Smart Chemistry

Chemical Classification & Clustering Algorithms
Conflict of Interest

Portions of this work were performed by the presenter in support of UL’s Cheminformatics Tool Kit.

I also work with CAAT and ToxTrack Inc
Data Source

Similarity

Graph Algorithms

Fingerprinter

0 1 0 0 1 1 0

Metric

\[ M(0 \ 1 0 \ldots, \ 0 0 \ 1 0 \ldots) = -\infty \leq x \leq \infty \]

\[ = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \]

Tanimoto - Heuristic
“Probabilistic hazard assessment for skin sensitization...” Luechtefeld, Maertens, McIm, Hartung, Kleensang, Sá-Rocha. *Journal of Applied Toxicology* 2015

Unlabeled EINECS ----- Labeled ANNEX

- 1387  ANNEX SMILES
- 33383 EINECS SMILES
Learning Similarity

Similarity is a respective measure
Classification

Hazard - Not Hazard
1000 Skin Sensitization Tests / Year
1000 Chemicals x 1000 Tests / Year

1000 Points VS 1,000,000 Points
Curate

Collect

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Endpoint</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 0 ...</td>
<td>Acute Oral</td>
<td>-1</td>
</tr>
<tr>
<td>1 0 1 ...</td>
<td>Acute Oral</td>
<td>1</td>
</tr>
</tbody>
</table>

Label

Featurize

Model

https://medium.com/@hacertilbec/
<table>
<thead>
<tr>
<th>Call</th>
<th>Prev</th>
<th>SE</th>
<th>SP</th>
<th>ACC</th>
<th>BAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclude</td>
<td>9%</td>
<td>56%</td>
<td>54%</td>
<td>56%</td>
<td>55%</td>
</tr>
<tr>
<td>Weak</td>
<td>14%</td>
<td>83%</td>
<td>80%</td>
<td>82%</td>
<td>81%</td>
</tr>
<tr>
<td>Moderate</td>
<td>20%</td>
<td>91%</td>
<td>94%</td>
<td>92%</td>
<td>93%</td>
</tr>
<tr>
<td>Strong</td>
<td>55%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
</tr>
</tbody>
</table>

DoA Evaluation on ULCT-Test-2.0.0
Cheminformatics

What is it good for?
**Non-sensitizer cyanate?**

Pubchem \( \ce{O=C=NC1=CC=C(Cl)C=C1} \)

**Sensitization sets**

100 compounds with \( \ce{N=C=O} \)
95 sensitizers
5 non-sensitizers

**HSDB**

> airway irritation, cough, dyspnea, asthma, sensitization, and reduced pulmonary...

---

**Skatole - Non acute oral hazard?**

Pubchem \( \ce{C=1C=CC2=C(C1)NC=C2C} \)

**Acute toxicity.** The acute oral \( \text{LD}_{50} \) value in rats was reported as \( 3.45 \pm 0.372 \text{ g/kg} \) and the acute dermal \( \text{LD}_{50} \) value in rabbits as \( < 5 \text{ g/kg} \) (McGee, 1974). The ip \( \text{LD}_{50} \) of skatole in mice

An oral dose of \( 0.3 \text{ g} \) skatole/kg given to goats produced diffuse pulmonary oedema which resulted in death, and oral or iv administration to cows caused pulmonary lesions. Death from pulmonary oedema and emphysema resulted in 3/3, 0/2 and 1/3 cows given 0.2 (oral), 0.1 (oral) and 0.06 g/kg...
P-anisidine
Non-Eye Irritant?

> Very irritating to skin, eyes, and mucous membranes... - HSDB

Bisphenol A diglycidyl ether (BADGE) non-Mutagen?

Mutagenic potential of bisphenol A diglycidyl ether (BADGE) and its hydrolysis-derived products in the Ames Salmonella assay.

Sueiro RA, Arajo M, Suárez S, Garrido MJ.

Author information
1 Laboratory of Microbiology, Institute of Food Analysis and Research (IIAA), University of Santiago, 15706 Santiago de Compostela, Spain. mprosaan@us.es

Abstract
The mutagenicity of bisphenol A diglycidyl ether (BADGE), its first and second hydrolysis products (the diol epoxide and bis-diol of BADGE, respectively) and the bis-chlorohydrin of BADGE were investigated using the Ames Salmonella assay with strains TA98, TA100, TA1535 and TA1537. The assays were performed in the absence and presence of various concentrations of rat liver S9 fraction. The results obtained confirm the mutagenic power of BADGE in strains TA100 and TA1535 and show a positive response to the diol epoxide of BADGE in these strains, although the latter compound was approximately 10 times less potent than the former. A lack of mutagenic activity of the bis-diol of BADGE and the chlorohydrin under study is also shown. These findings suggest that BADGE and, to a much lesser extent, the diol epoxide of BADGE may constitute a genotoxic hazard, but not the bis-diol or bis-chlorohydrin of BADGE.
Local Search - Green Chemistry

- **Tetrahydrofuran**
  - $H = 3.5$ Sim=1.0
  - Acute oral, Serious eye damage
  - Eye irritation, Resp. irritation
  - Carcinogen...

- **2-Methyltetrahydrofuran**
  - $H = 1.92$ Sim=0.762
  - Skin irritation, serious eye irritation
  - May cause respiratory irritation
  - May cause drowsiness or dizziness

- **Diethyl Ether**
  - $H = 0.75$ Sim=0.71
  - May cause drowsiness or dizziness

- **Butyl Acetate**
  - $H = 0.65$ Sim =0.61
  - May cause drowsiness or dizziness
5000 virtual solvents

- Improved Hazard Estimation
- Improved Utility Estimation
- Virtual Compound Generation
Integrated Testing

Value of Information

![Box plot graph showing balanced accuracy for different numbers of tests. The graph includes two strategies: conditional and static. The graph represents the distribution of accuracy with error bars for each strategy across various numbers of tests.]
Active Learning

PubChem

BIOASSAY RECORD
Luminescence-based cell-based high throughput dose response assay for agonists of heterodimerization of the mu 1 (OPRM1) and delta 1 (OPRD1) opioid receptors

All (229) Active (126) Inactive (103)

model.predict("OPRD1")
> "It's an agonist!"

Exploit

Explore

Integrated Testing
Use Cases

**Quality Control**

Non-sensitizer cyanate?

Sensitization sets
- 100 compounds with N=C=O
- 95 sensitizers
- 5 non-sensitizers

HSDB
- airway irritation, cough, dyspnea, asthma, sensitization, and reduced pulmonary...

**Skatole - Non acute oral hazard?**

<table>
<thead>
<tr>
<th>Organism</th>
<th>Test Type</th>
<th>Route</th>
<th>Dose</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>rat</td>
<td>LD50</td>
<td>oral</td>
<td>3450 mg/kg</td>
<td>Food and Cosmetics Toxicology, 14(863), 1976</td>
</tr>
</tbody>
</table>

Acute toxicity: The acute oral LD$_{50}$ value in rats was reported as 3.45 ± 0.372 g/kg and the acute dermal LD$_{50}$ value in rabbits as <5 g/kg (McGee, 1974). The ip LD$_{50}$ of skatole in mice is...

An oral dose of 0.3 g skatole/kg given to goats produced diffuse pulmonary oedema which resulted in death, and oral or iv administration to cows caused pulmonary lesions. Death from pulmonary oedema and emphysema resulted in 3/3, 0/2 and 1/3 cows given 0.2 (oral), 0.1 (oral) and 0.06 g/kg...

**Green Chemistry - Tetrahydrofuran**

Tetrahydrofuran
- H = 3.5 Sim = 1.0
- acute oral, serious eye damage
- Eye irritation, resp. irritation
- Carcinogen...

2-Methylytetrahydrofuran
- H = 1.92 Sim = 0.762
- skin irritation, serious eye irritation
- May cause respiratory irritation
- May cause drowsiness or dizziness

Diethyl Ether
- H = 0.75 Sim = 0.71
- May cause drowsiness or dizziness

Butyl Acetate
- H = 0.65 Sim = 0.61
- May cause drowsiness or dizziness

**Active Learning**

Integrated Testing

Value of Information

Balanced Accuracy

Testing Strategy
- conditional
- static

Number Of Tests

Exploit

Explore

Integrated Testing

model.predict("OPROI")

"It's an agonist!"

PubChem
Thank You

Craig Rowlands
Sr Toxicologist

Stacie Abraham
Sr Regulatory Specialist

Thomas Hartung
Professor
Center for Alternatives to Animal Testing

Emily Grant
Project Manager

Adam Sawyer
Director, Innovation

Johns Hopkins - CAAT

Toxtrack

Contact me at tom@toxtrack.com