Collaborative Acute Toxicity Modeling Suite (CATMoS)

Kamel Mansouri
LEAD COMPUTATIONAL CHEMIST, ILS IN SUPPORT OF NICEATM

SACATM – September 19-20, 2019

Disclaimer: ILS staff provide technical support for NICEATM, but do not represent NIEHS, NTP, or the official positions of any federal agency. (the author declares no conflict of interest)
Overview

• Project scope: acute oral toxicity
  – Regulatory use of these data
  – Endpoints selected for predictive modeling
  – Compiling inventory of rat acute oral LD50
  – Establishing training, evaluation, and prediction sets
  – Evaluation of submitted models

• International contributors

• Generation of consensus predictions

• Current status and public release
Toxicity prediction

Too many chemicals to test with standard animal-based methods
  – Cost, time, animal welfare

Alternative

- Organic pollutants with exposure potential accumulate in body tissues
  - Cause toxic effects to wild life and humans
- Existence of gaps in the experimental data for environmental endpoints
  - Need to fill the data gaps and bridge the lack of knowledge
- Regulatory requirements:
  - Reduce animal testing, time and costs
  - Methodology: use of QSAR/QSPR to predict the endpoints of interest.
ICCVAM Acute Toxicity Workgroup

- Identifies federal agency requirements, needs, and decision contexts for using acute systemic toxicity data
Agency-Based Modeling Endpoint Selection

Binary Models

- Highly toxic (≤50 mg/kg)
- Toxic (>50-5000 mg/kg)
- Nontoxic (>2000 mg/kg)

Continuous Model

- Point estimates of LD50 values

Categorical Models

- EPA Categories
  - I (≤ 50 mg/kg)
  - II (>50 ≤ 500 mg/kg)
  - III (>500 ≤ 5000 mg/kg)
  - IV (>5000 mg/kg)

- GHS Categories
  - I (≤ 5 mg/kg)
  - II (>5 ≤ 50 mg/kg)
  - III (>50 ≤ 300 mg/kg)
  - IV (>300 ≤ 2000 mg/kg)
  - NC (> 2000 mg/kg)

Hazard

Packing Group

- Group I (≤ 5 mg/kg)
- Group II (>5 ≤ 50 mg/kg)
- Group III (>50 ≤ 300 mg/kg)
- Group IV (>300 ≤ 2000 mg/kg)
- NC (≥ 2000 mg/kg)
Available data for modeling

Rat oral LD50s:
16,297 chemicals total
34,508 LD50 values

15,688 chemicals total
21,200 LD50 values

 QSAR-ready standardization
 Desalted, stereochemistry stripped,
tautomers and nitro groups standardized,
valence corrected, structures neutralized

11,992 chemicals with accurate structures

• Very toxic endpoint: 11,886 entries (binary, 0/1)
• Non-toxic endpoint: 11,871 entries (binary, 0/1)
• EPA endpoint: 11,755 entries (categorical, 4 categories)
• GHS endpoint: 11,845 entries (categorical, 5 categories)
• LD50 endpoint: 8,908 entries (continuous values)
**Aim of the workflow:**
- Combine different procedures and ideas
- Minimize the differences between the structures used for prediction
- Produce a flexible free and open source workflow to be shared

Wedelbye et al. Danish EPA Environmental Project No. 1503, 2013
Mansouri et al. (http://ehp.niehs.nih.gov/15-10267/)
Establishing Modeling Dataset

• **Training and evaluation sets:**

  • 11,992 chemicals from the final inventory of chemicals with QSAR-ready structures having rat oral acute toxicity data were split into training and test sets:
    • 75% training set: 8,994 chemicals
    • 25% evaluation set: 2,998 chemicals

  • All endpoints training data included in same structure file
  • Similar distributions and variability for values and categories
  • Similar distribution of chemical structures sources
Establishing Modeling Dataset

• Prediction set:

Included lists of regulatory interest:

• ToxCast/Tox21
• EDSP
• TSCA
• Substances on the market (EPA Dashboard list)

After QSAR-ready standardization:

48137 structures to be predicted (including the evaluation set)
ChemMaps landscape of CATMoS chemicals

http://www.chemmaps.com/chemmaps/DSSToxMap3D/
Consortium:

- **35 Participants/Groups** from around the globe representing academia, industry, and government contributed

(https://batchgeo.com/map/d06c5d497ed8f76ecfee500c2b0e1dfa)
Submitted Models

- Non-toxic: 33 models
- Very Toxic: 32 models
- GHS categories: 23 models
- EPA categories: 26 models
- LD50: 25 models

Total: 139 models
Evaluation procedure

**Qualitative evaluation:**
- Documentation
- Defined endpoint
- Unambiguous algorithm
- Availability of code
- Applicability domain definition
- Availability of data used for modeling
- Mechanistic interpretation

**Quantitative evaluation:**
- Goodness of fit: training statistics
- Evaluation set predictivity: statistics on the evaluation set
- Robustness: balance between (Goodness of fit) & (Test set predictivity)

\[ S = 0.3 \times (\text{Goodness of fit}) + 0.45 \times (\text{Test set predictivity}) + 0.25 \times (\text{Robustness}) \]

**Continuous models:**
- Goodness of fit = $R^2_T$
- Test set predictivity = $R^2_{Tst}$
- Robustness = $1 - |R^2_T - R^2_{Tst}|$

**Categorical models (binary and multi-class):**
\[
\text{Goodness of fit} = 0.7 \times (BA_{TR}) + 0.3 \times (1 - |Sn_{TR} - Sp_{TR}|)
\]
\[
\text{Test set predictivity} = 0.7 \times (BA_{Tst}) + 0.3 \times (1 - |Sn_{Tst} - Sp_{Tst}|)
\]
\[
\text{Robustness} = 1 - |BA_{TR} - BA_{Tst}|
\]
Coverage and concordance of the models
CATMoS consensus modeling

Steps of combining the single models into consensus

**Initial models & predictions**
- VT (32 models)
- NT (33 models)
- GHS (23 models)
- EPA (26 models)
- LD50 (25 models)

**Combining models**

**Step 1**
- Weighted average /majority rule

**Independent consensus models/predictions**
- VT
- NT
- GHS
- EPA
- LD50

**Step 2**
- Weight of Evidence approach (WoE)
  - Majority rule

**Consistent consensus models/predictions**
- VT
- NT
- GHS
- EPA
- LD50

A consensus model per endpoint (~20~30 models)

Consensus representing all ~140 models
WoE approach to combine the 5 endpoints

<table>
<thead>
<tr>
<th>molX</th>
<th>VT</th>
<th>NT</th>
<th>EPA</th>
<th>GHS</th>
<th>LD50</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Prediction</th>
<th>0</th>
<th>5</th>
<th>50</th>
<th>300</th>
<th>500</th>
<th>2000</th>
<th>5000</th>
<th>mg/kg</th>
</tr>
</thead>
<tbody>
<tr>
<td>VT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LD50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>316</td>
</tr>
</tbody>
</table>
WoE approach to combine the 5 endpoints

Variability range (log units) for LD50
WoE approach to combine the 5 endpoints

<table>
<thead>
<tr>
<th>Model Prediction</th>
<th>VT</th>
<th>NT</th>
<th>EPA</th>
<th>GHS</th>
<th>LD50</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>molX</th>
<th>VT</th>
<th>NT</th>
<th>EPA</th>
<th>GHS</th>
<th>LD50</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction</th>
<th>0</th>
<th>5</th>
<th>50</th>
<th>300</th>
<th>500</th>
<th>2000</th>
<th>5000</th>
<th>mg/kg</th>
</tr>
</thead>
<tbody>
<tr>
<td>VT</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NT</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EPA</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>GHS</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LD50</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WoE</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
WoE approach to combine the 5 endpoints

<table>
<thead>
<tr>
<th>VT</th>
<th>VT</th>
<th>EPA</th>
<th>GHS</th>
<th>LD50</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2.5</td>
</tr>
</tbody>
</table>

**Model Prediction**

<table>
<thead>
<tr>
<th>0</th>
<th>5</th>
<th>50</th>
<th>300</th>
<th>500</th>
<th>2000</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>VT</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NT</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>EPA</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GHS</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LD50?</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**WoE: consistent calls**

<table>
<thead>
<tr>
<th>VT</th>
<th>VT</th>
<th>EPA</th>
<th>GHS</th>
<th>LD50</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2.36</td>
</tr>
</tbody>
</table>

**Model**

Prediction: VT NT EPA GHS LD50

**molX**

Model Prediction

| molX | 0 | 0 | 2 | 3 | 2.5 |

**Winning bin**

<table>
<thead>
<tr>
<th>0</th>
<th>5</th>
<th>50</th>
<th>300</th>
<th>500</th>
<th>2000</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>VT</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NT</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>EPA</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GHS</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>LD50?</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**mg/kg**

How to adjust quantitative LD50? Avg of Lower CI and upper bin threshold

(160+300)/2 = 230mg/kg
## Consensus Model Statistics

<table>
<thead>
<tr>
<th></th>
<th>VT Train</th>
<th>VT Eval</th>
<th>NT Train</th>
<th>NT Eval</th>
<th>EPA Train</th>
<th>EPA Eval</th>
<th>GHS Train</th>
<th>GHS Eval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>0.87</td>
<td>0.70</td>
<td>0.88</td>
<td>0.67</td>
<td>0.81</td>
<td>0.62</td>
<td>0.80</td>
<td>0.58</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.99</td>
<td>0.97</td>
<td>0.97</td>
<td>0.90</td>
<td>0.92</td>
<td>0.86</td>
<td>0.95</td>
<td>0.90</td>
</tr>
<tr>
<td>Balanced Accuracy</td>
<td>0.93</td>
<td>0.84</td>
<td>0.92</td>
<td>0.78</td>
<td>0.87</td>
<td>0.74</td>
<td>0.88</td>
<td>0.74</td>
</tr>
<tr>
<td><strong>In vivo Balanced Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.81</td>
<td></td>
<td>0.89</td>
<td></td>
<td>0.82</td>
<td></td>
<td>0.79</td>
<td></td>
</tr>
</tbody>
</table>

The consensus predictions perform just as well as replicate *in vivo* data do at predicting oral acute toxicity outcome.
Extended CATMoS predictions

Weighted read-across

\[ d_1 = 0 \]
\[ \text{Pred}_i = N_i \]

\[ d_1 \neq 0 \]
\[ w_i = f(d_i) \]
\[ \text{Pred}_i = f(w_i, N_i) \]

- New chemical to be predicted
- Nearest neighbors \((N_i)\)

\(d_i\): Euclidean distance based on the selected descriptors for each endpoint

Automated, similarity-endpoint dependent read-across: weighted kNN
Generation of Consensus Predictions

- Models passing qualitative evaluation (requirement for transparency; description of approach was sufficient)

- Integrating only *in-domain* predictions across chemicals in the prediction set (48,137 chemicals) for each model, respectively
  - Categorical models: weighted majority rule
  - Continuous model: weighted average
Collaboration with ATWG partners and ICCVAM agencies

Evaluate and optimize CATMoS predictions based on lists of interest
Running CATMoS Consensus models

OPERA Standalone application

Command line

- Free, opensource & open-data
- Single chemical and batch mode
- Multiple platforms (Windows and Linux)
- Embeddable libraries (java, C, C++, Python)

Graphical user interface

https://github.com/NIEHS/OPERA

OPERA 1.5

Physchem & Environmental fate:

<table>
<thead>
<tr>
<th>Model</th>
<th>Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOH</td>
<td>Atmospheric Hydroxylation Rate</td>
</tr>
<tr>
<td>BCF</td>
<td>Bioconcentration Factor</td>
</tr>
<tr>
<td>BioHL</td>
<td>Biodegradation Half-life</td>
</tr>
<tr>
<td>RB</td>
<td>Ready Biodegradability</td>
</tr>
<tr>
<td>BP</td>
<td>Boiling Point</td>
</tr>
<tr>
<td>HL</td>
<td>Henry’s Law Constant</td>
</tr>
<tr>
<td>KM</td>
<td>Fish Biotransformation Half-life</td>
</tr>
<tr>
<td>KOA</td>
<td>Octanol/Air Partition Coefficient</td>
</tr>
<tr>
<td>LogP</td>
<td>Octanol-water Partition Coefficient</td>
</tr>
<tr>
<td>MP</td>
<td>Melting Point</td>
</tr>
<tr>
<td>KOC</td>
<td>Soil Adsorption Coefficient</td>
</tr>
<tr>
<td>VP</td>
<td>Vapor Pressure</td>
</tr>
<tr>
<td>WS</td>
<td>Water solubility</td>
</tr>
<tr>
<td>RT</td>
<td>HPLC retention time</td>
</tr>
</tbody>
</table>

**New in OPERA2:**

- **Physchem properties:**
  - General structural properties
  - pKa
  - Log D

- **ADME properties**
  - Plasma fraction unbound (FuB)
  - Intrinsic clearance (Clint)

- **Toxicity endpoints**
  - ER activity (CERAPP)
    https://ehp.niehs.nih.gov/15-10267/
  - AR activity (CoMPARA)
    https://doi.org/10.13140/RG.2.2.19612.80009
  - Acute toxicity (CATMoS)
    https://doi.org/10.1016/j.comtox.2018.08.002
CATMoS prediction examples

1,4-Dioxane
123-91-1 | DTXSID4020533
Molecular Formula: C_4H_8O_2
Average Mass: 88.106 g/mol
LD50: 4200 mg/kg
log10 LD50 = 3.62

Vitamin D3
67-97-0 | DTXSID6026294
Molecular Formula: C_27H_44O
Average Mass: 384.648 g/mol
LD50: 42 mg/kg
log10 LD50 = 1.62

https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID4020533
https://comptox.epa.gov/dashboard/dsstoxdb/results?search=DTXSID6026294

CATMoS predictions:

<table>
<thead>
<tr>
<th>MoleculeID</th>
<th>CATMoS_VT_pred</th>
<th>CATMoS_NT_pred</th>
<th>CATMoS_EPA_pred</th>
<th>CATMoS_GHS_pred</th>
<th>CATMoS_LD50_pred</th>
<th>AD_CATMoS</th>
<th>AD_index_CATMoS</th>
<th>Conf_index_CATMoS</th>
</tr>
</thead>
<tbody>
<tr>
<td>'123-91-1'</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>3.4053</td>
<td>1</td>
<td>1</td>
<td>0.9500</td>
</tr>
<tr>
<td>'67-97-0'</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1.2845</td>
<td>1</td>
<td>1</td>
<td>0.8694</td>
</tr>
</tbody>
</table>
Soon on NTP/ICE and EPA CompTox dashboard

https://ntp.niehs.nih.gov/

https://comptox.epa.gov/dashboard
The “3C” Concept at Work!

- Success of the project was due in great part to the use of the 3C concept as well as up-front and continuous engagement of regulators in the process.

[Image: A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States]

https://ntp.niehs.nih.gov/go/natl-strategy
Acknowledgements

THANK YOU!

• ICCVAM Acute Toxicity Workgroup

• EPA/NCCT
  – Grace Patlewicz
  – Jeremy Fitzpatrick

• ILS/NICEATM
  – Agnes Karmaus
  – Dave Allen
  – Shannon Bell
  – Patricia Ceger
  – Judy Strickland
  – Amber Daniel

• NTP/NICEATM
  – Nicole Kleinstreuer
  – Warren Casey

Feedback welcome: Kamel Mansouri (kmansouri@ils-inc.com)

Technical support was provided by ILS under NIEHS contract HHSN273201500010C.