13

Reliability of published QSAR models for the prediction of developmental and reproductive toxicity – a case study on pesticides

A. Weyrich[1], M. Joel[2], G. Lewin[2], T. Hofmann[1], M. Frericks[3]
[1]BASF SE, Experimental Toxicology and Ecology, Ludwigshafen am Rhein, Germany
[2]Preclinical Science – Föll, Mecklenburg & Partner GmbH, Münster, Germany
[3]BASF SE, Agricultural solutions – Toxicology CP, Limburgerhof, Germany

Introduction

The importance of in silico methods for predicting toxicity has increased significantly in recent years due to the 3 R principle. In addition to read across approaches, quantitative structure-activity relationship (QSAR) models play a major role. These use existing experimental toxicity data for a range of chemicals to create a model that relates experimentally observed toxicity to molecular descriptors or fingerprints to predict the toxicity of other chemicals. In addition to these statistical models, expert rule-based methods (structural alerts) are available. QSARs are also increasingly required for evaluations of substances that are toxic to development and reproduction.

Objectives

Aim of the study was to analyse freely available and commercial developmental and reproductive toxicity QSAR tools regarding their performance in a database with approx. 300 pesticides.

Materials & Methods

A database of 342 pesticides, which are or were approved in the EU, with information about their structure, pesticide class, mode of action and GHS classification by ECHA was prepared. The developmental and/or reproductive potential of the pesticides was predicted with different in silico tools: VEGA, OECD (Q)SAR Toolbox, Leadscope Model Applier and CASE Ultra by MultiCASE. The results were subsequently compared to the GHS classification by ECHA.

Results

In all models, a large proportion (up to 77 %) of all pesticides was classified as outside the applicability domain of the model. The evaluation of the prediction of the remaining pesticides resulted in a balanced accuracy of the models between 0.48 and 0.66.

Conclusion

The results show that the models' predictions are not fit for purpose, yet. A meaningful prediction of the toxicity of the tested pesticides was provided only rarely, often due to incompatibility of the chemical spaces. For the correct assessment of the prediction, all information in the underlying database must be considered, as well as potential modes of action and metabolism.

The models can be improved by expanding the existing, underlying databases with experimental standardized data. In addition, it would be necessary to generate an endpoint-specific database to allow for an in-depth evaluation of data.