NTP Workshop on Predictive Models for Acute Oral Systemic Toxicity

NTP Collaboration on Modeling Acute Systemic Oral Toxicity: Results and Ideas


aUniversity of North Carolina, Chapel Hill, USA
bFederal University of Goiás, Goiânia, Brazil
cOdessa National Polytechnic University, Ukraine
HOW TO SUCCEED IN COLLABORATION?

- Understanding of the problem;
- Use of additional data;
- Data curation;
- Rigorous External Validation;
- AD vs. 100% coverage;
- Consensus modeling;
- Experimental validation.
DATA CURATION

ORIGINAL DATASET

1. Chemical curation
2. Duplicate analysis
3. Analysis of intra- and inter-laboratorial variability
4. Exclusion of unreliable data sources
5. Detection and verification of activity cliffs
6. Calculation and tuning of modelability index
7. Consensus QSAR predictions to curate mislabeled compounds

INITIAL COMPOUND LIST

1. Removal of mixtures and inorganics
2. Structural conversion
   Cleaning/removal of salts
3. Normalization of specific chemotypes
4. Removal of duplicates
5. Manual inspection

Curated Dataset

<table>
<thead>
<tr>
<th>Original set</th>
<th>VT</th>
<th>NT</th>
<th>GHS</th>
<th>EPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>8,994</td>
<td>8,508</td>
<td>8,508</td>
<td>8,495</td>
<td>8,408</td>
</tr>
</tbody>
</table>

GENERAL WORFLOW

Original Dataset → Y-randomization → Curated Dataset → Curation

Descriptor calculation

External validation set

External validation with applicability domain

5-fold

Repeated for each set of descriptors and each fold.

Modeling set

Modeling using RF or MuDRA

Prediction of Prediction Set

Analysis of chemical space/applicability domain

Multi-Descriptor Read Across (MuDRA): a simple and transparent approach for developing accurate QSAR models

Vinicius M. Alves\textsuperscript{a,b}, Alexander Golbraikh\textsuperscript{a}, Stephen J. Capuzzi\textsuperscript{a}, Kammy Liu\textsuperscript{c}, Wai In Lam\textsuperscript{c}, Daniel Robert Korn\textsuperscript{c}, Diane Pozefsky\textsuperscript{c}, Carolina Horta Andrade\textsuperscript{b}, Eugene N. Muratov\textsuperscript{a,d,*}, Alexander Tropsha\textsuperscript{a,*}
$$S_{i,B}^j = 1 - d_{jac} = \frac{\sum_{j=1}^{p_j} x_i^j x_i^j}{\sum_{j=1}^{p_j} (x_i^j)^2 + \sum_{j=1}^{p_j} (x_i^j)^2 - \sum_{j=1}^{p_j} x_i^j x_i^j}$$

$$A_{i}^{pred,MCRA} = \frac{\sum_{j=1}^{D} \sum_{B_j=1}^{n_j} S_{i,B}^j A_{i,B_j}^j}{\sum_{j=1}^{D} \sum_{B_j=1}^{n_j} S_{i,B_j}^j}$$
## MuDRA vs. CERAPP MODELS

<table>
<thead>
<tr>
<th></th>
<th>CCR</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Agonist (n=6,319)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>CCR</td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>CERAPP (n = 5)</td>
<td>0.73 (± 0.05)</td>
<td>0.51 (± 0.13)</td>
<td>0.95 (± 0.05)</td>
</tr>
<tr>
<td>MuDRA</td>
<td>0.74</td>
<td>0.65</td>
<td>0.83</td>
</tr>
<tr>
<td><strong>Antagonist (n=6,532)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>CCR</td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>CERAPP (n = 4)</td>
<td>0.53 (± 0.02)</td>
<td>0.11 (± 0.09)</td>
<td>0.95 (± 0.05)</td>
</tr>
<tr>
<td>MuDRA</td>
<td>0.52</td>
<td>0.05</td>
<td>0.99</td>
</tr>
<tr>
<td><strong>Binding (n=7,283)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>CCR</td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>CERAPP (n = 9)</td>
<td>0.57 (± 0.02)</td>
<td>0.27 (± 0.11)</td>
<td>0.85 (± 0.08)</td>
</tr>
<tr>
<td>MuDRA</td>
<td>0.58</td>
<td>0.35</td>
<td>0.81</td>
</tr>
</tbody>
</table>

CERAPP compounds are provided by Dr. Mansourii.
### NON TOXIC MODELS

The Statistics is provided to participants by Dr. Mansouri.

<table>
<thead>
<tr>
<th></th>
<th>UNC (MuDRA)</th>
<th>FUG</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptors</strong></td>
<td>SiRMS, DRAGON, Morgan, RDKit</td>
<td>MACCS</td>
</tr>
<tr>
<td><strong>Algorithm</strong></td>
<td>MuDRA</td>
<td>RANDOM FOREST</td>
</tr>
<tr>
<td><strong>No. of compounds in training set</strong></td>
<td>4,834 toxic</td>
<td>2,298 toxic</td>
</tr>
<tr>
<td></td>
<td>3,661 not very toxic</td>
<td>2,298 not very toxic</td>
</tr>
</tbody>
</table>

![Graph showing TST metrics for UNC (MuDRA) and FUG]
Very Toxic Models

The statistics is provided to participants by Dr. Mansouri.
The Statistics is provided to participants by Dr. Mansourii.
## GHS Models

### UNC (QSAR) vs. FUG

<table>
<thead>
<tr>
<th></th>
<th>UNC (QSAR)</th>
<th>FUG</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptors</strong></td>
<td>SiRMS, DRAGON, Morgan, RDKit</td>
<td>MACCS</td>
</tr>
<tr>
<td><strong>Algorithm</strong></td>
<td>MuDRA</td>
<td>RANDOM FOREST</td>
</tr>
<tr>
<td><strong>No. of compounds in</strong></td>
<td>458 extreme (cat. I)</td>
<td>512 extreme/strong (cat. I and II)</td>
</tr>
<tr>
<td>training set</td>
<td>334 strong (cat. II)</td>
<td>512 moderate (cat. III)</td>
</tr>
<tr>
<td></td>
<td>70 moderate (cat. III)</td>
<td>512 non-toxic (cat. IV)</td>
</tr>
<tr>
<td></td>
<td>3,297 non-toxic (cat. IV)</td>
<td></td>
</tr>
</tbody>
</table>

The Statistics is provided to participants by Dr. Mansouri.
A Perspective and a New Integrated Computational Strategy for Skin Sensitization Assessment

Vinicius M. Alves, Stephen J. Capuzzi, Rodolpho C. Braga, Joyce V. B. Borba, Arthur C. Silva, Thomas Luechtefeld, Thomas Hartung, Carolina Horta Andrade, Eugene N. Muratov, and Alexander Tropsha
Correct identification and formulation of a problem is a must;

Use of additional data is extremely helpful;

Data curation and rigorous external validation is critical;

MuDRA is a simple, fast, and reliable approach that yields similar accuracy with complex modeling ensembles with 100% coverage of the prediction set;

We recommend use of AD for single models but 100% coverage for final consensus ensemble;

Comparison of the accuracy of the models must be made using the same compounds only;

Building smart consensus model is recommended – let the models help each other;

Only experimental validation could demonstrate predictivity and utility of a model.
ACKNOWLEDGMENT

- Organizers of NTP Collaboration;
- Organizers of the Workshop;
- Kamel Mansouri (NICEATM);
- Nicole Kleinstreuer (NICEATM);
- Alexey Zakharov (NCATS);
- Denis Fourches (NCSU).
Chembench: A Publicly Accessible, Integrated Cheminformatics Portal

Stephen J. Capuzzi,† Ian Sang-June Kim,† Wai In Lam,‡ Thomas E. Thornton,† Eugene N. Muratov,† Diane Pozefsky,*‡ and Alexander Tropsha*†‡

†Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, and ‡Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina 27599, United States

https://chembench.mml.unc.edu/mudra/
Chemical toxicity prediction for major classes of industrial chemicals: Is it possible to develop universal models covering cosmetics, drugs, and pesticides?

Vinicius M. Alves a, b, Eugene N. Muratov a, c, Alexey Zakharov d, Nail N. Muratov c, Carolina H. Andrade b, Alexander Tropsha a, *
Pred-Skin: A Fast and Reliable Web Application to Assess Skin Sensitization Effect of Chemicals

Rodolpho C. Braga, Vinicius M. Alves, Eugene N. Muratov, Judy Strickland, Nicole Kleinstreuer, Alexander Trospsha, and Carolina Horta Andrade

www.labmol.com.br/predskin