SAR and QSAR in Environmental Research

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gsar20

QSAR studies on the depuration rates of polycyclic aromatic hydrocarbons, polybrominated diphenyl ethers and polychlorinated biphenyls in mussels (Elliptio complanata)

F. Li [a], X. Liu [a b], L. Zhang [a b], L. You [a b], H. Wu [a], X. Li [c], J. Zhao [a] & J. Yu [a]

[a] Key Laboratory of Coastal Zone Environment Processes, CAS; Shandong Provincial Key Laboratory of Coastal Zone Environment Processes, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, P.R. China

[b] The Graduate School of Chinese Academy of Sciences, Beijing, China

[c] Key Laboratory of Industrial Ecology and Environmental Engineering (MOE), Department of Environmental Science and Technology, Dalian University of Technology, Dalian, China

Available online: 07 Jul 2011

To cite this article: F. Li, X. Liu, L. Zhang, L. You, H. Wu, X. Li, J. Zhao & J. Yu (2011): QSAR studies on the depuration rates of polycyclic aromatic hydrocarbons, polybrominated diphenyl ethers and polychlorinated biphenyls in mussels (Elliptio complanata), SAR and QSAR in Environmental Research, 22:5-6, 561-573

To link to this article: http://dx.doi.org/10.1080/1062936X.2011.569947

SAR and QSAR in Environmental Research
Vol. 22, Nos. 5–6, July–September 2011, 561–573

QSAR studies on the depuration rates of polycyclic aromatic hydrocarbons, polybrominated diphenyl ethers and polychlorinated biphenyls in mussels (*Elliptio complanata*)

F. Li[a], X. Liu[a,b], L. Zhang[a,b], L. You[a,b], H. Wu[a]*, X. Li[c], J. Zhao[a] and J. Yu[a]

[a]*Key Laboratory of Coastal Zone Environment Processes, CAS; Shandong Provincial Key Laboratory of Coastal Zone Environment Processes, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, P.R. China;* [b]*The Graduate School of Chinese Academy of Sciences, Beijing, China;* [c]*Key Laboratory of Industrial Ecology and Environmental Engineering (MOE), Department of Environmental Science and Technology, Dalian University of Technology, Dalian, China*

(*Received 21 December 2010; in final form 23 February 2011*)

Based on the mechanism of action, a quantitative structure–activity relationship (QSAR) model for the depuration rate constants (k_d) of 28 PAHs, 8 PBDEs and 28 PCBs in mussels (*Elliptio complanata*) was constructed by partial least squares (PLS) regression, following the guidelines for development and validation of QSAR models. For the training set of the QSAR model, $r^2 = 0.953$, the cross-validated regression coefficient (Q^2_{CUM}) was 0.947. The predicted $\log k_d$ values for the validation set were consistent with the observed values, with a standard error (*SE*) of 0.160 log units and a squared correlation coefficient (Q^2_{EXT}) of 0.892. Comparatively, the developed QSAR model had good robustness, predictive ability and extended applicability domain. The electrophilicity index (ω), molecular polarizability (α), the averages of the negative potentials on the molecular surface ($\bar{V}_s^-$) and the balance parameter of surface potential (τ) were the key parameters governing the $\log k_d$ values in the QSAR model, which indicated that the $\log k_d$ value was mainly related to the partition ability, electrostatic interactions, and van der Waals interactions of compounds.

Keywords: depuration rate; *Elliptio complanata*; PAHs; PBDEs; PCBs; quantitative structure–activity relationship (QSAR)

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs), polybrominated diphenyl ethers (PBDEs) and polychlorinated biphenyls (PCBs) are persistent organic pollutants (POPs) that can enter water bodies and eventually sink into the sediment through various transportation routes [1,2]. These POPs have been of great concern due to their elevated concentrations and wide distribution; they pose not only an environmental risk but also a human health risk through accumulation in human tissues and fluids [1–4].

Mussels accumulate POPs very efficiently, and are widely used as bio-indicators in the aquatic environment due to their wide geographical distribution, tolerance to various environmental conditions and availability throughout the year [5]. Mussels can be used to determine the time required to achieve chemical equilibrium, which can be calculated from

*Corresponding author. Email: hfwu@yic.ac.cn

ISSN 1062–936X print/ISSN 1029–046X online

DOI: 10.1080/1062936X.2011.569947
http://www.informaworld.com

the depuration rate constants (k_d) of chemicals in organisms [6]. The k_d is one of the most important kinetic parameters, which can be used to estimate bioconcentration factors (BCF) and the time to steady state [7–9]. Hence, it is of great importance to explore the depuration rate constants of hazardous compounds.

The guidelines for the experimental determination of k_d have been documented [10,11]; however, the experimental data are challenged by the reality that few empirical data are available for the thousands of commercial substances that require evaluation. Therefore an alternative approach, such as quantitative structure–activity relationship (QSAR) modelling, was proposed for chemical safety assessment by the new EU chemicals legislation REACH [12]. QSARs can fill the data gap of organic pollutants, decrease experimental expenses and, in particular, reduce animal testing [13]. They have been widely used in research on the acute toxicity [14], mixture toxicity [15], endocrine disrupting activities [16,17] and photoinduced toxicity [18,19] of organic compounds.

Several QSAR models for the k_d of compounds in mussels have been constructed using octanol/water partition coefficients (K_{OW}) [10,20] or quantum chemical parameters [6,21], which can be used to explore the inherent relations between molecular structures of chemicals and their k_d values. Previous studies have constructed QSAR models individually on PAHs [6] or PCBs [10,21], but no QSAR models have been developed on PBDEs or systematically on heterogeneous series of chemicals including PAHs, PCBs and PBDEs. In addition, the molecular structures of the compounds (planar or non-planar) may make contributions to the k_d. According to the guidelines for QSARs development and validation proposed by the Organization for Economic Co-operation and Development (OECD), QSARs for regulatory purposes should be associated with the following information: a defined endpoint; an unambiguous algorithm; a defined domain of applicability; appropriate measures of goodness-of-fit, robustness and predictive power; and a mechanistic interpretation, if possible [22]. Thus further QSAR development should follow these guidelines.

In this study, following the OECD guidelines [22], we attempted to develop a QSAR model, in freshwater mussels (*Elliptio complanata*), for the k_d of three categories of POPs (28 PAHs, eight PBDEs and 28 PCBs) with distinctive chemical structures by partial least squares (PLS) regression for the first time [23]. Based on the mechanism of action, appropriate molecular structural parameters computed by density function theory were adopted to construct QSAR models. In addition, the performance of the developed QSAR model was evaluated by external validation, and the critical molecular structural features related to the depuration rate were discussed.

2. Materials and methods

2.1 *Data compilation and chemical domain*

The k_d values of the three categories of POPs (28 PAHs, eight PBDEs and 28 PCBs) in mussels (*Elliptio complanata*) were taken from previous studies [10,11,20] and then converted into the form of $\log k_d$ values that ranged from -2.301 to -0.577 log unit.

2.2 *Mechanism consideration and molecular structural parameters selection*

As proposed by the OECD guidelines [22], QSAR models should be developed based on the mechanism of action. Drouillard et al. [20] and Baumard et al. [24] have reported that

the depuration rate is related to the equilibrium distribution interactions. The process may involve the following interactions: the partition ability in the bio-phase, dipolar–dipolar interactions, the hydrogen bond or electrostatic interactions. It has been generally accepted that equilibrium partitioning is the major factor determining the uptake and release rates of some lipophilic pollutions in gill-breathing aquatic animals [25]. Thus, a total of 13 theoretical molecular structural parameters were selected to characterize these interactions.

The molecular volume (V) and the average molecular polarizability (α) were selected to describe the cavity-forming interactions. Other parameters such as the energy of the highest occupied molecular orbital (E_{HOMO}), the energy of the lowest unoccupied molecular orbital (E_{LUMO}), the most positive hydrogen atom in the molecule ($q\mathrm{H}^+$), the most negative formal charge in the molecule (q^-), electrophilicity index (ω), the most positive and most negative values of the molecular surface potential ($V_{\mathrm{s,max}}$, $V_{\mathrm{s,min}}$), the averages of the positive and negative potentials on the molecular surface ($\bar{V}_{\mathrm{s}}^+$, $\bar{V}_{\mathrm{s}}^-$), the average deviation of surface potential (Π) and the balance parameter of surface potential (τ) were purposely selected to describe the hydrogen bond or electrostatic interactions.

The molecular volume V is defined as the volume inside a contour of 0.001 electrons/Bohr^3 density. The average molecular polarizability α is calculated as [18]:

$$\alpha = (\alpha_{xx} + \alpha_{yy} + \alpha_{zz})/3 \tag{1}$$

where α_{xx}, α_{yy} and α_{zz} are the diagonal elements in the standard orientation of molecular polarizability tensor.

ω was calculated as follows:

$$\mu = \frac{E_{\mathrm{LUMO}} + E_{\mathrm{HOMO}}}{2} \tag{2}$$

$$\eta = \frac{E_{\mathrm{LUMO}} - E_{\mathrm{HOMO}}}{2} \tag{3}$$

$$\omega = \frac{\mu^2}{2\eta} \tag{4}$$

where μ is the chemical potential, η is the chemical hardness.

Quantum chemical parameters such as E_{HOMO}, E_{LUMO}, $q\mathrm{H}^+$ and q^- were proved successful in many QSAR studies for characterizing intermolecular electrostatic interactions [26]. The electrophilicity index ω measures the ability of a compound to accept electrons. Roy et al. reported that the toxicity of aliphatic chemical compounds on *Tetrahymena pyriformis* correlated strongly with ω [27].

The potential derived parameters were calculated by the following equations [28]:

$$\bar{V}_S^+ = \frac{1}{\alpha}\sum_{i=1}^{\alpha} V_S^+(r_i) \tag{5}$$

$$\bar{V}_S^- = \frac{1}{\beta}\sum_{j=1}^{\beta} V_S^-(r_j) \tag{6}$$

$$\Pi = \frac{1}{\alpha+\beta}\sum_{i=1}^{\alpha+\beta}\left|V(r_i) - \bar{V}_s\right| = \frac{1}{\alpha+\beta}\sum_{i=1}^{\alpha+\beta}\left|V(r_i) - \frac{\alpha\bar{V}_s^+ + \beta\bar{V}_s^-}{\alpha+\beta}\right| \tag{7}$$

$$\tau = \frac{\sigma_+^2\sigma_-^2}{(\sigma_{\text{tot}}^2)^2} \tag{8}$$

$$\sigma_{\text{tot}}^2 = \sigma_+^2 + \sigma_-^2 = \frac{1}{\alpha}\sum_{i=1}^{\alpha}[V^+(r_i) - \bar{V}_s^+]^2 + \frac{1}{\beta}\sum_{j=1}^{\beta}[V^-(r_j) - \bar{V}_s^-]^2 \tag{9}$$

where s stands for molecular surface; α and β are the number of the points for the positive and negative potentials, respectively; $V^+(r_i)$ and $V^-(r_j)$ are the positive and negative potentials on the molecular surface, respectively; $\bar{V}_s^+$ and $\bar{V}_s^-$ are the averages of the positive and negative potentials on the molecular surface; σ_+^2, σ_-^2 and σ_{tot}^2 are the variance of values for the positive, negative and total surface potentials.

The molecular surface potential indicates the charge distribution in a molecule [29], which is a gauge of the basicity and nucleophilicity of a molecule [30]. These potential-derived parameters have been successfully used to rationalize the toxicities of chlorinated diphenyls [31]. The details of equations are listed in the supplementary materials available via the Supplementary Content tab on the article's online page at http://dx.doi.org/.

All the quantum chemical descriptors were computed by Gaussian 09 programs [32]. Initial geometries of the compounds were optimized by semi-empirical method PM3, then optimized at the hybrid Hartree–Fock density functional theory B3LYP/6-31G(d, p) level [33]. Solvent effects (water) were taken into consideration implicitly, including the integral equation formulation of the polarized continuum model [34]. A frequency analysis was performed on the optimized geometries to ensure that the systems had no imaginary vibration frequencies. The potentials on the molecular surface were calculated on the 0.001 electrons/bohr^3 surface of chemicals. Values for all the molecular structural parameters are listed in Table S1 of the supplementary materials (available via the Supplementary Content tab on the article's online page at http://dx.doi.org/10.1080/1062936X.2011.569947).

2.3 *QSAR development and validation*

The original data set was divided into a training set (80%) and a validation set (20%), as listed in Table 1. Partial least squares (PLS) regression was conducted for the model development as PLS is able to analyse data with strongly collinear, noisy and numerous predictor variables [23]. The software of Simca-S (Version 6.0, Umetri AB & Erisoft AB) was employed for the PLS analysis with leave-many-out cross-validation to determine the number of PLS components (A). Cross-validation simulates how well a model predicts new data, and gives a statistical Q_{CUM}^2 (the fraction of the total variation of the dependent variables that can be predicted by all the extracted components) for the final model. The PLS analysis was performed repeatedly so as to eliminate redundant molecular structural parameters, as done in previous studies [35].

Table 1. Logarithm of the observed and predicted depuration rate constants (log k_d) of the considered compounds.

	log k_d				log k_d		
Compounds[a]	*Observed*	*Predicted*	*Residuals*	*Compounds*[a]	*Observed*	*Predicted*	*Residuals*
Naphthalene	−0.654	−0.673	0.019	BDE-99	−2.000	−1.896	−0.104
2-Methylnaphthalene*	−0.686	−0.418	−0.269	BDE-153	−2.222	−2.190	−0.032
1-Methylnaphthalene	−0.604	−0.438	−0.166	BDE-138	−2.046	−2.029	−0.017
Biphenyl	−0.672	−0.745	0.073	BDE-190	−1.886	−2.062	0.176
2,6-Dimethylnaphthylene	−0.577	−0.527	−0.050	PCB-19*	−1.081	−1.151	0.070
Acenaphthylene	−0.734	−0.569	−0.165	PCB-22	−1.149	−1.215	0.066
Dibenzofuran	−0.635	−0.691	0.056	PCB-42	−1.347	−1.478	0.131
Acenaphthene*	−0.625	−0.417	−0.208	PCB-74	−1.553	−1.478	−0.075
2,3,5-Trimethylnaphthalene	−0.746	−0.642	−0.104	PCB-66	−1.509	−1.582	0.074
Fluorene	−0.721	−0.621	−0.100	PCB-95	−1.538	−1.689	0.152
1-Methylfluorene*	−0.903	−0.777	−0.126	PCB-91*	−1.553	−1.740	0.188
Dibenzothiophene	−0.793	−0.862	0.069	PCB-92	−1.678	−1.809	0.132
Phenanthrene	−0.768	−0.793	0.025	PCB-99	−1.658	−1.859	0.201
Anthracene	−0.747	−0.654	−0.093	PCB-97*	−1.638	−1.626	−0.012
1-Methylphenanthrene*	−0.858	−0.860	0.002	PCB-87	−1.602	−1.656	0.054
Fluoranthrene	−0.901	−0.964	0.063	PCB-85	−1.721	−1.700	−0.022
Pyrene	−0.786	−0.995	0.208	PCB-110	−1.602	−1.707	0.105
Benz[*a*]anthracene	−1.034	−1.083	0.049	PCB-118*	−1.854	−1.738	−0.116
Chrysene	−1.078	−1.202	0.124	PCB-105	−1.824	−1.738	−0.086
Benzo[*b*]fluoranthene	−1.082	−1.092	0.009	PCB-136	−1.745	−1.798	0.053
Benzo[*k*]fluoranthene*	−1.230	−1.184	−0.046	PCB-151	−2.000	−2.005	0.005
Benzo[*e*]pyrene	−1.138	−0.895	−0.243	PCB-149	−1.921	−1.903	−0.018
Benzo[*a*]pyrene	−1.122	−1.116	−0.006	PCB-134	−2.097	−2.062	−0.035
Perylene*	−1.376	−1.139	−0.237	PCB-146*	−2.000	−2.040	0.040
Indeno[123*cd*]pyrene	−1.327	−1.335	0.008	PCB-141	−2.155	−2.040	−0.115
Dibenz[*ah*]anthracene	−1.163	−1.349	0.186	PCB-130*	−2.155	−1.868	−0.287
Benzo[*ghi*]perylene	−1.223	−1.263	0.040	PCB-137	−2.222	−1.950	−0.272
Coronene	−1.300	−1.442	0.142	PCB-138	−2.155	−1.950	−0.205
BDE-28	−1.432	−1.494	0.062	PCB-128	−1.721	−1.803	0.082
BDE-75	−1.585	−1.647	0.062	PCB-156	−2.046	−1.906	−0.140
BDE-47*	−1.721	−1.648	−0.074	PCB-179	−2.301	−2.030	−0.271
BDE-100	−2.097	−2.049	−0.048	PCB-178	−2.222	−2.163	−0.059

[a]Compounds marked with asterisked (*) were selected to form the external validation set.

The model predictability was evaluated by external validation. The performance of external validation was characterized by the determination coefficient (r^2), standard error (SE) and external explained variance (Q^2_{EXT}), which are defined as below [36]:

$$r^2 = 1 - \sum_{i=1}^{n} (y_i^{\mathrm{fit}} - y_i)^2 \Big/ \sum_{i=1}^{n} (y_i - \overline{y})^2 \tag{10}$$

$$SE = \sqrt{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / n - 1} \tag{11}$$

$$Q^2_{\mathrm{EXT}} = 1 - \sum_{i=1}^{n_{\mathrm{EXT}}} (y_i - \hat{y}_i)^2 \Big/ \sum_{i=1}^{n_{\mathrm{EXT}}} (y_i - \bar{y}_{\mathrm{EXT}})^2 \tag{12}$$

where y_i^{fit} is the fitted $\log k_{\text{d}}$ value of the ith compound, $\bar{y}$ is the average response value in the training set, y_i and $\hat{y}_i$ are the observed and predicted values for the ith compound, respectively. $\bar{y}_{\text{EXT}}$ is the average response value of the validation set, n stands for the number of compounds in the training set, and n_{EXT} stands for the number of compounds in the validation set.

The applicability domain of the developed QSAR model was assessed by the Williams plot, i.e., the plot of standardized residuals (σ) versus leverage (Hat diagonal) values (h_i) [37]. The h_i value of a chemical in the original variable space and the warning leverage value (h^*) are defined as:

$$h_i = x_i^T (X^{\text{T}} X)^{-1} x_i \ (i = 1, \ldots, n) \tag{13}$$

$$h^* = 3(p+1)/n \tag{14}$$

where x_i is the descriptor vector of the considered compound and X is the model matrix derived from the training set descriptor values, and p is the number of predictor variables.

When the h value of a compound is lower than h^*, the probability of accordance between predicted and actual values is as high as that for the compounds in the training set. A chemical with $h_i > h^*$ will reinforce the model if the chemical is in the training set. Such a chemical in the validation set implies that it is structurally distant from compounds in the training set, and its predicted data may be unreliable. However, this chemical may not appear to be an outlier because its residuals may be low. Thus the leverage and the standardized residual should be combined for the characterization of the applicability domain.

3. Results and discussion

3.1 *Development and validation of the QSAR model for log k_d*

PLS analysis with the $\log k_{\text{d}}$ as the dependent variable and the molecular structural parameters as predictor variables resulted in the following optimal QSAR model:

$$\log k_{\text{d}} = 1.34 \times 10^{-1} - 1.07 \times 10^{3}\omega - 4.86 \times 10^{-3}\alpha - 7.65 \times 10\bar{V}_s^- - 4.31 \times 10^{-2}\tau$$

$$n(\text{training set}) = 51, \quad A = 2, \quad r^2 = 0.952, \quad Q^2_{\text{CUM}} = 0.947, \quad SE = 0.119,$$

$$n(\text{validation set}) = 13, \quad Q^2_{\text{EXT}} = 0.892, \quad SE = 0.160, \quad p < 0.0001$$

where p is the significance level.

The predicted $\log k_{\text{d}}$ values and residuals for compounds are listed in Table 1. The r^2 value of the QSAR model was 0.952, indicating a good goodness-of-fit of the model. Q^2_{CUM} of the QSAR is as high as 0.947, implying good robustness of the model. The differences between r^2 and Q^2_{CUM} (0.005) did not exceed 0.3, indicating no over-fitting in the model [38]. As show in Figure 1, the predicted $\log k_{\text{d}}$ values were consistent with the observed values for both the validation and training sets. The model revealed acceptable predictability with $Q^2_{\text{EXT}} = 0.892$, $SE = 0.160$. In summary, the developed QSAR model shows satisfactory performance.

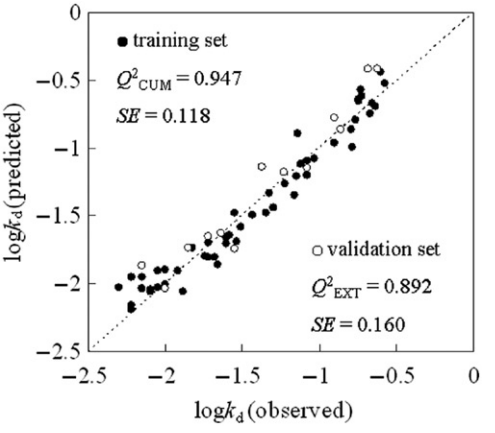

Figure 1. Plot of observed *versus* predicted $\log k_d$ values for the training and validation. Q^2_{CUM} is the leave-many-out cross-validation squared correlation coefficient for the training set; Q^2_{EXT} is the squared correlation coefficient for the validation set and *RMSE* is the root mean square error.

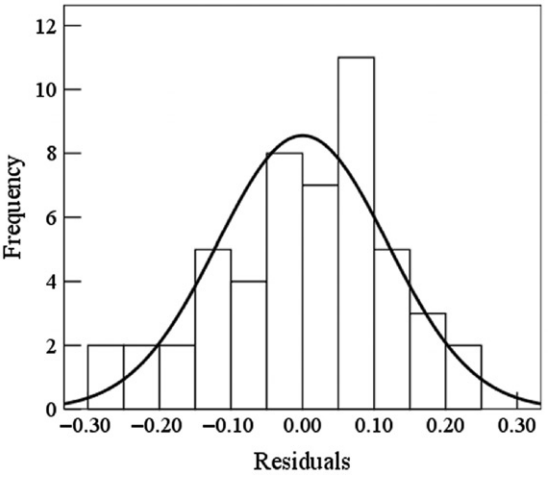

Figure 2. Distribution of the residuals for $\log k_d$ values. *SD* is the standard deviation.

3.2 *Applicability domain of the developed QSAR model*

The distribution of residuals is shown in Figure 2. Application of the Kolmogorov–Smirnov test for normality (at the 95% confidence level) confirms that the distribution of residuals is a distinctive bell-shaped pattern associated with a normal distribution (mean = 0.00, standard deviation = 0.12), which implies that the residuals are non-systematic and the applicability domain of the developed QSAR model can be visualized by the Williams plot.

As shown in the Williams plot (Figure 3), h_i values of all the compounds in the training and validation sets were lower than the warning value ($h^* = 0.294$), and all the compounds

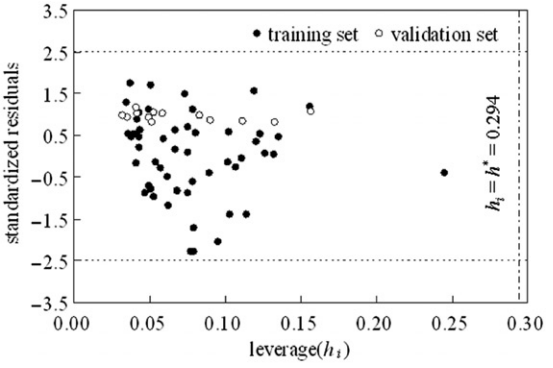

Figure 3. Plot of standardized residuals *versus* leverages. Dash lines represent ±2.5 standardized residual, dotted line represents warning leverage ($h^* = 0.294$).

Table 2. *VIP* values and PLS weights for the optimal PLS model.

	VIP	*w*c [1]*	*w*c [2]*
ω	1.253	−0.664	−0.045
$\bar{V}_s^-$	1.225	−0.648	−0.027
α	0.821	−0.299	−0.967
τ	0.506	−0.224	0.358

in both the training and validation sets were in the domain. None of the compounds were particularly influential in the model space and the training set was of great representativeness. For all the compounds in the training and validation sets, their standardized residuals were smaller than 2.5 standard deviation units (2.5σ); there were no outliers for the developed QSAR model. Thus, the developed QSAR model can be used to predict the $\log k_d$ of PAHs, PBDEs and PCBs in mussels (*Elliptio complanata*).

3.3 *Mechanistic implications of the developed QSAR model*

The developed PLS model extracted two PLS components which were loaded primarily upon four predictor variables. Values of the variable importance in the projection (*VIP*) and PLS weights (*w**) are listed in Table 2. The *w** values can be used to estimate how the predictor variables and the response variables combine in the projections (PLS components), and how they relate to each other [23].

The first PLS component was loaded primarily on the two descriptors, ω and $\bar{V}_s^-$ (Table 2). As ω and $\bar{V}_s^-$ are related to the electrophilicity of compounds, the PLS component mainly condenses information on the electrostatic interactions. The ω and $\bar{V}_s^-$ parameters remarkably govern the $\log k_d$ values, as indicated by their large *VIP* values among the predictor variables. Their negative PLS weights and coefficients in the current QSAR model indicated that the negative correlation between the predictor variables and

$\log k_d$. ω measures the ability of a compound to accept electrons [39]. Compounds with smaller ω values tend to depurate with difficulty. Hence, in the developed QSAR model, the $\log k_d$ increased with the decreasing ω values.

$\bar{V}_s^-$ stands for the average of the negative potentials on the molecular surface. As shown in the QSAR model, $\bar{V}_s^-$ negatively correlated with $\log k_d$. The molecules with lower $\bar{V}_s^-$ values tend to have stronger electrostatic interactions, and accordingly depurate with difficulty. For example, 1-methylnaphthalene and 2-methylnaphthalene were of the same ω value in this work. However, 1-methylnaphthalene ($\bar{V}_s^- = -0.0126$) showed a bigger $\bar{V}_s^-$ value than 2-methylnaphthalene ($\bar{V}_s^- = -0.0131$), which contributed to the bigger $\log k_d$ value for 1-methylnaphthalene.

The second PLS component was loaded primarily on α and τ. As α is correlated with $\log K_{OW}$ positively [40], compounds with higher $\log K_{OW}$ values tend to partition in the bio-phase easily, hence the $\log k_d$ values decrease with α values. The balance parameter τ is introduced as an indicator of the degree to which a molecule can interact approximately equally through both its positive and negative regions (whether that is strongly or weakly). Politzer et al. reported that the more similar were the magnitudes of σ_+^2 and σ_-^2, the higher was the value of τ, reaching a maximum value (0.25) when $\sigma_+^2 = \sigma_-^2$ [29]. In the developed QSAR model, the $\log k_d$ decreased with the increasing τ values, which indicated that the depuration rates of compounds were associated with a high level of internal charge separation within the molecules.

3.4 *Statistical performance compared with literature models*

In Table 3, the current QSAR model was compared with four published QSAR models for the depuration rates in mussels (*Elliptio complanata*) [6,10,11,20]. The first three QSAR models employed the unambiguous statistical algorithm of multivariate linear regression (MLR) and single parameter ($\log K_{OW}$) to predict the $\log k_d$ values, which showed comparable performances. O'Rourke et al. suggested that hydrophobicity was the main factor governing the depuration rates, which implied that the PCB compounds with stronger hydrophobicity depurated with more difficulty [10]. However, it was difficult to describe the molecular multi-dimension information by the single parameter ($\log K_{OW}$) [41]. Moreover, external validation was not mentioned in the three MLR models. Wu et al. employed PLS regression and three predictor variables to develop a QSAR model that was simple and reproducible for regulators and non-QSAR experts [6]. However, all the four mentioned QSAR models were constructed on the homologous series of chemicals such as PAHs and PCBs, and the applicability domains were not discussed [6,10,11,20]. Comparatively, the developed QSAR model in this study exhibited good robustness and predictive ability. Moreover, the applicability domain of the model has been extended to three categories of compounds (PAHs, PBDEs and PCBs).

4. Conclusion

Based on the mechanism of action, a QSAR model has been developed to characterize the depuration rates of PAHs, PBDEs and PCBs in freshwater mussels (*Elliptio complanata*) following the OECD guidelines. Electrostatic interactions and hydrophobicity were important factors for the depuration rates. The molecules with lower ω and $\bar{V}_s^-$, such as benz[*a*]anthracene, benzo[*a*]pyrene and coronene, tend to depurate with difficulty.

Table 3. Comparison with current QSAR models.

			Goodness-of-fit, robustness and predictivity[a]								
			Training set				*Validation set*				
No.	*Compounds*	*Algorithm*[a]	*n*	*K*	r^2	Q^2_{LMO}	*m*	Q^2_{EXT}	*SE*	*AD*[b]	*References*
1	PCBs	MLR	34	1	0.80	NM	NM	NM	NM	N	[10]
2	PBDEs	MLR	9	1	0.73	NM	NM	NM	NM	N	[20]
3	PAHs	MLR	46	1	0.83	NM	NM	NM	NM	N	[11]
4	PAHs	PLS	18	3	0.94	0.93	9	**0.66**	**0.14**	N	[6]
5	PAHs, PBDEs and PCBs	PLS	51	4	0.95	0.95	13	0.89	0.16	Y	This study

Note. The bold-faced values were not listed in the references and calculated for comparison using the supplementary data.
[a]*n* and *m* are the numbers of compounds in the training and validation sets, respectively; *K* is the number of molecular descriptors; r^2 is the squared correlation coefficient between observed and predicted values; Q^2_{LMO} is the leave-many-out cross-validation squared correlation coefficient for the training set, Q^2_{EXT} is squared correlation coefficient for the validation set. 'NM' means it was not mentioned in the reference. [b]Y and N denote the model is assessed with or without applicability domain (AD) discussion, respectively.

The partition ability of the molecules into bio-phase is also a significant parameter factor related to the depuration rates. In addition, compounds with planar structure (such as PAHs) tend to depurate faster. The developed QSAR model has shown good robustness, predictive ability and mechanism interpretability, which could be potentially applied to predict the depuration rates of other PAHs, PBDEs and PCBs.

Acknowledgements

This research was supported by The 100 Talents Program of the Chinese Academy of Sciences, the Key Laboratory of Industrial Ecology and Environmental Engineering, China Ministry of Education, Technology Development Program Projects of Shandong Province (2008GG20005006 and 2008GG3NS0700) and SDSFC (ZR2009CZ008) and in part by CAS innovation key project (KZCX2-YW-Q07-04), the CAS/SAFEA International Partnership Program for Creative Research Teams ''Representative environmental processes and resources effects in coastal zone''.

References

[1] M. Athanasiadou, S.N. Cuadra, G. Marsh, A. Bergman, and K. Jakobsson, *Polybrominated diphenyl ethers (PBDEs) and bioaccumulative hydroxylated PBDE metabolites in young humans from Managua, Nicaragua*, Environ. Health Perspect. 116 (2008), pp. 400–408.

[2] S. Khan, Q. Cao, A.J. Lin, and Y.G. Zhu, *Concentrations and bioaccessibility of polycyclic aromatic hydrocarbons in wastewater-irrigated soil using in vitro gastrointestinal test*, Environ. Sci. Pollut. Res. 15 (2008), pp. 344–353.

[3] H. Iwata, M. Watanabe, Y. Okajima, S. Tanabe, M. Amano, N. Miyazaki, and E.A. Petrov, *Toxicokinetics of PCDD, PCDF, and coplanar PCB congeners in Baikal seals,* Pusa sibirica*: Age-related accumulation, maternal transfer, and hepatic sequestration*, Environ. Sci. Technol. 38 (2004), pp. 3505–3513.

[4] P.M. Hoai, N.T. Ngoc, N.H. Minh, P.H. Viet, M. Berg, A.C. Alder, and W. Giger, *Recent levels of organochlorine pesticides and polychlorinated biphenyls in sediments of the sewer system in Hanoi, Vietnam*, Environ. Pollut. 158 (2010), pp. 913–920.

[5] S. Tanabe, K. Ramu, T. Isobe, and S. Takahashi, *Brominated flame retardants in the environment of Asia-Pacific: An overview of spatial and temporal trends*, J. Environ. Monit. 10 (2008), pp. 188–197.

[6] D. Wu, X.H. Liu, L. Wang, L. Wang, M.Z. Xu, T. Sun, Z.F. Yang, and J.L. Zhou, *QSARs on the depuration rate constants of polycyclic aromatic hydrocarbons in* Elliptio complanata, QSAR Comb. Sci. 28 (2009), pp. 537–541.

[7] H. Morrison, T. Yankovich, R. Lazar, and G.D. Haffner, *Elimination rate constants of 36 PCBs in zebra mussels (*Dreissena polymorpha*) and exposure dynamics in the Lake St Clair Lake Erie corridor*, Can. J. Fish. Aquat. Sci. 52 (1995), pp. 2574–2582.

[8] S. Uno, H. Shiraishi, S. Hatakeyama, and A. Otsuki, *Uptake and depuration kinetics and BCFs of several pesticides in three species of shellfish (*Corbicula leana*,* Corbicula japonica*, and* Cipangopludina chinensis*): Comparison between field and laboratory experiment*, Aquat. Toxicol. 39 (1997), pp. 23–43.

[9] A.V. Weisbrod, L.P. Burkhard, J. Arnot, O. Mekenyan, P.H. Howard, C. Russom, R. Boethling, Y. Sakuratani, T. Traas, T. Bridges, C. Lutz, M. Bonnell, K. Woodburn, and T. Parkerton, *Workgroup report: Review of fish bioaccurnulation databases used to identify persistent, bioaccumulative, toxic substances*, Environ. Health Perspect. 115 (1997), pp. 255–261.

[10] S. O'Rourke, K.G. Drouillard, and G.D. Haffner, *Determination of laboratory and field elimination rates of polychlorinated biphenyls (PCBs) in the freshwater mussel,* Elliptio complanata, Arch. Environ. Con. Toxicol. 47 (2004), pp. 74–83.

[11] W.A. Thorsen, D. Forestier, T. Sandifer, P.R. Lazaro, W.G. Cope, and D. Shea, *Elimination rate constants of 46 polycyclic aromatic hydrocarbons in the unionid mussel,* Elliptio complanata, Arch. Environ. Con. Toxicol. 47 (2004), pp. 332–340.

[12] European Commission. Enterprise & Industry Directorate General and Environment Directorate General. *REACH in brief*. Available online at: http://ecb.jrc.it/REACH/, 2002.

[13] J.W. Chen, X.H. Li, H.Y. Yu, Y. Wang, and X.L. Qiao, *Progress and perspectives of quantitative structure-activity relationships used for ecological risk assessment of toxic organic compounds*, Sci. China Ser. B 51 (2008), pp. 593–606.

[14] X. Li, T.A. Zhang, X.M. Min, and P. Liu, *Toxicity of aromatic compounds to* Tetrahymena *estimated by microcalorimetry and QSAR*, Aquat. Toxicol. 98 (2010), pp. 322–327.

[15] J. Neuwoehner, T. Zilberman, K. Fenner, and B.I. Escher, *QSAR-analysis and mixture toxicity as diagnostic tools: Influence of degradation on the toxicity and mode of action of diuron in algae and daphnids*, Aquat. Toxicol. 97 (2010), pp. 58–67.

[16] F. Li, Q. Xie, X.H. Li, N. Li, P. Chi, J.W. Chen, Z.J. Wang, and C. Hao, *Hormone activity of hydroxylated polybrominated diphenyl ethers on human thyroid receptor-beta: In vitro and in silico investigations*, Environ. Health Perspect. 118 (2010), pp. 602–606.

[17] F. Li, X.H. Li, J.P. Shao, P. Chi, J.W. Chen, and Z.J. Wang, *Estrogenic activity of anthraquinone derivatives: In vitro and in silico studies*, Chem. Res. Toxicol. 23 (2010), pp. 1349–1355.

[18] Y. Wang, J.W. Chen, F. Li, H.l. Qin, X. Qiao, and C. Hao, *Modeling photoinduced toxicity of PAHs based on DFT-calculated descriptors*, Chemosphere 76 (2009), pp. 999–1005.

[19] S.Y. Zhang, J.W. Chen, X.L.k. Qiao, L. Ge, X.Y. Cai, and G.S. Na, *Quantum chemical investigation and experimental verification on the aquatic photochemistry of the sunscreen 2-phenylbenzimidazole-5-sulfonic acid*, Environ. Sci. Technol. 2010 (2010), pp. 7484–7490.

[20] K.G. Drouillard, S. Chan, S. O'Rourke, G.D. Haffner, and R.J. Letcher, *Elimination of 10 polybrominated diphenyl ether (PBDE) congeners and selected polychlorinated biphenyls (PCBs) from the freshwater mussel,* Elliptio complanata, Chemosphere 69 (2007), pp. 362–370.

[21] M.Z. Xu, X.H. Liu, L. Wang, D. Wu, T. Sun, Z.F. Yang, and B.S. Cui, *Modeling the depuration rates of polychlorinated biphenyls in two mussel species with theoretical molecular descriptors*, Sci. China Ser. B 52 (2009), pp. 1281–1286.

[22] OECD, *Guidance document on the validation of (Quantitative) Structure–Activity Relationships [(Q)SARs] models*. Available online at: http://www.oecd.org/dataoecd/55/22/38131728.pdf, 2007.

[23] S. Wold, M. Sjostrom, and L. Eriksson, *PLS-regression: A basic tool of chemometrics*, Chemometr. Intel. Lab. Syst. 58 (2001), pp. 109–130.

[24] P. Baumard, H. Budzinski, and P. Garrigues, *PAHs in Arcachon Bay, France: Origin and biomonitoring with caged organisms*, Mar. Pollut. Bul. 36 (1998), pp. 577–586.

[25] M.J.S. Dewar and W. Thiel, *Ground-states of molecules .38. MNDO method – approximations and parameters*, J. Am. Chem. Soc. 99 (1977), pp. 4899–4907.

[26] L.M. Colosi, Q.G. Huang, and W.J. Weber, *Quantitative structure–activity relationship based quantification of the impacts of enzyme-substrate binding on rates of peroxidase-mediated reactions of estrogenic phenolic chemicals*, J. Am. Chem. Soc. 128 (2006), pp. 4041–4047.

[27] D.R. Roy, R. Parthasarathi, B. Maiti, V. Subramanian, and P.K. Chattaraj, *Electrophilicity as a possible descriptor for toxicity prediction*, Bioorgan. Med. Chem. 13 (2005), pp. 3405–3412.

[28] J. Padmanabhan, R. Parthasarathi, V. Subramanian, and P.K. Chattaraj, *Group philicity and electrophilicity as possible descriptors for modeling ecotoxicity applied to chlorophenols*, Chem. Res. Toxicol. 19 (2006), pp. 356–364.

[29] P. Politzer, L. Abrahmsen, and P. Sjoberg, *Effects of amino and nitro substituents upon the electrostatic potential of an aromatic ring*, J. Am. Chem. Soc. 106 (1984), pp. 855–860.

[30] C. Colominas, M. Orozco, F.J. Luque, J.I. Borrell, and J. Teixido, *A priori prediction of substituent and solvent effects in the basicity of nitriles*, J. Org. Chem. 63 (1998), pp. 4947–4953.

[31] A. Chana, M.A. Concejero, M. de Frutos, M.J. Gonzalez, and B. Herradon, *Computational studies on biphenyl derivatives. Analysis of the conformational mobility, molecular electrostatic potential, and dipole moment of chlorinated biphenyl: Searching for the rationalization of the selective toxicity of polychlorinated biphenyls (PCBs)*, Chem. Res. Toxicol. 15 (2002), pp. 1514–1526.

[32] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, O. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, and D.J. Fox, *Gaussian 09, Revision A.1*, Gaussian, Inc., Wallingford CT, 2009.

[33] S. Arulmozhiraja, F. Shiraishi, T. Okumura, M. Iida, H. Takigami, J.S. Edmonds, and M. Morita, *Structural requirements for the interaction of 91 hydroxylated polychlorinated biphenyls with estrogen and thyroid hormone receptors*, Toxicol. Sci. 84 (2005), pp. 49–62.

[34] P. Huetz, E.E. Kamarulzaman, H.A. Wahab, and J. Mavri, *Chemical reactivity as a tool to study carcinogenicity: Reaction between estradiol and estrone 3,4-quinones ultimate carcinogens and guanine*, J. Chem. Inf. Comput. Sci. 44 (2004), pp. 310–314.

[35] J.W. Chen, T. Harner, G.H. Ding, X. Quan, K.W. Schramm, and A. Kettrup, *Universal predictive models on octanol-air partition coefficients at different temperatures for persistent organic pollutants*, Environ. Toxicol. Chem. 23 (2004), pp. 2309–2317.

[36] G. Schüürmann, R.U. Ebert, J.W. Chen, B. Wang, and R. Kuhne, *External validation and prediction employing the predictive squared correlation coefficient - test set activity mean vs training set activity mean*, J. Chem. Inf. Model. 48 (2008), pp. 2140–2145.

[37] L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell, and P. Gramatica, *Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs*, Environ. Health Perspect. 111 (2003), pp. 1361–1375.
[38] A. Golbraikh and A. Tropsha, *Beware of q^2!*, J. Mol. Graph. Model. 20 (2002), pp. 269–276.
[39] P.K. Chattaraj, U. Sarkar, and D.R. Roy, *Electrophilicity index*, Chem. Rev. 106 (2006), pp. 2065–2091.
[40] T.H. Nguyen, U. Goss K, and W.P. Ball, *Polyparameter linear free energy relationships for estimating the equilibrium partition of organic compounds between water and the natural organic matter in soils and sediments*, Environ. Sci. Technol. 39 (2005), pp. 913–924.
[41] F.A.D. Ribeiro and M.M.C. Ferreira, *QSPR models of boiling point, octanol-water partition coefficient and retention time index of polycyclic aromatic hydrocarbons*, J. Mol. Struct. 663 (2003), pp. 109–126.