Docking and 3D-QSAR studies on the *Ah* receptor binding affinities of polychlorinated biphenyls (PCBs), dibenzo-*p*-dioxins (PCDDs) and dibenzofurans (PCDFs)

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 Environment Processes, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai 264003, PR China*
[b] *Key Laboratory of Industrial Ecology and Environmental Engineering (MOE), Department of Environmental Science and Technology, Dalian University of Technology, Linggong Road 2, Dalian 116024, PR China*
[c] *The Graduate School of Chinese Academy of Sciences, Beijing 100049, PR China*

ARTICLE INFO

ABSTRACT

Polychlorinated biphenyls (PCBs), dibenzo-*p*-dioxins (PCDDs) and dibenzofurans (PCDFs) binding with the aryl hydrocarbon receptor (AhR) have been correlated with many toxic responses. Hence, it is very necessary to study the interactions between these ligands and AhR for further understanding of the mechanism of toxicity. In this study, an integrated molecular docking and 3D-QSAR approach was employed to investigate the binding interactions between PCBs, PCDDs, PCDFs and AhR. From molecular docking, hydrogen-bonding and hydrophobic interactions were observed to be characteristic interactions between compounds and AhR. Based on the mechanism of interactions, an optimum 3D-QSAR model with good robustness ($Q_{CUM}^2 = 0.907$) and predictability ($Q_{EXT}^2 = 0.863$) was developed by partial least squares. Additionally, the developed QSAR model indicated that the molecular size, shape profiles, polarizability and electropological states of compounds were related to the binding affinities to AhR.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Polychlorinated dibenzo-*p*-dioxins (PCDDs), dibenzofurans (PCDFs) and biphenyls (PCBs) are persistent and widespread environmental contaminants, which can cause a great diversity of biological effects including hepatotoxicity, endocrine effects, immunotoxicity, body weight loss, teratogenicity, carcinogenicity and the induction of diverse enzymes such as aryl hydrocarbon hydroxylase (AHH) and 7-ethoxyresorufin *O*-deethylase (EROD) in various organisms (Chovancova et al., 2005; Domingo and Bocio, 2007). Most of their toxic effects are thought to be mediated through a specific protein complex known as the aryl hydrocarbon receptor (AhR) (Landers and Bunce, 1991).

The AhR belongs to the basic helix-loop-helix protein family, which is a ligand-dependent transcription factor located in the cytosol (Nie et al., 2001). The AhR has unique ligand

specificity and it can induce target gene transcription. Binding to AhR is a key step for contaminants exhibiting their toxicity (Hilscherova et al., 2000). Hence, AhR activated by the ligands plays a key role in adverse effects (Ohura et al., 2010). However, the structure of this multimeric protein complex is not known in detail.

Previous studies have demonstrated that several toxic and biochemical effects caused by dioxin-like chemicals are mediated through AhR (Lucier et al., 1993; Nebert et al., 1993). The relative affinities of individual PCBs, PCDDs and PCDFs for the receptors are related with many toxic responses such as thymic atrophy, body weight loss, immunotoxicity and acute lethality (Safe, 1990; Villeneuve et al., 2002; Olivero-Verbel et al., 2004; Mandal, 2005; Ohura et al., 2010). For example, 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) is a potent AhR ligand, which has been used as a reference standard for hazard and risk assessment of these environmental and dietary contaminants (Okey et al., 1994).

The number of PCB (209), PCDD (75) and PCDF (135) congeners is large, and it is impossible to determine the binding affinity of each compound to AhR. Hence, quantitative structure–activity relationship (QSAR) is suggested to model and predict the binding affinities of ligands to AhR by the new EU chemicals legislation REACH (European Commission, 2002), which has been successfully used to predict the toxicity of hydroxylated and quinoid PCB metabolites (Niu et al., 2007), endocrine disrupting activities (Li et al., 2010a, 2010b) and photoinduced toxicity (Wang et al., 2009a,b; Zhang et al., 2010) of organic compounds. Bandiera et al. (1983) developed linear free-energy relationships involving substituent constants and indicator variable for PCBs affinity data, and concluded that polarizability and electron-acceptor properties of the ligands can control the affinity of PCBs binding to the AhR. Zhao et al. (2008) established QSAR models for PCDD/Fs and suggested that dispersion and electrostatic interactions are equally important for the interaction of PCDD/Fs with the AhR. However, the detail mechanisms of these compounds binding associated with AhR remain unclear.

In this study, the three-dimension crystal structure of AhR was obtained by homologous modeling. Molecular docking was performed to define a model for the further understanding of the binding interactions between ligands and receptor interactions. Based on the mechanism of interaction, an optimal QSAR model of AhR binding affinity of PCBs, PCDDs and PCDFs was developed based on the experimental data taken from So and Karplus (1997) and partial least squares (PLS) regression (Wold et al., 2001). From the developed QSAR model, critical molecular structural features related to the AhR binding affinity of PCBs, PCDDs and PCDFs were identified. Furthermore, the developed model was externally validated and the applicability domain was depicted.

2. Materials and methods

2.1. Biological data and computational methods

The data set of this investigation was generated from 65 compounds including 24 PCDDs, 27 PCDFs, and 14 PCBs. Reported data for the relative affinity for binding of PCBs, PCDDs, and

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

Descriptors	Meaning
RG_{yr}	Radius of gyration (mass weight)
SE_{ig}	Absolute eigenvalue sum on geometry matrix
E_{1s}	1st component accessibility directional WHIM index/weighted by atomic electrotopological states
E_{1p}	1st component accessibility directional WHIM index/weighted by atomic polarizabilities
RDF_{065u}	Radial distribution function – 6.5/unweighted
Mor_{14u}	3D-MoRSE – signal 14/unweighted

PCDFs to the AhR are reproduced in Table 1 (Waller and Mckinney, 1992; Wagener et al., 1995; So and Karplus, 1997). The pEC_{50} value was defined as the concentration of the test chemicals reducing specific binding of $[^3H]$ TCDD to 50% of the maximal value in the absence of the competitor. The entire set of compounds was divided into two subsets: training set (80%) that was used to build the actual models, and test set (20%), consisting of molecules not found in the training set, which was used to validate the models once they were built. Members of each set were assigned randomly.

2.2. Homologous modeling

The amino acid residue sequence (sequence GI: 7304873) of AhR conservative domain (the number of residues is from 278 to 384) for *Mus musculus* was obtained by searching NCBI. Similar sequence searching was done with BLAST, and the sequence alignment was completed by ClustalW. The sequence analysis and molecular modeling were completed through both Internet resources and PCs. Homologous 3D model of AhR was built on SWISS-MODEL net server. The nuclear magnetic resonance (NMR) structure of the human PAS domain of the hypoxia-inducible factor 2R (HIF-2R) available in the Protein Data Bank (http://www.rcsb.org/pdb, PDB ID: 1P97) was used as the 3D coordinate template for the homology modeling of AhR. PROCHEEK carried out the structure rational evaluation of the simulated model.

2.3. Molecular docking

The binding mode for the PCBs, PCDDs and PCDFs to AhR was investigated by CDOCKER protocol which had been incorporated into Discovery Studio 2.5. In CDOCKER, random ligand conformations are generated through molecular dynamics, and a variable number of rigid-body rotations are applied to each conformation to generate initial ligand poses. The conformations are further refined by grid-based simulated annealing in the receptor active site, which makes the results accurate. The CDOCKER interaction energy between the compounds and AhR ($E_{binding}$) was finally computed. From the docking analysis, insights into the interactions between the ligands and the receptor were gained, which facilitated the selection of appropriate molecular parameters to characterize the interactions in the following QSAR studies.

2.4. Mechanism consideration and molecular structural parameters selection

As proposed by the OECD guideline, QSAR models should be developed based on the mechanism of action (OECD, 2007). Previous QSAR analyses suggested that steric, electrostatic, hydrophobic, hydrogen bonding and dispersion properties might be important for receptor binding affnity (Poland and Knutson, 1982; Tuppurainen and Ruuskanen, 2000; Giesy et al., 2002). Based on the mechanism consideration, three-dimension (3D) molecular structural descriptors that describe electronic and steric properties of molecules were selected to describe the interaction between compounds and AhR, which was calculated using the DRAGON 2.1 (Mauri et al., 2006).

The molecular structures of the chemicals were modeled with CS Chem3D Ultra (Version 6.0), and were optimized using EF (eigenvector following), a geometry optimization procedure within MOPAC (2000, Cambridge, UK). Molecular descriptor meanings and their calculation procedure are summarized in the DRAGON software, and explained in detail, with related literature references, in the *Handbook of Molecular Descriptors* by Todeschini and Consonni (2000).

2.5. QSAR development and validation

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 (Chen et al., 2004). When Q^2_{CUM} of a model is larger than 0.5, the model is believed to have a good predictive ability (Golbraikh and Tropsha, 2002). 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 evaluated by external validation. The performance of external validation was characterized by the determination coefficient (R^2), root mean standard error ($RMSE$) and external explained variance (Q^2_{EXT}), which are defined as follows (Schüürmann et al., 2008):

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

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

$$Q^2_{EXT} = 1 - \frac{\sum_{i=1}^{n_{EXT}}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{EXT}}(y_i - \bar{y}_{EXT})^2} \tag{3}$$

where y_i^{fit} is the fitted pEC_{50} 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:

$$h_i = x_i^T(X^TX)^{-1}x_i \quad (i = 1, \ldots, n) \tag{4}$$

$$h^* = \frac{3(p+1)}{n} \tag{5}$$

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. Homologous modeling and molecular docking analysis

The LBD of modeled AhR contained five β-sheets and one α-helix, which was in accordance with previous investigations (Denison et al., 2002; Pandini et al., 2007). A Ramachandran plot from PROCHECK (Fig. 1) validated the reliability of the model by checking the stereo-chemical structure of each chain and the dihedral angel information. PROCHECK analysis also confirmed that no irrational conformation was observed for the alignment of various chains, bond length, bond angel and coplanar properties of the constructed AhR model.

TCDD was chosen to display the binding mode with AhR due to its high AhR affinity, and the key residues in the binding process were indicated in Fig. 2. The residues like Gys294, His331, Phe281, Gln377, Thr283, His285 and Phe289 were crucial for orientating and locating the ligand. Specifically, the residue Gln377, which was discovered only in AhR comparing with other PAS proteins, formed hydrogen bond with chlorine atom in TCDD. 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 TCDD with His285, Cys327, Ile319, Gln377 and Thr283, as shown in Fig. 2. The results are consistent with the findings of relevant reports (Procopio et al., 2002; Pandini et al., 2007). Therefore, the modeled AhR could be used for the following mechanism exploration.

3.2. Development and validation of the QSAR model for the pEC_{50}

Forward stepwise regression was adopted to screen molecular descriptors, then 6 descriptors (RG_{yr}, SE_{ig}, E_{1s}, E_{1p}, RDF_{065u} and Mor_{14u}) were finally selected for model development, which are listed in Table 1 with their physical–chemical meanings.

PLS regression with pEC_{50} as the dependent variable and the selected molecular structural parameters as predictor
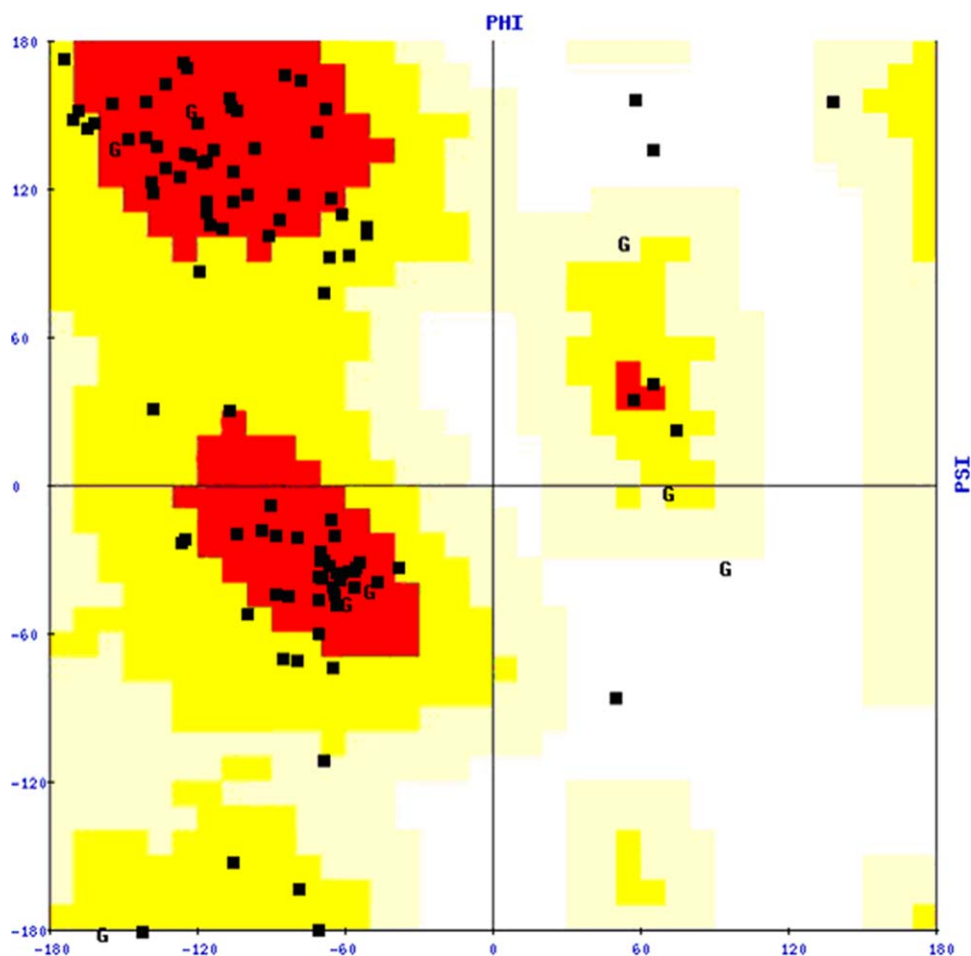
Fig. 1 – Ramachandran plot (the residues in the red area was the most reasonable, in yellow area was rational, whilst the residues in white area might be illogical). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)
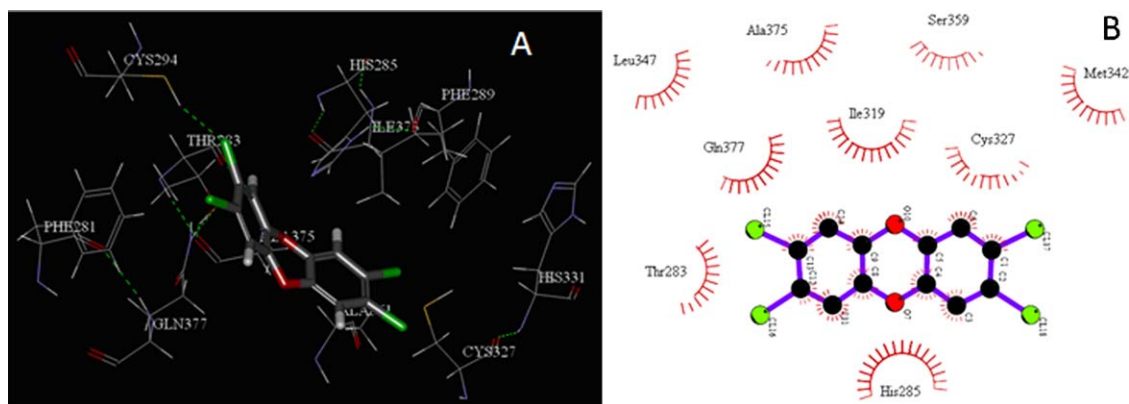
Fig. 2 – Docking views of TCDD in the binding site of AhR. (A) Green dotted line shows H-bonds between TCDD and basic groups. Carbon is colored in grey, oxygen red and nitrogen blue. (B) ●▬● : Ligand bond; ●▬● : non-ligand bond; ⌇ : non-ligand residues involved in hydrophobic contacts; ● : corresponding atoms involved in hydrophobic contacts. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

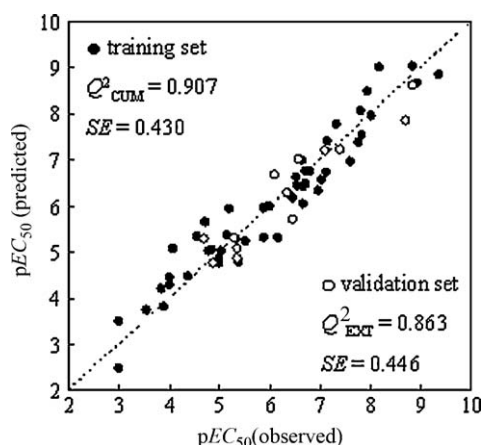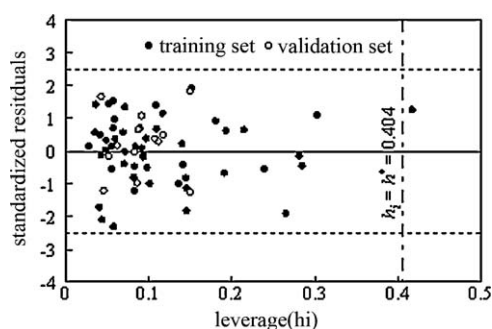**Fig. 3 – Plot of observed versus predicted pEC_{50} values for the training and validation.**

Fig. 4 – Plot of standardized residuals versus leverages. Dash lines represent ±2.5 standardized residual, dotted line represents warning leverage ($h^* = 0.404$).

variables resulted in the following optimal QSAR model:

$$pEC_{50} = 7.05 + 5.51E_{1p} + 0.633RG_{yr} - 0.331RDF_{065u}$$
$$- 0.610Mor_{14u} - 0.176SEig + 3.10E_{1s}$$

$$n\,(\text{training set}) = 52, \quad A = 3, \quad Q^2_{CUM} = 0.907, \quad R^2 = 0.922,$$
$$RMSE = 0.430\,(\text{training set}),$$

$$n\,(\text{validation set}) = 13, \quad Q^2_{EXT} = 0.863,$$
$$RMSE = 0.446\,(\text{validation set}), \quad p < 0.0001.$$

where A is the number of PLS components and p is the significance level.

The predicted pEC_{50} values and residuals for compounds are listed in Table 2 . The R^2 value of the QSAR model was 0.922, indicating a high goodness-of-fit of the model. Q^2_{CUM} of the QSAR was as high as 0.907, implying good robustness of the model. The differences between R^2 and Q^2_{CUM} (0.015) did not exceed 0.3, indicating no over-fitting in the model (Golbraikh and Tropsha, 2002). As shown in Fig. 3, the predicted pEC_{50} values were consistent with the observed values for both the validation and training sets. The model revealed acceptable predictability with $Q^2_{EXT} = 0.863$, $RMSE = 0.446$. 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 = 0.000, standard deviation = 0.434). 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. 4. As shown in the Williams plot (Fig. 4), h_i values of all the compounds in the training and validation sets were lower than the warning value ($h^* = 0.404$). Dibenzofurans in the training 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 2.5 standard deviation units (2.5σ). Thus there were no outliers for the developed QSAR model.

3.4. Mechanistic implications for the developed QSAR model

All the predictor variables, their VIP values that indicate the significance in explaining the variance of the dependent variable, and PLS weights (W^*) are listed in Table 3.

The first PLS component is loaded primarily on 4 descriptors, E_{1p}, RG_{yr}, RDF_{065u} and Mor_{14u}. E_{1p} belongs to the directional WHIM descriptors and is weighted by atomic polarizabilities (Todeschini and Consonni, 2000). E_{1p} remarkably governs pEC_{50}, as indicated by its VIP, the largest among all the predictor variables. RG_{yr} is a geometrical descriptor that is radius of gyration weighted by atomic mass. RDF_{065u} is a RDF descriptor, which could provide information about bond lengths, ring types, planar and nonplanar systems, atom types and molecular weight. Mor_{14u} belongs to 3D-MoRSE descriptors, which is the representation of the 3D structure of a molecule. E_{1p}, RG_{yr}, RDF_{065u} and Mor_{14u} relate to molecular size, and the PLS component mainly condenses information on the molecular size (volume). $W^*[1]$ and the coefficients in developed QSAR model indicated that E_{1p} and RG_{yr} were positively correlated with the pEC_{50} values, whilst the RDF_{065u} and Mor_{14u} were negatively correlated with the pEC_{50} values.

The second PLS components extract two descriptors, RDF_{065u} and SE_{ig}. SE_{ig} is absolute eigenvalue sum on geometry matrix, being within the geometrical descriptors. $W^*[2]$ and the coefficients in the current QSAR model indicated the negative correlation between SE_{ig} and pEC_{50}. The third PLS component is mainly loaded on the descriptor Mor_{14u}, SE_{ig} and

Table 2 – Observed and predicted pEC$_{50}$ of the considered compounds and molecular structural parameters.

No.	Compounds	pEC$_{50}$			E_{1p}	E_{1s}	RDF$_{065u}$	Mor$_{14u}$	RG$_{yr}$	SE$_{ig}$
		Obs.	Pred.	Res.						
1	2,3,7,8-Tetrachlorodibenzo-p-dioxin	8.00	7.95	0.05	0.76	0.80	2.51	−0.25	5.39	48.40
2	1,2,3,7,8-Pentachlorodibenzo-p-dioxin[a]	7.10	7.19	−0.08	0.74	0.76	3.65	−0.32	5.38	49.23
3	2,3,6,7-Tetrachlorodibenzo-p-dioxin	6.80	6.75	0.05	0.69	0.68	3.21	−0.24	5.16	48.53
4	2,3,6-Trichlorodibenzo-p-dioxin	6.66	6.05	0.61	0.62	0.54	2.60	−0.22	4.86	47.90
5	1,2,3,4,7,8-Hexachlorodibenzo-p-dioxin	6.55	6.42	0.13	0.71	0.72	4.71	−0.35	5.37	49.95
6	1,3,7,8-Tetrachlorodibenzo-p-dioxin[a]	6.10	6.67	−0.57	0.68	0.65	3.21	−0.37	5.16	48.66
7	1,2,4,7,8-Pentachlorodibenzo-p-dioxin	5.96	5.98	−0.02	0.66	0.62	4.44	−0.42	5.18	49.32
8	1,2,3,4-Tetrachlorodibenzo-p-dioxin	5.89	5.31	0.57	0.59	0.50	3.74	−0.19	4.90	48.39
9	2,3,7-Trichlorodibenzo-p-dioxin	7.15	7.39	−0.25	0.71	0.69	1.90	−0.20	5.14	47.88
10	1,2,3,4,7-Pentachlorodibenzo-p-dioxin	5.19	5.94	−0.75	0.65	0.62	4.36	−0.31	5.17	49.18
11	1,2,4-Trichlorodibenzo-p-dioxin[a]	4.89	4.75	0.14	0.53	0.40	3.22	−0.26	4.58	47.86
12	1,2,3,4,6,7,8,9-Octachlorodibenzo-p-dioxin	5.00	4.74	0.26	0.66	0.65	7.77	−0.45	5.37	51.42
13	1-Chlorodibenzo-p-dioxin	4.00	4.29	−0.29	0.47	0.28	1.47	−0.04	3.90	46.56
14	2,3,7,8-Tetrabromodibenzo-p-dioxin	8.82	9.03	−0.21	0.95	0.66	3.00	−0.83	6.24	49.53
15	2,3-Dibromo-7,8-dichlorodibenzo-p-dioxin[a]	8.83	8.61	0.22	0.86	0.72	2.61	−0.52	5.91	48.97
16	2,8-Dibromo-3,7-dichlorodibenzo-p-dioxin	9.35	8.86	0.49	0.86	0.73	1.89	−0.51	5.91	48.93
17	2-Bromo-3,7,8-trichlorodibenzo-p-dioxin	7.94	8.47	−0.53	0.81	0.76	2.10	−0.38	5.68	48.64
18	1,3,7,9-Tetrabromodibenzo-p-dioxin	7.03	6.56	0.47	0.66	0.44	2.28	−0.89	5.53	49.99
19	1,3,7,8-Tetrabromodibenzo-p-dioxin[a]	8.70	7.86	0.84	0.80	0.55	2.41	−0.87	5.89	49.82
20	1,2,4,7,8-Pentabromodibenzo-p-dioxin	7.77	7.38	0.39	0.77	0.54	2.87	−0.90	5.85	50.56
21	1,2,3,7,8-Pentabromodibenzo-p-dioxin	8.18	9.02	−0.84	0.90	0.64	1.94	−1.04	6.13	50.38
22	2,3,7-Tribromodibenzo-p-dioxin	8.93	8.65	0.28	0.87	0.58	1.71	−0.61	5.98	48.60
23	2,7-Dibromodibenzo-p-dioxin	7.81	8.06	−0.25	0.78	0.50	0.57	−0.41	5.62	47.92
24	2-Bromodibenzo-p-dioxin	6.53	6.61	−0.08	0.63	0.38	0.31	−0.17	4.98	46.89
25	2-Chlorodibenzofuran	3.55	3.74	−0.19	0.51	0.42	5.83	0.12	3.99	45.16
26	3-Chlorodibenzofuran	4.38	4.47	−0.09	0.53	0.46	4.67	0.09	4.11	45.16
27	4-Chlorodibenzofuran	3.00	3.49	−0.49	0.46	0.34	4.66	0.09	3.79	45.12
28	2,3-Dichlorodibenzofuran	5.33	5.27	0.06	0.59	0.58	4.78	0.09	4.48	45.72
29	1,3,6-Trichlorodibenzofuran[a]	5.36	4.87	0.49	0.51	0.46	3.40	−0.21	4.35	46.46
30	1,3,8-Trichlorodibenzofuran	4.07	5.07	−1.00	0.55	0.53	4.18	−0.18	4.47	46.63
31	2,3,4-Trichlorodibenzofuran	4.72	5.63	−0.91	0.59	0.58	3.71	0.06	4.61	46.25
32	2,3,8-Trichlorodibenzofuran	6.00	5.99	0.01	0.65	0.69	4.91	−0.02	4.76	46.29
33	2,6,7-Trichlorodibenzofuran[a]	6.35	6.28	0.07	0.63	0.68	3.52	−0.05	4.64	46.27
34	2,3,4,6-Tetrachlorodibenzofuran	6.46	6.16	0.29	0.60	0.60	2.44	−0.02	4.73	46.77
35	2,3,4,8-Tetrachlorodibenzofuran	6.70	6.49	0.21	0.65	0.71	3.63	−0.04	4.85	46.81
36	2,3,7,8-Tetrachlorodibenzofuran[a]	7.39	7.22	0.17	0.72	0.83	3.90	−0.04	5.03	46.84
37	1,2,4,8-Tetrachlorodibenzofuran	5.00	4.84	0.16	0.54	0.48	4.04	−0.13	4.53	47.01
38	1,2,4,7,9-Pentachlorodibenzofuran[a]	4.70	5.29	−0.59	0.52	0.48	1.97	−0.11	4.60	47.96
39	1,2,3,7,8-Pentachlorodibenzofuran	7.13	6.71	0.41	0.67	0.74	3.27	−0.09	4.95	47.80
40	1,2,4,7,8-Pentachlorodibenzofuran	5.89	5.95	−0.06	0.60	0.61	3.23	−0.17	4.80	47.70
41	2,3,4,7,8-Pentachlorodibenzofuran	7.82	7.54	0.28	0.71	0.82	2.68	−0.06	5.08	47.35
42	1,2,3,4,7,8-Hexchlorodibenzofuran	6.64	7.00	−0.36	0.66	0.73	2.12	−0.11	5.02	48.29
43	1,2,3,6,7,8-Hexchlorodibenzofuran[a]	6.57	7.02	−0.45	0.67	0.76	2.43	−0.11	5.02	48.29
44	2,3,4,6,7,8-Hexchlorodibenzofuran	7.33	7.77	−0.44	0.70	0.80	1.34	−0.04	5.12	47.86
45	2,3,6,8-Tetrachlorodibenzofuran	6.66	6.42	0.24	0.65	0.69	3.82	−0.14	4.86	46.82
46	1,2,3,6-Tetrachlorodibenzofuran[a]	6.46	5.69	0.77	0.58	0.60	3.46	−0.10	4.60	46.96
47	1,2,3,7-Tetrachlorodibenzofuran	6.96	6.33	0.62	0.63	0.69	3.35	−0.12	4.78	47.20
48	1,3,4,7,8-Pentachlorodibenzofuran	6.70	6.76	−0.06	0.63	0.69	2.06	−0.20	4.87	47.77
49	2,3,4,7,9-Pentachlorodibenzofuran	6.70	6.46	0.24	0.62	0.66	2.52	−0.20	4.86	47.79
50	1,2,3,7,9-Pentachlorodibenzofuran	6.40	6.25	0.15	0.59	0.61	1.64	−0.09	4.76	48.07
51	Dibenzofuran	3.00	2.46	0.54	0.43	0.26	5.72	0.24	3.34	44.57
52	2,3,4,7-Tetrachlorobiphenyl	7.60	6.95	0.65	0.66	0.73	2.83	−0.07	4.91	46.81
53	1,2,4,6,8-Pentachlorobiphenyl	5.51	5.21	0.29	0.54	0.50	3.17	−0.20	4.65	47.53
54	3,3′,4,4′-Tetrachlorobiphenyl	6.15	5.31	0.83	0.69	0.90	9.04	0.11	5.23	48.82
55	3,4,4′,5-Tetrachlorobiphenyl	4.55	5.34	−0.79	0.69	0.90	8.96	0.08	5.23	48.78
56	2′,3,4,4′,5-Pentachlorobiphenyl	4.85	5.03	−0.18	0.64	0.78	7.34	0.01	5.13	49.89
57	2,3,3′,4,4′-Pentachlorobiphenyl[a]	5.37	5.08	0.29	0.65	0.81	7.58	0.15	5.09	49.11
58	2,3′,4,4′,5-Pentachlorobiphenyl	5.04	5.01	0.03	0.65	0.82	8.13	0.06	5.10	49.11
59	2,3,4,4′,5-Pentachlorobiphenyl	5.39	4.79	0.60	0.64	0.78	8.12	0.14	5.09	49.04
60	2,3,3′,4,4′,5-Hexachlorobiphenyl	5.15	5.36	−0.21	0.66	0.83	7.24	0.11	5.21	49.65
61	2,3′,4,4′,5,5′-Hexachlorobiphenyl	4.80	5.03	−0.23	0.66	0.82	8.27	0.02	5.21	49.69

Table 2 (Continued)

No.	Compounds	pEC$_{50}$			E$_{1p}$	E$_{1s}$	RDF$_{065u}$	Mor$_{14u}$	RG$_{yr}$	SE$_{ig}$
		Obs.	Pred.	Res.						
62	2,3,3′,4,4′,5′-Hexachlorobiphenyl[a]	5.30	5.32	−0.02	0.67	0.85	7.60	0.02	5.25	50.44
63	2,2′,4,4′-Tetrachlorobiphenyl	3.89	3.80	0.08	0.54	0.59	6.83	0.26	4.67	48.71
64	2,3,4,5-Tetrachlorobiphenyl	3.85	4.21	−0.36	0.56	0.62	6.87	0.22	4.79	48.23
65	2,3′,4,4′,5′,6-Hexachlorobiphenyl	4.00	4.45	−0.44	0.58	0.68	6.65	0.14	4.97	49.95

[a] Compounds in the validation set.

Table 3 – VIP values and PLS weights.

	VIP	W*[1]	W*[2]	W*[3]
E$_{1p}$	1.231	0.549	0.177	0.236
RG$_{yr}$	1.118	0.494	−0.075	0.041
RDF$_{065u}$	1.091	−0.403	−0.787	−0.183
Mor$_{14u}$	1.020	−0.442	0.035	0.494
SE$_{ig}$	0.768	0.212	−0.631	−0.406
E$_{1s}$	0.644	0.228	0.064	0.865

E_{1s}. E_{1s} is a WHIM descriptor and weighted by atomic electropological states. The positive W*[3] and coefficient of E_{1s} in the QSAR model indicated the positive correlation between E_{1s} and pEC$_{50}$. In general, the current QSAR model indicated the pEC$_{50}$ value was related to molecular size, shape profiles and reactivity parameters such as polarizabilities and electropological states.

4. Conclusions

The 3D crystal structure of AhR was homologously modeled, and docking analysis showed that hydrogen bonding and hydrophobic interactions between compounds and AhR governed the binding affinities. Based on the mechanism of interactions, a QSAR was established to characterize the interactions and to model the relative binding affinity for PCBs, PCDDs, and PCDFs. Molecular size, shape profiles, polarizability and electropological were important factors for the binding interactions between compounds and AhR. The developed QSAR model had good robustness, predictive ability and mechanism interpretability, which could be applied to predict the binding affinity of other compounds.

Conflict of interest statement

Nothing declared.

Acknowledgments

REFERENCES

Bandiera, S., Sawyer, T.W., Campbell, M.A., Fujita, T., Safe, S., 1983. Competitive-binding to the cytosolic 2,3,7,8-tetrachlorodibenzo-para-dioxin receptor – effects of structure on the affinities of substituted halogenated biphenyls – a QSAR analysis. Biochem. Pharmacol. 32, 3803–3813.

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.

Chovancova, J., Kocan, A., Jursa, S., 2005. PCDDs, PCDFs and dioxin-like PCBs in food of animal origin (Slovakia). Chemosphere 61, 1305–1311.

Denison, M.S., Pandini, A., Nagy, S.R., Baldwin, E.P., Bonati, L., 2002. Ligand binding and activation of the Ah receptor. Chem-Biol. Interact. 141, 3–24.

Domingo, J.L., Bocio, A., 2007. Levels of PCDD/PCDFs and PCBs in edible marine species and human intake: a literature review. Environ. Int. 33, 397–405.

European Commission, Available online at:, http://ecb.jrc.it/REACH/, 2002.

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.

Giesy, J.P., Hilscherova, K., Jones, P.D., Kannan, K., Machala, M., 2002. Cell bioassays for detection of aryl hydrocarbon (AhR) and estrogen receptor (ER) mediated activity in environmental samples. Mar. Pollut. Bull. 45, 3–16.

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

Hilscherova, K., Machala, M., Kannan, K., Blankenship, A.L., Giesy, J.P., 2000. Cell bioassays for detection of aryl hydrocarbon (AhR) and estrogen receptor (ER) mediated activity in environmental samples. Environ. Sci. Pollut. R. 7, 159–171.

Landers, J.P., Bunce, N.J., 1991. The Ah receptor and the mechanism of dioxin toxicity. Biochem. J. 276, 273–287.

Li, F., Li, X.H., Shao, J.P., Chi, P., Chen, J.W., Wang, Z.J., 2010a. Estrogenic activity of anthraquinone derivatives: in vitro and in silico studies. Chem. Res. Toxicol. 23, 1349–1355.

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

Lucier, G.W., Portier, C.J., Gallo, M.A., 1993. Receptor mechanisms and dose–response models for the effects of dioxins. Environ. Health Perspect. 101, 36–44.

Mandal, P.K., 2005. Dioxin: a review of its environmental effects and its aryl hydrocarbon receptor biology. J. Comp. Physiol. B: Biochem. Syst. Environ. Physiol. 175, 221–230.

Mauri, A., Consonni, V., Pavan, M., Todeschini, R., 2006. Dragon software: an easy approach to molecular descriptor calculations. Match-Commun. Math. Co. 56, 237–248.

Nebert, D.W., Puga, A., Vasiliou, V., 1993. Role of the Ah receptor and the dioxin-inducible [Ah] gene battery in toxicity, cancer, and signal-transduction. Ann. N.Y. Acad. Sci. 685, 624–640.

Nie, M.H., Blankenship, A.L., Giesy, J.P., 2001. Interactions between aryl hydrocarbon receptor (AhR) and hypoxia signaling pathways. Environ. Toxicol. Pharmacol. 10, 17–27.

Niu, J.F., Long, X.X., Shi, S.Q., 2007. Quantitative structure–activity relationships for prediction of the toxicity of hydroxylated and quinoid PCB metabolites. J. Mol. Model. 13, 163–169.

OECD, Available online at:, http://appli1.oecd.org/olis/2007doc. nsf/linkto/env-jm-mono(2007)2, 2007.

Ohura, T., Morita, M., Kuruto-Niwa, R., Amagai, T., Sakakibara, H., Shimoi, K., 2010. Differential action of chlorinated polycyclic aromatic hydrocarbons on aryl hydrocarbon receptor-mediated signaling in breast cancer cells. Environ. Toxicol. 25, 180–187.

Okey, A.B., Riddick, D.S., Harper, P.A., 1994. The Ah receptor – mediator of the toxicity of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) and related-compounds. Toxicol. Lett. 70, 1–22.

Olivero-Verbel, J., Vivas-Reyes, R., Pacheco-Londono, L., Johnson-Restrepo, B., Kannan, K., 2004. Discriminant analysis for activation of the aryl hydrocarbon receptor by polychlorinated naphthalenes. J. Mol. Struc.-Theochem. 678, 157–161.

Pandini, A., Denison, M.S., Song, Y.J., Soshilov, A.A., Bonati, L., 2007. Structural and functional characterization of the aryl hydrocarbon receptor ligand binding domain by homology modeling and mutational analysis. Biochemistry 46, 696–708.

Poland, A., Knutson, J.C., 1982. 2,3,7,8-Tetrachlorodibenzo-para-dioxin and related halogenated aromatic-hydrocarbons – examination of the mechanism of toxicity. Annu. Rev. Pharmacol. 22, 517–554.

Procopio, M., Lahm, A., Tramontano, A., Bonati, L., Pitea, D., 2002. A model for recognition of polychlorinated dibenzo-p-dioxins by the aryl hydrocarbon receptor. Eur. J. Biochem. 269, 13–18.

Safe, S., 1990. Polychlorinated-biphenyls (PCBs), dibenzo-para-dioxins (PCDDs), dibenzofurans (PCDFs), and related-compounds – environmental and mechanistic considerations which support the development of toxic equivalency factors (TEFs). Crit. Rev. Toxicol. 21, 51–88.

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.

So, S.S., Karplus, M., 1997. Three-dimensional quantitative structure–activity relationships from molecular similarity matrices and genetic neural networks. 2. Applications. J. Med. Chem. 40, 4360–4371.

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

Tuppurainen, K., Ruuskanen, J., 2000. Electronic eigenvalue (EEVA): a new QSAR/QSPR descriptor for electronic substituent effects based on molecular orbital energies. A QSAR approach to the Ah receptor binding affinity of polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs). Chemosphere 41, 843–848.

Villeneuve, D.L., Khim, J.S., Kannan, K., Giesy, J.P., 2002. Relative potencies of individual polycyclic aromatic hydrocarbons to induce dioxinlike and estrogenic responses in three cell lines. Environ. Toxicol. 17, 128–137.

Wagener, M., Sadowski, J., Gasteiger, J., 1995. Autocorrelation of molecular-surface properties for modeling corticosteroid-binding globulin and cytosolic Ah receptor activity by neural networks. J. Am. Chem. Soc. 117, 7769–7775.

Waller, C.L., Mckinney, J.D., 1992. Comparative molecular-field analysis of polyhalogenated dibenzo-p-dioxins, dibenzofurans, and biphenyls. J. Med. Chem. 35, 3660–3666.

Wang, Y., Chen, J.W., Lin, J., Wang, Z., Bian, H.T., Cai, X.Y., Hao, C., 2009a. Combined experimental and theoretical study on photoinduced toxicity of an anthraquinone dye intermediate to *Daphnia magna*. Environ. Toxicol. Chem. 28, 846–852.

Wang, Y., Chen, J.W., Ge, L.K., Wang, D.G., Cai, X.Y., Huang, L.P., Hao, C., 2009b. Experimental and theoretical studies on the photoinduced acute toxicity of a series of anthraquinone derivatives towards the water flea (*Daphnia magna*). Dyes Pigments 83, 276–280.

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

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.

Zhao, Y.Y., Tao, F.M., Zeng, E.Y., 2008. Theoretical study of the quantitative structure–activity relationships for the toxicity of dibenzo-p-dioxins. Chemosphere 73, 86–91.