# Cheminformatics Modules

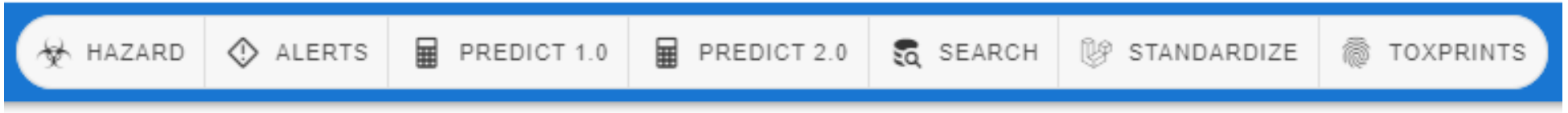
Todd Martin and Antony Williams
US EPA/ORD/CCTE

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

# Purpose of the Cheminformatics Modules

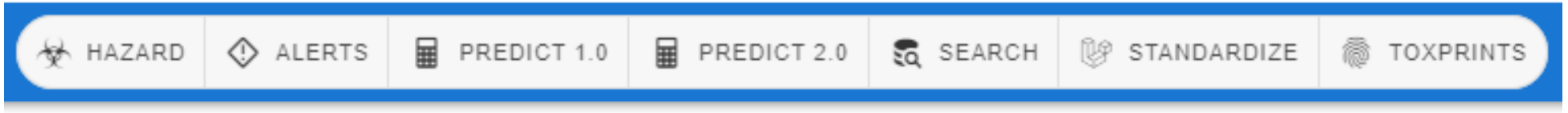
- PoCs are research software builds to **prove approaches** before moving into production software environments
- Assemble data, develop data model(s), **test** user interface **approaches**, work with test user base to **garner feedback**
- PoCs are internal access data refreshes and application updates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

# Cheminformatics Modules



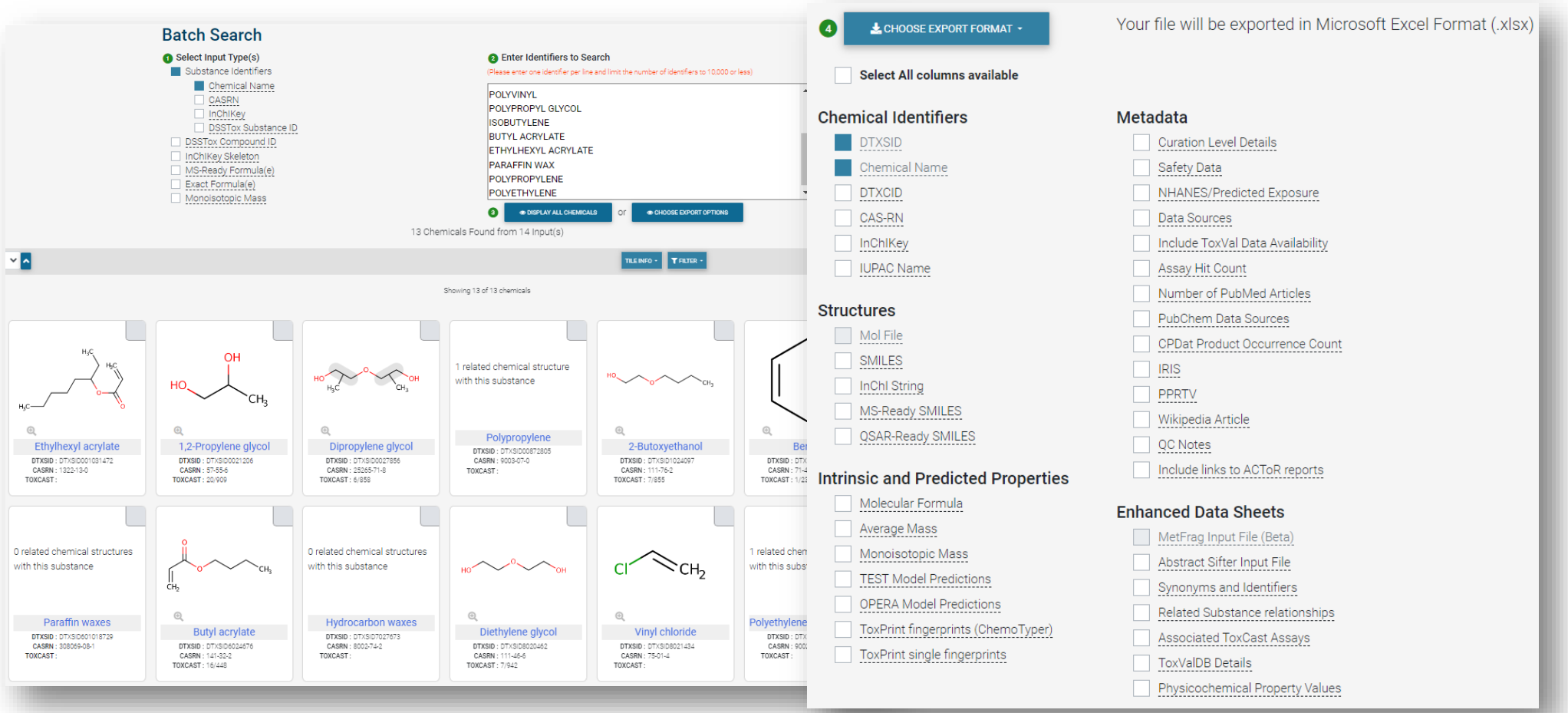

- Hazard Comparison module (MATURE: Background and Demo)
- Structure Alerts (DEVELOPMENT)
- PREDICT modules
  - 1.0 – Batch (MATURE)
  - 2.0 – New models (DEVELOPMENT)
- Structure/substructure/similarity searching (MATURE: Demo)
- Standardizer (DEVELOPMENT)
- ToxPrints (DEVELOPMENT)

# Cheminformatics Modules



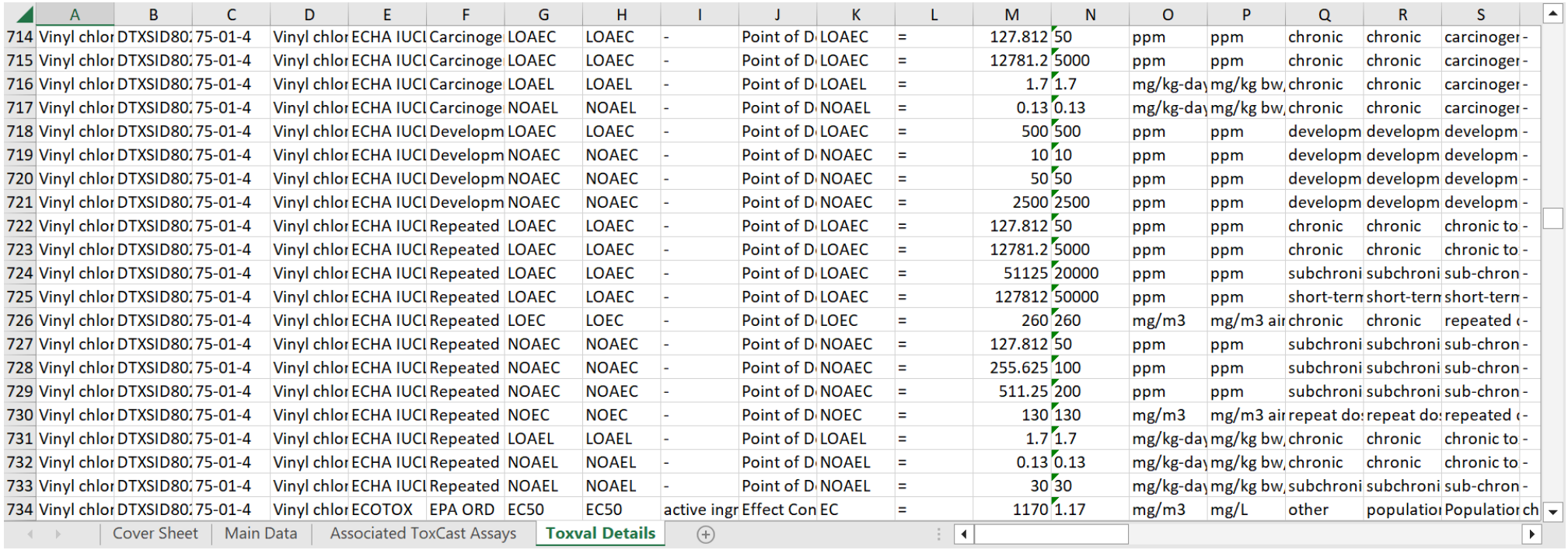

- Hazard Comparison module (MATURE: Background and Demo)
- Structure Alerts (DEVELOPMENT)
- PREDICT modules
  - 1.0 – Batch (MATURE)
  - 2.0 – New models (DEVELOPMENT)
- Structure/substructure/similarity searching (MATURE: Demo)
- Standardizer (DEVELOPMENT)
- ToxPrints (DEVELOPMENT)

# Hazard Module – Batch Toxicity Data



# 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-tern | short-tern | short-tern |
| 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 |

Cover Sheet | Main Data | Associated ToxCast Assays | **Toxval Details**

# 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**
  - ChemID*plus, 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*

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

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

## *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 | H | oral | | | Acute Tox. 3 | H301 | Toxic if swallowed | Score of H was assigned based on a hazard code of H301 | |
| Australia | H | 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 | H | 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 | H | 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

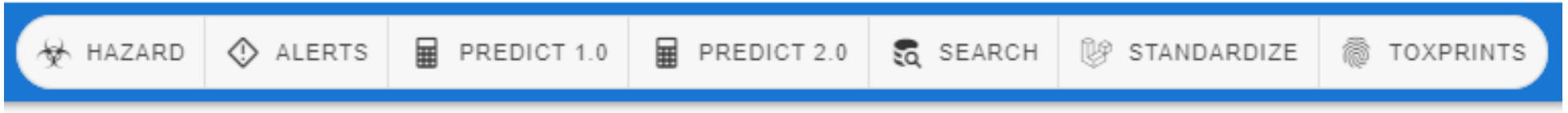
- Limitations
  - Data gaps
  - Automation limits the scope of data searching and quality assurance, particularly of primary sources

# Future Research Needs

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

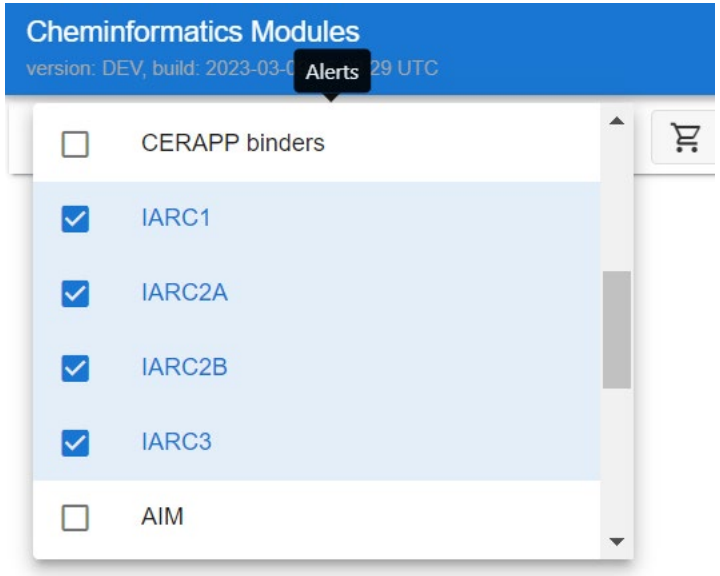
# Live DEMO

# Cheminformatics Modules

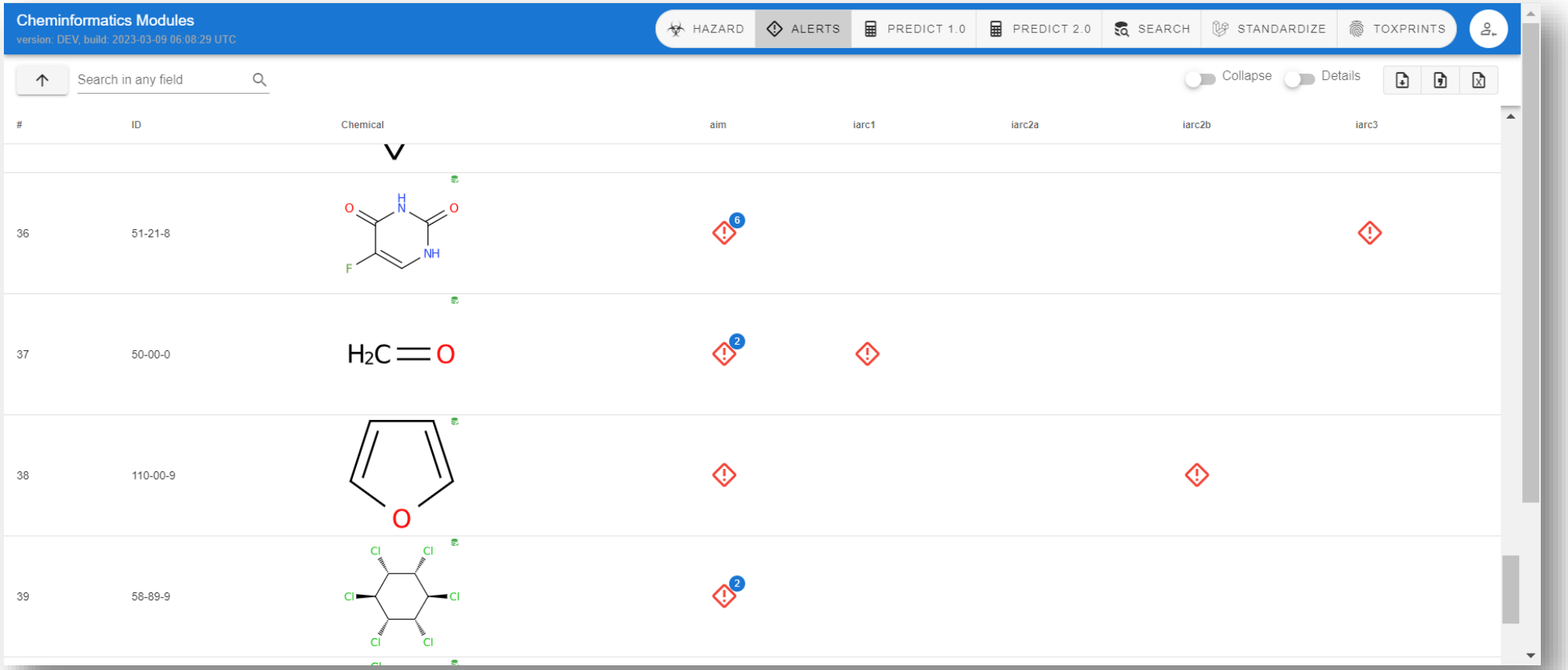



- Hazard Comparison module (MATURE: Background and Demo)
- Structure Alerts (DEVELOPMENT)
- PREDICT modules
  - 1.0 – Batch (MATURE)
  - 2.0 – New models (DEVELOPMENT)
- Structure/substructure/similarity searching (MATURE: Demo)
- Standardizer (DEVELOPMENT)
- ToxPrints (DEVELOPMENT)

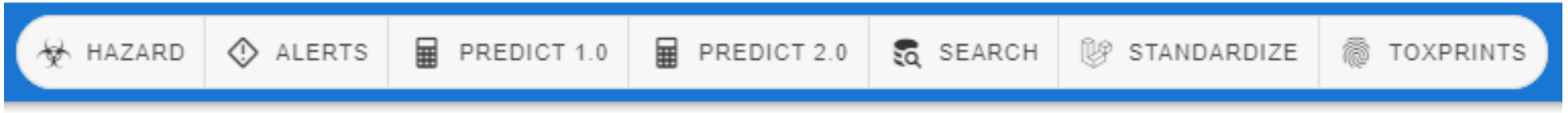
# Structure Alerts (Development)



# Structure Alerts (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)

# PREDICT 1.0 Batch Prediction

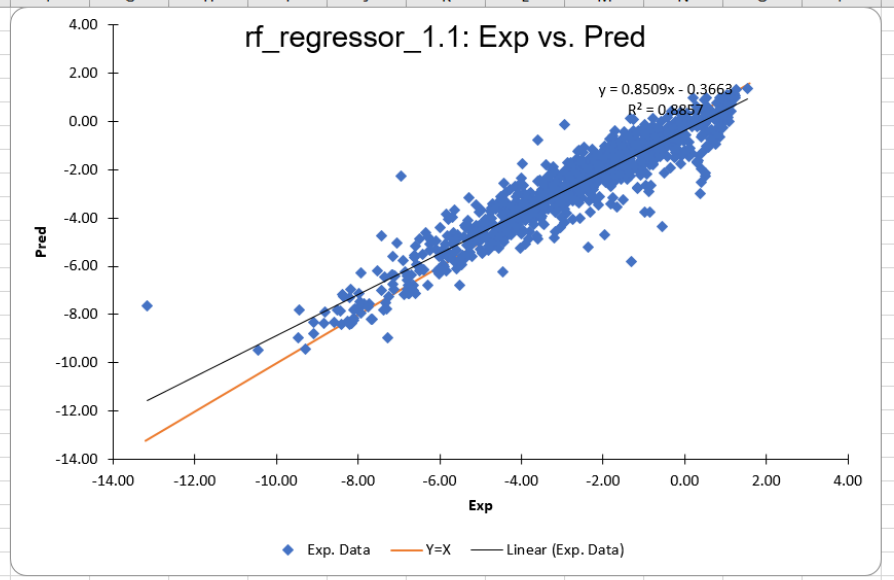
*Legend:*A = Active, N = Not active, I = Inconclusive, Experimental, *Predicted*

| Filters / Structure / Products<br>CAS<br>Name | 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 GBTM<br>Oxydemeton-met… | *3.35* | *5.81* | *I* | 3.91<br>*4.27* | *0.28* | *N* | A<br>*N* | *N* | | | *-5.19* | *50* | 1.31<br>*1.28* | 182<br>*186* | | | *I* | *1.10* |
| 115-29-7 AIGBT<br>Endosulfan | 6.10<br>*5.14* | 4.35<br>*3.35* | 8.48<br>*6.33* | 6.08<br>*5.36* | *I* | *N* | | N<br>*N* | | | *319* | -6.76<br>*-6.22* | 106<br>*90* | 1.94<br>*1.81* | 226<br>*227* | | | |
| 1646-75-9 GBT<br>Aldicarb oxime | *3.10* | *I* | *I* | 2.25<br>*2.39* | *I* | *N* | N<br>*N* | *N* | *I* | 210<br>*188* | *I* | 21<br>*28* | 1.00<br>*0.99* | 81<br>*70* | *31.43* | | | |
| 630-08-0 AGBT<br>Carbon monoxide | *I* | *I* | *I* | *I* | *I* | *I* | *I* | *I* | *I* | -192<br>*I* | 8.19<br>*I* | -205 | | *I* | *I* | *I* | *I* | *I* |
| 2312-35-8 IGBTM<br>Propargite | 6.21<br>*4.57* | 2.37<br>*1.93* | *5.82* | | *I* | *A* | | *N* | | *I* | *325* | -6.52<br>*-6.75* | *23* | 1.17<br>*1.16* | 226<br>*226* | | *I* | *I* |
| 115-90-2 GBTM<br>Fensulfothion | 2.19<br>*2.80* | 5.15<br>*4.52* | 3.85<br>*4.55* | *6.94* | N<br>*N* | *A* | *1.25* | *N* | | *I* | *I* | -4.30<br>*I* | *59* | 1.31<br>*1.37* | 198<br>*201* | | | *I* |

# PREDICT 2.0 – New model delivery

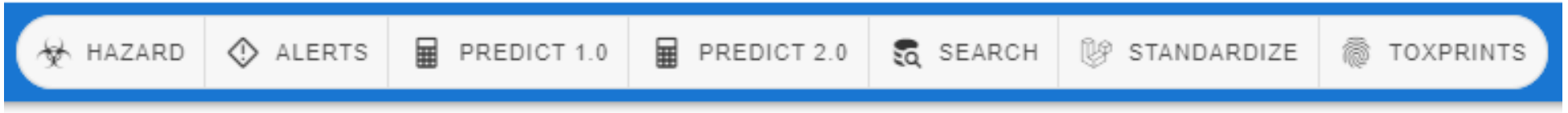
- Early development work for how we will deliver models in the future. The biggest benefit is data transparency

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Canonical QSAR Ready Smiles | Exp | Pre | Err |
| 2 | CCC(C(CC)C(O)=O)C(O)=O | -1.16 | -0.99 | 0.17 |
| 3 | CCOC1C=C(C=CC=1C1COC(=N1)C1C(F)=CC=CC=1F)C(C)(C)C | -6.69 | -6.37 | 0.33 |
| 4 | BrCCCBr | -2.08 | -2.06 | 0.02 |
| 5 | CC1C2=CC=CC=C2C2C=CC3=CC=CC4C=CC=1C=2C=43 | -8.52 | -7.83 | 0.70 |
| 6 | OCCNC1C=CC=CC=1 | -0.47 | -0.86 | 0.38 |
| 7 | CC1C=CC2=C(N=1)N(CC)C=C(C(O)=O)C2=O | -3.42 | -2.38 | 1.03 |
| 8 | CC1=CC2C=CC(Cl)=C(C(O)=O)C=2N=C1 | -1.66 | -3.53 | 1.87 |
| 9 | CCN(CC)C=O | 1.00 | 0.19 | 0.80 |
| 10 | ClC1(Cl)C2(Cl)C3CCC4C(CCC3C1(Cl)C(Cl)=C2Cl)C1(Cl)C(Cl)=C(Cl)C4(Cl)C1(Cl)Cl | -13.17 | -7.62 | 5.55 |
| 11 | CC(C)C1C=CC(=CC=1)NC(=O)N(C)C | -3.50 | -2.89 | 0.61 |
| 12 | CC1=CC(C)=NC(N)=N1 | -0.49 | -0.67 | 0.18 |
| 13 | [O-][N+](=O)C1C=C(C=CC=1)[N+]([O-])=O | -2.50 | -2.77 | 0.27 |
| 14 | CCCC1=CC2OCOC=2C=C1COCCOCCOCCCC | -4.37 | -4.06 | 0.31 |
| 15 | CC(C)OC(=O)C1(CCCC1)S(=O)(=O)C1C=CC(Cl)=CC=1 | -4.65 | -4.34 | 0.31 |
| 16 | S=C=NC1=CC=CC2=CC=CC=C21 | -4.60 | -4.52 | 0.08 |
| 17 | CC1C=C(Cl)C=CC=1OCCCC(=O)OCC | -4.41 | -3.54 | 0.87 |
| 18 | CNC(=O)N(C)C1=NC2C=CC=CC=2S1 | -3.57 | -2.99 | 0.58 |
| 19 | CCCN(CCC)C(=O)SCC1C=CC=CC=1 | -4.28 | -4.26 | 0.02 |
| 20 | CCCCOCCOC(N)=O | -0.51 | -0.60 | 0.09 |
| 21 | COC1C=CC(CO)=CC=1 | -1.84 | -0.75 | 1.09 |
| 22 | CC1C=C(Cl)C=CC=1NC(=S)N(C)C | -3.66 | -3.61 | 0.05 |
| 23 | OC1C=C(C=O)C=C(Br)C=1 | -3.25 | -2.01 | 1.24 |
| 24 | C=CC(Cl)CCl | -2.47 | -1.90 | 0.57 |
| 25 | CCC(C)(C)C1CC(=NO)C(CC1)=NO | -3.97 | -1.74 | 2.23 |
| 26 | CC(=O)OC12C3C(OC(=O)C4C=CC=CC=4)C(O)(CC3C34C5C(OC)C1C3N(CC5(COC)C | -3.32 | -2.67 | 0.65 |
| 27 | ClC1=C(C=C(Cl)C(Cl)=C1Cl)C1=C(Cl)C(Cl)=CC=C1Cl | -8.59 | -8.31 | 0.28 |




Cover sheet | Summary sheet | Training | Test | svm_regressor_1.1 | rf_regressor_1.1 | xgb_regre ...

# Cheminformatics Modules

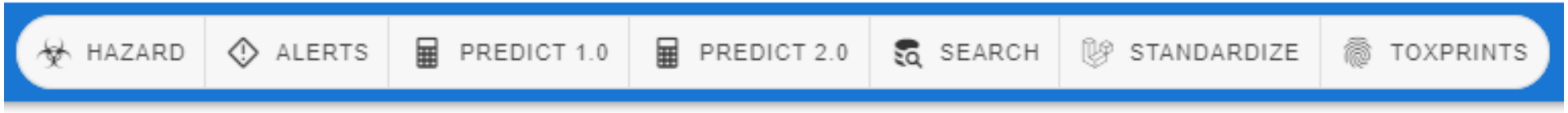



- Hazard Comparison module (MATURE: Background and Demo)
- Structure Alerts (DEVELOPMENT)
- PREDICT modules
  - 1.0 – Batch (MATURE)
  - 2.0 – New models (DEVELOPMENT)
- Structure/substructure/similarity searching (MATURE: Demo)
- Standardizer (DEVELOPMENT)
- ToxPrints (DEVELOPMENT)

# Structure/Substructure/Similarity

LIVE DEMO

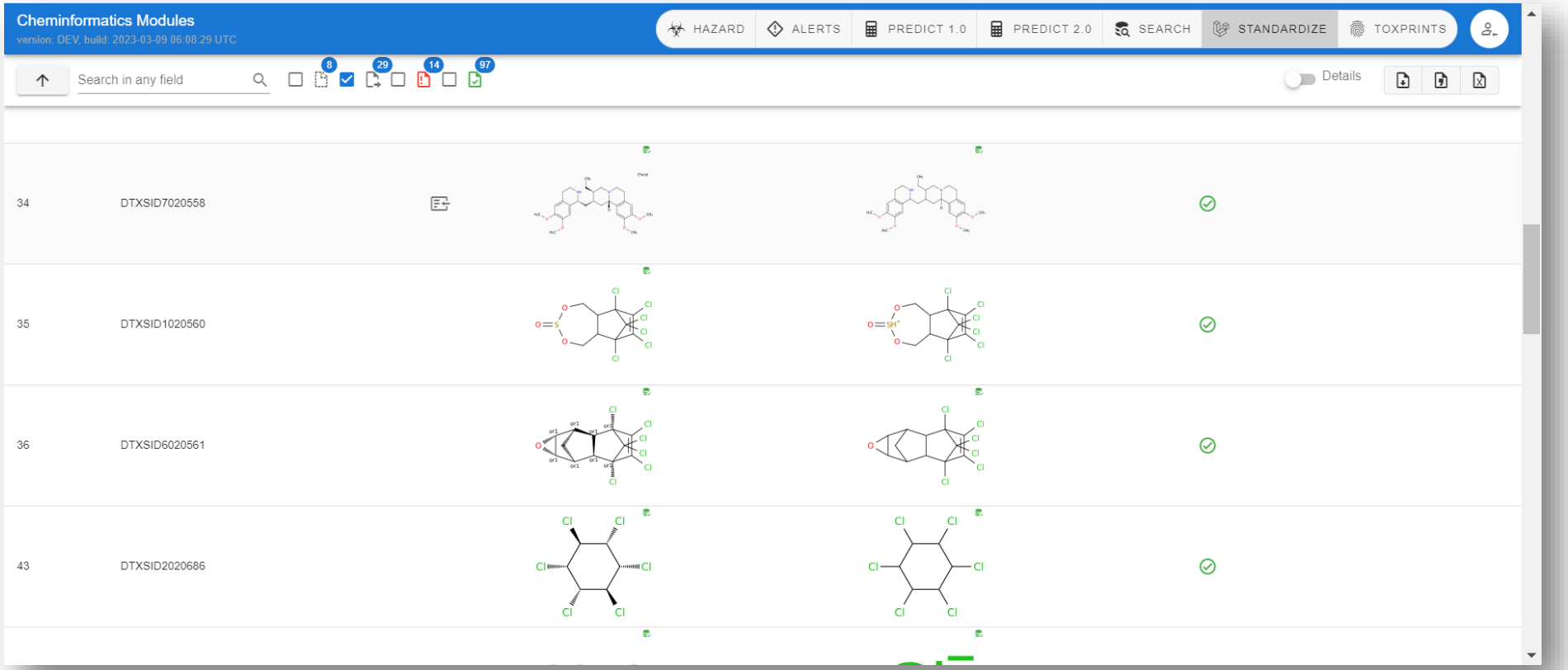
# Cheminformatics Modules



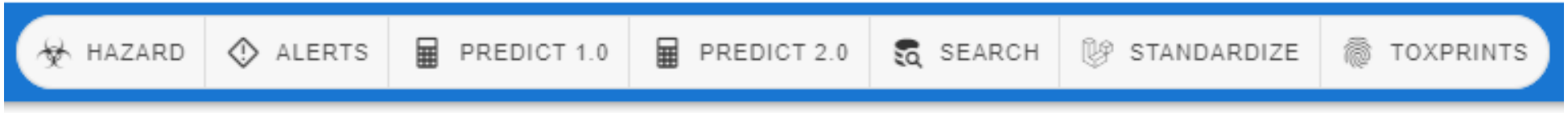

- Hazard Comparison module (MATURE: Background and Demo)
- Structure Alerts (DEVELOPMENT)
- PREDICT modules
  - 1.0 – Batch (MATURE)
  - 2.0 – New models (DEVELOPMENT)
- Structure/substructure/similarity searching (MATURE: Demo)
- Standardizer (DEVELOPMENT)
- ToxPrints (DEVELOPMENT)

# Generation of QSAR-Ready/MS-Ready

- Preparing chemicals for modeling and Linking in CCD

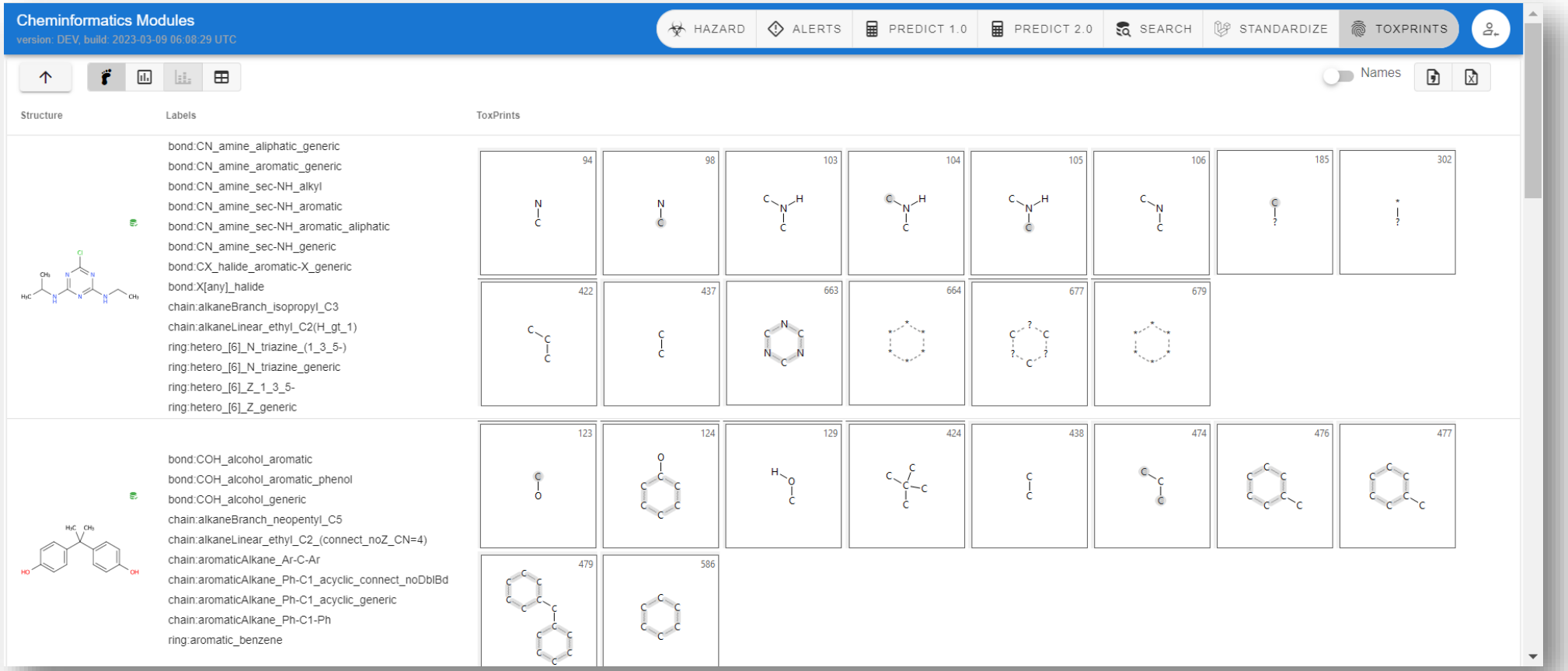


# Cheminformatics Modules

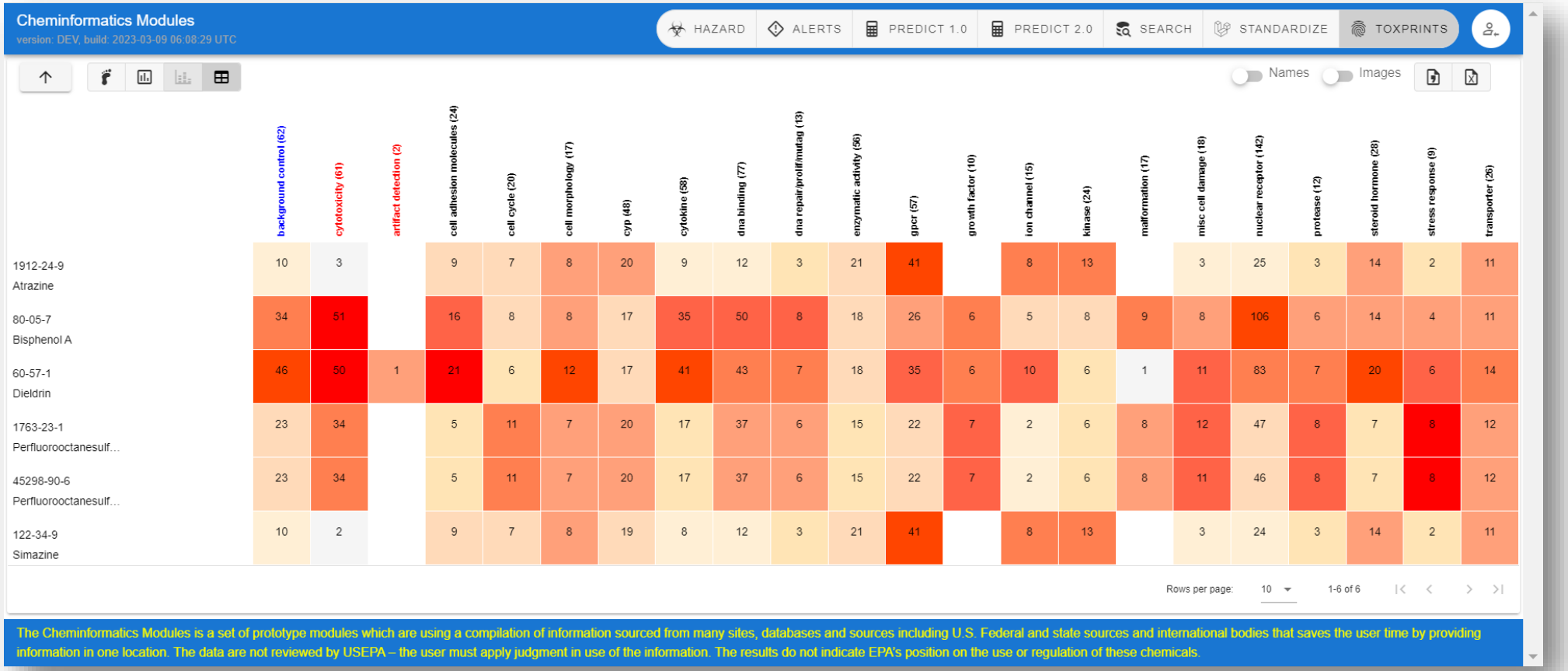



- Hazard Comparison module (MATURE: Background and Demo)
- Structure Alerts (DEVELOPMENT)
- PREDICT modules
  - 1.0 – Batch (MATURE)
  - 2.0 – New models (DEVELOPMENT)
- Structure/substructure/similarity searching (MATURE: Demo)
- Standardizer (DEVELOPMENT)
- ToxPrints (DEVELOPMENT)

# ToxPrint chemotype generation



# Examine ToxPrint Enrichment Statistics

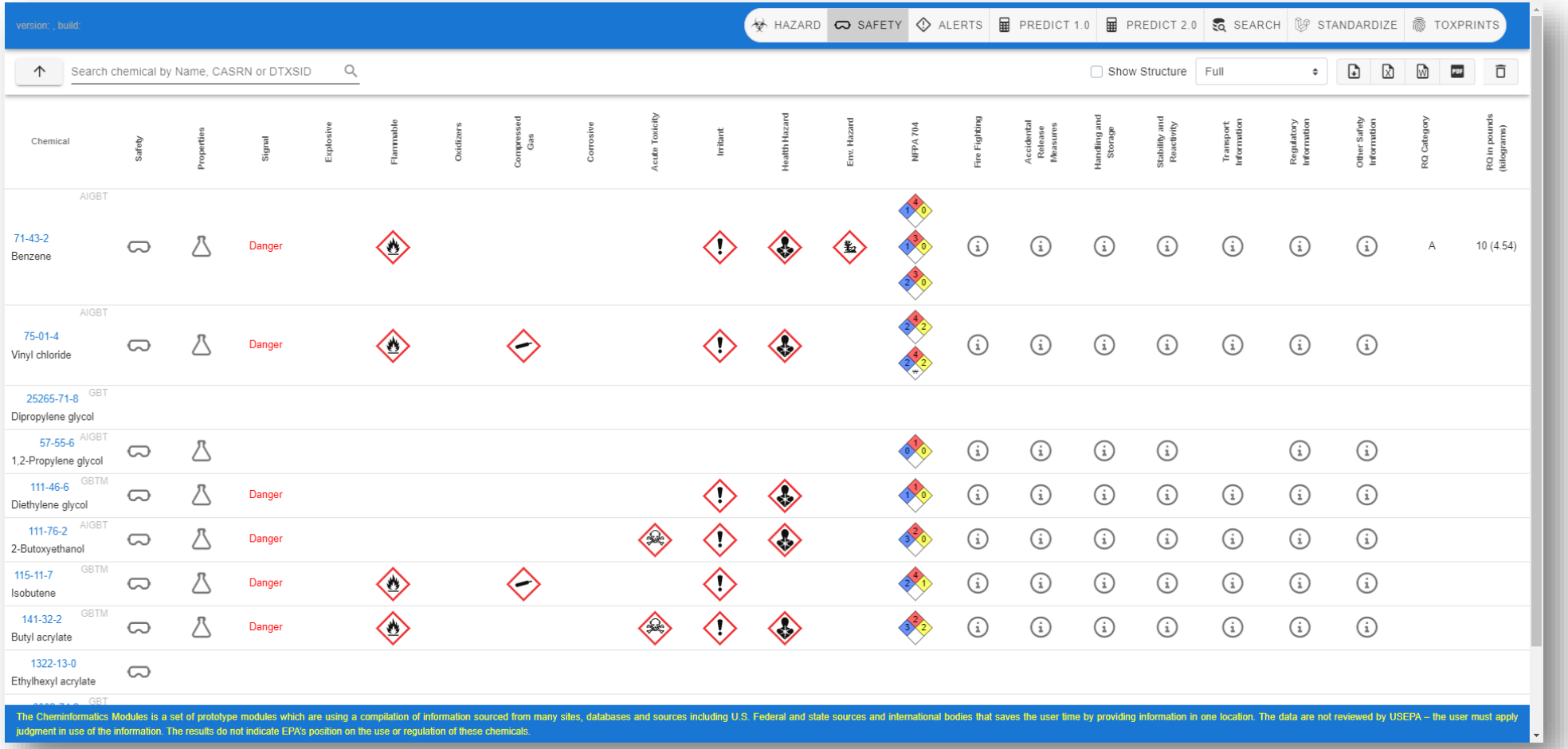


# Work-in-Progress

- Preparing for a new release which will include Safety Module

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

- Safety Module



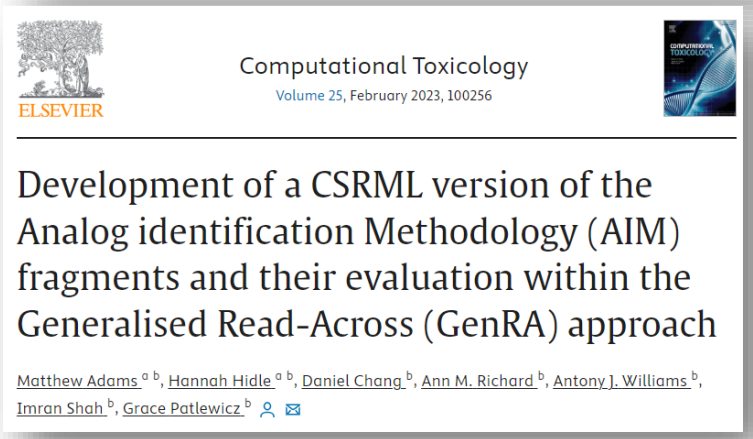
# Work-in-Progress

- 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

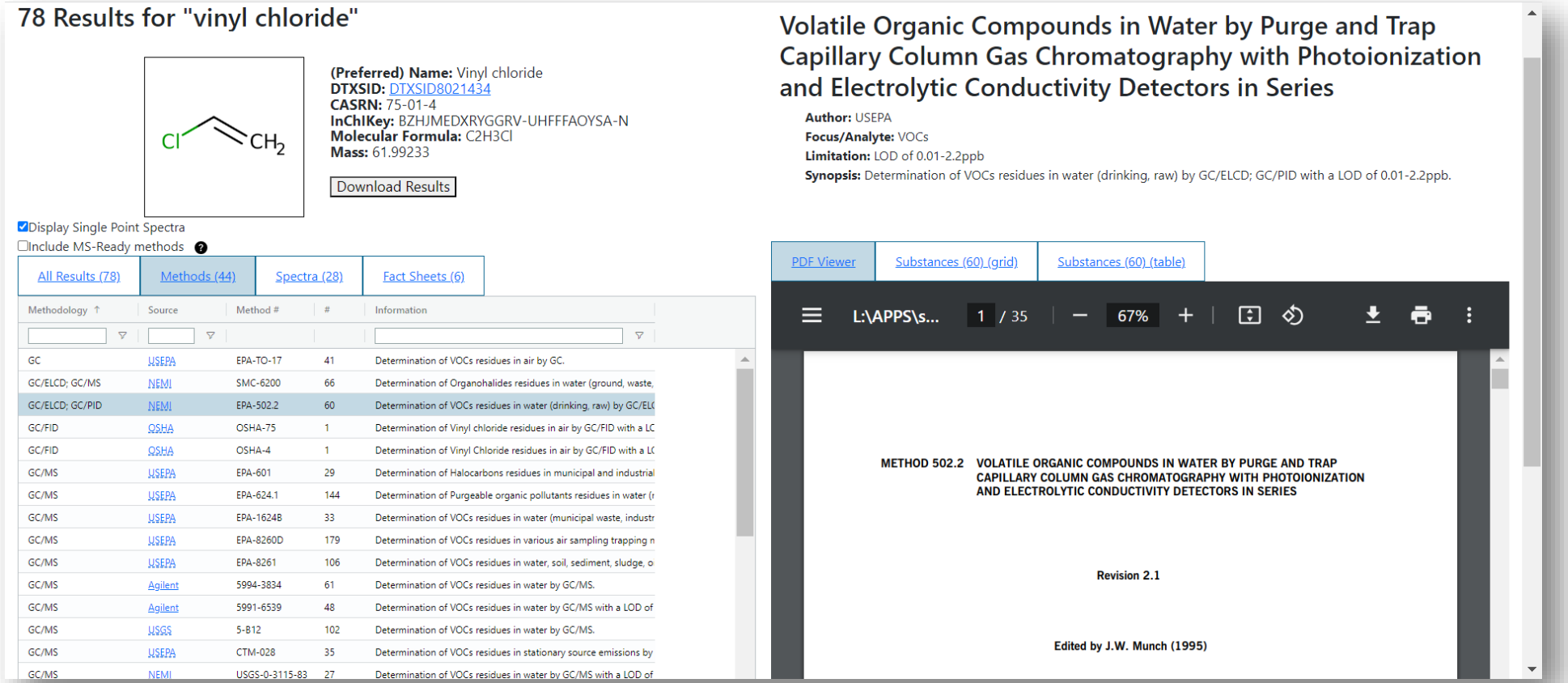


# Work-in-Progress

- 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

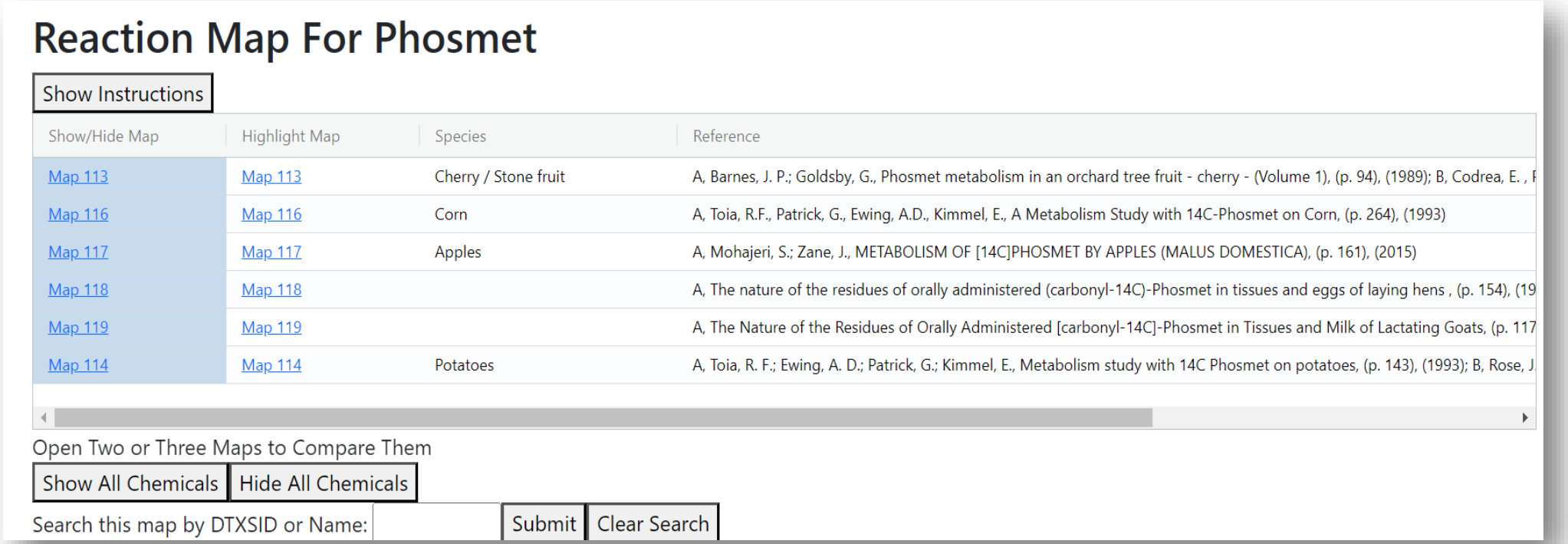
# Not all PoCs are public. Work-in-progress

- Analytical Methods and Open Spectral Database



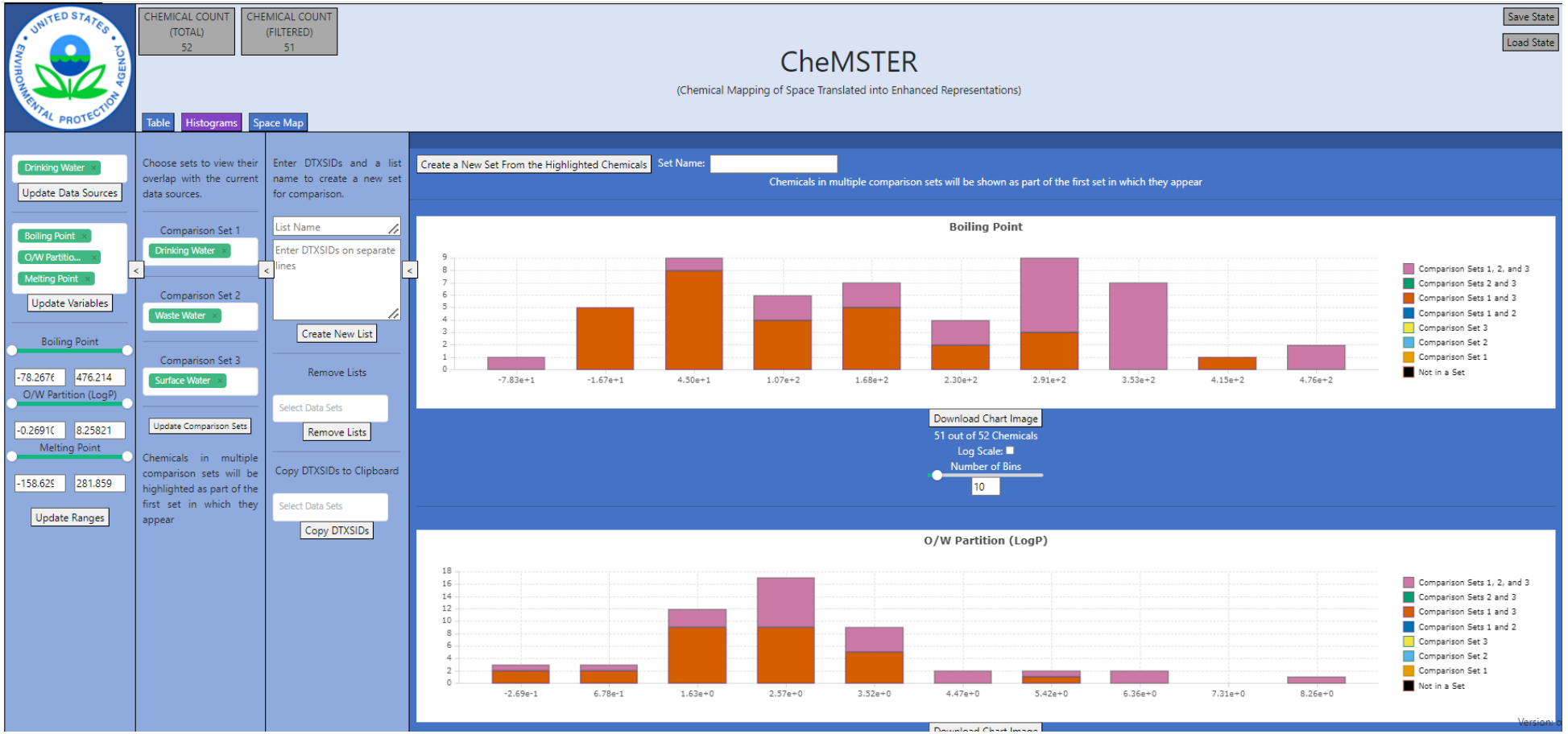
# Not all PoCs are public. Work-in-progress

- Chemical Transformation database (ChET)



# 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
- Todd Martin 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*

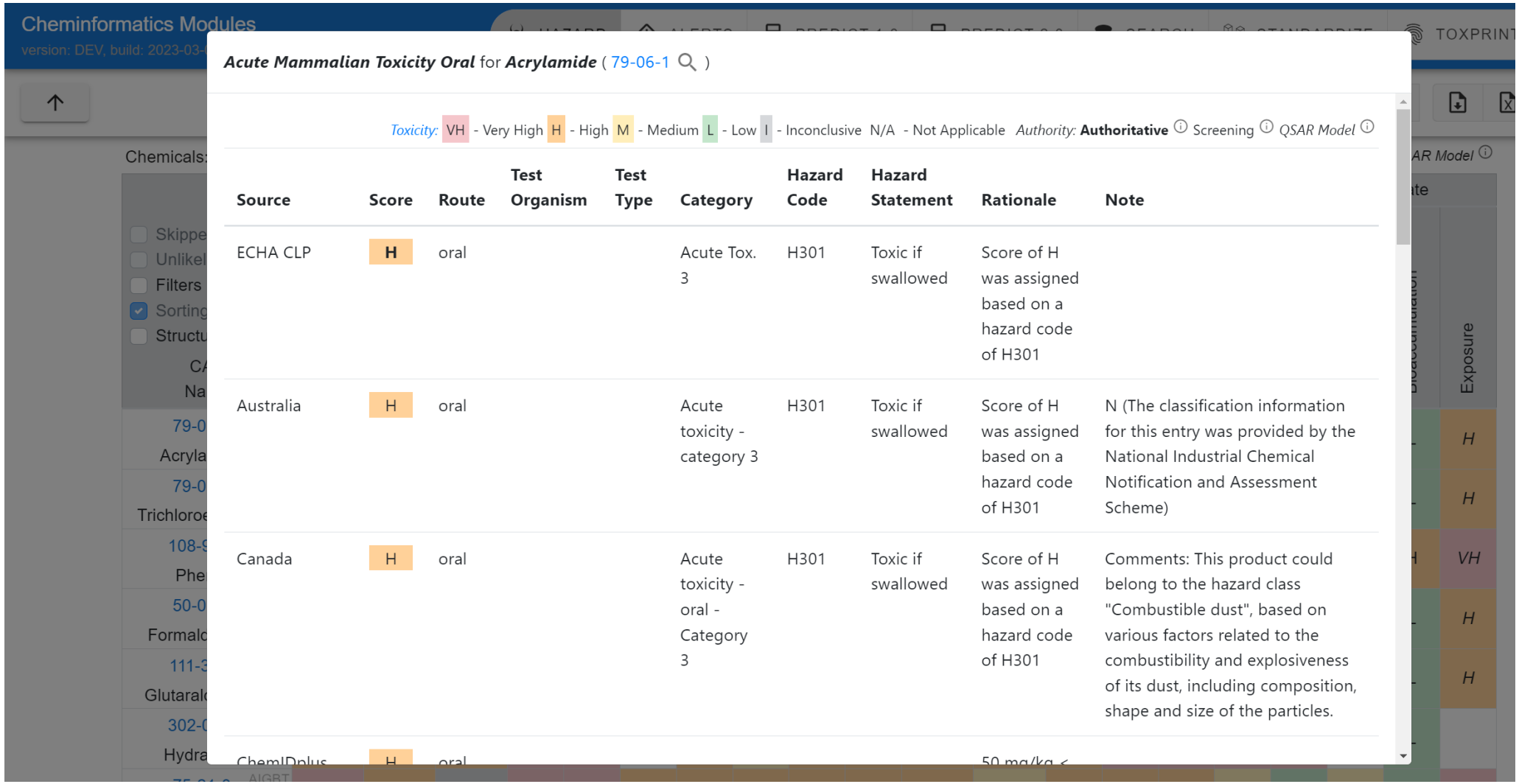
- Skipped (0)
- Unlikely (0)
- Filters (0)
- Sorting (0)
- Structure

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

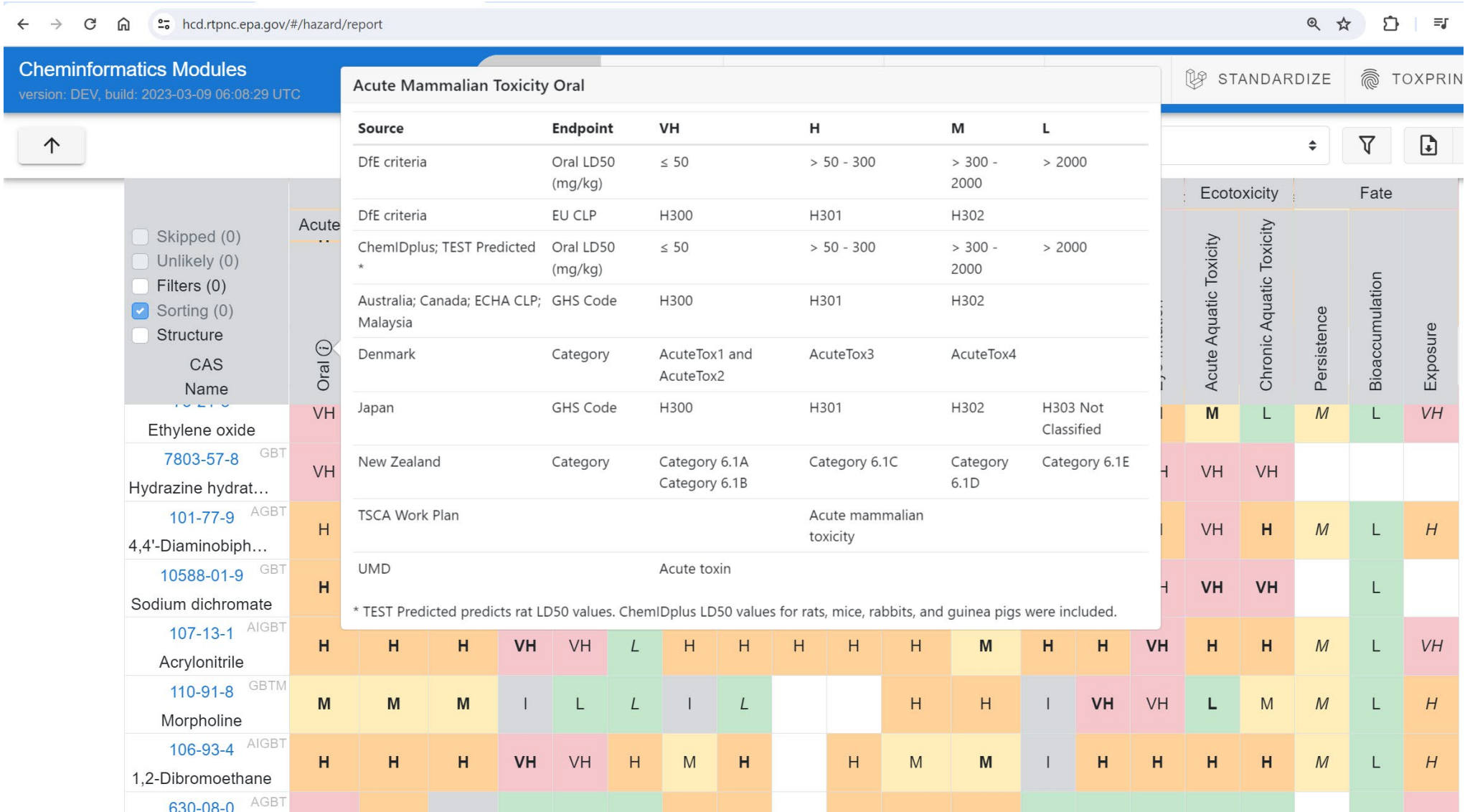
**Acrylamide | Acute Mammalian Toxicity Oral**

| | |
|---|---|
| **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** | |

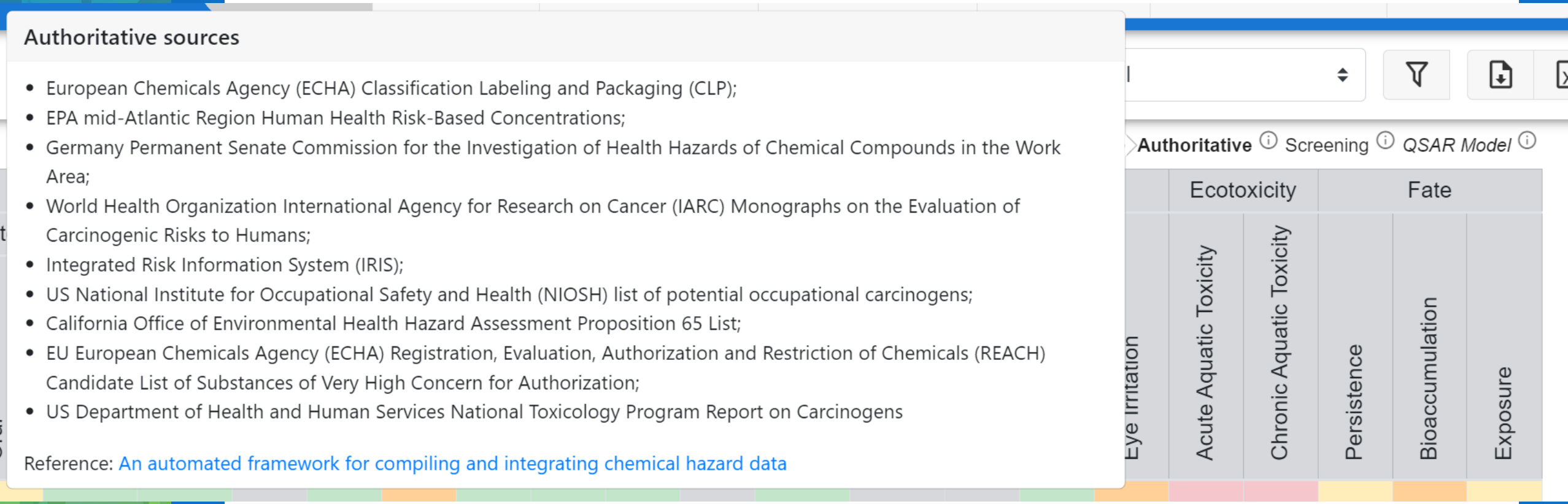
# All records for a given score on click



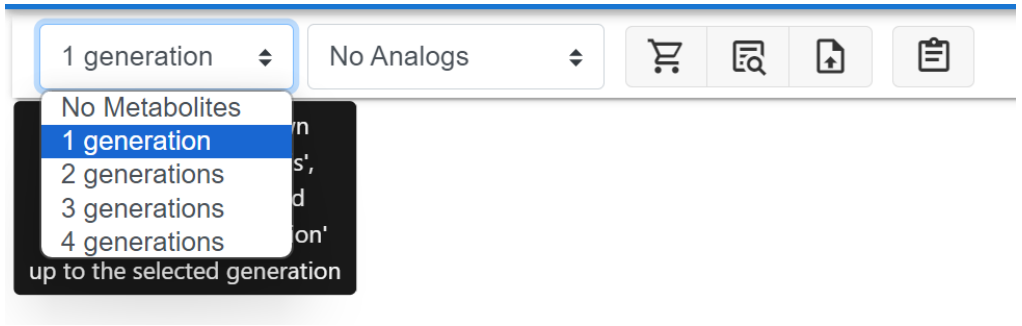
# Scoring Dictionaries available in headers



# List of sources by type from authority info buttons



# CTS Metabolite Feature




Chemicals: 3

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

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

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

# 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*

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

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

# Custom report feature