

Chemistry data: Distortion and dissemination in the Internet Era

*Antony Williams¹, Chris Grulke¹, Katie Paul-Friedman¹,
and Emma Schymanski²*

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

2) Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Campus Belval, Luxembourg


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

*August 2018
ACS Fall Meeting, Boston*

CompTox Dashboard

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





United States
Environmental Protection
Agency

HomeAdvanced SearchBatch SearchLists▼PredictionsDownloads

Share▼



ChemicalsProduct/Use CategoriesAssay/Gene

☐ Identifier substring search

762 Thousand Chemicals

See what people are saying, read the dashboard comments!
Cite the Dashboard Publication [click here](#)

Latest News

[Read more news](#)

YouTube video regarding using the Dashboard for Non-Targeted Analysis

ch 7th, 2018 at 9:43:36 AM

YouTube video discussing the application of the CompTox Chemistry Dashboard to support non-targeted analysis by mass spectrometry is available. This short video summarizes the advantages The dashboard in terms of data quality and focused data set for environmental non-targeted analysis. [View it here on Youtube.](#)

Ar
Mar



UNITED STATES
ENVIRONMENTAL PROTECTION AGENCY

Discover.

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

Connect.

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

Ask.

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


Underneath the Dashboard

View/Edit a Single Record Structure Search Browse/Curate Records Export DSSTox Chemotypes Manage Chemical Lists Manage Property Data Add Deleted Casm

Preferred Name matched
null
You are viewing the record associated with
DTXSID80198757
CASRN: 62885-41-0

4-Hydroxy-3-methoxy

Valid license cannot be found

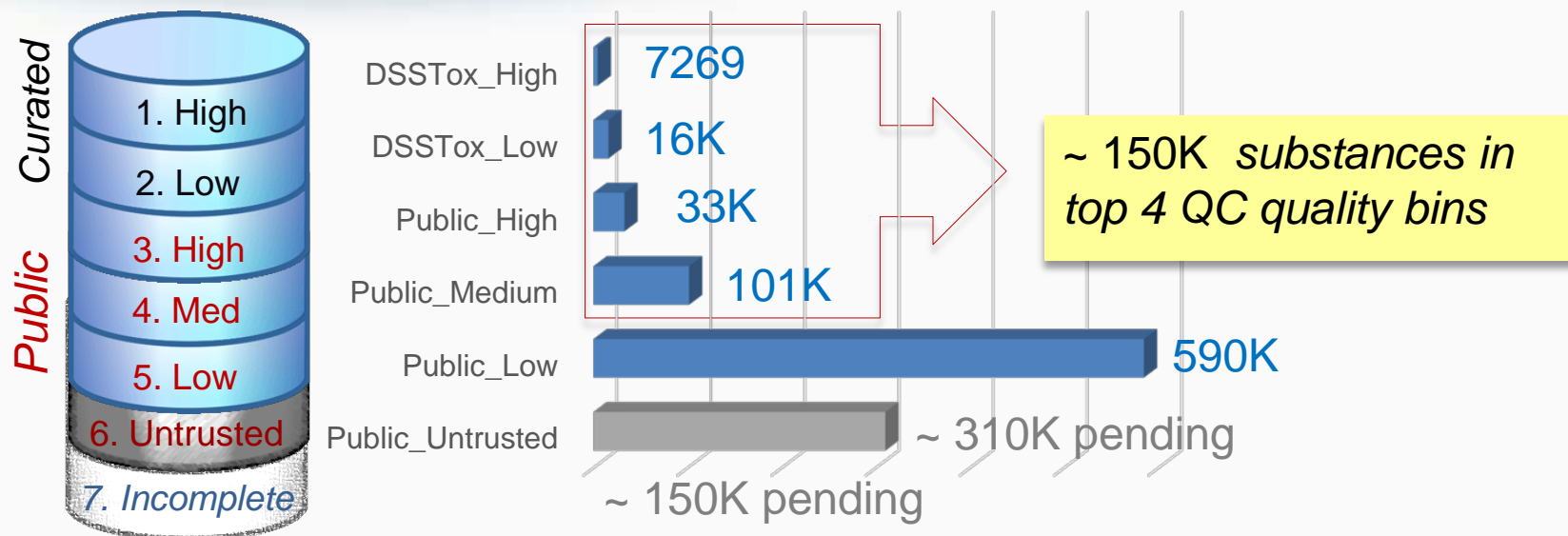


Chemical structure diagram of 4-Hydroxy-3-methoxypyridine, showing a pyridine ring with a hydroxyl group at position 4 and a methoxy group at position 3.

Calculate from Structure

Substance_ID:	DTXSID80198757	Compound_ID:	DTXCID40121248
CAS:	62885-41-0	Chemical Shown:	Tested Chemical
Name:	4-Hydroxy-3-methoxypyridine	Private Notes:	
Substance Type:	Single Compound	Source of CAS-Compound:	STN(DSSTox)
QC Level:	DSSTox_High	Double Stereo:	None
Data Source:	STN(DSSTox)	Chiral Stereo:	None
QC Notes:	CAS [50700-60-2] assigned by DSSTox to pyridin-one tautomer form, which resolves to hydroxy form thru InChI	Chemical Form:	Organic
		Organic Form:	Parent

Distribution of curated data



QC Levels

DSSTox_High:	Hand curated and validated
DSSTox_Low:	Hand curated and confirmed using multiple public sources
Public_High:	Extracted from EPA SRS and confirmed to have no conflicts in ChemID and PubChem
Public_Medium:	Extracted from ChemID and confirmed to have no conflicts in PubChem
Public_Low:	Extracted from ACToR or PubChem
Public_Untrusted:	Postulated, but found to have conflicts in public sources

Record Information Quality Flags

Bisphenol A


80-05-7 | DTXSID7020182

Searched by Approved Name.

Record Information



Citation: U.S. Environmental Protection Agency. Chemistry Dashboard. <https://comptox.epa.gov/dashboard/DTXSID7020182> (accessed Aug 20th, 2018), Bisphenol A.

 **Quality:**

Level 1: Expert curated, highest confidence in accuracy and consistency of unique chemical identifiers

Level 2: Expert curated, unique chemical identifiers using multiple sources

Level 3: Programmatically curated from high quality EPA source, unique chemical identifiers have no conflicts in ChemID and PubChem

Level 4: Programmatically curated from ChemID, unique chemical identifiers have no conflicts in PubChem

Level 5: Programmatically curated from ACToR or PubChem, unique chemical identifiers with low confidence, single public source

Level 4: Programmatically curated from ChemID, unique chemical identifiers have no conflicts in PubChem

Level 5: Programmatically curated from ACToR or PubChem, unique chemical identifiers with low confidence, single public source

ChemReg Curation

The ALANWOOD Pesticide Set

View/Edit a Single Record Structure Search Browse/Curate Records Export DSSTox Chemotypes **Manage Chemical Lists** Manage Property Data Add Deleted Casrns Welcome, Chris Logout

Welcome cgrulke

Editing Listname: ALANWOOD

External Check Results	
Description	Records
Curator Validated	1216
Resolved Duplicates	0
Ignored	0
Structure matched STRUCTURE Preferred Name matched NAME CAS-RN matched CASRN	2
Structure matched STRUCTURE Valid Synonym matched NAME CAS-RN matched CASRN	71
Structure matched STRUCTURE Unique Synonym matched NAME CAS-RN matched CASRN	106
Structure matched STRUCTURE Unique Synonym matched NAME Other CAS-RN matched CASRN	2
Structure matched	

Substance Mapping						
(1 of 5) 1 2 3 4 5 25						
	Source Casrn	Source Name	Hit Substance_ID	Hit Casrn	Hit Name	
1	88-82-4	2,3,5-tri-iodobenzoic acid	DTXSID4041317	88-82-4	2,3,5-Triiodobenzoic acid	Validate Mapping
2	50-31-7	2,3,6-TBA	DTXSID6040296	50-31-7	2,3,6-Trichlorobenzoic acid	Validate Mapping
3	122-88-3	4-CPA	DTXSID9034282	122-88-3	4-Chlorophenoxyacetic acid	Validate Mapping
4	126448-41-7	acibenzolar	DTXSID20155187	126448-41-7	Acibenzolar [ISO]	Validate Mapping
5	76636-10-7	amibuzin	DTXSID20227459	76636-10-7	Amibuzin [ISO]	Validate Mapping
6	3566-10-7	amobam	DTXSID0058067	3566-10-7	Ambam	Validate Mapping
7	86-88-4	antu	DTXSID8020919	86-88-4	1-(1-Naphthyl)-2-thiourea	Validate Mapping
8	52-46-0	apholate	DTXSID7073149	52-46-0	1,3,5,2,4,6-Triazatriphosphorine, 2,2,4,4,6,6-hexakis(1-aziridinyl)-2,2,4,4,6,6-hexahydro-	Validate Mapping
9	3586-60-5	asomate	DTXSID70189412	3586-60-5	Arsine, tris(dimethyldithiocarbamoyl)	Validate Mapping
10	28956-64-1	bentaluron	DTXSID30183153	28956-64-1	Bentaluron [ISO]	Validate Mapping
11	21564-17-0	benthiazole	DTXSID6032647	21564-17-0	2-(Thiocyanomethylthio)benzoic acid	Validate Mapping
12	1022-46-4	bentranil	DTXSID60144732	1022-46-4	4H-3,1-Benzoxazin-4-one, 2-phenyl-	Validate Mapping

The CompTox Dashboard for Structure Identification by MS

Advanced Search?

Mass Search?

±

Min/Max

Select Adduct: Neutral ▼

Mass

Da

±

Error Da

Da

ppm

Search Q

Molecular Formula Search?

Molecular Formula

☒ MS Ready Formula ?

☐ Exact Formula ?

Search Q

Generate Molecular Formula(e) ?

±

Min/Max

Mass

Da

±

Error

Da

ppm

Search Q

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]

- Mapping between our data (and websites) has resulted in collaborative data curation
- Collaboration with Emma Schymanski re. the NORMAN Suspects Exchange
<https://www.norman-network.com/?q=node/236>
- It has highlighted data quality issues – and these are not unexpected!

NORMAN Suspect Exchange

NORMAN

Network of reference laboratories, research centres and related organisations for monitoring of emerging environmental substances

[Home](#) | [NORMAN Network](#) | [Working Groups](#) | [Membership](#) | [NORMAN Bulletin](#) | [Success Stories](#) | [Publications](#) | [Job opportunities](#) | [Contact](#) | [Gallery](#) | [Members' Area](#) | [NORMAN GA meetings](#)

Menu

- Emerging Substances
- DATABASES
- Topics and Activities
- Workshops and Events
- QA/QC Issues
- Glossary
- Useful links
- Members' Area

User login

Username *

Password *

[Request new password](#)

Home

<http://www.norman-network.com/?q=node/236>

NORMAN Suspect List Exchange

In September 2014, NORMAN members expressed the need to exchange various lists of substances to improve their suspect screening efforts. This website was established as part of the 2015 Joint Programme of Activities as a central access point for NORMAN members (and others) to find suspect lists relevant for their environmental monitoring question. All suspect lists currently available are compiled in the table below and on the US EPA CompTox Chemistry Dashboard ([website](#), [downloads](#), [chemical lists](#)).

The "Link to full list" column below contains an excel or comma-separated file (csv) with all available information, e.g. as provided as supporting information for the publication, while the third column provides a list of the structures as InChIKeys only, which allows suspect searching using MetFrag or other workflows. The fourth column contains references for the data: please cite these references if you use the respective datasets.

Recent Suspect Exchange and Dashboard presentations/publications include: **ICCE Oslo 2017: NORMAN Suspects meet the Dashboard** and **NORMAN MassBank and Suspect Exchange**; SETAC Mixtures Denver: **Identifying Complex Mixtures with Cheminformatics and HR-MS**; ACS Fall 2017: **Markush Enumeration for UVCBs** and a [viewpoint article](#).

No.	Abbreviation	Description	Link to full list	Link to InChIKey list	References
	SUSDAT	Merged NORMAN Suspect List: SusDat	Interactive Data table (updating...)	MS-ready InChIKeys (1/03/2018)	A merged list of >40,000 structures from suspect lists. See interactive version . Compiled by Reza Aalizadeh, University of Athens, including RTI and toxicity values, support by Nikiforos Alygizakis, EI. <i>Work in progress ... please report any issues!</i>
S1	MASSBANK	NORMAN Compounds in MassBank	CSV, XLSX with Fragments (3/10/2017) CompTox MassBank EU Reference List CompTox MassBank EU Special Cases CompTox Fragment Download	MassBankEUInChIKeys (11/04/2017)	www.massbank.eu Stravs <i>et al.</i> 2013. DOI: 10.1002/jms.3131
S2	STOFFIDENT	HSWT/LfU STOFF-IDENT Database of Water-Relevant Substances	STOFF-IDENT Contents (6/09/2017) CompTox STOFF-IDENT List Further curation in progress...	STOFF-IDENT InChIKeys (6/09/2017)	The database enables the search for exact masses from target or unknown lists and the automatic use of a Retention Time Index. See: https://www.lfu.bayern.de/stoffident/#!home (single search for free; batch search after free registration).
S3	NORMANCT15	NORMAN Collaborative Trial Targets and Suspects	LC-MS: CSV, XLSX (3/10/2017) GC-MS: CSV, XLSX (3/10/2017) CompTox NORMANCT15 List	LC-MS InChIKeys (31/10/2016) GC-MS InChIKeys (31/10/2016)	Schymanski <i>et al.</i> 2015. DOI: 10.1007/s00216-015-8681-7

Example: NORMAN Priority List

S15	NORMANPRI	NORMAN Priority List	NORMAN Priority CSV (13/7/2017) CompTox NORMAN Priority List Further curation in progress...	NORMAN Priority InChIKeys (16/05/2017)	Priority substances from NORMAN WG-1 (Prioritisation), provided by Valeria Dulio.
-----	-----------	-----------------------------	--	---	---

mol_ID	Name	CAS_RN	SMILES
StNorman2	Microcystin-RR	CAS_RN: 111755-37-4	<chem>C[C@H]1[C@@H](NC(=O)[C@@H](NC(=O)[C@H]1C(=O)N)C(=O)N</chem>
StNorman3	Microcystin-YR	CAS_RN: 101064-48-6	<chem>C[C@H]1[C@@H](NC(=O)[C@@H](NC(=O)[C@H]1C(=O)N)C(=O)N</chem>
StNorman5	2,6-Di-tert-butylphenol	CAS_RN: 128-39-2	<chem>CC(C)(C)c1cccc(c1O)C(C)(C)C</chem>
StNorman6	Butylated hydroxyanisole	CAS_RN: 25013-16-5	<chem>COc1ccc(O)c(c1)C(C)(C)C</chem>
StNorman7	tert-Butylhydroquinone	CAS_RN: 1948-33-0	<chem>CC(C)(C)c1cc(O)ccc1O</chem>
StNorman8	Butylated hydroxytoluene	CAS_RN: 128-37-0	<chem>Cc1cc(c(O)c(c1)C(C)(C)C)C(C)(C)C</chem>
StNorman15	Diethylenetriaminepentaacetic acid	CAS_RN: 67-43-6	<chem>OC(=O)CN(CCN(CC(O)=O)CC(O)=O)CCN(CC(O)=O)CC(O)=O</chem>
StNorman16	Ethylenediaminetetraacetic acid	CAS_RN: 60-00-4	<chem>OC(=O)CN(CCN(CC(O)=O)CC(O)=O)CC(O)=O</chem>
StNorman17	Nitrilotriacetic acid	CAS_RN: 139-13-9	<chem>OC(=O)CN(CC(O)=O)CC(O)=O</chem>
StNorman18	Oxadixyl	CAS_RN: 77732-09-3	<chem>COCC(=O)N(N1CCOC1=O)c2c(C)cccc2C</chem>
StNorman19	Tetraacetylenediamine	CAS_RN: 10543-57-4	<chem>CC(=O)N(CCN(C(C)=O)C(C)=O)C(C)=O</chem>
StNorman26	2-(2-(4-Nonylphenoxy)ethoxy)acetic acid	CAS_RN: 106807-78-7	<chem>CCCCCCCCCc1ccc(OCCOCC(O)=O)cc1</chem>
StNorman29	4-Octylphenoxy acetic acid	CAS_RN: 15234-85-2	<chem>CCCCCCCCCc1ccc(OCC(O)=O)cc1</chem>
StNorman30	Cyanoformaldehyde	CAS_RN: 4471-47-0	<chem>NCC=O</chem>
StNorman31	Decabromodiphenyl ethane	CAS_RN: 84852-53-9	<chem>C(CC1=C(C(=C(C(=C1Br)Br)Br)Br)Br)C2=C(C(=C(Br)C1CC(Br)CC(Br)CC(Br)CC(Br)CC(Br)C1)C(=C2)Br</chem>
StNorman32	Hexabromocyclododecane	CAS_RN: 25637-99-4	<chem>BrC1CC(Br)CC(Br)CC(Br)CC(Br)CC(Br)C1</chem>
StNorman33	n-Nitrosodimethylamine	CAS_RN: 62-75-9	<chem>CN(C)N=O</chem>
StNorman34	Benzylbutylphthalate	CAS_RN: 85-68-7	<chem>CCCCOC(=O)c1cccc1C(=O)OCCc2ccccc2</chem>
StNorman35	Diethyl phthalate	CAS_RN: 84-66-2	<chem>CCOC(=O)c1cccc1C(=O)OCC</chem>
StNorman37	Di-n-butylphthalate	CAS_RN: 84-74-2	<chem>CCCCOC(=O)c1cccc1C(=O)OCCCC</chem>
StNorman39	Bisphenol A	CAS_RN: 80-05-7	<chem>CC(C)(c1ccc(O)cc1)c2ccc(O)cc2</chem>

Mapping on Two Identifiers Lookup Based on Name

LOOKUP BASED ON CHEMICAL NAME							
DTXSID	PREFERRED NAME	CASRN	INCHI KEY	IUPAC NAME	SMILES	MOL FORMULA	MONOISC
DTXSID40880085	Microcystin RR	111755-37-4	JIGDOBKZ	(5R,8S,11R)-2,6-Di-tert-butylphenyl 2,6-di-tert-butyl-4-hydroxy-3-methyl-5-pyridinecarboxylate	CC(C)(C)C1=CC=C(C=C1)C2=CC=CC=C2C3=CC=CC=C3C4=CC=CC=C4C5=CC=CC=C5C6=CC=CC=C6C7=CC=CC=C7C8=CC=CC=C8C9=CC=CC=C9C10=CC=CC=C10C11=CC=CC=C11C12=CC=CC=C12C13=CC=CC=C13C14=CC=CC=C14C15=CC=CC=C15C16=CC=CC=C16C17=CC=CC=C17C18=CC=CC=C18C19=CC=CC=C19C20=CC=CC=C20C21=CC=CC=C21C22=CC=CC=C22C23=CC=CC=C23C24=CC=CC=C24C25=CC=CC=C25C26=CC=CC=C26C27=CC=CC=C27C28=CC=CC=C28C29=CC=CC=C29C30=CC=CC=C30C31=CC=CC=C31C32=CC=CC=C32C33=CC=CC=C33C34=CC=CC=C34C35=CC=CC=C35C36=CC=CC=C36C37=CC=CC=C37C38=CC=CC=C38C39=CC=CC=C39C40=CC=CC=C40C41=CC=CC=C41C42=CC=CC=C42C43=CC=CC=C43C44=CC=CC=C44C45=CC=CC=C45C46=CC=CC=C46C47=CC=CC=C47C48=CC=CC=C48C49=CC=CC=C49C50=CC=CC=C50C51=CC=CC=C51C52=CC=CC=C52C53=CC=CC=C53C54=CC=CC=C54C55=CC=CC=C55C56=CC=CC=C56C57=CC=CC=C57C58=CC=CC=C58C59=CC=CC=C59C60=CC=CC=C60C61=CC=CC=C61C62=CC=CC=C62C63=CC=CC=C63C64=CC=CC=C64C65=CC=CC=C65C66=CC=CC=C66C67=CC=CC=C67C68=CC=CC=C68C69=CC=CC=C69C70=CC=CC=C70C71=CC=CC=C71C72=CC=CC=C72C73=CC=CC=C73C74=CC=CC=C74C75=CC=CC=C75C76=CC=CC=C76C77=CC=CC=C77C78=CC=CC=C78C79=CC=CC=C79C80=CC=CC=C80C81=CC=CC=C81C82=CC=CC=C82C83=CC=CC=C83C84=CC=CC=C84C85=CC=CC=C85C86=CC=CC=C86C87=CC=CC=C87C88=CC=CC=C88C89=CC=CC=C89C90=CC=CC=C90C91=CC=CC=C91C92=CC=CC=C92C93=CC=CC=C93C94=CC=CC=C94C95=CC=CC=C95C96=CC=CC=C96C97=CC=CC=C97C98=CC=CC=C98C99=CC=CC=C99C100=CC=CC=C100C101=CC=CC=C101C102=CC=CC=C102C103=CC=CC=C103C104=CC=CC=C104C105=CC=CC=C105C106=CC=CC=C106C107=CC=CC=C107C108=CC=CC=C108C109=CC=CC=C109C110=CC=CC=C110C111=CC=CC=C111C112=CC=CC=C112C113=CC=CC=C113C114=CC=CC=C114C115=CC=CC=C115C116=CC=CC=C116C117=CC=CC=C117C118=CC=CC=C118C119=CC=CC=C119C120=CC=CC=C120C121=CC=CC=C121C122=CC=CC=C122C123=CC=CC=C123C124=CC=CC=C124C125=CC=CC=C125C126=CC=CC=C126C127=CC=CC=C127C128=CC=CC=C128C129=CC=CC=C129C130=CC=CC=C130C131=CC=CC=C131C132=CC=CC=C132C133=CC=CC=C133C134=CC=CC=C134C135=CC=CC=C135C136=CC=CC=C136C137=CC=CC=C137C138=CC=CC=C138C139=CC=CC=C139C140=CC=CC=C140C141=CC=CC=C141C142=CC=CC=C142C143=CC=CC=C143C144=CC=CC=C144C145=CC=CC=C145C146=CC=CC=C146C147=CC=CC=C147C148=CC=CC=C148C149=CC=CC=C149C150=CC=CC=C150C151=CC=CC=C151C152=CC=CC=C152C153=CC=CC=C153C154=CC=CC=C154C155=CC=CC=C155C156=CC=CC=C156C157=CC=CC=C157C158=CC=CC=C158C159=CC=CC=C159C160=CC=CC=C160C161=CC=CC=C161C162=CC=CC=C162C163=CC=CC=C163C164=CC=CC=C164C165=CC=CC=C165C166=CC=CC=C166C167=CC=CC=C167C168=CC=CC=C168C169=CC=CC=C169C170=CC=CC=C170C171=CC=CC=C171C172=CC=CC=C172C173=CC=CC=C173C174=CC=CC=C174C175=CC=CC=C175C176=CC=CC=C176C177=CC=CC=C177C178=CC=CC=C178C179=CC=CC=C179C180=CC=CC=C180C181=CC=CC=C181C182=CC=CC=C182C183=CC=CC=C183C184=CC=CC=C184C185=CC=CC=C185C186=CC=CC=C186C187=CC=CC=C187C188=CC=CC=C188C189=CC=CC=C189C190=CC=CC=C190C191=CC=CC=C191C192=CC=CC=C192C193=CC=CC=C193C194=CC=CC=C194C195=CC=CC=C195C196=CC=CC=C196C197=CC=CC=C197C198=CC=CC=C198C199=CC=CC=C199C200=CC=CC=C200C201=CC=CC=C201C202=CC=CC=C202C203=CC=CC=C203C204=CC=CC=C204C205=CC=CC=C205C206=CC=CC=C206C207=CC=CC=C207C208=CC=CC=C208C209=CC=CC=C209C210=CC=CC=C210C211=CC=CC=C211C212=CC=CC=C212C213=CC=CC=C213C214=CC=CC=C214C215=CC=CC=C215C216=CC=CC=C216C217=CC=CC=C217C218=CC=CC=C218C219=CC=CC=C219C220=CC=CC=C220C221=CC=CC=C221C222=CC=CC=C222C223=CC=CC=C223C224=CC=CC=C224C225=CC=CC=C225C226=CC=CC=C226C227=CC=CC=C227C228=CC=CC=C228C229=CC=CC=C229C230=CC=CC=C230C231=CC=CC=C231C232=CC=CC=C232C233=CC=CC=C233C234=CC=CC=C234C235=CC=CC=C235C236=CC=CC=C236C237=CC=CC=C237C238=CC=CC=C238C239=CC=CC=C239C240=CC=CC=C240C241=CC=CC=C241C242=CC=CC=C242C243=CC=CC=C243C244=CC=CC=C244C245=CC=CC=C245C246=CC=CC=C246C247=CC=CC=C247C248=CC=CC=C248C249=CC=CC=C249C250=CC=CC=C250C251=CC=CC=C251C252=CC=CC=C252C253=CC=CC=C253C254=CC=CC=C254C255=CC=CC=C255C256=CC=CC=C256C257=CC=CC=C257C258=CC=CC=C258C259=CC=CC=C259C260=CC=CC=C260C261=CC=CC=C261C262=CC=CC=C262C263=CC=CC=C263C264=CC=CC=C264C265=CC=CC=C265C266=CC=CC=C266C267=CC=CC=C267C268=CC=CC=C268C269=CC=CC=C269C270=CC=CC=C270C271=CC=CC=C271C272=CC=CC=C272C273=CC=CC=C273C274=CC=CC=C274C275=CC=CC=C275C276=CC=CC=C276C277=CC=CC=C277C278=CC=CC=C278C279=CC=CC=C279C280=CC=CC=C280C281=CC=CC=C281C282=CC=CC=C282C283=CC=CC=C283C284=CC=CC=C284C285=CC=CC=C285C286=CC=CC=C286C287=CC=CC=C287C288=CC=CC=C288C289=CC=CC=C289C290=CC=CC=C290C291=CC=CC=C291C292=CC=CC=C292C293=CC=CC=C293C294=CC=CC=C294C295=CC=CC=C295C296=CC=CC=C296C297=CC=CC=C297C298=CC=CC=C298C299=CC=CC=C299C300=CC=CC=C300C301=CC=CC=C301C302=CC=CC=C302C303=CC=CC=C303C304=CC=CC=C304C305=CC=CC=C305C306=CC=CC=C306C307=CC=CC=C307C308=CC=CC=C308C309=CC=CC=C309C310=CC=CC=C310C311=CC=CC=C311C312=CC=CC=C312C313=CC=CC=C313C314=CC=CC=C314C315=CC=CC=C315C316=CC=CC=C316C317=CC=CC=C317C318=CC=CC=C318C319=CC=CC=C319C320=CC=CC=C320C321=CC=CC=C321C322=CC=CC=C322C323=CC=CC=C323C324=CC=CC=C324C325=CC=CC=C325C326=CC=CC=C326C327=CC=CC=C327C328=CC=CC=C328C329=CC=CC=C329C330=CC=CC=C330C331=CC=CC=C331C332=CC=CC=C332C333=CC=CC=C333C334=CC=CC=C334C335=CC=CC=C335C336=CC=CC=C336C337=CC=CC=C337C338=CC=CC=C338C339=CC=CC=C339C340=CC=CC=C340C341=CC=CC=C341C342=CC=CC=C342C343=CC=CC=C343C344=CC=CC=C344C345=CC=CC=C345C346=CC=CC=C346C347=CC=CC=C347C348=CC=CC=C348C349=CC=CC=C349C350=CC=CC=C350C351=CC=CC=C351C352=CC=CC=C352C353=CC=CC=C353C354=CC=CC=C354C355=CC=CC=C355C356=CC=CC=C356C357=CC=CC=C357C358=CC=CC=C358C359=CC=CC=C359C360=CC=CC=C360C361=CC=CC=C361C362=CC=CC=C362C363=CC=CC=C363C364=CC=CC=C364C365=CC=CC=C365C366=CC=CC=C366C367=CC=CC=C367C368=CC=CC=C368C369=CC=CC=C369C370=CC=CC=C370C371=CC=CC=C371C372=CC=CC=C372C373=CC=CC=C373C374=CC=CC=C374C375=CC=CC=C375C376=CC=CC=C376C377=CC=CC=C377C378=CC=CC=C378C379=CC=CC=C379C380=CC=CC=C380C381=CC=CC=C381C382=CC=CC=C382C383=CC=CC=C383C384=CC=CC=C384C385=CC=CC=C385C386=CC=CC=C386C387=CC=CC=C387C388=CC=CC=C388C389=CC=CC=C389C390=CC=CC=C390C391=CC=CC=C391C392=CC=CC=C392C393=CC=CC=C393C394=CC=CC=C394C395=CC=CC=C395C396=CC=CC=C396C397=CC=CC=C397C398=CC=CC=C398C399=CC=CC=C399C400=CC=CC=C400C401=CC=CC=C401C402=CC=CC=C402C403=CC=CC=C403C404=CC=CC=C404C405=CC=CC=C405C406=CC=CC=C406C407=CC=CC=C407C408=CC=CC=C408C409=CC=CC=C409C410=CC=CC=C410C411=CC=CC=C411C412=CC=CC=C412C413=CC=CC=C413C414=CC=CC=C414C415=CC=CC=C415C416=CC=CC=C416C417=CC=CC=C417C418=CC=CC=C418C419=CC=CC=C419C420=CC=CC=C420C421=CC=CC=C421C422=CC=CC=C422C423=CC=CC=C423C424=CC=CC=C424C425=CC=CC=C425C426=CC=CC=C426C427=CC=CC=C427C428=CC=CC=C428C429=CC=CC=C429C430=CC=CC=C430C431=CC=CC=C431C432=CC=CC=C432C433=CC=CC=C433C434=CC=CC=C434C435=CC=CC=C435C436=CC=CC=C436C437=CC=CC=C437C438=CC=CC=C438C439=CC=CC=C439C440=CC=CC=C440C441=CC=CC=C441C442=CC=CC=C442C443=CC=CC=C443C444=CC=CC=C444C445=CC=CC=C445C446=CC=CC=C446C447=CC=CC=C447C448=CC=CC=C448C449=CC=CC=C449C450=CC=CC=C450C451=CC=CC=C451C452=CC=CC=C452C453=CC=CC=C453C454=CC=CC=C454C455=CC=CC=C455C456=CC=CC=C456C457=CC=CC=C457C458=CC=CC=C458C459=CC=CC=C459C460=CC=CC=C460C461=CC=CC=C461C462=CC=CC=C462C463=CC=CC=C463C464=CC=CC=C464C465=CC=CC=C465C466=CC=CC=C466C467=CC=CC=C467C468=CC=CC=C468C469=CC=CC=C469C470=CC=CC=C470C471=CC=CC=C471C472=CC=CC=C472C473=CC=CC=C473C474=CC=CC=C474C475=CC=CC=C475C476=CC=CC=C476C477=CC=CC=C477C478=CC=CC=C478C479=CC=CC=C479C480=CC=CC=C480C481=CC=CC=C481C482=CC=CC=C482C483=CC=CC=C483C484=CC=CC=C484C485=CC=CC=C485C486=CC=CC=C486C487=CC=CC=C487C488=CC=CC=C488C489=CC=CC=C489C490=CC=CC=C490C491=CC=CC=C491C492=CC=CC=C492C493=CC=CC=C493C494=CC=CC=C494C495=CC=CC=C495C496=CC=CC=C496C497=CC=CC=C497C498=CC=CC=C498C499=CC=CC=C499C500=CC=CC=C500C501=CC=CC=C501C502=CC=CC=C502C503=CC=CC=C503C504=CC=CC=C504C505=CC=CC=C505C506=CC=CC=C506C507=CC=CC=C507C508=CC=CC=C508C509=CC=CC=C509C510=CC=CC=C510C511=CC=CC=C511C512=CC=CC=C512C513=CC=CC=C513C514=CC=CC=C514C515=CC=CC=C515C516=CC=CC=C516C517=CC=CC=C517C518=CC=CC=C518C519=CC=CC=C519C520=CC=CC=C520C521=CC=CC=C521C522=CC=CC=C522C523=CC=CC=C523C524=CC=CC=C524C525=CC=CC=C525C526=CC=CC=C526C527=CC=CC=C527C528=CC=CC=C528C529=CC=CC=C529C530=CC=CC=C530C531=CC=CC=C531C532=CC=CC=C532C533=CC=CC=C533C534=CC=CC=C534C535=CC=CC=C535C536=CC=CC=C536C537=CC=CC=C537C538=CC=CC=C538C539=CC=CC=C539C540=CC=CC=C540C541=CC=CC=C541C542=CC=CC=C542C543=CC=CC=C543C544=CC=CC=C544C545=CC=CC=C545C546=CC=CC=C546C547=CC=CC=C547C548=CC=CC=C548C549=CC=CC=C549C550=CC=CC=C550C551=CC=CC=C551C552=CC=CC=C552C553=CC=CC=C553C554=CC=CC=C554C555=CC=CC=C555C556=CC=CC=C556C557=CC=CC=C557C558=CC=CC=C558C559=CC=CC=C559C560=CC=CC=C560C561=CC=CC=C561C562=CC=CC=C562C563=CC=CC=C563C564=CC=CC=C564C565=CC=CC=C565C566=CC=CC=C566C567=CC=CC=C567C568=CC=CC=C568C569=CC=CC=C569C570=CC=CC=C570C571=CC=CC=C571C572=CC=CC=C572C573=CC=CC=C573C574=CC=CC=C574C575=CC=CC=C575C576=CC=CC=C576C577=CC=CC=C577C578=CC=CC=C578C579=CC=CC=C579C580=CC=CC=C580C581=CC=CC=C581C582=CC=CC=C582C583=CC=CC=C583C584=CC=CC=C584C585=CC=CC=C585C586=CC=CC=C586C587=CC=CC=C587C588=CC=CC=C588C589=CC=CC=C589C590=CC=CC=C590C591=CC=CC=C591C592=CC=CC=C592C593=CC=CC=C593C594=CC=CC=C594C595=CC=CC=C595C596=CC=CC=C596C597=CC=CC=C597C598=CC=CC=C598C599=CC=CC=C599C600=CC=CC=C600C601=CC=CC=C601C602=CC=CC=C602C603=CC=CC=C603C604=CC=CC=C604C605=CC=CC=C605C606=CC=CC=C606C607=CC=CC=C607C608=CC=CC=C608C609=CC=CC=C609C610=CC=CC=C610C611=CC=CC=C611C612=CC=CC=C612C613=CC=CC=C613C614=CC=CC=C614C615=CC=CC=C615C616=CC=CC=C616C617=CC=CC=C617C618=CC=CC=C618C619=CC=CC=C619C620=CC=CC=C620C621=CC=CC=C621C622=CC=CC=C622C623=CC=CC=C623C624=CC=CC=C624C625=CC=CC=C625C626=CC=CC=C626C627=CC=CC=C627C628=CC=CC=C628C629=CC=CC=C629C630=CC=CC=C630C631=CC=CC=C631C632=CC=CC=C632C633=CC=CC=C633C634=CC=CC=C634C635=CC=CC=C635C636=CC=CC=C636C637=CC=CC=C637C638=CC=CC=C638C639=CC=CC=C639C640=CC=CC=C640C641=CC=CC=C641C642=CC=CC=C642C643=CC=CC=C643C644=CC=CC=C644C645=CC=CC=C645C646=CC=CC=C646C647=CC=CC=C647C648=CC=CC=C648C649=CC=CC=C649C650=CC=CC=C650C651=CC=CC=C651C652=CC=CC=C652C653=CC=CC=C653C654=CC=CC=C654C655=CC=CC=C655C656=CC=CC=C656C657=CC=CC=C657C658=CC=CC=C658C659=CC=CC=C659C660=CC=CC=C660C661=CC=CC=C661C662=CC=CC=C662C663=CC=CC=C663C664=CC=CC=C664C665=CC=CC=C665C666=CC=CC=C666C667=CC=CC=C667C668=CC=CC=C668C669=CC=CC=C669C670=CC=CC=C670C671=CC=CC=C671C672=CC=CC=C672C673=CC=CC=C673C674=CC=CC=C674C675=CC=CC=C675C676=CC=CC=C676C677=CC=CC=C677C678=CC=CC=C678C679=CC=CC=C679C680=CC=CC=C680C681=CC=CC=C681C682=CC=CC=C682C683=CC=CC=C683C684=CC=CC=C684C685=CC=CC=C685C686=CC=CC=C686C687=CC=CC=C687C688=CC=CC=C688C689=CC=CC=C689C690=CC=CC=C690C691=CC=CC=C691C692=CC=CC=C692C693=CC=CC=C693C694=CC=CC=C694C695=CC=CC=C695C696=CC=CC=C696C697=CC=CC=C697C698=CC=CC=C698C699=CC=CC=C699C700=CC=CC=C700C701=CC=CC=C701C702=CC=CC=C702C703=CC=CC=C703C704=CC=CC=C704C705=CC=CC=C705C706=CC=CC=C706C707=CC=CC=C707C708=CC=CC=C708C709=CC=CC=C709C710=CC=CC=C710C711=CC=CC=C711C712=CC=CC=C712C713=CC=CC=C713C714=CC=CC=C714C715=CC=CC=C715C716=CC=CC=C716C717=CC=CC=C717C718=CC=CC=C718C719=CC=CC=C719C720=CC=CC=C720C721=CC=CC=C721C722=CC=CC=C722C723=CC=CC=C723C724=CC=CC=C724C725=CC=CC=C725C726=CC=CC=C726C727=CC=CC=C727C728=CC=CC=C728C729=CC=CC=C729C730=CC=CC=C730C731=CC=CC=C731C732=CC=CC=C732C733=CC=CC=C733C734=CC=CC=C734C735=CC=CC=C735C736=CC=CC=C736C737=CC=CC=C737C738=CC=CC=C738C739=CC=CC=C739C740=CC=CC=C740C741=CC=CC=C741C742=CC=CC=C742C743=CC=CC=C743C744=CC=CC=C744C745=CC=CC=C745C746=CC=CC=C746C747=CC=CC=C747C748=CC=CC=C748C749=CC=CC=C749C750=CC=CC=C750C751=CC=CC=C751C752=CC=CC=C752C753=CC=CC=C753C754=CC=CC=C754C755=CC=CC=C755C756=CC=CC=C756C757=CC=CC=C757C758=CC=CC=C758C759=CC=CC=C759C760=CC=CC=C760C761=CC=CC=C761C762=CC=CC=C762C763=CC=CC=C763C764=CC=CC=C764C765=CC=CC=C765C766=CC=CC=C766C767=CC=CC=C767C768=CC=CC=C768C769=CC=CC=C769C770=CC=CC=C770C771=CC=CC=C771C772=CC=CC=C772C773=CC=CC=C773C774=CC=CC=C774C775=CC=CC=C775C776=CC=CC=C776C777=CC=CC=C777C778=CC=CC=C778C779=CC=CC=C779C780=CC=CC=C780C781=CC=CC=C781C782=CC=CC=C782C783=CC=CC=C783C784=CC=CC=C784C785=CC=CC=C785C786=CC=CC=C786C787=CC=CC=C787C788=CC=CC=C788C789=CC=CC=C789C790=CC=CC=C790C791=CC=CC=C791C792=CC=CC=C792C793=CC=CC=C793C794=CC=CC=C794C795=CC=CC=C795C796=CC=CC=C796C797=CC=CC=C797C798=CC=CC=C798C799=CC=CC=C799C800=CC=CC=C800C801=CC=CC=C801C802=CC=CC=C802C803=CC=CC=C803C804=CC=CC=C804C805=CC=CC=C805C806=CC=CC=C806C807=CC=CC=C807C808=CC=CC=C808C809=CC=CC=C809C810=CC=CC=C810C811=CC=CC=C811C812=CC=CC=C812C813=CC=CC=C813C814=CC=CC=C814C815=CC=CC=C815C816=CC=CC=C816C817=CC=CC=C817C818=CC=CC=C818C819=CC=CC=C819C820=CC=CC=C820C821=CC=CC=C821C822=CC=CC=C822C823=CC=CC=C823C824=CC=CC=C824C825=CC=CC=C825C826=CC=CC=C826C827=CC=CC=C827C828=CC=CC=C828C829=CC=CC=C829C830=CC=CC=C830C831=CC=CC=C831C832=CC=CC=C832C833=CC=CC=C833C834=CC=CC=C834C835=CC=CC=C835C836=CC=CC=C836C837=CC=CC=C837C838=CC=CC=C838C839=CC=CC=C839C840=CC=CC=C840C841=CC=CC=C841C842=CC=CC=C842C843=CC=CC=C843C844=CC=CC=C844C845=CC=CC=C845C846=CC=CC=C846C847=CC=CC=C847C848=CC=CC=C848C849=CC=CC=C849C850=CC=CC=C850C851=CC=CC=C851C852=CC=CC=C852C853=CC=CC=C853C854=CC=CC=C854C855=CC=CC=C855C856=CC=CC=C856C857=CC=CC=C857C858=CC=CC=C858C859=CC=CC=C859C860=CC=CC=C860C861=CC=CC=C861C862=CC=CC=C862C863=CC=CC=C863C864=CC=CC=C864C865=CC=CC=C865C866=CC=CC=C866C867=CC=CC=C867C868=CC=CC=C868C869=CC=CC=C869C870=CC=CC=C870C871=CC=CC=C871C872=CC=CC=C872C873=CC=CC=C873C874=CC=CC=C874C875=CC=CC=C875C876=CC=CC=C876C877=CC=CC=C877C878=CC=CC=C878C879=CC=CC=C879C880=CC=CC=C880C881=CC=CC=C881C882=CC=CC=C882C883=CC=CC=C883C884=CC=CC=C884C885=CC=CC=C885C886=CC=CC=C886C887=CC=CC=C887C888=CC=CC=C888C889=CC=CC=C889C890=CC=CC=C890C891=CC=CC=C891C892=CC=CC=C892C893=CC=CC=C893C894=CC=CC=C894C895=CC=CC=C895C896=CC=CC=C896C897=CC=CC=C897C898=CC=CC=C898C899=CC=CC=C899C900=CC=CC=C900C901=CC=CC=C901C902=CC=CC=C902C903=CC=CC=C903C904=CC=CC=C904C905=CC=CC=C905C906=CC=CC=C906C907=CC=CC=C907C908=CC=CC=C908C909=CC=CC=C909C910=CC=CC=C910C911=CC=CC=C911C912=CC=CC=C912C913=CC=CC=C913C914=CC=CC=C914C915=CC=CC=C915C916=CC=CC=C916C917=CC=CC=C917C918=CC=CC=C918C919=CC=CC=C919C920=CC=CC=C920C921=CC=CC=C921C922=CC=CC=C922C923=CC=CC=C923C924=CC=CC=C924C925=CC=CC=C925C926=CC=CC=C926C927=CC=CC=C927C928=CC=CC=C928C929=CC=CC=C929C930=CC=CC=C930C931=CC=CC=C931C932=CC=CC=C932C933=CC=CC=C933C934=CC=CC=C934C935=CC=CC=C935C936=CC=CC=C936C937=CC=CC=C937C938=CC=CC=C938C939=CC=CC=C939C940=CC=CC=C940C941=CC=CC=C941C942=CC=CC=C942C943=CC=CC=C943C944=CC=CC=C944C945=CC=CC=C945C946=CC=CC=C946C947=CC=CC=C947C948=CC=CC=C948C949=CC=CC=C949C950=CC=CC=C950C951=CC=CC=C951C952=CC=CC=C952C953=CC=CC=C953C954=CC=CC=C954C955=CC=CC=C955C956=CC=CC=C956C957=CC=CC=C957C958=CC=CC=C958C959=CC=CC=C959C960=CC=CC=C960C961=CC=CC=C961C962=CC=CC=C962C963=CC=CC=C963C964=CC=CC=C964C965=CC=CC=C965C966=CC=CC=C966C967=CC=CC=C967C968=CC=CC=C968C969=CC=CC=C969C970=CC=CC=C970C971=CC=CC=C971C972=CC=CC=C972C973=CC=CC=C973C974=CC=CC=C974C975=CC=CC=C975C976=CC=CC=C976C977=CC=CC=C977C978=CC=CC=C978C979=CC=CC=C979C980=CC=CC=C980C981=CC=CC=C981C982=CC=CC=C982C983=CC=CC=C983C984=CC=CC=C984C985=CC=CC=C985C986=CC=CC=C986C987=CC=CC=C987C988=CC=CC=C988C989=CC=CC=C989C990=CC=CC=C990C991=CC=CC=C991C992=CC=CC=C992C993=CC=CC=C993C994=CC=CC=C994C995=CC=CC=C995C996=CC=CC=C996C997=CC=CC=C997C998=CC=CC=C998C999=CC=CC=C999C1000=CC=CC=C1000C1001=CC=CC=C1001C1002=CC=CC=C1002C1003=CC=CC=C1003C1004=CC=CC=C1004C1005=CC=CC=C1005C1006=CC=CC=C1006C1007=CC=CC=C1007C1008=CC=CC=C1008C1009=CC=CC=C1009C1010=CC=CC=C1010C1011=CC=CC=C1011C1012=CC=CC=C1012C1013=CC=CC=C1013C1014		

Mapping on Two Identifiers Lookup Based on CASRN

LOOKUP BASED ON CASRN							
DTXSID	PREFERRED NAME	CASRN	INCHI KEY	IUPAC NAME	SMILES	MOL FORMULA	MONOISOTOPIC WEIGHT
DTXSID40880085	Microcystin RR	111755-37-4	JIGDOBKZ	(5R,8S,11R)-2,6-Di-tert-butylphenoxy	CC(C)(C)C1=CC=C(C=C1)OC(=O)C2=CC=CC=C2	C49H75N13O12	1037.5658
DTXSID00880086	Microcystin YR	101064-48-6	OWHASZC	(5R,8S,11R)-2,6-Di-tert-butylphenoxy	CC(C)(C)C1=CC=C(C=C1)OC(=O)C2=CC=CC=C2	C52H72N10O13	1044.5280
DTXSID6027052	2,6-Di-tert-butylphenoxy	128-39-2	DKCPKDPY	2,6-Di-tert-butylphenoxy	CC(C)(C)C1=CC=C(C=C1)OC(=O)C2=CC=CC=C2	C14H22O	206.16706
DTXSID7020215	Butylated hydroxyanisole	25013-16-5	CZBZUDVE	2-tert-Butyl-4-methoxyphenol	COC1=CC=C(C=C1)C(C)(C)C	C22H32O4	360.23005
DTXSID6020220	tert-Butylhydroquinone	1948-33-0	BGNXCDM	2-tert-Butyl-1,4-benzenediol	CC(C)(C)C1=CC=C(C=C1)O	C10H14O2	166.09937
DTXSID2020216	Butylated hydroxytoluene	128-37-0	NLZUEZXR	2,6-Di-tert-butyl-4-methylphenol	CC1=CC=C(C=C1)C(C)(C)C	C15H24O	220.18271
DTXSID2023434	Pentetic acid	67-43-6	QPCDCPD	N,N-Bis{2-hydroxy-5-oxo-2-phenylpentanoic acid}	OC(=O)C1=CC=CC=C1C(=O)N(C(=O)C2=CC=CC=C2)C(=O)N(C(=O)C3=CC=CC=C3)C(=O)O	C14H23N3O10	393.13834
DTXSID6022977	Ethylenediaminetetraacetic acid	60-00-4	KCXVZYZY	2,2',2'',2'''-Ethylenediaminetetraacetic acid	OC(=O)C1=CC=CC=C1C(=O)N(C(=O)C2=CC=CC=C2)C(=O)N(C(=O)C3=CC=CC=C3)C(=O)O	C10H16N2O8	292.09066
DTXSID6020939	Nitrilotriacetic acid	139-13-9	MGFYIUZF	2,2',2''-Nitrilotriacetic acid	OC(=O)C1=CC=CC=C1C(=O)N(C(=O)C2=CC=CC=C2)C(=O)O	C6H9NO6	191.04298
DTXSID2032631	Oxadixyl	77732-09-3	UWVQIRO	N-(2,6-Dimethyl-4-oxocyclohex-2-en-1-yl)-2,6-dimethyl-4-oxocyclohex-2-en-1-amine	COCC(=O)C1=CC=CC=C1C(=O)N(C(=O)C2=CC=CC=C2)C(=O)O	C14H18N2O4	278.12665
DTXSID5040752	Tetraacetylenediamine	10543-57-4	BGRWYDF	N,N'-(Ethylenedioxy)bis(4-oxocyclohex-2-en-1-amine)	CC(=O)N(C(=O)C1=CC=CC=C1)C(=O)N(C(=O)C2=CC=CC=C2)C(=O)O	C10H16N2O4	228.11100
DTXSID70147779	2-(2-(4-Nonylphenoxy)ethyl)ethylamine	106807-78-7	RAQHOB	[2-(4-Nonylphenoxy)ethyl]ethylamine	CCCCCCCCC1=CC=CC=C1C(=O)N(C(=O)C2=CC=CC=C2)C(=O)O	C19H30O4	322.21440
DTXSID60165003	Acetic acid, (4-octylphenyl) ester	15234-85-2	DWUYSEM	(4-Octylphenoxy)acetic acid	CCCCCCCCC1=CC=CC=C1C(=O)N(C(=O)C2=CC=CC=C2)C(=O)O	C16H24O3	264.17254
DTXSID0021549	Cyanoformaldehyde	4471-47-0	TUHMQD	Oxoacetone	O=CC#N	C2HNO	55.005813
DTXSID2052732	1,1'-Ethane-1,2-diylbis(4-bromophenyl)	84852-53-9	BZQKBFHE	1,1'-(Ethane-1,2-diyl)bis(4-bromophenyl)	BrC1=CC=C(C=C1)C2=CC=C(C=C2)Br	C14H4Br10	961.21468
DTXSID8025383	Hexabromocyclododecane	25637-99-4		-	-	-	-
DTXSID7021029	N-Nitrosodimethylamine	62-75-9	UMFJAHH	N,N-Dimethylnitrosamine	CN(C)N=O	C2H6N2O	74.048012
DTXSID3020205	Benzyl butyl phthalate	85-68-7	IRIAEXORF	Benzyl butyl phthalate	CCCCOC(=O)C1=CC=CC=C1C(=O)O	C19H20O4	312.13615
DTXSID7021780	Diethyl phthalate	84-66-2	FLKPEMZC	Diethyl phthalate	CCOC(=O)C1=CC=CC=C1C(=O)O	C12H14O4	222.08920
DTXSID2021781	Dibutyl phthalate	84-74-2	DOIRQSBF	Dibutyl phthalate	CCCCOC(=O)C1=CC=CC=C1C(=O)O	C16H22O4	278.15180
DTXSID7020182	Bisphenol A	80-05-7	IISBACLAF	4,4'-(Prop-1-ene-2,2-diylidene)bis(phenol)	CC(C)(C1=CC=CC=C1)C2=CC=CC=C2	C15H16O2	228.11502

Mapping Quality Control

DTXSID_Name	DTXSID_CAS	DTXSID	DTXSID Equal	Name InChIKey	CAS InChIKey	Name Block1	CAS Block1	Name WCAS	Name WName	KeyOR Name	CAS Match	Num NoMatch
DTXSID2047064	DTXSID0020151	DTXSID0020151	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	0
DTXSID7042481	DTXSID2032683	DTXSID2032683	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	0
DTXSID5022853	DTXSID60110018	DTXSID60110018	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	0
DTXSID3023215	DTXSID7047803	DTXSID7047803	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	0
DTXSID1045033	DTXSID8044545	DTXSID8044545	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	0
DTXSID3041083	DTXSID9023386	DTXSID9023386	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	0
DTXSID8021480	DTXSID3030636	DTXSID3030636	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	0
DTXSID80185521	DTXSID90205325	DTXSID90205325	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	0
DTXSID8022826	DTXSID40872344	DTXSID40872344	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	0
DTXSID4021509	DTXSID7061277	DTXSID7061277	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	0
DTXSID20213251	DTXSID5060936	DTXSID5060936	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	0
DTXSID3021803	DTXSID20220667	DTXSID20220667	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	0
DTXSID80109343	DTXSID1034715	DTXSID1034715	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	0
DTXSID2023359	DTXSID9041073	DTXSID9041073	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	0
DTXSID6029094	DTXSID8073471	DTXSID8073471	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	0
DTXSID30860093	DTXSID7025219	DTXSID7025219	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	0
DTXSID1047524	DTXSID2041171	DTXSID2041171	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	0
DTXSID7041099	DTXSID2035726	-	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	0
-	DTXSID7027041	DTXSID7027041	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	1
-	DTXSID3032416	DTXSID3032416	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	1

Example: NORMAN Priority List



S15	NORMANPRI	NORMAN Priority List	NORMAN Priority CSV (13/7/2017) CompTox NORMAN Priority List Further curation in progress...	NORMAN Priority InChIKeys (16/05/2017)	Priority substances from NORMAN WG-1 (Prioritisation), provided by Valeria Dulio.
-----	-----------	-----------------------------	--	---	---

Select List

Show 10 entries

[Download](#)


Search:





List Acronym	List Name	Last Updated	Number of Chemicals	List Description
NORMANPRI	NORMAN Network Priority List	2017-07-14	922	NORMANPRI contains the list of priority substances determined by the NORMAN Network Working Group 1 on Prioritization, provided by Valeria Dulio, INERIS, France. Further details on the website.


Showing 1 to 1 of 1 entries (filtered from 67 total entries)




Example: NORMAN Priority List


S15	NORMANPRI	NORMAN Priority List	NORMAN Priority CSV (13/7/2017) CompTox NORMAN Priority List Further curation in progress...	NORMAN Priority InChIKeys (16/05/2017)	Priority substances from NORMAN WG-1 (Prioritisation), provided by Valeria Dulio.
-----	-----------	-----------------------------	--	---	---



 https://comptox.epa.gov/dashboard/chemical_lists/norman 80%  


 Search






United States
Environmental Protection
Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share  Search all data

NORMAN Network Priority List



☐ Substring search

List Details

Description: NORMANPRI contains the list of priority substances determined by the NORMAN Network [Working Group 1](#) on Prioritization, provided by Valeria Dulio, INERIS, France. Further details are available on the Working Group website. The original data is available on the [NORMAN Suspect List Exchange](#). This list is undergoing continuous curation/extension.

Number of Chemicals: 922

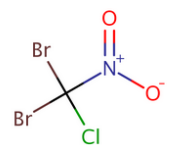
922 chemicals

Download / Send

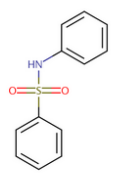
Show info: DTXSID CASRN TOXCAST Select all

Sort by: DTXSID ↑

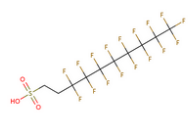
Filter by: Name or CASRN Hide



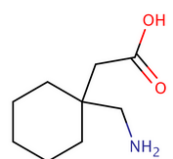
Dibromochloronitromethane
DTXSID:DTXSID00152114
CASRN:1184-89-0
TOXCAST:0



Benzenesulfonanilide
DTXSID:DTXSID00168371
CASRN:1678-25-7
TOXCAST:0



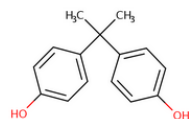
8:2 Fluorotelomer sulfonic acid
DTXSID:DTXSID00192353
CASRN:39108-34-4
TOXCAST:0



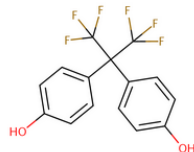
Gabapentin
DTXSID:DTXSID0020074
CASRN:60142-96-3
TOXCAST:0

14

>23 NORMAN Lists Available

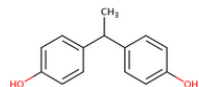


Bisphenol A
DTXSID:DTXSID7020182
CASRN:80-05-7
TOXCAST:166/812

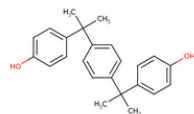


Bisphenol AF
DTXSID:DTXSID7037717
CASRN:1478-61-1
TOXCAST:248/639

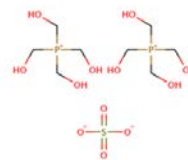
Bisphenol Compounds



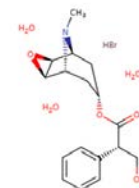
Bisphenol E
DTXSID:DTXSID3047891
CASRN:2081-08-5
TOXCAST:33/276



Bisphenol P
DTXSID:DTXSID0058693
CASRN:2167-51-3
TOXCAST:0

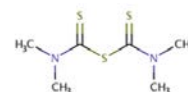


Tetraakis(hydroxymethyl)phosphonium su...
DTXSID:DTXSID0021331
CASRN:55566-30-8

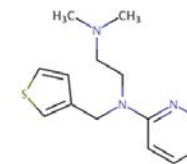


Scopolamine hydrobromide trihydrate
DTXSID:DTXSID0021258
CASRN:6533-68-2

KEMI List of Substances on the Market

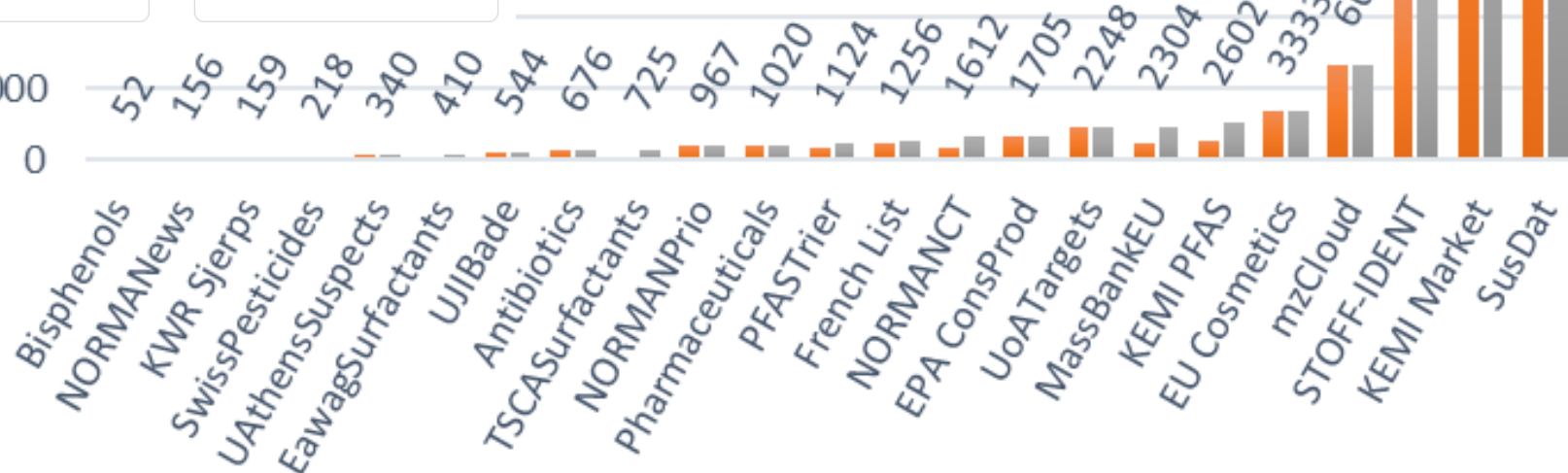


Tetramethylthiuram monosulfide
DTXSID:DTXSID0021333
CASRN:97-74-5
TOXCAST:74/453



Thenyldiamine
DTXSID:DTXSID0021335
CASRN:91-79-2
TOXCAST:0

5000
0



- We aggregate data from various sources
- Downloading is “easy”
- Databasing is “easy”
- Data curation is a laborious and painful task

EPI Suite Data - ISIS/Base & SDF

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

... *Updated September 15, 2010*

Basic Instructions:

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

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

Substructure Searching Files:

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

... [Click here to download EPI_ISIS_Data.zip](#) ... (about 11 MB)

... [Click here to download EPI_SDF_Data.zip](#) ... (about 10 MB)

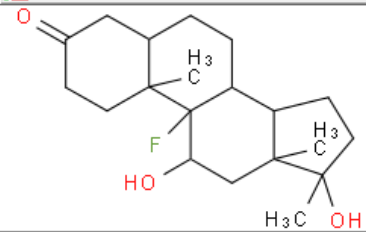
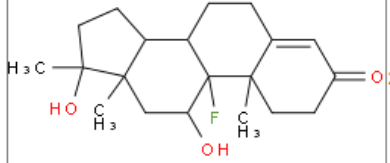
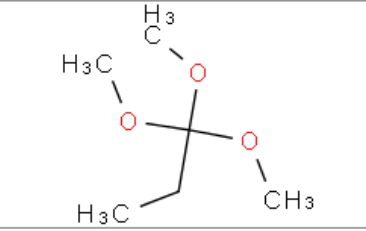
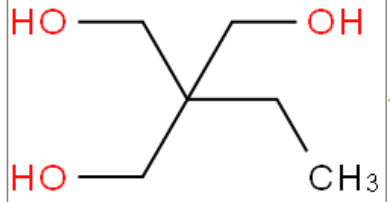
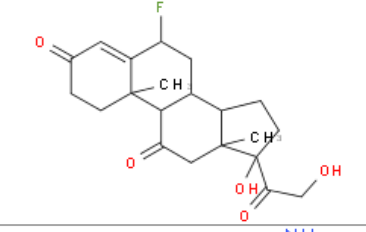
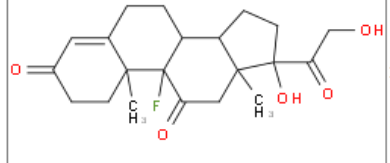
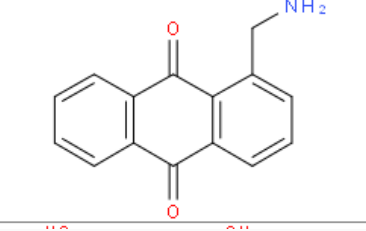
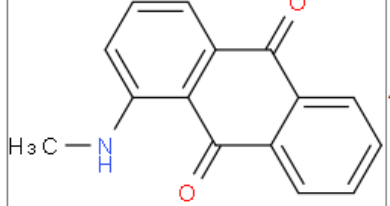
NOTE ... EPI Suite Data Files (some in Excel, Text, Word format) available at:

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

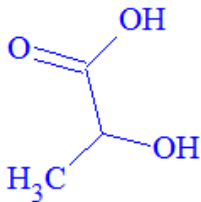
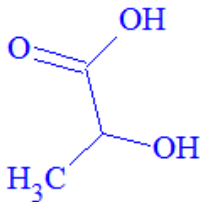
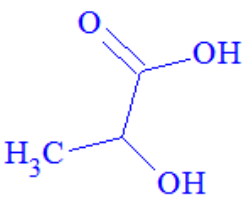
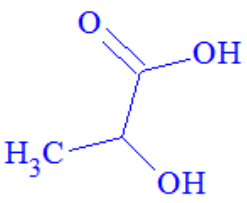
We Curated These Public Data to Build Prediction Models

Public data should be curated prior to modeling

Different Compounds

Mol Block	S CAS	S NAME	Smiles
	000076-43-7	FLUOXYMESTERONE	
	000077-99-6	1,1,1-TRIS(HYDROXYMETHYL)PROPANE	
	000079-60-7	CORTISONE-9A-FLUORO	
	000082-38-2	DISPERSE RED 9	

Duplicate Structures

Structure	Formula <	FW <	CAS <	NAME <	MP <	EstMP <	ErrorMP <
	C ₃ H ₆ O ₃	90.0779	000050-21-5	LACTIC ACID	1.6800000000000000e+001	2.2660000000000000e+001	5.8600000000000000e+000
	C ₃ H ₆ O ₃	90.0779	000079-33-4	L-LACTIC ACID	5.3000000000000000e+001	2.2660000000000000e+001	-3.0340000000000000e+001
	C ₃ H ₆ O ₃	90.0779	000598-82-3	A-HYDROXYPROPIONIC ACID	1.8000000000000000e+001	2.2660000000000000e+001	4.6600000000000000e+000
	C ₃ H ₆ O ₃	90.0779	010326-41-7	D-LACTIC ACID	5.2800000000000000e+001	2.2660000000000000e+001	-3.0140000000000000e+001

Covalent Halogens

Mol Block	S CAS	S NAME	Smiles
<p>Chemical structure of Benzyltrimethylammonium chloride. A central nitrogen atom (N⁺) is bonded to three methyl groups (H₃C) and a benzyl group (CH₂-C₆H₅). A chloride ion (Cl⁻) is shown as a counterion.</p>	000056-93-9	BENZYL TRIMETHYL AMMONIUM CHLORIDE	<p>Chemical structure of Benzyltrimethylammonium chloride. A central nitrogen atom (N) is bonded to three methyl groups (CH₃) and a benzyl group (CH₂-C₆H₅). A chloride ion (Cl) is shown as a counterion.</p>
<p>Chemical structure of Tetraethylammonium iodide. A central nitrogen atom (N⁺) is bonded to four ethyl groups (CH₃-CH₂). An iodide ion (I⁻) is shown as a counterion.</p>	000068-05-3	TETRAETHYL AMMONIUM IODIDE	<p>Chemical structure of Tetraethylammonium iodide. A central nitrogen atom (N) is bonded to four ethyl groups (CH₃-CH₂). An iodide ion (I) is shown as a counterion.</p>
<p>Chemical structure of Tetraethylammonium bromide. A central nitrogen atom (N⁺) is bonded to four ethyl groups (CH₃-CH₂). A bromide ion (Br⁻) is shown as a counterion.</p>	000071-91-0	TETRAETHYL AMMONIUM BROMIDE	<p>Chemical structure of Tetraethylammonium bromide. A central nitrogen atom (N) is bonded to four ethyl groups (CH₃-CH₂). A bromide ion (Br) is shown as a counterion.</p>

Curation to QSAR Ready Files

Property	Initial file	Curated Data	Curated QSAR ready
AOP	818	818	745
BCF	685	618	608
BioHC	175	151	150
Biowin	1265	1196	1171
BP	5890	5591	5436
HL	1829	1758	1711
KM	631	548	541
KOA	308	277	270
LogP	15809	14544	14041
MP	10051	9120	8656
PC	788	750	735
VP	3037	2840	2716
WF	5764	5076	4836
WS	2348	2046	2010

LogP dataset: 15,809 structures

- CAS Checksum: 12163 valid, 3646 invalid (>23%)
- Invalid names: 555
- Invalid SMILES 133
- Valence errors: 322 Molfile, 3782 SMILES (>24%)
- Duplicates check:
 - 31 DUPLICATE MOLFILES
 - 626 DUPLICATE SMILES
 - 531 DUPLICATE NAMES
- SMILES vs. Molfiles (structure check)
 - 1279 differ in stereochemistry (~8%)
 - 362 “Covalent Halogens”
 - 191 differ as tautomers
 - 436 are different compounds (~3%)



Journal

SAR and QSAR in Environmental Research >

Volume 27, 2016 - Issue 11: 17th International Conference on QSAR in Environmental and Health Sciences (QSAR 2016) - Part II. Guest Editors: C.G. Barber and G.J. Myatt

Enter keywords, authors, DOI etc.

258

Views

4

CrossRef citations

16

Altmetric

Articles

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

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

Pages 911-937 | Received 03 Sep 2016, Accepted 24 Oct 2016, Published online: 25 Nov 2016


Download citation

<http://dx.doi.org/10.1080/1062936X.2016.1253611>



OPERA Models: <https://github.com/kmansouri/OPERA>

Sourcing Chemical Hazard Data

 **United States
Environmental Protection
Agency**

Home Advanced Search Batch Search Lists ▼ Predictions Downloads

Copy ▼ Share ▼ Submit Comment

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES


SYNONYMS


▶ LITERATURE



LINKS

COMMENTS











Data Type












 Point of Departure ▼

 Download ▼

 Human  Eco

Columns ▼ 10 ▼

More	Priority	Toxval type	Subtype	Risk assessment class	Value	Units	Study type	Exposure route	Species	Subsource	Source
	5	BMDL-10	-	chronic	0.609	mg/kg-day	human	-	mouse	EFSA CEF	EFSA
	5	NOEL	Systemic	repeat dose	3.75	mg/kg-day	repeat dose toxicity : oral	oral	rat	-	ECHA
	6	NOAEL	-	reproductive	3.75	mg/kg-day	reproductive	oral	rat	-	HPVIS
	5	NOEL	Systemic	repeat dose	3.75	mg/kg-day	repeat dose toxicity : oral	oral	rat	-	ECHA
	5	NOEL	Systemic	repeat dose	4.5	mg/kg-day	repeat dose toxicity : oral	oral	mouse	-	ECHA
	5	NOEL	Systemic	repeat dose	4.5	mg/kg-day	repeat dose toxicity : oral	oral	mouse	-	ECHA
	7	LEL	-	subchronic	5	mg/kg-day	subchronic	oral	rat	unpublished_submission	ToxRefDB
	7	nel	-	chronic	5	mg/kg-day	reproductive multigeneration	oral	rat	open_lit	ToxRefDB
	5	NOAEL	-	chronic	5	mg/kg-day	human	-	mouse	EFSA AFC	EFSA
	7	nel	-	subchronic	5	mg/kg-day	subchronic	oral	rat	unpublished_submission	ToxRefDB

   1    4  5  6  7  

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

How can we curate our data?

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



- Submitted comments are reviewed by administrators and responded to

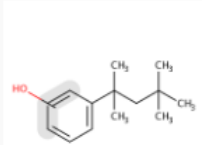
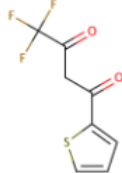
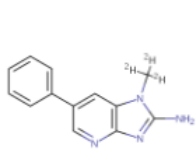
Public Crowdsourced Comments

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

Crowdsourced Comments

Show entries

Search:

Chemical	Structure	Date	Comment	Status
(1,1,3,3-Tetramethylbutyl)phenol		2017-07-15	Octylphenol redirects here, yet the name and related chemicals are 1,1,3,3-tetramethylbutylphenol - which is only a subset of all octylphenol isomers? Is this CAS only for these alkyl isomers?	★
1,3-Butanedione, 4,4,4-trifluoro-1-(2-thienyl)-		2017-03-30	Synonym: TTFA (Any way to bank these reCAPTCHAs so I don't have to do it everytime?)	★
1-(²H<sup>3</sup>Methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-amine		2017-05-06	1-(2H3Methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-amine 210049-13-1 DTXSID70670097 contains an error in the empirical formula due to an error in the deuterium representation and subsequent counting	★

Reviewer comments are public



3-[(3-Hydroxypropyl)disulfanyl]propanoic acid

663199-00-6 | DTXSID30792587

Searched by Approved Name.

[DETAILS](#)[EXECUTIVE SUMMARY](#)[PROPERTIES](#)[ENV. FATE/TRANSPORT](#)[HAZARD](#)[ADME](#)

The CAS is not correct. it is for Propanoic acid, 3-[(3-hydroxypropyl)dithio]-
[User comment posted about 1 year ago](#)

The CAS appears to match the structure shown of 3-[(3-Hydroxypropyl)disulfanyl]propanoic acid comparing with multiple other public sites, and with the name you supplied?
[Admin reply posted about 1 year ago](#)


- The majority of comments to date:
 - Structure and names/CASRN do not match
 - Add additional synonyms
 - Request to add specific property data
 - Structure layout/depiction needs improving

Crowdsourcing Comments


Single Cell Commenting added

- Highlight an alphanumeric text string

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



Crowdsourcing Comments




Details to be submitted with your comment:


Text selected: 149.999

Found On: August 11th 2018, 10:30:02 pm
Original Query: /dsstoxdb/results?search=BPA#toxicity-values
Browser: Chrome 68

There appears to be a rounding error in this ToxVal data

williams.antony@epa.gov


 I'm not a robot


reCAPTCHA
[Privacy](#) - [Terms](#)

Submit

In Vitro Bioassay Screening

ToxCast and Tox21

United States
Environmental Protection
Agency

HomeAdvanced SearchBatch SearchLists ▼PredictionsDownloads

Copy ▼Share ▼Submit CommentSearch all data

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

EXPOSURE

BIOACTIVITY

TOXCAST: SUMMARY

PUBCHEM

TOXCAST: DATA

TOXCAST: MODELS

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

LITERATURE

LINKS

COMMENTS

Assay Selection 1 Selected

☒ Active ☐ Inactive ☐ All

Filter

Filter assays

Assay Set: ER (1 of 18 Selected)

☒ ACEA_T47D_80hr_Positive

☐ ATG_ERE_CIS_up

☐ ATG_ERa_TRANS_up

☐ NVS_NR_bER

☐ NVS_NR_hER

☐ NVS_NR_mERa

☐ OT_ER_ERaERa_0480

☐ OT_ER_ERaERa_1440

☐ OT_ER_ERaERb_0480

☐ OT_ER_ERaERb_1440

☐ OT_ER_ERbERb_0480

☐ OT_ER_ERbERb_1440

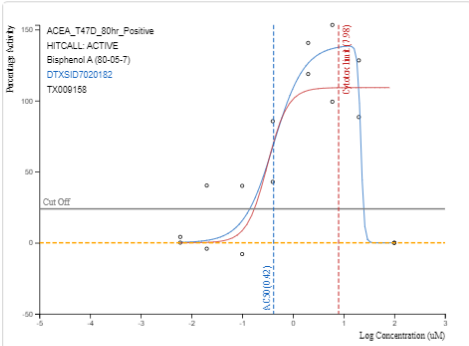
☐ OT_ERa_EREGFP_0120

☐ OT_ERa_EREGFP_0480

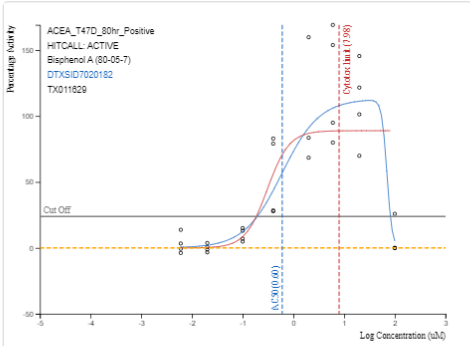
A Single Assay Can Have Multiple Charts

Number of Charts: 6

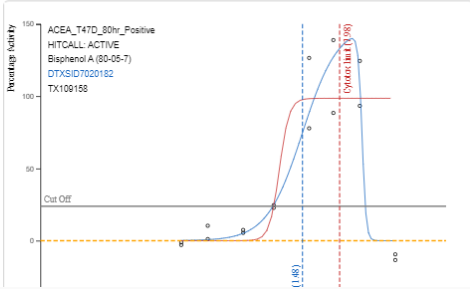
ACEA_T47D_80hr_Positive
HITCALL: ACTIVE
Bisphenol A (80-05-7)
DTXSID7020182
TX000158



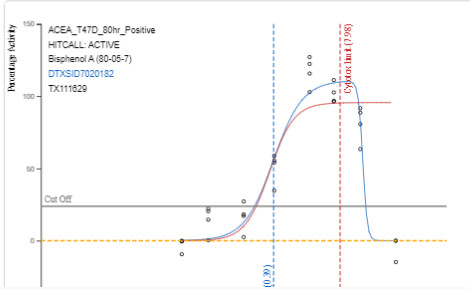
ACEA_T47D_80hr_Positive
HITCALL: ACTIVE
Bisphenol A (80-05-7)
DTXSID7020182
TX011529



ACEA_T47D_80hr_Positive
HITCALL: ACTIVE
Bisphenol A (80-05-7)
DTXSID7020182
TX100158

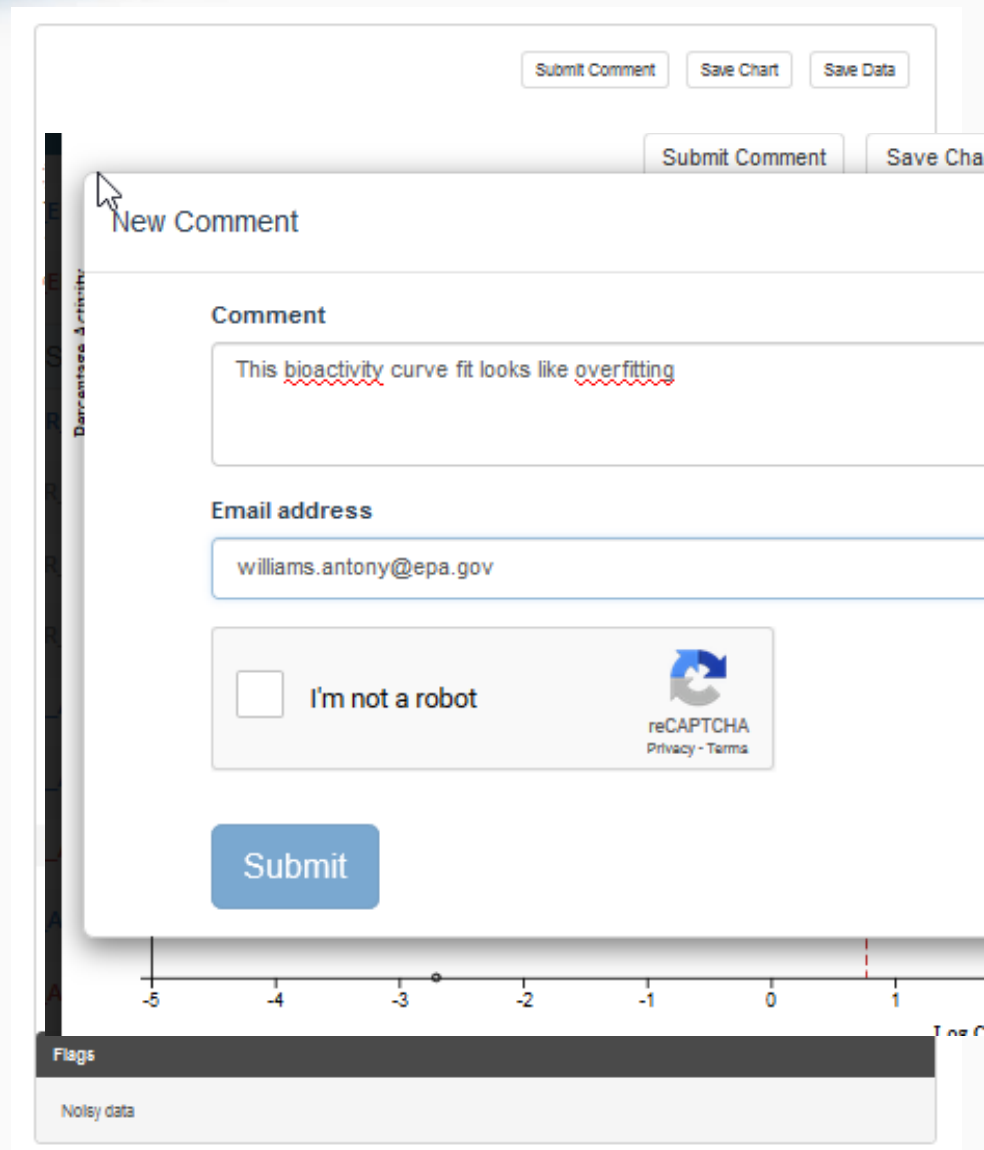


ACEA_T47D_80hr_Positive
HITCALL: ACTIVE
Bisphenol A (80-05-7)
DTXSID7020182
TX111529



Bioactivity Data

- 100s of thousands of bioactivity curves to review
- Impossible to review every one manually
- Now accepting public Crowdsourced Comments
- **Public crowdsourcing will not suffice!!!**



Submit Comment Save Chart Save Data

Submit Comment Save Chart

New Comment

Comment

This bioactivity curve fit looks like overfitting

Email address

williams.antony@epa.gov

☐ I'm not a robot

reCAPTCHA
Privacy - Terms

Submit

-5 -4 -3 -2 -1 0 1

Flags

Noisy data

Internal Review of 25,000 curves

Curve Review

[Download Flag Legend](#)[Download Fit Category Tree](#)

QA/QC Phase 1 Progress 19216 / 24633

[QA/QC](#)[Browse](#)[Batch](#)[Leaderboard](#)

Reviewer

Number of Curves to review

10

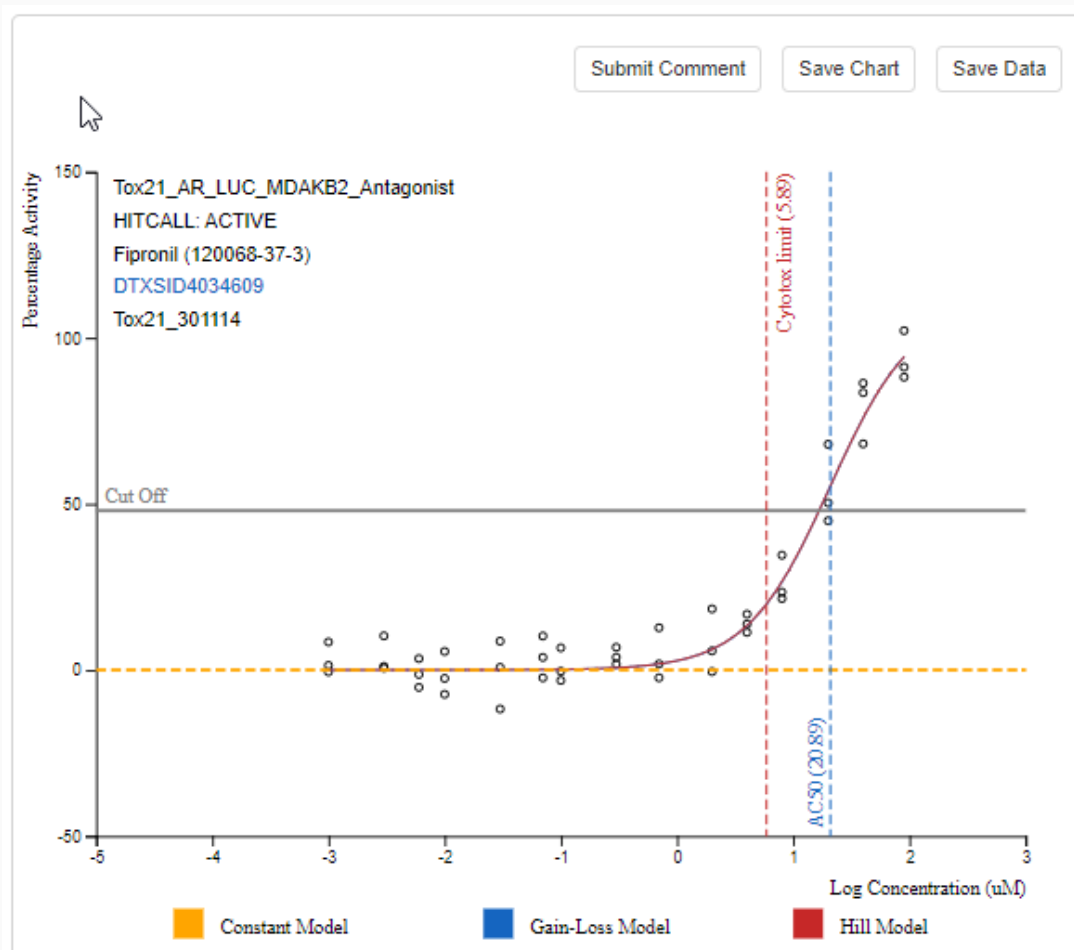
☒ Exclude curves that have already been reviewed

[Generate Curves For Review](#)

**Screenshot of entry page for Beta R Shiny Application for NCCT users
Brown & Paul-Friedman**

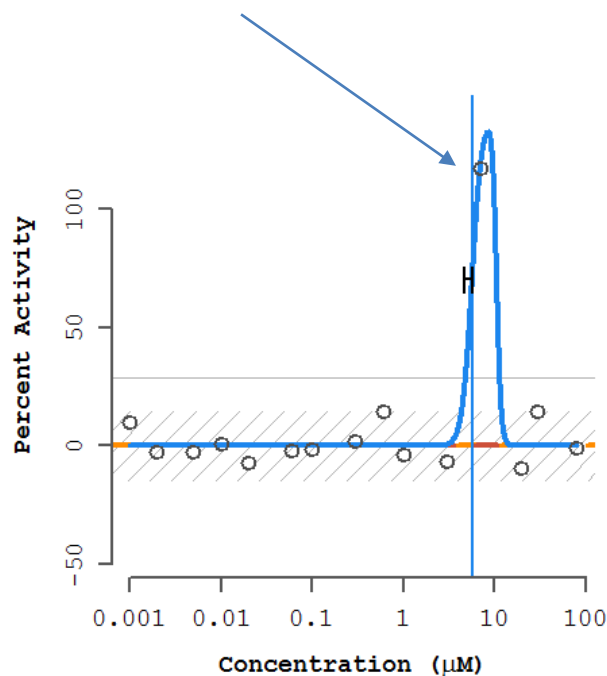
Internal Review of 25,000 curves

A “good fit” bioactivity curve



Internal Review of 25,000 curves

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



ASSAY: TOX21_RT_HEK293_FLO_40hr_viability

NAME: Tributylmethylammonium bis(trifluoromethylsulfo

CHID: 47905 CASRN: 405514-94-5

SPID(S): Tox21_200582

M4ID: 19917619

HILL MODEL (in red):

	tp	ga	gw
val:	104	2.4	7.97
sd:	NaN	NaN	NaN

GAIN-LOSS MODEL (in blue):

	tp	ga	gw	la	lw
val:	140	0.758	8	1.03	18
sd:	56.9	0.105	3.92	0.822	64.2

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

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

MAX_MEAN: 117 MAX_MED: 117 BMAD: 4.76

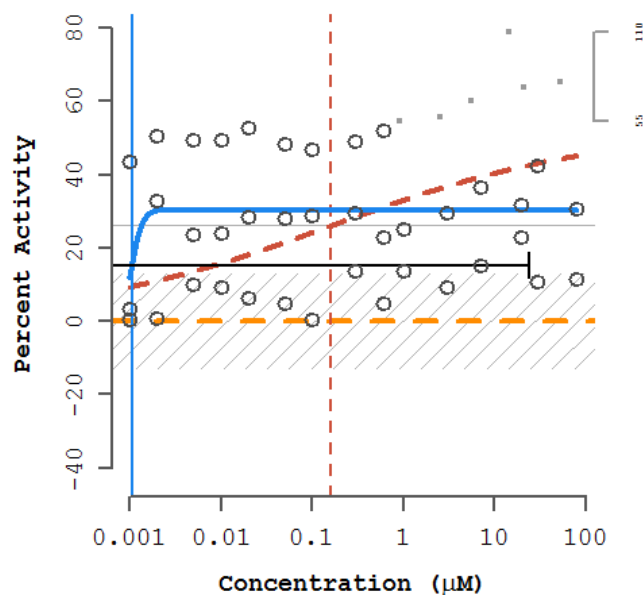
COFF: 28.5 HIT-CALL: 1 FITC: 50 ACTP: 1

FLAGS:

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

Internal Review of 25,000 curves

Abnormally High-Noise



ASSAY: TOX21_AR_BLA_Antagonist_ratio

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

HILL MODEL (in red):

	tp	ga	gw
val:	52	-0.797	0.3
sd:	NaN	NaN	NaN

GAIN-LOSS MODEL (in blue):

	tp	ga	gw	la	lw
val:	30.4	-2.97	7.76	3.77	4.44
sd:	3.58	0.211	59.3	7180	16800

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

MAX_MEAN: 55.2 MAX_MED: 42.4 BMAD: 4.35

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

FLAGS: 17; 11

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

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

tcpl: the ToxCast pipeline for high-throughput screening data

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

Bioinformatics, Volume 33, Issue 4, 15 February 2017, Pages 618–620,

<https://doi.org/10.1093/bioinformatics/btw680>

Published: 22 November 2016 **Article history** ▼

Chemical structures as text

Table 1. Metabolites found in the different incubations tested

Name	RT	m/z	Formula	m/z Diff (ppm)	Mass score	SMILES
Parent	3.13	455.2926	C ₂₇ H ₃₈ N ₂ O ₄	-3.48		<chem>N#CC(CCCN(C)CCc1ccc(OC)c(OC)c1)(C(C)C)c2ccc(OC)c(OC)c2</chem>
M6 -164	2.26	291.2077	C ₁₇ H ₂₆ N ₂ O ₂	-1.37	429	<chem>N(C)CCCC(C#N)(C(C)C)c1ccc(OC)c(OC)c1</chem>
M16 -14	3.06	441.2743	C ₂₆ H ₃₆ N ₂ O ₄	2.41	534	<chem>c1cc(CCNCCCC(C#N)(C(C)C)c2ccc(OC)c(OC)c2)cc(OC)c1OC</chem>
M14 +16	2.92	471.2866	C ₂₇ H ₃₈ N ₂ O ₅	-1.59	476	<chem>N#CC(CCCN(C)CC(O)c1ccc(OC)c(OC)c1)(C(C)C)c2ccc(OC)c(OC)c2</chem>
M9 -14	2.78	441.2761	C ₂₆ H ₃₆ N ₂ O ₄	-1.7	590	<chem>C(#N)C(CCCN(C)CCc1ccc(OC)c(OC)c1)(C(C)C)c2ccc(O)c(OC)c2</chem>
M11 -14	2.84	441.2742	C ₂₆ H ₃₆ N ₂ O ₄	2.57	473	<chem>Oc1ccc(CCN(C)CCCC(C#N)(C(C)C)c2ccc(OC)c(OC)c2)cc1OC</chem>
M12 +2	2.87	457.2707	C ₂₆ H ₃₆ N ₂ O ₅	-0.93	570	<chem>O(C)c1cc(ccc1OC)C(C#N)(CCCNCC(O)c2ccc(OC)c(OC)c2)C(C)C</chem>
M5 -178	2.2	277.1894	C ₁₆ H ₂₄ N ₂ O ₂	7.84	419	<chem>C(C)(C)C(C#N)(CCCN)c1ccc(OC)c(OC)c1</chem>
M8 +2	2.67	457.2708	C ₂₆ H ₃₆ N ₂ O ₅	-1.18	581	<chem>OC(CN(C)CCCC(C#N)(C(C)C)c1ccc(OC)c(OC)c1)c2ccc(O)c(OC)c2</chem>
M15 -14	2.92	441.2743	C ₂₆ H ₃₆ N ₂ O ₄	2.23	614	<chem>Oc1ccc(CCN(C)CCCC(C#N)(C(C)C)c2ccc(OC)c(OC)c2)cc1OC</chem>
M2 -259	0.73	196.1326	C ₁₁ H ₁₇ NO ₂	5.91	618	<chem>c1(CCNC)ccc(OC)c(c1)OC</chem>
M10 -28	2.8	427.2617	C ₂₅ H ₃₄ N ₂ O ₄	-4.79	487	<chem>c1(OC)cc(ccc1OC)C(C#N)(CCCNCCc2ccc(O)c(OC)c2)C(C)C</chem>
M7 +2	2.46	457.2717	C ₂₆ H ₃₆ N ₂ O ₅	-3.23	492	<chem>c1cc(CCN(O)CCCC(C#N)(C(C)C)c2ccc(OC)c(OC)c2)cc(OC)c1OC</chem>
M17 +16	3.21	471.2853	C ₂₇ H ₃₈ N ₂ O ₅	1.19	534	<chem>N#CC(CCCN(C)CCc1ccc(OC)c(OC)c1)(c2ccc(OC)c(OC)c2)C(C)(C)O</chem>
M4 -178	1.86	277.1927	C ₁₆ H ₂₄ N ₂ O ₂	-3.8	444	<chem>COc1cc(ccc1O)C(C#N)(CCCN)C(C)C</chem>
M1 -289	0.44	166.0858	C ₉ H ₁₁ NO ₂	5.93	136	<chem>c1(CCNC)ccc(=O)c(c1)=O</chem>
M13 -16	2.88	439.2603	C ₂₆ H ₃₄ N ₂ O ₄	-1.43	367	<chem>c1cc(CC=NCCCC(C#N)(C(C)C)c2ccc(OC)c(OC)c2)cc(OC)c1OC</chem>

How do I extract structures?

- Copy-Paste doesn't work

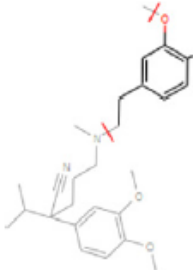
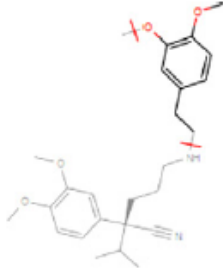
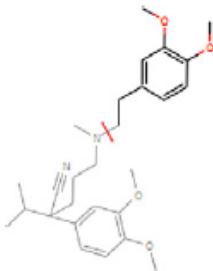
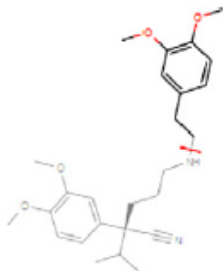
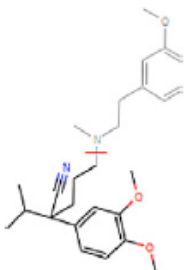
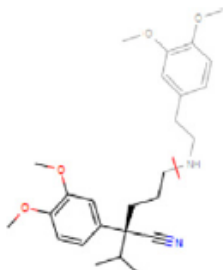
<chem>c1(CCNC)ccc(=O)c(c1)=O</chem>
<chem>c1cc(CC=NCCCC(C#N)(C(C)C)c2ccc(OC)c(OC)c2)cc(OC)c1OC</chem>

```
c1(CCNC)ccc( O)c(c1) O
c1cc(CC NCCCC(C#N)(C(C)C)c2ccc(OC)c(OC)c2)cc(OC)c1OC
```

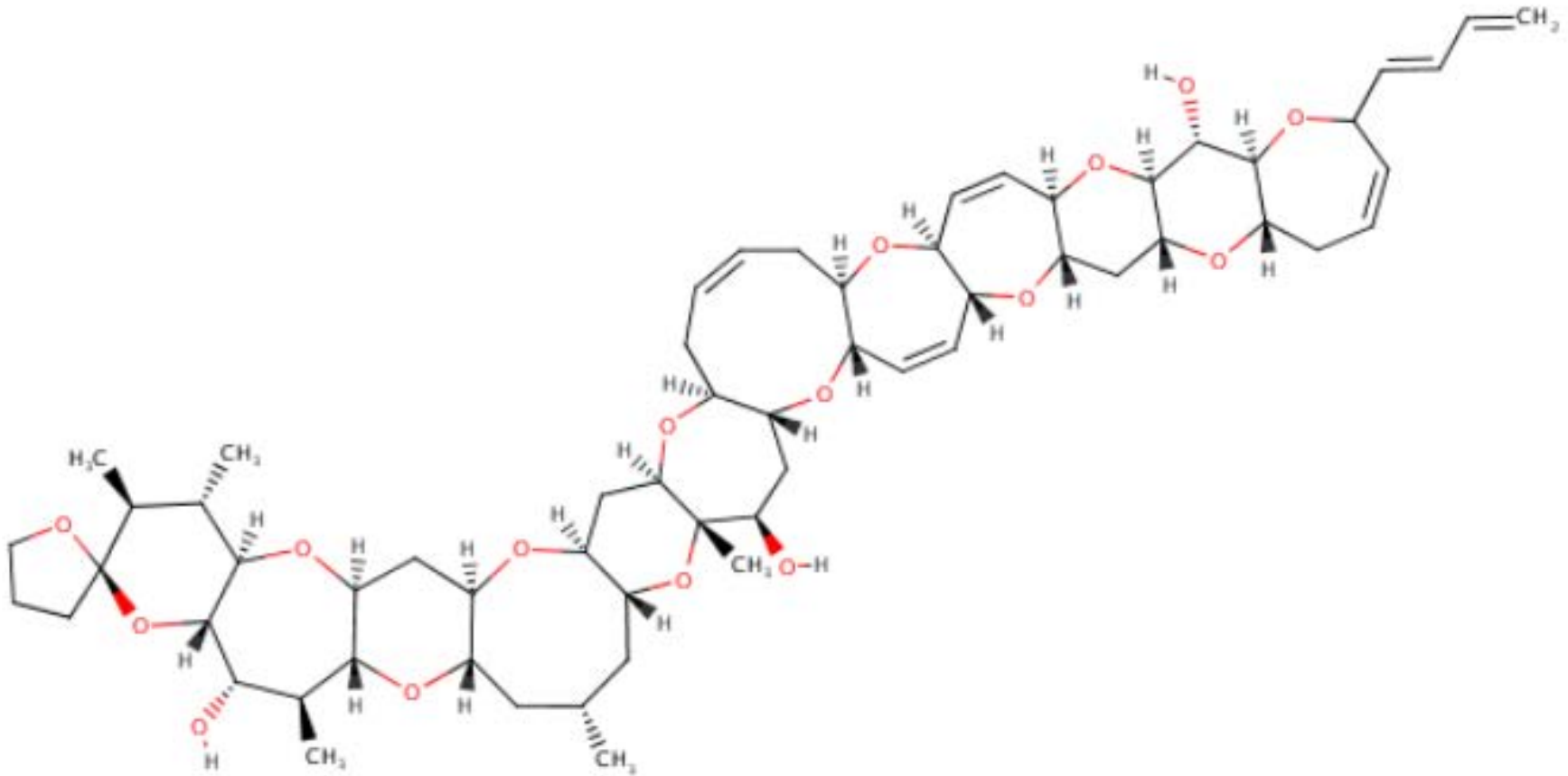
- This is not the way a publisher should deliver chemistry. But this is on the AUTHOR!

Extract Structure Drawings???

Table 2. Selection of fragments that help in the M16-16 metabolite structure elucidation

Sub. obs. m/z	Sub. cal. m/z	Sub. m/z diff. ppm	Substrate	Metabolite	Δ	Met. obs. m/z	Met. calc. m/z	Met. m/z diff. ppm
150.0664	150.0681	11.42			+0	150.0670	150.0681	7.25
165.0869	165.0916	28.22			+0	165.0892	165.0916	14.30
260.1637	260.1651	5.33			+0	260.1652	260.1651	-0.50

Try hand-drawing Algal Toxins!



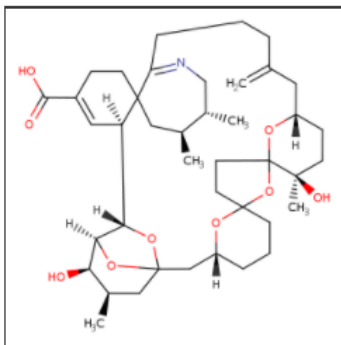
Think of files in multiple formats!

- SMILES are hyper-dependent on good layout algorithms. It's not easy!

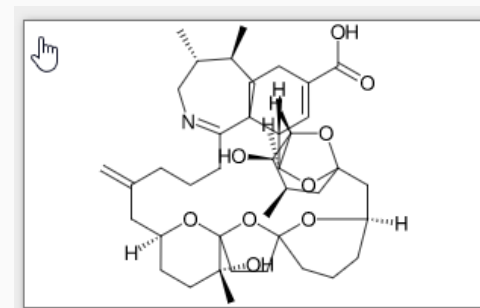
Pinnatoxin A

160759-36-4 | DTXSID40880101

© Searched by DSSTox_Substance_Id: Found 1 result for



[H][C@]12OC3(C[C@H](C)[C@H]1O)C[C@@]1([H])CCCC4(CC5(O4)O[C@]([H])(CC[C@]5(C)O)CC(=C)CCCC4=NC[C@H](C)[C@H](C)CC4C(C)C(=C[C@]4([H]))[C@]2([H])O3)C(O)=O)O1



Names and CASRNs are **NOT** structures

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

Attachment D (Method 3)

SIM quantitation ions and qualifiers for internal standards, references method analysis, and surrogates

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

And generally problematic...

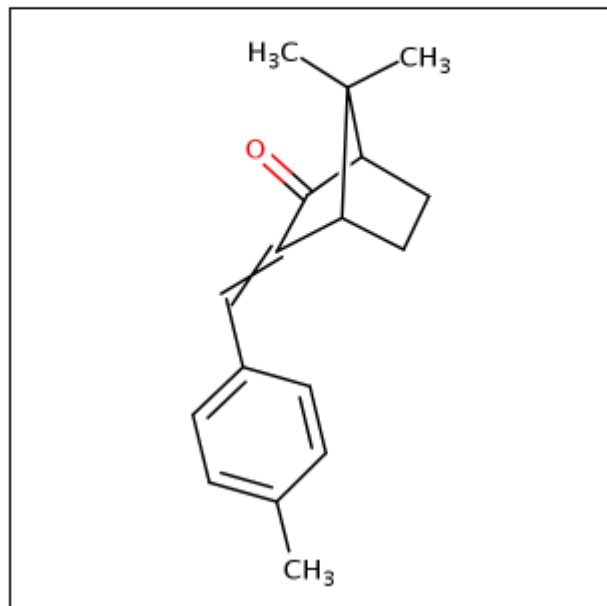
<u>Name of Compound</u>	<u>CAS No.</u>
Phenol-d6 (SS)	13187-88-3
Phenol	108-95-2
1,4-Dichlorobenzene	106-46-0
Acetophenone	98-86-2
Acenaphthene-d10 (IS)	15067-26-2
p-Cresol	106-44-5
Isophorone	78-59-1
Camphor	76-22-2
Isoborneol	124-76-5
Menthol	89, 78, 1
Naphthalene	91-20-3
Methyl salicilate	119-36-8

Active vs Deleted CASRN

Enzacamene

36861-47-9 | DTXSID8047896

Searched by Approved Name.



Synonym	Quality
Enzacamene	Valid
7,7-Dimethyl-3-[(4-methylphenyl)methylidene]bicyclo[2.2.1]heptan-2-one	Valid
Bicyclo[2.2.1]heptan-2-one, 7,7-dimethyl-3-[(4-methylphenyl)methylene]-	Valid
36861-47-9 Active CA 8-RN	Valid
Bicyclo[2.2.1]heptan-2-one, 1,7,7-trimethyl-3-[(4-methylphenyl)methylene]-	Valid
EINECS 253-242-8	Other
Eusolex 6300	Other
Uvinul MBC 95	Other
Parsol 5000	Other

UNII-8I3XWY40L9	Other
4-Methylbenzylidenecamphor	Other
p-Methylbenzylidenecamphor	Other
38102-62-4 Deleted CA 8-RN	Deleted
84055-85-2 Deleted CA 8-RN	Deleted

Tricky mapping by CASRN

This one has **316** Deleted CASRN

CAS Registry Number: 25068-38-6

(C₁₅ H₁₆ O₂ · C₃ H₅ Cl O)_x

Phenol, 4,4'-(1-methylethylidene)bis-, polymer with 2-(chloromethyl)oxirane


Polymer

Polymer Class Terms: Epoxy resin

Alternate CAS Registry Numbers: 26402-79-9

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

PLEASE use our curated data

 **EPA** United States
Environmental Protection
Agency

Home Advanced Search Batch Search Lists ▾ Predictions Downloads

Share ▾

Batch Search?

Step 1

Step 2

Step 3

Step 4

Step 5

Step 6

Step One: Select Input

Please enter one identifier per line ×

Select Input Type(s)

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

Enter Identifiers to Search (searches should be limited to <5000 identifiers)

<https://comptox.epa.gov/dashboard> Discover. Connect. Ask.

- What has changed?
 - Cheminformatics has progressed
 - The internet proliferates data access
 - Standards have progressed (InChI) and are improving
 - Anybody can access/download millions of structures!
- What hasn't changed?
 - Minor progress presenting structures in publications
 - Delivery via PDFs still dominates
 - Mandates on scientists are very unlikely

- The CompTox Dashboard provides access to data for ~760,000 chemicals
- Sourced and aggregated data from multiple resources
- Data quality is very random, source-to-source.
- We work hard to curate data - PLEASE USE IT!
- Our data is not perfect - we have a long way to go!

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>