Exploiting data on chemical substances to model environmental and (eco)toxicological properties

Emilio Benfenati
Istituto di Ricerche Farmacologiche Mario Negri IRCCS
Scientific and social aims

- Pro-active approach: plan safety
- Safer chemicals / identification of substitutes
- Integrated view: all hazardous properties
Methodological strategy

- Full data exploitation
- Increased reasoning on substances and their properties
- Much higher screening performance
The safety assessment, in the past

One substance, one assessment for each property, done experimentally
After tens of thousands substances registered, is the approach still the same?

Can we identify an heuristic process?

Can we take advantage of what we already know?
Technology: *in silico* models

(Q)SAR

Read-across
Technology: *in silico* models

*In the past*

- Individual models to predict properties of effects
Technology: *in silico* models

**Regulatory uses**

- **ECHA**
  - Projects from Italian Health and Environmental Ministries

- **EFSA**
  - Used by German UBA for prioritization and substance evaluation

---

**ECHA**

Preparation of an inventory of substances suspected to meet REACH Annex III criteria

Technical documentation

**EFSA**

Practical guide

How to use and report (Q)SARs

Version 3.1 – July 2016

---

**EFSA**

Guidance on the use of the weight of evidence approach in scientific assessments

Adopted: 12 July 2017
Growing architecture: Integrated system for prioritization, running about 50 models together, for CMR, PBT and ED
Technology: *in silico* models

- Integrated system for all REACH properties
- Integrated system for experimental data and predicted ones

Soon

LIFE
CONCERT
REACH
toDIVINE (UBA)
Optitox (EFSA)
Technology: *in silico* models

- Integrated system for hazard and exposure, to get automated risk assessment
  VERMEER
- System to suggest safer chemicals VERMEER

Soon
This can

- Speed up the planning of new chemicals
- Increase safety
- Reduce costs
Identification of risky substance
How?

Risk assessment

Exposure

Hazard
Software Predicting Human and Environmental Risk Assessment
A single platform for investigating multiple risk scenarios
JAVA stand-alone application

**Batch processing**
(large dataset and/or multiple endpoints)
includes a number of QSAR models
Includes an independent tool to assess reliability, through the **Applicability Domain Index**…

---

<table>
<thead>
<tr>
<th><strong>Global AD Index</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>AD index = 1</td>
</tr>
<tr>
<td>Explanation: the predicted compound is into the Applicability Domain of the model.</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th><strong>Similar molecules with known experimental value</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Similarity index = 1</td>
</tr>
<tr>
<td>Explanation: strongly similar compounds with known experimental value in the training set have been found.</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th><strong>Accuracy of prediction for similar molecules</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy index = 1</td>
</tr>
<tr>
<td>Explanation: accuracy of prediction for similar molecules found in the training set is good.</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th><strong>Concordance for similar molecules</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordance index = 1</td>
</tr>
<tr>
<td>Explanation: similar molecules found in the training set have experimental values that agree with the predicted value.</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th><strong>Atom Centered Fragments similarity check</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>ACF index = 1</td>
</tr>
<tr>
<td>Explanation: all atom centered fragment of the compound have been found in the compounds of the training set.</td>
</tr>
</tbody>
</table>
3.1 Applicability Domain:
Similar Compounds, with Predicted and Experimental Values

| Compound #1 | CAS: 62-53-3  
Dataset id: 4189 (Test set)  
SMILES: Nc1ccc(c1)N  
Similarity: 1  
Experimental value: NON-Mutagenic  
Predicted value: Possible NON-Mutagenic |
|-----------------|---------------------------------|
| Compound #2 | CAS: 100-61-8  
Dataset id: 2216 (Test set)  
SMILES: c1ccc(cc1)NC  
Similarity: 0.932  
Experimental value: NON-Mutagenic  
Predicted value: Possible NON-Mutagenic |
| Compound #3 | CAS: 108-44-1  
Dataset id: 1372 (Training set)  
SMILES: Nc1ccc(c1)C  
Similarity: 0.923  
Experimental value: NON-Mutagenic  
Predicted value: Possible NON-Mutagenic |
Modelling Exposure to chemicals for Risk assessment: a comprehensive Library of multimedia and PBPK models for Integration, prediction, uncertainty and sensitivity analysis.
Modelling Exposure to chemicals…

the exposure concept is extended from the environment to the internal tissues of the human body
...for **Risk assessment**: a comprehensive **Library of multimedia and PBPK models** ...

**Environmental Risk Assessments**

**Human Health Risk Assessments**
...for Integration, prediction...

integrated approach for ERA and HHRA
Identification of possible substitutes
To identify the chemical components and parameters associated with the adverse effect and to support substitution decision-making.
Software to assist user in making reproducible read across evaluations
similar chemicals

other information

structural alerts
Risky structural alerts

Possible substitutes
1. Food contact materials (FCM)
2. Biocides
3. Dispersant
4. Petroleum and oil fractions
5. Greener solvent
6. Cosmetics
Prioritization
Founded by the German Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety
Goal: **ranking** of chemicals on PBT bases

Including CMR and ED assessment
48 models running at the same time

Metabolites generation (microbial)
Toxicity assessment

Step 1
- Algae: L(E)C50 assessment, NOEC assessment
- Daphnia: L(E)C50 assessment, NOEC assessment
- Fish: L(E)C50 assessment, NOEC assessment

Step 2
- Algae: Acute / Chronic integration
- Daphnia: Acute / Chronic integration
- Fish: Acute / Chronic integration

Step 3
- Toxicity integration
- Reliability refinement
### Final scores

<table>
<thead>
<tr>
<th>No.</th>
<th>Metabolite</th>
<th>Id</th>
<th>SMILES</th>
<th>Label</th>
<th>Score(VPvE)</th>
<th>Score(SVHC)</th>
<th>Score(PBAT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Molecule 1</td>
<td></td>
<td>c1ccc(c1)Nc2 [.]</td>
<td>P B T - C M R E</td>
<td>0.463</td>
<td>0.893</td>
<td>0.577</td>
</tr>
<tr>
<td>2</td>
<td>Molecule 1 [M-01]</td>
<td></td>
<td>Oc2ccc(c(NNc1c1)c [.]</td>
<td>P B T - C M R E</td>
<td>0.261</td>
<td>0.836</td>
<td>0.383</td>
</tr>
<tr>
<td>3</td>
<td>Molecule 2</td>
<td></td>
<td>C1=C(C(C(2)c3c3)C(=O)[H]</td>
<td>P B T - C M R E</td>
<td>0.867</td>
<td>0.843</td>
<td>0.882</td>
</tr>
<tr>
<td>4</td>
<td>Molecule 3</td>
<td></td>
<td>c1(c)c(c(C)c1c1c1) [.]</td>
<td>P B T - C M R E</td>
<td>0.545</td>
<td>0.855</td>
<td>0.636</td>
</tr>
<tr>
<td>5</td>
<td>Molecule 4</td>
<td></td>
<td>C1c3ccc(c(c(c)c3)c3)O [.]</td>
<td>P B T - C M R E</td>
<td>0.833</td>
<td>0.929</td>
<td>0.663</td>
</tr>
<tr>
<td>6</td>
<td>Molecule 5</td>
<td></td>
<td>c1ccc(c1)c2ccc [.]</td>
<td>P B T - C M R E</td>
<td>0.874</td>
<td>0.698</td>
<td>0.815</td>
</tr>
<tr>
<td>7</td>
<td>Molecule 5 [M-01]</td>
<td></td>
<td>Oc1ccc(c1(O)c)[H]</td>
<td>P B T - C M R E</td>
<td>0.5</td>
<td>0.915</td>
<td>0.611</td>
</tr>
<tr>
<td>8</td>
<td>Molecule 6</td>
<td></td>
<td>c1ccc(c1)c3ccc [.]</td>
<td>P B T - C M R E</td>
<td>0.757</td>
<td>0.861</td>
<td>0.799</td>
</tr>
<tr>
<td>9</td>
<td>Molecule 6 [M-01]</td>
<td></td>
<td>Oc1ccc(c1(O)c) [.]</td>
<td>P B T - C M R E</td>
<td>0.251</td>
<td>0.866</td>
<td>0.362</td>
</tr>
</tbody>
</table>
To conclude:

Next future (5 years?)

- All physico-chemical, environmental, ecotoxicological and toxicological properties pre-assessed through *in silico* models, before preparation of a new substance.

- Ranking of new chemicals before their preparation for testing, based on *in silico* models.
To conclude (2):

- Larger set of properties used for final decision on new substances
- Information on the driving factors affecting a certain property, based on simulation

Next future (5 years?)
Thank you for your attention!

www.life-vermeer.eu
www.vegahub.eu