Food and Chemical Toxicology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Food and Chemical Toxicology

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

Predicting the bioconcentration factor of highly hydrophobic organic chemicals

Rajni Garg [a,*], Carr J. Smith [b]

[a] Computational Science Research Center, San Diego State University, San Diego, CA 92182-7720, USA
[b] Health, Safety & Environment, Albemarle Corporation, Baton Rouge, LA 70801-1765, USA

ARTICLE INFO

Article history:
Received 30 October 2013
Accepted 26 March 2014
Available online xxxx

Keywords:
QSAR
Bioconcentration factor
Risk assessment
Organic chemicals
Hydrophobicity
Octanol–water partition coefficient

ABSTRACT

Bioconcentration refers to the process of uptake and buildup of chemicals in living organisms. Experimental measurement of bioconcentration factor (BCF) is time-consuming and expensive, and is not feasible for a large number of chemicals of regulatory concern. Quantitative structure–activity relationship (QSAR) models are used for estimating BCF values to help in risk assessment of a chemical. This paper presents the results of a QSAR study conducted to address an important problem encountered in the prediction of the BCF of highly hydrophobic chemicals. A new QSAR model is derived using a dataset of diverse organic chemicals previously tested in a United States Environmental Protection Agency laboratory. It is noted that the linear relationship between the BCF and hydrophobic parameter, i.e., calculated octanol–water partition coefficient (ClogP), breaks down for highly hydrophobic chemicals. The parabolic QSAR equation, log BCF = 3.036 ClogP − 0.197 ClogP^2 − 0.808 MgVol (n = 28, r^2 = 0.817, q^2 = 0.761, s = 0.558) (experimental log BCF range = 0.44–5.29, ClogP range = 3.16–11.27), suggests that a non-linear relationship between BCF and the hydrophobic parameter, along with inclusion of additional molecular size, weight and/or volume parameters, should be considered while developing a QSAR model for more reliable prediction of the BCF of highly hydrophobic chemicals.

1. Introduction

Globally, regulatory agencies are developing methods and criteria for hazard and risk assessment of chemicals (ASTM, 1993; ECETOC, 1995; ECHA, 2012; OECD, 2007). Bioaccumulation and bioconcentration refer to the process of uptake and buildup of chemicals in living organisms. The bioaccumulation factor (BAF) parameter is used as a measure of a chemical's bioaccumulation potential. If the BAF value of a chemical is not available, its BCF is used to assess the bioaccumulation potential. Experimental measurement of BAF and BCF values is time-consuming and expensive, and is not feasible for a large number of chemicals of regulatory concern. Therefore, attention is focused on estimation of these values by using quantitative structure–activity relationship (QSAR) models. QSAR models are used as screening tools to assess the effect of a large number of chemicals on the environment and human health. These models establish empirical relationships between the molecular parameters (physico-chemical properties) of the organic chemicals and physiological responses in the organism. Based on a large number of QSAR studies, it has been noted that the dataset of chemicals should exhibit a wide range in their biological activities and parameter values for developing a robust QSAR model (Hansch and Leo, 1995; Hansch et al., 1995).

The most common method for estimating BCF value consists of developing QSAR models establishing correlations between BCF and hydrophobicity of a chemical as measured by the logarithm of the octanol–water partition coefficient (denoted by log$P$ or logK_{ow}). In regulatory context, the objective is to use parameters which are easy to calculate and compare (such as log$P$) and develop simple models which could be used to predict the most accurate BCF value (ASTM, 1993; ECETOC, 1995; ECHA, 2012; Mackay and Fraser, 2002; OECD, 2007).

Several QSAR models have been proposed for predicting the BCF of organic chemicals, which use a linear, parabolic, bilinear or polynomial relationship, extensively reviewed in Arnot and Gobas (2006), Devillers et al. (1998), Müller and Nendza (2009), and Pavan et al. (2006). Most of the QSAR models reported for the prediction of BCF within a regulatory context are based on the correlation of log BCF with logK_{ow}. For a chemical, the mechanistic basis underlying the relationship of BCF with logK_{ow} is the analogy

* Corresponding author. Address: 5500 Campanile Drive, San Diego, CA 92182-7720, USA. Tel.: +1 (858) 472 9132.
E-mail address: rg004747@yahoo.com (R. Garg).

http://dx.doi.org/10.1016/j.fct.2014.03.035
0278-6915/

between the partitioning process between a biological lipid membrane and water, and the partitioning process between n-octanol and water (Arnot and Gobas, 2003, 2006; Bintein et al., 1993; Dearden, 2004; Dearden and Hewitt, 2010; Devillers et al., 1998; Dimitrov et al., 2002; Jonker and van der Haijden, 2007; Kubinyi, 1976; Müller and Nendza, 2009; Pavan et al., 2006; USEPA, 2012; Veith et al., 1979). Few QSAR models have been reported using other experimentally derived parameters such as water solubility (S), and soil adsorption coefficients (Kenaga and Goring, 1980). However, their applicability is limited due to the problem of data availability. To avoid new tests, theoretical molecular descriptors (such as topological, connectivity indices, quantum, and other descriptors) have been used for developing BCF prediction models. Assessment of their ability to correctly predict the BCF value of a chemical resulted in a large number of incorrect classifications (Pavan et al., 2006).

Arnot and Gobas (2003) proposed a mechanistic QSAR model for predicting the BCF and BAF of organic chemicals in aquatic food webs. This model uses the $\log K_{ow}$ and a number of correction factors, but does not consider a chemical's molecular weight and size-related parameters. EPI Suite software from United States Environmental Protection Agency (USEPA) (2012) uses the BCFBAF program, based on the Arnot–Gobas model, to predict the BCF and BAF values of a chemical. The BCFBAF program uses two linear QSAR equations for predicting the BCF of a chemical. The first equation with a positive linear hydrophobic term indicates that the log BCF increases linearly with $\log K_{ow}$ values for $\log K_{ow} \leqslant 7.0$, while the second equation with a negative linear hydrophobic term shows a decreasing linear relationship for values of $\log K_{ow} > 7.0$. According to this model the decrease in BCF with increasing $\log K_{ow}$ (>7.0) for highly hydrophobic chemicals is mainly due to adsorption of chemical in the water phase and not due to biomagnification or steric factors affecting membrane permeability (Arnot and Gobas, 2003). In another study, Bintein et al. (1993) reported a comparative analysis of linear, parabolic and bilinear QSAR models to explain the nonlinear dependence of fish bioconcentration on $\log P$. These models indicate that the linear relationship between log BCF and hydrophobicity is unable to explain the low BCF of highly hydrophobic chemicals. The authors (Bintein et al., 1993) concluded that the parabolic model, and preferably the bilinear model (Kubinyi, 1976), is more useful.

The European Chemical Agencies (ECHA) guidance document indicates that the log BCF increases linearly with $\log K_{ow}$ values <5 and a decreasing linear relationship is observed for higher values of $\log K_{ow}$. It is noted that apart from experimental errors in the determination of BCF values for these very hydrophobic chemicals, reduced uptake due to the increasing molecular size may also be responsible for this relationship (ECHA, 2012). Dimitrov et al. (2002) established that the relationship between log BCF and $\log K_{ow}$ for highly hydrophobic chemicals can be explained by including the molecular size parameter in the QSAR model. The ECHA guidance document also suggests that the molecular weight parameter, even though not directly related to the molecular size of a compound, together with other information can be used to assess a chemical's bioaccumulation potential (ECHA, 2012). However, no experimental data have been reported to support a specific threshold for the molecular weight parameter.

To predict the BCF values of highly hydrophobic chemicals, we have derived a new QSAR model using a dataset of diverse organic chemicals whose experimental BCF values were measured in a USEPA laboratory (Veith et al., 1979). The developed model is validated using cross validation, Tropsha's metrics, r_m^2 metrics, y-randomization test, and applicability domain analysis. This new model is discussed below and also compared with other QSAR models reported in the literature.

2. Materials and methods

2.1. Selection of dataset

Experimental log BCF values of 29 chemicals used in this study are taken from Veith et al. (1979) (see Table 1). This study on a diverse group of organic chemicals tested for bioconcentration in fathead minnow (*Pimephales promelas*) was conducted at a USEPA Environmental Research Laboratory. This is a good dataset for QSAR study as it includes a diverse group of organic chemicals including halogenated, nonhalogenated, and phosphate containing chemicals displaying a wide range in the parameter values (experimental log BCF range = 0.44–5.29, ClogP range = 3.16–11.27). Earlier models based on this dataset are used as an example for BCF prediction in the European Union Technical Guidance Document on risk assessment (Pavan et al., 2006). Out of 55 chemicals for which the BCF data were reported (Veith et al., 1979), only 30 chemicals were tested at the USEPA laboratory and the others were taken from different sources. We have used the BCF data of chemicals tested in the USEPA laboratory. One chemical 'toluene diamine', out of these 30 chemicals, is not included in our study due to uncertainty as to its structure.

2.2. Calculation of molecular parameters

The log$P$ values listed in Table 1 are taken from Veith et al. (1979) and are provided here for comparison. They were estimated by the reverse phase HPLC method (Veith and Morris, 1978). The ClogP and MgVol parameter values are calculated and auto loaded from the C-QSAR Program (2006). The utility of the C-QSAR program in comparative correlation analysis has been discussed in Hansch and Leo (1995). Within chemical families of structural congeners, biological activity is well predicted from a chemical structure by the C-QSAR program. The parameters used in this report have been discussed in detail along with their applications in Hansch and Leo (1995). Briefly, ClogP is the calculated log$P$ and is a measure of hydrophobicity of a chemical (Leo et al., 1971; Leo, 1993), and MgVol is the molar volume calculated by the method of (Abraham, 1993; Zhao et al., 2003). Note that the ClogP values are for the neutral form of acids and bases that may be partially ionized. If the degree of ionization is about the same for a set of congeners, the ionization factor can be neglected; otherwise, good correlation can be obtained using electronic terms (Leo et al., 1971; Leo, 1993).

The correlation matrix for the parameters used in this study is given in Table 2. The correlation between experimental log$P$ and ClogP values for 13,815 compounds in the CLOG program, which is a part of the C-QSAR Program (2006), is 0.98 (experimental log$P$ = 1.00 ClogP − 0.03 (n = 13,815, r = 0.98, s = 0.35). Many programs are used for calculating octanol–water partition coefficients and are reviewed in Mannhold et al. (2009). However, we have used the ClogP parameter in this study as it has been widely used and cited by the QSAR community, both for environmental studies and drug design (Arnot and Gobas, 2006; Devillers et al., 1998; Garg et al., 1999; Hansch et al., 1989; Leo and Hansch, 1999; Müller and Nendza, 2009; Selassie et al., 2003; Smith et al., 2002, 2003, 2004, 2006), and a very high correlation (r = 0.98) between experimental log$P$ and ClogP gives confidence in using ClogP values whenever experimental log$P$ values are not available.

The QSAR stepwise multiple linear regression (MLR) analyses are executed with the C-QSAR program and all the parameters are auto loaded (C-QSAR, 2006). In all the QSAR equations reported in this report, n is the number of data points, r is the correlation coefficient, s is the standard deviation, and q^2 is the quality of fit of the data, calculated using Cramer et al.'s (1988) approach, which approaches the value of r^2 as the quality of fit improves.

3. Results and discussion

First, we developed a QSAR model for the whole dataset using stepwise MLR analysis. Next, we divided the whole dataset into a training set and a test set and performed internal and external validation studies. Cross validation techniques were utilized for internal validation, and the model developed using training set was used to predict the activity of test set chemicals. Tropsha's and r_m^2 metrics were also calculated to evaluate the internal and external predictive abilities of the QSAR model. To ensure the developed QSAR model is robust and not derived due to chance, the y-randomization test was performed. Lastly, the applicability domain of the developed QSAR model was evaluated to ascertain the reliability of the model.

3.1. Model development

Stepwise MLR analysis on whole dataset reported by Veith et al. (1979) (Table 1) resulted in Eqs. (1)–(3).

Please cite this article in press as: Garg, R., Smith, C.J. Predicting the bioconcentration factor of highly hydrophobic organic chemicals. Food Chem. Toxicol. (2014), http://dx.doi.org/10.1016/j.fct.2014.03.035