Automated Read-Across for REACH

Dr. Anne Bonhoff
April 25, 2017
UL Collaboration to Develop Predictive Toxicity Tool for Read-across

UL Collaborates with Prof. Dr. Thomas Hartung, Chair for Evidence-based Toxicology at Johns Hopkins Bloomberg School of Public Health, Baltimore, and University of Konstanz, Germany; also Director of their Centers for Alternatives to Animal Testing (CAAT)

and Thomas Luechtefeld, a PhD student,

and UL’s team of scientists and toxicologists.
REACH Data-rich substances registered 2010 and 2013:

- 75% of dossiers use read-across
- Other alternatives hardly used
- Expertise in industry low
- Low acceptance by EChA

Read-across Data gap filling concluding from (structurally) similar chemicals

Category approach Test only representatives of a group of similar chemicals or complex mixtures
The opportunity of Big Data & Bioinformatics
Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence - first machine learning, then deep learning, a subset of machine learning - have created ever larger disruptions.
Big biological data in toxicology

Tab. 3: Public databases of toxicity data

Unified and curated data-base
A marriage of technologies

Read-across
• Support weight of evidence
• Circumstantial
• Manual
• Unclear acceptability

(Q)SAR
• Data-mining by computer
• Broader applicability
• Can be validated with enormous consequences for acceptability

Automated Read-Across = “REACH-across”
• Mines local “similarity space”
• Comprehensive use of available data
• Expresses certainty
• Validation on the way
Welcome to REACHACross™ software a reliable digital assistant for REACH compliance.

Offering the best of both worlds, REACHACross™ software combines the best of an objective computational approach of a QSAR with the proven method of read-across systems. Generate REACH dossier compliant with just a few minutes.

Need assistance?
Contact us for technical support.

Chemical: (SMILES or CAS Registry Number. REACHACross™ 1.0.0 does not currently use European Inventory of Existing Commercial chemical Substances (EINECS) numbers to identify chemical structures.)

O=C(OC1=CC=CC=2OC(OC12)(C)C)NC

Endpoint Selection:
- Document
- Acute Dermal Irritation
- Acute Dermal Toxicity
- Acute Eye Irritation
- Acute Oral Toxicity
- Mutagenicity
- Skin Sensitization
REACHACROSS™ SOFTWARE PIPELINE

Data Source + Similarity \[ \similarityfunction \] = Graph Algorithms

Fingerprinter

\[ \text{Metric} \left( 0 \ 1 \ 0 \ 0 \ldots, \ 0 \ 0 \ 1 \ 0 \ldots \right) = -\infty \leq x \leq \infty \]
SIMPLE MODELS

Table 9. Table of accuracy metrics for KNN variant given different min-similarity parameters. For a given chemical, neighbors used for prediction are constrained to those with similarity $\geq$ the Min Similarity. Predictions are made for all chemicals with 1 or more neighbors.

<table>
<thead>
<tr>
<th>Min Similarity</th>
<th>Chemicals</th>
<th>TP</th>
<th>TN</th>
<th>FN</th>
<th>FP</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>BAC</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>525</td>
<td>429</td>
<td></td>
<td></td>
<td></td>
<td>0.68</td>
<td>0.96</td>
<td>0.82</td>
<td>0.92</td>
</tr>
<tr>
<td>0.9</td>
<td>1189</td>
<td>870</td>
<td></td>
<td></td>
<td></td>
<td>0.61</td>
<td>0.91</td>
<td>0.76</td>
<td>0.85</td>
</tr>
<tr>
<td>0.85</td>
<td>1738</td>
<td>1238</td>
<td>130</td>
<td>150</td>
<td></td>
<td>0.59</td>
<td>0.90</td>
<td>0.75</td>
<td>0.84</td>
</tr>
<tr>
<td>0.75</td>
<td>2288</td>
<td>224</td>
<td>1616</td>
<td>170</td>
<td>278</td>
<td>0.45</td>
<td>0.90</td>
<td>0.68</td>
<td>0.80</td>
</tr>
</tbody>
</table>

BAC = balanced accuracy, TP = true positives, TN = True negatives, FN = False Negatives, FP = False Positives
NETWORK EFFECT

Threshold
- 70% similar
- 80% similar
- 90% similar
INCREASING RELEVANCE

Unlabeled EINECS ----- Labeled ANNEX

- 1387 ANNEX SMILES
- 33383 EINECS SMILES
NETWORK FEATURES
Local similarity = prediction
Regional similarity = uncertainty
Positive = proximity to positives
Negative = proximity to negatives
Confidence = the more data the closer
Test performance characteristics are a choice: Test Sensitivity vs. Specificity

For REACH-across this is even more complex:
Two thresholds (proximity to positives and to negatives)
These choices influence coverage, i.e. for how many unknown chemicals a prediction can be made
Skin sensitization (Luechtefeld et al., 2016): Simple classification by nearest neighbor

<table>
<thead>
<tr>
<th>Min. Similarity</th>
<th>Chemicals</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>525</td>
<td>0.92</td>
</tr>
<tr>
<td>0.9</td>
<td>1189</td>
<td>0.85</td>
</tr>
<tr>
<td>0.85</td>
<td>1738</td>
<td>0.84</td>
</tr>
<tr>
<td>0.75</td>
<td>2288</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Accuracies better than different animal TG against each other, but coverage drops
### Status of the tool March 2017

<table>
<thead>
<tr>
<th>endpoint</th>
<th>tested</th>
<th>Se</th>
<th>Sp</th>
<th>labeled Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skin Sensitisation</td>
<td>5136</td>
<td>83%</td>
<td>55%</td>
<td>83%</td>
</tr>
<tr>
<td>Eye Irritation</td>
<td>15214</td>
<td>83%</td>
<td>54%</td>
<td>79%</td>
</tr>
<tr>
<td>Acute Oral</td>
<td>12342</td>
<td>82%</td>
<td>71%</td>
<td>77%</td>
</tr>
<tr>
<td>Mutagenicity</td>
<td>4077</td>
<td>80%</td>
<td>58%</td>
<td>81%</td>
</tr>
<tr>
<td>Skin Irritation / Corrosion</td>
<td>14718</td>
<td>88%</td>
<td>57%</td>
<td>64%</td>
</tr>
<tr>
<td>Acute Dermal</td>
<td>6732</td>
<td>89%</td>
<td>70%</td>
<td>59%</td>
</tr>
</tbody>
</table>

All endpoints for REACH 2018 <10 tons/a
Coming soon: acute fish toxicity

Accuracy similar to animal test

Internal validation for unprecedented number of chemicals

Works for most chemicals, to increase with more databases incorporated
OECD PRINCIPLES FOR THE VALIDATION, FOR REGULATORY PURPOSES, OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIP MODELS

These principles were agreed by OECD member countries at the 37th Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology in November 2004. The principles are intended to be read in conjunction with the associated explanatory notes which were also agreed at the 37th Joint Meeting.

To facilitate the consideration of a (Q)SAR model for regulatory purposes, it should be associated with the following information:

1) a defined endpoint
2) an unambiguous algorithm
3) a defined domain of applicability
4) appropriate measures of goodness-of-fit, robustness and predictivity
5) a mechanistic interpretation, if possible
A digital alternative to meet your REACH compliance needs:

UL REACHAcross™ software offers the best of both worlds: the objective computational approach of a QSAR combined with the proven robustness and acceptability of a read-across system.

ARE YOU LOOKING FOR A SMART, FAST APPROACH TO MEET YOUR REACH COMPLIANCE NEEDS?

This ground-breaking digital tool offers the best of both worlds: automated computational QSAR using one of the largest chemical toxicology databases available, combined with the reliability of read across. REACHAcross™ software works by building large networks of chemicals based on properties such as molecular structure and health endpoint interactions.
### REACHAcross™ Questions

**Chemical:** (SMILES or CAS Registry Number. REACHAcross™ 1.0.0 does not currently use European Inventory of Existing Commercial Chemical Substances (EINECS) numbers to identify chemical structures.)

<table>
<thead>
<tr>
<th>Smiles code for Methanol</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
</tr>
</tbody>
</table>

**Endpoint Selection:**

- [ ] Document
- [ ] Acute Dermal Irritation
- [ ] Acute Dermal Toxicity
- [ ] Acute Eye Irritation
- [ ] Acute Oral Toxicity
- [ ] Mutagenicity
- [ ] Skin Sensitization

[Select Endpoint]
Request Name: Webinar

Smiles code for Methanol: CO

Data Acceptance:

In case of any problem with this request we will communicate to the following email address. Please update this email address if you want us to use a different email:

sarah.partridge@ul.com

Accept

If all data entered is accurate, click 'Accept'

Home: To go back to My Requests, click 'Home'
# My Requests

<table>
<thead>
<tr>
<th>Request ID</th>
<th>Name</th>
<th>Creation Date</th>
<th>Endpoint Report Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1401702</td>
<td>Webinar</td>
<td>3/28/2017</td>
<td>ACOT</td>
</tr>
<tr>
<td>1401703</td>
<td>Methanol for Acute Oral Toxicity</td>
<td>3/28/2017</td>
<td>ACOT</td>
</tr>
</tbody>
</table>

Report Ready
REACHACROSS™ REPORT

REACHACROSS™ 1.0.0 estimates a 95% probability of Acute Oral Toxicity hazard for OC.

The below resources will aid in completing your IUCLID submission:
1. EC1A - How to use and report QSARs
2. REACHACROSS™ Documentation: http://ureachacross.com
3. REACHACROSS™ OMBE: http://ureachacross.com/Documents/reachacross-1.0.0-qmr.xml
4. REACHACROSS™ OPSE: http://ureachacross.com/Documents/reachacross-1.0.0-opse.txt

The below information is supplied to aid in completing an Acute Oral Toxicity submission in IUCLID:

ADMINISTRATIVE DATA

Type of information:
(Q)SAR

Reliability:
2 (reliable with restrictions)

Rationale for reliability:
Results derived from a valid (Q)SAR model and falling into its applicability domain, with adequate and reliable documentation / justification.

justification for type of information

Software:
http://ureachacross.com/

Model (incl. version number):
REACHACROSS™ v1.0.0

SMILES or other identifiers used as input for the model:
OC

Scientific validity of the (Q)SAR model:
- Defined endpoint: Acute Oral Toxicity
- Defined domain of applicability: REACHACROSS™ 1.0.0 defines a probabilistic domain of applicability. Substances predicted with sufficiently high or low probability are included in the domain of applicability.
- Appropriate measures of goodness-of-fit and robustness and predictivity: REACHACROSS™ 1.0.0 uses leave one out cross validation on the ECHA C&L database. These results are reported at http://ureachacross.com/Documents/reachacross-1.0.0-wp.pdf
- Mechanistic interpretation: N/A

Applicability domain:

Example Predictions

REACHACROSS 1.0.0 hazard estimates (a) for 12 REACH Annex compounds. Red bars show known hazards.

Figure 1: estimated hazard probability for 12 chemicals in annex 6 table 5.1
QUESTIONS?

ULREACHACROSS.COM
THANK YOU