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

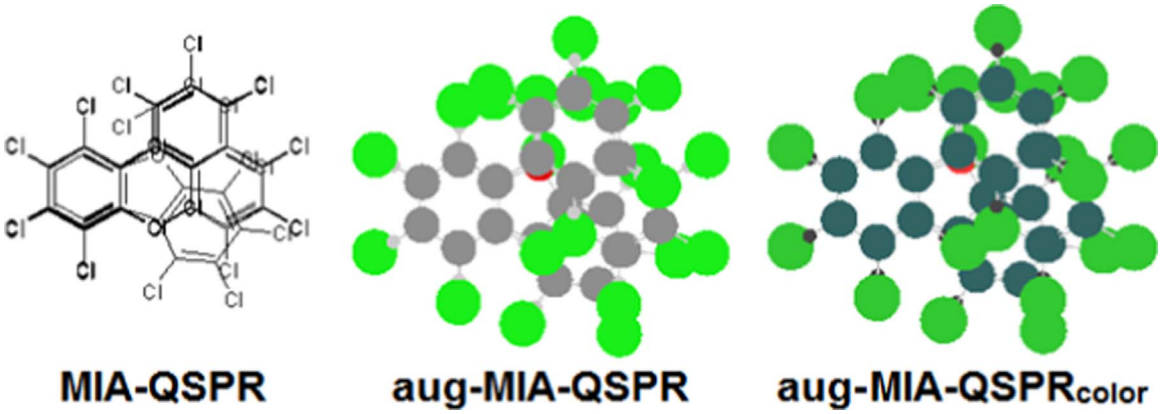


Fig. 1. Superposed images used for the building of the MIA-QSPR models (ring A is used for 2D alignment).

Table 2
Statistical parameters for the traditional MIA-QSPR model.[a]

Parameter	Test set 1	Test set 2	Test set 3	Test set 4	Test set 5	Average
RMSEcv	0.2453	0.2571	0.2022	0.2460	0.2653	0.2432
r^2cv	0.7730	0.7534	0.8391	0.7697	0.6917	0.7654
RMSEc	0.2112	0.2253	0.1807	0.2148	0.2256	0.2115
r^2cal	0.8306	0.8095	0.8711	0.8236	0.7746	0.8219
RMSE y-rand	0.4587	0.4560	0.4579	0.4664	0.4347	0.4547
r^2 Yrand	0.1945	0.2145	0.1693	0.1627	0.1611	0.1804
r^2p(*y*-rand)	0.6625	0.6244	0.7297	0.6696	0.6067	0.6586
RMSEp	0.2227	0.1650	0.3149	0.2191	0.1638	0.2171
r^2pred	0.7785	0.8677	0.6217	0.7864	0.9296	0.7968
r^2m(test)	0.5654	0.7949	0.3527	0.6398	0.7804	0.6266

[a] Test set 1: **3**; **6**; **11**; **16**; **19**; **24**; **32**; **37**. Test set 2: **4**; **7**; **12**; **20**; **25**; **33**; **38**; **39**. Test set 3: **2**; **5**; **10**; **15**; **18**; **23**; **28**; **31**. Test set 4: **8**; **13**; **21**; **26**; **29**; **34**; **35**; **40**. Test set 5: **1**; **9**; **14**; **17**; **21**; **27**; **30**; **36**.

Table 3
Statistical parameters for the aug-MIA-QSPR model.[a]

Parameter	Test set 1	Test set 2	Test set 3	Test set 4	Test set 5	Average
RMSEcv	0.236	0.278	0.230	1.287	0.209	0.4479
r^2cv	0.797	0.745	0.800	0.462	0.808	0.7222
RMSEc	0.198	0.199	0.192	0.194	0.185	0.1935
r^2cal	0.851	0.852	0.855	0.855	0.848	0.8524
RMSE y-rand	0.459	0.472	0.460	0.473	0.448	0.4625
r^2 Yrand	0.196	0.161	0.162	0.143	0.109	0.1541
r^2p(*y*-rand)	0.689	0.708	0.712	0.722	0.730	0.7122
RMSEp	0.199	0.197	0.224	0.239	0.288	0.2295
r^2pred	0.822	0.820	0.927	0.748	0.873	0.8381
r^2m(test)	0.688	0.756	0.540	0.451	0.784	0.6437

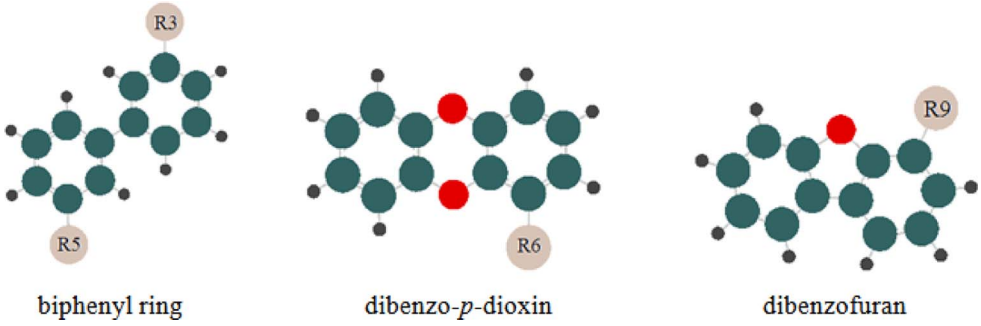
[a] Test set 1: **3**; **6**; **11**; **16**; **19**; **24**; **32**; **37**. Test set 2: **4**; **7**; **12**; **20**; **25**; **33**; **38**; **39**. Test set 3: **2**; **5**; **10**; **15**; **18**; **23**; **28**; **31**. Test set 4: **8**; **13**; **21**; **26**; **29**; **34**; **35**; **40**. Test set 5: **1**; **9**; **14**; **17**; **21**; **27**; **30**; **36**.

Table 4
Statistical parameters for the aug-MIA-QSPR$_{color}$ model.[a]

Parameter	Test set 1	Test set 2	Test set 3	Test set 4	Test set 5	Average
RMSEcv	0.2292	0.2561	0.6281	0.2498	0.2862	0.3299
r^2cv	0.8009	0.7657	0.5884	0.7657	0.6894	0.7220
RMSEc	0.1879	0.1909	0.1959	0.2006	0.1991	0.1949
r^2cal	0.8659	0.8631	0.8485	0.8461	0.8245	0.8496
RMSE y-rand	0.4628	0.4767	0.4712	0.4866	0.4362	0.4667
r^2 Yrand	0.1828	0.1450	0.1227	0.0934	0.1546	0.1397
r^2p(*y*-rand)	0.7157	0.7314	0.7228	0.7341	0.6748	0.7157
RMSEp	0.2310	0.2261	0.2048	0.1957	0.2061	0.2127
r^2pred	0.8451	0.7808	0.8719	0.8759	0.8978	0.8543
r^2m(test)	0.3529	0.6835	0.7957	0.6778	0.8796	0.6779

[a] Test set 1: **3**; **6**; **11**; **16**; **19**; **24**; **32**; **37**. Test set 2: **4**; **7**; **12**; **20**; **25**; **33**; **38**; **39**. Test set 3: **2**; **5**; **10**; **15**; **18**; **23**; **28**; **31**. Test set 4: **8**; **13**; **21**; **26**; **29**; **34**; **35**; **40**. Test set 5: **1**; **9**; **14**; **17**; **21**; **27**; **30**; **36**.

Table 5
Matrix of selected aug-MIA-QSPR$_{color}$ descriptors.



Cpd number	X1885 (*ca.* R5)	X2209 (*ca.* R5)	X9854 (*ca.* R3)	X11819 (*ca.* R6)	X13248 (*ca.* R9)
1	210	210	765	250	765
2	300	612	765	765	765
3	765	210	765	765	612
4	300	765	765	765	612
5	765	210	765	765	612
6	765	210	765	765	612
7	300	765	250	765	612
8	300	765	765	250	300
9	765	210	250	765	612
10	300	765	250	765	612
11	300	765	250	765	612
12	765	210	765	210	250
13	300	765	250	765	612
14	300	765	250	765	612
15	765	210	250	765	612
16	765	210	765	765	612
17	300	612	765	765	612
18	765	210	765	765	612
19	300	612	765	765	612
20	300	612	765	765	612
21	300	612	765	765	612
22	765	210	765	765	612
23	300	612	765	765	612
24	300	612	765	765	612
25	765	210	765	765	612
26	300	612	765	765	612
27	300	612	765	765	612
28	300	612	765	765	612
29	765	210	300	210	250
30	765	210	300	210	250
31	765	210	300	210	250
32	765	210	765	300	250
33	300	612	765	300	250
34	300	612	765	300	250
35	765	210	765	765	612
36	765	210	765	765	612
37	300	612	765	765	612
38	300	612	765	765	612
39	300	612	765	765	612
40	300	612	765	765	612

Table 6
Correlation matrix (expressed in terms of the determination coefficient) among aug-MIA$_{color}$ descriptors.

	X1885	X2209	X9854	X11819	X13248
X1885	1	0.81	0.01	0.03	0.08
X2209		1	0.00	0.07	0.03
X9854			1	0.01	0.02
X11819				1	0.70
X13248					1

polychlorinated aromatic compounds. Considering the range in logBMF along the series of *ca.* 2.6 and an error in calibration and external validation of *ca.* 0.2 ($< 10\%$), all the models can predict accurately the biomagnification of aromatic organochlorine derivatives. Special emphasis will be given further to the aug-MIA-QSPR$_{color}$ model, which was found to be predictive, robust and not prone to chance correlation, as attested by the statistical parameters depicted in Tables 2–4, as explained in the earlier section.

Thus, on the basis of the selected descriptors of Table 5, it is possible to analyze which substituents and positions most affect the logBMF values. The selected aug-MIA$_{color}$ descriptors are mostly uncorrelated (Table 6), while those presenting some redundancy (X1885 × X2209) lie on the region surrounding a same substituent (R5). Thus, either X1885 or X2209 could be removed from the data matrix, but this procedure reduced significantly the prediction ability of the model. Moreover, the presence of both descriptors could be advantageous to capture complimentary information on how substitution at the corresponding position affects the biomagnification factor. According to the

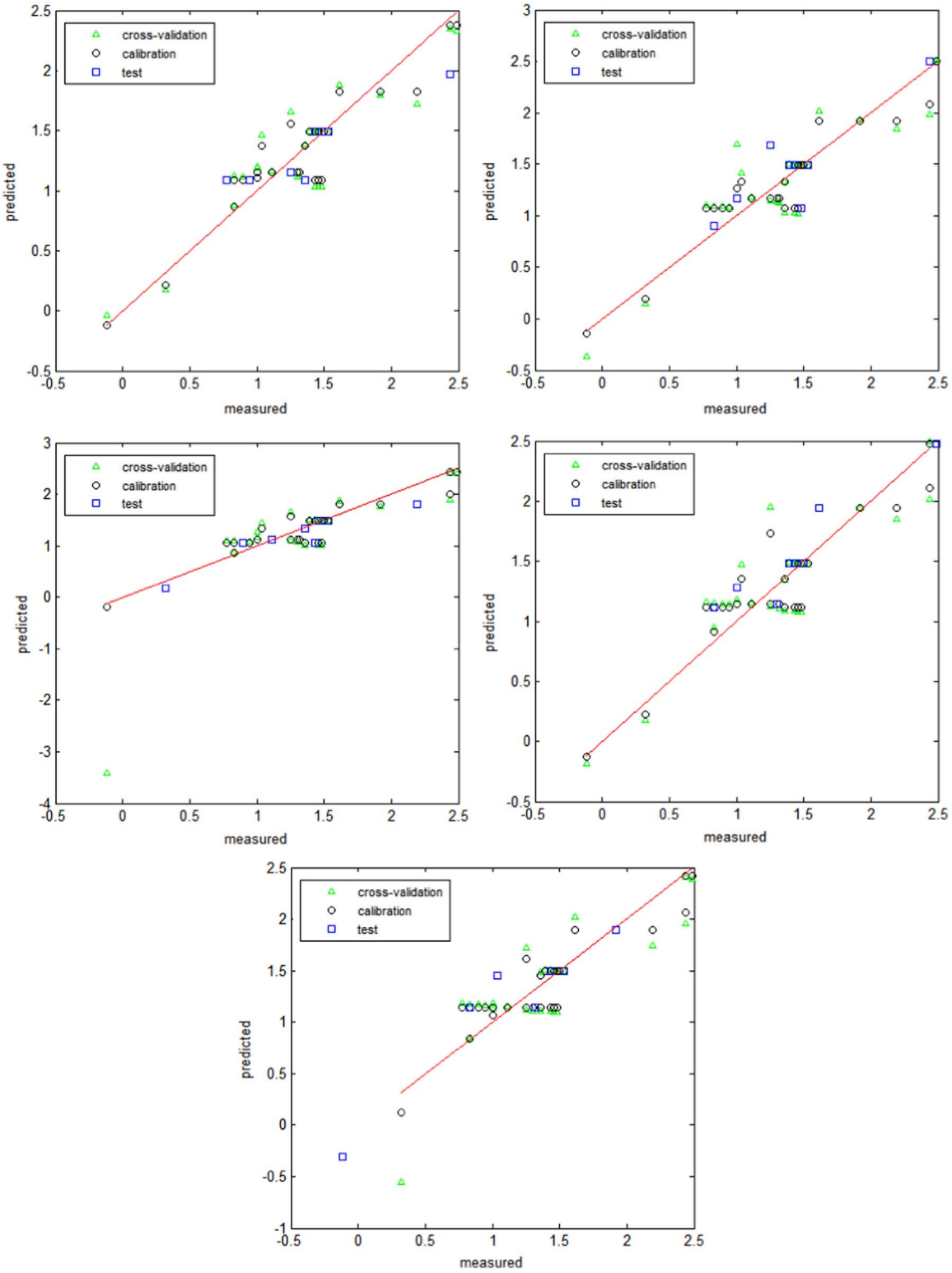


Fig. 2. Plots of experimental *vs.* predicted BMF for the aug-MIA-QSPR$_{color}$ models. Note that the model for Eq. 3 exhibits an anomalous behavior for compound 1 in the leave-one-out cross-validation, causing decrease in q^2.

most predictive aug-MIA-QSPR$_{color}$ model (see plots for experimental *versus* predicted logBMF values in Fig. 2), the substituent position surrounding the selected descriptors are: R5 (biphenyl ring A): X1885; R5 (biphenyl ring A): X2209; R5 (biphenyl ring B): X9854; R6 (dibenzo-*p*-dioxin): X11819; R9 (dibenzofuran): X13248.

The following MLR equations corresponding to the aug-MIA-QSPR$_{color}$ models give information on the weight of each selected descriptor and, therefore, of each substituent on the logBMF values. Consequently, it is possible to anticipate whether the absence of atoms at a given pixel/variable position (due to a long chemical bond or a small atom, giving rise to a blank region in the image and, consequently, yielding pixels 765 – white) or the presence of chlorine atoms (pixel 300, or long chemical bonds C-Cl with pixel value 612) contribute either to increase or decrease the logBMF

values. Most of the coefficients in the MLR equations are negative and, therefore, the occupation of coordinates X1885, X2209, X9854 and X13248 by a chlorine atom, instead of hydrogens or absence of atoms, lead to an increase of logBMF values. On the other hand, the presence of hydrogen (pixel 210) instead of chlorine at position X11819 is favorable, since the coefficient of this variable is positive in all five models. Consequently, considering prospective aromatic organochlorine compounds to be synthesized, we suggest avoiding the introduction of chlorine atoms particularly at positions R5 of the biphenyl ring A, R3 of the biphenyl ring B, R6 of the dibenzo-*p*-dioxin group, and R9 of the dibenzofuran motif. Considering the symmetry of biphenyls and that the crowding of chlorine substituents in PCBs increases properties related to biomagnification (Sabljic, 2001), the other meta-sites similar to R3 and R5 should also affect the logBMF values, but they were removed during the variable selection step probably because of the high correspondence with R3 and R5.

Equation 1 (model obtained from the training set without samples of test set 1):

$$\log BMF = 7.3172 - 0.0044 \times \mathbf{X1885} - 0.0041 \times \mathbf{X2209} - 0.0006 \times \mathbf{X9854} + 0.0046 \times \mathbf{X11819} - 0.0083 \times \mathbf{X13248}.$$

Equation 2 (model obtained from the training set without samples of test set 2):

$$\log BMF = 7.3135 - 0.0042 \times \mathbf{X1885} - 0.0039 \times \mathbf{X2209} - 0.0005 \times \mathbf{X9854} + 0.0044 \times \mathbf{X11819} - 0.0084 \times \mathbf{X13248}.$$

Equation 3 (model obtained from the training set without samples of test set 3):

$$\log BMF = 7.4719 - 0.0045 \times \mathbf{X1885} - 0.0042 \times \mathbf{X2209} - 0.0005 \times \mathbf{X9854} + 0.0047 \times \mathbf{X11819} - 0.0087 \times \mathbf{X13248}.$$

Equation 4 (model obtained from the training set without samples of test set 4):

$$\log BMF = 7.0724 - 0.0040 \times \mathbf{X1885} - 0.0037 \times \mathbf{X2209} - 0.0005 \times \mathbf{X9854} + 0.0043 \times \mathbf{X11819} - 0.0082 \times \mathbf{X13248}.$$

Equation 5 (model obtained from the training set without samples of test set 5):

$$\log BMF = 7.5851 - 0.0045 \times \mathbf{X1885} - 0.0043 \times \mathbf{X2209} - 0.0006 \times \mathbf{X9854} + 0.0050 \times \mathbf{X11819} - 0.0089 \times \mathbf{X13248}.$$

The mechanism of the biomagnification process is still unresolved, but there is clear evidence that, in aquatic organisms, it is related to bioconcentration and, consequently, to the propensity of hydrophobic organic chemicals to be absorbed through the respiratory and dermal surfaces Gobas and Morrison, 2000). Thus, hydrophobicity should play a decisive role for the biomagnification mechanism. Gobas et al. (1999) hypothesized a fugacity based model of the gastrointestinal absorption of contaminants (PCBs), suggesting that in food chains, lipid rich prey-items will not only result in a larger exposure of the predator to bioaccumulative substances, but also result in larger gastrointestinal magnification factors which can lead to higher bioaccumulation factors in the predator. In line with this, we have found that hydrophobic chlorine atoms bound to biphenyl rings, dibenzo-*p*-dioxin groups and dibenzofurans increase the logBMF values of organochlorine pollutants and, most importantly, that specific positions of chlorine atoms affect the biomagnification more than others, giving additional chemical insights for detailing the biomagnification mechanism. This result is consistent with previous reports in the literature (International Program on Chemical Safety, World Health Organization, 2003), and thus demonstrating the theoretical and practical contribution of the aug-MIA-QSAR method in BMF modeling.

Data reporting the effect of the position of chlorine atoms on the biomagnification factors of polycyclic aromatic hydrocarbons are scarce and, therefore, an straightforward analysis based on substituent positions has not been clear yet. QSPRs on similar biological factors mostly mention the important role of the amount of substituents, *e.g.* the study by de Melo (2012), in which the bioconcentration factors of polychlorinated biphenyls in fishes were found to increase with the number of chlorine substituents. One of the few studies reporting the influence of chlorine substitution pattern on biological properties shows that degradation of polychlorinated biphenyls depends on the bacterial strain (Bedard and Haberl, 1990). A systematic evaluation of some PCBs has shown that subtle structural changes in chlorine substitution pattern affected the aryl hydrocarbon hydroxylase (AHH) induction potency in rat: 3,3′,4,4′-tetra-, 3,3′,4,4′,5-penta- and 3,3′,4,4′,5,5′-hexachlorobiphenyls $>$ 3,4,4′,5-tetrachlorobiphenyl $\sim$ mono-*ortho* coplanar PCBs $>$ di-*ortho* coplanar PCBs (Safe et al., 1985). With respect to biomagnification, *ortho*-substituted PCB congeners with no unsubstituted *meta-para* positions have shown high biomagnification potential, while PCBs with low biomagnification have adjacent vicinal hydrogens, indicating that congeners with this feature may have been metabolically eliminated (Andersson et al., 2001). The present work contributes for a systematic insight of the influence of chlorine substitution pattern on the biomagnification factors of important classes of polychlorinated aromatic hydrocarbons. As an example of application and chemical interpretation from our findings, the selected R2 and R3 substituents are not at *ortho* position and, therefore, a coplanar structure (characteristic of dioxins and furans) is expected for biphenyls with high biomagnification factors. Other structural insights as mentioned earlier can provide useful information for the design of less toxic compounds.

4. Conclusion

All three MIA-QSPR models tested can be considered predictive, but aug-MIA-QSPR$_{color}$ descriptors offer advantages in terms of prediction, validation (including reliability) and interpretability of the model, since pixels are colored according to a chemically comprehensible property (electronegativity). Thus, logBMF for different aromatic organochlorine analogues can be predicted from the insights of this work and, in addition, the synthesis of prospective, more ecofriendly derivatives can be driven on the basis of the outcomes of the present QSPR modeling. That is, considering the outcomes of the aug-MIA-QSPR$_{color}$ models, it is environmentally desirable to avoid chlorine atoms at X1885, X2209, X9854, X11819 and X13248 pixel coordinates, corresponding to R5 of the biphenyl ring A, R3 of the biphenyl ring B, R6 of the dibenzo-*p*-dioxin group, and R9 of the dibenzofuran motif.

Acknowledgement

The authors are thankful to Fundação de Amparo à Pesquisa do Estado de Minas Gerais - FAPEMIG for the financial support and for a fellowship (to E.G.M.), as well as to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES (to M.H.D.) for the studentship and Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (to S.J.B., T.C.R. and M.P.F.) for fellowships. This work is a collaboration research project of members of the Rede Mineira de Química (RQ-MG) supported by FAPEMIG (Project: CEX - RED-00010-14).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ecoenv.2016.09.030.

References

ACD/ChemSketch Version 12.01, 2009. Advanced Chemistry Development, Inc, Toronto.

Andersson, P.L., Berg, A.H., Bjerselius, R., Norrgren, L., Olsén, H., Olsson, P.E., Orn, S., Tysklind, M., 2001. Bioaccumulation of selected PCBs in zebrafish, three-spined stickleback, and arctic char after three different routes of exposure. Arch. Environ. Contam. Toxicol. 40, 519–530.

Baird, C., Cann, M., 2012. Environmental Chemistry. W. H. Freeman and Company, New York.
Barigye, S.J., Freitas, M.P., 2016. Ten years of the MIA-QSAR strategy: historical developments and applications. Int. J. Quant. Struct. Prop. Relat. (1), 62–75.
Bedard, D.L., Haberl, M.L., 1990. Influence of chlorine substitution pattern on the degradation of polychlorinated biphenyls by eight bacterial strains. Microb. Ecol. 20, 87–102.
Bisson, A., Hontela, A., 2002. Cytotoxic and endocrine-disruspting potential of atrazine, diazinon, endosulfan, and mancozeb in adrenocortical steroidogenic cells of rainbow trout exposed in vitro. Toxicol. App. Pharmacol. 180, 110–117.
Dennington II, R.D., Keith, T.A., Millam, J.M., 2008. GaussView 5. Gaussian, Inc, Wallingford, p. 0.
Fatemi, M.H., Baher, E., 2009. A novel quantitative structure–activity relationship model for prediction of biomagnification factor of some organochlorine pollutants. Mol. Divers. 13, 343–352.
Font, J., Marsal, A., 1988. Determination of organochlorine pesticides in skins and leather by gas chromatography. J. Chromatogr. A 811, 256–260.
Freitas, M.P., Brown, S.D., Martins, J.A., 2005. MIA-QSAR: a 2D-image based approach for quantitative-structure activity relationship analysis. J. Mol. Struct. 738, 149–154.
Gobas, F.A.P.C., Morrison, H.A., 2000. Bioconcentration and biomagnification in the aquatic environment. In: Boethling, R.S., Mackay, D. (Eds.), Handbook of Property Estimation Methods for Chemicals: Environmental and Health Sciences. Lewis, Boca Raton, pp. 189–231.
Gobas, F.A.P.C., Wilcockson, J.B., Russell, R.W., Haffner, G.D., 1999. Mechanism of biomagnification in fish under laboratory and field conditions. Environ. Sci. Technol. 33, 133–141.
Golbraikh, A., Tropsha, A., 2002. Beware of q^2!. J. Mol. Graph. Model. 20, 269–276.
Hahn, M.E., 1998. The Aryl hydrocarbon receptor: a comparative perspective. Comp. Biochem. Physiol. C 121, 23–53.
Henny, C.J., Kaiser, J.L., Grove, R.A., Bentley, V.R., Elliott, J.E., 2003. Biomagnification factors (fish to Osprey eggs from Willamette River, Oregon, U.S.A.) for PCDDs, PCDFs, PCBs and OC pesticides. Environ. Monit. Assess. 84, 275–315.
International Programme on Chemical Safety, World Health Organization, 2003. Polychlorinated Biphenyls: Human Health Aspects. Concise International Chemical Assessment Document 55.
Mackay, D., Shiu, W.-Y., Ma, K.-C., 1997. Illustrated Handbook of Physical-Chemical Properties and Environmental fate for Organic Chemicals. Lewis Publishers, New York.
de Melo, E.B., 2012. A new quantitative structure–property relationship model to predict bioconcentration factors of polychlorinated biphenyls (PCBs) in fishes using E-state index and topological descriptors. Ecotoxicol. Environ. Saf. 75, 213–222.
Mitra, I., Saha, A., Roy, K., 2010. Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. Mol. Simul. 36, 1067–1079.
Newman, M.C., 2015. Fundamentals of Ecotoxicology: The Science of Pollution. CRC Press, Boca Raton.
Nunes, C.A., Freitas, M.P., Pinheiro, A.C.M., Bastos, S.C., 2012. Chemoface: a novel free user-friendly interface for chemometrics. J. Braz. Chem. Soc. 23, 2003–2010.
Roy, K., Chakraborty, P., Mitra, I., Ojha, P.K., Kar, S., Das, R.N., 2013. Some case studies on application of r_m^2 metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. J. Comput. Chem. 34, 1071–1082.
Sabljic, A., 2001. QSAR models for estimating properties of persistent organic pollutants required in evaluation of their environmental fate and risk. Chemosphere 43, 363–375.
Safe, S., Bandiera, S., Sawyer, T., Robertson, L., Safe, L., Parkinson, A., Thomas, P.E., Ryan, D.E., Reik, L.M., Levin, W., Denomme, M.A., Fujita, T., 1985. PCBs: structure-function relationships and mechanism of action. Environ. Health Perspect. 60, 47–56.
Todeschini, R., Consonni, V., 2000. Handbook of Molecular Descriptors. Wiley-VCH, Verlag.