Ecotoxicology and Environmental Safety 135 (2017) 130–136

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

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

Exploring MIA-QSPR's for the modeling of biomagnification factors of aromatic organochlorine pollutants

Estella G. da Mota, Mariene H. Duarte, Stephen J. Barigye, Teodorico C. Ramalho, Matheus P. Freitas *

Department of Chemistry, Federal University of Lavras, P.O. Box 3037, 37200-000 Lavras, MG, Brazil

ARTICLE INFO

Article history:
Received 18 April 2016
Received in revised form
28 September 2016
Accepted 29 September 2016

Keywords:
MIA-QSPR
Organochlorine pollutants
Biomagnification
MLR
Validation

ABSTRACT

Biomagnification of organic pollutants in food webs has been usually associated to hydrophobicity and other molecular descriptors. However, direct information on atoms and substituent positions in a molecular scaffold that most affect this biological property is not straightforward using traditional QSPR techniques. This work reports the QSPR modeling of biomagnification factors (logBMF) of a series of aromatic organochlorine compounds using three MIA-QSPR (multivariate image analysis applied to QSPR) approaches. The MIA-QSPR model based on augmented molecular images (described with atoms represented as circles with sizes proportional to the respective van der Waals radii and having colors numerically proportional to the Pauling's electronegativity) encoded better the logBMF data. The average results for the main statistical parameters used to attest the model's predictability were r^2=0.85, q^2=0.72 and r^2_{test}=0.85. In addition, chemical insights on substituents and respective positions at the biphenyl rings A and B, and dibenzo-*p*-dioxin and dibenzofuran motifs are given to aid the design of more ecofriendly derivatives.

1. Introduction

Organochlorine compounds are omnipresent pollutants in the environment and because of their high lipophilicity (they are stored in adipose tissues) and persistency, they tend to accumulate in the food chain. The toxicity of this class of compounds comes from their structural difference if compared to naturally occurring substances and, therefore, some contaminated organisms are not capable of metabolizing them, causing accumulation (Baird and Cann, 2012).

Polychlorinated biphenyls (PCBs) and dichloro diphenyl trichloroethane (DDT) are some examples of organochlorine compounds with capacity to bioaccumulate and produce harmful effects in ecosystems. Biomagnification refers to a progressive accumulation of substances from a trophic level to another along the food chain. Because of this phenomenon, the concentration of such micropollutants in the environment has increased at rates higher than their removal (such as degradation); studies have detected the presence of these compounds and the respective metabolites in several matrices, as a result of their accumulation in living organisms (Font and Marsal, 1988; Bisson and Hontela, 2002). The toxicology of PCBs is affected by the number and position of the chlorine atoms, as substitution in the ortho position hinders the rotation of the rings (PCBs without ortho substitution are referred to as coplanar and the others are noncoplanar) (Newman, 2015). Such structures bind to the aryl hydrocarbon receptor (*Ah*R) and may thus exert dioxin-like effects, namely impairment of the immune system, the developing nervous system, the endocrine system and reproductive functions (Hahn, 1998). The analysis of the effect of structural modification (*e.g.* substituent types and positions) on a given compound property (*e.g.* biomagnification) is within the field of Quantitative Structure-Property Relationships (QSPR).

Most QSPR studies for modeling environmental properties, such as soil sorption, bioacummulation and biomagnification, are based on octanol/water partition coefficients (logP) (Mackay et al., 1997), due to the hydrophobic properties of the living tissues where substances accumulate. Crowding of chlorine substituents, as well as specific substitution patterns, play an important role in partition of PCBs between water and octanol (Sabljic, 2001). Other physicochemical descriptors (Todeschini and Consonni, 2000) also provide valuable information on the molecular properties affecting the biomagnification in a general sense, but the inherent drawback of such analyses lies in the vague notion on the group types and/or molecular positions that most affect the biomagnification.

Thus, this work reports the modeling of biomagnification factors (BMF) of a series of aromatic organochlorine pollutants using three MIA-QSPR approaches. The MIA-QSPR (multivariate image analysis

Abbreviations: aug, augmented; MIA, Multivariate image analysis; QSPR, Quantitative Structure-Property Relationships; BMF, biomagnification; PCB, polychlorinated biphenyls

* Corresponding author.
E-mail address: matheus@dqi.ufla.br (M.P. Freitas).

http://dx.doi.org/10.1016/j.ecoenv.2016.09.030
0147-6513/

applied to QSPR) method is known for a decade as a QSPR technique capable of recognizing two-dimensional chemical structures and encoding atomic, stereochemical and connectivity properties using 2D projections of molecular images (in terms of pixels) as descriptors (Freitas et al., 2005; Barigye and Freitas, 2016). Consequently, particular structural features and/or positions responsible for enhanced or attenuated BMF of aromatic organochlorine compounds in living organisms can be rationalized, also contributing to driving the synthesis of more ecofriendly compounds.

2. Materials and methods

A series of aromatic organochlorine compounds with logBMF values experimentally available was obtained from the literature (Fatemi and Baher, 2009) (Table 1). The original chemical analysis reported by Henny et al. (2003) for these compounds was performed on osprey egg and whole fish composite samples that were collected from the Willamette River, USA. The data set molecules were drawn using either the ACD/ChemSketch program (2009) (for the traditional MIA-QSPR model) or the GaussView program (Dennington et al., 2008) (for the augmented MIA-QSPR models). For the aug-MIA-QSPR models, atoms were represented as spheres with sizes proportional to the van der Waals radii and colored differently to distinguish them, since different numbers are assigned to each color pixel (from 0 – black, to 765 – white), consistent with the RGB (red-green-blue) system of colors. For the aug-MIA-QSPR$_{color}$ model, the pixel values were numerically proportional to Pauling's electronegativity, in order to encode electrostatic interactions possibly ruling the biomagnification factors. The congruent chemical substructures were overlaid for 2D alignment purposes, in such a way that only variable motifs explain the variance in the y block. Each image was saved as bitmaps and converted to a numerical $x \times y$ matrix. Subsequently, the n images (compounds) were grouped to form a three-way array $n \times x \times y$, which was unfolded to an $n \times (x \times y)$ matrix. This matrix was divided into training (80% of compounds) and test (20% of compounds) set compounds. Five splits were performed to test the model's robustness. Because of the large data matrix obtained from this procedure (thousands columns), a search for the best 5 variable model for the logBMF was performed using the Multiple Linear Regression method coupled with the Genetic Algorithm (MLR-GA). Preliminary unsupervised feature selection based on Shannon's entropy (variables with less than 10% of entropy were discarded) and the variable correlation coefficients (X/X=0.98) was performed. The built QSPR models were validated using leave-one-out cross-validation (LOOCV) and external validation procedures. Other measures considered in the assessment of the quality of the built model include: the determination coefficient between actual and predicted logBMF (q^2 and r^2_{test}), root mean square error of prediction (RMSECV and RMSEP) and the modified r^2_{test} ($r_m{}^2$) parameter, according to the criteria established in the literature (Roy et al., 2013). In addition, the reliability of the model was attested using the y-randomization test [analyzed in terms of the corrected penalized r^2 ($^cr_p{}^2$)] (Mitra et al., 2010), in which the y-block is shuffled and regression performed to verify the inexistence of chance correlation. The image treatment and statistical analysis were performed using the Chemoface program (Nunes et al., 2012).

Table 1
Compounds used in the QSAR modeling and respective logBMF values.[a]

Cpd number	Name	Notation	logBMF$_{exp}$
1	2378TCDF	TCDF	−0.12
2	hexachlorobenzene	HCB	0.32
3	3,3,4,4-Tetrachlorobiphenyl	PCB77	0.77
4	2,4,4,5-Tetrachlorobiphenyl	PCB74	0.83
5	2,3,4,4-Tetrachlorobiphenyl	PCB60	0.90
6	2,2,3,4,5,6-Hexachlorobiphenyl	PCB149	0.95
7	2,2,3,3,4,5,6-Heptachlorobiphenyl	PCB174	1.00
8	1,2,3,4,6,7,8,9-Octachlorodibenzofuran	OCDF	1.00
9	2,3,3,4,6-Pentachlorobiphenyl	PCB110	1.04
10	2,2,4,4,5-Pentachlorobiphenyl	PCB99	1.11
11	2,2,4,5,5,-Pentachlorobiphenyl	PCB101	1.25
12	2,3,7,8-tetrachlorodibenzo-*p*-dioxin	TCDD	1.25
13	2,3,4,4,5-Pentachlorobiphenyl	PCB118	1.30
14	3,3,4,4,5,5-Hexachlorobiphenyl	PCB169	1.32
15	2,3,3,4,4-Pentachlorobiphenyl	PCB105	1.36
16	2,2,3,3,4,4,6-Heptachlorobiphenyl	PCB171	1.36
17	2,2,3,4,5,5,-Hexachlorobiphenyl	PCB141	1.43
18	2,2,3,4,4,5,6-Heptachlorobiphenyl	PCB183	1.43
19	2,2,3,3,4,4,5,5-Octachlorobiphenyl	PCB194	1.43
20	2,2,3,4,4,5,5,6-Octachlorobiphenyl	PCB203	1.43
21	3,3,4,4,5-Pentachlorobiphenyl	PCB126	1.43
22	2,2,3,4,4,5-Hexachlorobiphenyl	PCB138	1.46
23	2,2,4,4,5,5-Hexachlorobiphenyl	PCB153	1.46
24	2,2,3,4,5,5-Hexachlorobiphenyl	PCB146	1.48
25	2,2,3,3,4,5,6,6-Octachlorobiphenyl	PCB201	1.48
26	2,2,3,3,4,5,6,6-Octachlorobiphenyl	PCB200	1.50
27	2,2,3,3,4,5,5-Heptachlorobiphenyl	PCB172	1.53
28	2,2,3,4,4,5,5-Heptachlorobiphenyl	PCB180	1.53
29	1,1-Dichloro-2,2-(4-ClC_6H_4)ethane	*p,p*-DDD	1.61
30	Dichlorodiphenyltrichloroethane	DDT	1.92
31	1,1-Dichloro-2,2-(4-ClC_6H_4)ethene	*p,p*-DDE	2.19
32	1,2,3,6,7,8-Hexachlorodibenzo-*p*-Dioxin	H6CDD	2.44
33	1,2,3,4,6,7,8-Heptachlorodibenzo-*p*-Dioxin	H7CDD	2.44
34	1,2,3,4,6,7,8,9-Octachlorodibenzo-*p*-Dioxin	OCDD	2.49
35	2,3,4,4-Tetrachlorobiphenyl	PCB66	0.83
36	2,2,3,5,6-Pentachlorobiphenyl	PCB95	0.83
37	2,2,3,3,4,4,5-Heptachlorobiphenyl	PCB170	1.53
38	2,3,3,4,4,5,6-Heptachlorobiphenyl	PCB190	1.53
39	2,2,3,4,4,5,6-Heptachlorobiphenyl	PCB182	1.39
40	2,2,3,4,5,5,6-Heptachlorobiphenyl	PCB187	1.39

[a] The chemical structures are given in the Supplementary material.

3. Results and discussion

The predictive ability of the MIA-QSPR models for the 40 aromatic organochlorine compounds of Table 1 was evaluated using three approaches: 1) traditional MIA-QSPR, in which descriptors correspond to black and white pixels and chemical structures are represented as wireframes; 2) aug-MIA-QSPR, in which descriptors correspond to pixels colored according to the GaussView default for each atom (circles with sizes proportional to the van der Waals radii); 3) aug-MIA-QSPR$_{color}$, whose chemical structures are identical to the aug-MIA-QSPR model, but atom colors numerically proportional the corresponding electronegativity values. Fig. 1 shows the overlaid images representing these three models.

From the complete data matrix of thousands descriptors for each approach (MIA-QSPR, aug-MIA-QSPR and aug-MIA-QSPR$_{color}$), only five independent variables were selected for further regression against the logBMF values using multiple linear regression (MLR). Five QSPR models were built for each approach, differing by the test set compounds used for external validation, whose results are shown in Tables 2–4.

On the basis of the mean values for the statistical parameters of each model, particularly those related to external validation, which is considered the only way to establish a reliable QSPR model (Golbraikh and Tropsha, 2002), we found that models based on augmented images are more predictive than traditional MIA-QSPR. Moreover, the method that includes pixel colors proportional to the atomic electronegativity (aug-MIA-QSPR$_{color}$) showed to be slightly better. This small difference between the models obtained using aug-MIA descriptors indicates that steric (hydrophobicity) rather than electrostatic (encoded by the atoms electronegativity) effects are more effective to explain the biomagnification property of