs (such as ChemFinder).

... **Click here to download EPI_ISIS_Data.zip** ... (about 11 MB)

... **Click here to download EPI_SDF_Data.zip** ... (about 10 MB)

NOTE ... **EPI Suite Data Files (some in Excel, Text, Word format) available at:**

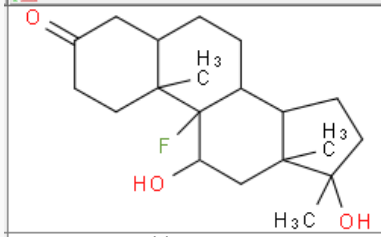
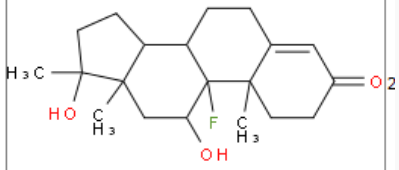
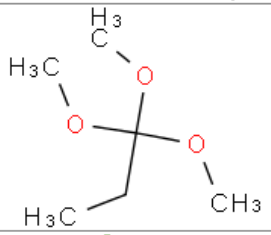
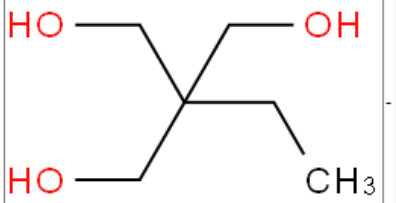
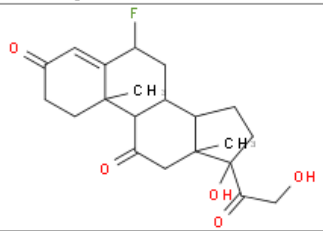
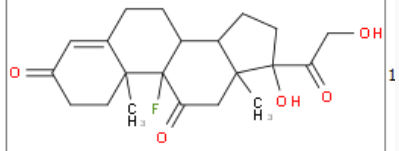
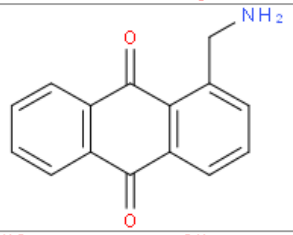
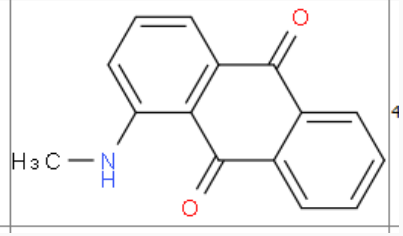
http://esc.syrres.com/interkow/EpiSuiteData.htm

We Curated These Public Data to Build Prediction Models

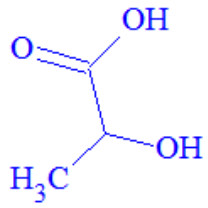
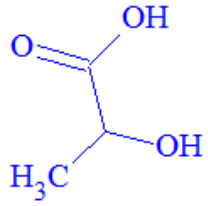
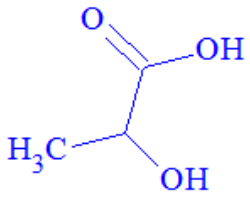
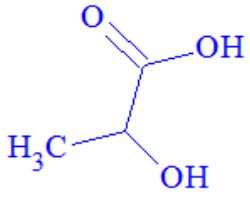
EPA United States Environmental Protection Agency

Public data should be curated prior to modeling

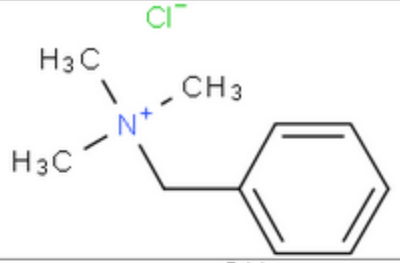
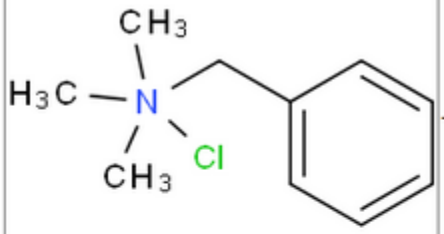
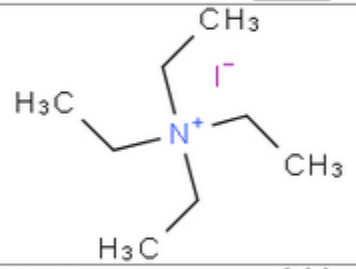
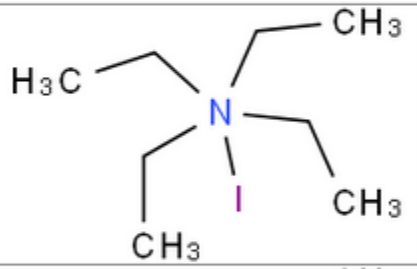
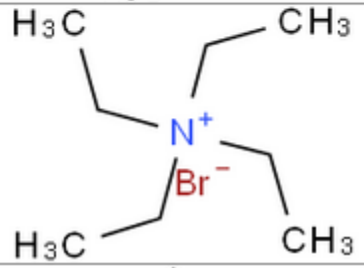
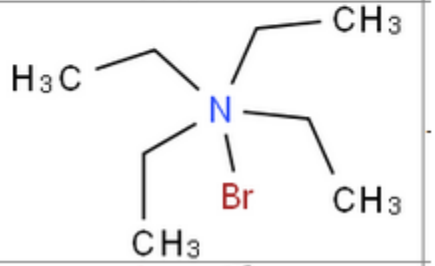
Different Compounds

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	

Duplicate Structures

Structure	Formula	FW	CAS	NAME	MP	EstMP	ErrorMP
	$C_3H_6O_3$	90.0779	000050-21-5	LACTIC ACID	1.68000000000000e+001	2.26600000000000e+001	5.86000000000000e+000
	$C_3H_6O_3$	90.0779	000079-33-4	L-LACTIC ACID	5.30000000000000e+001	2.26600000000000e+001	-3.03400000000000e+001
	$C_3H_6O_3$	90.0779	000598-82-3	A-HYDROXYPROPIONIC ACID	1.80000000000000e+001	2.26600000000000e+001	4.66000000000000e+000
	$C_3H_6O_3$	90.0779	010326-41-7	D-LACTIC ACID	5.28000000000000e+001	2.26600000000000e+001	-3.01400000000000e+001

Covalent Halogens

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

Curation to QSAR Ready Files

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

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%)

OPERA Predicted Properties

EPA United States Environmental Protection Agency

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

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

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

To link to th

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

Journal of Cheminformatics

RESEARCH ARTICLE Open Access

CrossMark

OPERA models for predicting physicochemical properties and environmental fate endpoints

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

Detailed OPERA Prediction Reports

EPA United States Environmental Protection Agency

Source	Result	Calculation Details
Experimental Values		

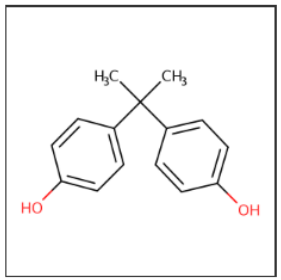
- PhysPropNCCT
- Predicted Values
- EPISUITE
- NICEATM
- ACD/Labs Conse
- ACD/Labs
- OPERA

OPERA Models: LogP: Octanol-Water

Save PDF

Bisphenol A

80-05-7 | DTXSID7020182



Model Results

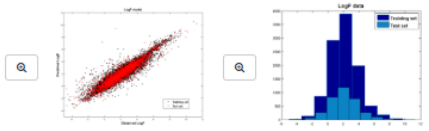
Predicted value: 3.35

Global applicability domain: Inside

Local applicability domain index: 0.88

Confidence level: 0.75

Model Performance

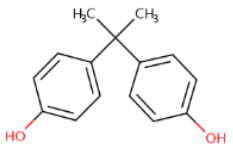


Weighted KNN model

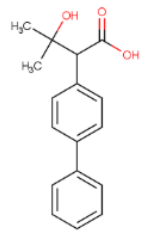
QMRF

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

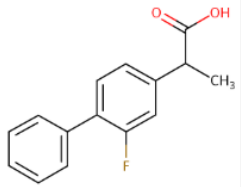
Nearest Neighbors from the Training Set



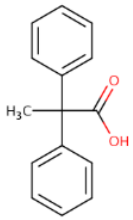
Bisphenol A
Measured: 3.32
Predicted: 3.35



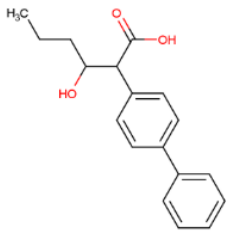
BUTANOIC ACID,2-(4-BIPHENYLYL)-3-HYDROX...
Measured: 3.25
Predicted: 3.45



Flurbiprofen
Measured: 4.16
Predicted: 3.83



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



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

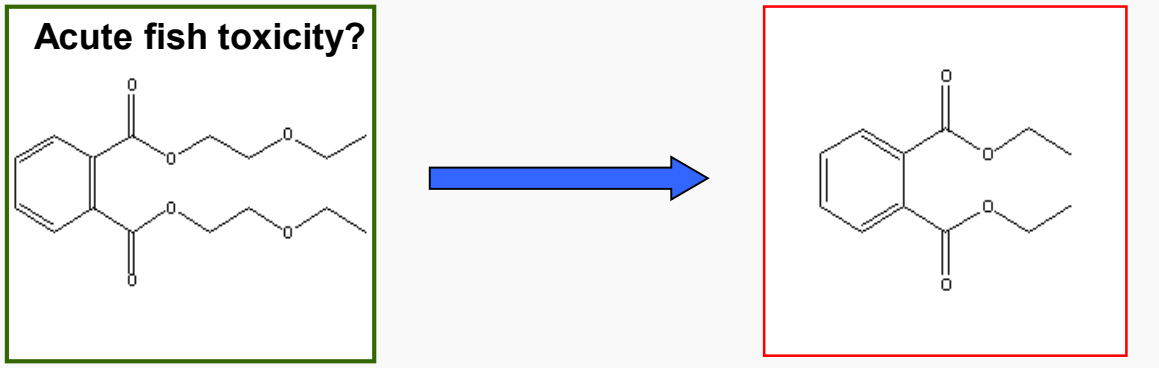
GenRA

Definitions: Read-Across

- Known information on the property of a substance (source) is used to make a prediction of the same property for another substance (target) that is considered “similar”

	Source chemical	Target chemical
Property	● →	○

● Reliable data
○ Missing data

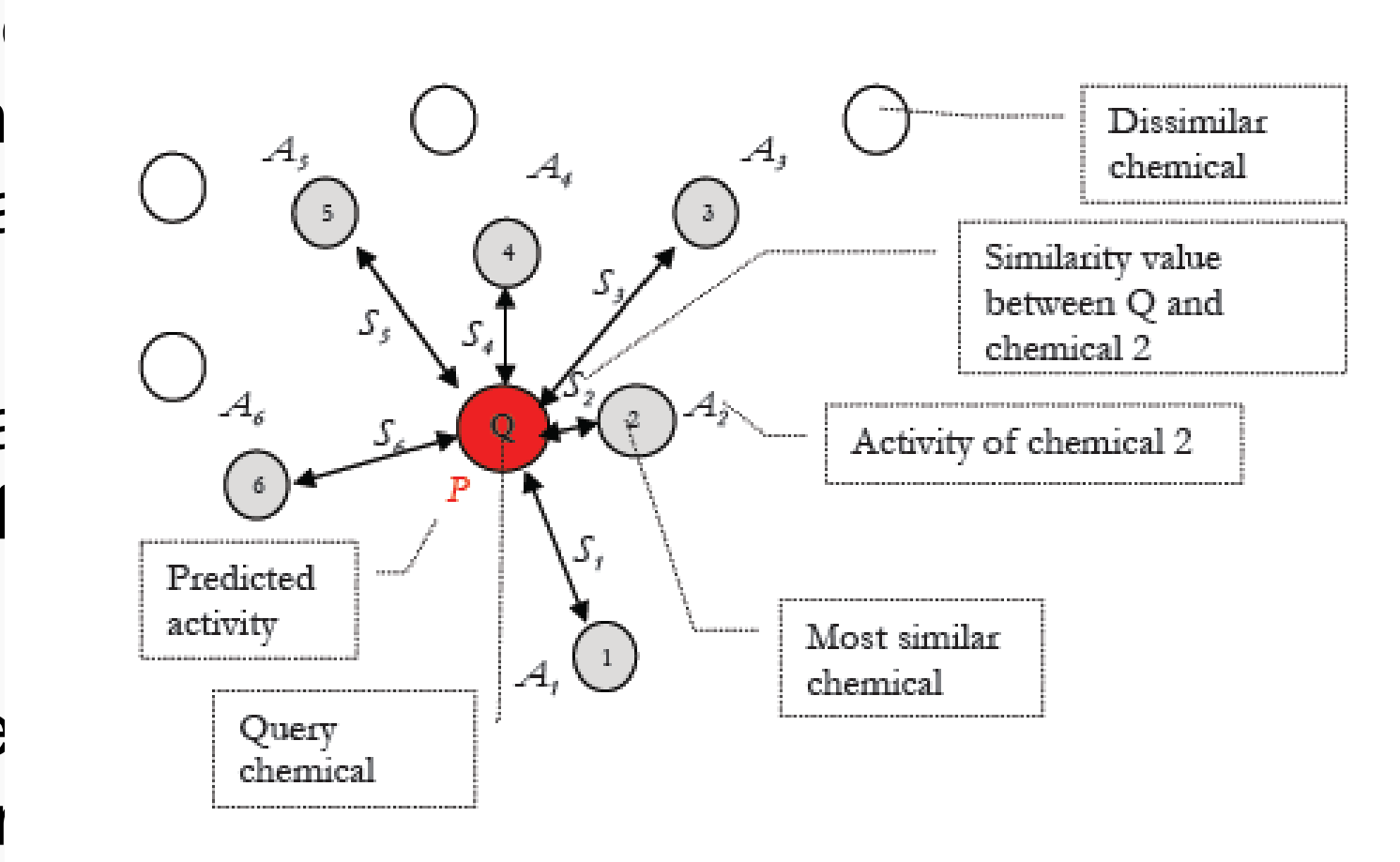


Known to be harmful

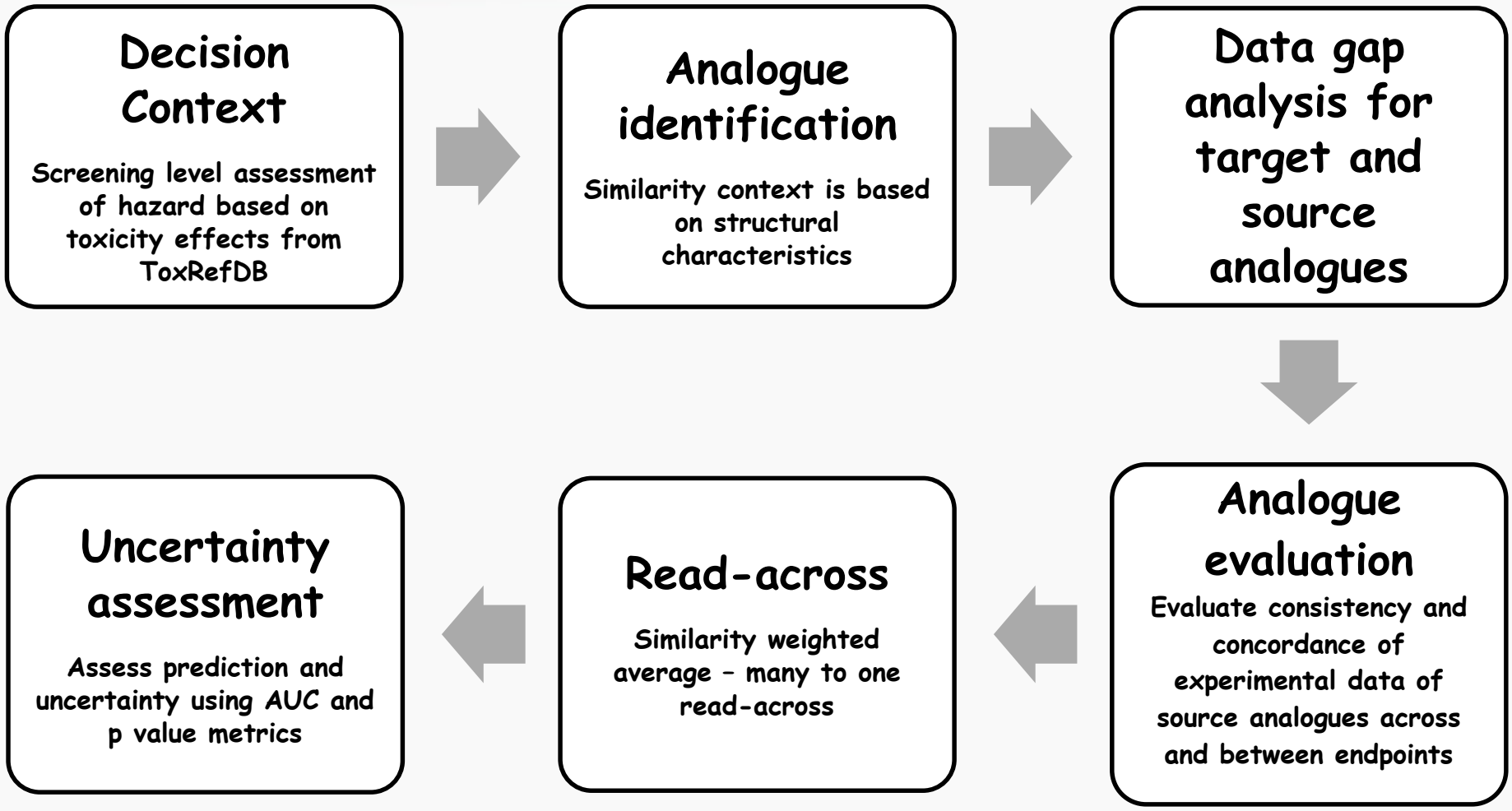
Predicted to be harmful

GenRA (Generalised Read-Across)

- Pre vity of n or bioa
- Goa data perf
- The e for r within specific study outcomes to be established

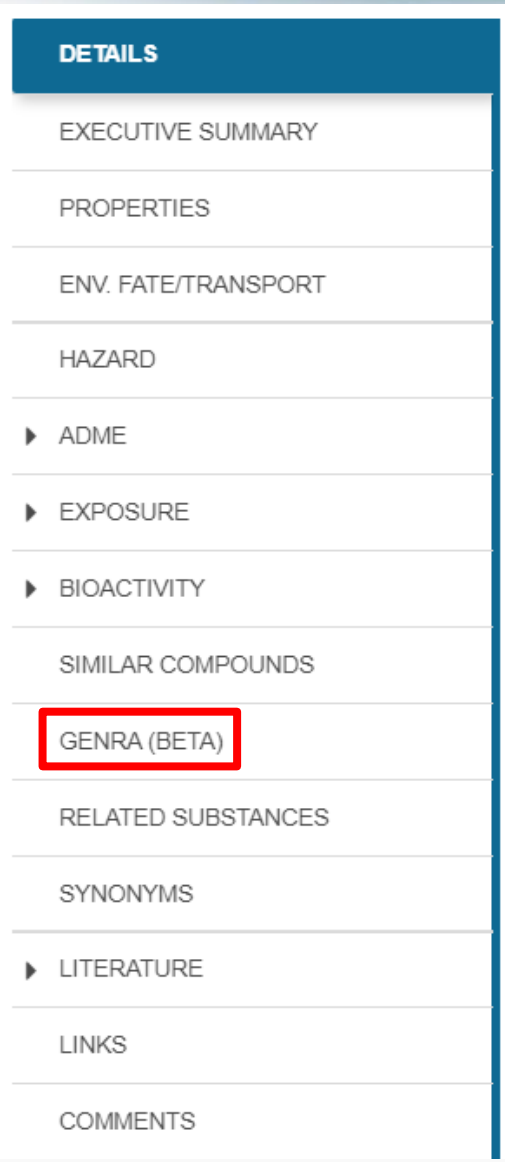


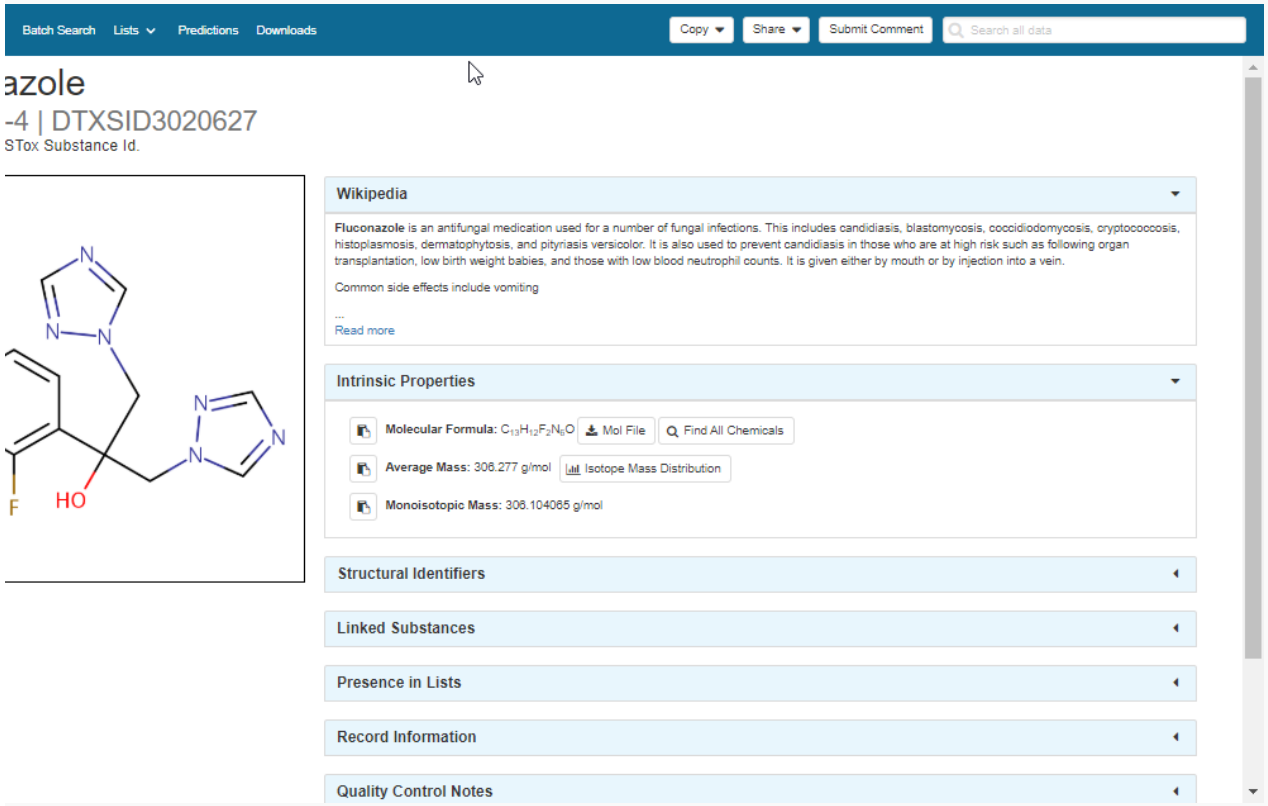
Read-across workflow in GenRA



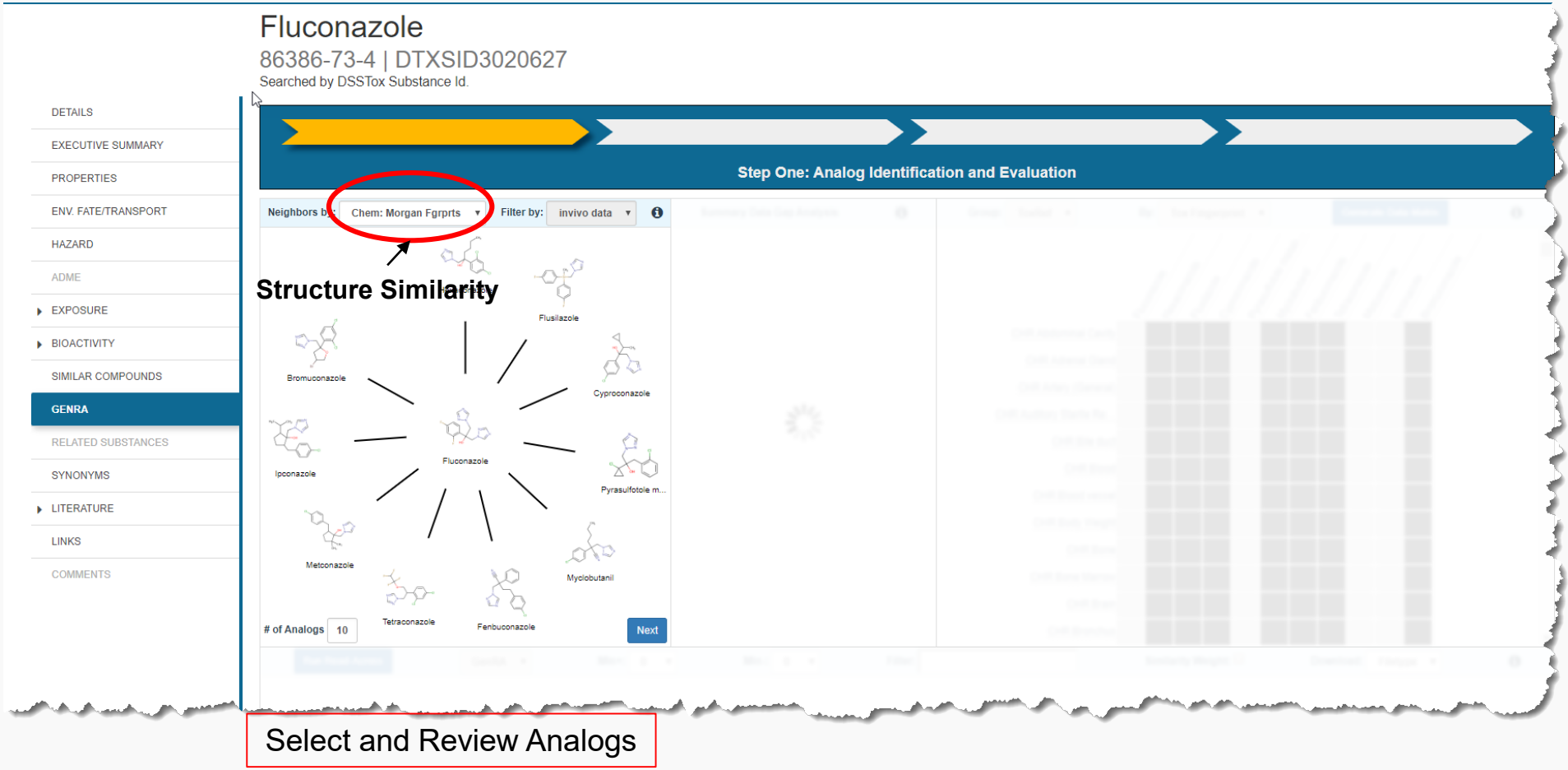
GenRA (Generalised Read-Across)

EPA United States Environmental Protection Agency





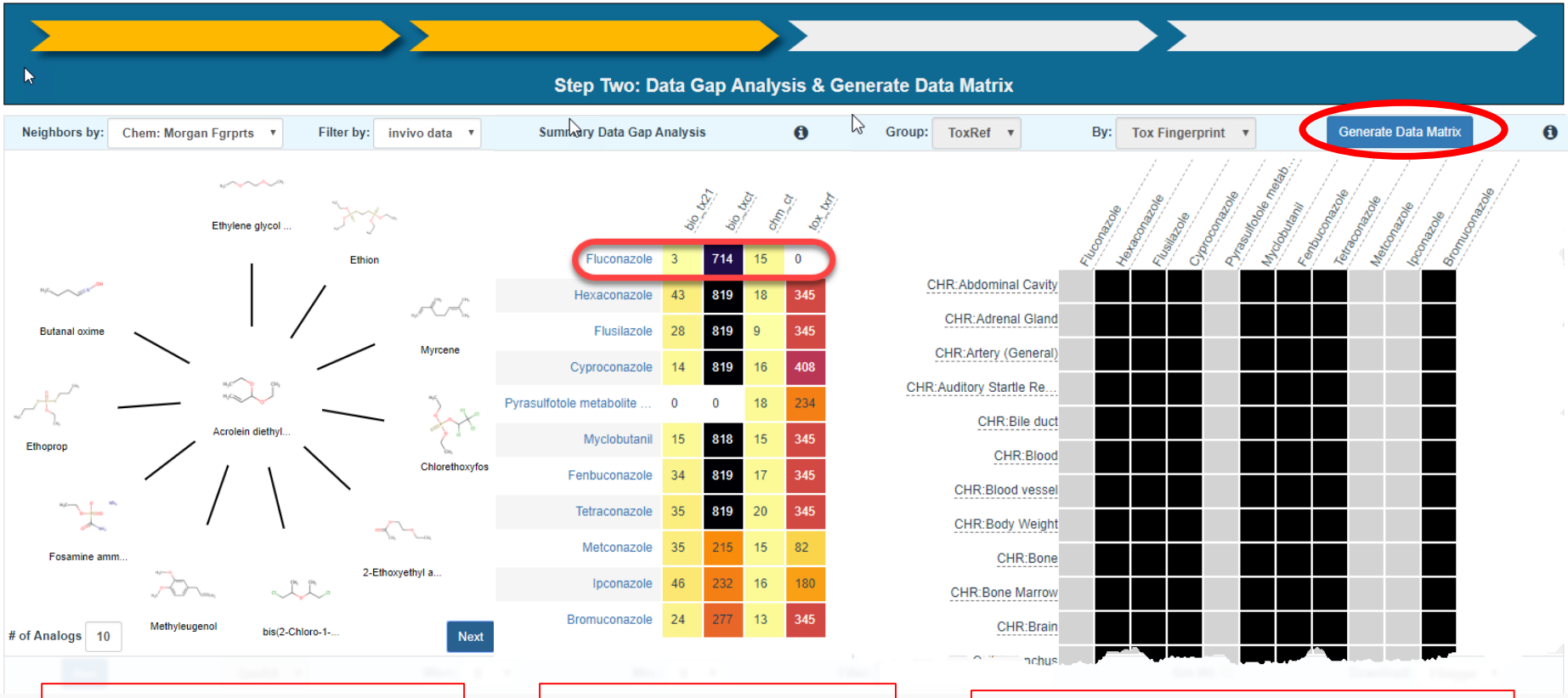
GenRA (Generalised Read-Across)



Select and Review Analogs

GenRA (Generalised Read-Across)

GenRA

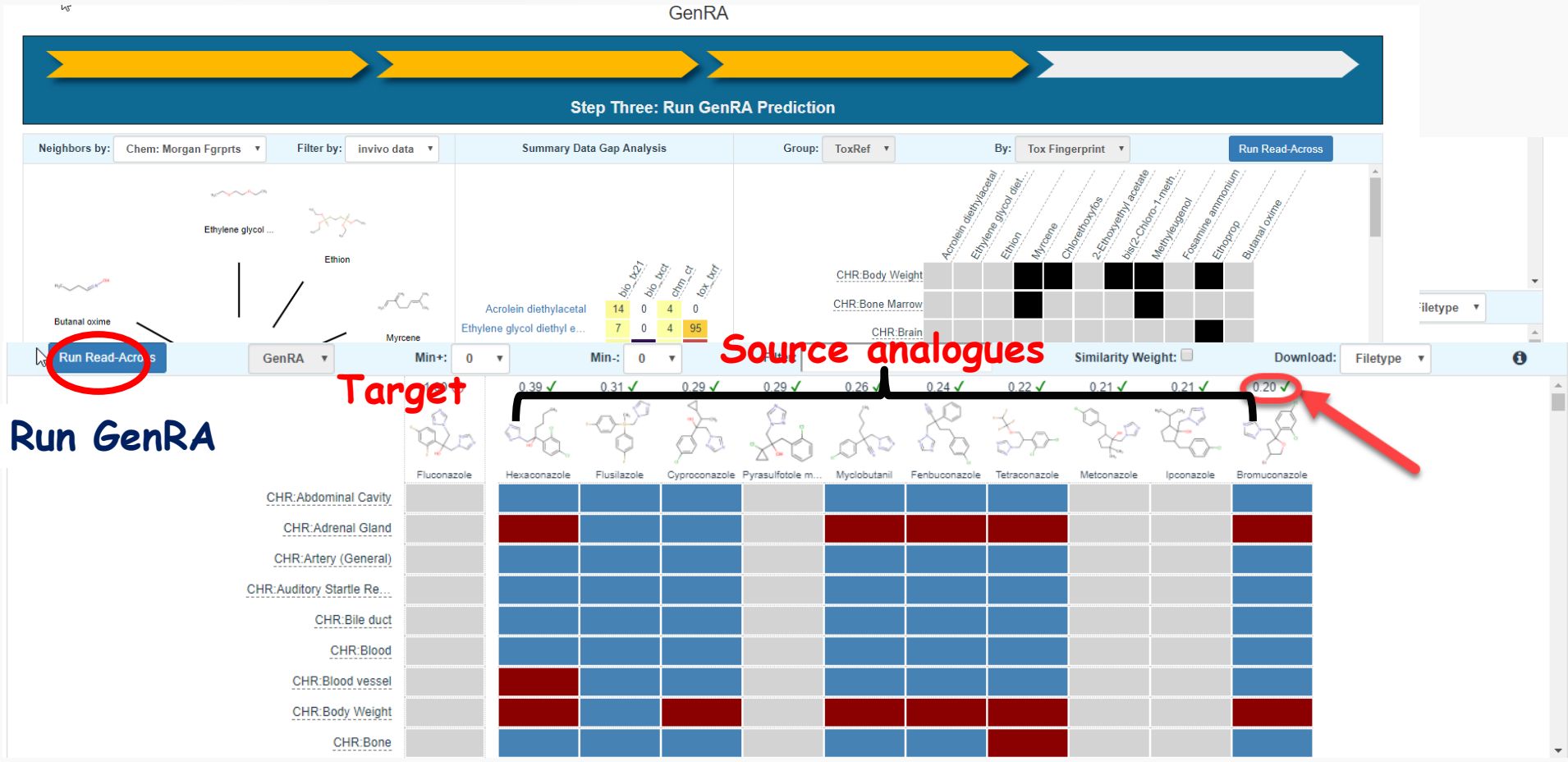


Select and Review Analogs

Review Available Data

Fingerprint indicating available data

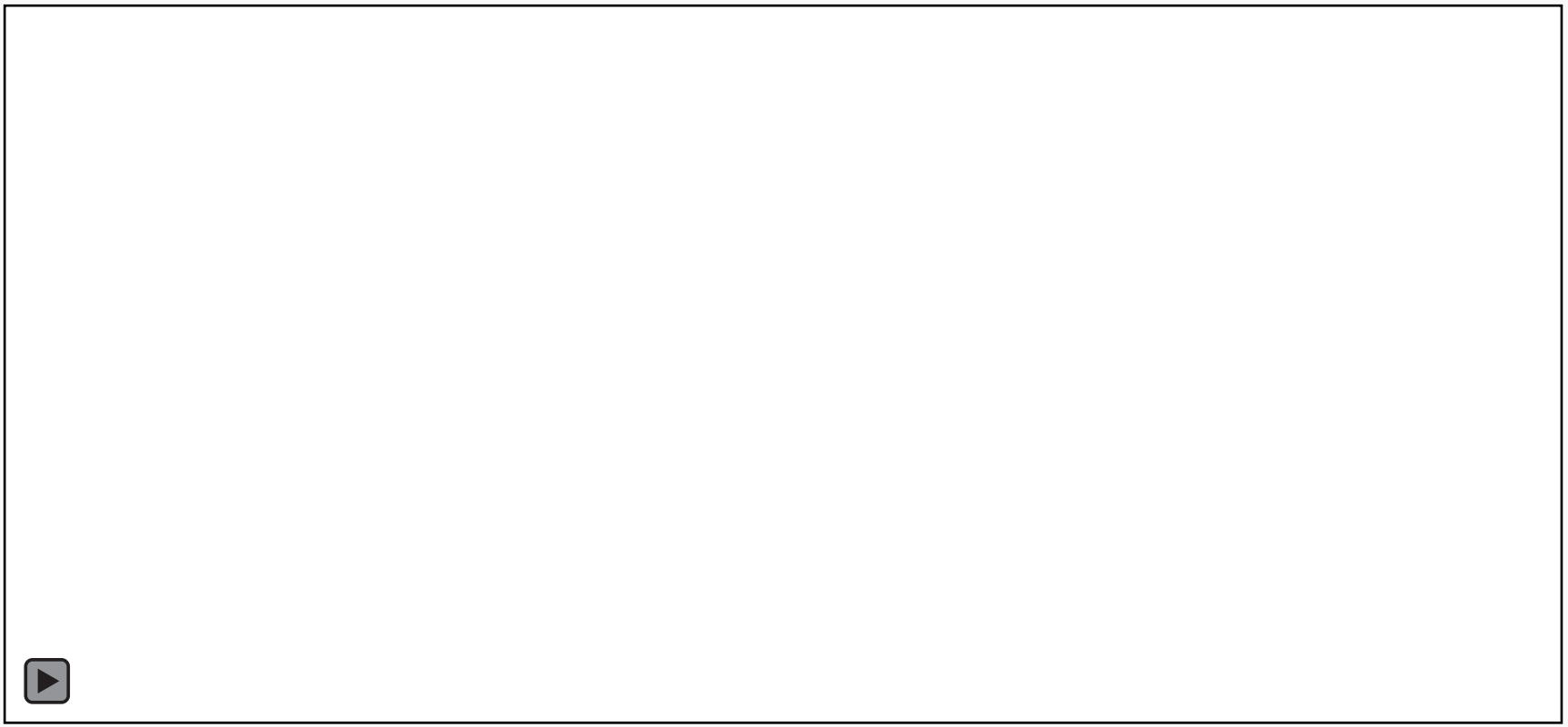
GenRA (Generalised Read-Across)



Red : Toxicity effects.
Blue: No Toxicity effects
Grey : Absence of data

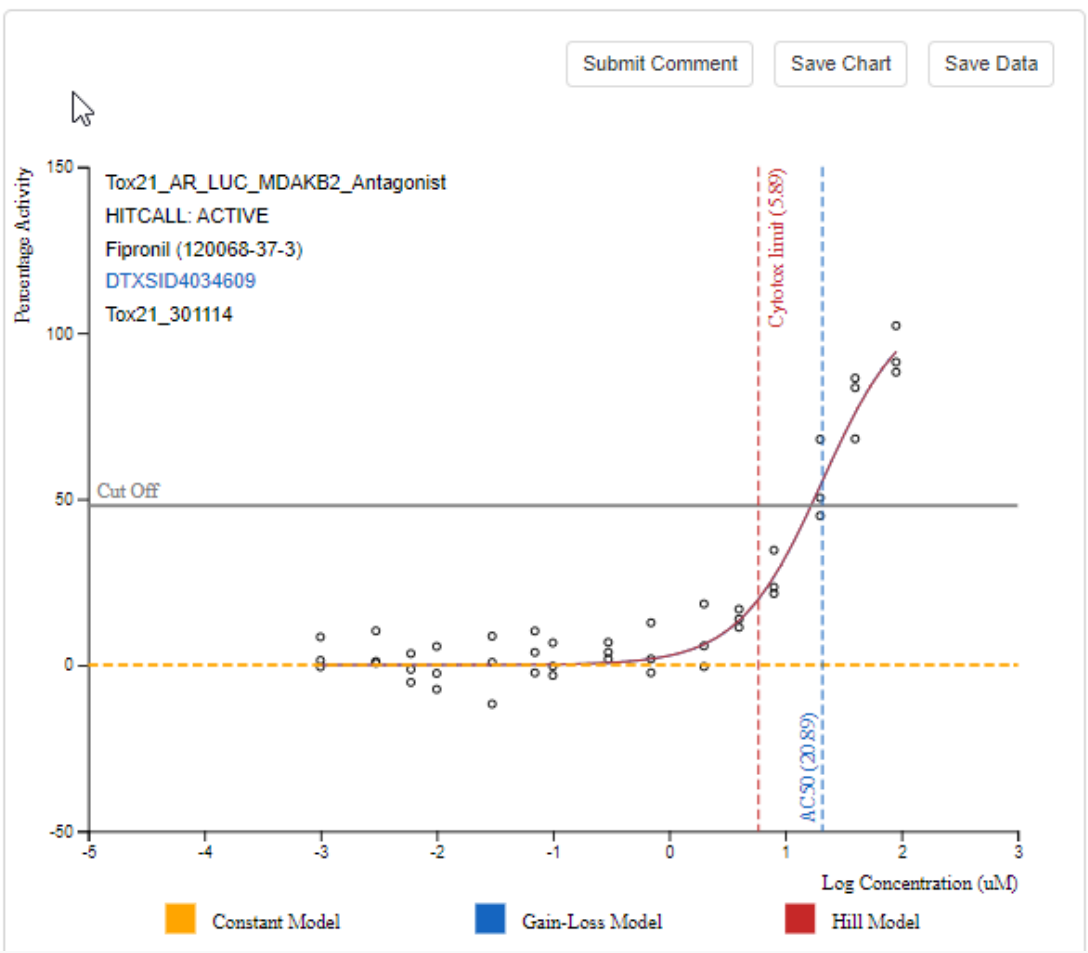
Demonstration

EPA United States Environmental Protection Agency



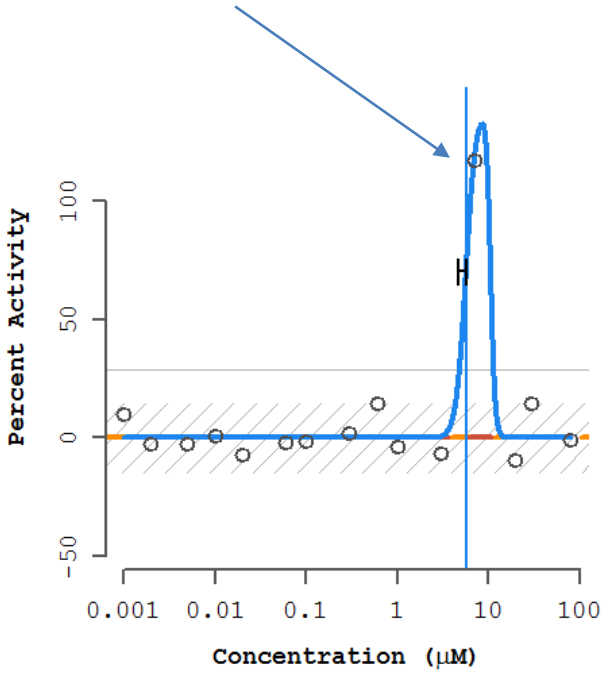
BIOACTIVITY DATA

Internal Review of 25,000 curves
A “good fit” bioactivity curve



Internal Review of 25,000 curves

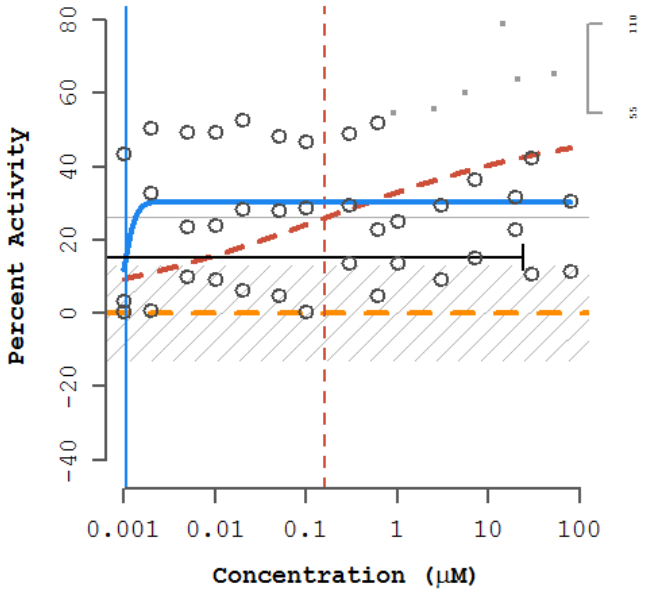
Single-Point in middle of concentration range drives ACTIVE Hit Call




```
ASSAY:    TOX21_RT_HEK293_FLO_40hr_viability

NAME:     Tributylmethylammonium bis(trifluoromethylsulfc
CHID:     47905    CASRN: 405514-94-5
SPID(S): Tox21_200582
M4ID:     19917619

HILL MODEL (in red):
       tp        ga        gw
val:   104       2.4       7.97
sd:    NaN       NaN       NaN

GAIN-LOSS MODEL (in blue):
       tp        ga        gw        la        lw
val:   140       0.758     8         1.03      18
sd:    56.9      0.105     3.92      0.822     64.2

       CNST         HILL         GNLS
AIC:   125.38       131.38       113.6
PROB:  0            0            1
RMSE:  31.04        31.04        7.1

MAX_MEAN: 117        MAX_MED: 117        BMAD: 4.76

COFF: 28.5    HIT-CALL: 1     FITC: 50    ACTP: 1

FLAGS:

HIT-PCT: 0.887  MED-GA: 0.7582  GA-CI: 0.0826
```

…and gain-loss fit with a cell viability assay (makes little sense)

Internal Review of 25,000 curves Abnormally High-Noise




```
ASSAY:    TOX21_AR_BLA_Antagonist_ratio

NAME:     1-Nitropyrene
CHID:     20983    CASRN: 5522-43-0
SPID(S): Tox21_200066
M4ID:     18181279  BRK

HILL MODEL (in red):
        tp        ga        gw
val:    52        -0.797    0.3
sd:     NaN       NaN       NaN

GAIN-LOSS MODEL (in blue):
        tp        ga        gw        la        lw
val:    30.4      -2.97     7.76      3.77      4.44
sd:     3.58      0.211     59.3      7180      16800

        CNST          HILL          GNLS
AIC:    464.32        423.78        423.68
PROB:   0             0.49          0.51
RMSE:   39.92         23.61         23.62

MAX_MEAN: 55.2       MAX_MED: 42.4       BMAD: 4.35

COFF: 26.1    HIT-CALL: 1     FITC: 46    ACTP: 1

FLAGS: 17; 11

HIT-PCT: 0.971   MED-GA: -1.5919   GA-CI: 6.2178
```

Internal Curve Review Results

- Internal curve review has resulted in:
  - Instances of correction of fitting procedures in the ToxCast Pipeline
  - Identification of issues with source data
  - Identification of additional flags or filters that could be used, depending on the application of ToxCast data
  - a beta implementation of quality assurance for HTS data
  - Brown & Paul-Friedman, **Uncertainty in ToxCast Curve-Fitting: Quantitative and Qualitative Descriptors Inform a Model to Predict Reproducible Fits** *(in preparation)*

tcpl: the ToxCast pipeline for high-throughput screening data FREE

Dayne L Filer, Parth Kothiya, R Woodrow Setzer, Richard S Judson, Matthew T Martin

Bioinformatics, Volume 33, Issue 4, 15 February 2017, Pages 618–620,
https://doi.org/10.1093/bioinformatics/btw680

Published: 22 November 2016 **Article history**

EPA United States Environmental Protection Agency

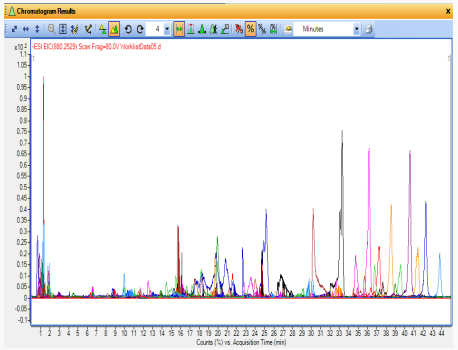
STRUCTURE ID BY MASS SPEC

Comparing Analysis Approaches

- 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!!



Structure Identification by MS

- An LC/MS/MS examination of a sol sample for tentatively identified compounds (TIC), also called suspect chemical screens, returns an exact mass of 222.22 g/mol. How can the dashboard assist in identifying and characterizing the environmental hazard of this “unknown”?

Advanced Search

- Search is based on the **Neutral Mass**
- Default error is +/-5ppm
- Search 222.22+/-5ppm – is it neutral mass??
  - 1 compound (Multicomponent)
  - Searching 67 million on ChemSpider – SAME compound
  - Common adduct searches
    - +H: 0 chemicals
    - +Na: 19 chemicals
    - +NH4: 0 chemicals
    - +K: 0 chemicals

Result

- It is a C13 linear or branched monoamine
- https://tinyurl.com/yae59vpd

- Based on sources maybe tridecylamine

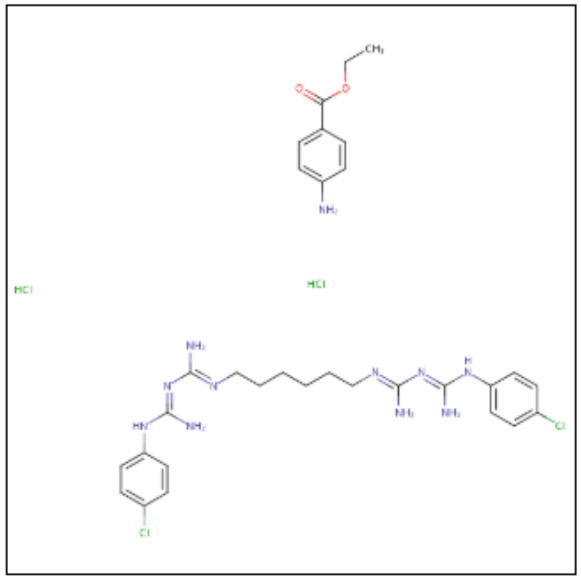
MS-Ready Mappings

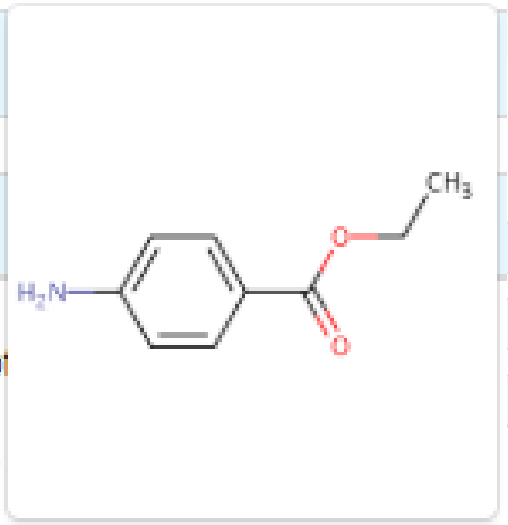
EPA United States Environmental Protection Agency

Progaron

108532-15-6 | DTXSID20148579

Searched by DSSTox Substance Id.

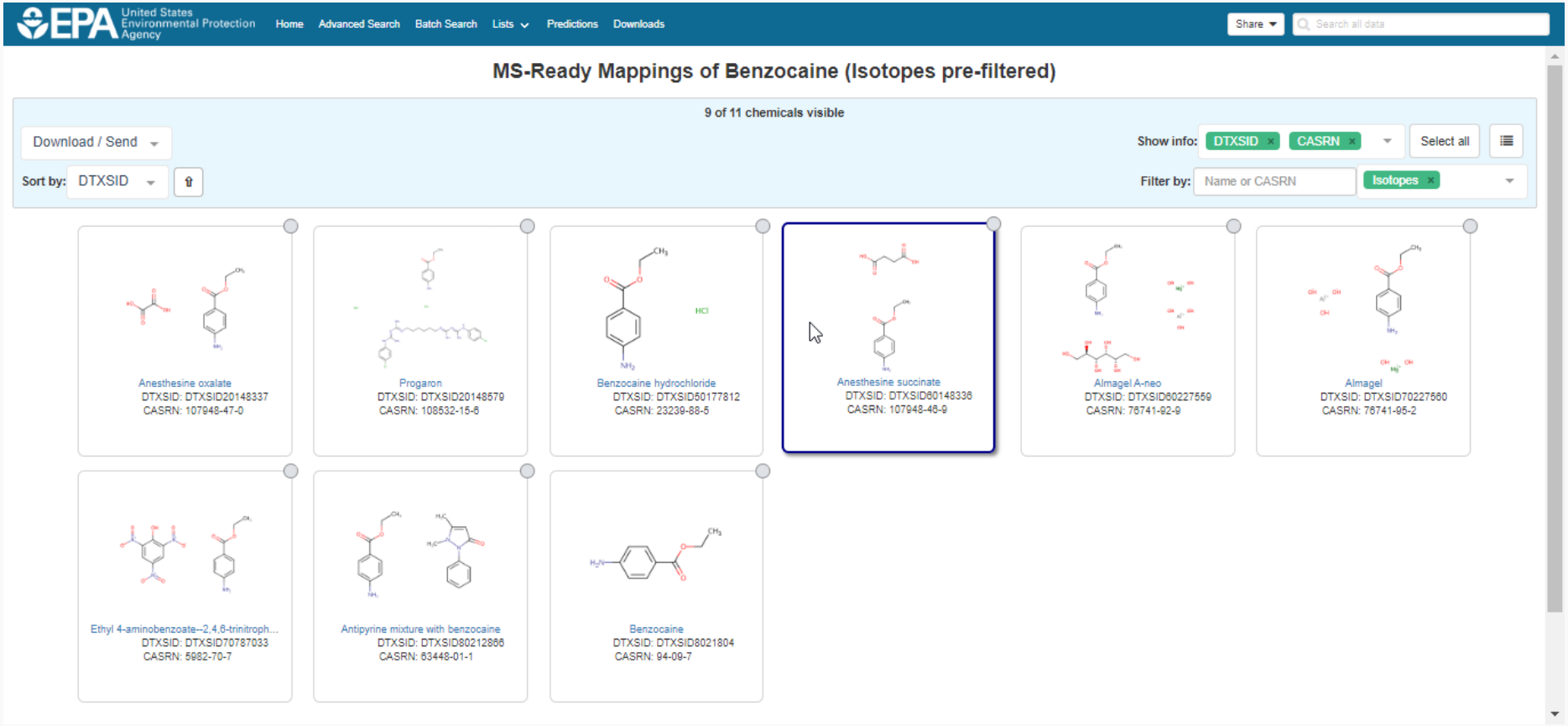




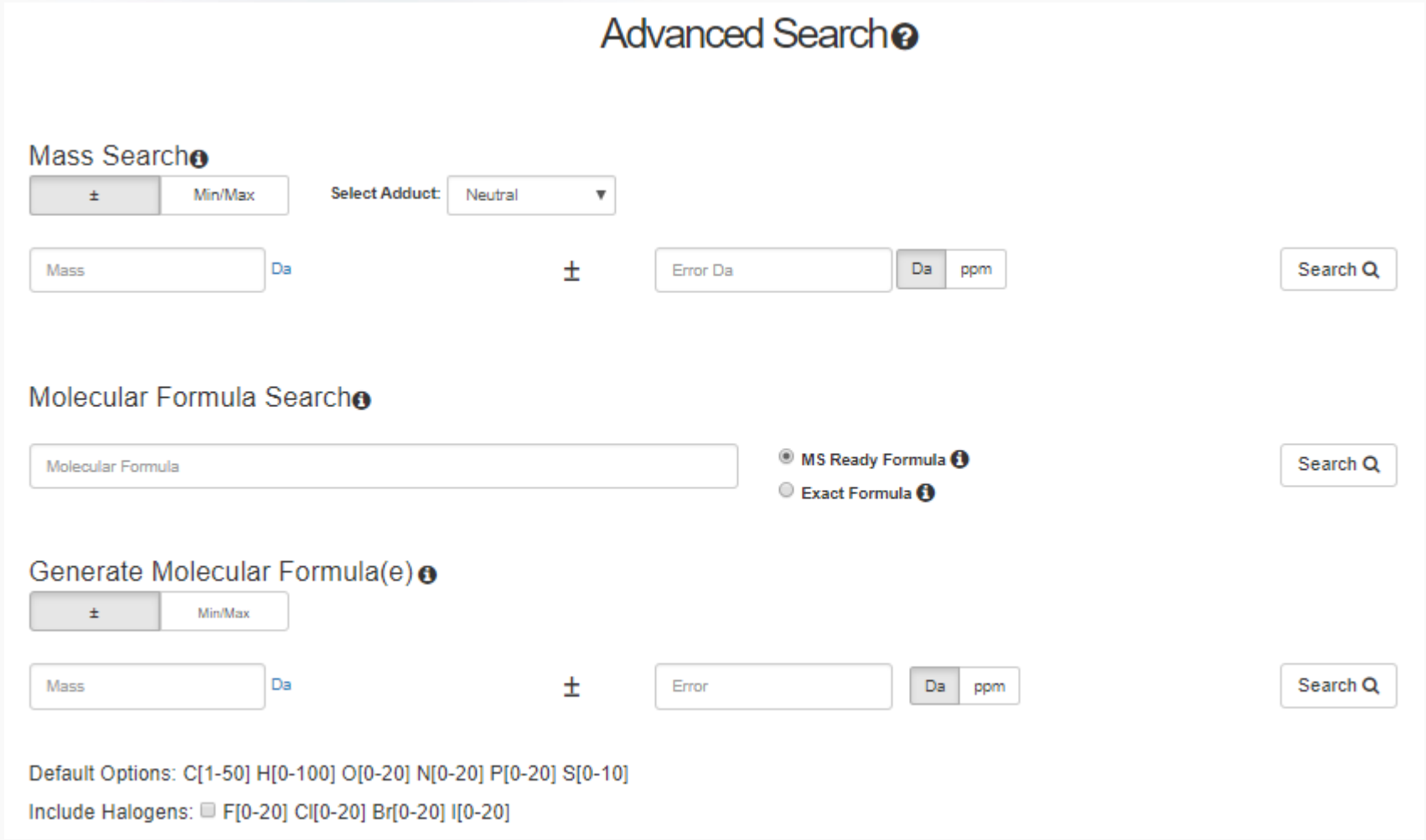
- Intrinsic Properties
- Structural Identifiers
- Linked Substances
  - Same Connectivity: 1 record (based on
  - Mixtures, Components and Neutralize
  - MS-Ready Mappings: DTXCID0013314
  - Similar Compounds: 0 records
- Presence in Lists
- Record Information
- Quality Control Notes

DTXCID301804:11 records; D

MS-Ready Mappings Set



Mass and Formula Searches Supporting Mass Spectrometry



Advanced Searches
Mass Based Search

EPA United States Environmental Protection Agency

Mass Search

± | Min/Max

M

191.131 Da ± 5 Da | ppm

Search

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

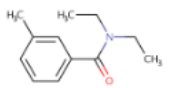
Search Results

Searched by Mass: 191.131 +/- 5.0 ppm.

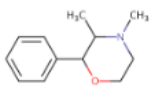
298 of 298 chemicals visible

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

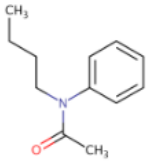
Sort by: Mass Difference | Filter by: Name or CASRN | Multicomponent Chemicals ×



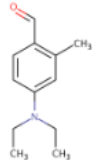
DEET
DTXSID: DTXSID2021995
CASRN: 134-62-3
TOXCAST: 14/663
Mass Diff: 0.000014



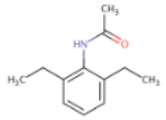
Phendimetrazine
DTXSID: DTXSID1023447
CASRN: 634-03-7
TOXCAST: 0
Mass Diff: 0.000014



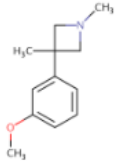
N-Butylacetanilide
DTXSID: DTXSID2042197
CASRN: 91-49-6
TOXCAST: 0
Mass Diff: 0.000014



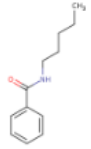
Benzaldehyde, 4-(diethylamino)-2-methyl-
DTXSID: DTXSID4059041
CASRN: 92-14-8
TOXCAST: 0
Mass Diff: 0.000014



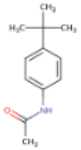
Acetanilide, 2',6'-diethyl-
DTXSID: DTXSID90168148
CASRN: 16665-89-7
TOXCAST: 0
Mass Diff: 0.000014



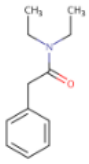
Azetidine, 1,3-dimethyl-3-(m-methoxyp...
DTXSID: DTXSID40173560
CASRN: 19832-26-9
TOXCAST: 0
Mass Diff: 0.000014



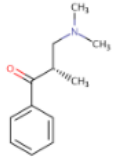
Benzamide, N-pentyl-
DTXSID: DTXSID20174196
CASRN: 20308-43-4



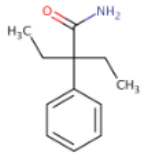
p-t-Butylacetanilide
DTXSID: DTXSID80174238
CASRN: 20330-45-4



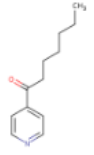
N,N-Diethylphenylacetamide
DTXSID: DTXSID00179048
CASRN: 2431-96-1



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

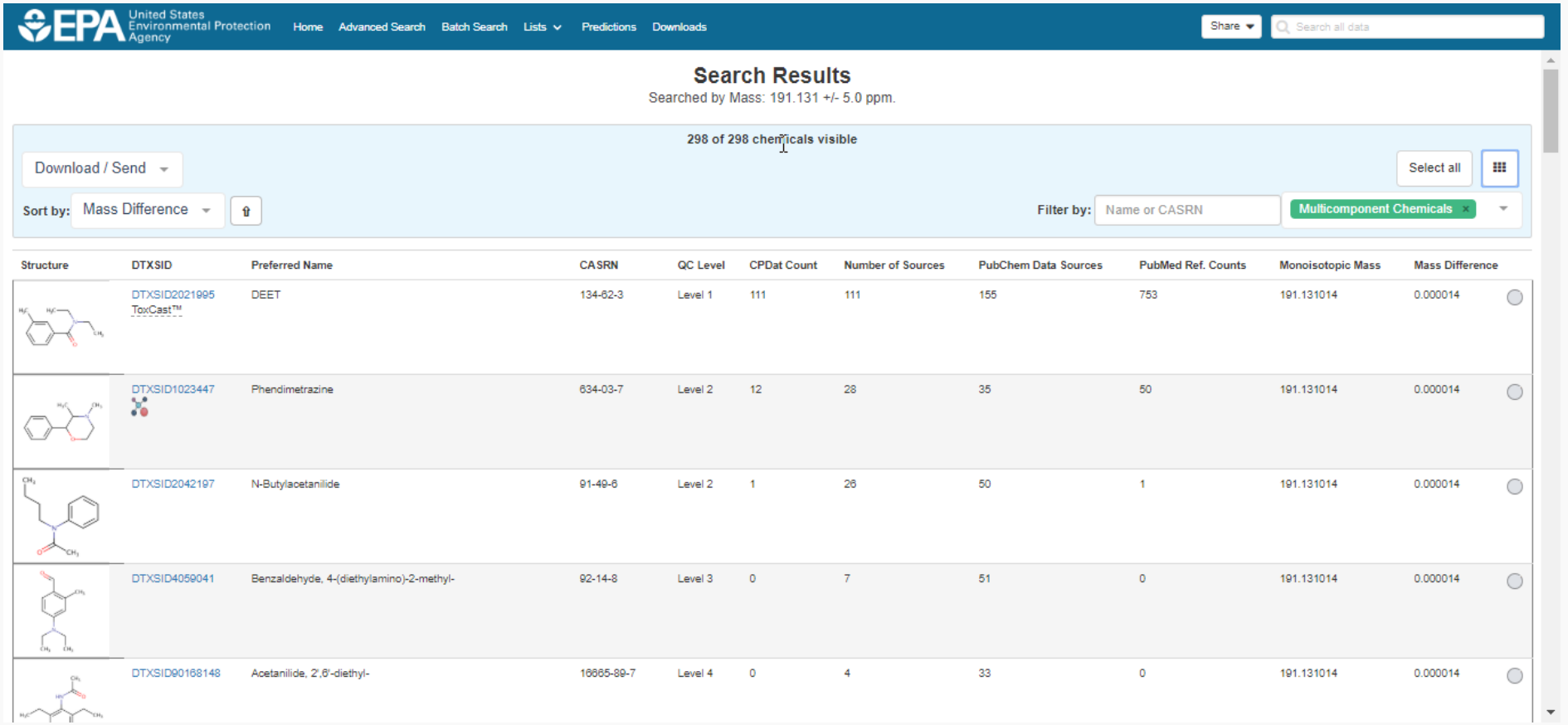


Butyramide, 2-ethyl-2-phenyl-
DTXSID: DTXSID60184653
CASRN: 30568-39-9



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

Advanced Searches
Mass Based Search



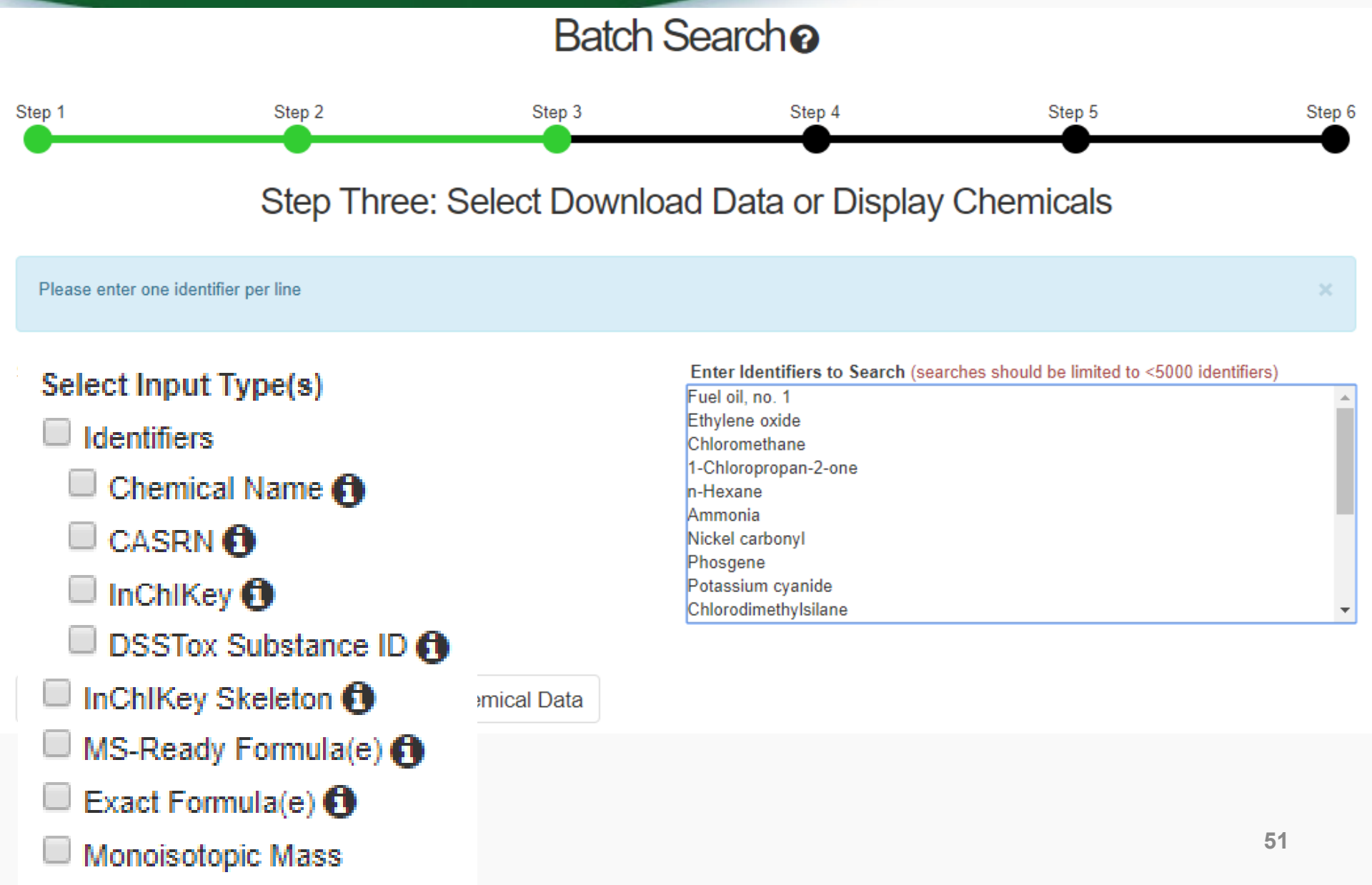
Batch Searching

- Singleton searches are useful but we work with thousands of chemicals!

- Typical questions
  - What is the list of chemicals for the formula $C_xH_yO_z$
  - What is the list of chemicals for a mass +/- error

  - Can I get chemical lists in Excel files? In SDF files?

Batch Searching



Batch Searching

EPA United States Environmental Protection Agency

Select Output Format:

Excel

Download

Customize Results

- Select All
- Select All in Lists

Chemical Identifiers

- DTXSID
- Chemical Name
- CAS-RN
- InChIKey
- IUPAC Name

Structures

- Mol File
- SMILES
- InChI String
- MS-Ready SMILES
- QSAR-Ready SMILES

Intrinsic And Predicted Properties

- Molecular Formula
- Average Mass
- Monoisotopic Mass
- TEST Model Predictions
- OPERA Model Predictions

Presence in Lists:

- ICCVAM test method evaluation report: in vitro ocular toxicity test methods
- 40CFR355
- A list of all PBDEs (Polybrominated diphenyl ethers)
- A list of all PCBs (Polychlorinated biphenyls)
- A list of polycyclic aromatic hydrocarbons
- Acute exposure guideline levels
- Algal Toxins
- Androgen Receptor Chemicals
- APCRA Chemicals for Prospective Analysis
- APCRA Chemicals for Retrospective Analysis
- APCRA Chemicals for Retrospective Analysis_App_List_448_Chemicals
- ATSDR Minimal Risk Levels (MRLs) for Hazardous Substances
- ATSDR Toxic Substances Portal Chemical List
- Bisphenol Compounds
- California Office of Environmental Health Hazard Assessment
- Chemicals with interesting names
- CMAP
- DNT Screening Library
- Drinking Water Suspects, KWR Water, Netherlands
- EDSP Universe
- EPA Chemicals associated with hydraulic fracturing

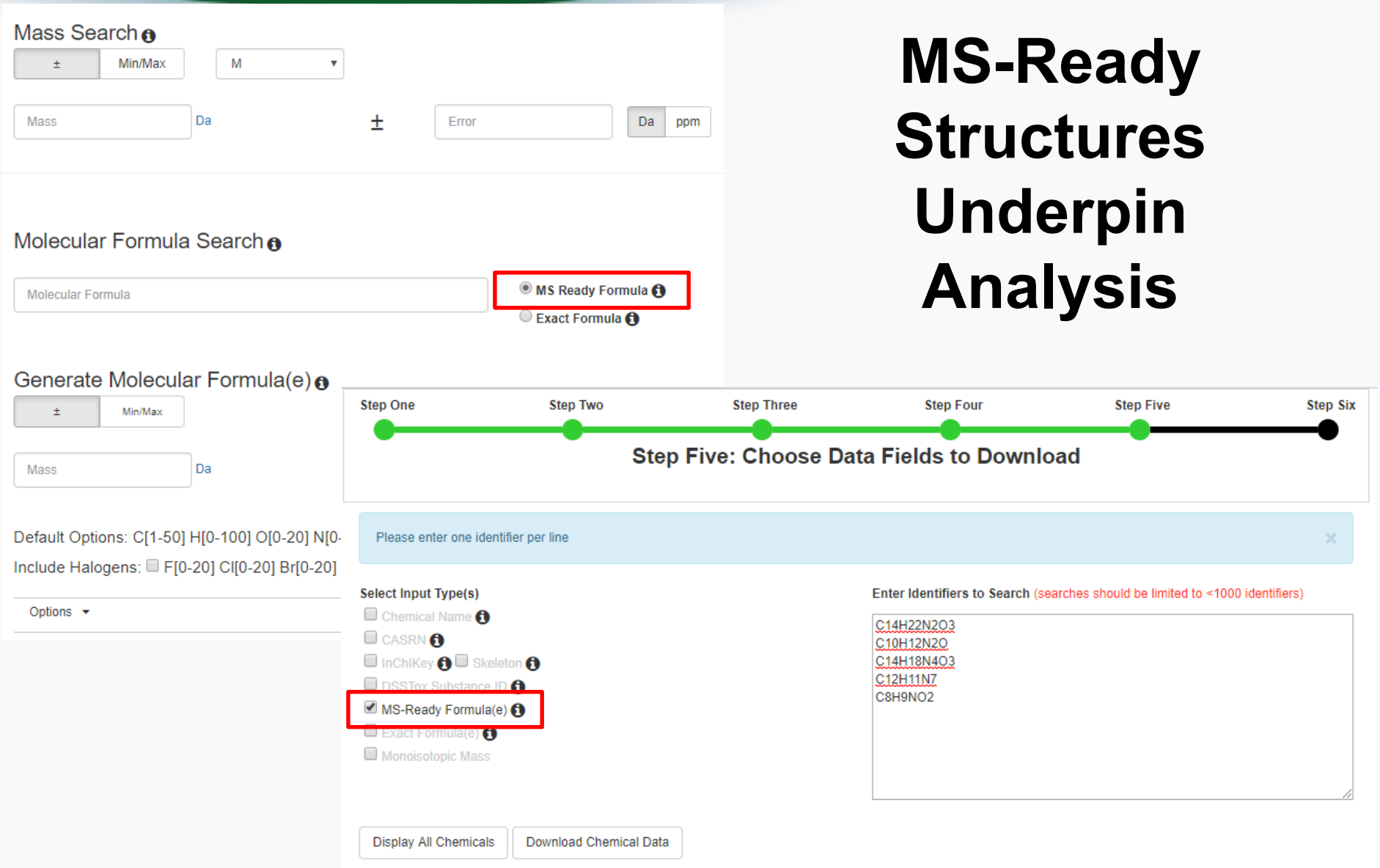
Excel Output

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

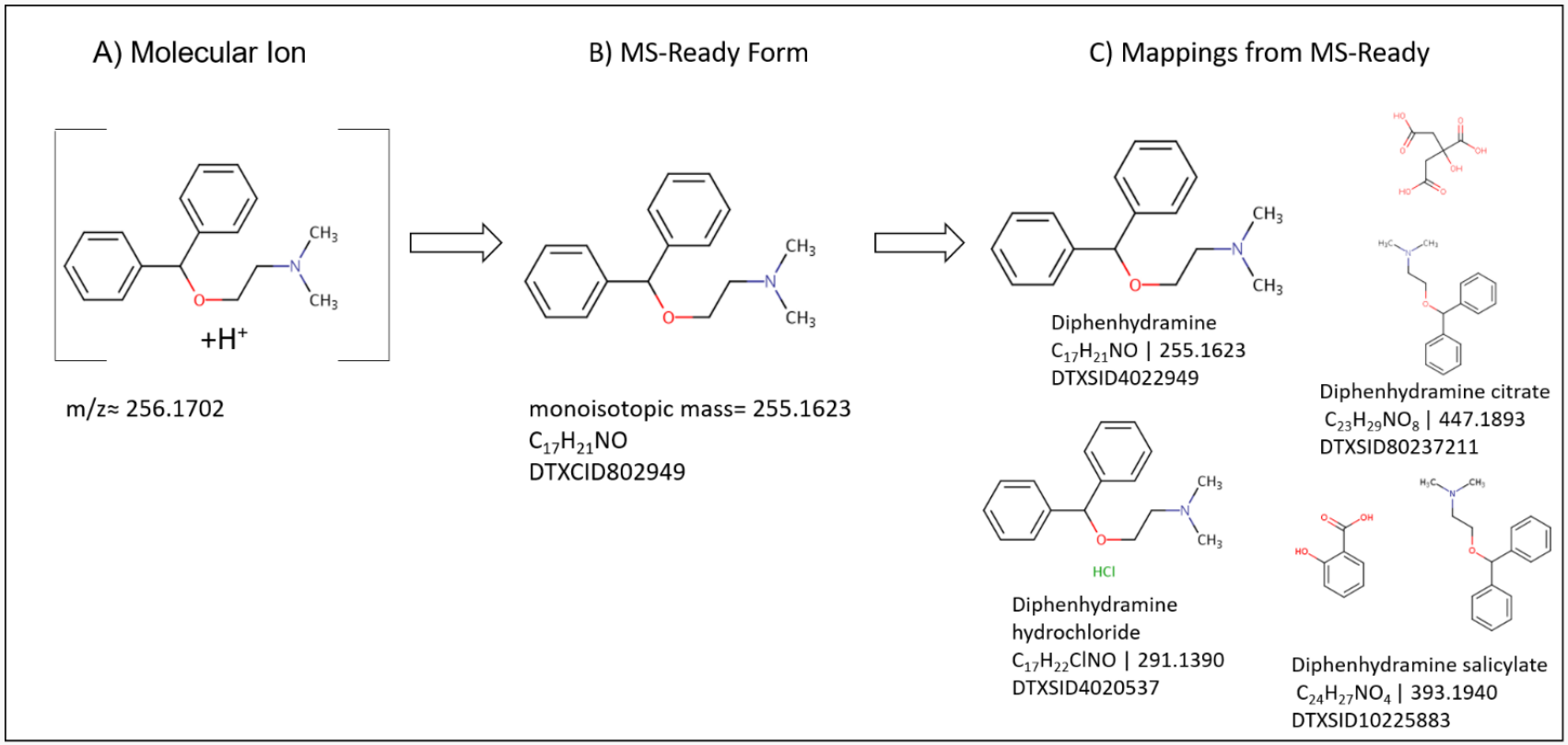
The Dashboard to Support MS-Analysis

EPA United States Environmental Protection Agency

MS-Ready Structures Underpin Analysis

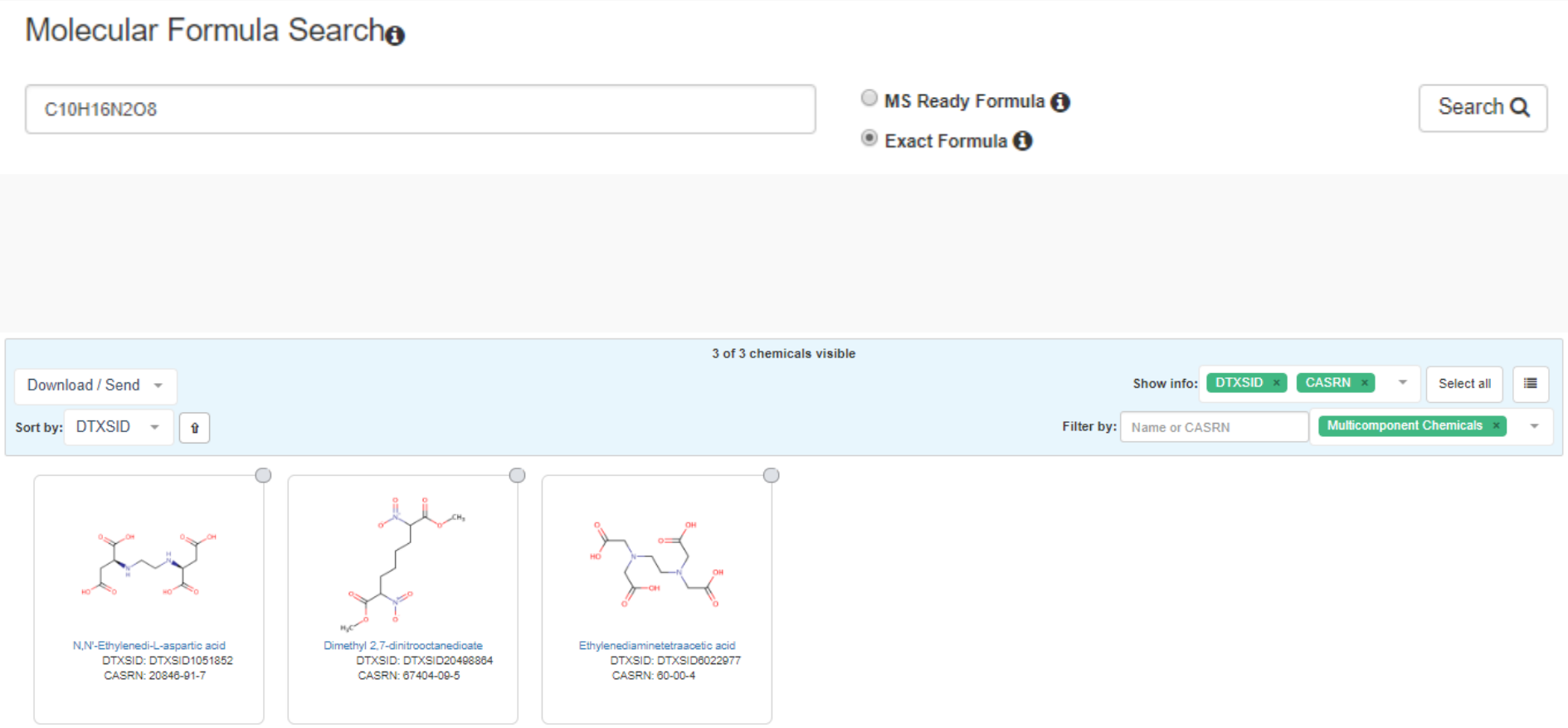


Specific Data-Mappings "MS-Ready Structures"



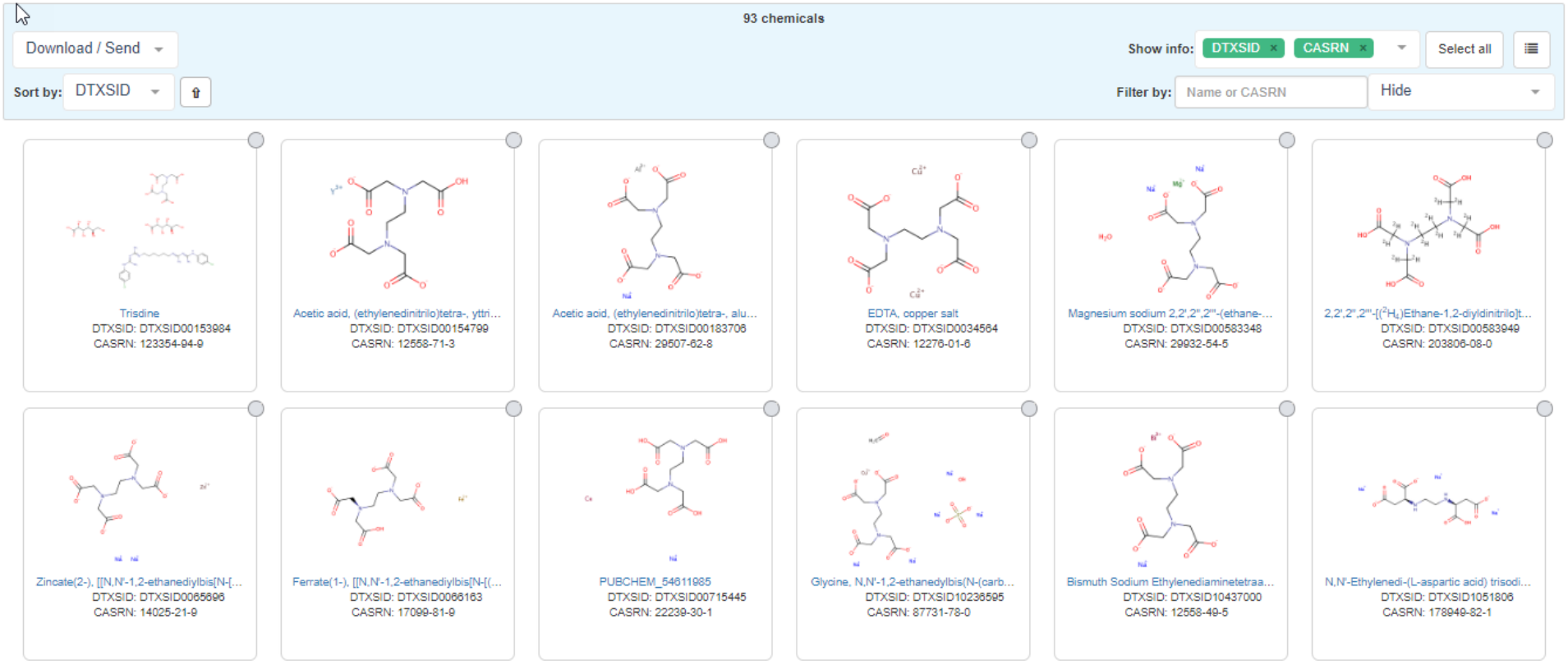
MS-Ready Mappings

- Input Formula: C10H16N2O8



MS-Ready Mappings

- **Same** Input Formula: C10H16N2O8
- MS Ready Formula Search: 93 Chemicals



MS-Ready Mappings

- 93 chemicals returned in total
  - Only 7 of the 93 are single component chemicals
  - Only 4 of the 93 are non-isotope-labeled
  - 3 are neutral compounds and 1 is charged

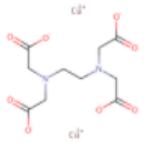
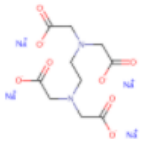
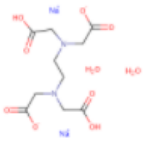
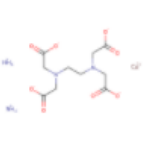
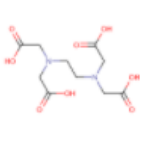
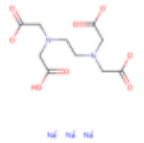
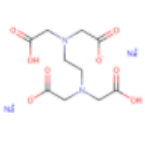
Complexity to Simplicity
93 Chemicals – 7 in EPAHFR

EPA United States Environmental Protection Agency

	INPUT	DTXCID_INDIVIDU	FORMULA	SMILES_I	DTXSID	CASRN	EXPOCAST_M	EXPOCAST	DATA_SOURCI	TOXVAL_D	TOXCAST_	TOXCAST	# OF_PUBMED	PUBCHEM	EPAHFR
2	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID6022977	60-00-4	7.96e-05	Y	71	Y	2.65	3/113	25251	158	Y
3	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID9027073	139-33-3	-	-	41	Y	-	-	25251	56	Y
4	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID3026350	64-02-8	-	-	37	Y	-	-	-	57	Y
5	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID7020556	150-38-9	-	-	30	Y	-	-	-	33	Y
6	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID5049609	67989-88-2	-	-	20	Y	-	-	-	8	Y
7	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID5049576	6381-92-6	-	-	19	Y	-	-	25251	31	Y
8	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID0034564	12276-01-6	-	-	11	-	-	-	-	8	Y
9	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID5027774	15708-41-5	-	-	48	Y	1.98	6/303	241	53	-
10	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID2036409	62-33-9	4.64e-06	Y	37	Y	0.0	0/64	25251	42	-
11	C10H16N2O8	DTXCID00197424	C10H16N2(	OC(=O)C(	DTXSID1051852	20846-91-7	-	-	36	Y	-	-	89	25	-
12	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID6042107	15375-84-5	-	-	25	Y	-	-	97	25	-
13	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID3036442	5964-35-2	-	-	23	Y	-	-	25251	25	-
14	C10H16N2O8	DTXCID00197424	C10H16N2(	OC(=O)C(	DTXSID1051806	178949-82-1	-	-	22	Y	-	-	-	5	-
15	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID0065696	14025-21-9	-	-	22	Y	-	-	-	43	-
16	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID9027813	20824-56-0	-	-	21	Y	-	-	-	12	-
17	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID9027815	21265-50-9	-	-	20	Y	-	-	241	24	-
18	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID5058272	17421-79-3	-	-	19	Y	-	-	25251	25	-
19	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID3058612	2001-94-7	-	-	18	Y	-	-	25251	19	-
20	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID8027820	22473-78-5	-	-	16	Y	-	-	-	11	-
21	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID8058324	17572-97-3	-	-	15	-	-	-	-	36	-
22	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID8028343	67859-51-2	-	-	14	Y	-	-	-	5	-
23	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID4051328	13235-36-4	-	-	14	-	-	-	-	18	-
24	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID6070980	68015-77-0	-	-	14	Y	-	-	-	13	-
25	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID9058317	15934-01-7	-	-	11	-	-	-	-	5	-
26	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID0066163	17099-81-9	-	-	11	-	-	-	241	14	-
27	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID1068988	54959-35-2	-	-	11	-	-	-	241	14	-
28	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID5074266	60816-63-9	-	-	11	-	-	-	1	10	-
29	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID4048197	39208-15-6	-	-	10	-	-	-	-	28	-
30	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID2065830	14931-83-0	-	-	10	-	-	-	47	9	-
31	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID70189997	36499-65-7	-	-	10	-	-	-	25298	26	-
32	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID7051420	61916-40-3	-	-	9	-	-	-	-	4	-
33	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID2051425	73513-47-0	-	-	8	Y	-	-	-	3	-
34	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID7051426	73637-19-1	-	-	8	Y	-	-	-	5	-
35	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID2051427	73637-20-4	-	-	8	Y	-	-	-	-	-
36	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID3058741	10378-23-1	-	-	8	Y	-	-	-	31	-
37	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID6065925	15708-48-2	-	-	8	-	-	-	-	19	-
38	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID20217976	6766-87-6	-	-	8	-	-	-	-	13	-
39	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID5065807	14689-29-3	-	-	7	-	-	-	-	12	-
40	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID6069408	60544-70-9	-	-	7	-	-	-	-	12	-
41	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID00153984	123354-94-9	-	-	7	-	-	-	2	6	-
42	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID70190705	37209-61-3	-	-	7	-	-	-	6	9	-
43	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID7051424	67401-50-7	-	-	6	-	-	-	-	4	-

Complexity to Simplicity
93 Chemicals – 7 in the list

EPA United States Environmental Protection Agency

Structure	DTXSID	Preferred Name	CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Source	PubMed Data Source	Monoisotopic Mass
	DTXSID0034564	EDTA, copper salt	12276-01-6	Level 1	10	11	8	0	413.918561
	DTXSID3026350	Ethylenediaminetetraacetic acid tetrasodium salt	64-02-8	Level 1	1227	37	57	0	380.018442
	DTXSID5049576	Disodium ethylenediaminetetraacetate dihydrate	6381-92-6	Level 1	93	19	31	25251	372.075683
	DTXSID5049609	Ethylenediaminetetraacetic acid, diammonium copper salt	67989-88-2	Level 2	9	20	8	0	387.057712
	DTXSID6022977	Ethylenediaminetetraacetic acid	60-00-4	Level 1	346	71	158	25251	292.090665
	DTXSID7020556	Trisodium ethylenediaminetetraacetate	150-38-9	Level 1	85	30	33	0	358.036498
	DTXSID9027073	Ethylenediaminetetraacetic acid, disodium salt	139-33-3	Level 1	1358	41	56	25251	336.054554

Searching batches
Formula (or mass) searching

EPA United States Environmental Protection Agency

	A	B	C	D	E	F	G
1	INPUT	DTXSID	CASRN	PREFERRED NAME	MOL FORMULA	MONOISOTOPIC MASS	DATA SOURCES
2	C14H22N2O3	DTXSID2022628	29122-68-7	Atenolol	C14H22N2O3	266.163042576	46
3	C14H22N2O3	DTXSID0021179	6673-35-4	Practolol	C14H22N2O3	266.163042576	32
4	C14H22N2O3	DTXSID4048854	841-73-6	Bucolome	C14H22N2O3	266.163042576	20
5	C14H22N2O3	DTXSID1045407	13171-25-0	Trimetazidine dihydrochloride	C14H24Cl2N2O3	338.116398	19
6	C14H22N2O3	DTXSID0045753	56715-13-0	R-(+)-Atenolol	C14H22N2O3	266.163042576	19
7	C14H22N2O3	DTXSID2048531	5011-34-7	Trimetazidine	C14H22N2O3	266.163042576	14
8	C14H22N2O3	DTXSID10239405	93379-54-5	Esatenolol	C14H22N2O3	266.163042576	12
9	C14H22N2O3	DTXSID50200634	52662-27-8	N-(2-Diethylaminoethyl)-2-(4-hydroxyphenoxy)acetamide	C14H22N2O3	266.163042576	7
10	C14H22N2O3	DTXSID4020111	51706-40-2	dl-Atenolol hydrochloride	C14H23ClN2O3	302.1397203	6
11	C14H22N2O3	DTXSID1068693	51963-82-7	Benzenamine, 2,5-diethoxy-4-(4-morpholinyl)-	C14H22N2O3	266.163042576	5
12	C18H34N2O6S	DTXSID3023215	154-21-2	Lincomycin	C18H34N2O6S	406.213757997	35
13	C18H34N2O6S	DTXSID7047803	859-18-7	Lincomycin hydrochloride	C18H35ClN2O6S	442.1904357	22
14	C18H34N2O6S	DTXSID20849438	1398534-62-7	PUBCHEM_71432748	C18H35ClN2O6S	442.1904357	1
15	C10H12N2O	DTXSID1047576	486-56-6	Cotinine	C10H12N2O	176.094963014	40
16	C10H12N2O	DTXSID8075330	50-67-9	Serotonin	C10H12N2O	176.094963014	22
17	C10H12N2O	DTXSID8044412	2654-57-1	4-Methyl-1-phenylpyrazolidin-3-one	C10H12N2O	176.094963014	18
18	C10H12N2O	DTXSID80165186	153-98-0	Serotonin hydrochloride	C10H13ClN2O	212.0716407	11
19	C10H12N2O	DTXSID2048870	29493-77-4	(4R,5S)-4-methyl-5-phenyl-4,5-dihydro-1,3-oxazol-2-amine	C10H12N2O	176.094963014	10
20	C10H12N2O	DTXSID10196105	443-31-2	6-Hydroxytryptamine	C10H12N2O	176.094963014	9
21	C10H12N2O	DTXSID90185693	31822-84-1	1,4,5,6-Tetrahydro-5-phenoxypyrimidine	C10H12N2O	176.094963014	7
22	C10H12N2O	DTXSID40178777	2403-66-9	2-Benzimidazolepropanol	C10H12N2O	176.094963014	7
23	C10H12N2O	DTXSID80157026	13140-86-8	N-Cyclopropyl-N'-phenylurea	C10H12N2O	176.094963014	6
24	C10H12N2O	DTXSID30205607	570-14-9	4-Hydroxytryptamine	C10H12N2O	176.094963014	6
25	C14H18N4O3	DTXSID5023900	17804-35-2	Benomyl	C14H18N4O3	290.137890456	68
26	C14H18N4O3	DTXSID3023712	738-70-5	Trimethoprim	C14H18N4O3	290.137890456	51
27	C14H18N4O3	DTXSID40209671	60834-30-2	Trimethoprim hydrochloride	C14H19ClN4O3	326.1145682	8
28	C14H18N4O3	DTXSID70204210	55687-49-5	Benzenemethanol, 4-((2,4-diamino-5-pyrimidinyl)methyl)-2,	C14H18N4O3	290.137890456	5
29	C14H18N4O3	DTXSID20152671	120075-57-2	6-Methoxy-4-(3-(N,N-dimethylamino)propylamino)-5,8-quin	C14H18N4O3	290.137890456	4
30	C14H18N4O3	DTXSID30213742	63931-79-3	1H-1,2,4-Benzotriazepine-3-carboxylic acid, 4,5-dihydro-4-	C14H18N4O3	290.137890456	3
31	C14H18N4O3	DTXSID30219608	69449-07-6	2,4-Pyrimidinediamine, 5-((3,4,5-trimethoxyphenyl)methyl)-	C14H20N4O4	308.14845514	3
32	C14H18N4O3	DTXSID20241155	94232-27-6	L-Aspartic acid, compound with 5-((3,4,5-trimethoxyphenyl	C18H25N5O7	423.175398165	3
33	C14H18N4O3	DTXSID80241156	94232-28-7	L-Glutamic acid, compound with 5-((3,4,5-trimethoxypheny	C19H27N5O7	437.191048229	3
34	C14H18N4O3	DTXSID20143781	101204-93-7	1H-Pyrido(2,3-e)-1,4-diazepine-2,3,5-trione, 4-(2-(diethylam	C14H18N4O3	290.137890456	3
35	C12H11N7	DTXSID6021373	396-01-0	Triamterene	C12H11N7	253.107593382	52
36	C12H11N7	DTXSID00204465	5587-93-9	Ampyrimine	C12H11N7	253.107593382	7
37	C12H11N7	DTXSID5064621	7300-26-7	Benzenamine, 4-azido-N-(4-azidophenyl)-	C12H9N7	251.091943318	4
38	C12H11N7	DTXSID00848025	90293-82-6	Sulfuric acid--6-phenylpteridine-2,4,7-triamine (1/1)	C12H13N7O4S	351.074973101	1
39	C12H11N7	DTXSID50575293	92310-83-3	(1E)-N-Phenyl-1,2-bis(1H-1,2,4-triazol-1-yl)ethan-1-imine	C12H11N7	253.107593382	1
40	C8H9NO2	DTXSID2020006	103-90-2	Acetaminophen	C8H9NO2	151.063328534	75
41	C8H9NO2	DTXSID6026567	134-20-3	Methyl 2-aminobenzoate	C8H9NO2	151.063328534	50

PFAS

A List of Lists of Chemicals

https://comptox.epa.gov/dashboard/chemical_lists

EPA United States Environmental Protection Agency

EPA United States Environmental Protection Agency | Home | Advanced Search | Batch Search | Lists | Predictions | Downloads | Search All Da

Chemistry Dashboard

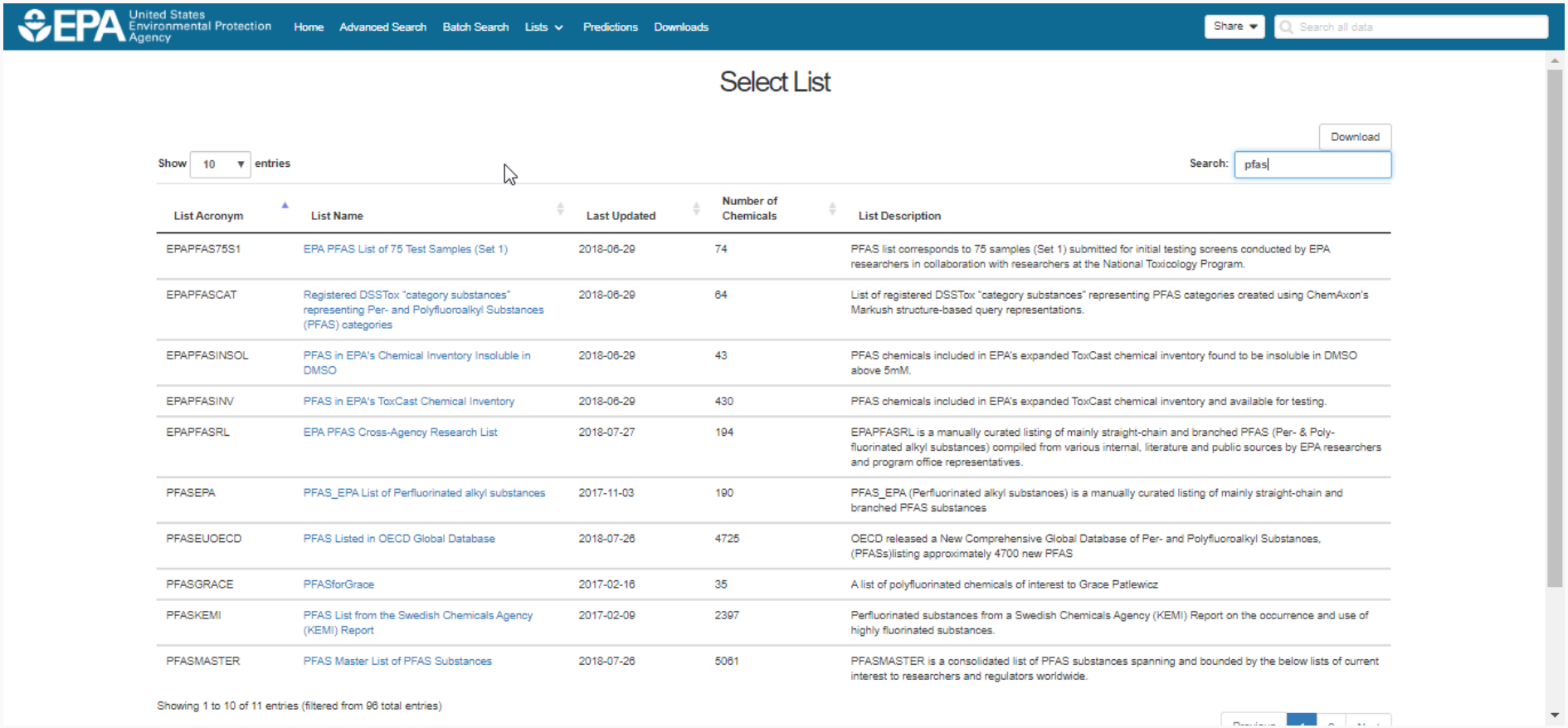
Aa | Aa | Aa

Select List

List Name	Number of Chemicals	List Description
40CFR355	354	Extremely Hazardous Substance List and Threshold Planning Quantities; Emergency Planning and Release Notification Requirements; Final Rule. (52 FR 13378)
Algal Toxins	54	A set of algal toxins of interest
Androgen Receptor Chemicals	110	The list of chemicals used to identify references with in vitro AR binding . From Kleinstrauer et al http://pubs.acs.org/doi/abs/10.1021/acs.chemrestox.6b00347
ATSDR Toxic Substances Portal Chemical List	200	The Agency for Toxic Substances and Disease Registry (ATSDR)is a federal public health agency of the U.S. Department of Health and Human Services.
Bisphenol Compounds	52	This list represents a collection of Bisphenol Compounds
California Office of Environmental Health Hazard Assessment	972	The OEHHA Chemical Database is a compilation of health hazard information including reference exposure levels, California public health goals, child-specific reference doses, Propos. 65 safe harbor numbers, soil-screening levels, and fish advisories
Chemicals with interesting names	17	This is a list of chemicals with interesting and fun names
EPA Integrated Risk Information System (IRIS)	510	EPA's IRIS Program identifies and characterizes the health hazards of chemicals found in the environment. Each IRIS assessment can cover a chemical, a group of related chemicals, or a complex mixture.
EPAHFR - EPA Chemicals associated with hydraulic fracturing	1640	EPAHFR lists chemicals associated with hydraulic fracturing from 2005-20013, as reported in EPA's Hydraulic Fracturing Drinking Water Assessment Final Report (Dec 2016)
EU Cosmetic Ingredients Inventory (Combined 2000/2006)	2878	EUCOSMETICS contains the Combined Inventory of Ingredients Employed in Cosmetic Products (2000, SCCNFP/0389/00 Final) and Revised Inventory (2006, Decision 2006/257/EC), prepared for NORMAN by P. von der Ohe (UBA) and R. Aalizadeh (Uni. Athens).
EU Toxrisk Dataset	230	Compounds of interest to the EU-ToxRisk Case Studies.
French Monitoring List	1171	FRENCHLIST contains substances for prospective monitoring activities in France, developed in cooperation with the NORMAN Network Working Group 1 on Prioritization. Provided by Valeria Dulio, INERIS, France. Further details on the website.

11 PFAS Lists

http://comptox-prod.epa.gov/dashboard/chemical_lists



List Acronym	List Name	Last Updated	Number of Chemicals	List Description
EPAPFAS75S1	EPA PFAS List of 75 Test Samples (Set 1)	2018-06-29	74	PFAS list corresponds to 75 samples (Set 1) submitted for initial testing screens conducted by EPA researchers in collaboration with researchers at the National Toxicology Program.
EPAPFASCAT	Registered DSSTox "category substances" representing Per- and Polyfluoroalkyl Substances (PFAS) categories	2018-06-29	64	List of registered DSSTox "category substances" representing PFAS categories created using ChemAxon's Markush structure-based query representations.
EPAPFASINSOL	PFAS in EPA's Chemical Inventory Insoluble in DMSO	2018-06-29	43	PFAS chemicals included in EPA's expanded ToxCast chemical inventory found to be insoluble in DMSO above 5mM.
EPAPFASINV	PFAS in EPA's ToxCast Chemical Inventory	2018-06-29	430	PFAS chemicals included in EPA's expanded ToxCast chemical inventory and available for testing.
EPAPFASRL	EPA PFAS Cross-Agency Research List	2018-07-27	194	EPAPFASRL is a manually curated listing of mainly straight-chain and branched PFAS (Per- & Poly-fluorinated alkyl substances) compiled from various internal, literature and public sources by EPA researchers and program office representatives.
PFASEPA	PFAS_EPA List of Perfluorinated alkyl substances	2017-11-03	190	PFAS_EPA (Perfluorinated alkyl substances) is a manually curated listing of mainly straight-chain and branched PFAS substances
PFASEUOECD	PFAS Listed in OECD Global Database	2018-07-26	4725	OECD released a New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances, (PFASs)listing approximately 4700 new PFAS
PFASGRACE	PFASforGrace	2017-02-16	35	A list of polyfluorinated chemicals of interest to Grace Patlewicz
PFASKEMI	PFAS List from the Swedish Chemicals Agency (KEMI) Report	2017-02-09	2397	Perfluorinated substances from a Swedish Chemicals Agency (KEMI) Report on the occurrence and use of highly fluorinated substances.
PFASMASTER	PFAS Master List of PFAS Substances	2018-07-26	5061	PFASMASTER is a consolidated list of PFAS substances spanning and bounded by the below lists of current interest to researchers and regulators worldwide.

Showing 1 to 10 of 11 entries (filtered from 96 total entries)

The OECD List of PFAS

http://www.oecd.org/chemicalsafety/portal-perfluorinated-chemicals/

EPA United States Environmental Protection Agency

The OECD releases a new list of PFASs

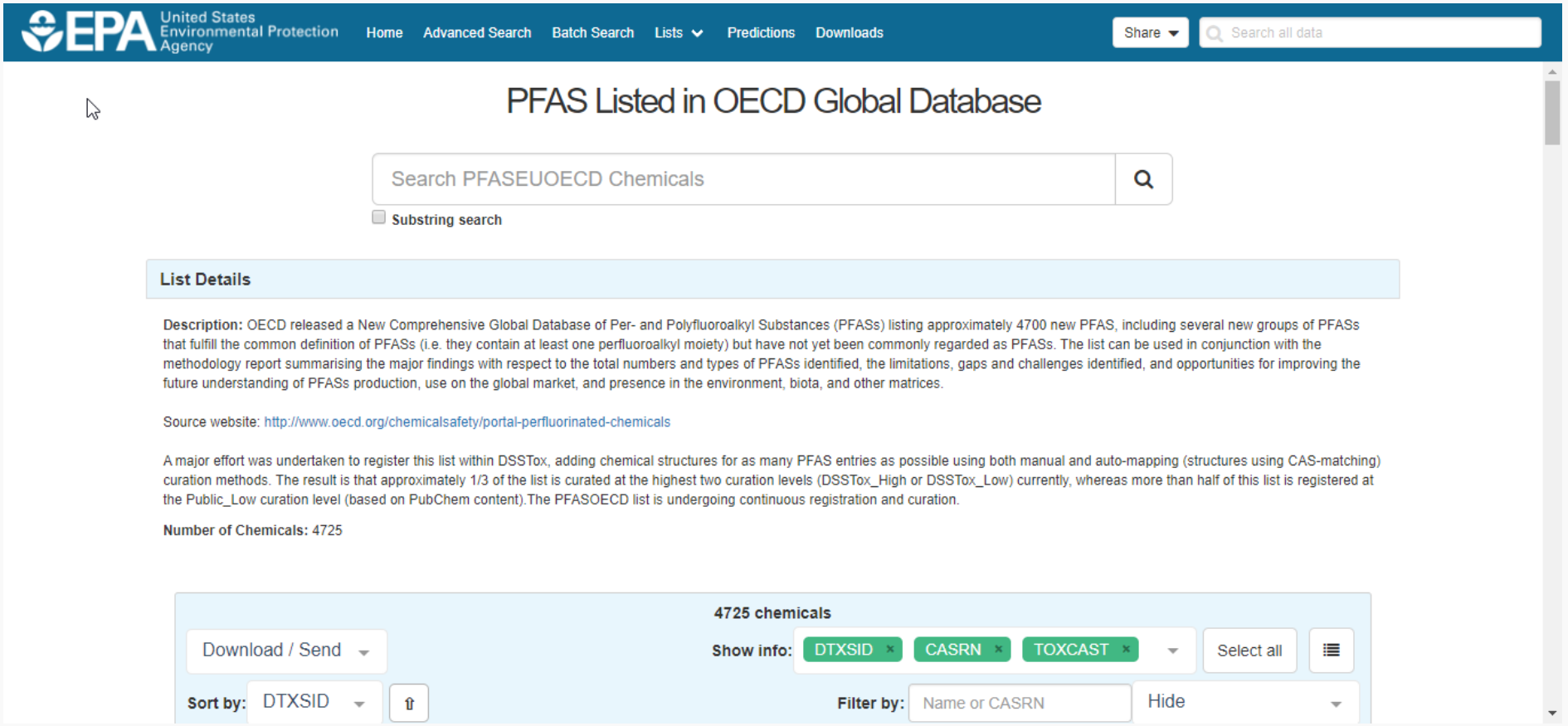
The OECD releases a new list of Per- and Polyfluoroalkyl Substances (PFASs) based on a comprehensive analysis of information available in the public domain. In total, 4730 PFAS-related CAS numbers have been identified and categorised in this study, including several new groups of PFASs that fulfil the common definition of PFASs (i.e. they contain at least one perfluoroalkyl moiety) but have not yet been commonly regarded as PFASs.

This work has been conducted under the OECD/UN Environment Global PFC Group in support of the Strategic Approach to International Chemicals Management (SAICM) and shifting to safer alternatives for PFASs.

The New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances (PFASs) comes with a methodology report also detailing the major findings with respect to the total numbers and types of PFASs identified, the limitations, gaps and challenges identified in the development of the new list, and opportunities for improving the future understanding of PFASs production, use on the global market, and presence in the environment, biota, and other matrices.

The OECD List of PFAS

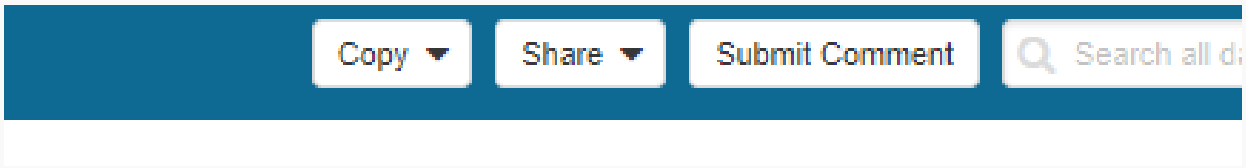
http://www.oecd.org/chemicalsafety/portal-perfluorinated-chemicals/



CROWDSOURCED ANNOTATION

How can be curate our data?

- Crowdsourcing is well proven nowadays
- Comments can be added at a record level



- Submitted comments are reviewed by administrators and responded to

Public Crowdsourced Comments

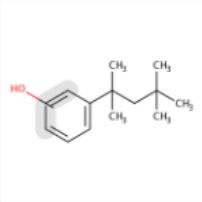
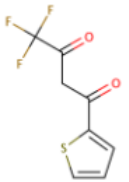
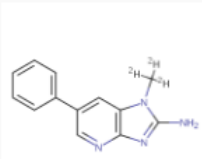
https://comptox.epa.gov/dashboard/comments/public_index

EPA United States Environmental Protection Agency

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

Crowdsourced Comments

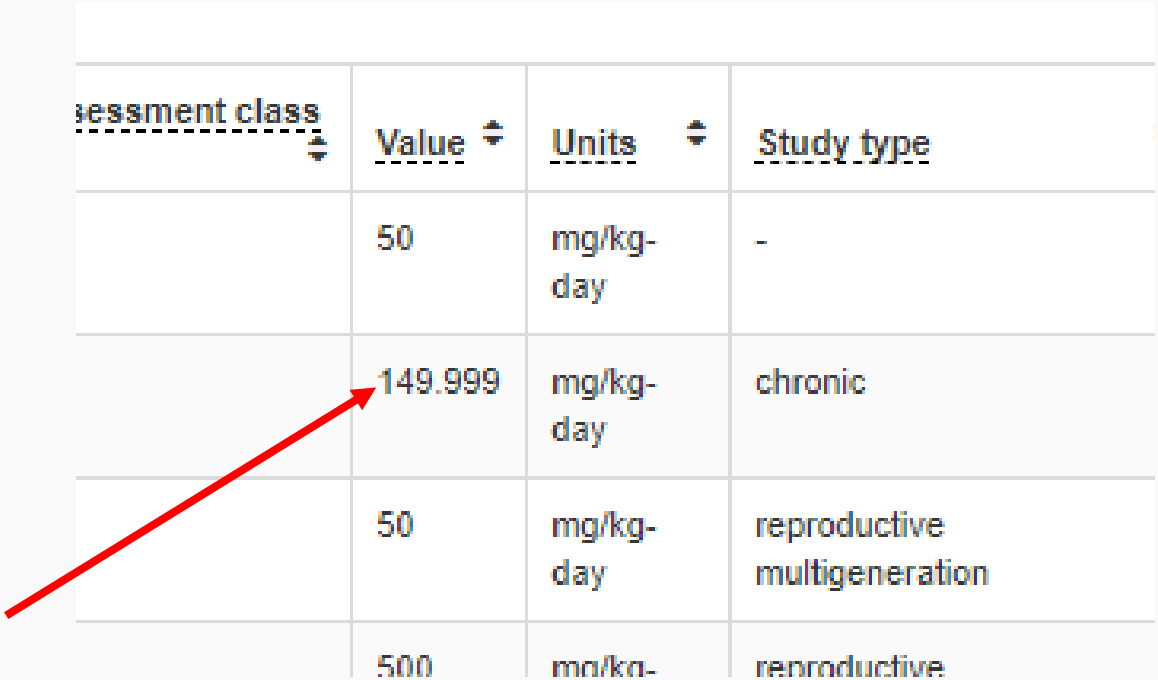
Show 10 entries | Search:

Chemical	Structure	Date	Comment	Status
(1,1,3,3-Tetramethylbutyl)phenol		2017-07-15	Octylphenol redirects here, yet the name and related chemicals are 1,1,3,3-tetramethylbutylphenol - which is only a subset of all octylphenol isomers? Is this CAS only for these alkyl isomers?	★
1,3-Butanedione, 4,4,4-trifluoro-1-(2-thienyl)-		2017-03-30	Synonym: TTFA (Any way to bank these reCAPTCHAs so I don't have to do it everytime?)	★
1-(²H₃)Methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-amine		2017-05-06	1-(2H3)Methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-amine 210049-13-1 | DTXSID70670097 contains an error in the empirical formula due to an error in the deuterium representation and subsequent counting	★

Reviewer comments are public

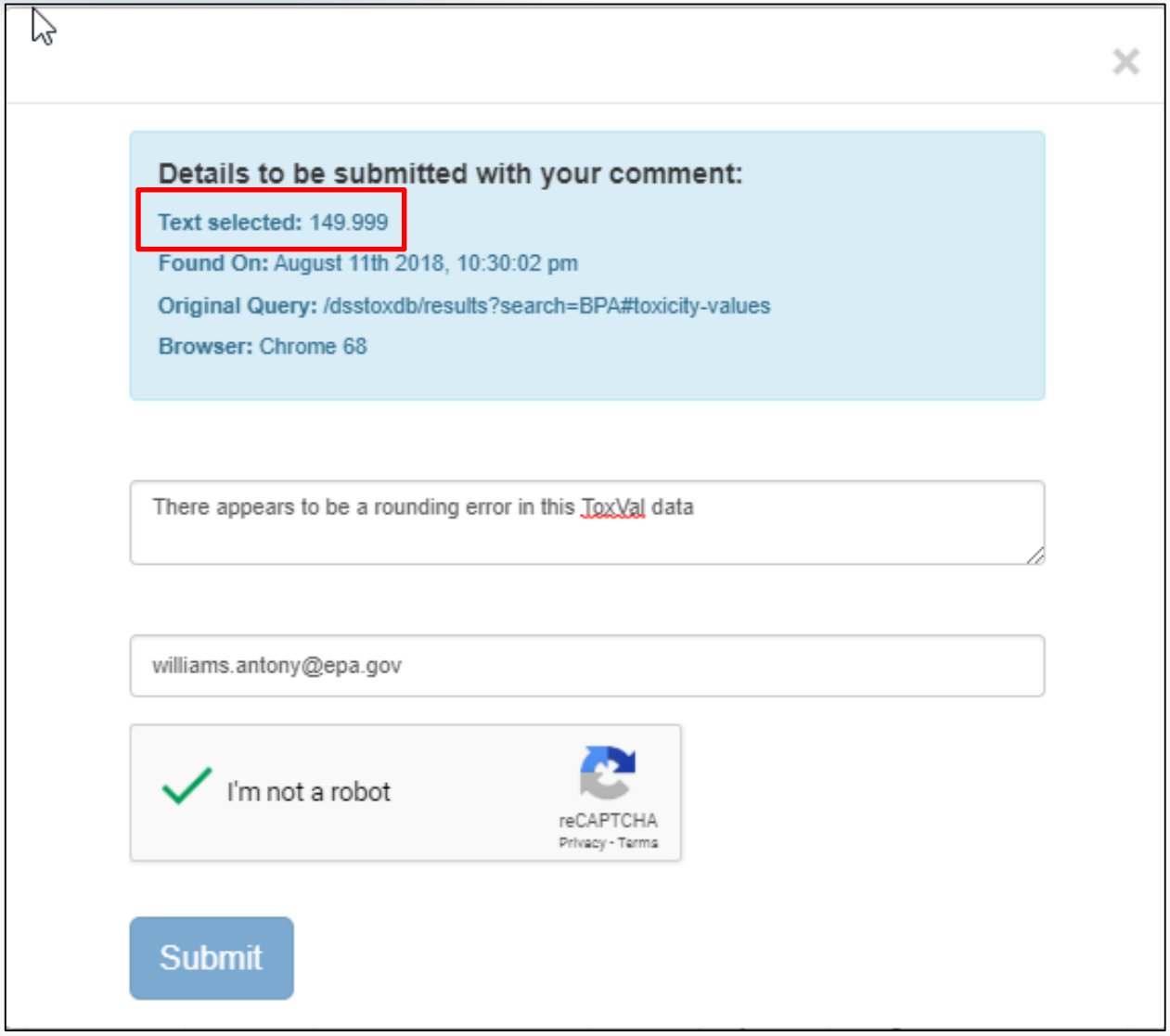
Crowdsourcing Comments
Single Cell Commenting added

- Highlight an alphanumeric text string



ssessment class	Value	Units	Study type
	50	mg/kg-day	-
	149.999	mg/kg-day	chronic
	50	mg/kg-day	reproductive multigeneration
	500	mg/kg-	reproductive

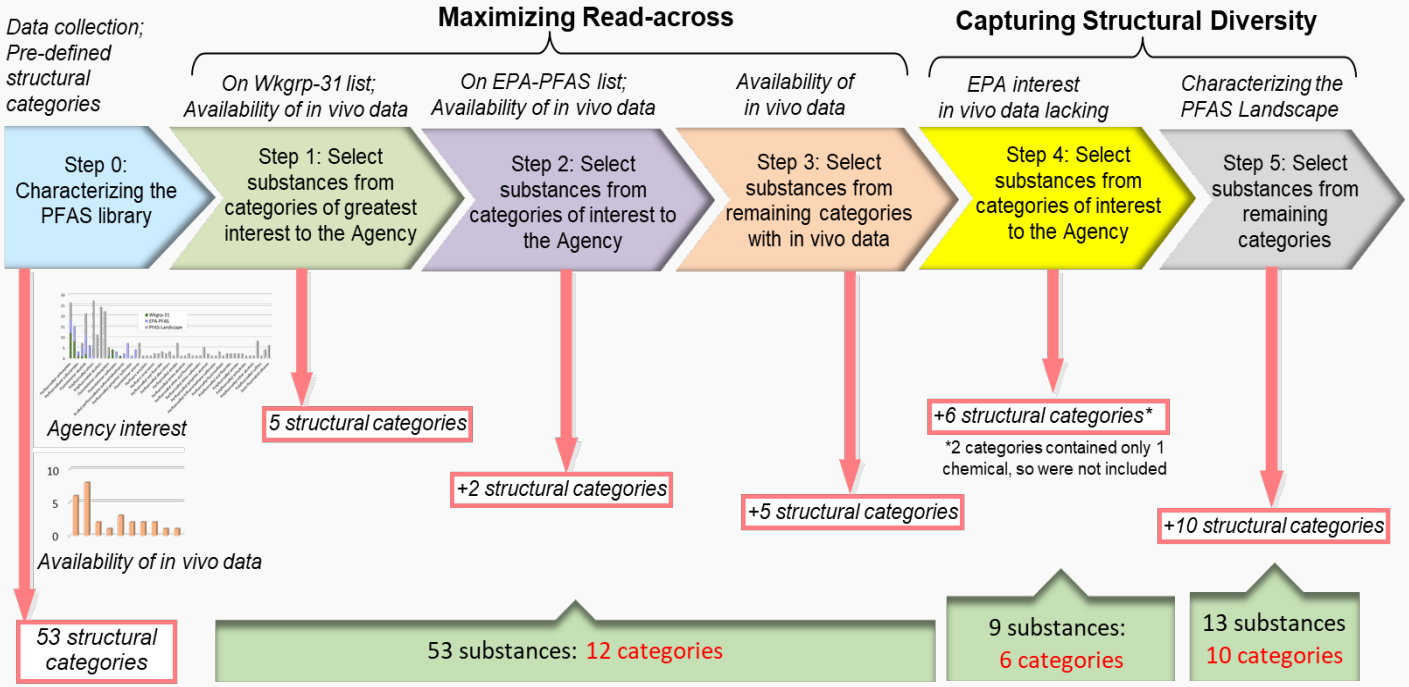
Crowdsourcing Comments



FUTURE WORK

Work in progress

- Development of a high-throughput screening library and collection of physical samples (~400)
- 75 PFAS chemicals for screening based on categories, diversity, exposure considerations, procurability and testability, availability of existing toxicity data



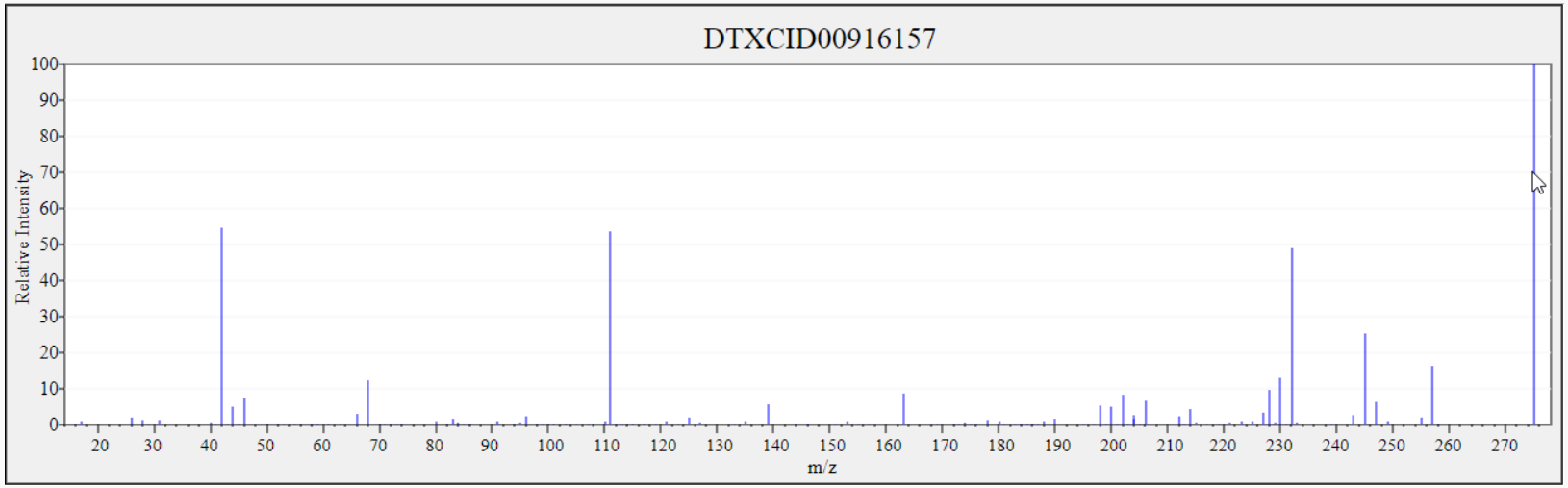
Work in Progress

- CFM-ID
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database
- Retention Time Index Prediction
- Structure/substructure/similarity search
- Generation of MS-ready structures:
  - Upload file, download results
  - Service based generation

Predicted Mass Spectra

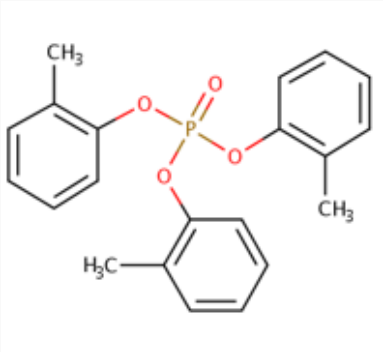
http://cfmid.wishartlab.com/

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

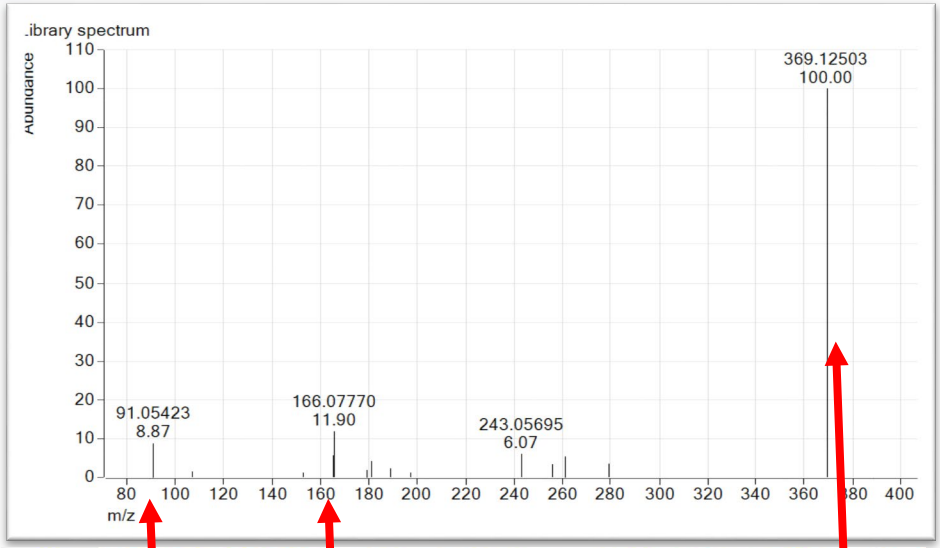


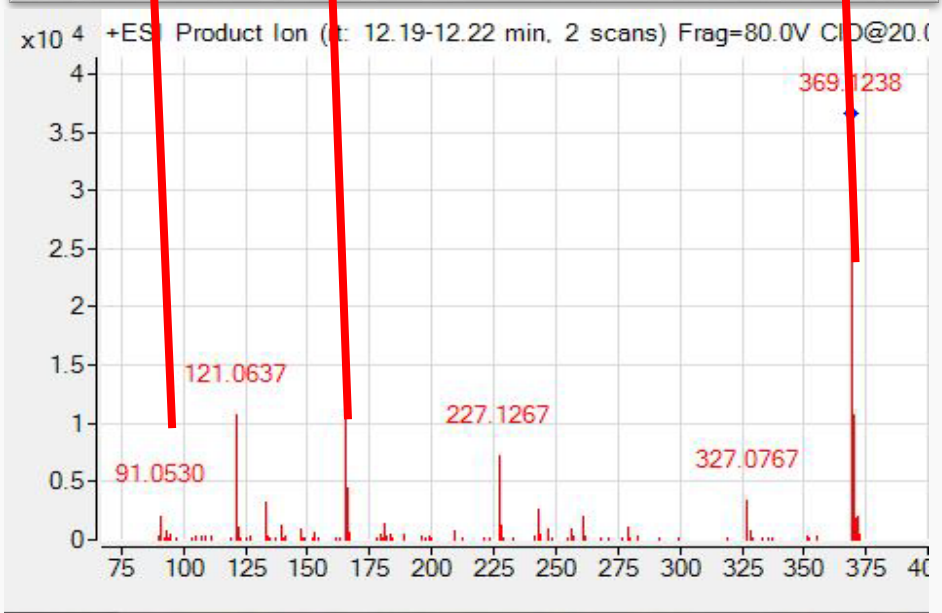
Predicted Mass Spectra

Library Fragmentation Spectra (20eV)



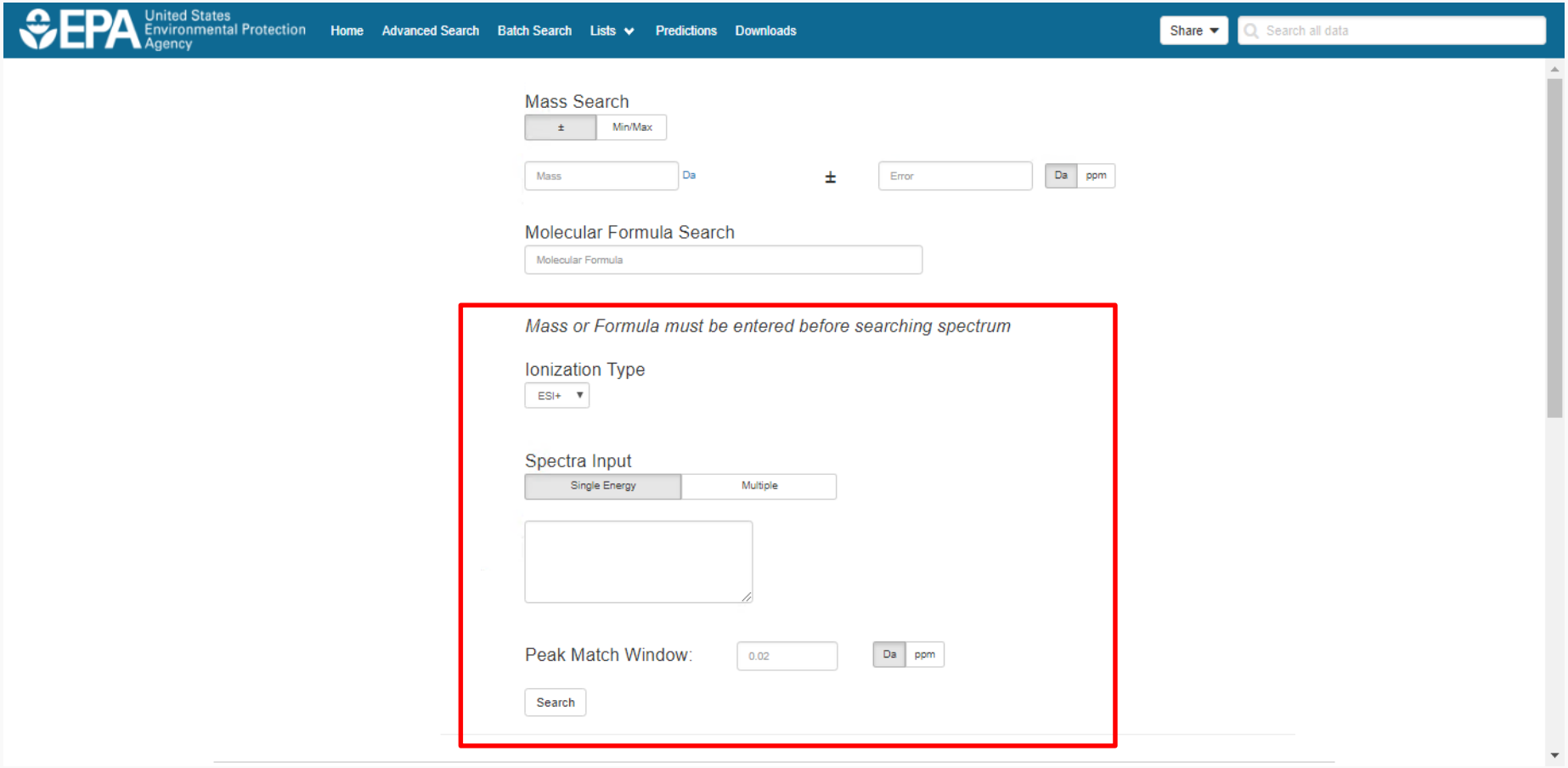
Observed Fragmentation Spectra (20eV)



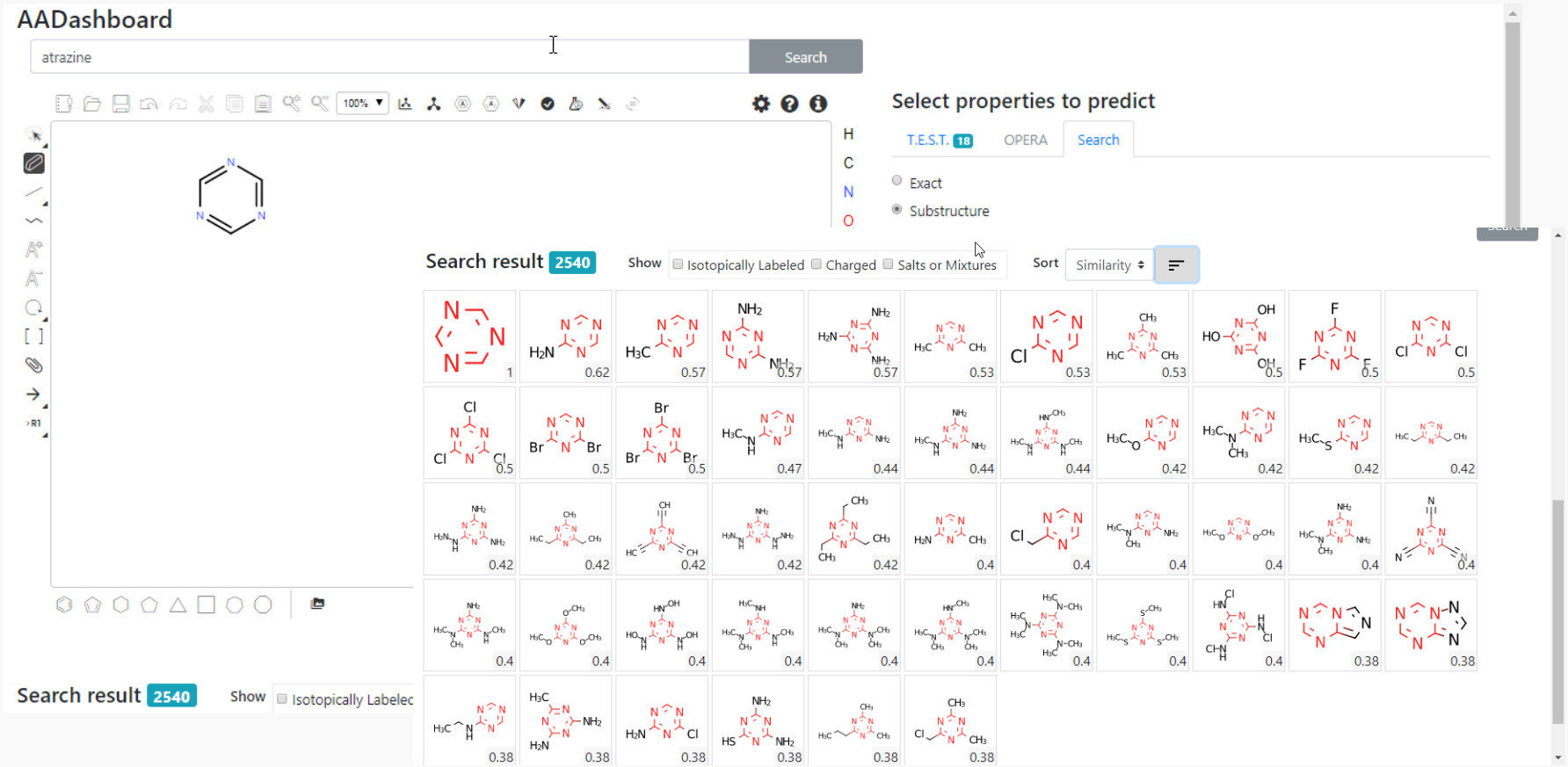


Match Score

Search Expt. vs. Predicted Spectra



Prototype Development



Prototype Development