

US-EPA Chemicals Dashboard – an integrated data hub supporting exposomics research

Antony Williams

Center for Computational Toxicology and Exposure, US-EPA, 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

NIH Metabolomics Special Interest Group April 20th 2021

Outline



- Quick overview of the dashboard
- Specific data of interest to this audience (it's not just Computational Toxicology)
- Support for Mass Spectrometry
- Data *quality* in the public domain
- Work in progress prototypes

CompTox Chemicals Dashboard

https://comptox.epa.gov/dashboard



	SEPA United States Environmental Protection Home A	Navanced Search Batch Search Lists 🛩 Predictions Downloads	Share 🔻		
	Charles and the second s	875 Thousand Chemicals Froduct/Use Categories Assay/Gene h for chemical by systematic name, synonym, CAS number, DTXSD or InChilley substring search	SEARCH		
Environmental Protection Home Advanced Search Batch Search Li	ists v Predictions Downloads	Share Q. Search al data	I Protection Home Advanced Search Batch Search Lists ♥ Predictions Downloads	Copy 🎽 Share 🎽 Submit Co	mment Q. Search all data
Sep 1 Sep 2		SEARCH Late	Bisphenol A 80-05-7 DTXSID7020182	ТО	X DATA
Step Four: :	Select Data Output Format and Choose Data Fields to Download		Searched by Expert Validated Synonym.	Hazard	
Select Input Type(s)	Enter Identifiers to Search (searches should be limited to <5000 identifiers)	PROPERTIES	DataType Toxicity Value		
S Identifiers	107-22-2 88-30-2 14024-55-6	en added	Toisoty Value	🛉 Human 💋 Eco	
CASEN O	53404-76-5 7784-40-9 1928-40-9 66332-96-5	RPENES HAZARD	La Downlead ▼ Columns >		
DSSTax Compound ID InChilkry Skeleton	2122-70-5 126833-17-8	► ADME	More [©] Type [©] Subtype [©] Risk assessme	ent class [®] Value [®] Units [©] Study type [®] Exposure route [®] Soscies [®] Subsource	+ Source
MS-Ready Formula(e) Exect Formula(e) Moneisotopic Mass	7786-81-4	BOACTIVITY	Miss Short-term Critical Air short-term		ry Exposure Guidelines Table DOD
Abirtadotepic Atala To Download Chemical Data		SIMILAR COMPOUNDS	MEG Short-term Marginal Air short-term		ry Exposure Guidelines Table DOD
Select Output Format:		GENRA (BETA)	MES Short-term Negligible Air short-term MES Soil Negligible Soil chronic		ry Exposure Guidelines Table DOD
Customize Results	Presence in Lists:	RELATED SUBSTANCES	MEG Long-Term, SL/d Negligible Water dhronic		ny Exposure Guidelines Table DOD
Select All Select All in Lists Chemical Identifiers	Pharmacousticals from ZINCLS C 400CR355 Extremely Hazardous Substance List and Threshold Planning Quantities C		ed States rommental Protection Heme Advanced Search Batch Search Läts ❤ Predictions Downloads ncy	Copy 👻 Share 🔻 Submit	Comment Q. Search all clata
	ALSES: Anote Expression Socialistical Exercise Anote Exercise Comparison of Comparison Com		Bisphenol A		CTIVIT
			Searched by Expert Validated Synonym.		
To find articles quickly, enter terms to sift abstracts.		DETAILS	10Y 010)	Chemical Activity Summary 🕕	ASSAY DETAILS
bisphenol estradiol onal bisphenol estradiol j onal Total PMID Year Title	Char Terris Journal	Rev PROPERTIES		ACS0 (k	M): 5.73
		controls - an official journal of the So ENV. FATE/TRANSP schenos - an official journal of the So ENV. FATE/TRANSP	20 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Assay E Assay C Grant F	mbolint Name: OT_ER_ERsERs_0480 escription: 742 mbol: ESR1
2 2 2 0 24409537 2014 Taxinity e Image: Comparison of the state o	valuation of biphenol A administered by gauge to Spregue Davley rats from petatil Delotes, Carancho, Lewis, Vanlandingham, Latendr Toxicological in like disruptive effects of detary exposure to biphenol A or 170 ethingl estadial in C Kondig, Buesing, Christie, Coolman, Gear, Hage International jo	contos: a notical journal of the Bo contos: an official journal of the Bo Umail of thisology I ADME	cell methology cell adhesion maleules 5 - 5 -	Organis	ex human
1 2 0 3 11351429 2001 Extrogen 1 1 1 3 30552454 2016 In vitro 41	is potnery of chemicali deletad in sewage heatment plant efficienti as determined b. Metcalle, Metcalle, Kepanisais, Koenig, Khan, Hugh. Environmental I environ extrapolation (VVVD by PDTX modeling for animal-their noix assessment app. Fabler, Garreet, Birk, Willinst, Hernandez, Hasse, Archives of toxi-	I loocology and chemistry Cology EXPOSURE	eytokine 19 -		ormat Type: cell-based al Process Target: protein stabilization an Technology: Protein-fragment Complementation Direction: positive
2 1 0 3 2007193 2017 Premain III 2 1 0 3 20890130 2017 Effects of IIII 2 1 1 4 27571134 2016 Impact of	Sophieton (perky opposite anis the transcriptione of the indexian la angegeau in al., paramala, sink, Palisau (perkytal bighend A sposure on the values of texnah)-denoptic nuclei of juvenil. A vanibula, Bucke, Cao, Palisau (aux of bas Carl Episoure to Bophend A (PPI) on the Necessari Rat Hypothalinin a		and any H-	Intende Descrip arralyzz	d Target Family: nuclear receptor tion: Data from the assay component OT_ER_ERAERa_0 I into 1 assay endpoint. This assay endpoint. RaERa_0480, was analyzed in the positive fitting cirecti
IFTER	son of Electrochemical Immunoenessan and Aplanemicen for Defection of Small Orga. Proc Shit Reliaberg, Noel, Anquelin Biosensors docrine and behavioral effects of matemal exposure to anal bisphenol A in female mice. Nauki: Floor, Matrix: Parmentier, Hardin Pouzet K. The Journal of a	av SSS Contrastre survey Contr	bedgound measurement DNA beding Surspecter 70 -		to DMSD as the negative control and baseline of activi f binding reporter, measures of receptor for gain-of-si
2 1 4 7 2500035 2013 Sex spec 2 5 0 3 20413007 2010 Neonats 2 5 0 3 20413007 2010 Neonats	Ac impact of pennals biphenel A (BPA) exposure over a range of stuly administer McCatthey, Jones, Maleer, Wess, Stevan, Patisaul Neurotenoology exposure to bisphenel a and espositure and endocrine alterations recentling the Ferdineter, Brougaginet, Luc-Lanton, Libertan Tenvironmental Section 2014 (Section 2014) Tenvironmental Section 2014 Section	It health perspectives TOXCAST/FOX Controls - in official journal of the So	21 de a biodog 0 -	activity they rel	an be used to understand the binding at the pathway to to the gene ESR1. Furthermore, this assay endpoint to as a primary readout, because the performed assay
		PUBCHEM	<u><u> </u></u>	produce relatable	d 1 assay endpoint. To generalize the intended target targets, this assay endpoint is annotated to the "nucl " intended target family, where the subtarrily is "decord."
HAZA/D	Step Three: Run GenRA Prediction	exced Search Batch Search Liats 🛩 1			
Neighbors by: Chem: Morgan Egrprts		effregrent * Coverse Data Matter	C	IMILARITY 📗	
			D7020102		
• EXPOSURE	of warman Pa	80-05-7 DTXSI Searched by Expert Validated S	Synonym.		
EROSURE	ночна. РО. 40.1330m	Searched by Expert Validated S	Synonym. Searched with a similarity threshold of 0.8		_
ERFOSURE	ночна. РО. 40.1330m	Source and by Export Validated 1	Synonym.		-
EDOLUH SIGLACTORY SIGLAC COMPONENT SIGLAC COMPONENT COMPARE TITE SIGLAC COMPONENT COMPARE TITE SIGLAC COMPONENT SIGLAC COMPONENT SIGLAC COMPARE TITE SIGLAC COMPONENT SIGLAC SIGLAC COMPONENT SIGLAC COMPONENT SIGLAC COMPONENT SIGLAC	ночна. РО. 40.1330m	1 1	Synonym. Searched with a similarity threshold of 0.8		-
100504 10050	ночна. РО. 40.1330m	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Synonym. Searched with a similarity threshold of 0.8		-
10/00/[10/01/17/	Annual Annual<		Synonym. Searched with a similarity threshold of 0.8		
10000H 10000H 10000H 10000H 10000H 10000H 10000H 10000H 1000H 100H 10H 10H 10H 10H 10H 10H	Annual		Searched with a similarity threshold of 0.8		
BOODUL CONTROL CO	Annual Annual Bit and annual Bit and annual Bit and annual Bit	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Searched with a similarity threshold of 0.8 THE 428 devices with THE 428 devices with		
100004 100004	Anti-Jahre Anti-Ja	Searched by Expert Validated $(-1)^{-1}$	Shown Searched with a similarity threshold of 0.8 THE VALUE AND ADDRESS OF THE SHORE ADDRESS		
	Anti-Jahre Anti-Ja	$\int \int \int dx = \int d$	Shown Searched with a similarity threshold of 0.8 THE VALUE AND ADDRESS OF THE SHORE ADDRESS		
EXCLUE EXCLUENT EX	Anti-Jahre Anti-Ja	Searched by Expert Validated 3	Shown Searched with a similarity threshold of 0.8 THE VALUE AND ADDRESS OF THE SHORE ADDRESS		2

•	ADME	ℍⅎℂ、ୢ୵ℂℍ₃	bisphenols, with two hydroxyphenyl groups. It is a coloriess solid that is soluble in organic solvents, but poorly soluble in water (0.344 wt % at 83 *C). BPA is a starting material for the synthesis of plastics, primarily certain polycarbonates 	
Þ	EXPOSURE		Read more Intrinsic Properties	•
Þ	BIOACTIVITY		Molecular Formula: C ₁₅ H ₁₆ O ₂ Mol File Q, Find All Chemicals	
	SIMILAR COMPOUNDS	но он	Monoisotopic Mass: 228.11503 g/mol	
	GENRA (BETA)		Structural Identifiers	4
			Linked Substances	•
	RELATED SUBSTANCES		Presence in Lists	•
	SYNONYMS		Record Information	4
Þ	LITERATURE		Quality Control Notes	4
	LINKS			
	COMMENTS			

Wikipedia

Detailed Chemical Pages

OF

Advanced Search Batch Search Lists v Predictions Downloads

80-05-7 | DTXSID7020182

Bisphenol A

Searched by DSSTox Substance Id.

DETAILS

PROPERTIES

HAZARD

ADME

EXECUTIVE SUMMARY

ENV. FATE/TRANSPORT



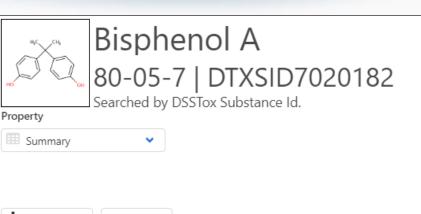
•

Copy 🔻 Share 💌 Submit Comment

Bisphenol A (BPA) is an organic synthetic compound with the chemical formula (CH₃)₂C(C₆H₄OH)₂ belonging to the group of diphenylmethane derivatives and

Q Search all data

Properties, Fate and Transport



Summary

📥 Download 🔻 🛛 🕻 C

Columns 🗡

Property 🗘	Experimental average 🗧 🗘	Predicted average 🗧 🗘	Experimental median 🗘	Predicted median *
LogKow: Octanol-Water	3.32 (1)	3.30		3.39
Melting Point	155 (7)	140	156	144
Boiling Point	200 (1)	360		355
Water Solubility	8.55e-4 (3)	8.78e-4	5.26e-4	7.56e-4
Vapor Pressure	-	6.83e-7		1.51e-7
Flash Point	-	190		190
Surface Tension	-	46.0		
Index of Refraction	-	1.60		
Molar Refractivity	-	68.2		
The standard have been shall be a set	and the second	and the second designed		the second se

wironmental Protection

Agency

Hazard



						Haz	zard				
ataType											
Point	of Departure	•									
						🛉 Humar	Eco				
L Down	and 💌 🔽	lumns v	10 👻								
Downi		iumns 🗸									
More 🗘	Priority 🗘	Type 🗘	Subtype 🗘	Risk assessment class 🗘	Value ^	Units 🗘	Study type 🗘	Exposure route 🗘	Species 🗘	Subsource 🗘	Source
More 🕈	Priority +	Type 🕈	Subtype 🗘	Risk assessment class \$ developmental neurotoxicity	Value ^ 0.015	Units 🕈 mg/kg-day	Study type 🗘	Exposure route 🗢	Species 🗢	Subsource 🕈 EPA ORD	Source ToxRefDB
		NOAEL	Subtype 🕈 - -			onics	5	•	opecies		
	4	NOAEL	-	developmental neurotoxicity	0.015	mg/kg-day	developmental neurotoxicity	oral	rat	EPA ORD	ToxRefDB
	4	NOAEL NEL	-	developmental neurotoxicity	0.015	mg/kg-day mg/kg-day	developmental neurotoxicity developmental neurotoxicity	oral oral	rat	EPA ORD	ToxRefDB ToxRefDB
	4 4 4	NOAEL NEL LOAEL LEL	· · ·	developmental neurotoxicity developmental neurotoxicity developmental neurotoxicity	0.015 0.015 0.015	mg/kg-day mg/kg-day mg/kg-day	developmental neurotoxicity developmental neurotoxicity developmental neurotoxicity	oral oral	rat rat	EPA ORD EPA ORD EPA ORD	ToxRefDB ToxRefDB ToxRefDB
	4 4 4 4 4	NOAEL NEL LOAEL	- -	developmental neurotoxicity developmental neurotoxicity developmental neurotoxicity developmental neurotoxicity	0.015 0.015 0.015 0.015	mg/kg-day mg/kg-day mg/kg-day mg/kg-day	developmental neurotoxicity developmental neurotoxicity developmental neurotoxicity developmental neurotoxicity	oral oral oral oral	rat rat rat rat	EPA ORD EPA ORD EPA ORD EPA ORD	ToxRefDB ToxRefDB ToxRefDB ToxRefDB

Sources of Exposure to Chemicals



	n Home Advanced Search Batch Search Lists 🗸 Pre	dicuons Downloads	Copy • Share •	Submit Comment Q Search all data
	Bisphenol A 80-05-7 DTXSID702018 Searched by DSSTox Substance Id.	2		
DETAILS		Product and Use	Categories (PUCs) 🚺	
EXECUTIVE SUMMARY	La Download 🔻			
PROPERTIES	Columns ~ 10 ~			Search query
ENV. FATE/TRANSPORT	Product or Use Categorization	Categorization type	◆ <u>Number o</u>	of Unique Products
HAZARD	manufacturing, metals	CPCat Cassette	17	
ADME	adhesive	CPCat Cassette	17	
EXPOSURE		CPCat Cassette	16	
EXPOSORE		CPCat Cassette	12	
PRODUCT 8	USE CATEGORIES	CPCat Cassette	11	
FRODUCT	USE CATEGORIES	CPCat Cassette	8	
		CPCat Cassette	8	
CHEMICAL V	VEIGHT FRACTION	CPCat Cassette	8	
		CPCat Cassette	7	
CHEMICAL F	FUNCTIONAL USE	CPCat Cassette	6	
TOXICS REL	EASE INVENTORY	First << < 1 2 3 4	5 6 7 8 9 10 > >>	Last
MONITORIN	G DATA			
EXPOSURE	PREDICTIONS			
	DN VOLUME			6

Identifiers to Support Searches

25 💙

📥 Download 🔻



_	 	

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

EXPOSURE

BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

LITERATURE LINKS

COMMENTS

Bisphenol A 80-05-7 | DTXSID7020182

Synonyms

Search query

Synonym	Quality
Bisphenol A	Valid
4,4'-(Propane-2,2-diyl)diphenol	Valid
Phenol, 4,4'-(1-methylethylidene)bis-	Valid
80-05-7 Active CAS-RN	Valid
BPA	Valid
4,4'-Propane-2,2-diyldiphenol	Valid
Phenol, 4,4'-(1-methylethylidene)bis-	Valid
4-06-00-06717 Belistein Registry Number	Beilstein
(4,4°-Dihydroxydiphenyl)dimethylmethane	Good
2,2-Bis(4'-hydroxyphenyl) propane	Good
2,2'-Bis(4-hydroxyphenyl)propane	Good
2.2-BIS-(4-HYDROXY-PHENYL)-PROPANE	Good
2,2-Bis(4-hydroxyphenyl)propane	Good
2,2-Bis(p-hydroxyphenyl)propane	Good
2,2-Di(4-Hydroxyphenyl) Propane	Good

Link Access



Bisphenol A 80-05-7 | DTXSID7020182

	Searched by Approved	Name.			
DETAILS	General	Toxicology	Publications	Analytical	Prediction
EXECUTIVE SUMMARY	EPA Substance Registry Service	▲ ACToR	Toxline	C FOR-IDENT	3 2D NMR HSQC/HMBC Prediction
PROPERTIES	🖑 Household Products Database	여 [,] DrugPortal	Environmental Health Perspectives	🌢 NEMI: National Environmental Methods Index	Carbon-13 NMR Prediction
PROPERTIES	Chemical Entities of Biological Interest	CCRIS	NIEHS	RSC Analytical Abstracts	Proton NMR Prediction
ENV. FATE/TRANSPORT	(ChEBI)	ChemView	National Toxicology Program	Tox21 Analytical Data	ChemRTP Predictor
HAZARD	PubChem	О СТР	G Google Books	MONA: MassBank North America	LSERD
HAZARD	Chemspider	SechemPortal	Google Scholar	and mzCloud	
► ADME	CPCat	Gene-Tox	G Google Patents	MST NIST IR Spectrum	
	🥔 DrugBank	HSDB	PPRTVWEB	NIST MS Spectrum	
EXPOSURE	hmp HMDB	ToxCast Dashboard 2	PubMed		
▶ BIOACTIVITY	W Wikipedia	TactMed	IRIS Assessments		
	Q MSDS Lookup	International Toxicity Estimates for Risk	🖲 EPA HERO		
SIMILAR COMPOUNDS	ChEMBL	ATSDR Toxic Substances Portal	NIOSH Skin Notation Profiles		
GENRA (BETA)	Q Chemical Vendors	Superfund Chemical Data matrix	🚾 NIOSH Pocket Guide		
	CalEPA Office of Environmental Health Hazard Assessment	🚾 NIOSH IDLH Values	RSC Publications		
RELATED SUBSTANCES	NIOSH Chemical Safety Cards	ACToR PDF Report			
SYNONYMS	ToxPlanet	Toxics Release Inventory	n Springer Materials		
	ACS Reagent Chemicals	CREST	Federal Register		
LITERATURE	W Wikidata	National Air Toxics Assessment	Regulations.gov		
LINKS	ChemHat: Hazards and Alternatives Toolbox		Bielefeld Academic Search Engine		
	Wolfram Alpha		🖆 CORE Literature Search		
COMMENTS	 ScrubChem 				
	ECHA Brief Profile				

Mass Spec Links

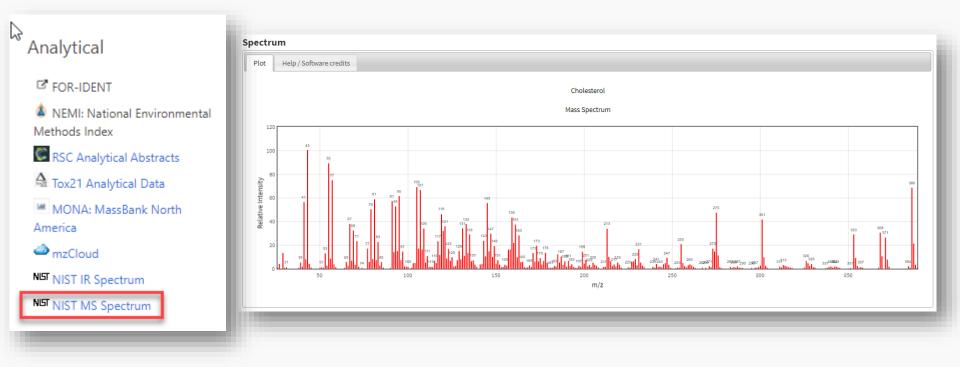


Analytical

RSC Analytical Abstracts A Tox21 Analytical Data MONA: MassBank North America anzCloud NIST IR Spectrum NIST MS Spectrum 🚚 MassBank A NEMI: National Environmental Methods Index NIST Antoine Constants IR Spectra on PubChem NIST Kovats Index values

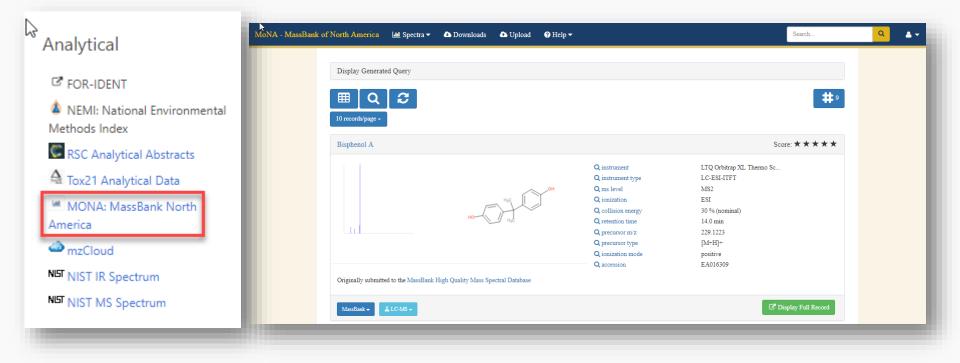
NIST WebBook https://webbook.nist.gov/chemistry/





MassBank of North America https://mona.fiehnlab.ucdavis.edu







Batch Searching

Aggregate data for a list of chemicals





Trends in Environmental Analytical Chemistry

Volume 20, October 2018, e00059





Opioid occurrence in environmental water samples—A review

Marina Celia Campos-Mañas ª, Imma Ferrer ^b ペ ⋈, E.Michael Thurman ^b, Ana Agüera ª

Show more

https://doi.org/10.1016/j.teac.2018.e00059

Get rights and content

Batch Search Names



Buprenorphine Codeine Dextromethorphan Dihydrocodeine Dihydromorphine Ethylmorphine Fentanyl Heroin Hydrocodone Hydromorphone Ketamine Meperidine Methadone Morphine Morphinone Naloxone Naltriben 0xycodone Oxymorphone Propoxyphene Sufentanil Tramadol

Step 1	Step 2	Step 3	Step 4	Step 5	Step
	Ste	ep Five: Choos	se Data Fields to Do	wnload	•
Please enter one i	identifier per line				×
Select Input Type(s) Identifiers Chemical Name CASRN CASRN DinChIKey DSSTox Substan DSSTox Compour InChIKey Skeletor MS-Ready Formul Exact Formula(e) (e () nce ID () nd ID () n () ila(e) ()		Enter Identifiers to Search Buprenorphine Codeine Dextromethorphan Dihydrocodeine Dihydromorphine Ethylmorphine Fentanyl Heroin Hydrocodone Hydrocodone	h (searches should be limited to	o <5000 identifiers)
 Monoisotopic Ma Display All C 		d Chemical Data			
		d Chemical Data			
		d Chemical Data	INPUT Buprenorphine	FOUND_BY	DTX SID
		d Chemical Data	Buprenorphine	Approved Name	DTXSID2022705
		d Chemical Data	Buprenorphine Codeine	Approved Name Approved Name	DTXSID2022705 DTXSID2020341
		d Chemical Data	Buprenorphine Codeine Dextromethorphan	Approved Name Approved Name Approved Name	DTXSID2022705 DTXSID2020341 DTXSID3022908
		d Chemical Data	Buprenorphine Codeine Dextromethorphan Dihydrocodeine	Approved Name Approved Name Approved Name Approved Name	DTXSID2022705 DTXSID2020341 DTXSID3022908 DTXSID5022936
		d Chemical Data	Buprenorphine Codeine Dextromethorphan Dihydrocodeine Dihydromorphine	Approved Name Approved Name Approved Name Approved Name Approved Name	DTXSID2022705 DTXSID2020341 DTXSID3022908 DTXSID5022936 DTXSID7048908
		d Chemical Data	Buprenorphine Codeine Dextromethorphan Dihydrocodeine Dihydromorphine Ethylmorphine	Approved Name Approved Name Approved Name Approved Name Approved Name Approved Name	DTXSID2022705 DTXSID2020341 DTXSID3022908 DTXSID5022936
		d Chemical Data	Buprenorphine Codeine Dextromethorphan Dihydrocodeine Dihydromorphine	Approved Name Approved Name Approved Name Approved Name Approved Name Approved Name Approved Name	DTXSID2022705 DTXSID2020341 DTXSID3022908 DTXSID5022936 DTXSID7048908 DTXSID1046760
	Chemicals Download		Buprenorphine Codeine Dextromethorphan Dihydrocodeine Dihydromorphine Ethylmorphine Fentanyl Heroin	Approved Name Approved Name Approved Name Approved Name Approved Name Approved Name Approved Name Synonym	DTXSID2022705 DTXSID2020341 DTXSID3022908 DTXSID5022936 DTXSID7048908 DTXSID1046760 DTXSID9023049
			Buprenorphine Codeine Dextromethorphan Dihydrocodeine Dihydromorphine Ethylmorphine Fentanyl	Approved Name Approved Name Approved Name Approved Name Approved Name Approved Name Approved Name	DTXSID2022705 DTXSID2020341 DTXSID3022908 DTXSID5022936 DTXSID7048908 DTXSID1046760 DTXSID9023049 DTXSID6046761
	Chemicals Download	cel	Buprenorphine Codeine Dextromethorphan Dihydrocodeine Dihydromorphine Ethylmorphine Fentanyl Heroin Hydrocodone	Approved Name Approved Name Approved Name Approved Name Approved Name Approved Name Approved Name Synonym Approved Name	DTXSID2022705 DTXSID2020341 DTXSID3022908 DTXSID5022936 DTXSID7048908 DTXSID1046760 DTXSID9023049 DTXSID6046761 DTXSID6046761 DTXSID8023131
	Chemicals Download	cel	Buprenorphine Codeine Dextromethorphan Dihydrocodeine Dihydromorphine Ethylmorphine Fentanyl Heroin Hydrocodone Hydrocodone	Approved Name Approved Name Approved Name Approved Name Approved Name Approved Name Synonym Approved Name Approved Name	DTXSID2022705 DTXSID2020341 DTXSID3022908 DTXSID5022936 DTXSID7048908 DTXSID1046760 DTXSID9023049 DTXSID6046761 DTXSID6046761 DTXSID8023131 DTXSID8023133
	Chemicals Download	cel	Buprenorphine Codeine Dextromethorphan Dihydrocodeine Dihydromorphine Ethylmorphine Fentanyl Heroin Hydrocodone Hydrocodone Hydromorphone Ketamine	Approved Name Approved Name Approved Name Approved Name Approved Name Approved Name Synonym Approved Name Approved Name Approved Name	DTXSID2022705 DTXSID2020341 DTXSID3022908 DTXSID5022936 DTXSID7048908 DTXSID1046760 DTXSID9023049 DTXSID6046761 DTXSID8023131 DTXSID8023133 DTXSID8023187

Add Other Data of Interest

OPERA Model Predictions ①



Chemical Identifiers						
🗹 DTXSID 🚯	-			I		
Chemical Name	INPUT	DTXSID	CASRN	MOLECULAR_F	MONOISOTOPIC	MS_READY_SMI
	Buprenorph	DTXSID202	52485-79-7	C29H41NO4	467.3035588	[H]C12CC3=C4C
DTXCID 🚯	Codeine	DTXSID202	76-57-3	C18H21NO3	299.1521435	[H]C12CC3=C4C
CAS-RN		DTXSID302		C18H25NO		[H]C12CC3=C(C=
CAS-KN	Dihydrocod	DTXSID502	125-28-0	C18H23NO3	301.1677936	[H]C12CC3=C4C
🗹 InChIKey 🚯	- ·	DTXSID704		C17H21NO3	287.1521435	[H]C12CC3=C4C
IUPAC Name 1	Ethylmorph	DTXSID104		C19H23NO3		[H]C12CC3=C4C
	Fentanyl	DTXSID902		C22H28N2O		CCC(=O)N(C1CC
Structures	Heroin	DTXSID604		C21H23NO5		[H]C12CC3=C4C
🗆 Mol File 🚯		DTXSID802		C18H21NO3		[H]C12CC3=C4C
_		DTXSID802		C17H19NO3		[H]C12CC3=C4C
SMILES 🚺	Ketamine	DTXSID802		C13H16CINO		CNC1(CCCCC1=
InChI String 1		DTXSID902		C15H21NO2		CCOC(=O)C1(CC
		DTXSID702		C21H27NO		CCC(=0)C(CC(C)
🗹 MS-Ready SMILES 🚯	Morphine	DTXSID902		C17H19NO3		[H]C12CC3=C4C
QSAR-Ready SMILES (1)		DTXSID501		C17H17NO3		[H]C12CC3=C4C(
	Naloxone	DTXSID802	465-65-6	C19H21NO4	327.1470582	[H]C12CC3=C4C
Intrinsic And Predicted Properties	Naltriben	-	-	-	-	•
🗹 Molecular Formula 🚯		DTXSID502		C18H21NO4		[H]C12CC3=C4C
		DTXSID502		C17H19NO4		[H]C12CC3=C4C
🗆 Average Mass 🚯		DTXSID102		C22H29NO2		CCC(=0)OC(CC1
Monoisotopic Mass 🚯			56030-54-7	C22H30N2O2S		CCC(=O)N(C1=C
	Tramadol	DTXSID908	27203-92-5	C16H25NO2	263.188529	COC1=CC=CC(=
🔲 TEST Model Predictions 🚯						



Chemical Lists of Interest...

290 Chemical Lists (and growing)

📩 Download 🔻

Columns ~



 Home
 Advanced Search
 Batch Search
 Lists
 Predictions
 Downloads

 Lists of Chemicals
 List of Assays
 List of Assays

Copy Filtered Lists URL

mass

List Acronym 🗘	List Name 🗘	Last Updated 🗘	Number of Chemicals 🕈	List Description +
HDXEXCH	MASSPECDB: Hydrogen Deuterium Exchange Standard Set - Under HDX Conditions	2018-11-07	592	Observed species (deuterated and undeuterated) from the HDXNOEX list under hydrogen deuterium exchange conditions (Ruttkies, Schymanski et al. in prep.)
HDXNOEX	MASSPECDB: Hydrogen Deuterium Exchange Standard Set - No Exchange	2018-11-07	765	Environmental standard set used to investigate hydrogen deuterium exchange in small molecule high resolution mass spectrometry (Ruttkies, Schymanski et al. in prep.)
MASSBANKEUSP	MASSPECDB: MassBank.EU Collection: Special Cases	2017-07-16	263	The MassBank.EU list contains curated chemicals (Schymanski/Williams) associated with the literature/tentative/unknown/SI spectra available on MassBank.EU that are not available as part of the full MassBank collection of reference standard spectra.
MASSBANKREF	MASSPECDB: MassBank Reference Spectra Collection	2017-07-13	1267	This MassBank list contains chemicals associated with the full MassBank collection of reference standard spectra available on MassBank.EU, MassBank.JP and MassBank of North America as well as the Open Data collection, curated by Williams/Schymanski.
MYCOTOXINS	MASSPECDB: Mycotoxins from MassBank.EU	2017-08-02	88	This is a set of mycotoxins, initiated by the contribution of spectra of 90 mycotoxins to MassBank.EU by Justin Renaud and colleagues from Agriculture and Agri-Food Canada, Government of Canada

"Volatilome" Human Breath



-

LIST: VOLATILOME: Human Breath

Q Search VOLATILOME Chemical

Identifier substring search

List Details

Description: This list is a subset of compounds detected in human breath and reported in the peer-reviewed literature and identified in experimental work at US-EPA. The bulk of the collection is extracted from the article "The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva" by de Lacy Costello et al in J. Breath Res. 8 (2014) 034001 (DDI:10.1088/1752-7155/8/3/034001), from the article "On-line analysis of exhaled breath", by Bruderer et al in Chemical Reviews (DDI:10.1021/acs.chemrev.9b00005) as well as an increasing number of chemicals identified in our own laboratory studies.

elect all 🛃 Download 🔻 Send to Batch Search	Default 👻 û CASRN X DTXSID X 👽	5 chemicals Hide chem	nicals that are: 👻 Filter by Name or CASRN 🔳
CH ₃		H ₂ C ₂	H ₂ C _N
H ₂ N Acetamide CASRN:60-35-5 DTXSID:DTXSID7020005	Acetonitrile CASRN:75-05-8 DTXSID:DTXSID7020009	Acrolein CASRN:107-02-8 DTXSID:DTXSID5020023	Acrylonitrile CASRN:107-13-1 DTXSID:DTXSID5020029

"Volatilome" Saliva



	LIST: VOLATILO	ME: Saliva	
Q Search VOL	TILSALIVA Chemicals		
Identifier subs	ring search		
Details			•
secription. This list is a subset of compounds detected in s	aliva and reported in the peer-reviewed literature and identified i	n experimental work at US-EPA. The collection is extracted	from the article "The human volatilome; volatile
	ns, urine, feces and saliva" by de Lacy Costello et al in J. Breath Re		nom the article. The numan volationie, volatile
answer of enerthedis, 507			
		-1-1	
ect all 📩 Download 🔻 Send to Batch Search	307 che Name ♥ 및 CASRN X DTXSID X ♥		cals that are: 👻 Filter by Name or CASRN 🔳
			0
н.с., сн.		H ₃ C CH ₃	H ₃ C CH ₃
		H ₃ C CH ₃	H ₃ C CH ₃
H ₃ C	H ₃ C-CH ₃	H ₃ C CH ₃	H ₃ C CH ₃
H ₃ C	H ₃ C-CH ₃	H ₃ C CH ₃	H ₃ C H ₃ C H ₃ C CH ₂
H ₃ C H ₃ C H ₃ C Santolina triene	H ₃ C-CH ₃	p-Menth-2-ene	p-lsopropyl-alpha-methylstyrene
	P-Xylene CASRN:106-42-3 DTXSID:DTXSID2021868	CH ₃	H ₃ C CH ₂

Disinfection By-Products



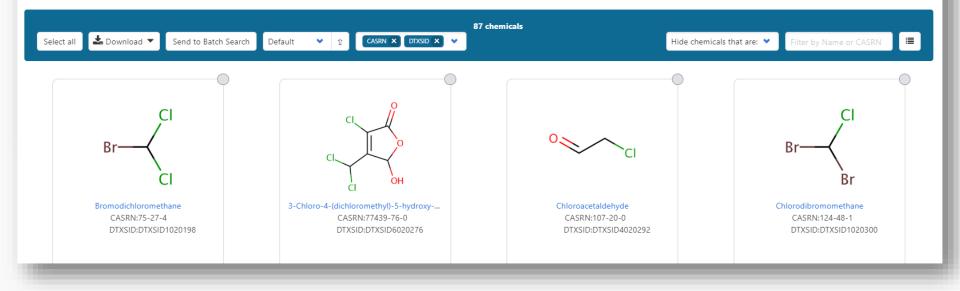
LIST: Disinfection By-Products

Identifier substring search

List Details

Description: Disinfection by-products (DBPs) result from chemical reactions between organic and inorganic matter in water with chemical treatment agents during the water disinfection process. DBPs are present in most drinking water supplies that have been subject to chlorination, chloramination, ozonation, or treatment with chlorine dioxide.

Number of Chemicals: 87



Tire Crumb Rubber (298)



Related Topics:	Safer	Chemica	ls Research
------------------------	-------	---------	-------------

CONTACT US SHARE

July 2019 Report: Tire Crumb Rubber Characterization

Key Takeaways:

- EPA is releasing a new report that addresses exposure (that is, chemicals and how people come in contact with these) to tire crumb rubber on synthetic turf fields. This report is not a risk assessment, nor can the information be used to identify a level above which health effects could occu. Tire Crumb Rubber
- In general, the findings findings findings findings findings

• Only Part 1 is being relea assessment.

List Details

- Part 1 of this report pres
- The scope of this study v

Q	Search T		1B Cher	micals	
🔲 Id	entifier s	ubstring	search		

Description: This chemical list is based on data contained within the <u>Federal Research Action Plan (FRAP) on Recycled Tire Crumb Used on Playing Fields and Playgrounds</u>. The chemical list is obtained from the <u>Toxicity reference information</u> <u>spreadsheet</u> compiled for the potential tire rrumb rubber chemical constituents identified in the State-of-Science Literature Review/Gaps Analysis. White Paper Summary of Results. Eleven sources of publicly available toxicity reference information were searched. It is important to recognize that not all potential chemical constituents identified through the literature search were confirmed through measurements made under the Federal Research Action Plan. Number of Chemicals: 298

Select all 📩 Download 🔻 Send to Batch Search Default 🌱 🖞 CASRN 🗙 DDXS	298 chemicals	Hide chemicals that are: Filter by Name or CASRN	:
N CH ₃ Acetonitrile CASRN:75-05-8 DTXSID:DTXSID7020009 Accolein CASRN:107-02-8 DTXSID:DTXSID5020023	H ₂ N	Azobenzene CASRN:103-33-3 DTXSI:DTXSID67X5108020123	•

Terpenes in Vape (37)



	LIST: Terper	nes in vape	
	earch VAPETERPENES Chemicals		
🗆 Ider	tifier substring search		
Details			
escription: Terpenes are organic compounds for	ind in the marijuana plant that give strains their distinct aromatic and f	flavor profiles. They are now being isolated and concentrated into	oils for individual vaping.
umber of Chemicals: 37			one for mannadar raping.
	37	7 chemicals	
lect all 🔹 Download 🔻 🛛 Send to Batch :	Search Default 💙 🕆 CASRN 🗙 DTXSID 🗙 💙	Hide chemica	Is that are: 👻 🛛 Filter by Name or CASRN 🛛 🔳
Ν			
			H ₃ C, CH ₃
H ₂ C ₂			
H ₃ C	CH ₃ CH ₃ O	H,C.	Hac
H ₃ C	HI, CHI, CHI, CHI,		H ₃ C
H ₃ C O	H, C C C C C C C C C C C C C C C C C C C		H ₃ C
0	H,C		H ₃ C H ₂ C
H ₃ C CH ₃	H,C	H ₁ C CH ₅ CH ₅ H ₁ C CH ₅ CH ₅	H ₃ C H ₂ C
H ₃ C CH ₃	Geranyl acetate	Nerolidal	H ₃ C H ₂ C
0	Geranyl acetate CASRN:105-87-3	Nerolidol CASRN:7212-44-4	beta-Caryophyllene CASRN:87-44-5
H ₃ C CH ₃ 1.8-Cineol			
H ₃ C CH ₃ 1.8-Cineol CASRN:470-82-6	CASRN:105-87-3	CASRN:7212-44-4	CASRN:87-44-5

Hydraulic Fracturing (1640)



Contact Us

EPA's Study of Hydraulic Fracturing and Its Potential Impact on Drinking Water Resources

Hydraulic Fracturing Study Home

Final Assessment

EPA Published Research

Fact Sheets

Questions & Answers about the final assessment

Multi-agency collaboration on unconventional oil and gas research

EPA Hydraulic Fracturing -Agency Main Page

Hydraulic Fracturing For Oil And Gas: Impacts From The Hydraulic

WATER|EPA; Chemicals associated with hydraulic fracturing

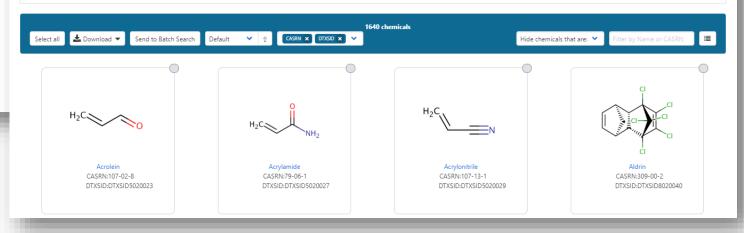
Q Search EPAHFR Chemicals

Identifier substring search

List Details

Description: Chemicals used in hydraulic fracturing fluids and/or identified in produced water from 2005-2013, corresponding to chemicals listed in Appendix H of EPA's Hydraulic Fracking Drinking Water Assessment Final Report (Dec 2016). Citation: U.S. EPA, Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States (Final Report). U.S. Environmental Protection Agency, Washington, D.C. EPA/600/R-16/236F, 2016. https://www.epa.gov/hfstudy

*Note that Appendix H chemical listings in Tables H-2 and H-4 were mapped to current DSSTox content, which has undergone additional curation since the publication of the original EPA HF Report (Dec 2016). In the few cases where a Chemical Name and CASRN from the original report map to distinct substances (as of Jan 2018), both were included in the current EPAHFR chemical listing for completeness; additionally, 34 previously unmapped chemicals in Table H-5 are now registered in DSSTox (all but 2 assigned CASRN) and, thus, have been added to the current EPAHFR listing. Number of Chemicals: 1640



Opioids and Metabolites (160)



	DRUGS: Opioids and r	elated metabolites	
Q Search Ol	IOIDS Chemicals		
Identifier su	bstring search		
List Details			•
	assembled primarily from public resources (e.g. Wikipedia, datab	ases and literature articles) and is under ongoing curation and	d expansion.
Number of Chemicals: 180			
Select all 🛃 Download 🔻 Send to Batch Search	Default ♥ û CASRN X DTXSID X ♥	hemicals Hide chemica	als that are: 👻 Filter by Name or CASRN 🔳
	0		
	jen,		H,C
	" J- AL		- Contraction of the second se
N-CH3	Č) ^{či}	""/ICH3	
HOM	Alfentanil		Anileridine
Codeine CASRN:76-57-3 DTXSID:DTXSID2020341	CASRN:71195-58-9 DTXSID:DTXSID9022570	Alphaprodine CASRN:77-20-3 DTXSID:DTXSID4022575	CASRN:144-14-9 DTXSID:DTXSID8022610
D173D:D173D2020341	017310:0173104022310	01/310:01/3104022373	D1X3ID:D1X3ID6022010
H ₂ C	но		
		но	
\searrow	HC O UNIT	\sim	HO CH3

Mycotoxins



-

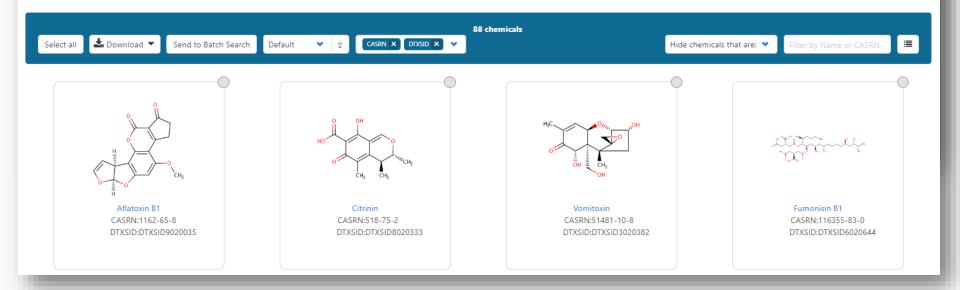
MASSPECDB: Mycotoxins from MassBank.EU

Q Search MYCOTOXINS Chemicals

Identifier substring search

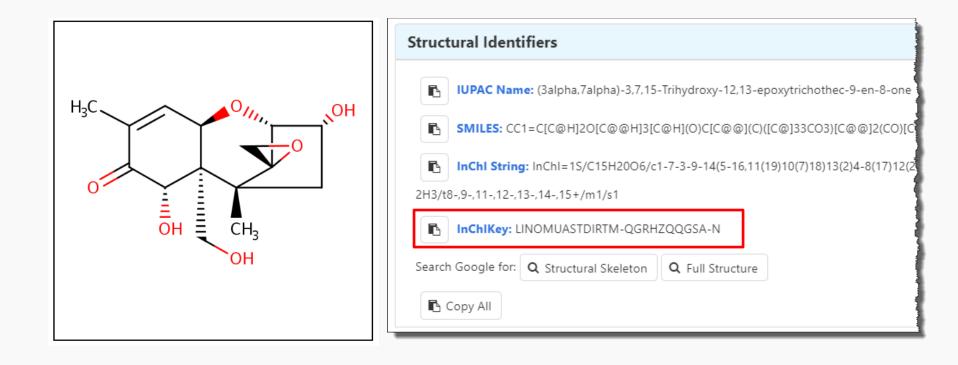
List Details

Description: This is a set of mycotoxins, initiated by the contribution of spectra of 90 mycotoxins to <u>MassBank.EU</u> by Justin Renaud and colleagues from Agriculture and Agri-Food Canada, Government of Canada. This list is also a part of the <u>MASSBANKREE</u> list and the <u>NORMAN Suspect Exchange</u> and will be expanded as new contributions arrive. Number of Chemicals: 88



Vomitoxin

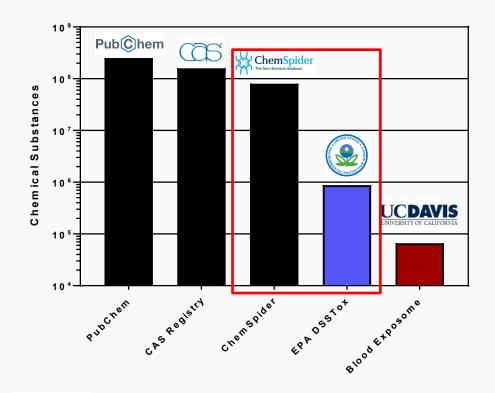




BIG databases are GREAT!



- Thanks to all of the public database efforts
- So much benefit from what's been done
- There are hundreds of them at this point...

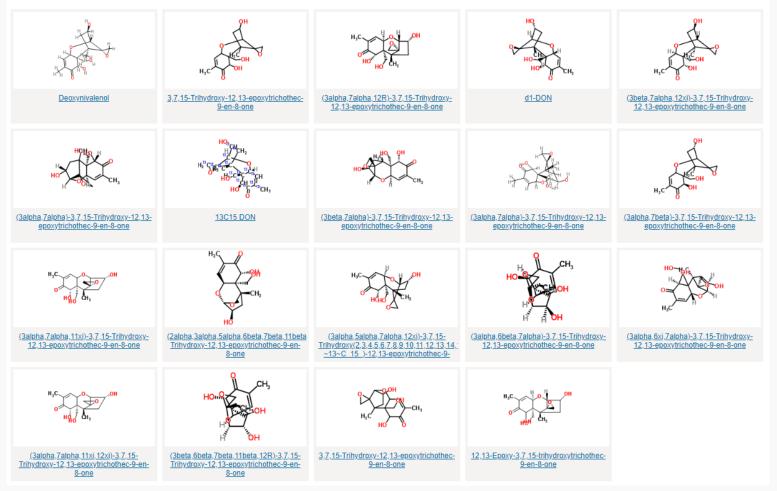


Vomitoxin - ChemSpider



19 "Vomitoxins" – 3 isotopically labeled

Search term: LINOMUASTDIRTM (Found by InChlKey (skeleton match))



Vomitoxin – PubChem



• 33 unique InChl Keys

Compound (33)	ds S	Substances (10)				
Searching chemica pages is not searc			ng IUPAC names and Ir	nChIKey	ys across the compour	nd collection. Note tha
33 results	- Filters		SORT BY	₽ Re	elevance	~
Not Vomitoxir	(3.Alpha Trichoth Methan Compour MF: C ₁₅ H ₂ InChIKey: IUPAC Na oxatricycle	a.,7.Alpha.)-; 1 necen-8-One; no-1-Benzoxep nd CID: 6432495 2006 MW: 296.31g LINOMUASTDIRTM me: (3R,10S)-3,10-	c-9-En-8-One, 12, 2,13-Epoxy-3.Alp LINOMUASTDIR in-10,2'-Oxirane] //mol M-LMJBVPRVSA-N ·dihydroxy-2-(hydroxy -5-ene-12,2'-oxirane]-	ha.,7. <mark>FM</mark> -LI , Tricl	Alpha.,15-Trihyd MJBVPRVSA-N; S hothec-9-En-8-O	roxy-9- Spiro[2,5- ne Deriv.

ChemSpider – lots of virtuals???



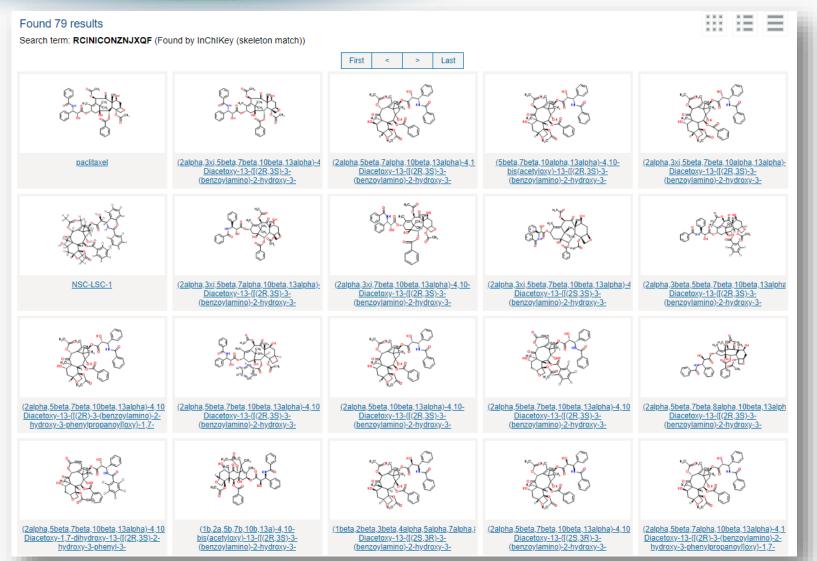


52 million chemicals from one vendor

Data Sources			
Data Source	<u>Count</u>	<u>Date</u> <u>Created</u>	<u>Last</u> <u>Updated</u>
Aurora Fine Chemicals	<u>51885566</u>	13/04/2009	09/01/2020
Chemspace	<u>14283313</u>	30/11/2016	04/12/2018
AKos	12326374	15/04/2008	09/10/2017
Mcule	<u>9299739</u>	21/01/2014	26/10/2018
Molport	<u>8200357</u>	09/02/2010	09/01/2020
<u>Enamine</u>	<u>3056649</u>	15/04/2008	15/10/2019

Taxol: 79 Results





Data Quality is important



Data quality in free web-based databases!





"MS-ready" structures

McEachran et al. J Cheminform (2018) 10:45 https://doi.org/10.1186/s13321-018-0299-2

Journal of Cheminformatics

METHODOLOGY



"MS-Ready" structures for non-targeted high-resolution mass spectrometry screening studies

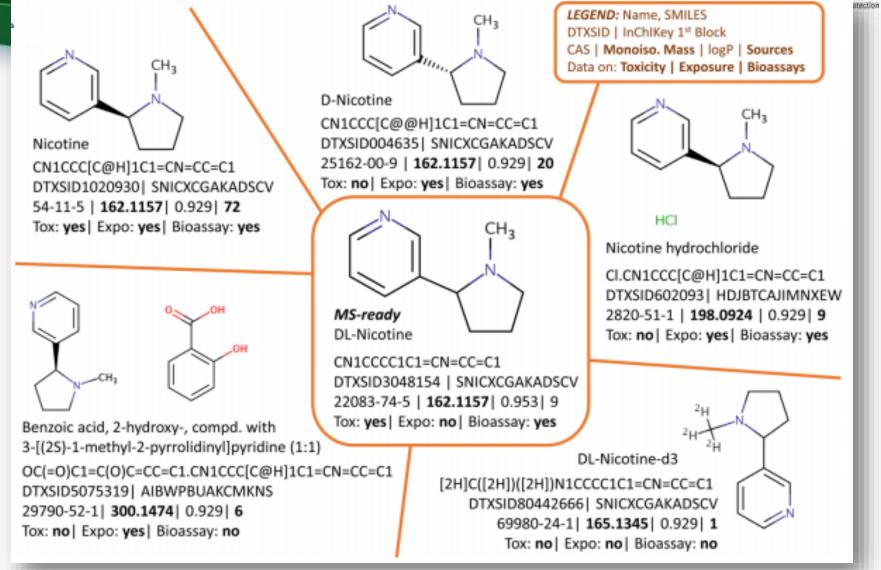
Andrew D. McEachran^{1,2*}, Kamel Mansouri^{1,2,3}, Chris Grulke², Emma L. Schymanski⁴, Christoph Ruttkies⁵ and Antony J. Williams^{2*}

Overview of MS-Ready Structures

SEPA

- All structure-based chemical substances are algorithmically processed to
 - Split multicomponent chemicals into individual structures
 - Desalt and neutralize individual structures
 - Remove stereochemical bonds from all chemicals
- MS-Ready structures are then mapped to original substances to provide a path between chemicals detected by mass spectrometry to original substances







Open Science for Identifying "Known Unknown" Chemicals Emma L. Schymanski*[†][®] and Antony J. Williams^{*†}[®]

Viewpoint

MS-Ready Mappings from Details Page



•

Perfluorooctanesulfonic acid 1763-23-1 | DTXSID3031864

Searched by Synonym from Valid Source.

F F	F F F F	F F F F	F F F F	F OH F O	=0

Wikipedia

Perfluorooctanesulfonic acid (conjugate base perfluorooctanesulfonate) (PFOS) is an anthropogenic fluorosurfactant and global pollutant. PFOS was the key ingredient in Scotchgard, a fabric protector made by 3M, and numerous stain repellents. It was added to Annex B of the Stockholm Convention on Persistent Organic Pollutants in May 2009. PFOS can be synthesized in industrial production or result from the degradation of precursors. PFOS levels that have been detected in wildlife

Read more

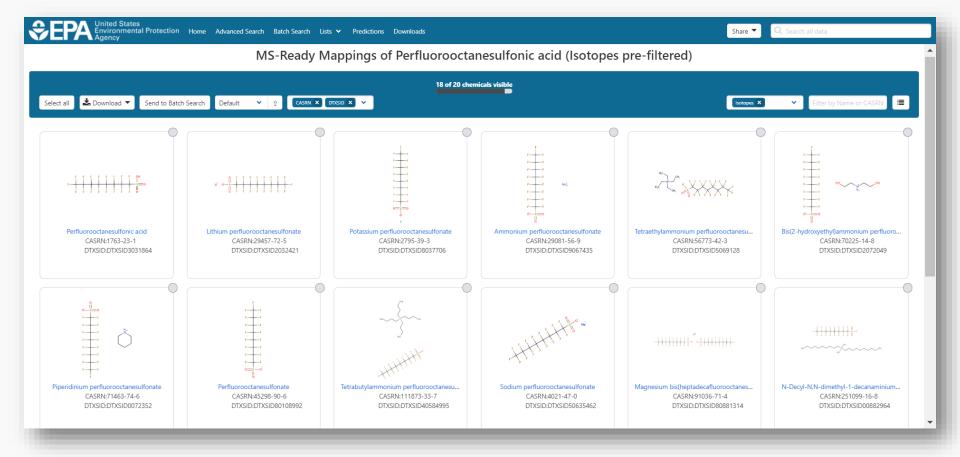
Quality Control Notes	•
ntrinsic Properties	•
Structural Identifiers	•
inked Substances	•
Same Connectivity: 4 records (based on first layer of InChI) Mixtures, Components and Neutralized Forms: 9 records (based on QSAR ready mappings and with the compound as a component of a mixture)	
MS-Ready Mappings: DTXCID1011864: 18 records;	

Similar Compounds: 83 records (based on Tanimoto coefficient >0.8)

ed_search/index

MS-Ready Mappings Set of 20 substances for "PFOS"



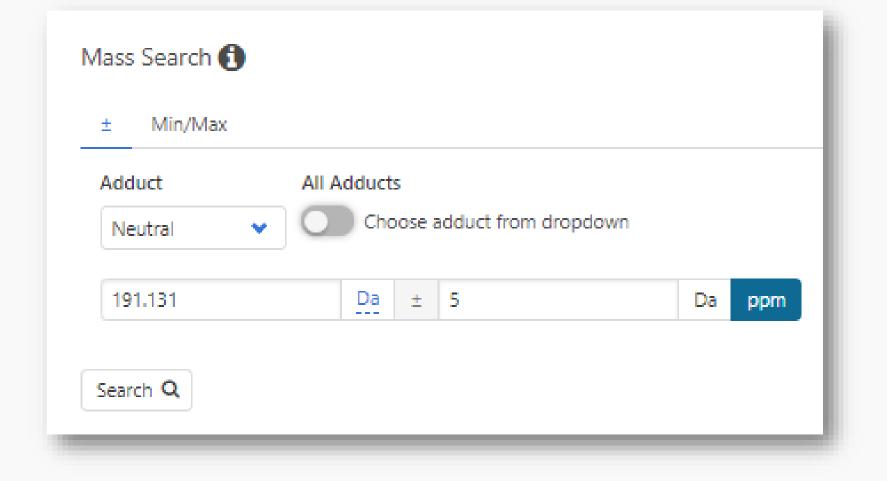




Mass and Formula Searching

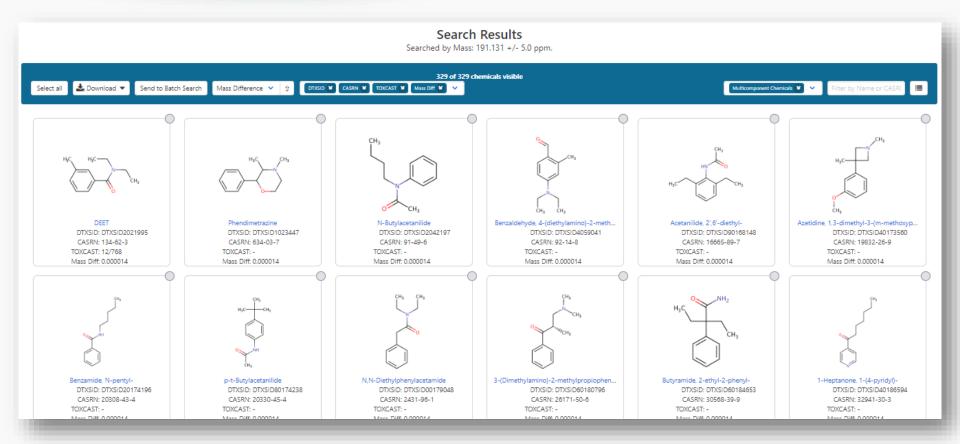
Advanced Searches Mass Search





Advanced Searches Mass Search





MS-Ready Structures for **Formula Search**



Molecular Formula Search 🚯

💿 MS Ready Formula 🚯 🔿 Exact Formula 🚯

Formula

Please use the format of the following example: C6H8O2 or C6H(8-10)O(0-2)

Search Q

MS-Ready Mappings



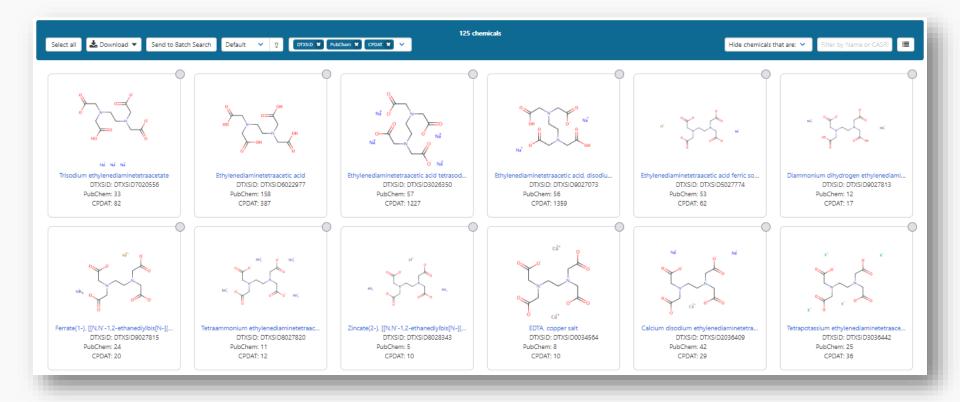
• EXACT Formula: C10H16N2O8: 3 Hits

	O MS R Formula C10H16	eady Formula 🚯 🧿 Exact Form	ula 🕦
,	Select all 🛃 Download 🔻 Send to B	atch Search Default 👻 🕆 DTXSID 🗙	3 of 3 chemi
	$ \begin{array}{c} & \stackrel{\mathbf{O}}{\underset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}{\overset{\mathbf{H}}{\overset{\mathbf{O}}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}}{\overset{\mathbf{O}}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}{\overset{\mathbf{O}}}{\overset{\mathbf{O}}}{\overset{\mathbf{O}}}{\overset{\mathbf{O}}{\overset{\mathcal{O}}{\overset{\mathcal{O}}{\overset{\mathcal{O}}{\overset{\mathcal{O}}{\overset{\mathcal{O}}{\overset{\mathcal{O}}{\overset{\mathcal{O}}{\overset{\mathcal{O}}{\mathcal{O$	$ \begin{array}{c} & \stackrel{\circ}{\rightarrow} \stackrel{\circ}$	$ \begin{array}{c} & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ & & \\ $

MS-Ready Mappings



- Same Input Formula: C10H16N2O8
- MS Ready Formula Search: 125 Chemicals







- 125 chemicals returned in total
 - 8 of the 125 are **single component** chemicals
 - 3 of the 8 are isotope-labeled
 - 3 are neutral compounds and 2 are charged
- Multiple components, stereo, isotopes and charge all collapsed and mapped through MS-Ready



Batch Searching mass and formula

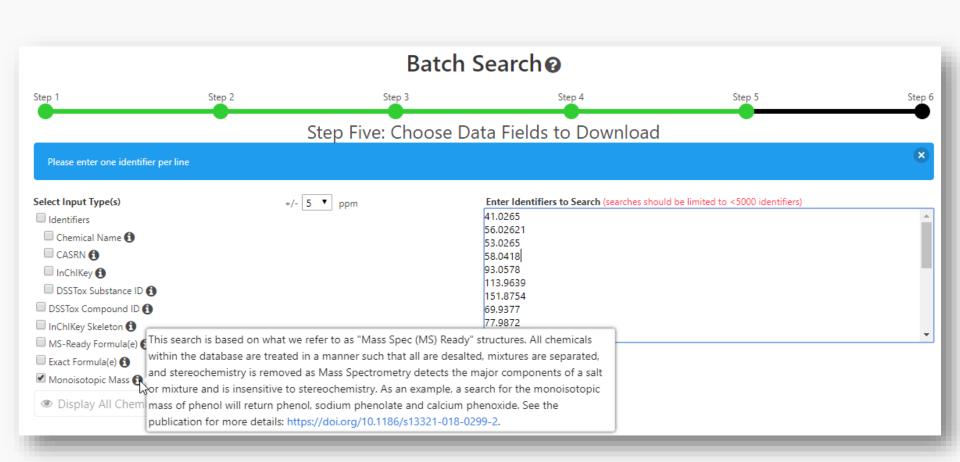
Batch Searching



• Singleton searches are useful but we work with **thousands** of masses and formulae!

- Typical questions
 - What is the list of chemicals for the formula $C_x H_y O_z$
 - What is the list of chemicals for a mass +/- error
 - Can I get chemical lists in Excel files? In SDF files?
 - Can I include properties in the download file?

Batch Searching Formula/Mass



inmental Protection

Searching batches using MS-Ready Formula (or mass) searching



	A	В	С	D	E	F	G
1	INPUT	DTXSID	CASRN	PREFERRED NAME	MOL FORMULA	MONOISOTOPIC MASS	DATA SOURCES
2	C14H22N2O3	DTXSID2022628	29122-68-7	Atenolol	C14H22N2O3		46
3	C14H22N2O3	DTXSID0021179	6673-35-4	Practolol			32
4	C14H22N2O3	DTXSID4048854	841-73-6	Bucolome	C14H22N2O3		20
5	C14H22N2O3	DTXSID1045407	13171-25-0	Trimetazidine dihydrochloride	C14H24Cl2N2O3		19
6	C14H22N2O3	DTXSID0045753	56715-13-0	R-(+)-Atenolol	C14H22N2O3		19
7	C14H22N2O3	DTXSID2048531	5011-34-7	Trimetazidine			14
8	C14H22N2O3	DTXSID10239405	93379-54-5	Esatenolol			12
9	C14H22N2O3	DTXSID50200634	52662-27-8	N-(2-Diethylaminoethyl)-2-(4-hydroxyphenoxy)acetamide		266.163042576	7
10	C14H22N2O3	DTXSID4020111	51706-40-2	dl-Atenolol hydrochloride	C14H23CIN2O3	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		35
13	C18H34N2O6S	DTXSID7047803	859-18-7	Lincomycin hydrochloride	C18H35CIN2O6S	442.1904357	22
14	C18H34N2O6S	DTXSID20849438	1398534-62-7	PUBCHEM_71432748	C18H35CIN2O6S	442.1904357	1
15	C10H12N2O	DTXSID1047576	486-56-6	Cotinine	C10H12N2O		40
16	C10H12N2O	DTXSID8075330					22
17	C10H12N2O	DTXSID8044412					18
18	C10H12N2O	DTXSID80165186	153-98-0	Serotonin hydrochloride			11
19	C10H12N2O	DTXSID2048870	29493-77-4	(4R,5S)-4-methyl-5-phenyl-4,5-dihydro-1,3-oxazol-2-amine			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		176.094963014	7
22	C10H12N2O	DTXSID40178777				176.094963014	7
	C10H12N2O	DTXSID80157026				176.094963014	6
24	C10H12N2O	DTXSID30205607	570-14-9			176.094963014	6
25	C14H18N4O3				C14H18N4O3		68
26	C14H18N4O3						51
27	C14H18N4O3	DTXSID40209671				326.1145682	8
28	C14H18N4O3	DTXSID70204210		Benzenemethanol, 4-((2,4-diamino-5-pyrimidinyl)methyl)-2,		290.137890456	5
29	C14H18N4O3	DTXSID20152671		6-Methoxy-4-(3-(N,N-dimethylamino)propylamino)-5,8-quina		290.137890456	4
30	C14H18N4O3	DTXSID30213742		1H-1,2,4-Benzotriazepine-3-carboxylic acid, 4,5-dihydro-4-		290.137890456	3
	C14H18N4O3	DTXSID30219608		2,4-Pyrimidinediamine, 5-((3,4,5-trimethoxyphenyl)methyl)-		308.14845514	3
	C14H18N4O3	DTXSID20241155		L-Aspartic acid, compound with 5-((3,4,5-trimethoxyphenyl		423.175398165	3
	C14H18N4O3	DTXSID80241156		L-Glutamic acid, compound with 5-((3,4,5-trimethoxypheny		437.191048229	3
	C14H18N4O3	DTXSID20143781		1H-Pyrido(2,3-e)-1,4-diazepine-2,3,5-trione, 4-(2-(diethylam		290.137890456	3
35	C12H11N7	DTXSID6021373					52
		DTXSID00204465				253.107593382	7
37	C12H11N7	DTXSID5064621				251.091943318	4
38	C12H11N7	DTXSID00848025				351.074973101	1
39	C12H11N7	DTXSID50575293				253.107593382	1
40	C8H9NO2	DTXSID2020006			C8H9NO2		75
11	C8H0NIO2	DTYSID6025567	13/1 201 3	Mothyl 2 aminohonzoato		161 063328634	50

Batch Search in specific lists

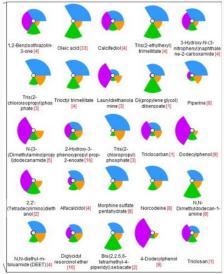


	INPUT	DTXSID	MASSBANKREF	NEMILIST	WRTMSD	NORMANPRI	SUSDAT
- T	Buprenorph	DTXSID202	-	-	Y	-	Υ
	Codeine	DTXSID202	Υ	Υ	Y	Υ	Y
	Dextrometh			Υ	Y	-	Υ
	Dihydrocod	DTXSID502	Υ	-	Y	Y	Υ
<u>N</u>	Dihydromor	DTXSID704	-	-	-	-	Υ
	Ethylmorph	DTXSID104	-	-	Y	-	Υ
	Fentanyl	DTXSID902	Υ	-	Y	-	Υ
💽 🚺	Heroin	DTXSID604	Υ	-	Y	Y	Υ
	Hydrocodor	DTXSID802	Υ	Υ	Y	Y	Υ
	Hydromorpł	DTXSID802	-	-	Y	-	Y
. 🗆 <u>N</u>	Ketamine	DTXSID802	Υ	-	Y	-	Υ
💌 N-	Meperidine	DTXSID902	Υ	-	Y	-	Υ
	Methadone	DTXSID702	Υ	Υ	Y	-	Υ
🗹 📐	Morphine	DTXSID902	Υ	Υ	Y	Y	Υ
	Morphinone	DTXSID501	-	-	-	-	Υ
	Naloxone	DTXSID802	-	-	Y	-	Y
	Naltriben	-	-	-	-	-	-
	Oxycodone	DTXSID502	Y	Υ	Y	Y	Y
	Oxymorpho	DTXSID502	-	-	Y	-	Υ
	Propoxyphe	DTXSID102	Y	Y	Y	-	Y
	Sufentanil	DTXSID602	-	-	Y	-	Y
	Tramadol	DTXSID908	Y	Υ	Y	Y	Y

Benefits of bringing it all together

- The true dashboard benefit is integration
- Rank potential candidates for toxicity using available data – hazard, exposure, *in vitro*





vironmental Protection



Candidate ranking using metadata



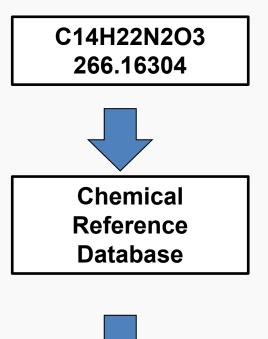
C American Society for Mass Spectrometry, 2011

J. Am. Soc. Mass Spectrom. (2012) 23:179-185 DOI: 10.1007/s13361-011-0265-y

RESEARCH ARTICLE

Identification of "Known Unknowns" Utilizing Accurate Mass Data and ChemSpider

- A mass and/or formula search is for an *unknown* chemical but it is a *known* chemical contained within a reference database
- Most likely candidate chemicals have the most associated data sources, most associated literature articles or both



Sorted candidate

structures



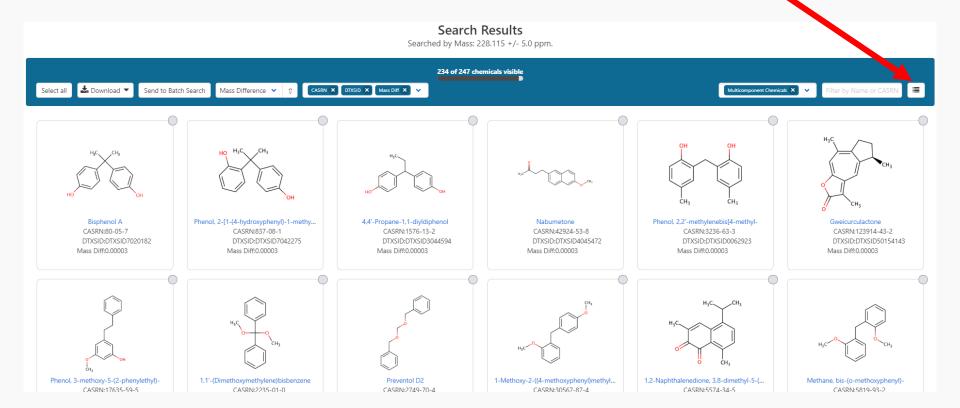
Data Streams for Ranking



- CompTox Dashboard Data Sources
- Pub©hem Data Source Count
- Publed.gov
 Reference Count
- Toxcast in vitro bioactivity
- Presence in CPDat database
- OPERA PhysChem Properties
- Other possibilities predicted media occurrence, frequency of InChIs online

Search 228.115 +/- 5.0 ppm 234 single component chemicals





54

Search 228.115 +/- 5.0 ppm 234 single component chemicals



CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Sources	PubMed Ref. Counts
80-05-7	Level 1	326	170	161	3850
42924-53-8	Level 2	14	45	138	342
87619-52-1	Level 5	0	2		0
87607-32-7	Level 5	0	2		0

The original ChemSpider work



Compound class	Number in class	Average rank	Number of compounds in each position rank-ordered				
			#1	#2	#3	#4	#5+
Pharmaceutical drug	72	1.4	55	9	6	2	
Industrial chemicals	42	5.5	28	6	3		5
Personal care products	8	6.1	3	1			4
Steroid hormones	7	1.0	7				
Perfluorochemicals	6	1.2	5	1			
Pesticides	12	2.3	6	2	3		1
Veterinary drugs	3	1.3	2	1			
Dyes	2	1.0	2				
Food product/natural compounds	4	3.8	2			1	1
Illicit drugs	2	2.0	1		1		
Misc. molecules	3 ^a	1.3	2	1			



- ChemSpider was 26 million chemicals for the original work
- Much BIGGER today
- Is bigger better??



Are there other metadata to use for ranking?

Comparing Search Performance

SEPA United States Environmental Protection

CrossMark

Änal Bioanal Chem (2017) 409:1729–1735 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

- When dashboard contained 720k chemicals
- Only 3% of ChemSpider size
- What was the comparison in performance?

SAME dataset for comparison



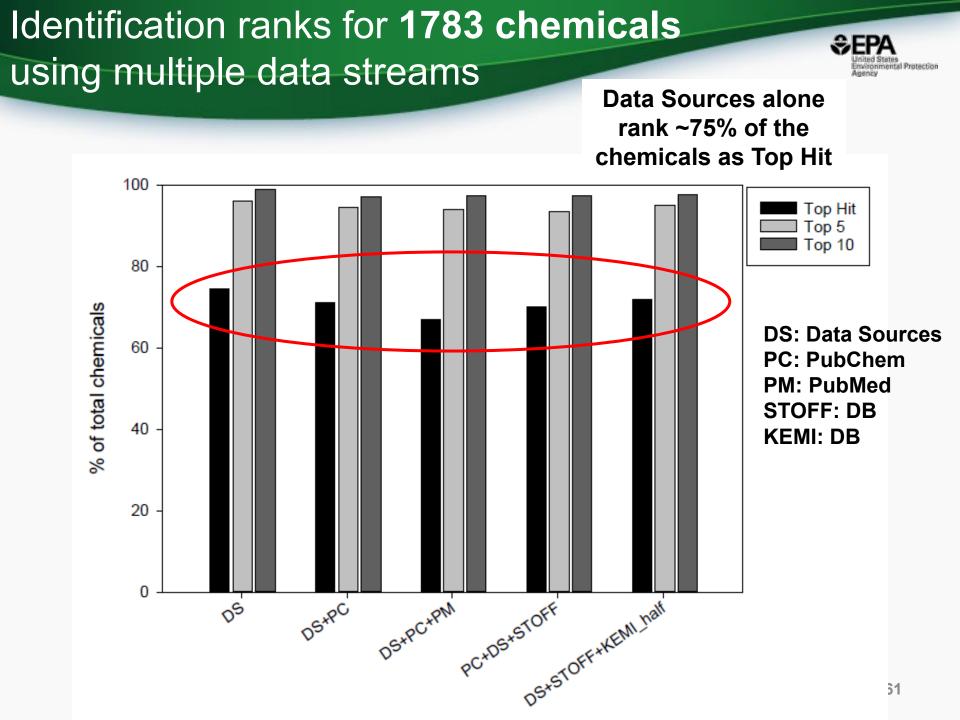
Compound class	Number in class	Average rank		ber of c ion rank		ds in ea 1	ch
			#1	#2	#3	#4	#5+
Pharmaceutical drug	72	1.4	55	9	6	2	
Industrial chemicals	42	5.5	28	6	3		5
Personal care products	8	6.1	3	1			4
Sterbid homones Perfluorochemicals		SAME	7 5) A	TA	SE	Т
Pesticides	12	2.3	6	2	3		1
Veterinary drugs	3	1.3	2	1			
Dyes	2	1.0	2				
Food product/natural compounds	4	3.8	2			1	1
Illicit drugs	2	2.0	1		1		
Misc. molecules	3 ^a	1.3	2	1			



	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

For the same 162 chemicals, Dashboard outperforms ChemSpider for both Mass and Formula Ranking



Prototype Work in Progress



• CFM-ID

- Viewing and Downloading pre-predicted spectra
- Search spectra against the database
- Structure/substructure/similarity search
- Access to API and web services
- Integration to EPA "Chemical Transformation Simulator"

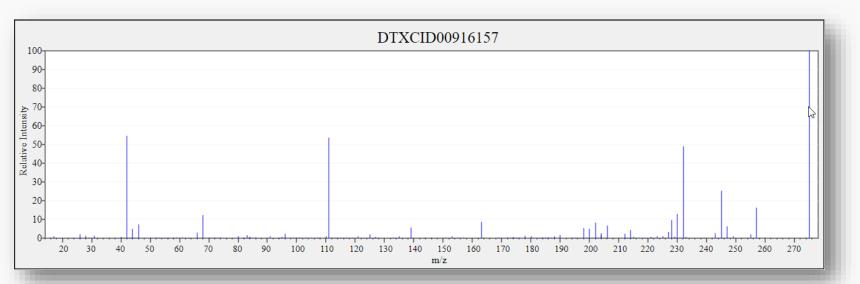
Predicted Mass Spectra

http://cfmid.wishartlab.com/





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



Search Expt. vs. Predicted Spectra



SEPA United States Environmental Protection Home Adv Agency	ranced Search Batch Search Lists 🗸 Prediction	ns Downloads	Share 🔻 🔍 Search all data
	Nor	n Target Analysis Prototype	Î
Mass Se ± 321.138493	Min/Max	± 0.000002 Da p	pm
Molecular F	ar Formula Search		
Mass or Ionizatio ESI+ ESI- EI Spectra		fore searching spectrum	
304.133200 198.09134	Ile Energy Multiple		
198.07569 218.10190 200.40800	59 6.538348292 04 5.269463115 51 4.700481978		
Peak Ma Search	atch Window: 0.02	Da ppm	

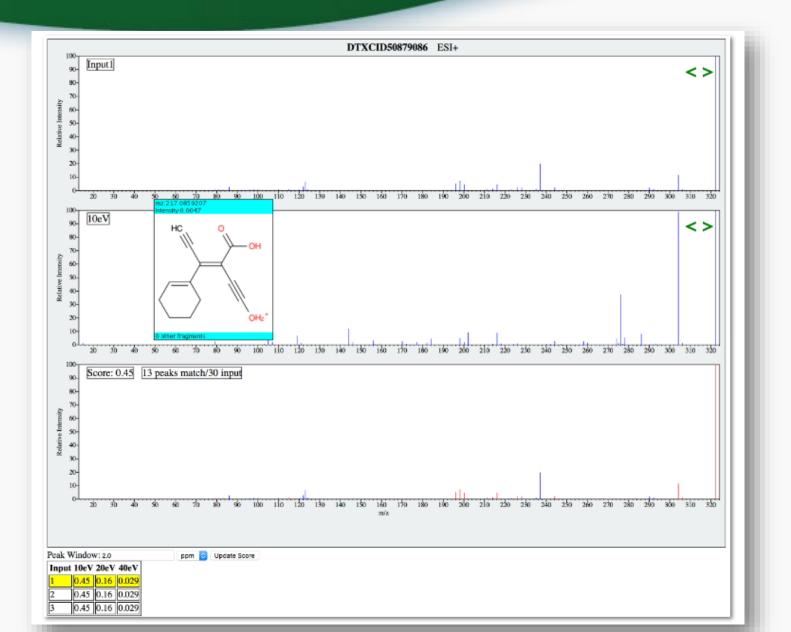
Search Expt. vs. Predicted Spectra



Sepa United State Environmer Agency	es ntal Protection Home Adv	anced Search Batch Search Lists 🛩 Predictions Downloads	Share •	Q Search all data
	Spectra Inpu Single Ener	Chemical Structure ID	Score (10eV)	Î
R	304.1332052 11.61 198.0913404 7.30 123.0440559 8.53 196.0756904 5.26	DTXCID101048191	0.22	
	216.1019051 4.70	DTXCID101181567	0.19	
	Peak Match	DTXCID50879086	0.17	
TSV	CSV Excel	DTXCID60686349	0.14	
	emical Structure ID	DTXCID00830900	0.13	n of Scores
	CID101181587 CID50879088	DTXCID10971176	0.12	
	CID60686349 CID00830900	DTXCID60301242	0.12	
	CID10971176 CID60301242	DTXCID40703048	0.11	
	CID60349982	DTXCID60349982	0.11	
	LID10316649	DTXCID10316649	0.09	1 2 3 4 Next

Spectral Viewer Comparison





Published: Alex Chao et al



Analytical and Bioanalytical Chemistry

RESEARCH PAPER

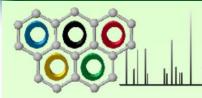
In silico MS/MS spectra for identifying unknowns: a critical examination using CFM-ID algorithms and ENTACT mixture samples

Alex Chao^{1,2} • Hussein Al-Ghoul^{1,2} • Andrew D. McEachran^{1,3} • Ilya Balabin⁴ • Tom Transue⁴ • Tommy Cathey⁴ • Jarod N. Grossman^{2,3} • Randolph Singh^{1,5} • Elin M. Ulrich² • Antony J. Williams⁶ • Jon R. Sobus²

Received: 4 October 2019 / Revised: 27 November 2019 / Accepted: 11 December 2019 © The Author(s) 2019

CASMI 2012-2017 revisited







Critical Assessment of Small Molecule Identification

The experimental and computational mass spectrometry communities are invited to participate in the fifth round of an open contest on the identification of small molecules from mass spectrometry data.

This year the contest will test the applicability of MS and MS/MS on natural products chemistry identifications. With 45 (Category 1) and up to 243 (Categories 2&3) natural products challenges - including a few tricky ones - there's something for everyone!

 Application of metadata candidate ranking and CFM-ID to all five years of CASMI data



Open Access Article

Revisiting Five Years of CASMI Contests with EPA Identification Tools

by 🔃 Andrew D. McEachran ^{1,*}			
📢 Christopher Grulke ² 🖾 💿, 📢	🕽 Jon R. Sobus ² 🖂 回 and 🏹	Antony J. Williams ^{2,*} 🖂	(D)

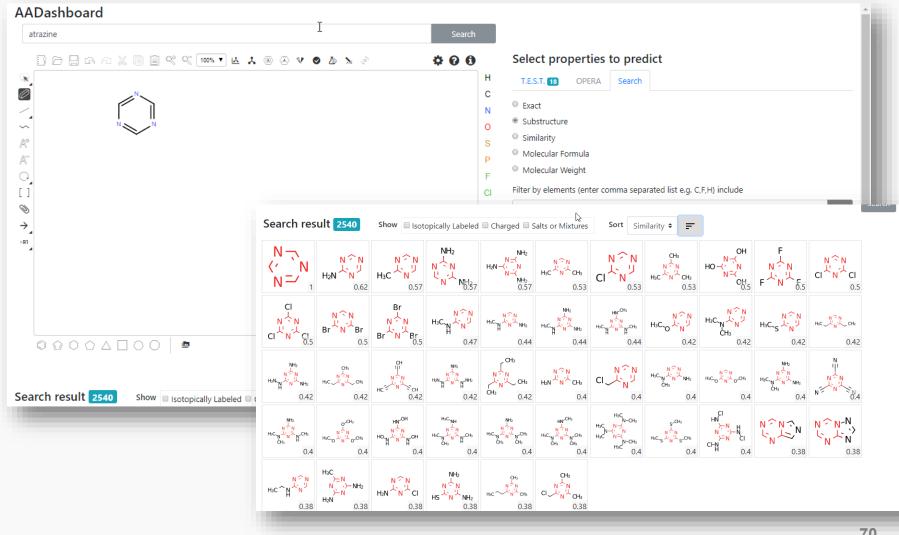
- ¹ Oak Ridge Institute for Science and Education (ORISE) Participant, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, USA
- ² Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, USA
- * Authors to whom correspondence should be addressed.

Metabolites 2020, 10(6), 260; https://doi.org/10.3390/metabo10060260

Received: 3 May 2020 / Revised: 3 June 2020 / Accepted: 17 June 2020 / Published: 23 June 2020

Prototype Development

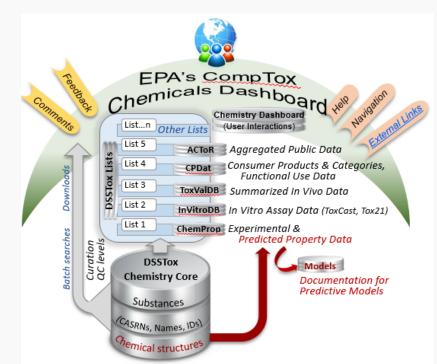




70

Conclusion

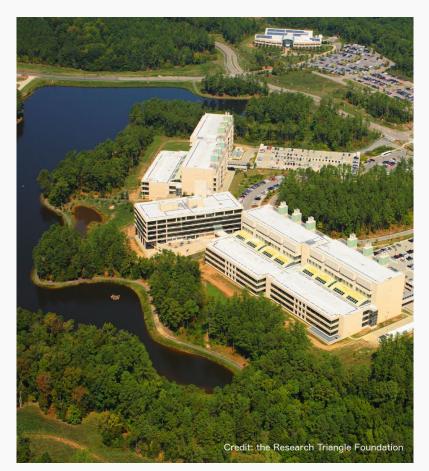
- Dashboard access to data for ~883,000 chemicals MS-Ready data facilitates structure identification
- Related metadata facilitates candidate ranking
- Relationship mappings and chemical lists of great utility
- Curation and mutual sharing of chemical lists is important (e.g. NORMAN)



nmental Protection

Acknowledgements





EPA ORD

Ann Richard Chris Grulke Jeremy Dunne Jeff Edwards Grace Patlewicz Alex Chao **Kristin Isaacs Charles Lowe** James McCord Seth Newton Katherine Phillips Jon Sobus Mark Strynar Elin Ulrich Joach Pleil

TEAMS

IT Development Team Curation Team

ILS Kamel Mansouri

<u>GDIT</u>

Ilya Balabin Tom Transue Tommy Cathey

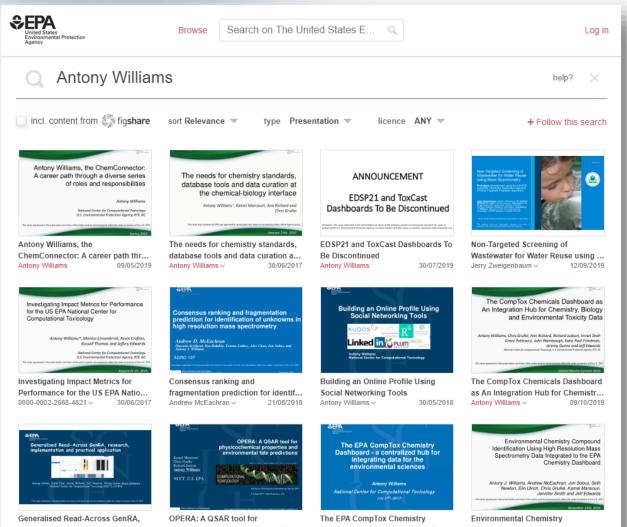
Collaborators

Emma Schymanski NORMAN Network Andrew McEachran

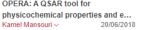
MANY presentations online

IDS://TIHYUFI.COM/W5hgs55





research, implementation and prac... Grace Patlewicz V 18/09/2018



Dashboard - a Centralized Hub for ... Antony Williams 05/07/2018

Compound

Identification Using High Resolutio... Antony Williams ~ 30/06/2017

Contact



Antony Williams CCTE, US EPA Office of Research and Development, Williams.Antony@epa.gov

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

[Williams et al. J Cheminform (2017) 9:61 DOI 10.1186/s13321-017-0247-6

Journal of Cheminformatics

DATABASE





The CompTox Chemistry Dashboard: a community data resource for environmental chemistry

Antony J. Williams^{1*}⁽¹⁾, Christopher M. Grulke¹, Jeff Edwards¹, Andrew D. McEachran², Kamel Mansouri^{1,2,4}, Nancy C. Baker³, Grace Patlewicz¹, Imran Shah¹, John F. Wambaugh¹, Richard S. Judson¹ and Ann M. Richard¹

https://doi.org/10.1186/s13321-017-0247-6



DEMO