The Use of QSAR Models to Predict Hazard Properties

By Ravi Menon, EHS³ Advisor

Date: December 12, 2018

Introduction

Afton's EHS³ (environmental protection, health, safety, security and stewardship) department routinely uses cutting-edge QSAR models to predict the hazards of both new and existing chemicals. This paper is an introduction to the use of such models and gives examples of the benefits delivered.

Quantitative Structure-Activity Relationship (QSAR) models are theoretical, computerized models that can be used to qualitatively or quantitatively predict the physicochemical, biological (e.g., ecotoxicological) and environmental fate (e.g., biodegradation, bioaccumulation) properties of chemical substances using features of the chemical structure.

Testing each chemical substance for all the hazard data required by regulatory authorities for registration can be very expensive and time-consuming. QSAR, when technically feasible and scientifically justifiable, is an effective means to generate the required hazard data avoiding unnecessary testing.

Regulatory agencies such as the European Chemical Agency (ECHA), U.S. Environmental Protection Agency (EPA) and Environment & Climate Change Canada (ECCC) will typically accept QSAR results when the following conditions are met:

- The QSAR model is validated according to the Organization of Economic Cooperation & Development (OECD) principles (see below),
- The prediction is fit for the specific regulatory purpose (e.g., classification & labeling, risk assessment, etc.),
- The prediction is reported in the proper format, and presented in a weight-of-evidence approach, along with other supporting information.

QSAR can also be an effective and efficient tool for screening new R&D molecules early in the development lifecycle, to look for 'show stoppers' like PBT (persistent, bioaccumulative and toxic) properties. This can help researchers correct course before extensive resources have been invested in development.

There are several QSAR models and platforms available, such as the US EPA's EPI Suite™ (includes models such as BIOWIN™, BCFBAF™ and ECOSAR™) and the QSAR Toolbox from OECD, which is an expert system that includes QSAR models as well as the functionality to find structural analogues to do read-across predictions. While both the EPI Suite and the OECD QSAR Toolbox can be downloaded for free, there are also proprietary models such as CATALOGIC from LMC, Bulgaria, that require licensing.

A QSAR model is developed using experimental data on a 'training set' of chemicals, which is then tested on a 'validation set'. For regulatory acceptance, QSAR models must satisfy the following five principles as required by OECD:

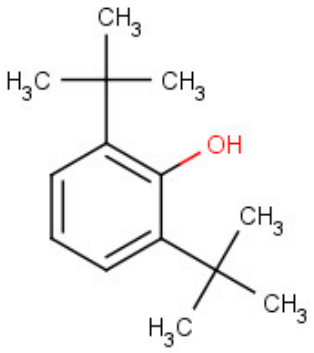
- A defined endpoint
- An unambiguous algorithm
- A defined domain of applicability
- Appropriate measures of goodness-of-fit, robustness and predictivity
- A mechanistic interpretation if possible

If the target chemical is within the applicability domain of the validated model, prediction can be made.

Examples of QSAR Use

Example chemical:

Name: 2,6-di-tertiary-butylphenol



CAS 128-39-2

SMILES (Simplified Molecular Input Line Entry System) code: Oc(c(ccc1)C(C)(C)C)c1C(C)(C)C

Biodegradation

Assessment of the biodegradation potential of chemicals is required under the chemical control regulation of many countries/regions, as part of PBT assessment. BIOWIN v4.10 in the EPI Suite platform from US EPA is a freely available model that is accepted by several regulatory agencies worldwide.

BIOWIN uses seven individual models to assess the biodegradation potential and predicts the ready biodegradability of the modeled chemical. Using the SMILES code as input, BIOWIN predicts the following results for the example chemical:

Probability of Rapid Biodegradation (BIOWIN v4.10):

- Biowin1 (Linear Model): 0.3973
- Biowin2 (Non-Linear Model): 0.0788

Expert Survey Biodegradation Results:

- Biowin3 (Ultimate Survey Model): 2.3753 (weeks-months)
- Biowin4 (Primary Survey Model): 3.2829 (days-weeks)

MITI Biodegradation Probability:

- Biowin5 (MITI Linear Model): 0.3249
- Biowin6 (MITI Non-Linear Model): 0.1588

Anaerobic Biodegradation Probability:

- Biowin7 (Anaerobic Linear Model): -0.5154

Overall Prediction Result: Not readily biodegradable

Bioaccumulation Prediction

Assessment of the bioaccumulation is also required by global regulatory agencies to evaluate the PBT characteristics of a chemical. To assess the bioaccumulation potential of a chemical, the BCF (bioconcentration factor in fish), is the typical data point required.

The conventional way of getting BCF information is through testing (OECD 305), but this can be quite expensive (USD 150,000+), time-consuming (6 months+ turnaround), and leads to the death of many fish (typically 100+). QSAR offers an effective alternate to testing, when technically feasible and scientifically justifiable.

The QSAR model, CATALOGIC BCF base-line model v.03.10 (LMC, Bulgaria) was used to estimate the BCF of the example chemical. The model satisfies the OECD validation principles as follows:

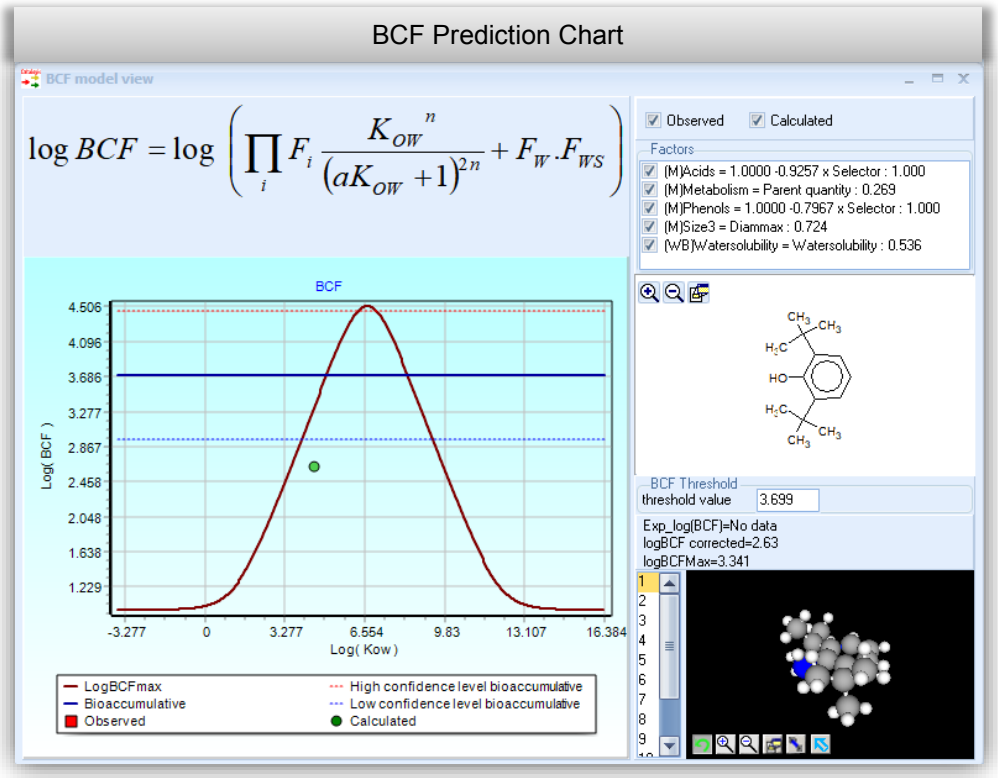
- A defined endpoint – BCF in fish, L/kg wet weight
- Unambiguous algorithm

$$\log BCF = \log\left(\prod_i F_i \frac{K_{OW}{}^{n}}{(aK_{OW}+1)^{2n}} + F_W.F_{WS}\right)$$

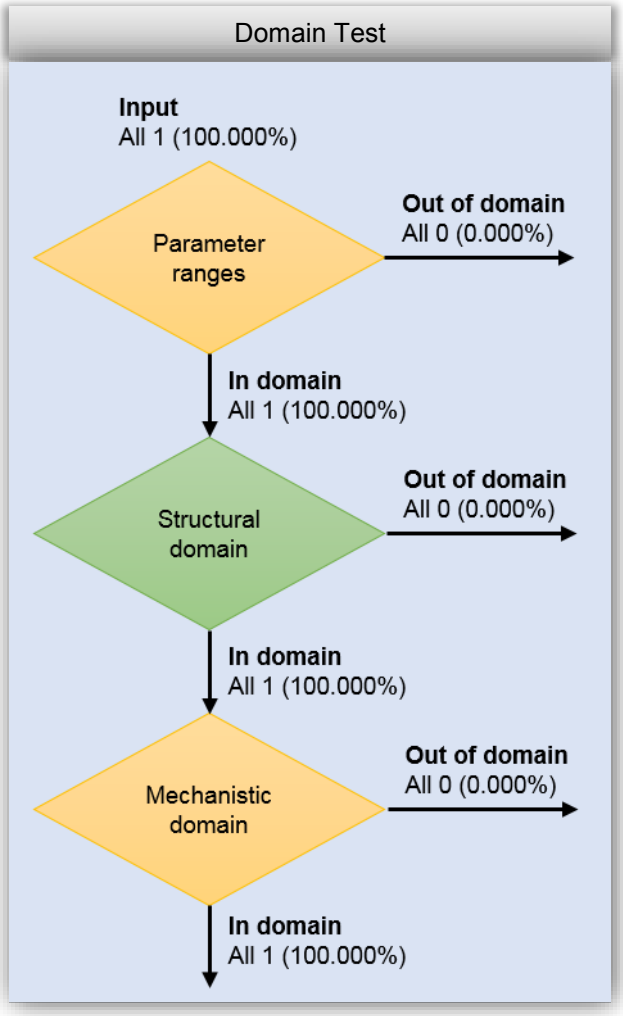
- Defined applicability domain – parametric (log Kow, MW & WS), structural (ACF) & mechanistic (passive diffusion)
- Appropriate measures of goodness-of-fit, robustness & predictivity – Training set: N=81, RSS=31, R2 = 0.85; validation set: N = 123; PRSS = 13; Q2 = 0.80
- A mechanistic interpretation, if possible – bioaccumulation of lipophilic organic chemicals in fish through the respiratory organs, mitigating factors

The predicted log BCF for the chemical is 2.63 (BCF 426.6 L/kg).

Overall Prediction: Not bioaccumulative



The chemical falls within the domains of applicability of the model, making this a reliable prediction.



Further, the CATALOGIC model also allows the generation of a report in the QSAR Prediction Reporting format (QPRF), which is customizable and can be submitted to the regulatory agency along with the details on the model itself. There is significant cost and time savings by using this non-testing method for the BCF endpoint and it is also in the interests of animal welfare.

Conclusions

Use of QSAR models is a proven non-testing means to predict hazard properties required by regulatory authorities for chemical registration/notification. However, to be acceptable to regulators, the model used must be validated according to OECD principles, the target chemical must be within the applicability domain, and the results reported in the proper format. When technically feasible and scientifically justifiable, the QSAR approach can provide a reliable estimate of the hazard endpoint, avoiding significant testing cost and time, and is also in the interests of animal welfare. Regulators are increasingly accepting QSAR as a means for generating hazard data.

References:

1. ECHA Practical Guide: How to use and report (Q)SARs 3.1, July 2016
2. Guidance document on the validation of (quantitative)structure-activity relationships [(Q)SAR] models, ENV/JM/MONO(2007)2, OECD, 2007