Noncovalent interactions between hydroxylated polycyclic aromatic hydrocarbon and DNA: Molecular docking and QSAR study

Fei Li[a], Xuehua Li[b], Xiaoli Liu[a,c], Linbao Zhang[a,c], Liping You[a,c], Jianmin Zhao[a], Huifeng Wu[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 264003, China
[b] Key Laboratory of Industrial Ecology and Environmental Engineering (MOE), School of Environmental Science and Technology, Dalian University of Technology, Linggong Road 2, Dalian 116024, China
[c] The Graduate School of Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT

Polycyclic aromatic hydrocarbons (PAHs) can be hydroxylated by CYP450-oxidases (1A1 and 1B1 mainly) and may cause DNA damage and cancer. However, the mechanism of such interactions has not been fully understood. In this study, an integrated molecular docking and QSAR approach was employed to further investigate the binding interactions between hydroxylated PAHs (HO-PAHs) and calf thymus DNA (CT-DNA). Molecular docking, hydrogen-bonding, hydrophobic and π–π interactions were observed to be characteristic interactions between HO-PAHs and DNA. An optimum QSAR model with good robustness and predictability was developed based on the molecular structural parameters calculated by the density function theory and partial least squares. Additionally, the developed QSAR model indicated that the molecular size, polarizability and electrostatic potential of HO-PAHs were related to the binding affinities to DNA.

1. Introduction

Polycyclic aromatic hydrocarbons (PAHs) are classed as persistent organic pollutants (POPs) under the Stockholm Convention. They are a large class of ubiquitous organic pollutants and have received great attention because of their carcinogenic, teratogenic and mutagenic properties (Gallegos et al., 2001; Kumar et al., 2001; Borosky and Laali, 2005; Xue and Warshawsky, 2005). In most cases, PAHs can be hydroxylated by CYP450-oxidases, which is a key step in the activation process to produce the polar biochemically reactive electrophilic species (ultimate carcinogenic metabolites) capable of interacting with cellular macromolecules, particularly nucleic acids and proteins (Zhou et al., 2003). Most recently, hydroxylated PAHs (HO-PAHs) have emerged and are causing increasing concern due to their detection in human hair (Schummer et al., 2009), urine (Campo et al., 2010; Li et al., 2010c), and expired air samples (Li et al., 2010c); these agents have even been found in the bile of deep-sea fish (Escartin and Porte, 1999).

Previous studies have shown that certain HO-PAHs can affect hormone homeostasis, as they act as potent ligands for binding to the aryl hydrocarbon receptor (AhR) and even interact with DNA (Wang et al., 2009a; Wenger et al., 2009; Ohura et al., 2010; Wei et al., 2010). For instance, it was reported that 5 HO-PAHs (2-hydroxychrysene, 2-hydroxyphenanthrene, 1-hydroxypyrene, 2-hydroxynaphthalene and 1-hydroxynaphthalene), which showed structural similarities to 17β-estrodiol, exhibit estrogenic activities (Wenger et al., 2009). Ohura et al. (2010) reported that HO-PAHs showed AhR-ligand binding activities by a recombinant yeast assay system, especially the hydroxylated derivatives of naphthalene. Additionally, a clear morphological change of calf thymus DNA (CT-DNA) from linear type to condensation form was observed after binding with 9-hydroxyfluorene in atomic force microscopy (Wang et al., 2009a). Recently, Wei et al. (2010) also reported that 11 PAHs metabolites predominantly interacted with human p53 DNA by intercalation instead of groove binding. However, the detail mechanisms of DNA binding associated with HO-PAHs remain unclear. Hence, it is of great importance to improve our understanding of the mechanism of interactions between HO-PAHs and DNA.

The formation of DNA adducts is a key step in DNA damage, which could lead to malignancy (Radwan and Ramsdell, 2008). The proven key carcinogenic product, benzo[a]pyrene-r-7,t-8-dihydrodiol-t-9,10-epoxide (BPDE), intercalates rapidly with DNA base pairs to form a complex, which undergoes protonation to yield an intercalated triol carbonium ion intermediate (Geacintov et al., 1981; Meehan et al., 1982). However, the standard samples are limited with respect to making experimental determinations of the binding constants (K_b) of each HO-PAHs to DNA. Therefore, an alternative approach, quantitative structure–activity relationship (QSAR), is suggested by the new EU chemicals legislation REACH (European Commission., 2002), which has been successfully used in acute toxicity studies (Christen et al., 2010; Li et al., 2010b), mixture toxicity (Arrhenius et al., 2004; Neuwoehner et al., 2010), endocrine disrupting activities (Li et al., 2009, 2010a) and photo-induced

toxicity (Wang et al., 2009b; Zhang et al., 2010) of organic compounds.

DNA sequences may be changed by the mutagens (Besaratinia and Pfeifer, 2006). Traditional experimental techniques are employed to determine the DNA damage, such as single cell gel electrophoresis (SCGE) (Lee and Steinert, 2003), the micronucleus assay (Grisolia, 2002) and other systems such as the sister-chromatid exchange assay (Tofilon et al., 1985). However, it is difficult to determine the base substitution and the position of mutation only by experimental approaches (Kozack and Loechler, 1999; Joung et al., 2009), which limit the comprehensive understanding of the mechanism of DNA damage. Consequently, further studies are required on the interactions between ligands and DNA by molecular simulation to clarify the mechanism of DNA damage (Kitchen et al., 2004; Moitessier et al., 2008). Molecular simulation, such as molecular docking, has become an important approach to elucidate the interactions between ligands and macromolecular targets, which is a rapid, low-cost detection system which has been successfully used in DNA–ligand interactions (Rabinowitz et al., 2009), xenoestrogen screening (Amadasi et al., 2009) and molecular recognition (Erickson et al., 2004).

In this study, an integrated molecular docking and QSAR approach was employed to investigate the binding interactions between HO-PAHs and CT-DNA. Molecular docking was performed to define a model for the comprehension of the binding interactions between ligands and receptor. By observing the mechanism of interactions, appropriate molecular structural parameters computed by the density function theory (DFT) were adopted to construct QSAR models. These developed QSAR models were externally validated and the applicability domain was depicted. Furthermore, from the developed QSAR models, critical molecular structural features related to DNA-adducts formation were identified.

2. Materials and methods

2.1. Data compilation and the chemical domain

The K_b values of 24 HO-PAHs with CT-DNA were taken from Wang et al. (2009a), and then converted into the form of log K_b (Table 2). The K_b values were determined by a previously established electrochemical displacement method. More details can be found in the previous study (Wang et al., 2009a).

2.2. Molecular docking

The binding mode for the HO-PAHs to CT-DNA was investigated by CDOCKER, which has been incorporated into Discovery Studio 2.5 (Accelrys Software Inc.) through the Dock Ligands protocol. CDOCKER is an implementation of a docking tool based on CHARMm forcefield that has been proved to be viable. The crystal structure of DNA (PDB entry code: 1DJD) was retrieved from the Brookhaven Protein Database (PDB http://www.rcsb.org/pdb). Hydrogen atoms were added and the crystallographic waters were removed. In CDOCKER,

random ligand conformations are generated through molecular dynamics (MD), and a variable number of rigid-body rotations/translations are applied to each conformation to generate initial ligand poses. The random conformations are refined by grid-based simulated annealing in the receptor active site, which improves accuracy. From the molecular docking analysis, insights into the interactions between the HO-PAHs and DNA were gained, which facilitated the selection of appropriate molecular parameters to characterize the interactions in the QSAR studies.

2.3. Mechanism consideration and molecular structural parameters selection

As proposed by the Organisation for Economic Co-operation and Development (OECD) guidelines, QSAR models should be developed based on the mechanism of action (OECD, 2007). The intercalation imposes structural alterations to DNA in order to open an intercalation gap between two consecutive base pairs, which may inhibit both transcription and DNA replication, causing toxicity and mutations (Hannon, 2007). Most intercalators are aromatic and planar molecules and they do not possess sequence selectivity since the binding of these ligands to DNA depends basically on π-stacking and stabilizing electrostatic interactions (Tse and Boger, 2004). Molecular structural descriptors that describe hydrophobic, electronic and steric properties of molecules were selected to describe the interaction between HO-PAHs and DNA, which was calculated using the DRAGON 2.1 (Todeschini and Consonni, 2000) and Gaussian 09 packages (Frisch et al., 2009).

All the initial geometries of the compounds were optimized by semi-empirical method PM3, then optimized at the hybrid Hartree-Fock DFT B3LYP/6-31G(d,p) level (Arulmozhiraja et al., 2005). Solvent (water) effects were taken into consideration implicitly, including the integral equation formulation of the polarized continuum model (IEFPCM) (Huetz et al., 2004). The frequency analysis was performed on the optimized geometries to ensure that the systems had no imaginary vibration frequencies.

The optimized molecular structures were imported to Dragon 2.1 (Talete Srl, Milano, Italy), and 1481 diverse descriptors (different functional groups, constitutional, geometrical, topological, Whim 3D, electronic, etc.) for each molecule were calculated. The quantum chemical descriptors, including 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 (qH^+), the most negative formal charge in the molecule (q^-) and electrophilicity index (ω), were computed by Gaussian 09 programs (Frisch et al., 2009).

Additionally, the molecular surface potential derived parameters, such as the most positive and most negative values of the molecular surface potential ($V_{s,max}$, $V_{s,min}$), the averages of the positive and negative potentials on the molecular surface ($\bar{V}_s^+$, $\bar{V}_s^-$), the average deviation of surface potential (Π) and the balance parameter of surface potential (ν), were purposely selected to describe the hydrogen bond or electro-

static interactions. These potential derived descriptors have been successfully used to rationalize the toxicities of chlorinated diphenyls (Chana et al., 2002), which are calculated by the following equations:

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

$$\bar{V}_S^- = \frac{1}{\beta}\sum_{j=1}^{\beta} V_S^-(r_j) \tag{2}$$

$$\Pi = \frac{1}{\alpha+\beta}\sum_{i=1}^{\alpha+\beta} \left|V(r_i) - \bar{V}_s\right| = \frac{1}{\alpha+\beta}\sum_{i=1}^{\alpha+\beta} \left|V(r_i) - \frac{\alpha\bar{V}_S^+ + \beta\bar{V}_S^-}{\alpha+\beta}\right| \tag{3}$$

$$\nu = \frac{\sigma_+^2\sigma_-^2}{(\sigma_{tot}^2)^2} \tag{4}$$

$$\sigma_{tot}^2 = \sigma_+^2 + \sigma_-^2 = \frac{1}{\alpha}\sum_{i=1}^{\alpha}[V^+(r_i) - \bar{V}_s^+]^2 + \frac{1}{\beta}\sum_{j=1}^{\beta}[V^-(r_j) - \bar{V}_s^-]^2 \tag{5}$$

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; $\bar{V}^+$ 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.

2.4. QSAR development and validation

The original data set was randomly divided into a training set (80%) and a validation set (20%), as listed in Table 2. Partial least squares (PLS) regression was performed for the model development as PLS can analyze data with strongly collinear, noisy and numerous predictor variables (Wold et al., 2001). The software of Simca-S (Version 6.0) was employed for the PLS analysis. Simca-S adopts 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^2_{CUM} (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 the previous studies (Chen et al., 2004; Li et al., 2010a).

The model predictability was also evaluated by Y-scrambling test and external validation. In the Y-scrambling test, the Y ($\log K_b$) data are randomly permuted keeping the descriptor matrix intact, followed by a PLS run. Each randomization and subsequent PLS analysis generates a new set of R^2 and Q^2 values, which are plotted against the correlation coefficient between the original Y values and the permuted Y values. The intercepts for the R^2 and Q^2_{CUM} lines in this plot are a measure of the overfit. A model is considered valid if $R^2_{int} < 0.4$ and $Q^2_{int} < 0.05$.

The performance of external validation was characterized by the determination coefficient (R^2), standard error (SE) and
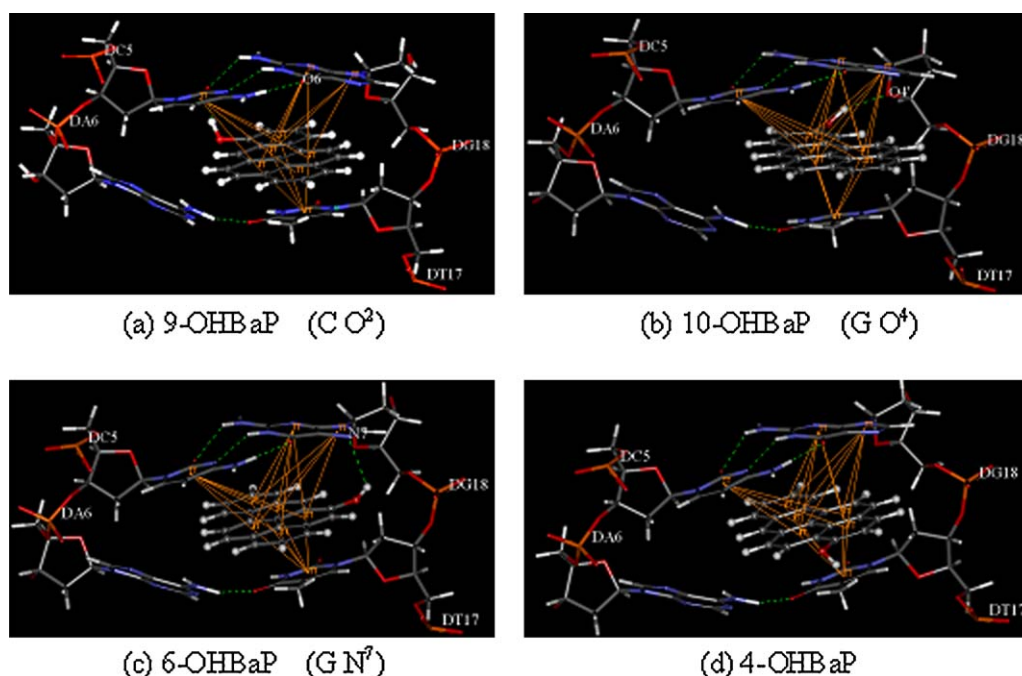
Fig. 1 – Docking views of (a) 9-OHBaP, (b) 10-OHBaP, (c) 6-OHBaP and (d) 4-OHBaP in the binding site of DNA. Green dotted line shows H-bonds between BaPs and basic groups. Carbon is colored in grey, oxygen red, and nitrogen blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

external explained variance (Q^2_{EXT}), which are defined as follows (Schüürmann et al., 2008):

$$R^2 = \frac{1 - \sum_{i=1}^{n}(y_i^{fit} - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (6)$$

$$SE = \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2/n - 1} \qquad (7)$$

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_{EXT}}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{EXT}}(y_i - \bar{y}_{EXT})^2} \qquad (8)$$

where y_i^{fit} is the fitted $\log K_b$ value of the i-th 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 i-th compound, respectively. $\bar{y}_{EXT}$ 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) (Eriksson et al., 2003). h_i value of a chemical in the original variable space and the warning leverage value (h^*) are defined as:

$$h_i = x_i^T(X^TX)^{-1}x_i(i = 1, \cdots, n) \qquad (9)$$

$$h^* = 3(p + 1)/n \qquad (10)$$

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

3. Results and discussion

3.1. Molecular docking analysis

Docking with program CDOCKER could successfully reproduce the X-ray pose of natural ligand with a root-mean-square deviation (RMSD) of 0.48 Å. Such proximity can be regarded as a good reproduction of the crystal structure. We therefore believe that the binding conformations of the HO-PAHs ligands analyzed here are reasonably well predicted by CDOCKER.

Fig. 1 shows the docking view of 4 representative HO-PAHs (9-hydroxybenzo[a]pyrene, 10-hydroxybenzo[a]pyrene, 6-hydroxybenzo[a]pyrene and 4-hydroxybenzo[a]pyrene) in the binding site of DNA. Wei et al. (2010) reported that the binding of the PAH metabolites to DNA showed sequence selectivity and the functional groups on the periphery of the PAH aromatic ring played crucial roles in regulating its binding affinity with DNA. As shown in Fig. 1, hydrogen-bondings are observed to be characteristic interactions. There are mainly three types of H-bonds: (a) H-bonds formed between the hydroxyl hydrogen of 9-OHBaP and the oxygen (O^2) of cytosine; (b) H-bonds between the hydroxyl hydrogen of 10-OHBaP with the oxygen (O^4) of guanine; and (c) H-bonds between the hydroxyl hydrogen of 6-OHBaP with the nitrogen (N^7) of guanine.

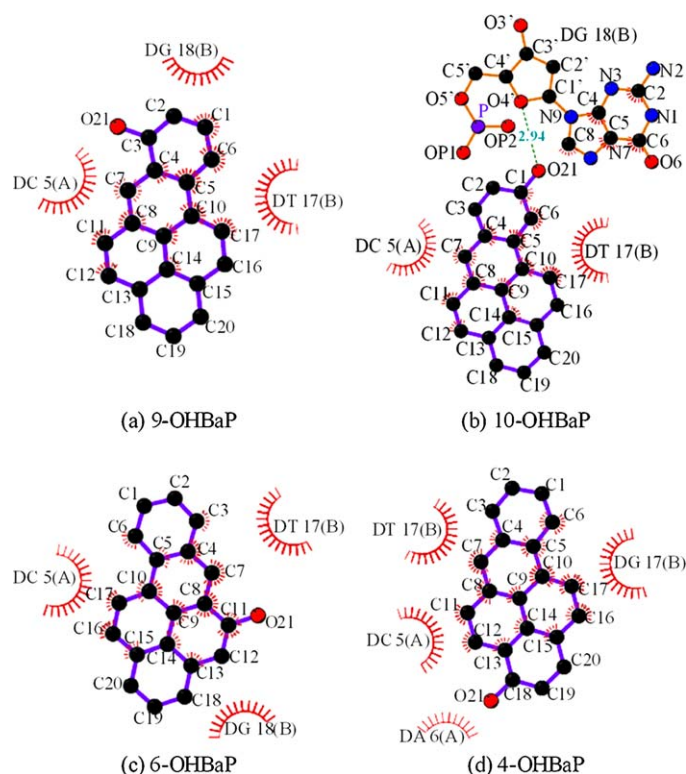Among the main DNA binding modes, intercalation is the most common way through which small and rigid aromatic

Fig. 2 – Hydrophobic interaction between HO-PAHs and DNA in the binding site (●—● ligand bond, ●—● non-ligand bond, ●··●·●● hydrogen bond and its length, ⌇ non-ligand residues involved in hydrophobic contacts, ● corresponding atoms involved in hydrophobic contacts).

molecules recognize DNA (Ricci and Netz, 2009). Acting as an 'anchor', the hydrogen-bonding intensely determines the 3D space position of the benzene ring in the binding pocket, and facilitates the hydrophobic interaction of the HO-PAHs with adenine (A), thymine (T), cytosine (C) and guanine (G), as shown in Fig. 2. Additionally, the 4 HO-PAHs intercalate between two adjacent base pairs, and there are also π–π interactions between the phenyl of HO-PAHs and guanine (G), cytosine (C) or thymine (T).

Hydrogen bonds play great importance in many fields of biological chemistry, such as genetic code. In DNA, the two helical chains of nucleotides are held together by purines forming hydrogen bonds to pyrimidines, with adenine (A) bonding only to thymine (T), and cytosine (C) bonding only to guanine (G) (Guerra et al., 2000). In many cases the produced specific DNA damage may either block the replication

and transcription or generate mutations by miscoding during replication. In the normal Watson–Crick pairs G–C, the nitrogen atoms (N^1) of G form hydrogen bond with the nitrogen atoms (N^3) of C. As shown in Fig. 1(a), 9-OHBaP formed hydrogen bond with the oxygen atom (O^2) of C, the charge distribution of ketone–enolate type may be caused, and the nitrogen atoms (N^4) of C may be deprotonated (Dipple, 1995), then hydrogen-bond may be formed between C and A instead
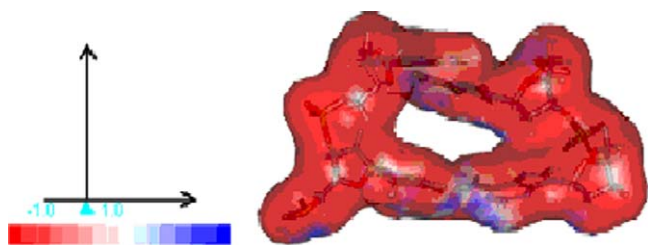

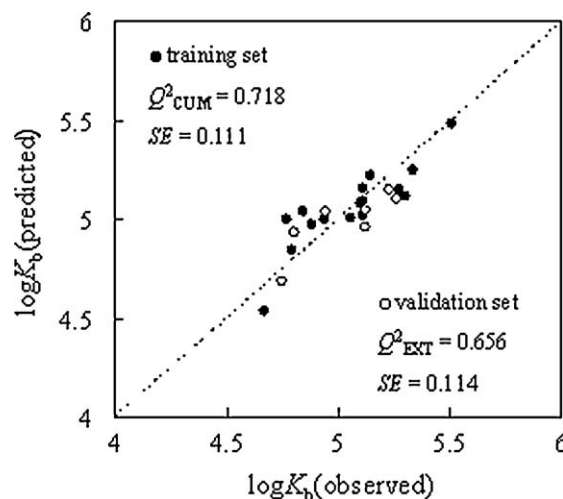**Fig. 3 – Electrostatic potential of the ligand binding site for DNA.**

Fig. 4 – Plot of observed versus predicted $\log K_b$ values for the training and validation.

Table 1 – Physical–chemical meanings of the descriptors used in the developed QSAR model.

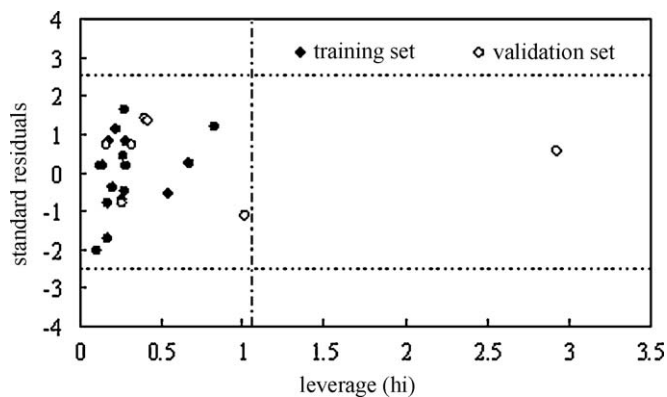| Descriptor	Chemical meanings
SIC_2	Structural information content index (neighborhood symmetry of 2-order)
$MATS_{5p}$	Moran autocorrelation – lag 5/weighted by atomic polarizabilities
Mor_{29e}	3D-MoRSE – signal 29/weighted by atomic Sanderson electronegativities
E_{2s}	2nd component accessibility directional WHIM index/weighted by atomic electrotopological states
Π	The average deviation of surface potential

Fig. 5 – Plot of standardized residuals versus leverages. Dash lines represent ±2.5 standardized residual, dotted line represents warning leverage ($h^* = 1.058$).

of C and G. Hence, adducts on C may cause the mispairing between C and A, eventually it results the G:C → A:T transversion mutations and might block the progression of DNA polymerases.

For the HO-PAHs with relatively planar highly conjugated aromatic structures, they binds with DNA depends on π-stacking and stabilizing electrostatic interactions (Tse and Boger, 2004). The molecular surface potential indicates the charge distribution in a molecule (Politzer et al., 1984), which gauge the basicity and nucleophilicity of a molecule (Colominas et al., 1998). As shown in Fig. 3, the binding site has negative potentials, from which it can also be concluded that the positive potentials of the HO-PAHs molecules facilitate them to bind with DNA.

3.2. Development and validation of the QSAR model for the $\log K_b$

Forward stepwise regression was adopted to screen molecular descriptors, then 5 descriptors (SIC_2, $MATS_{5p}$, Mor_{29e}, E_{2s} and Π) were finally selected for model development, which are listed in Table 1 with their physical–chemical meanings.

PLS analysis with $\log K_b$ as the dependent variable and the molecular structural parameters as predictor variables resulted in the following optimal QSAR model:

$$\log K_b = 5.77 - 2.76\,SIC_2 - 1.43MATS_{5p} + 2.62$$
$$\times\ 10^{-1}Mor_{29e} - 1.45E_{2s} + 5.36\Pi$$

$n(\text{training set}) = 17, A = 1, R^2 = 0.751, Q^2_{CUM} = 0.718, SE$

$\quad = 0.111(\text{training set}),$

$n(\text{validation set}) = 7, Q^2_{EXT} = 0.656, SE$

$\quad = 0.114(\text{validation set}), p < 0.0001,$

where p is the significance level.

The developed QSAR model was subjected to "Y-scrambling" statistical validation to test the possibility of chance correlation. We randomly permutated the Y-variable, re-built the statistical model, and observed the trends of the predictive power and goodness of fit at each step. One hun-

dred rounds of such reshuffling gave coherent decreases in both parameters and the extrapolated value of the Q^2 of -0.10, which shows that the separation model is statistically sound, and that its high predictability is not due to over-fitting of the data.

The predicted $\log K_b$ values and residuals for compounds are listed in Table 2. The R^2 value of the QSAR model was 0.751, indicating a high goodness-of-fit of the model. Q^2_{CUM} of the QSAR was as high as 0.718, implying good robustness of the model. The differences between R^2 and Q^2_{CUM} (0.033) did not exceed 0.3, indicating no over-fitting in the model (Golbraikh and Tropsha, 2002). As show in Fig. 4, the predicted $\log K_b$ values were consistent with the observed values for both the validation and training sets. The model revealed acceptable predictability with $Q^2_{EXT} = 0.656$, $SE = 0.114$. In summary, the developed QSAR model showed satisfactory performance.

3.3. Applicability domain of the developed QSAR model

Application of Kolmogorov-Smirnov test for normality (at the 95% confidence level) confirmed that the distribution of residuals was a distinctive bell-shaped pattern associated with a normal distribution (mean = 0.00, standard deviation = 0.11). Hence, the residuals were non-systematic, and the applicability domain of the developed QSAR model could be visualized by the Williams plot.

The applicability domain of the developed QSAR model is shown in Fig. 5. As shown in the Williams plot (Fig. 5), h_i values of all the compounds in the training and validation sets were lower than the warning value ($h^* = 1.058$). One compound (2-OHFLU) in the validation set was found with large leverage values ($h > h^*$), and it was predicted correctly, indicating that the developed QSAR model had good extrapolating ability. For all the compounds in the training and validation sets, their standardized residuals were smaller than 3 standard deviation units (3σ). Thus there were no outliers for the developed QSAR model.

No.	Name	$\log K_b$			E_{2s}	$MATS_{5p}$	Π	SIC_2	Mor_{29e}
		Observed	Predicted	Residuals					
1	1-OHNAP	4.792	4.846	−0.054	0.222	−0.091	0.019	0.018	−0.091
2	2-OHFLU[a]	4.748	4.697	0.051	0.101	−0.09	0.018	0.002	0.002
3	9-OHFLU	4.672	4.539	0.133	0.313	−0.114	0.016	−0.083	−0.114
4	2-OHPHE	5.146	5.225	−0.079	0.123	−0.152	0.020	−0.039	−0.152
5	3-OHPHE	5.111	5.092	0.019	0.171	−0.141	0.019	−0.017	−0.141
6	4-OHPHE[a]	5.124	4.959	0.165	0.252	−0.194	0.018	−0.109	−0.194
7	9-OHPHE[a]	5.121	5.014	0.107	0.302	−0.194	0.020	−0.041	−0.194
8	3-OHFLT	5.507	5.482	0.025	0.159	−0.296	0.020	0.174	−0.296
9	1-OHPYR	5.334	5.246	0.089	0.178	−0.207	0.020	−0.134	−0.207
10	2-OHBcPh[a]	4.806	5.010	−0.204	0.261	−0.089	0.020	−0.112	−0.089
11	3-OHBcPb[a]	4.944	5.057	−0.112	0.194	−0.097	0.020	−0.221	−0.097
12	4-OHBcPb	5.114	5.095	0.019	0.175	−0.092	0.020	0.567	−0.092
13	5-OHBcPh	4.771	4.998	−0.227	0.223	−0.127	0.019	−0.206	−0.127
14	3-OHBaP	5.276	5.151	0.125	0.171	−0.155	0.021	−0.267	−0.155
15	4-OHBaP	4.845	5.038	−0.193	0.259	−0.188	0.020	−0.263	−0.188
16	5-OHBaP	4.881	4.972	−0.092	0.271	−0.182	0.019	−0.236	−0.182
17	6-OHBaP	5.114	5.157	−0.043	0.232	−0.181	0.020	−0.273	−0.181
18	7-OHBaP[a]	5.100	5.080	0.020	0.186	−0.155	0.019	−0.174	−0.155
19	9-OHBaP[a]	5.228	5.144	0.084	0.175	−0.162	0.020	−0.209	−0.162
20	10-OHBaP	5.260	5.036	0.224	0.222	−0.194	0.018	−0.248	−0.194
21	12-OHBaP	5.057	5.009	0.048	0.283	−0.182	0.021	−0.295	−0.182
22	3-OHBkF	5.114	5.023	0.091	0.155	−0.087	0.019	0.023	−0.087
23	7-OHBbF	5.301	5.121	0.180	0.241	−0.203	0.018	0.109	−0.203
24	2-OHIPY	4.940	5.003	−0.063	0.286	−0.162	0.019	0.065	−0.162

Table 2 – Logarithm of the observed and predicted depuration rate constants ($\log K_b$) of the considered compounds.

[a] Compounds in the validation set.

3.4. Mechanistic implications

The developed PLS model extracted on 2 PLS components that were loaded primarily on 4 predictor variables. Values of the variable importance in the projection (*VIP*) and PLS weights (w^*) are listed in Table 3.

The established PLS model extracted one PLS component loaded primarily on 5 predictor variables, E_{2s}, $MATS_{5p}$, Π, SIC_2 and Mor_{29e} (Table 3). E_{2s} belongs to the directional WHIM descriptors and is weighted by atomic electrotopological states. $MATS_{5p}$ is a 2D autocorrelations descriptor and is weighted by atomic polarizabilities. E_{2s} and $MATS_{5p}$ relate to molecular size, and the PLS component mainly condenses information on the molecular size (volume). The negative w^*c and coefficient of E_{2s} and $MATS_{5p}$ in the current QSAR model also indicate the negative correlation between E_{2s}, $MATS_{5p}$ and $\log K_b$. For example, 1-OHNAP and 3-OHPHE have the same Π value (0.019), and the 3-OHPHE has a smaller E_{2s} value than 1-OHNAP, thus the $\log K_b$ value of 3-OHPHE is higher than 1-OHNAP.

As shown in the QSAR model, Mor_{29e} and Π positively correlated with $\log K_b$. Π stands for the average deviation of surface potential. Π presents the uniformity of electrostatic potential distribution on molecular surface, and it is of representative among all descriptors derived from molecular electrostatic potential. The larger Π value is, the less uniform the distribution of the surface electrostatic potentials of a molecule will be. The positive w^*c and coefficient of Π in the current QSAR model indicated that molecules with non-uniform charge distribution were more easily to interact with DNA. Mor_{29e} is a 3D-MoRSE descriptor representing 3D structure of molecular Sanderson electronegativities information. Additionally, the $\log K_b$ value is negatively correlated with SIC_2. The SIC_2 (structural informational content) is present and may be identifying steric qualities of the molecules, which deals with connectivity of the molecule. In general, the current QSAR model indicated the $\log K_b$ value was related to molecular size, polarizability and electrostatic potential.

4. Conclusion

Docking analysis showed that hydrogen bonding and π–π interactions between HO-PAHs molecules and DNA governed the binding affinities. A QSAR was established to characterize the interactions and to model the binding constants of the HO-PAHs. Molecular size, polarizability and electrostatic potential were important factors for the binding interactions between HO-PAHs and DNA. The HO-PAHs molecules with smaller molecular size and higher electrostatic interaction tended to have larger binding constants with DNA. The devel-

Table 3 – VIP values and PLS weights for the optimal PLS model.

	VIP	w^*c
E_{2s}	1.226	−0.548
$MATS_{5p}$	1.147	−0.513
Π	0.978	0.438
SIC_2	0.933	−0.417
Mor_{29e}	0.594	0.266

oped QSAR model had good robustness, predictive ability and mechanism interpretability, which could be applied to predict the binding constants of other HO-PAHs.

Conflict of interest

Nothing declared.

Acknowledgments

REFERENCES

Amadasi, A., Mozzarelli, A., Meda, C., Maggi, A., Cozzini, P., 2009. Identification of xenoestrogens in food additives by an integrated in silico and in vitro approach. Chem. Res. Toxicol. 22, 52–63.

Arrhenius, A., Gronvall, F., Scholze, M., Backhaus, T., Blanck, H., 2004. Predictability of the mixture toxicity of 12 similarly acting congeneric inhibitors of photosystem II in marine periphyton and epipsammon communities. Aquat. Toxicol. 68, 351–367.

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

Besaratinia, A., Pfeifer, G.P., 2006. Investigating human cancer etiology by DNA lesion footprinting and mutagenicity analysis. Carcinogenesis 27, 1526–1537.

Borosky, G.L., Laali, K.K., 2005. A computational study of carbocations from oxidized metabolites of dibenzo[a,h]acridine and their fluorinated and methylated derivatives. Chem. Res. Toxicol. 18, 1876–1886.

Campo, L., Rossella, F., Pavanello, S., Mielzynska, D., Siwinska, E., Kapka, L., Bertazzi, P.A., Fustinoni, S., 2010. Urinary profiles to assess polycyclic aromatic hydrocarbons exposure in coke-oven workers. Toxicol. Lett. 192, 72–78.

Chana, A., Concejero, M.A., de Frutos, M., Gonzalez, M.J., Herradon, B., 2002. 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, 1514–1526.

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

Christen, V., Hickmann, S., Rechenberg, B., Fent, K., 2010. Highly active human pharmaceuticals in aquatic systems: a concept for their identification based on their mode of action. Aquat. Toxicol. 96, 167–181.

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

Dipple, A., 1995. DNA-adducts of chemical carcinogens. Carcinogenesis 16, 437–441.

Erickson, J.A., Jalaie, M., Robertson, D.H., Lewis, R.A., Vieth, M., 2004. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. J. Med. Chem. 47, 45–55.

Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M., Gramatica, P., 2003. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ. Health Perspect. 111, 1361–1375.

Escartin, E., Porte, C., 1999. Hydroxylated PAHs in bile of deep-sea fish. Relationship with xenobiotic metabolizing enzymes. Environ. Sci. Technol. 33, 2710–2714.

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

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

Gallegos, A., Robert, D., Girones, X., Carbo-Dorca, R., 2001. Structure–toxicity relationships of polycyclic aromatic hydrocarbons using molecular quantum similarity. J. Comput. Aid. Mol. Des. 15, 67–80.

Geacintov, N.E., Yoshida, H., Ibanez, V., Harvey, R.G., 1981. Non-covalent intercalative binding of 7,8-dihydroxy-9,10-epoxybenzo(a)pyrene to DNA. Biochem. Biophys. Res. Co. 100, 1569–1577.

Golbraikh, A., Tropsha, A., 2002. Beware of q^2! J. Mol. Graph. Model. 20, 269–276.

Grisolia, C.K., 2002. A comparison between mouse and fish micronucleus test using cyclophosphamide, mitomycin C and various pesticides. Mutat. Res.-Gen. Toxicol. Environ. Mutagen. 518, 145–150.

Guerra, C.F., Bickelhaupt, F.M., Snijders, J.G., Baerends, E.J., 2000. Hydrogen bonding in DNA base pairs: reconciliation of theory and experiment. J. Am. Chem. Soc. 122, 4117–4128.

Hannon, M.J., 2007. Supramolecular DNA recognition. Chem. Soc. Rev. 36, 280–295.

Huetz, P., Kamarulzaman, E.E., Wahab, H.A., Mavri, J., 2004. 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, 310–314.

Joung, I.S., Cetinkol, O.P., Hud, N.V., Thomas, E., 2009. Molecular dynamics simulations and coupled nucleotide substitution experiments indicate the nature of A center dot A base pairing and a putative structure of the coralyne-induced homo-adenine duplex. Nucleic Acids Res. 37, 7715–7727.

Kitchen, D.B., Decornez, H., Furr, J.R., Bajorath, J., 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat. Rev. Drug Discov. 3, 935–949.

Kozack, R.E., Loechler, E.L., 1999. Molecular modeling of the major adduct of (+)-anti-B[a]PDE(N-2-dG) in the eight conformations and the five DNA sequences most relevant to base substitution mutagenesis. Carcinogenesis 20, 85–94.

Kumar, S., Chang, R.L., Wood, A.W., Xie, J.G., Huang, M.T., Cui, X.X., Kole, P.L., Sikka, H.C., Balani, S.K., Conney, A.H., Jerina, D.M., 2001. Tumorigenicity of racemic and optically pure bay region diol epoxides and other derivatives of the nitrogen heterocycle dibenz[a,h]acridine on mouse skin. Carcinogenesis 22, 951–955.

Lee, R.F., Steinert, S., 2003. Use of the single cell gel electrophoresis/comet assay for detecting DNA damage in aquatic (marine and freshwater) animals. Mutat. Res.-Rev. Mutat. 544, 43–64.

Li, F., Chen, J.W., Wang, Z., Li, J., Qiao, X.L., 2009. Determination and prediction of xenoestrogens by recombinant yeast-based assay and QSAR. Chemosphere 74, 1152–1157.

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

Li, X., Zhang, T.A., Min, X.M., Liu, P., 2010b. Toxicity of aromatic compounds to tetrahymena estimated by microcalorimetry and QSAR. Aquat. Toxicol. 98, 322–327.

Li, Z., Mulholland, J.A., Romanoff, L.C., Pittman, E.N., Trinidad, D.A., Lewin, M.D., Sjodin, A., 2010c. Assessment of non-occupational exposure to polycyclic aromatic hydrocarbons through personal air sampling and urinary biomonitoring. J. Environ. Monitor. 12, 1110–1118.

Meehan, T., Gamper, H., Becker, J.F., 1982. Characterization of reversible, physical binding of benzo[a]pyrene derivatives to DNA. J. Biol. Chem. 257, 479–485.

Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., Corbeil, C.R., 2008. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. Brit. J. Pharmacol. 153, S7–S26.

Neuwoehner, J., Zilberman, T., Fenner, K., Escher, B.I., 2010. 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, 58–67.

OECD, 2007. Guidance document on the validation of (Quantitative) Structure–Activity Relationships [(Q)SARs] models. Available online at: http://appli1.oecd.org/olis/2007doc.nsf/linkto/env-jm-mono(2007)2.

Ohura, T., Kurihara, R., Hashimoto, S., 2010. Aryl hydrocarbon receptor activities of hydroxylated polycyclic aromatic hydrocarbons in recombinant yeast cells. Toxicol. Environ. Chem. 92, 737–742.

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

Rabinowitz, J.R., Little, S.B., Laws, S.C., Goldsmith, M.R., 2009. Molecular modeling for screening environmental chemicals for estrogenicity: use of the toxicant-target approach. Chem. Res. Toxicol. 22, 1594–1602.

Radwan, F.F.Y., Ramsdell, J.S., 2008. Brevetoxin forms covalent DNA adducts in rat lung following intratracheal exposure. Environ. Health Perspect. 116, 930–936.

Ricci, C.G., Netz, P.A., 2009. Docking studies on DNA–ligand interactions: building and application of a protocol to identify the binding mode. J. Chem. Inf. Model. 49, 1925–1935.

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

Schummer, C., Appenzeller, B.M.R., Millet, M., Wennig, R., 2009. Determination of hydroxylated metabolites of polycyclic aromatic hydrocarbons in human hair by gas chromatography–negative chemical ionization mass spectrometry. J. Chromatogr. A 1216, 6012–6019.

Todeschini, R., Consonni, V., 2000. Handbook of Molecular Descriptors. Wiley-VCH, Weinheim, Germany.

Tofilon, P.J., Basic, I., Milas, L., 1985. Prediction of *in vivo* tumor response to chemotherapeutic-agents by the *in vitro* sister chromatid exchange assay. Cancer Res. 45, 2025–2030.

Tse, W.C., Boger, D.L., 2004. Sequence-selective DNA recognition: natural products and nature's lessons. Chem. Biol. 11, 1607–1617.

Wang, L.R., Wang, Y., Chen, J.W., Guo, L.H., 2009a. A structure-based investigation on the binding interaction of hydroxylated polycyclic aromatic hydrocarbons with DNA. Toxicology 262, 250–257.

Wang, Y., Chen, J.W., Li, F., Qin, H., Qiao, X.L., Hao, C., 2009b. Modeling photoinduced toxicity of PAHs based on DFT-calculated descriptors. Chemosphere 76, 999–1005.

Wei, Y., Lin, Y., Zhang, A., Guo, L., Cao, J., 2010. Evaluation of the noncovalent binding interactions between polycyclic aromatic hydrocarbon metabolites and human p53 cDNA. Sci. Total Environ. 408, 6285–6290.

Wenger, D., Gerecke, A.C., Heeb, N.V., Schmid, P., Hueglin, C., Naegeli, H., Zenobi, R., 2009. *In vitro* estrogenicity of ambient particulate matter: contribution of hydroxylated polycyclic aromatic hydrocarbons. J. Appl. Toxicol. 29, 223–232.

Wold, S., Sjostrom, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometr. Intell. Lab. 58, 109–130.

Xue, W.L., Warshawsky, D., 2005. Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: a review. Toxicol. Appl. Pharm. 206, 73–93.

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

Zhou, Z.G., Dai, Q.H., Gui, T., 2003. A QSAR model of PAHs carcinogenesis based on thermodynamic stabilities of biactive sites. J. Chem. Inf. Comput. Sci. 43, 615–621.