

Data Review and Clean-Up Using Crowdsourced Input via the US EPA CompTox Dashboard

*Antony Williams¹, Katie Paul-Friedman¹, Jason Brown²,
Chris Grulke¹, Emma L. Schymanski³ and Jeff Edwards¹*

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

2) ORAU

3) 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*

- An introduction to the EPA's CompTox Dashboard
- Our Data: experimental and predicted property data, high-throughput screening assay data and hazard and environmental exposure data, and lots more
- Data quality: high-quality data needed but challenging to produce
 - Millions of individual data points and annotations
 - 10'0s of 1000's of chemicals
- What is the role of user feedback?
- Our efforts to curate our ToxCast bioassay data

CompTox Portal

Environmental Topics

Laws & Regulations

About EPA

Search EPA.gov



Discover.

Accessibility

EPA Administrator

Connect.

Data.gov

Inspector General

Ask.

Contact Us


Hotlines

- A publicly accessible website delivering access:
 - New entry portal for all NCCT dashboards
 - ~**762,000** chemicals with related property data
 - **Searchable by chemical, product use, gene and assay (ToxCast)**
 - Experimental and predicted physicochemical property data
 - **“Bioactivity data” for the ToxCast/Tox21 project**
 - **Generalized Read-Across (GenRA) module**
 - Links to other agency websites and public data resources
 - “Literature” searches for chemicals using public resources
 - “Batch searching” for thousands of chemicals
 - **DOWNLOADABLE** Open Data for reuse and repurposing

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.](#)



Discover.

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


Connect.

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

Ask.

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


CompTox Dashboard Chemicals


 United States
Environmental Protection
Agency

Home Advanced Search Batch Search Lists ▼ Predictions Downloads


Share ▼

762 Thousand Chemicals


 **Chemicals** Product/Use Categories Assay/Gene




Bisphenol A
DTXSID7020182



Bisphenol A bis(2-hydroxyethyl ether) diacrylate
DTXSID6066991




Bisphenol A bis(2-hydroxyethyl ether) dimethacrylate
DTXSID1066992




Bisphenol A bis(2-hydroxypropyl) ether
DTXSID8051592


Bisphenol A carbonate polymer
DTXSID6027840




Bisphenol A diglycidyl ether
DTXSID6024624



Bisphenol A glycidyl methacrylate
DTXSID7044841




Bisphenol A propoxylate diglycidyl ether
DTXSID10399098



Bisphenol A propoxylate glycerolate diacrylate
DTXSID40400126

comptox-prod.epa.gov/dashboard


CompTox Dashboard Products and Use Categories



United States
Environmental Protection
Agency

Home Advanced Search Batch Search Lists ▼ Predictions Downloads

Share ▼



762 Thousand Chemicals

Chemicals **Product/Use Categories** Assay/Gene

CPDat PRODUCT category: personal care hair color
hair colors and dyes characterized as permanent

CPDat PRODUCT category: personal care hair color
hair colors and dyes characterized as for professional use


CPDat PRODUCT category: personal care hair color
hair colors and dyes characterized as temporary

CPDat PRODUCT category: personal care hair color
hair coloring products not otherwise categorized

CPDat PRODUCT category: personal care hair color activator
chemical activators for hair coloring products

CPDat PRODUCT category: personal care hair color developer
chemical developers for hair coloring products

CPDat PRODUCT category: personal care hair color toner
chemical toners for hair coloring products



Discover.

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


Connect.

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

Ask.

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


CompTox Dashboard Assays and Genes



United States
Environmental Protection
Agency

[Home](#) [Advanced Search](#) [Batch Search](#) [Lists](#) [Predictions](#) [Downloads](#)

Share



762 Thousand Chemicals

[Chemicals](#) [Product/Use Categories](#) [Assay/Gene](#)

GENE: ESR1
estrogen receptor 1


GENE: ESR2
estrogen receptor 2 (ER beta)

GENE: ESRRB
estrogen-related receptor alpha

GENE: ESRRA
estrogen-related receptor beta

GENE: ESRRG
estrogen-related receptor gamma

and curating data, major updates to the batch searching functionality and access to real time predictions for both physiochemical and toxicity endpoints. [A list of release notes is available](#) for your review. We look forward to your feedback.



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

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

Ask.
[Contact](#)
[Help](#)

Detailed Chemical Pages

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

► ADME

► EXPOSURE

► BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

► LITERATURE

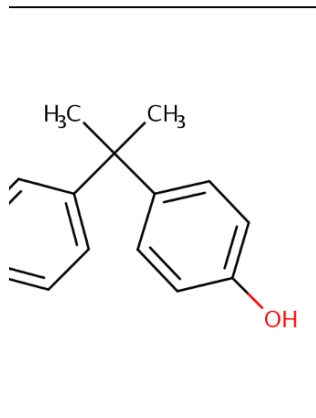
LINKS

COMMENTS

Bisphenol A

DTXSID7020182

3STox Substance Id.



Batch Search Lists Predictions Downloads

Copy

Share

Submit Comment

Search all data

Wikipedia

Bisphenol A (BPA) is an organic synthetic compound with the chemical formula $(\text{CH}_3)_2\text{C}(\text{C}_6\text{H}_4\text{OH})_2$ belonging to the group of diphenylmethane derivatives and bisphenols, with two hydroxyphenyl groups. It is a colorless solid that is soluble in organic solvents, but poorly soluble in water. It has been in commercial use since 1957.

BPA is a starting material for the synthesis of plastics, primarily

[Read more](#)

Intrinsic Properties

Structural Identifiers

Linked Substances

Presence in Lists

Record Information

Quality Control Notes

Physicochemical properties

Property



Summary



Summary

LogP: Octanol-Water

Melting Point

Boiling Point

Water Solubility

Vapor Pressure

Flash Point

Surface Tension

Index of Refraction

Molar Refractivity

Polarizability

Density

Molar Volume

Thermal Conductivity

Viscosity

Henry's Law


LogKoa: Octanol-Air

Summary

Search query

Average	Experimental median	Predicted median	Experimental range	Predicted range	Unit
		3.43	3.32	2.40 to 3.64	
	156	138	153 to 156	125 to 157	°C
		360	200	343 to 401	°C
		1.00e-3	5.26e-4	5.44e-4 to 1.31e-3	mol/L
		3.43e-7	-	6.83e-8 to 2.59e-6	mmHg
		190	-	188 to 192	°C
			-	46.0	dyn/cm
			-	1.60	
			-	68.2	cm ³
			-	27.0	Å ³
		1.17	-	1.14 to 1.20	g/cm ³
			-	200	cm ³
			-	150	mW/(m ² K)

Access to 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

In Vitro Bioassay Screening

ToxCast and Tox21

RELATED SUBSTANCES

SYNONYMS

► LITERATURE

LINKS

COMMENTS

Download ▼

Columns ▼ 10 ▼

Search query

Show Inactive

Show Background

Name	Modal	Description	SeqPASS	Gene Name	AOP	Event	Hit Call	Top	Scaled Top	AC50	logAC50	Intended Target Family
ACEA_T47D_80hr_Negative		-	-	-	-	-	ACTIVE	35.5	1.65	65.8	1.82	cell cycle
ACEA_T47D_80hr_Positive		2	NP_000116.2	ESR1	200	1181	ACTIVE	109	4.49	0.381	-0.419	nuclear receptor
APR_HepG2_CellLoss_24h_dn		-	-	-	-	-	ACTIVE	1.20	1.81	106	2.02	cell cycle
APR_HepG2_MitoMass_24h_dn		-	-	-	-	-	ACTIVE	0.874	1.76	109	2.04	cell morphology
APR_HepG2_MitoMembPot_24h_dn		-	-	-	-	-	ACTIVE	5.92	7.07	11.0	1.04	cell morphology
APR_HepG2_OxidativeStress_24h_up		-	-	-	-	-	ACTIVE	1.20	1.47	110	2.04	cell cycle
APR_HepG2_CellLoss_72h_dn		-	-	-	-	-	ACTIVE	4.49	5.05	95.2	1.98	cell cycle
APR_HepG2_MitoMembPot_72h_dn		-	-	-	-	-	ACTIVE	2.71	3.69	85.3	1.93	cell morphology
APR_HepG2_MitoticArrest_72h_up		-	-	-	-	-	ACTIVE	1.66	1.17	84.7	1.93	cell cycle
APR_HepG2_OxidativeStress_72h_up		-	-	-	-	-	ACTIVE	1.80	1.65	106	2.02	cell cycle

First << < 1 2 3 4 5 6 7 8 9 10 > >> Last

Showing 1 to 10 of 161 records

In Vitro Bioassay Screening

ToxCast and Tox21

EPA United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Copy Share Submit Comment Search 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

A Single Assay Can Have Multiple Charts

Number of Charts: 6

☒ 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

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

Percent Activity

Log Concentration (uM)

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

Percent Activity

Log Concentration (uM)

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

Percent Activity

Log Concentration (uM)

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

Percent Activity

Log Concentration (uM)

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

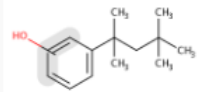
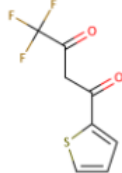
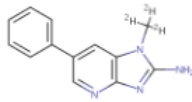
Public Crowdsourced Comments

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

Crowdsourced Comments

Show 10 entries

Search:

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

Reviewer comments are public



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

663199-00-6 | DTXSID30792587

Searched by Approved Name.

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

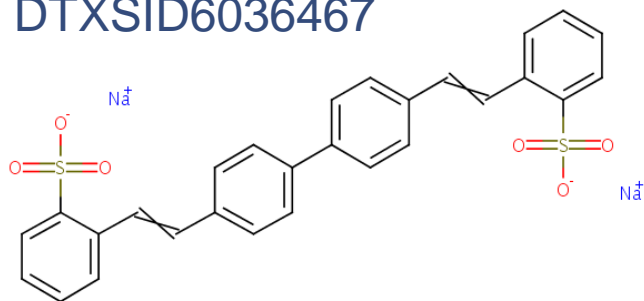
The CAS is not correct. it is for Propanoic acid, 3-[(3-hydroxypropyl)dithio]-
[User comment posted about 1 year ago](#)

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](#)

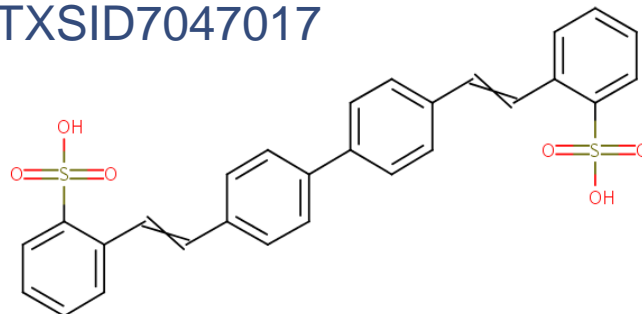
MassBank/CompTox Curation of External Data

○ A “nice” example: 4-4'-Bis(2-sulfostryl)biphenyl

Purchased: CAS: 27344-41-8
DTXSID6036467



Registered: CAS: 38775-22-3 (UFZ)
DTXSID7047017



Add A Comment

Comment from structure source: to my knowledge the stilbene-derived fluorescent whitening agents are all trans (E) isomers, as the cis (Z) isomers are not fluorescent (although they might undergo photo-isomerisation to the cis isomers under UV light, and clothing gets yellowish again then...). Thus I would consider the E,E form the correct thesized in a way that x % of the technical product are actually the Z,Z forms and thus the undefined stereo would be correct. In the Z (maybe also Z,Z) and thus both isomers occur, see: f8a CAS number on record was 27344-41-8 which is DTXSID6036467.

Related to comment just added to DTXSID7047017 - it would be useful for us to know if this CAS number is undefined stereochemistry (a technical mix) or defined ... we would update our records accordingly.

User comment posted 4 months ago

ed to the mixed E/Z-form of the chemical. The E/E form is 6036467 is ALSO the mixed EZ form

The CAS Number is indeed for the undefined stereochemistry form.

Admin reply posted 4 months ago

I have a stereo-defined version of this on record as Fluorescent brightener 351 c1cc(cc1)S(=O)(=O)O/C=C/c2ccc(cc2)c3ccc(cc3)/C=C/c4c(cccc4)S(=O)(=O)O SQAQVQFOMMLRPR-IWGRKNQJSA-N

User comment posted 4 months ago

There is NO indication of the E-defined stereochemistry in the registry and it appears to be a mix of isomers.

Admin reply posted 4 months 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

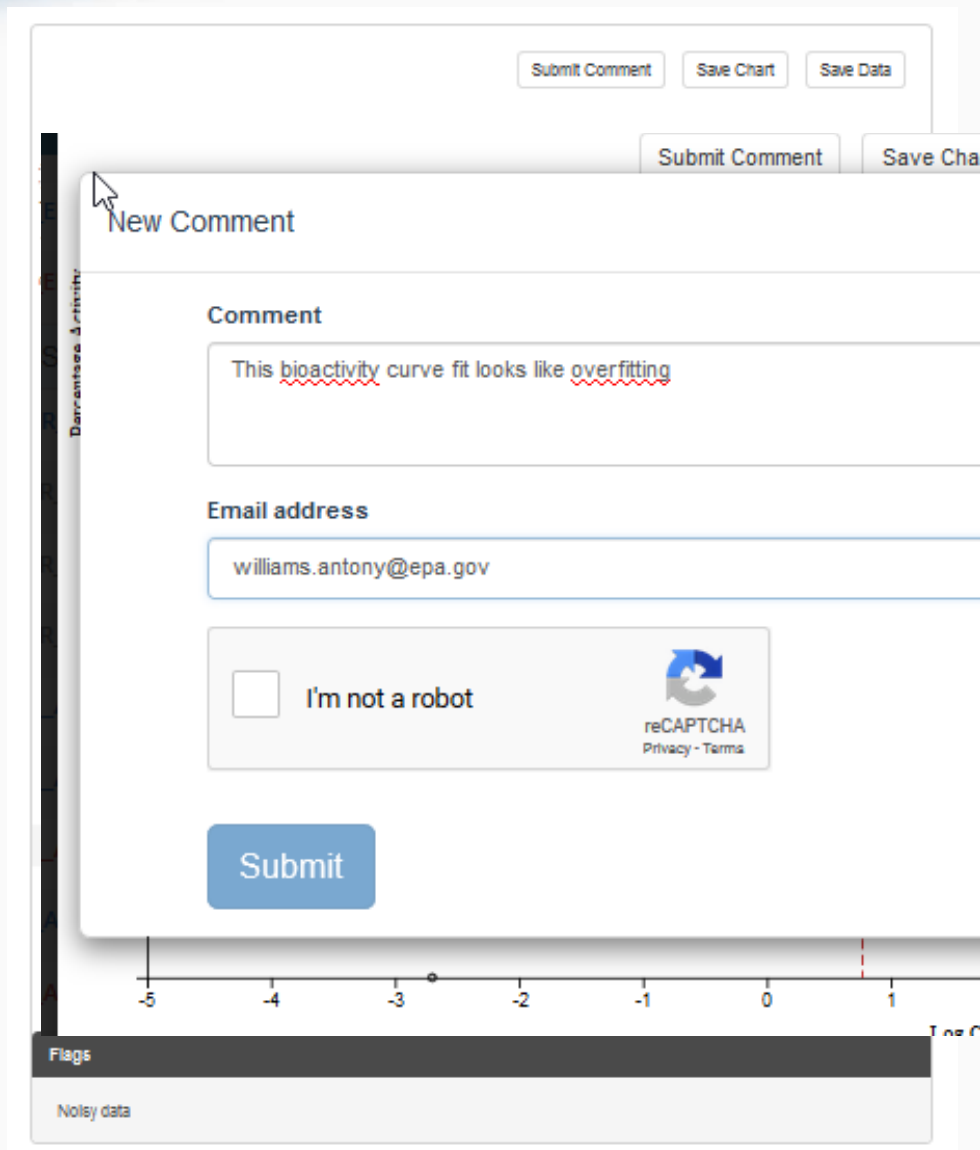
 I'm not a robot


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

Submit

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!!!



The screenshot displays a web interface for reviewing bioactivity data. At the top right, there are buttons for 'Submit Comment', 'Save Chart', and 'Save Data'. Below these, a 'New Comment' section contains a text area with the comment 'This bioactivity curve fit looks like overfitting', where 'bioactivity' and 'overfitting' are underlined in red. Below the comment is an 'Email address' field containing 'williams.antony@epa.gov'. A reCAPTCHA challenge is present with the text 'I'm not a robot' and a checkbox. A blue 'Submit' button is located below the reCAPTCHA. At the bottom, a 'Flags' section shows a dropdown menu with 'Noisy data' selected. In the background, a plot of 'Dose-Response Activity' is visible with a horizontal axis ranging from -5 to 1.

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

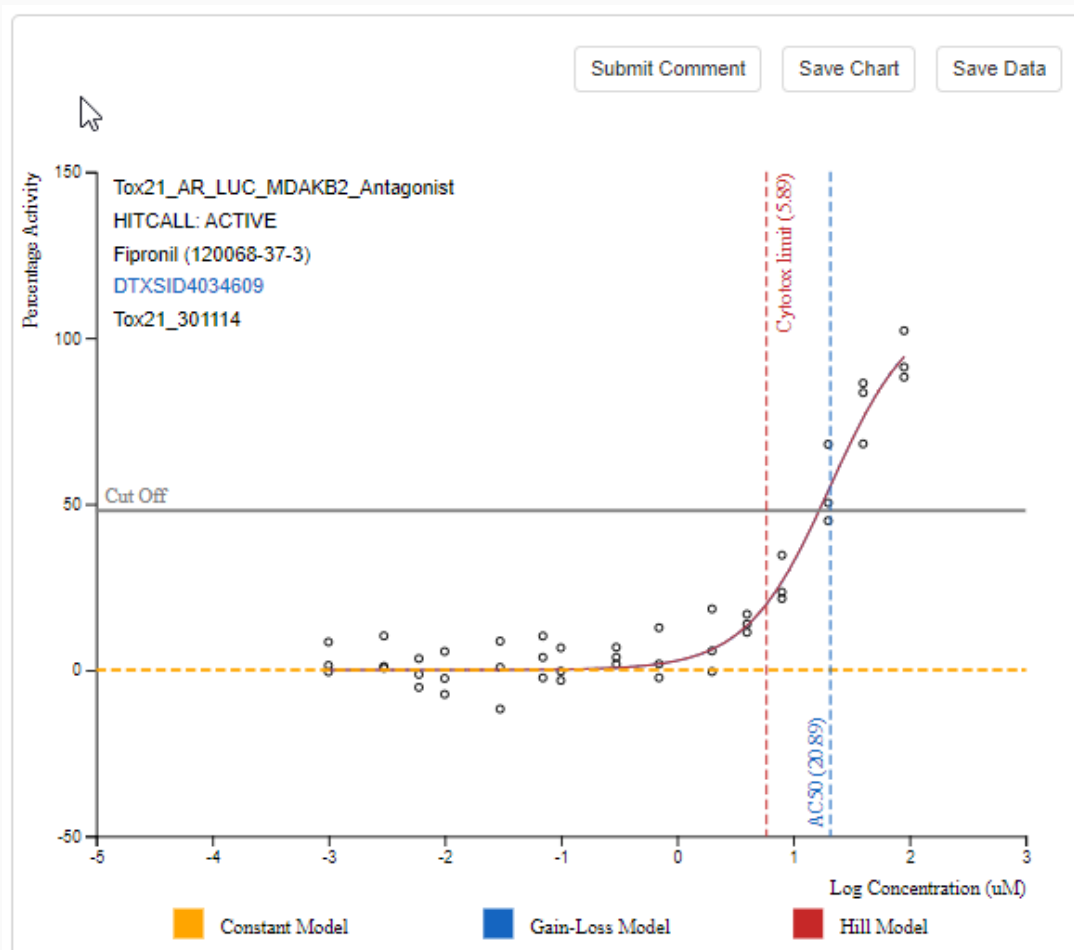
☒ Exclude curves that have already been reviewed

[Generate Curves For Review](#)

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

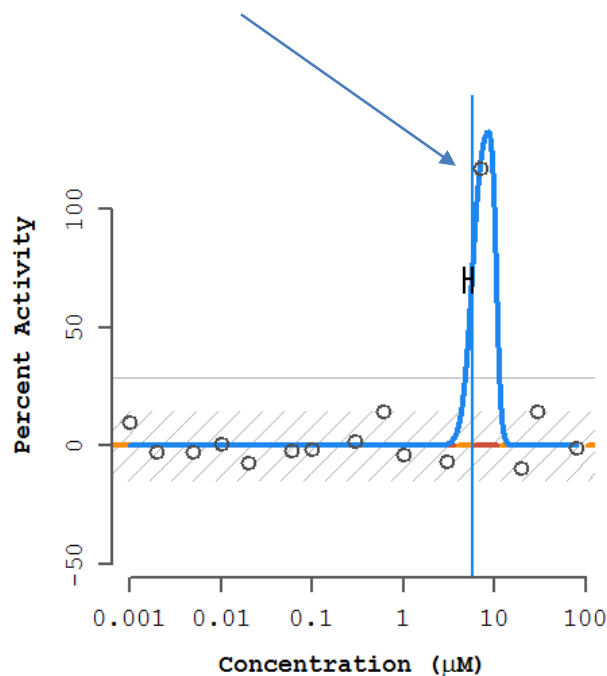
Internal Review of 25,000 curves

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

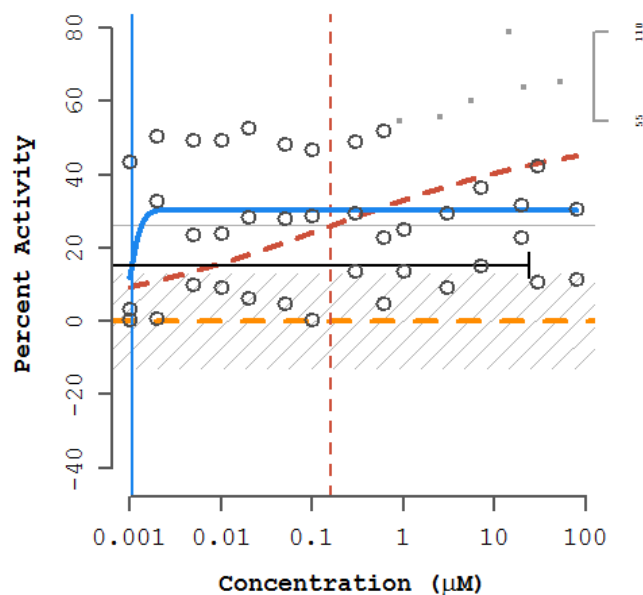
COFF: 28.5 HIT-CALL: 1 FITC: 50 ACTP: 1

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** ▼

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>
- Our process for mapping data is iterative

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)O)C(=O)O</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)O)C(=O)O</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	Tetraacetythylenediamine	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(C(=C2Br)Br)Br)Br)CC(Br)CC(Br)CC(Br)CC(Br)C1</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)c1ccccc1C(=O)OCCc2ccccc2</chem>
StNorman35	Diethyl phthalate	CAS_RN: 84-66-2	<chem>CCOC(=O)c1ccccc1C(=O)OCC</chem>
StNorman37	Di-n-butylphthalate	CAS_RN: 84-74-2	<chem>CCCCOC(=O)c1ccccc1C(=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 CHEMICAL NAME							
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-butyl-4-(2-cyano-2-methyl-5-oxo-5H-tetrazol-5-yl)phenol	CC(C)(C)C1=CC=C(C=C1C2=CC(=O)N3C#NC=N3)C4=CC(=O)N5C#NC=N54	C49H75N13O12	1037.5658
DTXSID00880086	Microcystin YR	101064-48-6	OWHASZC	(5R,8S,11R)-2,6-Di-tert-butyl-4-(2-cyano-2-methyl-5-oxo-5H-tetrazol-5-yl)phenol	CC(C)(C)C1=CC=C(C=C1C2=CC(=O)N3C#NC=N3)C4=CC(=O)N5C#NC=N54	C52H72N10O13	1044.5280
DTXSID6027052	2,6-Di-tert-butylphenol	128-39-2	DKCPKDPY	2,6-Di-tert-butylphenol	CC(C)(C)C1=CC=C(C=C1)C2=CC(=O)N3C#NC=N32	C14H22O	206.16706
DTXSID7020215	Butylated hydroxyanisole	25013-16-5	CZBZUDVE	2-tert-Butyl-4-methoxyphenol	COC1=CC=C(C=C1)C2=CC(=O)N3C#NC=N32	C22H32O4	360.23005
DTXSID6020220	tert-Butylhydroquinone	1948-33-0	BGNXCDM	2-tert-Butyl-1,4-dimethoxybenzene	CC(C)(C)C1=CC=C(C=C1)C2=CC(=O)N3C#NC=N32	C10H14O2	166.09937
DTXSID2020216	Butylated hydroxytoluene	128-37-0	NLZUEZXR	2,6-Di-tert-butyl-4-methylphenol	CC1=CC(=C(C=C1)C2=CC(=O)N3C#NC=N32)C4=CC(=O)N5C#NC=N54	C15H24O	220.18271
DTXSID2023434	Pentetic acid	67-43-6	QPCDCPD	N,N-Bis(2-hydroxy-5-oxo-5H-tetrazol-5-yl)-2,2',2''-nitroethane-1,1'-dicarboxylic acid	OC(=O)CN1C#NC=N1C2=CC(=O)N3C#NC=N32	C14H23N3O10	393.13834
DTXSID6022977	Ethylenediaminetetraacetic acid	60-00-4	KCXVZYZI	2,2',2'',2'''-nitroethane-1,1'-dicarboxylic acid	OC(=O)CN1C#NC=N1C2=CC(=O)N3C#NC=N32	C10H16N2O8	292.09066
DTXSID6020939	Nitrilotriacetic acid	139-13-9	MGFYIUFG	2,2',2''-nitroethane-1,1'-dicarboxylic acid	OC(=O)CN1C#NC=N1C2=CC(=O)N3C#NC=N32	C6H9NO6	191.04298
DTXSID2032631	Oxadixyl	77732-09-3	UWVQIRO	N-(2,6-Dimethoxyphenyl)-2,2,6,6-tetramethylpiperidine-1-oxide	COCC(=O)N1C#NC=N1C2=CC(=O)N3C#NC=N32	C14H18N2O4	278.12665
DTXSID5040752	Tetraacetylenehydrazide	10543-57-4	BGRWYDF	N,N'-(Ethane-1,1'-diyl)bis(2-oxo-2-azidoacetic acid)	CC(=O)N(C#N)C1=CC=C(C=C1)C2=CC(=O)N3C#NC=N32	C10H16N2O4	228.11100
DTXSID70147779	2-(2-(4-Nonylphenoxy)ethyl)ethyl acetate	106807-78-7	RAQHOB	[2-(4-Nonylphenoxy)ethyl]ethyl acetate	CCCCCCCCC1=CC=C(C=C1)C2=CC(=O)N3C#NC=N32	C19H30O4	322.21440
DTXSID60165003	Acetic acid, (4-octylphenyl)hydrazide	15234-85-2	DWUYSEM	(4-Octylphenyl)hydrazide	CCCCCCCCC1=CC=C(C=C1)C2=CC(=O)N3C#NC=N32	C16H24O3	264.17254
DTXSID0021549	Cyanoformaldehyde	4471-47-0	TUHMQD	Oxoacetic acid	O=CC#N	C2HNO	55.005813
DTXSID2052732	1,1'-Ethane-1,2-diylbis(4-bromobenzene)	84852-53-9	BZQKBFHE	1,1'-(Ethane-1,1'-diyl)bis(4-bromobenzene)	BrC1=C(Br)C=CC=C1	C14H4Br10	961.21468
DTXSID8025383	Hexabromocyclododecane	25637-99-4		-	-	-	-
DTXSID7021029	N-Nitrosodimethylamine	62-75-9	UMFJAH	N,N-Dimethylnitrosamine	CN(C)N=O	C2H6N2O	74.048012
DTXSID3020205	Benzyl butyl phthalate	85-68-7	IRIAEXOR	Benzyl butyl phthalate	CCCCOC(=O)C1=CC=C(C=C1)C2=CC(=O)N3C#NC=N32	C19H20O4	312.13615
DTXSID7021780	Diethyl phthalate	84-66-2	FLKPEMZC	Diethyl phthalate	CCOC(=O)C1=CC=C(C=C1)C2=CC(=O)N3C#NC=N32	C12H14O4	222.08920
DTXSID2021781	Dibutyl phthalate	84-74-2	DOIRQSBF	Dibutyl phthalate	CCCCOC(=O)C1=CC=C(C=C1)C2=CC(=O)N3C#NC=N32	C16H22O4	278.15180
DTXSID7020182	Bisphenol A	80-05-7	IISBACLA	4,4'-(Propylenebis(4-hydroxyphenyl))	CC(C)(C1=CC=C(C=C1)C2=CC(=O)N3C#NC=N32)	C15H16O2	228.11502

Mapping on Two Identifiers

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-dihydroxybenzene	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-hydroxyethyl)-N',N'-bis(2-hydroxypropyl)ethylenediamine	OC(=O)CN(CC(O)CC)CC(O)CC	C14H23N3O10	393.13834
DTXSID6022977	Ethylenediaminetetraacetic acid	60-00-4	KCXVZYZY	2,2',2'',2'''-Ethylenediaminetetraacetic acid	OC(=O)CN(CC(O)CC)CC(O)CC	C10H16N2O8	292.09066
DTXSID6020939	Nitrilotriacetic acid	139-13-9	MGFYIUZF	2,2',2''-Nitrilotriacetic acid	OC(=O)CN(CC(O)CC)CC(O)CC	C6H9NO6	191.04298
DTXSID2032631	Oxadixyl	77732-09-3	UWVQIRO	N-(2,6-Dimethyl-4-oxocyclohex-2-en-1-yl)-N',N'-bis(2-hydroxyethyl)ethylenediamine	COCC(=O)C1=CC=C(C=C1)O	C14H18N2O4	278.12665
DTXSID5040752	Tetraacetylenediamine	10543-57-4	BGRWYDF	N,N'-(Ethylenedioxy)bis(2-aminopropanoic acid)	CC(=O)N(CC(O)CC)CC(O)CC	C10H16N2O4	228.11100
DTXSID70147779	2-(2-(4-Nonylphenoxy)ethyl)propanoic acid	106807-78-7	RAQHOB	[2-(4-Nonylphenoxy)ethyl]propanoic acid	CCCCCCCCC1=CC=C(C=C1)OCC(=O)O	C19H30O4	322.21440
DTXSID60165003	Acetic acid, (4-octylphenyl) ester	15234-85-2	DWUYSEM	(4-Octylphenoxy)acetic acid	CCCCCCCCC1=CC=C(C=C1)OCC(=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(2-bromo-2-fluoroethyl)	84852-53-9	BZQKBFHE	1,1'-(Ethane-1,2-diyl)bis(2-bromo-2-fluoroethyl)	BrC1=C(Br)C(F)CC1	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=C(C=C1)OCC(=O)C2=CC=CC=C2	C19H20O4	312.13615
DTXSID7021780	Diethyl phthalate	84-66-2	FLKPEMZC	Diethyl phthalate	CCOC(=O)C1=CC=C(C=C1)OCC(=O)C2=CC=CC=C2	C12H14O4	222.08920
DTXSID2021781	Dibutyl phthalate	84-74-2	DOIRQSBF	Dibutyl phthalate	CCCCOC(=O)C1=CC=C(C=C1)OCC(=O)C2=CC=CC=C2	C16H22O4	278.15180
DTXSID7020182	Bisphenol A	80-05-7	IISBACLA	4,4'-(Prop-1-ene-2,3-diylidene)bis(phenol)	CC(C)(C1=CC=C(C=C1)O)C2=CC=CC=C2	C15H16O2	228.11502

Mapping Quality Control (I)

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
DTXSID40880085	DTXSID40880085	DTXSID40880085	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	0
DTXSID00880086	DTXSID00880086	DTXSID00880086	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	0
DTXSID6027052	DTXSID6027052	DTXSID6027052	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	0
DTXSID7020215	DTXSID7020215	DTXSID7020215	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	0
DTXSID6020220	DTXSID6020220	DTXSID6020220	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	0
DTXSID2020216	DTXSID2020216	DTXSID2020216	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	0
DTXSID2023434	DTXSID2023434	DTXSID2023434	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	0
DTXSID6022977	DTXSID6022977	DTXSID6022977	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	0
DTXSID6020939	DTXSID6020939	DTXSID6020939	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	0
DTXSID2032631	DTXSID2032631	DTXSID2032631	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	0
DTXSID5040752	DTXSID5040752	DTXSID5040752	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	0
DTXSID70147779	DTXSID70147779	DTXSID70147779	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	0
DTXSID60165003	DTXSID60165003	DTXSID60165003	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	0
DTXSID0021549	DTXSID0021549	DTXSID0021549	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	0
DTXSID2052732	DTXSID2052732	DTXSID2052732	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	0
DTXSID8025383	DTXSID8025383	DTXSID8025383	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	0
DTXSID7021029	DTXSID7021029	DTXSID7021029	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	0
DTXSID3020205	DTXSID3020205	DTXSID3020205	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	0
DTXSID7021780	DTXSID7021780	DTXSID7021780	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	0
DTXSID2021781	DTXSID2021781	DTXSID2021781	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	0

Mapping Quality Control (II)

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.
-----	-----------	-----------------------------	--	---	---



Home Advanced Search Batch Search Lists ▾ Predictions Downloads

Share ▾

Search all data

Select List

Show 10 ▾ entries

Download

Search: normanpri

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)

Previous 1 Next

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

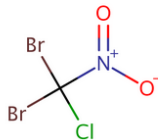
Download / Send ⌵

Show info: DTXSID ✕ CASRN ✕ TOXCAST ✕ ⌵

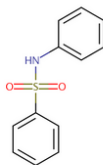
Select all 📄

Sort by: DTXSID ⌵ ⬆️

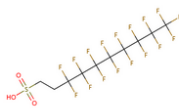
Filter by: Name or CASRN Hide ⌵



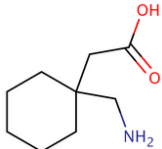
Dibromochloronitromethane
DTXSID:DTXSID00152114
CASRN:1184-89-0
TOXCAST:0



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

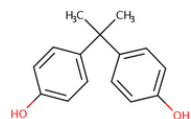


8:2 Fluorotelomer sulfonic acid
DTXSID:DTXSID00192353
CASRN:39108-34-4
TOXCAST:0

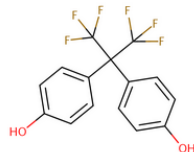


Gabapentin
DTXSID:DTXSID0020074
CASRN:60142-96-3
TOXCAST:0

>23 NORMAN Lists Available

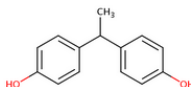


Bisphenol A
DTXSID:DTXSID7020182
CASRN:80-05-7
TOXCAST:166/812

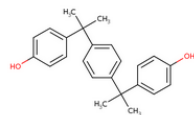


Bisphenol AF
DTXSID:DTXSID7037717
CASRN:1478-61-1
TOXCAST:248/639

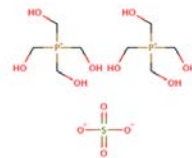
Bisphenol Compounds



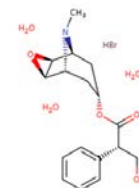
Bisphenol E
DTXSID:DTXSID3047891
CASRN:2081-08-5
TOXCAST:33/276



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

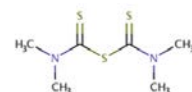


Tetraakis(hydroxymethyl)phosphonium su...
DTXSID:DTXSID0021331
CASRN:55566-30-8

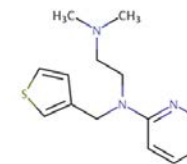


Scopolamine hydrobromide trihydrate
DTXSID:DTXSID0021258
CASRN:6533-68-2

KEMI List of Substances on the Market

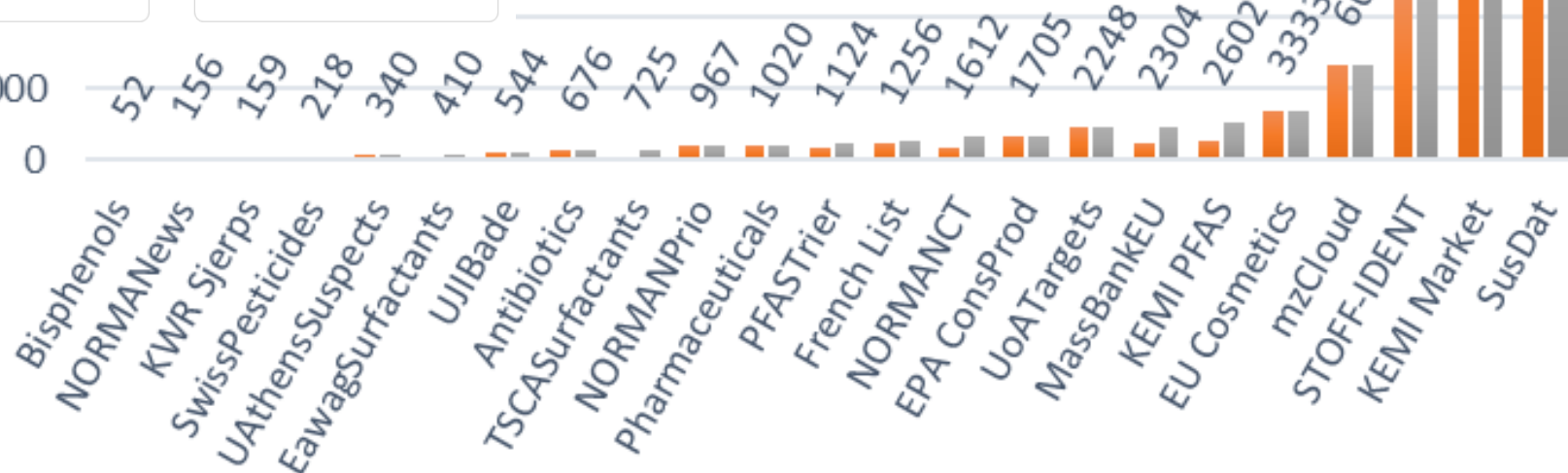


Tetramethylthiuram monosulfide
DTXSID:DTXSID0021333
CASRN:97-74-5
TOXCAST:74/453



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

5000
0



Progressive Curation on #Lists

Undefined mixtures (UVCBs)

Cleaning up lists to remove errors

Mol_ID	Name	EDITED NAMES FOR INPUT INTO SEARCH	CAS_RN	Merged DTXSiDs	DTXSiD Based on Name	Preferred Name
SA8750	By-Product	By-Product	NA	-	-	NO_MATCH
stpQQR1546	C10-DATS C10-Dialkyl tetra	C10-DATS C10-Dialkyl tetralin sulfonate 8	NA	-	-	NO_MATCH
SA2074	C10-LAS	C10-LAS	NA	-	-	NO_MATCH
stpQQR1582	C10LAS C10-linear alkylbe	C10LAS C10-linear alkylbenzyl sulfonate 4	NA	-	-	NO_MATCH
SA14931	C10-phosphonic	C10-phosphonic	NA	-	-	NO_MATCH
StpBB815	C12-15 ALKYL BENZOATE	C12-15 ALKYL BENZOATE	68411-27-8	-	-	NO_MATCH
SA13282	C12-AE5S	C12-AE5S	NA	-	-	NO_MATCH
stpQQR1548	C12-LAS C12-linear alkyl b	C12-LAS C12-linear alkyl benzene sulfonat	NA	-	-	NO_MATCH
stpQQR690	C14-SAS (TENTATIVE) tetra	C14-SAS (TENTATIVE) tetradecane-7-sulfo	NA	-	-	NO_MATCH
stpQQR1557	C16EOx C16EO2 C16-alcco	C16EOx C16EO2 C16-alcohol polyethoxyl	NA	-	-	NO_MATCH
stpQQR1556	C18EOx C18EO2 C18-alcco	C18EOx C18EO2 C18-alcohol polyethoxyl	4439-32-1	-	-	NO_MATCH
SA14932	C4-phosphonic	C4-phosphonic	NA	-	-	NO_MATCH
SA14929	C6-phosphonic	C6-phosphonic	NA	-	-	NO_MATCH
stpQQR1583	C7SPC C7-sulfophenyl car	C7SPC C7-sulfophenyl carboxylates 4-(de	NA	-	-	NO_MATCH
SA14930	C8-phosphonic	C8-phosphonic	NA	-	-	NO_MATCH
stpQQR1547	C8-SPC C8-Sulfophenyl ca	C8-SPC C8-Sulfophenyl carboxylic acid 4-(NA	-	-	NO_MATCH
stpQQR1576	CA5PE2C 7-{4-[2-(carboxy	CA5PE2C 7-{4-[2-(carboxymethoxy)ethoxy	NA	-	-	NO_MATCH
stpQQR1578	CA6PE2	CA6PE2	NA	-	-	NO_MATCH
stpQQR1577	CA6PE2C	CA6PE2C	NA	-	-	NO_MATCH
stpQQR1575	CA8PE2C	CA8PE2C	NA	-	-	NO_MATCH
SA9863	cacotheline	cacotheline	561-20-6	-	-	NO_MATCH
SAn15715	Caerulomycin A	Caerulomycin A	21802-37-9	-	-	NO_MATCH
SA5151	cafedrine	cafedrine	58166-83-9	-	-	NO_MATCH

(many) more registrations...

Conclusion

- The CompTox Dashboard provides access to data for ~760,000 chemicals
- Crowdsourced comments from users may help clean data
- Record level data curation - any alphanumeric string
- Improved data quality provides better data for modeling to underpin our prediction algorithms
- Collaborative data curation benefits extended communities of users
- Data curation is never complete – our data expands daily...

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>

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

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

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