NORMAN Databases Workshop:
UBA, Berlin, 8-9 June 2017

NORMAN Suspects meet the US EPA CompTox Chemistry Dashboard

Emma Schymanski, Eawag: Swiss Federal Institute of Aquatic Science and Technology
Antony J. Williams & Andrew McEachran, United States Environmental Protection Agency

What is the CompTox Chemistry Dashboard?

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

Aims of US EPA's Computational Toxicity Program:

- Integrate advances in biology, chemistry, and computer science to…
- Prioritize chemicals based on potential human and ecological health risks

The CompTox Chemistry Dashboard:

- Provides access to ~747,000 chemical substances & associated data
- Years of investment/experience in data validation and curation
- Data include: (plus a LOT more …)
  - Experimental and predicted physicochemical properties
  - ToxCast bioassay screening data
  - Product and functional use information and more
- Search capabilities include:
  - Mass or formula-based searching
  - Rank-ordering of results via functional use statistics to prioritize chemicals

The Dashboard in brief

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

Chemistry Dashboard

Search a chemical by systematic name, synonym, CAS number, or InChIKey

Single component search Ignore isotopes

See what people are saying, read the dashboard comments!

Need more? Use advanced search.

747 Thousand Chemicals

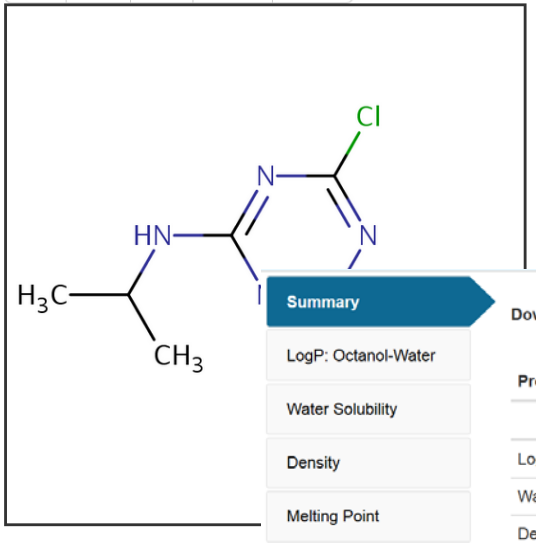
The Dashboard in brief

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

Atrazine

1912-24-9 | DTXSID9020112

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



Wikipedia

Atrazine is an herbicide of the triazine class. Atrazine is used to prevent pre- and postemergence broadleaf weeds in crops such as maize (corn) and sugarcane and on turf, such as golf courses and residential lawns. It is one of the most widely used herbicides in US and Australian agriculture. It was banned in the European Union in October 2003, when the EU found groundwater levels exceeding the limits set by regulators, and Syngenta could neither show that this could be prevented nor that… Read more

Chemical Properties | Literature | Comments

- Summary
- LogP: Octanol-Water
- Water Solubility
- Density
- Melting Point
- Boiling Point
- Surface Tension
- Vapor Pressure
- LogKoa: Octanol-Air
- Henry's Law
- Index of Refraction
- Molar Refractivity

Download as: TSV | Excel | SDF

Property	Average		Median		Range		Unit
	Experimental	Predicted	Experimental	Predicted	Experimental	Predicted	
LogP: Octanol-Water	2.61 (1)	2.76 (4)	2.61	2.76	2.61	2.50 to 3.05	-
Water Solubility	1.30e-04 (1)	1.46e-02 (4)	1.30e-04	1.46e-02	1.30e-04	1.50e-04 to 5.71e-02	mol/L
Density	-	1.27 (1)	-	1.27	-	-	g/cm^3
Melting Point	174 (6)	151 (3)	175	151	173 to 177	114 to 185	°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	4.47e-06	7.21e-11	2.06e-07 to 1.27e-05	mmHg
LogKoa: Octanol-Air	-	8.38 (1)	-	8.38	-	-	-
Henry's Law	-	4.20e-10 (1)	-	4.20e-10	-	-	atm-m3/mole
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	-	-	-

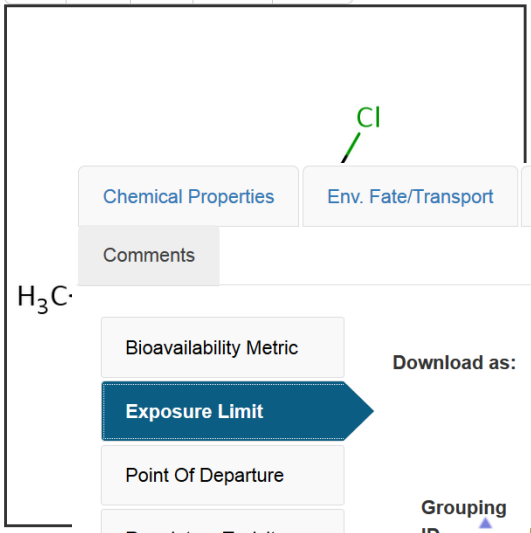
The Dashboard in brief

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

Atrazine

1912-24-9 | DTXSID9020112

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



Wikipedia

Atrazine is an herbicide of the triazine class. Atrazine is used to prevent pre- and postemergence broadleaf weeds in crops such as maize (corn) and sugarcane and on turf, such as golf courses and residential lawns. It is one of the most widely used herbicides

Chemical Properties | Env. Fate/Transport | Synonyms | External Links | Toxicity Values (Beta) | Exposure | Bioassays | Similar Molecules (Beta) | Literature | Comments

- Bioavailability Metric
- **Exposure Limit**
- Point Of Departure
- Regulatory Toxicity …
- Effect Level
- Misc Hazard Informa…
- Screening Level
- Uncertainty Factor

Download as: TSV | Excel

Exposure Limit

Grouping ID	Priority	Type	Subtype	Value	Units	Study Type	Exposure Route	Study Duration	Species	Media	Details	Source
5966	6	tolera…	-	2.00	ul/kg-d…	-	oral	-	-	-	RIVM …	RIVM
5967	6	tolera…	-	5.00	ul/kg-d…	-	oral	-	-	-	RIVM …	RIVM
48903	2	MCL-b…	-	1.90e-03	mg/kg…	-	-	-	-	-	RSL d…	RSL
176188	3	water …	drinkin…	0.7	mg/L	-	oral	chronic	-	drinkin…	EPA D…	ACToR
176190	3	MCL g…	drinkin…	3.00e-03	mg/L	-	oral	-	-	drinkin…	EPA D…	ACToR
176191	3	MCL	drinkin…	3.00e-03	mg/L	-	oral	-	-	drinkin…	EPA D…	ACToR

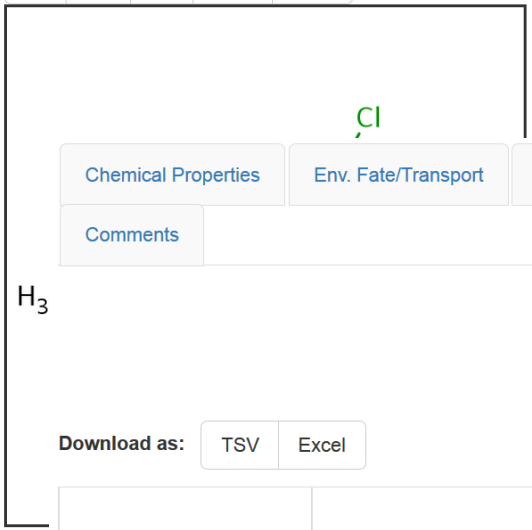
The Dashboard in brief

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

Atrazine

1912-24-9 | DTXSID9020112

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



Wikipedia

Atrazine is an herbicide of the triazine class. Atrazine is used to prevent pre- and postemergence broadleaf weeds in crops such as maize (corn) and sugarcane and on turf, such as golf courses and residential lawns. It is one of the most widely used herbicides

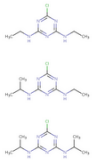
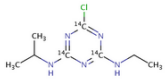
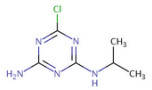
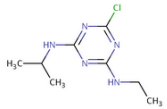
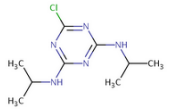
Chemical Properties | Env. Fate/Transport | Synonyms | External Links | Toxicity Values (Beta) | Exposure | Bioassays | Similar Molecules (Beta) | Literature | Comments

Similar Molecules

Searched with a similarity threshold of 0.82

Download as: TSV | Excel

Properties | Chemicals

						
Similarity Value		1	1	1	1	1
LogP: Octanol-Water	Experimental	-	-	1.51	2.61	2.93
	Predicted	-	2.77	1.58	2.76	3.15
Water Solubility	Experimental	-	-	0.00197	0.000130	0.0000374
	Predicted	-	0.0287	0.0108	0.0146	0.000514

Why should the Dashboard meet NORMAN?

Common Goals

- Much of the CompTox data is open – as is NORMAN data
  https://comptox.epa.gov/dashboard/downloads
- Increase access to data for use in other applications

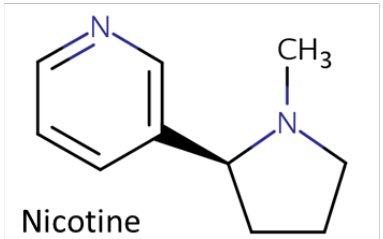
Mutual Benefits

- We have access to data they don't have, within a large European network
- We have additional predictive values & initiatives that are of interest
- They have access to data that we don't have
- They have years of investment/experience in data validation and curation
- They have the ability to provide services currently way beyond our means

Collaboration on Chemical Curation

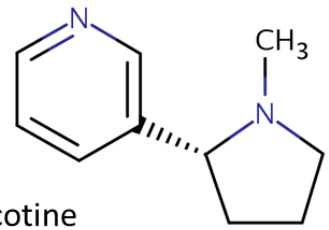
Schymanski & Williams, 2017, ES&T
DOI: 10.1021/acs.est.7b01908

LEGEND: Name, SMILES
DTXSID | InChIKey 1st Block
CAS | **Monoiso. Mass** | logP | **Sources**
Data on: **Toxicity | Exposure | Bioassays**



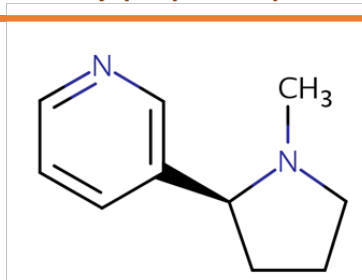
Nicotine

CN1CCC[C@H]1C1=CN=CC=C1
DTXSID1020930| SNICXCGAKADSCV
54-11-5 | **162.1157**| 0.929| **72**
Tox: **yes**| Expo: **yes**| Bioassay: **yes**



D-Nicotine

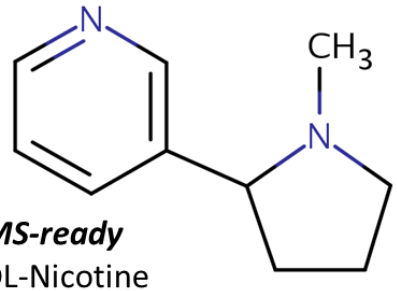
CN1CCC[C@@H]1C1=CN=CC=C1
DTXSID004635| SNICXCGAKADSCV
25162-00-9 | **162.1157**| 0.929| **20**
Tox: **no**| Expo: **yes**| Bioassay: **yes**



HCl

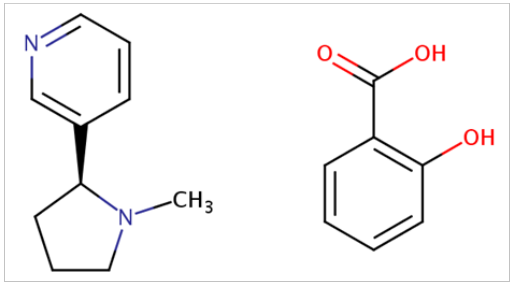
Nicotine hydrochloride

Cl.CN1CCC[C@H]1C1=CN=CC=C1
DTXSID602093| HDJBTCAJIMNXEW
2820-51-1 | **198.0924** | 0.929| **9**
Tox: **no**| Expo: **yes**| Bioassay: **yes**



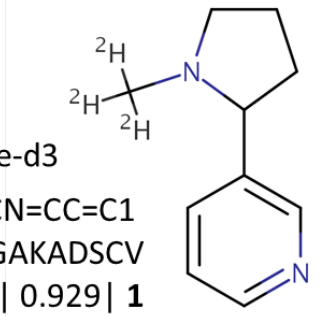
MS-ready
DL-Nicotine

CN1CCCC1C1=CN=CC=C1
DTXSID3048154 | SNICXCGAKADSCV
22083-74-5 | **162.1157**| 0.953| 9
Tox: **yes**| Expo: **no**| Bioassay: **yes**



Benzoic acid, 2-hydroxy-, compd. with 3-[(2S)-1-methyl-2-pyrrolidinyl]pyridine (1:1)

OC(=O)C1=C(O)C=CC=C1.CN1CCC[C@H]1C1=CN=CC=C1
DTXSID5075319| AIBWPBUAKCMKNS
29790-52-1| **300.1474**| 0.929| **6**
Tox: **no**| Expo: **yes**| Bioassay: **no**



DL-Nicotine-d3

[2H]C([2H])([2H])N1CCCC1C1=CN=CC=C1
DTXSID80442666| SNICXCGAKADSCV
69980-24-1| **165.1345**| 0.929| **1**
Tox: **no**| Expo: **no**| Bioassay: **no**

Collaboration on Chemical Curation of Lists

Pharmaceutical List with Consumption Data	SwissPharma_TableS2.csv	SwissPharma_TableS2_InChIKeys.txt	Singer *et al.* 2016. DOI: **10.1021/acs.est.5b03332**
Swiss Insecticides, Fungicides and TPs	SwissPesticides_TableS1.csv	SwissPesticides_TableS1_InChIKeys.txt	Moschet *et al.* 2013. DOI: **10.1021/ac4021598**
NormaNEWS for retrospective screening of new emerging contaminants	NormaNEWS_V4_26042017.csv	NormaNEWS_V4_InChIKeys.txt	**NormaNEWS** list provided by Nikiforos Alygizakis, Saer Samanipour and Kevin Thomas
Combined Inventory of Ingredients Employed in Cosmetic Products (2000) and Revised Inventory (2006)	Merged_CosmeticProducts_04052017.csv	Merged_CosmeticProducts_04052017_InChIKeys.txt	The scientific committee on cosmetic products and non-food products Intended for consumers - **SCCNFP/0389/00 Final** and Commission **Decision 2006/257/EC** amending the Decision 96/335/EC. Provided by Peter von der Ohe, UBA, curated by Reza Aalizadeh, University of Athens
PFAS Highly fluorinated substances list: KEMI	PFAS_Market_Kemi_EPA_1Feb2017.xlsx ~2,600 PFAS	Curation in progress: coming soon	Appendix 2 from Swedish Chemicals Agency **KEMI Report 7/15**. Provided by Stellan Fischer, KEMI
NORMAN Priority List 2015	NORMAN_PriorityList_2016.csv Further curation in progress...	NORMAN_PriorityList_2016_InChIKeys.txt	Priority substances from NORMAN WG-1 (Prioritisation), provided by Valeria Dulio
French Monitoring List	French_List_08052017.csv Further curation in progress	FrenchList_UniqueInChIKeys_08052017.txt	Provided by Valeria Dulio, curated by Reza Aalizadeh, University of Athens
KEMI Market List	KEMI_MarketList_12052017_MSready.xlsx 24,883 Substances (Expo, Hazard Scores)	KEMI_MarketList_12052017_MSready_InChIKeys.txt	Provided by Stellan Fischer, KEMI including Hazard and Exposure scores, documented **here**. Curated by Reza Aalizadeh, University of Athens.
TSCA Surfactants	Coming soon...		Provided by Lee Ferguson, sourced from James Little

KEMI PFAS List (Stellan Fischer)

PFAS Highly fluorinated substances list: KEMI	PFAS_Market_Kemi_EPA_1Feb2017.xlsx ~2,600 PFAS	Curation in progress: coming soon	Appendix 2 from Swedish Chemicals Agency **KEMI Report 7/15**. Provided by Stellan Fischer, KEMI

CASno	CASnr	ECno	DTXSID	PREFERRED NAME	CASRN	SMILES
422-63-9	422639	207-020-0	DTXSID9059969	1,1-Propanediol, 2,2,3,3,3-pentafluoro	422-63-9	OC(O)C(F)(F)C(F)(F)F
375-88-2	375882	206-799-4	DTXSID9059919	Heptane, 1-bromo-1,1,2,2,3,3,4,4,5,5,6	375-88-2	FC(F)(F)C(F)(F)C(F)(F)C(F)(F)C
375-62-2	375622	206-790-5	DTXSID9059917	Pentanoyl fluoride, nonafluoro-	375-62-2	FC(=O)C(F)(F)C(F)(F)C(F)(F)C(
375-16-6	375166	206-785-8	DTXSID9059915	Butanoyl chloride, heptafluoro-	375-16-6	FC(F)(F)C(F)(F)C(F)(F)C(Cl)=O
375-00-8	375008	206-781-6	DTXSID9059913	Butanenitrile, heptafluoro-	375-00-8	FC(F)(F)C(F)(F)C(F)(F)C#N
356-86-5	356865	206-608-4	DTXSID9059884	2,2,3,3,3-Pentafluoropropyl acrylate	356-86-5	FC(F)(F)C(F)(F)COC(=O)C=C
356-27-4	356274	206-602-1	DTXSID9059882	Ethyl heptafluorobutyrate	356-27-4	CCOC(=O)C(F)(F)C(F)(F)C(F)(F
338-83-0	338830	206-420-2	DTXSID9059834	1-Propanamine, 1,1,2,2,3,3,3-heptaflu	338-83-0	FC(F)(F)C(F)(F)C(F)(F)N(C(F)(F
335-99-9	335999	206-406-6	DTXSID9059832	1-Heptanol, 2,2,3,3,4,4,5,5,6,6,7,7-doc	335-99-9	OCC(F)(F)C(F)(F)C(F)(F)C(F)(F
33						F)C
35						(F)
89						(F)
57						(F)C
76						
38						
42						
73						C(F)
85						F)C
30						C(F)
89						F)C

Norman Network PFAS (KEMI Report)

Search SFISHFLUORO Chemicals

List Details

Description: This list of perfluorinated substances originated from Appendix 2 from Swedish Chemicals Agency Report 7/15 on the occurrence and use of highly fluorinated substances and alternatives (2015). The current KEMI PFAS list includes substances beyond the original report and was provided by Stellan Fisher.

Number of Chemicals: 970

NormaNEWS

NormaNEWS for retrospective screening of new emerging contaminants	**NormaNEWS_V4_26042017.csv**	**NormaNEWS_V4_InChIKeys.txt**	**NormaNEWS** list provided by Nikiforos Alygizakis, Saer Samanipour and Kevin Thomas

INPUT	DTXSID	PREFERRED NAME	CASRN	IUPAC NAM	SMILES
DTXSID40881097	DTXSID40881097	C11-LAS	NOCAS_881097	-	CCCCCCC(CCCC)C1=CC=C(C=C1)S(O)(=O)=O
DTXSID30860093	DTXSID30860093	4-(Dodecan-6-yl)benze	23003-92-1	4-(Dodecan-	CCCCCCC(CCCCC)C1=CC=C(C=C1)S(O)(=O)=O
DTXSID80881096	DTXSID80881096	C13-LAS	NOCAS_881096	-	CCCCCCCCCC(CCC)C1=CC=C(C=C1)S(O)(=O)=O
DTXSID20881095	DTXSID20881095	C14-LAS	NOCAS_881095	-	CCCCCCCCCCC(CCC)C1=CC=C(C=C1)S(O)(=O)=O
DTXSID60881094	DTXSID60881094	SPA-8C	NOCAS_881094	-	CCCC(CCCC(O)=O)C1=CC=C(C=C1)S(O)(=O)=O
DTXSID50865484	DTXSID50865484	10-hydroxycarbazepin	29331-92-8	10-Hydroxy-	NC(=O)N1C2=CC=CC=C2CC(O)C2=CC=CC=C12
DTXSID00881093	DTXSID00881093	Desacetyl diltiazem	42399-40-6	-	[H][C@]1(SC2=C(C=CC=C2)N(CCN(C)C)C(=O)[C@@H]

NormaNEWS: Norman Early Warning System

Search NORMANEWS Chemicals

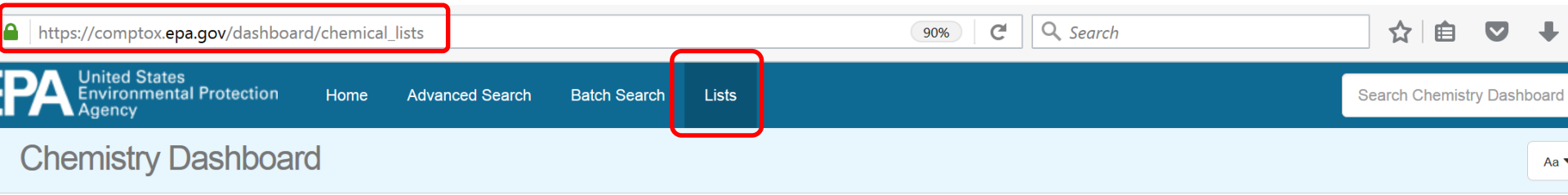
List Details

Description: The Norman Early Warning System (NormaNEWS) is a pilot network designed to investigate the spatial and temporal distribution of newly identified contaminants of emerging concern in environmental samples through performing retrospective suspect screening on HRMS data acquired using different instrumental platforms and data processing software. The NormaNEWS pilot study was performed through recruiting eight reference laboratories with available archived HRMS data with the goal of exploring the potential of an early warning network to rapidly establish the occurrence of newly-identified contaminants of emerging concern across Europe and beyond, through the use of retrospective suspect screening employing HRMS. The pilot study was referred to as the Norman Early Warning System, abbreviated to NormaNEWS.

Number of Chemicals: 131

List Functionality in the Dashboard

An overview of all the lists …



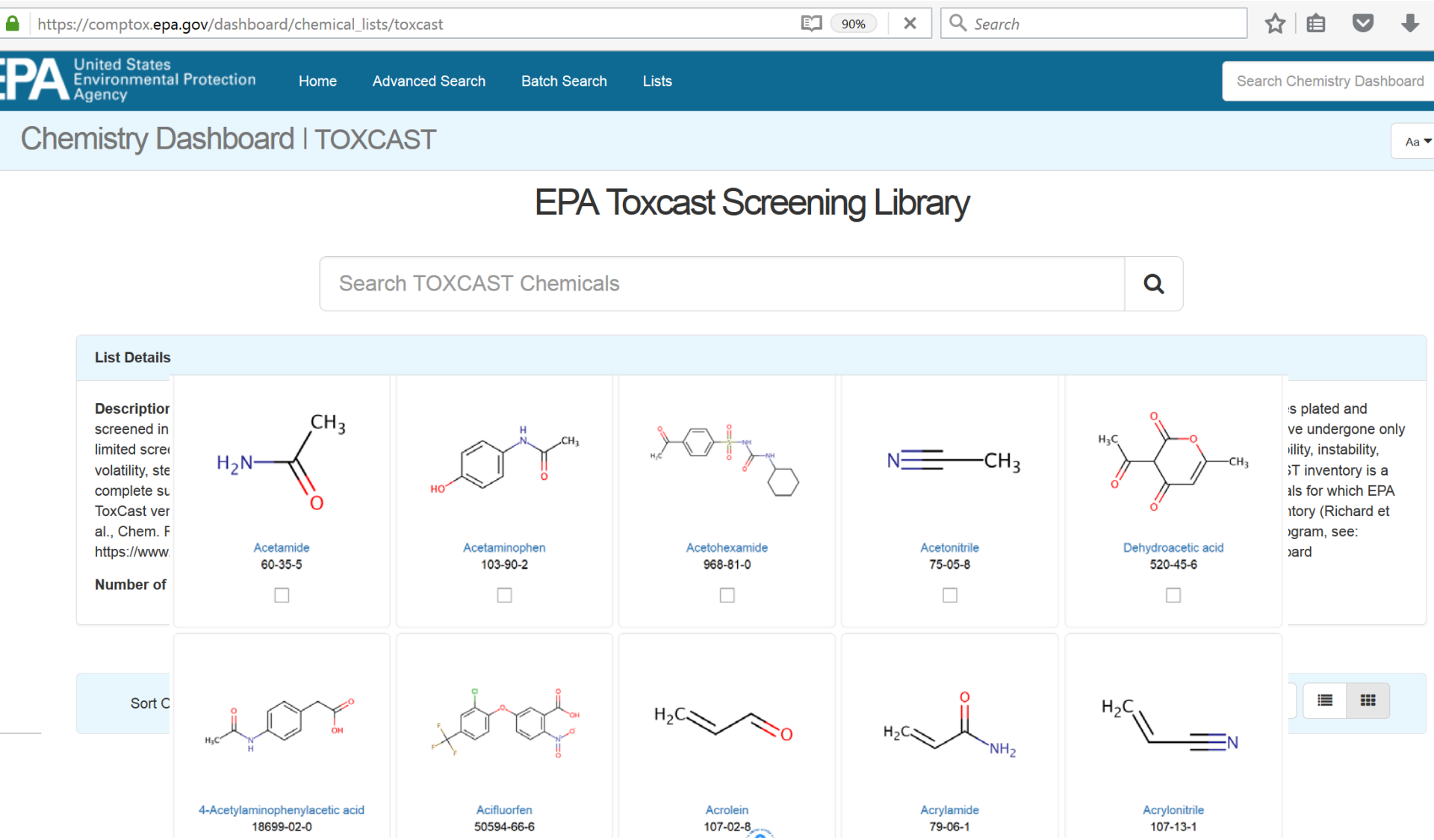
Select List

List Name	Number of Chemicals	List Description
CHEMINV: EPA Chemical Inventory for ToxCast (20170203)	5231	CHEMINV is full list of unique DSSTox substances mapped to historical chemical inventory of physical samples registered by EPA's ToxCast Chemical Contractor (Evotec) since launch of ToxCast program in 2007.
DNT Screening Library	1476	DNTSCREEN is a list of chemicals that is being used in medium- and high-throughput in vitro and zebrafish assays.
EPA Toxcast Screening Library	4736	TOXCAST includes all EPA-provided chemicals for which screening data have been generated in the ToxCast research program since 2007.
Tox21 Screening Library	8947	TOX21SL is list of unique substances in Tox21 multi-federal agency screening library, contributed by the EPA, National Toxicology Program (NTP), and National Center for Advances in Translational Science (NCATS).

More lists become available with every release

List Functionality in the Dashboard

Obtaining all chemicals from a given list (downloadable as XLS or SDF)



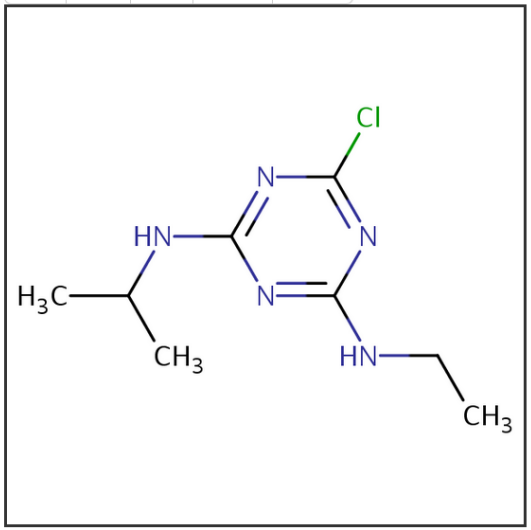
List Functionality in the Dashboard

Individual Chemical Level

Atrazine

1912-24-9 | DTXSID9020112

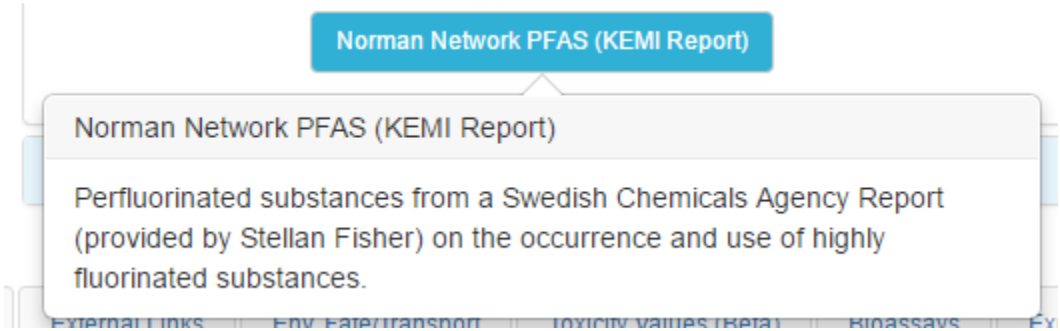
Searched by Approved Name: Found 1 result for 'atrazine'.

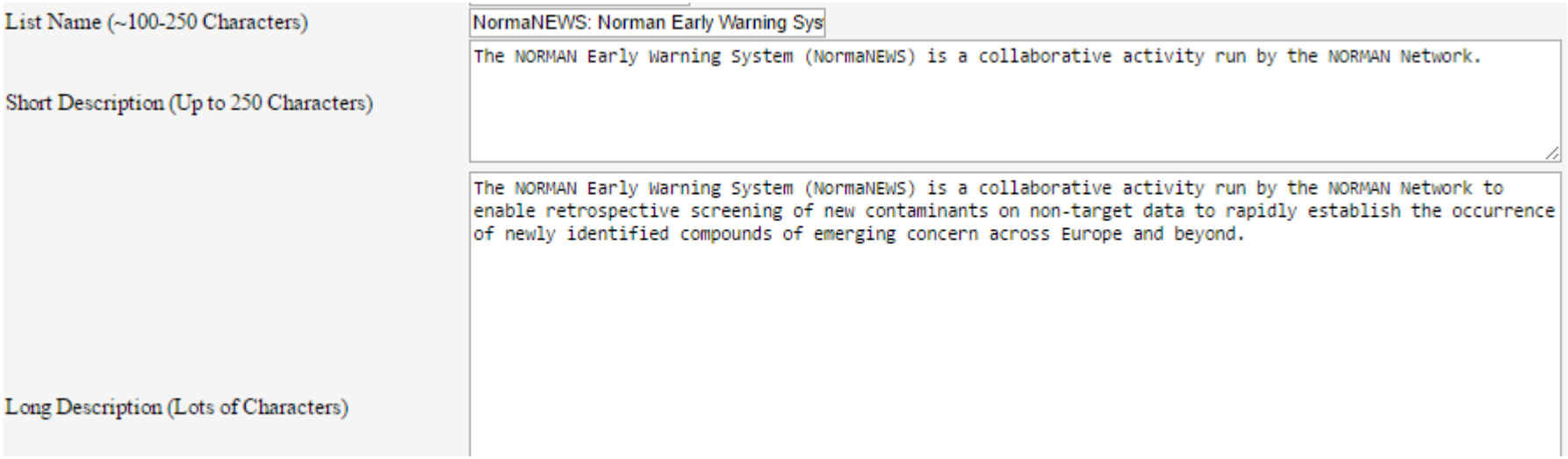


- Wikipedia
- Intrinsic Properties
- Structural Identifiers
- Related Compounds (Beta)
- Presence in Lists
  - DNT Screening Library
  - CHEMINV: EPA Chemical Inventory for ToxCast (20170203)
  - EPA Toxcast Screening Library
  - Tox21 Screening Library
- Record Information

NORMAN Lists in the Dashboard …

Coming soon …





Future work … integrating DTXSIDs into NORMAN &

Undefined mixtures (UVCBs)

Cleaning up lists to remove errors

Mol_ID	Name	EDITED NAMES FOR INPUT INTO SEARCH	CAS_RN	Merged DTXSIDs	DTXSID Based on Name	Preferred Name
SA8750	By-Product	By-Product	NA	-	-	NO_MATCH
stpQQR1546	C10-DATS|C10-Dialkyl tetr	C10-DATS|C10-Dialkyl tetralin sulfonate|8	NA	-	-	NO_MATCH
SA2074	C10-LAS	C10-LAS	NA	-		NO_MATCH
stpQQR1582	C10LAS|C10-linear alkylbe	C10LAS|C10-linear alkylbenzyl sulfonate|4	NA	-	-	NO_MATCH
SA14931	C10-phosphonic	C10-phosphonic	NA	-	-	NO_MATCH
StpBB815	C12-15 ALKYL BENZOATE	C12-15 ALKYL BENZOATE	68411-27-8	-	-	NO_MATCH
SA13282	C12-AE3S	C12-AE3S	NA	-	-	NO_MATCH
stpQQR1548	C12-LAS|C12-linear alkyl b	C12-LAS|C12-linear alkyl benzene sulfonat	NA	-	-	NO_MATCH
stpQQR690	C14-SAS (TENTATIVE)|tetr	C14-SAS (TENTATIVE)|tetradecane-7-sulfo	NA	-	-	NO_MATCH
stpQQR1557	C16EOx|C16EO2|C16-alcc	C16EOx|C16EO2|C16-alcohol polyethoxyl	NA	-	-	NO_MATCH
stpQQR1556	C18EOx|C18EO2|C18-alcc	C18EOx|C18EO2|C18-alcohol polyethoxyl	4439-32-1	-	-	NO_MATCH
SA14932	C4-phosphonic	C4-phosphonic	NA	-	-	NO_MATCH
SA14929	C6-phosphonic	C6-phosphonic	NA	-	-	NO_MATCH
stpQQR1583	C7SPC|C7-sulfophenyl car	C7SPC|C7-sulfophenyl carboxylates|4-(de	NA	-	-	NO_MATCH
SA14930	C8-phosphonic	C8-phosphonic	NA	-	-	NO_MATCH
stpQQR1547	C8-SPC|C8-Sulfophenyl ca	C8-SPC|C8-Sulfophenyl carboxylic acid|4-(	NA	-	-	NO_MATCH
stpQQR1576	CA5PE2C|7-{4-[2-(carboxy	CA5PE2C|7-{4-[2-(carboxymethoxy)ethoxy	NA	-	-	NO_MATCH
stpQQR1578	CA6PE2	CA6PE2	NA	-	-	NO_MATCH
stpQQR1577	CA6PE2C	CA6PE2C	NA	-	-	NO_MATCH
stpQQR1575	CA8PE2C	CA8PE2C	NA	-	-	NO_MATCH
SA9863	cacotheline	cacotheline	561-20-6	-	-	NO_MATCH
SAn15715	Caerulomycin A	Caerulomycin A	21802-37-9	-	-	NO_MATCH
SA5151	cafedrine	cafedrine	58166-83-9	-	-	NO_MATCH

(many) more registrations…

Handling of Undefined Mixtures

C10-12 chloroalkanes

https://comptox.epa.gov/dashboard/dsstoxdb/results?utf8=✓&search=C10-12+chloroalkanes

C10-12 chloroalkanes

108171-26-2|DTXSID10872316

Searched by DSSTox_Substance_Id: Found 1 result for 'DTXSID10872316'.

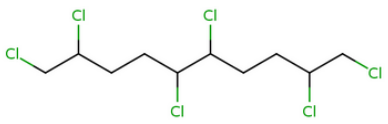
Presence in Lists

Record Information

Quality Control Notes

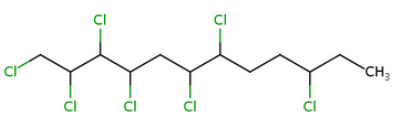
Related Chemicals

Found 3 chemicals



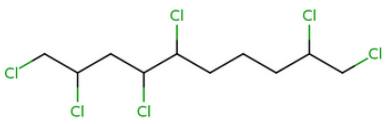
1,2,5,6,9,10-Hexachlorodecane

189350-94-5



1,2,3,4,6,7,10-Heptachlorododecane

1005111-47-6



1,2,4,5,9,10-Hexachlorodecane

890302-87-1

Acknowledgements

…after almost 1000 emails, dozens of phone calls, a LOT of extremely valuable input and uncountable hours of late night, after hours work, a HUGE thanks goes to Tony Williams and his team at the Dashboard

Stellan Fischer, KEMI

Kevin Thomas (UQ), Saer Samanipour (NIVA), Nikiforos Alygizakis (EI),

Andrew McEachran, US EPA

Questions?

Dashboard:

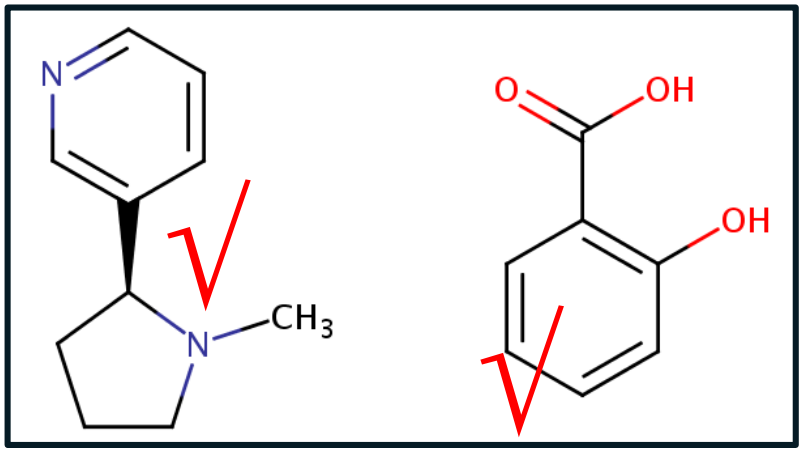
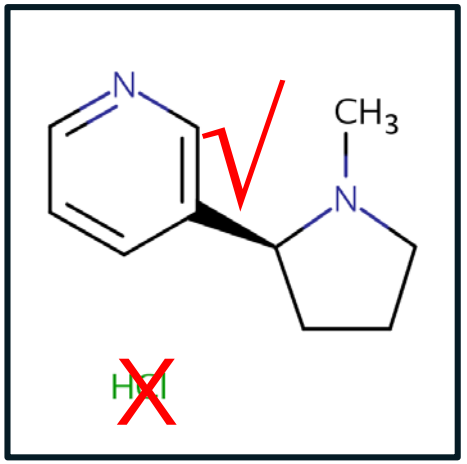
https://comptox.epa.gov/

Contact:

emma.schymanski@eawag.ch

EU Grant 603437

Mixtures (work in progress)

Mixture identification and curation

- Step 1: Identify all mixture components
- Step 2: Create SMILES, split into components, retain “non-salts”
- Step 3: Create MS-ready identifiers, compare retained components
  - Single structure => all the same, treat as individual
  - Otherwise save as Level 7 for validation …