Environmental Research 142 (2015) 529–534

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Comparison between bioconcentration factor (BCF) data provided by industry to the European Chemicals Agency (ECHA) and data derived from QSAR models

Maria I. Petoumenou [a,*], Fabiola Pizzo [a], Josep Cester [b], Alberto Fernández [b], Emilio Benfenati [a]

[a] IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Department of Environmental Health Sciences, Laboratory of Environmental Chemistry and Toxicology, Via La Masa 19, Milan, 20156 Italy

[b] URV – Universitat Rovira i Virgili, Departament d'Enginyeria Quimica, Av. Països Catalans 26, 43007 Tarragona, Catalunya, Spain

ARTICLE INFO

Article history:
Received 18 May 2015
Received in revised form
5 August 2015
Accepted 9 August 2015

Keywords:
QSAR
Bioconcentration factor
REACH
Applicability domain
Fish

ABSTRACT

The bioconcentration factor (BCF) is the ratio of the concentration of a chemical in an organism to the concentration in the surrounding environment at steady state. It is a valuable indicator of the bioaccumulation potential of a substance. BCF is an essential environmental property required for regulatory purposes within the Registration, Evaluation, Authorization and restriction of Chemicals (REACH) and Globally Harmonized System (GHS) regulations. *In silico* models for predicting BCF can facilitate the risk assessment for aquatic toxicology and reduce the cost and number of animals used. The aim of the present study was to examine the correlation of BCF data derived from the dossiers of registered chemicals submitted to the European Chemical Agency (ECHA) with the results of a battery of Quantitative Structure-Activity Relationship (QSAR). After data pruning, statistical analysis was performed using the predictions of the selected models. Results in terms of R^2 had low rating around 0.5 for the pruned dataset. The use of the model applicability domain index (ADI) led to an improvement of the performance for compounds falling within it. The variability of the experimental data and the use of different parameters to define the applicability domain can influence the performance of each model. All available information should be adapted to the requirements of the regulation to obtain a safe decision.

1. Introduction

In environmental assessments of the aquatic compartment, the bioconcentration factor (BCF) is of the utmost importance (Arnot and Gobas, 2006; Wang et al., 2014). The BCF refers to the likelihood of a chemical to concentrate and consequently accumulate in an aquatic organism. A related value is the bioaccumulation factor (BAF), which also takes the dietary exposure under consideration (Arnot and Gobas, 2004). Chemicals that are not degraded in the environment and bioaccumulate in biota may also exert long-term adverse effects (Pavan et al., 2008).

Demand for BCF data is increasing since they are required by the European Commission (EC) regulation Registration, Evaluation, Authorization and restriction of Chemicals (REACH) (EC Regulation 1907/2006) and they may be also useful within Globally Harmonized System (GHS). REACH aims to raise the level of protection for human and environmental health and to promote the free movement of substances, through early identification of the properties of chemicals. The GHS regulation is designed to build a global, harmonized architecture for classification and labeling of chemicals. In Europe the GHS is implemented through the Classification Labeling and Packaging (CLP) legislation, which is integrated into REACH (EC Regulation 1907/2006).

For the first REACH registration deadline (30 November 2010), registrants submitted dossiers with toxicological and exposure information (including BCF) for substances to the European Chemical Agency (ECHA) whose task is to check "at least 5%" of the total registration dossiers received for each tonnage band (Cesnaitis et al., 2014).

For the BCF assessment fish are exposed to a substance in the water phase only under controlled laboratory conditions (Costanza et al., 2012). The OECD test No. 305 has been updated recently including the dietary exposure aside from the aqueous exposure methodology. The experimental study for determination of the BCF for each chemical is relatively expensive and may require

* Corresponding author. Tel.: +39 0239014668.
E-mail address: maria.petoumenou@marionegri.it (M.I. Petoumenou).

http://dx.doi.org/10.1016/j.envres.2015.08.008
0013-9351/

hundreds of vertebrate test animals (Rodriguez-Sanchez et al., 2014). Since data-sharing plays a key role in avoiding unnecessary animal testing, the REACH legislation also requires data-sharing among the registrants (Spielmann et al., 2011). [1]

In addition, REACH encourages the development and the use of alternative methods for the assessment of hazard of substances (EC Regulation 1907/2006; Worth et al., 2007). The use of existing data, weight-of-evidence-based approaches, (quantitative) structure-activity relationship models ((Q)SAR), read-across approaches, *in vitro* techniques, and grouping of substances are indicated as alternatives to animal testing (Cesnaitis et al., 2014).

In silico models for predicting BCF properties for chemicals can be divided into linear relationships that correlate the BCF with the octanol–water partition coefficient (log *P*), models based on experimentally derived descriptors, models based on theoretical molecular descriptors and expert systems (Pavan et al., 2008). It is known that the correlation between log *P* and log BCF of chemicals is not completely linear; when log $P > 6$ the log BCF value of the chemical tends to decrease (Grisoni et al., 2015). Other studies used the mass-balance modeling approach (Arnot and Gobas 2003).

In the regulatory context, the main goals are to select and use simple and easy to calculate parameters (such as log *P*) and to develop models that can be helpful in the prediction of the most accurate BCF data (Garg and Smith, 2014). The requirements for the correct selection and use of QSAR models for regulatory purposes are listed in the Annex XI within the REACH legislation (Cesnaitis et al., 2014).

In the past the performance of several *in silico* models was assessed, using experimental values taken from the literature (Gissi et al., 2015). Here we present a study carried out within the European project CALEIDOS (http://www.caleidos-life.eu) to test the reliability of freely available QSAR models that provide values for BCF with the data provided by registrants to ECHA. We collected data from the ECHA database (http://echa.europa.eu/web/guest/information-on-chemicals/registered-substances) and assessed the performance of BCF models in the VEGA platform (http://www.vega-qsar.eu/) and EPI Suite (BCFBAF model) developed by US Environmental Protection Agency (EPA) (http://www.epa.gov).

2. Material and methods

2.1. The dataset

The workflow of this study follows:

- To gather all BCF data on registered substances within REACH;
- To verify the correctness of the chemical identity and the appropriateness for the QSAR models, selecting organic mono-constituents;
- To select data of good quality obtained according to the guidelines;
- To calculate the mean value of the experimental BCF data;
- To predict BCF values for the selected substances using a series of QSAR models;
- To calculate the statistics of the QSAR models.

We used the BCF data from the ECHA CHEM database, (http://echa.europa.eu/web/guest/information-on-chemicals/registered-substances) which reports information on chemical substances manufactured or imported in Europe. This information comes from the registration dossiers submitted by companies to ECHA under the European REACH regulation (EC), No 1907/2006. We started from 3092 studies on a total of 426 unique CAS RN. Further 186 studies had not CAS RN and they were deleted. We carefully checked the data in the ECHA CHEM database in order to retain only highly reliable and homogeneous studies. Studies related to BAF and biomagnification factor (BMF) were discarded. The database contained studies labeled with different reliability scores following Klimisch's criteria. Thus, to avoid less reliable data we selected only those with Klimish code 1 or 2; code 1 means that data are reliable without restrictions while code 2 means that data are reliable with restrictions (Klimisch et al., 1997). We eliminated all inorganic compounds, Substances of Unknown or Variable composition, Complex reaction products or Biological materials (UVCB), polymers and multi-constituent compounds. Only mono-constituent organic compounds with purity higher than 80% (where this information was available) were considered. To ensure an accurate dataset of chemicals for statistical analysis we applied a stepwise approach for data pruning using data from homogeneous studies, conducted according to the OECD Guideline 305; we only kept values from studies conducted on the following test species: *Brachidanio rerio*, *Pimephales promelas*, *Cyprinus carpio*, *Orizias latipes*, *Poecilia reticulata*, *Lepomis macrochirus*, *Oncorhyncus mykiss*, *Gasterosteus aculeatus*, as required by REACH regulation. We accepted studies using the whole body or only specific tissues (mussel) for the BCF calculation. The correspondence between CAS numbers and chemical structures was double-checked with Chem ID Plus (http://chem.sis.nlm.nih.gov/chemidplus/) and ChemSpider (http://www.chemspider.com/) for all the compounds. When several tautomers were available for the same structure, the most stable tautomer was considered using Marvin Sketch (version 5.9.2, 2012 ChemAxon) (http://www.chemaxon.com). The SMILES were modified deleting the ions and obtaining the neutralized structure. We finally obtained a dataset of 148 organic substances that we used for analysis.

This relatively severe data pruning reflects the heterogeneous sources of the data submitted to ECHA (Sobanska et al., 2014, Tarazona et al., 2014). The data unit was often not clearly expressed, or was expressed as L/kg or in logarithm units. Many figures were given as ranges, often with a wide spread between the minimum and maximum values within the same study. In these cases we calculated the arithmetic mean, based on logarithmic value, and reported as a punctual value. After that, 70% of the compounds had more than one value, resulting from different studies, so the arithmetic mean was calculated again for each compound.

2.2. QSAR models

The models we used were selected among those that gave the best results on datasets of good quality, on the basis of a previous study (Gissi et al., 2015), which are publicly and freely available. Thus, BCF was predicted using QSAR models on VEGA (http://www.vega-qsar.eu/-VEGANIC v. 1.0.8) and EPI Suite (http://www.epa.gov/oppt/exposure/pubs/episuitedl.htm-EPIWEB 4.11)

[1] Abbreviations: ADI, applicability domain index; B, bioaccumulative; BAF, bioaccumulation factor; BCF, bioconcentration factor; BMF, biomagnification factor; CAS RN, chemical abstracts service registry number; CLP, Classification Labeling and Packaging; EC, European Commission; ECHA, European Chemical Agency; EPA, Environmental Protection Agency; FN, false negative; FP, false positive; GHS, Globally Harmonized System; log P, logarithm of the octanol–water partition coefficient; MCC, Matthews' correlation coefficient; MW, molecular weight; nB, non-bioaccumulative; OECD, Organization for Economic Cooperation and Development; QSAR, quantitative structure activity relationship; R^2, coefficient of determination; REACH, Registration, Evaluation, Authorization and restriction of Chemicals; RMSE, root mean square error; TN, true negative; TP, true positive; UVCB, Unknown or Variable composition, Complex reaction products or Biological materials substances; y_i, experimental value; $\hat{y}_i$, predicted value; y_{avg}, mean of the experimental values;