http://www.orcid.org/0000-0002-2668-4821

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

*Antony Williams[1], Katie Paul-Friedman[1], Jason Brown[2], Chris Grulke[1], Emma L. Schymanski[3] and Jeff Edwards[1]*

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*

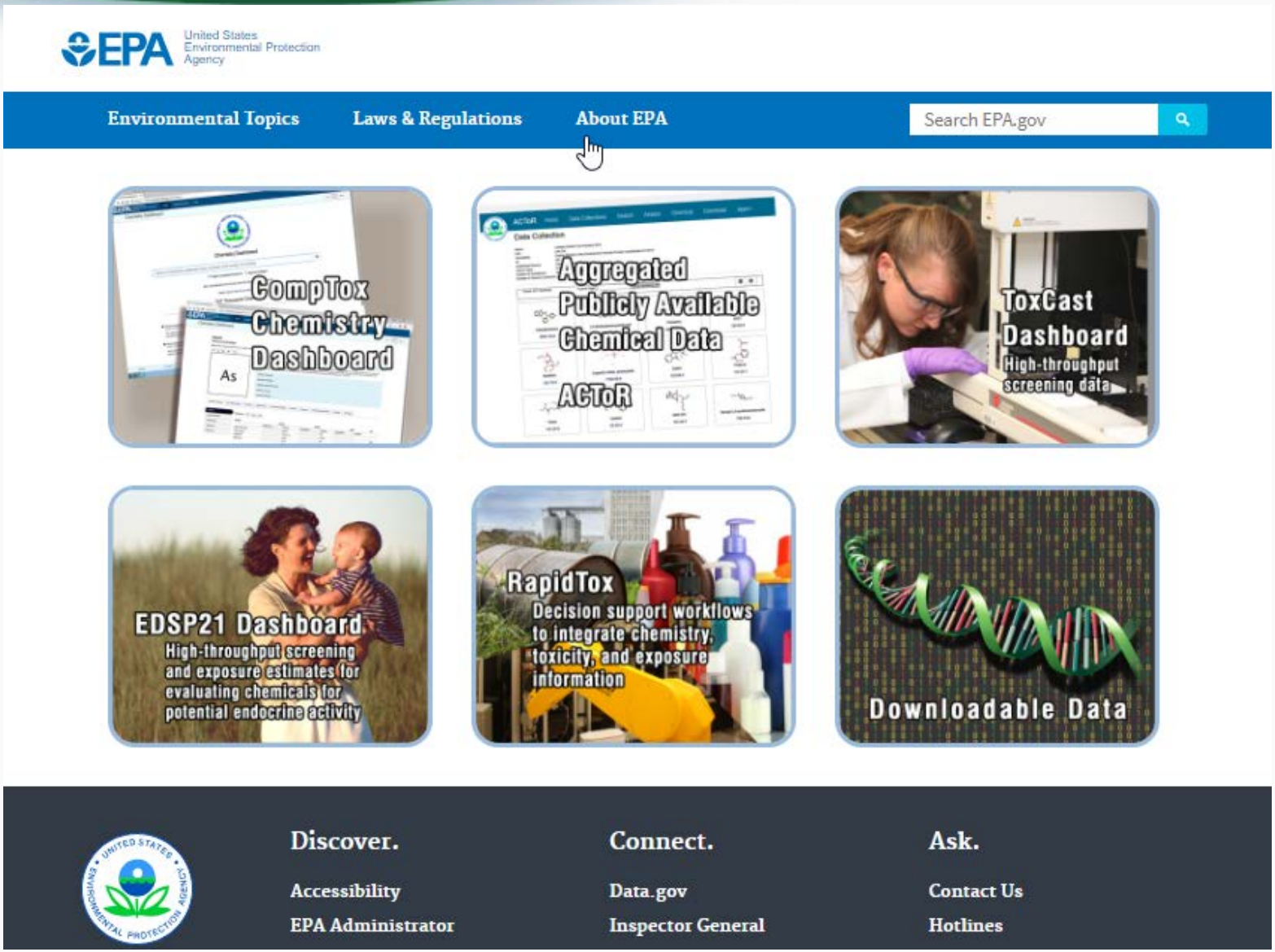


*August 2018*
*ACS Fall Meeting, Boston*

# Overview

- 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

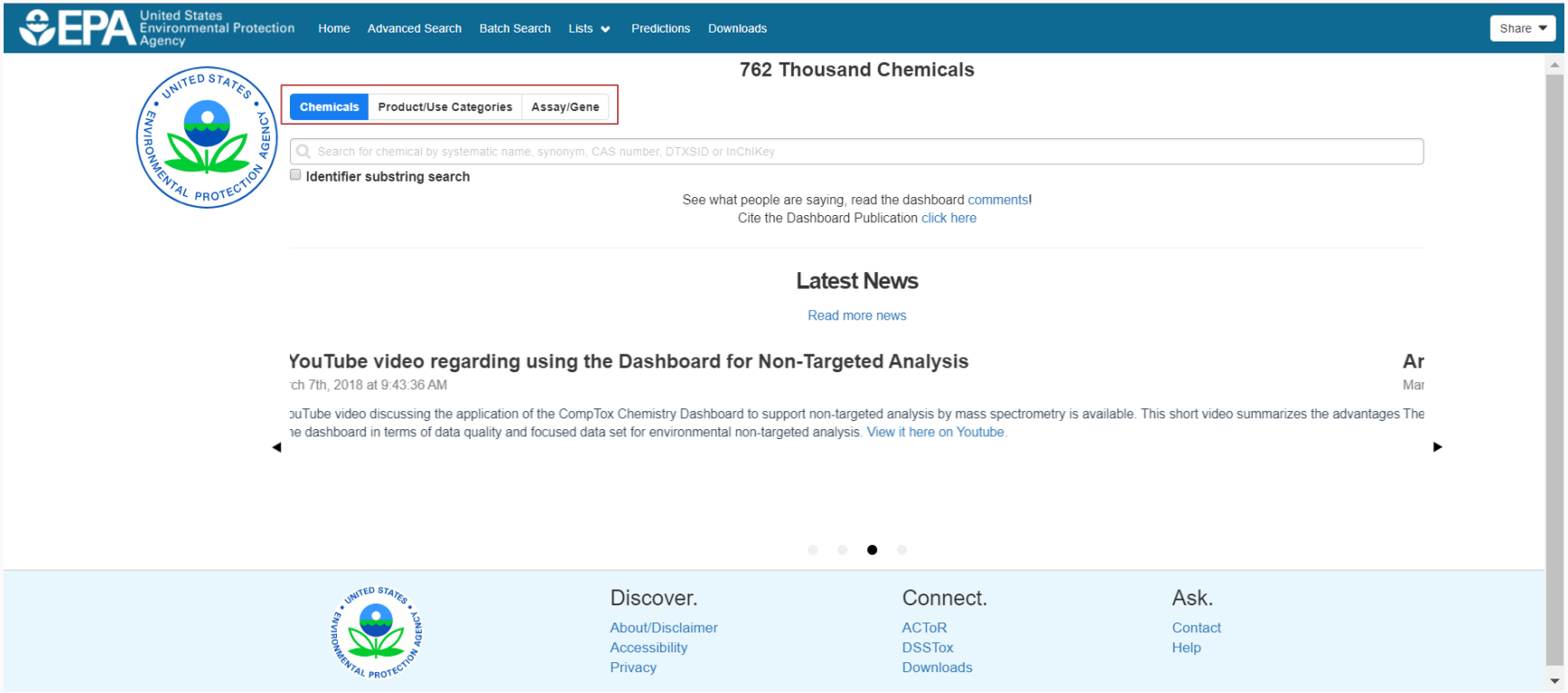


# The CompTox Dashboard

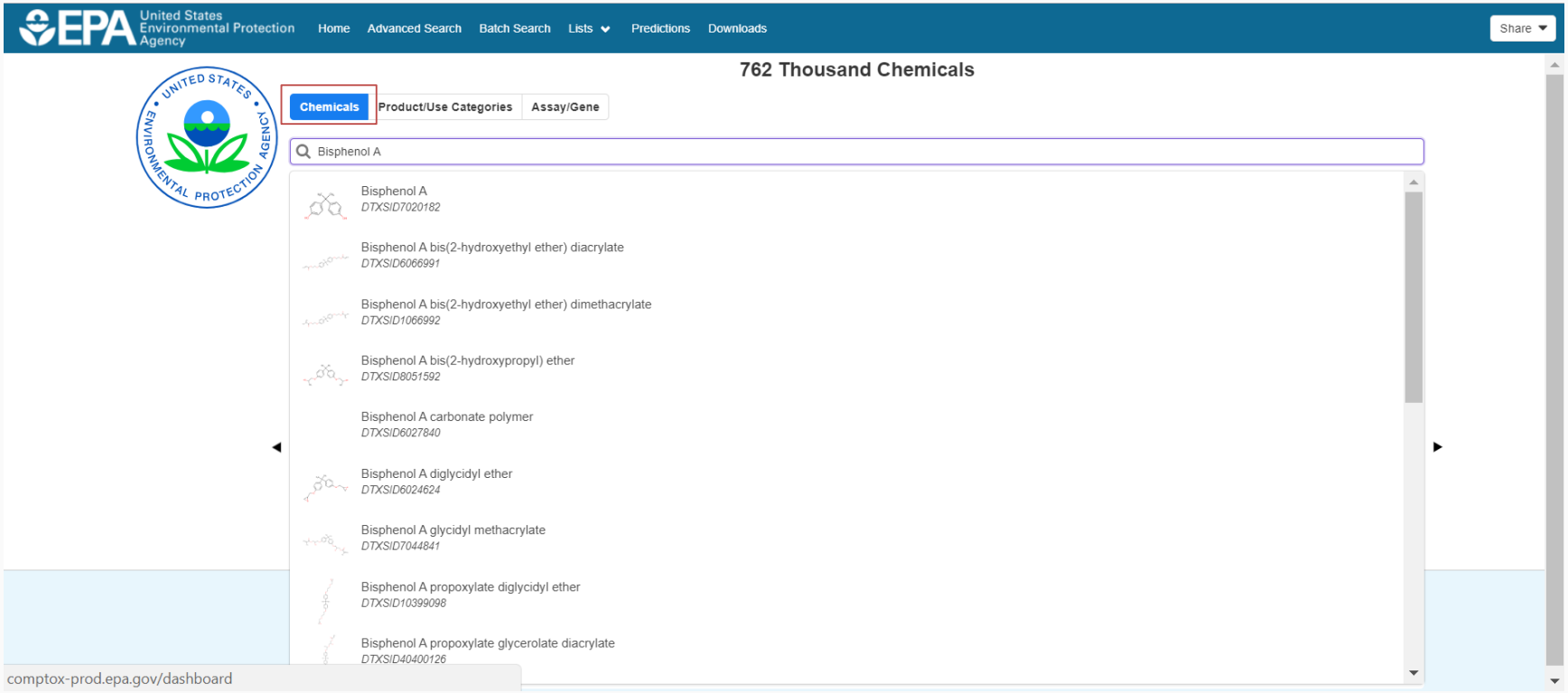
- A publicly accessible website delivering access:
  - New entry portal for all NCCT dashboards
  - ~**762,000** chemicals with related property data
  - **Searchable by chemical, product use, gene and assay (ToxCast)**
  - Experimental and predicted physicochemical property data
  - **“Bioactivity data” 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

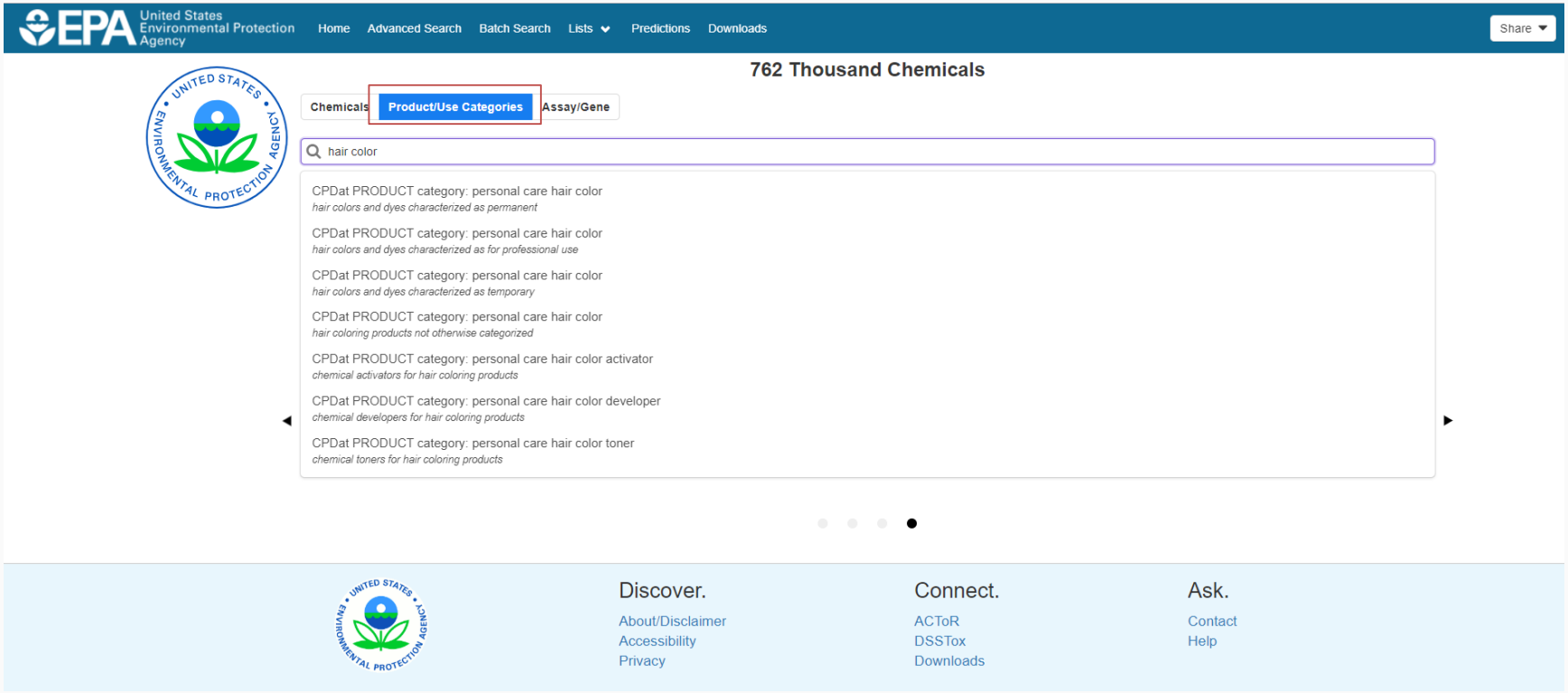
https://comptox.epa.gov/dashboard



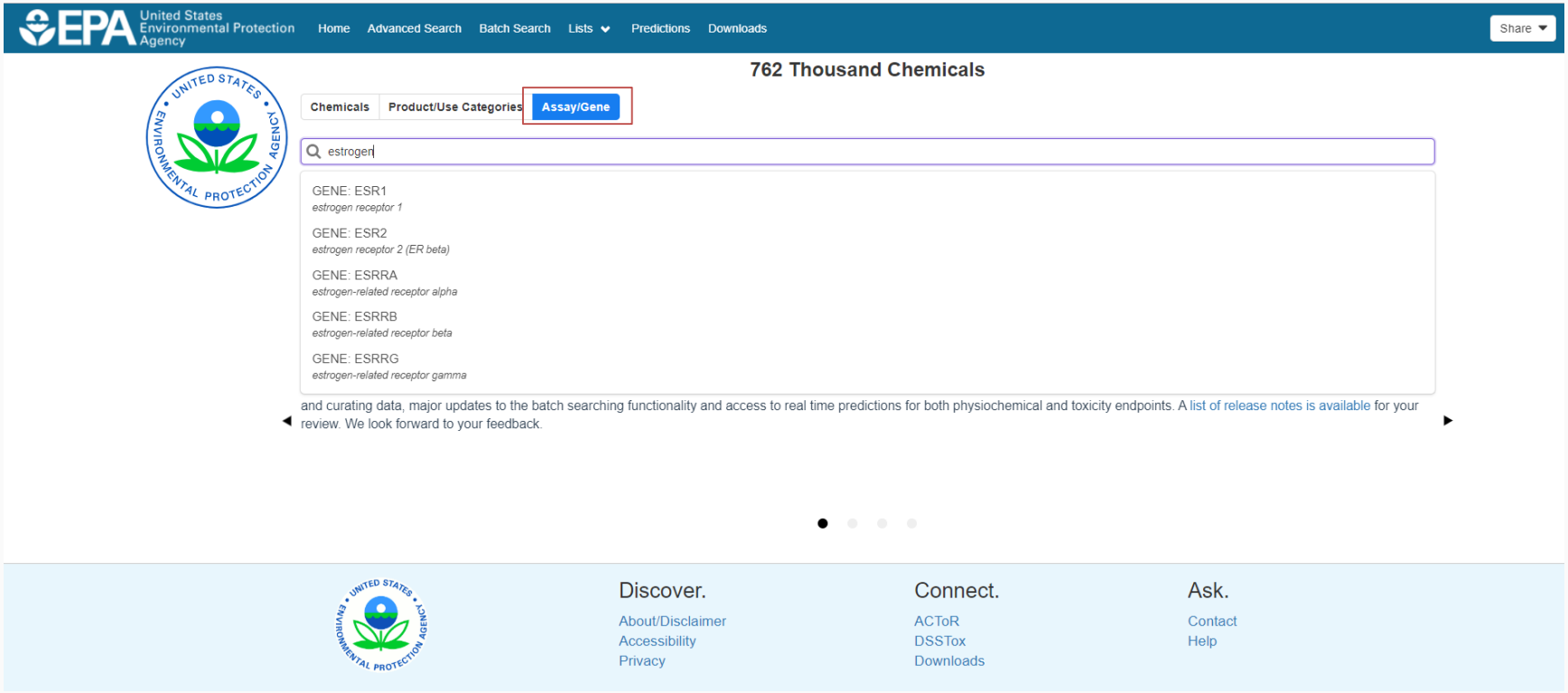
# CompTox Dashboard Chemicals



# CompTox Dashboard
# Products and Use Categories



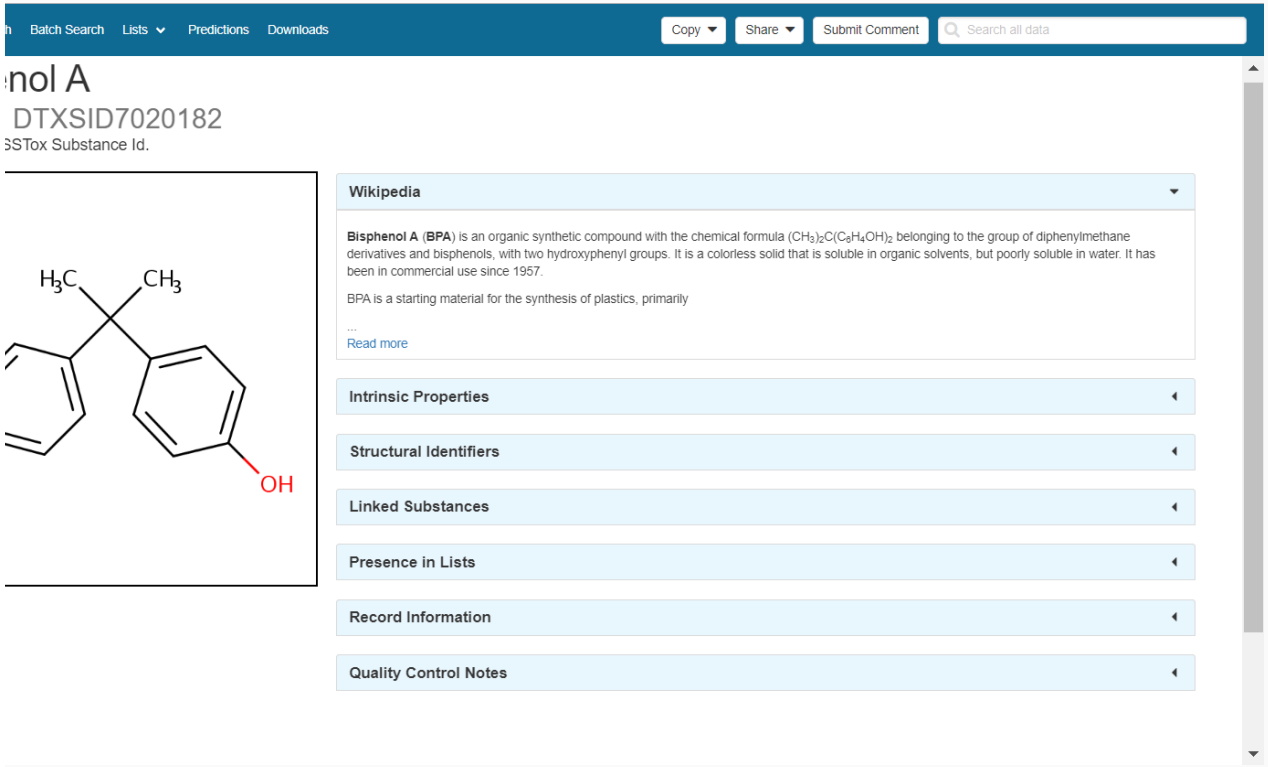
# CompTox Dashboard
# Assays and Genes



# Detailed Chemical Pages

EPA United States Environmental Protection Agency

- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- ADME
- EXPOSURE
- BIOACTIVITY
- SIMILAR COMPOUNDS
- GENRA (BETA)
- RELATED SUBSTANCES
- SYNONYMS
- LITERATURE
- LINKS
- COMMENTS



# Physicochemical properties

Property

Summary

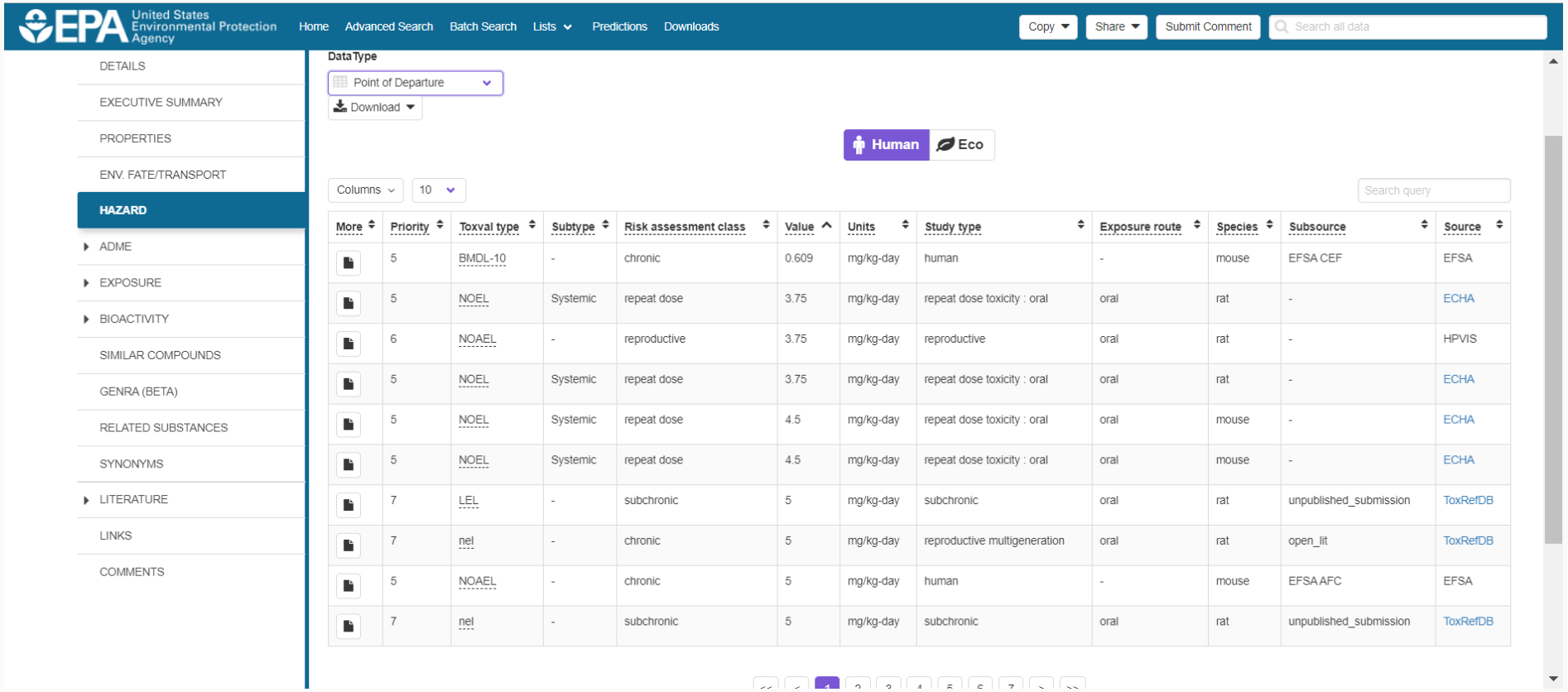
- Summary
- LogP: Octanol-Water
- Melting Point
- Boiling Point
- Water Solubility
- Vapor Pressure
- Flash Point
- Surface Tension
- Index of Refraction
- Molar Refractivity
- Polarizability
- Density
- Molar Volume
- Thermal Conductivity
- Viscosity
- Henry's Law
- LogKoa: Octanol-Air

## Summary

Search query

| verage | 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*K) |

# Access to Chemical Hazard Data



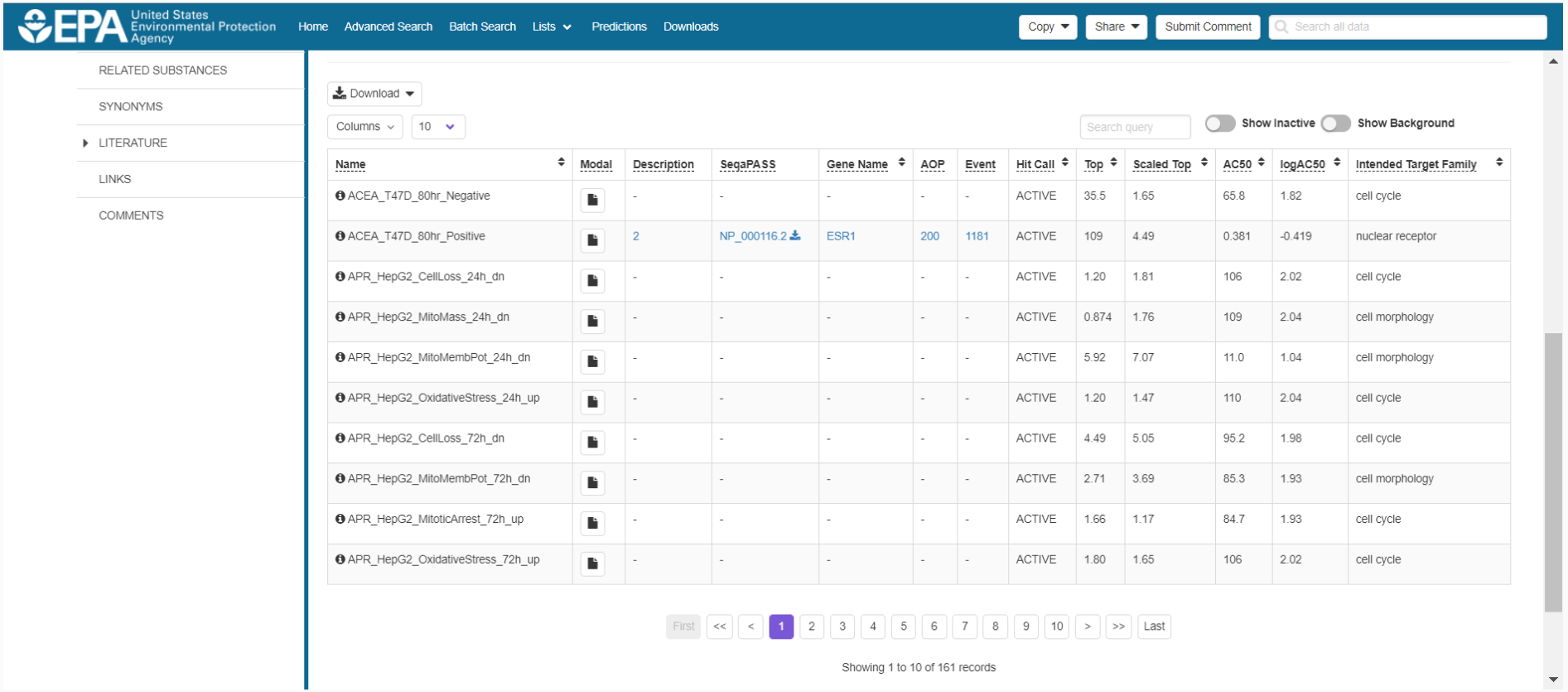

| More | Priority | Toxval type | Subtype | Risk assessment class | Value | Units | Study type | Exposure route | Species | Subsource | Source |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 5 | BMDL-10 | - | chronic | 0.609 | mg/kg-day | human | - | mouse | EFSA CEF | EFSA |
| | 5 | NOEL | Systemic | 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 |

# Hazard Data from “ToxVal_DB”

- 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

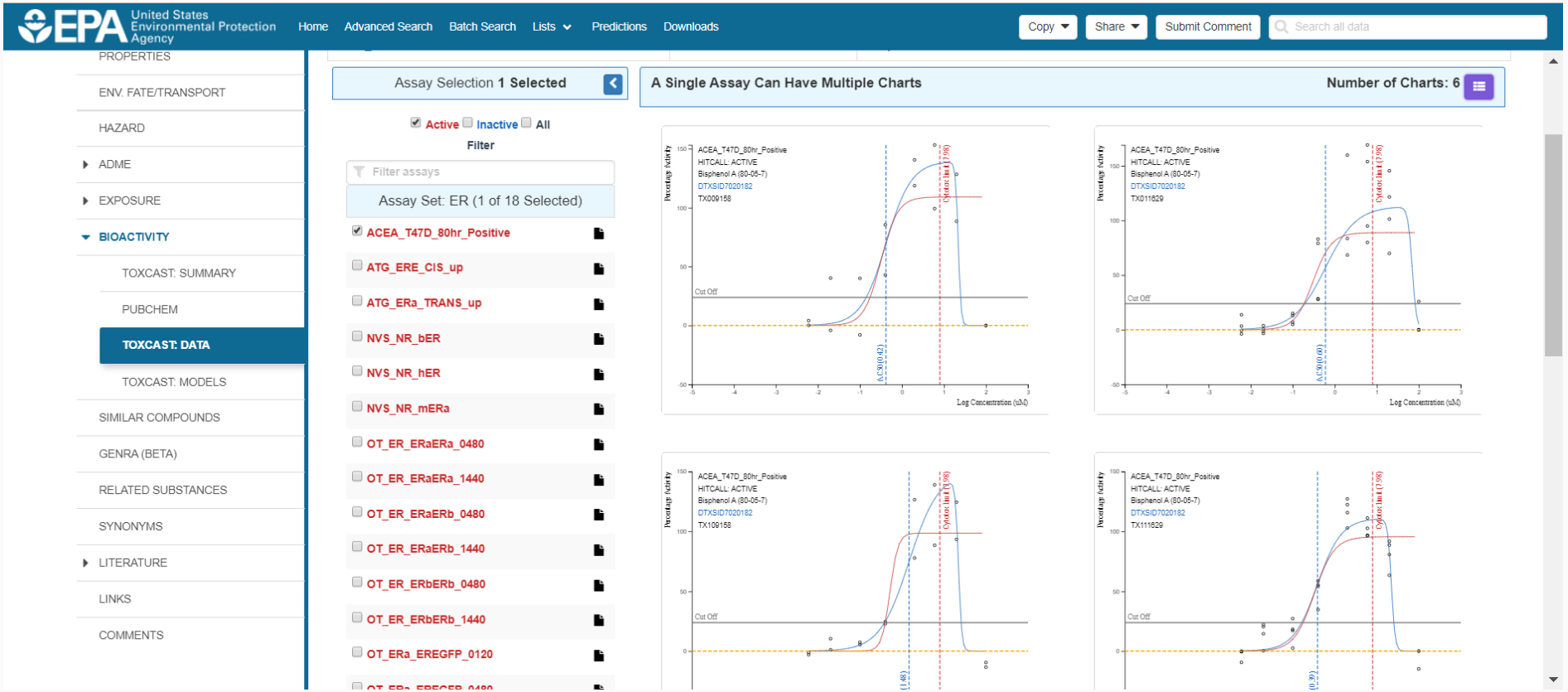
## ToxCast and Tox21




| Name | Modal | Description | SegaPASS | 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 |

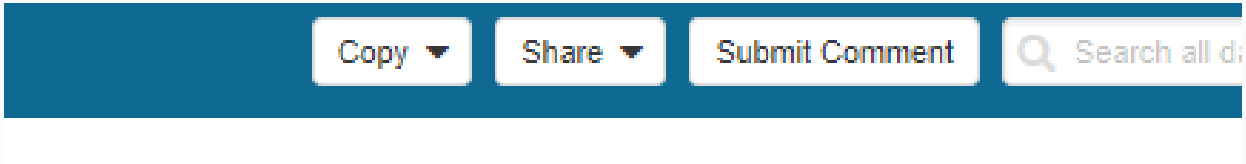
# *In Vitro* Bioassay Screening

## ToxCast and Tox21



# How can be 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

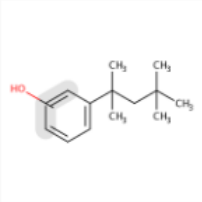
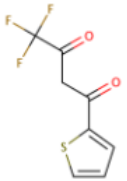
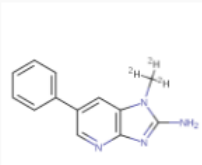
https://comptox.epa.gov/dashboard/comments/public_index

EPA United States Environmental Protection Agency

EPA United States Environmental Protection Agency | Home | Advanced Search | Batch Search | Lists | Predictions | Downloads | Share | Search all data

## Crowdsourced Comments

Show 10 entries | Search:

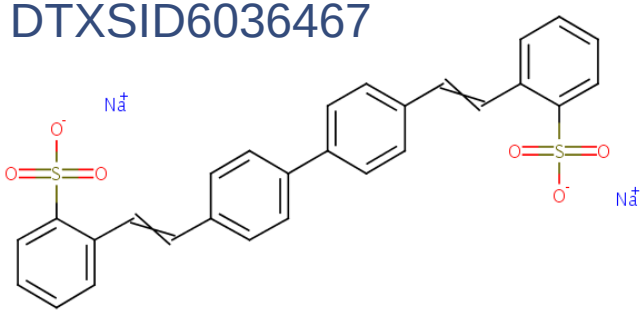
| Chemical | Structure | Date | Comment | Status |
|---|---|---|---|---|
| (1,1,3,3-Tetramethylbutyl)phenol |  | 2017-07-15 | Octylphenol redirects here, yet the name and related chemicals are 1,1,3,3-tetramethylbutylphenol - which is only a subset of all octylphenol isomers? Is this CAS only for these 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-(<sup>2</sup>H<sub>3</sub>)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

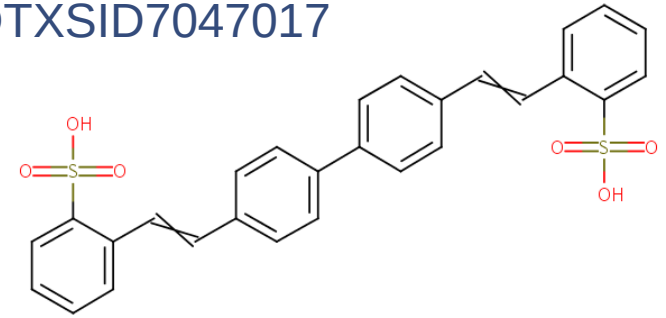
# MassBank/CompTox Curation of External Data

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

Purchased: CAS: 27344-41-8
DTXSID6036467




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




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:
48a CAS number on record was 27344-41-8 which is DTXSID6036467.

Add A Comment

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(c(cc1)S(=O)(=O)O)/C=C/c2ccc(cc2)c3ccc(cc3)/C=C/c4c(cccc4)S(=O)(=O)O SQAKQVFOMMLRPR-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.

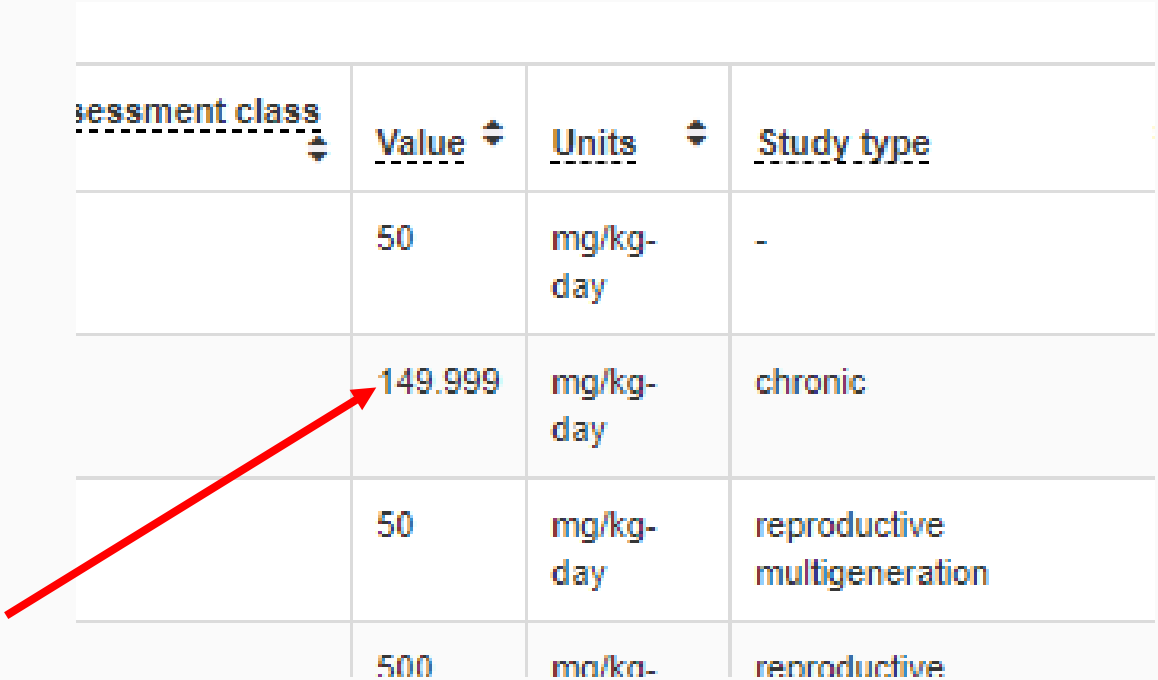
Admin reply posted 4 months ago

# Comments to date

- 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

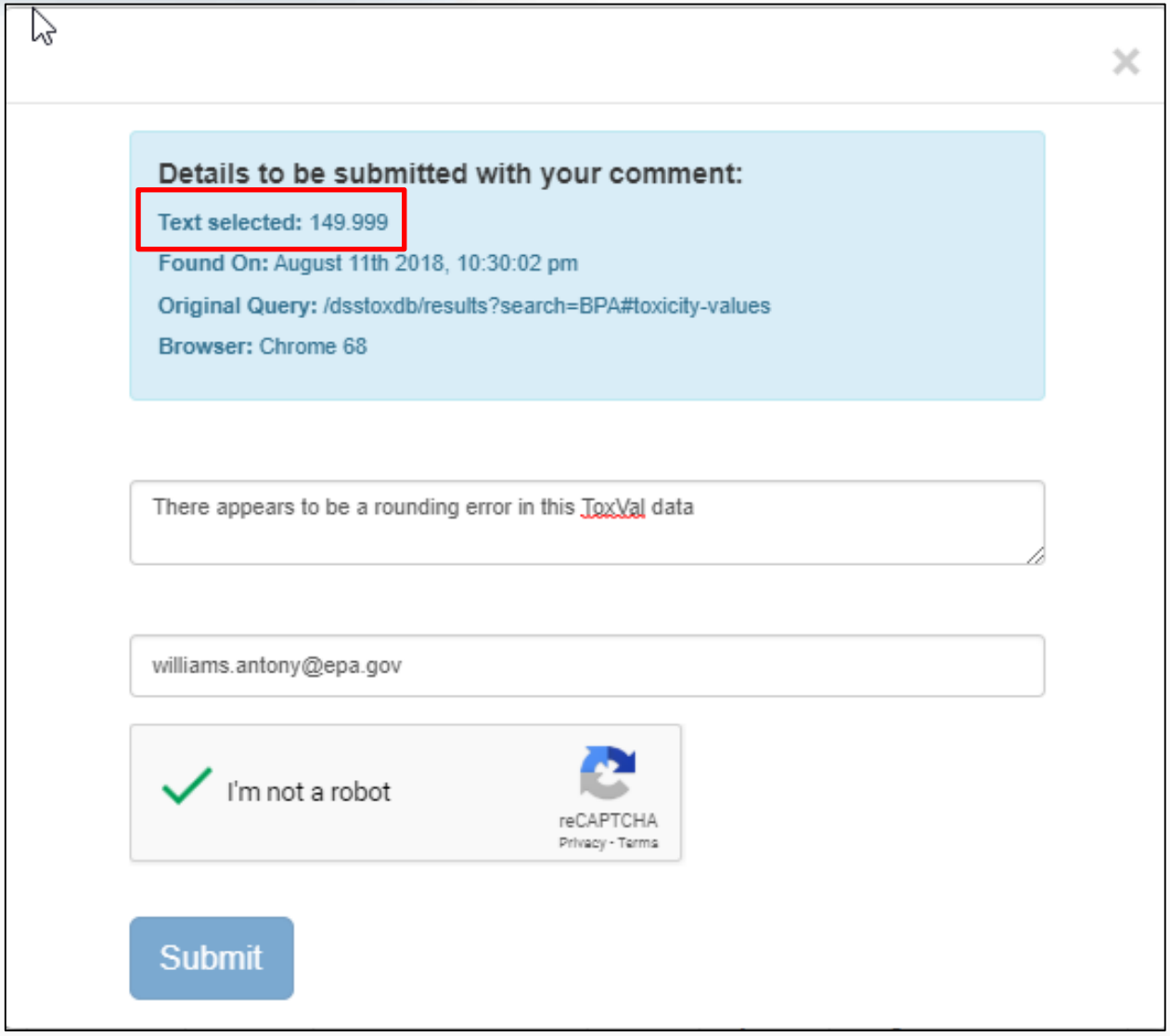
# Crowdsourcing Comments
# Single Cell Commenting added

- Highlight an alphanumeric text string



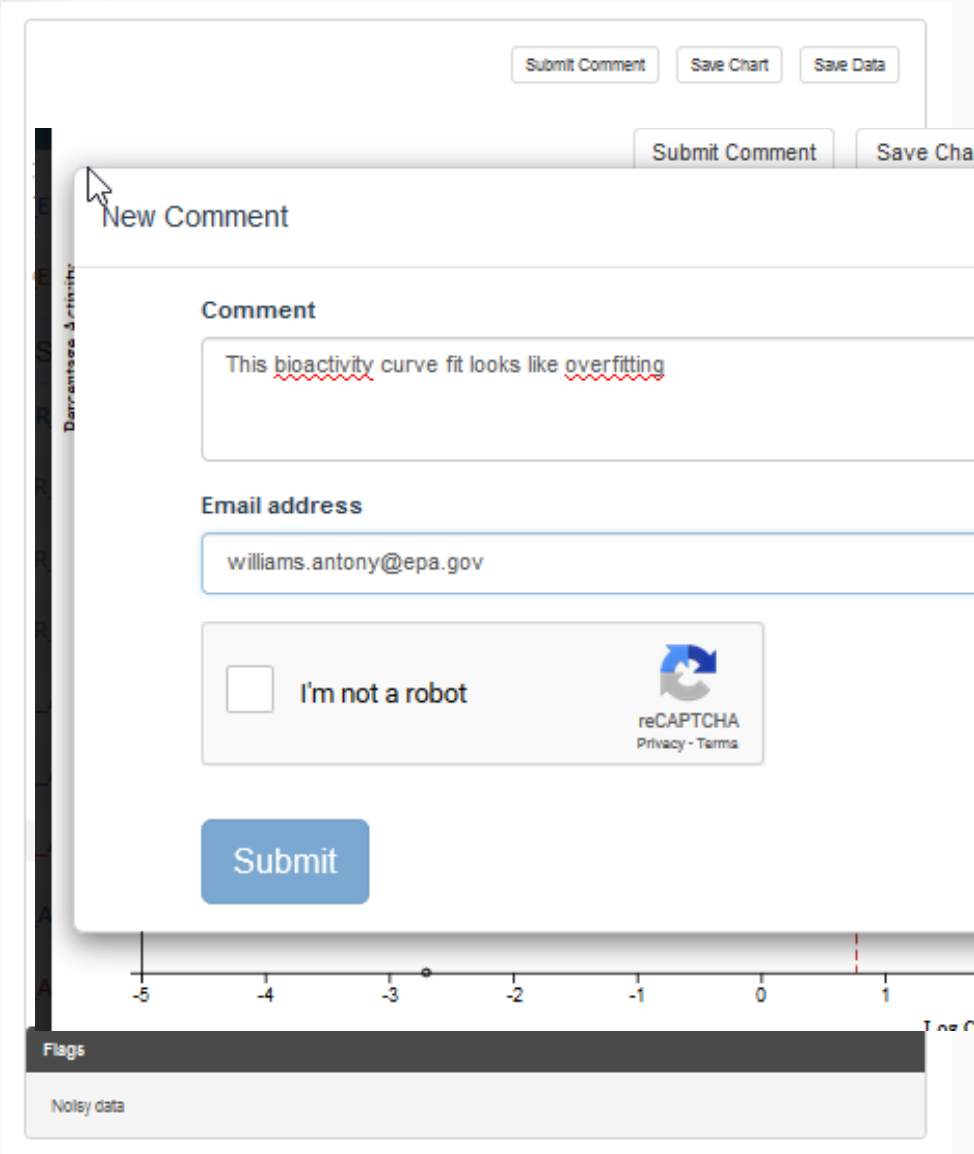
| ssessment class | Value | Units | Study type |
|---|---|---|---|
| | 50 | mg/kg-day | - |
| | 149.999 | mg/kg-day | chronic |
| | 50 | mg/kg-day | reproductive multigeneration |
| | 500 | mg/kg- | reproductive |

# Crowdsourcing Comments

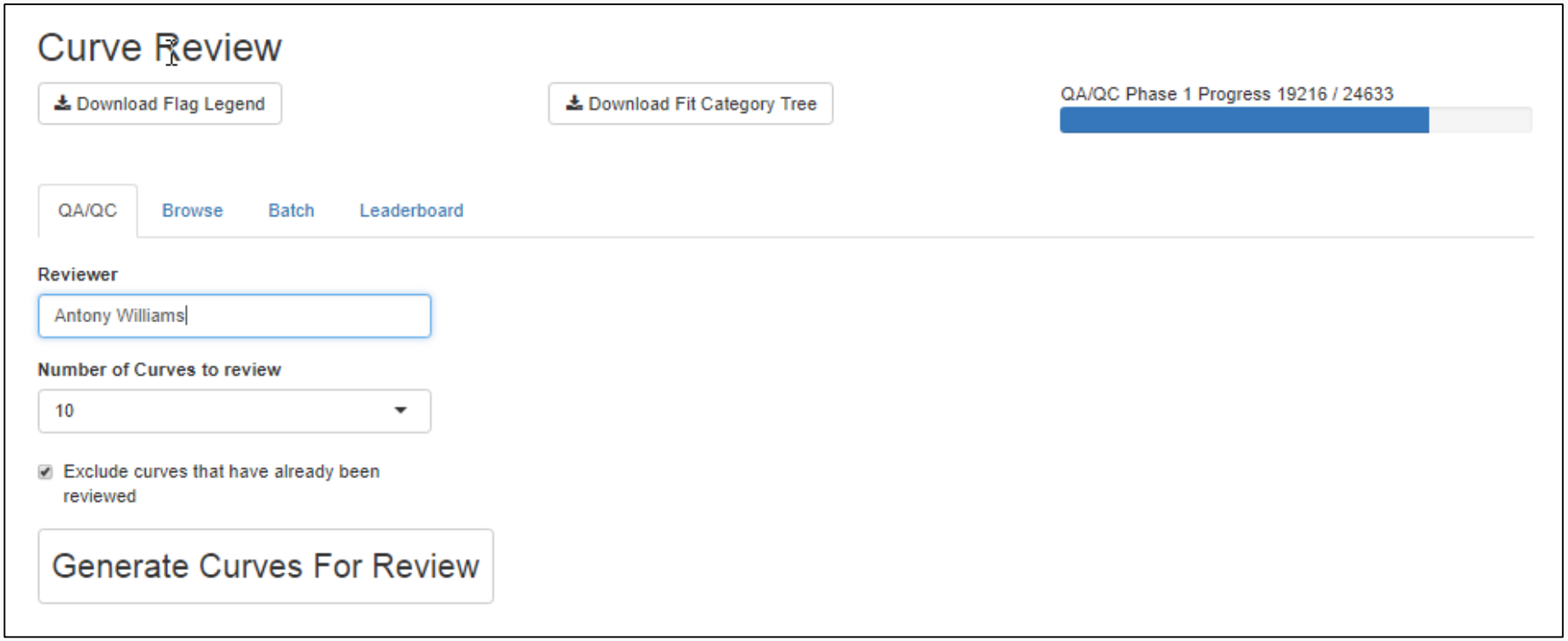


# Bioactivity Data

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

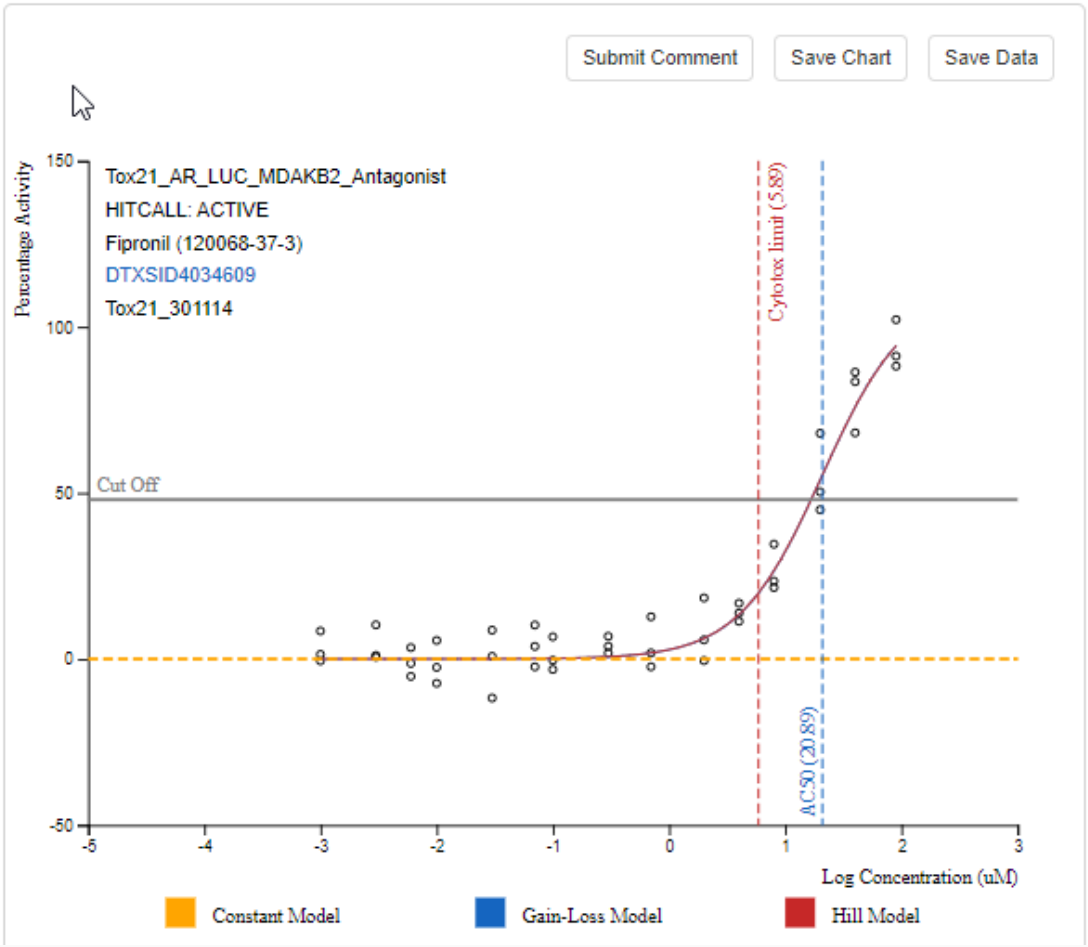


# Internal Review of 25,000 curves



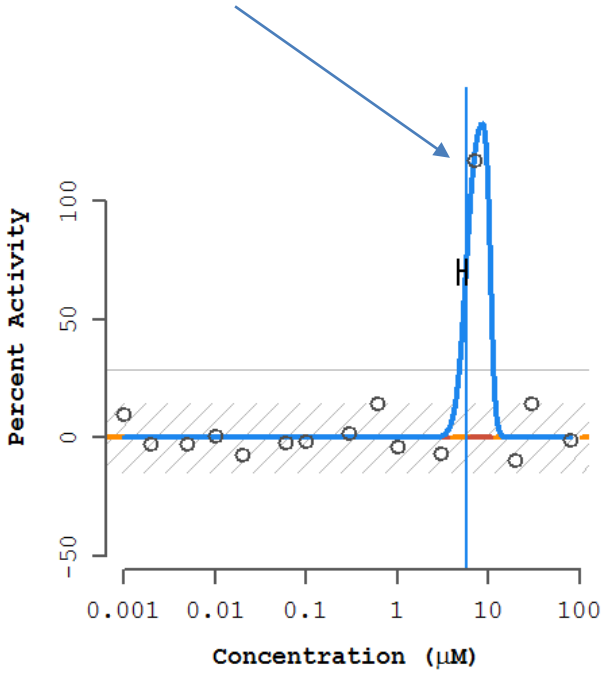

**Screenshot of entry page for Beta R Shiny Application for NCCT users**
**Brown & Paul-Friedman**

# Internal Review of 25,000 curves
# A “good fit” bioactivity curve



# Internal Review of 25,000 curves

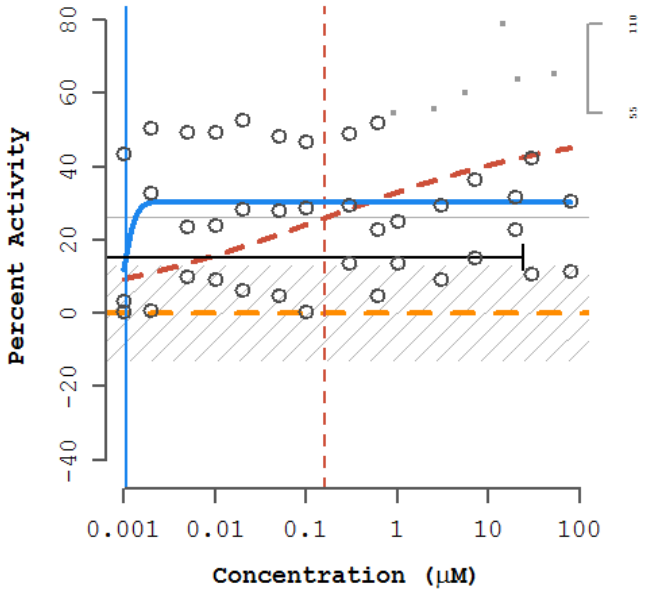
**Single-Point in middle of concentration range drives ACTIVE Hit Call**




```
ASSAY:    TOX21_RT_HEK293_FLO_40hr_viability

NAME:     Tributylmethylammonium bis(trifluoromethylsulfc
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

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
```

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

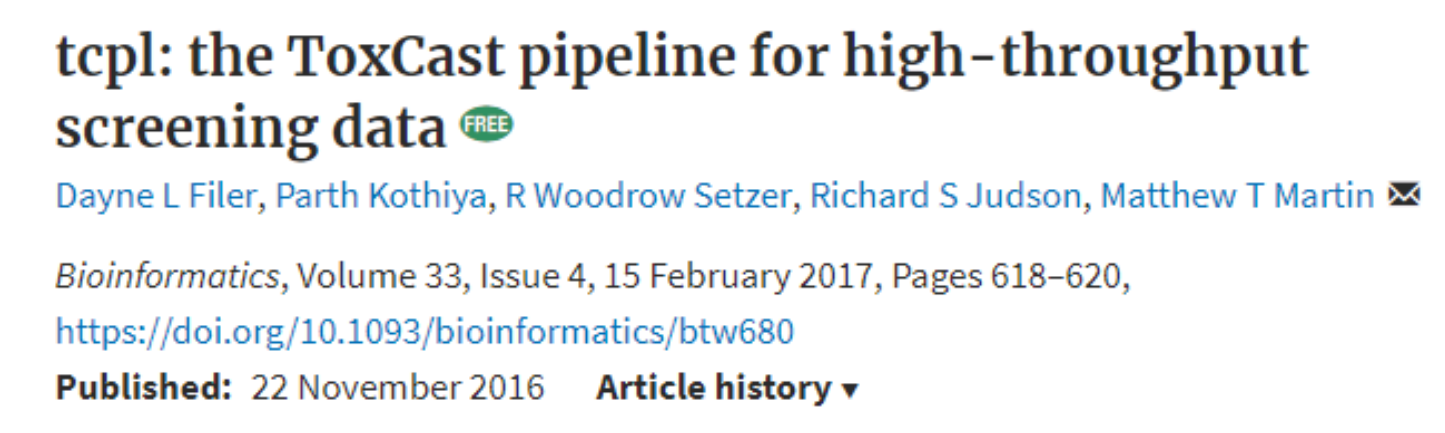
# Internal Review of 25,000 curves Abnormally High-Noise




```
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 Results

- 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)*



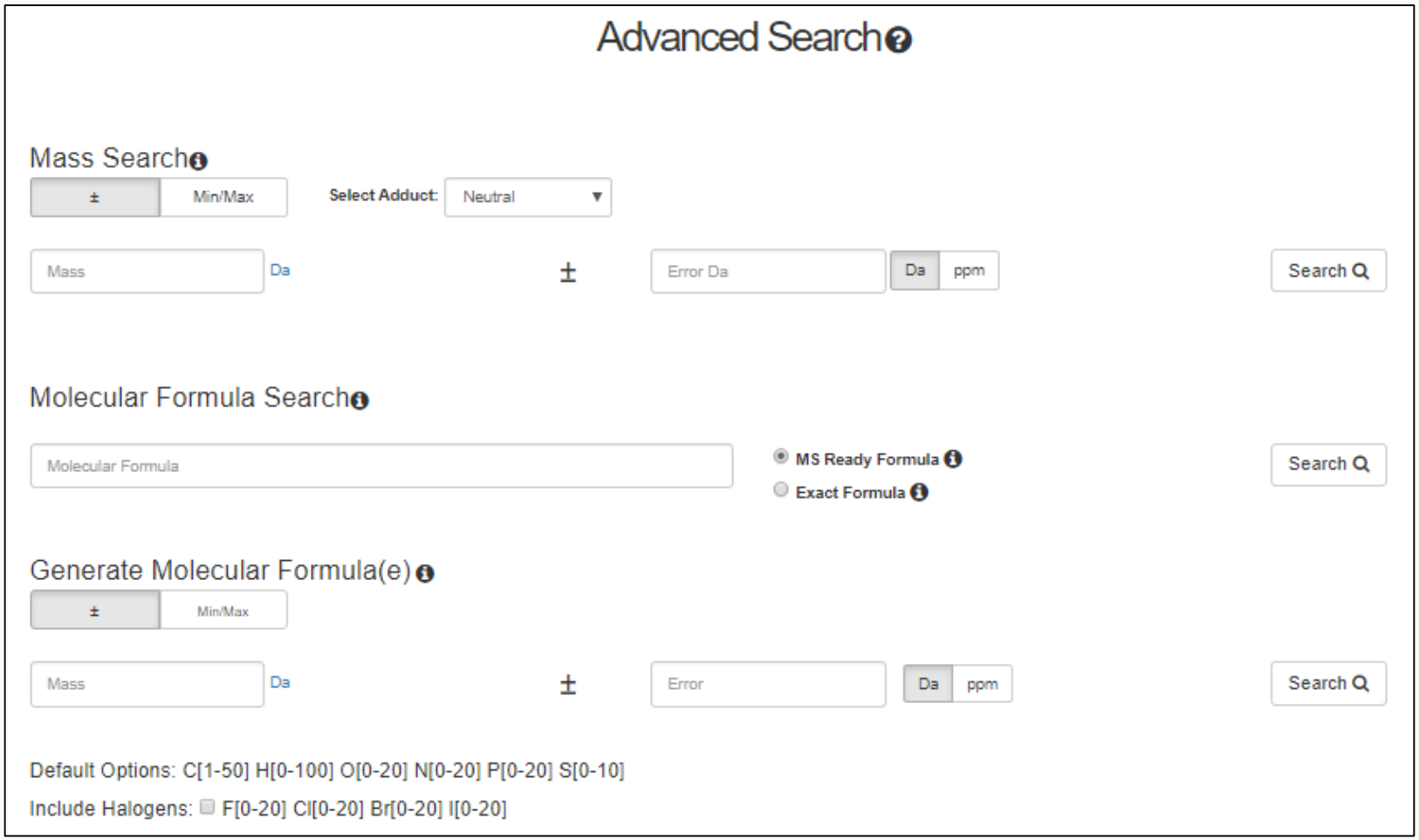
tcpl: the ToxCast pipeline for high-throughput screening data FREE

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



# Collaborative Data Curation

- 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

EPA United States Environmental Protection Agency

**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

Search

**User login**

Username *

Password *

Request new password

Log in

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 aspart 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 >40,000structures from suspect lists. See **interactive version**. Compiled by Reza Aalizadeh, University of Athens, including RTI and toxicity values, support by Nikiforos Alygizakis, EI. *Work in progress ... please report any issues!* |
| S1 | MASSBANK | **NORMAN Compounds in MassBank** | **CSV**, **XLSX** with Fragments (3/10/2017)<br>CompTox **MassBank EU Reference List**<br>CompTox **MassBank EU Special Cases**<br>CompTox **Fragment Download** | **MassBankEUInChIKeys** (11/04/2017) | **www.massbank.eu**<br>Stravs *et al.* 2013.<br>DOI: **10.1002/jms.3131** |
| S2 | STOFFIDENT | **HSWT/LfU STOFF-IDENT Database of Water-Relevant Substances** | **STOFF-IDENT Contents** (6/09/2017)<br>CompTox STOFF-IDENT List<br>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)<br>GC-MS: **CSV**, **XLSX** (3/10/2017)<br>CompTox **NORMANCT15 List** | **LC-MS InChIKeys** (31/10/2016)<br>**GC-MS InChIKeys** (31/10/2016) | Schymanski *et al.* 2015.<br>DOI: **10.1007/s00216-015-8681-7** |

# Example: NORMAN Priority List

| | | | | | |
|---|---|---|---|---|---|
| S15 | NORMANPRI | **NORMAN Priority List** | NORMAN Priority **CSV** (13/7/2017)<br>CompTox **NORMAN Priority List**<br>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 | C[C@H]1[C@@H](NC(=O)[C@@H](NC(=O)[C@ |
| StNorman3 | Microcystin-YR | CAS_RN: 101064-48-6 | C[C@H]1[C@@H](NC(=O)[C@@H](NC(=O)[C@ |
| StNorman5 | 2,6-Di-tert-butylphenol | CAS_RN: 128-39-2 | CC(C)(C)c1cccc(c1O)C(C)(C)C |
| StNorman6 | Butylated hydroxyanisole | CAS_RN: 25013-16-5 | COc1ccc(O)c(c1)C(C)(C)C |
| StNorman7 | tert-Butylhydroquinone | CAS_RN: 1948-33-0 | CC(C)(C)c1cc(O)ccc1O |
| StNorman8 | Butylated hydroxytoluene | CAS_RN: 128-37-0 | Cc1cc(c(O)c(c1)C(C)(C)C)C(C)(C)C |
| StNorman15 | Diethylenetriaminepentaace | CAS_RN: 67-43-6 | OC(=O)CN(CCN(CC(O)=O)CC(O)=O)CCN(CC(O)= |
| StNorman16 | Ethylenediaminetetraacetic | CAS_RN: 60-00-4 | OC(=O)CN(CCN(CC(O)=O)CC(O)=O)CC(O)=O |
| StNorman17 | Nitrilotriacetic acid | CAS_RN: 139-13-9 | OC(=O)CN(CC(O)=O)CC(O)=O |
| StNorman18 | Oxadixyl | CAS_RN: 77732-09-3 | COCC(=O)N(N1CCOC1=O)c2c(C)cccc2C |
| StNorman19 | Tetraacetylethylenediamine | CAS_RN: 10543-57-4 | CC(=O)N(CCN(C(C)=O)C(C)=O)C(C)=O |
| StNorman26 | 2-(2-(4-Nonylphenoxy)etho | CAS_RN: 106807-78-7 | CCCCCCCCCc1ccc(OCCOCC(O)=O)cc1 |
| StNorman29 | 4-Octylphenoxy acetic acid | CAS_RN: 15234-85-2 | CCCCCCCCc1ccc(OCC(O)=O)cc1 |
| StNorman30 | Cyanoformaldehyde | CAS_RN: 4471-47-0 | NCC=O |
| StNorman31 | Decabromodiphenyl ethane | CAS_RN: 84852-53-9 | C(CC1=C(C(=C(C(=C1Br)Br)Br)Br)Br)C2=C(C(=C |
| StNorman32 | Hexabromocyclododecane | CAS_RN: 25637-99-4 | BrC1CC(Br)CC(Br)CC(Br)CC(Br)CC(Br)C1 |
| StNorman33 | n-Nitrosodimethylamine | CAS_RN: 62-75-9 | CN(C)N=O |
| StNorman34 | Benzylbutylphthalate | CAS_RN: 85-68-7 | CCCCOC(=O)c1ccccc1C(=O)OCc2ccccc2 |
| StNorman35 | Diethyl phthalate | CAS_RN: 84-66-2 | CCOC(=O)c1ccccc1C(=O)OCC |
| StNorman37 | Di-n-butylphthalate | CAS_RN: 84-74-2 | CCCCOC(=O)c1ccccc1C(=O)OCCCC |
| StNorman39 | Bisphenol A | CAS_RN: 80-05-7 | CC(C)(c1ccc(O)cc1)c2ccc(O)cc2 |

# Mapping on Two Identifiers

| LOOKUP BASED ON CHEMICAL NAME | | | | | | | |
|---|---|---|---|---|---|---|---|
| DTXSID | PREFERRED NAME | CASRN | INCHI KEY | IUPAC NA | SMILES | MOL FORMULA | MONOISC |
| DTXSID40880085 | Microcystin RR | 111755-37-4 | JIGDOBKZI | (5R,8S,11R | CO[C@@H | C49H75N13O12 | 1037.5658 |
| DTXSID00880086 | Microcystin YR | 101064-48-6 | OWHASZC | (5R,8S,11R | CO[C@@H | C52H72N10O13 | 1044.5280 |
| DTXSID6027052 | 2,6-Di-tert-butylphen | 128-39-2 | DKCPKDPY | 2,6-Di-tert | CC(C)(C)C1 | C14H22O | 206.16706 |
| DTXSID7020215 | Butylated hydroxyani | 25013-16-5 | CZBZUDVE | 2-tert-But | COC1=CC= | C22H32O4 | 360.23005 |
| DTXSID6020220 | tert-Butylhydroquino | 1948-33-0 | BGNXCDM | 2-tert-But | CC(C)(C)C1 | C10H14O2 | 166.09937 |
| DTXSID2020216 | Butylated hydroxytol | 128-37-0 | NLZUEZXR | 2,6-Di-tert | CC1=CC(=C | C15H24O | 220.18271 |
| DTXSID2023434 | Pentetic acid | 67-43-6 | QPCDCPDI | N,N-Bis{2- | OC(=O)CN | C14H23N3O10 | 393.13834 |
| DTXSID6022977 | Ethylenediaminetetra | 60-00-4 | KCXVZYZYI | 2,2',2'',2'''- | OC(=O)CN | C10H16N2O8 | 292.09066 |
| DTXSID6020939 | Nitrilotriacetic acid | 139-13-9 | MGFYIUFZ | 2,2',2''-Nit | OC(=O)CN | C6H9NO6 | 191.04298 |
| DTXSID2032631 | Oxadixyl | 77732-09-3 | UWVQIRO | N-(2,6-Din | COCC(=O) | C14H18N2O4 | 278.12665 |
| DTXSID5040752 | Tetraacetylethylened | 10543-57-4 | BGRWYDH | N,N'-(Etha | CC(=O)N(C | C10H16N2O4 | 228.11100 |
| DTXSID70147779 | 2-(2-(4-Nonylphenox | 106807-78-7 | RAQHOBR | [2-(4-Nony | CCCCCCCC | C19H30O4 | 322.21440 |
| DTXSID60165003 | Acetic acid, (4-octylph | 15234-85-2 | DWUYSEN | (4-Octylph | CCCCCCCC | C16H24O3 | 264.17254 |
| DTXSID0021549 | Cyanoformaldehyde | 4471-47-0 | TUHMQDC | Oxoacetor | O=CC#N | C2HNO | 55.005813 |
| DTXSID2052732 | 1,1'-Ethane-1,2-diylbi | 84852-53-9 | BZQKBFHE | 1,1'-(Ethar | BrC1=C(Br | C14H4Br10 | 961.21468 |
| DTXSID8025383 | Hexabromocyclodod | 25637-99-4 | | - | - | - | - |
| DTXSID7021029 | N-Nitrosodimethylan | 62-75-9 | UMFJAHH | N,N-Dimet | CN(C)N=O | C2H6N2O | 74.048012 |
| DTXSID3020205 | Benzyl butyl phthalat | 85-68-7 | IRIAEXORF | Benzyl but | CCCCOC(= | C19H20O4 | 312.13615 |
| DTXSID7021780 | Diethyl phthalate | 84-66-2 | FLKPEMZC | Diethyl be | CCOC(=O) | C12H14O4 | 222.08920 |
| DTXSID2021781 | Dibutyl phthalate | 84-74-2 | DOIRQSBF | Dibutyl be | CCCCOC(= | C16H22O4 | 278.15180 |
| DTXSID7020182 | Bisphenol A | 80-05-7 | IISBACLAF | 4,4'-(Propa | CC(C)(C1= | C15H16O2 | 228.11502 |

# Mapping on Two Identifiers

| | | | LOOKUP BASED ON CASRN | | | | |
|---|---|---|---|---|---|---|---|
| DTXSID | PREFERRED NAME | CASRN | INCHI KEY | IUPAC NA | SMILES | MOL FORMULA | MONOISC |
| DTXSID40880085 | Microcystin RR | 111755-37-4 | JIGDOBKZI | (5R,8S,11R | CO[C@@H | C49H75N13O12 | 1037.5658 |
| DTXSID00880086 | Microcystin YR | 101064-48-6 | OWHASZC | (5R,8S,11R | CO[C@@H | C52H72N10O13 | 1044.5280 |
| DTXSID6027052 | 2,6-Di-tert-butylpheno | 128-39-2 | DKCPKDPY | 2,6-Di-tert | CC(C)(C)C1 | C14H22O | 206.16706 |
| DTXSID7020215 | Butylated hydroxyanis | 25013-16-5 | CZBZUDVE | 2-tert-But | COC1=CC= | C22H32O4 | 360.23005 |
| DTXSID6020220 | tert-Butylhydroquinon | 1948-33-0 | BGNXCDM | 2-tert-But | CC(C)(C)C1 | C10H14O2 | 166.09937 |
| DTXSID2020216 | Butylated hydroxytolue | 128-37-0 | NLZUEZXR | 2,6-Di-tert | CC1=CC(=C | C15H24O | 220.18271 |
| DTXSID2023434 | Pentetic acid | 67-43-6 | QPCDCPDI | N,N-Bis{2- | OC(=O)CN | C14H23N3O10 | 393.13834 |
| DTXSID6022977 | Ethylenediaminetetraa | 60-00-4 | KCXVZYZYI | 2,2',2'',2'''- | OC(=O)CN | C10H16N2O8 | 292.09066 |
| DTXSID6020939 | Nitrilotriacetic acid | 139-13-9 | MGFYIUFZ | 2,2',2''-Nit | OC(=O)CN | C6H9NO6 | 191.04298 |
| DTXSID2032631 | Oxadixyl | 77732-09-3 | UWVQIRO | N-(2,6-Din | COCC(=O) | C14H18N2O4 | 278.12665 |
| DTXSID5040752 | Tetraacetylethylenedia | 10543-57-4 | BGRWYDH | N,N'-(Etha | CC(=O)N(C | C10H16N2O4 | 228.11100 |
| DTXSID70147779 | 2-(2-(4-Nonylphenoxy) | 106807-78-7 | RAQHOBR | [2-(4-Nony | CCCCCCCC | C19H30O4 | 322.21440 |
| DTXSID60165003 | Acetic acid, (4-octylphe | 15234-85-2 | DWUYSEM | (4-Octylph | CCCCCCCC | C16H24O3 | 264.17254 |
| DTXSID0021549 | Cyanoformaldehyde | 4471-47-0 | TUHMQDC | Oxoacetor | O=CC#N | C2HNO | 55.005813 |
| DTXSID2052732 | 1,1'-Ethane-1,2-diylbis( | 84852-53-9 | BZQKBFHE | 1,1'-(Ethar | BrC1=C(Br | C14H4Br10 | 961.21468 |
| DTXSID8025383 | Hexabromocyclododec | 25637-99-4 | | - | - | - | - |
| DTXSID7021029 | N-Nitrosodimethylamir | 62-75-9 | UMFJAHH | N,N-Dimet | CN(C)N=O | C2H6N2O | 74.048012 |
| DTXSID3020205 | Benzyl butyl phthalate | 85-68-7 | IRIAEXORF | Benzyl but | CCCCOC(= | C19H20O4 | 312.13615 |
| DTXSID7021780 | Diethyl phthalate | 84-66-2 | FLKPEMZC | Diethyl be | CCOC(=O) | C12H14O4 | 222.08920 |
| DTXSID2021781 | Dibutyl phthalate | 84-74-2 | DOIRQSBF | Dibutyl be | CCCCOC(= | C16H22O4 | 278.15180 |
| DTXSID7020182 | Bisphenol A | 80-05-7 | IISBACLAF | 4,4'-(Propa | CC(C)(C1= | C15H16O2 | 228.11502 |

# Mapping Quality Control (I)

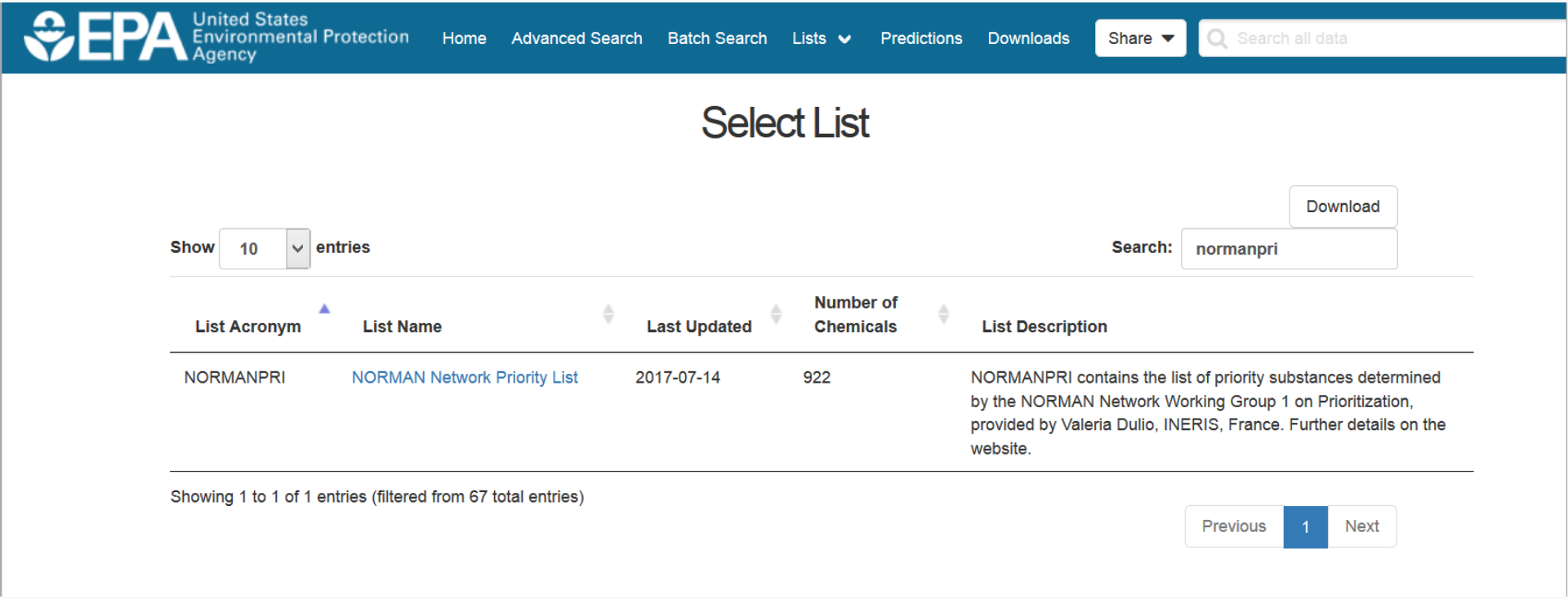
| DTXSID_Name | DTXSID_CAS | DTXSID | DTXSID Equal | Name InChIKey | CAS InChIKey | Name Block1 | CAS Block1 | Name WCAS | Name WName | KeyOR Name | CAS Match | Num NoMatch |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DTXSID40880085 | DTXSID40880085 | DTXSID40880085 | TRUE | FALSE | FALSE | 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 |

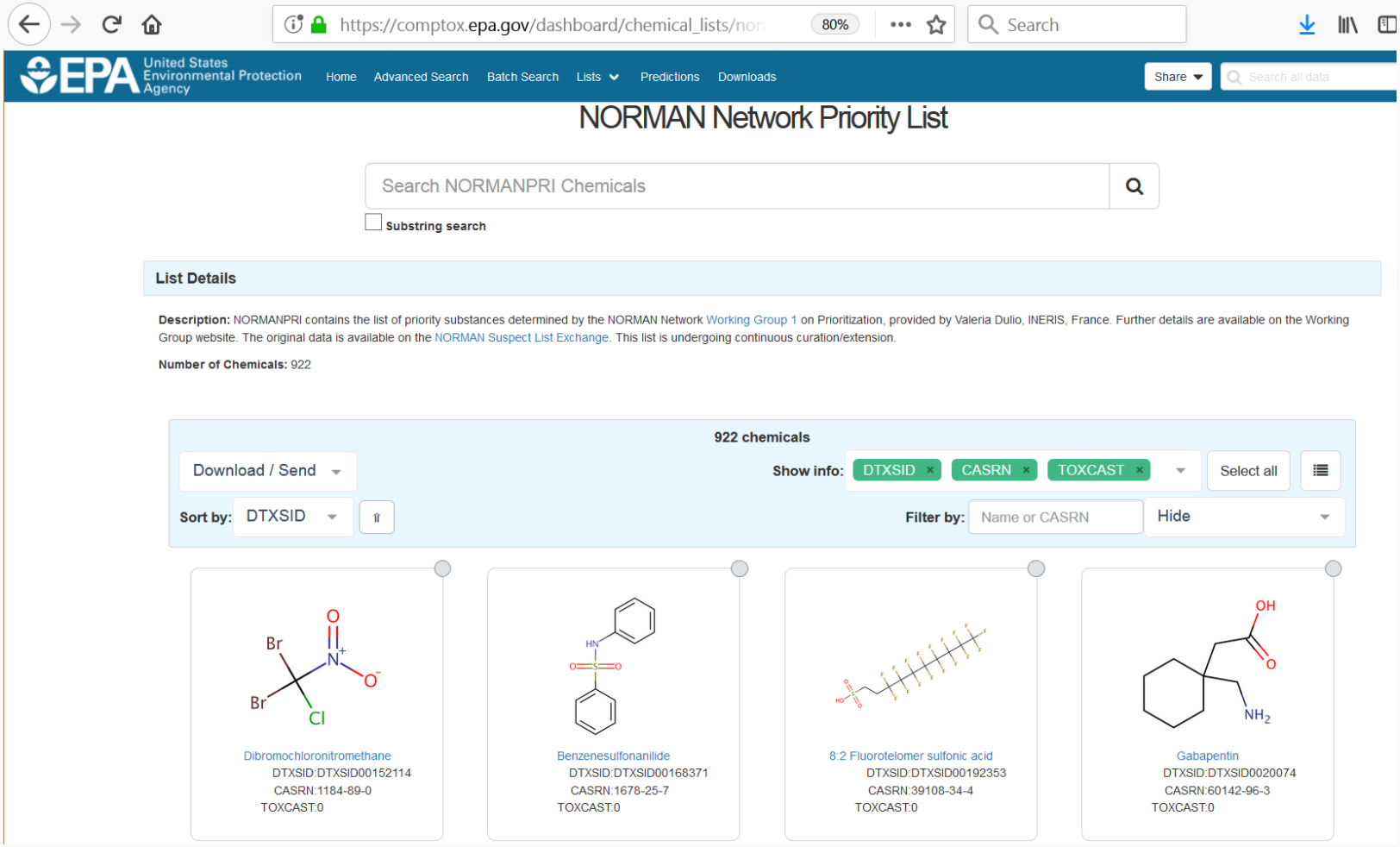
# Example: NORMAN Priority List

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



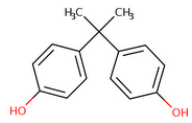
# Example: NORMAN Priority List

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

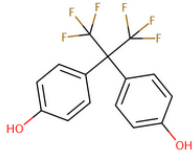


# >23 NORMAN Lists Available

EPA United States Environmental Protection Agency

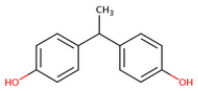


Bisphenol A
DTXSID:DTXSID7020182
CASRN:80-05-7
TOXCAST:166/812

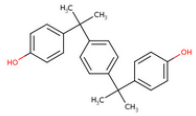


Bisphenol AF
DTXSID:DTXSID7037717
CASRN:1478-61-1
TOXCAST:248/639

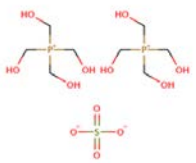
## Bisphenol Compounds



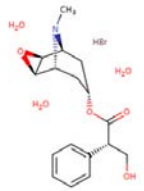
Bisphenol E
DTXSID:DTXSID3047891
CASRN:2081-08-5
TOXCAST:33/276



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

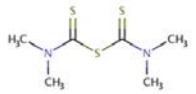


Tetrakis(hydroxymethyl)phosphonium su...
DTXSID:DTXSID0021331
CASRN:55566-30-8

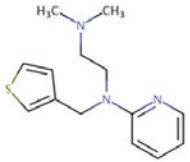


Scopolamine hydrobromide trihydrate
DTXSID:DTXSID0021258
CASRN:6533-68-2

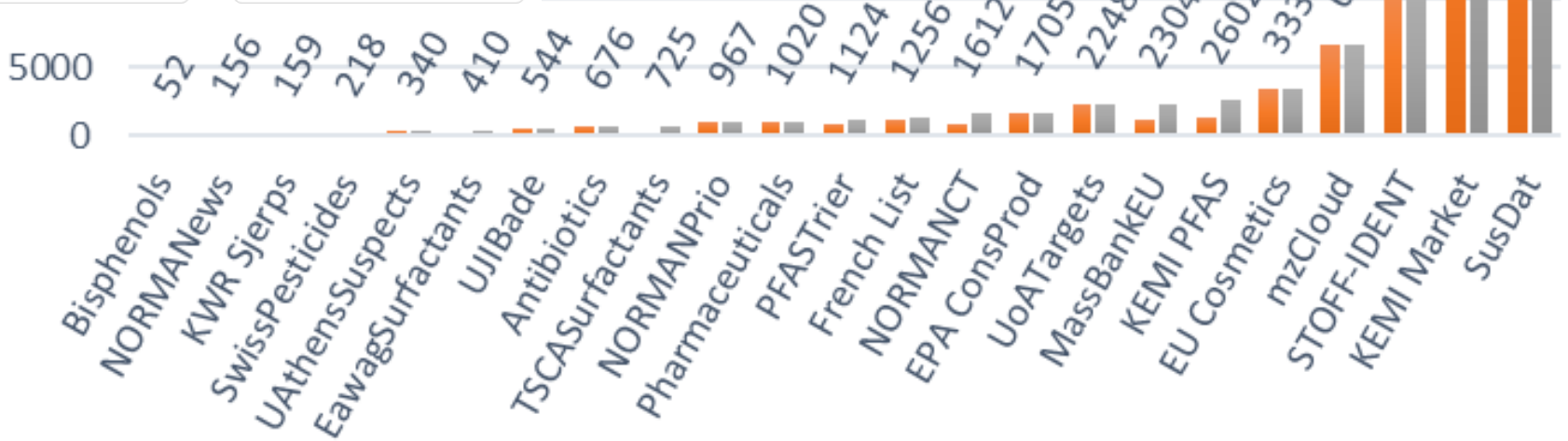
## KEMI List of Substances on the Market



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



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



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

(many) more registrations…

# 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…

# Contact

**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

# NORMAN Suspect Exchange

- 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

Search

**User login**

Username *

Password *

Request new password

Log in

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 aspart 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 >40,000structures from suspect lists. See **interactive version**. Compiled by Reza Aalizadeh, University of Athens, including RTI and toxicity values, support by Nikiforos Alygizakis, EI. *Work in progress ... please report any issues!* |
| S1 | MASSBANK | **NORMAN Compounds in MassBank** | **CSV**, **XLSX** with Fragments (3/10/2017)<br>CompTox **MassBank EU Reference List**<br>CompTox **MassBank EU Special Cases**<br>CompTox **Fragment Download** | **MassBankEUInChIKeys** (11/04/2017) | **www.massbank.eu**<br>Stravs *et al.* 2013.<br>DOI: **10.1002/jms.3131** |
| S2 | STOFFIDENT | **HSWT/LfU STOFF-IDENT Database of Water-Relevant Substances** | **STOFF-IDENT Contents** (6/09/2017)<br>CompTox **STOFF-IDENT List**<br>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)<br>GC-MS: **CSV**, **XLSX** (3/10/2017)<br>CompTox **NORMANCT15 List** | **LC-MS InChIKeys** (31/10/2016)<br>**GC-MS InChIKeys** (31/10/2016) | Schymanski *et al.* 2015.<br>DOI: **10.1007/s00216-015-8681-7** |