CAse studies on the Development and Application of in-Silico Techniques for Environmental hazard and Risk assessment

A QSAR-based compound prioritization for lab-testing for chemical safety assessment

Ullrika Sahlin[1], Barun Bhhatarai[2], Ester Papa[2], Tomas Öberg[1], Paola Gramatica[2]

[1]School of Natural Sciences, Linneaus University

[2]QSAR research unit in Environmental Chemistry and Ecotoxicology,
Department of Structural and Functional Biology, University of Insubria, Varese - Italy

Introduction & objective

REACH suggests that non-testing strategies, such as Quantitative Structure-Activity Relationships (QSARs), are useful to speed up assessments, reduce costs and avoid animal testing by

- Highlighting dangerous chemicals
- Filling data gaps
- Prioritize chemicals and focus experimental testing.

The practical integration of QSARs predictions into REACH can be discussed in terms of cost-efficient strategies of chemical testing.

On the design of testing strategies

Different QSAR-based strategies of prioritization for chemical testing result in various strengths of background information for chemical safety assessment.

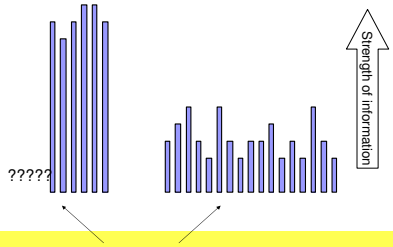
The strongest background information is provided by experimentally tested data for the endpoint in question. Strength decreases with number of extrapolations made. Strength increases for reliable QSAR models and experimental testing in concordance with the requirements set up by REACH (e.g. the OECD principles).

We suggest the design of strategies of compounds for lab-testing to be based on criteria related to

- Strength and spread of background information - There is a trade off between the strength of background information and the number of experimentally tested compounds (Fig 1). This criterion is important to motivate testing when the purpose is to build QSAR model to predict remaining untested compounds or when the purpose is to test compounds believed to be hazardous.
- Cost of testing - When similar, one should prioritize a less expensive chemical for testing, especially when this can lead to a reliable prediction (by extrapolation) of the more expensive one.

REACH = Registration Evaluation Authorization and Restriction of Chemicals

Figure 1. Information on many outweights the information on a few - Is it better to have weaker information on more compounds compared *to strong information on a few*?



An example of a QSAR-based prioritization

We discuss what is a cost-efficient strategy based on an example where the task was to select Per- and Polyfluorinated chemicals (PFCs) on the ECHA pre-registration list that should be prioritized for further lab-testing of eco-toxicological endpoints for CSA, given that

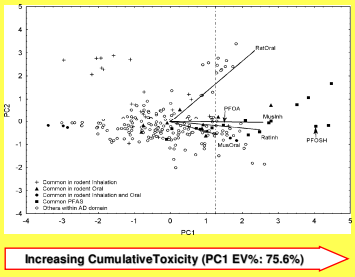
- Existing QSAR models are not reliably applicable for PFCs: they are mainly out of the AD, and
- No eco-toxicological data have been found for PFCs in reasonable amount for new QSAR model development.

A suggested strategy for prioritization:

1. Collect and prepare available data on PFCs
2. Predict endpoints for which reliable QSAR models exist or can be built upon available data (in this case LC50 inhalation and LD50 oral in Mouse and Rat, i.e. health endpoints).
3. Approximate target endpoint as toxicity trend based on an assumed correlation between the health endpoints and the eco-toxicological endpoints.
4. Select the most hazardous compounds for further testing. In this case 28 PFCs were prioritized by Inhalation study, 30 linear chain PFCs prioritized by Oral study and 22 PFCs prioritized by cumulative toxicity trend (PCA Fig 2).

In this way, starting from 50-58 experimental data on two health endpoints for PFCs, individual, externally predictive, models[1,2] were applied for predictions of 250-376 PFCs on the ECHA list for REACH (structural AD coverage of QSAR models: 75.6-90.9%).

Figure 2. PCA plot for increasing cumulative toxicity trend.



Discussion

Thus, most PFCs on the list could be assessed for prioritization but the strength of this background information for decision making is sensitive by the assumed correlation between health and eco-toxicological endpoint. Further, the final selection of compounds would have been different if the purpose was to build predictive QSAR model of the eco-toxicological endpoint, where selection should be based on experimental design with or without considering health endpoints or the cost of testing.

Conclusions

We suggest to design strategies for lab-testing considering **the strength and spread** of the resulting **background information**, such as the number of needed extrapolations to compare with experimental data on the target endpoint and the number of compounds that can be predicted with high reliability, and **differences in cost of testing** in the experimental design for QSAR modeling.
What strategy to use depend on the **purpose of testing** – is it to build predictive models, requiring lab-testing on hazardous as well as non-hazardous compounds, or to get stronger information on compounds believed to be hazardous, requiring known correlations between what is known and the target endpoint.

Acknowledgements
This study was funded by the FP7 project CADASTER (grant agreement 212668).
A CADASTER objective is to solve issues on how to practically integrate QSARs model predictions in risk assessment with the long-term goal to increase the use of non-testing information for regulatory decisions, while meeting the main challenges of quantifying and reducing uncertainty.

References
[1] Bhhatarai B, Gramatica P. Per- and Polyfluoro Toxicity (LC50 Inhalation) Study in Rat and Mouse Using QSAR Modeling. *Chem Res Toxicol.* 2010;23:528-39.
[2] Barun Bhhatarai, Paola Gramatica Oral LD50 Toxicity Modeling and Prediction of Per- and Polyfluorinated Chemicals on Rat and Mouse, *Molecular Diversity*, 2011,15 (2) Pages 467-476.