QSAR Toolbox as read-across/category building platform suitable for combining mechanistic data with other evidence

S. Dimitrov[1], C. Kuseva[1], T. Pavlov[1], O. Mekenyan[1]
T. Sobanski[2], D. Hirmann[2], T. Netzeva[2], A. Gissi[2], A. Martin Aparicio[2], Yuki Sakuratani[3], Romualdo Benigni[3]

1. *Laboratory of Mathematical Chemistry, University "Prof. As. Zlatarov", Bourgas, Bulgaria*
2. *Computational Assessment & Dissemination Unit, European Chemicals Agency*
3. *Environment Health and Safety Division, Environment Directorate, OECD*

INTRODUCTION

The OECD QSAR Toolbox is a computerised system for the hazard assessment of substances based on the category approach. The system incorporates theoretical knowledge, experimental data and computational tools organised in a logical workflow. The knowledge is used to define and mechanistically justify the selection of analogues of the target chemicals. The chemical category approach consists of grouping chemicals based on their similarity and using available experimental results from some members of the group (i.e. category) to fill data gaps for other members. The QSAR Toolbox is recommended by the OECD as the main tool for *in silico* predictions for regulatory purposes. Versions 3.x of the QSAR Toolbox have been downloaded over 8 600 times.

FUNCTIONALITIES

Two main pillars drive the processes of defining category and subsequently filling data gaps within the Toolbox:

1. knowledge coded into the system, organised in the so-called profilers;
2. the available experimental data, loaded from multiple databases.

The knowledge is used to build the category based on common structural features and mechanisms of interaction whereas the experimental data are used to fill the data gaps within category members.

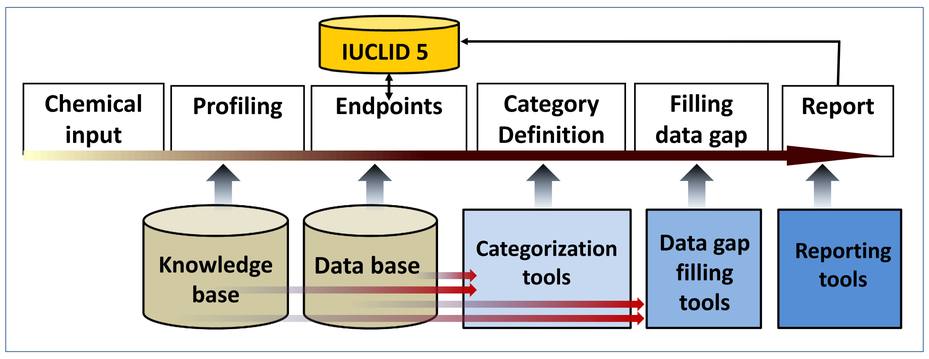


Fig. 1. Toolbox general scheme and workflow components

The system workflow consists of six main stages (**Fig. 1**):

- **Input** – to load the substance or list of substances into the system.
- **Profiling** – including identification of structural features, possible interactions mechanisms with macromolecules, etc. by making use of the **knowledge base**.
- **Endpoint** – gathering data from the available databases with about 60 000 substances having more than 1 500 000 data for different regulatory endpoints (Toolbox v.3.3.5)
- **Category definition** – defining a category for a target chemical based on its (structural) features. By making use of the knowledge base (profiling), analogues of target chemicals are selected. The analogues could be searched accounting for metabolism thus, identifying analogues having the same distribution profiles of interactions mechanisms with macromolecules.
- **Data gap filling** – three main approaches could be applied: read- across (RA), trend analysis (TA) and external QSARs. The RA assigns to the target the average or maximal toxicity values among the closest analogues; the TA searches a linear relationship between observed toxicities of analogues and the bioavailability parameter. External (Q)SAR models can be also loaded for various endpoints.
- **Reporting** – generation of a report of the predictions in a format similar to QPRF, accommodating the information from the QSAR Toolbox.

METABOLISM

A key feature in the QSAR Toolbox is the metabolism simulation and the possibility to take the metabolic activation into account when building categories. Profiling results could be found for target chemicals as well as for simulated/observed metabolites from:

- ✓ autoxidation, useful e.g. when the target endpoint is skin sensitisation
- ✓ rat liver S9 metabolism, relevant for *in vitro* genotoxicity predictions etc.

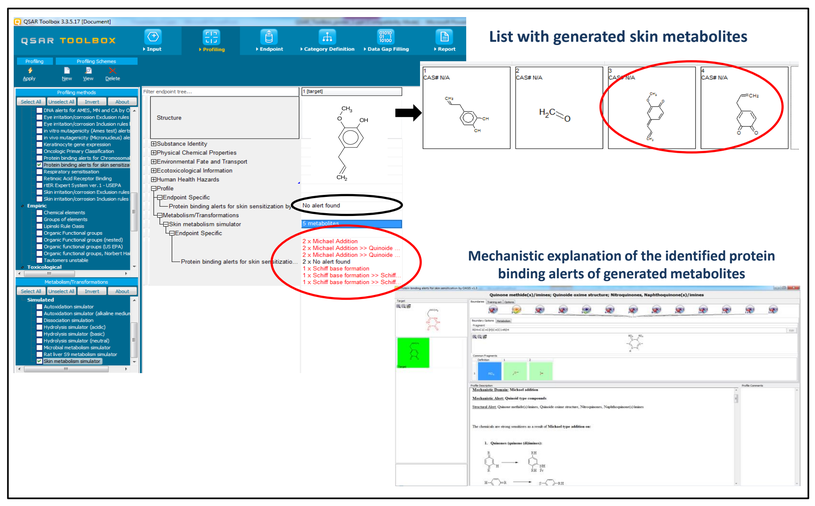


Fig. 2. Simulation of skin metabolism and additional explanation of interaction mechanisms of activated metabolites. In the example, no alert is found for the target (black circle). However, alerts associated to the skin sensitisation effect are found in generated metabolites (in red). Mechanistic explanations for the alerts is also available.

EXAMPLES

Application of the category approach for predicting *Ecotoxicity* and *Human health toxicity* by read-across are illustrated in Fig. 3 and Fig.4 respectively.

Prediction of ***short term toxicity to fish*** is illustrated in **Fig. 3**. A target chemical is classified as "***Aryl halide***" according to the "Organic functional group" (OFG) profiler.
Additional subcategorisations by using endpoint-specific profilers such as the ***Acute aquatic toxicity MOA*** and ***ECOSAR classification*** have been applied to eliminate analogues behaving differently from the target chemical which is classified as a "***Basesurface narcotic***".

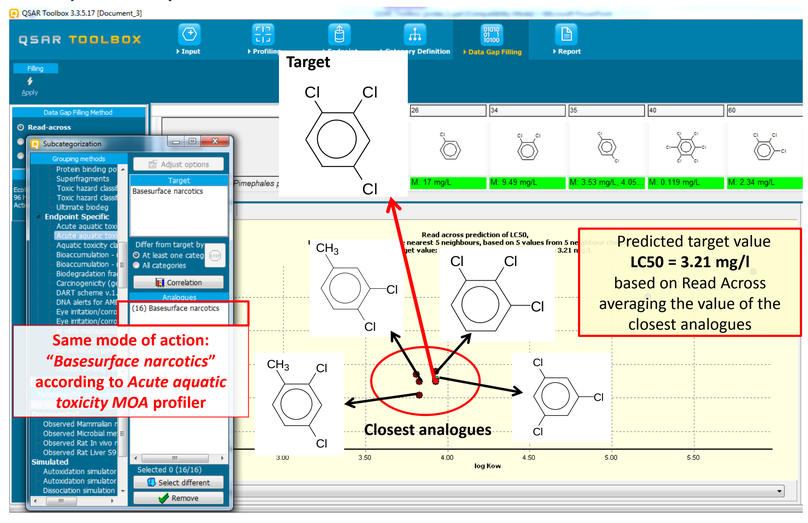
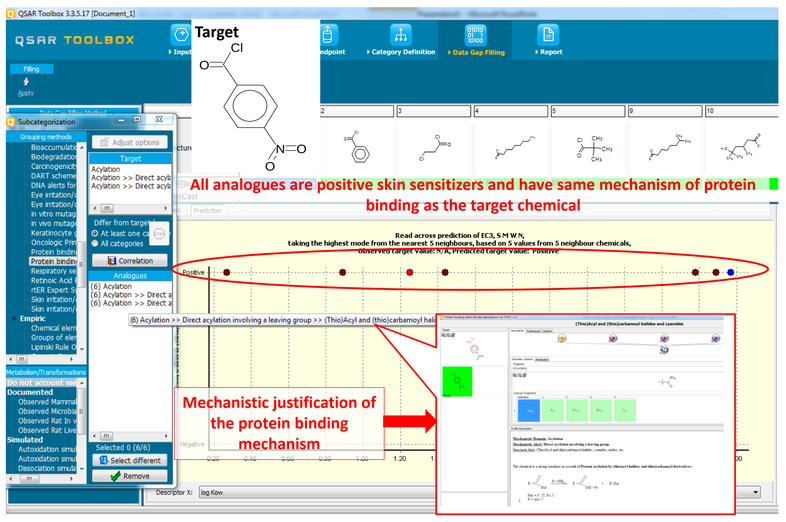


Fig. 3. Data gap filling based on read-across for short term toxicity to fish



Prediction of *in vivo* ***skin sensitisation*** is illustrated in **Fig. 4**. A target chemical is classified as "***Acyl halide***" according to the ***US EPA new chemical categories*** profiler. Additional subcategorisations by using the endpoint-specific profiler ***protein binding alerts for skin sensitisation*** have been applied to eliminate analogues behaving differently from the target chemical, which is classified as an ***"Acylating agent"***.

Fig. 4. Data gap filling based on read-across for *in vivo* skin sensitisation

The system incorporates many **additional functionalities** such as DB query searching, endpoint vs endpoint correlation, etc. These functionalities could support the read-across hypothesis.

The correlation between ***acute oral toxicity (LD50, rat)*** and ***acute fish toxicity (LC50, short term fish)*** data within the QSAR Toolbox environment is illustrated for alcohols as shown in **Fig. 5**. Statistics of the correlation are provided.

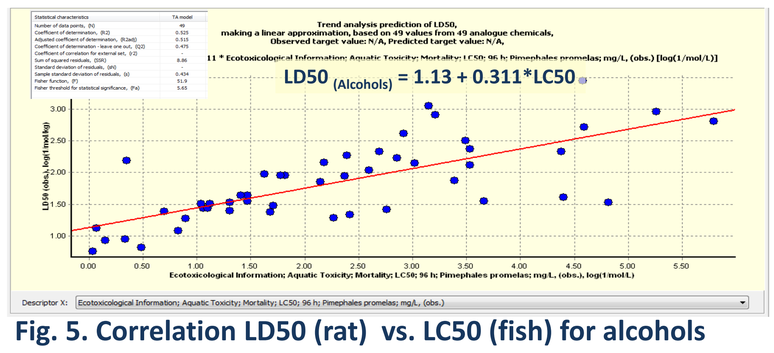


Fig. 5. Correlation LD50 (rat) vs. LC50 (fish) for alcohols

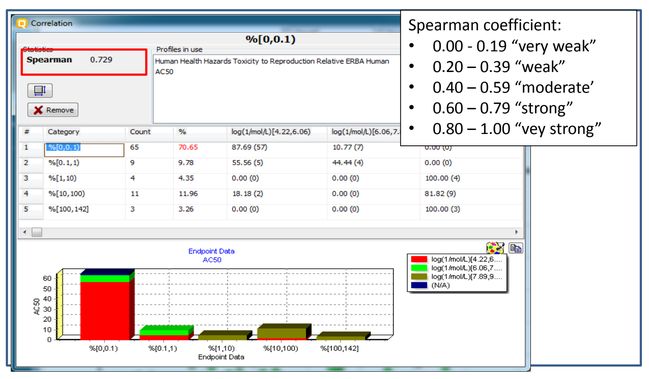


Fig. 6. Correlation of *Toxcast/ AC50 - NCGC Reporter Gene Assay ERa Agonist* and *human ERBA*

One of the newly added databases in the Toolbox is **ToxCast**. The challenge to use the ToxCast experimental data is how to link various enzyme activities to apical endpoints. This association could be explored by building correlation between enzymatic activities and endpoints like Estrogen receptor binding affinity (Fig.6).

SUMMARY

The OECD QSAR Toolbox is a unique software system that:

- **Provides uniform application of the category approach for hazard assessment of chemicals**
- **Uses mechanistic knowledge, structural features and parametric boundaries for category building.**
- **Accounts for metabolic activation during the categorisation process (e.g. for endpoints such as mutagenicity, skin sensitisation, RDT, carcinogenicity, etc.)**
- **Has a straightforward workflow for hazard assessment**

A completely new, re-designed Toolbox v4.0 will be available in spring 2017. This version will have many new features including: a new faster DB engine, the possibility of automating predictions and web services. It will also interface with Effectopedia (OECD), OncoLogic (US EPA), the Danish QSAR DB and CATALOGIC/TIMES (LMC).