



# Cheminformatics Modules

Todd Martin and Antony Williams  
US EPA/ORD/CCTE

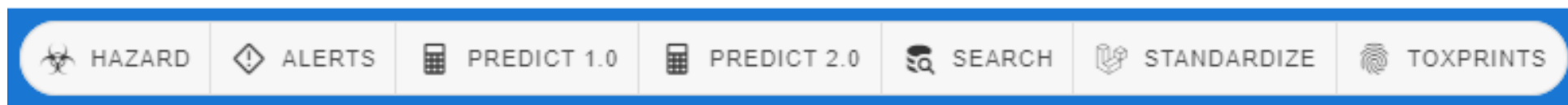
*Views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of USEPA.*

5/1/2024

# Purpose of the Cheminformatics Modules

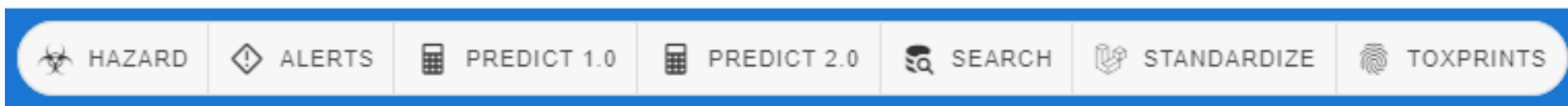
- PoCs are research software builds to **prove approaches** before moving into production software environments
- Assemble data, develop data model(s), **test** user interface **approaches**, work with test user base to **garner feedback**
- PoCs are internal access data refreshes and application updates can be more frequent

# Cheminformatics Modules



- Hazard Comparison module (MATURE: Background and Demo)
- Structure Alerts (DEVELOPMENT)
- PREDICT modules
  - 1.0 – Batch (MATURE)
  - 2.0 – New models (DEVELOPMENT)
- Structure/substructure/similarity searching (MATURE: Demo)
- Standardizer (DEVELOPMENT)
- ToxPrints (DEVELOPMENT)

# Cheminformatics Modules



- Hazard Comparison module (MATURE: Background and Demo)
- Structure Alerts (DEVELOPMENT)
- PREDICT modules
  - 1.0 – Batch (MATURE)
  - 2.0 – New models (DEVELOPMENT)
- Structure/substructure/similarity searching (MATURE: Demo)
- Standardizer (DEVELOPMENT)
- ToxPrints (DEVELOPMENT)

# Hazard Module – Batch Toxicity Data

**Batch Search**

1 Select Input Type(s)  
Substance Identifiers  
 Chemical Name  
 CASRN  
 InChIKey  
 DSSTox Substance ID  
 DSSTox Compound ID  
 InChIKey Skeleton  
 MS-Ready Formula(e)  
 Exact Formula(e)  
 Monoisotopic Mass

2 Enter Identifiers to Search  
(Please enter one identifier per line and limit the number of identifiers to 10,000 or less)

POLYVINYL  
POLYPROPYL GLYCOL  
ISOBUTYLENE  
BUTYL ACRYLATE  
ETHYLHEXYL ACRYLATE  
PARAFFIN WAX  
POLYPROPYLENE  
POLYETHYLENE

3  or

13 Chemicals Found from 14 Input(s)

4

Your file will be exported in Microsoft Excel Format (.xlsx)

Select All columns available

**Chemical Identifiers**

DTXSID  
 Chemical Name  
 DTXCID  
 CAS-RN  
 InChIKey  
 IUPAC Name

**Structures**

Mol File  
 SMILES  
 InChI String  
 MS-Ready SMILES  
 QSAR-Ready SMILES

**Intrinsic and Predicted Properties**

Molecular Formula  
 Average Mass  
 Monoisotopic Mass  
 TEST Model Predictions  
 OPERA Model Predictions  
 ToxPrint fingerprints (ChemoTyper)  
 ToxPrint single fingerprints

**Metadata**

Curation Level Details  
 Safety Data  
 NHANES/Predicted Exposure  
 Data Sources  
 Include ToxVal Data Availability  
 Assay Hit Count  
 Number of PubMed Articles  
 PubChem Data Sources  
 CPDat Product Occurrence Count  
 IRIS  
 PPRTV  
 Wikipedia Article  
 QC Notes  
 Include links to ACToR reports

**Enhanced Data Sheets**

MetFrag Input File (Beta)  
 Abstract Sifter Input File  
 Synonyms and Identifiers  
 Related Substance relationships  
 Associated ToxCast Assays  
 ToxValDB Details  
 Physicochemical Property Values

Showing 13 of 13 chemicals

**Ethylhexyl acrylate**  
DTXSID : DTXSID0001081472  
CASRN : 1322-13-0  
TOXCAS :

**1,2-Propylene glycol**  
DTXSID : DTXSID00021206  
CASRN : 57-35-6  
TOXCAS : 20/909

**Dipropylene glycol**  
DTXSID : DTXSID00027856  
CASRN : 25265-71-8  
TOXCAS : 6/858

**Polypropylene**  
DTXSID : DTXSID000872305  
CASRN : 9003-07-0  
TOXCAS :

**2-Butoxyethanol**  
DTXSID : DTXSID01024097  
CASRN : 111-76-2  
TOXCAS : 7/855

**Paraffin waxes**  
DTXSID : DTXSID061018729  
CASRN : 328069-08-1  
TOXCAS :

**Butyl acrylate**  
DTXSID : DTXSID06024676  
CASRN : 141-32-2  
TOXCAS : 16/448

**Hydrocarbon waxes**  
DTXSID : DTXSID07027673  
CASRN : 8002-74-2  
TOXCAS :

**Diethylene glycol**  
DTXSID : DTXSID08020462  
CASRN : 111-46-6  
TOXCAS : 7/942

**Vinyl chloride**  
DTXSID : DTXSID08021434  
CASRN : 75-01-4  
TOXCAS :

**Polyethylene**  
DTXSID : DTXSID09020462  
CASRN : 9003-07-0  
TOXCAS :

# Hazard Module – Batch Toxicity Data

- >1800 rows of hazard data for a dozen chemicals
- Is there an easier way to view complex hazard data?

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
714	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Carcinoge	LOAEC	LOAEC	-	Point of D	LOAEC	=		127.812	50	ppm	ppm	chronic	chronic	carcinoger
715	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Carcinoge	LOAEC	LOAEC	-	Point of D	LOAEC	=		12781.2	5000	ppm	ppm	chronic	chronic	carcinoger
716	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Carcinoge	LOAEL	LOAEL	-	Point of D	LOAEL	=		1.7	1.7	mg/kg-day	mg/kg bw,	chronic	chronic	carcinoger
717	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Carcinoge	NOAEL	NOAEL	-	Point of D	NOAEL	=		0.13	0.13	mg/kg-day	mg/kg bw,	chronic	chronic	carcinoger
718	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Developm	LOAEC	LOAEC	-	Point of D	LOAEC	=		500	500	ppm	ppm	developm	developm	developm
719	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Developm	NOAEC	NOAEC	-	Point of D	NOAEC	=		10	10	ppm	ppm	developm	developm	developm
720	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Developm	NOAEC	NOAEC	-	Point of D	NOAEC	=		50	50	ppm	ppm	developm	developm	developm
721	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Developm	NOAEC	NOAEC	-	Point of D	NOAEC	=		2500	2500	ppm	ppm	developm	developm	developm
722	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Repeated	LOAEC	LOAEC	-	Point of D	LOAEC	=		127.812	50	ppm	ppm	chronic	chronic	chronic to
723	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Repeated	LOAEC	LOAEC	-	Point of D	LOAEC	=		12781.2	5000	ppm	ppm	chronic	chronic	chronic to
724	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Repeated	LOAEC	LOAEC	-	Point of D	LOAEC	=		51125	20000	ppm	ppm	subchroni	subchroni	sub-chron
725	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Repeated	LOAEC	LOAEC	-	Point of D	LOAEC	=		127812	50000	ppm	ppm	short-term	short-term	short-term
726	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Repeated	LOEC	LOEC	-	Point of D	LOEC	=		260	260	mg/m3	mg/m3 air	chronic	chronic	repeated (-
727	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Repeated	NOAEC	NOAEC	-	Point of D	NOAEC	=		127.812	50	ppm	ppm	subchroni	subchroni	sub-chron
728	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Repeated	NOAEC	NOAEC	-	Point of D	NOAEC	=		255.625	100	ppm	ppm	subchroni	subchroni	sub-chron
729	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Repeated	NOAEC	NOAEC	-	Point of D	NOAEC	=		511.25	200	ppm	ppm	subchroni	subchroni	sub-chron
730	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Repeated	NOEC	NOEC	-	Point of D	NOEC	=		130	130	mg/m3	mg/m3 air	repeat do:	repeat do:	repeated (-
731	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Repeated	LOAEL	LOAEL	-	Point of D	LOAEL	=		1.7	1.7	mg/kg-day	mg/kg bw,	chronic	chronic	chronic to
732	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Repeated	NOAEL	NOAEL	-	Point of D	NOAEL	=		0.13	0.13	mg/kg-day	mg/kg bw,	chronic	chronic	chronic to
733	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECHA IUCL	Repeated	NOAEL	NOAEL	-	Point of D	NOAEL	=		30	30	mg/kg-day	mg/kg bw,	subchroni	subchroni	sub-chron
734	Vinyl chlor	DTXSID80; 75-01-4	Vinyl chlor	ECOTOX	EPA ORD	EC50	EC50	active ingr	Effect Con	EC	=		1170	1.17	mg/m3	mg/L	other	populatio	Populatio

# Hazard Module: Purpose

- Identify safer alternatives for chemicals of concern
  - Comparative Chemical Hazard Assessment
- OECD review\* identified need for “automated tools and methods to reduce hours of highly technical work”
- The **Hazard Module** aims to fill this gap
  - Enable users to readily compare alternatives
  - Display compiled chemical hazard data

\*Organization for Economic Cooperation and Development (OECD) (2013), *Current Landscape of Alternatives Assessment Practice: A Meta-Review*.



# Hazard comparison methodology

- Based on the DfE scoring system

## **Design for the Environment Program Alternatives Assessment Criteria for Hazard Evaluation**

**Version 2.0  
August 2011**

An automated framework for compiling and integrating  
chemical hazard data

[Leora Vegosen](#) & [Todd M. Martin](#) 

*Clean Technologies and Environmental Policy* **22**, 441–458 (2020) | [Cite this article](#)



# Ordinal scoring

- Scores of **Low**, **Medium**, **High**, and **Very High** L, M, H, VH

## Human Health

- Acute mammalian toxicity\*
- Carcinogenicity
- Mutagenicity\*
- Endocrine disruption\*
- Reproductive toxicity
- Developmental toxicity\*
- Neurotoxicity
- Systemic toxicity
- Skin sensitization
- Skin irritation
- Eye irritation

## Ecotoxicity

- Acute aquatic toxicity\*
- Chronic aquatic toxicity

## Fate

- Persistence
- Bioaccumulation\*

**\*Six endpoints can be predicted using Quantitative Structure-Activity Relationship (QSAR) models in WebTEST**

# Sources of Hazard Data

- **GHS H-codes**
  - Safe Work Australia Hazardous Chemical Information System (HCIS)
  - Canada CNESST Workplace Hazardous Materials Information System (WHMIS)
  - European Chemicals Agency (ECHA) Classification Labeling and Packaging (CLP)
  - National Institute of Technology and Evaluation (NITE) of Japan
  - Ministry of Human Resources Malaysia Industry Code of Practice on Chemicals Classification and Hazard Communication
- **Hazard categories**
  - Germany Permanent Senate Commission for the Investigation of Health Hazards of Chemical Compounds in the Work Area (MAK Commission)
  - New Zealand Environmental Protection Authority
- **Quantitative toxicity data**
  - ChemIDplus, ToxVal v9
- **Quantitative Structure-Activity Relationships (QSAR) Predictions**
  - WebTEST, OPERA
  - Ministry of Environment and Food of Denmark Advisory List for Self-Classification of Dangerous Substances

# Sources of Hazard Data

## Hazardous Chemical Lists

- Environment and Climate Change Canada Domestic Substance List (DSL)
- EPA mid-Atlantic Region Human Health Risk-Based Concentrations
- Health Canada Priority Substance Lists (Carcinogenicity and Reproductive Tox)
- International Agency for Research on Cancer (IARC) Monographs
- Integrated Risk Information System (IRIS)
- National Institute for Occupational Safety and Health (NIOSH) list of potential occupational carcinogens
- California Proposition 65
- ECHA Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) Candidate List of Substances of Very High Concern for Authorization
- Report On Carcinogens
- Chemsec Substitute It Now (SIN) List
- The Endocrine Disruption Exchange (TEDX) List of Potential Endocrine Disruptors
- Toxic Substances Control Act (TSCA) Work Plan
- University of Maryland (UMD) List of Acute Toxins, Teratogens, Carcinogens, or Mutagens

# Criteria for converting acute mammalian toxicity data into hazard scores

Source	Endpoint	Score				
		VH	H	M	L	N/A
DfE criteria	Oral LD50 (mg/kg)	≤ 50	> 50 - 300	> 300 - 2000	> 2000	
	Hazard Code	H300	H301	H302		
ChemIDplus; T.E.S.T. Predicted*	Oral LD50* (mg/kg)	≤ 50	> 50 - 300	> 300 - 2000	> 2000	
Australia; Canada; ECHA CLP; Japan**; Malaysia	Hazard Code	H300	H301	H302	H303	
Denmark	Category	AcuteTox1 and AcuteTox2	AcuteTox3	AcuteTox4		
New Zealand	Category	Category 6.1A	Category 6.1C	Category 6.1D	Category 6.1E	
		Category 6.1B				
TSCA Work Plan			Acute mammalian toxicity			
UMD		Acute toxin				

# Assigning the Overall Score

- Trumping Method: overall score is the most toxic score from the most authoritative source:
  - 1. Authoritative (e.g., ECHA CLP)**
  2. Screening (e.g., ChemIDplus)
  3. *Predicted* (e.g., WebTEST)



Full

Chemicals: 20 Toxicity: VH - Very High H - High M - Medium L - Low I - Inconclusive N/A - Not Applicable Authority: **Authoritative** Screening QSAR Model

CAS Name	Human Health Effects															Ecotoxicity		Fate		
	Acute Mammalian Toxicity			Carcinogenicity	Genotoxicity Mutagenicity	Endocrine Disruption	Reproductive	Developmental	Neurotoxicity		Systemic Toxicity		Skin Sensitization	Skin Irritation	Eye Irritation	Acute Aquatic Toxicity	Chronic Aquatic Toxicity	Persistence	Bioaccumulation	Exposure
	Oral	Inhalation	Dermal						Repeat Exposure	Single Exposure	Repeat Exposure	Single Exposure								
79-06-1 <small>AIGBT</small> Acrylamide	H	M	M	VH	VH	L	M	H	H	H	H	H	H	H	H	M	M	L	L	H
79-01-6 <small>AIGBT</small> Trichloroethylene	L	M	L	VH	VH	I	H	H	H	H	H	M	H	H	H	H	VH	H	L	H
108-95-2 <small>AIGBT</small> Phenol	H	H	H	H	H	H	H	H	H	H	M	H	H	VH	VH	H	H	L	H	VH
50-00-0 <small>AIGBT</small> Formaldehyde	H	H	H	VH	H	H		L			L	M	H	VH	VH	H	L	L	L	H
111-30-8 <small>AGBTM</small> Glutaraldehyde	H	VH	H	M	VH	H	H	L		H	H	M	H	VH	VH	VH	H	L	L	H
302-01-2 <small>IGBTP</small> Hydrazine	H	H	H	VH	VH		H	M	H	H	H	H	H	VH	VH	VH	VH	L	L	
75-21-8 <small>AIGBT</small> Ethylene oxide	VH	H	I	VH	VH	H	H	H	H	H	H	M	H	H	H	M	L	H	L	VH
7803-57-8 <small>GBT</small> Hydrazine hydrat...	VH	VH	VH	VH	VH		M	I	H	H	H	H	H	VH	VH	VH	VH			
101-77-9 <small>AGBT</small>																				

- Skipped (0)
- Unlikely (0)
- Filters (0)
- Sorting (0)
- Structure

**Acute Mammalian Toxicity Oral** for **Acrylamide** ( 79-06-1 🔍 )

Toxicity: VH - Very High H - High M - Medium L - Low I - Inconclusive N/A - Not Applicable Authority: **Authoritative** ⓘ Screening ⓘ QSAR Model ⓘ

Source	Score	Route	Test Organism	Test Type	Category	Hazard Code	Hazard Statement	Rationale	Note
ECHA CLP	<span style="background-color: #fff3cd;">H</span>	oral			Acute Tox. 3	H301	Toxic if swallowed	Score of H was assigned based on a hazard code of H301	
Australia	<span style="background-color: #fff3cd;">H</span>	oral			Acute toxicity - category 3	H301	Toxic if swallowed	Score of H was assigned based on a hazard code of H301	N (The classification information for this entry was provided by the National Industrial Chemical Notification and Assessment Scheme)
Canada	<span style="background-color: #fff3cd;">H</span>	oral			Acute toxicity - oral - Category 3	H301	Toxic if swallowed	Score of H was assigned based on a hazard code of H301	Comments: This product could belong to the hazard class "Combustible dust", based on various factors related to the combustibility and explosiveness of its dust, including composition, shape and size of the particles.
ChemIDplus	<span style="background-color: #fff3cd;">H</span>	oral						50 mg/kg < Oral LD50 <=300	

# Strengths and Limitations

- **Strengths**
  - Provides rapid way to compare chemicals and retrieve hazard data
  - Includes data from several sources including QSAR models
- **Limitations**
  - Data gaps
  - Automation limits the scope of data searching and quality assurance, particularly of primary sources

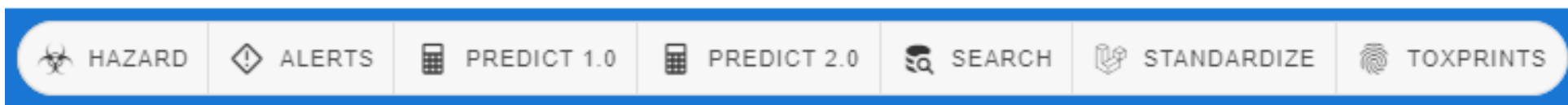


# Future Research Needs

- Update the data from each source
- Add additional data sources
- New/Improved QSAR models

# Live DEMO

# Cheminformatics Modules

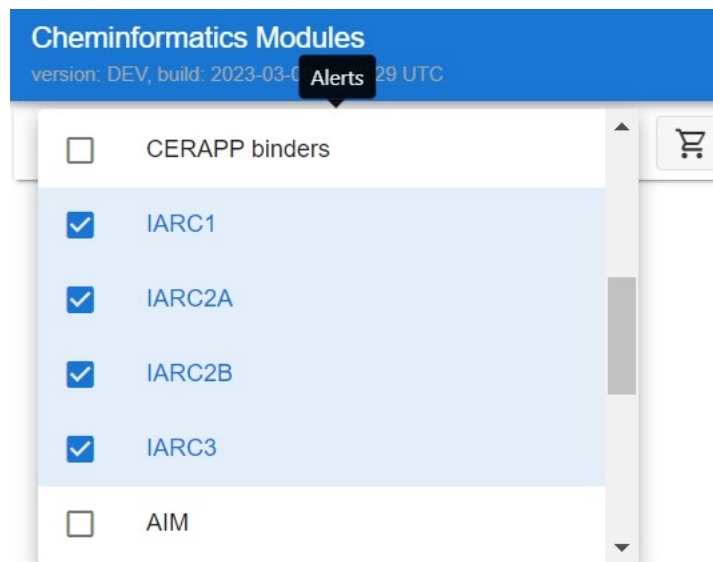


- Hazard Comparison module (MATURE: Background and Demo)
- **Structure Alerts (DEVELOPMENT)**
- PREDICT modules
  - 1.0 – Batch (MATURE)
  - 2.0 – New models (DEVELOPMENT)
- Structure/substructure/similarity searching (MATURE: Demo)
- Standardizer (DEVELOPMENT)
- ToxPrints (DEVELOPMENT)

# Structure Alerts (Development)

**Cheminformatics Modules**  
version: DEV, build: 2023-03-0 Alerts 29 UTC

- CERAPP binders
- IARC1
- IARC2A
- IARC2B
- IARC3
- AIM



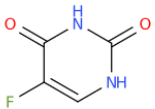




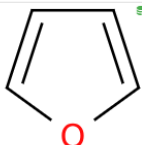


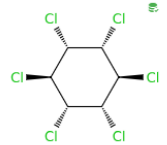

# Structure Alerts (Development)

Cheminformatics Modules  
version: DEV, build: 2023-03-09 06:08:29 UTC

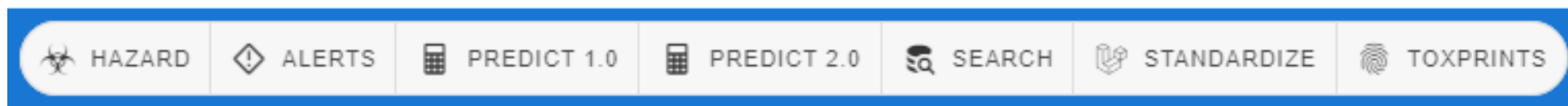
HAZARD ALERTS PREDICT 1.0 PREDICT 2.0 SEARCH STANDARDIZE TOXPRINTS

Search in any field

Collapse Details

#	ID	Chemical	aim	iarc1	iarc2a	iarc2b	iarc3
36	51-21-8						
37	50-00-0	$\text{H}_2\text{C}=\text{O}$					
38	110-00-9						
39	58-89-9						

# Cheminformatics Modules



- Hazard Comparison module (MATURE: Background and Demo)
- Structure Alerts (DEVELOPMENT)
- **PREDICT modules**
  - 1.0 – Batch (MATURE)
  - 2.0 – New models (DEVELOPMENT)
- Structure/substructure/similarity searching (MATURE: Demo)
- Standardizer (DEVELOPMENT)
- ToxPrints (DEVELOPMENT)

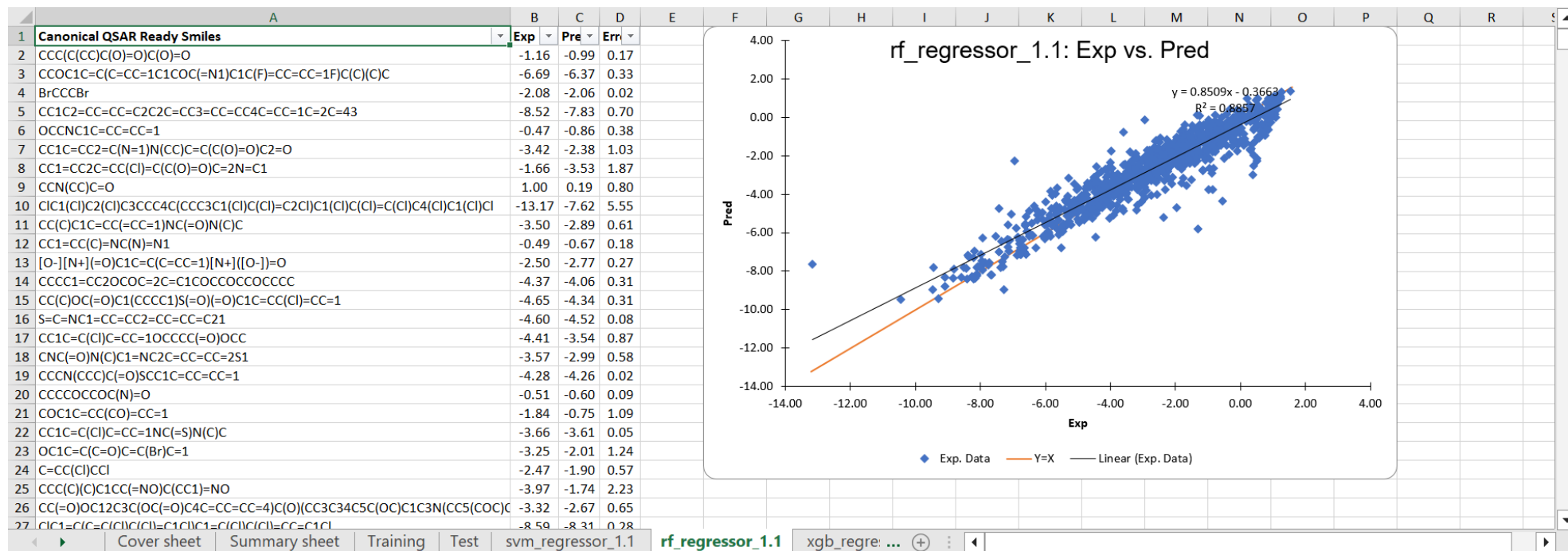
# PREDICT 1.0 Batch Prediction

Legend: A = Active, N = Not active, I = Inconclusive, Experimental, *Predicted*

CAS Name	Filters	Structure	Products	Fathead minnow LC50 (96 hr)	Daphnia magna LC50 (48 hr)	T. pyriformis IGC50 (48 hr)	Oral rat LD50	Bioconcentration factor	Developmental Toxicity	Mutagenicity	Estrogen Receptor Binding	Estrogen Receptor RBA	Normal boiling point	Vapor pressure at 25°C	Melting point	Density	Flash point	Surface tension at 25°C	Thermal conductivity at 25°C	Viscosity at 25°C	Water solubility at 25°C
301-12-2 Oxydemeton-met...	GBTM			<u>3.35</u>	<u>5.81</u>	<i>I</i>	<u>3.91</u> <u>4.27</u>	<u>0.28</u>	<i>N</i>	<u>A</u> <i>N</i>	<i>N</i>			<u>-5.19</u>	<u>50</u>	<u>1.31</u> <u>1.28</u>	<u>182</u> <u>186</u>			<i>I</i>	<u>1.10</u>
115-29-7 Endosulfan	AIGBT			<u>6.10</u> <u>5.14</u>	<u>4.35</u> <u>3.35</u>	<u>8.48</u> <u>6.33</u>	<u>6.08</u> <u>5.36</u>	<i>I</i>	<i>N</i>		<i>N</i> <i>N</i>			<u>319</u>	<u>-6.76</u> <u>-6.22</u>	<u>106</u> <u>90</u>	<u>1.94</u> <u>1.81</u>	<u>226</u> <u>227</u>			
1646-75-9 Aldicarb oxime	GBT			<u>3.10</u>	<i>I</i>	<i>I</i>	<u>2.25</u> <u>2.39</u>	<i>I</i>	<i>N</i>	<u>N</u> <i>N</i>	<i>N</i>	<i>I</i>	<u>210</u> <u>188</u>	<i>I</i>	<u>21</u> <u>28</u>	<u>1.00</u> <u>0.99</u>	<u>81</u> <u>70</u>	<u>31.43</u>			
630-08-0 Carbon monoxide	AGBT			<i>I</i>	<i>I</i>	<i>I</i>	<i>I</i>	<i>I</i>	<i>I</i>	<i>I</i>	<i>I</i>	<i>I</i>	<u>-192</u> <i>I</i>	<u>8.19</u> <i>I</i>	<u>-205</u>		<i>I</i>	<i>I</i>	<i>I</i>	<i>I</i>	<i>I</i>
2312-35-8 Propargite	IGBTM			<u>6.21</u> <u>4.57</u>	<u>2.37</u> <u>1.93</u>	<u>5.82</u>		<i>I</i>	<i>A</i>		<i>N</i>		<i>I</i>	<u>325</u>	<u>-6.52</u> <u>-6.75</u>	<u>23</u>	<u>1.17</u> <u>1.16</u>	<u>226</u> <u>226</u>		<i>I</i>	<i>I</i>
115-90-2 Fensulfothion	GBTM			<u>2.19</u> <u>2.80</u>	<u>5.15</u> <u>4.52</u>	<u>3.85</u> <u>4.55</u>	<u>6.94</u>	<i>N</i> <i>N</i>	<i>A</i>	<u>1.25</u>	<i>N</i>		<i>I</i>	<i>I</i>	<u>-4.30</u> <i>I</i>	<u>59</u>	<u>1.31</u> <u>1.37</u>	<u>198</u> <u>201</u>			<i>I</i>

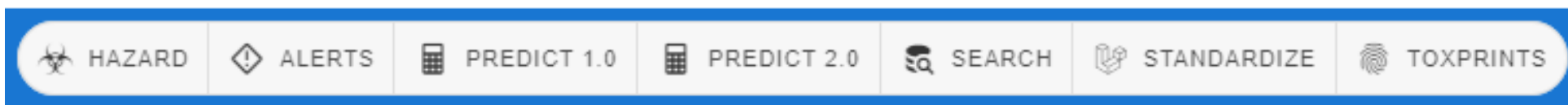
# PREDICT 2.0 – New model delivery

- Early development work for how we will deliver models in the future. The biggest benefit is data transparency





# Cheminformatics Modules

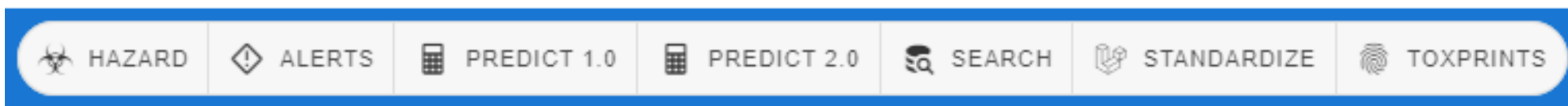


- Hazard Comparison module (MATURE: Background and Demo)
- Structure Alerts (DEVELOPMENT)
- PREDICT modules
  - 1.0 – Batch (MATURE)
  - 2.0 – New models (DEVELOPMENT)
- **Structure/substructure/similarity searching (MATURE: Demo)**
- Standardizer (DEVELOPMENT)
- ToxPrints (DEVELOPMENT)

Structure/Substructure/Similarity

**LIVE DEMO**

# Cheminformatics Modules



- Hazard Comparison module (MATURE: Background and Demo)
- Structure Alerts (DEVELOPMENT)
- PREDICT modules
  - 1.0 – Batch (MATURE)
  - 2.0 – New models (DEVELOPMENT)
- Structure/substructure/similarity searching (MATURE: Demo)
- **Standardizer (DEVELOPMENT)**
- ToxPrints (DEVELOPMENT)

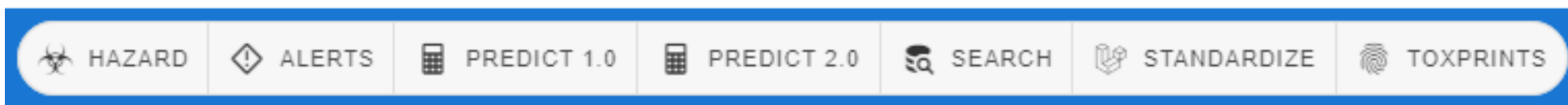
# Generation of QSAR-Ready/MS-Ready

- Preparing chemicals for modeling and Linking in CCD

The screenshot displays the Cheminformatics Modules interface. The top navigation bar includes 'HAZARD', 'ALERTS', 'PREDICT 1.0', 'PREDICT 2.0', 'SEARCH', 'STANDARDIZE', and 'TOXPRINTS'. A search bar is located below the navigation bar. The main content area shows a list of chemical structures, each with a corresponding processed version and a green checkmark indicating successful processing.

ID	DTXSID	Original Structure	Processed Structure	Status
34	DTXSID7020558			✓
35	DTXSID1020560			✓
36	DTXSID6020561			✓
43	DTXSID2020686			✓

# Cheminformatics Modules



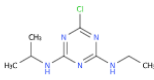
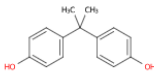
- Hazard Comparison module (MATURE: Background and Demo)
- Structure Alerts (DEVELOPMENT)
- PREDICT modules
  - 1.0 – Batch (MATURE)
  - 2.0 – New models (DEVELOPMENT)
- Structure/substructure/similarity searching (MATURE: Demo)
- Standardizer (DEVELOPMENT)
- **ToxPrints (DEVELOPMENT)**

# ToxPrint chemotype generation

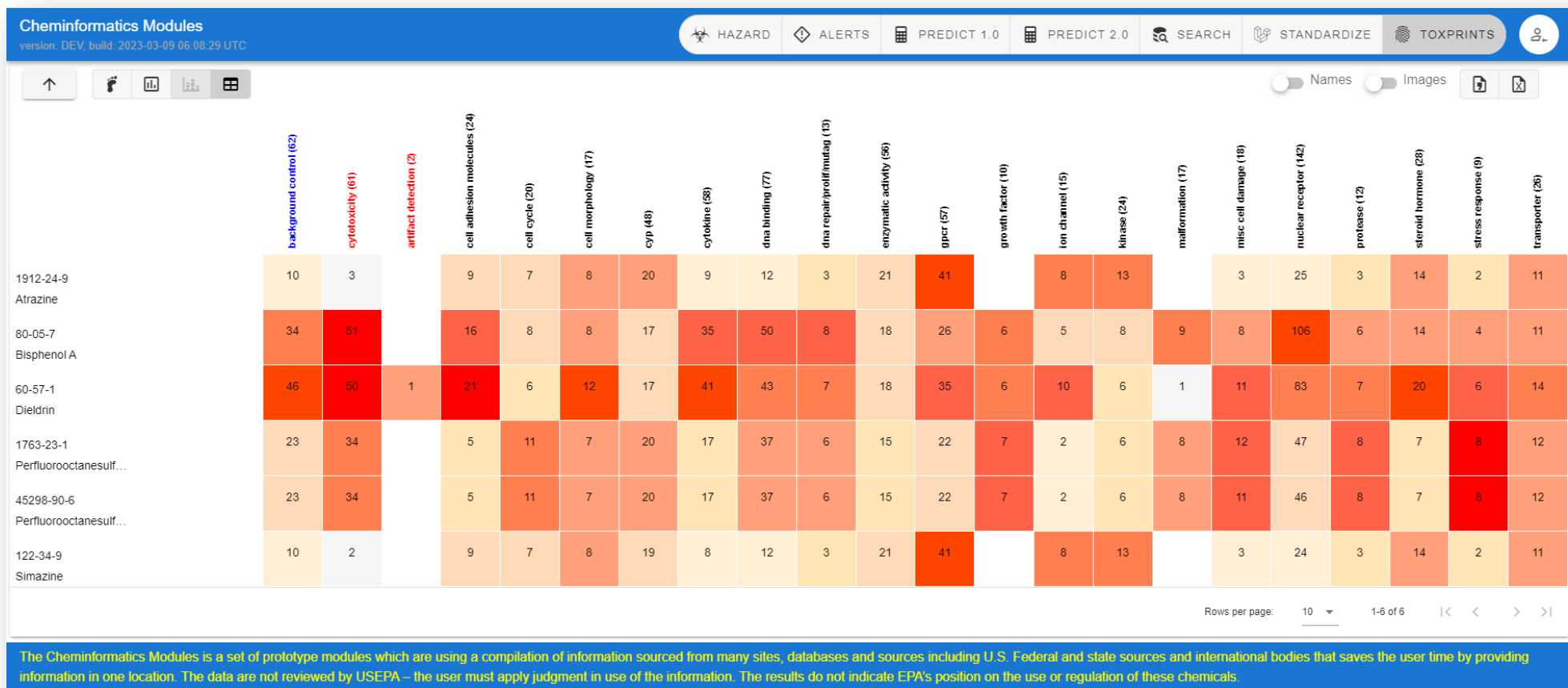
Cheminformatics Modules  
version: DEV, build: 2023-03-09 06:08:29 UTC

HAZARD ALERTS PREDICT 1.0 PREDICT 2.0 SEARCH STANDARDIZE TOXPRINTS

↑ Names

Structure	Labels	ToxPrints
	bond:CN_amine_aliphatic_generic	94
	bond:CN_amine_aromatic_generic	98
	bond:CN_amine_sec-NH_alkyl	103
	bond:CN_amine_sec-NH_aromatic	104
	bond:CN_amine_sec-NH_aromatic_aliphatic	105
	bond:CN_amine_sec-NH_generic	106
	bond:CX_halide_aromatic-X_generic	185
	bond:X[any]_halide	302
	chain:alkaneBranch_isopropyl_C3	422
	chain:alkaneLinear_ethyl_C2(H_gt_1)	437
ring:hetero_6_N_triazine_(1_3_5-)	663	
ring:hetero_6_N_triazine_generic	664	
ring:hetero_6_Z_1_3_5-	677	
ring:hetero_6_Z_generic	679	
	bond:COH_alcohol_aromatic	123
	bond:COH_alcohol_aromatic_phenol	124
	bond:COH_alcohol_generic	129
	chain:alkaneBranch_neopentyl_C5	424
	chain:alkaneLinear_ethyl_C2(connect_noZ_CN=4)	438
	chain:aromaticAlkane_Ar-C-Ar	474
	chain:aromaticAlkane_Ph-C1_acyclic_connect_noDbIBd	476
	chain:aromaticAlkane_Ph-C1_acyclic_generic	477
	chain:aromaticAlkane_Ph-C1-Ph	479
	ring:aromatic_benzene	586

# Examine ToxPrint Enrichment Statistics



# Work-in-Progress

- Preparing for a new release which will include Safety Module



# Not all PoCs are public. Work-in-progress

- Safety Module

version: , build: HAZARD SAFETY ALERTS PREDICT 1.0 PREDICT 2.0 SEARCH STANDARDIZE TOXPRINTS

Search chemical by Name, CASRN or DTXSID   Show Structure Full

Chemical	Safety	Properties	Signal	Explosive	Flammable	Oxidizers	Compressed Gas	Corrosive	Acute Toxicity	Irritant	Health Hazard	Env. Hazard	NFPA 704	Fire Fighting	Accidental Release Measures	Handling and Storage	Stability and Reactivity	Transport Information	Regulatory Information	Other Safety Information	RQ Category	RQ in pounds (kilograms)
<small>AIGBT</small> 71-43-2 Benzene			Danger																		A	10 (4.54)
<small>AIGBT</small> 75-01-4 Vinyl chloride			Danger																			
<small>GBT</small> 25265-71-8 Dipropylene glycol																						
<small>AIGBT</small> 57-55-6 1,2-Propylene glycol																						
<small>GBTM</small> 111-46-6 Diethylene glycol			Danger																			
<small>AIGBT</small> 111-76-2 2-Butoxyethanol			Danger																			
<small>GBTM</small> 115-11-7 Isobutene			Danger																			
<small>GBTM</small> 141-32-2 Butyl acrylate			Danger																			
<small>GBTM</small> 1322-13-0 Ethylhexyl acrylate																						

The Cheminformatics Modules is a set of prototype modules which are using a compilation of information sourced from many sites, databases and sources including U.S. Federal and state sources and international bodies that saves the user time by providing information in one location. The data are not reviewed by USEPA – the user must apply judgment in use of the information. The results do not indicate EPA's position on the use or regulation of these chemicals.

# Work-in-Progress

- Preparing for a new release which will include Safety Module
- Data will be updated with latest curated data PLUS support for chemicals with no structures (cannot do QSAR prediction)
- Structure alerts will be updated with support for:

- PFAS

## A New CSRML Structure-Based Fingerprint Method for Profiling and Categorizing Per- and Polyfluoroalkyl Substances (PFAS)

Ann M. Richard\*, Ryan Lougee, Matthew Adams, Hannah Hidle, Chihae Yang, James Rathman, Tomasz Magdziarz, Bruno Bienfait, Antony J. Williams, and Grace Patlewicz

- AIM



Computational Toxicology  
Volume 25, February 2023, 100256



Development of a CSRML version of the Analog identification Methodology (AIM) fragments and their evaluation within the Generalised Read-Across (GenRA) approach

Matthew Adams<sup>a, b</sup>, Hannah Hidle<sup>a, b</sup>, Daniel Chang<sup>b</sup>, Ann M. Richard<sup>b</sup>, Antony J. Williams<sup>b</sup>, Imran Shah<sup>b</sup>, Grace Patlewicz<sup>b</sup>  

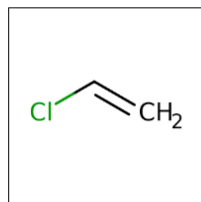
# Work-in-Progress

- All Hazard and Safety profiles, and TEST predictions will be pre-generated for faster response
- “Kubernetes-support” and “queueing” will spin up additional processors for faster predictions – should support 1000s of chemicals for TEST predictions
- Adding support for multiple other prediction engines such as OPERA

# Not all PoCs are public. Work-in-progress

- Analytical Methods and Open Spectral Database

78 Results for "vinyl chloride"



(Preferred) Name: Vinyl chloride  
DTXSID: [DTXSID8021434](#)  
CASRN: 75-01-4  
InChIKey: BZHJMEDXRYGGRV-UHFFFAOYSA-N  
Molecular Formula: C2H3Cl  
Mass: 61.99233

[Download Results](#)

Display Single Point Spectra

Include MS-Ready methods

[All Results \(78\)](#) [Methods \(44\)](#) [Spectra \(28\)](#) [Fact Sheets \(6\)](#)

Methodology	Source	Method #	#	Information
GC	<a href="#">USEPA</a>	EPA-TO-17	41	Determination of VOCs residues in air by GC.
GC/ELCD; GC/MS	<a href="#">NEMI</a>	SMC-6200	66	Determination of Organohalides residues in water (ground, waste,
GC/ELCD; GC/PID	<a href="#">NEMI</a>	EPA-502.2	60	Determination of VOCs residues in water (drinking, raw) by GC/EL
GC/FID	<a href="#">OSHA</a>	OSHA-75	1	Determination of Vinyl chloride residues in air by GC/FID with a LC
GC/FID	<a href="#">OSHA</a>	OSHA-4	1	Determination of Vinyl Chloride residues in air by GC/FID with a LC
GC/MS	<a href="#">USEPA</a>	EPA-601	29	Determination of Halocarbons residues in municipal and industrial
GC/MS	<a href="#">USEPA</a>	EPA-624.1	144	Determination of Purgeable organic pollutants residues in water (r
GC/MS	<a href="#">USEPA</a>	EPA-1624B	33	Determination of VOCs residues in water (municipal waste, industr
GC/MS	<a href="#">USEPA</a>	EPA-8260D	179	Determination of VOCs residues in various air sampling trapping n
GC/MS	<a href="#">USEPA</a>	EPA-8261	106	Determination of VOCs residues in water, soil, sediment, sludge, o
GC/MS	<a href="#">Agilent</a>	5994-3834	61	Determination of VOCs residues in water by GC/MS.
GC/MS	<a href="#">Agilent</a>	5991-6539	48	Determination of VOCs residues in water by GC/MS with a LOD of
GC/MS	<a href="#">USGS</a>	5-B12	102	Determination of VOCs residues in water by GC/MS.
GC/MS	<a href="#">USEPA</a>	CTM-028	35	Determination of VOCs residues in stationary source emissions by
GC/MS	<a href="#">NEMI</a>	USGS-0-3115-83	27	Determination of VOCs residues in water by GC/MS with a LOD of

## Volatile Organic Compounds in Water by Purge and Trap Capillary Column Gas Chromatography with Photoionization and Electrolytic Conductivity Detectors in Series

Author: USEPA

Focus/Analyte: VOCs

Limitation: LOD of 0.01-2.2ppb

Synopsis: Determination of VOCs residues in water (drinking, raw) by GC/ELCD; GC/PID with a LOD of 0.01-2.2ppb.

[PDF Viewer](#)

[Substances \(60\).\(grid\)](#)

[Substances \(60\).\(table\)](#)

L:\APPS\s... 1 / 35 67%

METHOD 502.2 VOLATILE ORGANIC COMPOUNDS IN WATER BY PURGE AND TRAP CAPILLARY COLUMN GAS CHROMATOGRAPHY WITH PHOTOIONIZATION AND ELECTROLYTIC CONDUCTIVITY DETECTORS IN SERIES

Revision 2.1

Edited by J.W. Munch (1995)

# Not all PoCs are public. Work-in-progress

- Chemical Transformation database (ChET)

## Reaction Map For Phosmet

Show Instructions

Show/Hide Map	Highlight Map	Species	Reference
<a href="#">Map 113</a>	<a href="#">Map 113</a>	Cherry / Stone fruit	A, Barnes, J. P.; Goldsby, G., Phosmet metabolism in an orchard tree fruit - cherry - (Volume 1), (p. 94), (1989); B, Codrea, E., f
<a href="#">Map 116</a>	<a href="#">Map 116</a>	Corn	A, Toia, R.F., Patrick, G., Ewing, A.D., Kimmel, E., A Metabolism Study with 14C-Phosmet on Corn, (p. 264), (1993)
<a href="#">Map 117</a>	<a href="#">Map 117</a>	Apples	A, Mohajeri, S.; Zane, J., METABOLISM OF [14C]PHOSMET BY APPLES (MALUS DOMESTICA), (p. 161), (2015)
<a href="#">Map 118</a>	<a href="#">Map 118</a>		A, The nature of the residues of orally administered (carbonyl-14C)-Phosmet in tissues and eggs of laying hens , (p. 154), (19
<a href="#">Map 119</a>	<a href="#">Map 119</a>		A, The Nature of the Residues of Orally Administered [carbonyl-14C]-Phosmet in Tissues and Milk of Lactating Goats, (p. 117
<a href="#">Map 114</a>	<a href="#">Map 114</a>	Potatoes	A, Toia, R. F.; Ewing, A. D.; Patrick, G.; Kimmel, E., Metabolism study with 14C Phosmet on potatoes, (p. 143), (1993); B, Rose, J.

Open Two or Three Maps to Compare Them

Show All Chemicals

Hide All Chemicals

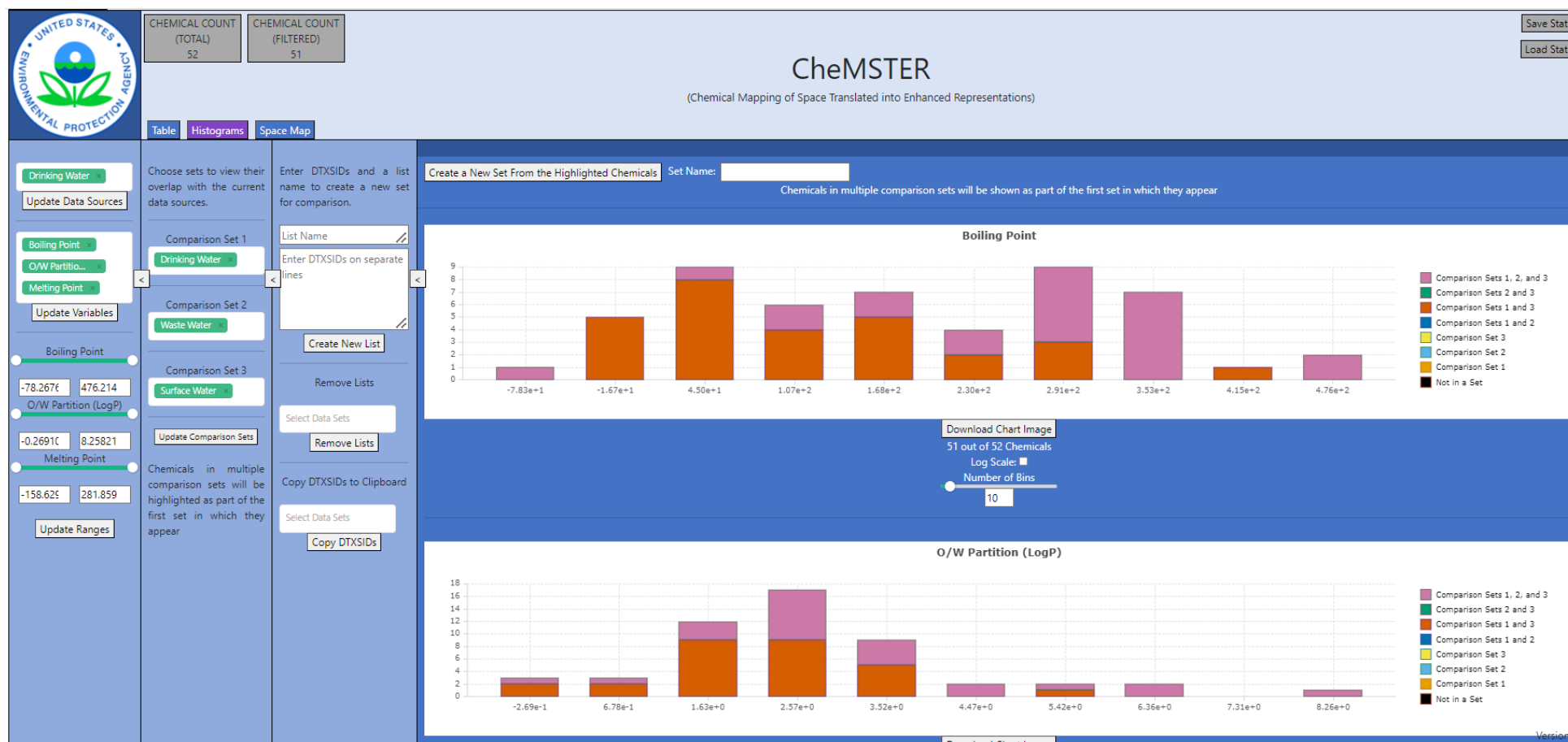
Search this map by DTXSID or Name:

Submit

Clear Search

# Not all PoCs are public. Work-in-progress

- ChemSTER: Chemical Mapping of Space  
Translated into Enhanced Representations



# Interested in a demo?

- Cheminformatics Modules:

<https://www.epa.gov/comptox-tools/cheminformatics>

- Contact for follow up demo(s):
- Antony Williams [williams.antony@epa.gov](mailto:williams.antony@epa.gov)
- Todd Martin [martin.todd@epa.gov](mailto:martin.todd@epa.gov)

Demo screen shots (just in case)



# Final record used to assign score on hover

Chemicals: 21

Toxicity: VH - Very High H - High M - Medium L - Low I - Inconclusive N/A - Not Applicable Authority: **Authoritative** ⓘ Screening ⓘ QSAR Model ⓘ

CAS Name	Human Health Effects															Ecotoxicity		Fate		
	Acute Mammalian Toxicity			Carcinogenicity	Genotoxicity Mutagenicity	Endocrine Disruption	Reproductive	Developmental	Neurotoxicity		Systemic Toxicity		Skin Sensitization	Skin Irritation	Eye Irritation	Acute Aquatic Toxicity	Chronic Aquatic Toxicity	Persistence	Bioaccumulation	Exposure
	Oral	Inhalation	Dermal						Repeat Exposure	Single Exposure	Repeat Exposure	Single Exposure								
<a href="#">79-06-1</a> AIGBT Acrylamide	H	M	M	VH	VH	I	M	H	H	H	H	H	H	H	H	M	M	M	L	H
<a href="#">79-01-6</a> AIGBT Trichloroethylene	L	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
<a href="#">108-95-2</a> AIGBT Phenol	H	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
<a href="#">50-00-0</a> AIGBT Formaldehyde	H	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
<a href="#">111-30-8</a> AGBTM Glutaraldehyde	H	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
<a href="#">302-01-2</a> IGBTP Hydrazine	H	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M
<a href="#">75-21-8</a> AIGBT Acetone	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M	M

**Acrylamide | Acute Mammalian Toxicity Oral**

**Source** ECHA CLP

**Authority** Authoritative

**Score** H

**Route** oral

**Category** Acute Tox. 3

**Hazard Code** H301

**Hazard Statement** Toxic if swallowed

**Rationale** Score of H was assigned based on a hazard code of H301

**Note**

# All records for a given score on click

Cheminformatics Modules  
version: DEV, build: 2023-03-10

Acute Mammalian Toxicity Oral for Acrylamide ( 79-06-1 )

Toxicity: VH - Very High H - High M - Medium L - Low I - Inconclusive N/A - Not Applicable Authority: Authoritative  ⓘ Screening ⓘ QSAR Model ⓘ

Source	Score	Route	Test Organism	Test Type	Category	Hazard Code	Hazard Statement	Rationale	Note
ECHA CLP	<span style="color: orange;">H</span>	oral			Acute Tox. 3	H301	Toxic if swallowed	Score of H was assigned based on a hazard code of H301	
Australia	<span style="color: orange;">H</span>	oral			Acute toxicity - category 3	H301	Toxic if swallowed	Score of H was assigned based on a hazard code of H301	N (The classification information for this entry was provided by the National Industrial Chemical Notification and Assessment Scheme)
Canada	<span style="color: orange;">H</span>	oral			Acute toxicity - oral - Category 3	H301	Toxic if swallowed	Score of H was assigned based on a hazard code of H301	Comments: This product could belong to the hazard class "Combustible dust", based on various factors related to the combustibility and explosiveness of its dust, including composition, shape and size of the particles.
Hydra	<span style="color: orange;">H</span>	oral						50 mg/kg <	

ChemIDplus AIGBT

42

# Scoring Dictionaries available in headers

hcd.rtpnc.epa.gov/#/hazard/report

Cheminformatics Modules  
version: DEV, build: 2023-03-09 06:08:29 UTC

STANDARDIZE TOXPRIN

Acute Mammalian Toxicity Oral

Source	Endpoint	VH	H	M	L
DfE criteria	Oral LD50 (mg/kg)	≤ 50	> 50 - 300	> 300 - 2000	> 2000
DfE criteria	EU CLP	H300	H301	H302	
ChemIDplus; TEST Predicted *	Oral LD50 (mg/kg)	≤ 50	> 50 - 300	> 300 - 2000	> 2000
Australia; Canada; ECHA CLP; Malaysia	GHS Code	H300	H301	H302	
Denmark	Category	AcuteTox1 and AcuteTox2	AcuteTox3	AcuteTox4	
Japan	GHS Code	H300	H301	H302	H303 Not Classified
New Zealand	Category	Category 6.1A Category 6.1B	Category 6.1C	Category 6.1D	Category 6.1E
TSCA Work Plan			Acute mammalian toxicity		
UMD	Acute toxin				

\* TEST Predicted predicts rat LD50 values. ChemIDplus LD50 values for rats, mice, rabbits, and guinea pigs were included.

↑

Skipped (0)  
 Unlikely (0)  
 Filters (0)  
 Sorting (0)  
 Structure

CAS Name

	Acute	Oral	Ecotoxicity	Fate																
			Acute Aquatic Toxicity	Chronic Aquatic Toxicity	Persistence	Bioaccumulation	Exposure													
Ethylene oxide	VH		M	L	M	L	VH													
Hydrazine hydrat...	VH		VH	VH																
4,4'-Diaminobiph...	H		VH	H	M	L	H													
Sodium dichromate	H		VH	VH		L														
Acrylonitrile	H	H	H	VH	VH	L	H	H	H	H	H	M	H	H	VH	H	H	M	L	VH
Morpholine	M	M	M	I	L	L	I	L			H	H	I	VH	VH	L	M	M	L	H
1,2-Dibromoethane	H	H	H	VH	VH	H	M	H		H	M	M	I	H	H	H	H	M	L	H

43

# List of sources by type from authority info buttons

## Authoritative sources





- European Chemicals Agency (ECHA) Classification Labeling and Packaging (CLP);
- EPA mid-Atlantic Region Human Health Risk-Based Concentrations;
- Germany Permanent Senate Commission for the Investigation of Health Hazards of Chemical Compounds in the Work Area;
- World Health Organization International Agency for Research on Cancer (IARC) Monographs on the Evaluation of Carcinogenic Risks to Humans;
- Integrated Risk Information System (IRIS);
- US National Institute for Occupational Safety and Health (NIOSH) list of potential occupational carcinogens;
- California Office of Environmental Health Hazard Assessment Proposition 65 List;
- EU European Chemicals Agency (ECHA) Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) Candidate List of Substances of Very High Concern for Authorization;
- US Department of Health and Human Services National Toxicology Program Report on Carcinogens

Reference: [An automated framework for compiling and integrating chemical hazard data](#)

Authoritative ⓘ Screening ⓘ QSAR Model ⓘ

	Ecotoxicity		Fate		
Eye Irritation	Acute Aquatic Toxicity	Chronic Aquatic Toxicity	Persistence	Bioaccumulation	Exposure

# CTS Metabolite Feature

1 generation ⌵ No Analogs ⌵    

No Metabolites  
**1 generation**  
 2 generations  
 3 generations  
 4 generations  
 up to the selected generation

Chemicals: 3

Toxicity: **VH** - Very High **H** - High **M** - Medium **L** - Low **I** - Inconclusive N/A - Not Applicable Authority: **Authoritative**  Screening  QSAR Model 

CAS Name	Human Health Effects															Ecotoxicity		Fate		
	Acute Mammalian Toxicity			Carcinogenicity	Genotoxicity Mutagenicity	Endocrine Disruption	Reproductive	Developmental	Neurotoxicity		Systemic Toxicity		Skin Sensitization	Skin Irritation	Eye Irritation	Acute Aquatic Toxicity	Chronic Aquatic Toxicity	Persistence	Bioaccumulation	Exposure
	Oral	Inhalation	Dermal						Repeat Exposure	Single Exposure	Repeat Exposure	Single Exposure								
115-86-6 <sup>HGBTM</sup> Triphenyl phosph...	M	L	L	I	L	H	L	L	L	I	L	I	I	L	H	<b>VH</b>	<b>VH</b>	M	H	M
838-85-7 <sup>M</sup> Diphenyl phosph... <b>LIKELY</b>	M				<b>VH</b>	L	M	H								L		M	L	H
108-95-2 <sup>AIGBT</sup> Phenol <b>LIKELY</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>H</b>	<b>M</b>	<b>H</b>	<b>H</b>	<b>VH</b>	<b>VH</b>	<b>H</b>	<b>H</b>	M	H	<b>VH</b>

- Skipped (0)
- Unlikely (3)
- Filters (0)
- Sorting (0)
- Structure

# Similar chemicals feature

Chemicals: 20

Toxicity: VH - Very High H - High M - Medium L - Low I - Inconclusive N/A - Not Applicable Authority: Authoritative ⓘ Screening ⓘ QSAR Model ⓘ

CAS Name	Human Health Effects															Ecotoxicity		Fate		
	Acute Mammalian Toxicity			Carcinogenicity	Genotoxicity Mutagenicity	Endocrine Disruption	Reproductive	Developmental	Neurotoxicity		Systemic Toxicity		Skin Sensitization	Skin Irritation	Eye Irritation	Acute Aquatic Toxicity	Chronic Aquatic Toxicity	Persistence	Bioaccumulation	Exposure
	Oral	Inhalation	Dermal						Repeat Exposure	Single Exposure	Repeat Exposure	Single Exposure								
<a href="#">115-86-6</a> <sup>HGBTM</sup> Triphenyl phosph...	M	L	L	I	L	H	L	L	L	I	L	I	I	L	H	VH	VH	M	H	M
<a href="#">4009-39-6</a> methyl phenyl hy 1.00	M				L	L		H								H			L	M
<a href="#">10113-28-7</a> Phosphoric acid, 1.00	M				H	L		H								VH			M	M
<a href="#">838-85-7</a> <sup>M</sup> Diphenyl phosph 1.00	M				VH	L	M	H								L		M	L	H
<a href="#">NOCAS_892675</a> Diphenyl hydroge 1.00	I				I	I		I								I			I	

- Skipped (0)
- Unlikely (0)
- Filters (0)
- Sorting (0)
- Structure

# Custom report feature

Chemicals: 22

Toxicity: **VH** - Very High **H** - High **M** - Medium **L** - Low **I** - Inconclusive **N/A** - Not Applicable **Au** - Available **Q** - QSAR Model

Emergency Response  
 Full  
 Custom  
**Emergency Response**  
 Site-Specific Screening

Skipped (0)  
 Unlikely (0)  
 Filters (0)  
 Sorting (0)  
 Structure

CAS Name	Human Health Effects										Ecotoxicity		
	Acute Mammalian Toxicity			Genotoxicity Mutagenicity	Neurotoxicity		Systemic Toxicity		Skin Sensitization	Skin Irritation		Eye Irritation	Acute Aquatic Toxicity
	Oral	Inhalation	Dermal		Single Exposure	Single Exposure	Single Exposure	Single Exposure					
79-06-1 Acrylamide	AIGBT	H	M	M	VH	H		H	H	H	M		
79-01-6 Trichloroethylene	AIGBT	L	M	L	VH	H		M	H	H	H		
108-95-2 Phenol	AIGBT	H	H	H	H	H		H	VH	VH	H		
50-00-0 Formaldehyde	AIGBT	H	H	H	H			M	H	VH	VH	H	
111-30-8 Glutaraldehyde	AGBTM	H	VH	H	VH	H		M	H	VH	VH	VH	
302-01-2 Hydrazine	IGBTP	H	H	H	VH	H		H	H	VH	VH	VH	
75-21-8	AIGBT												