Assessing acute toxicity by combining QSAR predictions in a quantitative weight-of-evidence approach

James Rathman1,2, Aleksandra Mostag-Szlichtyng1, Tomasz Magdziarz2, Bruno Bienfait3, Oliver Sacher3, Chihae Yang1,3

INTRODUCTION

Although acute oral toxicity testing on rats remains a regulatory requirement for many chemicals, there is growing interest in developing computational methods for predicting acute toxicity in order to reduce animal testing and enable more efficient compound screening and prioritization, and also as a screening tool to determine dose ranges for in vivo studies.

We present a workflow for evaluating acute toxicity potential in which evidence from multiple QSAR models is rigorously and quantitatively combined to arrive at a weight-of-evidence prediction with associated estimation of uncertainty.

This presentation focuses on acute toxicity related to food additives and cosmetics, and compliments our previously developed acute toxicity models for pesticides and drugs.

Objectives:
1) Establish QSAR methodology for acute classifications
2) Develop knowledgebase approach

TRAINING AND TEST SETS

Data sources
- SCCS/SCCP/SCCFNP (EU Scientific Committee): cosmetics and consumer products
- US FDA CFSAN PAFA (Center for Food Safety and Applied Nutrition): Priority based Assessment of Food Additives: food additives (direct and indirect) and colorants (food and cosmetics).
- ChemIDPlus toxicity data: LD50 values from ChemIDPlus (from RTECS, the Registry of Toxic Effects of Chemical Substances). Compound structures similar to cosmetics and food additives and with LD50 data were queried from ChemIDPlus.
- Data were also augmented by information from EFSA Journal, JECFA/WHO publications, and open literature.

LD50 and Classification

GHS (Global Harmonization System) scale:

<table>
<thead>
<tr>
<th>Category</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LD50 (mg/kg/bw)</td>
<td>&lt;3</td>
<td>3-30</td>
<td>30-300</td>
<td>300-3000</td>
<td>&gt;3000</td>
</tr>
</tbody>
</table>

Analysis based on oral Rat LD50 data:

- When multiple studies are available, the values used in regulatory setting were taken.
- LD50 values span 6 orders of magnitude (0.05 – 74,000 mg/kg bw).
- Total of 1034 structures in training set.

For regulator purposes, a binary classification is desirable. Binary classifications were based on the binning scheme:

- LD50 ≥ 500 mg/kg bw: positive (1)
- LD50 < 500 mg/kg bw: negative (0)

A rigorous decision-theory approach based on Dempster-Shafer theory (DST) has been developed to accomplish two key tasks essential for probabilistic modeling: 1) uncertainty estimation and 2) combination of multiple sources of evidence.

DEMPSTER-SHAFER THEORY: ESTIMATING UNCERTAINTY AND QUANTITATIVE COMBINATION OF EVIDENCE

A combination of multiple sources of evidence
- Evidence is formed based on the combination of multiple QSAR models, structural alerts, in vivo and in vitro assay data, and expert opinions...

Test Set

A test set of 138 compounds with available oral rat LD50 data was prepared for model validation. 28% of these compounds are positive (LD50 ≤ 500 mg/kg bw).

QSAR model building and cross validations were performed exclusively on the training set.

Predictions were obtained for 112 compounds in the test set; remaining 26 were out of domain.

Training and Test Set Comparisons

Relative Frequencies of Selected ToxPrint Chemotypes

<table>
<thead>
<tr>
<th>Chemotype</th>
<th>Training Set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>colorants/hair dyes</td>
<td>28%</td>
<td>36%</td>
</tr>
<tr>
<td>aromatic amines</td>
<td>139</td>
<td>39%</td>
</tr>
<tr>
<td>k nearest neighbors (KNN)</td>
<td>1034</td>
<td>20%</td>
</tr>
</tbody>
</table>

Training set: positive prediction value (probability that a POS prediction is correct)

AUC: binary classification model optimization metric (area under ROC curve)

PPV: positive prediction value (probability that a POS prediction is correct)

NPV: negative prediction value (probability that a NEG prediction is correct)

Performance on Test Set

The global, KNN, and appropriate local QSAR models are first applied to each compound in the test set and then the results combined using DST to quantitatively weight of evidence. The combination results are better than for any of the models applied individually.

DISTRIBUTIONS AND CORRELATIONS OF SELECTED PHYSICOCHEMICAL PROPERTIES

NEXT STEPS

- Develop and evaluate multilevel ordinal classification model based on LD50 GHS classification scheme
- Develop and evaluate model for continuous-valued LD50 data
- Improve models by including chemotype alerts
- Explore possibility of improving models by including in vitro and/or in vivo assay data

Software: 1. CORINA Symphony descriptors: https://www.mn-am.com/products/corinisymphony

Contact: james.rathman@mn-am.com / rathman.10@osu.edu