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

The EPA Comptox Chemistry Dashboard: A Web-Based Data Integration Hub

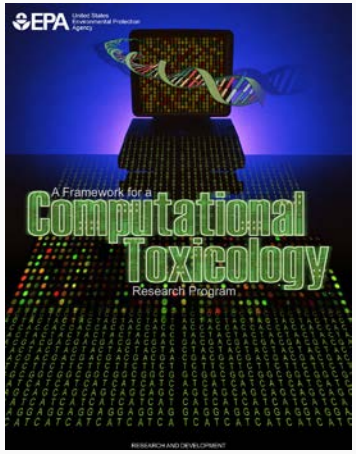
Antony Williams

U.S. Environmental Protection Agency, RTP, NC

NORMAN Network-UFZ

30th November 2017

National Center for Computational Toxicology



- National Center for Computational Toxicology established in 2005 to integrate:
  - High-throughput and high-content technologies
  - Modern molecular biology
  - Data mining and statistical modeling
  - Computational biology and chemistry
- Staffed by ~60 employees and contractors as part of EPA's Office of R&D
- Home of ToxCast & ExpoCast research efforts
- Key partner in U.S. Tox21 federal consortium
- Multiple cross-division collaborations (e.g. NERL, OPP, OPPT)

The NCCT Team

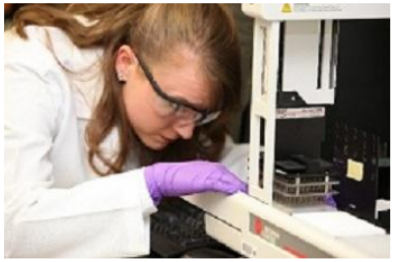
- Our team is broad in skills
  - Biologists, chemists and toxicologists
  - Bioinformaticians and cheminformaticians - modelers (QSAR, Deep Learning), text miners
  - Information technology team, software developers
  - A large IT support team for production applications (National Computing Center)

The need for data and derivative models and algorithms

- NCCT outputs: include a lot of data, models, algorithms and software applications
- We produce Open Data – we want people to interrogate it, learn from it, develop understanding

Toxicity Forecasting

Advancing the Next Generation of Chemical Evaluation



EPA needs rapid and efficient methods to prioritize, screen and evaluate thousands of chemicals. EPA's Toxicity Forecaster (ToxCast) generates data and predictive models on thousands of chemicals of interest to the EPA. ToxCast uses high-throughput screening methods and computational toxicology approaches to rank and prioritize chemicals. In fact, EPA's Endocrine Disruption Screening Program (EDSP) is working to use ToxCast to rank and prioritize chemicals.

- ToxCast has data on over 1,800 chemicals from a broad range of sources including industrial and consumer products, food additives, and potentially "green" chemicals that could be safer alternatives to existing chemicals.
- ToxCast screens chemicals in over 700 high-throughput assays that cover a range of high-

Downloadable Computational Toxicology Data

EPA's computational toxicology research efforts evaluate the potential health effects of thousands of chemicals. The process of evaluating potential health effects involves generating data that investigates the potential harm, or hazard of a chemical, the degree of exposure to chemicals as well as the unique chemical characteristics.

As part of EPA's commitment to share data, all of the computational toxicology data is publicly available for anyone to access and use.

High-throughput Screening Data

EPA researchers use rapid chemical screening (called high-throughput screening assays) to limit the number of laboratory animal tests while quickly and efficiently testing thousands of chemicals for potential health effects.

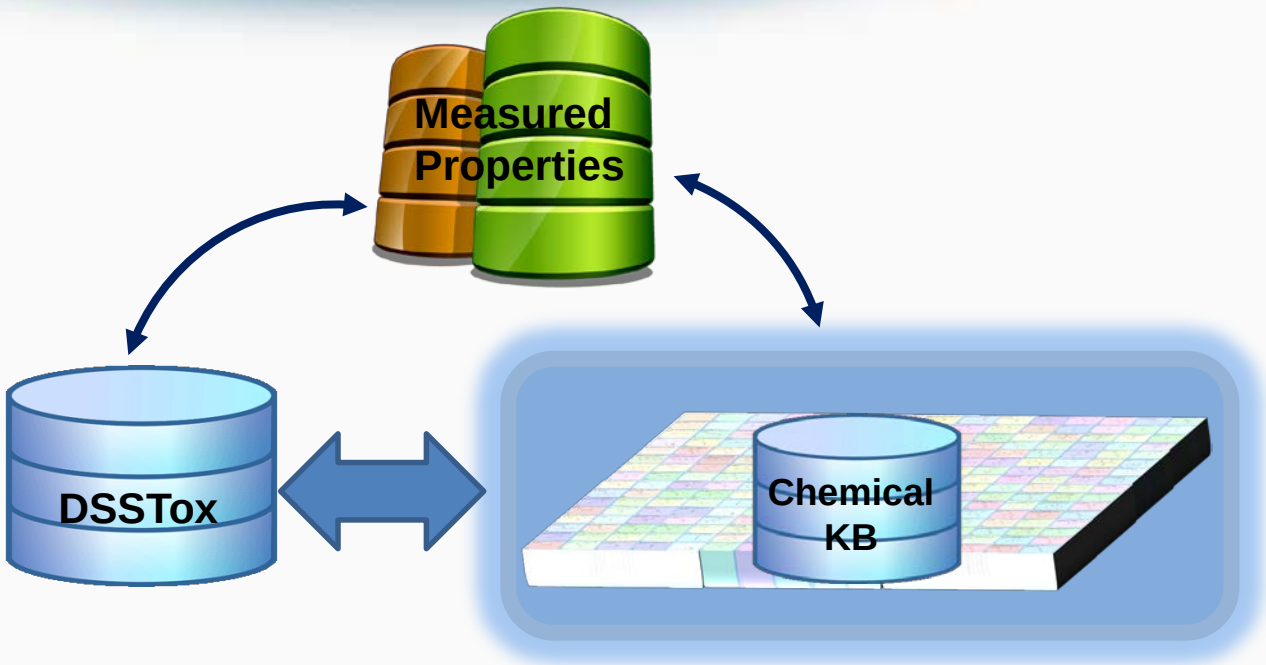
- ToxCast Data: High-throughput screening data on thousands of chemicals.

Rapid Exposure and Dose Data

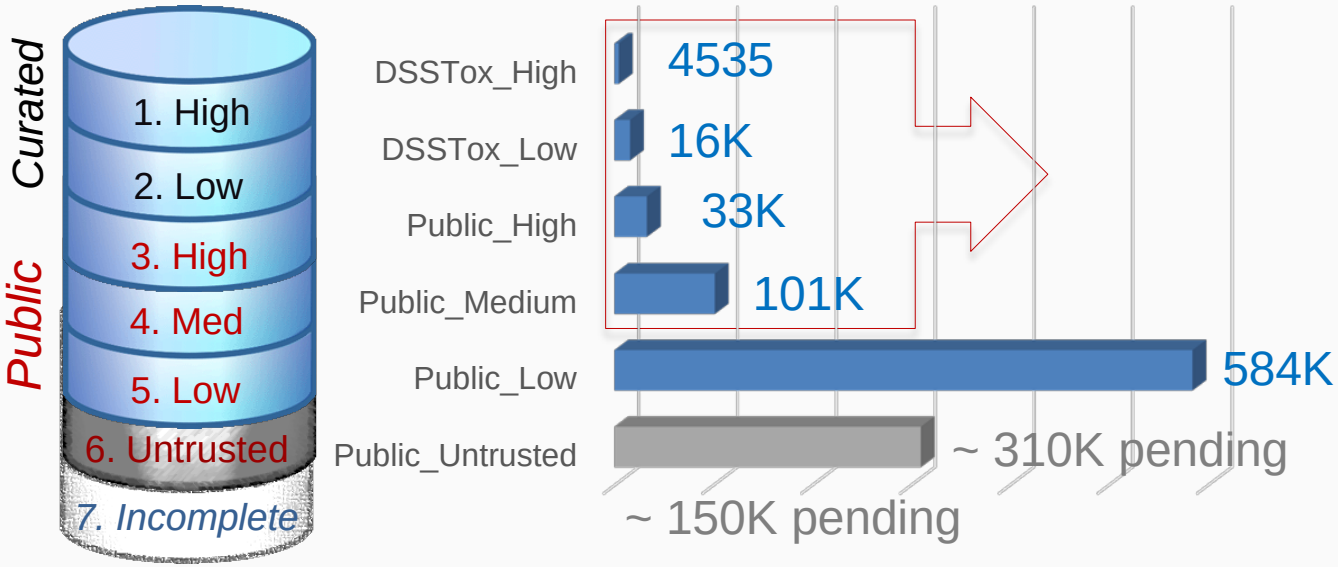
EPA researchers develop and use rapid exposure estimates to predict potential exposure for thousands of chemicals.

- High-throughput toxicokinetics data: It is important to link the external dose of a chemical to an internal blood or tissue concentration, this process is called toxicokinetics. EPA researchers measure the critical factors that determine the distribution

Underpinning with chemicals



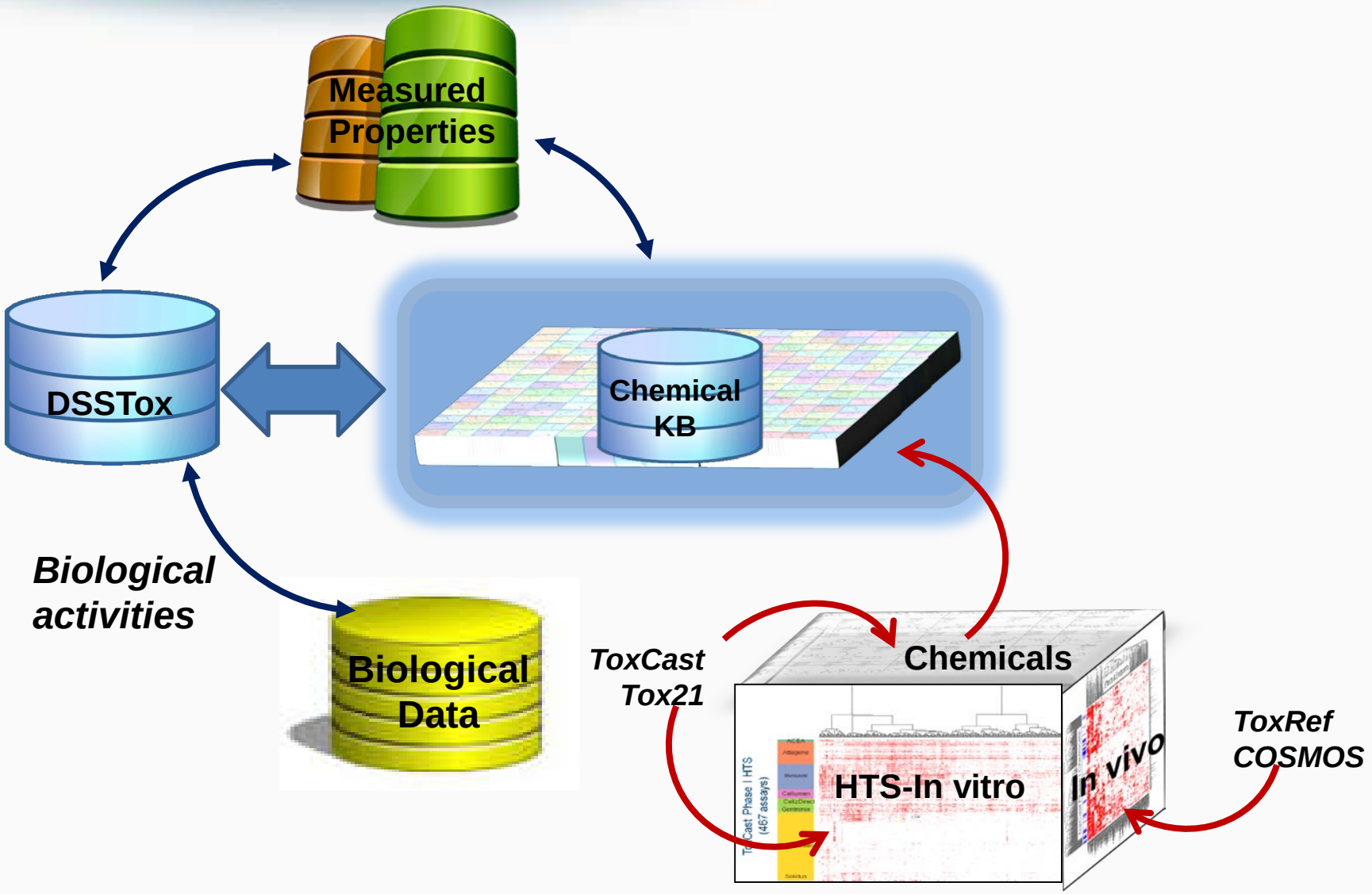
Approximately 17 Years of Data… Growing with daily curation



QC Levels

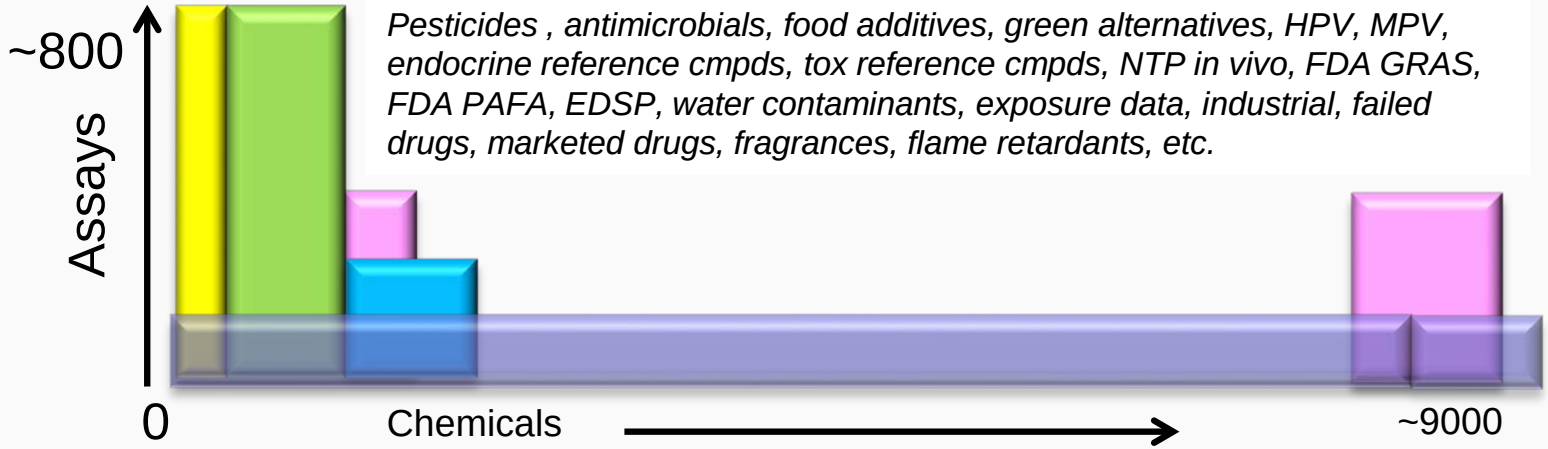
DSSTox_High:	Hand curated and **validated**
DSSTox_Low:	Hand curated and confirmed using multiple public sources
Public_High:	Extracted from EPA SRS and confirmed to have no conflicts in ChemID and PubChem
Public_Medium:	Extracted from ChemID and confirmed to have no conflicts in PubChem
Public_Low: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

Integrating *in vitro* and *in vivo* data

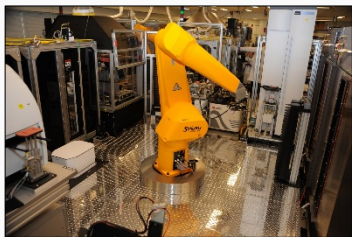


TEN YEARS of Assay Measurements: ToxCast & Tox21

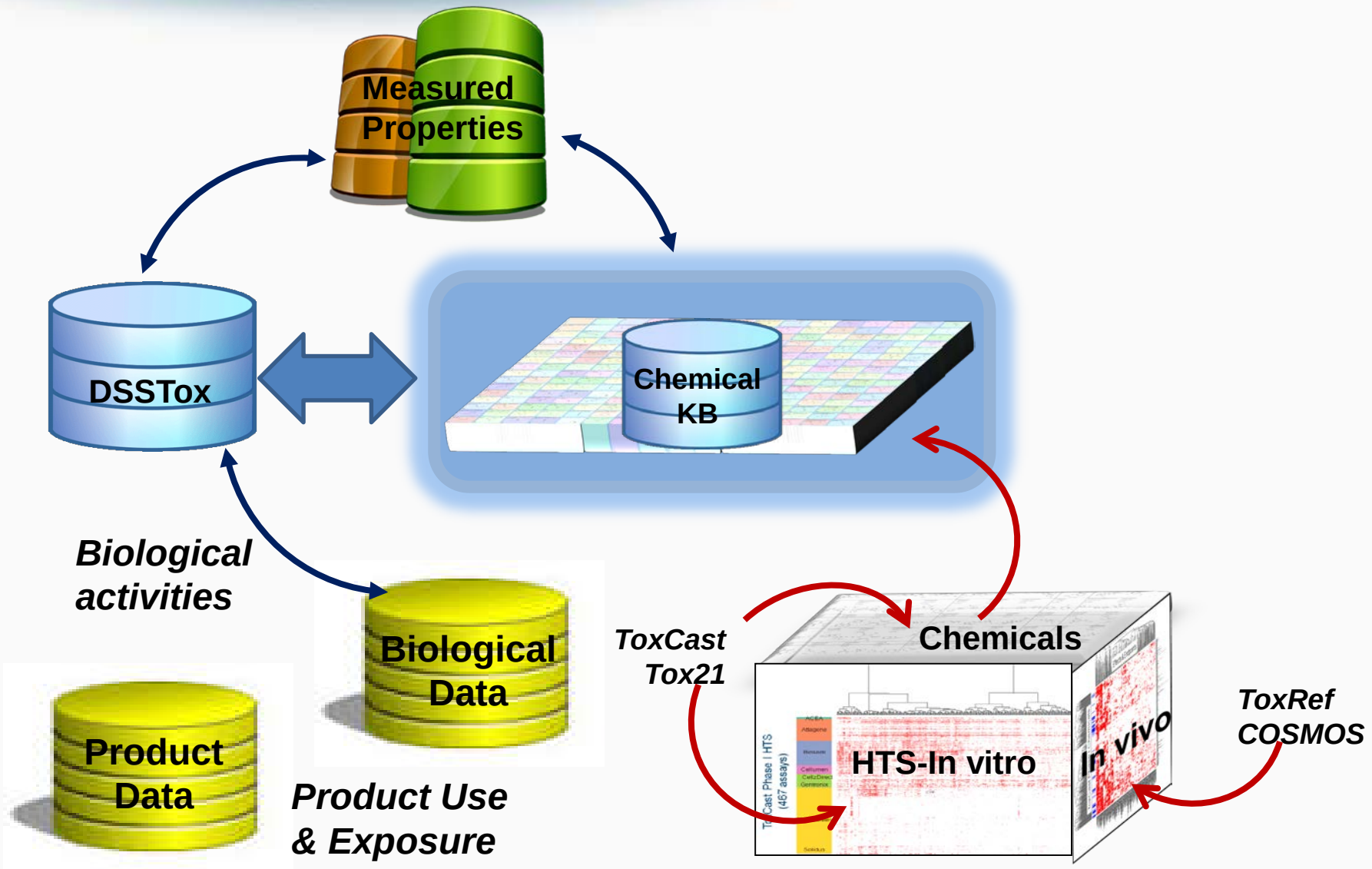
Set	Chemicals	Assays	Endpoints	Completion
ToxCast Phase I	293	~600	~700	2011
ToxCast Phase II	767	~600	~700	03/2013
ToxCast E1K	800	~50	~120	03/2013
ToxCast Phase III	~900	~300	~300	In progress
Tox21	~9000	~80	~150	In progress



Pesticides , antimicrobials, food additives, green alternatives, HPV, MPV, endocrine reference cmpds, tox reference cmpds, NTP in vivo, FDA GRAS, FDA PAFA, EDSP, water contaminants, exposure data, industrial, failed drugs, marketed drugs, fragrances, flame retardants, etc.



Adding Product Use and Exposure



High Throughput Measurement to Identify **Exposure**

EPA United States Environmental Protection Agency

EPA US Environmental Protection Agency

Español | 中文: 繁體版 | 中文: 简体版 | Tiếng Việt | 한국어

Learn the Issues | Science & Technology | Laws & Regulations | About EPA

Search EPA.gov

Related Topics: Safer Chemicals Research

Contact Us Share

Rapid Chemical Exposure and Dose Research

EPA is responsible for ensuring the safety of thousands of chemicals. Quantitative exposure data are available for only a small fraction of registered chemicals. This type of exposure data is needed to thoroughly evaluate chemicals for potential risks to humans, wildlife and ecosystems. EPA is developing innovative methods to develop exposure estimates for thousands of chemicals to better protect human health and the environment. These innovative methods are called rapid exposure and dose assessments.

Rapid Exposure Predictions

Rapid, also called high-throughput, exposure predictions or ExpoCast provide rapid exposure estimates for thousands of chemicals. ExpoCast quickly and efficiently looks at multiple routes of exposure to provide exposure estimates. ExpoCast uses and enhances two well-known exposure models to estimate chemical exposure.

Pictured Above: Farfield Exposure Examples

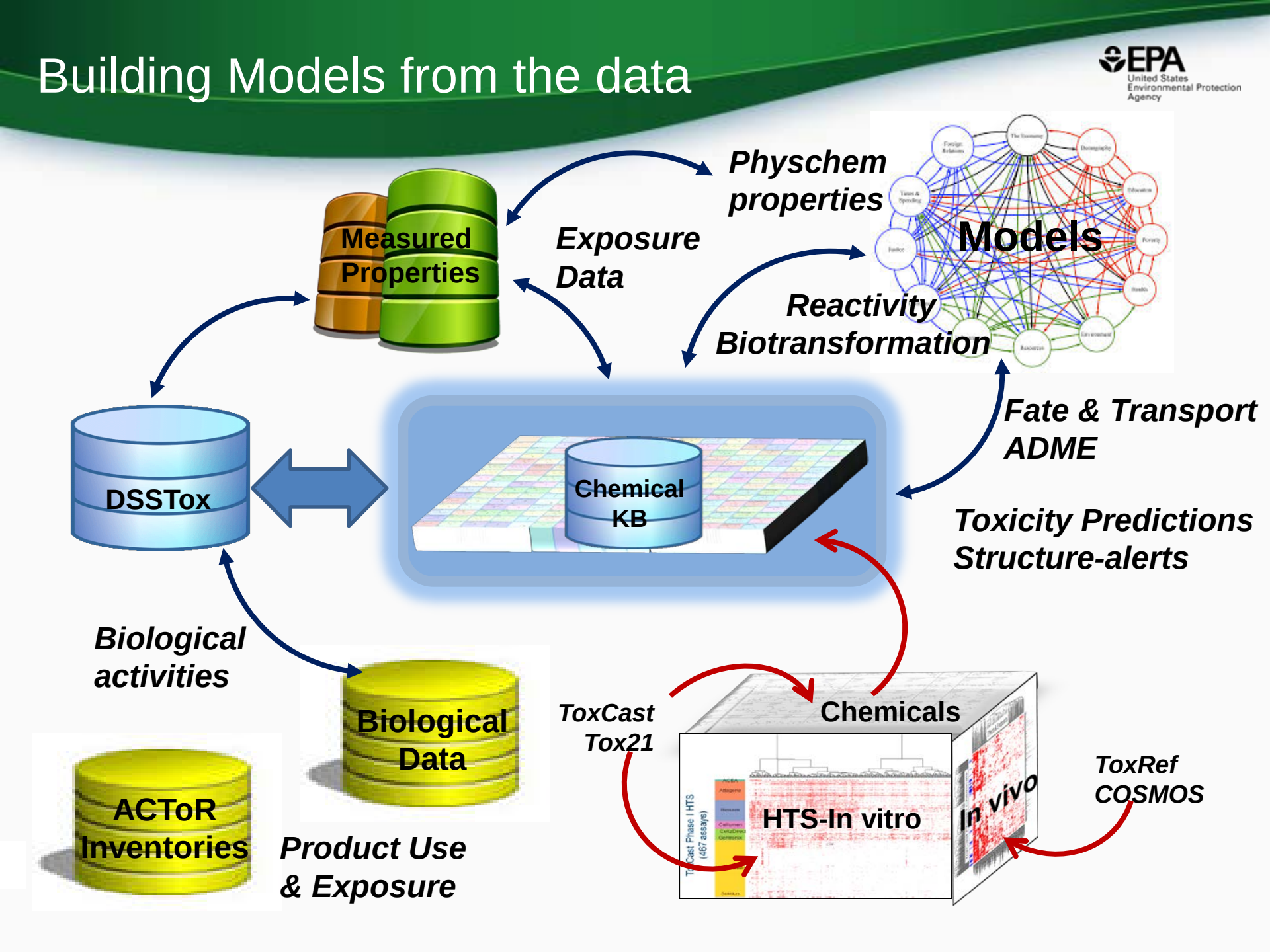
› Farfield Exposure Models

› Nearfield Exposure Models

Evaluating High-throughput Exposure Predictions

EPA is currently evaluating the effectiveness of high-throughput exposure models

Building Models from the data

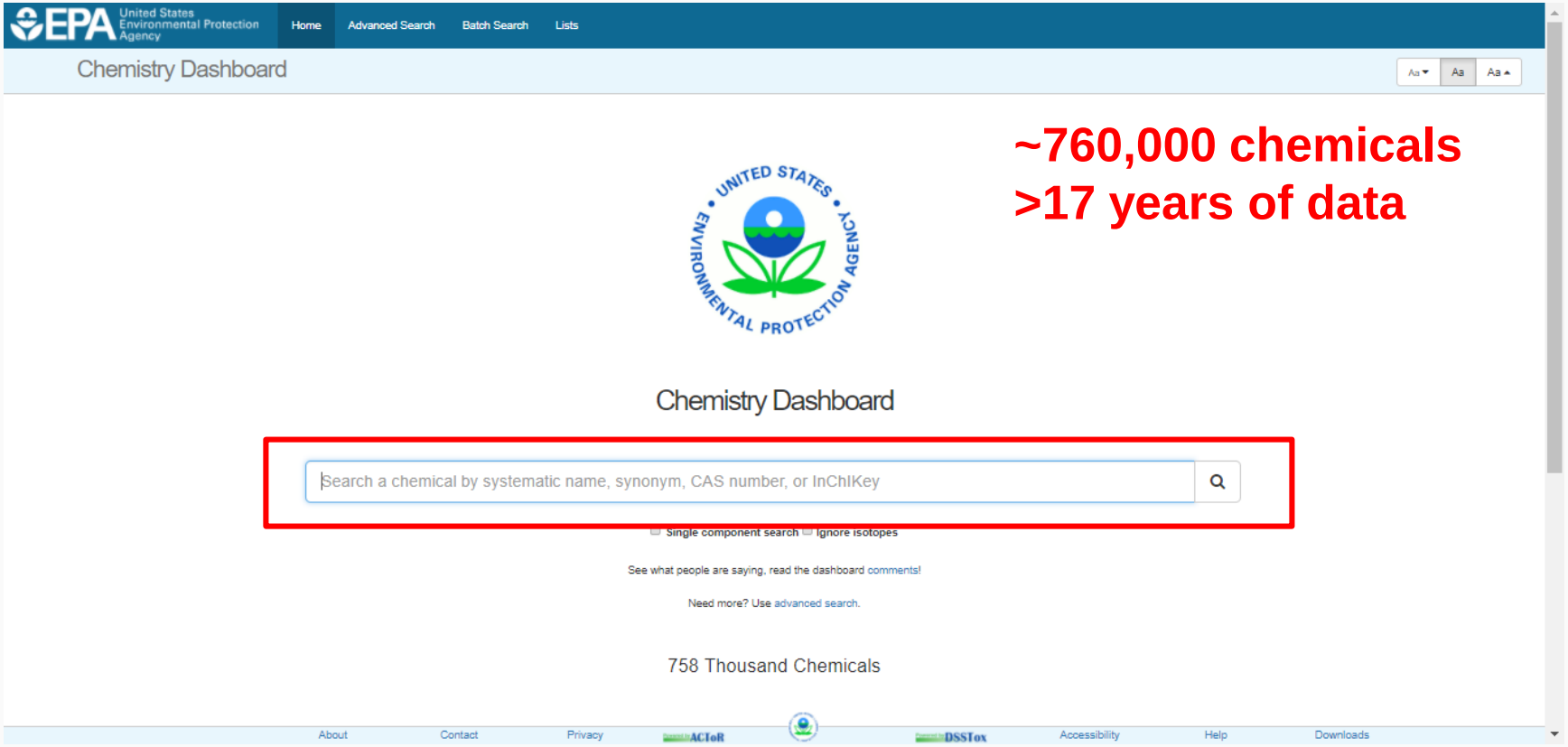


What we have learned…

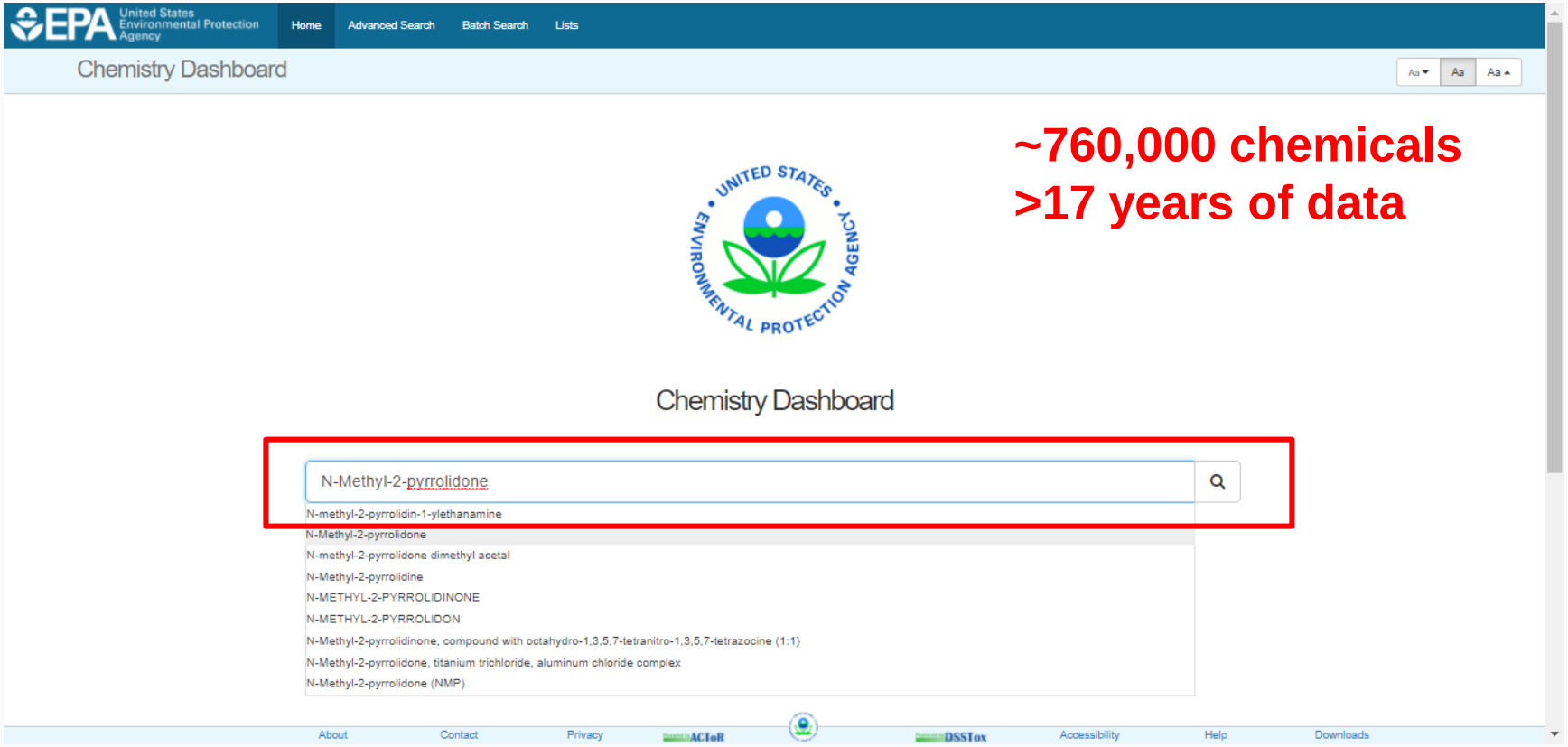
- Data curation, standardization and versioning is **essential**
- **Prototype** application development suffices **for research** projects
- **Production apps** need managed processes
- ODOSOS (Open Data, Open Source and Open Standards) endows many benefits

- We are building: CompTox Chemistry Dashboard

Comptox Chemistry Dashboard
https://comptox.epa.gov



Comptox Chemistry Dashboard
https://comptox.epa.gov

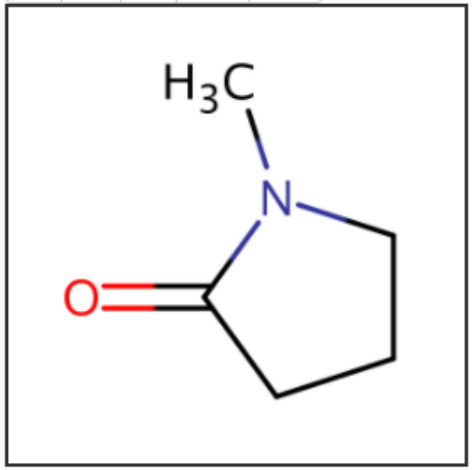


Chemical Page

N-Methyl-2-pyrrolidone

872-50-4 | DTXSID6020856

Searched by Synonym from Valid Source: Found 1 result for 'N-METHYLPYRROLIDONE'.



- Wikipedia
- Intrinsic Properties
  - **Molecular Formula:** C5H9NO — Find All Chemicals
  - **Average Mass:** 99.133 g/mol
  - **Monoisotopic Mass:** 99.068414 g/mol
- Structural Identifiers
- Related Compounds
- Presence in Lists
- Record Information

Chemical Properties | Env. Fate/Transport | Toxicity Values (Beta) | ADME (Beta) | Exposure | Bioassays | Similar Molecules (Beta) | Synonyms | Literature | External Links

Chemical Properties

EPA United States Environmental Protection Agency

- Summary
- LogP: Octanol-Water
- Water Solubility
- Density
- Flash Point
- Melting Point
- Boiling Point
- Surface Tension
- Thermal Conductivity
- Vapor Pressure
- Viscosity
- LogKoa: Octanol-Air

Download as: TSV | Excel | SDF

Property	Average		Median		Range		Unit
	Experimental	Predicted	Experimental	Predicted	Experimental	Predicted	
LogP: Octanol-Water	-0.380 (1)	-0.329 (5)	-0.380	-0.329	-0.380	-0.494 to -0.110	-
Water Solubility	10.1 (1)	6.68 (4)	10.1	6.68	10.1	1.48 to 12.8	mol/L
Density	-	1.02 (2)	-	1.02	-	1.01 to 1.03	g/cm^3
Flash Point	-	75.7 (2)	-	75.7	-	65.2 to 86.1	°C
Melting Point	-23.8 (8)	1.32e-01 (4)	-24.0	1.32e-01	-24.0 to -23.0	-10.2 to 25.9	°C
Boiling Point	203 (6)	199 (5)	204	199	202 to 204	191 to 202	°C
Surface Tension	-	33.8 (1)	-	33.8	-	-	dyn/cm
Thermal Conductivity	-	158 (1)	-	158	-	-	mW/(m*K)
Vapor Pressure	3.45e-01 (1)	5.21e-01 (4)	3.45e-01	5.21e-01	3.45e-01	1.71e-01 to 9.99e-01	mmHg
Viscosity	-	3.61 (1)	-	3.61	-	-	cP
LogKoa: Octanol-Air	-	3.84 (1)	-	3.84	-	-	-
Henry's Law	3.20e-09 (1)	9.15e-09 (1)	3.20e-09	9.15e-09	-	-	atm-m3/mole
Index of Refraction	-	1.47 (1)	-	1.47	-	-	-

Available Properties

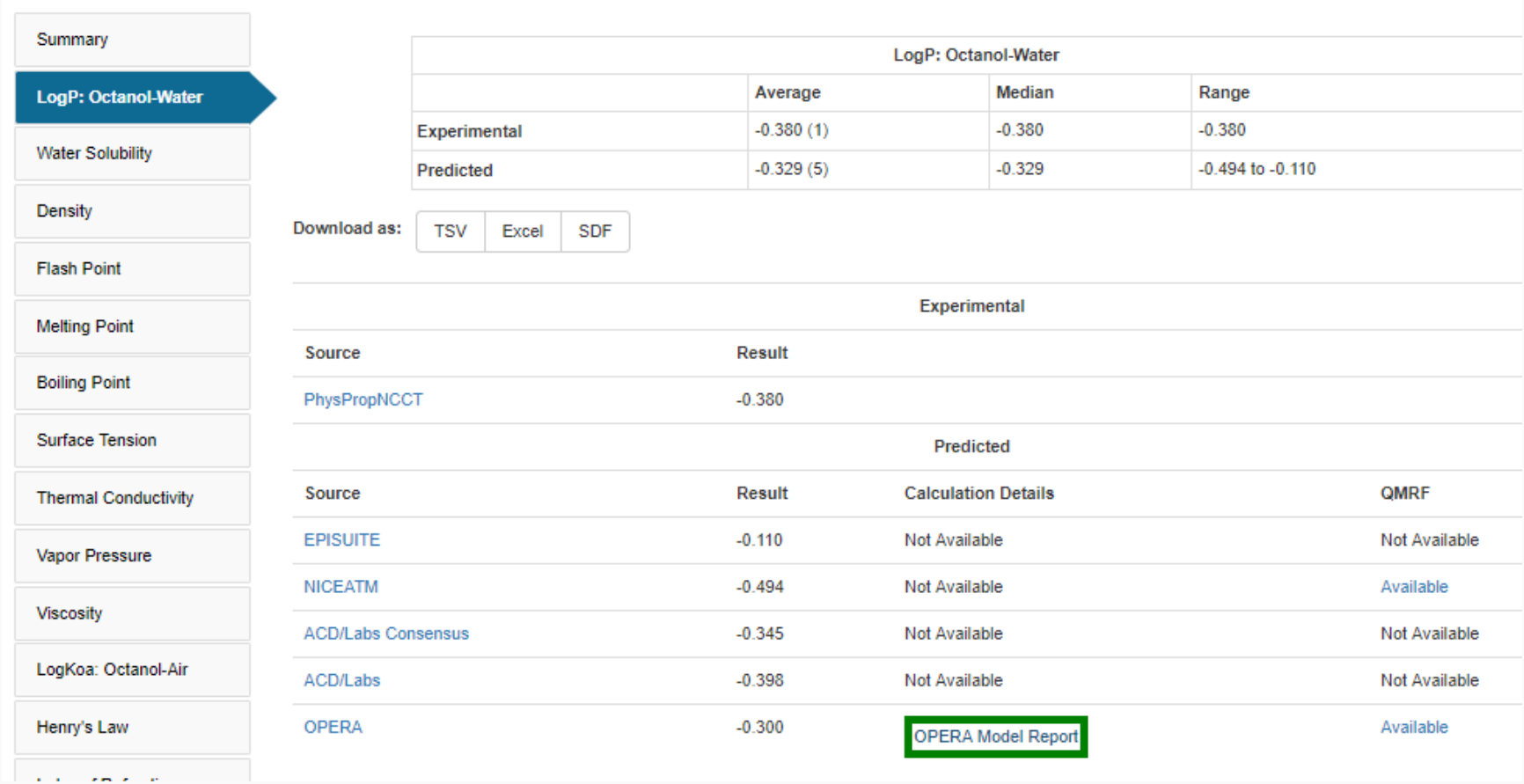
- Solubility
- Melting Point
- Boiling Point
- LogP (Octanol-water partition coefficient)
- Atmospheric Hydroxylation Rate
- LogBCF (Bioconcentration Factor)
- Biodegradation Half-life
- Henry's Law Constant
- Fish Biotransformation Half-life
- LogKOA (Octanol/Air Partition Coefficient)
- LogKOC (Soil Adsorption Coefficient)
- Vapor Pressure

- Data and models are updated ~90 days

Developing “NCCT Models”

- Our approach to modeling:
  - Obtain **high quality** training sets
  - Apply appropriate modeling approaches
  - **Validate** performance of models
  - Define the applicability domain and model limitations
  - Use models to predict properties across our full datasets
  - Release as **Open Data and Open Models**

Multiple Prediction Algorithms
Transparency is Important



- Summary
- LogP: Octanol-Water
- Water Solubility
- Density
- Flash Point
- Melting Point
- Boiling Point
- Surface Tension
- Thermal Conductivity
- Vapor Pressure
- Viscosity
- LogKoa: Octanol-Air
- Henry's Law

LogP: Octanol-Water			
	Average	Median	Range
Experimental	-0.380 (1)	-0.380	-0.380
Predicted	-0.329 (5)	-0.329	-0.494 to -0.110

Download as: TSV | Excel | SDF

Experimental

Source	Result
PhysPropNCCT	-0.380

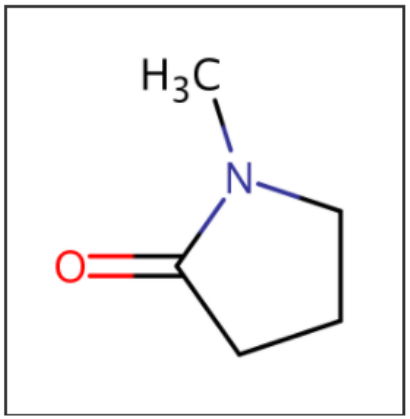
Predicted

Source	Result	Calculation Details	QMRF
EPISUITE	-0.110	Not Available	Not Available
NICEATM	-0.494	Not Available	Available
ACD/Labs Consensus	-0.345	Not Available	Not Available
ACD/Labs	-0.398	Not Available	Not Available
OPERA	-0.300	OPERA Model Report	Available

OPERA Models: LogP: Octanol-Water

N-Methyl-2-pyrrolidone

872-50-4 | DTXSID6020856



Model Results

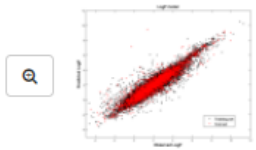
Predicted value: -0.300

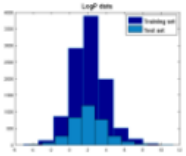
Global applicability domain: Inside

Local applicability domain index: 0.88

Confidence level: 0.81

Model Performance



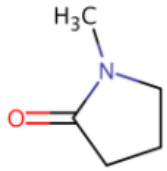


Weighted KNN model

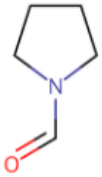
QMRF

5-fold CV (75%)		Training (75%)		Test (25%)	
Q2	RMSE	R2	RMSE	R2	RMSE
0.85	0.69	0.86	0.67	0.86	0.78

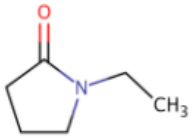
Nearest Neighbors from the Training Set



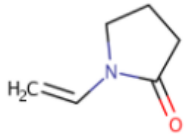
N-Methyl-2-pyrrolidone
Measured: -0.380
Predicted: -0.300



Prolinal
Measured: -0.320
Predicted: -0.226

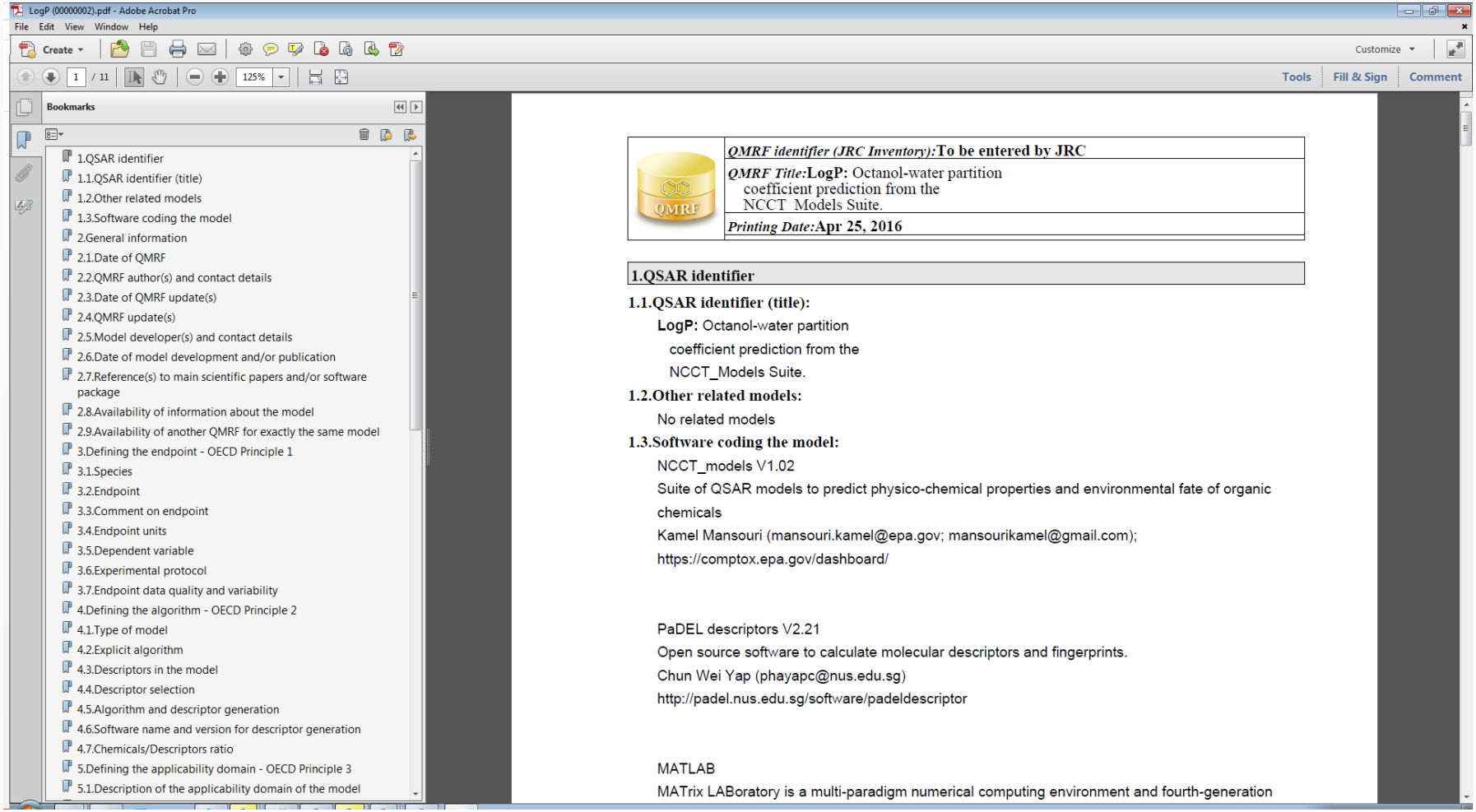


1-Ethyl-2-pyrrolidinone
Measured: -0.0400
Predicted: -0.211



N-Vinyl-2-pyrrolidone
Measured: 0.37
Predicted: 2.65e-01

QSAR Modeling Reporting Format



Workflow Details and Data

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

Enter keywords, authors, DOI etc.

258
Views
4
CrossRef citations
16
Altmetric

Articles

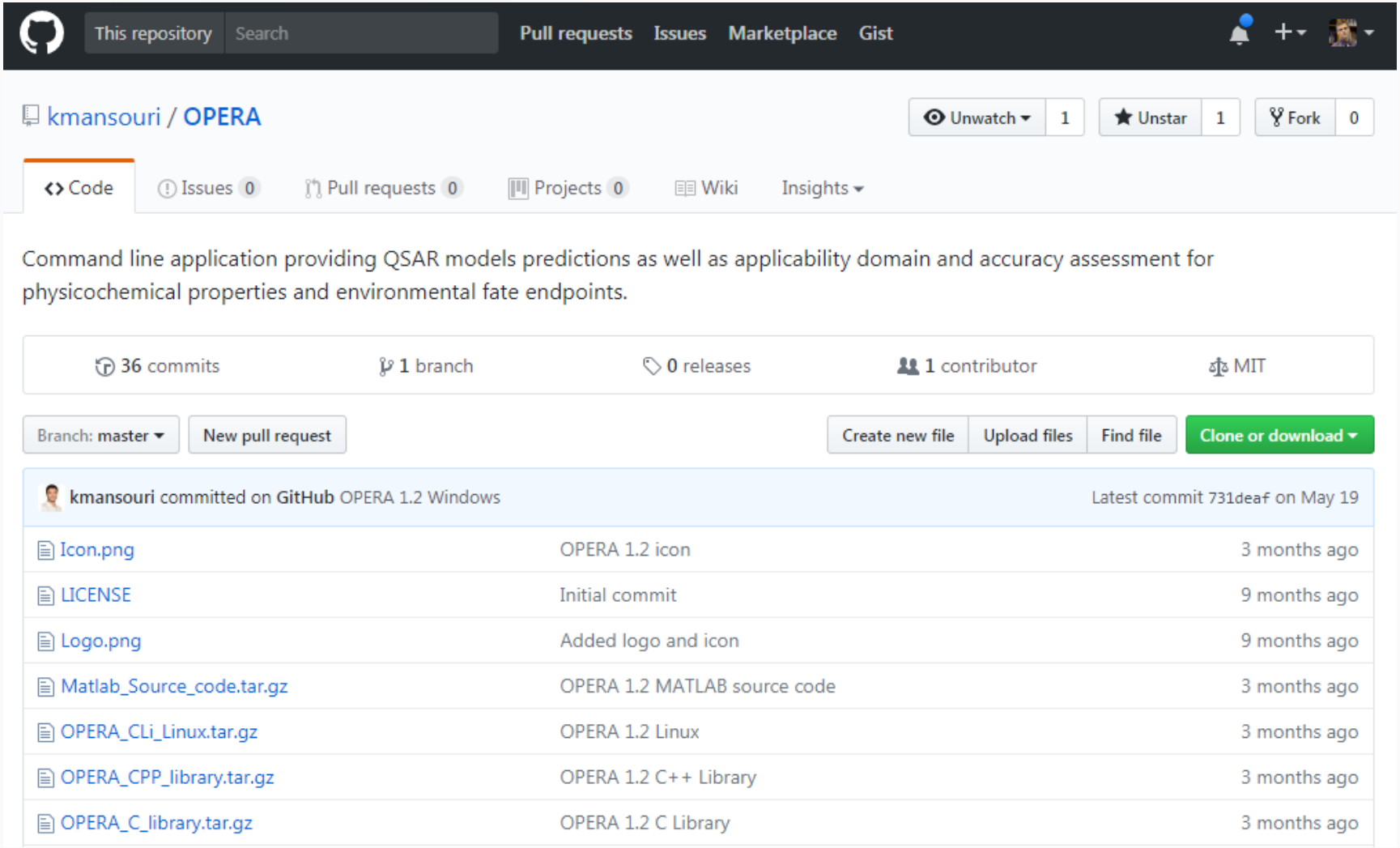
An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling $

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

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

Download citation http://dx.doi.org/10.1080/1062936X.2016.1253611 Check for updates

OPERA on GitHub



https://github.com/kmansouri/OPERA.git

Toxicity Values

EPA United States Environmental Protection Agency

- **Bioavailability Metric**
- Exposure Limit
- Point Of Departure
- Regulatory Toxicity Value
- Effect Level
- Misc Hazard Information
- Screening Level
- Uncertainty Factor

Download as: TSV | Excel

Type	Subtype	Value	Units	Study Type	Exposure Route	Study Duration	Species	Media	Details	Source
TD50	-	20.7	mM/kg-day	-	-	-	-	-	DSSTox C...	ACToR
TD50	-	2050	mg/kg-day	-	-	-	-	-	DSSTox C...	ACToR
LEL	systemic	619	mg/kg-day	subchronic	oral	subchronic	mouse	-	Study ID: ...	ToxRefDB
NEL	systemic	277	mg/kg-day	subchronic	oral	subchronic	mouse	-	Study ID: ...	ToxRefDB
LEL	systemic	25.0	mg/kg-day	subchronic	oral	subchronic	dog	-	Study ID: ...	ToxRefDB
NEL	systemic	25.0	mg/kg-day	subchronic	oral	subchronic	dog	-	Study ID: ...	ToxRefDB
LEL	systemic	173	mg/kg-day	chronic	oral	chronic	mouse	-	Study ID: ...	ToxRefDB
NEL	systemic	115	mg/kg-day	chronic	oral	chronic	mouse	-	Study ID: ...	ToxRefDB
LEL	systemic	678	mg/kg-day	chronic	oral	chronic	rat	-	Study ID: ...	ToxRefDB
NEL	systemic	283	mg/kg-day	chronic	oral	chronic	rat	-	Study ID: ...	ToxRefDB
LEL	systemic	1230	mg/kg-day	subacute	oral	subacute	rat	-	Study ID: ...	ToxRefDB
NEL	systemic	493	mg/kg-day	subacute	oral	subacute	rat	-	Study ID: ...	ToxRefDB
LEL	systemic	2130	mg/kg-day	subacute	oral	subacute	mouse	-	Study ID: ...	ToxRefDB
NEL	systemic	920	mg/kg-day	subacute	oral	subacute	mouse	-	Study ID: ...	ToxRefDB

Product Composition Details

EPA United States Environmental Protection Agency

- Product & Use Categor...
- Chemical Weight Fra...
- Chemical Functional Use
- Monitoring Data
- Exposure Predictions

Chemical Weight Fractions

Download as: TSV | Excel

Product Name	Product Use Category	Minimum Weight Fraction	Maximum Weight Fraction	Data Type	Source
10-02199- calico tip & ...	personal care: nail poli...	0.01	0.05	MSDS	Retail Product Categori...
6095-6096 minwax wo...	home maintenance: fin...	0.01	0.01	MSDS	Retail Product Categori...
6095/6096 minwax wo...	home maintenance: fin...	0.01	0.01	MSDS	Retail Product Categori...
ab artificial nail remove...	personal care: nail poli...	0.01	0.03	MSDS	Retail Product Categori...
artificial nail remover 7...	personal care: nail poli...	0.01	0.05	MSDS	Retail Product Categori...
calico tip & glue remov...	personal care: nail poli...	0.01	0.05	MSDS	Retail Product Categori...
citristrip canadian strip...	home maintenance: str...	0.65	0.7	MSDS	Retail Product Categori...
citristrip stripping gel q...	home maintenance: str...	0.4	0.55	MSDS	Retail Product Categori...
citristrip stripping gel q...	home maintenance: str...	-	-	MSDS	Retail Product Categori...

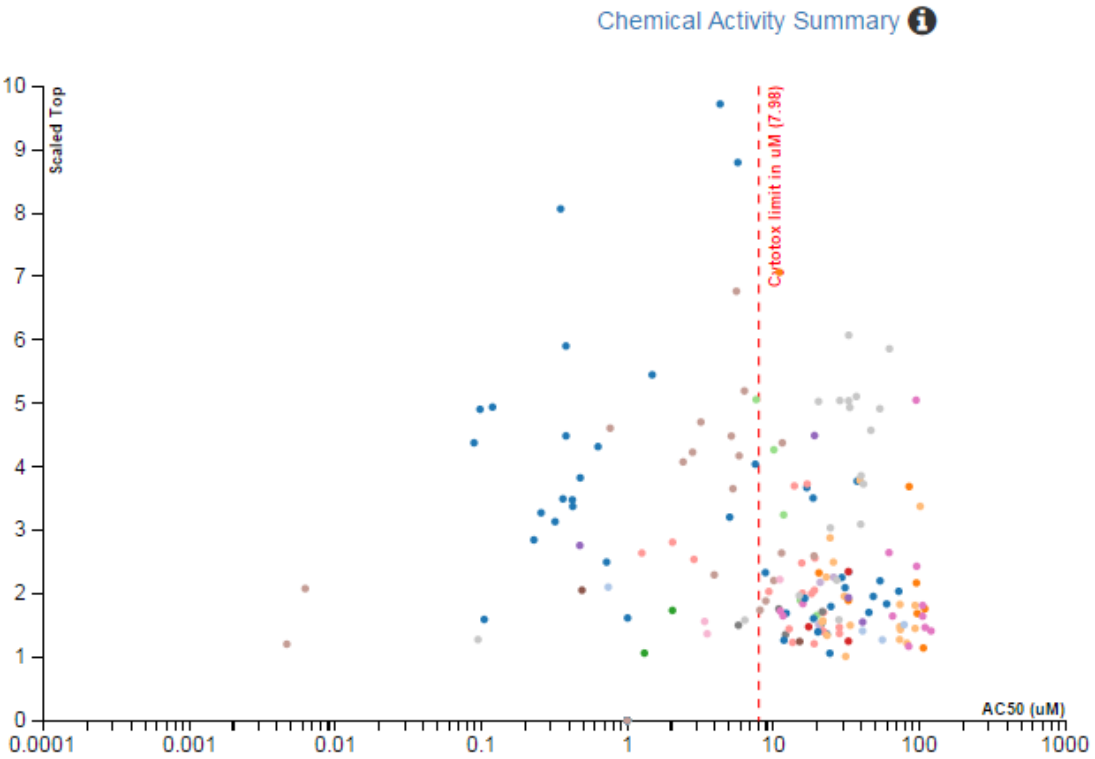
ToxCast and Tox21 Bioassays

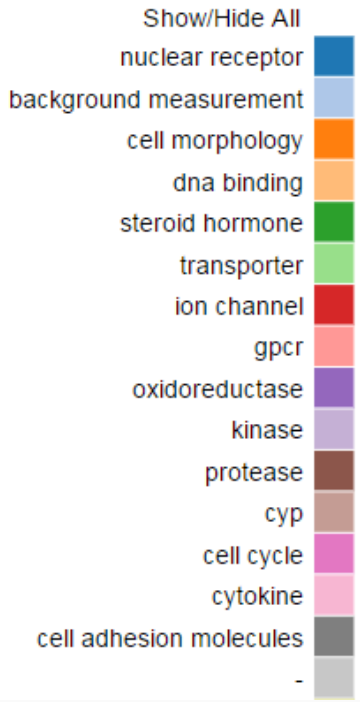
EPA United States Environmental Protection Agency

ToxCast

PubChem

ScrubChem (Beta)





Chemical Properties | Env. Fate/Transport | Synonyms | External Links | Toxicity Values (Beta) | Exposure | Bioassays | Similar Molecules (Beta) | Literature | Comments

ToxCast and Tox21 Bioassays

EPA United States Environmental Protection Agency

Download as: TSV | Excel Show: Inactive | Background

Assay Name	Hit Call	Top	Scaled Top	AC50	log AC50 ↓	Intended Target Family
APR_Hepat_CellLoss_48hr_dn	ACTIVE	1.41	1.41	120	2.08	cell cycle
APR_HepG2_MitoMass_24h_dn	ACTIVE	0.874	1.76	109	2.04	cell morphology
APR_HepG2_OxidativeStress_24h_up	ACTIVE	1.20	1.47	110	2.04	cell cycle
APR_Hepat_DNADamage_48hr_up	ACTIVE	1.84	1.14	107	2.03	cell morphology
APR_HepG2_CellLoss_24h_dn	ACTIVE	1.20	1.81	106	2.02	cell cycle
APR_HepG2_OxidativeStress_72h_up	ACTIVE	1.80	1.65	106	2.02	cell cycle
ATG_HSE_CIS_up	ACTIVE	1.59	3.38	102	2.01	dna binding

Names and Identifiers

EPA United States Environmental Protection Agency

Found 40 synonyms

Legend: **Valid Synonyms** *Good Synonyms* Other Synonyms Copy all Synonyms

N-Methyl-2-pyrrolidone
1-Methylpyrrolidin-2-one
2-Pyrrolidinone, 1-methyl-
872-50-4 Active CAS-RN
2-Pyrrolidinone, 1-methyl-
1-Methyl-2-pyrrolidinone
1-Methyl-2-pyrrolidon
1-Methyl-2-pyrrolidone
1-Methyl-5-pyrrolidinone
1-Methylazacyclopentan-2-one
1-Methylpyrrolidone
1-metil-2-pirrolidona
2-Pyrolidinone, 1-methyl-
2-Pyrrolidone, 1-methyl-
Microposit 2001
M-Pyrol

Chemical Properties | Env. Fate/Transport | Toxicity Values (Beta) | ADME (Beta) | Exposure | Bioassays | Similar Molecules (Beta) | Synonyms | Literature | External Links

Integrated Literature Searching

EPA United States Environmental Protection Agency

- Google Scholar
- PubMed Abstract S...
- PubChem Articles
- PubChem Patents

Select Term: Hazard **AND** RfD OR reference dose

Google "RfD OR reference dose" AND "Hazard" AND "872-50-4" OR "N-Methyl-2-py

Scholar About 22 results (0.05 sec)

Articles
Case law
My library

Any time
Since 2017
Since 2016
Since 2013
Custom range...

Sort by relevance
Sort by date

include patents
include citations

Create alert

Solvent Substitution: An Analysis of Comprehensive Hazard Screening Indices
M Debia, D Bégin, M Gérin - Journal of occupational and ..., 2011 - oeh.tandfonline.com
... Reliability and weighting factors were not used. Armenti and Moure-Eraso(22) used the same FHS index, using six **hazard** categories, to compare replacement options. ... **872-50-4 N-Methyl-2-pyrrolidone** 45.4 13 1.46E-03 11 4.55 6 20 4 ...
Related articles All 4 versions Cite Save

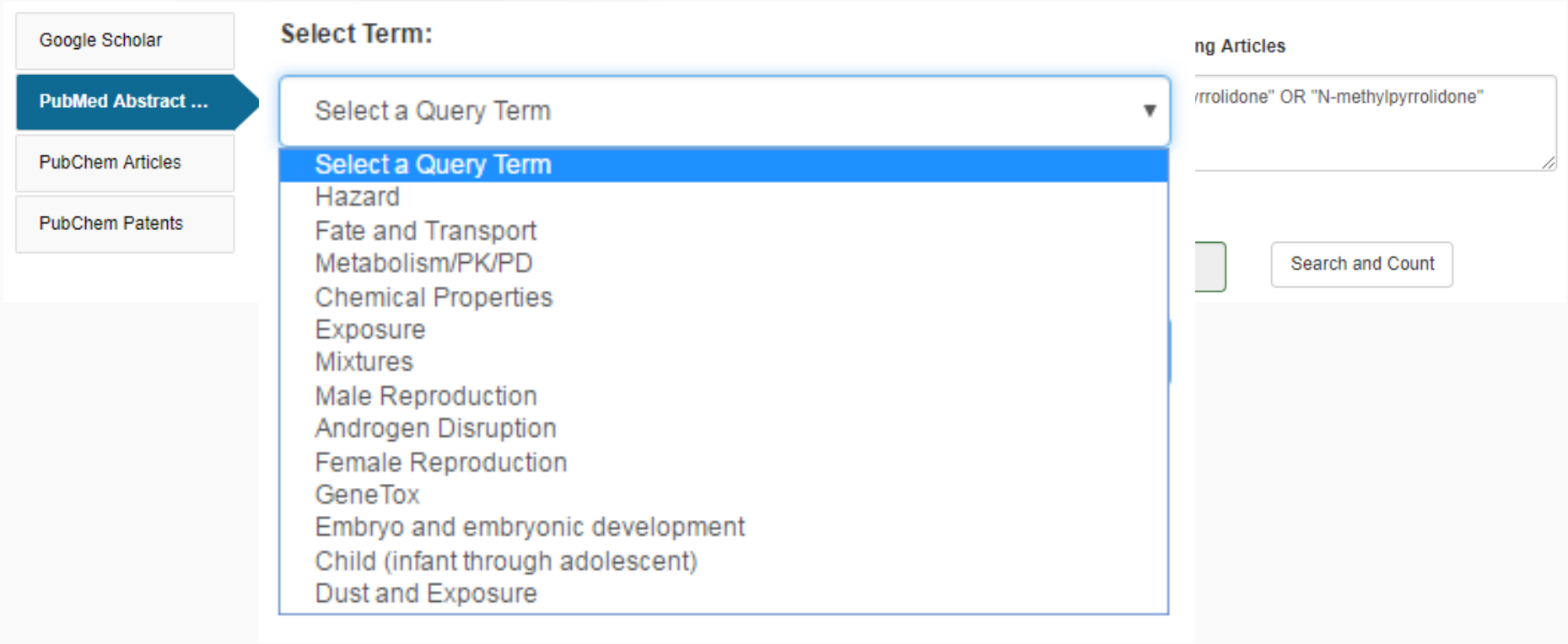
Effects of 'inactive'ingredients on bees
CA Mullin - Current Opinion in Insect Science, 2015 - Elsevier
... tallow amines, organosilicone ethoxylates and co-solvents such as **N-methyl-2-pyrrolidone** (NMP) [7 ... of tools to environmentally monitor residues and determine their relative **hazard** to bees. ... Defining a benchmark or **reference dose**, particularly for bee viruses [51], for what can ...
Cited by 9 Related articles All 2 versions Web of Science: 6 Cite Save More

Biological monitoring: exquisite research probes, risk assessment, and routine exposure measurement
WM Draper - Analytical chemistry, 2001 - ACS Publications
ADVERTISEMENT. ...
Cited by 19 Related articles All 5 versions Web of Science: 10 Cite Save

Evolution of chemical-specific adjustment factors (CSAF) based on recent international experience; increasing utility and facilitating regulatory acceptance
VS Bhat, ME Meek, M Valcke, C English... - Critical Reviews in ..., 2017 - Taylor & Francis
... As shown in the right side of Figure 2, uncertainty is taken into consideration at all tiers of **hazard** assessment. ... For decades, developing "safe doses" such as the oral **reference dose** (RfD), the inhaled reference concentration (RfC) or the acceptable or tolerable daily intake (ADI ...
All 3 versions Cite Save

Chemical Properties | Env. Fate/Transport | Toxicity Values (Beta) | ADME (Beta) | Exposure | Bioassays | Similar Molecules (Beta) | Synonyms | **Literature** | External Links

Integrated Literature Searching



Integrated Literature Searching

EPA United States Environmental Protection Agency

Google Scholar

PubMed Abstract Sifter

PubChem Articles

PubChem Patents

Edit the Query Before Retrieving Articles

("872-50-4" OR "N-Methyl-2-pyrrolidone" OR "N-methylpyrrolidone") AND (exposure OR near-field OR far-field OR SHEDS[tiab] AND ENVIRONMENTAL MONITORING)

1

0

0	0	0	0	24078144	2013	Biological monitoring and health effects of low-level exposure to N-methyl-2-pyrrolidone: a cross-sectional study.
0	0	0	0	23337464	2013	Biomonitoring of exposure to N-methyl-2-pyrrolidone in workers of the automobile industry.
0	0	0	0	19875680	2009	Quantitative risk analysis for N-methyl pyrrolidone using physiologically based pharmacokinetic and benchmark dos…
0	0	0	0	16897094	2006	Human volunteer study on the influence of exposure duration and dilution of dermally applied N-methyl-2-pyrrolidon…

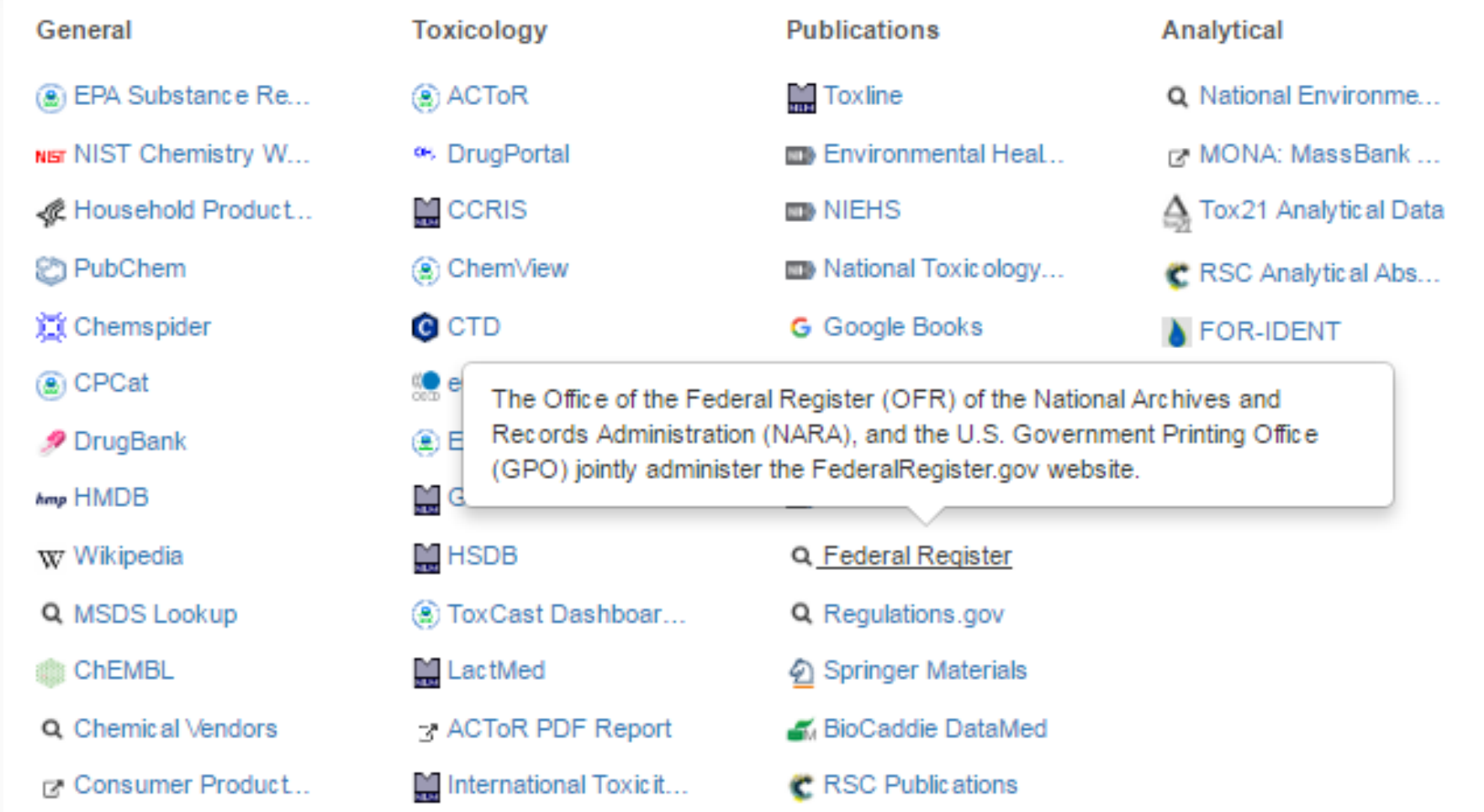
Record: 1 of 16

Title: [Not Available].

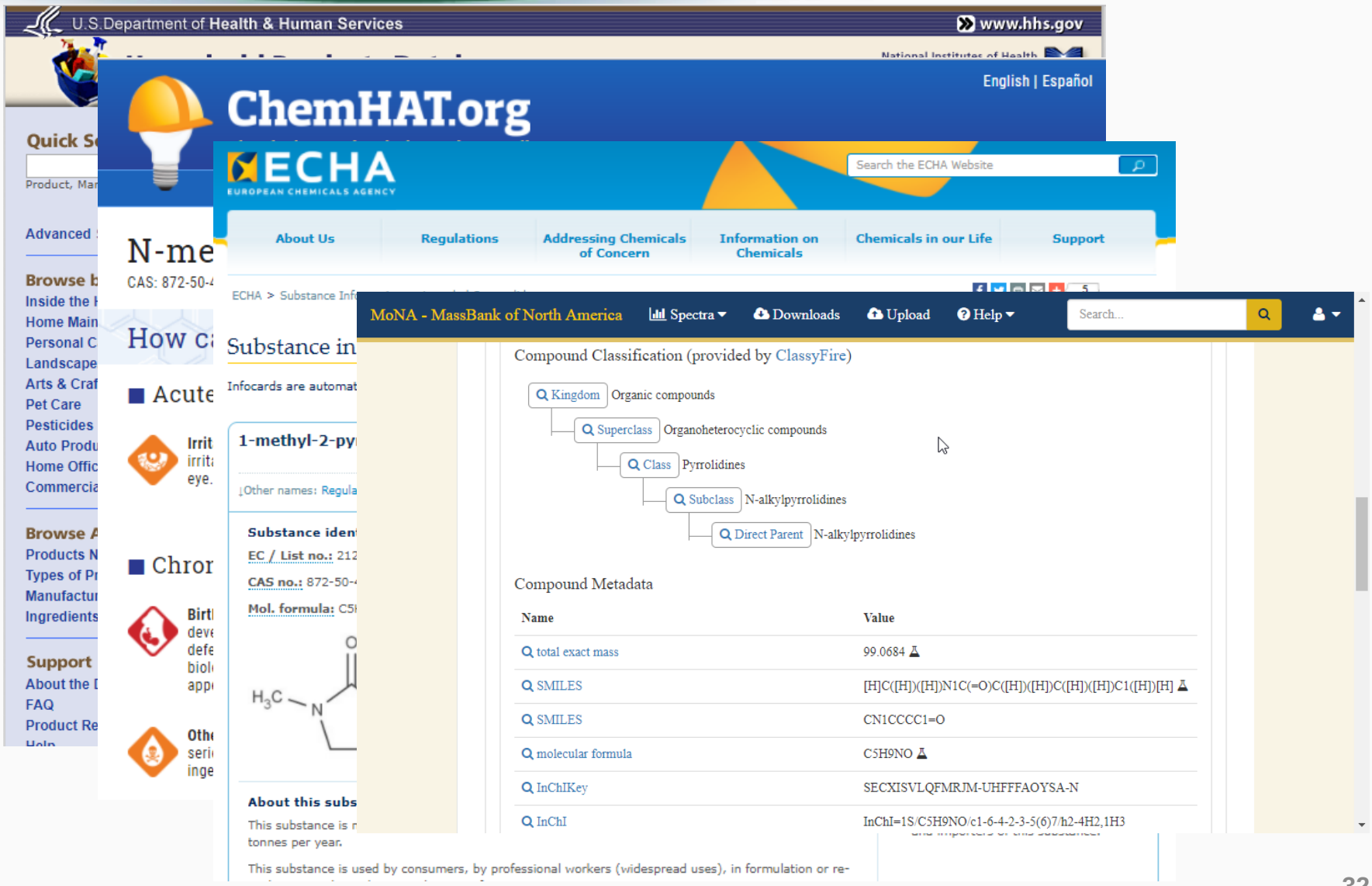
Abstract: N-Ethyl-2-pyrrolidone (NEP), a polar aprotic solvent, is used in many applications as substitute for the structural analogue N-methyl-2-pyrrolidone (NMP), e. g. for surface coatings, in cleaning agents and paint strippers. Monitoring studies indicate that individuals within the general public, without occupational exposure, may be exposed to NEP to an extent, which is comparable to NMP. As NMP, NEP presents a potential health hazard due to its developmental toxicity and teratogenicity. Exposure to NEP can be quantified by the determination of the excretion of its urinary metabolites 5-Hydroxy-N-ethyl-2-pyrrolidone (5-HNEP) and 2-Hydroxy-N-ethylsuccinimide (2-HESI). For the derivation of HBM values, the german Human Biomonitoring Commission (HBM commission) evaluated different toxicological endpoints and finally decided on the BMDL05 and the BMD10 for the endpoint "reduced grasp intensity" of a subchronic feeding study with rats as point of departure (POD) for further procedural steps. The resulting HBM-I and HBM-II values for the sum of the metabolites 5-HNEP and 2-HESI in the urine of children are 10 resp. 25 mg/l and in the urine of adults are 15 resp. 40 mg/l. If the HBM values are exceeded, a check-up will be necessary at first. Measurements above the HBM-II value give cause for concern, especially for pregnant women. Air meas ents to determine the source of exposure can be useful. The possibility of skin absorption from use of

Chemical Properties | Env. Fate/Transport | Toxicity Values (Beta) | ADME (Beta) | Exposure | Bioassays | Similar Molecules (Beta) | Synonyms | **Literature** | External Links

Links to Other Resources



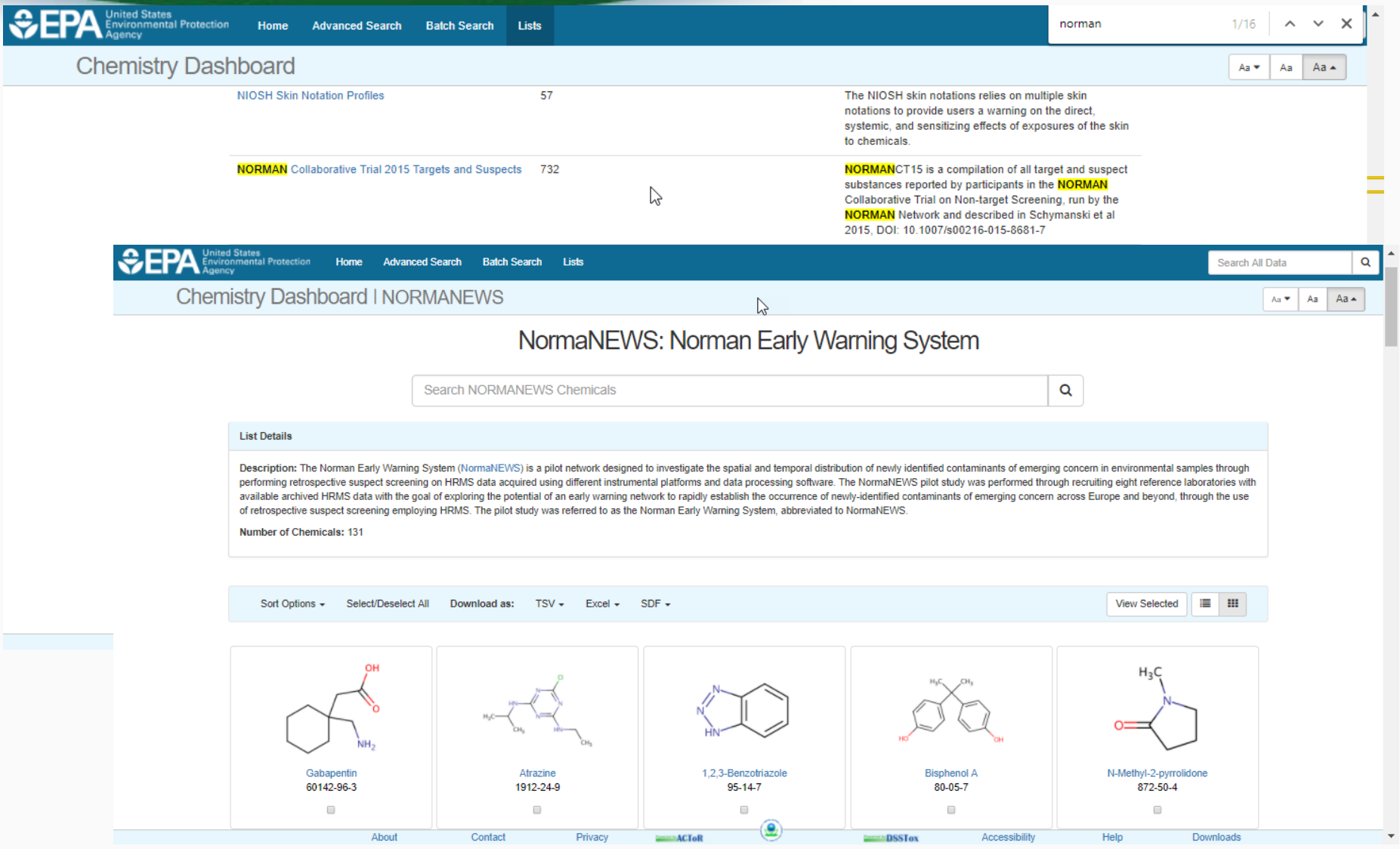
Example External Links…



Accessing Lists of Chemicals

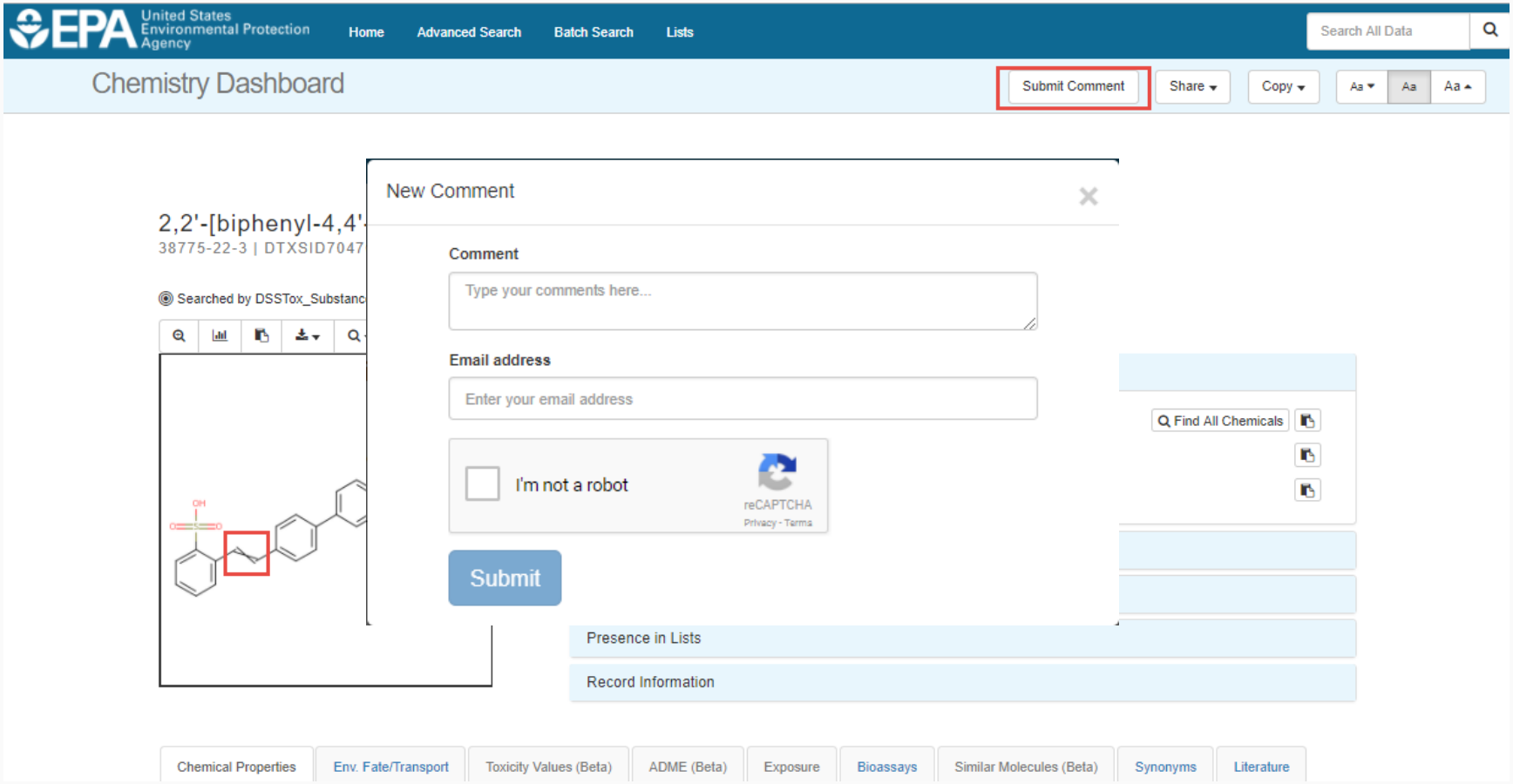
- Build out definitive “lists” of chemicals
  - Algal toxins
  - Poly/perfluorinated chemicals
  - Pesticides
  - Toxcast screening chemical collection
  - Public data sets:
    - NORMAN Network data collections
    - MASSBANK data

The Collection of Lists



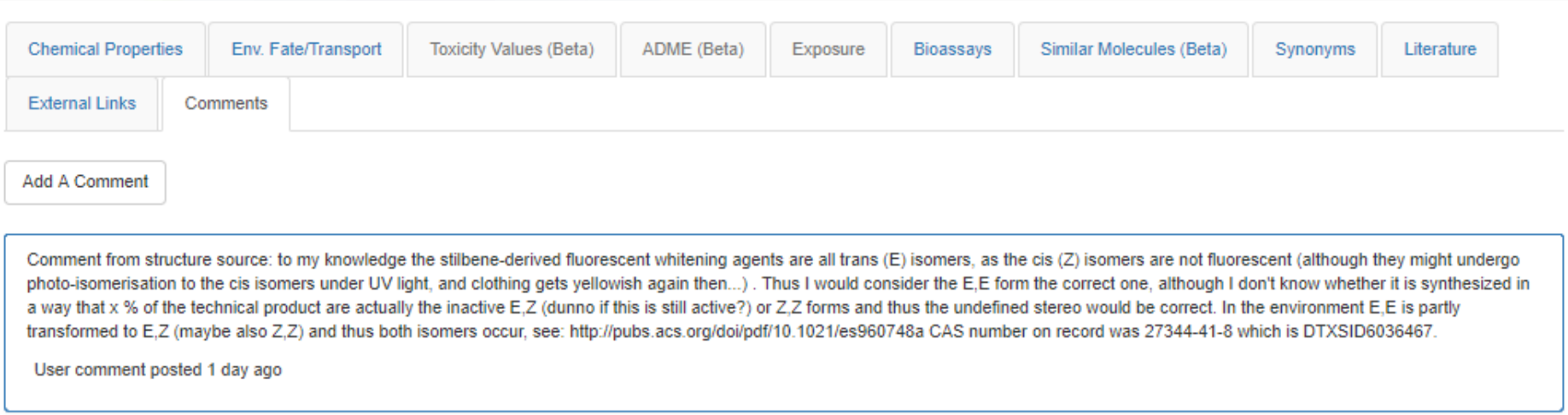
Crowdsourced Curation – HELP!

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



Crowdsourced Curation – HELP!

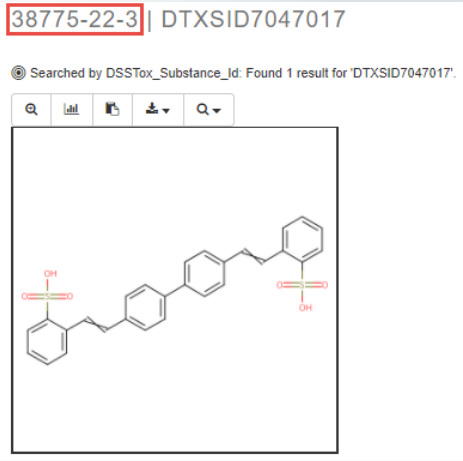
https://comptox.epa.gov/dashboard/comments/public_index



- The Question – is this a mixture of E/Z, E/E or Z/Z? Brighteners should be E/E…

- **Thanks for helping us Emma (& Martin)!**

Curation is laborious work



- So, it's clear what the correct E/Z orientation is yes???

CAS Registry Number 38775-22-3

~160 ~17

$C_{28}H_{22}O_6S_2$

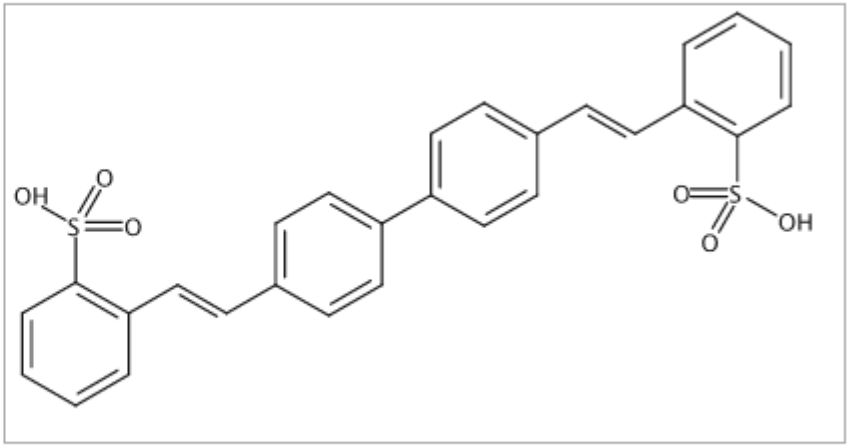
Benzenesulfonic acid, 2,2'-([1,1'-biphenyl]-4,4'-diyldi-2,1-ethenediyl) bis-

Molecular Weight
518.60

Density (Predicted)
Value: 1.414±0.06 g/cm3 | Condition: Temp: 20 °C Press: 760 Torr

pKa (Predicted)
Value: -0.92±0.27 | Condition: Most Acidic Temp: 25 °C

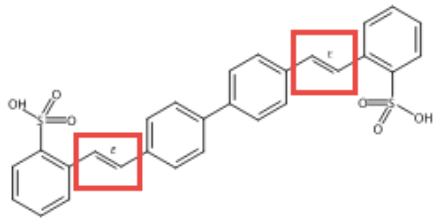
Other Names
2,2'-([1,1'-Biphenyl]-4,4'-diyldi-2,1-ethenediyl)bis[benzenesulfonic acid]
4,4'-Bis(2-sulfostyryl)biphenyl
4,4'-Bis[2-(2-sulfophenyl)ethenyl]biphenyl
C.I. 482200
C.I. Fluorescent Brightener 351



Curation is laborious work

1. **334756-45-5**

~5 ~1



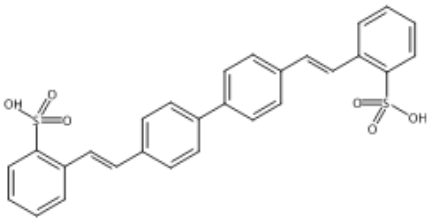
Double bond geometry as shown.

$C_{28}H_{22}O_6S_2$
Benzenesulfonic acid, 2,2'-[(1*E*)-[1,1'-biphenyl]-4,4'-diyldi-2,1-ethenediyl]bis- (9CI)

▸ **Key Physical Properties**

2. **38775-22-3**

~160 ~17



$C_{28}H_{22}O_6S_2$
Benzenesulfonic acid, 2,2'-([1,1'-biphenyl]-4,4'-diyldi-2,1-ethenediyl)bis-

▸ **Key Physical Properties**
Regulatory Information

Curation from YESTERDAY

NORMAN-SusDat: NORMAN Suspect List Exchange Merged Data Table

Reset search results

Show 100 entries

Mol_ID	Name	CAS_RN	ValidationLevel	SMILES
SA10004	garamicidin D	CAS_RN: 1405-97-6	Level 5	CC(C)C[C@@H](NC(=O)[C@H](C)NC(=O)CNC(=O)[C@@H](NC=O)C(C)C)C(=O)N[C@@H](C)C(=O)N[C@H](C(C)C)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)N[C@H](CC(C)C)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)N[C@H](CC(C)C)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)N[C@H](CC(C)C)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)NCCO

Curation from YESTERDAY

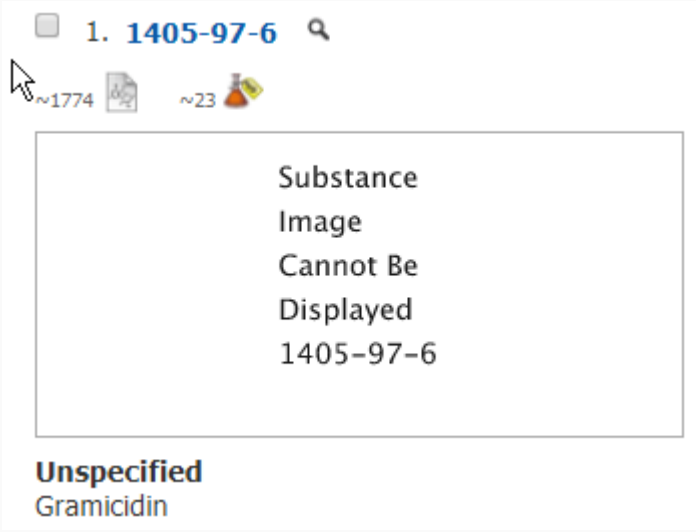
NORMAN-SusDat: NORMAN Suspect List Exchange Merged Data Table

Reset search results

Show 100 entries

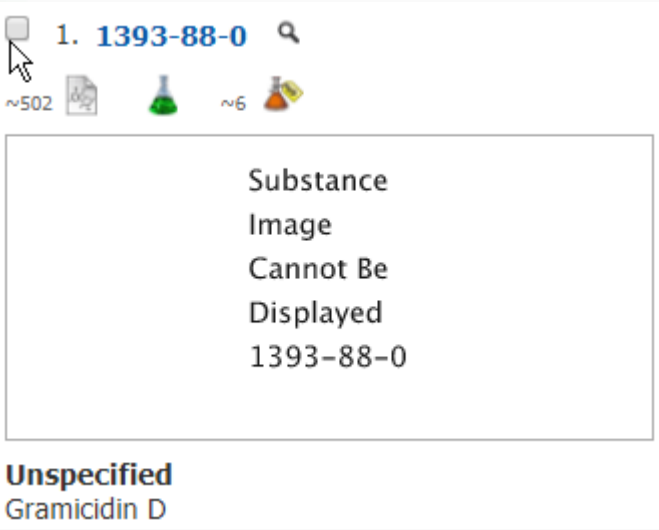
Mol_ID	Name	CAS_RN	ValidationLevel	SMILES
SA10004	garamicidin D	CAS_RN: 1405-97-6	Level 5	CC(C)C[C@@H](NC(=O)[C@H](C)NC(=O)CNC(=O)[C@@H](NC=O)C(C)C)C(=O)N[C@@H](C)C(=O)N[C@H](C(C)C)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](C(C)C)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)N[C@H](CC(C)C)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)N[C@H](CC(C)C)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)N[C@H](CC(C)C)C(=O)N[C@@H](Cc1c[nH]c2ccccc12)C(=O)NCCO

- Gramicidin D not Garamicidin D
- CASRN 1405-07-6



Curation from YESTERDAY

- Gramicidin D is 1393-88-0
- But neither CAS Number has a structure!



Gramicidin

From Wikipedia, the free encyclopedia

Gramicidin is a heterogeneous mixture of three antibiotic compounds, **gramicidins A**, **B** and **C**, making up 80%, 6% and 14%, respectively,[1] all of which are obtained from the soil bacterial species *Bacillus brevis* and called collectively **gramicidin D**. Gramicidin D contains linear pentadecapeptides, that is chains made up of 15 amino acids.[2] This is in contrast to gramicidin S, which is a cyclic peptide chain.

Answering Questions

- I have a 1000 CAS Numbers (or Names) – are there data available?
  - Has any Toxcast data been run?
  - Are there Toxicity Data values available?
  - Are there predicted exposure data (via Expocast)?
  - Can I get predicted physchem data for my model?

Batch Searching for Data for Thousands of Chemicals

EPA United States Environmental Protection Agency

Select Input Type(s)

- [x] Chemical Name
- [ ] CAS-RN
- [ ] InChIKey
- [ ] DSSTox Substance ID
- [ ] Exact Molecular Formula

Enter Identifiers to Search

Sodium l-glutamate
4-Hydroxy-3-nitrophenylarsonic acid
4-Nitro-1,2-phenylenediamine
Methoxypromazine
1-Phenyl-3-methyl-5-pyrazolone
1-Phenyl-2-thiourea
Phenylbutazone
1,4-Benzenediamine
Prednisolone
Probenecid

Display All Chemicals | Download Chemical Data

Select Output Format

Excel

Customize Results

- [ ] Select All

Chemical Identifiers

- [x] Chemical Name
- [x] DTXSID
- [ ] CAS-RN
- [ ] InChIKey
- [ ] IUPAC Name

Structures

- [ ] Mol File
- [ ] SMILES
- [ ] InChI String

Intrinsic Properties

- [ ] Molecular Formula
- [ ] Average Mass
- [ ] Monoisotopic Mass
- [ ] OPERA and TEST Model Predictions

Supporting NTA Applications

- Exposure surveillance
  - What chemicals are in food, products, dust, blood, etc.?

- Chemical prioritization
  - What are relevant chemicals & mixtures?

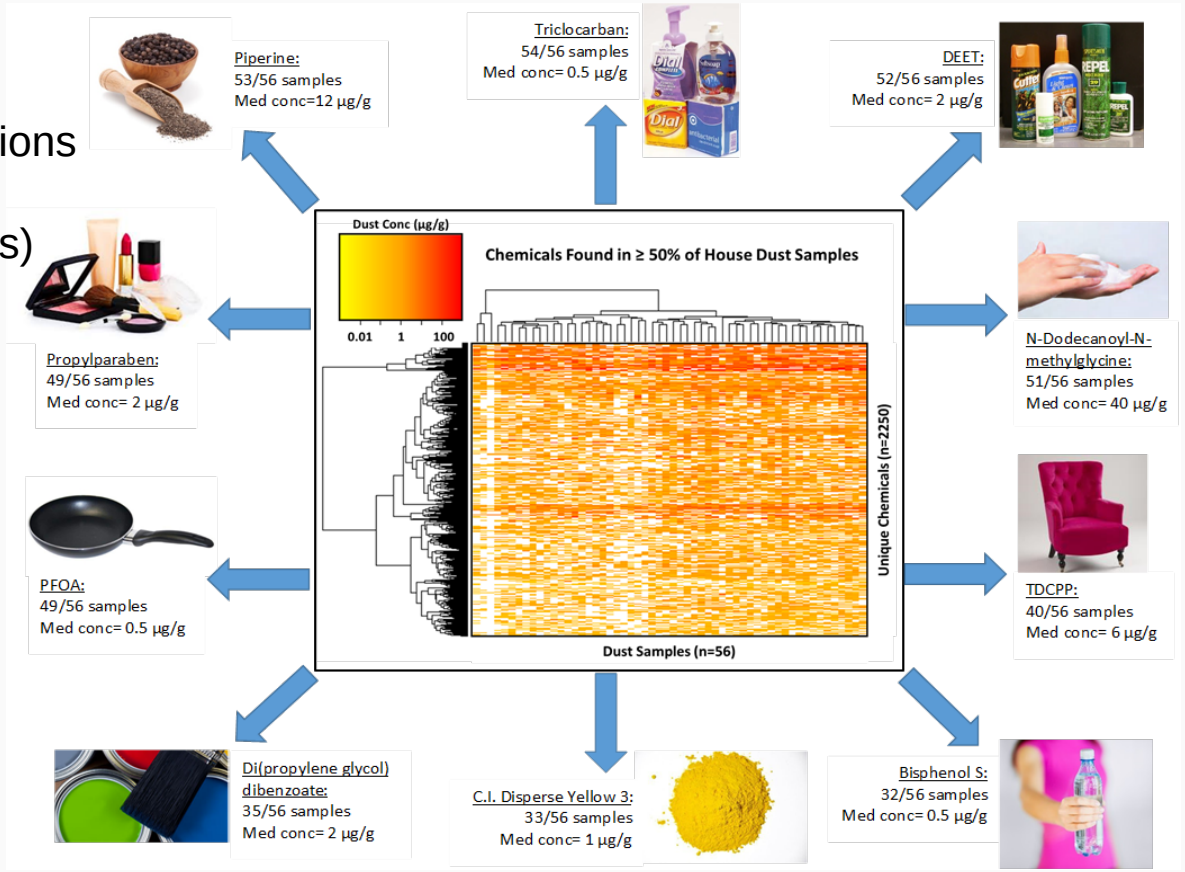
- Exposure forensics
  - What are chemical signatures of exposure sources?

- Effect-directed analysis
  - What are the biologically active chemicals in complex mixtures?

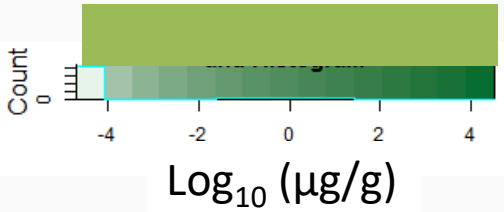
- Biomarker discovery
  - What chemicals are predictive of bioactivity/health impairment?

Suspect Screening Analysis
Chemicals in House Dust

- Chemical confirmation from ToxCast
- Detection frequency
- External calibration for concentrations
- Cluster analysis (homes/chemicals)
- Chemical functional use



Improving Exposure Estimates – Characterizing Commercial Products



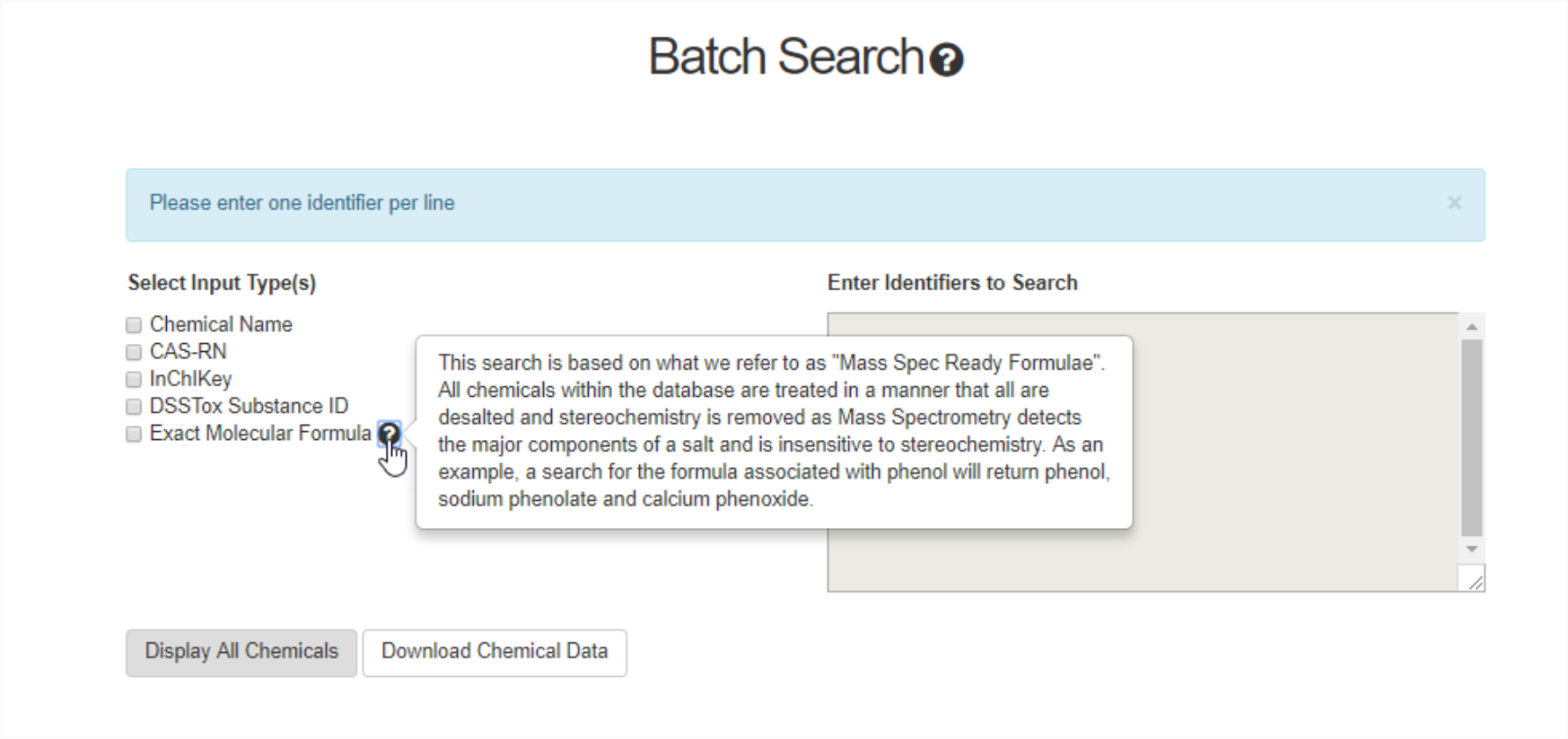
- Air freshener
- Baby soap
- Carpet
- Carpet padding
- Cereals
- Cotton clothing
- Deodorant
- Fabric upholstery
- Glass cleaners
- Hand soap
- Indoor house paint
- Lipstick
- Plastic children's toys
- Shampoo
- Shaving cream
- Shower curtain
- Skin lotion
- Sunscreen
- Toothpaste
- Vinyl upholstery

- **GCXGC-MS with DCM Extraction**
- **1606 tentative and confirmed chemical identifications**

- Common Chemical (n>19)
- ToxCast
- Flame Retardant
- Potent ER

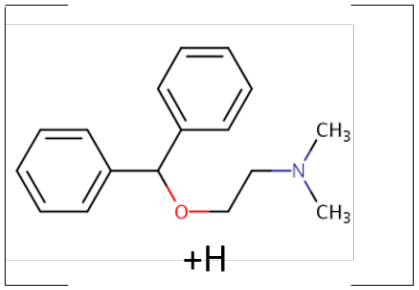
Wambaugh et al. Unpublished

Batch Searching MS-Ready Formulae



MS-Ready database searching

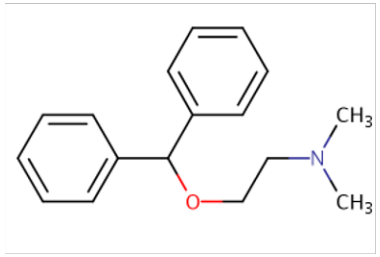
A) Molecular Ion



m/z≈ 256.170

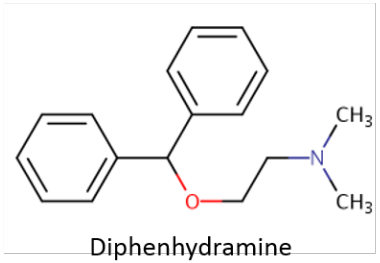
+

B) MS-Ready Form

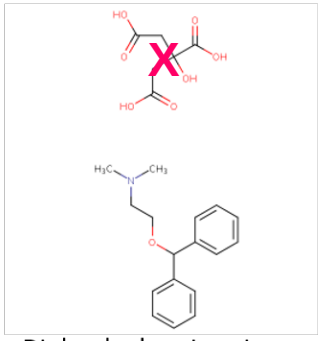


monoisotopic mass= 255.162
$C_{17}H_{21}NO$
DTXCID802949

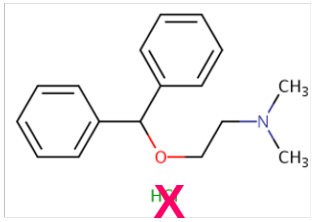
C) Mappings from MS-Ready



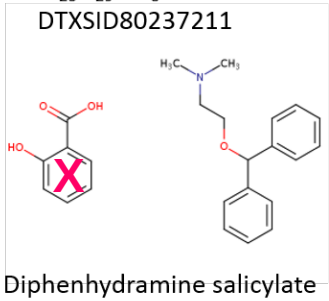
Diphenhydramine
$C_{17}H_{21}NO$
DTXSID4022949



Diphenhydramine citrate
$C_{23}H_{29}NO_8$
DTXSID80237211



Diphenhydramine
hydrochloride
$C_{17}H_{22}ClNO$
DTXSID4020537

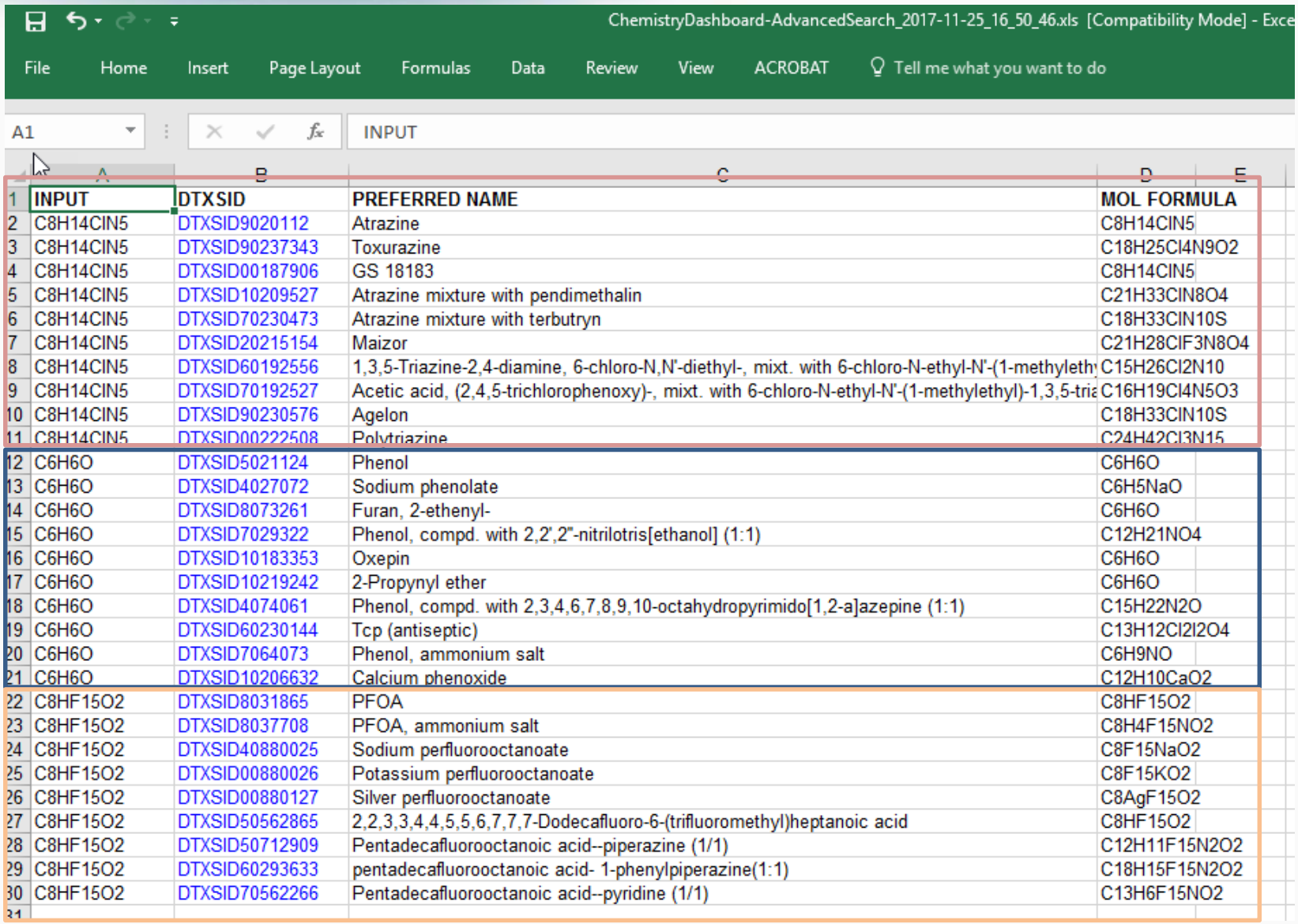


Diphenhydramine salicylate
$C_{24}H_{27}NO_4$
DTXSID10225883

Atrazine C8H14ClN5

Phenol C6H6O

PFOA C8HF15O2



ChemistryDashboard-AdvancedSearch_2017-11-25_16_50_46.xls [Compatibility Mode] - Exce

File Home Insert Page Layout Formulas Data Review View ACROBAT Tell me what you want to do

A1 INPUT

	A	B	C	D	E
1	INPUT	DTXSID	PREFERRED NAME	MOL FORMULA	
2	C8H14ClN5	DTXSID9020112	Atrazine	C8H14ClN5	
3	C8H14ClN5	DTXSID90237343	Toxurazine	C18H25Cl4N9O2	
4	C8H14ClN5	DTXSID00187906	GS 18183	C8H14ClN5	
5	C8H14ClN5	DTXSID10209527	Atrazine mixture with pendimethalin	C21H33ClN8O4	
6	C8H14ClN5	DTXSID70230473	Atrazine mixture with terbutryn	C18H33ClN10S	
7	C8H14ClN5	DTXSID20215154	Maizor	C21H28ClF3N8O4	
8	C8H14ClN5	DTXSID60192556	1,3,5-Triazine-2,4-diamine, 6-chloro-N,N'-diethyl-, mixt. with 6-chloro-N-ethyl-N'-(1-methyleth	C15H26Cl2N10	
9	C8H14ClN5	DTXSID70192527	Acetic acid, (2,4,5-trichlorophenoxy)-, mixt. with 6-chloro-N-ethyl-N'-(1-methylethyl)-1,3,5-tria	C16H19Cl4N5O3	
10	C8H14ClN5	DTXSID90230576	Agelon	C18H33ClN10S	
11	C8H14ClN5	DTXSID00222508	Polytriazine	C24H42Cl3N15	
12	C6H6O	DTXSID5021124	Phenol	C6H6O	
13	C6H6O	DTXSID4027072	Sodium phenolate	C6H5NaO	
14	C6H6O	DTXSID8073261	Furan, 2-ethenyl-	C6H6O	
15	C6H6O	DTXSID7029322	Phenol, compd. with 2,2',2"-nitrilotris[ethanol] (1:1)	C12H21NO4	
16	C6H6O	DTXSID10183353	Oxepin	C6H6O	
17	C6H6O	DTXSID10219242	2-Propynyl ether	C6H6O	
18	C6H6O	DTXSID4074061	Phenol, compd. with 2,3,4,6,7,8,9,10-octahydropyrimido[1,2-a]azepine (1:1)	C15H22N2O	
19	C6H6O	DTXSID60230144	Tcp (antiseptic)	C13H12Cl2I2O4	
20	C6H6O	DTXSID7064073	Phenol, ammonium salt	C6H9NO	
21	C6H6O	DTXSID10206632	Calcium phenoxide	C12H10CaO2	
22	C8HF15O2	DTXSID8031865	PFOA	C8HF15O2	
23	C8HF15O2	DTXSID8037708	PFOA, ammonium salt	C8H4F15NO2	
24	C8HF15O2	DTXSID40880025	Sodium perfluorooctanoate	C8F15NaO2	
25	C8HF15O2	DTXSID00880026	Potassium perfluorooctanoate	C8F15KO2	
26	C8HF15O2	DTXSID00880127	Silver perfluorooctanoate	C8AgF15O2	
27	C8HF15O2	DTXSID50562865	2,2,3,3,4,4,5,5,6,7,7,7-Dodecafluoro-6-(trifluoromethyl)heptanoic acid	C8HF15O2	
28	C8HF15O2	DTXSID50712909	Pentadecafluorooctanoic acid--piperazine (1/1)	C12H11F15N2O2	
29	C8HF15O2	DTXSID60293633	pentadecafluorooctanoic acid- 1-phenylpiperazine(1:1)	C18H15F15N2O2	
30	C8HF15O2	DTXSID70562266	Pentadecafluorooctanoic acid--pyridine (1/1)	C13H6F15NO2	
31					

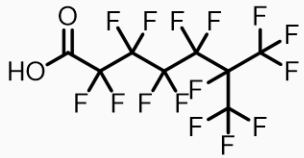
Chemistry Dashboard Output

$C_8HF_{15}O_2$

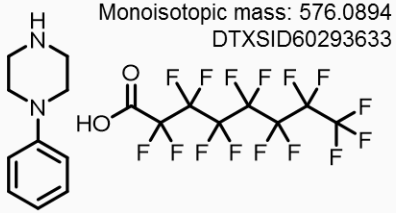
Predicted molecular feature, neutral formula

PFOA, **$C_8HF_{15}O_2$**
Monoisotopic mass: 413.9737
DTXSID8031865

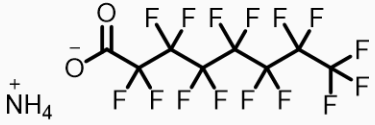
PFOA (isomer), $C_8HF_{15}O_2$
Monoisotopic mass: 413.9737
DTXSID50562865



PFOA-Ph-piperazine, $C_{18}H_{15}F_{15}N_2O_2$
Monoisotopic mass: 576.0894
DTXSID60293633

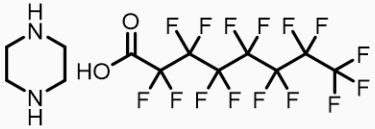


PFOA-NH_4^+, $C_8H_4F_{15}NO_2$
Monoisotopic mass: 431.0003
DTXSID8037708

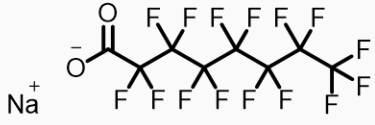


PFOA-Ag^+, $C_8AgF_{15}NO_2$
Monoisotopic mass: 519.8710
DTXSID00880127

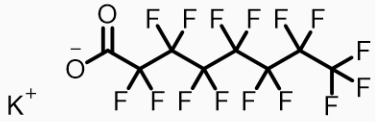
PFOA-piperazine, $C_{12}H_{11}F_{15}N_2O_2$
Monoisotopic mass: 500.0581
DTXSID50712909



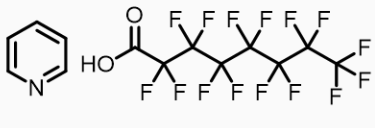
PFOA-Na^+, $C_8F_{15}NaO_2$
Monoisotopic mass: 435.9557
DTXSID40880025



PFOA-K^+, $C_8F_{15}KO_2$
Monoisotopic mass: 451.9296
DTXSID00880026



PFOA-pyridine, $C_{13}H_6F_{15}NO_2$
Monoisotopic mass: 493.0159
DTXSID70562266



Chemistry Dashboard Output

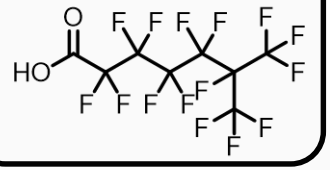
$C_8HF_{15}O_2$

Predicted molecular feature, neutral formula

Two structures to one formula

PFOA, **$C_8HF_{15}O_2$**
Monoisotopic mass: 413.9737
DTXSID8031865

PFOA (isomer), $C_8HF_{15}O_2$
Monoisotopic mass: 413.9737
DTXSID50562865



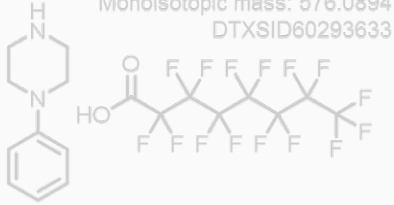
PFOA-NH_4^+, $C_8H_4F_{15}NO_2$
Monoisotopic mass: 431.0003
DTXSID8037708

PFOA-Na^+, $C_8F_{15}NaO_2$
Monoisotopic mass: 435.9557
DTXSID40880025

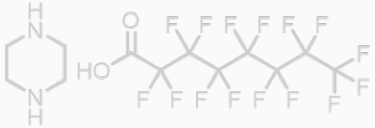
PFOA-Ag^+, $C_8AgF_{15}NO_2$
Monoisotopic mass: 519.8710
DTXSID00880127

PFOA-K^+, $C_8F_{15}KO_2$
Monoisotopic mass: 451.9296
DTXSID00880026

PFOA-Ph-piperazine, $C_{18}H_{15}F_{15}N_2O_2$
Monoisotopic mass: 576.0894
DTXSID60293633



PFOA-piperazine, $C_{12}H_{11}F_{15}N_2O_2$
Monoisotopic mass: 500.0581
DTXSID50712909



PFOA-pyridine, $C_{13}H_6F_{15}NO_2$
Monoisotopic mass: 493.0159
DTXSID70562266

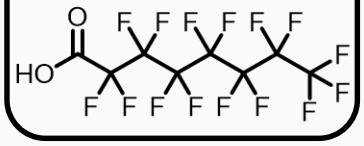
Chemistry Dashboard Output

$C_8HF_{15}O_2$

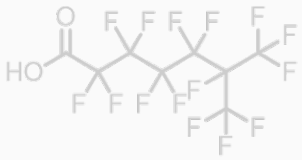
Exclude salts & counter-ions

Adds H to neutralize, maps salt form of structure to neutral formula

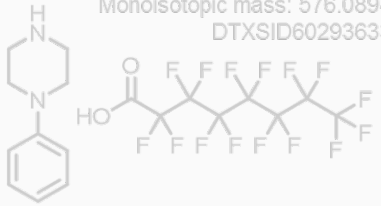
PFOA, **$C_8HF_{15}O_2$**
Monoisotopic mass: 413.9737
DTXSID8031865



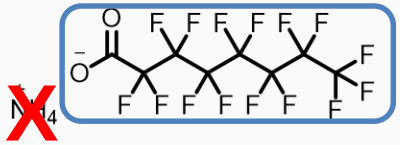
PFOA (isomer), $C_8HF_{15}O_2$
Monoisotopic mass: 413.9737
DTXSID50562865



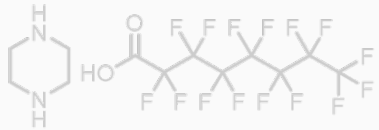
PFOA-Ph-piperazine, $C_{18}H_{15}F_{15}N_2O_2$
Monoisotopic mass: 576.0894
DTXSID60293633



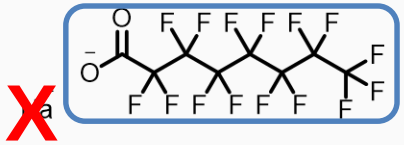
PFOA-NH_4^+, $C_8H_4F_{15}NO_2$
Monoisotopic mass: 431.0003
DTXSID8037708



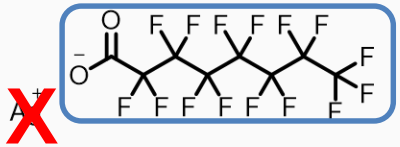
PFOA-piperazine, $C_{12}H_{11}F_{15}N_2O_2$
Monoisotopic mass: 500.0581
DTXSID50712909



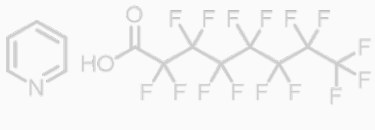
PFOA-Na^+, $C_8F_{15}NaO_2$
Monoisotopic mass: 435.9557
DTXSID40880025



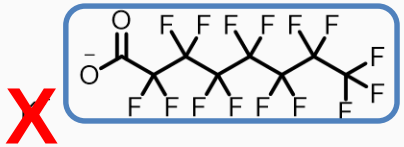
PFOA-Ag^+, $C_8AgF_{15}NO_2$
Monoisotopic mass: 519.8710
DTXSID00880127



PFOA-pyridine, $C_{13}H_6F_{15}NO_2$
Monoisotopic mass: 493.0159
DTXSID70562266



PFOA-K^+, $C_8F_{15}KO_2$
Monoisotopic mass: 451.9296
DTXSID00880026



Chemistry Dashboard Output

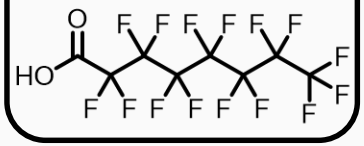
$C_8HF_{15}O_2$

Separates components within mixture

PFOA in mixture mapped to neutral formula,

Heterocycles are separate, searchable components

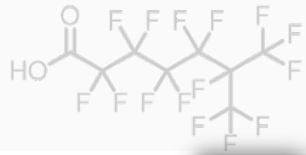
PFOA, **$C_8HF_{15}O_2$**
Monoisotopic mass: 413.9737
DTXSID8031865



PFOA-NH_4^+, $C_8H_4F_{15}NO_2$
Monoisotopic mass: 431.0003
DTXSID8037708

PFOA-Na^+, $C_8F_{15}NaO_2$
Monoisotopic mass: 435.9557
DTXSID40880025

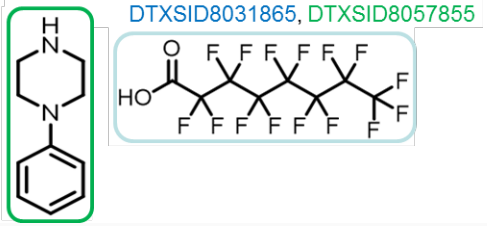
PFOA (isomer), $C_8HF_{15}O_2$
Monoisotopic mass: 413.9737
DTXSID50562865



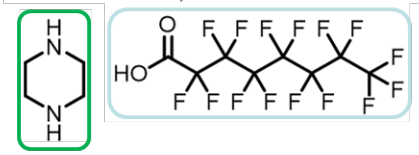
PFOA-Ag^+, $C_8AgF_{15}NO_2$
Monoisotopic mass: 519.8710
DTXSID00880127

PFOA-K^+, $C_8F_{15}KO_2$
Monoisotopic mass: 451.9296
DTXSID00880026

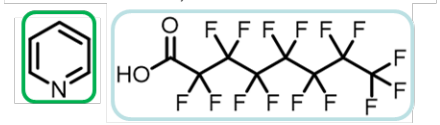
PFOA · Ph-piperazine, **$C_8HF_{15}O_2$** **$C_{10}H_{14}N_2$**
Monoisotopic masses: 413.9737 + 162.1157
DTXSID8031865, DTXSID8057855



PFOA · Piperazine, **$C_8HF_{15}O_2$** **$C_4H_{10}N_2$**
Monoisotopic masses: 413.9737 + 86.0844
DTXSID8031865, DTXSID1021164



PFOA · Pyridine, **$C_8HF_{15}O_2$** **$C_5H_5N_2$**
Monoisotopic masses: 413.9737 + 79.0422
DTXSID8031865, DTXSID9021924



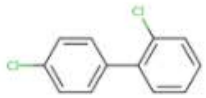
Chemical “Families”

- Sometimes the simplest of questions are difficult to answer!
  - What is the list of CAS Numbers for all PCBs?
  - Can I get an SDF file of all PCBs?
  - Do you have predicted properties for all PCBs?
  - What toxicity data is available for individual PCBS?
  - Have you measured ToxCast data for any PCBs?
  - Can I get all PCBs listed in an Excel Spreadsheet?

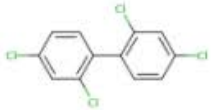
Chemical “Families”

EPA United States Environmental Protection Agency

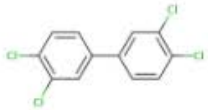
Download as: TSV | Excel | SDF



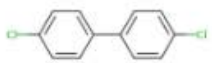
2,4'-Dichlorobiphenyl
34883-43-7



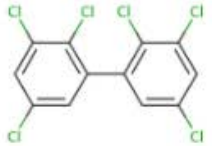
2,2',4,4'-Tetrachlorobiphenyl
2437-79-8



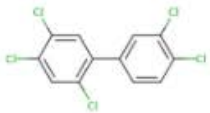
3,3',4,4'-Tetrachlorobinphenyl
32598-13-3



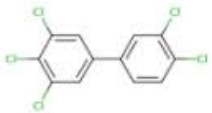
4,4'-Dichlorobiphenyl
2050-68-2



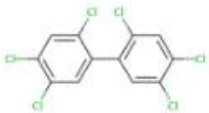
2,2',3,3',5,5'-Hexachlorobiphenyl
35694-04-3



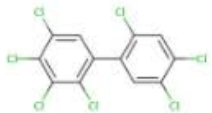
2,3',4,4',5-Pentachlorobiphenyl
31508-00-6



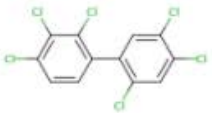
3,3',4,4',5-Pentachlorobiphenyl
57465-28-8



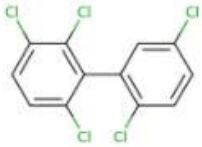
2,2',4,4',5,5'-Hexachlorobiphenyl
35065-27-1



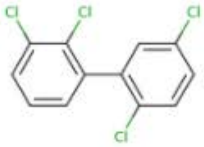
2,2',3,4,4',5,5'-Heptachlorobiphenyl
35065-29-3



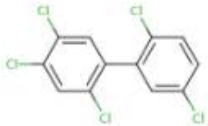
2,2',3,4,4',5'-Hexachlorobiphenyl
35065-28-2



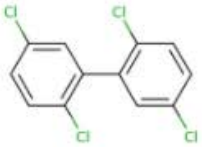
2,2',3,5',6-Pentachlorobiphenyl
38379-99-6



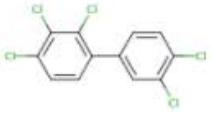
2,2',3,5'-tetrachlorobiphenyl
41464-39-5



2,2,4,5,5'-Pentochlorobiphenyl
37680-73-2



2,2',5,5'-Tetrachlorobiphenyl
35693-99-3



2,3,3',4,4'-Pentachlorobiphenyl
32598-14-4

One click download

	A	B	C	D	E	F	G	H	I
1	DTXSID	CASRN	PREFERRED NAME	IUPAC NAME	MOLECULAR F	MONOISOT	AVERAGE	SMILES	INCHI KEY
2	DTXSID9074779	70362-45-7	PCB 045	2,2',3,6-Tetrachloro-1,1	C12H6Cl4	289.92236	291.98001	ClC1=CC=C(Cl)C(=C1Cl)C1=CC=CC=C1Cl	VHGHHZZT
3	DTXSID9074777	38444-73-4	PCB 019	2,2',6-Trichloro-1,1'-biph	C12H7Cl3	255.96133	257.54001	ClC1=CC=CC(Cl)=C1C1=C(Cl)C=CC=C1	MVXIJRBB(
4	DTXSID9074228	74472-39-2	2,3',4',5',6-Pentachlorol	2,3',4',5',6-Pentachloro-	C12H5Cl5	323.88339	326.42001	ClC1=CC(=CC(Cl)=C1Cl)C1=C(Cl)C=CC=C1Cl	WAZUWHG
5	DTXSID9074226	74472-37-0	2,3,4,4',5-Pentachlorob	2,3,4,4',5-Pentachloro-1	C12H5Cl5	323.88339	326.42001	ClC1=CC=C(C=C1)C1=C(Cl)C(Cl)=C(Cl)C(Cl)=C1	SXZSFWH(
6	DTXSID9074224	74472-35-8	2,3,3',4,6-Pentachlorob	2,3,3',4,6-Pentachloro-1	C12H5Cl5	323.88339	326.42001	ClC1=CC=CC(=C1)C1=C(Cl)C(Cl)=C(Cl)C=C1Cl	XGQBSVV'
7	DTXSID9074222	74472-33-6	2,3,3',6-Tetrachlorobiph	2,3,3',6-Tetrachloro-1,1	C12H6Cl4	289.92236	291.98001	ClC1=CC=CC(=C1)C1=C(Cl)C(Cl)=CC=C1Cl	WZNAMGY
8	DTXSID9074220	74338-23-1	2,3',5',6-Tetrachlorobipl	2,3',5',6-Tetrachloro-1,1	C12H6Cl4	289.92236	291.98001	ClC1=CC(=CC(Cl)=C1)C1=C(Cl)C=CC=C1Cl	HDULUCZR
9	DTXSID9074199	68194-11-6	2,3,4',5,6-Pentachlorob	2,3,4',5,6-Pentachloro-1	C12H5Cl5	323.88339	326.42001	ClC1=CC=C(C=C1)C1=C(Cl)C(Cl)=CC(Cl)=C1Cl	ZDDZPDTV
10	DTXSID9074197	68194-08-1	2,2',3,4',6,6'-Hexachlor	2,2',3,4',6,6'-Hexachlor	C12H4Cl6	357.84442	360.85999	ClC1=CC(Cl)=C(C(Cl)=C1)C1=C(Cl)C(Cl)=CC=C1Cl	RPPNJBZN
11	DTXSID9074195	68194-04-7	2,2',4,6'-Tetrachlorobipl	2,2',4,6'-Tetrachloro-1,1	C12H6Cl4	289.92236	291.98001	ClC1=CC=C(C(Cl)=C1)C1=C(Cl)C=CC=C1Cl	WVHNUGR
12	DTXSID9074193	60233-25-2	2,2',3,4',6'-Pentachlorol	2,2',3,4',6'-Pentachloro-	C12H5Cl5	323.88339	326.42001	ClC1=CC(Cl)=C(C(Cl)=C1)C1=C(Cl)C(Cl)=CC=C1	GOFFZTAP
13	DTXSID9074191	60145-23-5	2,2',3,4,4',5,6'-Heptach	2,2',3,4,4',5,6'-Heptachl	C12H3Cl7	391.80544	395.31	ClC1=CC(Cl)=C(C(Cl)=C1)C1=C(Cl)C(Cl)=C(Cl)C(Cl)=C1	RXRLRYZU
14	DTXSID9074149	41411-61-4	2,2',3,4,5,6-Hexachloro	2,2',3,4,5,6-Hexachloro	C12H4Cl6	357.84442	360.85999	ClC1=C(C=CC=C1)C1=C(Cl)C(Cl)=C(Cl)C(Cl)=C1Cl	RUEIBQJF(
15	DTXSID9074147	40186-70-7	2,2',3,3',4,5',6-Heptach	2,2',3,3',4,5',6-Heptachl	C12H3Cl7	391.80544	395.31	ClC1=CC(Cl)=C(Cl)C(=C1)C1=C(Cl)C(Cl)=C(Cl)C=C1Cl	KJBDZJFS'
16	DTXSID9074145	39635-33-1	3,3',4,5,5'-Pentachlorob	3,3',4,5,5'-Pentachloro-	C12H5Cl5	323.88339	326.42001	ClC1=CC(=CC(Cl)=C1)C1=CC(Cl)=C(Cl)C(Cl)=C1	MXVAYAXI
17	DTXSID9074143	38444-76-7	2,3',6-Trichlorobiphenyl	2,3',6-Trichloro-1,1'-biph	C12H7Cl3	255.96133	257.54001	ClC1=CC=CC(=C1)C1=C(Cl)C=CC=C1Cl	VQOFJPFY
18	DTXSID9074141	38380-05-1	2,2',3,3',4,6'-Hexachlor	2,2',3,3',4,6'-Hexachlor	C12H4Cl6	357.84442	360.85999	ClC1=CC=C(Cl)C(=C1Cl)C1=C(Cl)C(Cl)=C(Cl)C=C1	OKBJVIVE
19	DTXSID9073599	65510-45-4	2,2',3,4,4'-Pentachlorob	2,2',3,4,4'-Pentachloro-	C12H5Cl5	323.88339	326.42001	ClC1=CC(Cl)=C(C=C1)C1=C(Cl)C(Cl)=C(Cl)C=C1	LACXVZHA
20	DTXSID9073541	52744-13-5	2,2',3,3',5,6'-Hexachlor	2,2',3,3',5,6'-Hexachlor	C12H4Cl6	357.84442	360.85999	ClC1=CC(=C(Cl)C(Cl)=C1)C1=C(Cl)C(Cl)=CC=C1Cl	UUTNFLRS
21	DTXSID9073410	16606-02-3	2,4',5-Trichlorobiphenyl	2,4',5-Trichloro-1,1'-biph	C12H7Cl3	255.96133	257.54001	ClC1=CC=C(C=C1)C1=C(Cl)C=CC(Cl)=C1	VAHKBZSA
22	DTXSID80873557	36559-22-5	2,2',3,4'-Tetrachloro-1,1	2,2',3,4'-Tetrachloro-1,1	C12H6Cl4	289.92236	291.98	ClC1=CC(Cl)=C(C=C1)C1=CC=CC(Cl)=C1Cl	ALFHIHDQ
23	DTXSID8074780	61798-70-7	PCB 131	2,2',3,3',4,6-Hexachlorc	C12H4Cl6	357.84442	360.85999	ClC1=CC(Cl)=C(Cl)C(Cl)=C1C1=CC=CC(Cl)=C1Cl	WDLTVNW
24	DTXSID8074239	74472-51-8	2,3,3',4,5,5',6-Heptachl	2,3,3',4,5,5',6-Heptachl	C12H3Cl7	391.80544	395.31	ClC1=CC(=CC(Cl)=C1)C1=C(Cl)C(Cl)=C(Cl)C(Cl)=C1Cl	ZUTDUGMN
25	DTXSID8074237	74472-49-4	2,2',3,4,5,6,6'-Heptachl	2,2',3,4,5,6,6'-Heptachl	C12H3Cl7	391.80544	395.31	ClC1=CC=CC(Cl)=C1C1=C(Cl)C(Cl)=C(Cl)C(Cl)=C1Cl	FGDPOTMI
26	DTXSID8074235	74472-47-2	2,2',3,4,4',5,6-Heptachl	2,2',3,4,4',5,6-Heptachl	C12H3Cl7	391.80544	395.31	ClC1=CC(Cl)=C(C=C1)C1=C(Cl)C(Cl)=C(Cl)C(Cl)=C1Cl	DJEUXBQA
27	DTXSID8074233	74472-44-9	2,3,3',4',5,6-Hexachlorc	2,3,3',4',5,6-Hexachlorc	C12H4Cl6	357.84442	360.85999	ClC1=CC=C(C=C1Cl)C1=C(Cl)C(Cl)=CC(Cl)=C1Cl	ZAGRQXM

How Did We Do This? DSSTox

EPA United States Environmental Protection Agency

ACToR-DSSTox Chemical Registration

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

Substance_ID: DTXSID5024267
CAS: 1336-36-3
Name: Polychlorinated biphenyls
Substance Type: Mixture/Formulation
QC Level: DSSTox_High
Data Source: STN(DSSTox)
QC Notes: biphenyl with multiple (unknown number) chlorines attached at unknown locations

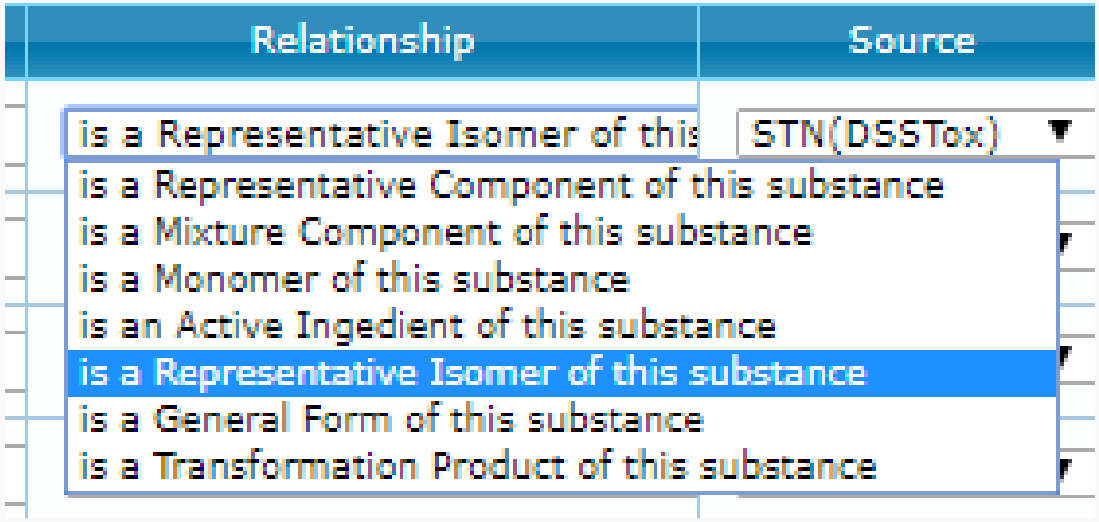
Compound_ID:
Chemical Shown: No Structure
Private Notes:
Source of CAS-Compound:
Double Stereo:
Chiral Stereo:
Chemical Form:

▸ Synonyms (31)
▸ Other Cas (0)
▾ Successor Substances (209)

	CAS-RN	Relationship	Source	Struc	Casrn	Comments
	32774-16-6	is a Representative Isomer of this	STN(DSSTox)	✓		structure shown 3,3',4,4',5,5
	2051-60-7	is a Representative Isomer of this	Public			
	2051-61-8	is a Representative Isomer of this	Public			
	2051-62-9	is a Representative Isomer of this	Public			
	13029-08-8	is a Representative Isomer of this	Public			
	16605-91-7	is a Representative Isomer of this	Public			
	25569-80-6	is a Representative Isomer of this	Public			
	33284-50-3	is a Representative Isomer of this	Public			

Relationship Mappings

- Various relationship mappings can be established. To this point all are manual.



- In progress – **metabolite** mappings (building metabolism competence into high-throughput assays)

UVCB Chemicals

- UVCB chemical examples
  - Surfactants with undefined composition
  - Petroleum Distillates
  - Gelatins, hydrozylates
  - Formaldehyde, reaction products with diethanolamine
  - Fatty acids, linseed-oil, compds. with triethylamine

Managing UVCB Relationships

EPA United States Environmental Protection Agency

Alkylbenzenesulfonate, linear

42615-29-2 | DTXSID3020041

Searched by Synonym: Found 1 result for 'Linear alkylbenzene sulfonate'.

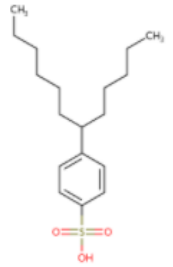
Presence in Lists

Surfactant List Screened in Swiss Wastewater (2014)

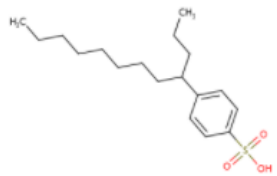
Surfactant List Screened in Swiss Wastewater (2014)

EAWAGSURF is a list of surfactants screened in Swiss wastewater effluents as part of a 2014 study. Structures/mixtures are being progressively curated and linked (Schymanski/Williams). Further details in Schymanski et al 2014, DOI: 10.1021/es4044374

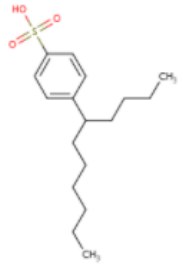
cals



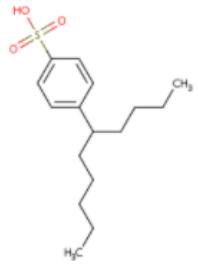
4-(Dodecan-6-yl)benzene-1-sulfon...
23003-92-1



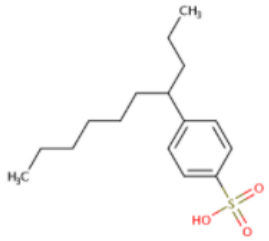
4-(dodecan-4-yl)benzene-1-sulfoni...
NOCAS_862870



C11-LAS
NOCAS_881097

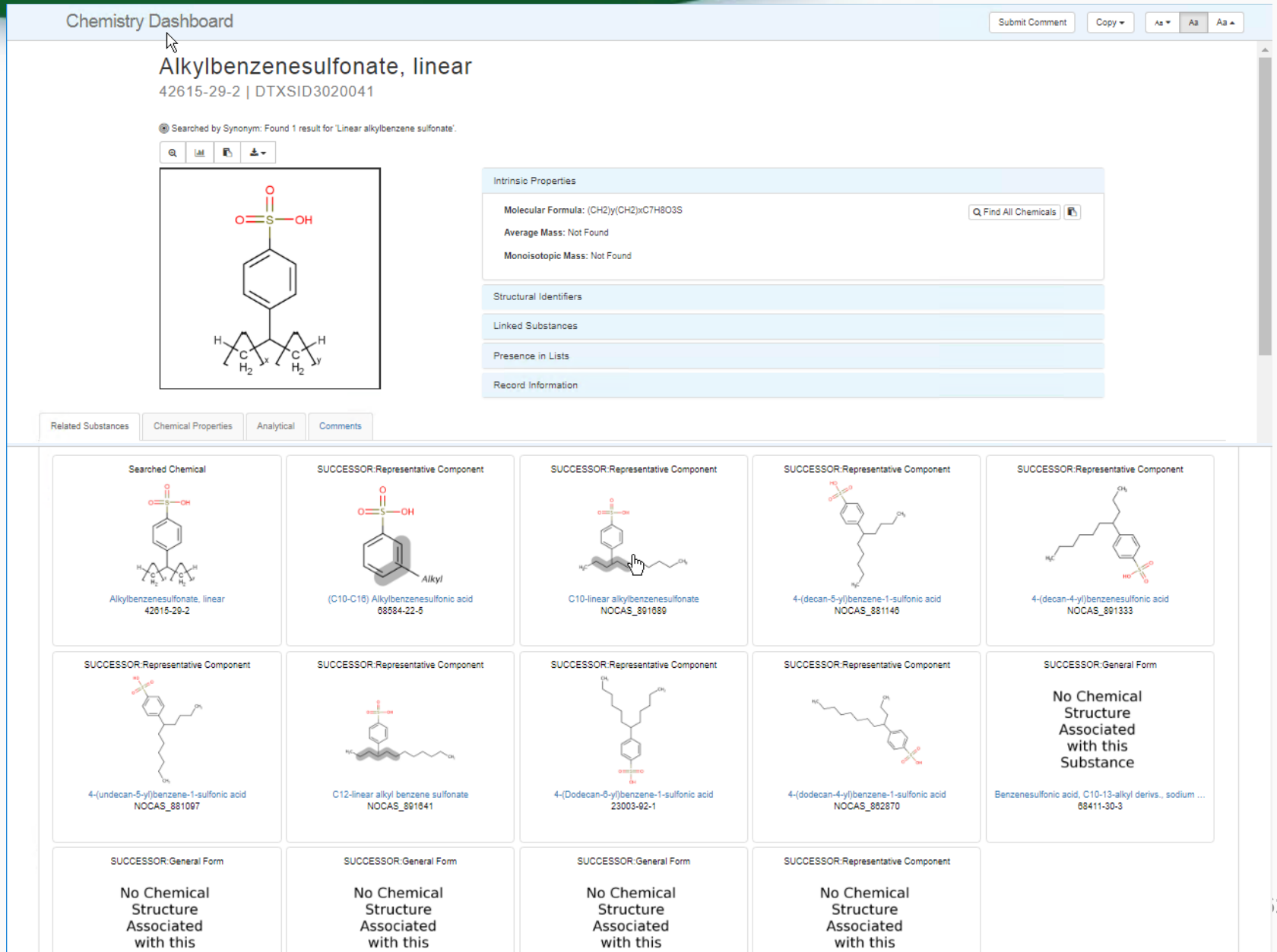


4-(decan-5-yl)benzene-1-sulfonic ...
NOCAS_881146

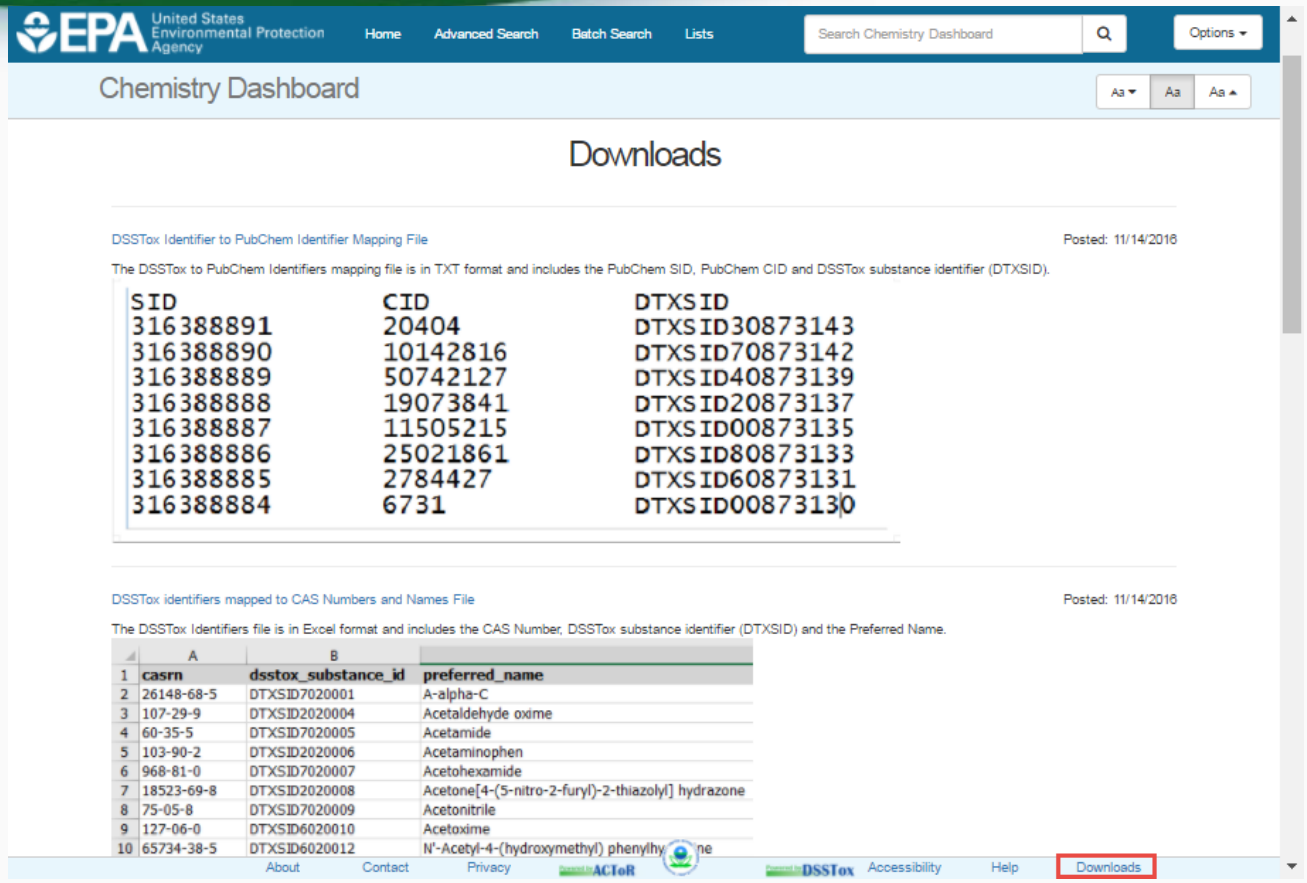


4-(decan-4-yl)benzenesulfonic acid
NOCAS_891333

Ambiguous Chemicals

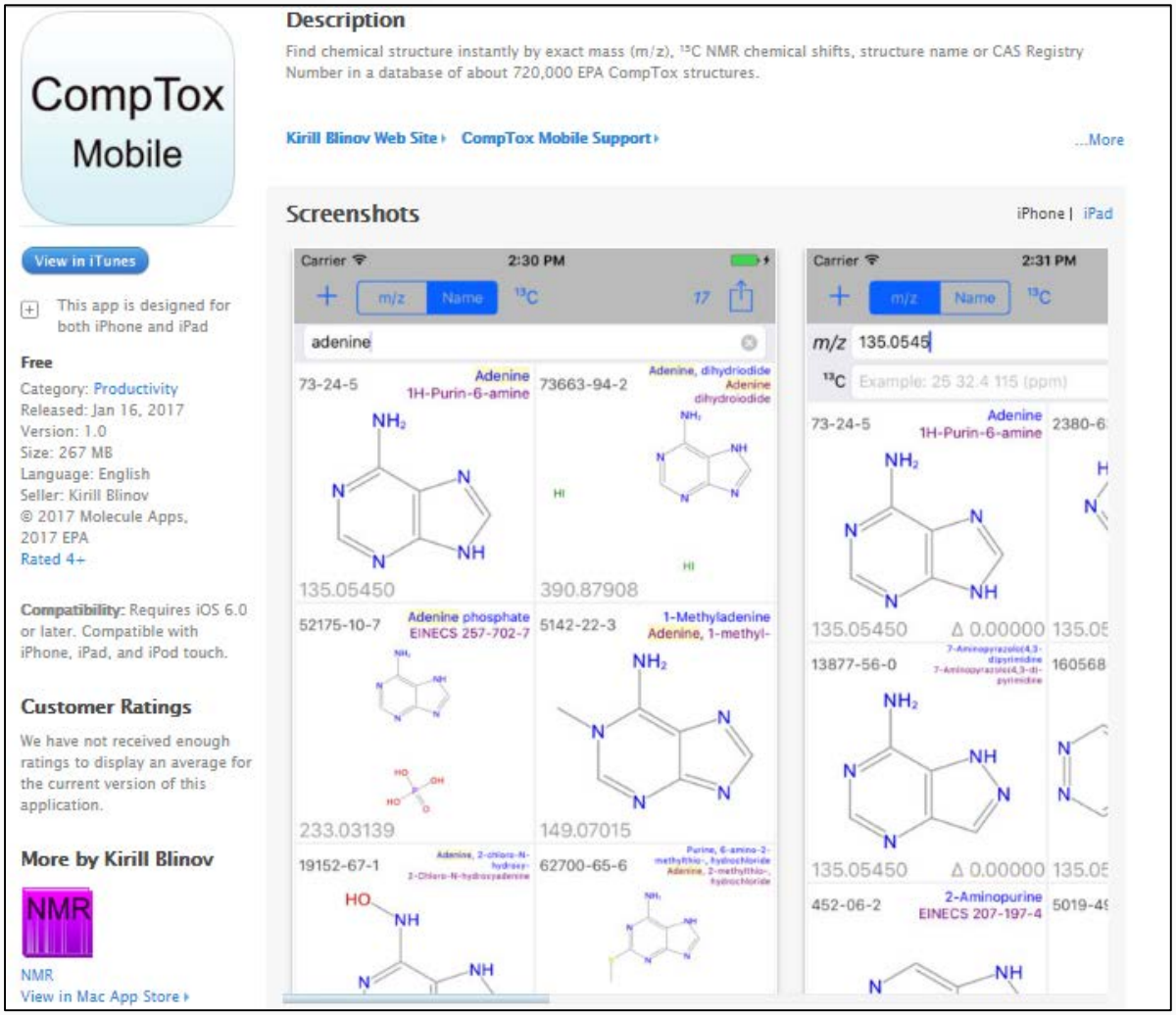


Delivering our Chemistry Data



- Various types of data at FTP download site: ftp://newftp.epa.gov/COMPTOX/Sustainable_Chemistry_Data/Chemistry_Dashboard

Open Data Reuse on iOS

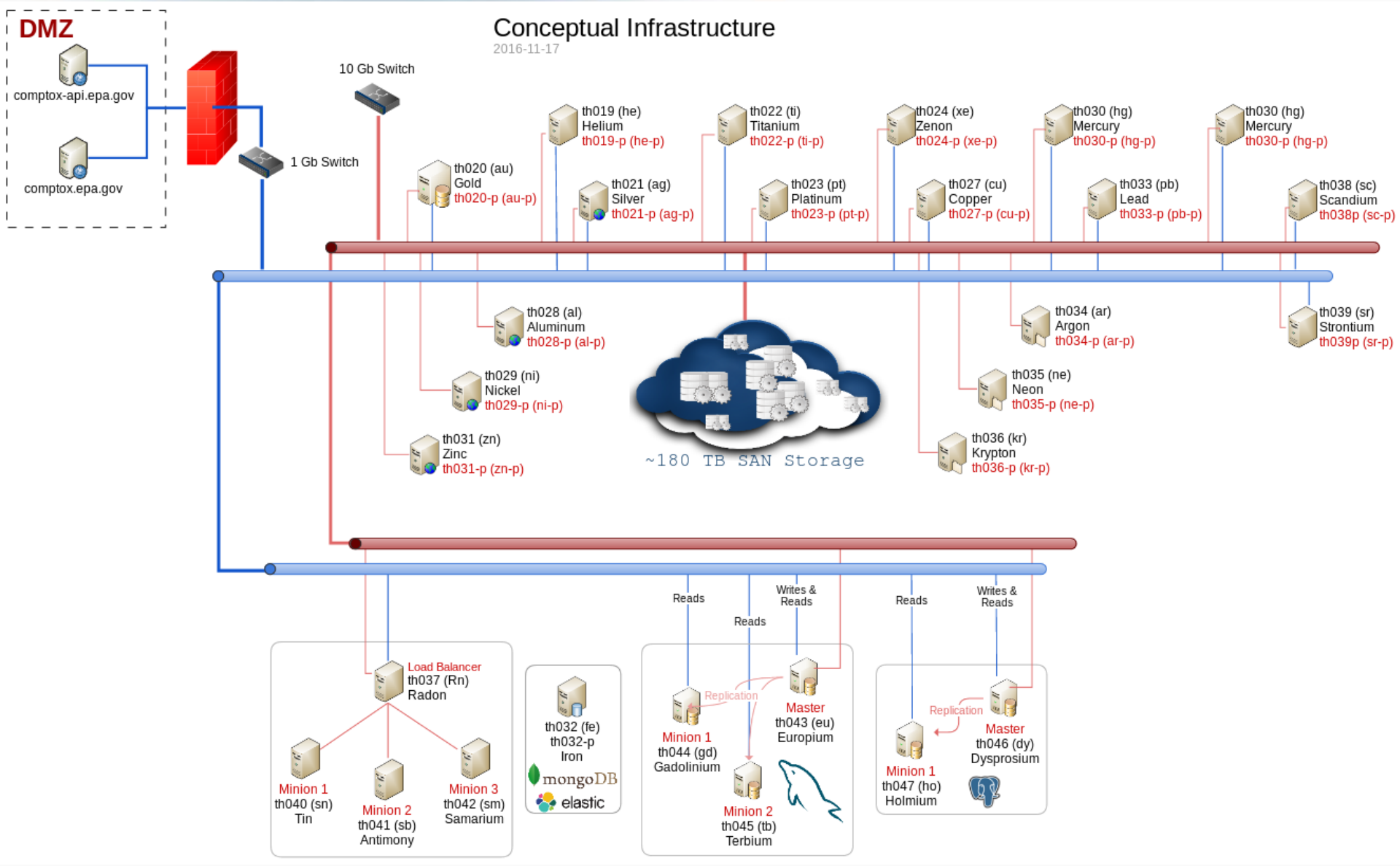


Computational Infrastructure and Processes

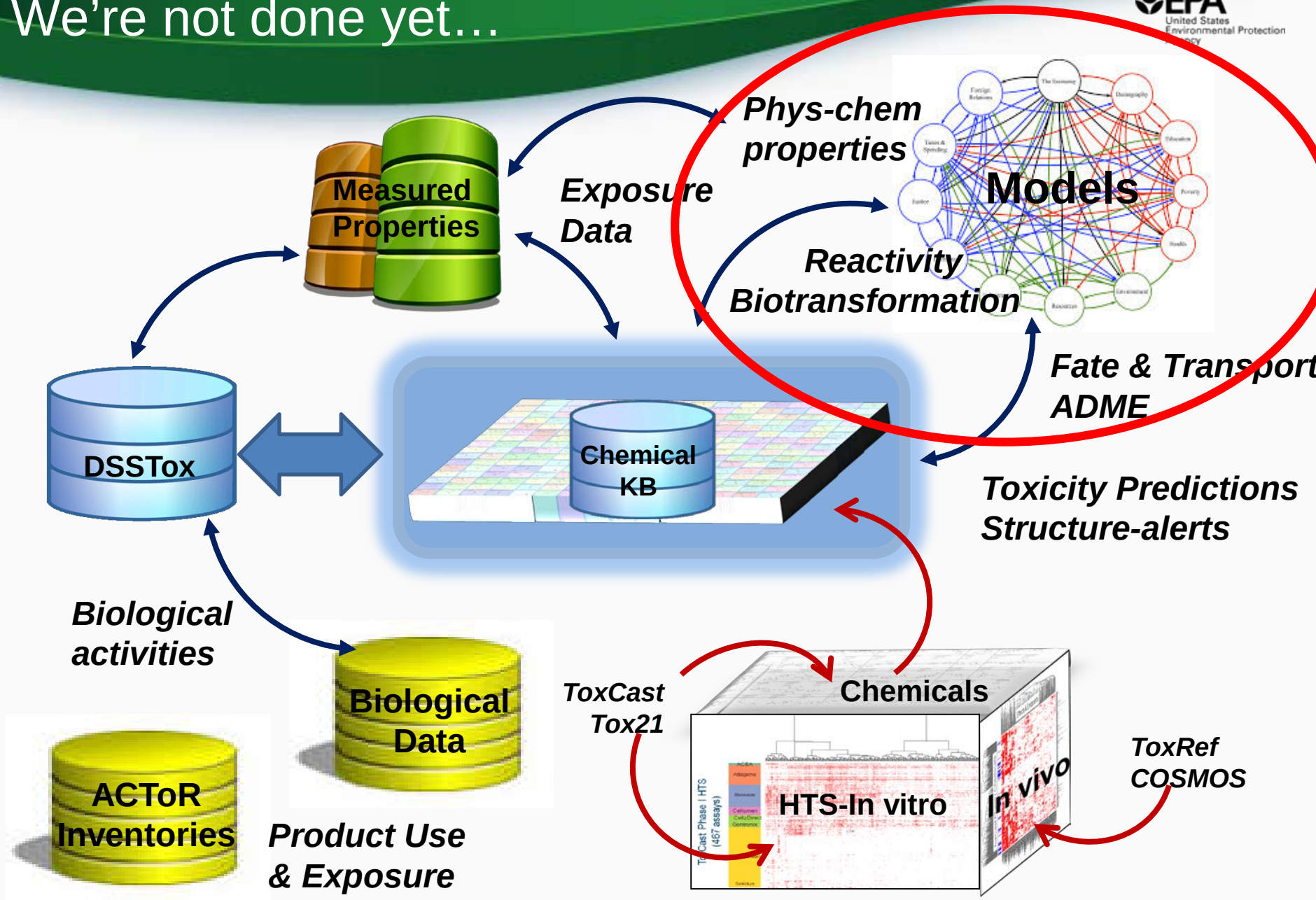
EPA United States Environmental Protection Agency

- 10 years of development – heterogeneous
  - MySQL, Java, R-code, Matlab, multiple Javascript environments, 100s of web services
  - Multiple projects, loose connectivity
- Migration to homogeneity for production Apps
  - MySQL
  - Java
  - JQuery/JavaScript
  - Versioned microservices
- Sprint methodology with Atlassian management tools
  - **Confluence** – knowledge management
  - **Jira** – ticketing/stories (work assignments)
  - **Fisheye** – collaborative peer code review
  - **Bitbucket** – code version control system

Computational Infrastructure



We're not done yet…



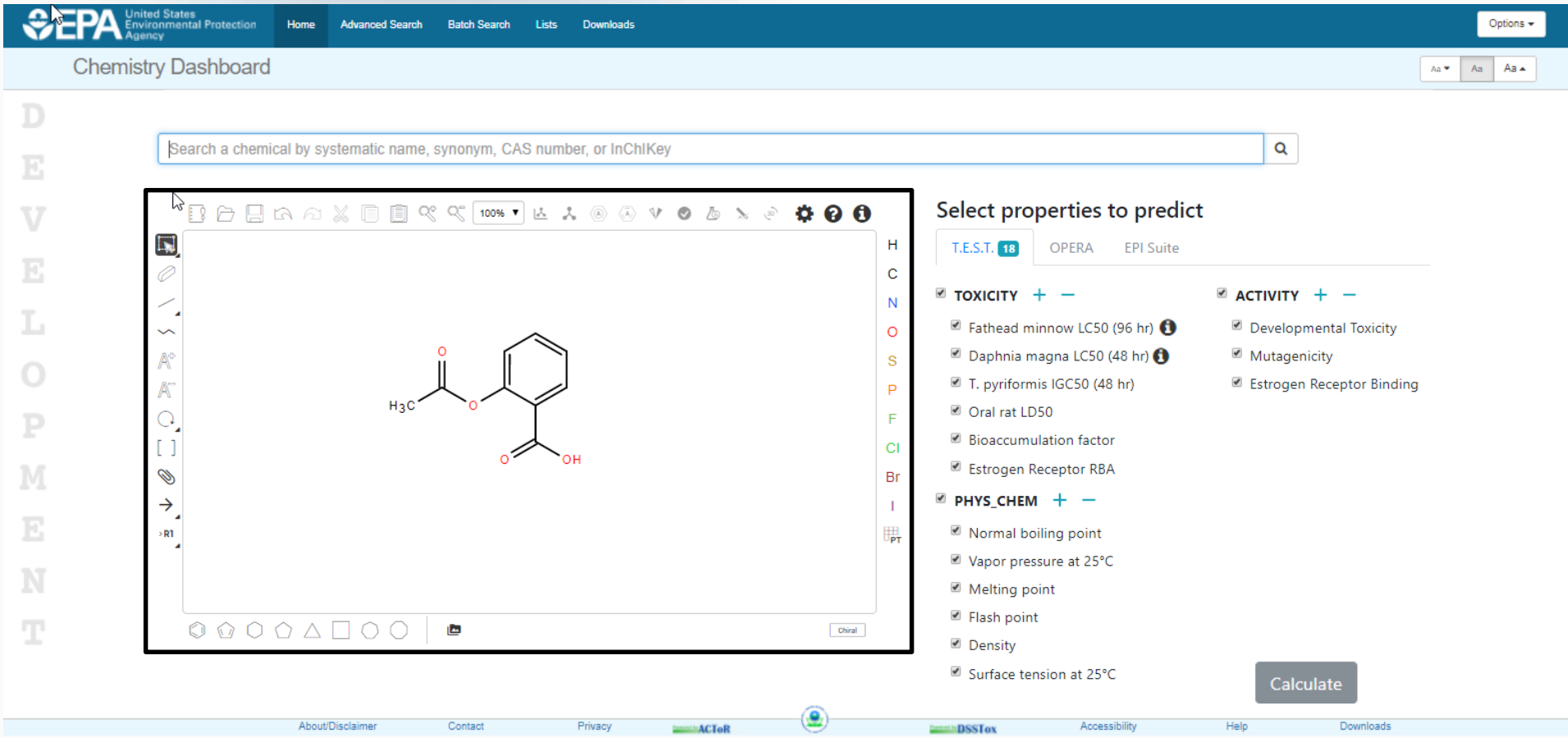
NTA Support Using Fragmentation Predictions

- Work in progress
  - Chemical structures for all chemicals processed into “MS-ready” form
  - MS fragmentation - +ve/-ve ESI spectra low (10V), medium (20V) and high (40V) collision energy levels plus EI data
  - All data generated and presently building spectral search interface

T.E.S.T services (ALPHA)

- 96hr fathead minnow 50% lethal concentration (LC50)
- 48hr daphnia magna 50% lethal concentration (LC50)
- Tetrahymena pyriformis 50% growth inhibition conc. (IGC50)
- Oral rat 50% lethal dose (LD50)
- Bioconcentration Factor (BCF)
- Developmental Toxicity (DevTox)
- Ames Mutagenicity (Mutagenicity)
- Normal boiling point, Flash point, Melting point
- Surface tension, Viscosity, Water Solubility
- Thermal Conductivity, Vapor Pressure, Density

Real time predictions



“RapidTox” prioritization

https://tinyurl.com/y7bkyxt3

Public Report on Approaches

https://www.regulations.gov/document?D=EPA-HQ-OPPT-2017-0586-0003

- "Identifying Potential Candidates for **Prioritization** for Existing Chemical Risk Evaluations Under the Amended Toxic Substances Control Act"
  - Use the data streams under the dashboard – experimental, in vitro, in vivo, predicted physchem, environmental fate, read-across
  - RapidTox: algorithmic-based prioritization of chemical datasets

Future Work

- Continuous updating of lists and ongoing curation
- Integration to other agency databases – **ECOTOX**
- **Structure and substructure searching** in development
- Release of **Analytical QC data** for ToxCast
- Further **Non-Targeted Analysis** support
- **Open** API and web services

- "CompTox ~~Chemistry~~ Dashboard" will integrate other dashboards over time (i.e. EDSP21 and ToxCast)

Acknowledgements

EPA-RTP

An enormous team of contributors from NCCT

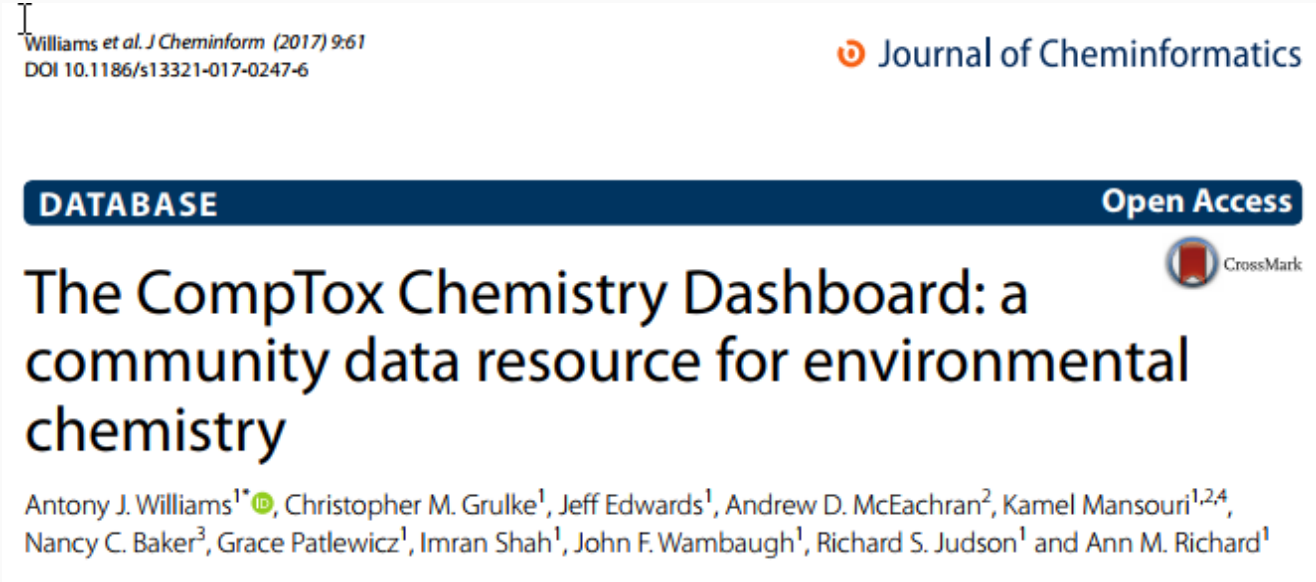
and collaborators from
NERL
NHERL
NRMRL

NORMAN Network

Emma Schymanski

Contact

Antony Williams
NCCT, US EPA Office of Research and Development,
Williams.Antony@epa.gov
ORCID: https://orcid.org/0000-0002-2668-4821



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