



ELSEVIER

Contents lists available at ScienceDirect

## Environmental Research

journal homepage: [www.elsevier.com/locate/envres](http://www.elsevier.com/locate/envres)

# Evaluation of QSAR models for predicting the partition coefficient ( $\log P$ ) of chemicals under the REACH regulation



Claudia Ileana Cappelli<sup>a</sup>, Emilio Benfenati<sup>a,\*</sup>, Josep Cester<sup>b</sup>

<sup>a</sup> Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, IRCCS-Istituto di Ricerche Farmacologiche "Mario Negri", Milan, Italy

<sup>b</sup> Departament d'Enginyeria Química, Universitat Rovira i Virgili, Av. Països Catalans 26, Catalunya, Tarragona 43007, Spain

## ARTICLE INFO

## Article history:

Received 3 July 2015

Received in revised form

18 September 2015

Accepted 21 September 2015

## Keywords:

 $\log P$  $\log K_{ow}$ 

REACH

QSAR

Applicability domain

## ABSTRACT

The partition coefficient ( $\log P$ ) is a physicochemical parameter widely used in environmental and health sciences and is important in REACH and CLP regulations. In this regulatory context, the number of existing experimental data on  $\log P$  is negligible compared to the number of chemicals for which it is necessary. There are many models to predict  $\log P$  and we have selected a number of free programs to examine how they predict the  $\log P$  of chemicals registered for REACH and to evaluate whether they can be used in place of experimental data. Some results are good, especially if the information on the applicability domain of the models is considered, with  $R^2$  values from 0.7 to 0.8 and root mean square error (RMSE) from 0.8 to 1.5.

© 2015 Published by Elsevier Inc.

## 1. Introduction

The partition coefficient ( $\log P$ ) is a physicochemical parameter defined as the ratio (as logarithm) between the equilibrium concentrations of a substance dissolved in a two-phase system consisting of a polar and less polar for the most part immiscible solvent. In the life and pharmaceutical sciences, these two phases are often restricted to two liquid solvents, water and *n*-octanol.

Four experimental methods are used to measure the *n*-octanol/water partition coefficient ( $\log K_{ow}$ ). Three are direct measurements (shake flask method, generator column method, and slow-stirring method), and one is indirect (reverse phase HPLC).

The shake flask method (OECD TG 107) is the default procedure. It is considered to give accurate results for low to medium hydrophobic chemicals. For repeatability, three replicates of the method should fall within  $\pm 0.3$  logarithmic units (OECD, 1995). The generator column method (Woodburn et al., 1984) is suitable for measuring the  $\log K_{ow}$  of more hydrophobic chemicals. There is no information on its repeatability in the literature. The slow-stirring method (OECD TG 123) is a more recent method developed as an alternative to the shake flask procedure. This

method can also be used for very hydrophobic chemicals; it takes a few days to reach equilibrium. An intra-laboratory median standard deviation of 0.15–0.3 logarithmic units has been estimated (OECD, 2006).

Reverse phase HPLC (OECD TG 117) is a relatively quick way of estimating  $\log K_{ow}$ . It is not measured directly, but from a correlation between  $\log K$  (capacity factor) and  $\log K_{ow}$  for a series of reference chemicals. Therefore it depends on the quality of the  $\log K_{ow}$  measurement of the reference chemicals (often measured by the shake flask method) (ECHA, 2012). In this case three replicates of the experimental procedure should fall within  $\pm 0.1$  logarithmic units (OECD, 2004a). Special attention is usually paid to OECD methods, but test data with equivalent methods from other organizations such as EPA-FIFRA, ASTM or ISO should equally be accepted.

There are certain structural or physico-chemical properties that can make it difficult to determine  $K_{ow}$  accurately. This is the case for substances that are poorly soluble, volatile, surface-active, or ionizable, or are subject to rapid degradation (e.g. phototransformation, hydrolysis, oxidation or biotic degradation).

$\log K_{ow}$  is moderately temperature-dependent and usually measured at 25 °C. It is used as input to assess a number of properties, such as environmental partitioning, sorption, bioavailability, bioconcentration and bioaccumulation, ecotoxicity, human toxicity and the kinetic and dynamic aspects of drug activity.  $\log K_{ow}$  is a critical parameter for chemical safety assessment, classification, labeling and packaging (CLP) (EC, 2008), and PBT assessment/screening (Fig. 1).

\* Correspondence to: Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, IRCCS – Istituto di Ricerche Farmacologiche "Mario Negri", via La Masa 19, 20156 Milan, Italy.

E-mail addresses: [claudia.cappelli@marionegri.it](mailto:claudia.cappelli@marionegri.it) (C.I. Cappelli), [emilio.benfenati@marionegri.it](mailto:emilio.benfenati@marionegri.it) (E. Benfenati), [josep.cester@urv.cat](mailto:josep.cester@urv.cat) (J. Cester).

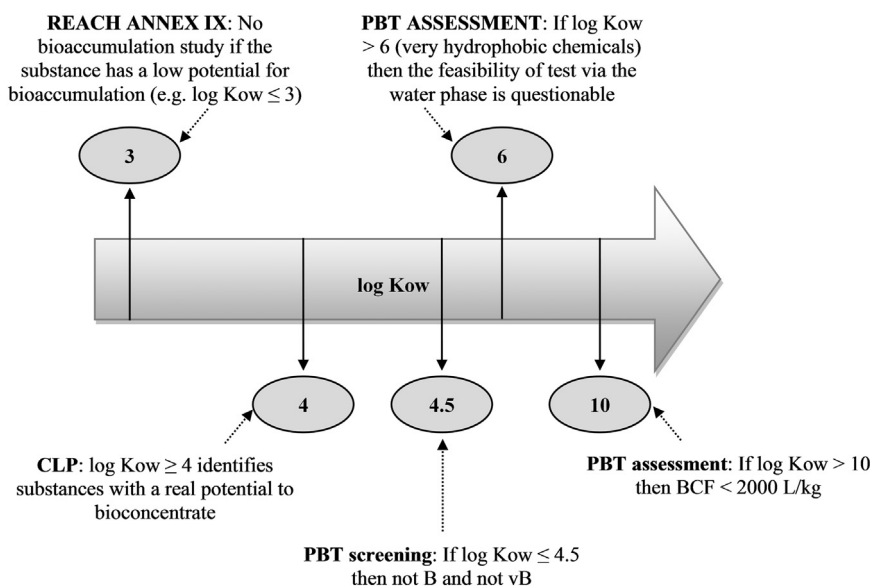


Fig. 1. Regulatory requirements for  $\log K_{ow}$ .

Within the REACH regulation (EC, 2006)  $\log P$  is required at all tonnage bands (i.e.  $> 1$  t/y; information requirements according to REACH, Annexes VII–X). For CLP, a cut-off of  $\log K_{ow} \geq 4$  is intended to identify substances with a potential to bioconcentrate. The potential of an organic substance to bioconcentrate is primarily related to its lipophilicity and a surrogate measure of lipophilicity is  $\log K_{ow}$  which is correlated with the bioconcentration factor (BCF), for lipophilic non-ionized organic substances, undergoing minimal metabolism or biotransformation within the organism (Dimitrov et al., 2005). A  $\log K_{ow} \geq 4$  can be used in place of an experimentally determined  $BCF \geq 500$  in combination with acute toxicity to evaluate chronic toxicity for classification purposes (ECHA, 2013).

The hazard categories for acute and chronic aquatic toxicity and their related criteria are set out in CLP, Annex I. As regards the long-term aquatic hazard,  $\log K_{ow}$  is used to classify substances for which there are no adequate chronic toxicity data.

The number of experimental values of  $\log K_{ow}$  is negligible compared to the number of chemicals for which it is needed, especially in the regulatory context. Experimental determination of  $\log K_{ow}$  is time- and material-consuming and in many cases only an estimate of the lipophilicity of the chemicals is required. The literature provides various figures for the same compound, which can differ by more than one order of magnitude (Fielding et al., 1992). Benfenati et al. (2003) compared experimental and calculated  $\log P$  values using a dataset of 235 pesticides and experimental values from four different sources. They reported that experimental values for the same chemical differed widely, by more than 0.5 log unit. Consequently, reliable predictive methods to compute the  $\log K_{ow}$  are really needed.

Within the EC funded project CALEIDOS (Chemical Assessment according to Legislation Enhancing the In silico Documentation and Safe use, LIFE11 ENV/IT/295, <http://www.caleidos-life.eu/index.php>) a set of six freely available software packages was selected to predict  $\log K_{ow}$  of 729 organic compounds submitted by industries to ECHA for REACH registration. Several statistical parameters were considered to assess the predictive performance of existing freely available QSAR models for  $\log K_{ow}$  prediction in order to promote and enhance their use for further REACH registrations. We show the results in Section 3.

## 2. Materials and methods

### 2.1. The dataset

Data on 2482 chemicals already registered for REACH were collected from the ECHA CHEM database (ECHA, 2014) and retrieved from the graphical user interface of the OECD QSAR Toolbox (<http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm>). The ECHA CHEM database includes information on chemical substances manufactured or imported in Europe. This information originates from the registration dossiers submitted by companies to ECHA in the framework of the European REACH regulation (EC, 2006).

The dataset was pruned, considering experimental values (calculated values and read-across evaluations were discarded, but experimental weight of evidence was retained) for organic monoconstituents (inorganic compounds, mixtures and UVCBs were eliminated) and studies with reliability 1 and 2 (Klimisch score) (Klimisch et al., 1997). Experimental data were selected on the basis of the use of the test guidelines recommended by the endpoint specific guidance for the implementation of REACH legislation (ECHA, 2012), in particular the shake flask method (OECD TG 107 and EPA OPPTS 830.7550), the correlation approach using HPLC method (OECD TG 117; EPA OPPTS 830.7570; ASTM E 1147–92), the slow-stirring method (OECD TG 123) and the generator column method (EPA OPPTS 830.7560). We took account of the experimental conditions of the test: the temperature, pH (to be sure that the experimental values of the substances refer to the non-ionized form) and the purity (high purity  $> 80\%$ ). Salts were excluded from the analysis.

The correspondence between SMILES, structures and CAS RN was checked and the chemicals without CAS RN or SMILES identified.

The final dataset contained data on 729 chemicals; few had more than one experimental value. In case of multiple experimental data for the same substance, the arithmetic mean was calculated.

### 2.2. The models

Six free software packages were used to predict  $\log K_{ow}$ ; Table 1 summarizes their main features.

In the VEGA platform (<http://www.vega-qsar.eu/>), AlogP and

Download English Version:

<https://daneshyari.com/en/article/6351794>

Download Persian Version:

<https://daneshyari.com/article/6351794>

[Daneshyari.com](https://daneshyari.com)