

Structure identification by Mass Spectrometry Non-Targeted Analysis using the US EPA's CompTox Chemistry Dashboard

*Antony Williams¹, Andrew D. McEachran², Chris Grulke¹,
Seth Newton³, Kristin Isaacs³, Katherine Phillips³,
Nancy Baker¹ and Jon R. Sobus³*

- 1) *National Center for Computational Toxicology, U.S. Environmental Protection Agency, RTP, NC*
- 2) *Oak Ridge Institute of Science and Education (ORISE) Research Participant, Research Triangle Park, NC*
- 3) *National Exposure Research Laboratory, U.S. Environmental Protection Agency, RTP, NC*

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

*March 2018
ACS Spring Meeting, New Orleans*

Two Years of Development for the CompTox Chemistry Dashboard

- The Chemistry Dashboard went online on April 1st 2016
- Two years later, with 10k users a week, it is fulfilling the promise with an underlying architecture for integrating CompTox data
- Used by our mass spectrometry team (in NERL) on a daily basis – suspect screening and non-targeted analysis

Published Many Times



ELSEVIER

Contents lists

Environ

journal hom

Journal of Exposure Science & Environmental Epidemiology



Altmetric: 4 Citations: 1

[More detail >>](#)

Suspect screening and non-targeted analysis filters[☆]

Seth R. Newton^{a,*}, Rebecca L. McMahan^b, Jon R. Sob
Andrew D. McEachran^{b,c}, Mark J. Strynar^a

Review Article | [OPEN](#)

Integrating tools for non-targeted analysis



[Environ. Sci. Technol.](#)

[Environ. Sci. Technol.](#)



[Browse the Journal](#)

[Articles ASAP](#)

[Cu](#)

[Environ. Sci. Technol.](#)

[Environ. Sci. Technol. Lett.](#)



[Browse the Journal](#)

[Articles ASAP](#)

[Current Issue](#)

[Multimedia](#)

[Submission & Review](#)

Article

Suspect Screening Analysis of

Katherine A. Phillips[†] , Alice Yau[‡], Kristin A. Grulke[‡], Ann M. Richard[‡], Antony J. Williams[‡]

Viewpoint

Open Science for Identifying “Known Unknown” Chemicals

Emma L. Schymanski^{††}  and Antony J. Williams^{††} 

- A **publicly accessible website** delivering access:
 - ~760,000 chemicals with related property data
 - Experimental and predicted physicochemical property data
 - Experimental Human and Ecological hazard data
 - Integration to “biological assay data” for 1000s of chemicals
 - Information regarding consumer products containing chemicals
 - Links to other agency websites and public data resources
 - “Literature” searches for chemicals using public resources
 - “Batch searching” for thousands of chemicals
 - Real time prediction of physchem and toxicity endpoints

CompTox Chemistry Dashboard

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



Home Advanced Search Batch Search Lists Predictions Downloads

Chemistry Dashboard

Aa Aa Aa



761 Thousand Chemicals

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

Identifier substring search

See what people are saying, read the dashboard comments!

Latest News

[Read more news](#)

A Movie Regarding how to Identify "Known Unknowns" Using the CompTox Dashboard

March 28th, 2017 at 7:35:41 PM

Recently we published a paper regarding [Identifying known unknowns using the US EPA's CompTox Chemistry Dashboard](#). Analytical and Bioanalytical Chemistry, March 2017, Volume 409, Issue 7, pp 1729–1735. A movie explaining the paper in full animated detail has been put on YouTube. Enjoy the movie interlude [here](#).



Discover.

[About/Disclaimer](#)
[Accessibility](#)

Connect.

[ACToR](#)
[DSSTox](#)

Ask.

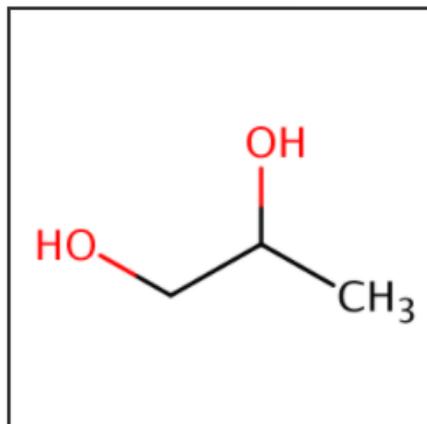
[Contact](#)
[Help](#)

[Submit Comment](#)[Copy](#)[Aa](#)[Aa](#)[Aa](#)

1,2-Propylene glycol

57-55-6 | DTXSID0021206

© Searched by DSSTox_Substance_Id: Found 1 result for 'DTXSID0021206'.



Wikipedia

Propylene glycol (IUPAC name: propane-1,2-diol) is a synthetic organic compound with the chemical formula C₃H₈O₂. It is a viscous colorless liquid which is nearly odorless but possesses a faintly sweet taste. Chemically it is classed as a diol and is miscible with a broad range of solvents, including water, acetone, and chloroform. It is produced on a large scale and is primarily used in the production of polymers, but also sees use in food...[Read more](#)

Intrinsic Properties

Structural Identifiers

Linked Substances

Presence in Lists

Record Information

Quality Control Notes

[Chemical Properties](#)[Env. Fate/Transport](#)[Hazard](#)[ADME \(Beta\)](#)[Exposure](#)[Bioassays](#)[Similar Compounds](#)[Related Substances](#)[Synonyms](#)[Literature](#)[Links](#)[Comments](#)

Access to Chemical Hazard Data

Chemistry Dashboard | EPAHFR

Submit Comment

Copy 

Aa  Aa  Aa 

Chemical Properties

Env. Fate/Transport

Hazard

ADME (Beta)

Exposure

Bioassays

Similar Compounds

Related Substances

Synonyms

Literature

Links

Comments

Exposure Limit

Lethality Effect Level

Point of Departure

Toxicity Value

Download table as: TSV Excel

Human

Eco

	Priority	Type	Subtype	Risk Assessment Class	Values	Units	Study Type	Exposure Route	Species	Subsource	Source
	8	NOEL	Cardiova...	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vailla et ...	PPRTV (...)
	8	NOEL	Endocrine	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vailla et ...	PPRTV (...)
	8	LOEL	Hematol...	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vailla et ...	PPRTV (...)
	8	LOEL	Hepatic	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vailla et ...	PPRTV (...)
	8	NOEL	Immune	immunot...	5000.0	mg/kg-day	subchronic	oral	rat	Vailla et ...	PPRTV (...)
	8	NOEL	Renal	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vailla et ...	PPRTV (...)
	8	LOEL	Systemic	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vailla et ...	PPRTV (...)
	8	NOEL	Hematol...	subchronic	1500.0	mg/kg-day	subchronic	oral	rabbit	Vailla et ...	PPRTV (...)
	8	NOEL	Systemic	subchronic	1500.0	mg/kg-day	subchronic	oral	rabbit	Vailla et ...	PPRTV (...)

In Vitro Bioassay Screening

ToxCast and Tox21

Chemistry Dashboard | EPAHFR

Submit Comment Copy Aa Aa Aa

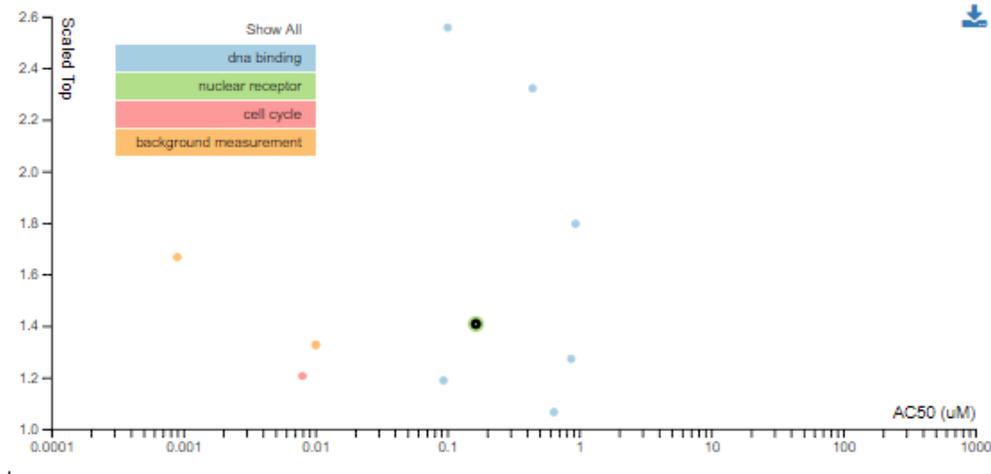
Chemical Properties Env. Fate/Transport Hazard ADME (Beta) Exposure Bioassays Similar Compounds Related Substances Synonyms Literature Links Comments

ToxCast: Summary

PubChem

Chemical Activity Summary ⓘ

ⓘ ToxCast Data



ⓘ Assay Details

AC50 (uM): 0.16
Scaled top: 1.41
Assay Name: NVS_NR_hFXR_Antagonist
Assay Description: 716
Gene Symbol: NR1H4
Organism: human
Tissue: NA
Assay Format Type: biochemical
Biological Process Target: receptor binding
Detection Technology: TR-FRET
Analysis Direction: positive
Intended Target Family: nuclear receptor
Description: Data from the assay component NVS_NR_hFXR_Antagonist was analyzed into 1 assay endpoint. This assay endpoint, NVS_NR_hFXR_Antagonist, was analyzed in the positive fitting direction relative to DMSO as the negative control and baseline of activity. Using a type of binding reporter, loss-of-signal activity can be used to understand changes in the binding as they relate to the gene NR1H4.

Download as: TSV Excel Show: Inactive Background

Assay Name Assa... ⓘ SeqA... AOP Link AOP Eve... ⓘ Hit Call ⓘ T... Scale... ⓘ ... log ... ⓘ Target Family

Sources of Exposure to Chemicals

Exposure [Bioassays](#) [Similar Compounds](#) [Related Substances](#) [Synonyms](#) [Literature](#) [Links](#) [Comments](#)

Product & Use Categories

Chemical Weight Fraction

Chemical Functional Use

Monitoring Data

Exposure Predictions

Production Volume

Product & Use Categories (PUCs)

Categorization type	Number of Unique Products
PUC	288
PUC	208
PUC	117
PUC	107
PUC	107
PUC	101
PUC	101
PUC	90
PUC	89

Identifiers to Support Searches

Chemical Properties

Env. Fate/Transport

Hazard

ADME (Beta)

Exposure

Bioassays

Similar Compounds

Related Substances

Synonyms

Found 78 synonyms

Legend: **Valid Synonyms** *Good Synonyms* *Other Synonyms*

 Copy all Synonyms

1,2-Propylene glycol

Propane-1,2-diol

1,2-Propanediol

57-55-6 Active CAS-RN

alpha-Propylene glycol

(+/-) 1,2-Propanediol

(RS)-1,2-Propanediol

dl-Propylene glycol

3-01-00-02142 Beilstein Registry Number

1,2-Propanediol

(+,-)-1,2-Propanediol

(+,-)-Propylene glycol

1,2-(RS)-Propanediol

1,2-DIHYDROXYPROPANE

1,2-PROPANDIOL

Literature Searches and Links

Google Scholar

PubMed Abstract Sifter

PubChem Articles

PubChem Patents

PPRTV

IRIS

Exposure Bioassays Similar Compounds Related Substances Synonyms Literature Links Comments

Query then 2) click on Retrieve. **i**

Retrieve Articles **i**

13 of 13 articles loaded...

Optionally, edit the query before retrieving.

("57-55-6" OR "1,2-Propylene glycol" OR "Propylene Glycol") AND (NOAEL or NOEL OR LOEL or Rfd OR "reference dose" OR "reference concentration" OR "adverse effect level"[tiab] OR "cancer slope factor"[tiab])

Download / Send to... **i** Download Sifter for Excel **i**

	Authors	Journal	Rev
...igarettes: a systematic review of available studies on hea...	Zulkifli; Abidin; Abidin; Amer Nordin; Praveena; Sye...	Reviews on environmental health	
...assessment of a prototype e-cigaret device and three fl...	Werley; Kirkpatrick; Oldham; Jerome; Langston; Lill...	Inhalation toxicology	
...monitoring Equivalents for selected E- and P-series glyco...	Poet; Ball; Hays	International journal of hygiene and environmental h...	
...mful health effects of inhaling nicotine-free shisha-pen v...	Kienhuis; Soeteman-Hernandez; Bos; Cremers; Kle...	Tobacco induced diseases	
...e: Developmental and reproductive toxicity potential of ...	Glynn; Jo; Minowa; Sanada; Nejishima; Matsuuchi; ...	Reproductive toxicology (Elmsford, N.Y.)	
...afety assessment of Eflinaconazole Solution (10%) for o...	Jo; Glynn; Nejishima; Sanada; Minowa; Calvarese; ...	Regulatory toxicology and pharmacology : RTP	
...id formulations for intravenous mouse pharmacokinetic ...	Thackaberry; Wang; Schweiger; Messick; Valle; De...	Xenobiotica: the fate of foreign compounds in biolog...	
...safety and pharmacokinetic evaluations of propylene gly...	Werley; McDonald; Lilly; Kirkpatrick; Wallery; Byron;...	Toxicology	
...in the safety assessment of methoxyisopropanol and m...		International journal of toxicology	
...logically-based pharmacokinetic modeling to address n...	Kiman; Sweeney; Corley; Gargas	Risk analysis : an official publication of the Society f...	
...of 2-methoxypropionic acid formed from beta-propylene...	Carney; Pottenger; Johnson; Liberacki; Tomesi; Dry...	Toxicological sciences : an official journal of the Soc...	

... to sift abstracts. **i**

Google Scholar

PubMed Abstract Sifter

PubChem Articles

PubChem Patents

PPRTV

IRIS

Select Term:

Exposure

Edit the Query Be

"Exposure" AND

"Exposure" AND "57-55-6" OR "1,2-Propylene glycol"

About 5,560 results (0.22 sec)

Submit

Cardiac morbidity and mortality associated with occupational exposure to 1, 2 propylene glycol dinitrate.

SA Forman, JC Helmkamp, CM Bone - Journal of occupational ..., 1987 - europepmc.org
Abstract Myocardial infarction and angina pectoris are conditions long associated with occupational **exposure** to nitroglycerin and related explosives. Cardiac sentinel events in selected munitions workers exposed to the related nitrated ester **1, 2 propylene glycol** ...

☆ ⓘ Cited by 14 Related articles All 6 versions Web of Science: 4 ⓘ

Experimental exposure to propylene glycol mist in aviation emergency training: acute ocular and respiratory effects

G Wieslander, D Norbäck, T Lindgren - ... and Environmental Medicine, 2001 - oem.bmj.com
... to high concentrations of propylene glycol in workplaces and public places may cause ocular and respiratory irritation, and that sensitive subjects should be protected or avoid extreme or prolonged **exposure**. Propylene glycol (PG) (1-2 propanediol; CAS nr **57-55-6**) is a ...

☆ ⓘ Cited by 140 Related articles All 14 versions Web of Science: 81 ⓘ

Evaluation of the neurophysiologic effects of 1, 2-propylene glycol dinitrate by quantitative ataxia and oculomotor function tests

EP Horvath, RA Ilka, J Boyd... - American journal of ..., 1981 - Wiley Online Library
... **1,2-Propylene glycol** dinitrate (PGDN), a nitrated ester found in the torpedo propellant Otto Fuel 11, has been suspected of causing neurologic and ... of chronic neurotoxicity was found, even among a subgroup of workers (CE~UB) with the longest total duration of **exposure** ...

☆ ⓘ Cited by 17 Related articles All 4 versions

[HTML] Propylene Glycol Dinitrate

C on Emergency, CEGL for Selected... - 2009 -.ncbi.nlm.nih.gov
... Toxicol. Lett. 43(1-3):51-65. [PubMed: 3051528]. Forman, SA, JC Helmkamp, and CM Bone. 1987. Cardiac morbidity and mortality associated with occupational **exposure** to **1,2 propylene glycol** dinitrate. J. Occup. Med. 29(5):445-450 ...

☆ ⓘ Related articles

External Links to Data and Services

[Chemical Properties](#)
[Env. Fate/Transport](#)
[Hazard](#)
[ADME \(Beta\)](#)
[Exposure](#)
[Bioassays](#)
[Similar Compounds](#)
[Related Substances](#)
[Synonyms](#)
[Literature](#)
[Links](#)

General

[EPA Substance Registry ...](#)
[Household Products Data...](#)
[PubChem](#)
[CPCat](#)
[DrugBank](#)
[Wikipedia](#)
[MSDS Lookup](#)
[ChEMBL](#)
[Chemical Vendors](#)
[NIOSH Chemical Safety ...](#)
[ToxPlanet](#)
[ACS Reagent Chemicals](#)
[Wikidata](#)
[ChemHat: Hazards and A...](#)
[Wolfram Alpha](#)

Toxicology

[ACToR](#)
[DrugPortal](#)
[CCRIS](#)
[ChemView](#)
[CTD](#)
[eChemPortal](#)
[Gene-Tox](#)
[HSDB](#)
[ToxCast Dashboard 2](#)
[LactMed](#)
[International Toxicity Esti...](#)
[ATSDR Toxic Substances...](#)
[ACToR PDF Report](#)
[CREST](#)

Publications

[Toxline](#)
[Environmental Health Per...](#)
[NIEHS](#)
[National Toxicology Progr...](#)
[Google Books](#)
[Google Scholar](#)
[Google Patents](#)
[PPRTVWEB](#)
[PubMed](#)
[IRIS Assessments](#)
[EPA HERO](#)
[RSC Publications](#)
[BioCaddie DataMed](#)
[Springer Materials](#)
[Federal Register](#)

Analytical

[RSC Analytical Abstracts](#)
[Tox21 Analytical Data](#)
[MONA: MassBank North ...](#)
[NIST NIST IR Spectrum](#)
[NIST NIST MS Spectrum](#)

Prediction

[2D NMR HSQC/HMBC Pr...](#)
[Carbon-13 NMR Prediction](#)
[Proton NMR Prediction](#)
[ChemRTP Predictor](#)
[LSERD](#)

Mass Spec Data

[Hazard](#) | [ADME \(Beta\)](#) | [Exposure](#) | [Bioassays](#) | [Similar Compounds](#) | [Related Subst](#)

MassBank of America (MoNA) is a metadata-centric, auto-curating repository designed for efficient storage and querying of mass spectral records. It intends to serve as the framework for a centralized, collaborative database of metabolite mass spectra, metadata and associated compounds. MoNA currently contains over 200,000 mass spectral records from experimental and in-silico libraries as well as from user contributions.

[Applications](#)

[Publine](#)

[Environment](#)

[IEHS](#)

[National Toxicology](#)

[Google Books](#)

[Google Scholar](#)

[QC/HMBC](#)

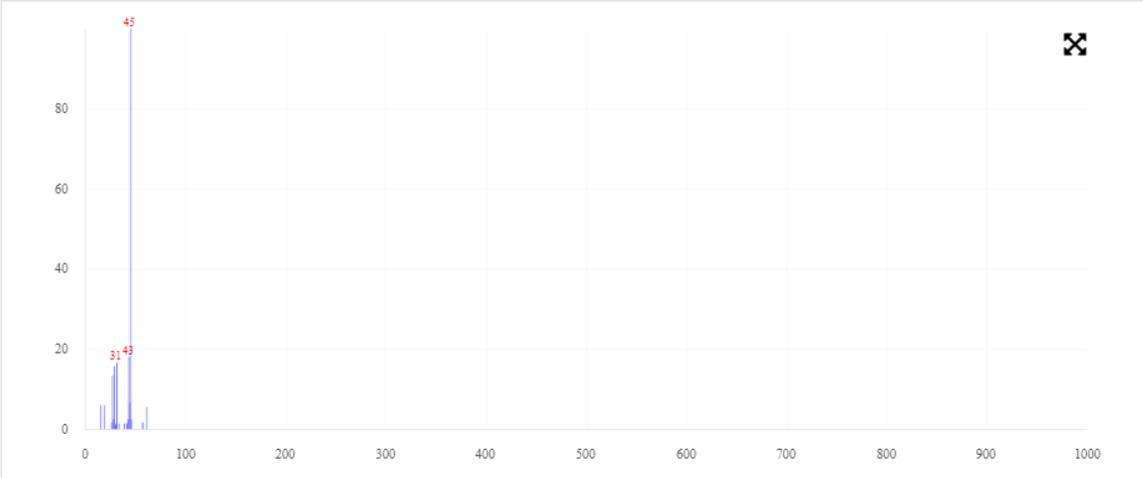
[MR-Pradi](#)

MoNA - MassBank of North America [Mass Spectra](#) [Downloads](#) [Upload](#) [Help](#)

[MassBank](#) [GC-MS](#) 

SPLASH: [Q splash10-0002-9000000000-6d2345a66e2cf8d2d379](#) Submitter: [Q Kimito Funatsu](#)

Mass Spectrum 



m/z	Relative Intensity
31	~15
43	~15
45	100

Ion Table / Peak Table 

- Structure Identification using the dashboard
 - Formula/mass-based searching – 1 chemical at a time

Advanced Search

Mass Search

Min/Max ▼

Mass Da Error Da ppm

Molecular Formula Search

MS Ready Formula 
 Exact Formula 

Generate Molecular Formula(e)

Min/Max

Mass Da Error Da ppm

Default Options: C[1-50] H[0-100] O[0-20] N[0-20] P[0-20] S[0-10]

Include Halogens: F[0-20] Cl[0-20] Br[0-20] I[0-20]

Options ▼

Advanced Searches

Mass Based Search

Mass Search

Min/Max

Da

Advanced Searches

 United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

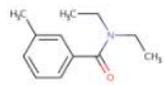
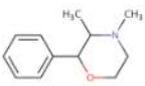
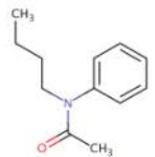
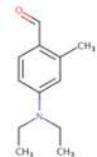
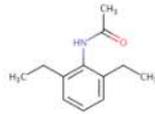
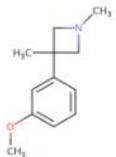
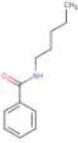
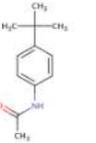
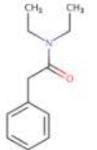
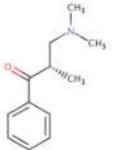
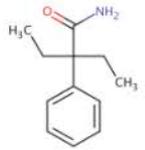
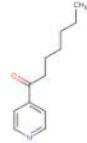
Search All Data

Chemistry Dashboard

Search Results
Searched by Mass: '191.031 +/- 5 ppm'.
298 of 298 chemicals visible

Download / Send Sort by: Mass Difference 

Hide: **Multicomponent Chemicals** 

 <p>DEET 134-62-3</p>	 <p>Phendimetrazine 634-03-7</p>	 <p>N-Butylacetanilide 91-49-6</p>	 <p>Benzaldehyde, 4-(diethylamino)-... 92-14-8</p>	 <p>Acetanilide, 2',6'-diethyl- 16665-89-7</p>	 <p>Azetidine, 1,3-dimethyl-3-(m-met... 19832-26-9</p>
 <p>Benzamide, N-pentyl- 20308-43-4</p>	 <p>p-l-Butylacetanilide 20330-45-4</p>	 <p>N,N-Diethylphenylacetamide 2431-96-1</p>	 <p>3-(Dimethylamino)-2-methylpropi... 26171-50-6</p>	 <p>Butyramide, 2-ethyl-2-phenyl- 30568-39-9</p>	 <p>1-Heptanone, 1-(4-pyridyl)- 32941-30-3</p>

Formula Searches

Molecular Formula Search

C₁₂H₁₇NO

MS Ready Formula 

Exact Formula 

Search 

Exact Formula Search: C₁₂H₁₇NO

298 Chemicals

 United States Environmental Protection Agency

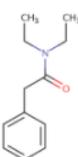
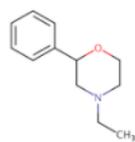
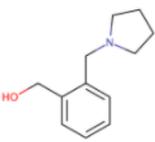
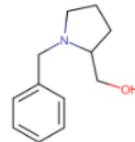
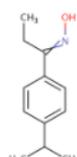
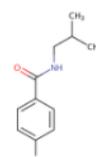
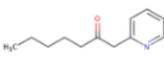
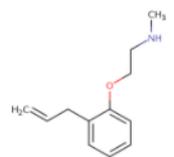
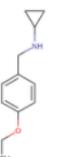
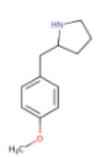
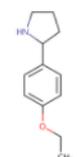
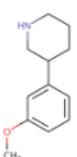
Home Advanced Search Batch Search Lists Predictions Downloads

Search All Data

Chemistry Dashboard

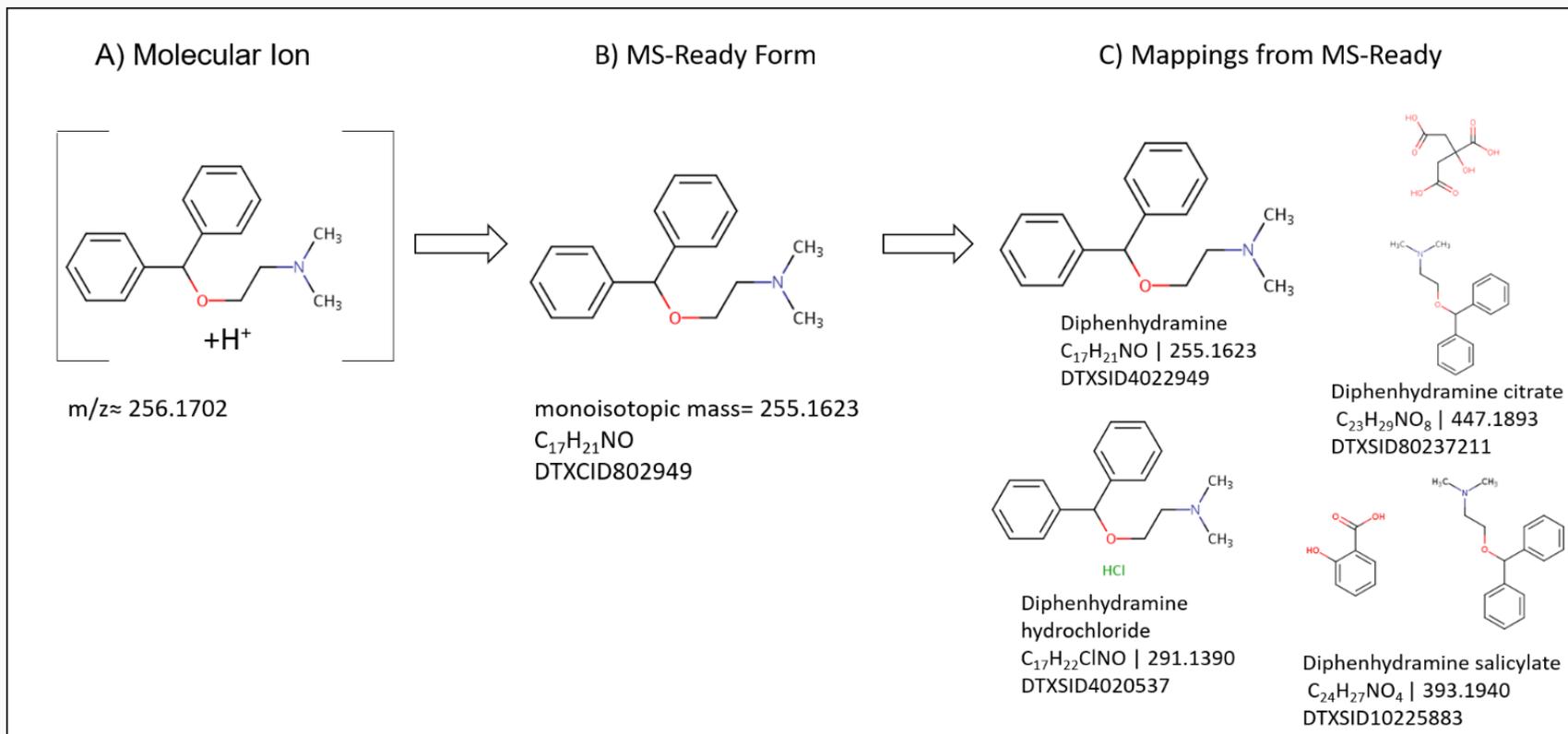
Search Results
Searched by Exact Molecular Formula: 'C₁₂H₁₇NO'.

Download / Send Sort by: DTXSID 298 of 298 chemicals visible Hide: Multicomponent Chemicals Select all

 <p>N,N-Diethylphenylacetamide 2431-96-1</p>	 <p>MLS002639407 83081-04-3</p>	 <p>{2-[(Pyrrolidin-1-yl)methyl]phenyl... 91271-58-8</p>	 <p>(1-benzylpyrrolidin-2-yl)methanol 67131-44-6</p>	 <p>1-Propanone, 1-[4-(1-methylethyl)... 78575-14-1</p>	 <p>4-methyl-N-(2-methylpropyl)benz... 88358-24-1</p>
 <p>2-Heptanone, 1-(2-pyridinyl)- 21211-98-3</p>	 <p>2-(2-ALLYLPHENOXY)-N-METH... 57162-94-4</p>	 <p>N-(4-ethoxybenzyl)cyclopropana... 892571-13-0</p>	 <p>2-[(4-methoxyphenyl)methyl]pyrr... 66162-38-7</p>	 <p>2-(4-ethoxyphenyl)pyrrolidine 383127-28-4</p>	 <p>3-(3-methoxyphenyl)piperidine 79601-21-1</p>

- Structure Identification using the dashboard
 - Formula/mass-based searching – 1 chemical at a time
 - Distilling structures into “MS-Ready form”

Specific Data-Mappings “MS-Ready Structures”



Diphenhydramine

15 Total MS-Ready Mappings

 United States Environmental Protection Agency

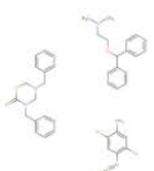
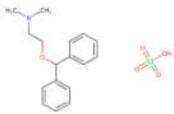
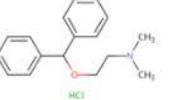
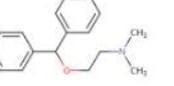
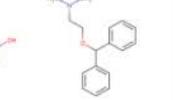
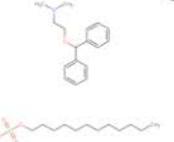
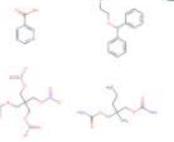
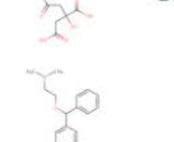
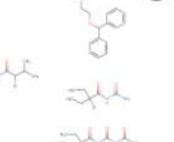
Home Advanced Search Batch Search Lists Predictions Downloads

Search All Data

Chemistry Dashboard

MS-Ready Mappings of Diphenhydramine (Isotopes pre-filtered)

Download / Send Sort by: DTXSID 15 of 15 chemicals visible Hide: Isotopes Select all

 <p>Dampa D 101052-67-9</p>	 <p>Perchloric acid-2-(diphenylmethoxy)ethane 17626-30-1</p>	 <p>Diphenhydramine salicylate 7491-10-3</p>	 <p>Diphenhydramine hydrochloride 147-24-0</p>	 <p>Diphenhydramine 58-73-1</p>	 <p>Acetic acid-2-(diphenylmethoxy)ethane 840517-01-3</p>
 <p>Diphenhydramine laurylsulfate 103659-13-8</p>	 <p>Diphenhydramine hydrochloride 8076-99-1</p>	 <p>Nitric acid-2-(diphenylmethoxy)ethane 17626-28-7</p>	 <p>Visano cor 81246-67-5</p>	 <p>Diphenhydramine citrate [USP] 88637-37-0</p>	 <p>n-(acetylcarbamoyl)-2-bromo-2-ethoxyethane 52232-04-9</p>

“MS Ready” Formula Search C12H17NO 354 Chemicals

 United States Environmental Protection Agency

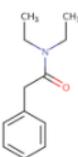
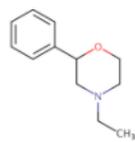
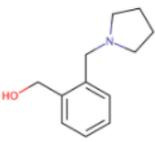
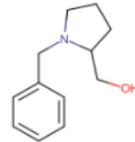
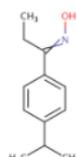
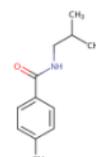
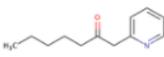
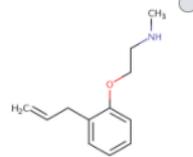
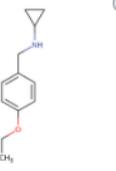
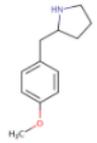
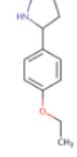
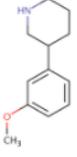
Home Advanced Search Batch Search Lists Predictions Downloads

Search All Data

Chemistry Dashboard

Search Results
Searched by MS Ready Formula: C₁₂H₁₇NO

Download / Send Sort by: DTXSID 354 chemicals Hide: Select all

 <p>N,N-Diethylphenylacetamide 2431-96-1</p>	 <p>MLS002639407 83081-04-3</p>	 <p>{2-[(Pyrrolidin-1-yl)methyl]phenyl}methanol 91271-58-8</p>	 <p>(1-benzylpyrrolidin-2-yl)methanol 67131-44-6</p>	 <p>1-Propanone, 1-[4-(1-methylethyl)phenyl] 78575-14-1</p>	 <p>4-methyl-N-(2-methylpropyl)benzamide 88358-24-1</p>
 <p>2-Heptanone, 1-(2-pyridinyl) 21211-98-3</p>	 <p>2-(2-allylphenoxy)-N-methylmethanamine 57162-94-4</p>	 <p>N-(4-ethoxybenzyl)cyclopropanamine 892571-13-0</p>	 <p>2-[(4-methoxyphenyl)methyl]pyrrolidine 66162-38-7</p>	 <p>2-(4-ethoxyphenyl)pyrrolidine 383127-28-4</p>	 <p>3-(3-methoxyphenyl)piperidine 79601-21-1</p>

- Structure Identification using the dashboard
 - Formula/mass-based searching – 1 chemical at a time
 - Distilling structures into “MS-Ready form”
 - Ranking based on metadata

Identifying Known Unknowns by reference ranking



[Journal of The American Society for Mass Spectrometry](#)
January 2012, Volume 23, [Issue 1](#), pp 179-185 | [Cite as](#)

Table 1

Searching ChemSpider by Elemental Composition then Sorting by Number of Associated References

Class of compounds	Number compounds in class	Position of compound sorted in descending order by number of references					
		#1	#2	#3	#4	#5	>#5
Drugs	45	43	1	1			
Pesticides	8	7	1				
Toxins	2	2					
Polymer antioxidants	15	15					
Polymer UV stabilizers	10	8	1	1			

Data source ranking using the Dashboard

Anal Bioanal Chem
DOI 10.1007/s00216-016-0139-z



RAPID COMMUNICATION

Identifying known unknowns using the US EPA's CompTox Chemistry Dashboard

Andrew D. McEachran¹ · Jon R. Sobus² · Antony J. Williams³

	Mass-based searching		Formula-based searching	
	Dashboard	ChemSpider	Dashboard	ChemSpider
Average rank position	1.3	2.2 ^a	1.2	1.4
Percent in #1 position	85%	70%	88%	80%

^a Average rank in ChemSpider shown here does not include an outlier where the rank was 201, when added the average rank position is 3.5

Additional Metadata Ranking

- US EPA CompTox Chemistry Dashboard Data Sources
- “CPDat” Consumer Product Database
- PubChem Data Source Count
- PubMed Reference Count

Additional Metadata Ranking

C12H17NO: 354 Chemicals

United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Search All Data

Chemistry Dashboard

Search Results

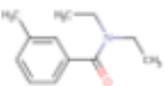
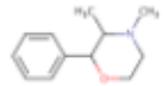
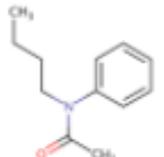
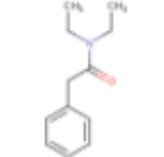
Searched by MS Ready Formula: 'C12H17NO'.

Download / Send Sort by: Sources 354 chemicals Hide: Select all

Structure	DTXSID	Preferred Name	CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Source	PubMed Data Source	Monoisotopic Mass
	DTXSID2021995	DEET	134-62-3	Level 1	111	104	155	753	191.131014
	DTXSID1023447	Phendimetrazine	634-03-7	Level 2	12	27	35	50	191.131014
	DTXSID2042197	N-Butylacetanilide	91-49-6	Level 2	1	26	50	1	191.131014
	DTXSID00179048	N,N-Diethylphenylacetamide	2431-96-1	Level 4	0	18	52	34	191.131014
	DTXSID60865298	N,N-Diethyl-2-methylbenzamide	2728-04-3	Level 1	0	11	49	0	191.131014
	DTXSID4016528	4-Aminoacetophenone	99227-75-2	Level 4	0	11	26	1	191.131014

Additional Metadata Ranking

C12H17NO: 354 Chemicals

Structure	DTXSID	Preferred Name	CPDat Count	Number of Sources	PubChem Data Source	PubMed Data Source
	DTXSID2021995	DEET	111	104	155	753
	DTXSID1023447	Phendimetrazine	12	27	35	50
	DTXSID2042197	N-Butylacetanilide	1	26	50	1
	DTXSID00179048	N,N-Diethylphenylacetamide	0	18	52	34

 United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Search All Data

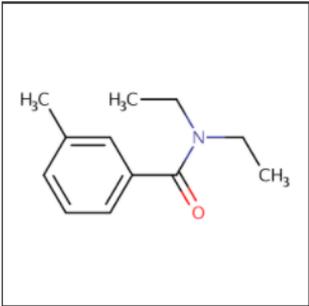
Chemistry Dashboard

Submit Comment Share Copy Aa Aa Aa

DEET

134-62-3 | DTXSID2021995

© Searched by DSSTox_Substance_Id: Found 1 result for 'DTXSID2021995'.



Wikipedia

N,N-Diethyl-m-toluamide, also called DEET () or diethyltoluamide, is the most common active ingredient in insect repellents. It is a slightly yellow oil intended to be applied to the skin or to clothing and provides protection against mosquitoes, ticks, fleas, chiggers, leeches and many biting insects. ...Read more

Intrinsic Properties

Structural Identifiers

Linked Substances

Presence in Lists

Record Information

Quality Control Notes

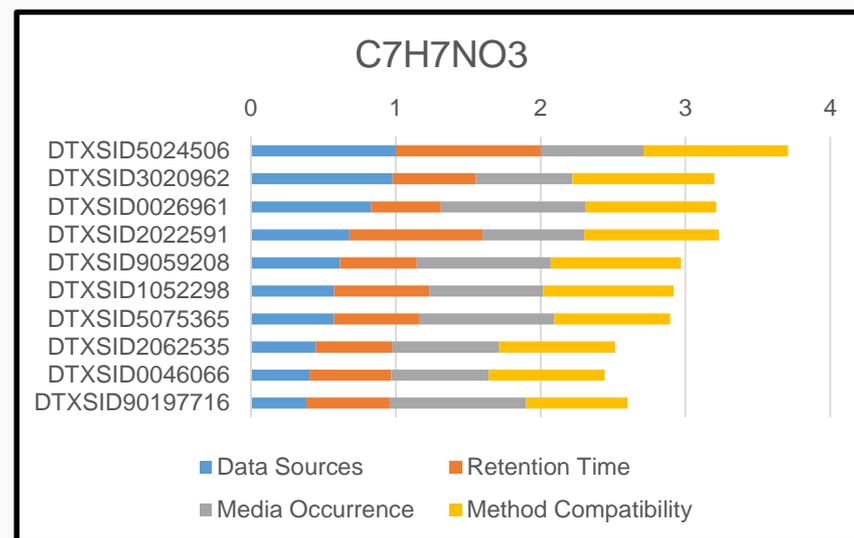
Executive Summary (Beta) Chemical Properties Env. Fate/Transport Hazard ADME (Beta) Exposure Bioassays Similar Compounds Related Substances Synonyms Literature Links Comments

Executive Summary

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

Additional data streams in development

- US EPA CompTox Chemistry Dashboard Data Sources
- “CPDat” Consumer Product Database
- PubChem Data Source Count
- PubMed Reference Count
- Retention Time Prediction $SC_{TOTAL} = SC_{DS} + SC_{PM} + SC_{RT} + SC_{MO} + \dots$
- Predicted Environmental Media Occurrence
- Presence in Lists



“Chemicals Detected in Water”

Chemistry Dashboard

Select List

List Name	Number of Chemicals	List Description
Drinking Water Suspects, KWR Water, Netherlands	136	KWRSJERPS is a list of prioritized suspects relevant for human health in drinking water from KWR Water in Nieuwegein, The Netherlands. The methods are detailed in Sjerps et al 2016, DOI: 10.1016/j.watres.2016.02.034
EPA Consumer Products Suspect Screening Results	1705	This is a compiled list of the suspects reported in the supporting information of Phillips et al 2018, DOI: 10.1021/acs.est.7b04781 - Suspect Screening Analysis of Chemicals in Consumer Products with GCxGC-TOF/MS.
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 2005-2013, as reported in EPA's Hydraulic Fracturing Drinking Water Assessment Final Report (Dec 2016)
STOFF-IDENT Database of Water-Relevant Substances	8885	STOFF-IDENT is a database of water relevant substances collated from various sources within the STOFF-IDENT and FOR-IDENT projects, hosted by LfU, HSWT and TUM. The database at https://www.lfu.bayern.de/stoffident/#!home has additional functional...
Superfund Chemical Data Matrix	220	The Superfund Chemical Data Matrix (SCDM) generates a list of the corresponding Hazard Ranking System (HRS) factor values, benchmarks, and data elements for a particular chemical.
Surfactant List Screened in Swiss Wastewater (2014)	122	EAWAGSURF is a list of surfactants screened in Swiss wastewater effluents as part of a 2014 study. Structures/mixtures are being progressively curated and linked (Schymanski/Williams). Further details in Schymanski et al 2014, DOI: 10.1021/es4044374

- Structure Identification using the dashboard
 - Formula/mass-based searching – 1 chemical at a time
 - Distilling structures into “MS-Ready form”
 - Ranking based on metadata
 - Batch searching of formulae and masses

Batch Search

Batch Search ?



Step One Step Two Step Three Step Four Step Five Step Six

Step One: Select Input

Select Input Type(s)

- Chemical Name i
- CASRN i
- InChIKey i Skeleton i
- DSSTox Substance ID i
- MS-Ready Formula(e) i
- Exact Formula(e) i
- Monoisotopic Mass

×

(Maximum number of identifiers should be limited to <1000 identifiers)

This search is based on what we refer to as 'Mass Spec Ready Formulae'. All chemicals within the database are treated in a manner that all are desalted and stereochemistry is removed as Mass Spectrometry detects the major components of a salt and is insensitive to stereochemistry. As an example, a search for the formula associated with phenol will return phenol, sodium phenolate and calcium phenoxide.

Display All Chemicals

Download Chemical Data

Batch Search

Select Input Type(s)

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

Enter Identifiers to Search (searches should be limited to <1000 identifiers)

C6H12O3
C7H7N3
C8H11NO
C7H5NOS
C9H15NO
C11H12O
C9H8O3
C6H12O5
C9H15NO2

Metadata

- Curation Level Details **i**
- Data Sources **i**
- Assay Hit Count **i**
- Include links to ACToR reports - SLOW! (BETA) **i**
- NHANES/Predicted Exposure **i**
- Include ToxVal Data Availability **i**
- Number of PubMed Articles **i**
- Abstract Sifter Input File (Beta) **i**
- MetFrag Input File(Beta)
- IRIS
- PPRTV
- PubChem Data Sources
- ToxPrint fingerprints **i**

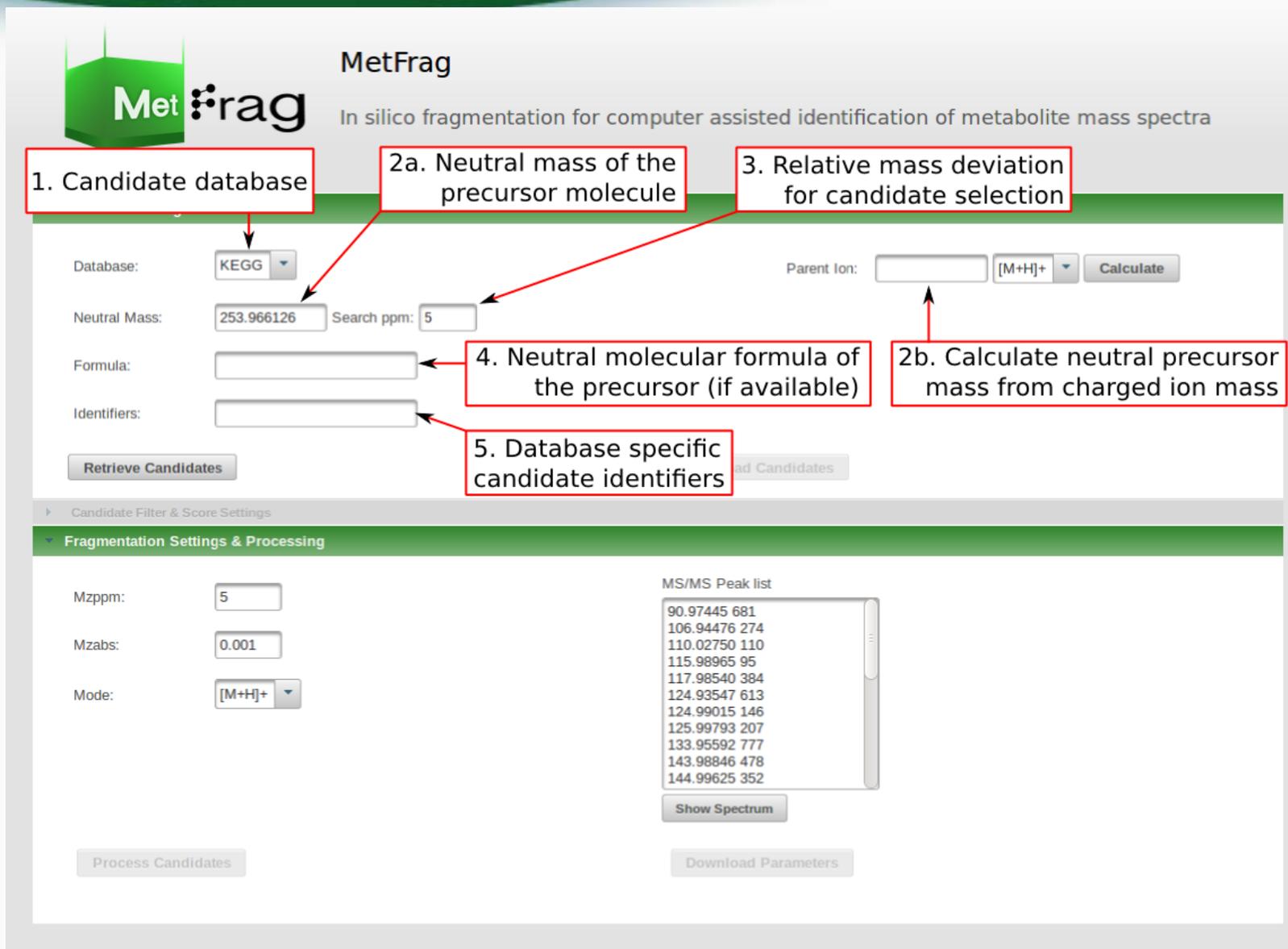
- NIOSH IDLH Values
- NIOSH International Chemical Safety Cards
- NIOSH Pocket Guide to Chemical Hazards
- NIOSH Skin Notation Profiles
- NORMAN Collaborative Trial 2015 Targets and Suspects
- Norman Network PFAS (KEMI Report)
- NORMAN Network Priority List
- NormaNEWS: Norman Early Warning System
- PFAS list provided by X.Trier et al
- Pharmaceutical List with EU, Swiss and US Consumption Data
- Provisional Peer Reviewed Toxicity Values
- Stockholm Convention on Organic Pollutants
- STOFF-IDENT Database of **Water**-Relevant Substances
- Superfund Chemical Data Matrix
- Surfactant List Screened in Swiss Wastewater (2014)

Excel Output

INPUT	FOUND_BY	DTXCID_IN	DATA_SQ	TOXVAL_D	TOXCAST	TOXCAST	NUMBER_C	PUBCHEM	STO
C6H12O3	MS Ready	DTXCID701:	51	Y	0.36	2/562	24	83	Y
C6H12O3	MS Ready	DTXCID003:	67	Y	0.36	1/276	376	80	Y
C6H12O3	MS Ready	DTXCID106:	65	Y	4.42	5/113	6	77	Y
C6H12O3	MS Ready	DTXCID105:	45	Y	0.0	0/163	3	94	-
C6H12O3	MS Ready	DTXCID901:	38	Y	-	-	14	110	Y
C6H12O3	MS Ready	DTXCID402:	34	Y	0.0	0/113	-	53	Y
C6H12O3	MS Ready	DTXCID202:	31	Y	-	-	-	36	Y
C6H12O3	MS Ready	DTXCID202:	30	-	2.54	7/276	-	54	-
C6H12O3	MS Ready	DTXCID109:	26	Y	-	-	-	46	-
C6H12O3	MS Ready	DTXCID202:	24	Y	0.0	0/113	-	47	-
C6H12O3	MS Ready	DTXCID303:	22	Y	-	-	-	89	-
C6H12O3	MS Ready	DTXCID302:	20	Y	-	-	2	25	Y
C6H12O3	MS Ready	DTXCID407:	19	Y	-	-	12	62	-
C6H12O3	MS Ready	DTXCID704:	17	Y	-	-	-	64	-
C6H12O3	MS Ready	DTXCID704:	16	Y	-	-	3	49	-

Batch Search Integration to MetFrag

<http://c-ruttkies.github.io/MetFrag/projects/metfragweb/>



MetFrag
In silico fragmentation for computer assisted identification of metabolite mass spectra

1. Candidate database (points to Database: KEGG)

2a. Neutral mass of the precursor molecule (points to Neutral Mass: 253.966126)

3. Relative mass deviation for candidate selection (points to Search ppm: 5)

4. Neutral molecular formula of the precursor (if available) (points to Formula:)

5. Database specific candidate identifiers (points to Identifiers:)

2b. Calculate neutral precursor mass from charged ion mass (points to Parent Ion: [M+H]⁺ Calculate)

Retrieve Candidates

Fragmentation Settings & Processing

Mzppm:

Mzabs:

Mode:

MS/MS Peak list

90.97445	681
106.94476	274
110.02750	110
115.98965	95
117.98540	384
124.93547	613
124.99015	146
125.99793	207
133.95592	777
143.98846	478
144.99625	352

Show Spectrum

Process Candidates

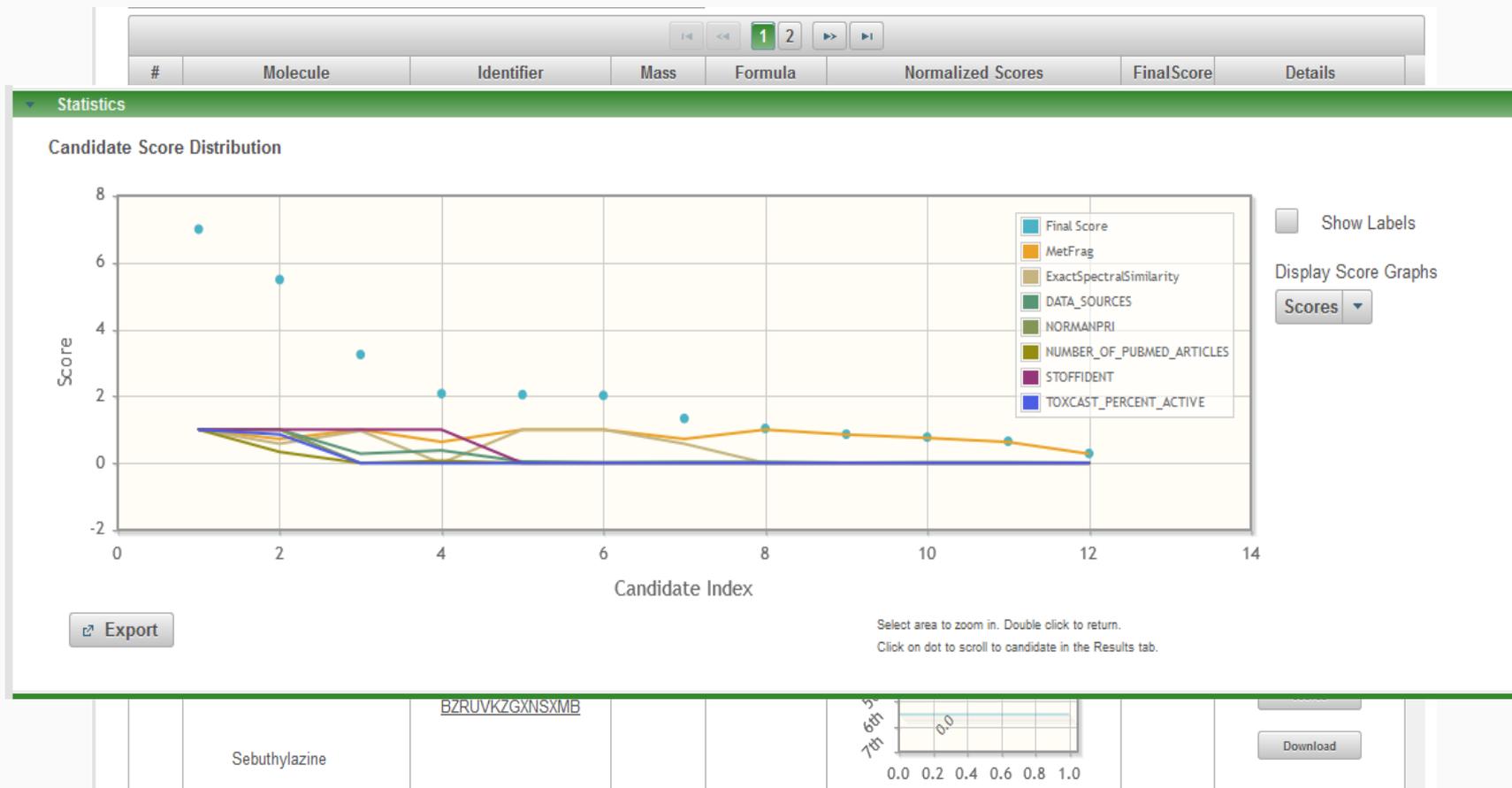
Download Parameters

Metadata

- Curation Level Details
- Data Sources
- Assay Hit Count
- Include links to ACToR reports - SLOW! (BETA)
- NHANES/Predicted Exposure
- Include ToxVal Data Availability
- Number of PubMed Articles
- Abstract Sifter Input File (Beta)
- MetFrag Input File(Beta)
- IRIS
- PPRTV
- PubChem Data Sources

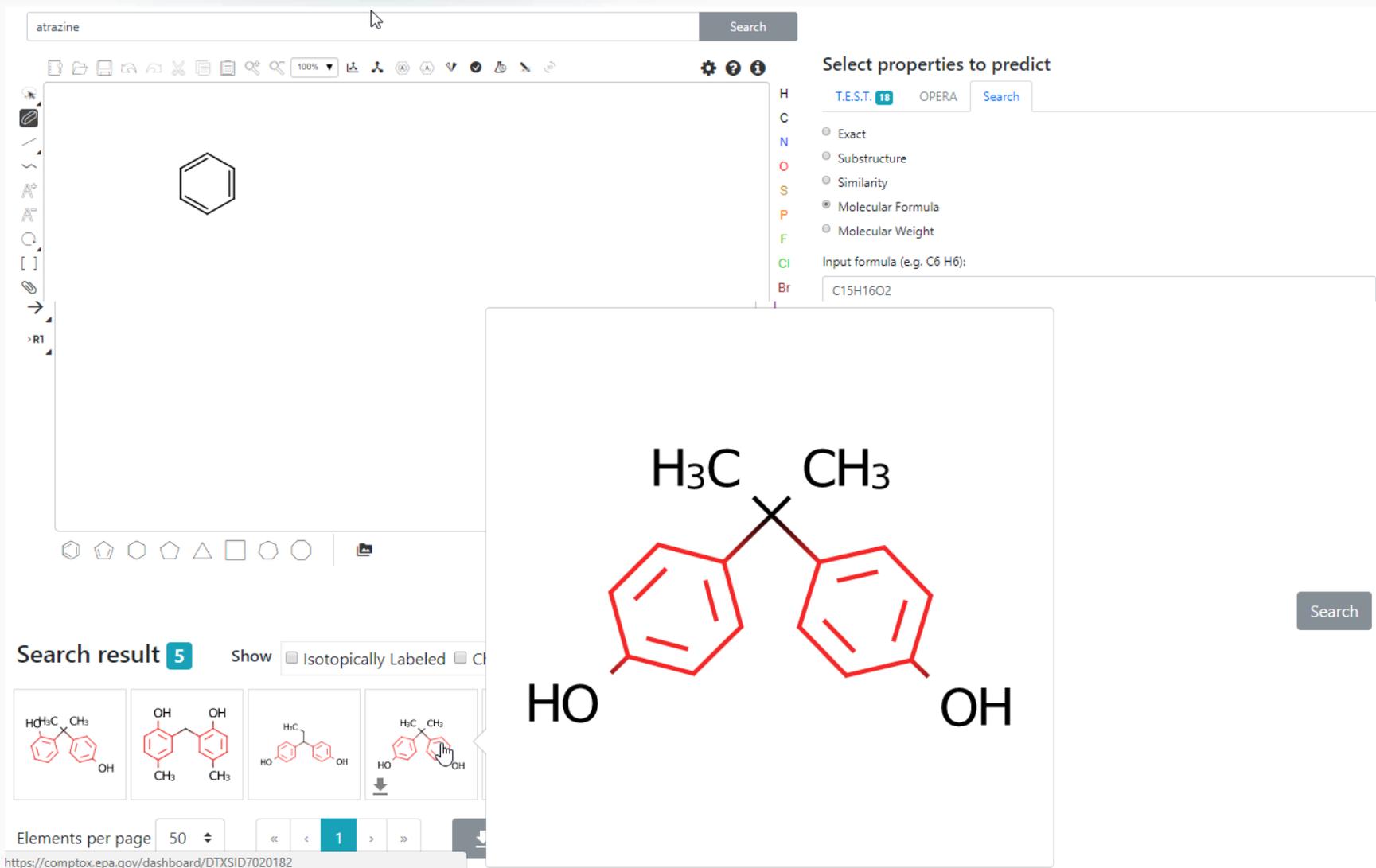
Batch Search Integration to MetFrag

<http://c-ruttikies.github.io/MetFrag/projects/metfragweb/>



Future Work: Combined Substructure/Formula Searching

atrazine Search



Select properties to predict

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

C

N

O

S

P

F

Cl

Br

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

Search

Search result 5 Show Isotopically Labeled Cl

Elements per page 50

1

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

Future Work: Searching Against Predicted Spectra

CFM-ID

★ Utilities ▾ Help Data Publications Contact Us

CFM-ID

Competitive Fragmentation Modeling for Metabolite Identification

Spectra Prediction

Predicts the spectra for a given input molecule. Spectra are computed for low (10V), medium (20V) and high (40V) collision energy levels and are represented by a list of 'mass intensity' pairs, each corresponding to a peak in the spectra.

Parent Compound Structure
InChI or SMILES format

InChI strings need to start with "InChI=" and are not expected to have any charge - an additional H+ will be added. Maximum compound size is 200 atoms. Load an [InChI example](#), [SMILES example](#), or another [SMILES example](#).

Spectra Type

Ion Mode

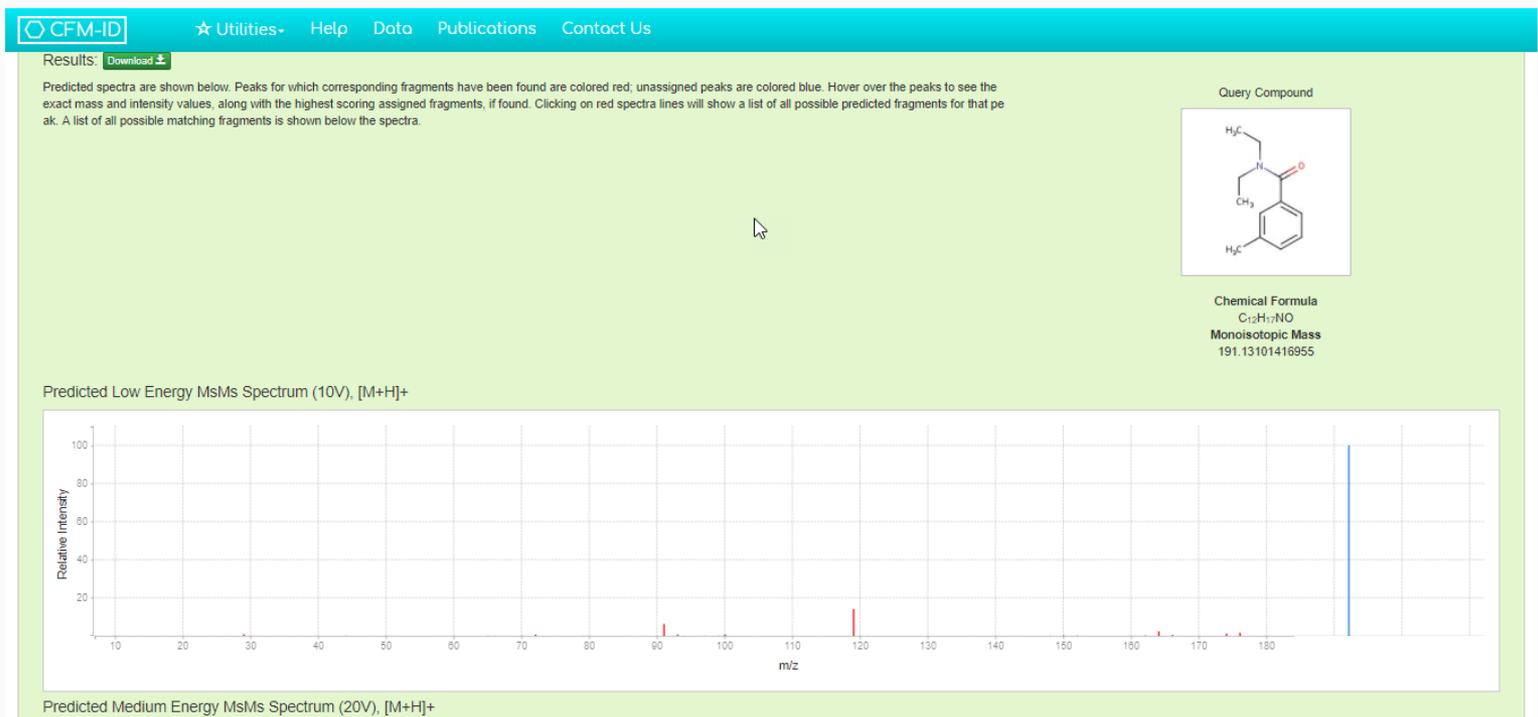
Adduct Type

If you wish to run multiple jobs, input larger query molecules, or customize the computation parameters, you can freely download the source code here: <http://sourceforge.net/projects/cfm-id>.

cfmid.wishartlab.com

Future Work: Searching Against Predicted Spectra

- CFM-ID predicted spectra generated for 700,000 chemicals
 - Positive ion, Negative ion, Electron Impact
 - Three energies



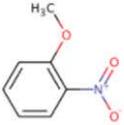
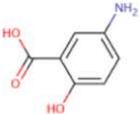
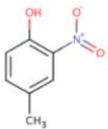
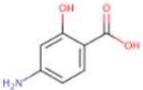
Future Work

Scoring scheme into results

Search Results

Searched by molecular formula: Found 188 results.

Download as: TSV Excel SDF

ID ↓	Structure	Preferred Name ↓	CAS-RN ↓	QC Level ↓	CPCat Count	Number of S...	PubChem D...	Monoisotopic Mass	Identification Score
DTXSID3020962 ToxCast™		1-Methoxy-2-nitrobenzene	91-23-6	Level 2: Expert curate...	0	58	111	153.042593	 3.6
DTXSID5024506 ToxCast™		5-Aminosalicylic acid	89-57-6	Level 2: Expert curate...	2	57	193	153.042593	 3.3
DTXSID0026961 ToxCast™		4-Methyl-2-nitrophenol	119-33-5	Level 2: Expert curate...	0	46	93	153.042593	 3.3
DTXSID2022591		Aminosalicic acid	65-49-6	Level 1: Expert curate...	0	40	177	153.042593	 2.5

$$SC_{TOTAL} = SC_{DS} + SC_{PM} + SC_{RT} + SC_{MO} + \dots$$

- The CompTox Chemistry Dashboard provides access to data for ~760,000 chemicals
- High quality curated data and rich metadata facilitates mass spec analysis
- “MS-Ready” processed data enables structure identification

- The CompTox Chemistry Dashboard team
- NERL colleagues:
 - Jon Sobus, Elin Ulrich, Mark Strynar, Seth Newton (NTA Analysis)
 - Katherine Phillips, Kathie Dionisio, Kristin Isaacs (Consumer Products Database)
- Emma Schymanski – Luxembourg Center for Systems Biomedicine (MS-ready/NTA)

Antony Williams

US EPA Office of Research and Development

National Center for Computational Toxicology (NCCT)

Williams.Antony@epa.gov

ORCID: <https://orcid.org/0000-0002-2668-4821>